Interprétation des modèles boîte noire de prévision

[1]**Département de mathématiques** - *Université du Québec à Montréal*
[2]**Institut Intelligence et Données** - *Université Laval*

Séminaire STATQAM
Université du Québec à Montréal - Montréal QC, Canada
29 Janvier 2026

Marouane Il Idrissi[12]

Marouane IL IDRISSI

il_idrissi.marouane@uqam.ca - marouanelidrissi.com

Doctorat - Université de Toulouse et EDF R&D (2021-2024)
Stagiaire postdoctoral (depuis août 2024)

☞ *Département de mathématiques, Université du Québec à Montréal*
☞ *Institut Intelligence et Données, Université Laval*

Bourse postdoctorale distinguée de l'INCASS (2025-2027)
Interprétation des modèles boîte noire
☞ Arthur Charpentier (UQÀM), Marie-Pier Côté (ULaval)

Thématiques de recherche :

Apprentissage statistique ● XAI ● Quantification de l'incertitude ● Analyse de sensibilité ● Théorie des probabilités ● Théorie des jeux coopératifs ● Analyse fonctionnelle

Thématique de recherche

Pourquoi les modèles de prévision “boîte noire” ne sont-ils pas largement utilisés pour modéliser des systèmes critiques ?

Système critique : Système dont la panne engendre des **conséquences dramatiques**

p. ex. barrage hydroélectrique, avion, marché financier, système judiciaire, un corps humain

Raison principale : La prise de décision doit être **justifiable** et **justifiée**.

p. ex. en utilisant des arguments statistiques

“Notre but est de s’assurer que le Québec soit à l’avant-garde du développement et de l’usage responsable et éthique de l’IA.”

Luc Sirois (2024), Directeur général du Conseil de l’Innovation du Québec

Les **méthodes d’IA explicable (XAI)** reposent généralement sur des **arguments empiriques**

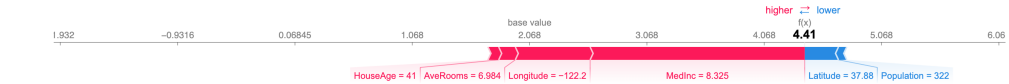
p. ex. méthodes populaires, tests sur des jeux de données limités...

☞ Ce n’est **pas suffisant pour convaincre** les autorités de sûreté ou de régulation...

Avant de choisir une méthode de XAI, il faut la comprendre théoriquement

Au programme de cette présentation

Méthodes **post-hoc**, **agnostiques au modèle**, basées sur les **jeux coopératifs**



Elles promettent **de quantifier l'influence des covariables**

p. ex. SHAP de Lundberg et Lee (2017)

Elles permettent de **décomposer des quantités d'intérêt**

p. ex. prévision ponctuelle, variance du modèle

Elles reposent beaucoup sur **la valeur de Shapley**

Il existe d'autres allocations pertinentes

Elles sont de **complexité exponentielle**

Mais il est possible de réduire le coût computationnel

☞ **Comment utiliser les jeux coopératifs pour extraire de l'information pertinente sur les modèles de prévision ?**

Cadre et notations

☞ $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé (abstrait)

☞ $X = (X_1, \ldots, X_d)$ est le vecteur des d covariables à valeur dans E $(\mathbb{R}^d)$

☞ $D = \{1, \ldots, d\}$ et $\mathcal{P}_D$ sont **l'ensemble des parties (power-set)** de D

☞ Pour chaque $A \in \mathcal{P}_D$, X_A est **le sous-ensemble des covariables d'indices dans** A

☞ $\widehat{f} : E \to \mathbb{R}$ est un **modèle de prévision boîte noire**, et $\widehat{f}(X)$ la **sortie aléatoire**

Remarque.

☞ Boîte-noire $\neq$ modèle complexe
Prise en compte d'une **large gamme de modèles** (peu d'hypothèses)

☞ Approche post-hoc
Le modèle est **déjà entraîné/estimé**, et on peut l'évaluer

Jeux coopératifs = Art du partage d'un gâteau

Deux éléments :

☞ Un ensemble de joueurs : $D = \{1, \ldots, d\}$
Et $\mathcal{P}_D$ représente **l'ensemble des coalitions de joueurs**

☞ Une fonction de valeur : $v : \mathcal{P}_D \to \mathbb{R}$
Elle **assigne une valeur à chaque coalition**

(D, v) définit formellement un **jeu coopératif** et $v(D)$ est la quantité à redistribuer (le gâteau)

Grande question :

☞ **Comment redistribuer $v(D)$ à chacun des d joueurs ?**

Exemple issu de la statistique - Indices LMG

Clouvel, looss, Chabridon, **Il Idrissi**, Robin. *Variance-based importance measures for linear regression*, Socio-Environmental Systems Modelling (2025)

Lindeman, Merenda et Gold (1980) : Contributions au R^2 dans une **régression linéaire**

☞ Joueurs : Les covariables $X_1, \ldots, X_d$

☞ Coalitions : Sous-ensembles des covariables

☞ **Fonction de valeur :** $v(A) = R^2_Y(X_A)$, le R^2 du modèle emboîté

☞ Le gâteau : Le coefficient de détermination $v(D) = R^2_Y(X)$ du modèle complet

Évaluer la **fonction de valeur** $\iff$ Calculer le R^2 des 2^d modèles emboîtés

Pour 20 covariables, c'est plus d'un million de coefficients

☞ **Comment synthétiser toute cette information ?**

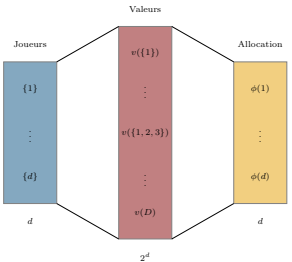
Allocations

Allocation : **Agrège** les évaluations de la **fonction de valeur**

Résume les 2^d valeurs en **une seule quantité pour chaque joueur**

C'est une fonction $\phi : D \to \mathbb{R}$, qui doit respecter le critère **d'efficacité** : $\sum_{i \in D} \phi(i) = v(D)$

Garantie de **bien redistribuer le gâteau**



☞ On veut étudier un modèle, avec d covariables

☞ On calcule la **fonction de valeur** pour chacun des 2^d sous-ensemble de covariables

☞ On résume ces 2^d quantités en d en utilisant une **allocation efficace**

☞ **Comment déterminer si une allocation est "bonne" ?**

Valeur de Shapley

Théorie des jeux coopératifs : Caractérise les **allocations** via leurs **axiomes**

Valeur de Shapley (1951) : **Unique** allocation $\mathrm{Shap} : D \to \mathbb{R}$ qui respecte

- **Efficacité :** $\sum_{i=1}^{d} \mathrm{Shap}(i) = v(D)$ ✔
- **Anonymité :** Si pour tout $A \in \mathcal{P}_{D \setminus \{i,j\}}$ $v(A \cup \{i\}) = v(A \cup \{j\})$, alors $\mathrm{Shap}(i) = \mathrm{Shap}(j)$ ✗
- **Joueur nul :** Si pour tout $A \in \mathcal{P}_D$, $v(A \cup \{i\}) = v(A)$, alors $\mathrm{Shap}(i) = 0$ ✗
- **Additivité :** La somme des Shap de (D, v) et (D, w), est égale à celle de $(D, v + w)$ ✗

Pour tout $i \in D$, on a :

$$\mathrm{Shap}(i) = \sum_{A \in \mathcal{P}_D \,:\, i \notin A} \frac{|A|!\,(d - |A| - 1)!}{d!} \left[v\left(A \cup \{i\}\right) - v(A) \right]$$

Résultat majeur qui a révolutionné la théorie des jeux coopératifs...
... mais qui n'est **pas forcément la solution** à **tous** les problèmes de XAI

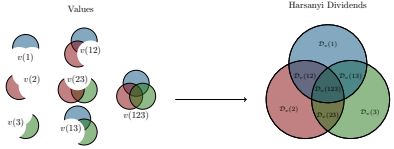
☞ **Peut-on aller plus loin que la valeur de Shapley ?**

Allocations vues comme un partage de dividendes

Il Idrissi, Bousquet, Gamboa, looss, Loubes. *On the coalitional decomposition of parameters*, Comptes Rendus. Mathématique (2023)

Dividendes de Harsanyi (1963) :

$$\mathcal{D}_v(A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} v(B), \quad \text{ou de manière équivalente,} \quad \mathcal{D}_v(A) = v(A) - \sum_{B \in \mathcal{P}_A} \mathcal{D}_v(B)$$



Mesurent la **plus-value d'une coalition** :
$\mathcal{D}_v(12) = v(12) - v(1) - v(2)$

On a également la **somme télescopique** :
$v(D) = \sum_{A \in \mathcal{P}_D} \mathcal{D}_v(A)$

Interprétation algébrique : **Inversion de Möbius** de la **fonction de valeur**

Proposition. *(Inversion de Möbius sur le treillis booléen (Rota 1964))*

Pour deux fonctions $v : \mathcal{P}_D \to \mathbb{A}$, $\mathcal{D} : \mathcal{P}_D \to \mathbb{A}$, où $\mathbb{A}$ est un groupe abélien, on a l'équivalence suivante :

$$\forall A \in \mathcal{P}_D, \quad v(A) = \sum_{B \in \mathcal{P}_A} \mathcal{D}(B), \quad \iff \quad \forall A \in \mathcal{P}_D, \quad \mathcal{D}(A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} v(B).$$

Valeur de Shapley : partage égalitaire des dividendes

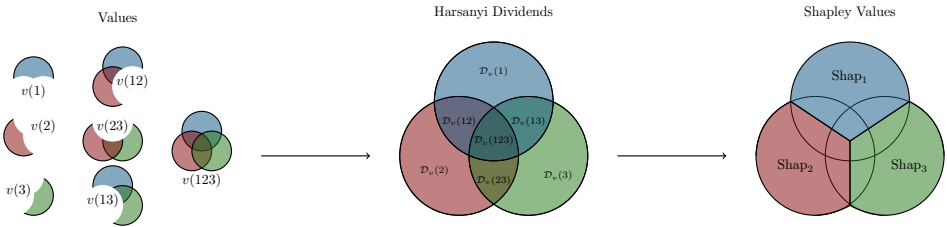
Allocations de Harsanyi : Famille d'**allocations efficaces**

Agrégation des **dividendes** :

$$\phi(i) = \sum_{A \in \mathcal{P}_D \,:\, i \in A} \lambda_i(A) \mathcal{D}_v(A), \quad \text{où} \quad \begin{cases} \forall i \in D, \forall A \in \mathcal{P}_D,\ \lambda_i(A) \geq 0, \\ \forall A \in \mathcal{P}_D,\ \sum_{i \in D} \lambda_i(A) = 1 \end{cases}$$

La famille est paramétrée par un **système de poids** $\lambda : D \times \mathcal{P}_D \to \mathbb{R}$

Valeur de Shapley : **Partage égalitaire** des **dividendes** ($\lambda_i(A) = 1/|A|$)



Allocations et ordres aléatoires

Allocations de Weber (1988) : Famille d'**allocations** efficaces, qui reposent sur la notion **d'ordre aléatoire**

Soit $\mathcal{S}_D$ l'ensemble **des permutations** $\pi = (\pi_1, \ldots, \pi_d)$ (c.-à-d. ordres) de joueurs
Pour un joueur $i \in D$, $\pi(i)$ est sa position dans la permutation π (c.-à-d. $\pi_{\pi(i)} = i$)

Espérance sur les ordres :

$$\begin{aligned}\phi(i) &= \mathbb{E}_{\pi\sim p}\left[v\left(\{\pi_1, \ldots, \pi_{\pi(i)}\}\right) - v\left(\{\pi_1, \ldots, \pi_{\pi(i)-1}\}\right)\right] \\ &= \sum_{\pi\in\mathcal{S}_D} p(\pi)\left[v\left(\{\pi_1, \ldots, \pi_{\pi(i)}\}\right) - v\left(\{\pi_1, \ldots, \pi_{\pi(i)-1}\}\right)\right]\end{aligned}$$

La famille est paramétrée par **une fonction de masse de probabilité** p sur $\mathcal{S}_D$

Valeur de Shapley : **Distribution uniforme** ($p(\pi) = 1/d!$)

$$\mathrm{Shap}(i) = \frac{1}{d!}\sum_{\pi\in\mathcal{S}_D}\left[v\left(\{\pi_1, \ldots, \pi_{\pi(i)}\}\right) - v\left(\{\pi_1, \ldots, \pi_{\pi(i)-1}\}\right)\right]$$

Notre position

Il Idrissi, Charpentier et Fernandes Machado. *Cooperative Games for the Interpretation of ML models*, IJCAI'25, Workshop on XAI (2025)

Trois étapes fondamentales pour utiliser les **jeux coopératifs** pour l'interprétation :

- Étape 1 : Choisir la **quantité d'intérêt**
  p. ex. une prévision $\widehat{f}(x)$, la variance de la sortie $\mathbb{V}\left(\widehat{f}(X)\right)$, une mesure d'ajustement $R^2_Y(X)$
  ☞ Guide **l'interprétation** que l'on fait de l'allocation

- Étape 2 : Choisir la **fonction de valeur** v
  par ex. $\mathbb{E}\left[\widehat{f}(X) \mid X_A = x_A\right]$ ou $\mathbb{E}_{X_{\overline{A}}}\left[\widehat{f}(x_A, X_{\overline{A}})\right]$ pour $\widehat{f}(x)$, $\mathbb{V}\left(\mathbb{E}\left[\widehat{f}(X) \mid X_A\right]\right)$ pour $\mathbb{V}\left(\widehat{f}(X)\right)$...
  ☞ **Cette étape est la plus importante !**

- Étape 3 : Choisir l'**allocation efficace**
  Permettra de **résumer l'information** contenue dans les 2^d évaluations de v
  ☞ Peut jouer un rôle dans la recherche de **propriétés intéressantes**

Sommaire

Cinq grands défis :

1. Le choix de la **fonction de valeur**
2. Le choix de l'**allocation**
3. Explorer la décomposition de nouvelles **quantités d'intérêt**
4. Aspects computationnels
5. Application aux systèmes critiques

Défi n°1 :
Le choix de la fonction de valeur

Choix de la fonction de valeur

Un **mauvais choix** peut entraîner **des interprétations fallacieuses**

p. ex. **identification** (Zhang et al. 2024), **pureté** (Köhler et al. 2024), **exogénéité** (Iooss et al. 2019)

$$\widehat{f}(X) = X_1 + X_2 + X_1X_2, \quad X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right), \quad x = X(\omega),\ \omega \in \Omega$$

Espérance conditionnelle / Projection orthogonale

$$v(A) = \mathbb{E}\left[\widehat{f}(X) \mid X_A = x_A\right]$$

$$\mathcal{D}_v(1) = x_1 + \rho(x_1 + x_1^2 - 1) \quad \mathcal{D}_v(2) = x_2 + \rho(x_2 + x_2^2 - 1)$$

$$\mathcal{D}_v(12) = x_1x_2 - \rho(x_1 + x_1^2 + x_2 + x_2^2 - 1)$$

$$\mathrm{Shap}(1) = x_1 + \frac{\rho}{2}(x_1 + x_1^2 - x_2 - x_2^2 - 1) + \frac{x_1x_2}{2}$$

$$\mathrm{Shap}(2) = x_2 + \frac{\rho}{2}(x_2 + x_2^2 - x_1 - x_1^2 - 1) + \frac{x_1x_2}{2}$$

Projection oblique

$$v(A) = \mathbb{M}_A\left[\widehat{f}(X)\right](x_A)$$

$$\mathcal{D}_v(1) = x_1 \quad \mathcal{D}_v(2) = x_2$$

$$\mathcal{D}_v(12) = x_1x_2$$

$$\mathrm{Shap}(1) = x_1 + \frac{x_1x_2}{2}$$

$$\mathrm{Shap}(2) = x_2 + \frac{x_1x_2}{2}$$

Projections obliques

El Idrissi, Bousquet, Gamboa, Iooss et Loubes. *Hoeffding decomposition of functions of random dependent variables*, Journal of Multivariate Analysis (2025)

Projection oblique : Généralise l'opérateur **d'espérance conditionnelle**

Théorème. Sous des hypothèses peu restrictives sur la loi de X, on a, pour chaque $A \in \mathcal{P}_D$,

$$\mathbb{L}^2(\sigma_A) = \bigoplus_{B \in \mathcal{P}_A} V_B, \quad \text{où } V_B = \left[\sum_{C \in \mathcal{P}_B, C \neq B} V_C \right]^{\perp_B}, \quad \text{et } \sigma_A = \sigma(X_A) \subset \mathcal{F}$$

☞ Espérance conditionnelle $\mathbb{E}[. \mid X_A]$: Projection sur $\mathbb{L}^2(\sigma_A)$ parallèle à $\mathbb{L}^2(\sigma_A)^{\perp}$

☞ Projection oblique $\mathbb{M}_A[.]$: Projection sur $\mathbb{L}^2(\sigma_A)$ parallèle à $\bigoplus_{B \in \mathcal{P}_D \setminus \mathcal{P}_A} V_A$

Corollaire. Toute v.a. $\widehat{f}(X) \in \mathbb{L}^2(\sigma_X)$ peut être s'écrire de façon **unique** comme

$$\widehat{f}(X) = \sum_{A \in \mathcal{P}_D} f_A(X_A), \quad \text{où} \quad f_A(X_A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} \mathbb{M}_B \left[\widehat{f}(X) \right] \in V_A$$

Travaux futurs

Perspective : Estimer les projections obliques

☞ **Champion et al. (2015) :** Estimateur lorsque $\widehat{f}$ est un **modèle de boosting**

☞ **Ferrere et al. (2025) :** Calcul des projections obliques lorsque X est **Bernoulli multivariée**

☞ **Benard (2025) :** Estimateur lorsque $\widehat{f}$ est **une forêt aléatoire**

Projets de recherche :

Estimateur(s) non-paramétrique(s), agnostique(s) au modèle




Projections obliques pour les modèles de boosting par arbres

Défi n°2 :
Le choix de l'allocation

Choix de l'allocation

Pour une **fonction de valeur** fixée, l'**allocation** peut entraîner des **propriétés intéressantes**

Exemple : Blague de Shapley (looss et Prieur 2019)

1. **Quantité d'intérêt :** $\mathbb{V}(\widehat{f}(X))$
2. **Fonction de valeur :** $v(A) = \mathbb{E}\left[\mathbb{V}(\widehat{f}(X) \mid X_{D\setminus A})\right]$
3. **Allocation :** Valeur de Shapley

$$\widehat{f}(X) = X_1 + X_2, \quad X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & \rho \\ 0 & 1 & 0 \\ \rho & 0 & 1 \end{pmatrix}\right),$$

$$\mathrm{Shap}(1) = 0.5 - \rho^2/4, \quad \mathrm{Shap}(2) = 0.5, \quad \mathrm{Shap}(3) = \rho^2/4$$

☞ **Non détection de l'exogénéité**

X_3 n'est **pas dans le modèle**, mais reçoit une **part du gâteau**

Effets marginaux proportionnels

Herin*, **Il Idrissi**, Chabridon, looss. *Proportional Marginal Effects for Global Sensitivity Analysis*, SIAM/ASA Journal on Uncertainty Quantification (2024)

La **valeur proportionnelle** (PV) (Ortmann 2000) :

$$p(\pi) = \frac{L(\pi)}{\sum_{\sigma \in \mathcal{S}_D} L(\sigma)}, \quad L(\pi) = \exp\left(- \sum_{j \in D} \log\left(v\left(\{\pi_1, \dots \pi_{\pi(j)}\} \right) \right) \right)$$

Effets marginaux proportionnels (PME) :

Extension continue de la PV pour $v(A) = \mathbb{E}\left[\mathbb{V}(\widehat{f}(X) \mid X_{D \setminus A}) \right]$

Proposition. *(Détection de l'exogénéité)*

$$\mathrm{PME}(i) = 0 \iff X_i \text{ n'est pas dans le modèle}$$

Alternative à la **valeur de Shapley**, qui **détecte l'exogénéité**

Travaux futurs

Perspective : Allocations dédiées à l'interprétation des modèles

- ☞ **Frye, Rowat et Feige (2020)** : Allocation et graphe causal
- ☞ **Fumagalli et al. (2023)** : Indices de Banzhaf
- ☞ **Miroshnikov et al. (2024)** : Valeur de Owen

Projets de recherche :

Allocations basées sur des modèles graphiques


Allocations minimisant un objectif

Défi n°3 :
Explorer la décomposition de nouvelles quantités d'intérêt

Attribution de l'incertitude

Il Idrissi, Fernandes Machado, Gallic, Charpentier. *Feature Contribution to Conformal Prediction Intervals*, prépublication

Idée : Décomposer **l'incertitude de prévision** d'un modèle de prévision
☞ **Quelles covariables contribuent à rendre une prévision incertaine ?**

Outil de choix : **Intervalle de prévision conforme** (Papadopoulos et al. 2002)

☞ Un jeu de données $\mathcal{D}$ échangeable tel que $\mathcal{D} = \mathcal{D}^{\mathrm{Tr}} \cup \mathcal{D}^{\mathrm{Cal}}$

☞ Un modèle $\widehat{f} : \mathcal{X} \to \mathcal{Y} \in \mathcal{Y}^{\mathcal{X}}$ entraîné sur $\mathcal{D}^{\mathrm{Tr}}$

☞ Un score $s(x, y, f)$, et l'ensemble $\mathcal{S} = \left\{ s(X_i, Y_i, \widehat{f}) : (X_i, Y_i) \in \mathcal{D}^{\mathrm{Cal}} \right\}$

☞ Pour une **nouvelle observation** (X_{n+1}, Y_{n+1}), **l'intervalle de prévision conforme est**

$$\widehat{C}(X_{n+1}) := \left\{ y \in \mathcal{Y} : s\left(X_{n+1}, y, \widehat{f}\right) \leq q_{1-\alpha}(\mathcal{S}) \right\}$$

où $q_{1-\alpha}(\mathcal{S})$ est le quantile empirique d'ordre $1-\alpha$ de $\mathcal{S}$

Proposition.

$$1 - \alpha \leq \mathbb{P}\left(Y_{n+1} \in \widehat{C}(X_{n+1})\right) \leq 1 - \alpha + \frac{1}{\#\mathcal{D}^{\mathrm{Cal}} + 1}$$

Attribution de l'incertitude

Il Idrissi, Fernandes Machado, Gallic, Charpentier. *Feature Contribution to Conformal Prediction Intervals*, prépublication

1. **Quantité d'intérêt :** Largeur de l'**intervalle de prévision conforme** (IPC)
2. **Fonction de valeur :** Largeur de l'**IPC**, du modèle emboîté avec X_A
3. **Allocation :** *Valeur de Shapley et Shapley proportionnelle*

Algorithm 1 Exact Computation Procedure

Require: Data $\mathcal{D} = \{(X_i, Y_i)\}_{i=1}^n$, new data $\mathcal{D}^{\text{New}} = \{(X_i, Y_i)\}_{i=n+1}^m$, miscoverage level $\alpha \in (0,1)$, regression algorithm f, conformity score algorithm s, and a weight assignment $\lambda : \mathcal{P}_D \to \mathbb{R}$ that associates a weight $\lambda(A)$ to each subset $A \subseteq D$, where $D = \{1, \ldots, d\}$ is the set of variable indices in X.

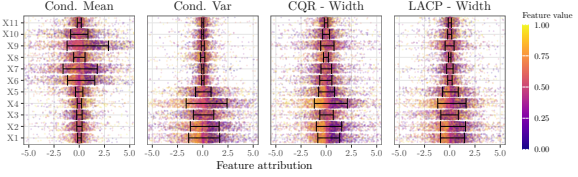
Randomly split $\mathcal{D}$ into two disjoint datasets $\mathcal{D}^{\text{Tr}}$ and $\mathcal{D}^{\text{Cal}}$
for $A \in \mathcal{P}_D$ **do**
  Define $\mathcal{D}_A^{\text{Tr}} = \{(X_{i,A}, Y_i)\}_{(X_i,Y_i)\in\mathcal{D}^{\text{Tr}}}$
  Define $\mathcal{D}_A^{\text{Cal}} = \{(X_{i,A}, Y_i)\}_{(X_i,Y_i)\in\mathcal{D}^{\text{Cal}}}$
  Define $\mathcal{D}_A^{\text{New}} = \{X_{i,A}\}_{X_i\in\mathcal{D}^{\text{New}}}$
  Train $\hat{f}_A$ on $\mathcal{D}_A^{\text{Tr}}$
  Compute conformity scores $\hat{s}_i$ for each $(X_i, Y_i) \in \mathcal{D}_A^{\text{Cal}}$
  for $X_i \in \mathcal{D}_A^{\text{New}}$ **do**
    Compute conformal prediction interval $C_A(X_i)$
    Compute associated value $v(A; X_i)$
  end for
end for
Initialize matrix $\Phi \in \mathbb{R}^{m\times d}$ with zeros
for $X_i \in \mathcal{D}^{\text{New}}$ **do**
  for $j \in D$ **do**
    Compute $\phi_i(j)$
    Store $\phi_i(j)$ in $\Phi_{i,j}$
  end for
end for
return Φ

☞ Méthode **d'attribution de l'incertitude**
Plus générale que la variance

☞ **Trois types de scores pour trois méthodes différentes**
Standard Mean Regression, *Local Adaptive Conformal Prediction*, *Conformalized quantile regression*

☞ **N'implique pas d'hypothèse additionnelle sur $\hat{f}$**

☞ Apporte une **information différente que celle des méthodes existantes**



Travaux futurs

Perspective : Décomposer d'autres quantités d'intérêt

☞ **da Veiga (2021)** : Mesures d'incertitude basées sur des méthodes à noyau

☞ **Watson et al. (2023)** : Mesures d'incertitude basées sur l'entropie

☞ **Lindholm et al. (2026)** : Mesures de discrimination

Projets de recherche :

Décomposition de mesures de discrimination par proxy



Décomposition des erreurs de calibration d'un modèle

Défi n°4 :
Aspects computationnels

Aspects computationnels

Les méthodes basées sur les jeux coopératifs sont **de complexité exponentielle**

☞ S'autoriser une **petite erreur** pour **réduire le temps de calcul**

Deux types d'approximations :

- Approcher la **fonction de valeur**
  Calculer chaque $v(A)$ plus vite
  p. ex. KernelSHAP (Lundberg et Lee 2017), TreeSHAP (Lundberg, Erion et Lee 2018)

- Approcher l'**allocation**
  Évaluer la **fonction de valeur** sur **moins de coalitions**
  p. ex. Monte Carlo (Štrumbelj et Kononenko 2014), FastSHAP (Jethani et al. 2022)

Approcher l'allocation - Monte Carlo et permutations

Il Idrissi, Fernandes Machado, Gallic, Charpentier. *Feature Contribution to Conformal Prediction Intervals*, prépublication

Idée : **Échantillonner les permutations** selon p (Monte Carlo)
Généralise l'estimateur de (Štrumbelj et Kononenko 2014)

☞ On veux approcher l'allocation

$$\phi(i) := \mathbb{E}_{\pi\sim p}\left[v\left(\{\pi_1,\dots,\pi_{\pi(i)}\}\right) - v\left(\{\pi_1,\dots,\pi_{\pi(i)-1}\}\right)\right]$$

☞ On tire un échantillon $\pi^{(1)},\dots,\pi^{(m)}$ i.i.d. selon p

Proposition.

$$\widehat{\phi}(i) = \frac{1}{m}\sum_{j=1}^{m}\left[v\left(\{\pi_1,\dots,\pi_{\pi(i)}\}\right) - v\left(\{\pi_1,\dots,\pi_{\pi(i)-1}\}\right)\right]$$

est un estimateur sans biais, fortement convergent, et asymptotiquement normal de $\phi(i)$.

☞ **On n'évalue v que pour $m \times d \ll 2^d$ coalitions (pire cas)**

Approcher l'allocation - Échantillonnage d'importance

Il Idrissi, Fernandes Machado, Gallic, Charpentier. *Feature Contribution to Conformal Prediction Intervals.*

Idée : **Recycler un échantillon** (Échantillonnage d'importance)

☞ On tire un échantillon $\pi^{(1)}, \dots, \pi^{(m)}$ i.i.d. selon p

☞ On s'intéresse à l'allocation

$$\phi'(i) := \mathbb{E}_{\pi \sim p'} \left[v\left(\{\pi_1, \dots, \pi_{\pi(i)}\}\right) - v\left(\{\pi_1, \dots, \pi_{\pi(i)-1}\}\right) \right]$$

selon la fonction de masse de probabilité p'

Proposition.

$$\widehat{\phi'}^{\mathrm{IS}}(j) = \frac{1}{m} \sum_{i=1}^{m} \frac{p'(\pi_i)}{p(\pi_i)} \left[v\left(\{\pi_1, \dots, \pi_{\pi(i)}\}\right) - v\left(\{\pi_1, \dots, \pi_{\pi(i)-1}\}\right) \right]$$

est un estimateur sans biais, fortement convergent, et asymptotiquement normal de $\phi'(i)$.

☞ **Deux estimations pour le prix d'une**

Travaux futurs

Perspective : Approximations et implémentations informatiques

☞ **Covert et Lee (2021)** : Approcher l'espérance conditionnelle par une régression linéaire

☞ **Jethani et al. (2022)** : Échantillonnage des coalitions pour Shapley (sans convergence)

☞ **Chen et al. (2023)** : Étude sur les algorithmes d'approximation de l'espérance conditionnelle

Projets de recherche :

Approximation de fonction de valeur et échantillonnage avancé

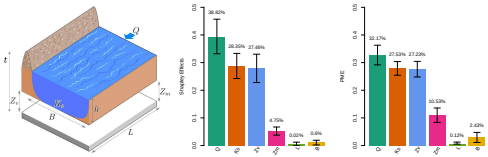

Paquetage `R` efficace


Défi n°5 :
Application aux systèmes critiques

Application aux systèmes critiques

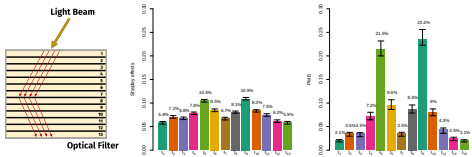
Survenance de crue en contexte industriel

(Il Idrissi et al. 2021, Il Idrissi et al. 2024)



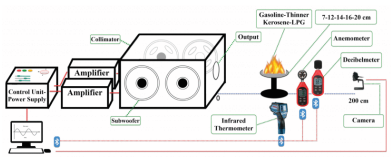
Performance de filtres optiques en aéronautique

(Herin et al. 2024)



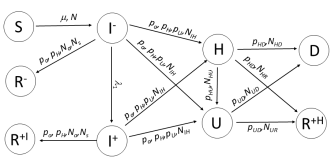
Extinction acoustique d'un feu

(Il Idrissi et al. 2024)



Modèle SIR pour la COVID-19 en France

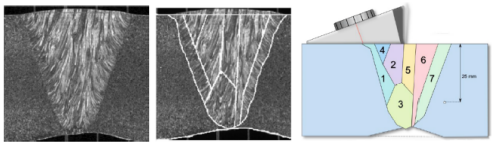
(Il Idrissi et al. 2021)



Application aux systèmes critiques

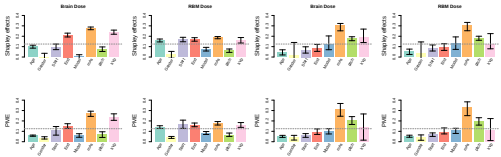
Contrôle ultrasonique de soudures industrielles

(Il Idrissi et al. 2021)



Doses de radiation absorbées après tomodensitométries pendant l'enfance

(Foucault et al. 2023)



- **Il Idrissi**, Chabridon, looss. *Developments and applications of Shapley effects to reliability-oriented sensitivity analysis with correlated inputs*, Environmental Modelling and Software (2021)
- **Il Idrissi**, Bousquet, Gamboa, looss, Loubes. *Quantile-constrained Wasserstein projections for robust interpretability of numerical and machine learning models*, Electronic Journal of Statistics (2024)
- Herin*, **Il Idrissi**, Chabridon, looss. *Proportional Marginal Effects for Global Sensitivity Analysis*, SIAM/ASA Journal on Uncertainty Quantification (2024)
- Foucault, **Il Idrissi**, looss, Ancelet, *Shapley and proportional marginal effects : application to computed tomography scan organ dose estimation*

Travaux futurs

Perspective : Étudier d'autres systèmes critiques

☞ **Demange-Chryst, Bachoc et Morio (2023)** : Modèles de propagation de feu de forêt

☞ **Koç et Uzmay (2024)** : Construction d'un indice de sécurité alimentaire

☞ **Dumon, Lebental et Perrin (2025)** : PME pour le suivi de la pollution de l'air

Projets de recherche :

Analyse de sensibilité pour mesurer la discrimination en tarification


Évaluation de la discrimination et méthodes d'interprétation pour la validation méthodologique



Conclusion

Méthodes d'interprétations inspirées des jeux coopératifs :

- Trois ingrédients pour mobiliser les jeux coopératifs en interprétabilité : la **quantité d'intérêt**, la **fonction de valeur**, et l'**allocation**
- La **valeur de Shapley** est **un exemple** d'allocation et **pas une solution universelle**
- Limite pratique principale : la **complexité exponentielle**
- Outils permettant **d'évaluer et de valider les modèles de prévision**

Interprétation des modèles boîte noire de prévision :

- Domaine de recherche en **phase de maturation**
- **Questions scientifiques** au croisement de **plusieurs domaines d'études**
- **Plus-value pratique forte** et un **engouement croissant**
- L'un des **grands défis** pour une **adoption responsable de l'intelligence artificielle**

References i

Axler, S. 2015. *Linear Algebra Done Right* [en en]. Undergraduate Texts in Mathematics. Cham : Springer International Publishing. ISBN : 978-3-319-11079-0 978-3-319-11080-6. https://doi.org/10.1007/978-3-319-11080-6. https://link.springer.com/10.1007/978-3-319-11080-6.

Benard, C. 2025. "Tree Ensemble Explainability through the Hoeffding Functional Decomposition and TreeHFD Algorithm". In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*. https://openreview.net/forum?id=dRLWcpBQxS.

Broto, B., F. Bachoc et M. Depecker. 2020. "Variance Reduction for Estimation of Shapley Effects and Adaptation to Unknown Input Distribution". *SIAM/ASA Journal on Uncertainty Quantification* 8 (2) : 693-716. ISSN : 2166-2525. https://doi.org/10.1137/18M1234631. https://epubs.siam.org/doi/10.1137/18M1234631.

Bryc, W. 1984. "Conditional expectation with respect to dependent sigma-fields". In *Proceedings of VII conference on Probability Theory,* 409-411. https://homepages.uc.edu/~brycwz/preprint/Brasov-1982.pdf.

———. 1996. "Conditional Moment Representations for Dependent Random Variables". Publisher : Institute of Mathematical Statistics and Bernoulli Society, *Electronic Journal of Probability* 1 (none) : 1-14. ISSN : 1083-6489, 1083-6489. https://doi.org/10.1214/EJP.v1-7. https://projecteuclid.org/journals/electronic-journal-of-probability/volume-1/issue-none/Conditional-Moment-Representations-for-Dependent-Random-Variables/10.1214/EJP.v1-7.full.

Champion, M., G. Chastaing, S. Gadat et C. Prieur. 2015. "L2-Boosting for sensitivity analysis with dependent inputs". *Statistica Sinica* 25 (4) : 1477-1502. ISSN : 10170405, 19968507, visité le 18 janvier 2026. http://www.jstor.org/stable/24721243.

References ii

Chen, H., I. C. Covert, S. M. Lundberg et S.-I. Lee. 2023. "Algorithms to estimate Shapley value feature attributions". *Nature Machine Intelligence* 5, n° 6 (juin) : 590-601. ISSN : 2522-5839. https://doi.org/10.1038/s42256-023-00657-x. https://doi.org/10.1038/s42256-023-00657-x.

Covert, I., et S.-I. Lee. 2021. "Improving KernelSHAP : Practical Shapley Value Estimation Using Linear Regression". In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, sous la direction d'A. Banerjee et K. Fukumizu, 130 : 3457-3465. Proceedings of Machine Learning Research. PMLR. https://proceedings.mlr.press/v130/covert21a.html.

da Veiga, S. 2021. "Kernel-based ANOVA decomposition and Shapley effects - Application to global sensitivity analysis". Working paper or preprint, janvier. https://hal.science/hal-03108628.

Dauxois, J, G. M Nkiet et Y Romain. 2004. "Canonical analysis relative to a closed subspace". *Linear Algebra and its Applications*, Tenth Special Issue (Part 1) on Linear Algebra and Statistics, 388 : 119-145. ISSN : 0024-3795. https://doi.org/10.1016/j.laa.2004.02.036. https://www.sciencedirect.com/science/article/pii/S0024379504001107.

Demange-Chryst, J., F. Bachoc et J. Morio. 2023. "SHAPLEY EFFECT ESTIMATION IN RELIABILITY-ORIENTED SENSITIVITY ANALYSIS WITH CORRELATED INPUTS BY IMPORTANCE SAMPLING" [en English]. Publisher : Begel House Inc. *International Journal for Uncertainty Quantification* 13 (3). ISSN : 2152-5080, 2152-5099. https://doi.org/10.1615/Int.J.UncertaintyQuantification.2022043692. https://www.dl.begellhouse.com/journals/52034eb04b657aea,796f39cb1acf1296,701a298578abbff3.html.

References iii

Dumon, M., B. Lebental et G. Perrin. 2025. *Variance-based variable selection in sensor calibration with strong interferents – application to air pollution monitoring with a carbon nanotube sensor array*. ArXiv :2507.05001 [stat], juillet. https://doi.org/10.48550/arXiv.2507.05001. http://arxiv.org/abs/2507.05001.

Feldman, B. E. 1999. "The proportional value of a cooperative game". *Manuscript. Chicago : Scudder Kemper Investments*.

Ferrere, B., N. Bousquet, F. Gamboa, J.-M. Loubes et J. Muré. 2025. *Multivariate Bernoulli Hoeffding Decomposition : From Theory to Sensitivity Analysis*. arXiv : 2510.07088 [stat.ML]. https://arxiv.org/abs/2510.07088.

Foucault, A., M. Il Idrissi, B. Iooss et S. Ancelet. 2023. "Shapley effects and proportional marginal effects for global sensitivity analysis : application to computed tomography scan organ dose estimation". Preprint. https://hal.science/hal-04114533.

Friedrichs, K. 1937. "On Certain Inequalities and Characteristic Value Problems for Analytic Functions and For Functions of Two Variables". Publisher : American Mathematical Society, *Transactions of the American Mathematical Society* 41 (3) : 321-364. ISSN : 0002-9947. https://doi.org/10.2307/1989786. https://www.jstor.org/stable/1989786.

Frye, C., C. Rowat et I. Feige. 2020. "Asymmetric Shapley values : incorporating causal knowledge into model-agnostic explainability". In *Advances in Neural Information Processing Systems*, 33 : 1229-1239. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2020/hash/0d770c496aa3da6d2c3f2bd19e7b9d6b-Abstract.html.

References iv

Fumagalli, F., M. Muschalik, P. Kolpaczki, E. Hüllermeier et B. Hammer. 2023. "SHAP-IQ : Unified Approximation of any-order Shapley Interactions" [en en]. *Advances in Neural Information Processing Systems* 36 (décembre) : 11515-11551. https://proceedings.neurips.cc/paper_files/paper/2023/hash/264f2e10479c9370972847e96107db7f-Abstract-Conference.html.

Galántai, A. 2004. *Projectors and Projection Methods*. Boston, MA : Springer US. ISBN : 978-1-4613-4825-2. https://doi.org/10.1007/978-1-4419-9180-5. http://link.springer.com/10.1007/978-1-4419-9180-5.

Harsanyi, J. C. 1963. "A Simplified Bargaining Model for the n-Person Cooperative Game". Publisher : [Economics Department of the University of Pennsylvania, Wiley, Institute of Social and Economic Research, Osaka University], *International Economic Review* 4 (2) : 194-220. ISSN : 0020-6598. https://doi.org/10.2307/2525487. https://www.jstor.org/stable/2525487.

Herin, M., M. Il Idrissi, V. Chabridon et B. Iooss. 2024. "Proportional Marginal Effects for Global Sensitivity Analysis" [en en]. *SIAM/ASA Journal on Uncertainty Quantification* 12, n° 2 (juin) : 667-692. ISSN : 2166-2525. https://doi.org/10.1137/22M153032X.

Hoeffding, W. 1948. "A Class of Statistics with Asymptotically Normal Distribution". *The Annals of Mathematical Statistics* 19 (3) : 293-325. ISSN : 0003-4851, 2168-8990. https://doi.org/10.1214/aoms/1177730196. https://projecteuclid.org/journals/annals-of-mathematical-statistics/volume-19/issue-3/A-Class-of-Statistics-with-Asymptotically-Normal-Distribution/10.1214/aoms/1177730196.full.

Il Idrissi, M., N. Bousquet, F. Gamboa, B. Iooss et J.-M. Loubes. 2024. "Quantile-constrained Wasserstein projections for robust interpretability of numerical and machine learning models". *Electronic Journal of Statistics* 18 (2) : 2721 -2770. https://doi.org/10.1214/24-EJS2268.

References v

Il Idrissi, M., V. Chabridon et B. Iooss. 2021. "Developments and applications of Shapley effects to reliability-oriented sensitivity analysis with correlated inputs". *Environmental Modelling and Software* 143 : 105115. ISSN : 1364-8152. https://doi.org/10.1016/j.envsoft.2021.105115.

Iooss, B., et P. Lemaître. 2015. "A Review on Global Sensitivity Analysis Methods". In *Uncertainty Management in Simulation-Optimization of Complex Systems : Algorithms and Applications*, sous la direction de G. Dellino et C. Meloni, 101-122. Springer US. https://doi.org/10.1007/978-1-4899-7547-8_5. https://doi.org/10.1007/978-1-4899-7547-8_5.

Iooss, B., et C. Prieur. 2019. "Shapley effects for Sensitivity Analysis with correlated inputs : Comparisons with Sobol' Indices, Numerical Estimation and Applications". *International Journal for Uncertainty Quantification* 9 (5) : 493-514.

Jethani, N., M. Sudarshan, I. C. Covert, S.-I. Lee et R. Ranganath. 2022. "FastSHAP : Real-Time Shapley Value Estimation". In *International Conference on Learning Representations*. https://openreview.net/forum?id=Zq2G_VTV53T.

Kallenberg, O. 2021. *Foundations of modern probability*. Probability theory and stochastic modelling. Cham, Switzerland : Springer. ISBN : 978-3-030-61871-1. https://doi.org/10.1007/978-3-030-61871-1.

Koklu, M., et Y. S. Taspinar. 2021. "Determining the Extinguishing Status of Fuel Flames With Sound Wave by Machine Learning Methods". Conference Name : IEEE Access, *IEEE Access* 9 : 86207-86216. ISSN : 2169-3536. https://doi.org/10.1109/ACCESS.2021.3088612.

Koç, G., et A. Uzmay. 2024. "Construction of a Farm-Level Food Security Index : Case Study of Turkish Dairy Farms" [en en]. *Social Indicators Research* 175, n° 2 (novembre) : 687-714. ISSN : 1573-0921. https://doi.org/10.1007/s11205-024-03406-8. https://doi.org/10.1007/s11205-024-03406-8.

References vi

Köhler, D., D. Rügamer et M. Schmid. 2024. *Achieving interpretable machine learning by functional decomposition of black-box models into explainable predictor effects*. ArXiv :2407.18650 [cs, stat], juillet. Visité le 26 août 2024. http://arxiv.org/abs/2407.18650.

Lindeman, R. H., P. F. Merenda et R. Z. Gold. 1980. *Introduction to Bivariate and Multivariate Analysis* [en English]. Scott, Foresman. ISBN : 978-0-673-15099-8. https://books.google.cz/books?id=-hfvAAAAMAAJ.

Lindholm, M., R. Richman, A. Tsanakas et M. V. Wüthrich. 2026. "Sensitivity-based measures of discrimination in insurance pricing". *European Journal of Operational Research*, ISSN : 0377-2217. https://doi.org/https://doi.org/10.1016/j.ejor.2026.01.021. https://www.sciencedirect.com/science/article/pii/S0377221726000433.

Lundberg, S. M., G. G. Erion et S.-I. Lee. 2018. "Consistent individualized feature attribution for tree ensembles". *arXiv preprint arXiv :1802.03888*.

Lundberg, S. M., et S.-I. Lee. 2017. "A Unified Approach to Interpreting Model Predictions". In *Advances in Neural Information Processing Systems*, sous la direction d'I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan et R. Garnett, t. 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf.

Miroshnikov, A., K. Kotsiopoulos, K. Filom et A. R. Kannan. 2024. *Stability theory of game-theoretic group feature explanations for machine learning models*. arXiv : 2102.10878 [cs.GT]. https://arxiv.org/abs/2102.10878.

Ortmann, K. M. 2000. "The proportional value for positive cooperative games". *Mathematical Methods of Operations Research (ZOR)* 51 (2) : 235-248.

References vii

Papadopoulos, H., K. Proedrou, V. Vovk et A. Gammerman. 2002. "Inductive Confidence Machines for Regression" [en en]. In *Machine Learning : ECML 2002,* sous la direction de T. Elomaa, H. Mannila et H. Toivonen, 345-356. Berlin, Heidelberg : Springer. ISBN : 978-3-540-36755-0. https://doi.org/10.1007/3-540-36755-1_29.

Rota, G. C. 1964. "On the foundations of combinatorial theory I. Theory of Möbius Functions". *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 2 (4) : 340-368. ISSN : 1432-2064. https://doi.org/10.1007/BF00531932.

Shapley, L. S. 1951. *Notes on the n-Person Game – II : The Value of an n-Person Game.* Research Memorandum ATI 210720. Santa Monica, California : RAND Corporation.

Sidák, Z. 1957. "On Relations Between Strict-Sense and Wide-Sense Conditional Expectations". *Theory of Probability & Its Applications* 2 (2) : 267-272. ISSN : 0040-585X. https://doi.org/10.1137/1102020. https://epubs.siam.org/doi/abs/10.1137/1102020.

Vasseur, O., M. Claeys-Bruno, M. Cathelinaud et M. Sergent. 2010. "High-dimensional sensitivity analysis of complex optronic systems by experimental design : applications to the case of the design and the robustness of optical coatings". *Chinese Optics Letters* 8(s1) : 21-24.

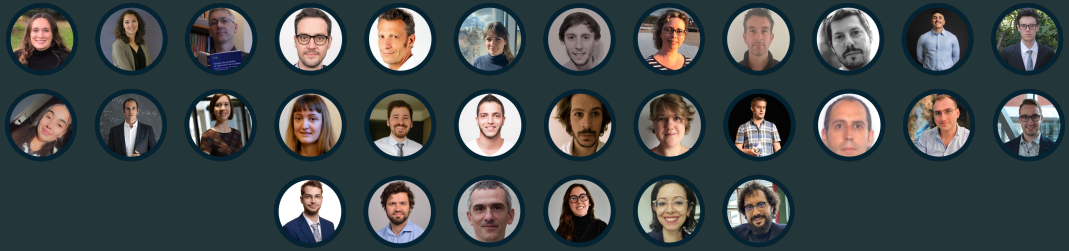
Watson, D., J. O'Hara, N. Tax, R. Mudd et I. Guy. 2023. "Explaining Predictive Uncertainty with Information Theoretic Shapley Values" [en en]. *Advances in Neural Information Processing Systems* 36 (décembre) : 7330-7350. https://proceedings.neurips.cc/paper_files/paper/2023/hash/16e4be78e61a3897665fa01504e9f452-Abstract-Conference.html.

Weber, R. J. 1988. "Probabilistic values for games". Chap. 7 in *The Shapley value : essays in honor of Lloyd S. Shapley,* sous la direction d'A. E. Roth, 101-120. New York, NY : Cambridge University Press.

References viii

Zhang, X., J. Martinelli et S. T. John. 2024. "Challenges in interpretability of additive models". In *Proceedings of the XAI Workshop @ IJCAI 2024*. ArXiv :2504.10169 [cs]. arXiv. https://doi.org/10.48550/arXiv.2504.10169. http://arxiv.org/abs/2504.10169.

Štrumbelj, E., et I. Kononenko. 2014. "Explaining prediction models and individual predictions with feature contributions". *Knowledge and Information Systems* 41, n° 3 (décembre) : 647-665. ISSN : 0219-3116. https://doi.org/10.1007/s10115-013-0679-x. https://link.springer.com/article/10.1007/s10115-013-0679-x.



Merci de votre attention !

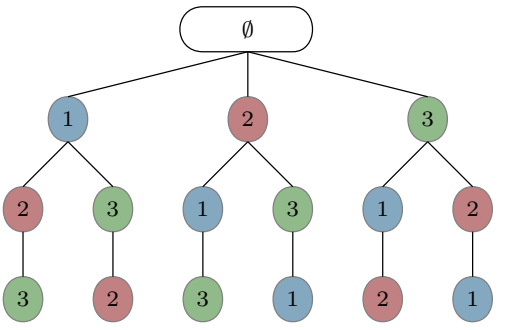
MAROUANEILIDRISSI.COM

Nous reconnaissons le soutien de l'Institut Canadien des Sciences Statistiques (INCASS) et du Conseil de Recherches en Sciences Naturelles et en Génie du Canada (CRSNG)

We acknowledge the support of the Canadian Statistical Sciences Institute (CANSSI) and the Natural Sciences and Engineering Research Council of Canada (NSERC)

Allocations de Weber

Random order allocations



$(1,2,3)$ $(1,3,2)$ $(2,1,3)$ $(2,3,1)$ $(3,1,2)$ $(3,2,1)$ $\mathcal{S}_D$

$$\Delta_\pi(i) := v\left(C_{\pi(i)}(\pi)\right) - v\left(C_{\pi(i)-1}(\pi)\right)$$

$$\phi_1 = \sum_{\pi \in \mathcal{S}_D} p(\pi)\Delta_\pi(1)$$

$$\phi_2 = \sum_{\pi \in \mathcal{S}_D} p(\pi)\Delta_\pi(2)$$

$$\phi_3 = \sum_{\pi \in \mathcal{S}_D} p(\pi)\Delta_\pi(2)$$

If v is **monotonic** (i.e., $\forall B \subseteq A \in \mathcal{P}_D, v(B) \leq v(A)$), every random order allocation is also **nonnegative**.

Projections Obliques

Generated σ-algebras

☞ Denote by σ_X the **σ-algebra generated by X** :

$$\sigma_X = \left\{ X^{-1}[B] \ : \ \forall B \in \bigotimes_{i \in D} \mathcal{E}_i \right\} \subset \mathcal{F}$$

☞ For every $A \in \mathcal{P}_D$, denote by σ_A the **σ-algebra generated by X_A** :

$$\sigma_A = \left\{ X_A^{-1}[B] \ : \ \forall B \in \bigotimes_{i \in A} \mathcal{E}_i \right\} \subset \mathcal{F}$$

☞ Denote by $\sigma_\emptyset$ the **$\mathbb{P}$-trivial σ-algebra** :

$$\sigma_\emptyset = \sigma\left[\{B \in \mathcal{F} \ : \ \mathbb{P}(B) = 0\}\right] \subset \mathcal{F}$$

Measurability and Lebesgue spaces

Lemme. *(Doob-Dynkin)*

Let Y be an $\mathbb{R}$-value random variable, and let X be random inputs.
If Y is σ_X-measurable, then there exists a function $f : E \to \mathbb{R}$ such that

$$Y = f(X) \text{ a.s.}$$

Lemme. *(Kallenberg (2021, Lemma 4.9))*

Let Y be an $\mathbb{R}$-value random variable.
If Y is $\sigma_\emptyset$-measurable, then it is constant a.s.

Définition. *(Lebesgue spaces $\mathbb{L}^2$)*

For a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, denote by $\mathbb{L}^2(\mathcal{G})$ the **Lebesgue space of square-integrable, $\mathbb{R}$-valued, $\mathcal{G}$-measurable** random variables.

It is a **Hilbert space** with the inner product, $\forall Z_1, Z_2 \in \mathbb{L}^2(\mathcal{G})$:

$$\langle Z_1, Z_2 \rangle = \mathbb{E}[Z_1 Z_2] = \int_\Omega Z_1(\omega) Z_2(\omega) d\mathbb{P}(\omega)$$

Non-perfect functional dependence

Assumption 1 *(Non-perfect functional dependence)*

- $\sigma_\emptyset \subset \sigma_i$, $i = 1, \ldots, d$ (inputs are not constant).
- For $B \subset A$, $\sigma_B \subset \sigma_A$ (inputs add information).
- For every $A, B \in \mathcal{P}_D$, $A \neq B$, $\sigma_A \cap \sigma_B = \sigma_{A \cap B}$.

☞ $\mathbb{L}^2(\sigma_\emptyset) \subset \mathbb{L}^2(\sigma_A)$, for every $A \in \mathcal{P}_D$
There are non-constant random variables in the Lebesgue spaces

☞ For $B \subset A$, $\mathbb{L}^2(\sigma_B) \subset \mathbb{L}^2(\sigma_A)$
There are functions of X_A that are not functions of X_B

☞ For any $A, B \in \mathcal{P}_D$, $\mathbb{L}^2(\sigma_A) \cap \mathbb{L}^2(\sigma_B) = \mathbb{L}^2(\sigma_{A \cap B})$
The functions of X_A and X_B are functions of $X_{A \cap B}$

Proposition. Under Assumption 1, for any $A, B \in \mathcal{P}_D$ such that $A \cap B \notin \{A, B\}$, **there is no mapping T such that** $X_B = T(X_A)$ a.e.

Non-perfect stochastic dependence

Définition. *(Friedrichs (1937) angle)*The cosine of Friedrichs' angle is defined as

$$c(M,N) := \sup \left\{ |\langle x, y \rangle| : \begin{cases} x \in M \cap (M \cap N)^{\perp}, \|x\| \leq 1 \\ y \in N \cap (M \cap N)^{\perp}, \|y\| \leq 1 \end{cases} \right\},$$

where the orthogonal complement is taken w.r.t. to H.

☞ Analogous to the **maximal partial dependence** between random elements (Bryc 1984, 1996; Dauxois, Nkiet et Romain 2004)

Définition. *(Feshchenko matrix)*Let Δ be the $(2^d \times 2^d)$, symmetric **set-indexed** matrix, defined element-wise, $\forall A, B \in \mathcal{P}_D$ as

$$\Delta_{AB} = \begin{cases} 1 & \text{if } A = B; \\ -c\left(\mathbb{L}^2(\sigma_A), \mathbb{L}^2(\sigma_B)\right) & \text{otherwise.} \end{cases}$$

Assumption 2 *(Non-degenerate stochastic dependence)*The Feshchenko matrix Δ of X is definite-positive.

Direct-sum decomposition

Définition. *Direct-sum decomposition (Axler 2015)*

Let H be a vector space and let $H_1, \ldots, H_n$ be proper subspaces of H.

H is said to admit a **direct-sum decomposition** if any $h \in H$ can be written **uniquely** as

$$h = \sum_{i=1}^{n} h_i \text{ where } h_i \in H_i \text{ for } i = 1, \ldots, n.$$

In this case, we write :

$$H = \bigoplus_{i=1}^{n} H_i.$$

Théorème. *Sidák (1957, Theorem 2)* Let $\mathcal{G}_1, \mathcal{G}_2 \subseteq \mathcal{F}$, then

- If $\mathcal{G}_1 \subset \mathcal{G}_2$, then $\mathbb{L}^2(\mathcal{G}_1) \subset \mathbb{L}^2(\mathcal{G}_2) \subseteq \mathbb{L}^2(\mathcal{F})$;
- $\mathbb{L}^2(\mathcal{G}_1) \cap \mathbb{L}^2(\mathcal{G}_2) = \mathbb{L}^2(\mathcal{G}_1 \cap \mathcal{G}_2)$.

Goal : Find a direct-sum decomposition of $\mathbb{L}^2(\sigma_X)$ w.r.t. the subspaces $\{\mathbb{L}^2(\sigma_A)\}_{A \in \mathcal{P}_D}$
Any real-valued $G(X)$ could be **uniquely** decomposed as a sum of **functions** of X_A, $A \in \mathcal{P}_D$

Generalized Hoeffding decomposition

Théorème.

Under Assumptions 1 and 2, for every $A \in \mathcal{P}_D$, one has that

$$\mathbb{L}^2(\sigma_A) = \bigoplus_{B \in \mathcal{P}_A} V_B.$$

where $V_\emptyset = \mathbb{L}^2(\sigma_\emptyset)$, and

$$V_B = \left[\sum_{C \in \mathcal{P}_B, C \neq B} V_C \right]^{\perp_B},$$

where $\perp_B$ denotes the orthogonal complement in $\mathbb{L}^2(\sigma_B)$.

Intuition of the proof :

Inductive functional centering

Intuition of the proof : One input

One input :

1. Let $i \in D$, and **fix** $\mathbb{L}^2(\sigma_i)$ **as the ambient space**
2. We have that $V_\emptyset := \mathbb{L}^2(\sigma_\emptyset)$ **is a closed subspace of** $\mathbb{L}^2(\sigma_i)$ (it is **complemented**)
3. Denote $V_i = [V_\emptyset]^{\perp_i}$, **the orthogonal complement of** $V_\emptyset$ **in** $\mathbb{L}^2(\sigma_i)$
4. One has that $\mathbb{L}^2(\sigma_i) = V_\emptyset \oplus V_i$

We just showed that any $f(X_i) \in \mathbb{L}^2(\sigma_i)$ can be written as

$$f(X_i) = \underbrace{\mathbb{E}[f(X_i)]}_{\in V_\emptyset} + \underbrace{\mathbb{E}[f(X_i) - \mathbb{E}[f(X_i)]]}_{\in V_i = \mathbb{L}^2_0(\sigma_i)}$$

And note that $\mathbb{L}^2(\sigma_i) = V_\emptyset \oplus V_i$ **hold for any** $i \in D$ **(induction)**

Intuition of the proof : Two inputs

Two inputs :

1. Let $i, j \in D$, and **fix** $\mathbb{L}^2(\sigma_{ij})$ **as the ambient space**
2. **Assumptions 1 and 2 imply that** $\mathbb{L}^2(\sigma_i) + \mathbb{L}^2(\sigma_j)$ **is closed in** $\mathbb{L}^2(\sigma_{ij})$ (it is **complemented**)
3. Notice **(previous step)** that $\mathbb{L}^2(\sigma_i) + \mathbb{L}^2(\sigma_j) = V_\emptyset + V_i + V_j$
4. Denote $V_{ij} = \left[V_\emptyset + V_i + V_j\right]^{\perp_{ij}}$, **the orthogonal complement in** $\mathbb{L}^2(\sigma_{ij})$
5. We thus have that $\mathbb{L}^2(\sigma_{ij}) = V_\emptyset + V_i + V_j + V_{ij}$

And note that the decomposition hold for any pair $i, j \in D$
We **"centered" a bivariate function from its "univariate and constant parts"**

We continue the induction up to d **inputs.**

Orthocanonical decomposition

Corollaire. *(Orthocanonical decomposition)*

Suppose that Assumptions 1 and 2 hold.

Then, any random variable $G(X) \in \mathbb{L}^2(\sigma_X)$ can be **uniquely decomposed** as

$$G(X) = \sum_{A \in \mathcal{P}_D} G_A(X_A),$$

where each $G_A(X_A) \in V_A$.

The term **"orthocanonical"** comes from **the choice** of the **orthogonal complement** in the "centering process"

The subspaces V_A contain **random variables that are functions of exactly** X_A
If one of their element can be expressed with fewer inputs, it is necessarily equal to 0

Is it possible to characterize the *representants* $G_A(X_A)$?

Projectors

☞ Let H be a Hilbert space, and let $P : H \to H$ be an operator
☞ $\mathrm{Ran}\,(P)$ is the range of P, and $\mathrm{Ker}\,(P)$ is its nullspace

If P is an idempotent ($P \circ P = P$), linear, and bounded operator, it is then called the **oblique projector onto $\mathbf{Ran}\,(P)$ parallel to $\mathbf{Ker}\,(P)$** and

$$H = \mathrm{Ran}\,(P) \oplus \mathrm{Ker}\,(P)$$

Conversely, suppose that

$$H = M \oplus N$$

then there exists a projector P such that $\mathrm{Ran}\,(P) = M$ and $\mathrm{Ker}\,(P) = N$ (Galántai 2004)
☞ P is called the **canonical oblique projector** w.r.t. to this direct-sum decomposition of H

If in addition $N = M^{\perp}$ (P is self-adjoint), then P is called the **orthogonal projector** onto M
☞ The orthogonal projector onto $\mathbb{L}^2(\sigma_A)$ is the conditional expectation $\mathbb{E}[. \mid \sigma_A]$

Orthocanonical projectors

From the direct-sum decomposition of $\mathbb{L}^2(\sigma_X)$:

$$G(X) = \sum_{A \in \mathcal{P}_D} G_A(X_A).$$

Oblique projection onto V_A

The operator

$$Q_A : \mathbb{L}^2(\sigma_X) \to \mathbb{L}^2(\sigma_X), \quad G(X) \mapsto G_A(X_A).$$

Q_A is the (canonical) **oblique projection** with

$$\mathrm{Ran}(Q_A) = V_A \text{ , and } \mathrm{Ker}(Q_A) = \bigoplus_{B \in \mathcal{P}_D : B \neq A} V_B$$

Orthogonal projections onto V_A

The projector

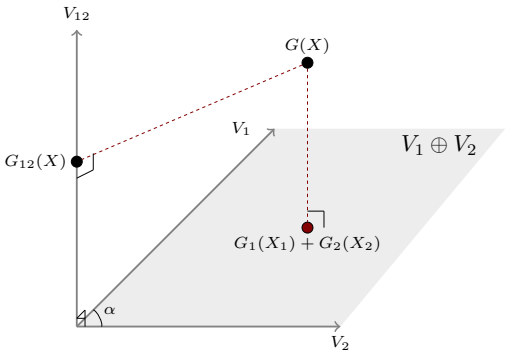
$$P_A : \mathbb{L}^2(\sigma_X) \to \mathbb{L}^2(\sigma_X), \text{ with } \mathrm{Ran}(P_A) = V_A \text{ and } \mathrm{Ker}(P_A) = [V_A]^{\perp}$$

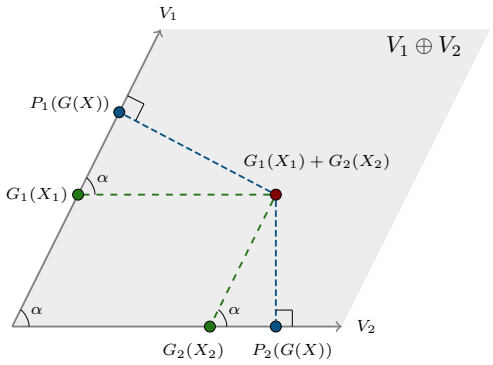
is the **orthogonal projection** onto V_A.

Illustration* $\mathbb{L}_0^2(\sigma_{12})$

Hence, for any $G(X) \in \mathbb{L}^2(\sigma_X)$, one has that, $\forall A \in \mathcal{P}_D$

$$G_A(X_A) = Q_A(G(X)).$$





The oblique projection Q_A usually differ from the orthogonal projections P_A

Oblique and orthogonal projections

Proposition.

Under Assumptions 1 and 2,

$$P_A(G(X)) = Q_A(G(X)) \text{ a.s. }, \forall A \in \mathcal{P}_D \quad \iff \quad X \text{ is mutually independent.}$$

This comes from the fact that **the subspaces V_A are all pairwise orthogonal if and only if the inputs are mutually independent**

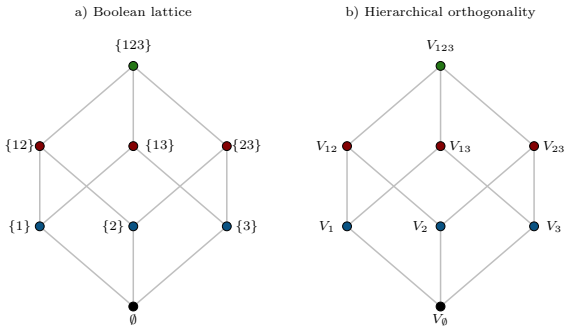
But, under Assumptions 1 and 2, they **may not be all orthogonal**

There is a more visual way to illustrate that

Boolean lattice and hierarchical orthogonality

Our decomposition is **over the power-set $\mathcal{P}_D$, and this is not trivial**

☞ Endowed with the **binary relation** $\subseteq$, $(\mathcal{P}_D, \subseteq)$ forms **a Boolean lattice**



The subspaces $\{V_A\}_{A \in \mathcal{P}_D}$ are **hierarchically orthogonal by design**

☞ They form the same algebraic structure **w.r.t. to** $\perp$

More projectors

Recall that :

- Q_A is the **canonical oblique projection** onto V_A
- P_A is the **orthogonal projection** onto V_A

But we're more familiar with projections onto $\mathbb{L}^2(\sigma_A)$...
☞ Conditional expectation operators, for example

(Canonical) oblique projection onto $\mathbb{L}^2(\sigma_A)$:

$$\mathbb{M}_A : \mathbb{L}^2(\sigma_X) \to \mathbb{L}^2(\sigma_X), \quad G(X) \mapsto \sum_{B \in \mathcal{P}_A} G_B(X_B)$$

Orthogonal projection onto $\mathbb{L}^2(\sigma_A)$:

$$\mathbb{E}_A : \mathbb{L}^2(\sigma_X) \to \mathbb{L}^2(\sigma_X), \text{ with } \mathrm{Ran}(\mathbb{E}_A) = \mathbb{L}^2(\sigma_A) \text{ and } \mathrm{Ker}(P_A) = \mathbb{L}^2(\sigma_A)^{\perp},$$

a.k.a **the conditional expectation w.r.t. to X_A** (i.e., $\mathbb{E}[. \mid X_A]$).

Can we characterize Q_A w.r.t. $\mathbb{M}_A$?

Generalized Möbius inversion

Because $(\mathcal{P}_D, \subseteq)$ forms a **Boolean lattice**, yes!

Corollaire. *(Möbius inversion on power-sets (Rota 1964))*

For any two set functions :

$$f : \mathcal{P}_D \to \mathbb{A}, \quad g : \mathcal{P}_D \to \mathbb{A},$$

valued in an abelian group $\mathbb{A}$, the following equivalence holds :

$$f(A) = \sum_{B \in \mathcal{P}_A} g(B), \quad \forall A \in \mathcal{P}_D \quad \Longleftrightarrow \quad g(A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} f(B), \quad \forall A \in \mathcal{P}_D.$$

☞ Analogous to the *inclusion-exclusion principle*

In our case, we have, **by definition of the oblique projection onto** $\mathbb{L}^2(\sigma_A)$, that

$$\mathbb{M}_A(G(X)) = \sum_{B \in \mathcal{P}_A} G_B(X_B), \quad \forall A \in \mathcal{P}_D,$$

which is equivalent to

$$G_A(X_A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} \mathbb{M}_B(G(X)), \quad \forall A \in \mathcal{P}_D$$

Generalized Hoeffding decomposition

Hoeffding (1948) found that **for mutually independent inputs** :

$$G_A(X_A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} \mathbb{E}_B [G(X)], \quad \forall A \in \mathcal{P}_D$$

Under Assumptions 1 and 2, we have that :

$$G_A(X_A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} \mathbb{M}_B [G(X)], \quad \forall A \in \mathcal{P}_D$$

In addition :

Proposition.

Under Assumptions 1 and 2,

$$\mathbb{M}_A [G(X)] = \mathbb{E}_A [G(X)] \text{ a.s. }, \forall A \in \mathcal{P}_D \iff X \text{ is mutually independent.}$$

Our approach generalizes Hoeffding's original decomposition!

Effets Proportionnels Marginaux

Proportional values

The **proportional values** (Ortmann 2000) can be interpreted as a redistribution such that

"[...] *each player gains in* ***equal proportion*** *to that which could be obtained by each alone.*" - B. Feldman (1999)

They are based on a **proportional allocation principle** for **positive games**.

If $\forall A \in \mathcal{P}_D, v(A) > 0$, the choice of p is :

$$p(\pi) = \frac{L(\pi)}{\sum_{\sigma \in \mathcal{S}_D} L(\sigma)}, \quad L(\pi) = \exp\left(- \sum_{j \in D} \log\left(v\left(C_j(\pi) \right) \right) \right)$$

They are uniquely caracterized as the only efficient allocation that respects the **equal proportional gains axiom** :

$$\forall i, j \in A \subseteq D, \ \frac{\mathrm{PV}_i(A, v)}{\mathrm{PV}_i(A_{-j}, v)} = \frac{\mathrm{PV}_j(A, v)}{\mathrm{PV}_j(A_{-i}, v)} \tag{1}$$

Proportional values extension

Théorème. *(PV extension to monotonic nonnegative games)*Let (D, v) be a nonnegative and monotonic game with value function $v : \mathcal{P}(D) \to \mathbb{R}^+$. Denote $\mathcal{K}$ the set of largest (w.r.t. their cardinality) zero coalitions, i.e., $\mathcal{K} = \underset{A \in \mathcal{P}(D)}{\operatorname{argmax}} \{|A| : v(A) = 0\}$. Additionally, the sets of largest zero coalitions that do not contain $i \in D$ is denoted by $\mathcal{K}_{-i}$, i.e., $\mathcal{K}_{-i} = \underset{A \in \mathcal{P}(D)}{\operatorname{argmax}} \{|A| : v(A) = 0, i \notin A\}$. Define, for any $A \in \mathcal{K}$, the positive set function :

$$\begin{aligned} v_A : \mathcal{P}(D \setminus A) &\to \mathbb{R}_*^+ \\ B &\mapsto v(B \cup A). \end{aligned}$$

Let $\mathrm{PV}^0((D, v)) = \left(\mathrm{PV}_1^0, \ldots, \mathrm{PV}_d^0\right)$ be the allocation defined as :

$$\mathrm{PV}_i^0 = \frac{\sum_{A \in \mathcal{K}_{-i}} R(D_{-i} \setminus A, v_A)^{-1}}{\sum_{A \in \mathcal{K}} R(D \setminus A, v_A)^{-1}} \quad \text{if } \mathcal{K}_{-i} \neq \emptyset \text{ and } \quad \mathrm{PV}_i^0 = 0 \text{ otherwise.} \tag{2}$$

Then, PV^0 is a continuous extension of PV to the set of nonnegative monotonic games, i.e., for a positive monotonic game (D, v),

$$\mathrm{PV}^0((D, v)) = \mathrm{PV}((D, v)).$$

Exogeneity

Définition. *L^2-exogeneity*

Let $X = (X_1, \dots, X_d)$ be random inputs of a model $G : \mathbb{R}^d \mapsto \mathbb{R}$ such that $Y = G(X)$, with Y the random output. Let $i \in D$. The random input X_i is said to be L^2-exogenous to G if, $\exists f \in L^2(P_{X_{D_{-i}}})$ such that $Y = f(X_{D_{-i}})$ a.s..

Moreover, if for $E \in \mathcal{P}(D)$, $\exists f \in L^2(P_{X_{D_{\overline{E}}}})$ such that $Y = f(X_{\overline{E}})$ a.s. then X_E is said to form an L^2-exogenous vector.

Assumption Let $E \in \mathcal{P}(D)$. If for every $i \in E$, X_i is exogenous, then X_E forms an exogenous vector.

Prévision conforme

Conformal prediction

Définition. Random variables $X_1, \dots, X_n$ are exchangeable if

$$(X_1, \dots, X_n) \overset{\mathcal{L}}{=} (X_{\pi_1}, \dots, X_{\pi_n})$$

for any permutation π of $\{1, \dots, n\}$.

Proposition. For any $z_1, \dots, z_n, t \in \mathbb{R}$, and quantile level $\tau \in [0, 1]$

$$\{t \leq q_\tau(z_1, \dots, z_n, t)\} \iff \left\{t \leq q_{\tau \frac{n+1}{n}}\right\}$$

Proposition. If $X_1, \dots, X_n$ are exchangeable, then for $i = 1, \dots, n$

$$\mathbb{P}(X_i \leq q_\tau(X_1, \dots, X_n)) \geq \alpha$$

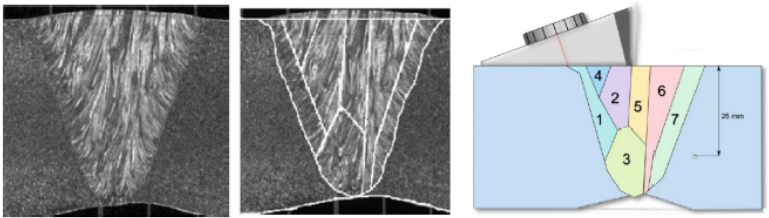
Cas d'application

Ultrasonic control of a weld

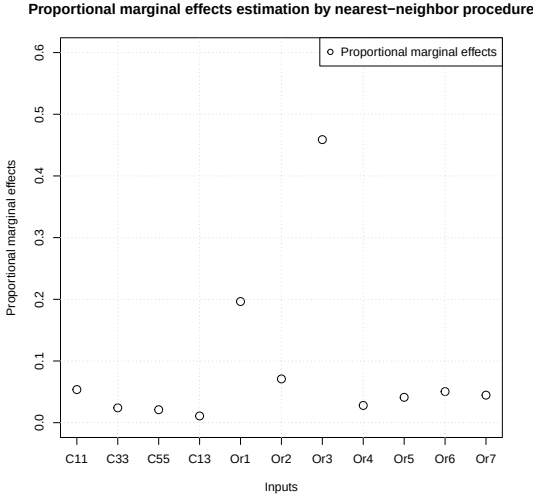
Non-destructive control of a weld defect, using the *ATHENA2D* numerical code (looss et Prieur 2019).

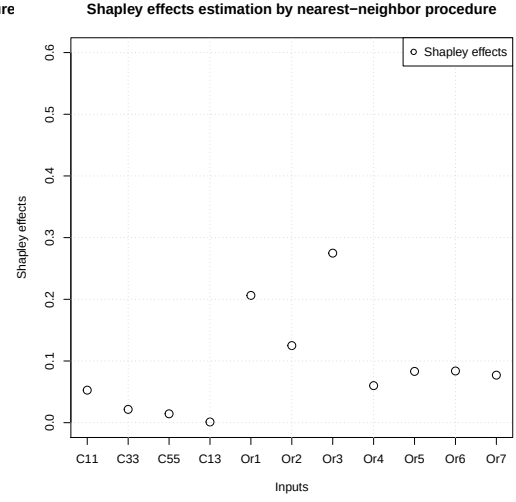
- 11 input variables :
  - 4 elastic coefficients related to the welding material ;
  - 7 columnar grain orientation relative to the 7 different zones.
- Output : wave amplitude after the weld defect ultrasonic inspection.

The inputs are assumed to be **Gaussian**, and **the 7 columnar grain orientation** are **highly correlated**.



Ultrasonic control of a weld





Acoustic Fire Extinguisher

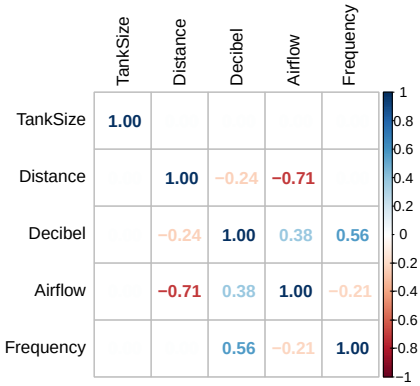
15390 experiments of sound wave fire extinguishing.

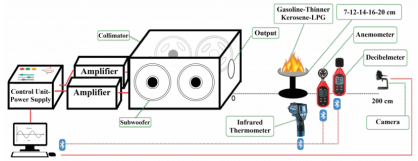
Classification task on 6 variables measured during the experiments.

- Tank Size (L)
- Fuel (Kerosene, Gasoline, Thinner)
- Fire source distance (m)
- Decibel
- Airflow
- Sound frequency

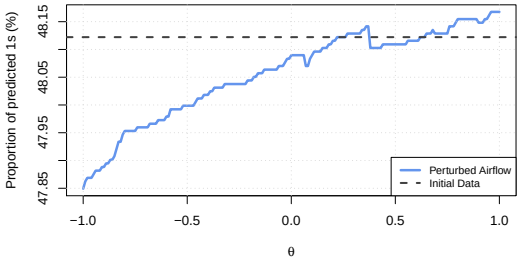
Black-box model : 1-layer neural network (Koklu et Taspinar 2021) trained with an accuracy of 95.15% (validation accuracy of 94.26%).

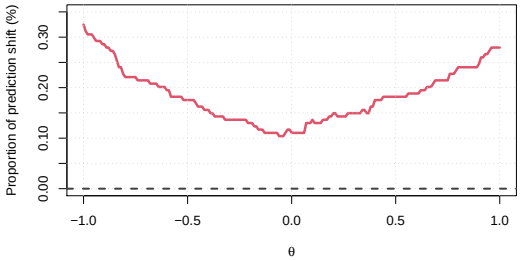
Perturbation scheme : shift of the Airflow 0.8-quantile : initial value at 12, shift between 9.5 ($\theta = -1$) and 14.5 ($\theta = 1$) by polynomial perturbation approximation of degree 9.

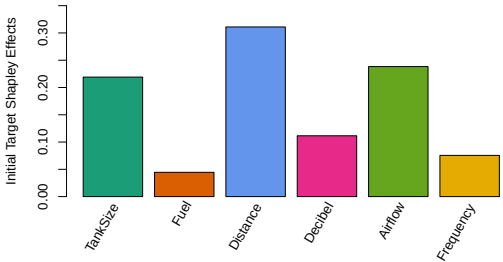


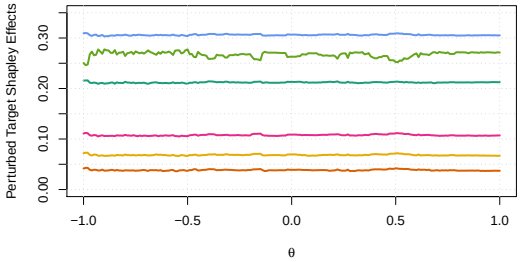


Global robustness









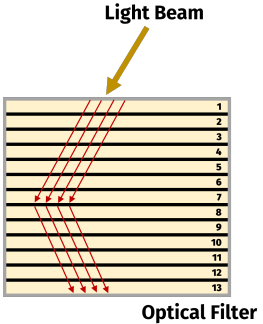
Optical filter transmittance - Feature selection

Transmittance performance of an **optical filter** composed of 13 consecutive layers (Vasseur et al. 2010).

The inputs $I_1, \ldots, I_{13}$ represent the **refractive index error** of each filter ($\mathcal{U}([-0.05, 0.05])$)

These errors are (highly) correlated due to the manufacturing process (Gaussian copula, $\rho = 0.95$).

The numerical model computes the **transmittance error w.r.t. the "perfect filter"** over several wavelengths.



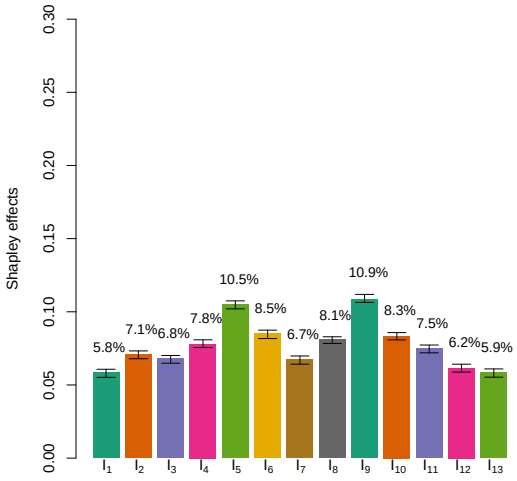
☞ **We only have access to an i.i.d. input-output sample** ($n = 1000$).

The indices are computed using a **nearest-neighbors approach** (Broto, Bachoc et Depecker 2020).

Parallelized implementation using the R package `sensitivity` ($\sim$ 4min runtine, 8 cores).

Arbitrarily chosen number of neighbors : 6.

Optical filter transmittance - Feature selection



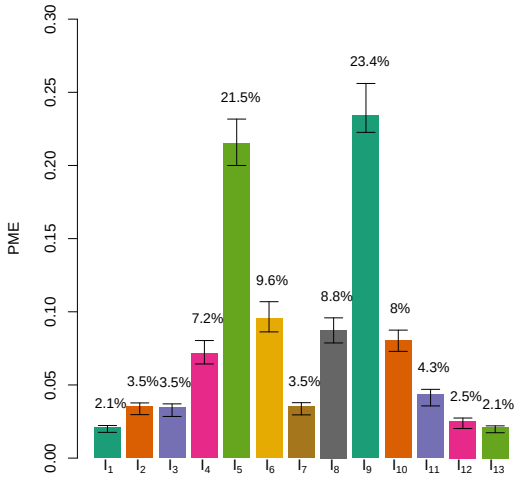
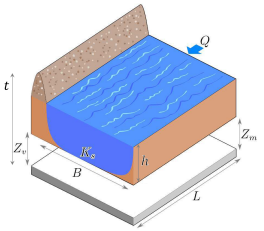


Illustration : River Water Level

Simplified **numerical model of a river water level** (Iooss et Lemaître 2015).

$$Y = Z_v + \left(\frac{Q}{BK_s \sqrt{\frac{Z_m - Z_v}{L}}} \right)^{3/5}$$

- Q : River maximum annual water flow rate.
- K_s **: Strickler riverbed roughness coefficient.**
- Z_v : Downstream river level.
- Z_m : Upstream river level.
- L : River length.
- B : River width.

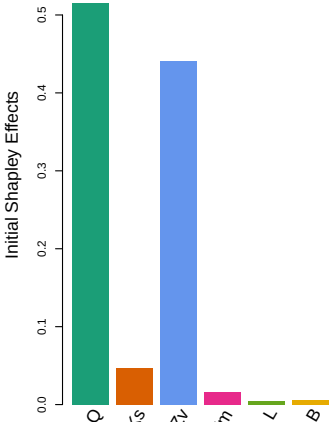


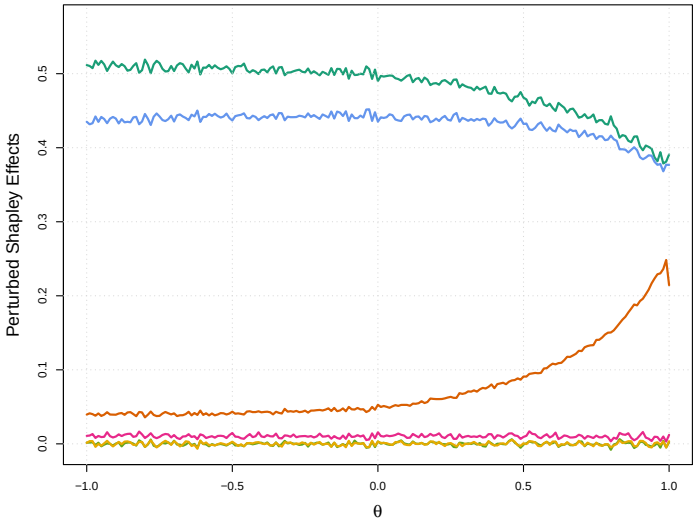
Structure probabiliste :

Input	Distribution	Support
Q	$\mathcal{G}(1013, 558)$ trunc.	[500, 3000]
K_s	$\mathcal{N}(30, 7)$ **trunc.**	[20, 50]
Z_v	$\mathcal{T}(49, 50, 51)$	[49, 51]
Z_m	$\mathcal{T}(54, 55, 56)$	[54, 56]
L	$\mathcal{T}(4990, 5000, 5010)$	[4990, 5010]
B	$\mathcal{T}(295, 300, 305)$	[295, 305]

The **inputs are correlated** by means of a Gaussian copula : $\rho(Q, K_s) = 0.5$ and $\rho(Z_v, Z_m) = \rho(L, B) = 0.3$.

Effects of the perturbation on the importance quantification





COVID-19 epidemiological model

Goal : quantify which uncertain parameters drive the risk of **ICU bed shortage**.

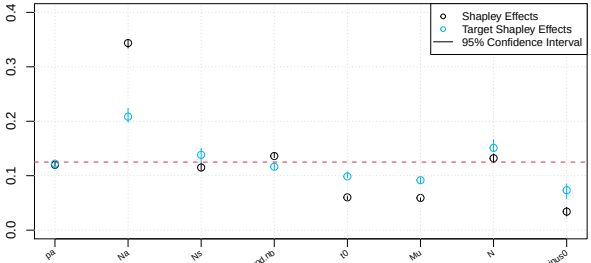
p_a	Probability of being *asymptomatic or mild* given infection (drives hidden transmission and detection).
N_a	Days until recovery if *asymptomatic* (infectious period length).
N_s	Days until recovery if *symptomatic* without hospitalization (infectious period length).
R_0	Basic reproduction number (initial transmissibility level).
t_0	Epidemic start date (initialization / alignment with observed wave).
μ	Decay rate of transmission after interventions (strength/speed of mitigation).
N	Date of effect of distancing/lockdown measures (timing of mitigation).
I_0^-	Initial number of infected *undetected* (hidden seeding of the wave).
k	ICU capacity threshold (beds available ; defines "shortage" event).

We have access to the data generated according to the posterior distribution, after a first screening step ($n = 5000$).

More details on the SIR model, and the calibrated compartmental model in **DaVeiga202empty citation**.

COVID-19 : Target Shapley effects for ICU shortage

	pa	Na	Ns	R0	t0	Mu	N	Iminus0
pa	1	−0.19	0.03	0.23	0.01	−0.18	−0.24	0.41
Na	−0.19	1	−0.53	0.45	0.04	0.38	0.04	0.14
Ns	0.03	−0.53	1	−0.1	0.06	0.03	0.12	0.14
R0	0.23	0.45	−0.1	1	0.11	0.04	−0.66	−0.36
t0	0.01	0.04	0.06	0.11	1	0.02	−0.06	−0.01
Mu	−0.18	0.38	0.03	0.04	0.02	1	0.69	0.22
N	−0.24	0.04	0.12	−0.66	−0.06	0.69	1	0.55
Iminus0	0.41	0.14	0.14	−0.36	−0.01	0.22	0.55	1



CT dose estimation (NCICT + PACS) : inputs, output, data

Goal : rank the main drivers of uncertainty in organ dose estimates ($d = 8$).

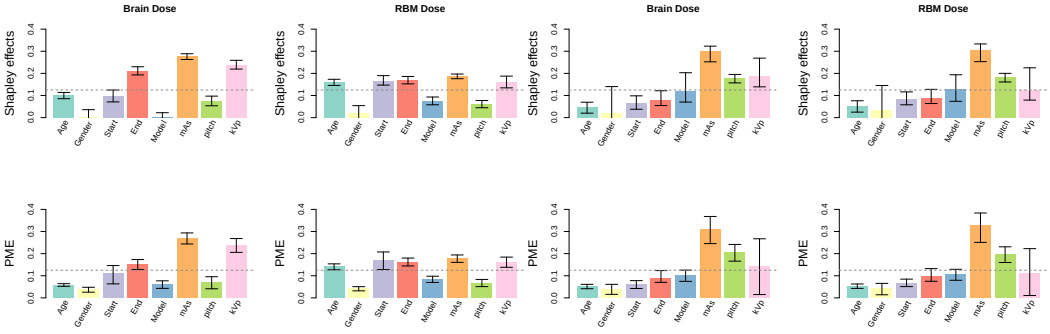
Input	Type	Meaning (PACS / acquisition)
Age	Disc.	Patient age (phantom selection).
Gender	Cat.	Patient sex (phantom selection).
Start / End	Disc.	Scan landmarks (scanned length).
mAs	Disc.	Tube current–time product (dose scale).
kvp	Disc.	Tube potential (beam energy).
Pitch	Cont.	Helical pitch ($CTDI_{vol}$ normalization).
Model	Cat.	Scanner model (CTDI library / hardware).

Output (mGy) : *estimated absorbed dose in the target organ (e.g., brain, RBM) for a given CT scan, obtained by aggregating slice-by-slice dose contributions over the scanned region.*

Data : $n = 8848$ CT images (PACS), France 2005–2014 + Spain 2000–2004.

Context : French pediatric CT cohort (**Bernier:2012**). Dose engine : NCICT v1.2 (**Lee:2015**) (restricted : *DC* treated as fixed).

CT dose estimation



Left : Head dose absorption ; Right : Chest dose absorption