



**HAL**  
open science

## Dynamiques épidémiques, risques et copules

Mohamad Ghassani

► **To cite this version:**

Mohamad Ghassani. Dynamiques épidémiques, risques et copules. Médecine humaine et pathologie. Université de Grenoble, 2012. Français. NNT : 2012GRENS027 . tel-00947345

**HAL Id: tel-00947345**

**<https://theses.hal.science/tel-00947345>**

Submitted on 15 Feb 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## THÈSE

Pour obtenir le grade de

### DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **MBS - Modèles, méthodes et algorithmes en biologie, santé et environnement**

Arrêté ministériel : 7 août 2006

Présentée par

**Mohamad GHASSANI**

Thèse dirigée par **Mr. Jacques Demongeot et Mr. Mustapha Rachdi**

préparée au sein du **Laboratoire AGIM (AGe-Imagerie-Modélisation)**  
dans l'**Ecole Doctorale Ingénierie pour la Santé, la Cognition et l'Environnement**

# Dynamiques épidémiques, risques et copules

Thèse soutenue publiquement le **30 Novembre 2012**,  
devant le jury composé de :

**Mr. Bedreddine Ainseba**

Professeur - Université Segalen - Bordeaux 2, Président

**Mr. Ali Gannoun**

Professeur - université Montpellier 2, Rapporteur

**Mr. Philippe Vieu**

Professeur - Université Paul Sabatier - Toulouse 3, Rapporteur

**Mr. Philippe Sabatier**

Professeur - Université Joseph Fourier - Grenoble 1, Examineur

**Mr. Pierre Magal**

Professeur - Université Segalen - Bordeaux 2, Examineur

**Mr. Alain Duhamel**

Professeur - Université Lille 2, Examineur

**Mr. Jacques Demongeot**

Professeur - Université Joseph Fourier - Grenoble 1, Directeur de thèse

**Mr. Mustapha Rachdi**

Professeur - Université Pierre-Mendès-France - Grenoble 2, Directeur de thèse





A ma mère  
A mon frère  
A mes soeurs



# Remerciements

En préambule de ce mémoire, je souhaite adresser mes remerciements les plus sincères aux personnes qui m'ont apporté leur aide et qui ont contribué à l'élaboration de ce mémoire ainsi qu'à la réussite de ces magnifiques trois années de doctorat.

Ce travail n'aurait pu être mené sans la forte implication d'un grand nombre de personnes que je voudrais remercier ici, même si la liste est trop longue pour être exhaustive.

Je veux tout d'abord remercier mes directeurs de thèses, Mr. Jacques Demongeot et Mr. Mustapha Rachdi, pour toute la confiance et le soutien offerts. Je leur suis profondément reconnaissant pour le temps conséquent qu'ils m'ont accordé, leurs conseils et encouragements ainsi que leurs qualités sur le plan scientifique, pédagogique et humain. J'ai énormément appris à leurs côtés et j'ai pris un grand plaisir à travailler avec eux.

Deuxièmement, je remercie les rapporteurs de cette thèse, Mr. Ali Gannoun et Mr. Philippe Vieu pour l'intérêt qu'ils ont porté à mon travail. Je souhaite ensuite remercier Mr. Philippe Sabatier, Mr. Bedreddine Ainseba, Mr. Pierre Magal et Mr. Alain Duhamel d'avoir accepté de participer au jury de cette thèse.

Je désire remercier en outre toute les personnes qui font partie du laboratoire AGIM et à tous mes collègues de travail avec qui j'ai partagé le bureau et de longues et enrichissantes conversations et aussi à Mme Caroline Messina-Dos-Santos Secrétaire du Laboratoire AGIM pour sa sympathie et son efficacité.

Un grand merci à mes amis, en particulier Hasnaa avec qui j'ai partagé les moments difficiles et les moments heureux, et surtout son grand soutien sa patience et sa confiance en moi tout au long de ma thèse.

Je voudrais remercier ma famille, surtout ma mère, pour son encouragement et son soutien, sans elle je ne serai jamais là où je suis. Je tiens à remercier mon frère Abbass et mes soeurs Zeinab, Fatima et Zahraa qui m'ont persuadé et encouragé à continuer mes études en France. Je remercie aussi mes oncles Hassan et Mihsen qui ont cru en moi et qui sont restés près de moi tout au long de cette thèse.



# Table des matières

<b>Remerciements</b>	<b>5</b>
<b>Résumé</b>	<b>13</b>
<b>Summary</b>	<b>14</b>
<b>Introduction générale</b>	<b>15</b>
<b>I Etat de l'art</b>	<b>21</b>
I.1 Introduction . . . . .	21
I.2 Les fonctions copules . . . . .	22
I.2.1 Définitions . . . . .	22
I.2.2 Théorème de Sklar . . . . .	24
I.2.3 Densité de la fonction copule . . . . .	26
I.2.4 Propriétés des fonctions copules . . . . .	27
I.2.5 Les classes de Fréchet . . . . .	28
I.2.6 Mesure de dépendance . . . . .	30
I.2.7 Les fonctions copules paramétriques . . . . .	33
I.3 Modèle de Cox . . . . .	38
I.3.1 Notion de risque instantané . . . . .	38
I.3.2 Principes généraux et définitions . . . . .	38
I.3.3 Interprétation des coefficients . . . . .	39
I.3.4 Modèle de fragilité . . . . .	40
I.4 Modèles épidémiologiques . . . . .	41
I.4.1 Modèle de Ross . . . . .	42
I.4.2 Modèle de Macdonald . . . . .	47
I.4.3 Modèle de Ross-McKendrick . . . . .	49
I.4.4 Stabilité des systèmes dynamiques . . . . .	50
I.5 Conclusion . . . . .	52
<b>II Demography in epidemics modelling: the copula approach</b>	<b>57</b>
II.1 Introduction . . . . .	57
II.2 Classical Epidemiology: The Ross-McKendrick Model . . . . .	57
II.3 Leslie model . . . . .	59



II.4	Usher model . . . . .	59
II.5	Cox model . . . . .	60
II.6	Our models . . . . .	60
II.6.1	Model 1: Model with three age classes . . . . .	61
II.6.2	Model 2: Model with two age classes . . . . .	64
II.6.3	Model 3: three age classes without mosquitoes . . . . .	68
II.7	Simulation studies . . . . .	68
II.7.1	The simulation of the model 2 . . . . .	68
II.7.2	The simulation of the model 3 . . . . .	70
II.8	Conclusion . . . . .	72
 <b>III Demography and Diffusion in epidemics: Malaria and Black Death spread</b>		<b>77</b>
III.1	Introduction . . . . .	77
III.2	Introduction of demographic dynamics . . . . .	77
III.2.1	Introduction of age classes . . . . .	77
III.2.2	Mathematical properties . . . . .	78
III.3	Introduction of a spatial dynamics . . . . .	81
III.4	Introduction of saturation kinetics and genetic drifts . . . . .	82
III.5	Black Death . . . . .	83
III.5.1	Introduction . . . . .	83
III.5.2	The raw data . . . . .	83
III.5.3	The Model . . . . .	85
III.6	The Malaria in Mali . . . . .	86
III.6.1	Introduction . . . . .	86
III.6.2	The Model . . . . .	86
III.7	Perspectives . . . . .	92
III.8	Conclusion . . . . .	94
 <b>IV Archimedean copula and contagion modelling in epidemiology</b>		<b>100</b>
IV.1	Introduction . . . . .	100
IV.2	Preliminaries . . . . .	102
IV.2.1	Archimedean copula . . . . .	102
IV.2.2	Concordance, Kendall's tau, Spearman's rho, Copula . . . . .	104
IV.2.3	Multidimensional conditional distribution function using copulas . . . . .	105
IV.3	Quantiles regression based on the system of transmission of Malaria . . . . .	106
IV.3.1	Interaction between compartments of hosts . . . . .	106
IV.3.2	Quantile regression using the bivariate Gumbel copula . . . . .	108
IV.4	Quantiles from the distribution functions . . . . .	110
IV.4.1	A case where the population sizes $S_1$ , $I_1$ , $S_2$ , $E_2$ and $I_2$ are Poissonian random variables . . . . .	112
IV.4.2	Quantile regression with Poissonian distributions . . . . .	114
IV.5	Conclusion . . . . .	114
IV.6	Appendix . . . . .	114
IV.6.1	Appendix A . . . . .	114

IV.6.2 Appendix B . . . . .	115
IV.6.3 Appendix C . . . . .	115
<b>Conclusions et Perspectives</b>	<b>120</b>
Conclusions . . . . .	120
Perspectives . . . . .	121
<b>Bibliographie générale</b>	<b>122</b>



# Table des figures

I.1	La fonction copule bivariée $C^\perp$ . . . . .	24
I.2	La fonction copule bivariée $C^-$ . . . . .	30
I.3	La fonction copule bivariée $C^+$ . . . . .	30
I.4	La fonction distribution de la fonction copule bivariée de Gumbel (resp. pour $\alpha = 2$ , $\alpha = 5$ et $\alpha = 10$ ) . . . . .	37
I.5	La fonction distribution de la fonction copule bivariée de Clayton (resp. pour $\alpha = 2$ , $\alpha = 5$ et $\alpha = 10$ ) . . . . .	37
I.6	Un anophèle femelle (cf. Sallet, 2010) . . . . .	41
I.7	Le modèle compartimental de Ross . . . . .	42
I.8	Prévalence en fonction de la capacité vectorielle. . . . .	46
I.9	Modèle de Ross-Macdonald . . . . .	48
II.1	Picture of D. Bernoulli (top left); interaction graphs of the Ross-McKendrick model having one (top right) and three (bottom) age classes, with identical $\beta_i$ 's and $\gamma_i$ 's and no fecundity in elderly classes $S_3$ and $I_3$ . . . . .	58
II.2	interaction graph of the model (7) with three age classes and two compartments of mosquitoes . . . . .	62
II.3	interaction graph of the model (12) with two age classes and two compartments of mosquitoes . . . . .	65
II.4	Interaction graph for the model (14) with three age classes . . . . .	69
II.5	Graph of the equilibrium line . . . . .	70
II.6	Graph of the functions $P_{B_1}(\lambda)$ (in red color), $P'_{B_1}(\lambda)$ (in blue color) and $P''_{B_1}(\lambda)$ (in green color) . . . . .	72
III.1	Graphs of the functions $P_{B_1}(\lambda)$ , $P'_{B_1}(\lambda)$ and $P''_{B_1}(\lambda)$ . . . . .	79
III.2	Graphs of $P_{B_2}(\lambda)$ and $P'_{B_2}(\lambda)$ . . . . .	80
III.3	Representation of the co-evolution of the zero-diffusion domains for interacting species $S$ (in blue color) and $A_i$ (in red color) in the case of isotropic diffusion (on the left). Asymptotic co-existence of $S$ and $A_i$ on their common least diffusion domain (on the right) . . . . .	81
III.4	The triple win game in which hosts and vectors use the viral genome for making evolve their own genomes, and the virus survives thanks to these later, which code for its proteins . . . . .	82

III.5	<i>Top left:</i> The spread of bubonic plague following sea and overland routes (cf. <a href="http://www.cosmovisions.com/ChronoPestesMA02.htm">http : //www.cosmovisions.com/ChronoPestesMA02.htm</a> ). <i>Top right:</i> Observed wave fronts after 1 (in red color), 2 (in blue color) and 3 (in green color) years of spread from the 2 initial entry ports Marseilles and Genoa (cf. Mocellin-Spicuzza, 2002); the black grid corresponds to the collected altitudes. <i>Bottom:</i> Simulation of the wave front after 3 (on the left) and 6 (on the right) months from the 2 initial entry ports Marseilles and Genoa . . . . .	83
III.6	World distribution of plague in 1998 (cf. WHO, 1999) . . . . .	85
III.7	Interaction graph ( <i>in top</i> ), and SIGR Diffusion (SIGRD) Equation system (9) both for hosts and for vectors ( <i>in bottom</i> ) for the Bankoumana model . . . . .	87
III.8	Results of the SGIRD simulated ( <i>on the top-right</i> ) and real data ( <i>on the middle-right</i> ) showing a good fit along a gradient parallel to Niger river from the southwest backwater zone ( <i>on bottom</i> ) to the village of Bankoumana (Mali) for the Bankoumana model . . . . .	88
III.9	Interaction graph of the Ross-Macdonald model . . . . .	93
IV.1	Interaction graph of the Ross-Macdonald model . . . . .	101
IV.2	Distribution functions of Clayton bivariate copula (resp. for $\alpha = 2$ , $\alpha = 5$ and $\alpha = 10$ ) . . . . .	103
IV.3	Distribution functions of Gumbel bivariate copula (resp. for $\alpha = 2$ , $\alpha = 5$ and $\alpha = 10$ ) . . . . .	103
IV.4	Simulation of social graphs representing a contagion network, with initial conditions (a) and asymptotic state in case of an homophilic graph (b), random graph (c), scale free graph (d) and small world graph (e). . . . .	107
IV.5	Interaction between the distribution functions of $S_2$ and $E_2$ using the Gumbel copula with the parameter of regression equal to 1, 3 and 5 respectively . . .	108
IV.6	Interaction between the distribution functions of $E_2$ and $I_2$ using the Gumbel copula with the parameter of regression equal to 1, 3 and 5 respectively . . .	108
IV.7	Interaction between the distribution functions of $S_2$ and $E_2$ using the Clayton copula with the parameter of regression equal to 1, 3 and 5 respectively . . .	109
IV.8	Interaction between the distribution functions of $E_2$ and $I_2$ using the Clayton copula with the parameter of regression equal to 1, 3 and 5 respectively . . .	109
IV.9	Interaction between the distribution functions of $S_2$ and $E_2$ using the Gumbel copula with $\alpha = 3$ , using the quantile regression (of $E_2$ on $S_2$ ) curves. . . . .	110

# Résumé

Les modèles stochastiques classiques comportent des copules d'interactions linéaires, exprimant en général des interactions de paire. Il sera envisagé d'étendre ces modèles à des interactions non linéaires de type saturation ou de type triplet, en vue de traiter des applications réalistes, comme les diffusions épidémiques.

Le but de cette thèse est d'introduire les fonctions copules en épidémiologie, et surtout d'appliquer ces fonctions sur le système de transmission de la Malaria afin de constater la dépendance entre les différents compartiments du système. Nous étudierons quelques modèles compartimentaux, qui sont une généralisation du modèle de Ross-Macdonald, en supposant que la population n'est pas constante et en prenant en compte des paramètres de transmission comme la fécondité, la mortalité et autres. Aussi, nous introduirons les classes d'âges dans certains de ces modèles compartimentaux, afin de trouver une relation entre les individus de ces classes d'âges à l'aide du modèle de Cox et des fonctions copules. Nous donnerons ensuite, deux exemples sur ces modèles : la Malaria au Mali et la peste en Europe au moyen-âge. Nous introduirons aussi les quantiles conditionnels et les fonctions copules archimédiennes, ce qui nous mènera à trouver une dépendance entre les différents compartiments des hôtes et des vecteurs.

# Summary

The stochastic classical models include linear interactions copulas, expressing in general pair interactions. It is planned to extend these models to nonlinear interactions of saturation type or triplet type, to treat realistic applications, as the epidemics diffusions.

The aim of this thesis is to apply the copulas functions in epidemiology, and especially to apply these functions in the transmission system of malaria to detect the dependence existing between compartments of the epidemic system. We will study some compartmental models, which are a generalization of the Ross-Macdonald model, assuming that the population is not constant and taking into account the transmission parameters such as fertility, mortality, etc. Also, we will introduce the age classes in some of these compartmental models, and study the relationships between individuals of these age classes, using the Cox model and the copulas functions. Then, we will give two examples of these models : the Malaria in Mali and the plague in Europe during the Middle Ages. We will introduce also the conditional quantiles and the Archimedean copulas functions, that will lead us to find dependencies between the different compartments of hosts and vectors.

# Introduction générale

Les fonctions copules sont devenues en quelques années un outil important, pour modéliser les risques multivariés (entre autres). Les fonctions copules permettent de “coupler” les lois marginales afin d’obtenir une loi multivariée, d’où le nom latin “copula” choisi par Abe Sklar en 1959. Ces fonctions sont simplement des fonctions de répartition en dimension quelconque dont les lois marginales sont uniformes sur  $[0, 1]$ . En effet, les copules, largement utilisées en statistique financière, fournissent des modèles de lois multivariées qui permettent de représenter les relations de dépendance entre les diverses variables utilisées.

Les mesures statistiques classiques de dépendance telles que le coefficient de Hoeffding, présentent une faiblesse, en ne prenant en compte que les distributions marginales. Des mesures de concordance, comme le tau de Kendall et le rho de Spearman, généralisent la corrélation linéaire et s’obtiennent en fonction des fonctions copules, prenant en compte les dépendances des distributions jointes.

Maintenant, les fonctions copules sont devenues un outil standard et souvent utilisées pour étudier la dépendance, les modèles de survie, etc.

Les modèles de survie, ont été développés en vue d’applications en biologie, en épidémiologie, etc. Parmi les différents modèles de durée de vie, figurent les modèles à risques proportionnels, en particulier le modèle de Cox. Il s’agit de modèles semi-paramétriques dans lesquels on modélise la fonction de risque, en considérant que les covariables agissent par effet multiplicatif via une fonction monotone à valeurs positives, sur une fonction de risque de base.

En épidémiologie, le premier travail de modélisation a été réalisé, à la fin du dix-huitième siècle, pour l’évaluation de l’efficacité de la variolisation, en termes de nombre de morts évités et de gain d’espérance de vie (cf. Bernoulli, 1760). Puis, au dix-neuvième siècle, dans le domaine de l’épidémiologie des maladies transmissibles, Ross fut le premier à mettre en place une analyse sur la planification de la lutte contre le paludisme et a développé le premier modèle mathématique sur celle-ci. Les fondements de l’épidémiologie basés sur les modèles compartimentaux ont été établis par des médecins de santé publique, comme W. Farr, Sir Ronald Ross, W.H. Hamer, A.G. McKendrick et W.O. Kermack entre les années 1865 et 1935.

La modélisation de la transmission du paludisme concerne les variations de la fréquence des infections chez les hôtes et les vecteurs. Elle est la première à avoir été développée depuis maintenant plus de 90 ans, et plus particulièrement à l’époque du programme mondial d’éradication. L’histoire de l’approche mathématique du paludisme est presque aussi vieille que la découverte de sa transmission. Si Ross a découvert la transmission par les anophèles en 1898, il a aussi publié le premier modèle compartimental, en 1911, de la transmission du paludisme. On peut dire que l’épidémiologie mathématique est née avec Ross.



L'objectif de cette thèse est d'étudier les modèles compartimentaux, précisément les modèles de transmission de la Malaria, en supposant que la population n'est pas constante, et en introduisant des paramètres démographiques comme la fécondité, la mortalité, etc. Ensuite, nous procéderons une généralisation de ces modèles avec les classes d'âges. Le but de cette thèse est aussi de trouver des interactions entre les compartiments des hôtes et des vecteurs, en utilisant les fonctions copules et en passant par le modèle de Cox.

Par la suite, nous présenterons un résumé de chaque chapitre :

Le chapitre 1, qui est un chapitre introductif, est divisé en trois parties : dans la première partie, nous débiterons par des notions préliminaires sur les fonctions copules et quelques résultats fondamentaux les concernant. Puis, nous expliquerons la notion de la mesure de dépendance pour parler du tau de Kendall et du rho de Spearman. Mais encore, nous définirons les fonctions copules paramétriques qui se composent de trois différentes copules : la fonction copule normale, la fonction copule de Student et les fonctions copules Archimédiennes. Tandis que dans la seconde partie, nous nous focaliserons sur le modèle de Cox, en donnant la notion de risque instantané et quelques principes généraux, la dernière section sera consacrée au modèle de fragilité lié à la fonction de survie.

Dans la troisième et dernière partie de ce chapitre, nous entamerons l'analyse des différents systèmes épidémiologiques en commençant par le modèle de Ross et sa modélisation. Puis, nous introduirons le modèle de Macdonald qui est une généralisation du modèle de Ross. Ensuite, nous exposerons le modèle de Ross-Mckendrick. Enfin, nous discuterons la stabilité de ces systèmes dynamiques.

Dans le deuxième chapitre, nous étudierons les modèles épidémiques classiques donnés par Ross et McKendrick, en considérant des facteurs démographiques (la fécondité, la mortalité et la migration) pour les compartiments des hôtes et des vecteurs de la population étudiée, et en prenant aussi en compte la diffusion et la mutation des agents infectieux. Ensuite, nous présenterons trois modèles épidémiques avec les classes d'âges en faisant une approche à l'aide des fonctions copules. A la fin de ce chapitre, nous ferons une simulation de certains modèles étudiés.

L'approche classique des modèles épidémiques est, en effet, faite en supposant que la population étudiée est constante pendant la vague épidémique. Pour cela, dans le troisième chapitre, nous exposerons deux exemples des modèles compartimentaux en prenant en compte le fait que la population n'est pas constante, c'est-à-dire en supposant qu'il y a des migrations dans les deux compartiments des hôtes et des vecteurs. Ces deux exemples concernent la Malaria au Mali et la peste en Europe au moyen-âge.

Le but du quatrième chapitre est, premièrement, de trouver des interactions entre les compartiments des hôtes dans le modèle de transmission de la Malaria de Ross-Macdonald. Pour rendre plus claire ces interactions, nous introduirons une mesure de dépendance, précisément le tau de Kendall et le rho de Spearman. En outre, puisque les compartiments de la population sont dépendants, nous calculerons leurs fonctions de distributions conditionnelles à l'aide des fonctions copules archimédiennes. Deuxièmement, nous diviserons la population des vecteurs en plusieurs parties dépendantes conditionnellement à la fécondité et aux paramètres de transmission. Nous montrerons aussi que l'on peut diviser cette même population en utilisant le  $p$ -ème quantile. Troisièmement, nous calculerons le  $p$ -ème quantile à partir de

la distribution de Poisson. Et enfin, nous introduirons le modèle à risques proportionnels de Cox dans le modèle de Ross-Macdonald, avec l'approche des fonctions copules, afin de trouver une relation entre les fonctions de survie des vecteurs.

Notons que les fonctions copules sont largement utilisées en finance, mais que c'est la première fois qu'elles seront utilisées en épidémiologie.



# Liste de Publications et communications aux congrès

1. Demography and Diffusion in epidemics : Malaria and Black Death spread,  
Avec : M. Rachdi, J. Demongeot, J. Mintsa, J. Waku,  
*Acta Biotheoretica, Volume 58, Issue 2-3, pages 277-305.*
2. Demographic and spatial factors as causes of an epidemic spread, the copula approach :  
Application to the retro-prediction of the Black Death epidemic of 1346,  
Avec : J. Demongeot, M. Rachdi, J. Gaudart, J. Mintsa, J. Waku, O.K. Doumbo,  
*IEEE Advanced Information Networking and Application, DOI 10.1109 /WAINA. 2010. 79 , pages 751-758.*
3. Archimedean copula and contagion modeling in epidemiology,  
Avec : M. Rachdi, J. Demongeot, I. Ouassou, C. Taramasco,  
*Networks and Heterogeneous Media, AIMS Journals, Soumis.*
4. Statistiques spatiales et copules d'interactions non-linéaires,  
Avec : J. Demongeot, M. Rachdi,  
*Journées Internationales Analyse statistique : Théorie et Applications, 04-06 JUIN 2012, OUJDA, MAROC.*



# Chapitre I

## Etat de l'art

### I.1 Introduction

L'étude des fonctions copules et leurs applications en statistiques est un phénomène qui est plutôt moderne, et encore plus en biologie. En effet, les fonctions copules, largement utilisées en statistique financière, sont des fonctions qui relient les fonctions de distributions multivariées à celles des distributions univariées. En fait, les fonctions copules sont principalement utiles pour la modélisation des dépendances entre les variables aléatoires.

Le concept de “copule” a été introduit par Sklar en 1959 [1], afin de résoudre un problème de probabilité qui a été énoncé par Fréchet [52, 53, 54]. A l'époque, Sklar et Schweizer travaillent sur les espaces de Menger, c'est-à-dire sur les espaces métriques aléatoires (Probabilistic Metric Space ou PMS), qui sont une généralisation de l'espace métrique usuel introduit par Fréchet en 1906. Même si les fonctions copules occupent une place importante dans l'oeuvre de Sklar et Schweizer, celles-ci ne constituent pas l'objet central de leurs recherches. L'utilisation des fonctions copules par Sklar et Schweizer est, en fait, assez originale : elles interviennent pour résoudre certains problèmes mais ne font pas, véritablement, l'objet d'études appropriées.

Pendant de nombreuses années, les fonctions copules ont été peu (ou pas) utilisées en statistiques. Il y a les travaux sur la dépendance dus à Kimeldorf et Sampson dans les années 1975 ou encore les recherches de Deheuvels à la fin des années 1970 [4, 5, 6]. Mais il faut attendre le milieu des années 1980 pour que celles-ci fassent l'objet d'une étude systématique par quelques statisticiens. Le point de départ est bien sûr l'article de Genest et MacKey [12], et ensuite suivront de nombreux travaux de Genest avec différents co-auteurs (MacKey, Louis-Paul Rivest,...). Maintenant, les fonctions copules sont devenues un outil standard et sont largement utilisées pour étudier la dépendance, les modèles de survie, etc.

D'un autre côté, dans le cadre de l'analyse des données de survie, le modèle le plus utilisé est celui de Cox. Celui-ci permet de modéliser les temps de survie avec des données censurées, et est donc très utilisé dans le domaine médical. Rappelons que le modèle de Cox se rapproche des modèles classiques de régression, dans le sens où l'on tente de relier un événement (modélisé par une date) à un certain nombre de variables explicatives. Il est aussi considéré comme un modèle semi-paramétrique basé sur l'hypothèse des risques proportionnels. Aussi, le modèle de régression de Cox est largement utilisé pour évaluer les associations entre les facteurs de risque potentiels et l'apparition de la maladie dans les études épidémiologiques.

Notons, par ailleurs, qu'en épidémiologie, le premier travail de modélisation a été réalisé,

à la fin du dix-huitième siècle, pour l'évaluation de l'efficacité de la variolisation, en termes de nombre de morts évités et de gain d'espérance de vie [21, 22]. Puis au dix-neuvième siècle, dans le domaine de l'épidémiologie des maladies transmissibles, Ross fut le premier à introduire l'analyse dans la planification de la lutte contre le paludisme et a développé le premier modèle mathématique sur celle-ci. Sa conviction fut qu' "*affirmer qu'une maladie est sous la dépendance de certains facteurs sert à bien peu de choses, à moins qu'il ne soit possible d'évaluer l'influence de chacun de ces facteurs sur le résultat final*". En effet, on sait que le paludisme est transmis par des anophèles. Ceci pousse à se poser la question suivante : "supprimer la moitié des anophèles suffit-il pour supprimer la moitié des cas du paludisme?". Malheureusement non! Pour appuyer ses déclarations, Ross a donc construit le modèle qui porte son nom, qui a permis de conforter ses hypothèses. C'est lui qui a introduit, pour la première fois, la notion de seuil. Ce seuil, qui sera rendu populaire plus tard par Kermack et Mckendrick (1933), sera désigné par  $R_0$ .

Comme les trois précédents paragraphes le montrent, ce chapitre, introductif, est divisé en trois parties. Notre but, dans ce chapitre, est de permettre au lecteur de mieux assimiler les notions élaborées par Sklar, Cox, Ross, Macdonald et autres.

Pour ce faire, nous entamerons la première partie par des notions préliminaires sur les fonctions copules et quelques résultats fondamentaux les concernant. Ensuite, nous expliquerons la notion de la mesure de dépendance afin d'introduire le tau de Kendall et le rho de Spearman. Mais encore, nous définirons les fonctions copules paramétriques qui se composent de trois différentes copules : la fonction copule normale, la fonction copule de Student et les fonctions copules Archimédiennes.

Tandis que, dans la seconde partie, nous nous focaliserons sur le modèle de Cox, en définissant la notion de risque instantané et quelques principes généraux, la dernière section sera consacrée au modèle de fragilité lié à la fonction de survie.

Dans la troisième et dernière partie de ce chapitre, nous entamerons l'analyse des différents systèmes épidémiologiques en commençant par le modèle de Ross et sa modélisation. Puis nous introduirons le modèle de Macdonald, qui est une généralisation du modèle de Ross. Ensuite, nous exposerons le modèle de Ross-Mckendrick. Enfin, nous discuterons la stabilité de ces systèmes dynamiques.

## I.2 Les fonctions copules

### I.2.1 Définitions

Les fonctions copules, introduites par Sklar [1], constituent un outil mathématique qui est relativement ancien.

**Définition I.1** Une fonction copule bidimensionnelle (ou 2-copula) est une fonction  $C$  qui possède les propriétés suivantes :

1.  $DomC = [0, 1] \times [0, 1]$  où  $DomC$  désigne le domaine de définition de  $C$
2.  $C(0, u) = C(u, 0) = 0$  et  $C(1, u) = C(u, 1) = u$ , pour tout  $u$  dans  $[0, 1]$
3.  $C$  est 2-increasing :

$$C(v_1, v_2) - C(v_1, u_2) - C(u_1, v_2) + C(u_1, u_2) \geq 0$$

pour tout  $(u_1, u_2) \in [0, 1]^2$ ,  $(v_1, v_2) \in [0, 1]^2$  tels que  $0 \leq u_1 \leq v_1 \leq 1$  et  $0 \leq u_2 \leq v_2 \leq 1$ .

A partir de cette définition, on peut dire que  $C$  est une distribution admettant des marges uniformes.

Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace probabilisé.

Soient  $U_1$  et  $U_2$  deux variables aléatoires uniformes sur  $\Omega$ . Considérons le vecteur aléatoire  $U = (U_1, U_2)$ . Nous avons :

$$C(u_1, u_2) = \mathbb{P}[U_1 \leq u_1, U_2 \leq u_2]$$

Le support de  $U$  est donc le carré unité  $[0, 1]^2$ . La propriété  $C(0, u) = C(u, 0) = 0$  implique que :

$$\mathbb{P}[U_1 \leq 0, U_2 \leq u] = \mathbb{P}[U_1 \leq u, U_2 \leq 0] = 0$$

$C$  doit donc vérifier la condition suivante :

$$\mathbb{P}[U_1 \leq 1, U_2 \leq u] = \mathbb{P}[U_1 \leq u, U_2 \leq 1] = u$$

Enfin,  $C$  est une distribution de probabilité, ce qui implique que :

$$\mathbb{P}[u_1 \leq U_1 \leq v_1, u_2 \leq U_2 \leq v_2] = C(v_1, v_2) - C(v_1, u_2) - C(u_1, v_2) + C(u_1, u_2) \geq 0$$

**Exemple I.1** *Il est facile de montrer que la fonction copule d'indépendance (ou la fonction copule produit)  $C^\perp(u_1, u_2) = u_1 u_2$ , est une fonction copule. Considérons uniquement la propriété 2-increasing. Comme nous avons  $v_2 - u_2 \geq 0$  et  $v_1 - u_1 \geq 0$ , nous en déduisons que :*

$$v_1(v_2 - u_2) \geq u_1(v_2 - u_2)$$

et

$$v_1 v_2 + u_1 u_2 - v_1 u_2 - u_1 v_2 \geq 0$$

Soient  $F_1$  et  $F_2$  deux lois de probabilité. Il est évident que  $C(F_1(x_1), F_2(x_2))$  définit une distribution de probabilité bidimensionnelle  $F(x_1, x_2)$  dont les marges sont données. En effet, nous savons que  $U_i = F_i(X_i)$  définit une transformation uniforme<sup>1</sup> (Probability Integral Transformation ou PIT). De plus, nous vérifions que :

$$\begin{aligned} C(F_1(x_1), F_2(\infty)) &= C(F_1(x_1), 1) \\ &= F_1(x_1) \end{aligned}$$

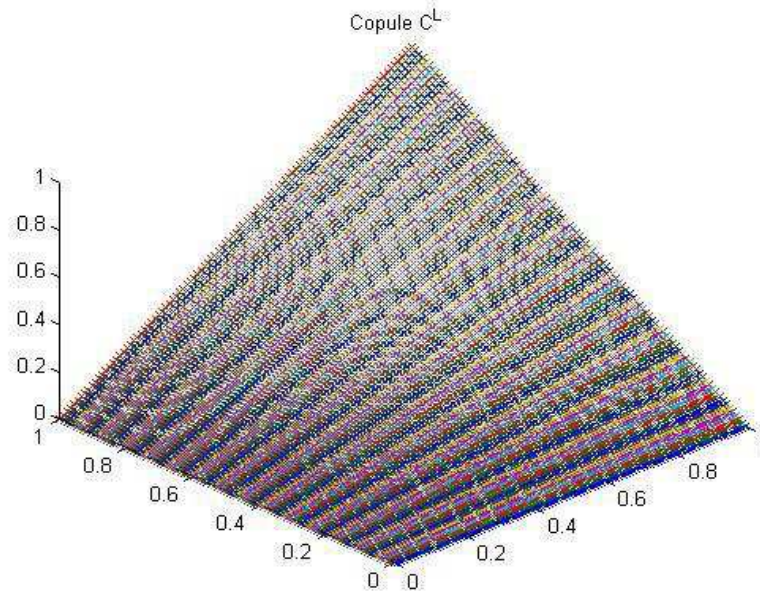
Les fonctions copules sont donc un outil très puissant permettant de construire des distributions multidimensionnelles dont les marges sont données.

---

1. Soit  $U = F(x)$ . Le support de  $U$  est évidemment  $[0, 1]$ . Et nous vérifions que la distribution de  $U$  est uniforme :

$$\begin{aligned} \mathbb{P}[U \leq u] &= \mathbb{P}[F(X) \leq u] \\ &= \mathbb{P}[X \leq F^{-1}(u)] \quad (\text{car } F \text{ est une fonction croissante}) \\ &= F(F^{-1}(u)) \\ &= u \end{aligned}$$



FIGURE I.1 – La fonction copule bivariable  $C^\perp$ 

**Définition I.2** Une fonction  $C : [0, 1]^n \rightarrow [0, 1]$  est une fonction copule  $n$ -dimensionnelle si elle satisfait les conditions suivantes :

1. Pour tout  $u$  dans  $[0, 1]^n$ ,  $C(u) = 0$  si au moins l'une des coordonnées de  $u$  est égale à 0, et  $C(u) = u_k$  si toutes les coordonnées de  $u$  sont égales à 1 à l'exception de  $u_k$ .
2. Pour tout  $a, b$  dans  $[0, 1]^n$  tel que  $a \leq b$ ,  $V_C([a, b]) \geq 0$ , où  $V_C([a, b])$  désigne le volume  $C$ , produit de  $n$  intervalles  $[a_i, b_i]$  défini par la formule suivante :

$$V_C([a, b]) = \Delta(a_n, b_n) \Delta(a_{n-1}, b_{n-1}) \dots \Delta(a_1, b_1) C(t)$$

avec

$$\Delta(a_k, b_k) C(t) = C(t_1, \dots, t_{k-1}, b_k, t_{k+1}, \dots, t_n) - C(t_1, \dots, t_{k-1}, a_k, t_{k+1}, \dots, t_n)$$

### I.2.2 Théorème de Sklar

Il s'agit d'un théorème fondamental dans la théorie sur les fonctions copules. En effet, à partir de la connaissance des lois marginales d'une fonction copule  $C$ , ce théorème (cf. Théorème I.1) permet d'obtenir la fonction de répartition multidimensionnelle. On peut, donc, affirmer que toute la théorie sur les fonctions copules est fondée sur le théorème suivant :

**Théorème I.1** Soit  $F$  une fonction de répartition  $n$ -dimensionnelle de fonctions de répartition marginales  $F_1, \dots, F_n$ . Alors, il existe une  $n$ -copule  $C$  telle que, pour tout  $x \in \mathbb{R}^n$  :

$$F(x) = F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)) \quad (1)$$

Si les fonctions  $F_1, \dots, F_n$  sont continues, alors  $C$  est unique. Dans le cas contraire,  $C$  est uniquement déterminée par le produit  $\text{Ran}F_1 \times \dots \times \text{Ran}F_n$  des images de  $F_1, \dots, F_n$ .

Inversement, si  $C$  est une  $n$ -copule et si  $F_1, \dots, F_n$  sont des fonctions de répartition univariées, alors la fonction  $F$  définie par (1) est une fonction de répartition  $n$ -dimensionnelle de fonctions de répartition marginales  $F_1, \dots, F_n$ .

La démonstration de ce théorème se trouve dans un travail récent de Rüschendorf [7].

Donc, le théorème de Sklar prouve que nous pouvons associer une fonction copule à chaque fonction de répartition multidimensionnelle. Nous pouvons, donc, décomposer une fonction de répartition multivariée en deux parties : d'une part les fonctions de répartition marginales, et d'autre part la fonction copule reliant ces fonctions. Ainsi, la fonction copule caractérise entièrement la structure de dépendance stochastique des variables aléatoires liées.

Notons aussi que le théorème de Sklar reste valable pour les fonctions de répartition multivariées de survie. En effet, en remplaçant  $F(x) = \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n)$  et  $F_i(x_i) = \mathbb{P}(X_i \leq x_i)$ ,  $i = 1, \dots, n$  par  $\bar{F}(x) = \mathbb{P}(X_1 \geq x_1, \dots, X_n \geq x_n)$  et  $\bar{F}_i(x_i) = \mathbb{P}(X_i \geq x_i)$ ,  $i = 1, \dots, n$ , nous obtenons :

$$\bar{F}(x) = \bar{C}(\bar{F}_1(x_1), \dots, \bar{F}_n(x_n)) \quad \text{pour tout } x \in \mathbb{R}^n$$

La fonction copule  $\bar{C}$  est appelée fonction copule de survie. Celle-ci est donnée en fonction de la fonction copule  $C$  par :

$$\bar{C}(u) = 1 + \sum_{1 \leq k \leq n} (-1)^k \sum_{i_1 \leq \dots \leq i_k} C_{i_1 \dots i_k}(u_{i_1}, \dots, u_{i_k}), \quad \text{pour tout } u = (u_{i_1}, \dots, u_{i_k}) \in [0, 1]^n \quad (2)$$

où les  $C_{i_1 \dots i_k}$  sont les  $k^{\text{èmes}}$  marges de la fonction copule  $C$

En particulier, dans le cas bivarié, nous obtenons  $\bar{C}(u_1, u_2) = 1 - u_1 - u_2 + C(u_1, u_2)$ .

Le théorème de Sklar fournit, donc, une méthode de construction des fonctions copules à partir des marges, en passant par les fonctions quantiles des lois marginales. Pour cela, nous rappelons la définition des quantiles associés à une fonction de répartition univariée.

**Définition I.3** Soit  $F$  une fonction de répartition univariée. Le quantile de  $F$  est une fonction, notée  $F^{-1}$ , de domaine de définition  $I = [0, 1]$ , telle que :

1. Si  $t$  est dans l'image  $\text{Ran}F$  de  $F$ , alors  $F^{-1}(t)$  est un nombre  $x$  tel que  $F(x) = t$ , c'est-à-dire, qui vérifie :

$$F(F^{-1}(t)) = t \quad \text{pour tout } t \in \text{Ran}F \quad (3)$$

2. Dans tous les cas, et en particulier si  $t$  n'est pas dans l'image  $\text{Ran}F$  de  $F$ ,  $F^{-1}(t)$  est définie par :

$$F^{-1}(t) = \inf \{x : F(x) \geq t\} = \sup \{x : F(x) \leq t\} \quad (4)$$

Si la fonction de répartition  $F$  est strictement croissante, alors sa fonction quantile  $F^{-1}$  coïncide avec l'inverse habituelle de  $F$  pour la loi de composition interne des applications.

**Corollaire I.2** Soit  $F$  une fonction de répartition  $n$ -dimensionnelle de fonctions de répartition marginales continues  $F_1, \dots, F_n$ . Alors, la fonction copule  $C$  associée à  $F$  est donnée par :

$$C(u_1, \dots, u_n) = F(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n)) \quad \text{pour tout } u \in I^n \quad (5)$$

**Exemple I.2** La fonction de répartition de la distribution logistique bivariée de Gumbel est donnée par :

$$F(x_1, x_2) = (1 + e^{-x_1} + e^{-x_2})^{-1} \quad (6)$$

qui est définie sur  $\mathbb{R}^2$ . Nous pouvons montrer que les fonctions de répartition marginales sont  $F_1(x_1) \equiv F(x_1, \infty) = (1 + e^{-x_1})^{-1}$  et  $F_2(x_2) = (1 + e^{-x_2})^{-1}$ . Les fonctions quantiles sont donc  $F_1^{-1}(u_1) = \ln u_1 - \ln(1 - u_1)$  et  $F_2^{-1}(u_2) = \ln u_2 - \ln(1 - u_2)$ . Nous en déduisons que la fonction copule est

$$\begin{aligned} C(u_1, u_2) &= F(F_1^{-1}(u_1), F_2^{-1}(u_2)) \\ &= \left(1 + \frac{1 - u_1}{u_1} + \frac{1 - u_2}{u_2}\right)^{-1} \\ &= \frac{u_1 u_2}{u_1 + u_2 - u_1 u_2} \end{aligned} \quad (7)$$

Cette fonction copule est appelée la fonction copule logistique de Gumbel (Gumbel Logistic copula).

### I.2.3 Densité de la fonction copule

Par analogie aux fonctions de répartition multivariées, les fonctions copules admettent des densités de probabilités. Si la densité  $c$  associée à la fonction copule  $C$  existe, alors elle est définie par :

$$c(u_1, \dots, u_n) = \frac{\partial^n C(u)}{\partial u_1 \dots \partial u_n} \quad (8)$$

Si la fonction de répartition multivariée  $F$  est absolument continue par rapport à la mesure de Lebesgue et en utilisant le théorème de Sklar, nous pouvons exprimer la densité d'un vecteur aléatoire  $(X_1, \dots, X_n)$  en fonction des densités marginales de sa fonction copule et de ses fonctions de répartition marginales  $F_1, \dots, F_n$  par :

$$f(x_1, \dots, x_n) = c(F_1(x_1), \dots, F_n(x_n)) \prod_{i=1}^n f_i(x_i) \quad (9)$$

A partir de la relation (9), nous pouvons calculer l'expression de la densité  $c$  de la fonction copule  $C$  via l'expression :

$$c(u_1, \dots, u_n) = \frac{f(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n))}{\prod_{i=1}^n f_i(F_i^{-1}(u_i))}$$

Cette dernière identité nous donne une seconde représentation canonique, mais qui porte désormais sur les densités. Ce résultat est important pour l'estimation des paramètres de la loi de probabilité d'un vecteur aléatoire  $(X_1, \dots, X_n)$  par la méthode du maximum de vraisemblance. Observons que la  $n$ -croissance de la copule  $C$  (cf. Définition I.2) correspond à la positivité de la densité :  $c(u_1, \dots, u_n) = \frac{\partial^n C(u_1, \dots, u_n)}{\partial u_1 \dots \partial u_n} \geq 0$  lorsque celle-ci existe.

**Exemple I.3** Reprenons la fonction copule logistique de Gumbel. L'expression de la densité bidimensionnelle est la suivante :

$$c(u_1, u_2) = \frac{2u_1 u_2}{(u_1 + u_2 - u_1 u_2)^3}$$

Il est alors facile de vérifier la propriété 2-increasing.

**Exemple I.4** La fonction copule  $C^\perp(u_1, \dots, u_n) = u_1 u_2 \dots u_n$  est la fonction copule produit ou la fonction copule d'indépendance (Product copula). Nous avons  $c^\perp = 1$ . Nous en déduisons que toute distribution construite avec cette fonction copule vérifie :

$$f(x_1, \dots, x_n) = \prod_{i=1}^n f_i(x_i)$$

La fonction copule produit caractérise donc l'indépendance entre les variables aléatoires.

Il faut noter que, pour certaines distributions, il n'existe pas d'expression analytique de la densité de la fonction copule. C'est par exemple le cas de la fonction copule normale.

### I.2.4 Propriétés des fonctions copules

Avant d'aborder deux propriétés fondamentales des fonctions copules, considérons quelques propriétés simples en théorie des probabilités. Nous avons :

$$\begin{aligned} \mathbb{P}[U_1 \leq u_1, U_2 > u_2] &= \mathbb{P}[U_1 \leq u_1] - \mathbb{P}[U_1 \leq u_1, U_2 \leq u_2] \\ &= u_1 - C(u_1, u_2) \end{aligned} \quad (10)$$

$$\begin{aligned} \mathbb{P}[U_1 \leq u_1 | U_2 \leq u_2] &= \frac{\mathbb{P}[U_1 \leq u_1, U_2 \leq u_2]}{\mathbb{P}[U_2 \leq u_2]} \\ &= \frac{C(u_1, u_2)}{u_2} \end{aligned} \quad (11)$$

$$\begin{aligned} \mathbb{P}[U_1 \leq u_1 | U_2 > u_2] &= \frac{\mathbb{P}[U_1 \leq u_1, U_2 > u_2]}{\mathbb{P}[U_2 > u_2]} \\ &= \frac{u_1 - C(u_1, u_2)}{1 - u_2} \end{aligned} \quad (12)$$

$$\begin{aligned} \mathbb{P}[U_1 > u_1, U_2 > u_2] &= \mathbb{P}[U_1 > u_1] - \mathbb{P}[U_1 > u_1, U_2 \leq u_2] \\ &= 1 - u_1 - u_2 + C(u_1, u_2) \end{aligned} \quad (13)$$

$$\begin{aligned} \mathbb{P}[U_1 > u_1 | U_2 > u_2] &= \frac{\mathbb{P}[U_1 > u_1, U_2 > u_2]}{\mathbb{P}[U_2 > u_2]} \\ &= \frac{1 - u_1 - u_2 + C(u_1, u_2)}{1 - u_2} \end{aligned} \quad (14)$$

La distribution conditionnelle de  $U_1$  sachant  $U_2$ ,  $C_{1|2}(u_1, u_2)$ , est définie par :

$$\begin{aligned} C_{1|2}(u_1, u_2) &= \mathbb{P}[U_1 \leq u_1 | U_2 = u_2] \\ &= \lim_{\Delta u \rightarrow 0^+} \frac{\mathbb{P}[U_1 \leq u_1, u_2 \leq U_2 \leq u_2 + \Delta u]}{\mathbb{P}[u_2 \leq U_2 \leq u_2 + \Delta u]} \\ &= \lim_{\Delta u \rightarrow 0^+} \frac{C(u_1, u_2 + \Delta u) - C(u_1, u_2)}{\Delta u} \\ &= \partial_2 C(u_1, u_2) \end{aligned} \quad (15)$$

où  $\partial_2 C(u_1, u_2) = \frac{\partial C(u_1, u_2)}{\partial u_2}$ .

A partir d'une fonction copule  $C$ , nous pouvons construire trois autres copules :

$$\begin{aligned}\bar{C}(u_1, u_2) &= u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2) \\ \tilde{C}(u_1, u_2) &= u_1 + u_2 - C(u_1, u_2) \\ C^*(u_1, u_2) &= 1 - C(1 - u_1, 1 - u_2)\end{aligned}\tag{16}$$

Les fonctions copules  $\bar{C}$ ,  $\tilde{C}$  et  $C^*$  sont appelées, respectivement, la fonction copule de survie (survival copula), la fonction copule duale (dual copula) et la co-copule (co-copula). Soient  $X_1$  et  $X_2$  deux variables aléatoires. Les interprétations probabilistes sont les suivantes :

$$\begin{aligned}\mathbb{P}[X_1 \leq x_1, X_2 \leq x_2] &= C(F_1(x_1), F_2(x_2)) \\ \mathbb{P}[X_1 \geq x_1, X_2 \geq x_2] &= \bar{C}(S_1(x_1), S_2(x_2)) \\ \mathbb{P}[X_1 \leq x_1 \text{ ou } X_2 \leq x_2] &= \tilde{C}(F_1(x_1), F_2(x_2)) \\ \mathbb{P}[X_1 \geq x_1 \text{ ou } X_2 \geq x_2] &= C^*(S_1(x_1), S_2(x_2))\end{aligned}\tag{17}$$

avec  $S(x) = 1 - F(x)$  qui désigne la fonction de survie.

Maintenant, nous sommes en mesure d'énoncer deux propriétés importantes :

**Théorème I.3** *Une fonction copule  $C$  est uniformément continue sur son domaine. En particulier, nous avons :*

$$|C(v_1, v_2) - C(u_1, u_2)| \leq |v_1 - u_1| + |v_2 - u_2|\tag{18}$$

**Théorème I.4** *Les dérivées partielles  $\partial_1 C$  et  $\partial_2 C$  existent (presque sûrement) pour tout  $(u_1, u_2)$  dans  $[0, 1]^2$ . Elles satisfont, également, les conditions suivantes  $0 \leq \partial_1 C(u_1, u_2) \leq 1$  et  $0 \leq \partial_2 C(u_1, u_2) \leq 1$ .*

La preuve du premier théorème est immédiate si l'on considère l'inégalité triangulaire (cf. [4]). La démonstration rigoureuse du second théorème est plus délicate (cf. [8]).

### I.2.5 Les classes de Fréchet

Les classes de Fréchet sont les classes des distributions multidimensionnelles avec des marges données. Ces marges peuvent être unidimensionnelles, multidimensionnelles et aussi conditionnelles. Soient, par exemple,  $X_1$ ,  $X_2$  et  $X_3$  trois variables aléatoires avec, respectivement,  $F_1$ ,  $F_2$  et  $F_3$  leurs fonctions de distribution marginales, et soient  $F_{12}$  et  $F_{23}$  les fonctions des distributions bivariées correspondant aux couples  $(X_1, X_2)$  et  $(X_2, X_3)$ . La classe de Fréchet  $\mathcal{F}(F_{12}; F_{23})$  est l'ensemble des distributions trivariées qui sont compatibles avec ces marges. Puisque nous connaissons  $F_{12}$  et  $F_{23}$ , nous connaissons aussi  $F_1$ ,  $F_2$  et  $F_3$ . Nous savons que toute distribution  $F$  de dimension 3 s'écrit :

$$F(x_1, x_2, x_3) = C(F_1(x_1), F_2(x_2), F_3(x_3))\tag{19}$$

Néanmoins, le fait que  $F \in \mathcal{F}$  impose certaines conditions sur la fonction  $C$ . L'étude des classes de Fréchet revient à étudier ces restrictions : les points extrémaux de  $\mathcal{F}$ , les bornes, etc.

**Définition I.4** On dit que  $F$  appartient à la classe de Fréchet  $(F_1, F_2)$  qu'on note par  $F \in \mathcal{F}(F_1, F_2)$  si, et seulement si, les marges de  $F$  sont  $F_1$  et  $F_2$  :

$$F(x_1, \infty) = F_1(x_1) \quad (20)$$

$$F(\infty, x_2) = F_2(x_2) \quad (21)$$

Caractériser la classe de Fréchet  $\mathcal{F}(F_1, F_2)$  revient à trouver l'ensemble  $\mathcal{C}$  des fonctions copules, puisque nous avons :

$$\mathcal{F}(F_1, F_2) = \{F ; F(x_1, x_2) = C(F_1(x_1), F_2(x_2)) \text{ où } C \in \mathcal{C}\} \quad (22)$$

Le problème de la caractérisation de la classe de Fréchet  $\mathcal{F}(F_1, F_2)$  est donc indépendant des marges  $F_1$  et  $F_2$ .

Les distributions extrémales (qui sont aussi les bornes)  $F^-$  et  $F^+$  de  $\mathcal{F}(F_1, F_2)$  sont données par :

$$F^-(x_1, x_2) = \max(F_1(x_1) + F_2(x_2) - 1, 0) \quad (23)$$

$$F^+(x_1, x_2) = \min(F_1(x_1), F_2(x_2)) \quad (24)$$

$F^-$  et  $F^+$  sont appelées, respectivement, les bornes inférieure et supérieure de Fréchet. Les fonctions copules qui en sont associées sont données par :

$$C^-(u_1, u_2) = \max(u_1 + u_2 - 1, 0) \quad (25)$$

$$C^+(u_1, u_2) = \min(u_1, u_2) \quad (26)$$

Dans le cas multidimensionnel, nous notons :

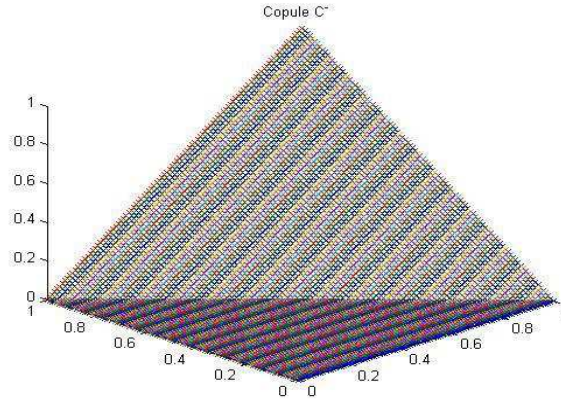
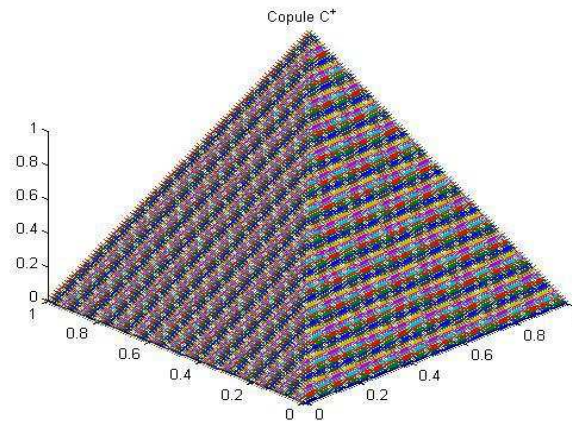
$$C^-(u_1, \dots, u_n) = \max\left(\sum_{i=1}^n u_i - n + 1, 0\right) \quad (27)$$

$$C^+(u_1, \dots, u_n) = \min(u_1, \dots, u_n) \quad (28)$$

Nous pouvons montrer que  $C^+$  est une fonction copule, mais que  $C^-$  n'est pas dans  $\mathcal{C}$ . Néanmoins,  $C^-$  est une borne inférieure atteignable. Cela veut dire que pour tout  $(u_1, \dots, u_n)$  appartenant à  $[0, 1]^n$ , il existe une fonction copule qui coïncide avec  $C^-$ . Cela veut dire aussi que  $\mathcal{F}(F_1, \dots, F_n)$  admet une fonction de distribution minimale si, et seulement si,  $\max(\sum_{i=1}^n F_i(x_i) - n + 1, 0)$  est une distribution de probabilité. Malheureusement, cela implique des restrictions très contraignantes (cf. [9]).

**Remarque I.1** Les bornes inférieures et supérieures de  $\mathcal{F}(F_1, \dots, F_n)$  sont, bien entendu, des points extrémaux de  $\mathcal{F}$ . Néanmoins, ce ne sont pas les seuls. En fait, il existe une infinité de points extrémaux qu'il est difficile de caractériser.

Sur les graphiques I.2 et I.3, nous représentons les distributions des fonctions copules  $C^-$  et  $C^+$ .

FIGURE I.2 – La fonction copule bivariable  $C^-$ FIGURE I.3 – La fonction copule bivariable  $C^+$ 

## I.2.6 Mesure de dépendance

Nous pouvons interpréter la fonction copule du vecteur aléatoire  $X = (X_1, X_2)$  comme une reparamétrisation ou une normalisation de la distribution conjointe  $F$ , après avoir éliminé les effets de marges  $F_1$  et  $F_2$ . Deheuvels (cf. [49, 50]) lui a donné le nom de fonction de dépendance de  $X$ . De plus, c'est une statistique exhaustive de cette dépendance. Considérons une statistique (non exhaustive) de la dépendance, c'est-à-dire une mesure (un résumé) de la dépendance. Celle-ci sera effectivement une mesure de dépendance, si elle s'exprime exclusivement à partir de la fonction copule.

Parmi toutes ces statistiques, nous pouvons considérer de façon générale les mesures dites de concordance.

**Définition I.5** Soient  $C_1$  et  $C_2$  deux fonctions copules. On dit que  $C_1$  est plus petite que  $C_2$  et on note  $C_1 \prec C_2$  si, et seulement si,  $C_1(u_1, u_2) \leq C_2(u_1, u_2)$  pour tout  $(u_1, u_2) \in [0, 1]^2$ .

**Remarque I.2** L'ordre  $\prec$  est appelé "ordre de concordance". Il correspond à la dominance stochastique du premier ordre sur les fonctions de distribution.

**Définition I.6** Une mesure numérique  $M$  d'association entre deux variables aléatoires continues  $X$  et  $Y$  dont la fonction copule est  $C$  (que nous notons par  $M_{X,Y}$  ou  $M_C$ ) est une mesure

de concordance, si elle satisfait les propriétés suivantes :

1.  $M_{X,Y}$  est définie pour toute paire  $(X,Y)$  de variables aléatoires continues
2.  $M_{X,Y} \in (-1,1)$  avec  $M_{X,X} = 1$  et  $M_{X,-X} = -1$
3.  $M_{X,Y} = M_{Y,X}$
4. Si  $X$  et  $Y$  sont indépendantes, alors  $M_{X,Y} = 0$
5.  $M_{-X,Y} = M_{X,-Y} = -M_{X,Y}$
6. Si  $C_1$  et  $C_2$  sont deux fonctions copules avec  $C_1 \prec C_2$  alors  $M_{C_1} \leq M_{C_2}$
7. Si  $\{(X_n, Y_n)\}_{n \in \mathbb{N}^*}$  une suite de variables aléatoires continues, dont la fonction copule  $C_n$  converge vers  $C$ , alors  $\lim_{n \rightarrow +\infty} M_{C_n} = M_C$

La propriété 6., ci-dessus, permet de comprendre le pourquoi de l'appellation de mesure de concordance puisque l'ordre de concordance implique l'ordre sur  $M$ . Parmi toutes les mesures de concordance deux mesures, très célèbres, jouent un rôle important en statistique non paramétrique : le tau de Kendall et le rhô de Spearman. Ces deux mesures peuvent, toutes les deux, s'exprimer à l'aide de la fonction copule.

### I.2.6.1 Tau de Kendall

Après la définition de la mesure de concordance, nous définissons maintenant la notion de concordance (cf. [10]) : deux observations  $(x_1, y_1)$  et  $(x_2, y_2)$  d'une paire  $(X, Y)$  de variables aléatoires continues sont dites concordantes, si les deux valeurs de la 1<sup>ère</sup> paire sont plus grandes que les valeurs correspondantes de la 2<sup>ème</sup> paire, c'est-à-dire si  $x_1 < x_2, y_1 < y_2$  ou  $x_1 > x_2, y_1 > y_2$ ; et elles sont dites discordantes si, pour une paire, une valeur est plus grande et l'autre est plus petite que les valeurs correspondantes de l'autre paire, c'est-à-dire si  $x_1 < x_2, y_1 > y_2$  ou  $x_1 > x_2, y_1 < y_2$ .

Donc le tau de Kendall est défini en termes de concordance comme suit :

**Définition I.7** Soient  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  un échantillon aléatoire de  $n$  observations à partir d'un vecteur aléatoire continu  $(X, Y)$ . Le tau de Kendall de cet échantillon aléatoire est défini par :

$$\tau = \frac{\text{nombre de paires concordantes} - \text{nombre de paires discordantes}}{\text{nombre total de paires}} \quad (29)$$

De façon équivalente, on peut définir  $\tau$  comme la différence de la probabilité de concordance et de la probabilité de discordance pour une paire d'observations  $(x_i, y_i)$  et  $(x_j, y_j)$  choisies aléatoirement à partir de l'échantillon. Le tau de Kendall pour un vecteur  $(X, Y)$  de variables aléatoires continues est défini de la même manière.

Soient  $(X_1, Y_1)$  et  $(X_2, Y_2)$  deux vecteurs aléatoires indépendants et identiquement distribués (i.i.d), chacun de fonction de distribution conjointe  $H$ . Alors, la probabilité pour que  $(X_1, Y_1)$  et  $(X_2, Y_2)$  soient concordants est équivalente à :

$$\mathbb{P}((X_1 - X_2)(Y_1 - Y_2) > 0)$$

et la probabilité pour que ces deux vecteurs soient discordants est équivalente à :

$$\mathbb{P}((X_1 - X_2)(Y_1 - Y_2) < 0)$$

Le tau de Kendall peut, alors, être défini comme suit :



**Définition I.8** *Le tau de Kendall est défini comme étant la différence entre la probabilité de concordance et la probabilité de discordance :*

$$\tau = \mathbb{P}((X_1 - X_2)(Y_1 - Y_2) > 0) - \mathbb{P}((X_1 - X_2)(Y_1 - Y_2) < 0) \quad (30)$$

**Théorème I.5** *Soient  $X$  et  $Y$  deux variables aléatoires continues dont la fonction copule est  $C$ . Alors, le tau de Kendall pour  $X$  et  $Y$  est donné par :*

$$\tau = 4 \int_0^1 \int_0^1 C(u, v) c(u, v) du dv - 1 \quad (31)$$

*Preuve :*

$$\begin{aligned} \tau &= \mathbb{P}\{(X_1 - X_2)(Y_1 - Y_2) > 0\} - \mathbb{P}\{(X_1 - X_2)(Y_1 - Y_2) < 0\} \\ &= 2 \mathbb{P}\{(X_1 - X_2)(Y_1 - Y_2) > 0\} - 1 \\ &= 2 \mathbb{P}\{(X_1 > X_2 ; Y_1 > Y_2) \cup (X_1 < X_2 ; Y_1 < Y_2)\} - 1 \\ &= 2 [\mathbb{P}\{(X_1 > X_2 ; Y_1 > Y_2)\} + \mathbb{P}\{(X_1 < X_2 ; Y_1 < Y_2)\}] - 1 \\ &= 4 \mathbb{P}\{(X_1 > X_2 ; Y_1 > Y_2)\} - 1 \\ &= 4 \int_x \int_y \mathbb{P}\{X_2 \leq x ; Y_2 \leq y \mid X_1 = x ; Y_1 = y\} f_{XY}(x, y) dx dy - 1 \\ &= 4 \int_x \int_y F_{XY}(x, y) f_{XY}(x, y) dx dy - 1 \\ &= 4 \int_x \int_y C(F_X(x), F_Y(y)) f_{XY}(x, y) dx dy - 1 \end{aligned}$$

En faisant le changement de variables  $u = F_X(x)$  et  $v = F_Y(y)$ , nous obtenons :

$$\tau = 4 \int_0^1 \int_0^1 C(u, v) c(u, v) du dv - 1$$

où  $c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v}$  ■

L'intégrale qui apparaît dans l'équation (31) peut être interprétée comme la valeur absolue de la fonction  $C(U, V)$ , où  $U$  et  $V$  sont des variables aléatoires uniformément réparties sur  $[0, 1]$ , et pour lesquelles la fonction de distribution est  $C$ , c'est-à-dire :

$$\tau = 4 \mathbb{E}(C(U, V)) - 1 = 4 \int_0^1 t f_X(t) dt - 1 \quad (32)$$

où  $f_X(t)$  est la fonction de densité de probabilité de  $X$ .

### I.2.6.2 Rhô de Spearman

Comme pour le tau de Kendall, le rhô de Spearman est un coefficient qui est basé sur la concordance et la discordance des valeurs. Soient  $(X_1, Y_1)$ ,  $(X_2, Y_2)$  et  $(X_3, Y_3)$  trois vecteurs aléatoires indépendants issus de variables aléatoires continues de fonction de distribution conjointe commune  $H$  (dont les marges sont  $F_X$  et  $F_Y$ ) et soit  $C$  une fonction copule. Alors,

le rhô de Spearman est égal à trois fois la différence entre la probabilité de concordance et celle de discordance des deux vecteurs  $(X_1, X_2)$  et  $(Y_1, Y_3)$ , c'est-à-dire :

$$\rho = 3 [\mathbb{P}((X_1 - X_2)(Y_1 - Y_3) > 0) - \mathbb{P}((X_1 - X_2)(Y_1 - Y_3) < 0)] \quad (33)$$

La paire  $(X_1, X_3)$  pourrait être utilisée tout aussi bien.

**Théorème I.6** Soient  $X$  et  $Y$  deux variables aléatoires continues dont  $C$  est la fonction copule. Alors, le rhô de Spearman est donné par :

$$\rho = 12 \int \int_{I^2} u_1 u_2 dC(u_1, u_2) - 3 \quad (34)$$

*Preuve :*

On a :

$$\rho = 3 [\mathbb{P}((X_1 - X_2)(Y_1 - Y_3) > 0) - \mathbb{P}((X_1 - X_2)(Y_1 - Y_3) < 0)]$$

A partir du Théorème I.5, on peut écrire le rhô de Spearman sous la forme suivante :

$$\rho = 3 \left[ 4 \int_x \int_y C(F_X(x), F_Y(y)) dC(F_X(x), F_Y(y)) - 1 \right]$$

En faisant le changement de variables  $u_1 = F_X(x)$  et  $u_2 = F_Y(y)$ , nous obtenons :

$$\rho = 12 \int_0^1 \int_0^1 C(u_1, u_2) dC(u_1, u_2) - 3$$

où  $dC(u_1, u_2) = \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2}$

Puisque  $X_2$  et  $Y_3$  sont indépendantes, alors  $C(u_1, u_2) = u_1 u_2$  et :

$$\rho = 12 \int_0^1 \int_0^1 u_1 u_2 dC(u_1, u_2) - 3$$

■

## I.2.7 Les fonctions copules paramétriques

Dans cette section, nous étudions quelques fonctions copules dites paramétriques, c'est-à-dire des fonctions qui dépendent d'un ou de plusieurs paramètres.

### I.2.7.1 La fonction copule normale

La fonction copule normale est la fonction copule associée à la distribution normale multidimensionnelle (cf. Définition I.9)

**Définition I.9** Soit  $\rho$  une matrice diagonale définie positive avec  $\text{diag}(\rho) = 1$ ,  $\Phi$  la distribution normale standard, et de matrice de corrélation  $\rho$ . La fonction copule normale est alors :

$$C(u_1, \dots, u_n; \rho) = \Phi_\rho(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n))$$

Utilisons la relation qui relie la densité multivariée, les densités univariées marginales et la densité de la copule :

$$f(x_1, \dots, x_n) = c(F_1(x_1), \dots, F_n(x_n)) \prod_{i=1}^n f_i(x_i)$$

et appliquons cette relation à la densité de la distribution normale multivariée :

$$\frac{1}{(2\pi)^{\frac{n}{2}} |\rho|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} x^t \rho^{-1} x\right) = c(\Phi(x_1), \dots, \Phi(x_n)) \times \left(\prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} x_i^2\right)\right)$$

Nous en déduisons que la densité de la fonction copule normale est :

$$c(u_1, \dots, u_n; \rho) = \frac{1}{|\rho|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \zeta^t (\rho^{-1} - I) \zeta\right)$$

où  $\zeta_i = \Phi^{-1}(u_i)$  et  $I$  est la matrice identité de dimension  $n \times n$ .

Dans le cas bivarié, nous obtenons :

$$c(u_1, u_2; \rho) = \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{\zeta_1^2 + \zeta_2^2 - 2\rho\zeta_1\zeta_2}{2(1-\rho^2)} + \frac{\zeta_1^2 + \zeta_2^2}{2}\right)$$

Nous en déduisons que :

$$C(u_1, u_2; \rho) = \int_0^{u_1} \int_0^{u_2} \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{\zeta_1^2 + \zeta_2^2 - 2\rho\zeta_1\zeta_2}{2(1-\rho^2)} + \frac{\zeta_1^2 + \zeta_2^2}{2}\right) dx_1 dx_2$$

avec  $\zeta_1 = \Phi^{-1}(x_1)$  et  $\zeta_2 = \Phi^{-1}(x_2)$ .

**Théorème I.7** *La distribution de la fonction copule normale bivariée est :*

$$C(u_1, u_2; \rho) = \int_0^{u_1} \Phi\left(\frac{\Phi^{-1}(u_2) - \rho\Phi^{-1}(u)}{\sqrt{1-\rho^2}}\right) du$$

**Preuve :**

Soient  $X = (X_1, X_2)$  un vecteur aléatoire normal dont la corrélation est  $\rho$  et  $X_3$  une variable normale indépendante de  $X_1$  et  $X_2$ . Nous avons :

$$\begin{aligned} \Phi(x_1, x_2; \rho) &= \mathbb{P}[X_1 \leq x_1, X_2 \leq x_2] \\ &= \mathbb{E}\left[\mathbb{P}\left(X_1 \leq x_1, \rho X_1 + \sqrt{1-\rho^2} X_3 \leq x_2 \mid X_1\right)\right] \\ &= \int_{-\infty}^{x_1} \Phi\left(\frac{x_2 - \rho x}{\sqrt{1-\rho^2}}\right) \Phi(x) dx \end{aligned}$$

L'expression de la distribution de la fonction copule normale est donc :

$$\begin{aligned} C(u_1, u_2; \rho) &= \int_{-\infty}^{\Phi^{-1}(u_1)} \Phi\left(\frac{\Phi^{-1}(u_2) - \rho x}{\sqrt{1-\rho^2}}\right) \Phi(x) dx \\ &= \int_0^{u_1} \Phi\left(\frac{\Phi^{-1}(u_2) - \rho\Phi^{-1}(u)}{\sqrt{1-\rho^2}}\right) du \end{aligned}$$

■

### I.2.7.2 La fonction copule t ou de Student

Comme pour la fonction copule normale, la fonction copule t (ou la fonction copule de Student) est la fonction de dépendance associée à la distribution t multidimensionnelle.

**Définition I.10** Soit  $\rho$  une matrice diagonale définie positive avec  $\text{diag}(\rho) = 1$ . Notons  $t_{\rho,\nu}$  la distribution de Student multivariée standard à  $\nu$  degrés de liberté et de matrice de corrélation  $\rho$ . La fonction copule de Student est alors définie de la façon suivante :

$$C(u_1, \dots, u_n; \rho, \nu) = t_{\rho,\nu} (t_\nu^{-1}(u_1), \dots, t_\nu^{-1}(u_n))$$

En utilisant la définition de la distribution de Student, nous obtenons :

$$t_{\rho,\nu}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} \frac{\Gamma(\frac{\nu+n}{2}) |\rho|^{-\frac{1}{2}}}{\Gamma(\frac{\nu}{2}) (\nu\pi)^{\frac{n}{2}}} \left( 1 + \frac{1}{\nu} y^t \rho^{-1} y \right)^{-\frac{\nu+n}{2}} dy$$

Nous pouvons montrer, également, que la densité de la fonction copule de Student est :

$$c(u_1, \dots, u_n; \rho) = |\rho|^{-\frac{1}{2}} \frac{\Gamma(\frac{\nu+n}{2}) [\Gamma(\frac{\nu}{2})]^n (1 + \frac{1}{\nu} \zeta^t \rho^{-1} \zeta)^{-\frac{\nu+n}{2}}}{[\Gamma(\frac{\nu+1}{2})]^n \Gamma(\frac{\nu}{2}) \prod_{i=1}^n (1 + \frac{\zeta_i^2}{\nu})^{-\frac{\nu+1}{2}}}$$

où  $\zeta_i = t_\nu^{-1}(u_i)$ .

Dans le cas bivarié, nous déduisons que l'expression de la fonction copule de Student est donnée par :

$$C(u_1, u_2; \rho, \nu) = \int_{-\infty}^{t_\nu^{-1}(u_1)} \int_{-\infty}^{t_\nu^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \times \left( 1 + \frac{x_1^2 + x_2^2 - 2\rho x_1 x_2}{\nu(1-\rho^2)} \right)^{-\frac{\nu+2}{2}} dx_1 dx_2$$

### I.2.7.3 Les fonctions copules archimédiennes

La classe la plus importante des fonctions copules est la classe des fonctions copules dites archimédiennes. Les fonctions copules archimédiennes demeurent associées au statisticien canadien Christian Genest. Ce n'est pas lui qui les a inventées, mais il est le premier à avoir proposé une analyse statistique de ces fonctions copules. Parmi ses nombreuses publications, les articles [11, 12, 13, 14] ont largement contribué à les faire connaître. En fait, ces copules archimédiennes dérivent des  $t$ -normes archimédiennes (cf. [15]). Comme il existe un lien très fort entre les fonctions copules et les  $t$ -normes, la construction de ces fonctions copules devient immédiate (cf. [16]).

**Définition I.11** Les fonctions copules archimédiennes multidimensionnelles sont définies comme suit :

$$C(u_1, \dots, u_n) = \begin{cases} \phi^{-1}(\phi(u_1) + \dots + \phi(u_n)) & \text{si } \phi(u_1) + \dots + \phi(u_n) \leq 0 \\ 0 & \text{sinon} \end{cases} \quad (35)$$

où  $\phi$  désigne une fonction deux fois continûment différentiable et qui vérifie :

$$\phi(1) = 0, \quad \phi^{(1)}(u) < 0 \quad \text{et} \quad \phi^{(2)}(u) > 0 \quad \text{pour tout } u \in [0, 1]^n$$

La fonction  $\phi(u)$  est appelée le générateur (ou la fonction génératrice) de la fonction copule et  $\phi^{(j)}$  désigne sa  $j^{\text{ème}}$  dérivée. Ce générateur est dit strict si  $\phi(0) = \infty$  et non strict si  $\phi(0) < \infty$ .

Dans [11, 12], Genest et MacKay donnent une caractérisation d'une fonction copule archimédienne en utilisant le critère d'Abel :

**Théorème I.8** *Une fonction copule  $C$  est dite archimédienne si elle possède deux dérivées partielles et s'il existe une application intégrable  $\zeta : [0, 1] \rightarrow [0, \infty[$ , telle que :*

$$\zeta(u_2) \frac{\partial C(u_1, u_2)}{\partial u_1} = \zeta(u_1) \frac{\partial C(u_1, u_2)}{\partial u_2} \quad (36)$$

pour tout  $(u_1, u_2) \in [0, 1]^2$ . Le cas échéant, l'application  $\phi$  qui engendre  $C$  est donnée (à une constante près) par :

$$\phi(u) = \int_u^1 \zeta(u) du$$

où  $0 \leq u \leq 1$

En voici d'autres propriétés (cf. [17, 18]) :

- La fonction copule  $C$  est symétrique, i.e.  $C(u_1, u_2) = C(u_2, u_1)$ .
- La fonction copule  $C$  est associative, i.e.  $C(u_1, C(u_2, u_3)) = C(C(u_1, u_2), u_3)$ .
- La section diagonale  $\delta(u) = C(u, u)$  vérifie  $\delta(u) < u$  pour tout  $u \in [0, 1]$ .
- Si une fonction copule  $C$  est associative et  $\delta(u) < u$  pour tout  $u \in [0, 1]$ , alors  $C$  est archimédienne.

Les fonctions copules archimédiennes jouent un rôle important, car elles présentent de nombreuses propriétés intéressantes et sont très maniables, par exemple le tau de Kendall  $\tau$  pour ces copules est égal à :

$$\tau = 1 + 4 \int_0^1 \frac{\phi(u)}{\phi^{(1)}(u)} du$$

**Exemple I.5** *Considérons la fonction copule de Clayton. Nous avons  $\phi(u) = u^{-\theta} - 1$  et  $\phi^{(1)} = -\theta u^{-\theta-1}$ . Nous en déduisons que :*

$$\begin{aligned} \tau &= 1 + 4 \int_0^1 \frac{1 - u^{-\theta}}{\theta u^{-\theta-1}} du = 1 + \frac{4}{\theta} \int_0^1 (u^{\theta+1} - u) du \\ &= 1 + \frac{4}{\theta} \left[ \frac{u^{\theta+2}}{\theta+2} - \frac{u^2}{2} \right]_0^1 = 1 + \frac{4}{\theta} \left( \frac{1}{\theta+2} - \frac{1}{2} \right) \\ &= \frac{\theta}{\theta+2} \end{aligned} \quad (37)$$

Pour une fonction copule archimédienne, la densité s'exprime de la manière suivante :

$$c(u_1, \dots, u_n) = (\phi^{-1})^{(n)}(\phi(u_1) + \dots + \phi(u_n)) \prod_{i=1}^n \phi^{(1)}(u_i)$$

Dans ce qui suit, nous donnons deux exemples de fonctions copules archimédiennes, qui sont la fonction copule de Gumbel et la fonction copule de Clayton. Nous nous intéresserons particulièrement au cas bivarié, et nous opérerons ensuite une généralisation au cas multivarié.

### La fonction copule de Gumbel

**Définition I.12** La fonction copule de Gumbel bivariée est définie comme suit :

$$C(u, v) = \exp\{-[(-\ln u)^\alpha + (-\ln v)^\alpha]^{\frac{1}{\alpha}}\}$$

où  $\alpha \geq 1$  est le paramètre de régression de la fonction copule.

Le générateur de la fonction copule de Gumbel est donné par :  $\phi(t) = (-\ln t)^\alpha$ .

Le tau de Kendall de cette fonction copule s'exprime de la manière suivante :  $\tau = (\alpha - 1)/\alpha$

Dans le cas multivarié, la fonction copule de Gumbel se définit de la façon suivante :

$$C(u_1, \dots, u_n) = \exp\left(-\left[\sum_{i=1}^n (-\ln u_i)^\alpha\right]^{\frac{1}{\alpha}}\right)$$

Malheureusement, dans le cas de cette fonction copule, la dérivée  $n^{\text{ème}}$  de l'inverse du générateur n'a pas d'expression explicite, donc a fortiori la densité de la fonction copule de Gumbel multivariée n'a pas une forme explicite.

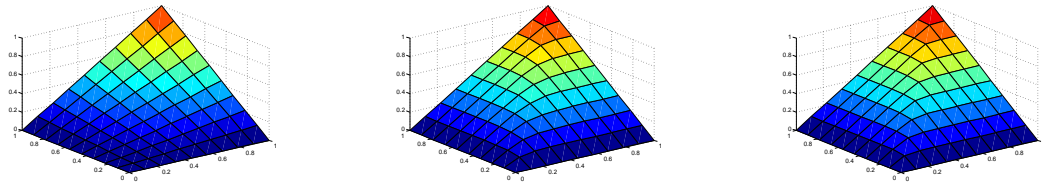


FIGURE I.4 – La fonction distribution de la fonction copule bivariée de Gumbel (resp. pour  $\alpha = 2$ ,  $\alpha = 5$  et  $\alpha = 10$ )

### La fonction copule de Clayton

**Définition I.13** La fonction copule de Clayton bivariée est définie par :

$$C(u, v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-1} \theta$$

où  $\theta > 0$  est le paramètre de régression.

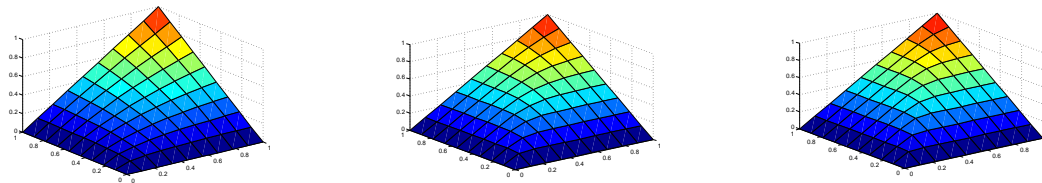


FIGURE I.5 – La fonction distribution de la fonction copule bivariée de Clayton (resp. pour  $\alpha = 2$ ,  $\alpha = 5$  et  $\alpha = 10$ )

Le générateur de la fonction copule de Clayton est donné par :  $\phi(u) = u^{-\theta} - 1$ .

Le tau de Kendall de cette fonction copule s'exprime de la manière suivante :  $\tau = \theta/(\theta + 2)$

Dans le cas multivarié, la fonction copule de Clayton se définit par :

$$C(u_1, \dots, u_n) = \left( \sum_{i=1}^n (u_i)^{-\theta} - n + 1 \right)^{\frac{-1}{\theta}}$$

Dans le tableau suivant, nous faisons un récapitulatif de quelques fonctions copules archimédiennes :

Copule	$\phi(u)$	$C(u_1, u_2)$
$c^\perp$	$-\ln u$	$u_1 u_2$
Gumbel	$(-\ln u)^\theta$	$\exp\left(-\left((-\ln(u_1))^\theta + (-\ln(u_2))^\theta\right)^{\frac{1}{\theta}}\right)$
Frank	$-\ln\left(\frac{e^{-\theta u} - 1}{e^{-\theta} - 1}\right)$	$\frac{-1}{\theta} \ln\left(1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1}\right)$
Clayton	$u^{-\theta} - 1$	$(u_1^{-\theta} + u_2^{-\theta} - 1)^{\frac{-1}{\theta}}$
Joe	$-\ln(1 - (1 - u)^\theta)$	$1 - \left((1 - u_1)^\theta + (1 - u_2)^\theta - (1 - u_1)^\theta(1 - u_2)^\theta\right)^{\frac{1}{\theta}}$

### I.3 Modèle de Cox

En épidémiologie étiologique, on est amené à prendre en compte simultanément les rôles de plusieurs facteurs de risque dans la survenue d'une maladie. Les modèles multivariés permettent de représenter la variable étudiée en fonction de plusieurs autres variables. L'un des principaux modèles utilisés est le modèle de Cox (cf. [19, 20]).

#### I.3.1 Notion de risque instantané

Nous nous sommes, pour l'instant, intéressés à la fonction de survie,  $S(t)$ . Cette fonction intègre l'ensemble des événements observés avant  $t$  et décrit mal la dynamique instantanée du processus.

La dynamique d'un processus peut s'exprimer sous la forme d'une fonction de risque instantané, traduisant le risque de l'occurrence d'un événement sur un intervalle de temps infinitésimal, conditionnellement au fait que cet événement ne s'est pas produit auparavant. Cette fonction de risque peut être paramétrable (exprimée sous forme d'une formule mathématique). C'est le cas du modèle exponentiel (qui suppose que le risque instantané est constant au cours du temps) et du modèle de Weibull.

On peut aussi exprimer cette fonction de manière non paramétrique (sans faire d'hypothèse sur son allure au cours du temps). Le plus souvent, dans ce cas, on estime la fonction de risque instantané  $h(t)$  par un estimateur de Kaplan-Meier :

pour chaque temps  $t_i$ , la proportion d'événements observés est  $h(t_i) = m_i/n_i$  où  $m_i$  est le nombre d'événements observés en  $t_i$  et  $n_i$  est le nombre de sujets exposés au risque d'événements juste avant  $t_i$ .

#### I.3.2 Principes généraux et définitions

Le modèle de Cox s'applique à toute situation où l'on étudie le délai de l'occurrence d'un événement  $E$ . Cet événement peut représenter la récurrence d'une maladie, la réponse à un traitement, etc. Les sujets entrent dans l'étude au fur et à mesure qu'elle se déroule. Pour chaque sujet, on connaît la date du début de l'observation (date d'origine), la date des

dernières nouvelles et l'état par rapport à l'événement étudié. Les sujets pour lesquels on ne connaît pas l'état à la date de fin de l'étude ou ceux dont la réponse par rapport à l'occurrence de l'événement est négative constituent des données censurées.

Les  $p$  valeurs des variables explicatives  $Z_j$  ( $j = 1, \dots, p$ ) sont notées, pour chaque sujet, à la date de son entrée dans l'étude. Ces variables peuvent être qualitatives ou quantitatives. La variable considérée  $T$  est le délai entre la date d'origine ou de départ et la date à laquelle l'événement étudié se produit. Le modèle de Cox permet d'exprimer le risque instantané de l'occurrence de l'événement en fonction de l'instant  $t$  et des variables explicatives  $Z_j$ . Ces variables peuvent, en effet, représenter des facteurs de risque, des facteurs pronostiques, des traitements, des caractéristiques intrinsèques au sujet, etc.

Le risque instantané  $h(t, Z_1, Z_2, \dots, Z_p)$  de l'occurrence de l'événement  $E$  représente la probabilité d'apparition de l'événement dans un intervalle de temps  $(t, t + \Delta t)$ , sachant que l'événement ne s'est pas réalisé avant l'instant  $t$ . Le modèle de Cox exprime ainsi  $h(t, Z_1, Z_2, \dots, Z_p)$  sous la forme :

$$h(t) = h(t, Z_1, Z_2, \dots, Z_p) = h_0(t) \exp \left[ \sum_{j=1}^p \beta_j Z_j \right] \quad (38)$$

Notons bien que cette formule appelle quelques commentaires :

- Le risque instantané se décompose en produit de 2 termes, dont l'un dépend du temps  $t$  et l'autre des variables  $Z_j$ .
- Si, par exemple, les variables  $Z_j$  représentent des facteurs de risque et si elles sont toutes égales à 0,  $h_0(t)$  est le risque instantané de sujets ne présentant aucun facteur de risque.
- La forme de  $h_0(t)$  n'étant pas précisée, c'est plutôt l'association entre les variables  $Z_j$  et l'apparition de l'événement  $E$  considéré qui constituent l'intérêt central du modèle. Cela revient à déterminer les coefficients  $\beta_j$ .
- Le rapport des risques instantanés de 2 individus, dont les caractéristiques respectives sont  $(Z_1, Z_2, \dots, Z_p)$  et  $(Z'_1, Z'_2, \dots, Z'_p)$  est donné par :

$$\frac{h(Z_1, Z_2, \dots, Z_p)}{h(Z'_1, Z'_2, \dots, Z'_p)} = \frac{\sum_{j=1}^p \beta_j Z_j}{\sum_{j=1}^p \beta_j Z'_j}$$

Remarquons que ce rapport ne dépend pas du temps. De plus, de tels modèles sont dits à risques proportionnels. C'est une hypothèse importante du modèle de Cox.

### I.3.3 Interprétation des coefficients

Le modèle de Cox ne fait aucune hypothèse concernant la distribution du risque de référence. On parle de risque proportionnel, dans la mesure où le ratio des taux de risque entre deux individus est constant et ne dépendant donc plus du temps.

Si la variable  $Z$  est dichotomique i.e., si  $Z$  prend les valeurs 0 ou 1 selon l'absence ou la présence de la caractéristique considérée. C'est le cas par exemple, d'être au chômage ou ne pas être au chômage. Le rapport des risques instantanés des sujets de la classe 1 par rapport à la classe 0 est :

$$\frac{h(t, 1)}{h(t, 0)} = e^\beta$$



On en déduit que le coefficient  $\beta$  est donc le logarithme du risque instantané relatif de la classe 1 par rapport à la classe 0.

De façon générale, les coefficients  $\beta_j$  représentent l'effet de la caractéristique  $Z_j$  sur l'apparition de l'événement  $E$ . Si  $\beta_j$  est nul, la  $j$ -ième caractéristique n'a pas d'influence sur l'événement considéré. Si  $\beta_j$  est positif et si 2 sujets ne diffèrent que par la  $j$ -ième caractéristique, alors des valeurs élevées de la  $j$ -ième caractéristique sont associées à un risque instantané plus élevé. Inversement, si  $\beta_j$  est négatif, alors des valeurs élevées de la  $j$ -ième caractéristique sont associées à un risque instantané plus faible.

### I.3.4 Modèle de fragilité

Les modèles à effets aléatoires sont importants dans les études biologiques et épidémiologiques, car ils présentent une méthode d'hétérogénéité de modélisation.

Un modèle à effet aléatoire qui est particulièrement adapté à l'analyse de survie multivariée est le modèle de fragilité. Pour décrire ce modèle, nous définissons la fonction de survie comme suit :

$$S(t) = \mathbb{P}(T > t) = 1 - F(t)$$

où  $F$  est la fonction de distribution de  $T$ .

La relation entre la fonction de risque instantané de Cox  $h(t)$  et la fonction de survie  $S(t)$  est donnée par :

$$h(t) = -\frac{\partial \ln S(t)}{\partial t} = \frac{f(t)}{S(t)}$$

où  $f(t)$  est la densité de probabilité de  $T$ .

Dans l'équation (38), posons  $\gamma = \exp \left[ \sum_{j=1}^p \beta_j Z_j \right]$ . Alors, on peut exprimer la fonction de survie par rapport à  $\gamma$  comme suit :

$$S(t|\gamma) = \exp \left[ - \int_0^t h(s) ds \right] = [B(t)]^\gamma \quad (39)$$

où  $B(t) = \exp \left[ - \int_0^t h_0(s) ds \right]$  est la fonction de survie correspondant au risque de base.

Les modèles de fragilité se posent lorsque  $Z_j$ , et donc  $\gamma$ , est inobservable. Le paramètre  $\gamma$  s'appelle le facteur de fragilité, et quand  $\gamma$  augmente, la fonction de survie diminue. Donc, on peut exprimer la fonction de survie moyenne par :

$$S(t) = \mathbb{E}_\gamma [S(t|\gamma)] \quad (40)$$

Pour les modèles de fragilité multivariés, si nous supposons que les durées de vie  $T_1, T_2, \dots, T_p$  sont indépendantes par rapport à  $\gamma$ , nous aurons :

$$\begin{aligned} \mathbb{P}[T_1 > t_1, \dots, T_p > t_p | \gamma] &= \mathbb{P}[T_1 > t_1 | \gamma] \dots \mathbb{P}[T_p > t_p | \gamma] \\ &= S_1(t_1 | \gamma) \dots S_p(t_p | \gamma) \\ &= [B_1(t_1)]^\gamma \dots [B_p(t_p)]^\gamma \end{aligned} \quad (41)$$

D'où la fonction de survie multivariée est :

$$\mathbb{P}[T_1 > t_1, \dots, T_p > t_p] = \mathbb{E}_\gamma [B_1(t_1) \dots B_p(t_p)]^\gamma \quad (42)$$

## I.4 Modèles épidémiologiques

Notons d'abord que les premières contributions à l'épidémiologie mathématique moderne sont dues à D. Bernoulli (cf. [21, 22]) sur les techniques de variolisation, à W. Farr et P.D. Enk'o entre 1760 et 1894. En fait, Bernoulli a considéré un modèle de la variole et a calculé le gain en terme de vie moyenne si la variolisation est appliquée.

Il faut dire aussi que les fondements de l'épidémiologie basée sur les modèles compartimentaux ont été établis par des médecins de santé publique comme Sir Ronald Ross, W. H Hamer, A.G. McKendrick et W.O. Kermack, entre les années 1900 et 1935 (cf. [23, 24, 25, 26, 27]). En fait, R. Ross peut-être considéré comme le fondateur des modèles compartimentaux. Il a reçu le prix Nobel en 1902, pour avoir apporté la preuve que le paludisme était transmis par les anophèles. Bien que son travail ait eu une reconnaissance immédiate dans la communauté médicale, ses conclusions sur le fait que le paludisme pouvait être contrôlé par une lutte contre les moustiques était très contestée (cf. [57]). G. Sallet (2010) disait : "Ses détracteurs considéraient qu'il était impossible de débarrasser une région de toute sa population anophélienne et que, dans tous les cas, les moustiques envahiraient de nouveau la région. C'est une des raisons qui a conduit R. Ross à construire et à étudier son modèle. Celui-ci estimait qu'il n'était pas nécessaire d'éliminer tous les anophèles pour éradiquer le paludisme, mais simplement de faire baisser cette population en dessous d'un certain seuil critique. Cette affirmation laissait sceptiques les paludologues de l'époque. C'est ainsi que R. Ross a bâti le premier modèle de transmission du paludisme en 1911, afin de renforcer ses arguments. Ce modèle est très simple, mais il contient tous les ingrédients de la modélisation".

Dans ce qui suit, nous allons discuter trois principaux modèles épidémiologiques : le modèle de Ross (cf. [57]), le modèle de Macdonald (cf. [57]) et le modèle de Ross-Mckendrick (cf. [58]). Les notations et les analyses du modèle de Ross et du modèle de Macdonald sont inspirées des travaux de G. Sallet 2010 (cf. [57]).



FIGURE I.6 – Un anophèle femelle (cf. Sallet, 2010)

### I.4.1 Modèle de Ross

#### I.4.1.1 Le contexte biologique

Le paludisme est transmis par les femelles d'une espèce particulière d'insectes : les anophèles. Les moustiques sont des insectes. Seules les femelles de certaines espèces sont hémato-phages (elles ont besoin de repas de sang). On peut distinguer dans les vecteurs les *anophèles*, les *culex* et les *aedes*. Seule une partie des anophèles peut transmettre le paludisme.

En fait, seule la femelle pique l'homme, généralement après le coucher du soleil. Un repas sanguin lui est nécessaire, avant de pondre ses œufs isolément sur des surfaces liquides. Les œufs donnent naissance à des larves, puis à des nymphes et enfin à l'insecte ailé. Rappelons que le vol du moustique ne dépasse pas, en principe, un ou deux kilomètres.

Il existe plus de 300 espèces d'anophèles et seules 60 sont vectrices des plasmodiums humains. Notons aussi qu'en Afrique, les trois principaux complexes d'espèces vectrices sont :

- A. Funestus
- A. Gambiae
- A. Arabiensis

Ce sont les vecteurs les plus redoutables. Tous ces moustiques vivent presque exclusivement en proximité de l'homme (anthropophilie). Particulièrement, la meilleure illustration de la puissance de *A. gambiae* comme vecteur est donnée par son introduction accidentelle par les bateaux, au Brésil, en 1931. Il y a suscité en 1938 la pire épidémie, jamais enregistrée, en causant 14000 morts en 8 mois. L'éradication de *A. gambiae* au Brésil a permis de réduire l'impact du paludisme, non seulement au Brésil, mais aussi aux Amériques. Le même scénario s'est produit en haute Egypte en 1941.

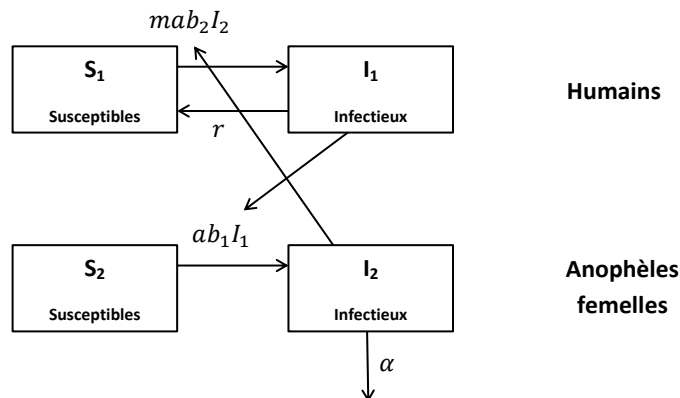


FIGURE I.7 – Le modèle compartimental de Ross

#### I.4.1.2 Modélisation

Dans le modèle compartimental de Ross (cf. Figure I.7), les flèches horizontales montrent les mouvements entre compartiments, avec comme coefficient la vitesse de transfert. Les flèches obliques indiquent l'effet d'un compartiment sur ce coefficient. C'est ainsi que le nombre des anophèles femelles conditionne le passage des humains susceptibles vers les humains infectieux (flèche oblique montante), et que le nombre des humains infectieux agit sur la vitesse de transfert des anophèles femelles sains vers les anophèles infectieux (flèche oblique descendante).

Pour expliquer le procédé de modélisation qui a guidé Ross, à l'époque, nous allons détailler ce modèle, qui peut être considéré comme pédagogiquement très intéressant. On verra aussi que ce modèle est très simple, et qu'il permet de formuler des conclusions non évidentes (surtout en 1911).

Rappelons que l'objectif de R. Ross était d'établir, dans la communauté épidémiologique (ce qui lui paraissait évident mais était alors un objet d'incrédulité), qu'il n'était pas nécessaire d'éradiquer totalement la population anophélienne pour éliminer le paludisme, mais que le fait de la faire baisser en dessous d'un certain seuil était suffisant. Il a donc introduit ce modèle en 1911 (cf. [23]). Il a adopté l'hypothèse implicite de l'absence de surinfection. En d'autres termes, un individu infecté ne peut plus déclencher une deuxième infection, i.e., il n'a que deux possibilités : rester infecté ou guérir. Dans le cas du paludisme, il n'y a pas d'immunité innée ou acquise permanente. Dans les pays de forte endémie, les personnes très souvent infectées finissent par s'immuniser ("immunité acquise"), devenant porteurs asymptomatiques, mais après une période sans contact avec le parasite, cette immunité disparaît. Le modèle est un modèle dit "compartimental", qui est devenu classique en biologie (cf. [23, 24]). On distingue dans les deux populations, deux compartiments (virtuels) : les infectieux (qui transmettent) et les susceptibles (qui peuvent être infectés). Les deux populations totales (moustiques et hommes) sont supposées constantes. Plus exactement, on suppose que, sur l'intervalle de temps considéré, cette hypothèse est acceptable. Par ailleurs, l'accroissement démographique, i.e., la différence entre les naissances et les décès, est considérée comme négligeable sur la période étudiée. De même, on estime que le moustique n'a pas le temps de guérir et reste infectieux le reste de sa vie (cette hypothèse est raisonnable, d'après les données entomologiques).

Un humain susceptible quittera le compartiment susceptible, quand il sera piqué par un moustique infectieux et développera une infection. Il quitte le compartiment infectieux, quand il guérit, à la vitesse  $r$ , autrement dit cela signifie que la durée de séjour dans le compartiment infectieux est exponentiellement distribuée avec un temps de séjour moyen de  $1/r$ . Donc, le taux de survie, dans le compartiment, au jour  $n$  est  $e^{-nr}$ . Avec cette hypothèse, si l'on estime à 6 mois le temps de guérison (n'être plus transmetteur), comparé à une espérance de vie, par exemple 50 années, on voit que le principe de négliger la mortalité revient à négliger (unité de temps la journée) :

$$\frac{1}{(360 \times 50)^{j-1}} \text{ devant } \frac{1}{180j-1} \text{ (valeur 100 fois plus grande)}$$

Si on désigne par  $a$  le nombre moyen de piqûres sur les humains par unité de temps (le biting rate of Macdonald),  $b_2$  la probabilité qu'une piqûre infectée donne une infection, alors s'il y a  $I_2$  moustiques infectés, un humain recevra, pendant le temps  $\Delta t$  :

$$b_2 a \frac{I_2}{N_H} \Delta t$$

piqûres donnant une infection.

L'ensemble des moustiques donnent  $a I_2 \Delta t$  piqûres, il faut donc diviser par la taille de la population humaine  $N_H$ , afin d'obtenir le nombre de piqûres par humain.

Pour obtenir une nouvelle infection, seuls les susceptibles sont concernés (pas de surinfection). Donc, s'il y a  $I_1(t)$  infectieux, alors il y aura  $(N_H - I_1(t))$  susceptibles. En effet, comme la population est supposée constante, il suffit de connaître l'effectif d'un seul compartiment.

On fait le bilan de “masse”, c’est-à-dire que l’on cherche à exprimer le nombre  $I_1(t + \Delta t)$  d’humains au temps  $t + \Delta t$  dans le compartiment des infectieux. Pour cela, on ajoute ceux qui rentrent par contamination et on retranche ceux qui sortent par guérison à la vitesse  $r$  :

$$I_1(t + \Delta t) = I_1(t) + b_2 a \frac{I_2}{N_H} \Delta t (N_H - I_1) - r I_1(t) \Delta t$$

En divisant par  $\Delta t$ , puis en faisant tendre  $\Delta t$  vers zéro, nous obtenons la fonction  $I_1(t)$  qui est régie par l’équation différentielle :

$$\frac{d}{dt} I_1(t) = b_2 a \frac{I_2(t)}{N_H} (N_H - I_1(t)) - r I_1(t)$$

Un raisonnement analogue sur la population vectorielle infectieuse  $I_2(t)$  donne :

$$\frac{d}{dt} I_2(t) = b_1 a \frac{I_1(t)}{N_H} (N_V - I_2(t)) - \mu I_2(t)$$

Soient  $x = \frac{I_1}{N_H}$  et  $y = \frac{I_2}{N_V}$  et posons  $m = \frac{N_V}{N_H}$ , où la densité  $m$  est la densité vectorielle.

En posant  $\dot{x} = \frac{dx}{dt}$  et  $\dot{y} = \frac{dy}{dt}$ , nous obtenons :

$$\begin{cases} \dot{x} = mab_2y(1-x) - rx \\ \dot{y} = ab_1x(1-y) - \mu y \end{cases} \quad (43)$$

Le système (43) est le fameux modèle de R. Ross. L’étude mathématique en a été faite par Ross (cf. [23, 24]) et par Lotka (cf. [30]). Nous donnons rapidement, dans ce qui suit, le cheminement mathématique et les conclusions qui en découlent.

Notons que cette équation peut s’écrire matriciellement sous la forme :

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} -r & mab_2(1-x) \\ ab_1(1-y) & -\mu \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad (44)$$

Chaque terme de la matrice  $2 \times 2$  dans le système (44) est majoré par le terme correspondant dans la matrice suivante :

$$J = \begin{pmatrix} -r & mab_2 \\ ab_1 & -\mu \end{pmatrix} \quad (45)$$

Ceci vient du fait que les quantités  $(1-x)$  et  $(1-y)$  sont des proportions, et sont donc positives. Cette matrice n’est autre que la matrice Jacobienne du système d’origine (généralisation de la dérivée dans des espaces multidimensionnels). Des théories mathématiques récentes montrent facilement (techniques dites de Lyapounov-Lasalle) que le point d’équilibre  $(0, 0)$  i.e., l’équilibre sans maladie, est globalement asymptotiquement stable si, et seulement si, le déterminant de la matrice  $J$  est positif ou nul. Soit :

$$\mu r - ma^2 b_1 b_2 > 0 \quad (46)$$

Dans le cas contraire, l’équilibre sans maladie est instable i.e., il existe un équilibre endémique  $(x_e, y_e)$  avec  $x_e > 0$  et  $y_e > 0$ , et cet équilibre endémique est globalement asymptotiquement stable sur le quadrant ouvert  $x > 0$  et  $y > 0$ .

En posant de façon équivalente :

$$R_0 = \frac{ma^2b_1b_2}{\mu r}$$

nous obtenons le théorème du seuil :

- Si  $R_0 \leq 1$ , alors l'épidémie disparaît
- Si  $R_0 > 1$ , alors dans ce cas la maladie est endémique

Si  $m$ , qui est le nombre de moustiques par humain, est suffisamment petit, alors  $R_0 \leq 1$  et le paludisme disparaît. C'est la prédiction de R. Ross.

Cette expression de  $R_0$  est appelée "le taux de reproduction de base". Ce taux  $R_0$  admet une interprétation biologique simple. Il s'agit du nombre moyen de cas secondaires produits par un individu infectieux pendant sa période d'infectiosité, quand il est introduit dans une population constituée totalement d'individus susceptibles. Ceci peut être expliqué facilement. En effet, pour un humain infectieux, la durée d'infectiosité est  $1/r$ , il reçoit  $m \times a$  piqûres par unité de temps, et de celle-ci une proportion  $b_1$  contamine les moustiques. Cet humain transmet donc l'infection à :

$$mab_1 \frac{1}{r}$$

moustiques. Ces moustiques vivront pendant une durée de  $1/\mu$  unités de temps, et piqueront "a" humains par unité de temps. Parmi ces piqûres, une proportion  $b_2$  de la population humaine sera infectante. En tout, la taille de la population humaine des "infectieux secondaires" à partir du premier cas, est égale à :

$$mab_1 \frac{1}{r} ab_2 \frac{1}{\mu}$$

C'est bien égal à  $R_0$ . On comprend donc, de façon intuitive, que si  $R_0 > 1$ , alors la maladie peut s'établir.

Ceci peut s'écrire autrement : comme la durée de survie moyenne d'un moustique est  $1/\mu$ , la probabilité de survie au temps  $t$  est  $e^{-\mu t}$ . Si on note  $p$  la probabilité de survie d'un moustique pour un jour, on peut l'exprimer à l'aide du logarithme népérien :  $\mu = -\ln(p)$ .

R. Ross propose, au lieu d'un système de deux équations différentielles, seulement une seule équation qui décrit l'évolution, au cours du temps, de la proportion "x" des humains infectieux. On peut donc prouver par les techniques mathématiques, dites de perturbations singulières, que cette équation se déduit du système (43).

L'idée est de considérer que l'équation qui gouverne la prévalence chez les moustiques évolue beaucoup plus rapidement que la prévalence chez les humains. Autrement dit, "y" décroît beaucoup plus vite que "x". Géométriquement, cela signifie que, dans le plan de coordonnées "xOy", les trajectoires se précipitent vers la courbe du plan où la dérivée de "y" s'annule. Sur cette courbe, "x" évolue plus lentement. On dit donc que la "courbe est lente". Pour trouver l'équation de cette courbe on résout :

$$\dot{y} = ab_1x(1-y) - \mu y = 0 \tag{47}$$

Ce qui donne, pour l'équation de cette courbe dans le plan xOy :

$$y = \frac{\frac{ab_1x}{\mu}}{1 + \frac{ab_1x}{\mu}}$$

Si on a :  $\frac{ab_1}{\mu} \ll 1$ , alors :

$$y = \frac{\frac{ab_1}{\mu}x}{1 + \frac{ab_1}{\mu}x} \approx \frac{ab_1}{\mu}x$$

Les trajectoires atteignent rapidement cette courbe, où la dynamique est plus lente. La dynamique s'obtient en remplaçant "y" par cette valeur, dans la première équation du modèle de R. Ross. Aussi, la dynamique de "x" est donnée par :

$$\dot{x} = Cx(1-x) - rx \quad (48)$$

où

$$C = \frac{ma^2b_1b_2}{\mu}$$

Ce terme désigne la capacité vectorielle définie par Garret-Jones (cf. [31]) avec  $b_1b_2 = 1$ .

Cette deuxième équation de R. Ross c'est-à-dire l'équation (48) exprime que le taux d'infection vecteur-humain peut être exprimé comme le produit de la proportion d'infectés dans la population humaine et de la capacité vectorielle  $C$  des anophèles (cf. [32]).

Si l'on ré-écrit cette équation sous la forme :

$$\dot{x} = Cx \left( 1 - x - \frac{r}{C} \right)$$

on s'aperçoit qu'il s'agit d'une équation logistique, dont toutes les trajectoires convergent vers l'équilibre  $x_e$  :

$$x_e = 1 - \frac{r}{C}$$

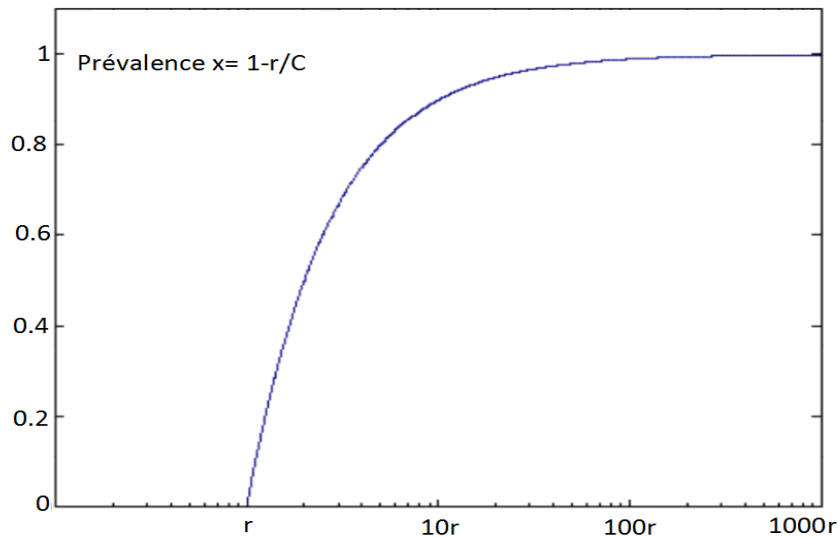


FIGURE I.8 – Prévalence en fonction de la capacité vectorielle.

Pour que cet équilibre ait un sens, il faut évidemment que  $x_e > 0$ , soit  $C > r$ . Il faut dire que, très vite, la population est proche de l'équilibre. Si l'on trace la prévalence à l'équilibre en fonction de  $C$  (cf. Figure I.8), on voit que cette simple relation implique (cf. [57]) :

- qu'il existe un niveau de capacité vectorielle non nul  $r$ , de capacité vectorielle, en dessous duquel le paludisme ne pourrait se maintenir.
- que, plus la durée de l'infection est courte (durée moyenne  $1/r$ ), plus cette capacité vectorielle critique est élevée (cela pourrait expliquer l'endémicité de *P. vivax* dans des zones où *P. falciparum* n'est pas endémique).
- la relation est fortement non linéaire : proche de son niveau critique, une légère augmentation de la capacité vectorielle s'accompagne d'une augmentation importante du taux de prévalence des infectés, qui est initialement rapide, puis plus lent (relation non linéaire), jusqu'à ce qu'il atteigne un plateau proche de 100%, où même de grands changements de la capacité vectorielle produisent peu de changements.

Ce modèle peut paraître trivial, mais, malgré sa simplicité caricaturale, il permet des prédictions réalistes. C'est ainsi que ces conclusions, proposées par R. Ross, à partir de ce modèle, sont confirmées par des paludologues avertis (cf. [33]).

### I.4.2 Modèle de Macdonald

McKendrick (1926), Kostitzin (1934) puis Macdonald (cf. [34]) firent l'hypothèse qu'une infection n'empêchait pas la surinfection et que plusieurs souches plasmodiales pouvaient infecter ensemble, de façon indépendante les unes des autres, un même individu. Le modèle développé par Macdonald à partir du modèle de Ross est présenté dans la Figure I.9.

Ce modèle repose sur les principales hypothèses suivantes (cf. [57]) :

- Les populations des humains et des vecteurs sont homogènes (pas de différence d'immunité acquise, de prise d'antipaludéens, ni d'attractivité des moustiques entre individus ; pas de différence de susceptibilité entre anophèles) et restent constantes.
- Les anophèles piquent l'homme au hasard (n'importe quel anophèle pique n'importe quel individu) et la probabilité de repas sur l'homme est identique quel que soit le repas et quel que soit l'anophèle.
- Un anophèle infecté devient infectant  $n$  jours après être infecté.
- Le taux de mortalité  $\mu = -\ln(p)$  des anophèles est constant quels que soient leurs âges et leurs infections, le taux d'émergence est constant, la population totale est à l'équilibre.
- Un humain infecté est infectant sans délai.
- Il n'y a pas de migration.

Comme pour le modèle de Ross, on peut calculer le taux de reproduction de base  $R_0$  du modèle de Ross-Macdonald, en évaluant le nombre d'infections secondaires à partir d'une seule personne infectée.

Macdonald propose :

$$R_0 = \frac{ma^2b_1b_2 e^{-\mu n}}{\mu r} \quad (49)$$

ou encore, comme  $\mu = -\ln(p)$  :

$$R_0 = \frac{ma^2b_1b_2 e^{-\mu n}}{-\ln(p)r} \quad (50)$$



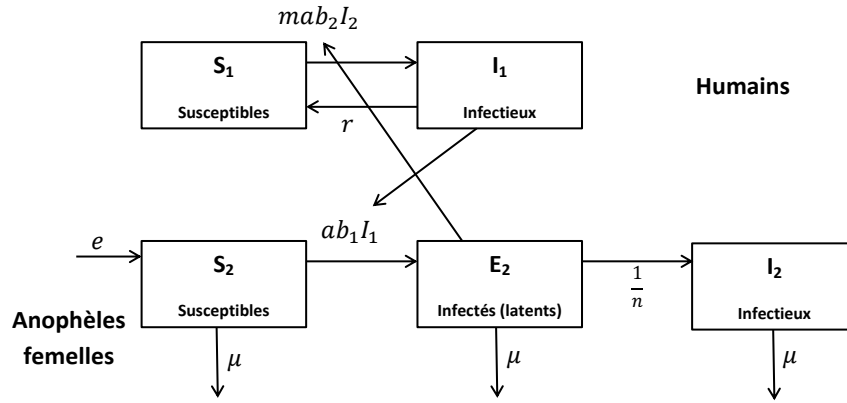


FIGURE I.9 – Modèle de Ross-Macdonald

La différence avec le modèle de “Ross” est la multiplication par le facteur  $p^n = e^{-\mu n}$  qui est tout simplement le taux de survie d’un anophèle pendant le cycle sporogonique de durée  $n$  jours.

Dans ce formalisme, la capacité vectorielle devient :

$$C = \frac{ma^2 e^{-\mu n}}{\mu} = \frac{ma^2 p^n}{-\ln(p)}$$

Rappelons que l’endémie palustre ne peut se maintenir que si  $R_0 > 1$ . De la formule (50), il est aisé de déduire que le niveau d’endémie dépend plus du taux quotidien de survie  $p$  et de la durée  $n$  du cycle sporogonique ( $n$  et  $\mu$  apparaissant dans une exponentielle) que de la proportion de repas sanguins faits sur l’homme (terme en  $a^2$ ). En revanche, il est plus efficace, pour faire baisser  $R_0$ , de faire baisser ce dernier terme  $a^2$  que de faire baisser la densité vectorielle  $m$  ou augmenter le taux de guérison  $r$ .

On peut donc en déduire que la lutte par insecticides dirigée contre les vecteurs adultes a potentiellement plus d’impact que par des larvicides. Le taux  $R_0$  peut aussi être exprimé en fonction de la capacité vectorielle (cf. [31, 36]) par :

$$R_0 = \frac{Cb_1 b_2}{r}$$

On retrouve ainsi, au coefficient  $b_1 b_2$  près, le concept de niveau critique de capacité vectorielle en dessous duquel le paludisme ne pourrait plus se maintenir.

Notons que la dérivation de  $R_0$  est ici heuristique. Si l’on considérait le modèle compartimental associé à la Figure I.9, une même analyse que dans le modèle de Ross conduirait à un système de 3 équations différentielles, avec pour  $R_0$  l’expression :

$$R_0 = \frac{ma^2 b_1 b_2}{\mu r} \frac{\rho}{\mu + \rho} \quad (51)$$

où  $r$  est la vitesse à laquelle les moustiques quittent le compartiment latent  $E_2$  et  $n = 1/r$ . Remarquons qu’il y a là une contradiction apparente. Mais, si l’on remarque qu’au premier ordre :

$$e^{-\mu n} = 1 - \mu n = \frac{1}{1 + \mu n} = \frac{\rho}{\mu + \rho}$$

on s’aperçoit qu’il s’agit bien de deux approximations équivalentes.

Par exemple, au Congo, avec comme unité de temps la journée, à température de 27 degrés

celsius,  $n = 10$  jours et  $\mu = 1/28j^{-1}$  pour An. Gambiae, nous obtenons :

$$e^{-\mu n} = 0.700 \quad \text{et} \quad \frac{\rho}{\mu + \rho} = 0.737$$

À partir de ces résultats, nous remarquons que la formule de Macdonald diminue la valeur du taux  $R_0$ .

Comme pour le modèle de “Ross”, il faut dire que ce modèle et sa formulation mathématique sont aisément compréhensibles. Cela a fait son succès au point d’en masquer les limites (cf. [37]). Tenant compte de l’hypothèse de surinfections chez l’homme, il serait logique de considérer que le risque d’infection de l’anophèle par l’homme est proportionnel au nombre d’infections portées par l’hôte humain.

“Dans son modèle, Macdonald fait cependant l’hypothèse implicite qu’il existerait un contrôle drastique de l’infectivité de l’hôte humain (quel que soit son niveau d’infection, son infectivité serait constante) et donc que la possibilité ou non de surinfection n’a aucune importance. Cette limite dans la formulation mathématique du modèle est à l’origine d’une grande partie de l’écart observé entre ses prédictions et les données (cf. [38]), particulièrement entre le taux de reproduction de base  $R_0$  calculé et le niveau d’endémie”, d’après G. Sallet (2010).

Une autre limite de cette famille de modèles vient de l’absence de prise en compte :

1. de l’hétérogénéité des populations humaines et vectorielles
2. de l’acquisition d’une immunité anti-plasmodiale avec l’âge

L’effet de ne pas prendre en compte l’hétérogénéité entraîne généralement une surévaluation grossière du seuil critique en dessous duquel l’endémie ne peut plus se maintenir. Cela pousse à tirer des conclusions exagérément optimistes sur les taux de couverture et les niveaux d’efficacité nécessaires pour que les mesures de lutte permettent l’extinction de l’endémie. L’hypothèse d’homogénéité du contact vecteur-homme entraîne une surestimation du taux de prévalence des infections dans les zones de forte endémie (où le taux de reproduction de base  $R_0$  est élevé) et une sous-estimation dans les zones où  $R_0$  est bas (cf. [39]).

### I.4.3 Modèle de Ross-McKendrick

Après l’introduction d’un modèle historique par Bernoulli (cf. [21, 22]) qui a été proposé afin d’expliquer la dynamique de la variole, de nombreuses discussions sont survenues sur l’efficacité de l’inoculation, d’une part, et sur la vaccination, d’autre part (cf. [48]). Dans le modèle princeps (cf. [55]), la population est divisée en une sous-population d’individus susceptibles (qui n’ont pas encore été infectés) et une autre sous-population d’individus immuns (immunisés pour le reste de leur vie après une infection), et les deux variables correspondantes sont  $u(a)$  (la probabilité pour un individu d’être susceptible et encore vivant à l’âge  $a$ ), et  $w(a)$  (la probabilité d’être immunisé et encore vivant à l’âge  $a$ ).

R. Ross (cf. [24]), McKendrick (cf. [27]), Kermack et McKendrick (cf. [25, 26]) ont proposé le modèle dit Susceptibles-Infectieux-Récupérés immunisés, ou modèle SIR, qui est représenté

par le système d'équations différentielles suivant :

$$\begin{aligned}\frac{dS}{dt} &= \delta S + \delta I + (\delta + \gamma)R - \beta SI - \delta S \\ \frac{dI}{dt} &= \beta SI - (\delta + \nu)I \\ \frac{dR}{dt} &= \nu I - (\delta + \gamma)R\end{aligned}$$

où  $S$  (resp.  $I$ ,  $R$ ) désigne la taille de la sous-population des individus Susceptibles (resp. Infectieux, Récupérés) avec  $S + I + R = N$ ;  $\beta$  (resp.  $\delta$ ,  $\gamma$ ,  $\nu$ ) est le taux de contagion (resp. décès/natalité, perte de résistance, immunisation).

Le taux de reproduction de base  $R_0 = \beta N / (\nu + \delta)$  est le nombre moyen de cas secondaires produits par chaque cas primaire dans une population totalement susceptible; s'il est plus grand que 1, on aura une apparition d'une vague épidémique.

Notons que les faiblesses de ce modèle réside dans les insuffisances suivantes :

- Lorsque la taille de la sous-population des individus susceptibles ou infectieux a tendance à être très grande, le terme quadratique  $SI$  doit être remplacé par un terme michaelian :

$$\frac{SI}{(K + S)(K' + I)}$$

- Les individus infectieux et immunisés ou les porteurs sains sont négligés
- La taille de la population totale est supposée constante (la fécondité est égale à la mortalité)
- Les variables et les paramètres ne dépendent pas de l'espace (pas de migration, ni de déplacements de la population)
- Les paramètres ne sont pas fonction du temps (pas d'adaptation génétique de l'agent infectieux ou de la population humaine, même très lente)

#### I.4.4 Stabilité des systèmes dynamiques

En général, pour un système d'équations différentielles, la notion de stabilité est essentielle. Rappelons ici que la stabilité est la capacité d'un système à se maintenir autour d'un point de fonctionnement.

Lorsqu'un système est représenté par des équations différentielles non linéaires, le problème de la vérification de la propriété de stabilité n'est pas trivial. Bien entendu, des outils théoriques, ainsi que des formulations mathématiques parfois relativement sophistiquées, sont à notre disposition pour étudier la stabilité de tels systèmes. En recherche mathématique, des propositions d'outils théoriques pour la stabilité sont apparues dans la littérature, avec, pour chacune d'elles, un aspect et des conditions d'application différentes. Parmi ces outils, la théorie de Lyapunov et ses célèbres fonctions dites de Lyapunov qui ont pris une importance considérable, quant à leur développement théorique et à leur utilisation pour résoudre des problèmes concrets; tels que la stabilité d'un système non linéaire.

Considérons une maladie modélisée par une équation différentielle de la forme :

$$dx_i/dt = f_i(x), \quad x_i(0) = c_{i0} \geq 0, \quad \text{avec } i = 1, \dots, n, \quad \text{où } x = (x_1, \dots, x_n)^t$$

Un point d'équilibre de ce système est une solution à l'état stationnaire,  $x_i \geq 0$ ,  $i = 1, \dots, n$ , c'est-à-dire solutions du système  $f_i(x) = 0$ .

**Définition I.14** (*Points d'équilibre endémique et DFE<sup>2</sup>*)

1. Un point d'équilibre endémique est une solution d'équilibre où la maladie persiste dans la population
2. Un point d'équilibre sans maladie, couramment appelé DFE, est un point d'équilibre où il n'y a pas de maladie dans la population

La stabilité locale de l'état d'équilibre sans maladie est gouvernée par le nombre de reproduction  $R_0$ . Le théorème suivant confirme ce résultat :

**Théorème I.9** *L'état d'équilibre sans maladie est localement et asymptotiquement stable lorsque  $R_0 < 1$ , tandis qu'il est instable pour  $R_0 > 1$ .*

Ce théorème montre que l'introduction d'un petit nombre d'individus sans maladie proche de l'état sans maladie ne conduit pas au développement d'une épidémie pour  $R_0 < 1$ .

La DFE est localement et asymptotiquement stable, si toutes les valeurs propres de la matrice Jacobienne du système ont chacune une partie réelle négative et instable si la matrice Jacobienne du système a au moins une valeur propre possédant une partie réelle positive.

**Proposition I.10** *Si  $R_0 > 1$ , alors il existe un point d'équilibre endémique.*

Pour chacun des modèles étudiés, il existe une condition pour la stabilité asymptotique et globale de l'équilibre endémique lorsque  $R_0 > 1$ .

#### I.4.4.1 Méthode de Lyapunov

**Définition I.15** (*Fonction de Lyapunov*)

On appelle fonction de Lyapunov en  $x_0$ , point d'équilibre du système  $\dot{x} = f(x)$ , une fonction  $V : \Omega \rightarrow \mathbb{R}$  telle que :

1.  $V(x_0) \geq 0$
2.  $V(x) = 0$  si, et seulement si,  $x = x_0$
3. il existe un voisinage de  $x_0$  tel que, sur ce voisinage, on ait :

$$\dot{V}(x) = \langle \nabla V(x) | f(x) \rangle \leq 0$$

Une fonction qui vérifie les deux premières propriétés est dite définie positive en  $x_0$ .

**Théorème I.11** (*de Lyapunov*)

Si  $x_0$  est un point d'équilibre du système  $\dot{x} = f(x)$ , et s'il existe une fonction de Lyapunov en  $x_0$  pour ce système, alors  $x_0$  est un point d'équilibre stable.

Si en outre,  $\dot{V}$  est définie négative, i.e., si  $\dot{V}(x) = 0$  si, et seulement si,  $x = x_0$ , alors  $x_0$  est un point d'équilibre asymptotiquement stable.

Le Théorème I.11 de Lyapunov nous dit que, pour montrer qu'un point d'équilibre est stable, il suffit de trouver une fonction de Lyapunov en ce point. Pour utiliser ce théorème, original, de Lyapunov dans le but de prouver qu'un point d'équilibre est asymptotiquement stable, nous devons trouver une fonction de Lyapunov stricte. Cette opération n'est en général pas chose facile. La condition sur la dérivée peut être allégée, en employant le principe d'invariance de LaSalle (cf. Théorème I.12).

---

2. Disease-Free Equilibrium - Equilibre sans maladie.

**Théorème I.12** (de LaSalle)

Si  $V$  est une fonction de Lyapunov propre, alors le plus grand ensemble invariant contenu dans :

$$\ell = \{x | \dot{V}(x) = 0\}$$

est un ensemble attractif.

Par ailleurs, si  $\ell = \{x_0\}$  alors  $x_0$  est globalement asymptotiquement stable.

Ce théorème a la particularité de ne pas exiger que  $\dot{V}$  soit définie négative. Cependant, il fournit juste une information sur l'attractivité du système considéré au point d'équilibre  $x_0$  (cf. [56]).

## I.5 Conclusion

Ce chapitre introductif, qui est composé de trois grandes sections, nous a permis de mieux comprendre les notions élaborées par Sklar, Cox, Ross, Macdonald, McKendrick, etc. Nous avons débuté par la définition des fonctions copules et du Théorème de Sklar, avec quelques résultats fondamentaux les concernant, et par l'explication de la notion de mesure de dépendance (tau de Kendall et rho de Spearman), et ensuite nous avons défini les fonctions copules paramétriques (copule Normale, copule de Student et copule archimédienne). En second lieu, nous avons expliqué le modèle de Cox, en introduisant la notion de risque instantané et de ses principes généraux, et aussi la définition du modèle de fragilité, qui est lié à la fonction de survie. Enfin, nous avons approfondi l'analyse de différents systèmes épidémiologiques : modèle de Ross, modèle de Macdonald et modèle de Ross-McKendrick, et nous avons discuté de la stabilité des systèmes dynamiques.

Dans les chapitres qui suivent, nous nous focaliserons sur la fonction copule de Gumbel, qui est une fonction copule archimédienne. Nous verrons ensuite les modèles épidémiologiques, avec l'introduction d'autres notions démographiques, comme l'âge, la fécondité, etc. Nous verrons plus en détail quelques modèles de classes d'âge, et nous proposerons pour ces modèles une approche avec les fonctions copules.

En utilisant le modèle de Cox et le modèle de fragilité, nous relierons la notion de fonctions copules archimédiennes aux modèles épidémiologiques.

Nous exposerons ensuite une étude par de simulations, qui nous permet de montrer la stabilité de certains modèles. Puis, avec les mesures de dépendance, nous calculerons les fonctions quantiles à partir des copules archimédiennes, et, à l'aide de ces fonctions quantiles, nous pourrions diviser la population étudiée en plusieurs sous-populations, ce qui nous mènera à des résultats plus précis.

# Bibliographie

- [1] A. Sklar, Fonctions de répartition à  $n$  dimensions et leurs marges, *Publications de l'Institut de Statistique de l'Université de Paris*, 8 (1959), pp. 229-231.
- [2] A. Sklar, Random variables, joint distribution functions and copulas, *Kybernetika*, 9 (1973), pp. 449-460.
- [3] A. Sklar, Random variables, distribution functions and copulas. A personal look backward and forward, in L. Ruschendorf, B. Schweizer et M.D. Taylor (Eds.), Distributions with Fixed Marginals and Related Topics, *Institute of Mathematical Statistics, Hayward, CA*, (1996).
- [4] S.Y. Zhang, L. Yu and P. Daszak, EcoHealth and the black death in the year of the rat, *Ecohealth*, 5 (2008), pp. 99-100.
- [5] P. Deheuvels, La fonction de dépendance empirique et ses propriétés Un test non paramétrique d'indépendance, *Académie Royale de Belgique, Bulletin de la Classe des Sciences 5me Série*, 65 (1979), pp. 274-292.
- [6] P. Deheuvels, A non parametric test for independence, *Publications de l'Institut de Statistique de l'Université de Paris*, 26 (1981), pp. 29-50.
- [7] L. Rüschendorf, On the distributional transform, Sklar's Theorem, and the empirical copula process, *Journal of Statistical Planning and Inference*, 139(11) (2009), pp. 3921-3927.
- [8] W. Hoeffding, Masstabinvariante Korrelationstheorie, *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin*, 5 (1940), pp. 181-233.
- [9] G. Dall'Aglio, Fréchet classes and compatibility of distribution functions, *Symposia Mathematica*, 9 (1972), pp. 131-150.
- [10] W. Kruskal, Ordinal measures of association, *Journal of the American Statistical Association*, 53 (1958), pp. 814-861.
- [11] C. Genest and J. MacKay, Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données, *American Statistician*, 40 (1986), pp. 280-283.

- [12] C. Genest and J. MacKay, The joy of copulas : Bivariate distributions with uniform marginals, *Canadian Journal of Statistics*, 14(2) (1986), pp. 145–159.
- [13] C. Genest and L-P. Rivest, A characterization of Gumbel’s family of extreme value distributions, *Statistical Probability Letters*, 8(3) (1989), pp. 207–211.
- [14] C. Genest and L-P. Rivest, Statistical inference procedures for bivariate Archimedean copulas, *Journal of the American Statistical Association*, 88 (1993), pp. 1034–1043.
- [15] C-H. Ling, Representation of associative functions, *Publicationes mathematicae Debrecen*, 12 (1965), pp. 189–212.
- [16] B. Schweizer and A. Sklar, Probabilistic Metric Spaces, Probability and Applied Mathematics, *North-Holland, New York*, (1983).
- [17] R.B. Nelsen, Copulas and Association. In : Advances in Probability Distributions With Given Marginals. Beyond the Copulas, *Kluwer, Dordrecht* (1991), pp. 51–74.
- [18] R.B. Nelsen, An Introduction to Copulas, Lectures Notes in Statistics, *Springer Verlag, New-York*, 139 (1999).
- [19] D.R. Cox, Regression Models and Life-Tables, *Journal of the Royal Statistical Society. Series B (Methodological)*, 34 (1972), pp.187–220.
- [20] C. Mason, Cox proportional hazard models, <http://www.demog.berkeley.edu/213/Week14/welcome.pdf>, (2005).
- [21] D. Bernoulli, Essai d’une nouvelle analyse de la mortalité causée par la petite vérole et des avantages de l’inoculation pour la prévenir, *Mem. Math. Phys. Acad. Roy. Sci., Paris*, (1760).
- [22] D. Bernoulli, Réflexions sur les avantages de l’inoculation, *Mercur de France*, Paris, (1760).
- [23] R. Ross, An Application of the Theory of Probabilities to the Study of a priori Pathometry. Part I, *Murray Ed., London*, (1911).
- [24] R. Ross, The prevention of malaria, *Proceedings of the Royal Society of London Series A*, 92 (1916), pp. 204–230.
- [25] W.O. Kermack and A.G. McKendrick, Contributions to the Mathematical Theory of Epidemics II. The Problem of Endemicity, *Proceedings of the Royal Society of London Series A*, 138 (1932), pp. 55-83.
- [26] W.O. Kermack and A.G. McKendrick, Contributions to the Mathematical Theory of Epidemics III. Further Studies of the Problem of Endemicity, *Proceedings of the Royal Society of London Series A*, 141 (1933), pp. 94–122.
- [27] A.G. McKendrick, Applications of mathematics to medical problems, *Proc. Edinburgh Mathematical Society*, 44 (1925), pp. 1–34.

- [28] P. Hougaard, Modelling Multivariate Survival, *Scand. J. Statist.*, 14 (1987), pp. 291–304.
- [29] P. Hougaard, A class of multivariate failure time distributions, *Biometrika*, 73 (1986), pp. 671–678.
- [30] A.K. Lotka, Contributions to the Analysis of Malaria Epidemiology, *Am. J. Hyg.*, 3 (1923), pp. 1–121.
- [31] C. Garret-Jones, The human blood index of malaria vectors in relation to epidemiological assessment, *Bull. World Health Organ*, 30 (1964), pp. 241–261.
- [32] L. Molineaux, The pros and cons of modelling malaria transmission, *Trans. R. Soc. Trop. Med. Hyg.*, 79 (1985), pp. 733–747.
- [33] B. Ruce and L.J. Chwatt, Mass drug administration for control of malarial, *ul. Lancet*, (1983).
- [34] G. Macdonald, The Epidemiology and Control of Malaria, *Oxford University Press, London*, (1957).
- [35] G. Macdonald, Dynamics of tropical diseases, *Oxford University Press, London*, (1973).
- [36] K. Dietz, Mathematical Models for Transmission and Control of Malaria. In Principles and Practice of Malariology, *Churchill Livingstone Ed., Edinburgh*, (1988).
- [37] M. Finepe, Superinfection - A problem in formulating a problem (an historical critique of Macdonald's theory), *Trop. Dis. Bull.*, 72 (1975), pp. 475–488.
- [38] J.A. Najera, A Critical Review of the field of applications of a mathematical model of malaria eradication, *Bull. World Health Organ*, 50 (1974), pp. 449–457.
- [39] C. Dye and G. Hasiberder, Population dynamics of mosquito-borne disease : effect of flies which bite some people more frequently than others, *Trans. R. Soc. Trop. Med. Hyg.*, 80 (1980), pp. 69–77.
- [40] T. Kulpa, Archimedean copulas and their application to the individual risk model, *In : Int. Congress Finance Modelling, Rizeni*, (2008).
- [41] D. Oakes, Bivariate Survival Models Induced by Frailties, *J. American Statistical Association*, 84 (1989), pp. 487–493.
- [42] D. Beaudoin and L. Lakhali-Chaieb, Archimedean copula model selection under dependent truncation, *Statistics in Medicine*, 27 (2008), pp. 4440–4454.
- [43] R.M. Cooke and O. Morales-Napoles, Competing risk and the Cox proportional hazard model, *Journal of Statistical Planning and Inference*, 136 (2006), pp. 1621–1637.
- [44] S.X. Chen and T. Huang, Nonparametric Estimation of Copula Functions for Dependent Modeling, *Canadian Journal of Statistics*, 35 (2007), pp. 265–282.



- [45] J. Demongeot and J. Waku, Counter-examples for the population size growth in demography, *Math. Pop. Studies*, 12 (2005), pp. 199–210.
- [46] A.W. Marshall and I. Olkin, Families of Multivariate Distribution, *Journal of the American Statistical Association*, (1988), pp. 199–210.
- [47] T. Roncalli, La gestion des risques financiers, *Economica, Paris*, (2004).
- [48] J. D’Alembert, Opuscles Mathématiques, *David, Paris*, (1761).
- [49] P. Deheuvels, Caractérisation complète des lois extrêmes multivariées et de la convergence des types extrêmes, *Publications de l’Institut de Statistique de l’Université de Paris*, 23 (1978), pp. 1–36.
- [50] P. Deheuvels, Propriétés d’existence et propriétés topologiques des fonctions de dépendance avec applications à la convergence des types pour des lois multivariées, *Comptes Rendus de l’Académie des Sciences de Paris, Série 1*, 288 (1979), pp. 145–148.
- [51] M. Fréchet, Généralisations du théorème des probabilités totales, *Fundamenta Mathematicae*, 25 (1935), pp. 379–397.
- [52] M. Fréchet, Sur les tableaux de corrélation dont les marges sont données. *Annales de l’Université de Lyon, Section A*, 9 (1951), pp. 53–77.
- [53] M. Fréchet, Sur les tableaux de corrélation et les programmes linéaires. *Revue de l’Institut International de Statistique*, 25 (1957), pp. 23–40.
- [54] M. Fréchet, Remarques au sujet de la note précédente. *Comptes Rendus de l’Académie des Sciences de Paris, Série 1 Math*, 246 (1958), pp. 2719–2720.
- [55] H.K. Beecher, The Powerful Placebo, *Journal of the American Medical Association*, 159 (1955), pp. 1602–1606.
- [56] J.P. LaSalle, Stability theory for ordinary differential equations. *J. Differ. Equations*, 41 (1968), pp. 57–65.
- [57] G. Sallet, Modélisation et simulation en épidémiologie, *INRIA, Paris*, (2010).
- [58] J. Demongeot, M. Ghassani, J. Mintsa, J. Waku and M.Rachdi, Demography and Diffusion in epidemics : Malaria and Black Death spread, *Acta Biotheoretica*, 58 (2010), pp. 277–305.
- [59] L.J. Bruce-Chwatt, Mass drug administration for control of malaria, *The Lancet*, 322 (1983), pp. 688.

## Chapter II

# Demography in epidemics modelling: the copula approach

**Abstract.** Classical models of epidemics by Ross and McKendrick have to be revisited in order to take into account the demography (fecundity, mortality and migration) both of host and vector populations and also diffusion and mutation of infectious agents. We will study three models along different age classes of human, with and without mosquitoes by using the copula function, and we will conduct a simulation study for two of these models.

**Keywords and Phrases:** Demographic dynamics, Copulas functions, Age classes, Cox model.

### II.1 Introduction

The practical use of epidemic models must rely heavily on the realism put into the models. This doesn't mean that a reasonable model can include all possible effects but rather incorporate the mechanisms in the simplest possible fashion so as to maintain major components that influence disease propagation. Great care should be taken before epidemic models are used for prediction of real phenomena. However, even simple models should, and often do, pose important questions about the underlying mechanisms of infection spread and possible means of control of the disease or epidemic.

This chapter is organized as follows: in the next section, we will discuss a classical epidemiological model which is the model by Ross-Macdonald. Then, we will present the model of Leslie, the model of Usher and the Cox model. Afterwards, we will expose our models with different age classes, with and without the mosquitoes compartments. Finally, we will conduct a simulation study for two of these models.

### II.2 Classical Epidemiology: The Ross-McKendrick Model

After the first historical model by D. Bernoulli (cf. [11, 12, 13, 14]) which is proposed for explaining the smallpox dynamics, many discussions occurred about the efficacy of firstly the inoculation and secondly the vaccination (cf. [15, 16, 17, 18, 19]). In the model Princeps, the population is divided into susceptibles (not yet been infected) and immunes (immunized for the rest of their life after one infection), and the two corresponding variables are  $u(a)$ , the

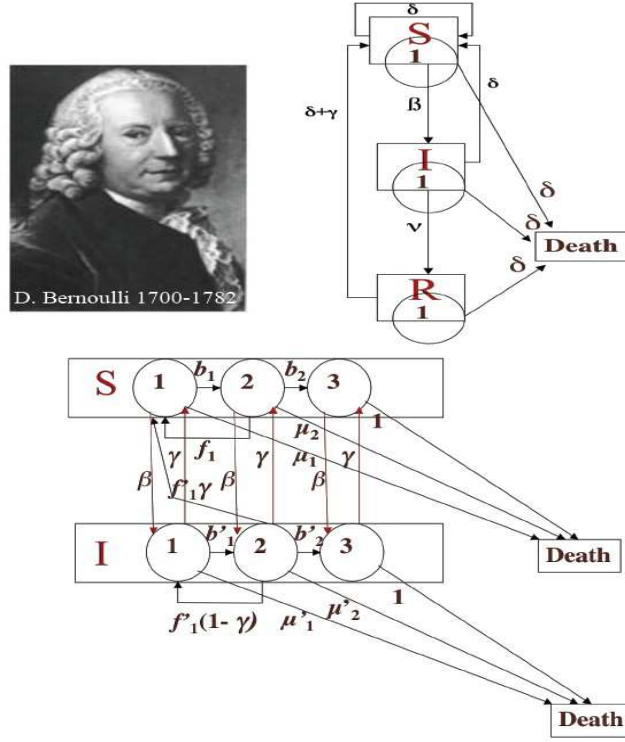


Figure II.1: Picture of D. Bernoulli (top left); interaction graphs of the Ross-McKendrick model having one (top right) and three (bottom) age classes, with identical  $\beta_i$ 's and  $\gamma_i$ 's and no fecundity in elderly classes  $S_3$  and  $I_3$

probability for a newborn individual of being susceptible and alive at age “ $a$ ”, and  $w(a)$ , the probability of being immune and alive at age “ $a$ ”. Ross in [20], Mckendrick in [23], Kermack and Mckendrick in [21, 22] proposed a model called Susceptible-Infectives-Recovered with immunity model (SIR), which is represented by the following equations:

$$\begin{aligned} \frac{dS}{dt} &= \delta S + \delta I + (\delta + \gamma)R - \beta SI - \delta S \\ \frac{dI}{dt} &= \beta SI - (\nu + \delta)I \\ \frac{dR}{dt} &= \nu I - (\delta + \gamma)R \end{aligned} \quad (1)$$

where  $S$  (resp.  $I$ ,  $R$ ) denotes the size of Susceptible (resp. Infective, Recovered) population with  $S + I + R = N$ , and where  $\beta$  (resp.  $\delta$ ,  $\gamma$ ,  $\nu$ ) are the contagion (resp. death/birth, loss of resistance, immunization) rate (cf. Figure II.1 and [3]). The epidemic parameter  $R_0 = \beta N / (\nu + \delta)$  is the mean number of secondary infecteds by one primary infective and predicts, if it is greater than 1, the occurrence of an epidemic wave. By defining the age classes denoted by  $S_i$ ,  $I_i$  and  $R_i$  ( $i = 1, \dots, n$ ) in each subpopulation of  $S$ ,  $I$  and  $R$ , we obtain at any stationary state  $(S^*, I^*, R^*)$  that:

$$u^*(i) = \frac{S_i^*}{S_1^*}, \quad v^*(i) = \frac{I_i^*}{I_1^*}, \quad w^*(i) = \frac{R_i^*}{R_1^*} \quad (2)$$

The relationships (2) between the probabilities for a newborn individual of being alive and either susceptible, infected or immune at age “ $a$ ” make the link between the Bernoulli and the

Ross-McKendrick models, but the weakness of the later still resides in many insufficiencies and approximations:

- when the population size of either susceptibles or infectives tend to be very large, the quadratic term  $SI$  has to be replaced by a Michaelian saturation term  $SI/(k+S)(k'+I)$
- the immunized infectives or healthy carriers are neglected
- the total population size is supposed to be “constant”, the fecundity just equalling the natural mortality. The Bernoulli model takes implicitly into account the fecundity, and explicitly the natural mortality. The model by d’Alembert improved the Bernoulli’s one by distinguishing the specific mortality due to the infectious disease from the natural one, being more widely applicable than the model by Bernoulli which was restricted to immunizing infections. In d’Alembert’s method the only task was to calculate the survival function after eliminating a particular cause of death (the infectious disease), but the Bernoulli’s approach provided much more insight for a mechanistic interpretation of infectious disease data
- variables and parameters are not depending on space (no migration nor population displacement)
- parameters are not depending on time (no genetic adaptation of infectious agent or human population, even very slow compared to the fast dynamics of epidemics).

## II.3 Leslie model

The population growth has been modelled by Leslie (cf. [34, 35]) using the “age pyramid” vector  $x(t) = x_i(t)$  for  $i = 1, \dots, n$ , where  $x_i(t)$  represents the size of the age class  $i$  at time  $t$ , with  $i$  ranging from the birth age 1 to the maximal death age  $n$ , whose discrete dynamics is governed by the matrix equation:

$$x(t) = Lx(t-1) \quad (3)$$

where

$$L = (L_{ij})_{i,j} = \begin{pmatrix} f_1 & f_2 & f_3 & \dots & \cdot & f_n \\ b_1 & 0 & 0 & \dots & \cdot & 0 \\ 0 & b_2 & 0 & \dots & \cdot & 0 \\ \dots & \dots & \dots & \dots & \cdot & \cdot \\ 0 & 0 & 0 & \dots & b_{n-1} & 0 \end{pmatrix} \quad (4)$$

and where  $b_i = 1 - \mu_i \leq 1$ , for all  $i = 1, \dots, n$ , is the survival probability between ages  $i$  and  $i + 1$  and  $f_i$  is the fecundity at age  $i$  (i.e., the mean number of offsprings from an individual at age  $i$ ).

The dynamical stability for the  $L^2$ -distance between the stationary age pyramid  $w$  and the current age pyramid is related to the quantity  $|\lambda - \lambda'|$  i.e, the modulus of the difference between the dominant and the sub-dominant eigenvalues of  $L$ , namely  $\lambda = e^r$  and  $\lambda'$  ( $r$  being the Malthusian growth rate), where  $w$  is the eigenvector of  $L$  corresponding to  $\lambda$ . The dynamical stability for the Kullback distance to the stationary distribution of the probabilities that the mother of a newborn be in age  $i$ , is related to the population entropy  $H$  (cf. [36]).

## II.4 Usher model

The possibility to remain in the same biological age (corresponding to an increase of the longevity) or to pass over a biological age state (corresponding to an acceleration of ageing)

between times  $t$  and  $t + dt$  has been modelled by Usher (cf. [37]), by using the vector  $x(t)$ , whose discrete dynamics is ruled by the matrix equation:

$$x(t) = U x(t - 1) \quad (5)$$

where

$$U = (U_{ij})_{i,j} = \begin{pmatrix} \alpha_1 + f_1 & f_2 & f_3 & \dots & f_{n-1} & f_n \\ \beta_1 & \alpha_2 & 0 & \dots & 0 & 0 \\ \gamma_1 & \beta_2 & \alpha_3 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \alpha_{n-1} & 0 \\ 0 & 0 & 0 & \dots & \beta_{n-1} & \alpha_n \end{pmatrix} \quad (6)$$

and where  $\alpha_i$  (resp.  $\beta_i$  and  $\gamma_i$ ) is the probability to remain in the state  $i$  (resp. to go to the state  $(i + 1)$  and  $(i + 2)$ ) between times  $t$  and  $t + 1$ , with:

$$\alpha_i + \beta_i + \gamma_i = 1 - \mu_i \leq 1, \quad \forall i = 1, \dots, n$$

Like in the Leslie model, the dynamical stability for the  $L^2$ -distance between the stationary age pyramid  $w$  and the current age pyramid is linked to the quantity  $|\lambda - \lambda'|$ .

## II.5 Cox model

The Cox regression model is a useful method that permit to determine the impact of explanatory variables on the survival time of patients. It applies to survival data, that is to say, a time variable, a variable of censorship (binary variable) and of explanatory variables (cf. Chapter 1).

The Cox model with proportional risk is for which the risk function is assumed given by (cf. [6]):

$$h(t, z) = \exp(\rho z) b(t)$$

where  $\rho$  is a regression parameter and  $b(t)$  is the baseline risk function.

Then, if we denote by  $q = \exp(\rho z)$  the fragility factor, then, the survival function (i.e., the probability to survive until the age  $t$ ) is given by:

$$S(t, q) = \exp\left(-\int_0^t h(s, z) ds\right) = B(t)^q$$

where  $B(t) = \exp\left(-\int_0^t b(s) ds\right)$  is the survival function corresponding to the baseline risk.

## II.6 Our models

In this chapter we focus our study on two models. The first one concerns the dynamic modelization where there are three classes of ages, the second focuses on the dynamics for two classes of ages.

## II.6.1 Model 1: Model with three age classes

### II.6.1.1 Introduction of age classes

By introducing age classes, we add new demographic parameters as the fecundity rate, which is equal to the mean number of offsprings a person in class  $i$  is sending to the class 1 between times  $t$  and  $t + dt$ , and the survival (resp. death) rate which is equal to the probability to survive from age  $i$  to age  $i + 1$  (resp. to die at age  $i$ ) between times  $t$  and  $t + dt$ . The equations of the extended Ross-Mckendrick model corresponding to 3 age classes with two compartments of mosquitoes (cf. Figure II.2) are the following:

$$\begin{aligned}
\frac{\partial S_m}{\partial t} &= -\frac{\zeta_1 S_m I_2}{N_H} + j I_m \\
\frac{\partial I_m}{\partial t} &= \frac{\zeta_1 S_m I_2}{N_H} - j I_m \\
\frac{\partial S_1}{\partial t} &= (\gamma_1 - \beta_{11}) E_1 - \beta_{12} S_1 E_2 - (b_1 + \mu_1)(1 - \beta_{11} - \beta_{12}) S_1 + f_1 S_2 + f'_1 \gamma_2 E_2 + f''_1 K_2 I_2 \\
&\quad - \frac{\zeta_2 S_1 I_m}{N_H} \\
\frac{\partial S_2}{\partial t} &= b_1(1 - \beta_{11} - \beta_{12}) S_1 + (\gamma_2 - \beta_{22} S_2) E_2 - \beta_{21} S_2 E_1 - (\mu_2 + b_2)(1 - \beta_{22} - \beta_{21} - \beta_{23}) S_2 \\
&\quad - \beta_{23} S_2 E_3 - \frac{\zeta_2 S_2 I_m}{N_H} \\
\frac{\partial S_3}{\partial t} &= b_2(1 - \beta_{22} - \beta_{21} - \beta_{23}) S_2 + (\gamma_3 - \beta_{33} S_3) E_3 - \mu_3 S_3 - \beta_{32} S_3 E_2 + \gamma_2 E_2 S_3 - \frac{\zeta_2 S_3 I_m}{N_H} \\
\frac{\partial E_1}{\partial t} &= (\beta_{11} E_1 + \beta_{12} E_2) S_1 - \gamma_1 E_1 + f'_1(1 - \gamma_2) E_2 - (b'_1 + \mu'_1)(1 - \gamma_1) E_1 - r_1 E_1 + K_1 I_1 \\
&\quad - r_{12} E_1 I_2 + f''_1 K_2 I_2 + \frac{\zeta_2 S_1 I_m}{N_H} \\
\frac{\partial E_2}{\partial t} &= (\beta_{21} E_1 + \beta_{22} E_2) S_2 - \gamma_2 E_2 - (\mu'_2 + b'_2)(1 - \gamma_2) E_2 - r_2 E_2 + b'_1(1 - \gamma_1) E_1 + K_2 I_2 \\
&\quad - r_{23} E_2 I_3 + \frac{\zeta_2 S_2 I_m}{N_H} \\
\frac{\partial E_3}{\partial t} &= b'_2(1 - \gamma_2) E_2 + \beta_{33} S_3 E_3 - (\mu'_3 + \gamma_3 + r_3) E_3 + K_3 I_3 + K_{23} I_2 E_3 + \beta_{23} S_2 E_3 + \frac{\zeta_2 S_3 I_m}{N_H} \\
\frac{\partial I_1}{\partial t} &= r_1 E_1 - K_1 I_1 + f''_1(1 - K_2) I_2 - (b''_1 + \mu''_1)(1 - K_1) I_1 - K_{12} I_1 E_2 \\
\frac{\partial I_2}{\partial t} &= r_2 E_2 + b''_1(1 - K_1) I_1 - K_2 I_2 - (b''_2 + \mu''_2)(1 - K_2) I_2 - K_{23} I_2 E_3 \\
\frac{\partial I_3}{\partial t} &= b''_2(1 - K_2) I_2 + r_3 E_3 + K_{23} E_2 I_3 - (K_3 + \mu''_3) I_3
\end{aligned} \tag{7}$$

where:

- $\mu_1, \mu_2$  and  $\mu_3$  are the mortality rates of the three age classes respectively of the vector susceptible population.
- $\mu'_1, \mu'_2$  and  $\mu'_3$  are the mortality rates of the three age classes respectively of the vector infected population.
- $\mu''_1, \mu''_2$  and  $\mu''_3$  are the mortality rates of the three age classes respectively of the vector infective population.
- $f_1, f'_1$  and  $f''_1$  are the fecundity rates in the second class of ages respectively of the vector susceptible, infected, infective populations.

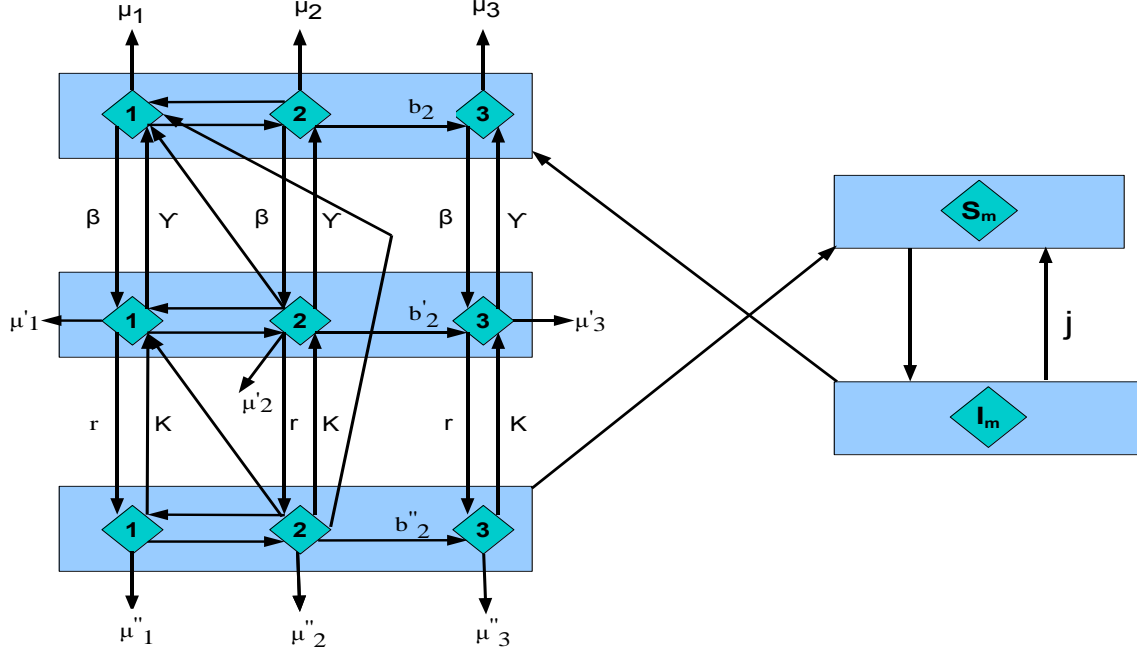


Figure II.2: interaction graph of the model (7) with three age classes and two compartments of mosquitoes

- $b_1$  is the speed of passage for the vectors from the age class 1 to the age class 2 in the susceptible state.
- $b_2$  is the speed of passage for the vectors from the age class 2 to the age class 3 in the susceptible state.
- $b'_1$  is the speed of passage for the vectors from the age class 1 to the age class 2 in the infected state.
- $b'_2$  is the speed of passage for the vectors from the age class 2 to the age class 3 in the infected state.
- $b''_1$  is the speed of passage for the vectors from the age class 1 to the age class 2 in the infective state.
- $b''_2$  is the speed of passage for the vectors from the age class 2 to the age class 3 in the infective state.
- $r$  is the speed of passage for the vectors from the infected state to the infective state.
- $K$  is the speed of passage for the vectors from the infective state to the infected state.
- $\beta$  is the speed of passage for the vectors from the susceptible state to the infected state.
- $\gamma$  is the speed of passage for the vectors from the infected state to the susceptible state.
- $j$  is the speed of passage for the hosts from the infected state to the susceptible state.
- $\zeta_1$  is the host contagion parameter.
- $\zeta_2$  is the vector contagion parameter.

It is very difficult, for the moment, to solve this system, but we can proceed (to understand the phenomenon) by treating some special cases of the system (7) which one particularly interesting, that is:

1. with two age classes and two compartments of mosquitoes (cf. section II.6.2).
2. with three age classes without any compartment of mosquitoes (cf. section II.6.3).

### II.6.1.2 The Copula approach

**Proposition II.1** *Assume that there exist three age classes into the host subpopulations whose sojourn times  $T_i$  for  $i = 1, 2, 3$ , are independent random variables defined on the probabilized space  $(\Omega, \mathfrak{F}, \mathbb{P})$ , then we can relate the survival functions  $S_j$  for  $j = 1, 2, 3$ , by:*

$$\mathbb{P}(T_i > t_i \text{ for } i = 1, 2, 3) = \exp \left[ - \left( \sum_{j=1,2,3} \left( -\ln(S_j(t_j))^{\frac{1}{\alpha}} \right) \right)^{\alpha} \right] \quad \text{for } t_1, t_2, t_3 > 0$$

where  $\alpha$  is a parameter of dependence.

**Proof.**

we define the mean survival function by:

$$S(t) = \mathbb{E}_q [B(t)^q]$$

where  $\mathbb{E}_q$  denotes the conditional expectation relatively to  $q$ .

Recall that the Laplace transform of a positive random variable  $q$  is defined by:

$$L_q(s) = \mathbb{E}_q [\exp(-sq)] = \int \exp(-st) dG_q(t)$$

where  $G_q$  is the distribution function of  $q$ . It is also the generating function evaluated at  $(-s)$ ; thus, the knowledge of  $L_q(s)$  determines entirely the distribution of  $q$ .

Using the Laplace transformation, we obtain:

$$\mathbb{E}_q (\exp(-sq)) = \exp(-s^\alpha) \quad (8)$$

On the other hand, we have that:

$$\begin{aligned} \mathbb{P}(T_i > t_i \text{ for } i = 1, 2, 3) &= \mathbb{E}_q \left[ \prod_{j=1,2,3} B_j(t_j)^q \right] \\ &= E_q \left[ \prod_{j=1,2,3} \exp(q \ln(B_j(t_j))) \right] \\ &= E_q \left[ \exp \left( \sum_{j=1,2,3} q \ln(B_j(t_j)) \right) \right] \end{aligned}$$

From the equation (8), we have:

$$\mathbb{P}(T_i > t_i \text{ for } i = 1, 2, 3) = \exp \left[ - \left( - \sum_{j=1,2,3} \ln(B_j(t_j)) \right)^{\alpha} \right]$$

Then, we have:

$$S_i(t_i) = \exp \left[ - \left( - \sum_{j=1,2,3} \ln(B_j(t_j)) \right)^{\alpha} \right]$$



Thus:

$$\left( - \sum_{j=1,2,3} \ln(B_j(t_j)) \right)^\alpha = \sum_{j=1,2,3} \left( - \ln(S_j(t_j))^{\frac{1}{\alpha}} \right)^\alpha$$

Therefore:

$$\mathbb{P}(T_i > t_i \text{ for } i = 1, 2, 3) = \exp \left[ - \left( \sum_{j=1,2,3} \left( - \ln(S_j(t_j))^{\frac{1}{\alpha}} \right) \right)^\alpha \right]$$

■

In another side, we defined the archimedean copula as follows (cf. chapter 1):

$$C(u_1, \dots, u_n) = \begin{cases} \phi^{-1}(\phi(u_1) + \dots + \phi(u_n)) & \text{if } \phi(u_1) + \dots + \phi(u_n) \leq 0 \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

where the generator of the copula  $\phi$  is a twice continuously differentiable function which satisfies:

$$\phi(1) = 0, \quad \phi^{(1)}(u) < 0 \quad \text{and} \quad \phi^{(2)}(u) > 0 \quad \text{for all } u \in [0, 1]^n$$

where  $\phi^{(i)}$  denotes the  $i$ -th order derivative of  $\phi$ .

Notice that a popular archimedean copula is the Gumbel-Hougaard copula that is defined as follows (cf. chapter 1):

$$\mathbf{C}(u_1, \dots, u_n) = \exp \left\{ - [(-\ln(u_1))^\alpha + \dots + (-\ln(u_n))^\alpha]^{\frac{1}{\alpha}} \right\} \quad (10)$$

where  $\alpha \geq 1$  and  $\phi(t) = (-\ln t)^\alpha$ .

So from Proposition II.1 and the equation (10), we obtain:

$$\begin{aligned} \mathbb{P}(T_i > t_i \text{ for } i = 1, 2, 3) &= \exp \left[ - \left( \sum_{j=1,2,3} \left( - \ln(S_j(t_j))^{\frac{1}{\alpha}} \right) \right)^\alpha \right] \\ &= \mathbf{C}(S_1, S_2, S_3) \end{aligned} \quad (11)$$

where  $\mathbf{C}$  is a Gumbel-Hougaard copula.

## II.6.2 Model 2: Model with two age classes

### II.6.2.1 Introduction of age classes

By taking the model (II.6.1) and removing one class of ages and the corresponding passage speeds, then we find a model with two age classes as in Figure II.3. Equations of this model are as given in the following system:

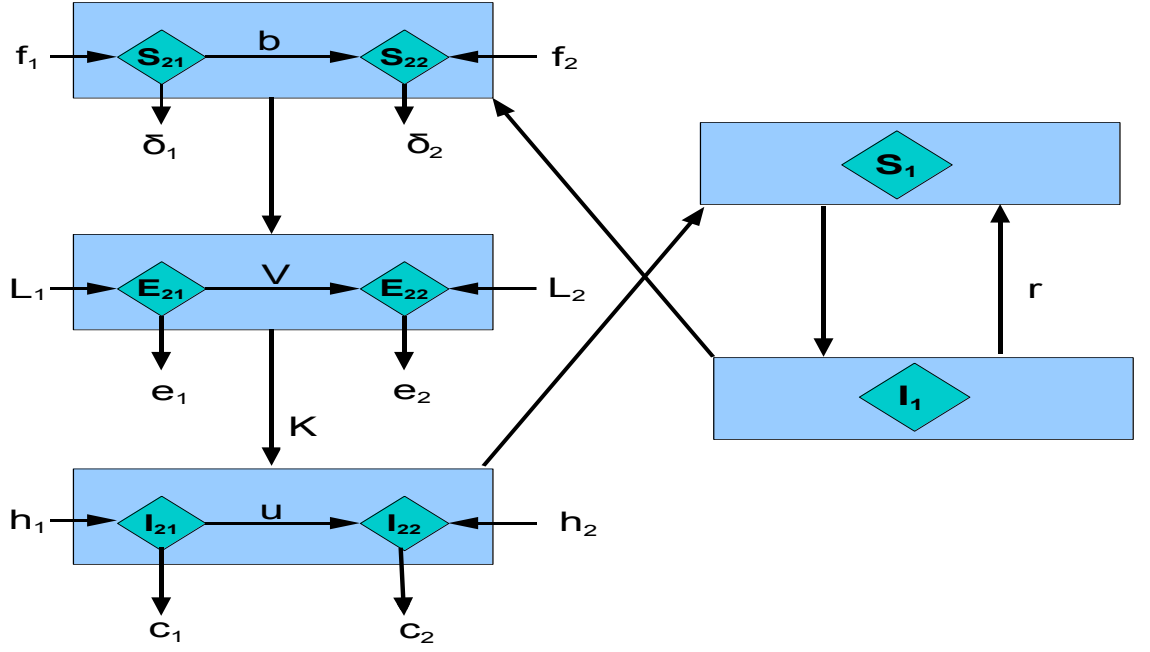


Figure II.3: interaction graph of the model (12) with two age classes and two compartments of mosquitoes

$$\begin{aligned}
 \frac{\partial S_1}{\partial t} &= \frac{-\beta_1 S_1 I_{22}}{N_H} + r I_1 \\
 \frac{\partial I_1}{\partial t} &= \frac{\beta_1 S_1 I_{22}}{N_H} - r I_1 \\
 \frac{\partial S_{21}}{\partial t} &= f_1 S_{21} - \frac{\beta_2 S_{21} I_1}{N_H} - \delta_1 S_{21} - b S_{21} \\
 \frac{\partial S_{22}}{\partial t} &= f_2 S_{22} + b S_{21} - \frac{\beta_2 S_{22} I_1}{N_H} - \delta_2 S_{22} \\
 \frac{\partial I_{21}}{\partial t} &= h_1 I_{21} - u I_{21} - c_1 I_{21} + K E_{21} \\
 \frac{\partial I_{22}}{\partial t} &= u I_{21} + h_2 I_{22} + K E_{22} - c_2 I_{22} \\
 \frac{\partial E_{21}}{\partial t} &= l_1 E_{21} - v E_{21} + \frac{\beta_2 S_{21} I_1}{N_H} - K E_{21} - e_1 E_{21} \\
 \frac{\partial E_{22}}{\partial t} &= v E_{21} + l_2 E_{22} + \frac{\beta_2 S_{22} I_1}{N_H} - K E_{22} - e_2 E_{22}
 \end{aligned} \tag{12}$$

where:

- $f_1$  is the fecundity rate in the first age class for the susceptibles
- $f_2$  is the fecundity rate in the second age class for the susceptibles
- $l_1$  is the fecundity rate in the first age class for the infected
- $l_2$  is the fecundity rate in the second age class for the infected

- $h_1$  is the fecundity rate in the first age class for the infectives
- $h_2$  is the fecundity rate in the second age class for the infectives
- $\delta_1$  is the mortality rate in the first age class for the susceptibles
- $\delta_2$  is the mortality rate in the second age class for the susceptibles
- $e_1$  is the mortality rate in the first age class for the infected
- $e_2$  is the mortality rate in the second age class for the infected
- $c_1$  is the mortality rate in the first age class for the infectives
- $c_2$  is the mortality rate in the second age class for the infectives
- $\beta_1$  is the host contagion parameter
- $\beta_2$  is the vector contagion parameter
- $N_H$  is the host population size
- $N_V$  is the vector population size
- $m = N_V/N_H$  is the vector density
- $K$  is the speed of passage for the vectors from the infected/not infective state to the infective state
- $r$  is the speed of passage for the hosts from the infected state to the susceptible state
- $b$  is the speed of passage for the vectors from the age class 1 to the age class 2 in the susceptible state
- $v$  is the speed of passage for the vectors from the age class 1 to the age class 2 in the infected state
- $u$  is the speed of passage for the vectors from the age class 1 to the age class 2 in the infective state

After some calculations, the value of  $R_0$  for the system (12) is then given by:

$$R_0 = \frac{K\beta_1(f_2 - \delta_2)N_V}{rN_H(h_2 - c_2)(l_2 - e_2 - K)}$$

Since at the equilibrium state:

$$E_{22} = \frac{(f_2 - \delta_2)S_{22}}{-l_2 + K + e_2}$$

and we get:

$$\begin{aligned} S_1 &= \frac{ri_1N_H(l_2h_2 - l_2c_2 - Kh_2 + Kc_2 - e_2h_2 + e_2c_2)}{\beta_1S_{22}K(f_2 - \delta_2)} \\ i_1 &= \frac{N_H(f_2 - \delta_2)}{\beta_2} \\ i_{22} &= \frac{S_{22}(f_2 - \delta_2)K}{l_2h_2 - l_2c_2 - Kh_2 + Kc_2 - e_2h_2 + e_2c_2} \end{aligned}$$

and  $S_{21} = E_{21} = i_{21} = 0$

Since these equilibrium points depend on  $S_{22}$ , then we do not obtain a point, but we obtain a line of equilibrium.

On the other hand, if  $f_2 - \delta_2 = T(t) = T$ ,  $h_2 - c_2 = M(t) = M$  and  $l_2 - e_2 = P(t) = P$ , then:

$$R_0 = \frac{K\beta_1TN_V}{rN_HM(P - K)} = \frac{K\beta_1Tm}{rM(P - K)}$$

Consequently, we obtain the following result:

**Proposition II.2**

If the rate  $P = l_2 - e_2$  is small with respect to the speed  $K$ , then:

$$\frac{K}{P - K} \approx -\exp\left(\frac{P}{K}\right)$$

where  $1/K$  is the mean sojourn time in the compartment  $E_2$  (sporogonic cycle duration), and if we assume that  $h_2 - c_2 = M$  is negative (the mortality of the second age class in the compartment of infected is greater than the fecundity), then  $R_0$  is given by:

$$R_0 \approx \frac{\beta_1 T m}{r M} \exp\left(\frac{P}{K}\right) \quad (13)$$

Remark that when  $M$  increases and  $T$  decreases, then  $R_0$  decreases.

**II.6.2.2 The copula approach**

According to  $R_0$  expression given by (13) and the Cox model (cf. Section II.5), we can assume that:

$$z = P, \quad \rho = \frac{1}{K} \quad \text{and} \quad b(t) = \frac{\beta_1 T m}{r M}$$

Also, from the Cox model (cf. Section II.5), we can suppose that:  $q = \exp(P/K)$ , and consequently the survival function corresponding to the baseline risk  $B(t)$  will be:

$$B(t) = \exp\left(-\int_0^t \frac{\beta_1 T(s) m}{r M(s)} ds\right) \approx \exp\left(-\int_0^t R_0(s) q ds\right)$$

Then, we can obtain the Proposition II.1 result from the same computational aspects:

$$\mathbb{P}(T_i > t_i \text{ for } i = 1, 2, 3) = \exp\left[-\left(\sum_{j=1,2,3} \left(-\ln(S_j(t_j))^{\frac{1}{\alpha}}\right)\right)^\alpha\right]$$

where  $S_j$  depends on the parameters that are in  $R_0$ .

The relationship between the different age classes is reduced by the archimedean copulas functions properties as in the equation (11), but with two age classes, we obtain:

$$\mathbb{P}(T_i > t_i \text{ for } i = 1, 2, 3) = \mathbf{C}(S_1, S_2)$$

where  $\mathbf{C}$  is an archimedean copula, precisely a Gumbel-Hougaard copula.

### II.6.3 Model 3: three age classes without mosquitoes

By taking the model II.6.1 and by eliminating the compartments of mosquitoes, we find the model illustrated in Figure II.4. The equations corresponding to this model are as follows:

$$\begin{aligned}
\frac{\partial S_1}{\partial t} &= (\gamma_1 - \beta_{11}) E_1 - \beta_{12} S_1 E_2 - (b_1 + \mu_1)(1 - \beta_{11} - \beta_{12}) S_1 + f_1 S_2 + f_1' \gamma_2 E_2 \\
&\quad + f_1'' K_2 I_2 \\
\frac{\partial S_2}{\partial t} &= b_1(1 - \beta_{11} - \beta_{12}) S_1 + (\gamma_2 - \beta_{22} S_2) E_2 - \beta_{21} S_2 E_1 - (\mu_2 + b_2)(1 - \beta_{22} \\
&\quad - \beta_{21} - \beta_{23}) S_2 - \beta_{23} S_2 E_3 \\
\frac{\partial S_3}{\partial t} &= b_2(1 - \beta_{22} - \beta_{21} - \beta_{23}) S_2 + (\gamma_3 - \beta_{33} S_3) E_3 - \mu_3 S_3 - \beta_{32} S_3 E_2 + \gamma_2 E_2 S_3 \\
\frac{\partial E_1}{\partial t} &= (\beta_{11} E_1 + \beta_{12} E_2) S_1 - \gamma_1 E_1 + f_1'(1 - \gamma_2) E_2 - (b_1' + \mu_1')(1 - \gamma_1) E_1 - r_1 E_1 \\
&\quad + K_1 I_1 - r_{12} E_1 I_2 + f_1'' K_2 I_2 \\
\frac{\partial E_2}{\partial t} &= (\beta_{21} E_1 + \beta_{22} E_2) S_2 - \gamma_2 E_2 - (\mu_2' + b_2')(1 - \gamma_2) E_2 - r_2 E_2 + b_1'(1 - \gamma_1) E_1 \\
&\quad + K_2 I_2 - r_{23} E_2 I_3 \\
\frac{\partial E_3}{\partial t} &= b_2'(1 - \gamma_2) E_2 + \beta_{33} S_3 E_3 - (\mu_3' + \gamma_3 + r_3) E_3 + K_3 I_3 + K_{23} I_2 E_3 + \beta_{23} S_2 E_3 \\
\frac{\partial I_1}{\partial t} &= r_1 E_1 - K_1 I_1 + f_1''(1 - K_2) I_2 - (b_1'' + \mu_1'')(1 - K_1) I_1 - K_{12} I_1 E_2 \\
\frac{\partial I_2}{\partial t} &= r_2 E_2 + b_1''(1 - K_1) I_1 - K_2 I_2 - (b_2'' + \mu_2'')(1 - K_2) I_2 - K_{23} I_2 E_3 \\
\frac{\partial I_3}{\partial t} &= b_2''(1 - K_2) I_2 + r_3 E_3 + K_{23} E_2 I_3 - (K_3 + \mu_3'') I_3
\end{aligned} \tag{14}$$

Since in the system (14), there is no mosquito, so it is not possible to calculate the basic reproduction rate. To be more precise, by introducing the age classes, in particular with the biological age (cf. [1]), then we have to recall the classical models used for modelling the population growth (cf. Sections II.3 and II.4).

## II.7 Simulation studies

### II.7.1 The simulation of the model 2

Let us consider the model given by (II.6.2) with two age classes and let us assume:

$$\begin{aligned}
f_1 &= 75k, \quad f_2 = 25k, \quad \delta_1 = 2k/3, \quad \delta_2 = 4k/5, \quad b = 98k/96, \quad l_1 = 30k \\
l_2 &= 101k, \quad v = 70k/96, \quad e_1 = k, \quad e_2 = 6k/5, \quad h_1 = 15k, \quad h_2 = 3k \\
u &= 50k/96, \quad c_1 = 4k/3, \quad c_2 = 2k, \quad \beta_1 = 4k^2/100, \quad \beta_2 = k^2/10, \quad K = 9k/10, \quad r = 2k/5.
\end{aligned}$$

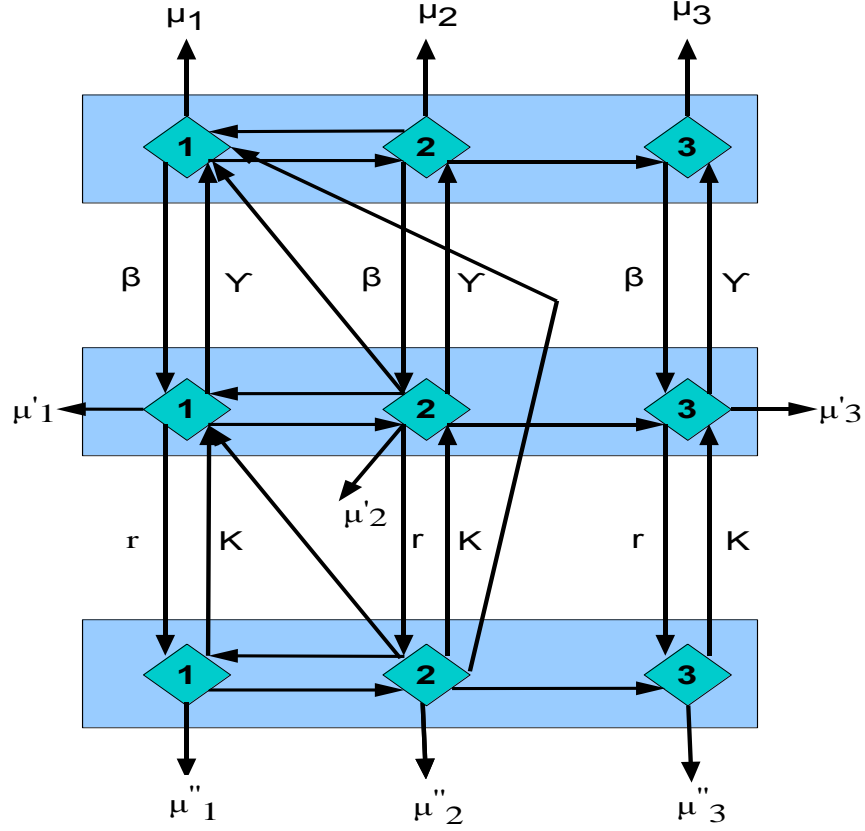


Figure II.4: Interaction graph for the model (14) with three age classes

Then, if  $k = 1$  and  $N_H = 10^6$ , the differential system ruling the population dynamics becomes as follows:

$$\begin{aligned}
 \frac{\partial S_1}{\partial t} &= \frac{-1}{25 \cdot 10^6} S_1 I_{22} + \frac{2}{5} I_1 \\
 \frac{\partial I_1}{\partial t} &= \frac{1}{25 \cdot 10^6} S_1 I_{22} - \frac{2}{5} I_1 \\
 \frac{\partial S_{21}}{\partial t} &= \frac{-1}{10^8} S_{21} I_1 - \frac{1173}{16} S_{21} \\
 \frac{\partial S_{22}}{\partial t} &= \frac{49}{48} S_{21} - \frac{-1}{10^8} S_{22} I_1 + \frac{121}{5} S_{22} \\
 \frac{\partial I_{21}}{\partial t} &= 0.9 E_{21} + \frac{631}{48} I_{21} \\
 \frac{\partial I_{22}}{\partial t} &= 0.9 E_{22} + \frac{25}{48} I_{21} + I_{22} \\
 \frac{\partial E_{21}}{\partial t} &= \frac{1}{10^8} S_{21} I_1 + \frac{6569}{240} E_{21} \\
 \frac{\partial E_{22}}{\partial t} &= \frac{35}{48} E_{21} + \frac{1}{10^8} S_{22} I_1 + \frac{79}{10} E_{22}
 \end{aligned} \tag{15}$$

The stationary points of the system (15) are then given by:

$$(s_1^*, i_1^*, s_{21}^*, s_{22}^*, i_{21}^*, i_{22}^*, e_{21}^*, e_{22}^*) = \left( \frac{79 \times 10^{15}}{9 S_{22}}, 242 \times 10^7, 0, S_{22}, 0, -\frac{1089 S_{22}}{395}, 0, \frac{-242 S_{22}}{79} \right)$$

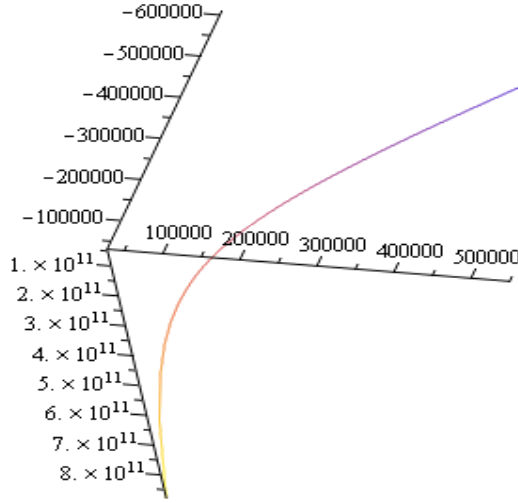


Figure II.5: Graph of the equilibrium line

and the Jacobian matrix  $B^*$  corresponding to the system (15) is given by:

$$B^* = \begin{pmatrix} \frac{-1089}{9875 \times 10^6} S_{22} & 2/5 & 0 & 0 & 0 & \frac{-316 \times 10^7}{9} S_{22} & 0 & 0 \\ \frac{1089}{9875 \times 10^6} S_{22} & -2/5 & 0 & 0 & 0 & \frac{-316 \times 10^7}{9} S_{22} & 0 & 0 \\ 0 & 0 & \frac{3929}{80} & 0 & 0 & 0 & 0 & 0 \\ 0 & -10^{-8} S_{22} & \frac{49}{48} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{631}{48} & 0 & \frac{9}{10} & 0 \\ 0 & 0 & 0 & 0 & \frac{25}{48} & 1 & 0 & \frac{9}{10} \\ 0 & 0 & \frac{121}{5} & 0 & 0 & 0 & \frac{6569}{240} & 0 \\ 0 & 10^{-8} S_{22} & 0 & \frac{121}{5} & 0 & 0 & \frac{35}{48} & \frac{79}{10} \end{pmatrix}$$

By varying  $S_{22}$  between  $10^4$  and  $2 \times 10^5$ , we find the equilibrium line, which depends of  $S_{22}$  (cf. Figure II.5). For the same values of  $S_{22}$ , we find also that the dominant eigenvalue of  $B^*$  is strictly positive and consequently, from the unstable endemic state, the susceptible population grows with a Malthusian parameter which is greater than 0 (the dominant eigenvalue of the Leslie matrix restricted to susceptibles is greater than 1), then  $N$  is not constant and the Ross-McKendrick framework is no more available.

### II.7.2 The simulation of the model 3

Let us consider the model (II.6.3) with three age classes, whose dynamics is driven by equations (16) and interaction graph is given in the Figure II.4. Various possibilities of the demographic evolution and the stability of the endemic state can be observed depending on the set of values fixed for the model parameters:

$$\begin{aligned} f_1 &= 499k/100, f'_1 = 0.2k, f''_1 = 0.02k, b_2 = b_1 = 98k/96, b'_2 = b'_1 = 4k/5, \\ b''_2 &= b''_1 = 2k/3, \mu_1 = \mu'_1 = \mu''_1 = 0, \mu_2 = 49k/100, \mu'_2 = \mu''_2 = 4k/5, \\ \mu_3 &= \mu'_3 = \mu''_3 = 2k/3, r_1 = \beta_{11} = 4k^2/100, K_1 = \gamma_1 = k/5. \end{aligned}$$

and all the other terms are assumed equal to zero. Then, if  $k = 1$ , the differential system ruling the population dynamics is then defined by:

$$\begin{aligned}
\frac{\partial S_1}{\partial t} &= \frac{4}{25}E_1 - \frac{49}{50}S_1 + \frac{499}{100}S_2 \\
\frac{\partial S_2}{\partial t} &= \frac{49}{50}S_1 - \frac{1813}{1200}S_2 \\
\frac{\partial S_3}{\partial t} &= \frac{49}{48}S_2 - \frac{2}{3}S_3 \\
\frac{\partial E_1}{\partial t} &= \frac{1}{25}E_1S_1 - \frac{9}{25}E_1 + 0.2E_2 + \frac{1}{5}I_1 \\
\frac{\partial E_2}{\partial t} &= -\frac{8}{5}E_2 + \frac{16}{25}E_1 \\
\frac{\partial E_3}{\partial t} &= \frac{4}{5}E_2 - \frac{2}{3}E_3 \\
\frac{\partial I_1}{\partial t} &= \frac{1}{25}E_1 - \frac{11}{15}I_1 + 0.02I_2 \\
\frac{\partial I_2}{\partial t} &= \frac{8}{15}I_1 - \frac{22}{15}I_2 \\
\frac{\partial I_3}{\partial t} &= \frac{2}{3}I_2 - \frac{2}{3}I_3
\end{aligned} \tag{16}$$

The two stationary points are given by :

$$(s_1, s_2, s_3, e_1, e_2, e_3, i_1, i_2, i_3) = (0, 0, 0, 0, 0, 0, 0, 0, 0) \tag{17}$$

and

$$(s_1^*, s_2^*, s_3^*, e_1^*, e_2^*, e_3^*, i_1^*, i_2^*, i_3^*) = (19.72, 12.79, 19.59, 278.20, 111.28, 133.54, 15.3, 5.57, 5.57) \tag{18}$$

We remark that both stationary points defined by (17) and (18) are locally unstable. For the first stationary point (17), it is easily proved, by calculating the Jacobian matrix  $B_1$  of the system (16) and by searching the roots of its characteristic polynomial, that the endemic state is unstable. For the second stationary point given by (18) and called the endemic state, we have:

$$\begin{aligned}
& B_1 - \lambda I \simeq \\
& \begin{pmatrix} -1 - \lambda & 5 & 0 & 4/25 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1813/1200 - \lambda & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2/3 - \lambda & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/25 & 0 & 0 & -9/25 - \lambda & 0.2 & 0 & 1/5 & 0 & 0 \\ 0 & 0 & 0 & 16/25 & -8/5 - \lambda & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4/5 & -2/3 - \lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/25 & 0 & 0 & -11/15 - \lambda & 0.02 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 8/15 & -22/15 - \lambda & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2/3 & -2/3 - \lambda \end{pmatrix}
\end{aligned}$$

Notice that the characteristic polynomial  $P_{B_1}$  of  $B_1$  verifies:

$$P_{B_1}(\lambda) = 7.25 \lambda + 18.47 \lambda^2 + 8.99 \lambda^3 - 33.17 \lambda^4 - 56.80 \lambda^5 + 0.87 - 30.82 \lambda^6 - 0.83 \lambda^7 + 4.31 \lambda^8 + \lambda^9$$



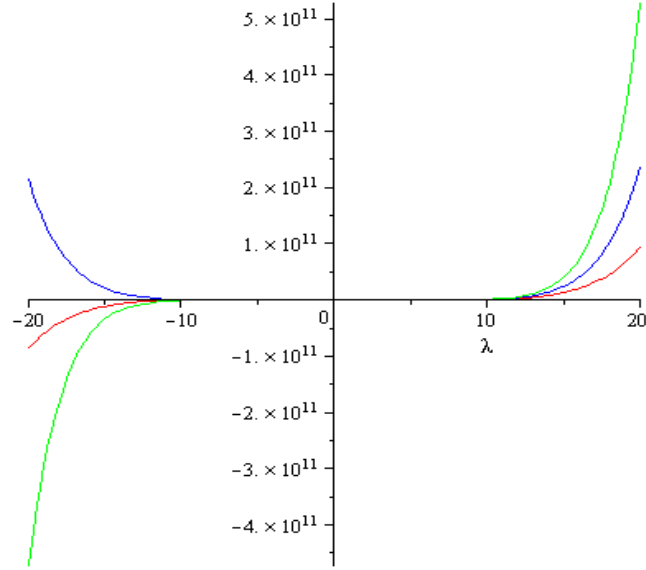


Figure II.6: Graph of the functions  $P_{B_1}(\lambda)$  (in red color),  $P'_{B_1}(\lambda)$  (in blue color) and  $P''_{B_1}(\lambda)$  (in green color)

Thus, the successive derivatives of the characteristic polynomial are given by:

$$P'_{B_1}(\lambda) = 7.25 + 36.95 \lambda + 26.97 \lambda^2 - 132.70 \lambda^3 - 284.02 \lambda^4 - 184.94 \lambda^5 - 5.85 \lambda^6 + 34.52 \lambda^7 + 9 \lambda^8$$

$$P''_{B_1}(\lambda) = 36.95 + 53.95 \lambda - 398.11 \lambda^2 - 1136.10 \lambda^3 - 924.74 \lambda^4 - 35.15 \lambda^5 + 241.68 \lambda^6 + 72 \lambda^7$$

It is clear that the characteristic polynomial coefficients are of different signs, and if we construct the Routh-Hurwitz table, then it appears that the elements of the first column of this table are also of different signs which indicate that the endemic state is unstable.

## II.8 Conclusion

We have considered, in this chapter, some natural extensions of the classical Ross-McKendrick-Macdonald approaches, adding in each compartment of vector the age classes. Two examples have been presented, which show the interest of the introduction of age classes into the classical equation, by presenting the interaction graph for each model. Then, we presented several models such as the Leslie model, the Usher model and the Cox model, all in order to perform a copula function approach with the various presented models. This copula approach allows us to find the relationship between the different age class to see how we can reduce the infectious contacts. Also, we discussed the equilibrium of some systems, and finally we did simulation studies to properly present the models and their equilibrium.

In the future, other realistic examples will be treated showing also the importance of the demography and the sociogeography.

# Bibliography

- [1] J. Demongeot, Biological boundaries and biological age, *Acta Biotheoretica*, 57 (2009), pp. 397–419.
- [2] W.E. Frees and E.A. Valdez, Understanding Relationships Using Copulas, *Actuarial Research Conference, at the University of Calgary, Calgary, Alberta, Canada*, (1997).
- [3] J. Gaudart, M. Ghassani, J. Mintsa, J. Waku, M. Rachdi, O.K. Doumbo and J. Demongeot, Demographic and spatial factors as causes of an epidemic spread, the copula approach, *IEEE Advanced Information Networking and Application (AINA '10, Perth April 2010), IEEE Proceedings, Piscataway*, (2010), pp. 751–758.
- [4] R.M. May and R.M. Anderson, Spatial heterogeneity and the design of immunization programs, *Math. Biosci.*, 72 (1984), pp. 83–111.
- [5] J. Demongeot and J. Waku, Counter-examples for the population size growth in demography, *Math. Pop. Studies*, 12 (2005), pp. 199–210.
- [6] W.O. Cox, Regression Models and Life-Tables, *Journal of the Royal Statistical Society Series B (Methodological)*, 34 (1972), pp. 187–220.
- [7] G. Macdonald, The epidemiology and control of malaria, *Oxford University Press, London*, (1957).
- [8] C. Mason, Cox proportional hazard models, <http://www.demog.berkeley.edu/213/Week14/welcome.pdf>, (2005).
- [9] R.B. Nelsen, Copulas and Association. In : Advances in Probability Distributions With Given Marginals. Beyond the Copulas, *Kluwer, Dordrecht*, (1991), pp. 51–74.
- [10] D. Bernoulli, Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'inoculation pour la prévenir, *Mém. Acad. Roy. Sci., Paris*, (1760).
- [11] E.C. Zeeman, Controversy in science: on the ideas of Daniel Bernoulli and René Thom, *Nieuw Arch. Wisk.* 11 (1993), pp. 257–270.
- [12] K. Dietz and J.A.P. Heesterbeek, Bernoulli was ahead of modern epidemiology, *Nature*, 408 (2000), pp. 513–520.
- [13] K. Dietz and J.A.P. Heesterbeek, Daniel Bernoulli's epidemiological model revisited, *Math. Biosci.*, 180 (2002), pp. 1–21.

- [14] J. d'Alembert, *Opuscules Mathématiques, David, Paris*, 2 (1761), pp. 35–38.
- [15] J.H. Lambert, *Beitrige zum Gebrauche der Mathematik und deren Anwendung, Dritter Theil, Berlin*, (1772).
- [16] G.J. Lépine, *Rapport de six des douze commissaires (contre l'inoculation), Paris*, (1764), pp. 40–43.
- [17] N. May, *Impartial remarks on the Suttonian method of inoculation, London*, (1770).
- [18] J.A. Murray, *Fata variolarum insitionis in Suecia, Göttingen*, (1763).
- [19] J. Trembley, *Recherches sur la mortalité de la petite vérole, Mém. Acad. Roy. Sci., Paris*, (1796).
- [20] R. Ross, *An Application of the Theory of Probabilities to the Study of a priori Pathometry. Part I, Proceedings of the Royal Society of London Series A*, 92 (1916), pp. 204–230.
- [21] W.O. Kermack and A.G. McKendrick, *Contributions to the Mathematical Theory of Epidemics II. The Problem of Endemicity, Proceedings of the Royal Society of London Series A*, 138 (1932), pp. 55–83.
- [22] W.O. Kermack and A.G. McKendrick, *Contributions to the Mathematical Theory of Epidemics III. Further Studies of the Problem of Endemicity, Proceedings of the Royal Society of London Series A*, 141 (1933), pp. 94–122.
- [23] A.G. McKendrick, *Applications of mathematics to medical problems, Proc. Edinburgh Mathematical Society*, 44 (1925), pp. 1–34.
- [24] A.W. Marshall and I. Olkin, *Families of multivariate distribution, Journal of the American Statistical Association*, (1988), pp. 199–210.
- [25] T. Roncalli, *La gestion des risques financiers, Economica, Paris*, (2004).
- [26] P. Deheuvels, *Caractérisation complète des lois extrêmes multivariées et de la convergence des types extrêmes, Publications de l'Institut de Statistique de l'Université de Paris*, 23 (1978), pp. 1–36.
- [27] P. Deheuvels, *Propriétés d'existence et propriétés topologiques des fonctions de dépendance avec applications à la convergence des types pour des lois multivariées, Comptes Rendus de l'Académie des Sciences de Paris, Série 1*, 288 (1979), pp. 145–148.
- [28] S.Y. Zhang, L. Yu and P. Daszak, *EcoHealth and the Black Death in the Year of the Rat, Ecohealth*, 5 (2008), pp. 99–100.
- [29] J. Gaudart, R. Giorgi, B. Poudiougou, S. Ranque, O. Doumbo and J. Demongeot, *Spatial cluster detection: principle and application of different general methods, Rev. Epidemiol. Santé Publique*, 55 (2007), pp. 297–306.

- [30] J. Gaudart, O. Touré, N. Dessay, A.L. Dicko, S. Ranque, L. Forest, J. Demongeot and O. Doumbo, Modelling malaria incidence with environmental dependency in a locality of Sudanese savannah area, *Mali. Malar. J.*, 8 (2009), pp. 61–80.
- [31] D. Beaudoin and L. Lakhal-Chaieb, Archimedean copula model selection under dependent truncation, *Statistics in Medicine*, 27 (2008), pp. 4440–4454.
- [32] J. Dutertre, Étude d'un modèle épidémiologique appliqué au paludisme, *Ann. Soc. Belge. Méd. Trop.*, 56 (1976), pp. 127–141.
- [33] R.M. Cooke and O. Morales-Napoles, Competing risk and the Cox proportional hazard model, *Journal of Statistical Planning and Inference*, 136 (2006), pp. 1621–1637.
- [34] P. Leslie, On the use of matrices in certain population mathematics, *Biometrika*, 33 (1945), pp. 184–212.
- [35] P. Leslie, Some further notes on the use of matrices in population mathematics, *Biometrika*, 35 (1948), pp. 213–245.
- [36] J. Demongeot and L. Demetrius, La dérive démographique et la sélection naturelle: Étude empirique de la France (1850-1965), *Population*, 2 (1989), pp. 231-248.
- [37] M.B. Leslie, A matrix model for forest management, *Biometrics*, 25 (1969), pp. 309-315.



## Chapter III

# Demography and Diffusion in epidemics: Malaria and Black Death spread<sup>1</sup>

**Abstract.** The general approach of the classical models of epidemics dynamics is indeed dealing with populations supposed to be constant during the epidemic wave, but the presently observed pandemics show duration of their spread during years imposing to take into account the host and the vector population changes as well as the transient or permanent migration and diffusion of hosts (Susceptible or Infected), as well as vectors and infectious agents. Two examples are presented in this chapter, one concerns the malaria in Mali and the other the plague at the middle-age.

**Keywords and phrases:** Epidemics modelling, Contagious diseases, Endemic state, Black Death, Demographic dynamics, Reaction-diffusion

### III.1 Introduction

Major advances in epidemics modelling have been done recently by introducing demographic aspects (i.e. consideration of host and vector populations whose global size changes during both epidemic and endemic histories) as well as spatial aspects about host, vector or infectious agent spread or genetic change (cf. Gaudart et al., 2010). Mathematical tools corresponding to these improvements have been recently introduced making the classical models more realistic, hence more convenient for prediction and anticipation (like vaccination or other measures of public health limiting the contagion). As examples of application, the dynamics of two infectious diseases will be studied: the malaria endemics in the South of Mali, and a well-known historic plague epidemics, the Black Death (1347-1352), which occurred at the Middle Age and whose demographic and socio-economic consequences were dramatic: about 25 millions deaths in Europe, and 25 millions in Asia (cf. Prentice, 2007).

Despite remaining simple, the models presented in this chapter account qualitatively for the morphology of the endemic spatial distribution and the epidemic front waves. Perspectives will be drawn concerning present epidemic risks, in which a model like those well

---

1. J. Gaudart, M. Ghassani, J. Mintsa, M. Rachdi, J. Waku and J. Demongeot, Demography and Diffusion in epidemics: Malaria and Black Death spread, *Acta Biotheoretica*, 58 (2010), 277-305.

retro-predicting the Black Death episode could be “mutatis mutandis” useful to predict the dynamical behavior of future epidemics.

## III.2 Introduction of demographic dynamics

### III.2.1 Introduction of age classes

By introducing age classes, we add new demographic parameters as the fecundity rate  $f_i$ , equals to the mean number of offsprings a person in class  $i$  is sending in class 1 between times  $t$  and  $t + dt$ , and the survival (resp. death) rate  $b_i$  (resp.  $\mu_i$ ) equals to the probability to survive from the age  $i$  to the age  $i + 1$  (resp. to die at age  $i$ ) between times  $t$  and  $t + dt$ . When the biological age is defined by the physiology of cells and tissues (cf. Demongeot, 2009b) with the possibility to remain in the same age between times  $t$  and  $t + dt$  (despite of the fact that the chronological age is increasing of  $dt$ ), then  $\beta + b_i + \mu_i < 1$ . The equations of the extended Ross-McKendrick model corresponding to 2 age classes are the following:

$$\begin{aligned} \frac{dS_1}{dt} &= (\gamma_1 - \beta_{11}S_1)I_1 - \beta_{12}S_1I_2 - (b_1 + \mu_1)(1 - \beta_{11} - \beta_{12})S_1 + f_1S_2 + f'_1\gamma_2I_2 \\ \frac{dS_2}{dt} &= b_1(1 - \beta_{11} - \beta_{12})S_1 + (\gamma_2 - \beta_{22}S_2)I_2 - \beta_{21}S_2I_1 - \mu_2S_2 \\ \frac{dI_1}{dt} &= (\beta_{11}I_1 + \beta_{12}I_2)S_1 - \gamma_1I_1 + f'_1(1 - \gamma_2)I_2 - (b'_1 + \mu'_1)(1 - \gamma_1)I_1 \\ \frac{dI_2}{dt} &= (\beta_{21}I_1 + \beta_{22}I_2)S_2 - \gamma_2I_2 + b'_1(1 - \gamma_1)I_1 - \mu'_2I_2 \end{aligned} \quad (1)$$

Herein  $\mu'_i$  incorporates the mortality rate due to the disease,  $\beta_{ji}$  is the “efficient contagion rate” of susceptibles  $S_j$  by infectives  $I_i$ ,  $1/\gamma_i$  is the duration of the infective stage,  $f_i^z$  for  $i = 1, 2$  and  $z = S, I$ , denote fertility rates satisfying  $0 \leq f_i^I \leq f_i^S$  and finally  $0 \leq \theta_1, \theta_2 \leq 1$  are the probabilities of vertical transmission. To be more precise in introducing the age classes, in particular with the biological age, we have before to recall the classical models used for modelling the population growth (cf. Doliger, 2006 and Demongeot, 2009b).

### III.2.2 Mathematical properties

Let us assume that the last fecundity parameter  $f_n$  is strictly positive (which is for example the case for both host and vector populations, if the last age class keeps at least one fertile individual), then because the subdiagonal is supposed to be strictly positive,  $L$  (Leslie matrix) and  $U$  (Usher matrix) (cf. Chapter 2) are irreducible and nonnegative. Then, from the Frobenius Theorem's,  $L$  and  $U$  have a strictly positive and a simple dominant eigenvalue  $\lambda$ , with an associated strictly positive eigenvector (stable age pyramid). Notice that the population is constant (resp. in explosion, in extinction) if the Malthusian parameter  $r$  satisfies:  $r = \ln \lambda = 0$  (resp.  $> 0, < 0$ ). For example, let us consider the model with two age classes both for hosts and vectors, whose dynamics is driven by equations (1). Let us denote  $s_i = kS_i$  and  $i_i = kI_i$ . Various possibilities of demographic evolution and stability of the endemic state can be observed depending on the set of values fixed for the model parameters:

- 1)  $f_2^S = 499k/100$ ,  $f_2^I = 0.2k$ ,  $b_1 = 98k/96$ ,  $b'_1 = 4k/5$ ,  $\mu_1 = \mu'_1 = 0$ ,  $\mu_2 = 49k/100$ ,  $\mu'_2 = 4k/5$ ,  $\beta_{11} = 4k^2/100$ ,  $\gamma_1 = k/5$ ,  $\beta_{22} = \beta_{12} = \beta_{21} = 0$ ,  $\gamma_2 = 0$ .

Then, if  $k = 1$ , the differential system ruling the population dynamics is defined by:

$$\begin{aligned}\frac{dS_1}{dt} &= \frac{(20 - 4S_1)}{100}I_1 - \frac{98}{100}S_1 + \frac{499}{100}S_2 \\ \frac{dS_2}{dt} &= \frac{98}{100}S_1 - \frac{49}{100}S_2 \\ \frac{dI_1}{dt} &= \frac{4}{100}I_1S_1 - \frac{1}{5}I_1 - \frac{4}{5}I_1 + 0.2I_2 \\ \frac{dI_2}{dt} &= \frac{4}{5}I_1 - \frac{4}{5}I_2\end{aligned}\tag{2}$$

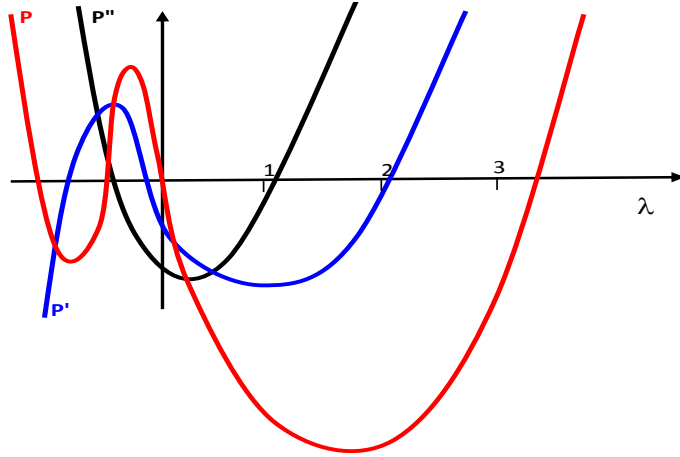


Figure III.1: Graphs of the functions  $P_{B_1}(\lambda)$ ,  $P'_{B_1}(\lambda)$  and  $P''_{B_1}(\lambda)$

The two stationary points are then:

$$(s_1^*, s_2^*, i_1^*, i_2^*) = (0, 0, 0, 0) \text{ and } (s_1^{**}, s_2^{**}, i_1^{**}, i_2^{**}) = (20, 40, 15, 15)$$

and they are both locally unstable (cf. Chapter 1 for the definition). The instability of the first stationary point is easily proved by calculating the Jacobian matrix  $B_1$  of the system (2) and searching for the roots of its characteristic polynomial. For the second stationary point, called the endemic state, we have:

$$B_1 - \lambda I \approx \begin{pmatrix} -1.6 - \lambda & 5 & -0.6 & 0 \\ 1 & -0.5 - \lambda & 0 & 0 \\ 0.8 & 0 & -0.2 - \lambda & 0.2 \\ 0 & 0 & 0.8 & -0.8 - \lambda \end{pmatrix}$$

Its characteristic polynomial  $P_{B_1}$  verifies:

$$\begin{aligned}P_{B_1}(\lambda) &= (-0.8 - \lambda)[(-0.2 - \lambda)[(-0.5 - \lambda)(-1.6 - \lambda) - 5] - 0.48(0.5 + \lambda)] \\ &\quad - 0.16[(-0.5 - \lambda)(-1.6 - \lambda) - 5]\end{aligned}$$

Because  $P_{B_1}(0) > 0$  and  $P_{B_1}(\infty) > 0$ , thus  $\lambda > 0$  and because  $P_{B_1}(1) < 0$ , therefore  $\lambda > 1$ .



We have more:

$$\begin{aligned}
P_{B_1}(\lambda) &= (-0.8 - \lambda) [-\lambda^3 + 1.9\lambda^2 + 4.94\lambda + 0.6] - 0.16\lambda^2 - 0.336\lambda + 0.672 \\
&= \lambda^4 - 1.1\lambda^3 - 3.42\lambda^2 - 4.552\lambda - 0.48 - 0.16\lambda^2 - 0.336\lambda + 0.672 \\
&= \lambda^4 - 1.1\lambda^3 - 3.58\lambda^2 - 5.188\lambda + 0.192
\end{aligned}$$

Thus, the successive derivatives of the characteristic polynomial are given by:

$$\begin{aligned}
P'_{B_1}(\lambda) &= 4\lambda^3 - 3.3\lambda^2 - 7.16\lambda - 5.188 \\
P''_{B_1}(\lambda) &= 12\lambda^2 - 6.6\lambda - 7.16 \\
P'''_{B_1}(\lambda) &= 24\lambda - 6.6
\end{aligned}$$

Therefore, the dominant eigenvalue of  $P_{B_1}$  is strictly positive and consequently, from the unstable endemic state, the susceptibles population grows with a Malthusian parameter greater than 0 (the dominant eigenvalue of the Leslie matrix restricted to susceptibles is greater than 1)(cf. Figure III.1), then  $N$  is not constant and the Ross-McKendrick framework is no more available.

- 2)  $f_2^S = k/2$ ,  $f_2^I = 3k/2$ ,  $b_1 = b'_1 = k/3$ ,  $\mu_1 = \mu'_1 = 0$ ,  $\mu_2 = \mu'_2 = k$ ,  $\beta_{11} = k^2/2$ ,  $\gamma_1 = k/2$ ,  $\beta_{22} = \beta_{12} = \beta_{21} = 0$ ,  $\gamma_2 = 0$ . Then, if  $k = 1$ , the differential system ruling the population dynamics becomes:

$$\begin{aligned}
\frac{dS_1}{dt} &= \frac{(1 - S_1)}{2}I_1 - \frac{1}{3}S_1 + \frac{1}{2}S_2 \\
\frac{dS_2}{dt} &= \frac{1}{3}S_1 - S_2 \\
\frac{dI_1}{dt} &= \frac{1}{2}I_1S_1 - \frac{1}{2}I_1 - \frac{1}{3}I_1 + \frac{3}{2}I_2 \\
\frac{dI_2}{dt} &= \frac{1}{3}I_1 - I_2
\end{aligned} \tag{3}$$

The stationnary point is then:

$$(s_1^*, s_2^*, i_1^*, i_2^*) = (2/3, 2/9, 2/3, 2/9) \tag{4}$$

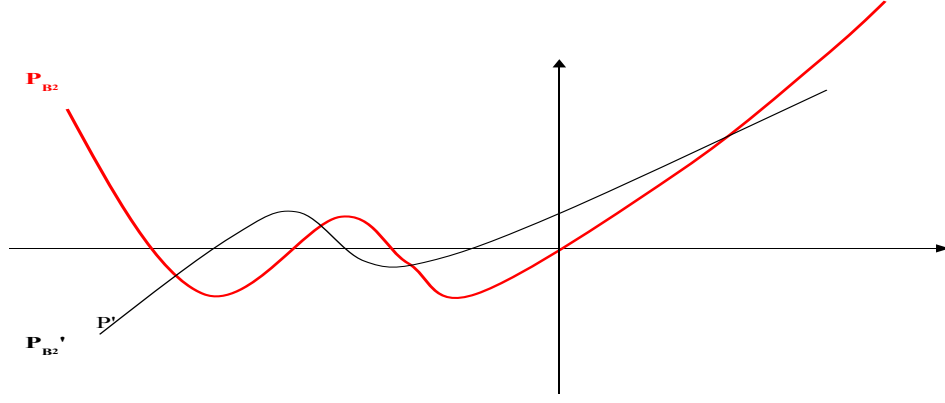
The Jacobian matrix  $B_2$  of the system (3) calculated at the stationary state (4) verifies:

$$6B_2 - 6\lambda I = \begin{pmatrix} -4 - 6\lambda & 3 & 3 & 0 \\ 4 & -6 - 6\lambda & 0 & 0 \\ 2 & 0 & -3 - 6\lambda & 0 \\ 0 & 0 & 2 & -6 - 6\lambda \end{pmatrix}$$

The characteristic polynomial  $P_{B_2}$  of the endemic state of the system (3) verifies:

$$6^4 P_{B_2}(\lambda) = (-6 - 6\lambda) [-216\lambda^3 - 468\lambda^2 - 324\lambda]$$

Then  $P_{B_2}(0) = 0$ ,  $P_{B_2}(\infty) > 0$  and  $P_{B_2}(x) > 0$ , if  $x > 0$  (cf. Figure III.2), ensuring that the dominant eigenvalue  $\lambda$  is equal to zero. Thus the endemic state is locally stable in the Lyapunov sense.

Figure III.2: Graphs of  $P_{B_2}(\lambda)$  and  $P'_{B_2}(\lambda)$ 

- 3)  $f_2^S = 109k/100$ ,  $f_2^I = 0.2k$ ,  $b_1 = 98k/96$ ,  $b'_1 = k$ ,  $\mu_1 = \mu'_1 = 0$ ,  $\mu_2 = 49k/100$ ,  $\mu'_2 = 4k/5$ ,  $\beta_{11} = 4k^2/100$ ,  $\gamma_1 = k/5$ ,  $\beta_{22} = \beta_{12} = \beta_{21} = 0$ ,  $\gamma_2 = 0$ .

Then if  $k = 1$ , we have:

$$\begin{aligned} \frac{dS_1}{dt} &= (20 - 4S_1)I_1/100 - 98S_1/100 + 109S_2/100 \\ \frac{dS_2}{dt} &= 98S_1/100 - 49S_2/100 \\ \frac{dI_1}{dt} &= 4I_1S_1/100 - I_1/5 - 4I_1/5 + 0.2I_2 \\ \frac{dI_2}{dt} &= 4I_1/5 - 4I_2/5 \end{aligned} \quad (5)$$

The characteristic polynomial  $P$  of the Jacobian matrix of the system (5), of the endemic state  $(s_1^{**}, s_2^{**}, i_1^{**}, i_2^{**}) = (20, 40, 2, 2)$ , satisfies:

$$\begin{aligned} P(\lambda) &= (-0.8 - \lambda)[(-1.8 - \lambda)[(-0.5 - \lambda)(-1 - \lambda) - 1] - 0.05(0.5 + \lambda)] \\ &\quad - 0.16[(-0.5 - \lambda)(-1 - \lambda) - 1] \end{aligned}$$

Then  $P(0) < 0$  and  $P(\infty) > 0$ , ensuring that the dominant eigenvalue  $\lambda$  is strictly positive, and  $P(1) > 0$ , then  $\lambda < 1$ . The endemic state is unstable and all populations are in extinction.

- 4)  $f_2^S = k$ ,  $f_2^I = 3k/4$ ,  $b_1 = b'_1 = 3k/4$ ,  $\mu_1 = \mu'_1 = 0$ ,  $\mu_2 = \mu'_2 = k$ ,  $\beta_{11} = k^2/4$ ,  $\gamma_1 = k/4$ ,  $\beta_{22} = \beta_{12} = \beta_{21} = 0$ ,  $\gamma_2 = 0$ .

Then if  $k = 1$ , we have:

$$\begin{aligned} \frac{dS_1}{dt} &= (1 - S_1)I_1/4 - 9S_1/16 + S_2 \\ \frac{dS_2}{dt} &= 9S_1/16 - S_2 \\ \frac{dI_1}{dt} &= I_1S_1/4 - I_1/4 - 9I_1/16 + 3I_2/4 \\ \frac{dI_2}{dt} &= 9I_1/16 - I_2 \end{aligned} \quad (6)$$

The characteristic polynomial  $P$  of the Jacobian matrix of the system (6), of the endemic state  $(s_1^{**}, s_2^{**}, i_1^{**}, i_2^{**}) = (1/4, 9/64, 0, 0)$ , satisfies:

$$16^4 P(\lambda) = (-16 - 16\lambda)[(-12 - 16\lambda)[(-16 - 16\lambda)(-9 - 16\lambda) - 144]] - 108[(-9 - 16\lambda)(-16 - 16\lambda) - 144]$$

Then  $P(0) = 0$ ,  $P(\infty) > 0$  and  $P(x) > 0$ , if  $x > 0$ , ensuring that the dominant eigenvalue  $\lambda$  is equal to 0. All populations are locally stable only in the Lyapunov sense, but are asymptotically unstable.

### III.3 Introduction of a spatial dynamics

The introduction of the space parameter in the Ross-McKendrick models can be made through stochastic spatial Markovian or renewal models (cf. Demongeot and Fricot, 1986), or deterministic Partial Differential Equations (PDE) in which the diffusion of hosts or vectors is modelled by the Laplacian operator  $\Delta$  or possibly the d'Alembertian  $\square$ , when some subpopulations can present an accelerated ageing (cf. Demongeot, 2009b). These models are called SIGR with Diffusion (SIGRD) (cf. de Magny et al., 2005). The Bankoumana model is a double SIGRD model (cf. Gaudart et al., 2007, 2009 and 2010) whose PDE equations have spatial initial conditions essentially determined by the spawning zones of mosquitos in backwater places.

These zones are depending on the rainfall, e.g. the spawning places of *Anopheles gambiae*—the malaria vector—are located on backwater perimeter, whose length is equal to 0 in absence of rain (stable dry season), tends to infinity when backwater is progressively fulfilled by water (fractal transient phase during the season transition) and diminishes until  $2\pi R$ , where  $R$  is the radius of the final backwater mare (stable rainy season). During the susceptible host and infective *Anopheles* spread, the maximum of contagion is observed on the common zones of least diffusion of both hosts and vectors, which can be asymptotically confounded: as for the morphogens interaction in morphogenesis, this common zero-diffusion domain allows a maximum of contagious contacts between interacting species (cf. Abbas et al., 2009).

During the stable rainy season, taking into account the diffusion of all vector subpopulations  $A_s, A_g$  and  $A_i$  (*Anopheles* susceptible, infected/non infective and infective) until the human subpopulations  $S, G, I$  and  $R$  (susceptible, infective, infected/non infective and recovered), it is possible to simulate the model and compare its numerical results to the data recorded on the ground, showing a good fit. In order to improve this fit, contagion parameters are chosen depending on space, e.g. maximum in zones where diffusion of infective vectors and hosts (whose concentration is respectively  $A_i$  and  $G$ ) is minimum and in zones where concentration of susceptibles ( $A_s$  and  $S$ ) is maximum ensuring locally a large coexistence time, hence a high contagion rate between large interacting subpopulations (cf. Dutertre, 1976).

In case of isotropic diffusion, the zero Laplacian (or zero curvature or maximal gradient) lines of the concentration surfaces of the concerned populations becomes, if they intersect, a contagion frontier, where hosts, vectors and infectious agents interact. These lines correspond to regions where the mean Gaussian curvature of the concentration surfaces  $S$  and  $A_i$  vanishes, these regions being defined respectively by:

$$\frac{\partial^2 S}{\partial x^2} \frac{\partial^2 S}{\partial y^2} - \left( \frac{\partial^2 S}{\partial x \partial y} \right)^2 = 0 \text{ and } \frac{\partial^2 A_i}{\partial x^2} \frac{\partial^2 A_i}{\partial y^2} - \left( \frac{\partial^2 A_i}{\partial x \partial y} \right)^2 = 0$$

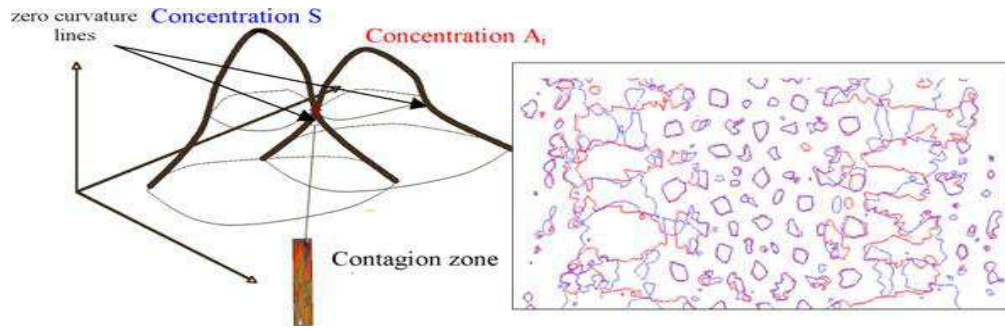


Figure III.3: Representation of the co-evolution of the zero-diffusion domains for interacting species  $S$  (in blue color) and  $A_i$  (in red color) in the case of isotropic diffusion (on the left). Asymptotic co-existence of  $S$  and  $A_i$  on their common least diffusion domain (on the right)

Figure III.3 shows the possibility of such an intersection on only one tangency point or two intersection points (on the left) and on whole zero-diffusion curves asymptotically confounded (on the right) for a convenient value of the ratio between the diffusion coefficients  $D_S/D_{A_i}$  (cf. Abbas et al., 2009).

### III.4 Introduction of saturation kinetics and genetic drifts

As noticed in Horie et al., (2010), the viral genome is easily mutating and transferring new genes to both hosts and vectors, these latter being often also hosts and rarely neutral healthy carriers. The vast majority of these new genes apparently do nothing, but some still produce working proteins or contribute to code for small regulatory RNAs (siRNAs or miRNAs) or for RNA-binding oligo-peptides, important traduction factors. It is yet impossible to know what these RNAs, peptides or proteins exactly do. But our ancestors have domesticated their viral interlopers to act as partners of our cells (cf. Figure III.4). In the example discovered by Horie et al., (2010), bornaviruses have clearly taken part to the evolution of mammals.

Taking into account, in the models, of viral epidemics these genetic transfer could allow to explain the apparition of resistances both in hosts and vectors diminishing their ability to build viral proteins, and also on the viral side, could render explicit certain strategies for escaping the host immunologic defences (cf. Demongeot et al., 2009; Thuderoz et al., 2010). A way to incorporate this triple win game (wins for hosts, vectors and also infectious agents which have survived and coexist together during the evolution) consists in rendering dependent the contagion, fecundity, longevity and death parameters of both hosts and vectors on the level of contact represented by the term  $\beta SA_i$ . This dependence is supposed to decrease  $\beta, \sigma, \gamma$  and  $\mu$ , and at the same time increase  $f$ : indeed, the largest is the contact number, the most adapted in terms of low susceptibility, high fecundity and longevity are the population of hosts and vectors offering to the infectious agent numerous targets to survive, and to hosts and vectors a way to evolve rapidly in order to increase their adaptive power. The decreasing functions could be linear between two thresholds, the upper corresponding to the situation of a new infectious agent whose virulence is maximal, and the lower to endemics resulting from a long cohabitation between the infectious disease actors. The advantage for vectors would be for example clear, if the host disease causes also a disease in the vector, because evolved vectors with resistance would be healthier and would have an adaptive advantage.

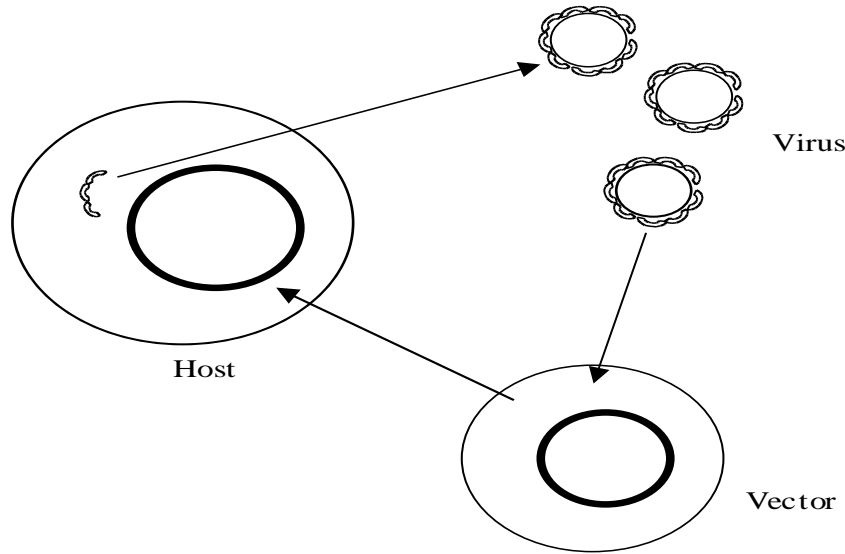


Figure III.4: The triple win game in which hosts and vectors use the viral genome for making evolve their own genomes, and the virus survives thanks to these later, which code for its proteins

The strategy of infectious agents would be then to evolve around the vector's and host's defences, circumventing and overcoming them (cf. Baum et al., 2004). The observation in model simulations of a periodic time evolution of the parameter values with the possibility to randomly reset them at their upper threshold values (representing the mutation of an ancient virus or the occurrence of a new one) could render the model more realistic hence more adapted to simulated scenarios for testing public health policies and anticipating real epidemics or pandemics.

## III.5 Black Death

### III.5.1 Introduction

Plague was considered as endemic in the steppes of Southern-Russia where Mongols originated (cf. Zhang et al., 2008). Born in the Caspian sea area (probably triggered by contacts between Mongolian and Genoa sailors and warriors in wars around 1346), the European epidemic wave went through the mean of Mediterranean routes (cf. Figure III.5). It reached ports like Marseilles in France and Genoa in Italy at the end of the year 1347. During 5 years it was spread widely in Europe from these two large commercial cities and come back to the Caspian reservoir. A simple Susceptibles-Infectives model with Diffusion (SID) explains the essential of the observed front wave dynamics during years between 1348 and 1350. The model uses only 3 coefficients: (1) a local viscosity proportional to the altitude, (2) a contagion parameter and (3) a death/recovering parameter (representing the future of infecteds as dead or immunized after being cured of the plague).

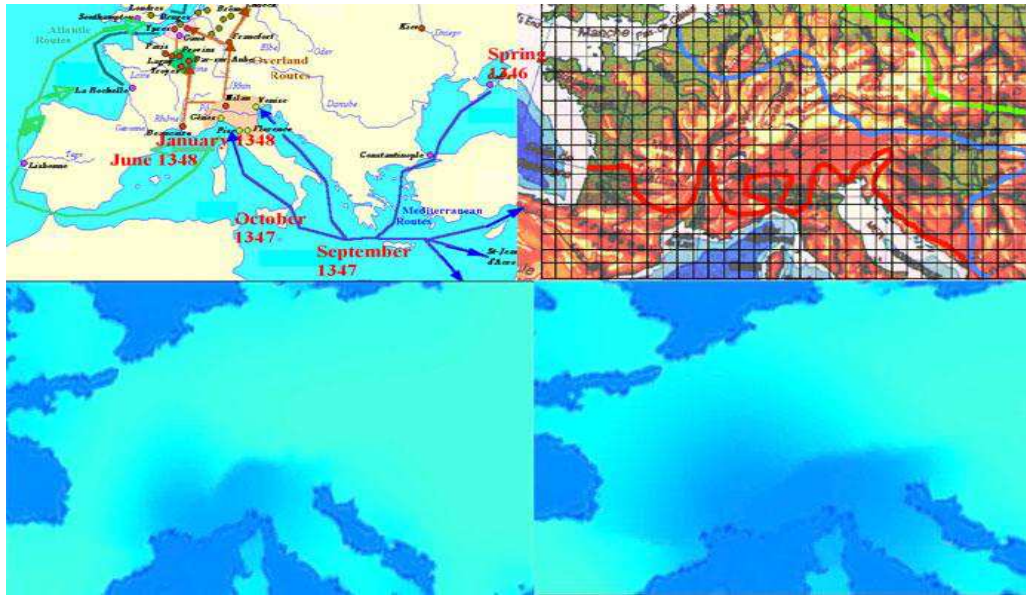


Figure III.5: *Top left:* The spread of bubonic plague following sea and overland routes (cf. [http : //www.cosmovisions.com/ChronoPestesMA02.htm](http://www.cosmovisions.com/ChronoPestesMA02.htm)). *Top right:* Observed wave fronts after 1 (in red color), 2 (in blue color) and 3 (in green color) years of spread from the 2 initial entry ports Marseilles and Genoa (cf. Mocellin-Spicuzza, 2002); the black grid corresponds to the collected altitudes. *Bottom:* Simulation of the wave front after 3 (on the left) and 6 (on the right) months from the 2 initial entry ports Marseilles and Genoa

### III.5.2 The raw data

Data are coming from numerous different sources, like parish, bishop, monastery and hospital registers, abbey cartularies, town council registers, e.g. *riformagioni* in Italy (cf. Carpentier, 1993), cemeteries,... For example, a part of data comes from a monastic order, the Hospitaller Order of St Anthony founded at La Motte (presently Saint Antoine) in the Dauphiné (France) in 1095 near Grenoble by Gaston du Dauphiné, whose son was struck by a fungal disease, known in the Middle Ages as Saint Anthony's fire (ergotism), caused by a transformation of the grain (often rye) into enlarged, hard, brown to black spur-like structures that constitute the source of the drug ergot in flour and causes convulsions often leading to death. The members of this Order were specialised in curing patients suffering from this disease. The Order was approved by Pope Urban II during the Council of Clermont in 1095. Later in 1218, Pope Honorius III permitted the brothers to take the vows of obedience, poverty and chastity. In the thirteenth century the Order spread over the whole of Western Europe with about 370 hospitals in the fourteenth century, able to treat about 4000 patients. This organization permitted to the order to receive about 1500 patients suffering of the plague and since 1339 has been in relationship with the University of Grenoble under the Dauphin Humbert.

The origine of the Black Death epidemics is uncertain (cf. Wheelis, 2002). Wars between Mongols and Chinese contributed to its dissemination in Asia. In 1334, in the North-Eastern Chinese province of Hopei, the plague was particularly virulent and killed about 90% of the population - some 5 millions people. Then it went in Europe from east, striking Caucasia and Crimea (cf. Wheelis, 2002). In 1346, Tatars attacked the port of Caffa, presently in Ukraine but belonging at this time to Genoa. After an agreement between Genoa and Tatars, the

conflict ceased and ships from Caffa transmitted the disease in each ports at which they stop. Hence in 1347, the Black Death arrived first in Constantinople, then in the Mediterranean trade cities: Messine in Sicily, and after Genoa (where commercial boats were sent back for a time) and Marseilles (where boats have been accepted for commercial reasons) at the start of the year 1348.

The diffusion of the plague is probably due to rat infestations and abundant fleas in trade ships, transmitting plague to city rat populations (cf. Wheelis 2002). From bubonic plague, the outbreak continuation appears to have been mainly due to the direct pneumonic transmission. From Marseilles, plague devastated Provence reaching Avignon - 100 km far from Marseilles - in 1 month, respecting the estimation of the front speed given in Murray (2002) and went through the Rhône valley until Paris. Some says the maximal velocity was 75 km a day, i.e. 87 cm/s, which is notably larger than the estimation of 5 cm/s made in Murray (2002). This maximal velocity probably occurred only in zones with diffusion maximum, i.e. with viscosity minimum, like the Rhône Valley (maximal human density and commercial transactions). During the next 3 years of the epidemics, it spreads northwards, reaching Norway and crossing to England and from there to Scotland, Ireland, Iceland and Greenland. Mortality of the pandemic was terrible: at least 25 millions people, that is 25-75% of the European population (cf. Russell, 1948) are estimated to have died, e.g. at Givry in Burgundy for about 1500 inhabitants, the parish register shows 649 funerals in 1348, whose 630 from June to September. The parish having normally only 40 funerals the year, the specific mortality rate due to plague was equal to  $(649-40)/1500 = 40,6\%$ . Another example is a sample of 235 deaths from the bishop's registers of Coventry and Lichfield, the only English register to list both date of death and date of institution, showing that the Black Death swept through very rapidly to local areas (cf. Wood, 2003).

The influence of climate on the outbreak is controversial. It is likely that a harsh climate, combined with the poverty, population and war (the Hundred Years' War), has been an important risk factor (cf. Zhang et al., 2008). Furthermore, emerging from the dark ages, overland and sea trade routes have been developed and the population density increased in the cities, favouring epidemics. It is now believed that bubonic plague (*Yersinia pestis*) is the infectious agent of the Black Death (cf. Raoult et al., 2000). Sometime in the past, *Yersinia pestis* lost a set of genes expressed as adhesins, binding the bacteria to intestinal crypts (cf. Orent, 2001 and 2004). Now, by suppressing signals between immune cells, plague spreads through the lymphatic system, invading organs such as spleen, lungs, and especially the liver. Bubonic plague is transmitted indirectly (mainly by flies), has an incubation period of 2-6 days and a mortality rate between 50 and 90% (if untreated). Pulmonary plague can be secondary to a bubonic plague or primary after direct contamination. Highly contagious, the primary pulmonary plague occurs from an aerial contagion (direct by respiratory droplets) and if not treated the disease is fatal in most cases. Its Incubation period is between 2 or 4 days, with  $R_0$  estimated varying between 0.8 and 3, with mean 1.3 (cf. Gani and Leach, 2004).

Knowledge about the demographic dynamics needs data about the population growth in middle-age cities (cf. Renouard, 1948; Russel, 1972; Mocellin and Experton, 1992; Brossolet and Mollaret, 1994; Horrox, 1994; Mischlewski, 1995; Drancourt et al., 1998; Eckert, 2000; Cantor, 2001; Wood, 2003; Mocellin-Spicuzza G., 2002; Cohn, 2002; Scott and Duncan, 2004; Benedictow, 2004; Christakos et al., 2005; Kelly, 2005; Barry and Gualde, 2006) like Florence in Italy, whose population passed from about 100000 inhabitants in 1338 - 90000 in 1336 for Villani (2001) - to 50000 in 1351. Parallely, during this period of time, between 60 to 70% of Hamburg's and Bremen's population died and in Provence, Dauphiné or Normandy, historians observed a decrease of 60% of fiscal hearths in French cities of these regions

(cf. <http://www.answers.com/topic/> and <http://www.io.com/~john/demog.htm>). In some regions, two thirds of the population were annihilated. About half of Perpignan's population died in several months (only two of the eight physicians survived the plague). England lost 70% of its population, which passed from 7 millions to 2 millions in 1400. Big European cities ranged from 12000 to 100000 peoples, with some exceptional cities exceeding this scale. Some historical examples before Black-Death included London (25000-40000), Paris (50000-80000), Genoa (75000-100000) and Venice (100000). Moscow in the fifteenth century had only a population in excess of 200000! No complete population censuses were taken until the eighteenth century, thus estimates of population levels are notoriously unreliable. Estimated levels vary as a number of "multiplier" factors which often have to be taken into account: estimated population density, ages of marriage, and perhaps most importantly the number of people denoted by a "hearth" in medieval tax surveys that do provide hard numbers.

### III.5.3 The Model

The Fisher equation (cf. Fisher, 1937 and Murray, 2002) has been firstly used for representing the evolution of the host and vector sub-populations during the spread of the Black Death:

$$\frac{\partial S}{\partial t} = \lambda S \left(1 - \frac{S}{S_0}\right) + k \frac{\partial^2 S}{\partial x^2} \quad (7)$$

where  $\lambda$  in the logistic term is the growth rate at the homogeneous limit ( $S$  independent of the space variable  $x$ ),  $S_0$  is a saturation size and  $k$  is the diffusion coefficient (the inverse of the viscosity). Equation (7) admits propagating wave solutions of the form  $S(x, t) = S(x - \nu_{front} t)$ , with  $\nu_{front} \approx 2(k\lambda)^{1/2}$  as speed of the front (cf. Murray, 2002). The case of a heterogeneous medium is treated in Mendez et al., (2003). Murray quoted a diffusion coefficient  $k$  of about  $10^3 \text{ m}^2/\text{s}$  and a reaction (growth) rate of  $15 \text{ year}^{-1}$ , corresponding to  $\lambda \sim 5 \cdot 10^{-7} \text{ s}^{-1}$  and giving an estimation for  $\nu_{front}$  of about  $5 \text{ cm/s}$ , i.e. about  $1500 \text{ km/year}$  (cf. Brandenburg and Multamaki, 2004).

The model used in this chapter for modelling the Black-Death spread is a SIRD model as in the Bankoumana study (cf. Gaudart et al., 2007, 2009 and 2010), but without vector terms and has for its reaction term the form of a Lotka-Volterra Ordinary Differential Equation (ODE) of dimension 3, plus a diffusion term:

$$\begin{aligned} \frac{dS}{dt} &= \varepsilon \Delta S - \beta SI \\ \frac{dI}{dt} &= \varepsilon \Delta I + \beta SI - \gamma R \\ \frac{dR}{dt} &= \varepsilon \Delta R + \gamma R \end{aligned} \quad (8)$$

where the  $\beta SI$  term comes from the "law of mass action", assuming homogeneous mixing between susceptibles and infecteds,  $\beta$  is the rate of transition from the susceptible to the infected state, calculated per infected and per susceptible,  $\gamma$  is the rate of transition from infected to post-infected state (e.g., death or immunity) per infected person and  $\varepsilon$  is the diffusion coefficient. By taking the viscosity (the inverse of  $\varepsilon$ ) proportional to the altitude, the simulated front waves are more similar to the observed ones (cf. Figure III.5) than in the previous simulations (cf. Murray, 2002). The initial population size of susceptibles in the main middle age cities has been fixed following the demographic data. The results of simulations (cf. Figure III.5 on the bottom) are in agreement with the data observed in the 370 hospitals of the order of St Anthony (cf. Figure III.5 on the top right). Improvements



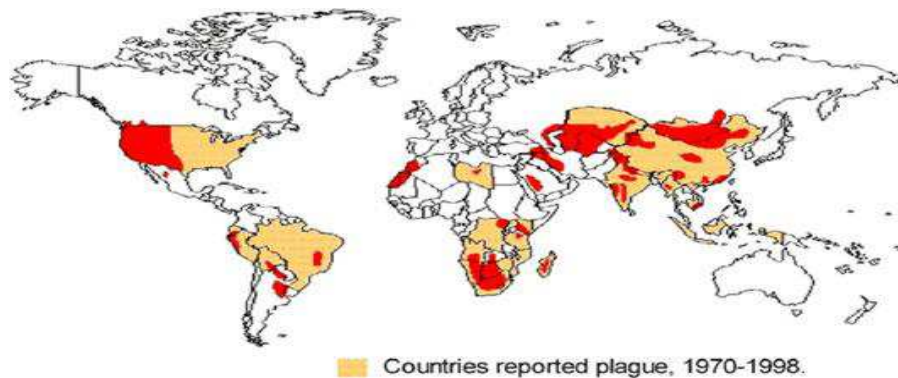


Figure III.6: World distribution of plague in 1998 (cf. WHO, 1999)

could come from considering multiple entrance points (ports like Barcelona reached in June 1348 or La Rochelle, Rouen and Dover reached later in 1348), and taking into account all the commercial sea (Mediterranean and Atlantic) and overland routes (cf. Figure III.5 on the top left) as well as the demography (fecundity and natural mortality, as well as more sophisticated notions as demographic potential and Hamiltonian demographic energies (cf. Maupertuis (1745); reed (1965); Thom (1972); Demongeot and Demetrius (1989); Porte (1994); Demongeot et al. (2007a and b); Forest et al. (2007); Glade et al. (2007)).

The present endemic state (cf. Figure III.6) could be explained by a new model taking into account the air routes (cf. *La peste humaine* (1997); WHO (1999)). Plague is still important because it could be considered as a re-emerging disease (cf. Stenseth et al., 2008): *Yersinia pestis* still causes several thousand human cases per years and about hundred human deaths are reported each year. Plague is present in all continents, as human and enzootic disease, particularly in Africa, North and South America, and Asia (cf. Prentice and Rahalison, 2007). Democratic Republic of the Congo and Madagascar are particular places, accounting most of the reported cases (cf. Neerinckx et al., 2008). Most of the present cases correspond to bubonic plague, but outbreaks of pneumonic plague still occur. Environmental, geographical and social characteristics are particularly favourable for a broad diffusion of plague in Africa (cf. Neerinckx et al., 2008), in spite of the focal nature of the transmission. *Yersinia pestis* is also an attractive agent for bioterrorism (cf. Wheelis (2002); Prentice and Rahalison (2007); Zhang et al. (2008)). Furthermore, climate change might modify the dynamics of plague transmission and cause outbreaks in endemic regions but also in non-endemic regions (cf. Raoult et al. (2000); Stenseth et al. (2008)). Crisis-management approach is considered as insufficient (cf. Orent (2001 and 2004)) and prevention action would be more effective. An efficient prediction from simulations of a realistic model taking into account the new aerial routes with minimal viscosity (cf. Khan et al., 2009) could serve this cause.

## III.6 The Malaria in Mali

### III.6.1 Introduction

The malaria is a parasitic infectious disease whose agent belongs to the genus *Plasmodium* (essentially *P. falciparum*). Malaria is carried by the mosquitos of the genus *Anopheles* and

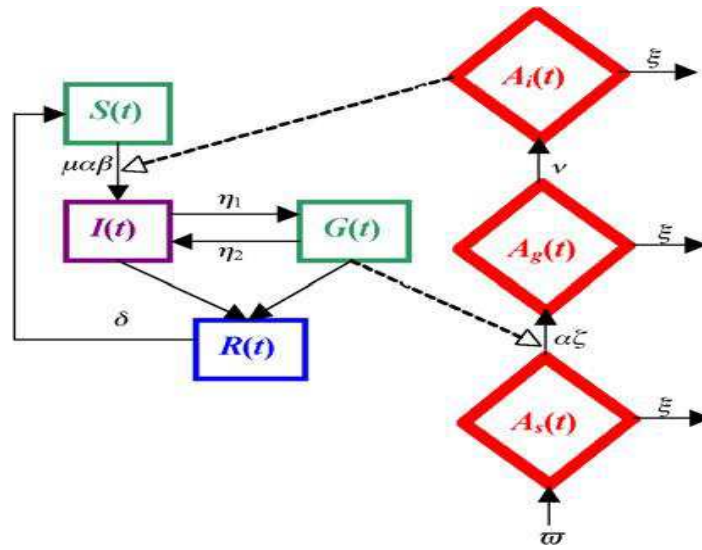


Figure III.7: Interaction graph (in top), and SIGR Diffusion (SIGRD) Equation system (9) both for hosts and for vectors (in bottom) for the Bankoumana model

the vector in Mali is *Anopheles funestus* or *Anopheles gambiae* (cf. Depinay et al. (2004); Huang et al. (2006); Gaudart et al. (2009)). Symptoms of malaria include fever, headache and vomiting, and usually appear between 10 and 15 days after the mosquito bite. Untreated, malaria becomes life-threatening by disrupting the blood supply to vital organs. In southwest of Mali in particular in the region near the river Niger each rainy season triggers annual malaria epidemic. The WHO's statistics say malaria accounts for 17% of child deaths in Mali. One in five Malian children die before their fifth birthday. Of the one to three million people worldwide who die of malaria every year, 90% live in sub-Saharan Africa. Malaria kills an African child every 30 s, according to WHO. Of those, several hundred thousand live in the Sahel region of West Africa, which encompasses Mali, Mauritania, Niger, Burkina Faso, Chad and parts of Senegal, Togo, Benin and Nigeria. The Bankoumana village is a locality of Sudanese savannah area in Mali in which the disease has been carefully studied and recorded since 15 years. At each evaluation on the ground (each 2 months during the rainy season and each 3 months during the dry one) a blood sample is collected on each child of the village and the parasitemia is studied for *Plasmodia falciparum*, *malariae* and *ovale*, as well as the gametocytemia (for *P. falciparum*), with Giemsa technique (Dumbo 2005).

### III.6.2 The Model

The model has been drawn in order to take into account the known mechanism of the disease and to qualitatively fit the empirical observations. The equations are given in Figure III.7, without age classes for host nor for vector, but with diffusion for vector (supposed to be much larger than the host diffusion). The contagion parameters  $\alpha\beta$  and  $\alpha\zeta$  can be chosen depending on space, in particular  $\alpha$ , the mean bite number per mosquito and per night. During the stable rainy season, taking into account the diffusion of all vector subpopulations

$A_s$ ,  $A_g$  and  $A_i$  (*Anopheles* susceptible, infected/non infective and infective) until the human subpopulations  $S$ ,  $G$ ,  $I$  and  $R$  (susceptible, infective, infected/non infective and recovered), it is possible to simulate the model and compare its numerical results to the data recorded on the ground, showing a good fit. For improving the fit, contagion parameters have been chosen depending on space, e.g. fixed at a value maximum in zones where diffusion of infective vectors and hosts ( $A_i$  and  $G$ ) is minimum, and in zones where the concentration of susceptibles ( $A_s$  and  $S$ ) is maximum.

$$\begin{aligned}
\frac{dS}{dt} &= -\mu\alpha\beta SA_i + \delta R \\
\frac{dI}{dt} &= \mu\alpha\beta SA_i - (\eta_1 + \gamma)I + \eta_2 G \\
\frac{dG}{dt} &= \eta_1 I - (\eta_2 + \gamma)G \\
\frac{dR}{dt} &= \gamma(I + G) - \delta R \\
\frac{\partial A_s}{\partial t} &= D_s \Delta A_s + \varpi - \alpha\varsigma GA_s - \xi A_s \\
\frac{\partial A_g}{\partial t} &= D_g \Delta A_g + \alpha\varsigma GA_s - (\xi + \nu)A_g \\
\frac{\partial A_i}{\partial t} &= D_i \Delta A_i - \xi A_i + \nu A_g
\end{aligned} \tag{9}$$

The variables of the Bankoumana model are:

$S(t)$ : size of the sub-population of Susceptible hosts at time  $t$

$I(t)$ : size of the sub-population of Infected not Infective hosts (positive parasitemia and negative gametocytemia) at time  $t$

$G(t)$ : size of the sub-population of infective hosts by Gametocytes hosts (positive gametocytemia) at time  $t$

$R(t)$ : size of the sub-population of Resistant hosts, i.e. treated and resistant to the disease, or immunized, died or displaced at time  $t$

$A_s(t)$ : size of the sub-population of Susceptible *Anopheles* at time  $t$

$A_g(t)$ : size of the sub-population of infected (but not infective) *Anopheles* by Gametocytes at time  $t$

$A_i(t)$ : size of the sub-population of Infective *Anopheles* at time  $t$

$N(t)$ : total size of hosts at time  $t$

$M(t)$ : total size of *Anopheles* at time  $t$

Then, parameters of the Bankoumana model are:

$\delta$ : rate of immunization loss in host ( $1/\delta$  is the mean duration of the resistance)

$\eta_1$ : rate of gametocytes occurrence in host ( $1/\eta_1$  is the mean duration of the time interval between the primo-infection and the first appearance of gametocytes in an infected individual)

$\eta_2$ : rate of gametocytes loss in host

$\gamma$ : rate of resistance occurrence

$\mu$ : Anophelian density, i.e. *Anopheles* number per host

$\alpha$ : mosquito bite rate per mosquito and per night ( $\mu\alpha$  is called the vector aggressivity)

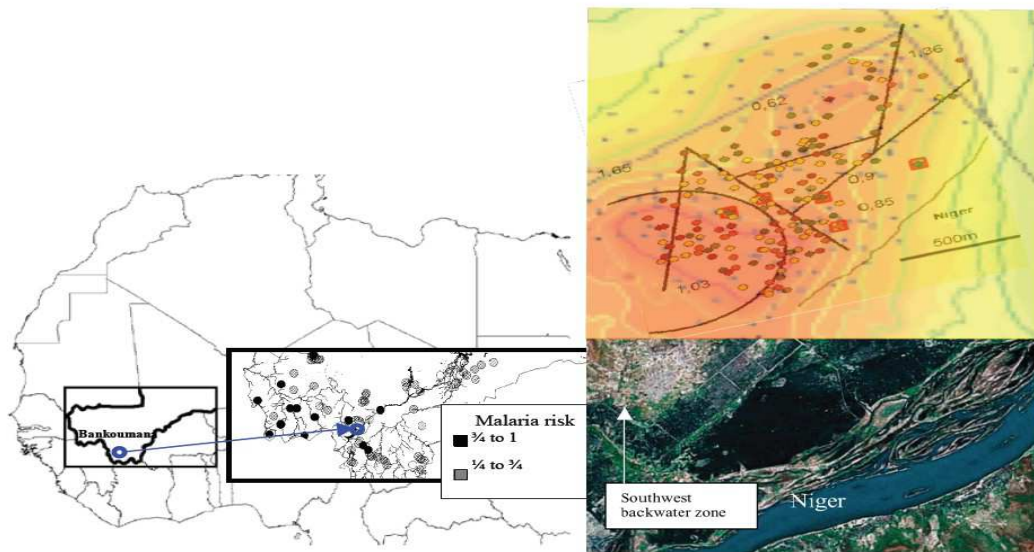


Figure III.8: Results of the SGIRD simulated (*on the top-right*) and real data (*on the middle-right*) showing a good fit along a gradient parallel to Niger river from the southwest backwater zone (*on bottom*) to the village of Bankoumana (Mali) for the Bankoumana model

In the model, a Susceptible can become Plasmodic Infected (non Infective). A Plasmodic Infected can shift to the Gametocytic state, or recover and acquire an immunization, or recover without immunization, i.e. become a new susceptible. The immunized state can also naturally disappear (e.g. due to an intercurrent disease paralysing the immune system). The introduction of the space in the model could be done by using stochastic spatial Markovian or renewal models (cf. Demongeot and Fricot, 1986) or deterministic Partial Differential Equations (PDE). Such models are of SIGRD type (cf. de Magny et al., 2005).

The Bankoumana model (cf. Figure III.7 on bottom, system of Eq. (9)) is a double SIGRD model (cf. Gaudart et al., 2007 and 2009) whose interaction graph (cf. Figure III.7 on the top) corresponds to PDE equations given in (9) with spatial initial conditions essentially determined by the spawning zones in backwater places (cf. Figure III.6 on the bottom). These zones are depending on the rainfall hence have a seasonal occurrence (cf. Balenghien et al. (2006); Bicout et al. (2002); Bicout and Sabatier (2004); Ndiaye et al. (2006); Porphyre et al. (2004)): the spawning places of *Anopheles gambiae*—one of the malaria vectors—are located on the backwater perimeter, whose length is equal to 0 in absence of rain (stable dry season), tends to infinity when backwater holes/ponds are progressively fulfilled by water (fractal transient phase during the season of transition) and diminishes until  $2\pi R$ , where  $R$  is the radius of the final backwater hole (stable rainy season). During the stable rainy season, taking into account the diffusion of all vector subpopulations  $A_s$ ,  $A_g$  and  $A_i$  (*Anopheles* susceptible, infected/non infective and infective) until the human subpopulations  $S, G, I$  and  $R$  (susceptible, infective, infected/non infective and recovered) supposed to be fixed, we can simulate and compare the numerical results to the data recorded on the ground, showing a good fit (cf. Figure III.8). For improving this fit, contagion parameters  $\beta$  and  $\zeta$  can be chosen depending on space, e.g. maximum in zones, which constitute overlaps

between domains where diffusion of infective vectors and hosts ( $A_i$  and  $G$ ) is minimum and domains where concentration of susceptible ( $A_s$  and  $S$ ) is maximum, ensuring locally a large coexistence time, hence a high contagion rate between these interacting subpopulations (cf. Abbas et al., 2009).

If we simplify the malaria model in (9), by considering only the population of infected and contagious vectors, we can introduce in the equations a delay  $\theta$  in order to take into account the long incubation period both in host and vector, due in particular to host and vector migrations and climatic changes (which explains in part that malaria is reappearing in south of Europe). In this new model, the variables are denoted as follows:

$S(t)$  represents the size of the susceptible host population at time  $t$

$I(t)$  represents the size of the infected but not contagious host population at time  $t$

$G(t)$  represents the size of the infected and contagious host population at time  $t$

$R(t)$  represents the size of the resistant host population at time  $t$

$A_i(t)$  represents the size of the infected and contagious vector population at time  $t$

$VI(t)$  represents the Normalized Difference Vegetation Index (NDVI), i.e. a simple numerical indicator coming from remote sensing measurements assessing whether an observed zone contains live green vegetation or not.

The parameters of the new model are defined in the Table III.1.

The transition from susceptible to infected state depends on host (resp. vector) population size, but also on climatic factors represented by the variable  $i(t)$  (resp.  $i_m(t)$ ) in the following equations:

$$\begin{aligned}
\frac{dS(t)}{dt} &= -i(t)S(t) + \delta R(t) \\
\frac{dI(t)}{dt} &= i(t)S(t) - (\eta_1 + \gamma)I(t) + \eta_2 G(t) \\
\frac{dG(t)}{dt} &= \eta_1 I(t) - (\eta_2 + \gamma)G(t) \\
\frac{dR(t)}{dt} &= \gamma[I(t) + G(t)] - \delta R(t) \\
\frac{dA_i(t)}{dt} &= i_m(t) \left[ \exp \left\{ \frac{-\xi v}{1 + \chi_{\{VI(t-\theta) \geq \tau\}} VI(t-\theta)} \right\} - A_i(t) \right] - \frac{\xi}{1 + \chi_{\{VI(t-\theta) \geq \tau\}} VI(t-\theta)} A_i(t)
\end{aligned} \tag{10}$$

with:

$$\begin{cases} i(t) &= \mu \alpha \beta \chi_{\{VI(t-\theta) \geq \tau\}} VI(t-\theta) A_i(t) \\ i_m(t) &= \alpha \varsigma (\chi_{\{VI(t-\theta) \geq \tau\}} VI(t-\theta)) G(t) \end{cases}$$

By denoting:

$$\begin{cases} a &= \mu \alpha \beta [\chi_{\{VI(t-\theta) \geq \tau\}} VI(t-\theta)] \\ b &= \alpha \varsigma [\chi_{\{VI(t-\theta) \geq \tau\}} VI(t-\theta)] \end{cases} \quad \text{and} \quad \Delta = \frac{-\xi}{1 + \chi_{\{VI(t-\theta) \geq \tau\}} VI(t-\theta)}$$

Parameters	Definition
$\alpha$	Mosquito bite rate per mosquito and per night ( $\mu\alpha$ is called the vector agressivity)
$\beta$	Probability for a susceptible human of becoming infected after a single bite
$\gamma$	Rate of resistance occurrence
$\delta$	Rate of immunization loss in host ( $1/\delta$ is the mean duration of the resistance)
$\zeta$	Probability for a susceptible <i>Anopheles</i> of becoming infected after a single bite on an infected human
$\eta_1$	Rate of gametocytes occurrence in host ( $1/\eta_1$ is the mean duration of the time interval between the primo-infection and the first appearance of gametocytes in an infected individual)
$\eta_2$	Rate of gametocytes loss in host
$\theta$	Latency period for the normalized vegetation index
$\mu$	Anophelian density, i.e. <i>Anopheles</i> number per host
$v$	Average duration of the gonotrophic cycle
$\xi$	Mortality rate of the susceptible <i>Anopheles</i>
$\tau$	NDVI lowest threshold value conditioning the <i>Anopheles</i> behavior

Table III.1: Definitions of the parameters in the Bankoumana model

the previous equations become the following system:

$$\begin{aligned}
\frac{dS(t)}{dt} &= -aA_i(t)S(t) + \delta R(t) \\
\frac{dI(t)}{dt} &= aA_i(t)S(t) - (\eta_1 + \gamma)I(t) + \eta_2 G(t) \\
\frac{dG(t)}{dt} &= \eta_1 I(t) - (\eta_2 + \gamma)G(t) \\
\frac{dR(t)}{dt} &= \gamma[I(t) + G(t)] - \delta R(t) \\
\frac{dA_i(t)}{dt} &= bG(t)[\exp(\Delta v) - A_i(t)] + \Delta A_i(t)
\end{aligned} \tag{11}$$

The two stationary states of (11) are the healthy state  $E_0 = (S_0, 0, 0, 0, 0)$  and the endemic state  $E^* = (S^*, I^*, G^*, R^*, A_i^*)$ , where:

$$\begin{aligned}
S_0 &= \frac{1}{R_0} \\
S^* &= \frac{1}{R_0 \left( 1 - \frac{A_i^*}{\exp\left(\frac{-\xi v}{1 + \chi_{\{VI(t-\theta) \geq \tau\}} VI(t-\theta)}\right)} \right)} \\
I^* &= \frac{\eta_2 + \gamma}{\eta_1} G^* \\
G^* &= \frac{\mu\alpha\beta[\chi_{\{VI(t-\theta) \geq \tau\}} VI(t-\theta)]}{\gamma \left[ \frac{\eta_1 \eta_2 + \gamma}{\eta_1} \right]} A_i^* S^* \\
R^* &= \frac{\gamma}{\delta} \left[ \frac{\eta_1 + \eta_2 + \gamma}{\eta_1} \right] G^*
\end{aligned}$$

with

$$R_0 = \frac{\mu\alpha^2\beta\zeta[\chi_{\{VI(t-\theta)\geq\tau\}}VI(t-\theta)]^2 \exp\left(\frac{-\xi v}{1+\chi_{\{VI(t-\theta)\geq\tau\}}VI(t-\theta)}\right)}{\frac{\xi}{1+\chi_{\{VI(t-\theta)\geq\tau\}}VI(t-\theta)}\gamma\left[\frac{\eta_1+\eta_2+\gamma}{\eta_1}\right]}$$

$VI(t-\theta)$  is supposed to be constant equal to  $\tau$  and  $\Delta = -\xi/(1+\tau)$ , when  $t$  is sufficiently large. We can notice that:  $S_0 > S^*$ .

We will now study the stability of the first steady state  $E_0$  by linearizing the system (11) and doing the change of variables:  $x_1(t) = S(t) - S_0(t)$ ;  $x_2(t) = I(t)$ ;  $x_3(t) = G(t)$ ;  $x_4(t) = R(t)$  and  $x_5(t) = A_i(t)$ , we get the following equation:

$$\frac{dx}{dt} = \begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \\ \dot{x}_5 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & \delta & -aS_0 \\ 0 & -(\eta_1 + \gamma) & \eta_2 & 0 & aS_0 \\ 0 & \eta_1 & -(\eta_2 + \gamma) & 0 & 0 \\ 0 & \gamma & \gamma & -\delta & 0 \\ 0 & 0 & b \exp(\Delta v) & 0 & -\Delta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = B_1(x) \quad (12)$$

The characteristic polynomial  $P_{B_1}$  of  $B_1$  is given by:

$$\det(B_1 - \lambda I) = \begin{vmatrix} -\lambda & 0 & 0 & \delta & -aS_0 \\ 0 & -(\lambda + (\eta_1 + \gamma)) & \eta_2 & 0 & aS_0 \\ 0 & \eta_1 & -(\lambda + (\eta_2 + \gamma)) & 0 & 0 \\ 0 & \gamma & \gamma & -(\lambda + \delta) & 0 \\ 0 & 0 & b \exp(\Delta v) & 0 & -(\lambda - \Delta) \end{vmatrix}$$

$$P_{B_1}(\lambda) = \lambda(\lambda + \delta) \{-[\lambda + (\eta_1 + \gamma)][\lambda + (\eta_2 + \gamma)](\lambda - \Delta) + \eta_1\eta_2(\lambda - \Delta) + \eta_1 abS_0 \exp(\Delta v)\}$$

Hence:

$$P_{B_1} = -\lambda^5 - A\lambda^4 - (B + C)\lambda^3 - (\delta B + D)\lambda^2 - \frac{1}{\delta}D\lambda$$

where:

$$\begin{aligned} A &= \eta_1 + \eta_2 + 2\gamma - \Delta + \delta \\ B &= \gamma(\eta_1 + \eta_2 + \gamma) - \Delta(\eta_1 + \eta_2 + 2\gamma) \\ C &= \delta(\eta_1 + \eta_2 + 2\gamma - \Delta) \\ D &= \gamma\Delta(\eta_1 + \eta_2 + 2\gamma) - \eta_1 abS_0 \exp(\Delta v) \end{aligned}$$

All coefficients of the characteristic polynomial being negative, the largest eigenvalue is 0 and the Hessian dominant eigenvalue is strictly positive, then  $E_0$  is unstable.

For the second steady state  $E^*$ , after linearizing the system (11) and changing variables as:

$$x_1(t) = S(t) - S^*; \quad x_2(t) = I(t) - I^*; \quad x_3(t) = G(t) - G^*; \quad x_4(t) = R(t) - R^*; \quad x_5(t) = A_i(t) - A_i^*$$

we get the following equation:

$$\frac{dx}{dt} = \begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \\ \dot{x}_5 \end{pmatrix} = \begin{pmatrix} -aA_i^* & 0 & 0 & \delta & -aS^* \\ aA_i^* & -(\eta_1 + \gamma) & \eta_2 & 0 & aS^* \\ 0 & \eta_1 & -(\eta_2 + \gamma) & 0 & 0 \\ 0 & \gamma & \gamma & -\delta & 0 \\ 0 & 0 & b(\exp(\Delta v) - A_i^*) & 0 & -(bG^* - \Delta) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = B_2(x) \quad (13)$$

The characteristic polynomial  $P_{B_2}$  of  $B_2$  is then given by:

$$\begin{aligned}
P_{B_2}(\lambda) &= \det(B_2 - \lambda I_5) \\
&= \begin{vmatrix} -(\lambda + aA_i^*) & 0 & 0 & \delta & -aS^* \\ aA_i^* & -(\lambda + (\eta_1 + \gamma)) & \eta_2 & 0 & aS^* \\ 0 & \eta_1 & -(\lambda + (\eta_2 + \gamma)) & 0 & 0 \\ 0 & \gamma & \gamma & -(\lambda + \delta) & 0 \\ 0 & 0 & b(\exp(\Delta v) - A_i^*) & 0 & -(\lambda + (bG^* - \Delta)) \end{vmatrix} \\
&= K_1 + K_2
\end{aligned}$$

where:

$$\begin{aligned}
K_1 &= -(\lambda + aA_i^*) \begin{vmatrix} -(\lambda + (\eta_1 + \gamma)) & \eta_2 & 0 & aS^* \\ \eta_1 & -(\lambda + (\eta_2 + \gamma)) & 0 & 0 \\ \gamma & \gamma & -(\lambda + \delta) & 0 \\ 0 & b(\exp(\Delta v) - A_i^*) & 0 & -(\lambda + (bG^* - \Delta)) \end{vmatrix} \\
K_2 &= -aA_i^* \begin{vmatrix} 0 & 0 & \delta & -aS^* \\ \eta_1 & -(\lambda + (\eta_2 + \gamma)) & 0 & 0 \\ \gamma & \gamma & -(\lambda + \delta) & 0 \\ 0 & b(\exp(\Delta v) - A_i^*) & 0 & -(\lambda + (bG^* - \Delta)) \end{vmatrix}
\end{aligned}$$

After some calculations, we get:

$$K_1 = -D_0\lambda^5 - D_1\lambda^4 - D_2\lambda^3 - D_3\lambda^2 - D_4\lambda - D_5$$

where:

$$\begin{aligned}
D_0 &= 1 \\
D_1 &= \eta_1 + \eta_2 + \gamma + bG^* - \Delta + aA_i^* + \delta \\
D_2 &= (\eta_1 + \eta_2 + \gamma)(bG^* - \Delta) + \gamma(\eta_1 + \eta_2 + \gamma + bG^* - \Delta) \\
&\quad + (aA_i^* + \delta)(\eta_1 + \eta_2 + 2\gamma + bG^* - \Delta) + aA_i^*\delta \\
D_3 &= \gamma(\eta_1 + \eta_2 + \gamma)(bG^* - \Delta) - \eta_1 abS^*[A_i^* - \exp(\Delta v)] \\
&\quad + (aA_i^* + \delta)[(\eta_1 + \eta_2 + \gamma)(bG^* - \Delta) + \gamma(\eta_1 + \eta_2 + \gamma + bG^* - \Delta)] \\
D_4 &= (\gamma aA_i^* + \gamma\delta + aA_i^*\delta)[(\eta_1 + \eta_2 + \gamma)(bG^* - \Delta) + \eta_1 abS^*(A_i^* - \exp(\Delta v))] \\
&\quad + aA_i^*\delta\gamma(\eta_1 + \eta_2 + \gamma + bG^* - \Delta) \\
D_5 &= aA_i^*\delta\gamma(\eta_1 + \eta_2 + \gamma)(bG^* - \Delta) + \eta_1 a^2 bA_i^* S^* \delta [A_i^* - \exp(\Delta v)]
\end{aligned}$$

and  $K_2 = -T_0\lambda^2 - T_1\lambda - T_2$  with:

$$\begin{aligned}
T_0 &= aA_i^*\delta\gamma \\
T_1 &= aA_i^*\delta\gamma(\eta_1 + \eta_2 + \gamma + bG^* - \Delta) + \eta_1 a^2 bA_i^* S^* \delta [A_i^* - \exp(\Delta v)] \\
T_2 &= aA_i^*\delta\gamma(\eta_1 + \eta_2 + \gamma)(bG^* - \Delta) + \eta_1 a^2 bA_i^* S^* \delta [A_i^* - \exp(\Delta v)]
\end{aligned}$$

Then we have:

$$P_{B_2}(\lambda) = -D_0\lambda^5 - D_1\lambda^4 - D_2\lambda^3 - (D_3 + T_0)\lambda^2 - (D_4 + T_1)\lambda - (D_5 + T_2)$$

If  $VI(t - \theta)$  is supposed to be constant equal to a large value  $\tau$ , which corresponds to a saturated contagion from hosts  $G(t)$  for a given number of vector *Anopheles*, then  $\exp(\Delta v)$  is small and all the coefficients of the characteristic polynomial are strictly negative, which is the necessary condition for the application of the Routh-Hurwitz criterion; by building the Routh-Hurwitz matrix, all the elements of its first column are positive, which corresponds to the fact that  $E^*$  is locally stable.



### III.7 Perspectives

The last improvements come from the Macdonald  $SI$  model of malaria spread (cf. Macdonald, 1957), another extension of the Ross model, which has the interaction graph given in Figure III.9 and the equations system:

$$\begin{aligned}
 \frac{\partial S_1}{\partial t} &= \frac{-\beta_2 S_1 I_2}{N_H} + r I_1 \\
 \frac{\partial I_1}{\partial t} &= \frac{\beta_2 S_1 I_2}{N_H} - r I_1 \\
 \frac{\partial S_2}{\partial t} &= \omega + f S_2 - \frac{\beta_1 S_2 I_1}{N_V} - \delta S_2 \\
 \frac{\partial I_2}{\partial t} &= f I_2 + K E_2 - \delta I_2 \\
 \frac{\partial E_2}{\partial t} &= f E_2 + \frac{\beta_1 S_2 I_1}{N_V} - K E_2 - \delta E_2
 \end{aligned} \tag{14}$$

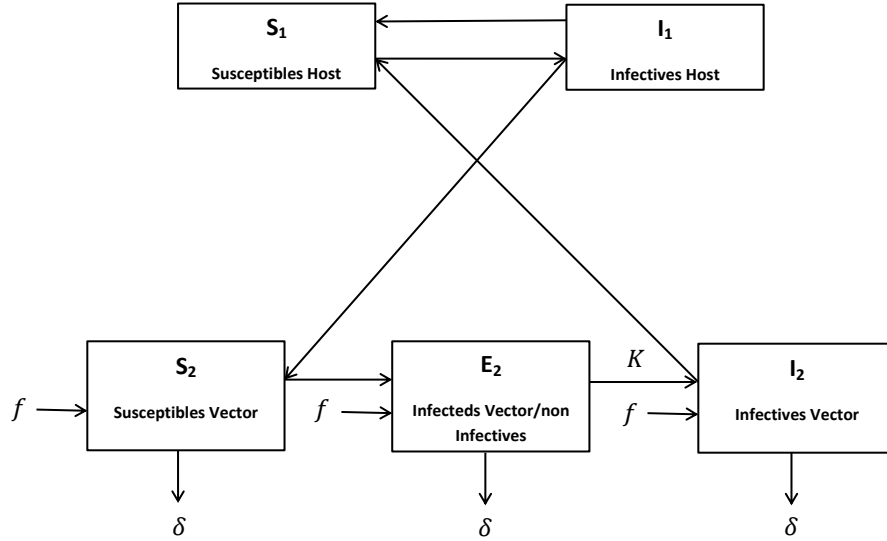


Figure III.9: Interaction graph of the Ross-Macdonald model

where  $f$  (resp.  $\delta$ ) is the fecundity (resp. mortality) rate of the vector population (susceptible, infected and infective vectors being supposed to have the same fecundity and mortality),  $\beta_1$  (resp.  $\beta_2$ ) is the host (resp. vector) contagion parameter,  $N_H$  (resp.  $N_V$ ) is the host (resp. vector) population size, the ratio  $m = N_V/N_H$  is the vector/host ratio,  $K$  (resp.  $r$ ) is the vector (resp. host) speed of passage from the infected/not infective (resp. infected) state to the infective (resp. susceptible) state. If  $f = \delta - \mu$  (the fecundity compensating partly the mortality), the value of  $R_0$ , the mean number of secondary infected vectors for one infective host, is equal to:

$$R_0 = \frac{\beta_1 \beta_2 K}{N_H N_V \mu r (K + \mu)}$$

If  $R_0 > 1$ , assuming that  $\omega = 1$ ,  $N_H = 1/\mu$  and  $m(0) = N_V(0)/N_H = 1$ , the stationary state  $(0, 0, 0, 0, 0)$  is unstable and the endemic stable stationary state is reached after a

transient epidemic wave for the values:

$$i_1^* = \frac{I_1^*}{N_H}, \quad i_2^* = \frac{I_2^*}{N_V}, \quad e_2^* = \frac{E_2^*}{N_V}$$

with:

$$\begin{aligned} i_1^* &= \frac{R_0 - 1}{R_0 + \frac{\beta_1}{\mu}} \\ i_2^* &= \frac{i_1^* r}{m\beta_2(1 - i_1^*)} \\ e_2^* &= \frac{i_1^* \mu r}{Km\beta_2(1 - i_1^*)} \end{aligned}$$

If  $\mu$  is small with respect to  $K$ , then  $K/(K + \mu) \approx e^{-\mu/K}$ , where  $1/K$  is the mean sojourn time in the compartment  $E_2$  (sporogonic cycle duration) and:

$$R_0 = \left[ \frac{\beta_1 \beta_2}{N_V N_H \mu r} \right] e^{-\mu/K}$$

Let us consider now the Cox model with proportional risk (cf. Cooke and Morales-Napoles, 2006) and suppose that the risk function would be given by  $h(t, z) = e^{\rho z} b(t)$ , where  $\rho$  is a regression parameter and  $b(t)$  the baseline risk function. Then, by denoting  $u = e^{\rho z}$ , the survival function  $S(t, u)$  (i.e. the probability to survive until the age  $t$  with a risk  $u$ ) is given by:

$$S(t, u) = \exp\left(-\int_0^t h(s, z) ds\right) = B(t)^u \quad (15)$$

where  $B(t) = \exp\left(-\int_0^t b(s) ds\right)$ .

In the Macdonald model, for calculating the survival function  $S_2$  of the subpopulation  $E_2$ , it is possible to identify:

$$z = \log\left(\frac{\beta_1 \beta_2}{N_V N_H \mu r}\right), \quad \rho = \frac{-K}{\mu}, \quad t = \frac{1}{K}, \quad b(s) = cste = \mu, \quad B\left(\frac{1}{K}\right) = e^{-\mu/K}$$

and:

$$\begin{aligned} R_0 &= \left(\frac{\beta_1 \beta_2}{N_V N_H \mu r}\right) e^{-\mu/K} \\ &= \exp\left(\log\left(\frac{\beta_1 \beta_2}{N_V N_H \mu r}\right)\right) e^{-\mu/K} \\ &\approx S_2\left(\frac{1}{K}, \left(\frac{\beta_1 \beta_2}{N_V N_H \mu r}\right)^{-K/\mu}\right) \end{aligned}$$

if  $\beta_1 \beta_2 / N_V N_H \mu r$  is close to 1.

If there exist  $n$  age classes into the vector subpopulation  $E_2$  whose sojourn times  $T_i$  for  $i = 1, \dots, n$  are independent random variables related to the survival functions  $S_i$ , we have:

$$P(T_i > t_i \text{ for } i = 1, \dots, n | u) = \prod_{i=1}^n S_i(t_i, u) = \prod_{i=1}^n B_i(t_i)^u \quad (16)$$

If  $z$  is a random variable, then  $u = e^{\rho z}$  is also random variable and we define the mean survival function as  $S(t) = \mathbb{E}_u[B(t)^u]$ . If we consider now the Laplace transform defined by:  $\mathbb{E}_u[e^{-vu}] = \exp(-vp) = L(v)$ , where  $p$  is a parameter depending on the probability distribution of  $u$ , we can write:

$$\begin{aligned} P(T_i > t_i \text{ for } i = 1, \dots, n) &= \exp \left[ - \sum_{j=1}^n (-\log[S_j(t_j)^{1/p}]) \right]^p \\ &= C(S_1, \dots, S_n) \end{aligned}$$

where  $C$  is an archimedean copula (cf, Beaudoin and Lakhali-Chaieb, 2008).

By introducing now a demographic dynamics and by using the archimedean copula methodology, we can deal with a proportional risk increasing for example with the biological age (cf. Demongeot, 2009b). Such an approach would be more realistic than the Macdonald's one by taking into account the resistance of both vectors and hosts to infectious diseases, which is highly varying between young or elderly animals and humans; hence, it could be possible to give a better prediction of the efficacy of public health policies like vector eradication, vaccination, quarantine or other preventive actions in the different age classes of the populations of vectors and hosts.

### III.8 Conclusion

We have considered in this chapter some natural extensions of the classical Ross-McKendrick-Macdonald approaches, in order to account for demographic and spatial dependencies of the contagion parameters on the host age and on the vector spread. Two examples have been presented, the first concerning the malaria incidence with environmental dependency in Bankoumana, a locality of Sudanese savannah area in Mali, and the second concerning a retro-prediction of the medieval Black-Death epidemics between 1348 and 1350 in Western Europe. Both examples show the interest of the introduction of space and age classes into the classical equations. In the future, some realistic examples (like Sexually Transmitted Diseases, STD) will be treated showing also the importance of the demography (the sexual relationships depending on the age of the partners) and of the socio-geography (conditioning the sexual behavior). Eventually, based on the knowledge of the new aerial routes (cf. Khan et al., 2009), the study of the Black-Death could also be revisited for the prediction of new possible plague pandemics from the Central Asia reservoir, with a viscosity vanishing no more on maritime but on aerial routes.

# Bibliography

- [1] L. Abbas, J. Demongeot and N. Glade, Synchrony in reaction-diffusion models of morphogenesis: applications to curvature-dependent proliferation and zero-diffusion front waves, *Philos. Trans. R Soc. A*, 367 (2009), pp. 4829-4862.
- [2] T. Balenghien, F. Fouque, P. Sabatier and D.J. Bicout, Horse-, bird-, and human-seeking behavior and seasonal abundance of mosquitoes in a West Nile virus focus of southern France, *J. Med. Entomol.*, 43 (2006), pp. 936-946.
- [3] S. Barry and N. Gualde, The biggest epidemics of history, *L'Histoire*, 310 (2006), pp. 38-49.
- [4] T.P. Baum, N. Pasqual, F. Thuderoz, V. Hierle, D. Chaume, M.P. Lefranc, E. Jouvin-Marche, P. Marche and J. Demongeot, IMGT/GeneInfo: enhancing V(D)J recombination database accessibility, *Nucleic Acids Res.*, 32 (2004), pp. 51-54.
- [5] D. Beaudoin and L. Lakhil-Chaieb, Archimedean copula model selection under dependent truncation, *Stat. Med.*, 27 (2008), pp. 4440-4454.
- [6] O.J. Benedictow, The black death 1346-1353: the complete history, *Boydell Press, Woodbridge*, (2004).
- [7] D. Bernoulli, Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'inoculation pour la prévenir, *Mém. Acad. Roy. Sci., Paris*, (1760).
- [8] D.J. Bicout and P. Sabatier, Mapping rift valley fever vectors and prevalence using rainfall variations, *Vector-Borne Zoonotic Dis*, 4 (2004), pp. 33-42.
- [9] D.J. Bicout, K. Chalvet-Monfray and P. Sabatier, Infection persistence time of Aedes breeding habitats, *Physica A Stat. Mech. Appl.*, 305 (2002), pp. 597-603.
- [10] A. Brandenburg and T. Multamaki, How long can left and right handed life forms coexist? *Int. J. Astrobiol.*, 3 (2004), pp. 209-219.
- [11] J. Brossollet and H. Mollaret, Pourquoi la peste ? Le rat, la puce et le bubon, *Gallimard, Paris*, (1994).
- [12] N. Brouhns and M. Denuit, Risque de longévité et rente viagère, *Institut de Statistique Université Catholique, Lowain*, Discussion Paper 0137, (2001).

- [13] N.F. Cantor, In the wake of the plague: the black death and the world it made, *Free Press, New York*, (2001).
- [14] E. Carpentier, Une ville devant la peste, Orvieto et la peste noire de 1348, *De Boeck Université, Bruxelles*, (1993).
- [15] G. Christakos, R.A. Olea, M.L. Serre, H.L. Yu and L.L. Wang, Interdisciplinary public health reasoning and epidemic modelling: the case of black death, *Springer, Berlin*, (2005).
- [16] S.K. Cohn, The black death: end of a paradigm, *Am. Hist. Rev.*, 107 (2002), pp. 1-54.
- [17] R.M. Cooke and O. Morales-Napoles, Competing risk and the cox proportional hazard model, *J. Stat. Plan. Inference*, 136 (2006), pp. 1621-1637.
- [18] J. D'Alembert, Opuscules Mathématiques, *David, Paris*, (1761).
- [19] G.C. de Magny, C. Paroissin, B. Cazelles, M. de Lara, J.F. Delmas and J.F. Guégan, Modeling environmental impacts of plankton reservoirs on cholera population dynamics, *Esaim*, 14 (2005), pp. 156-173.
- [20] L. Demetrius, Relations between demographic parameters, *Demography*, 16 (1979), pp. 329-338.
- [21] J. Demongeot, Thesis, *Université J. Fourier, Grenoble*, (1983).
- [22] J. Demongeot, Biological boundaries and biological age, *Acta Biotheor.*, 57 (2009b), pp. 397-419.
- [23] J. Demongeot and L. Demetrius, La dérive démographique et la sélection naturelle: Etude empirique de la France (1850-1965), *Population*, 2 (1989), pp. 231-248.
- [24] J. Demongeot and J. Fricot, Random fields and renewal potentials, *NATO ASI Serie F*, 20 (1986), pp. 71-84.
- [25] J. Demongeot, N. Glade and L. Forest, Liénard systems and potential-Hamiltonian decomposition. I Methodology, *Comptes Rendus Mathématique*, 344 (2007a), pp. 121-126.
- [26] J. Demongeot, N. Glade and L. Forest, Liénard systems and potential-Hamiltonian decomposition. II Algorithm, *Comptes Rendus Mathématique*, 344 (2007b), pp. 191-194.
- [27] J. Demongeot, N. Glade, A. Moreira and L. Vial, RNA relics and origin of life, *Int. J. Mol. Sci.*, 10 (2009), pp. 3420-3441.
- [28] J.M.O. Depinay, C.M. Mbogo, G. Killeen, B. Knols, J. Beier, J. Carlson, J. Dusho, P. Billingsley, H. Mwambi, J. Githure, A.M. Toure and F.E. McKenzie, A simulation model of African Anopheles ecology and population dynamics for the analysis of malaria transmission, *Malaria J.*, 3:29 (2004).

- [29] K. Dietz and J.A.P. Heesterbeek, Bernoulli was ahead of modern epidemiology, *Nature*, 408:513 (2000).
- [30] K. Dietz and J.A.P. Heesterbeek, Daniel Bernoulli's epidemiological model revisited, *Math. Biosci.*, 180 (2002), pp. 1-21.
- [31] C. Doliger, Démographie, fécondité et croissance économique en France, Thesis, *Université Louis Pasteur, Strasbourg*, (2006).
- [32] O.K. Doumbo, It takes a village: medical research and ethics in Mali, *Science*, 307 (2005), pp. 679-681.
- [33] M. Drancourt, G. Aboudharam, M. Signoli, O. Dutour and D. Raoult, Detection of 400-year-old *Yersinia pestis* DNA in human dental pulp: an approach to the diagnosis of ancient septicemia, *Proc. Natl. Acad. Sci. USA*, 95 (1998), pp. 2637-2640.
- [34] J. Dutertre, Étude d'un modèle épidémiologique appliqué au paludisme, *Ann. Soc. Belge Méd. Trop.*, 56 (1976), pp. 127-141.
- [35] E.A. Eckert, The black death and the transformation of the West, *Bull. Hist. Med.*, 74 (2000), pp. 356-357.
- [36] R.A. Fisher, The wave of advance of advantageous genes, *Ann. Eugenics.*, 7 (1937), pp. 353-369.
- [37] L. Forest, N. Glade and J. Demongeot, Liénard systems and potential-Hamiltonian decomposition. Applications, *C R Acad. Sci. Biol.*, 330 (2007), pp. 97-106.
- [38] R. Gani and S. Leach, Epidemiologic determinants for modelling pneumonic plague outbreaks, *Emerg. Infect. Dis.*, 10 (2004), pp. 608-614.
- [39] J. Gaudart, R. Giorgi, B. Poudiougou, S. Ranque, O.K. Doumbo and J. Demongeot, Spatial cluster detection: principle and application of different general methods, *Rev. Epidemiol. Santé Publique*, 55 (2007), pp. 297-306.
- [40] J. Gaudart, O. Touré, N. Dessay, A.L. Dicko, S. Ranque, L. Forest, J. Demongeot and O.K. Doumbo, Modelling malaria incidence with environmental dependency in a locality of Sudanese savannah area, Mali, *Malaria J.*, 8:61 (2009).
- [41] J. Gaudart, M. Ghassani, J. Mintsa, J. Waku, M. Rachdi, O.K. Doumbo and J. Demongeot, Demographic and spatial factors as causes of an epidemic spread, the copule approach. Application to the retroprediction of the black death epidemic of 1346, *IEEE AINA' 10 and BLSMC' 10. IEEE Press, Piscataway*, (2010), pp. 751-758.
- [42] N. Glade, L. Forest and J. Demongeot, Liénard systems and potential-Hamiltonian decomposition. III Applications in biology, *Comptes Rendus Mathématique*, 344 (2007), pp. 253-258.

- [43] M. Horie, T. Honda, Y. Suzuki, Y. Kobayashi, T. Daito, T. Oshida, K. Ikuta, P. Jern, T. Gojobori, J.M. Coffin and K. Tomonaga, Endogenous non-retroviral RNA virus elements in mammalian genomes, *Nature*, 463 (2010), pp. 84-88.
- [44] R. Horrox, The black death, *Manchester University Press, Manchester*, (1994).
- [45] J. Huang, E.D. Walker, P.E. Otienoburu, F. Amimo, J. Vulule and J.R. Miller, Laboratory tests of oviposition by the african malaria mosquito, *Anopheles gambiae*, on dark soil as influenced by presence or absence of vegetation, *Malaria J.*, 5:88 (2006).
- [46] J. Kelly, The great mortality: an intimate history of the black death, the most devastating plague of all time, *Harper Collins, New York*, (2005).
- [47] W.O. Kermack and A.G. McKendrick, Contributions to the mathematical theory of epidemics. II. The problem of endemicity, *Proc. R Soc. Lond. A*, 138 (1932), pp. 55-83.
- [48] W.O. Kermack and A.G. McKendrick, Contributions to the mathematical theory of epidemics. III. Further studies of the problem of endemicity, *Proc. R Soc. Lond. A*, 141 (1933), pp. 94-122.
- [49] K. Khan, J. Arino, W. Hu, J. Raposo, J. Sears, F. Calderon, C. Heidebrecht, M. Macdonald, J. Liauw, A. Chan and M. Gardam, Spread of a novel Influenza A (H1N1) virus via global airline transportation, *N Engl. J. Med.*, 361 (2009), pp. 212-214.
- [50] G.J. L'Épine, Rapport de six des douze commissaires (contre l'inoculation), *Quillau, Paris*, (1764), pp. 40-43.
- [51] La peste humaine en 1997, Relevé Epidémiologique Hebdomadaire, *OMS*, 74 (1999), pp. 340-344.
- [52] J.H. Lambert, Beiträge zum Gebrauche der Mathematik und deren Anwendung, *Dritter Theil, Berlin*, (1772).
- [53] P.H. Leslie, On the use of matrices in certain population mathematics, *Biometrika*, 33 (1945), pp. 183-212.
- [54] G. Macdonald, The epidemiology and control of malaria, *Oxford University Press, London*, (1957).
- [55] P.L. Maupertuis, Moreau de, Venus physique, in Oeuvres, *Georg. Olms., Hildesheim*, (1745, reed. 1965).
- [56] N. May, Impartial remarks on the Suttonian method of inoculation, *Tilley, Wheble and Brown, London*, (1770).
- [57] R.M. May and R.M. Anderson, Spatial heterogeneity and the design of immunization programs, *Math. Biosci.*, 72 (1984), pp. 83-111.
- [58] A.G. McKendrick, Applications of mathematics to medical problems, *Proc. Edinburgh Math Soc.*, 44 (1925), pp. 1-34.

- [59] V. Mendez, J. Fort, H.G. Rotstein and S. Fedotov, Speed of reaction-diffusion fronts in spatially heterogeneous media, *Phys. Rev. E*, 68 (2003).
- [60] A. Mischlewski, Un ordre hospitalier au Moyen Age. Les chanoines réguliers de Saint-Antoine en Viennois, *PUG, Grenoble*, (1995).
- [61] G. Mocellin and I. Experton, Saint-Antoine et l'Ordre des Antonins aux XVIIème et XVIIIème siècles, *Ed. Musée, St-Antoine l'Abbaye*, (1992).
- [62] G. Mocellin-Spicuzza, Chroniques d'une abbaye au Moyen-Age. Guérir l'âme et le corps, *Ed. Musée, St-Antoine l'Abbaye*, (2002).
- [63] J.A. Murray, Fata variolarum insitionis in Suecia, *Doctoral dissertation, Universtiy Göttingen, Göttingen*, (1763).
- [64] J.D. Murray, Mathematical biology I and II, *Springer Verlag, Berlin*, (2002).
- [65] P.I. Ndiaye, D.J. Bicout, B. Mondet and P. Sabatier, Rainfall triggered dynamics of Aedes mosquito aggressiveness, *J. Theor. Biol.*, 243 (2006), pp. 222-229.
- [66] S.B. Neerinckx, A.T. Peterson, H. Gulinck, J. Deckers and H. Leirs, Geographic distribution and ecological niche of plague in sub-Saharan Africa, *Int. J. Health Geogr.*, 7:54 (2008).
- [67] W. Orent, Will the black death return? *Discover*, 22 (2001), pp. 1-10.
- [68] W. Orent, Plague: the mysterious past and mystifying future of the world's most dangerous disease, *Free Press, New York*, (2004).
- [69] T. Porphyre, D.J. Bicout and P. Sabatier, Modelling the abundance of mosquito vectors versus flooding dynamics, *Ecol. Modell.*, 183 (2004), pp. 173-181.
- [70] M. Porte, Passion des formes. A René Thom, *ENS Editions, Paris*, (1994).
- [71] M.B. Prentice and L. Rahalison, Plague, *Lancet*, 369 (2007), pp. 1196-1207.
- [72] D. Raoult, G. Aboudharam, E. Crubézy, G. Larrouy, B. Ludes and M. Drancourt, Molecular identification of Suicide PCR of Yersinia pestis as the agent of medieval black death, *Proc Natl Acad Sci USA*, 97 (2000), pp. 12800-12803.
- [73] Y. Renouard, Conséquences et intérêt démographiques de la Peste noire de 1348, *Population*, 3 (1948), pp. 459-466.
- [74] R. Ross, An application of the theory of probabilities to the study of a priori pathometry. Part I, *Proc. R Soc. Lond. A*, 92 (1916), pp. 204-230.
- [75] J.C. Russell, British medieval population, Thesis, *Univ. of New Mexico, Albuquerque*, (1948).



- [76] J.C. Russell, Population in Europe. In: Cipolla CM (ed) The Fontana economic history of Europe, vol. I, the middle ages, *Collins/Fontana, Glasgow*, (1972), pp. 25-71.
- [77] S. Scott and C. Duncan, Return of the black death: the world's greatest serial killer West Sussex, *J. Wiley, New York*, (2004).
- [78] N.C. Stenseth, B.B. Atshabar, M. Begon, S.R. Belmain, E. Bertherat, E. Carniel, K.L. Gage, H. Leirs and L. Rahalison, Plague: past, present, and future, *PLOS Med.*, 5 (2008).
- [79] R. Thom, Stabilité structurelle et Morphogénèse, *Benjamin, New York*, (1972).
- [80] F. Thuderoz, M.A. Simonet, O. Hansen, A. Dariz, T.P. Baum, V. Hierle, J. Demongeot, P.N. Marche and E. Jouvin-Marche, From the TCRAD rearrangement quantification to the computational simulation of the locus behavior, *PloS Comp. Biol.*, 6 (2010).
- [81] J. Trembley, Recherches sur la mortalité de la petite vérole, *Mém. Acad. Roy. Sci., Paris*, (1796).
- [82] M.B. Usher, A matrix model for forest management, *Biometrics*, 25 (1969), pp. 309-315.
- [83] G. Villani, Florence, 1336-1338, the chronicle of Giovanni Villani. In: Medieval trade in the mediterranean world, Lopez, R.S. and I.W. Raymond Trads, *Columbia Un. Press, New York*, (2001).
- [84] M. Wheelis, Biological warfare at the 1346 Siege of Caffa, *Emerg. Infect Dis.*, 8 (2002), pp. 971-975.
- [85] WHO, La peste humaine en 1997. Relevé Epidémiologique Hebdomadaire, *OMS*, 74 (1999), pp. 340-344.
- [86] J.W. Wood, The temporal dynamics of the fourteenth-century black death: new evidence from english ecclesiastical records, *Hum. Biol.*, 75 (2003), pp. 427-448.
- [87] E.C. Zeeman, Controversy in science: on the ideas of Daniel Bernoulli and René Thom, *Nieuw Arch. Wisk.*, 11:257 (1993).
- [88] S.Y. Zhang, L. Yu and P. Daszak, EcoHealth and the black death in the year of the rat, *Ecohealth*, 5 (2008), pp. 99-100.

## Chapter IV

# Archimedean copula and contagion modelling in epidemiology<sup>1</sup>

**Abstract.** The aim of this chapter is, first, to find interactions between compartments of hosts in the Ross-Macdonald Malaria transmission system. So, to make clearer this association we introduce the concordance measure and then the Kendall's tau and Spearman's rho. Moreover, since the population compartments are dependent, we compute their conditional distribution function using the Archimedean copula. Secondly, we get the vector population partition into several dependent parts conditionally to the fecundity and to the transmission parameters and we show that we can divide the vector population by using  $p$ -th quantiles and test the independence between the subpopulations of susceptibles and infecteds. Third, we calculate the  $p$ -th quantiles with the Poisson distribution. Fourth, we introduce the proportional risk model of Cox in the Ross-Macdonald model with the copula approach to find the relationship between survival functions of compartments.

**Keywords and phrases:** Archimedean copula, Quantile, Malaria, Transmission system, Measure of concordance.

### IV.1 Introduction

Advances in epidemics modelling have been done recently by introducing demographic aspects (i.e. consideration of host populations whose global size changes during the epidemic and the endemic history) as well as spatial aspects about vector or infectious agents spread. As examples of application, malaria endemics in South Mali or recurrent seasonal influenza epidemics with irregular pandemics, are modelled by the same type of mathematical framework, coming from the Ross-Macdonald tradition. We will focus in this chapter on some improvements of this classical model, in terms of:

- mechanism of contacts, supposed to be assimilated to shocks, like in stochastic chemistry.
- demography of non constant population, with a fecundity and mortality parameters specific to considered populations of hosts and vectors.
- differential infectious risk caused by a given population of hosts or vectors, in comparison with other populations less or more subjected to infectious risk.

---

1. J. Demongeot, M. Ghassani, M. Rachdi, I. Ouassou and C. Taramasco, Archimedean Copula and Contagion Modeling in Epidemiology, Networks and Heterogeneous Media, (submitted).

These improvements come from the fact that the Ross-Macdonald model  $SI$  has the following interaction graph (cf. [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13] and references therein):

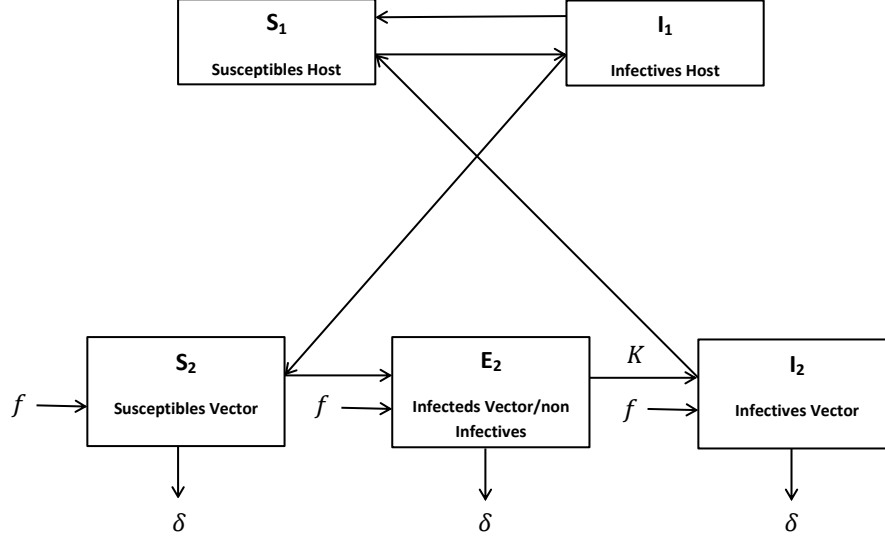


Figure IV.1: Interaction graph of the Ross-Macdonald model

The equations of the Ross-Macdonald model corresponding to the interaction graph in Figure IV.1 are as follows:

$$\begin{aligned}
 \frac{\partial S_1}{\partial t} &= \frac{-\beta_2 S_1 I_2}{N_H} + r I_1 \\
 \frac{\partial I_1}{\partial t} &= \frac{\beta_2 S_1 I_2}{N_H} - r I_1 \\
 \frac{\partial S_2}{\partial t} &= \omega + f S_2 - \frac{\beta_1 S_2 I_1}{N_V} - \delta S_2 \\
 \frac{\partial I_2}{\partial t} &= f I_2 + K E_2 - \delta I_2 \\
 \frac{\partial E_2}{\partial t} &= f E_2 + \frac{\beta_1 S_2 I_1}{N_V} - K E_2 - \delta E_2
 \end{aligned} \tag{1}$$

where  $f$  (resp.  $\delta$ ) is the fecundity (resp. the mortality) rate of the vector population (susceptible, infected and infective vectors being supposed to have the same fecundity and mortality),  $\beta_1$  (resp.  $\beta_2$ ) is the host (resp. the vector) contagion parameter,  $N_H$  (resp.  $N_V$ ) is the host (resp. the vector) population size, the ratio  $m = N_V/N_H$  is the vector/host ratio,  $K$  (resp.  $r$ ) is the vector (resp. the host) speed of passage from the infected/not infective (resp. the infected) state to the infective (resp. to the susceptible) state. If  $f = \delta - \mu$  (the fecundity compensating partly the mortality), the value of  $R_0$ , the mean number of secondary infected vectors for one infective host, is equal to:

$$R_0 = \frac{\beta_1 \beta_2 K}{N_H N_V \mu r (K + \mu)}$$

If  $R_0 > 1$ , assuming that  $\omega = 1$ ,  $N_H = 1/\mu$  and  $m(0) = N_V(0)/N_H = 1$ , then the stationary state  $(0, 0, 0, 0, 0)$  is unstable and the endemic stable stationary state is reached

after a transient epidemic wave for the values:

$$i_1^* = \frac{I_1^*}{N_H}, \quad i_2^* = \frac{I_2^*}{N_V}, \quad e_2^* = \frac{E_2^*}{N_V}$$

with

$$\begin{aligned} i_1^* &= \frac{R_0 - 1}{R_0 + \frac{\beta_1}{\mu}} \\ i_2^* &= \frac{i_1^* r}{m\beta_2(1 - i_1^*)} \\ e_2^* &= \frac{i_1^* \mu r}{Km\beta_2(1 - i_1^*)} \end{aligned}$$

If  $\mu$  is small with respect to  $K$ , then  $K/(K + \mu) \approx e^{-\mu/K}$ , where  $1/K$  is the mean sojourn time in the compartment  $E_2$  (the sporogonic cycle duration) and  $R_0 = (\beta_1\beta_2/N_V N_H \mu r) e^{-\mu/K}$ .

In this chapter, we are interested in analyzing interactions between vector compartments in system (1), by using a copula approach. In fact, partitioning into several compartments the studied population must be done in the model in a rigorous way. For this aim, we focus on estimating the conditional quantiles of several compartments of the vector population. This can be made by estimating the marginal distribution of the sizes of susceptible and infected not infectious vector populations which may be done through an Archimedean copula approach. It must be said that copula functions permit to modelize compartments associations and then highlight their dependence structure which allows a better stratification of the population. So, to make clearer these associations we introduce the concordance measure and then the Kendall's tau and Spearman's rho. Afterward, we show that we can divide the population by using the  $p$ -th percentiles, where  $p \in ]0, 1[$ . Remark that it is known that the copula functions are used essentially in finance and this is the first time we use the information it brings to explain an epidemiological phenomena. We believe that this study could open a window on several future developments.

Recall that, in statistics, a copula is used as a general way of formulating a multivariate distribution in such a way that various general types of dependence can be represented. The approach to formulating a multivariate distribution using a copula is based on the idea that a simple transformation can be made on each marginal variable in such a way that each such transformed variable has a uniform distribution. Once this is done, the dependence structure can be expressed as a multivariate distribution of the obtained uniform variables, and a copula is precisely the multivariate distribution of these marginally uniform random variables. There are many families of copulas which differ in the detail of the dependence they represent. A parameterized family of distribution functions will typically depend on several parameters which relate to the strength and form of the dependence. A typical use of copulas consists in choosing such a family in order to define the adequate multivariate distribution fitting the empirical distribution observed from a sample of data, and then to derive the copula corresponding to this multivariate distribution.

A copula is a multivariate joint distribution defined on the  $n$ -dimensional unit cube  $[0, 1]^n$  such that every marginal distribution be uniform on the interval  $[0, 1]$ . Let consider a copula function  $C$  for an  $n$ -dimensional random variable  $X = (X_1, \dots, X_n)$  defined on a probability space  $(\Omega, \Sigma, \mathbb{P})$  with a joint distribution function  $F_X$  such that for any  $x_1, \dots, x_n \in \mathbf{R}$ , (cf.

[14, 15, 16, 17, 18, 19, 20, 21] and references therein), we have:

$$F_X(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n))$$

where  $F_i$  is the marginal distribution function of  $X_i$  for  $i = 1, \dots, n$ .

This chapter is organized as follows. In Section II, we give the definition of the Archimedean copula and provide two examples of the Gumbel's and Clayton's copulas, we define and discuss the Kendall's tau, the Spearman's rho, and then, we propose a measure of concordance. At the end of this Section, we will give some formulas of conditional distribution functions using Archimedean copulas. In Section III, we discuss the interaction between the vector compartments, susceptible and infected not infectious, of the system (1), by providing a link between the Kendall's tau and the regression parameter of copulas. After, we will conduct a simulation study to highlight the interaction between the distribution functions of these two compartments. Eventually, we will discuss the conditional quantiles from copulas. In Section IV, we study an application in which we calculate the  $p$ -quantiles with any distribution function and after with, specifically, the Poisson distribution. In the last Section, we introduce the proportional risk model of Cox in the Ross-Macdonald model (1) with the copula approach to quantify the relationships existing between the survival functions of the considered compartments.

## IV.2 Preliminaries

### IV.2.1 Archimedean copula

The most important family of copulas is the Archimedean one (cf. [2]). This latter is defined as follows:

$$C(u_1, \dots, u_n) = \begin{cases} \phi^{-1}(\phi(u_1) + \dots + \phi(u_n)) & \text{if } \phi(u_1) + \dots + \phi(u_n) \leq 0 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where the generator of the copula  $\phi$  is a twice continuously differentiable function which satisfies:

$$\phi(1) = 0, \quad \phi^{(1)}(u) < 0 \quad \text{and} \quad \phi^{(2)}(u) > 0 \quad \text{for all } u \in [0, 1]^n$$

where  $\phi^{(i)}$  denotes the  $i$ th order derivative of  $\phi$ .

Notice that, there are some popular Archimedean copulas:

– Clayton copula:

$$C(u, v) = (u^{-\alpha} + v^{-\alpha} - 1)^{-\frac{1}{\alpha}} \quad \text{where } \alpha > 0 \quad \text{and} \quad \phi(t) = t^{-\alpha} - 1$$

– Gumbel copula:

$$C(u, v) = \exp\{-[(-\ln u)^\alpha + (-\ln v)^\alpha]^{\frac{1}{\alpha}}\} \quad \text{where } \alpha \geq 1 \quad \text{and} \quad \phi(t) = (-\ln t)^\alpha$$

On the other hand, let consider that  $U$  and  $V$  are two random variables defined on the probability space  $(\Omega, \Sigma, \mathbb{P})$  and uniformly distributed on  $[0, 1]$ , and let  $X$  be the random

variable  $C(U, V)$  valued in  $[0, 1]$ . Then, the real  $\alpha$  corresponds to a regression parameter. In the case of Archimedean copula, the distribution function of  $X$  is given by:

$$F_X(t) = t - \frac{\phi(t)}{\phi^{(1)}(t)}$$

In the case of the Gumbel copula, we have  $\phi(t) = (-\ln t)^\alpha$ , so  $\phi^{(1)}(t) = \frac{-\alpha}{t}(-\ln t)^{\alpha-1}$ . Then:

$$\begin{aligned} F_X(t) &= t - \frac{(-\ln t)^\alpha}{\frac{-\alpha}{t}(-\ln t)^{\alpha-1}} \\ &= t + \frac{t}{\alpha}(-\ln t)^{\alpha-\alpha+1} \\ &= t - \frac{t \ln t}{\alpha} \end{aligned}$$

and the probability density function of  $X$  is the given by:

$$f_X(t) = 1 - \frac{1}{\alpha} - \frac{\ln t}{\alpha} \tag{3}$$

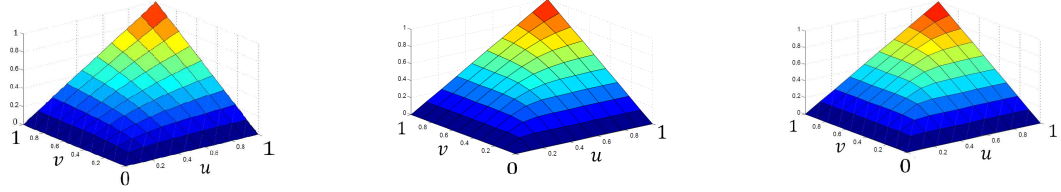


Figure IV.2: Distribution functions of Clayton bivariate copula (resp. for  $\alpha = 2$ ,  $\alpha = 5$  and  $\alpha = 10$ )

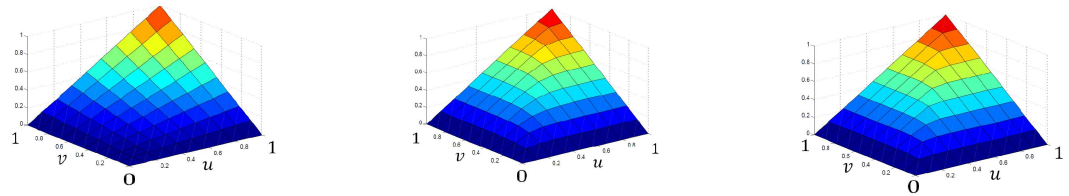


Figure IV.3: Distribution functions of Gumbel bivariate copula (resp. for  $\alpha = 2$ ,  $\alpha = 5$  and  $\alpha = 10$ )

### IV.2.2 Concordance, Kendall's tau, Spearman's rho, Copula

In order to define the concordance measure, we begin by defining the concordance (cf. [18]): two observations  $(x_1, y_1)$  and  $(x_2, y_2)$  of a pair  $(X, Y)$  of continuous random variables are said to be concordant if both values of one pair are greater than the corresponding values

of the other pair, that is if  $x_1 < x_2$  and  $y_1 < y_2$  or if  $x_1 > x_2$  and  $y_1 > y_2$ ; and they are said to be discordant if for one pair one value is greater and the other is smaller than the corresponding value of the other pair, that is if  $x_1 < x_2$  and  $y_1 > y_2$  or  $x_1 > x_2$  and  $y_1 < y_2$ . The simple version of the measure of association known as Kendall's tau is defined in terms of concordance as follows (cf. [18]): let  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  denote a random sample of  $n$  observations from a continuous random vector  $(X, Y)$ . The Kendall's tau for the latter random sample is defined by:

$$\tau = \frac{\text{number of concordant pairs} - \text{number of discordant pairs}}{\text{total number of pairs}} \quad (4)$$

Equivalently,  $\tau$  may be written and interpreted as the difference between the empirical probabilities of concordance and discordance for a pair of observations  $(x_i, y_i)$  and  $(x_j, y_j)$ , chosen randomly from the sample. The multivariate version of Kendall's tau will be defined similarly.

Let  $(X_1, Y_1)$  and  $(X_2, Y_2)$  be independent and identically distributed random vectors, each with joint distribution function  $H$ . So, the probability that  $(X_1, Y_1)$  and  $(X_2, Y_2)$  are concordant is equal to  $\mathbb{P}((X_1 - X_2)(Y_1 - Y_2) > 0)$  and the probability that these two vectors are discordant is equal to  $\mathbb{P}((X_1 - X_2)(Y_1 - Y_2) < 0)$ , and we have:

$$\begin{aligned} \mathbb{P}((X_1 - X_2)(Y_1 - Y_2) > 0) &= \frac{\text{number of concordant pairs}}{\text{total number of pairs}} \\ \mathbb{P}((X_1 - X_2)(Y_1 - Y_2) < 0) &= \frac{\text{number of discordant pairs}}{\text{total number of pairs}} \end{aligned}$$

Then the multivariate generalization of Kendall's tau is defined as the difference between the probability of concordance and the probability of discordance:

$$\tau = \mathbb{P}((X_1 - X_2)(Y_1 - Y_2) > 0) - \mathbb{P}((X_1 - X_2)(Y_1 - Y_2) < 0)$$

#### IV.2.2.1 Kendall's tau and copula relation

**Theorem IV.1** *Let  $X$  and  $Y$  be continuous random variables whose copula is  $C$ . Then the multivariate version of Kendall's tau for  $X$  and  $Y$ , denoted by  $\tau_{X,Y}$  or  $\tau_C$  is given by:*

$$\tau_C = 4 \int_0^1 \int_0^1 C(u, v) c(u, v) du dv - 1 \quad (5)$$

The proof of Theorem IV.1 is given in Appendix A.

For computational purposes, there are alternative expressions for  $\tau_C$ . The integral which appears in (5) can be interpreted as the expected value of the function  $C(U, V)$  where  $U$  and  $V$  are random variables uniformly distributed on  $[0, 1]$  for which the joint distribution function is  $C$ , i.e.:

$$\tau_C = 4 \mathbb{E}(C(U, V)) - 1 = 4 \int_0^1 t f_C(t) dt - 1 \quad (6)$$

where  $f_C(t)$  denotes the density function of  $C$ .

### IV.2.2.2 Spearman's rho

The multivariate version of Spearman's rho is based on the concordance and discordance values. In order to obtain the population value of this measure (cf. [18]), let  $(X_1, Y_1)$ ,  $(X_2, Y_2)$  and  $(X_3, Y_3)$  be three independent random vectors whose components are continuous random variables with a common joint distribution function  $H$  (whose marginal distribution are  $F$  and  $G$ ) and let  $C$  be a copula function. The multivariate version of Spearman's rho is three time the difference between the probability of concordance and the probability of discordance of the two vector pairs  $(X_1, X_2)$  and  $(Y_1, Y_3)$ , i.e.:

$$\rho = 3 \left( \mathbb{P}((X_1 - X_2)(Y_1 - Y_3) > 0) - \mathbb{P}((X_1 - X_2)(Y_1 - Y_3) < 0) \right)$$

The pair  $(X_1, X_3)$  could be used equally as well.

On the other hand, notice that while the joint distribution function of  $(X_1, Y_1)$  is  $H(x, y)$ , the joint distribution function of  $(X_2, Y_3)$  is  $F(x)G(y)$  (because  $X_2$  and  $Y_3$  are independent), thus the copula of  $X_2$  and  $Y_3$  is the independence copula, i.e.,  $C(u_1, u_2) = u_1 u_2$ .

**Theorem IV.2** *Let  $X$  and  $Y$  be continuous random variables whose copula function is  $C$ . Then the multivariate version of Spearman's rho for  $X$  and  $Y$ , denoted by  $\rho_{X,Y}$  or  $\rho_C$ , is given by:*

$$\rho_C = 12 \int \int_{I^2} u_1 u_2 dC(u_1, u_2) - 3$$

where  $I = [0, 1]$ .

The proof of theorem IV.2 is given in Appendix B.

### IV.2.2.3 Measure of concordance

**Definition IV.1** *A measure of association, denoted  $M$ , between two continuous random variables  $X$  and  $Y$  whose copula function is  $C$ , is said to be a measure of concordance, if it satisfies the following properties (in which we denote  $M$  by  $M_{X,Y}$  or  $M_C$  when convenient):*

- $M_{X,Y}$  is defined for every pair  $(X, Y)$  of continuous random variables
- $M_{X,Y} \in (-1, 1)$  with  $M_{X,X} = 1$  and  $M_{X,-X} = -1$
- $M_{X,Y} = M_{Y,X}$
- If  $X$  and  $Y$  are independent, then  $M_{X,Y} = 0$
- $M_{-X,Y} = M_{X,-Y} = -M_{X,Y}$
- if  $C_1$  and  $C_2$  are copulas with  $C_1 \leq C_2$  then  $M_{C_1} \leq M_{C_2}$
- if  $\{(X_n, Y_n)\}_{n \in \mathbb{N}^*}$  is a sequence of pairs of continuous random variables with copula functions  $C_n$  and if  $C_n$  converges in pointwise to  $C$  then,  $\lim_{n \rightarrow +\infty} M_{C_n} = M_C$

**Theorem IV.3** *The multivariate versions of Kendall's tau and Spearman's rho are two measures of concordance.*

The proof of Theorem IV.3 is given in Appendix C.

### IV.2.3 Multidimensional conditional distribution function using copulas

Let  $X_1, X_2, \dots, X_k$  be random variables defined on a probability space  $(\Omega, \Sigma, \mathbb{P})$ , where the joint probability density function  $f_k$  of the vector  $X = (X_1, X_2, \dots, X_k)$  is assumed to exist. By using the Archimedean copula construction, the joint probability density function of  $X$  may be written as follows:



$$\begin{aligned}
f_k(x_1, \dots, x_k) &= \frac{\partial^k}{\partial x_1 \dots \partial x_k} \phi^{-1} \left\{ \phi(F_1(x_1)) + \dots + \phi(F_k(x_k)) \right\} \\
&= \phi^{-1(k)} \left\{ \phi(F_1(x_1)) + \dots + \phi(F_k(x_k)) \right\} \prod_{j=1}^k \phi^{(1)}(F_j(x_j)) F_j^{(1)}(x_j)
\end{aligned} \tag{7}$$

for all  $(x_1, \dots, x_k) \in \mathbb{R}^k$ , where  $F_j$  denotes the marginal distribution function of  $X_j$ .

Thus, the conditional density of  $X_k$  given  $X_1, \dots, X_{k-1}$  is given by:

$$\begin{aligned}
f_k(x_k | x_1, \dots, x_{k-1}) &= \frac{f_k(x_1, \dots, x_k)}{f_{k-1}(x_1, \dots, x_{k-1})} \\
&= \phi^{(1)}(F_k(x_k)) F^{(1)}(x_k) \frac{\phi^{-1(k)} \left\{ \phi(F_1(x_1)) + \dots + \phi(F_k(x_k)) \right\}}{\phi^{-1(k-1)} \left\{ \phi(F_1(x_1)) + \dots + \phi(F_{k-1}(x_{k-1})) \right\}}
\end{aligned} \tag{8}$$

Further, the conditional distribution function of  $X_k$  given  $X_1, \dots, X_{k-1}$  is also given by:

$$\begin{aligned}
F_k(x_k | x_1, \dots, x_{k-1}) &= \int_{-\infty}^{x_k} f_k(x | x_1, \dots, x_{k-1}) dx \\
&= \frac{\phi^{-1(k-1)} \left\{ \phi(F_1(x_1)) + \dots + \phi(F_k(x_k)) \right\}}{\phi^{-1(k-1)} \left\{ \phi(F_1(x_1)) + \dots + \phi(F_{k-1}(x_{k-1})) \right\}} \\
&= \frac{\phi^{-1(k-1)} \left\{ c_{k-1} + \phi(F_k(x_k)) \right\}}{\phi^{-1(k-1)}(c_{k-1})}
\end{aligned} \tag{9}$$

where  $c_k = \phi(F_1(x_1)) + \dots + \phi(F_k(x_k))$ .

### IV.3 Quantiles regression based on the system of transmission of Malaria

Recall that the regression function is the most widely used tool for describing multivariate relationships. Then, copula functions can help to understand the full joint distribution and thus be used to address some important applications, which we tackle to explain in the sequel. In this part, we will use the Gumbel copula.

### IV.3.1 Interaction between compartments of hosts

Let  $X$  be the random variable  $C(U, V)$  defined on the probability space  $(\Omega, \Sigma, \mathbb{P})$ , where  $F_X$  is the distribution function of  $X$ . From (3) we obtain (cf. [19]):

$$\begin{aligned}
 \mathbb{E}(X) &= \int_0^1 t f_X(t) dt \\
 &= \left(1 - \frac{1}{\alpha}\right) \int_0^1 t dt - \frac{1}{\alpha} \int_0^1 t \ln t dt \\
 &= \frac{1}{2} - \frac{1}{2\alpha} - \frac{1}{\alpha} \left(\frac{-1}{4}\right) \\
 &= \frac{2\alpha - 1}{4\alpha}
 \end{aligned} \tag{10}$$

Then, from the equation (6) the Kendall's tau is expressed in the following manner:

$$\begin{aligned}
 \tau_C &= 4 \left(\frac{2\alpha - 1}{4\alpha}\right) - 1 \\
 &= \frac{\alpha - 1}{\alpha}
 \end{aligned} \tag{11}$$

where  $C$  denotes the Gumbel copula.

Notice that the parameter  $\alpha$  measures the degree of dependence between two variables. Recall that the higher it is, the stronger is the dependence between the two studied variables.

#### IV.3.1.1 Relationship between the Kendall's tau and the equilibrium of the system of transmission of Malaria

Let  $(S_{21}, E_{21}), (S_{22}, E_{22}), \dots, (S_{2n}, E_{2n})$  be calculated from  $n$  successive independent observations of samples of  $m$  individuals extracted from the part of the vector population made of the two compartments, susceptible and infected non infectious:  $S_{2i}$  (resp.  $E_{2i}$ ) is the number of susceptible (resp. infected non infectious) vectors, counted in the  $i$ -th sample observed in mosquitoes population transmitting malaria (cf. [22, 23, 24, 25, 26, 27, 28]).

From the equation (4), if the number of concordant pairs is equal to the number of discordant pairs, so we will have the Kendall's tau equal to 0 and the regression parameter equal to 1 ( $\tau_C = 0$  and  $\alpha = 1$ ), then in this case the two compartments will be non concordant (notion similar to uncorrelated), property observed in the case of stochastic independence, and for example when the system is in equilibrium (but also in case of deterministic chaos, cf. [29]). However, if the number of concordant pairs is equal to the total number of pairs, so we will have  $\tau_C = 1$  and then we will have the perfect concordance and so the system is in transient state. The sampling used to calculate  $\tau_C$  can also be obtained by simulating Individual Based Models (IBM) in which social networks allow to simulate all the possibilities of contact between hosts and vectors, by using the same simulation techniques as in stochastic molecular kinetics (cf. [30, 31, 32, 33, 34, 35, 36]). An example of such a social network has been simulated for explaining the contagion process of a social disease, the obesity (cf. [37]). This social network has been simulated on Figure IV.4, in which the homophilic rule is based on the fact that individuals tend to interact with those who resemble them in terms of social behaviour.

In the system (1), when the contagion parameter  $\beta_1$  or the kinetic parameter  $K$  between the two compartments  $S_2$  and  $E_2$  increases, then the number of concordant pairs increases,

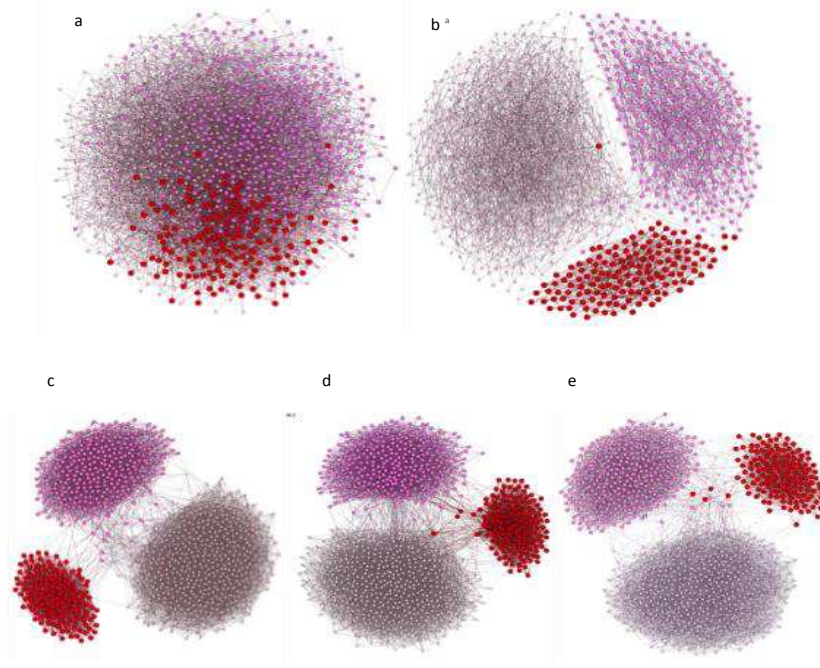


Figure IV.4: Simulation of social graphs representing a contagion network, with initial conditions (a) and asymptotic state in case of an homophilic graph (b), random graph (c), scale free graph (d) and small world graph (e).

because in both cases the size of  $S_2$  diminishes and the size of  $E_2$  is growing, thus the Kendall's tau increases, so we could expect that the system is still in a transient state.

Let consider a simplified version of the system (1), given by:

$$\begin{aligned}
 \frac{\partial S_1}{\partial t} &= -\beta_2 S_1 I_2 + r I_1 \\
 \frac{\partial I_1}{\partial t} &= \beta_2 S_1 I_2 - r I_1 \\
 \frac{\partial S_2}{\partial t} &= \omega - \beta_1 S_2 I_1 - \mu S_2 \\
 \frac{\partial E_2}{\partial t} &= \beta_1 S_2 I_1 - (\mu + K) E_2 \\
 \frac{\partial I_2}{\partial t} &= K E_2 - \mu I_2
 \end{aligned} \tag{12}$$

If we calculate its Jacobian matrix  $J$ , we get:

$$J = \begin{pmatrix} -\beta_2 I_2 & r & 0 & 0 & -\beta_2 S_1 \\ \beta_2 I_2 & -r & 0 & 0 & \beta_2 S_1 \\ 0 & -\beta_1 S_2 & -\beta_1 I_1 - \mu & 0 & 0 \\ 0 & \beta_1 S_2 & \beta_1 I_1 & -(\mu + K) & 0 \\ 0 & 0 & 0 & K & -\mu \end{pmatrix}$$

0 is a trivial eigenvalue and the rest of the spectrum is given by the eigenvalues of the

submatrix  $A$ :

$$A = \begin{pmatrix} -r & 0 & 0 & rI_1/I_2 \\ -\beta_1 S_2 & -\beta_1 I_1 - \mu & 0 & 0 \\ \beta_1 S_2 & \beta_1 I_1 & -(\mu + K) & 0 \\ 0 & 0 & K & -\mu \end{pmatrix}$$

If we neglect the mortality  $\mu$  and the recovering parameter  $r$ , the eigenvalues can be calculated at the stationary state by making explicit the characteristic polynomial:

$$\lambda(K - \lambda)(\lambda - \beta_1(I_1^* - S_2^*)) = \lambda(K - \lambda)(\lambda - \beta_1(I_1^* - \frac{\omega}{\beta_1 I_1^*}))$$

If  $\beta_1$  increases, we have seen in the Introduction section that  $R_0$  increases, then  $I_1^*$  tends to 1 if  $K \gg \beta_1 - \omega$ , then the rate of convergence to the stationary state being given by  $\log(K/(\beta_1 - \omega))$ , which causes a long transient. We observe the same behaviour, if  $\beta_1 \gg K - \omega$ .

For example, let us take a sample of tail 1000 in each compartment. Figures IV.5, IV.6, IV.7 and IV.8 show that the interaction between the distribution functions  $S_k$  and  $E_k$  in each compartment with the Gumbel and Clayton copulas. In each figure we have three graphs. In the first graph on the left we assume that there is neither a contagion nor a fertility for the two compartments so that the number of concordant pairs is equal to the number of discordant pairs: it is the case of independence between the two compartments, then the regression parameter is equal to 1 and the Kendall's tau is equal to 0, so in this case the equilibrium of the system is reached. In the graph that is in the middle, we assume that there is a dependency between the two compartments, with a regression parameter  $\alpha$  equal to 3, so the Kendall's tau increases and we have an unbalanced system. In the graph on the right, we increased the part of dependency between the two compartments, by taking the regression parameter equal to 5, thus the Kendall's tau increases a little more and we have a system increasingly unbalanced. To summarize, when  $\tau_C = 1$ , the system is totally unbalanced and when  $\tau_C = 0$ , the system is totally balanced.

In the graphs that are in the middle and right, it is clear that we are not in the case of total dependence, so there are individuals who are dependent and individuals who are independent.

In the next section we will use the quantiles to divide the population into several parts, where individuals inside each part are dependent, in relation to the Kendall's tau.

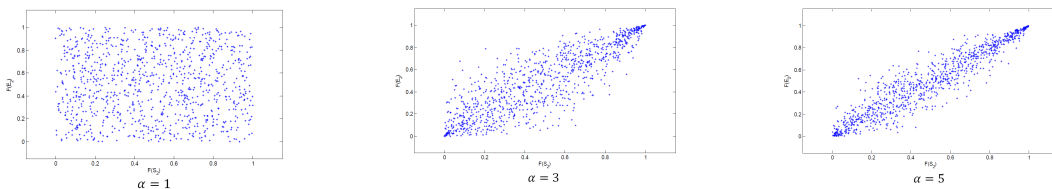


Figure IV.5: Interaction between the distribution functions of  $S_2$  and  $E_2$  using the Gumbel copula with the parameter of regression equal to 1, 3 and 5 respectively

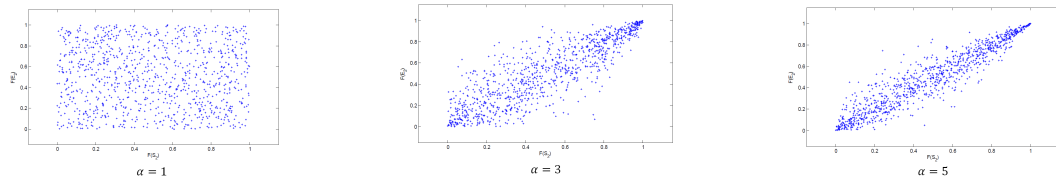


Figure IV.6: Interaction between the distribution functions of  $E_2$  and  $I_2$  using the Gumbel copula with the parameter of regression equal to 1, 3 and 5 respectively

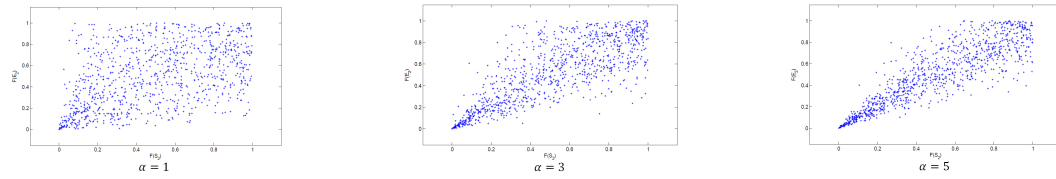


Figure IV.7: Interaction between the distribution functions of  $S_2$  and  $E_2$  using the Clayton copula with the parameter of regression equal to 1, 3 and 5 respectively

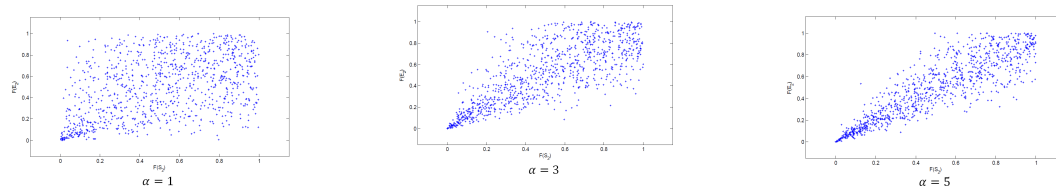


Figure IV.8: Interaction between the distribution functions of  $E_2$  and  $I_2$  using the Clayton copula with the parameter of regression equal to 1, 3 and 5 respectively

### IV.3.2 Quantile regression using the bivariate Gumbel copula

In general, it is known that the calculation of the regression function is tedious. As an alternative, copulas are well-suited in the concept of quantile regression. Instead of examining the mean of a conditional distribution, one looks at the median or some other quantiles (for instance, percentiles) of this distribution. For  $p \in (0, 1)$ , the  $p$ -th quantile is defined as the solution  $t_p$  of the equation:

$$p = F_{X_k}(t_p | X_1 = x_1, \dots, X_{k-1} = x_{k-1})$$

That we will denote, for the sake of simplicity, in the following by:

$$p = F_k(t_p | x_1, \dots, x_{k-1})$$

Let  $F_{S_2}$  and  $F_{E_2}$  be the distribution functions of the two vector compartments, susceptible and infected respectively. The bivariate Gumbel copula for these two compartments is defined as follows:

$$C(F_{S_2}, F_{E_2}) = \exp\{-[(-\ln F_{S_2})^\alpha + (-\ln F_{E_2})^\alpha]^\frac{1}{\alpha}\} \quad (13)$$

where  $\alpha$  is the parameter of regression. So, we can write the equation (13) as follows:

$$[-\ln C(F_{S_2}, F_{E_2})]^\alpha = (-\ln F_{S_2})^\alpha + (-\ln F_{E_2})^\alpha \quad (14)$$

For simplicity, we denote  $C(F_{S_2}, F_{E_2})$  by  $C$ . Now, we take the partial derivative with respect to  $F_{S_2}$  of both sides of the equation (14) to get:

$$\frac{(-\ln C)^{\alpha-1}}{C} \frac{\partial C}{\partial F_{S_2}} = \frac{(-\ln F_{S_2})^{\alpha-1}}{F_{S_2}}$$

Then, we extract the first partial derivative denoted  $C_1 = \partial C / \partial F_{S_2}$  as follows:

$$\begin{aligned} C_1 &:= C_1(F_{S_2}, F_{E_2}) \\ &:= \frac{\partial C}{\partial F_{S_2}} \\ &= \frac{\partial C(F_{S_2}, F_{E_2})}{\partial F_{S_2}} \\ &= \left(\frac{\ln F_{S_2}}{\ln C}\right)^{\alpha-1} \frac{C}{F_{S_2}} \end{aligned} \quad (15)$$

By symmetry, we get also the second partial derivative as follows:

$$C_2 := \frac{\partial C}{\partial F_{E_2}} = \left(\frac{\ln F_{E_2}}{\ln C}\right)^{\alpha-1} \frac{C}{F_{E_2}} \quad (16)$$

Using the conditional distribution function from (9) and (15), we obtain the  $p$ -th quantile,  $e_{2,p}(s)$  as follows:

$$\begin{aligned} p &= F_{E_2}(e_{2,p}(s) | S_2 = s) \\ &= \frac{\phi^{-1}\{c_1(s, e_{2,p}(s)) + \phi(F_{E_2}(e_{2,p}(s)))\}}{\phi^{-1}(c_1(s, e_{2,p}(s)))} \\ &= \frac{\phi^{-1}\{\phi(F_{S_2}(s)) + \phi(F_{E_2}(e_{2,p}(s)))\}}{\phi^{-1}(\phi(F_{S_2}(s)))} \\ &= C_1(F_{S_2}(s), F_{E_2}(e_{2,p}(s))) \end{aligned} \quad (17)$$

where  $C_1$  is the first partial derivative of the Archimedean copula,  $\phi$  is the generator of the Gumbel copula, i.e.,  $\phi(t) = (-\ln t)^\alpha$ , and  $F_{E_2}(e_{2,p}(s))$  is the distribution function of the infected compartment where  $e_{2,p}(s)$  denotes the  $p$ -th conditional quantile conditioned by  $S_2 = s$ .

Notice that, in the case of the Gumbel-Hougaard copula, one can use Equation (15) to get the distribution function  $F_{E_2}(e_{2,p})$ :

$$p = \left[ \frac{\ln(F_{S_2})}{\ln C(F_{S_2}, F_{E_2}(e_{2,p}))} \right]^{\alpha-1} \frac{C(F_{S_2}, F_{E_2}(e_{2,p}))}{F_{S_2}} \quad (18)$$

Finding the  $p$ -th quantile  $e_{2,p}$  permits to divide the population into several parts where the

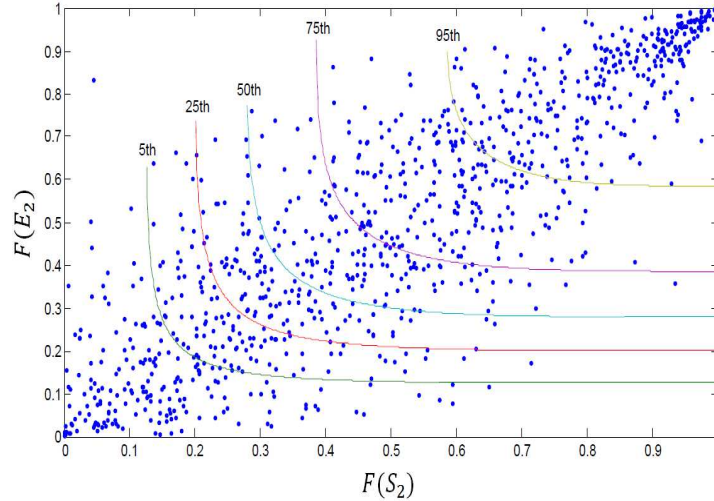


Figure IV.9: Interaction between the distribution functions of  $S_2$  and  $E_2$  using the Gumbel copula with  $\alpha = 3$ , using the quantile regression (of  $E_2$  on  $S_2$ ) curves.

individuals in each part are dependent. In Figure IV.9, there are 500 individuals in each compartment and then there are 1000 individuals in the population. Taking into account the interaction between the two distribution functions  $F_{S_2}$  and  $F_{E_2}$  of the two compartments susceptible and infected respectively, we can divide the population, with the conditional quantiles  $e_{2,p}$ 's, taken for different  $p$ 's, for instance for  $p \in \{0.25, 0.5, 0.75\}$ . For tracing the separation lines of the Figure IV.9, we must calculate the former conditional quantiles related to the two compartments susceptible and infected. Then, by computing these quantiles, we can analyze each part to find the dependence between the two vector compartments. With the relations between the two compartments, we can find links between the various parameters studied systems, which leads us to put conditions on these parameters to lower the basic reproduction rate  $R_0$ .

#### IV.4 Quantiles from the distribution functions

The purpose of this Section is to calculate the  $p$ -th percentile, as in equation (18), from the distribution functions of the susceptible and infective host populations sizes  $F_{S_1}$  and  $F_{I_1}$ , by writing:

$$F_{S_1}(s) = 1 - R_\gamma(s) \quad (19)$$

$$F_{I_1}(k) = 1 - R_\nu(k) \quad (20)$$

where  $R_\gamma(s)$  and  $R_\nu(k)$  are the rests at order  $k$  of the distribution functions  $F_{S_1}$  and  $F_{I_1}$  respectively.

Then:

$$\ln(F_{S_1}(s)) = \ln(1 - R_\gamma(s))$$

$$\ln(F_{I_1}(k)) = \ln(1 - R_\nu(k))$$

Consequently, using the Gumbel copula in (13), we will have :

$$C\left(F_{S_1}(s), F_{I_1}(k)\right) = \exp \left\{ - \left[ [-\ln(1 - R_\gamma(s))]^\alpha + [-\ln(1 - R_\nu(k))]^\alpha \right]^{\frac{1}{\alpha}} \right\} \quad (21)$$

or by a second order Taylor's expansion we will obtain when  $k$  tends to infinity:

$$\ln(1 - R_\gamma(s)) = -R_\gamma(s) - \frac{1}{2}R_\gamma^2(s) + o(R_\gamma^2(s))$$

and

$$\ln(1 - R_\nu(k)) = -R_\nu(k) - \frac{1}{2}R_\nu^2(k) + o(R_\nu^2(k))$$

Then :

$$\begin{aligned} C(F_{S_1}(s), F_{I_1}(k)) &= \exp \left\{ - \left[ \left[ R_\gamma(s) + \frac{1}{2}R_\gamma^2(s) + o(R_\gamma^2(s)) \right]^\alpha + \left[ R_\nu(k) + \frac{1}{2}R_\nu^2(k) + o(R_\nu^2(k)) \right]^\alpha \right]^{\frac{1}{\alpha}} \right\} \\ &= \exp \left\{ - \left[ R_\gamma^\alpha(s) \left( 1 + \frac{\alpha}{2}R_\gamma(s) + o(R_\gamma(s)) \right) + R_\nu^\alpha(k) \left( 1 + \frac{\alpha}{2}R_\nu(k) + o(R_\nu(k)) \right) \right]^{\frac{1}{\alpha}} \right\} \\ &= \exp \left\{ - \left[ R_\gamma^\alpha(s) + R_\nu^\alpha(k) \right]^{\frac{1}{\alpha}} \left[ 1 + \frac{\alpha}{2} \frac{R_\gamma^{\alpha+1}(s) + R_\nu^{\alpha+1}(k)}{R_\gamma^\alpha(s) + R_\nu^\alpha(k)} \right. \right. \\ &\quad \left. \left. + \frac{R_\gamma^\alpha(s)}{R_\gamma^\alpha(s) + R_\nu^\alpha(k)} o(R_\gamma^\alpha(s)) + \frac{R_\nu^\alpha(k)}{R_\gamma^\alpha(s) + R_\nu^\alpha(k)} o(R_\nu^\alpha(k)) \right]^{\frac{1}{\alpha}} \right\} \\ &= \exp \left\{ - \left[ R_\gamma^\alpha(s) + R_\nu^\alpha(k) \right]^{\frac{1}{\alpha}} \left[ 1 + \frac{1}{2} \frac{R_\gamma^{\alpha+1}(s) + R_\nu^{\alpha+1}(k)}{R_\gamma^\alpha(s) + R_\nu^\alpha(k)} + o(R_\gamma(s) + R_\nu(k)) \right] \right\} \\ &= \exp \left\{ - \left[ R_\gamma^\alpha(s) + R_\nu^\alpha(k) \right]^{\frac{1}{\alpha}} \left[ 1 - \frac{1}{2} \frac{R_\gamma^{\alpha+1}(s) + R_\nu^{\alpha+1}(k)}{(R_\gamma^\alpha(s) + R_\nu^\alpha(k))^{1-\frac{1}{\alpha}}} \right. \right. \\ &\quad \left. \left. - (R_\gamma^\alpha(s) + R_\nu^\alpha(k))^{\frac{1}{\alpha}} o(R_\gamma(s) + R_\nu(k)) \right] \right\} \end{aligned}$$

We denote by:  $L_\alpha(s, k) = R_\gamma^\alpha(s) + R_\nu^\alpha(k)$ . For simplicity, we will replace in what follows  $L_\alpha(s, k)$  by  $L_\alpha$ .

Because  $o(R_\gamma(s) + R_\nu(k))$  tends to 0 when  $s \rightarrow \infty$  and  $k \rightarrow \infty$  then:

$$C(F_{S_1}(s), F_{I_1}(k)) \approx \left[ 1 - \frac{1}{2}L_{\alpha+1}L_\alpha^{\frac{1-\alpha}{\alpha}} \right] e^{-L_\alpha^{\frac{1}{\alpha}}} \quad (22)$$

Thus, the  $p$ -th percentile of the conditional distribution of the random variable  $I_1$  given the random variable  $S_1$  is given by a formula similar to (18) by:

$$p \approx \frac{[\ln(1 - R_\gamma(s))]^{\alpha-1}}{\left[ -L_\alpha^{\frac{1}{\alpha}} + \ln \left( 1 - \frac{1}{2}L_{\alpha+1}L_\alpha^{\frac{1-\alpha}{\alpha}} \right) \right]^{\alpha-1}} \frac{\left[ 1 - \frac{1}{2}L_{\alpha+1}L_\alpha^{\frac{1-\alpha}{\alpha}} \right] e^{-L_\alpha^{\frac{1}{\alpha}}}}{1 - R_\gamma(s)} \quad (23)$$

where  $L_\alpha = R_\gamma^\alpha(s) + R_\nu^\alpha(k_\nu(p))$  and  $k_\nu(p) = \inf \{m \in \mathbb{N}, R_\nu(m) \leq 1 - p\}$ .



On the other hand, because:

$$\ln(1 - R_\gamma(s)) \approx -R_\gamma(s)$$

and

$$\ln\left(1 - \frac{1}{2}L_{\alpha+1}L_\alpha^{\frac{1-\alpha}{\alpha}}\right) \approx -\frac{1}{2}L_{\alpha+1}L_\alpha^{\frac{1-\alpha}{\alpha}}$$

Then:

$$[\ln(1 - R_\gamma(s))]^{\alpha-1} \approx (-R_\gamma(s))^{\alpha-1} \quad (24)$$

and

$$\begin{aligned} \left[-L_\alpha^{\frac{1}{\alpha}} + \ln\left(1 - \frac{1}{2}L_{\alpha+1}L_\alpha^{\frac{1-\alpha}{\alpha}}\right)\right]^{\alpha-1} &\approx \left[-L_\alpha^{\frac{1}{\alpha}} - \frac{1}{2}L_{\alpha+1}L_\alpha^{\frac{1-\alpha}{\alpha}}\right]^{\alpha-1} \\ &\approx (-L_\alpha)^{\frac{\alpha-1}{\alpha}} \left(1 + \frac{1}{2} \frac{L_{\alpha+1}}{L_\alpha}\right)^{\alpha-1} \\ &\approx (-L_\alpha)^{\frac{\alpha-1}{\alpha}} \left(1 + \frac{\alpha-1}{2} \frac{L_{\alpha+1}}{L_\alpha}\right) \end{aligned} \quad (25)$$

Therefore :

$$p \approx (-1)^{\alpha+\frac{1}{\alpha}} \frac{(R_\gamma(s))^{\alpha-1} \left(1 - \frac{1}{2}L_{\alpha+1}L_\alpha^{\frac{1-\alpha}{\alpha}}\right) e^{-L_\alpha^{\frac{1}{\alpha}}}}{1 - R_\gamma(s) L_\alpha^{\frac{\alpha-1}{\alpha}} \left(1 + \frac{\alpha-1}{2} \frac{L_{\alpha+1}}{L_\alpha}\right)} \quad (26)$$

In the case of the independence between  $I_1$  and  $S_1$  in the Gumbel copula,  $\alpha = 1$ , and we obtain:

$$p \approx \frac{\left(1 - \frac{1}{2}L_2\right) e^{-L_1}}{1 - R_\gamma(s)}$$

This formula is general and can be applied to any distribution function. Particularly, in the next section, we will apply it on the Poisson distribution.

#### IV.4.1 A case where the population sizes $S_1$ , $I_1$ , $S_2$ , $E_2$ and $I_2$ are Poissonian random variables

A simple way to obtain a Poissonian behaviour for the sizes  $S$  and  $I$  is to consider only the host population with two compartments of random sizes  $S$  and  $I$ , and assume that the population size  $N = S + I$  is constant (e.g., by supposing that fecundity equilibrates mortality for both susceptibles and infecteds). Moreover, if we suppose that there is at least one event (contact, birth, death or recovering) during the time interval  $(t, t + dt)$ , we get:

$$\begin{aligned} &\mathbb{P}(S(t + dt) = k, I(t + dt) = N - k) = \\ &\mathbb{P}(S(t) = k, I(t) = N - k) \left[ 1 - \beta k(N - k) - fk - \delta(k + 1) - \rho(N - k) \right] dt \\ &+ \left[ \beta(k + 1)(N - k - 1) \mathbb{P}(S(t) = k + 1, I(t) = N - k - 1) \right. \\ &+ (f(k - 1) + r(N - k + 1)) \mathbb{P}(S(t) = k - 1, I(t) = N - k + 1) \\ &\left. + \delta(k + 1) \mathbb{P}(S(t) = k + 1, I(t) = N - k - 1) \right] dt \end{aligned} \quad (27)$$

and by multiplying each equation by  $k$  and summing over  $k$  we get:

$$\begin{aligned}
& \sum_{N \geq k \geq 0} \frac{k \left[ \mathbb{P}(S(t+dt) = k, I(t+dt) = N-k) - \mathbb{P}(S(t) = k, I(t) = N-k) \right]}{dt} \\
&= \frac{d \left[ \sum_{N \geq k \geq 0} k \mathbb{P}(S(t) = k, I(t) = N-k) \right]}{dt} \\
&= \sum_{N \geq k \geq 0} \left[ -\beta k^2 (N-k) \mathbb{P}(S(t) = k, I(t) = N-k) \right. \\
&\quad + \beta k(k+1)(N-k-1) \mathbb{P}(S(t) = k+1, I(t) = N-k-1) \\
&\quad + k(f(k-1) + r(N-k+1)) \mathbb{P}(S(t) = k-1, I(t) = N-k+1) \\
&\quad - (fk^2 + rk(N-k)) \mathbb{P}(S(t) = k, I(t) = N-k) - \delta k^2 \mathbb{P}(S(t) = k, I(t) = N-k) \\
&\quad \left. + \delta k(k+1) \mathbb{P}(S(t) = k+1, I(t) = N-k-1) \right] \\
&= - \sum_{N \geq k \geq 0} \left[ \beta k(N-k) \mathbb{P}(S(t) = k, I(t) = N-k) \right. \\
&\quad \left. - (fk + r(N-k)) \mathbb{P}(S(t) = k, I(t) = N-k) + \delta k \mathbb{P}(S(t) = k, I(t) = N-k) \right] \tag{28}
\end{aligned}$$

Hence, only by supposing  $V(S)$  (variance of  $S$ ), is negligible, we get the following expectation equation:

$$\begin{aligned}
\frac{d\mathbb{E}(S)}{dt} &= f\mathbb{E}(S) - \beta\mathbb{E}(S(N-S))dt - \delta\mathbb{E}(S) + r(N - \mathbb{E}(S)) \\
&\approx \mathbb{E}(S) (-\beta\mathbb{E}(N-S) + r) \quad \text{if } f = \delta
\end{aligned} \tag{29}$$

Then, we get the macroscopic equation for  $S$ :

$$\begin{aligned}
\frac{dS}{dt} &= -\beta S(N-S) + r(N-S) \\
&= -\beta SI + rI
\end{aligned} \tag{30}$$

and in a similar way, the equation for  $I$ :

$$\frac{dI}{dt} = \beta SI - rI \tag{31}$$

By replacing  $k$  by  $k^2$  in the above equations, we obtain the differential system ruling  $\mathbb{E}(S^2(t))$  and  $\mathbb{E}(I^2(t))$ , the not centred moments of order 2, then we get the differential system ruling the variances  $V(S(t))$  and  $V(I(t))$ , where  $V(S(t)) = \mathbb{E}(S^2(t)) - \mathbb{E}^2(S(t))$ , and we can draw the confidence cylinder (or viability tube) around the expected trajectory. By replacing  $k$  by  $s^k$ , we get the differential system ruling the generating function  $\psi(s)$  of the distribution of  $S$ :

$$\frac{d\psi}{dt} \approx k\psi - \beta \frac{d\psi}{ds} \tag{32}$$

where  $k = -\beta N + f - \delta + r$ .

This last equation has a stationary solution:

$$\psi = \exp\left(\frac{K(s-1)}{\beta}\right)$$

and which corresponds to a Poissonian distribution for  $S$ , with expectation and variance equal to  $K/\beta$ , which is in general not negligible and explains the differences observed between the random and the deterministic  $SI$  models.

#### IV.4.2 Quantile regression with Poissonian distributions

In this Section we assume that  $F_{S_1}$  and  $F_{I_1}$  follow Poisson distributions, whose parameters  $\gamma$  and  $\nu$  are respectively the expected values of the susceptible and infective host populations sizes.

The distribution functions  $F_{S_1}(s)$  and  $F_{I_1}(k)$  of the Poisson distributions are defined as:

$$F_{S_1}(s) = \sum_{i=0}^s \frac{\gamma^i}{i!} e^{-\gamma} = 1 - R_\gamma(s) \quad (33)$$

where:  $R_\gamma(s) = \sum_{i=s+1}^{+\infty} \frac{\gamma^i}{i!} e^{-\gamma}$

$$F_{I_1}(k) = \sum_{i=0}^k \frac{\nu^i}{i!} e^{-\nu} = 1 - R_\nu(k) \quad (34)$$

where:  $R_\nu(k) = \sum_{i=k+1}^{+\infty} \frac{\nu^i}{i!} e^{-\nu}$

Then the  $p$ -th quantile is as in equation (26) with  $L_\alpha = R_\gamma^\alpha(s) + R_\nu^\alpha(k(p))$ , and  $k(p) = \inf \left\{ m \in \mathbb{N}, \sum_{i=m+1}^{\infty} \frac{\nu^i}{i!} e^{-\nu} \leq 1 - p \right\}$ .

### IV.5 Conclusion

In this chapter, we have given some definitions of Archimedean copulas used to define conditional quantiles permitting to analysis interactions between compartments of vectors and hosts in a system of transmission of Malaria. By using the bivariate Gumbel copula, we have calculated explicitly conditional quantiles and applied it when the compartments sizes are supposed to be random and Poissonian distributed.

We focus on two of the direct applications of this work:

1. If the calculation of the regression parameter  $\alpha$  of the Gumbel copula allows to the conclusion that  $\alpha$  is close to 1, then the sizes of the concerned populations (e.g., susceptibles and infecteds) are not concordant or uncorrelated, which is in favour of a total population size non constant, hypothesis rarely done in the epidemiologic studies, especially in Malaria spread modelling (cf. [22, 29]), which leads to incorporate a refined demographic part in the epidemic model.

2. The calculation of conditional quantiles with the general formula from the empiric distribution functions of the observed population sizes allows to reconstruct their density, hence allows to test a posteriori their Poissonian character, which only authorizes the use of the specific simplified formulae. The Poisson distribution hypothesis has the interest to link the epidemic interaction mechanism to the stochastic molecular kinetics (cf. [30, 36]), which is a way to model the contagious contacts.

## IV.6 Appendix

### IV.6.1 Appendix A

$$\begin{aligned}
\tau &= \mathbb{P}\{(X_1 - X_2)(Y_1 - Y_2) > 0\} - \mathbb{P}\{(X_1 - X_2)(Y_1 - Y_2) < 0\} \\
&= 2\mathbb{P}\{(X_1 - X_2)(Y_1 - Y_2) > 0\} - 1 \\
&= 2\mathbb{P}\{(X_1 > X_2 ; Y_1 > Y_2) \cup (X_1 < X_2 ; Y_1 < Y_2)\} - 1 \\
&= 2[\mathbb{P}\{(X_1 > X_2 ; Y_1 > Y_2)\} + \mathbb{P}\{(X_1 < X_2 ; Y_1 < Y_2)\}] - 1 \\
&= 4\mathbb{P}\{(X_1 > X_2 ; Y_1 > Y_2)\} - 1 \\
&= 4 \int_x \int_y \mathbb{P}\{X_2 \leq x ; Y_2 \leq y \mid X_1 = x ; Y_1 = y\} f_{XY}(x, y) dx dy - 1 \\
&= 4 \int_x \int_y F_{XY}(x, y) f_{XY}(x, y) dx dy - 1 \\
&= 4 \int_x \int_y C(F_X(x), F_Y(y)) f_{XY}(x, y) dx dy - 1
\end{aligned} \tag{35}$$

where  $F_{XY}$  (resp.  $f_{XY}$ ) is the joint distribution function (resp. the joint density function) of  $X$  and  $Y$ .

In making the change of variables  $u = F_X(x)$  and  $v = F_Y(y)$ , we get:

$$\tau = 4 \int_0^1 \int_0^1 C(u, v) c(u, v) du dv - 1$$

where  $c(u, v) = \partial^2 C(u, v) / \partial u \partial v$

### IV.6.2 Appendix B

We have that:

$$\rho = 3 [\mathbb{P}((X_1 - X_2)(Y_1 - Y_3) > 0) - \mathbb{P}((X_1 - X_2)(Y_1 - Y_3) < 0)]$$

So, according to the Appendix A, we can write:

$$\rho = 3 \left[ 4 \int_x \int_y C(F_X(x), F_Y(y)) dC(F_X(x), F_Y(y)) - 1 \right]$$

In making the change of variables  $u_1 = F_X(x)$  and  $u_2 = F_Y(y)$ , we get:

$$\rho = 12 \int_0^1 \int_0^1 C(u_1, u_2) dC(u_1, u_2) - 3$$

where  $dC(u_1, u_2) = \partial^2 C(u, v) / \partial u_1 \partial u_2$

Since  $X_2$  and  $Y_3$  are independent random variables, then  $C(u_1, u_2) = u_1 u_2$ . So:

$$\rho = 12 \int_0^1 \int_0^1 u_1 u_2 dC(u_1, u_2) - 3$$

### IV.6.3 Appendix C

For both coefficients  $\tau$  and  $\rho$ , the first six properties in Definition IV.1 are obvious from properties of the Kendall's tau and the Spearman's rho. For the seventh property, we note that the Lipschitz condition implies that any family of copulas functions is equicontinuous<sup>2</sup>. Thus the convergence of  $C_n$  to  $C$  is uniform.

#### The Lipschitz condition:

Let  $C'$  be a copula function. If for every  $(u_1, u_2), (v_1, v_2)$  in  $DomC'$

$$|C'(u_2, v_2) - C'(u_1, v_1)| \leq |u_2 - u_1| + |v_2 - v_1|$$

then  $C'$  is uniformly continuous on its domain.

Where  $DomC'$  denotes the domain of definition of  $C'$ .

---

2. Let  $X$  and  $Y$  be two metric spaces, and  $F$  a family of functions from  $X$  to  $Y$ . The family  $F$  is equicontinuous at a point  $x_0 \in X$  if for every  $\epsilon > 0$ , there exists a  $\delta > 0$  such that  $d(F(x_0), F(x)) < \epsilon$  for all  $f \in F$  and all  $x$  such that  $d(x_0, x) < \delta$ . The family is equicontinuous on  $X$  if it is equicontinuous at each point of  $X$ .

# Bibliography

- [1] J. Demongeot, Biological boundaries and biological age, *Acta Biotheoretica*, 57 (2009), pp. 397–419.
- [2] W.E. Frees and E.A. Valdez, Understanding Relationships Using Copulas, *Actuarial Research Conference, at the University of Calgary, Calgary, Alberta, Canada*, (1997).
- [3] J. Gaudart, M. Ghassani, J. Mintsa, J. Waku, M. Rachdi, O.K. Doumbo and J. Demongeot, Demographic and spatial factors as causes of an epidemic spread, the copule approach. Application to the retroprediction of the black death epidemy of 1346, *IEEE AINA' 10 and BLSMC' 10. IEEE Press, Piscataway*, (2010), pp. 751-758.
- [4] J. Gaudart, M. Ghassani, J. Mintsa, J. Waku, M. Rachdi and J. Demongeot, Demography and Diffusion in Epidemics: Malaria and Black Death Spread, *Acta Biotheoretica*, 58 (2010), pp. 277–305.
- [5] E.C. Zeeman, Controversy in science: on the ideas of Daniel Bernoulli and René Thom, *Nieuw Arch Wisk*, 11:257 (1993).
- [6] J. Demongeot and J. Waku, Counter-examples for the population size growth in demography, *Math. Pop. Studies*, 12 (2005), pp. 199–210.
- [7] A.W. Marshall and I. Olkin, Families of Multivariate Distribution, *Journal of the Amercian Statistical Association*, (1988), pp. 199–210.
- [8] W.O. Kermack and A.G. McKendrick, Contributions to the mathematical theory of epidemics. II. The problem of endemicity, *Proc. R. Soc. Lond. A*, 138 (1932), pp. 55-83.
- [9] W.O. Kermack and A.G. McKendrick, Contributions to the mathematical theory of epidemics. III. Further studies of the problem of endemicity, *Proc. R. Soc. Lond. A*, 141 (1933), pp. 94-122.
- [10] G. Macdonald, The epidemiology and control of malaria, *Oxford University Press, London*, (1957).
- [11] R.B. Nelsen, Copulas ans Association. In : Advances in Probability Distributions With Given Marginals. Beyond the Copulas, *Kluwer, Dordrecht*, (1991), pp. 51-74.
- [12] R. Ross, An application of the theory of probabilities to the study of a priori pathometry. Part I, *Proc. R. Soc. Lond. A*, 92 (1916), pp. 204-230.

- [13] A.G. McKendrick, Applications of mathematics to medical problems, *Proc. Edinburgh Math. Soc.*, 44 (1925), pp. 1-34.
- [14] T. Roncalli, La gestion des risques financiers, *Economica, Paris*, (2004).
- [15] P. Hougaard, Modelling Multivariate Survival, *Scand. J. Statist.*, 14 (1987), pp. 291–304.
- [16] P. Hougaard, A class of multivariate failure time distributions, *Biometrika*, 73 (1986), pp. 671–678.
- [17] W. Hoeffding, On the distribution of the rank correlation coefficient  $\tau$  when the variates are not independent, *Biometrika*, 34 (1947), pp. 183–196.
- [18] W. Kruskal, Ordinal measures of association, *Journal of the American Statistical Association*, 53 (1958), pp. 814–861.
- [19] R.B. Nelsen, An Introduction to Copulas, Lectures Notes in Statistics, *Springer Verlag, New-York*, 139 (1999).
- [20] R.M. Cooke and O. Morales-Napoles, Competing risk and the cox proportional hazard model, *J. Stat. Plan. Inference*, 136 (2006), pp. 1621-1637.
- [21] D. Beaudoin and L. Lakhhal-Chaieb, Archimedean copula model selection under dependent truncation, *Stat. Med.*, 27 (2008), pp. 4440-4454.
- [22] J.C. Koella and R. Antia, Epidemiological models for the spread of anti-malarial resistance, *Malaria J.*, 2 (2003).
- [23] D.L Smith and F.E. McKenzie, Statics and dynamics of malaria infection in Anopheles mosquitoes, *Malaria J.*, 3 (2004).
- [24] P. Pongsumpun and I.M. Tang, Mathematical Model for the transmission of Plasmodium Vivax Malaria, *Int. J. Math. Models Methods in Appl. Sciences*, 1 (2007), pp. 117–121.
- [25] M.F. Boni, C.O. Buckee and N.J. White, Mathematical Models for a New Era of Malaria Eradication, *PLoS Medicine*, 5 (2008).
- [26] A. Ducrot, S.B. Sirima, B. Som and P. Zongo, A mathematical model for malaria involving differential susceptibility, exposedness and infectivity of human host, *J. Biol. Dynamics*, 3 (2009).
- [27] J. Gaudart, O. Touré, N. Dessay, A.L. Dicko, S. Ranque, L. Forest, J. Demongeot and O.K. Doumbo, Modelling malaria incidence with environmental dependency in a locality of Sudanese savannah area, Mali, *Malaria J.*, 8:61 (2009).
- [28] S. Mandal, R.R. Sarkar and S. Sinha, Mathematical models of malaria -a review, *Malaria J.*, 10 (2011).
- [29] W.A. Brock, J.A. Scheinkman, W.D. Dechert and B. LeBaron, A test for independence based on the correlation dimension, *Econometric Reviews*, 15 (1996), pp. 197–235.

- [30] A.F. Bartholomay, Stochastic models for chemical reactions: I. Theory of the unimolecular reaction process, *Bull. Math. Biophys.*, 20 (1958), pp. 175–190.
- [31] A.F. Bartholomay, Stochastic models for chemical reactions: II. The unimolecular rate constant, *Bull. Math. Biophys.*, 21 (1959), pp. 363–373.
- [32] M. Delbrück, Statistical fluctuations in autocatalytic reactions, *J. Chem. Phys.*, 8 (1940), pp. 120–124.
- [33] D.T. Gillespie, A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions, *Journal of Computational Physics*, 22 (1970), pp. 403–434.
- [34] D.A. McQuarrie, Kinetics of small systems, *J. Chem. Phys.*, 38 (1963), pp. 433–436.
- [35] D.A. McQuarrie, C.J. Jachimowski and M.E. Russell, Kinetics of small systems. II., *J. Chem. Phys.*, 40 (1964), pp. 2914–2921.
- [36] D.A. McQuarrie, Stochastic approach to chemical kinetics, *J. Appl. Prob.*, 4 (1967), pp. 413–478.
- [37] J. Demongeot, O. Hansen, A.S. Jannot and C. Taramasco, Random modelling of contagious (social and infectious) diseases: examples of obesity and HIV and perspectives using social networks, *IEEE Advanced Information Networking and Application (AINA '12, Fukuoka, March 2012)*, *IEEE Proceedings, Piscataway*, (2012), pp. 101–108.





# Conclusions et Perspectives

## Conclusions

L'objectif de cette thèse était d'étudier les modèles compartimentaux, précisément les modèles de transmissions de la Malaria, en supposant que la population n'est pas constante, et en introduisant des paramètres démographiques comme la fécondité, la mortalité, etc. Un autre objectif était d'effectuer une généralisation de certains modèles compartimentaux à classes d'âge, et de trouver des interactions entre les compartiments des hôtes et des vecteurs, en utilisant les fonctions copules et en passant par le modèle de Cox.

Nous avons entamé le premier chapitre par la définition des fonctions copules et par le théorème de Sklar, avec quelques résultats fondamentaux les concernant, et par l'explication de la mesure de dépendance (tau de Kendall et rho de Spearman), et ensuite nous avons défini les copules paramétriques (copule Normale, copule de Student et copule archimédienne). Puis nous avons expliqué le modèle de Cox, en définissant la notion de risque instantané et en introduisant ses principes généraux, et en donnant la définition du modèle de fragilité, qui est lié à la fonction de survie. A la fin de ce chapitre, nous avons approfondi l'analyse des différents systèmes épidémiologiques : modèle de Ross, modèle de Macdonald et modèle de Ross-McKendrick, et nous avons discuté la stabilité des systèmes dynamiques.

Dans le deuxième chapitre, nous avons examiné certaines extensions naturelles des modèles épidémiologiques classiques de Ross-McKendrick-Macdonald, en ajoutant des classes d'âge dans chaque compartiment de vecteurs. Les exemples qui ont été présentés, montrent l'intérêt de l'introduction des classes d'âges dans les équations classiques. Puis, nous avons présenté plusieurs modèles, comme le modèle de Leslie, le modèle d'Usher et le modèle de Cox. Tout cela nous a amené à introduire la notion de fonction copule dans les différents modèles présentés. En outre, nous avons discuté l'équilibre de certains systèmes, et nous avons effectué une étude par simulation, afin de bien valider les modèles et leurs équilibres.

L'approche classique des modèles épidémiologiques est, en effet, faite en supposant que la population étudiée est constante pendant la vague épidémique. Pour cela, dans le troisième chapitre, nous avons présenté deux exemples de modèles compartimentaux, en prenant en compte le fait que la population n'est pas constante, et en supposant qu'il y a des migrations dans les compartiments des hôtes et des vecteurs. Ces deux exemples concernent la Malaria au Mali et la peste noire en Europe au moyen-âge.

Dans le quatrième chapitre, nous avons présenté quelques définitions de copules archimédiennes, afin de les utiliser pour introduire les quantiles conditionnels et donc d'analyser les interactions entre les différents compartiments de vecteurs et d'hôtes, dans les systèmes de

transmission du paludisme.

Dans ce chapitre, deux des applications directes du travail sont :

- Si le calcul du paramètre de régression  $\alpha$  de la fonction copule de Gumbel donne que  $\alpha$  est proche de 1, alors les susceptibles et les infectés de la population étudiée ne sont pas concordants ou ne sont pas corrélés, et cela est en adéquation avec l’hypothèse que la taille de la population n’est pas constante. Cette hypothèse est rarement faite dans les études épidémiologiques, particulièrement dans la modélisation de la propagation du paludisme, ce qui nous a conduit à incorporer une partie démographique dans les modèles épidémiques.
- Le calcul de quantiles conditionnels avec la formule générale des fonctions de distribution empiriques permet de reconstruire leur densité, ce qui donne alors la possibilité de tester leur caractère poissonien, ce qui autorise uniquement l’utilisation des formules spécifiques simplifiées. L’hypothèse de Poisson a l’intérêt de lier le mécanisme d’interaction épidémique à la cinétique moléculaire stochastique, ce qui est une manière de modéliser les contacts contagieux.

## Perspectives

Pour conclure les travaux de cette thèse, nous exposons, dans ce qui suit, quelques développements futurs possibles, en vue d’améliorer et détendre nos résultats. Ces développements sont résumés comme suit :

- L’approche par des fonctions copules, faite dans cette thèse, peut être appliquée à d’autres modèles, épidémiologiques ou non épidémiologiques.
- D’autres fonctions copules peuvent être utilisées, pour effectuer d’autres approches ou d’autres applications en épidémiologie.
- A la suite des exemples que nous avons présentés, quelques autres exemples réalistes, comme les maladies sexuellement transmissibles, peuvent être traités en insistant également sur l’importance de la démographie et de la socio-géographie.
- Les calculs de quantiles conditionnels peuvent nous aider, dans le futur, à trouver des conditions sur les paramètres des modèles étudiés afin de diminuer les contacts contagieux.
- A partir du travail présenté dans cette thèse, nous pourrions lancer des applications dans le cadre spatial afin de trouver une relation entre les différents compartiments des vecteurs et des hôtes dans plusieurs populations.

Nous ne souhaitons pas trop charger cette partie, puisqu’en effet, d’autres perspectives pourraient être traitées sur le long terme.

C’est la première fois que les fonctions copules sont utilisées en épidémiologie, nous espérons donc avoir ouvert une fenêtre, pour qu’il y ait d’autres applications de ces fonctions que celles que nous avons proposées, dans le domaine de l’épidémiologie.

# Bibliographie générale

- L. Abbas, J. Demongeot and N. Glade, Synchrony in reaction-diffusion models of morphogenesis: applications to curvature-dependent proliferation and zero-diffusion front waves, *Philos. Trans. R. Soc. A*, 367 (2009), pp. 4829-4862.
- T. Balenghien, F. Fouque, P. Sabatier and D.J. Bicout, Horse-, bird-, and human-seeking behavior and seasonal abundance of mosquitoes in a West Nile virus focus of southern France, *J. Med. Entomol.*, 43 (2006), pp. 936-946.
- S. Barry and N. Gualde, The biggest epidemics of history, *L'Histoire*, 310 (2006), pp. 38-49.
- A.F. Bartholomay, Stochastic models for chemical reactions: I. Theory of the uni-molecular reaction process, *Bull. Math. Biophys.*, 20 (1958), pp. 175-190.
- A.F. Bartholomay, Stochastic models for chemical reactions: II. The unimolecular rate constant, *Bull. Math. Biophys.*, 21 (1959), pp. 363-373.
- T.P. Baum, N. Pasqual, F. Thuderoz, V. Hierle, D. Chaume, M.P. Lefranc, E. Jouvin-Marche, P. Marche and J. Demongeot, IMG/GenInfo: enhancing V(D)J recombination database accessibility, *Nucleic Acids Res.*, 32 (2004), pp. 51-54.
- D. Beaudoin and L. Lakhal-Chaieb, Archimedean copula model selection under dependent truncation, *Stat. Med.*, 27 (2008), pp. 4440-4454.
- H.K. Beecher, The Powerful Placebo, *Journal of the American Medical Association*, 159 (1955), pp. 1602-1606.
- O.J. Benedictow, The black death 1346-1353: the complete history, *Boydell Press, Woodbridge*, (2004).
- D. Bernoulli, Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'inoculation pour la prévenir, *Mém. Acad. Roy. Sci., Paris*, (1760).
- D. Bernoulli, Réflexions sur les avantages de l'inoculation, *Mercure de France, Paris*, (1760).
- D.J. Bicout and P. Sabatier, Mapping rift valley fever vectors and prevalence using rainfall variations, *Vector-Borne Zoonotic Dis*, 4 (2004), pp. 33-42.
- D.J. Bicout, K. Chalvet-Monfray and P. Sabatier, Infection persistence time of *Aedes* breed-

ing habitats, *Physica A Stat. Mech. Appl.*, 305 (2002), pp. 597-603.

M.F. Boni, C.O. Buckee and N.J. White, Mathematical Models for a New Era of Malaria Eradication, *PLoS Medicine*, 5 (2008).

A. Brandenburg and T. Multamaki, How long can left and right handed life forms coexist?, *Int. J. Astrobiol.*, 3 (2004), pp. 209-219.

W.A. Brock, J.A. Scheinkman, W.D. Dechert and B. LeBaron, A test for independence based on the correlation dimension, *Econometric Reviews*, 15 (1996), pp. 197-235.

J. Brossollet and H. Mollaret, Pourquoi la peste ? Le rat, la puce et le bubon, *Gallimard, Paris*, (1994).

N. Brouhns and M. Denuit, Risque de longévité et rente viagère, *Institut de Statistique Université Catholique, Louvain*, (2001).

L.J. Bruce-Chwatt, Mass drug administration for control of malaria, *The Lancet*, 322 (1983), pp. 668.

N.F. Cantor, In the wake of the plague: the black death and the world it made, *Free Press, New York*, (2001).

E. Carpentier, Une ville devant la peste, Orvieto et la peste noire de 1348, *De Boeck Université, Bruxelles*, (1993).

S.X. Chen, and T. Huang, Nonparametric Estimation of Copula Functions for Dependent Modeling, *Canadian Journal of Statistics*, 35 (2007), pp. 265-282.

G. Christakos, R.A. Olea, M.L. Serre, H.L. Yu and L.L. Wang, Interdisciplinary public health reasoning and epidemic modelling: the case of black death, *Springer, Berlin*, (2005).

S.K. Cohn, The black death: end of a paradigm, *Am. Hist. Rev.*, 107 (2002), pp. 1-54.

R.M. Cooke and O. Morales-Napoles, Competing risk and the cox proportional hazard model, *J. Stat. Plan. Inference*, 136 (2006), pp. 1621-1637.

D.R. Cox, Regression Models and Life-Tables, *Journal of the Royal Statistical Society. Series B (Methodological)*, 34 (1972), pp. 187-220.

G. Dall'Aglio, Fréchet classes and compatibility of distribution functions, *Symposia Mathematica*, 9 (1972), pp. 131-150.

J. D'Alembert, Opuscles Mathématiques, *David, Paris*, (1761).

G.C. de Magny, C. Paroissin, B. Cazelles, M. de Lara, J.F. Delmas and J.F. Guégan, Modeling environmental impacts of plankton reservoirs on cholera population dynamics, *ESAIM*, 14

(2005), pp. 156-173.

P. Deheuvels, Caractérisation complète des lois extrêmes multivariées et de la convergence des types extrêmes, *Publications de l'Institut de Statistique de l'Université de Paris*, 23 (1978), pp. 1–36.

P. Deheuvels, La fonction de dépendance empirique et ses propriétés Un test non paramétrique d'indépendance, *Académie Royale de Belgique, Bulletin de la Classe des Sciences 5me Série*, 65 (1979), pp. 274–292.

P. Deheuvels, Propriétés d'existence et propriétés topologiques des fonctions de dépendance avec applications à la convergence des types pour des lois multivariées, *Comptes Rendus de l'Académie des Sciences de Paris, Série 1*, 288 (1979), pp. 145–148.

P. Deheuvels, A non parametric test for independence, *Publications de l'Institut de Statistique de l'Université de Paris*, 26 (1981), pp. 29–50.

M. Delbrück, Statistical fluctuations in autocatalytic reactions, *J. Chem. Phys.*, 8 (1940), pp. 120–124.

L. Demetrius, Relations between demographic parameters, *Demography*, 16 (1979), pp. 329-338.

J. Demongeot, Thesis, *Université J. Fourier, Grenoble*, (1983).

J. Demongeot, Biological boundaries and biological age, *Acta Biotheor.*, 57 (2009b), pp. 397-419.

J. Demongeot and L. Demetrius, La dérive démographique et la sélection naturelle: Etude empirique de la France (1850-1965), *Population*, 2 (1989), pp. 231-248.

J. Demongeot and J. Fricot, Random fields and renewal potentials, *NATO ASI Serie F*, 20 (1986), pp. 71-84.

J. Demongeot, J. Gaudart, J. Mintsa and M. Rachdi, Demography in epidemics modelling, *Communication on Pure and Applied Analysis*, 11(1) (2012), pp. 61–82.

J. Demongeot, M. Ghassani, M. Rachdi, I. Ouassou and C. Taramasco, Archimedean copula and contagion modeling in epidemiology, *Networks Heterogeneous Media*, submitted for publication, (2012).

J. Demongeot, N. Glade and L. Forest, Liénard systems and potential-Hamiltonian decomposition. I Methodology, *Comptes Rendus Mathématique, série 1, Paris*, 344 (2007a), pp. 121-126.

J. Demongeot, N. Glade and L. Forest, Liénard systems and potential-Hamiltonian decomposition. II Algorithm, *Comptes Rendus Mathématique, Série 1, Paris*, 344 (2007b) pp. 191-194.

- J. Demongeot, N. Glade, A. Moreira and L. Vial, RNA relics and origin of life, *Int. J. Mol. Sci.*, 10 (2009), pp. 3420-3441.
- J. Demongeot, O. Hansen, A.S. Jannot and C. Taramasco, Random modelling of contagious (social and infectious) diseases: examples of obesity and HIV and perspectives using social networks, *IEEE Advanced Information Networking and Application (AINA'12, Fukuoka March 2012)*, *IEEE Proceedings, Piscataway*, (2012), pp. 101–108.
- J. Demongeot and J. Waku, Counter-examples for the population size growth in demography, *Math. Pop. Studies*, 12 (2005), pp. 199–210.
- J.M.O. Depinay, C.M. Mbogo, G. Killeen, B. Knols, J. Beier, J. Carlson, J. Dusho, P. Billingsley, H. Mwambi, J. Githure, A.M. Toure and F.E. McKenzie, A simulation model of African Anopheles ecology and population dynamics for the analysis of malaria transmission, *Malaria J.*, 3:29 (2004).
- K. Dietz, Mathematical Models for Transmission and Control of Malaria. in Principles and Practice of Malariology, *Churchill Livingstone Ed., Edinburgh*, (1988).
- K. Dietz and J.A.P. Heesterbeek, Bernoulli was ahead of modern epidemiology, *Nature*, 408:513 (2000).
- K. Dietz and J.A.P. Heesterbeek, Daniel Bernoullis epidemiological model revisited, *Math. Biosci.*, 180 (2002), pp. 1-21.
- C. Doliger, Démographie, fécondité et croissance économique en France, Thesis, *Université Louis Pasteur, Strasbourg*, (2006).
- O.K. Doumbo, It takes a village: medical research and ethics in Mali, *Science*, 307 (2005), pp. 679-681.
- M. Drancourt, G. Aboudharam, M. Signoli, O. Dutour and D. Raoult, Detection of 400-year-old *Yersinia pestis* DNA in human dental pulp: an approach to the diagnosis of ancient septicemia, *Proc. Natl. Acad. Sci. USA*, 95 (1998), pp. 2637-2640.
- A. Ducrot, S.B. Sirima, B. Som and P. Zongo, A mathematical model for malaria involving differential susceptibility, exposedness and infectivity of human host, *J. Biol. Dynamics*, 3 (2009).
- J. Dutertre, Etude d'un modèle épidémiologique appliqué au paludisme, *Ann. Soc. Belge Méd. Trop.*, 56 (1976), pp. 127-141.
- C. Dye and G. Hasiberder, Population dynamics of mosquito-borne disease: effect of flies which bite some people more frequently than others, *Trans. R. Soc. Trop. Med. Hyg.*, 80 (1980), pp. 69–77.
- E.A. Eckert, The black death and the transformation of the West, *Bull. Hist. Med.*, 74 (2000), pp. 356-357.

- M. Finepe, Superinfection - A problem in formulating a problem (an historical critique of Macdonald's theory), *Trop. Dis. Bull.*, 72 (1975), pp. 475–488.
- R.A. Fisher, The wave of advance of advantageous genes, *Ann. Eugenics*, 7 (1937), pp. 353–369.
- L. Forest, N. Glade and J. Demongeot, Liénard systems and potential-Hamiltonian decomposition. Applications, *C. R. Acad. Sci. Biol.*, 330 (2007), pp. 97–106.
- M. Fréchet, Généralisations du théorème des probabilités totales, *Fundamenta Mathematicae*, 25 (1935), pp. 379–397.
- M. Fréchet, Sur les tableaux de corrélation dont les marges sont données. *Annales de l'Université de Lyon, Section A*, 9 (1951), pp. 53–77.
- M. Fréchet, Sur les tableaux de corrélation et les programmes linéaires. *Revue de l'Institut International de Statistique*, 25 (1957), pp. 23–40.
- M. Fréchet, Remarques au sujet de la note précédente. *Comptes Rendus de l'Académie des Sciences de Paris, Série 1 Math*, 246 (1958), pp. 2719–2720.
- W.E. Frees and E.A. Valdez, Understanding Relationships Using Copulas, *Actuarial Research Conference, at the University of Calgary, Calgary, Alberta, Canada*, (1997).
- R. Gani and S. Leach, Epidemiologic determinants for modelling pneumonic plague outbreaks, *Emerg. Infect. Dis.*, 10 (2004), pp. 608–614.
- C. Garret-Jones, The human blood index of malaria vectors in relation to epidemiological assessment, *Bull. World Health Organ.*, 30 (1964), pp. 241–261.
- J. Gaudart, R. Giorgi, B. Poudiougou, S. Ranque, O.K. Doumbo and J. Demongeot, Spatial cluster detection: principle and application of different general methods, *Rev. Epidemiol. Santé Publique*, 55 (2007), pp. 297–306.
- J. Gaudart, O. Touré, N. Dessay, A.L. Dicko, S. Ranque, L. Forest, J. Demongeot and O.K. Doumbo, Modelling malaria incidence with environmental dependency in a locality of Sudanese savannah area, Mali, *Malaria J.*, 8:61 (2009).
- J. Gaudart, M. Ghassani, J. Mintsa, J. Waku, M. Rachdi, O.K. Doumbo and J. Demongeot, Demographic and spatial factors as causes of an epidemic spread, the copule approach. Application to the retroprediction of the black death epidemic of 1346, *IEEE AINA' 10 and BLSMC' 10. IEEE Press, Piscataway*, (2010), pp. 751–758.
- J. Gaudart, M. Ghassani, J. Mintsa, J. Waku, M. Rachdi and J. Demongeot, Demography and Diffusion in Epidemics: Malaria and Black Death Spread, *Acta Biotheoretica*, 58 (2010), pp. 277–305.
- C. Genest and J. MacKay, Copules archimédiennes et familles de lois bidimensionnelles



dont les marges sont données, *American Statistician*, 40 (1986), pp. 280–283.

C. Genest and J. MacKay, The joy of copulas : Bivariate distributions with uniform marginals, *Canadian Journal of Statistics*, 14(2) (1986), pp. 145–159.

C. Genest and L-P. Rivest, A characterization of Gumbel's family of extreme value distributions, *Statistics and Probability Letters*, 8(3) (1989), pp. 207–211.

C. Genest and L.P. Rivest, Statistical inference procedures for bivariate Archimedean copulas, *Journal of the American Statistical Association*, 88 (1993), pp. 1034–1043.

D.T. Gillespie, A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions, *Journal of Computational Physics*, 22 (1970), pp. 403–434.

N. Glade, L. Forest and J. Demongeot, Liénard systems and potential-Hamiltonian decomposition. III Applications in biology, *Comptes Rendus Mathématique, série 1, Paris*, 344 (2007), pp. 253-258.

W. Hoeffding, Masstabinvariante Korrelationstheorie, *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin*, 5 (1940), pp. 181–233.

W. Hoeffding, On the distribution of the rank correlation coefficient  $\tau$  when the variates are not independent, *Biometrika*, 34 (1947), pp. 183–196.

M. Horie, T. Honda, Y. Suzuki, Y. Kobayashi, T. Daito, T. Oshida, K. Ikuta, P. Jern, T. Gojobori, J.M. Coffin and K. Tomonaga, Endogenous non-retroviral RNA virus elements in mammalian genomes, *Nature*, 463 (2010), pp. 84-88.

R. Horrox, The black death, *Manchester University Press, Manchester*, (1994).

P. Hougaard, Modelling Multivariate Survival, *Scand. J. Statist.*, 14 (1987), pp. 291–304.

P. Hougaard, A class of multivariate failure time distributions, *Biometrika*, 73 (1986), pp. 671–678.

J. Huang, E.D. Walker, P.E. Otienoburu, F. Amimo, J. Vulule and J.R. Miller, Laboratory tests of oviposition by the african malaria mosquito, *Anopheles gambiae*, on dark soil as influenced by presence or absence of vegetation, *Malaria J.*, 5:88 (2006).

J. Kelly, The great mortality: an intimate history of the black death, the most devastating plague of all time, *Harper Collins, New York*, (2005).

W.O. Kermack and A.G. McKendrick, Contributions to the mathematical theory of epidemics II. The problem of endemicity, *Proc. R. Soc. Lond. A*, 138 (1932), pp. 55-83.

W.O. Kermack and A.G. McKendrick, Contributions to the mathematical theory of epidemics III. Further studies of the problem of endemicity, *Proc. R. Soc. Lond. A*, 141 (1933), pp. 94-122.

- K. Khan, J. Arino, W. Hu, J. Raposo, J. Sears, F. Calderon, C. Heidebrecht, M. Macdonald, J. Liauw, A. Chan and M. Gardam, Spread of a novel Influenza A (H1N1) virus via global airline transportation, *N. Engl. J. Med.*, 361 (2009), pp. 212-214.
- J.C. Koella and R. Antia, Epidemiological models for the spread of anti-malarial resistance, *Malaria J.*, 2 (2003).
- W. Kruskal, Ordinal measures of association, *Journal of the American Statistical Association*, 53 (1958), pp. 814–861.
- T. Kulpa, Archimedean copulas and their application to the individual risk model, *in : Int. Congress Finance Modelling, Rizeni*, (2008).
- J.H. Lambert, Beiträge zum Gebrauche der Mathematik und deren Anwendung, *Dritter Theil, Berlin*, (1772).
- J.P. LaSalle, Stability theory for ordinary differential equations. *J. Differ. equations*, 41 (1968), pp. 57–65.
- G.J. L'Epine, Rapport de six des douze commissaires (contre l'inoculation), *Quillau, Paris*, (1764), pp. 40-43.
- La peste humaine en 1997, Relevé Epidémiologique Hebdomadaire, *OMS*, 74 (1999), pp. 340-344.
- P.H. Leslie, On the use of matrices in certain population mathematics, *Biometrika*, 33 (1945), pp. 183-212.
- P. Leslie, Some further notes on the use of matrices in population mathematics, *Biometrika*, 35 (1948), pp. 213–245.
- M.B. Leslie, A matrix model for forest management, *Biometrics*, 25 (1969), pp. 309-315.
- C-H. Ling, Representation of associative functions, *Publicationes mathematicae Debrecen*, 12 (1965), pp. 189–212.
- A.K. Lotka, Contributions to the Analysis of Malaria Epidemiology, *Am. J. Hyg.*, 3 (1923), pp. 1–121.
- G. Macdonald, The epidemiology and control of malaria, *Oxford University Press, London*, (1957).
- G. Macdonald, Dynamics of tropical diseases, *Oxford University Press, London*, (1973).
- S. Mandal, R.R. Sarkar and S. Sinha, Mathematical models of malaria -A Review, *Malaria J.*, 10 (2011).

- A.W. Marshall and I. Olkin, Families of Multivariate Distribution, *Journal of the American Statistical Association*, (1988), pp. 199–210.
- C. Mason, Cox proportional hazard models, <http://www.demog.berkeley.edu/213/Week14/welcome.pdf>, (2005).
- P.L. Maupertuis, Moreau de, Venus physique, in Oeuvres, *Georg. Olms.*, Hildesheim, (1745, reed. 1965).
- N. May, Impartial remarks on the Suttonian method of inoculation, *Tilley, Wheble and Brown, London*, (1770).
- R.M. May and R.M. Anderson, Spatial heterogeneity and the design of immunization programs, *Math. Biosci.*, 72 (1984), pp. 83–111.
- A.G. McKendrick, Applications of mathematics to medical problems, *Proc. Edinburgh Math. Soc.*, 44 (1925), pp. 1–34.
- D.A. McQuarrie, Kinetics of small systems, *J. Chem. Phys.*, 38 (1963), pp. 433–436.
- D.A. McQuarrie, Stochastic approach to chemical kinetics, *J. Appl. Prob.*, 4 (1967), pp. 413–478.
- D.A. McQuarrie, C.J. Jachimowski and M.E. Russell, Kinetics of small systems. II., *J. Chem. Phys.*, 40 (1964), pp. 2914–2921.
- V. Mendez, J. Fort, H.G. Rotstein and S. Fedotov, Speed of reaction-diffusion fronts in spatially heterogeneous media, *Phys. Rev. E*, 68:041105 (2003).
- A. Mischlewski, Un ordre hospitalier au Moyen Âge. Les chanoines réguliers de Saint-Antoine en Viennois, *PUG, Grenoble*, (1995).
- G. Mocellin and I. Experton, Saint-Antoine et l'Ordre des Antonins aux XVIIème et XVIIIème siècles, *Ed. Musée, St-Antoine l'Abbaye*, (1992).
- G. Mocellin-Spicuzza, Chroniques d'une abbaye au Moyen-Age. Guérir l'âme et le corps, *Ed. Musée, St-Antoine l'Abbaye*, (2002).
- L. Molineaux, The pros and cons of modelling malaria transmission, *Trans. R. Soc. Trop. Med. Hyg.*, 79 (1985), pp. 733–747.
- J.A. Murray, Fata variolarum insitionis in Suecia, *Doctoral dissertation, Universtiy Göttingen, Göttingen*, (1763).
- J.D. Murray, Mathematical biology I and II, *Springer Verlag, Berlin*, (2002).
- J.A. Najera, A Critical Review of the field of applications of a mathematical model of malaria

- eradication, *Bull. World Health Organ*, 50 (1974), pp. 449–457.
- P.I. Ndiaye, D.J. Bicout, B. Mondet and P. Sabatier, Rainfall triggered dynamics of *Aedes* mosquito aggressiveness, *J. Theor. Biol.*, 243 (2006), pp. 222-229.
- S.B. Neerinckx, A.T. Peterson, H. Gulinck, J. Deckers and H. Leirs, Geographic distribution and ecological niche of plague in sub-Saharan Africa, *Int. J. Health Geogr.*, 7 (2008), pp. 54.
- R.B. Nelsen, Copulas and Association. In : Advances in Probability Distributions With Given Marginals. Beyond the Copulas, *Kluwer, Dordrecht*, (1991), pp. 51-74.
- R.B. Nelsen, An Introduction to Copulas, Lectures Notes in Statistics, *Springer Verlag, New-York*, 139 (1999).
- D. Oakes, Bivariate Survival Models Induced by Frailties, *J. A. Statist. Asso.*, 84 (1989), pp. 487–493.
- W. Orent, Will the black death return?, *Discover*, 22 (2001), pp. 1-10.
- W. Orent, Plague: the mysterious past and mystifying future of the world's most dangerous disease, *Free Press, New York*, (2004).
- P. Pongsumpun and I.M. Tang, Mathematical Model for the transmission of Plasmodium Vivax Malaria, *Int. J. Math. Models Methods in Appl. Sciences*, 1 (2007), pp. 117–121.
- T. Porphyre, D.J. Bicout and P. Sabatier, Modelling the abundance of mosquito vectors versus flooding dynamics, *Ecol. Modell.*, 183 (2004), pp. 173-181.
- M. Porte, Passion des formes. A René Thom, *ENS Editions, Paris*, (1994).
- M.B. Prentice and L. Rahalison, Plague, *Lancet*, 369 (2007), pp. 1196-1207.
- M. Rachdi, J. Mintsa and J. Demongeot, Stochastic approach in modelling epidemic spread, *IEEE Advanced Information Networking and Application*, (2011), pp. 478–482.
- D. Raoult, G. Aboudharam, E. Crubézy, G. Larrouy, B. Ludes and M. Drancourt, Molecular identification of Suicide PCR of *Yersinia pestis* as the agent of medieval black death, *Proc. Nat. Acad. Sci. USA*, 97 (2000), pp. 12800-12803.
- Y. Renouard, Conséquences et intérêt démographiques de la Peste noire de 1348, *Population*, 3 (1948), pp. 459-466.
- T. Roncalli, La gestion des risques financiers, *Economica, Paris*, (2004).
- R. Ross, An application of the theory of probabilities to the study of a priori pathometry. Part I, *Proc. R. Soc. Lond. A*, 92 (1916), pp. 204-230.

- R. Ross, The prevention of malaria, *Proc. R. Soc. Lond. A*, 92 (1916), pp. 204–230.
- B. Ruce et L.J. Chwatt, Mass drug administration for control of malarial, *Ul. Lancet*, (1983).
- L. Rüschendorf, On the distributional transform, Sklar's Theorem, and the empirical copula process, *J. Stat. Plan. Inference*, 139(11) (2009), pp. 3921-3927.
- J.C. Russell, British medieval population, Thesis of the *Univ. of New Mexico, Albuquerque*, (1948).
- J.C. Russell, Population in Europe. In: Cipolla CM (ed) The Fontana economic history of Europe, vol. I, the middle ages, *Collins/Fontana, Glasgow*, (1972), pp. 25-71.
- C. Schöolzell and P. Friederichs, Multivariate non-normally distributed random variables in climate research introduction to the copula approach, *Nonlin. Processes Geophys.*, 15 (2008), pp. 761–772.
- B. Schweizer and A. Sklar, Probabilistic Metric Spaces, Probability and Applied Mathematics, *North-Holland, New York*, (1983).
- S. Scott and C. Duncan, Return of the black death: the world's greatest serial killer West Sussex, *J. Wiley, New York*, (2004).
- A. Sklar, Fonctions de répartition à  $n$  dimensions et leurs marges, *Publications de l'Institut de Statistique de l'Université de Paris*, 8 (1959), pp. 229-231.
- A. Sklar, Random variables, joint distribution functions and copulas, *Kybernetika*, 9 (1973), pp. 449–460.
- A. Sklar, Random variables, distribution functions and copulas A personal look backward and forward, in L. Ruschendorf, B. Schweizer et M.D. Taylor (Eds.), Distributions with Fixed Marginals and Related Topics, *Institute of Mathematical Statistics, Hayward, CA*, (1996).
- D.L Smith and F.E. McKenzie, Statics and dynamics of malaria infection in Anopheles mosquitoes, *Malaria J.*, 3 (2004).
- N.C. Stenseth, B.B. Atshabar, M. Begon, S.R. Belmain, E. Bertherat, E. Carniel, K.L. Gage, H. Leirs and L. Rahalison, Plague: past, present, and future, *PLOS Med.*, 5 (2008).
- R. Thom, Stabilité structurelle et Morphogenèse, *Benjamin, New York*, (1972).
- F. Thuderoz, M.A. Simonet, O. Hansen, A. Dariz, T.P. Baum, V. Hierle, J. Demongeot, P.N. Marche and E. Jouvin-Marche, From the TCRAD rearrangement quantification to the computational simulation of the locus behavior, *PloS Comp. Biol.*, 6 (2010).
- J. Trembley, Recherches sur la mortalité de la petite vérole, *Mém. Acad. Roy. Sci., Paris*, (1796).

- M.B. Usher, A matrix model for forest management, *Biometrics*, 25 (1969), pp. 309-315.
- G. Villani, Florence, 1336-1338, the chronicle of Giovanni Villani. In: Medieval trade in the mediterranean world, Lopez, R.S. and I.W. Raymond Trads, *Columbia Un. Press, New York*, (2001).
- M. Wheelis, Biological warfare at the 1346 Siege of Caffa, *Emerg. Infect. Dis.*, 8 (2002), pp. 971-975.
- WHO, La peste humaine en 1997. Relevé Epidémiologique Hebdomadaire, *OMS*, 74 (1999), pp. 340-344.
- J.W. Wood, The temporal dynamics of the fourteenth-century black death: new evidence from english ecclesiastical records, *Hum. Biol.*, 75 (2003), pp. 427-448.
- E.C. Zeeman, Controversy in science: on the ideas of Daniel Bernoulli and René Thom, *Nieuw Arch. Wisk.*, 11:257 (1993).
- S.Y. Zhang, L. Yu and P. Daszak, EcoHealth and the black death in the year of the rat, *Ecohealth*, 5 (2008), pp. 99-100.