

# 24ème Forum des jeunes mathématiciennes et mathématiciens

20-22 nov. 2024

Montpellier

France

# Table des matières

Échantillonnage de polytopes et indices de théorie de l'information pour l'analyse de réseaux trophiques marins, Brehelin Jacques [et al.]	1
Characterization of sources of microbiological contamination in rivers using inverse modeling, Cartier Yoann [et al.]	6
A spatio-temporal weather generator for the temperature over France, Cognot Caroline [et al.]	12
Modeling of intracellular compartmentalized GPCR signaling, Darrigade Léo [et al.]	17
Surviving in a shifting and size changing environment in presence of selection, Dhaouadi Nessim [et al.]	21
Modèle dynamique pour la représentation de données d'expression des gènes en cellules uniques, Gaillard Mathilde	25
Sélection de variables par approximation de la norme L0 dans un modèle de Poisson log-normal, Kioye Togo Jean Yves [et al.]	31
Survie des colonies d'Apis mellifera et prix du miel : modélisation et perspectives d'application, Kohli Mathieu	36
Étude mathématique de la propagation et du blocage dans les maladies inflammatoires chroniques de l'intestin, Latrach Saoussen	40
Modélisation spatio-temporelle des voies de signalisation : impact de la compartimentation endosomale et application aux récepteurs aux gonadotrophines., Weckel	

Chloé [et al.] 44

**THE COASTAL-CARBON MODEL, A GENERIC WETLAND MODEL THAT  
PREDICTS CARBON DYNAMICS, Xaus Lucila 50**

Liste des auteurs 59

---

# ÉCHANTILLONNAGE DE POLYTOPES ET INDICES DE THÉORIE DE L'INFORMATION POUR L'ANALYSE DE RÉSEAUX TROPHIQUES MARINS

Jacques Bréhélin <sup>1</sup> & Valérie Girardin <sup>2</sup> & Théo Grente <sup>3</sup>  
& Nathalie Niquil <sup>4</sup> & Philippe Regnault <sup>5</sup>

<sup>1,2</sup> *Laboratoire de Mathématiques Nicolas Oresme, Université de Caen Normandie,  
CNRS-6139, BP5186, Caen, 14032, France. breheli241@unicaen.fr;  
valerie.girardin@unicaen.fr*

<sup>3</sup> *GREYC, Université de Caen Normandie, CNRS 6072, Caen, 14000, France.  
theo.grente@unicaen.fr*

<sup>4</sup> *Laboratoire Biologie des Organismes et Ecosystemes Aquatiques, Université de Caen  
Normandie, CNRS-8067, Esplanade de la Paix, Caen, 14000, France.  
nathalie.niquil@unicaen.fr*

<sup>5</sup> *Laboratoire de Mathématiques de Reims, Université de Reims Champagne-Ardenne,  
CNRS-9008, BP1039, Reims, 51687, France. philippe.regnault@univ-reims.fr*

**Résumé.** Un réseau trophique marin est constitué de l'ensemble des interactions d'ordre alimentaire entre les différentes espèces d'un écosystème marin. Ces interactions sous forme d'échanges de biomasse peuvent être représentées en modélisation linéaire inverse sous formes de graphes pondérés et orientés. Les valeurs des flux inconnues sont soumises à un certain nombre de contraintes écologiques qui déterminent un ensemble de solutions sous forme d'un polytope de grande dimension. L'étude de ce polytope de scénarios possibles est essentielle pour les écologues pour comprendre l'écosystème. La pertinence écologique des scénarios et la stabilité de l'écosystème peuvent être étudiées avec des indices d'analyse de réseaux écologiques (ENA), issus de la théorie de l'information.

Leur étude systématique nécessite un outil rapide et performant d'échantillonnage par méthode de Monte-Carlo par chaînes de Markov (MCMC) de points du polytope. Nous avons pour cela développé une version optimisée et mise à jour du package R `{limSolve}` utilisé par les écologues depuis 2009 malgré sa lenteur d'exécution; voir Van Den Meersche et al. (2009). Le nouveau package `{samplelim}` apporte un gain de temps de calcul important qui permet la prise en compte de systèmes de plus grande dimension, ou l'étude systématique de systèmes relaxés, plus proches des conditions réelles des écosystèmes trophiques; voir Girardin et al. (2024) et Bréhélin (2024). Nos travaux actuels portent sur l'étude de la relaxation de certaines contraintes sur le réseau trophique et de ses conséquences mesurées par les ENA, pour un écosystème de la mer du Nord.

**Mots-clés.** Réseaux métaboliques et trophiques, Modélisation linéaire inverse, Indices d'analyse des réseaux écologiques (ENA), Polytopes de grande dimension, Algorithmes réflexifs MCMC, Package R

---

**Abstract.** A marine trophic network is the set of food-related interactions between the different species in a marine ecosystem. Such interactions, in the form of biomass exchanges, can be represented in inverse linear modeling as a weighted, directed graph. Although the values of these flows are unknown, they are subject to ecological constraints that define a high-dimensional polytope of possible solutions. Its study is therefore crucial for ecologists for a better understanding of the ecosystem. The ecological relevance of scenarios and ecosystem stability can be studied using ecological network analysis (ENA) indices, derived from information theory.

Their systematic study requires a fast, high-performance tool for Markov Chain Monte Carlo (MCMC) sampling of polytope points. To achieve this, we have developed an optimized and updated version of the R package `{limSolve}` used by ecologists since 2009 despite its slow execution; see Van Den Meersche et al. (2009). The new `{samplelim}` package saves a considerable amount of computing time, making it possible to take into account larger systems, or the systematic study of relaxed systems, closer to the actual conditions of trophic ecosystems; see Girardin et al. (2024) and Bréhelin (2024). Our current work focuses on the study of the relaxation of certain constraints on the trophic network and its consequences measured by ENA, for a North Sea ecosystem.

**Keywords.** Metabolic and trophic networks, Linear inverse modeling, Ecological network analysis (ENA) indices, High-dimensional polytopes, Reflexive MCMC algorithms, R package.

## Réseaux trophiques et modélisation

Utilisés notamment en écologie marine, les réseaux trophiques représentent les différentes interactions entre des proies et des prédateurs au sein d'un écosystème. Ces interactions sont généralement exprimés en quantité de biomasse par unité de surface par une unité de temps. Un réseau trophique peut être représenté par un graphe pondéré et orienté  $G = (V, E, f)$ , où  $V$  est l'ensemble des sommets représentant les espèces,  $E \subseteq V \times V$  l'ensemble des arêtes orientées, avec  $n = |E|$ , et  $f = (f_{ij})_{(i,j) \in E}$  le vecteur de valeurs de flux à chaque arête.

Les valeurs de flux sont inconnues mais obéissent à des contraintes établies à partir de connaissances biologiques issues de sources diverses. Parmi ces contraintes biologiques on retrouve notamment les équations de bilan de masse (mass balance equations). Leur principe est que les quantités de matière organique restent constantes au sein d'une espèce. ce qui s'exprime sous forme d'équations,

$$\sum_{j \in V: ji \in E} f_{ij} - \sum_{j \in V: ij \in E} f_{ji} = 0, \quad i \in V.$$

Globalement, ces contraintes prennent la forme d'équations et d'inéquations linéaires dont l'ensemble des solutions peut s'écrire

$$\mathcal{S} = \{f = (f_{ij})_{i,j \in E} \in \mathbb{R}^n : Af = b, Gf \leq h\},$$

---

où  $A$  est une matrice de dimension  $m \times n$ ,  $b$  un vecteur de dimension  $m < n$  avec  $n$  le nombre de flux,  $G$  une matrice de dimension  $k \times n$  et  $h$  un vecteur de dimension  $k$ . C'est une intersection d'hyperplans et de demi-espaces, et les flux ont des valeurs bornées. Un tel ensemble est connu sous le nom de polytope.

Notons que d'autres types de réseaux de fonctionnement de systèmes aboutissent à l'étude de tels polytopes, comme les réseaux métaboliques biochimiques en biologie, ou urbains en gestion des villes, voir Girardin et al. (2024).

## Échantillonnage de polytope et package `samplelim`

Le grand nombre de variables (flux) et les grandes variations de leurs ordres de grandeur (anisotropie), rendent le polytope difficile à étudier analytiquement. Les approches déterministes ont tendance à ne fournir qu'un nombre très restreint de solutions et par conséquent à ne pas être représentatives de la diversité des scénarios possibles, point de vue privilégié par les écologues. Pour cette raison, ils préfèrent des méthodes stochastiques d'échantillonnage, comme les méthodes de Monte Carlo par chaînes de Markov (MCMC).

Parmi ces méthodes, on retrouve notamment les algorithmes réflexifs, comme l'algorithme du miroir (mirror walk algorithm) ; voir Van Den Meersche et al. (2009). Leur principe est d'explorer le polytope en partant d'un point intérieur, puis de tirer une direction et une distance aléatoires et de conserver le point obtenu, en itérant la procédure. Pour ne conserver que les points qui sont dans le polytope, beaucoup d'algorithmes d'échantillonnage utilisent la méthode du rejet, ce qui perd un grand nombre de points et a un coup algorithmique conséquent. Dans les algorithmes réflexifs, lorsque la trajectoire suivie croise une frontière, elle est réfléchiée avec un angle de réflexion égal à l'angle d'incidence, comme la lumière est réfléchiée sur un miroir, et ainsi de suite, de sorte que tous les points tirés restent dans le polytope.

En terme d'implémentation, l'outil le plus utilisé par les écologues était le package R `{limSolve}`; voir Van Den Meersche et al. (2009). Il est cependant relativement ancien et a pour défaut important d'être très lent pour des polytopes anisotropiques de grande dimension. Nous en avons développé une version optimisée et mise à jour nommée `{samplelim}`, voir Girardin et al. (2024) et Bréhélin (2024). Ce package a une interface R pour sa facilité d'utilisation par les écologues, mais les parties les plus importantes du code sont écrites en C++, qui est un langage beaucoup plus performant et permet des gains de temps conséquents. Pour des polytopes de dimension importante, les temps de calcul peuvent passer de plusieurs jours à quelques heures. Cette rapidité d'exécution permet d'étudier systématiquement des systèmes relaxés, un point de vue plus conforme à la réalité des données.

---

## ENA et dégradation d'équations

Pour étudier les écosystèmes, les écologues utilisent des indices d'analyse des réseaux écologiques (ENA), donnant des informations sur la stabilité de l'écosystème et sur les flux les plus importants dans le fonctionnement de l'écosystème. Les ENA sont basés sur la théorie de l'information. Par exemple l'indice de Mc Arthur est équivalent à l'entropie de Shannon, et les divergences du type Kullback-Leibler ont également des équivalents;

- Mc Arthur index 1955 = Entropie de Shannon 1948

$$C(F) = \mathcal{S}(F) = - \sum_{ij \in E} \left( \frac{F_{ij}}{F_{..}} \right) \log_2 \left( \frac{F_{ij}}{F_{..}} \right)$$

- Ascendency 1984 = Information mutuelle 1951

$$A(f) = \sum_{ij \in E} \frac{F_{ij}}{F_{..}} \log_2 \left( \frac{F_{ij} F_{..}}{F_{i.} F_{.j}} \right)$$

Nos travaux en cours portent sur l'étude de l'impact de la relaxation de certaines contraintes sur ces indices et leurs conséquences sur le réseau trophique. Ces relaxations portent sur les contraintes d'égalité, sur lesquelles il peut y avoir des incertitudes de mesure. Celles-ci sont alors remplacées par deux inégalités. Précisément, en dégradant la  $s$ -ième équation, soit  $\sum_{(i,j) \in E} a_{ij}^s f_{ij} = b^s$ , on obtient deux inéquations,

$$\sum_{(i,j) \in E} a_{ij}^s f_{ij} \leq (1 + \delta) b^s \quad \text{et} \quad \sum_{(i,j) \in E} -a_{ij}^s f_{ij} \leq (\delta - 1) b^s,$$

où  $\delta$  est le facteur de relaxation, par exemple  $\delta = 0.1$  ou  $\delta = 0.3$ .

Une étude systématisée des ENA pour différents systèmes relaxés nous a permis de mettre en évidence les interactions ayant le plus gros impact sur le fonctionnement du réseau trophique d'un écosystème de la mer du Nord, zone dite Sylt-Rømø Bight à la frontière entre l'Allemagne et le Danemark, comme on le verra dans l'exposé.

## Bibliographie

- Bréhélin, J. (2024) Vérification et didacticiel d'un package d'échantillonnage de polytopes en analyse de réseaux trophiques. Mémoire de master informatique, Université de Caen Normandie.
- Bréhélin, J., Dien, M., Girardin, V., Grente, T., Niquil, N., et Regnault, P. (2024) Impact of Data Degradation on Information Theory Indices: The Case of the LIM Approach to Food Web Analysis. Poster présenté à la 43ème édition de MaxEnt. [https://theogrente.fr/wp-content/uploads/2024/09/poster\\_maxent2024\\_brehelin.pdf](https://theogrente.fr/wp-content/uploads/2024/09/poster_maxent2024_brehelin.pdf)
- Girardin, V., Grente, T., Niquil, N., et Regnault, P. (2024). Comparing and Updating R packages using MCMC Algorithms for Linear Inverse Modeling of Metabolic Networks, *HAL Preprint* <HAL-04455831>.
- Van Den Meersche, K., Soetaert, K., et Van Oevelen, D. (2009). `xsample()` : An R Function for Sampling Linear Inverse Problems. *Journal of Statistical Software, Code Snippets*, 30, 1–15.

---

# CHARACTERIZATION OF SOURCES OF MICROBIOLOGICAL CONTAMINATION IN RIVERS USING INVERSE MODELING.

Yoann Cartier <sup>1,2</sup> & Arthur Guillot-Le Goff <sup>1,2</sup> & Rémi Carmigniani <sup>2</sup> & David Métivier <sup>3</sup> & Thomas Einfalt <sup>4</sup> & Brigitte Vinçon-Leite <sup>1</sup>

<sup>1</sup> *LEESU, Ecole des Ponts, Univ Paris-Est Créteil, Champs-sur-Marne, France  
yoann.cartier@enpc.fr, arthur.guillot-legoff@enpc.fr, b.vincon-leite@enpc.fr*

<sup>2</sup> *LHSV, Ecole des Ponts, EDF R&D, Chatou, France remi.carmigniani@enpc.fr*

<sup>3</sup> *INRAE, Mistea, Institut Agro, Univ Montpellier, Montpellier, France  
david.metivier@inrae.fr*

<sup>4</sup> *hydro&meteo, Lübeck, Germany einfalt@hydrometeo.de*

**Résumé.** Les rivières fournissent de l'eau douce à l'être humain et abritent de nombreux écosystèmes. Cependant, dans les zones urbaines denses, elles sont contaminées par les déversements d'eaux usées. En étudiant le comportement du réseau d'assainissement en fonction du facteur déclenchant, la pluie, il est possible d'anticiper ces déversements. De plus, en utilisant une méthode inverse, il semble possible de déterminer la quantité de contaminants rejetés au moment du déversement. Cette caractérisation des sources permettrait d'anticiper les contaminations et d'orienter les décisions pour les endiguer.

**Mots-clés.** Modélisation, Rivière, Qualité d'eau, Problème inverse, Statistiques

**Abstract.** Rivers provide freshwater for human use and are home to numerous ecosystems. However, in dense urban areas, they are contaminated by wastewater discharges. By studying the behavior of the wastewater drainage network with regard to the triggering factor, rainfall, it is possible to anticipate these spills. In addition, using an inverse method, it seems possible to determine the quantity of contaminants discharged at the time of the spill. This characterization of sources would make it possible to anticipate contamination and guide decisions to contain it.

**Keywords.** Modeling, River, Water quality, Inverse problem, Statistics

## 1 Context and stakes

Rivers are central to human activities. They provide water for basic needs (drinking, fishing, agriculture), economic development (transportation, hydropower, defense among others in the past), and well-being (recreational bathing, surrounding cooling)(Mishra and Saxena 2024). They are also the habitat of numerous ecosystems (Giller and Malmqvist 1998). However, the water quality of these rivers has deteriorated as a result of human activity. The anthropogenic context defines the type of contaminants to be found in



the watercourses. When it comes to bathing purpose, the European Union commission selected the Fecal Indicator Bacteria (FIB) to characterize water quality (European Parliament and Council of the European 2006). FIB indicate the presence of pathogens, also present in the intestines of warm-blooded animals. These pathogens being harmful to human health. Modeling water quality would enable to better understand and anticipate, with the aim of better managing: (i) bathing times, (ii) pumping times for drinking water, (iii) human pressures on biodiversity.

## 2 Study case

The study site covers the city of Paris and the department of Val-de-Marne (see Figure 3). It is crossed by two rivers: the Marne and the Seine, whose confluence occurs a few kilometers before entering Paris. This area is densely urbanized. Within dense urban watersheds, fecal contamination is due to sewage drainage network, whose functioning is illustrated in Figure 1 (colored symbols make reference to data location in Figure 3).

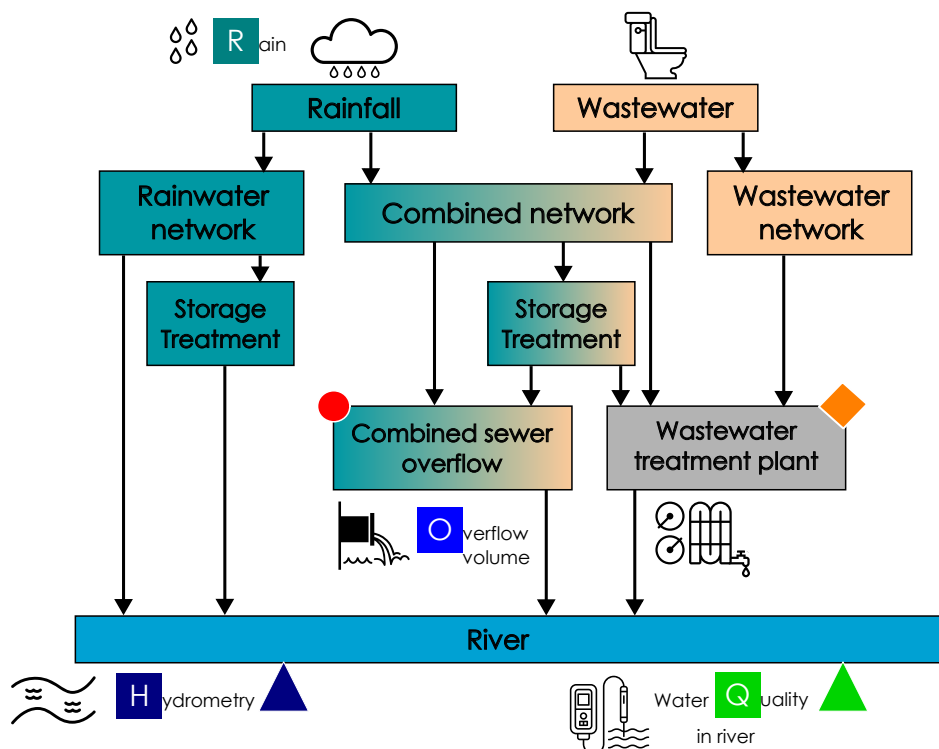


Figure 1: Functioning of wastewater drainage network (modified from Chocat et al. 2021)

The purpose of such network is to transport wastewater from its point of production (homes and workplaces) to the nearest treatment plants. During rainfall events, combined

network (mixing rainwater and wastewater) is saturated. Certain infrastructures relieve it in such cases. Part of the water can be redirected to storage basins. This allows part of the transit volume to be stored, to be returned to the network once the rain event is over. Part of the transit volume can also be discharged directly into surface waters via combined sewer overflows (CSOs). Furthermore, because of the high flow rate entering the treatment plant, which exceeds the nominal flow rate, the quality of the water discharged by these treatment plants is degraded.

### 3 Objectives and methodology

Rainfall events saturate the wastewater drainage system, causing wastewater to be discharged into the receiving environment. This leads to a deterioration in water quality in rivers. The objective of the current work is to **quantify the contribution of different sources of microbiological contamination**. To achieve this objective, we follow the methodological framework illustrated in Figure 2 (colored letters make reference to data description in Figure 1). The first step is to break down the data into events. Contamination occurs during rainfall events that need to be isolated. An event begins with the cause, the rain, and ends with the consequence, the contamination.

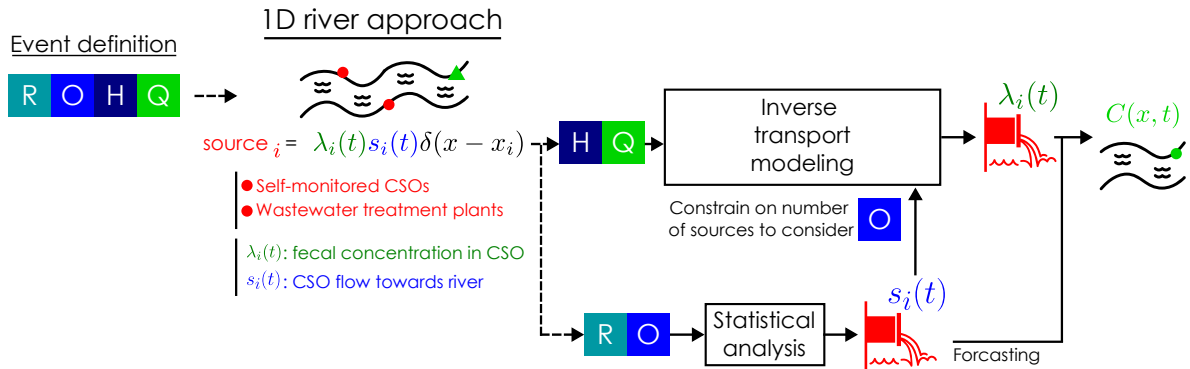


Figure 2: Methodological framework

A one-dimensional river approach is chosen to begin. The sources considered are monitored storm overflows (monitoring criterion based on contamination potential) and wastewater treatment plant discharges. Sources are modeled as follows:

$$source_i = \lambda_i(t) s_i(t) \delta(x - x_i), \quad (1)$$

for source  $i$  in position  $x_i$  discharging contaminated water to the receiving environment at a flow  $s_i(t)$  with a concentration  $\lambda_i(t)$  in contaminants,  $\delta$  a Gaussian distribution. Discharge data are available for the monitored CSOs, but not the contaminant concentration. The modeling is a two-step process in two steps:

- 
1. **Model the function linking flow to precipitation and watershed data.** A first approach consists of a binary output function indicating flow or not. The aim is twofold: predictive capacity and limiting the number of sources to be considered for contaminant concentration studies.
  2. **Estimate the concentration of bacteria in the spill based on downstream microbiological measurements.** This will be based on downstream microbiological measurements.

## 4 Results

A first step was to collect the data shown in Figure 3: concerning sources (location, hourly discharge volume for those in Paris), river (hydrometric data, water quality from ColiMinder instrument (Cazals 2019)), study site (urban catchment areas) and rainfall (radar rain on the catchment areas and a rain gauge to the south-east of Paris). As contamination is due to rainfall events, we broke down the temporal data to reflect this notion of events (based on threshold consideration and a minimum inter-event duration between events).

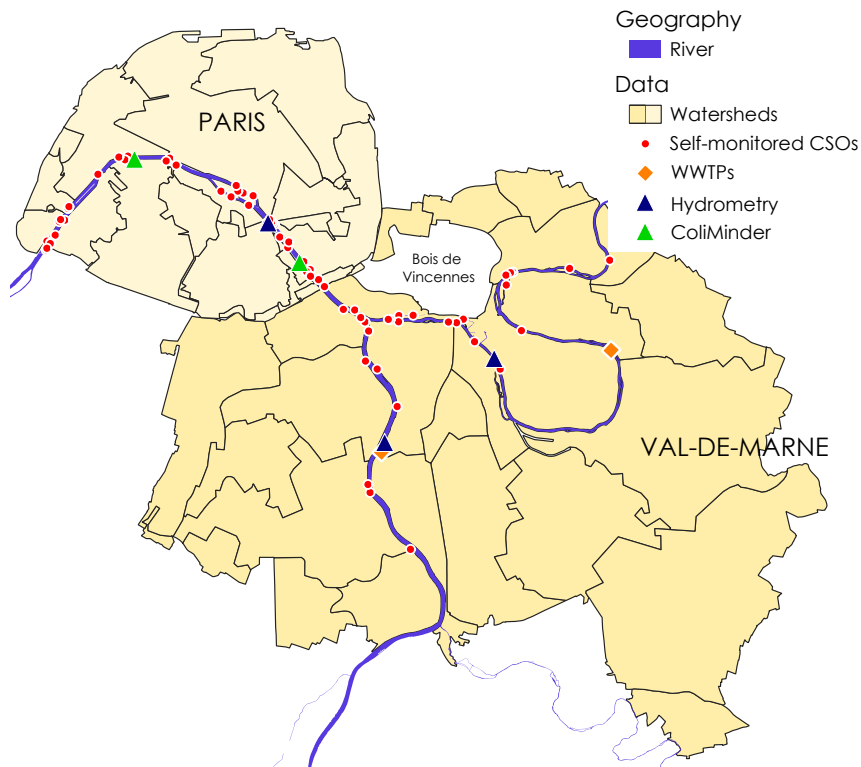


Figure 3: Map of the study site with data location

---

Based on sewer network map, we created a directional graph representing the wastewater transport from one watershed to another. This graph highlights which rainfall variables to consider regarding discharge location. From the event breakdown, we extracted rainfall event characteristics (duration, total accumulation, maximum intensity, average intensity, percentage of dry time within event) and associated consequences (total volume spilled). A first idea from the literature is to look at the episodes on a figure representing maximum rain intensity as a function of rain height (Chun et al. 2024). For the moment, no clear trend is emerging. Statistical exploration of explanatory variables is underway, and some good leads have been identified.

Regarding source contaminant concentration modeling, a first approach is to consider constant  $\lambda_i$  per event. As these data are not available, we propose to retrieve this information using inverse modeling (Barati Moghaddam et al. 2021). It involves studying the cause (wastewater discharge) from the effect (degradation of water quality). We developed a 1D numerical river model, based on the advection-dispersion transport equation:

$$\frac{\partial C}{\partial t} + u \frac{\partial C}{\partial x} - D \frac{\partial^2 C}{\partial x^2} + k_T C = \sum_{i=1}^N \text{source}_i \quad , \quad (2)$$

with  $u$  the flow velocity ( $\text{m s}^{-1}$ ),  $D$  the diffusion coefficient ( $\text{m}^2 \text{s}^{-1}$ ),  $K$  the comprehensive decay coefficient ( $\text{s}^{-1}$ ),  $N$  the number of contaminant sources. This numerical model enables to evaluate methods for solving the inverse problem with synthetic data, before applying it to the real case. The first method, Covariance Matrix Adaptation - Evolution Strategy (Hansen 2023), gives preliminary results.

## References

- Barati Moghaddam, M., Mazaheri, M., and Mohammad Vali Samani, J. (2021). Inverse modeling of contaminant transport for pollution source identification in surface and groundwaters: A review. *Groundwater for Sustainable Development*, 15:100651.
- Cazals, M. (2019). Application d’une nouvelle technologie de détection enzymatique pour le suivi en quasi-temps réel de la dynamique d’*Escherichia coli* dans des eaux récréatives. Master’s thesis, Polytechnique Montréal.
- Chocat, B., Bertrand-Krajewski, J.-L., and Barraud, S. (2021). Pollution des rejets urbains de temps de pluie. In *Techniques de l’Ingénieur*, pages article W 6800 v2, 38 p.
- Chun, S. B., Zhu, Z., and Ghodsi, S. H. (2024). Predicting combined sewer overflow occurrences using rainfall depth and maximum intensity: A case study of Buffalo, USA. *Urban Water Journal*, 21(2):244–250.

---

European Parliament and Council of the European (2006). Directive 2006/7/EC of the European Parliament and of the Council of 15 February 2006 concerning the management of bathing water quality and repealing Directive 76/160/EEC.

Giller, P. S. and Malmqvist, B. (1998). Life in running waters. In Giller, P. S. and Malmqvist, B., editors, *The Biology of Streams and Rivers*, page 0. Oxford University Press.

Hansen, N. (2023). The CMA Evolution Strategy: A Tutorial.

Mishra, R. R. and Saxena, S. (2024). Chapter 1 - Cities and rivers: A symbiotic relationship. In Shinde, V. R., Mishra, R. R., Bhonde, U., and Vaidya, H., editors, *Managing Urban Rivers*, pages 3–24. Elsevier.

---

# A SPATIO-TEMPORAL WEATHER GENERATOR FOR THE TEMPERATURE OVER FRANCE

Caroline Cognot <sup>1&2</sup> & Liliane Bel <sup>2</sup> & Sylvie Parey <sup>1</sup> & David Métivier <sup>3</sup>

<sup>1</sup> EDF R&D, 91120, Palaiseau, France - caroline.cognot@edf.fr

<sup>2</sup> Université Paris-Saclay, AgroParisTech, INRAE, UMR MIA Paris-Saclay, 91120, Palaiseau, France

<sup>3</sup> UMR MISTEA, 34000, Montpellier, France

**Résumé.** Les générateurs de temps sont des simulateurs qui permettent de reproduire la variabilité climatique et générer un grand nombre de situations pour des variables météorologiques selon un modèle statistique ajusté sur les observations. La plupart des générateurs de temps fournissent des simulations pour une ou plusieurs variables météorologiques site par site. Dans ce travail, nous nous proposons de concevoir un générateur spatialisé pour une seule variable météorologique. Nous nous focalisons sur la température. La moyenne et la variance sont décomposées chacune en une tendance et une saisonnalité. La partie stochastique est modélisée par un champ gaussien avec une fonction de corrélation spatio-temporelle non séparable. Ces deux étapes permettent de simuler sur une grille, sur n'importe quelle période de temps. La validation du modèle a été réalisée à la fois sur le jeu de données des stations et sur la grille. En calculant plusieurs indicateurs liés à la structure spatiale, à la structure temporelle ou aux extrêmes, on constate que le générateur fournit des simulations adéquates et permet la génération de séries de températures spatialement cohérentes, y compris dans des extrêmes élevés.

**Mots-clés.** Générateur de temps, processus spatio-temporel, processus Gaussiens

**Abstract.** Weather generators are simulators useful to reproduce climate variability and generate a great number of situations for meteorological variables according to a statistical model fitted on observations. Most weather generators provide simulations for one or several meteorological variables site by site, we aim in this work to design a spatial weather generator for one meteorological variable. We focus on temperature. The mean and variance are each modeled as a trend and a seasonality function. The stochastic part is modeled as a Gaussian field with a non-separable spatio-temporal correlation function. These two steps allow for simulation on a grid, over any time period. The validation of the model was conducted both on the stations dataset and the grid. By computing several indicators related to the spatial structure, the temporal structure or extremes, it is found that the generator provides adequate simulations and allows the generation of spatially coherent temperature series, including high extremes.

**Keywords.** Weather generator, Spatio-temporal process, Gaussian Processes

## 1 Introduction

Weather generators consider the observations of a meteorological variable to be one of the infinite possible trajectories of a stochastic model. The main objective is to allow the simulation of many plausible sequences of a meteorological variable, that share common statistical properties with the observations. The parameters are estimated on the observations. Simulations are fast and allow for a more complete sampling of the climate variables than physical models. Since the first weather generator of Richardson [1981] many others have been proposed using various statistical tools, for univariate and multivariate series. In the case of multisite temperature-focused models, using regular multivariate tools leads to a high number of parameters (for example in the auto-regressive model of Dubrovsky et al. [2020]). Strategies to reduce this amount include Empirical Orthogonal Functions [Sparks et al., 2018], and Gaussian fields [Kleiber et al., 2012]. The latter allows the model to include spatial information. In this work, we

---

focused on building a spatio-temporal Stochastic Weather Generator (SWG) for the temperature using a Gaussian field. The model is built in two steps: first, the mean and variance are represented as a sum of trend and seasonality functions. Trends are depicted as non-parametric functions, whereas the seasonality components are expressed using trigonometric polynomials. Then, a Gaussian field model is fitted on the residuals.

## 2 Methodology

### 2.1 Model

Let  $X(s, t)$  be the temperature at location  $s$  and time  $t$ . Following [Hoang \[2010\]](#),  $X(s, t)$  is written as a combination of deterministic and stochastic terms :

$$X(s, t) = T_m(s, t) + S_m(s, t) + \sqrt{T_{\sigma^2}(s, t)}\sqrt{S_{\sigma^2}(s, t)}Z(s, t), \quad (1)$$

$T_m$  and  $S_m$  are the trend and the seasonality in the mean,  $T_{\sigma^2}$  and  $S_{\sigma^2}$  are the trend and seasonality in the variance.  $Z$  is the stochastic part called residuals. The deterministic terms are interpreted as representing long-term climate variations while the stochastic part represents the intra-annual climate variability.

In order to better capture the long-term dynamics, we use a LOESS (LOcally Estimated Scatterplot Smoothing, [Cleveland \[1979\]](#)) regression which fits linear regression lines locally to the data according to a suitably chosen smoothing parameter. The seasonality terms are modeled as trigonometric polynomials of low order, not exceeding 5, and obtained through linear regression.

We assume the spatio-temporal residuals  $Z(\cdot)$  is a Gaussian 0 mean, stationary, isotropic random field. Under these assumptions, it is entirely described by its covariance function  $C(h, u)$  where  $h$  is a space lag and  $u$  is a time lag. Simplest choices of covariance functions do not perform well, we turn to non-separable models and use the Gneiting-Matérn class of [Gneiting \[2002\]](#), popular for meteorological variables.

$$C(h, u) = \frac{\sigma^2}{\left(\left(\frac{u}{a}\right)^{2\alpha} + 1\right)^\tau} \mathcal{M}\left(\frac{h}{\sqrt{\left(\left(\frac{u}{a}\right)^{2\alpha} + 1\right)^b}}; r; \nu\right), \quad (2)$$

where  $\mathcal{M}(d, r, \nu)$  is the value of the Matérn covariance function at distance  $d$ , with smoothness parameter  $\nu$  and scale parameter  $r$ . A nugget effect is added to the covariance models, representing small scale variability and measurement error. The parameters of this model are estimated using a composite pairwise likelihood approach with a grid search for the separability parameters.

### 2.2 Simulation

Simulations are performed over a grid covering the country, according to (1), where the deterministic components are evaluated on each grid point, and the residuals are simulated as a spatio-temporal Gaussian field with covariance function determined by (2).

The seasonality coefficients  $\beta_m(s), \beta_\sigma(s)$  are obtained by ordinary kriging [[Cressie, 1991](#)] of the coefficients calculated at locations  $\{s_1, \dots, s_{n_s}\}$  using a Gaussian variogram. The annual cycles are then obtained multiplying these predicted coefficients by the corresponding sines and cosines.

---

To provide the trends at each grid point, we first correct the trend in mean to account for elevation and then use kriging with a linear variogram.

Simulation of the gridded Gaussian vector have been done in two ways. The first possibility is to build an auto-regressive model of order  $\ell$  using the conditionnal distribution of the Gaussian process. The other choice was to use the spectral method of [Allard et al. \[2020\]](#).

### 2.3 Validation

To assess the validity of the weather generator, several indicators are computed on the simulations and compared to those obtained from the observations.

For spatially-focused models, easy to compute indicators are cross-correlation between pairs of stations  $(s_i, s_j)$  both in the temperature  $Cor(X_i, X_j)$  and in the residuals  $Cor(Z_i, Z_j)$ . They can be compared with the covariance function.

Another crucial point of interest is the behavior of the generator in high or low temperatures compared to the observations. This can be evaluated by looking at pairwise conditional threshold exceedances probabilities for high or low quantiles, the proportion of locations exceeding a quantile at any given time (spatial hot days), or the distribution of the length of spatial hot days of a given size (heat events).

## 3 Results

The model was fitted on data from 41 weather stations from the ECA&D dataset [[Klein Tank, 2002](#)]. These stations are regularly distributed in space and located in areas of low elevation (less than 500m).

For the covariance model, parameters were estimated by extended season (October, November, December, January, February, March - ONDJFM, April, May, June, July, August, September - AMJJAS) to account for possible remaining seasonality using the R package GeoModels [[Bevilacqua et al., 2024](#)].

Temperature series were simulated 100 times at each of the 41 fitting data points by first simulating 31 summers and 31 winters from the fitted spatio-temporal covariance model (2) and then adding the deterministic parts in (1).

The Gneiting-Matérn model was simulated using an iterative method and a spectral method. The iterative method was also used to simulate from two additional models. A purely temporal version of this covariance model (same values when  $h = 0$ , else 0) represents a spatially uncorrelated model. Its objective is to understand how much of the spatial information is carried by the deterministic trend and seasonality as opposed to the spatial model in the stochastic part. A separable exponential model is used to highlight the shortcomings of a too simple model.

Figure 1 (left) shows all of the simulated pairwise correlation coefficients between stations, on the right for the residuals and on the left for the temperature. It shows that the model without spatial dependence (green) produces simulation that greatly underestimate the correlation in the temperature, while the model with spatial dependence (all other colors) reproduce well the observed values. The separable model performs as well as the non-separable model, except that it slightly overestimates low correlations.

Figure 1 (right) shows the pairwise conditional probabilities of threshold exceedance in the observations compared to the simulations. For all models except the spatially uncorrelated model, the values in simulation are in the range of the observations. For the highest quantile (0.99), the small probabilities are over-estimated, which means the model over-estimates the extremal dependance.

This model can also be used on the grid corresponding to the E-OBS dataset. We generated 40 simulation over the same 31 years, and found that it is able to simulate sequences of spatial hot days with similar spatial size or length in time as the 2003 heat wave.



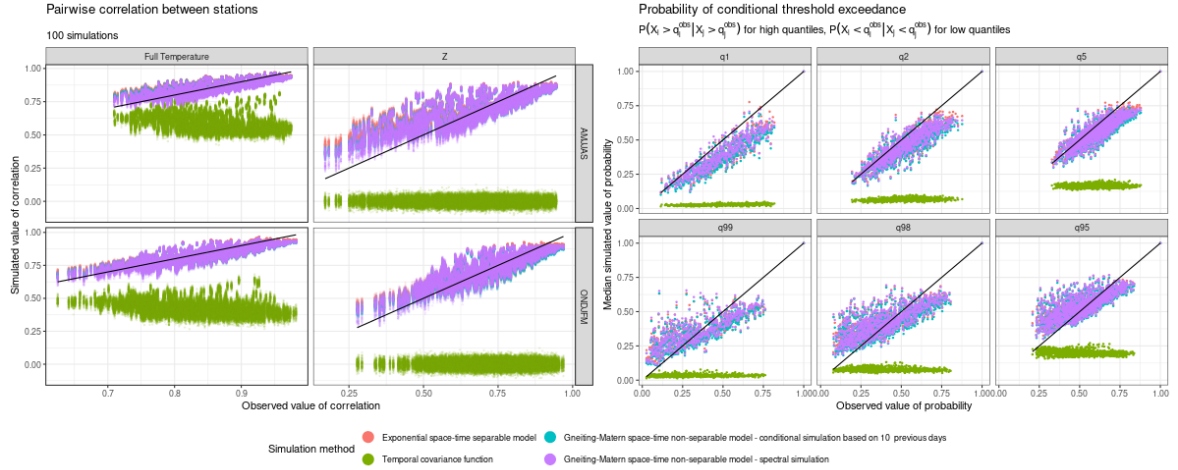


Figure 1: Left : Pairwise correlation coefficient for all stations, first in the full temperature and in the reduced series. Top line represents the result in the extended summer, second line in the winter. Right : Pairwise conditional probabilities of exceeding a given quantile given another station does. q1 denotes the 1% quantile, q2 the 2%, etc.

## 4 Conclusion

The spatio-temporal stochastic weather generator presented in this study is a tool to generate realistic temperature series across a large area. Making use of spatial statistics, it is able to reproduce the spatial correlation, which makes it possible to use it to simulate plausible large-scale events such as heat waves across the study area. This feature makes it possible to use the model to generate large sample for use in impact studies. Future work will combine this model with other variables, such as precipitation.

## References

- D. Allard, X. Emery, C. Lacaux, and C. Lantuéjoul. Simulating space-time random fields with nonseparable Gneiting-type covariance functions. *Statistics and Computing*, 30(5):1479–1495, Sept. 2020. ISSN 0960-3174, 1573-1375. doi: 10.1007/s11222-020-09956-4.
- M. Bevilacqua, V. Morales-Oñate, and C. Caamaño-Carrillo. *GeoModels*, May 2024. URL <https://CRAN.R-project.org/package=GeoModels>.
- W. S. Cleveland. Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association*, 1979.
- N. A. C. Cressie. *Statistics for spatial data*. Wiley series in probability and mathematical statistics. Wiley, New York, 1991. ISBN 978-0-471-84336-8.
- M. Dubrovsky, R. Huth, H. Dabhi, and M. W. Rotach. Parametric gridded weather generator for use in present and future climates: focus on spatial temperature characteristics. *Theoretical and Applied Climatology*, 139(3):1031–1044, Feb. 2020. ISSN 1434-4483. doi: 10.1007/s00704-019-03027-z.

- 
- T. Gneiting. Nonseparable, Stationary Covariance Functions for Space-Time Data. Journal of the American Statistical Association, 97(458):590–600, June 2002. ISSN 0162-1459, 1537-274X. doi: 10.1198/016214502760047113.
- T. T. H. Hoang. Modeling of non-stationary, non-linear time series. Application to the definition of trends on mean, variability and extremes of air temperatures in Europe. phdthesis, Université Paris Sud - Paris XI, Jan. 2010. URL <https://theses.hal.science/tel-00531549>.
- W. Kleiber, R. W. Katz, and B. Rajagopalan. Daily spatiotemporal precipitation simulation using latent and transformed Gaussian processes: PRECIPITATION SIMULATION USING GAUSSIAN PROCESSES. Water Resources Research, 48(1), Jan. 2012. ISSN 00431397. doi: 10.1029/2011WR011105. URL <http://doi.wiley.com/10.1029/2011WR011105>.
- A. a. C. Klein Tank. Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. Int. J. of Climatol., 22, 1441-1453, 2002.
- C. W. Richardson. Stochastic simulation of daily precipitation, temperature, and solar radiation. Water Resources Research, 17(1):182–190, 1981. ISSN 1944-7973. doi: 10.1029/WR017i001p00182. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/WR017i001p00182>.
- N. J. Sparks, S. R. Hardwick, M. Schmid, and R. Toumi. IMAGE: a multivariate multi-site stochastic weather generator for European weather and climate. Stochastic Environmental Research and Risk Assessment, 32(3):771–784, Mar. 2018. ISSN 1436-3259. doi: 10.1007/s00477-017-1433-9.

---

# MODELING OF INTRACELLULAR COMPARTMENTALIZED GPCR SIGNALING

Léo Darrigade <sup>1,2</sup> & Frédéric Jean-Alphonse <sup>1</sup> & Romain Yvinec <sup>1,2</sup>

<sup>1</sup> *Université Paris-Saclay, Inria Saclay-Île-de-France Saclay, Palaiseau, France* <sup>2</sup> *PRC, INRAE, CNRS, Université de Tours, Centre INRAE Val de Loire*

**Abstract.** G-protein coupled receptors (GPCRs) constitute a family of cell membrane receptors crucial for intercellular communication. In mammals, they play a pivotal role in regulating fundamental processes, including reproduction and social interactions. Upon activation by hormones at the cell membrane, GPCRs initiate a cascade of intracellular reactions culminating in their internalization. Notably, GPCRs can subsequently transmit signals from intracellular vesicles, with these signals exerting distinct effects on cell physiology compared to signals originating from the membrane. This study is dedicated to quantitatively analyzing the signals generated by GPCRs within various cellular compartments.

To model the chemical reactions occurring at the membrane and in intracellular vesicles, we construct a measure-valued piecewise deterministic Markov process. The deterministic part encompasses a system of ODEs representing a chemical reactions network (CRN) within each compartment, while the stochastic component accounts for the birth and death of compartments. We study the asymptotic convergence of the process, and we use numerical simulations to glean qualitative insights into the system.

**Keywords.** PDMP, CRN, molecular biology, intracellular trafficking

## 1 Content

### 1.1 Toy-models of GPCR signaling

Let  $m_t, a_t$  be respectively the random quantities of ligand-receptor complex at the membrane and of cAMP. The first model we consider is that the ligand-receptor complex at the plasmic membrane  $m_t$  is internalized randomly and desensitized. We assume that upon endocytosis, a random fraction  $\alpha$  of ligand-receptor complex at the membrane is internalized. For example, this would be the case if the repartition of ligand-receptor complex at the membrane in the plasma membrane is homogeneous and an endosome of size  $\alpha$  with respect to the total surface of the plasma membrane is formed. More precisely, we assume that

$$m_t = m_0 - \int_0^t \int_0^1 \alpha m_s N(ds, d\alpha) \tag{M1}$$

---

with  $N(ds, d\alpha)$  a random Poisson measure of intensity  $n(ds, d\alpha) = \lambda ds A(d\alpha)$ . Note that, with  $(T_n)_n$  being the sequence of jump times of  $N(ds, d\alpha, du)$ , we have :

$$m_{T_n} = m_0 + \sum_{i=1}^n -\alpha_i m_{T_i^-} \quad \Leftrightarrow \quad m_{T_{n+1}} = (1 - \alpha_{n+1}) m_{T_n} \quad \Leftrightarrow \quad m_{T_n} = m_0 \prod_{i=1}^n (1 - \alpha_i).$$

In the second model we consider that complex ligand-receptor can be recycled after internalization and resensitized. For each endosome, the waiting time to recycling is an independent exponential random variable of parameter  $\gamma$ . More precisely, we assume that

$$\begin{aligned} m_t &= m_0 + \int_0^t \int_0^1 \int_{\mathbb{R}^+} (-\alpha m_{s^-} + \mathbb{1}_{\{s+u \leq t\}} \alpha m_{s^-}) N(ds, d\alpha, du) \\ &= m_0 - \int_0^t \int_0^1 \int_{\mathbb{R}^+} \mathbb{1}_{\{s+u > t\}} \alpha m_{s^-} N(ds, d\alpha, du) \end{aligned} \tag{M2}$$

with  $N(ds, d\alpha, du)$  a random Poisson measure of intensity  $n(ds, d\alpha, du) = \lambda ds A(d\alpha) \gamma e^{-\gamma u} du$ . Then

$$m_{T_{n+1}} = \underbrace{(1 - \alpha_{n+1})}_{\text{endocytosis}} \left( m_{T_n} + \underbrace{\sum_{i=1}^n \mathbb{1}_{\{T_n < T_i + U_i < T_{n+1}\}} \alpha_i m_{T_i^-}}_{\text{recycling}} \right).$$

And  $m_0 > 0$ . In both models, we assume that there is no internalized receptor at  $t = 0$ . Moreover, we will have

$$\begin{cases} d_t a_t = k^+ m_t - k^- a_t, \\ a_0 = 0. \end{cases} \tag{1}$$

Models M1 and M2 are inspired by standard models of pharmacokinetics (see Hoare 2020). Then with standard computation on marked Poisson process, we can obtain some explicit formula about the dynamics of the system. Let  $p \in \mathbb{N}$  and  $\tilde{\alpha}_p := \mathbb{E}[(1 - \alpha)^p]$ . In model M1,

$$\mathbb{E}[m_t^p] = \mathbb{E}[m_0^p] e^{-\lambda(1 - \tilde{\alpha}_p)t}. \tag{2}$$

Regarding the dynamic of cAMP, we have the following results. In model M1,

$$\mathbb{E}[a_t] = \mathbb{E}[m_0] \frac{e^{-\lambda \mathbb{E}[\alpha]t} - e^{-k^- t}}{-\lambda \mathbb{E}[\alpha] + k^-}, \tag{3}$$

and

$$\begin{aligned}
\mathbb{E} [a_{T_n}^2] &= \mathbb{E} [a_0^2] \frac{\lambda^n}{(\lambda + 2k^-)^n} + \sum_{i=1}^n \mathbb{E} [m_0^2] \mathbb{E} [(1 - \alpha)^2]^{i-1} \frac{k^{+2} \lambda^{n-i}}{(\lambda + k^-)(\lambda + 2k^-)^{n-i+1}} \\
&\quad + \sum_{i=1}^n \mathbb{E} [a_0] \mathbb{E} [m_0] \mathbb{E} [1 - \alpha]^i \frac{k^+ \lambda^{n+1}}{(\lambda + k^-)^{i+1} (\lambda + 2k^-)^{n-i+1}} \\
&\quad + 2 \sum_{\substack{i,j=1 \\ i < j}}^n \mathbb{E} [m_0^2] \mathbb{E} [(1 - \alpha)^2]^{i-1} \mathbb{E} [1 - \alpha]^{j-i} \frac{k^{+2} \lambda^{n-i}}{(\lambda + k^-)^{j-i+1} (\lambda + 2k^-)^{n-j+1}}.
\end{aligned} \tag{4}$$

In the case of model (M2), we can also obtain explicit formula for the moments of  $m_t$  but not for  $a_t$ .

## 1.2 A general framework with exponentially contracting deterministic flows

We introduce here a general model which couples signaling and trafficking of receptors between the membrane and endosomes. We consider a process  $(z_t = (x_t, \mu_t))_{t \geq 0} \in E := \mathbb{R}_+^d \times \mathcal{M}_P(\mathbb{R}_+^d)$  where  $x_t$  represents chemical quantities at the membrane and  $\mu_t$  chemical quantities in each endosome.

Let  $F_1$  and  $F_2 : \mathbb{R}_+^d \rightarrow \mathbb{R}_+^d$  be the rates of chemical reactions at the membrane and in endosomes respectively. Typically, they are polynomial functions of the vector coordinate and thus, we assume that for each initial condition, there exists a unique global solution of the associated ODE. Let  $\Phi_1(t, x)$  and  $\Phi_2(t, x)$  be the associated flows. Write test functions

$$\Psi_\psi(x, \mu) := \Psi(x, \langle \mu, \psi \rangle)$$

with  $\Psi : \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}$  and  $\psi : \mathbb{R}^d \rightarrow \mathbb{R}$  smooth enough. For a counting measure  $\nu$ , let

$$\nu^{\otimes \downarrow 2}(dy^1, dy^2) = \nu(dy^1)(\nu - \delta_{y^2})(dy^2).$$

The infinitesimal generator of  $z_t$  is

$$\begin{aligned}
L\Psi_\psi(x, \mu) &= \lambda \int_0^1 (\Psi_\psi((1 - \alpha)x, \mu + \delta_{\alpha x}) - \Psi_\psi(x, \mu)) A(d\alpha) \\
&\quad + \kappa \int_{(\mathbb{R}_+^d)^2} (\Psi_\psi(x, \mu + \delta_{y_1+y_2} - \delta_{y_1} - \delta_{y_2}) - \Psi_\psi(x, \mu)) \mu^{\otimes \downarrow 2}(dy_1, dy_2) \\
&\quad + \phi \int_{\mathbb{R}_+^d} \int_0^1 (\Psi_\psi(x, \mu + \delta_{\epsilon y} + \delta_{(1-\epsilon)y} - \delta_y) - \Psi_\psi(x, \mu)) B(d\epsilon) \mu(dy) \\
&\quad + \gamma \int_{\mathbb{R}_+^d} (\Psi_\psi(x + y, \mu - \delta_y) - \Psi_\psi(x, \mu)) \mu(dy) \\
&\quad + \langle \nabla_1 \Psi(x, \langle \mu, \psi \rangle), F_1(x) \rangle + \partial_2 \Psi(x, \langle \mu, \psi \rangle) \int_{\mathbb{R}_+^d} \langle \nabla \psi(y), F_2(y) \rangle \mu(dy).
\end{aligned} \tag{5}$$

---

The first four terms are random and models the dynamic of endosomes. The first term term models formation of a new endosome , the second one coagulation of two endosomes, the third one fragmentation of an endosome, the forth one recycling of an endosome.

We study the long-term behaviour of this model. We make the hypothesis that the deterministic chemical reactions at the membrane and in the endosomes have a single globally exponentially attractive equilibrium, that is to say, for  $i = 1, 2$ , there exists  $\beta > 0$  tel que,  $\forall x, \tilde{x} \in \mathbb{R}^d$ ,

$$\langle x - \tilde{x}, F_i(x) - F_i(\tilde{x}) \rangle \leq -\beta|x - \tilde{x}|^2.$$

Under this hypothesis, we prove the exponential convergence of the process to a single invariant measure. The proof is based on a coupling argument and is inspired by the work of Benaim (2020) which applies it argument to switching ODEs.

## Bibliographie

Benaim, Michel et al (2012), Quantitative ergodicity for some switched dynamical systems, *Electronic Communications in Probability*.

Hoare, Sam R. J. et al (2020). Analyzing kinetic signaling data for G-protein-coupled receptors, *Scientific Reports*.

Jean-Alphonse, Frédéric et al. (Feb. 2014). “Spatially Restricted G Protein-coupled Receptor Activity via Divergent Endocytic Compartments”. In: *Journal of Biological Chemistry*

---

# SURVIVING IN A SHIFTING AND SIZE CHANGING ENVIRONMENT IN PRESENCE OF SELECTION

Nessim Dhaouadi<sup>1</sup> & Matthieu Alfaro<sup>2</sup> & Adel Blouza<sup>3</sup>

<sup>1</sup> *Univ. Rouen Normandie, CNRS, LMRS UMR 6085, F-76000 Rouen, France, nessim.dhaouadi@univ-rouen.fr*

<sup>2</sup> *Univ. Rouen Normandie, CNRS, LMRS UMR 6085, F-76000 Rouen, France, matthieu.alfaro@univ-rouen.fr*

<sup>3</sup> *Univ. Rouen Normandie, CNRS, LMRS UMR 6085, F-76000 Rouen, France, adel.blouza@univ-rouen.fr*

**Résumé.** Ce travail cherche à mieux comprendre les interactions entre une population et son environnement qui est lui aussi dynamique. Nous considérons un modèle décrivant l'adaptation d'une population soumise à une sélection qui évolue périodiquement en temps, se diffusant dans un environnement qui peut translater, s'allonger ou bien rétrécir. Cet ajout permet d'obtenir des résultats très différents du modèle classique dans lequel le domaine est fixe. Dans un premier temps, nous étudions les éléments propres périodiques du problème parabolique sous-jacent, avant de s'intéresser à l'éventuelle survie ou bien extinction de la population qui sont la conséquence des effets combinés du domaine mobile et de la sélection périodique en temps. Finalement, nous proposons une approximation de type 'élément fini'. Nous montrons son caractère bien posé puis effectuons des test numériques qui confirment les résultats théoriques, les complètent et ouvrent des perspectives.

**Mots-clés.** dynamique évolutive, comportement asymptotique, éléments propres périodique parabolique, méthodes des éléments finis. . .

**Abstract.** In this work, we aim at better understanding how the dynamics of a population can be altered by the evolution of the environment itself. We consider a model to describe the adaptation of a diffusing population facing a time periodic selection and whose environment may shift and/or change size, which causes sharp differences with classical results on fixed domains. After studying the underlying periodic parabolic principal eigenelements, we consider the extinction vs. persistence issue, underlying the combined effects of the moving habitat and the time periodic selection. Then, we propose a finite element approximation, prove its well posedness, and perform numerical explorations that confirm, complete the analysis, and raise new issues.

**Keywords.** dynamics of adaptation, long time behavior, periodic parabolic eigenelements, finite difference methods.

---

# 1 Long summary

When it comes to ecology or population dynamics, we often revolve around models which are used to describe a phenomenon that we want to understand better, set in a fix environment. In this work, we investigate the intricacies between our studied population and its dynamic environment. On one hand, the reasons for such moving range boundaries, see [4], could be the consequences of flooding, forest fire, etc. This also connects to the issue of ecological niches shifted by some external factors, such as *Global Warming*, and which has received a lot of attention, see [12], [14], [5] and the references therein. On the other hand, we consider the growth depends on the location in the environment: the maximal growth rate  $r > 0$  is reached at the time dependent position (2), that can turn out to be outside the domain and, away from this optimal position, the growth decays quadratically. Furthermore, we will consider that the optimal position is time periodic to reflect, see [3], some seasonal variations in temperature, water level, etc.

Changing perspective, in the context of evolutionary biology, diffusion models the mutation process, where  $x$  denotes a phenotypic trait (see [8], [9] among many other references). In this framework, the fitness (reproductive success) of a phenotype  $x$  is described by a function that decreases away from the optimum and, here, we use Fisher's geometrical phenotype-to-fitness model, see [15], [11], [1] and the references therein. Precisely, the fitness function admits a unique maximum and decreases quadratically away from it. Furthermore, while the phenotypic space changes through functions  $A$  and  $L$ , the location of the optimum  $x = x_{opt}(t)$  also moves through the periodic function  $\beta$ , and the intensity of selection varies through the periodic function  $\alpha$ . This connects to the issue of moving optima studied in [13], [10], among others, while the conjugate effects of periodic moving optimum and intensity of selection were analyzed in [6, 7].

Hence the following model (1) can be considered in both frameworks of ecology and evolutionary biology and, in the following, we may alternatively refer to habitat, spatial dispersal, growth or to phenotypic space, mutation, selection.

We consider the solutions  $u = u(t, x)$  to the boundary value problem

$$\begin{cases} u_t = du_{xx} + \left( r - \frac{\alpha(t)}{2} (x - x_{opt}(t))^2 \right) u, & t > 0, A(t) < x < A(t) + L(t), \\ u(t, A(t)) = u(t, A(t) + L(t)) = 0, & t > 0, \end{cases} \quad (1)$$

where

$$x_{opt}(t) := A(t) + \beta(t)L(t), \quad (2)$$

and, obviously, we use the shortcuts  $u_t$ ,  $u_{xx}$  for  $\partial_t u$ ,  $\partial_{xx} u$ . This model serves as a linear model to describe the adaptation of a diffusing population, whose mobility is measured by the constant  $d > 0$ . One of the originalities of (1) stands in the fact that it is posed on an interval that may not only shift, through the function  $A \in C^2([0, +\infty))$ , but also change size, through the positive function  $L \in C^2([0, +\infty))$ . Both functions  $\alpha$  which represent the strength of the selection and  $\beta$  the optimal position, are assumed to be



---

Hölder continuous and  $T$ -periodic for some given  $T > 0$ , and  $\alpha$  is positive. The boundary conditions at  $x = A(t)$ ,  $x = A(t) + L(t)$  are of the Dirichlet type. We aim at exploring the long time behavior of the solutions to (1).

In a series of very recent works, Allwright [4, 2, 3] has initiated the analysis of reaction-diffusion problems posed on shifting and/or changing size domain, revealing sharp differences with the case of fixed domains. Among other things (such as the role of competition and/or the dimension), she proved that, when  $\alpha \equiv 0$ , it may happen that the population survives although the habitat is always strictly smaller than the critical one, a phenomenon only possible in presence of a moving habitat.

Nevertheless, in these works, all individuals are assumed to be identical ( $\alpha \equiv 0$ ) and selection is thus ignored. The goal of the present work is to take this phenomenon into account ( $\alpha > 0$ ) in presence of a moving phenotypic space.

The organization of the present work is as follows. In a first Section, we consider the case of a fixed domain, and provide estimates on the periodic principal eigenvalue, revealing contrasted outcomes depending on the fluctuations of the location of the optimum and of the intensity of selection. In a second Section, we consider the moving habitat case and study the issue of extinction vs. survival for the solutions to (1). To do so, we first switch to a fixed domain and construct adequate sub- and supersolutions whose long time behavior can be understood thanks to the first Section. Last, the final Section is devoted to the numerical analysis of (1). We propose a weak formulation in both the continuous and discrete settings. We prove well-posedness of the finite element scheme, and perform numerical explorations that enlight complex possible behaviors for the solution.

## References

- [1] M. ALFARO AND R. CARLES, *Replicator-mutator equations with quadratic fitness*, Proc. Amer. Math. Soc., 145 (2017), pp. 5315–5327.
- [2] J. ALLWRIGHT, *Exact solutions and critical behaviour for a linear growth-diffusion equation on a time-dependent domain*, Proceedings of the Edinburgh Mathematical Society, 65 (2022), pp. 53–79.
- [3] —, *Reaction-diffusion problems on time-periodic domains*, arXiv preprint arXiv:2210.11516, (2022).
- [4] —, *Reaction-diffusion on a time-dependent interval: refining the notion of ‘critical length’*, Commun. Contemp. Math., 25 (2023), pp. Paper No. 2250050, 11.
- [5] H. BERESTYCKI, O. DIEKMANN, C. J. NAGELKERKE, AND P. A. ZEGELING, *Can a species keep pace with a shifting climate?*, Bull. Math. Biol., 71 (2009), pp. 399–429.

- 
- [6] S. FIGUEROA IGLESIAS AND S. MIRRAHIMI, *Long time evolutionary dynamics of phenotypically structured populations in time-periodic environments*, SIAM J. Math. Anal., 50 (2018), pp. 5537–5568.
- [7] —, *Selection and mutation in a shifting and fluctuating environment*, Commun. Math. Sci., 19 (2021), pp. 1761–1798.
- [8] M. KIMURA, *A stochastic model concerning the maintenance of genetic variability in quantitative characters.*, Proc. Natl. Acad. Sci. USA, 54 (1965), pp. 731–736.
- [9] R. LANDE, *The maintenance of genetic variability by mutation in a polygenic character with linked loci*, Genetics Research, 26 (1975), pp. 221–235.
- [10] F. LAVIGNE, *Adaptation of an asexual population with environmental changes*, Math. Model. Nat. Phenom., 18 (2023).
- [11] G. MARTIN AND T. LENORMAND, *The fitness effect of mutations across environments: Fisher’s geometrical model with multiple optima*, Evolution, 69 (2015), pp. 1433–1447.
- [12] A. B. POTAPOV AND M. A. LEWIS, *Climate and competition: the effect of moving range boundaries on habitat invasibility*, Bull. Math. Biol., 66 (2004), pp. 975–1008.
- [13] L. ROQUES, F. PATOUT, O. BONNEFON, AND G. MARTIN, *Adaptation in general temporally changing environments*, SIAM Journal on Applied Mathematics, 80 (2020), pp. 2420–2447.
- [14] L. ROQUES, A. ROQUES, H. BERESTYCKI, AND A. KRETZSCHMAR, *A population facing climate change: joint influences of allee effects and environmental boundary geometry*, Population Ecology, 50 (2008), pp. 215–225.
- [15] O. TENAILLON, *The utility of Fisher’s geometric model in evolutionary genetics*, Annual Review of Ecology, Evolution, and Systematics, 45 (2014), pp. 179–201.

---

# MODÈLE DYNAMIQUE POUR LA REPRÉSENTATION DE DONNÉES D'EXPRESSION DES GÈNES EN CELLULES UNIQUES

Mathilde Gaillard<sup>1</sup> & Ulysse Herbach<sup>1</sup>

<sup>1</sup> *Université de Lorraine, CNRS, Inria, IECL, F-54000, Nancy, France*  
{*mathilde.gaillard, ulysse.herbach*}@inria.fr

**Résumé.** Une première simplification du mécanisme d'expression des gènes considère qu'un gène est transcrit en ARN messager, lui-même traduit à son tour en protéine. Les données issues de cellules individuelles (*single-cell*) ont permis de révéler la présence d'une variabilité biologique entre des cellules de génome et environnement identiques, mettant en lumière non seulement les aspects épigénétiques mais également le caractère stochastique de l'expression des gènes. Cette stochasticité soulève deux enjeux d'un point de vue statistique. Tout d'abord, il faut pouvoir estimer la variabilité entre les cellules (ou entre les gènes) en utilisant par exemple l'entropie de Shannon [Gandrillon et al., 2021]. D'autre part, dans le contexte des réseaux de régulation qui sous-tendent les états et types cellulaires, il s'agit de construire un modèle qui tienne compte à la fois de la stochasticité et de l'interaction des gènes entre eux. On s'intéresse ici à un modèle dynamique d'expression des gènes, formulé comme un processus de Markov déterministe par morceaux (PDMP) et décrivant un nombre arbitraire de gènes en interaction [Herbach et al., 2017]. Ce modèle stochastique reproduit la variabilité biologique mesurée expérimentalement [Ventre et al., 2023], tout en étant compatible avec l'observation d'un pic transitoire dans l'entropie de Shannon au cours de processus de différenciation comme l'hématopoïèse [Dussiau et al., 2022]. Nous justifions d'abord l'existence et l'unicité d'une loi invariante pour ce modèle et estimons la vitesse de convergence vers cette dernière. Cette loi admet alors une expression analytique pour toute une classe de paramètres et peut s'interpréter comme un modèle graphique probabiliste de type champ de Markov caché, avec de bonnes propriétés pour l'inférence des interactions.

**Mots-clés.** Processus de Markov déterministes par morceaux, Distance de Wasserstein, Champs aléatoires de Markov

**Abstract.** A first simplification of the gene expression mechanism considers that a gene is transcribed into messenger RNA, which in turn is translated into protein. Single-cell data have revealed the presence of biological variability between cells of identical genome and environment, highlighting not only epigenetic aspects but also the stochastic nature of gene expression. This stochasticity raises two issues from a statistical perspective. First, we need to be able to estimate the variability between cells (or between genes) using, for example, the Shannon entropy [Gandrillon et al., 2021]. On the other hand, in the context of regulatory networks underlying cell states and types, we need to build a model that takes into account both stochasticity and the interaction of genes with each other. Here we focus on a dynamical model of gene expression, formulated as a piecewise-deterministic Markov process (PDMP) and describing an arbitrary number of interacting genes [Herbach et al., 2017]. This stochastic model is able to reproduce the biological variability measured experimentally [Ventre et al., 2023], while being compatible with the observation of a transient peak in

Shannon entropy during differentiation processes such as hematopoiesis [Dussiau et al., 2022]. We first justify the existence and uniqueness of an invariant distribution for this model, and estimate the rate of convergence to it. This distribution then admits an analytical expression for a wide class of parameters and can be interpreted as a probabilistic graphical model of the hidden Markov field type, with good properties for the inference of interactions.

**Keywords.** Piecewise-deterministic Markov processes, Wasserstein distance, Markov random fields

## 1 Présentation du modèle

On s'intéresse à un modèle approché qui ne tient pas compte de l'ARN messenger, c'est-à-dire qu'il considère uniquement les quantités de protéines, associées à  $n$  gènes, dans une cellule, au cours du temps. Ces quantités sont modélisées par un PDMP noté  $(Z_t)_{t \geq 0}$  et à valeurs dans  $(\mathbb{R}_+)^n$ . En dehors des instants de saut, la quantité de protéines est donnée par la solution d'un système d'équations différentielles. À des instants aléatoires, les coordonnées du processus effectuent des sauts dont la taille est distribuée selon une loi exponentielle dont le paramètre est propre à chaque coordonnée. Le taux de saut dépend de la valeur du processus à chaque instant, traduisant les interactions entre les gènes.

Le processus  $(Z_t)_{t \geq 0}$  est caractérisé par son générateur infinitésimal :

$$Lf(z) = \sum_{i=1}^n \left( -d_{1,i} z_i \partial_{z_i} f(z) + \alpha_i(z) \int_0^{+\infty} (f(z + h e_i) - f(z)) c_i e^{-c_i h} dh \right) \quad (1)$$

où  $(e_1, \dots, e_n)$  est la base canonique de  $\mathbb{R}^n$  et où  $d_{1,i}$  et  $c_i$  correspondent respectivement au taux de dégradation et au paramètre de la loi des sauts du gène  $i$ . On considère l'application  $\alpha$  définie par

$$\alpha : \begin{cases} \mathbb{R}^n & \longrightarrow \mathbb{R}^n \\ x & \longmapsto (\alpha_1(x), \dots, \alpha_n(x)) \end{cases}$$

et on suppose que  $\alpha$  est  $\ell$ -lipschitzienne :

$$\forall x, y \in \mathbb{R}^n, \quad \sum_{i=1}^n |\alpha_i(x) - \alpha_i(y)| = \|\alpha(x) - \alpha(y)\|_1 \leq \ell \|x - y\|_1.$$

On suppose également que pour tout  $i \in \{1, \dots, n\}$ ,  $\alpha_i$  est bornée :

$$\forall x \in (\mathbb{R}_+)^n, \quad 0 \leq \underline{\alpha}_i \leq \alpha_i(x) \leq \bar{\alpha}_i.$$

La figure 1 montre un exemple de trajectoire du processus dans le cas d'un gène isolé. Intuitivement, le modèle (1) fait l'hypothèse que les interactions entre les gènes proviennent exclusivement de la fonction  $\alpha$ , qui caractérise la fréquence des sauts de chaque protéine en fonction des quantités présentes de la protéine elle-même et des autres protéines.

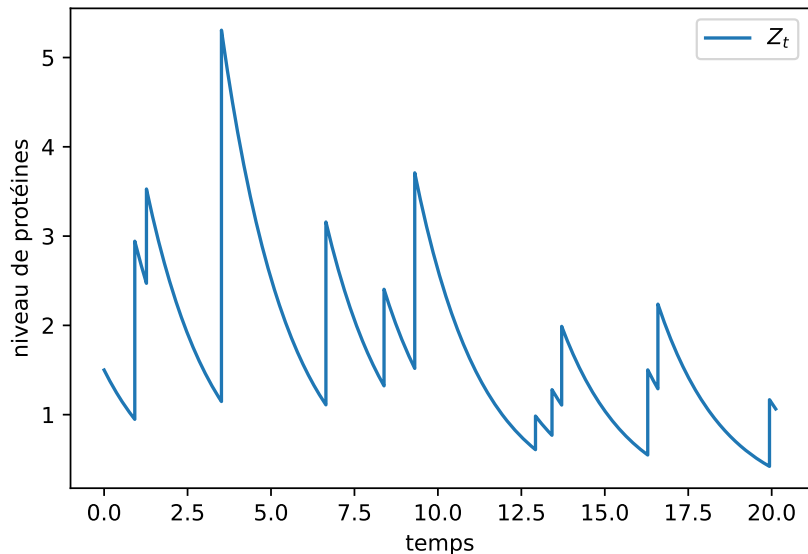


FIGURE 1 – Un exemple de trajectoire du processus  $(Z_t)_{t \geq 0}$  dans le cas d'un gène isolé.

## 2 Couplage

On cherche à justifier l'existence d'une loi invariante pour ce modèle et à estimer la vitesse de convergence vers cette dernière. Pour cela, on utilise une méthode de couplage markovien : on considère le processus  $(Z_t^1, Z_t^2)_{t \geq 0}$  où  $(Z_t^1)_{t \geq 0}$  et  $(Z_t^2)_{t \geq 0}$  sont deux copies du processus qui nous intéresse, mais avec une loi jointe particulière. On pourra voir  $(Z_t^1)_{t \geq 0}$  et  $(Z_t^2)_{t \geq 0}$  comme deux cellules que l'on essaie de synchroniser afin de minimiser  $\mathbb{E}[\|Z_t^1 - Z_t^2\|_1]$  pour  $t \geq 0$ . En effet, on sait que pour  $t \geq 0$ ,  $\mathbb{E}[\|Z_t^1 - Z_t^2\|_1]$  est un majorant de la distance de Wasserstein  $W_1$  entre la loi de  $Z_t^1$  et celle de  $Z_t^2$ . Comme l'ensemble des mesures de probabilité sur  $(\mathbb{R}_+)^n$  admettant un moment d'ordre 1 est complet pour cette distance, en majorant cette espérance par une fonction de  $t$  qui tend vers 0 lorsque  $t$  tend vers  $+\infty$ , on obtient l'existence et l'unicité d'une limite, c'est-à-dire l'existence et l'unicité de la loi invariante de notre modèle, ainsi qu'une estimation de la vitesse de convergence.

On donne maintenant les grandes lignes de la construction de la majoration qui nous donne le résultat. Notre objectif est donc de majorer  $\mathbb{E}[\|Z_t^1 - Z_t^2\|_1]$  où  $(Z_t^1)_{t \geq 0}$  et  $(Z_t^2)_{t \geq 0}$  sont deux copies du même processus  $(Z_t)_{t \geq 0}$ . On se place d'abord dans le cas d'un seul gène, c'est-à-dire que  $(Z_t)_{t \geq 0}$  est à valeurs dans  $\mathbb{R}_+$ . L'élément clé est que, en dehors des instants de saut, nos deux copies de  $(Z_t)_{t \geq 0}$  tendent toutes les deux vers 0 à une vitesse exponentielle. En particulier, elles se rapprochent l'une de l'autre à vitesse exponentielle. De plus, si  $(Z_t^1)_{t \geq 0}$  et  $(Z_t^2)_{t \geq 0}$  sautent au même instant, la loi des sauts étant la même, on peut coupler les sauts pour qu'ils aient exactement la même hauteur. Tout l'enjeu est donc d'arriver à synchroniser les instants de saut. La première étape consiste à construire un couplage qui rapproche les trajectoires de  $(Z_t^1)_{t \geq 0}$  et  $(Z_t^2)_{t \geq 0}$ . Plus précisément, comme les trajectoires se rapprochent

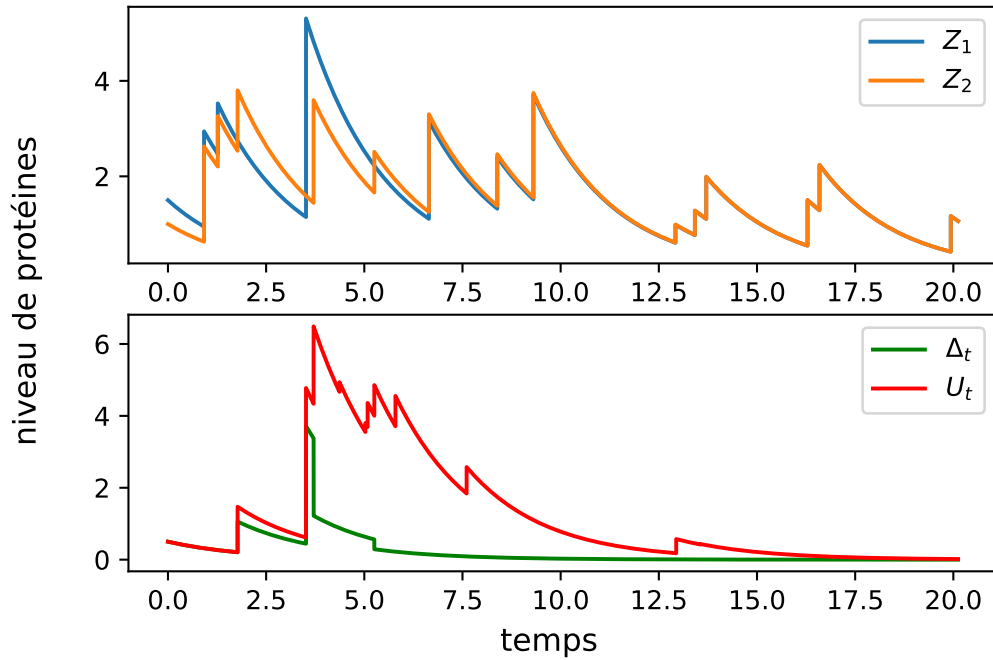


FIGURE 2 – Une trajectoire du couplage des différents processus.

spontanément, on veut un couplage tel que plus les trajectoires sont proches, plus les deux processus sautent simultanément avec forte probabilité.

Une fois un tel couplage construit, on regarde le comportement du processus

$$\Delta_t = |Z_t^1 - Z_t^2|$$

pour tout  $t \geq 0$ . Cependant, ce processus n'étant pas un processus markovien, il nous faut ajouter un nouveau processus à notre couplage, que l'on appelle « processus compagnon » noté  $(U_t)_{t \geq 0}$ , qui majore  $|Z_t^1 - Z_t^2|$  p.s. et dont la construction s'inspire de [Benaïm et al., 2012]. La figure 2 montre un exemple de trajectoire des différents processus dont il est question dans le cas d'un gène isolé. Le processus compagnon  $(U_t)_{t \geq 0}$  vérifie :

$$\forall t \geq 0, \quad \mathbb{E}[|Z_t^1 - Z_t^2|] \leq \mathbb{E}[U_t].$$

On cherche maintenant à calculer  $\mathbb{E}[U_t]$  pour tout  $t \geq 0$ . Il s'avère que le processus compagnon a un comportement particulier : il arrête de sauter presque sûrement en temps fini. Or, lorsque  $(U_t)_{t \geq 0}$  ne saute pas, il tend exponentiellement vite vers 0. On peut donc obtenir une borne sur  $\mathbb{E}[U_t]$  qui tend vers 0 lorsque  $t$  tend vers  $+\infty$ . La borne obtenue avec cette construction est intéressante car elle est valable *quelle que soit* la force des interactions entre les gènes, caractérisée par la constante de Lipschitz  $\ell$  de la fonction  $\alpha$ .

L'ingrédient principal de la preuve est que le processus  $(U_t)_{t \geq 0}$  arrête de sauter presque sûrement. Pour montrer cela, on étudie la loi du temps d'attente avant le prochain saut de

$(U_t)_{t \geq 0}$ . Ensuite, on note  $H$  l'instant du dernier saut et on fixe  $t \geq 0$ . On va conditionner  $\mathbb{E}[U_t]$  en deux évènements  $\{H < t\}$  et  $\{H \geq t\}$ . D'une part, si  $H$  est « petit », alors  $U_t$  ne peut pas être « trop grand » avant d'arrêter de sauter, puis  $(U_t)_{t \geq 0}$  tend vers 0 à vitesse exponentielle. Pour justifier que  $(U_t)_{t \geq 0}$  n'est pas « trop grand », on majore  $(U_t)_{t \geq 0}$  par un processus qui saute à taux constant. D'autre part, la probabilité que  $H$  soit grand est petite : on utilise alors le fait que  $\mathbb{E}[e^{sH}]$  est petite pour certaines valeurs de  $s$  positives.

### 3 Loi invariante

Dans la section précédente, on a donné les grandes lignes de la preuve de l'existence et l'unicité d'une loi invariante pour le PDMP caractérisé par (1). Ce qui nous intéresse maintenant c'est que cette loi admet une expression analytique dans certains cas. En effet, on a le résultat suivant.

**Théorème 1.** *S'il existe une application  $V : (\mathbb{R}_+)^n \rightarrow \mathbb{R}$  telle que*

$$\forall i \in \{1, \dots, n\}, \quad \partial_{x_i} V(x) = -\frac{\alpha_i(x)}{d_{1,i} x_i},$$

alors la loi invariante des protéines  $(Z_t)_{t \geq 0}$  s'écrit :

$$\mu(x) = \frac{1}{A} \exp(-V(x)) \prod_{i=1}^n x_i^{-1} c_i e^{-c_i x_i}, \quad (2)$$

où  $A = \int_{(\mathbb{R}_+)^n} \exp(-V(x)) \prod_{i=1}^n x_i^{-1} c_i e^{-c_i x_i} dx$  est la constante de normalisation.

Dans certains cas, on peut interpréter cette loi comme un champ aléatoire de Markov sur un graphe qui coïncide avec le graphe des interactions entre les gènes dans le modèle (1). En effet, soit  $\mathcal{G}$  un graphe non orienté à  $n$  sommets. Pour  $i, j \in \{1, \dots, n\}$ , on note  $i \sim j$  si  $\mathcal{G}$  contient l'arête  $(i, j)$ . On a alors le résultat suivant.

**Théorème 2.** *Sous les hypothèses du théorème 1 et si  $V$  peut s'écrire sous la forme*

$$V(x) = \sum_{i=1}^n f_i(x_i) + \sum_{\substack{1 \leq i < j \leq n \\ i \sim j}} f_{ij}(x_i, x_j),$$

alors la loi invariante des protéines (2) est un champ de Markov pour le graphe  $\mathcal{G}$  et pour tout  $i, j \in \{1, \dots, n\}$ , on a  $i \sim j$  si et seulement si la fonction  $\alpha_i$  dépend de  $x_j$ .

La figure 3 montre la structure typique des jeux de données *single-cell RNA-seq* considérés, sur lesquels on peut appliquer le théorème précédent. Chaque groupe de cellules collectées au même instant expérimental  $t_k$  constitue un « instantané » (*snapshot*) de l'hétérogénéité biologique à l'instant  $t_k$ . En raison de la nature destructive du processus de mesure, ces instantanés proviennent de cellules différentes. Grâce aux résultats ci-dessus, on dispose d'une

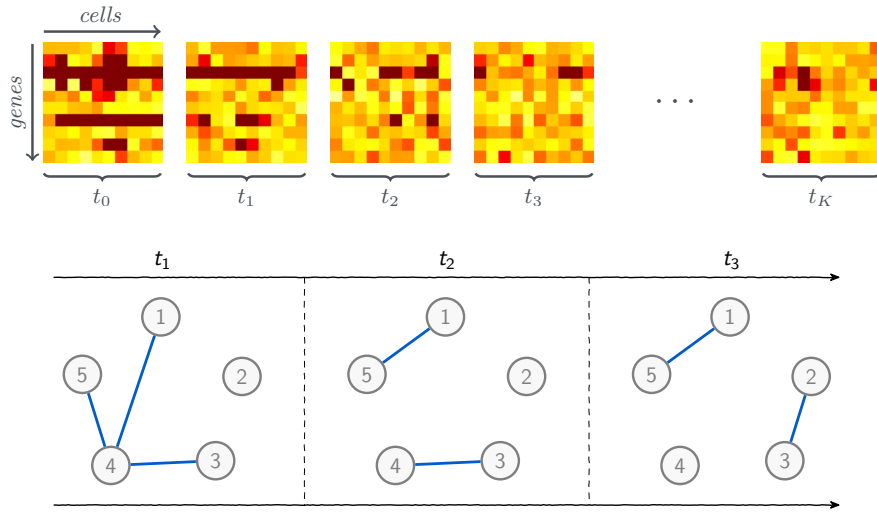


FIGURE 3 – Structure des jeux de données considérés, et modélisation de la structure de dépendance à chaque instant de mesure grâce à un champ de Markov issu du théorème 1.

famille naturelle de modèles statistiques de type champ de Markov permettant de décrire la variabilité, et surtout la structure de dépendance entre les gènes, à chaque instant de mesure  $t_k$  : le point clé est que la propriété de Markov sur les graphes (un gène est indépendant des autres conditionnellement à ses voisins) coïncide *systématiquement* avec la structure de dépendance de la fonction  $\alpha$  et a fortiori du modèle dynamique associé.

## Références

- [Benaïm et al., 2012] Benaïm, M., Le Borgne, S., Malrieu, F., and Zitt, P.-A. (2012). Quantitative ergodicity for some switched dynamical systems. *Electronic Communications in Probability*, 17(56) :1–14.
- [Dussiau et al., 2022] Dussiau, C., Boussaroque, A., Gaillard, M., Bravetti, C., Zaroili, L., Knosp, C., Friedrich, C., Asquier, P., Willems, L., Quint, L., Bouscary, D., Fontenay, M., Espinasse, T., Plesa, A., Sujobert, P., Gandrillon, O., and Kosmider, O. (2022). Hematopoietic differentiation is characterized by a transient peak of entropy at a single-cell level. *BMC Biology*, 20(1) :60.
- [Gandrillon et al., 2021] Gandrillon, O., Gaillard, M., Espinasse, T., Garnier, N. B., Dussiau, C., Kosmider, O., and Sujobert, P. (2021). Entropy as a measure of variability and stemness in single-cell transcriptomics. *Current Opinion in Systems Biology*, 27 :100348.
- [Herbach et al., 2017] Herbach, U., Bonnaffoux, A., Espinasse, T., and Gandrillon, O. (2017). Inferring gene regulatory networks from single-cell data : a mechanistic approach. *BMC Systems Biology*, 11(1) :105.
- [Ventre et al., 2023] Ventre, E., Herbach, U., Espinasse, T., Benoit, G., and Gandrillon, O. (2023). One model fits all : Combining inference and simulation of gene regulatory networks. *PLOS Computational Biology*, 19(3) :e1010962.



---

# SÉLECTION DE VARIABLES PAR APPROXIMATION DE LA NORME $L_0$ DANS UN MODÈLE DE POISSON LOG-NORMAL

Togo Jean Yves KIOYE<sup>1</sup>, Paul-Marie GROLLEMUND<sup>1,2</sup>, Jocelyn CHAUVET<sup>3,4</sup>,  
Christophe CHASSARD<sup>1</sup>

<sup>1</sup> *Unité Mixte de Recherche sur le Fromage (UMRF), France*  
*togo\_jean\_yves.kioye@uca.fr, christophe.chassard@inrae.fr*

<sup>1,2</sup> *Laboratoire de Mathématiques Blaise Pascal (LMBP), France*  
*paul\_marie.grollemund@uca.fr*

<sup>3</sup> *Centre de recherche de l'ICES, France*

<sup>4</sup> *Laboratoire Angevin de Recherche en Ingénierie des Systèmes (LARIS), France*  
*jchauvet@ices.fr*

**Résumé.** Les méthodes de sélection de variables sont nécessaires en modélisation statistique pour identifier et inclure uniquement les prédicteurs les plus pertinents, facilitant ainsi l'interprétabilité du modèle. Ces méthodes de sélection de variables sont généralement utilisées dans les modèles de régression, comme c'est le cas dans ce travail pour le modèle de Poisson Log-Normal (PLN, [Chiquet et al., 2021](#)). Ce modèle vise à expliquer des données de comptage multivariées en utilisant des variables dépendantes, et son utilité a été démontrée dans des domaines scientifiques tels que l'écologie et l'agronomie. Dans le cas du modèle Poisson Log-Normal (PLN), les travaux les plus récents se concentrent sur l'inférence de réseaux parcimonieux en combinant une borne inférieure de la vraisemblance avec une pénalisation  $L_1$  appliquée à la matrice de précision. Dans ce travail, nous proposons de nous appuyer sur une méthode de pénalisation récente (SIC, [O'Neill and Burke, 2023](#)), qui consiste à approcher de manière lisse la pénalisation  $L_0$  et qui évite la calibration d'un paramètre de régularisation par une procédure de validation croisée. De plus, ce travail se concentre sur la matrice des coefficients du modèle PLN et propose une procédure d'inférence qui assure une sélection efficace des variables. Ainsi, le modèle ajusté obtenu explique les données de comptage multivariées en utilisant exclusivement les variables explicatives pertinentes. Notre approche consiste à mettre en œuvre une procédure intégrant l'algorithme de pénalisation SIC ( $\varepsilon$ -telescoping) et l'algorithme d'ajustement du modèle PLN (algorithme EM variationnel). Afin de valider notre approche, nous menons des études de simulation destinées à évaluer la performance de la méthode proposée et procédons à son application sur des données réelles.

**Mots-clés.** Sélection de variables, données de comptage multivariées, algorithme EM variationnel, critères d'information

**Abstract.** Variable selection methods are required in practical statistical modeling, to identify and include only the most relevant predictors, and then improving model inter-

---

pretability. Such variable selection methods are typically employed in regression models, for instance in this article for the Poisson Log Normal model (PLN, [Chiquet et al., 2021](#)). This model aim to explain multivariate count data using dependent variables, and its utility was demonstrating in scientific fields such as ecology and agronomy. In the case of the PLN model, most recent papers focus on sparse networks inference through combination of the evidence lower bound of the likelihood with a  $L_1$ -penalty on the precision matrix. In this paper, we propose to rely on a recent penalization method (SIC, [O’Neill and Burke, 2023](#)), which consists in smoothly approximating the  $L_0$ -penalty, and that avoids the calibration of a tuning parameter with a cross-validation procedure. Moreover, this work focuses on the coefficient matrix of the PLN model and establishes an inference procedure ensuring effective variable selection performance, so that the resulting fitted model explaining multivariate count data using only relevant explanatory variables. Our proposal involves implementing a procedure that integrates the SIC penalization algorithm ( $\varepsilon$ -telescoping) and the PLN model fitting algorithm (a variational EM algorithm). To support our proposal, we perform simulation studies to assess the method, which is also applied on real datasets.

**Keywords.** Variable selection, Multivariate count data, Variational EM algorithm, Information criteria

## 1 Introduction

L’analyse des données de comptage multivariées joue un rôle central dans la modélisation statistique, offrant des perspectives dans divers domaines. Parmi les applications, on trouve l’étude de l’abondance simultanée de différentes espèces dans les communautés écologiques, la modélisation de l’occurrence de diverses maladies dans une population ([Zhang et al., 2017](#)), et l’analyse des abondances relatives des microorganismes dans un écosystème spécifique ([Chiquet et al., 2018](#)). Toutefois, l’analyse des données de comptage multivariées présente des défis particuliers, notamment en ce qui concerne la gestion des dépendances entre variables et le traitement de la surdispersion. Ces dernières années, plusieurs approches statistiques avancées ont été développées pour relever ces défis, telles que la régression de Poisson multivariée ([Chiquet et al., 2021](#)), la régression binomiale négative multivariée ([Shi and Valdez, 2014](#)), ainsi que les modèles basés sur les copules ([Nikoloulopoulos and Karlis, 2009](#)).

Dans ce travail, nous nous concentrons sur le modèle Poisson Log-Normal (PLN), qui offre une grande flexibilité pour envisager des extensions supplémentaires ([Chiquet et al., 2021](#)). Ce modèle repose sur une couche latente, ce qui complexifie son inférence. L’approximation variationnelle est une méthode efficace pour surmonter cette difficulté, et

---

le modèle est ainsi ajusté à l'aide de l'algorithme *Expectation-Maximization* variationnel (VEM). Toutefois, il est également possible d'effectuer l'inférence de ce modèle en utilisant des approches basées sur Monte Carlo (Chib and Winkelmann, 2001; Karlis and Xekalaki, 2005) ou la méthode de Laplace (Choi et al., 2017).

L'utilisation du modèle Poisson Log-Normal (PLN) permet d'étudier les structures de réseau à travers les coefficients estimés de la matrice de précision dans des contextes pratiques. Elle offre également la possibilité d'examiner l'impact des variables dépendantes sur les données de comptage en estimant les coefficients de régression. Dans ce contexte, il est pertinent de concevoir une procédure d'estimation associée à une méthode de sélection de variables pour garantir que le modèle ajusté conserve uniquement les coefficients significativement différents de zéro.

Le principal enjeu de la sélection de variables réside dans le compromis à établir entre la qualité de l'ajustement du modèle et la parcimonie, souvent formalisé comme un problème d'optimisation. L'utilisation de pénalités de type  $L_1$ , telles que la méthode Lasso (Tibshirani, 1996), constitue une solution classique pour faciliter cette optimisation. Cependant, ces méthodes nécessitent la calibration d'un paramètre de régularisation, ce qui est souvent réalisé par des méthodes coûteuses telles que la validation croisée.

Récemment, O'Neill and Burke (2023) ont introduit une nouvelle approche de pénalisation visant à approcher la norme  $L_0$ , permettant une sélection de variables efficace sans nécessiter la calibration du paramètre de régularisation par des procédures telles que la validation croisée. Cette approche, dénommée *Smooth Information Criterion* (SIC), a été appliquée aux modèles de régression distributionnelle.

Dans ce travail, nous proposons d'adapter cette approche au modèle Poisson Log-Normal (PLN) afin d'obtenir une estimation plus parcimonieuse des coefficients de régression et de réduire le coût computationnel associé à l'ajustement du paramètre de régularisation. La section suivante présente le modèle Poisson Log-Normal (PLN), détaille la méthode d'estimation de ses paramètres, et décrit la fonction de pénalisation utilisée pour obtenir des estimations parcimonieuses des coefficients de régression.

## 2 Modélisation

**Modèle de Poisson Log-Normal (PLN).** Le modèle PLN multivarié a été introduit par Aitchison and Ho (1989) dans le but de modéliser des données de comptage conjointes à l'aide de facteurs environnementaux, tout en tenant compte de la structure de dépendance entre les comptages. Chaque vecteur de comptage observé  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{ip}) \in \mathbb{N}^p$  est relié à un vecteur latent  $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{ip}) \in \mathbb{R}^p$ , distribué selon une loi gaussienne multivariée avec une matrice de covariance  $\Sigma$ . Les comptages  $Y_{ij}$  observés suivent une loi de Poisson avec un paramètre  $\exp(o_{ij} + \mathbf{x}_i^\top \boldsymbol{\beta}_j + Z_{ij})$ , où  $o_{ij}$  est considéré comme l'effort

d'échantillonnage de l'espèce  $j$  sur le site  $i$  et  $\mathbf{x}_i \in \mathbb{R}^d$  représente le vecteur des covariables pour le site  $i$ . Le modèle peut s'écrire de la façon suivante (Chiquet et al., 2018) :

$$\begin{aligned} Y_{ij} | Z_{ij} &\sim \mathcal{P}(\exp(o_{ij} + \mathbf{x}_i^\top \boldsymbol{\beta}_j + Z_{ij})) && \text{(espace observé)} \\ \mathbf{Z}_i &\sim N_p(0, \boldsymbol{\Sigma}) && \text{(espace latent),} \end{aligned} \quad (1)$$

Dans ce modèle,  $\boldsymbol{\beta}_j = (\beta_{1j}, \dots, \beta_{dj})^\top$  encode l'effet des  $d$  covariables environnementales sur l'abondance de l'espèce  $j$ . Notons par  $\mathbf{Y} \in \mathbb{N}^{n \times p}$  la matrice de comptage,  $\mathbf{X} \in \mathbb{R}^{n \times d}$  la matrice des variables environnementales,  $\mathbf{O} \in \mathbb{R}^{n \times p}$  la matrice des décalages (ou efforts d'échantillonnage), et  $\mathbf{B} \in \mathbb{R}^{d \times p}$  la matrice des paramètres de régression. Nous notons aussi par  $\boldsymbol{\theta} = \{\mathbf{B}, \boldsymbol{\Sigma}\}$  l'ensemble des paramètres du modèle PLN et par  $\mathbf{Z} \in \mathbb{R}^{n \times p}$  la matrice constituée des vecteurs latents  $\mathbf{Z}_i$ .

**Méthodes d'estimation.** L'estimation des paramètres du modèle PLN nécessite l'utilisation de l'algorithme EM. Cependant, l'étape E requiert le calcul de l'espérance conditionnelle sous la distribution  $p_{\boldsymbol{\theta}}(\mathbf{Z}_i | \mathbf{Y}_i)$ , qui est intraitable. Pour contourner ce problème, Chiquet et al. (2021) proposent de recourir à une stratégie d'inférence variationnelle pour approcher  $p_{\boldsymbol{\theta}}(\mathbf{Z}_i | \mathbf{Y}_i)$  par une distribution  $q_{\boldsymbol{\psi}}(\mathbf{Z}_i)$  qui minimise la divergence de Kullback-Leibler (KL) entre  $q_{\boldsymbol{\psi}}(\mathbf{Z}_i)$  et  $p_{\boldsymbol{\theta}}(\mathbf{Z}_i | \mathbf{Y}_i)$ . Cette stratégie conduit à la maximisation d'une borne inférieure de la log-vraisemblance appelée ELBO (*Evidence Lower Bound*) :

$$J(\mathbf{Y}, \boldsymbol{\theta}, \boldsymbol{\psi}) = \log(p_{\boldsymbol{\theta}}(\mathbf{Y})) - \text{KL}[q_{\boldsymbol{\psi}}(\mathbf{Z} | \mathbf{Y}) || P_{\boldsymbol{\theta}}(\mathbf{Z} | \mathbf{Y})]. \quad (2)$$

Chiquet et al. (2021) ont considéré l'approche champ moyen dans laquelle :

$$q_{\boldsymbol{\psi}}(\mathbf{Z} | \mathbf{Y}) = \prod_i^n \mathcal{N}(\mathbf{Z}_i; \mathbf{m}_i, \text{diag}(\mathbf{s}_i^2))$$

L'ensemble des paramètres variationnels sont rassemblés dans le vecteur  $\boldsymbol{\psi} = (\mathbf{M}, \mathbf{S})$ , où  $\mathbf{M} = (\mathbf{m}_1^\top, \dots, \mathbf{m}_n^\top)^\top$  et  $\mathbf{S} = (\mathbf{s}_1^2, \dots, \mathbf{s}_n^2)^\top$ .

**Sélection de variables.** Nous considérons la ELBO du modèle PLN (2) en nous focalisant sur le paramètre d'intérêt, à savoir la matrice des paramètres de régression  $\mathbf{B}$ , pour laquelle nous cherchons à obtenir une estimation parcimonieuse. À cette fin, nous ajoutons une pénalité  $\phi_\varepsilon$  à l'ELBO pour contraindre certains coefficients de régression à être exactement égaux à 0. La fonction de pénalisation  $\phi_\varepsilon$  a été introduite par O'Neill and Burke (2023) dans le but d'approcher la norme  $L_0$ .

La fonction objective pénalisée correspondante est alors

$$J_{\text{pen}}(\mathbf{Y}; \mathbf{B}, \boldsymbol{\Sigma}, \boldsymbol{\psi}) = J(\mathbf{Y}; \mathbf{B}, \boldsymbol{\Sigma}, \boldsymbol{\psi}) - \frac{\lambda}{2} [\phi_\varepsilon(\mathbf{B}) + k], \quad (3)$$

---

où  $\lambda = \log(n)$ ,  $k$  le nombre de paramètres qu'on ne souhaite pas pénaliser et

$$\phi_\varepsilon(x) = \frac{x^2}{x^2 + \varepsilon^2}.$$

Notons que la fonction  $\phi_\varepsilon$  est dérivable pour  $\varepsilon > 0$ , et on a :  $\lim_{\varepsilon \rightarrow 0} \phi_\varepsilon(x) = \|x\|_0$ .

Durant l'exposé, nous présenterons la formulation du SIC comme une distribution a priori dans le cadre bayésien, l'algorithme d'inférence résultant du couplage du SIC et du VEM, ainsi que des résultats numériques basés à la fois sur des données simulées et réelles.

## References

- Aitchison, J. and Ho, C. (1989). The multivariate poisson-log normal distribution. *Biometrika*, 76(4):643–653.
- Chib, S. and Winkelmann, R. (2001). Markov chain monte carlo analysis of correlated count data. *Journal of Business & Economic Statistics*, 19(4):428–435.
- Chiquet, J., Mariadassou, M., and Robin, S. (2018). Variational inference for probabilistic poisson pca. *The Annals of Applied Statistics*, 12(4):2674–2698.
- Chiquet, J., Mariadassou, M., and Robin, S. (2021). The poisson-lognormal model as a versatile framework for the joint analysis of species abundances. *Frontiers in Ecology and Evolution*, 9:588292.
- Choi, Y., Coram, M., Peng, J., and Tang, H. (2017). A poisson log-normal model for constructing gene covariation network using rna-seq data. *Journal of Computational Biology*, 24(7):721–731.
- Karlis, D. and Xekalaki, E. (2005). Mixed poisson distributions. *International Statistical Review/Revue Internationale de Statistique*, pages 35–58.
- Nikoloulopoulos, A. K. and Karlis, D. (2009). Modeling multivariate count data using copulas. *Communications in Statistics-Simulation and Computation*, 39(1):172–187.
- O'Neill, M. and Burke, K. (2023). Variable selection using a smooth information criterion for distributional regression models. *Statistics and Computing*, 33(3):71.
- Shi, P. and Valdez, E. A. (2014). Multivariate negative binomial models for insurance claim counts. *Insurance: Mathematics and Economics*, 55:18–29.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.
- Zhang, Y., Zhou, H., Zhou, J., and Sun, W. (2017). Regression models for multivariate count data. *Journal of Computational and Graphical Statistics*, 26(1):1–13.

---

# SURVIE DES COLONIES D'*Apis mellifera* ET PRIX DU MIEL : MODÉLISATION ET PERSPECTIVES D'APPLICATION

Mathieu Kohli <sup>1</sup>

<sup>1</sup> *Chercheur indépendant à Dijon, France, mathieu.kohli@protonmail.ch*

**Résumé.** Une colonie d'abeilles mellifères, en butinant le nectar des fleurs qu'elle transforme en miel accumule ce miel en vue d'avoir assez d'énergie pour chauffer la ruche pendant l'hiver. Plus elle possède de réserves en miel, plus sa chance de survivre à l'hiver augmente. On présentera un modèle centré sur la probabilité de survivre au prochain hiver et sur le coût induit par une récolte de miel qui se répercute dans le prix du miel.

**Mots-clés.** Abeille mellifère, miel, prix, optimisation.

**Abstract.** A honeybee colony forages nectar to transform it into honey that is then stockpiled by the bees to ensure there is enough energy to heat the hive during winter. The more honey the colony holds, the likelier it is to survive all winter long. We present a model that focuses on the probability a bee colony has to survive the next winter and on the cost that harvesting honey induces. This cost is embedded in the price of honey.

**Keywords.** Honeybee, honey, price, optimization.

## 1 Probabilité de survie d'une ruche et prix du miel

Une colonie d'abeilles domestiques *Apis mellifera* constitue une société organisée qui passe l'année à se préparer à l'hiver en accumulant des réserves de miel. Le miel permet aux abeilles de maintenir une chaleur suffisante dans la ruche durant la saison froide. Plus spécifiquement, c'est une seule abeille, la reine, qui donne naissance à toutes les autres et qui doit être préservée de la rudesse de l'hiver. Pour cela toutes les autres abeilles forment autour d'elle une grappe dense qu'elles chauffent activement grâce à l'énergie contenue dans le miel.

Le modèle présenté dans la suite se base sur Kohli et Anstett (2024) et Kohli (2024). On n'oublie pas qu'il existe déjà des modèles classiquement utilisés qui, contrairement au présent modèle, sont plus centrés sur les variables extensives de taille de la colonie. On citera ainsi BEEHAVE de Becher et al. (2014)

Chaque kilogramme de miel accumulé dans la ruche augmente la probabilité que la colonie passe l'hiver et a contrario chaque kilogramme récolté pour être vendu sur le marché diminue cette probabilité de survie et constitue donc un coût qu'il faudra compenser par l'achat de nouveaux essaims et répercuter dans le prix de vente du miel.

$$Prix = \alpha \frac{\partial \mathbb{P}}{\partial M} \quad (1)$$

où  $\mathbb{P}$  est la probabilité de survivre à l'hiver suivant,  $M$  est la quantité de miel présente dans la ruche et  $\alpha$  est le prix d'un essaim.

Il est donc essentiel de comprendre la probabilité  $\mathbb{P}$  de survivre au prochain hiver pour appréhender l'économie de la ruche.

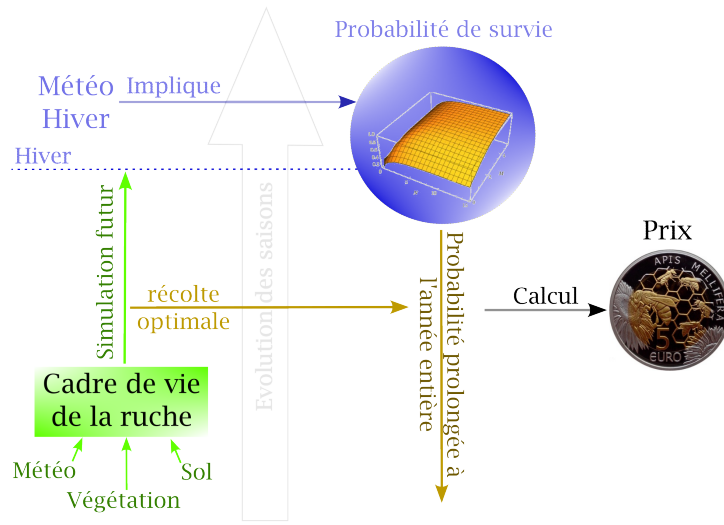


Figure 1: Prolongement de  $\mathbb{P}$  à l'année entière.

## 1.1 Au début de l'hiver

A l'entrée en hivernage, pourvu que la ruche soit en bonne santé et sachant les statistiques météorologiques au lieu où se trouve une ruche,  $\mathbb{P}$  ne dépend que de la quantité de miel  $M$  et du nombre d'abeilles  $N$  dans la ruche.

$$\mathbb{P}(M, N) = \phi\left(\nu \frac{M}{N^{\frac{1}{3}}}\right) \psi\left(\theta N^{\frac{2}{3}}\right) \quad (2)$$

où  $\nu$  et  $\theta$  sont des constantes expérimentales et où  $\phi$  et  $\psi$  sont respectivement les fonctions de répartition des variables aléatoires :

$$\begin{aligned} \{\text{hivers}\} &\longrightarrow \mathbb{R} & \{\text{hivers}\} &\longrightarrow \mathbb{R} \\ \text{hiver} &\longmapsto \int_{\text{hiver}} T_{\text{centre}} - T_{\text{ext}} & \text{hiver} &\longmapsto T_{\text{centre}} - T_{\text{ext,min}}. \end{aligned} \quad (3)$$

---

## 1.2 Durant la saison apicole

Afin de prolonger la probabilité de survivre au prochain hiver en fonction des réserves de miel et du nombre d'abeilles on suit le fil conducteur décrit en figure 1.

Pour la simulation de la croissance des plantes qui fournissent les ressources (nectar et pollen) aux abeilles, on agrège des données de météorologie, de cartographie de la flore et des observations satellite de phénologie (lien entre la saison et la pousse des plantes). Des méthodes d'apprentissage statistique devront ensuite être mises en oeuvre pour décrire le développement annuel des plantes.

On modélise la récolte du nectar et du pollen par les abeilles butineuses en supposant que celles-ci privilégient la qualité de la ressource c'est à dire la ressource nette rapportée à la ruche par unité de temps qui se calcule ainsi

$$\text{Qualité d'une ressource} = \frac{\text{ressource récoltée} - \text{énergie dépensée}}{\text{temps d'un cycle de butinage}}. \quad (4)$$

Cela est rendu possible par le fait que les abeilles domestiques n'effectuent pas une marche aléatoire lors de leur vol de butinage mais se communiquent entre elles les lieux où se trouvent les meilleures ressources grâce à une "danse" à l'intérieur de la ruche d'après Von Frisch (1967)

Remarquons que dans la formule définissant la qualité d'une ressource, la différence entre la récolte et la dépense fait sens quand les abeilles récoltent du nectar qui sera transformé en miel et utilisé comme énergie. Par contre quand les butineuses récoltent du pollen qui sera utilisé comme protéines constitutives des larves, il faut définir un "taux de change"  $\tau$  entre miel et pollen. Chaque taux de change potentiel donnera lieu à une qualité potentielle de la ressource en nectar qui impliquera une répartition potentielle des butineuses sur les ressources aux alentours. Si on considère alors que la colonie optimise sa probabilité de survie à l'hiver suivant, le taux de change  $\tau$  réel sera celui qui maximise  $\mathbb{P}(M(t+dt), N(t+dt))$  où  $dt$  correspond par exemple à un jour. Cela donne lieu au système différentiel suivant :

$$\begin{cases} \left\langle \frac{\partial}{\partial \tau} \begin{pmatrix} M'(t) \\ N'(t) \end{pmatrix}, \overrightarrow{\text{grad}}(\mathbb{P}) \right\rangle = 0, \\ N'(t) \frac{\partial \mathbb{P}}{\partial N} + H'(t) \frac{\partial \mathbb{P}}{\partial H} = -\frac{\partial \mathbb{P}}{\partial t}, \end{cases} \quad (5)$$

qui peut être résolu numériquement.

## 2 Applications

Parmi les applications potentielles d'un tel modèle se trouvent la cartographie des zones propices à l'apiculture. En effet, il peut être judicieux de placer des ruches là où le miel



---

est moins cher à produire, bien que d'autres paramètres entrent en jeu tels que le type de fleurs présentes dans la région pour les miels monofloraux.

Si on souhaite savoir où placer des ruches, il est tout aussi important de savoir *quand* récolter le miel pour minimiser la perturbation des abeilles. Comme le modèle ici présenté fournit un cours du miel prévisionnel durant l'année, la meilleure opportunité de récolte se situe là où le prix calculé est le plus bas.

Une dernière application notable se base non pas sur le prix du miel mais sur la fonction de probabilité de survie  $\mathbb{P}$  dont découle ce prix. La reproduction naturelle des colonies d'abeilles domestiques se déroule lors d'un processus nommé essaimage : après la naissance d'une nouvelle reine, la reine plus âgée part avec la moitié des abeilles de la colonie pour trouver une cavité dans un arbre où s'installer. Les conditions seront donc favorables à un essaimage lorsqu'il y a plus de chance qu'au moins une des deux moitiés de la colonie initiale survive au prochain hiver qu'il n'y a de chance que la colonie initiale elle-même y survive. La condition nécessaire d'essaimage s'exprime donc :

$$\left(1 - \mathbb{P}\left(t, \frac{M}{2}, \frac{N}{2}\right)\right)^2 \leq 1 - \mathbb{P}(t, M, N). \quad (6)$$

La pratique apicole consistant à devancer l'essaimage en divisant une colonie dans deux ruches distinctes serait facilitée par la vérification numérique de cette condition.

## Bibliographie

- Becher M. et al. (2014) *BEEHAVE: a systems model of honeybee colony dynamics and foraging to explore multifactorial causes of colony failure* Journal of Applied Ecology, 51: 470-482. <https://doi.org/10.1111/1365-2664.12222>
- Kohli, M. et Anstett M.-C. (2024) *A mathematical modelization of the energy content of an Apis mellifera beehive* soumis et en cours de review.
- Kohli, M. (2024) *Price structure of traded honey, pollen and wax linked to winter survival of a honey bee colony* preprint HAL <https://hal.science/hal-04469620/>
- Von Frisch K. (1967) *The Dance Language and Orientation of Bees*. Harvard University Press. ISBN 978067419051

---

# Étude mathématique de la propagation et du blocage dans les maladies inflammatoires chroniques de l'intestin

Saoussen Latrach, Eric Ogier-Denis, Nicolas Vauchelet, et Hatem Zaag

October 18, 2024

## 1 Introduction générale aux MICI

Les maladies inflammatoires chroniques de l'intestin (MICI), principalement représentées par la colite ulcéreuse (RCH) et la maladie de Crohn, sont des maladies du système digestif qui se caractérisent par une inflammation chronique de l'intestin. Ces maladies touchent une proportion croissante de la population mondiale, surtout dans les pays industrialisés. Le système digestif, directement lié à l'alimentation, est fortement affecté par des facteurs environnementaux et alimentaires.

Les MICI ont commencé à émerger de manière significative au 20e siècle, d'abord dans les pays développés comme l'Amérique du Nord et l'Europe. Cette croissance rapide a poussé les chercheurs à explorer les causes possibles, notamment l'impact de l'environnement moderne et de l'alimentation transformée. Cette tendance suggère une forte interaction entre les modes de vie modernes et l'apparition de ces maladies.

### 1.1 L'alimentation comme cause potentielle des MICI

L'évolution des pratiques alimentaires modernes est souvent liée à l'augmentation des cas de MICI. En particulier, l'apparition et l'usage généralisé du réfrigérateur au 20e siècle a permis la conservation d'aliments pour de plus longues périodes, ce qui a profondément modifié la manière dont les gens consomment des produits alimentaires. Cela a notamment réduit la consommation d'aliments fermentés naturellement, qui jouent un rôle crucial dans l'équilibre du microbiote intestinal. Ce changement pourrait être un facteur déclencheur dans l'apparition de l'inflammation intestinale chez certaines personnes ( voir [1]).

Certaines études épidémiologiques ont montré une corrélation entre une alimentation occidentale riche en graisses, sucres raffinés et aliments transformés, et l'apparition de maladies inflammatoires. Ces aliments transformés sont souvent pauvres en fibres et contiennent des additifs qui pourraient perturber l'équilibre des bactéries bénéfiques dans l'intestin. Un déséquilibre dans le microbiote, appelé dysbiose, est souvent observé chez les personnes atteintes de MICI.

La modernisation de l'agriculture joue également un rôle clé. L'utilisation croissante de pesticides, la production de masse d'aliments et les organismes génétiquement modifiés (OGM) affectent la qualité nutritionnelle des aliments. Cela pourrait avoir des effets délétères

---

sur le système immunitaire, augmentant ainsi les risques de maladies inflammatoires comme les MICI.

## 1.2 Conséquences alimentaires pour les malades des MICI

Les personnes atteintes de MICI doivent suivre des régimes spécifiques pour réduire l'inflammation et les symptômes. Cela implique souvent l'élimination d'aliments comme le gluten, les produits laitiers, les épices et les aliments transformés, qui peuvent aggraver la maladie. Des régimes pauvres en FODMAP sont couramment recommandés pour limiter les fermentations dans l'intestin et réduire les douleurs. Une alimentation riche en fibres provenant de produits non transformés aide également à équilibrer la flore intestinale. Ainsi, une agriculture durable et une alimentation naturelle pourraient non seulement améliorer la gestion des MICI, mais aussi contribuer à leur prévention.

## 1.3 L'impact de l'environnement

L'environnement joue un rôle clé dans la santé intestinale et le développement des MICI. L'exposition à des toxines, pesticides et produits chimiques dans l'eau, l'air et les aliments peut perturber le microbiote intestinal et favoriser l'inflammation chronique. De plus, la réduction de l'exposition à la nature due à l'urbanisation affaiblit le système immunitaire, tandis qu'un mode de vie urbain et une alimentation transformée réduisent la diversité du microbiote. Les changements climatiques et la pollution aggravent également la situation, soulignant la nécessité de protéger les écosystèmes pour prévenir les MICI.

## 2 Modélisation mathématique des MICI

La modélisation mathématique des maladies inflammatoires chroniques de l'intestin (MICI) est un outil puissant pour mieux comprendre les mécanismes complexes qui sous-tendent l'apparition, la progression et la gestion de ces maladies. En combinant des concepts de biologie, de médecine et de mathématiques, ces modèles permettent de décrire de manière quantitative les interactions entre les bactéries intestinales, les cellules immunitaires, et les autres facteurs biologiques influençant l'inflammation chronique.

Les modèles mathématiques de MICI reposent souvent sur des équations différentielles, qu'elles soient ordinaires ou partielles, pour représenter les dynamiques d'interaction entre les bactéries du microbiote et les réponses immunitaires. Par exemple, des modèles de réaction-diffusion peuvent être utilisés pour explorer la propagation de l'inflammation dans les tissus intestinaux.

Dans nos recherches, nous présentons un modèle de réaction-diffusion inspiré des travaux de Nadin et al. (2021), (voir [2]) qui modélise les interactions entre un pathogène et les cellules immunitaires, afin de décrire la propagation dans un milieu homogène dans le cadre de la colite ulcéreuse (CU), également connue sous le nom de rectocolite hémorragique (RCH). Nous étudierons ensuite les conditions dans lesquelles la propagation peut être stoppée dans un milieu hétérogène.

---

Le système d'équations que nous étudions dans [3] est le suivant :

$$\begin{cases} \partial_t u - D\partial_{xx}u = r_b u \left(1 - \frac{u}{b_i}\right) + \gamma \left(1 - \frac{u}{b_i}\right)v - \frac{au}{s_b+u}v - du, \\ \partial_t v - D\partial_{xx}v = \beta u - r_c v. \end{cases} \quad (1)$$

Dans ce modèle,  $u$  représente les bactéries, et  $v$  désigne les cellules immunitaires. Ce modèle utilise 7 paramètres positifs :

- $r_b$  est le taux de reproduction des bactéries,
- $d$  représente le taux de mortalité bactérienne,
- $b_i$  est la densité bactérienne dans la lumière,
- $a$  est le coefficient proportionnel au taux de phagocytose,
- $\gamma$  est une mesure de l'effet négatif de la concentration en phagocytes sur la résistance épithéliale,
- $\beta$  est le taux de réponse immunitaire,
- $r_c$  est le taux de mortalité intrinsèque des phagocytes.

## 2.1 Analyse Théorique de la Propagation et du Blocage

### 2.1.1 Propagation dans un Milieu Homogène :

Dans un premier temps, nous étudions la propagation de l'inflammation dans un milieu homogène, afin de déterminer les conditions sous lesquelles une onde de propagation est possible. Les résultats montrent que dans un milieu homogène, l'inflammation peut se propager indéfiniment. Ce résultat est particulièrement pertinent pour l'UC, où l'inflammation a tendance à se propager de manière continue.

### 2.1.2 Blocage de la propagation dans un milieu hétérogène :

Dans un deuxième temps, nous introduisons une hétérogénéité dans le modèle en variant le taux de mortalité des bactéries, afin d'explorer les conditions dans lesquelles la propagation de l'inflammation peut être bloquée.

## 2.2 Simulations Numériques :

Les simulations numériques illustrent les comportements observés théoriquement. Par exemple, les ondes de propagation montrent une progression stable dans des milieux homogènes, tandis que dans des milieux hétérogènes, des blocages apparaissent, simulant des zones où l'inflammation est restreinte. Ces résultats sont présentés sous forme de graphiques qui permettent de visualiser les différents scénarios de propagation et de blocage.

## References

- [1] L'émergence de la maladie de crohn pourrait être liée au développement de la réfrigération, 2003.

- 
- [2] Grégoire Nadin, Eric Ogier-Denis, Ana I Toledo, and Hatem Zaag. A turing mechanism in order to explain the patchy nature of crohn's disease. *Journal of Mathematical Biology*, 83:1–17, 2021.
- [3] Nicolas Vauchelet S. Latrach, E. Ogier-Denis and Hatem Zaag. Mathematical study of the spread and blocking in inflammatory bowel disease. *Mathematical Biosciences*, pages 1–27, 2024. submitted.

---

# MODÉLISATION SPATIO-TEMPORELLE DES VOIES DE SIGNALISATION : IMPACT DE LA COMPARTIMENTATION ENDOSOMALE ET APPLICATION AUX RÉCEPTEURS AUX GONADOTROPHINES.

Chloé Weckel <sup>1,2,\*</sup> & Romain Yvinec <sup>1,2</sup>

<sup>1</sup> BIOS, PRC, INRAE Centre Val de Loire 37380 Nouzilly

<sup>2</sup> MUSCA Inria Saclay 91120 Palaiseau

\* *chloe.weckel@inrae.fr*

**Résumé.** Les cellules communiquent entre elles en s'envoyant des ligands extra-cellulaires tels que des signaux hormonaux. Les ligands d'intérêt sont les hormones de la reproduction qui une fois reconnues par leur récepteur à la membrane plasmique enclenchent au sein de la cellule une cascade de signalisation. Le couple ligand-récepteur est ensuite internalisé dans différents compartiments appelés endosomes qui vont eux aussi produire une réponse cellulaire ayant un rôle physiologique différent. Cette notion de compartimentage est modélisée mathématiquement à travers deux approches : un modèle structuré basé sur des systèmes d'équations différentielles ordinaires et un modèle individu-centré à l'aide des processus de Markov déterministes par morceaux. Cette modélisation permet l'étude de l'effet de l'internalisation et du recyclage sur la quantité de réponse cellulaire. L'intérêt final du projet est de caractériser plus finement l'action pharmacologique des ligands en contribuant à une meilleure compréhension des voies de signalisation des hormones de la reproduction.

**Mots-clés.** Signalisation cellulaire, Compartimentage, Modélisation, Réseaux de réactions chimiques, Systèmes hybrides, Endosomes, Processus de Markov déterministes par morceaux.

**Abstract.** Cells communicate with each other by sending extra-cellular ligands such as hormones. The ligands of interest are reproductive hormones which, once recognized by their receptors at the plasma membrane, trigger a signaling cascade within the cell. These receptors are internalized in different compartments called endosomes, which produce a cellular response with a different physiological role. This notion of compartmentalization is mathematically modeled using two approaches: a structured model based on systems of ordinary differential equations, and an individual-centered model using piecewise deterministic Markov processes. This modeling enables us to study the effect of internalization and recycling on the quantity of cellular response. The ultimate aim of the project is to further characterize the pharmacological action of ligands, contributing to a better understanding of gonadotropin signaling pathways.

**Keywords.** Cell signaling, Compartmentalization, Modeling, Chemical reaction networks, Hybrid systems, Endosome, Piecewise deterministic Markov processes.

---

## 1 Enjeux sociétaux en maîtrise de la reproduction

La maîtrise de la reproduction dans les élevages s'appuie largement sur l'utilisation d'hormones d'origines extractives telles que les gonadotrophines (eCG, LH ou FSH) et des progestagènes afin de synchroniser les cycles ovariens et d'induire les ovulations. Ces manipulations pharmacologiques permettent notamment des reproductions à contre-saison et plus généralement une conduite rationnelle des élevages. Toutefois, l'utilisation de ces hormones n'est pas sans conséquences sur le plan sanitaire, éthique mais aussi environnemental. En particulier, les analogues stéroïdiens à longue demi-vie, abondamment utilisés par les élevages, sont libérés et accumulés dans l'environnement où ils se comportent comme des perturbateurs endocriniens affectant à la fois la faune mais représentant aussi une menace pour la santé humaine. Bien qu'il existe une volonté des pouvoirs publics de promouvoir les pratiques s'inscrivant dans un développement durable, pour l'heure, il n'existe pas d'alternatives pharmacologiques à l'utilisation de ces hormones pour la maîtrise de la reproduction dans l'élevage. Le projet contribue à une meilleure compréhension des voies de signalisations des gonadotrophines, permettant de sélectionner des molécules innovantes pour un développement pharmacologique ultérieur.

## 2 Signalisation des récepteurs membranaires (RCPG)

Afin de coordonner leurs actions et d'accomplir des fonctions physiologiques, les cellules communiquent entre elles à l'aide de signaux extracellulaires, tels que des signaux hormonaux à longue portée. Au niveau des cellules individuelles, ces signaux induisent des réponses biologiques en déclenchant des cascades de signalisation complexes contenant une série de molécules et de réactions biochimiques. Une cascade de signalisation est typiquement activée par l'interaction entre un signal extracellulaire et son récepteur. La compréhension des événements de signalisation cellulaire et de leur dynamique spatio-temporelle est une étape clé pour développer des approches pharmacologiques. Les récepteurs qui nous intéressent dans ce projet sont les récepteurs des gonadotrophines, qui appartiennent à la famille des récepteurs couplés aux protéines G (RCPG), constituant une large gamme de cibles pour les médicaments car ils contrôlent de nombreux systèmes tels que la reproduction, le développement embryonnaire ou le système neuronal.

## 3 Signalisation compartimentée

Une fois activée par leur signal extracellulaire correspondant, les RCPG induisent une réponse physiologique au niveau de la membrane plasmique mais aussi dans des vésicules intracellulaires (endosomes) [2]. Le nombre de vésicules dans la cellule, leur taille mais aussi leurs caractéristiques spécifiques ont une évolution dynamique qui façonne la cascade de signalisation (Fig.1). Cette spatialité et cette intensité de la réponse ont un intérêt

physiologique et suggèrent de prendre en compte l'ensemble de la réponse intracellulaire pour le développement de molécules pharmacologiques. Un aspect essentiel de la structure spatiale de la cascade de signalisation est le trafic des récepteurs, qui vont et viennent de la membrane plasmique aux vésicules de manière dynamique [1].

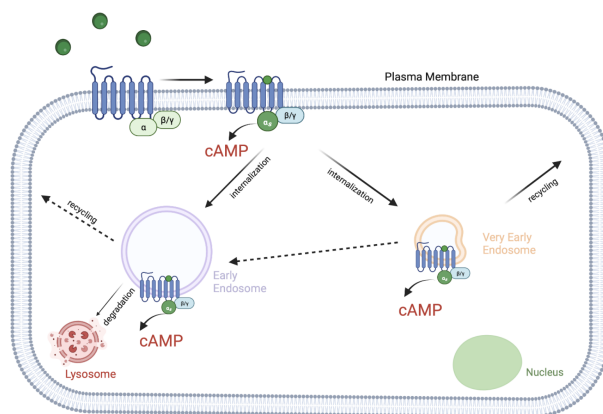


Figure 1: Schéma biologique représentant le trafic des RCPG suite à l'activation de ce dernier par un ligand. Le schéma a été réalisé à l'aide de Biorender. Le RCPG une fois activé par son ligand recrute de nouvelles molécules ( $\alpha_s$  et  $\beta/\gamma$  qui vont produire un effecteur, ici *cAMP*, enclenchant au sein de la cellule une réponse cellulaire. Les récepteurs sont internalisés dans plusieurs types de vésicules (endosomes très précoces et endosomes précoces) et vont ensuite être recyclés et/ou dégradés. Les récepteurs présents dans les vésicules produisent aussi l'effecteur *cAMP*, marqueur de la réponse cellulaire.

## 4 Modélisation mathématique

L'objectif général du projet est de modéliser le trafic des récepteurs, afin d'améliorer les connaissances sur les voies de signalisation et aider à la quantification des expériences biologiques *in vitro* en caractérisant la dynamique spatiale et temporelle des voies de signalisation induites par les RCPG.

### 4.1 Réseaux de réactions biochimiques et modèle compartimenté

Les réseaux de réactions chimiques permettent de traduire les cascades de signalisation en un modèle mathématique. Ce formalisme permet de traduire un ensemble de réactions chimiques en un graphe orienté qui définit de manière univoque un système d'équations différentielles ordinaires (EDO) [6]. En particulier, la communauté a beaucoup utilisé ce formalisme pour la modélisation de l'interaction ligand-récepteur [4].

Dans un premier temps, les vésicules et la membrane plasmique peuvent être considérées comme des compartiments fixes et le trafic des récepteurs peut alors être traduit



comme une réaction, comprise comme une transition spatiale d'un compartiment à l'autre. Cette approche, appelée modèle compartimenté ou structuré, considère une globalisation des quantités dans les différents types de compartiments (Fig.??). Cette méthode permet d'utiliser la théorie des réseaux de réactions chimiques pour étudier le comportement du modèle [6]. Nous avons écrit différents modèles avec deux ou trois compartiments incluant des paramètres d'intérêt : taux de recyclage, d'internalisation, de dissociation et d'association. Nous nous sommes appuyés sur diverses hypothèses mathématiques motivées par des considérations biologiques ([5], [3]). Grâce à des expressions analytiques in-

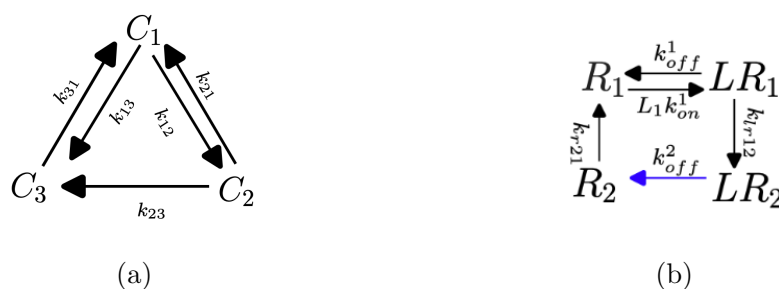


Figure 2: Schémas de quelques modèles étudiés. Fig.2a : Un modèle simplifié représentant le trafic entre trois compartiments, noté  $C_i$ . Fig.2b : Un modèle représentant les interactions ligands-récepteurs ainsi que leur trafic pour deux types de compartiments.

terprétables, nous avons alors étudié l'influence des paramètres révélant que la dynamique de la réponse de signalisation peut être influencée de manière distincte et opposée par le trafic des récepteurs en fonction des paramètres d'interaction ligand-récepteur (Fig.3).

## 4.2 Modèle individu-centré et PDMP

L'une des limites actuelles de l'approche par réseaux de réactions chimiques est de ne pas considérer les dynamiques individuelles des vésicules intracellulaires. De plus, cette approche n'indique pas la quantité de réponse cellulaire produite par chaque vésicule individuelle. Les modèles individu-centré ont une plus grande flexibilité de modélisation et permettent de représenter des phénomènes plus complexes, au prix d'une plus grande complexité pour étudier analytiquement leur comportement. On a alors recours à la simulation numérique. J'ai développé un premier modèle hybride basé sur les processus de Markov déterministes par morceaux (PDMP). Les événements liés au trafic des récepteurs et à la taille des vésicules sont aléatoires, mais les réactions biochimiques, y compris l'interaction ligand-récepteur, sont toujours supposées déterministes. D'une certaine manière, ces hypothèses sont plus proches de la réalité biologique. Nous avons écrit un algorithme afin de simuler ce modèle [7]. L'avantage de cette approche est d'estimer le nombre de vésicules en fonction du temps (Fig/4a) [8] mais aussi de suivre la quantité de réponse cellulaire produite par chaque vésicule (Fig.4b).

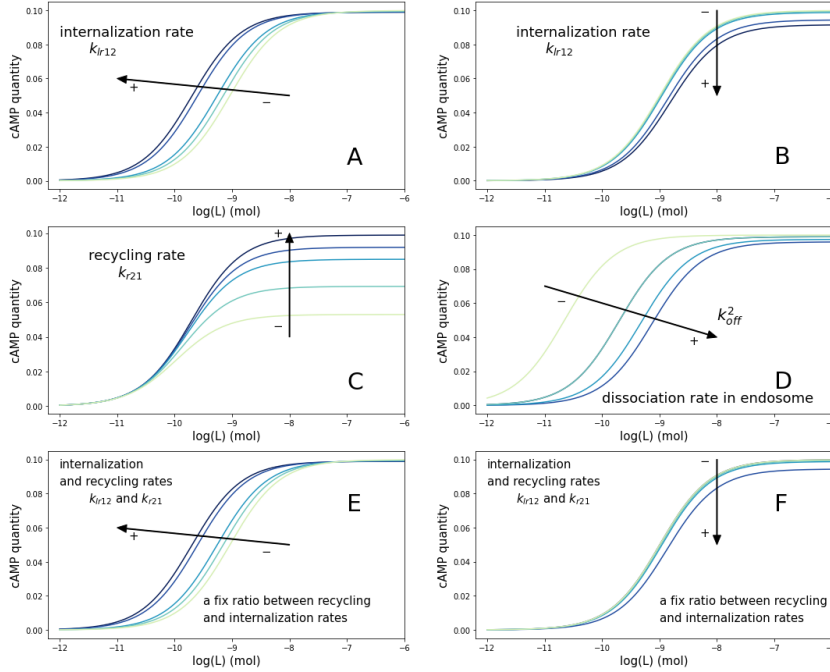


Figure 3: Influence des paramètres lorsque les taux de production et de dégradation sont égaux dans les deux compartiments afin de se concentrer sur l'effet des autres paramètres.

## 5 Perspectives

La première perspective est de calibrer les modèles développés avec des données biologiques récoltés au sein de l'équipe BIOS à l'aide d'une approche statistique d'estimation des paramètres. Ensuite, nous allons chercher à caractériser les états stationnaires du modèle individu-centré, en particulier les conditions d'existence et d'unicité d'une mesure invariante, son support etc. Nous chercherons alors à comparer les approches déterministe et stochastiques. Une possibilité serait d'utiliser la méthode de fermeture des moments sur le modèles individu-centré, afin d'obtenir un ensemble réduit d'EDO, plus facile à analyser et à comparer aux EDO provenant du modèle à compartiments fixes.

Enfin, de nombreuses extensions du modèle peuvent être envisagées, en incluant de nouveaux acteurs moléculaires qui interagissent avec la machinerie de trafic des récepteurs, ou en considérant la diffusion dans l'espace. Dans la poursuite du projet, d'autres formalismes seront développés issus des communautés probabilistes et informatiques. Les formalismes informatiques (P-systems ou processus hybrides) peuvent se rapprocher sous l'angle des PDMP), et fournissent un bon compromis entre complexité et fidélité, permettant notamment de modéliser finement des micro-environnements distincts, qui ont un impact in fine sur la réponse cellulaire.

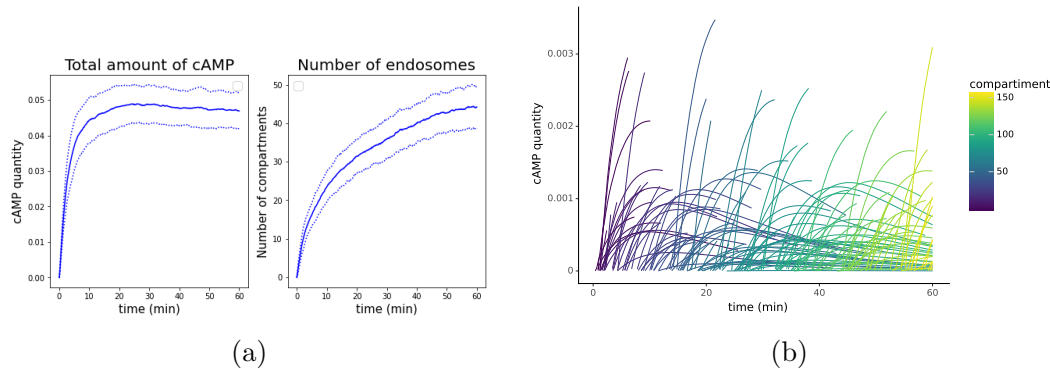


Figure 4: 4a Simulations PDMP. Cette figure est une moyenne de 100 réalisations de l’algorithme. Les lignes en pointillé représentent l’écart-type. (A) représente la quantité totale d’AMPc en fonction du temps et (B) représente le nombre d’endosomes. 4b Quantité de cAMP pour chaque endosome. L’un des avantages du modèle stochastique est de suivre la dynamique individuelle des compartiments. Chaque endosome produit sa propre quantité de cAMP.

## References

- [1] F. Jean-Alphonse, et al. (2014). Spatially Restricted G Protein-coupled Receptor Activity via Divergent Endocytic Compartments. *Journal of Biological Chemistry*, 289 (7): 3960-3977.
- [2] S.E. Crilly, et al. (2021). Compartmentalized GPCR Signaling from Intracellular Membranes. *The Journal of Membrane Biology*, 254 (3): 259-271.
- [3] M. R. Birtwistle and B. N. Kholodenko. (2009). Endocytosis and signalling: A meeting with mathematics. *Molecular Oncology*, 3(4): 308-320.
- [4] J.W. Black and P.Leff. (1983). Operational models of pharmacological agonism. *Proceedings of the Royal Society of London*
- [5] S.R.J. Hoare et al. (2018). Kinetic operational models of agonism for G-protein-coupled receptors. *Journal of Theoretical Biology*, 446 : 168-204.
- [6] M. Feinberg. Foundations of Chemical Reaction Network Theory. (2019) *Applied Mathematical Sciences*, 202.
- [7] M.H.A. Davis. (1984). Piecewise-deterministic markov processes: A general class of non-diffusion stochastic models. *Journal of the Royal Statistical Society, Series B (Methodological)*
- [8] L. Foret et al. (2014) Theory of cargo and membrane trafficking, 46(3): 353–388.

---

THE COASTAL-CARBON MODEL, A GENERIC WETLAND MODEL THAT PREDICTS  
CARBON DYNAMICS

Xaus, Lucila<sup>1,2</sup>; Polsenaere, Pierre<sup>3</sup>; Afonso, Marine<sup>4</sup>; Amman, Benjamin<sup>1</sup>; Dubillot, Bénédicte<sup>1</sup>; Perdrau, Anaïs<sup>1</sup>; Lacoue-Labarthe, Thomas<sup>1</sup>; Gaucherel, Cédric<sup>2</sup>; Dupuy, Christine<sup>1</sup>

<sup>1</sup> LIENSs, UMR7266 CNRS- La Rochelle University, 17000, La Rochelle, France,

[lucila.xaus@univ-lr.fr](mailto:lucila.xaus@univ-lr.fr), [christine.dupuy@univ-lr.fr](mailto:christine.dupuy@univ-lr.fr)

<sup>2</sup> AMAP, CNRS, 34000, Montpellier, France, [cedric.gaucherel@cirad.fr](mailto:cedric.gaucherel@cirad.fr)

<sup>3</sup> Ifremer, LITTORAL, F-17390 La Tremblade, France,

<sup>4</sup> La Rochelle Ville, 17000, La Rochelle, France

## Résumé

Les écosystèmes « Carbone Bleu », tels que les zones humides (marais salants et mangroves), sont essentiels à la réduction des émissions de carbone, en raison de leur capacité potentielle de piégeage du carbone. Cependant, leur effet sur le stockage global du carbone et la dynamique des émissions a été récemment reconnu. Des études menées sur la côte atlantique française ont montré l'importante capacité de puits de carbone ( $-483 \text{ g C m}^{-2} \text{ an}^{-1}$ ) et de piégeage du carbone ( $-345 \text{ g C m}^{-2} \text{ an}^{-1}$ ) de leurs zones humides. Dans ce contexte, les modèles mathématiques sont de plus en plus utilisés pour estimer la dynamique du carbone et les impacts de la séquestration, ainsi que la dynamique sociale. Cette recherche a utilisé l'approche du réseau écologique à événements discrets (EDEN), un modèle dynamique, qualitatif et non déterministe, qui dépend du formalisme des événements discrets. Cette approche a permis d'étudier la dynamique du modèle en se concentrant sur les changements radicaux du système. Les objectifs de cette étude sont : (1) créer un modèle intégré des zones humides de la région métropolitaine de La Rochelle en utilisant le cadre EDEN, incluant les 7 topologies de zones humides, (2) voir si les trajectoires écologiques décrites entre les zones humides sont trouvées et (3) étudier toutes les voies possibles dans lesquelles les zones humides peuvent se comporter comme des puits de carbone. La compréhension de la dynamique écologique de ces environnements permet de prévoir la dynamique du carbone et d'éclairer les stratégies de gestion. Le cadre EDEN a permis de prédire avec succès l'évolution des zones humides dans le temps, tandis que le modèle Coastal-Carbon a permis de prédire la dynamique du carbone dans sept types de zones humides. Il s'agit du premier modèle qualitatif et possibiliste permettant d'évaluer la dynamique du carbone dans un écosystème social générique de zone humide.

**Mots-clés.** Modèle qualitative ; modèle à événements discrets; possibiliste; dynamiques de carbone; écosystèmes aquatiques

## Abstract

“Blue Carbon” ecosystems such as wetlands (i.e., salt marshes and mangroves), are essential in reducing carbon emissions, due to their potential carbon sequestration capacity. However, their effect on global carbon storage and emission dynamics has been recently recognized. Studies at the French Atlantic coast, showed the important carbon sink ( $-483 \text{ g C m}^{-2} \text{ yr}^{-1}$ ), and carbon sequestration ( $-345 \text{ g C m}^{-2} \text{ yr}^{-1}$ ) capacity of their wetlands. In this context, mathematical models are increasing their use for estimating carbon dynamics and

---

sequestration impacts as well as social dynamics. This research used the Ecological Discrete-Event Network (EDEN) approach, a dynamic, qualitative and non-deterministic model, which depend on discrete-event formalism. This approach allowed to study the model dynamics by focusing in drastic changes on the system. The aims of this study are to: (1) create an integrated wetlands model of la Rochelle metropolitan area using the EDEN framework, including all 7 wetlands topologies, (2) see if the describes ecological trajectories between wetlands are found and (3) study every possible pathway in which wetlands may behave as carbon sinks. Understanding the ecological dynamics of these environments helps predict carbon dynamics and inform management strategies. The EDEN framework successfully predicted wetland changes over time, while the Coastal-Carbon model predicted carbon dynamics across seven wetland types. This is the first qualitative, possibilistic model for assessing carbon dynamics in a generic wetland social-ecosystem.

**Keywords.** qualitative modelling; discrete-event model; possibilism; carbon dynamics; aquatic ecosystems

## 1. Introduction

In order to reduce anthropogenic carbon emissions, “Blue Carbon” ecosystems such as wetlands (i.e., salt marshes and mangroves), are essential due to their potential carbon sequestration capacity (Chmura et al., 2003; Duarte et al., 2005; Greiner et al., 2013; Mcleod et al., 2011). Wetlands are crucial from an ecological point of view, by favoring biodiversity through increasing carbon storage in their soils, in addition to acting as regulatory systems of economy and society, thru minimizing flooding risks and help improving water quality (Boutwell and Westra, 2016; Duarte et al., 2013, 2005; Ghermandi et al., 2008; Monnoyer-Smith, 2019; Nayak and Bhushan, 2022; Sutton-Grier et al., 2014). However, anthropogenic impacts may be causing important changes in wetlands water regime and soil alteration, impacting highly their physic-chemical cycling (Bertness et al., 1987; Fitzpatrick, 2004; Snow and Vince, 1984).

Mathematical models are becoming a frequently use tool for estimating GHG and carbon sequestration impacts on landscape management (Baustian et al., 2021; Gilhespy et al., 2014; Zhan et al., 2024). A limited number of studies have modeled costal environments as a result of their large physicochemical, ecological, and temporal variability (Alizad et al., 2016; Barbier, 2013; Barbier et al., 2011; Costanza, 1997; Fagherazzi and Priestas, 2010; Ganju et al., 2017; Ward et al., 2020), especially their carbon dynamics. Even though quantitative modeling method is the most frequently used, its procedure often becomes complicated, due to their requirements of precise information (e.g., shape and intensity of a determinate ecological parameter) (McCallum, 2008), which can create problems regarding scop and accuracy between simplification and precision and scop problems (Jakeman and Letcher, 2003; Janssen and van Ittersum, 2007; Kanter et al., 2018). Conversely, qualitative and discreet-event modeling is able to bypass lacking data and high dimension limitations (Gaucherel et al., 2020), reason why its applicability in this context is appropriated. The Ecological Discrete-Event Network (EDEN) approach, is a dynamic, qualitative and non-deterministic model, which depend on discrete-event formalism. This approach allows to study the model dynamics by focusing in drastic changes on the system. EDEN has already been validated and it has given insightful contribution on Trophic Network dynamic modelling (Gaucherel et al., 2024), on how to represent dynamics of social-ecosystem (Gaucherel et al., 2024 b) and in farm dynamic sustainable crop-livestock integration (Cosme

---

et al., 2023).

This study aims to (1) create an integrated wetlands model of la Rochelle metropolitan area using the EDEN framework, including all wetlands topologies (seven), (2) see if the describes ecological trajectories between wetlands are found and (3) study every possible pathway in which wetlands may behave as carbon sinks.

## 2. Materials and Methods: Modeling wetland ecosystems with the EDEN framework

The generic wetland (Coastal-Carbon) model was created using the Ecological Discrete-Event Network (EDEN) modelling framework (Gaucherel and Pommereau, 2019; Pommereau et al., 2022). Variables in this model are Boolean, thus they can only take two values: either “*functionally present*” (“+”) or “*functionally absent*” (“-”), and together they created the system state. System states are computed by “*if-then*” rules (Table S1), which combined variables involved into specific (e.g., physicochemical, ecological or biological) processes. Once a rule fulfills its condition, it can be executed, following a *condition* >> *consequence* dynamic (Gaucherel et al. 2024 b). For example, the rule *sun+*, *nutrient+* >> *phytoplankton+*, indicates that the presence of phytoplankton is conditioned by sun and nutrient presences.

As “*if-then*” rules are updated asynchronously, they apply independently and one at a time, and so, alternative trajectories might be computed from any given state. Furthermore, none of the trajectories are chosen and all possible are computed, opening the widest range of the system dynamic. Therefore, this model is non-deterministic and non-probabilistic; the EDEN framework is thus defined as *possibilistic* (Gaucherel et al. 2020). The representation of the system dynamics in EDEN is called a State-Transition Graph (STG), where system states (here wetlands) and transitions (switch from one state to another) are represented by nodes and edges respectively. Three main topological structures can be represented at a STG: cyclic behavior (or “structural stability”), transient and stable states. Structural stabilities are characterized by a subset of mutually reachable states (Gaucherel and Pommereau 2019), meaning that any change inside the structural stability is reversible (strongly connected component, SCC). When a structural stability cannot be left, it is called a “terminal stability” or “deadlock”, when there is only a single state. Transient states are called “basins”, and are a set of states who shared a common precedent and subsequent state (Gaucherel and Pommereau 2019). Additional information about this method, its illustration and technical details are found in previous works (Cosme et al., 2024; Gaucherel et al., 2024 b; Gaucherel and Pommereau, 2019).

### 2.1. Defining the Coastal-Carbon model

To build the Coastal-Carbon model, 19 variables were chosen with a bibliographic survey, meetings between the scientific experts and field work. The main objective of these models were to answer three main questions:

Q1. Were all (observed) wetland topologies predicted by the model?

Q2. Does the model predict observed transitions (Fig. 2) and theoretical trajectories between wetlands?

Q3. Can all wetland topologies behave as carbon sink wetlands?

Wetlands were defined by specific sets of variables (data not shown).

---

## 2.2. Application and analysis

The model analysis was held in a Jupiter environment, with a user-friendly platform that relies on Python 3.7.5. The result of each EDEN model (such as the Coastal-Carbon model) is an output of state space computed from a specific initial state and that is generated by the application of each possible rule.

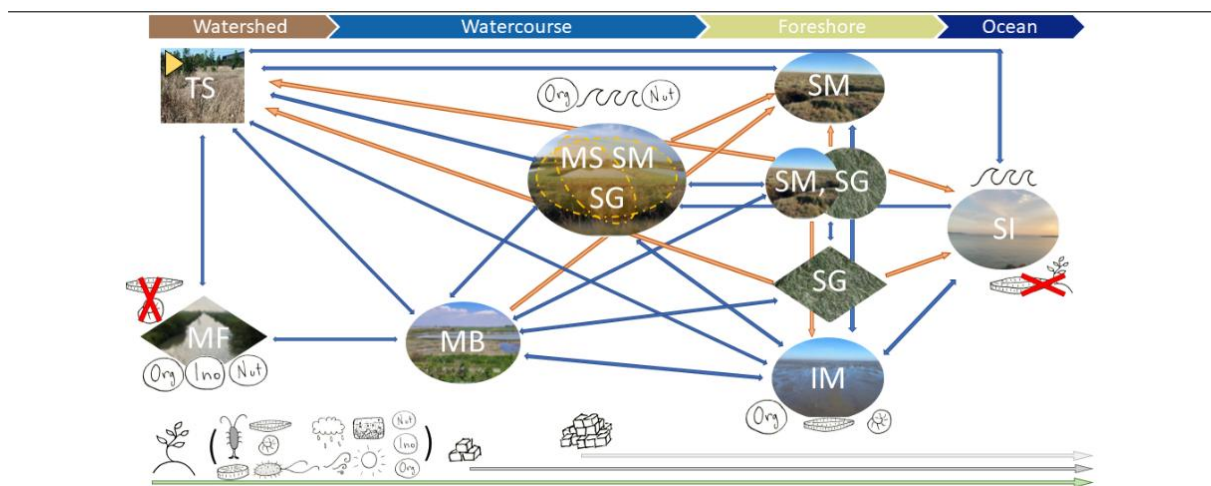
To validate the correct functioning of the Coastal-Carbon model, wetlands transitions, trajectories and their carbon dynamics displayed into the STG were analyzed by temporal logic. This method expressed discrete dynamics as dynamic properties (Cordier et al., 2014; Cosme et al., 2024; Hélias, 2003; Largouet and Cordier, 2020). In particular, the Computation Tree Logic (CTL) (Clarke et al., 2018) aims to test whether a temporal property holds for some or for all possible futures from the initial state (Thomas et al., 2022). This would mean, that CTL is able to tell whether a particular state of a system will be found in the computed STG and will reach a specific future (Cosme et al., 2024). For instance, CTL helps elucidating if a specific wetland topology, like freshwater marsh (MF), never, possibly or necessarily leads to another wetland topology, like brackish marsh (MB).

To highlight chronological sequences the temporal operators U (Until) and F (Finally) were used too. Furthermore, the Boolean logical operators AND ( $\wedge$ ), OR ( $\vee$ ) and NOT ( $\neg$ ) were included in the tested formulae. In addition, an extensional operator ( $\exists$ ) was used. Combining these operators, complex dynamical properties like:  $(\exists(\{MF\} U \{MB\}))$  may be formulated, which translates as MF exist until MB appears. For more information about this methodology see (Cosme et al. 2024). The existence of each wetland, as well as their connections and carbon dynamics were studied using CTL formulae on a STG with an automated model-checker tool (Thierry-Mieg, 2015).

## 3. Results: The Coastal-Carbon model, a dynamic and generic model for wetlands

The model was able to predict every wetland from La Rochelle metropolitan area. Each wetland was reached under the specific set of variables previously defined, except for saltwater marsh, salt meadow and seagrass bed, which were sometimes associated to one another (circle shared by MS, SM and SG; Fig. 1). The analysis of the simplified diagram of wetland topology clarified which wetland was a strongly connected component (SCC), transitory basin or deadlock (Fig. 1).

A merged (simplified) STG was computed to illustrate the main transitions and trajectories between wetland topologies (Fig. 1). The transition on the land-sea continuum (Fig. 1) was clearly observed in the Coastal-Carbon dynamics ( $MF \rightarrow MB \rightarrow MS \rightarrow SM \rightarrow IM \rightarrow SG \rightarrow SI$ , Fig. 1). In addition to expected transitions and trajectories that are validating the model, a few unexpected ones were found. For example, a brackish marsh was not expected to become a salt meadow, a seagrass meadow or even an intertidal mudflat. The meaning of this result will be discussed in the next section.

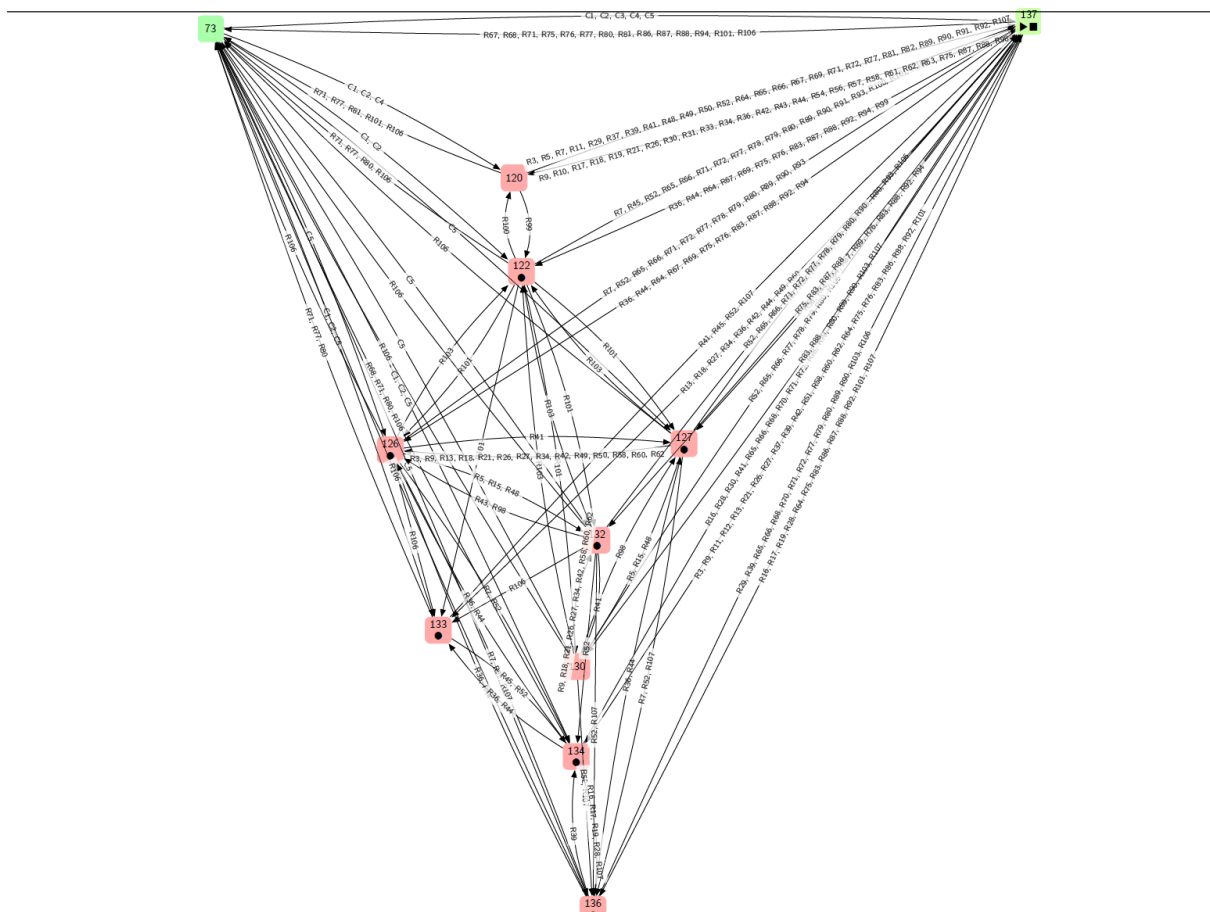


**Figure 5:** Simplified State-Transition Graph of the coastal-carbon model trajectories. Edges illustrate transitions between wetlands (orange: irreversible transition, blue: reversible transition). Variables can be either systematically present (illustrations beside each wetland topology) or absent (red cross). Wetland's topologies appear as freshwater marsh (MF), brackish marsh (MB), salt meadows (MS), seagrass beds (SG), intertidal mudflat (IM), sea inshore (SI), and the transitory state (TS). Saltwater marsh is associated with only seagrass beds (MS, SG) or seagrass beds and salt meadows (MS, SG, SM). Grey, white and green arrows represent variables present from the initial state (yellow triangle). Facultative variables are inside brackets. Photography: Afonso, M.; Dessir, A.; Meslin, S.

### 3.1. Carbon dynamic at each wetland and brief interpretation

Concerning the carbon dynamics in model (controlled by interface fluxes), both carbon source and sink behaviors were observed for each wetland topology, for salt- and fresh-water environments. Moreover, no stability for carbon sink environments was observed, meaning that wetlands can behave as carbon sink and eventually become carbon source (reversibility). Consequently, to minimize carbon source environments, some processes should be evaded. However, wetlands were able to remain as carbon sink environments between each transition (Fig. 2).





**Figure 2:** Aggregated wetlands' State Transition Graph for carbon sink wetlands. The state 73 includes every unrealistic/impossible state (before constraints). Wetland's topologies appear as freshwater marsh (120), brackish marsh (122), salt meadows (132), seagrass beds (130), intertidal mudflat (134), sea inshore (136), and the transitory state (137). Saltwater marsh is associated with only seagrass beds (127) or seagrass beds and salt meadows (126). Circles represents structural stabilities, the triangle the initial state and the square a deadlock. Rules appear connecting each carbon sink state at each wetland

Carbon source wetlands were characterized by respiration processes. Conversely, carbon sink environments were mainly influenced by the death of living organisms, which increase organic carbon (Org+) when sediment settle (Sed-) (data not shown), enabling the carbon sequestration processes (data not shown). These processes resulted in carbon fluxes going either from the atmosphere to the water or to the sediment (Faw- or Fas-, respectively). In addition, rules related to food web processes were also important for achieving carbon sink behavior (data not shown).

#### 4. Conclusion

Wetland transition may occur faster than expected, due to climate change, sea level rise, high temperatures and/or drought. By knowing how these environments ecological dynamic may change in time, predictions about carbon dynamic and management strategies may become clearer. The EDEN framework presented a great capacity of predicting wetland changes in time. Furthermore, the Coastal-Carbon model was able to predict carbon dynamic in 7

different wetland topologies. The Coastal-Carbon model is the first qualitative and possibilistic model ever, for assessing carbon dynamic in a generic wetland ecosystem as a whole. Moreover, using the EDEN framework, it will likely be possible to address carbon neutrality questions.

## References

- Alizad, K., Hagen, S.C., Morris, J.T., Medeiros, S.C., Bilskie, M.V., Weishampel, J.F., 2016. Coastal wetland response to sea-level rise in a fluvial estuarine system. *Earth's Future* 4, 483–497. <https://doi.org/10.1002/2016EF000385>
- Barbier, E.B., 2013. Valuing Ecosystem Services for Coastal Wetland Protection and Restoration: Progress and Challenges. *Resources* 2, 213–230. <https://doi.org/10.3390/resources2030213>
- Barbier, E.B., Hacker, S.D., Kennedy, C., Koch, E.W., Stier, A.C., Silliman, B.R., 2011. The value of estuarine and coastal ecosystem services. *Ecological Monographs* 81, 169–193. <https://doi.org/10.1890/10-1510.1>
- Baustian, M.M., Stagg, C.L., Perry, C.L., Moss, L.C., Carruthers, T.J.B., 2021. Long-Term Carbon Sinks in Marsh Soils of Coastal Louisiana are at Risk to Wetland Loss. *Journal of Geophysical Research: Biogeosciences* 126, e2020JG005832. <https://doi.org/10.1029/2020JG005832>
- Bertness, M.D., Wise, C., Ellison, A.M., 1987. Consumer pressure and seed set in a salt marsh perennial plant community. *Oecologia* 71, 190–200. <https://doi.org/10.1007/BF00377284>
- Boutwell, J.L., Westra, J.V., 2016. The Role of Wetlands for Mitigating Economic Damage from Hurricanes. *JAWRA Journal of the American Water Resources Association* 52, 1472–1481. <https://doi.org/10.1111/1752-1688.12473>
- Chmura, G.L., Anisfeld, S.C., Cahoon, D.R., Lynch, J.C., 2003. Global carbon sequestration in tidal, saline wetland soils. *Global Biogeochemical Cycles* 17. <https://doi.org/10.1029/2002GB001917>
- Clarke, E.M., Henzinger, T.A., Veith, H., Bloem, R. (Eds.), 2018. *Handbook of Model Checking*. Springer International Publishing, Cham. <https://doi.org/10.1007/978-3-319-10575-8>
- Cordier, M.-O., Largouët, C., Zhao, Y., 2014. Model-Checking an Ecosystem Model for Decision-Aid, in: 2014 IEEE 26th International Conference on Tools with Artificial Intelligence. Presented at the 2014 IEEE 26th International Conference on Tools with Artificial Intelligence, pp. 539–543. <https://doi.org/10.1109/ICTAI.2014.87>
- Cosme, M., Koné, A., Pommereau, F., Gaucherel, C., 2024. Improving livelihood through crop-livestock integration: Insights from a farm trajectory model. *Agricultural Systems* 219, 103949. <https://doi.org/10.1016/j.agsy.2024.103949>
- Cosme, M., Pommereau, F., Gaucherel, C., 2023. Farm Trajectories in the South Sudanian Zone of Burkina Faso: How to Achieve a Persistent Crop-Livestock Integration? <https://doi.org/10.2139/ssrn.4635820>
- Costanza, R., 1997. Chapter 3: Valuation and Management of Wetland Ecosystems.
- Duarte, C.M., Losada, I.J., Hendriks, I.E., Mazarrasa, I., Marbà, N., 2013. The role of coastal plant communities for climate change mitigation and adaptation. *Nature Clim Change* 3, 961–968. <https://doi.org/10.1038/nclimate1970>
- Duarte, C.M., Middelburg, J.J., Caraco, N., 2005. Major role of marine vegetation on the oceanic carbon cycle. *Biogeosciences* 2, 1–8. <https://doi.org/10.5194/bg-2-1-2005>
- Fagherazzi, S., Priestas, A.M., 2010. Sediments and water fluxes in a muddy coastline:

- 
- interplay between waves and tidal channel hydrodynamics. *Earth Surface Processes and Landforms* 35, 284–293. <https://doi.org/10.1002/esp.1909>
- Fitzpatrick, R.W., 2004. Changes in soil and water characteristics of natural, drained and re-flooded soils in the Mesopotamian marshlands: Implications for land management planning.
- Ganju, N.K., Defne, Z., Kirwan, M.L., Fagherazzi, S., D’Alpaos, A., Carniello, L., 2017. Spatially integrative metrics reveal hidden vulnerability of microtidal salt marshes. *Nat Commun* 8, 14156. <https://doi.org/10.1038/ncomms14156>
- Gaucherel, C., Cosme, M., Noûs, C., Pommereau, F., 2024. A single changing hypernetwork to represent (social-)ecological dynamics. <https://doi.org/10.1101/2023.10.30.564699>
- Gaucherel, Cedric, Fayolle, S., Savelli, R., Philippine, O., Pommereau, F., Dupuy, C., 2024. Diagnosis of planktonic trophic network dynamics with sharp qualitative changes. *Peer Community Journal* 4. <https://doi.org/10.24072/pcjournal.417>
- Gaucherel, C., Pommereau, F., 2019. Using discrete systems to exhaustively characterize the dynamics of an integrated ecosystem. *Methods in Ecology and Evolution* 10, 1615–1627. <https://doi.org/10.1111/2041-210X.13242>
- Gaucherel, C., Pommereau, F., Hély, C., 2020. Understanding Ecosystem Complexity via Application of a Process-Based State Space rather than a Potential Surface. *Complexity* 2020, e7163920. <https://doi.org/10.1155/2020/7163920>
- Ghermandi, A., van den Bergh, J.C.J.M., Brander, L.M., de Groot, H.L.F., Nunes, P.A.L.D., 2008. The Economic Value of Wetland Conservation and Creation: A Meta-Analysis. <https://doi.org/10.2139/ssrn.1273002>
- Gilhespy, S.L., Anthony, S., Cardenas, L., Chadwick, D., del Prado, A., Li, C., Misselbrook, T., Rees, R.M., Salas, W., Sanz-Cobena, A., Smith, P., Tilston, E.L., Topp, C.F.E., Vetter, S., Yeluripati, J.B., 2014. First 20 years of DNDC (DeNitrification DeComposition): Model evolution. *Ecological Modelling* 292, 51–62. <https://doi.org/10.1016/j.ecolmodel.2014.09.004>
- Greiner, J.T., McGlathery, K.J., Gunnell, J., McKee, B.A., 2013. Seagrass Restoration Enhances “Blue Carbon” Sequestration in Coastal Waters. *PLOS ONE* 8, e72469. <https://doi.org/10.1371/journal.pone.0072469>
- Hélias, A., 2003. Agrégation/abstraction de modèles pour l’analyse et l’organisation de réseaux de flux : application à la gestion des effluents d’élevage à La Réunion (These de doctorat). Montpellier, ENSA.
- Jakeman, A.J., Letcher, R.A., 2003. Integrated assessment and modelling: features, principles and examples for catchment management. *Environmental Modelling & Software, Applying Computer Research to Environmental Problems* 18, 491–501. [https://doi.org/10.1016/S1364-8152\(03\)00024-0](https://doi.org/10.1016/S1364-8152(03)00024-0)
- Janssen, S., van Ittersum, M.K., 2007. Assessing farm innovations and responses to policies: A review of bio-economic farm models. *Agricultural Systems, Special Section: sustainable resource management and policy options for rice ecosystems* 94, 622–636. <https://doi.org/10.1016/j.agsy.2007.03.001>
- Kanter, D.R., Musumba, M., Wood, S.L.R., Palm, C., Antle, J., Balvanera, P., Dale, V.H., Havlik, P., Kline, K.L., Scholes, R.J., Thornton, P., Tittonell, P., Andelman, S., 2018. Evaluating agricultural trade-offs in the age of sustainable development. *Agricultural Systems, Agricultural Systems Perspectives on Global Food Security* 163, 73–88. <https://doi.org/10.1016/j.agsy.2016.09.010>
- Largouet, C., Cordier, M.-O., 2020. Timed Automata Model to Improve the Classification of a Sequence of Images.
- McCallum, H., 2008. *Population Parameters: Estimation for Ecological Models*. John Wiley

---

& Sons.

- Mcleod, E., Chmura, G.L., Bouillon, S., Salm, R., Björk, M., Duarte, C.M., Lovelock, C.E., Schlesinger, W.H., Silliman, B.R., 2011. A blueprint for blue carbon: toward an improved understanding of the role of vegetated coastal habitats in sequestering CO<sub>2</sub>. *Frontiers in Ecology and the Environment* 9, 552–560. <https://doi.org/10.1890/110004>
- Monnoyer-Smith, L., 2019. La séquestration de carbone par les écosystèmes en France. Commissariat général du développement durable.
- Nayak, A., Bhushan, B., 2022. Wetland Ecosystems and Their Relevance to the Environment: Importance of Wetlands, in: *Handbook of Research on Monitoring and Evaluating the Ecological Health of Wetlands*. IGI Global, pp. 1–16. <https://doi.org/10.4018/978-1-7998-9498-8.ch001>
- Pommereau, F., Thomas, C., Gaucherel, C., 2022. EDEN Framework for Interactive Analysis of Ecosystems Models. Presented at the International Workshop on Petri Nets and Software Engineering (PNSE 2022), p. 119.
- Snow, A.A., Vince, S.W., 1984. Plant Zonation in an Alaskan Salt Marsh: II. An Experimental Study of the Role of Edaphic Conditions. *Journal of Ecology* 72, 669–684. <https://doi.org/10.2307/2260075>
- Sutton-Grier, A.E., Moore, A.K., Wiley, P.C., Edwards, P.E.T., 2014. Incorporating ecosystem services into the implementation of existing U.S. natural resource management regulations: Operationalizing carbon sequestration and storage. *Marine Policy* 43, 246–253. <https://doi.org/10.1016/j.marpol.2013.06.003>
- Thierry-Mieg, Y., 2015. Symbolic Model-Checking Using ITS-Tools, in: Baier, C., Tinelli, C. (Eds.), *Tools and Algorithms for the Construction and Analysis of Systems*. Springer, Berlin, Heidelberg, pp. 231–237. [https://doi.org/10.1007/978-3-662-46681-0\\_20](https://doi.org/10.1007/978-3-662-46681-0_20)
- Thomas, C., Cosme, M., Gaucherel, C., Pommereau, F., 2022. Model-checking ecological state-transition graphs. *PLOS Computational Biology* 18, e1009657. <https://doi.org/10.1371/journal.pcbi.1009657>
- Ward, N.D., Bianchi, T.S., Martin, J.B., Quintero, C.J., Sawakuchi, H.O., Cohen, M.J., 2020. Pathways for Methane Emissions and Oxidation that Influence the Net Carbon Balance of a Subtropical Cypress Swamp. *Front. Earth Sci.* 8. <https://doi.org/10.3389/feart.2020.573357>
- Zhan, J., Wang, C., Wang, H., Zhang, F., Li, Z., 2024. Pathways to achieve carbon emission peak and carbon neutrality by 2060: A case study in the Beijing-Tianjin-Hebei region, China. *Renewable and Sustainable Energy Reviews* 189, 113955. <https://doi.org/10.1016/j.rser.2023.113955>

# Liste des auteurs

Alfaro Matthieu, 21–24

Bel Liliane, 12–16

Blouza Adel, 21–24

Brehelin Jacques, 2–5

Carmigniani Rémi, 6–11

Cartier Yoann, 6–11

Chassard Christophe, 31–35

Chauvet Jocelyn, 31–35

Cognot Caroline, 12–16

Darrigade Léo, 17–20

Dhaouadi Nessim, 21–24

Einfalt Thomas, 6–11

Gaillard Mathilde, 25–30

Girardin Valérie, 2–5

Grente Théo, 2–5

Grollemund Paul-Marie, 31–35

Guillot-Le Goff Arthur, 6–11

Jean-Alphonse Frédéric, 17–20

Kioye Togo Jean Yves, 31–35

Kohli Mathieu, 36–39

Latrach Saoussen, 40–43

Métivier David, 6–16

Niquil Nathalie, 2–5

Parey Sylvie, 12–16

Regnault Philippe, 2–5

Vinçon-Leite Brigitte, 6–11

Weckel Chloé, 44–49

Xaus Lucila, 50–58

Yvinec Romain, 17–20, 44–49

