



**HAL**  
open science

## Simulation de la densité limite de champ moyen pour des populations hétérogènes avec interactions

Antonin Della Della Noce, Amélie Mathieu, Paul-Henry P.-H. Cournède

► **To cite this version:**

Antonin Della Della Noce, Amélie Mathieu, Paul-Henry P.-H. Cournède. Simulation de la densité limite de champ moyen pour des populations hétérogènes avec interactions. 22ème Rencontre du Non-Linéaire, Mar 2019, Paris, France. hal-02425831

**HAL Id: hal-02425831**

**<https://centralesupelec.hal.science/hal-02425831>**

Submitted on 31 Dec 2019

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

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

# Simulation de la densité limite de champ moyen pour des populations hétérogènes avec interactions

Della Noce Antonin<sup>1</sup>, Mathieu Amélie<sup>2</sup> & Cournède Paul-Henry<sup>1</sup>

<sup>1</sup> Laboratoire MICS, CentraleSupélec, Université Paris-Saclay, 9 rue Joliot-Curie, 91190, Gif-sur-Yvette

<sup>2</sup> UMR ÉcoSys, INRA AgroParisTech, Route de la ferme, 78850 Thiverval-Grignon

antonin.della-noce@centralesupelec.fr

**Résumé.** Les modèles de populations hétérogènes considérés ici consistent en des systèmes dynamiques dont les conditions initiales sont des variables aléatoires indépendantes et identiquement distribuées, avec autant d'équations qu'il y a d'individus dans la population, et dont les interactions sont exprimées par un potentiel régulier. La distribution de champs moyen, caractérisant la dynamique d'une population avec une infinité d'individus, est alors bien définie. Une méthode pour la simulation numérique, basée sur la régression par processus gaussiens, est présentée puis appliquée sur un modèle de plantes en compétition pour la lumière.

**Abstract.** The heterogeneous population models considered here consist in dynamical systems, whose initial conditions are independent and identically distributed random variables, representing trajectories of individuals interacting with each other through a potential function. The mean-field distribution characterising the dynamics of an infinitely-crowded population is then well-defined. A methodology to estimate this distribution, based on Gaussian processes regression, is presented and applied to a models of plants population in competition for light.

## Introduction

Soit  $\mathcal{X}$  un espace euclidien et  $\Theta$ , un espace Polonais de dimension finie. On considère une fonction  $g : \mathcal{X} \times \Theta \times \mathcal{X} \times \Theta \rightarrow \mathcal{X}$ , que nous appelons dans la suite fonction de transition.  $\mu_0$  est une mesure de probabilité sur  $\mathcal{Z} = \mathcal{X} \times \Theta$ . Une population hétérogène de taille  $N > 1$  est représentée par un système dynamique dont les conditions initiales sont des variables aléatoires échantillonnées de manière indépendante selon la distribution  $\mu_0$  et dont la dynamique est décrite par la fonction  $g$ .

$$(X_i^0, \theta_i)_{1 \leq i \leq N} \sim \mu_0^{\otimes N} \text{ le produit tensoriel de la mesure } \mu_0$$

$$\forall i \in \{1, \dots, N\}, \begin{cases} X_i(0) = X_i^0 \\ \forall t \in \mathbb{R}_+, \frac{dX_i}{dt}(t) = \frac{1}{N-1} \sum_{1 \leq j \leq N, j \neq i} g(X_i(t), \theta_i, X_j(t), \theta_j) \end{cases} \quad (1)$$

Dans certain cas,  $g$  peut être vue comme une force résultant de l'interaction entre deux particules dans la population. Cette interaction dépend de l'état  $X_i$  de l'individu  $i$  (e.g. sa position, sa vitesse, ...), mais également de caractéristiques intrinsèques à l'individus regroupées dans la variable  $\theta_i \in \Theta$  (e.g. sa masse, sa couleur,...), qui sont variables d'un individu à l'autre selon la distribution  $\mu_0$  mais que l'on considère constantes au cours du temps. Si la distribution  $\mu_0^\theta$  de la variable aléatoire  $\theta$  est non réduite à une distribution de Dirac, la population est dite hétérogène dans le sens où elle met en relation des individus différents. Ce formalisme est particulièrement pertinent quand le système modélisé consiste en un ensemble d'êtres vivants, avec une plus ou moins grande diversité génétique.

L'équation (1) implique une équation de transport sur la mesure empirique de la population, définie par  $\mu_N[t] = \frac{1}{N} \sum_{i=1}^N \delta_{X_i(t)}$ . Cette équation s'exprime sous sa forme faible de la manière suivante : pour toute fonction test  $\varphi : \mathbb{R}_+ \times \mathcal{Z} \rightarrow \mathbb{R}$  (i.e. infiniment dérivable et à support compact) et tout temps  $T \in \mathbb{R}_+$

on a

$$\int_0^T \int_{\mathcal{Z}} \left( \frac{\partial \varphi}{\partial t}(t, X, \theta) + \frac{\partial \varphi}{\partial X}(t, X, \theta) \cdot \left( \frac{N}{N-1} \int_{\mathcal{Z}} g(X, \theta, X', \theta') \mu_N[t](dX', d\theta') - \frac{1}{N-1} g(X, \theta, X, \theta) \right) \right) \mu_N[t](dX, d\theta) dt = \int_{\mathcal{Z}} \varphi(T, X, \theta) \mu_N[T](dX, d\theta) - \int_{\mathcal{Z}} \varphi(0, X, \theta) \mu_N[0](dX, d\theta) \quad (2)$$

Ainsi la dynamique de la population avec un nombre infini d'individus ( $N = +\infty$ ) peut être vue comme la solution du problème dont la formulation faible est obtenue en passant formellement à la limite en  $N$  dans l'équation (2). On cherche alors l'application  $t \in \mathbb{R}_+ \mapsto \mu[t] \in \mathcal{P}(\mathcal{Z})$  à valeurs dans l'ensemble des mesures de probabilité sur  $\mathcal{Z}$ , telle que pour toute fonction test  $\varphi$  et pour tout temps  $T \in \mathbb{R}_+$

$$\begin{aligned} & \int_0^T \int_{\mathcal{Z}} \left( \frac{\partial \varphi}{\partial t}(t, X, \theta) + \frac{\partial \varphi}{\partial X}(t, X, \theta) \cdot \int_{\mathcal{Z}} g(X, \theta, X', \theta') \mu[t](dX', d\theta') \right) \mu[t](dX, d\theta) dt \\ &= \int_{\mathcal{Z}} \varphi(T, X, \theta) \mu[T](dX, d\theta) - \int_{\mathcal{Z}} \varphi(0, X, \theta) \mu_0(dX, d\theta) \end{aligned} \quad (3)$$

Si l'on est en mesure de démontrer que pour tout temps  $t$  la mesure  $\mu[t]$  est absolument continue par rapport à la mesure de Lebesgue  $\lambda^{\otimes d_z}$  sur  $\mathcal{Z}$ , et que l'on note  $(X, \theta) \in \mathcal{Z} \mapsto f(t, z) \in \mathbb{R}_+$  la densité associée à  $\mu[t]$ , et que celle-ci est suffisamment régulière, la formulation (3) entraîne, par intégration par parties, la formulation forte suivante bien plus évocatrice, assez similaire à l'équation de Vlasov pour les plasmas.

$$\forall (X, \theta) \in \mathcal{Z}, \begin{cases} f(0, X, \theta) = f_0(X, \theta) \text{ la densité associée à } \mu_0 \\ \forall t \in \mathbb{R}_+, \frac{\partial f}{\partial t}(t, X, \theta) + \operatorname{div}_X \left( f(t, X, \theta) \int_{\mathcal{Z}} g(X, \theta, X', \theta') f(t, X', \theta') \lambda^{\otimes d_z}(dX', d\theta') \right) = 0 \end{cases} \quad (4)$$

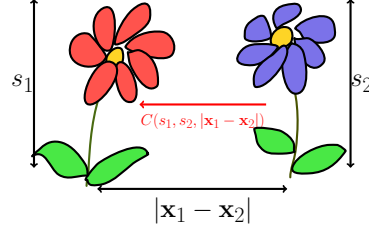
Les équations (1), (2) et (3) correspondent à différents niveaux d'observation du système : l'équation (1) est à une échelle individuelle ; l'équation (2) décrit la dynamique d'un nuage de points où les individus ne sont plus distingués entre eux ; enfin l'équation (3) se place à une échelle *macroscopique* où l'ensemble de la population est perçue comme un continuum. Être en mesure de résoudre numériquement le problème (3) peut permettre d'approcher la dynamique d'une grande population sans avoir à simuler l'ensemble des individus la constituant. La section 1 donne un exemple de modèle de population hétérogène issu de [2]. La section 2 donne une résolution théorique du problème (3), basée sur la méthode des flots caractéristiques [1], dans le cas particulier du modèle présenté en section 1. Ces développements orientent vers une certaine méthodologie pour la simulation numérique décrite dans la section 3.

## 1 Exemple du modèle de Schneider pour la croissance de plantes en compétition

Nous spécifions les propos de l'introduction sur un exemple simple. L'article [2] propose une série de modèles pour décrire la croissance de populations de plantes (*Arabidopsis thaliana*) en compétition. Nous avons choisi le modèle qui avait le plus de pertinence statistique, tout en ayant de bonne propriété de régularité. Ces modèles de compétition ont ensuite été repris dans [3] pour rendre compte de la dynamique d'une forêt sur plusieurs décénies.

Dans ce modèle, dont le paramétrage est donné dans la figure 1, la croissance d'une plante isolée est donnée par l'équation différentielle suivante sur sa taille  $s$  :

$$\begin{cases} s(0) = s^0 \\ \frac{ds}{dt}(t) = \gamma(S - s(t)) \\ \Rightarrow s(t) = S - (S - s^0)e^{-\gamma t} \end{cases} \quad (5)$$



**Figure 1.** Paramétrage du modèle de Schneider : la *compétition* exercée par la plante 2 sur la plante 1 est notée  $C(s_1, s_2, |\mathbf{x}_1 - \mathbf{x}_2|)$ . Elle dépend des tailles des parties aériennes des deux plantes,  $s_1$  et  $s_2$ , et de leur distance  $|\mathbf{x}_1 - \mathbf{x}_2|$ .

Cette croissance est paramétrée par des caractéristiques propres que sont  $\gamma$ , le taux de croissance, et  $S$  la taille d'équilibre de la plante. En présence de plantes concurrentes, l'équation (5) est modifiée : la taille de plante  $i \in \{1, \dots, N\}$  est influencée par les tailles des autres plantes dans la population via l'équation

$$\forall i \in \{1, \dots, N\}, \begin{cases} s_i(0) = s^0 \\ \frac{ds_i}{dt}(t) = \gamma_i \left( S_i \left( 1 - \frac{1}{N-1} \sum_{1 \leq j \leq N, j \neq i} C(s_i(t), s_j(t), |\mathbf{x}_i - \mathbf{x}_j|) \right) - s_i(t) \right) \end{cases} \quad (6)$$

avec  $C(s_1, s_2, |\mathbf{x}_1 - \mathbf{x}_2|) = \frac{s_2}{2S_M \left( 1 + \frac{|\mathbf{x}_1 - \mathbf{x}_2|^2}{\sigma_x^2} \right)} \left( 1 + \tanh \left( \frac{s_2 - s_1}{\sigma_s} \right) \right)$

Le terme de compétition  $C$  est normalisé de manière à être une grandeur sans dimension à valeur dans  $[0; 1]$  : la compétition exercée par la plante 2 sur la plante 1 est d'autant plus importante que la plante 2 est grande dans l'absolue, est grande par rapport à la plante 1 et enfin qu'elle est proche de la plante 1. Les paramètres propres à l'individu dans ce modèle sont  $\theta = (\mathbf{x}, S, \gamma) \in \Theta = [0; L]^2 \times [S_m; S_M] \times [\gamma_m; \gamma_M]$  avec  $L, S_m, S_M, \gamma_m, \gamma_M$  des constantes connues. On suppose qu'ils sont distribués dans la population comme des échantillons i.i.d. d'une densité  $p_0^\theta$ . Nous avons choisi cette expression pour la densité  $p_0^\theta$  :

$$\forall (x, y, S, \gamma) \in \Theta, p_0^\theta(x, y, S, \gamma) = \frac{\mathbb{I}\{0 \leq x, y \leq L\}}{L^2} \cdot \frac{\mathbb{I}\{S_1(x) \leq S \leq S_2(x)\}}{\sigma_s} \cdot \frac{\mathbb{I}\{\gamma_1(y) \leq \gamma \leq \gamma_2(y)\}}{\sigma_\gamma} \quad (7)$$

où  $S_1(x) = S_m + \frac{x}{L}(S_M - \sigma_s - S_m)$ ,  $S_2(x) = S_1(x) + \sigma_s$

$\gamma_1(y) = \gamma_m + \frac{y}{L}(\gamma_M - \sigma_\gamma - \gamma_m)$ ,  $\gamma_2(y) = \gamma_1(y) + \sigma_\gamma$

Cette densité signifie que les plantes sont réparties uniformément sur le carré  $[0; L]^2$ , que les plantes proches de  $x = L$  ont tendance à avoir une grande taille d'équilibre et que les plantes proches de  $y = L$  ont tendance à avoir une croissance rapide. Ainsi la distribution initiale de la population est  $\mu_0 = \delta_{s^0} \otimes (p_0^\theta \lambda^{\otimes 4})$  car la taille initiale est constante égal à  $s^0$ .

## 2 Résolution théorique et propagation du chaos

Par souci de concision, nous présenterons la résolution théorique de l'équation (3) dans le cas du modèle de Schneider, qui ne correspond pas tout à fait au cas général, car la distribution marginale de l'état initial est un Dirac (centré en  $s^0$ ). On a dans ce cas particulier  $g : (s, \theta, s', \theta') \in \mathbb{R} \times \Theta \times \mathbb{R} \times \Theta \mapsto \gamma(S(1 - C(s, s', |\mathbf{x} - \mathbf{x}'|)) - s)$ .

Pour résoudre le problème de transport (3), on peut considérer l'équation de l'évolution de la taille  $s(t, \theta)$  d'un individu en compétition avec un continuum d'individus [1].

$$\forall \theta \in \Theta, \begin{cases} \mathbf{s}(0, \theta) = s^0 \\ \forall t \in \mathbb{R}_+, \frac{\partial s}{\partial t}(t, \theta) = \int_{\Theta} g(s(t, \theta), \theta, s(t, \theta'), \theta') p_0^\theta(\theta') \lambda^{\otimes 4}(d\theta') \end{cases} \quad (8)$$

On peut voir en cette équation la version continue de l'équation (6), où la moyenne empirique est remplacée par l'espérance théorique. On peut démontrer que l'équation fonctionnelle (8) admet une unique solution  $(t, \theta) \in \mathbb{R}_+ \times \Theta \mapsto s(t, \theta)$ . La preuve consiste en une généralisation du théorème Cauchy-Lipschitz à des équations différentielles dont la condition initiale est une distribution de probabilité, et dont l'inconnue évolue dans un espace euclidien classique. Le flot caractéristique  $t \in \mathbb{R}_+ \mapsto s(t, \theta)$  est la taille d'un individu au sein d'une population infinie, cet individu ayant une position, une taille d'équilibre isolée et une vitesse de croissance données par  $\theta$ . Il permet donc de déterminer la dynamique de l'ensemble du continuum. On peut ainsi montrer que si  $\mu[t]$  est la distribution de la variable aléatoire  $(s(t, \theta), \theta)$ , avec  $\theta$  une variable aléatoire de densité  $p_0^\theta$ , alors  $t \mapsto \mu[t]$  est l'unique solution du problème (3) dans le cas du modèle de Schneider. Il découle de ce résultat une expression de l'action de  $\mu[t]$  sur les fonctions test  $\varphi$ .

$$\int_{\mathbb{R} \times \Theta} \varphi(s, \theta) \mu[t](ds, d\theta) = \int_{\Theta} \varphi(s(t, \theta), \theta) p_0^\theta(\theta) \lambda^{\otimes 4}(d\theta) \quad (9)$$

On pose  $s_1(t; \theta_1, \dots, \theta_N)$  la variable aléatoire donnant la taille de l'individu 1 dans une population finie de caractéristiques  $(\theta_1, \dots, \theta_N)$ , elles-mêmes variables aléatoires indépendantes de densité  $p_0^\theta$ . On a alors la convergence presque sûre suivante

$$s_1(t; \theta_1, \dots, \theta_N) \xrightarrow[N \rightarrow +\infty]{p.s.} s(t, \theta_1) \quad (10)$$

Cela implique en particulier que les individus interagissant avec une population infinie sont indépendants, alors qu'ils ne le sont pas pour une population de taille finie. Cette propriété de factorisation asymptotique, appelée propagation du chaos, offre un autre attrait à la simulation de la distribution de champ moyen  $\mu[t]$ , pouvant permettre de simplifier l'inférence statistique sur un grand système complètement corrélé du fait de l'interaction de ses constituants.

### 3 Méthode numérique

La méthode numérique choisie ici s'inspire largement de la méthode des caractéristiques utilisées dans la résolution théorique de l'équation (8). Cette méthodologie n'a pas recours à un maillage, car l'espace  $\Theta$  est de dimension 4. La preuve de consistance de ce schéma est un travail en cours. Bien que les simulations numériques semblent encourageantes, et malgré notre ferme intuition sur la consistance de cette méthode, nous incitons à une certaine prudence quant aux résultats présentés. Le fait que  $g$  soit non-linéaire par rapport aux états  $s$  et  $s'$  rend la tâche plus complexe : dans le cas opposé, si  $g$  est linéaire par rapport aux variables d'états, on peut démontrer que le flot caractéristique est une fonction linéaire de l'état ; la résolution numérique revient alors à identifier les coefficients.

Partons de l'équation (8) et considérons le schéma d'Euler explicite associé à cette équation intégral-différentielle. Pour un pas de temps  $\Delta t > 0$  choisi suffisamment petit (la condition CFL n'est pas encore déterminée), on considère la suite de fonctions  $(s_n)_{n \in \mathbb{N}}$  donnée par l'équation de récurrence suivante :

$$\forall \theta \in \Theta, \begin{cases} s_0(\theta) = s^0 \\ \forall n \in \mathbb{N}, s_{n+1}(\theta) = s_n(\theta) + \Delta t \int_{\Theta} g(s_n(\theta), \theta, s_n(\theta'), \theta') p_0^\theta(\theta') \lambda^{\otimes 4}(d\theta') \end{cases} \quad (11)$$

Cette suite de fonctions ne peut être calculée exactement en pratique du fait que l'intégrale n'est pas analytique. On peut en revanche avoir une estimation précise de  $s_n(\theta)$  en remplaçant l'intégrale par une

moyenne empirique sur un échantillon de la loi  $p_0^\theta$ . Soit  $\omega_N = (\theta_i^\omega)_{1 \leq i \leq N}$  une collection de variables aléatoires de même loi  $p_0^\theta$  et mutuellement indépendantes. On considère la suite de fonctions aléatoires paramétrées par  $\omega_N$ ,  $(s_n(\cdot, \omega_N))_{n \in \mathbb{N}}$ , qui est définie par l'équation de récurrence suivante.

$$\forall \theta \in \Theta, \begin{cases} s_0(\theta, \omega_N) = s^0 \\ s_n(\theta, \omega_N) = s_n(\theta, \omega_N) + \frac{\Delta t}{N} \sum_{i=1}^N g(s_n(\theta, \omega_N), \theta, s_n(\theta_i^\omega), \theta_i^\omega) \end{cases} \quad (12)$$

Cette fonction  $s_n(\cdot, \omega_N)$  est a priori un bon estimateur de la fonction déterministe  $s_n$  pour une grande taille d'échantillon  $N$ , mais elle ne permet pas d'avoir une vision globale de la fonction  $s_n$  sur l'ensemble de l'espace  $\Theta$ , car avoir une estimation de la valeur de  $s_{n_0}(\theta_0)$  pour un certain entier  $n_0$  et un certain  $\theta_0$  nécessite de calculer toutes les valeurs antérieures  $s_1(\theta_0, \omega_N), \dots, s_{n_0-1}(\theta_0, \omega_N)$ . Nous allons donc restreindre l'évaluation de la fonction  $s_n(\cdot, \omega_N)$  sur un ensemble fixe de taille raisonnable  $\Theta_{1:M} = (\theta_j)_{1 \leq j \leq M}$ , que nous appellerons ensemble d'apprentissage, puis étendre ces valeurs à l'ensemble de l'espace  $\Theta$  par interpolation. Le choix de la base de fonctions d'interpolation est important si l'on souhaite que l'information se propage efficacement des points  $\Theta_{1:M}$  à l'espace  $\Theta$ . Dans notre cas, ce choix a été motivé par une approximation de la distribution de la variable aléatoire  $s_n(\theta, \omega_N)$ , approximation permettant de placer le problème d'interpolation dans le cadre de la régression par processus gaussiens [4]. Nous faisons donc l'approximation suivante pour  $N$  suffisamment grand :

$$\forall n \in \mathbb{N}^*, \forall \theta \in \Theta, s_n(\theta, \omega_N) \sim \mathcal{N} \left( s_n(\theta), \frac{\Delta t^2}{N} \text{Var}_{\theta'}(g(s_{n-1}(\theta), \theta, s_{n-1}(\theta'), \theta')) \right) \quad (13)$$

Cette approximation semble justifiée par le théorème de la limite centrale, cela est vrai pour  $n = 1$ , mais faux pour  $n \geq 2$ . En revanche, cette approximation remplit un certain nombre de conditions cohérentes avec l'intuition que l'on peut se faire de la suite  $(s_n(\cdot, \omega_N))_{n \in \mathbb{N}}$ , essentiellement au niveau de l'expression de la variance : les fluctuations de la variable  $s_n(\theta, \omega_N)$  sont d'autant plus faible que  $\Delta t$  est petit et que  $N$  est grand, enfin cette expression de la variance serait exacte si la fonction  $g$  et  $s_{n-1}$  étaient linéaires. Le rôle de cette approximation est plus d'orienter le choix de la base d'interpolation. On choisit ainsi le noyau d'interpolation suivant.

$$\forall \theta, \theta' \in \Theta, k_n(\theta, \theta') = \frac{\Delta t^2}{N} \hat{\text{Cov}}_{\theta_1}(g(m^s(\theta, \beta_{n-1}), \theta, m^s(\theta_1, \beta_{n-1}), \theta_1), g(m^s(\theta', \beta_{n-1}), \theta', m^s(\theta_1, \beta_{n-1}), \theta_1)) \quad (14)$$

où  $\hat{\text{Cov}}_{\theta_1}$  est la covariance empirique que l'on peut évaluer en utilisant l'échantillon  $\omega_N$ , et  $m_n^s$  est une fonction paramétrique identifiée en minimisant l'erreur de moindre carré sur l'échantillon  $\omega_N$ .

$$\beta_n = \underset{\beta}{\text{argmin}} \sum_{i=1}^N (m^s(\theta_i^\omega, \beta) - s_n(\theta_i^\omega, \omega_N))^2 \quad (15)$$

Nous avons choisi dans notre cas une fonction quadratique pour  $m^s$ , soit  $\beta = (b, \mathbf{v}, B) \in \mathbb{R} \times \mathbb{R}^4 \times \mathcal{M}_4(\mathbb{R})$  et  $m^s(\theta, \beta) = b + \mathbf{v}^\top \theta + \text{Tr}(B^\top \theta \theta^\top)$ . Cette première approximation paramétrique a l'avantage de capturer précisément les variations de la fonction  $\theta \mapsto s_n(\theta, \omega_N)$  et d'être simple à identifier. On fait ainsi un nouveau niveau d'approximation en substituant les fonctions  $m^s(\cdot, \beta_n)$  aux fonctions  $s_n$  dans la distribution supposée dans l'équation (13). En fin de compte si l'on considère la distribution jointe du vecteur aléatoire  $(s_n(\theta, \omega_N); s_n(\Theta_{1:M}, \omega_N))$  pour  $\theta \in \Theta$  et  $s_n(\Theta_{1:M}, \omega_N) = (s_n(\theta_j, \omega_N))_{1 \leq j \leq M}$ , on a

$$\begin{pmatrix} s_n(\theta, \omega_N) \\ s_n(\Theta_{1:M}, \omega_N) \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} m^s(\theta, \beta_n) \\ m^s(\Theta_{1:M}, \beta_n) \end{pmatrix}, \begin{pmatrix} k_n(\theta, \theta) & k_n(\theta, \Theta_{1:M}) \\ k_n(\theta, \Theta_{1:M}) & k_n(\Theta_{1:M}, \Theta_{1:M}) \end{pmatrix} \right) \quad (16)$$

À partir du calcul des valeurs de  $s_n(\Theta_{1:M}, \omega_N)$ , on peut ainsi en déduire la distribution de  $s_n(\theta, \omega_N)$  connaissant ces valeurs en effectuant un conditionnement au sein d'un vecteur gaussien, soit

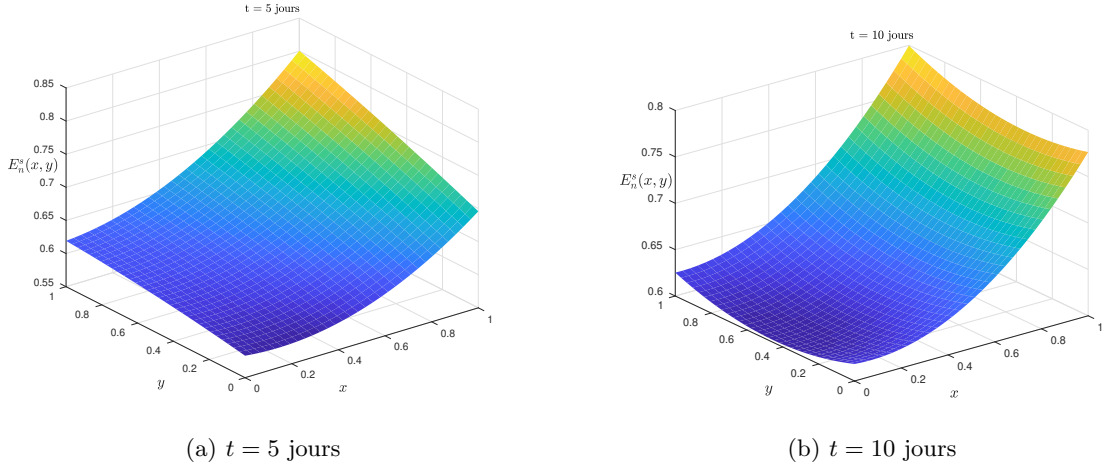
$$s_n(\theta, \omega_N) | s_n(\Theta_{1:M}, \omega_N) \sim \mathcal{N}(m^s(\theta, \beta_n) + k_n(\theta, \Theta_{1:M}) k_n(\Theta_{1:M}, \Theta_{1:M})^{-1} (s_n(\Theta_{1:M}, \omega_N) - m^s(\Theta_{1:M}, \beta_n)), k_n(\theta, \theta) - k_n(\theta, \Theta_{1:M}) k_n(\Theta_{1:M}, \Theta_{1:M})^{-1} k_n(\Theta_{1:M}, \theta)) \quad (17)$$

L'estimateur final de la fonction  $s_n$  peut être pris comme la moyenne de cette distribution conditionnelle, que l'on peut noter  $\tilde{s}_n(\theta, \omega_N)$  et qui a la forme suivante

$$\tilde{s}_n(\theta, \omega_N) = m^s(\theta, \beta_n) + \sum_{j=1}^M \alpha_j k_n(\theta, \theta_j) \quad (18)$$

$$\text{où } (\alpha_1, \dots, \alpha_M)^\top = k_n(\Theta_{1:M}, \Theta_{1:M})^{-1} (s_n(\Theta_{1:M}, \omega_N) - m^s(\Theta_{1:M}, \beta_n))$$

Les graphes qui suivent présentent l'évolution de la surface  $(x, y) \in [0; L]^2 \mapsto E_n^s(x, y) = \mathbb{E}_\theta (\tilde{s}_n(\theta, \omega_N) | x, y)$ , l'espérance de la taille des plantes conditionnellement à leur position, avec l'itération du schéma  $n$ .



**Figure 2.** Ces simulations ont été réalisées avec la configuration suivante :  $\Delta t = 0.1$  jours,  $N = 100$ ,  $M = 100$ ,  $L = 1$  m,  $S_m = 0.8$  m,  $S_M = 1$  m,  $\gamma_m = 0.1$  jours<sup>-1</sup>,  $\gamma_M = 1$  jours<sup>-1</sup>,  $s^0 = 0.5$  m

## Conclusion et perspectives

Les simulations réalisées sur l'exemple simple du modèle de Schneider sont prometteuses : la comparaison de  $\tilde{s}_n(\theta, \omega_N)$  et de  $s_n(\theta, \omega_N)$  sur un ensemble test de  $\theta$ , n'ayant pas servi à l'interpolation, donnent des erreurs relatives inférieures à  $10^{-3}\%$  en moyenne. En revanche, l'analyse de stabilité et de consistance du schéma numérique proposé reste à faire. Dans le cas plus complexe où l'état initial admet une densité de probabilité, il ne s'agit pas seulement de reconstruire le flot caractéristique, mais également sa fonction réciproque, et la question du choix d'une *bonne* base de fonctions d'interpolation se pose. Cette méthode permettant de simuler en temps raisonnables les distributions limite de champ moyen ouvre des perspectives pour l'inférence statistique de modèle de populations avec interactions utilisant des approximations variationnelles.

## Références

1. GOLSE, F. On the Dynamics of Large Particle Systems in the Mean Field Limit (2013)
2. SCHNEIDER M., LAW R., ET ILIAN J. Quantification of neighbourhood-dependent plant growth by Bayesian hierarchical modelling. *Journal of Ecology* (2006)
3. NAKAGAWA Y., YOKOZAWA M., ET HARA T. Competition among plants can lead to an increase in aggregation of smaller plants around larger ones. *Ecological Modelling* (2015)
4. RASMUSSEN, C. E. Gaussian processes in machine learning. *Springer* (2003)