Apprentissage de données multi-fidélités par mélange de processus gaussiens. Learning of multi-fidelity data with a mixture of Gaussian processes.

Matthias De Lozzo[1,2] & Loic Le Gratiet[3,4]

[1] *ONERA (DTIM), 2 av. É. Belin, 31055 Toulouse, France ; matthias.de_lozzo@onera.fr*
[2] *EPSILON, Portes Sud - Bât. 3, 12 rue L. C. de Viçose, 31100 Toulouse, France*
[3] *Université Paris Diderot - Paris VII 75205 Paris Cedex 13*
[4] *CEA, DAM, DIF, F-91297 Arpajon, France ; loic.le-gratiet@cea.fr*

Résumé. Afin d'approcher un phénomène physique, nous disposons d'une source d'observations haute-fidélité (HF) coûteuse en utilisation et de plusieurs sources de données basse-fidélités (BF) plus économiques. Nous construisons un mélange de processus gaussiens multi-fidélités basé sur ces observations dont l'espérance fournit le modèle de substitution de la source HF et la variance l'erreur quadratique du modèle. Les processus gaussiens multi-fidélités sont ici des co-krigeages construits via une "agrégation par validation croisée", chacun associé à une des sources BF et à la source HF. Une classification dure basée sur un critère de prédiction est effectuée de sorte à associer à tout point du domaine des paramètres d'entrée un unique co-krigeage moyenné. Cette classification est par la suite considérée comme la réalisation d'une variable aléatoire afin de rendre compte de la continuité du phénomène physique.

Mots-clés. Observations multi-fidélités, processus gaussien, co-krigeage, classification, mélange de modèles, agrégation, validation croisée.

Abstract. In order to approximate a physical phenomenom, we consider a high-fidelity (HF) but expensive data source and several economical but low-fidelity (LF) data sources. We build a mixture of multi-fidelity Gaussian processes from these observations ; its mean is our surrogate model for the HF source and its variance is the quadratic error of the model. A multi-fidelity Gaussian process is an averaged co-kriging based upon a cross-validation aggregation. A hard classification is realized thanks to a prediction criterion in order to associate an unique averaged co-kriging model for each input parameter. The classification is considered as the realization of a random variable so as to respect the continuity of the physical phenomenon.

Keywords. Multi-fidelity observations, Gaussian process, co-kriging, classification, mixture of models, aggregation, cross-validation.

1 Introduction

La méthode développée consiste à approcher un phénomène physique d'intérêt par apprentissage d'observations de fidélités diverses. Ces dernières proviennent de s sources différentes : des codes de calcul fins, des versions dégradées de ces codes (résultats non convergés, maillages grossiers, schémas numériques moins précis, hypothèses physiques simplifiées, ...), des modèles physiques réduits, des mesures expérimentales, ... Une de ces sources est considérée de haute-fidélité (HF) et les $s-1$ autres de basse-fidélité (BF), chacune pouvant avoir une "région de fiabilité" possiblement non-connexe où sa sortie est proche de celle de la source HF.

Pour mener à bien cette approximation, on considère le modèle de co-krigeage présenté dans [1] qui utilise une relation auto-régressive à l'ordre 1 entre les s sources de données de fidélité croissante. Celle de plus basse fidélité est approchée par un premier processus gaussien et les écarts entre deux sources de fidélités successives sont également modélisés par des processus gaussiens. Ce modèle à s niveaux n'étant valable que dans le cas de sources de données hiérarchisées selon leurs fidélités, il n'est pas applicable dans notre cas où les $s-1$ différentes sources BF possèdent des régions de fiabilité distinctes et des fiabilités globales non comparables. Ainsi, nous sommes amenés à construire $s-1$ modèles de co-krigeage à deux niveaux de fidélité que nous cherchons ensuite à assembler.

Pour cela, plusieurs méthodes de mélange de modèles ont été proposées dans la littérature (voir par exemple [2], [3] et [4]). Elles reposent sur une partition de l'espace des paramètres d'entrée en fonction du comportement de la sortie via l'utilisation d'arbres de classification (e.g. CART, voir [2]) ou de mélanges de gaussiennes sur l'espace des entrées/sorties [4]. Un modèle de substitution global est alors constitué par l'assemblage de modèles de substitution locaux construits sur chaque élément de la partition ; on parle aussi de mélange d'experts. Ce type de méthode a notamment été utilisé pour des modèles de régression par processus gaussiens dans le cadre d'approximation de codes de calcul [3].

Dans nos travaux, nous utilisons un découpage de l'espace des paramètres d'entrée basé sur la partition de Voronoï associée aux observations HF. Contrairement aux approches classiques qui se fondent sur le comportement de la sortie de la source HF ou sur la densité des observations HF, notre méthode consiste à associer à la cellule de Voronoï d'un point HF le co-krigeage ayant la meilleure capacité de prédiction en ce point. Cette capacité de prédiction est évaluée à l'aide d'un critère de validation croisée ; disposant de peu d'observations HF, on utilise sa version *leave-one-out* (LOO). Le choix de ce critère repose sur l'hypothèse que l'erreur de prédiction LOO représente correctement l'erreur locale des co-krigeages. Ceci n'est *a priori* pas le cas pour le modèle de co-krigeage de [1] qui interpole aux points d'apprentissage. Ainsi plutôt que d'utiliser le co-krigeage de [1], nous proposons un co-krigeage régressant de type "agrégation par validation croisée" effectuant la moyenne des co-krigeages construits par la méthode LOO. Par ailleurs lorsque le nombre d'observations HF est faible, on observe que ce co-krigeage est plus robuste que celui de [1]. Cette méthode est présentée dans [5] et est semblable aux approches *bagging* [6] et *boosting* [7] basées sur les techniques de *bootstrap*.

2 Modèle de co-krigeage moyenné

Nous considérons ici une source HF notée $z(x)$ et une source BF notée $z_{BF}(x)$, où $x \in Q \subset \mathbb{R}^d$. De plus, on note $\mathbf{z}^n$ (resp. $\mathbf{z}_{BF}^{n_{BF}}$) les réponses de la source $z(x)$ (resp. $z_{BF}(x)$) sur le plan d'expériences $\mathbf{D}$ de taille n (resp. $\mathbf{D}_{BF}$ de taille n_{BF}). Dans un cadre de co-krigeage, on considère que la source $z(x)$ (resp. $z_{BF}(x)$) est une instance du processus gaussien $Z(x)$ (resp. $Z_{BF}(x)$) et on impose la contrainte $\mathbf{D} \subset \mathbf{D}_{BF}$. Nous notons $\mathbf{Z}^n := Z(\mathbf{D})$, $\mathbf{Z}_{BF}^{n_{BF}} := Z_{BF}(\mathbf{D}_{BF})$ et $\mathbf{x}'$ la transposée de $\mathbf{x}$.

Nous supposons que $Z_{BF}(x)$ est un processus gaussien de moyenne $\mathbf{f}'_{BF}(x)\boldsymbol{\beta}_{BF}$, $\mathbf{f}_{BF}(x)$ étant un vecteur de fonctions de base que l'on se donne, et de structure de covariance $\sigma_{BF}^2 r_{BF}(x,\tilde{x};\theta_{BF})$, $\forall(x,\tilde{x}) \in Q^2$. Les équations classiques du modèle de krigeage associé à la source BF s'obtiennent à partir de la loi $[Z_{BF}(x)|\mathbf{Z}_{BF}^{n_{BF}} = \mathbf{z}_{BF}^{n_{BF}}]$ (voir [1]). Une approche alternative pour leur obtention consiste à considérer $\tilde{Z}_{BF}(x)$, un processus gaussien de moyenne nulle et de structure de covariance $\sigma_{BF}^2 r_{BF}(x,\tilde{x};\theta_{BF})$, et à définir le processus $Z_{BF}^{n_{BF}}(x)$ par :

$$Z_{BF}^{n_{BF}}(x) = \mu_{BF}(x) - \tilde{\mu}_{BF}(x) + \tilde{Z}_{BF}(x) \tag{1}$$

où :

$$\mu_{BF}(x) = \left(\mathbf{f}'_{BF}(x)\hat{\boldsymbol{\beta}}_{BF} + \mathbf{r}'_{BF}(x)\mathbf{R}_{BF}^{-1}\left(\mathbf{z}_{BF}^{n_{BF}} - \mathbf{F}_{BF}\hat{\boldsymbol{\beta}}_{BF}\right)\right),$$

$$\tilde{\mu}_{BF}(x) = \left(\mathbf{f}'_{BF}(x)\tilde{\boldsymbol{\beta}}_{BF} + \mathbf{r}'_{BF}(x)\mathbf{R}_{BF}^{-1}\left(\tilde{Z}_{BF}(\mathbf{D}_{BF}) - \mathbf{F}_{BF}\tilde{\boldsymbol{\beta}}_{BF}\right)\right),$$

$\hat{\boldsymbol{\beta}}_{BF} = (\mathbf{F}'_{BF}\mathbf{R}_{BF}^{-1}\mathbf{F}_{BF})^{-1}\mathbf{F}'_{BF}\mathbf{R}_{BF}^{-1}\mathbf{z}_{BF}^{n_{BF}}$, $\tilde{\boldsymbol{\beta}}_{BF} = (\mathbf{F}'_{BF}\mathbf{R}_{BF}^{-1}\mathbf{F}_{BF})^{-1}\mathbf{F}'_{BF}\mathbf{R}_{BF}^{-1}\tilde{Z}_{BF}(\mathbf{D}_{BF})$, $\mathbf{F}_{BF} := \mathbf{f}'_{BF}(\mathbf{D}_{BF})$, $\mathbf{r}_{BF}(x)' := r_{BF}(x,\mathbf{D}_{BF};\theta_{BF})$ et $\mathbf{R}_{BF} = r_{BF}(\mathbf{D}_{BF},\mathbf{D}_{BF};\theta_{BF})$. En tant que transformation linéaire du processus gaussien $\tilde{Z}_{BF}(x)$, ce processus $Z_{BF}^{n_{BF}}(x)$ est gaussien. De plus, il suit la loi $[Z_{BF}(x)|\mathbf{Z}_{BF}^{n_{BF}} = \mathbf{z}_{BF}^{n_{BF}}]$. $Z_{BF}^{n_{BF}}(x)$ est notre modèle de krigeage pour la source BF ; son espérance constitue le modèle de substitution de la source $z_{BF}(x)$ tandis que sa variance représente son erreur quadratique.

On suppose ensuite la relation suivante entre $Z(x)$ et ce processus $Z_{BF}^{n_{BF}}(x)$:

$$\begin{cases} Z(x) = \rho Z_{BF}^{n_{BF}}(x) + \delta(x) \\ Z_{BF}^{n_{BF}}(x) \perp \delta(x) \end{cases} \tag{2}$$

où $\delta(x)$ suit la loi d'un processus gaussien de moyenne $\mathbf{f}'_\delta(x)\boldsymbol{\beta}_\delta$, $\mathbf{f}_\delta(x)$ étant un vecteur de fonctions de base, et de structure de covariance $\sigma_\delta^2 r_\delta(x,\tilde{x};\theta_\delta)$. Le processus $\delta(x)$ modélise le biais entre la source HF et la source BF pondérée par ρ. On souhaite l'apprendre à partir des observations HF et BF, et ce de façon robuste par rapport aux quelques observations HF. On considère pour cela $\tilde{\delta}(x)$, un processus gaussien centré et de structure de covariance $\sigma_\delta^2 r_\delta(x,\tilde{x};\theta_\delta)$. Nous proposons alors le modèle de co-krigeage moyenné suivant :

$$\bar{Z}^n(x) = \rho Z_{BF}^{n_{BF}}(x) + \frac{1}{n}\sum_{i=1}^{n} \delta_{-i}^n(x) \tag{3}$$

où de la même façon que dans (1), on pose :

$$\delta_{-i}^n(x) = \mu_{\delta,-i}(x) - \tilde{\mu}_{\delta,-i}(x) + \tilde{\delta}(x) \tag{4}$$

avec $\mu_{\delta,-i}(x) = \mathbf{f}'_\delta(x)\boldsymbol{\beta}_\delta + \mathbf{r}'_{\delta,-i}(x)\mathbf{R}^{-1}_{\delta,-i,-i}(\mathbf{z}^n_{-i} - \mathbf{F}_{\delta,-i}\boldsymbol{\beta}_\delta - \rho z_{BF}(\mathbf{D}_{-i}))$, $\tilde{\mu}_{\delta,-i}(x) = \mathbf{r}'_{\delta,-i}(x)\mathbf{R}^{-1}_{\delta,-i,-i}\tilde{\delta}(\mathbf{D}_{-i})$, $(\rho, \boldsymbol{\beta}_\delta)' \sim \mathcal{N}\left(\boldsymbol{\Sigma}_{\rho,\boldsymbol{\beta}_\delta}\mathbf{H}'\mathbf{R}_\delta^{-1}\mathbf{z}^n, \sigma_\delta^2\boldsymbol{\Sigma}_{\rho,\boldsymbol{\beta}_\delta}\right)$, $\boldsymbol{\Sigma}_{\rho,\boldsymbol{\beta}_\delta} = (\mathbf{H}'\mathbf{R}_\delta^{-1}\mathbf{H})^{-1}$, $\mathbf{F}_\delta := \mathbf{f}'_\delta(\mathbf{D})$, $\mathbf{r}_\delta(x)' := r_\delta(x, \mathbf{D}_\delta; \theta_\delta)$, $\mathbf{R}_\delta = r_\delta(\mathbf{D}, \mathbf{D}; \theta_\delta)$, $\mathbf{h}' = [\mu_{BF}(x)\ \mathbf{f}'_\delta(x)]$, $\mathbf{H} = [z_{BF}(\mathbf{D})\ \mathbf{F}_\delta]$ et $(\rho, \boldsymbol{\beta}_\delta)$ indépendant de $\tilde{\delta}(x)$ et $Z_{BF}^{n_{BF}}(x)$. Par ailleurs, $\mathbf{A}_{-i}$ est la matrice $\mathbf{A}$ sans la $i^{\text{ème}}$ ligne et $\mathbf{A}_{-i,-j}$ est la matrice $\mathbf{A}$ sans la $i^{\text{ème}}$ ligne et la $j^{\text{ème}}$ colonne.

Dans (3), la loi $[\delta^n_{-i}(x)|\rho, \boldsymbol{\beta}_\delta]$ est égale à la loi $[\delta(x)|\mathbf{Z}^n_{-i} = \mathbf{z}^n_{-i}, Z_{BF}(\mathbf{D}_{-i}) = z_{BF}(\mathbf{D}_{-i}), \rho, \beta_\delta]$. On note que la somme $\sum_{i=1}^n \delta^n_{-i}(x)$ suit une loi gaussienne en tant que transformation linéaire du processus $\tilde{\delta}(x)$ et du vecteur $(\rho, \boldsymbol{\beta}_\delta)$. Nous avons choisi la forme particulière (3)-(4) pour modéliser le biais entre les sources HF et BF afin de ne pas interpoler les points HF et de rendre plus robuste le co-krigeage (2).

Le modèle de co-krigeage moyenné (3) que l'on propose est une agrégation de n modèles de co-krigeage (2) après conditionnement avec les observations $(\mathbf{D}_{-i}, \mathbf{z}^n_{-i})_{i=1,\dots,n}$ et $(\mathbf{D}_{BF}, \mathbf{z}_{BF}^{n_{BF}})$. Dans un cadre de co-krigeage, le modèle de substitution $\hat{f}(x)$ est donné par l'espérance (5) de (3) et sa covariance nous renseigne sur l'incertitude de modèle :

$$\hat{f}(x) := \mathbb{E}[\bar{Z}^n(x)] = \hat{\rho}\mu_{BF}(x) + \mathbf{f}'_\delta(x)\hat{\boldsymbol{\beta}}_\delta + \frac{1}{n}\sum_{i=1}^n \left(\mathbf{r}'_{\delta,-i}(x)\mathbf{R}^{-1}_{\delta,-i,-i}(\mathbf{z}^n_{-i} - \mathbf{F}_{\delta,-i}\hat{\boldsymbol{\beta}}_\delta - \hat{\rho} z_{BF}(\mathbf{D}_{-i}))\right) \tag{5}$$

où $(\hat{\rho}, \hat{\beta}_\delta) = \boldsymbol{\Sigma}_{\rho,\boldsymbol{\beta}_\delta}\mathbf{H}'\mathbf{R}_\delta^{-1}\mathbf{z}^n$. On peut également donner une expression analytique pour la structure de covariance de (3) que nous n'explicitons pas par souci de clarté.

Les paramètres $\boldsymbol{\psi} = (\sigma_\delta^2, \sigma_{BF}^2, \theta_\delta, \theta_{BF})$ sont estimés par maximum de vraisemblance à partir de toutes les observations HF et BF. Une ré-estimation des paramètres $(\rho, \boldsymbol{\beta}_\delta, \sigma_\delta^2, \theta_\delta)$ pour chaque jeu de données $(\mathbf{D}_{-i}, \mathbf{z}^n_{-i})_{i=1,\dots,n}$ est aussi possible ; les équations se généralisent alors naturellement.

Remarque : La structure de modèle présentée en (2) et introduite par [1] est répandue dans la littérature du "Computer Experiments" et a été largement éprouvée sur des cas industriels. Cependant, son défaut majeur est d'être interpolante et donc de manquer de souplesse lorsque le nombre d'observations HF est faible. L'approche classique pour régresser dans un cadre de krigeage consiste à ajouter un terme de régularisation via l'intégration d'un "effet pépite" ; ceci revient à considérer les observations comme bruitées. Néanmoins, dans le cadre des codes de calcul déterministes ce bruit est inexistant et cet outil devient donc délicat à justifier. Le co-krigeage moyenné que nous proposons n'interpole pas les points HF et permet d'accroitre la capacité de généralisation du co-krigeage classique dans le cas d'un faible nombre d'observations HF. Par ailleurs, contrairement à une approche de type "effet pépite", la méthode proposée prend en compte le niveau d'erreur locale (au travers de la procédure LOO) dans la procédure de régularisation. Enfin, quand le nombre d'observations HF grandit, le modèle proposé tend à les interpoler. Ceci est naturel dans un contexte de "Computer Experiments" où l'on cherche souvent à remplacer un code non bruité par un méta-modèle.

3 Assemblage du modèle multi-fidélité

On note $z(x),\ x \in Q \subset \mathbb{R}^d$, une source HF et $z_{BF,1}(x)$ et $z_{BF,2}(x)$ deux [1] sources BF auxquelles on associe les plans d'expériences $\mathbf{D}$, $\mathbf{D}_{BF,1}$ et $\mathbf{D}_{BF,2}$. Le modèle de co-krigeage moyenné introduit précédemment nécessite d'imbriquer le plan d'expériences HF dans l'intersection des plans d'expériences BF.

Nous décrivons ci-dessous les étapes de construction du modèle multi-fidélité par mélange de co-krigeages moyennés.

1. On construit les plans d'expériences $\mathbf{D}$, $\mathbf{D}_{BF,1}$ et $\mathbf{D}_{BF,2}$ pour la source HF et les deux sources BF de telle sorte que $\mathbf{D} \subset \mathbf{D}_{BF,1} \cap \mathbf{D}_{BF,2}$.
2. Pour chaque source $z_{BF,j}(x),\ j = 1,2$:
   (a) on détermine sa zone de fiabilité, *i.e.* les points $x_i \in \mathbf{D}$ tels que $|z(x_i) - z_{BF,j}(x_i)|/(\max \mathbf{z}^n - \min \mathbf{z}^n) < \alpha$ où α est un seuil fixé par l'utilisateur ;
   (b) on apprend sur cette zone le modèle de co-krigeage (2) faisant intervenir la source $z_{BF,j}(x)$ aux points de $\mathbf{D}_{BF,j}$;
   (c) on construit le modèle de co-krigeage moyenné $\bar{Z}_j^n(x)$ à partir des observations de $z(x)$ en $\mathbf{D}$ et de $z_{BF,j}(x)$ en $\mathbf{D}_{BF,j}$ (voir équation (3)).
3. On effectue une classification dure :
   (a) on évalue $\forall x \in \mathbf{D}$ les erreurs de prédiction de $\hat{f}_1(x)$ et $\hat{f}_2(x)$, les modèles de substitution associés respectivement à $\bar{Z}_1^n(x)$ et $\bar{Z}_2^n(x)$;
   (b) on associe à la cellule de Voronoï du point $x \in \mathbf{D}$ le modèle de co-krigeage moyenné pour lequel l'erreur de prédiction est minimale ;
   (c) on définit C_j comme l'union des cellules de Voronoï associées au modèle de co-krigeage moyenné $\bar{Z}_j^n(x),\ j = 1,2$.
   (d) On note $\partial\bar{C}$ la frontière entre C_1 et C_2.
4. On considère la frontière entre C_1 et C_2 comme la réalisation de la variable aléatoire : $\partial C = \partial\bar{C} + \sigma\varepsilon$ où $\varepsilon \sim \mathcal{N}(0, I_d)$; l'écart-type σ est proportionnel à la distance moyenne entre les points connexes (au sens de Delaunay) de $\mathbf{D}$.
5. On génère N réalisations $(\partial C^{(k)})_{k=1,\dots,N}$ de cette variable aléatoire. Pour chaque instance k on obtient alors une partition $(C_1^{(k)}, C_2^{(k)})$ de Q sur laquelle est défini le modèle multi-fidélité suivant :
   $$[\bar{Z}(x)|\partial C = \partial C^{(k)}] = \bar{Z}_1^n(x)\mathbf{1}_{x \in C_1^{(k)}} + \bar{Z}_2^n(x)\mathbf{1}_{x \in C_2^{(k)}}$$
6. On calcule la moyenne et la covariance *a posteriori* de $\bar{Z}(x)$:
   $$\mathbb{E}[\bar{Z}(x)] = \mathbb{E}[\mathbb{E}[\bar{Z}(x)|\partial C]] = \mathbb{P}(x \in C_1)\mathbb{E}[\bar{Z}_1^n(x)] + \mathbb{P}(x \in C_2)\mathbb{E}[\bar{Z}_2^n(x)]$$
   où $\mathbb{P}(x \in C_j)$ est estimée par $\sum_{k=1}^N \mathbf{1}_{x \in C_j^{(k)}}/N$. $\hat{\bar{f}}(x) := \mathbb{E}[\bar{Z}(x)]$ est alors le modèle de substitution de la source HF $z(x)$ construit par mélange de processus gaussiens et $\mathrm{var}[\bar{Z}(x)]$ nous renseigne sur son erreur.

1. On se restreint au cas de deux sources BF par souci de clarté. La méthode est généralisable à un nombre quelconque de sources de données BF.

4 Discussion

La méthode proposée permet d'approcher la sortie d'une source HF au moyen d'un mélange de co-krigeages moyennés construits à partir d'observations provenant d'une source d'observations HF et de plusieurs sources BF. Elle permet d'obtenir une forme analytique du modèle de substitution et de son erreur quadratique. Ce mélange de modèles étend le co-krigeage de [1] au cas où il n'existe pas de hiérarchie entre les différentes sources de données BF. Pour se faire, nous proposons une classification dure basée sur l'erreur de prédiction plutôt qu'une basée sur la vraisemblance entre les entrées et les sorties comme dans [3] et [4] afin de sélectionner localement le meilleur co-krigeage moyenné associé à une source BF. Cette approche est avantageuse par rapport aux méthodes présentées en [3] et [4] qui nécessitent un grand nombre de données HF, ce qui est irréaliste dans notre cas.

Par ailleurs, nous considérons la précédente classification comme la réalisation d'une variable aléatoire afin de tenir compte de la continuité du phénomène physique que nous cherchons à approcher. L'espérance du modèle en fonction de la loi du classifieur nous donne le modèle de substitution de la source HF et sa variance nous donne l'erreur quadratique moyenne du modèle. Nous avons choisi une classification dure aléatoire plutôt qu'une classification floue (i.e. où C_1 et C_2 se superposent) pour conserver une formulation analytique du modèle de substitution et de son erreur quadratique. En effet, dans le cas où l'intersection des deux classes est non vide, l'utilisation d'une classification floue nécessite de déterminer la covariance entre les deux modèles de co-krigeage. En pratique, ceci se révèle extrêmement difficile, voire irréalisable.

La méthode présentée est utilisée pour résoudre un cas industriel d'ingénierie thermique où la source de données HF est un code de calcul.

Bibliographie

[1] Kennedy, M. C. & O'Hagan, A. (2000), Predicting the output from a complex computer code when fast, approximations are available, Biometrika, 87, 1-13.

[2] Friedman, J. & Hastie, T. & Tibshirani, R. (2008), The Elements of Statistical Learning - Data Mining, Inference and Prediction, Springer, New-York.

[3] Gramacy, R.B. & Lee, H.K.H. (2008), Bayesian Treed Gaussian Process Models With an Application to Computer Modeling, Journal of the American Statistical Association, 103 :483, 1119-1130.

[4] Bettebghor, D. & Bartoli, N. & Grihon, S. & Morlier, J. & Samuelides, S. (2011), Surrogate modeling approximation using a mixture of experts based on EM joint estimation, Structural and Multidisciplinary Optimization, 43, 243-259.

[5] Lecué, G. (2012), Oracle inequalities for cross-validation type procedures, Electronic Journal of Statistics, 6, 1803-1837.

[6] Breiman, L. (1996), Bagging Predictors, Machine learning, 24, 123-140.

[7] Freund, Y. (1995), Boosting a weak learning algorithm by majority, Inform. and Comput., 121 (2), 256-285.