UNIVERSITE PARIS-SUD
FACULTE DES SCIENCES D'ORSAY

Présentée pour obtenir
LE GRADE DE DOCTEUR EN SCIENCES
DE L'UNIVERSITE PARIS XI

Spécialité : Mathématiques

par

Thi Thu Huong HOANG

MODELISATION DE SERIES CHRONOLOGIQUES NON STATIONNAIRES, NON LINEAIRES

Application à la définition des tendances sur la moyenne, la variabilité et les extrêmes de la température de l'air en Europe

Soutenue le 28 Janvier 2010 devant la commission d'examen :

Mme. Elisabeth Gassiat	(Président du jury)
M. Didier Dacunha-Castelle	(Directeur de thèse)
Mme. Sylvie Parey	(Responsable d'EDF)
M. Anestis Antoniadis	(Rapporteur)
Mme. Anne-Laure Fougères	(Rapporteur)
M. Grégory Benmenzer	

Table des matières

1 Introduction **3**

2 Observed versus model produced data of temperature and warming effect **15**
2.1 Introduction 16
2.2 State of the art 17
2.3 The data used 19
2.3.1 ECA & D series 19
2.3.2 European ENSEMBLES project gridded dataset . . . 21
2.3.3 ERA40 reanalysis 21
2.3.4 Simulation models 23

3 Nonparametric estimation **24**
3.1 Introduction 25
3.2 Application of loess 27
3.2.1 Estimation of mean trend m 28
3.2.2 Estimation of variance trend s^2 32
3.2.3 Bootstrap confidence intervals 35
3.3 Automatic selection of the smoothing parameters for correlated data 36
3.3.1 The existing automatic selections in the presence of dependence 38
3.3.1.1 Modified cross-validation (MCV) 39
3.3.1.2 Partitioned cross-validation (PCV) 40
3.3.1.3 Block bootstrap (BB) 40
3.3.1.4 Plug-in method 41
3.3.2 A new algorithm : Modified partitioned cross-validation (MPCV) 42
3.3.3 Simulation study 45
3.3.3.1 Simulation study for all methods 45
3.3.3.2 Detailed simulation study on MPCV 47
3.3.3.3 Application of MPCV to the temperature series 51

3.3.4 Conclusion on the new algorithm . . . 52
3.4 Application of splines . . . 52
3.4.1 The estimators . . . 53
3.4.2 Asymptotic properties . . . 55
3.4.2.1 Convergence of the backfitting procedure . . 55
3.4.3 Estimation algorithm for multi-parameters with bandwidth selection . . . 57
3.4.3.1 Iteratively reweighted least squares . . . 58
3.4.3.2 Iterative cross-validation . . . 58
3.4.3.3 A new algorithm . . . 59
3.4.3.4 Simulation study . . . 60

4 Mean and variance evolutions of the hot and cold temperatures in Europe 63
4.1 Introduction . . . 64
4.2 Presentation of the used data series . . . 65
4.2.1 ECA & D series . . . 65
4.2.2 European ENSEMBLES project gridded dataset . . . 66
4.2.3 ERA40 reanalysis . . . 66
4.3 Statistical tools . . . 66
4.3.1 Trend definition . . . 66
4.3.2 Trend estimation . . . 67
4.3.2.1 Estimation methods . . . 67
4.3.2.2 Bootstrap confidence intervals . . . 69
4.3.3 Possible link between the estimated trends $m(t)$ and $s^2(t)$ . . . 70
4.4 Results for the different observational datasets . . . 70
4.4.1 ECA & D temperature series . . . 70
4.4.1.1 ENSEMBLES gridded dataset . . . 71
4.4.2 The ERA40 reanalysis . . . 72
4.4.3 Results summary . . . 72
4.4.3.1 Cold season . . . 72
4.4.3.2 Hot season . . . 73
4.5 Results for 6 European climate models . . . 73
4.6 Conclusion and perspectives . . . 74

5 Non stationary extreme models 87
5.1 Introduction . . . 88
5.2 Results from probability theory . . . 89
5.2.1 Basic extreme theory for i.i.d sequences . . . 89
5.2.2 Extreme for stationary sequences with correlation . . . 92
5.2.2.1 The extremal index θ for GEV . . . 94
5.2.3 Non-stationary EVT models . . . 95
5.3 Statistics of extremes . . . 96

5.3.1 Basic methods for i.i.d data 96
5.3.2 Estimation methods for the non-stationary case 98
5.3.2.1 Parametric estimation 100
5.3.2.2 Nonparametric or semiparametric estimation 101
5.3.3 Confidence intervals 103
5.3.4 Application to the series of temperature 104
5.3.4.1 Discussion on the behaviour of the shape parameter ξ 104
5.3.4.2 Estimation of extreme parameters 105

6 A new test of stationarity. Links between central field and extreme trends 114
6.1 Introduction 115
6.2 Hypothesis K : stationarity of extremes of reduced data . . . 117
6.3 A new test for stationarity. Application to extreme models . 119
6.3.1 Asymptotic theory of the test 119
6.3.2 Numerical study of the test 125
6.4 Reconstruction and departures from stationarity in the tails of Y 126
6.5 Results of application to data series of Europe 127
6.5.1 Links between trends in central field and in extremes . 127
6.6 Conclusion 132

7 Multidimensional trends: the example of temperature 134
7.1 Introduction 135
7.2 Statistical framework 137
7.2.1 Trend in mean 137
7.2.2 Trend in variance 138
7.2.3 Trend in extremes 139
7.2.4 Measure of the similarity of trends 140
7.3 Results for the European temperature series 141
7.3.1 Mean and variance trends 141
7.3.2 The residuals 143
7.3.3 Trend in extremes 145
7.4 Trends in extremes and trends in mean and variance 145
7.4.1 The K hypothesis 145
7.4.2 How to test K ? 148
7.5 Conclusion 150

8 Different ways to compute temperature Return Levels in the climate change context 151
8.1 Introduction 152
8.2 Mathematical tools 154
8.2.1 General framework 154

8.2.2 Trends in the parameters of the extreme value distributions . . . 155
8.2.2.1 Trend identification . . . 155
8.2.2.2 Return levels in a non stationary context . . 156
8.2.2.3 Associated confidence intervals . . . 157
8.2.3 Link between trend in extremes and trends in the whole series : use of the K hypothesis . . . 159
8.2.3.1 Non parametric tools to derive trends . . . 159
8.2.3.2 Link with the trends in extremes . . . 159
8.2.3.3 The use of this link to compute return levels and their associated confidence intervals . . . 160
8.3 Example : extrapolation of trends in extreme value distribution parameters . . . 161
8.3.1 The block maxima method . . . 162
8.3.2 The POT method . . . 165
8.3.2.1 Trend extrapolation . . . 165
8.3.2.2 Effect of a change in the location parameter of the GEV or in the intensity of the Poisson process . . . 167
8.3.3 Research of a break point in the evolution of the mean 168
8.4 Example : use of the K hypothesis . . . 170
8.4.1 The K hypothesis . . . 170
8.4.2 Non stationary return levels . . . 171
8.4.3 Stationary return levels . . . 172
8.4.3.1 Mean and variance trends extrapolation . . . 173
8.4.3.2 Use of climate model results . . . 173
8.5 Discussion and perspectives . . . 174

9 Seasonality dynamics with warming effect 187
9.1 The seasonality in statistical study . . . 188
9.2 Transformation of the seasonality under global growing of the temperature . . . 189
9.3 How to detect the seasonality from the time series ? . . . 192
9.4 Conclusion . . . 195

10 From continuous to discrete-time models taking into account extremes 196
10.1 Introduction . . . 197
10.2 Theoretical results on continuous-time diffusions and their discrete time approximations . . . 200
10.2.1 Extremes of diffusion with inaccessible finite boundaries 200
10.2.1.1 Main definitions and Berman result . . . 201

10.2.1.2 Bounded case 202
10.2.2 Diffusion approximations for discrete samples 204
10.2.2.1 Discretization, imbedding and limits 205
10.2.2.2 The exact discretization and their limits . . . 205
10.2.3 Approximate discretizations and their statistics . . . 207
10.2.4 The estimate procedure with the constraints of extremes 209
10.2.4.1 Estimation of the drift coefficient 209
10.2.4.2 Estimation of diffusion coefficient 210
10.2.4.2.1 Non-parametric estimates 210
10.2.4.3 Parametric estimate 211
10.2.5 Simulation study on the properties of the model with a bounded a 212
10.2.5.1 Study on the properties of Model (10.27) . . 213
10.2.5.2 Difference in the extreme field with an Orstein-Ulhenbeck model 216
10.2.5.3 Discretization properties of Model (10.27) with bounded a 219
10.2.5.4 Example 1 220
10.2.5.5 Example 2 221
10.2.6 Example 3 221
10.2.6.1 Discussion 223

11 Modeling observed temperature series, criteria of validation, simulations 228
11.1 Introduction 229
11.2 Reduced temperatures and their characteristics 231
11.2.1 How obtain the reduced series ? 231
11.2.2 Seasonality effect on the reduced series 232
11.3 Modeling reduced series of air temperatures 237
11.3.0.1 Validation of the model and comparison with simpler other models 240
11.3.0.2 Modeling on daily fixed-hour temperatures . 256
11.3.1 Modeling on daily mean, maximum and minimum temperatures 259
11.4 Conclusion 261

Chapitre 1

Introduction

In this thesis, we address some statistical problems linked with the warming of atmospheric temperatures. The significant increase in their average during at least the last thirty years, is today considered to be one of the most important problem for humanity. This observed temperature increase is probably caused by increasing concentrations of greenhouse gases due to human activities. This phenomenon can lead to serious, both short-termed and long-termed, consequences for our planet. "An increase in global temperature will cause sea levels to rise and will change the amount, and pattern, of precipitation, probably including expansion of subtropical deserts" (Lu et al., [94]). Other likely effects include increases in the intensity of extreme weather events, species extinctions, and changes in agricultural yields. The problem of extreme events is therefore a strong preoccupation. Heatwaves, storms, or hurricanes, have become a common problem to many countries.

Besides these general topics, the security of industrial installations are considered in this new framework. Cooling size has to be designed for the 100 future years taking into account the prediction of the temperature. This is in particular the case for existing and forthcoming nuclear plants. Their safety not only depends on the cooling of their own installations but also on ecological problems, such as the temperature of rivers and seas. As often, the problem of large risks presents a meta-mathematic part. It involves the existence of scenarios for the emission of greenhouse gases but absolutely needs a probabilistic approach and a reflection on the statistical methods. This last point needs to be more specifically developed. It allows to clarify some discussions, and some interpretations, about different predictions at middle and long horizon.

Facing up to these problems, climate numerical models are used to make predictions. Numerical models of the atmospheric circulation are the main tools in climatology. These models of fluid mechanics are based on coupled PDE (partial differential equations). Sciences in this field are of course in progress, however, some important difficulties still remain, as the problem of clouds representation for instance. Models are deterministic. Variability is approached in two different ways. The first is due to the existence of numerous models with quite different results. This variability is taken into account in the GIEC work and doesn't concern mathematics or even sciences. The second comes from simulations. An often small number of trajectories is obtained from small variations of initial conditions. The pseudo-randomness comes from the chaotic properties of these dynamical systems.

The simulation models are continually improving, but for some particular points, these models still present some weaknesses. Concretely, they give an inadequate representation of the variability of local temperatures for the models are defined for a large spatial scale. To deduce the temperature in

some precise points, some statistical works on the observed temperature are necessary in order to make corrections on model data. Besides the variability of the data, it is commonly admitted that extreme values are not correctly approached in simulation models. The methods concerning the problems of confidence intervals still need to be significantly improved. Some of these points will be detailed in Chapter 2.

This thesis work will then focus on the understanding of the recent evolutions of extreme temperatures, related to the evolutions of the mean and the variance. The aim is then to propose a methodology to infer the future extreme levels, and on the other hand, to statistically generate coherent temperature series in the climate change context. Our work is mainly a work on a very specific kind of time series. Some important features of temperature series, apparently unknown until today, have been discovered in this direction and published in different climate journals. These results are useful from the climatologist's point of view. They can be considered as one criterion to check the goodness of numerical models. These specific properties of observed temperatures have been checked on some data produced by numerical models.

We mainly study series of observed temperatures. On the other hand, some works in this thesis are linked with reanalysis data and data from simulation models. We discuss in Chapter 2 the nature of the observed data that we will use in our works (source of the observations as well as models, length and quality of the series, geographic distribution, etc...).

The length of data is a central parameter in order to speak about "temperature increase". The increase needs to refer to a precise period. Everyone knows that there are "natural pseudo cycles" of temperatures as those linked to sun activity and we are mainly interested in the recent, very probably anthropic effect due to greenhouse gas emissions. We have worked mainly on the more or less last fifty years but also on longer series in order to clarify results as well as methods.

A statistical, quite rigorous, approach of the notion of "trends" is the main goal of Chapter 3 on non-parametric methods. As said, "trends" are not an intrinsic notion if the length of the observations is not fixed. Once the length is fixed, it is possible to search for an intrinsic definition of trends, not only on the intuitive mean but also, for instance, on the variability. In any case, trend definition or computation have to respect an heuristic principle : the latter evolution and the addition of new data must not change the trends already defined.

Trends are, physically, a slow long-term deterministic movement, some-

thing related to low frequencies of a signal. But pay attention, trends usually do not contain seasonal cycles. These cycles have not a simple definition in climate. The facts are that their own characteristics also have variations at a similar speed as the trend on mean temperature for instance.

So the definition implies an intrinsic definition of smoothness (roughness of a curve extracted or estimated from discrete data). This smoothness is from a very important statistical interpretation linked to some scales of time. We present the tools of non-parametric statistic which can be used for our specific time series evolutions. Our basic model, more complicated than almost all model classically studied in theoretical non parametric analysis is of the following form :

$$X_t = m(t) + S(t) + S_v(t)s(t)Z_t \tag{1.1}$$

where t is the date, X_t the observations, $m(t)$ the trend in mean, $S(t)$ the seasonal component and $S_v^2(t)$, $s^2(t)$ are respectively the seasonal variance and the long term trend in variance. What about Z_t ? It is a stochastic process, centred, normed, supposed or hoped to have at least a stationary seasonal covariance. "Stationary seasonal covariance" means that the covariance is a periodic function of the time interval between observations.

To begin, we have to estimate at least two trends m, s and two seasonalities S and S_v.

In Chapter 3, in order to avoid complicated characteristics of the seasonality, and only consider basic trends themselves, our study is limited to rather homogeneous temperature series, such as daily maximum temperatures in summer, or daily minimum ones in winter. Our model (1.1) then transforms into :

$$X_t = m(t) + s(t)Y_t \tag{1.2}$$

Y_t is also supposed to have at least a stationary covariance. Y is non-gaussian, its tails are light and in general of bounded support. For close (in time) observations, it appears to be a weakly dependent, near of an ARMA process but with seasonal coefficients, colored innovation and a complex structure of the variance. The problem is to know what statistical non-parametric methods can be used (perhaps can be "saved") in such a situation. Data can be thought as "long" series with from 5000 to 10000 observations in general.

The non-parametric literature is very large. It is often classified by the main method studied. There are roughly two kinds of methods : approximation by projection on a suitable basis as Fourier, Haar or wavelets, where the choice of model order gives a threshold for the dimension of the image

space ; and another kind of methods, called "regularization methods" as kernels, loess, lasso or splines. They use a tuning parameter (regularization parameter) to transform an ill-posed problem, as least squares or maximum likelihood, into a well-posed problem by adding a penalization term, which depends on the tuning parameter and the sample size. In fact, splines can be classified in the first family but is closer to the second one.

In our case, a logic and appropriate approach is to consider a trend as linked to an "appropriate local window". This bandwidth parameter has a significant sense in our context because it is clearly linked to the "memory" of the warming process. Loess (local polynomial estimation) is chosen because it presents many advantages. Asymptotically efficient estimations by loess for both trends in mean and in variance will be considered giving some theoretical arguments. The choice of the tuning parameter, traditionally, can be done using different criteria as cross-validation (CV) and many other criteria of model choice. Non asymptotic bounds for different risks are now known and asymptotic results are defined. But as the data are complex, the situation quickly becomes not simple. Non constant variance for the noise is already a problem. Moreover, the noise Y is correlated, which makes the situation even more difficult. Particularly, all intuitive extensions lead to important difficulties when the choice of the tuning parameter is addressed. For instance, CV has to be deeply modified to keep a sense. In Chapter 3, a study on the existing automatic selectors in the presence of correlated errors will be carried out through simulations. Our contribution consists in a new automatic algorithm for the choice of regularization parameters, named "modified partition cross-validation" (MPCV). It will be proposed in specific circumstances, where the datasets are long, the conditional variance is non constant, and the noise is colored with an unknown correlation function. Our method is asymptotically efficient and has advantages concerning computation times.

Still in Chapter 3, the estimation of the functional multi-parameters models is considered. The context is different than the previous one. Now, the law of the data is known, and the independence of the observations is supposed. However, we have the same purpose : describe the dynamics of the data through the parameters. Therefore, nonparametric estimations are also needed. Both approaches by loess and by spline are possible. In loess, the estimation of the parameters leads to maximize the local likelihood (LL), whereas in spline, it leads to maximize the penalized likelihood(PL). In this situation, the spline estimation is preferred, it shows its advantages in the flexibility and also the numerical implementation. Here the non-orthogonal (in the statistical sense) parameters in the distribution function once more set difficulties on the choice of smoothing parameters. Again, facing the lack of methods in the literature, we propose a new algorithm for an automa-

tic choice of smoothing parameters. This method is based on the iteratively weighted least squares and iterate cross validation of Gu ([61]). Its asymptotic performance is really complicated to show. We prefer to check it through a simulation study. This part prepares tools for chapter 5 which deals with non-stationary extreme models.

Chapter 4 is one of our articles. In this paper, studies on ordinary trends, $m(t)$ in mean and $s^2(t)$ in variance, of these homogeneous temperature series are realized. A strong dynamic relation between these deterministic trends, which can change with the considered season, is discovered. We measure this dynamic relation using different functional analysis tool and check it on a very large sample of stations in Europe. This relation is also checked on some simulation models and the results show that the models do not take into account this characteristic of the observed temperature.

Then chapter 5 is dedicated to extreme models in non stationary situation. The basic theory assumes that the series studied are stationary, they do not present any cycle nor trend. However, when dealing with climatic data, this assumption has to be carefully considered. An interesting subject is to study how climate changes might affect the occurrence of high and low temperatures. This leads us to non-stationary extreme models where the trends in samples of extreme values are estimated. We extend GEV and POT classical models to our situation. Let us consider only here GEV models. If we take one block of successive dates, say one month, long term variations are negligible on a such short period. Asymptotics provided by probability theory can be applied on the observations which are weakly dependent. Once the seasonal effects are eliminated, the distribution of the maxima in a given block depends on the label of the block and not on the trend effect into the block, which is negligible. So the maximum of a block can be supposed to have an approximate GEV distribution with parameters depending on the block label. Of course, as usually, we do not check if the regularity conditions required to apply the asymptotic theory are verified. The quality of this approximation is the same as in the stationary case, not worse ! In this context, the extreme parameters of the distribution of the maxima are functions of the date (or of the block label). An alternative algorithm to estimate the extreme parameters accompanied by an automatically updated choice of smoothing parameters is introduced, using the results of Chapter 3. To make more sense, the application of non-stationary extreme models are usually used on the daily maximum temperature in summer or the daily minimum temperature in winter.

Until now, the previous chapters, without taking into account the seasonality, give us complete features of the trends of the "central field", which corresponds to the mean and variance functions of the whole time series and

the trends of the extreme field, which in turn corresponds to the mean and variance functions of extreme values. It is then interesting if we can find links between these factors. Consider the H hypothesis, which states that the centered and scaled data, called reduced data through the thesis, $Y_t = \frac{X_t - m(t)}{s(t)}$, follows a stationary distribution. If H is valid, the trends in the extreme field would only be due to translation/scaling effects of the central field and it is easy to show that these trends can be computed only using the mean $m(t)$, the variance $s^2(t)$ and the constant parameters of the extreme model of Y_t. Instead of H, we can consider a weaker K hypothesis : the extreme model ext_Y of Y_t is stationary. This avoids taking into account some small dynamic deformations of the central part of the distribution of the reduced data whose physical interest is in general not evident. The K hypothesis is checked by a statistical test. The extreme parameters are first estimated as constants, noted $\hat{\theta}_1$, $\hat{\theta}_1$ is multidimensional. Then we estimate them nonparametrically and note the nonparametric estimator $\hat{\theta}_2$. Calculate some L_2 distance between two these estimators $\Delta = \parallel \hat{\theta}_1 - \hat{\theta}_2 \parallel$. In order to test if Δ is significantly small, which means that K is probably true, we built the table of Δ distribution using simulations (or bootstrap). From that we can calculate the $p - value$ of this test to accept, or not, the validation of K. In the non-valid case of K, Δ is shown to converge to some $C > 0$. The power of test is also derived in the same way. The theoretical asymptotic properties of this test will be proved. Then the K hypothesis will be checked on many stations in Europe. The results show a strong significance of K on the temperature series : more for the daily maximum temperature in summer and less for the daily minimum one in winter. All these contents are in Chapter 6.

Chapter 7 corresponds to one of our published article. This paper give the synthesis of all the previous chapters. It gives a general look on what we called "multidimensional trends". Multidimensional trends imply the trends in the central field and those in extremes. Multidimensional trends contain in themselves strong links between their trends, more or less strongly depending on stations. This relation is checked by the K hypothesis.

We finish the first part of the thesis : detection of trends in the central as well as extreme field and tests for the stationarity of extremes, by a very important application : the prediction of return levels. In Chapter 8, which is also a submit paper, we present the return levels in the non-stationary context. The definition of return levels in the non stationary context is extended from the one of the stationary case. To apply it in practice, we need some extrapolations of the model of temperature extremes. Extrapolation for non stationary cases is, of course, a metamathematical problem. What kind of "persistence" one has to choose ? Too often, the idea of extrapolating the estimated functional trends in the future time is used. This behaviour often leads to non physical sense for many parametric families as polynomials.

We propose some reflections about this problem using some new asymptotic theory.

A parametric approach is necessary for this extrapolation. The extreme parameters in this application can be estimated by polynomials of high degree or continuous piecewise linear functions (CPL). Modeling by linear piecewise functions needs to be associated to a specific mode of extrapolation, which is not necessarily the extension of the last period, in order to avoid edge effects. Dealing with the polynomial estimation, CPL, although without mathematical support, seems in practice to have the advantages to give coherent physical results. Nevertheless, these classical methods show their weakness, concretely, the "sampling effect". A very hot (or very cold) last period can affect seriously, in a negative way, the results of the return levels. In order to obtain more robust return levels, a new method is proposed. This method is based on the fact that K is usually valid on the temperature series. Through K, we can estimate the return levels by passing the elements of the central field ($m(t)$ and $s(t)$) which give the results much more stable and robust. The uncertainty of these estimates will be approached through bootstrap confidence intervals.

The next chapter opens a new paradigm : taking into account the seasonality, a temperature series is considered as a realization of model (1.1). The purpose is to model the "reduced series" Z_t. The first question one can put on the series Z_t is : is Z_t stationary (at least in some useful sense). This question is important for many reasons :

1- Validation of non-parametric estimations, done in previous chapters, required the second order stationary (stationarity of covariance) or at least some boundedness of the covariance. In fact this requirement is more concerned with theories than with applications.

2- Validation of a model, to be used in particular to build a simulation model, requires **the generation of the date by a stationary white noise** to be able to use bootstrap validation or simulation one.

3- Of course, non stationarity for initial non reduced data, can be extended without restriction to every characteristic as skewness, kurtosis, high quantiles, extremes. But this extension has to have a statistical meaning. Data are large but not huge and neither infinite. So non-stationary properties without any statistical evidence have, in this framework, no practical interest.

In our case, the stationarity needs to be extended to a broader sense : "periodic stationarity". In fact, the seasonality shape has deformations due

to the warming effect. We give a discussion on this seasonality change in Chapter 9. To catch this variable seasonality, without confusion with trend components, is not simple at all. Therefore, for instance, we satisfy ourselves with the traditional long term annual cycles, which are constant year after year, and accept that the seasonality still remains in the residuals. Effective treatments of this changing seasonality remain a very interesting work for the future.

We now skip to the second part of the thesis . We will study the random part, once the deterministic one has been extracted as the "main signal". The random part is the centred and scaled temperature.

To model this reduced temperature, which, as we shall see, has bounded support for its marginal densities, we need to develop some theoretical points in Chapter 10. The temperature, for continuous time, is of course a continuous process. For quite evident physical reasons, as detailed in the literature, it has the Markov property and so is of a diffusion type. From the boundedness of the distributions, this diffusion has inaccessible boundaries. It is not stationary, but only periodically stationary. The drift and the volatility are periodic functions. In fact, the boundedness of the distributions is provided by the study of extremes, only using extreme theory, without the need of a model of reduced series.

As usual, we only can observe the temperature process at discrete times of the form $k\Delta$, with Δ being the lag of observations. We get data for $\Delta = 3$ hours. We do not have continuous filter observations of the form, for instance, $\int K(t-x)T_x dx$ with K compact. Discrete observations are checked as being Markovian series. We also study functions of these observations, as daily maximum or minimum, which can lose this markovian property, as mean daily temperature do in practice. Now the discrete process, skeleton of the continuous one, for equispaced observations, is a Markov chain, periodically stationary and observed at dates $k\Delta$. The exact likelihood is very complicated and one has to use an approximated likelihood for Δ near zero, which also appears too complicated in our situation, and nobody knows if Δ is small or not ! So we have to look for a simpler, but of course, also less accurate, model.

We use approximations by first order Euler schemes. In the case where we neglect the seasonality, they have the form :

$$Z_n = b(Z_{n-1}) + a(Z_{n-1})\varepsilon_n \tag{1.3}$$

These processes are sometimes called functional autoregressive processes with conditional heteroscedasticity : FARCH.

The next step consists in estimating the drift and the diffusion coefficient of these processes and to precise their properties. First we give new results on the theory of extremes of stationary continuous time diffusions, in the specific case of finite inaccessible boundary points, to obtain the behaviour of the diffusion coefficient near the boundary, knowing that the boundedness of marginal densities implies that the diffusion coefficient is null outside the interval defined by the upper and lower bounds.

We translate these results in constraints for the coefficients of the discrete process we have to estimate. Then we study non parametric and parametric estimators. For discrete observations, many methods can be used. Using different non-parametric estimators, we can see that the drift b is linear, as often found in the literature, and the diffusion coefficient (or the volatility) a is almost linear in the central part.

We have to consider ε_n as Gaussian for coherence with the continuous time process and also for physical reasons. If we think of the continuous process driven by a Brownian motion, it is physically logical to keep the gaussianity of the noise. There exists a large literature about models given by (1.3). The geometric ergodicity, and thus the existence of a stationary solution, are proved under some hypothesis as strict positivity of the noise density, strict positivity or uniformity of the lipchitzian character of the diffusion coefficient, and some other weaker conditions.

Such hypothesis are not satisfied in our case, so we give slight modifications in order to keep the statistical properties of the model, including extremes behaviour but allowing gaussian noise and stationarity. We check our construction using simulations in different cases and bootstrap. We take a parametric piecewise polynomial model for a^2 and then use it for simulations.

There exists, in the literature, many models of simulation of the temperature. Most of them are done, with an econometric purpose, to study energy prices for instance. Only a few ones are in a climate scope. All these models have a linear drift and there are many kinds of residuals. It seems that none of them give a good reproduction of distributions, specifically in the tails. Our model gives, from a statistical point of view, almost correct extremes and good quantiles between 1% and 99%. These models can be useful in order to compute different risks for very hot or cold periods, as waves of heat or very cold winters (in different meaning).

We also make a crude study, by simulations, of the effect of Δ size on the model. It is difficult to extend to diffusion Pitterbag's theory of the effect of the size of the mesh Δ on the comparison between extremes in continuous

time and extremes for dates t of the form $t = k\Delta$. Pitterbag's work concerns only stationary gaussian processes. Extensions to a diffusion remain an open problem.

Chapter 11 discusses the modeling of the reduced temperatures, noted Z_t, given by equation (1.1). As discussed above, the seasonality still remains in Z_t and it affects both the central and extreme fields of these reduced series. This characteristic will be shown in the chapter. On the other hand, Z_t has usually a short memory never exceeding 5, which needs to be taken into account. Then, a model combining this facts with the extreme theory in Chapter 10 for Z_t is :

$$Z(t) = \sum_{k=1}^{p} \left[\theta_{0,k} + \sum_{j=1}^{p1} \left(\theta^j_{1,k} \cos \frac{2j\pi t}{365} + \theta^j_{2,k} \sin \frac{2j\pi t}{365} \right) \right] Z(t-k) + a(t, Z_{t-1})\varepsilon_t$$
$$\varepsilon_t \sim N(0,1) \tag{1.4}$$

where $a(t, Z_{t-1})$ is the diffusion coefficient which depends on both the dates t and the states Z_{t-1}.

In practice, a^2 is estimated as a polynomial of high degree with seasonal coefficients, and the constraints at the boundaries that a is canceled outside the bounded support are taken into account. The upper and lower bounds are estimated using extreme theory and depend on months, thus the seasonality is also considered for extremes.

Therefore, in our model the seasonal coefficients are used both in the mean and the variance terms of Z_t. The conditional variance is canceled out of the boundaries to guarantee the boundedness of Z_t in theory. These constraints are also seasonal.

The models will be applied on many kinds of temperatures : at a fixed-hour (the discrete version of the temperature process) and daily mean, maximum or minimum temperature. The goodness-of-fit of these models will be considered through simulations from many points of view : the residuals including their squared values, density functions, quantiles, extreme parameters. These quantities are evaluated at a precise date or precise month. Our model, in general, performs well and shows its better goodness-of-fit compared with other models. However it has some limits that we try to discuss.

Thus modeling the temperature is a very interesting, but very complicated subject, especially considering the recent greenhouse effects. Our model for

instance gives a quite good response to this problem. An important remaining problem is that our model has many parameters, which can become numerically heavy for simulations. A perspective work is to find a simpler model, in the sense of less parameters, which can still take into account all these features (complex seasonality, boundedness, volatility depending on state and on time).

We were obliged to limit the data treated in the thesis. We do not discuss problems involving several stations and problems specified to very long observed series for instance. This work is needed to check the generality of our methods, that means it must not be valid only on one particular station, one particular climate in Europe or one specific length of series.

It remains an important work to do in order to improve the model, if data are provided with a shorter interval of sampling, which has not been the case so far.

To summarize, this thesis gives a complete treatment on time series : trends, seasonalities, extremes, links between these factors, and is finished by the construction of a simulation model. The statistical methodologies proposed are general. Here we apply them on air temperatures, which is a popular subject of this last century. Some of our methodologies have been applied to water temperatures or precipitations. Many interesting phenomena of the temperature are revealed throughout this study and give us a more complete view about the strong complexity of this stochastic process.

Chapitre 2

Observed versus model produced data of temperature and warming effect

2.1 Introduction

For more than a century, human activities have contributed to the release of growing quantities of greenhouse gases in the atmosphere, such as carbon dioxide, methane, nitrous oxide or halocarbon. These greenhouse gases reinforce the natural greenhouse effect of the atmosphere and contribute to the modification of the earth climate. The global mean surface temperature has risen by +0.74C (+-0.18C) when estimated by a linear trend over 1906-2005 [4th IPCC assessment report], with an acceleration of the trend over the last 50 years. Most of this temperature increase, supported by changes in other climatic indicators like the northern hemisphere snow coverage or the sea level rise, is very likely due to the observed increase in anthropogenic greenhouse gas concentration, according to the last 2007 IPCC assessment report.

Air temperature has a great influence on EDF's activities : electricity consumption is strongly linked to air temperature, especially in winter due to an important part of electrical heating but increasingly in summer with the development of air conditioning installations. Electricity production is also conditioned by air and water temperature, particularly in summer. Therefore, the evolution of temperature extremes, very low winter temperatures as well as very high summer ones, is of crucial interest in the climate change context.

The impact of climate changes on extreme weather events is a growing subject of research studies in the scientific community. Some evidence on the evolution of the distribution of extreme and very high values can be found in papers of Understanding changes in weather and climate extremes (2000 in Bulletin of the American Society : Easterling et al.,[40], Meehl et al.,[99]). These papers also include a large bibliography on physical aspects of the problem and on data and studies for various areas in the world.

Many recent works focus on regions of different size, from parts of Europe to the entire world. They are concerned with the evolution of universally accepted indices proposed in order to represent high values of climate variables, such as temperature or precipitation.

Based on particular standards, these indices do not examine very high, rare events, but enable fair evaluations for very large scales (see Yan et al. ([154]), Frich et al. ([56])). Until recently, these studies suffered from a lack of spatial coherence, especially over Europe, due to the lack of a high time resolution dataset for this continent. This drawback has now been overcome, as the European project ECA&D (European Climate Assessment and Dataset) has provided a coherent dataset of long time series, checked for homogeneity, over the whole European continent (Klein Tank et al.,[90] ;

Alexander et al.,[4]).

These studies however focus, as mentioned, on an internationally agreed set of extreme indexes which describe rare events, but not the extremely rare values needed in the industrial dimensioning context. As a matter of fact, dimensioning of civil engineering works is based on the statistical evaluation of return levels for some defined return periods (20, 50 or 100 years depending on the required reliability). For example, air conditioning for some sensitive industrial sites is based on 100-year return levels of high local air temperature. Thus the main goal of this thesis work is to understand the possible evolution of temperature extremes and to propose methodologies to infer future long period return levels. To do so, it seemed relevant to first carefully study how the past warming tendency affected the occurrence and level of extreme temperature events. Then, from this knowledge, a statistical simulation of the temperature evolution, taking climate change into account and correctly reproducing the behavior of the tails of the distributions, has been investigated. As a matter of fact, such an evolution could be of great interest in the electricity consumption forecast, as temperature is one of the leading explaining variable.

2.2 State of the art

The climate change is in fact complex. It has an effect, with statistical evidence, on both the central field and extreme events of the temperature. Therefore, the analysis of past temperature distribution evolution need to be approached by a careful statistical non parametric study of the evolutions of the mean, variance and extremes. As recalled in the introduction, the temperature increase is bound to continue, and changes in mean temperature as well as in variability and extremes are predicted. Katz et al.([85]) have already mentioned that the variability change has more impact on extremes than the change in the mean. Some authors have also looked at the impact of climate changes on both the mean and the variability. Schär et al.([133]) pointed out the role of increasing variability in the occurrence of the 2003 heat wave in Europe. Ferro et al.([50]) proposed a technique to explore changes in probability distributions and applied it to climate model simulations.

The other methodological point in these studies concerns the trend estimation. Generally, trends in mean are derived using ordinary least square regression methods (Angell [84], Percival & Rothrock [34]). On the other hand, trends are usually studied separately : papers are devoted either to trends in mean or in variance, or to trends in extreme events. The trends

in extreme events are analyzed using linear least squares fit on the series of the previously mentioned "extreme index" (Kiktev et al. [89], Klein & Können [5]). Mudelsee et al.([96]) use kernel fitting to study flood risk in a nonparametric way. In this study, we would like to address the non parametric derivation of trends, as well as the links between trends in different quantities such as mean, variance and extremes. Generally, a trend is computed (not defined) as a slow, and thus regular component of a time series, leaving a quite stationary series, with less variable residuals. In other words, it consists in extracting some deterministic signal from a noisy dataset. In most studies, as stated before, this is done using ordinary least squares fitting, which will be referred to as classical trends.

Another problem is usually seen when working with the temperature data is the choice of scale for the time. There are many such scales possible, their choice depend on the purpose of the researchers. The long term one is traditionally the year effect, which can give an objective meaning to the measure of the unity, to measure "long time" effects (or low frequencies). The next effect is clearly the seasonal one in the period of one year. Then we can study higher frequency data if we have observations for instance every three hours, or every hour, and so on. All these scales : Short, medium and large are affected by the warming effect. Moreover, there exists the correlation between the observations, which is natural for climate factors, and this correlation is stronger on small scale temperatures. These facts have of course social consequences and need a careful statistical treatment.

Climate models describe the climate system of the earth in the physical-mathematical equations that can be solved numerically. It is used to reconstruct past climate and to study its future. The law governing the behavior of the atmosphere, soil, oceans, sun, and their interactions are translated into computer language. We finally get numerical description of climate phenomena and prediction following different scenarios. However careful studies of strong qualitative scaling of observed data and model simulations showed the difference between data from model simulations and observations.

This shows uncertainty in model simulations. They often do not correctly reproduce the stochastic items such as variability of the temperature, the chaotic nature of the climate system, and the influence of human being(Dessai and Hulme,[37]). Schwierz et al. ([136]) present an overview of the uncertainties in climate model projections that arise from various sources. They identified uncertainties stemming from the complexity and non-linearity of the climate system, its irregular evolution and the changing climate sensitivity, the emission scenarios selection and their implications for the radiative forcing, and the inevitably incomplete and inadequate representation of the climate system in a weather or climate model. The links

between trends and the extreme characteristics are also not satisfactorily represented in model simulations. A statistical simulation of the temperature evolution, taking climate change into account and correctly reproducing the behavior of the tails of the distribution could thus be helpful and needs to be investigated.

All these problems will be treated with care throughout our thesis. In the following, we will give more details on our datasets, the observations, reanalysis and model simulations.

2.3 The data used

The identification of changes in extremes requires the availability of long, reliable and homogeneous observation series on a daily basis. Our study concentrate on the series of temperature in Europe. For Europe, the datasets of daily temperature was lacking, only monthly datasets were available, and this statement of fact leads to the creation of the European Climate Assessment and Dataset project in 1998. Since 2002, this project provides a database of long and criticized series for Europe and the Middle East (Klein Tank et al.,[42]).

2.3.1 ECA & D series

The initial temperature series used in this study consist of 55 of the longest and most homogeneous daily temperature series (mean, maximum and minimum) made available by the ECA & D project (Klein Tank et al.,[42]). For each of the collected series, the ECA & D project tested for homogeneity using different test to identify breaks. The results are summarized for each series over different periods in classifying it as "suspect", "useful" or "undetermined" and in indicating the years when breaks have been identified. In the first place, only series quoted as "useful" over the 1946-2006 and the 1901-2006 periods have been retained. Then, among these series we keep only those for which no break has been identified with the three methods used in the ECA & D project. Table 2.1 summaries the series and their period lengths whereas Figure 2.1 shows their geographical location in Europe. The period lengths are the maximum periods of available (non-missing) data since 1946 (for 53 series) or 1901 (for the 2 Deols and Dresden series "useful" over this period).

Location	Country	Period for Tx	Period for Tn
Alger	Algeria	1946-1998	1946-1998
Armavir	Russia	1946-2003	1946-2003
Birr	Ireland	1954-2006	1954-2006
Bourges	*France*	*1946-2001*	*1945-2000*
Bremen	Germany	1946-2001	1946-2001
Calarasi	Romania	1946-2006	1946-2006
Cognac	*France*	*1946-2006*	*1947-2006*
Deols	*France*	*1901-2006*	*1921-2006*
Dresden	Germany	1917-2006	1917-2006
Elatma	Russia	1946-2003	1946-2003
Elista	Russia	1946-1999	1946-1999
Erfurt	Germany	1951-2006	1951-2006
Halle	Germany	1946-2006	1946-2006
Helgoland	Germany	1953-2006	1953-2006
Hopen	*Norway*	*1949-2006*	*1946-2006*
Hoseda	Russia	1946-2003	1946-2003
Hurbanovo	Czech Republic	1948-2006	1948-2006
Ile de Groix	France	1949-2006	1949-2006
Kaliningrad	Russia	1947-2006	1947-2006
Karasjok	*Norway*	*1951-2006*	*1946-2006*
Kaunas	Lithuania	1946-2006	1946-2006
Kem	Russia	1946-2005	1946-2005
Kempten	Germany	1952-2006	1952-2006
Kleine Brogel	Belgium	1954-2006	1954-2006
Kojnas	Russia	1946-2003	1946-2003
La Rochelle	*France*	*1946-2006*	*1955-2006*
Leipzig	Germany	1946-2006	1946-2006
List	Germany	1948-2006	1948-2006
Lugansk	Ukraine	1946-1996	1946-1996
Magdeburg	Germany	1947-2006	1947-2006
Moermansk	Russia	1946-2003	1946-2003
Narjan Mar	Russia	1946-2006	1946-2006
Onega	Russia	1946-2006	1946-2006
Orenburg	Russia	1946-2003	1946-2003
Orleans	France	1946-2006	1946-2006
Osijek	Hungary	1946-2006	1946-2006
Petrozawodsk	Russia	1946-2003	1946-2003

Petsjora	Russia	1946-2003	1946-2003
Potsdam	Germany	1946-2001	1946-2001
Siauliai	Lithuania	1946-2006	1946-2006
Smolensk	Russia	1946-2003	1946-2003
Sortavala	Russia	1946-2006	1946-2006
Syktyvar	Russia	1946-2003	1946-2003
Troitzko	Russia	1946-2003	1946-2003
Uman	Ukraine	1946-2006	1946-2006
Uzhgorod	Ukraine	1946-2006	1946-2006
Valley	United Kingdom	1946-2001	1946-2001
Vardoe	Norway	1951-2006	1951-2006
Verona	Italy	1951-2006	1951-2006
Vichy	France	1946-2006	1956-2006
Vilnius	Lithuania	1946-2006	1946-2006
Vilsandi	*Estonia*	*1949-2006*	*1946-2006*
Voru	Estonia	1946-2004	1946-2004
Vytegra	Russia	1946-2006	1946-2006
Wologda	Russia	1946-2006	1946-2006

TABLE 2.1 – Location and period length of the ECA&D selected daily temperature series.The series with different period lengths for maximum temperature(Tx) and minimum temperature (Tn) are written in italics

2.3.2 European ENSEMBLES project gridded dataset

The production of gridded daily datasets from observations is one of the deliverables of the European ENSEMBLES project (http ://www.ensembles-eu.org) : the research team RT5.1 of the project is due to produce observational daily gridded datasets for temperature and precipitations. The ENSEMBLES project is supported by the European Commissions 6th Framework Program as a 5-year Integrated Project from 2004 to 2009 under the Thematic Sub-Priority "Global Change and Ecosystems". Different datasets are now available, for daily minimum, maximum and mean temperature and daily precipitation amount, covering the period from 1950 to 2006 on a 0.25 and 0.5 regular grid as well as on a 0.22 and 0.44 rotated pole grid. The datasets for daily minimum (Tn) and daily maximum (Tx) temperature on a 0.5 regular grid have been used in this study.

2.3.3 ERA40 reanalysis

The European Centre for Medium range Weather Forecasts (ECMWF : http :// www.ecmwf.int) has conducted a reanalysis project, in order to promote the use of global analyses of the state of the atmosphere, land and

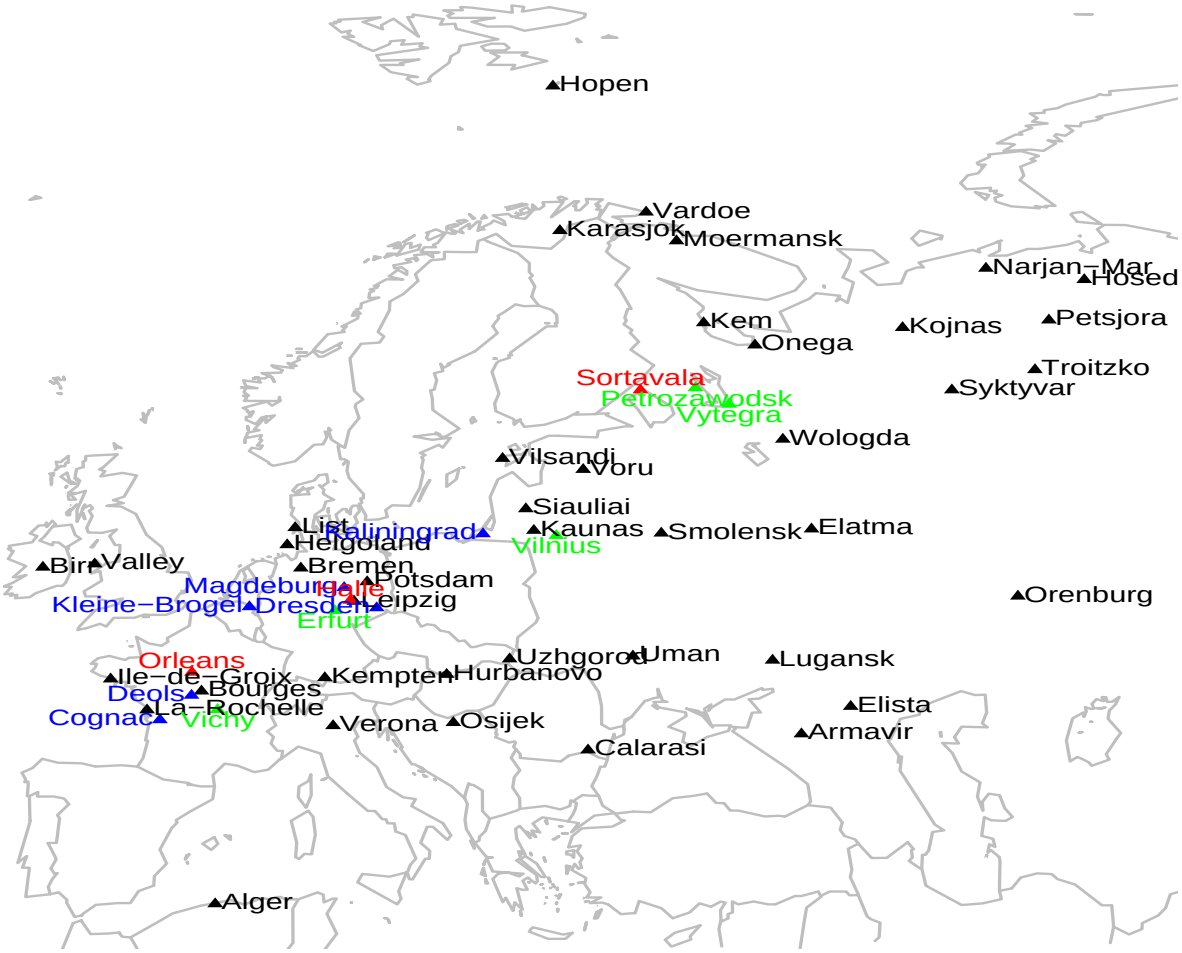


FIGURE 2.1 – Locations of the selected ECA&D temperature series

surface conditions over the period from mid-1957 to 2001. The three dimensional variational technique has been applied using the T159L60 version of

the Integrated Forecasting System to produce the analyses every six hours. The 2m-temperature dataset over a 2.5 regular grid has been used in this study. From this dataset, the maximum (minimum) of the four daily values has been considered for the daily temperature maximum (minimum). This leads to slightly under-estimated values compared to the "true" daily maximum or minimum, but with a comparable order of magnitude. From these different datasets, the analysis has been made on a spatial window covering Europe between longitudes 10 and 65 East and between latitudes 35 and 80 North. The hot season has been defined as the 100 days between the 14th of June and the 21st of September, whereas the cold season has been defined as the 90 days between the 1st of December and the 28th of February. These periods have been selected because most extreme events occur between these dates.

In Chapter 4, we work on the daily maximum or minimum temperature in the purpose of finding the global evolution of extreme temperature in the summer and the winter. In Chapter 6, we work on the same kind of datasets but we concentrate on the maxima (or minima) of these series to construct the extreme models. In Chapter 11, we work on the temperature with different frequencies from 3 hours to 24 hours to study a discretized version of diffusion process.

2.3.4 Simulation models

The data from simulation models are not our central data. They are only used sometimes for the purpose of giving a general view of some behaviours and also the uncertainty of simulation models.

Six different climate models will be used in our work. These models are :

• bccr-bcm2-0 of the university of Bergen in Norway

• cnrm-cm3 of the French meteorological office research centre (Centre de Recherches Mtorologiques de Mto-France)

• ipsl-cm4 of the Institut Pierre-Simon Laplace in Paris, France

• mpi-echam5 of the Max Planck Institute for meteorology in Hamburg, Germany

• ingv-echam4 of the National Institute of Geophysics and Volcanology in Bologna, Italy

• ukmo-hadgem1 of the United Kingdom Meteorological Office Hadley Centre, over the period 1950-1999

Simulations of these models obtained in the framework of the fourth IPCC assessment are available from the PCMDI web portal.

Chapitre 3

Nonparametric estimation

3.1 Introduction

This chapter is devoted to the study of all the non parametric methods used in this thesis. Some classical results are most often repeated when they need modifications to be applied to temperature series. On the other hand, we will develop some theoretical points when the existing literature is not sufficient. Non-constant variability, correlated errors and non-orthogonal parameters are all known to provide difficult problems in stochastic modeling and estimation. Temperature series present all these characteristics. In this feature, the data length is also an important element.

Temperature series X_t presents also quite evident year and day seasonalities but in a complex form. All the coefficients chosen to represent these seasonalities and the dynamic of the residuals are affected by the warming effect.

In this chapter, to simplify the exposition, we neglect the seasonalities. Later we come back to the seasonality and justify our advance. This assumption is acceptable when the datasets are rather homogeneous, for example when only the daily maximum temperature in summer are considered, but this kind of assumption needs to be considered with caution.

In this case, the question is how to choose a representation of the series simple enough to have a statistical meaning while taking into account the main features of the series. We can think about a model of the forms :

$$X_t = m_t + s_t\varepsilon_t \quad \text{or} \quad X_t = m_t + \eta_t \tag{3.1}$$

where m_t is the mean trend, s_t^2 the variance one and ε_t or η_t the residual process.

There are many ways to estimate m_t in a quite objective sense. First, the most naive way is to compute a mean for every "year" (can be a season) a : $\bar{m}_a = \sum X_t/N_a$ where N_a being the length of the year (or season). Then $\bar{m}_a$ can be used as a nonparametric estimator of m. This method is often used in climate applications. Other simple methods are to use a linear regression or piecewise linear regression instead of a nonparametric estimation .

Over a long period, say a century, such methods are quite arbitrary because they force the regression function to take a defined-in-advance parametric form. They avoid a statistical choice of the parameter of smoothness, but replace it by the degree (in general high for long series) of a polynomial or a piece-wise linear line. Nonparametric methods present the advantage of giving an objective approach of "what is a low frequency for this trend ?". Non parametric regression surfaces can consume a large number of degrees of freedom in the patterns of complex data. In our case, a logic and appro-

priate way is to find out what is "the significant window" one has to choose to make a (weighted) moving average of the observations. This bandwidth parameter is significant in our context because it is clearly linked to the "memory" of the warming process. There exists a large number of non parametric methods linked with this kind of "window" parameter. The simplest is the (weighted) moving average method, then we have the kernel method and the most popular is the loess (local polynomial regression) method. The fact that the model has a trend in the variance, and that evidently ε_t or η_t is not a white noise make the use of nonparametric approaches difficult and dangerous without caution in choosing the bandwidth parameter. In this context, we will use loess to estimate the trends in model (3.1). The asymptotic properties of loess, and an estimate procedure for this kind of model, will be detailed below.

If we work without seasonalities, as we previously said, one can suppose that ε_t is a stationary process or at least of stationary covariance, and for us weakly dependent. Of course this is a first assumption ; there are no statistical evidence of non stationarity. In the simplest case, we therefore need to estimate at least $m(t), s(t)$, the covariance function of ε_t or that of η_t and perhaps, in very particular cases, the distribution of ε_t, to have a complete description of the model (3.1).

In the literature, there are no asymptotic results, neither are there good and well justified algorithm which would be able to answer this problem in the general case of model (3.1). So we try to give a partial answer, to estimate (3.1). These trends $m(t)$ and $s(t)$ represent the main characteristics of the central part of the distribution of X_t. From now on, we will call this part "the central field" for the sake of lightness.

Another important subject of this thesis is the modeling of the extreme temperatures. Roughly speaking, the problems related to this subject are linked with the construction of probabilistic and statistical properties related to very high or very low temperatures. In order to understand how climate change might affect such extremes, it is necessary to explore the changes that have occurred both in the past and in the present. That means we need to build the trends of extremes. In this case, they are represented as the functional parameters of the extreme value distribution. In order to have flexible representation for trends, we estimate the extreme parameters non parametrically. In this case where we know the distribution of the noise, or at least an approximation, likelihood based methods of approximation are then preferred. In practice, both the two nonparametric methods loess and splines can be used. For numerical considerations, the spline approach gives a faster algorithm. Moreover, when we use local likelihood (loess), sometimes, a rupture in the curve of the estimates can happen when the tuning

parameter is small. For these reasons, when working with the extreme models, we will use cubic splines to estimate the extreme trends.

The choice of smoothing parameters is really important, but very difficult. Moreover in our case, the problems cited above, such as correlated and non stationary data or estimations of non-orthogonal parameters, complicate the choice of smoothing parameters. There exists a number of available works dealing with these problems, in general with poor theoretical basis. They do not always give good results or need really heavy numerical computations for large datasets. In this chapter, we develop some new algorithms to solve optimization problems in two different contexts : estimation of the trends in the central field by loess with the presence of correlated errors, and estimation of non-orthogonal extreme parameters in extreme models by splines. An important part of this chapter concerns the validation of these algorithms, which are mainly justified through simulations.

3.2 Application of loess

In this section, we will develop some points on asymptotic properties for our specific context : the variance is not constant and the errors are correlated.

The basic idea of loess is similar to the one of the kernel estimation : estimate the mean function by using the information of the nearest neighborhood. However instead of approximating locally the mean function by a constant, the loess estimator is obtained by locally fitting a pth degree polynomial to the data via weighted least squares. Throughout our work, the local linear fit is used, that means $p = 1$ is taken.

The earlier works on local least squares regression estimators are due to Stone ([145]) and Cleveland ([21]). Advantages of loess compared with kernel regression estimation are its better boundary behavior, its adaptation to estimate regression derivatives and its good minimax properties. Some significant references are Fan ([45], [46]), Hastie and Loader ([74]), Fan and Gijbels ([47]). In these papers, the independence of the observations is assumed. The statistical properties of local polynomial regression for dependent stationary data, have been studied in recent works of Masry and Fan ([98]) and Francisco-Fernández and Vilar-Fernández ([53],[54]).

In our case, fixed designs and short-range dependence nonparametric regression models are considered.

3.2.1 Estimation of mean trend m

In this part, we start from results of Ruppert and Wand ([131]), Francisco-Fernández and Vilar-Fernández ([53]). They studied the loess estimation of the regression function in the context of dependent errors but constant variance, and second-order stationary process. We give appropriate extensions to our more general model.

Let X_t be the observations. We consider first the model :

$$X_t = m(t) + s(t)\varepsilon_t,\ t = 1, ..., T \tag{3.2}$$

where $m(t)$ and $s(t)$ are regular (in general twice derivable functions), $s(t) > 0$ and ε_t is a stationary process, or at least the model

$$X_t = m(t) + \eta_t,\ t = 1, ..., T$$

with some constraints of boundedness.

As for parametric regressions (see significant examples in chapter 5 on extremes), no asymptotic theory exists for this kind of model, for $T \to \infty$, because nothing has been said on the behavior of m and s out of $[1, T]$. So in general, we study its corresponding asymptotic properties on a compact set, say [0,1], by the time transformation $t \longrightarrow t/T$. Therefore, the asymptotics of the estimates $\hat{m}(t)$ and $\hat{s}(t)$ are in fact the asymptotics on $\hat{m}_n(t/n)$ and $\hat{s}_n(t/n)$. We note however $\hat{m}(t)$ and $\hat{s}(t)$.

The only remaining problem for us (see chapter 10 on the diffusion) is the possible confusion with discretization problems of a continuous time process. We discuss this point in the chapter on diffusion processes.

We now give some notations to make the description of the asymptotic properties easier.

Let $x \in (0, 1)$, $x_k = k/n$, $k \in \mathbb{N}$, n is often omitted. Let m and s be functions defined on $(0, 1)$. For every n, let M_n (respectively S_n) be the vector $(m(x_1), m(x_2), ..., m(x_n))^T$ (respectively $(s(x_1), s(x_2), ..., s(x_n))^T$). The vector of observations is $\vec{X}_n = (X_1, ..., X_n)^T$.

Let K be a kernel and K_n defined by $K_n(x) = \frac{1}{h_n} K\left(\frac{x}{h_n}\right)$ where h_n is the bandwidth.

We shall need the following quantities : $\mathfrak{X}_n(x)$ is the $n \times 2$ design matrix

$$\mathfrak{X}_x = \frac{1}{n} \begin{pmatrix} 1 & x - x_1 \\ \vdots & \vdots \\ 1 & x - x_n \end{pmatrix}$$

$W_n(x)$ is the $n \times n$ weight matrix defined by :

$$W_n(x) = (n)^{-1} diag\left[K_n(x - x_1), \dots, K_n(x - x_n)\right]$$

For $r \in \mathbb{N}$, let $u_n^r(x) = \frac{1}{n}\sum_n^{i=1}(x - x_i)^r K_n(x - x_i)$. $U_n(x)$ is the 2×2 matrix defined for $0 \leq i, j \leq 1$ by $U_n^{i,j}(x) = u_n^{i+j-2}(x)$.

Let $v_n^r(x) = \frac{1}{n}\sum_n^{i=1}(x - x_i)^r K_n(x - x_i)X_i$ and $V_n(x) = (v_n^0(x), v_n^1(x))^T$.

In what follows, x is omitted when there is no ambiguity. We note that $U_n = \mathfrak{X}_n^T W_n \mathfrak{X}_n$ and $V_n = \mathfrak{X}_n W_n \vec{X}_n$.

The local linear estimator of $m(x)$ is constructed from :

$$\hat{\beta}(x) = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} [X_i - \beta_0 - \beta_1(x - x_i)^p]^2 K_n(x - x_i) \tag{3.3}$$

The minimum is reached for $\hat{\beta}_x = (\hat{\beta}_{0,x}, \hat{\beta}_{1,x})^T$. $\hat{\beta}$ is given by :

$$\hat{\beta} = \left(\mathfrak{X}_n^T W_n \mathfrak{X}_n\right)^{-1} \mathfrak{X}_n^T W_n \mathfrak{X}_n \vec{X} = U_n^{-1} V_n \tag{3.4}$$

Then we have $\hat{m}(x) = \hat{\beta}_0(x)$. Let $M_n = \mathfrak{X}_n\, \beta$ the vector $(m(x_1), ..., m(x_n))^T$.

Now we investigate the asymptotic properties for the bias and variance of the estimator $\widehat{m}(x)$. The following assumptions will be needed in our analysis :

A.1.(kernel) K is symmetric, with a bounded support, and uniformly Lipschitz. The odd moments of K are null.
A.2.(bandwidth) $h_n \in (0, 1)$, $h_n \to 0$ and $nh_n \to \infty$.

A.3.(m and s) $m \in C^2(0, 1)$, $s \in C^2(0, 1)$.

A.4. (correlations) Denote $\eta_j = s_j \varepsilon_j$, then
$\sup_n 1/n \sum \sum_{1 \leq i,j \leq n} \mid cov(\eta_i, \eta_j) \mid = D < \infty$.

A'.4 $\lim 1/n \sum \sum_{1 \leq i,j \leq n} \mid cov(\eta_i, \eta_j) \mid < \infty$.
A.5 $\sup_n 1/n \sum \sum_{1 \leq i,j \leq n} \mid cov(\eta_i^2, \eta_j^2) \mid = D_2 < \infty$.
The assumption A4 is linked to non-stationarity. If ε_j is stationary, with autocovariance function γ, A'4 is implied by $\sum_k k \mid \gamma(k) \mid < \infty$. It is a stronger hypothesis than $\sum \mid \gamma(k) \mid < \infty$ which leads to

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma(j - i) = c < \infty$$

but do not imply A4.

In order to explore the asymptotic properties of the estimator $\hat{m}$, we need some lemmas. The proof of the following lemma is based on two remarks proceeding from A1.

K with bounded support implies Card$\{i, K_n(x_i - x) \neq 0\} = O\left(\frac{1}{nh_n}\right)$ for every n and Card$\{l, K_n(x_i - x)K_n(x_l - x) \neq 0\} = O\left(\frac{1}{nh_n}\right)$ for every x and every i.

Note $K^*(t) = e_1^T U_n^{-1}(1,t)^T K(t)$ with $e_1^T = (1, 0...., 0)$. $K^*(t)$ is called *equivalent kernel* (see Fan & Jijbels,[47]). We have $\widehat{m}(x) = \sum_{i=1}^n K^*(x_i - x)X_i$,, $\sum_{i=1}^n K^*(x_i - x) = 1$ and $\sum_{i=1}^n (x_i - x)K^*(x_i - x) = 0$.
Let $\mu_j^r = \int u^j K^r(u)du$, $(j,r) \in \mathbb{N}^2$ and let

$$w_n^{j,r} = \frac{1}{n}\sum_{i=1}^{n}(x - x_i)^j K_n^r(x - x_i)$$

then :

Lemma 3.2.1 *1/* $\lim_n h_n^{-j} w_n^{j,1} = \int u^j K(u)du = \mu_j^1$

2/ $\lim_n h_n^{-2j}(w_n^{j,1})^2 = \int u^{2j} K^2(u)du = \mu_{2j}^2$

3/ Let $\bar{v}_n^r(x) = v_n^r(x) - Ev_n^r(x)$ *and* $\bar{V}_n(x) = (\bar{v}_n^0(x), \bar{v}_n^1(x))^T$. *So* $\bar{V}_n(x) = \mathfrak{X}_n(x)W_n(x)(\vec{X}_n - M_n)$ *then :*

$$\lim_n cov(h_n^{-l}\bar{v}_n^{-l}(x), h_n^{-r}\bar{v}_n^{-r}(x)) \leq D\mu_{l+r}^1$$

The proof of 3 is straightforward using A1, A4 and point 2 of the lemma. From this lemma we cannot have a limit, only a majoration of the variance except if A'4 is satisfied.

The asymptotics of the estimator $\hat{m}(x)$ will be studied. The following lemma is a consequence of the theorem 4.1 of Ruppert and Wand ([130]).

Lemma 3.2.2

$$M_n = \mathfrak{X}_n(x)\beta_n(x) + \frac{m"(x)}{2}\begin{bmatrix} (x - x_1)^2 \\ (x - x_n)^2 \end{bmatrix} + o(h_n^2)$$

So

$$\begin{aligned} E(\hat{\beta}_n(x)) &= E(V_n^{-1}W_n\mathfrak{X}_n\vec{X}) \\ &= U_n^{-1}\mathfrak{X}_n^T W_n\mathfrak{X}\beta_n(x) + U_n^{-1}\frac{m''(x)}{2}\begin{bmatrix} u_n^2 \\ u_n^3 \end{bmatrix} + o(h_n^2) \end{aligned} \tag{3.5}$$

Directly from the definitions, we have : $U_n^{-1}(x)\bar{V}_n(x) = \hat{\beta}_n(x) - E(\hat{\beta}_n(x))$. *Then*

$$\hat{\beta}_n(x) - E(\hat{\beta}_n(x)) = U_n^{-1}\bar{V}_n(x) + U_n^{-1}\frac{m''(x)}{2}\begin{bmatrix} u_n^2 \\ u_n^3 \end{bmatrix} + o(h_n^2)$$

We get the following theorem from Lemma 3.2.1 and Lemma 3.2.2.

Theorem 3.2.1

$$\begin{bmatrix} 1 \\ h_n \end{bmatrix} E(\hat{\beta}_n - \beta(x)) = \frac{m''(x)}{2} h_n^2 U_n^{-1} \begin{bmatrix} \mu_2^1 \\ \mu_3^1 \end{bmatrix} + o(h_n^2)\mathbb{I}_2 \tag{3.6}$$

where $\mathbb{I}_p = (1, ..., 1)^T$ *: unit vector of dimension* p.

$$Var\left(\begin{bmatrix} 1 \\ h_n \end{bmatrix} \hat{\beta}_n(x)\right) \leq \frac{D}{nh_n} U_n^{-1} G_2 U_n^{-1} \tag{3.7}$$

$$\text{where } G_2 = \begin{bmatrix} \mu_0^2 & \mu_1^2 \\ \mu_1^2 & \mu_2^2 \end{bmatrix}$$

If $\frac{1}{n}\sum_i \sum_j E\eta_i\eta_j \xrightarrow{n\to\infty} C < \infty$ *then*

$$Var\left(\begin{bmatrix} 1 \\ h_n \end{bmatrix} \hat{\beta}_n(x)\right) = \frac{C}{nh_n} U_n^{-1} G_2 U_n^{-1} + o\left(\frac{1}{nh_n}\right) \tag{3.8}$$

Remark : This kind of theorem gives a trade-off between bias and variance which highly depends on the derivability hypothesis made on m. In this sense, its practical interest is quite relative ; it is more an indication than a true asymptotic result. If $m^{(p)}$ is supposed to be continuous of course the bias becomes of order h_n^{p+1}.

By writing (3.6) and (3.8) under the form of K^* and then using the fact that $K^* = K$ in the linear case ($p = 1$), we get the mean square error of the estimator $\hat{m}(x)$:

Theorem 3.2.2

$$MSE(\hat{m}_n(x)) = \left[\frac{1}{2}h_n^2 m^{(2)}(x)\mu_2^1\right]^2 + \frac{1}{nh_n}\mu_0^2 C + o\left(\frac{1}{nh_n} + h_n^4\right) \tag{3.9}$$

where $C = \lim_n \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^n E\eta_i\eta_j$ *if it exists.*
and MSE$(\hat{m}_n(x))$ *is bounded by the same quantity when we replace* C*, if it does not exist, by* $\sup_n \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^n |E\eta_i\eta_j|$.

These asymptotic results are valid for an interior point at, at least, an h_n distance from the boundary. What is the difference with a point in the boundary ? Does there exist a boundary bias ?

Let us choose a left boundary point in the form $x = ch$, with $c \geq 0$, whereas we select a right boundary point of the form $x = 1 - ch$. Let us also assume that $m^{(2)}(x)$ and $s(x)$ are right continuous at the point 0 and left continuous at the point 1.

Modifying the results of Fan([45]), we can write the MSE of the estimator $\hat{m}$ at the left boundary point $x = ch$ as :

$$MSE(\hat{m}_n(x)|X) \sum = h^4 \alpha(c) \left[\frac{m^{(2)}(0^+)}{2} \right]^2 + \beta(c) \frac{C_n^+}{nh} + o(h^4) \qquad (3.10)$$

where $C_n^+ = 1/n \sum_{k=0}^{n} \sum_{i=1}^{n-k} E(\eta_i \eta_j)$ and with $\mu_{j,c} = \int_{-\infty}^{c} t^j K(t) dt$,

$$\alpha(c) = \frac{\mu_{2,c}^2 - \mu_{1,c}\mu_{3,c}}{\mu_{0,c}\mu_{2,c} - \mu_{1,c}^2}$$

$$\beta(c) = \frac{\int_{-c}^{\infty} (\mu_{2,c} - u\mu_{1,c})^2 K^2(u) du}{(\mu_{0,c}\mu_{2,c} - \mu_{1,c}^2)^2} \qquad (3.11)$$

From (3.9) and (3.10), we note that the squared bias of the local linear regression estimator is smaller at the boundary point than at an interior point, but the variance tends to be larger at the boundary because there are fewer observations in estimation. However the MSE of a boundary point has the same form as the one of an interior point. The expression (3.9) of an interior point corresponds to (3.10) with $c = \infty$. We can conclude that the linear local polynomial estimation does not suffer from boundary effects. Usually, we say that local polynomial estimation has an automatic boundary bias correction.

3.2.2 Estimation of variance trend s^2

As presented in Chapter 2, the variance of temperature changes with time. For example, when we consider the daily maximum temperature for the last 30 years, the variance has an increasing tendency with time. For this reason, a non constant variance needs to be taken into account in our modeling. It is of common interest to estimate variances in a variety of statistical applications. In our case, this variability is identified through the model :

$y_i = m(x_i) + s(x_i)\varepsilon_i$ by estimating $s(x)$ by loess from squared residuals. The literature on this type of model is rather sparse. There were some works on the heteroscedasticity model with kernel-based methods such as Müller & Stadtmüller([101]), Hall & Carroll([67]), Brown & Levine([13]) ; with local polynomials such as Ruppert and al.([131]), Fan and Yao([48]).

From the existing works, it is well-known that without knowing the regression function m, we can still estimate the variance asymptotically as well as if the regression were given. The effect of preliminary estimation of the mean has only a second-order effect on the asymptotic performance of $\hat{s}^2(x)$ (see Ruppert et al., [131]). We will show in this section that by adding appropriate assumptions, these properties are still true when the noise is correlated.

If m is given, one can solve the problem of estimating s^2 by using the following nonparametric regression residual model given by :

$$r_i^2 = (X_i - m(x_i))^2$$

and

$$r_i^2 = s^2(x_i) + e_i$$

It can be considered as equivalent, in some precise sense, to :

$$r_i = s(x_i)\varepsilon_i$$

in which the relation $E(r_i^2) = s^2(x_i)$ is kept as in the previous one.

For least square consideration, the two parameters m and s play the same role. Now, if m is unknown, an intuitive approach is to first estimate m and to then plug it, in order to estimate the new sample $\hat{r}_i^2 = (X_i - \hat{m}(x_i))^2$. The loess method is then applied to this series to estimate s^2.

Suppose that m has been estimated as in the previous sub-section by $\hat{m}(x) = \hat{\beta}_0(x)$ with a bandwidth h_1. Now we estimate $s^2(x)$ by $\hat{s}^2(x) = \hat{\alpha}_0$ by the same method with a bandwidth h_2 where :

$$(\hat{\alpha}_0, \hat{\alpha}_1) = \underset{\alpha_0,\alpha_1}{\operatorname{argmin}} \sum_{i=1}^{n} [\hat{r}_i^2 - \alpha_0 - \alpha_1(x_i - x)]^2 K\left(\frac{x_i - t}{h_2}\right) \tag{3.12}$$

This estimator of s^2 is called residual-based estimator.

We need to show that the asymptotic performance of the estimate $\hat{s}^2$ from this plug-in method is the same as if the mean function m were known. We follow the work of Ruppert et al. [131] to study the asymptotic properties

of this residual-based estimator of s^2, but we have to extend the hypothesis.

We keep the same notations as in the previous sub-section, note h_n^1 is the bandwidth in the estimate of m and h_n^2 is the one in the estimate of s^2.

We make on s the same assumptions as on m. Now we have to control the correlation of r^2.

Let $\kappa^{i,j} = cov(r_i^2, r_j^2) = cov(\eta_i^2, \eta_j^2)$ and suppose satisfied the condition :

A.5. $\sup_n \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \mid \kappa^{i,j} \mid = D_2 < \infty$

Let Γ_n be the covariance matrix $(\kappa^{i,j}_{1\leq i,j\leq n}\ )$, $S_n^2 = (s^2(x_1), ..., s^2(x_n))^T$ and $\hat{S}_n^2 = (\hat{s}^2(x_1), ..., \hat{s}^2(x_n))^T$. Let L_1, L_2 be the smoother matrix corresponding respectively to the estimators $\hat{m}$ and $\hat{s}^2$. That means $\hat{m} = L_1 \vec{X}_n$ and $\hat{s}^2 = L_2 \hat{r}^2$.

Consider the lemma below :

Lemma 3.2.3

$$E(\hat{s}_n^2(x) - s^2(x)) = (L_2 - I)S_n^2 + L_2 b_n^1 (b_n^1)^T + diagonal(L_1 \Sigma_n L_1^T - 2L_1 \Sigma_n)$$

where $b_n^1(x) = (L_1 - I)M_n$ *and*

$$cov(\hat{S}_n^2) = L_2(L_1 - I)(L_1 - I)\Gamma_n(L_1 - I)^T(L_1 - I)^T L_2^T \tag{3.13}$$

Then we obtain the following theorem :

Theorem 3.2.3 *If*

$$\{(h_n^1)^4 + (nh_n^1)^{-1}\} = o((h_n^2)^2) \tag{3.14}$$

as $n \to \infty$. *Then :*

$$E[\hat{s}^2(x) - s^2(x)] = \frac{(h_n^2)^2}{2} \mu_2^1 (s^2)"(x) + o((h_n^2)^2) \tag{3.15}$$

and

$$var[\hat{s}^2(x)] \leq \frac{1}{nh_n^2} \mu_0^2 D_2 + o(\frac{1}{nh_n^2}) \tag{3.16}$$

If $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \kappa^{i,j} = l$ *exists and* $D_2 < \infty$, *we can put* l *instead of* D_2.

The proof is the same as that of the theorem on the estimate of m.

From this theorem, we can see that the leading bias and variance terms of $\hat{s}^2$ estimated by a local linear estimation are analogous to those obtained by the mean estimator $\hat{m}$, except that the terms linked with m are now replaced by s^2. The asymptotic properties of the estimator of the variance are the same as those in the case where m is known. These terms depend only on the bandwidth h_n^2, indicating that the initial bandwidth h_n^1 has only a second-order effect on the asymptotic performance of the variance estimator $\hat{s}^2$, which is determined by (3.14). So there is no loss in asymptotic efficiency due to the estimation of m if the condition (3.14) is fulfilled.

The only problem left in the argument of Ruppert et al.([131]) is the condition (3.14) in the choice of bandwidth. We can use an automatic selector to choose h_n^1 but this is not possible for h_n^2 because of the second-order effect of h_n^1 on h_n^2. The choice of h_n^2 must therefore depend on h_n^1. Ruppert et al. proposed an alternative strategy to conserve the optimal asymptotic properties : first choose h_n^1 through an automatic bandwidth selection and then apply the same bandwidth to estimate squared-based variance. In the end, (3.14) will be satisfied.

Fan and Yao ([48]) applied a local linear regression to the squared residuals and demonstrated that, without knowing the regression function, we can still estimate the variance asymptotically as well as if the regression were given. Their basic idea is analogous to that in Ruppert et al.[131], however the errors are not demanded independent and condition of second-order effect of h_n^1 on h_n^2 are not needed. The choice of bandwidth for the estimation of variance is therefore more flexible. The assumption required in their study is that the observations are made from a strictly stationary and absolutely regular process. Their assumption cannot be satisfied by our temperature data. As often, asymptotic results are interesting to prove the consistence of the estimators. But their lack of adaptivity, in particular in the choice of the order of derivatives which controls the bias value, limits their practical usefulness.

3.2.3 Bootstrap confidence intervals

It is now necessary to quantify the uncertainties associated with the estimation of the mean and variance evolutions. A confidence interval based on the residual bootstrap resampling will be constructed. It is simple, and has been chosen because it does not rely on the evaluation of quantities according to some asymptotic property. There exists a widely developed theory for asymptotic confidence intervals (uniform or pointwise confidence intervals) such as Hall ([65],[66]), Faraway ([49]) ; Haïrdle and Mammen([71]) ;

Neumann([103]). Recently, a complete treatment on the uniform confidence bands, for both the mean and the variance for the spline method, has been done in Wang and Yang([147]). We do not discuss about uniform bands here. Our choice will be to construct a pointwise confidence interval based on bootstrap percentiles. It is well adapted to our estimation method because it does not artificially assume normality and it is simple to implement with our large datasets.

A classic bootstrap can not be satisfied because of the presence of dependence in the datasets. We use here the moving blocks bootstrap (Lahiri, [91]) with an appropriate size of block.

The procedure can be described as :

• Estimate $m(x)$ and $s(x)$ and calculate the residuals $\hat{\varepsilon}_i = \frac{X_i - \hat{m}_i}{\hat{s}_i}, 1 \leq i \leq n$.

• Construct B bootstrap samples of same size of $\hat{\varepsilon}$: by taking randomly $[\frac{n}{b}]$ blocks, of chosen size b, of the elements in $\hat{\varepsilon}$ with replacement, we obtain ε^{*j}, $j = 1, ..., B$.

• Construct the bootstrap samples of X by using : $X_i^{*j} = \hat{m}_i + \hat{s}_i \varepsilon_i^{*j}$. Estimate the new trends from these samples m^{*j} and s^{*j} with $j = 1, ..., B, i = 1, ..., n$.

• At each $i, 1 \leq i \leq n$, take the 5th percentile and 95th percentile from the vectors $(m_i^{*1}, ..., m_i^{*B})$ and $(s_i^{*1}, ..., s_i^{*B})$. We then obtain the confidence interval of 90% for $\hat{m}(x)$ and $\hat{s}(x)$.

3.3 Automatic selection of the smoothing parameters for correlated data

The use of nonparametric regression involves the choice of a smoothing parameter which controls the balance between goodness of fit to the data and smoothness of the regression function. In order to avoid subjectivity and to simplify the work of the statisticians while not using unrealistic asymptotic results, the use of an automated selection method is desirable. In the previous section, we showed some rules of parameter choice for the local and smoothing spline regressions. When the data are independent, these methods often perform well. However it is well known that the presence of correlated errors greatly affects the selection of smoothing parameters. If the data are dependent, the smoothing parameter selectors designed for independent

data will produce poor results. If we consider the model $X_i = m(x_i)+\varepsilon_i$, the bandwidth selection methods "perceive" the trend in the data due to the mean function m, and attempts to incorporate that trend into its estimate. When the data are uncorrelated, this "perception" is valid, but it breaks down in the presence of correlation (Opsomer et al., [108]).

In the different contexts, many authors such as Opsomer([107]), Altman([106]) and Chu and Marron ([20]), Hart ([82]), in the context of kernel regression, Francisco-Fernández and Vilar-Fernández ([53]), in the context of loess, and Wang ([148]), in the context of splines, showed that the effect of ignoring even minor amounts of error correlation when selecting a smoothing parameter could potentially be severe. An analysis of the problem and a complete literature on automatic selections for correlated data are carried out in Opsomer et al. ([108]). When the data are (mostly) positively correlated, the smoothing parameter is underestimated by an automatic selection (for independent case) which produces then a rough estimate. On the other hand, when data are (mostly) negatively correlated, a larger bandwidth will be chosen which results in an oversmooth estimate. More details about the performance of ordinary cross-validation when data are correlated is detailed in Hart ([72]).

In the heteroscedastic model :

$$y_i = m(x_i) + s(x_i)\varepsilon_i \tag{3.17}$$

when we estimate the trends by local linear polynomial, the effect of dependence on the smoothing parameter can be seen in the form of an asymptotically optimal bandwidth.

As presented in the previous subsection , supposing the appropriate hypothesis are satisfied, we have asymptotic mean squared error :

$$MSE(\hat{f}(x)|X) = \left[\frac{1}{2}h_n^2 m^{(2)}(x)\mu_2^1\right]^2 + \frac{1}{nh_n}\mu_0^2 C + o(\frac{1}{nh_n} + h_n^4) \tag{3.18}$$

Then by minimizing MSE, we can obtain an asymptotically "optimal" local bandwidth at each point x

$$h_n^{opt}(x) = \left[\frac{C\mu_0^2}{n(\mu_2^2)^2(m^{(2)}(x))^2}\right]^{1/5} \tag{3.19}$$

When the data are independent, we obtain the same form of h_n^{opt} but C is replaced by the variance $s^2(x)$. So if the autocovariance function is very often positive, C is large and the optimal bandwidth is larger than the one in the independent case. So, when we wrongly assume independence of the

errors, by using the ordinary automatic bandwidth selection, we obtain a bandwidth that is far from the optimal bandwidth.

We place our study in the local linear estimation context because this estimation will be applied in the next sections. We will present some existing automatic selections which could correct the correlation effect. A simulation study will be added to compare these methods. We will propose a new approach to solve this problem. It is a modified version of the partitioned cross validation(PCV) of Marron ([97]). The main idea of PCV is to partition the data into g subgroups by taking every gth values. Then PCV is derived from the average value of ordinary cross-validation of each group. This method is really efficient to remove the dependence within data, but it has been shown that it fails to converge to the optimal bandwidth by Chu and Marron([20]). Our method, called modified PCV, gives a correction to PCV. Its asymptotic forms show that the bandwidth chosen by MPCV is convergent if g is correctly chosen. A practical strategy will be given for the choice of g. And some simulation studies show a better performance of MPCV compared with other methods for heteroscedastic correlated data. Moreover, it has a slight numerical implementation, which is an advantageous point when working with large samples.

3.3.1 The existing automatic selections in the presence of dependence

Many methods to solve the problem of correlation were proposed independently. Chu and Marron([20]) proposed two new cross validations, *modified cross-validation* and *partitioned cross-validation*, based on the idea of leaving out the neighbour observations from two sides of a point in estimating the mean function at this point. When the correlation is short ranged, an appropriate choice of the number of left out points will greatly decrease the effect of correlation in the cross-validation. More recently, Hall et al.([68]) proposed a bootstrap based approach, called *block bootstrap* which can be applied even in the long range dependence case. Another idea for the choice of bandwidth in the dependent case is to estimate the correlation structure parametrically (by supposing the errors are an ARMA process). One method proposed by Hart [73] following this approach is *time series cross validation*. Another fully nonparametric method without, specifying the correlation function, is the *plug-in* method. Plug in bandwidth selection is performed by estimating the unknown quantities in the expression of asymptotic optimal bandwidth and replacing them by their estimators.

These methods are studied in the different contexts : kernel, loess and spline estimations and give usually good results in the asymptotic point

of view. However, they were studied only in the model with a constant variance : $X_i = m(x_i) + \varepsilon_i$ where ε is assumed to follow a covariance stationary process. We will study them in a larger model with the presence of a variance term $X_i = m(x_i) + s(x_i)\varepsilon_i$ where $s(x)$ is the scale function, ε is also assumed to follow a covariance stationary process. In some methods, we need to do some modifications to take into account the presence of non constant variance term.

Time series cross-validation will not be mentioned here because our residuals has a variance term, not placed in ARMA case. Of course, if one wants to use it in this model, a modification of TSCV is possible. The performance of these automatic selections will be studied for our specific model through a simulation study. This study is also for a comparison between our new method with the others.

3.3.1.1 Modified cross-validation (MCV)

This technique is a modification of the classical cross-validation method. MCV is proposed and studied in detail by Chu and Marron ([20]).It is simply the "leave -$(2l+1)$-out" version of CV criterion, for each $l \geq 0$.

For any $l \geq 0$, the MCV criterion can be written as

$$MCV_l(h) = \frac{1}{n}\sum_{i=1}^{n}[X_i - \hat{m}_h^{i,l}(x_i)]^2 \tag{3.20}$$

where $\hat{m}^{i,l}(x)$ is the "leave-$2l+1$-out" (remove ith observation and l observations of his left and his right) estimate of $m(x)$. When $l = 0$, MCV is ordinary cross-validation.

For the linear smoother (kernel,loess, splines...), we remark that : $\hat{m}(x) = \frac{\sum_{i=1}^{n} w_i(x)X_i}{\sum_{i=1}^{n} w_i(x)}$, so the leave-$2l+1$-out estimator is :

$$\hat{m}^{i,l}(x) = \frac{\sum_{k\notin[i-l,i+l]} w_k(x_i)X_k}{\sum_{k\notin[i-l,i+l]} w_k(x_i)}$$

After some calculations, we get :

$$X_i - \hat{m}_h^{i,l}(x_i) = \frac{X_i - \hat{m}_h(x_i)}{1 - \sum_{k\in[i-l,i+l]} w_k(x_i)}$$

So we can write (3.20) more simply :

$$MCV_l(h) = \frac{1}{n}\sum_{i=1}^{n}[\frac{X_i - \hat{m}_h(x_i)}{(1-\sum_{k=i-l}^{i+l} S_{ik})}]^2 \tag{3.21}$$

with S is the smoother matrix n x n in the formula : $\hat{m} = S\overrightarrow{X}$

l is a parameter to choose, such that X_t and X_{t+l+l} are practically independent. From the asymptotic property of MCV(see [20]) we can get the relative choice of l in practice : l can be chosen empirically as the smallest value such that $c(l) \simeq 0$ where $c(k)$ is the k order autocovariance of ε. It is however not evident because the autocovariance function of ε is unknown and so have to be estimated as in plug-in method.

3.3.1.2 Partitioned cross-validation (PCV)

Partitioned cross-validation was first proposed by Marron (1987,[97]). While in *partitioned cross-validation*(PCV), the observations are partitioned into g subgroups by taking every gth observations, for example the first subgroup consists of the observations $1,1+g,1+2g,...$, the second subgroup consists of the observations $2,2+g,2+2g,....$ Cross validation is performed for each subgroup, and the bandwidth estimate for all the observations is a simple function of the optimal bandwidth which minimize the average value of subgroup CV values :

$$PCV_g(h) = \frac{1}{g}\sum_{k=1}^{g} CV_{0,k}(h) \tag{3.22}$$

where $CV_{0,k}(h)$ is the ordinary cross-validation score for the k^{th} subgroup.

Note $\hat{h}_0$ is the minimizer of (3.22). Chu and Marron ([20]) proved that $\hat{h}_0$ is of the order $(n/g)^{-1/5}$ and since the optimal bandwidth is of the order $n^{-1/5}$ (under the previous choice of the second derivative), the partitioned cross-validation bandwidth is defined to be the rescaled $\hat{h}_0$:$\hat{h}_{PCV} = g^{-1/5}\hat{h}_0$.

When $g = 1$, PCV is simply ordinary cross-validation. g must be chosen with care. Chu and Marron showed that the asymptotic mean of the bandwidth from PCV is significantly different with the optimal bandwidth which minimizes the mean average square error with large g.

3.3.1.3 Block bootstrap (BB)

Hall et al. ([68]) proposed block bootstrap approach to choose the smoothing parameter. The main idea is to estimate directly the mean average

squared errors through resampling of "blocks" of residuals from a pilot smooth function. Hall et al.(([68])) showed that the block bootstrap has interesting empirical properties, even under long-range dependence.

The global bootstrap bandwidth used here is obtained as follows :

Let $\hat{m}_1$ and $\hat{m}_2$ denote two loess estimators of m corresponding respectively two bandwidths h_1 and h_2. $\hat{m}_1$ is used to compute centred residuals and $\hat{m}_2$ to generate bootstrap data.

Put $\hat{\varepsilon}_{1i} = X_i - \hat{m}_1(x_i)$, $\bar{\hat{\varepsilon}}_1 = 1/n \sum \hat{\varepsilon}_{1i}$ and $\hat{\varepsilon}_i = \hat{\varepsilon}_{1i} - \bar{\hat{\varepsilon}}_1$. These are the centered residuals. Construct the block bootstrap errors ε_i^* following the method of Hall et al.(([68])). We can describe it simply in our case where $x \in [0,1], x_1 = 0, x_n = 1$:

- Take l the block length where l should be of smaller order than nh_1.

- Take b which satisfies $(b-1)l < n \leq bl$. We take here $b = [\frac{n}{l}]$.

- Note $B_i = \{\hat{\varepsilon}_i, ..., \hat{\varepsilon}_{i+l-1}\}$ the block of centred residuals that starts from position i, with $i = 1, 2, ..., n-l$. Resample randomly, with replacement, b time from the sequence of all B_i, we then obtain b blocks B_j^*, with $1 \leq j \leq b$.

- Put the elements of these blocks B_j^* into a string of length bl and then we obtain a new series ε^* called *block bootstrap errors.*

Then let $X_i^* = \hat{m}_2(x_i) + \varepsilon_i^*$ and $\hat{m}_h^*(x)$ is the regression estimator from the sequence y^* by loess corresponding to the bandwidth h.

Then we look for h^* that minimizes the estimated MASE :

$$\hat{MASE}(h) = 1/n \sum [\hat{m}_h^*(x) - \hat{m}_2(x)]^2$$

Realize this procedure N times (N chosen), then we get $\hat{h}_{BB}$ as the average value of N obtained h^*.

3.3.1.4 Plug-in method

Plug-in method replace the CV-based criteria by a plug-in approach. Plug-in bandwidth selection is performed by estimating the unknown quantities in the mean squared errors and then replacing them by the estimators and minimizing the estimate mean squared errors with respect to the smoothing parameter h.

From the expression (3.19), an asymptotically optimal global bandwidth , which does not depend to the position of x, can be obtained :

$$h_0 = \left[\frac{\mu_0^2 C}{n(\mu_2^2)^2 \int (m^{(2)}(x))^2 dx}\right]^{1/5} \tag{3.23}$$

In the expression (3.23), there are two unknown elements : C and $\int (m^{(2)}(x))^2 dx$. We can estimate these quantities by using a pilot bandwidth h_{pilot}. See Francisco-Fernáldez and Vilar-Fernández ([53]) for other ways to estimate these unknown parameters.

As described in Fan et al. [45], the estimator of $m^{(r)}(x)$ the rth derivative of m at x, obtained from fitting a pth-degree local polynomial based on the bandwidth $h > 0$, for $r \leq p$ is :

$$\hat{m}^{(r)}(x;h) = r! e_{r+1}^T (\mathfrak{X}_n^T(x) W_n(x) \mathfrak{X}_n(x))^{-1} \mathfrak{X}_n^T(x) W_n(x) \vec{X}_n$$

where e_{r+1} is the $(p+1)$ x 1 vector with 1 in its $r+1$ coordinate and 0 elsewhere.

So if we want to estimate $m^{(2)}$, a fit of pth degree local polynomial with $p \geq 2$ is demanded. If we use a local fit of $p = 2$, then the estimator $\hat{m}^{(2)}$ can be written explicitly as :

$$\hat{m}^{(2)}(x) = \sum_{i=1}^{n} W_2\left(\frac{x_i - x}{h}\right) X_i \tag{3.24}$$

where $W_2(t) = 2! e_3^T U_n^{-1} (1, ht, ht^2)^T K(t)$ and $U_n = (U_{n,i+j-2})_{1 \leq i,j \leq 2}$, $U_{n,l} = \sum_{j=1}^{n} (x_j - x)^l K\left(\frac{x_j - x}{h}\right)$.

So a plug in approach will be : Estimate $m(x)$ and $m^{(2)}(x)$ by modeling the observations y by loess ($p = 2$) with bandwidth h_{pilot}. Estimate $s(x)$ by using loess from the squared residuals with same bandwidth h_{pilot}. The autocovariance function of ε can be estimated from the residuals $\frac{X - \hat{m}}{\hat{s}}$. From these $\hat{s}(x)$ and $\hat{c}(k)$, we can estimate $C(s, \varepsilon)$.

3.3.2 A new algorithm : Modified partitioned cross-validation (MPCV)

These previous methods haven been shown to work rather well for dependent data with constant variance, when the parameters (size of blocks, pilot bandwidth) are chosen "appropriately". But it is not clear or there is

not yet an algorithm to find appropriate parameters. Moreover, the numerical computations of most of these methods are heavy, especially when the dataset is large (5000 values for example). This is problematic because our series of temperature have at least a length of 5000. These troubles motivate us to find another approach which must be efficient in both the asymptotic aspect and time of implementation aspect.

We propose here a modified version of partitioned cross-validation. The computation of PCV is slight, however the choice of parameter g is not clear, poor results can be obtained if g is not appropriate.

We will first give the asymptotic results of Chu and Marron in the case of constant variance and present modified partititioned cross-validation with the appropriate form of g. Then we modify the form of g in the heteroscedastic model and the use of MPCV in pratice will be mentionned.

We remind that in PCV, $\hat{h}_0$ is the minimizer of :

$$PCV_g(h) = \frac{1}{g}\sum_{k=1}^{g} CV_{0,k}(h)$$

where $CV_{0,k}(h)$ is the ordinary cross-validation score for the k^{th} subgroup.

Inheriting the results of Chu and Marron and ignoring the weight function W (added to avoid the boundary effect) for model (3.17) with s constant, we have :

$$\hat{h}_0 = C_{PCV(g)} g^{1/5} n^{-1/5}(1 + o(1)) \tag{3.25}$$

On the other hand, they showed that the optimal bandwidth h_M, as the minimizer of the mean average square error (MASE), can be asymptotically expressed as :

$$h_M = Cn^{-1/5}(1 + o(1)) \tag{3.26}$$

where $C_{PCV(g)} = [V_{PCV(g)}/4B_2]$, $C = [V/(4B_2)]$ with :

$$B_2 = \frac{1}{4}\mu_2^2 \int (m")^2, V = \sum_{k=-\infty}^{\infty} c(k)s^2\nu_0,\ V_{PCV} = \sum_{k=-\infty}^{\infty} c(gk)s^2\nu_0 - 4s^2K(0)\sum_{k>0} c(gk)$$

Chu and Marron define $h_{PCV} = h_0 g^{-1/5}$ but C_{PCV} is significantly different from C for all values of g, h_{PCV} then is significantly different from h_M. This difference can be seen in the simulation study in the next subsection.

To modify this problem, we define $h_{MPCV} = h_0$ and to approach $\hat{h}_{MPCV}$ and h_M, we have to approach $C_{PCV(g)}g^{1/5}$ with C, that means we must have :

$$\left| \sum_{k=-\infty}^{\infty} c(k)s^2\nu_0 - g^{1/5}\left(\sum_{k=-\infty}^{\infty} c(gk)s^2\nu_0 - 4s^2K(0)\sum_{k>0} c(gk)\right)\right| \approx 0 \quad (3.27)$$

When s varies with time,B_2 is the same but $V_{PCV(g)}$ changes :

$$\begin{aligned} V_{PCV(g)} =& 1/n \sum_{0\leq i\leq n} \sum_{k=[(1-i)/g]}^{[n/g]} Cov(s(x_i)\varepsilon_i, s(x_{i+gk})\varepsilon_{i+gk})\nu_0 \\ & - 4/nK(0) \sum_{0\leq i\leq n} \sum_{k=1}^{[(n-i)/g]} Cov(s(x_i)\varepsilon_i, s(x_{i+kg})\varepsilon_{i+kg}) \end{aligned}$$

We develop this expression into the autocovariance function of ε :

$$\begin{aligned} C_2(g,s,\varepsilon) &= 1/n \sum_{0\leq i\leq n} \sum_{k=1}^{[(n-i)/g]} Cov(s(x_i)\varepsilon_i, s(x_{i+kg})\varepsilon_{i+kg} \\ &= 1/n \sum_{i=1}^{n-g} \sum_{k=1}^{[(n-i)/g]} s_i s_{i+kg} c(kg) \end{aligned} \quad (3.28)$$

$$\begin{aligned} C_1(g,s,\varepsilon) &= 1/n \sum_{0\leq i\leq n} \sum_{k=[(1-i)/g]}^{[n/g]} Cov(s(x_i)\varepsilon_i, s(x_{i+gk})\varepsilon_{k+gj}) \\ &= 1/n \sum_{i=1}^{n} s_i^2 c(0) + 2C_2(g,s,\varepsilon) \end{aligned} \quad (3.29)$$

And the optimal g satisfies :

$$\left| C\nu_0 - g^{1/5}(C_1(g,s,\varepsilon)\nu_0 + 4K(0)C_2(g,s,\varepsilon))\right| \approx 0 \quad (3.30)$$

In practice, we execute the modified PCV approach as follows :

- Estimate h^g_{MPCV} with different values of g.

- Retain the values of g for which $\hat{h}^g_{MPCV}$ is not absurde, for example not too near to 0.

- For each retained value g, estimate the trends m and s by loess with bandwidth $\hat{h}^g_{MPCV}$, the estimation procedure is explained in section 3.2, we then obtain an estimator $\hat{\varepsilon}$ and the estimator of autocovariance $c(k)$.

- With these estimators for each retained g, estimate $C, C_1(g,s,\varepsilon)$ and $C_2(g,s,\varepsilon)$. Calculate the expression $\mid C\nu_0 - g^{1/5}(C_1(g,s,\varepsilon)\nu_0 + 4K(0)C_2(g,s,\varepsilon)) \mid$ and keep the value of g which minimizes this expression and corresponding $\hat{h}^g_{MPCV}$

3.3.3 Simulation study

A simulation study will be carried out to compare the bandwidth selections presented in the previous section. Different regression models are considered : less variable and more variable, constant variance and variable variance. The values are taken on an equally spaced design in the unit interval. The correlated noise is taken from the ARMA family.

We take some criteria to compare the goodness-of-fit of these methods :

i/ Optimal bandwidth from the measurement of discrepancy :

$$MASE(h) = \frac{1}{n}\sum_{i=1}^{n}(\hat{m}_h(x_i) - m(x_i))^2 \tag{3.31}$$

By minimizing (3.31), we can find out the optimal bandwith with respect to MASE. By taking 100 samples, we obtain h_{MASE} the average value of 100 optimal bandwidths with respect to MASE of each sample.

ii/ Mean average squared errors : we calculate (3.31) for each bandwidth obtained from the different methods.

3.3.3.1 Simulation study for all methods

For each method, we take 50 samples of size n, and the retained bandwidth is the average value of the obtained bandwidth from each sample. For the methods which demand a pilot parameter (bandwidth or size of blocks), we try it with some values and the best result is retained (with respect to the value of MASE).

• For h_{MCV} : the choice of l to remove $2l+1$ in each smoothing window is important. We try with $l = 2, 5, 10$.

• For h_{PCV} : Take g respectively from 3 to 10.

• For h_{BB} : Let B be the number of bootstrap replicas, take $B = 50$. We try the size of block $l = 2, 5, 10, 20$

• For h_{MPCV} : Execute the procedure describe above with g varying from 3 to 10.

Model 1 : $m(x) = 5 + sin10x$, $s(x) = 0.25$, $n = 500$, ε is an AR(1) with $\rho = 0.5$.

Below we obtain the results of these selection methods :

	MASE	MCV	PCV	BB	PI	MPCV
h	0.166	0.190	0.125	0.145	0.176	0.170
MSE		0.004	0.007	0.007	0.006	0.003

TABLE 3.1 – $h_{MASE}, h_{MCV}, h_{PCV}, h_{BB}, h_{PI} and h_{MPCV}$ with their estimated mean squared errors in Model 1

In Table 3.1, the asymptotic optimal bandwidth and the bandwidths obtained by the automatic selections with their estimated mean squared errors are shown.

Model 2 : Same $m(x)$, errors ε as in Model 1 but variable s : $s(x) = 0.4x + 1$, $n = 500$

The results are presented in Table 3.2 :

	MASE	MCV	PCV	BB	PI	MPCV
h	0.290	0.350	0.220	0.280	0.240	0.310
MSE		0.090	0.100	0.075	0.300	0.080

TABLE 3.2 – $h_{MASE}, h_{MCV}, h_{PCV}, h_{BB}, h_{PI} and h_{MPCV}$ with their estimated mean squared errors in Model 2

Conclusion

In our study, we use only the methods which do not have any assumption on the correlation structure. In fact, a specific assumption for the errors can give negative results because in our model, the variance is variable and itself needs to be estimated. A misspecification on the structure of ε can give a bad result in estimating $s(x)$. In our case, the selectors that do not suppose any parametric structure for the errors could have better performance.

Modified cross-validation and block bootstrap work rather well in both cases : constant variance and variable variance. Block bootstrap however is better compared with MCV in the case of variable variance and works much more rapidly. MCV takes a much heavier computation time with respect to

PCV and block bootstrap due to the computation of the smoother matrix. Block bootstraps demands the pilot parameters, however it is rather stable with the different pilot bandwidths and size of block l.

Partitioned cross-validation does not give good results even when we choose g with care (We tried many values of g and retained the best). This confirms the remark in Chu and Marron ([20]) : this method is effective in removing the dependence but there is a significant distance between the h_{PCV} and the optimal bandwidth.

Plug-in gives a good result when the variance is constant, however the result for the case of variable variance is bad. The reason for this is that when s is constant, the loss can be found in the estimator of second derivative of m and autocovariance function of ε, but when s is not constant, the loss can be found also in the estimator of s. It is difficult to correctly estimate at the same time all these unknown quantities, and the result is that we can obtain a poor bandwidth.

In both models, modified partitioned cross-validation give good results. Moreover, the optimal value of g is rather stable with different simulations. And the advantage of MPCV, like PCV, is that it works numerically very quickly. A very important remark for MPCV is that a bad choice of g immediately gives an explosion in the results. Concretely, for model 1, the optimal g is 5. If we take $g = 4$, $g = 6$, we obtain respectively $h_{MPCV} = 0.145$ and $h_{MPCV} = 0.18$. But when we take $g = 7$, we obtain an absurd value $h_{PCV} = 0.05$. This underestimated bandwidth is also obtained with ordinary cross-validation.

The MPCV method, if it works correctly, is really promising due to its numerical convenience. Now in oder to confirm the efficiency of MPCV and also its stability with the number of observations, we will develop a detailed simulation study with different trends m, s, correlation structures of ε and different values of n. Then an application will be done on the temperature series.

3.3.3.2 Detailed simulation study on MPCV

$\star$ Simulation study

In the following tables, in order to make the lecture easier, bad results will be marked in bold.

Model 3 : $m(x) = 10+sinx$, $s(x) = 2+sinx$ and $n = 300, 500, 1000, 5000$. In this model, m and s are rather linear.

We will consider different correlation structures :

• ε is an AR(1) with different φ, φ = 0.1,0.2,0.3,0.4,0.5,0.6 .

In Table 3.3, the optimal bandwidth h_{MASE} and the results when applying MPCV : bandwidth h_{MPCV}, the optimal value of g and the loss function (MASE) with $\hat{m}$ corresponding to h_{MPCV} are given.

n	φ	h_{MASE}	g	h_{PMCV}	MASE
300	0.1	0.9	3	0.85	0.07
300	0.2	0.88	3	0.84	0.1
300	0.3	0.89	3	0.76	0.16
300	0.4	0.9	3	0.7	0.33
300	**0.5**	**0.9**	**3**	**0.45**	**1**
300	**0.6**	**0.9**	**3**	**0.24**	**2.26**
500	0.1	0.9	5	0.87	0.048
500	0.2	0.89	5	0.91	0.044
500	0.3	0.89	5	0.89	0.06
500	0.4	0.91	5	0.84	0.1
500	0.5	0.91	5	0.824	0.22
500	0.6	0.9	6	0.72	0.34
1000	0.1	0.897	7	0.895	0.02
1000	0.2	0.9	7	0.88	0.03
1000	0.3	0.91	7	0.9	0.034
1000	0.4	0.9	9	0.9	0.05
1000	0.5	0.9	9	0.89	0.05
1000	0.6	0.9	9	0.89	0.1
5000	0.1	0.88	8	0.9	0.004
5000	0.2	0.9	8	0.89	0.005
5000	0.3	0.88	9	0.87	0.007
5000	0.2	0.9	10	0.89	0.009
5000	0.4	0.88	11	0.89	0.01
5000	0.6	0.9	12	0.91	0.028

TABLE 3.3 – h_{MPCV} and their estimated mean squared errors with different n in Model 3 with an AR(1) errors

• ε is an AR(3) with $\varphi = (0.8, -0.4, 0.3)$. The results are presented in Table 3.4

n	h_{MASE}	h_{PMCV}	MSE
300	**0.92**	**0.1**	**4.08**
500	**0.91**	**0.45**	**1.34**
1000	0.88	0.85	0.15
5000	0.89	0.88	0.02

TABLE 3.4 – h_{MPCV} and their estimated mean squared errors with different n in Model 3 with an AR(3) errors

Model 4 : $m(x) = 5+sin10x$, $s(x) = 1+0.4x$ and $n = 300, 500, 1000, 5000$

The functions $m(x)$ and $s(x)$ are more variable than those of Model 3. We will consider different correlation structures :

• ε is an AR(1) with different $\varphi = 0.1, 0.6$. The results are presented in Table 3.5.

• ε is an AR(3) with $\varphi = (0.8, -0.4, 0.3)$. The results are presented in Table 3.6.

Conclusion

From the previous results, we can draw the following interesting conclusions :

The change of the optimal bandwidth got from the mean squared errors with the sample size is not really significant when the true function m oscillates slightly . However when m *oscillates much*, for example in Model 4, h_{MASE} has a tendency to increase when the correlation increases and it decreases with the sample size n.

For different mean, scale functions and models of the errors, the results show that MPCV performs well. When the size n increases, h_{MPCV} approaches the optimal bandwidth of MASE h_{MASE}. From a value of n larger than 1000, we can obtain good results from modified partitioned cross-validation.

When the correlation is weak, although with a small sample, MPCV can draw a rather good result. However, when the correlation of the data is strong, we can see that MPCV has difficulties to approach the optimal bandwidth when there are not enough observations. When the correlation increases, MPCV demands naturally a higher g but if we work with a small sample, this means that in each partition there are few values. So when ap-

n	φ	h_{MASE}	g	h_{PMCV}	MASE
300	0.1	0.266	2	0.26	0.05
300	0.2	0.27	3	0.29	0.06
300	0.3	0.287	3	0.278	0.087
300	0.4	0.3	3	0.236	0.176
300	0.5	0.32	3	0.217	0.29
300	**0.6**	**0.35**	**3**	**0.11**	**0.54**
500	0.1	0.237	3	0.25	0.039
500	0.2	0.249	3	0.248	0.048
500	0.3	0.26	4	0.279	0.054
500	0.4	0.276	5	0.29	0.077
500	0.5	0.287	5	0.284	0.1
500	0.6	0.326	6	0.32	0.13
1000	0.1	0.212	2	0.2	0.018
1000	0.2	0.217	3	0.219	0.025
1000	0.3	0.23	4	0.24	0.03
1000	0.4	0.244	4	0.246	0.037
1000	0.5	0.268	5	0.266	0.06
1000	0.6	0.275	6	0.277	0.08
5000	0.1	0.158	2	0.159	0.005
5000	0.2	0.155	2	0.154	0.0056
5000	0.3	0.164	3	0.168	0.007
5000	0.2	0.177	4	0.167	0.008
5000	0.4	0.196	5	0.192	0.011
5000	0.6	0.2	6	0.195	0.018

TABLE 3.5 – h_{MPCV} and their estimated mean squared errors with different n in Model 3 with an AR(1) errors

n	h_{MASE}	h_{PMCV}	MSE
300	**0.42**	**0.09**	**0.88**
500	**0.40**	**0.24**	**0.37**
1000	0.32	0.3	0.12
5000	0.25	0.24	0.028

TABLE 3.6 – h_{MPCV} and their estimated mean squared errors with different n in Model 4 with the AR(3) errors

plying ordinary cross-validation in each partition, because of a lack of data, the bandwidth cannot be well estimated. For example when we consider Model 3 with an AR(1) of coefficient $\varphi = 0.6$ and sample size $n = 300$, when we take $g > 3$, we obtain $h_{MPCV} = 0.05$. The best result obtained in this case is when $g = 3$ but even the best bandwidth is still far from the optimal bandwidth.

The asymptotically optimal number of partitions g can change with the sample size. The change of g depends on the mean function m. For Model 3, when m has a positive growth with time, g increases rapidly in function of n, but for a oscillating trend as in Model 4, g is almost constant with n.

3.3.3.3 Application of MPCV to the temperature series

This paragraph is devoted to the application of MPCV on the temperature series. We will consider two series with different length : the series of daily maximum temperature in Orleans (1946-2006) and the series of daily maximum temperature in Déols (1901-2006). The applied procedures will be explained in detail for each series. The data consist of the maximum values of temperature in each summer day (summer is defined as 100 days : from 14th June to 21st September)

- Temperature series of Orleans :

The temperature in Orleans is measured over 60 years. The temperature series comprises 6100 observations, its length is smaller than the other series for Déols. We apply MPCV on this series to estimate the mean trend. g varies from 1 to 20 and for each value of g, we calculate its corresponding h_{MPCV}, then we have the following results :

h_{MPCV} = 0.05, 0.05, 0.29, 0.29, 0.32, 0.56, 0.56, 0.56, 0.56, 0.56, 0.56, 0.56, 0.56, 0.56, 0.56, 0.62, 0.62, 0.62, 0.62, 0.62.

From this result, we will consider $g = 3, 4, 5, 6, 10, 16$ and calculate the expression C_g , the left term in (3.30) for each g. We obtain then :

g	3	4	5	6	10	16
C_g	22.23	2.53	1.4	5.88	26.08	59.3

From that, $g = 5$ is retained and the bandwidth chosen from MPCV is 0.32.

- Temperature series of Déols :

We realize the same procedure to the temperature series of Déols but this time the series is longer, it contains 13100 observations. Applying MPCV with g varies from 1 to 20, we obtain :

h_{MPCV}= 0.05, 0.05, 0.05, 0.05, 0.05, 0.08, 0.17, 0.17, 0.17, 0.17, 0.20, 0.32, 0.32, 0.32, 0.53, 0.47, 0.53, 0.47, 0.53, 0.53.

We will consider then $g = 7, 10, 11, 12, 14, 15$ and calculate the corresponding quantity C_g, we get :

g	7	10	11	12	14	15
C_g	2.98	7.63	6.17	3.90	12.02	4.14

$g = 7$ minimizes C_g, so we retain $g = 7$ and the chosen bandwidth is 0.17.

3.3.4 Conclusion on the new algorithm

Compared with the other selection methods, our new approach shows its clear advantage. For the asymptotic aspect, with a sample size large enough, the bandwidth given by MPCV approaches the optimal bandwidth. This is true for different kinds of mean, variance functions and structures of correlations of the errors (in the ARMA family). On the other hand, it shows the numerical convenience which allows us to work with large samples, as we have seen in the case of temperature series in Déols with more than 10000 observations.

For these positive points, we will use MPCV to choose appropriate bandwidths when applying loess to estimate the trends of the temperature series in Europe.

3.4 Application of splines

We place ourselves here in a different case than in the previous one. The distribution of the i.i.d observations $X_1, \ldots, X_n$ is known. The approach by maximizing the likelihood function is preferred. For nonparametric estimations, it can be a local likelihood function for loess, or a penalized likelihood function for splines. In numerical implementation, when applying these two approaches to estimate the extreme parameters in extreme models (GEV or POT), splines (with maximizing penalized likelihood) shows its advantages in computing time compared to loess (with maximization the local likelihood). Moreover, when the bandwidth in loess is too small, we can have breaks characterized by some sudden jumps in the values of the estimators (see Davison and Ramesh, [33] for extreme-events modeling). Therefore, we will use splines to estimate extreme parameters.

Whittaker ([149]) originated the idea of smoothing spline. The modern formulation of smoothing spline as a nonparametric regression estimator and its asymptotic properties are developed by Schoenberg ([134]), Reinsch([124]), Wahba([146])and recently, we have the works of Hastie and Tibshirani ([75]), Green and Silverman ([59]), Gu([61]), Cox and O'Sullivan([26],[27]). The basic idea of smoothing spline is to fit the functions by piecewise polynomials. An important point in smoothing splines is the presence of *roughness (or smoothness) penalty*. Roughness penalties are motivated by the fact that without imposing restrictions on the smooth functions, for example in fitting a model by least squares, the residual sum of squares can always be reduced to zero by choosing a function that interpolates the data. The roughness penalty function commonly taken is $\int_0^1 f''(t)^2 dt$.

Natural cubic splines are defined for all piecewise polynomials with the additional constraint that the function is linear beyond the boundary points (zero second and third derivatives at the boundaries). When maximizing the penalized likelihood in the family of natural cubic splines, we find a unique natural cubic spline solution. When there are at least two parameters to estimate, the solution tends to solve a set of estimating equations by a backfitting procedure with a Fisher scoring or a Newton- Raphson algorithm. However, when the knots x_i are not regularly distributed and the number of observations n is large, we can face difficulties when estimating with a natural cubic spline. One solution is to restrict the choice of smooth functions to a space of basis functions. *Cubic B-splines* is often chosen in numerical functions in R. Both approaches are equivalent, but cubic B-splines gives faster computation. For more details about cubic B-splines, see Green and Silverman([59]).

Also in this section, an estimate algorithm for multi-parameters combining with the choice of smoothing parameters will be proposed. This algorithm is based on the equivalence of the iteratively reweighted least squares with backfitting within a Fisher scoring or a Newton- Raphson algorithm, and performance-oriented iteration of Gu ([60]) (can be considered as indirect cross-validation). The empirical performance of this estimation algorithm will be illustrated through simple simulations.

3.4.1 The estimators

Suppose we have the real numbers x_i called *knots* $0 < x_1 < x_2 < ... < x_n < 1$. A function g defined on $[0, 1]$ is a *cubic spline* if it satisfies the two following conditions :

1/ On each interval $(0, x_1), (x_1, x_2), ..., (x_n, 1)$, g is a cubic polynomial.

2/ g and its first and second derivatives are continuous at each x_i and hence on $[0,1]$.

g can be written in the form :

$$g(x) = \theta_{0,i}+\theta_{1,i}(x-x_i)+\theta_{2,i}(x-x_i)^2+\theta_{3,i}(x-x_i)^3 \text{ for } x_i \leq x \leq x_{i+1} \quad (3.32)$$

where $i = 0, .., n+1, x_0 = 0$ and $x_{n+1} = 1$.

Suppose that we observe a sequence $X_1, ..., X_n$ of i.i.d random variables with known distribution.

We immediately attack the case of multi-parameters, which means we have at least two parameters θ and φ to estimate . Fixing two smoothing parameters λ, β, the problem is equivalent to maximize the penalized log-likelihood depending on unknown functional parameters θ and φ :

$$l_{n,\lambda,\beta} = \frac{1}{n}\sum_{i=1}^{n} l(X_i,\theta,\varphi) - \frac{1}{2}\lambda \int \theta"(t)^2 dt - \frac{1}{2}\beta \int \varphi"(t)^2 dt \qquad (3.33)$$

If we use natural cubic splines, the discretized form of (3.33) is :

$$l_{n,\lambda_n,\beta_n} = \frac{1}{n}\sum_{i=1}^{n} l(X_i,\theta,\varphi) - \frac{1}{2}\lambda\theta^T K\theta - \frac{1}{2}\beta\varphi^T K\varphi \qquad (3.34)$$

where $K = \Delta^T C\Delta$.

Let $h_i = t_{i+1} - t_i$ for $i = 1, ..., n-1$. Δ is a $n \times (n-2)$ matrix with entries Δ_{ij}, for $i = 1, ..., n$ and $j = 1, ..., n-2$, given by :

$$\Delta_{jj} = \frac{1}{h_j}, \Delta_{j,j+1} = -\frac{1}{h_j} - \frac{1}{h_{j+1}}, \Delta_{j,j+2} = -\frac{1}{h_{j+1}}$$

and $\Delta_{ij} = 0$ for $\mid i-j \mid > 1$. C is a symmetric $(n-1)\times(n-1)$ matrix with elements :

$$C_{jj} = \frac{1}{3}(h_{j-1}+h_j), C_{j,j+1} = C_{j+1,j} = \frac{1}{6}h_j$$

and $C_{ij} = 0$ for $\mid i-j \mid > 1$.

The solution of this maximization problem leads to use an iterative Fisher scoring or an iterative Newton- Raphson algorithm.

In Fisher scoring, the following process must keep being iterated :

$$\begin{bmatrix} W_{\theta\theta}+\lambda K & W_{\theta\varphi} \\ W_{\varphi\theta} & W_{\varphi\varphi}+\beta K\beta \end{bmatrix} \begin{bmatrix} \theta^{new}-\theta \\ \varphi^{new}-\varphi \end{bmatrix} = \begin{bmatrix} u_\theta - \lambda K\theta \\ u_\varphi - \beta K\varphi \end{bmatrix}$$

where $W_{\theta\varphi} = diag((E\left[-\frac{\partial^2 l}{\partial\theta_1\partial\varphi_1}\right], ..., E\left[-\frac{\partial^2 l}{\partial\theta_n\partial\varphi_n}\right])^T)$ and $u_\theta = (\frac{\partial l}{\partial\theta_1}, ..., \frac{\partial l}{\partial\theta_n})^T$

For the Newton Raphson algorithm, W is replaced by the observed minus second derivative $W_{\theta\varphi} = diag((-\frac{\Sigma^2 l}{\partial\theta_1\partial\varphi_1}, ..., -\frac{\Sigma^2 l}{\partial\theta_n\partial\varphi_n})^T)$.

It turns out to be the alternation of the two estimate processes solving repeatedly for θ and φ respectively. This procedure is called the backfitting algorithm (just for multi-parameters). Backfitting combining with Newton-Raphson (Fisher scoring) converges to the penalized maximum likelihood estimate. The convergence of backfitting algorithm will be discussed in the next subsection.

The smoother matrix in the final iteration , for example for $\hat{\theta}$ corresponding to the smoothing parameters λ, is calculated by :

$$S_\theta = W_{\theta\theta}(W_{\theta\theta} + \lambda K)^{-1} \tag{3.35}$$

The degree of freedom of the smooth $\hat{\theta}$ is $trace(S_\theta)$.

3.4.2 Asymptotic properties

There are two problems of convergence that need to be considered : the first one is the existence and convergence of backfitting with Newton- Raphson (Fisher scoring) algorithm to the solution of maximization (3.34), the second one is the convergence or asymptotic properties of the penalized likelihood type estimators (with respect to the true parameters).

3.4.2.1 Convergence of the backfitting procedure

The condition of convergence of the backfitting algorithm is explained in Hastie et al.([75]). We can summarize the main results for the two-parameter case, noted S_θ and S_φ respectively the true smoother matrix of $\hat{\theta}_\lambda$ and $\hat{\varphi}_\beta$, we have :

-If $\| S_\theta S_\varphi \| < 1$, then the backfitting algorithm converges to the unique solution, which is independent on the initial values of the parameters.

-If S_θ and S_φ are symmetric with eigenvalues in (-1,1], then the backfitting algorithm converge to one solution which depends on the starting parameters.

The asymptotic properties of the penalized likelihood estimators are rather complicated because they depend not only on the characteristics of the true functions -their smoothness properties, of course, and their behavior at the boundaries - but also the behaviour of the smoothing parameters. For the sake of simplicity, we only discuss here the case of one parameter θ. The extension to the multiparameter case presents no difficulty.

The asymptotic properties of the penalized likelihood estimator are discussed in details in the work of Cox and O'Sullivan([27]). Here we just summarize their results applied to our special case : x is equispaced fixed design on $[0, 1]$.

We begin with some basic assumptions :

B.1. $X_i, 1 \leq i \leq n$ are i.i.d elements with X_i taking values in $\mathbb{R}$.

B.2. The parameter space $\theta : [0, 1] \to \mathbb{R}$ with inner product $< ., . >$ and norm $\| . \|$ is a Sobolev space :
$W_2^s([0, 1]; \mathbb{R}) = \{\theta : [0, 1] \to \mathbb{R} \mid \theta^{(1)}\}$ *are absolutely continuous and* $\theta^{(p)} \in L_2[0, 1]$ *with* $p \leq s\}$.

B.3. For some bounded strictly positive constants K_1 and K_2, the penalty function $J(\theta) = \int (\theta^{(2)})^2$ satisfies :

$$K_1 \parallel \theta \parallel^2 \leq J(\theta) + \int_0^1 \theta(t)^2 dt \leq K_2 \parallel \theta \parallel^2$$

for all $\theta \in \Theta$.

B.4. The true function parameter θ_0 is in W_2^s for some s satisfying $3/2 < s \leq 2$.

We note that $\theta_0 \in \Theta$ is not required, which means that the estimates can converge to a function which is not as smooth as the functions in Θ.

We have the following theorem :

Theorem 3.4.1 *(derived from Theorem 1 in [27])*
Suppose there is an α *satisfying :*

$$1/2 < \alpha < s/4 - 1/4 \tag{3.36}$$

and a sequence $\lambda_n \to 0$ *satisfying*

$$n^{-1} \lambda_n^{-(2\alpha+1/2)} \to 0 \tag{3.37}$$

Then under assumptions B.1 through B.4, given $\varepsilon > 0$ *, there exist constants* M *and* n_0 *such that* $\forall n > n_0$*,*

$$P\{\exists\ \theta_{n,\lambda_n} \text{ such that the first derivative } Dl_{n,\lambda_n}(\theta_{n,\lambda_n}) = 0, \text{ and } \forall b \in [0,\alpha], \\ \| \theta_{n,\lambda_n} - \theta_0 \|_b \leq M[\lambda_n^{(s/2-b)/2} + n^{-1/2}\lambda_n^{-(b+1/4)/2}]\} > 1 - \varepsilon \tag{3.38}$$

The first term on the right side of the inequality in (3.38) gives an upper bound on the order of the asymptotic bias and the second term gives the order of the asymptotic standard deviation. The optimal upper bound of the convergence rate is obtained when the equality is set. The order of λ then :

$$\lambda_n \approx n^{-2/(2s+1)} \tag{3.39}$$

which results in

$$\| \theta_{n\lambda} - \theta_0 \|_b^2 = O(n^{-2(s/2-b)/(2s+1)}) \tag{3.40}$$

We can choose α so that both (3.36) and (3.37) hold when λ_n is given by (3.39).

Following these results, we can see that the asymptotic properties of penalized likelihood estimators for interior points depend on the characteristics of the smoothness properties of the true parameters. The rate of convergence is still held near the boundaries if θ satisfies the natural boundary condition $\theta''(0) = \theta''(1) = \theta^{(3)}(0) = \theta^{(3)}(1) = 0$(Rice and Rosenblatt [125]) . These conditions are in fact satisfied for the case of cubic splines.

3.4.3 Estimation algorithm for multi-parameters with bandwidth selection

This subsection is devoted to an algorithm which allows us to estimate two or more parameters by cubic splines. In the previous parts, we always presented the estimate procedure in the context of given smoothing parameters. But these parameters are unknown and an automatic selection is then required. In our algorithm, the automatic choice of the smoothing parameters is also proposed. This estimate algorithm is a combination of iteratively reweighted least squares and performance -oriented iteration cross-validation of Gu ([60]). We will first present the basic rules of these points in the one parameter case.

3.4.3.1 Iteratively reweighted least squares

As introduced in the introduction of this section, a procedure equivalent to the Fisher scoring or Newton-Raphson algorithm is the iteratively reweighted least squares. In the single parameter case (θ), when fixing the smoothing parameter λ, the problem of maximizing penalized likelihood can be solved via **iterating** on minimizing the penalized weighted least squares (see O'Sullivan et al., [109]) :

$$\sum_{i=1}^{n} \tilde{w}(t_i)(\tilde{z}(t_i) - \theta(t_i))^2 + \lambda \int (\theta_j''(t))^2 dt \tag{3.41}$$

where $\tilde{w}(t_i) = I_{\theta\theta}(\tilde{\theta}(t_i)), \tilde{z}(t_i) = \tilde{\theta}(t_i) + \frac{\partial l}{\partial \theta}(\tilde{\theta}(t_i))/\tilde{w}(t_i)$, $W_{\theta\theta}$ is the Fisher information of θ and $\tilde{\theta}(t_i)$ is the estimation of the preceding iteration. The $\{\tilde{w}(t_i)\}$ and $\{\tilde{z}(t_i)\}$ are called the pseudo data based on $\tilde{\theta}$.

This process is very interesting because when using iteratively reweighted least squares, we can obtain numerically faster algorithms and the bandwidth selectors can be more easily applied.

3.4.3.2 Iterative cross-validation

In the bandwidth selection of penalized likelihood-type estimations, a general method is to compute the cross-validation (CV) in the final iteration and the minimum of those CV gives the optimal smoothing parameters as suggested by O'Sullivan, Yandell and Raynor ([109]). However, the cost of numerical computation of this method is expensive. Another method, proposed by Gu ([60]), uses the cross-validation score (or maybe another selection criterion) through the penalized weighted least square (3.41) which is calculated from the pseudo data based on the previous estimate in a so-called performance-oriented iteration. We call this selector "iterative cross-validation". Iterative CV tracks an appropriate loss in an indirect manner and while it may not be the most effective, the simultaneous updating of smoothing parameter makes it numerically efficient.

Performance-oriented iteration, in the case of one parameter θ, can be described as follows :

- Choose the initial values for θ.

- Iterate on (3.41) :

• In each k^{th} step :

i/ choose a new $\lambda^{(k)}$ minimizing the cross-validation which is calculated from the pseudo data based on the previous estimate $\theta^{(k-1)}$ that could be written as :

$$CV(\lambda \mid \tilde{z}) = \sum_{i=1}^{n} \frac{(\tilde{z}_i - \theta_{\lambda,i})^2 \tilde{w}_i}{(1 - A(\lambda)_{ii})^2} \tag{3.42}$$

ii/ update θ which corresponds to optimal $\lambda^{(k)}$.

• Repeat i/, ii/ until θ and λ converge.

The condition of convergence : θ is said to be convergent if :

$$\sqrt{\frac{(\theta^{new} - \theta^{old})^2}{\theta^{old^2}}} < \delta \approx 0 \tag{3.43}$$

For example $\delta = 10^{-7}$

3.4.3.3 A new algorithm

Combining the two procedures above, we derive an algorithm to estimate multi-parameters of penalized likelihood type. The basic idea is to use backfitting with reweighted least squares alternatively for these parameters, and in each iteration, for each parameter, we use iterative cross-validation to update its corresponding smoothing parameter. Suppose θ and φ are two parameters to estimate. This algorithm can be described as :
i/ Initialize $\theta = \theta^0, \varphi = \varphi^0$.

ii/ Fix φ and minimize iterative CV (3.42) until convergence to find out optimal θ(corresponding to fixed φ). Update θ.

iii/ Fix θ(the value obtained from step ii/) and find out corresponding optimal φ with the same method as for θ. Update φ.

Repeat ii/, iii/ until the convergence of the two parameters following the condition (3.43).

This optimization procedure can overcome the effect of correlation between the estimate parameters in bandwidth selection. This procedure is usually convergent if the initial parameters are suitably chosen. The non-convergence can occur if the starting value is far from the optimal region. On the other hand, when the initial values are reasonably chosen, the algorithm converges rapidly. As discussed in Gu ([60]), the larger the smoothing parameters, the more quadratic (3.41) will be. For a large fixed smoothing

parameter, one can in general insure the convergence of the Newton iteration starting from an arbitrary point. Therefore we would start with rather "sufficiently smooth" functional parameters, corresponding to large smoothing parameters, for example constant θ^0 and φ^0. The empirical performance of this estimate algorithm will be justified through simulation.

3.4.3.4 Simulation study

To verify the empirical performance of the previous estimate algorithm, we will realize some simple simulations. We will consider here the simulations of the non-stationary GEV distribution. The reason to try this distribution is that we will apply this algorithm to extreme models and in any case, it is one of the most complicated distribution with three parameters.

We consider three GEV models with different time-varying μ, σ, ξ. We first consider a model with constant ξ : in one ξ is positive and in another ξ is negative. Another model with all three time-varying parameters will also be considered.

$$\begin{cases} \mu(t) = 5 + sin10t \\ \sigma(t) = t \\ \xi(t) = -0.1 \\ \text{where } t = (1:500)/500 \end{cases} \tag{3.44}$$

$$\begin{cases} \mu(t) = 5 + sin10t \\ \sigma(t) = t \\ \xi(t) = 0.2 \\ \text{where } t = (1:500)/500 \end{cases} \tag{3.45}$$

and

$$\begin{cases} \mu(t) = 10 + t + t^2 + 5t^4 \\ \sigma(t) = 2 - t + t^2 \\ \xi(t) = -0.3 - 0.1t \\ \text{where } t = (1:500)/500 \end{cases} \tag{3.46}$$

For each model, we realize 50 simulation samples. Then, on each sample, we estimate the extreme parameters following the estimate algorithm which is described above. The empirical confidence intervals at 90% for μ, σ and ξ are constructed from those estimates. To consider the quality of the estimators, we calculate the average difference between the confidence intervals from the simulations with the true values of μ, σ and ξ(with I_{sup} is the upper confidence band and I_{inf} is the lower confidence band). The results from three models are illustrated below :

Model	Average $I\mu_{sup} - \mu$	Average $I\mu_{inf} - \mu$	Average $I\sigma_{sup} - \sigma$	Average $I\sigma_{inf} - \sigma$	Average $I\xi_{sup} - \xi$	Average $I\xi_{inf} - \mu$
(3.44)	0.09	0.08	0.06	0.09	0.08	0.06
(3.45)	0.07	0.08	0.062	0.09	0.09	0.065
(3.46)	0.26	0.21	0.19	0.11	0.06	0.04

The algorithm can give us the good confidence intervals for the parameters even when the true parameters are rather variable like in Model (3.46). The fact that ξ is negative and positive does not change the empirical performance of the estimation method. A little improvement can be remarked in the estimation of μ and σ in model (3.44) where ξ is constant compared to the ones of model (3.45) where ξ slightly varies.

These results justify the asymptotic convergence of this estimate procedure. This approach is not only valid for extreme models, but can be applied for all distributions where we need to estimate the parameters nonparametrically.

In Figure 3.1 for visualization purposes, we present the confidence intervals for Model (3.44) with a constant ξ and Model (3.46) with a varying ξ.

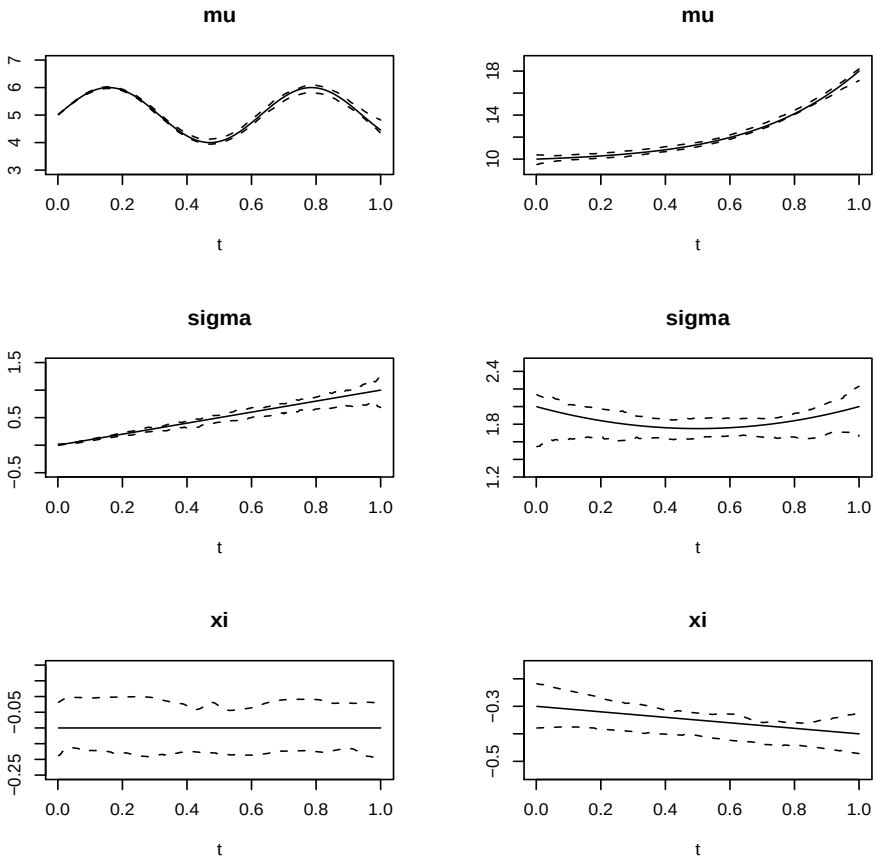


FIGURE 3.1 – Estimates of extreme parameters from the simulated samples.Left panel. Model (3.44). Right panel. Model (3.46). The dark black lines are the true parameters.

Chapitre 4

Mean and variance evolutions of the hot and cold temperatures in Europe

This chapter corresponds to our article [112]. In this paper, we examine the trends of temperature series in Europe, for the mean as well as for the variance in hot and cold seasons. To do so, we use as long and homogeneous series as possible, provided by the European Climate Assessment and Dataset project (ECA&D) for different locations in Europe, as well as the European ENSEMBLES project gridded dataset and the ERA40 reanalysis. The seasonality in a homogeneous season is in general not very important. When the purpose is not to model the residuals, this seasonality can be ignored. Moreover, in our trend identification method (in section 3.2), there is any strong assumption on the noise. The noise can follow any distribution law, it can not a white noise. Only assumption is that its covariance is stationary. This assumption can be check by a Fourier transformation which will be show in appendix. We provide a definition of trends that we keep as intrinsic as possible and apply non parametric statistical methods to analyze them. Obtained results show a clear link between trends in mean and variance of the whole series of hot or cold temperatures : in general, variance increases when the absolute value of temperature increases, i.e. with increasing summer temperature and decreasing winter temperature. This link is reinforced in locations where winter and summer climate has more variability. In very cold or very warm climates, the variability is lower and the link between the trends is weaker. We performed the same analysis on outputs of six climate models proposed by European teams for the 1961-2000 period (1950-2000 for one model), available through the PCMDI portal for the IPCC fourth assessment climate model simulations. The models generally perform poorly and have difficulties in capturing the relation between the two trends, especially in summer.

4.1 Introduction

Global and regional warming is observed and is now well established for the twentieth century, at least for the mean temperature. This temperature increase is bound to continue, and changes in mean temperature as well as in variability and extremes are predicted. These have important implications in terms of adaptation. Katz et al.([85]), have already mentioned that the variability change has more impact on extremes than the change in the mean. Some authors have also looked at the impact of climate change on both mean and variability. Schär et al.([133]), pointed out the role of increasing variability in the occurrence of the 2003 heat wave in Europe. Ferro et al.([50]), proposed a technique to explore changes in probability distributions and applied it to climate model simulations. Here, we would like to focus on observed datasets, in order to carefully study the observed properties of the mean and variance of temperature series, before investigating climate model simulations. As a matter of fact, climate models are designed to represent

and simulate the possible evolution of global and regional mean temperature, but their ability to reproduce observed trends, variability and extremes, at least at local scales where they influence industrial activities, is not yet fully demonstrated. Therefore, a careful study of the link between trends in mean and variance of daily minimum and maximum temperature, is a first step to analyze in detail the relationships between these trends and the trends in extremes. In this context, we focused on daily maximum summer temperatures and daily minimum winter temperatures. To do so, we used 55 as long and homogeneous temperature series as possible, for Europe. Those were obtained from the European Climate Assessment and Dataset project (ECA & D) (Klein et al., [42]). The study has then been conducted using the ENSEMBLES project gridded dataset (Haylock et al.,[76]) and the ERA40 reanalysis. Trend identification is carried out using non parametric methods, in order to avoid any assumption on the shape of trends. Then, the same analysis is performed using the results of six climate models proposed by European teams for the period 1961-2000 (1950-2000 for 1 model), obtained in the framework of the fourth IPCC assessment report and made available through the PCMDI portal.

The paper is organized as follows : section 4.2 presents the temperature series used, then section 4.3 describes the general statistical methods of smoothing which are used for all series in this paper, for the mean as well as for the variance. In section 4.4 we discuss the results obtained from the different observed datasets. Section 4.5 gives the results obtained from the climate simulation models, before concluding in section 4.6.

4.2 Presentation of the used data series

4.2.1 ECA & D series

The initial temperature series used in this study consist of 55 of the longest and most homogeneous daily minimum and maximum temperature series made available by the European Climate Assessment and Dataset project (ECA & D) (Klein et al.,[42]). For each of the collected series, the ECA & D project tested for homogeneity using different test to identify breaks. The results are summarized for each series over different periods in classifying it as “suspect”, “useful” or “undetermined” and in indicating the years when breaks have been identified. In the first place, only series quoted as “useful” over the 1946-2006 and the 1901-2006 periods have been retained. Then, among these series we keep only those for which no break has been identified with the 3 methods used in the ECA & D project. The period lengths are the maximum periods of available (non missing) data since 1946 (for 53 series) or 1901 (for the 2 Déols and Dresden series “useful” over this period).

4.2.2 European ENSEMBLES project gridded dataset

The production of gridded daily datasets from observations is one of the deliverables of the European ENSEMBLES project (www.ensembles-eu.org) : the research team RT5.1 of the project is due to produce observational daily gridded datasets for temperature and precipitations. The ENSEMBLES project is supported by the European Commission's 6th Framework Program as a 5 year Integrated Project from 2004 to 2009 under the Thematic Sub-Priority "Global Change and Ecosystems".Different datasets are now available, for daily minimum, maximum and mean temperature and daily precipitation amount, covering the period from 1950 to 2006 on a 0.25 and 0.5 degree regular grid as well as on a 0.22 and 0.44 degree rotated pole grid. The datasets for daily minimum (Tn) and daily maximum (Tx) temperature on a 0.5 degree regular grid have been used in this study.

4.2.3 ERA40 reanalysis

The European Centre for Medium range Weather Forecasts (ECMWF : www.ecmwf.int) has conducted a reanalysis project, in order to promote the use of global analyses of the state of the atmosphere, land and surface conditions over the period from mid 1957 to 2001. The three dimensional variational technique has been applied using the T159L60 version of the Integrated Forecasting System to produce the analyses every six hours. The 2-meter temperature dataset over a 2.5 degree regular grid has been used in this study. From this dataset, the maximum (minimum) of the 4 daily values has been considered for the daily temperature maximum (minimum). This leads to slightly under-estimated values compared to the "true" daily maximum or minimum, but with a comparable order of magnitude.

From these different datasets, the analysis has been made on a spatial window covering Europe between longitudes 10 and 65 East and between latitudes 35 and 80 North. The hot season has been defined as the 100 days between the 14th of June and the 21st of September, whereas the cold season has been defined as the 90 days between the 1st of December and the 28th of February. These periods have been selected because most extreme events occur between these dates. Daily maximum and minimum temperature have been considered in the hot and cold season respectively.

4.3 Statistical tools

4.3.1 Trend definition

We first give some precision about the definition of a trend.

If $X(t)$ is the observation series whose mean, variance and probability density function are respectively $m(t)$, $s^2(t)$ and $X(t)$ can be expressed as $X(t) = m(t) + s(t)Y(t)$, with $Y(t)$ centred with variance 1. $Y(t)$ can be a

stationary process or not, in particular its distribution $g(t, x)$ can evolve with t or not. For instance, even if the third moment depends on t ; $m(t)$ and $s^2(t)$ are the usual or central trends of $X(t)$. It can be proved that the different methods of estimation discussed below are valid even if the process $Y(t)$ presents some dependence properties and some non stationarity properties. If g is still time dependent, it just means that the detrended process has a remaining deformation trend. In this study, our aim is only to estimate the central trends $m(t)$ and $s^2(t)$.

4.3.2 Trend estimation

4.3.2.1 Estimation methods

Time series associated with climate variables such as temperature are non-stationary and non-Gaussian. In order to estimate the central trends $m(t)$ and $s^2(t)$ as intrinsically as possible, in the first place we will avoid using either moving average with a window length h or linear regression. The window length h fixes the time scale taken into account, that is the period over which the average is computed, and linearity is a strong assumption concerning trends.

Parametric modeling such as linear regression is widely used to analyze trends in mean or in variance. However, it can constrain excessively the range of possible fits for exploratory modeling, it loses a large amount of information and might not be adapted to complex issues arising from climate studies, except for some particular cases.

In an other work (Yiou et al.,[110]), we have seen that seasonal means and variances are correlated for some areas in Europe in winter and summer. Here, we would like to go further in looking for the time evolutions of the means and variances, and their possible connections, in the most general way as possible, thus in using non-parametric methods. In the present work, time scale has to be thought of as linked to the speed of evolution of annual data. The evolution during a fixed year is mainly due to seasonality and reveals only a small component of the large temporal scale evolution. Thus, the choice of the set of functions able to describe trends is a quite difficult problem. The idea is to fix a degree of smoothness such that the trend does not show too many extrema and inflexions for a twice differentiable function. In fact the choice has to be adapted to the analyzed data to keep the trend identification as intrinsic as possible. A discussion of this topic can be found in Wu et al.([151]) concerned with the application of specific methods based on physical characteristics, but this is out of the scope of this paper and can not take into account a variation of the variance. Classical non parametric methods such as cubic splines, penalized least squares, loess, wavelets etc. have adaptive variants in the sense of Bickel ([11]).

Each method relies on some “technical” parameters and the adaptive

character consists in an optimal (in a statistical sense) estimation of these parameters from the data. Each of these methods allows a visual description of the trend in the resulting plot of the response versus time, making them very flexible. Here we use two popular methods : cubic splines smoothing(Green and Silverman, [59]) and local polynomial estimators (loess) (Fan and Gijbel, [47]). Cubic splines smoothing consists in looking for a twice differentiable function which represents optimally the evolution of the mean of the observed data. The smoothing parameter happens to be a penalization term which does not allow too many inflexions.

Loess consists in looking for the polynomial of degree d that corresponds best to the representation of the regression between the variable and the covariate (here, the covariate is time) on an optimal window length. The optimal window length is defined by the parameter λ, which corresponds, if smaller than 1, to the proportion of the total number of data used for the local regression. Then, local regression is applied with a tri-cubic weighting proportional to

$$(1 - (dist/h)^3)^3$$

where $dist$ is the distance between two values(1 day in our application) and h the window length. This allows adding more weight to neighbour points and less to more separated ones.

Thus, cubic splines give a trend function optimized on the whole series length giving the best curvature, whereas loess follows more locally the regression between the two variables.

We used both cubic splines and loess to model functions of mean and variance of the whole observation set ($m(t)$ and $s^2(t)$), and both methods give very similar results. Therefore, in the following, we will only present the results for the loess, which is the most convenient. The proper choice of the smoothing parameters for splines or loess is crucial. Adaptive procedures such as cross-validation procedures help in making the right choice, but their efficiency is hindered by highly auto-correlated series such as temperature series. Cross-validation has to be modified in order to eliminate the dependence. Here, the dependence between the dates can be considered as zero if the dates are distant by more than 5 days. We thus used a cross validation method on data sampled every 10 days, and an optimal parameter corresponding to a window length of around 11 to 15 years has been found. Another approach is to choose a smoothing parameter using an empirical and efficient way based on learning. We used this alternative as a visual control of our previous choice.

A common problem with smoothing methods is related to boundary problems at the beginning and end of the observation period. The data are all on one side of the fitting point in the boundary case. Thus, using the same algorithm at the boundaries and in the interior leads to higher variability. Loess and splines can automatically provide boundary bias correction. Fit-

ting local polynomials at values in the boundary region, the bias is small, and can therefore be ignored.

The calculations are conducted as follows :

• selection of the days of the studied season over the total period length (this leads to a series of nyear100 days for summer and nyear90 days for winter, nyear being the number of years in the total period length)

• computation of $m(t)$ from $X(t)$ using loess

• computation of $[X(t) - m(t)]^2$

• computation of $s(t)$ from $[X(t) - m(t)]^2$ using loess

where the hat notation corresponds to estimates.

Thus, the time evolution of the variance takes into account the time evolution of the mean. The estimation is not done using the likelihood(unknown), but using least squares. This methodology has a theoretical support (Ruppert et al.,[131]).

In order to illustrate the role of the smoothing parameter, the of figure 4.1a presents the trends $m(t)$ identified in summer temperatures for the long summer temperature series of Déols between 1901 and 2006 using a linear local regression with a smoothing parameter corresponding to a 15-year window (15 summers here, i.e. 1500 days) and using a moving average on the same window. The two trends are superimposed over the evolution of the summer means computed from the data series (for each day in a summer, the temperature is set to the mean summer temperature). It can be seen that loess is able to reproduce better small data variations, where moving averages tend to lead to a higher smoothing (top panel). To capture the same details as obtained with loess, the moving average window must be twice as small, that is around 7 years, as shown in the bottom panel.

On the other hand, the choice of linear local regression has been compared to the one of degree 2 local regression (figure 4.1b). It can be seen that the details are enhanced in using quadratic local regression, but it has been judged sufficient for the study to use linear local regression.

4.3.2.2 Bootstrap confidence intervals

It is now necessary to quantify the uncertainties associated with the estimation of the mean and variance evolutions. To do so, the construction of a confidence interval is based on the residual bootstrap resampling, which is simple and chosen because it does not rely on the evaluation of quantities according to some asymptotic property.

There exists a widely developed theory for asymptotic confidence intervals (Hall [65], [66] ; Faraway [49] ; Haïrdle and Mammen [71] ; Neumann [104], [103]).

We choose to construct a pointwise confidence interval based on bootstrap percentiles. It is well adapted to our estimation method because it does not artificially assume normality and it is simple to implement for our large

datasets.

Here because of the dependence of the data, we use moving blocks bootstrap (Lahiri [91]) with a size of 10 days, and we simulate 2000 samples. The results obtained for the evolution of the mean of summer daily maximum temperature for the long series of Déols (19012006) are plotted in Figure 4.2.

4.3.3 Possible link between the estimated trends $m(t)$ and $s^2(t)$

Using the previously presented non parametric methods, the function $m(t)$ and $s^2(t)$ are estimated by $\hat{m}(t)$ and $\hat{s}^2(t)$. Suppose that these estimations are two functions with a strong link, measured for instance by a scalar product close to one (which indicates more a linear link). Could this link be produced by some statistical artefacts ?

Firstly, for a large dataset, all the procedures used here give consistent estimators, roughly speaking the speed of estimation is of the order of N^{-a} (N being the number of values in the dataset), $a > 0$ and depending on the chosen degree of smoothness for the trend. Besides, we suppose that the trend functions are three times continuously differentiable.

The estimation errors $m(t) - \hat{m}(t)$ and $s(t) - \hat{s}^2(t)$ are asymptotically correlated but their correlation is too weak in N^{-a} to be the source of the identified link. In fact, for a large dataset (see the appendix and Ruppert et al., [131]), we can first estimate $m(t)$ and then $s^2(t)$,as we do, and this is as efficient as a direct estimation of the pair $(m(t), s^2(t))$.

Now can the link be induced by the density probability functions $f(x, t)$ of the data itself ? As a matter of fact, all the information on the data X(t) is given by its probability density function. However, a link between $m(t)$ and $s^2(t)$ does not necessarily imply some defined form of the probability density function.

All the calculations are made using the free R statistical software.

4.4 Results for the different observational datasets

4.4.1 ECA & D temperature series

The first task has been to choose the smoothing parameter. As we want to study a large number of series, it is either necessary to automate the adaptive choice of the smoothing parameter or to fix it in advance. As previously mentioned, a parameter corresponding to a window length of around 11-15 years has been found in most cases. Figure 4.3 shows the mean evolution for long daily maximum temperature series of Déols (1901-2006) using different window lengths : 10, 15 and 20 years (or summers as we only keep the 100 summer days each year). As expected, the larger the window is, the smoother

the curve appears. In the following, the results for a 15-year window length will be shown in order to simplify the exposition.

The loess technique with a 15-year window length is applied to compute the evolution of mean and variance of daily minimum temperature in winter and daily maximum temperature in summer. The results for the series of Smolensk in winter and La Rochelle in summer are presented in Figure 4.4, together with the 90% confidence intervals . These results suggest a strong link between the evolution of mean and variance, with the variance increasing at the same time as the mean in summer and decreasing as the mean increases in winter. In particular, there is a quite perfect correspondence between the local extremes. To examine the dependence between those two series, we can compute correlation coefficients which correspond to the scalar product (divided by the Euclidean distances). Another way to look at the resemblance between two curves is to compare their parallel slopes, thus we can compute the correlation coefficients between their first derivatives too.

We compute correlation coefficients between the two curves or their first derivatives for all the 55 time series. Correlation coefficients are thought as we said previously as a measure of the linear relation between two function $m(t)$ and $s^2(t)$, say of the possibility to get a good approximation of $s^2(t)$ by an affine function $a + bm(t)$. This similarity could be evaluated through other measures, such as for example the link between their first derivatives, as stated before. Thus here, the link between the two curves, is measured by the correlation coefficient and has been considered as significant if the correlation coefficient exceeds 0.2 in summer and is lower than -0.2 in winter. According to this criterion, all correlations but one (for the Petsjora series) are significant in winter and all but 8 (Alger, Elatma, Kem, Smolensk, Sortavala and Vilnius series) in summer. The top panel of Figure 4.5 shows the maps of the correlations coefficients between the evolution of mean and variance of daily minimum winter temperatures on the right and the same significant correlation coefficients between their first derivatives on the left. The bottom panel shows the same results in the same way for daily maximum summer temperatures. Correlation coefficients between $m(t)$ and $s^2(t)$ or between their first derivatives give very similar results : they are rather strong in winter, except along the coasts and in the most northern regions. In summer, correlations are strongest for locations experiencing more temperate climate than for those with more continental climate. In the following, we will only show the results for the correlation coefficients between the evolutions of the mean and the variance ($m(t)$ and $s^2(t)$).

4.4.1.1 ENSEMBLES gridded dataset

In order to extend and verify the previous statement, the same analysis of correlation coefficients between the evolutions of mean and variance

for daily minimum temperature in winter and daily maximum temperature in summer has been performed using the ENSEMBLES gridded dataset. As previously mentioned, this dataset is on a 0.5 degree regular grid and has been constructed from the best available observation datasets, including those used in the ECA & D project. Many grid point series show an important number of missing values (all values missing in some cases) thus the analysis has been made only for those series presenting less than 10% of missing values. The remaining missing values have been replaced by the climatological mean of the corresponding day. Results for correlation coefficients between the evolution of mean and variance of daily minimum winter temperatures and daily maximum summer temperatures are presented in figure 4.6. The results are consistent with the previous ones, showing strong negative correlations in winter for the central part of Europe except some of the coastal areas. In summer, the picture is more contrasted : high positive correlations are found for France, the south of Great Britain and some areas in south east of Europe, but weaker, although still positive, correlation elsewhere.

4.4.2 The ERA40 reanalysis

The last test has been done using the results of the ECMWF ERA40 reanalysis project, in order to verify the results on a dataset whose homogeneity is ensured by construction. Results are presented figure 4.7. They are very similar to the previous ones and confirm the relationships : the variance decreases when the mean increases in winter for most of the central part of Europe and the variance increases when the mean increases in summer in France and Great Britain and in some parts of central Europe near the Black Sea.

4.4.3 Results summary

4.4.3.1 Cold season

All datasets lead at similar conclusions with strong negative links between the evolution of mean and variance of minimum daily temperature in winter over a large part of central and northern Europe. However some regions around the Mediterranean and the Black Sea show positive correlation coefficients, in the ENSEMBLES gridded dataset (top of figure 4.6) as well as in the ERA 40 reanalysis (top of figure 4.7). Figure 4.8 shows evolutions of mean and variance for a grid point in the ENSEMBLES gridded dataset near Smolensk, where the correlation coefficient is strongly negative, and a grid point in Spain, where the correlation coefficient is positive, in winter. The plot for the grid point near Smolensk strongly resembles the one for the ECA & D Smolensk series (Figure 4.4), which is not surprising, as the same series have been used to produce the ENSEMBLES gridded dataset.

For the point in Spain with positive correlation coefficient, it can be noted that the variance evolution is rather weak and the mean is rather warm, which could explain our results. The same behaviour is found for all rather strongly positive correlation coefficients in winter.

4.4.3.2 Hot season

In the summer season the picture is more contrasted, with rather high correlation coefficients in France and Great Britain and in some parts around the Black Sea. Correlation coefficients remain positive over a large part of Northern and central Europe, although lower than 0.5. Here, the points showing strong negative correlation coefficients correspond to two different types of behaviour : most of them correspond to very hot summers, and others seem to correspond to points where the effect of altitude is strong, with rather mild summers (points in Switzerland and northern Italy). Figure 4.9 shows the evolution of mean and variance of maximum daily temperature in summer for two points of the ENSEMBLES gridded dataset : a point near La Rochelle, with a strong positive correlation coefficient, and a point in Spain with a strong negative coefficient. The evolution of the mean and variance for the grid point near La Rochelle is very similar to those of the series of La Rochelle in Figure 4.4 for the same reason as explained in the previous section. For the point in Spain, the temperature is warm and seems to be less variable when it is warmer.

The method used here to derive trends describes the evolution over a continuous set of time scales and it is difficult to infer which scale is the most responsible for the observed link between mean and variance (day to day, intra seasonal scale or inter seasonal scale). This could be further investigated using for example wavelet analysis. However, from an analogous study made on seasonal mean and variance (Yiou et al., [110]), it can be seen that the intra seasonal variability plays an important role. This means that in some regions, a very cold winter or a very warm summer is more variable. This could be due to non-permanent hot or cold episodes, interrupted by milder ones, and this could be the explanation for a strong link in France and Great Britain during the summer period for example. This hypothesis should be further investigated by studying episodes for some particular series. For example in Spain, where summers are hot, the variance increases when the mean is lower, and one could wonder if there exists a summer mean threshold above which the behaviour reverses.

4.5 Results for 6 European climate models

The study on different observation datasets shows some robust relationships between the evolution of mean and variance of daily minimum temperature in winter and daily maximum temperature in summer. The aim

of this section is then to investigate if climate models reproduce a similar link. To do so, we used results of the simulation for the period 1961-2000 conducted with five different climate models and for the period 1950-1999 for one other climate model, elaborated by European research teams. The models are :

- bccr-bcm2-0 of the university of Bergen in Norway
- cnrm-cm3 of the French meteorological office research centre (Centre de Recherches Mtorologiques de Mto-France)
- ipsl-cm4 of the Institut Pierre-Simon Laplace in Paris, France
- mpi-echam5 of the Max Planck Institute for meteorology in Hamburg, Germany
- ingv-echam4 of the National Institute of Geophysics and Volcanology in Bologna, Italy
- ukmo-hadgem1 of the United Kingdom Meteorological Office Hadley Centre, over the period 1950-1999

Simulations of these models obtained in the framework of the fourth IPCC assessment are available from the PCMDI web portal.

We applied the same methods as described above and obtained results summarized in Figure 4.10 for winter and Figure 4.11 for summer. A comparison of these figures with those obtained from the observational datasets (figures 4.6 and 4.7) shows that the models generally fail to reproduce the observed link between the evolutions of the mean and the variance of temperature. Except for cnrm-cm3, the models seem to perform better in winter than in summer, especially the ingv-echam4 model and the bccr-cm2-0 model. Curiously, these models share their atmospheric component (although in a different version) with other ones (mpi-echam5 for ingv-echam4 and cnrm-cm3 for bccr-bcm2-0) whose results are less good. In summer, all models fail to capture the observed link correctly. This puts in doubt the representation of temperature variability by current climate models.

4.6 Conclusion and perspectives

In this paper we tried to use as far as possible the properties of non parametric methods in statistics to obtain general qualitative properties on the time evolution of temperatures in Europe over periods of at least 50 years.

The first conclusion is that the mean and variance for hot and cold seasons have a very similar evolution : the variance increases as the mean increases in absolute value, i.e. it increases when daily maximum summer temperatures increase and when daily minimum winter temperatures decrease. This link seems to be more general in winter for a large part of central Europe, even though it is weaker in coastal areas, whereas in summer it is limited to areas where summers can experience heat waves without

being in average too mild or too warm. This result is found whatever the observational dataset used, the ECA & D data series, the ENSEMBLES daily gridded dataset or the ECMWF ERA40 reanalysis, each available on different period lengths over the twentieth and the beginning of the twenty first century. Accordingly, the result seems to be robust. This could be explained by the fact that in these areas, cold winters or warm summers are associated with some very cold or very warm episodes, interrupted by more "normal" conditions, in relationship with the large scale atmospheric circulation. This hypothesis will be investigated by advanced studies on episodes during summer/winter seasons.

On the other hand, climate models generally fail to correctly reproduce this link, although their behaviour seems to be better in winter than in summer. The ingv-echam4 model, and to a lesser extend the bccr-bcm2-0 model show rather good results in winter, whereas all the studied models fail to correctly reproduce the observed link in summer. Thus, the models cannot be trusted for the reproduction of temperature variability in their current versions.

A further step will then be to carefully study the link between evolutions of mean and variance and the evolution of extremes, using statistical extreme value theory and similar techniques to derive non parametric evolutions. A first analysis of this link is presented in Nogaj et al. ([105], but a more systematic study is needed to obtain more robust results.

Acknowledgments We acknowledge the climate dataset from the EU-FP6 project ENSEMBLES (http ://www.ensembles-eu.org) and the data providers in the ECA & D project (http ://eca.knmi.nl). We acknowledge the modeling groups for making their simulations available for analysis, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) for collecting and archiving the CMIP3 model output, and the WCRP's Working Group on Coupled Modelling (WGCM) for organizing the model data analysis activity. The WCRP CMIP3 multi-model dataset is supported by the Office of Science, U.S. Department of Energy.

Appendix

We have defined $Y(t) = [X(t) - m(t)]/s(t)$. Deformation trends are trends able to partially describe the remaining non-stationarity in the distribution of $Y(t)$. A first result concerns the non-stationarity of the extremes of $Y(t)$ and is exposed in Nogaj et al. ([105]). Other remaining trends can be found using the probability density function of the residuals $Y(t)$. This is illustrated in Fig. 4.12 with two estimates of this probability density function for the station of Déols in summer. The first one is constructed from the data during the first 15 years of observation and the second one during the last 15 years. We also can read in Table 4.1 the evolution of the skewness (here simply the third moment because Y is centred, with variance 1 and kurtosis, here the fourth moment, minus 3). These quantities do not seem

to evolve with time, the relative variation being always less than 1/10. We mentioned in Sect. 3.3 that our method can be applied for non-stationary Y if the fourth moment is bounded. It can be seen that it is the case here, it is bounded by 4 and moreover the estimation of the parameters of extreme distribution shows that this distribution is bounded for Y (this can also be seen on the previous estimation of the probability density function in Fig.4.12). The correlation between the Y(t) is zero for time distances larger than 4, so loess can safely be applied.

	First 15 summer	Last 15 summer
Mean	-0.0066	-0.016
Median	-0.13	-0.08
Variance	0.9932	0.9987
Skewness	0.3940	0.2814
Kurtosis	-0.2587	-0.2745

TABLE 4.1 – Moments of the probability density function of $Y(t)$ estimated from the first 15 summers and the last 15 summers for the long daily maximum temperature series of Déols, 19012006

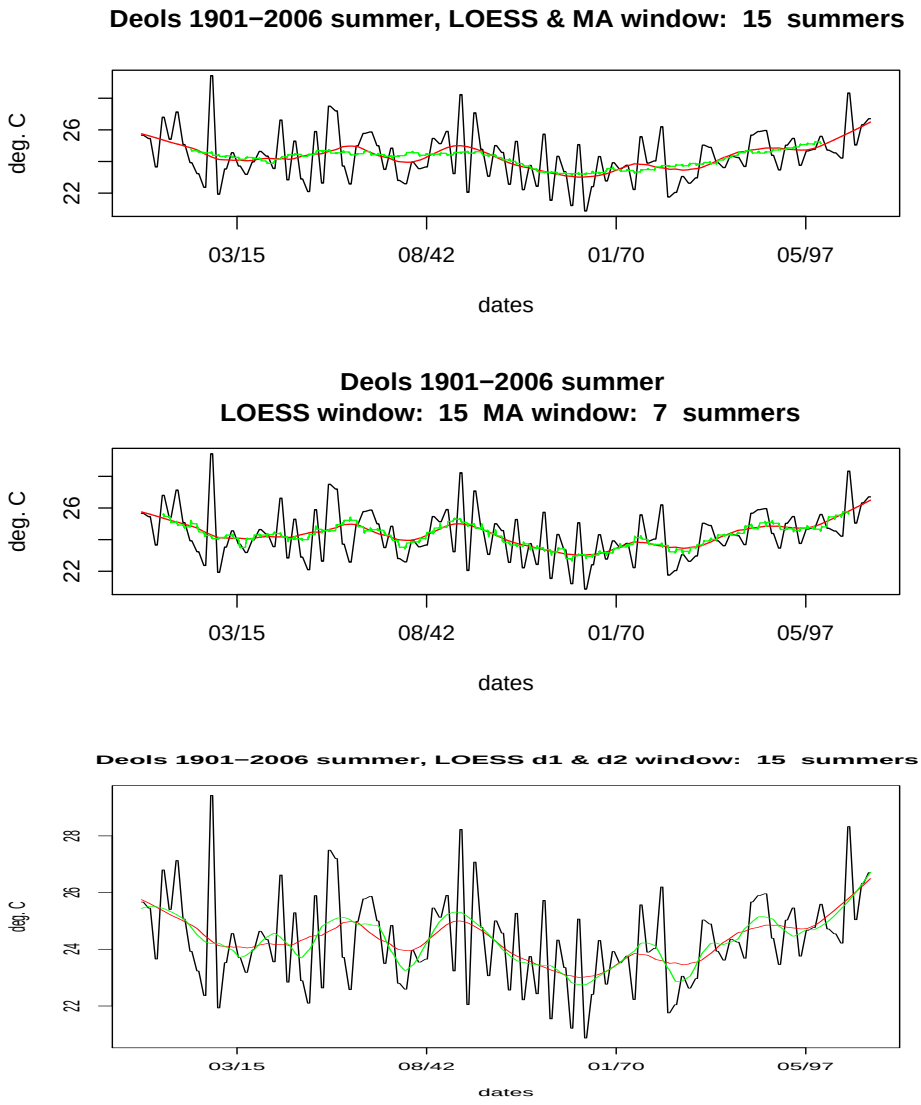


FIGURE 4.1 – **Upper panel**. Mean summer daily maximum temperature in Déols over the 19012006 period (black line) and the corresponding trends computed with LOESS with a window length of 15 years (red line) compared to the same trend computed using a 15-year moving average window (*green line*) (*top panel*) and to the same trend computed using a 7-year moving average window (*green line*)(*bottom panel*). **Lower panel.** Mean summer daily maximum temperature in Déols over the 19012006 period (*black line*) and the corresponding trends computed with LOESS with a window length of 15 years with linear local regression (red line) and quadratic local regression (*green line*)

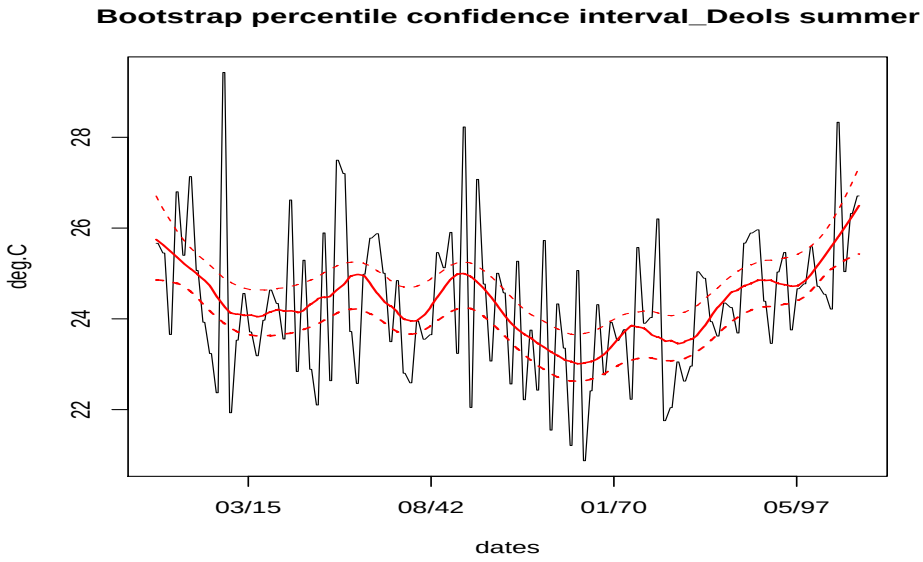


FIGURE 4.2 – Mean summer daily maximum temperature in Deols over the 19012006 period (*black line*) and the corresponding trends computed with LOESS with a window length of 15 years and its 90% confidence intervals (*red lines*)

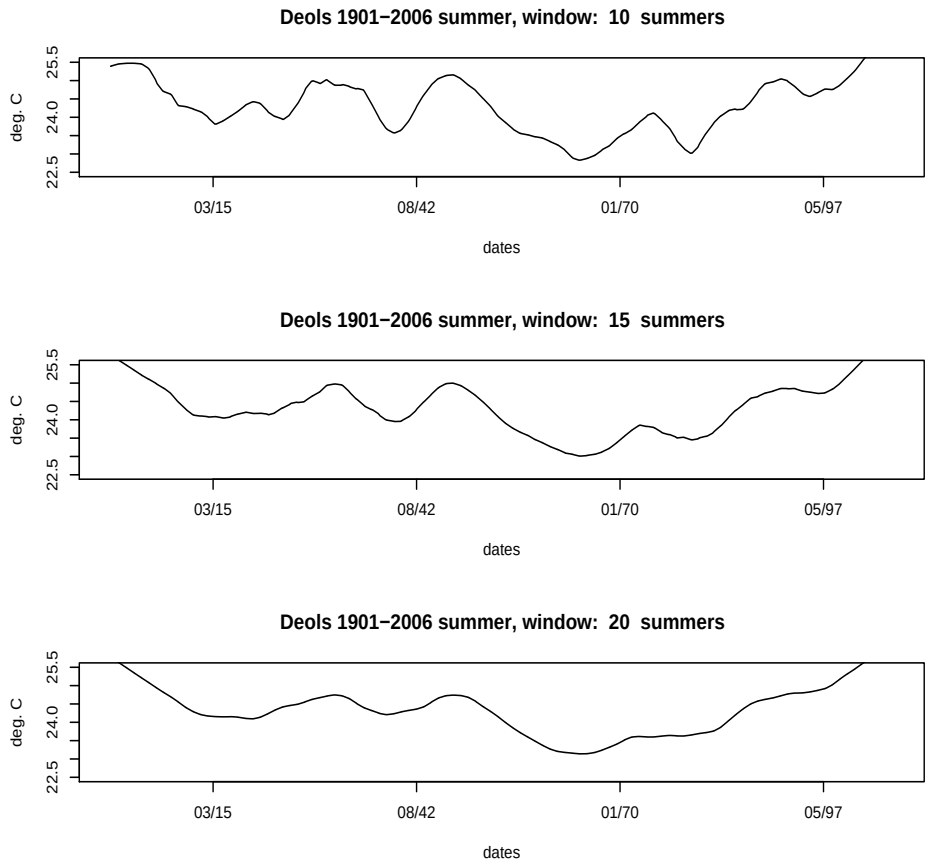


FIGURE 4.3 – Time evolution of daily maximum temperature mean in summer for the long 19012006 Déols series using three different window lengths for the smoothing parameter : 10, 15 and 20 years (or summers)

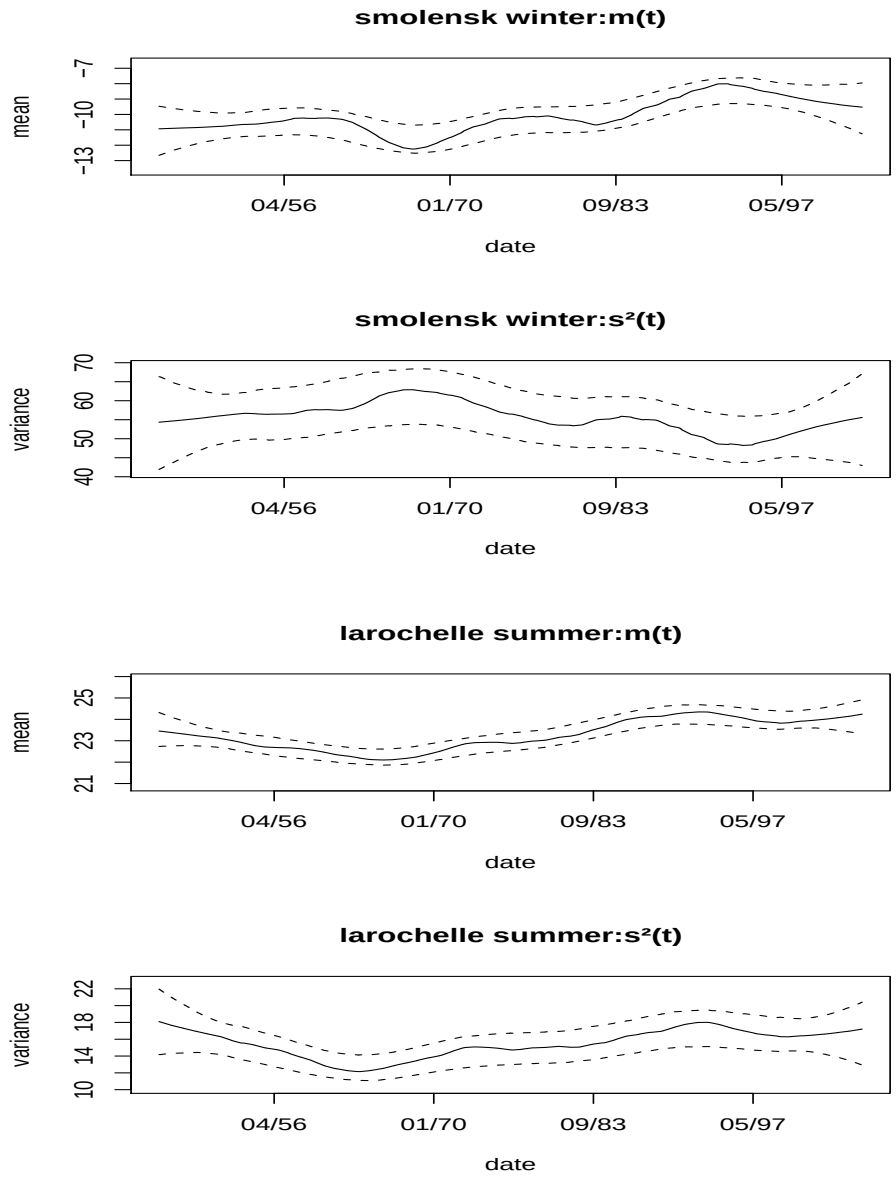


FIGURE 4.4 – Time evolution of mean ($m(t)$) and variance ($s^2(t)$) of minimum daily temperature in winter in Smolensk (*top*) and of daily maximum temperature in summer in La Rochelle (*bottom*) as computed using the loess technique with a smoothing parameter corresponding to a 15-year window and their 90% confidence intervals (*dashed lines*)

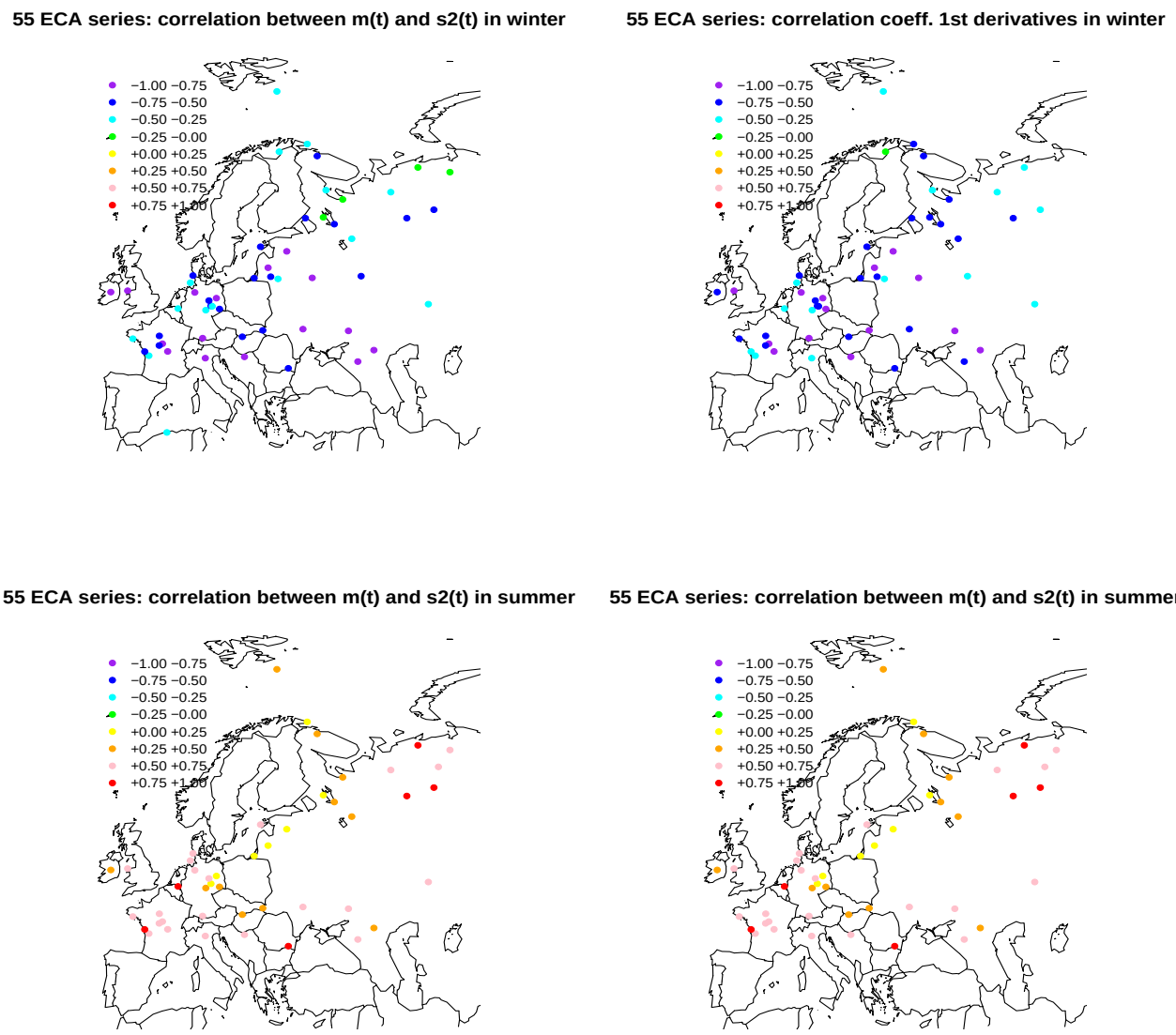
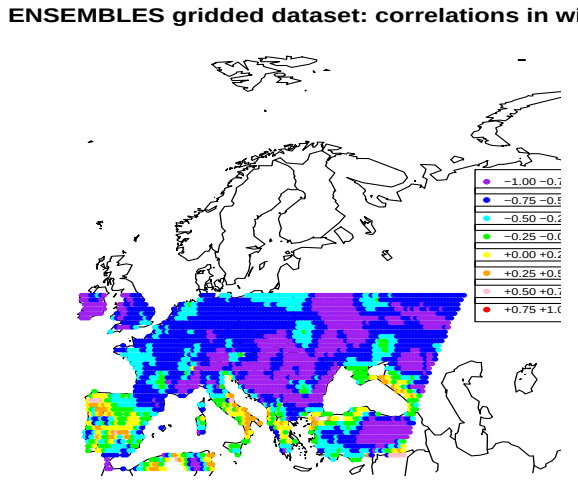
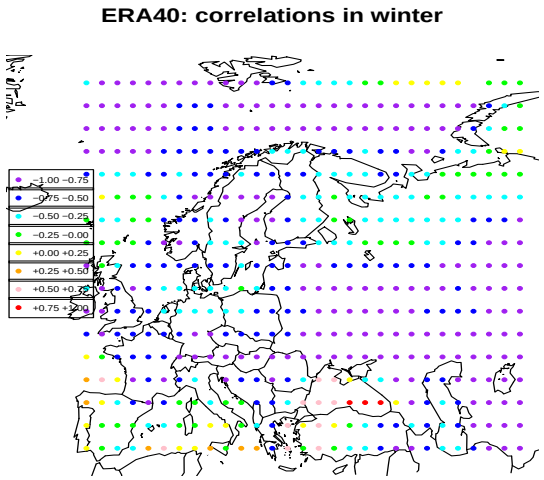
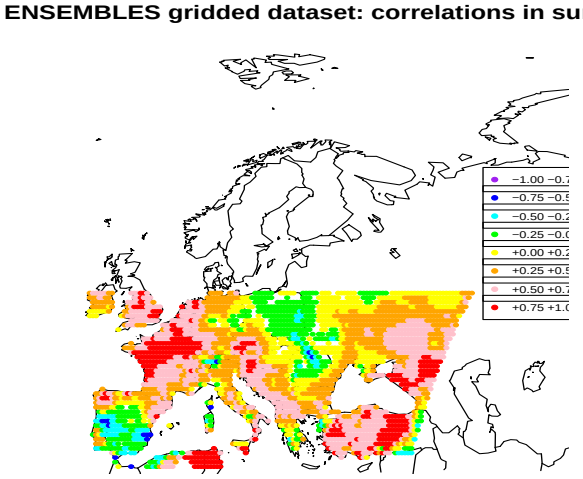


FIGURE 4.5 – Correlation coefficients between the evolution of the mean and the variance (left) and between the first derivatives of mean and variance evolutions (right) of daily minimum temperature in winter (top) and daily maximum temperature in summer (bottom) for the 55 ECA & D series







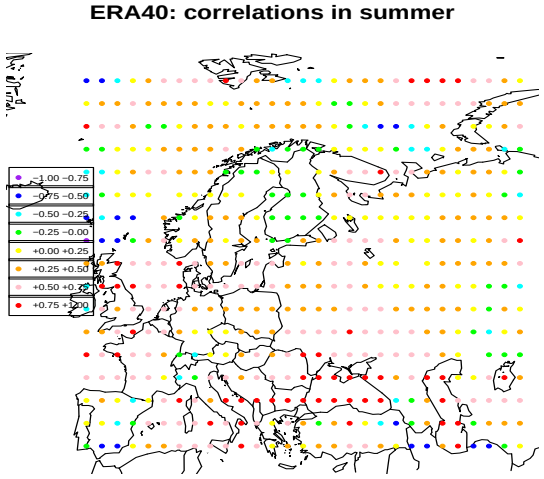


FIGURE 4.6 – Correlation coefficients between the evolution of mean and variance of daily minimum temperature in winter (*top*) and daily maximum temperature in summer (*bottom*) for the ENSEMBLES gridded dataset

FIGURE 4.7 – Correlation coefficients between the evolution of mean and variance of daily minimum temperature in winter (*top*) and daily maximum temperature in summer (*bottom*) for the ERA40 reanalysis dataset

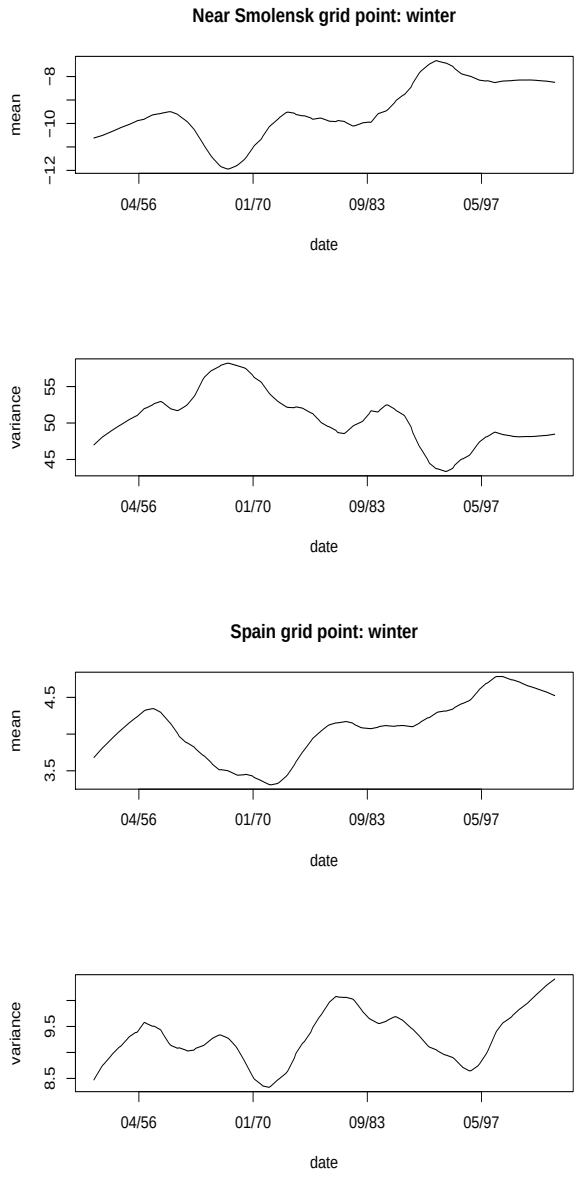


FIGURE 4.8 – Time evolution of mean and variance of the daily minimum temperature in winter for a grid point of the ENSEMBLES gridded dataset near Smolensk (*top*) and in Spain (*bottom*)

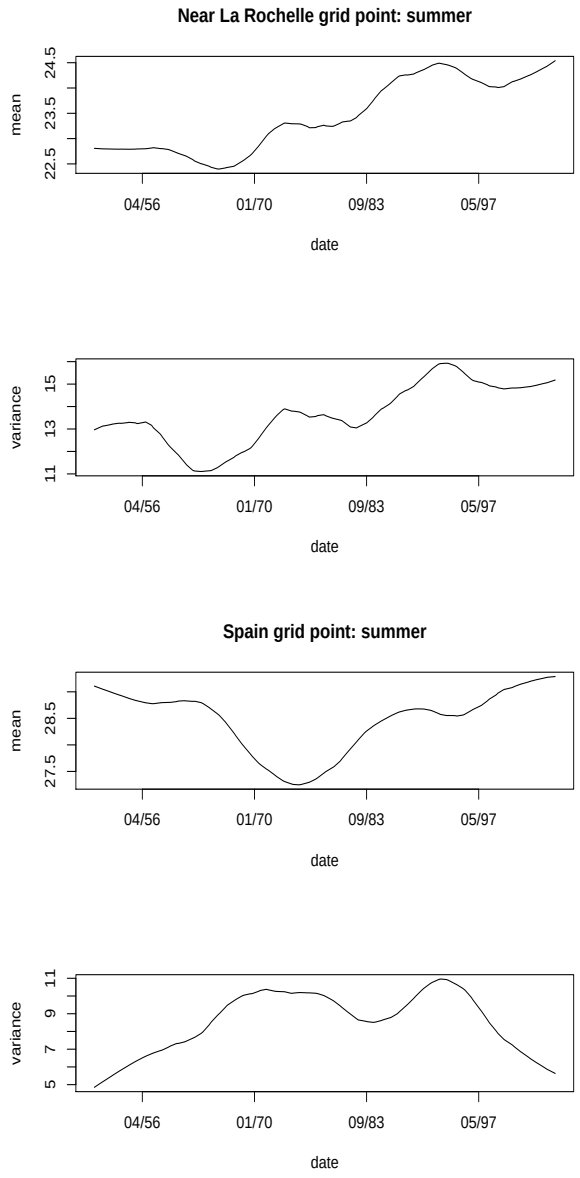


FIGURE 4.9 – Time evolution of mean and varian of daily maximum temperature in summer for a gr point of the ENSEMBLES gridded dataset near L Rochelle (*top*) and in Spain (*bottom*)

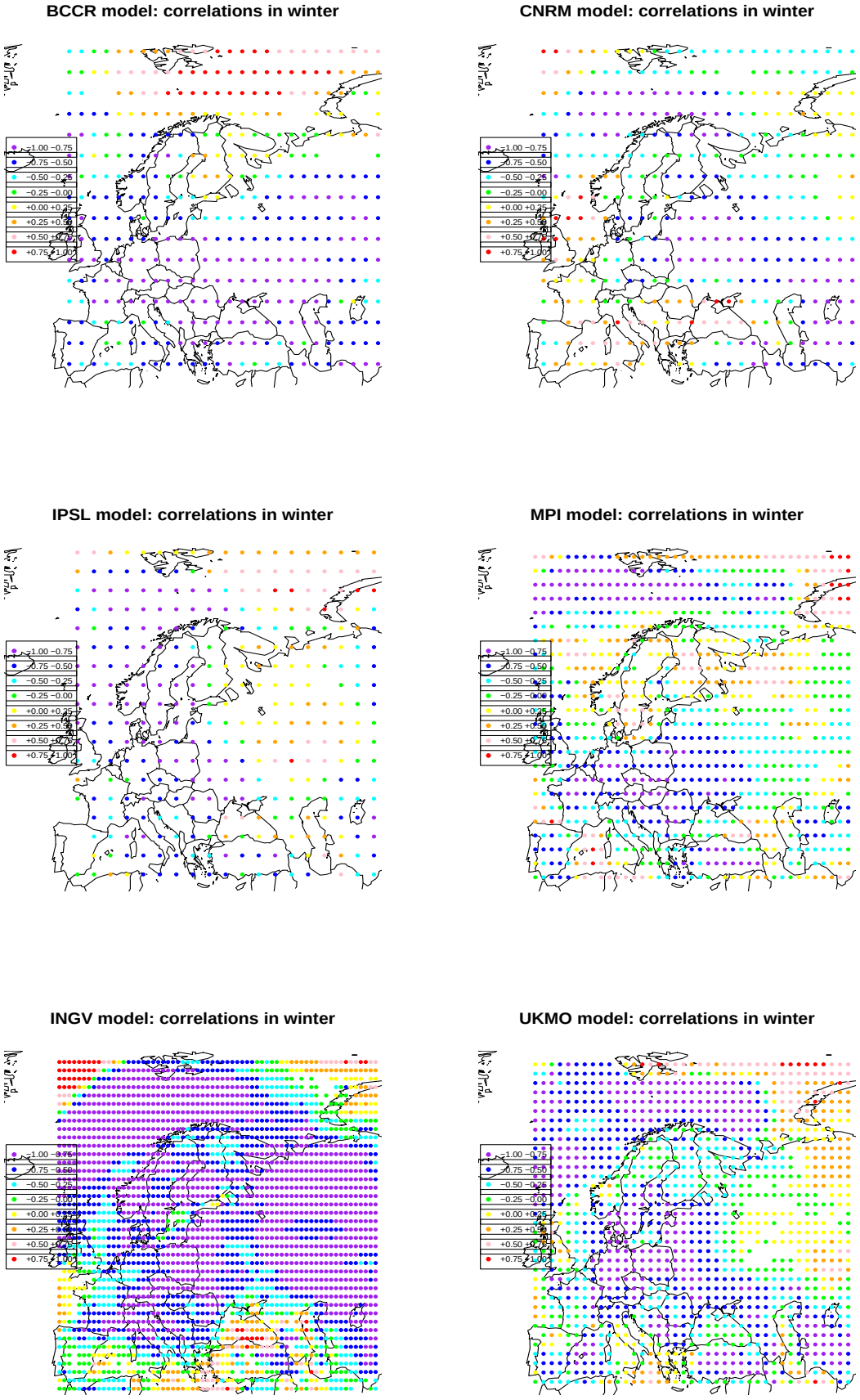


FIGURE 4.10 – Correlation coefficients between the evolution of mean and variance of the daily minimum temperature in winter for each of the six models

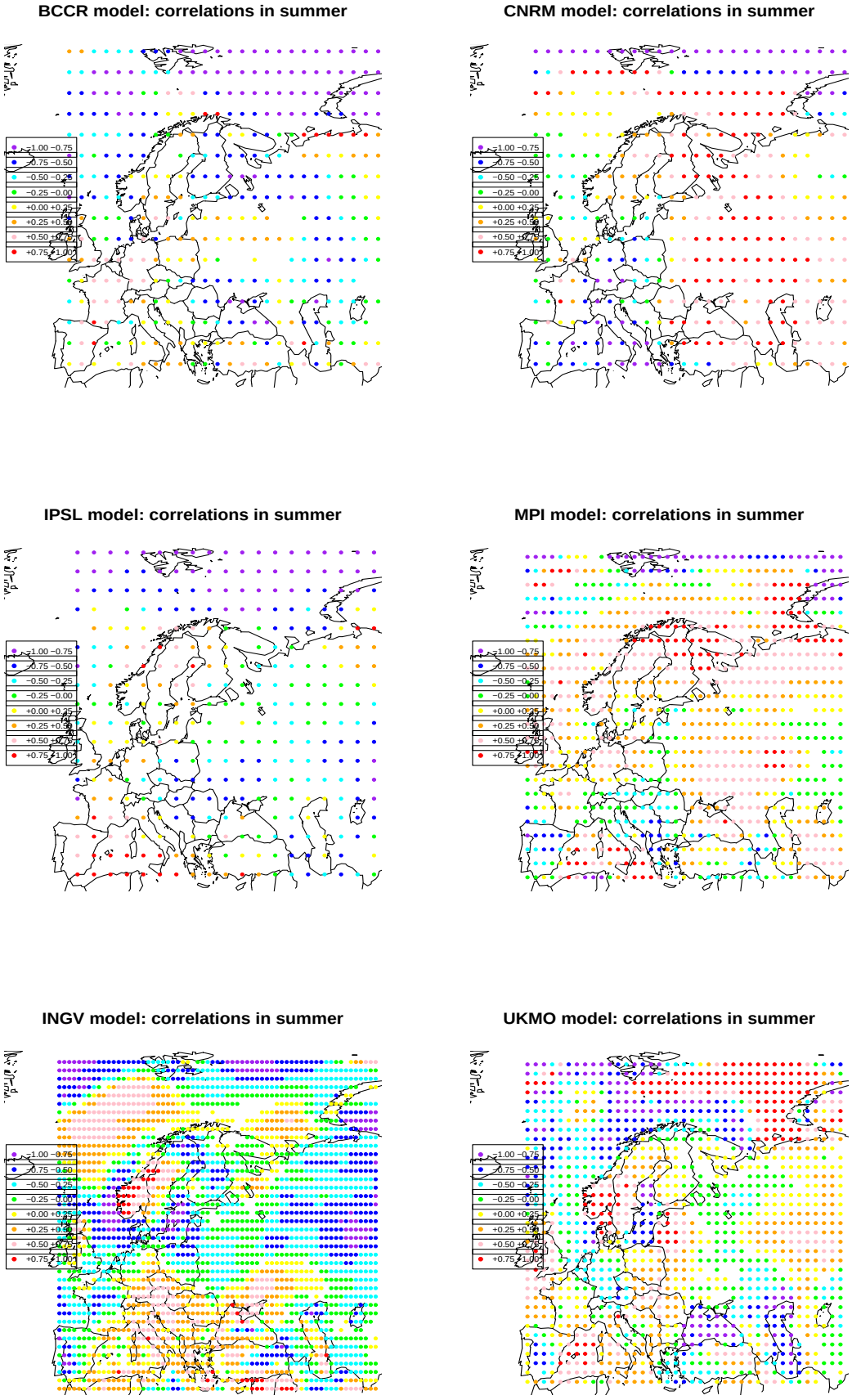


FIGURE 4.11 – Correlation coefficients between the evolution of mean and variance of the daily maximum temperature in summer for each of the six models

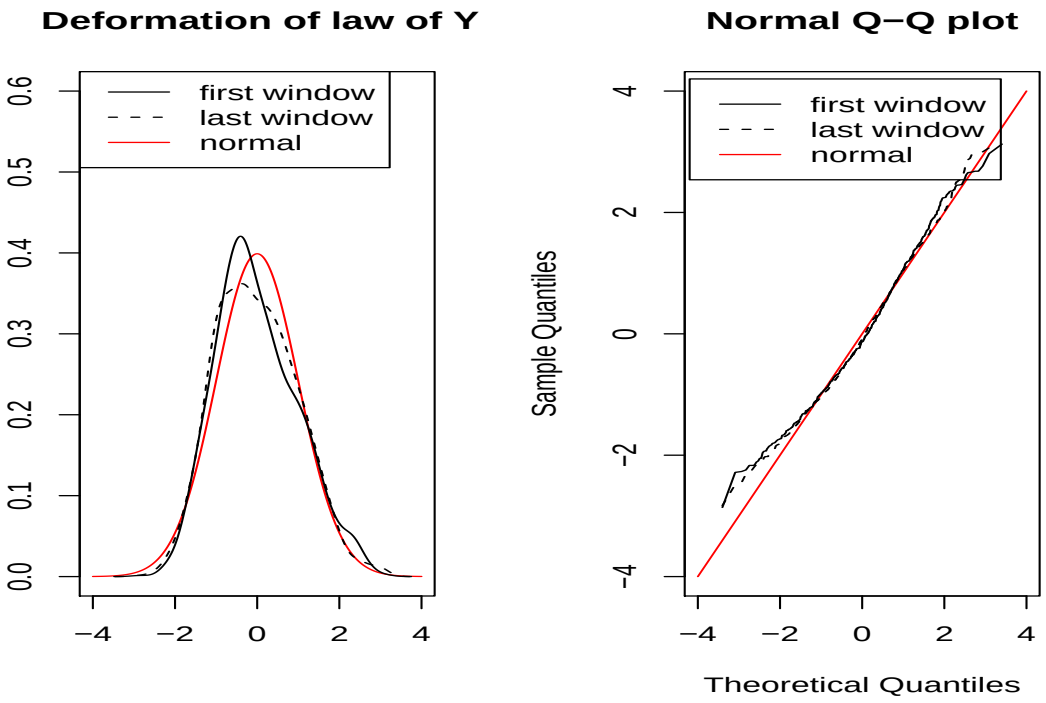


FIGURE 4.12 – *Left panel* probability density functions for $Y(t)$ estimated from the first 15 (*black line*) and the last 15 summers (*dashed line*) of the long daily maximum temperature series of Déols, 19012006 and the normal probability density function with mean 0 and variance 1 $N(0,1)$(*red line*) ; right panel associated quantilequantile plot

Chapitre 5

Non stationary extreme models

5.1 Introduction

This chapter addresses extreme value theory. This theory is concerned with probabilistic and statistical questions related to very high, or very low values. The statistical analysis of extreme events, linked with statistics of the risks, is important in many fields, including insurance, finance, environment and, of course, climate studies. The statistical Extreme Value Theory (EVT) is commonly used by engineers to predict the intensity of meteorological extreme events to which buildings or industrial installations have to resist. The EVT relies on two general definitions of extreme events. The first one considers extreme events as maxima of observations on given blocks of time and then describes them with the Generalized Extreme Value (GEV) distribution. The second considers those extreme events as exceedances over a defined high threshold. The model of these exceedances is called Peak over threshold (POT). Their distribution is a Generalized Pareto Distribution, whereas their dates of occurrences are given by the jumps of a Poisson process.

In this chapter, we will first shortly review the basic theory, including the properties of the two classical models in the dependent and non-stationary cases. The basic theory assumes that the series studied are stationary, that they do not present any cycle nor trend. However, when dealing with climatic data, this assumption has to be carefully considered. Moreover, our subject is to study how climate changes might affect the occurrence of high and low temperatures. This leads us to estimate the trends in samples of extreme values.

As a matter of fact, climatic data also show seasonal cycles. Seasonality and extremes are difficult problems in modeling the temperature.

The traditional approaches, to model trends for extremes, are parametric. The main reason for this is that extremes are most often used to predict return levels and so need to be extrapolated. The second reason is that the seasonality and trends are incorporated, in a linear form, to basic theory. However, due to the need to make assumptions on the parameters, more specific for trends, these approaches reach limits rather quickly. Although trends, or other features of a process, may vary smoothly, there are many applications in which such variations do not have the form of low-order polynomials or, more generally, can not be described by a parametric model.

In problems of this type, it is commonplace to use nonparametric or semi parametric techniques to smooth data. Nonparametric, or semi parametric, modeling is more flexible. It does not rely on any assumption of specific form

for the trends. Furthermore, nonparametric fitting provides a natural way to choose adequate parametric models.

In our context, we use both approaches. An important part of this chapter deals with the nonparametric estimation for functional extreme parameters. A new automatic selection of smoothing parameter, especially for extreme models, is proposed. The demonstration for its asymptotic convergence seems complicated, simulations will be considered to test the performance of the method.

For the parametric approach, the parametric form is assumed here as polynomials or continuous piecewise linear functions(CPL). Although CPL is not the best if one wants to obtain an optimal description of the evolution of extreme events, it does give way to an interesting application in calculating the return levels by extrapolating recent reasonable linear trend.

For the nonparametric aspect, there are two main approaches : local likelihood (LL) and penalized likelihood estimations (PL). PL method is retained for its advantages, which will be detailed later. Penalized likelihood estimation is especially attractive for us to estimate trends in extremes for datasets with numerous series over a large geographical area.

In our context, we model extreme temperatures both by GEV and POT models. Unfortunately, no work has been done on the choice of smoothing parameters in penalized likelihood approaches for GEV models. *For instance, there is the reparametrization method of Chavez-Demoulin concerned POT models with two independent Poisson and Pareto sub-models.* However it seems difficult for GEV models because of the very complicated computations on partial differential equations that it demands for three parameters. The new procedure that will be proposed in this chapter has already been introduced in Chapter 3. Here, we will concentrate on its application to extreme values.

5.2 Results from probability theory

5.2.1 Basic extreme theory for i.i.d sequences

There are some very complete books about the extreme value theory in the probabilistic asymptotic framework. Embrechts et al.([41]), de Hann and Ferreira ([36]) are such exhaustive books. Leadbetter et al.([92]) and Resnick([138]) are seminal books dealing with all the theory.

Let us first give some notations which are often used in the statistics of extremes.

Let $\xi \in \mathbb{R}$ be called the shape parameter. The basic extreme value distribution G_ξ is defined by :

$$\begin{aligned} &\text{if } \xi \neq 0, G_\xi(z) = \exp -(1+\xi z)^{-1/\xi} \text{ when } 1+\xi z > 0\,; \\ &\qquad\qquad = 1 \text{ if not} \\ &\text{if } \xi = 0, G_\xi(z) = \exp\left(-\exp\left(-z\right)\right). \end{aligned} \tag{5.1}$$

The expectation and variance of the distribution $G_\xi(z)$ are expressed as :

$$E(Z_\xi) = 1/\xi(\Gamma(1-\xi)-1); V(Z_\xi) = 1/\xi^2(\Gamma(1-2\xi) - \Gamma^2(1-\xi)) \tag{5.2}$$

where Z_ξ has the distribution G_ξ.

With this shape parameter ξ, the basic Pareto distribution H_ξ is defined by :

$$\begin{aligned} &\text{if } \xi \neq 0, H_\xi(z) = 1 - (1+\xi z)^{-1/\xi} \\ &\text{if } \xi = 0, G_\xi(z) = 1 - \exp(-z). \end{aligned} \tag{5.3}$$

The expectation and variance of the distribution $H_\xi(z)$ are :

$$\begin{aligned} E(Z_\xi) &= \frac{1}{1-\xi} \\ Var(Z_\xi) &= \frac{1}{(1-\xi)^2(1-2\xi)} \end{aligned} \tag{5.4}$$

Generalized extreme value distributions (GEV) are defined as the distribution of $\mu + \sigma Z_\xi$. $\mu \in \mathbb{R}$ is a location parameter, $\sigma \in \mathbb{R}^+$ is the scale parameter and Z_ξ has the distribution G_ξ

If M follows distribution $G_\xi(\mu, \sigma)$, we have :

$$E(M) = \mu + \sigma E(Z_\xi); V(M) = \sigma^2 V(Z_\xi) \tag{5.5}$$

and we address the model $M = \mu + \sigma Z_\xi$.

With the same principal, *generalized Pareto distributions* (GPD) $H_\xi(\sigma, u)$ are defined as the distribution of $u + \sigma Z_\xi$ where u is the threshold, σ is the scale parameter and Z_ξ has the distribution H_ξ.

Then the expectation and variance of $Y = X - u$ from the distribution $H_\xi(\sigma, u)$ are :

$$E(Y) = \sigma E(Z_\xi);\ V(Y) = \sigma^2 V(Z_\xi) \tag{5.6}$$

and the conditional expectation is :

$$E(Y - y|Y > y > 0) = \frac{\sigma + \xi y}{1 - \xi} \tag{5.7}$$

For GEV, as for GPD, the case $\xi > 0$ is "heavy-tailed" or Fréchet, the case $\xi = 0$ is "medium-tailed", or Gumbel, and the case $\xi < 0$ is "short-tailed", or Weibull, the distribution has finite upper endpoint at $\mu - \sigma/\xi$ for GEV and $u - \sigma/\xi$ for GPD.

A function L is called slow varying if and only if for $t > 0$:

$$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1. \tag{5.8}$$

A function F is said to have a regular $\alpha-$ variation (or to be a Karamata function) if and only if there exists α such that :

$$\lim_{x\to\infty} \frac{F(tx)}{F(x)} = t^{\alpha} \tag{5.9}$$

For details see de Hann et al. [35].

Now consider the basic theory on extreme values.

Theorem 5.2.1 *(The GEV limit theorem)(see de Hann et al.[35])*

Let F be a distribution of probability and let $X_1, ..., X_n$ be a sequence of i.i.d variables with distribution F. Let $M_n = max(X_1, ..., X_n)$, then we have $P(M_n \leq x) = F(x)^n$. If there exists two sequences $(B_n) \in \mathbb{R}$ and $(A_n) \in R_*^+$ such that $\frac{M_n - B_n}{A_n}$ converges in distribution to some W then :

- There exists $\xi \in \mathbb{R}$ and $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^+$ such that $W = G_\xi(\mu, \sigma)$, F is then said to belong to the domain of max attraction D_ξ of G_ξ.

The following condition for the tail of F are necessary and sufficient in order that $F \in D_\xi$.

• $\xi > 0$: The distribution F belongs to the domain of max attraction of Fréchet with the index ξ if and only if $1 - F(x) = x^{-1/\xi} L(x)$ where L is a slowly varying function.

• $\xi < 0$: The distribution F belongs to the domain of max attraction of Weibull with the index ξ if and only if $1 - F(r_s - \frac{1}{x}) = x^{1/\xi} L(x)$ where r_s is the finite endpoint and L is a slowly varying function.

• $\xi = 0$:

The distribution F belongs to the domain of max attraction of Gumbel if and only if for some auxiliary function b for every $v > 0$

$$\frac{1 - F(y + b(y)v)}{1 - F(y)} \to e^{-v} \tag{5.10}$$

as $y \to \infty$. *Then*

$$\frac{b(y + vb(y))}{b(y)} \to 1. \tag{5.11}$$

as $y \to \infty$.

A stationary POT(I,σ, ξ) model is defined by the following model :

- $\mathcal{P}$ is an homogeneous Poisson process with intensity I. Let $t_1, ..., t_n$ be the jumps of $\mathcal{P}$ and $Y = \{Y_{t_1}, ..., Y_{t_n}\}$ is a sequence of pointwise i.i.d random variables with Pareto distribution $H_\xi(\sigma)$. The sequence Y and $\mathcal{P}$ are independent.

Theorem 5.2.2 *(The POT limit theorem)*
Let $X_1, ...X_j, ..., X_n$ *be a sequence of i.i.d random variables with distribution* $F \in D_\xi$. *Let* u_n *be a sequence of thresholds,* $u_n \to \infty$.
Consider the set of $\{j, X_j \geq u_n\}$. *Its cardinal is* N, *and the conditional distribution given by* $P(X_j > y + u_n \mid X_j > u_n) = F_n(y)$.
Let $n \to \infty$, *then if and only if* $nu_n \to e^{-I}$ *the point process defined on* $(0, 1)$ *by the Poisson variable* $N_n(w)$ *and* $(\frac{t_1}{N}, ..., \frac{t_N}{N})$ *converges to a Poisson process on (0,1) of the intensity* I. *The conditional distribution* $F_n(y)$ *converges in distribution to some Pareto distribution* H_ξ.

Remark : It has been shown (Pickands,[118]) that asymptotically, the excess values above a high level will follow a generalized Pareto distribution (GPD) if and only if the parent distribution belongs to the domain of attraction of one of the extreme value distributions (GEV). The assumption of a Poisson process model for the exceeding times combined with GP distributed excesses can be shown to lead to the GEV distribution for corresponding extremes.

In fact, we can relate parameters of GEV(μ, σ, ξ) to parameters of POT $(u; \lambda, \sigma^*, \xi^*)$ by :

$$\begin{aligned} \xi^* &= \xi \\ \sigma^* &= \sigma + \xi(u - \mu) \\ \log \lambda &= -\frac{1}{\xi} \log \left(1 + \xi \frac{u - \mu}{\sigma}\right) \end{aligned} \tag{5.12}$$

5.2.2 Extreme for stationary sequences with correlation

Let us begin with an example from de Hann et al. [35].

Example Let X_i be a stationary centered Gaussian process with correlation r_n. Let (A_n) and (B_n) be the sequences of normalization for a sequence $Y_1, ..., Y_n$ of standard Gaussian variables. Let (A_n), (B_n) be the sequences such that

$$\lim_{n\to\infty} P\left(\frac{max(Y_1, ..., Y_n) - B_n}{A_n} < x\right) \to exp(-e^{-x})$$

If

$$\lim_{n\to\infty} r_n logn = \gamma \geq 0$$

then

$$\lim_{n\to\infty}\left(\frac{max(X_1, ..., X_n) - B_n}{A_n} \leq x\right) = P(M + N\sqrt{2\gamma} - \gamma \leq x)$$

where M and N are independent, N is normal and M is standard Gumbel distribution. So if $r_n = o(1/logn)$, we obtain a Gumbel distribution.

The basic definitions and asymptotic results for a stochastic stationary process whose marginal distribution $F \in \mathcal{D}_\xi$ for some ξ is due to Leadbetter ([92]). Let (A_n) and (B_n) be the normalized sequences of the maxima of a sequence $Y_1, ..., Y_n$ of i.i.d variables with distribution F. Then Leadbetter ([92]) gives conditions (D_n) and (D'_n) such that :

$$\frac{max(X_1, ..., X_n) - B_n}{A_n} \xrightarrow{d} G_\xi \tag{5.13}$$

Let $u_n(x) = B_n + A_n x$. Let $l, p \in \mathbb{N}$. Let $F_{i_1,...,i_p}$ the multivariate distribution of $X_{i_1}, ..., X_{i_p}$, $D(u_n)(x)$ is satisfied if and only if $\forall p, \forall 1 \leq i_1 \leq ... \leq i_p < j_1 < ... < j_p \leq n$, there exists a sequence $l_n \to \infty$ and $l_n = o(u_n)$ with $j_1 - i_p \geq l_n$ such that :

$$\mid F_{i_1,...,i_p,j_1,...,j_p}(u_n) - F_{i_1,...,i_n}(u_n)...F_{j_1,...,j_p}(u_n) \mid \leq x_{n,l_n} \tag{5.14}$$

for $x_{n,l_n} \to 0$.

It is easier to check condition $D'(u_n)$: there exists $u_n \to \infty$ such that :

$$\lim_{n\to\infty} \sup\, n \sum_{j=1}^{[n/k]} P(X_1 > u_n, X_j > u_n) \to 0 \tag{5.15}$$

Theorem 5.2.3 *For $D(u_n)$ and $D'(u_n)$ true then* (5.13) *is true*

Many important results on extremes are given in Resnick ([138]), Davis ([122]) and Rootzén([127]) about the case of linear process, especially in cases where the innovation belongs to the Gumbel domain of attraction or

has an heavy tail. For positive bounded innovation, see the paper of Davis and McCormick([31]).

Non linear process results are given for ARCH and GARCH processes, autoregressive processes with multiple threshold, and Markov chains (see Smith et al.[142] and Perfeckt [116]). In chapter 8, we are concerned with a class of Markov process or multidimensional Markov processes.

5.2.2.1 The extremal index θ for GEV

In fact, for GEV modelling, the effect of dependence is often a replacement of the limit distribution G by its θ-power G^θ where θ is the "extremal index", $0 < \theta \leq 1$. More precisely, if G corresponds to the GEV distribution with parameters (μ, σ, ξ) and $\xi \neq 0$ then :

$$\begin{aligned} G^\theta(z) &= exp\left\{-\theta\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \\ &= exp\left\{-\left[1+\xi^*\left(\frac{z-\mu^*}{\sigma^*}\right)\right]^{-1/\xi^*}\right\} \end{aligned}$$

where

$$\mu^* = \mu - \frac{\sigma}{\xi}(1-\theta^\xi),\ \sigma^* = \sigma\theta^\xi \text{ and } \xi^* = \xi$$

Accordingly, the location and scale parameters of the two distributions are different but their shape parameters are equal. For dependent observations, the maxima are often in the same domain of attraction, as for the independent case. We give a reference to Leadbetter et al. ([92]) for a detailed mathematical treatment and in fact the main result has a simple heuristic interpretation.

Another way to interpret the extremal index of stationary series is in term of the propensity of the process to clusters at extreme levels. More precisely, θ is the inverse value of the average cluster size in the limiting point process of exceeding times over high thresholds. When $\theta = 1$, the dependence is negligible at *asymptotically high* levels. When $0.5 \leq \theta \leq 1$, the dependence is not strong, the average value of cluster length is lower than 2, and the parameter θ can then be ignored. (The term θ^ξ which plays the important role in the difference of μ^*, σ^* with μ, σ is rather close to 1 with $|\xi| < 0.5$).

In order to estimate θ, the three common approaches are the block methods, the runs method and the inter-exceedance times method. The two first methods identify clusters and construct estimates for θ based on these clusters. For each, there are two decisive parameters : a threshold and a

cluster identification scheme parameter. The third method is based on inter-exceedance times and does not need a cluster identification scheme parameter. The blocks method is used, in our case, to estimate the extremal index. For more details and more bibliographies about other methods, one can see Robert et al.([126]).

5.2.3 Non-stationary EVT models

In this section, we present some existing theories about non-stationary extreme models and the different methods to estimate their parameters. Especially, we will propose a new algorithm to estimate nonparametrically the extreme parameters. The algorithm is based on the Fisher scoring algorithm and accompanied with an automatic choice of smoothing parameter in each iteration. This algorithm is introduced in Chapter 3 as a general method. In this chapter, it is used particularly for extreme models.

Non stationarity on distribution

There are few results which that can be applied. The most known is due to Mejzler (see de Hann et al.[35]) :

Theorem 5.2.4 *Let $X_1, ..., X_n$ be a sequence of independent random variables with distribution $F_1, ..., F_n$. If there exists sequences (A_n), (B_n) such that*

$$\frac{max(X_1, ..., X_n) - B_n}{A_n} \to G$$

and if

$$| \log A_n | + | B_n | \to \infty, | \frac{A_{n+1}}{A_n} | \to 1 \text{ and } | \frac{B_{n+1} - B_n}{A_n} | \to 0$$

$x^*(G)$ is the upper bound of G then :
- *$\log G(x)$ is convex if $x^*(G) = \infty$*
- *$\log G(x^* - e^{-x})$ is convex if $x^*(G) < \infty$.*

Any G with this property is a limit in the previous formulation.

This result proves that without the hypothesis of identical distribution, the set of limits G is very large and of nonparametric character.

This kind of results is a limit for statistical applications in non stationary situations.

Definition 5.2.1 *A non-stationary GEV model is a sequence $(M_n)_{n\in\mathbb{N}}$ of independent random variables whose distribution is given by $G_{\xi_n}(\mu_n, \sigma_n)$. The parameters then are the functions of time.*

A non-stationary POT model is given by :

a/ a Poisson process $\mathcal{P}(t)$ *on* $(0,1)$ *with the intensity* $I(t)$*.*

b/ If $\mathcal{P}(1) = N$*, and if* $t_1, ..., t_N$ *are the jumps of* $\mathcal{P}$*, there exists a sequence* $X_1, ..., X_k, ..., X_n$ *such that* $(X_{t_1}, ., X_{t_j}, .., X_{t_N})$ *are independent, have a Pareto distribution* $H_{\xi_{t_j}}(\sigma_{t_j})$ *and are independent of* $\mathcal{P}$*.*

5.3 Statistics of extremes

5.3.1 Basic methods for i.i.d data

For i.i.d random variables $X_1, ..., X_n$ with distribution $F, F \in \mathcal{D}_\xi$, there is a well-founded theory to evaluate the shape parameter ξ (Hill and other classical estimators , see de Hann et al. [35]) based on the behaviour of the ordered statistics of the whole sample. In the same way, the upper bound of F can be estimated from the whole sample as extreme quantiles.

From a practical point of view, this theory is not sufficient. Even if ξ is well estimated, many applications require a complete estimation of the distribution of the maxima M_n, or of the distribution of exceedances and intensity for a POT approach. At this point, the statistical theory encounters some difficulties, serious enough to be contested by some people. The traditional aspect is to approach the distribution of $\frac{M_n - B_n}{A_n}$ by a GEV distribution $G_\xi(\mu, \sigma) \sim \mu + \sigma G_\xi(0,1)$, so to consider the approximation $M_n \sim G_\xi(B_n, A_n)$ as good enough to be used in the following way.

Let $B_1, ..., B_j, ..., B_{N(n)}$ be disjointed adjacent blocks of $(1, ..., n)$, $\mid B_j \mid = b(n)$. If $b(n) \to \infty$, the previous consideration can be applied to every block, which is considered as very large. But of course the number of blocks $N(n) = n/b(n)$ has to tend to infinity. And so we have to approximate simultaneously $N(n)$ the distribution of the $N(n)$ variables $M_{j,n}$ which are the maxima of $X_{\alpha_j}, ..., X_{\alpha_j + b(n)}$ for $B_j = (\alpha_j, \alpha_j + b(n))$. This requires a good convergence of the empirical distribution F_M^n of $M_{1,n}, ..., M_{N(n),n}$ to that of GEV distribution $G_\xi(\mu, \sigma)$. This step is almost ignored in most practical applications, for it is very difficult to take into account.

The traditional use of GEV models just consists in taking the distribution of M_n as $G_\xi(\mu, \sigma)$ and to then apply classical methods of parametric estimation, such as maximum likelihood (m.l), moments or Bayes estimates. In estimating the extreme parameters, the maximum likelihood is often preferred. An advantage of maximum likelihood over other techniques of parameter estimation is its adaptability to changes in model structure. That is, although the estimating equations change if a model is modified, the underlying methodology essentially stays unchanged. Nevertheless, a potential difficulty with the use of likelihood methods for extreme models concerns the

regularity conditions, that are required for the usual asymptotic properties associated with the maximum likelihood estimator to be valid. Smith([140]) has obtained the following results :

• when $\xi > -0.5$, maximum likelihood estimators are regular, in the sense of having the usual Gaussian asymptotic properties.

• when $-1 < \xi < -0.5$, maximum likelihood estimators are obtainable, but do not have the standard asymptotic properties. The speed of convergence is larger than $\sqrt{n}$ in this case.

• when $\xi < -1$, maximum likelihood estimators can not be applied.

Some care is needed to ensure that such algorithms do not move to parameter combinations violating the condition $1 - \xi(x - \mu)/\sigma > 0$ in GEV or $1 - \xi(x - u)/\sigma > 0$ in POT.

Mutatis mutandis the same approximations can be done for POT models. The size block $b(n)$ is now replaced by the threshold $u_n \to \infty$. The exceedances over $u(n)$ are modeled in the stationary i.i.d case by a Pareto distribution $H_\xi(\sigma)$ and the intensity is taken as constant. The situation is simpler than for GEV because in this case the approximation is global.

The choice of threshold u_n plays an important role in the POT approach. To make a good choice, two methods are available : the first one is an explanatory technique from the characteristic of the expectation of the model, and the other is an assessment of the stability of the shape parameter

• *Stability of the shape parameter* :

If the distribution of $(Y_u|Y_u > 0)$ is $\mathrm{GPD}(\sigma_u, \xi)$ for a threshold u ($Y_u = X - u$), then with a higher threshold $v > u$, the distribution of $(Y_v|Y_v > 0)$ is a $\mathrm{GPD}(\sigma_u + \xi(v - u), \xi)$.

This means that ξ does not change from some high value of the threshold.

In practice, we can try many values of u to estimate the parameters of the GPD, then take a sufficiently high value from which ξ does not change a lot.

• *Mean excess over a threshold :*

If X follows a generalized Pareto $H(u; \sigma, \xi)$, when $\xi > -1$ and $\sigma - u\xi > 0$,

we have :

$$E(X - u|X \geq u) = \frac{\sigma - u\xi}{1 + \xi} \tag{5.16}$$

In practice, we can choose the threshold from this linearity by a graphical technique plotting the mean excess against u. The threshold is chosen as the smallest possible one which satisfies both the precedent properties.

In similar manner as with the block maxima approach, we need some treatments when the dependence remains among exceedances over a high threshold. Due to the clustered form of the exceedances showing a correlation between themselves, the exceeding dates do not appear as a sample of a Poisson process. Various suggestions, with different degrees of sophistication, have been made to deal with this problem. Among them, the most widely-adopted method is **declustering**, which corresponds to a filtering of the dependent observations to obtain a set of threshold excesses approximately independent.

We have a "cluster" of length k, which begins at the point $t + 1$ over a certain threshold v, if there exists a set of $k + 2$ consecutive observations so that $X_t < v$, $X_{t+i} > v$ with $1 \leq i \leq k$ and $X_{t+k+1} < v$.

From this type of cluster, we describe the declustering procedure : in each cluster U let $t + i^*$ be the moment of the maximum Y_{t+i^*} in U. The set of all Y_{t+i^*} of all the clusters is a sequence D,called "declustered" sequence, of independent values. Then a sample of exceedances over a threshold $u(u > v)$ is a subset of D satisfies $Y_{t+i^*} > u$.

The method is simple, but has its limitations. In particular, results can be sensitive to the arbitrary choices made in cluster determination. Information can be wasted while discarding all data except the cluster maxima(Coles,[22]).

5.3.2 Estimation methods for the non-stationary case

In general, the only way to get asymptotic results in a nonparametric framework is to reduce the time on [0,1] by the transformation $j \to j/N(n)$. This is what we shall do.

In the non-stationary context of GEV, if in every block the distribution is stationary but the distribution depends on the block label, then for block j, we take the approximation $G_{\xi_j}(\mu_j, \sigma_j)$.

At this step, as is often the case for statistical approaches of non-stationary models, we can choose a parametric or nonparametric representation for the function $j \longrightarrow \phi(j)$ where $\phi(j) = (\xi_j, \mu_j, \sigma_j)$.

Now the distribution of the data is often non-stationary with "slow" changes. Two approaches can be taken :

1/ Neglect the "slow" changes inside a given block.

2/ Suppose that there exists a function on $[0,1] : t \longrightarrow \phi(t)$ in a given block. The maximum value is reached for the jth position, then we approach its distribution by a distribution $G(\bar{\xi}(j), \bar{\mu}(j), \bar{\sigma}(j))$ where $\bar{\phi}(j) = \phi(j/N(n))$.

Indeed, these two points of view can be confounded if one uses, as a second type of probabilistic approximation, the approximation of the distribution of the variables $X_{j+1}, ..., X_{j+N(n)}$ by some F_j, and then that of M_j by $G_{\xi_j}(\mu(j), \sigma(j))$. These approaches are heuristically serious and can be checked by simulations. There nevertheless remains that some mathematical results are required.

For the non-stationary POT models, there are difficult problems concerning the Poisson intensity. One important problem of what is the asymptotic is discussed in the appendix of Chapter 8 on return levels. In fact, for the exceedances, the whole set of observations can be considered as a set of observations of random cardinal of the distribution $H_{\xi(t)}(\sigma(t))$ in the previous sense. We consider the observations $Z_{t_1}, ..., Z_{t_N}$ of distribution $H_{\xi(t_1)}(\sigma(t_1)), ..., H_{\xi(t_N)}\sigma(t_N)$ and then we can apply parametric representations and use classical statistical tools.

Of course, combination of non independence and non stationarity can be "justified" in the same way as in GEV. It seems easier to do this work for POT models than for GEV ones. For the latter one, we need a theory of maxima of triangular arrays of blocks which seems fastidious to write.

An implicit difficulty in any extreme value analysis is the limited amount of data for the estimation of models. From this point of view, POT is more advantageous than GEV. Modeling only block maxima (classical size of block is a year) is really a wasteful approach because a lot of information is lost.

5.3.2.1 Parametric estimation

For the parametric approach, some simple studies of the theoretical properties of particular classes of non-stationary processes are described in Leadbetter et al. ([92]). Then, improvements and applications have been proposed in different fields : Smith ([141]) gives basic examples for environmental applications, Coles ([22]) and Katz et al.([86]) developed the EVT for the fields of oceanography and hydrology. In these works, extreme models are fitted with polynomial trends. Recently, Parey et al.([114]) proposed to approach the problem of the form of the trends, for air temperature extremes, by continuous piecewise linear functions.

The advantage of a parametric approach is that it can sometimes give a direct asymptotic as $j \to \infty$, but only sometimes as is shown on an example in the appendix of Chapter 8.

In our parametric approach, we assume most often the parameters of GEV, or POT, given as polynomials or continuous piecewise linear function (CPL).

CPL models are very useful in practice. First for the usual mean and variance, they allow to split the observation periods into homogeneous parts following increasing or decreasing criteria (see Chapter 8 on return levels). For return levels also, they are probably the best tool to extrapolate parametric models.

However difficult problems, practical as well as theoretical, remain when using CPL in our situation. The first problem is linked to what can be named here "a sampling effect due to the boundaries", for instance, in our case due to some rare events in the first or last period. For example, when considering the temperature series until 2003, we can obtain a very strong linear trend because of the heat wave of 2003. We get catastrophic return levels if we use the last and short piece as trend to extrapolate the parameter.

Another theoretical and practical difficulty is that CPL models are not identifiable, i.e the same model can be defined by different sets of parameters. The classical theory of likelihood is not valid anymore and one has to use a specific theory similar to that given for mixture models as in Dacunha-Castelle and Gassiat ([30]). This issue results in problems of estimation of parameters that can have no consistent estimators. And as a matter of fact, the theory of test of likelihood for nested models is no more applicable, the limits are not the usual χ^2 ones. This can be avoided under the constraint of separation of angular nodes, e.g. by 10 years.

The choice of the optimal degree with the previous constraint in polynomial modeling are based on the Akaike criterion. This is no more valid for CLP modeling, again because of a lack of identifiability. Nevertheless its use can be justified if one finds a "clear" break (see the same problem for mixture in [30]).

5.3.2.2 Nonparametric or semiparametric estimation

Nonparametric and semiparametric smooth are well-known and have been applied in many fields. There are some works exploring these techniques in the extreme value context. Davison and Ramesh([33]), Hall and Tajvidi([69]), Ramesh and Davison([123]) proposed a local likelihood approach. Rosen and Cohen ([128]) instead suggested a penalized likelihood for the location parameter of a Gumbel model. This idea was then developed by Chavez-Demoulin([17]). In this latter work, in order to apply efficiently the automatic selection for smoothing parameters, a particular reparametrization to obtain orthogonal parameters is proposed. Related works were performed by Pauli & Coles ([115]) and Chavez-Demoulin & Davison([18]) who applied a similar approach and used bootstrap simulation, or a Bayesian approach, to calculate confidence intervals for the assessment of uncertainty.

Two approaches, local likelihood estimation (LL) and penalized likelihood estimation (PL), based on completely different principles, both give, in practice, rather similar results with the appropriate smoothing parameters. However, estimation of extreme models by penalized likelihood is nevertheless clearly more advantageous than by local likelihood for different reasons :

• PL requires much less computations compared with LL. For example, in local linear method, when estimating location and scale parameter, it requires computation of four-valued parameter estimators (rather than two). Moreover, LL estimation could become heavily saturated, particularly when design points are sparse (see Hall and Tajvidi, [69]).

• Automatic parameter selection does not work well in LL for local linear cases ([123]). To overcome this problem, Hall and Tajvidi ([69]) used local constant fits, but this approach can give serious down-weighted problems at the boundaries.

For those reasons, penalized likelihood estimation is especially attractive for us to estimate trends in extremes for datasets with numerous series over a large geographical area. In our context, we model extreme temperatures by both GEV and POT models.

The algorithm

The estimation procedure, which is introduced in Chapter 3, will be used here to estimate extreme parameters. Using penalized likelihood to estimate the extreme parameters, we also need the regular condition that $\xi > -0.5$.

We use here expected second derivatives, rather than observed second derivatives, to avoid the negative weight in iterative reweighted least squares. The performance of this approach has been justified through simulation, as shown in Chapter 3.

Let us recall that the estimate procedure is a combination of the equivalence of penalized likelihood maximization with iterative reweighted least squares minimization and iterative cross-validation of Gu. We realize that this procedure concerns three parameters $\mu, \phi = \log\sigma$ and ξ for GEV and on two parameters $\phi = \log\sigma$ and ξ for POT.

In the case of the GEV method, noting $I_{\theta\theta}$ the expected second derivative of θ, the estimation procedure is described as :
i/ Initialize : $\mu = \mu^0$, $\phi = \log(\sigma^0)$, $\xi = \xi^0$
ii/ Optimize μ *by CV cycle with fixed* ϕ, ξ *:*
$\mu^{(1)} = \mu$
Repeat the optimal procedure combining by CV : In the k^{th} *step, update :* $\tilde{w}_i = I_{\mu\mu}(\mu_i^{(k-1)})$, $\tilde{z}_i = \mu_i^{(k-1)} + \frac{\partial l}{\partial \mu}(\mu_i^{(k-1)})/\tilde{w}_i$ *and find out optimal* $\lambda^{(k)}$ *and* $\mu^{(k)}$ *by minimizing the cross-validation term :*

$$CV(\lambda \mid \tilde{z}) = \sum_{i=1}^{n} \frac{(\tilde{z}_i - \theta_{\lambda,i})^2 \tilde{w}_i}{(1 - A(\lambda)_{ii})^2} \tag{5.17}$$

Stop when $\lambda^{(k)}$ *and* μ^k *converge. Update* μ.
iii/ Optimize ϕ *by CV cycle with fixed* μ, ξ *:*
The same as optimization of μ *in ii/. Update* ϕ.
iv/ Optimize ξ *by CV cycle with fixed* μ, ϕ *:*
The same as optimization of μ *in ii/. Update* ξ.
v/ Alternate : step ii/, iii/, iv/ until convergence.

When considering ξ as constant, the step iv/, optimization of ξ with given μ, ϕ, results in choosing the ξ which maximizes the log-likelihood with these fixed μ, ϕ.

This procedure is usually convergent if the initial parameters are suitably chosen. From our experience, one can use the constant extreme parameters as initial values. They are estimated from stationary extreme models.

For the intensity of the POT model, we will simply estimate it by a kernel method through density estimation from the dates of exceedances. It is preferred to estimate the intensity by cubic splines with penalized likelihood method. However in the likelihood function of Poisson, there exists an integral term that must be approximated. This approach is more complicated. The kernel method can be sufficient for our purpose. The bandwidth in the kernel estimation will be chosen by cross-validation (see Schäbe, [132]).

5.3.3 Confidence intervals

In any statistical analysis, it is important to obtain measures of uncertainty on the estimates. This is especially true for extreme value modeling, where quite small model changes can be greatly magnified on extrapolation. Here, we propose an assessment of uncertainty based on confidence intervals.

In the stationary context, confidence intervals of estimators can be derived from the approximate normality of the maximum likelihood estimators . This approach which is based on the likelihood function still can be used in non stationary inference. Here, we will calculate the point-wise confidence intervals for extreme parameter based on a bootstrap resampling. We do not give here the numerical results on uniform confidence bands. Important theoretical, as well as practical, works on the confidence bands for spline estimators were given in Song & Yang ([143]).

For the block maxima method, once the trend has been identified, the block maxima series can be expressed as :

$$M_k = \hat{\mu}_k + \hat{\sigma}_k e_k \tag{5.18}$$

where M_k are the block maxima, $\hat{\mu}_k$,$\hat{\sigma}_k$ are the estimators of the location and shape parameters and e_k are residuals having a GEV distribution of location 0 and scale 1.

Bootstrap on the residuals to obtain the bootstrap samples M^* of observed block maxima. From these bootstrap samples, use the same trend identification method to estimate the extreme parameters. We obtain samples μ^*, σ^*, from which a $100(1-2\alpha)\%$ point-wise confidence interval can be derived by taking two empirical quantiles q_α and $q_{1-\alpha}$.

For the POT method, a quite similar procedure can be applied.

The number of dates of threshold exceedances follows a Poisson distribution.

For GPD of exceedances, we consider the following process for the values

V^u over threshold u :

$$V_i^u = u + \hat{\sigma}_i e_i \tag{5.19}$$

where $\hat{\sigma}$ is the estimated scale parameter and e the residual following a GPD of scale 1.

The point-wise confidence interval of $\hat{\sigma}$ is built by the bootstrap technique in the same way as for the parameters in GEV.

For the intensity $I(t)$ of the non-homogeneous Poisson process, a different work is required. A reference on the confidence intervals for the intensity is Cowling et al.([25]).

Let $j = 1, ..., S$ be the indices of S samples of simulations .

• Simulate N variables from the Poisson distribution $P(\Lambda)$ with $\Lambda = \int_0^1 \hat{I}(t)dt$. Let note N_u^j this sample.

• Take a sample $t_1^j, ..., t_{N_u^j}^j$ from the density $\frac{1}{\Lambda} I(t)$.

• Estimate a Poisson intensity $\hat{I}(t)^j$ from the jumps $t_1^j, ..., t_{N_u^j}^j$.

• For every t fixed, take the 5% and 95% quantiles of $(\hat{I}^1(t), ..., \hat{I}^S(t))$ as the confidence interval of $I(t)$.

5.3.4 Application to the series of temperature

5.3.4.1 Discussion on the behaviour of the shape parameter ξ

To illustrate the approaches outlined above, an application on the series of temperature will be developed. Both GEV and POT models will be considered in the non-stationary context. In both models, there is not a statistical evidence of the variation of the shape parameter ξ. In Nogaj et al.([105]) as in Chavez-Demoulin et Davison ([18]), the parameter ξ is considered as constant, after a long discussion to decide if it is possible to consider that the asymptotic approximation is of the same quality for both models (constant or not ξ) . The estimate of ξ in the first cited paper was obtained using polynomial trends and in the second one using nonparametric estimation. We also refer to Parey et al.([114]) for a detailed discussion about the constancy of ξ.

In this chapter, we discuss the behavior of ξ in a more precise analysis, using complete non parametric methods in GEV model. The role of the scale parameter in POT is similar to the one it has in GEV. Finally a constant ξ will be kept. Then both parametric and nonparametric approaches will be used to estimate the parameters in extreme models (GEV and POT). A comparison between the estimators from different methods allows us to

assess their empirical performance.

In the following, we call "algorithm 1" the algorithm which corresponds to the estimation of three "optimal" time-varying parameters from the previous estimate procedure, while "algorithm 2" corresponds to the estimation of optimal time-varying $\mu(t), \sigma(t)$ and a constant ξ from the optimization procedure.

First we take two distant stations, La Rochelle (France) and Uman(Ukraine), and we will consider GEV models on the block maxima of daily maximum temperature in summer (sequence X) in La Rochelle and block minima of daily minimum temperature in winter in Uman. We consider the behaviour of ξ and its influence on the other extreme parameters.

In figure 5.1, the left panel shows the study on the temperature in summer in La Rochelle(France), and the right one is for the temperature in winter in Uman(Ukraine). For both cases, we illustrate the estimators of the three parameters from algorithm 1 with the confidence intervals at the level of 90 percent based on the normality of the estimators when the number of observations is large. On the other hand, we add in the same graphics the optimal estimators of the three parameters when ξ is constant (algorithm 2) . We can see that the curves are very close. Moreover, the constant ξ lies completely in the confidence interval of the varying ξ. After this study, it seems reasonable to consider ξ as a constant.

In Chapter 6 on K hypothesis, we propose an other method to test the constancy of ξ. Its asymptotic properties and methodology will be detailed in that chapter.

5.3.4.2 Estimation of extreme parameters

For a more complete study, we take here two series of temperature : a long and a short one. The long one will be the series of daily maximum temperature in Déols(1901-2006) and the short one will be the series of daily maximum temperature in Bourges(1946-2001). We take these two geographically close stations on purpose to consider the quality of the estimators. If the estimators are correct, we must obtain the similar evolutions in the two stations for the same period . In both series, we consider only the observations in summer : 14th June to 21st September, including 100 days.

GEV application

In the block maxima approach, a study of the correlation structure of extreme events over a fixed threshold is needed. More precisely, we calculate

the cluster rate θ, which can be estimated by the ratio of the number of clusters with the number of values above a fixed threshold. An initial threshold is chosen as the threshold which corresponds to 8% of the dataset. The value of θ found for the series in Déols and Bourges is respectively 0.51 and 0.54, which corresponds to an average cluster length smaller than 2. The dependence in extreme values is then not very important and, we can ignore it.

In the parametric approach, firstly we use polynomials of degree 0 to 5 for μ and of degree 0 to 3 for σ. The optimal polynomials for μ and σ are determined by AIC and likelihood-ratio test. Secondly, we estimate μ by one to three piecewise linear function with the optimal polynomial estimator of σ(constant in our case), in the same principal as for polynomials, we can retain the optimal CPL. The two piecewise estimator of σ with the optimal polynomial estimator of μ is also considered. In the nonparametric approach, we use the estimate procedure with fixed ξ(algorithm 2) where the estimators are optimized by iterated CV. With a size of block as 25, we obtain the results which are shown in Figure 5.2.

For Déols, when estimated by polynomials, the degrees of optimal polynomials retained for μ and σ are respectively 5 and 0. For Bourges, they are respectively 3 and 0. Observing Figure 5.2, the first remark is that the estimate trends in Déols and Bourges are rather similar in the same period. The series in Déols is much longer than in Bourges, which leads to much more variable trends. In block maximum temperatures of Bourges, there are two values significantly small because of the interpretation of a cold year (1982), but a good news is that the extreme estimators are not affected by them, and so give correct evolutions globally.

On the other hand, the different approaches are close to each other, although they do not give us exactly the same curves. Nonparametric estimators are very flexible and adapted to the characteristics of the datasets : more variable for Déols' temperature and smoother for Bourges. Comparing with nonparametric estimators, polynomial estimators are rather flexible, but have difficulties to follow some sudden changes in extreme events. CPL works rather well for μ, but not very well for σ. More precisely, when the change occurs in the first or last years, CPL gives a too high-slope straight line because of the edge effect. When using CPL to extrapolate the parameters in the future, one should take the previous piecewise if the last piecewise increases or decreases brutally.

POT application

We take a threshold corresponding to 1% of the dataset to decluster the series of temperature. Using two strategies of threshold selection which were discussed in section 5.3.1, we take respectively the threshold 30.9 for the

series of Déols (corresponding to a number of exceedances NU=450) and the threshold 30.5 for Bourges (corresponding to NU=250). With a fixed ξ, we will use all three approaches like in GEV to estimate σ in GPD and the intensity I in the Poisson process. The results are shown in Figure 5.3.

For Déols, for the estimation by polynomials, the degrees of optimal polynomials retained for I and σ are respectively 2 and 1. For Bourges, they are respectively 2 and 0.

As in GEV, the different estimators are close to each other and the estimated trends in Déols and Bourges are rather similar over the same period. Nonparametric estimation for the intensity follows the variability of the exceedances in a better way, but we must take care of the choice of the smoothing parameters. Here, the smoothing parameters are chosen by cross-validation, but of course it does not always work well. The estimator of I in Déols by kernel seems too variable, which means a too small bandwidth.

For CPL, we can clearly see the boundary effect in the estimation of σ for the last part. In general, we can ignore this kind of trend.

APPENDIX

1. Some developments of GEV

In the estimation procedure of extreme parameters, we use these following informations about the first and second derivatives of the likelihood function of GEV.

All the terms below are calculated at a point $t : \mu(t) = \mu$, $\sigma(t) = \sigma$ and $\xi(t) = \xi$.

Note $a = 1 + \xi/\sigma(x - \mu)$, with $a > 0$ we have :

Log-likelihood :

$$l(x; \mu, \sigma, \xi) = \begin{cases} -\log\sigma - \left(1 + \frac{1}{\xi}\right)\log a - a^{-1/\xi}, \text{if } \xi \neq 0 \\ -\log(\sigma) - \frac{x-\mu}{\sigma} - exp\left(-\frac{x-\mu}{\sigma}\right), \text{if } \xi = 0 \end{cases}$$

First derivatives :

$$\begin{aligned}
l^{\mu} &= \left(\frac{\xi}{\sigma}+\frac{1}{\sigma}\right)\frac{1}{a}-\frac{1}{\sigma}a^{-1/\xi-1}, \text{if } \xi \neq 0 \\
&= \frac{1}{\sigma}-exp\left(-\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma}, \text{if } \xi = 0 \\
l^{\sigma} &= -\frac{1}{\sigma}+\left(1+\frac{1}{\xi}\right)\frac{a-1}{a\sigma}-\frac{1}{\xi\sigma}(a-1)a^{-1/\xi-1}, \text{if } \xi \neq 0 \\
&- \left[\frac{1}{\sigma}-\frac{x-\mu}{\sigma^2}+exp\left(-\frac{x-\mu}{\sigma}\right)*\left(\frac{x-\mu}{\sigma^2}\right)\right], \text{if } \xi = 0 \\
l^{\xi} &= \frac{1}{\xi^2}\log(a)-\left(1+\frac{1}{\xi}\right)\frac{x-\mu}{\sigma a}-a^{-1/\xi}\frac{\log a}{\xi^2}-\frac{x-\mu}{\xi\sigma a}, \text{if } \xi \neq 0
\end{aligned}$$

Second derivatives or components of minus observed information matrix. (The cross second derivatives $l^{\mu\xi}$ and $l^{\sigma\xi}$ are complicated and as it is not in our interest, we will not develop them here)

$$\begin{aligned}
l^{\mu\mu} &= \left(\frac{\xi^2+\xi}{\sigma^2}\right)\frac{1}{a^2}-\frac{\xi}{\sigma^2}\left(1+\frac{1}{\xi}\right), \text{if } \xi \neq 0 \\
&= -\left[exp\left(-\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma}\right], \text{if } \xi = 0 \\
l^{\sigma\sigma} &= \frac{1}{\sigma^2}-\left(1+\frac{1}{\xi}\right)\frac{a^2-1}{a^2\sigma^2}-\frac{1}{\xi\sigma^2}\left[2(1-a)a^{-1/\xi-1}+\left(1+\frac{1}{\xi}\right)(1-a)^2a^{-1/\xi-2}\right], \text{if } \xi \neq 0 \\
&= -exp\left(-\frac{x-\mu}{\sigma}\right)\left(\frac{x-\mu}{\sigma^2}\right)^2+exp\left(-\frac{x-\mu}{\sigma}\right)\left(\frac{2(x-\mu)}{\sigma^3}\right)+\frac{1}{\sigma^2}-\frac{2(x-\mu)}{\sigma^3}, \text{if } \xi = 0 \\
l^{\xi\xi} &= -\frac{2\log(a)}{\xi^3}+\frac{2}{\xi^2}\frac{x-\mu}{\sigma a}+\left(1+\frac{1}{\xi}\right)\frac{(x-\mu)^2}{\sigma^2a^2}-a^{1/\xi}\frac{\log(a)}{\xi^2}+\frac{x-\mu}{\xi^2\sigma^2a^2} \\
&+ a^{-1/\xi}\left[\frac{-2\log(a)}{\xi^3}+\frac{2(x-\mu)}{\sigma\xi^2a}+\frac{(x-\mu)^2}{\xi\sigma^2a^2}\right], \text{if } \xi \neq 0 \\
l^{\mu\sigma} &= -\frac{1+\xi}{\sigma^2a^2}+\frac{a^{-1/\xi-1}}{\sigma^2}+\frac{(1+\xi)(1-a)a^{-1/\xi-2}}{\xi\sigma^2}, \text{if } \xi \neq 0 \\
&= -\left[\frac{1}{\sigma^2}+exp\left(-\frac{x-\mu}{\sigma}\right)\frac{(x-\mu)}{\sigma^3}-exp\left(-\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma^2}\right]
\end{aligned}$$

Components of Fisher expected information matrix : (revealed by Pres-

cott and Walden(1980,[121]))

$$I^{\mu\mu} = E(-l^{\mu\mu}) = \frac{n}{p^2}$$
$$I^{\sigma\sigma} = E(-l^{\sigma\sigma}) = \frac{n}{\sigma^2\xi^2}(1 - 2\Gamma(2+\xi) + p)$$
$$I^{\xi\xi} = E(-l^{\xi\xi}) = \frac{n}{\xi^2}\{\frac{\pi^2}{6} + (1 - \gamma + \frac{1}{\xi})^2 - \frac{2q}{\xi} + \frac{p}{\xi^2}\}$$
$$I^{\mu\sigma} = E(-l^{\mu\sigma}) = -\frac{n}{\sigma^2\xi}\{p - \Gamma(2+\xi)\}$$
$$I^{\mu\xi} = E(-l^{\mu\xi}) = -\frac{n}{\sigma\xi}\left(q - \frac{p}{\xi}\right)$$
$$I^{\sigma\xi} = E(-l^{\sigma\xi}) = \frac{n}{\sigma\xi^2}\left[1 - \gamma + \frac{1-\Gamma(2+\xi)}{\xi} - q + \frac{p}{\xi}\right]$$

where $p = (1+\xi)^2\Gamma(1+2\xi)$, $q = \Gamma(2+\xi)\{\psi(1+\xi) + (1+\xi)/\xi\}$, $\gamma = 0.5772157$: Euler's constant. ψ is the digamma function $\psi(r) = \log'\Gamma(r)$.

2. Some developments of POT

We show here the expressions of fist and second derivatives of GPD process, which contains only two parameters, thus is much simpler than in GEV. All the terms below are calculated at a point $\sigma(t) = \sigma$ and $\xi(t) = \xi$ with $\sigma > 0, a = 1 - \xi(x-u)/\sigma > 0$.

Log-likelihood :

$$l(u;\sigma,\xi) = \begin{cases} -(x-u)/\sigma - \log\sigma, \text{if } \xi = 0 \\ -\left(-\frac{1}{\xi} + 1\right)\log a - \log\sigma, \text{if } \xi \neq 0 \end{cases}$$

First derivatives :

$$l^{\sigma} = -\frac{1}{\sigma} + \left(\frac{1}{xi} + 1\right)(\frac{a(1-a)}{\sigma}, \text{if } \xi \neq 0$$
$$= -\frac{1}{\sigma} - \frac{x-u}{\sigma^2}, \text{if } \xi = 0$$
$$l^{\xi} = -\left(\frac{1}{xi} + 1\right)(1-a)/(\xi a) + \frac{1}{xi^2}\log a, if\ \xi \neq 0.$$

Minus components of the observed information matrix :

Note $z = u/\sigma$

$$l^{\sigma\sigma} = \frac{1}{\sigma^2} - \left(\frac{1}{xi} + 1\right)\frac{2(a-1)}{a\sigma^3} - \frac{(1-a)^2}{\sigma^2 a^2}, \text{if } \xi \neq 0$$
$$= \frac{1}{\sigma^2} - \frac{2(x-u)}{\sigma^3}, \text{if } \xi \neq 0$$
$$l^{\sigma\xi} = \left(\frac{1}{xi} + 1\right)(\frac{1}{\sigma^2}\frac{x-u}{a} - \frac{(1-a)^2}{\xi\sigma a^2} - \frac{1}{xi^2}\frac{1-a}{\sigma a}, \text{if } \xi \neq 0$$
$$l^{\xi\xi} = \left(\frac{1}{xi} + 1\right)\frac{(1-a)^2}{\xi^2 a^2} + \frac{2}{xi^2}\frac{1-a}{\xi a} - \frac{2}{\xi^3}\log a, \text{if } \xi \neq 0$$

Components of the expected information matrix :

$$\begin{aligned} I^{\sigma\sigma} &= \frac{1}{\sigma^2 + 2\xi} \\ I^{\sigma\xi} &= \frac{-1}{\sigma(1+\xi)(1+2\xi)} \\ I^{\xi\xi} &= \frac{2}{(1+\xi)(1+2\xi)} \end{aligned} \tag{5.20}$$

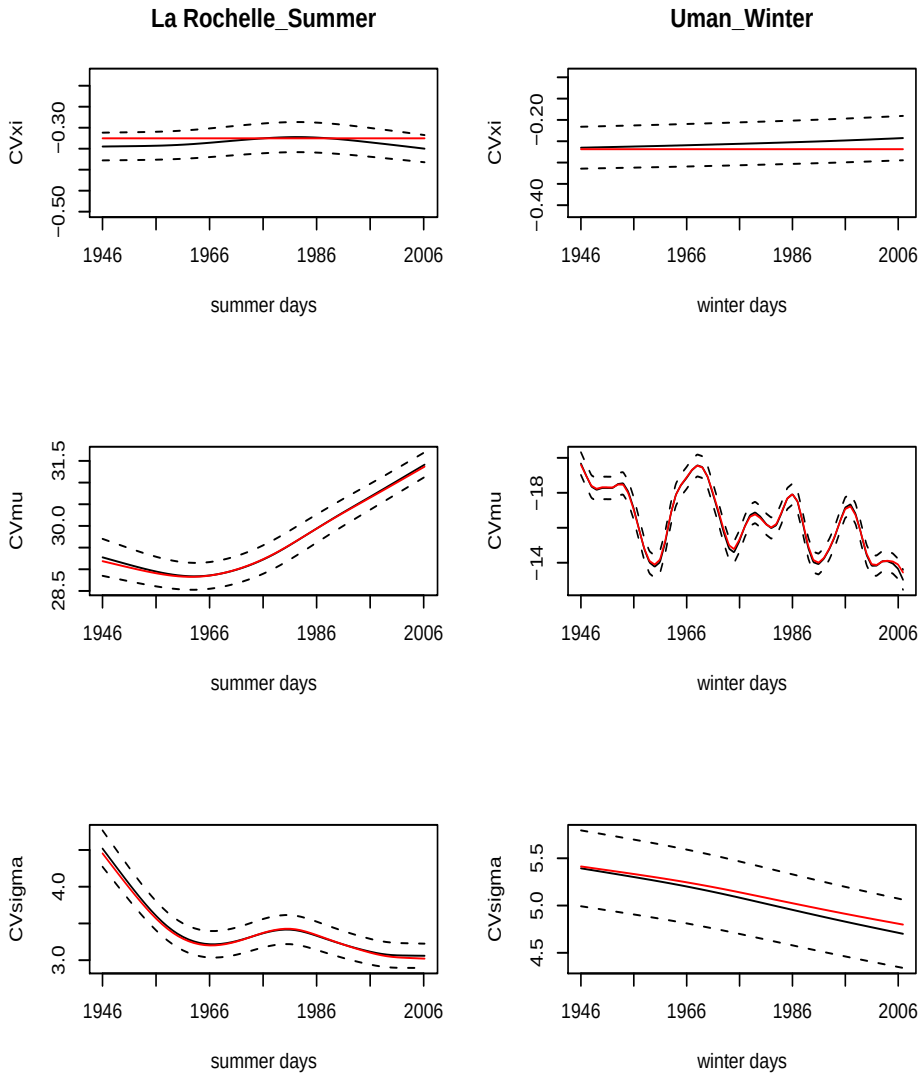


FIGURE 5.1 – The behaviour of the shape parameter ξ. Left panel. the winter in Uman. Right panel : the summer in La Rochelle. In black, optimal estimators of the parameters by algorithm 1 and in red, optimal estimators of the parameters with constant ξ from algorithm 2.

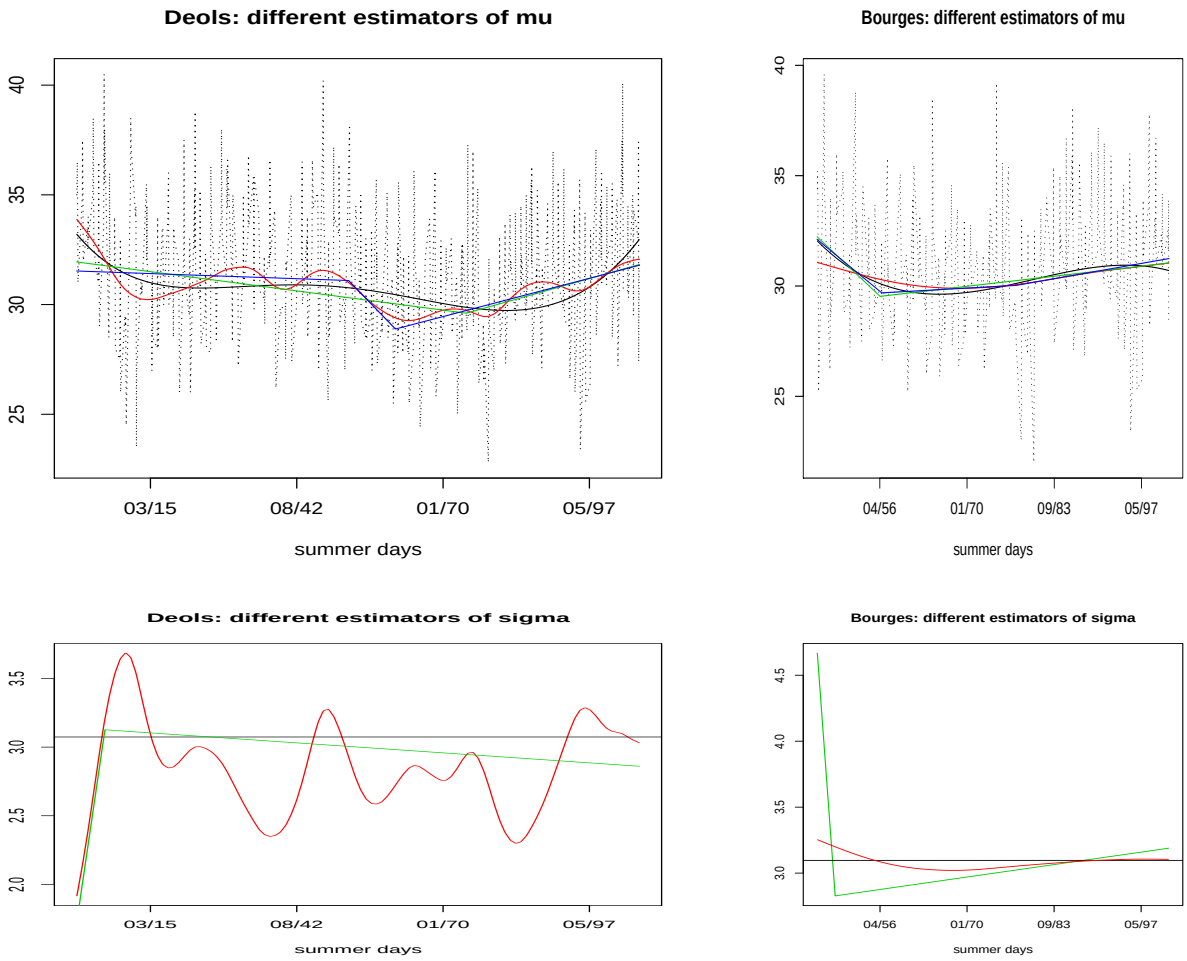


FIGURE 5.2 – Estimation of extreme parameters in GEV by different methods : optimal polynomials (black line), two piecewise linear function (green line), three piecewise linear function (blue line), cubic spline (red line) and the block maxima (dashed line)

FIGURE 5.3 – Estimation of extreme parameters in GEV by different methods : optimal polynomials(*black line*), two piecewise linear function(*green line*), three piecewise linear function(*blue line*),nonparametric(*red line*)

Chapitre 6

A new test of stationarity. Links between central field and extreme trends

6.1 Introduction

The studies in the non-stationary context allow us to understand better the evolution of the mean and the variance, which are considered to be an evolution of the "central field", and those of the mean and the variance of extreme events. It seems natural to ask if the temporal dynamics of the entire data set, and that of extremes, are closely related. If they really are, it is interesting to use this fact in the statistical studies of these variables. More precisely, we can perhaps extract more information about the central field in modeling extreme events. Consider the model :

$$X(t) = m(t) + s(t)Y(t) \tag{6.1}$$

We have :

$$P(Y_t < M) = P(\frac{X_t - m_t}{s_t} < M) = P(X_t < s_t M + m_t)$$

Assuming that the maxima of X_t and Y_t converge to an extreme value distribution, we can derive in the case of GEV ($\text{Max}X_t \sim \text{GEV}(\mu_X, \sigma_X, \xi_X)$, $\text{Max}Y_t \sim \text{GEV}(\mu_Y, \sigma_Y, \xi_Y)$),

$$\begin{cases} \xi_X(t) = \xi_Y(t) \\ \sigma_X(t) = \sigma_Y(t) s_X(t) \\ \mu_X(t) = m_X(t) + \mu_Y(t) s_X(t) \end{cases}$$

and in the case of POT (X with threshold u,parameters σ_X, ξ_X, intensity I ; Y with threshold v, parameters σ_Y, ξ_Y, intensity J),

$$\begin{cases} \xi_X(t) = \xi_Y(t) \\ \sigma_X(t) = \sigma_Y(t) + \xi(t)(s(t)v + m(t) - u) \\ I(t) = \left(1 + \frac{\xi(t)}{\sigma_X(t)}(s(t)v + m(t) - u)\right)^{-1} J(t) \end{cases}$$

where $s_t v + m(t) - u > 0$. For the case where $s_t v + m_t - u > 0$, we can find analogue relations for σ_Y and J by inverting X with Y and u with v.

From these systems of equations, we find that if the extremes of Y_t are stationary, that means its extreme parameters are constant, the extreme parameters of X are obtained by translation/scaling of the parameters m, s of the central field. From this characterization, it motives us to study the K hypothesis : "extremes of the centred and normed series Y are stationary" and especially to study how to test K on the datasets.

Of course, in practice, if the answer is yes, K can not be rejected, this just means that with the amount of data we have, it is not possible to have a statistical evidence for the non-stationarity of the extremes of Y.

The previous results on nonparametric statistics are applied to the case where M_{t_k} is the sequence of maxima (minima) in blocks of reduced temperature $Y_t = \frac{X_t - m(t)}{s(t)}$, where X_t is the maximal (minimal) daily temperature, $m(t)$ its mean and $s^2(t)$ its variance. We then discuss the physical meaning for observed temperatures in Europe as well as for data provided by climate models.

We only use the GEV approach, and not the POT, because it is easier to apply to a large number of datasets (from European stations for instance). The asymptotics of K are studied specifically on Weibull models where the extremes are bounded with $\xi < 0$, which is the case of the temperature.

Looking at the temperatures in Europe, if we estimate the mean and the variance of of the daily minimum (or maximum) temperatures as suitably smoothed functions of time (thought as continuous quantities, as intrinsic as possible), and if the same quantities are estimated for monthly minimum or maximum, what kind of links do we expect to find between these quantities ?

It is interesting to measure these similarities with something which looks like a distance between functions of time. A more precise question can then be asked : if the statistical extreme value models present a trend for some parameters, is it possible to describe it, or explain it, at least in part, by the parameters of the distribution of the whole dataset ? How can this idea be presented in the statistical language ?

Our work is based on homogeneous periods. The analyzed time series will be noted X_t, and the time evolution of its mean and variance respectively m and s^2.

As a first step, the links between the different trends were cautiously addressed in a previous work (Parey et al.,[112]) dedicated to the strong and complex link between mean and variance for summer or winter data. This link is an important feature of temperature series. Inheriting this idea, we will consider in this paper the links between the trends of the whole time series (that we call central field) and the trends (functional parameters have characteristics as mean and variance) of the extreme field. The extremes are described using the statistical extreme value theory (EVT) and the results will be controlled using the generalized extreme values models (GEV) (see Embrechts et al.,[41]). The peaks over threshold models (POT) are not mentioned in our paper, we just consider POT on some datasets to verify the methodologies. For more details on POT models, see Embrechts et al. ([41]), Chavez-Demoulin ([17], [18]) and Nogaj et al.([105]).

At first, it seems quite natural to address the H hypothesis which states that the centered and scaled data, $Y_t = (X_t - m(t))/s(t)$, also called reduced data, follow a stationary distribution H, but this is usually not verified in general. If H was true, the trends in the extreme field would only be due to translation/scaling effects of the central field and as it is easy to show these trends can be computed using only the mean m_X, the variance s_X and the constant parameters of the extreme model of Y_t.

This so-called H hypothesis has been often discussed. Katz and Brown ([85]) introduced "local properties" as the (relative) sensibility of the probability of extreme events to a change in mean, or variance, of the whole observation set. In a more global perspective, Ferro and Stephenson ([50]) proposed a procedure for an analogous test of H using the estimates of a limited set of percentiles at two different times. Then Nogaj et al. ([105]) consider the validation of H by introducing two ways of estimating the same model of extremes of X_t. The direct approach is an application of the non-stationary EVT to X_t, ignoring the information of the whole observation set. The second approach is an indirect one consisting in estimating during a first step m and s from the whole data set, then fitting a (non)stationary model for Y_t in a second one and finally reconstructing the extreme parameters of X_t in a final step. Then their differences are discussed to analyze the validating circumstances of H.

6.2 Hypothesis K : stationarity of extremes of reduced data

Unlike these papers, we do not study H but the weaker K hypothesis : " *the extreme model of* Y_t*, noted* ext_Y*, is stationary*" . This avoids taking into account some small dynamic deformations of the central part of the distribution of the reduced data whose physical interest is in general not evident. We have improved on the statistical tools in order to measure the departure from K. The main tool is a test of K based on a simple mathematical fact written here in an asymptotic framework.

Let $\theta \in \Theta$ be a vector of functional parameters of a family of densities with respect to the Lebesgue measure, where $\theta(t)$ is the value of θ for some $t, 0 < t < 1$. We consider a sequence Y_{t_k} of observations. Then for K either false or true, any "good" nonparametric estimator of $\theta(t)$ is convergent. On the contrary, if we try to estimate $\theta(t)$ as a constant by the maximum likelihood method, we obtain a good estimator if $\theta(t)$ is constant, and not if it is not. So we built a statistical test choosing a distance between these two estimators. Asymptotic theory allows approximating the level and the power of the test and easy simulations for a finite sample. We then apply these

general results to families of GEV densities with a negative shape parameter which needs some specific modifications for the proofs. Their difficulties are due to the support.

We now consider GEV models. The considered observations are divided into n consecutive blocks of days of same length. We select in each block the maximum value of the observations in order to obtain the sequence $M_1, M_2, \dots\ , M_k, \dots, M_n$. The distribution of M_k is approximated by the distribution of $\mu_k + \sigma_k Z_{\xi_k}$ where Z_{ξ_k} is a sequence of independent variables with a GEV distribution G_{ξ_k} of zero location parameter and unit scale parameter. We address here the model $M_k = \mu_k + \sigma_k Z_{t_k}$.

All those functions of the discrete time k will be considered as restrictions of functions defined for the continuous time t. Here we do not consider discrete time k but the date t_k of the maximum in the block k, for instance $\mu_k(t_k)$.

In what follows, the statistical models for the extremes of the variables X_t and Y_t are respectively denoted ext_X and ext_Y.

The K hypothesis supposes that the extreme model ext_Y of $Y_t = (X_t - m_(t))/s_(t)$ is stationary.

First, let us consider the relations (8.2.3) between the parameters of ext_X and ext_Y. $\sigma_Y(k)$ appears as the multiplier allowing to pass from the scale function s of the global sample at time t_k (corresponding to the maximum of block k), to the scale parameter $\sigma_X(k)$. The same role for $\mu(k)$ is played by $\mu_Y(k)$ through a translation of $m(k)$ and a multiplication with s_X.

From the formula (8.2.3), we can see that if Y_t is stationary, which is the H hypothesis, then the trends of the extremes can be exactly computed using m ,s and the constant parameters of ext_Y. There follows that the time dependency of ext_X is completely explained by m and s. A test for the stationarity of Y_t is considered in Ferro et al,([50]). However, H is too strong if we stress on the problem : "slow deformation of tails, with about the same speed as the global mean". For us this is the main question. We therefore suggest to test the K hypothesis "invariance in the extremes of Y_t", avoiding to focus on the possible weak deformations in the centre of the Y_t distribution. In other words, we try to see if the hypothesis that "the tails of Y_t are stationary" is verified or not. This hypothesis is translated under the form of the K hypothesis : "the parameters of ext_Y are constant".

If $\mu_Y(k)$ and $\sigma_Y(k)$ are constant, which corresponds to the validation of the K hypothesis, we can see that the difference of behavior between the trends of the whole sample and those of the extremes are completely due to

m and s. From the formula of the mean and variance of extremes, and from (8.2.3), we can express the mean and the variance functions of extremes differently as :

$$\begin{aligned} m_{extX}(k) &= m(k) + s(k)(\mu_Y(k) + \sigma_Y(k)E(Y)) \\ s_{extX}(k) &= s(k).\sigma_Y(k)/\xi.(V(Y))^{-1/2} \end{aligned} \tag{6.2}$$

If K is true, then this formula shows that the strong relation described in Parey et al.([112]) between m and s is transported to a relation between the mean function m of the whole observations and the mean function m_X of the extremes. Thus the verification of K implies a strong relation between the trends in the central field and the trends of the extremes. This point gives the "connection" between the K hypothesis and the study of the correlation between the trends of the central field and extremes in our previous paper (Hoang et al., [79]).

6.3 A new test for stationarity. Application to extreme models

In this section, we propose a new test for the stationarity. Our approach is different from that of Foutz and Strisvatava ([52]).The asymptotic properties and numerical applications of the test will be described in the following.

6.3.1 Asymptotic theory of the test

The set of possible evolutions of ext_Y parameters is very large. Of course it can be due to a possible evolution of ξ, that we have previously more or less excluded, or to more subtle deformations. So a direct test suffers from the absence of natural alternatives. This is one of the reason why we prefer the use of the distance Δ between functions of time to study K. If we estimate a function of time f by g, $\Delta(f,g) = (\int_{t\in D}(f(t)-g(t))^2dt)^{-1/2}$ is a measure of the quality of g as an estimate of f. We have two estimates of the parameters of ext_Y : $\tilde{f}$, a non parametric estimate, and $\hat{f}$ a constant estimate obtained if K is true, this means that ext_Y is stationary. Now, in any case (K true or false), one can prove that $\tilde{f}$ converges to f when the sample size T tends to infinity and the smoothing parameter λ tends to 0.

The situation is of course different for $\hat{f}$. If K is true, $\hat{f}$ converges to f with a speed of order $\sqrt{T}$ and in this case $\Delta(\hat{f},\tilde{f})$ is, for a large sample, very close to $\Delta(f,\tilde{f})$. On the contrary, if K is false, $\Delta(f,\hat{f})$ does not tend to zero and the same is true for $\Delta(\tilde{f},\hat{f})$. The intuitive reason is that we try to find f in a set of functions "far away" from f when K is false. This argument could be translated in an asymptotic result.

Let $X_t, t = 1...N$ be a sequence of observations with distribution $G(\theta(t))$ where $\theta(t)$ is a functional parameter with values in $\mathbb{R}^p$ and $G(\alpha)$ is a known parametric family of distributions for $\alpha \in C \subset \mathbb{R}^p$. $\theta(t) \in \Theta$, with Θ a set of functions such that $G(c) \in \Theta$ for every constant function $c \in C$. Let $g(c, x)$ the density of $G(c)$ (supposed to exist).
In this part we want to test the following hypothesis :

$\mathbf{H}_0$: $\theta(t)$ is a constant , or

$\mathbf{H}_0'$: some coordinate of $\theta(t)$ is constant.

If the variables are not independent, we test if the marginal distribution of order 1 stationarity. If they are independent, we test their identically distributed property.
As an alternative hypothesis H_1, we consider $\inf_c \|\theta(t) - c\| > 0$ where $\|\|$ is some norm on Θ.
As an example of such a model we can consider a GEV model with parameters $\theta(t) = (\mu(t), \sigma(t), \xi(t))$.

Let $\widehat{\theta}_n(t)$ be a nonparametric consistent estimator of $\theta(t)$, for instance, a spline estimator (studied in section 3.2 of Chapter 3).Its speed of convergence can be given when Θ is included in some Besov space.

We suppose that $\|\widehat{\theta}_n(t) - \theta^0(t)\| \to 0$ where $\theta^0(t)$ is the true value of the parameter $\theta(t)$. We will use this kind of convergence result which requires N large enough and also some regularity of functions in Θ.

If $\theta^0(t) = c$ for some c, then $\|\widehat{\theta}_n(t) - c\| \to 0$.

Now we estimate $\theta(t)$ as a constant, even if it is not. Let $\widehat{c_n}$ be the maximum likelihood estimate. Of course it does not converge to θ if θ is not constant. We will prove that under some hypothesis, $\widehat{c_n}$ converges to $\underset{c}{\operatorname{argmin}}\, K(\overline{G}(\theta^0), G(c))$ or at least has a convergent sub-sequence (adherence points) whose limit belongs to a set F with :

$$F = \left\{c, c = \operatorname{argmin}\ K(\overline{G}(\theta^0), G(c))\right\}$$

where K means the Kullback information $K(f, h) = \int [\log f(x)/g(x)] f(x) dx$, $K \geq 0$ and $\overline{G}(\theta_0)$ is a distribution function whose density is defined by

$$\overline{g}(x) = \int g(x, \theta(t)) dt$$

We have always $K(G(\theta^0), \overline{G}(\theta^0)) > 0$ if θ^0 is not constant.

Now suppose that Θ, $\Theta \subset [C_\alpha(1, T)]^p$, $\alpha \geq 2$, $p \geq 1$, is a vector of functional parameters of a family of densities $g(x, \theta(t))$ with respect to the Lebesgue measure where $\theta(t)$ is the value of θ for some $t \in [0, 1]$.

We suppose that the following conditions A and B are satisfied :

A1-$g(x, \theta(t)$ is uniformly continuous in (x, t).

A2- Let C be a compact set of constant functions denoted by vectors c, $C \subset \Theta$ and $\{g(x,\theta(t)), x \in R^p, t \in (0,1)\} \subset C$.
In general,we can easily extend the model to satisfy these conditions. From now on, we suppose $t \in (0,1)$ in order to use classical asymptotic non-parametric theory. We consider a sequence $X_{t_k^n}$ of observations, $t_k^n = \frac{k}{n}$ instead of X_t.

B1- $\sup_{t,\theta(t)} |\log(g(x,\theta(t))| < h(x), h \in L^1(\bar{g})$.

B2- $\int [\log g(x,\theta_0(t))g(x,c)]^2 dx < \infty$ for every t.

B3- $g(x,\theta(t)) > 0$ for every (x,t).

The last condition will be relaxed later.

We have $\hat{c}_n = \underset{c\in C}{\operatorname{argmax}} \sum_{k=1}^{T_n} \log(g(X_{t_k}, c))$.

Let $\hat{\theta}_\lambda^n$ the non parametric estimator by splines given by :

$$\hat{\theta}_\lambda^n(t_k) = \underset{\theta\in\Theta}{\operatorname{argmax}} \left[\sum_{k=1}^{T_n} \log g(X_{t_k}, \theta_{t_k}) - \sum_{j=1}^{p} \lambda_j \int_1^T (\theta_j(t))^2 dt \right], \qquad (6.3)$$

with $\theta = (\theta_1, ..., \theta_p)$ and $\lambda = (\lambda_1, ..., \lambda_p)$ are the smoothing parameters.

We know that when $n \to \infty$ and $\lambda \to 0$, $\hat{\theta}_\lambda^n(t) \to \theta(t)$; the speed of convergence depends on λ (see Section 3.2). Our first purpose is to prove that when the hypothesis K : "$\theta(t) = c$ *for some* c" is false, the distance between $\hat{\theta}_\lambda^n$ and $\hat{c}_n$ tends to a positive constant, significantly different from zero, and given in the following lemma.

Lemma 6.3.1 *Let* $L_n(\theta) = \sum_{k=1...n} \log g(X(t_k^n), \theta(t_k^n))$, $L_n(c) = \sum_{k=1...n} \log g(X(t_k^n), c)$.
We have then :

$$\frac{1}{n}(L_n(\theta_0) - L_n(c)) \longrightarrow \int K(g(\theta_0(t)), \bar{g})dt + K(\bar{g}, g(c)) \qquad (6.4)$$

and $K(\bar{g}, g(c))$ *is continuous on* C.

The first term in the limit measures the *lack of stationarity* of the true model, and the second the *lack of convexity* of the global statistical model. This first term is strictly positive for a non constant $\theta(t)$. This is basic to test the stationarity.

Proof

From the Riemann convergence, we have :

$$\frac{1}{|D_n|} \sum_{t_k^n \in D_n} g(x, \theta_0(t_k^n)) \to g(x) = \int_1^T g(x, \theta_0(t))dt,$$

then,

$$\frac{1}{n} E(L_n(c)) = \frac{1}{n} \sum_k \int \log g(x,c) g(x, \theta_0(t_k^n))dx.$$

From the continuity of $g(x, \theta_0(t))$ in t for every x and by Lebesgue theorem,

$$\frac{1}{n}E(L_n(c)) \to \int \log g(x,c)\bar{g}(x)dx \tag{6.5}$$

Using the same trick and condition B2, it is easy to show that :

$$E\left(\frac{1}{n}L_n(c) - \int \log g(x,c)\bar{g}(x)dx\right)^2 = o(1)$$

So

$$\frac{1}{|n|}L_n(c) \xrightarrow{\mathrm{P}} \int \log g(x,c)\bar{g}(x)dx$$

In the same manner, we get

$$\frac{1}{|n|}\int \sum_k \log g(x,\theta_0(t_k^n))g(x,\theta_0(t_k^n\theta) \xrightarrow{\mathrm{P}} \int\int_1^T \log g(x,\theta_0(t))g(x,\theta_0(t))dtdx \tag{6.6}$$

By addition and subtraction of $\int \log \bar{g}(x)\bar{g}(x)dx$ to (6.6) and (6.6), we get Lemma 6.3.1 .

□

Lemma 6.3.2 *$F^1 = \{ \lim_{n\to\infty} \hat{c}_{n_k}$ where $\hat{c}_{n_k}$ is a convergent sub-sequence of $\hat{c}_n\} \subset F = \{c, K(\bar{g}, g(c))$ is minimum $\}$.*

Proof (extension of the proof of the mle consistency, see Dacunha-Castelle and Duflo ([29],t2,p93)).
Let c and c' be in C .

Let $w_0(\eta) = \sup_{|c-c']<\eta} |\log g(c,.) - \log g(c',.)|$

So $w_0(\eta,.) \leq 2h(.)$, $log(c,.)$ is $\bar{g}$ a.s in x uniformly continuous for $c \in C$, so $\lim_{\eta\downarrow 0} w_0(\eta) = 0$ implies $\lim_{\eta\downarrow 0} E w_0(\eta) = 0$.

Let $\nu_n(\eta) = \sup_{|c-c'|<\gamma} |\upsilon_n(c) - \upsilon_n(c')|$ where

$$\upsilon_n(c) = \frac{1}{n}\sum_{k=1}^{n}\left[\log g(c, X_{\frac{k}{n}}) - \log g\left(\theta_0\left(\frac{k}{n}\right), X_{\frac{k}{n}}\right)\right]$$

$\nu_n(\eta) \leq w_0(\eta)$, so $\lim_{\eta\downarrow 0} E\nu_n(\eta) = 0$.
Then there exists a sequence (ε_k) which decreases to 0, so that for all k :

$$\lim_{n\to\infty} P\left(\nu_n\left(\frac{1}{k}\right) \geq \epsilon_k\right) = 0 \tag{6.7}$$

Let now $\varepsilon > 0$ given and the set :

$$B_\epsilon = \{c, K(\bar{g}, c) \leq \inf_{c' \in F} K(\bar{g}, c') + 2\epsilon\} \tag{6.8}$$

Take k so that $\varepsilon_k < \varepsilon$ and let $C \setminus B_\epsilon$ covered by a finite number $N(p)$ balls with center c_i and radius $1/k$ (in $\mathbb{R}^p$), $1 \leq i \leq N(p)$. We now study the behaviour of the maximum likelihood in C. Consider $c \in C \setminus B_\epsilon$. We have :

$$\upsilon_n(c) \geq \upsilon_n(ci) - |\upsilon_n(c) - \upsilon_n(c_i)|$$

So

$$\inf_{c \in C | B_\epsilon} \upsilon_n(c) \geq \inf_{1 \leq i \leq N(p)} \upsilon_n(c_i) - \nu_n(1/k) \tag{6.9}$$

On the other hand, $\hat{c}_n = \underset{c \in C}{\operatorname{argmax}} \sum_{k=1}^n \log g(X_{t_k}, c) = \underset{c \in C}{\operatorname{argmin}} \upsilon_n(c)$.

$$\Rightarrow (\hat{c}_n \notin B_\epsilon) \subset \{\inf_{c \in C \setminus B_\epsilon} \upsilon_n(c) < \inf_{c \in F} \upsilon_n(c)\} \tag{6.10}$$

Inequations (6.9) and (6.10) bring to :

$$P(\hat{c}_n \notin B_\epsilon) \leq P\left(\inf_{1 \leq i \leq N(p)} \upsilon_n(c) - \inf_{c \in F} \upsilon_n(c) \leq \nu_n(1/k)\right)$$

So

$$P(\widehat{\theta_{K,n}} \notin B_\epsilon) \leq P(\nu_n(1/k) > \varepsilon) + P(\left(\inf_{1 \leq i \leq N(p)} \upsilon_n(c) - \inf_{c \in F} \upsilon_n(c) \leq \varepsilon\right) \tag{6.11}$$

But $\upsilon_n(c_i) - \upsilon_n(c) \to K(\bar{g}, c_i) - K(\bar{g}, c)$, $c \in F$ (according to lemma 6.3.1).

$K(\bar{g}, c_i) > m + 2\epsilon$ where $m = \inf_{c' \in C} K(\bar{g}, c')$, then $\lim_{n \to \infty} (\inf_{1 \leq i \leq N(p)} \upsilon_n(c_i) - \inf_{c \in F} \upsilon_n(c)) \geq 2\epsilon$ and $\lim_{n \to \infty} P(\nu(1/k) > \epsilon) = 0$

So $P(\hat{c}_n \notin B_\epsilon) \to 0$ .Then by the definition of the infimum, it exists a sub-sequence of $\hat{c}_n$ which converges to a value in F.

So we have $F^1 = \{\lim_{n \to \infty} \hat{c}_{n_k}, \hat{c}_{n_k}$ *convergent sub-sequence of* $\hat{c}_n\} \subset F$.

The lemma 6.3.2 is proven.

□

Theorem 6.3.1 *For every a such that* $a \leq \min(\| \theta_0 - c \|, c \in F)$ *then a.s.* $\| \hat{\theta}_\lambda^n - \hat{c}_n \| > \frac{a}{2}$ *for n large enough.*

Proof

We know that with appropriate conditions, $\hat{\theta}_\lambda^n$ is such that $\| \hat{\theta}_\lambda^n - \theta_0 \| \to 0$ when $n \to \infty$ and $\lambda \to 0$ (see Section 3.2 of Chapter 3).

We have : $\| \hat{\theta}_\lambda^n - \hat{c}_n \| \geq \| \theta_0 - \hat{c}_n \| - \| \hat{\theta}_\lambda^n - \theta_0 \|$

In fact, as mentioned, $\| \hat{\theta}_\lambda^n - \theta_0 \| \to 0$ and using Lemma 6.3.2, we have :

$$\| \theta_0 - \hat{c}_n \| \to \inf_{c \in F^1} \| \theta_0 - c \| \geq \inf_{c \in F} \| \theta_0 - c \| \tag{6.12}$$

Because of the non-constancy of θ_0, $\inf_{c \in F} \| \theta_0 - c \| > a > 0$.

From that, $\| \hat{\theta}_\lambda^n - \hat{c}_n \| > \frac{a}{2}$ for n large enough. Theorem 6.3.1 is thus proved.

□

This theorem can be applied in various directions. For testing H_0', we have to take $\hat{c}_n^1$ and $\hat{\theta}_\lambda^{2,n}$ instead of $\hat{c}_n$, with $\hat{c}_n^1$ a constant estimator for the parameter we want to test, and $\hat{\theta}_\lambda^n = (\hat{\theta}_\lambda^{1,n}, \hat{\theta}_\lambda^{2,n})$. The proof for this case is almost similar to that of Theorem 6.3.1 and we do not detail it here.

We now have to extend it to the case of GEV densities g whose support is of the form $(-\infty, b(g))$ where $b(g) < \infty$, the support is bounded on the right side when the shape parameter ξ is strictly negative.

Now we need to start with a compact set C defined as $C \subset \mathbb{R}^3$ and $\left\{(\mu, \sigma, \xi),\ -M \leq \mu \leq M;\ 0 < \varepsilon \leq \sigma \leq M, -\frac{1}{2} + \varepsilon \leq \xi \leq -\varepsilon\right\}$ where ε and M are arbitrary constants. We choose this compact set because it will be the usefull one for the studied application. The condition $\xi > -\frac{1}{2}$ is not necessary but is usual for giving the existence of the m.l.e. in the regular case. In fact $\xi > -1$ is necessary.

With these conditions, it is not difficult to check conditions A1, B1, B2. Of course A1 and B3 are not satisfied. We need to find out un compact $C_1 \subset C$ where A1 and B3 can be valid. This means that $g(x, c)$ exists $(0 < g(x, c) < \infty)$ for any $cinC_1$.

In order to extend the previous theorem to the case of GEV distributions, we need that for every $c \in C$, $K(\bar{g}, g_c) < \infty$. This leads to prove the following lemma.

Lemma 6.3.3 *For n large enough,*

$$P\{\text{argmax} \sum_{k=1}^{n} g(c, M_{k/n} \in C_1\} = 1$$

where $C_1 = \{c, b(c) >= \max_{t \in (0,1)} b(\theta_0(t))$ with $b(c)$ the upper bound of $g(c, x)$ and $b(\theta(t))$ that of $g(\theta(t), x)$.

If the lemma is true, then we have to consider only c such that $g(c)$ is absolutely continuous with respect to all $g(\theta(t))$, $t \in (0,1)$ and also with respect to $\bar{g}$.

From the form of the GEV density (ξ fixed), we check easily that $k(\bar{g}, g_c) < \infty$ if $c \in C_1$. The proof of Lemma 6.3.3 is shown here.

Proof

Suppose that there exists some t_c such that $b(c) < b(\theta_0(t_c))$.

Then from the continuity of the map $t \longrightarrow b(\theta(t))$, this is still true for $t \in I_c$ with I_c an open interval and $t_c \in I_c$. Thus there exists an open interval $J_c \subset I_c$ and $\alpha > 0$ such that

$$b(c) < b(\theta(t)) - \alpha \text{ for } t \in J_c$$

The density of $G_\xi(\mu, \sigma)$ near the upper bound $b = \mu - \sigma/\xi$ is equivalent to $D(\sigma, \xi)(b - x)^{-1/\xi}$ for some constant $D > 0$ depending on σ and ξ.

So there exist $D_1, D_2 > 0$ such that

$$\sum_{k=1}^{n} P(Y_{k/n}) > b(\theta_0\left(\frac{k}{n}\right) - \alpha) \geq D_1 \alpha^{-1/\xi}$$

and

$$\sum_{k=1}^{n} P(Y_k) > b(\theta_0\left(\frac{k}{n}\right) - \alpha) \geq D_2 n$$

From the dependence of the variables $M_{k/n}$ and the Borel Cantelli lemma, there exists n_0 such that for every subsequence $n_k > n_0$

$$P\left(\sum_{k=1}^{n_k} \log g(c, Y_{k/n_k}^{n_k})\right) = -\infty) = 1$$

and thus if $\hat{c}_n = \text{argmax} \sum_{k=1:n} \log(c, Y_k^n)$, $P(\hat{c}_n \in C_1)$ for $n > n_0$. Then the lemma is proved.

□

6.3.2 Numerical study of the test

In order to make the previous test, there are different things to consider.

We see that the statistical test is based on some distance $\| \hat{\theta}_\lambda^n - \hat{c}_n \|$. If the distance is a norm then

$$\| \hat{\theta}_\lambda^n - \hat{c}_n \| \geq \| \theta_0 - \hat{c}_n \| - \| \hat{\theta}_\lambda^n - \theta_0 \| \tag{6.13}$$

and

$$\| \hat{\theta}_\lambda^n - \hat{c}_n \| \leq \| \theta_0 - \hat{c}_n \| + \| \hat{\theta}_\lambda^n - \theta_0 \| \tag{6.14}$$

The statistical error $\| \hat{\theta}^n_\lambda - \theta_0 \|$ is always $O(\frac{1}{n^\alpha})$ where α depends on the hypothesis made on Θ(see Theorem 3.4.1 in Section 3.2).

When K is true, the term $\| \theta_0 - \hat{c}_n \|$ is $O(\frac{1}{\sqrt{n}})$. So in the case where $\alpha < \frac{1}{2}$, following (6.14), the distance $\| \hat{\theta}^n_\lambda - \hat{c}_n \|$ is mainly the statistical error in nonparametric estimation $\| \hat{\theta}^n_\lambda - \theta_0 \|$.

When K is false, the main error in the right term of (6.13) is $\| \theta_0 - \hat{c}_n \| > a > 0$. Then the distance $\| \hat{\theta}^n_\lambda - \hat{c}_n \|$ is significantly different from zero.
It is preferable not to use asymptotic distribution which highly depends on properties of Θ and to proceed using bootstrap and simulation if it is possible.

The formula $M_k = \mu_k + \sigma_k Z_{\xi_k}$ allows to bootstrap the estimates of the parameters if and only if ξ is constant. In this case we have a model with i.i.d residuals and the classical bootstrap can be used to create the empirical distribution for the distance $\Delta = \| \hat{\theta}^n_\lambda - \hat{c}_n \|$. Simulation technique can be also used.

If $\xi(t)$ is a function of time, we have to use simulations in order to compute the distribution of Δ.

6.4 Reconstruction and departures from stationarity in the tails of Y

The previous results can be illustrated on the original data X_t. Using the relations (6.2) written in the previous paragraph, we can recover the GEV model $(ext_X) = G(\mu_X, \sigma_X, \xi_X)$ of X_t from the knowledge we can get from the GEV model $(ext_Y) = G(\mu_Y, \sigma_Y, \xi_Y)$ linked to Y_t and that of m_X and s_X.

Here we have three estimates of the parameters of ext_X : the direct non parametric estimate $\hat{g}$, and the indirect ones using Y : $\hat{f}$ when the parameter in the model ext_Y is stationary and $\tilde{f}$ without this constraint. $\tilde{f}$ and $\hat{g}$ are always convergent estimators of f . For $\hat{f}$ if K is satisfied, the error in estimation is, for a size of observations large enough, smaller than the other errors obtained for $\hat{g}$ and $\tilde{f}$. On the contrary, as already said if K is false, $\hat{f}$ estimates f with a bias of $\| \hat{f} - f \| > \alpha > 0$.

In summary, we can compare three estimates of ext_X, the direct and usual one and two indirect estimates obtained from m, s and ext_Y, constrained or not. We proceed to the following sequence of estimations.

- Direct estimation of μ_X, σ_X, ξ_X of the GEV
- Estimation of m, s and computation of Y_t

• Estimation of μ_Y, σ_Y, ξ_Y of the GEV for both cases : with and without K.

6.5 Results of application to data series of Europe

We work here on 55 data series from the ECA & D project (European Climate Assessment and Dataset project, (Klein et al., 2002)). On the other hand, some applications will be also made for ERA40 reanalysis data. The summer has been defined as the 100 days between the 14th of June and the 21st of September, whereas the winter has been defined as the 90 days between the 1st of December and the 28th of February. These periods have been selected because most extreme events occur between these dates. Daily maximum and minimum temperature have been considered in the summer and winter respectively.

In the summer, there are 100 observations for each year, so we use a block size of 25 days for the method of block maxima (GEV) and in the cold season, there are 90 observations for each year, then we use a block size of 30 days.

The constancy of ξ is studied in Chapter 5. In this chapter, we see that this constancy can also be tested by the previous test when we estimate ξ as a constant and as a function of time at the same time. This test however is not executed here for its heavy computation (in the case where ξ is functional). After the study in Chapter 5, we accept the constancy of ξ. Then we place ourselves here in the semiparametric case when estimating GEV models for temperature series : μ, σ are estimated nonparametrically and ξ is a constant.

6.5.1 Links between trends in central field and in extremes

Our aim is to graphically compare m with m_{extX}, and s with s_{extX} in a systematic way. m_{extX} and s_{extX} are estimated in using (6.2) with the estimators of extreme parameters obtained by algorithm 2 above, while m_X and s_X are estimated locally by loess (see Parey et al., [112] for details).

We first estimate these quantities from the datasets of five stations from different places in Europe. Both seasons are considered. The results are shown in figure 6.1.

The first remark is that for several series, the trends in mean and the trends in variance of the central field and extreme events are very similar. Intuitively the relation between m and m_X seems stronger than that between s and s_X.

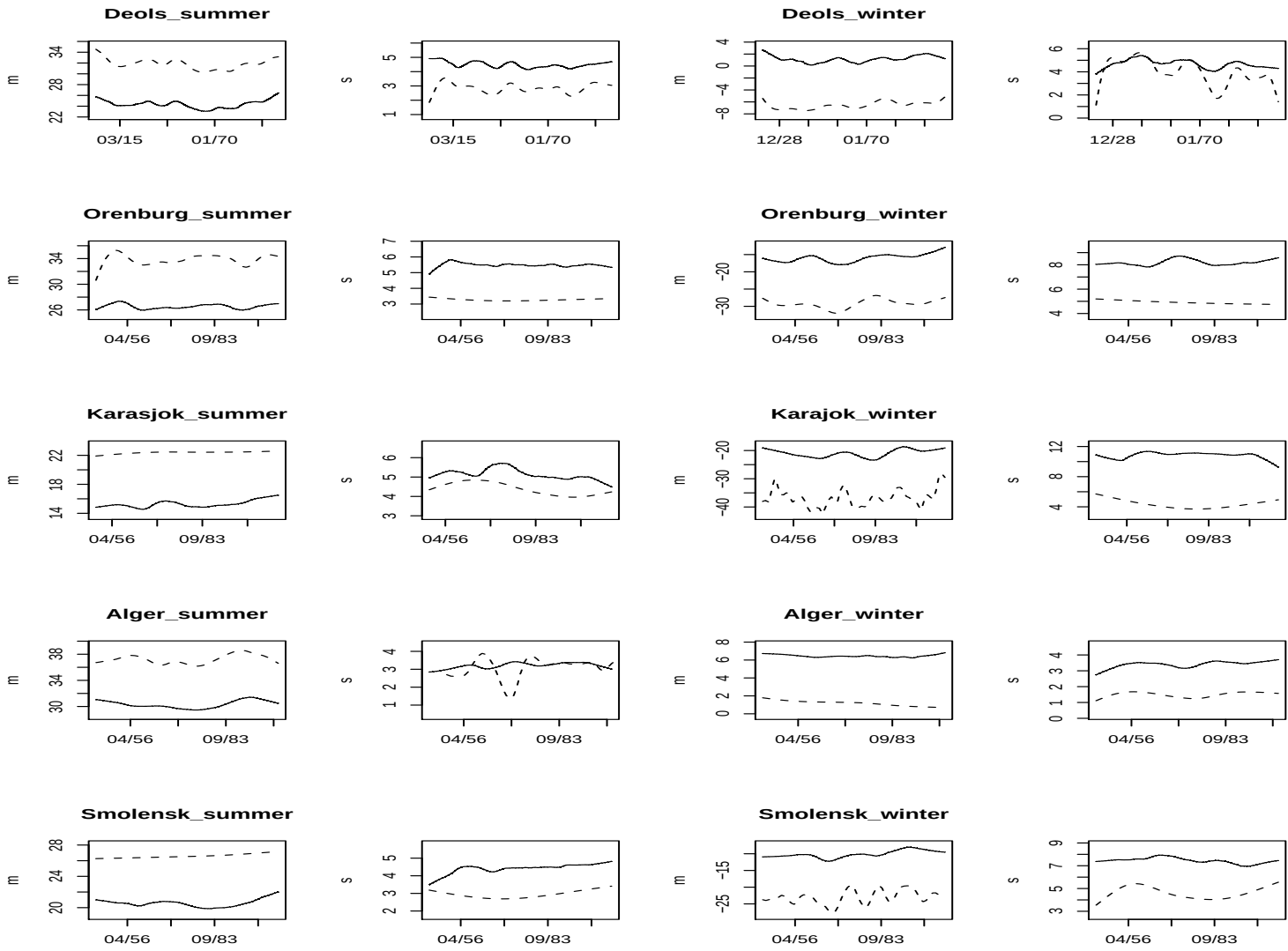


FIGURE 6.1 – Mean function, scale function of the whole observations (solid line) and mean function, scale function of extreme events (dashed line) for different seasons at the different stations.

Now we will study the K hypothesis for all the European ECA &D temperature series. The test is realized at the level of 90 percent, which means that if the true values of Δ are found inferior to the 90th percentile of the simulated Δ or in other words, if the p-value =P(D>true value) where D is the variable of the distribution of delta is superior to 0.1 for both μ and σ , K cannot be rejected. In the case where K is rejected by our test, by constructing the distribution of Δ with the nonparametric estimators (which

are supposed not to satisfy K), the power of the test can be calculated. These powers are high enough (more than 80%), which shows the credibility of our test. One example is for the summer in Ile-de-groix where K is refused with respect to μ (p-value = 0) and σ(p-value=0.05) and the power of test for μ is 96.7%, for σ is 91.5%.

In this sense, we have the results for the ECA & D temperature series (figure 6.2).

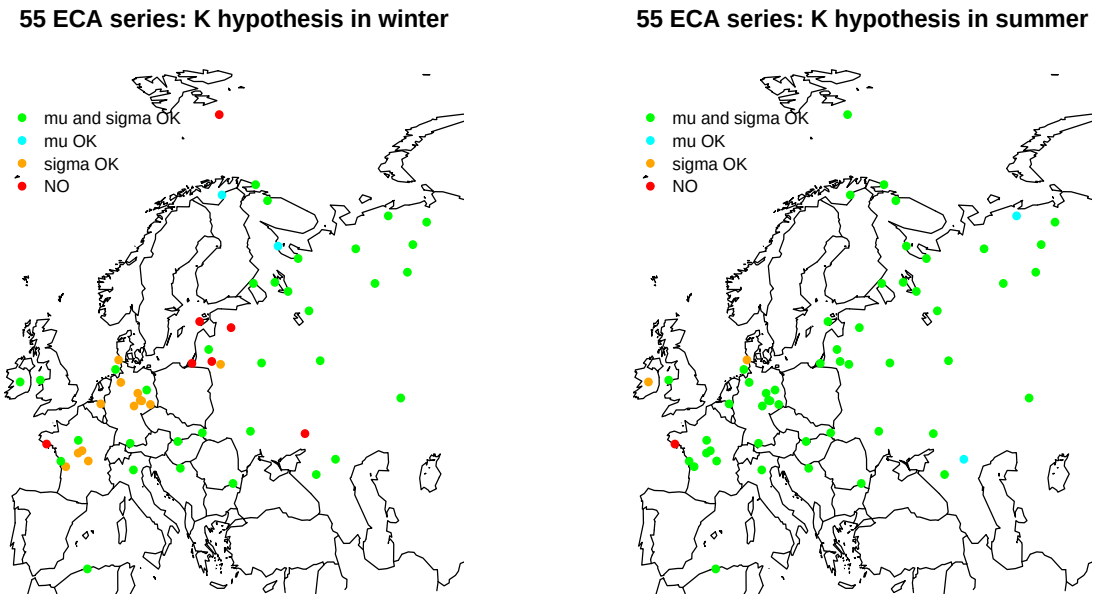


FIGURE 6.2 – Validation of K on ECA & D data series

The K hypothesis is generally rather well accepted in summer, for both μ and σ, whereas in winter it is less obvious. This is coherent with the fact that the link between mean and variance of the central field has been found stronger in summer than in winter (Parey et al., [112]). There seems then to be a different behavior in the two seasons concerning the evolutions of the mean and variance and that of the extremes. One possible explanation could be that in summer, the extremes and the variance are linked to the occurrence of heat waves in a stronger way than in winter with cold waves. In winter, the variability is more influenced by the large scale circulation, through the weather regimes for example. This hypothesis will need to be

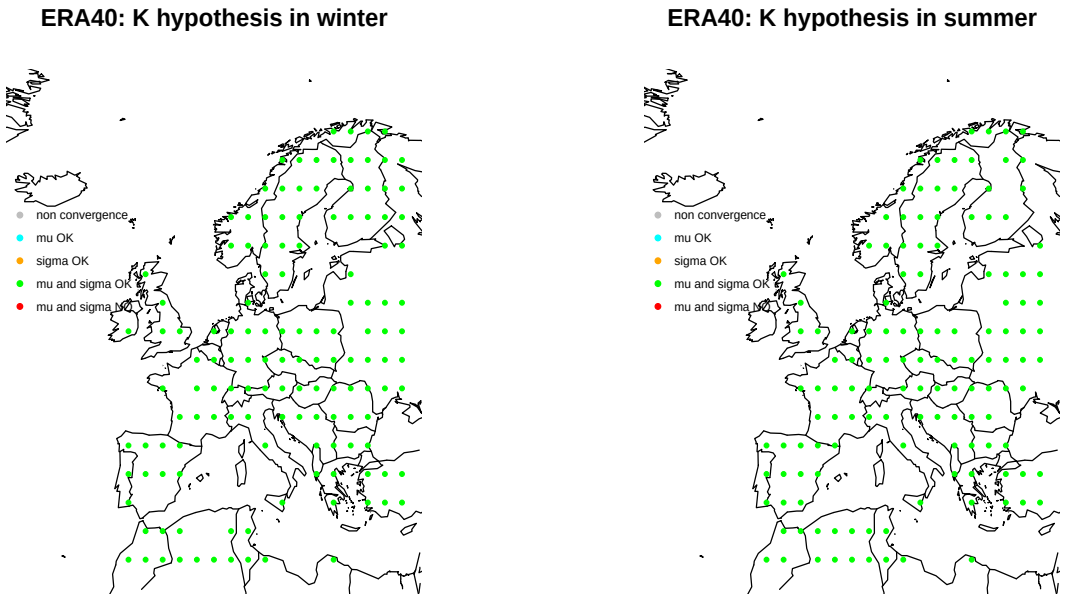


FIGURE 6.3 – Validation of K on ERA reanalysis data series

further investigated.

The study of K can be extended to the reanalysis data. The results of the validation of K on the daily maximum temperature in summer and daily minimum temperature in winter are shown in figure 6.3. For reanalysis datasets, K is always valid.

The K hypothesis, if satisfied, can provide a good way to evaluate the extreme parameters of X from those of Y and the mean and variance evolutions of X. Let us give one example in the case of the summer temperature in Deols, where K is accepted in Figure 6.4. The reconstructions of μ with K and without K are quite good ; both of them approach the direct estimator. The reconstructions of σ are less good ; the reconstruction without K cannot follow the other ones where the peaks are found. Once more, this is due to the weak robustness of the estimators of the scale function.

The fact that the K hypothesis can generally be accepted in summer could have an interesting application. As climate models currently have difficulties to correctly reproduce the parameters of the tail of the temperature distribution ([111]), this result could be used to estimate the GEV parameters for a future period from those of the current one (evaluated from the observed

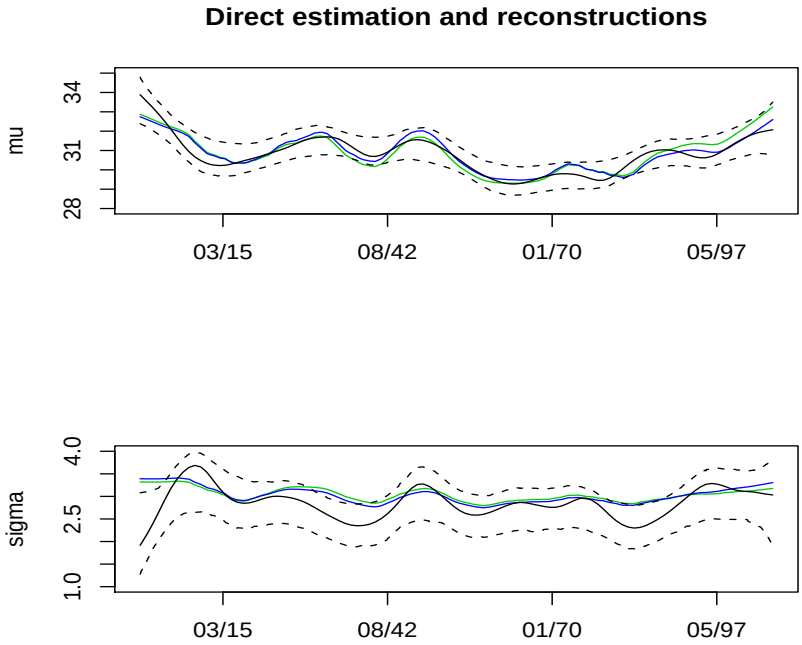


FIGURE 6.4 – Direct estimations (in black), their 90% bootstrap confidence intervals (dashed lines) and reconstructed parameters under K (in green) and without H(in blue) for the GEV distribution fitted on the maxima of the temperature series in summer in Deols.

series) and from the mean and variance change of the central field as projected by the climate models for the desired future period.

6.6 Conclusion

In this paper we have tried to use, as far as possible, the properties of the non parametric methods in statistics to obtain general qualitative properties on the time evolution of the temperatures in many stations in Europe during periods of the last century. One has to understand that the conclusion depends on the amount of data. Some subtle physical behaviors cannot receive a statistical evidence with the observations provided by meteorological measures so far and thus sentences as "the effect of warming is stronger on the extremes than on the mean" do not have any strong statistical support in large parts of Europe even if the very recent climate behavior could suggest the opposite.

The first conclusion is that the mean and variance of the central field and the mean and variance of the extremes for both summer and winter have a very similar evolution. These relations inherit one part of the relation between the mean and variance which is described in Parey et al.,([112]). This analysis could be extended to other stations in other areas of the world in order to see if it is a general behavior.

In order to evaluate to what extent the evolution of the extremes is due to the evolutions of the mean and variance of the central field, we proposed a methodology to test the so-called K hypothesis : the null hypothesis is the stationarity of the model against the non stationarity, parameters being smooth functions of time. As a particular case we studied the stationarity of the extremes of the reduced series Y_t obtained from the initial X_t series. This hypothesis can generally be accepted for a majority of the studied European temperature series in summer. However, if the geographical repartition of the satisfaction of the K hypothesis is coherent with that of the link between evolutions in mean and variance in the central field in summer, it is less obvious in winter. In summer, for stations experiencing more variable conditions, the variance increases when the mean increases, and these increases explain a large part of the evolution of the extremes. In winter, the picture is more complicated and further analyses would be needed in order to better understand this behavior.

For the stations and season where the general validity of the K hypothesis is confirmed, this could be used to estimate future return levels from the General Circulation Models projections. As a matter of fact, these models

currently perform better in simulating the average behaviour of the temperature than their extremes. However, the evolution of the variance has been shown to have an impact on the evolution of both the mean and the variance of extremes. Thus, a correct reproduction of the variance is also an important factor in order to have a reliable estimation of the parameters of extremes for a future period in using the K hypothesis. Thus, this point should be emphasized in the development and validation of the climate models.

Chapitre 7

Multidimensional trends: the example of temperature

This chapter is our article [79]. It can be considered as the synthesis of the results from the previous chapters. It gives us a general look for the trends in the central field and in extremes, so-called "multidimensional trend" in the climate change context. The definition for the trends in mean and in extremes in the literature have not been defined clearly yet. In this paper, we will discuss about the notion of multidimensional trend. For us, trend's notion is linked with nonparametric estimation. "Trend" classically is understood as a straight line, so increasing trend or decreasing trend. This comprehension is really limited. Whereas the flexibility of nonparametric estimation allows us to give a much larger definition for trend without any assumption on its form.

Non parametric trends in mean, variance and extremes will be derived for the temperature series in Europe. Besides that, a criterion to compare the different trends will be proposed. The relation between trends in the central field and themselves with trends is extremes will be studied to give us more comprehension in the global environment of the temperature.

7.1 Introduction

Climate change is generally presented and discussed in term of trends in the mean of climate variables over different spatial scales, from individual observation stations to spatial means over geographical areas or the entire globe. The study of temperature variability, confidence intervals and trends in extremes are also worth of interest. Generally, trends in mean are derived using ordinary least square regression methods [84, 34]. On the other hand, trends are usually studied separately : papers are devoted either to trends in mean, or to trends in extreme events. The trends in extreme events are analyzed using linear least squares fit on the series of so called "extreme indices" [89, 5]. Mudelsee[96] uses kernel fitting to study flood risk in a nonparametric way.

In this paper, we would like to address the non parametric derivation of trends, as well as the links between trends in different quantities such as mean, variance and extremes, and then illustrate the point by considering temperature series in Europe [42].

Generally, a trend is computed (not defined) as a slow and thus regular component of a time series, superimposed on a series of quite stationary and less variable residuals. In other words, computing a trend consists in extracting some deterministic signal from noisy data. In most studies, as stated before, this is done using ordinary least squares fitting, which will be referred to as "classical" trends.

The point and main topic considered here can be formulated as follows : once the trend in mean has been estimated, are there other trends which could describe the temporal evolution of the series and how many are si-

gnificant ? Are they "independent" in some sense ? If we leave out of this study, for simplicity, the usual seasonality and all periodic phenomena, our goal is then to look for multidimensional deterministic trends, which, when removed, leave an as stationary as possible residual series. In practice, as we will illustrate on temperature series, the search for stationarity has to be stopped when the physical meaning of the trends becomes too weak or when the trends are too strongly linked.

As our goal is to derive non parametric trends in the most objective and general possible way, we first mention the general qualitative properties, requirements and assumptions made to identify a trend. They can be listed as follows :

- **Choice of the time scale** : When observations are given over a time period T one has to decide which time scale (less than T) is of interest to study the signal variation ; to simplify we will call it "the window length" in this work. For example, if the topic of interest is the possible anthropogenic effect on climate change, if one has to consider data over the industrial period, the window length is more or less arbitrarily chosen as 30 years if two to three centuries of observations are available. Over the most recent period, the decade is often chosen as window length if the period length of observations is at most fifty years. The window length for a given observation period is indeed often chosen in a quite arbitrary way and this choice is very important for instance in the case of the increasing mean temperature. We will show in the paper that such a choice based on informal or subjective ideas (heuristics) can be improved when using more intrinsic (mathematical) considerations.

- **Almost invariance by data extension and localization** : Once the length T of the observations has been fixed, if we add new observations whose period length is a significant fraction of T ($T/2$ for example), then the modifications on the previously computed trend on any subinterval of the initial dataset have to be small. Of course, if we consider a trend estimated on a 50-year period of observations, and then include this period in a larger period of say 300 years, the trend is obviously modified and the intrinsic smoothness will change. Nevertheless, intrinsic choices allow to give an objective interpretation of this modification and can help to define changes or breaks in the trend behaviour.

- **Monotonicity** : In order to easily interpret the physical phenomena, the monotonicity can be imposed to a trend. For instance if trends are used in order to define an extrapolation necessary to compute return levels, monotonicity is required in general. Variants are in constraints on the number of intervals where the trend is monotone or convex. There are general tools to obtain constrained estimates as isotonic regression (see [151] for example).

Starting from these considerations, the paper is devoted to the identification of trends in mean, variance and extremes of a time series, together with their possible links. It is organized as follows : first, the statistical frame-

work is given in section 7.2 before coming to the results for trends in mean, variance and extremes of temperature series in Europe in section 7.3. The link between the trends in mean and variance and the trends in extremes is discussed in section 7.4 and a methodology is proposed to test if the trends in extremes are due to trends in mean and variance, before concluding in section 7.5.

7.2 Statistical framework

7.2.1 Trend in mean

We first discuss the trend in mean, with the supposition that there is no seasonal trend (or any trend with an obvious shorter window, this point can be checked for example using a wavelet analysis).

The basic approach uses a moving average procedure, with a moving window length L. In this simplest case as for the more complicated forthcoming, we define the model, here :

$$X(t) = m(t) + e(t) \tag{7.1}$$

where $X(t)$ is the observation series with unknown mean $m(t)$ and $e(t)$ is a centred process, which is expected in general to be stationary and uncorrelated.

This "signal+noise" approach is first improved in using non parametric statistics, and then often followed by the choice of a regression model in parametric statistics.

Non parametric methods satisfy the previously given requirements, but now the window length of interest is chosen in an intrinsic way and has a precise definition. The main lines are the following :

-We suppose that $m(t)$ is in a space M, for instance that of the two continuously differentiable functions.

-We suppose that $e(t)$ is stationary. An important point must not be forgotten : the statistical tools used to identify $m(t)$ depends on the properties of $e(t)$, for instance of its correlation or of its distribution.

-A criterion is chosen to define a "good estimate" of $m(t)$, for instance the minimization of the integrated quadratic error.

$$Err = \int_{t \in D} [m(t) - \hat{m}(t)]^2 \tag{7.2}$$

which requires the control of the bias and of the variance of the estimator (it exists of course the same but local criteria for fixed t).

There are two families of methods : smoothers (splines, kernels, loess, lasso) and approximations (Fourier, orthonormal systems as Fourier series, wavelets families). See [151, 47, 59, 131, 139] for a general presentation of these basic statistical tools.

In all cases, the **intrinsic** window length is associated to a **tuning parameter** of the method (penalization parameter or kernel width in the first case, threshold for coefficients in the second one). This parameter is estimated from the data (adaptive estimation), normally with *independent data*, using methods such as the cross validation (CV) or the generalized cross validation (GCV) or other criteria as AIC (see [93, 75, 103]). Besides we have another parameter, the **degrees of freedom**, calculated by the trace of the smoother matrix $df = tr(S)$ (see [75]) which is directly linked to the tuning parameter. The more df is large, the less the function is smooth. The degree of freedom is the only common parameter which allows to compare different smoothing methods such as loess and splines. It fixes the window length which becomes intrinsic because the degree of freedom only depends on the data without any a priori choice [151]. Similar notions can be inferred for approximation methods. Details are given in [28] about the technical use of CV and so the mathematical choice of the degrees of smoothness.

The methods are developed and deeply studied in the case $e(t)$ is a sequence of independent and equidistributed variables. In the case of *dependent data*, they can be used with suitable and precise specifications and modifications depending on the hypothesis made on the process $e(t)$: for instance $e(t)$ has a (short) memory (autocorrelation) or the variance of $e(t)$ is not constant. Examples are given for temperatures in our previous work [112]. It remains to obtain more general results in mathematical statistics to improve the intrinsic character of the tools.

Confidence bands are obtained by bootstrap using also modifications (very close to that necessary for the choice of the tuning parameter) to take into account the properties of $e(t)$ [66, 91, 104].

7.2.2 Trend in variance

To derive the trend in variance, we first choose a model, here :

$$X(t) = m(t) + s(t)\varepsilon(t). \tag{7.3}$$

We shall use sequential "plug in" methods : once the trend $m(t)$ for the mean has been estimated (as $\hat{m}(t)$), it is then natural to look for a trend for the variance $s^2(t)$ of the same observations with the same window length of interest.

The model is now :

$$Y(t) = X(t) - \hat{m}(t) = s(t)\varepsilon(t) \tag{7.4}$$

where $\varepsilon(t)$ is a centred process with variance one, once again expected to be stationary and uncorrelated. The purpose then is to estimate $s(t)$.

For a large class of hypotheses on $\varepsilon(t)$ (like for $m(t)$), we can proceed as for $m(t)$ but on the data $Y^2(t)$. It is worth noticing here that the fact

to plug $\hat{m}(t)$ instead of $m(t)$ in the data in order to obtain $Y^2(t)$ does not modify, at least for large datasets, the efficiency of the usual non parametric methods [96]. Of course, here as for $m(t)$, statistical results are needed to extend to very general $\varepsilon(t)$ the use of the tools.

Periodic seasonality and global warming are not independent, global warming strongly depends on the season (even on the month) of the year(see [78]). Technically, this is a difficult point .What happens for seasonal trends if these trends in mean and variance exist ? We do not discuss this problem here, as it is neither a problem of usual separation of frequencies nor a problem of space time decomposition usually solved using wavelets for instance. Here the main problems are that of the identification of parameters because seasonality is not invariant for the time scale defined by the mean and variance trends, and the mean and variance trends are highly seasonal. In a parametric context it is a simple problem but it needs tricky solutions in a non parametric context. We shall discuss it in a forthcoming paper. $\varepsilon(t)$, as defined above, is not a stationary process because in general, and it is the case for temperatures, its dynamics depends on the season. Of course all the difficulties previously mentioned disappear, or at least are much attenuated, if one limits the study to a homogeneous part of the year, for instance the 2 hottest months.

7.2.3 Trend in extremes

It is suspected that if a warming effect is observed for temperature, more precisely increasing trends in mean and/or variance, then some linked phenomena has to be expected for the extremes. For the extremes, the statistical procedures are based on an asymptotic result coming from the probability theory [135], which requires a large amount of data (the convergence is slow) and some regularity conditions on the tails of distributions, in general quite impossible to check but accepted for their weakness. Two other conditions have to be verified in order to have a limit theory for the extremes :

- does the temporal structure of the process $\varepsilon(t)$ adjust to a theory of extremes for this kind of process ?

- what is determinant for the extremes (what "makes" the extremes ?) : the deterministic trend $m(t)$ or the random fluctuations of $\varepsilon(t)$?

In our applications, the effect of the deterministic trend $m(t)$ during a short period (some months) can be neglected, with a statistical evidence, with respect to the random fluctuation $\varepsilon(t)$, whose distribution has parameters which are the extreme trends. Now what happens when we consider a larger time scale ?

We will consider here the classical block maxima approach [135]. The time period of observations is divided into n consecutive blocks of days of same length. Then, we select into each block of time the maximum value of the observations $X(t)$ obtaining a sequence $M_1, M_2, , M_k, , M_n$ of the

maxima. We suppose that the independence or at least a weak dependence of the values is insured if the length of the blocks is sufficiently large. In this case, the probability theory allows to suppose that there exist sequences μ_k and $\sigma_k > 0$ such that the distribution of M_k is approximated by one of $\mu_k + \sigma_k Z_{\xi_k}$ where Z_{ξ_k} is a sequence of independent variables with a distribution G_{ξ_k}.

$$\begin{array}{l} \text{if } \xi_k \neq 0, G_{\xi_k}(z) = \exp -(1+\xi_k z)^{-1/\xi} \text{ when } 1+\xi_k z > 0\,; =1 \text{ if not} \\ \text{if } \xi_k = 0, G_{\xi_k}(z) = \exp\left(-\exp\left(-z\right)\right). \end{array} \tag{7.5}$$

Then the discrete model of extremes is : $M_k = \mu_k + \sigma_k Z_{\xi_k}$ and the trends for the extremes are μ_k and σ_k (k being the block index) which will be extended into continuous functions of time t.

The expectation and variance of the distribution are expressed as :

$$E(Z) = \frac{1}{\xi_k}(\Gamma(1-\xi_k) - 1); V(Z) = \frac{1}{\xi_k^2}(\Gamma(1-2\xi_k) - \Gamma^2(1-\xi_k)) \tag{7.6}$$

So if M_k is the maximum in a block k with distribution G_{ξ_k}, we have :

$$E(M_k) = \mu_k + \sigma_k E(Z); V(M_k) = \sigma_k^2 V(Z) \tag{7.7}$$

where μ_k, σ_k and ξ_k are functions of the block index k and respectively location parameter, scale parameter and shape parameter of G_{ξ_k}. The location parameter of the GEV models is not a "natural" parameter, it mixes mean and variance. Thus we often consider in the following the mean and the variance of extremes, which are more easily interpretable.

We first supposed that ξ does not depend on time (see the discussion on this hypothesis in [22]). μ_k and $\sigma_k > 0$ depending on the block index k are then the only trends. To be compared with the trends in mean and variance, they must be defined for every time t. Thus, we now have to estimate the trends $\mu_X(t)$ and $\sigma_X(t)$ (subscript X referring to the observation series $X(t)$) and also ξ_X in using a non parametric method based in any case on the likelihood, and not on least squares like for the mean trend (see [105, 18]). We use cross validations to choose the smoothing parameters for μ_k and σ_k. These problems of optimization and choice of the tuning parameters are quite difficult because of the correlation between the parameters (see [18]).

7.2.4 Measure of the similarity of trends

The different trends considered here, for instance trends in the temperature mean and variance, show some similarities and are supposed to be both affected by climate change. Thus, for physical reasons, they have to be considered on the same time scale (or with the same window length).

Furthermore, the "dependence" between the different trends can be measured, as for random series, by a formal correlation defined as the scalar product of their reduced form, respectively called nf and ng, centred around their mean $\overline{f} = \int_{t\in D} f(t)dt$ and normalized by $\sqrt{\int_{t\in D} (f(t)-\overline{f})^2}$:

$$r(f,g) = \int_{t\in D} \langle nf(t), ng(t)\rangle dt \tag{7.8}$$

D is the ensemble of the dates in the total time period T.

This correlation is a measure of the linear dependence of f and g and of the quality of the approximation of $g(t)$ by a function of the form $a+bf(t)$. To analyze more accurately the links between trends, we can also calculate the correlation coefficient between the first derivatives of f and g which is a measure of the quality of the approximation once the best linear trend has been removed. Another indicator of the similarity of trends is the number and the distances between the locations of local extremes of the curves. Other quantities can be used in order to measure the likeliness between functions f and g including non linear transformations, for instance one can estimate θ such that the correlation between f and $\theta(g)$ is maximal.

7.3 Results for the European temperature series

7.3.1 Mean and variance trends

First we computed the trends in mean and variance of daily minimum temperature in winter and daily maximum temperature in summer for 2 observation series in France : Paris-Montsouris over the 1873-2003 period (131 years) and Strasbourg over the 1949-2005 period (57 years). This avoided to deal with the seasonal effect, as previously discussed in section 7.2.2. The summer is defined as the 100 days between the 14th of June and the 21st of September and the winter as the 90 days between the 1st of December and the 28th of February. These periods have been chosen because the extremes mainly occur during these times. In our applications to temperature series in Europe [112, 110], we found degrees of freedom corresponding to a local window length of 20% of the total length, it means 11 to 15 years depending on the station. Therefore, the trends are computed here with loess and a smoothing parameter corresponding to an 11-year window length for Strasbourg and a 26-year window length for Paris-Montsouris. The steps are then the following :

- selection of the days of the studied season over the total period length (this leads to a series of nyear100 days for summer and (nyear-1)90 days for winter, nyear being the number of years in the total period length)
- computation of $\hat{m}(t)$ from $X(t)$ using loess

- computation of $[X(t) - \hat{m}(t)]^2$
- computation of $\hat{s}^2(t)$ from $[X(t) - \hat{m}(t)]^2$ using loess.

where the hat notation corresponds to estimates.

When computing the trends in mean and variance, it came out that their evolutions seemed linked : variance increasing when mean increases in summer and when mean decreases in winter. In order to verify this visual link, the standard deviation evolution is super-imposed with that obtained from a linear regression regarding the mean evolution. The results are shown in figure 7.1 (the length of the plotting area is proportional to the period length of each series). Although the linear reconstruction is far from perfect, it globally shows a similar evolution. In summer, the correlation is positive : when mean increases, variance increases ; whereas in winter, the correlation is negative : when mean increases, variance decreases.

In particular, in summer, the linear relationship seems stronger for the last 50 years in Paris-Montsouris than for the beginning of the period, coherently with the result found for Strasbourg for which only the last period is available. While in winter, this relationship is similar for whole times.

In order to verify if this is a common behaviour, the correlation coefficients between the 2 functions of time $\hat{m}(t)$ and $\hat{s}(t)$ have been computed for a set of 55 observation series in Europe provided by the European Climate Assessment and Dataset (ECA&D) project [42], with period lengths between 40 and 100 years. As mentioned in section 7.2.4, the correlation coefficients are used as a measure of the likeliness of the curves. The results are plotted in figure 7.2, and show that this is a quite systematic behaviour in winter(correlation coefficients generally lower than -0.5, represented by black dots), while it is also true in summer, although less systematically (most points show a correlation coefficient larger than 0.5, represented by black dots). A more detailed study of this link can be found in [112].

7.3.2 The residuals

Now we consider that the data have been reduced, since the seasonal effect is avoided by selecting the days of the hot and cold seasons respectively and the mean and variance trends have been removed ; it remains a process $\varepsilon(t)$ defined as :

$$\varepsilon(t) = \frac{X(t) - \hat{m}(t)}{\hat{s}(t)} \tag{7.9}$$

where $\varepsilon(t)$ as random process can have a dynamic which is not stationary.

Once $\varepsilon(t)$ has been estimated as residuals, it is necessary to check the hypothesis done on $e(t)$ or $\varepsilon(t)$ which allows the application of the detrending procedures. So one has to study the stationarity of $\varepsilon(t)$, its autocorrelation, the boundedness and the stationarity of its moments of order larger than 2, and then, if possible, to select a parametric model for $\varepsilon(t)$, but this is not at

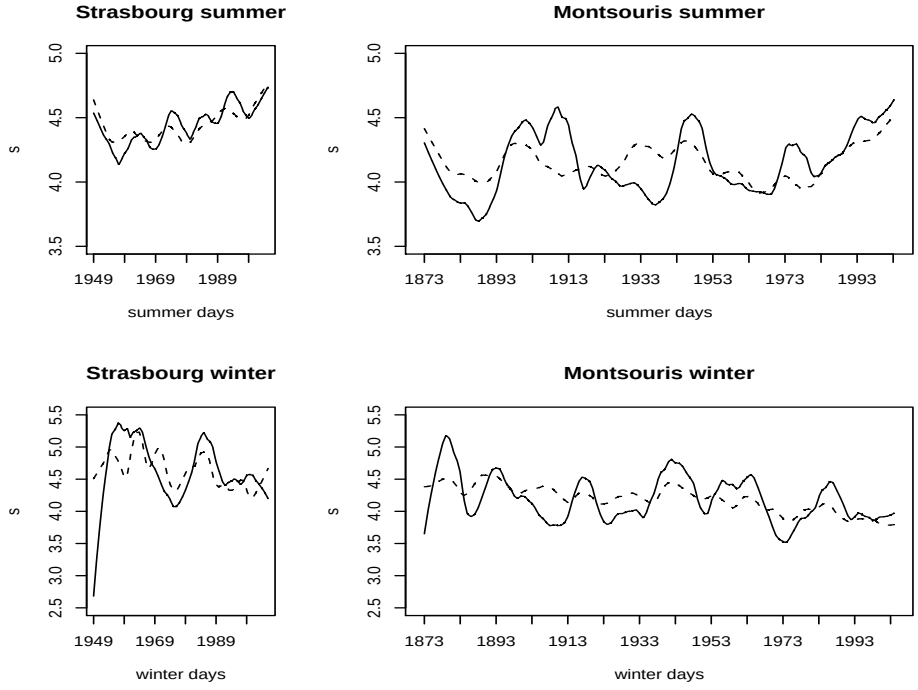


FIGURE 7.1 – Standard deviation function $\hat{s}(t)$ (solid line) and the estimate of $\hat{s}(t)$ as a linear function of $\hat{m}(t)$: $\hat{s}(t) = a\hat{m}(t) + b$ (dashed line) for summer (top panels) and winter (bottom panels) and for Strasbourg (left) and Paris-Montsouris (right). Note that for the summer $a > 0$ and for the winter $a < 0$.

all a requirement in order to apply the detrending procedures. An amount of non stationarity is possible in order to use the mathematical basis of the method of estimation of trends.

For temperature, $\varepsilon(t)$ is often modeled in a quite convenient way, by considering it as stationary, with an autocorrelation of slow order p, AR(p), chosen by AIC criterion. As an example see figure 7.3 representing partial-autocorrelation in the case of the Strasbourg temperature series in summer, for which p=4.

Thus, one can fit for instance a model which has the form :

$$X(t) = \sum_{k=1}^{p} X(t-k) + a(X(t))\eta(t) \tag{7.10}$$

where a is an estimated non linear function bounded and uniformly Lipschitz, which is zero for X large enough in the summer, and η a sequence of independent centred random variables. But in general this modeling can-

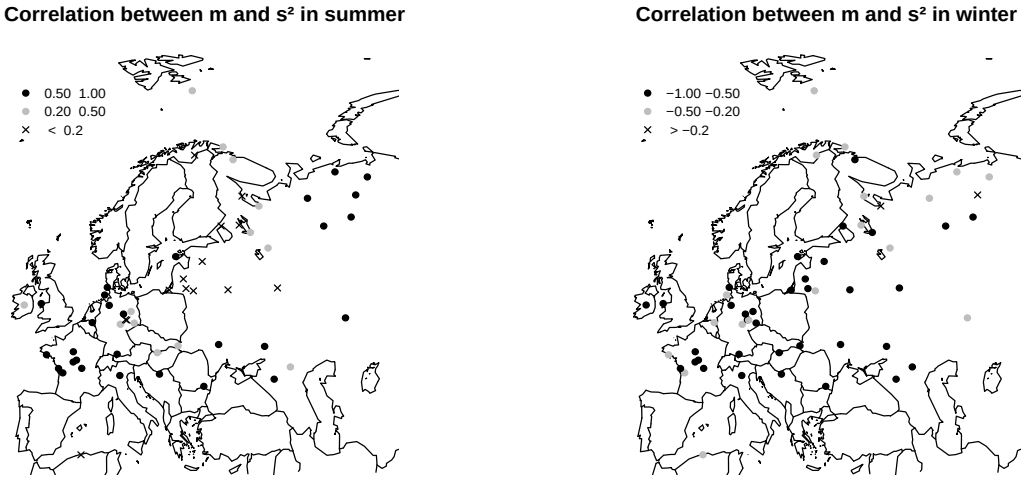


FIGURE 7.2 – Correlation coefficients between the evolution of the mean and the variance of daily minimum temperature in summer (left) and daily maximum temperature in winter (right) for the 55 ECA&D series.

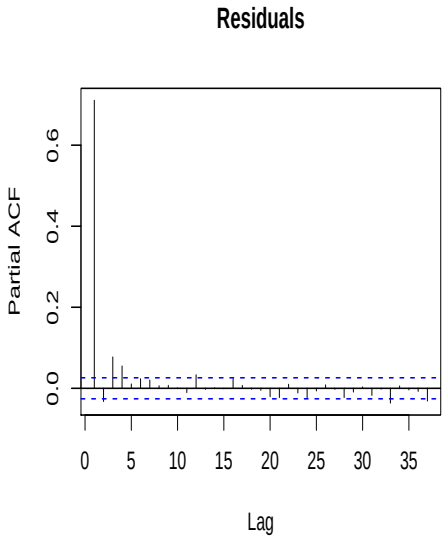


FIGURE 7.3 – Partial- autocorrelation of the residuals $\varepsilon(t)$ for the temperature series of Strasbourg in summer, the lag being expressed in days

not take completely into account the behaviour of the extremes even if it explains the behaviour of the very large (or low) percentiles (see [110]).

7.3.3 Trend in extremes

Applying the cross-validation with a specific algorithm to the time evolution of the parameters of the GEV model (7.5) for the European temperature time series, we find degrees of freedom corresponding to the local (intrinsic) window length of 15 to 20 years. It is larger than the window length of the central field ($m(t)$ and $s(t)$). This can be explained by the fact that the extremes are more variable.

We then compute the trends for the mean of the series and of the extremes ($m(t)$ and $m_X(t)$), where m_X refers to the mean of the extremes, estimated from the GEV parameters as stated in formulas (7.6) and (7.7) and for their variances ($s^2(t)$ and $s^2_X(t)$) , s^2_X being as for the variance of the extremes, expressed from the GEV parameters in (7.6) and (7.7) . Here $X(t)$ are the summer daily maximum temperatures or the winter daily minimum temperatures observed in Paris-Montsouris over the 1873-2003 period and in Strasbourg over the 1949-2005 period, defined in the same way as previously. In computing these trends, we obtained very parallel temporal evolutions. Thus here again, we chose to plot the evolution of the mean or standard deviation function of the extremes together with their estimation via a linear regression with the mean or standard deviation of the whole series. The results are shown in figure 7.4.

We can see that the evolutions of the mean of the whole observations and that of the extremes are strongly linked, whereas the same is true, although in a less systematic way, for the standard deviations. In the same way as previously, we verified the generality of the link in computing the correlation coefficients between the curves for $m(t)$ and $m_X(t)$ for the 55 ECA&D series over Europe. The results shown in figure 7.5 confirm that the two means are strongly linked, as well in winter as in summer, with correlation coefficients generally larger than 0.5 (black dots).

7.4 Trends in extremes and trends in mean and variance

7.4.1 The K hypothesis

As can be seen from the previous results, trends in total mean and in the mean of the extremes, as well as trends in total variance (or standard deviation) and in the variance (or standard deviation) of the extremes look very similar. The question is then : is there some systematic link between these trends, and how can this be tested ? Consider now Y(t) the reduced

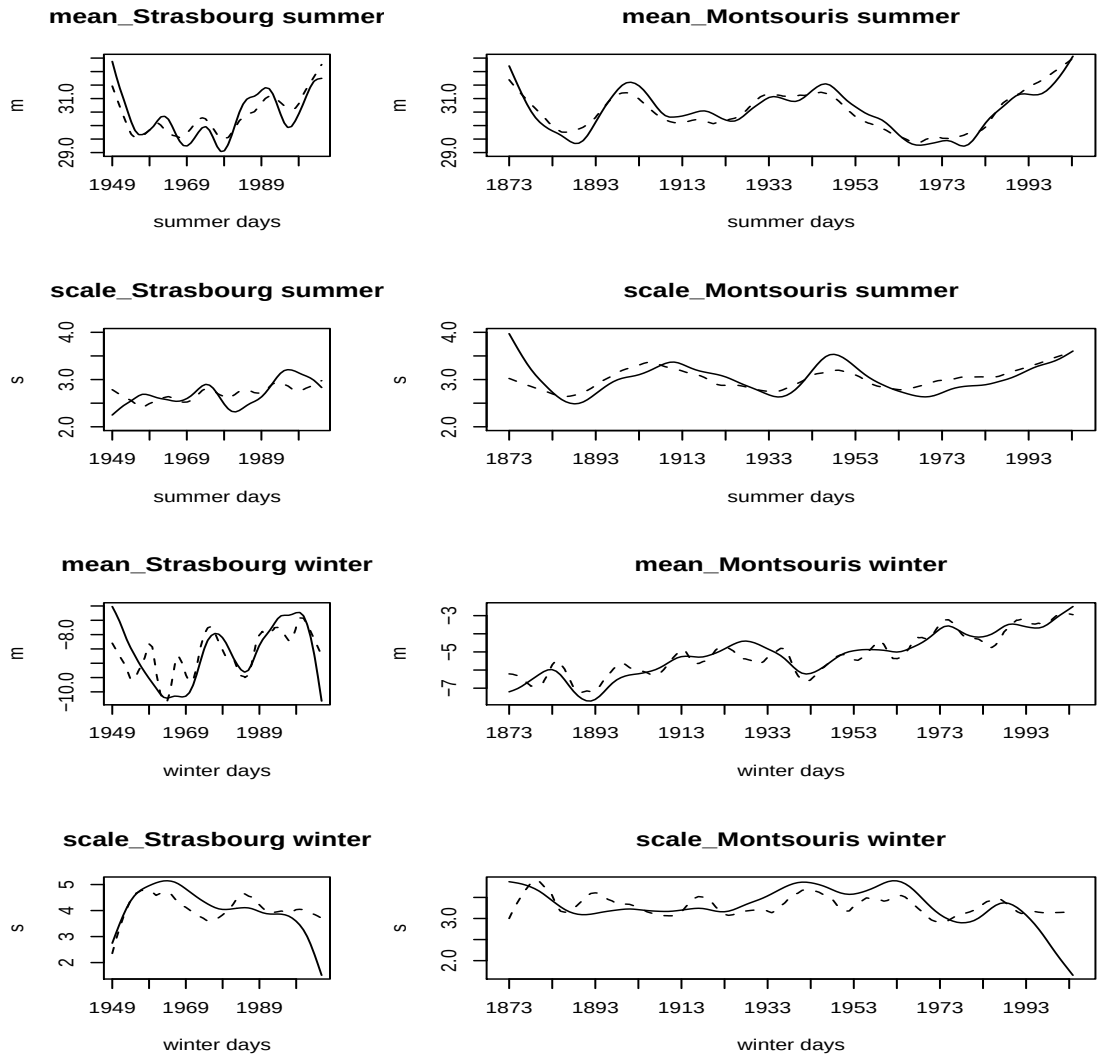


FIGURE 7.4 – Estimates of mean function and standard deviation function of the extremes and their linear estimates in function of the mean function, standard deviation function of the whole observations $\hat{m}_X(t) = a\hat{m}(t) + b$, $\hat{s}_X(t) = a\hat{s}(t)+b$ (dashed line) for summer (top) and winter (bottom) seasons in Strasbourg (left) and Paris-Montsouris (right)

data : $Y(t) = \frac{X(t)-m(t)}{s(t)}$. Some works [50, 85] address the so called H hypothesis : $Y(t)$ is a stationary sequence. It is, as we shall see, a too strong hypothesis and so we choose to study the weaker K hypothesis : the tails of $Y(t)$ are stationary, which is equivalent to say that the parameters of ext_Y : μ_Y, σ_Y, ξ_Y are constant, where ext_X (μ_X, σ_X, ξ_X) and ext_Y (μ_Y, σ_Y, ξ_Y)

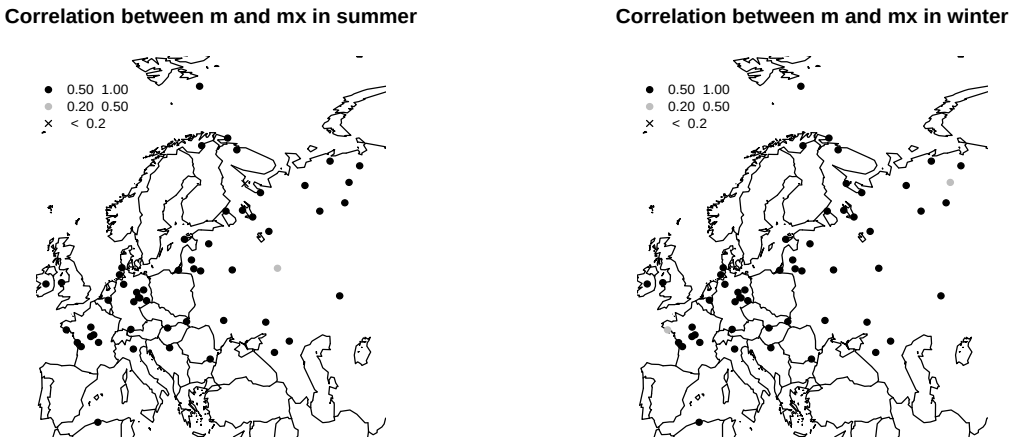


FIGURE 7.5 – Correlation coefficients of the mean of the central field against the one of the extremes in summer (left) and in winter (right) for the 55 European ECA&D temperature series.

are respectively the GEV models for $X(t)$ and $Y(t)$.

Their parameters are linked in the following way :

$$\begin{aligned} \xi_X(k) &= \xi_Y(k) = \xi \\ \sigma_X(k) &= \sigma_Y(k)s(k) \\ \mu_X(k) &= m(k) + \mu_Y(k)s(k) \end{aligned} \tag{7.11}$$

k being the block index. $\sigma_Y(k)$ appears as the multiplier allowing to pass from the standard-deviation $s(t)$ of the global sample at time k of the maximum of the temperature in block k, to the scale parameter of extremes. In fact, as $s(t)$ is a smooth function, $s(k)$ is the standard deviation for the days of block k. The same role for $\mu_X(k)$ is played by $\mu_Y(k)$ but here again multiplied by $s(k)$.

If $\mu_Y(k)$ and $\sigma_Y(k)$ can be considered as constant, which corresponds to the validity of the K hypothesis, we can see that the difference of behaviour between the trends of the whole sample and those of the extremes is completely due to the mean and variance trends. And then, according to (7.6) and (7.7) , we have :

$$\begin{aligned} m_X(k) &= m(k) + s(k)(\mu_Y(k) + \sigma Y(k)E(Z)) \\ s_X(k) &= s(k)\sigma_Y(k)/\xi.(V(Z))^{-1/2} \end{aligned} \tag{7.12}$$

If K is true, these formulas show that the strong almost affine relation found between $m(t)$ and $s^2(t)$ implies a strong relation between the mean $m(t)$ and the mean function $m_X(t)$ of the extremes.

7.4.2 How to test K ?

This is an example of a general method for testing the absence of trends.

The set of possible evolutions of ext_Y parameters is very large. We have previously more or less excluded a possible evolution of ξ [33], but there are many subtle deformations of the distribution of $Y(t)$. So any test suffers from the absence of natural alternatives. It is why we prefer the use of a distance Δ between two functions of time, defined as :

$$\Delta(f,g) = \left(\int_{t \in D} (f(t) - g(t))^2 \right)^{1/2} \tag{7.13}$$

to test the K hypothesis. If we estimate a function of time f by g, $\Delta(f,g)$ is a measure of the quality of g as an estimate of f. Here, we have two estimates of the parameters of ext_Y : $\tilde{f}$, the non parametric estimate, and $\hat{f}$ the constant estimate obtained if K is true, that means if ext_Y is stationary. Now, in any case, K being true or false, one can prove that $\tilde{f}$ converges to f when the sample size T tends to infinity, the speed of convergence depends on the supposed smoothness of f (see [133]). The situation is of course different for $\hat{f}$, if K is true it converges to f with a speed of the order of $\sqrt{T}$ and in this case $\Delta(\hat{f},\tilde{f})$ is, for a large sample, very close to $\Delta(f,\tilde{f})$. On the contrary if K is false, $\hat{f}$ converges in general to a constant (this is a theoretical result, [85, 28]) which is of course different from f and even if it does not converge ; $\Delta(\hat{f},f)$ does not tend to zero and remains larger than some $A > 0$ and the same is true for $\Delta(\hat{f},\tilde{f})$. The intuitive reason is that we try to find f in a set of functions "far away" from f if K is false. These ideas could be translated in an asymptotic result. We prefer the use of a more numerical approach based on simulation. See [28] for details.

Our proposed solution is then to construct by simulation or bootstrap the distribution of $\Delta(\hat{f},\tilde{f})$ under the K hypothesis, that is the distribution of the distances between the non parametric estimates and the constants in case we know that the parameters are constant. To do this, we simulate 1000 samples of the stationary GEV (μ_Y,σ_Y,ξ_Y) distribution with the same size as the series of the maxima of $Y(t)$ (an alternative will be to use bootstrap of the residuals for the model of extremes with very similar results). From each sample, we estimate the extreme parameters in two ways : first, in considering them as constant ; secondly, in considering them as functions of time. Then we calculate the distances between these two estimations of the parameters and so we obtain a distribution of the statistical error of estimation under the K hypothesis. If our true distances are found lower

than the 90th percentile, the K hypothesis is considered as satisfied : the distances are only due to statistical errors.

What happens if ext_Y is not stationary (from the test) ? This is a very difficult question, probably without any precise answer with these data. For example, when $\mu_Y(k)$ can be considered as constant and $\sigma_Y(k)$ as a function of k, so the upper bound of the distribution of $Y(k)$ which is the same as that of $X(k)$ (taken at time t which is the date of the maximum in block k) is $\mu_Y(k) - \sigma_Y(k)/\xi$. The deformation of the tail of the distribution can be described by this trend on the upper bound (of course to keep a mass one to the probability this is compensated by a slight modification of the central part of the distribution). We tested the K hypothesis using this method for

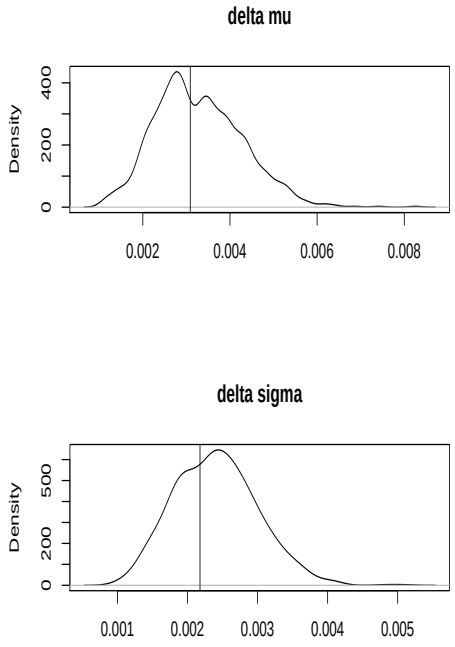


FIGURE 7.6 – The distribution of distances under the K hypothesis with the real values of the same distance for the observations (vertical lines) in summer for the series of Déols and for the location parameter μ(top panel) and the scale parameter σ(bottom panel).

the series in Paris-Montsouris, Strasbourg and all European temperature series. An example is shown in figure 7.6 for the Déols temperature series in summer (ECA&D series over the period 1901-2006). The distances computed from the observed series (for both the location μ and scale σ parameters) lie inside the distribution of the distances when K is satisfied, thus in this

case, the K hypothesis can be accepted. When considering the 55 series over Europe, the hypothesis can be accepted in about 80% of the cases, but differences appear between summer and winter which need to be further analyzed.

7.5 Conclusion

In this paper, we discussed the requirements and assumptions currently associated to the identification of trends in data series and applied non parametric statistical methods to derive trends in European temperature series. We could then show that the trends in mean and variance are linked, and that this link induces a strong relation between trends in the whole dataset and trends in extremes. We developed a testing procedure to verify if the extremes of the centred and normalized series are stationary. If the hypothesis is verified, then the trends in extremes are only due to trends in mean and variance of the whole dataset. This analysis has been conducted for an important number of European temperature series, and showed that this is the case for the majority of the series in Europe. Further detailed analyses have then to be conducted in order to analyze the different behaviours and to identify some physical explanations.

Chapitre 8

Different ways to compute temperature Return Levels in the climate change context

The climate change context has raised new problems in the computation of temperature return levels in using the statistical extreme value theory. As a matter of fact, it is not yet possible to accept the hypothesis that the series of maxima or of high level values are stationary, without at least verifying the assumption. Thus, in this chapter which correspond to one of our articles [113], different approaches are tested and compared to derive high order return levels in the non stationary context. These return levels are computed in extrapolating identified trends, and a bootstrap method is used to estimate confidence intervals. The identification of trends can be made either in the parameters of the extreme value distributions or in the mean and variance of the whole series. Then, a methodology is proposed to test if the trends in extremes can be explained by the trends in mean and variance of the whole dataset. If this is the case, thus the future extremes can be derived from the stationary extremes of the centered and normalized variable and the changes in mean and variance of the whole dataset. The return level can then be estimated as non stationary or as stationary for fixed future periods. The work is done for both extreme value methods : block maxima and peak over threshold, and will be illustrated with the example of a long observation time series for daily maximum temperature in France.

8.1 Introduction

The statistical Extreme Value Theory (EVT) is commonly used by engineers to evaluate the intensity of meteorological extreme events to which industrial installations or buildings have to resist. These events are evaluated as long period return levels, which correspond to very rare events.

The problem raised in this paper concerns the prediction of such very rare events in a non stationary context. As a matter of fact, climate change induces non stationarity in the climatic series, at least for temperature. In this context of statistical evaluation, this means first to identify an evolution of some parameters with time and then to extrapolate this evolution to predict the possible future return levels. Then the question of the definition of trends, which can reasonably be extrapolated in the future, is raised. Once some parametric form, easy to extrapolate, has been identified, its extrapolation means that the observed behavior is supposed to remain valid in the future, the same causes leading to the same consequences. This is a strong hypothesis attached to these methodologies, but difficult to overcome. It is thus important to determine such trends in an as intrinsic manner as possible, that is in avoiding high frequency variations or sample effects. The best trend in a mathematical sense may then not be the optimal one for the extrapolation purpose. Thus, the trend used for extrapolation may not be the identified one over the total observation period length, it can be defined on a sub-period or modified to rid of sample effects. This will be illustrated

by examples.

The EVT exists since the mid 20th century (Gumbel ([62]), Leadbetter et al. ([92])). Then, advancement has been proposed in different fields : Smith (1989) gives basic examples for environmental applications, Coles ([22]) and Katz ([86]) developed the EVT for the fields of oceanography and hydrology. The EVT relies on two general definitions of extreme events : extreme events can be considered as maxima of given blocks of time and then, they are described by the Generalized Extreme Value (GEV) distribution, or they can be considered as exceedances over a defined high threshold, and then, when independent and in sufficient quantity, they follow a Generalized Pareto Distribution, whereas their dates of occurrence follow the trajectory of a Poisson process.

Probability theory assumes that the studied series are stationary (not necessarily independent), so they do not present any cycle nor trend, quite evident form of non stationarity. However, when dealing with climatic data, this assumption has to be carefully considered. As a matter of fact, climatic data often show a seasonal cycle. This can be sometimes overcome if the study is restricted to the season when the extremes of the studied variable mainly occur, but still needs some basic verification. Concerning trends, as previously mentioned the actual climate change context carries with it the possibility of recent trends in meteorological data, especially in the temperature series. Thus, the evaluation of temperature extremes can now hardly be conducted without considering this issue of trends. Extensions have been proposed to the EVT to deal with trends, and can be found in Smith ([141]) as well as in Coles ([22]). Probability theory is applied for every block separately and so gives an estimate of the parameters depending on the block. Then one can constraint these parameters to some predetermined analytical dependence as polynomial or Continuous Piecewise Linear functions (CPL).

In Parey et al. ([114]), a methodology is proposed to estimate high temperature return levels in a non stationary context from observed temperature series in France, in extrapolating recent observed trends. Some insights are given on the choice of the trend and the impact of observed extremes like the 2003 heat wave, based on many observation series in France. Based on these extensions, the present paper aims at further studying the identification and extrapolation of trends in estimating temperature return levels. The first point concerns the trend identification. Once the optimal trend in a statistical sense has been identified, is it possible to determine if this trend corresponds to a background climatic evolution or is strongly influenced by the sample effect ? The sample effect is here mainly due to the position, for instance, at the end of the period of observation of several consecutive very hot or very cold years. This position is considered as a pure random effect. The second point concerns the confidence interval estimation : how can the confidence interval for the return level be evaluated in this non stationary context ? Theses questions will be carefully investigated in this paper in

applying methodologies to identify trends in both the extreme distribution parameters and the mean and variance of the whole series. A methodology is proposed to test the link between these two types of trends, for the two EVT methods, block maxima and peak over threshold and use it to compute future return levels.

The paper is organized as follows : first the mathematical tools are described in section 2, then the main issues are illustrated on the example of the observed daily maximum temperature at the station of Déols in France for the period 1901-2006 (Klein-Tank et al., [42]). In section 3, the extrapolation of trends in the extreme value distribution parameters is discussed whereas the verification and the use of the link between trends in extremes and trends in the mean and variance of the whole dataset are exposed in section 4, before coming to the discussion in section 5. The appendix gives some asymptotic results on the extrapolation of the parameters in the non-stationary context for return levels computation.

8.2 Mathematical tools

8.2.1 General framework

The general framework of the study is the statistical EVT. As stated before, the statistical EVT consists of two methods : the block maxima and the peak over threshold. According to the block maxima method, the block maxima M_j of block j are modeled by independent variables with a GEV distribution. The GEV distribution has 3 parameters : the location parameter μ, the scale parameter σ and the shape parameter ξ. Regarding the Peak Over Threshold (POT) method, the distribution of independent values selected over a fixed high threshold asymptotically converges to a Generalized Pareto Distribution (GPD), whereas their dates converge to a Poisson process. The GPD has two parameters : the scale parameter σ and the shape parameter ξ. In theory, the shape parameter of the GEV and the GPD distributions for the same series are identical. Then, the intensity of the Poisson process is another parameter of the POT method.

In applying each of the methods, choices have to be made in order to verify the assumptions. For the block maxima method, the block length has to be determined in such a way that it is sufficiently long to allow the asymptotic convergence of the maxima distributions and the independence between two consecutive maxima. For the POT method, the threshold has to be selected in such a way that the process of the dates of occurrence follows a Poisson process and that the property of threshold stability of the GPD is fulfilled (Smith, [141]). On the other hand, the values have to be independent, and a declusterization technique has often to be used. Goodness of fit tests allow to check the fit of the obtained distribution of exceedances to the GPD, the belonging of the set of dates to a Poisson process, and vali-

date the hypotheses of independence. The threshold selection can be based on the properties of the GPD, such as the constancy of the shape parameter or the linearity of the mean of the exceedances over an adequate threshold value.

8.2.2 Trends in the parameters of the extreme value distributions

8.2.2.1 Trend identification

Non stationarity is translated for EVT models in the fact that parameters are non constant functions of time. So once the general statistical framework is set, trends can be considered on each parameter for both methods : the location, scale and shape parameters of the GEV and the scale and shape parameters of the GPD as well as the intensity of the Poisson process are no more considered as constant, but are allowed to evolve with time (or with any other covariate). The shape parameter ξ is however the most delicate to estimate, and it could be tricky to differentiate possible evolutions from estimation errors. Tests on different periods of a long observation series have proved that this parameter does not significantly evolve with time (Parey et al., [114]). On the other hand, more sophisticated non parametric studies lead to the same conclusion (Hoang et al. ([79]). Thus, in the following, the shape parameter ξ will be considered as constant for both GEV and GPD distributions.

Therefore, trends will have to be identified for the other parameters. As the ultimate goal is then to extrapolate such trends in order to compute return levels, parametric forms given by closed formulas have to be considered. Firstly, non parametric approaches allow to suggest the general form of these trends, and it can be reasonably approximated by polynomials or other family like the family of Continuous Piecewise Linear functions (CPL) (Parey et al. ,[114]). These families of functions are often ordered (or partially ordered) by an integer d as the polynomial degree or the number of pieces of a CPL.

The choice of d is then based on the likelihood ratio test or on another procedure of model choice based on likelihood as the Akaike criterion (Davison ,[32]).

For the GEV distribution, the procedure is the following : for every pair of integers $(dmu, dsigma)$, one computes the coefficients which maximizes the log-likelihood for the block maxima M_j with a location parameter $\mu(j)$ and a scale parameter $\sigma(j)$ considered as polynomials of degree dmu for $\mu(j)$ and $dsigma$ for $\sigma(j)$ with these coefficients. To choose the optimal model in this polynomial class for $\mu(j)$ and $\sigma(j)$, one of the previously mentioned procedures can be used. In fact because the pairs of integers are not totally ordered, the criteria valid only on nested models can not be applied and in

this case the chosen solution is the one with the maximum likelihood. The tests are at level $\alpha = 0.05$ and the Akaike penalization is $2(dmu + dsigma)$. The same procedure can be applied for CPL functions with some restrictions due to mathematical complications.

For the POT method, the same principle is applied : for every integer d ($d \geq 1$), the set of coefficients which maximizes the log-likelihood for data X_i with a Poisson intensity I depending on time as a polynomial of degree d with these coefficients is computed. To choose the optimal model for $I(t)$, the best model for d is tested against the best model for $d + 1$ using a likelihood ratio test or a Akaike criterion and an associated level of confidence of the test, for instance $\alpha = 0.05$. Degree $d + 1$ is chosen if the test rejects degree d at level α. The same holds true for the scale parameter σ of the GPD. Here again, this procedure is suitable for CPL functions too.

The advantage of CPL models is to capture more details and to derive linear trends which seems, as we shall see later, to be more appropriate for the Return Level extrapolations. For instance for long periods, instead of a quadratic trend (3 parameters) we can find a model with three pieces (6 parameters) which is much more informative. Still, statistical optimization is harder for CPL functions than for polynomials and the mathematical theory for model choice and estimation is more difficult (see Dacunha-Castelle & Gassiat, [30]) for details on this point). The main difficulty of CPL functions lies in their non identifiability, i.e the same model can be defined by different sets of parameters. The classical theory of likelihood and so the use of AIC criterion are not valid anymore. This issue results in problems of estimation of parameters that can have no consistent estimators. This can be avoided under the constraint of separation of angular nodes, e.g. by 10 years. A misuse of CPL models could however be to systematically conduct the prediction in extrapolating the last linear piece. The slope of this linear fragment may be too dependent on the distribution of the last observed values in the last period. We thus define a specific statistical procedure in order to control these edge effects and to choose the extrapolated trend.

It thus seems a practical advantage to benefit from both classes of functions, polynomials and CPL, to avoid sampling artifacts.

8.2.2.2 Return levels in a non stationary context

Once the model for the evolution of the distribution parameters with time has been selected, the return level has to be estimated in extrapolating these trends. This necessitates that the parameters are monotonous functions, thus often polynomials will be restricted to degree 2. In the stationary case, the return level z_a for a years is the level for which the probability of exceedances every year is equal to $1/a$. In the non-stationary case, knowing the identified trend, z_a will then be re-defined as the unique level such that the expectation of the number of exceedances over z_a in the next a years

will be 1. We shall define the return level z_a for a years starting from the date t_0 , so the return level has to be thought as a function of the initial date and of the number of years taken into account to make the projection. Let $D(t_0, a)$ be the set of all days that belong to $t_0, t_0 + 365a$. Then, for the block maxima method, return level z_a is such that :

$$\frac{1}{nb}\sum_{t=t_0} t_0 + 365a \left\{ 1 - exp\left[-\left(1 + \frac{\xi}{\sigma(t)}(z_a - \mu(t)) \right)^{-1/\xi} \right] \right\} = 1 \quad (8.1)$$

nb is the number of days in each block, as each day in a block shares the same probability of being the maximum.

For the POT method, z_a is such that :

$$\sum_{t \in D(t_0,a)} \left(1 + \frac{\xi}{\sigma(t)}(z_a - u) \right)^{-1/\xi} I(t) = 1 \quad (8.2)$$

$I(t)$ is the mathematical expectation of the number of exceedances at date t and the Pareto term $\left(1 + \frac{\xi}{\sigma(t)}(z_a - u)\right)^{-1/\xi}$ is the mathematical expectation that the exceedance is larger than the level z_a, u being the threshold.
The definition for the stationary case is a particular case of this one.
In each case, z_a is then the result of a minimization procedure of :

$$\frac{1}{nb}\sum_{t=t_0} t_0 + 365a \left\{ 1 - exp\left[-\left(1 + \frac{\xi}{\sigma(t)}(z_a - \mu(t)) \right)^{-1/\xi} \right] \right\} - 1 \text{ for GEV} \quad (8.3)$$

and

$$\sum_{t \in D(t_0,a)} \left(1 + \frac{\xi}{\sigma(t)}(z_a - u) \right)^{-1/\xi} I(t) - 1 \text{ for POT} \quad (8.4)$$

We will consider that the optimization procedure properly converges if the value of the expressions (8.3) and/or (8.4) for z_a are lower than 10^{-5}.

8.2.2.3 Associated confidence intervals

In the stationary context, confidence intervals can be derived from the standard deviation of the estimated return levels, using the approximate normality of the maximum likelihood estimator or of any function of it. This so-called “delta method” will be used here to evaluate the confidence interval in the stationary case.

If a trend is identified, the confidence interval has to take both sampling errors and trend identification errors into account. The proposed method is based here on a bootstrap technique, which will be described first for the block maxima method.

Once the trend has been identified, the block maxima series can be expressed as :

$$M_j = \hat{\mu}(j) + \hat{\sigma}(j)e_j \tag{8.5}$$

where M_j are the block maxima, $\hat{\mu}(j)$ and $\hat{\sigma}(j)$ the estimated optimal functions for the location and scale parameters and e_j are residuals having a GEV distribution of location parameter 0 and scale parameter 1.

Then, the residuals are computed from the observed maxima and the estimated optimal trends. This residuals series is bootstrapped and a new sample of block maxima is computed. From this new sample, the trend identification method allows to compute the optimal trend of the same form as the original one and a new return level is estimated. In applying this procedure a high number of times (500 times here), a distribution of the estimated return level is produced, from which a confidence interval can be derived for a defined level, for example 90% .

For the POT method, a similar procedure can be applied, but with the help of simulation. The number of dates of threshold exceedances follows a Poisson distribution. Thus, first a new number of dates N^* will be sampled from a Poisson distribution on integers with parameter $\bar{I}$ corresponding to the mean value of $I(t)$ over the observation period. Then, a new series of N^* dates of exceedances is drawn (simulated) as a sample of the probability density function giving sets of new dates $t_1, ..., t_{N^*}$ and so new estimates $I^*(t)$ of the intensity.

At these new dates, we shall get new values of V_u over the threshold u. Observed values can be expressed as :

$$V_u = u + \pi\hat{\sigma}_t \tag{8.6}$$

where $\hat{\sigma}$ is the estimated scale parameter and π a residual following a GPD of scale 1. A new set of values $V^*(u)$ can be produced in bootstrapping the i.i.d. π. Then, values sampled in this way are associated to the new set of dates of exceedances, while for the other dates of the series, a value is sampled from the original values under the threshold u. This procedure is of course applied on the "declusterized" series, that is the series from which the dependent values over the threshold and their dates have been eliminated. Here again, in applying this sampling procedure a high number of times (for instance 500), it is possible to estimate a high number of new sets of parameters $I*(t)$, $\hat{\xi}$ and $\hat{\sigma}^*(t)$ from which a distribution of the return level can be derived by the previous formula. Thus, the confidence interval can be computed as the lower and upper percentiles of this distribution according to a fixed level, for instance 90%.

8.2.3 Link between trend in extremes and trends in the whole series : use of the K hypothesis

8.2.3.1 Non parametric tools to derive trends

In Parey et al. ([112]), non parametric methods have been introduced and used to derive trends in mean and variance of the summer daily maximum (and winter daily minimum) temperature series. A link has been identified between the trend in mean and the trend in variance, the variance increasing when mean increases in summer for different European location, and especially in France. The local polynomial estimators (loess) will be used here, with the same smoothing parameter, to compute the trends in mean $m(t)$ and in variance $s^2(t)$ of the observed summer daily maximum temperature $X(t)$.

8.2.3.2 Link with the trends in extremes

The question then is whether a link can be identified between these non parametric trends in mean and variance of the whole dataset and the trends in extremes. To study it, $Y(t)$ is defined as the centred and normed variable computed from the observed series $X(t)$ as :

$$Y(t) = \frac{X(t) - m(t)}{s(t)} \tag{8.7}$$

For GEV, the parameters for $X(t)$ can be obtained from those of $Y(t)$ using the following relationships :

$$\begin{cases} \xi_X = \xi_Y \\ \sigma_X(t) = \sigma_Y(t)s((t) \\ \mu_X(t) = m(t) + \mu_Y(t)s(t) \end{cases}$$

If the extreme value model for $Y(t)$ is stationary, then the return level of $X(t)$ can be obtained from the stationary return level for $Y(t)$ and the trends in mean and variance of the whole dataset. Thus the trend in extremes is mainly explained by the trends in mean and variance of the whole series. Then, the trends are to be identified from the whole ensemble of data, which let the identification be more robust as obtained from a much larger dataset.

A methodology to test the so called "K hypothesis", which states that the extremes of $Y(t)$ can be considered as stationary, has been proposed and detailed in the GEV case in Hoang et al. ([78]). Only the general principles are recalled here. The test consists in comparing the distance between the two functions of time corresponding to the stationary and non stationary estimations of the parameters of the extreme distribution of $Y(t)$ to a table of such distances constructed in using a known stationary extreme value distribution with the same parameters. The steps are then the following :

1) compute a non parametric trend for the mean $m(t)$ of the observed values $X(t)$, using loess : $\hat{m}(t)$.

2) compute the variance $var(t)$ as $(X(t)-\hat{m}(t))^2$ and its non parametric trend using the same loess method : $\hat{s}^2(t)$.

3) compute $Y(t) = \frac{X(t)-\hat{m}(t)}{\hat{s}(t)}$.

4) estimate $\hat{\mu}_j$ and $\hat{\sigma}_j$, j being the block index, the location and scale parameters of the GEV distribution :

- as constant in time $\hat{\mu}_0$ and $\hat{\sigma}_0$.
- as time varying $\hat{\mu}(j)$ and $\hat{\sigma}(j)$.
- and their distances :

$$\Delta\mu = \left(\int_{t\in D} (\hat{\mu}(t) - \hat{\mu}_0)^2\right)^{1/2} \text{ and } \Delta\sigma = \left(\int_{t\in D} (\hat{\sigma}(t) - \hat{\sigma}_0)^2\right)^{1/2}$$

D being the set of days in the series.

5) simulate 1000 samples of the same number of maxima following GEV($\hat{\mu}_0$,$\hat{\sigma}_0, \hat{\xi}$). Then compute the 1000 distances Δ between the parameters estimated as constant and time varying from these samples, which are considered as empirical distributions of distances for μ and σ.

6) situate $\Delta\mu$ and $\Delta\sigma$ in these empirical distributions of distances, obtained from the stationary distribution. Finally, decide to accept or reject the hypothesis.

If the distance computed from the observed series lies inside the distributions, for instance between the 5th and 95th percentiles, the K hypothesis can be accepted and the distance is only due to statistical errors. Of course this acceptance means no more that with this amount of data, it is impossible to have a statistical evidence for the non-stationarity.

The same procedure can be applied for the POT method in computing ΔI and $\Delta\sigma$ and comparing them to the distributions of such distances coming from a sample of a known stationary situation. The main difference lies in the fact that the non parametric evolution of $I(t)$ is computed using a kernel method rather that loess.

8.2.3.3 The use of this link to compute return levels and their associated confidence intervals

Once the K hypothesis has been tested and can be accepted, one can imagine different ways to use it in the computation of future return levels. On one hand, the return levels and associated confidence intervals can be evaluated in the non stationary context, but in extrapolating the trends in mean and variance of the whole dataset instead of the trends in the parameters of the extreme value distributions. The trends in mean and variance of the whole dataset will be computed here as a two-parts continuous piecewise linear function, based on the identification of one break point in their

evolutions. This will be done in using the so called "brute force" procedure proposed by Mudelsee ([100]). Thus, $m(t)$ and $s(t)$, respectively the time evolutions of the mean and standard deviation of the whole dataset, are considered as two linear parts, and the second part is extrapolated to compute the non stationary return levels, as follows :

$$\frac{1}{nb}\sum_{t=t_0} t_0 + 365a\left\{1 - exp\left[-\left(1+\frac{\xi}{\sigma_Y}(\frac{z_a - m(t)}{s(t)} - \mu_Y)\right)^{-1/\xi}\right]\right\} - 1 = 0 \text{ for GEV} \tag{8.8}$$

and

$$\sum_{t\in D(t_0,a)}\left(1+\frac{\xi}{\sigma_Y}(\frac{z_a - m(t)}{s(t)} - u)\right)^{-1/\xi} I_Y - 1 = 0 \text{ for POT} \tag{8.9}$$

Then, the confidence intervals are constructed in applying a bootstrap procedure :

- bootstrap $Y(t)$ to obtain a high number of samples $Y^*(t)$ and compute the stationary parameters for the extreme value distribution of each $Y^*(t)$.
- compute $X^*(t)$ from each $Y^*(t)$ and $m(t)$ and $s(t)$.
- from each $X*(t)$, estimate two piecewise linear trends with the previously identified break point, to obtain $m^*(t)$ and $s^*(t)$.
- compute each return levels z_a^* using (8.8) or (8.9).
- get the confidence interval as the desired inter-quintile interval from the sample z_a^* according to the chosen confidence level (5th and 95th percentiles for the 90% confidence interval).

Another way is to consider that the return levels can be computed in the stationary context for fixed periods of time, for example 30 years. Thus, the present day return levels are the stationary ones evaluated over the most recent period of the observed series. Then, using the K hypothesis, the future return levels can be computed from the stationary distribution of the extremes of $Y(t)$ and the values of mean m and standard deviation s over a future 30-year period. The confidence interval is computed in using the delta method, taking both the errors in the extreme value distribution parameters of $Y(t)$ and those of m and s, these last ones being evaluated using a bootstrap technique. The values for m and s for a future period can be retrieved either from the trend extrapolations or from climate model simulations.

8.3 Example : extrapolation of trends in extreme value distribution parameters

The temperature series chosen to illustrate the previously described methodology to compute return levels in the non stationary context is provided

by the European Climate Assessment and Dataset (ECA&D) project (Klein Tank et al. ([42]). It consists of daily maximum temperature observed at the station named Déols (in the center of France) over the period 1901–2006 (106 years). In order to limit the problem linked to the seasonal cycle in temperatures, the high temperature return levels will be computed from the sub-series corresponding to the summer days, when high temperatures are susceptible to occur. The summer period has been defined from empirical statistical analyses regarding the occurrence of very high temperatures as the 100 days between the 14th of June and the 21st of September. Thus a series of 106 times 100 days is extracted from the observation series to produce the series of daily maximum temperatures in summer in Déols between 1901 and 2006. The observed maximum occurred on the 2nd of August 1906 with a temperature of 40.5C.

8.3.1 The block maxima method

The study has firstly been conducted using the block maxima method. The first step consisted in choosing the block length and verifying the identical distribution of the selected maxima between the blocks. To do so, the repartition of the block maxima and of the annual maxima within each block have been compared for two block sizes : 25 days (4 blocks per summer) and 50 days (2 blocks per summer). The results are summarized in table 8.1. The maxima are reasonably well represented in each block. However, for the 25-day blocks, annual maxima seems to occur more rarely at the beginning (block 1) and at the end of the summer (block 4), revealing the remaining of a seasonal cycle, even though the summer season only has been retained. In order to avoid a fastidious deseasonalisation procedure, a 50-day block length will be chosen, for which the repartition in each block is more similar. This leads to 2 maxima per summer, thus 212 values over the period and is a good compromise between the length of the series of extremes and the length of the block able to insure independence and identical distribution of the selected maxima.

The trend identification procedure for polynomials leads to a quadratic trend for the location parameter $\mu(j)$, j being the block number, whereas the scale parameter σ can be considered as constant. Figure 8.1 shows the optimal identified trend superimposed on the block maxima evolution. The extrapolation of this trend gives the following results for the 30-, 50- and 100-year return levels (RL30, RL50 and RL100 respectively), with their 90% confidence intervals : RL30 = 42,0C [40,0 43,5] ; RL50 = 44,6C [41,2-46,3] ; RL100 : 54,01C.

The 50- and 100-year return levels are in italics as the convergence of the optimization is not robust. This example shows how polynomial trends may lead to very high, physically unrealistic temperature values. The quadratic form is due to the occurrence of high temperature values at the beginning

	25-day blocks in summer				50-day blocks in summer	
	Block 1	Block 2	Block 3	Block 4	Block 1	Block 2
Block maxima	92	115	105	112	102	110
Summer maxima	16	44	37	9	60	46

TABLE 8.1 – Repartition of the 424 block maxima and the 106 summer maxima in the 4 blocks for 25-day blocks and of the 212 block maxima and 106 summer maxima in the 2 blocks for 50-day blocks. For the 25-day blocks, the dates of the blocks are : block 1 from the 14th of June to the 8th of July ; block 2 from the 9th of July to the 2nd of August ; block 3 from the 3rd of August to the 27th of August and block 4 from the 28th of August to the 21st of September. For the 50-day blocks, the dates of the blocks are : block 1 from the 14th of June to the 2nd of August ; block to from the 3rd of August to the 21st of September

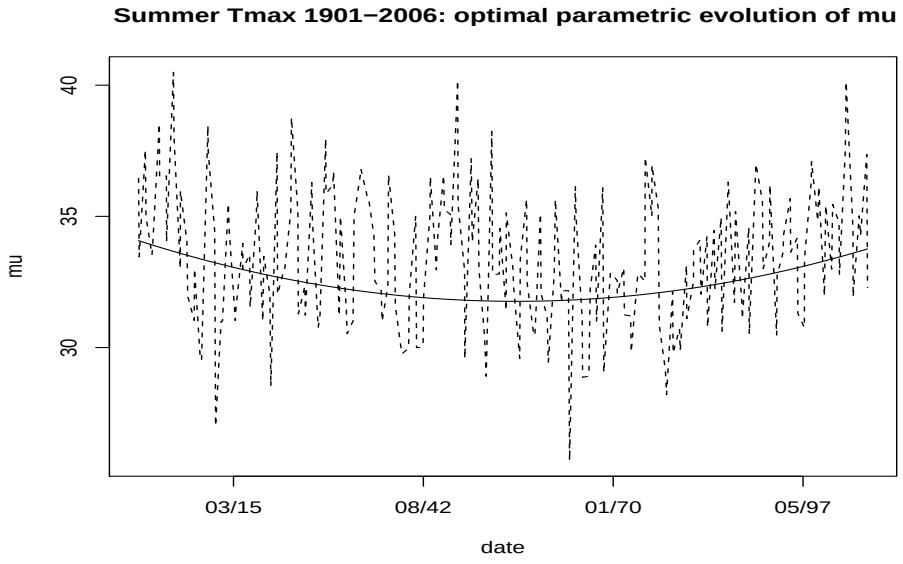


FIGURE 8.1 – Optimal polynomial trend for the location parameter $\mu(j)$ (j is the block index) superimposed on the evolution of the block maxima, for the daily maximum temperature in summer in Déols between 1901 and 2006

and at the end of the sample.

Another proposed way of evaluating the return levels is to look for the optimal CPL trend, and use it to define the period over which the trend should better be considered, if one wants to avoid high frequency fluctuations contained in the sample. The optimal CPL trend for the daily maximum temperature in summer in Déols is found in 2 parts for the location parameter $\mu(j)$, while the scale parameter remains constant. The identi-

fied break point is in year 1974. Figure 8.2 shows this optimal CPL trend, superimposed on the block maxima evolution.

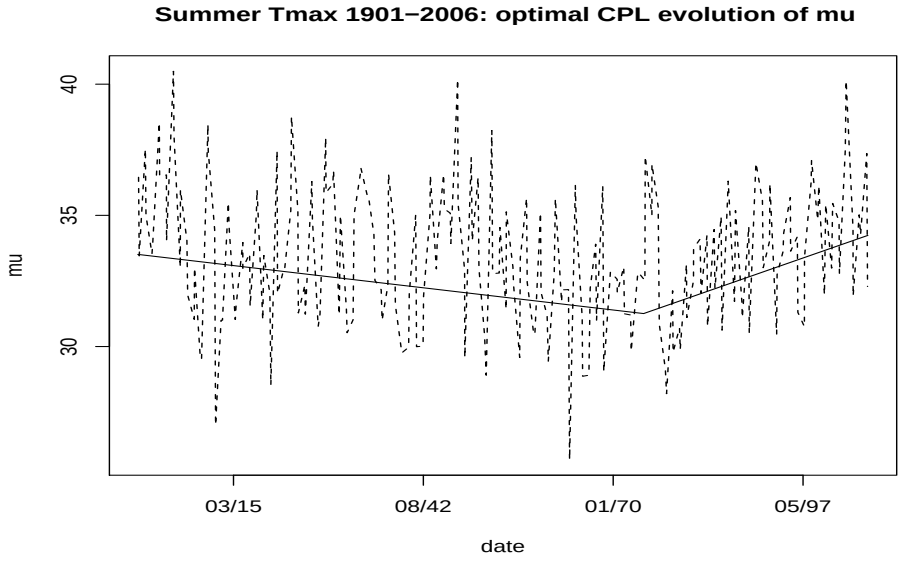


FIGURE 8.2 – Optimal CPL trend for the location parameter $\mu(j)$ (j is the block index) superimposed on the evolution of the block maxima, for the daily maximum temperature in summer in Déols between 1901 and 2006.

The extrapolation of this last linear part, which could be judged as more representative of the recent warming trend, leads to the following return levels :

RL30 = 42,0C [39,7 43,9]
RL50 = 44,1C [40,6 45,7]
RL100 = 49,3C [41,4-51,7]

Here again, although the convergence is better achieved, the optimization is still difficult for the 100-year return level. The two approaches, using the optimal polynomial trend or the last part of the optimal CPL trend, lead to very similar results for the 30- and 50-year return levels. The extrapolation to 100-year return levels remains however not reasonable, probably because of the very high values encountered at the end of the sample, whereas when beginning after the break point, the sub-sample starts with the lowest block temperatures reached in the sample. It thus does not seem reasonable to assume an unchanged trend on such a long period.

8.3.2 The POT method

8.3.2.1 Trend extrapolation

The same analysis has then been conducted with the POT method.

The first task is here to choose the threshold and the declusterization procedure. The declusterization procedure used is a simple one, consisting in retaining only the maximum value when several exceedances concern consecutive days. The threshold selection is then guided by the properties of the GPD, as stated in section 8.2.1, to which we add the necessity for the identified trends in the Poisson process intensity and in the scale parameter of the GPD to remain constant. This leads to a threshold of 31,4C, with 400 independent exceedances for the whole series length.

Then, the optimal polynomial trend is identified for both the Poisson process intensity $I(t)$ and the scale parameter of the GPD $\sigma(t)$: the result is a quadratic trend for $I(t)$ and a constant σ. The trend in the Poisson process intensity is illustrated in figure 8.3, superimposed on the frequency of threshold exceedances each summer. The 30-, 50- and 100-year return levels obtained in extrapolating this trend are the following :

RL30 = 39,6C [38,8 40,9]
RL50 = 40,1C [39,3 41,4]
RL100 = 40,7C

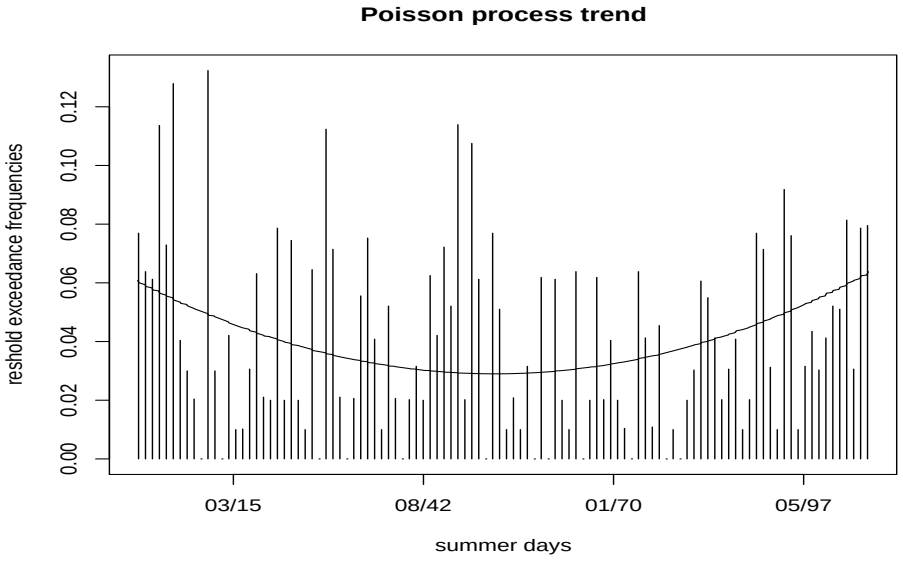


FIGURE 8.3 – Optimal polynomial trend for the Poisson process intensity superimposed on the frequencies of threshold exceedances each summer, for the daily maximum temperature in summer in Déols between 1901 and 2006

Here again, for the 100-year return level, the convergence criterion is not fulfilled. The obtained 30- and 50-year return levels are lower than those obtained previously with the block maxima method.

As previously, the optimal CPL trend has then been identified. Here, the results are different, since the optimal trend for the Poisson process intensity has 3 parts, with the last one strongly influenced by the edge effect of the recent hot summers in France, as can be seen in figure 8.4, with break points respectively in 1965 and 1999. The scale parameter is still constant. The best fitted 2 parts evolution, although not optimal, identifies the break point in 1997, which confirms the strong edge effect of the last hot summers in the series. In order to overcome this effect, the first idea could be to extrapolate the second part rather than the last one. This trend is however very low, and probably artificially low, as the effect of the recent high values is totally excluded. It has thus been decided to cut the sample at the date of the identified second break point, and to look for the same optimal trends over the 1965-2006 period. The optimal trend is then linear for the intensity of the Poisson process, while the scale parameter of the GPD remains constant. The extrapolated return levels are then the following :

RL30 = 39,3C [37,5-41,3]
RL50 = 39,7C [37,7-42,2]
RL100 = 40,4C [37,8-44,5]

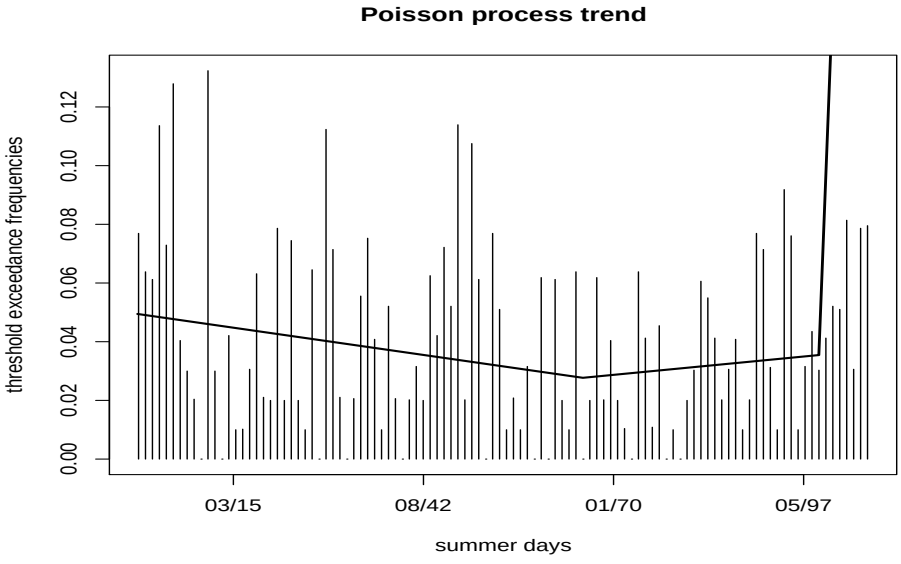


FIGURE 8.4 – optimal CPL trend for the Poisson process intensity superimposed on the frequencies of threshold exceedances each summer, for the daily maximum temperature in summer in Déols between 1901 and 2006

8.3.2.2 Effect of a change in the location parameter of the GEV or in the intensity of the Poisson process

It seems from the previous results that an evolution in the Poisson process intensity of the POT method has a lower impact on the return level than an evolution of the location parameter of the GEV distribution. In order to confirm this observation, the derivation of the return level for the two methods against these parameters are computed. The calculations here are conducted for the stationary return levels, as their analytical expressions are easier to derivate than the proposed expressions for the non stationary return levels (equations (8.3) and (8.4) above). The only aim is to understand the effects of an increase in the location parameter of the GEV distribution or in the intensity of the Poisson process on the estimated return level.

For GEV, the T-year stationary return level ZT writes :

$$Z_T = \mu - \frac{\sigma}{\xi}\left[1 - ((log(1-T)^{-\xi})\right] \tag{8.10}$$

in the case where $\xi \neq 0$ like here, and then $\frac{\partial Z_T}{\partial \mu} = 1$.

For POT, the same T-year stationary return level Z_T writes :

$$Z_T = u + \frac{\sigma}{\xi}\left[(Tn_yI)^{\xi} - 1\right] \tag{8.11}$$

where n_y is the number of exceedances each year. n_y is thus the mean number of declusterized exceedances of the threshold per year. Then :

$$\frac{\partial Z_T}{\partial I} = \sigma(Tn_yI)^{\xi-1} \text{ and } \frac{\partial Z_T}{\partial \sigma} = \frac{1}{\xi}\left[(Tn_yI)^{\xi} - 1\right]$$

With the values of the parameters in the studied case, this leads to the results summarized in table 8.2. One can see that the impact of an increase in I on the return level is much lower and decreases when the return period increases, whereas the impact of an increase of the scale parameter σ is much higher and increases with the return period. This explains the observed differences between the return level extrapolation according to the two extreme value methods.

Return period	30-year	50-year	100-year
$\frac{\partial Z_T}{\partial I}$	0.0077	0.0041	0.0017
$\frac{\partial Z_T}{\partial \sigma}$	2.87	3.03	3.22

TABLE 8.2 – Values of the derivation of the return level against the Poisson process intensity I and the scale parameter of the Pareto distribution σ for the 30-, 50- and 100-year return periods

With the values of the parameters in the studied case, this leads to the results summarized in table 8.2. One can see that the impact of an

increase in I on the return level is much lower and decreases when the return period increases, whereas the impact of an increase of the scale parameter σ is much higher and increases with the return period. This explains the observed differences between the return level extrapolation according to the two extreme value methods.

8.3.3 Research of a break point in the evolution of the mean

Until now, the trends have been identified for the whole sample of extremes, that is over the entire period, and the CPL form of the trends lead us to use sub-periods. Another idea could be to look for a break in the evolution of the mean daily maximum temperatures in summer, and then to consider only the period beginning after this break. In order to do this, the method used consists in adjusting a 2 pieces linear trend to the series of daily maximum temperatures in summer and identifying the break point as the one for which the standard deviation of the residuals is the minimum. All dates of the series are tested as potential break points, except the 5 first and last summers in order to minimize the edge effects. This follows the method proposed by Mudelsee ([100]). This procedure leads to a break in the trend of mean daily maximum temperature evolution in 1956, as illustrated in figure 8.5.

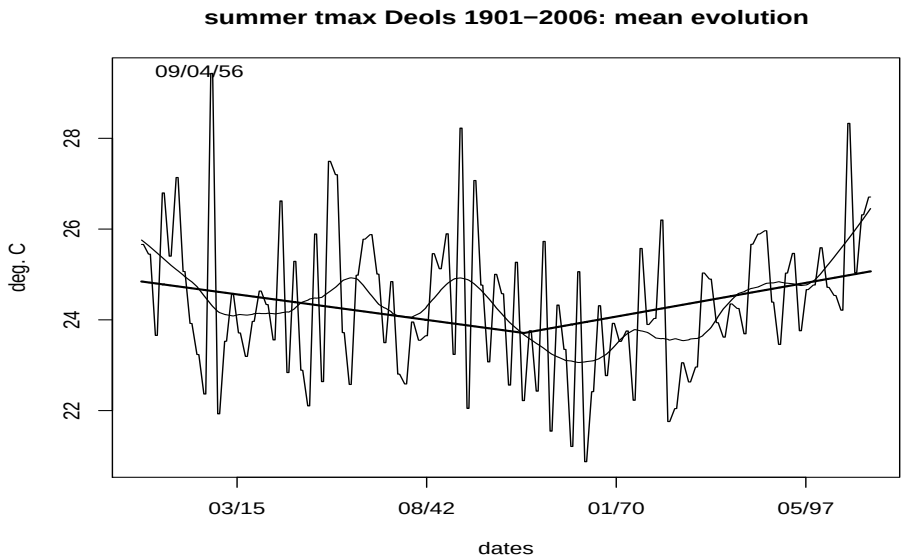


FIGURE 8.5 – Non parametric (loess, thin smooth line) and optimal 2 pieces linear trends (bold) for the summer daily maximum temperature in Déols between 1901 and 2006, superimposed on the evolution of the seasonal means. The date of the break point is written in the top left corner

Starting from this result, the evolution of the parameters of the extreme value distributions for the GEV and POT methods has been computed over the period 1957-2006. For GEV, the optimal polynomial trends are of degree 1 (linear) for the location parameter ? and 0 (stationary) for the scale parameter σ. The extrapolated return levels and their 90% bootstrap confidence interval are then :

RL30 = 40,8C [39,3 42,1]
RL50 = 42,2C [40,2 43,5]
RL100 : 45,6C [40,8-46,5]

These values are lower than those obtained previously for the trend identified over the total period length or over the 1975-2006 period after the break in the trend of the location parameter, as the 50-year return level now is comparable to the 30-year return level previously obtained.

When using the POT method, first the threshold has been selected in the same way as previously as 31,0C, corresponding to 200 independent exceedances. The optimal trends are of degree 2 (quadratic) for the intensity I of the Poisson process and 1 (linear) for the scale parameter of the Pareto distribution σ. Then the extrapolated return levels and their 90% bootstrap confidence intervals are the following :

RL30 = 41,3C [35,9 38,6]
RL50 = 42,9C [35,9 39,9]
RL100 = 46,9C [36,0 42,1]

What appears first is the fact that the extrapolated return levels lie outside their bootstrap confidence intervals. This is due to the fact that the 2003 maximum is far above the other high levels and over this shorter period, it alone implies the obtained linear trend for the scale parameter. Figure 8.6 shows the selected independent values over the threshold on the left, and the identified optimal trend of the scale parameter together with all the different trends computed from the bootstrap samples in the evaluation of the confidence interval (dotted lines) on the right. Very few trends in the bootstrap samples are as steep as the optimal trend of the original sample (some are even decreasing, depending on the location of the 2003 event in the bootstrap sample). And, as seen before, the evolution of the scale parameter has the greatest influence on the computed return level in the POT method. This explains why the extrapolated return levels are outside the bootstrap confidence intervals. This shows then that this methodology of extrapolation of the trends in the parameters of the POT distributions is very sensitive to unusually high values encountered at the end of the period, and cannot be reliably applied in such cases. In this case, it is therefore recommended to deal with the longest possible series to avoid such a drawback.

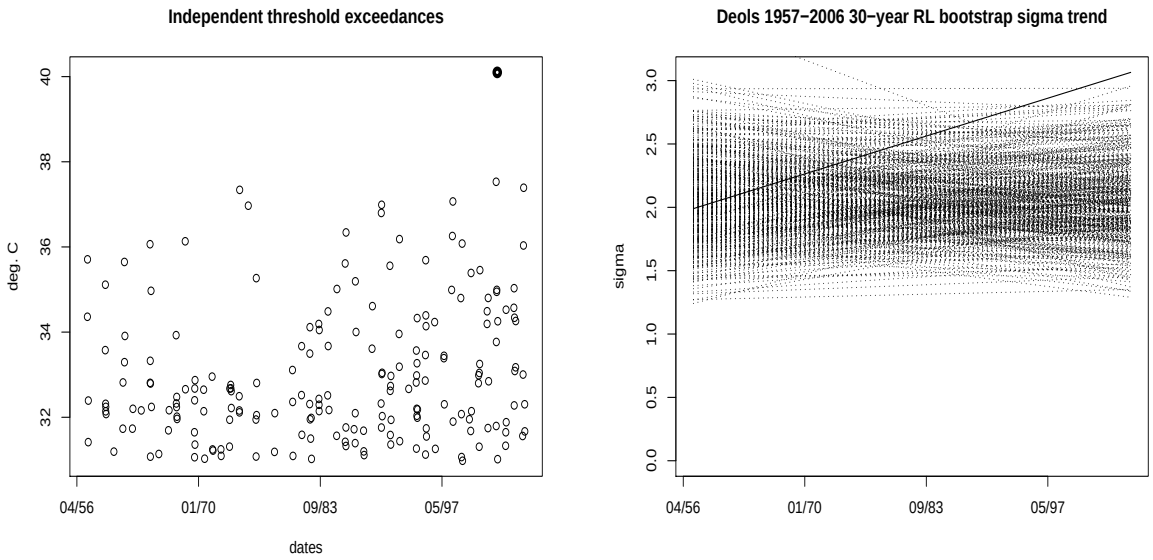


FIGURE 8.6 – Threshold exceedances over the 1957-2006 period with the 2003 maximum in bold (left panel) and the optimal scale parameter trend (solid line) together with the 500 bootstrap scale parameter trends (dotted lines) (right panel)

8.4 Example : use of the K hypothesis

As discussed in section 8.2.3, previous studies opened the possibility to use the link between the evolution of the mean and variance of the whole summer daily maximum temperature series and the trend in extremes to derive future extreme levels. The steps are first to check the K hypothesis for the Déols series and then to chose the way to evaluate mean and variance in the desired future period.

8.4.1 The K hypothesis

As stated before, the K hypothesis supposes that the extremes of the centered and normed series $Y(t)$ are stationary. It has to be tested in comparing the distance between the evaluation of the parameters of the extreme value distribution as constants and as time varying to a distribution of such distances obtained from a similar distribution known as stationary, as described in section 8.2.3. The results for the Déols series in summer in the case of block maxima and POT are illustrated in figure 8.7 and show that the K hypothesis can be accepted.

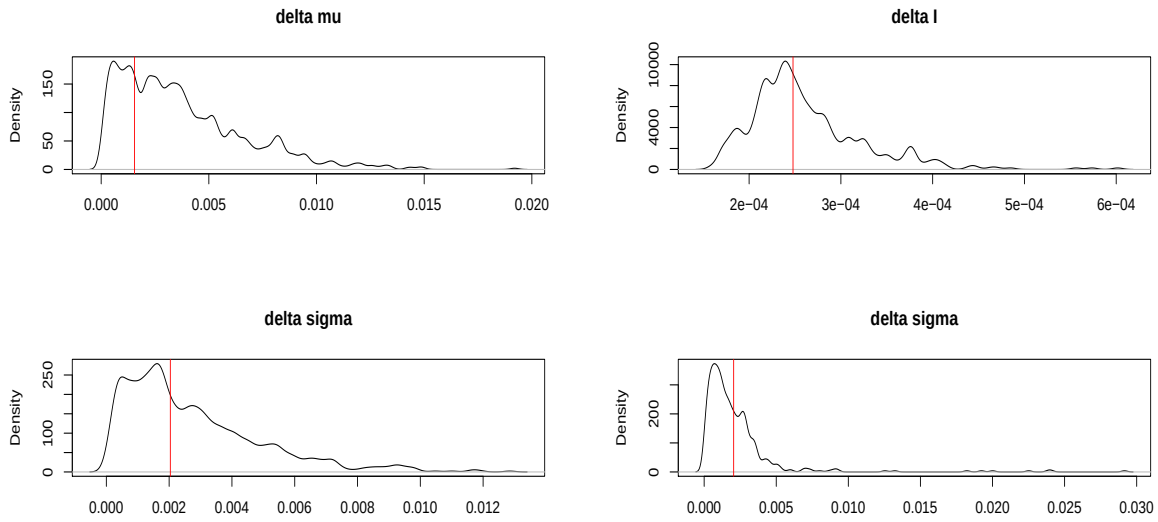


FIGURE 8.7 – distributions of the distances between constant and time varying parameters of the extreme value distributions in the case of constant distribution and the same distances for $Y(t)$ in summer in Déols (red lines), for GEV (location parameter μ and scale parameter σ ; left panel) and for POT (intensity I of the Poisson process and scale parameter σ of the GPD ; right panel)

8.4.2 Non stationary return levels

As stated before, the K hypothesis can first be used to compute return levels in the non stationary context, but in extrapolating the trends in mean and variance of the whole data series instead of those of the extreme value distribution parameters. In order to do so, it is first necessary to identify a parametric trend in the mean and variance of the observed summer temperatures.

In section 8.3.3, a method has been used to identify a break point in the trend in mean daily maximum temperature in summer in Déols. The same method is applied to the evolution of the variance of the same temperatures, and leads to the identification of a break point in the same 1956 year. These two parts linear trends will then be used to compute the return levels in extrapolating the second parts (figure 8.8).

The 30-, 50- and 100-year return levels are then computed following (9) and (10) and their 90% confidence intervals using the bootstrap technique described in section 8.2.3.3 above. The results are summarized in table 8.3. Both methods lead to comparable values.

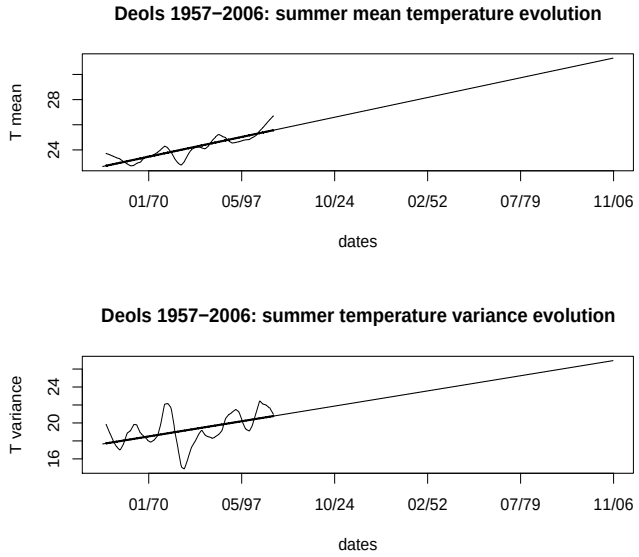


FIGURE 8.8 – Extrapolation of the second part of the 2-part linear trends in mean and variance of the daily maximum temperatures in summer in Déols, beginning 1956.

	30-year RL	50-year RL	100-year RL
GEV	40.3[38.9-42.0]	41.2[39.4-43.3]	43.3 [40.5-46.3]
POT	39.9[38.4-41.0]	40.8[38.9-42.2]	42.9 [40.0-45.4]

TABLE 8.3 – Non stationary return levels using the K hypothesis for GEV and POT, with their bootstrap 90% confidence intervals

8.4.3 Stationary return levels

Another proposed approach consists in supposing that over fixed periods of time, the return levels can be estimated in the stationary context. These periods have however to be long enough to allow reliable return level estimations, but not too long to be able to neglect the trends. The period length will be chosen here as 30 years, the length being suggested by the World Meteorological Organization to define climatology. Thus, the present day climate return levels will be computed for the last 30 years of the observation series, that is here for the 1977–2006 period. Then, two future 30-year fixed periods will be defined in accordance with the current outputs of the climate simulations : the periods 2021–2050 and 2046–2065. Over these future periods, the return levels will be computed from the stationary return levels of $Y(t)$, defined over the observed period, and the changes in mean and standard deviation between each future period and the present one. It

must be recalled here that the definition of the return level in this stationary context differs a bit from the one used in the previously set non stationary context : here, the A-year return level is the value reached or exceeded in average ones every A years over the studied period, whereas in the non stationary context, the same level corresponds to a value whose probability to be exceeded in the next A years equals 1, that is the value which is expected to be reached or exceeded only one day in the A future years.

Two ways of computing the changes in mean and standard deviation of the summer daily maximum temperatures will be tested : the extrapolation of the previously identified linear trends, and the use of climate model simulations.

8.4.3.1 Mean and variance trends extrapolation

The previously identified trends in mean and variance of the observed summer series are extrapolated, and the mean and standard deviations over the periods 2021–2050 and 2046–2065 respectively are calculated. The results are summarized in table 8.4. Thus the future return levels are estimated using the stationary return levels for the centred and normalized variable $Y(t)$, z_Y, and the future values of the mean (m_F) and standard deviation (s_F) as : $z_X = z_Y * s_F + m_F$ where z_X is the desired return level for the original series $X(t)$. The 90% confidence intervals are computed in using the delta-method, where the variances for the mean and variance are added to the variance of z in the computation of the covariance matrix. The results are shown in table 8.5. Not surprisingly, the two extreme value methods give similar results in that case.

	1977–2006	2021–2050	Increase	2046–2065	Increase
Mean	24.56	25.87	1.31	26.55	1.99
Standard deviation	4.50	4.68	0.18	4.79	0.29

TABLE 8.4 – Mean, standard deviation and their respective increases for the periods 1977–2006, 2021–2050 and 2046–2065 computed from the observed series of summer daily maximum temperatures in Déols (1977–2006) and from the extrapolation of the second part of the identified 2-part linear trends for the future 2021–2050 and 2046–2065 periods

8.4.3.2 Use of climate model results

In the framework of the European project ENSEMBLES, different Regional Climate Models (RCM), forced using different General Circulation Models (GCM) have been used to simulate future climate under different greenhouse gas evolution scenarios. We will use here 6 of these models, with a 25 km resolution and using the Intergovernmental Panel on Climate Change

		30-year RL	50-year RL	100-year RL
2021–2050	GEV	40.3 [39.5-41.1]	40.7 [39.8-41.6]	41.1 [40.1-42.1]
	POT	40.3 [39.1-41.5]	40.7 [39.4-41.9]	41.1 [39.7-42.5]
2046–2065	GEV	41.2 [40.3-42.0	41.5 [40.6-42.3]	42.0 [41.0-43.0]
	POT	41.1 [40.0-42.3]	41.5 [40.3-42.8]	42.0 [40.6-43.4]

TABLE 8.5 – 30-, 50- and 100-year Return Levels and their 90% confidence intervals estimated in the stationary context according to the K hypothesis for the 2 future periods, in extrapolating the mean and standard deviation trends derived from the observation series, with the GEV and POT methods

(IPCC) Special Report on Emission Scenarios (SRES) A1B scenario. The simulations cover the period 1950–2100 (1960–2100 for one of the models). In order to assess return levels for the two selected future periods, the mean and variance changes will be computed between the 2021–2050 and the 1977–2006 periods, and the 2046–2065 and 1977–2006 periods respectively, for the modeled series corresponding to the nearest grid point of the station Déols in France, and added to the observed 1977–2006 mean and standard deviation. Table 8.6 presents the simulations used and Table 8.7 summarizes the mean and standard deviation changes for the two periods. As expected, there is an important dispersion in the climate model results, as well for the change in mean as for the change in standard deviation, and the values previously obtained in extrapolating the observed trends lie in the range of the model values. For the 2021–2050 period, the change in mean given by the models is rather lower than the extrapolated one, except for DMI-ARPEGE. For the 2046–2065 period however, the change in mean computed from the extrapolated trend is rather low compared to model values. A comparison between the series of observed annual temperature maxima and modeled ones for the nearest grid points (not shown) have shown that the DMI-ARPEGE model exhibits a warm bias, DMI-ECHAM5, SMHI-BCM and SMHI-ECHAM5 a rather cold one and KNMI-RACMO and MPI-ECHAM5 give correct results. The use of these changes, together with the stationary return levels for $Y(t)$ as previously described, lead to the 30-, 50- and 100-year return levels (with their 90% confidence intervals) summarized in table 8.8 for both GEV and POT methods. Here again, the two extreme value methods lead to comparable results, and besides DMI-ECHAM5 which gives much higher levels (but has been seen to have a warm bias over the current period), all models give comparable orders of magnitude for the future return levels.

8.5 Discussion and perspectives

In this paper, we proposed different methodologies to estimate return levels with the statistical EVT in a non stationary context. Starting from the

Acronym	DMI-AEPEGE	DMI-ECHAM5	KNMI-RACMO	MPI-ECHAM5	SMHI-BCM	SMHI-ECHAM5
Institute	DMI	DMI	KNMI	MPI	SMHI	SHMI
Scenario	A1B	A1B	A1B	A1B	A1B	A1B
GCM	AEPEGE	ECHAM5	ECHAM5	ECHAM5	BCM	ECHAM5
RCM	HIRHAM	HIRHAM	RACMO	REMO	RCA	RCA
Resolution	25 km	25 km	25 km	25 km	25 km	25 km
Period	1950–2100	1950–2100	1950–2100	1950–2100	1950–2100	1950–2100

TABLE 8.6 – List of the ENSEMBLE project model simulations used

Model	2021–2050		2046–2065	
	Change in mean	Change in std	Change in mean	Change in std
DMI-ARPEGE	2.03	0.29	3.57	0.81
DMI-ECHAM5	0.47	0.01	1.47	-0.02
KNMI-RACMO	0.89	0.17	2.62	0.33
MPI-ECHAM5	0.70	0.09	2.20	1.16
SMHI-BCM	-0.03	-0.11	1.38	0.46
SMHI-ECHAM5	1.05	0.21	2.49	0.31

TABLE 8.7 – Changes in mean and standard deviation between each future period and the current 1977-2006 period given by the climate simulations of each model for the nearest grid point of Déols

identification of trends in the parameters of the extreme value distributions, algorithms are given to compute the return levels and evaluate the confidence interval taking the uncertainties of the trend into account, in using a bootstrap procedure, both for the block maxima and the POT methods. The return level must be re-defined here as the value expected to be reached or exceeded only once in a future period. The approach is illustrated on the example of the daily maximum summer temperature in Déols, center of France, over the 1901-2006 period. This example shows that the extrapolation of the optimal parametric trends as polynomials or CPL functions generally leads to similar results, but extrapolation sometimes gives unrealistic levels for long return periods, especially for quadratic evolutions. The use of the CPL functions is helpful to identify sub-periods over which the trend can be more reasonably extrapolated, but lets the sample begin with the lowest identified temperature levels.

Another way to select such sub-periods is to search for a break point in the linear evolutions of the mean and variance of the whole dataset and select the second part as the studied sub-period. In this non stationary context, the block maxima method generally leads to higher return levels than the POT one, as the influence of an increase in the location parameter for the GEV distribution is higher than that of an increase in the Poisson process intensity for the POT method. On the other hand, the POT method has

		30-year RL	50-year RL	100-year RL
2021–2050	GEV			
DMI-ARPEGE		41.4 [40.6-42.3]	41.8 [40.9-42.7]	42.2 [41.3-43.2]
DMI-ECHAM5		39.0 [38.2-39.8]	39.4 [38.5-40.2]	39.8 [38.8-40.7]
KNMI-RACMO		39.9 [39.1-40.7]	40.3 [39.7-41.7]	40.7 [39.7-41.7]
MPI-ECHAM5		39.5 [38.7-40.3]	39.9 [39.0-40.7]	40.3 [39.3-41.2]
SMHI-BCM		38.1 [37.3-38.9]	38.5 [37.6-39.3]	38.9 [37.9-39.8]
SMHI-ECHAM5		40.2 [39.4-41.0]	40.6 [39.7-41.4]	41.0 [40.0-42.0]
2021–2050	POT			
DMI-ARPEGE		41.4 [40.1-42.8]	41.8 [40.4-43.3]	42.3 [40.8-43.9]
DMI-ECHAM5		39.0 [37.7-40.3]	39.4 [38.0-40.8]	39.8 [38.3-41.4]
KNMI-RACMO		39.9 [38.6-41.2]	40.3 [38.9-41.7]	40.8 [39.3-42.3]
MPI-ECHAM5		39.5 [38.2-40.8]	39.9 [38.5-41.8]	40.3 [38.8-41.8]
SMHI-BCM		38.1 [36.8-39.4]	38.5 [37.1-39.9]	38.9 [37.4-40.4]
SMHI-ECHAM5		40.2 [38.9-41.5]	40.6 [39.2-42.0]	41.1 [39.6-42.6]
2046–2065	GEV			
DMI-ARPEGE		44.6 [43.7-45.5]	45.0 [44.1-46.0]	45.5 [44.4-46.5]
DMI-ECHAM5		39.9 [39.1-40.7]	40.3 [39.4-41.1]	40.7 [39.7-41.6]
KNMI-RACMO		42.2 [41.3-43.0]	42.5 [41.6-43.4]	43.0 [42.0-43.9]
MPI-ECHAM5		41.2 [40.4-42.0]	41.6 [40.7-42.5]	42.0 [41.0-43.0]
SMHI-BCM		41.3 [40.5-42.1]	41.7 [40.8-42.6]	42.1 [41.1-43.1]
SMHI-ECHAM5		42.0 [41.1-42.8]	42.3 [41.5-43.2]	42.8 [41.8-43.8]
2046–2065	POT			
DMI-ARPEGE		44.6 [43.3-45.9]	45.0 [43.6-46.4]	45.6 [44.1-47.1]
DMI-ECHAM5		39.9 [38.6-41.2]	40.3 [38.9-41.7]	40.7 [39.2-42.2]
KNMI-RACMO		42.2 [40.8-43.5]	42.6 [41.2-44.0]	43.0 [41.5-44.5]
MPI-ECHAM5		41.2 [39.9-42.5]	41.6 [40.2-43.0]	42.1 [40.6-43.6]
SMHI-BCM		41.3 [40.0-42.6]	41.7 [40.3-43.1]	42.2 [40.7-43.7]
SMHI-ECHAM5		42.0 [40.7-43.3]	42.4 [41.0-43.8]	42.9 [41.3-44.4]

TABLE 8.8 – 30-, 50- and 100-year Return Levels and their 90% confidence intervals computed in the stationary context according to the K hypothesis in using climate simulation results to calculate the mean and standard deviation changes in the future, with both GEV and POT methods

been shown to be very sensitive to very high values located at the end of the sample.

Another proposed way of estimating return levels in the non stationary context is through the use of the link between the evolution of extremes and the evolution of the mean and variance of the whole dataset. A method to carefully study this link has been detailed in Hoang et al. ([78]). The application of the described test to the summer temperature series of Déols showed that the so called K hypothesis, under which the extremes of the

centered and normalized data can be considered as constant, is accepted for both GEV and POT approaches. Then, again two ways can be used to estimate the future return levels. First, in keeping the non stationary return level definition, one can compute the value reached or exceeded once in a future period of time in extrapolating the trends in mean and standard deviation of the whole data series instead of those of the extreme value distribution parameters. On the other hand, the return levels for future fixed periods (2021–2050 and 2046–2065 in this paper) can be obtained from the stationary extremes of the centered and normalized quantity and the mean and standard deviation of the summer daily maximum temperature in the future period, in a stationary context. The return level is then the stationary one, that is the value reached or exceeded in average once over the period.

The computation of future mean and standard deviation is proposed in two ways : first, in extrapolating the identified trends in mean and variance of the observed series, and then in using climate model simulations. This has been tested for a selection of RCM simulations conducted in the framework of the European ENSEMBLES project.

Table 8.9 summarizes the different return levels obtained with the non stationary definition and table 8.10 those obtained with the stationary one for fixed future periods. Globally, the extrapolation of trends in the parameters of the extreme value distributions have limitations when levels associated to high return periods are concerned. As a matter of fact, the assumption that the trend will remain the same becomes very strong for long future periods. On the other hand, the trends in EVT are derived from quite limited samples regarding the total number of values in the series, and are therefore certainly less robust than trends identified from the whole sample. In that sense, the use of the K hypothesis seems more promising. The best way could then be to derive the mean and standard deviation changes from an as wide as possible range of climate simulations, to get a distribution of possible return levels for different future periods. However, further studies should be devoted to the analysis of the climate model behaviour regarding mean, and especially variance evolution at the local scale. This could then be used to weight the different simulations in the distribution of obtained return levels.

		30-year RL	50-year RL	100-year RL
Trends in EVT parameters				
GEV	1901–2006	42.0C [40.0 43.5]	*44.6C [41.2-46.3]*	*54.01C*
	1957–2006	40.8C [39.3 42.1]	42.2C [40.2 43.5]	*45.6C [40.8-46.5]*
POT	1901–2006	39.6C [38.8 40.9]	40.1C [39.3 41.4]	*40.7C*
	1965–2006	39.3C [37.5-41.3]	39.7C [37.7-42.2]	40.4C [37.8-44.5]
	1957–2006	41.3C [35.9 38.6]	42.9C [35.9 39.9]	46.9C [36.0 42.1]
Trends in total mean and				
GEV		40.3C [38.9-42.0]	41.2C [39.4-43.3]	43.3C [40.5-46.3]
POT		39.9C [38.4-41.0]	40.8C [38.9-42.2]	42.9C [40.0-45.4]

TABLE 8.9 – 30-, 50- and 100-year non stationary Return Levels computed in extrapolating the identified trends in the parameters of the extreme value distributions or the identified trend in mean and variance of the whole sample according to the K hypothesis

APPENDIX

Slow and regular non-stationarity and parameters extrapolation for return levels computation. (one part of a forthcoming paper with Dacunha-Castelle, D. and Malek, F.)

As we have seen, the computation of return levels from the observations of a time series requires an extrapolation of the parameters of the extreme models. The following result seems useful in order to choose an extrapolation as well as to precise in what conditions it can be applied, in particular when the sample of observations is long enough. We illustrate our results on a case of the Poisson intensity as a parameter of a POT model. The same thing could be done for other parameters of the POT or GEV model. The framework is the increase of the temperature. It is represented by functions which tends slowly to infinity as a signal of very weak frequency. The term "slowly" has a physical sense which can be translated as : the increase during a large period of observation is small with respect to the usual variability of these observations. To illustrate our purpose let us consider the following examples.

Let us consider the family of Poisson processes with linear intensities :

$$I(t) = a + bt \tag{8.12}$$

Let us consider the physical asymptotic $T \to \infty$, which means that we add data, and that we consider the non-stationarity shape of the parameters as persistent. This is quite different as the asymptotics considered for non-parametric statistics, in particular where the time is "sent" over $(0, 1)$, and the new observations are injected "between" the others, in order to tend to a dense set of dates.

		30-year RL	50-year RL	100-year RL
2021–2050	GEV			
Observed trends		40.3 [39.5-41.1]	40.7 [39.8-41.6]	41.1 [40.1-42.1]
DMI-ARPEGE		41.4 [40.6-42.3]	41.8 [40.9-42.7]	42.2 [41.3-43.2]
DMI-ECHAM5		39.0 [38.2-39.8]	39.4 [38.5-40.2]	39.8 [38.8-40.7]
KNMI-RACMO		39.9 [39.1-40.7]	40.3 [39.7-41.7]	40.7 [39.7-41.7]
MPI-ECHAM5		39.5 [38.7-40.3]	39.9 [39.0-40.7]	40.3 [39.3-41.2]
SMHI-BCM		38.1 [37.3-38.9]	38.5 [37.6-39.3]	38.9 [37.9-39.8]
SMHI-ECHAM5		40.2 [39.4-41.0]	40.6 [39.7-41.4]	41.0 [40.0-42.0]
2021–2050	POT			
Observed trends		40.3 [39.1-41.5]	40.7 [39.4-41.9]	41.1 [39.7-42.5]
DMI-ARPEGE		41.4 [40.1-42.8]	41.8 [40.4-43.3]	42.3 [40.8-43.9]
DMI-ECHAM5		39.0 [37.7-40.3]	39.4 [38.0-40.8]	39.8 [38.3-41.4]
KNMI-RACMO		39.9 [38.6-41.2]	40.3 [38.9-41.7]	40.8 [39.3-42.3]
MPI-ECHAM5		39.5 [38.2-40.8]	39.9 [38.5-41.8]	40.3 [38.8-41.8]
SMHI-BCM		38.1 [36.8-39.4]	38.5 [37.1-39.9]	38.9 [37.4-40.4]
SMHI-ECHAM5		40.2 [38.9-41.5]	40.6 [39.2-42.0]	41.1 [39.6-42.6]
2046–2065	GEV			
Observed trends		41.2 [40.3-42.0]	41.5 [40.6-42.3]	42.0 [41.0-43.0]
DMI-ARPEGE		44.6 [43.7-45.5]	45.0 [44.1-46.0]	45.5 [44.4-46.5]
DMI-ECHAM5		39.9 [39.1-40.7]	40.3 [39.4-41.1]	40.7 [39.7-41.6]
KNMI-RACMO		42.2 [41.3-43.0]	42.5 [41.6-43.4]	43.0 [42.0-43.9]
MPI-ECHAM5		41.2 [40.4-42.0]	41.6 [40.7-42.5]	42.0 [41.0-43.0]
SMHI-BCM		41.3 [40.5-42.1]	41.7 [40.8-42.6]	42.1 [41.1-43.1]
SMHI-ECHAM5		42.0 [41.1-42.8]	42.3 [41.5-43.2]	42.8 [41.8-43.8]
2046–2065	POT			
Observed trends		41.1 [40.0-42.3]	41.5 [40.3-42.8]	42.0 [40.6-43.4]
DMI-ARPEGE		44.6 [43.3-45.9]	45.0 [43.6-46.4]	45.6 [44.1-47.1]
DMI-ECHAM5		39.9 [38.6-41.2]	40.3 [38.9-41.7]	40.7 [39.2-42.2]
KNMI-RACMO		42.2 [40.8-43.5]	42.6 [41.2-44.0]	43.0 [41.5-44.5]
MPI-ECHAM5		41.2 [39.9-42.5]	41.6 [40.2-43.0]	42.1 [40.6-43.6]
SMHI-BCM		41.3 [40.0-42.6]	41.7 [40.3-43.1]	42.2 [40.7-43.7]
SMHI-ECHAM5		42.0 [40.7-43.3]	42.4 [41.0-43.8]	42.9 [41.3-44.4]

TABLE 8.10 – 30-, 50- and 100-year stationary Return Levels computed for two future periods in evaluating the changes in mean and variance of summer daily maximum temperature in extrapolating the observed trends or in using climate model simulations, according to the K hypothesis

Suppose that the true value $b^0 > 0$, then the maximum likelihood estimators (mle) $\widehat{a}$, $\widehat{b}$ are consistent and the pair ($\sqrt{T}(\widehat{b} - b^0)$, $\sqrt{\log T}(\widehat{a} - a^0)$) converges in distribution to a bivariate Gaussian centered with covariance Γ calculated below.

Let us now consider the quadratic case :

$I(t) = a + bt + ct^2$ where $I(t) \geq 0$ for $t \in [1, T]$ and the true $c_0 > 0$ then the mle $\widehat{c}$ and other estimator $\widetilde{b}$ are consistent and the pair $(T^{3/2}(\widehat{c} - c^0)$, $T^{1/2}(\widetilde{b} - b^0))$ is asymptotically Gaussian but it does not exist any consistent estimator $\widehat{a}$ for a.

Suppose that in the two cases, a is considered as the value of the intensity for the date 0. We see that a is very slowly convergent or not convergent and in some sense forgotten. The error on $\hat{a}$ is of course translated on all the calculations of the return levels, and, moreover, the result of the following theorem proves that the determinant parameter for a long observation is the one with the highest degree. This is more or less almost evident. This can imply the following thing : we try to parameter the "rate" of the increase, more than other details of the observations. For instance, it should be better for us to work with models as $t^\alpha \ln t^\beta$ than polynomials.

Now we pass to the mathematical framework.

The usual technique using likelihoods and Kullback information has to be modified in order to change the contrast theory as it is exposed, for instance, in Dacunha-Castelle et Duflo ([28], chapter 3). In fact the asymptotic information does not allow to identify the model but only some of its parameters. The model is asymptotically non-identifiable.

Let the intensity of the Poisson process P be of the form :

$$I(t) = \sum_{k=1}^{K} \lambda_k f_k(t) \tag{8.13}$$

with $\int_0^T f_K(t) \to \infty$ and $\frac{f_k(t)}{f_K(t)} \to 0$ for $t \to \infty$ and $k = 1, ...K - 1$.

λ_K associated to the largest parameter $f_K(t)$ has always a consistent mle estimator. For other parameters, the existence of convergent estimators depends on the behavior of $\Psi_k(T) = \int_0^T \frac{f_k^2(t)}{f_K(t)} dt$, which can be considered as the residual information remaining to estimate λ_k, once λ_K is estimated.

Under very weak regularity hypothesis for the functions $f_k(t)$, the condition $\Psi_k(T) \to \infty$ is necessary and sufficient to assure the existence of a consistent estimator $\widehat{\lambda_k}$. Once λ_K is estimated, we use, for every other parameters, a profiled contrast depending on $\widehat{\lambda_K}$. It always leads to a least square procedure. The speed of estimation is $\sqrt{\Psi_k(T)}$, the estimator being asymptotically efficient. The set of consistent estimators $H \leq K$ has for limit distribution a multidimensional Gaussian.

We shall use the following notations and results :

Let $I(h) = \int_0^T h(u) I(u) du$, I be the intensity of the Poisson process $P(t)$, h be a function such that $\int_0^\infty |h(t)| \, I(t) dt < \infty$.

$$P(h) = \int_0^T h(u) dP(u) = \sum_{i=1}^{P(T)} h(t_i)$$

where $t_1, ..., t_{P(T)}$ is the sequence of $P(t)$ jumps for $0 < t \leq T$.

Elemental properties of stochastic integral imply :

$$E(P(h)) = I(h) \ ; \ Var(P(h)) = I(h^2) \ ; \ Cov(P(h), P(g)) = I(hg)$$

Let a set of intensities of the linear form $I_\theta = \sum_{k=1}^{K} \lambda_k f_k(t)$ be with $\theta \in \Theta$. The properties of the functions f_k will be detailed below.

Let θ^0 be the true value of θ and $L_T(\theta^0, \theta)$ be the log likelihood of $P(t)$ with intensity I_θ with respect to the Poisson process whose intensity is I_{θ^0}, then (cf Dacunha-Castelle and Duflo ,[28])

$$L_T(\theta^0, \theta) = P(\log(I_{\theta^0}) - \log(I_\theta)) + \int_0^T (I_\theta(u) - I_{\theta^0}(u))du$$

H) We suppose I_θ twice continuously derivable in θ.

$$grad(L_T(\theta^0, \theta)) = P(grad(\log[I_\theta])) - \int_0^T grad(I_\theta(u))du$$

so $E_\theta\left[grad(L_T(\theta^0, \theta))\right] = 0$

The Fisher information matrix for $\theta = \theta^0$ is

$$\begin{aligned} J_{\theta^0}(T) =& E_{\theta^0}\left[grad(L_T(\theta^0) \otimes grad(L_T(\theta^0))\right] \\ =& - E_{\theta^0}[\bigtriangledown L_T(\theta^0)] \\ =& (\int_0^T \frac{\partial I_{\theta^0}(u)}{\partial \theta_i} \frac{\partial I_{\theta^0}(u)}{\partial \theta_j} \frac{1}{I_{\theta^0}(u)} du_{i=1,..K \ , \ j=1,...K}) \end{aligned}$$

Kullback information is given by :

$$K_T(\theta^0, \theta) = \int_0^T (\log(I_{\theta^0}(u)) - \log(I_\theta(u)))I_{\theta^0}(u)du + \int_0^T (I_\theta(u) - I_{\theta^0}(u))du$$

Theorem 8.5.1 *Let the family of Poisson process with intensity* $I_\theta(t) = \sum_{k=1}^{K} \lambda_k f_k(t)$, $\theta = \{\lambda_k, \ k = 1, ...K\} \in \Theta$ *be open and bounded in* $\mathbb{R}^K$.

Suppose the following hypothesis are verified :

h1- *For every* $k = 1, ...K$, f_k *is bounded on every compact set, denote* $F_k(T) = \int_0^T f_k(t)dt$

h2- $\int_0^t f_K(u)du \to \infty$

h3- *Let* $\varphi(t) = \frac{f_k(t)}{f_K(t)}$ *then* φ *is bounded on* $\mathbb{R}^+$ *and* $\lim_{t\to\infty} \varphi(t) = 0$.

These conditions imply that $\lambda_K > 0$.

Then :

c1) $\sqrt{F_K(T)}(\widehat{\lambda_K} - \lambda_K^0)$ *converges to* $\mathcal{N}(0, \lambda_K^0)$.

where $\widehat{\lambda_K}$ *is the mle estimator .*

c2) *for $k < K$, there exists estimators $\widehat{\lambda_k}$ of λ_k convergent, asymptotically Gaussian and efficient, if and only if the condition $\Psi_k(T) = \int_0^T \frac{f_k^2(t)}{f_K(t)}dt \to \infty$ is verified . In this case $\sqrt{\Psi_k(T)}\ (\widehat{\lambda_k} - \lambda_k^0) \ \stackrel{w}{\to}\ \mathcal{N}(0, (\lambda_k^0)^2)$.*

If the condition on $\Psi_k(T)$ is not verified, there does not exist any consistent estimator.

c3) *Let H be the set of $k < K$ such that $\Psi_k(T) = \int_0^T \frac{f_k^2(t)}{f_K(t)}dt \to \infty$. then*

$$\left\{ \sqrt{\Psi_k(T)}\ (\widehat{\lambda_k} - \lambda_k^0), k \in H\ ; \sqrt{F_K(T)}\ (\widehat{\lambda_K} - \lambda_K^0) \right\}$$

has a limit Gaussian law (centred) of the covariance matrix :

$$\Gamma = (\gamma_{ij})_{i,j} \ \text{ with } \gamma_{ij} = \begin{cases} \gamma_{ii} = \lambda_K^0 & \text{if} \quad 1 \le i \le K \\ \gamma_{iK} = \lim\limits_{T\to\infty} \frac{F_i^2(T)}{\Psi_i(T)F_K(T)} & \text{if} \quad 1 \le i \le K-1 \\ \gamma_{ij} = \lim\limits_{T\to\infty} \frac{\int_0^T \frac{f_i(t)f_j(t)}{f_K(t)}dt}{\Psi_i(T)\Psi_j(T)} & \text{if} \ \ 1 \le i < j \le K-1 \end{cases}$$

if and only if Γ has a limit.

If not, we only have that the distributions are tight, and that all the limits of convergent subsequences are Gaussian.

Proof

Consider first the case **K=2**

$I_\theta(t) = a_1 f_1(t) + \alpha_2 f_2(t)$ and $\theta = (a_1, a_2)$.

The log-likelihood is

$$L_T(a_1, a_2) = \int_0^T \log(\frac{a_1^0 f_1(t) + a_2^0 f_2(t)}{a_1 f_1(t) + a_2 f_2(t)})dP + F_1(T)(a_1 - a_1^0) + F_2(T)(a_2 - a_2^0)$$

and

$$K_T(a_2^0, a_2) = a_2^0 \log(\frac{a_2^0}{a_2}) + a_2 - a_2^0$$

Let $b = \frac{a_1}{a_2}$ then

$$L_T(a_1, a_2) = \int_0^T \log(\frac{a_2^0}{a_2}) + \log(\frac{1 + b^0\varphi(t)}{1 + b\varphi(t)})dP + F_1(T)(a_1 - a_1^0) + F_2(T)(a_2 - a_2^0)$$

$a_2 > 0$ and b is bounded, so $\log(\frac{1+b^0\varphi(t)}{1+b\varphi(t)}) = O(\varphi(t))$ and tends towards 0 from the hypothesis.

Let $K_T(\theta_1^0, \theta) = E(L_T((\theta_1^0, \theta)))$ be the Kullback information of the Poisson processes on $[0, T]$, whose intensities are I_{θ_0} and I_θ.

$$\begin{aligned} \lim_{T\to\infty} \frac{1}{F_2(T)} K_T(a_1, a_2) = \lim_{T\to\infty} [&(a_1 - a_1^0)\frac{F_1(T)}{F_2(T)} + \\ &\frac{1}{F_2(T)} \int_0^T \log(\frac{1 + b^0\varphi(t)}{1 + b\varphi(t)})(a_1^0 f_1(t) + a_2^0 f_2(t))dt + K_T(a_2^0, a_2)] \\ &= K(a_2^0, a_2) \end{aligned}$$

Now we have

$$Var(L_T(a_1,a_2)) = \int_0^T \log^2(\frac{1+b^0\varphi(t)}{1+b\varphi(t)})(a_1^0 f_1(t) + a_2^0 f_2(t))dt$$

so

$$\frac{1}{F_2^2(T)} Var(L_T(a_1,a_2)) = O(\frac{1}{F_2(T)})$$

thus

$$\frac{1}{F_2(T)} L_T(a_1,a_2) \xrightarrow{P} K(a_2^0,a_2)$$

We shall use now an extension to non asymptotically identifiable models of the theorem on the sequence of contrast functions to prove that the mle estimator $\widehat{a_2}$ converges to a_2^0. Let $\widehat{\alpha_i}^T$ be the mle estimators for T fixed, $i = 1, 2$.

Choose as contrast family for the parameter a_2, the profiled likelihood $\frac{1}{F_2(T)} L_T(\widehat{\alpha_1}^T, a_2)$.

It converges to $K(a_2^0, a_2)$, $K(a_2^0, a_2) = 0$ if and only if $a_2 = a_2^0$, thus $\widehat{a_2}^T$ converges to a_2.

In general, we shall see $\widehat{a_1}^T$ is not consistent.

Let

$$\overline{L}_T(a_2^0, a_1) = \int_0^T \log(\frac{1+b^0\varphi(t)}{1+b\varphi(t)})dP(t) + (a_1 - a_1^0)F_1(T)$$

and

$$E(\overline{L}_T(a_2^0,a_1)) = \int_0^T \log(\frac{a_2^0 + a_1^0\varphi(t)}{a_2^0 + a_1\varphi(t)})(a_1^0 f_1(t) + a_2^0 f_2(t))dt + (a_1 - a_1^0)F_1(T)$$

For $\varphi(t) \to 0$,

$$\log(\frac{a_2^0 + a_1^0\varphi(t)}{a_2^0 + a_1\varphi(t)}) = \varphi(t)(a_1^0 - a_1)(a_2^0)^{-1} - \frac{1}{2}\varphi^2(t)((a_1^0)^2 - a_1^2)(a_2^0)^{-2} + o(\varphi^2(t))$$

Let

$$\Psi(T) = \int_0^T \varphi^2(t)dt = \int_0^T \frac{f_1^2(t)}{f_2(t)}dt$$

$$E(\overline{L}_T(a_2^0,a_1)) = ((a_1^0 - a_1)(a_2^0)^{-1}a_1^0\Psi(T) - \frac{1}{2}((a_1^0)^2 - a_1^2)(a_2^0)^{-1}\Psi(T))(1+o(1))$$

If $\Psi(T) = \int_0^T \varphi^2(t) f_2(t)dt \to \infty$ then $\Psi(T)^{-1}E(\overline{L}_T(a_2^0,a_1)) \to \frac{1}{2a_2^0}(a_1 - a_1^0)^2$

Let us study the limit of $\sqrt{F_2(T)}(\widehat{a_2} - a_2^0)$.

Note ∂_β^n for $\frac{\partial^n}{\partial\beta^n}$.

$$\partial_{a_1}^n \partial_{a_2}^m L_T(a_1,a_2) = \int (-1)^{n+m} \frac{f_1^n(t) f_2^m(t)dt}{(a_1 f_1(t) + a_2 f_2(t))^{n+m}} dP(t) + 1_{n=1}F_1(T) + 1_{m=1}F_2(T)$$

For $n+m>1$, because we can choose α_2 in an open neighborhood of a_2^0, for T large enough so that :

$$E\left|\partial_{a_1}^n\partial_{a_2}^m L_T(a_1,a_2)\right| \leq C\int_0^T \frac{|f_1(t)|^{2n}}{f_2(t)^{2n-1}}dt$$

$$E\left|\partial_{a_1}^n\partial_{a_2}^m L_T(a_1,a_2)\right|^2 \leq C(\int_0^T \frac{|f_1(t)|^{n}}{f_2(t)^{n-1}}dt)^2 + \int_0^T \frac{|f_1(t)|^{2n}}{f_2(t)^{2n-1}}dt$$

for some $C>0$.
thus $E\left|\partial_{a_2}^3(L_T(a_1^0,a_{2**}))\right| \leq 2CF_2(T)$.
and $E(\partial_{a_2}\partial_{a_1}(L_T(a_{1*},a_2^0))^2 \leq CF_1(T)^2+\Psi(T)$.
and

$$\begin{aligned}\partial_{a_2}(L_T(\widehat{a_1}^T,\widehat{a_2}^T)) &= \partial_{a_2}(L_T(a_1^0,a_2^0)) + (\widehat{a_2}^T-a_2^0)\partial_{a_2}^2(L_T(a_1^0,a_2^0))\\ &\quad+\frac{1}{2}(a_{2*}-a_2^0)^2\partial_{a_2}^3(L_T(a_1^0,a_{2**})) + (\widehat{a_1}^T-a_1^0)(\widehat{a_2}^T-a_2^0)\partial_{a_2}^2\partial_{a_1}(L_T(a_{1*},a_2^0))\\ &= 0\end{aligned}$$

for a_{2*} and $a_{2**}\in(a_2^0,\widehat{a_2}^T)$, $a_{1*}\in(a_1^0,\widehat{a_1}^T)$.
$\partial_{a_2}(L_T(a_1^0,a_2^0)) = \int_0^T \frac{f_2(t)}{a_1f_1(t)+a_2^0f_2(t)}dP - F_2(T)$ is a square integrable martingale for $Var(\partial_{a_2}(L_T(a_1^0,a_2^0))) = F_2(T)(\frac{1}{a_2^0}+o(1))$

Therefore, the central limit theorem for martingales gives :

$$\frac{1}{\sqrt{F_2(T)}}\partial_{a_2}(L_T(a_1^0,a_2^0)) \overset{law}{\to} N(0,\frac{1}{a_2^0}).$$

The strong law for martingales gives now :

$$\frac{1}{F_2(T)}\partial_{a_2}^2(L_T(a_1^0,a_2^0)) \to \frac{1}{a_2^0}$$

$$\begin{aligned}\sqrt{F_2(T)}(\widehat{a_2}^T-a_2^0) &= -\frac{\frac{1}{\sqrt{F_2(T)}}\partial_{a_2}(L_T(a_1^0,a_2^0))}{\frac{1}{F_2(T)}\partial_{a_2}^2(L_T(a_1^0,a_2^0))}\\ &\quad+\frac{\frac{1}{2}(a_{2*}-a_2^0)^2\partial_{a_2}^3(L_T(a_1^0,a_{2**})) + (\widehat{a_1}^T-a_1^0)(\widehat{a_2}^T-a_2^0)\partial_{a_2}^2\partial_{a_1}(L_T(a_{1*},a_2^0))}{\frac{1}{F_2(T)}\partial_{a_2}^2(L_T(a_1^0,a_2^0))}\end{aligned}$$

$$\left|\frac{\frac{1}{2}(a_{2*}-a_2^0)^2\partial_{a_2}^3(L_T(a_1^0,a_{2**})) + (\widehat{a_1}^T-a_1^0)(\widehat{a_2}^T-a_2^0)\partial_{a_2}^2\partial_{a_1}(L_T(a_{1*},a_2^0))}{\frac{1}{F_2(T)}\partial_{a_2}^2(L_T(a_1^0,a_2^0))}\right|$$
$$= o_P((\widehat{a_2}^T-a_2^0)\frac{F_1(T)}{\sqrt{F_2(T)}}) + o_P((a_{2*}-a_2^0)^2\sqrt{F_2(T)})$$

$$F_1(T) = \int_0^T \frac{f_1(t)}{\sqrt{f_2(t)}}\sqrt{f_2(t)}dt < \sqrt{\Psi(T)}\sqrt{F_2(T)}$$

Thus $(\widehat{a_2}^T - a_2^0)\frac{F_1(T)}{\sqrt{F_2(T)}} = O((\widehat{a_2}^T - a_2^0)\sqrt{\Psi(T)}) = o(\sqrt{F_2(T)}(\widehat{a_2}^T - a_2^0))$

Then $\sqrt{F_2(T)}(\widehat{a_2} - a_2^0) \overset{loi}{\to} N(0, a_2^0)$.

Let us now study the convergence in distribution of the pair $(\widehat{a_1}^T - a_1^0, \widehat{a_2} - a_2^0)$, as $\Psi(T) \to \infty$:

$$\begin{aligned}\partial_{a_1}(L_T(\widehat{a_1}^T, \widehat{a_2}^T)) &= \partial_{a_1}(L_T(a_1^0, a_2^0)) + (\widehat{a_1}^T - a_1^0)\partial^2_{a_1}(L_T(a_1^0, a_2^0)) \\ &+ \frac{1}{2}(a_{1*} - a_1^0)^2\partial^3_{a_1}(L_T(a_1^0, \widehat{a_2}^T)) + (\widehat{a_1}^T - a_1^0)(\widehat{a_2}^T - a_2^0)\partial_{a_2}\partial^2_{a_1}(L_T(a_{1*}, a_{2*}) \\ &= 0\end{aligned}$$

$E(\left|\partial^3_{a_1}(L_T(a_1^0, \widehat{a_2}))\right|) = 0(\Psi(T))$
$E(\left|\partial_{a_2}\partial^2_{a_1}(L_T(a_{1*}, a_{2*}))\right| = 0(\Psi(T))$
So as in the case of $(\widehat{a_2} - a_2^0)$,

$$\sqrt{\Psi(T)}(\widehat{a_1}^T - a_1^0) \overset{law}{\to} N(0, a_2^0)$$

But $F_1^{-1}(T)(\partial_{a_2}\partial_{a_1}(L_T(a_{1*}, a_{2*})) \to (a_2^0)^{-1}$ and from $Cov(P(h), P(k)) = I(hk)$ we get

$$\lim_{T\to\infty} cov((\widehat{a_1}^T - a_1^0, \widehat{a_2} - a_2^0) = \lim_{T\to\infty} \frac{F_1(T)}{\sqrt{\Psi(T)}\sqrt{F_2(T)}}a_2^0$$

if this limit exists.

This correlation is in general different from 0 for instance if $f_1 = t^a$ and $f_2 = t^b$, $0 < a < b$. Moreover for monotone functions with very long oscillations, the limit of the covariance can not exist.

Suppose now $\Psi_k(T) < C < \infty$ when $T \to \infty$. For bounded λ_k, every estimator of λ_k has a finite variance. Cramer-Rao inequality shows that $\Psi_k^{-1}(T)$ is a lower bound of this variance and so the estimator can not be consistent.

Let us now consider $K > 2$. $I_\theta(t) = \sum_{k=1}^K \lambda_k f_k(t)$. The proof made for $K = 2$ can be still applied for the behaviors of $\widehat{\lambda_K}$.

The next step consists to use separately for every $k < K$ the contrast defined by :

$$\widetilde{L}(\lambda_k) = \int_0^T \log(\frac{\lambda_k^0 f_k(t) + \widehat{\lambda_K} f_K(t)}{\lambda_k f_k(t) + \widehat{\lambda_K} f_K(t)})dP_T(t) + (\lambda_k - \lambda_k^0)F_k(T)$$

$$E[\widetilde{L}(\lambda_k, \lambda_k^0)] = \int((\frac{\lambda_k^0 f_k(t)}{\lambda_K^0 f_K(t)} - \frac{\lambda_k f_k(t)}{\lambda_K^0 f_K(t)}) - \frac{1}{2}(\frac{\lambda_k^0 f_k(t)}{\lambda_K^0 f_K(t)})^2 + (\frac{\lambda_k^0 f_k(t)}{\lambda_K^0 f_K(t)})^2(1 + o(1)))$$

As previously, we get :

$\Psi_k^{-1}(T)E[\widetilde{L}(\lambda_k, \lambda_k^0)] \to \frac{1}{2\lambda_k^0}(\lambda_k^0 - \lambda_k)^2$ if $\Psi_k(T) \to \infty$
and
$\Psi_k^{-1}(T)(\widetilde{L}(\lambda_k, \lambda_k^0) - \widetilde{L}(\lambda_k, \widehat{\lambda_k^0})) = o_P(1)$.

So we obtain the convergence part of the theorem. Its Gaussian part is obtained through the convergence in distribution of linear combinations of the type $\sum_{k=1}^{H} a_k \sqrt{\Psi_k(T)}\ (\widehat{\lambda_k} - \lambda_k^0)$, which can be obviously treated as the pair $(\sqrt{\Psi_k(T)}\ (\widehat{\lambda_k} - \lambda^0), \sqrt{\Psi_k(T)}\ (\widehat{\lambda_k} - \lambda_k^0))$.

□

Chapitre 9

Seasonality dynamics with warming effect

9.1 The seasonality in statistical study

In statistics, a time series exhibits cyclic or periodic fluctuations known as "seasonality". The concept of seasonality shares a feature with many other concept used in daily language. For example when we work with the daily temperature, there exists daily and annual seasonalities. For a regular deterministic cyclic variation, we can model the seasonality simply by a periodic or a trigonometric function. For a stochastic variation, it seems natural to consider the autocorrelation structure or the spectrum of the series. We can quote important contributions on the subject : Nerlove([102]) and Granger([58]) which are popular in econometric. Here we discuss some main problems usually met with model adjustment of the seasonality for temperature series.

Seasonality plays an important role in climate studies. When working with temperatures, an accurate statistical model of the variability of the daily temperature, or even of hourly temperature (in the case where we have this kind of datasets), are needed.

The daily temperatures : daily mean, maximum or minimum or the temperature at some fixed hour during the whole year present a strong seasonal component that we need to remove in order to obtain a more homogeneous series. The literature shows that it is not easy to completely detect the seasonality from the temperature. In fact, the works done so far are usually based on the assumption that the observed time series, or their logarithm, can be meaningfully divided into at least two components : a deterministic one (a trend, a cycle or a seasonal component) and irregular components. However, this is too strong an assumption.

Many current published models are shown to be inaccurate for stations that show strong seasonality in the probability distribution and autocorrelation structure of temperature (see see Jewson & Caballero, [83] for bibliographies and discussion). A bad seasonal adjustment can also affect the results of the estimation of models in a negative way. In particular, dynamic relationships can be distorted by a seasonal overadjustment (see Hylleberg([80])). As a matter of fact, the seasonality must be adjusted with caution.

It is well-known that for temperature series, the seasonality exists not only in the mean, but also in the variance, and in fact in the whole distribution. However, even after removing the seasonality from the mean and the variance, the usual stationarity hypothesis on the residual series can still not be acceptable. The seasonality still remains, for example in the autocorrelation function. We can see it in many works on the weather derivatives such as Cao and Wei ([16]),Campbell([14]), Jewson & Caballero ([83]). The assumption on the stationarity can severely underestimate the characteristics in certain seasons and overestimate them otherwise. Even more seriously, even when we take into account this fact and remove the seasonality from

the residuals, this "famous" seasonality still exists (see Jewson & Caballero ([83]) for more details).

It is evident that the modeling of daily temperatures based on simpler models has certain potential advantages. Such modeling, however, can not catch the whole richness and complexity of the climate variability, and in particular, the seasonality.

Facing this situation, the question : "Does the seasonality change along with time ?" does not seem illogical. This question is for today pertinent under the greenhouse effect. To respond to this question, we come to the next section.

9.2 Transformation of the seasonality under global growing of the temperature

Climate change involves not only changes in the mean, but also in the characteristics of the corresponding seasonal cycle. The seasonality profile is really changing, summers are "longer" and winters are "shorter", if they are defined by the presence of high, or low, temperatures. In the paper of Frich et al.([56]), a study on the temperature series of Europe and Australia in 1946-1999 period pointed out a reduction in the number of frost days and a significant lengthening of the thermal growing season.

Recently, there are some significant studies on the variability of seasonality of the temperature. Among them, there are the works of The Climate Research Committee (USA) ([23]), Pezzulli et al.([117]),Wu et al.([152]) and Barbosa ([8]). In the essay of Diaz and Brasley in [23], the seasonal change for the decaded-scale in temperature anomalies (after removing the average value of each months) were shown. Pezzulli et al. used Census Method II Seasonal Adjustment Program (X-11) which is usually met in econometrics to model trends and varying seasonalities. Wu et al. proposed "modulated annual cycle" to catch of the dynamics of the seasonality. Differently, Barbosa modeled the daily mean temperature series by autoregressive family with high order p. All of them confirmed the change in the amplitude of the annual cycle or seasonal cycle.

Here, a simple study will be carried out on some series of temperature to get a better understanding of the transformation of the seasonality with time. The study is applied to detrended series. In order to keep the original values of temperature when studying the seasonality, the trend, which is removed, is estimated from the centred series (subtract the average value of the series). This trend is estimated by loess (the estimation method is describe in Chapter 3). Three kinds of temperature will be considered at different stations : daily mean, maximum or minimum temperature. The method used for the study is to choose a window of N-year size with N rather large. Then consecutive windows are obtain by shifting the first window each Δ

year until the end of the series. For each window, we estimate the seasonality of its corresponding sub-series by fitting a trigonometric function :

$$\theta_0 + \sum_{i=1}^{k} \left(\theta_{i,1} \cos \frac{2\pi t}{365} + \theta_{i,2} \sin \frac{2\pi t}{365} \right)$$

where t is the dates, k is the number of trigonometric terms. k is taken rather high to get flexible estimates. We take here k=4.

We consider here three different stations which different length of observations. The temperature series in Strasbourg consists of 57 years (1949–2005), while the one in Déols contains 106 years(1901–2006). The temperature series in Paris-Montsouris is longest with 131-year observations (1873–2003).

Three different types of temperature are studied : daily mean temperature in Strasbourg, daily maximum temperature in Déols and daily minimum temperature in Montsouris. For the one of Strasbourg, N will be taken as 30 years and Δ is taken as 4 years. For the one of Déols, N, Δ are respectively equal to 50 years and 4 years. While for the one of Montsouris, $N = 50$ years and Δ=10 years are considered. The results are shown in Figure 9.1, 9.2 and 9.3.

Discussion

- In general, after observing these three figures, we can conclude that the seasonality change slowly with time. Then with a longer observed time, this change can be better seen. The change of the seasonality seems clearest in Montsouris.

- The seasonality in Strasbourg changes slowly in the last fifty years. Most remarkable are the changes in summers. The more we advance in time, the higher the peak of the summer. This means the global warming is reflected in the seasonality.

- The seasonality also changes slowly with time. However for the last fifty year, the change is remarkable. Comparing with the seasonality in the first fifty years, the winter is really warmer.

- The seasonality of the minimum temperature in Montsouris remarkably changes. The average difference between the first and the last seasonalities is nearly 2.5C, which is really significant. It is colder with time in Montsouris.

- In conclusion, the seasonality is changing in a complicated way under the greenhouse effect. The difference of temperatures in a day or between seasons is larger. More precisely, in a day, the maximum temperature tends to higher and the minimum one tends to smaller. Then in modeling the seasonality, a model which does not take into account this changing seasonality can not give a complete estimation for the seasonality. The result of this is that the seasonality still remains in the residuals.

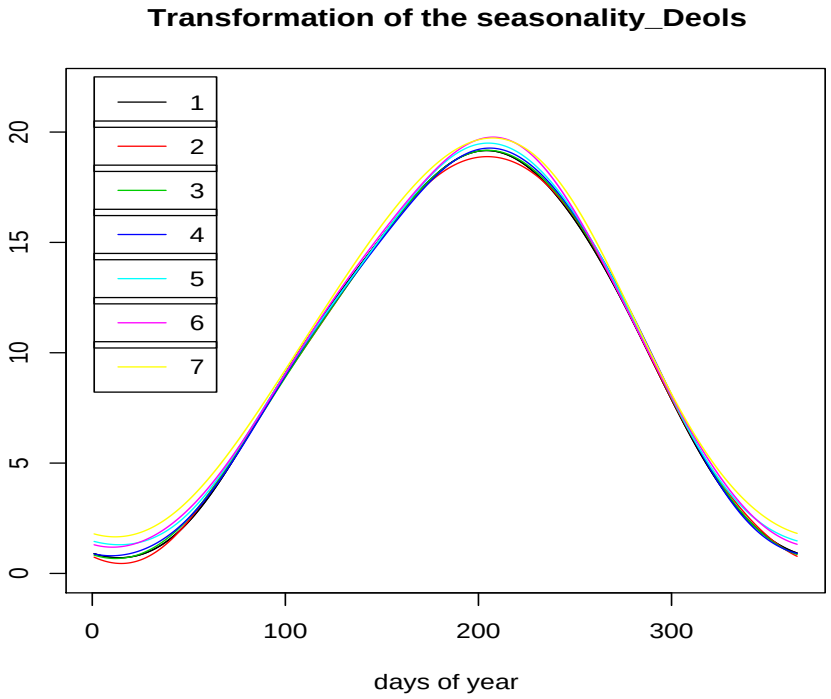


FIGURE 9.1 – Transformation of the seasonality in Strasbourg with size of window N=30 years. The numbers correspond to the number of windows

9.3 How to detect the seasonality from the time series ?

We all know that the long term traditional annual cycles, which are taken to be exact repeating of themselves years after years, are no more appropriate. For these modeling, an implicit a priori assumption is that temperatures produce a constant seasonal cycle. This assumption is nowadays questionable in the face of a nonlinear climate system under the external forcing.

For this reason, a better way is to define a seasonal cycle which can change with time. In order to take into account the dynamics of the seasonality, a reasonable way is to use time-varying cycles in climate series, allowing to retrieve fluctuations in the amplitude and phase of the periodic components and to assess their statistical significance. The argument is easy to understand, but it is very difficult to find a simple model which can adjust this situation. The reason is that estimating the seasonality without mixing up long-term changes in the mean and low-frequency variations in the seasonal pattern, is not simple. We need a model which captures the slow

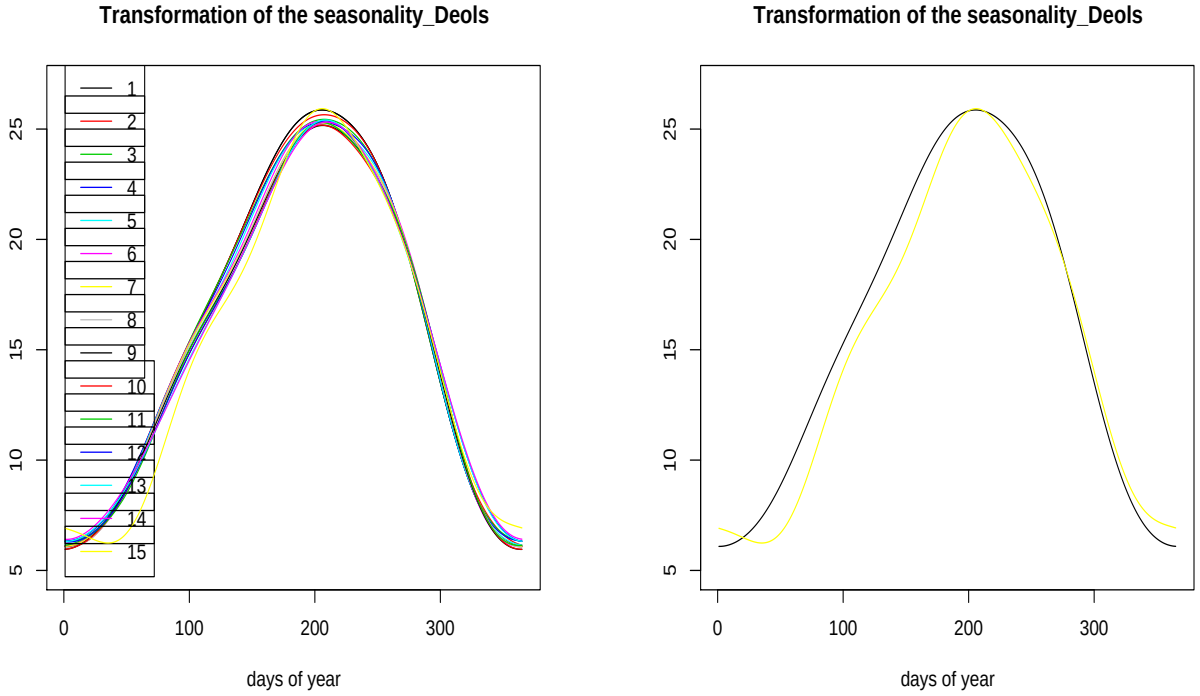


FIGURE 9.2 – Transformation of the seasonality in Déols with the size of window $N = 50$ years. *Left panel.* The estimated seasonalities in all the windows. *Right panel.* The estimates in the first and the last window

variation changes in probability distribution and correlation from season to season, but not much more rapid changes (as mean, variance function).

For example, one can think of a simple and reasonable parametric approach, to find a trigonometric polynomial which has coefficients slowly changing with time. For example :

$$\sum_{k=1}^{p} \theta_{k,1}(a) cos \frac{2\pi t}{q} + \theta_{k,2}(a) cos \frac{2\pi t}{q} \tag{9.1}$$

where t is the date, a is the year and q is the period length (q=365 for the temperature of the whole year).

The coefficients θ's change with years.

If one estimate the parameter $\theta(a)$ from the observations of the year a, the estimated seasonality also leads to take the higher frequency change in mean and variance. This way of estimation does not only detect the slow variation.

There are not many available works on time series decomposition methods which take into account the seasonality change. Barbosa ([8]) modeled

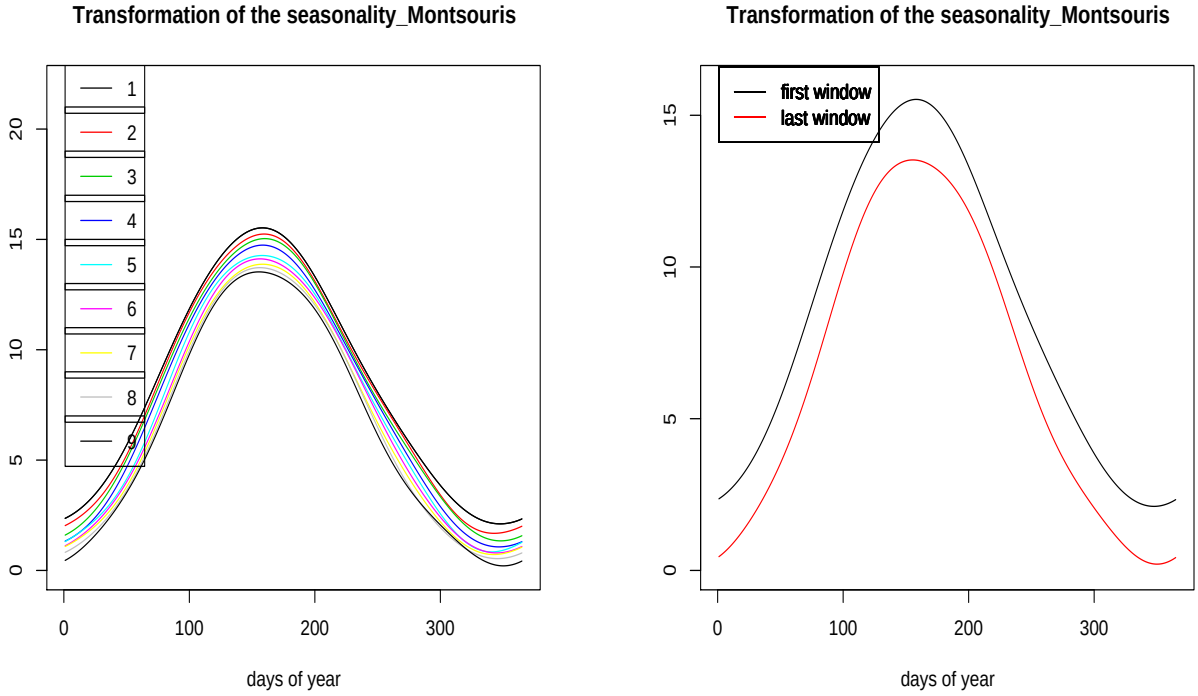


FIGURE 9.3 – Transformation of the seasonality in Montsouris with the size of window $N = 50$ years. *Left panel*. The estimated seasonalities in all the windows. *Right panel*. The estimates in the first and the last window.

the seasonality by an AR(p) with high value of p (for example p can equals to 31). Recently, Wu et al.([152]) gave an important work on the seasonality change and they proposed the modulated annual cycle (MAC) which allows the change of both amplitude and frequency of the seasonality. "MAC shows the response of the climate system to the external forcing, reflects the internal complexity of the climate system as well as the forcing".

Wu et al. extracted MAC using the ensemble empirical mode decomposition (EEMD). The detailed description of the EEMD method can be found in Wu & Huang ([150]).In practice, the EEMD is implemented through a sifting process that uses only local extrema.

Using modulated annual cycle is a good idea, however its calculus is complicated and has the heavy numerical computation. This method is not convenient when someone works with large datasets.

In summary, there are not any methods adaptable to our case. Facing this situation, another point of view can be taken. We use a traditional long term annual cycle which is constant year after year by accepting that this stationary annual cycle may not reflect well the intrinsic nonlinearity of the

temperature. Then knowing that the seasonality still remains in the reduced series (after removing the trend and seasonality components), we estimate this reduced series by a periodic stationary process. This kind of process will be studied in Chapter 9 on modeling the reduced temperatures.

Now placing ourselves in a case of a constant cycle for the seasonality, the literature propose many methods, both parametric and nonparametric, to perform estimations. In any way, we need a flexible estimation, as nonparametric ones, to have objective estimators. For our temperature series, we use the STL method, a rather simple method based on loess, which can however give robust estimates on trend and seasonality. On the other hand, the nonparametric estimators can be considered as good measures for the goodness of parametric fits.

We will show here one example on the daily mean temperature in Bordeaux during the whole year, note X_t. STL will be used to detect the seasonality component in mean of X_t. A parametric approach by trigonometric functions is chosen. The number p of trigonometric terms is chosen by the Akaike criterion.

We show an example on the daily mean temperature in Bordeaux. The results of the estimated seasonality by different methods in Figure 9.4 show the similarity of these two estimates : nonparametric by STL method and parametric by trigonometric functions. The parametric estimation for a invariant seasonality is correct.

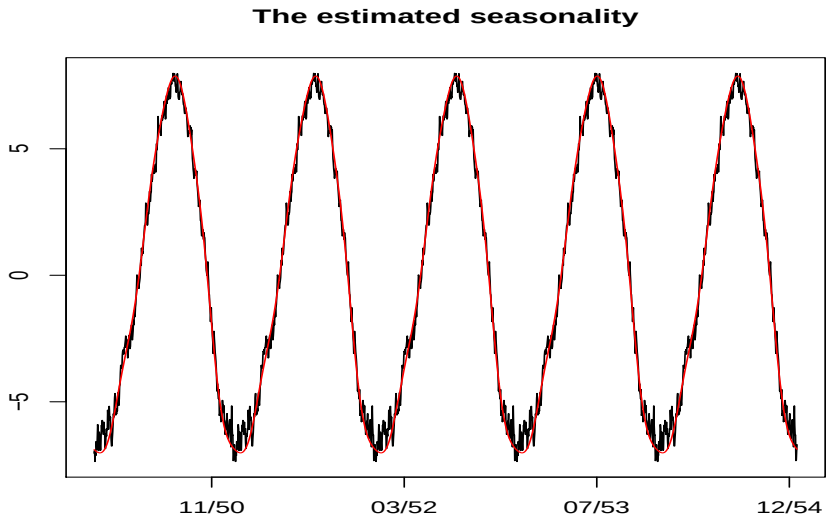


FIGURE 9.4 – Estimators of the seasonality in the mean : parametric one (*red line*), nonparametric one (*black line*)

9.4 Conclusion

This chapter discusses about the dynamics of the seasonality, but does not give a solution to this problem. The change in seasonality of the climate is well-known nowadays. However efficient methods to detect this dynamic seasonality are not found. While the trend and seasonality components are unobservable, it is a really complicated problem. The reason is that it is not obvious to completely remove the seasonality without confounding with the trends. More deeper and more detailed studies are necessary to reveal the characteristics of the seasonality.

In our case, because of lack of methods, we consider the seasonality as constant year after year and model it parametrically. Then the seasonality in the residuals will be continuously treated. The effect of the seasonality is still strong in the central field (mean, variance) and extreme field (extreme parameters), which will be shown in the next chapter.

Chapitre 10

From continuous to discrete-time models taking into account extremes

10.1 Introduction

Air temperatures are measured with different lags, from 10 minutes to 24 hours. Many statistical problems, linked with the mean, the extreme values, or the construction of simulation models, require physical coherence to consider the observations as a discrete sample of a continuous process.

We address the following situation : first we remove trends and seasonal components to obtain reduced discrete observations. They are provided by a nonlinear stationary continuous time process, which obviously possesses the Markov property, so it is a diffusion. Discrete observations, made for instance at a fix hour, have to keep this Markovian property, and if the diffusion is stationary the discrete chain have the same invariant density. One can hope that the reduced variables are modeled by a stationary diffusion, stationary in an extended sense : there can remain the seasonality in its coefficients. For a short period of the year, thus without seasonality, the observations, done at a fixed hour every day, can proceed from a stationary diffusion. The data of interest, as daily maximum, minimum, or mean value for a day, are functions of discrete observations with a possible loss of markovianity, for example for the mean which is not only non Markovian but anticipative.

The boundedness of temperature distributions is implied by the statistics of extreme values of the discrete reduced (or not) observations. This characteristic is implied by negative shape indices. This boundedness property is translated into a particular property for the underlying continuous time diffusion : it has inaccessible bounded boundaries.

The starting point of the applied literature for models of simulation of temperatures is a discretization of an Ornstein-Uhlenbeck process using an Euler first order scheme, with a seasonal deterministic component, leading to the so-called Vasicek model given by :

$$dX_t = d\beta(t) + \alpha(\beta(t) - X_t)dt + \sigma(t)dW_t \tag{10.1}$$

where β and σ are deterministic seasonal components for the drift and the diffusion coefficient respectively, α allows the return to equilibrium for this stationary diffusion, W is a Brownian motion.

The discrete version is :

$$X_n - \beta_n = (X_{n-1} - \beta_{n-1})(1 - \alpha) + \sigma_n \varepsilon_n \tag{10.2}$$

where ε_n is a standard Gaussian white noise.

Since the first works (see Dornier and Querel, [44] for instance), some other propositions have been made, (see Alaton et al. [2], Brody et al. [12],

Benth and Saltyte-Benth [39]). Campbell and Diebold [14] propose changes starting from the discrete version (10.2) where αX_{n-1} is replaced by an AR(p) process (the statistical procedure leads to $k = 3$), so a longer memory and secondly ε_n is no more a white noise but a GARCH process $\varepsilon_n^2 = \gamma\varepsilon_{n-1}^2\eta_{n-1}^2$, where η is a Gaussian white noise. ARMA or more generally classical time series models (see Roustand et al. [129], Moreno [95], and Cao et al. [15]) are not adapted to our problematic.

All these models only take what we call the "central part or field" component of the whole dataset into account. In fact, tests about the quality of the simulations, do not consider the extreme values, even if sometimes clustering properties are considered. Nothing is said about the quality of the model for tails.

In our specific application for very high (or low) temperatures, the non gaussianity is evidently seen from data graphics. The choice of a linear model as an Ornstein-Uhlenbeck or a Vasicek one, whose diffusion coefficients do not depend on the state X_t, implies that extremes have a Gumbel limit distribution for the imbedded discrete model is Gaussian and so is not appropriate.

An important difficulty is to take into account the boundedness of the distributions. Problems of the estimation of the support of a distribution are always difficult in statistics. In our case, they are critical because the estimation of the support is one of the determinant elements to estimate the risks of rare events.

Literature is much more developed on heavy tails, often linked to long memory properties, than on bounded ones. So a large work remains to be done. At least, working with hour time scale (the day or three hours), such correlation properties never appear.

We first detail and give some new precisions on the specific features of the extreme theory for stationary diffusions (see Berman [9],[10], Davis [122]), for the bounded case, in order to apply them to estimate extreme parameters from a discrete sample. A theoretical development will be carried out for the bounded case with inaccessible boundaries.

If we do not take into account the seasonality, the dynamics of a stationary diffusion process are governed by the conditional mean (drift) and variance (diffusion coefficient or volatility). A continuous-time diffusion process can be written as :

$$dX_t = b(X_t)dt + a(X_t)dW_t \tag{10.3}$$

where b is the drift and a is the diffusion coefficient of the diffusion process. W_t is a standard Brownian motion.

The temperature is a continuous process indexed by continuous time, but observed, at least for our data, at discrete dates, in general regularly spaced by a fixed lag Δ. Is Δ small or not ? This question cant make sense in itself from a mathematical point of view.

It is then tempting, for the next step, to work with the stationary bounded discrete-time Markov chain, the skeleton of the continuous time diffusion for $t = k\Delta$. However, this leads, as usually in this domain, to many difficulties. Indeed, we do not know how large Δ can be and still allow the use of these approximations. Moreover, one can be faced with complex computations to estimate the parameters passing by the transition operator. So simpler approximations for discretely sampled diffusions are needed. Nothing can make us think, a priori, that they are sufficient for the physical situation.

A natural and simpler discretization is to use a first order Euler scheme of the diffusion on the reduced series. This first-order discrete process is given by :

$$X_n = b(X_{n-1}) + a(X_{n-1})\varepsilon_n \tag{10.4}$$

where b and a are suitable functions depending on some parameters and ε_n is a white noise.

In fact the dynamic is seasonal so b and a are periodic functions for a given value of the temperature. They must be written as $b([n], X_n)$ where a general date n is written as $n = (year, [n])$. We omit to mention this seasonality in this chapter but we have to remember that it is the source of difficult problems from the practical problematic !

There are many papers about these functional autoregressive processes (the autoregressive order can be higher than one) with a non constant conditional heteroscedastic variance (FARCH). In most of the papers, conditions for geometric ergodicity are given. The literature will be detailed later. However our situation is more complicated. We need a model for a stationary process whose distribution has a bounded support due to extreme properties. We shall prove that for the associated continuous models the diffusion coefficient a is not lipschitzian at the boundary and null out of a bounded interval. There are too much constraints for a discrete Markov chain, except if we work with an (ε_n) bounded (we can give the bounds), which seems physically difficult to admit.

So there are no Euler schemes which can approximate the discrete Markov chain if we want to keep all the features suggested by the continuous time diffusion and specifically the gaussianity of the generating noise. We propose another solution, only motivated by statistical applications and the need to have a good simulation tool. If we keep the gaussianity, the boundedness is lost, but also, as will be detailed later, the process has some deterministic

part into its trajectory.

We choose, in the second step, the function a as a strictly positive function but almost negligible out of a bounded support. In fact for very long observations we should lose the boundedness of the distribution. Nevertheless, we keep a bounded estimate for a where the bounds are estimated from our data, or from an arbitrary long dataset with a fixed length. a can be now adapted to the data length.

This is not of course a completely satisfactory solution, but we get a good fit by simulation studies even for extremes. This study proves the limit of the first-order Euler scheme when a rough discretization of diffusions, in a stationary framework, reaches to the finite inaccessible boundary problem. Many complicated discretization schemes can perhaps be used but nothing in theory proves the existence of a scheme that satisfies all our needs. If we want something clearly simpler than the exact discrete Markov chain, a possible way, as we indicated, is to look for another approximation of the density transition. This does not strictly demand an approximation for a small enough Δ, but this needs, in any case, a very important and heavy numerical work for optimizations.

These points will be studied and discussed more in detail below. Then some of these points will be clarified by a simulation study.

10.2 Theoretical results on continuous-time diffusions and their discrete time approximations

In this section, we will remind the classical theory of the continuous-time diffusion with inaccessible boundaries r_1, r_2. We develop then some specific theoretical points for the bounded case, where r_1 and r_2 are finite inaccessible boundary points. After that, a discussion on the diffusion approximations for discrete samples wile be given. A specific model for discrete bounded data will be proposed and its statistical properties will be detailed, accompanied with a simulation study. The estimation for the diffusion coefficient with the boundary constraints, inheriting the extreme theory for the bounded continuous process, is also mentioned

10.2.1 Extremes of diffusion with inaccessible finite boundaries

Let us state some results from probability theory concerned by the maximum on an interval for a stationary diffusion when the length of the interval

increases. The main contributions on this subject are Berman([9],[10]) and Davis([122]). The basic results are due to Berman ([9]) but the version of Davis ([122]) fits a statistical framework better.

10.2.1.1 Main definitions and Berman result

Let X_t be a recurrent diffusion defined on the open interval (r_1, r_2) with Lipschitzian coefficients a and b and the endpoints r_1 and r_2 are inaccessible

$$dX_t = b(X_t) + a(X_t)dW_t \tag{10.5}$$

where b is the drift and a is the diffusion coefficient of the diffusion process. W_t is a Brownian motion.

Let $T_a = \inf\{t > 0 : X_t = a\}$ denote the first passage time to state $a \in (r_1, r_2)$. Let s be the scale function of the process :

$$s = \int^x e^{\int^u -2\frac{b(v)}{a^2(v)}dv} du \tag{10.6}$$

If $r_1 < a \leq x \leq b < r_2$, then we have :

$$P^x\{T_a < T_b\} = \frac{s(b) - s(x)}{s(b) - s(a)} \tag{10.7}$$

where $P^x(.)$ is the underlying probability measure of the process given that $X_0 = x$.

Since the process is recurrent,

$$\lim_{b\uparrow r_2} P^x(T_a < T_b) = \lim_{b\uparrow r_2} \frac{s(b) - s(x)}{s(b) - s(a)} = P^x(T_a < \infty) = 1 \tag{10.8}$$

we must have $s(b) \to \infty$ as $b \to r_2$. Similarly, $s(a) \to \infty$ as $a \to r_1$. From the form of s, we can deduce that $a(b) \to 0$ as $b \to r_1$ or $b \to r_2$. Conversely, if $a(x) \to 0$ as $x \to r_1$ or $x \to r_2$, one can show that r_1 and r_2 are inaccessible.

It is one important property of the extremes of diffusion with inaccessible boundaries that we will use in the next sections.

Now consider the theorem 3.2 in Davis [122] on the behaviour of the maxima of the process.

Theorem 10.2.1 *Suppose Z_t is a recurrent diffusion on (r_1, r_2) with scale function $s(x)$ and finite speed measure m. If the endpoints r_1 and r_2 are inaccessible, then for any sequence $w_t \to r_2$ and any $x \in (r_1, r_2)$,*

$$\mid P^x(M_t^Z \leq w_t) - e^{-1/(s(w_t)\mid m\mid)} \mid \to 0 \text{ as } t \to \infty \tag{10.9}$$

From the previous theorem, and without loss of generality, suppose that the total mass of speed measure $\mid m \mid = 1$, we have the lemma

Lemma 10.2.1 *Define the function F by :*

$$F(x) = exp(-1/s(x)) \tag{10.10}$$

Let $M_T = max(X_t,\ 0 < t < T)$. If there exists some functions $A_T \in \mathbb{R}^*$ and $B_T \in \mathbb{R}$ such that

$$\frac{M_T - B_T}{A_T}$$

converges in distribution to G, then G is a GEV distribution such that F is in the extreme domain of attraction of G.

The GEV distribution G associated with the maximum value in a large block of consecutive observations at discrete times is not in general the same GEV distribution H associated to a sample of the marginal density . H and G may even have different shape parameters (see Davis [122] for an example). In what follows, we shall give a condition linked to the boundedness of the distribution of this model.

From the lemma (10.2.1), we have :

Suppose a and b are continuous bounded Lipschitzian functions on the open interval (r_1, r_2) and we suppose $a > 0$ on the open interval (r_1, r_2) and $a = 0$ at the boundary points, b is continuous on the closed interval, (because $s(x)$ tends to infinity as x tends to r_1 or r_2, and so a tends to 0) and

$$\frac{s"}{s} = \frac{-2b}{a^2} \tag{10.11}$$

10.2.1.2 Bounded case

In this paragraph, we develop the extreme theory into a specific framework : bounded process, i.e r_1 and r_2 are finite and the maxima of process is in the domain of attraction of Weibull with the shape parameter $\xi < 0$.So F is in the extreme domain of attraction of some GEV distribution G with shape parameter $\xi < 0$. We can obtain the following lemma :

Lemma 10.2.2 *Suppose that F is in the extreme domain of attraction of some GEV distribution G with shape parameter $\xi < 0$, let r_2 the common upper bound of F and G.*

We have the following behavior of a near the upper bound r_2,

$$a^2(t) \approx \frac{-2b(t)(r_2 - t)}{1 - \frac{1}{\xi}} \tag{10.12}$$

Remark : In the case $\xi > 0$ the results are slightly different and more complicated when $\xi = 0$.

This lemma thus proves that if two diffusions have the same drift, the same $\xi < 0$ and the same upper bound, then necessarily near this bound the diffusion coefficients are equivalent.

The proof is based on the proposition 0.7 in Resnick, [138].

Theorem 10.2.2 *(Resnick). Suppose $U : R^+ \to R^+$ is absolutely continuous with density u so that*

$$U(x) = \int_0^x u(t)du \tag{10.13}$$

If $U \in RV_\varphi$, that means U is regularly varying with index φ ($U(x)$ equivalent to $L(x)x^\varphi$), $\varphi \in R$ and u is monotone then

$$\lim_{x\to\infty} \frac{xu(x)}{U(x)} = \varphi \tag{10.14}$$

and if $\varphi \neq 0$ then (sgn φ)$u(x) \in RV_{\varphi-1}$

Following the proposition 1.13 in Resnick, [137], if F is in the domain of attraction of the Weibull distribution with $\xi < 0$, then $1 - F(r_2 - x^{-1})$ is regularly varying with index $1/\xi$ when $x \to \infty$.

From the precedent theorem of Davis, we have

$1 - F(r_2 - x^{-1}) = 1 - exp(-1/s(r_2 - x^{-1})) \propto 1/s(r_2 - x^{-1})$ when $x \to \infty$

Then $1/s(r_2 - x^{-1})$ is regularly varying with index $1/\xi$, so $s(r_2 - x^{-1})$ is regularly varying with index $-1/\xi$.

Note $s^*(x) = s(r_2 - x^{-1})$, $s^*(x) \in RV_{-1/\xi}$, the first and second derivative of $s^*(x)$ are monotone. Applying twice the previous proposition of Resnick,

we have :

$$\lim_{n\to\infty} \frac{xs^{*"}(x)}{s^{*'}(x)} = -\frac{1}{\xi} - 1 \Leftrightarrow \lim_{n\to\infty} \frac{x\left[s"(r_2 - x^{-1}\frac{1}{x^2} - \frac{2}{x}s'(r_2 - x^{-1}))\right]}{s'(r_2 - x^{-1})} = -\frac{1}{\xi} - 1$$

$$\Leftrightarrow \lim_{n\to\infty} \frac{1}{x}\frac{s"(r_2 - x^{-1}}{s(r_2 - x^{-1})} = 1 - \frac{1}{\xi}$$

Combining with (10.11), we can deduce the behavior of a near the upper bound with $t = r_2 - x^{-1}, x \to \infty$:

$$a^2(t) \approx \frac{-2b(t)(r_2 - t)}{1 - \frac{1}{\xi}} \tag{10.15}$$

From this lemma, we can deduce the value of the first derivative of a^2 at the upper bound r_2 :

$$(a^2)'(r_2) = \frac{2b(r_2)}{1 - 1/\xi} \tag{10.16}$$

and the same for the lower bound. *square*

These values will be used as constraints at the boundary. From the lemma and the formula

$$\nu(x) = \frac{1}{\mid m \mid a^2(x)e^{\int^u -2\frac{b(v)}{a^2(v)}}} \tag{10.17}$$

elementary computations proves that the tail of $\nu(x)$ has the same behaviour as the tail of F given by

$$F(x) = exp(-1/s(x))$$

In fact, from (10.15), we have when $v \to r_2$, $e^{\int^u 2\frac{b(v)}{a^2(v)}}$ is equivalent to $(r_2 - t)^{1-1/\xi}$ with $t \to 0$, then $\nu(v)$ is equivalent to $C(r_2 - t)^{-1/\xi}$ (C is an appropriate constant) which in turn proves that the extremes computed on a sample of i.i.d random variables of distribution $\nu(x)dx$ have the same shape coefficient as that of the diffusion.

In summary, the tail of the stationary distribution has the same shape parameter as that of F.

10.2.2 Diffusion approximations for discrete samples

When studying the criteria of a good discrete approximation of a continuous time process, the question : "what are the statistical properties of the discretization" is a very important problem, specifically for processes as

diffusions whose trajectories are very irregular, and thus unobservable. This is also true for numerical aspects of stochastic differential equations. We show here the exact discretization of diffusion process and also their limits in applications.

10.2.2.1 Discretization, imbedding and limits

Often the continuous time processes are built as limits or as sequences of interpolations of discrete time ones. This is true, of course, for the Brownian motion as for many diffusions starting, for example, with as popular discrete time processes as a GARCH processes. There is a quite large literature, not unified, on these topics. Another point of view, probably closer to the physical measures, is that it is given by the regularization of trajectories, say by convolution of the continuous data with an infinitely derivable positive with compact support function. This function is chosen in order to represent the inertia of the effective system of measurement (see Azais and Wschebor, [7]). For more regular processes, as ARMA or more generally CAR (continuous time autoregressive), or threshold processes, there are also many works, see for instance Souchet ([144]) and Guyon & Souchet ([63]).

For extremes, the discretization statistical effect is not completely understood even today. There are, as explained in the chapter 5 on extremes, a lot of results on the limit distribution of the maxima of discrete time processes and their possible use in statistics. For maxima of continuous time processes a large bibliography, and many important results, can be found in Azais and Wschebor ([7]). The link is also made between their work on maxima of processes with irregular trajectories, as diffusions, and their regularization. For this link between extreme of discretization and continuous time process, the seminal work is the one of Pitterbag([119]). A general framework for extensions can be found in Anderson & Turkman ([6]) and Albin ([3]) but with no significant applications until today.

We do not use in our work the results concerning imbedding of a discrete time process as the "skeleton" of a continuous one.

10.2.2.2 The exact discretization and their limits

The use of the exact transition density $p_\Delta(x, y)$ can be tempting even if we have to use approximations with respect to Δ_n. Formulas giving p_Δ as functions of a and b are complicated. Dacunha-Castelle and Florens-Zmirou [29] give a method to approximate p_Δ in order to obtain a consistent asymptotic theory if $n \to \infty$, $n\Delta_n \to \infty$, $\Delta_n \to 0$. This method has been supplied by Ait – Sahalia [153] and the basic lemma was used to build exact simulation methods in Beskos and Roberts[1].

Thus if we assume here that the diffusion process (10.5) is stationary (and ergodic), and that a discrete observation $(X_{k\Delta_1 \leq k \leq n+1})$ of the sample path is available, an asymptotic framework is available for $\Delta_n \to 0$ (high-frequency data) while $n\Delta_n \to \infty$ as $n \to \infty$.

p_Δ expression as a function of a and b is based on the following formulas :(Dacunha-Castelle and Florens-Zmirou [29])

$$
\begin{aligned}
&p_\Delta(x,y) = A_\Delta(x,y) E L_\Delta(x,y) \\
&\text{with } A_\Delta(x,y) = \frac{1}{\sqrt{(2\pi\Delta)a(y)}} exp\left[-\frac{(\nu(y)-\nu(x))^2}{2\Delta} + H(y) - H(x)\right] \\
&\nu(x) = \int_0^x \frac{du}{a(u)},\ H(x) = \int_0^{\nu(x)} C(u)du \\
&\text{with } C(\nu(x)) = \frac{b}{a} - \frac{1}{2}a' \\
&L_\Delta = exp\left(\Delta \int_0^1 [g(1-us(x)+us(y)) + \sqrt{\Delta}B_u]du\right) \\
&\text{with } g = -\frac{1}{2}(C^2 + C")
\end{aligned}
\tag{10.18}
$$

where B is the Brownian bridge and with the condition $g(x) = O(|x|^2)$, $x \to \infty$

Then the asymptotics are based on Taylor formulas applied to A_Δ and L_Δ as $\Delta \to 0$.

Its use in parametric statistics is almost practically impossible in general cases. It is possible that for very simple parametric problems, we need heavy computations for these approximations.

Contributions to this estimation problem can be found in Prakasa-Rao([120]), Yoshida[155]). A complete bibliography is available in Souchet([144]). Most works concern drift estimation. They give asymptotic results for $n\Delta_n \to \infty$, $\Delta_n \to 0$.

There exists a very important literature on the estimation of discrete-time diffusion and the asymptotic properties of the estimates for a fixed lag, such as Florens-Zmirou ([51], Hansen et al.([70]), Kessler and Sorensen([88]) and Gobet et al.([57]).

The practical use of these results remains difficult for complicated models, as that linked to the temperature because they are linked to non realistic hypothesis, as reflexions at the boundaries. We do not present in this thesis the numerical work done with these estimators.

10.2.3 Approximate discretizations and their statistics

The simplest approximation is the one given by a FARCH model (functional, heteroscedastic autoregressive model) associated to the equation :

$$X_{n\Delta} = b(X_{(n-1)\Delta})dt + a(X_{(n-1)\Delta})\varepsilon_n \tag{10.19}$$

where ε_n is a white noise or more precisely a sequence of i.i.d random variables with density E. These models can be extended in an obvious manner to :

$$X_n = \sum_{i=1}^{p} b_i(X_{n-i}) + \sum_{j=1}^{q} a_j(X_{n-j})\varepsilon_{n-j+1} \tag{10.20}$$

There is an important literature about these processes. The most general approach can be found in Doukhan ([38]) where various criteria are given for the existence of a stationary solution and for geometric ergodicity. Some papers (see Chen and An [19] , Wang [64] give more specific conditions. But in general, if b is linear, these conditions are mainly : $a > 0$, a is bounded on $\mathbb{R}$ and the density of E is almost *everywhere positive*. So bounded supported noise are excluded.

Another approach is that of Wang [64]. Now a is *uniformly lipschitzian,* the existence of unique stationary solutions, as well as the fact that the geometrical decrease of the covariance as strong mixing properties, has been proved.

The geometric decrease of the covariance allows the applicance of the results of Leadbetter et al [92] about extremes of this kind of processes.

As remarked by Wang, the lipchitz condition can be weakened for some explicit a but no results are given in this direction.

It seems that general results on distributions, and particularly some detailed results concerning the case of the boundedness of the support of the invariant measure, have not yet been given.

In our case, a^2 is estimated using the information provided by extremes theory, is not uniformly lipschitzian for $\xi < 0$. We have to look for a model with bounded marginals, but unfortunately we have the following almost evident negative result.

Lemma 10.2.3 *If E has a non bounded support (for instance E Gaussian) then there are not bounded supported stationary solutions of* (10.19).

It can exist bounded solution of (10.19) *for a bounded noise.*

In our applications, we want, for the noise, gaussianity as a condition of physical coherence. Nevertheless, from Lemma 10.2.3, for every fixed Δ, the

boundedness of the observations implies a non Gaussian noise with bounded support. It is thus impossible to have such an approximate discrete model with the following wanted properties : it is stationary with bounded inaccessible boundary and has a Gaussian innovation. Then what are the possible choices for this situation ?

The process (10.19) for bounded data, with parameters a and b, should have the following properties if the noise is Gaussian :

• it is a markovian process , it is recurrent for the potential of open bounded intervals being infinite whatever be the initial measure.

• suppose $J = (r_1, r_2)$ be the support of a. We see that the transition density is such that :

$$p(x,y) = \frac{1}{\sqrt{2\pi}a(x)} \exp\left(-\frac{1}{2}\frac{(y-b(x)^2}{a(x)^2}\right) \text{ for } x \in J$$

$$p(x, b(x)) = 1 \text{ for } x \in J^C \text{ and } p(x,y) = 0 \text{ for } x \in J^C \text{ and } y \neq b(x)$$

Does this recurrent process has an invariant density and so a stationary version ?

In order to answer this question, a convenient approximation for a statistical purpose will be studied. It will be a FARCH process, called $a(\eta)$ approximation of the process defined before.

Let $a(\eta)$ defined by :

- $a(\eta, x) = \delta(x)$ for $x > r_2 - \eta$ or $x < r_1 + \eta > x > -\infty$ with $\eta > \delta(x) > 0$ for $x > r_2 - \eta$ or $x < r_1 + \eta$ and $\delta \to 0$ as $x \to \infty$.

- $a(\eta, x)$ is three times continuously differentiable in x and uniformly lipschitzian.

This process, defined by $a(\eta, x)$, has evidently from Wang results a stationary version and is geometrically ergodic and weakly mixing.

η and $\delta(x)$ will be chosen depending on the sample size N for practical applications.

Theoretically, the extremes of this process are in the Gumbel domain of attraction. But if we choose η small enough, for instance if $N = 10^5$ then $\eta = 10^{-5}$. Therefore, for a *statistical point of view*, the extremes of this process are the same as those of the bounded continuous time process(the "exact extremes"). Of course, this is only true form a statistical point of view ! We shall see through simulations if we have the "same extend" as the one of a discrete-time process associated to the original $a(x)$ with almost (in a very strong sense) all the observations in (r_1, r_2). Some more precise practical considerations would be done using the elemental large deviation theory.

To finish with these considerations, let us consider how the continuous-time process could be considered as a limit of discrete ones. First we can remark that the convergence of the discrete model, given by (10.19), to a

diffusion signifies that with parameters a and b, the noise ε_n^Δ has to depend on Δ and to converge to a Gaussian noise as Δ tends to zero. A posibility is to use coefficients a and b depending on Δ and on the sample size N.

For some Δ, let the equation :

$$X_{(n+1)\Delta} = b(X_{n\Delta}) + a(X_{n\Delta})\varepsilon_{(n+1)\Delta} \tag{10.21}$$

Let $X_t^\Delta = X_n^\Delta$ for $n\Delta \leq t \leq (n+1)\Delta$.

Suppose ε_n^Δ Gaussian or converges with a convenient speed to a Gaussian noise. Then if Δ_n is a sequence tending to zero, we shall see that the process $X_t^{\Delta_n}$ converges on the Skorohod space $\mathbb{D}$ to the solution of (10.19). We call this process an FARCH gaussian approximation of the diffusion with loss $|a_\Delta - a|$. This is true in our case when the noise is Gaussian and $a(\eta)$ is a strictly positive sequence of bounded strictly positive functions. These situations can be extended to a sequence $a_{\Delta_n}(\eta_n)$ converging uniformly on $\mathbb{R}$ to an a with the previous properties. This is a consequence of a lot of classical results such as those given in Inge Helland [81].

We do not know if it is true for the original a with a bounded support (r_1, r_2), but it is true with $a(\eta_n)$ with $\eta_n \to 0$.

A complete theory of functional autoregressive processes, with variances depending on the state, remains to be done. We do not enter it in this paper because we do not want to study strong approximations of the "skeleton' of a continuous time diffusion and so we are, in this work, out of the scope of this interesting problem.

10.2.4 The estimate procedure with the constraints of extremes

This subsection is devoted to the estimation of the coefficients in a discrete version of a bounded diffusion. In our work, for caution, we first estimate both the drift coefficient and diffusion coefficient by nonparametric methods, without any reference to any specific model. Both coefficients are then estimated by parametric models. Parametric modeling will be necessary for the simulation procedure at a later stage (see Comte and Rozenholc, [43] for an extended bibliography).

10.2.4.1 Estimation of the drift coefficient

$$X_{n+1} - X_n = b(X_n) + a(X_n)\varepsilon_n, \varepsilon_n \sim N(0,1) \tag{10.22}$$

As we already said, the estimation of the coefficients of a diffusion process has been considered in the literature for many years, with some of the papers being concerned by continuously sampled data.

We estimate the drift coefficient b in the supplied model first by local smoothing loess with b being the conditional expectation $E(X_n - X_{n-1}|X_{n-1})$. In general, we can see that the drift b is mostly linear. The drift function has a quite clear physical meaning as the elastic part of the basic oscillator which is the deterministic justification of the Ornstein – Uhlenbeck first model. So we suppose that this linear drift is valid up to the boundaries). So in our paper we consider only the linear drift $b(x) = \alpha x + \beta$, with $\alpha < 0$ and $\beta = 0$.

10.2.4.2 Estimation of diffusion coefficient

The form of the diffusion coefficient a is not monotonic as the drift b, but more complicated. For this reason, many nonparametric methods will be first considered to give flexible estimators for a. Then from that, we will propose a specific parametric form for a.

10.2.4.2.1 Non-parametric estimates The diffusion coefficient a, in model (10.22) can be estimated by a kernel method. If K is a kernel (Epanechnikov or Gaussian are used) and h is the corresponding bandwidth, then the conditional density $f(x, y)$ is estimated as

$$\hat{f}_n(x/y) = \frac{\sum_{n=2}^{N} \frac{1}{h_N^2} K\left(\frac{X_n - x}{h_N}\right) K\left(\frac{X_{n-1} - y}{h_N}\right)}{\sum_{n=2}^{N} \frac{1}{h_N} K\left(\frac{X_{n-1} - y}{h_N}\right)} \tag{10.23}$$

For its mean and variance, the summand in the numerator is multiplied by the adequate term. The kernel estimation uses local data. At the tails the data become asymmetric, so there is boundary bias. In our case, firstly because of the complexity of boundary correction for kernel, and secondly because the kernel estimator is only used to compare with other estimators in the central part and won't be kept, we don't mention the boundary correction for kernel estimation.

The estimator of a^2 can also be obtained by the penalized likelihood method. We start from the density

$$\frac{1}{a(X_{n-1})} \exp\left[-\frac{1}{2}\frac{(X_n - X_{n-1} - b(X_{n-1}))^2}{a^2(X_{n-1})}\right]$$

So the penalized likelihood is given as a function of a by :

$$L(N, a^2, \hat{b}) = \sum_{i=2}^{N} \left[-\frac{1}{2}\frac{(X_i - X_{i-1} - \hat{b}(X_{i-1}))^2}{a^2(X_{i-1})} - \frac{1}{2}\log a^2(X_{i-1})\right] - \frac{1}{2}\lambda \int [(a^2)"(x)]^2 dx \tag{10.24}$$

with a convenient regularization parameter λ, where $\hat{b}$ is estimated by linear least-squares : $\hat{b}(X_{n-1}) = \alpha X_{n-1}$ as previously. The maximization of L gives a as a cubic spline function.

The cubic spline estimation often has less boundary bias compared with the kernel one because of its natural boundary condition. It is constructed to have zero second and third derivatives at the boundaries, which reduces the bias near the boundary (see Simonoff [139]) .

For non-parametric estimates, the usual boundary corrections seem difficult to apply in estimating the volatility. Moreover, in our applications, the variation of a is naturally strong near the boundary, and so difficult to capture. So the problem of estimation of a near the boundaries has to be done with the help of the theory of extremes.

10.2.4.3 Parametric estimate

The parametric model is suggested by the previous results, the one on extremes and the non parametric estimates. This analysis suggests using a polynomial of high degree, with the constraints fixed by extreme theory at the boundaries to get a parametric model convenient also for simulations.

Supposing that the ε_t in (10.22) is Gaussian, for $a > 0$, the theory has be done by Kessler ([87]). It seems that the results could be extended to the case where a has a bounded support if the bounds are fixed. The reason is that the parameter set is defined by a^2 which can be chosen differentiable with respect to the parameters and x, despite the fact that a is not lipschitzian. Then the statistical problem is regular and the Fisher information is finite. We do not enter into details of the asymptotic results for $a(\eta)$. We can use Kessler results for instance.

We estimate a^2 by **maximum likelihood method with the boundary constraints** with given linear $\hat{b}$. Thus we maximize :

$$L(N, a^2, \hat{b}) = \sum_{i=2}^{N} \left[-\frac{1}{2} \frac{(X_i - X_{i-1} - \hat{b}(X_{i-1}))^2}{a^2(X_{i-1})} - \frac{1}{2} \log a^2(X_{i-1}) \right] \quad (10.25)$$

The constraints used at the boundaries are provided by the results in section 2 : at first, a^2 equals zero at the upper and lower bounds. Then, a^2 satisfies the condition (10.16) of first derivative at the two bounds, for example the first derivative of a^2 at the upper bound r_2 must be :

$$(a^2)'(r_2) = \frac{2b(r_2)}{1 - 1/\xi}$$

and a similar constraint for the lower bound.

We have then four possible constraints. The degree of polynomials is chosen by AIC criteria. Usually an appropriated degree is 8, which allows to give a flexible estimator satisfying these constraints. Outside the intervals (r_1, r_2), $\widehat{a}^2$ is forced to take zero-value.

In this chapter, to simulate discrete-time diffusions, the values of ξ and the two bounds are given. In the next chapter, these values are directly estimated from the data. To estimate the upper and lower bounds, we apply stationary usual GEV distribution for stationary X_t. The upper bound (and change the sign, for the lower bound) can be estimated from a function of the estimators of extreme parameters of GEV distribution : $r_2 = \mu - \sigma/\xi$.

With $\widehat{a^2}$ obtained from (10.25), we re-estimate b with the linear form by maximizing :

$$L(N, \widehat{a^2}, b) = \sum_{i=2}^{N} \left[-\frac{1}{2} \frac{(X_i - X_{i-1} - b(X_{i-1}))^2}{\widehat{a^2}(X_{i-1})} - \frac{1}{2} \log \widehat{a^2}(X_{i-1}) \right] \quad (10.26)$$

The results obtained show the stability of the estimation procedure for a^2 and b.

10.2.5 Simulation study on the properties of the model with a bounded a

Let us consider the model :

$$X_{n+1} - X_n = b(X_n) + a(X_n)\varepsilon_n, \varepsilon_n \sim N(0, 1) \quad (10.27)$$

A reasonable way for us is to study the following practical problem : what are the statistical properties of the process in (10.22) if a is constrained at the boundaries using the extreme theory, and ε is Gaussian ?

The difference in the extreme field of a Orstein-Ulhenbeck with a constant diffusion coefficient, and our model with the diffusion coefficient null out of the boundaries interval will be considered through simulations.

Another interesting point is to study the discretization schema of this kind of continuous-time model with finite inaccessible boundaries. It is complicated to simulate a continuous model, then we study on a discrete-time model whose marginal distribution has not really a bounded support (because of a Gaussian generating noise), but conditionally with data inside the given boundary points interval r_1, r_2. We will consider discrete-time model

(10.27) and simulate a large number N of its realizations, and then consider its sub-samples as its discrete versions.

However, when studying on the real data, as the data are provided by a bounded continuous-time process, they have naturally a bounded support.

10.2.5.1 Study on the properties of Model (10.27)

Our trick, as said before, is to use an $a(\eta)$ which has very small values out of the boundaries, not zero but really negligible. The question is : do two kinds of models, where a is practically the same, but for one, a is zero outside the support, and for the other, a is nearly zero, give practically the same realizations ? Or if they are different, this difference is significant or not ? In this part, we try to answer this question through simulations.

Then for the diffusion coefficient a, these two kinds are considered. We will study the simulations of (10.22) with an a positive everywhere, but very small outside of (r_1, r_2), and those with an a which equals zero outside (r_1, r_2). The constraints 10.16 about the derivatives of a^2 at the two bounds are also considered.

We take, in the two cases, the same a which is rather linear in the central part. This form of a is considered to give a coherence with a of temperature series, whose volatility is not constant but slightly linear (see in the next chapter).

- In model i/, take the shape parameter of extremes $\xi_1 = \xi_2 = -0.2$ and the lower and upper bounds respectively $r_1 = -6, r_2 = 7$. a^2 is taken as a polynomial of degree 8 with four constraints at the boundary which are mentioned in Section 10.2.4.2 : $a^2(x) = 0$ when $x \notin (r_1, r_2)$ and a^2 satisfies the condition (10.16).

- Model ii/ is similar as model i/ except for a equals 10^{-4} for $x > r_2 - 10^{-5}$ or for $x < r_1 + 10^{-5}$ and the first derivatives of $a^2(r_1 + 10^{-5})$ and $a^2(r_2 - 10^{-5})$ satisfying the condition (10.16).

These a are shown in figure 10.1, we cannot distinguish these two models . a in Model ii/ is in fact positive everywhere.

We always take a linear drift coefficient. Here $b(x) = -0.2x$. When a is negligible outside the boundaries, model ii/ can be replaced by model i/. Through the simulation, we want to verify if the process (10.22) from these two kinds of a are not different in the extreme field.

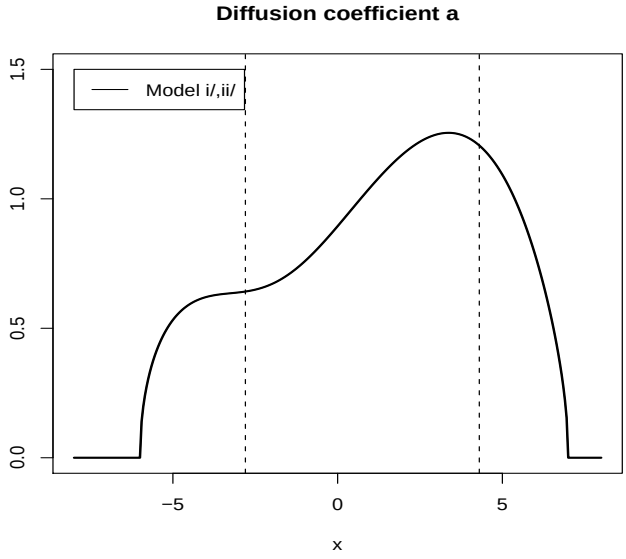


FIGURE 10.1 – Models of diffusion coefficient . The vertical lines are the 1% and 99% empirical quantiles

Simulate 100 samples of N values from model (10.22) for each model of a with N varies : $N = 5000, 10000, 50000$.We will study the difference in the tail behaviour through the average values of three parameters of GEV model, the upper bound and the number d of values which exceed initial r_2 of the samples of simulation following i/, ii/ and with the initial parameters. The extreme parameters are estimated using blocks of size 32. We consider only the right tail, the same conclusion hold for the left tail.

Initial parameters	N	Model i/	Model ii/
$\xi = -0.2$ $r_2 = 7$	5000	$E(\hat{\mu}) = 2.28$ $E(\hat{\sigma}) = 1.24$ $E(\hat{\xi}) = -0.156$ $E(\hat{r}_2) = 12$ $E(d) = 0.24$ $(4.8.10^{-3}\%)$	$E(\hat{\mu}) = 2.26$ $E(\hat{\sigma}) = 1.237$ $E(\hat{\xi}) = -0.152$ $E(\hat{r}_2) = 11.43$ $E(d) = 0.24$ $(4.8.10^{-3}\%)$
	10000	$E(\hat{\mu}) = 2.286$ $E(\hat{\sigma}) = 1.243$ $E(\hat{\xi}) = -0.159$ $E(\hat{r}_2) = 10.55$ $E(d) = 0.41$ $(4.1.10^{-3}\%)$	$E(\hat{\mu}) = 2.274$ $E(\hat{\sigma}) = 1.247$ $E(\hat{\xi}) = -0.154$ $E(\hat{r}_2) = 10.73$ $E(d) = 0.53$ $(5.3.10^{-3}\%)$
	50000	$E(\hat{\mu}) = 2.27$ $E(\hat{\sigma}) = 1.237$ $E(\hat{\xi}) = -0.153$ $E(\hat{r}_2) = 10.42$ $E(d) = 1.82$ $(3.6.10^{-3}\%)$	$E(\hat{\mu}) = 2.27$ $E(\hat{\sigma}) = 1.236$ $E(\hat{\xi}) = -0.15$ $E(\hat{r}_2) = 10.55$ $E(d) = 2.1$ $(4.2.10^{-3}\%)$

The table shows that the tail behaviors of the simulated samples from model i/ and ii/ of a are similar. These results show that we can use model i/ where $a = 0$ outside the boundary in the place of model ii/.

The simulations from two models are nearly the same for extremes. It is not surprising that, for a long simulated sample, some simulated values overpass the higher upper bound, which is estimated from the GEV distribution $\hat{\mu} - \hat{\sigma}/\hat{\xi}$.

Now we will check the property of the unique solution of model (10.22) with model i/ of a. We simulate a large sample S of 1 000 000 values from this model. We then re-estimate a and b parametrically following the parametric methods described in Section 10.2.4.2 . If this model has an unique solution, normally we should obtain after this estimation, $\hat{a}$ and $\hat{b}$ identical with initial a and b.
And equally, the final residuals $\hat{\varepsilon}_t$ must be a normal distribution of zero mean and unit variance.

Effectively, $\hat{b}$ is identical to b, $\hat{a}$ is identical to a in the interval (q_1, q_{99}) where q_1 q_{99} are the 1st percentile and 99th percentile of the sample S. However outside of this quantile interval, $\hat{a}$ is different because of the difference of estimated extreme parameters with the initial parameters. $\hat{\varepsilon}_t$ is accepted as the Gaussian noise following the Kolmogorov test (p-value = 0.87).

Now, we want to see if the difference between these estimators, with initial parameters, can give a practical difference in simulations, or not. To study this fact, from $\hat{\varepsilon}_t$ and parametric estimators $\hat{a}$ and $\hat{b}$, we build a sample of bootstrap B of 1 000 000 values. Then, the statistical properties of S and B will be compared.

Two samples have similar densities when we estimate them by the kernel method. The estimators of extreme parameters are also considered :

	Simulations	**Bootstrap**
$\hat{\mu}_1$	1.97	1.96
$\hat{\sigma}_1$	0.71	0.71
$\hat{\xi}_1$	-0.21	-0.24
$\hat{r}_1$	-5.30	-4.85
$\hat{\mu}_2$	2.27	2.26
$\hat{\sigma}_2$	1.25	1.24
$\hat{\xi}_2$	-0.15	-0.14
$\hat{r}_2$	10.30	11.42

There is not a significant difference in the shape parameters and the bounds.

After this study, we can see that practically there exists a unique solution for this kind of models, with a canceled outside the support. So it is reasonable to apply it in modeling the temperature series in the next chapter.

10.2.5.2 Difference in the extreme field with an Orstein-Ulhenbeck model

We rewrite Model (10.27) :

$$X_{n+1} - X_n = b(X_n) + a(X_n)\varepsilon_n, \varepsilon_n \sim N(0,1)$$

Here we take b with a linear form : $b(x) = \alpha x$ with $\alpha < 0$. a is firstly considered as a constant $a(x) = ac$ which gives Model 1, and then as a constant $a = ac$ in the centre but goes down to zero at the boundaries, which gives Model 2. Two boundaries are taken as $r_1 = -5, r_2 = 5$. One example of a in Model 2 is illustrated in figure 11.5.

In order to take into account four boundary constraints , a^2 is estimated by a polynomial of degree 8. Then $\hat{a}$ gives not exactly but nearly a constant in the centre.

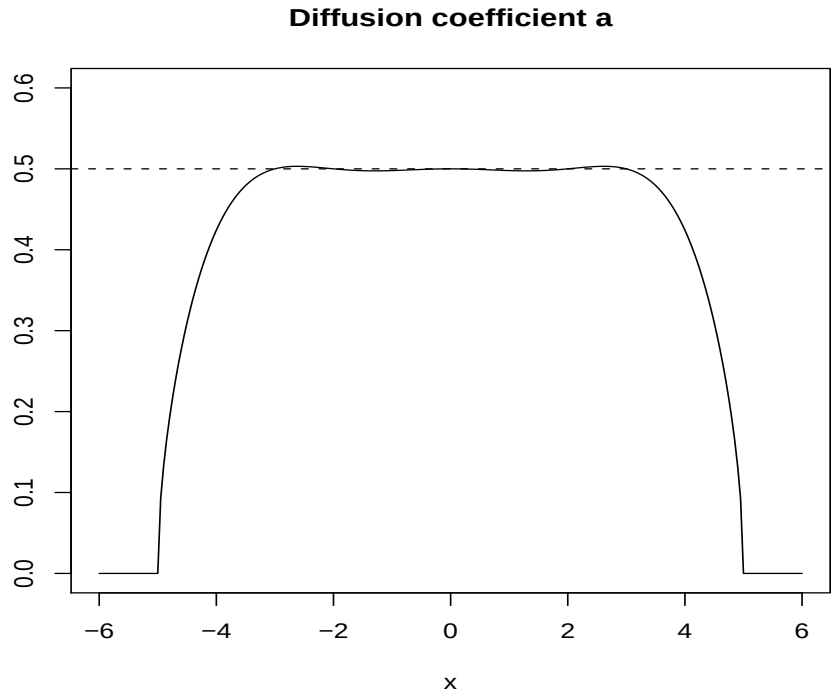


FIGURE 10.2 – Diffusion coefficient of Model 2

Now we simulate 50 samples of size N=10000 from these two models, then we consider their empirical density and extreme parameters from the GEV model with different values of α and ac.

Estimate the density functions of these two simulation samples by the kernel method, we obtain two different curves, which are shown in figure 10.3.

More concretely, we have the empirically statistical characteristics from these simulated samples in Table 10.1

Conclusion

Some interesting conclusion can be drawn from these results. When α is negative and near to zero, which means that the correlation between two consecutive values is strong, and that we are near an explosive situation, or at least near a transient process, the difference in the extreme field of Model 2 with respect to Model 1 is clearer. And this difference is proportional with the value of ac. For instance, with the maximum value of ac (1.0), the simulation sample of Model 2 is more bounded, with smaller variance,

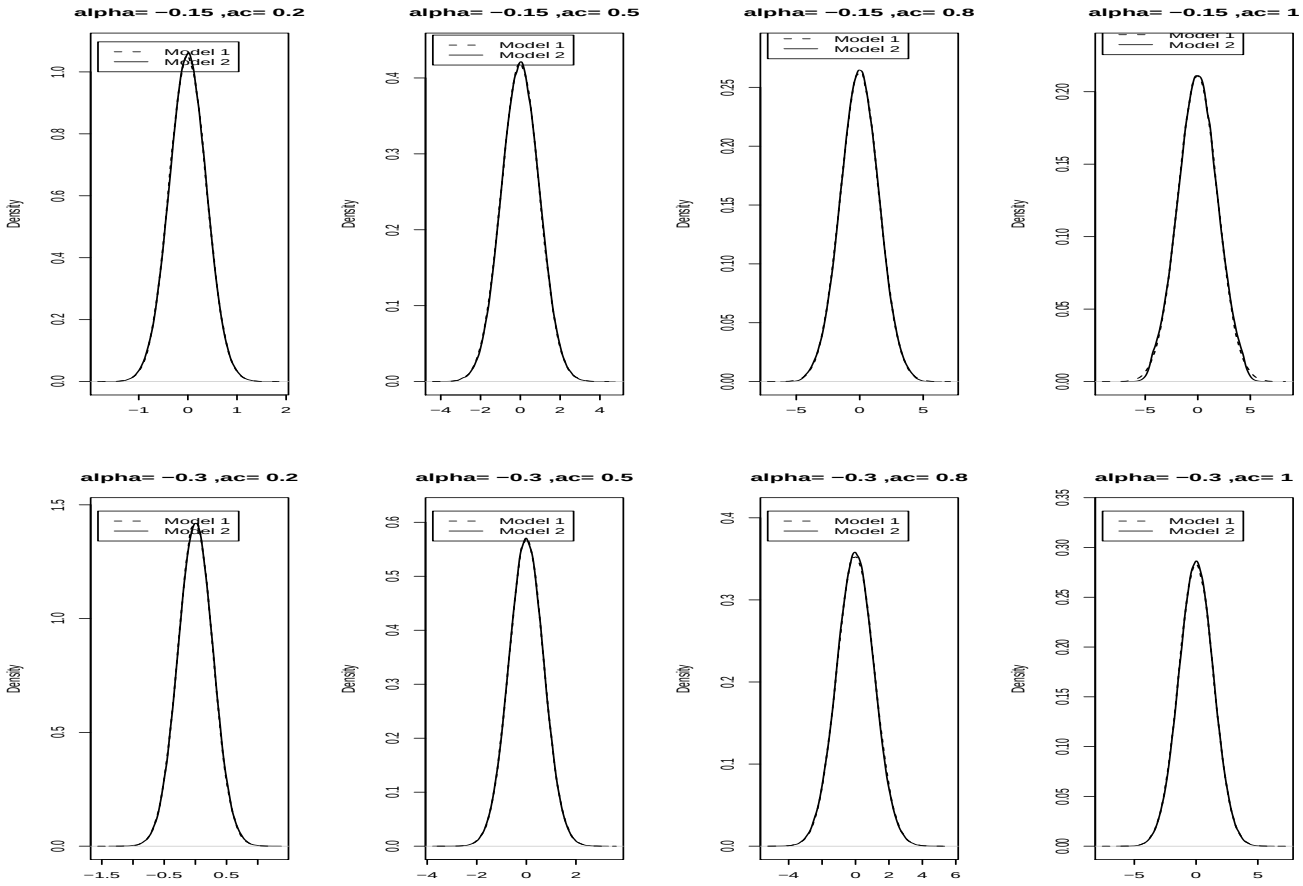


FIGURE 10.3 – Empirical density functions of simulation samples from Model 1 and Model 2

	Model 1				Model 2			
	Mean	Variance	Min	Max	Mean	Variance	Min	Max
$\alpha = -0.15$,ac=0.2	-0.001	0.144	-1.428	1.440	-0.001	0.144	-1.419	1.453
$\alpha = -0.15$,ac=0.5	-0.002	0.902	-3.500	3.560	0.006	0.901	-3.555	3.610
$\alpha = -0.15$,ac=0.8	-0.010	2.300	-5.755	6.835	-0.008	2.270	-5.100	5.157
$\alpha = -0.15$,ac=1.0	-0.004	3.580	-7.190	7.013	0.002	3.452	-5.853	5.873
$\alpha = -0.3$,ac=0.2	0.000	0.078	-1.085	1.061	0.002	0.079	-1.073	1.087
$\alpha = -0.3$,ac=0.5	-0.001	0.490	-2.719	2.705	-0.002	0.490	-2.705	2.686
$\alpha = -0.3$,ac=0.8	0.003	1.256	-4.272	4.346	-0.009	1.248	-4.343	4.262
$\alpha = -0.3$,ac=1.0	-0.002	1.961	-5.520	5.415	0.006	1.938	-5.117	5.117

TABLE 10.1 – Comparison the statistical characteristics of simulation samples from two models

higher minimum and smaller maximum. The estimated values of the shape parameter of ξ in Model 2 is evidently more negative than those in Model 1.

However in the case where α approaches -1, which means that the correlation between two consecutive values is small, the difference in the extreme field of the two models is negligible. This remark can be seen in values of the variance, the local extrema (minimum and maximum).

When the value of ac is small, which corresponds to a small volatility, the simulations are then automatically more bounded.

In summary, when the correlations between the observations are strong, a model with a constant volatility gives a heavier tail. So in this case, the constraints which reduce the value of a near the boundaries are necessary. For the temperature series, the correlation between sequential values are rather strong. Then the constraints in the boundary for the diffusion coefficient a^2 is advised to obtain a better fit to extreme events.

10.2.5.3 Discretization properties of Model (10.27) with bounded a

In this section, we will study the properties on the central field and extreme events from the discretization of diffusion process. We will address the influence of the size mesh working with simulations. The different sub sequences are extracted from the higher frequency sequence. The size of the initial sequence must be very large.

We will study the changes in the behaviour of the central and extreme parameters with the mesh of discretization for the bounded support diffusion coefficient a in (10.27). The diffusion coefficient in this case satisfies the constraints at the boundaries. We will consider two models of a, constant or slightly linear in the center. Different models for b and also AR(1)and AR(3) processes are compared .

Following the theory in Section 10.2.1.2, the marginal law and the shape parameters must be invariant with the discretization.

It is very difficult to simulate the data from a continuous-time diffusion with such a variance term. So we consider only a sample with a very large size in order that a high-frequency data can be considered as an approach of a continuous-time data. First, simulate a sample X of size N with $N = 100000$, then we consider its extracted subsequences $X_{k\Delta}$ corresponding to

different mesh of discretization Δ, $\Delta = 1, 2, ..., 10$. The density functions estimated from these sub-sequences can be compared with the theoretical density function of the continuous-time diffusion process :

$$dX_t = b(X_{t-1}) + a(X_{t-1})dW_t \tag{10.28}$$

whose form is :

$$\nu(x) = \frac{1}{\mid m \mid a^2(x) e^{\int^u -2\frac{b(v)}{a^2(v)}}} \tag{10.29}$$

which can be calculated when the functions a and b are known.

The different moments can be calculated based on the formula (10.29).

10.2.5.4 Example 1

Here, an a constant in the center and first-order autoregressive AR(1) for b are considered. We take $\alpha = -0.15, ac = 1$, the lower and upper bounds $r_1 = -5, r_2 = 5$. As before, this a^2 is a polynomial of degree 8 with the boundary constraints. Fix the shape parameters of the two tails $\xi_1 = \xi_2 = -0.2$.

We estimate the densities by the kernel method. Its bandwidth is chosen by cross-validation. Figure 10.4 illustrates these estimated densities, compared with the continuous-time marginal one from the expression (10.29).

All of them are similar and approach the theoretical one.

More details about the empirical moments and the theoretical moments are shown in Table 10.2.

Δ	Mean	Variance	Skewness	Kurtosis
Continuous	0.000	3.137	0.000	-0.190
1	-0.008	3.449	0.015	-0.262
2	-0.006	3.455	0.011	-0.270
3	-0.012	3.445	0.012	-0.260
4	-0.009	3.457	0.011	-0.281
5	-0.006	3.451	0.012	-0.274
6	-0.016	3.481	0.016	-0.286
7	-0.011	3.492	-0.001	-0.251
8	-0.003	3.400	0.001	-0.267
9	-0.008	3.408	0.017	-0.269
10	-0.009	3.445	0.027	-0.297

TABLE 10.2 – The empirical moments of discretized samples of example 1

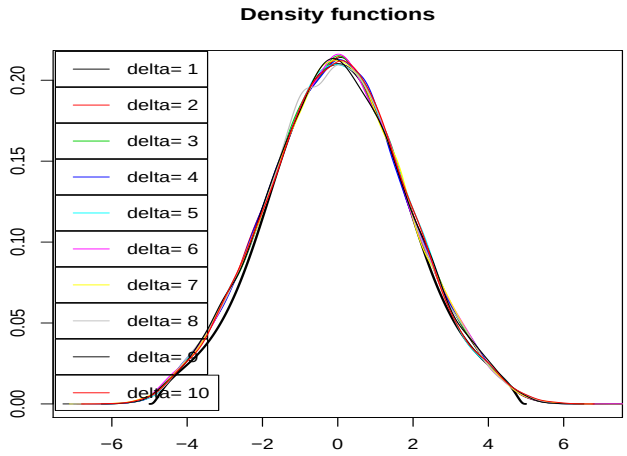


FIGURE 10.4 – Marginal law with the discretization for example 1. The theoretical density function is the black bold line

The tail behaviour will be considered through the values of the shape parameters in stationary GEV models applying to block maxima. These parameters (in the left and right tails) are estimated from all discretized subsequences of the series X. We take the block size of 40, then for $\Delta = 10$, the corresponding sub-series still have 250 values to apply the extreme theory.

10.2.5.5 Example 2

Here, linear a in the center and first-order autoregressive AR(1) for b are considered. $b = \alpha x$ with $\alpha = -0.15$. The values of bounds and shape parameters are the same as in example 1 : r_1 = -5, r_2=5 and $\xi_1 = \xi_2 = -0.2$. The same extreme parameters are taken on purpose to compare the tail behaviours of two examples. a is always a polynomial of degree 8 with the constraints (see Figure 10.6). The same procedure than in example 1 is realized. We also estimate the scale parameters in both the right and left tails. The results are in Figure 10.7, Table 10.3 and Figure 10.8.

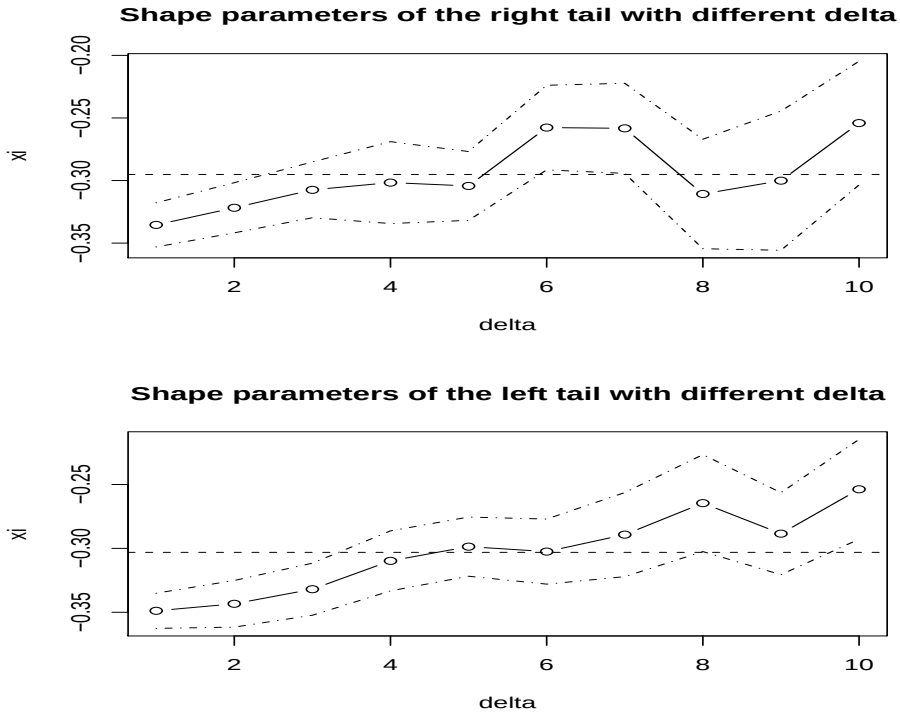


FIGURE 10.5 – Behaviour of shape parameters with the discretization for example 1. Upper panel : the right tail. Lower panel : the left tail. The dashed horizontal line is the average value of ξ's.(The dotted dashed lines are the 90% confidence interval

10.2.6 Example 3

Here the same a as in example 2 and third-order autoregressive AR(3) for b are considered.

$$X_t = \sum_{i=1}^{3} \theta_i X_{t-i} + a(X_{t-1})\varepsilon_t \tag{10.30}$$

For instance, we take for the coefficients $\theta_1, \theta_2, \theta_3$ the values 0.85, -0.4, 0.3, respectively. This model is not a usual discretized approximation of the continuous-time diffusion process (10.28). In modeling the temperature, it is usually more suitable to adjust the temperature by an autoregressive model with an order higher than one (three for daily mean temperature). We then consider the discretization characteristics of (10.30). The same procedure will be explored with the aim to determine if the discretization properties of the diffusion process are still valuable. In this case, we don't have an explicit formula to calculate the theoretical statistical elements because there is not

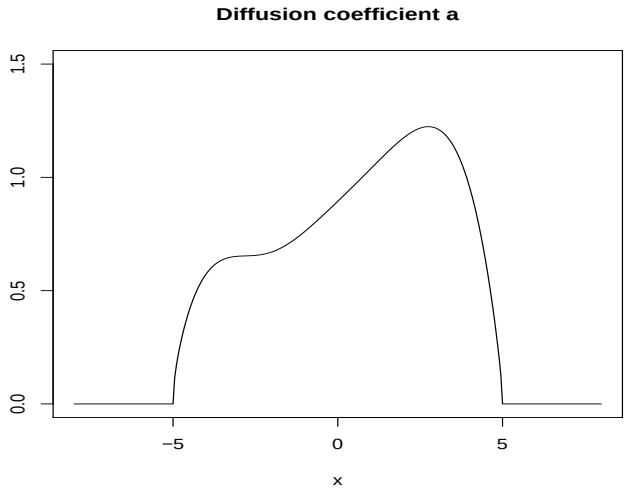


FIGURE 10.6 – Diffusion coefficient of example 2

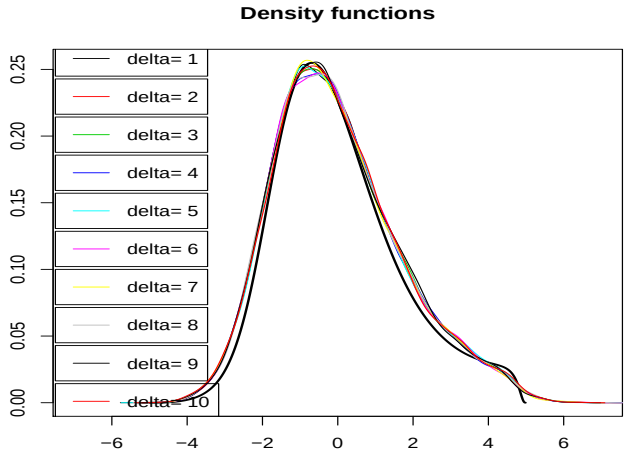


FIGURE 10.7 – Marginal law with the discretization for example 2. The theoretical density is the black bold line.

an Euler scheme to connect the autoregressive model with a continuous-time diffusion. The results of this discretization can be found in Figure 10.9, Table 10.4 and Figure 10.10

Δ	Mean	Variance	Skewness	Kurtosis
Continuous	0.000	2.500	0.640	0.300
1	-0.019	2.871	0.582	0.127
2	-0.019	2.866	0.581	0.129
3	-0.021	2.861	0.576	0.116
4	-0.016	2.863	0.588	0.123
5	-0.020	2.865	0.594	0.165
6	-0.017	2.872	0.584	0.138
7	-0.025	2.836	0.563	0.105
8	-0.025	2.849	0.603	0.182
9	-0.019	2.870	0.564	0.068
10	-0.030	2.846	0.612	0.202

TABLE 10.3 – The empirical moments of discretized samples of example 2

Δ	Mean	Variance	Skewness	Kurtosis
1	0.0159	1.802	0.499	0.328
2	0.0168	1.809	0.496	0.311
3	0.014	1.804	0.491	0.363
4	0.0217	1.806	0.470	0.210
5	0.008	1.780	0.508	0.368
6	0.021	1.826	0.484	0.340
7	0.009	1.821	0.508	0.301
8	0.014	1.809	0.479	0.174
9	0.012	1.795	0.484	0.354
10	-0.006	1.760	0.526	0.438

TABLE 10.4 – The empirical moments of discretized samples of example 3

10.2.6.1 Discussion

The marginal densities do not change in their central part with the discretization of different kinds of models : first-order autoregressive models AR(1), longer memory process (AR(3)) for b, and a constant or linear in the central part. Moreover, in examples 1 and 2, where the models are a discrete version of the continuous-time diffusion process by the Euler scheme, the estimated density functions are really close to the theoretical density function. Usually the theoretical density has a more bounded support than the discretized versions.

The estimators of the shape parameters in the different extracted subsequences are close to one another in both tails. They vary around their average value with a small variance (nearly zero).

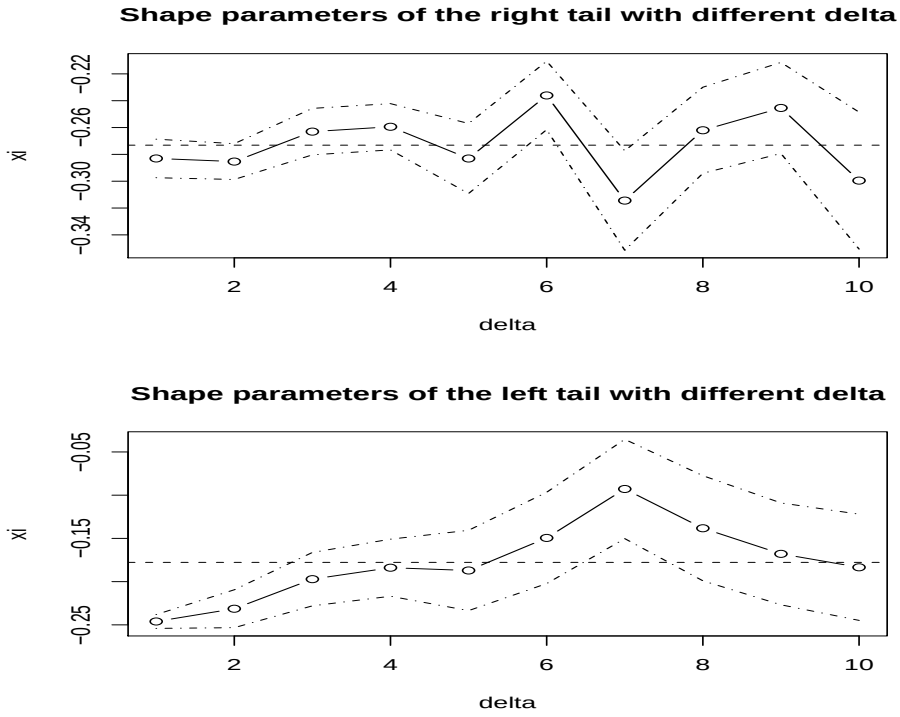


FIGURE 10.8 – Behaviour of shape parameters with the discretization for example 2. Upper panel : the right tail. Lower panel : the left tail. The horizontal line is the average value of these ξ. The dotted dashed lines are the 90% confidence interval

The tail behaviour not only depends on the boundary constraints, but also on conditional mean b and conditional variance a^2. As we have seen, in all three examples, the boundary constraints are the same, however each example gives different values of shape parameters. In the first example, when a is symmetric, the behaviour of both tails is the same. On the other hand, in other examples, the two tails are different due to the asymmetry of a. Both examples, 2 and 3, have the same a, but different conditional mean (AR(1) and AR(3)), and this also results in a difference in the tail behaviour. Example 3 gives heavier tails than example 2.

We can insist on this (almost) trivial conclusion : If one wants to have a good estimation in the tails, he must have a good estimation in the central field.

In the next chapter, we will apply a discrete-time diffusion, exactly

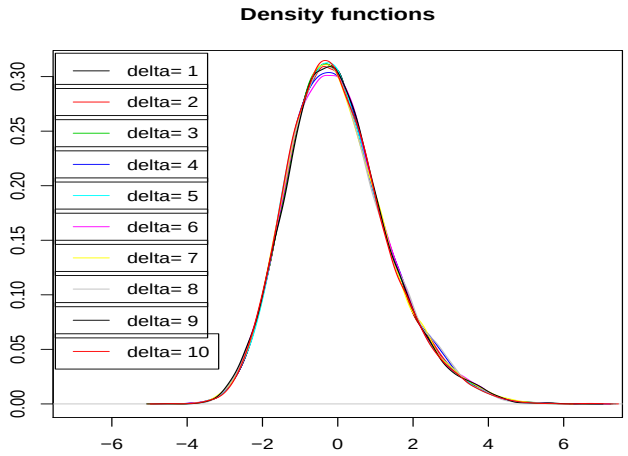


FIGURE 10.9 – Marginal law with the discretization for example 3. The theoretical density is the black bold line

FARCH model (functional autoregressive model with heteroscedastic variance), on the temperature series. A complete estimation for this kind of model, where the seasonality and the boundedness of the data are taken into account, will be described in detail. This chapter prepared theoretical bases for the next chapter.

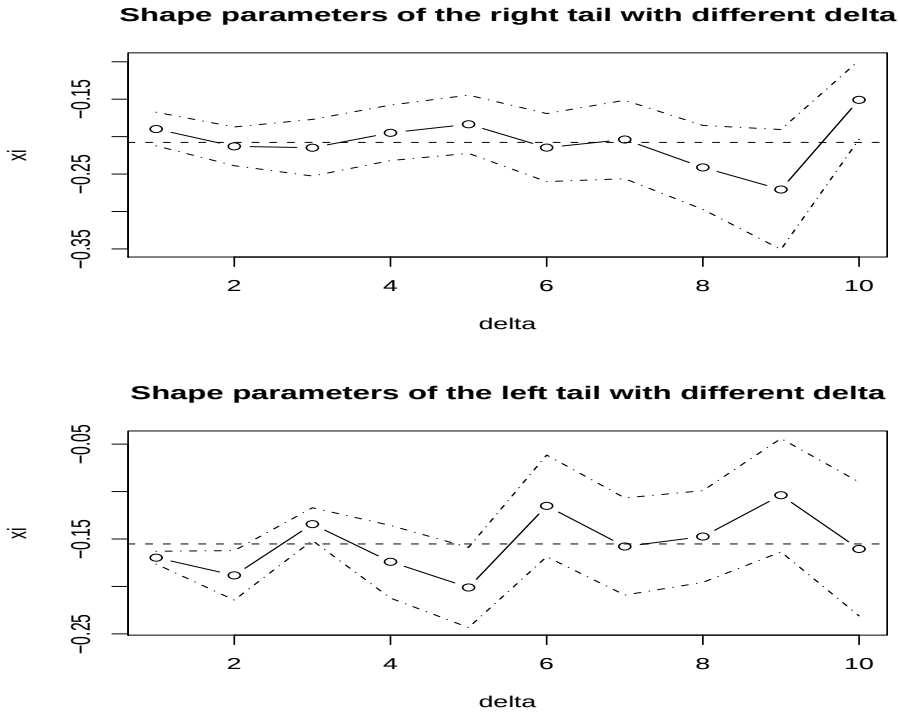


FIGURE 10.10 – Behaviour of shape parameters with the discretization for example 3. Upper panel : the right tail. Lower panel : the left tail. The horizontal line (dashed line) is the average value of these ξ. The dotted dashed lines are the 90% confidence interval

Chapitre 11

Modeling observed temperature series, criteria of validation, simulations

11.1 Introduction

When working on temperature time series, we separate a deterministic and a random component. The deterministic component can be a trend which presents a long-term evolution, or a periodic function, or a composition of these. The random component will be named from now on "the reduced series", after having been normalized by the standard deviation. It is a random process with zero mean, unit variance and has some stationary properties, but in the extended sense. The random component reflects the intrinsic variability of the observed phenomenon around his trend and seasonality, and perhaps errors of measurement.

Starting with the purpose of building models for the temperatures, the general method is, as often, to find a white noise who generates the reduced series. The point is that when we have such a "noise", we have extracted all the useful information on the mechanism generating the temperature. The mechanism is here a purely statistical relation without physical content.

Air temperatures are measured only at discrete times, which results in discrete time series. There are a lot of theories on continuous time series, that have not an easy translation in terms of discrete ones. Therefore, the model for the reduced series must be chosen with care and its goodness-of-fit needs to be checked through simulations.

As previously, we consider temperature series as a realization of additive process with a heteroscedastic variance :

$$X_t = m(t) + S(t) + s(t)S_V(t)Z_t, \tag{11.1}$$

where $m(t)$ is a slowly changing function, called "trend", $S(t)$ and $S_V^2(t)$ are seasonal components in mean and variance with chosen period length. s_t^2 is the trend in variance and Z_t is the reduced series of zero mean and unit variance.

The first step is to understand if there remains statistically in Z some significant characteristics which are non stationary. And the first question is : is Z_t "periodic stationary" ? In this case we generalize the stationarity by the annual (or monthly) periodicity. As discussed in Chapter 9, the seasonality is difficult to completely remove by classical methods. Indeed, many statistics of Z_t are affected by periodic phenomenon, which will be shown by the following detailed studies. There are not many efficient statistical tools for testing the stationarity of this kind of time series even if there exists numerous tests for the classical stationarity. They can be generalized for the periodic stationarity when cutting the series in seasons. But these tests, in general, are not often robust in this domain (a detailed study is Cordery and

Nazemosadat, [24]). Another effective way to test the periodic stationarity of Z_t, which is considered in this chapter, is to model Z_t by a periodic stationary model and then consider the goodness-of-fit of the model through simulations (or bootstrap).

Let X_t be an observed series of temperature. We are interested in the following series :

1/ temperatures at a fixed hour.

2/ daily mean temperatures.

3/ daily maximum temperatures.

4/ daily minimum temperatures.

When applying the discrete-time model on these kinds of temperatures, the lag of discretization Δ shows less physical senses. For the temperature measured at a fixed hour, Δ still has its meaning. But the other series, for example daily mean temperatures, where we take the average value of all measures in a day, cannot be considered as a discrete-time temperature series with a fixed lag Δ. In practice, this fact has some influences on the choice of models.

Basic features are the dynamics of the seasonality and the non-linear dependence on the state of the variability (often called now in other fields "volatility") and the boundedness of the temperature (as mentioned in the previous chapter). When the dynamics of reduced temperatures is seasonal, the conditional mean and variance need to be periodic functions for a given value of the temperature. They depend on both the dates $[t]$ in the year (seasonality) and the state (temperature) Z_t. Moreover, constraints on the boundaries of the distribution of Z_t have to be translated in terms of zero-variability outside the interval of possible values of Z_t. These problems are often ignored but one must realize that they are the main source of practical difficulties. All these points will be taken into account in our model. The performance of the model will be seen through simulations.

11.2 Reduced temperatures and their characteristics

11.2.1 How obtain the reduced series ?

We call back the procedure to obtain the reduced series Z. We rewrite the process as :

$$X_t = m(t) + S(t) + s_t S_V(t) Z_t$$

where t is the dates, $m(t), s^2(t)$ are respectively the trends in mean and in variance. $S(t)$ and $S_V^2(t)$ are the seasonality components in mean and in variance.

We then obtain the reduced series by the following steps :

• estimate the seasonality in mean by a trigonometric polynomial where the degree p of trigonometric terms is chosen by Akaike criterion.

• estimate m_t by loess from the series $X(t) - \hat{S}(t)$. The estimate method is described in Chapter 3.

• estimate the seasonal term $S_V^2(t)$ from the squared residuals $(X(t) - \hat{S}(t) - \hat{m}(t))^2$ with the same way as $S(t)$.

• estimate the variance function $s^2(t)$ by loess from the squared residuals

$$\left(\frac{X(t) - \hat{S}(t) - \hat{m}(t)}{S_V(t)} \right)^2$$

.

•

$$Z(t) = \frac{X(t) - \hat{m}(t) - \hat{S}(t)}{\hat{s}(t)\hat{S}_V(t)} \tag{11.2}$$

is the reduced series with a zero mean and a unit variance.

Remark : For the purpose of constructing simulation models for the temperature, many writers modeled the trend m by a parametric model, often by a linear trend. Following our studies, this linear trend shows its insufficiency and inadequacy. Even when a seasonally linear trend (with seasonal coefficients) $\alpha(t)t + \beta(t)$ is used as an estimator, the trend in mean is not completely detected. The result of this is that Z_t still has a significant trend, which can be seen by simple graphics.

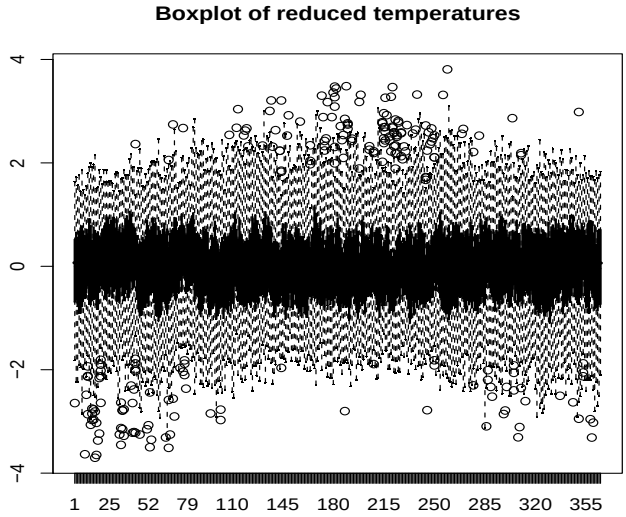


FIGURE 11.1 – Boxplot of the reduced temperature Z_t for a year

11.2.2 Seasonality effect on the reduced series

The reduced series Z_t is in fact not stationary in an usual way, but in an extended sense : cyclic or periodic seasonality as we show in the following studies.

We use the daily mean temperature series in Bordeaux during the period 1950-2008. The previous estimation procedure will be used to obtain the reduced series $Z(t)$.

As mentioned in Chapter 9, the seasonality has its dynamic which cannot be easily removed. Therefore, after removing the trend and seasonality components in the mean and the variance of the observed series X_t, the seasonality still remains in the reduced series Z_t. This behaviour can be firstly seen in the boxplot of Z_t (Figure 11.1) for a year.

The central field of Z_t is reasonably modeled by an autoregressive model of order p, AR(p), because Z_t has a conditional mean *linearly dependent* on its past values. For instance, an AR(p) with $p = 3$ (chosen by the Akaike criterion) is generally suitable to model daily mean temperature series. An intuitive way to see better this linear dependence is to model Z_t by a generalized additive model (GAM) of the form :

$$Z_t = f_1(Z_{t-1}) + f_2(Z_{t-2}) + f_3(Z_{t-3}) + \varepsilon_t \tag{11.3}$$

where f_1, f_2, f_3 are the non-parametric regression functions, which will be

estimated by spline method.

The linear dependence of Z_t on its past values is seen in Figure 11.2.

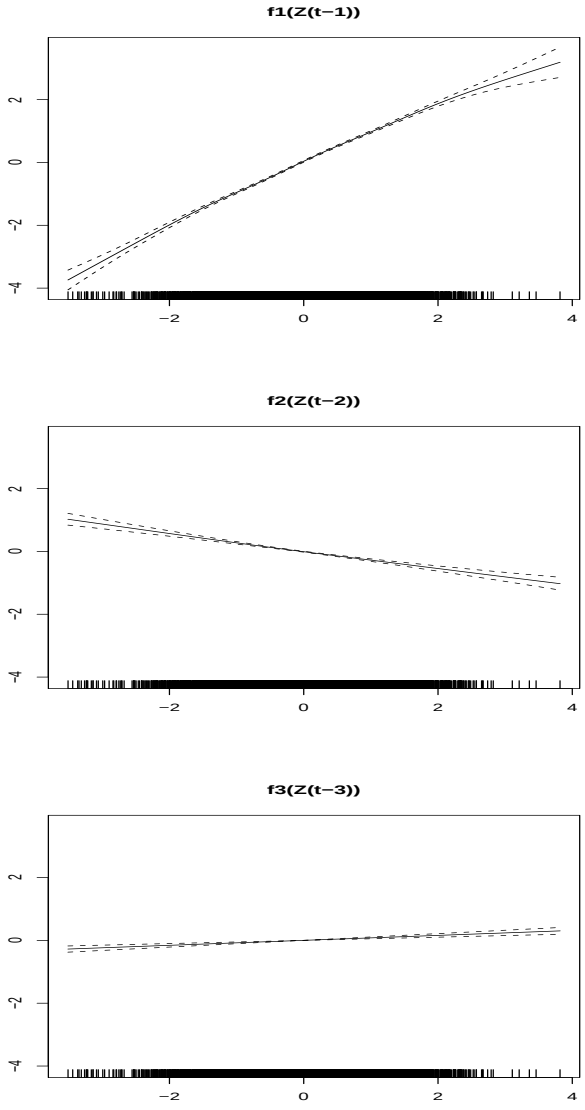


FIGURE 11.2 – Estimation by GAM method (by cubic splines) of the dependence of Z_t on $Z_{t-1}, Z_{t-2}, Z_{t-3}$.

Now, to consider the seasonality in the central field, we divide $Z(t)$ in twelve months, then model each sub-series by an AR(3). The curves of estimates for 12 months are shown in Figure 11.3.

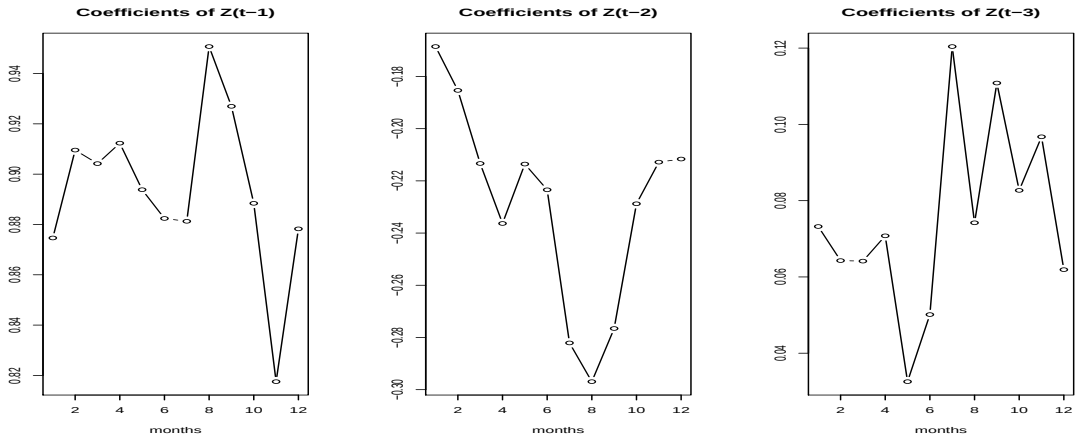


FIGURE 11.3 – Estimation of AR(3) for reduced temperatures of 12 months.

Figure 11.3 shows that the coefficients of the AR(3) change with time. The maximum difference between the coefficients of $Z(t-1)$ and those of $Z(t-2)$, that we can find, is 1.2 ; for those of $Z(t-3)$, it is 0.08. This confirms that the seasonality still remains in the mean of reduced temperatures. With this characteristic, Z can be modeled by a periodic autoregressive model of order p, PAR(p). We can write the model of Z_t as follows :

$$Z_t = \sum_{k=1}^{p} \left[\theta_{0,k} + \sum_{j=1}^{d} \left(\theta_{1,k}^{j} \cos \frac{2j\pi t}{365} + \theta_{2,k}^{j} \sin \frac{2j\pi t}{365} \right) \right] Z_{t-k} + \varepsilon_t \qquad (11.4)$$

where $\theta_{.,k}^{j}, 1 \leq k \leq p$, $1 \leq j \leq d$, are seasonally-varying autoregressive coefficients, p is the autoregressive order, d is the number of trigonometric terms and ε_t is a white noise.

For the series of daily mean temperature in Bordeaux, we get $d = 3$ and $p = 3$. Figure 11.4 illustrates the seasonality of the dynamics of the autoregressive coefficients. The estimated curves for the seasonal autoregressive coefficients are similar to the ones that we obtained when we divided the reduced series in 12 months. This signifies that this model is reasonable.

The periodic phenomenon is also seen in the variance term. If we consider the variance of each monthly series as constant, estimating it from the squared residuals of the AR(3) of the series corresponding to this month, this constant varies with time, which can be seen in Figure 11.5.

When we consider the conditional variance as a function of Z_{t-1}, noted $\mathrm{Var}(Z_t \mid Z_{t-1})$, by estimating it by cubic splines, we obtain different curves

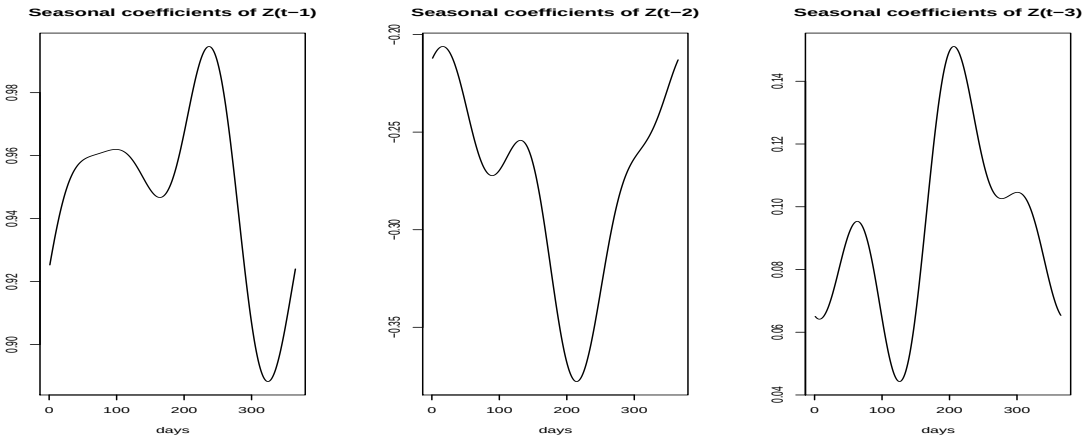


FIGURE 11.4 – Estimation of the coefficients of seasonal AR(3) for the reduced temperature Z_t.

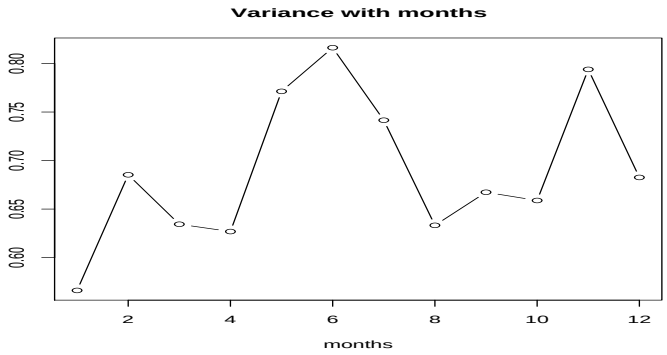


FIGURE 11.5 – Estimation of the variance (supposed constant) from the squared residuals of AR(3) applied for each month.

for different months (see Figure 11.6). In hotter months (May, June, July, August), a is linear with an increasing tendency, whereas in colder months (November, December, January, February) it has a decreasing trend. Of course, these estimators do not have much sense near the observed or estimated bounds due to the lack of data. One point is important for the application : the variance of the hot extreme events is much higher than the variance of the cold extreme events in June and July. This difference in these two months is much more important than in the remaining months of the year.

Consider now the tail behaviour of Z_t in every months to see if the seasonality has its effect on extreme events. In each month, the series are rather

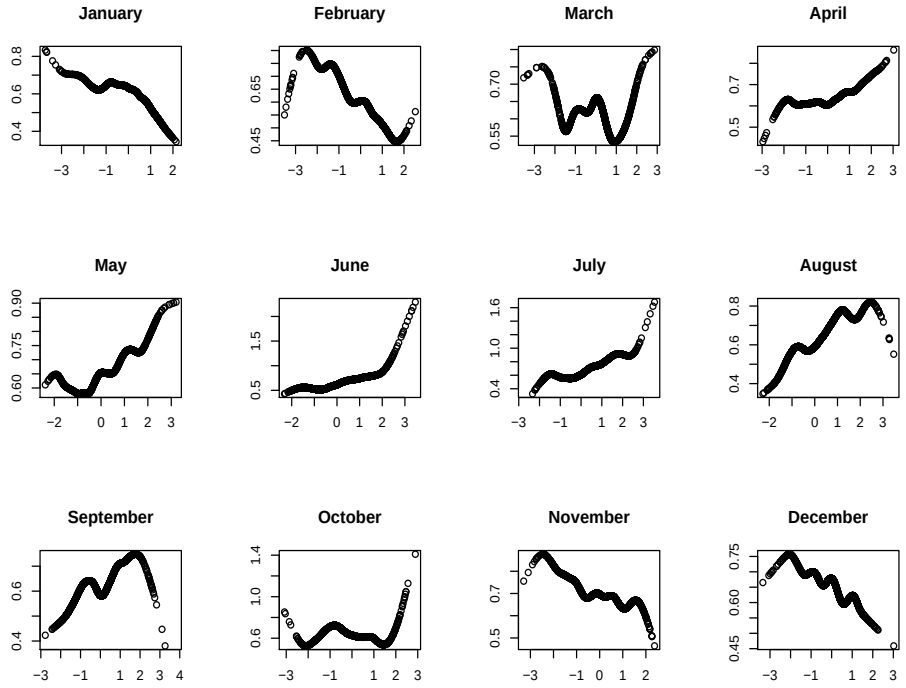


FIGURE 11.6 – Nonparametric estimation (cubic splines) of the conditional variance $Var(Z_t \mid Z_{t-1})$ for each month from the squared residuals of AR(3).

stationary and extreme theory can be applied. For each monthly sub-series, we model block maxima and block minima by a stationary GEV(μ, σ, ξ). The size of blocks here is 30. Indeed, in order to apply extreme theory, the size of block must be sufficiently high, however in our case for this size, we usually have only 59 values (for 59 years), which is not sufficient to build an extreme model. Therefore, for each monthly series, we take the data of 15 days before and after this month to estimate extreme parameters for this month. This method is used throughout the chapter.

Applying a stationary GEV model for every sub-series, we obtain successfully estimators of the three parameters μ, σ, ξ for these sub-series. The curves of estimators and their 95% confidence interval (built using the asymptotic normality of the extreme parameters) in Figure 11.7 show that the seasonality still remains in extreme events. ξ has a very weak seasonality.

This seasonality effect in extremes also needs to be taken into account in the models.

After these preliminary studies, we can see that the periodic cycle still remains in the reduced series. So when considering the stationarity of Z_t,

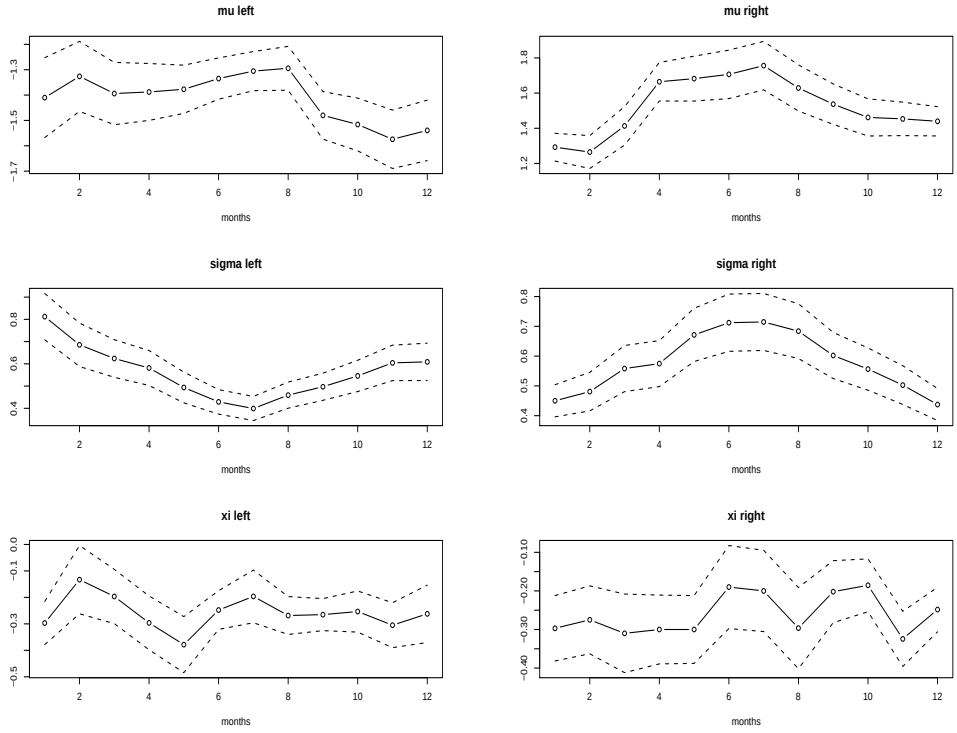


FIGURE 11.7 – Estimators of GEV extreme parameters(solid lines) for reduced temperatures of 12 months and their 95% confidence intervals (dashed lines). The left panel shows those for the left tail. The right panel shows those for the right tail.

it is necessary of take into account its periodic characteristic. Model (11.4) for Z_t is in fact called "periodic autoregressive model"(PAR) (see Franses and Paap, [55]). Periodic time series of this kind are in fact non-stationary models as the parameters take different values at different dates.

11.3 Modeling reduced series of air temperatures

Some discrete models have been presented in the previous chapter. In this section, we will apply these models on different kinds of temperature series. A detailed study of the specific characteristics of these series, using simulations, will be given. The behaviour of the parameters in discrete-time diffusion models is different for each series. This complete study will give us a general look on the dynamics and different intrinsic characteristics of the temperature.

We consider in this section temperature series in Bordeaux. Bordeaux is a city in the South-West of France. Its climate is of oceanic-type, characterized by mild winters and hot summers. The summer rainfall is often caused by storms heat. Bordeaux has in average 15 to 20 days in summer when temperatures exceed 30°C. In the summer of 2003, temperatures reached 41°C and there were 12 consecutive days when the maximum reached or exceeded 35°C.

We have applied the same method to other series, with different climate and different length including series covering two centuries of observations. We show here in the framework of our thesis only the detailed results of temperatures in Bordeaux. The other temperature series give corresponding results.

We will consider different kinds of temperatures : at a fixed hour, daily mean, daily maximum or daily minimum in Bordeaux. After removing the trend and seasonal components, reduced series are obtained. Our purpose is to find a suitable model for these reduced series.

After the previous preliminary study on the characteristics of the reduced series, we can say that a reasonable model of the temperature series is :

$$Z(t) = \sum_{k=1}^{p} \left[\theta_{0,k} + \sum_{j=1}^{p_1} \left(\theta_{1,k}^{j} \cos \frac{2j\pi t}{365} + \theta_{2,k}^{j} \sin \frac{2j\pi t}{365} \right) \right] Z(t-k) + a(t, Z_{t-1})\varepsilon_t,$$
$$\varepsilon_t \sim \mathcal{N}(0,1) \tag{11.5}$$

where $a(t, Z_{t-1})$ needs to be taken as a function of both t and Z_{t-1}. If a is taken as a parametric model, this model needs to have the seasonal coefficients.

We choose as an estimator of a^2 a high-order algebraic polynomial (of degree 5) with respect to the state (the temperature) to allow flexible fits. The function a^2 has seasonal coefficients and then has the following form :

$$a^2(t, Z_{t-1}) = \sum_{k=0}^{5} \sum_{j=1}^{p_2} \left(\theta_{1,k}^{j} \cos \frac{2j\pi t}{365} + \theta_{2,k}^{j} \sin \frac{2j\pi t}{365} \right) Z_{t-1}^{k}$$

In order to estimate this process, we follow the following steps :

• estimate the coefficients in AR(p) by the least squares method, by considering a as constant. The parameters p and p_1 chosen are those which give the minimum AIC value for the model.

• estimate $a^2(t, Z_{t-1})$ by the maximum likelihood method, supposing $\varepsilon_t \sim \mathcal{N}(0,1)$. Sometimes, due to too-much-parameters fact, this estimation

does not converge. In this case, the least squares method can be used by regressing on the squared residuals from the previous step. This way of estimate still guarantees that the noise has zero mean and unit variance.

In order to take the boundedness into account, we take zero as values for a outside $(r_1,\ r_2)$. Concretely, the values of the function $a(t, Z_{t-1})$ for each t are that estimated by maximum of likelihood inside the interval $((r_1+m_1)/2, (r_2+m_2)/2)$ with r_1, r_2, m_1, m_2 are respectively the lower, upper bound and observed minimum and maximum values. Outside this interval, we take (in order to simplify this step) a polynomial $g(x) > 0$ of degree 2 for each tail with three constraints to connect with the central previous polynomial a^2 and to satisfy the condition at the bounds. As an example, for the right tail, these constraints are $g(r_2) = 0$, $g((r_2+m_2)/2) = a^2(t, (r_2+m_2)/2)$ and $g'((r_2 + m_2)/2) = \min((a^2)'(t, (r_2 + m_2)/2), 0)$. The constraints on the first derivative gives the continuity at the connected points and guarantee that after these points, a will continue to decrease until the boundary. The constraints on the first derivatives of a^2 at the bounds, which can be computed from the theory in the case of a fixed hour temperature (see [77]) are ignored here for they are not clearly given in the case of the mean temperature.

Let us now come back to the bounds. As we have seen before, the extreme parameters are seasonal. Thus r_1 and r_2 are also seasonal when one calculate them from stationary GEV models (for $\xi < 0$) by, for example, $\hat{r}_2 = \hat{\mu} - \hat{\sigma}/\hat{\xi}$. Then the constraints on g must be seasonal. Here we let r_1, r_2 and m_1, m_2 change with months. r_1, r_2 can be estimated by applying a stationary GEV model to 12 sub-series corresponding to 12 months. m_1 and m_2 are taken as the maximum and minimum values of each month. Outside the interval (r_1, r_2), $\hat{a}$ equals to zero. This monthly approximation is done only as a simplification avoiding calculations for each t.

Consider now the form of this estimator of $a(t, Z_{t-1})$ (with "the boundary extension") for different dates t. We take the tenth day in each month and draw corresponding volatility $\hat{a}$ in Figure 11.8. The estimated curves of a are rather similar than the nonparametric estimators of monthly sub-series in Figure 11.6, except for regions near the boundaries. As we mentioned before, the nonparametric estimators have less sense near the boundaries due to the lack of data.

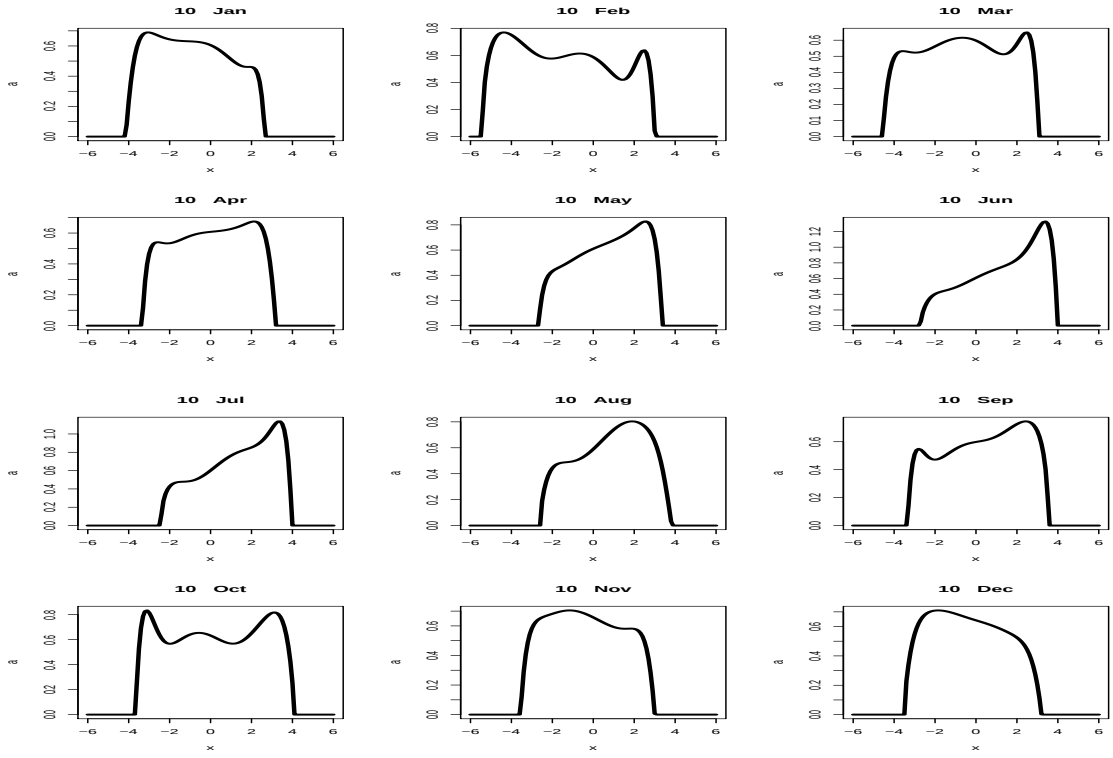


FIGURE 11.8 – Parametric estimators of $a(t, Z_{t-1})$ for different dates t with the boundary extension.

11.3.0.1 Validation of the model and comparison with simpler other models

The validation of the model will be considered through the residuals $\hat{\varepsilon}_t$ (compared with a Gaussian noise) and the quality of the simulations. Both central and extreme fields will be checked for the simulations.

Let us now consider, at the same time, other models : constant a or a^2 is a trigonometric function $f(t)$ which only depends on the dates, and not on the state $Z(t-1)$,

$$f(t) = \theta_0 + \sum_{k=1}^{p} \left(\theta_{k,1} \cos \frac{2\pi kt}{365} + \theta_{k,2} \sin \frac{2\pi kt}{365} \right)$$

The optimal p is chosen by AIC.

Our model will be compared with these two models. As we will see, following the criteria of the validation, our model really gives a better fit compared with the other models.

First consider the residuals. If the model is good, the residuals must be a Gaussian noise $\mathcal{N}(0,1)$. We will consider its normal Q-Q plot, autocorrelation functions of the residuals and also *squared residuals*. It is well-known that usually the correlation still remains in the variance of the residuals (see [14]). Then a Kolmogorov normality test will also be considered.

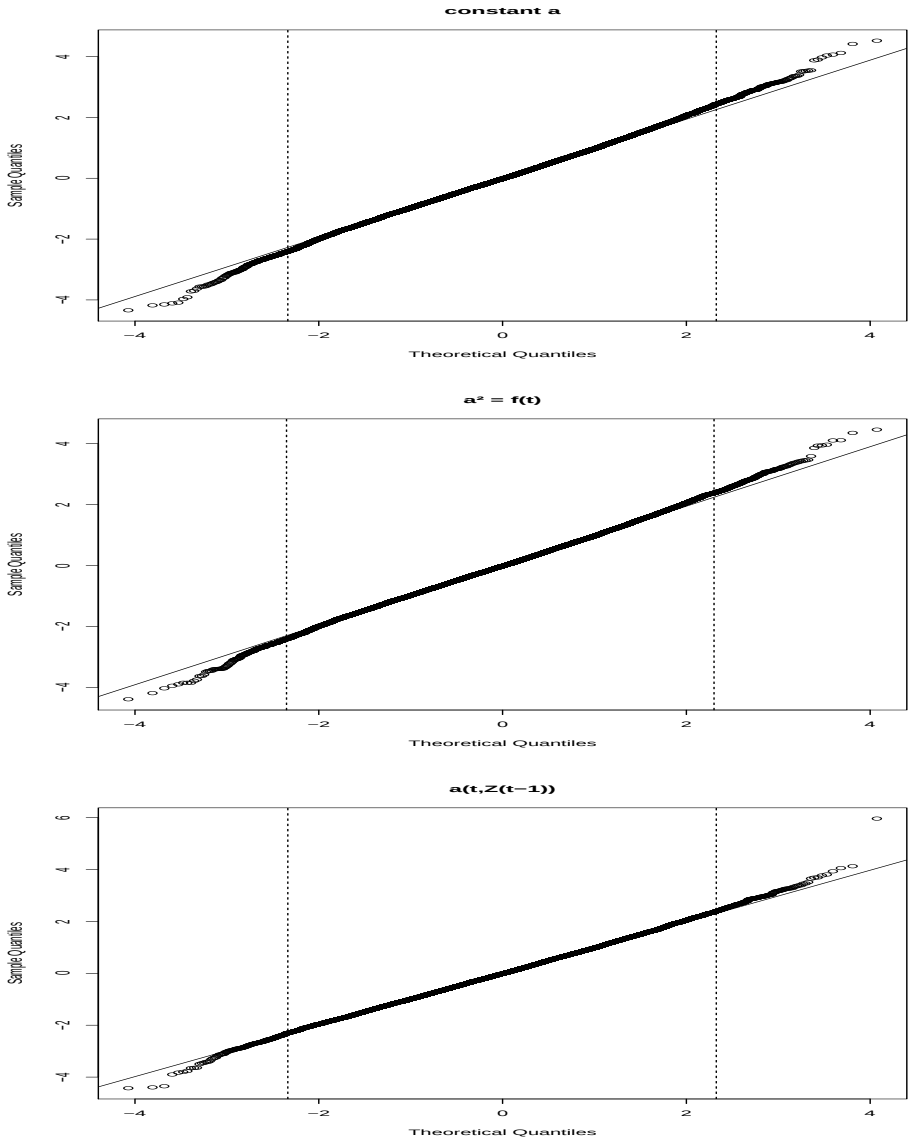


FIGURE 11.9 – Normal QQ-plot of the residuals for different models of a with the 99th and 1st percentiles (*dashed lines*) of $\mathcal{N}(0,1)$.

The normal Q-Q plots in Figure 11.9 show that the residuals of the case of a, depending on the dates and the state, $a(t, Z(t-1))$, have closer quantiles with a normal distribution. Their tail quantiles approach better those of a Gaussian noise.

Now consider the autocorrelation function of the residuals and that of their squared values in Figure 11.10.

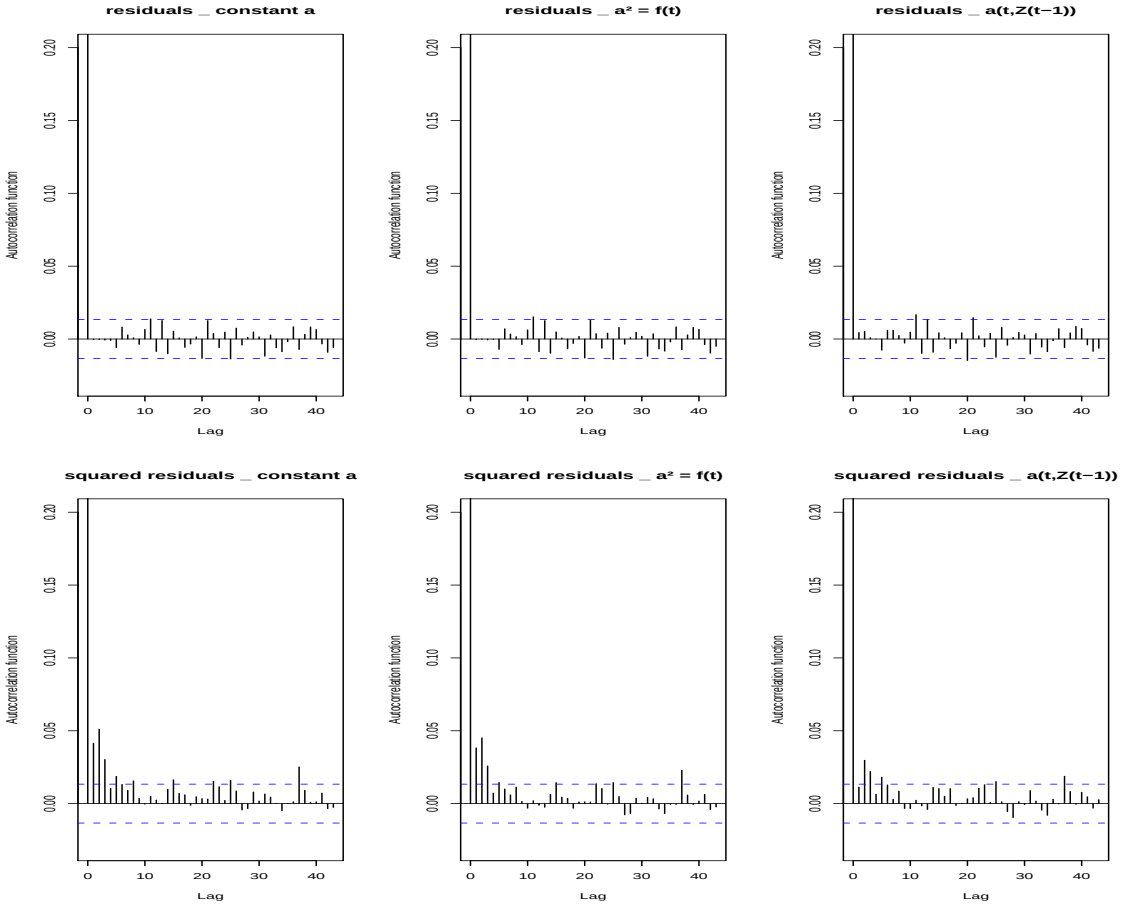


FIGURE 11.10 – Autocorrelation functions of the residuals and their squared values for different a.

All the residuals are white. However, in the cases where a is a constant or a function of dates $f(t)$, it still clearly remains the correlation in the squared residuals. For this reason, when modeling the weather, in many works, a combination of a GARCH noise is proposed, see for example Campbell & Diebold ([14]). In fact this solution is not at all adequate as it gives heavy tail residuals. Our model, with $a(t, Z_{t-1})$, removes effectively the correlation

in the variance of the residuals and this correlation becomes negligible. The most significant value of autocorrelation is 0.025 where the confidence interval for the autocorrelation function of a white noise is [-0.015,0.015]. The squared residuals in this case are nearly white.

The Kolmogorov test for the normality accepts the residuals of $a(t, Z_{t-1})$ as a normal realization with p-value=0.12. It refuses the normality of the other residuals with p-value=0.02 for the constant a and p-value=0.03 for $a^2 = f(t)$.

Now in order to validate these models, we will simulate 100 samples of Z_t with the same length. 100 simulated samples of X_t can be obtained by adding the estimators $\hat{m}(t)$, $\hat{S}(t)$, and multiplying by the estimators $\hat{s}(t)$, $\hat{S}_V(t)$, the simulated samples of Z_t.

To test the quality of the model, we have to determine whether the simulation model is an accurate representation of the observations. The comparison is based on the following items for the series Z_t and sometimes X_t and their simulated samples :
• Density function and empirical moments for whole year or a month
• Quantiles for whole year or a month
• Temperature of a fixed date
• GEV parameters for the tails
• Clusters distribution : number of clusters and length for observation larger than threshold given by the 2% or 98% quantiles.

⋆ *Density function and empirical moments*
Let us first make a very important remark for the practical interpretation of our results. We have

$$X_t = m(t) + S(t) + s_t S_V(t) Z_t$$

thus the distribution of X is a mixture of that of Z with other parameters. The mixing parameters are the location $m(t)$ and the scale $s(t)$. The mixing parameters can be thought as varying very slowly. This can lead to false interpretations. First for instance, the cold extremes come from the cold extremes of the earlier years of the sample rather than those of the last years. Because of the warming effect, these last extremes are "hotter". An other example is that bimodalities can appear at the center of the distribution as a mixture effect , they have no physical sense, only a statistical artefact. Even if we consider temperatures of a fixed date of each year, for example each 10th of January, we cannot obtain something "stationary". Look at Figure 11.11 for an example, temperatures of the 10th of January varies strongly around their mean and it seems warmer in the last years.

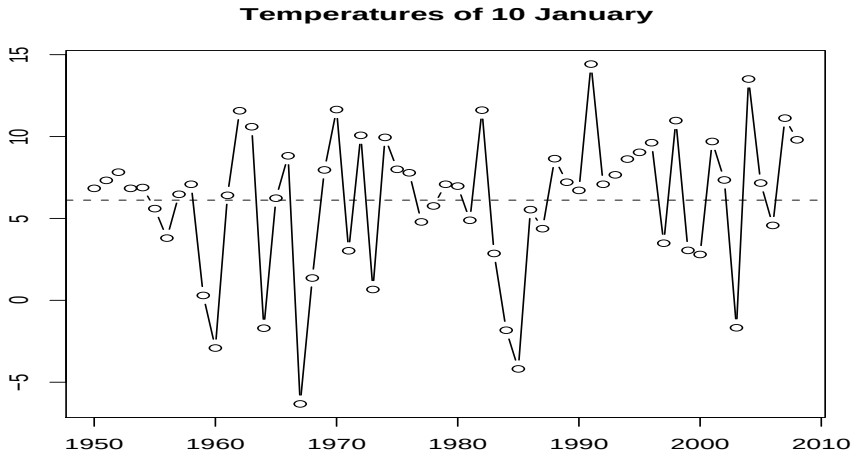


FIGURE 11.11 – Temperatures of each 10th of January of X_t

For this reason, to validate a model, studying the reduced series Z_t gives much more sense than the observed series X_t. There is a second level of mixing in this hierarchical model on time scales due to the seasonality. Year distribution for Z then has no more physical sense, it is a mixture of months distributions to be concrete, or more accurately of days distributions. Corresponding to mixtures, there are different levels of variability : in a day, a month or a year and we use as a practical tool, without justification, values of Z_t in a month as homogeneous data in some procedures or graphs. Sometimes we give graphs of the characteristics of X_t and its simulations in order to validate the model with "clearer visualizations", but we emphasize on the difficulties and possible mistakes in interpretation.

First we consider the density of X_t, thus for a whole year, and also the density functions of a fixed month, for example January (in winter) and July (in summer). The plots of estimated densities and their 95% confidence intervals, built from the different simulation types of X_t, are given in Figure 11.12. Apparently, the confidence interval of the density of a whole year constructed from the simulations of $a(t, Z_{t-1})$ follows better the bimodal characteristic of X_t in the center, and the tail behaviour of these observations. Equally these simulations give a much better feature for the seasonality of the observations. The density of observed temperature in a month are found inside their confidence intervals, which is not the case for the simulations of the other models.

These density functions are considered to give a general look on how these models adjust the observed temperature X_t. However, as mentioned

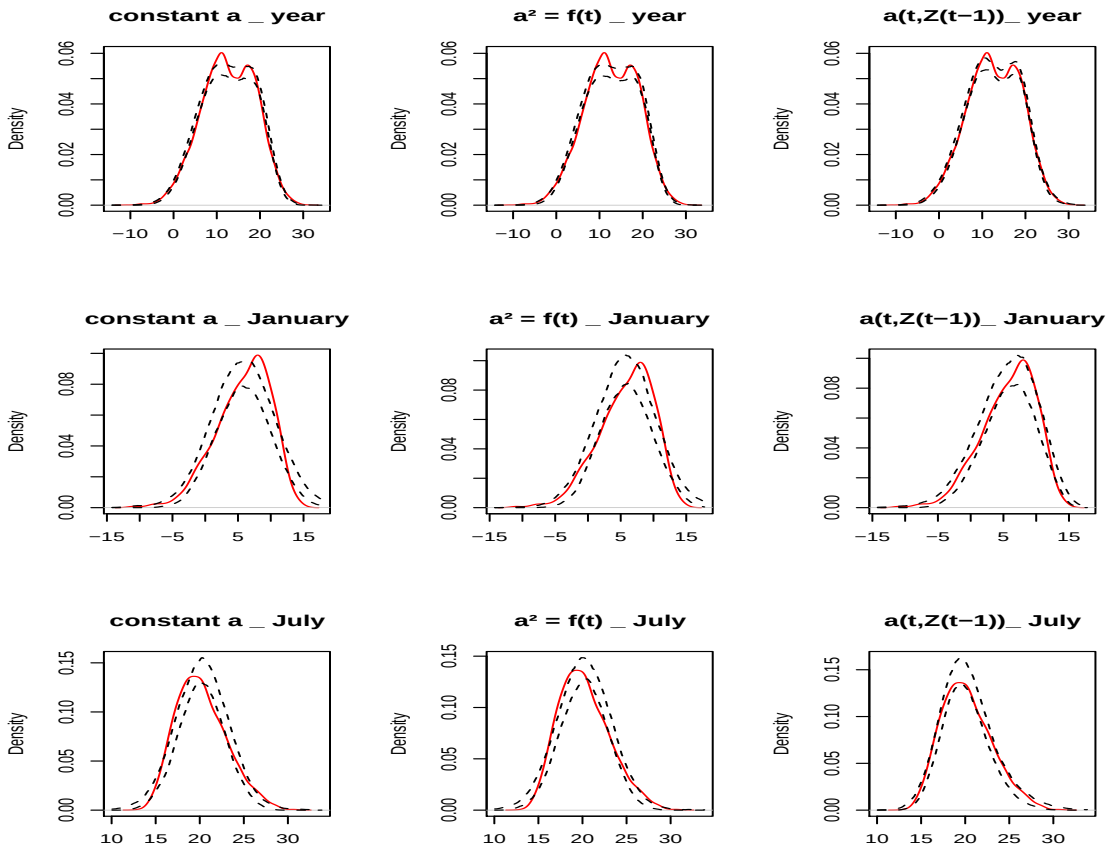


FIGURE 11.12 – Density functions of X_t for whole year, January, July and their different 95% confidence interval with different simulations.

before, the density of X_t does not really have a sense in this situation. In the observed series X_t, which is not stationary, the seasonalities and the trends are mixed. The bimodality that we saw in the center of X_t, in fact, represents the mixture of the seasons : hotter seasons and colder seasons. In this kind of density, we can recognize the hot extremes in the right tail and cold extremes in the left tail, but in the center, all kinds of temperature are mixed. Even for one month of X_t, the trends still exist, then the density estimation can not give a real statistical sense.

A more reasonable validation criterion is to consider the same quantities but on Z_t, which is rather stationary (periodicals stationary). The estimated densities of Z_t for whole year, January and July are shown in Figure 11.13.

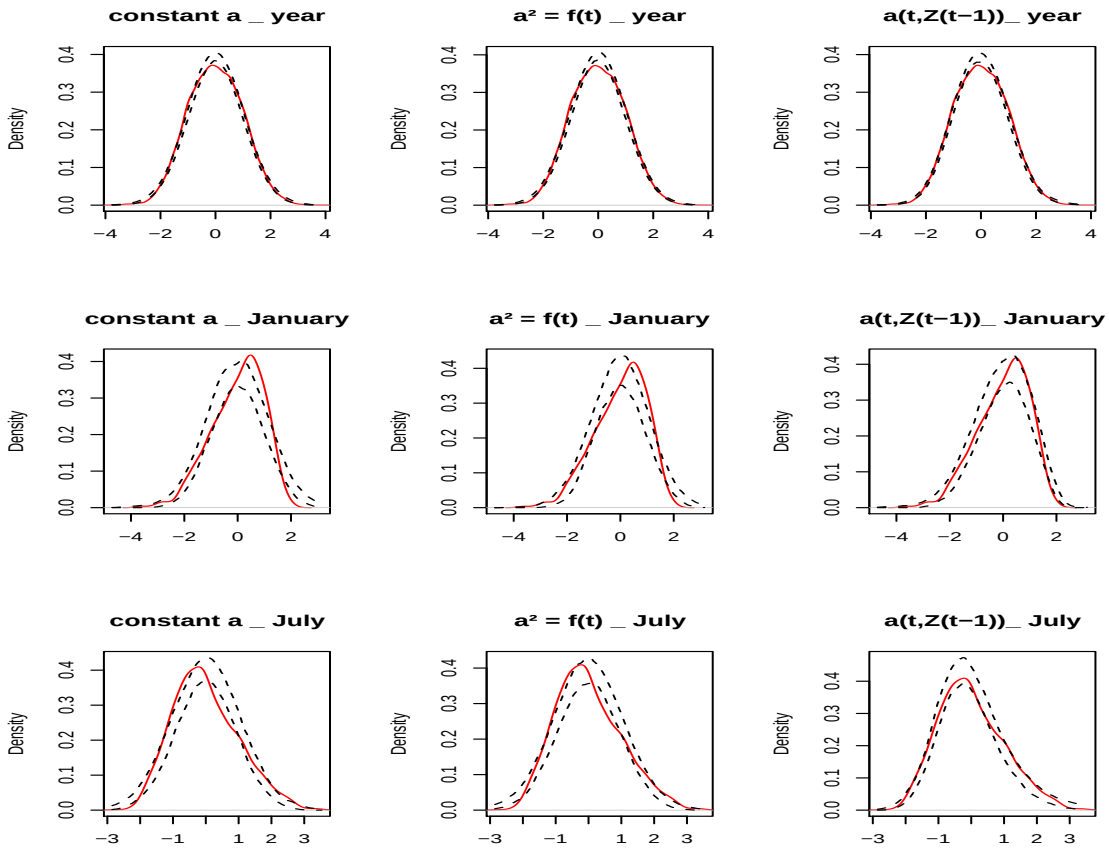


FIGURE 11.13 – Density functions of Z_t for a whole year, January, July and their different 95% confidence interval with different simulations.

Observing Figure 11.13, we see that in the whole year case, the simulations with $a(t, Z_{t-1})$ are a little better than the other simulations, but the difference is difficult to see. However, for a fixed month, the simulations or our model show their clear superiority when almost all the density values of Z_t are found in the simulated confidence intervals.

For a better appraisal, the empirical skewness (due to non- linear effects) and kurtosis (due to large values following a large number of simulations) of these simulations, for January and July of Z_t, are shown in Table 11.1. The empirical skewness of simulations with $a(t, X_{t-1})$ is closer the one of the reduced temperature than those of simulations from two other models. However, none of these simulations can give a good approximation for the kurtosis.

Time	Series	Skewness	Kurtosis
January	Z_t	-0.567	0.009
	Sim., constant a	0.005	0.030
	Sim., $a^2 = f(t)$	-0.004	-0.019
	Sim.,$a(t, Z_{t-1})$	-0.363	-0.093
July	Z_t	0.444	-0.117
	Sim., constant a	-0.003	-0.006
	Sim., $a = f(t)$	-0.011	0.000
	Sim.,$a(t, Z_{t-1})$	0.640	0.690

TABLE 11.1 – Comparison of the empirical skewness and kurtosis for different kinds of simulation.

⋆ *Quantiles*

Firstly, for a general view, we build from the simulated samples the distribution for quantiles of X_t. Consider the distributions built from three models for the 1%, 2%, 3%, 10%, 30%, 50%, 70%, 90%, 95%, 97%, 98% and 99% quantiles. The observed quantiles with their different distributions from different simulations are shown in Figure 11.14. For instance, if the observed quantile is found in the simulated quantile distribution, we say that the simulations give approximately a good quantile distribution. In this sense, we can see that our model give better quantile distributions for the observed quantiles. Particularly, for higher quantiles (97%, 98%, 99% ones), the observed quantiles are found outside of other simulated distributions, but are always found inside the quantile distribution built from the simulations with $a(t, Z_{t-1})$. The quantiles for the center (30%, 50% ones) are not very good, probably due to the mixture of the seasonality and trend. However, after all, the difference is not really high. The empirical value for these quantiles calculated from the simulations of our model has a difference of 0.12°C with the observed quantiles.

For more statistical sense, we consider now the quantiles in a fixed month, for example February and July for Z_t, to have a stationary distribution. The results are shown in Figure 11.15 and 11.16. The results for our model are clearly better, for both February and July.

⋆ *Temperature of a fixed date*

We now pass on another criterion of validation. From the simulations, we will build a distribution for the observed temperature of a fixed date. For example, we fix t as 10th of January 2000. The method to build the distribution for its temperature value is :
- take in 100 simulated samples of reduced temperatures Z_t all values cor-

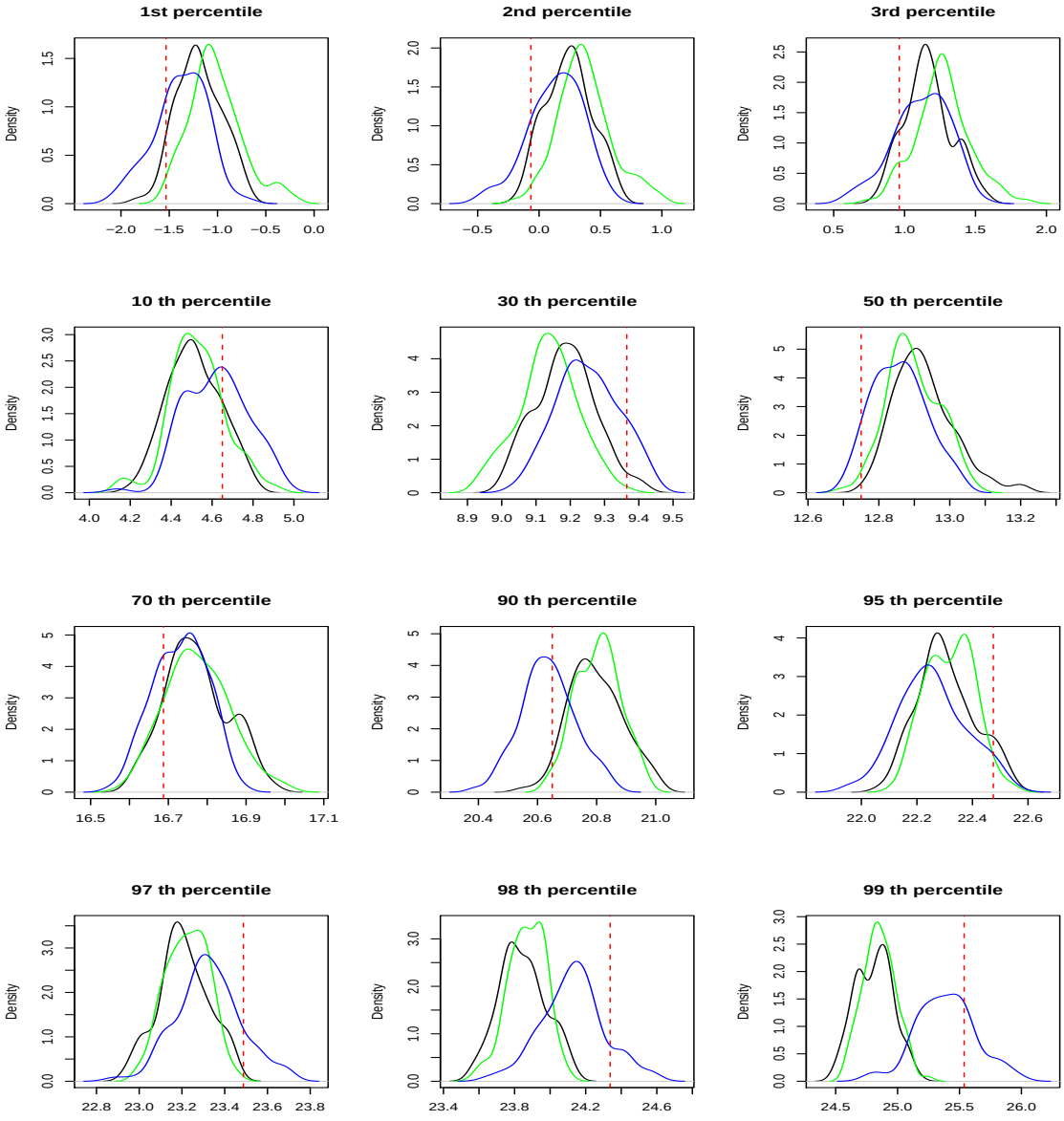


FIGURE 11.14 – Observed quantiles (red vertical lines) for whole year of X_t and their distributions built from the simulations of different models : in black, model with constant a, in green : models with $a^2 = f(t)$, in blue, model with $a(t, Z_{t-1})$.

responding to the date 10th of January.
- add and multiply this sequence with the deterministic components corresponding to $t =$ 10th of January 2000, $\hat{m}(t), \hat{S}(t), \hat{s}(t), \hat{S}_V(t)$.

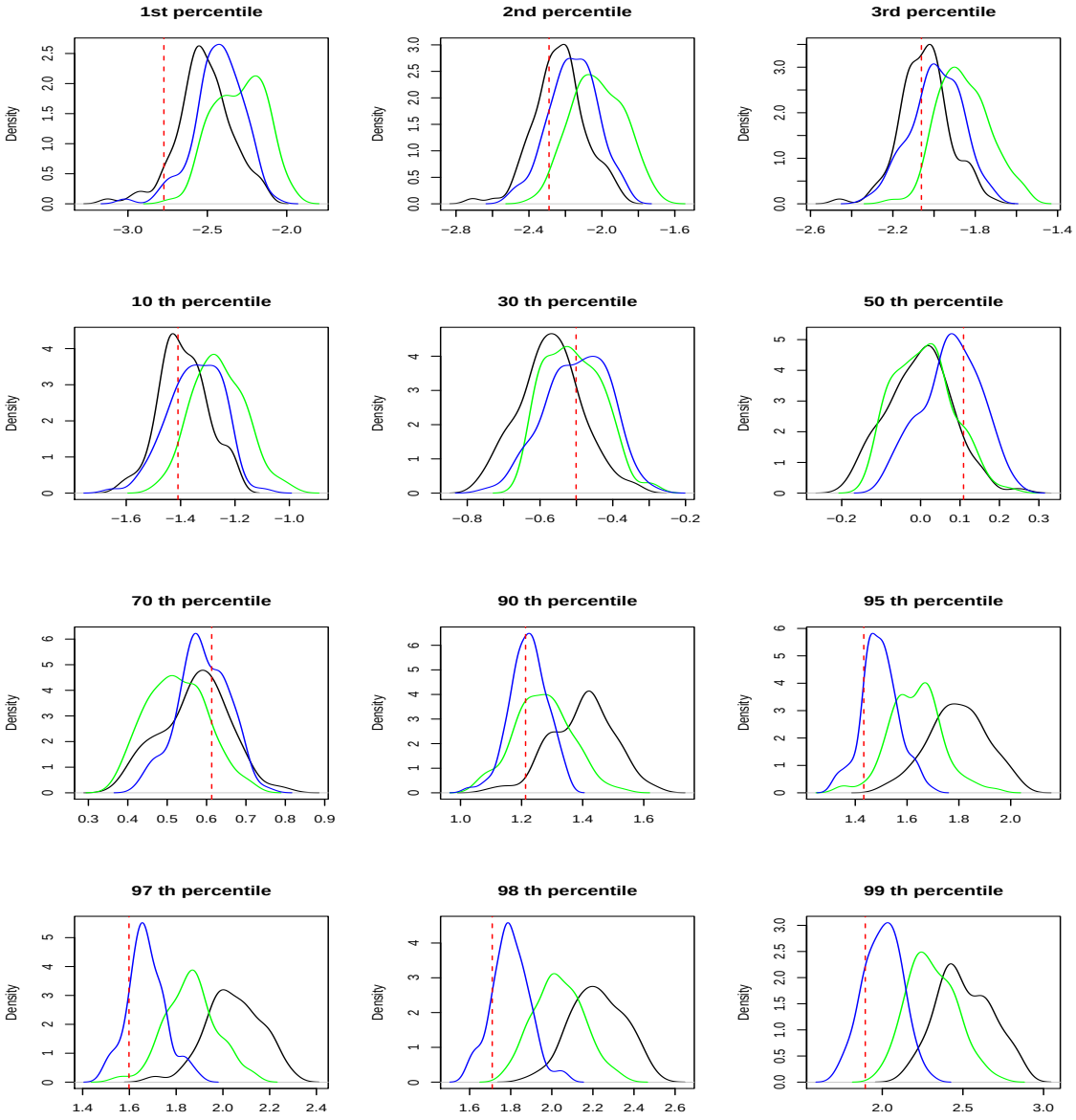


FIGURE 11.15 – Quantiles (red vertical lines) for February of Z_t and their distributions built from the simulations of different models : in black, model with constant a, in green, model with $a^2 = f(t)$, in blue, model with $a(t, Z_{t-1})$.

- from the obtained sequence, estimate the density function for the temperature at date t by kernel method.

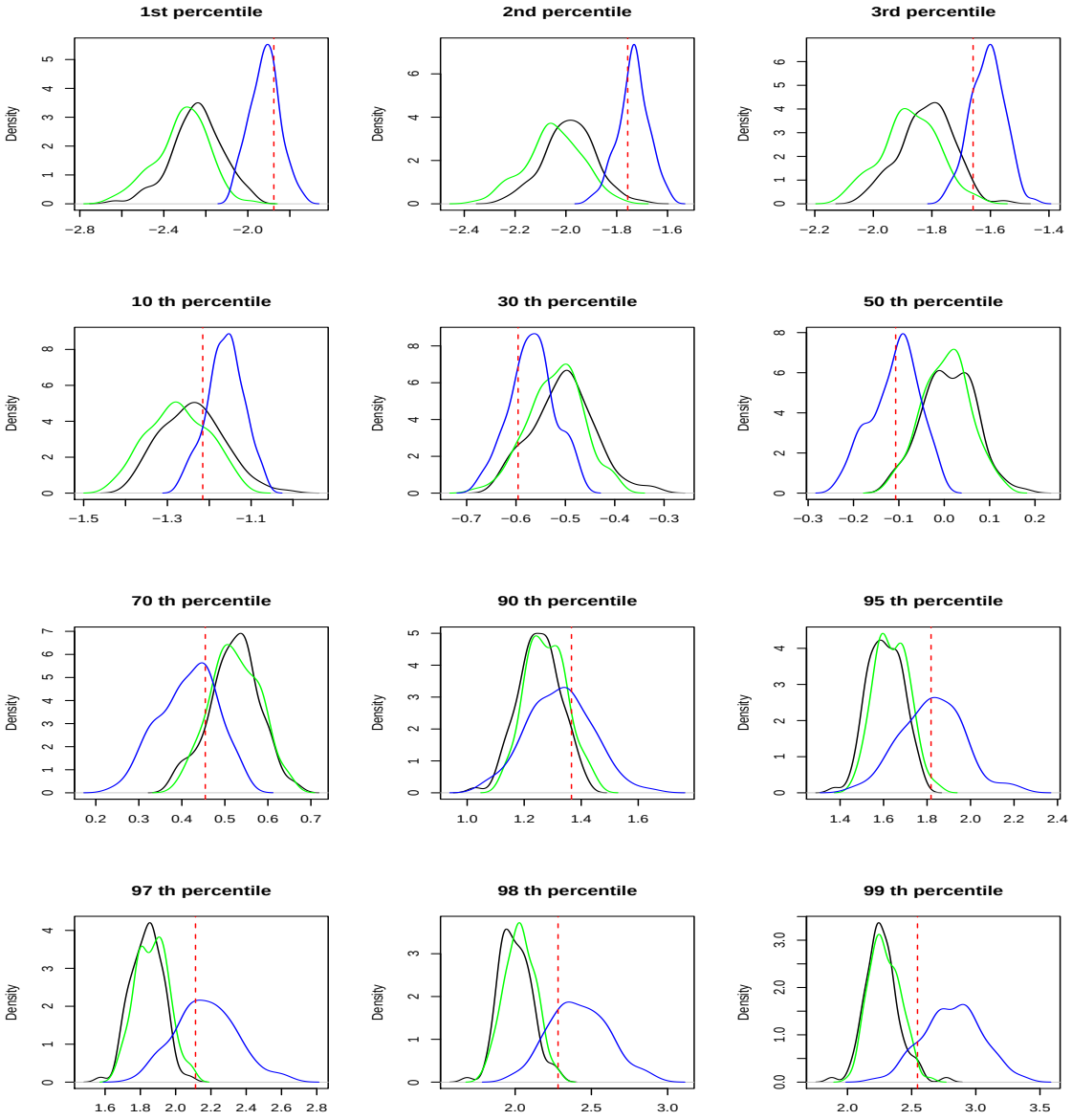
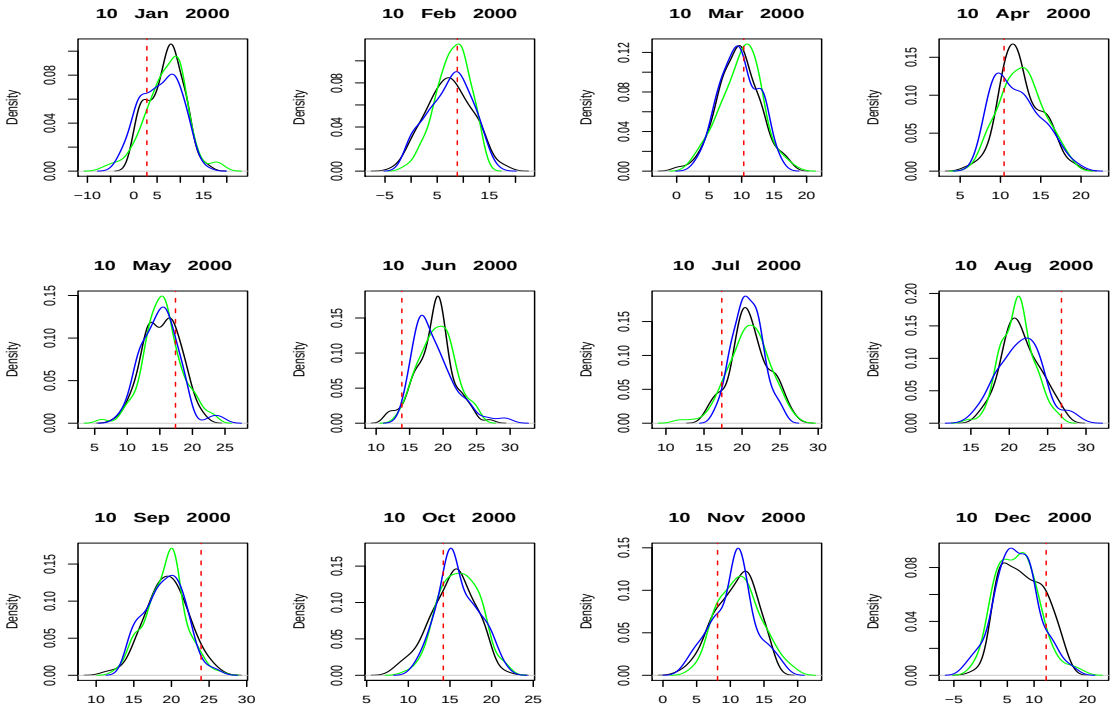


FIGURE 11.16 – Quantiles (red vertical lines) for July of Z_t and their distributions built from the simulations of different models : in black, model with constant a, in green, model with $a^2 = f(t)$, in blue, model with $a(t, Z_{t-1})$.

We consider respectively t as the tenth day of each month in 2000. The results are shown in Figure 11.17.

For this criterion, we cannot see which model performs better. In gene-

FIGURE 11.17 – Observed temperatures X_t at fixed dates (in red) and their different distributions from different simulations : with constant a (black lines), $a^2 = f(t)$(red lines) and $a(t, Z_{t-1})$(blue lines)



ral, this method gives a distribution with high variance for the temperature at a fixed date, so usually a "normal" observed value is found inside the distribution. "Normal" here means that an usual temperature in general for this date. However, for an abnormal temperature, three models can not give a good distribution. For example, for 10 June 2000, the temperature is about 14°C, which is cold for a summer day, or for 10 August 2000, the temperature is about 27°, which is apparently hot for this date, these models do not perform well. This remark is interesting because it shows a weakness of our model.

⋆ *Extreme parameters*

The performance of the simulations in the extreme fields also needs to be considered. If the extreme behaviour in simulations of Z_t is correct, the one in simulations of the observed series X_t will be correct. As the reduced series Z_t is rather stationary, but not X_t, it is simpler to apply stationary extreme

models on Z_t. We show in Table 11.2 the estimated extreme parameters (from GEV models) of Z_t and their different 95% confidence intervals from different simulated samples. This table shows the results for the right tail. Similar results must be obtained for the left one.

	$\hat{\mu}$	Simulated $I\mu$			$\hat{\sigma}$	Simulated $I\sigma$			$\hat{\xi}$	Simulated $I\xi$		
	value	constant	$f(t)$	$a(t, Z_{t-1})$	value	constant	$f(t)$	$a(t, Z_{t-1})$	value	const.	$f(t)$	$a(t, Z_{t-1}$
Jan	1.29	(1.38,1.63)	(1.27,1.55)	(1.24,1.48)	0.45	(0.59,0.84)	(0.56,0.81)	(0.44,0.69)	-0.29	(-0.42, -0.02)	(-0.40, -0.12)	(-0.48, -0.21)
Feb	1.26	(1.36,1.65)	(1.30,1.51)	(1.22,1.45)	0.48	(0.59,0.79)	(0.55,0.72)	(0.44,0.63)	-0.27	(-0.36, -0.08)	(-0.33, -0.12)	(-0.39, -0.14)
Mar	1.41	(1.37,1.60)	(1.35,1.54)	(1.29,1.54)	0.56	(0.56,0.74)	(0.53,0.70)	(0.51,0.67)	-0.31	(-0.36, -0.08)	(-0.39, -0.01)	(-0.31, -0.10)
Apr	1.65	(1.40,1.65)	(1.40,1.62)	(1.40,1.65)	0.57	(0.50,0.67)	(0.48,0.68)	(0.56,0.72)	-0.30	(-0.34, -0.07)	(-0.34, -0.09)	(-0.30, -0.06)
May	1.68	(1.40,1.63)	(1.41,1.62)	(1.42,1.68)	0.67	(0.49,0.67)	(0.52,0.68)	(0.61,0.80)	-0.30	(-0.34, -0.09)	(-0.30, -0.03)	(-0.26, 0.00)
Jun	1.70	(1.39,1.61)	(1.43,1.62)	(1.46,1.72)	0.71	(0.51,0.68)	(0.51,0.67)	(0.65,0.86)	-0.19	(-0.34, -0.11)	(-0.30, -0.09)	(-0.19, 0.05)
Jul	1.74	(1.38,1.61)	(1.42,1.65)	(1.39,1.74)	0.71	(0.42,0.70)	(0.51,0.67)	(0.66,0.85)	-0.20	(-0.34, -0.10)	(-0.34, -0.09)	(-0.21, 0.01)
Aug	1.63	(1.39,1.64)	(1.39,1.62)	(1.38,1.70)	0.68	(0.52,0.68)	0.51,0.68	0.62,0.83	-0.29	-0.32, -0.08	(-0.32, -0.08)	(-0.31, -0.07)
Sep	1.54	(1.40,1.63)	(1.40,1.61)	(1.39,1.67)	0.60	(0.53,0.69)	(0.49,0.68)	(0.57,0.75)	-0.20	(-0.33, -0.08)	(-0.31, -0.08)	(-0.27, -0.04)
Oct	1.46	(1.36,1.61)	(1.40,1.68)	(1.39,1.63)	0.55	(0.49,0.66)	(0.50,0.67)	(0.50,0.65)	-0.18	(-0.31, -0.09)	(-0.35, -0.10)	(-0.30, -0.07)
Nov	1.45	(1.35,1.55)	(1.43,1.66)	(1.40,1.57)	0.50	(0.48,0.63)	(0.52,0.67)	(0.44,0.60)	-0.32	(-0.33, -0.09)	(-0.32, -0.09)	(-0.36, -0.14)
Dec	1.44	(1.35,1.57)	(1.44,1.65)	(1.40,1.57)	0.44	(0.49,0.63)	(0.51,0.68)	(0.43,0.60)	-0.24	(-0.32, -0.08)	(-0.34, -0.10)	(-0.38, -0.15)

TABLE 11.2 – Estimated extreme parameters of Z_t and their 95% confidence intervals calculated from different simulations.

The simulations from the model with $a(t, Z_{t-1})$ performs much better than the others for the tail behaviour. The estimated extreme parameters of Z_t are mostly found in the confidence intervals for this kind of simulations. This model shows clearly a better fit, especially for the summer and the winter, where the estimated parameters of Z_t are usually found outside the confidence intervals of the other simulations.

⋆ *Cluster properties*

Studying on the clusters gives us a general look about the distribution and the length of consecutive days where the temperatures are larger (smaller) than a given threshold. We will consider the clusters over the threshold of 98th percentile and 2nd percentile and the results from the simulations will be compared with the observations (Table 11.3). This study can apply on Z_t or X_t. We give here the results for X_t.

	Observations	Sim. const. a	Sim.$a^2 = f(t)$	Sim. $a(t, Z_{t-1})$
		2% quantile		
Threshold	-0.066	0.239	0.367	0.131
Length	Distribution	Distribution	Distribution	Distribution
1	0.419	0.499	0.492	0.485
2	0.256	0.234	0.241	0.243
3	0.100	0.113	0.113	0.118
4	0.069	0.063	0.063	0.062
5	0.044	0.035	0.036	0.034
6	0.050	0.019	0.021	0.021
7	0.025	0.014	0.01	0.014
8	0.000	0.008	0.008	0.009
9	0.006	0.004	0.005	0.005
10	0.006	0.003 0.004	0.003	
Rate of declust.	0.371	0.461	0.461	0.457
		98% quantile		
Threshold	24.332	23.850	23.872	24.132
Length	Distribution	Distribution	Distribution	Distribution
1	0.512	0.483	0.488	0.483
2	0.244	0.262	0.263	0.243
3	0.107	0.123	0.120	0.119
4	0.054	0.060	0.057	0.060
5	0.029	0.031	0.030	0.037
6	0.029	0.017	0.016	0.021
7	0.005	0.011	0.012	0.014
8	0.005	0.005	0.006	0.008
9	0.000	0.003	0.004	0.006
10	0.005	0.002	0.002	0.004
Rate of declust.	0.477	0.481	0.483	0.455

TABLE 11.3 – The cluster properties of X_t and its different simulations

All three models give correct fits for the clusters. The difference between these models cannot be seen.

Now the last work is to check if the estimation procedure for the parameters in our model is robust. We will use the same estimation method for each sample of simulations and then, build the 95% confidence intervals for the estimated parameters from the observations.

The estimators of the deterministic functions $\hat{m}, \hat{s}, \hat{S}, \hat{S}_V$ with their confidence intervals are found in Figure 11.18. The simulations give good confidence intervals for these estimators.

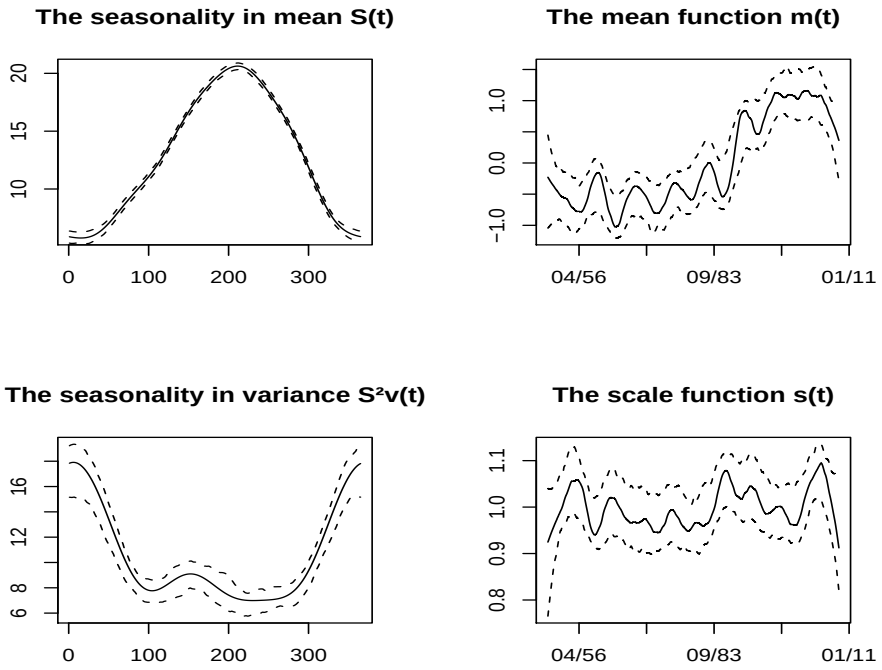


FIGURE 11.18 – Estimators of trends and seasonalities with their confidence intervals built from the simulations.

With the same rule, we construct the confidence intervals for the coefficients of AR(3) (functions of time) and for $\hat{a}(t, Z_{t-1})$ (with extension in the boundaries) with different fixed t, for example the twentieth day of the January, Mars, August and November (Figure 11.19 and 11.20).

These results show that the estimation procedure for our model is robust.

Conclusion

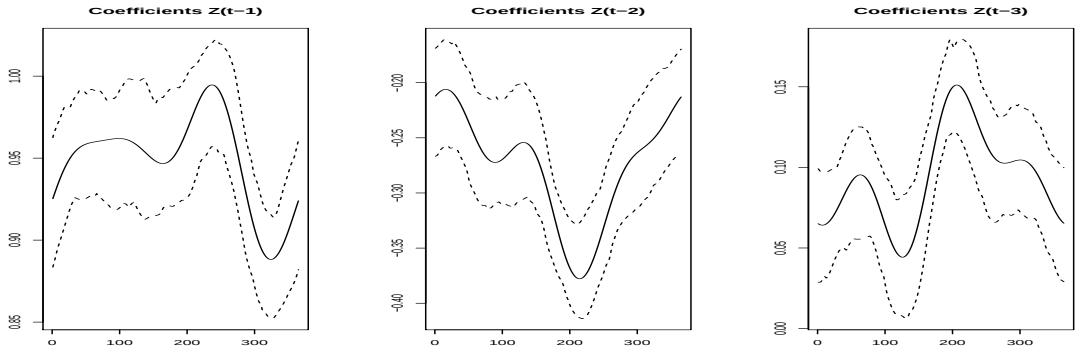


FIGURE 11.19 – Estimators of seasonal coefficients of AR(3) with their confidence intervals built from the simulations.

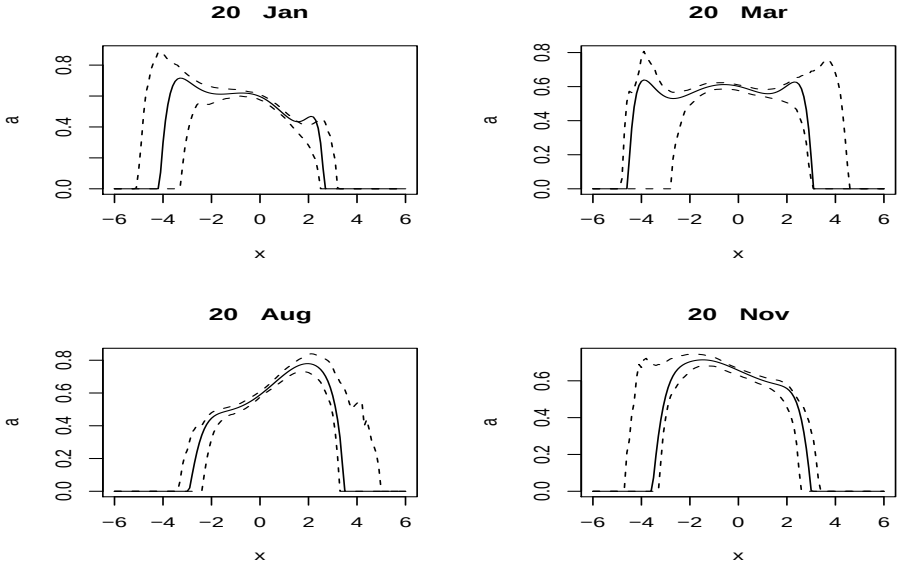


FIGURE 11.20 – Estimators of $a(t, Z_{t-1})$ at fixed date t with their confidence intervals built from the simulations

The very interesting results obtained from the model with $a(t, Z_{t-1})$, which takes into account both the dates and the state, show the adequacy of this model for the temperature. It shows its quality in both the central and extreme fields, dealing with the other models of the diffusion coefficient a. The good performance of this model means that the seasonality in the reduced temperature is really significant and absolutely needs to be taken

into account in simulation models.

The volatility a is not constant but linear, and its sign changes with different seasons. The considered constraints in the boundaries for a give good results in extremes. These are necessary conditions when working with bounded data, especially when the innovation is taken as Gaussian.

The estimation procedure for the model is robust and stable.

This study confirms the goodness-of-fit of our model. Now we will use it to model different kinds of temperature in Bordeaux and consider the goodness-of-fit of the model dealing with different temperatures.

From now on, we will keep these following criteria for the validation of a model :

• The normality and the whiteness of the residuals, the whiteness of the squared residuals.

• Density functions for a fixed month of Z_t (we take here 3 months with different levels of temperatures : January, July, September).

• Quantiles of Z_t for fixed months.

• GEV extreme parameters of Z_t for 12 months.

When the model gives a good fit for Z_t, normally it should also give a good fit for X_t when the deterministic functions (trend, seasonality) are correctly estimated.

11.3.0.2 Modeling on daily fixed-hour temperatures

The series, at our disposal, for which the intervals between the observations are the shortest, are temperature series measured each three hours (at 0h, 3h,..., 21h). Their global dynamic is really complicated : they have two evident seasonalities, daily cycle and annual cycle, which need to be removed. Of course with more data one can hope and have to get a better model. But it seems very difficult to treat all the problems we consider in this thesis into the framework of this very heavy model.

Nevertheless, preliminary studies show that except for heavy computation time, our method is also convenient with these kinds of biperiodic series. The daily temperature at one fixed hour is simpler and however could be

considered as one discrete-time temperature series with one-day fixed lag.

Each temperature series of a 'fixed hour' type contains in itself all the characteristics of general temperatures : strong seasonality, trends in mean and in variance, as well as correlations between the observations. We show in figure 11.21 the curves of daily average temperatures of each fixed-hour temperature series. These series are taken from average values of fixed-hour temperatures at the same dates of every years.

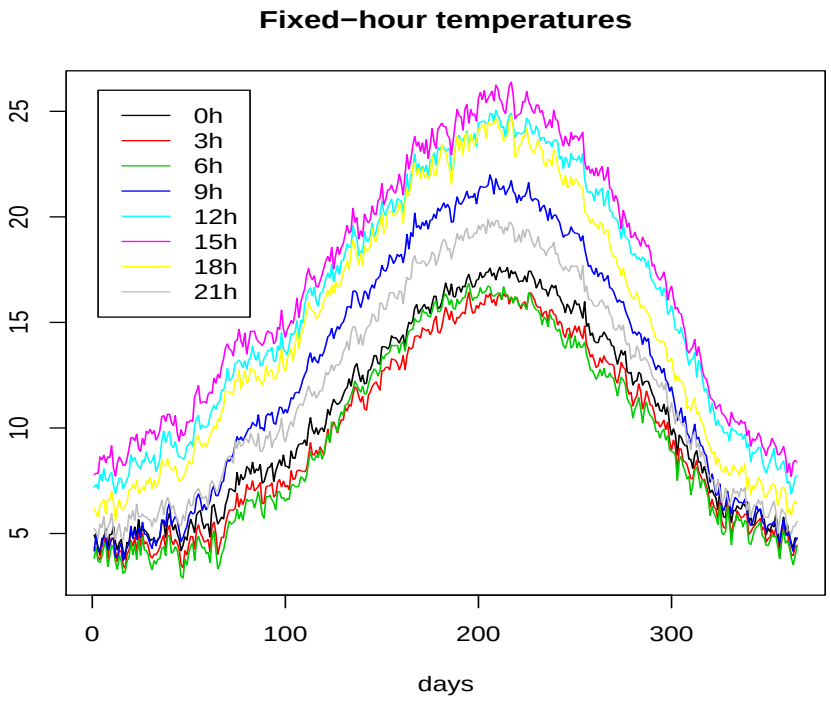


FIGURE 11.21 – Fixed-hour temperatures

We will not make a study of all these eight series, but only some of them, for example three temperature series of 0h, 12h and 18h. Normally, at 0h it is colder, and at 12h, it is hotter. Thus these series have probably different characteristics. It is interesting to see if our model can adjust all these different categories or if it shows some weaknesses to some specific characteristics.

The estimation procedure is the same as before. To validate the models, we simulate again 100 samples for each reduced series of these fixed-hour temperatures. The residuals and different statistical elements of the simula-

tions will be considered.

⋆ *The residuals*

All the obtained residuals are white, and the correlation in their squared values are negligible. The quantiles of the residuals of all series comparing with those of a normal distribution $\mathcal{N}(0,1)$ are correct.

The Kolmogorov normality test accept the normality for these three residuals. The normality of the residuals can be accepted following p-values of this test, respectively 0.34, 0.05 and 0.07 for the series of 0h, 12h and 18h.

We consider now the performance of the simulations of these series.

⋆ *Density functions*

We build 95% confidence intervals for the density function for a month (January, July or September)of each series from simulated samples. The results are illustrated in Figure 11.22.

The simulations give in general good intervals for the density functions of Z_t in a fixed month. Except for July of Z_t of 18h series, the data in this case is not always found in the interval of the simulations. The simulations have difficulties to catch of the asymmetry of the data.

⋆ *Quantiles*

We build now the 95% confidence intervals for the quantiles. We take the 1%, 2%, 30%, 50%, 60%, 80%, 98% and 99% quantiles . The quantiles of Z_t for January, July, September and their confidence intervals for different fixed-hour temperature series are respectively presented in Table 11.4.

We marked in bold where the observed quantile is outside the confidence interval. The results are good almost everywhere. For January, the lower quantiles are really good, but the higher quantiles (98% ,99% ones) are not good, which is less important for a cold month. The results of July and September are better.

⋆ *Extreme parameters*

We repeat the same work as in the previous section to check the quality of the extreme field in the simulations. Table 11.5 shows the results for the right tail, while extreme parameters for the left tail are in Table 11.6.

Discussion

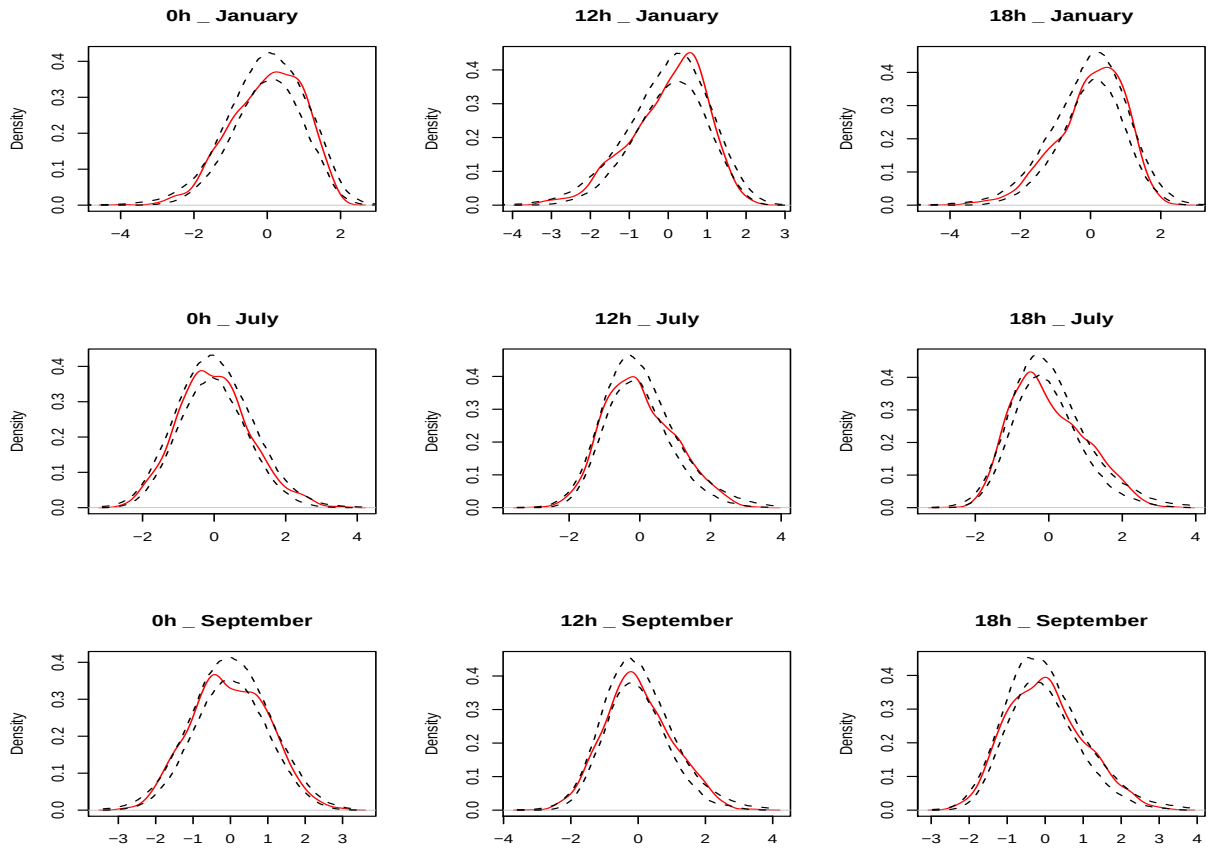


FIGURE 11.22 – Density functions for fixed months of Z_t and their 95% confidence interval from simulations (dashed lines) of fixed-hour temperature series

The model for fixed-hour temperatures is correct for both the distributions and the quantiles. For extremes, it seems that the simulations perform better for the left tail than the right one. Sometimes, the extreme parameters are found outside the confidence intervals of simulations. However, this comparison is relative, not exact. The standard deviations for the estimators of μ and ξ are rather high about 0.05 in average for μ and ξ, for σ it is about 0.04 in average.

11.3.1 Modeling on daily mean, maximum and minimum temperatures

We now apply Model (11.5) on the daily mean (Tmean), maximum (Tmax) and minimum temperatures (Tmin). The validation of the model is

January	1%	2%	30%	50%
0h	-2.49([-2.88,-2.17])	-2.15([-2.48,-1.93])	-0.50([-0.61,-0.38])	0.09([-0.06,0.16])
12h	-2.7([-2.89,-2.24])	-2.25([-2.51,-1.94])	-0.44([-0.57,-0.32])	0.17([-0.01,0.21])
18h	-2.72([-3.06,-2.27])	-2.28([-2.57,-1.96])	-0.39([-0.63,-0.32])	0.13([-0.07,0.18])
January	60%	80%	98%	99%
0h	0.36([0.17,0.43])	0.89([0.73,1.00])	1.67([1.67,1.95])	**1.78([1.88,2.17])**
12h	0.41 ([0.23,0.42])	0.84([0.76,0.96])	**1.62([1.67,1.91])**	**1.80([1.83,2.11])**
18h	0.38 ([0.17,0.42])	0.86([0.71,0.93])	**1.59([1.69,2.00])**	**1.80([1.90,2.23])**
July	1%	2%	30%	50%
0h	-2.07([-2.28,-1.97])	-1.92([-2.03,-1.78])	-0.57([-0.63,-0.44])	-0.04([-0.14,0.07])
12h	-2.00 ([-2.1,-1.81])	-1.78([-1.88,-1.62])	-0.60([-0.65,-0.5])	-0.1([-0.18,0.00])
18h	-1.83 ([-2.02,-1.77])	-1.64([-1.81,-1.63])	-0.62([-0.65,-0.49])	-0.14([-0.19,-0.01])
July	60%	80%	98%	99%
0h	0.22([0.12,0.32])	0.81([0.72,0.94])	2.34([1.93,2.37])	2.58([2.17,2.69])
12h	0.15 [0.05,0.27])	0.89([0.64,0.93])	2.20([2.09,2.64])	**2.40([2.44,3.06])**
18h	0.15([0.05,0.25])	0.91([0.63,0.91])	2.18([2.13,2.64])	**2.38([2.46,3.15])**
September	1%	2%	30%	50%
0h	-2.14([-2.46,-2.13])	-1.99([-2.16,-1.87])	-0.59([-0.61,-0.44])	-0.04([-0.10,0.09])
12h	-2.15([-2.18,-1.89])	-1.93([-1.96,-1.69])	-0.54 ([-0.64,-0.46])	-0.08([-0.19,0.03])
18h	-1.98([-2.04,-1.78])	-1.77([-1.86,-1.61])	-0.59 ([-0.66,-0.49])	-0.05([-0.21,0.00])
September	60%	80%	98%	99%
0h	0.26 ([0.17,0.35])	0.91([0.77,0.95])	2.01([1.92,2.18])	2.27([2.14,2.50])
12h	0.19([0.05,0.28])	0.88([0.65,0.92])	2.16([2.00,2.63])	2.38([2.43,3.13])
18h	0.20([0.04,0.26])	0.87([0.66,0.94])	2.24([2.10,2.64])	2.46([2.46,3.21])

TABLE 11.4 – Quantiles of Z_t for fixed months and their 95% confidence intervals of fixed-hour temperatures

as follows.

⋆ *The residuals*

The residuals are white and nearly white for their squared values. Their quantiles are close to those of a normal distribution between 1st and 99th percentiles. This can be seen from Figure 11.23.

The normality of these residuals can be accepted by a Kolmogorov test with p-values 0.11, 0.05, 0.03 respectively for Tmean, Tmax and Tmin .

We simulate 100 samples for each series. The performance of the models will be studied through simulations.

⋆ *Density functions*

Figure 11.24 shows the confidence intervals of the densities from the simulation samples of the temperature series Tmean, Tmax and Tmin for fixed months.

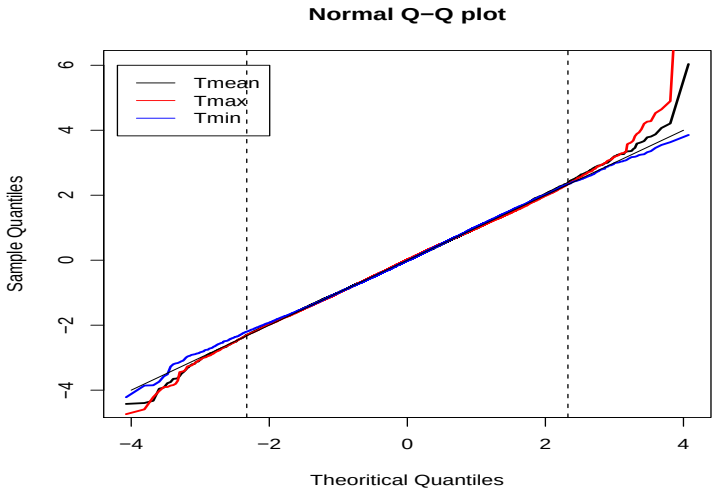


FIGURE 11.23 – Normal Q-Q plot of the residuals when modeling Tmean, Tmax and Tmin. The vertical lines are the 1% and 99% quantiles of $\mathcal{N}(0,1)$ distribution.

⋆ *Quantiles*

The quantiles of fixed months are estimated with their 95% confidence intervals (Table 11.7).

The case where the quantiles of reduced temperatures are outside their confidence intervals are in bold. The simulations give in general good confidence intervals. The model of Tmin is not really good in the higher quantiles.

⋆ *Extreme parameters*

Extreme parameters of GEV models will be calculated from the subseries of each months from the reduced series Z_t respectively for Tmax, Tmean, Tmin. Their confidence intervals for both left and right tails are built from the simulations, shown in Table 11.8 and 11.9.

11.4 Conclusion

After these studies, we find out some very interesting phenomena :

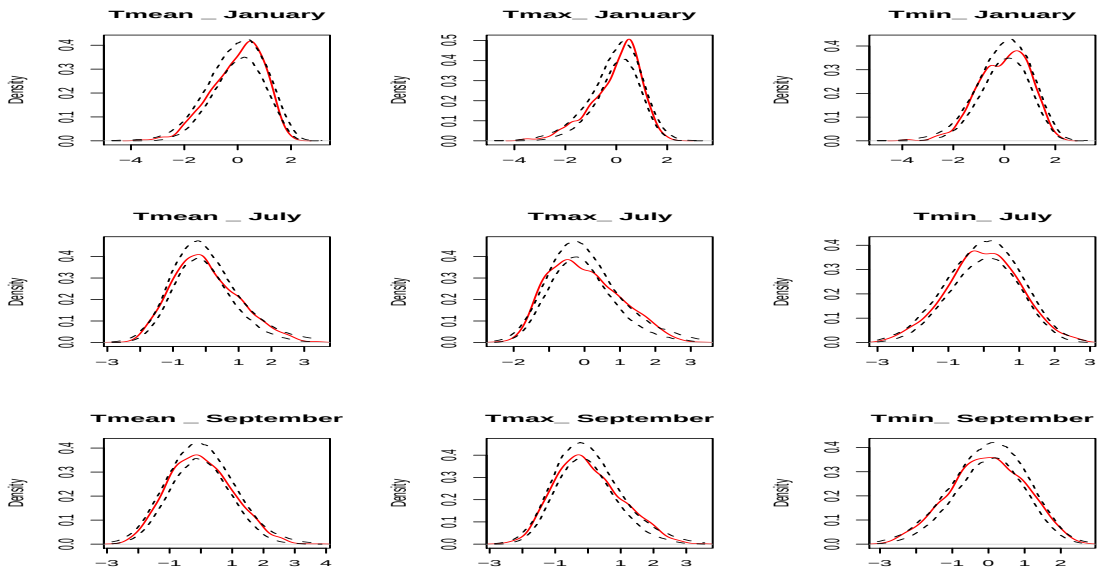


FIGURE 11.24 – Density functions of Tmean, Tmax, Tmin for January, July, September of reduced temperatures and their confidence intervals from simulated samples

- It is obvious that all the statistical elements of reduced temperatures are affected by the seasonality (mean, variance of the central field and also extreme events).

- Especially, the volatility a depends both on the dates and the state. Particularly, in summer, when it is hot, a has an increasing trend with the temperatures, whereas in winter, when it is cold, a has a tendency to decrease with the temperatures. Thus to have a good modeling for the temperature, this characteristic of a needs to be taken into account.

- The temperature is bounded. Here we proposed an estimation of a^2 using seasonal constraints on the finite inaccessible boundaries of the temperature.

- All these points show the complicated features of the temperature. Our model takes all of them into account which give a rather complete model. The closeness of the residuals to a normal distribution and the validation of the models by simulations show the good performance of the model.

- The model give a good, although not yet complete, representation both for the central and extreme fields. The model catches of rather well the sea-

sonality in reduced series of different kinds of temperatures. Of course, some weaknesses remain in our model : sometimes it cannot completely catch of some complicated characteristics of the temperature such as the asymmetry of the data and some exceptional temperature values. The lack of theories about the volatility near the boundaries force us to find a way to complete it. And this way is maybe not sufficient. More studies are needed to obtain a better "connection" for this conditional variance.

- Our model is complete because it takes into account both the seasonality of all statistical factors and the boundedness of the temperature. However, it is also complex with many parameters. Therefore, it becomes numerically heavy for the simulation. A future work is to find a model which keeps the same spirit than this model but contains less parameters.

Months	0h	12h	18h
January			
$\mu(I\mu)$	1.40(1.40,1.62)	1.33(1.31,1.51)	1.30(1.23,1.45)
$\sigma(I\sigma)$	**0.38(0.48,0.70)**	0.49(0.48,0.75)	**0.45(0.51,0.77)**
$\xi(I\xi)$	-0.27(-0.50,-0.21)	-0.26(-0.36,-0.09)	-0.26(-0.32 ?-0.06)
February			
$\mu(I\mu)$	**1.37(1.42,1.62)**	1.33(1.30,1.52)	1.37(1.25,1.48)
$\sigma(I\sigma)$	0.45(0.45,0.63)	0.54(0.50,0.69)	0.53(0.53,0.72)
$\xi(I\xi)$	-0.33(-0.39,-0.18)	-0.19(-0.27,0.05)	**-0.29(-0.27,-0.02)**
March			
$\mu(I\mu)$	1.51(1.48,1.68)	1.53(1.41,1.64)	1.58(1.40,1.66)
$\sigma(I\sigma)$	0.43(0.43,0.61)	0.64(0.56,0.78)	0.60(0.60,0.78)
$\xi(I\xi)$	-0.24(-0.36,-0.12)	**-0.22(-0.19,0.10)**	**-0.35()-0.28,-0.05**
April			
$\mu(I\mu)$	1.66(1.53,1.72)	1.78(1.51,1.78)	1.81(1.53,1.83)
$\sigma(I\sigma)$	0.52(0.44,0.60)	**0.56(0.59,0.78)**	**0.55(0.60,0.80)**
$\xi(I\xi)$	-0.30(-0.34,-0.10)	-0.23(-0.23,0.05)	**-0.35(-0.30,-0.04)**
May			
$\mu(I\mu)$	1.69(1.58,1.76)	1.77(1.57,1.84)	1.78(1.57,1.84)
$\sigma(I\sigma)$	0.61(0.48,0.63)	0.63(0.61,0.78)	0.60(0.60,0.80)
$\xi(I\xi)$	-0.31(-0.32,-0.08)	**-0.30(-0.24,-0.01)**	**-0.30(-0.22,0.05)**
June			
$\mu(I\mu)$	1.74(1.58,1.78)	1.75(1.56,1.86)	1.75(1.55,1.82)
$\sigma(I\sigma)$	0.67(0.51,0.68)	0.64(0.62,0.82)	0.60(0.60,0.80)
$\xi(I\xi)$	-0.27(-0.30,-0.06)	**-0.33(-0.22,0.03)**	-0.25(-0.25,0.00)
July			
$\mu(I\mu)$	1.76(1.53,1.78)	1.80(1.56,1.87)	1.81(1.56,1.87)
$\sigma(I\sigma)$	0.66(0.52,0.71)	0.62(0.62,0.84)	0.60(0.60,0.83)
$\xi(I\xi)$	-0.25(-0.32,-0.05)	-0.27(-0.27,0.02)	-0.25(-0.25,0.05)
August			
$\mu(I\mu)$	1.68(1.56,1.78)	1.73(1.54,1.87)	1.73(1.60,1.88)
$\sigma(I\sigma)$	0.61(0.52,0.66)	**0.57(0.65,0.83)**	0.61(0.60,0.80)
$\xi(I\xi)$	-0.25(-0.36,-0.05)	-0.26(-0.26,0.02)	-0.25(-0.24,0.03)
September			
$\mu(I\mu)$	1.62(1.59,1.79)	1.64(1.50,1.80)	1.65(1.52,1.81)
$\sigma(I\sigma)$	0.53(0.46,0.62)	0.60(0.60,0.80)	0.60(0.58,0.80)
$\xi(I\xi)$	-0.27(-0.34,-0.07)	-0.25(-0.27,0.00)	-0.26(-0.27,0.00)
October			
$\mu(I\mu)$	**1.50(1.57,1.74)**	1.58(1.45,1.71)	1.52(1.421.69)
$\sigma(I\sigma)$	0.48(0.45,0.58)	0.56(0.52,0.72)	0.57(0.57,0.76)
$\xi(I\xi)$	-0.26(-0.38,-0.10)	-0.21(-0.27,-0.04)	-0.24(-0.25,-0.02)
November			
$\mu(I\mu)$	1.52(1.48,1.69)	1.55(1.46,1.65)	1.50(1.42,1.68)
$\sigma(I\sigma)$	0.43(0.43,0.57)	0.55(0.49,0.65)	0.52(0.51,0.66)
$\xi(I\xi)$	-0.32(-0.38,-0.13)	-0.30(-0.30,-0.07)	-0.20(-0.30,-0.06)
December			
$\mu(I\mu)$	1.52(1.46,1.66)	1.45(1.44,1.65)	1.46(1.43,1.66)
$\sigma(I\sigma)$	**0.38(0.43,0.57)**	0.52(0.46,0.63)	0.46(1.48,0.65)
$\xi(I\xi)$	-0.24(-0.37,-0.11)	-0.26(-0.32,-0.07)	-0.17(-0.28,-0.04)

TABLE 11.5 – Estimated extreme parameters of Z_t in the right tail and their 95% confidence intervals for fixed-hour series.

Months	0h	12h	18h
January			
$\mu(I\mu)$	-1.46(-1.70,-1.42)	-1.54(-1.63,-1.38)	-1.46(-1.62,-1.32))
$\sigma(I\sigma)$	0.78(0.60,0.89)	0.76(0.63,0.84)	0.83(0.62,0.88)
$\xi(I\xi)$	-0.25(-0.36,-0.06)	-0.34(-0.39,-0.08)	-0.26(-0.29,0.00)
February			
$\mu(I\mu)$	-1.44(-1.71,-1.43)	-1.44(-1.58,-1.34)	-1.35(-1.54,-1.28)
$\sigma(I\sigma)$	0.55(0.55,0.73)	0.70(0.55,0.76)	0.74(0.57,0.74)
$\xi(I\xi)$	-0.01(-0.29,-0.01)	-0.20(-0.32,-0.06)	-0.16(-0.28,0.00)
March			
$\mu(I\mu)$	-1.56(-1.74,-1.51)	-1.44(-1.58,-1.40)	-1.38(-1.51,-1.30)
$\sigma(I\sigma)$	**0.51(0.53,0.70)**	0.60(0.45,0.63)	0.61(0.44 ?0.61)
$\xi(I\xi)$	-0.15(-0.30,-0.03)	-0.19(-0.36,-0.10)	-0.16(-0.34,-0.06)
April			
$\mu(I\mu)$	-1.60(-1.78,-1.58)	-1.41(-1.59,-1.41)	-1.35(-1.56,-1.39)
$\sigma(I\sigma)$	0.53(0.48,0.64)	0.51(0.38,0.52)	0.47(0.38,0.50)
$\xi(I\xi)$	-0.27(-0.30,-0.07)	-0.25(-0.33,-0.12)	-0.25(-0.36,-0.12)
May			
$\mu(I\mu)$	-1.57(-1.77,-1.56)	-1.46(-1.60,-1.46)	-1.40(-1.57,-1.43)
$\sigma(I\sigma)$	0.47(0.46,0.62)	0.47(0.37,0.50)	0.39(0.36,0.47)
$\xi(I\xi)$	-0.24(-0.34,-0.12)	-0.34(-0.36,-0.12)	-0.36(-0.38,-0.11)
June			
$\mu(I\mu)$	-1.57(-1.71,-1.51)	-1.48(-1.58,-1.43)	**-1.34(-1.55,-1.42)**
$\sigma(I\sigma)$	0.44(0.44,0.58)	0.44(0.38,0.49)	0.35(0.35,0.46)
$\xi(I\xi)$	-0.32(-0.37,-0.16)	-0.24(-0.38,-0.12)	-0.17(-0.35,-0.08)
July			
$\mu(I\mu)$	-1.57(-1.70,-1.50)	-1.46(-1.57,-1.41)	-1.41(-1.54,-1.41)
$\sigma(I\sigma)$	0.46(0.43,0.58)	0.42(0.35,0.48)	0.35(0.34,0.46)
$\xi(I\xi)$	-0.36(-0.40,-0.13)	-0.21(-0.37,-0.12)	-0.15(-0.36,-0.11)
August			
$\mu(I\mu)$	**-1.51(-1.69,-1.53)**	-1.54(-1.57,-1.41)	-1.36(-1.54,-1.36)
$\sigma(I\sigma)$	0.50(0.43,0.57)	**0.49(0.36,0.47)**	0.37(0.33,0.47)
$\xi(I\xi)$	-0.35(-0.36,-0.09)	**-0.35(-0.34,-0.11)**	-0.22(-0.35,-0.10)
September			
$\mu(I\mu)$	-1.67(-1.75,-1.54)	-1.58(-1.58,-1.40)	-1.47(-1.55,-1.36)
$\sigma(I\sigma)$	0.45(0.45,0.61)	0.51(0.40,0.53)	0.41(0.36,0.51)
$\xi(I\xi)$	-0.28(-0.36,-0.07)	-0.23(-0.33,-0.07)	-0.25(-0.39,-0.12)
October			
$\mu(I\mu)$	-1.74(-1.80,-1.57)	-1.54(-1.63,-1.42)	-1.45(-1.56,-1.39)
$\sigma(I\sigma)$	0.50(0.50,0.66)	0.49(0.45,0.62)	0.53(0.44,0.58)
$\xi(I\xi)$	**-0.34()-0.27,-0.06**	-0.14(-0.32,0.04)	-0.20(-0.31,-0.01)
November			
$\mu(I\mu)$	-1.67(-1.75,-1.53)	-1.66(-1.68,-1.47)	-1.58(-1.63,-1.43)
$\sigma(I\sigma)$	0.54(0.51,0.68)	0.60(0.53,0.67)	0.58(0.49,0.68)
$\xi(I\xi)$	**-0.33(-0.29,-0.03)**	**-0.30(-0.24,0.00)**	-0.22(-0.24,-0.01)
December			
$\mu(I\mu)$	-1.61(-1.74,-1.49)	-1.63(-1.71,-1.48)	-1.52(-1.69,-1.44)
$\sigma(I\sigma)$	0.53(0.52,0.69)	0.62(0.54,0.72)	0.62(0.51,0.71)
$\xi(I\xi)$	-0.23(-0.28,-0.01)	**-0.27(-0.25,-0.01)**	-0.22(-0.25,0.03)

TABLE 11.6 – Estimated extreme parameters of Z_t in the right tail and their 95% confidence intervals for fixed-hour series

January	1%	2%	30%	50%
Tmean	-2.77([-2.83,-2.22])	-2.17([-2.49,-1.93])	-0.45([-0.66 ,-0.33])	0.15([-0.06 ,0.22])
Tmax	-2.73([-2.98,-2.35])	-2.16([-2.61,-2.06])	-0.48([-0.54 ,-0.27])	0.15([0.00,0.20])
Tmin	-2.64([-2.99,-2.26])	-2.14([-2.54 ,-2.00])	-0.48([-0.59,-0.36])	0.15([-0.07,0.17])
January	60%	80%	98%	99%
Tmean	0.39([0.19,0.45])	0.89([0.77,1.00])	**1.57([1.65,1.87])**	**1.72([1.81,2.02])**
Tmax	0.39([0.23,0.41])	0.88([0.7,0.88])	1.57([1.55,1.81])	1.73([1.71,2.00])
Tmin	0.39([0.18,0.42])	0.87([0.74,0.98])	**1.57([1.67,1.87])**	**1.73([1.82,2.06])**
July	1%	2%	30%	50%
Tmean	-1.88([-2.05,-1.78])	-1.76([-1.86,-1.62])	-0.61([-0.66,-0.48])	-0.12([-0.21,-0.02])
Tmax	-1.88([-2.02,-1.74])	-1.77([-1.80,-1.57])	-0.60 [-0.64,-0.50])	-0.11 ([-0.20,-0.02])
Tmin	-1.88([-2.54,-2.16])	-1.76([-2.27,-1.96])	-0.60 ([-0.63,-0.41])	**-0.11([-0.08,0.13])**
July	60%	80%	98%	99%
Tmean	0.12([0.02,0.23])	0.86([0.62,0.92])	2.26([2.07,2.82])	2.55([2.43,3.22])
Tmax	0.13([0.02,0.24])	0.88([0.60,0.91])	2.25([2.17,2.66])	2.54([2.50,3.10])
Tmin	0.14([0.19,0.38])	0.88([0.77,0.96])	**2.28([1.81,2.04])**	**2.55([2.00,2.32])**
September	1%	2%	30%	50%
Tmean	-2.11([-2.34,-1.99])	-1.91([-2.10,-1.80])	-0.59([-0.66,-0.42])	-0.04([-0.13,0.10])
Tmax	**-2.10([-2.0,-1.78])**	**-1.91([-1.86,-1.60])**	-0.59([-0.63,-0.46])	-0.05([-0.18,0.01])
Tmin	-2.09([-2.59,-2.23])	-1.86([-2.33,-1.97])	-0.60([-0.60,-0.38])	-0.06([-0.06,0.16])
September	60%	80%	98%	99%
Tmean	0.25([0.14,0.36])	0.89([0.74,0.97])	2.15([1.89,2.44])	2.48([2.20,2.80])
Tmax	0.24([0.05,0.26])	0.88([0.65,0.91])	2.13([1.99,2.52])	2.48([2.26,2.87])
Tmin	0.22([0.19,0.42])	0.87([0.77,0.98])	**2.13([1.77,2.00])**	**2.49([1.96,2.26])**

TABLE 11.7 – Quantiles of the reduced temperature of Tmean, Tmax, Tmin for fixed months and their 95% confidence intervals.

Months	Tmean	Tmax	Tmin
January			
$\mu(I\mu)$	1.29(1.24,1.48)	1.30(1.20,1.38)	1.44(1.33,1.54)
$\sigma(I\sigma)$	0.45(0.44,0.69)	0.51(0.42,0.71)	0.41(0.41,0.64)
$\xi(I\xi)$	-0.29(-0.48,-0.21)	-0.22(-0.25,-0.03)	-0.36(-0.44,-0.17)
February			
$\mu(I\mu)$	1.26(1.22,1.45)	1.35(1.23,1.43)	1.43(1.37,1.56)
$\sigma(I\sigma)$	0.48(0.44,0.63)	0.60(0.48,0.67)	0.46(0.41,0.54)
$\xi(I\xi)$	-0.27(-0.39,-0.14)	**-0.25(-0.20,0.00)**	-0.35(-0.44,-0.21
March			
$\mu(I\mu)$	1.41(1.29,1.54)	1.61(1.39,1.67)	1.52(1.43,1.60)
$\sigma(I\sigma)$	0.56(0.51,0.67)	0.68(0.61,0.83)	**0.41(0.39,0.53)**
$\xi(I\xi)$	-0.31(-0.31,-0.10)	**-0.20(-0.18,0.03)**	-0.35(-0.43,-0.17)
April			
$\mu(I\mu)$	1.66(1.40,1.65)	1.72(1.54,1.86)	1.57(1.48,1.65)
$\sigma(I\sigma)$	0.57(0.56,0.72)	**0.61(0.66,0.86)**	0.46(0.39,0.51)
$\xi(I\xi)$	-0.30(-0.30,-0.06)	**-0.33(-0.24,-0.01)**	-0.34(-0.36,-0.12)
May			
$\mu(I\mu)$	1.68(1.42,1.68)	1.69(1.56,1.82)	1.53(1.49,1.67)
$\sigma(I\sigma)$	0.67(0.61,0.80)	0.62(0.61,0.81)	0.48(0.40,0.52)
$\xi(I\xi)$	**-0.30(-0.26,0.00)**	**-0.30(-0.26,0.02)**	-0.27(-0.37,-0.13)
June			
$\mu(I\mu)$	1.70(1.46,1.72)	1.69(1.53,1.84)	1.59(1.49,1.67)
$\sigma(I\sigma)$	0.71(0.65,0.86)	0.65(0.65,0.83)	**0.53(0.42,0.55)**
$\xi(I\xi)$	-0.19(-0.19,0.05)	-0.25(-0.25,0.18)	-0.31(-0.40,-0.15)
July			
$\mu(I\mu)$	1.74(1.39,1.74)	1.80(1.54,1.84)	1.60(1.49,1.68)
$\sigma(I\sigma)$	0.71(0.66,0.85)	0.71(0.66,0.87)	**0.62(0.43,0.56)**
$\xi(I\xi)$	-0.20(-0.21,0.01)	-0.31(-0.32,-0.05)	-0.35(-0.41,-0.13)
August			
$\mu(I\mu)$	1.62(1.38,1.70)S	1.71(1.56,1.83)	1.50(1.50,1.69)
$\sigma(I\sigma)$	0.68(0.62,0.83)	0.60(0.60,0.80)	0.51(0.41,0.56)
$\xi(I\xi)$	-0.29(-0.31,-0.07)	-0.24(-0.31,-0.02)	-0.27(-0.36,-0.09)
September			
$\mu(I\mu)$	1.53(1.39,1.67)	1.62(1.50,1.80)	1.56(1.50,1.67)
$\sigma(I\sigma)$	0.60(0.57,0.75)	0.68(0.59,0.79)	**0.39(0.41,0.54)**
$\xi(I\xi)$	-0.20(-0.27,-0.04)	-0.24(-0.29,-0.03)	-0.27(-0.36,-0.14)
October			
$\mu(I\mu)$	1.46(1.36,1.58)	1.57(1.44,1.75)	1.51(1.47,1.65)
$\sigma(I\sigma)$	0.55(0.50,0.65)	0.57(0.57,0.76)	0.46(0.39,0.53)
$\xi(I\xi)$	-0.18(-0.30,-0.07)	-0.22(-0.30,-0.08)	-0.35(-0.40,-0.16)
November			
$\mu(I\mu)$	1.45(1.40,1.57)	1.50(1.43,1.68)	1.55(1.38,1.59)
$\sigma(I\sigma)$	0.50(0.44,0.60)	0.61(0.53,0.69)	0.47(0.42,0.54)
$\xi(I\xi)$	-0.32(-0.36,-0.14)	-0.31(-0.31,-0.08)	-0.36(-0.38,-0.12)
December			
$\mu(I\mu)$	1.44(1.40,1.57)	1.40(1.39,1.61)	1.52(1.37,1.56)
$\sigma(I\sigma)$	0.44(0.43,0.60)	0.56(0.49,0.64)	0.45(0.42,0.55)
$\xi(I\xi)$	-0.24(-0.38,-0.15)	-0.29(-0.31,-0.10)	-0.34(-0.34,-0.10)

TABLE 11.8 – Estimated extreme parameters from the reduced series of Tmean, Tmax Tmin and their 95% confidence intervals calculated from simulations.

Months	Tmean	Tmax	Tmin
January			
$\mu(I\mu)$	-1.41(-1.59,-1.29)	-1.54(-1.60,-1.28)	-1.40(-1.68,-1.42)
$\sigma(I\sigma)$	0.81(0.61,0.87)	0.84(0.65,0.94)	0.76(0.59,0.88)
$\xi(I\xi)$	-0.30 (-0.38,-0.12)	-0.27(-0.33,-0.05)	-0.25(-0.30,-0.05)
February			
$\mu(I\mu)$	-1.33(-1.54,-1.32)	-1.35(-1.54,-1.28)	-1.45(-1.70,-1.45)
$\sigma(I\sigma)$	0.68(0.58,0.81)	0.72(0.54,0.72)	0.60(0.57,0.79)
$\xi(I\xi)$	-0.13(-0.41,-0.08)	-0.11(-0.24,-0.03)	-0.11(-0.30,-0.05)
March			
$\mu(I\mu)$	-1.39(-1.57,-1.34)	-1.35(-1.48,-1.33)	-1.57(-1.80,-1.53)
$\sigma(I\sigma)$	0.62(0.55,0.72)	0.59(0.42,0.60)	0.56(0.54,0.74)
$\xi(I\xi)$	-0.20(-0.39,-0.13)	-0.17(-0.29,-0.04)	-0.21(-0.31,-0.10)
April			
$\mu(I\mu)$	-1.39(-1.58,-1.36)	**-1.35(-1.53,-1.37)**	-1.66(-1.83,-1.59)
$\sigma(I\sigma)$	0.58(0.47,0.62)	0.44(0.37,0.52)	0.55(0.52,0.70)
$\xi(I\xi)$	-0.30(-0.36,-0.12)	-0.26(-0.36,-0.12)	-0.32(-0.34,-0.09)
May			
$\mu(I\mu)$	-1.38(-1.52,-1.34)	-1.40(-1.55,-1.40)	-1.70(-1.90,-1.60)
$\sigma(I\sigma)$	0.49(0.39,0.53)	0.38(0.36,0.48)	0.50(0.50,0.70)
$\xi(I\xi)$	-0.36(-0.38,-0.17)	-0.30(-0.40,-0.14)	-0.30(-0.37,-0.14)
June			
$\mu(I\mu)$	-1.33(-1.50,-1.33)	**-1.37(-1.55,-1.39)**	-1.71(-1.81,-1.53)
$\sigma(I\sigma)$	0.42(0.39,0.42)	0.37(0.35,0.46)	0.54(0.53,0.68)
$\xi(I\xi)$	-0.25(-0.42,-0.19)	-0.28(-0.41,-0.13)	**-0.35(-0.34,-0.12)**
July			
$\mu(I\mu)$	-1.31(-1.48,-1.31)	**-1.32(-1.53,-1.37)**	-1.75(-1.82,-1.56)
$\sigma(I\sigma)$	0.40(0.38,0.50)	0.34(0.34,0.46)	0.54(0.52,0.68)
$\xi(I\xi)$	-0.20(-0.44,-0.18)	-0.16(-0.37,-0.13)	-0.35(-0.38,-0.14)
August			
$\mu(I\mu)$	**-1.29(-1.50,-1.33)**	-1.35(-1.52,-1.35)	-1.69(-1.80,-1.56)
$\sigma(I\sigma)$	0.46(0.39,0.53)	0.40(0.34,0.46)	0.52(0.52,0.69)
$\xi(I\xi)$	-0.27(-0.44,-0.17)	-0.20(-0.38,-0.13)	-0.36(-0.39,-0.12)
September			
$\mu(I\mu)$	-1.48(-1.58,-1.36)	-1.47(-1.52,-1.35)	-1.75(-1.84,-1.56)
$\sigma(I\sigma)$	0.50(0.46,0.62)	0.46(0.37,0.49)	**0.51(0.55,0.69)**
$\xi(I\xi)$	-0.26(-0.37,-0.12)	-0.23(-0.38,-0.11)	-0.40(-0.48,-0.01)
October			
$\mu(I\mu)$	-1.51(-1.65,-1.41)	-1.43(-1.57,-1.40)	-1.71(-1.80,-1.56)
$\sigma(I\sigma)$	0.55(0.53,0.68)	0.53(0.45,0.60)	0.56(0.56,0.72)
$\xi(I\xi)$	-0.25(-0.34,-0.12)	-0.17(-0.29,-0.05)	-0.32(-0.33,-0.06)
November			
$\mu(I\mu)$	-1.57(-1.72,-1.46)	-1.61(-1.68,-1.44)	-1.57(-1.78,-1.53)
$\sigma(I\sigma)$	0.60(0.54,0.73)	0.62(0.55,0.73)	**0.55(0.59,0.77)**
$\xi(I\xi)$	-0.30(-0.33,-0.09)	-0.24(-0.27,-0.05)	-0.27(-0.29,-0.04)
December			
$\mu(I\mu)$	-1.54(-1.67,-1.44)	-1.62(-1.76,-1.44)	-1.56(-1.75,-1.45)
$\sigma(I\sigma)$	0.61 (0.55,0.74)	0.65(0.61,0.82)	**0.56(0.60,0.77)**
$\xi(I\xi)$	-0.26(-0.33,-0.11)	-0.24(-0.27,-0.03)	-0.26(-0.27,-0.05)

TABLE 11.9 – Estimated left extreme parameters from the reduced series of Tmean, Tmax Tmin and their 95% confidence intervals calculated from simulations

Bibliographie

[1] Beskos A. and Roberts G.O. Exact simulation of diffusions. *JRSS serie B*, pages 332 –382, 2006.

[2] P. Alaton, Djehiche B., and Stillberger D. On modelling and pricing weather derivatives. *Applied Mathematical Finance*, 9, 2002.

[3] J.M.P Albin. On extremal theory for stationary processes. *Ann. Probab.*, 18 :92–128, 1990.

[4] L.V. Alexander et al. Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research*, 111, 2006.

[5] Klein Tank A.M.G. and Knnen G.P. Trends in indices of daily temperature and precipitation extremes in europe, 1946-99. *J. Clim*, 16 :3665–3680, 2003.

[6] C.W. Anderson and K.F. Turkman. Continuous-and discrete-time extremes : some theoretical and practical aspects. under preparation, 2008.

[7] J.-M. Azaïs and M. Wschebor. Régularité de la loi du maximum de processus gaussiens réguliers. *C.R. Acad. Sci. Paris*, 328 :333 – 336, 1999.

[8] S.M. Barbosa. Changing seasonality in europes air temperature. *Eur. Phys. J. Special Topics*, 174 :81–89, 2009.

[9] S. M. Berman. Extreme sojourns of diffusion processes. *Ann. Probab.*, 16 :361–374, 1988.

[10] S.M. Berman. *Sojourns and extremes of stochastic processes*. Waddsworth & Brooks/Cole statistics/probability series, 1992.

[11] P.J. Bickel and Ritov Y. Testing for goodness of fit : a new approach. *In : AKMdE Saleh (ed) Nonparametric statistics and related topics*, pages 51–57, 1992.

[12] D. Brody, J. Syroka, and Zervos M. Dynamical pricing of weather derivatives. *Quantitative Finance*, 2 :189–198, 2002.

[13] L.D. Brown and M. Levine. Variance estimation in nonparametric regression via the difference sequence method. *Ann. Stat.*, 35 :2219–2232, 2007.

[14] S. D. Campbell and F. X. Diebold. Weather forecasting for weather derivatives. *Journal of the American Statistical Association*, 100, 2002.

[15] M. Cao, A. Li, and J. Wei. Precipitation modelling and contract valuation : A frontier in weather derivatives. *The Journal of Alternative Investments*, 2004.

[16] M. Cao and J. Wei. Pricing weather derivatives : an equilibrium approach. Working paper, Queens University Kingston and University of Toronto, Ontario, Canada., 2000.

[17] V. Chavez-Demoulin. *Two problems in environmental statistics : capture- recapture analysis and smooth extremal models.* PhD thesis, Ecole Polytechnique Fdrale de Lausanne, 1999.

[18] V. Chavez-Demoulin and A.C. Davison. Generalized additive models for sample extremes. *Appl. Statist.*, 54 :207–222, 2005.

[19] M. Chen and H. An. The probabilistic properties of the nonlinear autoregressive model with conditional heteroskedasticity. *Acta Mathematicae applicatae sinica*, 15, 1999.

[20] C.K. Chu and J.S. Marron. Comparison of two bandwidth selectors with dependent errors. *Ann. Stat.*, 19 :1906–1918, 1991.

[21] W.S. Cleveland. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74 :829–836, 1979.

[22] S. Coles. *An Introduction to Statistical Modelling of Extreme Values.* Springer, 2001.

[23] Climate Research Committee. *Natural Climate Variability on Decade-to-Century Time Scales.* National Academy Press, 1995.

[24] I. Cordery, R. Mehrotra, and M.J. Nazemosadat. How reliable are standard indicators of stationarity ? *Stoch. Environ. Res. Risk Assess.*, 21 :765771, 2007.

[25] A. Cowling, P. Hall, and M.J. Phillips. Bootstrap confidence regions for the intensity of a poisson point process. *Journal of the American Statistical Association*, 91(15161524), 1996.

[26] D.D. Cox and F. O'Sullivan. Asymptotic analysis of penalized likelihood and related estimators. *Ann. Stat.*, 18 :1676–1695, 1990.

[27] D.D. Cox and F. O'Sullivan. Penalized likelihood-type estimators for genaralized nonparametric regression. *Journal of multivariate analysis*, 516 :185–206, 1996.

[28] D. Dacunha-Castelle and M. Duflo. *Probabilités et Statistiques vol 2.* Springer, 1984.

[29] D. Dacunha-Castelle and D. Florens-Zmirou. Estimation of the coefficients of a diffusion from discrete observations. *Stochastics*, 19 :263–284, 1986.

[30] D. Dacunha-Castelle and E. Gassiat. Testing the order of a model using locally parameterization : population mixtures and stationary arma processes. *Ann. Stat.*, 4, 1997.

[31] R. Davis and W. McCormick. Estimation for first-order autoregressive processes with positive or bounded innovations. *Stochastic Processes and their Applications*, 31 :237–250, 1989.

[32] A.C. Davison. *Statistical models.* Cambridge series in probability and statistics, 2003.

[33] A.C. Davison and N.L. Ramesh. Local likelihood smoothing of sample extremes. *J.R. Statist. Soc. B.*, 19, 1999.

[34] Percival D.B. and Rothrock D.A. 'Eyeballing' trends in climate time series : A cautionary note. *J. Clim.*, 18 :886–891, 2005.

[35] L. de Hann. On regular variation and its application to the weak convergence of sample extremes. *Thesis, University of Amsterdam/ Mathematical Centre Tract*, 1970.

[36] L. de Hann and A. Ferreira. *Extreme value theory : An introduction.* Springer series in operations research and financial engineering, 2006.

[37] S. Dessai and M. Hulme. Does climate policy need probabilities ? *Tyndall Centre Working Paper No.34*, page 43, 2003.

[38] P. Doukhan. *Mixing : properties and examples.* Springer-Verlag, 1994.

[39] Benth F. E. and J. Saltyte-Benth. Stochastic modelling of temperature variations with a view towards weather derivatives. *Statistical Research Report*, 2004.

[40] D.R. Easterling, G.A. Meehl, C. Parmesan, S.A. Changnon, T.R. Karl, and L.O. Mearns. Climate extremes : Observations, modeling, and impacts. *Science*, 289 :2068–2074, 2000.

[41] P. Embrechts, C. Klppelberg, and T. Mikosch. *Modelling extremal events for insurance and finance.* Chapman & Hall, 1997.

[42] Klein Tank A.M.G. et al. Daily datasets of 20th-century surface air temperature and precipitation series for the european climate assessment. *Int. J. of Clim.*, 22 :1441–1453, 2002.

[43] Comte F. and Rozenholc Y. A new algorithm for fixed design regression and denoising. *Annals of the Institute of Statistical Math*, 56 :449–473, 2004.

[44] Dornier F. and Querel M. Pricing weather derivatives by marginal value. *Quantitative Finance*, 1, 2000.

[45] J. Fan. Design-adaptive nonparametric regression. *Journal of the American Statistical Association*, 87 :998–1004, 1992.

[46] J. Fan. Local linear regression smoothers and their minimax efficiency. *Ann. Stat.*, 21 :196–216, 1993.

[47] J. Fan and I. Gijbels. *Local polynomial modelling and its application.* Chapman and Hall, 1996.

[48] J. Fan and Q. Yao. Efficient estimation of conditional variance functions in stochastic regression. *Biomeka*, 85 :645–660, 1998.

[49] J. Faraway. Bootstrap selection of bandwidth and confidence bands for nonparametric regression. *J. Stat. Comput. Simul.*, 37 :3744, 1990.

[50] C. A. Ferro, A. Hannachi, and D. Stephenson. Simple non-parametric techniques for exploring changing probability distributions of weather. *J. Clim.*, 18 :4344 4354, 2005.

[51] D. Florens-Zmirou. Estimation de la variance d'une diffusion partir d'une observation discrétisée. *C.R.A.S., t.309, Srie I*, pages 195–200, 1989.

[52] R.V Foutz and R.C. Srivastava. Statistical inference for markov processes when the model is incorrect. *Advances in Applied Probability*, 11 :737–749, 1979.

[53] M. Francisco-Fernández and J.M. Vilar-Fernández. Local polynomial regression estimation with correlated error. *Communications in Statistics : Theory and Methods*, 30(7) :1271–1293, 2001.

[54] M. Francisco-Fernández and J.M. Vilar-Fernández. Bandwidth selection for the local polynomial estimator under dependence : a simulation study. *Computational Statistics*, 20 :539–558, 2005.

[55] P.H. Franses and R. Paap. *Periodic Time Series Models.* Oxford University Press, 2004.

[56] P. Frich et al. Observed coherent changes in climatic extremes during the second half of the twentieth century. *Climate Research*, 19(193-212), 2002.

[57] E. Gobet, M. Hoffmann, and M. Reib. Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Stat.*, 32 :2223–2253, 2004.

[58] J. Granger, C.W. Seasonality : Causation, interpretation and implications. *Seasonal Analysis of econometric time series*, 1978.

[59] P.J. Green and Silverman B.W. *Nonparametric regression and generalized linear models : A roughness penalty approach.* Chapman and Hall, 1994.

[60] C. Gu. Cross-validating non- gaussian data. *Journal of Computational and Graphical Stat.*, 1 :169–179, 1992.

[61] C. Gu. *Smoothing spline ANOVA models.* Springer, 2002.

[62] E.J. Gumbel. *Statistics of Extremes.* Columbia University Press, New York, 1958.

[63] X. Guyon and Souchet S. Estimation de Yule-Walker d'un CAR(p) observ temps discret. *Ann. IHP, Proba. et Stat.*, 38, 2002.

[64] Wang H. Nonlinear ARMA models with functional MA coefficients. *Journal of Time series analysis*, 29 :1032–1056, 2008.

[65] P. Hall. Edgeworth expansions for nonparametric density estimators, with applications. *Statistics*, 22 :215232, 1992.

[66] P. Hall. On bootstrap confidence intervals in nonparametric regression. *Ann. Statist.*, 20 :695–711, 1992.

[67] P. Hall and R.J. Caroll. Variance function estimation in regression : the effect of estimating the mean. *J.Roy.Statist.Soc.Ser. B*, 51 :3–14, 1989.

[68] P. Hall, S.N. Lahiri, and J. Polzehl. On bandwidth choice in nonparametric regression with both short and long-range dependent errors. *Ann. Stat.*, 23 :1921–1936, 1995.

[69] P. Hall and N. Tajvidi. Nonparametric analysis of temporal trend when fitting parametric models to extreme value data. *Statistical Science*, 15 :153–167, 2000.

[70] L.P. Hansen, J.A. Scheinkman, and N. Touzi. Indentification of scalar diffusions using eigenvectors. *Journal of econometrics*, 86 :1–32, 1998.

[71] E. Härdle, W.and Mammen. Comparing nonparametric versus parametric regression fits. *Ann. Stat.*, 21 :19261947, 1993.

[72] W. Härdle, P. Hall, and J.S. Marron. How far are automatically chosen regression smoothing parameters from their optimum ? *Journal of The American Statistical Association*, 83 :86–95, 1988.

[73] J.D. Hart. Automated kernel smoothing of dependent data by using time series cross-validation. *Journal of the Royal Statistical Association, Series B*, 56 :529–542, 1994.

[74] T. Hastie and C. Loader. Local regression : automatic kernel carpentry (with discussion). *Statistical Science*, 4 :120–143, 1993.

[75] T.J. Hastie and R.J. Tibshirani. *Generalized Additive Models*. Chapman and Hall, 1990.

[76] M.R. Haylock, N. Hofstra, A.M.G. Klein Tank, E.J. Klok, P.D. Jones, and M. New. A european daily high-resolution gridded data set of surface temperature and precipitation for 1950-2006. *Journal of Geophysical Research*, 113, 2008.

[77] T.T.H Hoang. *Séries chronologiques non stationnaires non linéaires. Le cas des séries de températures en Europe.* PhD thesis, Universit Paris Sud 11, 2009.

[78] T.T.H Hoang, D. Dacunha-Castelle, and Parey S. Testing non stationnarity. Application to trends and extremes for temperature in climate models. Submitted to *Extremes*, 2009.

[79] T.T.H Hoang, S. Parey, and D. Dacunha-Castelle. Multidimensional trends : the example of temperature. *European Physical Journal-Special Topics*, 174 :113–124, 2009.

[80] S. Hylleberg. *Seasonality in Regression.* Academic Press, Orlando, FL., 1986.

[81] S.H. Inge. Minimal conditions for weak convergence to a diffusion process on the line. *Ann. Probab.*, 9, 1981.

[82] Hart J.D. Kernel regression estimation with time series errors. *Journal of the Royal Statistical Society (B)*, 53 :173–187, 1991.

[83] S. Jewson and R. Caballero. Seasonality in the statistics of surface air temperature and the pricing of weather derivatives. *Meteorol. App.*, 10 :377–389, 2003.

[84] Angell J.K. Effect of exclusion of anomalous tropical stations on temperature trends from a 63-station radiosonde network, and comparison with other analyses. *J. Clim.*, 16 :2288–2295, 2003.

[85] R. Katz and Brown B. Extreme events in a changing climate : variability is more important than averages. *Clim. Change*, 21 :289 302, 1992.

[86] Richard W. Katz, M. B. Parlange, and P. Naveau. Statistics of extremes in hydrology. *Adv. Water Ressource*, 25 :12871304, 2002.

[87] M. Kessler. Estimation of an ergodic diffusion from discrete observations. *Scandinavial journal of Statistics*, 24 :211–229, 1997.

[88] M. Kessler and M. Sorensen. Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli*, 5(299-314), 1999.

[89] D. Kiktev, D.M.H. Sexton, L. Alexander, and Folland C.K. Comparison of modeled and observed trends in indices of daily climate extremes. *J. Clim.*, 16 :3560–3571, 2003.

[90] A. M. G. Klein Tank, G. P. Können, and F. M. Selten. Signals of anthropogenic influence on european warming as seen in the trend patterns of daily temperature variance. *Int. J. Climatol.*, 25 :1–16, 2005.

[91] S.N. Lahiri. *Resampling methods for dependent data.* Springer, 2003.

[92] M. R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and related properties of random sequences and series,.* Springer Verlag, New York, 1983.

[93] C. Loader. *Local Regression and Likelihood.* Springer, 1999.

[94] J. Lu, G. A. Vecchi, and T. Reichler. Expansion of the hadley cell under global warming. *Geophysical Research Letter*, 34, 2007.

[95] Moreno M. Evaluation des dérivés climatiques. *Speedwell Weather Derivatives Limited*, 2000.

[96] Mudelsee M., Brngen M. Deutsch M., and Tetzlaff G. Trends in flood risk of the River Werra (Germany) over the past 500 years. *Hydr. Sciences J.*, 51 :833, 2006.

[97] J.S. Marron. Partiontioned cross-validation. *Econometric Rev.*, 6 :271–284, 1987.

[98] E. Masry and J. Fan. Local polynomial estimation of regression function for mixing processes. *Scandinavian Journal of Statistics*, 24 :165–179, 1997.

[99] G.A. Meehl, F. Zwiers, J. Evans, T. Knutson, L.O. Mearns, and P. Whetton. Trends in extreme weather and climate events : Issues related to modeling extremes in projections of future climate change. *Bulletin of the American Meteorological Society*, 81 :427–436, 2000.

[100] M. Mudelsee. Break function regression. *The European Physical Journal Special Topics*, 174(49-63), 2009.

[101] H.G. Müller and U. Stadtmüller. Estimation of heteroscedasticity in regression analysis. *Ann. Stat.*, 15 :610–625, 1987.

[102] M. Nerlove. Spectral analysis of seasonal adjustment procedures. *Econometrica*, 32 :241–286, 1964.

[103] M.H. Neumann. Automatic bandwidth choice and confidence intervals in nonparametric regression. *Ann. Stat.*, 23 :1937–1959, 1995.

[104] M.H. Neumann. Pointwise confidence intervals in nonparametric regression with heteroscedastic error structure. *Statistics*, 29 :1–36, 1997.

[105] M. Nogaj, S. Parey, and D. Dacunha Castelle. Non stationary extreme models and a climatic application. *Nonlin. Processes. Geophys.*, 14 :305 316, 2007.

[106] Altman N.S. Kernel smoothing of data with correlated errors. *Journal of the American Statistical Association*, 85 :749–759, 1990.

[107] J.D. Opsomer. Nonparametric regression in the presence of correlated errors. *Modelling Longitudinal and Spatially correlated data : Methods, Application and Future directions*, pages 339–348, 1997.

[108] J.D Opsomer, Y. Wang, and Y. Yang. Nonparametric regression with correlated errors. *Statistical Science*, 16 :134–153, 2001.

[109] F. OSullivan, B.S. Yandell, and W.J. Raynor. Automatic smoothing of regression functions in generalized linear models. *Journal of the American Statistical Association*, 81 :96–103, 1986.

[110] Yiou P., Dacunha-Castelle D., Parey S., and Hoang T.T.H. Statistical representation of temperature mean and variability in europe. *Geophysical Research Letters*, 36, 2009.

[111] S. Parey. Extremely high temperatures in france at the end of the century. *Climate Dynamics*, 30 :99–112, 2008.

[112] S. Parey, D. Dacunha-Castelle, and T.T.H. Hoang. Mean and variance evolutions of the hot and cold temperatures in europe. *Climate Dynamics*, 2009.

[113] S. Parey, T.T.H Hoang, and D. Dacunha-Castelle. Different ways to compute temperature return levels in the climate change context. Submited to *Environmetrics*, 2009.

[114] S. Parey, F. Malek, C. Laurent, and D. Dacunha-Castelle. Trends and climate evolution : Statistical approach for very high temperatures in france. *J. Clim.*, 81 :331–352, 2007.

[115] F. Pauli and S.G. Coles. Penalized likelihood inference in extreme value analyses. *Journal of Applied Statistics*, 28 :547–560, 2001.

[116] R. Perfeckt. Extremal behaviour of stationart markov chains and application. *The Annals of Applied Probability*, 4 :529–548, 1994.

[117] S. Pezzulli, D.B. Stephenson, and A. Hannachi. The variability of seasonality. *J. Clim.*, 18 :71–88, 2005.

[118] J. Pickands. Statistical inference using extreme order statistics. *Ann. Stat.*, 3 :119–131, 1975.

[119] V. Piterbarg. Discrete and continuous time extremes of gaussian processes. *Extremes*, 7 :161–177, 2004.

[120] B.L.S. Prakasa-Rao. Statistical inference from sampled data for stochastic processes. *Contemp. Math.*, 80 :249–284, 1988.

[121] P. Prescott and A.T. Walden. Maximum likelihood estimation of the parameters of the generalized extreme value distribution. *Biometrika*, 67 :723–724, 1980.

[122] Davis R.A. Maximum and minimum of one-dimensional diffusions. *Stochastic Processes and their applications*, 13 :1–9, 1982.

[123] N. I. Ramesh and A. C. Davison. Local models for exploratory analysis of hydrological extremes. *Journal of Hydrology*, 256 :106 119, 2002.

[124] C. Reinsch. Smoothing by spline functions. *Numerische Mathematik*, 10 :177–183, 1967.

[125] J. Rice and M. Rosenblatt. Smoothing splines : regression, derivatives and deconvolution. *Ann. Stat.*, 11 :141–156, 1983.

[126] C.Y. Robert, J. Segers, and C.A.T. Ferro. A sliding blocks estimator for the extremal index. *Electronic Journal of Statistics*, 3 :993–1020, 2009.

[127] H. Rootzén. Weak convergence of the tail empirical function for dependent sequences. *Stoch. Proc. Appl.*, 119 :468–490, 2009.

[128] G. Rosen and A. Cohen. Extreme percentile regression. *Statistical Theory and Computational Aspects of Smoothing : Proceedings of the COMPSTAT '94 satelilite meeting*, pages 200–214, 1994.

[129] O. Roustand, Laurent J.P., X. Bay, and L. Carraro. Model risk in the pricing of weather derivatives. *Ecole des Mines, Saint- Etienne and Ecole ISFA, Villeurbanne*, 2003.

[130] D. Ruppert and M.P. Wand. Multivariate locally weighted least squares regression. *Ann. Stat.*, 22 :1346–1370, 1994.

[131] D. Ruppert, M.P. Wand, U. Holst, and O. Hössjer. Local polynomial variance function estimation. *Technometrics*, 39 :262–273, 1997.

[132] H. Schäbe. Nonparametric estimation of intensities of nonhomogeneous poisson processes. *Statistical Papers*, 34 :113–131, 1993.

[133] C. Schär, P.L. Vidale, D. Lüthi, C. Frei, C. Hberli, M.A. Liniger, and Appenzeller C. The role of increasing temperature variability for european summer heat waves. *Nature*, 427 :332–336, 2004.

[134] I.J. Schoenberg. Spline functions and the problem of graduation. *Proceedings of the National academy of sciences of the United States*, 52 :947–950, 1964.

[135] C.D. Schönwise, Fuchs andT. Rapp, J., and M. Denhard. Observed climate trends in europe 1891-1990. *Meteorologische Zeitschrift*, 3 :22–28, 1994.

[136] C. Schwierz, C. Appenzeller, H.C. Davies, M.A. Liniger, W. Müller, T.F. Stocker, and M. Yoshimori. Challenges posed by and approaches to the study of seasonal-to-decadal climate variability. *Climate change*, 2006.

[137] Resnick S.I. *Extreme values, regular variation and point processes.* Springer- Verlag, 1987.

[138] Resnick S.I. *Heavy-Tail Phenomena : Probabilistic and Statistical Modeling.* Springer- Verlag, 2007.

[139] J.S. Simonoff. *Smoothing methods in Statistics.* Springer, 1996.

[140] R.L. Smith. Maximum likelihood estimation in a class of a non-regular cases. *Biometrika*, 72 :67–90, 1985.

[141] R.L. Smith. Extreme value analysis of environmental time series : an example based on ozone data (with discussion). *Statistical Science*, 4 :367–393, 1989.

[142] R.L. Smith, J. Tawn, and S. Coles. Markov chain models for threshold exceedances. *Biometrika*, 84 :249–268, 1997.

[143] Q. Song and L. Yang. Spline confidence bands for variance funcions. *Journal of Nonparametric Statistics*, 21 :589–609, 2009.

[144] S. Souchet. *Estimation des paramtres dune diffusion ergodique observe temps discret*. PhD thesis, lUniversit Paris 1, 1998.

[145] C.J. Stone. Consistent nonparametric regression. *Ann. Stat.*, 5 :595–620, 1977.

[146] G. Wahba. Smoothing noisy data with spline functions. *Numerische Mathematik*, 24 :383–393, 1975.

[147] J. Wang and L. Yang. Polynomial spline confidence bands for regression curves. *Statist. Sinica*, 19 :325–343, 2009.

[148] Y. Wang. Smoothing spline models with correlated random errors. *Journal of the American Statistical Association*, 93 :341–348, 1998.

[149] E. Whittaker. On a new method of graduation. *Proceedings of the Edinburgh Mathematical society*, 41 :63–75, 1923.

[150] Z. Wu and N.E. Huang. Ensemble empirical mode decomposition : a noise-assisted data analysis method. *Adv. Adaptive Data Anal. (in press)*, 2008.

[151] Z. Wu, N.E. Huang, S.R. Long, and C.K. Peng. On the trend, detrending, and variability of nonlinear and nonstationary time series. *proc. Nat. Acad. Sci.*, 18, 2007.

[152] Z. Wu et al. The modulated annual cycle : an alternative reference frame for climate anomalies. *Climate Dynamics*, 31 :823–841, 2008.

[153] AitSahalia Y. Maximum-likelihood estimation of discretely-sampled diffusions : A closed-form approximation approach. *Econometrica*, 70 :223–262, 2002.

[154] Z. Yan et al. Trends of extreme temperatures in europe and china based on daily observations. *Climatic change*, 53(355-392), 2002.

[155] N. Yoshida. Estimation for diffusion processes form discree observations. *J. Multivar. Anal.*, 41 :220–242, 1992.