

Exploration des capacités de Persalys

O. Braydi¹, A. Dumas¹, A. Aroles¹, J. Schueller¹, T. Yalamas¹, M. Baudin²

¹ Phimeca, Courmon d'Auvergne, braydi@phimeca.com

² EDF (R&D), Chatoux

Résumé — Persalys est une interface graphique dédiée au traitement des incertitudes et à la gestion des variabilités. Le logiciel est un outil à la frontière de la simulation numérique, des analyses probabilistes et des data sciences. Dans ce travail, nous présentons Persalys, ces fonctionnalités et son cycle de développement.

Mots clés — Persalys, Incertitudes, Variabilité, Optimisation.

1 Introduction

Persalys est développé dans le cadre d'un partenariat entre Phimeca et EDF R&D. Il s'agit d'une interface graphique dédiée au traitement des incertitudes et à la gestion des variabilités. Le logiciel est un outil à la frontière de la simulation numérique, des analyses probabilistes et des data sciences. Persalys permet de :

- Créer des modèles mathématiques : analytique de type scalaire ou champs, en langage Python ou couplage avec un modèle externe (éléments finis, ...), FMU.
- Analyser la variabilité des paramètres grâce à un grand nombre de méthodes et d'outils de visualisation : propagation d'incertitudes, optimisation, calcul de fiabilité, ...
- Analyser les données de mesures statistiquement, inférer des distributions de probabilités ou encore créer des métamodèles.

2 Partenariat et développement de Persalys

2.1 Partenariat

Le partenariat entre **EDF R&D** et **Phimeca** a débuté en 2015 avec un objectif clair : faciliter l'usage des méthodologies d'incertitude auprès des ingénieurs et chercheurs d'EDF R&D, en s'appuyant sur la puissance d'OpenTURNS [1]. Bien que le logiciel *PhimecaSoft*®, développé par Phimeca, répondait partiellement à ces besoins, il ne couvrait pas l'ensemble des attentes d'EDF R&D. C'est dans ce contexte que les deux entités ont décidé de s'associer pour concevoir une interface graphique dédiée, plus intuitive et accessible, entièrement basée sur OpenTURNS. Ce projet a donné naissance à **Persalys**, une solution *open source* pensée pour démocratiser l'analyse d'incertitude et rendre ces outils plus facilement exploitables dans les environnements industriels et de recherche.

2.2 Développement

Le logiciel Persalys est développé en langage C++, en se basant sur la librairie *Qt* pour la partie graphique et OpenTURNS pour le noyau de calcul. Il est disponible en version autonome ou bien intégré à la plateforme SALOME. Conçu comme un outil générique, il n'est pas limité à une application spécifique. Il permet l'analyse de données, prend en charge une grande variété de modèles, peut être couplé à des codes externes, et fonctionne en parallèle localement. L'interface graphique est disponible en anglais et en français, afin de répondre aux besoins d'une communauté d'utilisateurs internationales. Persalys est distribué sous licence **LGPL** (*Lesser General Public License*), ce qui permet une utilisation libre, y compris dans des contextes industriels, tout en garantissant l'ouverture du code source. Le calendrier de publication prévoit une nouvelle version deux fois par an, assurant ainsi une évolution régulière de l'outil

en fonction des besoins des utilisateurs et des avancées technologiques.

2.3 Communauté et accompagnement

La communauté autour de Persalys est active et engagée. Une journée annuelle des utilisateurs est organisée afin de favoriser les échanges, le partage d'expériences et la présentation des nouveautés. Le site officiel, accessible à l'adresse persalys.fr, fournit des ressources utiles, tandis qu'un forum communautaire est accessible à l'adresse persalys.discourse.group pour échanger entre utilisateurs. Bien que Persalys soit *open source*, Phimeca propose un support à l'utilisation du logiciel, et propose des formations ou accompagnement sur projets, ainsi que le développement de versions personnalisées adaptées aux problématiques des utilisateurs. Des fonctionnalités spécifiques ont également été développées pour des acteurs comme Thales ou Naval Group, à des coûts souvent inférieurs à ceux des logiciels dédiés.

3 Fonctionnalités

Persalys propose plusieurs fonctionnalités, selon le type de modèle défini. Les types de modèles proposés dans Persalys sont classifiés sous trois catégories :

- Fonctions vectorielles :
 - Modèle Symbolique : créer un modèle mathématique avec des fonctions symboliques.
 - Modèle Python : créer un modèle plus complexe en langage Python.
 - Modèle FMI : importer un modèle système au format Functional Mock-up Interface.
 - Modèle de couplage : réaliser un couplage avec un code de calcul externe.
- Fonctions de champs :
 - Modèle de champs symbolique : créer un modèle mathématique simple dont la ou les sorties est un champ 1-D.
 - Modèle de champs Python : créer un modèle Python ayant pour sortie des champs 1-D.
- Échantillons de données :
 - Modèle de données : importer un jeu de mesures pour l'analyser.
 - Modèle de champs de données : importer un jeu de champs de mesures pour l'analyser.

la figure 1 donne une vues globale des modèles disponibles dans Persalys.

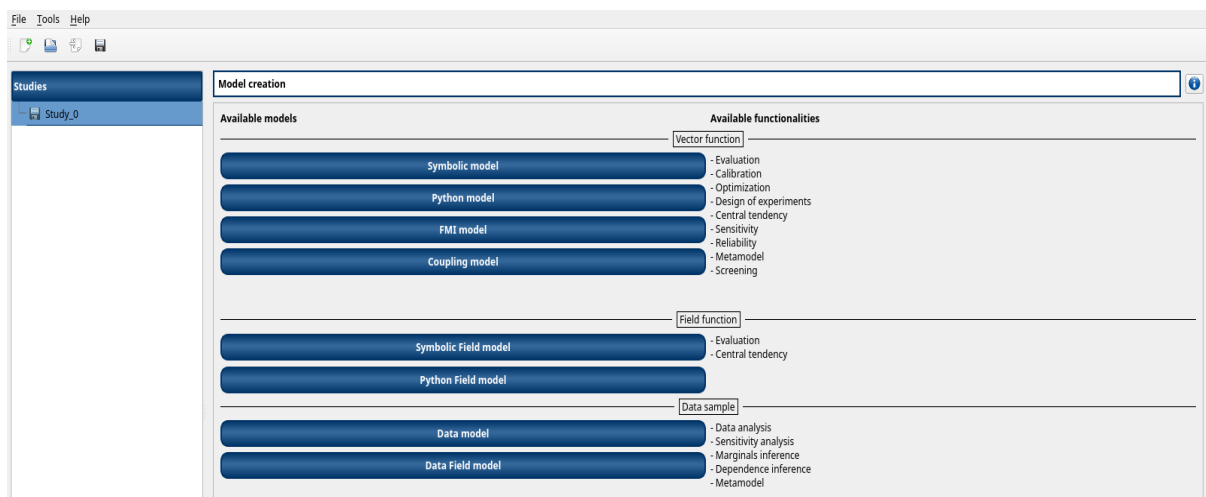


FIGURE 1 – Les modèles dans Persalys

3.1 Définition du modèle probabiliste

La définition des modèles probabiliste est disponible pour les modèles de types fonctions vectorielles ou fonctions de champs. La définition du modèle probabiliste permet de définir les lois de probabilités pour chacun des paramètres, voir figure 2. Un grand nombre de lois sont disponibles, à la fois continues

et discrètes. Il est possible également de définir une corrélation entre les variables à l'aide des copules, ou encore d'importer des résultats d'inférence effectuée sur un jeu de données.

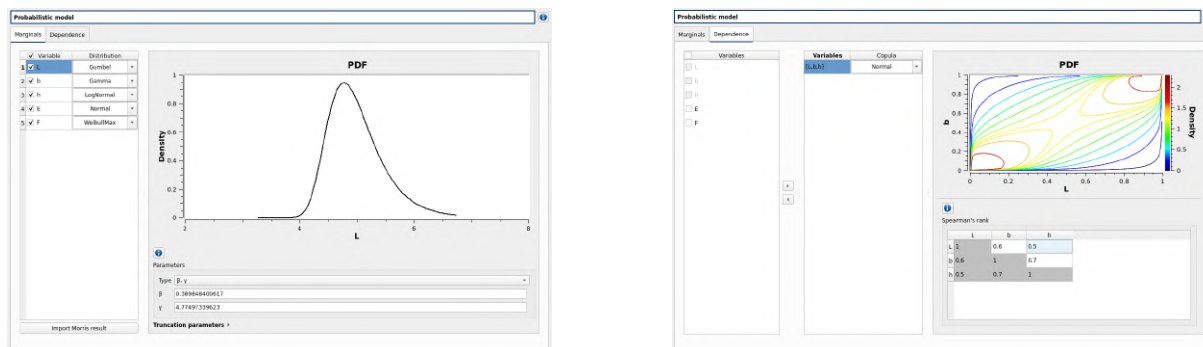


FIGURE 2 – Définition du modèle probabiliste

3.2 Construction d'un plan d'expériences

Persalys permet également d'établir des plans d'expériences, voir figure 3. Ces plans permettent de visualiser le comportement du modèle via des graphes de dispersion, mais également de construire un métamodèle qui pourra remplacer un modèle physique coûteux. Plusieurs types de plans d'expériences sont proposés :

- Plan factoriel complet
- Plan probabiliste (hypercube latin (LHS) [3], optimisé [4] ou non, Monte-Carlo ou quasi Monte-Carlo)
- Plan importé depuis un fichier

En plus d'un résumé des grandeurs statistiques de l'échantillon, Persalys propose de nombreux affichages (courbe des densités de probabilités ou fonction de répartition, graphique boîtes à moustaches (*boxplot*), graphique en coordonnées parallèles, graphes de dispersion 2D, tableau de corrélation).

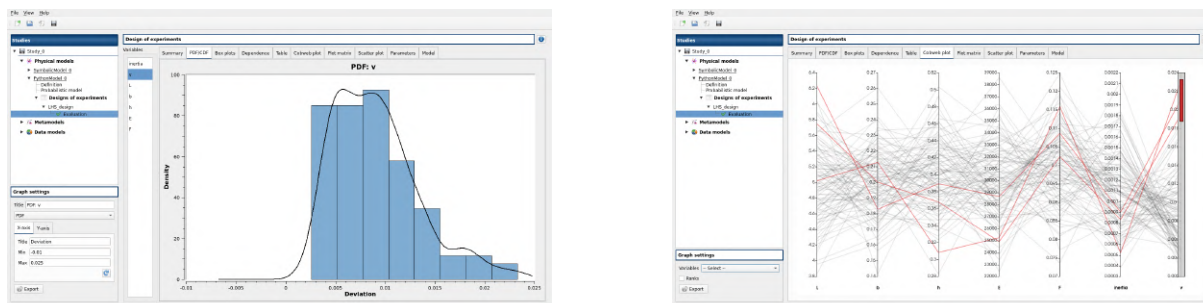


FIGURE 3 – Exemple de visualisation de plan d'expériences

3.3 Criblage

Le criblage est une fonctionnalité de Persalys permettant d'explorer rapidement le comportement des sorties d'un code. Cette fonctionnalité est disponible pour les modèle de type fonctions vectorielles. Cette analyse se fait grâce à la méthode de Morris [2], voir figure 4. Elle permet de distinguer le type d'évolution des sorties (linéarités, monotonie, avec ou sans interactions) par rapport aux paramètres d'entrée. De plus, les paramètres d'entrée jugés non influents sur les sorties peuvent être considérés constants, réduisant le nombre de paramètres à considérer dans les analyses probabilistes.

3.4 Métamodèle

Un métamodèle est construit à partir d'un plan d'expériences construit dans Persalys ou à partir d'un modèle de données. Le but du métamodèle est de remplacer le modèle initial, en particulier lorsque celui-ci est très coûteux. Persalys propose trois méthodes :

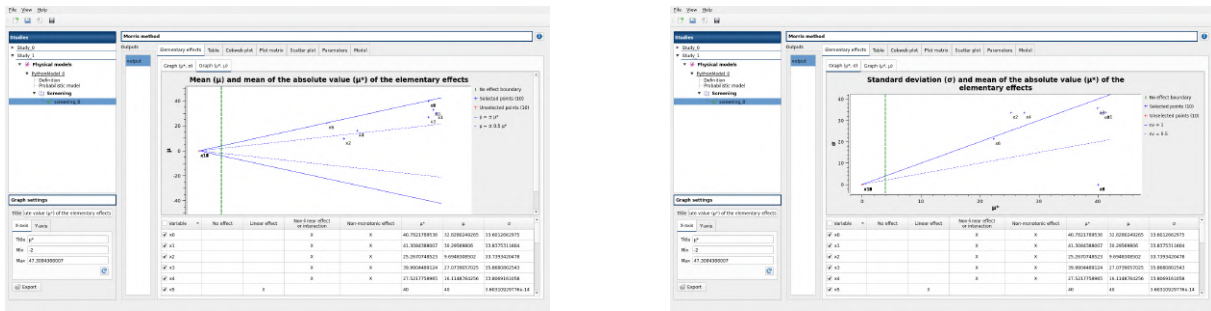


FIGURE 4 – Exemple de résultats du criblage

- Régression linéaire
- Chaos polynomial
- Krigage, ou processus gaussien

Les méthodes de validation croisées implémentées dans Persalys permettent de valider la construction du métamodèle. Dans le cas du chaos polynomial, les indices de Sobol [5] sont également fournis par défaut. Après avoir créé et validé le métamodèle, il est possible de le convertir en modèle physique. Toutes les fonctionnalités de Persalys seront applicables sur celui-ci. Un exemple est donné dans la figure 5.

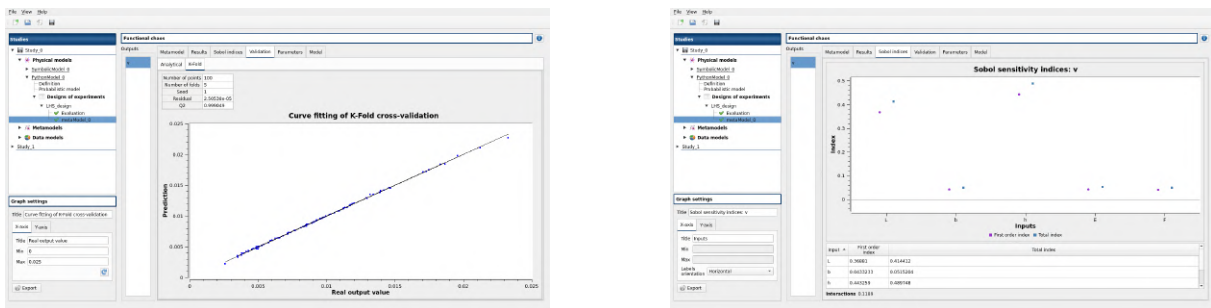


FIGURE 5 – Exemple de validation de métamodèle et des indices de Sobol (chaos polynomial).

3.5 Tendances centrale

Une fois le modèle probabiliste défini, l'analyse en tendance centrale permet de calculer les principales statistiques des grandeurs de sorties. Persalys propose deux méthodes pour effectuer une analyse en tendance centrale :

- Décomposition de Taylor
- Simulation de Monte-Carlo

La Simulation de Monte-Carlo permet d'obtenir une plus grande variété de résultats, similaires à ceux obtenus à partir d'un plan d'expériences.

3.6 Analyse de sensibilité

Persalys permet de réaliser des analyses de sensibilité afin de déterminer quelles sont les paramètres d'entrée influents sur les sorties.

Deux types d'indices sont proposés :

- Indice de Sobol : ce sont des indices définis à partir de l'ANOVA du modèle [5]. Dans le cas où plusieurs sorties sont choisies, les indices agrégés sont également calculés.
- Indice SRC (Standard Regression Coefficient) : ces indices sont obtenus en approximant le modèle réel par un modèle linéaire de degré 1.

Ces indices (compris entre 0 et 1) permettent de déterminer quelles entrées ont ou non de l'influence sur la sortie. Plus l'indice se rapproche de 1, plus il est influent. Dans le cas des indices SRC, des indices signés (positif ou négatif) permettent de déterminer l'effet croissant ou décroissant d'une variable d'entrée sur la sortie. Un exemple est donné dans la figure 6.

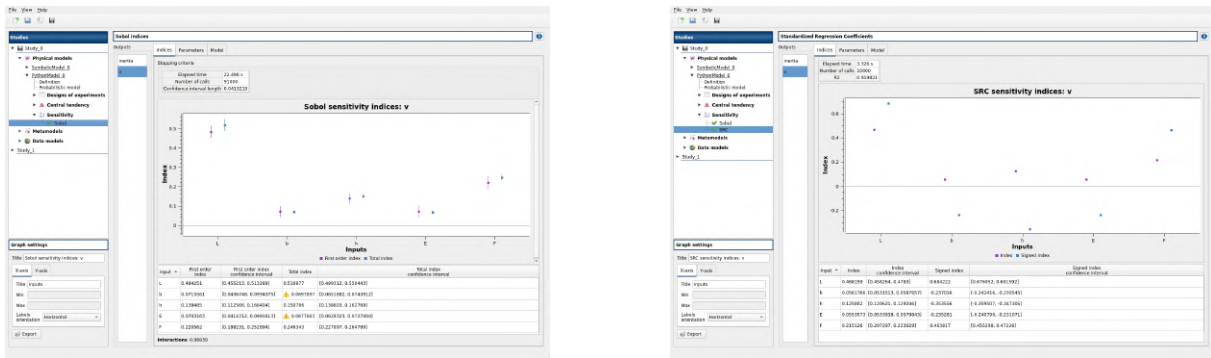


FIGURE 6 – Exemple des résultats des indices de Sobol et des indices SRC

3.7 Fiabilité

Persalys permet de calculer la fiabilité des structures ou des systèmes, le but est de calculer une probabilité de défaillance ou probabilité de dépassement de seuil d'une grandeur d'intérêt. Pour cela, il vous suffit de définir un état-limite spécifiant si le système est défaillant lorsqu'il est supérieur ou inférieur à une valeur critique. Persalys propose 4 méthodes :

- Simulation de Monte Carlo
- Approximation FORM /SORM (First and Second Order Reliability Methods) [6]
- Simulation par tirages d'importance (Importance Sampling) après l'approximation FORM [7].

En plus de la probabilité de défaillance, les facteurs d'importance ou les sensibilités des paramètres de la distribution à la probabilité sont également fournis. Un exemple est donné dans la figure 7.

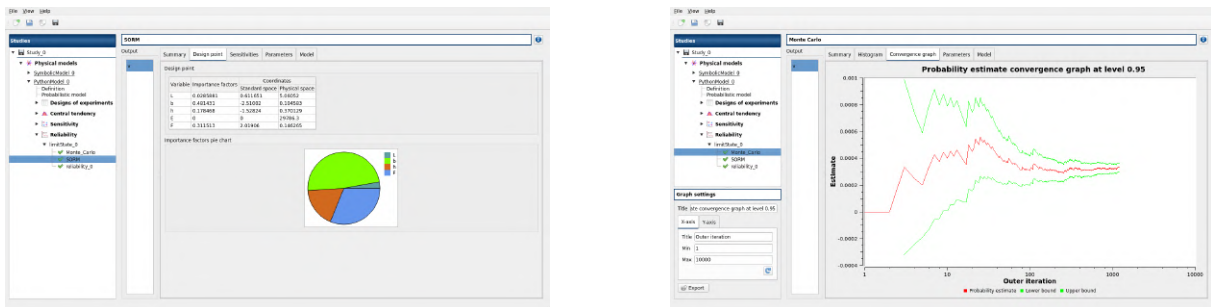


FIGURE 7 – Exemple des résultats des méthodes de calcul de fiabilité

3.8 Optimisation

Persalys permet d'optimiser les sorties des modèles, deux types d'optimisation sont proposés :

- Optimisation mono-objective
- Optimisation multi-objective

De nombreux algorithmes sont disponibles : méthodes à gradients, sans gradients, évolutionnaires, locaux, globaux. Les résultats fournissent les valeurs optimisées ainsi que le graphique de convergence de la fonction objectif. Pour l'optimisation multi-objectif, les différents niveaux du front de Pareto sont également fournis. Un exemple est donné dans la figure 8.

3.9 Calage des modèles paramétrés

Le calage a pour but d'optimiser des paramètres d'entrée inconnus d'un modèle à partir de données de mesures ou d'observations (entrée et sorties). Pour effectuer un calage, il faut d'abord importer une plage de données (fichier texte) contenant au moins une entrée et une sortie. Le but sera ensuite de caler des paramètres d'entrée du modèle pour les sorties calculées et importées correspondent le mieux possibles, voir l'exemple figure 9. Il existe deux types de méthodes pour effectuer un calage et chacune faisant une hypothèse de linéarité ou non sur le modèle physique :

- Méthode des moindres carrés (linéaire ou non linéaire)

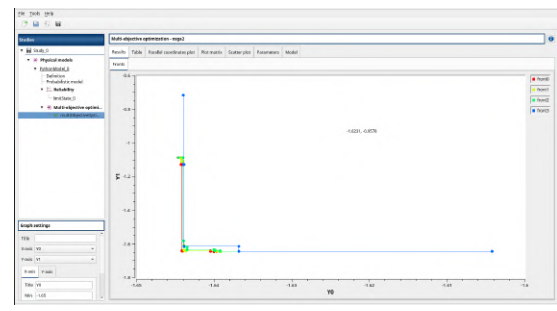
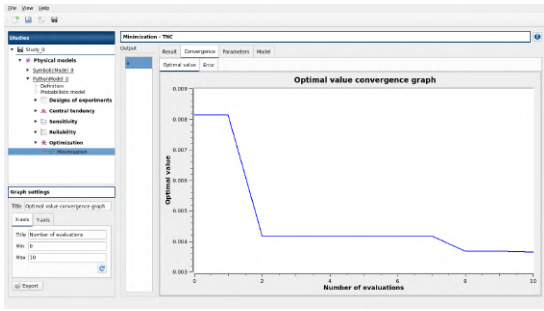


FIGURE 8 – Exemple des résultats d'optimisation

— Méthode bayésienne avec hypothèse gaussienne (linéaire ou non linéaire)

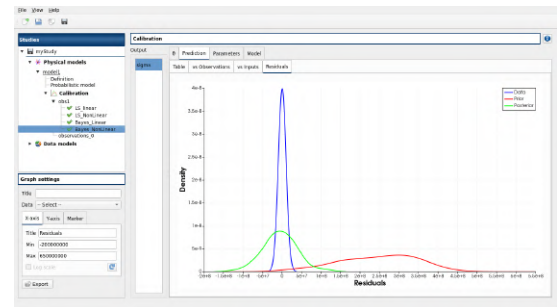
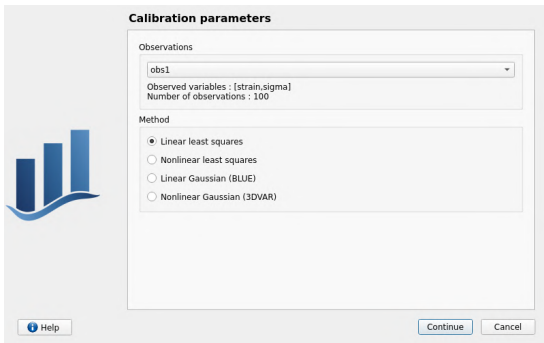


FIGURE 9 – Menu et résultats du calage

3.10 Fonctionnalité relative à l'analyse de données

Les fonctionnalités citées ci-dessus sont disponibles pour les modèles de type fonctions vectorielles. Avec un jeu de données importé dans Persalys avec un fichier CSV, on peut appliquer les fonctionnalités suivantes :

- Réaliser une analyse statistique des données
- Inférer d'une distribution paramétrique ou d'une copule (*i.e.* corrélation) sur les données
- Réaliser une analyse de sensibilité de Sobol du premier ordre.
- Calculer des quantiles pour différents niveaux de probabilités.
- Construire un métamodèle.

Analyse de données

Cette fonctionnalité permet d'analyser les données, et fournir des résultats similaires à ceux des plans d'expériences évalués. Des exemples des résultats fournis sont affichés dans la figure 10.

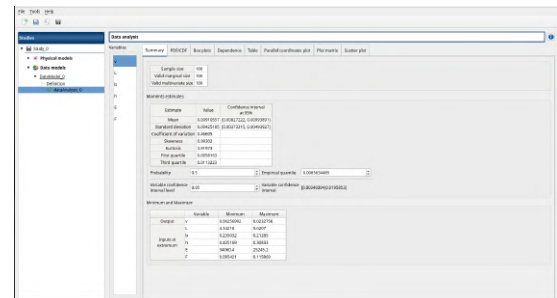
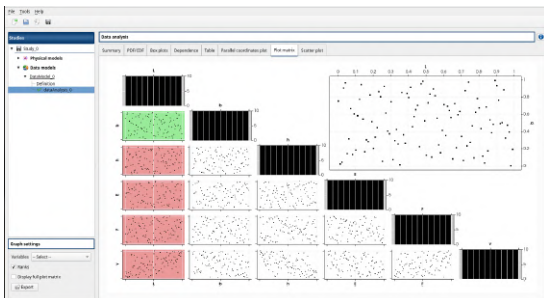


FIGURE 10 – Exemples de résultats de l'analyse de données

Inférences des distributions marginales et de la dépendance

Persalys permet d'inférer les distributions marginales sur chacun des paramètres, voir l'exemple de la figure 11. Plusieurs types de distributions peuvent être choisis, et les résultats obtenus fournissent le critère BIC [8] attribué à chaque type de distribution ainsi que les résultats des tests de Kolmogorov-Smirnov[9] ou de Lilliefors[10].

Le logiciel permet également d'inférer une copule caractérisant la dépendance entre les paramètres, ce qui permet de prendre en compte la corrélation entre eux. Plusieurs types de copules sont proposés, telles que les copules normale, student, Clayton, Gumbel et autres.

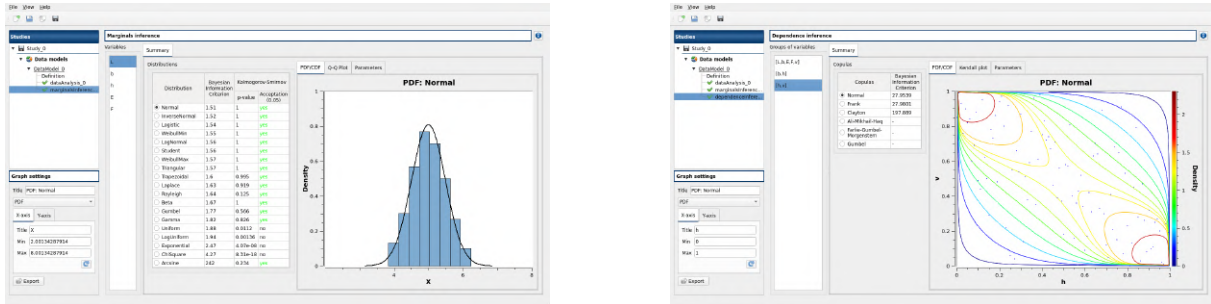


FIGURE 11 – Exemples de résultats d'inférence des distributions de données et de leurs dépendances

Analyse de sensibilité

La méthode de calcul des indices de Sobol d'ordre 1 est implémenté dans Persalys, la méthode est proposée par Gamboa et al. [11]. Elle permet d'estimer ces indices à partir des rangs des données. L'intervalle de confiance de l'estimation des indices de Sobol est également calculé par Bootstrap.

Les indices SRC sont également estimés en se basant sur les données. Des exemples de résultats des deux méthodes sont données dans la figure 12.

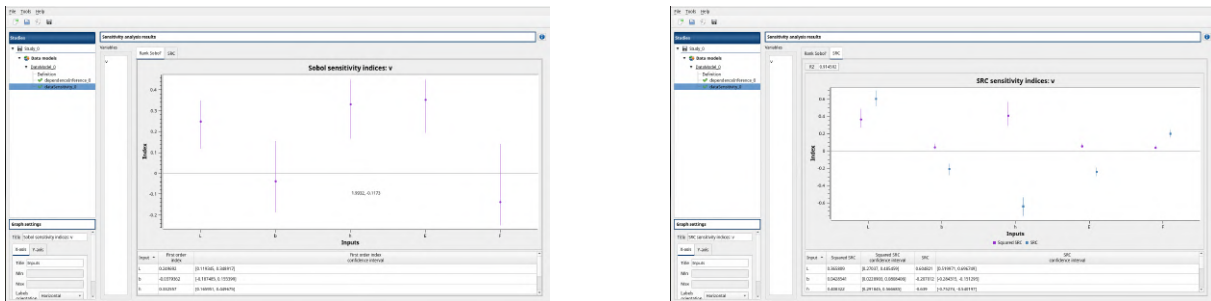


FIGURE 12 – Exemples de résultats de l'analyse de sensibilité sur un modèle de données

Calcul de quantile

Persalys permet d'estimer les quantiles sur chacun des paramètres, l'estimation peut être effectuée sur n'importe quelle marginale d'un modèle de données selon deux méthodes :

- Monte-Carlo : le quantile empirique est calculée à partir des données ainsi que son intervalle de confiance théorique.
- Méthode des excès via la de Pareto généralisée : la distribution de la queue est inférée en utilisant la théorie des valeurs extrêmes.

La figure 13 montre le menu disponible pour l'analyse de quantile et un exemple de résultats.

Création de métamodèle

Sur un modèle de données, Persalys propose de construire un métamodèle, on aura la même fonctionnalité présenté dans la section 3.4.

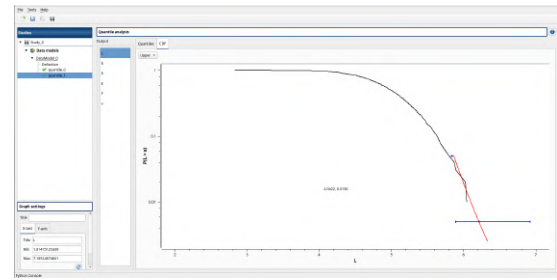
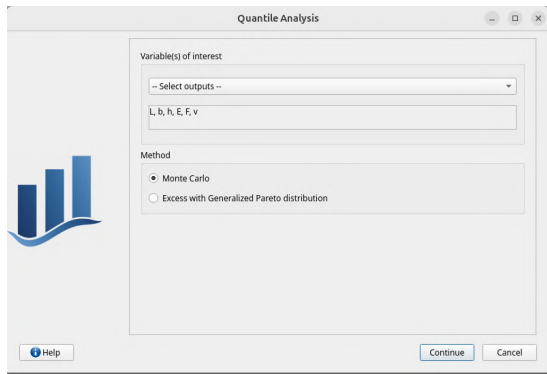


FIGURE 13 – Exemple d'analyse de quantile

4 Modèle de champs et jeu de données de champs

Le modèle de champs permet de construire un modèle dont la ou les sorties sont des champs 1-D. Il est possible d'importer un maillage ou de le construire localement dans Persalys. L'analyse disponible est la simulation de Monte Carlo, permettant une analyse statistique des trajectoires, et fournissant également une décomposition de Karhunen-Loève des trajectoires calculées ou importées.

5 Conclusion

Persalys est une interface graphique dédiée au traitement des incertitudes et à l'analyse statistique. Il est développé et maintenu en continu par une équipe composée de collaborateurs d'EDF et de Phimeca. Il propose plusieurs fonctionnalités et est conçu pour être accessible même aux utilisateurs non familiers avec l'analyse de données et le traitement des incertitudes.

Références

- [1] Dutfoy A, Dutka-Malen I, Pasanisi A, Lebrun R, Mangeant F, Gupta JS, Pendola M, Yalamas T. Open-TURNS, an Open Source initiative to Treat Uncertainties, Risks' N Statistics in a structured industrial approach. In41èmes Journées de Statistique, SFdS, Bordeaux 2009.
- [2] Morris, Max D. "Factorial sampling plans for preliminary computational experiments." *Technometrics* 33.2. 1991.
- [3] Stein M. Large sample properties of simulations using Latin hypercube sampling. *Technometrics*. 1987 May 1;29(2) :143-51.
- [4] Damblin G, Couplet M, Iooss B. Numerical studies of space-filling designs : optimization of Latin Hypercube Samples and subprojection properties. *Journal of Simulation*. 2013 Nov 1;7(4) :276-89.
- [5] Sobol, I. M. Sensitivity analysis for non-linear mathematical model *Math. Modelling Comput. Exp.*, 1993, 1, 407-414.
- [6] Haldar A, Mahadevan S. First-order and second-order reliability methods. In *Probabilistic Structural Mechanics Handbook : theory and industrial applications* 1995 (pp. 27-52). Boston, MA : Springer Us.
- [7] Shayanfar MA, Barkhordari MA, Roudak MA. An adaptive importance sampling-based algorithm using the first order method for structural reliability. *Int. J. Optim. Civil Eng.* 2017 Jan 10;7(1) :93-107.
- [8] Volinsky CT, Raftery AE. Bayesian information criterion for censored survival models. *Biometrics*. 2000 Mar;56(1) :256-62.
- [9] Massey Jr FJ. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association*. 1951 Mar 1;46(253) :68-78.
- [10] Lilliefors HW. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American statistical Association*. 1967 Jun 1;62(318) :399-402.
- [11] Gamboa F, Gremaud P, Klein T, Lagnoux A. Global sensitivity analysis : A novel generation of mighty estimators based on rank statistics. *Bernoulli*. 2022 Nov ;28(4) :2345-74.