Quantile oriented sensitivity analysis: why and how?

Kevin Elie-dit-Cosaque*, Véronique Maume-Deschamps, Clémentine Prieur**, Ri Wang.

Institut Camille Jordan, Lyon 1 Université
* SCOR, Paris
** Laboratoire Jean Kuntzmann, Université Grenoble Alpes

ENBIS 2026

Plan

Plan

Variance based indices

$\mathbf{X} = (X_1, \dots, X_d)$ a random vector of input variables, $Y = f(\mathbf{X})$ an output model. If the X_i's are independent, the Sobol indices, Sobol (1993)[1], quantify the impact of the X_i's on the variability of Y:

$$S_i = \frac{\mathrm{Var}\left(\mathbb{E}[Y|X_i]\right)}{\mathrm{Var}(Y)}.$$

Sensitivity is considered only with respect to the mean (since it involves variances). What about other interesting quantities such as quantile?

[1] IM Sobol' (1993). In: *Math. Model. Comput. Exp.*

From Sobol indices to QOSA

Sobol indices rewrite:

$$S_i = \frac{\mathrm{Var}\left(\mathbb{E}[Y|X_i]\right)}{\mathrm{Var}(Y)}$$

$$S_i = \frac{\mathbb{E}\left[(Y-\mathbb{E}[Y])^2\right] - \mathbb{E}\left(\mathbb{E}\left[(Y-\mathbb{E}[Y|X_i])^2 \,|X_i\right]\right)}{\mathbb{E}\left[(Y-\mathbb{E}[Y])^2\right]}$$

$$S_i = \frac{\min_\theta \mathbb{E}\left[(Y-\theta)^2\right] - \mathbb{E}\left(\min_\theta \mathbb{E}\left[(Y-\theta)^2 \,|X_i\right]\right)}{\min_\theta \mathbb{E}\left[(Y-\theta)^2\right]}$$

Quantile oriented sensitivity analysis

QOSA: Quantile Oriented Sensitivity Analysis index: (Fort *et al.* 2016)

$$S_i^\alpha = \frac{\min_{\theta \in \mathbb{R}} \mathbb{E}\left[\psi_\alpha(Y,\theta)\right] - \mathbb{E}\left[\min_{\theta \in \mathbb{R}} \mathbb{E}\left[\psi_\alpha\left(Y,\theta\right)|X_i\right]\right]}{\min_{\theta \in \mathbb{R}} \mathbb{E}\left[\psi_\alpha(Y,\theta)\right]}$$

$$S_i^\alpha = \frac{\mathbb{E}\left[\psi_\alpha\left(Y, q_\alpha(Y)\right)\right] - \mathbb{E}\left[\psi_\alpha\left(Y, q_\alpha(Y|X_i)\right)\right]}{\mathbb{E}\left[\psi_\alpha(Y, q_\alpha(Y))\right]}$$

with the contrast function $\psi_\alpha : (y,\theta) \mapsto (y-\theta)(\alpha - \mathbf{1}_{y\leq\theta})$, $\alpha \in [0,1]$. ψ_α is like an asymetric distance with the property that $\arg\min_{\theta\in\mathbb{R}} \mathbb{E}\left[\psi_\alpha\left(Y,\theta\right)|X_i\right] = q_\alpha(Y|X_i)$ (α-level quantile).
Properties of S_i^α: $0 \leq S_i^\alpha \leq 1$,
$S_i^\alpha = 0 \iff Y$ and X_i are independent; $S_i^\alpha = 1 \iff Y$ is X_i measurable.

Limits of QOSA indices

- Not automatically normalized (sum less than one for additive model with independent inputs)
- No intrinsic decomposition
- Interpretation not so easy and meaningless in case of dependencies.

$\Longrightarrow$ proposition of QOSA-Shpaley effects or QOSE.

Variance based Shapley effects

$$v^i = \sum_{u \subseteq \{1,\ldots,d\} \setminus \{i\}} \frac{(d-|u|-1)!|u|!}{d!} \left(c\left(u \cup \{i\}\right) - c\left(u\right) \right) ,$$

$c(\cdot)$ is a generic cost function which maps the exploratory power generated by each subset $u \subseteq \{1, \ldots, d\}$. Variance-based sensitivity measures: Shapley effect of the i-th input, Sh_i defined using the cost function[2]: $c(u) = \mathbb{E}\left[\mathrm{Var}\left[Y|\mathbf{X}_{-u}\right]\right] / \mathrm{Var}(Y)$[3];

$Sh_i \geqslant 0$ and $\sum_{i=1}^{d} Sh_i = 1$.

[2] Art B Owen (2014). In: *SIAM/ASA Journal on Uncertainty Quantification* Eunhye Song, Barry L Nelson, and Jeremy Staum (2016). In: *SIAM/ASA Journal on Uncertainty Quantification*

[3] Or equivalently $\tilde{c}(u) = \mathrm{Var}\left[\mathbb{E}\left[Y|\mathbf{X}_u\right]\right] / \mathrm{Var}(Y)$

A cost function for quantile oriented sensitivity analysis

Quantile based cost function option: consider[4]

$$c(u) = \frac{\mathbb{E}\left[\psi_\alpha\left(Y, \; q_\alpha\left(Y | \mathbf{X}_{-u}\right)\right)\right]}{\mathbb{E}\left[\psi_\alpha(Y, q_\alpha(Y))\right]}.$$

$c(u)$ has non negative increments: $c\left(u \cup \{i\}\right) - c\left(u\right) \geq 0$, use it as a cost function $\Longrightarrow$ QOSE indices denoted S_i^α.

Fairly distributes input effects.

Should be usefull to better understand blackbox models such as machine learning tools.

Can be interpreted as a percentage of the global cost $\mathbb{E}\left[\psi_\alpha(Y, q_\alpha(Y))\right]$.

[4] Kévin Elie-Dit-Cosaque and Véronique Maume-Deschamps (2022). In: *SIAM/ASA Journal on Uncertainty Quantification*

Plan

Estimating QOSA / QOSE

Estimating Sobol' index may avoid the estimation of the conditional distribution by using $\mathrm{Var}\left(\mathbb{E}[Y|X_i]\right) = \mathrm{Cov}(Y, Y')$ with

$$Y' = f(\mathbf{X}'),\ \mathbf{X}' = (X'_1, \ldots, X'_{i-1}, X_i, X'_{i+1}, \ldots, X'_n)$$

X'_j independent copy of X_j.
No such formula for $\mathbb{E}\left[\psi_\alpha\left(Y,\ q_\alpha\left(Y \middle| \mathbf{X}_{-u}\right)\right)\right]$.

Estimating QOSA / QOSE

The estimation of QOSA / QOSE' indices requires to estimate the conditional distribution $Y|X_i$. For QOSA,

- Kernel methods[5] optimal window width difficult to calibrate, requires a large number of calls to the costly function f.
- Random Forest method Less calls to f, time consuming nevertheless[6].
- We combine quantile oriented random forests and projection[7], works also for QOSE.

---

[5] Véronique Maume-Deschamps and Ibrahima Niang (2018). In: *Statistics & Probability Letters*

Thomas Browne et al. (2017). In:

[6] Kevin Elie-Dit-Cosaque and Véronique Maume-Deschamps (2024). In: *Computational Statistics*

[7] Ri Wang, Véronique Maume-Deschamps, and Clémentine Prieur (2026). In: *Journal of Statistical Planning and Inference*

Véronique Maume-Deschamps et al. (2026). Work in progress

Quantile random forests

From a random sample $(\mathbf{X}^j, Y^j)$, $j = 1, \dots, n$, recursive construction of a tree.
For a node A of size n_A, split on (k, z) that minimizes[8]:

$$\Delta_{k,z} = \frac{n_L}{n_A} \cdot \frac{1}{n_L} \sum_{j \in A_L} \psi_\alpha(Y^j, \hat{q}_L) + \frac{n_R}{n_A} \cdot \frac{1}{n_R} \sum_{j \in A_R} \psi_\alpha(Y^j, \hat{q}_R) \text{ where:}$$

$A_L = \{\mathbf{X}^j \in A : X_k^j \leq z\}$ is of size n_L and $A_R = \{\mathbf{X}^j \in A : X_k^j > z\}$ is of size n_R;
$\hat{q}_L, \hat{q}_R$ are empirical α-quantiles of Y in A_L, A_R.
It maximizes the empirical counterpart of
$\mathbb{E}[\psi_\alpha(Y, q_A)|\mathbf{X} \in A] - \mathbb{P}[\mathbf{X} \in A_L|X \in A]\mathbb{E}[\psi_\alpha(Y, q_L)|\mathbf{X} \in A_L] - \mathbb{P}[\mathbf{X} \in A_R|\mathbf{X} \in A]\mathbb{E}[\psi_\alpha(Y, q_R)|\mathbf{X} \in A_R]$.

[8] Harish S Bhat, Nitesh Kumar, and Garnet J Vaz (2015). In: *2015 IEEE International Conference on Big Data (Big Data)*. IEEE

How to obtain conditional informations from a tree?

QOSA / QOSE estimation require to estimate quantities like $\mathbb{E}\left[\psi_\alpha\left(Y,\ q_\alpha\left(Y \middle| \mathbf{X}_u\right)\right)\right]$, $u \subset \{1, \dots, d\}$, it may be achieved by using projected random forests[9]. For a query point $x_u \in \mathcal{X}_u$:

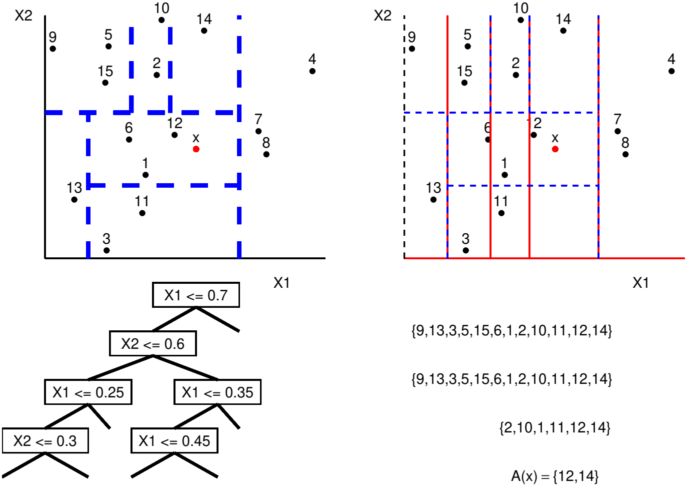
- Project the forest's partition onto the subspace spanned by X_u.
- For each tree, follow the path of x_u ignoring splits on variables $\notin u$.

The estimations (conditional quantiles, conditional means) are done in the projected leaves.

[9] Clément Bénard et al. (2022). In: *International Conference on Artificial Intelligence and Statistics*. PMLR

How to obtain conditional informations from a tree?

Example in dimension $p = 2$ and with $u = \{1\}$.



Importance sampling, estimation of QOSE

Reduce the estimation cost (for d large) due to $\sum_{u \subset \{1,\dots,d\} \backslash \{i\}}$ in the definition of Shapley effects, by considering only some u's: sample uniformely[9] K subsets u vs sample with respect to some importance in the trees construction[10]. We improved[11] this importance sampling notion.

- For each variable j, record the occurrences, of j as a splitting variable in any path, and any tree of the forest $\rightarrow$ m_j.
- Define subset u occurrence as: $f(u) = \min_{j \in u} m_j$
- Normalize to get $\tilde{p}_{B,n}(u)$, a probability distribution on $\mathcal{P}(\{1, \dots, d\})$.

---

[9] Eunhye Song, Barry L Nelson, and Jeremy Staum (2016). In: *SIAM/ASA Journal on Uncertainty Quantification*

[10] Clément Bénard et al. (2022). In: *International Conference on Artificial Intelligence and Statistics*. PMLR

[11] Véronique Maume-Deschamps et al. (2026). Work in progress

Importance sampling, estimation of QOSE

Reduce the estimation cost (for d large) due to $\sum_{u \subset \{1,\ldots,d\} \setminus \{i\}}$ in the definition of Shapley effects, by considering only some u's: sample uniformely[9] K subsets u vs sample with respect to some importance in the trees construction[10]. We improved[11] this importance sampling notion.

- Covers all subsets (not just ordered ones).
- Preserves importance trends (e.g., subsets with influential variables have higher probability).
- Empirical validation: More accurate QOSE estimates in simulations.

[9] Eunhye Song, Barry L Nelson, and Jeremy Staum (2016). In: *SIAM/ASA Journal on Uncertainty Quantification*

[10] Clément Bénard et al. (2022). In: *International Conference on Artificial Intelligence and Statistics*. PMLR

[11] Véronique Maume-Deschamps et al. (2026). Work in progress

Plan

Example: Gaussian inputs and lognormal output

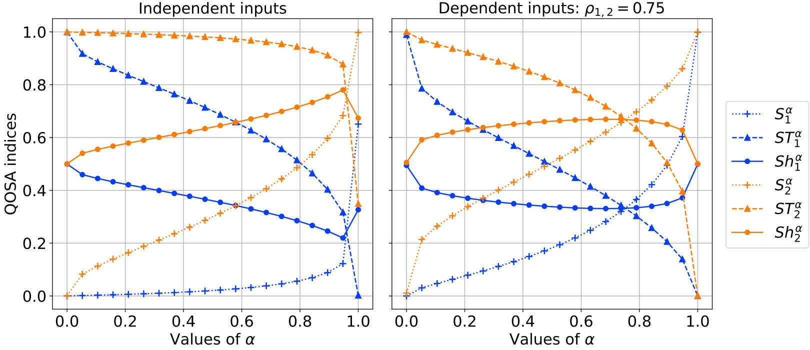
Consider $Y = f(\mathbf{X}) = \exp\left(\beta_0 + \beta^\mathsf{T}\mathbf{X}\right)$ with $\beta_0 \in \mathbb{R}$, $\beta \in \mathbb{R}^d$ and $\mathbf{X} \sim \mathcal{N}(\mu, \Sigma)$ with $\Sigma \in \mathbb{R}^{d\times d}$ a semi-positive definite symetric matrix.

Closed form formulas for QOSA indices and QOSA-Shapley effects[12].

Example in dimension 2 with $\mu_1 = \mu_2 = 0, \beta_1 = \beta_2 = 1, \sigma_1 = 1$ and $\sigma_2 = 2$, for independent ($\rho_{1,2} = 0$) inputs and correlated inputs with $\rho_{1,2} = 0.75$.

[12] Kévin Elie-Dit-Cosaque and Véronique Maume-Deschamps (2022). In: *SIAM/ASA Journal on Uncertainty Quantification*

Example: Gaussian inputs and lognormal output



Sandwich effect: QOSE between order one and total QOSA indices.
Same ranking for all α-levels, QOSE is easier to interpret, it properly condenses all the information (dependence and effects on quantiles).

Estimation of QOSE: leafsize issue

The estimation of QOSE is sensitive to leafsize. An example (linear Gaussian model): $Y = \beta^\intercal \boldsymbol{X}$, $d = 3$.

Parameters

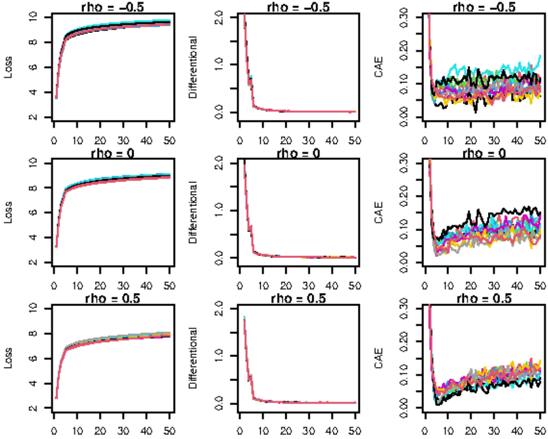
- $\beta = (1, 1, 1)^\intercal$
- $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- $\boldsymbol{\mu} = (0, 0, 0)^\intercal$

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \rho \\ 0 & \rho & 4 \end{pmatrix}, \quad \rho \in \{-0.5, 0, 0.5\}$$

Estimation of QOSE: leafsize issue

The estimation of QOSE is sensitive to leafsize.

Figure for $\alpha = 0.5$, 20 repetitions. Remark that for the linear Gaussian model, QOSE indicies do not depend on α.



A heuristic for the leafsize selection

For fixed K, a set of subsets $\{1, \ldots, d\} = \{u_1, \ldots, u_K\}$, and a minimum leafsize m, consider the loss function:

$$L(m) = \sum_{i=1}^{K} \hat{c}(u_i | m)$$

where $\hat{c}$ is the estimator of c, with m as the minimum leafsize in the projected trees. Recall

$$c(u) = \frac{\mathbb{E}\left[\psi_\alpha\left(Y,\ q_\alpha\left(Y | \mathbf{X}_{-u}\right)\right)\right]}{\mathbb{E}\left[\psi_\alpha(Y, q_\alpha(Y))\right]}.$$

Select: $m_0 =$ leaf.size the smallest value where $L(m)$ stagnates (stops decreasing).

A real dataset

Bias between the predictions from **MOCAGE** (**M**odèle de **C**himie **A**tmosphérique à **G**rande **E**chelle) and the observed ozone concentration.

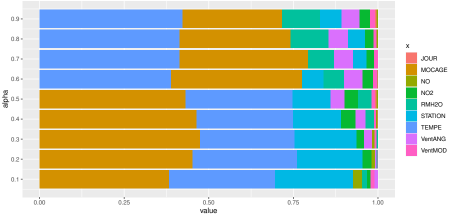
This dataset[12] contains 10 variables with 1041 observations.

O3obs: observed ozone concentration will be explained by the 9 other variables.

JOUR: type of day (holiday vs no holiday) MOCAGE: ozone concentration predicted by a fluid mechanics model RMH2O: humidity ratio VentMOD: wind force NO: nitric oxide concentration	STATION: site of observations (5 different sites) TEMPE: officially predicted temperatures NO2: nitrogen dioxide concentration VentANG: wind direction

[12] Philippe Besse et al. (2007). In: *Pollution atmosphérique*

Application on a real dataset: results



Evolution of the ranking of the QOSA indices in function of the levels α.

Considering the central effects leads[13] to consider MOCAGE and TEMPE as the most influencial variables, then STATION and NO2. We see that for quantile levels ≥ 0.6 wind and humidity are also important.

[13] Philippe Besse et al. (2007). In: *Pollution atmosphérique*
Baptiste Broto, Francois Bachoc, and Marine Depecker (2020). In: *SIAM/ASA Journal on Uncertainty Quantification*

Plan

Conclusion.

- Goal oriented random forests give better results than CART based random forests for QOSA. Projection of quantile random forests allows estimation of QOSE.
- QOSE are interesting sensitivity indices, give informations on the sensitivity around quantiles, fair repartition of the impact of each variable, even with dependent inputs.
- Work in progress for optimized code (R / cpp).
- Consistency results for QOSA and QOSE estimations[14]. Asymptotic distributions / confidence intervals for QOSA/E indices?

[14] Kevin Elie-Dit-Cosaque and Véronique Maume-Deschamps (2024). In: *Computational Statistics*

Ri Wang, Véronique Maume-Deschamps, and Clémentine Prieur (2026). In: *Journal of Statistical Planning and Inference*

Véronique Maume-Deschamps et al. (2026). Work in progress

References I

Bénard, Clément et al. (2022). "SHAFF: Fast and consistent SHApley eFfect estimates via random Forests". In: *International Conference on Artificial Intelligence and Statistics*. PMLR, pp. 5563–5582.

Besse, Philippe et al. (2007). "Comparaison de techniques de «Data Mining» pour l'adaptation statistique des prévisions d'ozone du modèle de chimie-transport MOCAGE". In: *Pollution atmosphérique* 49.195, pp. 285–292.

Bhat, Harish S, Nitesh Kumar, and Garnet J Vaz (2015). "Towards scalable quantile regression trees". In: *2015 IEEE International Conference on Big Data (Big Data)*. IEEE, pp. 53–60.

Broto, Baptiste, Francois Bachoc, and Marine Depecker (2020). "Variance reduction for estimation of Shapley effects and adaptation to unknown input distribution". In: *SIAM/ASA Journal on Uncertainty Quantification* 8.2, pp. 693–716.

Browne, Thomas et al. (2017). "Estimate of quantile-oriented sensitivity indices". In:

Elie-Dit-Cosaque, Kévin and Véronique Maume-Deschamps (2022). "Goal-oriented Shapley effects with special attention to the quantile-oriented case". In: *SIAM/ASA Journal on Uncertainty Quantification* 10.3, pp. 1037–1069.

Elie-Dit-Cosaque, Kevin and Véronique Maume-Deschamps (2024). "Random forest based quantile-oriented sensitivity analysis indices estimation". In: *Computational Statistics*, pp. 1–31.

Fort, Jean-Claude, Thierry Klein, and Nabil Rachdi (2016). "New sensitivity analysis subordinated to a contrast". In: *Communications in Statistics-Theory and Methods* 45.15, pp. 4349–4364.

Maume-Deschamps, Véronique and Ibrahima Niang (2018). "Estimation of quantile oriented sensitivity indices". In: *Statistics & Probability Letters* 134, pp. 122–127.

Maume-Deschamps, Véronique et al. (2026). "Statistical inference of Quantile Oriented Shapley Effects and application to meteorological data". Work in progress.

References II

Owen, Art B (2014). "Sobol'Indices and Shapley Value". In: *SIAM/ASA Journal on Uncertainty Quantification* 2.1, pp. 245–251.

Sobol', IM (1993). "Sensitivity estimates for nonlinear mathematical models". In: *Math. Model. Comput. Exp.* 1, p. 407.

Song, Eunhye, Barry L Nelson, and Jeremy Staum (2016). "Shapley effects for global sensitivity analysis: Theory and computation". In: *SIAM/ASA Journal on Uncertainty Quantification* 4.1, pp. 1060–1083.

Wang, Ri, Véronique Maume-Deschamps, and Clémentine Prieur (2026). "Quantile oriented sensitivity analysis with random forest based on pinball loss". In: *Journal of Statistical Planning and Inference* 246, p. 106447.

Thanks


The 11th International Congress
on Industrial and Applied Mathematics (ICIAM)
ICIAM 2027
The Hague
The Netherlands
Mathematics for a sustainable future
12-16 July 2027
www.iciam2027.org
IMPORTANT DATES
Registration open:
SEPTEMBER 2026
Early-bird registration deadline:
9 FEBRUARY 2027
Abstract submission deadline:
7 DECEMBER 2026
Regular registration deadline:
14 JUNE 2027
Grant application deadline:
7 DECEMBER 2026
Congress dates:
12-16 JULY 2027
WHY ATTEND?
• Engage with leading experts in industrial and applied mathematics
• Discover cutting-edge developments across disciplines
• Build international collaborations and expand your network
Be part of the world congress of applied mathematics.