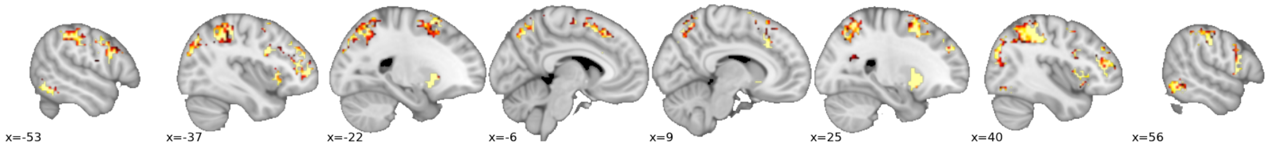
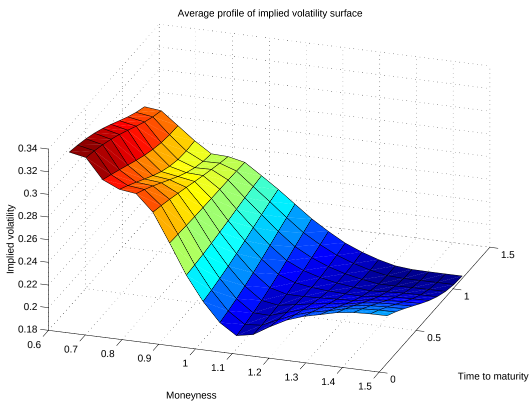
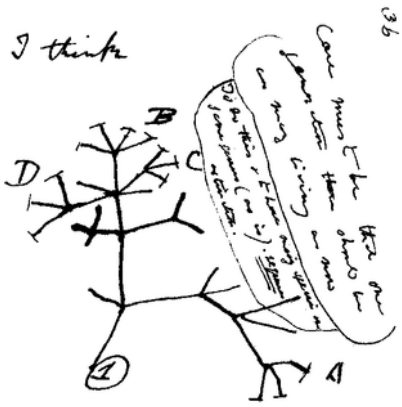
La $\mathbb{R}$echerche en mathématiques appliquées







Pourquoi? Par qui? Comment?

Cette brochure a été financée par les Projets Exploratoires Premier Soutien (PEPS) Égalité « Probabilités et Statistiques pour l'Écologie » et « Approximation de Stein : approche par calcul de Malliavin et applications à la gestion des risques financiers ». Ces projets font partie du projet européen INTEGER cofinancé par la Commission Européenne dans le cadre du 7ème PCRD (GA n°266638). Nous remercions le CNRS et tout particulièrement la Mission pour la place des femmes au CNRS.

Nous remercions également les chercheurs et chercheuses qui ont contribué à l'élaboration de ce fascicule.

Camille Coron, Caroline Hillairet et Amandine Véber

Edito

Cette brochure s'adresse aux lycéens, aux étudiants et à toute personne désireuse d'en savoir plus sur la recherche en mathématiques appliquées. Vous y trouverez 7 projets de recherche issus de différents domaines des mathématiques :

Modéliser l'évolution génétique d'une population
Micro-natation, ou comment nager dans du miel
Déterminer le « juste » prix d'un produit financier
Estimer l'impact du changement climatique
Les citoyens peuvent-ils mesurer la biodiversité ?
Comment faire émerger un consensus dans un réseau
L'algèbre tropicale en renfort des appels d'urgence parisiens.

Des informations pratiques sur les métiers liés aux mathématiques et notamment sur celui d'enseignant(e)-chercheur(se) sont données en fin de brochure.

Qu'est-ce que la modélisation mathématique ?

Dans tous les exemples décrits dans ce fascicule, la démarche est la même et correspond à celle du modélisateur en mathématique. Le système « réel » que l'on cherche à comprendre est souvent très complexe et nécessite donc de réfléchir à l'approche la plus pertinente pour le décrire :

(*i*) Tout d'abord, il est important de bien définir le ou les phénomènes qui nous intéressent ainsi que les variables quantifiables qui permettront de les étudier : toxicité d'un antibiotique, nombre d'individus de chaque espèce, valeur d'un actif financier... En particulier, à quelle échelle (temporelle, spatiale, individuelle) choisit-on d'étudier ces phénomènes ?

(*ii*) Une deuxième étape consiste à trouver les bons paramètres à prendre en compte : taux de transfert d'ADN entre bactéries, variabilité de la dispersion spatiale d'un individu, variabilité du cours de l'actif financier...

Ces deux étapes impliquent un compromis entre précision et simplicité. En effet, on souhaite pouvoir décrire le système étudié de la manière la plus exacte qui soit et donc à prendre en compte des variables précises et détaillées et un nombre a priori important de paramètres. En revanche, plus le nombre de variables et de paramètres est grand et plus les équations d'évolution qui les relient sont complexes et difficiles à étudier.

(*iii*) Finalement, il faut également choisir ou développer les outils mathématiques adaptés aux phénomènes considérés : équations différentielles et analyse fonctionnelle, processus stochastiques et théorie des probabilités, estimateurs et statistique...

Une fois ces trois étapes réalisées, on obtient une description simplifiée du système étudié, écrite dans une langue universelle, qui permet d'en dégager les comportements principaux et de répondre de manière qualitative ou quantitative à des questions d'intérêt. C'est ce que l'on appelle un ***modèle mathématique***.

Modéliser l'évolution génétique d'une population

Je m'appelle Amandine Véber, je suis chargée de recherche CNRS au Centre de Mathématiques Appliquées de l'École Polytechnique. Je travaille à comprendre comment la composition génétique d'une population évolue au cours du temps, en particulier lorsque celle-ci a une structure spatiale.

L'évolution de la diversité génétique d'une population est une question qui apparaît dans de nombreux contextes :

- ◇ en médecine, par exemple pour comprendre et mieux anticiper la résistance des bactéries aux antibiotiques ;
- ◇ en agronomie, par exemple pour évaluer les possibilités de contamination de cultures « classiques » ou « biologiques » par des cultures OGM voisines ;
- ◇ en gestion de l'environnement, par exemple pour comprendre l'effet de la construction d'une route ou d'une usine qui vient fragmenter l'habitat d'une espèce donnée.

Lexique

(1) **Gène** : portion de chromosome ayant un sens fonctionnel (codant pour une protéine).

(2) **Allèle** : l'une des versions possibles d'un gène (par exemple, bleu ou brun pour le gène codant pour la couleur des yeux).

(3) **Choisir uniformément au hasard** : dans un ensemble de taille finie, on choisit un élément uniformément au hasard si chaque élément de l'ensemble a la même probabilité d'être choisi.

(4) **Se fixer** : on dit qu'un allèle se fixe dans la population si à partir d'un moment donné tous les individus portent cet allèle.

(5) **Mutation** : erreur de copie de l'ADN qui peut modifier l'allèle d'un individu par rapport à celui de son parent.

Un modèle très simple mais néanmoins très utilisé est le *modèle de Wright-Fisher*. Il suppose que la population est de taille fixe et qu'elle évolue par générations discrètes. La première hypothèse correspond à une population exploitant au mieux les ressources de son environnement, de sorte que celui-ci ne puisse pas contenir plus d'individus qu'il n'y en a déjà. La seconde hypothèse est valable pour des espèces se reproduisant de manière saisonnière (à chaque printemps...), elle est nettement plus discutable pour des espèces telles que l'Homme (qui ont des enfants à n'importe quel moment de l'année).

Supposons que l'on s'intéresse à un gène(1) particulier, ayant plusieurs allèles(2) différents. On part d'une génération 0 dans laquelle chaque individu a un allèle donné. On passe alors de la génération k à la génération $k+1$ de la manière suivante : à chaque individu de la génération $k+1$, on attribue un parent choisi uniformément au hasard(3) parmi la génération précédente dont il adopte l'allèle. Ainsi, si les deux seuls allèles possibles sont A et a et si p_k est la proportion des individus A à la génération k, alors en notant N la taille (constante) de la population on obtient que la probabilité qu'il y ait exactement $j \in \{0, \dots, N\}$ individus d'allèle A à la génération $k+1$ vaut

$$\mathbb{P}(p_{k+1} = j/N) = \binom{N}{j} p_k^j (1-p_k)^{N-j}.$$

En effet, $\binom{N}{j} = \frac{N!}{j!(N-j)!}$ est le nombre de manières de choisir les j individus qui seront d'allèle A parmi les N individus de la génération $k+1$. Ensuite, la probabilité que ces j individus précisément choisissent (indépendamment) un parent d'allèle A et les $N-j$ autres un parent d'allèle a vaut $p_k^j(1-p_k)^{N-j}$.
Grâce à des résultats mathématiques sur ce type d'évolution aléatoire, on peut montrer que forcément (autrement dit, avec probabilité 1) l'un des deux allèles va se fixer(4) dans la population après un temps qui, lui, est aléatoire. Par ailleurs, la probabilité que ce soit l'allèle A qui envahisse la population vaut p_0, la proportion initiale de A.

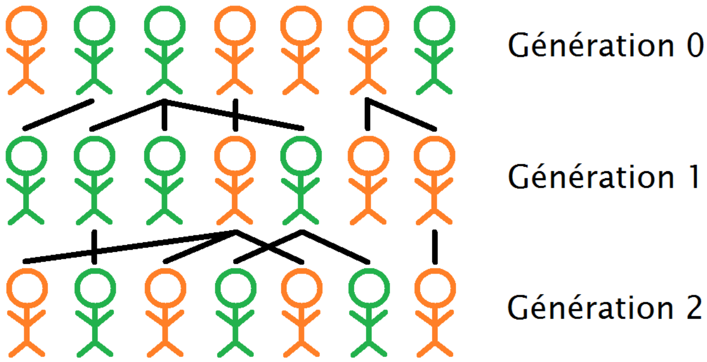


FIGURE 1 – Illustration du modèle de Wright-Fisher. Les traits noirs représentent les liens de parenté entre individus de deux générations successives. Les individus de la génération 0 ont tous un allèle donné (orange ou vert), qui est transmis à leurs descendants lorsqu'ils en ont.

Il existe de nombreuses généralisations de ce modèle, afin de prendre en compte la possibilité de mutations(5) ou le fait que certains allèles peuvent conférer un avantage reproductif (on parle alors de *sélection naturelle*) à leurs porteurs. Une autre direction de généralisation est d'inclure une structure au sein de la population, de sorte que tous les individus ne peuvent pas être parents de n'importe qui dans la génération suivante. L'étude mathématique de ces modèles permet alors de comprendre le comportement en temps long de la proportion de chaque allèle dans la population et donc celui de sa diversité génétique.

Pour plus d'informations et pour me contacter :
`http://www.cmap.polytechnique.fr/~veber/`

Micro-natation, ou comment nager dans du miel

Je m'appelle Laetitia Giraldi, je suis postdoctorante à l'École nationale supérieure de techniques avancées (ENSTA-Paristech). Avec François Alouges (Centre de mathématiques appliquées de l'Ecole Polytechnique), je m'intéresse à la micro-natation, i.e. à la compréhension de la nage des spermatozoïdes, bactéries et autres micro-organismes. Le grand défi technologique associé à ce sujet est la conception d'un micro-robot nageur commandé. De nos jours, il n'existe pas d'appareil microscopique auto-propulsé commandé par l'homme. La conception d'une telle machine au service de la médecine pourrait être un outil révolutionnaire.

Cependant, la natation à cette échelle se heurte à des obstructions qui diminuent singulièrement son efficacité. Du fait de leur petite taille, les micro-organismes qui se déplacent dans l'eau ont la sensation qu'ils évoluent dans un fluide très visqueux. On peut relier leur expérience à la nôtre en essayant de s'imaginer nager dans du miel. Pour un tel régime, les forces inertielles sont négligeables devant les forces visqueuses. La modélisation du déplacement d'un micro-organisme dans un fluide en considérant avec attention les particularités de la situation est donc un enjeu pour la recherche.

Le domaine se développe à partir des années 1950 avec le travail de George Taylor, qui traite de la modélisation du déplacement de spermatozoïdes ayant des flagelles infinis. Plus tard, en 1977, la contribution de E. M. Purcell « Life at Low Reynolds number » révolutionne le sujet. Il explique pourquoi certaines stratégies utilisées à l'échelle humaine pour nager ne permettent pas le déplacement d'organismes microscopiques. Ce résultat mathématique est connu sous le nom du « théorème de la coquille Saint-Jacques ». Il stipule qu'un système microscopique n'ayant qu'un seul degré de liberté et faisant un mouvement périodique ne peut pas se déplacer dans de l'eau. C'est le cas d'une coquille Saint-Jacques, puisqu'elle ne peut qu'ouvrir et fermer sa coquille (voir Figure 2). Ce théorème implique que si les coquilles Saint-Jacques étaient microscopiques, elles ne pourraient pas avancer.

Récemment, les mathématiciens se sont intéressés à ces questions. Ils ont permis, par exemple, l'émergence de connexions entre la théorie du contrôle mathématique et la nage à l'échelle microscopique.

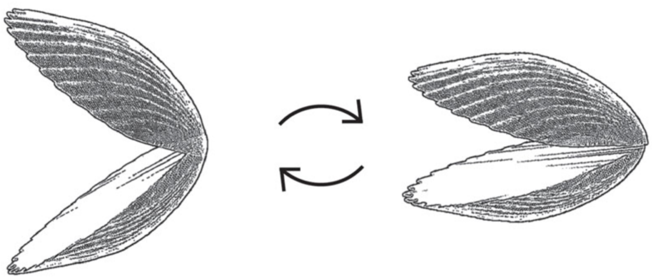


FIGURE 2 – Stratégie de déformation de la coquille Saint-Jacques pour se déplacer.

Pour illustrer ce lien, nous nous sommes intéressés au problème de savoir si l'environnement extérieur peut affecter la mobilité d'un micro-organisme. De nombreuses études biologiques traitent de cette question. Les expériences pratiquées sur des micro-organismes particuliers laissent penser que le bord de l'environnement pourrait avoir des effets attractifs sur le micro-nageur. Récemment, des travaux mathématiques ont contribué significativement à décrire les effets des bords sur la mobilité d'un micro-organisme. En utilisant la théorie du contrôle, on montre que la présence d'une paroi n'affecte pas la capacité de mobilité du nageur : au contraire, elle enrichit ses directions accessibles. Cependant, nous n'arrivons toujours pas à expliquer tous les phénomènes qui sont responsables de la tendance des micro-organismes à s'agglutiner sur le pourtour de leur milieu. Il est probable que la présence d'un mur demande au micro-organisme plus d'efforts pour nager dans certaines directions plutôt que dans d'autres.

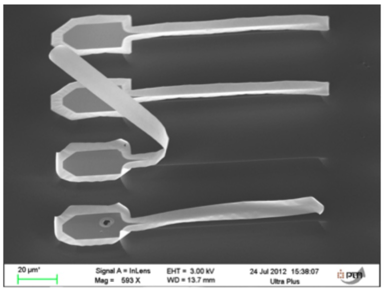


FIGURE 3 – Micro-robots nageurs réagissant sous l'effet d'un champ magnétique extérieur. Ils sont construits au laboratoire *Spintec* du CEA de Grenoble et mesurent quelques dizaines de micromètres.

Théorie du contrôle

La théorie du contrôle est une théorie mathématique permettant de déterminer les lois de guidage, d'action, sur un système donné. Un système de contrôle est un système dynamique sur lequel on peut agir au moyen d'une commande, appelée *fonction de contrôle*. Les systèmes de ce type sont très nombreux : conduire une voiture, piloter un drone, guider un satellite. Dans le cas d'une voiture, l'état est la position de la voiture tandis que les fonctions de contrôle sont l'angle du volant et les pressions exercées sur le frein et sur l'accélérateur. Le problème de contrôlabilité consiste à déterminer l'existence de fonctions de contrôle pour que le système puisse atteindre un état cible désiré. L'objectif peut également être de réaliser cette tâche en minimisant un critère, comme par exemple l'essence consommée dans le cas de la voiture.

Dans l'exemple de la micro-natation, on suppose que le nageur contrôle la déformation de son corps. Il s'agit alors d'étudier les stratégies de nage à utiliser pour que le nageur atteigne une position désirée. On peut également supposer que le corps du micro-nageur réagit sous l'effet d'un contrôle extérieur, comme par exemple lorsqu'il a une charge électrique et qu'il peut se déformer sous l'effet d'un champ magnétique extérieur. La question consiste alors à comprendre comment agir sur l'aimant pour que le nageur se déplace. De tels micro-robots nageurs sont construits au CEA de Grenoble par les équipes de recherche du laboratoire *Spintec* sous la direction de Bernard Dieny (voir Figure 3).

Pour plus d'informations et pour me contacter :
`http://perso.ensta-paristech.fr/~giraldi/`

Déterminer le « juste »prix d'un produit financier

Je m'appelle Caroline Hillairet, je suis maître de conférences au Centre de Mathématiques Appliquées de l'École Polytechnique. Avec Ying Jiao (Institut de Science Financière et d'Assurances, Lyon) et Anthony Réveillac (Institut national des sciences appliquées de Toulouse), nous nous intéressons à la modélisation et à l'estimation des risques financiers. Pour cela, nous utilisons des outils probabilistes.

Le marché du risque financier

En 1987 est créé à Paris le premier super-marché de produits financiers : le MATIF (maintenant Euronext). On y trouve essentiellement des produits d'assurance (produits dérivés) contre les évolutions défavorables des cours des titres financiers. Trois ans après, les salles de marché ne recrutent plus que des ingénieurs (plutôt que des élèves d'écoles de commerce). Pourquoi la finance de marché relève-t-elle plus des mathématiques appliquées que de l'économie ?

> *« Si, à l'égard de plusieurs questions traitées dans cette étude, j'ai comparé les résultats de l'observation à ceux de la théorie, ce n'était pas pour vérifier des formules établies par les méthodes mathématiques, mais pour montrer seulement que le marché, à son insu, obéit à une loi qui le domine : la loi de la probabilité. »*
>
> **L. Bachelier, *Théorie de la Spéculation*** *(1900)*

Les spécialistes de la finance ont en effet recours à des outils mathématiques sophistiqués (*martingales, intégrales stochastiques*) pour la description des phénomènes et la mise au point de méthodes de calcul. Néanmoins, l'utilisation du calcul des probabilités en modélisation financière n'est pas récente et remonte à Bachelier (1900) dans sa thèse sur la « théorie de la spéculation ». Elle s'est largement développée avec les travaux de Black, Scholes et Merton en 1973.

Le mouvement brownien : le même objet en finance et en physique

L'évolution d'un cours financier est une trajectoire continue mais très erratique, caractérisée par une tendance variable et des fluctuations autour de cette tendance. Y a-t-il un modèle pour représenter cela ? Oui, le mouvement brownien !
C'est le botaniste Robert Brown (1773-1858) qui décrit le premier le mouvement erratique de fines particules en suspension dans un fluide. En 1905, Einstein construit un modèle probabiliste pour décrire le mouvement d'une particule qui diffuse ; ses travaux sont à l'origine du grand intérêt porté au mouvement brownien.

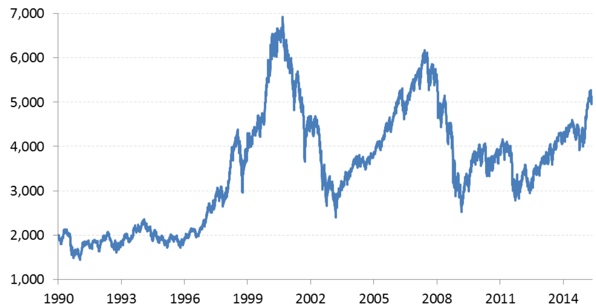


FIGURE 4 – L'indice d'action français (CAC40) entre janvier 1990 et mai 2015

Un problème d'assurance

Dans les années 1970, les entreprises industrielles et commerciales sont soumises à des risques accrus, suite à la déréglementation des marchés et à des taux de change très variables. Pour les aider et plus généralement pour permettre aux compagnies d'assurance et aux banques de couvrir ces nouveaux risques, des marchés organisés sont créés, les autorisant à intervenir massivement pour échanger des produits assurant contre les variations des cours : ce sont les produits dérivés. Notons que l'idée d'un lieu d'échanges où différents acteurs peuvent échanger leur risque en contrepartie d'une prime n'est pas nouvelle. On peut citer par exemple le marché des bulbes de tulipes en Hollande dès le 17$^{\text{ième}}$ siècle, où des grossistes s'engageaient, moyennant une prime, à acheter une certaine quantité de bulbes à une date future donnée et à un prix fixé. La question est alors de trouver quel est le *juste* prix de cette prime, au sens où ce prix doit garantir une forme d'équilibre entre les agents sur le marché. Il n'est pas facile de déterminer mathématiquement la valeur de cette prime qui permette de donner un sens à des termes aussi vagues que « juste » et « équilibre » : les négociants de l'époque ne disposaient pas des outils mathématiques qui ne seront introduits que trois siècles plus tard et suite à un hiver trop rigoureux, les cours s'effondrèrent, forçant de nombreux négociants à mettre la clé sous la porte.

Un autre exemple, plus moderne, est celui d'une compagnie aérienne qui veut se protéger contre une hausse du prix du kérosène dans le futur (indexé sur le cours du baril de pétrole S_T, maturité T=6 mois par exemple). Pour ce faire, elle achète une garantie de niveau K (en général proche de S_0), appelée option d'achat ou *Call* (une option de vente est appelée un *Put*). Si à la date T le prix d'exercice K est supérieur au cours S_T, le détenteur de l'option n'a pas intérêt à exercer son option. Par contre, si $S_T > K$, l'exercice de l'option permet de réaliser un profit égal à $(S_T - K)$. Ainsi, en T, la valeur de l'option d'achat est $(S_T - K)^+ = \max(S_T - K, 0)$.

Modélisation de l'évolution du cours d'un actif

$$\frac{dS_t}{S_t} = \mu_t dt + \sigma dW_t$$

W_t est un bruit brownien, μ=tendance, σ=volatilité

Au moment de la vente de l'option (que l'on peut prendre pour origine des temps), le cours S_T est inconnu et deux questions se posent :

1. Quel est le prix de cette option, i.e., quelle est la prime que doit payer l'acheteur de l'option à l'instant $t = 0$? C'est le problème de *pricing*.
2. Comment le vendeur peut-il réduire ses risques, i.e. comment doit-il investir la prime (qu'il touche à l'instant $t = 0$) afin de produire la richesse $(S_T - K)^+$ à la date T ? C'est le problème de *couverture*.

Black et Scholes ont été les premiers à proposer un modèle conduisant à une formule explicite pour le prix de l'option et à une stratégie de gestion qui permette au vendeur de se couvrir parfaitement, c'est-à-dire d'éliminer le risque. La formule obtenue ne dépend que d'un paramètre non directement observable sur le marché, appelé *volatilité*. De nombreuses extensions des méthodes de Black et Scholes ont été développées par la suite.

Pour conclure, soulignons l'importance des hypothèses de modélisation et de calcul pour la gestion des risques financiers. En effet, cette période de crise a mis à jour des risques souvent sous-estimés (risque de liquidités, risque systémique..) qui en période normale paraissent effectivement marginaux, mais qui en cas de situations extrêmes et de fluctuations inhabituelles des cours financiers deviennent très importants.

Pour plus d'informations et pour me contacter :
`http://www.cmap.polytechnique.fr/~hillaire/`

Estimer l'impact du changement climatique

Je m'appelle Gaël Raoul, je suis chargé de recherche CNRS au Centre de Mathématiques Appliquées de l'École Polytechnique. Je m'intéresse à l'effet du changement climatique sur des populations naturelles.

Les outils de prédiction développés par les météorologues permettent d'anticiper l'évolution du climat au cours du siècle à venir. Ainsi, nous savons qu'en France métropolitaine les températures vont augmenter et les précipitations diminuer. En se basant sur ces prédictions, des biologistes se sont intéressés à l'impact de ces changements climatiques. Leurs premiers résultats ont été alarmants : une étude a notamment prédit la disparition des hêtres (une espèce emblématique des forêts françaises) d'ici 2080. Comment faut-il réagir à ce résultat (évolution des stratégies de gestion des forêts, introduction de nouvelles essences...) ? Quels seraient les effets des politiques de préservation d'espèces que l'on pourrait adopter ?

L'étude mentionnée ci-dessus et de nombreux autres travaux récents se basent sur la notion d'*enveloppe climatique*, qui étudie les corrélations statistiques entre aires de répartition(1) des espèces et données environnementales. En d'autres termes, on cherche à quantifier les liens entre la distribution géographique des espèces et certains facteurs environnementaux tels que la température ou la composition des sols. Cet outil théorique performant repose sur des approximations et des simplifications : le mouvement des individus n'est pas vraiment pris en compte et la diversité génétique de ces populations est négligée. Avec des collègues biologistes et mathématiciens, nous essayons de développer une autre approche, fondée sur des approximations différentes. Nous espérons que ces nouveaux modèles pourront apporter un point de vue intéressant et complémentaire aux approches existantes.

Notre travail est basé sur l'idée suivante : simplifions à l'extrême le climat et l'environnement, pour nous concentrer sur une description explicite du mouvement des individus et de leur diversité génétique. Ce modèle est essentiellement qualitatif et ne permettra donc pas d'étudier l'effet quantitatif du changement climatique sur un territoire précis, mais il pourrait permettre d'étudier les effets de politiques de préservation d'espèces. Par exemple, des résultats préliminaires semblent indiquer que créer un site naturel protégé à l'extrémité méridionale de l'aire de répartition d'une espèce pourrait favoriser l'adaptation de l'ensemble de l'espèce à l'augmentation des températures.

Lexique

(1) **Aire de répartition d'une espèce** : zone délimitant la répartition géographique des individus de l'espèce.

(2) **Quantité conservée** : quantité qui reste constante au cours du temps.

(3) **Déterministe** : non aléatoire.

Les populations biologiques sont diverses et les outils mathématiques à considérer le sont donc également : les espèces asexuées peuvent être modélisées par des équations de réaction-diffusion (qui ont été abondamment utilisées pour décrire des réactions chimiques), tandis que les modèles servant à décrire les populations sexuées sont proches des modèles utilisés pour décrire les fluides (équations de Boltzmann ou de Navier-Stokes). Cependant, une caractéristique des modèles de biologie est l'absence de quantités conservées(2) telles que l'énergie ou la quantité de mouvement, qui jouent un rôle important dans bien des problèmes issus de la physique ou de la chimie. Une autre caractéristique est l'importance des effets aléatoires, car les systèmes biologiques sont généralement de taille beaucoup plus petite que les systèmes de particules ou de réactifs (rendant par exemple leurs rencontres plus aléatoires). Ces spécificités des modèles biologiques par rapport aux modèles physiques déjà largement étudiés impliquent qu'il est nécessaire de développer de nouvelles approches mathématiques, mêlant processus aléatoires et équations déterministes(3).

Un modèle de dynamique des populations

Un modèle simple de dynamique des populations consiste à décrire une espèce par sa densité $n = n(t,x)$, où les individus sont caractérisés par leur latitude $x \in \mathbb{R}$ au temps $t \geq 0$. On suppose que le taux de reproduction de la population dépend du climat, via une fonction $r : \mathbb{R} \mapsto \mathbb{R}_+$ et que ce climat se déplace vers le nord à une vitesse constante $c > 0$. De cette manière, on peut par exemple modéliser le fait que l'on retrouve les températures les plus fraîches de plus en plus au nord à cause du réchauffement climatique. On obtient alors l'équation suivante :

$$\partial_t n(t,x) - \Delta_x n(t,x) = (r(x-ct) - n(t,x))n(t,x),$$

où $\partial_t n(t,x)$ est la dérivée de $(t,x) \mapsto n(t,x)$ par rapport à la première variable et $\Delta_x n(t,x)$ est la dérivée seconde de $(t,x) \mapsto n(t,x)$ par rapport à la deuxième variable. Cette équation nous dit que la variation temporelle de la densité d'individus à la latitude x est due au déplacement spatial des individus au voisinage de x (modélisé par le terme de dérivée seconde en espace) et à un terme de reproduction. Ce dernier modélise le fait que la population locale de densité $n(t,x)$ se reproduit à un taux dicté par son climat au temps t (soit $r(x-ct)$) moins l'impact de la compétition pour les ressources entre les individus, qui diminue la capacité des nouveaux individus à réellement s'établir dans la population.

A partir de ce modèle, on peut par exemple montrer qu'il existe une vitesse critique c^* telle que la population survit si $c < c^*$ et s'éteint si $c \geq c^*$. En d'autres termes, si le climat change suffisamment lentement la population a le temps de s'adapter, tandis que s'il change trop vite elle ne peut pas survivre.

Pour plus d'informations et pour me contacter :
`http://www.cmap.polytechnique.fr/~raoul/`

Les citoyens peuvent-ils mesurer la biodiversité ?

Je m'appelle Camille Coron, je suis maître de conférences à l'Université Paris-Sud (Orsay). J'effectue mon enseignement à l'IUT de Sceaux et je fais partie de l'équipe Probabilités et Statistiques du Laboratoire de Mathématiques d'Orsay. Avec Clément Calenge (Office National de la Chasse et de la Faune Sauvage), Christophe Giraud (Université Paris-Sud) et Romain Julliard (Museum National d'Histoire Naturelle), je m'intéresse au suivi de la biodiversité à l'aide de données issues de programmes de sciences participatives (voir encadré). Certains de ces programmes imposent des protocoles d'observations très stricts mais fournissent de ce fait relativement peu de données (on parle de données *standardisées*), tandis que d'autres programmes laissent l'observateur très libre, ce qui apporte en général un très grand nombre de données mais pour lesquelles nous manquons d'informations (on parle de données *opportunistes* ; en particulier on ne connaît généralement pas le temps total d'observation). Sur la carte ci-dessous, nous représentons par exemple les positions d'observations pour deux tels jeux de données.

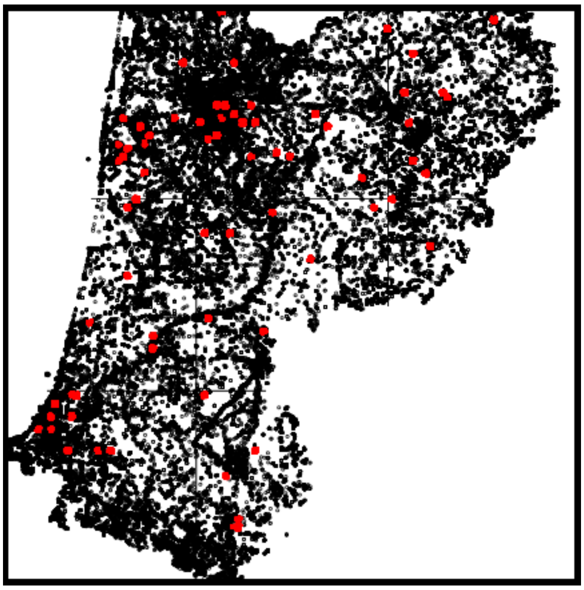


FIGURE 5 – Sur cette carte de l'Aquitaine, nous montrons les emplacements d'observations d'oiseaux communs pour des données standardisées (en rouge, collectées par Vigie Nature), et opportunistes (en noir, collectées par la Ligue pour la Protection des Oiseaux).

Ces nouveaux types de données génèrent beaucoup de nouvelles questions auxquelles les probabilités et les statistiques permettent de répondre :

- ◇ Peut-on suivre la biodiversité à l'aide de données participatives ?
- ◇ Comment gérer les erreurs d'observation ou d'identification faites par les observateurs ?
- ◇ Quel protocole les scientifiques devraient-ils imposer aux observateurs ?
- ◇ Les données opportunistes peuvent-elles aider à l'estimation de la biodiversité ?
- ◇ ...

Notre démarche consiste à combiner deux jeux de données, l'un standardisé et l'autre opportuniste, pour comparer l'abondance d'une espèce à différents moments ou différents endroits : le jeu de données standardisées sert ici à calibrer le jeu de données opportunistes qui est plus riche mais moins précis.

La première étape de notre travail consiste à modéliser mathématiquement les données d'observations auxquelles nous avons accès.

L'espace est divisé en J zones, indexées par $j \in [\![1, J]\!]$. On considère I espèces, indexées par $i \in [\![1, I]\!]$, et l'on note N_{ij} le nombre d'individus de l'espèce i sur la zone j. Notre but est d'estimer N_{ij}/N_{i1} pour tous i et j, donc de comparer les abondances de chaque espèce, sur différentes zones. Nous disposons pour cela de deux jeux de données, indexés par $k \in \{0, 1\}$: $\{k = 0\}$ correspond au jeu de données standardisées (jeu de référence), et $\{k = 1\}$ correspond au jeu de données opportunistes.

Par ailleurs, nous souhaitons prendre en compte le fait que d'une part, les espèces ont des probabilités de détection différentes les unes des autres et que d'autre part, certaines zones sont plus visitées que d'autres par les observateurs. On note P_{ik} la probabilité de détection de l'espèce i pour le jeu de données k et E_{jk} l'intensité d'observation dans le site j pour le jeu de données k. Remarquons que l'on peut supposer que E_{j0}/E_{10} est connu pour tout j, car l'on connaît le temps passé par chaque observateur à chaque endroit (c'est justement l'avantage des données standardisées).

Sciences participatives

Les programmes de sciences participatives (ou sciences citoyennes) sont mis en place par des scientifiques et mobilisent la participation de citoyens volontaires, amateurs ou éclairés. Le but de ces programmes, très utilisés en sciences naturelles notamment, est de rassembler les observations faites par ces citoyens, pour constituer des jeux de données que les chercheurs peuvent ensuite exploiter. Dans notre travail, nous utilisons des jeux de données de Vigie Nature et de la Ligue pour la Protection des Oiseaux (LPO). Pour le premier jeu de données (Vigie Nature), l'observateur doit se rendre à un lieu donné, et noter l'espèce de chaque oiseau qu'il aura pu voir, durant un laps de temps qui est fixé par le protocole. Pour le deuxième jeu de données (LPO), l'observateur peut se rendre où il veut, passer le temps qu'il veut à observer et rapporter les observations qu'il veut.

Nous supposons alors que le nombre X_{ijk} d'observations de l'espèce i dans le site j pour le jeu de données k suit une loi de Poisson (voir encadré) :

$$X_{ijk} \sim \mathcal{P}oisson(N_{ij} P_{ik} E_{jk}).$$

Dans ce modèle, l'estimation des paramètres N_{ij}/N_{i1} peut alors être ramenée à un problème classique d'estimation statistique, que le langage de programmation R nous permet de mettre en oeuvre.

Nous obtenons alors une estimation des abondances relatives N_{ij}/N_{i1} et nous prouvons que la combinaison des deux jeux de données, lorsque les E_{j1} sont assez grands, mène à une estimation qui est plus précise que celle obtenue en utilisant seulement les données standardisées.

Loi de Poisson

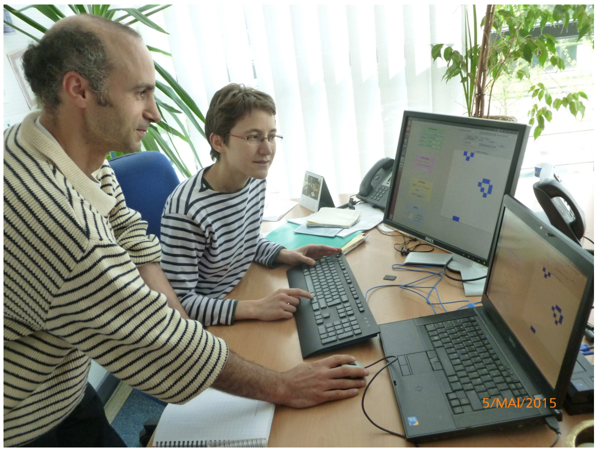
On dit qu'une variable aléatoire X suit une loi de Poisson de paramètre λ si pour tout entier positif ou nul k,

$$\mathbb{P}(X = k) = \exp(-\lambda)\frac{\lambda^k}{k!}$$

Pour plus d'informations et pour me contacter :
`http://www.math.u-psud.fr/~ccoron/`

Comment faire émerger un consensus dans un réseau

Je m'appelle Irène Marcovici et je suis maître de conférences en mathématiques appliquées à l'Université de Lorraine. J'effectue mon enseignement à la Faculté des sciences et technologies de Nancy et je fais partie de l'équipe de probabilités et statistiques de l'Institut Élie Cartan de Lorraine, qui regroupe les chercheurs en mathématiques de Metz et Nancy.



Mes travaux sont aussi étroitement liés à l'informatique. J'ai d'ailleurs effectué mon doctorat au sein du LIAFA, laboratoire d'informatique théorique de l'Université Paris Diderot et actuellement, je poursuis un travail en collaboration avec Nazim Fatès, chercheur en informatique au centre Inria Nancy - Grand Est. Mes recherches sont de nature fondamentale, mais reposent sur des motivations concrètes comme comprendre dans quelle mesure un programme peut être robuste à des erreurs aléatoires, ou encore développer des procédures de régulation pour des réseaux informatiques.

Considérons un réseau (informatique, biologique...) constitué d'un certain nombre d'entités, appelées des *cellules*, qui peuvent prendre plusieurs états possibles (par exemple saines ou infectées) et interagissent entre elles de manière locale. On peut modéliser l'évolution de ce réseau par un *automate cellulaire probabiliste* : on suppose que les différentes cellules sont régies par exactement la même loi et qu'à chaque pas de temps, le nouvel état d'une cellule dépend seulement des états qu'avaient ses voisines au temps précédent.

Je m'intéresse aux comportements collectifs qui peuvent *émerger* quand des entités se comportent ainsi selon une même règle locale simple, déterministe ou probabiliste (autrement dit, avec ou sans aléa). Il s'agit alors :

- d'étudier des automates cellulaires probabilistes spécifiques utilisés pour décrire certains phénomènes. J'ai par exemple travaillé sur un modèle de formation d'essaims d'oiseaux, introduit pour comprendre comment à partir d'un état désordonné une direction commune de vol est choisie.
- de concevoir de nouvelles dynamiques ayant un comportement particulier. En effet, les systèmes informatiques actuels sont constitués d'un nombre d'entités toujours plus grand et sont souvent distribués : ils ne possèdent pas d'autorité centrale permettant de contrôler et de réguler l'ensemble du fonctionnement. Le développement de nouveaux protocoles permettant d'effectuer certaines tâches est donc un enjeu d'actualité.

Un problème sur lequel je travaille avec Nazim Fatès est le *problème de la majorité*, aussi connu sous le nom de *problème de la classification de la densité*. Considérons par exemple un réseau ayant la forme d'une grille carrée, dans lequel chaque cellule a deux états possibles, représentés par un carré bleu ou un carré blanc. Le problème consiste à concevoir un automate cellulaire capable de décider si une configuration initiale contient plus de carrés bleus que de carrés blancs, en convergeant vers la configuration contenant uniquement l'élément majoritaire, comme représenté en Figure 6. L'esprit général du problème est celui de l'émergence d'un consensus dans un système distribué : il s'agit de déterminer une information globale en collectant seulement une information locale à chaque étape. La difficulté est double : premièrement, il n'est pas possible de centraliser l'information (les cellules sont indistinguables) ; deuxièmement, il n'est

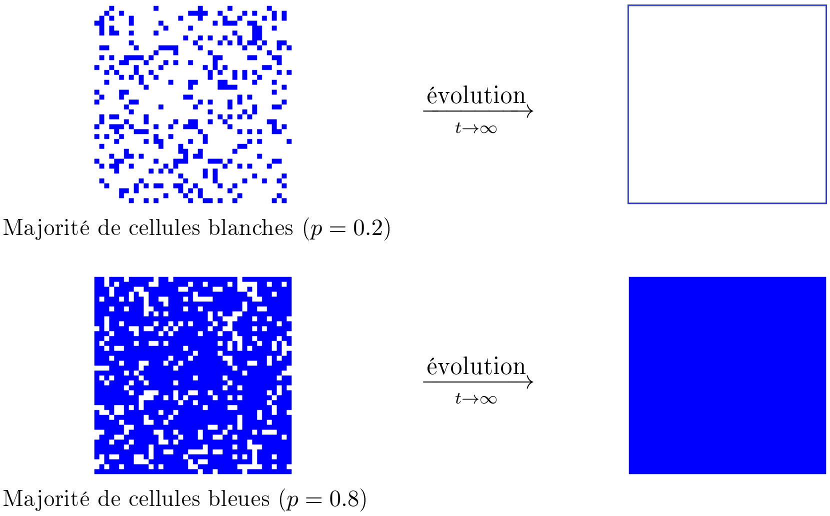


FIGURE 6 – Illustration du problème de la majorité.

pas possible d'utiliser des techniques de comptage classique (le contenu des cellules contient seulement une information binaire).
Pour les grilles finies, nous avons généralisé un résultat prouvant qu'il n'existe pas d'automate cellulaire capable de résoudre *parfaitement* le problème de la majorité, c'est-à-dire qui aie la propriété de converger vers la configuration toute bleue à partir de toutes les configurations initiales comportant une majorité de cellules bleues, ou vers la configuration toute blanche à partir de toutes les configurations qui comportent une majorité de cellules blanches.
Nous cherchons maintenant à construire des automates cellulaires qui permettent de résoudre le problème avec une probabilité d'erreur arbitrairement petite.
Sur une grille infinie, nous avons étendu le problème de la majorité en cherchant un automate cellulaire qui converge vers la configuration toute bleue si initialement chaque cellule est coloriée indépendamment en bleu avec probabilité $p > 1/2$ et vers la configuration toute blanche pour $p < 1/2$. Nous avons alors démontré que l'*automate cellulaire de Toom* avait le comportement souhaité. La règle de cet automate cellulaire est la suivante : à chaque instant t, chaque cellule regarde sa propre couleur, la couleur de sa voisine Nord et la couleur de sa voisine Est et choisit comme nouvelle couleur au temps $t+1$ la couleur qui était majoritaire parmi ces trois couleurs. Notre démonstration utilise des résultats de *percolation* (voir encadré). Etudier le même problème pour un ensemble de cellules disposées sur une ligne plutôt que sur la grille est toujours un défi d'actualité...

Percolation

Considérons une grille infinie où chaque carré est colorié en bleu indépendamment avec probabilité p. La théorie de la percolation s'intéresse aux propriétés des *composantes connexes* de cellules bleues : ce sont les amas de cellules bleues tels qu'on peut passer d'une cellule à une autre par un chemin qui n'emprunte que des cellules bleues. Si on considère des chemins tels qu'à partir d'une cellule donnée, on a le droit de se déplacer vers l'une des quatre voisines (Nord, Sud, Est, Ouest), mais aussi en diagonale vers la cellule Nord-Ouest ou la cellule Sud-Est, on peut montrer que si $p < 1/2$ les amas correspondants de cellules bleues sont tous finis, alors que pour $p > 1/2$ il existe un amas infini (avec probabilité 1). Nous avons utilisé ce résultat pour démontrer que l'automate cellulaire de Toom fournit une solution au problème de la majorité.

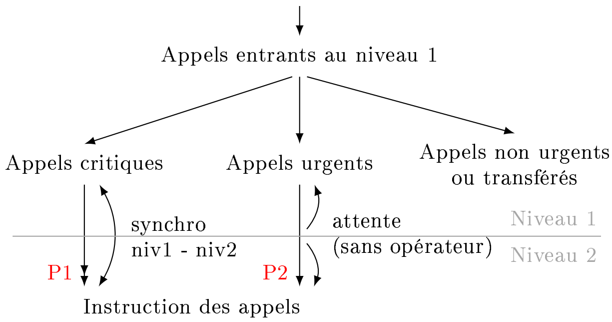
Pour plus d'informations et pour me contacter :
`http://iecl.univ-lorraine.fr/~Irene.Marcovici/`

L'algèbre tropicale en renfort des appels d'urgence parisiens

Je m'appelle Vianney Boeuf. Je prépare une thèse en mathématiques, en collaboration avec les pompiers de Paris, en tant qu'ingénieur de l'État (ingénieur des Ponts, des Eaux et des Forêts). Je travaille dans une équipe de recherche commune au Centre de Mathématiques Appliquées de l'École Polytechnique et à l'INRIA. Le but de ma thèse est de proposer aux pompiers des méthodes d'optimisation de leurs procédures de traitement des urgences. Avec Xavier Allamigeon et Stéphane Gaubert, je travaille sur la problématique de la réception des appels d'urgence.

La préfecture de police prépare une réforme qui vise à unifier le traitement des appels d'urgence des pompiers et de la police à Paris. Dans l'organisation projetée, un premier niveau d'opérateurs téléphoniques traite les appels parasites (il y a énormément d'appels faits par erreur au numéro d'urgence 112), gère les appels non urgents (si votre véhicule est à la fourrière, ce n'est pas une urgence !) et transmet les appels urgents aux opérateurs de niveau 2. Parmi les appels urgents, les policiers et les pompiers veulent identifier les appels critiques : personne en détresse vitale, victime d'un braquage ou incendie menaçant de se propager. L'appel est alors transmis en priorité. L'opérateur de niveau 1 maintient la communication jusqu'à ce que l'opérateur de niveau 2 ait pris l'appel (synchronisation).

Les travaux de recherche se justifient d'abord par l'intérêt concret : on cherche des résultats qui confirment ou infirment la conception du centre d'appels prévue par les responsables et qui les aident à fixer le bon nombre d'opérateurs à chaque niveau selon la densité d'appels. C'est aussi scientifiquement intéressant car complexe, à cause de la priorisation des appels critiques et de la synchronisation entre les opérateurs de niveaux 1 et 2.



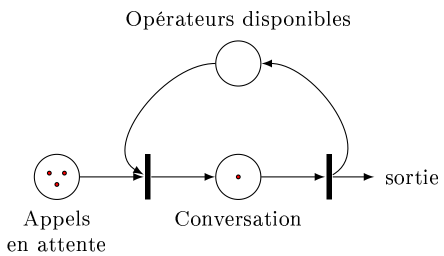


FIGURE 7 – Une modélisation simple de standard téléphonique par un réseau de Pétri. Les cercles sont des places et les rectangles noirs des transitions.

Nous pouvons modéliser ce schéma d'appels par un *réseau de Pétri*, tel que celui représenté sur la Figure 7. Dans ce modèle, des *jetons* (ici, les opérateurs et les appels) circulent dans des *places* où ils restent au minimum un certain temps appelé *temps de séjour*. Lors de *transitions*, ils sont ensuite consommés et/ou de nouveaux jetons sont produits. Une transition donnée possède un certain nombre de places entrantes (les jetons nécessaires pour que la transition ait lieu, par exemple les appels en attente et les opérateurs disponibles pour la transition la plus à gauche sur la Figure 7) et de places sortantes (les conversations, dans notre exemple).

Si l'on impose que chaque transition se produit dès qu'elle peut, on peut écrire des équations entre des fonctions qui sont les *compteurs* de jetons : on note $x_p(t)$ le nombre de jetons passés par la place p jusqu'au temps t et $z_q(t)$ le nombre d'occurrences de la transition q jusqu'au temps t. Ainsi, si une place p est sortante

Algèbre tropicale

Les mathématiciens qui étudient l'algèbre *tropicale*, ou *max-plus*, se placent dans un monde parallèle dans lequel le "plus" est un "fois" et le "max" est un "plus". Plus précisément, ils exploitent les similarités entre les opérateurs $(+, \times)$ et $(\max, +)$ pour adapter des théorèmes classiques en des théorèmes max-plus (et en établir de nouveaux !). Par exemple, $a \times (b+c) = (a \times b) + (a \times c)$: c'est l'associativité, propriété que l'on retrouve en algèbre max-plus : $a + \max(b, c) = \max(a + b, a + c)$.
En algèbre max-plus, on écrit donc "$2 + 3 = 3$" et "$2 \times 3 = 5$". Ainsi, le polynôme $x^2 + 3y$ en les deux variables x et y correspond à l'expression $\max(2x, 3 + y)$ en algèbre classique.
L'algèbre tropicale a été introduite historiquement pour modéliser les phénomènes de synchronisation (par exemple de chaînes de montage de différentes pièces en usine). Aujourd'hui, elle intéresse aussi les mathématiciens purs.

pour deux transitions q_1, q_2, on a $x_p(t) = z_{q_1}(t) + z_{q_2}(t)$. Si q est une transition qui a deux places entrantes p_1 et p_2, alors le nombre de fois où elle s'est produite avant l'instant t, $z_q(t)$, n'est autre que le minimum de $x_{p_1}(t - t_1)$ et $x_{p_2}(t - t_2)$, où t_1 et t_2 désignent les temps de séjour dans les deux places.

Le circuit des appels téléphoniques est donc modélisé par un réseau de Pétri, dans lequel les appels sont pris au plus tôt. Si l'on omet les règles de priorité, les équations vérifiées par les fonctions compteurs s'écrivent à l'aide du minimum et de l'addition, comme ci-dessus. Pour analyser la performance de ce système, on se place dans le cadre de l'algèbre *tropicale* ou *max-plus* (ou plutôt min-plus ici). On parvient à un système monotone : les dates de fin de traitement d'un appel dépendent de manière croissante des dates d'arrivée des appels. En exploitant le caractère affine par morceaux des équations, on peut calculer le débit d'appels auquel le centre peut répondre.

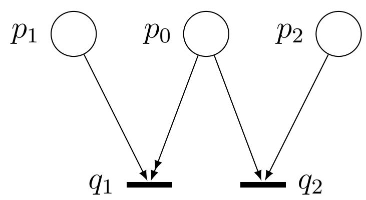


FIGURE 8 – Une priorité, représentée par une double flèche.

Lorsqu'on rajoute des règles de priorité, on obtient des équations différentes. Comme illustré dans la Figure 8, si la place p_0 envoie ses jetons en priorité vers la transition q_1, les équations deviennent :

$$z_{q_1}(t) = \min\big(x_{p_1}(t - t_1)\, ,\, x_{p_0}(t - t_0) - \lim_{s \to t^-} z_{q_2}(s)\big)$$
$$z_{q_2}(t) = \min\big(x_{p_2}(t - t_2)\, ,\, x_{p_0}(t - t_0) - z_{q_1}(t)\big).$$

A cause de l'opération $-$, qui correspond à la « division tropicale », on perd la propriété de monotonie (un appel critique peut doubler un appel urgent). Ceci rend plus difficile le calcul du débit.

Avec mes encadrants de thèse, nous avons mis en évidence trois phases du système, selon le nombre d'opérateurs de niveau 2 : soit tous les appels passent, soit une partie des appels urgents sont perdus, soit enfin même des appels critiques sont perdus. Cela rejoint l'intuition. En outre, il est possible de calculer les frontières entre ces différentes phases en fonction des paramètres du système (temps de séjour) et donc le bon nombre d'opérateurs de niveau 2 requis pour assurer la transmission des appels entre les deux niveaux. Cela pourra aider les pompiers et les policiers de Paris à ne perdre aucun appel d'urgence.

Pour plus d'informations : voir l'article de recherche sur `http://tinyurl.com/ABG2015`

Pour me contacter : `http://www.cmap.polytechnique.fr/~boeuf/`

Les métiers des mathématiques

Les métiers liés aux mathématiques sont nombreux et accessibles à différents niveaux d'études :

Bac+3 (assistant ou technicien supérieur, modélisateur...)

Bac+5 (ingénieur, enseignant, consultant...)

Bac+8 (enseignant-chercheur, chercheur en entreprise...).

Les sites (1) et (2) et les brochures ONISEP (3) et (4) ci-dessous donnent un large panel de débouchés possibles. Le rapport récent (5) décrit l'impact socio-économique des mathématiques en France et montre le rôle important qu'elles tiennent dans le développement industriel et l'innovation. Par ailleurs, le livre (6) de médiation scientifique illustre la variété des problèmes scientifiques dans lesquels la recherche mathématique actuelle joue un rôle important.

Le métier d'enseignant(e)-chercheur(se)

Pour devenir enseignant-chercheur il faut préparer et soutenir une thèse (Bac+8). Ce travail de thèse est l'apprentissage du métier de chercheur et dure généralement 3 ans. Il est entamé après s'être spécialisé dans un domaine particulier des mathématiques grâce à l'obtention d'un diplôme de Master (Bac+5). Les postes de maître de conférences ou de chargé de recherche sont ensuite obtenus par concours.

L'enseignant-chercheur travaille dans un laboratoire et enseigne principalement en université ou en école d'ingénieurs. En dehors des enseignements, il mène ses recherches (seul ou en équipe, parfois avec des collaborateurs internationaux), publie ses résultats dans des articles de revues spécialisées, les expose lors de conférences nationales et internationales ; il organise également des séminaires et des rencontres scientifiques. Ces rencontres permettent ainsi d'échanger avec les spécialistes de son domaine et de nourrir de nouvelles idées pour ses recherches futures. Ces nouvelles idées peuvent également provenir de collaborations avec des industriels, qui pourront ensuite utiliser les outils développés. Plus de renseignements sont disponibles sur la fiche ONISEP (7).

Références

(1) `www.cmap.polytechnique.fr/~giraud/MetiersMaths.html`

(2) `http://www.onisep.fr/Toute-l-actualite-nationale/Decouvrir-les-metiers/Actus-2014/Decembre-2014/Quels-metiers-pour-les-matheux-et-les-matheuses`

(3) Zoom Métiers « Les métiers de la statistique » de l'ONISEP et de la SFDS, 2011. Disponible à l'adresse `http://www.sfds.asso.fr/images/zoom-statistique-2011.pdf`

(4) Zoom Métiers « Les métiers des mathématiques et de l'informatique » de l'ONISEP, Mars 2015. Disponible à l'adresse `http://metiers-mathsinfo.fr/`

(5) *Etude de l'impact socio-économique des Mathématiques en France* (Mai 2015). Disponible à l'adresse `http://www.agence-maths-entreprises.fr/a/eisem`.

(6) *Brèves de maths*, Martin Andler, Liliane Bel, Sylvie Benzoni-Gavage, Thierry Goudon, Cyril Imbert, Antoine Rousseau. Ed. Nouveau Monde, 2014.

(7) `www.onisep.fr/Ressources/Univers-Metier/Metiers/enseignant-e-chercheur-euse`

Vous avez d'autres questions sur ce métier ? Vous pouvez nous contacter à l'adresse suivante :
`camille.coron@math.u-psud.fr`

Crédits illustrations :

Page de garde. En haut : cartographie de la zone moyenne du cerveau activée lors d'un calcul mental (réalisée par Hao Xu).
Milieu : à gauche, surface de volatilité ; à droite, esquisse d'arbre phylogénétique de Darwin tirée de *First Notebook on Transmutation of Species* (1837) (source : wikipedia.org).

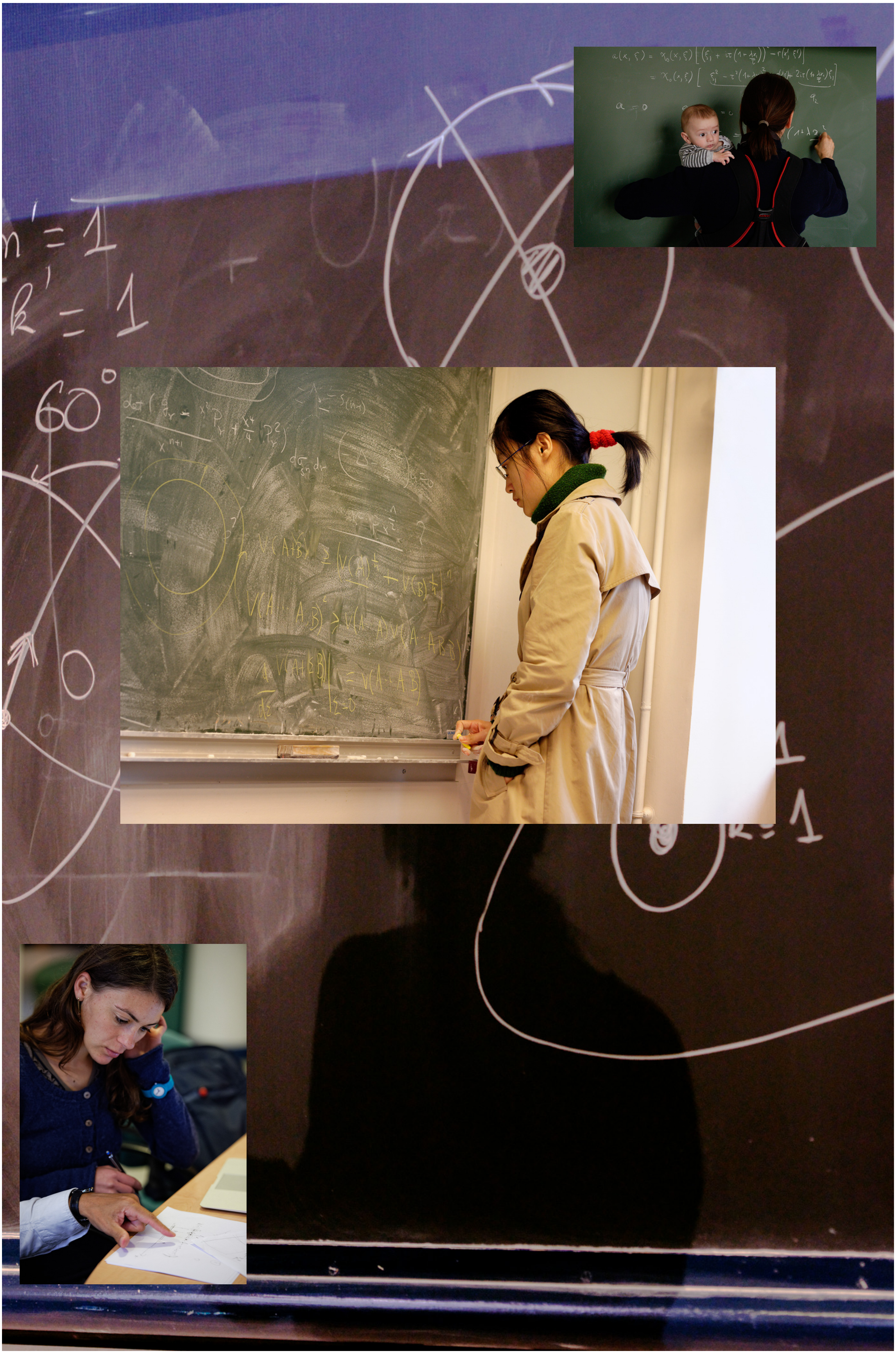
Quatrième de couverture. Photographies de chercheur(se)s tirées du livre *La maison des mathématiques* de C. Villani, J.-P. Uzan et V. Moncorgé (2014). Nous remercions particulièrement l'Institut Henri Poincaré et Vincent Moncorgé de nous avoir accordé l'autorisation d'utiliser ces portraits.

ihp Institut Henri Poincaré

VINCENT MONCORGÉ photographe

cnrs dépasser les frontières

INTEGER
INstitutional Transformation for Effecting Gender Equality in Research


60°