

METHODES STOCHASTIQUES DE MODELISATION DE DONNEES : APPLICATION A LA RECONSTRUCTION DE DONNEES NON REGULIERES POUR L'ANALYSE DE RISQUES

Jean Baccou
IRSN
Centre de Cadarache
13115 St Paul-Lez-Durance

jean.baccou@irsn.fr

Jacques Liandrat
LATP/ECM
Technopôle de Château-Gombert
13451 Marseille Cedex 20

jacques.liandrat@ec-marseille.fr

Contexte et objectifs :

Les méthodes de modélisation de données ressortent naturellement en analyses de risques puisqu'on est amené à analyser des données (concentration d'un polluant, température d'une gaine de réacteur,...) qui sont le plus souvent discrètes car obtenues via des mesures ponctuelles (capteurs répartis sur le territoire français) ou via un processus numérique (code de calcul simulant un accident hypothétique dans un réacteur). Parmi les méthodes de modélisation, les approches stochastiques sont souvent privilégiées car elles permettent d'associer une précision à toute quantité qui sera estimée à partir des données via le modèle probabiliste choisi. Cependant, ces méthodes supposent le plus souvent que le phénomène que l'on cherche à expliquer est continu. Or, en pratique, les phénomènes étudiés sont souvent non réguliers voire discontinus. L'objet de la thèse est donc d'abord de proposer de nouvelles techniques stochastiques pour une meilleure modélisation de données non régulières. Contrairement aux méthodes classiques, les méthodes proposées seront construites en deux étapes : une étape d'analyse de la structure des données à partir des observations (il s'agit en particulier de définir un outil pour détecter les zones de non régularité du phénomène) et une étape de prédiction qui intégrera l'information collectée dans la première étape. Cette démarche est très liée aux techniques de traitement de données utilisées pour la compression de signaux ou d'images. On cherchera ensuite à appliquer les nouvelles méthodes de traitement de données non régulières à la prédiction de phénomènes réels (cartographie d'un polluant,...) ou à la représentation de codes de calcul (surface de réponse) pour l'analyse d'incertitude. Dans ce dernier cas, il est aussi possible de choisir la série de points (plan d'expérience,[1]) où l'on veut connaître la réponse du code et qui servira à la construction du modèle. Cette thèse doit donc permettre aussi d'intégrer l'information a posteriori fournie par la détection des zones de non régularité pour la construction de nouveaux plans d'expérience. En effet, les plans d'expérience classiques sont basés sur des critères conduisant à une occupation uniforme du domaine de variation des paramètres. L'uniformité de la répartition n'est pas adaptée pour la représentation de phénomènes non réguliers pour lesquels l'information pertinente se situe dans les zones de non régularité. Cette information n'est accessible qu'après la segmentation des données observées réalisée dans la première étape des méthodes que l'on propose de construire. Cette optimisation de la répartition des points du plan d'expérience permettra un gain de calcul conséquent tout en décrivant précisément le phénomène étudié.

Déroulement

La première étape de la thèse consistera d'abord à réaliser une étude bibliographique sur les différentes techniques de modélisation de données (en particulier les techniques géostatistiques de krigeage [2]) puis d'explorer de nouvelles pistes en vue de l'amélioration des outils déjà existants. Cela concerne : l'étude de procédés d'enrichissement du plan d'expérience intégrant des approches venant du traitement d'images (détection de contours et segmentation, [3]), la construction de méthodes locales d'interpolation basées sur la sélection d'un stencil d'interpolation dépendant de la régularité locale des données dans un voisinage du point à estimer ([4]). Pour ce travail, le doctorant bénéficiera de l'appui de l'Ecole Centrale Marseille (Prof. J. Liandrat).

Dans un deuxième temps, le doctorant appliquera les méthodes construites et analysées dans la première étape à des cas synthétiques puis à un cas réel issu des différents projets sur lesquels l'IRSN travaille.

Bibliographie

- [1] G. Box et N. Draper. *Empirical model-building and response surfaces*, Wiley series in probability and mathematical statistics, 1987.
- [2] N. Cressie. *Statistics for spatial data*, Wiley series in probability and mathematical statistics, 1993.
- [3] T. Chan et L. Vese, *Active contours without edges*, IEEE Trans. On Image Processing, 10(2), 266-277, 2001
- [4] F. Arandiga et R. Donat. *Non linear multi-scale decompositions: the approach of A. Harten*, Numer. Algo., 23:175-216,2000.