

Analyse de très grands ensembles de données satellites

Bertrand Chapron, Laboratoire d'Océanographie Spatiale, IFREMER

Les données sont de plus en plus nombreuses, provenant des modèles, des observations in-situ, mais surtout des satellites. Tous les jours, nous avons accès à 300 Go de données satellites supplémentaires. Des données qu'il faut pouvoir analyser pour en tirer une information pertinente d'où une utilisation grandissante de méthodes de fouille de données. Cependant nous nous intéressons également à de nombreuses autres méthodes, qui font appel à d'autres thèmes mathématiques, telles que l'assimilation de données image, l'interpolation des données manquantes, l'analyse géométrique.

Enjeux environnementaux

La possibilité d'avoir accès à autant de données est avant tout une prouesse technique, un défi technologique. Pourtant, à l'aide ces observations, nous pouvons espérer résoudre des questions scientifiques majeures. Notamment, les données permettent d'observer des phénomènes à de très petites échelles (de l'ordre de 5km) et des phénomènes non linéaires qui ne sont pas résolus par les modèles. C'est le cas notamment des processus d'échanges entre océan et atmosphère que l'on observe au niveau des déferlantes grâce aux technologies « Glitter ». Ces phénomènes ont un impact important sur le climat global. C'est en effet là que les échanges gazeux ont lieu, c'est là que le carbone est « piégé » dans l'océan. C'est aussi là que l'énergie de l'océan est transmise aux cyclones.

Besoin mathématiques

Ces nouvelles données permettent de se poser des questions dont la réponse étaient jusqu'à maintenant inaccessibles. Mais pour tirer profit de ces données, il faut construire de nombreux outils mathématiques dont quelques exemples sont cités ci-dessous :

- L'analyse de données passe par le développement de méthodes d'assimilation d'images, avec par exemple pour défi, la gestion des données manquantes tels que les nuages. Sur ce point spécifique, des travaux sur l'interpolation optimale dynamique sont nécessaires. Le traitement d'image devient également une part importante de l'analyse de données, des travaux récents sur le traitement des données lagrangiennes par exposant de Lyapunov sont en cours avec l'équipe INRIA d'Etienne Mesmin, à Rennes.
- L'assimilation de données doit aussi pouvoir prendre en compte la multi-résolution, notamment pour intégrer des observations à haute résolution dans des modèles à plus basse résolution.
- Se dessine également le développement de nouveaux outils d'analyse géométrique basés sur les contours dynamiques, ou sur les iso-lignes. Ces analyses nécessitent le développement de plus d'interactions avec la géométrie, avec le besoin de méthodes de géométrie aléatoire des contours et la nécessité de pouvoir gérer le conditionnement de la géométrie haute résolution par un champ à basse résolution.
- Enfin, la multiplication des données et leur mutualisation permettent d'avoir des jeux de données pluridisciplinaires accessibles et ouvertes, ce qui encourage la mise en place de scénarios de tests et de validations de différents modèles.

Au delà de ces questions très scientifiques, se pose une question d'ordre technique. Comment archiver toutes ces données ? Deux niveaux d'archivage seraient judicieux. Un premier niveau, pratique, où les données seraient décomposées et réduites selon le mode d'analyse jugé pertinent à l'heure actuelle. Et un autre niveau, vision au long terme, dans lequel toutes les données serait mémorisée. Notre compréhension des phénomènes étant sans cesse en progression, certaine donnée se révélerait précieuse dans le futur.