

# LOI *A PRIORI* SUR LES PARAMÈTRES DU KRIGEAGE TRANS-GAUSSIEN

Joseph Muré <sup>1,2</sup> & Josselin Garnier <sup>2</sup> & Loic Le-Gratiet <sup>1</sup> & Anne Dutfoy <sup>1</sup>

<sup>1</sup> *Département de Management des Risques Industriels (MRI), EDF R&D, Chatou et Saclay, France / [prenom.nom@edf.fr](mailto:prenom.nom@edf.fr)*

<sup>2</sup> *Laboratoire de Probabilités et Modèles Aléatoires (LPMA), Université Paris Diderot (Paris 7), Paris, France / [garnier@math.univ-paris-diderot.fr](mailto:garnier@math.univ-paris-diderot.fr)*

**Résumé.** Nous cherchons à étendre le métamodèle de krigeage à des processus non gaussiens. Nous examinons le krigeage trans-gaussien, qui consiste à supposer que le processus est gaussien après application d'une certaine transformation. Cette méthode pose un problème d'estimation de paramètres qui est traité par une approche bayésienne. Cette approche dissocie paramètre gouvernant la transformation et paramètres gouvernant le processus gaussien sous-jacent, dans le but de faire porter toute l'information *a priori* sur le paramètre gouvernant la transformation.

**Mots-clés.** Métamodèle, processus gaussien, krigeage trans-gaussien, estimation de paramètres, approche bayésienne, loi *a priori*, information.

**Abstract.** We aim at extending the Kriging metamodel to non-Gaussian processes. We consider trans-Gaussian Kriging, in which the process is supposed to be Gaussian after application of a certain transformation. This method generates a problem of parameter estimation which can be tackled using a Bayesian approach. This approach separates the transformation parameter from the parameters of the Gaussian process in order to make the transformation parameter carry all available *a priori* information.

**Keywords.** Metamodel, Gaussian process, trans-Gaussian Kriging, parameter estimation, Bayesian approach, *a priori* distribution, information.

## 1 Introduction

Le krigeage est un métamodèle couramment utilisé pour émuler un code informatique. En plus d'être facile d'utilisation, il est capable d'assortir ses prédictions d'intervalles de confiance. Cependant, ce procédé repose sur l'hypothèse selon laquelle le code à simuler est une réalisation d'un processus gaussien stationnaire paramétré par un jeu de paramètres  $\boldsymbol{\eta}$ . Bien sûr, cette hypothèse devient de moins en moins restrictive à mesure que le nombre de points d'observation augmente car la loi gaussienne est prise conditionnellement aux observations, ce qui permet au krigeage de s'y adapter. Mais quand les observations sont rares (par exemple dans le cas où le code simulé est coûteux en temps de calcul), cette

hypothèse doit être vérifiée afin de garantir les intervalles de confiance fournis.

Il arrive que l'hypothèse soit clairement fautive. Quand le code simulé ne peut délivrer que des valeurs positives, il est déraisonnable de s'attendre à ce qu'il suive une loi gaussienne. C'est pourquoi nous travaillons sur le krigeage trans-gaussien, une méthode qui suppose qu'un certain membre d'une famille de transformations indexée par un paramètre  $\lambda$  doit être appliqué à la sortie scalaire du code afin d'obtenir un processus gaussien stationnaire. Se pose donc le problème de l'estimation du paramètre de transformation  $\lambda$ , en plus de celle du jeu de paramètres  $\boldsymbol{\eta}$ . Une bonne estimation de ces paramètres permettra d'appliquer le krigeage trans-gaussien à un code développé par EDF (Code\_Carmel3D) pour estimer la probabilité de détection de défaut de certains composants d'une centrale nucléaire et fournir des intervalles de confiance fiables.

## 2 Krigeage trans-gaussien

Définissons quelques notations concernant le krigeage classique. Notons  $\boldsymbol{x}$  la variable d'entrée (multidimensionnelle) du code et  $f(\boldsymbol{x})$  sa sortie scalaire.  $f(\boldsymbol{x})$  est supposée suivre une loi gaussienne dont la fonction d'espérance est (le plus souvent) une combinaison linéaire de  $p$  fonctions connues  $h_1(\boldsymbol{x}), h_2(\boldsymbol{x}), \dots, h_p(\boldsymbol{x})$ . Les coefficients de la combinaison linéaire sont inconnus et stockés dans le vecteur  $\boldsymbol{\beta}$ . Le noyau de covariance est inconnu mais supposé être le produit du paramètre réel positif inconnu de variance  $\sigma^2$  et d'un membre inconnu d'une famille connue de fonctions de corrélation stationnaires indexée par un paramètre (vraisemblablement multidimensionnel)  $\boldsymbol{\theta}$ . Le code est observé en  $n > p$  points  $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n$  et  $\boldsymbol{y}$  est le vecteur de taille  $n$  qui recense les observations (*i.e.*  $f(\boldsymbol{x}_i)$  est la  $i$ -ème composante du vecteur  $\boldsymbol{y}$ ).  $\boldsymbol{H}$  est la matrice connue de taille  $n \times p$  et de rang plein dont les  $p$  colonnes contiennent les valeurs des fonctions  $h_1, h_2, \dots, h_p$  en chaque point d'observation (*i.e.*  $\boldsymbol{H}_{ij} = h_j(\boldsymbol{x}_i)$ ). Ainsi,  $\boldsymbol{H}\boldsymbol{\beta}$  est le vecteur d'espérance du vecteur gaussien  $\boldsymbol{y}$ . Le modèle de krigeage est caractérisé par le paramètre d'espérance  $\boldsymbol{\beta}$ , le paramètre de variance  $\sigma^2$  et le paramètre de corrélation  $\boldsymbol{\theta}$ . Autrement dit,  $\boldsymbol{\eta} = (\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta})^\top$ . Notons  $\boldsymbol{\Sigma}_\theta$  la matrice de corrélation du vecteur gaussien  $\boldsymbol{y}$  puisqu'elle dépend évidemment du paramètre de corrélation  $\boldsymbol{\theta}$ . Avec ces notations, la vraisemblance de  $\boldsymbol{y}$  s'écrit :

$$L(\boldsymbol{y}|\boldsymbol{\eta}) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} |\boldsymbol{\Sigma}_\theta|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}_\theta^{-1}(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{\beta})\right\}. \quad (1)$$

En krigeage trans-gaussien, on suppose que l'image de la sortie de code  $f(\boldsymbol{x})$  par une certaine transformation est gaussienne stationnaire. Notons  $g_\lambda$  un membre générique de la famille de transformations et  $J_\lambda$  le déterminant jacobien de  $g_\lambda$  aux points d'observation.

Si  $\mathbf{z}$  est le vecteur contenant les observations, sa vraisemblance s'écrit :

$$L(\mathbf{z}|\boldsymbol{\eta}, \lambda) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} |\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2\sigma^2}(g_{\lambda}(\mathbf{z}) - \mathbf{H}\boldsymbol{\beta})^{\top} \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}(g_{\lambda}(\mathbf{z}) - \mathbf{H}\boldsymbol{\beta})\right\} J_{\lambda}(\mathbf{z}). \quad (2)$$

### 3 Loi *a priori* sur les paramètres

Comment estimer les paramètres  $\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}, \lambda$ ? Utiliser un estimateur du maximum de vraisemblance (EMV) peut sembler raisonnable, mais n'est pas un choix très robuste, surtout avec peu de points d'observation. En effet, l'EMV se révèle en pratique très sensible à la variation des données d'entrée. Ce problème pourra être corrigé en ayant recours à une loi *a priori* sur les paramètres  $\boldsymbol{\eta}$  et  $\lambda$  afin de stabiliser l'EMV. Deux choix s'offrent à nous quant à l'usage de cette loi *a priori*.

L'un d'eux est de l'utiliser comme pondération de la vraisemblance. En notant  $\pi(\boldsymbol{\eta}, \lambda)$  la loi *a priori*, nous maximiserons le produit  $L(\mathbf{z}|\boldsymbol{\eta}, \lambda)\pi(\boldsymbol{\eta}, \lambda)$ . Cette approche consiste essentiellement à maximiser la loi *a posteriori*, (*i.e.* la loi jointe sur  $\mathbf{z}$ ,  $\boldsymbol{\eta}$  et  $\lambda$  conditionnellement à  $\mathbf{z}$ ) plutôt que la vraisemblance.

L'autre est d'utiliser la loi *a posteriori* pour réaliser la prédiction et calculer les intervalles de confiance. Cette approche est dite *intégralement bayésienne*. C'est elle qui donne théoriquement les meilleurs résultats, mais elle est coûteuse en temps de calcul.

Quoi qu'il en soit, le problème de la détermination de la loi *a priori* se pose. Il nous semble à ce stade raisonnable de séparer le jeu de « paramètres gaussiens »  $\boldsymbol{\eta}$  et le « paramètre de transformation »  $\lambda$ , et que la loi *a priori* sur le premier devrait être exprimée conditionnellement au second. Cela signifie écrire la loi *a priori* sur les paramètres sous la forme  $\pi(\boldsymbol{\eta}|\lambda)\pi(\lambda)$ . Il y a deux raisons de faire cela : premièrement, le jeu de « paramètres gaussiens »  $\boldsymbol{\eta}$  est beaucoup plus facile à interpréter si le paramètre de transformation  $\lambda$  est fixé ; deuxièmement, cela permet de définir le facteur  $\pi(\boldsymbol{\eta}|\lambda)$  de façon générique, puisqu'il ne dépend pas de la famille de transformations choisie.

Nous pensons que toute l'information connue sur le processus devrait être portée par le choix de la famille de transformations d'une part et par la loi *a priori* sur le paramètre de transformation  $\pi(\lambda)$  d'autre part. La détermination de  $\pi(\lambda)$  devra donc se faire au cas par cas puisqu'elle dépend du choix de la famille de transformations.

En revanche, comme aucune information n'est (dans l'idéal) censée être contenue par la loi conditionnelle  $\pi(\boldsymbol{\eta}|\lambda)$ , il est possible de la choisir de façon générique.

On est ainsi ramené au problème de la détermination d'une loi *a priori* « objective » dans un modèle gaussien. Ce problème a été traité par [1] et [2]. Ils recommandent d'utiliser la loi *a priori* de référence de Bernardo (cf. [3] pour sa définition) en considérant

le paramètre d'espérance  $\beta$  comme un paramètre de nuisance relativement aux paramètres de variance et corrélation  $\sigma^2, \theta$ . Cela signifie concrètement que, dans l'algorithme de calcul du prior de référence de Bernardo, on calculera d'abord  $\pi(\beta|\sigma^2, \theta, \lambda)$ , puis  $\pi(\sigma^2, \theta|\lambda)$  afin de reconstituer  $\pi(\eta|\lambda) = \pi(\beta|\sigma^2, \theta, \lambda)\pi(\sigma^2, \theta|\lambda)$ .

Nous pensons que les paramètres de variance et de corrélation devraient, eux aussi, être séparés, car il n'y a pas de raison que le paramètre de variance  $\sigma^2$  joue un rôle similaire à celui du paramètre de corrélation  $\theta$ . Précisément, nous considérons  $\sigma^2$  comme un paramètre de nuisance relativement à  $\theta$ . Cela implique de calculer  $\pi(\sigma^2|\theta, \lambda)$  puis  $\pi(\theta|\lambda)$  afin de reconstituer  $\pi(\sigma^2, \theta|\lambda) = \pi(\sigma^2|\theta, \lambda)\pi(\theta|\lambda)$ .

Bien sûr, le plus difficile est de déterminer une loi *a priori* sur le paramètre de corrélation  $\theta$  sachant  $\lambda$  dans le cas où  $\theta$  est multidimensionnel.

Pour le faire, nous nous concentrons sur le cas où la fonction de covariance appartient à la classe de Matérn tensorisée avec régularité fixée. Dans ce contexte,  $\theta$  est le vecteur contenant les longueurs de corrélation suivant les différentes directions d'un repère orthonormé.

La loi *a priori* de Bernardo sur le paramètre  $\theta$  sachant  $\lambda$  est, à condition de ne pas hiérarchiser les différentes composantes de  $\theta$ , la loi *a priori* de Jeffreys sur  $\theta$  sachant  $\lambda$  relativement à la vraisemblance intégrée  $L^I(\mathbf{z}|\theta, \lambda) = \int L(\mathbf{z}|\eta, \lambda)\pi(\beta|\sigma^2, \theta, \lambda)\pi(\sigma^2|\theta, \lambda)d\beta d\sigma^2$ . Par souci de simplicité, nous ne donnons ici son expression que dans le cas où  $\beta$  est connu (krigeage simple), c'est-à-dire le cas où  $\pi(\beta)$  est une mesure de Dirac. Il s'agit de la racine carrée du déterminant de la matrice  $I_\theta$ , dont un élément générique  $(i, j)$  est donné par

$$[I_\theta]_{ij} \propto \text{Tr} \left[ \left( \frac{\partial}{\partial \theta_i} \Sigma_\theta (\Sigma_\theta)^{-1} \right) \left( \frac{\partial}{\partial \theta_j} \Sigma_\theta (\Sigma_\theta)^{-1} \right) \right] - \frac{1}{n} \text{Tr} \left[ \frac{\partial}{\partial \theta_i} \Sigma_\theta (\Sigma_\theta)^{-1} \right] \text{Tr} \left[ \frac{\partial}{\partial \theta_j} \Sigma_\theta (\Sigma_\theta)^{-1} \right]. \quad (3)$$

On peut remarquer que cette loi *a priori* coïncide avec celle donnée par [1] et par [2] dans le cas où  $\theta$  est monodimensionnel. Mais dans le cas où  $\theta$  est multidimensionnel, elle souffre du problème classique des lois *a priori* de Jeffreys multidimensionnelles, celui précisément auquel la loi *a priori* de référence de Bernardo veut remédier : elle ne vérifie pas le principe de vraisemblance (cf. [4]). Il paraît donc judicieux de séparer les composantes de  $\theta$ . Nous exposerons et analyserons plusieurs méthodes pour le faire.

## Bibliographie

- [1] James O. Berger, Victor De Oliveira, and Bruno Sansò. Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association*, 2001, 96 :1361–1374.

- [2] Rui Paulo. Default priors for Gaussian processes. *Annals of Statistics*, 2005, Vol. 33, No. 2, 556–582.
- [3] James O. Berger, José M. Bernardo, and Dongchu Sun. The formal definition of reference priors. *Annals of Statistics*, 2009, Vol. 37, No. 2, 905-938.
- [4] Christian P. Robert. Le choix bayésien : Principes et pratique. *Springer Science & Business Media*, 2006.