Uncertainty quantification in numerical models, with a focus on Sensitivity Analysis

Clémentine PRIEUR
Université Grenoble Alpes

Prévisibilité dans les sciences de l'atmosphère, des océans, et du climat, 2-3 Octobre 2023, IHP

Outline

Step 1
Input uncertainty
probability distribution
of $\mathbf{X}$

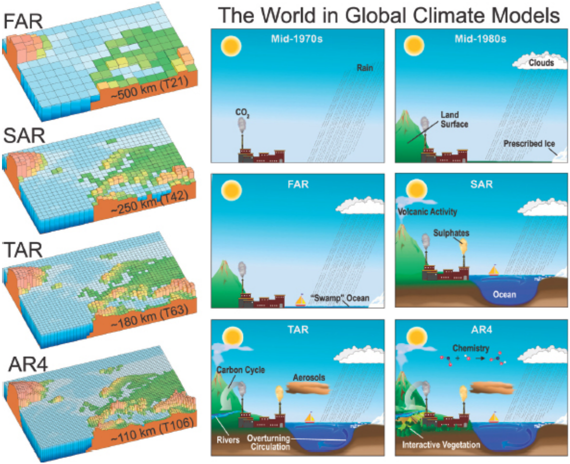
Step 0 (4)
Model specification
$Y = \mathcal{M}(\mathbf{X})$
input parameters $\mathbf{X}$, output Y

Step 2
Uncertainty propagation
tendency, quantiles...
of Y

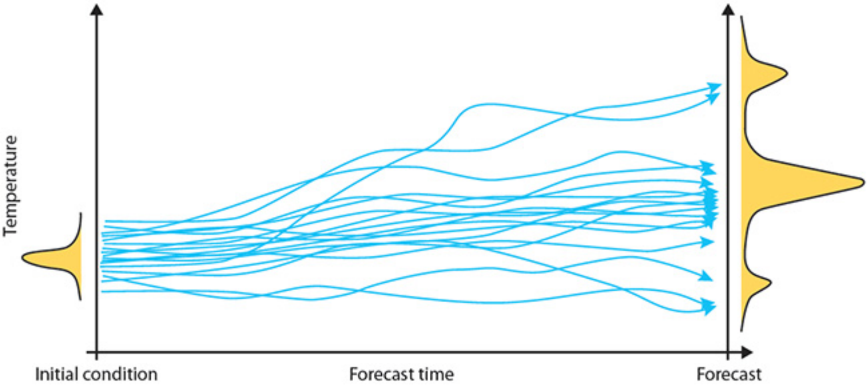
Step 3
Sensitivity analysis
- local variational approach
- global sensitivity analysis

Step 0: model specification

Evolution of climate models through last century



https://www.windows2universe.org/earth/climate/climate_modeling.html



https://www.ecmwf.int/en/about/media-centre/news/2017/twenty-five-years-ensemble-forecasting

Step 1

Step 2

Concerning **Step 1**, uncertainty may be classified into two categories

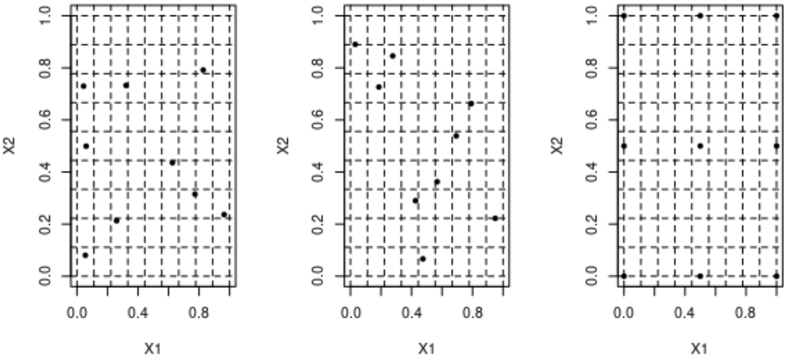
- aleatoric (aka statistical) uncertainty refers to the notion of randomness, that is, the variability in the outcome of an experiment,
- epistemic (aka systematic) uncertainty refers to uncertainty caused by a lack of knowledge.

Examples:

- meteorological inputs are random,
- bathymetry.

Classification is not always easy.

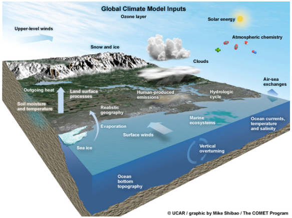
How to explore "at best" input parameter space for **Step 2** of uncertainty propagation?



Uniform design | Latin Hypercube Sampling | Factorial design

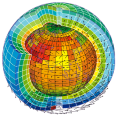
It is important, e.g., for parameter perturbation in view of ensemble forecast.

Step 1 misspecified input parameters

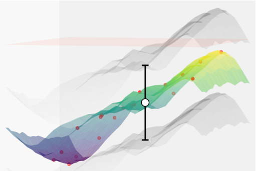


Step 0 complex models, $\mathcal{O}(10^{10})$ degrees of freedom

$$
\begin{cases}
\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + \dot{\sigma}\frac{\partial u}{\partial \sigma} - f(v - v_g) + \frac{\partial \phi}{\partial x}|_p + F_u = 0 \\
\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + \dot{\sigma}\frac{\partial v}{\partial \sigma} + f(u - u_g) + \frac{\partial \phi}{\partial y}|_p + F_v = 0 \\
\frac{\partial \phi}{\partial (T/\theta)} = -C_p\theta(1 + 0.85q_v) \\
\frac{\partial p_\star}{\partial t} = -\int_0^1 [\frac{\partial (p_\star u)}{\partial x} + \frac{\partial (p_\star v)}{\partial y}] d\sigma \\
\frac{\partial \sigma}{\partial t} = \frac{1}{p_\star}\{\sigma \int_0^1 [\frac{\partial (p_\star u)}{\partial x} + \frac{\partial (p_\star v)}{\partial y}] d\sigma - \int_0^\sigma [\frac{\partial (p_\star u)}{\partial x} + \frac{\partial (p_\star v)}{\partial y}] d\sigma\} \\
\frac{\partial \theta}{\partial t} + u\frac{\partial \theta}{\partial x} + v\frac{\partial \theta}{\partial y} + \dot{\sigma}\frac{\partial \theta}{\partial \sigma} + \frac{\partial R_{rad}}{\partial \sigma} + F_\theta + p_\theta = 0 \\
\frac{\partial q_v}{\partial t} + u\frac{\partial q_v}{\partial x} + v\frac{\partial q_v}{\partial y} + \dot{\sigma}\frac{\partial q_v}{\partial \sigma} + F_{q_v} + p_{q_v} = 0
\end{cases}
$$

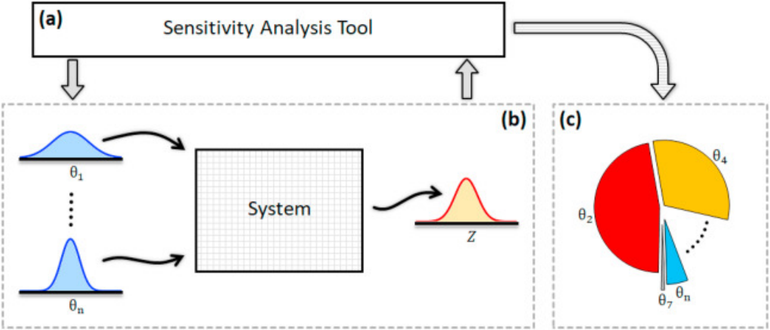


Step 2 uncertain predictions



Outline

Step 3: sensitivity analysis (Razavi *et al.*, 2021)

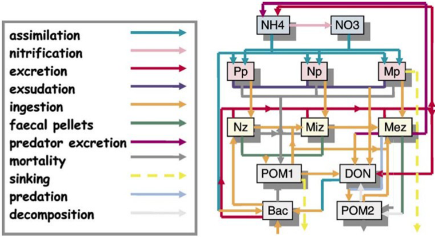


Aim of Sensitivity Analysis: find how model outputs vary with input changes.

Application to a biogeochemical model:
ecosystem model (MODECOGeL) of the Ligurian Sea

Joint work with IGE Lab (Grenoble, FRANCE)

MODECOGeL is a one-dimensional coupled hydrodynamical-biological model.



• hydrodynamic model: 1-D vertical simplification of primitive equations for the ocean, 5 state variables;
• ecosystem model: marine biogeochemistry, 12 biological state variables.

Inputs/Outputs:

▷ 74 scalar input parameters;

▷ spatio-temporal outputs.

Main issue: calibration of the model.

Sensitivity Analysis is a preliminary step to this calibration task.

GSA for convection-permitting Numerical Weather Prediction (NWP) models Wimmer *et al* (2022)

Aim: to determine the most influential parameters on the forecast of different near-surface variables.

Model:
convective-scale AROME model.

Input parameters:

Scheme	Parameter	Physical meaning	Default	Range
Radiation	RSWINHF	Shortwave inhomogeneity factor	1	0.6 - 1
	RLWINHF	Longwave inhomogeneity factor	1	0.6 - 1
Microphysics	RCRIAUTI	Snow Autoconversion threshold	0.2e-3	0.2e-4 - 0.25e-3
	RCRIAUTC	Rain Autoconversion threshold	1e-3	0.4e-3 - 1e-3
	VSIGQSAT	Constant for subgrid condensation	0.02	0 - 0.1
Turbulence	XLINI	Minimum mixing length	0	0 - 0.2
	XCTD	Constant for dissipation of temperature and vapor pressure fluctuations	1.2	0.98 - 1.2
	XCTP	Constant for temperature and vapor presure correlations	4.65	1.035 - 22.22
	XCEP	Constant for wind-pressure correlations	2.11	0.225 - 4.0
	XCED	Constant for dissipation of TKE	0.85	0.4 - 2
	XPHI_LIM	Threshold value for Sc^{-1} and Pr^{-1}	3	1 - 4.5
	XCET	Constant for transport of TKE	0.4	0.072 - 1.512
Diffusion	SLHDEPSH	Strength of SLHD	0.060	0.01 - 0.09
	SLHDKMIN	Diffusion function minimum	0	-1 - 1
	SLHDKMAX	Diffusion function maximum	6	4 - 12
Surface	XRIMAX	Critical Richardson Number	0.2	0 - 0.3
	XFRACZ0	Coefficient of orographic drag	5	2 - 10
Convection	XCMF	Closure coefficient at bottom level	0.065	0 - 0.1
	XABUO	Coefficient of the buoyancy	1	0.7 - 1.5
	XBDETR	Coefficient of the detrainment	1e-6	0 - 1
	XENTR_DRY	Coefficient for dry entrainment	0.55	0.1 - 0.699

Scalar outputs: either an averaged forecast field or a performance metric such as mean bias, root-mean square error and mean absolute error.

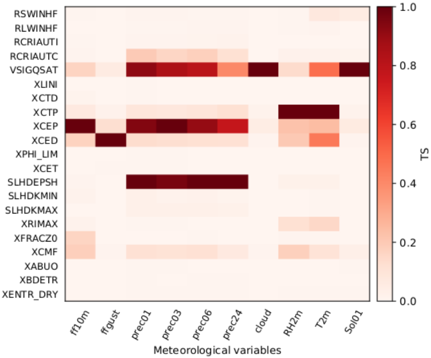
$$\text{Bias} = \frac{1}{n}\sum_{k=1}^{n}(y_k - o_k)\,, \ \text{RMSE} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(y_k - o_k)^2}\,,$$

$$\text{MAE} = \frac{1}{n}\sum_{k=1}^{n}|y_k - o_k|\,.$$

These scores are computed using SYNOP (surface synoptic observations) and the French real-time meteorological observation network (Tardieu & Leroy, 2003).

Notation: n is the number of in-situ measures, y_k is the model output and o_k is the k-th observation.

On the figure below (Wimmer *et al*, 2022), outputs are spatial-averaged scores computed for 10-meter wind speed ff10m, 10-meter wind gust ffgust, 1-hourly, 3-hourly, 6-hourly and 24-hourly accumulated precipitation prec01, prec03, prec06, prec24, total cloud cover cloud, 2-meter relative humidity RH2m, 2-meter temperature T2m and 1-hourly downward global solar radiation Sol01. Total Sobol' indices measure the sensitivity of each of these outputs with respect to input parameters.



The higher the Total Sobol' Indice is, the darker the colour is, the more influential the parameter is.

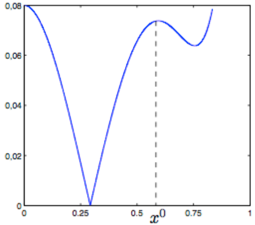
Outline

$$\mathcal{M} : \begin{cases} \mathbb{R}^d & \rightarrow & \mathbb{R} \\ \mathbf{x} & \mapsto & y = \mathcal{M}(x_1, \ldots, x_d) \end{cases}$$

Local Sensitivity Analysis is based on Taylor approximation:

$\mathcal{M}(\mathbf{x}) \approx \mathcal{M}(\mathbf{x^0}) + \sum_{i=1}^{d} \left(\frac{\partial \mathcal{M}}{\partial x_i}\right)_{\mathbf{x^0}} (x_i - x_i^0)$.

First order sensitivity index for input i: $\left(\frac{\partial \mathcal{M}}{\partial x_i}\right)_{\mathbf{x^0}}$.



Pros: Low computational cost even for large d if one uses the adjoint stae method (see, e.g., Plessix, 2006).

Cons: Local analysis, not well-suited for highly nonlinear models.

Global Sensitivity Analysis with Screening

Main objective: to screen among a large amount of inputs which ones are non influential on the quantity of interest (QoI).

Advantages: moderate computational cost.

Drawbacks: partial information, no hierarchisation.

A OAT screening method : Morris, 1991

OAT One At a Time we vary the factors one by one.
The screening method proposed by Morris is a global OAT approach.

Model $Y = \mathcal{M}(\mathbf{X})$, $\mathbf{X} = (X_1, \ldots, X_d)$ *with the* X_i*s independent uniform random variables on* $[0, 1]$.

More details on the method :

- input discretization on a grid with p values $\left\{0, \frac{1}{p-1}, \ldots, 1\right\}$.
- Δ a multiple of $1/(p-1)$, fixed once for all.
- $\Omega := \left\{0, \frac{1}{p-1}, \ldots, 1\right\}^d$.
- $\Omega_i^\Delta := \{x \in \Omega \text{ such that } (x_1, \ldots, x_{i-1}, x_i + \Delta, x_{i+1}, \ldots, x_d) \in \Omega\}$.

Definition

Elementary effect of X_i computed at $\mathbf{x} \in \Omega_i^\Delta$,

$$d_i(\mathbf{x}) = \frac{1}{\Delta} \left\{\mathcal{M}(x_1, \ldots, x_{i-1}, x_i + \Delta, x_{i+1}, \ldots, x_d) - \mathcal{M}(\mathbf{x})\right\} .$$

There are $p^{d-1}(p - \Delta(p-1))$ elementary effects to compute.

Steps :

- one draws uniformly a r-sample in Ω_i^Δ : $\mathbf{x}^1, \ldots, \mathbf{x}^r$;
- one computes $d_i(\mathbf{x}^j)$, $j = 1, \ldots, r$, $i = 1, \ldots, d$;
- one computes

$$\mu_i = \frac{1}{r}\sum_{j=1}^{r} d_i(\mathbf{x}^j)\,, \quad \sigma_i^2 = \frac{1}{r}\sum_{j=1}^{r}(d_i(\mathbf{x}^j) - \mu_i)^2.$$

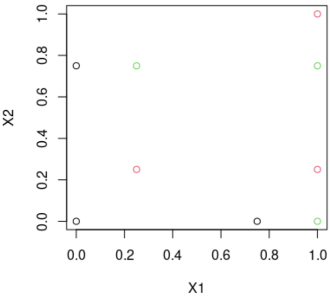
	σ_i^2 low	σ_i^2 high
$\lvert\mu_i\rvert$ low	non influential	nonlinearities and/or interactions
$\lvert\mu_i\rvert$ high	influential	nonlinearities and/or interactions

Elementary effect of X_i computed at $\mathbf{x} \in \Omega_i^\Delta$,

$$d_i(\mathbf{x}) = \frac{1}{\Delta}\left\{\mathcal{M}(x_1, \ldots, x_{i-1}, x_i + \Delta, x_{i+1}, \ldots, x_d) - \mathcal{M}(\mathbf{x})\right\} .$$

The efficiency of the method "number of elementary effects computed / number of model runs" is equal to $1/2$.

Morris (1991) presents an adaptation with an efficiency equal to $d/(d+1)$, with d the input space dimension.



$r = 3$ Morris trajectories with $p = 5$, $\Delta = 3/4$.

A toy example Advection-reaction-diffusion equation with Dirichlet boundary condition :

$$\begin{cases} \dfrac{\partial u}{\partial t} = -r.u - a\dfrac{\partial u}{\partial x} + \lambda\dfrac{\partial^2 u}{\partial x^2} + f \quad x \in [0, L],\ t \in [0, T] \\ u(x = 0, t) = \Psi_1(t) \quad t \in [0, T] \\ u(x = L, t) = \Psi_2(t) \quad t \in [0, T] \\ u(x, t = 0) = g(x) \quad x \in (0, L)\,. \end{cases}$$

A : energy norm of the solution at time $t = T$.

Sensitivity of A with respect to (a, r, λ)? Uncertain input parameters are modeled as $a, r \sim \mathcal{U}([0.4, 0.6])$, $\lambda \sim \mathcal{U}([0.04, 0.06])$.

Scheme : 2-stemps Adams-Moulton, sample size equals 2^{13}.

Sensitivity measures based on variance : $S_a = 0.0188$, $S_\lambda = 0.7299$, $S_r = 0.2488$, $S_a + S_\lambda + S_r = 0.988$.

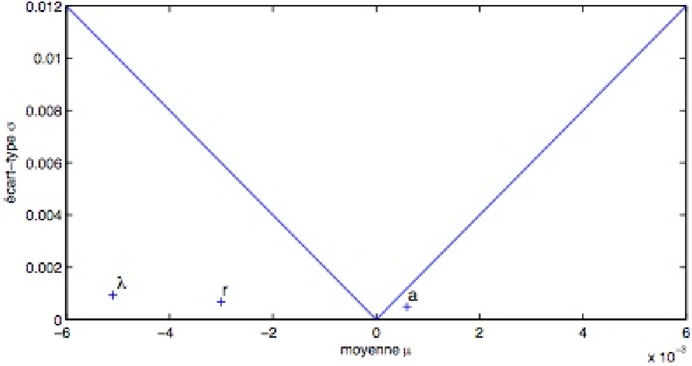


Figure: Morris with $p = 50$, $\Delta = 25/49$.

Variance-based Global Sensitivity Analysis

Independent framework: $P(d\boldsymbol{x}) = P_1(dx_1)\dots P_d(dx_d)$

Let

$$\mathcal{M}: \begin{cases} \mathbb{R}^d & \rightarrow & \mathbb{R} \\ \boldsymbol{x} = (x_1, \dots, x_d) & \mapsto & y = \mathcal{M}(\boldsymbol{x}) \end{cases}$$

Does the output Y vary more or less when fixing one of its input parameters? $V\left[Y|X_i = x_i\right]$, how to choose x_i?

$\longrightarrow E\left[V\left(Y|X_i\right)\right] = V\left[Y\right] - V\left[E\left(Y|X_i\right)\right].$

First-order Sobol' indices: $0 \leq S_i = \dfrac{V\left[E\left(Y|X_i\right)\right]}{V\left[Y\right]} \leq 1.$

The more this quantity is close to 1, the more fixing X_i reduces the variance of Y: the input X_i is influential.

More generally,

$$S_i = \frac{V\left[\mathbb{E}\left[Y|X_i\right]\right]}{V\left[Y\right]}, \quad 1 \leq i \leq d$$

$$S_{i,j} = \frac{V\left[\mathbb{E}\left[Y|X_i, X_j\right]\right] - V\left[\mathbb{E}\left[Y|X_i\right]\right] - V\left[\mathbb{E}\left[Y|X_j\right]\right]}{V\left[Y\right]}, \quad 1 \leq i \neq j \leq d \ \ldots$$

We have $1 = \sum_{i=1}^{d} S_i + \sum_{i \neq j} S_{i,j} + \ldots + S_{1,\ldots,d}$

Factors Prioritization (FP): which factor should one try to determine first to get the largest expected reduction in the variance of the model output? $\longrightarrow$ first order Sobol' indices do the job.

Total Sobol' indices:

$i = 1, \ldots, d \qquad S_i^{\text{tot}} = \sum_{\mathbf{u} \subseteq \{1,\ldots,d\}\,,\,\mathbf{u} \cap \{i\} \neq \emptyset} S_{\mathbf{u}}$

Factors Fixing (FF): which input factors can be fixed, anywhere in their range of variation, without sensibly affecting a specific output of interest? $\longrightarrow$ total Sobol' indices do the job.

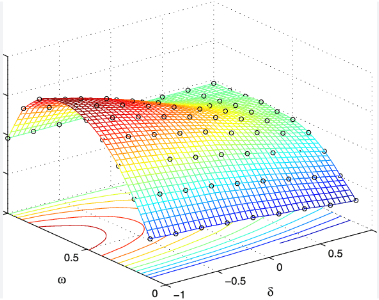
We have:

$$S_i^{\text{tot}} = \frac{E\left[V\left[Y|\mathbf{X}_{-i}\right]\right]}{V\left[Y\right]} = 1 - \frac{V\left[E\left[Y|\mathbf{X}_{-i}\right]\right]}{V\left[Y\right]}$$

with $\mathbf{X}_{-i} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_d)$.

Sobol' index estimation

- Sobol' indices can be estimated from input/ouput samples $\left(\mathbf{X}^{(\mathbf{i})}, Y^{(i)} = \mathcal{M}(\mathbf{X}^{(\mathbf{i})})\right)$, $1 \leq i \leq n$.
- Metamodels can be built to speed up computations.



Some additional issues

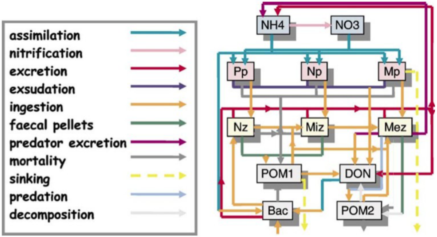
- visualization for complex outputs,
- correlated inputs,
- stochastic models. . .

Outline

Application to MODECOGeL

see Prieur *et al*, 2019

MODECOGeL is a one-dimensional coupled hydrodynamical-biological model.



• hydrodynamic model: 1-D vertical simplification of primitive equations for the ocean, 5 state variables;
• ecosystem model: marine biogeochemistry, 12 biological state variables.

▷ 74 independent scalar parameters

Index	Name	Unit	Pdf	Mean	Std	Std/Mean
1	**PicP** max growth rate	t^{-1}	$\Gamma(25,0.12)$	3.	0.6	20%
2	**NanP** max growth rate	t^{-1}	$\Gamma(25,0.1)$	2.5	0.5	20%
3	**MicP** max growth rate	t^{-1}	$\Gamma(25,0.08)$	2.	0.4	20%
4	dependence of **NO3** limitation to **NH4**	C^{-1}	$\Gamma(400,0.00365)$	1.46	0.073	5%
5	**NO3** semisaturation for **PicP**	C	$\Gamma(4,0.125)$	0.5	0.25	50%
6	**NO3** semisaturation for **NanP**	C	$\Gamma(4,0.175)$	0.7	0.35	50%
7	**NO3** semisaturation for **MicP**	C	$\Gamma(4,0.25)$	1.0	0.5	50%
8	**NH4** semisaturation for **PicP**	C	$\Gamma(4,0.075)$	0.3	0.15	50%
9	**NH4** semisaturation for **NanP**	C	$\Gamma(4,0.125)$	0.5	0.25	50%
10	**NH4** semisaturation for **MicP**	C	$\Gamma(4,0.175)$	0.7	0.35	50%
11	optimal PAR for **PicP**	I	$\Gamma(25,0.4)$	10.	2.	20%
12	optimal PAR for **NanP**	I	$\Gamma(25,0.6)$	15.	3.	20%
13	optimal PAR for **MicP**	I	$\Gamma(25,0.8)$	20.	4.	20%
14	variation of light limitation for **PicP**	–	$-\Gamma(4,0.2)$	-0.8	0.4	50%
15	variation of light limitation for **NanP**	–	$-\Gamma(4,0.175)$	-0.7	0.35	50%
16	variation of light limitation for **MicP**	–	$-\Gamma(4,0.15)$	-0.6	0.3	50%
17	optimal temperature for **PicP**	T	$\mathcal{N}(15,3^2)$	15.	3.	20%
18	optimal temperature for **NanP**	T	$\mathcal{N}(15,3^2)$	15.	3.	20%
19	optimal temperature for **MicP**	T	$\mathcal{N}(16,3.2^2)$	16.	3.2	20%
20	variation of temp. limitation for **PicP**	–	$-\Gamma(4,0.125)$	-0.5	0.25	50%
21	variation of temp. limitation for **NanP**	–	$-\Gamma(4,0.125)$	-0.5	0.25	50%
22	variation of temp. limitation for **MicP**	–	$-\Gamma(4,0.1375)$	-0.55	0.275	50%
23	bacteria growth limitation	–	$\Gamma(4,0.15)$	0.6	0.3	50%
24	semisaturation for **BAC** growth	C	$\Gamma(4,0.125)$	0.5	0.25	50%
25	exudation ratio for **PicP**	–	$\Gamma(4,0.015)$	0.06	0.03	50%
26	exudation ratio for **NanP**	–	$\Gamma(4,0.0125)$	0.05	0.025	50%
27	exudation ratio for **MicP**	–	$\Gamma(4,0.01)$	0.04	0.02	50%
28	max ingestion rate for **NanZ**	t^{-1}	$\Gamma(25,0.12)$	3.	0.6	20%
29	max ingestion rate for **MicZ**	t^{-1}	$\Gamma(25,0.08)$	2.	0.4	20%
30	max ingestion rate for **MesZ**	t^{-1}	$\Gamma(25,0.06)$	1.5	0.3	20%
31	threshold ingestion for **NanZ**	C	$\Gamma(4,0.0125)$	0.05	0.025	50%
32	threshold ingestion for **MicZ**	C	$\Gamma(4,0.0075)$	0.03	0.015	50%
33	threshold ingestion for **MesZ**	C	$\Gamma(4,0.0025)$	0.01	0.005	50%
34	semisaturation for ingestion by **NanZ**	C	$\Gamma(4,0.125)$	0.5	0.25	50%
35	semisaturation for ingestion by **MicZ**	C	$\Gamma(4,0.1875)$	0.75	0.375	50%
36	semisaturation for ingestion by **MesZ**	C	$\Gamma(4,0.25)$	1.	0.5	50%
37	efficiency of **MesZ** on **MicP**	–	$\beta(4.2,1.05)$	0.8	0.16	20%
38	efficiency of **NanZ** on **BAC**	–	$\beta(4.2,1.05)$	0.8	0.16	20%
39	efficiency of **MicZ** on **NanZ**	–	$\beta(4.2,1.05)$	0.8	0.16	20%
40	efficiency of **MesZ** on **MicZ**	–	$\beta(4.2,1.05)$	0.8	0.16	20%
41	efficiency of **MicZ** on **MOP1**	–	$\beta(19.8,79.2)$	0.2	0.04	20%
42	efficiency of **MesZ** on **MOP1**	–	$\beta(19.8,79.2)$	0.2	0.04	20%
43	efficiency of **MesZ** on **MOP2**	–	$\beta(19.8,79.2)$	0.2	0.04	20%
44	mortality rate for **PicP**	t^{-1}	$\Gamma(4,0.015)$	0.06	0.03	50%
45	mortality rate for **NanP**	t^{-1}	$\Gamma(4,0.0125)$	0.05	0.025	50%
46	mortality rate for **MicP**	t^{-1}	$\Gamma(4,0.01)$	0.04	0.02	50%
47	mortality rate for **NanZ**	t^{-1}	$\Gamma(4,0.015)$	0.06	0.03	50%
48	mortality rate for **MicZ**	t^{-1}	$\Gamma(4,0.0125)$	0.05	0.025	50%
49	mortality rate for **MesZ**	t^{-1}	$\Gamma(4,0.0075)$	0.03	0.015	50%
50	mortality rate for **BAC**	t^{-1}	$\Gamma(4,0.015)$	0.06	0.03	50%
51	threshold for predation	C	$\Gamma(4,0.005)$	0.02	0.01	50%
52	maximum predation rate on **MesZ**	t^{-1}	$\Gamma(4,0.25)$	1.	0.5	50%
53	semisaturation for predation on **MesZ**	C	$\Gamma(4,0.25)$	1.	0.5	50%
54	excreted fraction of predation on **MesZ**	–	$\beta(2.33,4.67)$	0.333	0.167	50%
55	fraction of grazing used for growth of **NanZ**	–	$\beta(4.2,1.05)$	0.8	0.16	20%
56	fraction of grazing used for growth of **MicZ**	–	$\beta(4.2,1.05)$	0.8	0.16	20%
57	fraction of grazing used for growth of **MesZ**	–	$\beta(4.2,1.05)$	0.8	0.16	20%
58	fraction of POM used for growth of **MicZ**	–	$\beta(12,12)$	0.5	0.1	20%
59	fraction of POM used for growth of **MesZ**	–	$\beta(12,12)$	0.5	0.1	20%
60	excretion rate for **NanZ**	t^{-1}	$\Gamma(4,0.0375)$	0.15	0.075	50%
61	excretion rate for **MicZ**	t^{-1}	$\Gamma(4,0.025)$	0.1	0.05	50%
62	excretion rate for **MesZ**	t^{-1}	$\Gamma(4,0.0125)$	0.05	0.025	50%
63	excretion rate for **BAC**	t^{-1}	$\Gamma(4,0.0375)$	0.15	0.075	50%
64	temperature variation of excretion for **NanZ**	–	LogGamma	1.05	0.0525	5%
65	temperature variation of excretion for **MicZ**	–	LogGamma	1.05	0.0525	5%
66	temperature variation of excretion for **MesZ**	–	LogGamma	1.02	0.051	5%
67	temperature variation of excretion for **BAC**	–	LogGamma	1.04	0.052	5%
68	fraction of excretion as **DOM**	–	$\beta(2.75,8.25)$	0.25	0.125	50%
69	POM1 decomposition rate	t^{-1}	$\Gamma(4,0.01625)$	0.065	0.0325	50%
70	POM2 decomposition rate	t^{-1}	$\Gamma(4,0.015)$	0.06	0.03	50%
71	sedimentation velocity for **MicP**	V	$\Gamma(4,0.25)$	1.	0.5	50%
72	nitrification rate	t^{-1}	$\Gamma(4,0.0075)$	0.03	0.015	50%
73	light attenuation coefficient in sea water	–	$\Gamma(25,0.0016)$	0.04	0.008	20%
74	fraction of photosynthetically active radiation	–	$\Gamma(25,0.02)$	0.5	0.1	20%

State variables

The ecosystem model provides a 12-component description of the ecosystem of the Ligurian Sea.

Variable	Acronym	Name
C_1	**NO3**	Nitrate
C_2	**NH4**	Ammonium
C_3	**PicP**	Picophytoplankton
C_4	**NanP**	Nanophytoplankton
C_5	**MicP**	Microphytoplankton
C_6	**NanZ**	Nanozooplankton
C_7	**MicZ**	Microzooplankton
C_8	**MesZ**	Mesozooplankton
C_9	**BAC**	Bacteria
C_{10}	**DON**	Dissolved organic nitrogen
C_{11}	**POM1**	Particulate organic matter (size 1)
C_{12}	**POM2**	Particulate organic matter (size 2)

The time evolution of each state variable is governed by the equation:

$$\frac{\partial C_i}{\partial t} = \mathtt{ADV}_i + \mathtt{DIFF}_i + \mathtt{SMS}_i \quad \text{with} \quad \mathtt{SMS}_i = \sum_{j \neq i} \mathtt{FLUX}(C_j \to C_i)$$

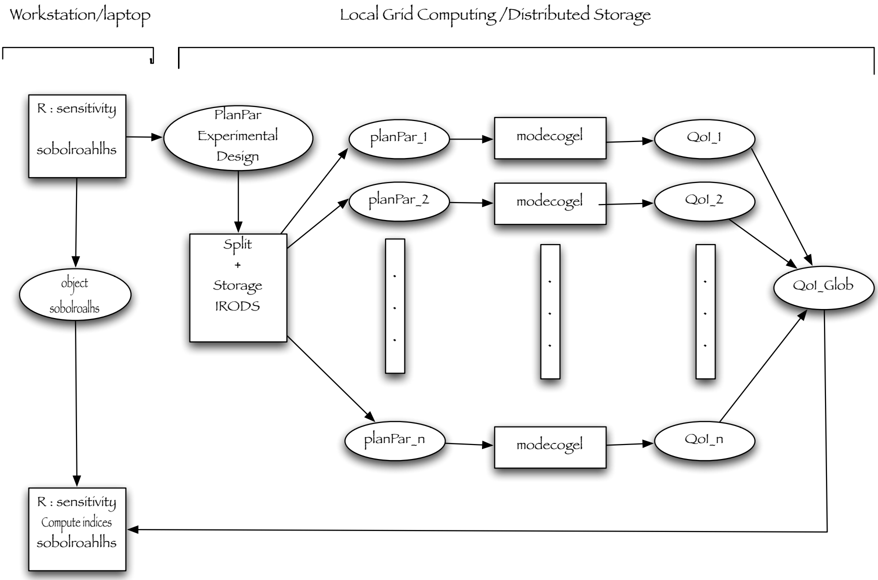
where $\mathtt{ADV}_i$ and $\mathtt{DIFF}_i$ are advection and diffusion terms, and $\mathtt{SMS}_i$ is the “source minus sink” term summing up the fluxes ($\mathtt{FLUX}(C_j \to C_i)$) between the various components of the ecosystem. We also introduce chlorophyll concentration $C_0 = \alpha(C_3 + C_4 + C_5)$.

QoIs

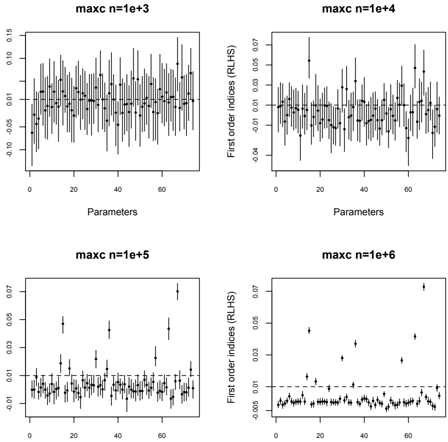
Index j	Name	Definition
1	surface maximum	$\max_t C_i(0,t)$
2	time of surface maximum	$\text{argmax}_t C_i(0,t)$
3	maximum of vertical average	$\max_t \frac{1}{Z}\int_0^Z C_i(z,t)\,dz$
4	time of maximum of vertical average	$\text{argmax}_t \frac{1}{Z}\int_0^Z C_i(z,t)\,dz$
5	time and vertical average	$\frac{1}{ZT}\int_0^T\int_0^Z C_i(z,t)\,dz\,dt$

Quantities of interest Y_{ij}. The maximum depth for averaging is $Z = 40\,\text{m}$, and T is the total duration of the experiment.

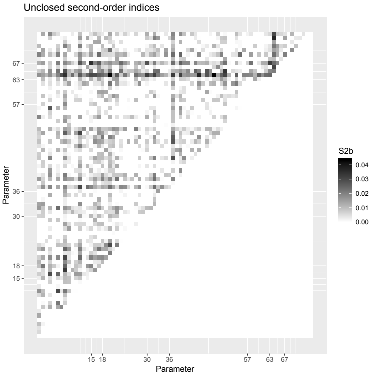
Processing chain



How the results look like?



Estimated first-order indices (y-axis) with their 95% confidence interval for the 74 model parameters (x-axis), for $n = 10^3, 10^4, 10^5$ and 10^6, in the case of the output Y_{01}. The dashed horizontal line corresponds to a threshold arbitrarily chosen to be 0.01. Confidence intervals were obtained with a bootstrap procedure and a bootstrap sample size of 100.



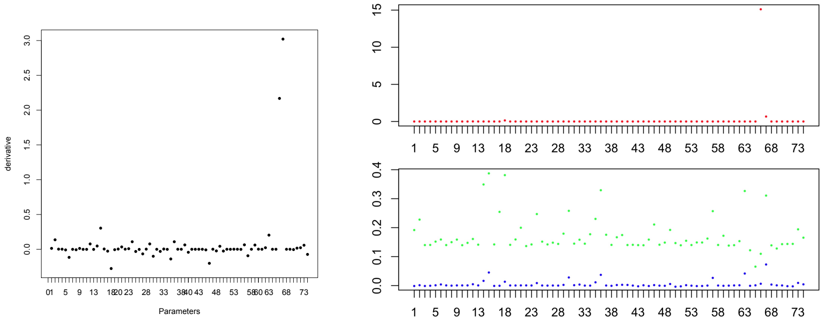
Map (74 $\times$ 74) of the second-order unclosed Sobol indices for QoI Y_{01}. The x and y axes correspond to the number of the parameters, and the grey scale to the value of the index. Note that the numbers indicated on the axes correspond to parameters with high first-order indices.

Top eight ranking of the local derivative $\partial Y / \partial X_j$, and first-order and total Sobol' indices S_j and S_j^{tot}.

j	2	14	15	18	30	35	36	46	57	63	66	67
$\partial Y / \partial X_j$	8^{th}		3^{rd}	4^{th}		7^{th}		6^{th}		5^{th}	2^{nd}	1^{st}
$S_{\{j\}}$		7^{th}	2^{nd}	8^{th}	5^{th}		4^{th}		6^{th}	3^{rd}		1^{st}
$S_{\{j\}}^{\text{tot}}$		3^{rd}	1^{st}	2^{nd}	7^{th}		4^{th}		8^{th}	5^{th}		6^{th}

We can normalize local derivatives

$$S_j^{\text{loc}} = \frac{V[X_j]}{V[Y]} \left(\frac{\partial Y}{\partial X_j} \right)^2 .$$



Derivative $\frac{\partial Y_{01}}{\partial X_j}$ (left), non dimensional derivative $S_j^{\text{loc}} = \frac{V[X_j]}{\text{Var}(Y_{01})}\left(\frac{\partial Y_{01}}{\partial X_j}\right)^2$ (right, upper panel), and first-order and total Sobol' indices (right, lower panel) as functions of the number of the parameter (x-axis). The derivatives are computed for $(x_1, \ldots, x_d) = (\mathrm{E}(X_1), \ldots, \mathrm{E}(X_d))$.

Outline

Why is the independent framework not always the right one?

Let us come back to the example of agro-climatic model for the water status management of vineyard.

The soil texture was initially described by 3 scalar parameters: the percentages of argil, sand and silt.

These parameters are not independent as

$$\% \text{ argil } + \% \text{ sand } + \% \text{ silt } = 100\% \,.$$

In the study, this set of parameters has been replaced by a unique parameter `aSoil` describing the influence of the soil texture on its evaporation capacity.

Daily precipitations, solar radiation, mean air temperature and potential evapotranspiration are temporal correlated inputs.

We chose to use kind of scenario approach: it consists in grouping the 4 temporal inputs into a single input factor, defining a weather scenario.

Why is the independent framework not always the right one?

Let us come back to the example of agro-climatic model for the water status management of vineyard.

The soil texture was initially described by 3 scalar parameters: the percentages of argil, sand and silt.

These parameters are not independent as

$$\%\text{ argil } + \%\text{ sand } + \%\text{ silt } = 100\%\,.$$

In the study, this set of parameters has been replaced by a unique parameter `aSoil` describing the influence of the soil texture on its evaporation capacity.

Daily precipitations, solar radiation, mean air temperature and potential evapotranspiration are temporal correlated inputs.

We chose to use kind of scenario approach: it consists in grouping the 4 temporal inputs into a single input factor, defining a weather scenario.

Why is the independent framework not always the right one?

Let us come back to the example of agro-climatic model for the water status management of vineyard.

The soil texture was initially described by 3 scalar parameters: the percentages of argil, sand and silt.

These parameters are not independent as

$$\%\text{ argil } + \%\text{ sand } + \%\text{ silt } = 100\% \, .$$

In the study, this set of parameters has been replaced by a unique parameter `aSoil` describing the influence of the soil texture on its evaporation capacity.

Daily precipitations, solar radiation, mean air temperature and potential evapotranspiration are temporal correlated inputs.

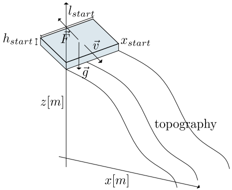
We chose to use kind of scenario approach: it consists in grouping the 4 temporal inputs into a single input factor, defining a weather scenario.

Sometimes, dependencies are due to a more complex simulation setting and cannot be handled by grouping inputs or with a scenario approach.

A snow avalanche model, joint work with INRAE (Grenoble, FRANCE)

Model based on depth-averaged Saint-Venant equations (see Heredia *et al.*, 2022 for more details)

$$
\begin{aligned}
\frac{\partial h}{\partial t} + \frac{\partial hv}{\partial x} &= 0 \\
\frac{\partial hv}{\partial t} + \frac{\partial}{\partial x}\left(hv^2 + \frac{h^2}{2}\right) &= h\,(g\sin\theta - \mathrm{F})
\end{aligned}
$$



with $v = \|\vec{\mathbf{v}}\|$ the flow velocity, h the flow depth, θ the local angle, t the time, g the gravity constant and $\mathrm{F} = \|\vec{\mathbf{F}}\|$ a frictional force. The model uses the Voellmy frictional force $\mathrm{F} = \mu g cos\theta + g/(\xi h)v^2$, where μ and ξ are friction parameters.

Equations are solved with a finite volume scheme Naaim *et al.* (98) . The topography is the one of a path located in Bessans, France.

Let us present one of the two scenarii presented in Heredia *et al.* (2022).

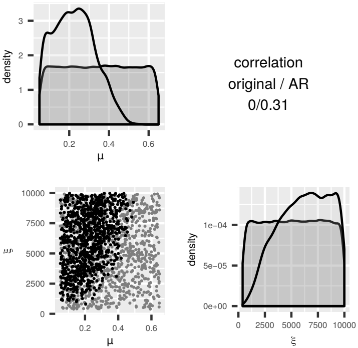
Input	Description	Distribution
μ	Static friction coefficient	$\mathcal{U}[0.05, 0.65]$
ξ	Turbulent friction [m/s^2]	$\mathcal{U}[400, 10000]$
l_{start}	Length of the release zone [m]	$\mathcal{U}[5, 300]$
h_{start}	Mean snow depth in the release zone [m]	$\mathcal{U}[0.05, 3]$
x_{start}	Release abscissa [m]	$\mathcal{U}[0, 1600]$

Let's $vol_{start} = l_{start} \times h_{start} \times 72.3/\cos(35^\circ)$ instead of h_{start} and l_{start}.

AR rules:

- avalanche simulation is flowing in $[1600m, 2412m]$,
- $vol > 7000m^3$,
- runout distance $< 2500m$ (end of the path).

From $n_0 = 100\,000$ initial runs, we keep $n_1 = 6152$ constrained ones.



An alternative, the Shapley effects

We define

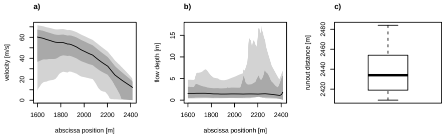
$$\phi_i = \frac{1}{d} \sum_{\mathbf{u} \subseteq -\{i\}} \binom{d-1}{|\mathbf{u}|}^{-1} \left(\mathtt{val}(\mathbf{u}+\mathtt{i}) - \mathtt{val}(\mathbf{u})\right)$$

with the characteristic function $\mathbf{u} \mapsto V[E[Y|\mathbf{X_u}]]/V[Y]$. The ϕ_is have been introduced as the Shapley effects in [Owe14].

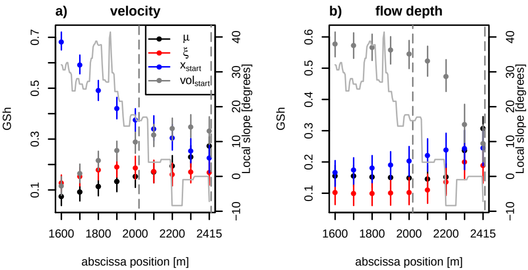
Interpretation

If we consider the inputs $X_1, X_2, \ldots, X_d$ as the team members trying to explain the variance of the output Y, then the set of Shapley effects $\{\phi_1, \ldots, \phi_d\}$ is the unique way to allocate $V[Y]$ to all players characterized by desirable properties known as Shapley axioms (see [Sha53] for more details).

How does it look like for the avalanche application?



Aggregated Shapley effects of velocity and flow depth curves calculated over space intervals $[x, 2412m]$ where $x \in \{1600m, 1700m, \ldots, 2412m\}$



We have $n = 6152$, $N_{\text{tot}} = 2002$, $B = 500$. Effects are estimated using the first (2, *resp.* 4) fPCs (see Yao *et al.*, 2005, Ramsay *et al.*, 2005) explaining more than 95% of the variance. Local slope is drawn with a gray line. A gray dotted rectangle is drawn at $[2017m, 2412m]$ where avalanche return periods vary from 10 to 10 000 years.

Outline

Conclusion, perpectives

UQ is an essential phase of forecasting.

SA helps:

- in understanding model behavior,
- as a preliminary step to model calibration,
- as a tool for decision support.

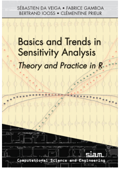
It may help in many other tasks:

- it is possible to use GSA for the construction of ensembles based on parameter perturbation (see Meryl Wimmer's PhD, 2021);
- it is possible to combine a GSA and a recursive Bayesian filtering approach for data-driven data assimilation (see, e.g., Hirvoas *et al*, 2022).

Important issue: how to perform UQ with as few as possible model runs?

Even so, SA from initial model can be prohibitive. Thus the importance of metamodelling: Gaussian Process Regression (kriging), Polynomial Chaos, Physics-Informed Neural Networks. . .

Today I only presented the basics of SA. A deeper review with practical implementation in R can be found in



with codes freely available at `https://bookstore.siam.org/cs23/bonus`

Thanks for your attention!

Questions?

A short bibliography I

[Bor07] E. Borgonovo. A new uncertainty importance measure. *Reliability Engineering and System Safety*, 92(6):771–784, 2007.

[Cha20] S. Chatterjee. A new coefficient of correlation. *Journal of the American Statistical Association*, 0(0):1–21, 2020.

[DV21] S. Da Veiga. Kernel-based anova decomposition and shapley effects–application to global sensitivity analysis. *arXiv preprint arXiv:2101.05487*, 2021.

[GGKL] F. Gamboa, P. Gremaud, T. Klein, and A. Lagnoux. Global sensitivity analysis: a new generation of mighty estimators based on rank statistics. *To appear in Bernoulli*.

[GJK+16] F. Gamboa, A. Janon, T. Klein, A. Lagnoux, and C. Prieur. Statistical inference for sobol pick-freeze monte carlo method. *Statistics*, 50(4):881–902, 2016.

[HPE22] M. B. Heredia, C. Prieur, and N. Eckert. Global sensitivity analysis with aggregated shapley effects, application to avalanche hazard assessment. *Reliability Engineering & System Safety*, 222:108420, 2022.

[JKL+14] A. Janon, T. Klein, A. Lagnoux, M. Nodet, and C. Prieur. Asymptotic normality and efficiency of two sobol index estimators. *ESAIM: Probability and Statistics*, 18:342–364, 2014.

[LMM11] M. Lamboni, H. Monod, and D. Makowski. Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. *Reliability Engineering and System Safety*, 96(4):450–459, 2011.

A short bibliography II

[MDN18] V. Maume-Deschamps and I. Niang. Estimation of quantile oriented sensitivity indices. *Statistics & Probability Letters*, 134:122–127, 2018.

[Mor91] M. D. Morris. Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2):161–174, 1991.

[Naa98] M. Naaim. Dense avalanche numerical modeling: interactFangion between avalanche and structures. In *25 years of snow avalanche research, Voss, NOR, 12-16 May 1998*, pages 187–191, Norway, 1998.

[Owe14] A. B. Owen. Sobol' indices and shapley value. *SIAM/ASA Journal on Uncertainty Quantification*, 2(1):245–251, 2014.

[Ple06] R.-E. Plessix. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, 167(2):495–503, 2006.

[PVBB19] C. Prieur, L. Viry, E. Blayo, and J-M Brankart. A global sensitivity analysis approach for marine biogeochemical modeling. *Ocean Modelling*, 139:101402, 2019.

[RJS+21] S. Razavi, A. Jakeman, A. Saltelli, C. Prieur, et al. The future of sensitivity analysis: An essential discipline for systems modeling and policy support. *Environmental Modelling and Software*, 137:104954, 2021.

[RS05] J. O. Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer Series in Statistics. Springer, 2nd edition, June 2005.

A short bibliography III

[Sal02] A. Saltelli. Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145:280–297, 2002.

[SCS00] A. Saltelli, K. Chan, and E. M. Scott. *Sensitivity Analysis*. John Wiley & Sons, 2000.

[Sha53] L. S. Shapley. A value for n-person games. In H. W. Kuhn and A. W. Tucker, editors, *Contribution to the Theory of Games II (Annals of Mathematics Studies 28)*, pages 307–317. Princeton University Press, Princeton, NJ, 1953.

[SK09] I. M. Sobol' and S. Kucherenko. Derivative based global sensitivity measures and the link with global sensitivity indices. *Mathematics and Computers in Simulation*, 79:3009–3017, 2009.

[TL03] J. Tardieu and M. Leroy. Radome, le réseau temps réel d'observation au sol de météo-france. *La Météorologie*, 2003.

[Wim21] M. Wimmer. *Représentation des erreurs de modélisation dans le système de prévision d'ensemble régional PEARO*. PhD thesis, Université Paul Sabatier-Toulouse III, 2021.

[WRD+22] M. Wimmer, L. Raynaud, L. Descamps, L. Berre, and Y. Seity. Sensitivity analysis of the convective-scale arome model to physical and dynamical parameters. *Quarterly Journal of the Royal Meteorological Society*, 148(743):920–942, 2022.

[YMW05] F. Yao, H.-G. Müller, and J.-L. Wang. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590, 2005.

A short bibliography IV

[ZCPM20] O. Zahm, P. Constantine, C. Prieur, and Y. Marzouk. Gradient-based dimension reduction of multivariate vector-valued functions. *SIAM Journal on Scientific Computing*, 42(1):A534–A558, 2020.