



Phimeca Engineering

Centre d'Affaires du Zénith
34 rue de Sarliève
63800 Cournon d'Auvergne

Tél: +33 (0) 4 73 28 93 66

contact@phimeca.com

Fax: +33 (0) 4 73 28 95 78

<http://www.phimeca.com/>

Benchmark GdR MASCOT NUM - Construction de métamodèles prédictifs

Rapport de recherche

Société anonyme au capital de 200 000 €

RCS de Clermont-Ferrand

SIRET : 439 210 477 00055 / APE : 7490 B

Antenne Ile-de-France

241 rue de Bercy
75012 Paris

Tél. : +33 (0) 1 58 51 18 02 Fax : +33 (0) 4 73 28 95 78

Antenne Sud-Est

Bât. F 1er étage – Technopole Var Matin
Route de la Seyne – 83191 Ollioules

Tél. : +33 (0) 4 94 62 51 95 Fax : +33 (0) 4 94 62 59 47

Certification ISO 9001

<p>Client : Interne</p> <p>Adresse : -</p> <p>Affaire : 9999BST005</p> <p>Commande n° : -</p>
<p>Type de document : Rapport de recherche</p> <p>Titre abrégé : Benchmark MASCOT NUM</p> <p>Référence : RT-9999BST005-002B</p> <p>Accessibilité : Libre</p> <p>Mots Clés : métamodèles, sensibilités, krigeage, SVM, chaos polynomial</p>

Version finale

Production

Indice	Date	Rédacteur(s)		Vérificateur/Approbateur	
		Nom(s)	Visa	Nom(s)	Visa
B	18/06/09	F. Deheeger		B. Sudret	

Diffusion

Nom(s)	Société	Date	Nombre
	www.gdr-mascotnum.fr	12/03/10	

Suivi des indices

Indice	Date	Évolutions
A	18/06/09	Création
B	25/02/10	Mise à jour avant diffusion

Synthèse

Contexte

Ce document présente les résultats du benchmark “Construction de métamodèles prédictifs” du GdR MASCOT NUM (www.gdr-mascotnum.fr) sur les données MARTHE (exemple fourni par le CEA Cadarache).

Objectifs de l'étude

L'objectif est d'identifier les paramètres d'entrée du code influents sur le résultat de la simulation. Les méthodes statistiques d'analyse d'incertitudes et de sensibilité sont utilisées. Le code MARTHE étant complexe et gourmand en temps de calcul, une première étape consiste à construire un métamodèle à partir d'un nombre limité de réalisations du code de calcul. Dans l'exemple traité, les expériences numériques sont imposées (calculs déjà faits).

Méthodologie

Trois méthodes de construction de méta-modèles ont été utilisées pour effectuer l'analyse de sensibilité :

- la régression par support vector machine (SVR),
- le krigeage,
- le chaos polynomial creux.

Résultats

L'objectif du benchmark est de comparer les moments statistiques et les indices de sensibilité obtenus par simulation sur les différents méta-modèles construits. Les quantités d'intérêts sont les suivantes : la moyenne et l'écart-type des 10 variables de sortie et finalement, les indices de Sobol' par rapport à chacune des variables d'entrée.

Conclusions et perspectives

Les 3 approches donnent toujours des résultats cohérents sur la hiérarchisation des variables. Les coefficients de partage volumique ($kd1$ ou $kd2$, selon les sorties) sont les paramètres les plus importants. Vient ensuite l'infiltration $i3$. Les variables $i1$, $i2$ et $i3$ étant liées, il est pertinent de regarder la somme des trois indices de sensibilité associés. En terme de validation (coefficient de validation croisée $Q2$), SVR et gPC donnent alternativement les meilleurs résultats (6 fois pour SVR et 4 fois pour gPC). L'erreur de généralisation est donc plus grande pour le krigeage, qui inversement interpole les données d'apprentissage.

Sommaire

1	Le problème physique	9
2	La problématique	10
3	Méthodologie de mise en œuvre	12
3.1	Régression et Support Vector Machine	12
3.2	Krigeage	13
3.3	Chaos polynomial	13
4	Mise en œuvre	14
5	Résultats	15
5.1	Variable de sortie 1 - <i>p102K</i>	15
5.2	Variable de sortie 2 - <i>p104</i>	17
5.3	Variable de sortie 3 - <i>p106</i>	19
5.4	Variable de sortie 4 - <i>p2 – 76</i>	21
5.5	Variable de sortie 5 - <i>p29K</i>	23
5.6	Variable de sortie 6 - <i>p31K</i>	25
5.7	Variable de sortie 7 - <i>p35K</i>	27
5.8	Variable de sortie 8 - <i>p37K</i>	29
5.9	Variable de sortie 9 - <i>p38</i>	31
5.10	Variable de sortie 10 - <i>p4b</i>	33
6	Conclusions	35
7	Références	36

1 Le problème physique

La présentation du cas d'étude est tirée de la présentation de B. Iooss et A. Marrel Iooss et Marrel (2008) réalisée en 2008 lors de la journée du GdR MASCOT NUM. Le cas traité ici est celui proposé par le CEA Cadarache dénommé "données MARTHE".

"En 2005, le CEA et l'Institut Kurchatov (Russie) ont collaboré au développement d'une modélisation du transport de strontium 90 (^{90}Sr) en milieu poreux saturé en eau, pour le cas d'un site de stockage temporaire de déchets radioactifs (STDR) à Moscou. Le but principal était de modéliser le transport de ^{90}Sr à des fins prédictives entre 2002 (où le terme source était connu) et 2010 afin de déterminer le degré de contamination potentielle de la nappe. La simulation numérique du transport de ^{90}Sr , espèce la plus mobile du terme source, dans l'aquifère supérieur du site a été réalisée à l'aide du code de calcul MARTHE (BRGM). La figure 1.1 illustre l'évolution du panache de concentration en ^{90}Sr sur le site. La figure de gauche correspond au champ de concentration initiale (dédit de mesures), la figure de droite correspond au champ de concentration calculé par le code. Les petits rectangles blancs correspondent à l'emplacement sur le site des différents piézomètres (sondes poreuses permettant de prélever l'eau en profondeur et de relever le niveau de la nappe)."

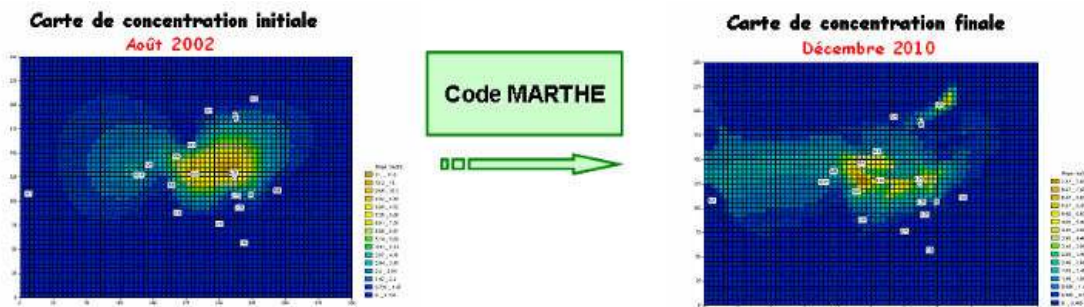


FIG. 1.1 – Illustration des cartes de concentration en ^{90}Sr sur le site.

2 La problématique

L'objectif est d'identifier les paramètres d'entrée du code influents sur le résultat de la simulation. Les méthodes statistiques d'analyse d'incertitudes et de sensibilité sont utilisées. Le code MARTHE étant complexe et gourmand en temps de calcul, une première étape consiste à construire un métamodèle à partir d'un nombre limité de réalisations du code de calcul. Vingt paramètres d'entrée sont retenus dans le cadre de cette analyse d'incertitudes. Ces variables et leur description probabiliste sont synthétisées dans le tableau 2.1.

Paramètres d'entrée	Indicateur	Distribution	Paramètres
Perméabilité couche 1	<i>per1</i>	Uniforme	1 - 15
Perméabilité couche 2	<i>per2</i>	Uniforme	5 - 20
Perméabilité couche 3	<i>per3</i>	Uniforme	1 - 15
Perméabilité zone 1	<i>perz1</i>	Uniforme	1 - 15
Perméabilité zone 2	<i>perz2</i>	Uniforme	1 - 15
Perméabilité zone 3	<i>perz3</i>	Uniforme	1 - 15
Perméabilité zone 4	<i>perz4</i>	Uniforme	1 - 15
Dispersivité longitudinale couche 1	<i>d1</i>	Uniforme	0,05 - 2
Dispersivité longitudinale couche 2	<i>d2</i>	Uniforme	0,05 - 2
Dispersivité longitudinale couche 3	<i>d3</i>	Uniforme	0,05 - 2
Dispersivité transversale couche 1	<i>dt1</i>	Uniforme	$0,01 * d1 - 0,1 * d1$
Dispersivité transversale couche 2	<i>dt2</i>	Uniforme	$0,01 * d2 - 0,1 * d2$
Dispersivité transversale couche 3	<i>dt3</i>	Uniforme	$0,01 * d3 - 0,1 * d3$
Coeff. partage volumique couche 1	<i>kd1</i>	Weibull	$\alpha = 1.1597,$ $\beta = 19.9875$
Coeff. partage volumique couche 2	<i>kd2</i>	Weibull	$\alpha = 0.7591597,$ $\beta = 24.4455$
Coeff. partage volumique couche 3	<i>kd3</i>	Weibull	$\alpha = 1.27363,$ $\beta = 22.4986$
Porosité de toutes les couches	<i>poros</i>	Uniforme	0,3 - 0,37
Infiltration type 1	<i>i1</i>	Uniforme	0 - 0,0001
Infiltration type 2	<i>i2</i>	Uniforme	<i>i1</i> - 0,01
Infiltration type 3	<i>i3</i>	Uniforme	<i>i2</i> - 0,1

TAB. 2.1 - Définition des variables aléatoires.

Le coût de calcul du modèle de simulation MARTHE n'a permis de réaliser que 300 simulations. Ces 300 points sont issus de 3 plans carré latin (LHS) indépendants de 100 points. La phase de construction d'un métamodèle s'est avéré difficile en raison des caractéristiques du problème :

- échantillon de taille limité (300 points) et dimension de l'espace des entrées importante (20 variables aléatoires);
- forte non-linéarité de la physique : présence d'effets de seuil et autres phénomènes du modèle hydrogéologique.

Dans le cadre de ce benchmark, seules 10 variables de sortie sont conservées pour étude : *p102K, p104, p106, p2-76, p29K, p31K, p35K, p37K, p38, p4b*.

L'objectif du benchmark est de proposer et comparer différentes méthodes de constructions de métamodèles avec pour objectif la hiérarchisation des variables d'entrée par rapport à leur influence sur les variables de sortie du modèle MARTHE.

3 Méthodologie de mise en œuvre

Trois méthodes de construction de méta-modèles ont été utilisées pour effectuer l'analyse de sensibilité :

- la régression par support vector machine (SVR),
- le krigeage,
- le chaos polynomial creux.

Les principes de chacune de ces techniques sont brièvement rappelés ci-après. Le point de départ est la donnée d'une base d'apprentissage obtenue par le plan d'expérience numérique dans l'espace des paramètres d'entrée et l'évaluation de la réponse du code de calcul associée, soit :

$$\{(\mathbf{x}_i, \mathbf{y}_i = \mathcal{M}(\mathbf{x}_i)), i = 1, \dots, N_e\} \quad (3.1)$$

où N_e est la taille de la base de données, et \mathcal{M} le code de calcul, ici le modèle MARTHE.

L'objectif de la construction de métamodèle est d'obtenir une formulation analytique $\tilde{\mathcal{M}}$ qui reproduit au mieux le vrai modèle \mathcal{M} .

3.1 Régression et Support Vector Machine

Les support vector machines sont à l'origine une technique statistique de classification de données. Une généralisation des approches *soft margin* pour les classifications non linéaires a mené à l'extension *support vector regression* qui permet de construire un métamodèle à partir d'une base d'apprentissage.

Partant d'une fonction noyau (par exemple, le noyau gaussien $K(\mathbf{x}, \mathbf{y}) = \phi_n(\|\mathbf{x} - \mathbf{y}\|)$, où ϕ_n est la densité de probabilité multinormale en dimension n), on cherche une approximation de la forme :

$$\tilde{\mathcal{M}}(\mathbf{x}) = \sum_{i=1}^{N_e} (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + b \quad (3.2)$$

où les coefficients b , α_i et α_i^* sont solution d'un problème d'optimisation quadratique (Smola et Schölkopf, 2006). Par construction ce métamodèle est creux, au sens où seulement $n < N_e$ termes $\alpha_i - \alpha_i^*$ sont non nuls : les \mathbf{x}_i correspondant sont les *vecteurs supports* du métamodèle.

3.2 Krigeage

Le krigeage est inspiré de méthodes géostatistiques et suppose que la réponse d'un modèle de simulation dont les paramètres d'entrée sont regroupés dans le vecteur \mathbf{x} est une réalisation d'un processus gaussien stationnaire :

$$\tilde{\mathcal{M}}(\mathbf{x}) = Z(\mathbf{x}, \omega) \quad (3.3)$$

dont la fonction d'autocorrélation est ici choisie sous la forme :

$$R(\mathbf{x}, \mathbf{y}) \propto \exp\left[-\sum_{i=1}^n \frac{(x^i - y^i)^2}{\theta_i^2}\right] \quad (3.4)$$

où $\{x^i, i = 1, \dots, n\}$ désigne les coordonnées du vecteur \mathbf{x} . Les paramètres du modèle sont calculés par maximum de vraisemblance, en utilisant la boîte à outils DACE (Lophaven *et al.*, 2002).

3.3 Chaos polynomial

Soit \mathbf{X} le vecteur aléatoire des paramètres d'entrée du modèle, de dimension n , et $y = \mathcal{M}(\mathbf{X})$ la quantité d'intérêt (ici, chacune des 10 variables d'intérêt mesurées). Si cette quantité est de variance finie, on peut la représenter sur la base du chaos polynomial :

$$Y = \sum_{\alpha \in \mathbb{N}^n} \gamma_{\alpha} \Psi_{\alpha}(\mathbf{X}) \quad (3.5)$$

Dans cette équation, $\{\Psi_{\alpha}, \alpha \in \mathbb{N}^n\}$ désigne la base du chaos polynomial (polynômes multivariés en les variables d'entrée rassemblées dans \mathbf{X}) et $\{\gamma_{\alpha}, \alpha \in \mathbb{N}^n\}$ désigne les coefficients à calculer. Le méta-modèle est obtenu en tronquant la série, par exemple en ne retenant que les polynômes de degré total inférieur ou égal à p .

$$\tilde{\mathcal{M}}(\mathbf{X}) = \sum_{|\alpha| \leq p} \gamma_{\alpha} \Psi_{\alpha}(\mathbf{X}) \quad (3.6)$$

Une méthode de régression permet de calculer ces coefficients à partir d'un plan d'expérience numérique et des résultats associés, (Berveiller, 2005). Des travaux récents (Blatman et Sudret, 2010) ont permis de mettre au point une méthode adaptative de construction de bases polynomiales creuses, ce qui permet d'appliquer la méthode de régression à des problèmes de grande taille (comme ici, $n = 20$) avec peu de points dans le plan d'expériences. En effet, un chaos polynomial complet à l'ordre 2 contiendrait ici 462 coefficients et nécessiterait donc un nombre encore plus grand de points dans le plan d'expérience pour réaliser le calcul de régression. La technique des bases creuses permet par contre de construire un méta-modèle même avec un nombre limité de points dans la base de données (ici $N_e = 300$). Les coefficients sont calculés par un algorithme de type *Least Angle Regression*, Efron *et al.* (2004), tel que modifié par Blatman (2009)

4 Mise en œuvre

L'objectif du benchmark est de comparer les moments statistiques et les indices de sensibilité obtenus par simulation sur les différents méta-modèles construits. Les quantités d'intérêts sont les suivantes : la moyenne et l'écart-type des 10 variables de sortie et finalement, les indices de Sobol' par rapport à chacune des variables d'entrée.

Les variables d'entrée sont normées. Les variables de type uniformes sont ramenées sur l'intervalle $[-1, 1]$, les variables de type Weibull sont transformées en variables gaussiennes centrées réduites, donc approximativement sur l'intervalle $[-3, 3]$. Pour obtenir des échantillons des variables de sorties, nous générons des entrées dans cet espace normé.

Certaines lois de paramètres dépendent d'autres entrées. Ces dépendances ne permettent pas de conclure sur les sensibilités.

Pour chacune des variables de sortie, on représente les densités tracées à partir de simulations de Monte-Carlo sur les 3 métamodèles obtenus (100 000 simulations sont utilisées). Un graphique de type "boîte à moustache"¹ permet de comparer également les moments statistiques (moyenne, écart-type, quartile) de ces échantillons générés.

Les indices de Sobol' sont obtenus par simulation (utilisation de suites «quasi Monte-Carlo» avec 10 000 tirages) pour les méta-modèles calculés par krigeage ou Support Vector Machine. Pour le chaos polynomial, les indices de Sobol' se calculent par combinaison des coefficients du développement polynomial.

L'analyse des erreurs de régression est également présentée. L'erreur du métamodèle est évaluée d'une part sur la base d'apprentissage (les 250 premiers points de la base de calcul), et d'autre part sur la base de validation (les 50 derniers points de la base de calcul). Un tracé, de type *scatter plot*², est affiché pour chacune des bases (apprentissage et validation). Le diagramme de type boîte à moustache représente également la distribution des valeurs de l'erreur et sa statistique pour les trois types de métamodèle utilisés.

¹ Cette représentation est composée, pour chaque coefficient :

- d'une boîte dont l'étendue correspond à l'interquartile (l'intervalle de confiance à 75%);
- d'un repère central correspondant à la médiane;
- de moustaches qui s'étendent sur une plage valant 1.5 fois l'interquartile;
- de valeurs suspects en dehors des moustaches.

²En abscisse la valeur de la réponse de la simulation, en ordonnée, la prédiction du métamodèle. Si le nuage de points est proche de la première bissectrice, alors la prédiction du métamodèle est satisfaisante

5 Résultats

5.1 Variable de sortie 1 - $p102K$

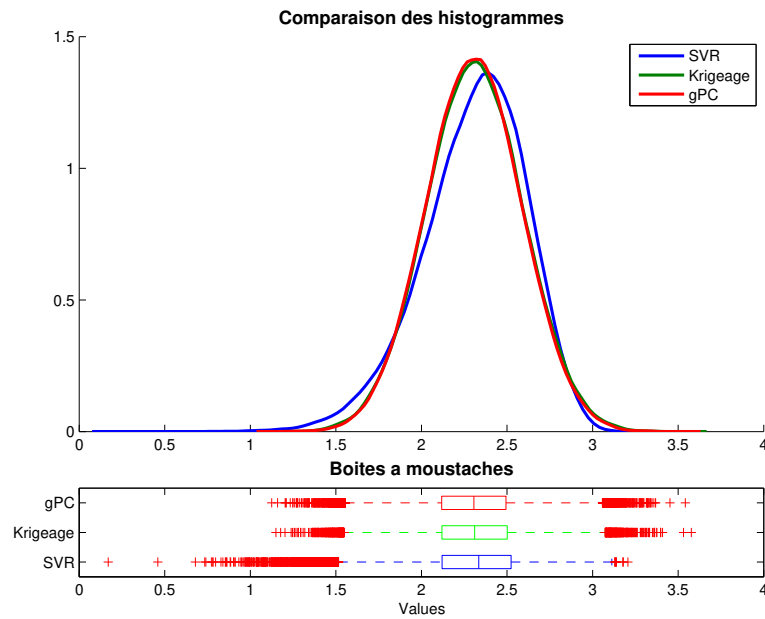


FIG. 5.1 - Comparaison des densités empiriques obtenues par simulation sur les métamodèles obtenus - Variable de sortie 1 - $p102K$.

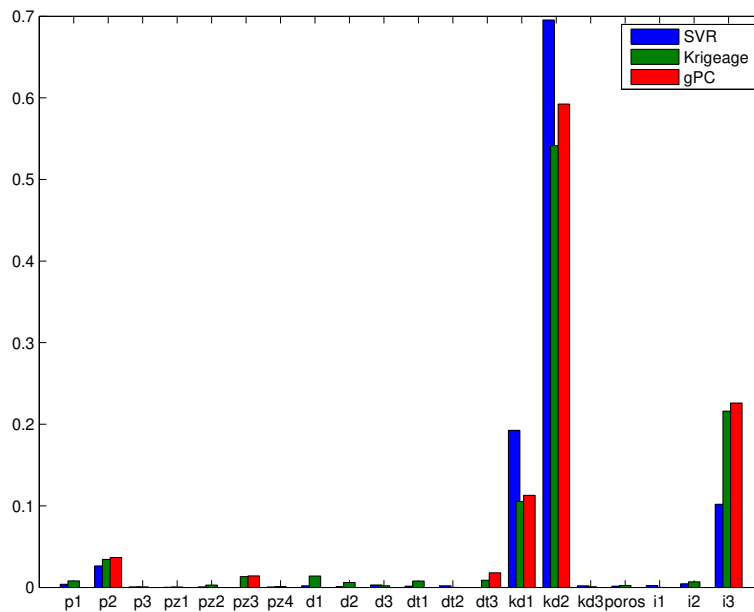


FIG. 5.2 - Comparaison des indices de Sobol' totaux mesurés sur les métamodèles obtenus - Variable de sortie 1 - $p102K$.

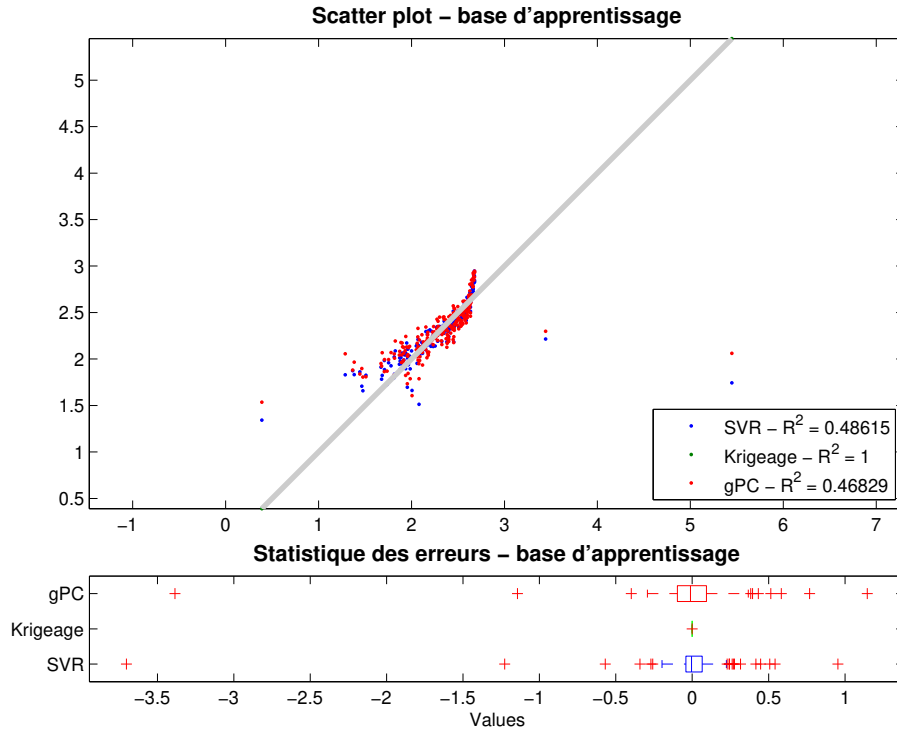


FIG. 5.3 – Analyse des erreurs de la base d'apprentissage (250 points) - Variable de sortie 1 - $p102K$.

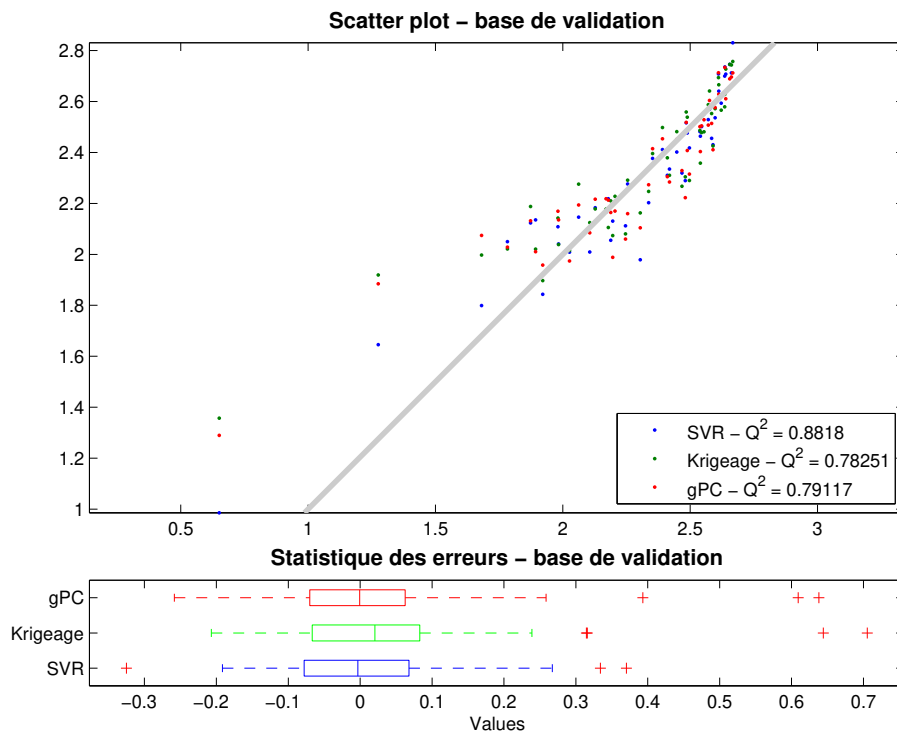


FIG. 5.4 – Analyse des erreurs de la base de validation (50 points) - Variable de sortie 1 - $p102K$.

5.2 Variable de sortie 2 - p104

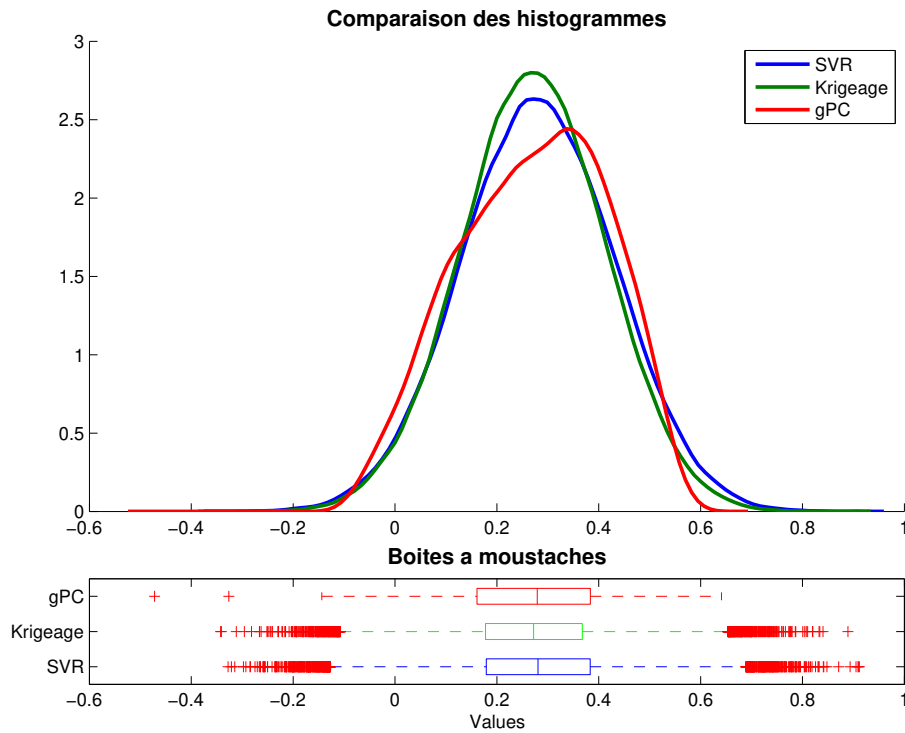


FIG. 5.5 - Comparaison des densités empiriques obtenues par simulation sur les métamodèles obtenus - Variable de sortie 2 - p104.

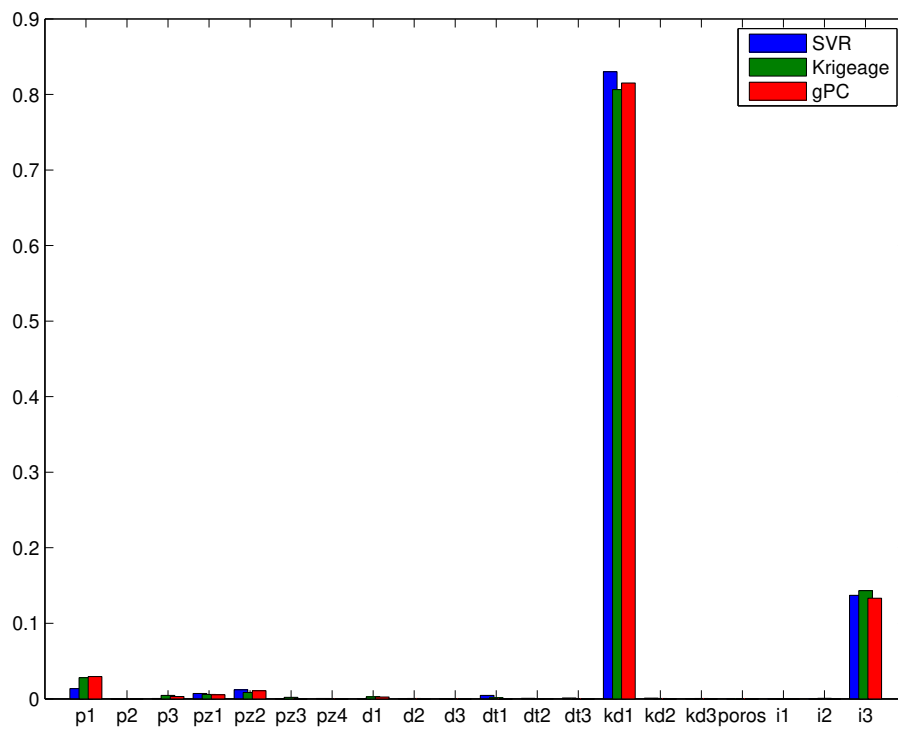


FIG. 5.6 - Comparaison des indices de Sobol' totaux mesurés sur les métamodèles obtenus - Variable de sortie 2 - p104.

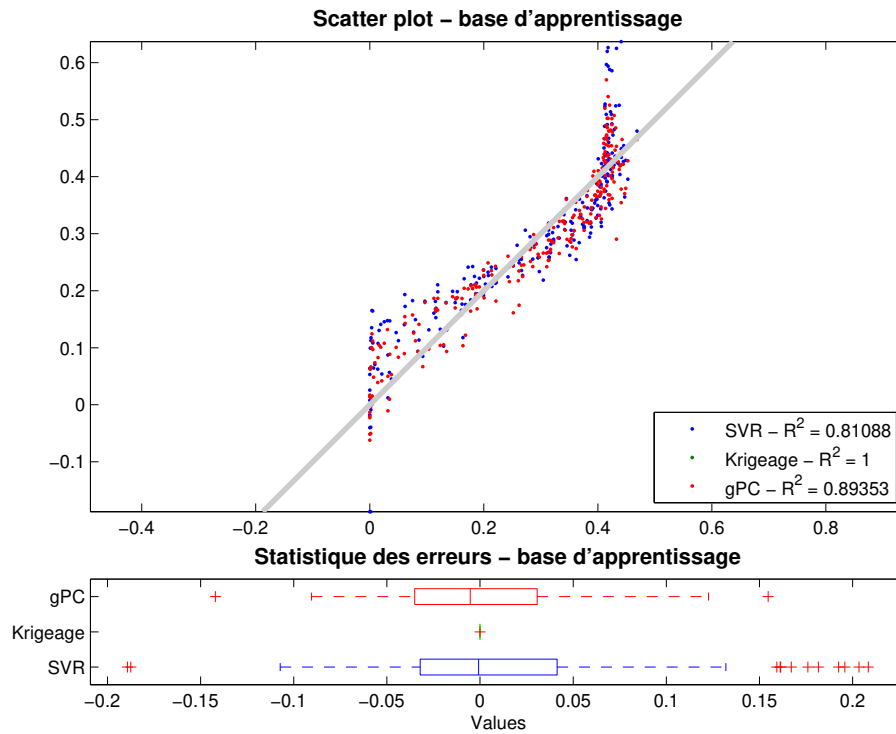


FIG. 5.7 – Analyse des erreurs de la base d'apprentissage (250 points) - Variable de sortie 2 - $p104K$.

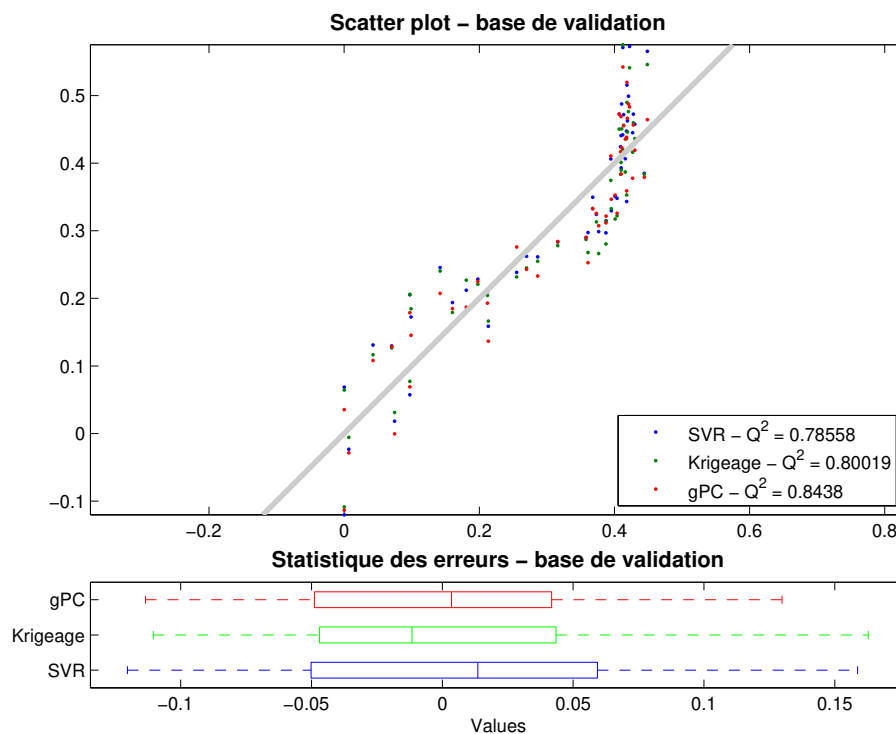


FIG. 5.8 – Analyse des erreurs de la base de validation (50 points) - Variable de sortie 2 - $p104K$.

5.3 Variable de sortie 3 - p106

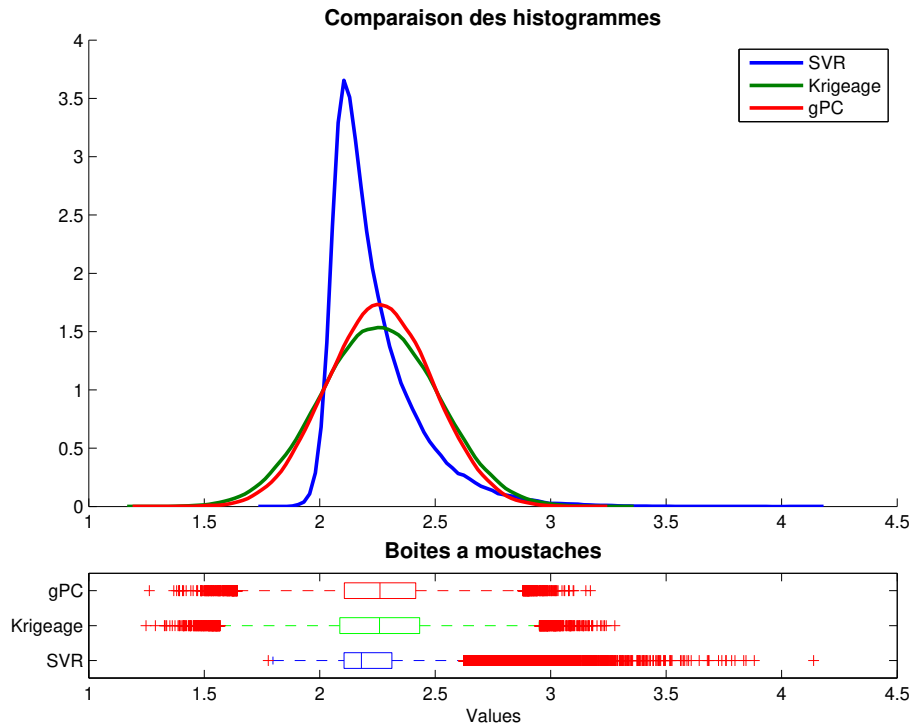


FIG. 5.9 - Comparaison des densités empiriques obtenues par simulation sur les métamodèles obtenus - Variable de sortie 3 - p106.

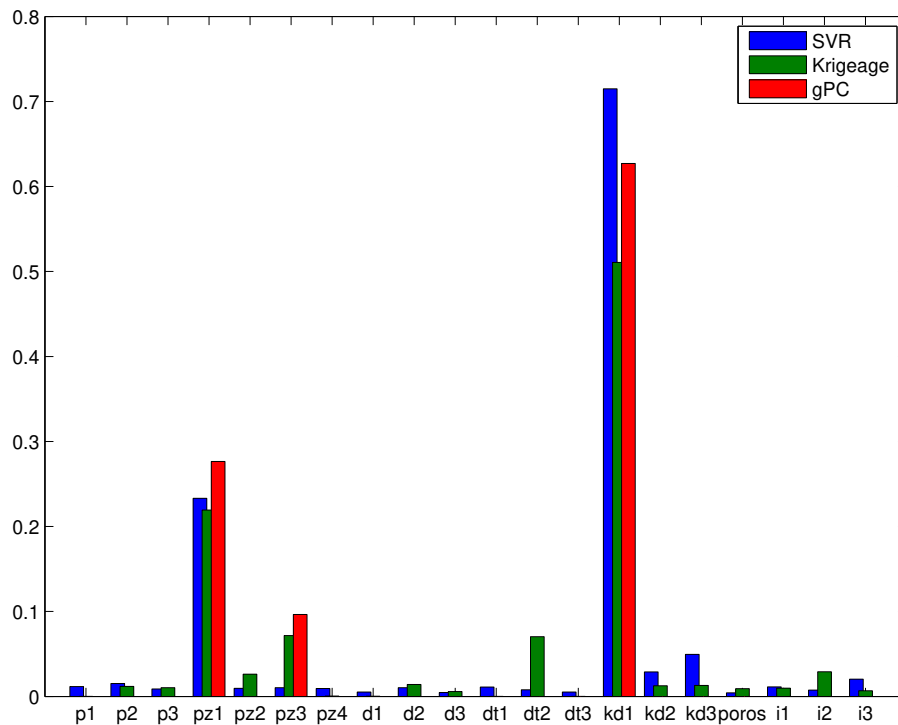


FIG. 5.10 - Comparaison des indices de Sobol' totaux mesurés sur les métamodèles obtenus - Variable de sortie 3 - p106.

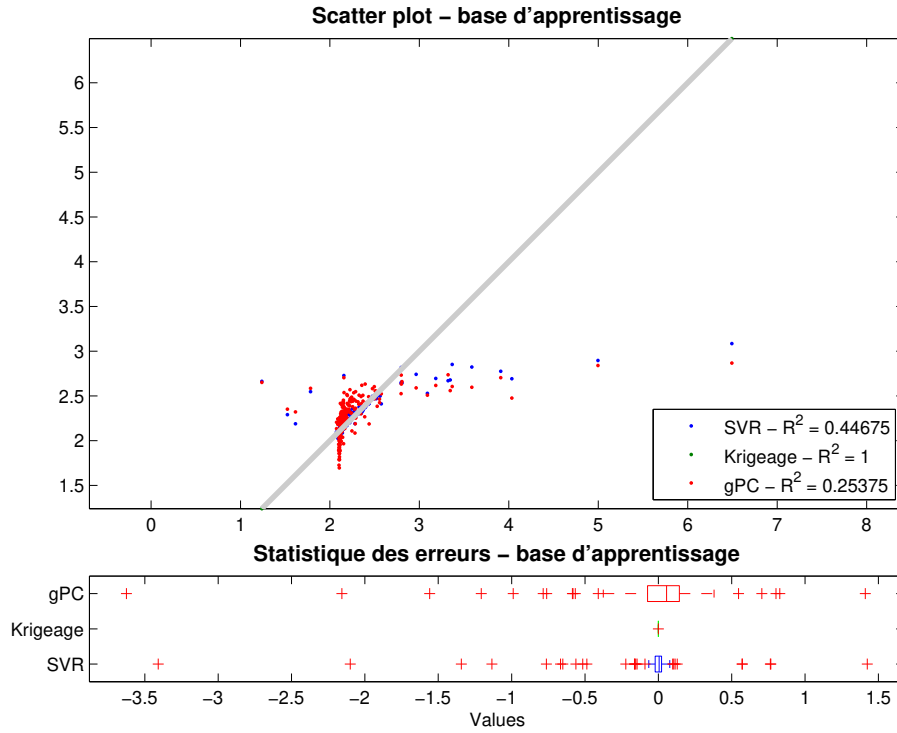


FIG. 5.11 – Analyse des erreurs de la base d'apprentissage (250 points) - Variable de sortie 3 - p106.

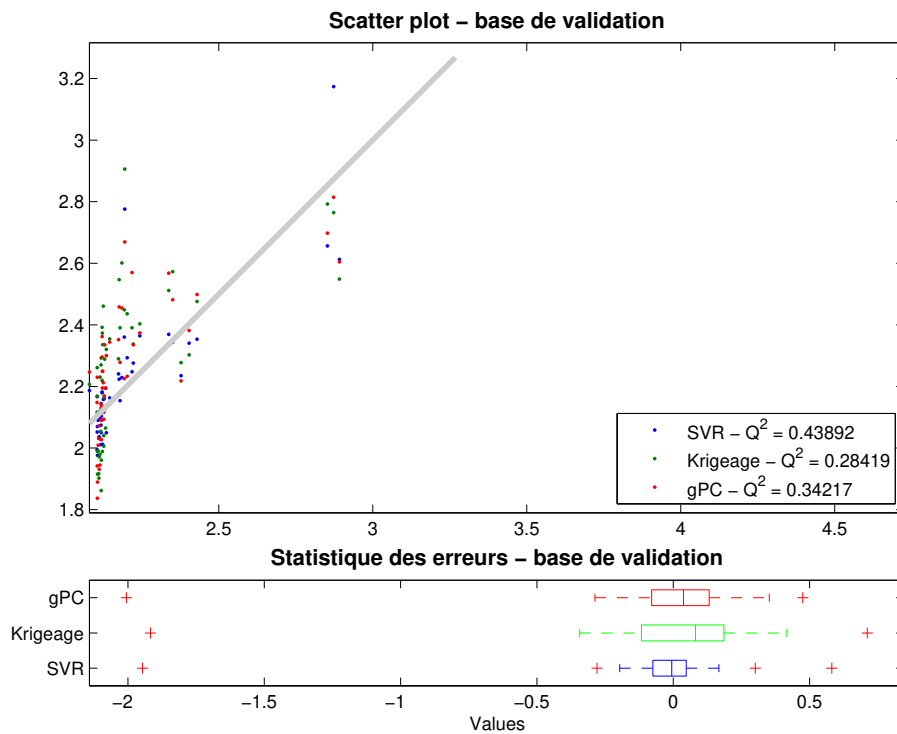


FIG. 5.12 – Analyse des erreurs de la base de validation (50 points) - Variable de sortie 3 - p106.

5.4 Variable de sortie 4 - $p2 - 76$

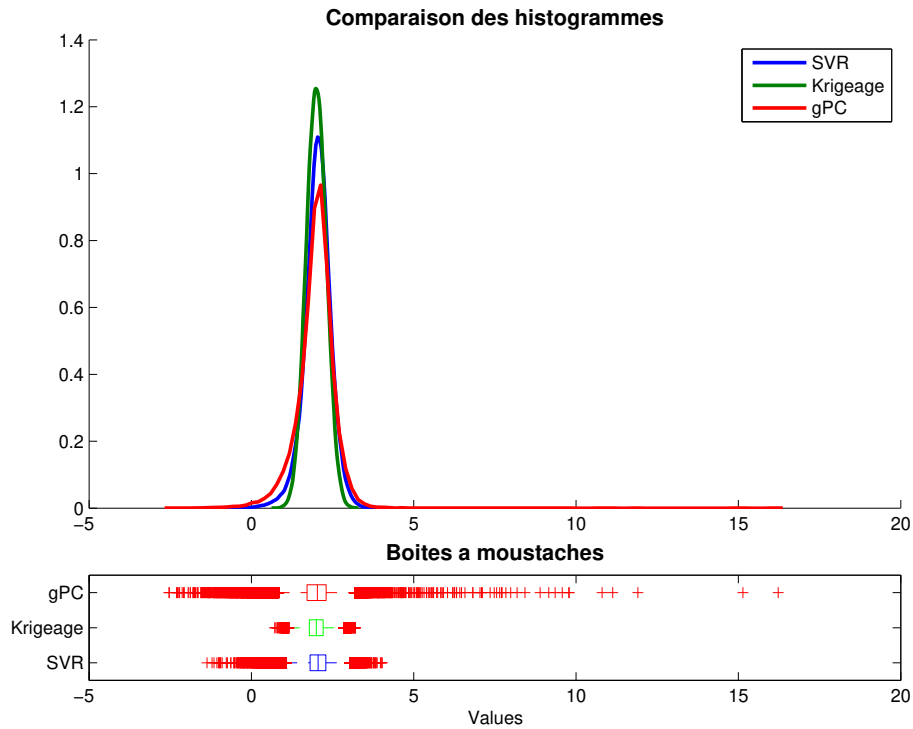


FIG. 5.13 - Comparaison des densités empiriques obtenues par simulation sur les métamodèles obtenus - Variable de sortie 4 - $p2 - 76$.

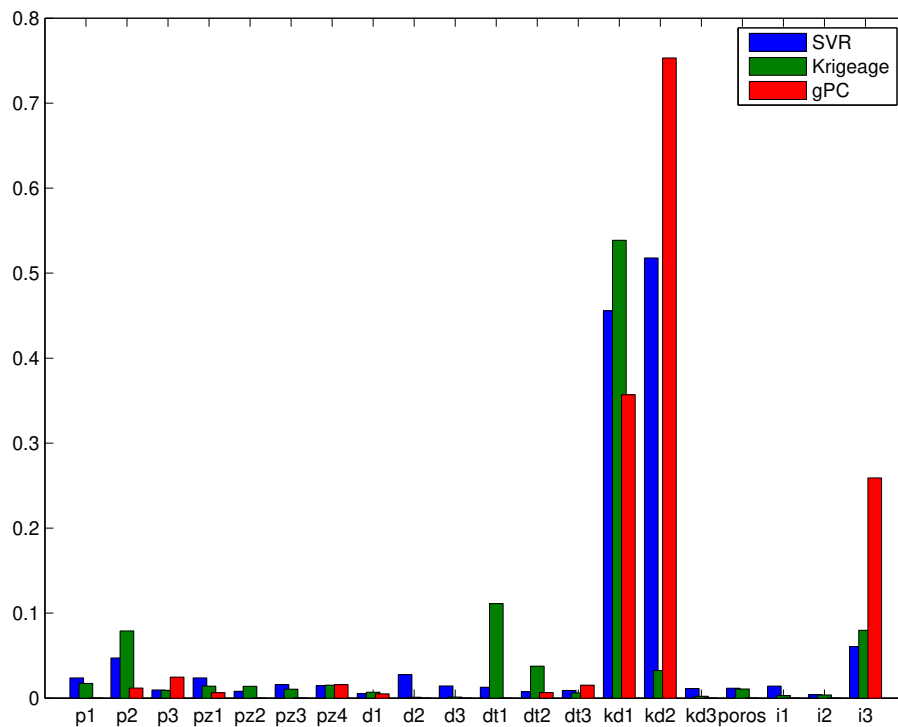


FIG. 5.14 - Comparaison des indices de Sobol' totaux mesurés sur les métamodèles obtenus - Variable de sortie 4 - $p2 - 76$.

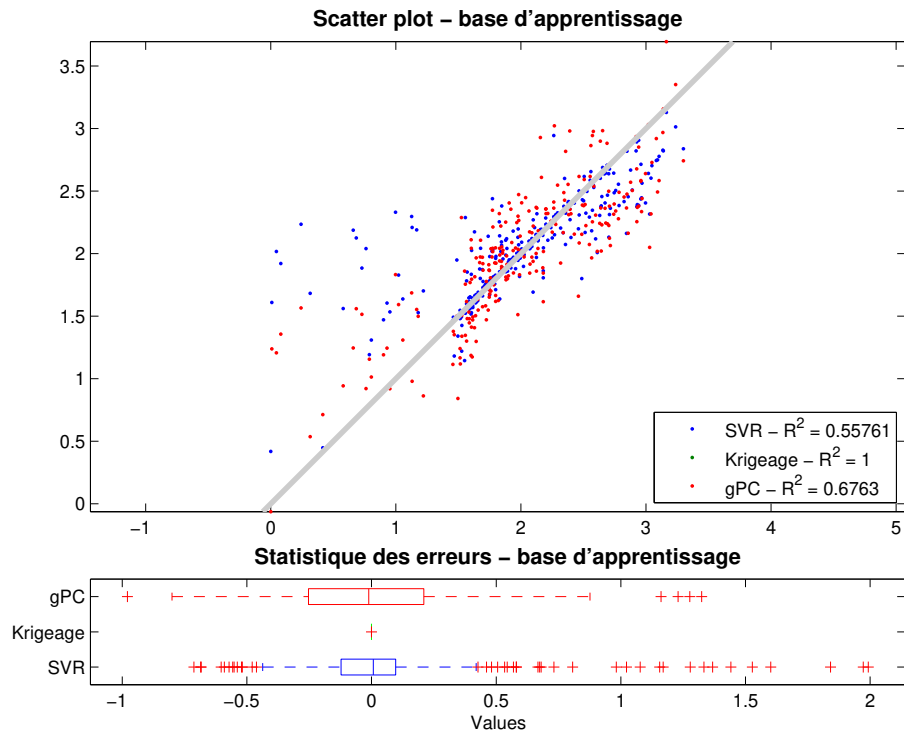


FIG. 5.15 – Analyse des erreurs de la base d'apprentissage (250 points) - Variable de sortie 4 - $p2 - 76$.

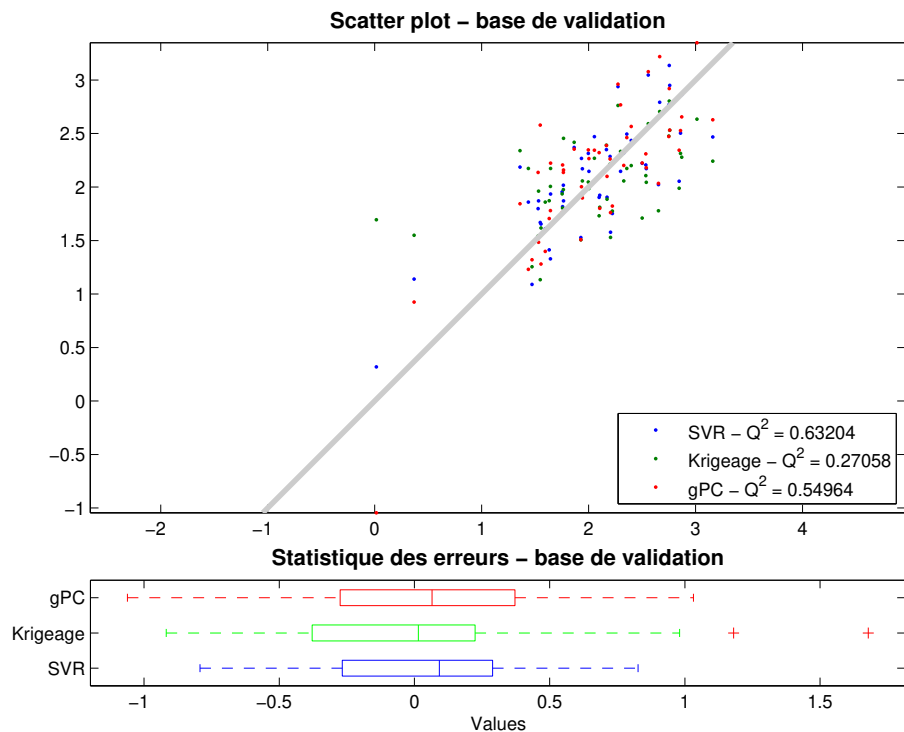


FIG. 5.16 – Analyse des erreurs de la base de validation (50 points) - Variable de sortie 4 - $p2 - 76$.

5.5 Variable de sortie 5 - p29K

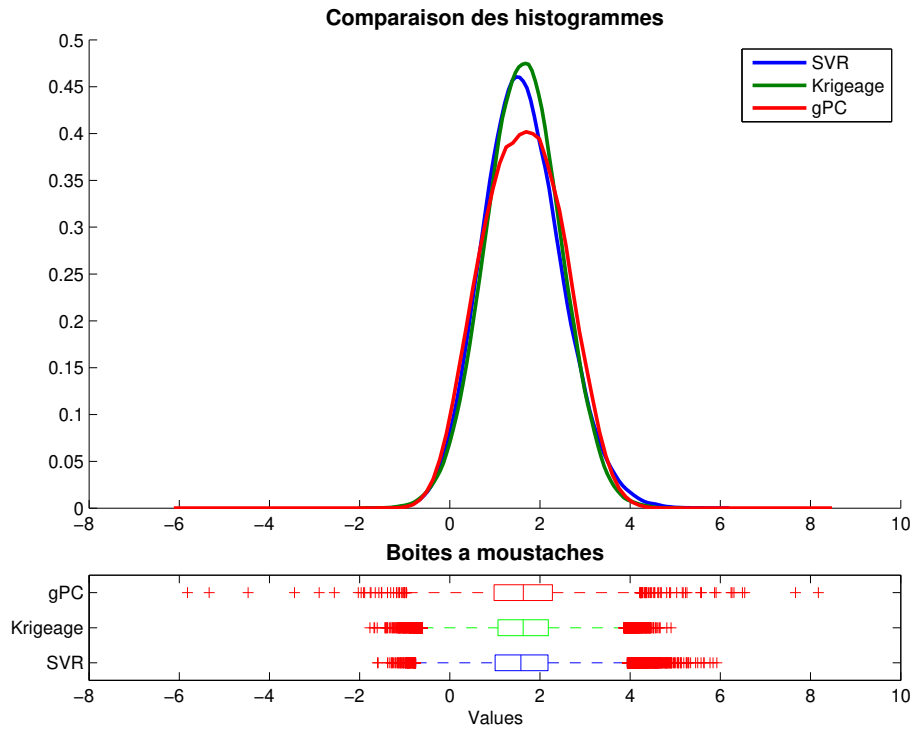


FIG. 5.17 - Comparaison des densités empiriques obtenues par simulation sur les métamodèles obtenus - Variable de sortie 5 - p29K.

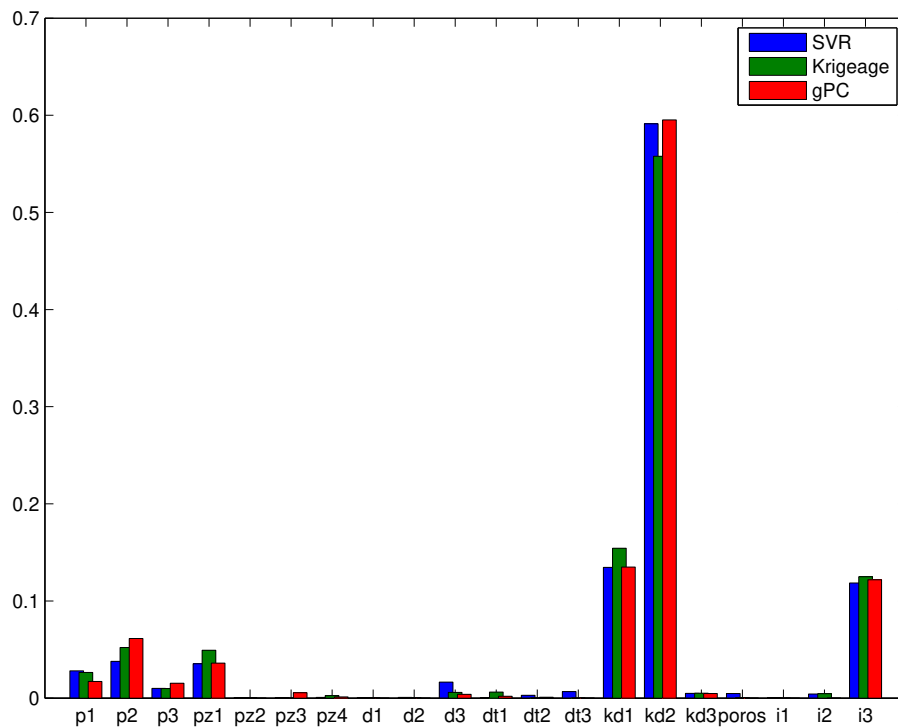


FIG. 5.18 - Comparaison des indices de Sobol' totaux mesurés sur les métamodèles obtenus - Variable de sortie 5 - p29K.

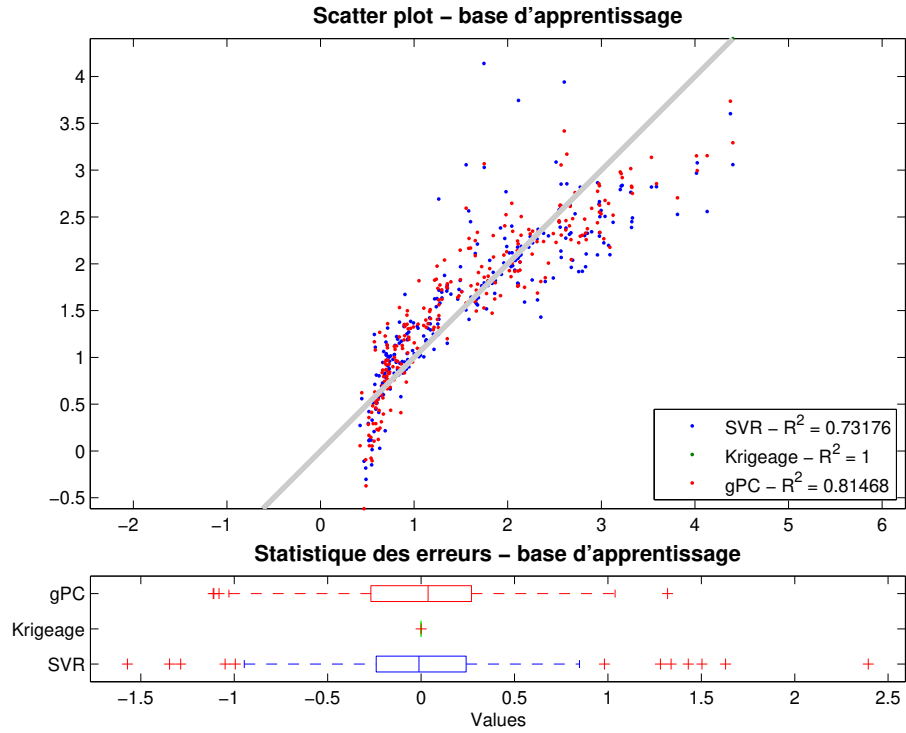


FIG. 5.19 – Analyse des erreurs de la base d'apprentissage (250 points) - Variable de sortie 5 - *p29K*.

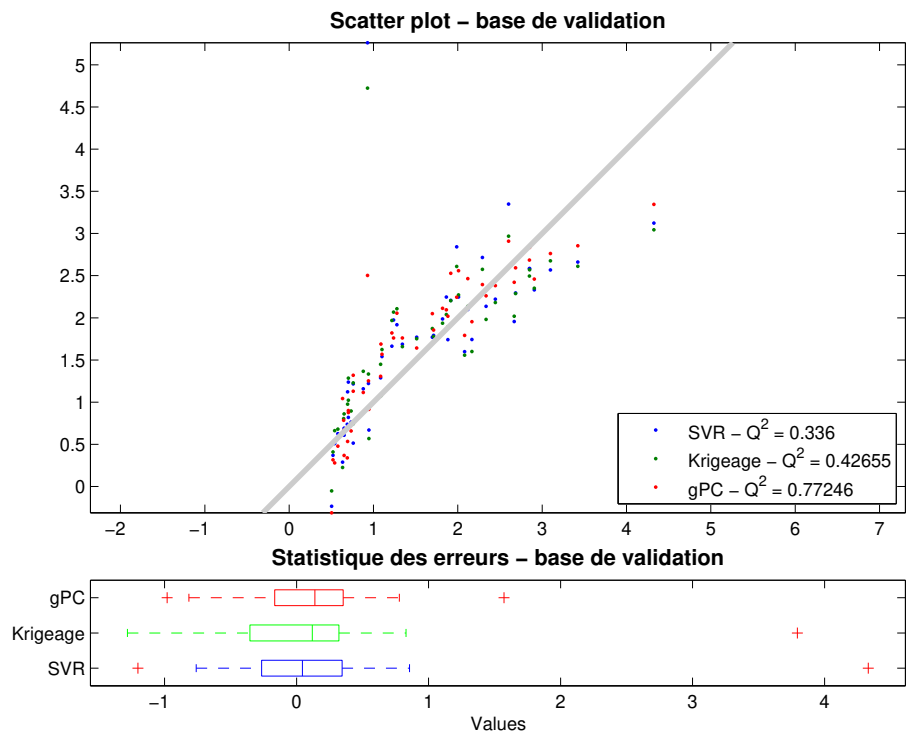


FIG. 5.20 – Analyse des erreurs de la base de validation (50 points) - Variable de sortie 5 - *p29K*.

5.6 Variable de sortie 6 - $p31K$

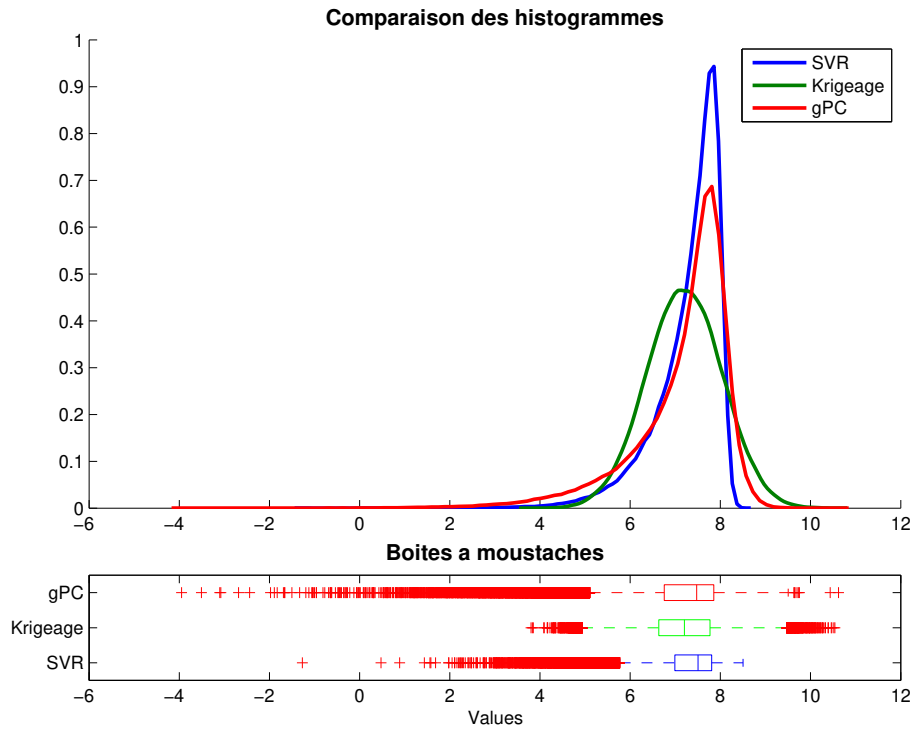


FIG. 5.21 - Comparaison des densités empiriques obtenues par simulation sur les métamodèles obtenus - Variable de sortie 6 - $p31K$.

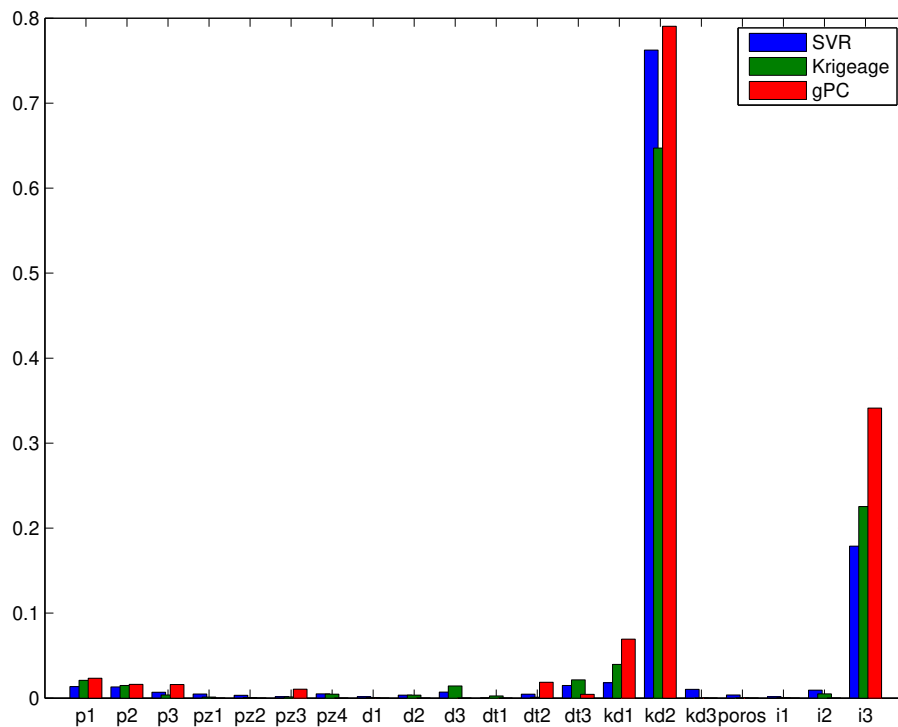


FIG. 5.22 - Comparaison des indices de Sobol' totaux mesurés sur les métamodèles obtenus - Variable de sortie 6 - $p31K$.

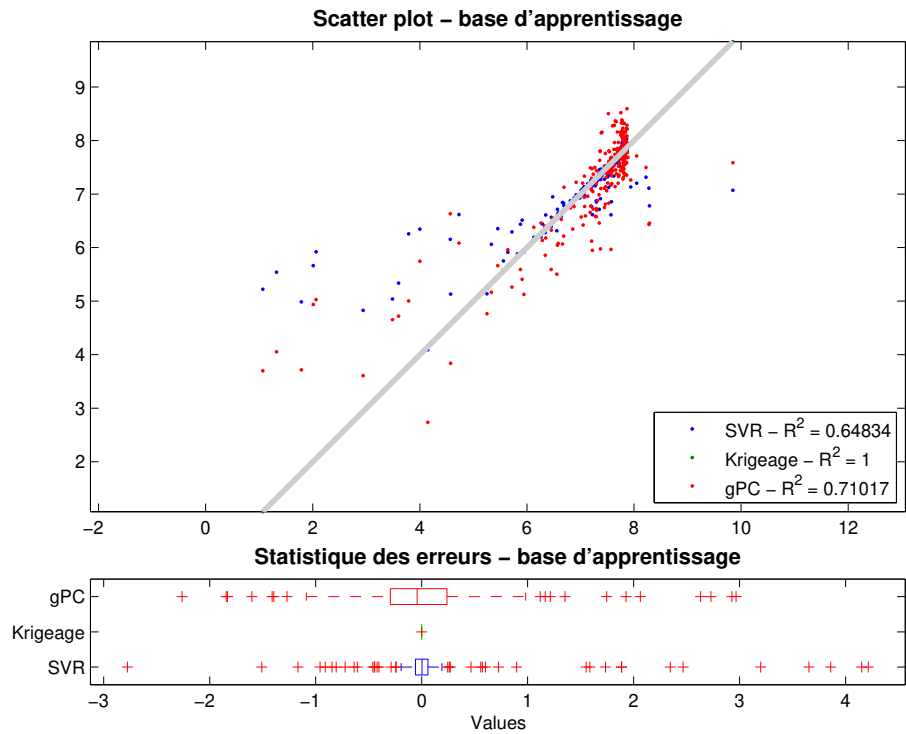


FIG. 5.23 – Analyse des erreurs de la base d'apprentissage (250 points) - Variable de sortie 6 - $p31K$.

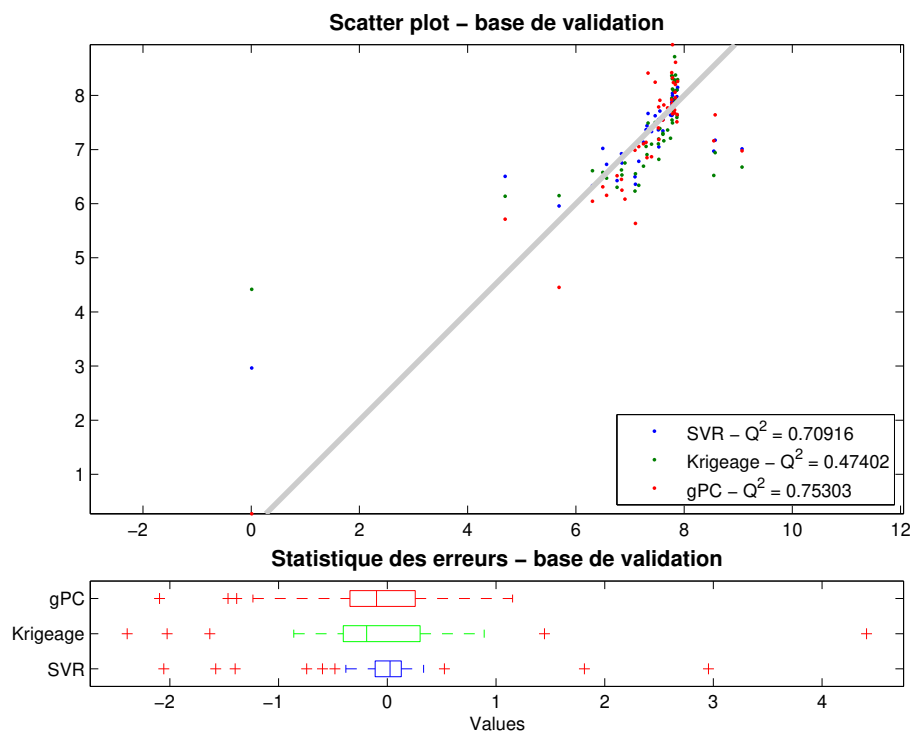


FIG. 5.24 – Analyse des erreurs de la base de validation (50 points) - Variable de sortie 6 - $p31K$.

5.7 Variable de sortie 7 - $p35K$

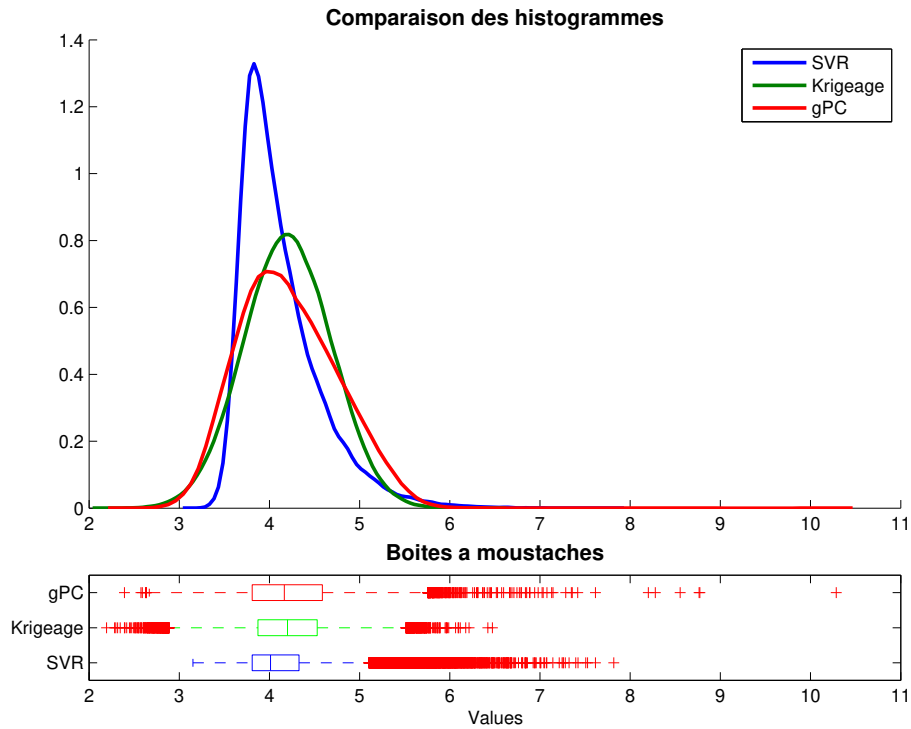


FIG. 5.25 - Comparaison des densités empiriques obtenues par simulation sur les métamodèles obtenus - Variable de sortie 7 - $p35K$.

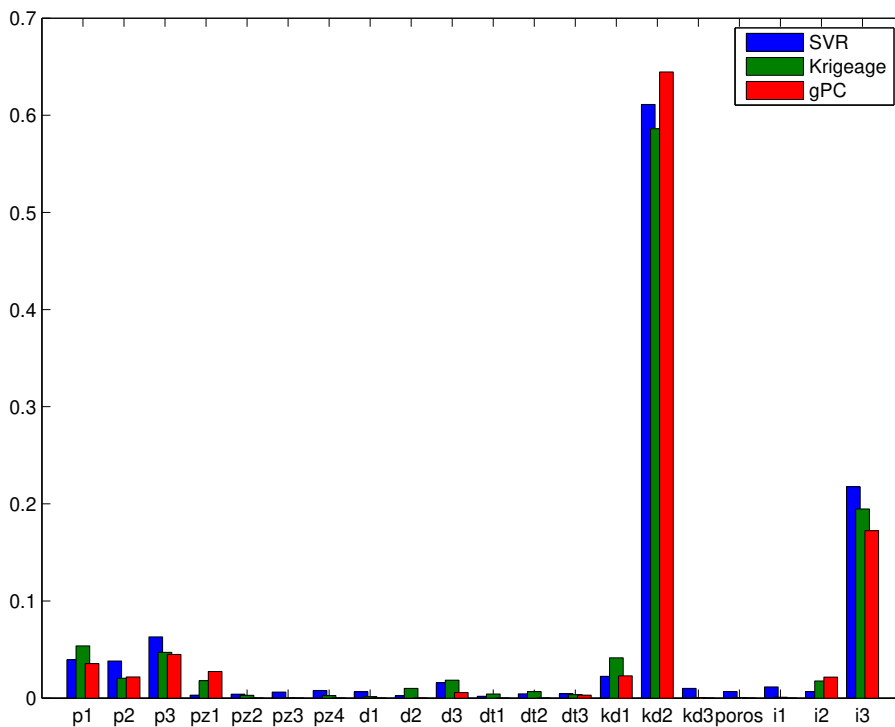


FIG. 5.26 - Comparaison des indices de Sobol' totaux mesurés sur les métamodèles obtenus - Variable de sortie 7 - $p35K$.

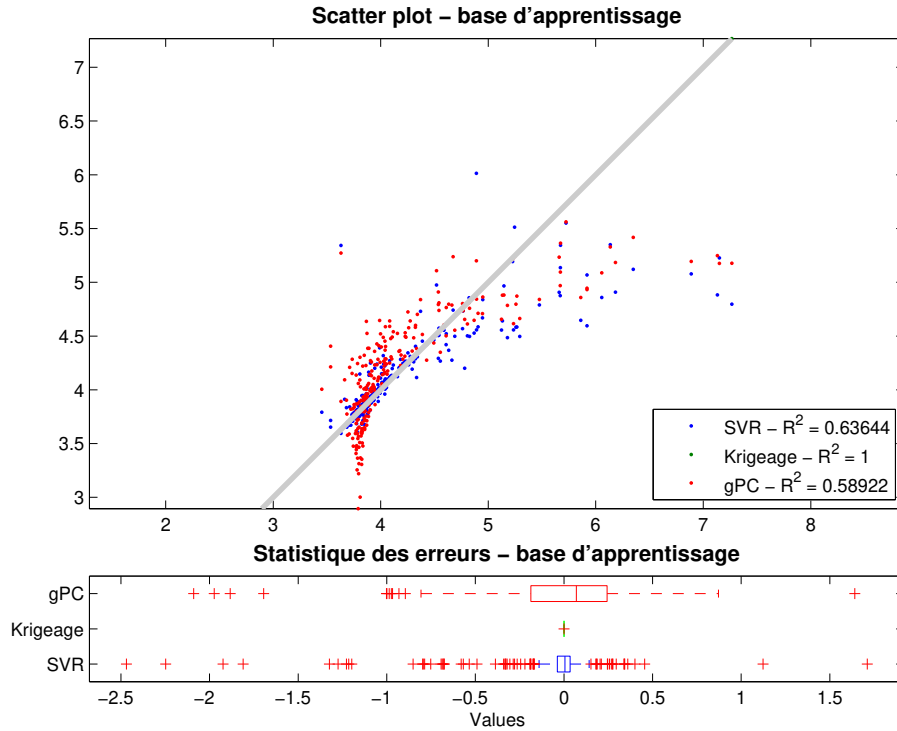


FIG. 5.27 – Analyse des erreurs de la base d'apprentissage (250 points) - Variable de sortie 7 - $p35K$.

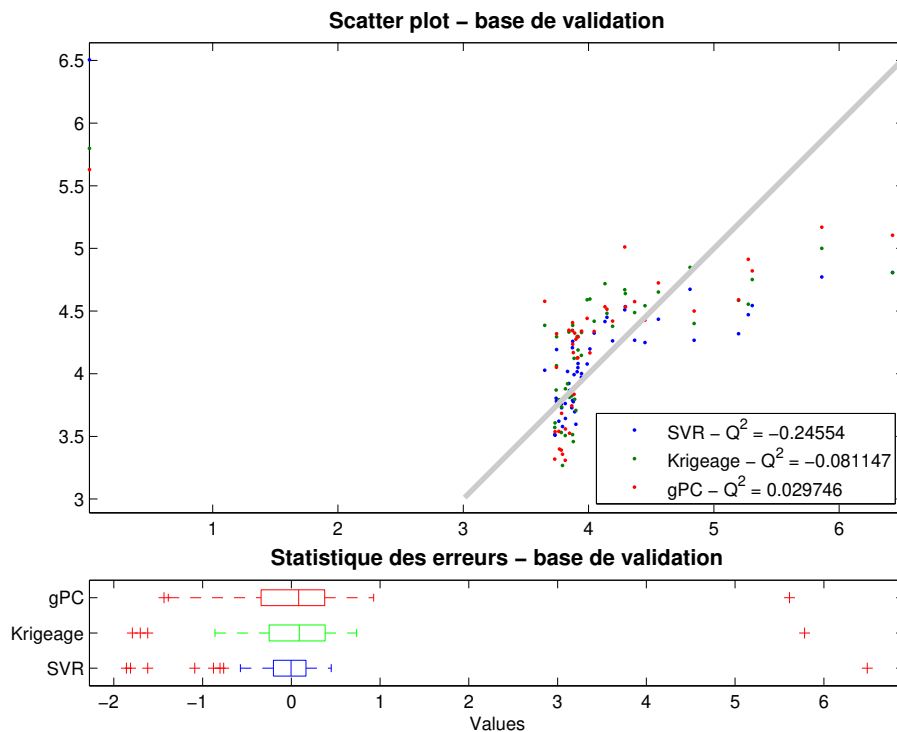


FIG. 5.28 – Analyse des erreurs de la base de validation (50 points) - Variable de sortie 7 - $p35K$.

5.8 Variable de sortie 8 - $p37K$

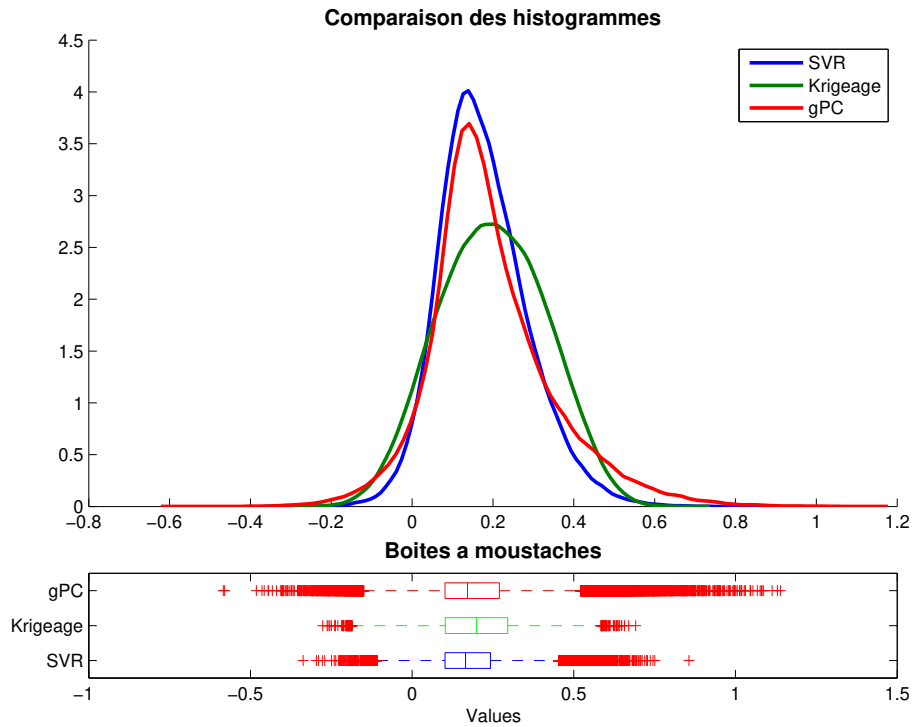


FIG. 5.29 - Comparaison des densités empiriques obtenues par simulation sur les métamodèles obtenus - Variable de sortie 8 - $p37K$.

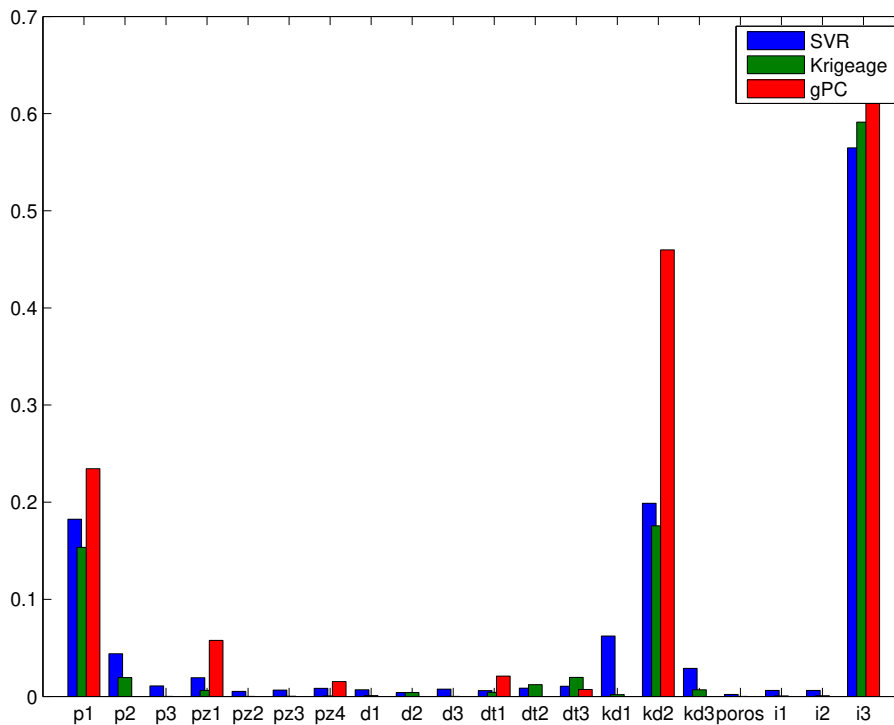


FIG. 5.30 - Comparaison des indices de Sobol' totaux mesurés sur les métamodèles obtenus - Variable de sortie 8 - $p37K$.

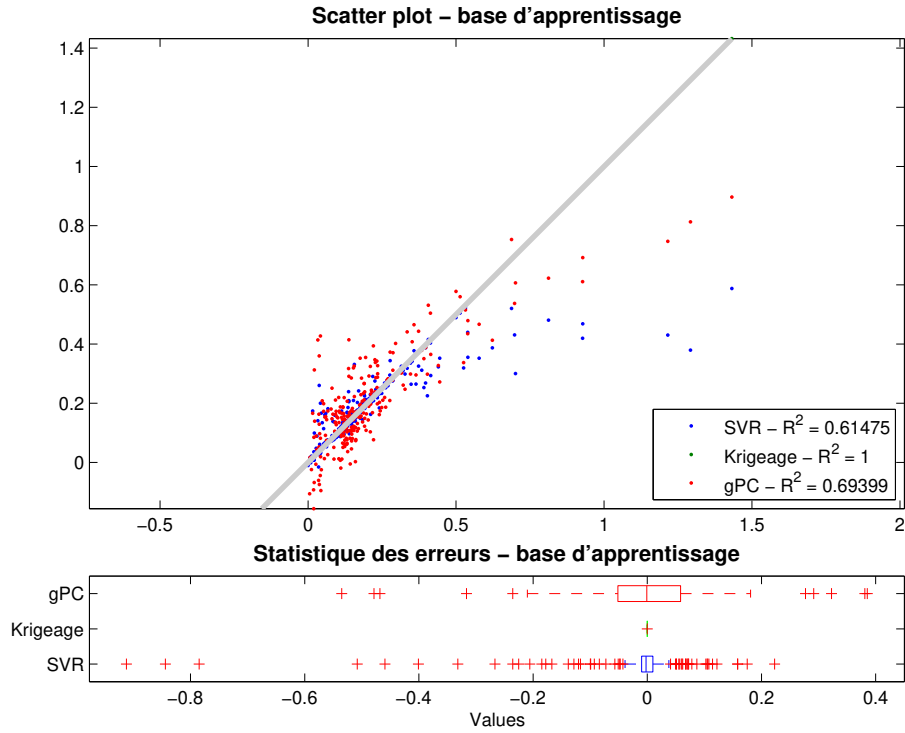


FIG. 5.31 – Analyse des erreurs de la base d'apprentissage (250 points) - Variable de sortie 8 - $p37K$.

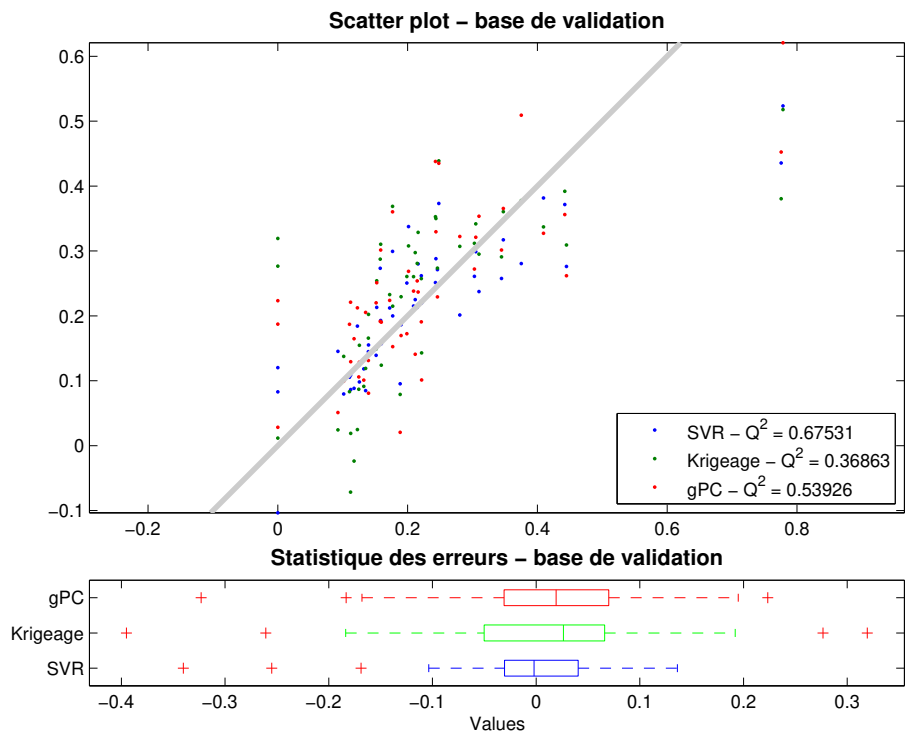


FIG. 5.32 – Analyse des erreurs de la base de validation (50 points) - Variable de sortie 8 - $p37K$.

5.9 Variable de sortie 9 - p38

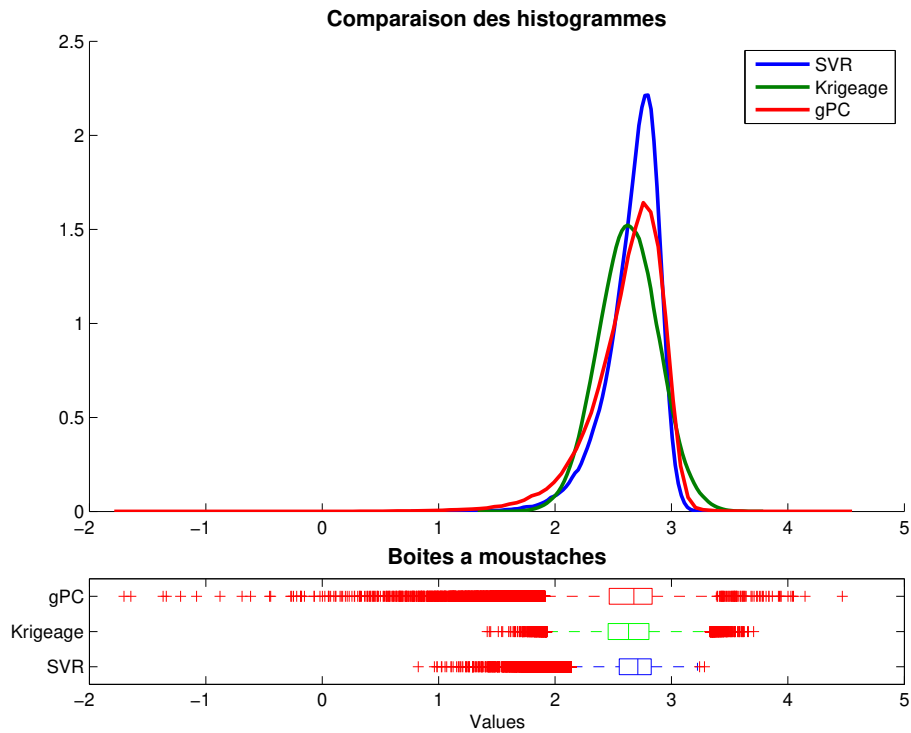


FIG. 5.33 - Comparaison des densités empiriques obtenues par simulation sur les métamodèles obtenus - Variable de sortie 9 - p38.

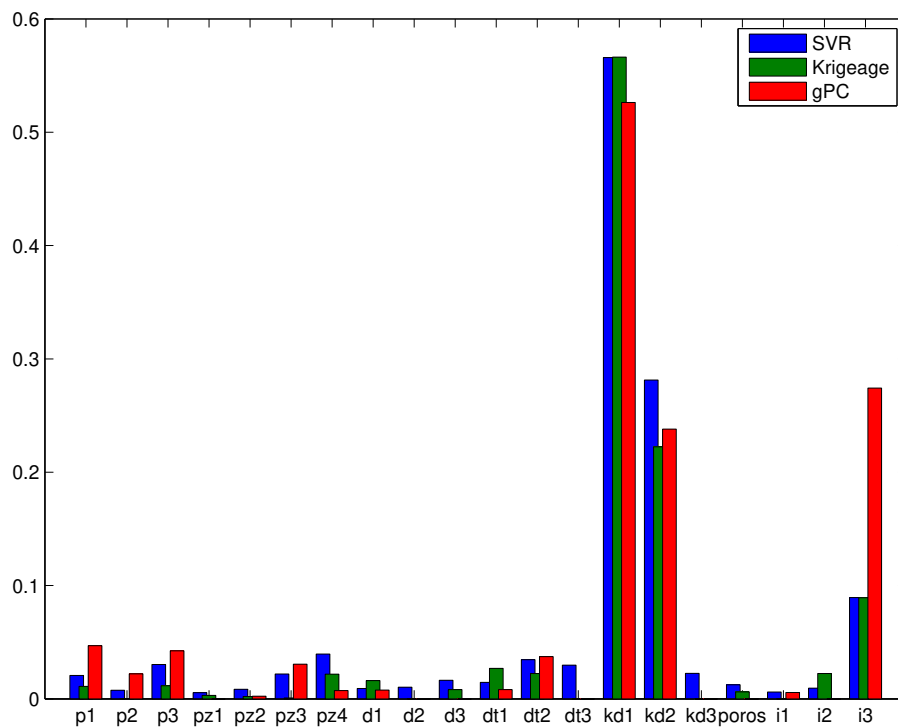


FIG. 5.34 - Comparaison des indices de Sobol' totaux mesurés sur les métamodèles obtenus - Variable de sortie 9 - p38.

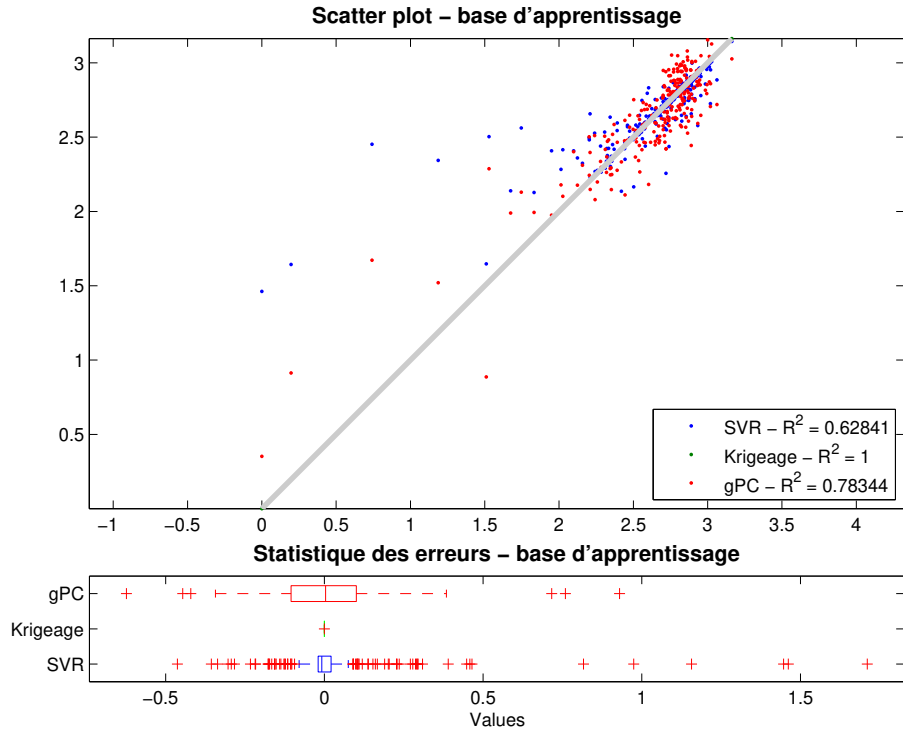


FIG. 5.35 – Analyse des erreurs de la base d'apprentissage (250 points) - Variable de sortie 9 - p38.

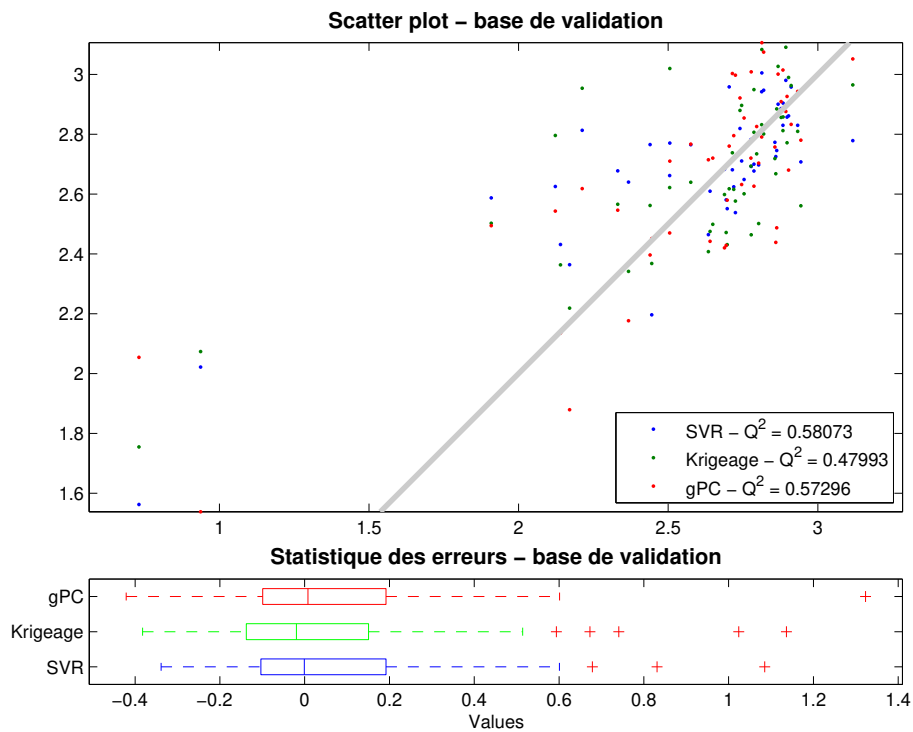


FIG. 5.36 – Analyse des erreurs de la base de validation (50 points) - Variable de sortie 9 - p38.

5.10 Variable de sortie 10 - p4b

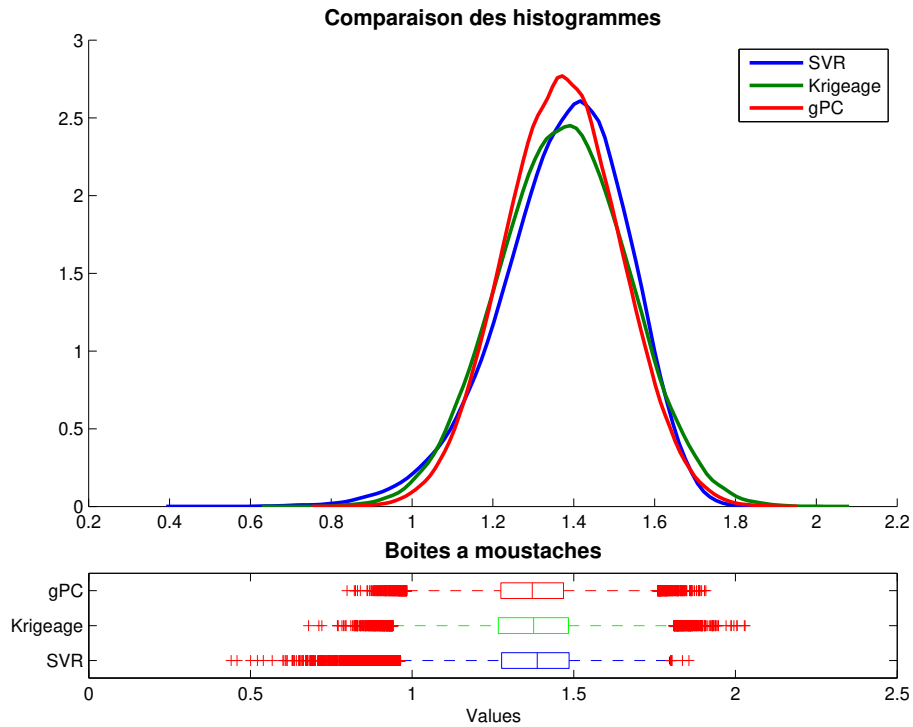


FIG. 5.37 - Comparaison des densités empiriques obtenues par simulation sur les métamodèles obtenus - Variable de sortie 10 - p4b.

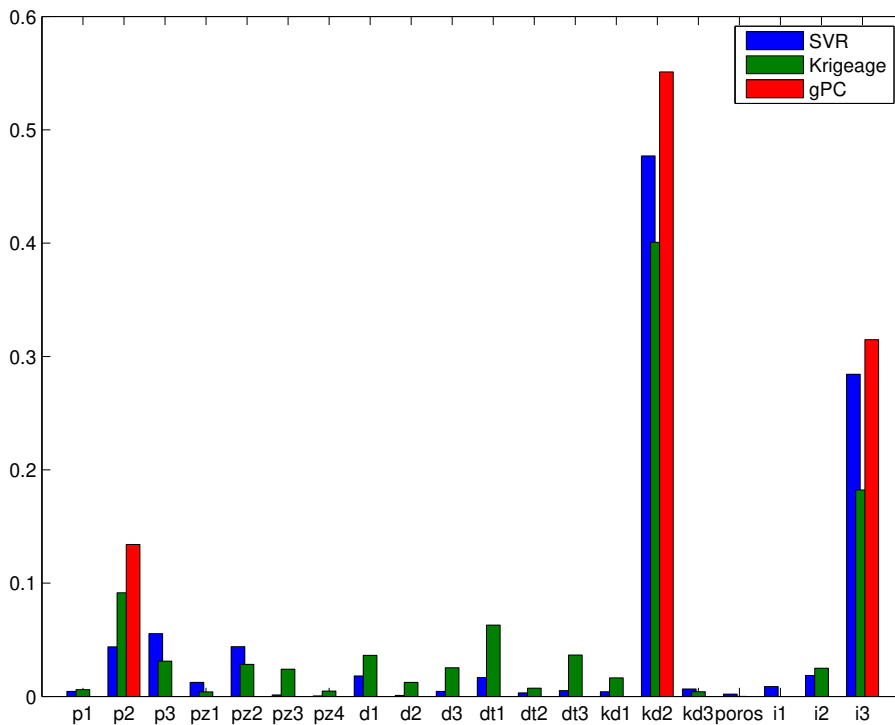


FIG. 5.38 - Comparaison des indices de Sobol' totaux mesurés sur les métamodèles obtenus - Variable de sortie 10 - p4b.

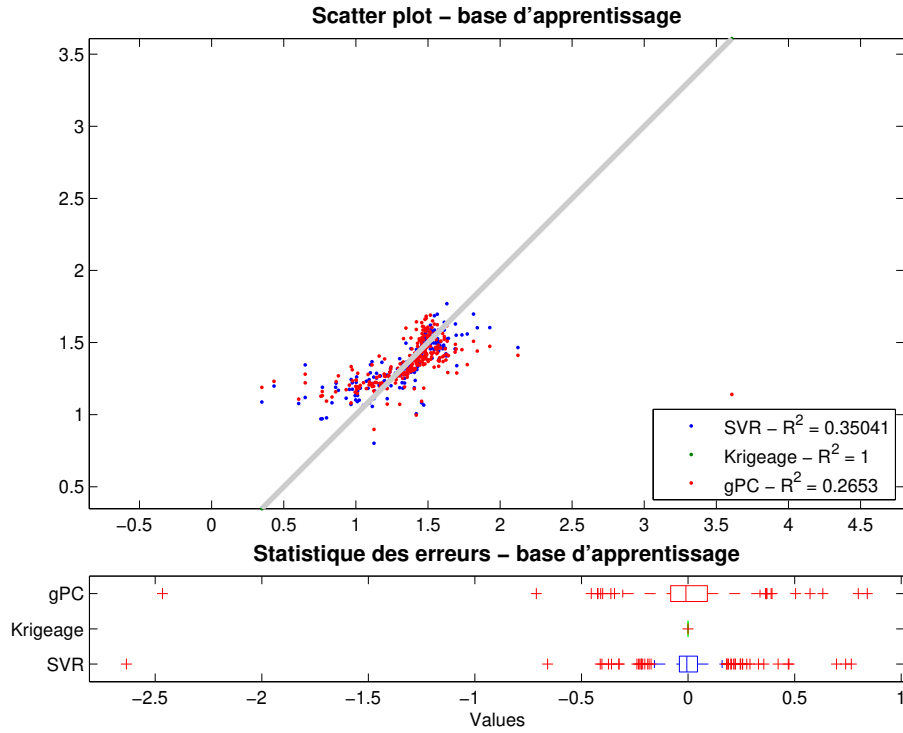


FIG. 5.39 – Analyse des erreurs de la base d'apprentissage (250 points) - Variable de sortie 10 - *p4b*.

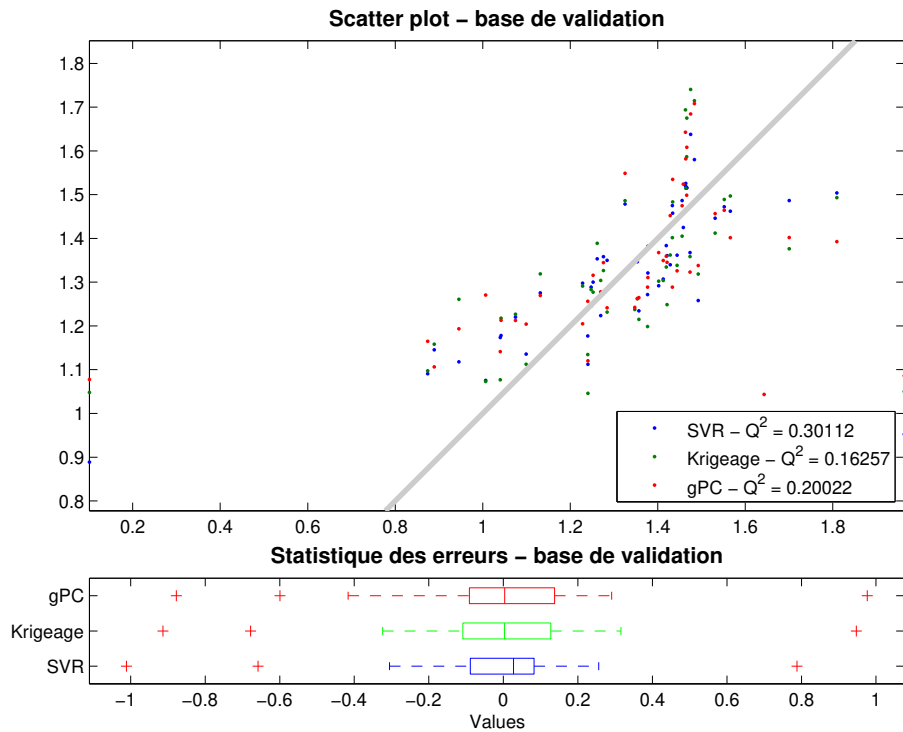


FIG. 5.40 – Analyse des erreurs de la base de validation (50 points) - Variable de sortie 10 - *p4b*.

6 Conclusions

Des figures précédentes, on peut tirer les conclusions suivantes.

Les trois approches donnent toujours des résultats cohérents sur la hiérarchisation des variables :

- Les coefficients de partage volumique ($kd1$ ou $kd2$, selon les sorties) sont les paramètres les plus importants. Vient ensuite l'infiltration $i3$. Comme énoncé en introduction, les variables $i1$, $i2$ et $i3$ étant liées, il est pertinent de regarder la somme des trois indices de sensibilité associés.
- Enfin, les perméabilités $p1$, $p2$ et $p3$ ont un poids plus faible variant de 0.065 à 0.25 selon les sorties considérées.
- La porosité et les variables décrivant la dispersité ne jouent aucun rôle.

En ce qui concerne les distributions des quantités d'intérêt, la forme et le mode sont en général similaires pour les différentes approches. Les queues de distribution sont parfois relativement distantes, surtout dans les cas où la distribution est asymétrique ($p35k$, $p37k$, $p38$).

En termes d'indices de sensibilité, les résultats sont d'autant plus cohérents entre les trois approches qu'il y a peu de paramètres importants e.g. $p104K$).

En terme de validation (Q2), SVR et gPC donnent alternativement les meilleurs résultats (6 fois pour SVR et 4 fois pour gPC). L'erreur de généralisation est donc plus grande pour le krigeage, qui inversement interpole les données d'apprentissage.

7 Références

- Berveiller, M. (2005). *Stochastic finite elements : intrusive and non intrusive methods for reliability analysis*. Ph. D. thesis, Université Blaise Pascal, Clermont-Ferrand.
- Blatman, G. (2009). *Adaptive stochastic finite element methods and metamodels for uncertainty propagation*. Ph. D. thesis, Université Blaise Pascal, Clermont-Ferrand.
- Blatman, G. et B. Sudret (2010). An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis. *Prob. Eng. Mech.*
- Efron, B., T. Hastie, I. Johnstone, et R. Tibshirani (2004). Least angle regression. *Annals of Statistics* 32, 407-499.
- Iooss, B. et A. Marrel (2008). Benchmark construction de métamodèles prédictifs.
- Lophaven, S., H. Nielsen, et J. Sondergaard (2002). Dace - a Matlab kriging toolbox, version 2.0. Technical Report IMM-TR-2002-12, Technical University of Denmark.
- Smola, A. et B. Schölkopf (2006). A tutorial on support vector regression. *Stat. Comput.* 14, 199-222.