

Modélisation mathématique et assimilation de données pour les sciences de l'environnement

Maëlle Nodet(*) & Antoine Rousseau(**)

Introduction

Les études actuelles en sciences de l'environnement sont réalisées autour d'une multitude de situations environnementales concrètes et locales, auxquelles l'Homme doit faire face par nécessité. Parmi les grands problèmes actuels, on peut citer le changement climatique bien sûr, mais aussi les risques environnementaux, qu'ils soient météorologiques/océanographiques (tempêtes, pollution), hydrologiques (inondations, état des ressources en eau, pollution), géophysiques (glissements de terrain, avalanches), ... La multiplicité des causes, des échelles (spatiales et temporelles), des conséquences (sociales, économiques, biologiques, zoologiques, etc.) rendent difficiles d'appréhender, comprendre, modéliser et tenter de résoudre les problèmes dans leur globalité.

Les géophysiciens spécialisés dans ces disciplines s'intéressent naturellement à ces problèmes, dont ils cherchent à mieux comprendre les processus d'une part et à prévenir et limiter les conséquences d'autre part. Les mathématiciens appliqués jouent ici un rôle important, car ils travaillent de concert avec les physiciens pour leur apporter des outils adaptés. Deux branches principales de recherche en mathématiques pour l'environnement sont la modélisation et l'assimilation de données.

La modélisation consiste à décrire en termes mathématiques et numériques le processus physique étudié, afin de pouvoir faire des simulations et l'étudier en profondeur. C'est la première étape du travail, qui nécessite une synergie entre physiciens et mathématiciens. L'assimilation de données consiste en une utilisation optimale de toutes les informations disponibles sur le processus (équations issues de la modélisation, mesures et observations, entre autres) afin par exemple d'améliorer les modèles, ou bien de pouvoir faire des prévisions dans le futur, ou encore de faire des choix entre diverses politiques environnementales.

Cet article présente brièvement ces deux aspects de la recherche en mathématiques appliquées pour l'environnement.

Modélisation

La modélisation mathématique est l'étape initiale de la simulation d'un processus physique, chimique, biologique, etc. Elle consiste à traduire en équations la

(*) Université Joseph Fourier Grenoble 1 & INRIA MOISE & Laboratoire Jean Kuntzmann.

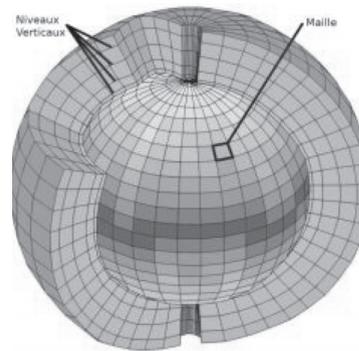
(**) INRIA MOISE & Laboratoire Jean Kuntzmann, Grenoble. <http://team.inria.fr/moise>

description phénoménologique du processus considéré (voir encadré ci-dessous), puis à en faire l'analyse mathématique (existence, unicité, régularité des éventuelles solutions).

La mise en équations repose par exemple (c'est le cas en géophysique) sur l'écriture d'un bilan des forces à partir de grandes lois de la physique. En fonction de la précision de la description du phénomène étudié, le modèle mathématique qui en découle peut être plus ou moins complexe. Les modèles les plus simples font intervenir des systèmes linéaires d'équations différentielles, mais la recherche actuelle se situe plutôt autour d'équations aux dérivées partielles non linéaires (voir par exemple [3]). Il y a donc un compromis à trouver entre des modèles extrêmement détaillés (mais qui sont mathématiquement trop complexes pour être analysés) et des modèles mathématiquement simples à étudier (mais qui sont physiquement irréalistes). Les chercheurs(euses) en mathématiques appliquées travaillent donc à améliorer la « compréhension mathématique » de processus de plus en plus complexes, que l'on peut même coupler entre eux (comme dans le cas du système océan-atmosphère).

À titre d'exemple, souvenons-nous que les équations de Navier-Stokes qui décrivent l'écoulement de l'eau (ou d'un fluide en général) dans l'espace à 3 dimensions n'ont toujours pas livré tous leurs secrets mathématiques⁽¹⁾. Et pourtant, elle coule !

Les solutions des équations que l'on considère, en admettant qu'elles ont toutes les bonnes propriétés (existence, unicité, régularité), sont très rarement explicites : il convient donc de les simuler numériquement pour pouvoir étudier leur comportement. C'est l'objectif des modèles numériques : ils sont obtenus à partir des équations continues en effectuant une discrétisation du domaine de calcul (voir figure ci-contre.).



Un exemple de domaine de calcul discrétisé (surface terrestre et couche atmosphérique)

Une fois le domaine de calcul « coupé en morceaux », les opérateurs différentiels qui apparaissent dans les équations sont approchés par des opérateurs matriciels (grâce notamment à la méthode d'Euler, voir encadré ci-dessous) qui peuvent ensuite être traités par ordinateur. Il y a là aussi toute une communauté de chercheurs qui travaillent à améliorer la précision des modèles numériques sans en augmenter le coût de calcul. En effet, étant données la complexité des phénomènes modélisés et la précision de calcul recherchée pour ces modèles, on arrive facilement à des matrices de taille 10^7 ou plus. Dès lors, on comprend tout à fait la nécessité d'être astucieux dans le développement de méthodes numériques, même sur des super-calculateurs !

(1) Voir la description du prix (1M\$) sur le site de l'institut Clay : http://www.claymath.org/millennium/Navier-Stokes_Equations/

Piste d'utilisation en classe : modélisation de l'équation de transport

• **Description**

« Une information notée $u(x,t)$ se déplace à vitesse constante $c > 0$, où x et t représentent les variables d'espace et de temps. »

• **L'équation**

$$\frac{\partial u}{\partial t}(x,t) + c \frac{\partial u}{\partial x}(x,t) = 0. \quad (1)$$

• **En classe** (nécessite la connaissance de la méthode d'Euler)

– **Modélisation**

Traduisons en termes mathématiques la description ci-dessus : pour toute position x et tout instant t , considérons un laps de temps Δt durant lequel l'information se déplace. Puisque l'information se déplace à vitesse constante c , alors à l'instant $t + \Delta t$ et au point x , l'information est la même qu'à l'instant t et à la position $x - c\Delta t$, d'où :

$$u(x, t + \Delta t) = u(x - c\Delta t, t).$$

Retranchons maintenant $u(x,t)$ de part et d'autre de l'égalité ci-dessus, et divisons par Δt . On obtient après calcul :

$$\frac{u(x, t + \Delta t) - u(x, t)}{\Delta t} + c \frac{u(x, t) - u(x - c\Delta t, t)}{c\Delta t} = 0.$$

Considérons la première de ces deux fractions. Pour tout x fixé (quitte à considérer la fonction d'une seule variable $f_x(t) = u(x,t)$), faisons tendre Δt vers 0. Par définition de la dérivée (ici par rapport au temps), le taux d'accroissement

$$\frac{u(x, t + \Delta t) - u(x, t)}{\Delta t} \text{ tend effectivement vers } \frac{\partial u}{\partial t}(x, t)$$

On obtient finalement l'équation (1) en agissant de même sur la seconde fraction, dans laquelle cette fois-ci t est fixé et où l'on peut définir $\Delta x = c\Delta t$.

– **Méthode d'Euler**

Afin de résoudre numériquement l'équation (1), on se donne un pas de temps Δt et un pas d'espace Δx , et l'on définit pour tout $0 \leq n \leq N$ et tout $0 \leq j \leq J$ le réel u_j^n par l'approximation de la fonction $u(x,t)$ au temps $t = n\Delta t$ et au point $x = j\Delta x$ (voir [1] et [2] pour les détails de ces définitions).

Si l'on se dote d'une donnée initiale $(u_j^0)_{0 \leq j \leq J}$ ainsi que d'une condition aux limites périodique $u_0^n = u_J^n$ (comme pour simuler une hola dans un stade par

exemple), il est facile de calculer les itérées successives u_j^n grâce à la formule suivante (qui repose sur la méthode d'Euler explicite appliquée à la fois en espace et en temps) :

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} + c \frac{u_j^n - u_{j-1}^n}{\Delta x} = 0, \quad \forall 0 \leq n \leq N-1, \forall 1 \leq j \leq J. \quad (2)$$

– **Algorithme** (calcul des u_j^n par récurrence)

La donnée initiale $(u_j^0)_{0 \leq j \leq J}$ étant connue, on calcule les u_j^1 pour $j \geq 1$ grâce à la formule (2), et l'on n'oublie pas la condition aux limites qui nous donne $u_0^n = u_J^n$

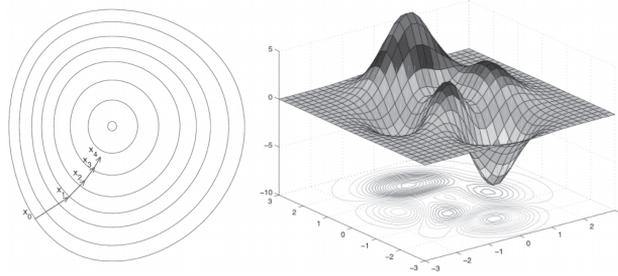
Ainsi, tous les $(u_j^1)_{0 \leq j \leq J}$ sont connus et l'on peut procéder au calcul des $(u_j^2)_{0 \leq j \leq J}$ et ainsi de suite jusqu'à $n = N$.

Assimilation de données

L'assimilation de données est l'ensemble des méthodes mathématiques et numériques permettant de combiner de manière optimale l'information dont on dispose sur un système physique donné, à savoir les équations mathématiques du modèle, les observations ou mesures physiques du système, et les statistiques sur les erreurs commises. Il s'agit d'un problème inverse ! En effet, pour un problème direct, on connaît les variables dites d'état du système (par exemple la vitesse des courants, la température, la salinité dans l'océan) et on en calcule des observations, qui sont des fonctions (en général non surjectives, et non injectives) de ces variables d'état, par exemple la densité ou des trajectoires de flotteurs dérivants. Pour un problème inverse, on fonctionne à rebours : on observe des mesures, fonctions indirectes non bijectives des variables d'état, et on cherche à reconstituer au mieux ces variables, sachant que des erreurs de mesures ont été commises, et que le système obéit à un ensemble de lois physiques et d'équations mathématiques (le modèle). C'est ce que l'on appelle le couplage modèle-données, ou encore l'assimilation de données.

Le problème de couplage modèle-données peut se formuler comme un problème d'optimisation (voir par exemple [4] et [5]). La première étape pour formuler ce problème est d'identifier les paramètres les plus influents sur la qualité de ce que l'on cherche. Par exemple, en météorologie, pour faire des prévisions, on cherche le meilleur état initial possible, en raison notamment de l'effet papillon, autrement dit du caractère chaotique de la dynamique de l'atmosphère, qui présente une grande sensibilité aux conditions initiales. On appelle x le vecteur des paramètres influents. On forme ensuite l'opérateur d'observation G , qui utilise entre autres le modèle pour envoyer un vecteur de paramètres x dans l'espace des observations. Si y est le vecteur des observations, y et $G(x)$ sont donc de même nature, et on va chercher le meilleur x , au sens où l'écart entre les observations y et leur équivalent issu du modèle $G(x)$ soit le plus petit possible, autrement dit on minimise le critère $J(x) = \|G(x) - y\|^2$..

La minimisation se fait de manière itérative : on commence avec une ébauche de x , et on ajuste x par une méthode de descente de type quasi-Newton, de sorte que l'écart $J(x)$ décroît à chaque itération. On s'arrête quand on s'estime assez proche du minimum.



Exemples d'optimisation en dimension 2 :
à gauche une méthode de descente itérative,
à droite un exemple de fonction présentant des optimums locaux

Dans le cas particulier de la météorologie, ce problème est à la fois complexe et coûteux numériquement. La complexité vient de l'opérateur G , qui peut être fortement non-linéaire, de sorte que la fonction $J(x)$ peut être non quadratique et présenter des minima locaux, qui compliquent la recherche du minimum global par méthode quasi-Newton. Un autre problème majeur est que les dimensions des vecteurs mis en jeu sont très grandes : les tailles de x et y peuvent être de l'ordre de 10^6 à 10^9 . Dans ce cas, il est nécessaire de trouver un compromis entre la qualité du résultat obtenu et le temps de calcul. En effet, une prévision météo d'excellente qualité pour le temps qu'il fera demain ne nous sera d'aucun usage s'il nous faut une semaine pour faire le calcul !

Conclusion

En raison de la complexité des processus impliqués dans les sciences de l'environnement, les enjeux pour la recherche actuelle et à venir sont nombreux, à la fois pour la modélisation (amélioration des connaissances, meilleurs outils numériques) mais aussi pour l'assimilation de données (prise en compte de nouveaux types de données comme les images issues des données de satellites). Plus récemment, l'analyse de l'incertitude des modèles systèmes de prévision est aussi devenue un thème de recherche très important, qui nécessite des connaissances mathématiques nouvelles (dans le domaine des probabilités et de la statistique). En effet, même si les prévisions que les modèles outils numériques sont capables de produire ne peuvent pas être parfaites, il est important de pouvoir fournir des indices de confiance sur la qualité de la prévision effectuée.

Sur tous ces aspects (modélisation, assimilation de données, analyse d'incertitudes), la communauté des chercheurs en mathématiques appliquées est très active, à la fois en France et à l'étranger. Il nous reste encore bien des choses à découvrir avant de pouvoir, avec certitude, dire quel temps il fera demain matin !

Références

- [1] J.-P. Demailly, Analyse numérique et équations différentielles, Presses Universitaires de Grenoble, 3^e édition 2006.
- [2] D. Euvrard, Résolution numérique des équations aux dérivées partielles de la physique, de la mécanique et des sciences de l'ingénieur, Masson, 1994.
- [3] L. C. Evans, Partial differential equations, American Mathematical Society, 2nd edition 2010.
- [4] P. G. Ciarlet, Introduction à l'analyse numérique matricielle et à l'optimisation, Dunod, 5^e édition 2007.
- [5] F. Bonnans, J.-C. Gilbert, C. Lemaréchal, C. Sagastizábal, Optimisation Numérique : Aspects Théoriques et Pratiques, Springer, 1997.