On nonparametric techniques for analyzing nonstationary signals

Douglas David Baptista de Souza

UNIVERSITÉ DE GRENOBLE

ÉCOLE DOCTORALE EEATS

Electronique, électrotechnique, automatique et traitement du signal

T H È S E

pour obtenir le titre de

DOCTEUR de l'Université de Grenoble

Mention : TRAITEMENT DU SIGNAL

Présentée et soutenue par

M. Douglas David Baptista de Souza

titre :

On nonparametric techniques for analyzing nonstationary signals

Directeurs de thèse : Jocelyn CHANUSSOT et Anne-Catherine FAVRE

préparée au laboratoire Grenoble, images, parole, signal, automatique(GIPSA-Lab) et sotenue le 08/10/2013

Jury :

Rapporteurs :	M. José BERMUDEZ	LPDS, Electrical Engineering Department, UFSC, Brésil
	M. Cédric RICHARD	Laboratoire Lagrange, UMR CNRS 729
Directeur :	M. Jocelyn CHANUSSOT	GIPSA-lab, Grenoble INP
Co-directrice :	Mme. Anne-Catherine FAVRE	LTHE, Grenoble INP
Président :	M. Patrick FLANDRIN	Laboratoire de Physique, ENS Lyon UMR CNRS 5672
Examinateurs :	M. Pierre BORGNAT	Laboratoire de Physique, ENS Lyon UMR CNRS 5672
	M. Philippe NAVEAU	Laboratoire des Sciences du Climat et l'Environnement (LSCE) CNRS
	Mme Marie CHABERT	INP - ENSEEIHT Toulouse

Abstract

Testing stationarity is important in signal processing and in many areas of environmental sciences. In real world applications, a special attention should be given to the nonparametric methods, and also to the fact that we often do not know whether a change has occurred, nor do we have any idea where the possible change point(s) could be. The purpose of this thesis is to develop methods to test stationarity and to estimate the change point of real world signals, more specifically, environmental ones. In this work, the stationarity test and the change point detection are treated as two distinct steps.

The stationarity tests are performed by using two different frameworks: a modified version of an existing technique and a novel one. The existing technique makes use of surrogate resampling for testing stationarity. We propose different contributions to the original method for improving its performance, and for evaluating the robustness of the test against changing outcomes. For the novel framework, we propose to use empirical mode decomposition and block bootstrapping for testing stationarity. This new stationarity test detects trends or evolutions of the local energy of the signal. By comparing with other approaches in the literature, the new method allows for a better detection of slowly-varying nonstationarities of first and second-order.

The tests developed in this thesis reject the stationarity of the whole observation interval. Thus, we propose a framework for change point detection in nonstationary signals, which is based on an existing method called Robust Singular Spectrum Transform (RSST). The method that we develop is nonparametric, data-driven, can detect multiple change points and has shown to work for first and second-order nonstationarities.

Finally, we apply all the developed techniques to an environmental dataset, which corresponds to rainfall signals generated by the *Canadian Regional Climate Model* (CRCM), a very realistic model that allows for the study of climate change. The consistency of the obtained results confirms the potential of the developed methods.

Key words: stationarity test, time-frequency analysis, marginals, distances, trend detection, empirical mode decomposition, bootstrapping, change point detection, robust singular spectrum transform

Résumé étendu

Dans l'analyse des signaux d'origine naturelle, nous sommes souvent confrontés à des situations où nous ne savons pas si un changement s'est produit, ni où le possible point de changement peut être localisé. Cependant, diverses méthodes en traitement du signal reposent implicitement sur une hypothèse de stationnarité, car le cas stationnaire est bien défini dans une perspective théorique. D'un autre côté, tous les processus du monde réel sont a priori non-stationnaires et, dans la majorité des cas, cette supposition se révèle vraie. Etant donné qu'il existe de nombreuses façons par lesquelles la propriété de stationnarité peut être enfreinte, différents tests de stationnarité ont été développés pour tester les différentes formes de non-stationnarité. Cette thèse se concentre sur la conception et l'amélioration des techniques qui peuvent être appliquées aux signaux environnementaux, plus spécifiquement, les signaux hydrométéorologiques. Les techniques qui ont été développées présentent certaines caractéristiques qui sont préférables pour tester les données environnementales (*i.e.* être non-parametrique, être capable d'extraire automatiquement les informations des données disponibles, être capable d'identifier un changement dans les moments statistiques du premier et du second ordre). Dans cette thèse, le test de stationnarité et la détection de point de changement ont été abordés séparément: les tests de stationnarité rejettent la stationnarité de tout l'intervalle d'observation, tandis que pour détecter les points de changement, nous testons les signaux pour les quels la stationnarité a déjà été rejetée. Dans ce manuscrit, de nombreuses contributions et de nouvelles approches de ces sujets sont proposées.

Dans la première partie de la thèse, plusieurs contributions sont proposées au test de stationnarité par la méthode des substituts. Cette méthode consiste à échantilloner des signaux stationnaires disposant de la même densité d'énergie que l'observation. Après avoir décrit les principes de la technique considérée, nous effectuons une étude exploratoire pour identifier des métriques appropriéés pour caractériser la distance d'un spectre instantané à la densité spectrale du signal. Puis, nous appliquons la méthode en utilisant les métriques choisies sur différents exemples de séries temporelle non-stationnaires, où nous constatons les limitations de cette aproche. En particulier, la méthode difficilement détecter des non-stationnarités du premier ordre et celles cor-

respondant à des évolutions lentes. Ansi, une modification de la méthodologie originale est proposée. Cette modification consiste à accroitre le contraste du test grâce à une forme de métrique dépendant de l'énergie instantanée du signal à analyser. Cette métrique est neutre dans la caractérisation d'un signal stationnaire et met en évidence les effets d'un comportement non-stationnaire que se traduit également sur l'énergie instantanée. La dernière contribution du chapitre consiste à développer une méthode de bootstrap sur les vecteurs de paramètres extraits des substituts, afin d'évaluer l'évolution des résultats qui peuvent apparaître en appliquant le test plusieurs fois.

Indépendamment des améliorations apportées par les modifications proposées, la méthode des substitus présente toujours des limitations que impactant la détection des changements lents des moments du premier et du second ordre. Néanmoins, la détection de ces formes de nonstationnarité est d'une grande importance pour de nombreuses applications du monde réel. Par conséquent, dans la deuxième partie de la thèse, nous proposons un test original de stationnarité pour tester les dérives lentes. Cette nouvelle approche consiste à décomposer en tendances une estimée lissée de l'énergie instantanée du signal analysé (ou la marginale en temps de sa représentation temps-fréquence) en utilisant de la décomposition modale empirique, qui est une méthode de décomposition non-paramétrique dictée par la série temporelle analysée). Après avoir expliqué les principes de la nouvelle approche, nous définissons comme statistique décisionnelle la fraction des variances du signal d'énergie instantanée sur ce qu'il en reste lorsqu'on y a neutralisé les possibles tendances non-stationnaires. Il est démontré que, en cas d'un signal non-stationnaire, la statistique décisionnelle prendra une valeur supérieure à l'unité, sa valeur nominale en présence d'un signal stationnaire. Par ailleurs, il est vérifié que la distribution de la statistique décisionnele sous l'hypothèse de stationarité peut être modélise par une distribution de type generalized extreme value (GEV). Enfin, la nouvelle approche est comparée à deux techniques paramétriques de la littérature, et de très bons résultats sont obtenus.

Aprés avoir présenté les deux tests de stationarité qui sont abordés dans cette thèse, nous proposons une méthode pour la détection de changements dans les signaux qui auraient été qualifiés de non-stationnaires. Cette partie de la thèse est destinée à compléter l'étude développée dans les deux chapitres précédents, où les tests de stationnarité proposés sont faits pour rejeter la stationnarité de tout l'intervalle d'observation. Par conséquent, un algortihme spécifique pour détecter les points de changement se fait nécessaire. L'approche proposée est une évolution de la méthode non-paramétrique qui s'appelle *Robust Singular Spectrum Transform* (RSST). Cette dernière consiste à fenêtrer les données en amont et en aval de l'instant de temps consideré, de manière à définir un sous-espace généré par les données en amont et l'autre par les données en aval. Pour designer si l'instant courant est ou non un point de changement, nouns prenons en compte la valeur d'un paramètre d'intrication des deux sous-espaces. Nous proposons

d'analyser la valeur du degré d'intrication en fonction des deux fenêtres choisies pour analyser le signal, où le domaine de variation des deux fenêtres est automatiquement défini au moyen d'un critère empirique. Les instants de changement sont sélectionnés en considérant que l'incertitude associée aux degrés d'intrication observés devra augmenter si l'instant courant est un point de changement. Enfin, la méthode proposée est comparée à l'autre approche non-paramétrique de la littérature.

La dernière partie de la thèse consiste à appliquer toutes les approaches développées sur des données environnementales. Les séries temporelles considérées sont les précipitations journalières maximales par année, sur une période de 139 annés et en 1631 points distribués sur la province du Québec. Les données ont été générées par le *Canadian Regional Climate Model* (CRCM), un modéle très réaliste qui prend en compte de nombreuses interactions physiques complexes. Les donnés générées par ce modèle ont été utilisées pour les prévisions climatiques sur différentes échelles de temps. Les simulations sont effectuées en deux réalisations distinctes du CRCM. La cohérence des résultats obtenus confirme le potentiel des approches proposées au regard des approches concurrentes. Finalement, nous présentons le lieux où des séries non-stationnaires ont été détectées, ainsi que les instants de rupture correspondants.

Mots clés: test de stationnarité, analyse temps-fréquence, marginales, distances, détection de tendance, décomposition modale empirique, bootstrapping, détection de point de changement, robust singular spectrum transform

Acknowledgments

And then three years have passed! I remember it was a thursday evening, 22nd of april, 2010, and I was working at the lab in Florianópolis. While I was reading a paper, an email popped up in my mailbox saying that my application has been selected for a doctoral scholarship within the Erasmus Mundus program. In the days that followed that 22nd of april, I found myself contemplating a new and crazy future in France, which was drastically different from what I had planned up to that point. I was excited and at the same time scared about moving to a far country, without knowing the language or any person. After all, it was a three-year project abroad, far from where my family, love, friends, and old plans were. However, when unique opportunities are given, one should take advantage of them. So I jumped into this new chapter of my life, and with the support of those I care most about, the PhD has suddenly passed.

Even a quite individual journey such as writing a PhD thesis is all about team work. I am enormously grateful to three people, without whom my research would have been impossible. First, I thank my supervisor Jocelyn Chanussot, for giving me the life-changing opportunity to embark on this amazing doctoral journey, for his support, and for continuously conveying a spirit of excitement about my research. Second, I thank my co-supervisor Anne-Catherine Favre, for helping me in many practical issues that I faced in my PhD, and for kindly making herself available whenever I needed (even in non-working days). I am lucky to have had such friendly supervisors as Jocelyn and Anne-Catherine, who have guided and supported me with patience and knowledge, whilst giving me the freedom to work on my own way. Also, I would like to express my sincere gratitude to Pierre Borgnat, whose collaboration was instrumental for my research. His insights helped me a lot in particular aspects of the thesis. Working with him was a very educational experience and I hope that our collaboration continues.

I would like to express my gratitude to Patrick Flandrin for having accepted to preside the jury of my thesis. My research was heavily influenced by his papers and his book, and I believe that anyone writing a thesis about nonstationary signals or time-frequency analysis could not wish for a more illustrious jury president. Also, I would like to thank the other members of the thesis committee for coming to my defense and

for their helpful comments and ideas: Marie Chabert, Philippe Naveau, Cédric Richard and José Bermudez.

Some people have played an important role before and during my thesis. I cannot thank enough my first advisor in Brazil, Sidnei Noceti Filho, a wonderful person who is the responsible for me to be in the signal processing and in the research world. Also, I thank the few but good friends that I made in Grenoble: Humberto, Aditya, Berekmeri and Giorgio, for helping me getting through the lonely days at my department. I should also thank to all my friends in Brazil (now, it would be impossible to list all of them here), who have always cheered me up independently of my location (in time or in space). I am fortunate to have made so many good friends in life, who I will always look forward to see again.

Finally and most importantly, I dedicate this PhD thesis to my family and to my wonderful fiancée, Chris, for their endless support and love. They have always believed in me and always encouraged me to reach my goals. My family, Alceu, Dione, Rosane, Diego, Therezinha and my late grandfather, Luiz Henrique (my unending inspiration, the kindest person one could ever meet), I owe them everything and I hope that this work makes them proud. Lastly, I have no words to thank my wonderful and beloved fiancée, Chris, for everything she has been through during these three years and for all her support. Her love and encouragement meant everything to me and allowed me to finish this journey. From the very bottom of my heart, thank you!

Contents

Notations

Some notations used in the Thesis.

- j: $\sqrt{-1}$
- t: time variable
- f: frequency variable
- I: number of surrogates, or number of IMFs (depending on the application)
- T: length of the observed signal
- N: length of the time marginal (or number of points in time where the spectrograms are computed). Also, number of of points in time where the change point scores are computed (depending on the application)
- $h_k(t), k = 1, ..., K$: first K Hermite window functions
- n_{h}: length in time of the Hermite window functions
- $\mathbb{E}\{\cdot\}$: stochastic expectation
- $\mathbb{V}\mathrm{ar}\{\cdot\}$: stochastic variance
- $\mathbb{C}\mathrm{ov}\{\cdot\}$: stochastic covariance
- $\mathbf{W}_x(t, f)$: Wigner-Ville Spectrum
- $D(.,.)$: dissimilarity measure in frequency
- $c_n^{(x)}$: vector in time containing values of $D(.,.)$ computed for the original signal
- $c_n^{(s_i)}$: vector in time containing values of $D(.,.)$ computed for the i^{th} surrogate

- $\widetilde{c}_n^{(x)}$: weighted version of $c_n^{(x)}$
- $\widetilde{c}_n^{(s_i)}$: weighted version of $c_n^{(s_i)}$
- ξ_1 and ξ_2: initial and final parameter values for the Lombard's nonstationarity models
- $S(t,f)$: time-varying spectrum
- $S_K(t,f)$: multitaper spectrogram
- $S^{h_k}(t,f)$: short-time Fourier transform using Hermite functions as windows
- Θ_0: vector of variances of the distances between two spectra under null hypothesis
- INS: index of nonstationarity (for the surrogate-based test of Chapter 2)
- γ: threshold for the gamma hypothesis test
- n_{h0}: free parameter of the surrogate-based test
- $\lfloor x \rfloor$: floor function, *i.e.*, the largest integer not greater than x
- $\lceil x \rceil$: ceil function, *i.e.*, the smallest integer not less than x
- $y(t)$: time marginal
- $y_{\mathrm{dt}}(t)$: detrended time marginal
- $c(t)$: trend component
- $r(t)$: stationary (detrended) fluctuation
- E_{x}: energy of the signal x (*i.e.* $\langle x, x \rangle$)
- $m^{(i)}(t)$: i^{th} IMF
- $\rho^I(t)$: EMD residual
- P: number of bootstrap resamples
- B_k: k^{th} block of the block bootstrap technique
- ℓ: length of the block B_k
- $\widehat{\theta}_{\mathrm{TI}}$: trend importance estimator
- $\mathcal{I}_{\mathrm{NS}}$: index of nonstationarity (for the new stationarity test of Chapter 3)

- B^2: Zempleni statistic
- Θ_{TI}: vector of the estimates of the importance of the trend under null hypothesis
- $\mathcal{T}$: GEV threshold for the hypothesis test (for the new stationarity test of Chapter 3)
- $z(t, w, n)$: change point score at time t and window values w and n
- p_i: proportion of change point scores at the i^{th} iteration
- ζ: point at which p_i achieves its *maximum* (stopping point)
- $\mathcal{T}_z$: change score threshold for significant event
- $\mathcal{Z}$: final collection of scores that is used for estimating the change point
- $\mathbf{Z}_t$: matrix with change point scores at time t
- $\mathbf{Y}_t$: matrix with change point scores at time $t-1$
- $H(\cdot|\cdot)$: conditional entropy
- i.i.d.: independent and identically distributed
- WGN(0,1): white Gaussian noise with zero mean and unity variance
- AR(1): first-order autoregressive process
- FFT: fast Fourier transform
- PSD: power spectrum density
- EMD: empirical mode decomposition
- IMF: intrinsic mode function
- GEV: generalized extreme value distribution
- AMD: absolute mean deviation
- CP: change point
- SST: singular spectrum transform
- RSST: robust singular spectrum transform
- SSA: singular spectrum analysis
- PCA: principal component analysis

- CRCM: Canadian Regional Climate Model
- i, l, q: dummy indices used throughout the Thesis

Chapter 1

Introduction

Contents

This chapter presents an introduction to the concepts and applications concerned in the PhD's research. In this chapter, the importance of testing for stationarity is discussed and the nonstationary analysis is put in the context of environmental science. A brief

overview of the common approaches available in the literature is presented, and based on the application of interest we outline the required profile of the stationarity tests developed in the PhD's research.

1.1 Motivation

Over the last few decades much has been done in developing methods to test signals with respect to stationarity. Considering stationarity is important not only in signal processing and in many areas of technical science (*e.g.* telecommunications, electronics, automatic control, power systems) but also in the domain of environmental science (*e.g.* the study of biological and physical processes, the analysis of time series corresponding to rainfalls, global warming indicators or sunspot records...), to cite just a few. Many techniques devoted for analyzing the signals encountered in these fields simply assume the process to be stationary [1]. However, the assumption of stationarity often fails to be true, and the physical character of the signal requires a nonstationary approach [2].

One could expect a large number of stationarity tests to be available in the literature because stationarity is an important property of many signals. Nevertheless, one could suggest that a simple way to check the stationarity of a given signal is to investigate the physical mechanism underlying this signal. If the mechanism is time-invariant, then the resulting signal is stationary [3]. Unfortunately, drawing conclusion based on physical considerations is often impossible due to the following limitations:

- the inaccessibility of such information,
- the limited knowledge due to the unavoidable finite observation interval,
- the lack of precision in evaluating the time-invariance of the generative process.

While the latter point seems to be a technical or engineering problem, the first and second ones are inherent limitations when analyzing real world process. Therefore, we are forced to rely on statistical methods to evaluate the stationarity [3].

This thesis focuses on nonparametric methods for analyzing nonstationary signals that are suitable for real world processes, more specifically, hydro-meteorological ones. Probing this kind of data with respect to stationarity is important in many regards, as environmental time series can undergo changes in first and second-order statistics that reflect the behavior of natural phenomena in particular ways. Historically, the interest on practical tests for stationarity used to be under the scope of engineering and related disciplines. However, it has been long ago that the nonstationarity analysis

has came to the attention of the environmental sciences, mostly due to the increasingly interest in climate change over the past two decades. Thus, developing methods for identifying particular nonstationary behaviors is of a high importance in many areas of both technical and environmental sciences.

1.2 Why is it important to test stationarity?

The classical interpretation of stationarity refers to the invariance of the statistical properties relative to an absolute time. Even though the stationarity property could be defined for all orders, the most important properties of a stationary process are those of second order, which means that only the statistical properties of first and second orders should be time-invariant.

In practice, we often consider only the wide-sense stationarity (WSS). The conditions for a random signal $x(t)$ to be WSS are as follows:

1. its expectation is a constant, independent of the time

$$\mathbb{E}\{x(t)\} = \mu_x, \tag{1.1}$$

2. its autocovariance function depends only on the difference of the two considered instants

$$\mathbb{E}\{x(t)x^*(s)\} = \gamma_x(t-s). \tag{1.2}$$

Assuming the classical definition of stationarity for many processes, specially those corresponding to real world phenomena, is an illusion and so far from the reality that one could think at a first moment that signal analysis techniques relying on the stationary assumption are the exception, rather than the default. In reality, a large amount of work assume stationarity for applying standard algorithms. Moreover, while the stationary case is well-defined from a theoretical perspective, the nonstationarity is not a well-defined concept [4]. Even without a rigorous definition, a basic truth has to be acknowledged: all real world processes are *a priori* nonstationary, and in most cases this assumption turns out to be true. We may wonder what does it mean that a signal is nonstationary? This question has no unique answer, because the stationarity property can be violated in many ways [5]. Hence, instead of trying to define what a nonstationary process is in a general sense, we analyze how its properties differ from those of a stationary one [4]. Thus, having stationarity tests that are sensitive to different

forms of nonstationarites is required in order to approach the analysis of nonstationary phenomena [6].

1.2.1 Application to environmental data

1.2.1.1 Motivation

Scientists are concerned that the impact of human activity on nature may cause major changes [5]. This concern has been growing in popularity, as climate change became a prominent topic in both scientific and public discourses. In different areas of environmental science, researchers are interested in methods for identifying nonstationarity in hydrologic and climatic time series [7]. One climate-related aspect of nonstationarity that has been reported, for example, is that a change in the rain and snow dynamics can be seen [8]. Over the last years, the nonstationary analysis has been the subject of much discussion in the hydrological community, specially after the much cited paper in *Science* entitled *"Stationarity is Dead: Whither Water Management?"* [9], which has cast doubt on the value of historical data, and has defended the importance of recognizing the inherent nonstationary nature of the hydroclimatic processes.

1.2.1.2 Nonstationarities seen in environmental data

In environmental applications, we are mostly interested in testing for first and second-order stationarity. Although identifying changes in high-order statistics might be of particular interest (*e.g.* a change in the skewness might indicate a modification in the annual cycles induced by climate change, while a change in the kurtosis might indicate a modification in the behavior of the tails, which is related to the occurrence of floods), this Thesis is focused on methods for identifying changes specifically in the mean and the variance. When processing real world data, one has to consider not only the existence of many different types of nonstationarities, but also the fact that an important quantity of information can be found in the nonstationarity itself [4], like modulations, trends, abrupt changes and change points. Recognizing such nonstationary behaviors is of major importance in many areas of environmental science. For example, in climatology, the identification of a change point might indicate the beginning of a possible climate change [10]. In meteorology, characterizing periodic seasonal effect in time series of monthly atmospheric temperature can improve already existing forecasting models [11]. In hydrology, a proper identification of trends in hydrological processes is a crucial point for water resources management [12]. It is well-known that hydrological behaviors can also undergo abrupt changes [8], as the one that occurred in the series of the annual

volume discharge from the Nile River, which has been widely analyzed in the literature [13, 14, 15], and referred to as a consequence of an abrupt change that occurred in the rainfall regime near the turn of the 20th century [10, 16].

1.3 What we require from a method for analyzing real world signals?

1.3.1 A brief overview of the literature

In the analysis of real world signals, we are often in a situation where we do not know *whether a change occurred* nor do we have any idea where the possible *change point could be*. Putting in this form, one could clear envision the problem in two separate phases: the stationarity test followed by the change point detection. However, it is common practice in the literature to have methods encompassing both stages, which are often presented simply as algorithms for change point detection [17, 5, 18, 19]. In a slightly different point of view, some methods are concerned in deciding whether a segment of the observed time series is nonstationary [20, 21], where the change points are deemed as the start and end points of the nonstationary segment. One could carry this analysis further, by studying separately the two parts of the signal (before and after the nonstationary segment), and trying to decide (automatically) whether these two parts can be considered stationary, or whether another one can be discovered [6]. This practice is very important in recognition-oriented signal processing, and refers to the automatic segmentation of signals, *i.e.* the automatic decomposition of a given signal in local stationary pieces, the length of which varies according to the local characteristics of the signal [22, 23, 24].

Despite the alternatives presented, in real world applications, where often none information is given about the data, one needs to probe signals with respect to more general forms of stationarity, for larger classes of signals. In this regard, the choice of an efficient representation is a crucial point, since a good representation can provide much valuable information about the structure of the underlying signal in the absence of an external specification [4]. To this end, time-frequency (TF) approaches have been considered as powerful tools, as opposed to time-based techniques, which are more commonly found in the statistical literature. Regardless of the domain of analysis, the techniques can be broadly classified as parametric or nonparametric, depending on whether or not the probability distribution of the process can be completely specified by a finite number of parameters. Considering the categories of interest, we present in Table 1.1 a short

list of examples of applications and problems of methods for analyzing nonstationary signals.

Table 1.1: Comparison between different approaches for analyzing nonstationary signals. We list common applications and problems of parametric and nonparametric methods that analyze the process in the time or in the time-frequency domain.

Parametric approaches	**Nonpararametric approaches**
• **Time domain**	• **Time domain**
Applications: Detection of slowly-varying nonstationarities [20], and abrupt changes in the mean and in the variance [25]. Tests for trend stationarity [26, 27]. Tests for change in the mean [28, 29].	**Applications:** Detection of a changing mean [17, 30]. Detection of smooth and abrupt changes in the mean and in the variance [10, 18]. Detection of change points with linear trends [31].
Problems: In cases of signals presenting dependence structure. The necessity of strong assumptions about the statistical distribution of the data. Lack of robustness under unstable parameters.	**Problems:** In some cases, with signals dependence structure. Lack of sensitivity to outliers in the data.
• **Time-frequency domain**	• **Time-frequency domain**
Applications: Analysis of narrowband data, such as vibration signals [32], short signals with high frequency resolution [33], and slowly time-varying processes [22]. Segmentation in local stationary pieces [34, 35], applications in speech processing [23, 36]. Analysis of AM and FM signals [37]	**Applications:** Detection of abrupt changes [21]. Detection of transients embedded in noise [38, 39, 40]. Analysis of AM and FM signals [41, 42].
Problems: The successful application, without accurate *a priori* information, is very difficult in practice. Analysis of transient signals.	**Problems:** The necessity of a larger representational space than used for the original data. The inherent tradeoff related to the TF localization.

1.3.2 Outlining the desired technique

Having taken a look at some of the methods available in the literature, let us now have an idea of the type of time series we are aiming at testing. In Chapter 5, we apply the techniques developed in this Thesis and other ones found in the literature to rainfall time series generated by a climate model. More specifically, time series corresponding to the annual maximum daily precipitations simulated by the Canadian Regional Climate Model (CRCM) [43], corresponding to 1631 grid points in Canada. The signals we are interested in have a short length, which refers to a time span of 140 years (1961–2100).

This dataset is explained in details in Chapter 5, but in Fig. 1.1 we plotted as an example three annual maximum daily precipitation time series together with the corresponding location in Canada.

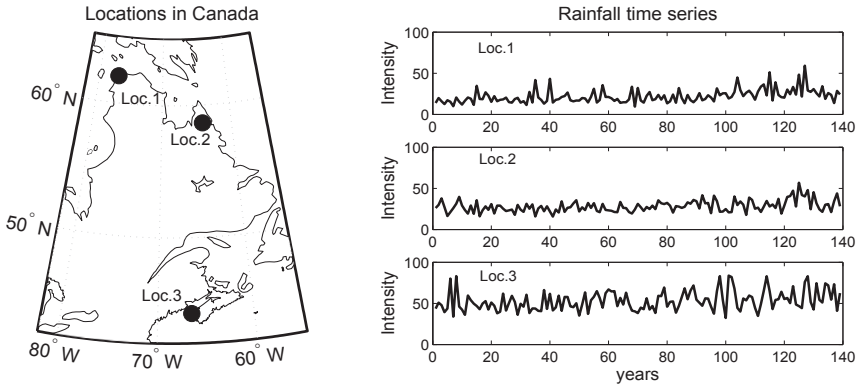

Figure 1.1: An illustration of the rainfall data analyzed in this Thesis: three annual maximum daily precipitation time series and the corresponding location in Canada.

Usually, visual inspection is the first step in the analysis of stationarity, as it can reveal or suggest possible nonstationary behaviors in the data, and make it easier to choose appropriate tests for stationarity [7]. However, only by visual inspection of the rainfall signals in Fig. 1.1, it is extremely difficult to identify any particular temporal structure, such as a sudden (abrupt) change, a modulation, or a trend, for example. In this case, representing the signal in TF domain is a good choice, as TF representations could tell us something about the structure of the signal that cannot be seen at a first glance: the time variation of the frequency content. It could be seen in Table 1.1 that a variety of options are given by TF techniques, which allow for testing more general forms of stationarity. Also, we shall confine ourselves to the class of *nonparametric approaches*, as we do not have any *a priori* knowledge about the functional forms of the signals, and all the information has to be determined entirely from the data [3]. Finally, we should have at our disposal methods that are robust against possible serial correlations in the data. Thus, some desired characteristics of a method for testing the signals shown in Fig. 1.1 could be listed as follows:

- Nonparametric,
- Possibly be given in TF domain, so as to offer a richer representation of the signal,
- Fully data-driven,
- Work for short samples,

- Allow for the (off-line) detection of possible change points or nonstationary segment,
- Sensitive to different types of nonstationarity,
- Account for potential serial correlations in the data.

Developing a framework that attend all the items shown above is particularly difficult, as there exist some conflicting requirements. For example, different from parametric techniques, nonparametric ones are not well-suited to test datasets of small size. Furthermore, in general, techniques that allow for the characterization of change points or nonstationary segments are either i) time-based and nonrobust under serial dependence [10, 18], ii) parametric and suited only to specific types of nonstationarities [20, 21], or iii) are not very reliable when testing small data records [6]. In order to take into account the different criteria listed above, the stationarity test and the change point problem are approached separately in this Thesis. In the following, we give a brief overview of the techniques proposed in this Thesis.

1.4 Stationarity tests

In the statistical literature, testing for stationarity usually means testing the null hypothesis of no change, against the alternative that there exists a time instant when the distribution of the time series changes. Stationarity tests in the TF domain usually consist in testing if a given observed characteristic in the time-varying spectrum of the signal is in accordance with what would be expected in a stationary situation. Thus one can usually compute a "distance from the stationary case" [4], which can be used as a statistic for testing the null hypothesis of stationarity.

1.4.1 First approach: Testing stationarity with surrogates

1.4.1.1 A brief description

The method of Chapter 2 tests if $\widehat{S}(t_n, f)$ – the time-varying spectrum estimated at N different time positions $\{t_n, n = 1, ..., N\}$ – is statistically similar to the global spectrum obtained by time averaging. The reason is that, under the null hypothesis, the time varying spectrum reduces to a stationary power spectral density (PSD): $\widehat{S}(t, f) \simeq \text{PSD}(f)$. Ideally, *stationarity* should correspond to an equivalence of the local spectra $\widehat{S}(t_n, f)$ evaluated at every instant $\{t_n, n = 1, ..., N\}$, and the global spectrum $\langle \widehat{S}(t_n, f) \rangle_n$ obtained by

time averaging. Thus, the *distance from the stationary case* could be measured by making use of the l_2-distance:

$$L(c_n, \langle c_n \rangle_n) = \frac{1}{N} \sum_{n=1}^{N} (c_n - \langle c_n \rangle_n)^2, \tag{1.3}$$

where c_n is a given dissimilarity measure $D(.,.)$ between local ($\widehat{S}(t_n, f)$) and global spectra ($\langle \widehat{S}(t_n, f) \rangle_n$) computed for each time point: $c_n = D(\widehat{S}(t_n, f), \langle \widehat{S}(t_n, f) \rangle_n)$, and $\langle c_n \rangle_n$ is computed by averaging c_n over all $n = 1, ..., N$. In practice, even for stationary sitatuations, the equality of (1.3) with zero for all cases will not exist. Fluctuation will always exist, and we should rather expect a statistical similarity between local and global spectra under the null hypothesis of stationarity. The stationarity test is performed by gathering a collection of (1.3) in stationary situations (by making use of *surrogates* resampling), and deriving its distribution under the null hypothesis. This distribution allows for the characterization of a given threshold γ, above which the null hypothesis is rejected. Hence, the test is based on the statistics $\Theta_1 = L(c_n, \langle c_n \rangle_n)$, and could be written as a one-sided test [44] as follows:

$$\begin{cases} 1 & \text{if } \Theta_1 > \gamma, \text{ "nonstationary"}, \\ 0 & \text{if } \Theta_1 \leq \gamma, \text{ "stationary"}. \end{cases} \tag{1.4}$$

1.4.1.2 Advantages of the method and contributions

This *nonparametric* method was originally proposed in [41], and uses a data-driven resampling technique (surrogates) for testing stationarity relatively to a *global* horizon of observation. The original surrogate method presents a good performance in detecting second-order nonstationarities, modulations and sudden changes. In Chapter 2, we present a number of contributions to the original framework:

- we perform an empirical study using many distances $D(.,.)$ of different classes, so as to have an idea of the appropriateness of each distance in testing different types of nonstationarity,

- we propose a modification of the method that allows for the detection of first-order nonstationarities, and improves the detection of slowly-varying nonstationarities in certain cases,

- we develop a measure of the robustness of the test, which is useful when only a single realization of the process is available.

1.4.2 Second approach: Testing for trend-based nonstationarity

1.4.2.1 A brief description

The stationarity test developed and described in Chapter 3 is a new technique suited to detect trends and slow nonstationary evolutions. The basic idea of the method is to identify the presence of a trend and/or an evolution of the local energy of the signal. We thus test for trends in the time marginal $y(t)$ obtained by summing the time-varying spectrum $S(t,f)$ over all frequencies: $y(t) = \sum_f S(t,f)$. In the framework we are interested in, we cannot specify any *a priori* model to $y(t)$. Thus, we propose to use the empirical mode decomposition (EMD) to estimate trends in $y(t)$ (according to the EMD interpretation). We consider that *stationarity* should correspond to the case of a *trendless* time marginal. On the other hand, the cases where a significant trend is detected in the time marginal should correspond to *nonstationarity*. The *distance from the stationary case* could be evaluated by using the following measure:

$$\theta_{\text{TI}} = \frac{\mathbb{V}\text{ar}\{y(t)\}}{\mathbb{V}\text{ar}\{y(t) - c(t)\}}, \tag{1.5}$$

which evaluates the trend importance in $y(t)$. In (1.5), $c(t)$ is the trend component. We should have $\theta_{\text{TI}} \geq 1$ for the case of a trended time marginal (nonstationarity), and $\theta_{\text{TI}} \approx 1$ for the case of a trendless time marginal (stationarity). The stationarity test is performed by gathering a collection of (1.5) under the null hypothesis (by making use of block bootstrap technique), and obtaining its distribution in order to characterize a given threshold $\mathcal{T}$, above which the null hypothesis of stationarity is rejected. The test itself is based on the statistics θ_{TI}, and could be written as a one-sided test as follows:

$$\begin{cases} 1 & \text{if } \theta_{\text{TI}} > \mathcal{T}, \text{ "nonstationary"}, \\ 0 & \text{if } \theta_{\text{TI}} \leq \mathcal{T}, \text{ "stationary"}. \end{cases} \tag{1.6}$$

1.4.2.2 Advantages of the new stationarity test

The original method developed in Chapter 3 presents a good performance in detecting trends and slowly nonstationary evolutions. In comparison to other approaches described in the literature devoted to detect such type of nonstationarity, the proposed method has the advantage of being nonparametric and data-driven. Moreover, this approach exhibits an overall better performance in detecting nonstationarities of first and second-order. Also, it is faster than other TF approaches (as the one of Chapter 2), as the computation of many TF representation is not required (the time marginal is

estimated directly by numerical convolution with the chosen window). Finally, the proposed method works even in the case of short time series.

1.5 Change point detection

The common procedures for change point estimation are based on statistical hypothesis testing [5]. Usually, the same hypothesis described in the first paragraph of Section 1.4 is considered, but it is assumed that a change does take place in the sample of observation. The problem of estimating the change point could be written as follows: let $\{x_t, t = 1, ..., T\}$ be a time series with conditional density $p_\theta(x_t|x_{t-1}, ..., x_1)$ with parameter $\theta = \theta_0$ before the unknown change time t_0, and parameter $\theta = \theta_1$ after t_0. The problem consists in estimating the change point t_0 from $\{x_t, t = 1, ..., T\}$ with the maximum possible accuracy [25]. Thus, the common statistical approach for change point detection can be seen as a typical estimation problem for a discrete parameter. Hence, the accuracy of the estimation is dictated by the probability that the estimate belongs to a given confidence interval, or by the bias and variance of the change point estimator.

The approach described above assumes the presence of at most one change point [25, 45]. However, if the time series is long, or other change points are plausible, such strategy can lead to erroneous results, as the effects of the first change point are heavily biased when other unaccounted change points occur [45]. In this thesis, we have not considered the traditional statistical approach for estimating the change points. The framework proposed in Chapter 3 allows for the detection of multiple change points, and aims at analyzing those signals in which stationarity have already been rejected by the hypothesis tests of Chapter 2 and Chapter 3. Also, we consider a more general interpretation of change point, which is not restricted to the one defining the time instant where a change in the conditional density occurs. Instead, we consider change points as *"time discontinuities in a time series that can be induced from changes in observation locations, equipment, measurement techniques, environmental changes, and so on"* [45].

1.5.1 Detection by using the robust singular spectrum transform

1.5.1.1 A brief description

The framework for change point detection of Chapter 4 is based on the robust singular spectrum transform (RSST). The RSST is an improved version of the singular spectrum transform (SST), which consists of using principal component analysis (PCA) to measure the anomaly between past and future patterns of the signal around a given time instant

t. This anomaly measure is called change point score and is given by $z(t)$. The RSST has two free parameters (w and n), which can be interpreted as two windows of analysis. In Chapter 4, we propose to represent the output of the RSST in the space spanned by its two parameters and the time, as different change patterns can be captured if we let w and n vary. The result will be a modified transform $\mathcal{T} : x(t) \rightarrow z(t, w, n)$, where the anomaly measure between the past and the future is represented in a space spanned by the t (time), w and n (the two windows of analysis). For estimating the change points, we first propose a procedure for filtering significant values $z(t, w, n)$ (significant change patterns). Then, we compute the conditional entropy (conditioned on the past values) of the observed change scores in time. The conditional entropy is chosen because we consider that the uncertainty associated with the change patterns captured by different windows w and n should increase around the change point.

1.5.1.2 Advantages of the proposed framework

This framework is nonparametric, data-driven and has shown to be sensitive to both changes in the mean and in the variance. Also, the proposed framework does not require the observations of the data to be independent. Thus, the proposed method is very flexible and suited to real world applications. Another advantage is that the method offers a richer representational space for the signal, allowing us to visualize the change patterns as function of time and window of analysis.

Chapter 2

New perspectives on testing stationarity with surrogates

Contents

As the first part of the Thesis, Chapter 2 presents various contributions to the surrogate approach for testing stationarity. It starts in Section 2.1 with an introduction on the stationarity test, explaining which aspects of the original framework this Thesis has focused on, and briefly describing the interpretation of stationarity associated with the framework. In Section 2.2, the stationarity test itself is present. Then, in Section 2.4, 2.5 and 2.7, the contributions to the original framework are proposed, which not only improve the detection of nonstationarity, but also allow for a better understanding of the surrogate-based approach.

2.1 Introduction

As pointed out in Section 1.2, real world signals are predominantly nonstationary, which require the introduction of time as explicit parameter for the analysis. Also, in real world applications we have to deal with complex processes with many spectral components, which are often given without any *a priori* information. In such situations, TF techniques are natural tools, which consider the temporal dependence of the spectral content of the process [4].

Given its applicability, a variety of options in nonstationary signal analysis are offered by TF methods. Different from the PSD of WSS signals, there are several ways to represent the time-varying spectrum of a nonstationary process. The advantages and disadvantages of different TF representations are discussed in [4, 46]. Among the emerging alternatives in the literature [47, 38, 48], one could easily categorize parametric [33, 32, 20] and nonparametric approaches [49, 50, 51] for analyzing nonstationary signals. In the latter category, the estimates of the *Wigner-Ville Spectrum* (WVS) for representing the time-varying spectrum are often considered [52, 53, 54]. In the past, the Wigner-Ville Spectrum has already been used for the analysis of nonstationary processes [2, 42]. Recently, a novel stationarity test based on WVS estimates was published [41]. The test is performed by gathering a collection of surrogates as stationary versions of the signal [55, 56]. Then the surrogates are used to learn the statistics of a null hypothesis of stationarity. The key point of this method is to use the local and global spectral properties of the time-varying spectrum for identifying the stationarity. As these local and global characteristics will not be completely identical, even for stationary situations, a hypothesis test is designed to specify whether or not the observed fluctuations are due to a nonstationarity. The stationarity test is described in details in Section 2.2.

The method presented in [41] considers that stationarity should not be regarded as

an absolute property. Instead, a more empirical interpretation of stationarity is adopted for taking into account some practical considerations often faced in real world applications. For instance, it is assumed that stationarity not only refers to stochastic processes with time-invariant statistics, but also to deterministic signals with spectral properties that are time-invariant. Also, since stationarity is a property being evaluated over a finite time interval, it should be given relatively to a scale of time or observation. Moreover, we should be able to test for stationarity given only one observable realization of the process, as it is generally the case in many applications. This generalized notion of stationarity has been referred to as *operational stationarity* , which can be more conveniently tested in practice.

2.1.1 Detailing the contributions of this chapter

The original framework leaves room for further improvements in various aspects. As it will be seen in the following Section, the procedure contains many different steps, so there are many points in the original methodology deserving a deeper investigation. It should be remarked that [41] proposes two approaches for testing stationarity with surrogates: i) the use of suitably chosen distances between local and global spectra for measuring the spectral fluctuation, ii) the implementation of a one-class classifier that uses the surrogates as a learning set for stationarity. All the contributions presented in this Chapter are made considering only the first case. Although putting the stationarity test in the context of learning theory is an interesting alternative, we have not led this work towards that direction. Instead, we have carried out a deeper analysis on the use of a proper distance for comparing local and global time-frequency features. In this regard, we have proposed three contributions to the original method:

- We study in detail the effects of using different distances. The chosen ones are described in Section 2.3, and the results of using the selected distances are shown in Section 2.4. This study highlights how difficult it is for the original framework to detect first-order nonstationarities, and nonstationarities varying slowly as a trend.

- We propose a modification that improves the performance of the test in certain cases, and increases the detection rate of first-order nonstationarities. In Section 2.5, we start by defining what we expect from the modified method in an ideal case. Then, we establish a condition for the parameters of the modified test to attend. In Section 2.6, we propose a procedure for changing the test that would attend the defined condition. In Appendix A.1, we demonstrate why this proposed procedure works.

- Finally, we develop a measure to evaluate the robustness of the test due to the different results that may appear by applying the method sequentially. This contribution is described in Section 2.7.

2.2 Testing stationarity with surrogates

For developing a framework in TF domain, while taking into account the considerations on the operational stationarity presented in Section 2.1, one should design a method that works for one realization of finite duration of a given signal. As far as the representation is concerned, it is common to consider the general class of estimators belonging to the Cohen's class. All distributions in this class can be interpreted as estimators for the WVS [4]. The WVS of a given signal $\{x(t), t \in \mathbb{R}\}$ is expressed as:

$$\mathbf{W}_x(t,f) = \int_{-\infty}^{\infty} \mathbb{E}\left\{x\left(t+\frac{\tau}{2}\right)x^*\left(t-\frac{\tau}{2}\right)\right\} e^{-j2\pi\tau f} d\tau \tag{2.1}$$

where $\mathbb{E}\{\cdot\}$, t and f stand for the expectation operator, time and frequency, respectively. The WVS possesses many good properties that are extensively explained in the literature [4, 46, 2, 57]. Thus, estimates of the WVS are commonly used for representing the time-varying spectrum of nonstationary signals. For estimating the WVS, one can make use of the spectrogram (a member of Cohen's class), or its modified (improved) version, the multitaper spectrogram:

$$S_K(t,f) = \frac{1}{K}\sum_{k=0}^{K-1} S^{h_k}(t,f), \tag{2.2}$$

which is obtained by averaging K different spectrograms $(S^{h_k}(t,f))$ at the time instants $\{t_n, n = 1, ..., N\}$. The spectrograms are the squared magnitude of the Short-Time Fourier Transform (STFT) of the signal, computed by using K different window functions $h_k(t)$:

$$S^{h_k}(t,f) = \left| \int x(s) h_k(s-t) e^{-j2\pi f s} ds \right|^2. \tag{2.3}$$

The multitaper spectrogram reduces the variance of the estimation by projecting the observation on a family of orthonormal basis functions $\{h_k(t), k \in \mathbb{N}\}$ [49, 58]. In [41], the *Hermite functions* were chosen as such basis. These functions are not only orthonormal, but also maximally concentrated in TF domains with elliptic symmetry. By construction, the length n_h of the windows $h_k(t)$ should contain an odd number of samples. For computing the time positions t_n, the spacing $\Delta_t = t_{n+1} - t_n$, and the number of points in frequency (N_FFT) used for the multitaper spectrograms, we need to specify the parameter $n_{\mathrm{h}0}$, which is one degree of freedom of this method. The

expression of n_{h0} associates the length of the signal (T) with an adjustable fraction of the chosen window (n_h) [41]:

$$n_h = 2\text{nint}\left(\frac{Tn_{h0}}{2}\right) - 1, \tag{2.4}$$

where the function nint(x) returns the nearest integer to x. Then, for obtaining N_{FFT} and Δ_t, we use, respectively, the following expressions:

$$N_{FFT} = 2^{\lceil \log_2 n_h \rceil}, \tag{2.5}$$

$$\Delta_t = \left\lfloor \frac{n_h + 1}{8} \right\rfloor, \tag{2.6}$$

where the functions $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ are the floor and ceiling functions, respectively.

The rationale behind this technique is the following: for a given signal $x(t)$, testing for stationarity in the operational sense amounts for testing the temporal invariance of $S_K(t, f)$ over the chosen observation interval T. Quantitatively, it means that the local spectra $S(t_n, f)$ at every time instant $\{t_n, n = 1, ..., N\}$ are statistically equal to the global average spectrum

$$\langle S(t_n, f) \rangle_n := \frac{1}{N} \sum_{n=1}^{N} S(t_n, f). \tag{2.7}$$

In practical applications we will always observe fluctuations in the local spectra. The originality of the method consists in giving significance to these fluctuations by constructing stationarized references of the signal using only the available data. The stationary signals are computed by means of the surrogate technique, a resampling method introduced originally in nonlinear physics [55, 56]. Each surrogate $s(t)$ has the same global PSD as the original signal while being stationary.

A comparison between the stationary surrogates and the original signal is performed. The idea is to elaborate a hypothesis test to check if fluctuations of the local time-varying spectra around the global average spectrum are consistent with what is expected under stationarity. If it is not, we say that the variation between local and global spectra is probably due to the nonstationarity of the original signal.

Nonstationary processes have their spectral content spread in time differently from stationary ones. Hence, for an identical marginal spectrum over the same observation interval, we may expect some time-organized structure in nonstationary signals that are not present in stationary ones. Thus, we obtain the surrogates by destroying this organized phase structure controlling the supposed nonstationarity [41]. More precisely, for a given signal observed in discrete time $\{x[n], n = 1, ..., N_T\}$, the corresponding surro-

gate $s[n]$ is obtained by first computing the Fourier transform of the observed signal: $X[k] = A[k]e^{-j\Phi[k]}$, and then by replacing its phase $\Phi[k]$ by the random sequence $\Psi[k]$ which is uniformly distributed over $[-\pi, \pi]$. By applying the inverse Fourier transform we can obtain as many stationary surrogates as phase randomizations are performed:

$$x[n] = \frac{1}{N_{\mathrm{T}}} \sum_k X[k] e^{j2\pi nk/N_{\mathrm{T}}} \rightarrow s[n] = \frac{1}{N_{\mathrm{T}}} \sum_k A[k] e^{j\Psi[k]} e^{j2\pi nk/N_{\mathrm{T}}}. \tag{2.8}$$

In [41], it is shown that by adopting $\Psi[k]$ as phase sequence, the covariance function of the surrogate $s[n]$ reduces to:

$$\mathbb{C}\mathrm{ov}(s[n], s[m]) = \frac{1}{N_{\mathrm{T}}^2} \sum_k \mathbb{E}\{A^2[k]\} e^{j2\pi(n-m)k/N_{\mathrm{T}}}, \tag{2.9}$$

and the stationarity of the surrogate $s[n]$ can be verified since its covariance is function of $n - m$ only. In order to perform the hypothesis test we create a collection of I surrogates $\{s_i(t), i = 1, ..., I\}$, and then we compute a distance between local and global spectra for each surrogate. The result will be a collection of distance vectors being function of both time indices and randomizations:

$$\{c_n^{(s_i)} := D\left[S_{s_i,K}(t_n, f), \langle S_{s_i,K}(t_n, f)\rangle_n\right], n = 1, ..., N, i = 1, ..., I\}. \tag{2.10}$$

In (2.10), the distance $D(.,.)$ is a dissimilarity measure in frequency, being one of the principal aspects of the method under consideration in this work. The null hypothesis of stationarity will be characterized by the statistical distribution of the estimate of (2.10) in time. For the signal itself, the distances between spectra are given in the following vector:

$$\{c_n^{(x)} := D\left[S_K(t_n, f), \langle S_K(t_n, f)\rangle_n\right], n = 1, \cdots, N\}, \tag{2.11}$$

corresponding to the same time instants where the multitaper spectrograms are being computed. Now, the variance of the distance vector in (2.11) is taken as a measure of the spectral fluctuation of the time-varying spectrum of the signal itself:

$$\Theta_1 = \mathbb{V}\mathrm{ar}\left\{c_n^{(x)}\right\}, \tag{2.12}$$

where in practice, (2.12) is estimated by:

$$\widehat{\Theta}_1 = \frac{1}{N-1} \sum_{n=1}^{N} \left[c_n^{(x)} - \widehat{\mu}_{c_x}\right]^2, \tag{2.13}$$

where $\widehat{\mu}_{c_x}$ stands for the sample mean of $c_n^{(x)}$. Putting it in the form of a one-sided test, this variance will be the test statistic.

The hypothesis test is given by:

$$d(x) = \begin{cases} 1 & \text{if } \Theta_1 > \gamma, \text{ "nonstationary"}, \\ 0 & \text{if } \Theta_1 \leq \gamma, \text{ "stationary"}. \end{cases} \tag{2.14}$$

The threshold γ above which the null hypothesis of stationarity is rejected is obtained from (2.10), the collection of distances of the surrogate set. To compute γ, the variance of each distance vector $\{c_n^{(s_i)}, i = 1, ..., I\}$ is calculated, leading to a vector of I variances:

$$\left\{ \Theta_0(i) = \mathbb{V}\text{ar}\{c_n^{(s_i)}\}, i = 1, ..., I \right\}, \tag{2.15}$$

In practice, the individual variances $\mathbb{V}\text{ar}\{c_n^{(s_i)}\}$ in (2.15) are estimated by

$$\widehat{\sigma}^2_{s_i} = \frac{1}{N-1} \sum_{n=1}^{N} \left[c_n^{(s_i)} - \widehat{\mu}_{c_{s_i}} \right]^2, i = 1, ..., I,$$

where $\widehat{\mu}_{c_{s_i}}$ is the sample mean of the i^{th} distance vector in the collection $\{c_n^{(s_i)}, i = 1, ..., I\}$. We do not use the crude histograms from (2.15) for approximating the distribution under the null hypothesis. Instead, it is proposed to model the distribution of $\Theta_0(i)$ by a gamma pdf, where its two parameters are estimated by maximum likelihood . An empirical study carried out by [44] has shown that the distribution of $\Theta_0(i)$ can be fairly well approximated by a gamma pdf. Assuming the gamma model holds, a threshold γ is derived above which the null hypothesis of stationarity is rejected [41].

Finally, an index of nonstationarity (INS) is proposed. It is a function of the ratio between the test statistic Θ_1 and the mean value of $\Theta_0(i)$, obtained from the surrogate set. If the signal happens to be stationary, INS is expected to take a value close to unity. On the other hand, the more nonstationary the signal, the larger the index [41].

$$\text{INS} = \sqrt{\frac{\Theta_1}{\langle \Theta_0(i) \rangle_i}} \tag{2.16}$$

An application of the stationarity test is illustrated in Fig. 2.1, where we have considered a false alarm rate of 5%. The test was applied to two time series with length $T = 300$. The first one is a zero-mean stationary Gaussian process with $\sigma^2 = 1$ (Fig. 2.1(a)). The second one is a nonstationary Gaussian process with $\mu = 0$ and an ever-growing variance increasing linearly in time from $t = 1$ to $t = 300$ (Fig. 2.1(c)).

Notice that the value of Θ_1 (the test statistic estimated from the original signal), is in the middle of the distribution in the stationary case (Fig. 2.1(b)). On the other hand, for the nonstationary signal, the value of Θ_1 appears clearly as an outlier (Fig. 2.1(d)). Notice that the INS value is significantly larger for the nonstationary signal. Also, one can verify the gamma fit applied to the histograms of Θ_0. As pointed in [41], we only need a reduced number of surrogates ($I \approx 50$) for obtaining a reasonable approximation with the gamma model.

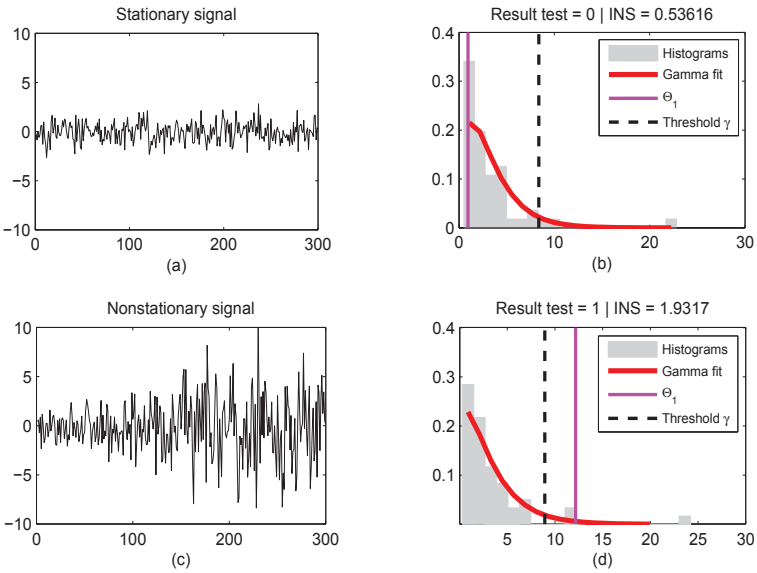

Figure 2.1: An illustration of the proposed stationarty test. (a) Stationary Gaussian process. (b) Results of the test: the test statistic Θ_1 is in the middle of the distribution. (c) Gaussian process with a nonstationary variance. (d) Results of the test: the test statistic Θ_1 is beyond the threshold γ (which corresponds to 95%).

Having presented the stationarity test of [41], it can be seen that the framework contains many different steps (*i.e.* surrogate resampling, computation of a distance between local and global spectra, gamma modeling), which call for a deeper investigation. In this regard, in the next Section, we investigate the effects of using different distances. This effort is the first contribution of this Thesis, and aims at finding an appropriate distance for a given application.

2.3 The influence of choosing an adequate distance

The dissimilarity measure between the local and the global spectra is a significant aspect of the proposed approach. In fact, the fluctuation of the distances is the key point to reject or accept the hypothesis of stationarity. Consequently, for the same signal and test parameters, different results can be obtained by using different distances, be they

completely different or similar in nature [6].

Choosing the most appropriate distance for a given test application is not an easy task. Insofar, one would consider many aspects to support the choice of a given distance. Without entering into many details, we could roughly list them as: the features from the signal being tested (its length, spectral content, etc.), which kind of nonstationarity we are aiming at detecting (exclusively first or second-order, or up to the second-order).

2.3.1 A robust distance

One may suggest to always use a dissimilarity measure that leads to a good performance. Unfortunately, the notion of "good performance" itself might be dubious at certain occasions. For example, by applying the test to a nonsynthetic signal with no *a priori* information about it or its generative system, we cannot be sure if the signal is being correctly classified as being stationary/nonstationary. Nevertheless, the result of the test may vary at every application due to the randomness of the surrogate set, even if we are testing the same realization of the signal. Considering all the drawbacks, we could at least search for a distance that leads to less fluctuations in the results, be the test signal synthetic or not. So, in the sense of test consistency, we propose here to call it a *robust* distance. A deeper analysis on robust distances is carried out in Section 2.7, where different measures for evaluating the robustness are presented. By now, it should be remarked that repeating the test with different distances and comparing the results is highly advisable in any situation.

2.3.2 The different classes of distances

Dissimilarity measures of different types have been commonly used in signal processing for solving problems such as pattern recognition, segmentation, classification, detection, etc. Hence, there is a variety of distances available in the literature [6]. The various distances could be classified in different ways, the classical manner being based on probability vs. frequency-based or symmetric vs. nonsymmetric classes [6]. In this section we briefly present the different classes of distances that are considered in this work.

2.3.2.1 Frequency and probability-based distances

Probability-based distances (often referred to as divergences) quantify the dissimilarity between two statistical objects, as probability distributions, for instance. Although the distances are essentially computed in the frequency domain, dissimilarity measures of

probability nature can be conceived due to the probabilistic interpretation of TF distributions [4]. As the two spectra to be compared are always positive, they only need to be normalized to the unity for characterizing a probability distribution. The remoteness of two probability distributions is commonly measured by using *divergence* functions [59]. The divergence in the context of statistics establishes a notion of distance different from the classical one, as the divergence not necessarily satisfy metric requirements. There exist different classes (or families) of divergences, the most common one being the so-called *f-divergence* class [60]. In this work, all the probability-based distances belong to the family of *f-divergences* [6].

Frequency-based distances are computed directly from the spectra, and consequently their understanding is more intuitive. However, even though spectral distances are easily defined in the frequency domain, they are most of the time computed numerically, without reference to this domain [6]. Spectral distances are aimed at comparing distributions in both shape and level. It is an advantage, since in the application considered here, the spectra from where the distances are being computed share the same band.

2.3.2.2 Symmetric and nonsymmetric distances

It should be remarked that *distance* in the context being used in this work, refers to a measure of how far away from each other two spectra are (with the inherent probabilistic interpretation). So, a given distance may not satisfy the necessary conditions to be a metric (*i.e.* positive-definiteness conditions, symmetry condition and triangle inequality [61]). In fact, many statistical distances are not metric, since they do not attend one or more requisites to be classified as so. One of the metric properties that are not presented by all distances is the symmetry. Here, we used symmetric and nonsymmetric distances. However, it should be noted that a nonsymmetric distance between two spectra H and G could be easily symmetrized as follows [6]:

$$d^{(q)}(H,G) = \frac{1}{2}\left\{d(H,G)^q + d(G,H)^q\right\}^{1/q}, \text{ with } q \geq 1. \tag{2.17}$$

2.3.2.3 Mixed domain distances

Finally, a mixture of distances could also be performed, combining symmetric, nonsymmetric, frequency and probability-based ones, so as to take advantages of the different classes. Usually, one sets a trade-off parameter in order to control the significance given to a particular distance.

Table 2.1: Name, expressions and classes of the different distances used in this study

PROBABILITY-BASED DISTANCES

Distance	Expression
Kullback-Leibler (symmetric)	$D_{\mathrm{KL}}(\widetilde{G},\widetilde{H}) = \sum_f \left[\widetilde{G}(f) - \widetilde{H}(f)\right] \log\frac{\widetilde{G}(f)}{\widetilde{H}(f)}$
Kolmogorov (symmetric)	$D_{\mathrm{KM}}(\widetilde{G},\widetilde{H}) = \sum_f \left\lvert \widetilde{H}(f) - \widetilde{G}(f) \right\rvert$
Matusita (symmetric)	$D_{\mathrm{MT}}(\widetilde{G},\widetilde{H}) = \sum_f \left[\sqrt{\widetilde{H}(f)} - \sqrt{\widetilde{G}(f)}\right]^2$

FREQUENCY-BASED DISTANCES

Distance	Expression
Itakura-Saito (nonsymmetric)	$D_{\mathrm{IS}}(G,H) = \sum_f \left[\frac{G(f)}{H(f)} - \log\frac{G(f)}{H(f)} - 1\right]$
Log-Spectral (symmetric)	$D_{\mathrm{LS}}(G,H) = \sum_f \left\lvert \log\frac{G(f)}{H(f)} \right\rvert$
Diffusion (symmetric)	$D_{DF}(G,H) = \sum_{l=0}^{L} \lvert\lvert d_l(f) \rvert\rvert$, where $d_l(f) = \begin{cases} \lvert G(f) - H(f) \rvert \text{ for } l = 0, \\ d_{l-1}(f) * \left[\frac{e^{-f^2/2\sigma^2}}{(2\pi)^{1/2} f}\right] \text{ for } l = 1, ..., L. \end{cases}$
Itakura-Saito (symmetric)	$D_{\mathrm{SIS}}(G,H) = \frac{1}{2}\sum_f \left[\frac{G(f)}{H(f)} + \frac{H(f)}{G(f)} - 2\right]$

MIXED DOMAIN DISTANCES

Distance	Expression
Combined (symmetric)	$D_{\mathrm{CB}}(G,H) = D_{\mathrm{KL}}(\widetilde{G},\widetilde{H})\,[1 + D_{\mathrm{LS}}(G,H)]$

2.3.3 Selected distances

There are many distances in the literature belonging to the previous classes. In this work, we have selected a total of eight distances, covering a variety of classes (three probability and four frequency-based distances, one symmetrized and one combined distance). The chosen dissimilarity measures, in comparison with others of the same class, present a better performance. They are presented in Table 2.1. The expressions include two positive spectra $H(f)$ and $G(f)$, and their modified versions $\widetilde{H}(f)$ and $\widetilde{G}(f)$, which were normalized to unity in order to be in accordance with those distances of probability nature [6].

The distances of probability nature in Table 2.1 belong to the *f-divergence* class; the well-known Kullback-Leibler divergence (D_{KL}), the Kolmogorov distance (D_{KM}), which is probably the simplest one, and a generalization of the Kolmogorov case, which is the Matusita distance (D_{MT}). Usually, the Kullback-Leibler divergence is used as a nonsymmetric measure, whereas here D_{KL} is symmetric, as originally adopted in [41].

The chosen distances of frequency nature were the Itakura-Saito (D_{IS}), Log-Spectral

(D_{LS}), Diffusion (D_{DF}) and the symmetrized version of the Itakura-Saito (D_{SIS}). The Itakura-Saito distance measures the percentual difference between a spectrum and its approximation. It is a nonsymmetric dissimilarity measure, however we can easily symmetrize it by means of (2.17) (with $q = 1$), leading to the symmetrized Itakura-Saito distance. The Log-Spectral deviation is defined by the L_q norm of the difference between the logarithms of the spectra, and this is probably the oldest one used in speech processing [6]. The diffusion distance treats the difference between $G(f)$ and $H(f)$ as an isolated temperature field and considers the diffusion process on that field. For discretizing the continuous diffusion process, we use a Gaussian pyramid with $l = 0, ..., L$ different layers $[d_0(f), ..., d_L(f)]$, and a constant standard deviation (σ). The diffusion distance is then considered as the sum of norms over all pyramid layers [62].

Finally, the chosen combined distance D_{CB} stands for the dissimilarity measure originally used in [41]. Performing a mixture of distances is considered as a way to take advantage of the different characteristics of distances belonging to distinct classes. The two distances combined are the Kullback-Leibler and Log-Spectral ones.

2.4 Testing with different distances

After presenting the selected distances in Table 2.1, we show the results of the stationarity test described in Section 2.2 by using different measures. Various nonstationary signals were created to assess the performance of the test. We considered nonstationary time series corresponding to various situations: nonstationarity of the mean, of the variance, of both the mean and the variance, following the *Abrupt-change* and the *Onset-of-trend* model. These are special cases of the more general *smooth-change* model proposed by Lombard [18]. The temporal patterns represented by the Lombard's model is similar to what is observed in many areas of environmental science, such as hydrology and climatology. More specifically, given the time series $\mathcal{X} = (x_1, ..., x_T)$ of length T, with observations taking equal time intervals, where $F_i(x) = \mathrm{P}(x_i \leq x)$. If one is interested in changes of a given parameter $\theta_i = \Theta(F_i)$, $i \in [1, T]$, where Θ is a functional of a particular interest, *e.g.* the mean or the variance, the series is said to follow a *smooth-change* model if

$$\theta_i = \begin{cases} \xi_1, & i \in \{1, ..., \kappa_1\}, \\ \xi_1 + \dfrac{(i - \kappa_1)(\xi_2 - \xi_1)}{(\kappa_2 - \kappa_1)}, & i \in \{\kappa_1 + 1, ..., \kappa_2 - 1\}, \\ \xi_2, & i \in \{\kappa_2, ..., T\}. \end{cases} \tag{2.18}$$

where ξ_1 and ξ_2 are the parameter values *before* and *after* change, respectively [10]. The *abrupt-change* model is obtained when $\kappa_1 = \kappa$ and $\kappa_2 = \kappa_1 + 1$, where the κ represents the change point:

$$\theta_i = \begin{cases} \xi_1, & i \in \{1, ..., \kappa\}, \\ \xi_2, & i \in \{\kappa + 1, ..., T\}. \end{cases} \tag{2.19}$$

On the other hand, the *onset-of-trend* model represents a gradual transition from the initial parameter value ξ_1 to the final value ξ_2. This refers to the special case when one sets $\kappa_2 = T$:

$$\theta_i = \begin{cases} \xi_1, & i \in \{1, ..., \kappa\}, \\ \xi_1 + \dfrac{(i-\kappa)(\xi_2 - \xi_1)}{(T-\kappa)}, & i \in \{\kappa + 1, ..., T-1\}. \\ \xi_2, & i = T, \end{cases} \tag{2.20}$$

The synthetic signals have a Gaussian distribution and two different lengths $T = 300$ and $T = 1050$. The short signals were chosen in order to be in accordance with several areas of applications like biomedicine, climatology and hydrology. Moreover, by testing short series we can evaluate the test performance under unfavorable situations. Three different sets of signals were generated by considering the *Abrupt-change* and the *Onset-of-trend* models. The first one consists of time series having a *fixed variance* ($\sigma^2 = 1$) and a *varying mean* ranging from $\xi_1 = 0$ to $\xi_2 = 4$. The second set is also formed by Gaussian series, but now with a *fixed mean* ($\mu = 0$) and a *varying variance* taking values from $\xi_1 = 1$ to $\xi_2 = 4$[1]. The third one contains signals having both a varying mean *and* variance, taking the same parameter values of the first and second sets. Realizations of the synthetic signals with $T = 300$ and $T = 1050$ are shown in Fig. 2.2(a) and Fig. 2.2(b), respectively.

The stationarity test has been applied to 1000 realizations of the signals presented in Fig. 2.2 and also to a stationary AR(1) process. For the simulation, we have set $n_{h0} = 0.1$, as this value has given a good classification accuracy. The results are shown in Table 2.2 as percentage (%) of "nonstationary" outcomes and arithmetic averages ($\langle \text{INS} \rangle$) of INS over all realizations. As it can be seen, the stationarity test could not detect properly those signals whose nonstationarities follow the *Onset-of-trend* model, *i.e.* signals with statistical properties evolving gradually as an ever-growing "trend". On the other hand, the nonstationary processes following the *Abrupt-change* model could be detected more easily. Moreover, note that the performances of the frequency-based

[1]These values were chosen empirically. We observed that, in general, the choice of parameter values ranging from $\xi_1 = 0$ to $\xi_2 = 4$ gives a good compromise between keeping the detectability of the test for all the adopted models, while having small variations in the series.

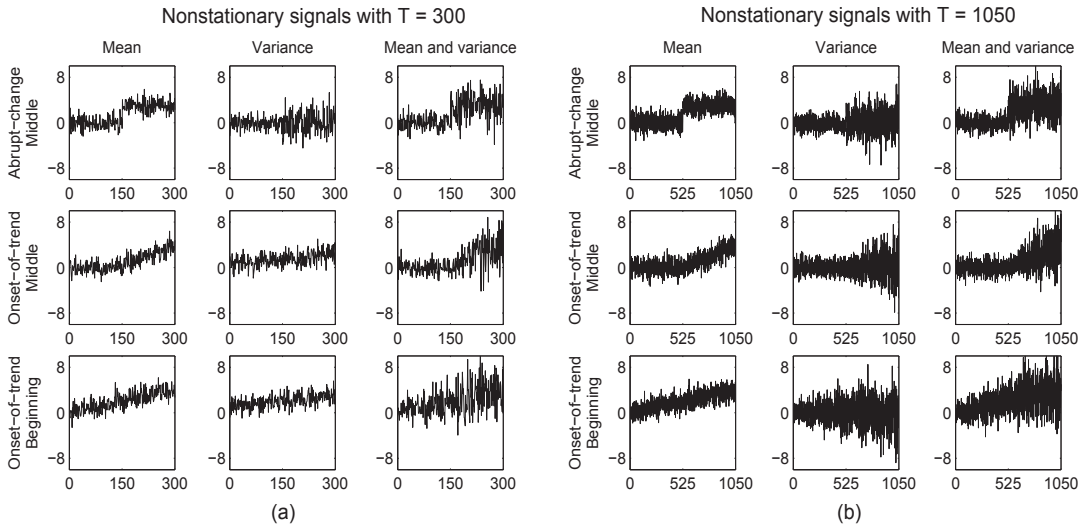

Figure 2.2: Synthetic signals following the *Abrupt-change* and the *Onset-of-trend* models. (a) Signals with length $T = 300$. (b) Signals with length $T = 1050$.

distances were notably better, specially in detecting those signals following the *Onset-of-trend* model. Most importantly, the overall performance in detecting a nonstationary mean was considerably poor. In the next Section, we propose a modification in the methodology of the stationarity test for improving its performance. This modification allows for a better detection of first-order evolutions and nonstationarities evolving as a trend.

2.5 Modifying the method

As it can be seen in Table 2.2, the test could not detect some nonstationarities even by using different distances. Not surprisingly, first-order nonstationarities were the most difficult ones to be detected. However, even second-order nonstationarities could not be detected properly by using some distances. Clearly, the performance of the test has to be improved, and it can be achieved by modifying the method according to the procedure proposed in this Section.

An interesting way to start with is by defining what we expect from the stationarity test after modifying it. Consider that the test has been accepting the null hypothesis for a given signal, and we want to increase the sensitivity to nonstationarity. Ideally, we search for a procedure that improves the method in a sense that:

- If the process is *nonstationary*, it will be more likely for the test to reject the null hypothesis of stationarity,

Table 2.2: Results of applying the stationarity test to 1000 realizations of each process shown in Fig. A.4. The results are given as percentage (%) of "nonstationary" outcomes and arithmetic averages (⟨INS⟩) of INS over all realizations.

Nonstationary signals																		
Distance			KL		KM		MS		IS		LS		DF		SIS		CB	
Length			300	1050	300	1050	300	1050	300	1050	300	1050	300	1050	300	1050	300	1050
Varying mean	*Abrupt-change*	%	0	0	0	0	0	0	85.0	88.5	10.0	15.0	97.0	100	10.0	10.0	5.00	0
	Middle	⟨INS⟩	0.50	0.48	0.64	0.63	0.65	0.51	1.61	1.65	1.20	1.10	1.67	1.75	1.01	1.07	0.58	0.55
	Onset-of-trend	%	0	0	0	0	0	0	1.00	2.00	1.00	3.00	0	0	1.00	0	0	0
	Middle	⟨INS⟩	0.58	0.46	0.74	0.61	0.59	0.43	0.85	0.90	0.85	0.93	0.69	0.70	0.66	0.76	0.49	0.52
	Onset-of-trend	%	0	0	0	0	0	0	0	5.00	0	3.00	0	0	3.00	0	1.00	0
	Beginning	⟨INS⟩	0.59	0.47	0.77	0.61	0.64	0.46	0.85	0.93	0.87	0.93	0.70	0.66	0.65	0.69	0.52	0.45
Varying variance	*Abrupt-change*	%	17.0	12.0	10.0	14.0	15.0	15.0	100	100	100	100	100	100	99.0	100	90.0	100
	Middle	⟨INS⟩	1.05	1.02	1.03	1.01	1.03	1.02	4.39	7.85	3.82	7.11	2.04	2.16	12.1	21.9	2.40	3.41
	Onset-of-trend	%	15.5	14.0	6.00	12.0	8.00	10.0	100	100	100	100	100	100	97.5	100	45.0	97.0
	Middle	⟨INS⟩	1.02	1.05	0.98	1.03	0.99	1.01	3.44	5.41	2.22	3.48	2.63	3.88	4.56	6.54	1.62	2.10
	Onset-of-trend	%	5.00	5.00	10.0	12.0	8.00	10.0	100	100	100	100	98.5	100	99.0	100	35.0	83.0
	Beginning	⟨INS⟩	0.95	0.97	1.01	1.02	1.03	0.99	2.40	3.71	2.03	3.42	1.74	2.12	3.49	6.00	1.47	1.75
Varying mean and variance	*Abrupt-change*	%	100	100	99.0	100	100	100	100	100	94.0	100	100	100	100	100	100	100
	Middle	⟨INS⟩	2.20	3.32	1.74	2.40	2.37	3.25	3.70	7.46	3.07	6.62	1.71	1.67	9.21	21.3	5.28	11.3
	Onset-of-trend	%	15.0	30.0	14.0	40.0	22.5	30.0	100	100	100	100	97.0	100	95.5	100	63.0	100
	Middle	⟨INS⟩	1.06	1.14	1.05	1.13	1.15	1.16	3.24	5.00	2.00	3.33	2.14	1.82	3.46	5.56	1.86	2.58
	Onset-of-trend	%	60.0	100	55.0	100	46.0	100	95.5	100	97.0	100	68.0	87.0	92.0	100	78.5	100
	Beginning	⟨INS⟩	1.45	2.23	1.25	1.62	1.34	2.00	2.01	3.44	1.60	3.07	1.55	1.62	2.66	5.22	2.25	4.90
Stationary signals																		
Distance			KL		KM		MS		IS		LS		DF		SIS		CB	
Length			300	1050	300	1050	300	1050	300	1050	300	1050	300	1050	300	1050	300	1050
Type:		%	7.00	4.50	7.00	5.00	7.00	3.00	9.50	5.50	7.00	6.00	7.00	5.00	5.00	5.00	6.50	4.00
AR(1) process		⟨INS⟩	0.95	0.96	0.96	0.91	0.95	0.97	0.98	0.97	0.98	0.98	0.92	0.98	0.93	0.95	0.93	0.94

- If the process is *stationary*, the test will keep accepting the null hypothesis.

This expected profile of the modified method can be put in terms of the gamma model. More specifically, let us denote k, θ, γ, Θ_1 and k', θ', γ', Θ_1' the gamma parameters (shape and scale), threshold and test statistic obtained *before* and *after* modifying the method, respectively. Thus, we will more likely reject the stationarity if, in comparision to Θ_1 and γ, Θ_1' is closer to or greater than γ'. Hence, in this case we should have:

$$F(\Theta_1';k',\theta') - F(\gamma';k',\theta') > F(\Theta_1;k,\theta) - F(\gamma;k,\theta) \tag{2.21}$$

where $F(x;k,\theta)$ is the gamma cdf of a random variable x. Conversely, we will keep accepting the null hypothesis if, after modifying the test we have:

$$F(\Theta_1';k',\theta') - F(\gamma';k',\theta') \leq F(\Theta_1;k,\theta) - F(\gamma;k,\theta). \tag{2.22}$$

Note that for a stationary signal Θ_1 is expected to stand in the middle of the distribution and $F(\Theta_1;k,\theta) - F(\gamma;k,\theta)$ should be *negative*. Since the gamma cdf evaluated at the chosen threshold (say, 0.95) should be necessarily $F(\gamma;k,\theta) = F(\gamma';k',\theta') = 0.95$, we

can rewrite (2.21) and (2.22) as $F(\Theta_1';k',\theta') > F(\Theta_1;k,\theta)$ and $F(\Theta_1';k',\theta') \leq F(\Theta_1;k,\theta)$, respectively. So, now we can rephrase what we expect from the test after modifying it in terms of the gamma distribution:

- If the process is nonstationary, obtain $F(\Theta_1';k',\theta') > F(\Theta_1;k,\theta)$, which means that the test is more susceptible to reject stationarity,
- If the process is stationary, obtain $F(\Theta_1';k',\theta') \leq F(\Theta_1;k,\theta)$ which means that the test keeps accepting the null hypothesis.

Therefore, a methodology for improving the stationarity test can be envisioned by developing a procedure that gives $F(\Theta_1';k',\theta') > F(\Theta_1;k,\theta)$ if the signal is indeed nonstationary, or keeps $F(\Theta_1';k',\theta') \leq F(\Theta_1;k,\theta)$ otherwise. This scheme is illustrated in Fig. 2.3, where the conditions above and the gamma cdf are shown. Now, it is straight-

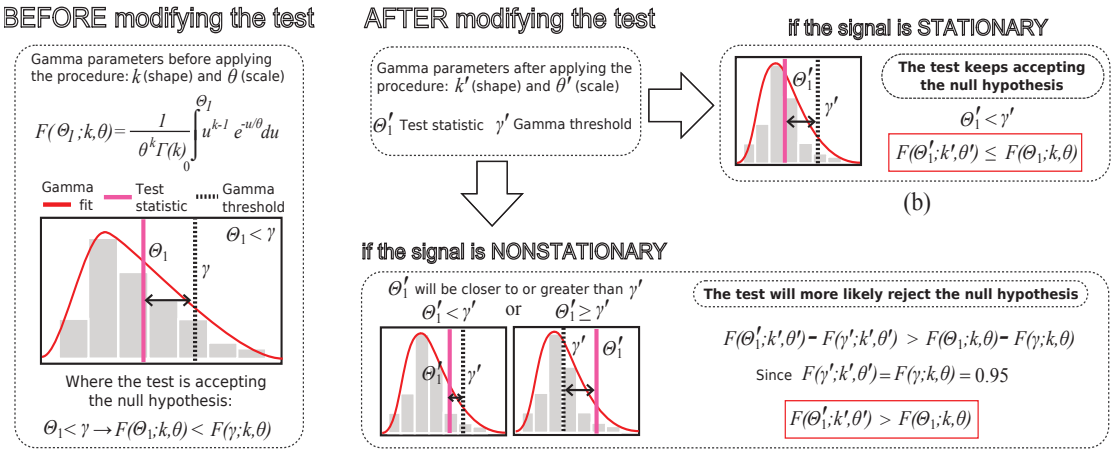

Figure 2.3: Enhancing the stationarity test. (a) Example of the situation before modifying the test and the original gamma parameters. (b) After modifying the test, if the signal is stationary, the test will keep accepting the null hypothesis. (c) After modifying the test, if the signal is nonstationary, the test will more likely reject the null hypothesis.

forward to extend these conditions to the gamma parameters. To do so, we represent the gamma cdf evaluated at Θ_1 as:

$$F(\Theta_1,k,\theta) = \frac{1}{\theta^k\Gamma(k)}\int_0^{\Theta_1} u^{k-1}e^{-\frac{u}{\theta}}du = \frac{\gamma(k,\frac{\Theta_1}{\theta})}{\Gamma(k)}, \tag{2.23}$$

where $\gamma(k,\frac{\Theta_1}{\theta})$ is the lower incomplete gamma function. Then, we can express the condition $F(\Theta_1';k',\theta') > F(\Theta_1;k,\theta)$ for the gamma cfds simply as:

$$\frac{\gamma\left(k',\frac{\Theta_1'}{\theta'}\right)}{\Gamma(k')} > \frac{\gamma\left(k,\frac{\Theta_1}{\theta}\right)}{\Gamma(k)}. \tag{2.24}$$

A key point is that the function $\gamma(k,s)/\Gamma(s)$ (if one sets $s = \Theta_1/\theta$), decreases with k and increases with s. It can be seen in Fig. 2.4, where we have plotted $\gamma(k,s)/\Gamma(s)$ for $k, s > 0$. It can be also verified by representing the incomplete gamma functions in (2.24) by their homomorphic expansions, and then by rewriting (2.24) as follows:

$$\sum_{i=0}^{\infty} \frac{(\Theta_1'/\theta')^{i+k'} e^{-\Theta_1'/\theta'}}{\Gamma(k'+i+1)} > \sum_{i=0}^{\infty} \frac{(\Theta_1/\theta)^{i+k} e^{-\Theta_1/\theta}}{\Gamma(k+i+1)}, \tag{2.25}$$

where the decaying with k could be verified due to the rapid growth of $\Gamma(k+i+1)$ as k grows. Hence, based on the behavior of the incomplete gamma function, a sufficient but not necessary condition for the inequality (2.24) to hold is given if *both* relations are attended:

$$k' < k \quad \text{and} \quad \frac{\Theta_1'}{\theta'} > \frac{\Theta_1}{\theta}. \tag{2.26}$$

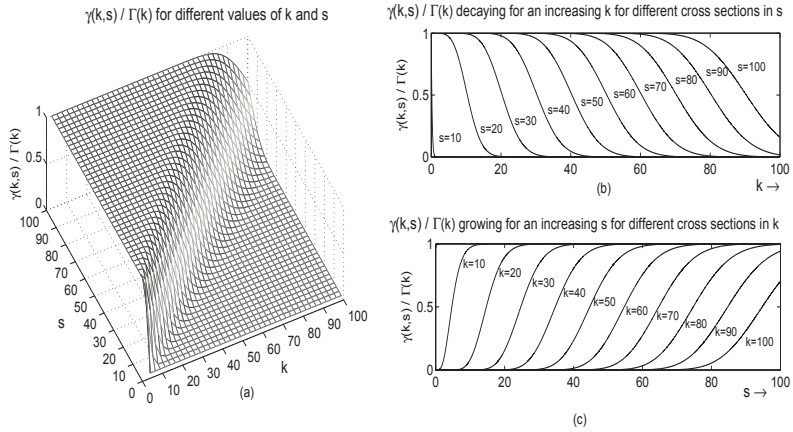

Figure 2.4: Behavior of $\gamma(k,s)/\Gamma(s)$, *i.e.* decreases with k and increases with $s = \Theta_1/\theta$. (a) $\gamma(k,s)/\Gamma(s)$ as function of k and s. (b) $\gamma(k,s)/\Gamma(s)$ decreasing with k for different "slices" over s. (c) $\gamma(k,s)/\Gamma(s)$ increasing with s for different "slices" over k.

For improving the nonstationarity detection, one could change the stationarity test in a sense that, after its application, the new parameters of the approximated gamma model (k', θ') and the new test statistic Θ_1' follow the conditions shown in (2.26) *if* the underlying signal is nonstationary, but not necessarily otherwise. Note that, while stationarity is a well-defined property, nonstationarity can be seen as a non-property [4], and there are many ways by which the stationarity can be violated. Thus, in practice, as we do not know *a priori* about the stationarity/nonstationarity of the signal, we have to search instead for a *nonstationary characteristic*. More specifically, for enhancing the

nonstationarity detection, one could develop a procedure that gives:

$$k' < k \text{ and } \frac{\Theta_1'}{\theta'} > \frac{\Theta_1}{\theta} \quad \begin{array}{l} \textit{If} \text{ the underlying process presents a clear} \\ \textbf{nonstationary characteristic}. \text{ This condi-} \\ \text{tion means that the modified test will more} \\ \text{likely reject the stationarity,} \end{array} \tag{2.27}$$

where many different types of nonstationary behaviors can be imagined. In this work, we analyze if the time-varying spectrum undergoes a structured evolution in time. Having extended the conditions for improving the test to the gamma parameters, we define in the following Section the procedure developed to guarantee (2.27).

2.6 Weighted distances

When computing Θ_0, the vector of variances of the distances between local and global spectra (from where the gamma distribution is approximated), *no consideration was taken about the spectral content of the signal*. Different nonstationary processes can have a spectral content spread in time in a particular way. For example, let us consider the signals and TF representations shown in Fig. 2.5, corresponding to Gaussian processes exhibiting an exclusively varying mean (Fig. 2.5(a)), varying variance (Fig. 2.5(b)) and a jointly varying mean and variance (Fig. 2.5(c)). In general, a signal with a varying mean is not expected to have a spectral content as significant at high frequencies as a signal with varying variance. However, if the mean varies sufficiently fast at a given time location, as it is shown in Fig. 2.5(a), there will be strong high frequency components around that local in the TF spectrum. Nevertheless, even for signals possessing a slowly varying variance, gradually changing according to the *onset-of-trend* model, the time-varying spectrum will undergo a structured evolution in time, as it can be seen in Fig. 2.5(b). Therefore, we propose to use this temporal information of the TF spectrum to modify the distance vectors (and consequently Θ_0), according to the following procedure:

1. Given an observed signal $x(t)$ with TF spectrum $S(t_n, f)$ evaluated at N time positions $\{t_n, n = 1, ..., N\}$, compute the time marginal series $y(t_n) = \sum_f S(t_n, f)$ and normalize it to the unity $[\widetilde{y}(t_n) = y(t_n) / \max y(t_n)]$ to obtain the weighting vector $\{\widetilde{y}_n, n = 1..., N\}$.

2. Use $\widetilde{y}_n$ to weight the distance vectors multiplicatively, *i.e.* $\{\widetilde{c}_n^{(x)} = \widetilde{y}_n c_n^{(x)}, n = 1, ..., N\}$ and $\{\widetilde{c}_n^{(s_i)} = \widetilde{y}_n c_n^{(s_i)}, n = 1, ..., N\}$, where $\widetilde{c}_n^{(x)}$ and $\widetilde{c}_n^{(s_i)}$ are, respectively, the *weighted distance vectors* for the original signal and the i^{th} surrogate.

3. Collect the vector of weighted distances for $i = 1, ..., I$ surrogates and compute their variances. The result will be a new vector $\{\Theta'_0(i), i = 1, ..., I\}$. Differently from Θ_0, the vector Θ'_0 carries information about the time location of the spectral content.

If we modify the stationarity test according to the procedure above, we attend the conditions for k, θ and Θ_1 shown in (2.26), and consequently enhance the nonstationarity detection (mainly for first-order nonstationarity). The demonstration of why the weighting procedure improves the stationarity test is given in Appendix A.1. It should be remarked, however, that in most of the cases the improvement in the performance of the test is only achieved by using *frequency-based* distances. By weighting the probability-based distances according to the procedure above, we end up affecting significantly the performance of the method. The explanation of why the weighting procedure is not appropriate for probability-based distances is also given in Appendix A.2.

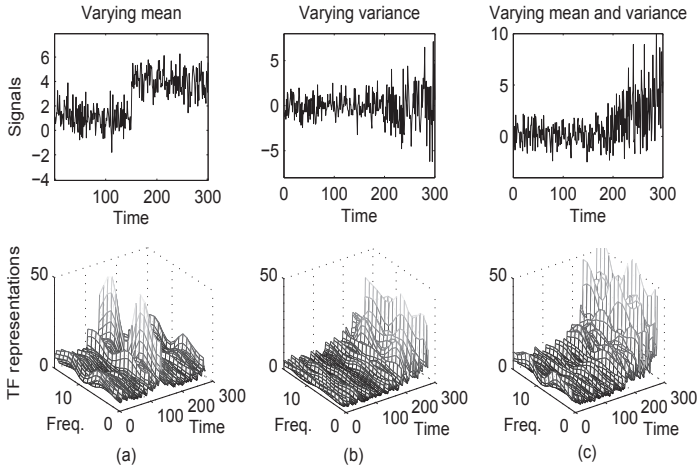

Figure 2.5: Different nonstationary signals and their particular TF representation. (a) Varying mean. (b) Varying variance. (c) Varying mean and variance.

One could have noted that there are other ways to obtain the required information from the TF plane, and consequently, to modify the method. Recall that the distances are computed from local spectra at specific time instants. Hence, while the distances take as input functions of frequency, their output are functions of time. Thus, there are two strategies to create the weighting vector: either to build it in time (as proposed above) *or* in frequency.

If we sum over time (which gives the marginal in frequency), and use the resulting vector to multiply the spectra at the input of a given distance, we will be giving more importance to the most significant frequencies of the TF representation (globally). However, the spectral content of the surrogates has to be *randomly* spread over the TF plane,

and this randomness is affected by forcing the spectra of the surrogates to be concentrated around a certain frequency. Moreover, the marginal spectrum of the surrogates *will not be equal to the one of the original signal*. The procedure for choosing the correct marginal is shown in Fig. 2.6. Notice that temporal structure exhibited by the TF spectrum of Fig. 2.6(a) can be view in the time marginal illustrated in Fig. 2.6(b). Also, it can be seen in Fig. 2.6(d), that by weighting the spectra of the surrogates the equality between the marginal spectrum of the signal and the ones of the surrogate is lost.

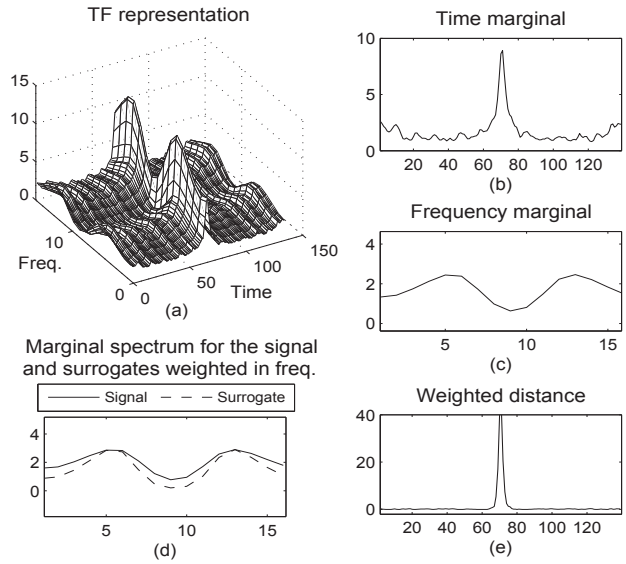

Figure 2.6: Two different ways in which the marginals can be used for building a weight vector. (a) Time-frequency representation of a given nonstationary signal. (b) Time marginal obtained by summing over frequency. (c) Frequency marginal obtained by summing over time. (d) By weighting the spectra at input of the distance, the marginal spectrum of the surrogate will be different from the one of the signal. (e) Weighted distance obtained by choosing the appropriate marginal.

2.6.1 Testing the modified approach

For evaluating the modified method, we reapply the stationarity test to the same signals shown in Fig. 2.2. The results are presented in Table 2.3 as percentage of observed "non-stationary" outcomes over all realizations, for both the *original* and *modified* stationarity test. Note that, due to the aforementioned problem involving weighting probability-based distances (see Appendix A.2), the performance of the modified test was indeed much worse for distances of probability nature than it was before or compared to the distances of frequency nature. Nevertheless, it can be seen that the *overall performance* of the modified approach was improved for the *frequency-based* distances (these cases are highlighted in Table 2.3), at the expense of a slight increase in the misclassification

rate of the statationary signals. In order to verify if the estimated gamma parameters really attended the conditions set in (2.27) for improving the test, we show in Table 2.4, for the frequency-based distances, the arithmetic average of $(k, \Theta_1/\theta)$ and $(k', \Theta'_1/\theta')$ computed for the 1000 realizations of the signals with $T = 300$ and $T = 1050$ shown in Fig. 2.2. Note that, despite a few exceptions, the proposed procedure gives $k' < k$ and $\Theta'_1/\theta' > \Theta_1/\theta$ in the average.

Finally, it can be seen that the detection of trend-based nonstationarities represented by the *onset-of-trend* models still has to be improved. One could point out that the surrogate approach is not the most appropriate choice for testing slowly-varying nonstationarities that appear as trends, while there exist *parametric* methods for testing this kind of nonstationarity [27, 20]. As model-based approaches often fails at representing real world processes, we propose in Chapter 3 a new nonparametric stationarity test that allows for the detection of the trend-based nonstationarities that could not be detected in Table 2.3.

Table 2.3: Results of applying the *original* (shown as c_n (%)) and *modified* (shown as $\widetilde{c}_n$ (%)) stationarity test to 1000 realizations of each process shown Fig. 2.2. The results are given as percentage of observed "nonstationary" outcomes over all realizations.

Nonstationary signals																		
Distance			KL		KM		MS		IS		LS		DF		SIS		CB	
Length			300	1050	300	1050	300	1050	300	1050	300	1050	300	1050	300	1050	300	1050
Varying mean	*Abrupt-change*	c_n (%)	0	0	0	0	0	0	85.0	88.5	10.0	15.0	97.0	100	10.0	10.0	5.00	0
	Middle	$\widetilde{c}_n$ (%)	2.00	5.50	0	0	0	2.50	0	**100**	**100**	**60.0**	**80.0**	**100**	**100**	**62.0**	2.00	3.00
	Onset-of-trend	c_n (%)	0	0	0	0	0	0	1.00	2.00	1.00	3.00	0	0	1.00	0	0	0
	Middle	$\widetilde{c}_n$ (%)	0	0	0	0	0	0	**38.0**	**24.0**	**4.50**	**5.00**	**37.5**	**21.0**	**2.00**	**0**	0	0
	Onset-of-trend	c_n (%)	0	0	0	0	0	0	0	5.00	0	3.00	0	0	3.00	0	1.00	0
	Beginning	$\widetilde{c}_n$ (%)	0	0	0	0	0	0	**31.0**	**18.5**	**3.00**	**6.00**	**32.0**	**4.00**	**3.00**	**0**	0	0
Varying variance	*Abrupt-change*	c_n (%)	17.0	12.0	10.0	14.0	15.0	15.0	100	100	100	100	100	100	99.0	100	90.0	100
	Middle	$\widetilde{c}_n$ (%)	8.00	7.50	3.00	5.00	9.00	7.00	**100**	**100**	**100**	**100**	**100**	**100**	**56.0**	**45.0**	13.0	12.0
	Onset-of-trend	c_n (%)	15.5	14.0	6.00	12.0	8.00	10.0	100	100	100	100	100	100	97.5	100	45.0	97.0
	Middle	$\widetilde{c}_n$ (%)	3.00	4.50	2.00	5.00	6.00	0	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	18.0	45.0
	Onset-of-trend	c_n (%)	5.00	5.00	10.0	12.0	8.00	10.0	100	100	100	100	98.5	100	99.0	100	35.0	83.0
	Beginning	$\widetilde{c}_n$ (%)	7.00	3.00	4.00	7.00	5.00	8.00	**100**	**100**	**87.5**	**100**	**100**	**100**	**39.5**	**54.0**	12.0	14.0
Varying mean and variance	*Abrupt-change*	c_n (%)	100	100	99.0	100	100	100	100	100	94.0	100	100	100	100	100	100	100
	Middle	$\widetilde{c}_n$ (%)	15.0	25.0	4.00	8.50	14.0	25.0	**100**	**100**	**100**	**59.0**	**91.5**	**97.0**	**100**	**80.5**	37.0	100
	Onset-of-trend	c_n (%)	15.0	30.0	14.0	40.0	22.5	30.0	100	100	100	100	97.0	100	95.5	100	63.0	100
	Middle	$\widetilde{c}_n$ (%)	0	0	3.00	4.00	0	0	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	17.0	21.5
	Onset-of-trend	c_n (%)	60.0	100	55.0	100	46.0	100	95.5	100	97.0	100	68.0	87.0	92.0	100	78.5	100
	Beginning	$\widetilde{c}_n$ (%)	0	3.00	2.00	3.00	0	2.00	**100**	**100**	**100**	**78.0**	**100**	**100**	**97.0**	**93.0**	14.5	76.0
Stationary signals																		
Distance			KL		KM		MS		IS		LS		DF		SIS		CB	
Length			300	1050	300	1050	300	1050	300	1050	300	1050	300	1050	300	1050	300	1050
Type: AR(1) process		c_n (%)	7.00	4.50	7.00	5.00	7.00	3.00	9.50	5.50	7.00	6.00	7.00	5.00	5.00	5.00	6.50	4.00
		$\widetilde{c}_n$ (%)	1.00	0	0	0	5.50	4.00	14.5	9.00	7.50	4.50	16.5	14.5	0	0	7.00	5.00

Table 2.4: Verifying the condition of (2.27) after applying the weighting procedure. The results are given as arithmetic averages $(\langle k\rangle, \langle \Theta_1/\theta\rangle)$ and $(\langle k'\rangle, \langle \Theta_1'/\theta'\rangle)$ of the estimates of $(k, \Theta_1/\theta)$ and $(k', \Theta_1'/\theta')$, over 1000 realizations of the signals with length $T = 1050$ and $T = 300$ tested in Table 2.3.

Nonstationary signals																		
Distance			IS				LS				DF				SIS			
Parameters			$\langle k\rangle$	$\langle k'\rangle$	$\langle \Theta_1/\theta\rangle$	$\langle \Theta_1'/\theta'\rangle$	$\langle k\rangle$	$\langle k'\rangle$	$\langle \Theta_1/\theta\rangle$	$\langle \Theta_1'/\theta'\rangle$	$\langle k\rangle$	$\langle k'\rangle$	$\langle \Theta_1/\theta\rangle$	$\langle \Theta_1'/\theta'\rangle$	$\langle k\rangle$	$\langle k'\rangle$	$\langle \Theta_1/\theta\rangle$	$\langle \Theta_1'/\theta'\rangle$
Varying mean	*Abrupt-change*	$T = 300$	7.47	2.50	21.3	44.9	12.6	3.39	13.2	14.1	9.19	2.50	25.5	37.3	3.19	1.64	3.47	5.61
	Middle	$T = 1050$	11.7	3.85	30.5	43.7	16.1	6.52	17.2	18.3	8.06	2.17	24.7	36.1	5.99	2.64	6.37	8.22
	Onset-of-trend	$T = 300$	7.31	4.08	5.58	6.39	11.5	5.98	8.69	9.01	5.87	3.54	3.22	6.43	2.39	1.80	1.26	0.79
	Middle	$T = 1050$	12.2	7.08	10.2	11.1	15.7	10.4	14.1	9.90	5.25	3.48	2.59	5.40	4.05	3.20	2.48	1.92
	Onset-of-trend	$T = 300$	6.86	4.17	5.16	6.80	10.8	6.33	8.41	8.86	5.04	3.43	2.52	5.85	2.21	1.77	1.18	1.20
	Beginning	$T = 1050$	11.8	7.07	10.4	11.1	15.7	10.7	14.1	10.3	4.42	3.12	1.97	4.25	3.57	2.90	1.84	1.94
Varying variance	*Abrupt-change*	$T = 300$	5.77	3.28	11.3	59.7	9.93	7.85	85.2	21.3	6.88	5.16	30.2	78.9	2.39	1.54	3.80	5.67
	Middle	$T = 1050$	11.6	10.3	7.19	13.7	16.2	3.68	8.18	68.9	10.1	19.8	47.6	172	5.47	5.12	2.64	9.91
	Onset-of-trend	$T = 300$	5.63	2.17	69.9	137	10.3	4.97	51.92	52.41	6.81	3.53	48.6	99.8	2.41	1.18	54.1	64.6
	Middle	$T = 1050$	11.3	6.32	337	451	15.6	14.1	19.2	21.4	9.83	9.66	151	241	5.24	3.32	23.3	46.8
	Onset-of-trend	$T = 300$	5.73	3.03	34.6	40.5	10.1	6.16	42.5	13.4	7.09	4.34	22.6	50.5	2.37	1.58	34.2	36.3
	Beginning	$T = 1050$	11.6	8.29	16.1	18.3	16.3	20.1	19.1	35.1	10.5	11.8	48.8	96.5	5.33	4.11	2.03	9.45
Varying mean and variance	*Abrupt-change*	$T = 300$	5.89	3.27	60.7	81.4	10.6	8.00	96.3	98.2	7.91	5.74	24.4	59.6	2.61	1.70	78.3	79.4
	Middle	$T = 1050$	11.8	11.4	63.8	95.8	16.1	14.1	7.03	4.41	8.55	7.23	24.2	7.89	5.64	5.39	25.7	33.1
	Onset-of-trend	$T = 300$	6.00	2.26	65.4	121	10.3	4.78	41.1	45.4	6.77	3.31	33.3	64.1	2.49	1.21	33.3	43.4
	Middle	$T = 1050$	11.7	6.19	299	366	15.7	14.1	17.5	6.73	4.79	2.93	16.7	28.8	5.08	2.99	16.1	36.8
	Onset-of-trend	$T = 300$	5.71	2.96	24.1	37.8	10.5	6.03	13.2	27.6	13.2	7.21	4.28	18.5	2.52	1.62	20.8	4.63
	Beginning	$T = 1050$	16.6	7.35	14.1	72.7	15.7	14.3	15.7	28.4	6.43	5.96	17.9	35.7	5.11	3.52	14.6	15.7

2.7 On the robustness of the test

One drawback of the surrogate-based approach is that the outcome ("stationary" or "nonstationary") can vary at every application of the test, even if one is testing the same realization of the same process. The fluctuation of the results can be a critical point if we are testing real world signals, where in the absence of *a priori* information, we would at least require consistent outcomes from one application of the test to another. According to the point view of distances that has been adopted in this work, it has been defined in Section 2.3.1 that a *robust distance* is the one that leads to less variation of the results.

As illustration, in Table 2.5, we applied the stationarity test repeatedly (1000 times) to a *single* realization of each test signal. The results are given as percentage of nonstationary outcomes. Considering the definition of robustness given in Section 2.3.1, it can be noticed that: i) the case of a nonstationary mean is the least robust one, ii) the probability-based distances (KL, KM and GM) are considerably less robust than the frequency-based ones, and iii) for a few cases, we lose the robustness by weighting the distance. In this Section, we propose a method to evaluate the robustness of the distances that uses classical bootstrap analysis. Before deriving the robustness measure, however, we analyze how the changing outcomes occur.

Table 2.5: Results of applying the *original* (shown as c_n (%)) and *modified* (shown as $\widetilde{c}_n$ (%)) stationarity test *repeatedly* (1000 times) to a *single* realization of each process shown in Fig. A.4. The results are given as percentage of observed "nonstationary" outcomes.

Nonstationary signals																	
Distance			KL		KM		MS		IS		LS		DF		SIS		CB	
Length			300	1050	300	1050	300	1050	300	1050	300	1050	300	1050	300	1050	300	1050
Varying mean	*Abrupt-change*	c_n (%)	0	0	0	0	0	0	95.0	100	30.0	35.0	97.0	100	0	0	0	0
	Middle	$\widetilde{c}_n$ (%)	0	0	0	0	0	0	100	100	52.0	56.0	100	100	95.0	98.0	0	0
	Onset-of-trend	c_n (%)	0	0	0	0	0	0	4.00	5.00	0	0	2.00	3.00	0	0	0	0
	Middle	$\widetilde{c}_n$ (%)	0	0	0	0	0	0	80.0	88.0	0	0	2.00	5.00	0	0	0	0
	Onset-of-trend	c_n (%)	0	0	0	0	0	0	3.00	2.00	0	0	0	0	0	0	0	0
	Beginning	$\widetilde{c}_n$ (%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Varying variance	*Abrupt-change*	c_n (%)	0	2.00	0	0	0	0	100	100	95.0	100	100	100	98.0	100	0	0
	Middle	$\widetilde{c}_n$ (%)	0	0	0	0	0	0	100	100	100	100	100	100	90.0	95.0	0	0
	Onset-of-trend	c_n (%)	0	0	0	0	0	0	50.0	100	20.0	100	45.0	100	15.0	100	0	100
	Middle	$\widetilde{c}_n$ (%)	0	0	0	0	0	0	5.00	100	5.00	100	8.00	100	5.00	100	0	90.0
	Onset-of-trend	c_n (%)	0	5.00	0	0	0	0	5.00	100	0	100	10.0	100	0	100	5.00	100
	Beginning	$\widetilde{c}_n$ (%)	0	0	0	0	0	0	90.0	0	5.0	100	0	100	0	0	0	5.00
Varying mean and variance	*Abrupt-change*	c_n (%)	0	0	0	0	0	0	5.00	100	5.00	100	0	100	0	100	0	100
	Middle	$\widetilde{c}_n$ (%)	0	0	0	0	0	0	100	100	100	100	100	100	100	100	100	100
	Onset-of-trend	c_n (%)	0	0	0	0	0	0	45.0	100	85.0	100	90.0	100	0	100	0	100
	Middle	$\widetilde{c}_n$ (%)	7.00	0	49.0	0	55.0	0	100	100	5.00	100	100	100	0	100	82.0	0
	Onset-of-trend	c_n (%)	0	100	0	100	0	100	83.0	100	44.0	100	97.0	95.0	55.0	100	5.00	100
	Beginning	$\widetilde{c}_n$ (%)	0	0	0	0	0	0	100	100	100	10.0	100	100	0	100	0	100
Stationary signals																		
Distance			KL		KM		MS		IS		LS		DF		SIS		CB	
Length			300	1050	300	1050	300	1050	300	1050	300	1050	300	1050	300	1050	300	1050
Type:		c_n (%)	0	0	0	0	0	0	0	0	0	0	4.00	3.00	2.00	3.00	0	0
AR(1) process		$\widetilde{c}_n$ (%)	0	0	0	0	0	0	4.00	5.00	5.00	2.00	6.50	5.50	6.00	7.00	2.00	2.50

2.7.1 Understanding the changing outcomes

The different outcomes are due to the randomness of the surrogate set, and as a consequence, at every application of the test, a different vector Θ_0 of variances is obtained. Since the gamma fit is performed on Θ_0, the approximated gamma model changes at every test. The effect of this varying distribution is observed as fluctuations of the threshold γ, which would lead to different results if the test statistics Θ_1 is close enough to the *range of fluctuations*. Since Θ_1 is fixed for all test applications (Θ_1 is a non-random quantity computed directly from the signal), one could evaluate *how likely* it is for Θ_1 to be inside the range of fluctuations of γ. This scheme is illustrated in Fig. 2.7, where the distance d between Θ_1 and the threshold γ, the range of fluctuation of γ, and the gamma fit are shown. In this case, it is desired to have d stretching outside the fluctuation limits for both stationary (Fig. 2.7(a)) and nonstationary (Fig. 2.7(b)) cases. On the other hand, having Θ_1 embraced by the fluctuation range (Fig. 2.7(c)) *might lead to changing outcomes*. The latter happens every time the threshold crosses the value of Θ_1 (Fig. 2.7(d)).

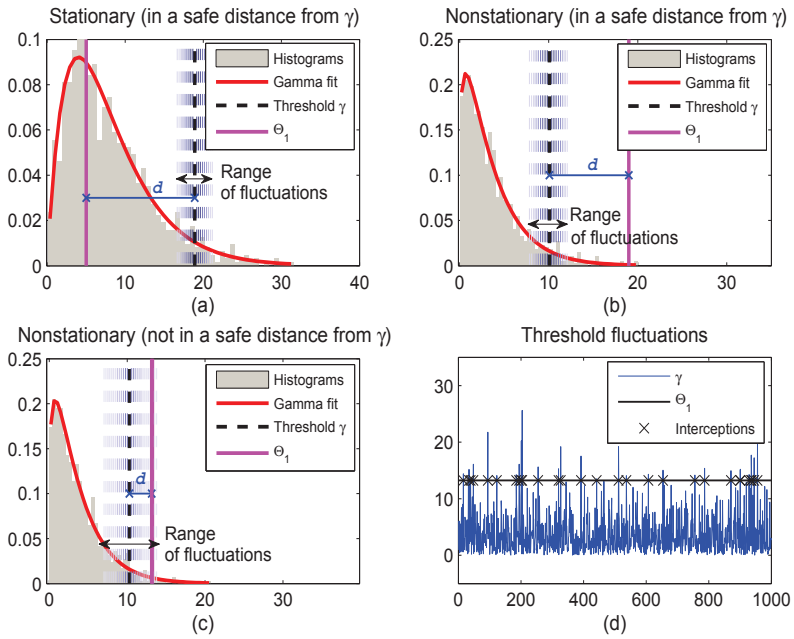

Figure 2.7: Scheme representing how the fluctuations of the threshold may change the results. In (a) and (b), the value of Θ_1 is in a safe distance d from the fluctuations of γ for both stationary and nonstationary cases, respectively. In (c), Θ_1 can be reached by the threshold line, causing changing outcomes. In (d), it is shown that such situation happens whenever γ crosses the fixed value of Θ_1.

2.8 Deriving a robustness measure

The fluctuations of the threshold γ need to be taken into account when evaluating the distances. In this sense, one could suggest, for instance, to compute the histograms of γ, and to analyze how likely it is for Θ_1 to be inside the range of fluctuations of γ. To do so, we could check if the value of Θ_1 has a low probability of being crossed by the threshold, when compared with the histograms. If it does, the dissimilarity measure being used is *robust*.

For the nonstationary signals following the *Onset-of-trend* model, we computed the histograms of the values taken by the threshold γ, when using the Itakura-Saito, Symmetrized Itakura-Saito and Combined distances. For the sake of simplicity the analysis was carried out with the original distances, not the weighted ones. The histograms are shown in Fig. 2.8, where the dashed line in black stands for the 95% line of the empirical distributions approximated by the histograms. If Θ_1 (red line) is beyond this threshold (to the right-hand side), it means that the corresponding distance seldom leads to changing results, being consistent with the definiton of robustness. Notice in Fig. 2.8 for the case of a nonstationary mean (first column), that all values of Θ_1 are lying in

the left-hand side of the histograms, outside of the range of fluctuation of γ. This is in accordance with the results shown in Table 2.5, where most of the distances could not detect properly the nonstationarity exclusively in the mean. Moreover, as also observed in Table 2.5, by symmetrizing the Itakura-Saito distance we lost the robustness of its original form. However, for the case of a nonstationarity exclusively in the variance, for example, the Symmetrized Itakura-Saito distance is more robust than the Combined one.

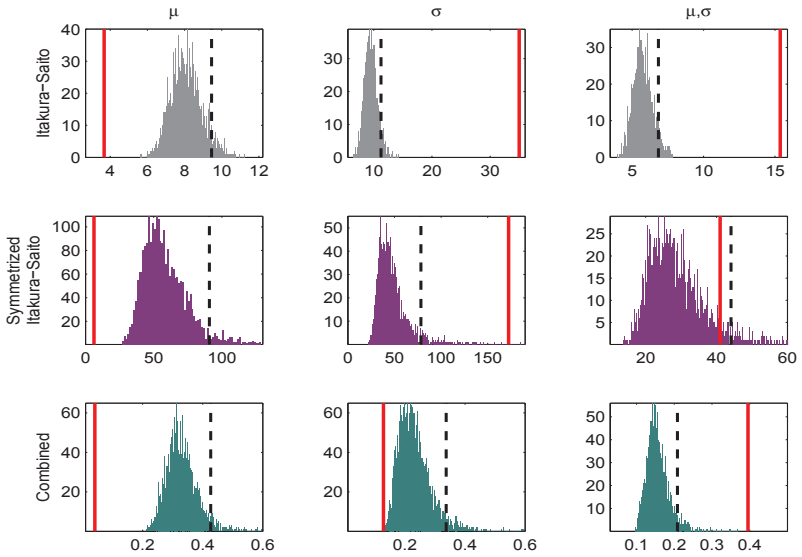

Figure 2.8: Analyzing the robustness of three distances: Itakura-Saito, Symmetrized Itakura-Saito and Combined. The nonstationary signals follow the *Onset-of-trend* model and exhibit three types of nonstationarities: exclusively in the mean (μ), exclusively in the variance (σ), and both in the mean and in the variance (μ,σ). The dashed line in black corresponds to the 95% line of the empirical distribution approximated by the histograms. The solid line in red represents Θ_1, which is expected to lie beyond the dashed line (to the right-hand side), for a robust distance when analyzing a nonstationary signal.

Unfortunately, evaluating the robustness like in Fig. 2.8, *i.e.*, by using the crude histograms of γ, can be considerably time-consuming. The reason is because the stationarity test has to be applied a number of times, and at every test application, a TF representation is computed for each surrogate signal of the collection. Ideally, we would like to be able to evaluate the robustness after the first realization. Doing so, the time-consuming part of computing the TF spectrum repeatedly would be avoided. Evaluating the robustness while having only one record of the threshold γ, would require a statistical characterization of this variable, and some extra knowledge regarding the method. We propose in the following Section, an approache where these two points are

addressed, providing a useful measure of robustness.

2.8.1 An approach based on bootstrapping

For evaluating the robustness like in Fig. 2.8, we have estimated different thresholds γ from a collection of vectors Θ_0. We have then used the crude histograms to check whether the fluctuations of γ could or could not change the results. Now, we propose an alternative (and simpler) method, which consists in computing the vector Θ_0 *once*, and estimating the threshold values by *resampling with bootstrap*.

The classical bootstrap technique [63, 64] is well-suited to this end, as the samples of Θ_0 are i.i.d. The method is performed by randomly sampling the data with replacement, and then gathering equally-sized resamples. Such procedure is done thousands of times and treated as repeated experiments [65]. By bootstrapping we simulate repeated observations from an unknown population using only the obtained sample as basis. It allows us to estimate different threshold values from the same basis vector Θ_0, skipping the time-consuming part of computing different TF spectra. Also, bootstrap replicates can be easily computed and the empirical confidence intervals are straightforwardly obtained even for complex estimators. Finally, the method does not require a large number of observations in order to invoke asymptotic results. In fact, theoretical and practical works have shown that bootstrap techniques can outperform large-sample ones [65].

With a collection of bootstrap replicates of the threshold γ, we can compute empirical confidence intervals or observe the fluctuation of γ using histograms. Then, we can evaluate the robustness by checking how likely it is for Θ_1 to be crossed by γ. However, a natural question arises: How many bootstrap replicates are necessary? In order to estimate the required number of replicates, it would be reasonable to first derive a proper robustness measure, and then check the effect of using different sample sizes.

The surrogates are random sequences that are stationarity, and due to their "well behaved" randomness we could characterize some properties of interest. For instance, since Θ_0 is computed directly from the surrogate set, it should be a stationary random series as well. Also, it is intuitive to address smooth changes to its mean value. Thus, we propose to measure how far Θ_1 (a fixed quantity) is from the mean value of the γ estimates obtained by bootstrapping, by computing the *absolute mean deviation* (AMD) for Θ_1. The AMD expression is given as follows:

$$\text{AMD} = \frac{|\Theta_1 - M|}{M}, \tag{2.28}$$

where M stands for the sample mean of the γ estimates. Note in (2.28), that the ex-

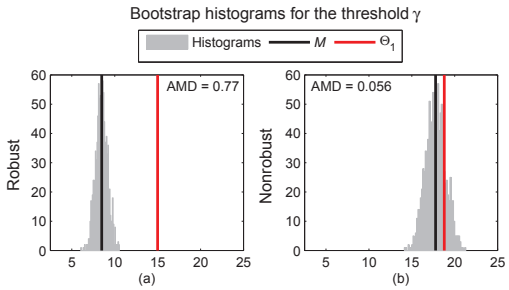

Figure 2.9: Robustness evaluation with bootstraps. (a) A robust case and its AMD value. (b) A nonrobust case and its AMD value.

pression was divided by M in order to avoid discrepancies when comparing cases with large differences between M and Θ_1. Otherwise, a small value of M could lead to a small value of AMD. An illustration of the use of (2.28) for measuring the robustness is given Fig. 2.9, where the histograms of the bootstrap replicates, M, Θ_1 and the AMD value are shown for a robust (Fig. 2.9(a)), and a nonrobust case (Fig. 2.9(b)), respectively. Notice that Θ_1 is far from M for the robust case, which gives a larger value of AMD. On the other hand, for the nonrobust case, Θ_1 is closer to M.

After defining the robustness measure, we need to check how many bootstrap replicates are necessary. To this end, we have bootstrapped Θ_0 and estimated the values of γ for three samples of different sizes: 1000, 5000 and 10,000. Although bootstrapping is faster than the procedure shown in Fig. 2.8, we have observed that the computation time increases notably by using more than 10000 replicates. For illustration, in Fig. 2.10, we show the results and the running time (in seconds) for the Itakura-Saito, Log-Spectral and Diffusion distances when testing the signals with a nonstationary mean following the *onset-of-trend* model (which was the least robust case according to Table 2.5). Notice that, in general, the gain in accuracy obtained by using the largest number of replicates cannot be identified clearly, while, on the other hand, the gain in the running time is well noticed. Thus, we have considered 1000 replicates as a good trade-off between computational time and robustness assessment. Increasing the number of replicates does not increases the amount of information from the data, but reduces the random sampling error from bootstrapping.

Having presented the approach to evaluate the robustness, now we compute the AMD for all the distances, the original and the weighthed ones. Naturally, the largest AMD values are expected to occur for the least changing cases, be they stationary or nonstationary. The results are shown in Table 2.6. Note that, in general, the nonrobust cases for each distance have low AMD values. Furthermore, as expected, the frequency-

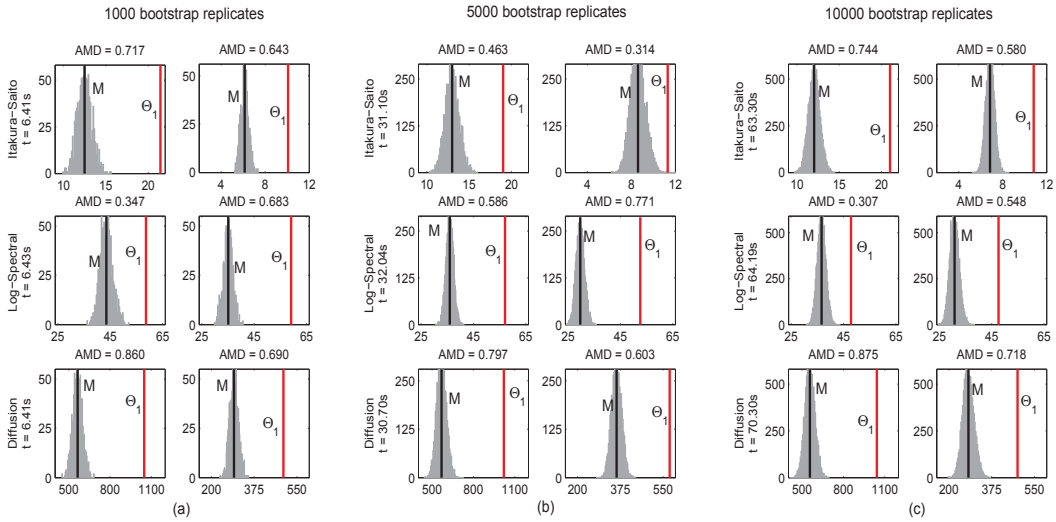

Figure 2.10: For the Itakura-Saito, Log-spectral and Diffusion distances, and signals with varying mean, the gamma threshold fluctuations obtained by bootstrapping, the running time and the AMD values computed by means of (2.28). The mean of the fluctuations is given by M.

based distances presented most of the largest AMD values (especially the Itakura-Saito and the Diffusion distances). As previously observed, for this class of distance we had the least changing (nonstationary) outcomes. Finally, note that for some cases, we have large AMD values for the probability-based distances, which reflects how constantly the null hypothesis is accepted for this class of distances.

2.8.2 Conclusions

We have seen that many signals found in the real world have a rich spectral content, are very often nonstationary, and are frequently given without any *a priori* information about their structure. Given these circumstances, TF techniques excel at providing a useful representation for nonstationary signals, and different approaches have emerged for testing stationarity in TF domains. In this Chapter, we have proposed various contributions to one of these approaches, more specifically, to the surrogate-based method presented in [41]. The surrogate method makes use of the so-called surrogates for characterizing the null hypothesis of stationarity, whereas the possible nonstationary behavior is verified by comparing the local and global spectral properties of the TF representation.

The original framework presented in [41] leaves room for improvements in various points. So in this Chapter, we have proposed several contributions to some aspects of the original method that deserved a deeper investigation. We have carried out the ana-

Table 2.6: AMD values computed for the *original* (shown as c_n) and *modified* (shown as $\widetilde{c}_n$ stationarity test.

Nonstationary signals																	
Distance			KL		KM		MS		IS		LS		DF		SIS		CB	
Length			300	1050	300	1050	300	1050	300	1050	300	1050	300	1050	300	1050	300	1050
Varying mean	*Abrupt-change*	c_n AMD	0.93	0.86	0.75	0.79	0.81	0.79	1.17	0.81	0.41	0.38	0.96	0.87	0.13	0.56	0.84	0.88
	Middle	$\widetilde{c}_n$ AMD	0.95	0.69	0.83	0.63	0.87	0.83	9.72	1.20	0.47	4.49	5.12	7.12	3.42	0.76	0.85	0.74
	Onset-of-trend	c_n AMD	0.74	0.71	0.67	0.69	0.49	0.84	0.51	0.61	0.42	0.23	0.74	0.71	0.70	0.67	0.24	0.93
	Middle	$\widetilde{c}_n$ AMD	0.89	0.82	0.79	0.71	0.42	0.82	0.63	0.63	0.40	0.37	0.85	0.61	0.30	0.28	0.29	0.94
	Onset-of-trend	c_n AMD	0.84	0.92	0.69	0.76	0.23	0.89	0.48	0.49	0.81	0.14	0.91	0.77	0.90	0.68	0.99	0.92
	Beginning	$\widetilde{c}_n$ AMD	0.76	0.87	0.57	0.63	0.88	0.89	0.94	0.33	0.69	0.32	0.41	0.47	0.72	0.63	0.89	0.88
Varying variance	*Abrupt-change*	c_n AMD	0.32	0.33	0.36	0.22	0.64	0.16	15.3	36.4	6.22	27.7	0.22	3.20	29.7	25.4	0.98	4.52
	Middle	$\widetilde{c}_n$ AMD	0.48	0.42	0.50	0.18	0.44	0.36	25.5	10.1	1.14	0.41	3.65	5.92	0.54	2.77	0.23	0.41
	Onset-of-trend	c_n AMD	0.54	0.33	0.05	0.60	0.55	0.09	9.73	26.7	0.94	6.00	3.59	7.77	6.64	14.3	1.20	1.74
	Middle	$\widetilde{c}_n$ AMD	0.94	0.12	0.65	0.60	0.23	0.72	42.8	78.6	1.21	1.74	17.2	12.1	1.47	2.14	0.87	4.68
	Onset-of-trend	c_n AMD	0.27	0.48	0.32	0.20	0.06	0.35	1.70	7.18	2.45	7.02	1.08	3.62	0.11	16.1	0.75	0.84
	Beginning	$\widetilde{c}_n$ AMD	0.21	0.68	0.62	0.35	0.85	0.56	3.16	4.65	0.12	0.29	4.13	6.19	0.21	0.07	8.46	0.67
Varying mean and variance	*Abrupt-change*	c_n AMD	1.77	6.92	0.19	3.96	0.63	7.18	7.02	50.5	5.54	33.0	1.08	38.7	115	324	22.5	75.7
	Middle	$\widetilde{c}_n$ AMD	0.06	0.11	0.27	0.41	0.57	0.29	4.11	5.08	0.33	8.53	8.29	3.00	1.63	1.96	0.37	2.76
	Onset-of-trend	c_n AMD	0.57	0.18	0.54	0.25	0.05	0.22	3.84	10.2	3.31	6.89	1.01	8.15	3.93	15.1	0.17	1.62
	Middle	$\widetilde{c}_n$ AMD	0.70	0.86	0.32	0.50	0.46	0.75	14.0	19.4	2.34	1.86	14.1	11.7	1.21	4.20	0.92	0.72
	Onset-of-trend	c_n AMD	0.58	1.67	0.60	0.74	0.44	0.98	2.68	6.95	0.47	6.05	0.10	1.16	2.15	13.2	5.51	1.16
	Beginning	$\widetilde{c}_n$ AMD	0.75	0.70	0.43	0.47	0.69	0.47	7.61	4.10	2.48	0.72	1.61	2.59	0.42	1.33	0.80	0.22
Stationary signals																		
Distance			KL		KM		MS		IS		LS		DF		SIS		CB	
Length			300	1050	300	1050	300	1050	300	1050	300	1050	300	1050	300	1050	300	1050
Type:		c_n AMD	0.93	0.94	0.56	0.54	0.63	0.82	2.03	1.33	0.67	0.58	2.01	2.88	1.39	1.45	0.89	0.23
AR(1) process		$\widetilde{c}_n$ AMD	0.98	0.97	0.72	0.74	0.88	0.93	1.44	1.39	0.89	0.37	2.56	1.01	2.78	0.88	0.44	0.34

lysis from the point of view of the distances between local and global spectra. The first contribution of the Chapter consisted of an investigation on the effect of using distances of different natures (frequency and probability-based). In this first part, it was observed that the original framework could hardly detect a changing mean and nonstationarities varying slowly as a trend, regardless of the chosen distance. Hence, as a second contribution, we have proposed to modify the methodology and to weight the distances by a vector built from the time marginal of the TF representation. Doing so, we could improve the performance of the test and increase the detection rate of first-order nonstationarities, but only for the frequency-based distance. The last contribution consisted in evaluating the robustness of the test against the changing results that may appear by applying the test sequentially to the same realization of the signal. We have proposed a robustness measure that makes use bootstrapping for estimating a collection gamma thresholds from a single realization of the process. Doing so, we have successfully assessed the least and most robust cases.

Although this Chapter contributed to the search of a "good" distance, and to modify the method in a way that its performance is improved for the frequency-based distances, one could say that the classification accuracies were not high for some cases. The

surrogate-based method does not work particularly well for the types of nonstationarity that were the most problematic in this Chapter, *i.e.*, first-order evolutions and slowly-varying nonstationarities. For detecting such types of nonstationarities, we develop in the next Chapter a new nonparametric, data-driven technique for testing stationarity.

Chapter 3

A new nonparametric test for trend-based nonstationarity

Contents

In Chapter 3, a new nonparametric approach for detecting slowly-varying nonstationarities is proposed. Different from the surrogate-based method, the proposed technique does not require the computation of many TF representation for assessing stationarity, which fastens the procedure. Instead, the idea is to quickly estimate the time marginal (Section 3.2) and to identify temporal patterns (Section 3.3). In comparison to other approaches in the literature, the developed framework shows to be more versatile, due to its sensitivity to first and second-order nonstationarities and its nonparametric nature.

3.1 Introduction

The basic idea of testing stationarity in TF domain with surrogates was to use this data-driven resampling method for learning the statistics of a null hypothesis of stationarity. While the original test in [41] was designed for second-order stationarity and stationarity relatively to a global observation scale, we have proposed some modifications to the original framework for allowing the characterization of nonstationarities that also appear in the mean, and for only portions of the time series (*i.e.* nonstationarities with particular start and end points, such as $\kappa_1 = T/2$ and $\kappa_2 = T$ in Fig. 2.2, respectively). As it could be seen, the proposed modified approach presented an overall improved performance, but only for frequency-based distances. Regardless of the improvements brought by the modifications proposed in the last Chapter, a drawback faced by nonparametric TF methods using surrogate resampling is their *computational load*, as one need to compute many TF representations, one for each surrogate. Also, the framework discussed in the previous Chapter could not detect properly a *varying mean* and *slowly-varying nonstationaries* simply because it was not designed to this end. The surrogate-based technique works particularly well for modulations, and it has been seen in Section 2.6 that signals following the *abrupt-change* model could be fairly well detected.

Although many practical problems arising in signal processing are actually concerned in detecting abrupt-changes [25], the detection of slowly-varying nonstationarities, specially those of second-order and for short time series, remains a challenge. Detecting slowly-varying changes of first and second-order statistics is a critical point

when analyzing real world time series, as changes of a given order could reflect very specif behaviors of natural phenomena. This kind of time series has usually a short length, and as consequence, only a few stationarity tests compatible to this application can be found in the literature. A first suggestion would be to apply the classical KPSS test [27]. However, the latter is not suited to detect changes of second-order (see Section 3.9). Other alternatives are also given by recent parametric methods, such as the nonstationarity detector of S. Kay [20], which is conveniently suited to short time series. Unfortunately, the performance of a parametric method depends on the accuracy of the chosen model, which is hardly assessed for real world processes (as it will be verified in Chapter 5). Therefore, the purpose of this Chapter is to propose a new nonparametric stationarity test, which is particularly designed for testing trend-based and slow nonstationary evolutions. The proposed method is data-driven and more sensitive to first-order nonstationarities than other nonparametric methods.

3.1.1 Detailing the contributions of this chapter

For slowly-varying cases, full TF representations are not needed to assess for nonstationary behaviors. The stationarity test that is developed in this Chapter is more efficient in this regard. We designed the test for a specific nonstationary pattern: presence of a trend and/or an evolution of the local energy of the signal. The proposed approach works for short time series and allows for the detection of first and second-order nonstationarities of slowly-varying signals. The proposed methodology is illustrated in Fig. 3.1. More specifically, in Section 3.2.1, we recall some elements of TF analysis for assessment of stationarity, and we explain the method of [66] for extracting the trend in the time marginal of the TF representation. Further, we propose a measure to quantify the trend contamination. We then compare this quantity with the ones that are likely to be found in stationary references, which are obtained by block bootstrapping the original signal. In Section 3.7, we show that the distribution of possible trend contaminations can be approximated by a generalized extreme value (GEV) model, allowing us to design a hypothesis test in Section 3.8 to detect these second-order trend nonstationarities. We also derive a typical index of nonstationarity. The experimental study is shown in 3.9.

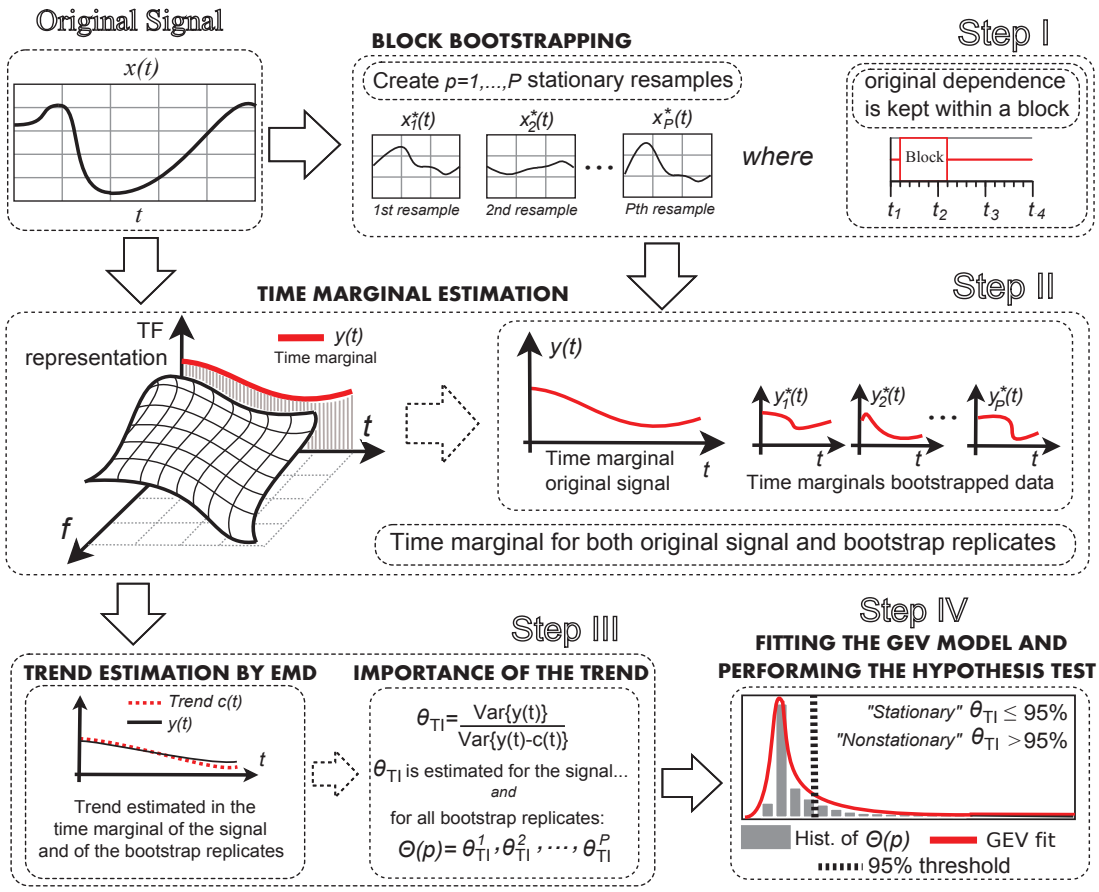

Figure 3.1: Flowchart representing the procedure for testing operational stationarity by evaluation of the trend importance in the time marginal.

3.2 On a test suited to trends and slow nonstationary evolutions

3.2.1 Background Elements

Consider that a given random process $\{x(t), t = 1, ..., T\}$ possesses a slowly varying type of nonstationarity, in a sense that its power spectral density (PSD) varies slowly in time, as opposed to an abrupt change, for instance. For example, let us assume that $x(t)$ is a nonstationary Gaussian process with a slowly-varying variance starting at $t = T/2$ following the *onset-of-trend* model, and $x_a(t)$ and $x_w(t)$ are two strict stationary signals, an autoregressive AR(1) process and a white Gaussian noise WGN(0,1), respectively. Realizations of $x_a(t)$, $x_w(t)$ and $x(t)$ are shown in Fig. 3.2 (a), while their TF spectra estimated by multitaper spectrogram (see (2.2)) are shown in Fig. 3.2 (b). The time marginals obtained by integration of the TF spectra over frequency are shown inFig. 3.2 (c), and a temporal structure in the time marginal of the nonstationary signal can be clearly observed through the fitted trend curve (this trend fitting will be explained further in this

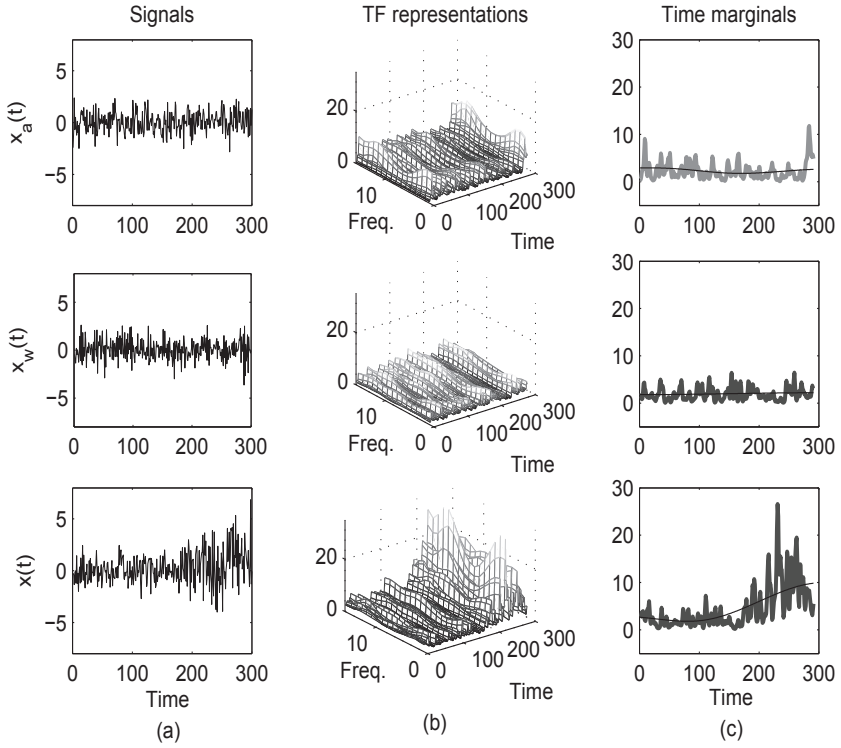

Figure 3.2: (a) Two stationary processes ($x_a(t)$ and $x_w(t)$) and a nonstationary one ($x(t)$). (b) Estimated TF spectra of all signals. (c) Time marginal distribution obtained by integration in the frequency (in gray, thick line), and estimated trend (in black, thin line).

Chapter).

For this type of nonstationary signals, the TF spectrum should generally read as:

$$S(t,f) = \alpha(t)^2\delta(f) + \beta(t)(\widetilde{\mathrm{PSD}}(f) + \epsilon(t,f)) \tag{3.1}$$

where $\widetilde{\cdot}$ is the normalization of the PSD. In (3.1), $\alpha(t)$ can be interpreted as a first-order trend (time-varying mean), $\beta(t)$ is a second-order trend (time-varying variance), $\widetilde{\mathrm{PSD}}(f)$ is possibly the PSD of the stationary component (or, if the signal is nonstationary, it is the frequency marginal of the TF spectrum) and $\epsilon(t,f)$ are the fluctuations in the TF domain that might code for other types of nonstationarities, such as frequency modulations [41]. We are interested here only in second-order trend (non)stationarities, where α and β evolve possibly in time, but changes in frequency as coded in $\epsilon(t,f)$ are not relevant.

The type of nonstationarities we are interested in reduce to test for the presence of time evolving α or β. The time marginal of eq. (3.1) is obtained by summing $S(t,f)$ over all frequencies and it reduces to: $y(t) = \alpha(t)^2 + \beta(t)$ as $\sum_f \epsilon(t,f) \simeq 0$ as there are no changes in frequency. The situation is not that simple as we cannot specify *a priori* a model for the time marginal in general situations. However, it is possible to use trend

estimation to first extract trends, before testing its statistical relevance.

Reducing the study of trend stationarity to the study of the time marginals of the TF spectrum is able to considerably fasten the computations needed in that situation: instead of computing an estimate of the full TF spectrum by means of the multitaper spectrogram given in (2.2), one needs only an estimate of $y(t)$. Thanks to the well-known marginal properties of TF distributions [4], this is easily computed by numerical convolution of the signal $x(t)$ by the window of analysis. The latter was chosen to be the same family of Hermite functions $\{h_k(t), k \in \mathbb{N}\}$ with a length n_{h} used in Section 2.2. Then, the time marginal is computed by squaring $x(t) * h_k(t)$ averaging over the different K, *i.e.*:

$$y(t) = \frac{1}{K} \sum_{k=1}^{K} \left[x(t) * h_k(t) \right]^2 , \tag{3.2}$$

this solves one drawback of the nonparametric stationarity test in TF domain, which is the computational load of having to estimate full TF representations. Nevertheless, for dealing with any type of real world signal, we need a model-free approach for extracting the trend. Among the different nonparametric approaches for trend estimation, the Singular Spectrum Analysis (SSA) [67] and the Empirical Mode Decomposition (EMD) [66] are known for decomposing the time series into oscillatory components. Both methods have their qualities, however, the SSA depends on a free parameter, while the EMD is fully data driven [66]. Due to the latter, and to the simplicity of the EMD algorithm, we have chosen the EMD technique for trend estimation.

3.3 Estimating trends in the time marginal

3.3.1 Trends in the time marginal: Definition

Before explaining the EMD technique, however, we should define the trend component properly. Since "trend" itself is a concept that is context-dependent, we shall adopt a definition that is in consonance with the proposed framework. The definition of trend adopted in this work is that of *a smooth additive component that carries information about global changes in the time series* [68]. Also, it is assumed that the detrended series should have generic stationary features [66]. We represent trend component by $c(t)$, the stationary (detrended) fluctuation by $r(t)$, and the trended time series by $y(t)$. Then, we consider the following additive model to hold:

$$y(t) = c(t) + r(t). \tag{3.3}$$

Testing for second-order trend stationarity can be done by first estimating a trend $c(t)$ from the time marginal $y(t)$ (see Section 3.3.2), then by proposing a way to quantify whether $c(t)$ is relevant or not (see Section 3.4). In the rest of this Chapter, we discuss stationarity in regard to trend stationarity only.

3.3.2 Trends in the time marginal: Estimation

The EMD is an algorithm introduced in [69] for decomposing time series into a superposition of oscillatory terms known as intrinsic mode functions (IMFs). The IMFs are computed iteratively and must satisfy two conditions regarding the number of zero crossing versus extrema and the mean values of local envelopes [70]. The algorithm for extracting the IMFs (known as *sifting process* [69]) can be described as follows [71]:

1. Identify all the local extrema in the time series $y(t)$
2. Interpolate all the local maxima and minima by a cubic spline to produce an upper envelope $e_{\text{up}}(t)$ and a lower $e_{\text{lo}}(t)$ envelope, respectively
3. Compute the mean $m(t) = [e_{\text{up}}(t) - e_{\text{lo}}(t)]/2$
4. Compute the detail $d(t) = y(t) - m(t)$
5. Repeat steps 1 to 4 until the detail $d(t)$ can be considered to be a zero-mean signal according to some stopping criterion. If so, $d(t)$ is called an IMF and the procedure continues by iterating on the residual $m(t)$.

The sifting process stops when the slowly-varying residual function has no more oscillations. We represent this last residual function by $\rho^I(t)$, and the result of the sifting algorithm is a collection of I IMFs $\{m^{(i)}, i = 1, ..., I\}$ plus $\rho^I(t)$. For the time marginal $y(t)$, the EMD gives the following representation:

$$y(t) = \sum_{i=1}^{I} m^{(i)}(t) + \rho^I(t), \tag{3.4}$$

where the IMFs ranging from $m^{(1)}(t)$ to $m^{(I)}(t)$ represent local oscillations going from the shortest period ($m^{(1)}(t)$) to the longest one ($m^{(I)}(t)$) [66]. There are two important points concerning the IMFs that we further recall in this work:

1. each IMF is a zero-mean waveform by construction [70],
2. for all practical purposes, the consecutive IMFs can be considered to be locally orthogonal to each other [69].

As we have defined trend as a slowly-varying component, we shall follow here the same idea of [66], which considers the trend component as being the superposition of the last few IMFs and the residual $\rho^I(t)$. Therefore, estimating the trend will be equivalent to estimating the index $i = i^*$ in (3.4) that gives the best approximation of the trend component $c(t)$:

$$c(t) = \sum_{i=i^*}^{I} m^{(i)}(t) + \rho^I(t). \tag{3.5}$$

The criterion for obtaining i^* is to select the *smallest common index of two independent approaches*, namely the Ratio and the Energy approach. The former verifies deviations of the expected ratio of zero crossing performed by successive IMFs. The latter considers that the energy of a given IMF generally *increases* for i near to i^* [66].

In Fig. 3.2(c) we have illustrated the EMD approximation of trends components estimated from the three time marginals of the stationary and nonstationary signals. Now, for the same test signals, we present in Fig. 3.3 the time marginals $y(t)$ computed with (3.2) by using windows $h_k(t)$ of different lengths ($n_{\mathrm{h}} = 3, 5, 7, 9$), and also the corresponding trend components $c(t)$ estimated by the EMD-based method. Notice that the trend estimated by the EMD does not need to be monotonic. Having defined the procedure of estimating the trend component $c(t)$, we now have to quantify its contamination in the time marginal and to decide whether or not it is significant enough to reject stationarity.

3.4 Testing for trend relevance

3.4.1 Overview

Existence of an additive trend in $y(t)$ conveys a breakdown in the trend stationarity we are interested in. A key point of the method is to estimate the importance of that trend, and further test for stationarity by means of a hypothesis test. More specifically, we build a one-sided test where the null hypothesis of no trend (stationarity) refers to the situation where trends in the time marginal cannot be distinguished statistically from those induced by random fluctuations.

We need to approximate different realizations of the signal to perform the hypothesis test. To do so, we apply the block bootstrap technique to the original signal to obtain a collection of virtual realizations (Section 3.5). The bootstrap replicates are stationary, so trends in their time marginals are unlikely to be found. We analyze the resamples individually by measuring the importance of the trend in $y(t)$ by means of the *trend importance estimator* (Section 3.4.2). Then, by studying the behavior of this estimator

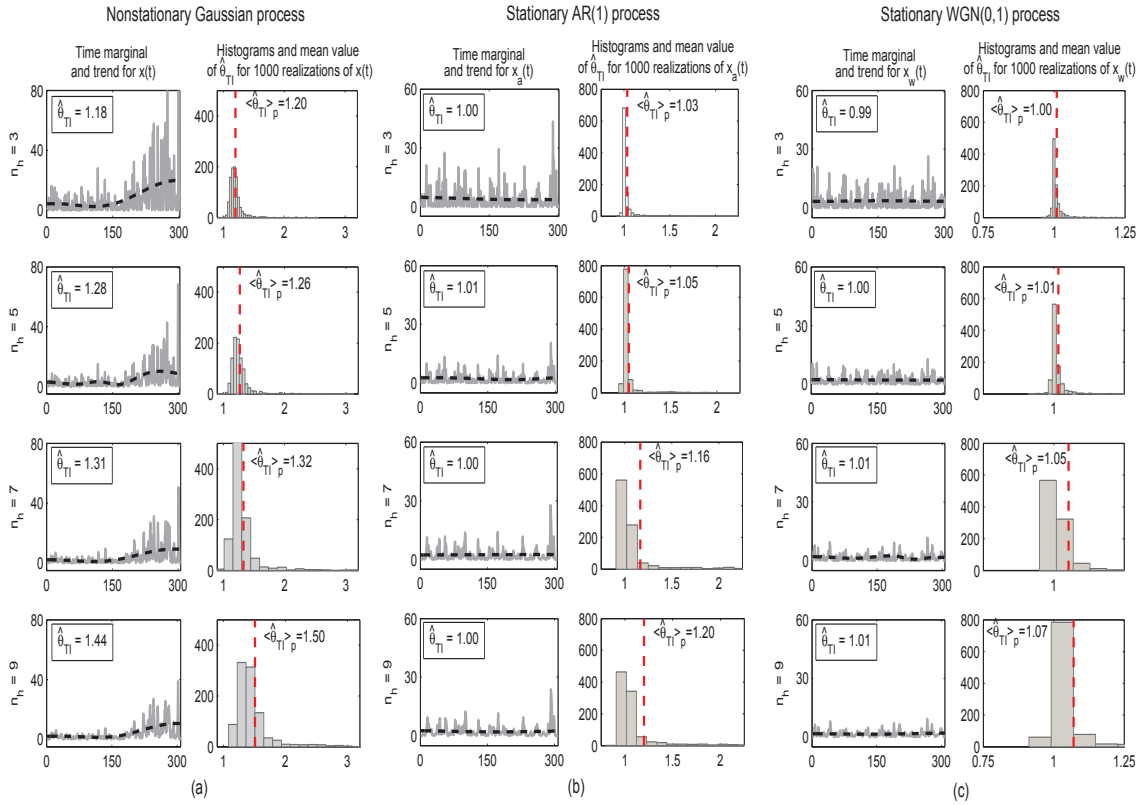

Figure 3.3: Time marginal $y(t)$ (in gray), trend component $c(t)$ (dashed line in black), histograms of $\widehat{\theta}_{\text{TI}}$ and its average value (dashed line in red) for the signals shown in Fig. 3.2. (a) Nonstationary Gaussian process. (b) Stationary AR(1) process. (c) Stationary WGN(0,1) process.

(Section 3.6), we see that its distribution can be approximated by a generalized extreme value (GEV) model (Section 3.7.1).

3.4.2 Estimating the importance of the trend

Here, we propose the following measure to quantify the importance of the trend:

$$\theta_{\text{TI}} = \frac{\mathbb{V}\text{ar}\{y(t)\}}{\mathbb{V}\text{ar}\{y_{\text{dt}}(t)\}} = \frac{\mathbb{V}\text{ar}\{y(t)\}}{\mathbb{V}\text{ar}\{y(t) - c(t)\}}, \tag{3.6}$$

where $y_{\text{dt}}(t) = y(t) - c(t)$ stands for the detrended time marginal series. Such expression measures the importance of the trend component in the data in comparison to the total variance, by computing the fraction of the original variance of $y(t)$ accounted for by the approximated trend [72]. The estimation of (3.6) is carried out by the trend

importance estimator:

$$\widehat{\theta}_{\mathrm{TI}} = \frac{\sum_{t=1}^{N} \left[y(t) - \widehat{\mu}_{\mathrm{y}}\right]^2}{\sum_{t=1}^{N} \left[y_{\mathrm{dt}}(t) - \widehat{\mu}_{\mathrm{dt}}\right]^2}, \tag{3.7}$$

where $\widehat{\mu}_{\mathrm{y}}$ and $\widehat{\mu}_{\mathrm{dt}}$ are the sample mean of the time marginal series and its detrended counterpart, respectively. In Fig. 3.3 we illustrate the performance of (3.7) in evaluating trended time marginals. Besides of computing the time marginals and trends components for different window lenghts ($n_{\mathrm{h}} = 3, 5, 7, 9$), we have estimated the importance of the trend for $P = 1000$ realizations of each process, so in Fig. 3.3 we also show the histograms of $\widehat{\theta}_{\mathrm{TI}}$ and its average value ($\langle \widehat{\theta}_{\mathrm{TI}} \rangle_P$) over all the realizations.

It can be noticed in Fig. 3.3 that $\widehat{\theta}_{\mathrm{TI}}$ is greater for the nonstationary process ($x(t)$) than it is for the stationary ones ($x_a(t)$ and $x_w(t)$). Also, the value $\widehat{\theta}_{\mathrm{TI}}$ tends to decrease when the window of analysis becomes shorter, or to increase otherwise. One could suggest to always use larger windows, but larger values of $\widehat{\theta}_{\mathrm{TI}}$ are not desirable for the case of stationarity. Hence, as usual for methods developed in TF domain, there exists a necessary trade-off regarding the choice of the window of analysis.

More precisely, a trend (a global and slowly-varying component) will be more likely accused by using larger windows, whereas a narrow window of analysis $h_k(t)$ will capture better the local variability in time of the spectra, which will give in return a non-smooth time marginal with an abundance of spikes. In this case, the portion of the variance of $y(t)$ accounted for by the slowly varying component $c(t)$ (the approximated trend) will be smaller, which will decrease the value of $\widehat{\theta}_{\mathrm{TI}}$. One could also interpret this behavior in terms of the IMFs, as the EMD of a time marginal with many spikes and fluctuations will likely use more energetic IMFs to represent fast oscillation modes, rather than slowly varying ones. The analysis of $\widehat{\theta}_{\mathrm{TI}}$ from the point of view of the IMFs and their energies will be carried out in Section 3.4.2.

In practice, one needs a window sufficiently large for visualizing any significant trend component $c(t)$ in $y(t)$ (whereas the significance of $c(t)$ is to be assessed), while keeping the best possible time resolution. By now, we have to consider that there is no *a priori* value for n_{h}, but allowing the window length vary is an extra degree of freedom of the proposed method.

3.5 Generating virtual realizations with bootstrapping

Among the various resamples techniques, the classical bootstrap introduced by Efron [63] remains as one of the most popular methods to derive the distribution of a given estimator. Its attractiveness lies in its simplicity, as the bootstrap can be easily computed [65]. Unfortunately, the classical bootstrap is not applicable for dependent time series, since the original resampling technique could destroy the underlying correlation structure. Therefore, for obtaining approximated realizations of the data while preserving any possible correlation, we propose to use the block boostrap method.

The rationale behind the block bootstrap technique is to divide the time series into blocks prior to resampling. Doing so, we preserve the dependence in the original time series within a block, without having to guess the correlation structure.

There are different approaches available for blocking the time series [73, 74]. They can be broadly categorized into *overlapping* and *nonoverlapping* block bootstrap. In this Chapter, we chose the latter method in order to obtain bootstrap replicates containing as many different samples as possible, while approximating (or equaling) the original length of the time series. The latter is an important point for the time marginal computation. Also, both blocking techniques have the same amount of bias asymptotically for a given block length [74].

3.5.1 The block boostrap technique

For generating the stationary counterparts, let us assume that the original time series is given by $\mathcal{X} = (x_1, ..., x_T)$. The nonoverlapping approach consists in dividing the data into b disjoint blocks, where $b = \lfloor T/\ell \rfloor$, T and ℓ stands for the length of the time series and the length of the block, respectively. The k^{th} block is then given by

$$B_k = [x_{(k-1)\ell+1}, ..., x_{k\ell}] \quad \text{for } 1 \leq k \leq b.$$

To build the bootstrap replicate, we select blocks $\{B_k^*, k = 1, ..., b\}$ by randomly resampling with replacement the blocks $\{B_k, k = 1, ..., b\}$ gathered from the original data. The symbol "*" is used to make a distinction between the resampled blocks and the original ones. After resampling, the bootstrap replicate is given by [75]:

$$\mathcal{X}^* = (x_{1_1}^*, ..., x_{\ell_1}^*, x_{1_2}^*, ..., x_{\ell_2}^*, ..., x_{1_b}^*, ..., x_{\ell_b}^*),$$

where $x_{1_1}^*$ and $x_{\ell_1}^*$ are the first and the ℓ^{th} samples of B_1^* (the first resampled block), $x_{1_2}^*$ and $x_{\ell_2}^*$ are the first and the ℓ^{th} samples of B_2^* (the second resampled block), and so on.

Clearly, the choice of the block length ℓ is a crucial point. Roughly speaking, we want blocks long enough to account for a significant portion of the correlation, while short enough to reduce the variability [76]. Thus, the optimal block length depends strongly on the context. In this work, we have adopted the criterion presented in [75], which states that if one wants to estimate a one-sided distribution (which is the case here), the mean squared error of the block bootstrap estimator is asymptotic to:

$$T^{-1}(C_1\ell^{-2} + C_2T^{-1}\ell^2), \tag{3.8}$$

where C_1 and C_2 are two positive numbers that do not depend on the length of the time series (T), the number of blocks (b) and their length (ℓ). It is shown in [75] that the optimal block length that minimizes (3.8) is of order of $T^{1/4}$. It can be simply checked in (3.8) by minimizing the expression over ℓ. Thus, in this work, we have chosen $\ell = T^{1/4}$.

For each bootstrap replicate $\mathcal{X}^*$, we estimate the trend importance in the time marginal by means of (3.7), which allows for the determination of the distribution of trend contaminations under the null hypothesis of stationarity. We remark that using the empirical distribution to elaborate the hypothesis test is not judicious, as it would require many computations for deriving a reliable threshold from the crude histograms. Fortunately, we have observed that the empirical distribution of (3.7) can be approximated by a generalized extreme value (GEV) model, which simplifies the selection of a threshold. To understand the choice of a GEV model, we take a closer look at the behavior of $\widehat{\theta}_{\mathrm{TI}}$ in the following Section.

3.6 Behavior of the trend importance estimator

We analyze the behavior of $\widehat{\theta}_{\mathrm{TI}}$ in two situations. The first one stands for the case of a *trendless* time marginal (which indicates stationarity), where a significant trend cannot be detected in $y(t)$. The second situation happens if we have a *trended* time marginal (which indicates nonstationarity), where a significant trend can be detected in $y(t)$.

3.6.1 Trendless time marginal

For the case of a *trendless* time marginal we have $\widehat{\theta}_{\mathrm{TI}} \approx 1$, as $y(t) \approx y_{\mathrm{dt}}(t)$. In this case, the EMD returns only the residual $\rho^I(t)$ as an approximation of the trend component $c(t)$. Hence, the detrended time marginal $y_{\mathrm{dt}}(t)$ series is given only by:

$$y_{\mathrm{dt}}(t) = y(t) - c(t) = \sum_{i=1}^{I} m^{(i)}(t) + \rho^I(t) - \rho^I(t) = \sum_{i=1}^{I} m^{(i)}(t) \tag{3.9}$$

For illustration, we could write the quantity that we want to estimate – the importance of the trend given in (3.6) – according to the EMD notation:

$$\theta_{\mathrm{TI}} = \frac{\mathbb{V}\mathrm{ar}\{y(t)\}}{\mathbb{V}\mathrm{ar}\{y_{\mathrm{dt}}(t)\}} = \frac{\mathbb{V}\mathrm{ar}\left\{\sum_{i=1}^{I} m^{(i)}(t) + \rho^{I}(t)\right\}}{\mathbb{V}\mathrm{ar}\left\{\sum_{i=1}^{I} m^{(i)}(t)\right\}}, \tag{3.10}$$

where the slowly-varying residual $\rho^{I}(t)$ shown in the numerator of (3.10) is nearly a constant in comparison to the sum of the IMFs, specially on the absence of trends. Now, if we recall that for a given constant a, we have

$$\mathbb{V}\mathrm{ar}\left\{\sum_{i=1}^{I} m^{(i)}(t) + a\right\} = \mathbb{V}\mathrm{ar}\left\{\sum_{i=1}^{I} m^{(i)}(t)\right\},$$

and it could be easily verified in (3.10) that $\theta_{\mathrm{TI}} \approx 1$. This is the most common case in the bootstrapped data[1].

3.6.2 Trended time marginal

For the case of a *trended* time marginal we have $\widehat{\theta}_{\mathrm{TI}} \geq 1$. Now, the EMD might not only use the residual $\rho^{I}(t)$, but also different IMFs to approximate $c(t)$. Thus, the algorithm returns an index $i^{*} \leq I$ as the one which gives the best approximation of the trend. In this case, the detrended time marginal series $y_{\mathrm{dt}}(t) = y(t) - c(t)$ is given by:

$$y_{\mathrm{dt}}(t) = \sum_{i=1}^{I} m^{(i)}(t) + \rho^{I}(t) - \left[\sum_{i=i^{*}}^{I} m^{(i)}(t) + \rho^{I}(t)\right] = \sum_{i=1}^{i^{*}-1} m^{(i)}(t). \tag{3.11}$$

By using (3.11), we can rewrite (3.7) as:

$$\widehat{\theta}_{\mathrm{TI}} = \frac{\sum_{t=1}^{N}\left[y(t) - \widehat{\mu}_{\mathrm{y}}\right]^{2}}{\sum_{t=1}^{N}\left[y_{\mathrm{dt}}(t) - \widehat{\mu}_{\mathrm{dt}}\right]^{2}} = \frac{\sum_{t=1}^{N}\left[\sum_{i=1}^{I} m^{(i)}(t) + \rho^{I}(t) - \widehat{\mu}_{\mathrm{y}}\right]^{2}}{\sum_{t=1}^{N}\left[\sum_{i=1}^{i^{*}-1} m^{(i)}(t) - \widehat{\mu}_{\mathrm{dt}}\right]^{2}}, \tag{3.12}$$

[1] It should be noted that due to approximation errors in the EMD algorithm, we might obtain values of θ_{TI} that are slightly lower than 1. In anyway, the approximation $\theta_{\mathrm{TI}} \approx 1$ is still reasonable.

where $\widehat{\mu}_{\mathrm{y}}$ and $\widehat{\mu}_{\mathrm{dt}}$ are, respectively, the sample mean of $y(t)$ and $y_{\mathrm{dt}}(t)$ as shown in (3.11). The length of $y(t)$ and $y_{\mathrm{dt}}(t)$ is given by N. We can expand (3.12), thus obtaining:

$$\widehat{\theta}_{\mathrm{TI}} = \frac{\sum_{t=1}^{N}\left[\sum_{i=1}^{I} m^{(i)}(t)\right]^2 + 2\sum_{t=1}^{N}\left[\rho^I(t) - \widehat{\mu}_{\mathrm{y}}\right]\left[\sum_{i=1}^{I} m^{(i)}(t)\right] + \sum_{t=1}^{N}\left[\rho^I(t) - \widehat{\mu}_{\mathrm{y}}\right]^2}{\sum_{t=1}^{N}\left[\sum_{i=1}^{i^*-1} m^{(i)}(t) - \widehat{\mu}_{\mathrm{dt}}\right]^2} \tag{3.13}$$

Now, let us define $\widehat{\mu}_i$ as the sample mean of the i^{th} IMF:

$$\widehat{\mu}_i = \frac{1}{n}\sum_{t=1}^{N} m^{(i)}(t) \text{ for } i = 1, ..., i^*, ..., I. \tag{3.14}$$

If we recall that each IMF is a zero-mean waveform by construction, it becomes straightforward to verify that

$$\begin{aligned} &\widehat{\mu}_i = 0 \text{ for } i = 1, ..., i^*, ..., I \\ &\widehat{\mu}_{\mathrm{dt}} = \widehat{\mu}_1 + \cdots + \widehat{\mu}_{i^*-1} = 0 \\ &\widehat{\mu}_{\mathrm{y}} = \widehat{\mu}_1 + \cdots + \widehat{\mu}_I + \widehat{\mu}_\rho = \widehat{\mu}_\rho, \end{aligned} \tag{3.15}$$

where $\widehat{\mu}_\rho$ is the sample mean of the residual $\rho^I(t)$. We can rewrite (3.13) by considering (3.15), which gives:

$$\widehat{\theta}_{\mathrm{TI}} = \frac{\sum_{t=1}^{N}\left[\sum_{i=1}^{I} m^{(i)}(t)\right]^2 + 2\sum_{t=1}^{N}\left[\rho^I(t) - \widehat{\mu}_\rho\right]\left[\sum_{i=1}^{I} m^{(i)}(t)\right] + \sum_{t=1}^{N}\left[\rho^I(t) - \widehat{\mu}_\rho\right]^2}{\sum_{t=1}^{N}\left[\sum_{i=1}^{i^*-1} m^{(i)}(t)\right]^2}. \tag{3.16}$$

If we expand the squared terms in (3.16), we obtain:

$$\widehat{\theta}_{\mathrm{TI}} = \frac{\sum_{t=1}^{N}\sum_{i=1}^{I}\left[m^{(i)}(t)\right]^2 + 2\sum_{t=1}^{N}\sum_{i<l}^{I} m^{(i)}(t)m^{(l)}(t) + 2\sum_{t=1}^{N}\sum_{i=1}^{I}\left[\rho^I(t) - \widehat{\mu}_\rho\right] m^{(i)}(t) + \sum_{t=1}^{N}\left[\rho^I(t) - \widehat{\mu}_\rho\right]^2}{\sum_{t=1}^{N}\sum_{i=1}^{i^*-1}\left[m^{(i)}(t)\right]^2 + 2\sum_{t=1}^{N}\sum_{i<l}^{i^*-1} m^{(i)}(t)m^{(l)}(t)} \tag{3.17}$$

By noting that $\widehat{\mu}_i = 0$ for the $i = 1, ..., I$ IMFs, one can verify that (3.17) is given as function of the sample variances $\{\widehat{\sigma}_i^2, i = 1, ..., I\}$ and covariances $\{\widehat{\gamma}_{i,l}, i < l, i, l = 1, ..., I\}$ between the IMFs, and the sample variance $\widehat{\sigma}_\rho^2$ and covariances $\{\widehat{\gamma}_{\rho,i}, i = 1, ..., I\}$ between

the residual $\rho^I(t)$ and the IMFs:

$$\begin{aligned}
\sum_{i=1}^{I}\widehat{\sigma}_i^2 &= \frac{1}{N-1}\sum_{i=1}^{I}\sum_{t=1}^{N}\left[m^{(i)}(t)-\widehat{\mu}_i\right]^2 = \frac{1}{N-1}\sum_{t=1}^{N}\sum_{i=1}^{I}\left[m^{(i)}(t)\right]^2\\
\sum_{i<l}^{I}\widehat{\gamma}_{i,l} &= \frac{1}{N-1}\sum_{i<l}^{I}\sum_{t=1}^{N}\left[m^{(i)}(t)-\widehat{\mu}_i\right]\left[m^{(l)}(t)-\widehat{\mu}_l\right] = \frac{1}{N-1}\sum_{t=1}^{N}\sum_{i<l}^{I}m^{(i)}(t)m^{(l)}(t)\\
\sum_{i=1}^{I}\widehat{\gamma}_{\rho,i} &= \frac{1}{N-1}\sum_{i=1}^{I}\sum_{t=1}^{N}\left[\rho^I(t)-\widehat{\mu}_\rho\right]\left[m^{(i)}(t)-\widehat{\mu}_i\right] = \frac{1}{N-1}\sum_{t=1}^{N}\sum_{i=1}^{I}\left[\rho^I(t)-\widehat{\mu}_\rho\right]m^{(i)}(t)\\
\widehat{\sigma}_\rho^2 &= \frac{1}{N-1}\sum_{t=1}^{N}\left[\rho^I(t)-\widehat{\mu}_\rho\right]^2,
\end{aligned} \tag{3.18}$$

Thus, according to (3.18), we can express (3.17) simply as:

$$\widehat{\theta}_{\mathrm{TI}} = \frac{\sum_{i=1}^{I}\widehat{\sigma}_i^2 + 2\sum_{i<l}^{I}\widehat{\gamma}_{i,l} + 2\sum_{i=1}^{I}\widehat{\gamma}_{\rho,i} + \widehat{\sigma}_\rho^2}{\sum_{i=1}^{i^*-1}\widehat{\sigma}_i^2 + 2\sum_{i<l}^{i^*-1}\widehat{\gamma}_{i,l}}, \tag{3.19}$$

where the terms $1/(N-1)$ were canceled out in (3.19) due to the division. In (3.19), the covariance estimate $\widehat{\gamma}_{i,l}$ between consecutive i^{th} and l^{th} IMFs should be close to zero, as for all practical purposes the IMFs can be considered to be locally orthogonal to each other [69]. Hence, the following could be assumed:

$$\sum_{i=1}^{I}\widehat{\sigma}_i^2 \gg 2\sum_{i<l}^{I}\widehat{\gamma}_{i,l} \quad \text{and} \quad \sum_{i=1}^{i^*-1}\widehat{\sigma}_{\rho'}^2 \gg 2\sum_{i<l}^{i^*-1}\widehat{\gamma}_{i,l}.$$

Then, one could approximate (3.19) as:

$$\widehat{\theta}_{\mathrm{TI}} = \frac{\sum_{i=1}^{i^*-1}\widehat{\sigma}_i^2 + \sum_{i=i^*}^{I}\widehat{\sigma}_i^2 + \widehat{\sigma}_\rho^2 + 2\sum_{i=1}^{I}\widehat{\gamma}_{\rho,i}}{\sum_{i=1}^{i^*-1}\widehat{\sigma}_i^2} = 1 + \frac{\sum_{i=i^*}^{I}\widehat{\sigma}_i^2 + \widehat{\sigma}_\rho^2}{\sum_{i=1}^{i^*-1}\widehat{\sigma}_i^2} + 2\frac{\sum_{i=1}^{I}\widehat{\gamma}_{\rho,i}}{\sum_{i=1}^{i^*-1}\widehat{\sigma}_i^2}, \tag{3.20}$$

where the terms corresponding to the trended ($i = i^*, ..., I$) and to the detrended ($i = 1, ..., i^*-1$) counterparts were rearranged. It is simple to verify that (3.20) could be expressed as function of the individual energies of the IMFs and the residual. To do so, one could first define the slightly shifted slowly varying residual $\rho'^I(t) = \rho(t) - \widehat{\mu}_\rho$. Then, by recalling the zero mean property of the IMFs, one can easily verify that:

$$\widehat{\sigma}_i^2 = \sum_{t=1}^{N}\left[m^{(i)}(t)\right]^2 = E_i \quad \text{and} \quad \widehat{\sigma}_i^2 = \sum_{t=1}^{N}\left[\rho'^I(t)\right]^2 = E_{\rho'}, \tag{3.21}$$

where E_i and $E_{\rho'}$ are the energies of the i^{th} IMF and of $\rho'^I(t)$, respectively. Also, one can verify that:

$$\sum_{i=1}^{I} \widehat{\gamma}_{\rho,i} = \sum_{i=1}^{I}\sum_{t=1}^{N} \left[\rho'^I(t) m^{(i)}(t)\right] = \sqrt{E_i}\sqrt{E_{\rho'}}\cos\alpha_i, \tag{3.22}$$

where α_i is the angle of the i^{th} scalar product. The terms $1/(N-1)$ were omitted in (3.21) and (3.22) since they are anyway canceled out in the expression of $\widehat{\theta}_{\text{TI}}$. According to (3.21) and (3.22), we can express (3.20) as:

$$\widehat{\theta}_{\text{TI}} = 1 + \frac{E_{i^*} + \cdots + E_I + E_{\rho'}}{E_1 + \cdots + E_{i^*-1}} + 2\frac{\sqrt{E_1}\sqrt{E_{\rho'}}\cos\alpha_1 + \cdots + \sqrt{E_I}\sqrt{E_{\rho'}}\cos\alpha_I}{E_1 + \cdots + E_{i^*-1}}, \tag{3.23}$$

where, in general, the angles α_i are approximately 90°, and the terms $\cos\alpha_i$ are close to zero due to the aforementioned properties of the IMFs, and to the fact that the residual is nearly constant in comparison to the other IMFs, which capture faster oscillation modes. In this regard, recall that the scalar product terms come from the covariance estimates $\widehat{\gamma}_{i,\rho}$, and in a limit case, for a given constant a, we should have $\mathbb{C}\text{ov}\{m^{(i)}(t), a\} = 0$. In Table 3.1, we present, for different window lengths, the angles α_i (in degrees) obtained in the estimation of θ_{TI} for the nonstationary Gaussian signal shown in Fig. 3.2. The EMD-based trend filtering algorithm has returned $I = 6$ IMFs and the residual. In (3.23), the energies of the IMFs of the trend component ($E_{i^*}, ..., E_I$) and residual ($E_{\rho'}$), are being divided by those of the detrended one ($E_1, ..., E_{i^*-1}$). According to [66], the energies are in general increasing for IMFs with index $i \geq i^*$. Hence, if the energy criterion for choosing i^* holds, the ratio in (3.23) should be maximized, increasing the value of $\widehat{\theta}_{\text{TI}}$, which indicates the presence of a trend.

Table 3.1: Verifying the angles (in degrees) of the scalar product terms in the estimation of θ_{TI} for the nonstationary Gaussian signal shown in Fig. 3.2.

n_h / IMF	α_1	α_2	α_3	α_4	α_5	α_6
$n_h = 3$	89.9°	90.0°	92.0°	88.0°	87.0°	91.1°
$n_h = 5$	93.0°	95.0°	87.0°	87.1°	89.0°	91.1°
$n_h = 7$	90.0°	92.1°	89.2°	88.1°	90.0°	92.3°
$n_h = 9$	89.9°	91.3°	88.5°	88.5°	90.5°	89.5°

3.7 The generalized extreme value distribution

Having analyzed the behavior of the trend importance estimator for the cases of stationarity (trendless time marginal) and nonstationarity (trended time marginal), it is now necessary to derive the statistical distribution of $\widehat{\theta}_{\text{TI}}$ under the null hypothesis for per-

forming the hypothesis test. Drawing conclusions from the crude histogram of $\widehat{\theta}_{\mathrm{TI}}$ is not practical, so we propose to model this distribution of $\widehat{\theta}_{\mathrm{TI}}$ by an appropriate model.

Due to their stationarity, the bootstrap resamples are supposed to return trendless time marginals. Therefore, we should have necessarily $\widehat{\theta}_{\mathrm{TI}} \approx 1$ in most of the cases. Thus, any resample with a significantly trended time marginal could be interpreted as an *extreme event*, very unlikely to be found in such stationary references. Nevertheless, for cases where trended time marginals occur, it is very likely to have large values of $\widehat{\theta}_{\mathrm{TI}}$ due to the EMD methodology. A heuristic argument suggests that the distribution of $\widehat{\theta}_{\mathrm{TI}}$ could be modeled by a heavy tailed pdf with a peak and a lower bound around the unity. These features match those from the Fréchet or GEV type II distribution [77]. This type of pdf is one of the three belonging to the GEV family, which is governed by its shape, scale and location parameters (ϵ, σ and μ, respectively). The type II, or the Fréchet case occurs for $\epsilon > 0$, and its cdf is given in (3.24). This distribution has no upper bound, however, a lower limit is given by its location parameter μ [78], which should be close to the unity in this application.

$$F(x;\epsilon,\sigma,\mu) = \begin{cases} 0 & \text{if } x < \mu \\ \exp\left[-\left(\frac{x-\mu}{\sigma}\right)^{-\epsilon}\right] & \text{if } x \geq \mu. \end{cases} \tag{3.24}$$

The three parameters of the Fréchet distribution are estimated by maximum likelihood from the collection of the values of $\widehat{\theta}_{\mathrm{TI}}$ obtained from $p = 1, ..., P$ (stationary) bootstrap resamples. We represent this collection by:

$$\left\{\Theta_{\mathrm{TI}}(p) = \widehat{\theta}_{\mathrm{TI}}^{(p)}, p = 1, \cdots, P\right\}, \tag{3.25}$$

where the GEV pdf fitted to $\Theta_{\mathrm{TI}}(p)$ allows for the selection of a threshold above which the null hypothesis of trendless time marginal is rejected. An example of the GEV modeling is illustrated in Fig. 3.4, where the importance of the trend was estimated in the time marginals of $P = 1000$ and $P = 100$ bootstrap replicates of a stationary AR(1) process with length $T = 1050$ and pole at $a = 0.8$. The approximated GEV pdf and the histograms are shown in Fig. 3.4(a) for the case of $P = 1000$, and in Fig. 3.4(b) for the case of $P = 100$.

3.7.1 Adherence of the GEV fit: Asymptotic regime analysis

The three parameters of the Fréchet pdf are estimated by maximum likelihood. In this section, we illustrate the results of an empirical study on the behavior of the parameter

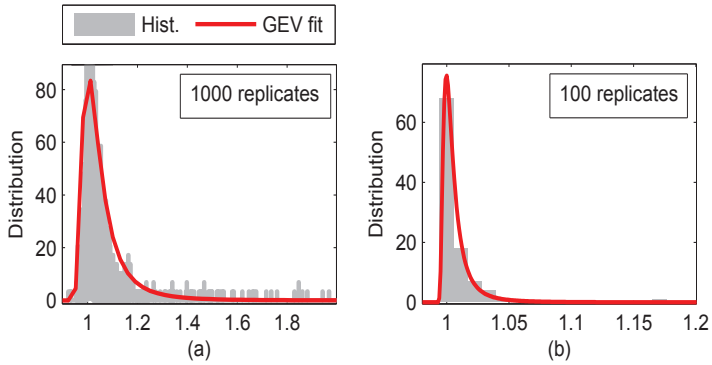

Figure 3.4: Fitting a GEV pdf of type II. (a) $P = 1000$ bootstrap resamples. (b) $P = 100$ bootstrap resamples.

estimators ($\widehat{\epsilon}$, $\widehat{\sigma}$ and $\widehat{\mu}$) as a function of the number of bootstrap replicates. The simulations were conducted using 1000 realizations of an AR(1) process of length $T = 1050$. The results are presented for windows $h_k(t)$ of different sizes $n_\mathrm{h} = (3, 5, 7)$. The mean and the standard deviation of the parameter estimators are shown in Fig. 3.5. The normalized bias ($\widetilde{\mathrm{Bias}}$) and the normalized variance ($\widetilde{\mathrm{Var}}$) are shown in Fig. 3.6. In this work, we make use of (3.26) and (3.27) for computing $\widetilde{\mathrm{Bias}}$ and $\widetilde{\mathrm{Var}}$, respectively. To do so, the expected value of 1000 realizations (approximately in the asymptotic regime) was considered as a reasonable approximation of the real parameter value.

$$\widetilde{\mathrm{Bias}}\left[\widehat{\theta}(n_\mathrm{h}, P)\right]\Big|_{\widehat{\theta}=\widehat{\epsilon},\widehat{\sigma},\widehat{\mu}} = \frac{\mathbb{E}\left\{\widehat{\theta}(n_\mathrm{h}, P)\right\} - \mathbb{E}\left\{\widehat{\theta}(n_\mathrm{h}, 1000)\right\}}{\mathbb{E}\left\{\widehat{\theta}(n_\mathrm{h}, 1000)\right\}} \tag{3.26}$$

$$\widetilde{\mathrm{Var}}\left[\widehat{\theta}(n_\mathrm{h}, P)\right]\Big|_{\widehat{\theta}=\widehat{\epsilon},\widehat{\sigma},\widehat{\mu}} = \frac{\mathbb{E}\left\{\left\{\widehat{\theta}(n_\mathrm{h}, P) - \mathbb{E}\left[\widehat{\theta}(n_\mathrm{h}, P)\right]\right\}^2\right\}}{\left\{\mathbb{E}\left\{\widehat{\theta}(n_\mathrm{h}, 1000)\right\}\right\}^2} \tag{3.27}$$

It can be seen in Fig. 3.5 and Fig. 3.6 that, as the number of bootstrap replicates grows, we have a convergence towards the asymptotic regime for all simulations. This behavior is in favor of the assumption that the distribution of $\widehat{\theta}_\mathrm{TI}$ can be fairly well modeled by a Fréchet pdf. Moreover, it can be seen that the normalized bias and variance are weakly dependent on the length of the window. Finally, notice that by using 100 bootstrap replicates we would have a reasonable precision, and the values taken by (3.26) and (3.27) become inferior to 5% for more than 100 replicates. This reduced number of bootstrap resamples can be considered for saving the computational time, since it leads to an acceptable approximation of the asymptotic distribution [44]. For now on, we shall

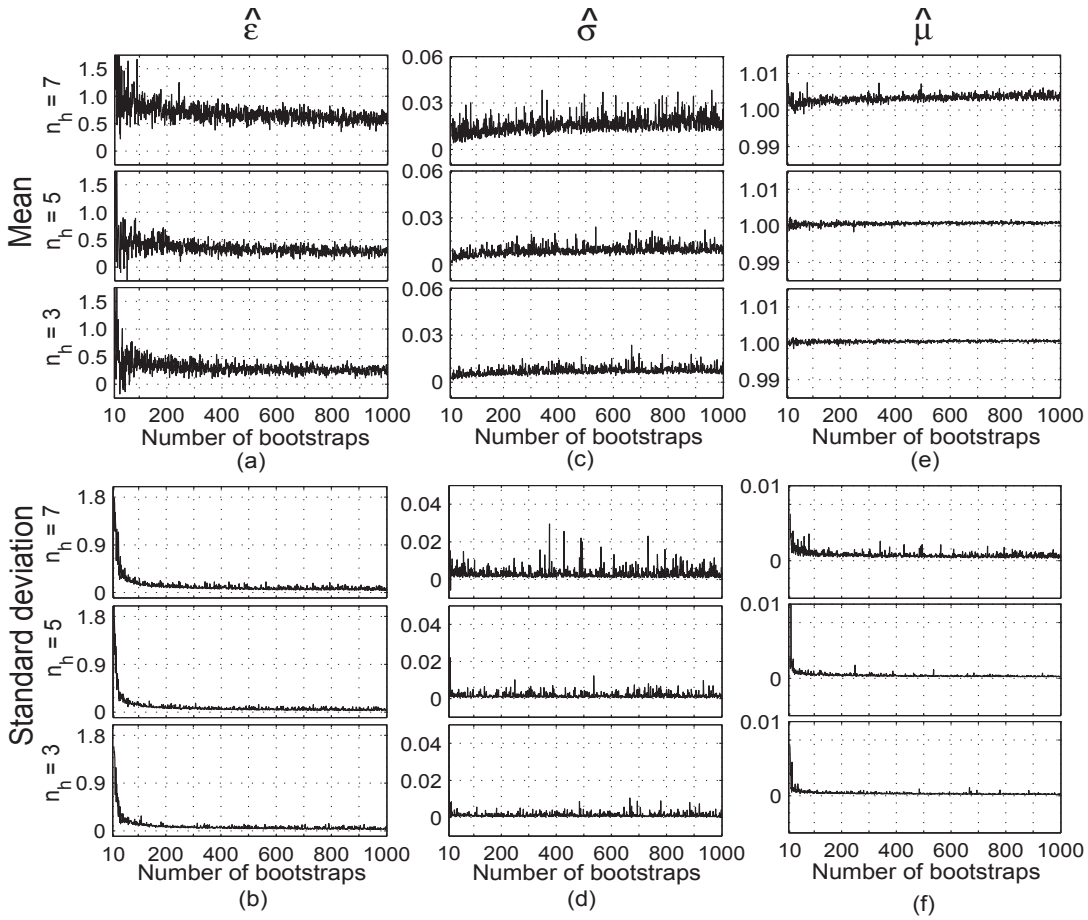

Figure 3.5: Analysis of the parameter estimators as function of the number of bootstrap replicates for windows $h_k(t)$ of different lengths (as indicated in the legend). The mean and standard deviation of the three parameter estimators are shown in (a), (c), (d), and (b), (d), (f), respectively.

use this number of bootstrap resamples to perform the stationarity test.

It is also interesting to confirm if the proposed framework verifies the null hypothesis of stationarity. To do so, we have used Monte Carlo simulations to verify the reproduction of the prescribed false alarm rate. The analysis was conducted on 1000 independent realizations of the stationary AR(1) process used in Fig. 3.6 and Fig. 3.5. We have chosen a threshold of 95% for the test, corresponding to a false alarm rate fixed *a priori* to 5%. The results are shown in Fig. 3.7, as function of the number of bootstrap resamples. It can be seen that the real level of confidence remains aproximately as low as 6.8%, if we use 100 or more bootstrap replicates. We have thus obtained an acceptable false alarm rate, but the proposed method is a little pessimistic, as the rejection of the null hypothesis occurs a little more frequently than necessary.

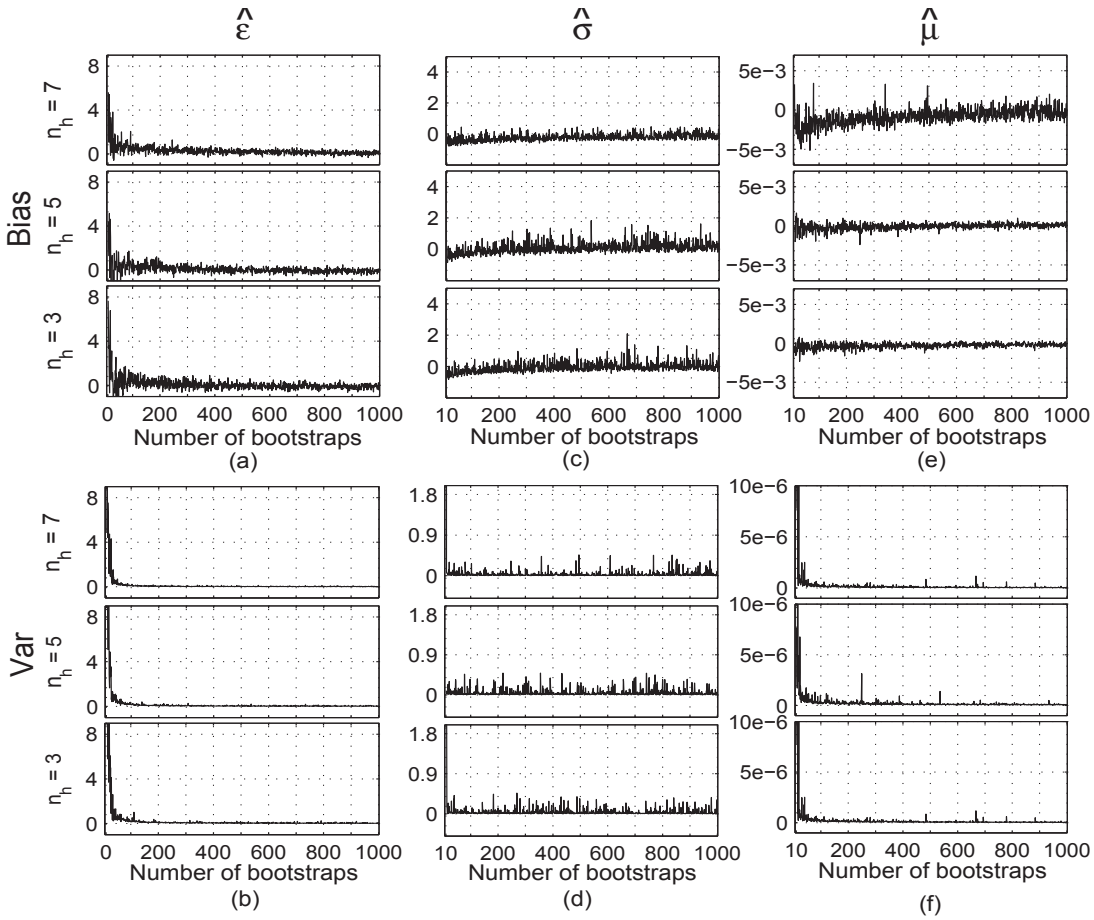

Figure 3.6: Analysis of the parameter estimators as function of the number of bootstrap replicates for windows $h_k(t)$ of different lengths (as indicated in the legend). The normalized bias and normalized variance of the three parameter estimators are shown in (a), (c), (d), and (b), (d), (f), respectively.

3.7.2 Adherence of the GEV fit: the Zempléni test

There are just a few goodness-of-fit tests for GEV distributions available in the literature. A test based on the stability property of the GEV pdf is proposed in [79]. One could also apply the Anderson-Darling test, which is not specifically made for testing GEV distributions. However, a reasonable performance can be nevertheless obtained, since the Anderson-Darling test gives more weight to the tails of the distribution [80]. Still, in many cases, only one of the tails of the distribution is important. This is the case of the Fréchet pdf. In this sense, Zempléni proposed a modification of the Anderson-Darling test that accounts for discrepancies at the *relevant* tail of the distribution [81]. The Zempléni statistic B^2 allows for testing the hypothesis that a cdf $F \in$ GEV against the null hypothesis of F being an arbitrary continuous cdf. The statistic B^2 is computed

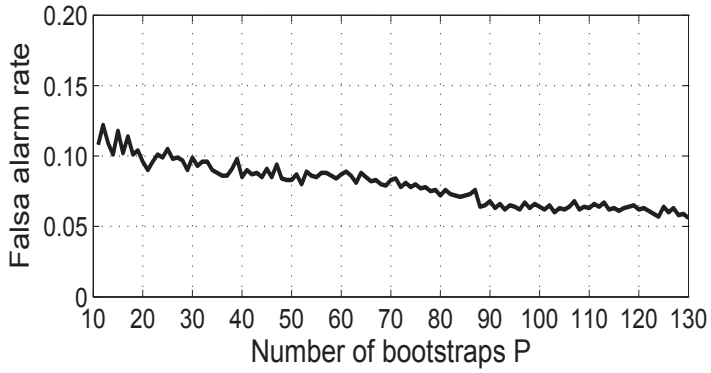

Figure 3.7: Verifying the null hypothesis of stationarity. For a false alarm rate fixed *a priori* as 5% and 1000 realizations of the AR(1) process of Fig. 3.6 and Fig. 3.5, a Monte Carlo estimation of the number of observed "nonstationary" outcomes is performed as function of the number P bootstrap resamples. The obtained level of confidence remains under 6.8% for $P \geq 100$.

as:

$$B^2 = -P/2 - \sum_{i=1}^{P}(2i-1)\log(1 - z_{[P+1-i]})/P - \sum_{i=1}^{P} z_{[i]}/P \tag{3.28}$$

where $z_{[i]} = F(x_{[i]})$ is the cdf evaluated at the i^{th} ordered sample of the collection $\{\Theta_{\text{TI}}(p), p = 1, \cdots, P\}$. The critical values of B^2 for rejecting $F \notin$ GEV are given in [81] for different values of the shape parameter ($\epsilon = -0.6, -0.2, 0.2, 0.5$). The critical values for a significance level of 5% are reproduced in Table 3.2. Notice in Table 3.2 that, in general, the critical values increase with the sample size, but *there is no significant variation* for $\epsilon = 0.2$ or $\epsilon = 0.5$. In Fig. 3.8, we show the histograms of the estimates of ϵ, computed for all different types of stationary and nonstationary signals tested in this Chapter. It can be seen that the values of $\widehat{\epsilon}$ are around 0.5 for all signals. In this regard, notice Fig. 3.5 that the average value $\widehat{\epsilon}$ tends indeed to 0.5 in asymptotic regime. Most importantly, in all cases the average value of the statistic B^2 was much larger than the ones shown in Table 3.2 (as it will be seen in Section 3.9). This supports the assumption that the distribution of $\widehat{\theta}_{\text{TI}}$ can be approximated by a GEV model.

Table 3.2: Critical values of B^2 for different values of ϵ.

ϵ / sample size	25	50	100	200	400
0.2	0.307	0.311	0.317	0.323	0.324
0.5	0.297	0.308	0.312	0.316	0.320
-0.2	0.313	0.320	0.331	0.333	0.340
-0.6	0.812	0.390	0.383	0.385	0.389

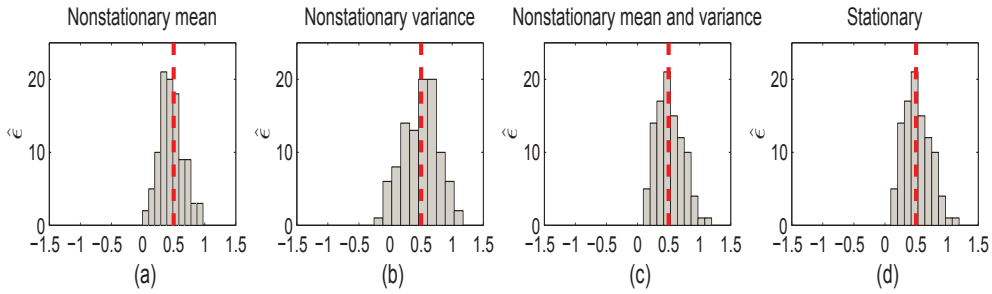

Figure 3.8: Histograms of $\hat{\epsilon}$ computed for all the different types of stationary and nonstationary signals tested in this Chapter. The dashed line in red stands for the 0.5 point. (a) Signals with a nonstationary mean. (b) Signals with a nonstationary variance. (c) Signals with a nonstationary mean and variance. (d) Stationary signals.

3.8 Hypothesis test and index of nonstationarity

It was shown that the Fréchet model approximates fairly well the distribution of the trend contaminations in the stationary references. Now, it is simple to derive a threshold above which the null hypothesis of stationarity is rejected under a given prescribed level. Putting it in the form of a one-sided test, we use as test statistic the estimated importance of the trend in the original signal. The hypothesis test is given in (3.29), where the threshold $\mathcal{T}$ is computed given a false alarm rate of 5%.

$$d(x) = \begin{cases} 1 & \text{if } \widehat{\theta}_{\text{TI}} > \mathcal{T}, \text{ "nonstationary"}, \\ 0 & \text{if } \widehat{\theta}_{\text{TI}} \leq \mathcal{T}, \text{ "stationary"}. \end{cases} \tag{3.29}$$

If the null hypothesis of stationarity is rejected, it is also interesting to derive an *index of nonstationarity*. To do so, we could compute the minimum value of $\widehat{\theta}_{\text{TI}}$ and then compare with the one returned by the method. Due to the fact different IMFs can be used to approximate the trend at each realization of the process, the minimization of $\widehat{\theta}_{\text{TI}}$ should take into account the particular energy profile of the IMFs. Since $\widehat{\theta}_{\text{TI}}$ is given in (3.23) as function of the individual energies of the IMFs, we define the energy vector $\mathbf{e} = [E_1, ..., E_{i^*}, ..., E_I, E_{\rho'}]$ and consider the problem of computing $\min \widehat{\theta}_{\text{TI}}(\mathbf{e})$ given that:

$$\widehat{\theta}_{\text{TI}}(\mathbf{e}) = 1 + \frac{E_{i^*} + \cdots + E_I + E_{\rho'}}{E_1 + \cdots + E_{i^*-1}}. \tag{3.30}$$

where $\min \widehat{\theta}_{\text{TI}}(\mathbf{e})$ could be computed without using the scalar product terms, since in general $\cos \alpha_i \approx 0$. However, the terms $\{\cos \alpha_i, i = 1, ..., I\}$ have been used as arguments to parametrize the objective function. We treat the minimization of $\widehat{\theta}_{\text{TI}}(\mathbf{e})$ as a

constrained optimization problem. There are different algorithms for solving the constrained optimization problem [82]. For the simulations presented in this work we used the Quasi-Newton algorithm. Due to the nature of the EMD, we can define a set equality and inequality constraints that need to be satisfied by the energy terms given in $\mathbf{e}$. We consider that $f(\mathbf{e}) = \widehat{\theta}_{\mathrm{TI}}(\mathbf{e})$, the scalar valued objective function to be minimized. The constrained optimization problem is given as follow:

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{e}) \\ \text{subject to} \quad & g_l(\mathbf{e}) \leq 0,\, l = 1, ..., p_{\mathrm{IN}} \\ & h_l(\mathbf{e}) = 0,\, l = 1, ..., p_{\mathrm{EQ}} \\ & 0 \leq \mathbf{e} \leq E_{\mathrm{T}}, \end{aligned} \tag{3.31}$$

where E_{T} is the total energy of the decomposition (see (3.33)), $\{g_l, l = 1, ..., p_{\mathrm{I}}\}$ and $\{h_l, l = 1, ..., p_{\mathrm{E}}\}$ are the set of p_{I} inequality and p_{E} equality constraints that need to be satisfied, respectively. The inequalities are obtained by sorting the vector $\mathbf{e} = [E_1, ..., E_{i^*}, ..., E_I, E_{\rho'}]$ in the ascending order (which, in general, is already the case for $\mathbf{e}_1, ..., \mathbf{e}_{i^*}$). We then define the inequality considering this energy profile of the representation:

$$g_l(\mathbf{e}) = \mathbf{e}_{l+1} - \mathbf{e}_l, \tag{3.32}$$

where $g_1(\mathbf{e}) = \mathbf{e}_2 - \mathbf{e}_1 \leq 0$, $g_2(\mathbf{e}) = \mathbf{e}_3 - \mathbf{e}_2 \leq 0$, and so on. Clearly, the number of inequality constraints p_{IN} depends on the number of IMFs. On the other hand, we have only $p_{\mathrm{EQ}} = 1$ equality constraint:

$$h_1(\mathbf{e}) = \mathbf{e}_1 + \cdots + e_{I+1} = E_{\mathrm{T}}, \tag{3.33}$$

which states that the sum of $\mathbf{e}$ should be equal to the total energy of the decomposition. The index of nonstationarity $\mathcal{I}_{\mathrm{NS}}$ is given by:

$$\mathcal{I}_{\mathrm{NS}} = \frac{\widehat{\theta}_{\mathrm{TI}}(\mathbf{e})}{\min \widehat{\theta}_{\mathrm{TI}}(\mathbf{e})}. \tag{3.34}$$

In Fig. 3.9, we illustrate the performance of (3.34) in evaluating different nonstationary processes if different window lengths n_{h} are chosen. In Fig. 3.9(a) and (b), we show, respectively, the stationary Gaussian process ($x_1(t)$) and the nonstationary one ($x_2(t)$) that were originally shown in Fig. 3.2. The signal ($x_2(t)$) has a variance that changes slowly from $\sigma_1^2 = 1$ to $\sigma_2^2 = 2$. In Fig. 3.9(c) and (d) we increase the values of σ_2^2 to $\sigma_2^2 = 3$, and then to $\sigma_2^2 = 4$, respectively. It can be observed that $\mathcal{I}_{\mathrm{NS}}$ follows the increase

of the level of nonstationarity.

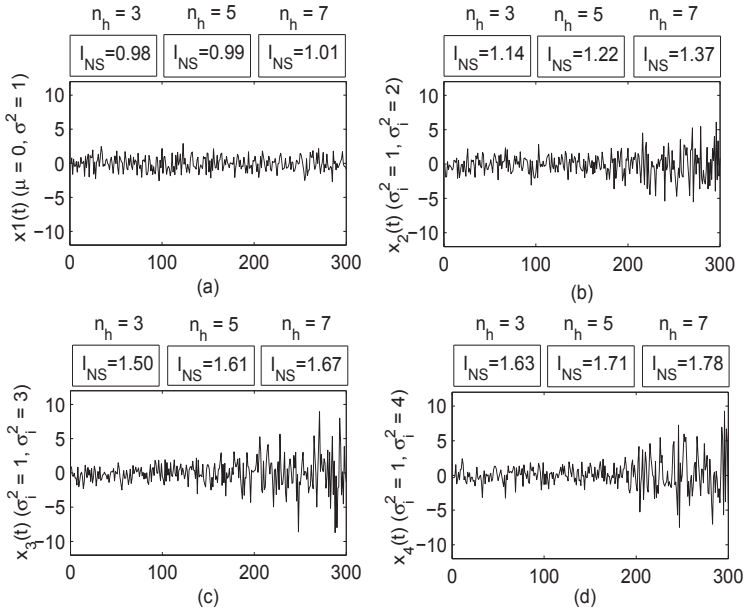

Figure 3.9: Results of using $\mathcal{I}_{\mathrm{NS}}$ to evaluate different nonstationary processes with an increasing transition of the parameters.

The convexity of $f(\mathbf{e})$ is verified automatically by testing the positive semi-definiteness of the Hessian matrix at the solution vector. We have observed that in the vast majority of the cases $f(\mathbf{e})$ is convex. Having explained the methodology of the stationarity test, we shall now present the experimental study. In the following section, we test different nonstationary signals, evaluate the performance of our method and compare with the ones of other approaches.

3.9 Testing the new stationarity test

Different nonstationary signals were generated to assess the performances of the test. We have tested stationary signals and nonstationary ones with a varying mean, varying variance, and varying mean *and* variance, following the *onset-of-trend* model of (2.20). Two configurations were chosen for (2.20): $\kappa = 1$ and $\kappa = T/2$, corresponding to a nonstationarity starting at the beginning or at the middle of the signal with length T.

The test signals are gamma or Gaussian processes of two lengths ($T = 300$ or $T = 1050$). In particular, the gamma processes were chosen because many natural phenomena are modeled by a gamma pdf. Realizations of the synthetic signals with $T = 300$ and $T = 1050$ are shown in Fig. 3.10(a), (b) and Fig. 3.10(c), (d), respectively. Notice that three cases were considered for the Gaussian processes: i) a *fixed variance*

($\sigma^2 = 1$) and a *varying mean* (first row), ii) a *fixed mean* ($\mu = 0$) and a *varying variance* (second row), and iii) a varying mean *and* variance (third row). The gamma signals (fourth row) also have a varying mean and variance, since the shape and scale parameters of the gamma pdf jointly determine the first and second-order moments.

The parameter values for the Gaussian and gamma processes vary according to two different ranges (see (2.20)):

	Gaussian		gamma	
Parameters	Mean	Variance	Shape	Scale
Short range	$\xi_1 = 0$ to $\xi_2 = 3$	$\xi_1 = 1$ to $\xi_2 = 3$	$\xi_1 = 0.355$ to $\xi_2 = 0.532$	$\xi_1 = 19.94$ to $\xi_2 = 29.91$
Long range	$\xi_1 = 0$ to $\xi_2 = 5$	$\xi_1 = 1$ to $\xi_2 = 5$	$\xi_1 = 0.355$ to $\xi_2 = 0.710$	$\xi_1 = 19.94$ to $\xi_2 = 39.88$,

(3.35)

where parameter values were chosen to give a good trade-off between test sensitivity and slowly-varying behavior, in a sense that the resulting nonstationary signals could be hardly detected by others stationarity tests. In (3.35), the gamma values of ξ_1 came from real world applications (*i.e.* gamma models applied to rainfall series [83]). Thus, we are aiming at detecting nonstationarity if a gradual increase of 50% (short range) or 100% (long range) occurs in the parameters of gamma models used in practice.

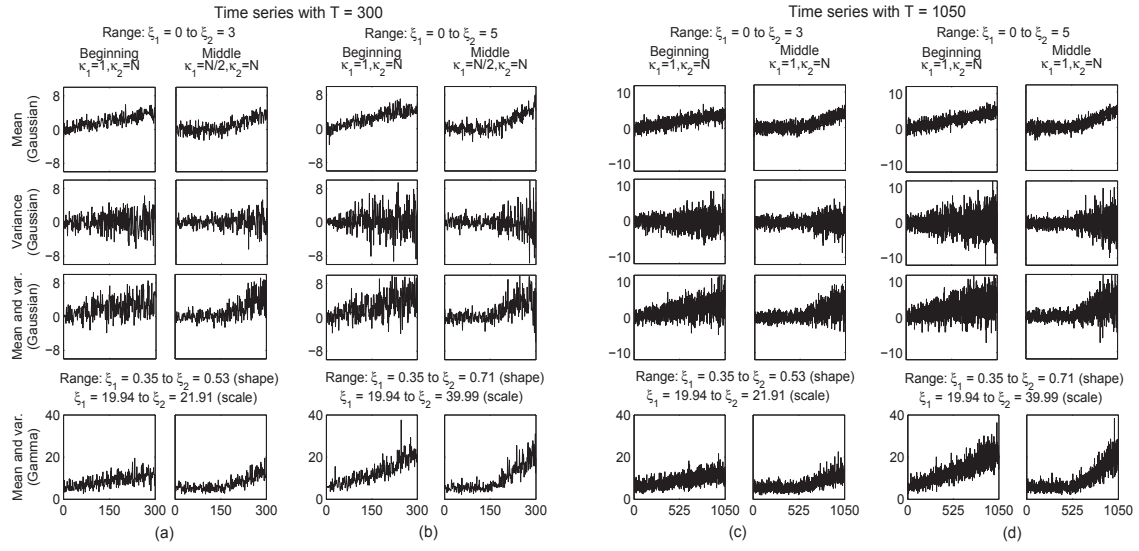

Figure 3.10: Realizations of the synthetic signals following the *onset-of-trend* model for two different ranges shown in (3.35) and two different start points: $\kappa = 1$ and $\kappa = T/2$. Two different lengths ($T = 300$ and $T = 1050$) and two different distributions (Gaussian and gamma) were also considered.

We have compared proposed approach with the nonstationarity detector proposed by S. Kay [20] and the classical KPSS test [27, 26]. These approaches are commonly used for testing trends and slowly-varying nonstationarities. These approaches are briefly

explained below.

3.9.1 Kay's nonstationarity detector

The Kay's approach aims at determining the stationary segment of the signal. The idea is to make use of a time-varying autoregressive model (TVAR) for representing the alternative hypothesis. The TVAR process is given as follows [84]:

$$x[n] = \sum_{i=1}^{p} a_i[n-i]x[n-i] + b[n]w[n], \tag{3.36}$$

where $w[n]$ is white Gaussian noise with unity variance and p is the order of the model. The TVAR parameters in (3.36) are given as:

$$a_i[n] = \sum_{j=0}^{m} a_{i,j} f_j[n] \quad b[n] = \sum_{j=0}^{m} b_j f_j[n] \quad i = 1, ..., p, \tag{3.37}$$

for some collection of basis functions $\{f_0[n], f_1[n], ..., f_m[n]\}$ [20]. A nonstationary process will result whenever the parameters in (3.37) are nonzero for $j = 1, ..., m$, otherwise the TVAR process in (3.36) reduces to a stationary AR process. Then, the Kay's method consists in verifying whether the TVAR parameters in (3.37) are equal to zero.

3.9.2 KPSS test

The KPSS test is applied for testing the null hypothesis that a signal is trend stationary against the alternative that it is a nonstationary unit-root process. For the KPSS test, a given time series $x(t)$ is expressed as the sum of deterministic trend $c_{\mathrm{d}}(t)$, random walk $r_{\mathrm{w}}(t)$, and a stationary error $\epsilon_{\mathrm{s}}(t)$:

$$x(t) = c_{\mathrm{d}}(t) + r_{\mathrm{w}}(t) + \epsilon_{\mathrm{s}}(t),$$

where the test itself is the Lagrange multiplier test of the hypothesis that the random walk has zero variance.

3.9.3 Experimental study

We have tested 1000 realizations of a stationary AR(1) process, a white Gaussian noise WGN(0,1), and the nonstationary signals illustrated in Fig. 3.10. The results are shown in Table 3.3, where "New", "KPSS" and "Kay" correspond to the stationarity test proposed in this work, the KPSS test, and the Kay's nonstationarity detector, respectively.

The experimental study has been carried out with a significance level of 5%. For the proposed approach, we have considered four different windows $n_{\mathrm{h}} = \{3, 5, 7, 9\}$. The results are given as percentage of observed "nonstationarity" outcomes. We also give the average value of $\mathcal{I}_{\mathrm{NS}}$ computed by means of (3.34). In Table 3.4, we show the average values of the Zempléni statistic B^2 and the GEV shape parameter (ϵ), computed for all realizations and different windows. Note that, in accordance to what was shown in Section 3.7.2, the value of $\overline{\epsilon}$ fluctuates around 0.5, and the averages of the Zempléni statistic are *much larger* than the critical values shown in Table 3.2, which indicates the adherence of the GEV model.

It can be seen in Table 3.3 that the method proposed in this work ("New") allows for a better detection of different slowly-varying nonstationarities, as we have achieved an overall higher classification accuracy of nonstationary signals following the *onset-of-trend* model (which is a variation of the *smooth-change* model). In particular, notice that the proposed technique is able to detect first and second-order nonstationarities. This is not the case for the traditional approaches. For example, the nonstationarity detector of [20] ("Kay") performed poorly in detecting a nonstationary variance for different configurations of the *onset-of-trend* model, and simply could not reject the stationarity of the nonstationary gamma processes. The classical KPSS test also could not detect properly signals with nonstationary exclusively in the variance, and those signals with nonstationarities that start at the beginning of the observation interval. Therefore, we can see that the global performance of the proposed stationarity test outperforms those of the traditional tests used in the literature for detecting slowly-varying nonstationarities.

Finally, a trade-off concerning the size of the window n_{h} could be observed, specially for the short time series ($T = 300$). Notice that, in general, the shorter the window, the better the detection of nonstationary signals and vice versa. The overall performance in detecting stationary signals also becomes poorer for shorter windows. As mentioned before, we have to consider that there is no *a priori* value for n_{h}, but allowing it to vary is an extra degree of freedom of the proposed method.

3.10 Conclusions

In this Chapter we have proposed a new stationarity test which has been designed to detect the presence of a trend or an evolution of the local energy of the signal. The assessment of this kind of nonstationarity is an important issue in many areas of environmental science. Although, some contributions that have been made to the surrogate approach could improve the detection of trend-based nonstationarities, the overall per-

Table 3.3: Results of applying the proposed stationarity test ("New"), the KPSS test ("KPSS"), and the nonstationarity detector of [20] ("Kay"). The outcomes are given as percentage "nonstationary" outcomes observed over 1000 realizations of the process shown in Fig. 3.10. The average value of $\mathcal{I}_{\text{NS}}$ over all realizations is also shown.

Nonstationary signals										
Length			$T = 300$				$T = 1050$			
Type			Beginning	Middle	Beginning	Middle	Beginning	Middle	Beginning	Middle
Range			Short		Long		Short		Long	
Mean Gaussian	KPSS		7.50%	100%	5.50%	100%	7.50%	100%	5.50%	100%
	Kay		98.0%	99.0%	87.0%	100%	98.0%	99.0%	87.0%	100%
	New	$n_{\text{h}} = 3$	**99.0%**, 1.00	**90.0%**, 1.14	**95.0%**, 1.10	**85.0%**, 1.34	**100%**, 1.10	**100%**, 1.25	**100%**, 1.22	**92.0%**, 1.45
		$n_{\text{h}} = 5$	**100%**, 1.12	**85.0%**, 1.35	**100%**, 1.28	**87.0%**, 1.51	**100%**, 1.45	**98.0%**, 1.50	**100%**, 1.45	**95.0%**, 1.77
		$n_{\text{h}} = 7$	**65.0%**, 1.00	**85.0%**, 1.01	**94.0%**, 1.00	**84.0%**, 1.14	**95.0%**, 0.99	**98.0%**, 1.09	**100%**, 1.10	**97.0%**, 1.28
		$n_{\text{h}} = 9$	**61.0%**, 0.99	**67.0%**, 1.12	**90.0%**, 1.05	**88.0%**, 1.31	**97.0%**, 1.06	**98.5%**, 1.27	**99.0%**, 1.22	**99.0%**, 1.58
Variance Gaussian	KPSS		5.50%	7.00%	7.00%	8.50%	6.00%	5.00%	6.00%	4.50%
	Kay		12.0%	21.5%	12.5%	20.5%	15.5%	27.0%	12.5%	30.0%
	New	$n_{\text{h}} = 3$	**90.0%**, 1.06	**67.5%**, 1.12	**80.0%**, 1.07	**55.0%**, 1.14	**100%**, 1.08	**80.0%**, 1.24	**100%**, 1.17	**80.0%**, 1.19
		$n_{\text{h}} = 5$	**65.0%**, 1.02	**45.0%**, 1.07	**65.0%**, 1.04	**25.0%**, 1.13	**80.0%**, 1.07	**34.5%**, 1.14	**81.0%**, 1.09	**45.0%**, 1.16
		$n_{\text{h}} = 7$	**25.0%**, 1.00	**15.0%**, 1.01	**25.0%**, 1.04	**15.0%**, 1.05	**40.5%**, 1.05	**10.0%**, 1.00	**52.0%**, 1.10	**20.0%**, 1.13
		$n_{\text{h}} = 9$	**41.0%**, 0.99	**26.0%**, 1.00	**42.0%**, 1.01	**25.5%**, 1.03	**52.0%**, 1.05	**11.0%**, 1.06	**55.0%**, 1.00	**15.0%**, 1.00
Mean and var. Gaussian	KPSS		6.50%	100%	6.50%	100%	7.00%	100%	5.50%	100%
	Kay		73.5%	73.0%	55.5%	58.5%	100%	100%	100%	100%
	New	$n_{\text{h}} = 3$	**70.0%**, 1.14	**65.0%**, 1.24	**75.0%**, 1.20	**69.5%**, 1.22	**95.0%**, 1.15	**54.0%**, 1.16	**100%**, 1.33	**82.5%**, 1.55
		$n_{\text{h}} = 5$	**69.5%**, 1.09	**63.0%**, 1.08	**72.0%**, 1.12	**65.0%**, 1.15	**96.5%**, 1.13	**94.0%**, 1.23	**98.5%**, 1.28	**100%**, 1.45
		$n_{\text{h}} = 7$	**62.5%**, 1.07	**45.5%**, 1.08	**63.0%**, 1.12	**41.0%**, 1.16	**85.0%**, 1.09	**80.5%**, 1.09	**85.0%**, 1.16	**75.3%**, 1.22
		$n_{\text{h}} = 9$	**33.0%**, 1.01	**34.5%**, 1.03	**47.0%**, 1.05	**53.5%**, 1.04	**45.0%**, 1.03	**78.5%**, 1.02	**76.5%**, 1.09	**77.5%**, 1.10
Mean and var. Gamma	KPSS		11.5%	100%	59.0%	100%	29.5%	100%	98.0%	100%
	Kay		0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	5.00%
	New	$n_{\text{h}} = 3$	**79.0%**, 1.05	**54.0%**, 1.09	**82.0%**, 1.03	**23.0%**, 1.02	**100%**, 1.11	**20.0%**, 1.02	**100%**, 1.14	**5.00%**, 1.04
		$n_{\text{h}} = 5$	**79.0%**, 1.14	**21.0%**, 1.03	**92.5%**, 1.05	**21.5%**, 1.16	**100%**, 1.00	**10.0%**, 1.05	**100%**, 1.15	**5.00%**, 1.01
		$n_{\text{h}} = 7$	**55.5%**, 1.06	**17.0%**, 1.04	**63.0%**, 1.09	**15.5%**, 1.00	**75.0%**, 1.20	**15.0%**, 0.99	**95.0%**, 1.13	**5.00%**, 1.03
		$n_{\text{h}} = 9$	**47.0%**, 1.18	**12.5%**, 1.01	**54.0%**, 1.08	**10.0%**, 1.00	**65.0%**, 1.05	**20.0%**, 1.09	**100%**, 1.16	**3.00%**, 0.99

Stationary signals						
Length			$T = 300$		$T = 1050$	
Type			AR(1)	WGN(0,1)	AR(1)	WGN(0,1)
Stationary WGN and AR(1)	KPSS		3.00%	4.50%	6.00%	7.00%
	Kay		5.80%	1.00%	8.00%	1.00%
	New	$n_{\text{h}} = 3$	**10.0%**, 1.05	**8.00%**, 1.09	**9.00%**, 0.98	**10.0%**, 1.00
		$n_{\text{h}} = 5$	**9.00%**, 1.09	**8.00%**, 1.07	**7.00%**, 0.99	**6.00%**, 0.98
		$n_{\text{h}} = 7$	**7.00%**, 0.98	**5.00%**, 0.97	**7.00%**, 1.01	**5.00%**, 0.99
		$n_{\text{h}} = 9$	**5.00%**, 1.00	**3.00%**, 0.97	**4.00%**, 0.99	**4.00%**, 1.01

formance was not satisfactory. Thus, the stationarity test that has been developed in this Chapter appears as an useful alternative to that application. This new technique has shown to be more sensitive to first and second-order nonstationarities. Many techniques devoted to test for trend-based nonstationarities are parametric [20, 11], hence they are not convenient to be applied to real world time series, as the performance of a parame-

Table 3.4: Mean values of the GEV shape parameter ϵ and the Zempleni statistic B^2 over 1000 realizations of the processes shown in Fig. 3.10.

Length	$T = 300$								$T = 1050$							
Type	Beginning		Middle		Beginning		Middle		Beginning		Middle		Beginning		Middle	
Range	Short				Long				Short				Long			
Average values	$\overline{\epsilon}$	$\overline{B}^2$	$\overline{\epsilon}$	$\overline{B}^2$	$\overline{\epsilon}$	$\overline{B}^2$	$\overline{\epsilon}$	$\overline{B}^2$	$\overline{\epsilon}$	$\overline{B}^2$	$\overline{\epsilon}$	$\overline{B}^2$	$\overline{\epsilon}$	$\overline{B}^2$	$\overline{\epsilon}$	$\overline{B}^2$
Mean ($\mathcal{N}$)	0.40	97.3	0.54	96.9	0.41	97.1	0.51	97.1	0.53	97.2	0.80	96.7	0.55	97.3	0.60	96.1
Var. ($\mathcal{N}$)	0.56	97.4	0.51	97.5	0.49	97.4	0.47	98.0	0.57	97.3	0.52	97.6	0.50	98.1	0.49	96.8
Mean and var. ($\mathcal{N}$)	0.59	96.5	0.44	95.2	0.53	97.7	0.55	97.1	0.48	97.2	0.51	97.6	0.51	95.9	0.58	98.8
Mean and var. (Γ)	0.53	89.3	0.49	92.5	0.48	93.7	0.51	91.1	0.47	93.4	0.53	96.8	0.53	88.9	0.55	91.3

tric method ultimately depends on the accuracy of the chosen model. The method that has been designed in this Chapter, on the other hand, is nonparametric and has shown to be more versatile than other approaches, while keeping a reasonable detection rate.

For the proposed approach, we have skipped the time consuming part of computing full TF representations, by directly estimating the time marginal and testing for trend contamination in its structure. The idea is that a stationary signal should not exhibit a structured pattern in the time marginal, like a trend. The trend itself has been estimated by means of the EMD, and the importance of the trend has been estimated by using the so-called trend importance estimator ($\widehat{\theta}_{\mathrm{TI}}$). It has been shown that, due to the properties of the IMFs, $\widehat{\theta}_{\mathrm{TI}}$ can be expressed solely as fraction of the energies of the IMFs. Moreover, the expression for $\widehat{\theta}_{\mathrm{TI}}$ should be maximized in case of trend. In order to obtain stationary references of the signal to perform the hypothesis test, we have proposed to use block bootstrapping, which allows for the conservation of the correlation structure (if any) within a block. The distribution of $\widehat{\theta}_{\mathrm{TI}}$ under the null hypothesis of stationarity has been approximated by a GEV model, for which adherence has been verified by performing a goodness-of-fit test and an analysis in asymptotic regime. Also, an index of nonstationarity ($\mathcal{I}_{\mathrm{NS}}$) has been proposed.

We have applied the new stationarity test to a variety of nonstationary signals and also to stationary ones. Although some parametric methods have presented a higher classification accuracy for first-order nonstationarity, the new technique has presented an overall better performance regarding all types of nonstationarities tested, being globally more sensitive to for first and second-order evolutions. Such generality is exactly what is searched for a test suited to real world applications. One could regard the method that has been designed in this Chapter as a complement to the surrogate-based method of Chapter 2, as the latter performs better in detecting second-order nonstationarities, and the former in detecting first-order nonstationarities. Also, both techniques test for

stationarity relative to the whole observation interval T. Thus, the information about the time of change is unattainable if the nonstationarity appears only for a portion of the series, for instance. For estimating the points of change of the nonstationary signals, we propose in the next Chapter a nonparametric and data-driven framework, which not only allows for the estimation of multiple change points, but also offers a visual representation of the change dynamics of the signal.

Chapter 4

Change point detection

Contents

In this chapter we present a framework for change point detection in nonstationary signals that is based on a recent technique known as the robust singular spectrum transform (RSST). The rationale behind the RSST is to use principal component analysis (PCA) to quantify the anomaly between past and future patterns of the signal around a given time instant, by considering only two fixed input variables that act like windows of analysis. In this work, we rather propose to let these windows vary, as we consider that particular change patterns are more likely to be captured by using many different windows of analysis. We consider that the uncertainty associated with the change patterns captured by different windows should increase around the change point. The estimation

of the change point is carried out by measuring the conditional entropy (conditioned on the past values) of the observed change scores in time. The proposed method is fully data-driven, hence better suited to be applied for real world signals.

4.1 Introduction

The change point (CP) detection is considered to be a basic step in the frame of non-stationary signal analysis. Although the stationarity property can be violated in many ways, one is often interested in changes that occur either in the mean or in the variance (second-order nonstationarity) [85]. These kinds of changes are of particular interest in the analysis of natural phenomena (*e.g.* climate change [86]), as many real world time series can undergo first and second-order changes over time that reflect the impact of human activity on nature [5].

Methods for detecting CP in real world signals should ideally satisfy some essential points. More specifically, the technique should allow for the detection of first and second-order changes, should be nonparametric and data adaptive. Moreover, we expect the method to be robust against measurement noise [87]. Unfortunately, as pointed in [88], many of the available CP discovery methods either detect only one type of change (*e.g.* CUSUM for changes in the mean [25]), assume a parametric underlying model (*e.g.* autoregressive model [89]), or need ad-hoc tuning for each time series (*e.g.* wavelet analysis [90]).

Several years ago, an algorithm for CP detection suited for real world applications was proposed in [91]. The method is called singular spectrum transform (SST) and consists in using principal component analysis (PCA) to measure the anomaly between past and future patterns of the signal around a given time instant. If we take the whole observation scale, the result will be a new time series representing the CP score at every instant. Different from other conventional techniques, the SST does not assume a generative model and can detect different kinds of changes while being data adaptative. Although these features make the SST better suited for real world applications, its actual use remains limited, as the traditional SST is not robust against noise and depends on five different parameters. In order to solve these problems, different modifications to the original SST algorithm have been proposed. In [87], the conventional SST has been extended to the multivariate SST, which main advantage is to improve the robustness against noise. However, the selection of various parameters is still needed [92]. In order to alleviate this problem, the robust singular spectrum transform (RSST) has been proposed by [88]. This approach is more robust against noise and depends only on two

parameters. In this Chapter we propose an alternative method for CP detection which is based on the RSST. We show that, due to the properties of the RSST, we can derive a framework that is fully data-driven, almost parameter-free, and sensitive to both changes in the mean and the variance.

4.1.1 Detailing the contributions of this chapter

The improvements brought by the RSST allow us to have a richer understanding about the change dynamics of the signal. The contributions of this Chapter to the framework of CP detection can be listed as follows:

- The RSST depends on two parameters, for which slight changes affect significantly the outcomes of the algorithm. In Section 4.3.1, we propose to represent the output of the RSST in the space spanned by its two parameters, as different change patterns can be captured if we let the two parameters vary. Further, in Section 4.3.2, we propose a stopping criterion for sweeping over that two parameters, which is a crucial point for reducing the computational time of the algorithm.

- In Section 4.3.3, we present a way for selecting significant change patterns in the signal. The proposed strategy works well for filtering the most significant CP scores (the ones representing actual changes) from noise

- For estimating the CP, we analyze the intervals in the space spanned by the two parameters of the RSST that might encode major changes. More specifically, in Section 4.4, we propose to use the measure of the uncertainty related to the CP scores in time for estimating the CP. We propose to compute the conditional entropy at each time instant, and to calculate the CP by searching for the instants where the local *maxima* of the conditional entropy vector occur.

4.2 SST and RSST, background elements

The main idea of the SST is to compute, for a given time t, an anomaly metric between the past and the future representative patterns of the time series $\{x(t), t = 1, ..., T\}$. The RSST encodes those past and future patterns in the subsequences $\mathbf{b}$ and $\mathbf{f}$ of length w, respectively:

$$\mathbf{b}(t) = [x(t-w), ..., x(t-1)]^T \quad \text{and} \quad \mathbf{f}(t) = [x(t), ..., x(t+w-1)]^T \tag{4.1}$$

The RSST uses (4.1) for constructing two trajectory matrices corresponding to the past and future of the signal, respectively:

$$B(t) = \begin{bmatrix} x(t-n-w+1) & x(t-n-w+2) & \cdots & x(t-w) \\ \vdots & \vdots & \ddots & \vdots \\ x(t-n) & x(t-n+1) & \cdots & x(t-1) \end{bmatrix}, \tag{4.2}$$

$$F(t) = \begin{bmatrix} x(t) & x(t+1) & \cdots & x(t+n-1) \\ \vdots & \vdots & \ddots & \vdots \\ x(t+w-1) & x(t+w) & \cdots & x(t+n+w-2) \end{bmatrix}, \tag{4.3}$$

where n controls how deep we look into the past and future of the analyzed time series. The matrices $B(t)$ and $F(t)$ represent, respectively, the dynamics of the points before and after the instant t, and can be considered to contain various change patterns within the range of w points [91]. Having constructed the trajectory (or Hankel) matrices, we apply singular value decomposition (SVD) to (4.2) and (4.3) to obtain the left singular vectors and the corresponding squared singular values, which are sorted in a descending order of magnitude. The notations used for the SVD outputs of $B(t)$ and $F(t)$ are, respectively, the following:

$$\begin{aligned} &l_b \text{ and } l_f : \text{number of singular values used in the decomposition,} \\ &\{\lambda_l^{[B(t)]}, l = 1, ..., l_b\} \text{ and } \{\lambda_l^{[F(t)]}, l = 1, ..., l_f\} : \text{squared singular values,} \\ &\{\mathbf{u}_l^{[B(t)]}, l = 1, ..., l_b\} \text{ and } \{\mathbf{u}_l^{[F(t)]}, l = 1, ..., l_f\} : \text{left singular vectors.} \end{aligned} \tag{4.4}$$

Larger singular values are related to dominant patterns, while smaller ones are related to noise [91]. We use the left singular vectors to define the *representative patterns*. The procedure to extract pattern by applying SVD to a Hankel matrix is known as singular

spectrum analysis (SSA). The number of singular values corresponds to the number of representative patterns to be used in the analysis, which is an important point of the method. The traditional SST considers $l_f = 1$ and fixes the parameter l_b to a value chosen by the user. The RSST, however, computes both l_b and l_f automatically according to the procedure described in [88]. This automatic computation of l_b and l_f not only reduces the noise in the final results, but also the number of input parameters of the SST.

The key-point of the methodology is that, if the dynamics of the signal do not change, then the l_f representative patterns of the future of the signal should be similar to the l_b representative patterns of the past, in a sense that the l^{th} singular vector $\mathbf{u}_l^{[F(t)]}$ should lie close to hyperplane $U^{[B(t)]}$ spanned by the vectors $\{\mathbf{u}_l^{[B(t)]}, l = 1, ..., l_b\}$. To quantify how far a given representative pattern is from the hyperplane, the RSST computes for each $\{\mathbf{u}_l^{[F(t)]}, l = 1, ..., l_f\}$ the normalized projection onto $U^{[B(t)]}$ as:

$$\alpha_l(t) = \frac{U^{[B(t)]}\mathbf{u}_l^{[F(t)]}}{\left\| U^{[B(t)]}\mathbf{u}_l^{[F(t)]} \right\|}, l \leq l_f. \tag{4.5}$$

By using (4.5), the RSST computes the change score at point t as the following weighted sum:

$$z(t) = \frac{\sum_{l=1}^{l_f} \lambda_l \times (1 - \alpha_l(t)^T \mathbf{u}_l^{[F(t)]})}{\sum_{l=1}^{l_f} \lambda_l}, \tag{4.6}$$

where the weights λ_l are the eigen value of the matrix $F(t)$. The output of the algorithm is given by $z(t)$, which varies from zero to one and represents the final CP score [93].

4.3 Modifying the RSST algorithm

4.3.1 Representing the CP scores in a different space

Since the development of the traditional SST [91], the different SST-based algorithms [88, 87] have shared a common point in their methodology: the fact that the output of the algorithm needs to be a time series of CP scores. Hence, the SST-based methods simply provide the nonlinear transformation $\mathcal{T} : x(t) \rightarrow z(t)$. In this work, we propose to represent the output of SST-based methods in a *different space*. We have noted that the improvements brought by the RSST allow for the development of the modified transform $\mathcal{T} : x(t) \rightarrow z(t, w, n)$, where the change scores are represented in an augmented space spanned by the variables t (time), w (length of the subsequence of the signal) and n (how deep we look into the past and future of the signal). Notice that w is the number of rows and n is the number of columns of the trajectory matrices given in (4.2) and (4.3).

Since the RSST requires the specification of only two parameters (w and n), it is possible to compute $z(t)$ for a *large collection* of w and n values at every instant t. The proposed procedure is illustrated in Fig. 4.1. Notice that, for a given signal $\{x(t), t = 1, ..., T\}$, the CP scores are computed via RSST at each value of w and n for $t = 1, ..., N$, where $N \leq T$ (usually $N = T - 1$). Doing so, we can represent different change patterns in the space spanned by t, w and n.

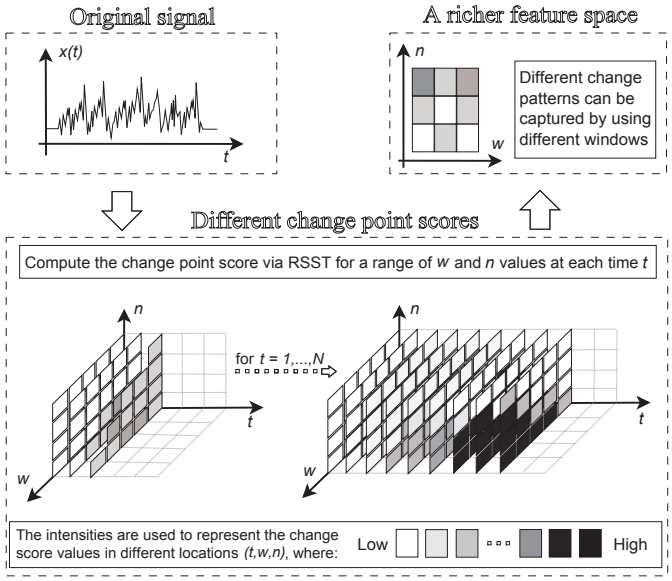

Figure 4.1: Representation of the output of RSST in an augmented space. Different change patterns are captured by varying w and n.

4.3.2 Defining a stopping criterion for the computation of the CP scores

Ideally, one should use as different values of w and n as possible for capturing the largest possible range of $z(t, w, n)$ scores for estimating the CP. The only (and obvious) limitation is that $w \leq T$ and $n \leq T$, where T is the length of the signal. However, regardless of the gain in computational time brought by the RSST, performing this kind of exhaustive search becomes time consuming for larger signals or wider windows of analysis. Thus, to reduce the time required for computing the collection of $z(t, w, n)$ scores, we have developed a stopping criterion for sweeping over t, w and n. More specifically, we have observed that, in general, the total number of scores captured by the windows $w \times n$ *decreases* after a given w and n. In this regard, we propose the following algorithm:

1. For the i^{th} iteration, define $h_i = [1, ..., w_0 + i; 1, ..., n_0 + i]$ as the range of possible w and n values to be used for the analysis, where the default is $w_0 = n_0 = 0$.

2. Compute $z(t, w, n)$ for $t = 1, ..., N$, by using every value of $w = 1, ..., w_0 + i$ and $n = 1, ..., n_0 + i$ given h_i, that has not been yet considered for the previous ranges $h_{i-1}, ..., h_1$.

3. Compute the proportion p_i of CP scores for all t, w and n over the area spanned by $(w_0 + i) \times (n_0 + i)$, *i.e.*:

$$p_i = \frac{1}{(w_0 + i) \times (n_0 + i)} \sum_{t=1}^{N} \sum_{w=1}^{w_0+i} \sum_{n=1}^{n_0+i} z(t, w, n).$$

4. Expand the range h_i to $h_{i+1} = [1, ..., w_0 + i + 1; 1, ..., n_0 + i + 1]$ and evaluate steps 2 and 3 again. If $p_{i+1} < p_i$, set $\zeta = i$ and go to the next step.

5. Set the final collection of scores that shall be used for estimating the CP as:

$$\mathcal{Z} = \{z(t, w, n), t = 1, ..., N, w = 1, ..., w_0 + \zeta, n = 1, ..., n_0 + \zeta\} \tag{4.7}$$

   and stop the algorithm.

The procedure starts at the first iteration ($i = 1$), and goes until $i = \zeta$, when the algorithm stops because p_ζ was achieved. We propose this algorithm because we have observed that, in practically all studied cases, p_i increases monotonically with the size of the window only until a given $i = \zeta$. For $i > \zeta$, p_i starts to decrease monotonically and the influence of the new scores added on the CP estimation starts to be not relevant[1]. Hence,

[1] In practice, however, the algorithm performs a short investigation for some values $i > \zeta$. We can be a little more certain that p_ζ is not a local *maximum* by allowing the procedure to continue for a few more iterations.

ζ is the point that gives the *maximum* proportion of CP scores over the area spanned by w and n, and we define the ζ^{th} iteration as a stop point for the computation of the chart shown in Fig. 4.1. Note that $\mathcal{Z}$ is an array of dimension $N \times (w_0 + \zeta) \times (n_0 + \zeta)$. Finally, one could set other values for w_0 and n_0, as long as $0 \geq w_0 \geq N$ and $0 \geq n_0 \geq N$ are verified.

It should be remarked that the efficiency of the procedure above has been only verified empirically, and a proper theoretical basis can be developed in the future. One could suggest to increase w and n not evenly at each iteration, but rather verifying if p_ζ could be reached through other ways, *e.g.* by using distinct values of w and n. However, the proposed procedure aims at being as simple as possible to fasten the computation of the CP, while keeping the choice of w and n as general as possible. The algorithm above has shown to be suitable for this end.

In Fig. 4.2 we illustrate the application of the proposed algorithm for different types of signals. In Fig. 4.2(a),(b),(c) and (d), we test nonstationary signals with a length $T = 150$ that have a varying mean and variance following the *onset-of-trend* and *abrupt-change* models (see (2.19) and (2.20)) with a CP at $t = T/2$. The signals with a varying mean have $\xi_1 = 0$ and $\xi_2 = 3$. The signals with a varying variance have $\xi_1 = 1$ and $\xi_2 = 3$. In Fig. 4.2(e) and (f), we test two real world time series, more specifically, rainfall time series (see Chapter 5 for data description). For each case, we show the original time series, p_i as function of the iteration i and the point ζ that leads to the *maximum*, and the different ranges h_i from $[1, ..., 10; 1, ..., 10]$ to $[1, ..., 40; 1, ..., 40]$. Notice that for larger windows the proportion of the CP scores over the area spanned by w and n does not tend to increase. It can be seen that p_i grows monotonically until $i = \zeta$ for all time series. Having defined an automatic procedure for computing the collection of CP scores, it can be noted that the proposed algorithm is better suited for real world applications, where the common strategy for choosing the SST parameters (visualization and domain knowledge [94]) are not always possible. In Section 4.3.3, we present the criterion to estimate the CP.

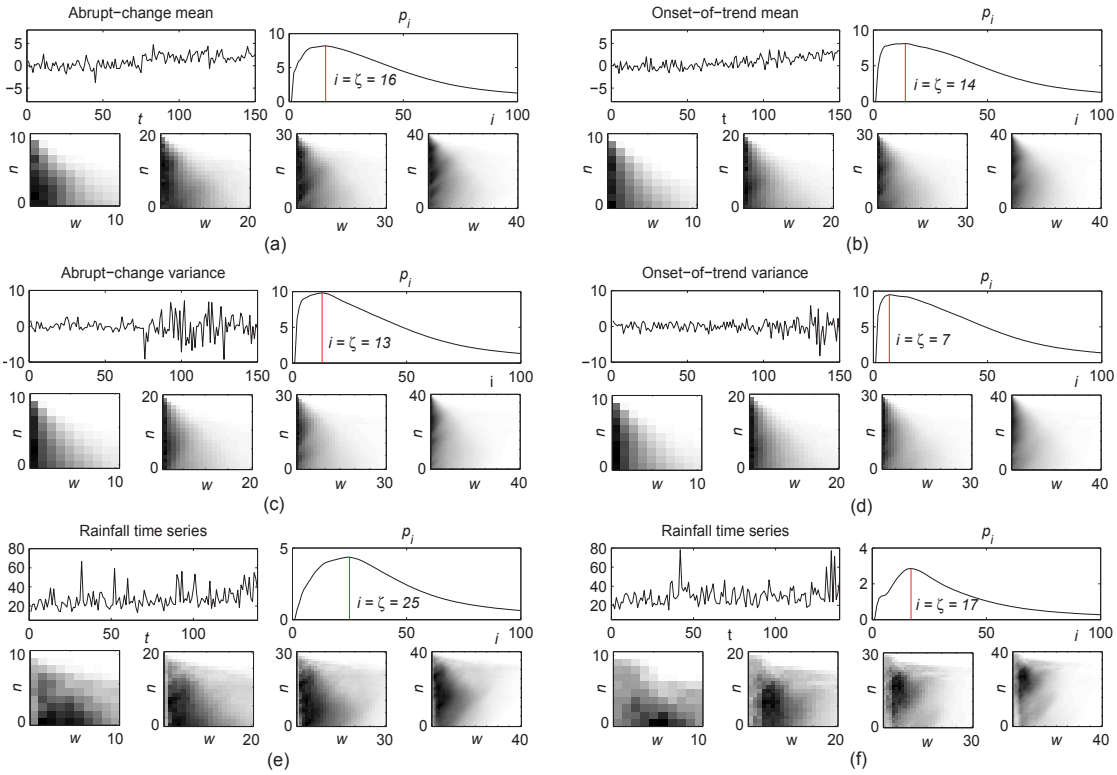

Figure 4.2: Illustrating the automatic procedure for stopping the algorithm for different types of signal. The original time series is plotted in the upper left. The proportion p_i and the point ζ that leads to the *maximum* are shown in the upper right. In the bottom, we illustrated the CP scores in gray levels over different areas given by $h_i = [1, ..., w_0 + i; 1, ..., n_0 + i]$. These plots are displayed for the following cases: (a) signal with a mean following the *abrupt-change* model. (b) signal with a variance following the *abrupt-change* model. (c) signal with a mean following the *onset-of-trend* model. (d) signal with a variance following the *onset-of-trend* model. (e) and (f) real world signals.

4.3.3 Selecting significant patterns

Having defined $\mathcal{Z}$ as the collection of scores to be used for CP detection, we can build a matrix $\mathbf{Z}_t$ containing the scores for different w and n at each $t \in [1, N]$:

$$\mathbf{Z}_t = \begin{bmatrix} z_{1,1}^{(t)} & z_{1,2}^{(t)} & \cdots & z_{1,n_0+\zeta}^{(t)} \\ z_{2,1}^{(t)} & z_{2,2}^{(t)} & \cdots & z_{2,n_0+\zeta}^{(t)} \\ \vdots & \vdots & \ddots & \vdots \\ z_{w_0+\zeta,1}^{(t)} & z_{w_0+\zeta,2}^{(t)} & \cdots & z_{w_0+\zeta,n_0+\zeta}^{(t)} \end{bmatrix}, \text{ for } t = 1, ..., N, \tag{4.8}$$

where $z_{1,1}^{(t)} = z(t, 1, 1)$, $z_{1,2}^{(t)} = z(t, 1, 2)$ and so on until $z_{w_0+\zeta,n_0+\zeta}^{(t)} = z(t, w_0 + \zeta, n_0 + \zeta)$, according to the values of $z(t, w, n)$ given in (4.7). Small CP scores happen if the representative patterns of the future are sufficiently similar to some frequent patterns of the past. Hence, if there is no significant change in the dynamics of the signal in

comparison to the past, it is fair to say that the probability of a given $z_{i,j}^{(t)} \in \mathbf{Z}_t$ taking a large value (*i.e.* being close to one) is *low*. Let us represent this "large value" by $\mathcal{T}_z$, which is called the *score threshold* for significant events. The criterion to select $\mathcal{T}_z$ depends on the significance of the change patterns captured by (4.7) that one is interested in. A simple way to select $\mathcal{T}_z$ is to set it according to a given percentage of the *maximum* score value observed in (4.7), *i.e.* $\mathcal{T}_z = (1 - a) \max \mathcal{Z}$, where $a = 0.1$ or $a = 0.05$, for instance. This strategy works well for filtering significant change patterns. We have considered that significant changes are represented by large values of the CP score, which stand for actual changes in the signal. Recall that large values of the CP score occur whenever the singular vectors representing the change patterns of the future of the signal, do not lie close to the hyperplane spanned by the singular vectors representing the change patterns of the past of the signal (see (4.5) and (4.6)). For example, in Fig. 4.3 we use different values of $\mathcal{T}_z$ for selecting change patterns in the $\mathcal{Z}$ collection computed for the nonstationary signal shown in Fig. 4.2(a), which presents an abrupt change in the mean at $t = 75$. Notice that the intervals before and after the change are separated well, specially for $\mathcal{T}_z = 0.6 \max \mathcal{Z}$.

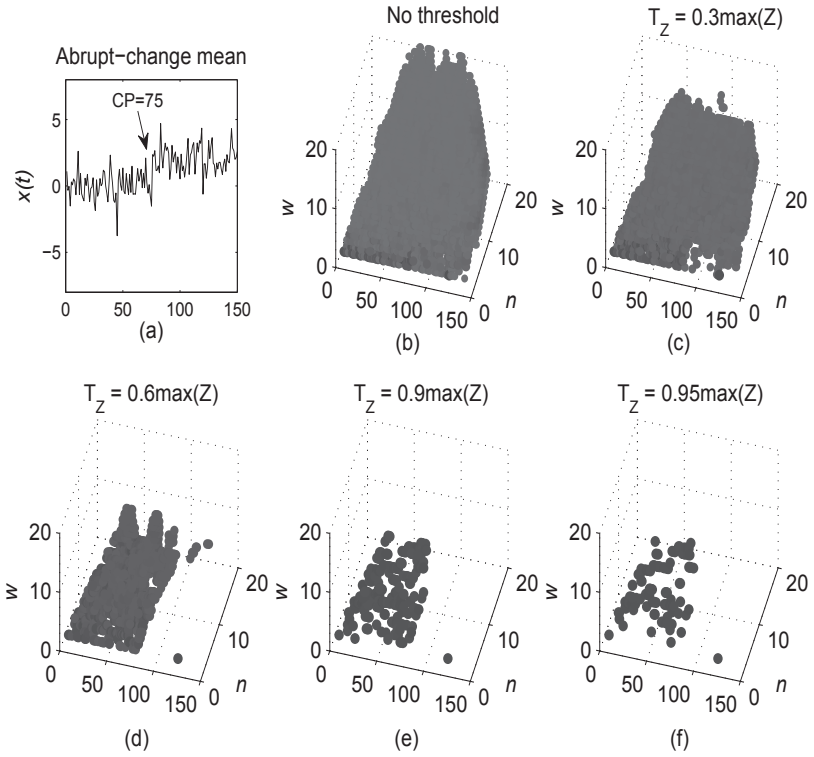

Figure 4.3: Using $\mathcal{T}_z$ for selecting significant change patterns. (a) Nonstationary signal presenting an abrupt-change in the mean at $t = 75$. (b) Collection $\mathcal{Z}$ of scores to be used for CP estimation with no threshold. (c) $\mathcal{T}_z = 0.3 \max \mathcal{Z}$. (d) $\mathcal{T}_z = 0.6 \max \mathcal{Z}$. (e) $\mathcal{T}_z = 0.9 \max \mathcal{Z}$. (f) $\mathcal{T}_z = 0.95 \max \mathcal{Z}$.

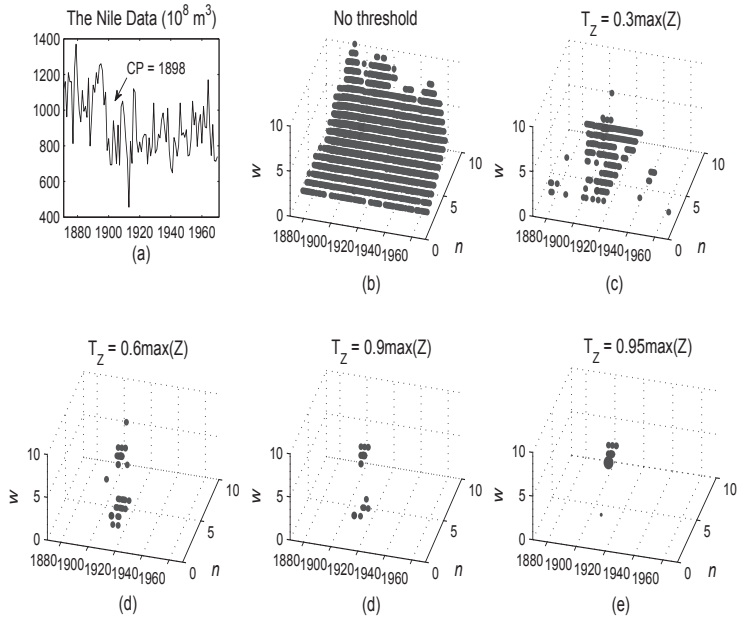

Figure 4.4: Using $\mathcal{T}_z$ for selecting significant change patterns. (a) Annual volume of the Nile River discharge at Aswan, Egypt from 1871 to 1960, which presents an abrupt change at 1898. (b) Collection $\mathcal{Z}$ of scores to be used for CP estimation with no threshold. (c) $\mathcal{T}_z = 0.3 \max \mathcal{Z}$. (d) $\mathcal{T}_z = 0.6 \max \mathcal{Z}$. (e) $\mathcal{T}_z = 0.9 \max \mathcal{Z}$. (f) $\mathcal{T}_z = 0.95 \max \mathcal{Z}$.

In Fig. 4.4, we have tested a well-known case in hydrology of a time series that has been proven to present an abrupt-change at a specific point. This series correspond to the annual volume discharge from the Nile River at Aswan, Egypt, from 1871 to 1960. The Nile data has been studied by many authors, and an abrupt-change in the year 1898 has been detected by using different approaches (*e.g.* by means of a likelihood ratio test [95], by using nonparametric methods [13, 14, 15], or in a semiparametric setting [96, 97]). As pointed in [16], this abrupt-change might be associated to a sudden change in the rainfall regime that occurred near to the year 1900, which is also observed in the rainfall reports of most tropical weather stations [10]. By using the method for CP detection that will be proposed in the following Section, we could detect a CP at year 1896. In Fig. 4.4, notice that by using different values of $\mathcal{T}_z$ we can select the most significant change patterns, and obtain a group of CP scores around the year where the change occurs.

4.4 A new CP detection framework

To estimate the CP, we first threshold $\mathbf{Z}_t$ for $t = 1, ..., N$ according to the chosen score threshold $\mathcal{T}_z$, *i.e.*:

$$\widetilde{\mathbf{Z}}_t = [\widetilde{z}_{i,j}^{(t)}]_{(w_0+\zeta)\times(n_0+\zeta)} = \begin{cases} \widetilde{z}_{i,j}^{(t)} = 0 & \text{if } z_{i,j}^{(t)} \leq \mathcal{T}_z \\ \widetilde{z}_{i,j}^{(t)} = z_{i,j}^{(t)} & \text{if } z_{i,j}^{(t)} > \mathcal{T}_z \end{cases}, \text{ for } t = 1, ..., N, \tag{4.9}$$

where $\widetilde{\mathbf{Z}}_t$ is the thresholded matrix, and $z_{i,j}^{(t)}$ and $\widetilde{z}_{i,j}^{(t)}$ are elements of $\mathbf{Z}_t$ and $\widetilde{\mathbf{Z}}_t$, respectively. In order to estimate the CP, we have assumed the following:

- the *uncertainty* related to the observed CP scores in $\widetilde{\mathbf{Z}}_t$ should increase from $(t-1)$ to t, if t is a CP.

We evaluate the uncertainty by computing the entropy of $\widetilde{\mathbf{Z}}_t$ conditioned on the past values $\widetilde{\mathbf{Y}}_t = \widetilde{\mathbf{Z}}_{t-1}$, *i.e.*, by computing the conditional entropy $H[\widetilde{\mathbf{Z}}_t|\widetilde{\mathbf{Y}}_t]$. To do so, we first create two vectors $\widetilde{\mathbf{z}}_t$ and $\widetilde{\mathbf{y}}_t$ containing, respectively, all possible elements of the matrices $\widetilde{\mathbf{Z}}_t$ and $\widetilde{\mathbf{Y}}_t$ for a given time instant t. Then, by computing the empirical distribution of $\widetilde{\mathbf{Z}}_t$ and $\widetilde{\mathbf{Y}}_t$, we define the individual probabilities $\mathrm{p}(\widetilde{z}_l)$ and $\mathrm{p}(\widetilde{y}_l)$ of a particular score value $\widetilde{z}_l \in \widetilde{\mathbf{z}}_t$ and $\widetilde{y}_l \in \widetilde{\mathbf{y}}_t$, respectively. The conditional entropy is thus computed by:

$$H(\widetilde{\mathbf{Z}}_t|\widetilde{\mathbf{Y}}_t) = -\sum_{l=1}^{L_y} \mathrm{p}(\widetilde{y}_l) \sum_{l=1}^{L_z} \mathrm{p}(\widetilde{z}_l|\widetilde{y}_l) \log_2 \mathrm{p}(\widetilde{z}_l|\widetilde{y}_l), \text{ for } t = 1, ..., N, \tag{4.10}$$

where L_z and L_y are, respectively, the numbers of possible score values in $\mathbf{z}_t$ and $\mathbf{y}_t$. We choose the CPs as the points where the local *maxima* of $H(\widetilde{\mathbf{Z}}_t|\widetilde{\mathbf{Y}}_t)$ occur. The CP estimation depends on the chosen significance of the change patterns. Note that the most general way would be to compute (4.10) by using $\mathcal{T}_z = (1-a)\max \mathcal{Z}$ for many different values of $a \in [0, 1]$.

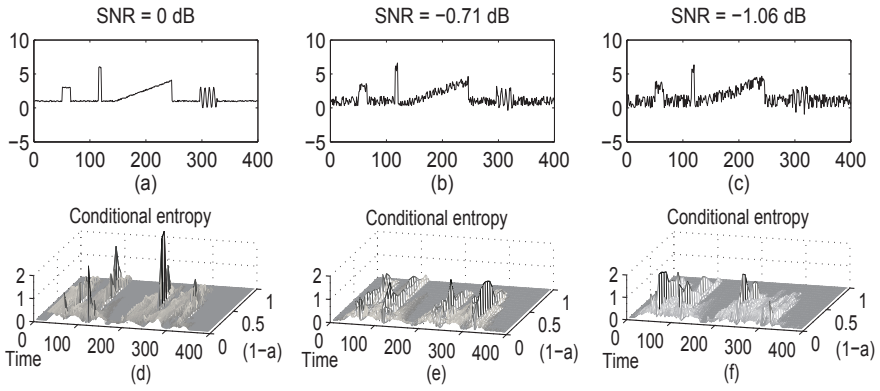

Figure 4.5: Signals with a changing mean following three different patterns embedded in background noise. Conditional entropy for each signal, computed by considering different values of $(1-a)$ for $\mathcal{T}_z = (1-a)\max \mathcal{Z}$.

4.5 Testing the new approach

To illustrate the modified RSST, we present in Fig. 4.5 three signals with a time-varying mean that undergoes three consecutive patterns in the same series: i) two abrupt-changes with amplitude values (x_a) set to $x_a = 3$ and $x_a = 6$, ii) a piece-wise linear trend with slope $m = 0.03$ and largest amplitude value set to $x_a = 4$, and iii) an oscillation with four cycles per 36 samples with peak-to-peak value (x_{pp}) set to $x_{pp} = 4$. We show the effect of adding uniform white noise with the following signal-to-noise ratio (SNR) values: 0 dB (Fig. 4.5(a)), -0.71 dB (Fig. 4.5(b)) and -1.06 dB (Fig. 4.5(c)). These values of SNR correspond to a background noise with peak-to-peak values of 0%, 60% and 90% in comparison to the original signal. In Fig. 4.5(d), (e) and (f) we show $H(\widetilde{\mathbf{Z}}_t|\widetilde{\mathbf{Y}}_t)$ computed by means of (4.10) for different values of $(1-a)$ in $\mathcal{T}_z = (1-a)\max \mathcal{Z}$. Notice that the peaks of $H(\widetilde{\mathbf{Z}}_t|\widetilde{\mathbf{Y}}_t)$ in time follow the changes in the signal, and the peaks corresponding to those patterns varying abruptly can be seen for a wider range of $(1-a)$, reflecting the significance of those change patterns. We have also tested signals with a varying variance and both a varying mean and variance following the *abrupt-change* model with initial and final parameter values ranging from $\xi_1 = 0$ to $\xi_2 = 4$. We have considered three different CPs for the nonstationary signals: CP = 150, 225, 300. The nonstationary signals and their conditional entropy for different levels of significance $(1-a)$ are shown in Fig. 4.6 and Fig. 4.7, for the cases of nonstationary variance and nonstationary mean and variance, respectively.

The CP estimation has been carried out by computing the local *maxima* of $H(\widetilde{\mathbf{Z}}_t|\widetilde{\mathbf{Y}}_t)$ by means of the MATLAB function `findpeaks`. The routine takes as input the *maxima* of $H(\widetilde{\mathbf{Z}}_t|\widetilde{\mathbf{Y}}_t)$ over all $(1-a)$ and returns the time instants where the peaks occur. We

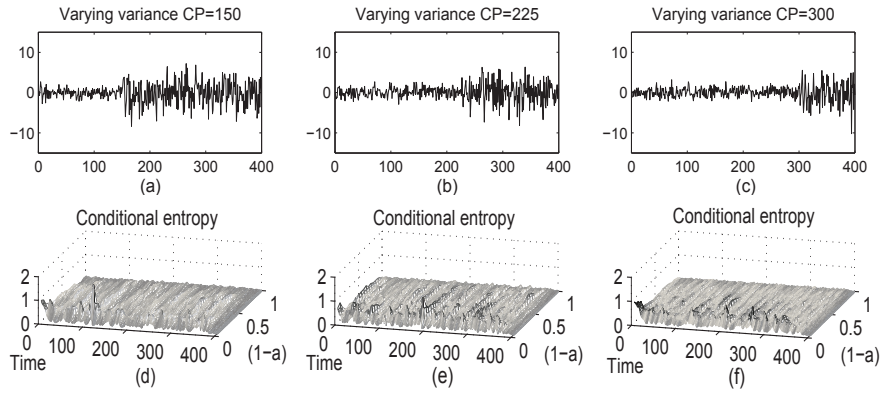

Figure 4.6: Signals exhibiting an abrupt-change in the variance at three different points. Conditional entropy for each signal, computed by considering different values of $(1-a)$ for $\mathcal{T}_z = (1-a)\max \mathcal{Z}$.

have assumed that consecutive peaks should be separated by a short interval, at least as short as $\Delta_t = 0.1N$, otherwise they are considered to represent the same change pattern. Notice in Fig. 4.5 that the large values of $H(\widetilde{\mathbf{Z}}_t|\widetilde{\mathbf{Y}}_t)$ are indeed spread over a small neighborhood.

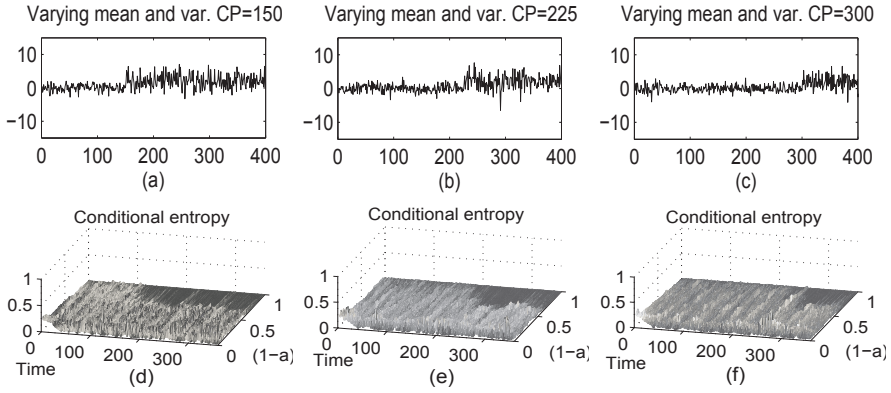

Figure 4.7: Signals exhibiting an abrupt-change in the mean and in the variance at three different points. Conditional entropy for each signal, computed by considering different values of $(1-a)$ for $\mathcal{T}_z = (1-a)\max \mathcal{Z}$.

In Table 4.1, we show the true and the estimated CPs for the signals shown in Fig. 4.5 and Fig. 4.6, respectively. It can be seen in Table 4.1 that the proposed approach can reasonably detect the CPs for all values of SNRs. Also, one can verify that the method cannot detect properly the beginning of the linear trend at CP = 145. In general, the framework is not suited for detecting the CP of slowly-varying nonstationarities such as a linear trend. Furthermore, in many real world applications, we do not have the ground truth available. Thus, we do not know how many CPs are to be detected in a given time series. By sorting (4.10) in a descending order from $\max H(\widetilde{\mathbf{Z}}_t|\widetilde{\mathbf{Y}}_t)$ to $\min H(\widetilde{\mathbf{Z}}_t|\widetilde{\mathbf{Y}}_t)$, and

Table 4.1: Results of applying the proposed algorithm to the signals shown in Fig. 4.5 and Fig. 4.6.

Changing mean

True CP	50	115	145	246	300
SNR (dB)	Estimated CP				
0	62	118	not detected	**243**	309
-0.71	62	111	not detected	**244**	310
-1.06	63	114	not detected	**242**	312

Changing variance

True CP	150	225	300
Estimated CP	**155**	**229**	**299**

Changing mean and variance

True CP	150	225	300
Estimated CP	**162**	**235**	**305**

ordering the time points accordingly, one could use $t_{\text{MAX}} = \underset{t}{\text{argmax}}[H(\widetilde{\mathbf{Z}}_t|\widetilde{\mathbf{Y}}_t)]$ as the time instant that is most likely the CP. In Table 4.1, the most likely CPs are highlighted.

4.5.1 Comparison with other approaches

We have compared the proposed approach with the algorithm for CP detection presented in [10], which is based on the Lombard's *smooth-change* model and makes use of the test statistics proposed in [18]. The algorithm of [10] is also nonparametric and can detect changes both in the mean and in the variance. These characteristics make the Lombard's technique suitable for being applied to real world data and to be compared with the approach proposed in this work. The Lombard's method will be briefly explained below.

4.5.1.1 Lombard's technique

Consider a time series $\mathcal{X} = (x_1, ..., x_n)$ containing independent samples. Let us assume that $F_i(x) = \mathrm{p}(x_i \leq x)$, $i \in [1, n]$ are continuous. Now, let us define the normalize rank of x_i as $R_i = F_n(x_i)$, where $F_n(x) = \frac{1}{n+1}\sum_{l=1}^{n} \mathbb{I}(x_i \leq x)$ stands for the re-scaled empirical distribution function and $\mathbb{I}(\cdot)$ is the indicator of a set. Now, for a square-integrable score function $\phi : [0, 1] \to \Re$ and for each $i \in [1, n]$, we define the rank score $Z_i = \{\phi(R_i - \overline{\phi})\}/\sigma_\phi$, where $\overline{\phi}$ and σ_ϕ^2 are the sample mean and the sample variance of the collection $\{\phi(R_i), i = 1, ..., n\}$. The Lombard's approach is based on the measure:

$$\mathcal{L}_{\kappa_1,\kappa_2} = \sum_{j=\kappa_1+1}^{\kappa_2} \sum_{l=1}^{j} Z_l, \tag{4.11}$$

from which different test statistics can be derived for detecting different nonstationary patterns, like the *abrupt-change* model and the *onset-of-trend* model defined in (2.19) and

Table 4.2: Results of applying the Lombard's technique to the signals shown in Fig. 4.5 to Fig. 4.7.

Changing mean

True CP	50	115	145	246	300
SNR (dB)	Estimated CP				
0	not detected	not detected	not detected	246	not detected
−0.71	not detected	not detected	not detected	245	not detected
−1.06	not detected	not detected	not detected	249	not detected

Changing variance

True CP	150	225	300
Estimated CP	**149**	**225**	**298**

Changing mean and variance

True CP	150	225	300
Estimated CP	**210**	**149**	**299**

(2.20), respectively. In order to use the test statistics derived from (4.11) for detecting CPs in time series, we can rely either on estimated or on asymptotic critical values obtained from their limiting distributions [10].

4.5.1.2 Experimental study using the Lombard's technique

In Table 4.2 we show the results obtained by applying the Lombard's technique to the signals shown in Fig. 4.5 to Fig. 4.7. It can be seen that Lombard's test does not allow for the detection of multiple CPs in one time series, contrary to the developed method. Instead, only the most likely CP is detected for the signals with a changing mean. Also, the algorithm could not detect properly the CPs of the signals with both varying mean and variance. In fact, such performance is expected, as the algorithm does not perform well when a change in the mean occurs if one is testing for a change in the variance. This problem is worse when the mean is large, becoming more apparent as the sample size grows [10]. Other disadvantage of the Lombard's approach is that the observations contained in time series should be independent. This assumption is often not true for real world data. Moreover, for detecting the CP in the mean or in the variance, we should first set the algorithm according to the nonstationarity we are aiming at detecting. This is a crucial point, since in many cases we do not know *a priori* which kind of nonstationary are present in the signal. It can be a problem when the visual inspection of the time series is unfeasible. Despite these limitations, it should be noted that the Lombard's technique is much faster than the approach proposed here, and should be considered if one is testing larger time series, or if one is particularly interested in the patterns included in Lombard's model (*e.g. abrupt-change*, *onset-of-trend*) [10].

4.6 Conclusions

In the frame of nonstationary signal analysis, many techniques are devoted to detect the time instants (if any), where a change in a given parameter occurs. Change point assessment is a very important topic in many areas such as climatology, where the first CP might be an indicator of the beginning of a climate change [10]. Also, since the stationarity tests presented in this work are made to reject stationarity of the whole observation interval, one would need a separate algorithm to detect CPs. Many algorithms for CP discovery are parametric or require an ad-hoc adjustment for every signal. These characteristics, however, are not in consonance with those that are expected from a CP detector suitable for being applied to real world signals, which was our motivation here.

In this Chapter, we have proposed a modified framework based on an existing technique called RSST, which presents the necessary profile to be applied to real world signals. However, the original RSST suffers from its own generality: the CP detection depends on the choice of two parameters (two "windows" of analysis) that need to be chosen by visualization or domain knowledge, which are often impracticable. The first contribution that we have made in this Chapter avoided the problematic part of having to set a specific pair of windows for the analysis. The procedure consisted in representing the output of the RSST in the space spanned by its two parameters. This has shown to be a good strategy, as different change patterns can be captured if we sweep over the two parameters. Then, we have proposed a stopping criterion for the sweep, which is a crucial point for reducing the computational time of the algorithm. Having a representation of the change dynamics of the signals for different windows and time instants, we then presented a way for selecting significant change patterns. Finally, the CP estimation has been carried out by evaluating the uncertainty related to the change patterns, more precisely, by computing the conditional entropy at each time instant. The CP itself has been obtained by searching for the instants where the local *maxima* of the conditional entropy vector occur.

Having presented the framework, we compared the proposed approach with another one available in the literature, which is also used for testing real world signals. The proposed approach is more versatile, and allows us to visualize the intervals where major changes in the signal occur as a function of the window of analysis and time. However, the technique presented in this Chapter is considerably time consuming if one is testing larger time series. On the other hand, there are many applications like in biomedicine and in hydrology, where short time series need to be analyzed by nonparametric techniques in a richer feature space. This is the case of the dataset that will be studied in

the next Chapter, for which we not only applied the technique proposed in this Chapter, but also the other stationarity tests presented earlier in this manuscript.

Chapter 5

Experimental study on real world data

Contents

Chapter 5 presents the results of applying the various methods developed in this Thesis to environmental data.

5.1 Introduction

5.1.1 The Canadian Regional Climate Model (CRCM)

The environmental data tested in this Chapter are rainfall time series corresponding to annual *maximum* daily precipitation obtained from the Canadian Regional Climate Model (CRCM). The CRCM data have been generated and supplied by Ouranos [43].

The CRCM is a realistic limited-area nested model that considers many complex multi-scale interactions and feedbacks. The multitude of variables that this climate model takes into account ranges from physical parameters such as radiation scheme, cloud formations, and atmospheric boundary layer mixing [98], to social considerations like possible future scenarios involving population growth, economic development, and technological change [99]. The CRCM allows for the study of climate change and variability, and also allows for the understanding of the large number of processing governing the climate system. The data generated by this model have been used for climate predictions on time scales ranging from seasons to decades.

5.1.2 The rainfall data

The region covered by the CRCM consists of 200×192 grid points over the North-American domain with a 45 km horizontal grid-size mesh. The rainfall data tested in this Chapter correspond to a smaller area of that region: 1631 locations spread over the province of Quebec (QC), Canada. This area is illustrated in Fig. 5.1 together with each grid point considered in the simulations.

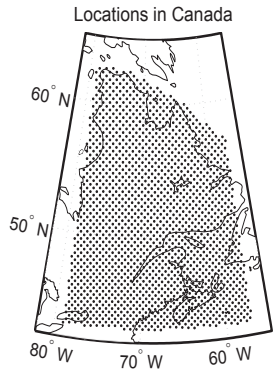

Figure 5.1: Map showing the grid points in Canada considered for the simulations.

The data are the annual *maximum* values obtained from the simulated daily precipitations ranging from December 1960 to November 2100. For the simulations, we had at our disposal two datasets, known as *aet* [100] and *aev* [101] simulations. These were obtained from two distinct realizations of the same model configuration, meaning that only the initial conditions are slightly different. These simulations were performed by using the version 4.2.3 of the CRCM model, and were driven by the data CGCM3 [102]. Also, the simulations considered SRES A2 scenario for years 2001-2100. In Fig. 5.2, we show for several grid points chosen randomly in the province of Quebec, the respective annual maximum daily rainfall signals for *aet* and *aev* simulation.

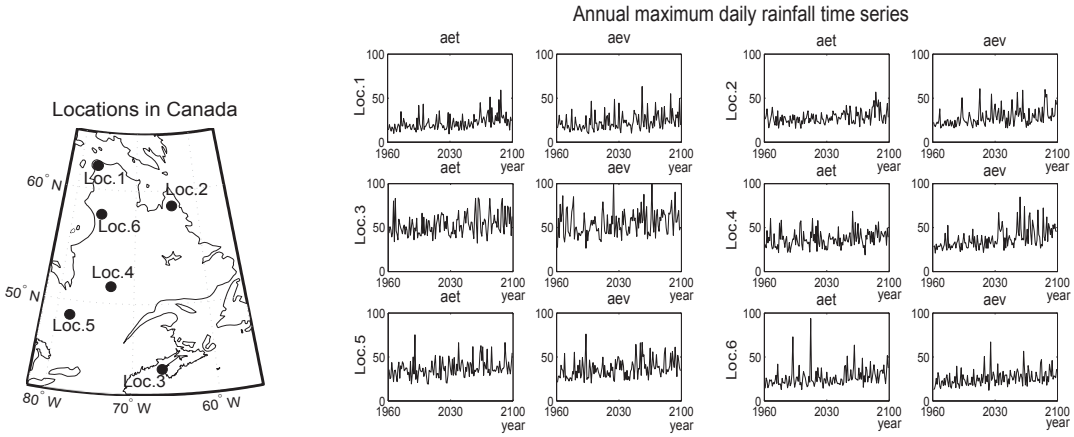

Figure 5.2: Example of six grid points randomly chosen and their corresponding annual maximum daily rainfall time series for both simulations (*aet* and *aev*).

Table 5.1: Statistical characteristics of the rainfall data

Location	1		2		3		4		5		6	
Simulation	*aet*	*aev*	*aet*	*aev*	*aet*	*aev*	*aet*	*aev*	*aet*	*aev*	*aet*	*aev*
maximum	59.16	63.22	56.89	60.75	83.74	100.36	68.46	84.60	75.11	75.96	93.70	66.97
mean	22.06	23.08	28.26	28.35	53.15	54.94	37.11	37.58	37.03	36.87	26.81	25.91
std	8.31	9.65	7.17	9.82	12.01	14.24	9.41	11.85	10.14	10.92	10.91	8.67
skewness	1.61	1.43	0.92	1.39	0.70	0.71	0.72	1.41	1.04	0.92	2.71	1.45
kurtosis	6.47	5.31	4.27	4.73	2.97	3.58	3.37	5.07	4.57	3.71	14.39	6.63

5.1.2.1 Exploratory analysis

In this Section, we perform a short exploratory analysis of the rainfall dataset. To this end, we have selected the time series illustrated in Fig. 5.2, and we have computed their *maxima*, the sample mean, standard deviation, skewness, and kurtosis. The results are shown in Table 5.1 for both the *aet* and *aev* simulations. Note that the data are positively skewed and also have positive kurtoses, which indicate heavy tails on the right side of the distribution. In this regard, we present in Fig. 5.3 the histograms of each signal, where we can see that the data have high peaks, and thus a heavy-tailed distribution. Such behavior was expected, since the data are the annual *maximum* daily precipitation (extreme values), which are not normally distributed. The significant presence of outliers and the heavy tailedness can be verified in the bloxplots of the data shown in Fig. 5.4. As it could be seen, such dataset is difficult to analyze by using classical approaches, which usually either rely on the stationarity or in the Gaussianity of the underlying process, or are not robust against the presence of outliers. However, the techniques developed in this Thesis are suited to analyze such dataset.

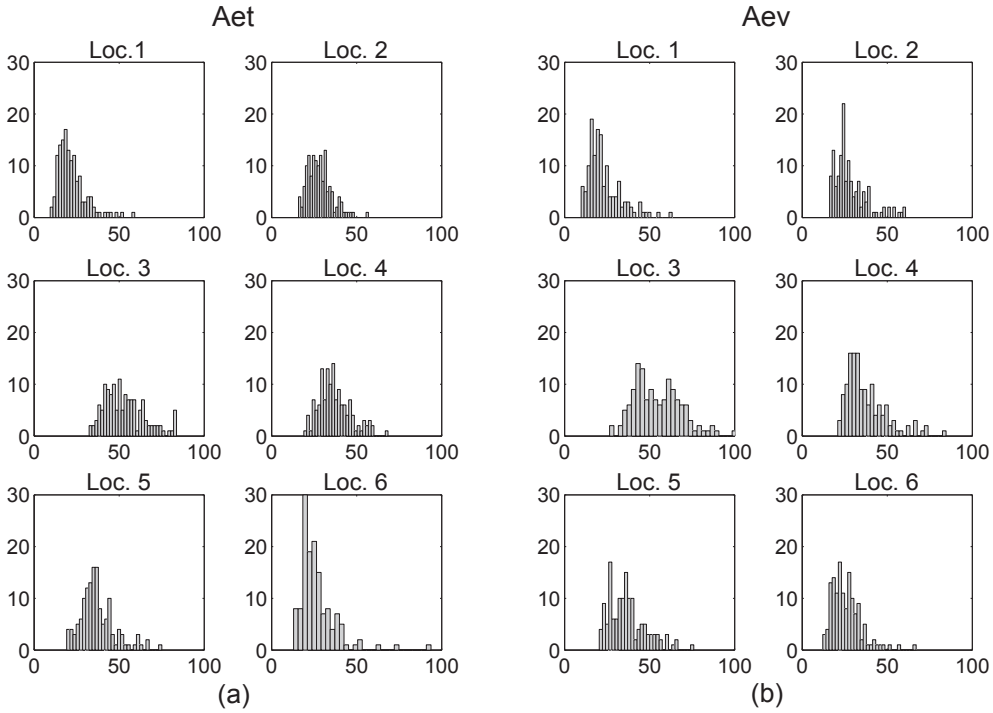

Figure 5.3: Histograms of the rainfall time series illustrated in Fig. 5.2. (a) (*aet* simulation. (b) *aev* simulation.

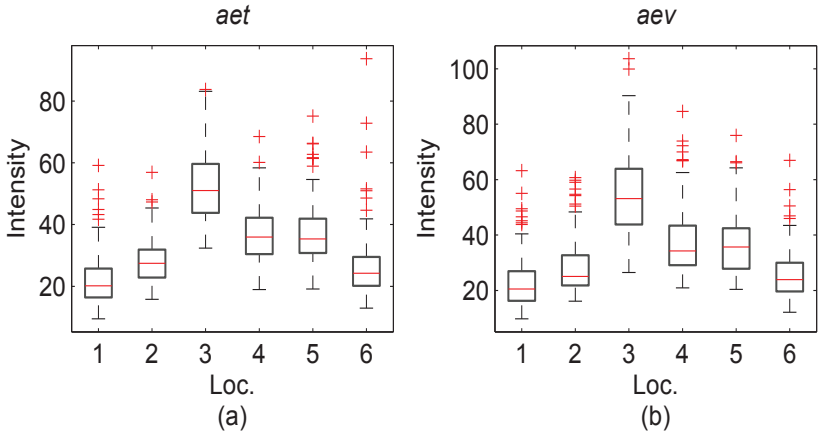

Figure 5.4: Boxplots of the rainfall time series illustrated in Fig. 5.2. (a) (*aet* simulation. (b) *aev* simulation.

5.2 Applying the stationarity tests

In this Section, we show the results of applying the stationarity tests developed in this Thesis, and also other nonstationarity tests available in the literature. For all the stationarity tests, a confidence level of 95% was used. In Section 5.2.1, we present the results for the surrogate-based approach and its modified version. In Section 5.2.2, we illustrate

the performance of the new stationarity test proposed in Chapter 3 for detecting trend-based nonstationarities. The results of applying other methods available in the literature are shown in Section 5.2.3.

5.2.1 Results for the surrogate-based technique

For applying the stationarity test described in Chapter 2 we have to specify the parameter n_{h0} (see (2.4)). For the datasets we considered the following values:

$$n_{h0} = [0.06, 0.08, 0.10, 0.12, 0.14, 0.16, 0.18, 0.20, 0.22, 0.24, 0.26, 0.28].$$

These values lead to Hermite windows with the following length:

$$n_h = [7, 11, 13, 15, 19, 21, 25, 27, 29, 33, 35, 37].$$

5.2.1.1 Testing with distances of probability nature

We have tested the annual maximum daily rainfall data by applying all distances shown in Table 2.1. However, as expected, the distances of probability nature did not lead to a satisfactory performance. Although the stationarity was rejected for some grid points by applying the original method, after weighting the probability-based distances the test could not reject the null hypothesis for almost all grid points and n_{h0} values. Moreover, if we consider the outcomes for all the different grid points and values of n_{h0}, and check if there exist "nonstationary" grid points intersecting all of them, we get a null result for both original and modified method. These results are in accordance with the problems described in Section 2.6 and Section A.2 involving the use of probability-based distances. That was not the case, on the other hand, for the distances of frequency nature. Thus for the remaining of this Section, we shall analyze the performances of the frequency-based distances only.

5.2.1.2 Testing with distances of frequency nature

By applying the method using the frequency-based distances, we have observed that the number of "nonstationary" grid points obtained for each simulation (*aet* and *aev*) *tends to be closer to each other*, for all considered values of n_{h0} and all distances. This situation is shown in Table 5.2, where it can be seen that the number of nonstationary results increases substantially by decreasing the value of n_{h0}, or by weighting the distances. However, the number of nonstationarities does not increase or decrease considerably by changing from the *aet* to the *aev* simulation.

Table 5.2: Number of grid points classified as "nonstationary" for each distance, simulation and value of n_{h0}.

Distance		Itakura-Saito		Log-spectral		Diffusion		Symmetrized Itakura-Saito	
Simulation		*aet*	*aev*	*aet*	*aev*	*aet*	*aev*	*aet*	*aev*
$n_{h0} = 0.06$	Original	970	924	342	310	1092	1071	175	168
	Weighted	1519	1482	1024	976	1615	1608	422	364
$n_{h0} = 0.08$	Original	892	828	427	401	999	970	255	232
	Weighted	1352	1300	970	914	1573	1545	484	437
$n_{h0} = 0.10$	Original	869	812	465	432	954	924	288	276
	Weighted	1297	1241	925	869	1540	1519	497	427
$n_{h0} = 0.12$	Original	831	779	477	458	907	862	320	300
	Weighted	1233	1170	869	802	1499	1479	461	391
$n_{h0} = 0.14$	Original	796	744	551	522	846	799	371	383
	Weighted	1106	1045	758	695	1442	1396	428	340
$n_{h0} = 0.16$	Original	784	717	551	548	798	763	387	384
	Weighted	1044	983	719	646	1400	1342	396	299
$n_{h0} = 0.18$	Original	707	617	521	520	751	687	379	391
	Weighted	919	860	613	530	1293	1238	336	229
$n_{h0} = 0.20$	Original	695	633	516	533	732	674	381	381
	Weighted	891	796	607	488	1262	1231	311	212
$n_{h0} = 0.22$	Original	682	601	522	513	689	631	389	385
	Weighted	823	743	540	442	1204	1165	278	184
$n_{h0} = 0.24$	Original	647	582	489	501	653	600	377	386
	Weighted	757	644	450	368	1126	1117	218	142
$n_{h0} = 0.26$	Original	628	578	490	508	639	574	384	378
	Weighted	706	573	418	335	1072	1050	192	133
$n_{h0} = 0.28$	Original	632	539	501	495	630	553	383	337
	Weighted	694	550	394	313	1061	1019	170	116

In Fig. 5.5, we show, for *aet* simulation, the grid points corresponding to the *maximum* annual daily rainfall time series where the stationarity was rejected for every distances and values of n_h (*i.e.* the intersections). The number of grid points correspond to 36 for the original method, and to 95 for the modified one. The results for *aev* simulation are shown in Fig. 5.6, where the number of intersections of the original and modified method is 20 and 65, respectively. Notice that most of the intersections are obtained only for the modified method. In general, after applying the weighting procedure the number of nonstationary cases returned by the method increases (note that it was the opposite for the case of the probability-based distances). The final intersection of nonstationary locations for the *aet* and *aev* simulations is shown in Fig. 5.7. For this experimental study, we have also verified if the conditions for the gamma parameters shown in (2.27) for improving the stationarity test are reached (in the average). In Table 5.3, it can be seen that, in the majority of the nonstationary cases, the two inequalities are verified, *i.e.* $k' < k$ and $\Theta_1'/\theta' > \Theta_1/\theta$ (except for the Log-spectral distance).

Relying only on the possible "nonstationarity" / "stationarity" outcome of the method may not be judicious, as we are applying the method to the annual *maximum* daily rainfall data corresponding to only two simulations of the CRCM. Thus, making use of some extra information for evaluating the nonstationary locations, such as the robustness measure AMD (see (2.28)) is recommended.

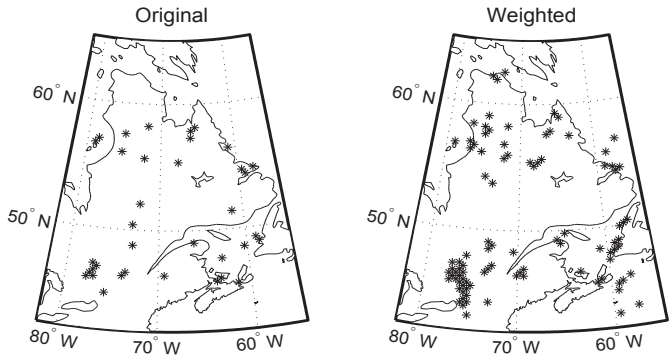

Figure 5.5: Grid points for *aet* simulation that were classified as "nonstationary" for every distance and value of n_{h}.

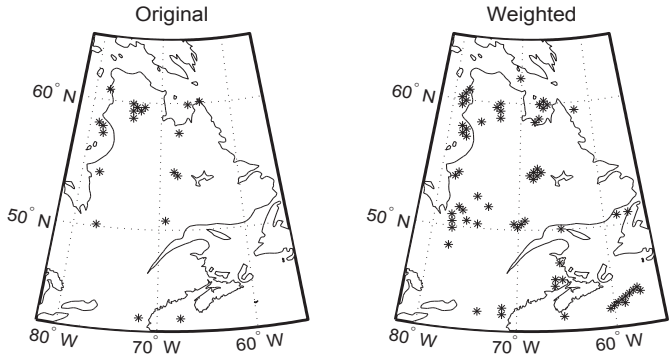

Figure 5.6: Grid points for *aev* simulation that were classified as "nonstationary" for every distance and value of n_{h}.

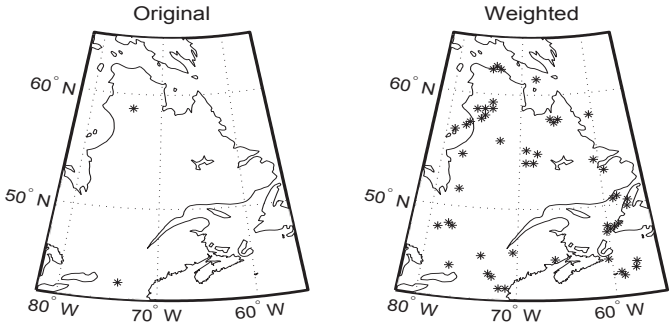

Figure 5.7: Final intersection of nonstationary locations of *aet* and *aev* simulations. All distances and values of n_{h} were considered.

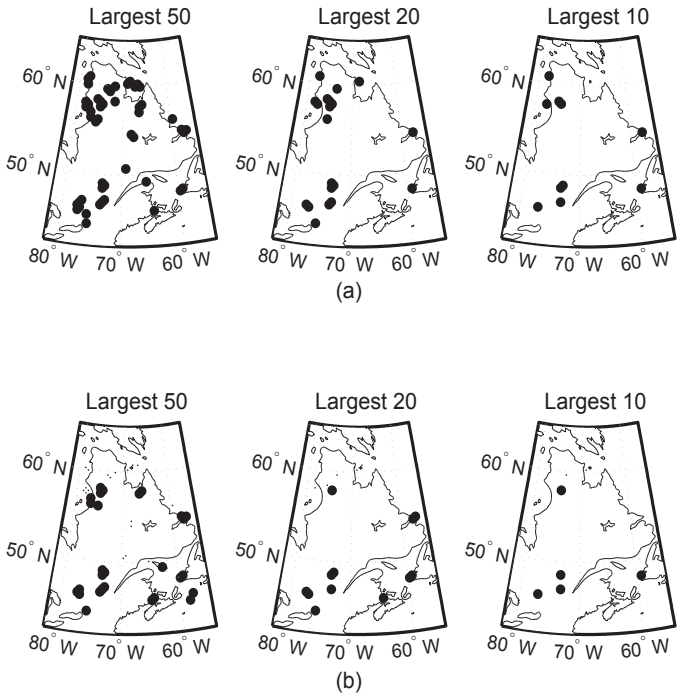

Figure 5.8: Grid points corresponding to the largest 50, 20, and 10 AMD values for the Itakura-Saito distance, by using (a) the original method, (b) the weighted one.

To this end, we have selected from both *aet* and *aev* simulations, the largest 50, 20, and 10 AMD values returned by the method for each distance. Then we have checked from which grid point the time series leading to the largest AMD values came from. The results are illustrated in Fig. 5.8, Fig. 5.9, Fig. 5.10, and Fig. 5.11 for the Itakura-Saito, Log-Spectral, Diffusion and Symmetrized Itakura-Saito distances, respectively.

To this end, we have plotted in the map the nonstationary cases given by using each distance and value of n_{h0}, whereas the larger the value of AMD evaluated from the data of a given grid point is, the bigger the marker used to represent that grid point on the map. For *aet* and *aev* simulations, the results issued from the AMD values are shown in the following figures according to the chosen distance: i) Itakura-Saito Fig. 5.12 (*aet*) and Fig. 5.13 (*aev*), ii) Log-spectral Fig. 5.14 (*aet*) and Fig. 5.15 (*aev*), iii) Diffusion Fig. 5.16 (*aet*) and Fig. 5.17 (*aev*), iv) Symmetrized Itakura-Saito Fig. 5.18 (*aet*) and Fig. 5.19 (*aev*). Notice that, in general, there are many nonstationary grid points spread over the North and the South of the map, but not in the central area. More specifically, for most of the cases, there exists a concentration of nonstationary grid points with large values of AMD clustered in the Southeast (over the Nova Scotia province), Southwest (over the North of Ottawa) and Northwest (in the South of the Ungava Peninsula).

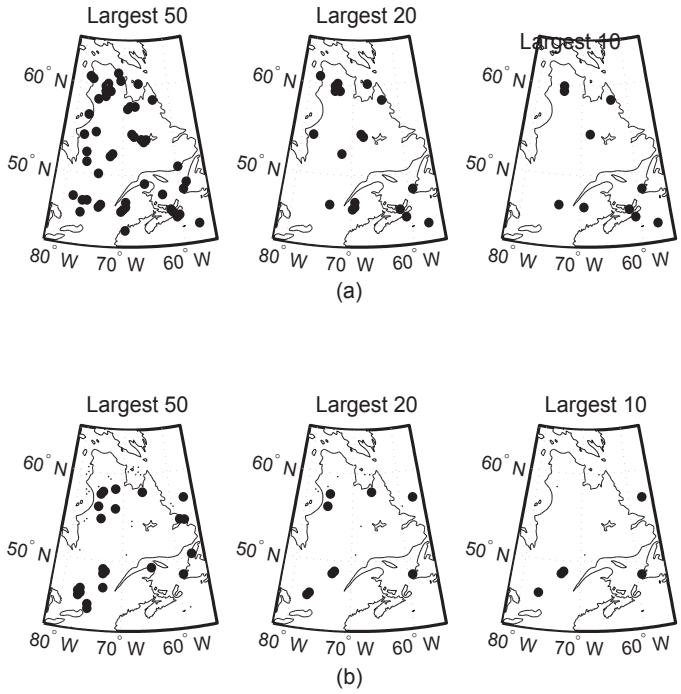

Figure 5.9: Grid points corresponding to the largest 50, 20, and 10 AMD values for the Log-spectral distance, by using (a) the original method, (b) the weighted one.

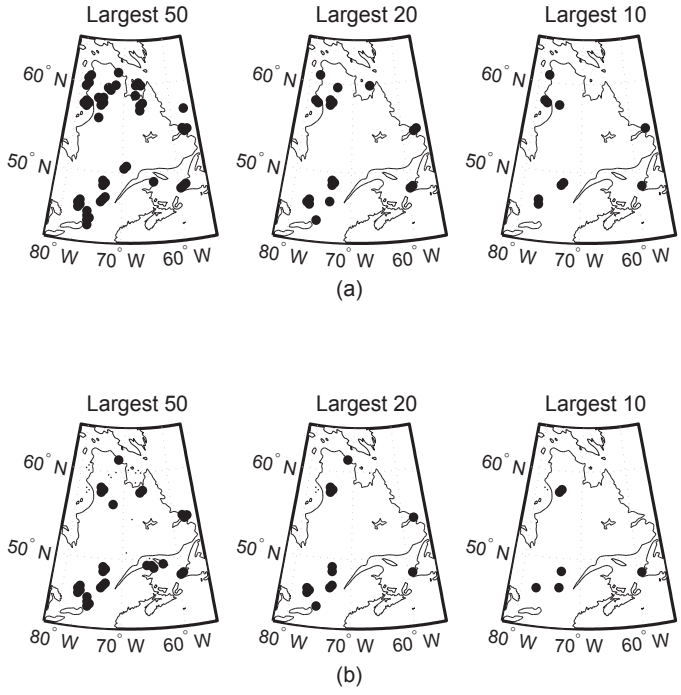

Figure 5.10: Grid points corresponding to the largest 50, 20, and 10 AMD values for the Diffusion distance, by using (a) the original method, (b) the weighted one.

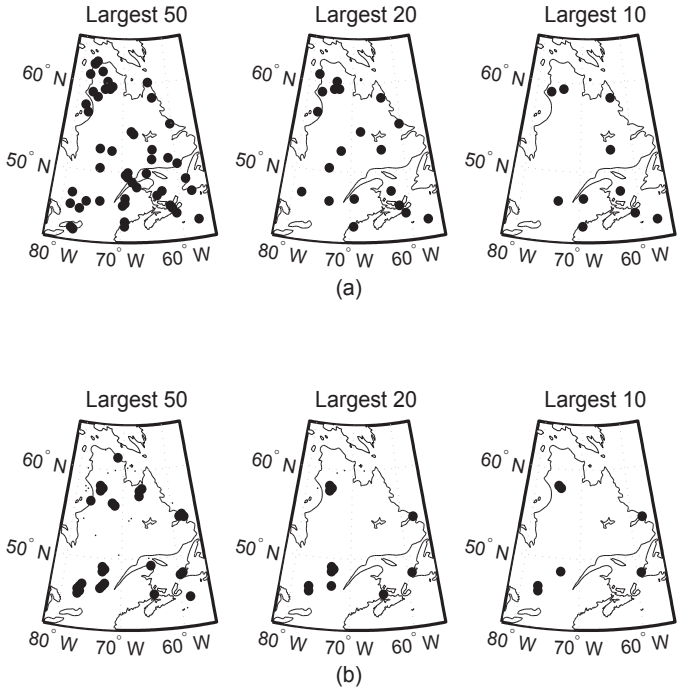

Figure 5.11: Grid points corresponding to the largest 50, 20, and 10 AMD values for the Symmetrized Itakura-Saito distance, by using (a) the original method, (b) the weighted one.

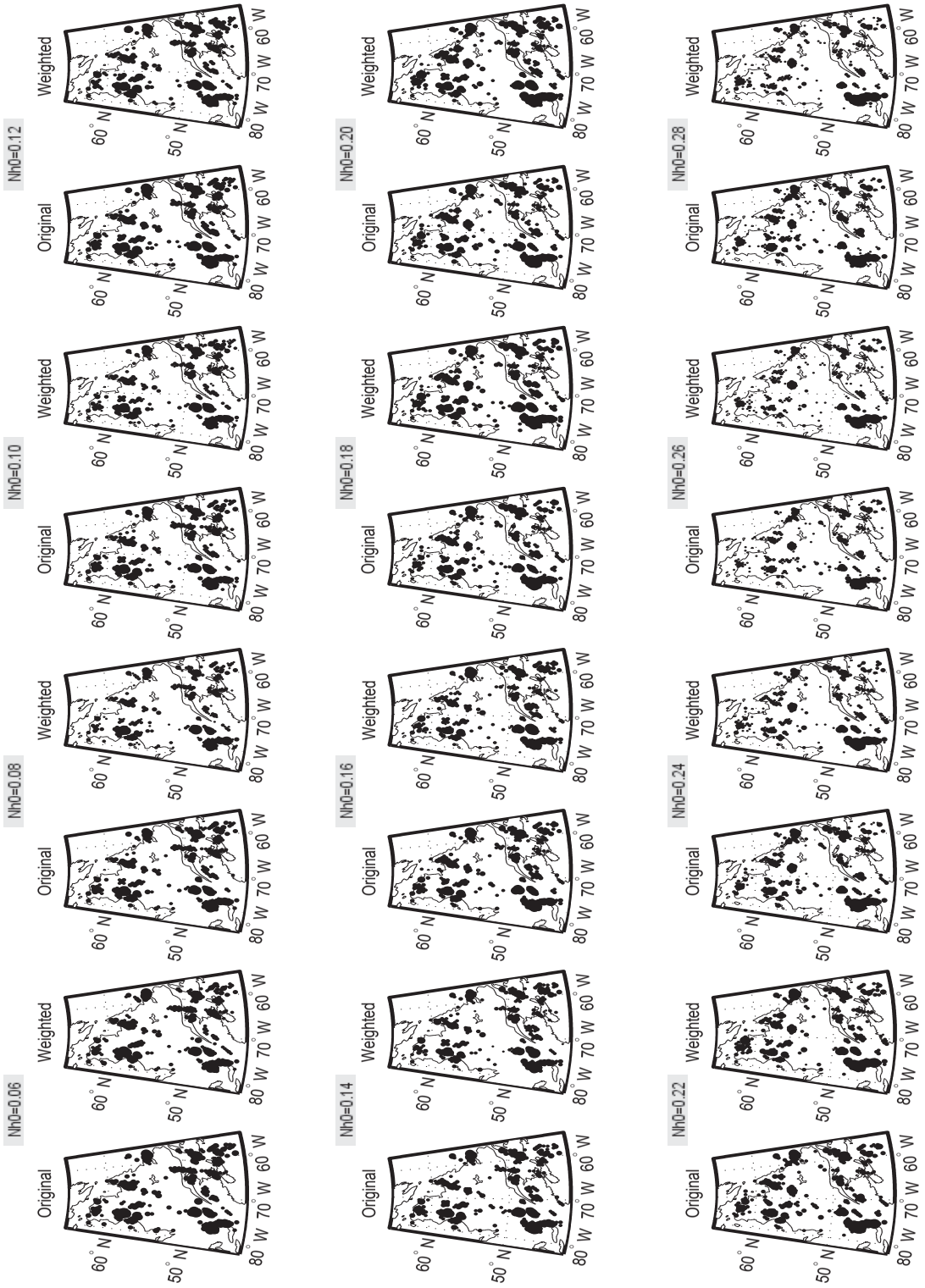

Figure 5.12: For *aet* simulation, nonstationary locations given by the Itakura-Saito distance. The marker showing each grid point is scaled according to the respective AMD value computed for the location.

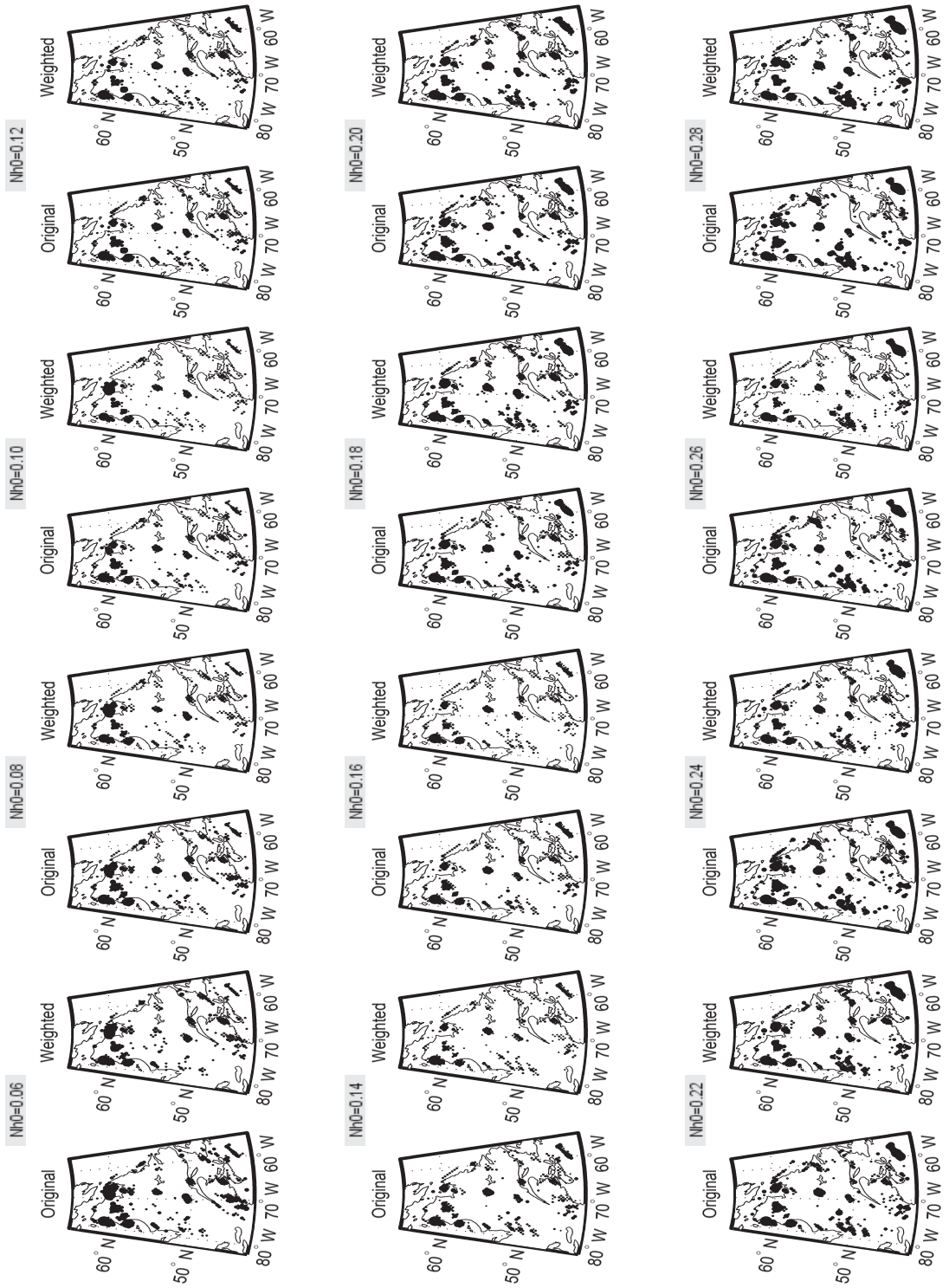

Figure 5.13: For *aev* simulation, nonstationary locations given by the Itakura-Saito distance. The marker showing each grid point is scaled according to the respective AMD value computed for the location.

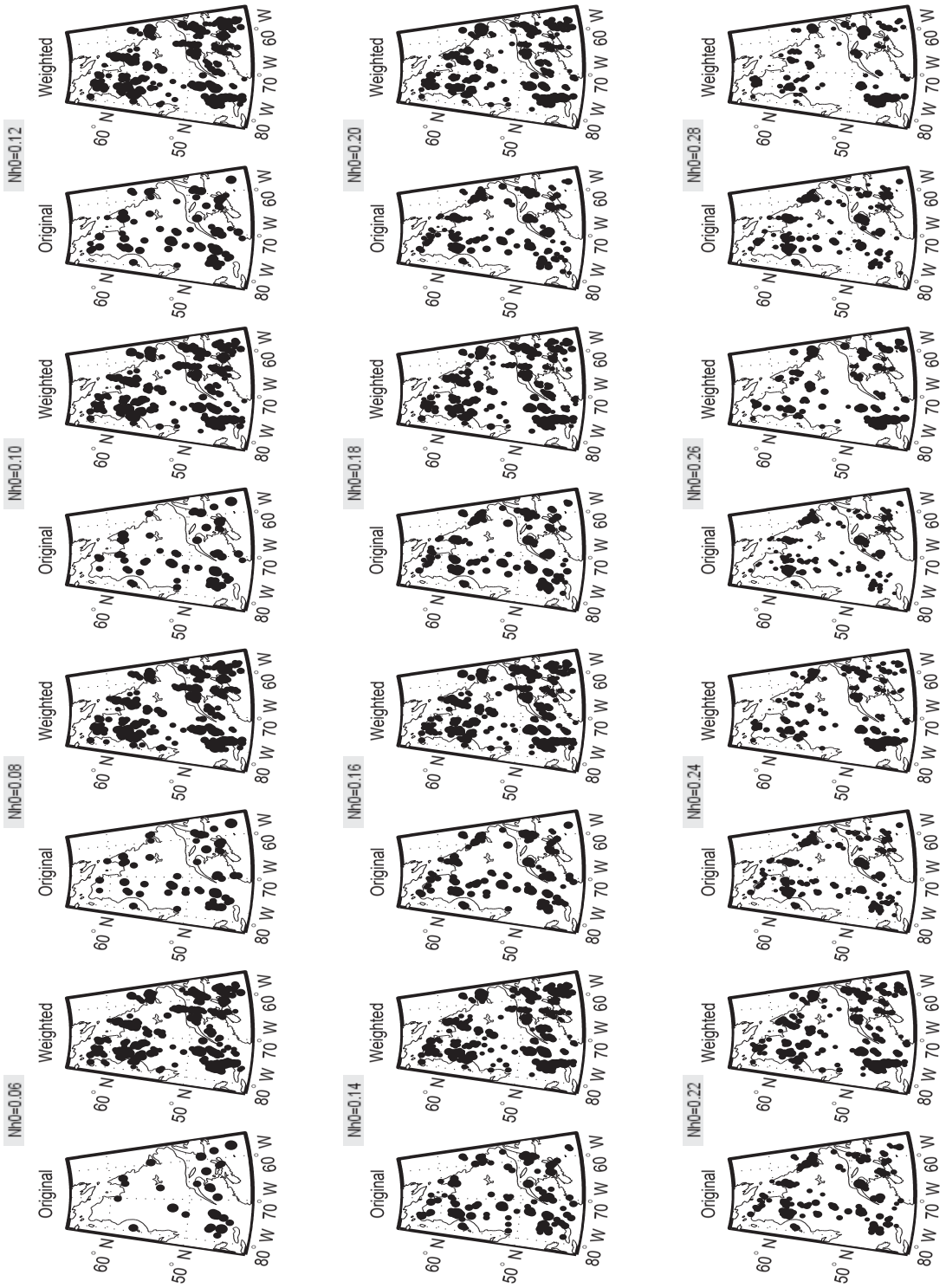

Figure 5.14: For *aet* simulation, nonstationary locations given by the Log-spectral distance. The marker showing each grid point is scaled according to the respective AMD value computed for the location.

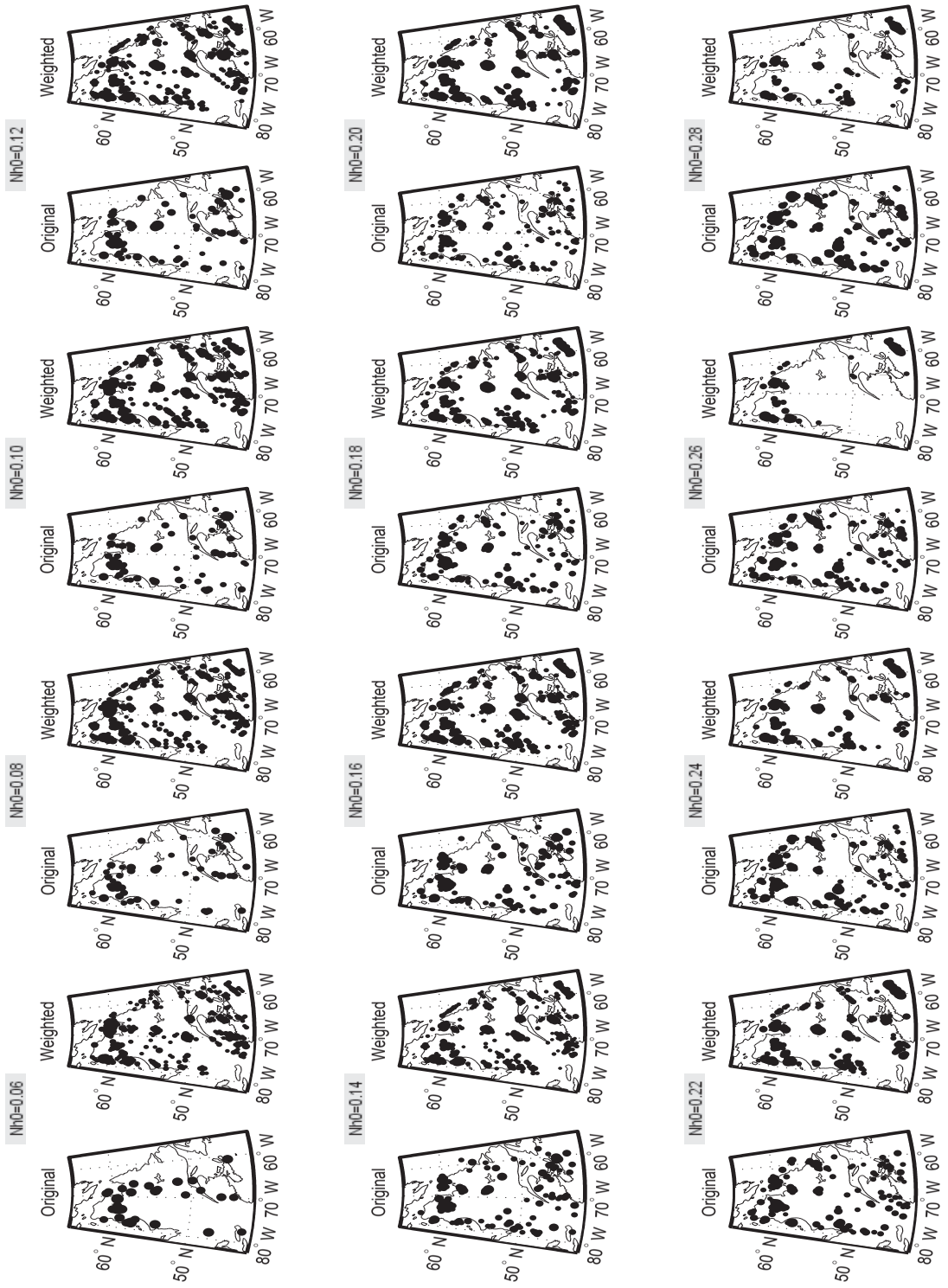

Figure 5.15: For *aev* simulation, nonstationary locations given by the Log-spectral distance. The marker showing each grid point is scaled according to the respective AMD value computed for the location.

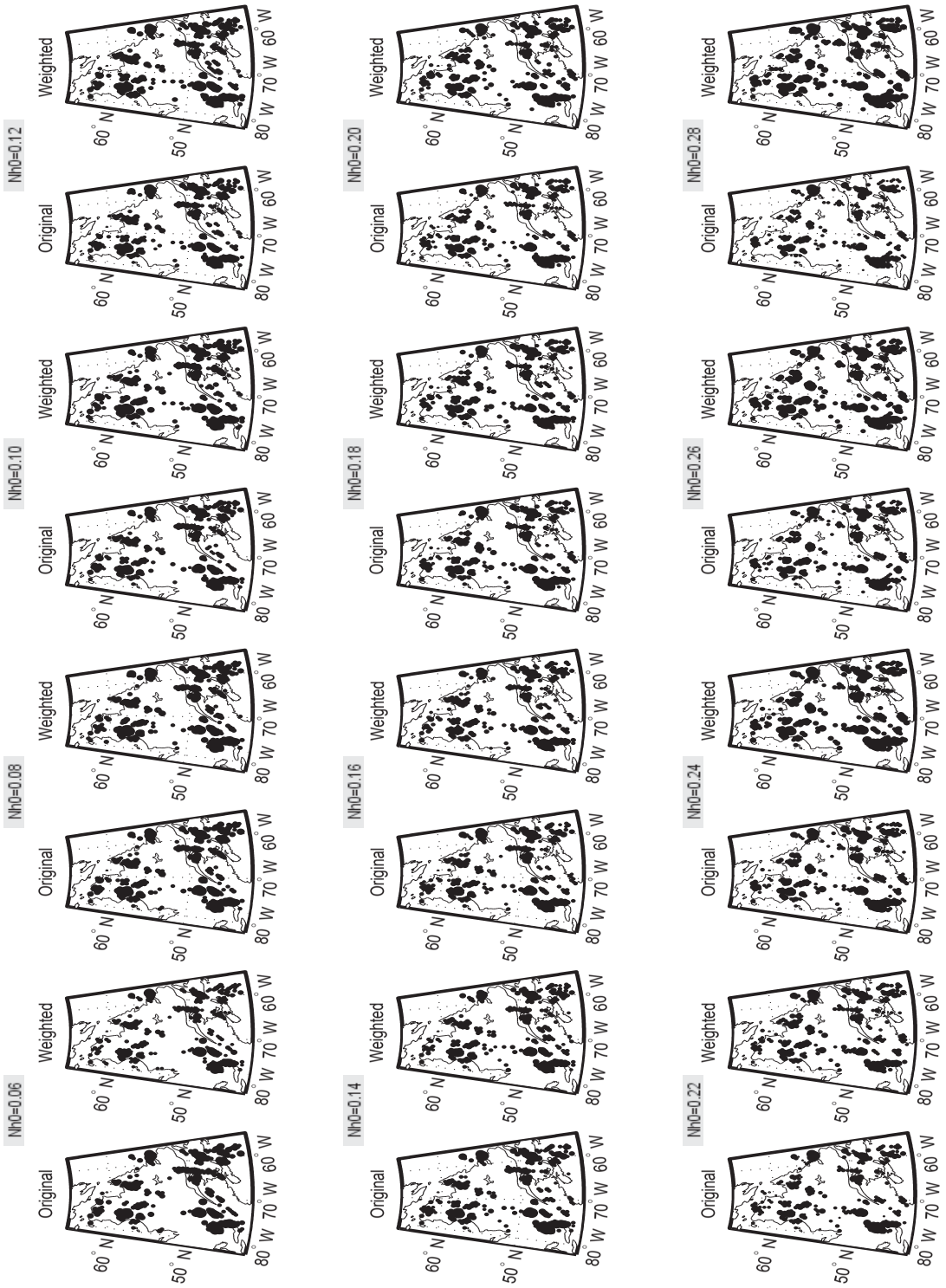

Figure 5.16: For *aet* simulation, nonstationary locations given by the Diffusion distance. The marker showing each grid point is scaled according to the respective AMD value computed for the location.

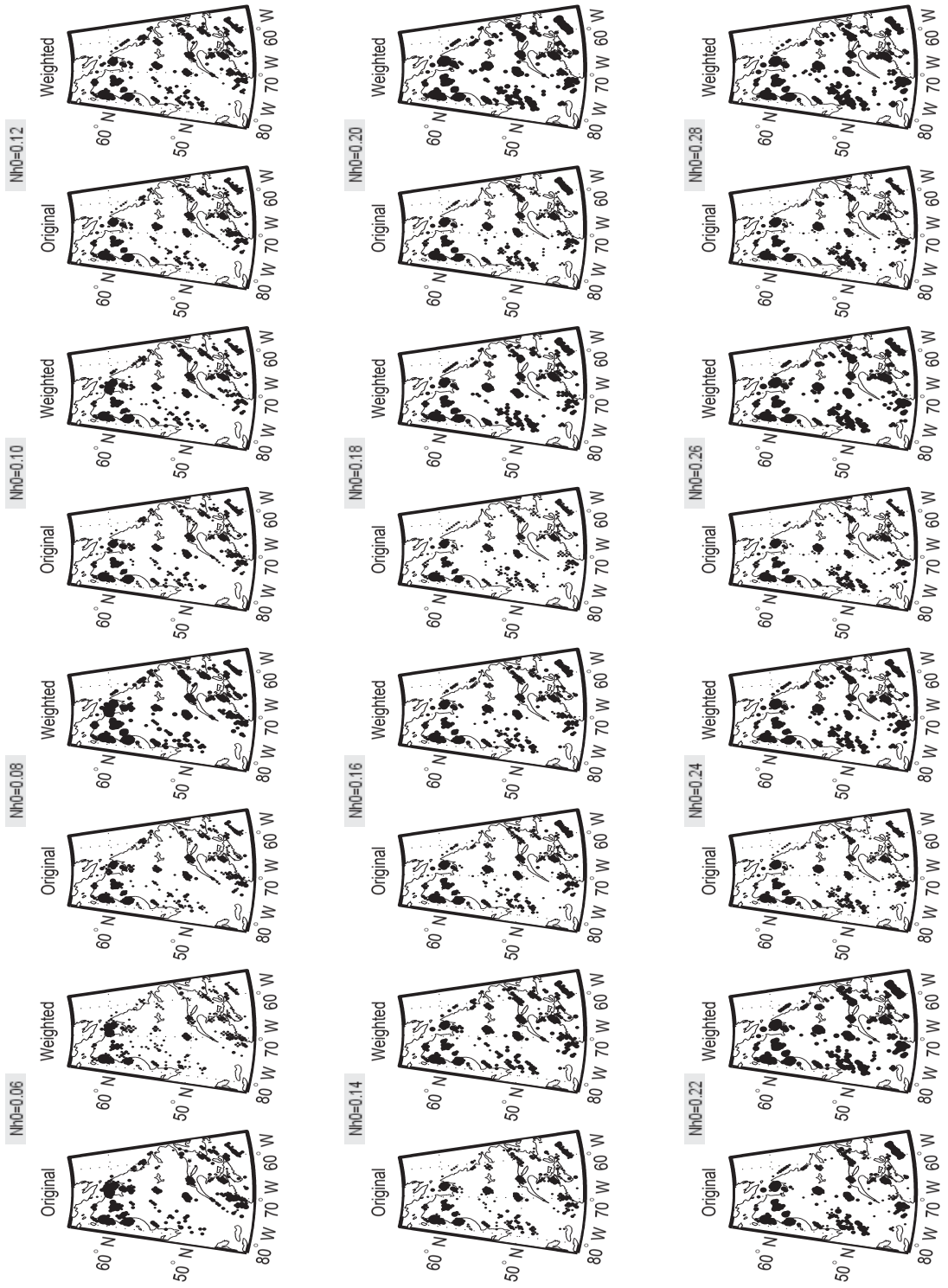

Figure 5.17: For *aev* simulation, nonstationary locations given by the Diffusion distance. The marker showing each grid point is scaled according to the respective AMD value computed for the location.

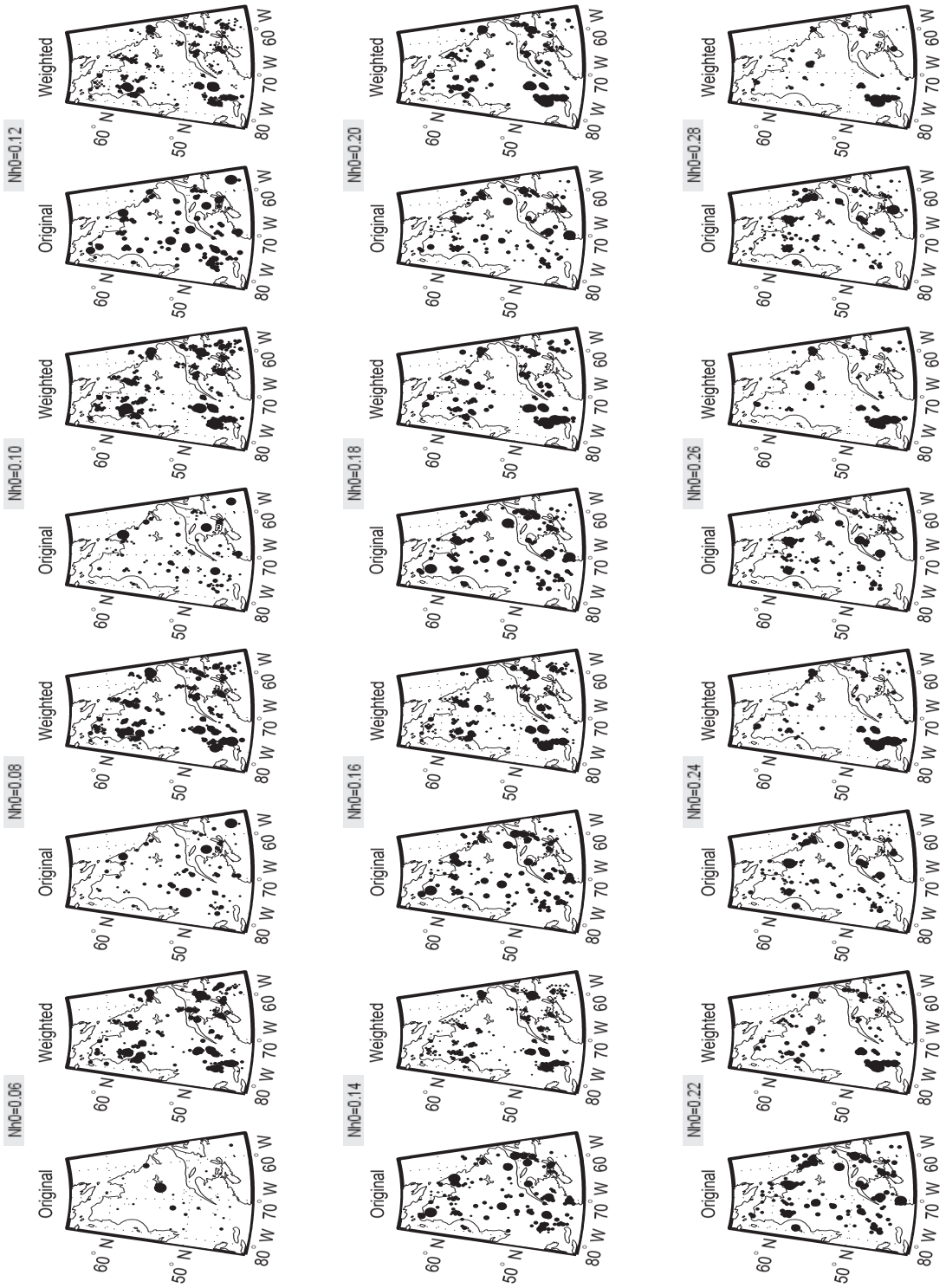

Figure 5.18: For *aet* simulation, nonstationary locations given by the Symmetrized Itakura-Saito distance. The marker showing each grid point is scaled according to the respective AMD value computed for the location.

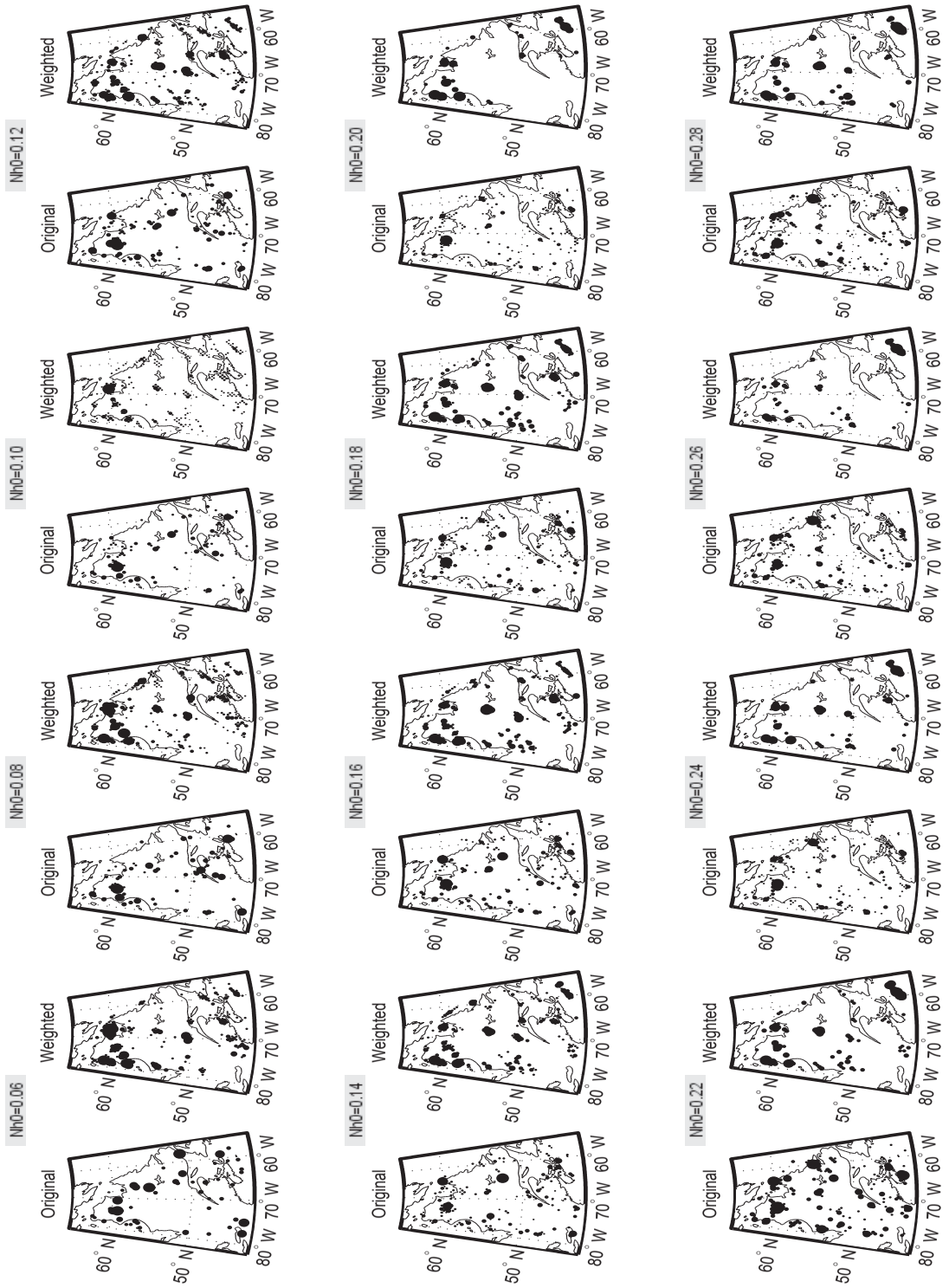

Figure 5.19: For *aev* simulation, nonstationary locations given by the Symmetrized Itakura-Saito distance. The marker showing each grid point is scaled according to the respective AMD value computed for the location.

Table 5.3: Verifying the condition of (2.27) after applying the weighting procedure. The results are given as arithmetic averages $(\langle k\rangle, \langle \Theta_1/\theta\rangle)$ and $(\langle k'\rangle, \langle \Theta_1'/\theta'\rangle)$ of the estimates of $(k, \Theta_1/\theta)$ and $(k', \Theta_1'/\theta')$, over all nonstationary and stationary grid points for both *aet* and *aev* simulation, for each distance (Itakuro-Saito [IS], Log-spectral [LS], Diffusion [DF], Symmetrized Itakuro-Saito [SIS]) and value of n_{h0}.

Nonstationary grid points																
Distance		IS				LS				DF				SIS			
n_{h0}	Real.	$\langle k\rangle$	$\langle k'\rangle$	$\langle \Theta_1/\theta\rangle$	$\langle \Theta_1'/\theta'\rangle$	$\langle k\rangle$	$\langle k'\rangle$	$\langle \Theta_1/\theta\rangle$	$\langle \Theta_1'/\theta'\rangle$	$\langle k\rangle$	$\langle k'\rangle$	$\langle \Theta_1/\theta\rangle$	$\langle \Theta_1'/\theta'\rangle$	$\langle k\rangle$	$\langle k'\rangle$	$\langle \Theta_1/\theta\rangle$	$\langle \Theta_1'/\theta'\rangle$
0.06	*aet*	5.15	1.76	17.5	51.8	11.2	2.65	15.4	9.39	6.80	2.62	21.8	59.6	1.31	0.74	2.53	7.40
	aev	5.15	1.80	15.8	45.7	10.9	2.74	15.0	9.34	6.83	2.69	20.1	54.8	1.30	0.75	2.36	6.54
0.08	*aet*	1.31	1.52	13.4	37.2	8.24	2.54	1.28	9.71	5.99	2.70	19.0	41.2	1.29	0.77	3.44	8.36
	aev	3.29	1.53	12.0	33.7	8.20	2.59	12.7	9.56	6.00	2.75	17.5	38.4	1.28	0.78	3.26	7.31
0.10	*aet*	2.91	1.45	12.4	33.1	6.75	2.43	11.3	9.59	5.10	2.64	17.0	36.8	1.29	0.81	3.69	8.31
	aev	2.88	1.47	10.1	30.7	6.73	2.48	11.4	9.50	5.05	2.65	15.2	34.2	1.28	0.88	3.69	7.91
0.12	*aet*	2.61	1.40	11.5	29.6	5.60	2.32	10.3	9.42	4.44	2.57	15.1	32.8	1.26	0.81	4.08	8.60
	aev	2.58	1.43	10.2	26.1	5.62	2.38	10.1	9.30	4.43	2.60	13.5	29.9	1.26	0.81	3.93	7.06
0.14	*aet*	2.34	1.38	11.1	25.8	4.54	2.31	9.65	9.60	3.77	2.57	13.2	28.5	1.28	0.83	4.92	7.41
	aev	2.37	1.43	10.1	22.8	4.58	2.36	9.64	9.13	3.77	2.58	11.2	26.1	1.28	0.86	5.17	6.91
0.16	*aet*	2.29	1.38	10.9	24.3	4.31	2.35	9.52	9.67	3.56	2.55	12.2	26.1	1.27	0.87	5.07	7.52
	aev	2.31	1.43	9.81	20.5	4.32	2.39	9.58	9.77	3.58	2.56	11.5	24.2	1.28	0.87	5.41	6.78
0.18	*aet*	2.19	1.38	10.6	20.9	3.85	2.31	9.29	9.35	3.22	2.51	1.13	22.9	1.22	0.85	5.13	6.41
	aev	2.23	1.41	9.64	17.6	3.94	2.35	9.41	9.01	3.24	2.50	10.6	21.5	1.24	0.85	6.26	6.40
0.20	*aet*	2.16	1.39	10.4	19.7	3.73	2.35	9.13	9.12	3.17	2.54	11.1	22.5	1.24	0.87	5.55	6.47
	aev	2.22	1.43	9.74	16.7	3.82	2.40	9.45	9.07	3.19	2.53	10.3	20.4	1.21	0.87	6.59	6.00
0.22	*aet*	2.09	1.37	10.2	18.3	3.61	2.34	9.10	9.12	3.05	2.53	10.6	20.8	1.22	0.87	5.49	6.19
	aev	2.14	1.40	9.41	15.6	3.65	2.33	9.27	9.71	3.09	2.52	9.90	19.3	1.22	0.87	6.69	6.07
0.24	*aet*	2.01	1.35	9.99	16.4	3.32	2.31	9.07	8.96	2.88	2.50	10.2	19.2	1.17	0.87	5.98	6.08
	aev	2.02	1.38	9.09	14.0	3.30	2.33	9.09	8.57	2.89	2.48	9.25	17.2	1.17	0.86	8.16	5.74
0.26	*aet*	1.94	1.35	9.51	15.4	3.11	2.27	8.57	8.50	2.79	2.47	9.81	17.9	1.13	0.85	6.12	5.84
	aev	1.95	1.35	9.13	13.7	3.21	2.35	9.02	8.65	2.78	2.45	9.08	16.6	1.16	0.86	8.75	5.83
0.28	*aet*	1.91	1.34	9.31	14.0	3.07	2.31	8.69	8.61	2.76	2.46	9.90	17.2	1.13	0.85	6.20	5.86
	aev	1.93	1.35	8.92	12.2	3.11	2.34	8.97	8.24	2.76	2.51	9.06	16.5	1.18	0.86	9.41	5.68

Table 5.4: Number of grid points classified as "nonstationary" for *aet* and *aev* simulations and different values of window size n_{h}.

Simulation / n_{h}	3	5	7	9	11
aet	92	95	109	110	125
aev	99	125	130	178	118

5.2.2 Results for the new stationarity test

We show in this Section the results of applying the stationarity test developed in Chapter 3 for testing trend-based and slowly-varying nonstationarity. This framework has a free parameter, which is the size of the Hermite window n_{h} chosen for estimating the time marginal. For the datasets, we considered the following values $n_{\mathrm{h}} = [3, 5, 7, 9, 11]$. By applying the proposed method for detecting trend-based nonstationarities, we have observed that the total number of nonstationary outcomes was smaller than the one obtained for the surrogate-based approach. In Table 5.4, we show the number of nonstationary cases obtained per value of n_{h} for the *aet* and *aev* simulations.

In Fig. 5.20, we show the locations in which signals were classified as "nonstationary" by the method. The results are shown according to the chosen window. Also, we have computed the value of the index of nonstationarity $\mathcal{I}_{\mathrm{NS}}$, proposed in (3.34). So, in Fig. 5.20, the bigger the marker used to represent the grid point is, the greater is the respective value of $\mathcal{I}_{\mathrm{NS}}$.

It should be noticed, that we did not find any intersection between the nonstationary locations given by the surrogate-based method, and the new technique proposed in Chapter 3. It could be explained by the fact that the two stationarity tests are designed for detecting different types of nonstationarities, and the nonstationary behavior detected by one at a given grid point, will be not necessarily detected by the other.

5.2.3 Results for other stationarity tests

We have applied the KPSS test and the Kay's nonstationarity procedure to the annual maximum daily rainfall data. These stationarity tests were compared with the new approach for testing slowly-varying nonstationarities described in Chapter 3. By applying the nonstationarity procedure of S. Kay, we could not find any nonstationary location. Although this method works for short sample sizes, and performed well in accusing the synthetic nonstationary signals shown in Fig. 3.10, it is still a parametric approach, whose performance depends on the chosen model. Possibly, the model assumed for the nonstationary signals (a time-varying autoregressive model) was not a good choice with the datasets under study.

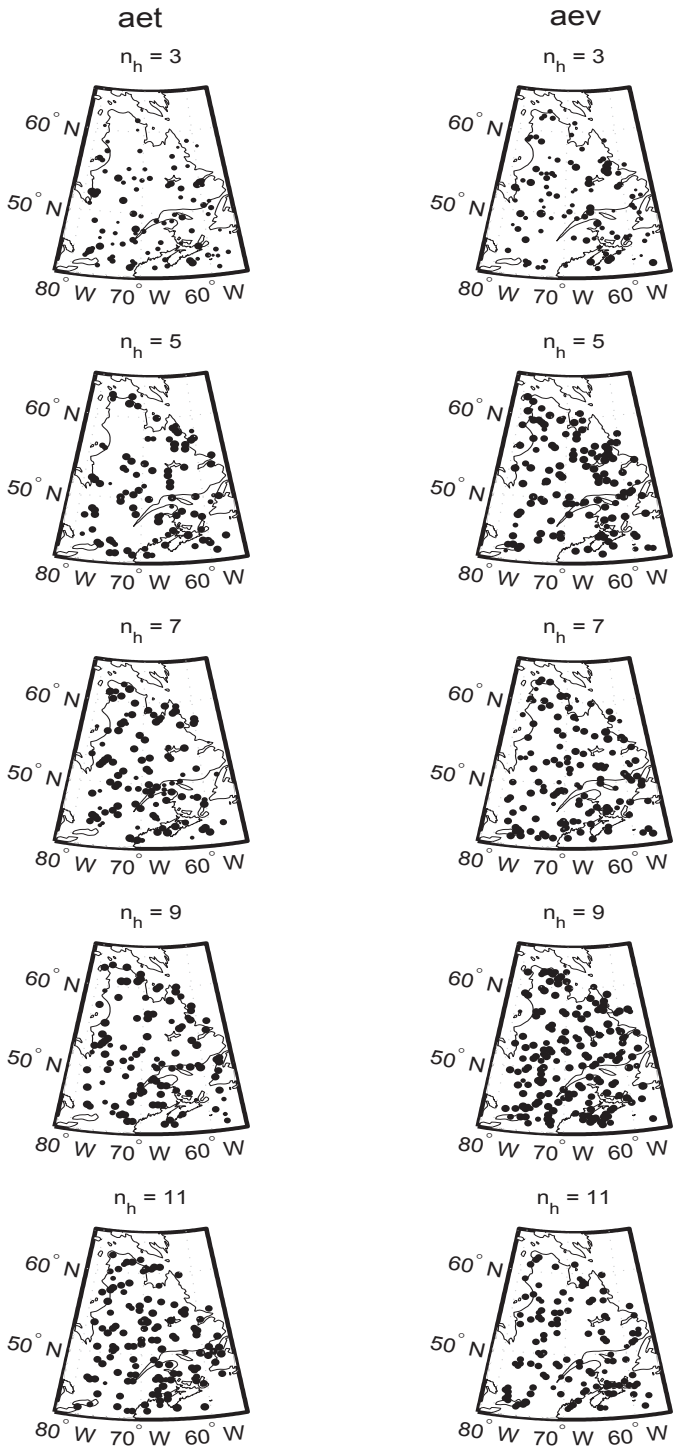

Figure 5.20: For *aet* and *aev* simulation, nonstationary locations given by the new stationarity test proposed in Chapter 3. The marker showing each grid point is scaled according to the respective $\mathcal{I}_{\mathrm{NS}}$ value computed for the location.

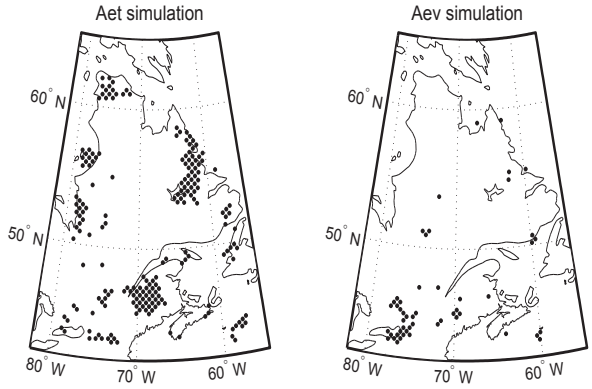

Figure 5.21: Results of applying the KPSS test to the maximum annual daily rainfall time series corresponding to *aet* and *aev* simulations.

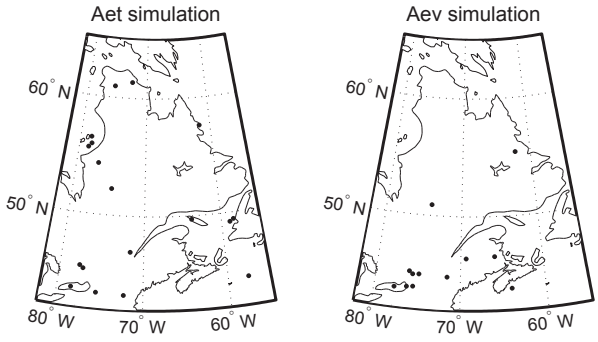

Figure 5.22: Intersection between the nonstationary grid points found by the KPSS test and the new stationarity test, by considering a window of length $n_{\mathrm{h}} = 9$.

By using the KPSS test, on the other hand, we have found 201 and 50 nonstationary grid points for *aet* and *aev* simulations, respectively. These locations are shown in Fig. 5.21. We have verified if there was any intersection between the locations found by the KPSS test and those accused by the new stationarity test of Chapter 3. If we consider for the latter all values of n_{h}, none intersection can be found. However, by choosing a specific window length n_{h}, several nonstationary grid points can be found (17 and 11 for *aet* and *aev* simulations, respectively). Fig. 5.22 illustrates the intersections between the grid points shown in Fig. 5.20 for $n_{\mathrm{h}} = 9$, and the ones classified as "nonstationary" by the KPSS test.

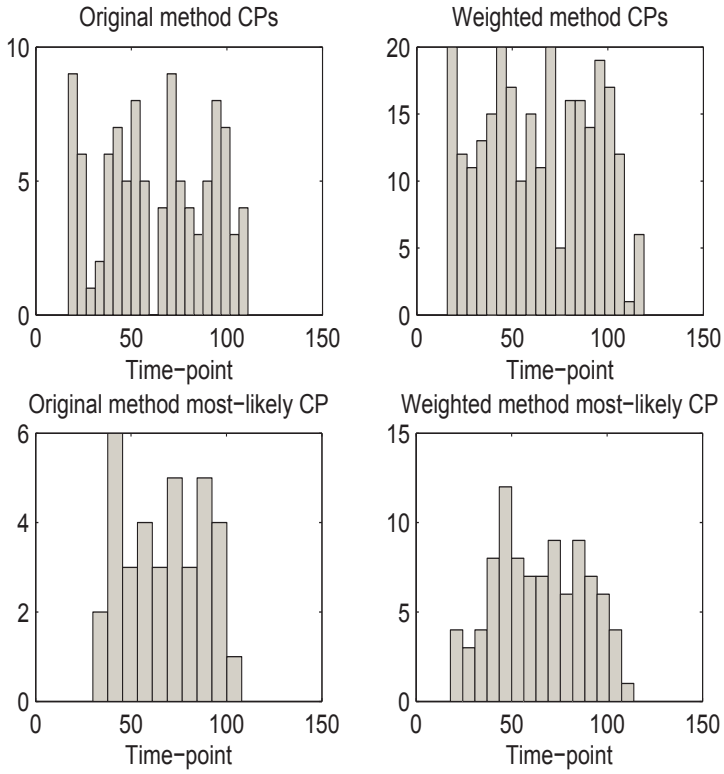

Figure 5.23: Histograms of the CPs and most likely CPs returned by the method. The nonstationary locations tested are those of *aet* simulation found by the *surrogate-based technique*.

5.3 Applying the change point detection algorithm

In this Section, we show the results of applying the CP detection framework proposed in Chapter 4. We have tested all grid points where the stationarity was rejected by the tests of Chapter 2 and Chapter 3. In Fig. 5.23 we present the histograms of the estimated CPs and most likely CPs for the nonstationary grid points found by the surrogate-based technique when testing the *aet* simulation. The corresponding results for *aev* simulation are illustrated in Fig. 5.24. The results for the new stationarity test are shown in Fig. 5.25 and Fig. 5.26 for *aet* and *aev* simulations, respectively.

It can be seen that the distributions of possible CPs and most likely CPs are too wide, and it is difficult to draw some conclusions from them. However, if we compute the mean values of the most likely CP for the cases described above, we can see that they are around $t = 60$ (or the year 2020) for each case. The estimated mean and mode from the histograms of the most likely CP are shown in Table 5.5.

We could narrow the scope of the search even further, and analyze, for instance, the CPs of the grid points that correspond only to the largest values of AMD (see Fig. 5.8,

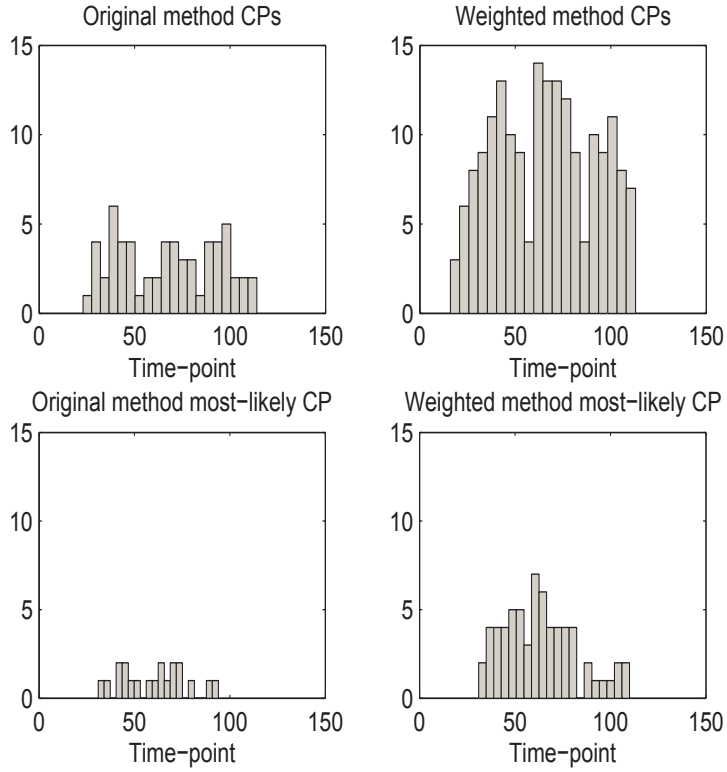

Figure 5.24: Histograms of the CPs and most likely CPs returned by the method. The nonstationary locations tested are those of *aev* simulation found by the *surrogate-based technique*.

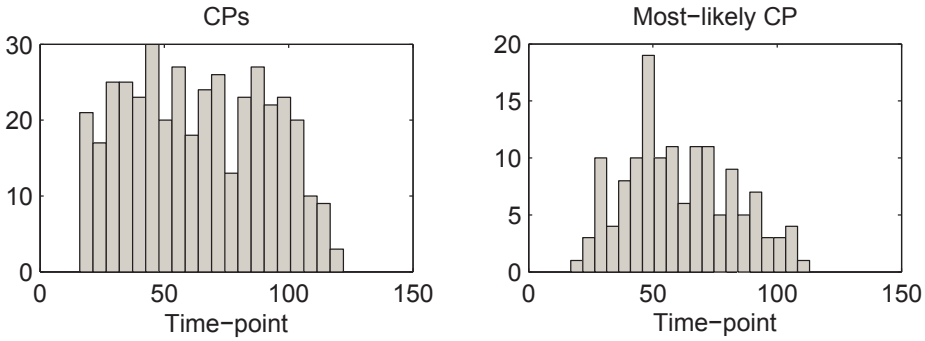

Figure 5.25: Histograms of the CPs and most likely CPs returned by the method. The nonstationary locations tested are those of *aev* simulation found by the new stationarity test.

Table 5.5: Mean and mode of the estimated CPs (most likely CP) of the nonstationary locations found by the two stationarity tests developed in this Thesis.

Stationarity test used	Surrogate-based technique		New test for trend-based nonstationarity	
Simulation	*aet*	*aev*	*aet*	*aev*
Mean	67	60	61	63
Mode	41	65	47	55

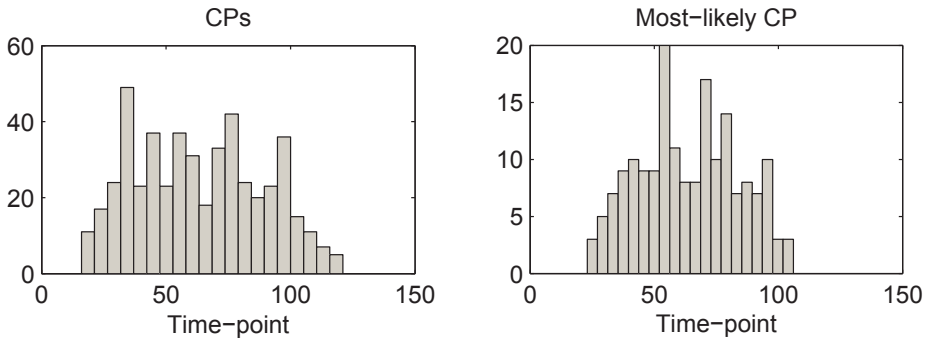

Figure 5.26: Histograms of the CPs and most likely CPs returned by the method. The nonstationary locations tested are those of *aev* simulation found by the new stationarity test.

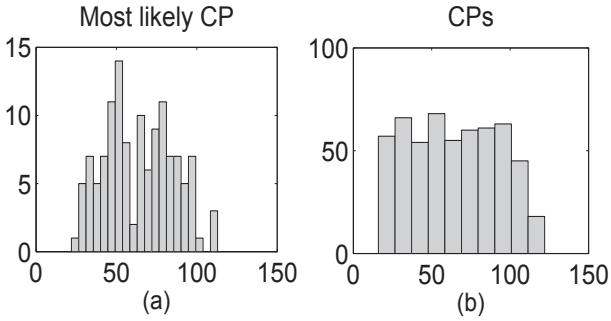

Figure 5.27: For the largest AMD cases: (a) histograms of the most likely CPs, (b) histograms of the CPs.

Fig. 5.9, Fig. 5.10 and Fig. 5.11). Fig. 5.27 illustrates the histogram of the estimated CPs and most likely CPs for the largest AMD cases. Notice in Fig. 5.27(a) that we have indeed a concentration around 60 for the most likely CP. In fact, the estimated mean and mode for Fig. 5.27(a) and (b), were, respectively, 63 and 73, and 64 and 74. Thus, the estimated mean value of the CP is in accordance with the ones of Table 5.5.

For observing the behavior of the change scores for different values of w, n and t, we have selected 8 locations belonging to the largest AMD ones for all the distances, and we have computed the CP scores for the selected range of the parameters. The coordinates of the chosen grid points are given in Table 5.6.

In Fig. 5.28 and Fig. 5.29, we illustrate the collection of CP scores over w, n and t,

Table 5.6: Coordinates of 8 grid points that lead to the largest AMD values for all distances.

Location	1	2	3	4	5	6	7	8
Latitude	60.8304	44.1138	48.1606	45.9826	48.5020	59.6040	57.8619	58.3774
Longitude	-77.8633	-57.3258	-59.2250	-64.4139	-73.3617	-73.5587	-74.4232	-75.5866

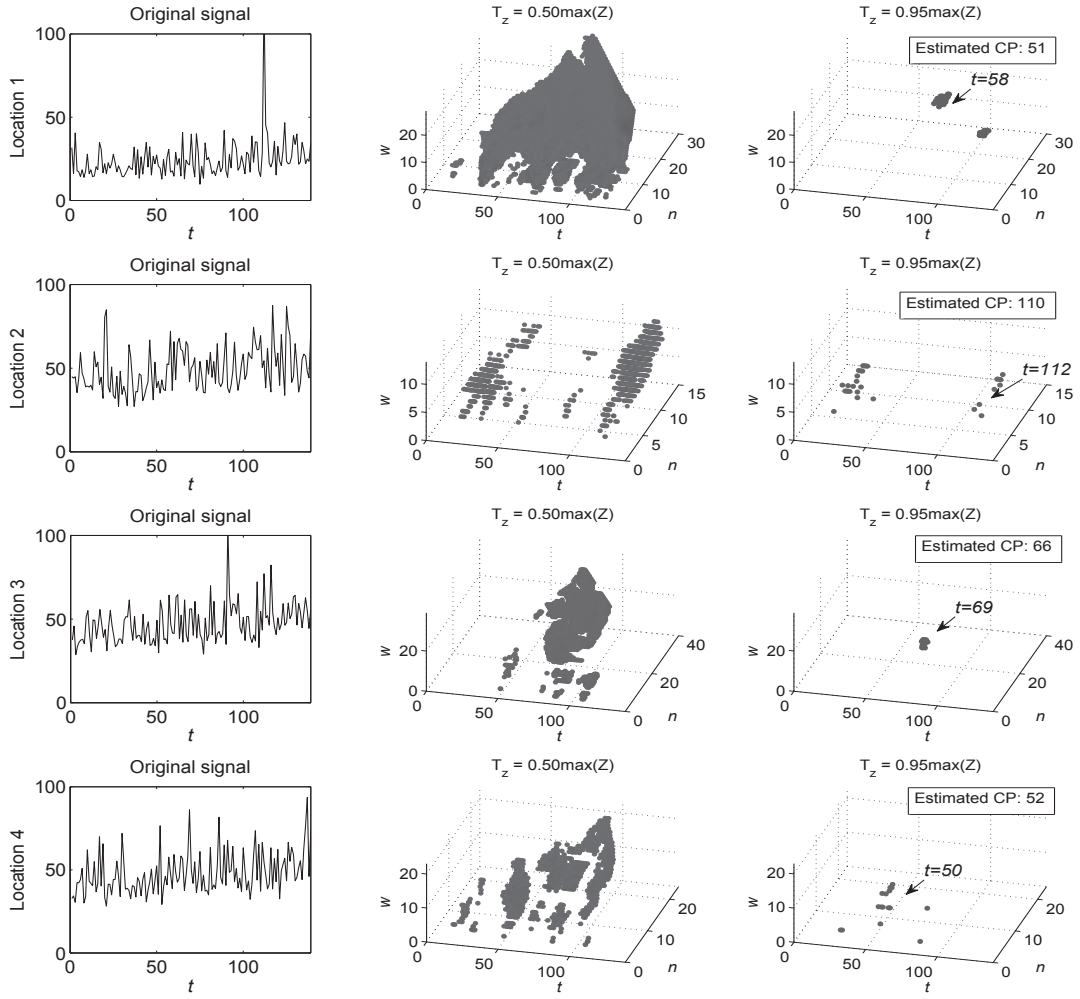

Figure 5.28: Change scores over w, n and t for the first 4 locations of Table 5.6 for different levels of significance.

after applying the procedure for selecting significant patterns, for the 8 locations shown in Table 5.6. The most likely CPs obtained for these locations are shown in the plots for each case. Notice that after choosing the significance of the change pattern, there is a concentration of CP scores around the respective time point relative to the CP.

5.3.1 Conclusions

In this Chapter, we have applied the techniques developed in this Thesis for testing the annual *maximum* daily precipitation obtained from two simulations of the Canadian Regional Climate Model (CRCM 4.2.3). The rainfall time series tested specifically in this Chapter correspond to 1631 grid points spread over the province of Quebec (QC), Canada, and have a time span of 139 years (from December 1960 to November 2100).

An interesting result is that the total number of nonstationary grid points found for

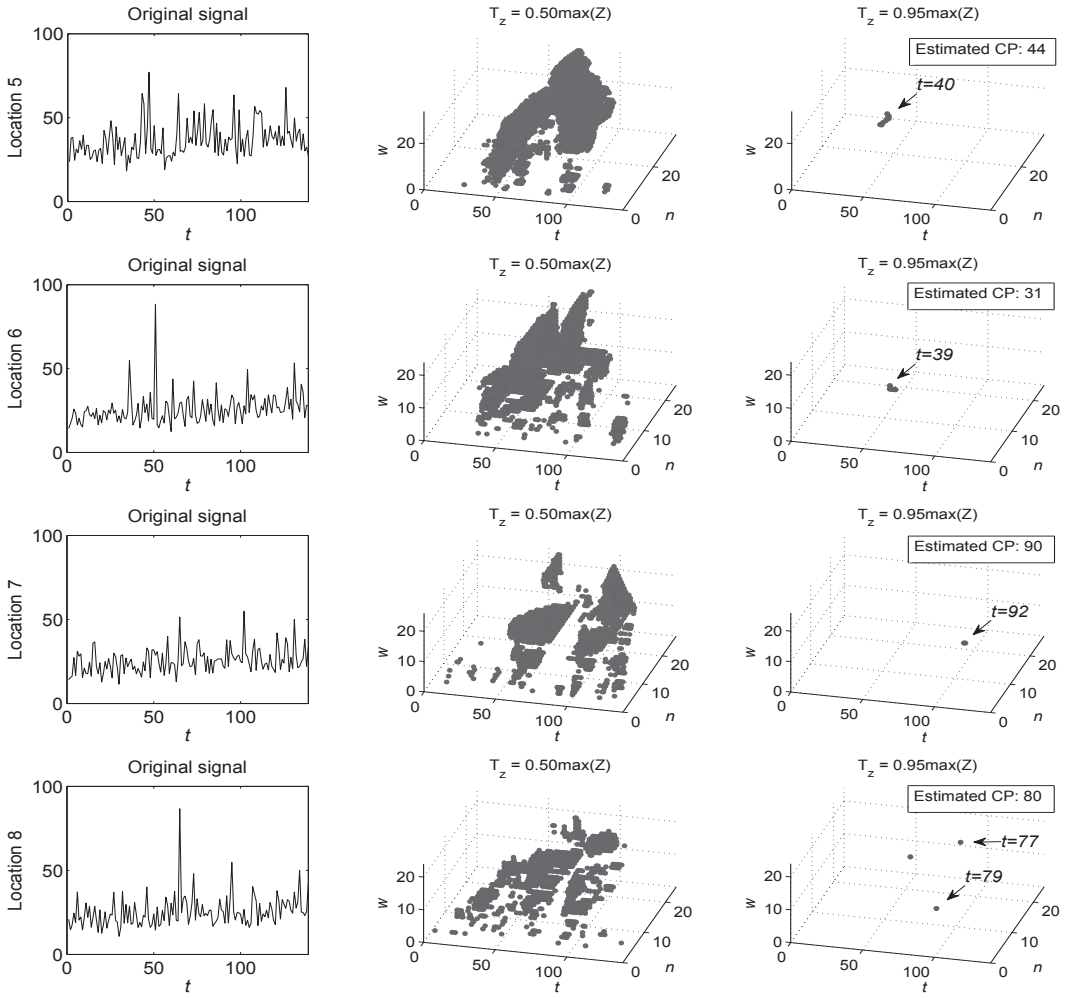

Figure 5.29: Change scores over w, n and t for the last 4 locations of Table 5.6 for different levels of significance.

each simulation (*aet* and *aev*) tends to be close to each other for the stationarity tests considered. Moreover, the weighting procedure that is proposed for the surrogate-based method, has reasonably fulfilled the required conditions for improving the stationarity test.

About the performance of the stationarity test based on surrogates, we have observed that, in general, there exist many nonstationary grid points spread over the North and the South of Quebec, but not in the central area. More precisely, for the majority of the cases, there is a concentration of nonstationary grid points having large AMD values (*i.e.* which are unlikely to change if we apply the test again), clustered in the following regions of Quebec: Southeast (over the Nova Scotia province), Southwest (over the North of Ottawa) and Northwest (in the South of the Ungava Peninsula). These findings call for a more careful analysis of the climatic time series over that areas.

The result of applying the new stationarity test for slowly-varying nonstationarities is a little different from what we have obtained for the surrogate-based techniques. There are many nonstationary grid points over the central area of Quebec. Such difference could be possibly explained by the fact that the two stationarity tests are designed for detecting different types of nonstationarities, and the nonstationary behavior detected by one at a given grid point, will be not necessarily detected by the other. However, even for the new stationarity test, a number of nonstationary locations over the regions mentioned above was detected. Other stationarity tests for detecting slowly-varying nonstationarities did not accuse as much nonstationary locations as the method developed in Chapter 3 (the Kay's test could not detect any grid point, and the KPSS test had only a few grid points intersecting the ones accused by the proposed approach).

The CP analysis over the nonstationary locations has indicated that there exists a high variability in the CP detected for each grid point. However, the expected value of the CP has shown to be around $t = 60$, *i.e.* the year 2020, for both stationarity tests and simulations. Finally, for the cases with the largest AMD values, the CP scores could be easily isolated by selecting a large significance level for the change pattern.

Chapter 6

Conclusions and perspectives

In the analysis of real world signals, we are often faced with a situation where we do not know whether a change occurred nor do we have any idea where the possible change point could be. However, a large amount of work assumes stationarity for applying standard algorithms, and the stationary case is well-defined from a theoretical perspective. On the other hand, all real world processes are *a priori* nonstationary, and in the majority of the cases this assumption turns out to be true. Hence, this has led to the development of techniques designed specifically for nonstationary situations. In this Thesis, we have focused on methods suitable for being applied to environmental processes, more specifically, hydro-meteorological ones. The techniques that have been developed in this Thesis, present some characteristics that are expected from a method to be applied to real world data (*i.e.* be nonparametric, data-driven, sensitive to first and second-order changes). Furthermore, in real world applications, we not only want to test for stationarity, but also to detect the time points (if any) where the changes take place in the data. Thus, in this Thesis, we have approached the stationarity test and the change point problem separately, by proposing a number of contributions and new approaches to these topics.

As the first part of the thesis, in Chapter 2, we have studied an existing stationarity test developed in TF domain that is nonparametric and data-driven. This method makes use of surrogates resampling for characterizing the statistics of the null-hypothesis of stationarity. In Chapter 2, we have proposed various contributions to the original approach, which allowed for:

- The identification of a proper class of distances to be used by the method.
- An improved detection of nonstationarities (when using the proper class of distances), specially first-order nonstationarities.

- The development of a measure to evaluate the changing results that may appear by applying the stationarity test sequentially.

Regardless of the improvements brought by the modifications proposed in Chapter 2, computing a TF representation for each surrogate resample still requires a large amount of computational work. Also, it could be pointed out that the surrogate-based technique does not present high classification accuracies when testing short time series with first-order or slowly-varying nonstationarities, simply because the surrogate-based framework has not been designed to this end. However, detecting such forms of nonstationarity is of major importance in many real world applications. Therefore, in Chapter 3 we have proposed a novel stationarity test, which is nonparametric, data-driven, and more sensitive sensitive to first-order evolutions than other nonparametric methods. Also, the proposed technique works for short signals.

The new stationarity test has been designed for testing a specific nonstationary behavior: the presence of a trend and/or an evolution in the local energy of the signal. To do so, we have proposed to test for trends in the time marginal, which is estimated directly by numerical convolution with a proper window of analysis (in this case, the Hermite functions). The trend itself is estimated by means of the EMD, and the importance of the trend seen in the time marginal is measured by making use of the so-called trend importance estimator ($\widehat{\theta}_{\mathrm{TI}}$). In order to perform the hypothesis test, we have proposed to use block bootstrapping to obtain the distribution of $\widehat{\theta}_{\mathrm{TI}}$ under the null hypothesis. Such distribution has been approximated by a pdf belonging to the GEV family. The adherence of this pdf has been verified by means of an analysis in asymptotic regime and a goodness-of-fit test. An index of nonstationarity ($\mathcal{I}_{\mathrm{NS}}$) has also been proposed.

The method that has been developed in Chapter 3 still leaves room for many possible improvements. As the proposed methodology encompasses a number of aspects, we have categorized some possible points that could be addressed in the future as follows:

- Index of nonstationarity ($\mathcal{I}_{\mathrm{NS}}$): one could identify the theoretical conditions for which the convexity of $f(\mathbf{e})$ (see (3.31) and (3.30)) is guaranteed. In general, further work needs to be done in deriving a robust index of nonstationarity for the stationarity test based on EMD. A possibility would be to improve the characterization of the terms that are used as arguments to parametrize the objective function.

- The distribution of $\widehat{\theta}_{\mathrm{TI}}$ under the null hypothesis: the adherence of the GEV model has been verified empirically, so the determination of the theoretical distribution of $\widehat{\theta}_{\mathrm{TI}}$ still remains an open question. A possibility would be to work with the

expanded expression for $\widehat{\theta}_{\mathrm{TI}}$ in (3.23), while taking into consideration some properties of the IMFs reported in [103].

- Speed up the computational time: although the computational time was reduced by skipping the part of computing a TF representation for each surrogate, the algorithm still takes a while to run. The most time-consuming part is by far the estimation of the trend by means of EMD, so one could think of a way to speed up the EMD algorithm. Other alternative would be to perform the test with an approximated asymptotic distribution of $\widehat{\theta}_{\mathrm{TI}}$ under the null hypothesis, since we have observed that parameters of the GEV distribution often tend to specific values (0.5, 0.02, and 1 for ϵ, σ and μ, respectively), as the number of bootstrap resamples increases.

After presenting the contributions to the nonparametric tests for stationarity, we have proposed in Chapter 4 an alternative framework for CP detection, which is based on the RSST. This technique is nonparametric and allows for the detection of multiple CPs. However, the full application of the RSST to real world data has been held captive by some limitations. For example, the RSST requires the specification of two parameters, whose values are chosen by means of visualization or domain knowledge. Unfortunately, these are often impracticable when testing environmental data. Thus, in Chapter 4 we have proposed a modified framework based on the RSST which is better suited to be a applied to real world signals. More precisely, we have proposed the following modifications:

- To represent the output of the RSST in the space spanned by its two parameters, as different change patterns can be captured if we sweep over w and n. Then, we have proposed a stopping criterion for the sweep, which is a crucial point for reducing the computational time of the algorithm.

- Better results are obtained by selecting significant change patterns in the signal. Hence, we have proposed a simple strategy to filter the most significant CP scores (the ones representing actual changes).

- For estimating the CP, we have proposed to measure the uncertainty related to the CP scores at each temporal "slice" in the space spanned by the two parameters of the RSST. To do so, we have proposed to compute the conditional entropy at each time instant, and to compute the CP by searching for the instants where the local *maxima* of the conditional entropy vector occur.

The modified approach for CP detection is more versatile than other ones in the literature, since it allows for: i) the detection of multiple CPs, ii) the visualization of the intervals where major changes in the signal occur as function of the windows of analysis and time, and iii) the detection of first and second-order changes. However, the proposed methodology calls for a deeper analysis or further improvements in many points, for example:

- The technique presented in Chapter 4 is considerably time consuming if one is testing larger time series. Solutions to fasten the methodology need to be searched. A possibility would be to find another stopping criterion for the sweep.
- The efficiency of the procedure to sweep over the two parameters (w and n) has been only verified empirically. Thus, a proper theoretical basis could be developed in the future. For instance, one could demonstrate why p_i decays after attaining the *maximum* at ζ. Also, one could suggest to increase w and n not evenly, but instead verifying if $\max p_i$ could be reached faster through other ways.

Finally, in Chapter 5, we have applied the various methods presented in this Thesis to an environmental dataset, more specifically, the time series corresponding to annual *maximum* daily precipitation simulated from the Canadian Regional Climate Model (CRCM 4.2.3). The time series have a time span of 139 years (December 1960 to November 2100), and correspond to 1631 grid points spread over the province of Quebec (QC), Canada. In the experimental study, we have observed that the two stationarity tests presented in the Thesis could detect nonstationary grid points in different areas (possibly because the two tests are designed for detecting different forms of nonstationarity). However, in general, there exist a concentration of nonstationary locations in the Southeast, Southwest and Northwest of Quebec, which calls for a deeper investigation of the climatic time series over that areas. About the detection of CPs, despite of the high variability observed in the outcomes, the expected value of the CP has shown to be around the year 2020 for all cases.

Appendix A

Explaining the weighting technique

A.1 Why weighting distances improves the performance of the stationarity test?

Let us consider again the nonstationary Gaussian processes plotted in Fig. 2.5 together with their TF representation. Theses plots are reproduced again Fig. A.1.

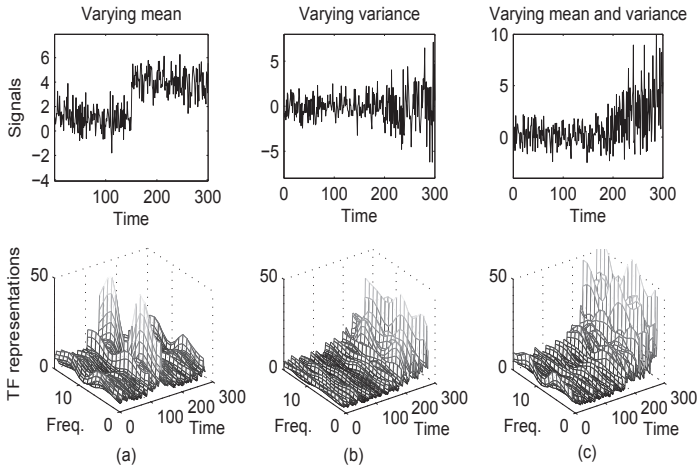

Figure A.1: Different nonstationary signals and their particular TF representation. (a) Varying mean. (b) Varying variance. (c) Varying mean and variance.

Although the time-varying spectra shown in Fig. A.1 refer to different nonstationary signals, one could identify a particular pattern: a significant portion of the spectral content is concentrated over a given time interval. For a given TF spectrum $\{S(t_n, f), n = 1, ..., N\}$, let us define a vector containing these time instants as $\tau = t_i, ..., t_f$, where $[t_i, t_f] \in [1, N]$. We consider τ as the vector presenting the time instants where majority of the spectral content of the signal. Then, we assume to have an accumulation or a

concentration of the spectral content of $S(t,f)$ in τ, if the local spectra are significantly greater than the marginal spectrum, in a way that the distance between them are greater for $\tau = t_i, ..., t_f$ than for the rest of the observation interval. More precisely, if we define the total excess of the local spectra over the global marginal spectrum by $\sum_f [S(t_n,f) - \langle S(t,f)\rangle_n]$, where $\langle S(t,f)\rangle_n$ is obtained by marginalization over all time instants, we say that $S(t_n,f)$ is concentrated around τ if:

$$\sum_f [S(t_n,f) - \langle S(t,f)\rangle_n]_{t_n\in\tau} \geq \sum_f [S(t_n,f) - \langle S(t,f)\rangle_n]_{t_n\notin\tau} \tag{A.1}$$

while the values taken by the distance vector should also increase in τ:

$$D\left[S(t_n,f), \langle S(t,f)\rangle_n\right]_{t_n\in\tau} \geq D\left[S(t_n,f), \langle S(t,f)\rangle_n\right]_{t_n\notin\tau}. \tag{A.2}$$

It is easy to see that this temporal structure is reflected on the time marginal $y(t_n)$, as (A.1) can be written as:

$$y(t_n)\Big|_{t_n\in\tau} \geq y(t_n)\Big|_{t_n\notin\tau}, \tag{A.3}$$

since $y(t_n) = \sum_f S(t_n,f)$, and to the fact that $\sum_f \langle S(t,f)\rangle_n$ is a constant invariant in time. Note that (A.1) and (A.2) simply say that the largest values of the distance vector occur where the TF spectrum is most concentrated. Notice that this idealized scheme is fairly well followed by the nonstationary signals shown Fig. 2.5. A simple example of this scheme is also illustrated in Fig. A.2, where a given time-varying spectrum exhibiting the aforementioned temporal structure is shown. The "slices" in time representing the local spectra over τ, and the relation with the distance in time are also illustrated.

The key point of using the same vector $\widetilde{y}_n$ to weight the distances $\{c_n^{(s_i)}, n = 1..., N\}$ of the surrogate set, is that one is in fact *adding a dependence structure* to the collection of $i = 1, ..., I$ distances, which is dictated by the time marginal. Furthermore, and also important, there should exist a *positive covariance* between consecutive weighted distances $\{\widetilde{c}_n^{(s_i)}, i = 1..., I\}$, whereas, the more nonstationary the signal (or the stronger the observed temporal structure in $\widetilde{y}_n$), the greater the covariance should be. In other words, for any $i, l \in [1, I]$, greater of values of $\widetilde{c}_n^{(s_i)}$ will mainly correspond with the greater values of $\widetilde{c}_n^{(s_l)}$ (as both random vectors are being weighted by the same vector $\widetilde{y}_n$), which characterizes a positive covariance. The latter can be also easily demonstrated by using the expression for the covariance. Let $\{\widetilde{c}_n^{(s_i)} = c_n^{(s_i)}\widetilde{y}_n, n = 1, ..., N\}$ and $\{\widetilde{c}_n^{(s_l)} = c_n^{(s_l)}\widetilde{y}_n, n = 1, ..., N\}$ be two random variables, where $c_n^{(s_i)}$ and $c_n^{(s_l)}$ are i.i.d. and the weighting vector $\widetilde{y}_n$ is independent of $c_n^{(s_i)}$ and $c_n^{(s_l)}$. Hence, the covariance can be

expressed as follows:

$$\mathbb{C}\text{ov}\{\widetilde{c}_n^{(s_i)}, \widetilde{c}_n^{(s_l)}\} = \mathbb{E}\{\widetilde{c}_n^{(s_i)}\widetilde{c}_n^{(s_l)}\} - \mathbb{E}\{\widetilde{c}_n^{(s_i)}\}\mathbb{E}\{\widetilde{c}_n^{(s_l)}\},$$

$$\mathbb{C}\text{ov}\{\widetilde{c}_n^{(s_i)}, \widetilde{c}_n^{(s_l)}\} = \mathbb{E}\{c_n^{(s_i)}\widetilde{y}_n c_n^{(s_l)}\widetilde{y}_n\} - \mathbb{E}\{c_n^{(s_i)}\widetilde{y}_n\}\mathbb{E}\{c_n^{(s_i)}\widetilde{y}_n\},$$

$$\mathbb{C}\text{ov}\{\widetilde{c}_n^{(s_i)}, \widetilde{c}_n^{(s_l)}\} = \mathbb{E}\{\widetilde{y}_n^2\}\mathbb{E}\{c_n^{(s_i)}c_n^{(s_l)}\} - \mathbb{E}^2\{\widetilde{y}_n\}\mathbb{E}\{c_n^{(s_i)}\}\mathbb{E}\{c_n^{(s_i)}\},$$

where, since $c_n^{(s_i)}$ and $c_n^{(s_l)}$ are i.i.d., we have

$$\mathbb{C}\text{ov}\{\widetilde{c}_n^{(s_i)}, \widetilde{c}_n^{(s_l)}\} = [\mathbb{E}\{\widetilde{y}_n^2\} - \mathbb{E}^2\{\widetilde{y}_n\}]\mathbb{E}\{c_n^{(s_i)}\}\mathbb{E}\{c_n^{(s_i)}\} = \mathbb{V}\text{ar}\{\widetilde{y}_n\}\mathbb{E}^2\{c_n^{(s_i)}\},$$

which should be positive.

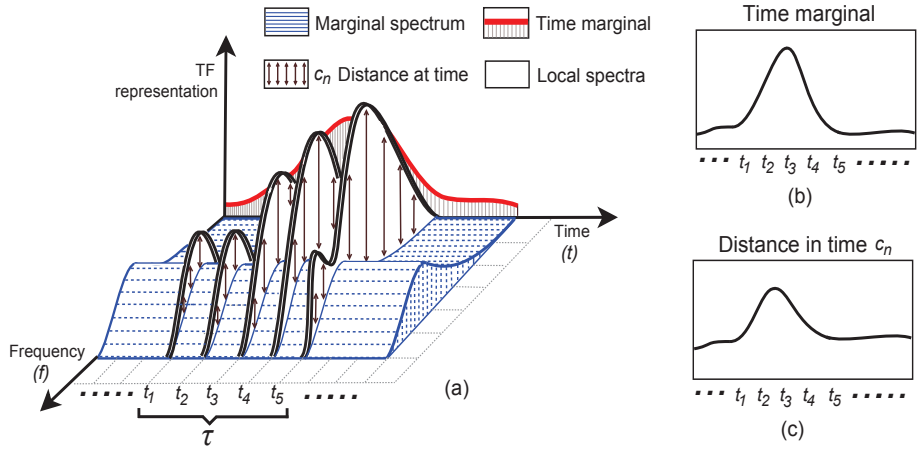

Figure A.2: Example of a TF representation exhibiting a concentration of the spectrum and the temporal structure observed in the time marginal and in the distance in time.

Based on the considerations above, one can show that the conditions presented in (2.27) for the gamma parameters are attended by weighting the distances with $\widetilde{y}_n$. First, recall that the gamma parameters k and θ are estimated from $\{\Theta_0(i) = \mathbb{V}\text{ar}\{\widetilde{c}_n^{(s_i)}\}, i = 1, ..., I\}$ by maximum likelihood. It can be shown that, for the gamma distribution, the log-likelihood function is given by:

$$\ell(k, \theta) = (k-1)\sum_{i=1}^{I} \ln \Theta_0(i) - \frac{1}{\theta}\sum_{i=1}^{I}\Theta_0(i) - Ik\ln(\theta) - I\ln\Gamma(k) \tag{A.4}$$

By taking the derivative with respect to θ and setting it to zero, we can easily find the maximum likelihood estimator of the θ parameter:

$$\widehat{\theta} = \frac{1}{kI}\sum_{i=1}^{I}\Theta_0(i) = \frac{1}{kI}\sum_{i=1}^{I}\mathbb{V}\text{ar}\{\widetilde{c}_n^{(s_i)}\}. \tag{A.5}$$

By using some basic variance rules, we can see that:

$$\mathbb{V}\mathrm{ar}\left\{\sum_{i=1}^{I} c_n^{(s_i)}\right\} = \sum_{i=1}^{I} \mathbb{V}\mathrm{ar}\{c_n^{(s_i)}\} + 2\sum_{i<l}^{I} \mathbb{C}\mathrm{ov}\{c_n^{(s_i)}, c_n^{(s_l)}\}. \tag{A.6}$$

Hence we could rewrite (A.5) as follows:

$$\widehat{\theta} = \frac{1}{kI}\sum_{i=1}^{I} \mathbb{V}\mathrm{ar}\{c_n^{(s_i)}\} = \frac{1}{kI}\mathbb{V}\mathrm{ar}\left\{\sum_{i=1}^{I} c_n^{(s_i)}\right\} - \frac{2}{kI}\sum_{i<l}^{I} \mathbb{C}\mathrm{ov}\{c_n^{(s_i)}, c_n^{(s_l)}\}. \tag{A.7}$$

It is simple to express the condition of (2.26) for Θ_1/θ and Θ_1'/θ' in terms of variances. To do so, note that $\Theta_1 = \mathbb{V}\mathrm{ar}\{c_n^x\}$ and $\Theta_1' = \mathbb{V}\mathrm{ar}\{\widetilde{c}_n^x\}$, where c_n^x and $\widetilde{c}_n^x$ are, respectively, the original and the modified distance vectors of the signal itself. Thus, (2.26) can be written as:

$$\frac{\mathbb{V}\mathrm{ar}\{\widetilde{c}_n^x\}}{\frac{1}{k'I}\mathbb{V}\mathrm{ar}\left\{\sum_{i=1}^{I} \widetilde{c}_n^{(s_i)}\right\} - \frac{2}{k'I}\sum_{i<l}^{I} \mathbb{C}\mathrm{ov}\{\widetilde{c}_n^{(s_i)}, \widetilde{c}_n^{(s_l)}\}} > \frac{\mathbb{V}\mathrm{ar}\{c_n^x\}}{\frac{1}{kI}\mathbb{V}\mathrm{ar}\left\{\sum_{i=1}^{I} c_n^{(s_i)}\right\} - \frac{2}{kI}\sum_{i<l}^{I} \mathbb{C}\mathrm{ov}\{c_n^{(s_i)}, c_n^{(s_l)}\}} \tag{A.8}$$

Since the surrogates are i.i.d.,the original distance vectors $\{c_n^{(s_i)}, i = 1..., I\}$ computed for consecutive surrogates, should not present a significance dependence structure, in a way that the terms $\{\mathbb{C}\mathrm{ov}\{c_n^{(s_i)}, c_n^{(s_l)}\}, i < j\}$ in (A.8) should be close to zero, independently of the stationarity/nonstationarity of the signal. This is not the case for the weighted distances, as the terms $\mathbb{C}\mathrm{ov}\{\widetilde{c}_n^{(s_i)}, \widetilde{c}_n^{(s_l)}\}$ should take large positive values for the case of nonstationarity (as seen in (A.2) and (A.3)). By replacing the covariance terms, one could rewrite (A.8) as follows:

$$\frac{\mathbb{V}\mathrm{ar}\left\{\sum_{i=1}^{I} c_n^{(s_i)}\right\}}{\mathbb{V}\mathrm{ar}\left\{\sum_{i=1}^{I} \widetilde{c}_n^{(s_i)}\right\} - \sum_{i<l}^{I} \mathbb{C}\mathrm{ov}\{\widetilde{c}_n^{(s_i)}, \widetilde{c}_n^{(s_l)}\}} > \frac{k}{k'}\frac{\mathbb{V}\mathrm{ar}\{c_n^x\}}{\mathbb{V}\mathrm{ar}\{\widetilde{c}_n^x\}}. \tag{A.9}$$

Although the right-hand side of (A.9) has to be greater than one, and also should increase for for the case of nonstationarity (where $k' > k$), the left-hand side should take *much larger values* if the time-varying spectrum undergoes a structured evolution in time. In this case, not only the variance computed in the numerator will grow, but also the covariance terms $\mathbb{C}\mathrm{ov}\{\widetilde{c}_n^{(s_i)}, \widetilde{c}_n^{(s_l)}\}$ will keep the value of the denominator low, guarantying that the inequality holds. According to this, the condition $\Theta_1'/\theta' > \Theta_1/\theta$ shown (2.26) should hold if the TF representation exhibits a structured evolution in time.

In order to verify the condition for the shape parameter ($k' < k$), we first substitute

(A.5) in (A.10) to obtain:

$$\ell(k) = (k-1)\sum_{i=1}^{I} \ln\left[\Theta_0(i)\right] - Ik - Ik \ln\left[\frac{1}{kI}\sum_{i=1}^{I}\Theta_0(i)\right] - I\ln\Gamma(k). \tag{A.10}$$

Now, by taking the derivative with respect to k and equaling to zero, we obtain:

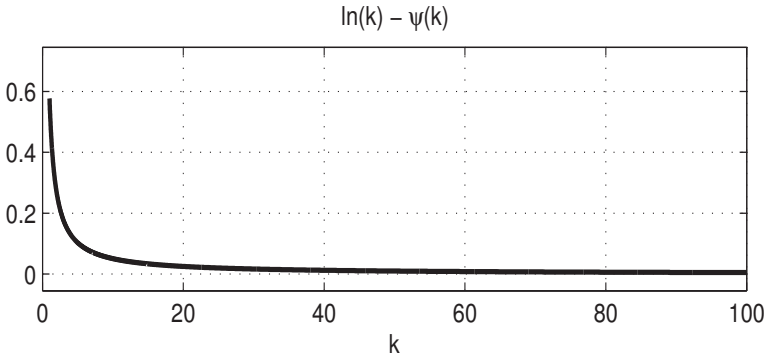

Figure A.3: Plot of $f(k) = \ln(k) - \psi(k)$ for $k > 0$.

$$\ln(k) - \psi(k) = \ln\left[\frac{1}{I}\sum_{i=1}^{I}\Theta_0(i)\right] - \frac{1}{I}\sum_{i=1}^{I}\ln\left[\Theta_0(i)\right], \tag{A.11}$$

where $\psi(k) = \Gamma'(k)/\Gamma(k)$ is the digamma function. Different from (A.5), in (A.11) there is no analytical solution for k. However, if we define $f(k) = \ln(k) - \psi(k)$, it could be verified that $f(k)$ decreases monotonically for $k > 0$ (see Fig. A.3). Therefore, for any positive $k_1, k_2 \in \Re$ where $k_1 < k_2$, the inequality $f(k_1) < f(k_2)$ should hold. Note that (A.11) can be rewritten as:

$$\ln(k) - \psi(k) = f(k) = \ln\left[\frac{\frac{1}{I}\sum_{i=1}^{I}\Theta_0(i)}{\prod_{i=1}^{I}\Theta_0(i)^{\frac{1}{I}}}\right]. \tag{A.12}$$

The right-hand side of (A.12) stands for the logarithm of the arithmetic mean of $\Theta_0(i)$ divided by its geometric mean. Since $\{\Theta_0(i), i = 1, ..., I\}$ is formed by I variances of the distances computed for the surrogates, we can express (A.12) as:

$$\ln\left[\frac{\frac{1}{I}\sum_{i=1}^{I}\Theta_0(i)}{\prod_{i=1}^{I}\Theta_0(i)^{\frac{1}{I}}}\right] = \ln\left[\frac{\frac{1}{I}\sum_{i=1}^{I}\mathbb{V}\mathrm{ar}\{c^{(s_i)}\}}{\prod_{i=1}^{I}\mathbb{V}\mathrm{ar}\{c^{(s_i)}\}^{\frac{1}{I}}}\right]. \tag{A.13}$$

Moreover, if we use (A.6), it can be seen that (A.13) can be given as:

$$\ln\left[\frac{\frac{1}{I}\sum_{i=1}^{I}\mathbb{V}\mathrm{ar}\{c_n^{(s_i)}\}}{\prod_{i=1}^{I}\mathbb{V}\mathrm{ar}\{c_n^{(s_i)}\}^{\frac{1}{I}}}\right]=\ln\left[\frac{\frac{1}{I}\mathbb{V}\mathrm{ar}\left\{\sum_{i=1}^{I}c_n^{(s_i)}\right\}-\frac{2}{I}\sum_{i<l}^{I}\mathbb{C}\mathrm{ov}\{c_n^{(s_i)},c_n^{(s_l)}\}}{\prod_{i=1}^{I}\mathbb{V}\mathrm{ar}\{c_n^{(s_i)}\}^{\frac{1}{I}}}\right]. \tag{A.14}$$

According to (2.26), we need to have $k' < k$ for the case of nonstationarity. As $f(k) = \ln(k) - \psi(k)$ is monotonically decreasing for $k > 0$ (see Fig. A.3), the condition for k can be expressed as:

$$\ln\left[\frac{\frac{1}{I}\mathbb{V}\mathrm{ar}\left\{\sum_{i=1}^{I}\widetilde{c}_n^{(s_i)}\right\}-\frac{2}{I}\sum_{i<l}^{I}\mathbb{C}\mathrm{ov}\{\widetilde{c}_n^{(s_i)},\widetilde{c}_n^{(s_l)}\}}{\prod_{i=1}^{I}\mathbb{V}\mathrm{ar}\{\widetilde{c}^{(s_i)}\}^{\frac{1}{I}}}\right]>\ln\left[\frac{\frac{1}{I}\mathbb{V}\mathrm{ar}\left\{\sum_{i=1}^{I}c_n^{(s_i)}\right\}-\frac{2}{I}\sum_{i<l}^{I}\mathbb{C}\mathrm{ov}\{c_n^{(s_i)},c_n^{(s_l)}\}}{\prod_{i=1}^{I}\mathbb{V}\mathrm{ar}\{c^{(s_i)}\}^{\frac{1}{I}}}\right] \tag{A.15}$$

Expanding the product of variances in the denominator of (A.15) is a tedious task, as there will be many cross terms. On the other hand, we can express the product of variances between two consecutive distances $c_n^{(s_i)}$ and $c_n^{(s_l)}$ as follows:

$$\mathbb{V}\mathrm{ar}\{c_n^{(s_i)}\}\mathbb{V}\mathrm{ar}\{c_n^{(s_l)}\}=\frac{\left[\mathbb{V}\mathrm{ar}\{c_n^{(s_i)}\}+\mathbb{V}\mathrm{ar}\{c_n^{(s_l)}\}\right]^2-\left[\mathbb{V}\mathrm{ar}^2\{c_n^{(s_i)}\}+\mathbb{V}\mathrm{ar}^2\{c_n^{(s_l)}\}\right]}{2}, \tag{A.16}$$

where, by using (A.6), we could see that:

$$\mathbb{V}\mathrm{ar}\{c_n^{(s_i)}\}\mathbb{V}\mathrm{ar}\{c_n^{(s_l)}\}=\frac{\left[\mathbb{V}\mathrm{ar}\{c_n^{(s_i)}+c_n^{(s_l)}\}-2\mathbb{C}\mathrm{ov}\{c_n^{(s_i)},c_n^{(s_l)}\}\right]^2-\left[\mathbb{V}\mathrm{ar}^2\{c_n^{(s_i)}\}+\mathbb{V}\mathrm{ar}^2\{c_n^{(s_l)}\}\right]}{2}. \tag{A.17}$$

The product of variances in the denominator of (A.15), can thus be expressed as:

$$\prod_{i=1}^{I}\mathbb{V}\mathrm{ar}\{c_n^{(s_i)}\}^{\frac{1}{I}}=$$

$$\underbrace{\left[\mathbb{V}\mathrm{ar}\{c_n^{(s_1)}\}\mathbb{V}\mathrm{ar}\{c_n^{(s_I)}\}\right]^{\frac{1}{I}}\left[\mathbb{V}\mathrm{ar}\{c_n^{(s_2)}\}\mathbb{V}\mathrm{ar}\{c_n^{(s_{I-1})}\}\right]^{\frac{1}{I}}\cdots\left[\mathbb{V}\mathrm{ar}\{c_n^{(s_{I/2})}\}\mathbb{V}\mathrm{ar}\{c_n^{(s_{I/2+1})}\}\right]^{\frac{1}{I}}}_{I/2\text{ terms}}=$$

$$\prod_{i=1}^{I/2}\left\{\frac{\left[\mathbb{V}\mathrm{ar}\{c_n^{(s_i)}+c_n^{(s_{I-i+1})}\}-2\mathbb{C}\mathrm{ov}\{c_n^{(s_i)},c_n^{(s_{I-i+1})}\}\right]^2-\left[\mathbb{V}\mathrm{ar}^2\{c_n^{(s_i)}\}+\mathbb{V}\mathrm{ar}^2\{c_n^{(s_{I-i+1})}\}\right]}{2}\right\}^{\frac{1}{I}}. \tag{A.18}$$

By replacing (A.18) in (A.15) we obtain the following:

$$\ln \frac{\frac{1}{I}\mathbb{V}\mathrm{ar}\left\{\sum_{i=1}^{I}\widetilde{c}_n^{(s_i)}\right\} - \frac{2}{I}\sum_{i<l}^{I}\mathbb{C}\mathrm{ov}\{\widetilde{c}_n^{(s_i)}, \widetilde{c}_n^{(s_l)}\}}{\underbrace{\prod_{i=1}^{I/2}\left\{\frac{\left[\mathbb{V}\mathrm{ar}\{\widetilde{c}_n^{(s_i)} + \widetilde{c}_n^{(s_{I-i+1})}\} - 2\mathbb{C}\mathrm{ov}\{\widetilde{c}_n^{(s_i)}, \widetilde{c}_n^{(s_{I-i+1})}\}\right]^2 - \left[\mathbb{V}\mathrm{ar}^2\{\widetilde{c}_n^{(s_i)}\} + \mathbb{V}\mathrm{ar}^2\{\widetilde{c}_n^{(s_{I-i+1})}\}\right]}{2}\right\}^{\frac{1}{I}}}_{f(k')}} >$$

$$\ln \frac{\frac{1}{I}\mathbb{V}\mathrm{ar}\left\{\sum_{i=1}^{I}c_n^{(s_i)}\right\} - \frac{2}{I}\sum_{i<l}^{I}\mathbb{C}\mathrm{ov}\{c_n^{(s_i)}, c_n^{(s_l)}\}}{\underbrace{\prod_{i=1}^{I/2}\left\{\frac{\left[\mathbb{V}\mathrm{ar}\{c_n^{(s_i)} + c_n^{(s_{I-i+1})}\} - 2\mathbb{C}\mathrm{ov}\{c_n^{(s_i)}, c_n^{(s_{I-i+1})}\}\right]^2 - \left[\mathbb{V}\mathrm{ar}^2\{c_n^{(s_i)}\} + \mathbb{V}\mathrm{ar}^2\{c_n^{(s_{I-i+1})}\}\right]}{2}\right\}^{\frac{1}{I}}}_{f(k)}}, \tag{A.19}$$

where we should have $f(k') \geq f(k)$ in order to guarantee the condition in (2.26) (i.e. $k' < k$). The terms $\mathbb{C}\mathrm{ov}\{\widetilde{c}_n^{(s_i)}, \widetilde{c}_n^{(s_{I-i+1})}\}$ in (A.19) take large positive values in case of nonstationarity, due to the dependence imposed by weighting with $\widetilde{y}_n$. On the other hand, $\mathbb{C}\mathrm{ov}\{c_n^{(s_i)}, c_n^{(s_{I-i+1})}\}$ should be much closer to zero, in a sense that one could verify $\mathbb{C}\mathrm{ov}\{\widetilde{c}_n^{(s_i)}, \widetilde{c}_n^{(s_{I-i+1})}\} \gg \mathbb{C}\mathrm{ov}\{c_n^{(s_i)}, c_n^{(s_{I-i+1})}\}$ for $i = 1, ..., I/2$. Note that the greater the covariances, the smaller the denominator of (A.19). As the terms in the denominator of (A.19) are multiplied out $I/2$ times, the influence of the covariance terms will be boosted significantly. Hence, the left-hand side of (A.19) will take much larger values whenever the values of the covariances are large (i.e. in case of nonstationarity), and one will have $f(k') > f(k)$. Hence, the condition $k' < k$ is verified.

A.2 Problems with weighting probability-based distances

As it has been shown in Table 2.3, by weighting the probability-based distances according to the procedure described in Section 2.6, we end up affecting the detection of nonstationarity. It happens due to the normalization of the spectra that is required for this class of distances. By normalizing the spectra $G(f)$ and $H(f)$ to the unity, we reduce drastically the range of possible values taken by the corresponding distances, which turns out to be a critical point for the method.

For all practical purposes it can be assumed that the distances of probability nature considered in this work will take values in the range $[0, 1]$. As illustration, we show in Fig. A.4 the histograms of the values taken by the probability-based distances ($D_{KL}(\cdot,\cdot)$, $D_{KM}(\cdot,\cdot)$ and $D_{GM}(\cdot,\cdot)$), before weighting, for all the cases tested with the surrogate

approach in this manuscript.

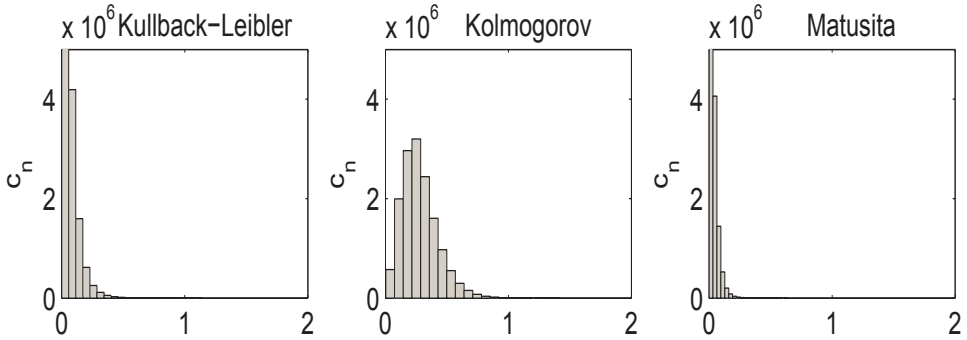

Figure A.4: Histograms of the values taken by the probability-based distances for all cases tested with the surrogate approach.

Notice that only a few outliers are greater than one, and as a matter of fact, the percentage of cases where $D(\cdot,\cdot) \geq 1$ is less than 0.05% for all cases. Hence, we can assume that $0 \leq D_{KL}(\cdot,\cdot) \leq 1$, $0 \leq D_{KM}(\cdot,\cdot) \leq 1$, and $0 \leq D_{GM}(\cdot,\cdot) \leq 1$. Thus, if we choose any of these distances for computing the vector $\{c_n, n = 1, ..., N\}$, we should have $0 \leq c_n \leq 1$ for every $n = 1, ..., N$. From this, if we define $\mathbb{E}\{c_n\} = \nu$, we could easily demonstrate the following:

$$\begin{gathered} 0 \leq c_n \leq 1 \Rightarrow 0 \leq c_n^2 \leq c_n, \\ 0 \leq \mathbb{E}\{c_n^2\} \leq \mathbb{E}\{c_n\} \Rightarrow \mathbb{E}\{c_n^2\} - \mathbb{E}^2\{c_n\} \leq \mathbb{E}\{c_n\} - \mathbb{E}^2\{c_n\}, \\ \mathbb{V}\mathrm{ar}\{c_n\} \leq \mathbb{E}\{c_n\}(1 - \mathbb{E}\{c_n\}) \Rightarrow \mathbb{V}\mathrm{ar}\{c_n\} \leq \nu(1-\nu). \end{gathered} \tag{A.20}$$

Also, since $0 \leq \nu \leq 1$, it is simple to verify the following:

$$\underset{\nu}{\mathrm{argmax}}[\nu(1-\nu)] = 0.5 \quad \text{and} \quad \max \nu(1-\nu) = 0.25, \tag{A.21}$$

Hence, we have $0 \leq \mathbb{V}\mathrm{ar}\{c_n\} \leq 0.25$ whenever $0 \leq c_n \leq 1$. The problem with weighting such distances, is that the weighting vector itself is given in $0 \leq \widetilde{y}_n \leq 1$. Thus, for the weighted distance $\{\widetilde{c}_n = \widetilde{y}_n c_n, n = 1..., N\}$, we necessarily verify the following:

$$0 \leq \widetilde{c}_n \leq c_n \leq 1 \Rightarrow 0 \leq \mathbb{E}\{\widetilde{c}_n\} \leq \mathbb{E}\{c_n\} \leq 1 \quad \text{for } n = 1, ..., N, \tag{A.22}$$

and then one could use (A.20) for showing that $\mathbb{V}\mathrm{ar}\{\widetilde{c}_n\} \leq \mathbb{E}\{\widetilde{c}_n\}(1 - \mathbb{E}\{\widetilde{c}_n\})$. The latter can be expressed as $\mathbb{V}\mathrm{ar}\{\widetilde{c}_n\} \leq \nu'(1-\nu')$ by letting $\mathbb{E}\{\widetilde{c}_n\} = \nu'$. Note that $\max \nu'(1-\nu') \leq 0.25$, as $\nu' \leq \nu$. Consequently, the variance of the weighted distance is confined in $0 \leq \mathbb{V}\mathrm{ar}\{\widetilde{c}_n\} \leq \max \nu'(1-\nu') \leq 0.25$. Thus, the vector of variances computed for the surrogate dataset $\{\Theta_0'(i) = \mathbb{V}\mathrm{ar}\{\widetilde{c}_n^{(s_i)}\}, i = 1, ..., I\}$ (where I is the number of surrogates),

will also be affected, and then we will have the following situation:

$$0 \leq \Theta_0'(i) \leq \max \mathbb{E}\{\widetilde{c}_n^{(s_i)}\}(1 - \mathbb{E}\{\widetilde{c}_n^{(s_i)}\}) \leq 0.25 \quad \text{for } i = 1, ..., I.$$

In practice, the estimated variance of the weighted distances will take values of the order of 10^{-5} (in the average). The latter can be seen in Fig. A.5, where we illustrated the estimates of $\mathbb{V}\text{ar}\{\widetilde{c}_n^{(s_i)}\}$ for the nonstationary signals illustrated in Fig. 2.2 with $T = 300$. Due to the very small range for which $\mathbb{V}\text{ar}\{\widetilde{c}_n\}$ and $\mathbb{V}\text{ar}\{\widetilde{c}_n^{(s_i)}\}$ are given, the round off errors and approximations in the estimation of the gamma parameters will be a critical problem, as well as the variance of the MLE estimators of θ and k (for which the latter needs to be computed numerically).

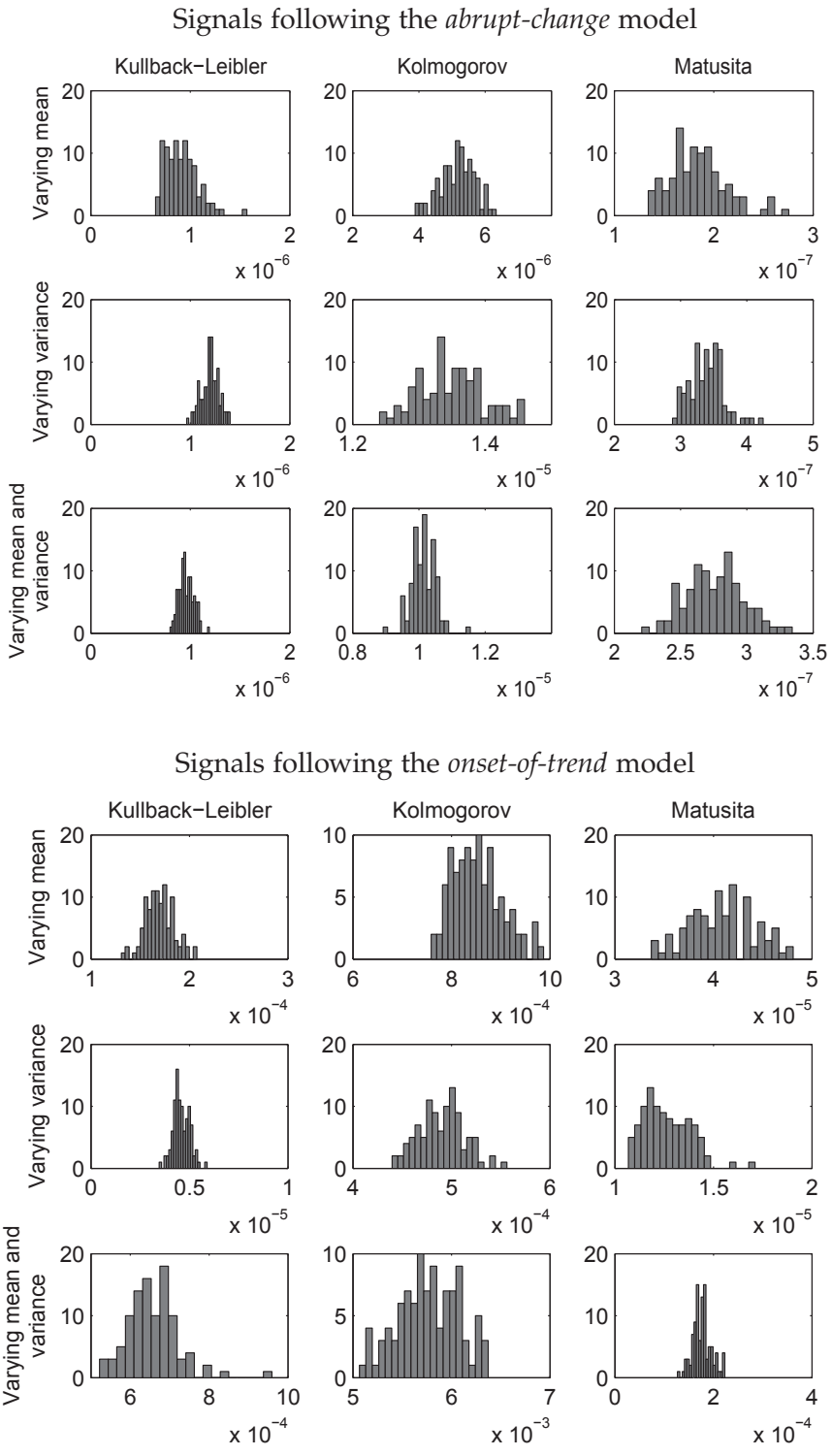

Figure A.5: Histograms of the values taken by the estimated variances forming the vector Θ'_0, for the probability-based distances and for the nonstationary signals shown in Fig. 2.2 with $T = 300$.

Bibliography

[1] G. B. Giannakis, *Cyclostationarity Signal Analysis, Digital Signal Processing Handbook*. CRC Press LLC, 1999.

[2] W. Martin and P. Flandrin, "Wigner-Ville spectral analysis of nonstationary processes," *IEEE Trans. on Acoustics, Speech and Signal Process.*, vol. 33, no. 6, pp. 1461–1470, 1985.

[3] D. G. Manolakis, V. K. Ingle, and S. M. Kogon, *Statistical and Adaptive Signal Processing: Spectral estimation, Signal Modeling, Adaptative Filtering and Array Processing*. London: Artech House, 2005, 796 p.

[4] P. Flandrin, *Time-Frequency/Time-Scale Anaylsis*, ser. Wavelet anlysis and its applications. San Diego, CA: Academic Press, 1999.

[5] D. Jarušková, "Some problems with application of change-point detection methods to environmental data," *Environmetrics*, vol. 8, no. 5, pp. 469–483, 1997.

[6] M. Basseville, "Distances measures for signal processing and pattern recognition," *Signal Process.*, vol. 18, no. 4, pp. 349–369, 1989.

[7] A. R. Rao, K. H. Hamed, and H.-L. Chen, *Nonstationarities in Hydrologic and Environmental Time Series (Water Science and Technology Library)*. Springer, Jul. 2003.

[8] R. M. Hirsch, "A perspective on nonstationarity and water management," *JAWRA Journal of the American Water Resources Association*, vol. 47, no. 3, pp. 436–446, 2011.

[9] P. C. D. Milly, J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, and R. J. Stouffer, "Stationarity Is Dead: Whither Water Management?" *Science*, vol. 319, no. 5863, pp. 573–574, Feb. 2008.

[10] J.-F. Quessy, A.-C. Favre, M. Saïd, and M. Champagne, "Statistical inference in Lombard's *smooth-change* model," *Environmetrics*, vol. 22, no. 7, pp. 882–893, 2011.

[11] S. H. Shih and C. P. Tsokos, "A new forecasting model for nonstationary environmental data," *Nonlinear Analysis, Theory, Methods and Applications*, vol. 71, no. 12, pp. e1209–e1214, 2009.

[12] H. Lins and T. Cohn, "Stationarity: Wanted dead or alive?" *JAWRA Journal of the American Water Resources Association*, vol. 47, no. 3, pp. 475–480, 2011.

[13] E. Carlstein, "Nonparametric Change-Point Estimation," *The Annals of Statistics*, vol. 16, no. 1, pp. 188–197, 1988.

[14] L. Dumbgen, "The asymptotic behavior of some nonparametric change-point estimators," *The Annals of Statistics*, vol. 19, no. 3, pp. 1471–1495, 1991.

[15] H. G. Müller, "Change-points in nonparametric regression analysis," *The Annals of Statistics*, vol. 20, no. 2, pp. 737–761, 1992.

[16] G. W. Cobb, "The problem of the nile: Conditional solution to a changepoint problem," *Biometrika*, vol. 65, no. 2, pp. 243–251, 1978.

[17] A. N. Pettitt, "A Non-Parametric Approach to the Change-Point Problem," *Applied Statistics*, vol. 28, no. 2, pp. 126–135, 1979.

[18] F. Lombard, "Rank test for changepoint problems," *Biometrika*, vol. 74, no. 3, pp. 615–624, 1987.

[19] E. Brodsky and B. Darkhovsky, *Nonparametric Methods in Change Point Problems*. Springer, Jan. 1993.

[20] S. Kay, "A new nonstationary detector," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1440–1451, 2008.

[21] H. Laurent and C. Doncarli, "Stationarity index for abrupt changes detection in the time-frequency plane," *Signal Processing Letters, IEEE*, vol. 5, no. 2, pp. 43–45, feb. 1998.

[22] S. Mallat, G. Papanicolaou, and Z. Zhang, "Adaptive covariance estimation of locally stationary processes," *Ann. Stat.*, vol. 24, no. 1, pp. 1–47, 1998.

[23] R. Andre-obrecht, "A new statistical approach for the automatic segmentation of continuous speech signals," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 36, no. 1, pp. 29–40, 1988.

[24] M. Davy and S. Godsill, "Detection of abrupt spectral changes using support vector machines an application to audio signal segmentation," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, vol. 2, 2002, pp. II–1313–II–1316.

[25] M. Basseville and I. Nikiforov, *Detection of Abrupt Changes - Theory and Application*. Englewood Clifs, NJ: Prentice-Hall, 1993.

[26] D. Kwiatkowski, P. C. B. Phillips, P. Schmidt, and Y. Shin, "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root," *J. Econom.*, vol. 54, pp. 159–178, 1992.

[27] B. Hobijn, P. H. Franses, and M. Ooms, "Generalizations of the kpss-test for stationarity," *Statistica Neerlandica*, vol. 58, no. 4, pp. 483–502, 2004.

[28] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, pp. 100–114, 1954.

[29] A. Sen and M. Srivastava, "On tests for detecting changes in mean," *Ann. Statist.*, vol. 3, pp. 98–108, 1975.

[30] T. Juhl, "A Nonparametric adjustment for test changing mean," *Economics Bulletin*, vol. 3, no. 34, pp. 1–11, 2004.

[31] N. Sugiura and R. T. Ogden, "Testing change-points with linear trend," *Communications in Statistics - Simulation and Computation*, vol. 23, no. 2, pp. 23–287, 1994.

[32] Y. Yang, Z. Peng, G. Meng, and W. Zhang", "Characterize highly oscillating frequency modulation using generalized warblet transform," *Mechanical Systems and Signal Processing*, vol. 26, pp. 128–140, 2012.

[33] Z. Leonowicz, *Parametric methods for time-frequency analysis of electric signals*. Faculty of Electrical Engineering, Wroclaw University of Technology, 2006.

[34] C. Hory, N. Martin, and A. Chehikian, "Spectrogram segmentation by means of statistical features for non-stationary signal interpretation," *Signal Processing, IEEE Transactions on*, vol. 50, no. 12, pp. 2915–2925, 2002.

[35] P. Djuric, S. Kay, and G. Boudreaux-Bartels, "Segmentation of nonstationary signals," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, vol. 5, 1992, pp. 161–164.

[36] S. Tahir, A. Shaameri, and S. Salleh, "Time-varying autoregressive modeling approach for speech segmentation," in *Signal Processing and its Applications, Sixth International, Symposium on. 2001*, vol. 2, 2001, pp. 715–718.

[37] F. Gianfelici, G. Biagetti, P. Crippa, and C. Turchetti, "Multicomponent AM-FM Representations: An Asymptotically Exact Approach," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 15, no. 3, pp. 823–837, 2007.

[38] A. Bracale, G. Carpinelli, K. Wozniak, T. Sikorski, and Z. Leonowicz, "Time-frequency analysis of non-stationary phenomena in electrical engineering," in *Modern Electric Power Systems. MEPS '06 Proc. Int. Symposium*, Wroclaw, Poland, Sep. 2006, pp. 373–378.

[39] P. Borgnat, P. Flandrin, P. Honeine, C. Richard, and J. Xiao, "Stationarization via surrogates," *J. Stat. Mech.: Th. and Exp.*, January 2009.

[40] J. Xiao and P. Flandring, "Multitaper time-frequency reassignment for nonstationary spectrum estimation and chirp enhancement," *Signal Processing, IEEE Transactions on*, vol. 55, no. 6, pp. 2851–2860, june 2007.

[41] P. Borgnat, P. Flandrin, P. Honeine, C. Richard, and J. Xiao, "Testing stationarity with surrogates: A time-frequency approach," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3459–3470, 2010.

[42] W. Martin and P. Flandrin, "Detection of changes of signal structure by using the Wigner-Ville spectrum," *Signal Processing*, vol. 8, no. 2, pp. 215–233, 1985.

[43] B. Music and D. Caya, "Evaluation of the hydrological cycle over the mississippi river basin as simulated by the canadian regional climate model (CRCM)," *J. Hydrometeor*, vol. 5, no. 8, pp. 968–988, 2007.

[44] J. Xiao, "Contributions to nonstationary spectrum estimation and stationarity tests in the time-frequency plane," Ph.D. dissertation, Ecole Normale Supérieure de Lyon, 2008.

[45] J. Reeves, J. Chen, X. L. Wang, R. Lund, and Q. Q. Lu, "A review and comparison of changepoint detection techniques for climate data," *Journal of Applied Meteorology and Climatology*, vol. 46, no. 6, pp. 900–915, 2007.

[46] B. Boashash, *Time Frequency Signal Analysis and Processing: A Comprehensive Reference*. Elsevier, 2003.

[47] A. Jung, G. Tauböck, and F. Hlawatsch, "Compressive nonstationary spectral estimation using parsimonious random sampling of the ambiguity function," in *Proc. IEEE Stat. Signal Process. Workshop (SSP)*, Cardiff, U.K., Sept. 2009, pp. 642–645.

[48] H. Hindberg, Y. Birkelund, T. A. Øigård, and A. Hanssen, "Kernel-based estimators for the Kirkwood-Rihaczek time-frequency spectrum," in *Proc. European Signal Processing Conference (EUSIPCO)*, Florence, Italy, Sep. 2006.

[49] M. Bayram and R. Baraniuk, *Nonlinear and Nonstationary Signal Processing*. Cambridge Univ. Press, 2001, ch. Multiple Window Time-Varying Spectrum Estimation, pp. 292–316.

[50] A. Jung, G. Taubock, and F. Hlawatsch, "Compressive spectral estimation for nonstationary random processes," *Information Theory, IEEE Transactions on*, vol. 59, no. 5, pp. 3117–3138, 2013.

[51] M. Alam, M. Rahman, N. Parvin, and M. Sobhan, "Time-frequency representation of a signal through non-stationary multipath fading channel," in *Informatics, Electronics Vision (ICIEV), 2012 International Conference on*, 2012, pp. 1130–1135.

[52] Z. Leonowicz, T. Lobos, and P. Schegner, "Modern spectral analysis of nonstationary signals in electrical power systems," in *Proc. Power System Computation Conference (PSCC)*, Sevilla, Spain, Jun. 2002, pp. 115–119.

[53] Z. Leonowicz and T. Lobos, "Time-frequency analysis of non-stationary three phase signals," in *15th Triennial World Congress of the International Federation of Automatic Control*, no. 302, Barcelona, Spain, Jul. 2002.

[54] R. Iqbal, T. Abhayapala, J. Ahmed, and T. Lamahewa, "Wigner-ville distribution of a type of non-stationary mobile rayleigh fading channels," in *Multitopic Conference, 2009. INMIC 2009. IEEE 13th International*, 2009, pp. 1–6.

[55] T. Schreiber and A. Schmitz, "Surrogate time series," *Physica D*, vol. 142, no. 3-4, pp. 346–382, 2000.

[56] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, "Testing for nonlinearity in time series: the method of surrogate data," *Physica D*, vol. 58, no. 1–4, pp. 77–94, 1992.

[57] G. Matz and F. Hlawatsch, "Wigner distributions (nearly) everywhere: time-frequency analysis of signals, systems, random processes, signal spaces and frames," *Signal Processing*, vol. 83, no. 7, pp. 1355–1378, 2003.

[58] J. Xiao and P. Flandrin, "Multitaper time-frequency reassignment for nonstationary spectrum estimation and chirp enhancement," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2851–2860, 2007.

[59] F. Liese and I. Vajda, "On divergences and informations in statistics and information theory," *Information Theory, IEEE Transactions on*, vol. 52, no. 10, pp. 4394–4412, 2006.

[60] M. Basseville, "Divergence measures for statistical data processing," Research Report PI-1961, Nov. 2010.

[61] R. L. Allen and D. W. Mills, *Signal analysis: time, frequency, scale, and structure*. Piscataway: IEEE Press, 2004.

[62] H. Ling and K. Okada, "Diffusion distance for histogram comparison," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1. Washington, DC, USA: IEEE Computer Society, June 2006, pp. 246–253.

[63] B. Efron, "Bootstrap methods: Another look at the jackknife," *The Annals of Statistics*, no. 7, pp. 1–26, 1979.

[64] ——, "Computers and the theory of statistics: Thinking the unthinkable," *SIAM Review*, no. 4, pp. 460–480, 1979.

[65] A. M. Zoubir and B. Boashash, "The bootstrap and its application in signal processing," *IEEE Signal Processing Magazine*, vol. 15, pp. 56–76, 1998.

[66] A. Moghtaderi, P. Flandrin, and P. Borgnat, "Trend filtering via empirical mode decompositions," *Computational Statistics and Data Analysis*, vol. 58, pp. 114–126, feb 2013.

[67] R. Vautard and M. Ghil, "Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series," *Physica D: Nonlinear Phenomena*, vol. 35, no. 3, pp. 395 – 424, 1989.

[68] T. Alexandrov, S. Bianconcini, E. B. Dagum, P. Maass, and T. S. McElroy, "A review of some modern approaches to the problem of trend extraction," *Econometric Reviews*, vol. 31, no. 6, pp. 593–624, 2012.

[69] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 454, pp. 903–995.

[70] P. Flandrin, G. Rilling, and P. Goncalves, "Empirical mode decomposition as a filter bank," *Signal Processing Letters, IEEE*, vol. 11, no. 2, pp. 112–114, feb. 2004.

[71] G. Rilling, P. Flandrin, and P. Gonçalvès, "On empirical mode decomposition and its algorithms," in *Proceedings of the 6th IEEE/EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP '03), Grado, Italy*, 2003.

[72] P. Guhathakurta and M. Rajeevan, "Trends in the rainfall pattern over india," *Int. J. Climatol.*, vol. 28, pp. 1453–1469, 2008.

[73] A. C. Davison and D. V. Hinkley, *Bootstrap Methods and their Application (Cambridge Series in Statistical and Probabilistic Mathematics)*, 1st ed. Cambridge University Press, 1997.

[74] S. N. Lahiri, "Theoretical comparisons of block bootstrap methods," *Annals of Statistics*, vol. 27, pp. 386–404, 1999.

[75] P. Hall, J. L. Horowitz, and B.-Y. Jing, "On blocking rules for the bootstrap with dependent data," *Biometrika*, vol. 82, no. 3, pp. 561–574, 1995.

[76] M. Sherman, F. M. Speed, and F. M. Speed, "Analysis of tidal data via the blockwise bootstrap," *Journal of Applied Statistics*, vol. 25, no. 3, pp. 333–340, 1998.

[77] S. Coles, *An Introduction to Statistical Modeling of Extreme Values*. Springer, 2001.

[78] S. Kotz, *Extreme Value Distributions: Theory and Applications*. World Scientific Publishing Company, 2001.

[79] A. Zempléni, "Inference for generalized extreme value distributions," *Journal of Applied Statistical Science*, vol. 4, no. 3, pp. 107–122, 1996.

[80] M. A. Stephens, "Edf statistics for goodness of fit and some comparisons," *Journal of the American Statistical Association*, vol. 69, no. 347, pp. 730–737, 1974.

[81] A. Zempléni, "Goodness-of-fit test in extreme value applications," München, Discussion paper 383, 2004.

[82] P. A. Jensen and J. F. Bard, *Operations Research Models and Methods*. Wiley, 2002.

[83] H. Aksoy, "Use of gamma distribution in hydrological analysis," *Turkish J. Eng. Env. Sci.*, vol. 24, no. 6, pp. 419–428, 2000.

[84] T. S. Rao, "The fitting of nonstationary time-series models with time-dependent parameters," *J. Royal Statist. Soc.*, vol. 32, no. 2, pp. 312–322, 1970.

[85] D. de Souza, J. Chanussot, A. Favre, and P. Borgnat, "A modified time-frequency method for testing wide-sense stationarity," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, march 2012, pp. 3409–3412.

[86] J.-F. Quessy, M. Saïd, and A.-C. Favre, "Multivariate Kendall's tau for change-point detection in copulas," *Canadian Journal of Statistics*, pp. 1–18, 2012.

[87] T. Tokunaga *et al.*, "Detecting precursory events in time series data by an extension of singular spectrum transformation," in *Proceedings of the 10th WSEAS international conference on Applied computer science*, ser. ACS'10, 2010, pp. 366–374.

[88] Y. Mohammad and T. Nishida, "Robust singular spectrum transform," in *Next-Generation Applied Intelligence*, ser. Lecture Notes in Computer Science, B.-C. Chien, T.-P. Hong, S.-M. Chen, and M. Ali, Eds. Springer Berlin Heidelberg, 2009, vol. 5579, pp. 123–132.

[89] E. Gombay, "Change detection in autoregressive time series," *J. Multivar. Anal.*, vol. 99, no. 4, pp. 451–464, 2008.

[90] S. Kadambe and G. F. Boudreaux-Bartels, "Application of the wavelet transform for pitch detection of speech signals," *Information Theory, IEEE Transactions on*, vol. 38, no. 2, pp. 917–924, mar 1992.

[91] T. Idé and K. Inoue, "Knowledge discovery from heterogeneous dynamic systems using change-point correlations," in *Proceedings of the 2005 SIAM International Data Mining Conference*, 2005.

[92] N. Itoh and N. Marwan, "An extended singular spectrum transformation (SST) for the investigation of kenyan precipitation data," *Nonlinear Processes in Geophysics*, vol. 20, no. 4, pp. 467–481, 2013.

[93] Y. Mohammad and T. Nishida, "On comparing SSA-based change point discovery algorithms," in *System Integration (SII), 2011 IEEE/SICE International Symposium on*, dec. 2011, pp. 938 –945.

[94] V. Moskvina and A. Zhigljavsky, "An algorithm based on singular spectrum analysis for change-point detection," *Communications in Statistics - Simulation and Computation*, vol. 32, no. 2, pp. 319–352, 2003.

[95] E. Gombay and L. Horváth, "Asymptotic distributions of maximum likelihood tests for change in the mean," *Biometrika*, vol. 77, no. 2, pp. 411–414, 1990.

[96] J.-Y. Koo, "Spline estimation of discontinuous regression functions," *Journal of Computational and Graphical Statistics*, vol. 6, no. 3, pp. 266–284, 1997.

[97] Z. Guan, "A semiparametric changepoint model," *Biometrika*, vol. 91, no. 4, pp. 849–862, 2004.

[98] H. Côté *et al.*, "Description of the canadian regional climate model," *Water, Air, and Soil Pollution*, no. 1-2, pp. 477–482, 1995.

[99] N. Nakicenovic *et al.*, *Special Report on Emissions Scenarios: A Special Report of Working Group III of the Intergovernmental Panel on Climate Change*, 1st ed. Cambridge University Press, Jul. 2000.

[100] "CRCM 4.2.3 monthly data (aet run) by Ouranos Climate Simulation Team," http://www.cccma.ec.gc.ca/data/crcm423/crcm423_aet_sresa2.shtml.

[101] "CRCM 4.2.3 monthly data (aev run) by Ouranos Climate Simulation Team," http://www.cccma.ec.gc.ca/data/crcm423/crcm423_aev_sresa2.shtml.

[102] J. F. Scinocca, N. A. McFarlane, M. Lazare, J. Li, and D. Plummer, "Technical note: The cccma third generation agcm and its extension into the middle atmosphere," *Atmospheric Chemistry and Physics*, vol. 8, no. 23, pp. 7055–7074, 2008.

[103] Z. Wu and N. E. Huang, "A study of the characteristics of white noise using the empirical mode decomposition method," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 460, no. 2046, pp. 1597–1611, 2004.