L’analyse de sensibilité est une méthode qui permet de quantifier l’influence des paramètres d’entrée sur les variables de sorties. Elle est très utilisée dans l’étude des modèles complexes car elle permet de cibler les incertitudes importantes à prendre en compte dans une évaluation de la robustesse des sorties du modèle. Elle permet aussi de classer les paramètres par ordre d’influence pour l’interprétation du mécanisme de prédiction du modèle. Elle peut aussi être utilisée pour trouver un petit nombre de covariables maximisant la précision et ainsi faire du factor fixing, c’est-à-dire choisir les entrées à fixer.

La plupart des méthodes d’analyse de sensibilité utilisées repose sur une hypothèse d’indépendance entre les paramètres d’entrée, ce qui, on le sait, est rarement le cas dans les modèles. Nous avons donc étudié différentes méthodes d’analyse de sensibilité prenant en compte la dépendance entre les paramètres d’entrée (Shapley et Shapley HSIC[[1]](#footnote-1)) et nous avons choisi d’utiliser la méthode Sobol-MDA[[2]](#footnote-2). Cette méthode repose sur la construction d’un RF et les indices de sensibilité sont calculés à partir de l’estimation de l’importance de chaque covariables (paramètres du modèle complexe).

Dans cette étude, nous avons étudié la sensibilité d’une pêcherie langoustinière paramétrée avec ISIS-Fish. ISIS-Fish est un modèle de simulation des dynamiques des pêches utilisé pour appréhender une gestion écosystémique spatialisée des pêches.

Après avoir fait une hypothèse de dépendance sur les 5 paramètres incertains d’ISIS-Fish, nous avons construit un plan d’expérience et réalisé les simulations de ce plan avec ISIS-Fish. Nous avons ensuite ajusté un random forest à chaque variable de sortie du modèle (biomasse, biomasse de géniteurs et captures) et estimé les indices de Sobol totaux avec l’algorithme Sobol-MDA.

Tout d’abord, en se basant sur le *R²*, le coefficient de corrélation partielle qui correspond au pouvoir descriptif ou explicatif d’un métamodèle, on note que les 3 random forests s’ajustent bien . On remarque ensuite que le classement des paramètres par ordre d’influence change en fonction de la sortie étudiée. De plus, il évolue en fonction des années, par exemple, pour la biomasse, la fécondité est 4ème sur 5 la 1ère année (peu d’influence) mais elle est 2ème sur 5 la 4ème année (nettement plus influente). Ce qui peut s’expliquer par le fait que la première année de simulation est caractérisée par une très forte abondance de jeunes langoustines de l’année (recrutement) dont le poids augmente avec le temps.

Pour évaluer l’erreur que l’on aurait pu faire en ne prenant pas en compte la dépendance entre les paramètres, nous avons aussi calculé les indices de sensibilité pour des données dont les paramètres d’entrée sont indépendants. Nous avons remarqué des différences dans les valeurs des indices, surtout pour les paramètres sur lesquels la dépendance a été mise. Néanmoins, dans la plupart des cas, cette différence n’était pas assez grande pour changer le classement par ordre d’influence des paramètres. On en déduit donc que l’on aurait eu des résultats assez similaires sans prendre en compte la dépendance.

Finalement, nous avons voulu comprendre comment l’intensité de la dépendance influait sur l’analyse de sensibilité. Nous avons montré qu’une augmentation de la corrélation entre les paramètres peut changer le classement par ordre d’influence des paramètres. La valeur des indices de Sobol MDAdes paramètres corrélés diminue significativement lorsque la corrélation augmente. Comme le Sobol-MDA mesure l’augmentation de l’erreur du random forest lorsque l’un des paramètres est retiré du processus de prédiction, il est normal que l’indice de Sobol total d’un paramètre fortement corrélé à un autre diminue puisque l’information qu’il apportait est contenue dans un autre paramètre. Dans une configuration à forte corrélation, il faut néanmoins rester prudent si l’on souhaite faire du factor fixing basée sur une sélection des paramètres à faible indice de sensibilité. En effet, si l’on a 3 paramètres très corrélés par exemple, leurs indices de sensibilité mesuré avec une méthode prenant en compte la dépendance sont très petits. On peut alors être tenté de fixer ces 3 paramètres mais on perdra alors beaucoup plus en précision que la somme des trois indices car il n’en resterait aucun résumant l’information contenue dans les trois. Pour éviter cela, il faut n’en fixer que deux et en garder un permettant d’apporter presque autant d’information que les 3 (grâce à leur corrélation).

1. Da Veiga, Sébastien, Fabrice Gamboa, Bertrand Iooss, and Clémentine Prieur. 2021. Basics and Trends in Sensitivity Analysis: Theory and Practice in r. SIAM. [↑](#footnote-ref-1)
2. Bénard, Clément, Sébastien Da Veiga, and Erwan Scornet. 2022. “Mean Decrease Accuracy for Random Forests: Inconsistency, and a Practical Solution via the Sobol-MDA.” Biometrika 109 (4): 881–900. [↑](#footnote-ref-2)