

Comment, à partir d'une animation satellite, le météorologue fait-il pour déduire les prévisions pour le lendemain ? Ni l'instinct des experts ni le calcul mental ne suffisent à assurer cette transition délicate : depuis les années quatre-vingt, elle est confiée à des supercalculateurs. Leur activité principale : l'« assimilation de données ».

Prévoir l'état de l'océan

Si l'on est familier des prévisions météorologiques, les prévisions océanographiques sont, elles, beaucoup moins connues. Pourtant, les deux font peu ou prou appel aux mêmes méthodologies, notamment l'utilisation conjointe d'un modèle mathématique et d'observation. La physique de l'océan, ses courants, ses marées et ses vagues obéissent à des lois sur lesquelles se sont penchés des scientifiques réputés tels le Norvégien Harald Sverdrup et l'Autrichien Walter Munk, auxquels on doit les prévisions d'état de la Manche au moment du

Laurent Bertino dirige le groupe de modélisation et d'assimilation de données au Nansen Environmental and Remote Sensing Center (NERSC) à Bergen, en Norvège.

Hans Wackernagel est enseignant-chercheur au centre de géostatistique de l'École des mines de Paris. hans.wackernagel@ensmp.fr

débarquement de juin 1944 [1]. Ils nous ont légué des équations fondamentales qui traduisent les principes de conservation de la masse et de la quantité de mouvement, et qui relient la vitesse du courant à la température de l'eau, à sa salinité et au niveau de sa surface. Il s'agit hélas d'équations peu commodes à résoudre : équations différentielles non linéaires aux dérivées partielles... À défaut de trouver des solutions exactes, les océanographes établissent des méthodes donnant des solutions approchées et laissent les calculs à des supercalculateurs.

Modèle numérique

Les ordinateurs ne connaissant pas les espaces continus, on considère le domaine d'étude comme une grille tridimensionnelle contenant des millions de cellules. Les modèles de circulation de l'eau dans les océans ainsi obtenus sont suffisamment fidèles pour simuler des comportements connus sur une grande échelle : courants, tourbillons, marées et mélanges de masses d'eau [fig. 1]. Cependant, ils ne le sont pas assez pour décrire les turbulences de petite taille. Autre problème : quelles valeurs initiales attribuer aux températures, aux salinités et aux courants pour les différentes cellules de la grille ?

Il y a certes les observations : les satellites nous informent sur les processus de surface, les flotteurs sur les profondeurs inférieures à 2000 mètres et la tomo-

graphie acoustique sur les grandes profondeurs. Mais la couverture reste insuffisante. Afin de combler les vides, on fait appel à l'assimilation de données. De quoi s'agit-il ? D'incorporer (d'« assimiler ») des observations dans un modèle numérique au fur et à mesure qu'elles arrivent, pour en améliorer les prédictions.

Moyennes pondérées

De tout temps les océanographes ont eu recours à l'assimilation de données en traçant les cartes à la main d'après leur expertise du terrain. À cette approche (l'« analyse subjective ») on oppose depuis les années soixante-dix l'« analyse objective », qui consiste à faire appel à des méthodes classiques de statistiques spatiales pour déduire au mieux les valeurs inconnues à partir des mesures. Concrètement, on apporte des corrections à la vitesse du courant, à la température et à la salinité en calculant des moyennes pondérées des erreurs commises par le modèle aux points d'observation.

Coefficients de corrélation

La question devient alors celle du meilleur système de pondération, c'est-à-dire celui pour lequel l'écart entre les valeurs vraies et celles interpolées est le plus petit. On peut le déterminer en toute objectivité, à condition de pouvoir définir un autre modèle, non plus physique mais proba-



HARALD SVERDRUP a dirigé le Scripps Institute of Oceanography, en Californie, de 1936 à 1948. S'étant fait une spécialité de la circulation des océans, il a réalisé de nombreuses prévisions océaniques pour les débarquements des Alliés en Europe, en Afrique et dans le Pacifique.

© ARCHIVES OF THE SCRIPPS INSTITUTION OF OCEANOGRAPHY

biliste. Ce modèle doit préciser les corrélations spatiales entre une température mesurée et celle calculée en un point où l'on ne dispose pas de données. Il doit également préciser les corrélations entre cette température et les autres variables qui lui sont directement liées, comme la salinité ou la hauteur de la surface de l'eau. Le tout forme une gigantesque matrice de coefficients de corrélation.

Afin de simplifier le problème, on pose une hypothèse importante, la « stationnarité » : les corrélations entre deux points sont les mêmes en tout point de l'océan et ne dépendent que de la distance qui sépare ces deux points. Hélas, les océans se prêtent mal à cette hypothèse, car les eaux de différentes profondeurs et de différentes latitudes s'organisent différemment.

Pour dépasser l'aspect statique de cette méthode, des spécialistes du contrôle optimal des systèmes dynamiques – ceux qui évoluent dans le temps – se sont penchés sur le problème de l'assimilation de données en se concentrant sur l'idée d'un pionnier des prévisions météorologiques, le Norvégien Vilhelm Bjerknes [2] : si le modèle physique part d'un état initial correct, alors il donnera la bonne solution. D'où la mise au point de techniques d'assimilation de données dites « variationnelles », dont le principe est de corriger pas à pas l'état initial jusqu'à ce que le modèle reproduise au mieux les données.

Inversion du temps

On minimise une fonction de l'erreur de prédiction, dont les paramètres sont les valeurs initiales des températures et des courants en tout point du domaine. Le problème est alors perçu comme la résolution d'un système d'équations dont on doit sortir une solution optimale et unique, l'état initial du système. À cette fin, il faut inverser le temps, c'est-à-dire faire remonter une information actuelle dans le passé, à l'aide d'un modèle dit « adjoint ».

Reste que les méthodes variationnelles, qui supplantent peu à peu l'analyse objective dans les différents centres de météo, ont encore besoin de matrices

Fig.1 Un exemple de prévision

CARTE DES TEMPÉRATURES DE SURFACE DE L'OcéAN ATLANTIQUE telles qu'elles étaient prédites pour le 9 septembre 2003 par le système Topaz [7] utilisant le filtre de Kalman d'ensemble et le modèle d'océan Hycom. Les couleurs varient du bleu, pour les eaux froides, à l'orange, pour les plus chaudes. Le Gulf Stream se détache de la côte Est des États-Unis.

© INFOGRAPHIES : PROJET TOPAZ

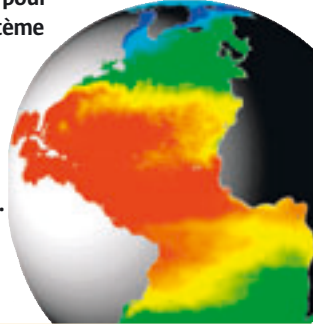
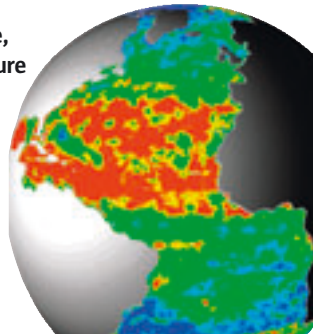


Fig.2 L'incertitude liée à l'estimation

CARTE DE L'INCERTITUDE, anticipée par la modélisation statistique, pour la prévision de température de la figure 1. En rouge apparaissent les régions où la température de la surface de l'océan est la plus incertaine (plus de 1 °C pour une portion d'océan de 30 kilomètres de côté et de 10 mètres de profondeur).



de corrélation et de l'hypothèse de stationnarité. Elles permettent cependant de prendre en compte un plus grand nombre d'observations et de les associer au moyen du modèle adjoint à des phénomènes complexes de grande échelle.

Filtre de Kalman

Également inspirée de la commande automatique des systèmes dynamiques, mais de nature plus statistique, une méthode différente, dite « séquentielle », permet de s'affranchir de l'hypothèse de stationnarité.

L'idée est, cette fois, de faire évoluer les matrices de corrélation à l'aide des erreurs de prédiction du modèle physique en attribuant une certaine marge d'incertitude aux paramètres mal connus. Comme on filtre en quelque sorte les erreurs, cette technique a reçu le nom de « filtre de Kalman », d'après les travaux de l'ingénieur Hongrois

Rudolph Kalman [3]. Contrairement à ce qu'il faisait en analyse objective ou avec les méthodes variationnelles, l'océanographe ne spécifie plus les corrélations spatiales de l'erreur de prédiction ; c'est le modèle lui-même qui s'en charge, en propageant les incertitudes de ses paramètres internes. Ainsi, la matrice de corrélation respecte automatiquement les lois de l'océanographie, puisqu'elle n'est que l'expression de ces mêmes lois. Le filtre de Kalman permet aussi de mettre aisément le doigt sur les imperfections du modèle et des réseaux d'observation en analysant les cartes d'incertitude [fig. 2].

Milieus distincts mais complémentaires

Les méthodes variationnelles et séquentielles ne s'excluent pas les unes les autres. La première s'appuyant sur un modèle « adjoint » et la seconde sur les caractéristiques statistiques des faiblesses du modèle, elles sont complémentaires et rien n'interdit en théorie de les panacher. Mais, en pratique, deux milieux distincts et concurrents se sont développés : les météorologues utilisent les méthodes variationnelles [4] et les océanographes les méthodes séquentielles [5], et chaque milieu a ses outils, son vocabulaire, voire sa philosophie. Rien ne permet de dire aujourd'hui que les deux milieux seront amenés à se rejoindre.

Il existe cependant un lien entre les deux, car toutes les méthodes reposent sur une interpolation spatiale telle qu'elle est pratiquée en géostatistique [6]. Si les matrices de corrélation ont été jusqu'ici au centre des préoccupations, elles ne constituent qu'une partie du modèle sous-jacent à l'interpolation. Et, dans l'avenir, on continuera à développer les modèles en utilisant les concepts et méthodes de la géostatistique. ■■

POUR EN SAVOIR PLUS

www.larecherche.fr

■ Lire notre présentation plus détaillée des techniques d'assimilation de données séquentielles et de la géostatistique.

[1] W. Munk et D. Day, *Oceanography*, 15, 7, 2002.

[2] www.bjerknes.uib.no/conference2004

[3] www.cs.unc.edu/~welch/kalman

[4] www.meteo.fr/meteonet/decouvr/a-z/html/90_initie.htm

[5] L. Bertino et al., *Int. Stat. Review*, 71, 223, 2003.

[6] C. de Fouquet, « Des statistiques contre la pollution », *La Recherche*, octobre 2003, p. 92.

[7] <http://topaz.nersc.no>