Sensitivity Analysis of Spatio-Temporal Models Describing Nitrogen Transfers, Transformations and Losses at the Landscape Scale

Pierre BARBILLON[1]

Joint work with:
Jordi FERRER-SAVALL[2], Damien FRANQUEVILLE[2], Cyril BENHAMOU[2], Patrick DURAND[3], Marie-Luce TAUPIN[4], Hervé MONOD[4] and Jean-Louis DROUET[2]

[1] UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay
[2] UMR ECOSYS, INRA, AgroParisTech, Université Paris-Saclay
[3] UMR SAS, INRA, Agrocampus Ouest
[4] UMR MaIAGE, INRA, Université Paris-Saclay

UQSay #03, June 13, 2019

Outline

Outline

1. ESCAPADE ANR and NitroScape Computer Model
2. Design of Experiments
3. Sensitivity Analyses
4. Analyses of some Results
   - Cluster Sensitivity Analysis
   - Time-Series Analysis
   - Spatial Map Analysis

Nitrogen Cascade in the Landscape

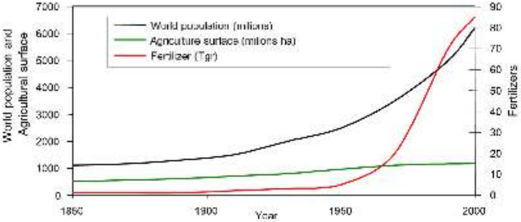


photo: Mark Sutton

photo: Gilles Billen

ANR ESCAPADE

Funded: February, 1 2013 - July, 31 2017

PI: Jean-Louis Drouet (ECOSYS)

Workpackages:

- T1: design, implement and analyze scenarii of nitrogen management in landscapes and in larger areas.
- T2: knowledge building, Nitrogen cascade modeling in landscapes.
- T3: transfer knowledge built on landscapes to larger areas.
- T4: data acquisition for parametrization, calibration and evaluation of developed models.

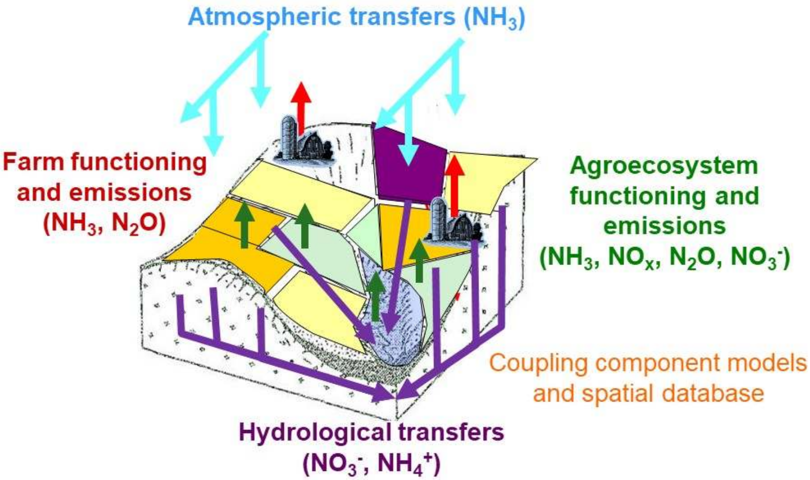
http://www.n-escapade.fr/

NitroScape Computer Model

[Duretz et al., 2011]

- Deterministic, spatially distributed and dynamic model describing N_r transfers and transformations in rural landscapes.
- couples four modules characterizing:
  1. farm management,
  2. biotransformations;
  3. transfers by the atmospheric pathways,
  4. transfers by the the hydrological pathways.
- Simulates the concentrations and fluxes, including the losses, of different forms of N_r within and between several landscape compartments:
  - the atmosphere,
  - the hydro-pedosphere (soil, water table, groundwater and streams),
  - the terrestrial agroecosystems (livestock buildings, croplands, grasslands and semi-natural areas)

NitroScape Computer Model

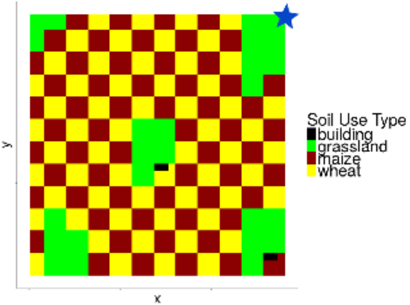


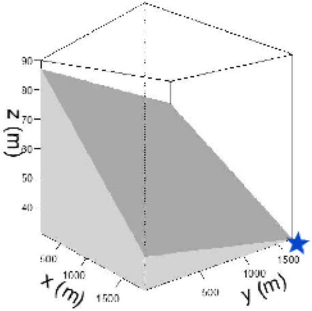
Outline

Virtual Landscape

Simplified theoretical landscape of 300 ha corresponding to an intensive rural area:

- succession of maize and wheat crops in a checkerboard distribution (125 ha each),
- pig farming buildings (two separate buildings, one ha each),
- unmanaged grasslands (5 plots scattered, 48 ha in total),
- Topography: linear slope with a gradient of 50 m,
- Meteorology: humid climatic conditions and little temperature contrasts.





Blue Star is the catchment outlet of the landscape.

Considered Factors

Factor	Description	Levels	Unit
A	Grid cell width (horizontal resolution)	12.5, 25, 50	m
B	Soil layer depth (vertical resolution)	0.02, 0.05, 0.1	m
C	Soil lateral transmissivity	2, 8, 15	m^2/day
D	Depth of exponential decrease of soil transmissivity	0.001, 0.01, 0.1	m
E	Surface layer (HS) depth	0.2, 0.3, 0.4	m
F	Soil porosity of the surface layer	0.12, 0.24, 0.48	-
G	Ratio of soil microporosity to macroporosity	0.5, 1, 1.2	-
H	Intermediate layer (HI) depth	0.6, 0.9, 1.2	m
I	Ratio of microporosity between layers (HI/HS)	1, 0.75, 0.5	-
J	Type of nitrogen fertilization	OL, OF, INO	-
K	Amount of nitrogen in fertilizer	$X \pm 20\%$	kg N ha^{-1}

Table: Input factors of NitroScape that were varied in the numerical experiments. OL: organic liquid manure, OF: organic solid fertilizer, INO: inorganic mineral fertilizer. The amounts of nitrogen in fertilizer were set at three levels: a fixed value (X) that depends on the type of fertilization, the number of applications and the type of crops (average value: 180 $kg\ N\ ha^{-1}\ yr^{-1}$), and two values at $\pm 20\%$ of the fixed value. Input factors from C to I were set as constant throughout the landscape and input factors J and K took non-zero values for fertilization events only.

Design of Experiments

- Input factors are a mix of quantitative and qualitative variables.
- Screening-design approach using a fractional factorial design (FFD) for the 11 factors [Saltelli et al., 2000] with 3 levels each.
- 3 levels for enabling the detection of non-monotonic effects for quantitative factors.
- Chosen resolution V to make possible the estimation of main effects and pairwise interactions without confounding effect in an ANOVA model [Box and Draper, 1987].
- Saturated design $\Rightarrow$ No residual degree of freedom to estimate the variance.
- Generated by the R package *planor* [Kobilinsky et al., 2012].

Model Settings

- Simulations performed at a daily time step and integrated over a five-year period, starting from January 1, 2007.
- First two years of simulation were used for model initialization and the sensitivity analysis used the results provided by the last three years of simulation only.
- Daily outputs were sampled from the variables simulated at the catchment outlet and monthly outputs were sampled from results obtained at different locations within the landscape.
- Atmospheric dispersion, transfer and deposition were not taken into account in this exploratory study since running the atmospheric component of NitroScape is very time-consuming.
- Each spatial output was considered at the highest resolution by dividing the coarse grid cell and copying the values of the concentration and dividing the values of the flows.

Outputs of NitroScape

29 N_r-related model outputs:

- 5 variables describing the fluxes and concentrations of N_r at the catchment outlet (*e.g.* daily NO_3^- concentration and amount),
- 9 spatially-distributed variables describing the fluxes at the interface between compartments (*e.g.* evapotranspiration, amount of mineralized NH_4^+ or NO_3^-),
- 15 spatially-distributed variables describing the local state of the compartments (*e.g.* NH_4^+ or NO_3^- concentration in groundwater or in soil).

Spatial, Temporal and Spatio-Temporal Aggregations

Spatially-distributed outputs formed large sets of data that were difficult to handle with conventional statistical tools: each output was described by a matrix of 243 rows and up to more than 7.10^5 columns.

- Each row corresponds to a configuration of the FFD,
- each column corresponds to an output variable in each grid cell at a given date of the theoretical landscape.
  For instance, for the highest horizontal resolution, the theoretical landscape included 19,600 grid cells, each characterized by the value of the 36 simulated monthly output variables, which resulted in 705,600 columns

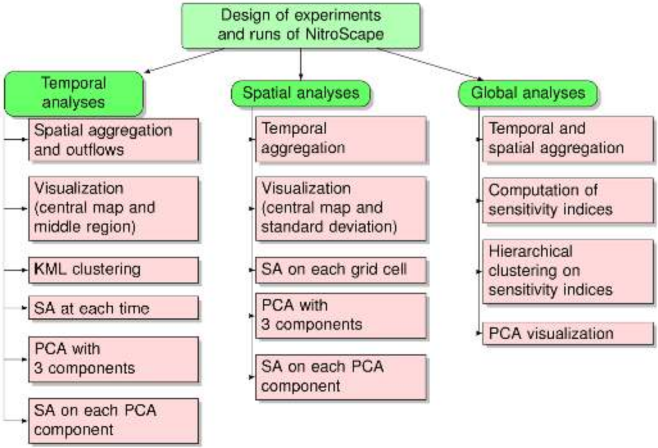
Output variables were spatially- or temporally-aggregated to produce different types of data sets:

- time series describing spatially-aggregated outputs or outputs at the outlet catchment,
- maps of temporally-aggregated outputs were used for spatial sensitivity analysis,
- synthetic view of the sensitivity of model outputs to input factors by spatial and temporal aggregation.

N.B. Flows were summed up and concentration were averaged.

Outline

Worflow of Sensitivity Analyses



For $1 \leq i \leq 243$ a configuration of the FFD, we assume an ANOVA model for an output of the model: Y_i (which may result from an aggregation or a projection (PCA))

$$Y_i = f(x_{i,1}, \ldots, x_{i,p}) = \mu + \sum_{j=1}^{p} \alpha^{(j)}_{x_{i,j}} + \sum_{1 \leq j < j' \leq p} \beta^{(j,j')}_{x_{i,j}, x_{i,j'}} + E_i$$

where $\alpha^{(j)}_{x_{i,j}}$ is the main effect of factor j ($1 \leq j \leq p = 11$) on the output and $\beta^{(j,j')}_{x_{i,j}, x_{i,j'}}$ is the pairwise second order interactions between factors j and j' on the output, with $1 \leq j < j' \leq p$.
$x_{i,j}$ stands input factor j of the configuration i of the FFD. The three different levels of each factor j are denoted by k ($k = 1, 2, 3$).

- Effects were estimated by using the least squares method,
- FFD being saturated, the residual terms E_i were all zero. The residual variance could not be therefore estimated.
- The residual variance would have only corresponded to interactions of order higher than two.

Sensitivity Indices

- Total Sum of Squares: $TSS = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$.
- Main Effect of factor j:

$$mSI_j = \sum_{k=1}^{3} \#\mathcal{X}_j^{(k)} \cdot (\bar{Y}_j^{(k)} - \bar{Y})^2 \Big/ TSS$$

where $\bar{Y} = \frac{1}{n} Y_i$ is the overall average of Y_i's, $\mathcal{X}_j^{(k)} = \{1 \leq i \leq n : x_{i,j} = k\}$ are the sets of configurations i such that the factor j has level k, $\#$ denotes the cardinal of a set, $\bar{Y}_j^{(k)} = 1/\#\mathcal{X}_j^{(k)} \cdot \sum_{i \in \mathcal{X}_j^{(k)}} Y_i$ are the means for the levels k of factor j

- Pairwise interaction effects of factors $j \neq j'$:

$$SI_{j,j'} = \sum_{k,k'=1}^{3} \#\mathcal{X}_{j,j'}^{(k,k')}(\bar{Y}_{j,j'}^{(k,k')} - \bar{Y}_j^{(k)} - \bar{Y}_{j'}^{(k')} + \bar{Y})^2 \Big/ TSS$$

where $\mathcal{X}_{j,j'}^{(k,k')} = \{1 \leq i \leq n : x_{i,j} = k \text{ and } x_{i,j'} = k'\}$ are the sets of configurations i such that the factor j (resp. j') has level k (resp. k') and $\bar{Y}_{j,j'}^{(k,k')} = 1/\#\mathcal{X}_{j,j'}^{(k,k')} \cdot \sum_{i \in \mathcal{X}_{j,j'}^{(k,k')}} Y_i$.

Sensitivity Indices

- Sum of pairwise interactions:

$$iSI_j = \sum_{j':j' \neq j} SI_{j,j'}$$

- Total effect: an index describing the total (*i.e.* main and interaction) effect of factor j:

$$tSI_j = mSI + iSI_j$$

- Total interactions: and an index describing the sum of interactions between all factors:

$$i_{tot} = \sum_{1 \leq j < j' \leq p} SI_{jj'}$$

FFD saturated $\Rightarrow \sum_j mSI_j + i_{tot} = 100\%$.

Principal Component Analysis

2 uses of PCA.

1. PCA was used to reduce data redundancy and identify features linked to model structure such as
   - seasonality in time series,
   - land use in maps.

   PCA was applied to
   - $\mathbf{Y}$ (= $(Y_{it})_{1 \le i \le 243,\ 1 \le t \le 36}$) data set that describes temporal outputs simulated at the catchment outlet or spatially-aggregated outputs
   - $\mathbf{Y}$ (= $(Y_{is})_{1 \le i \le 243,\ 1 \le s \le nc}$ data set that describes spatially-distributed and temporally-aggregated outputs. nc=19,600 grid cells of size 12.5 m x 12.5 m.

2. PCA was applied to ensemble of sensitivity indices of the ensemble of temporally- and spatially-aggregated outputs. PCA was applied to the data set $\mathbf{S}$ (=$(S_{kl})_{1 \le k \le 29,\ 1 \le l \le 66}$), in which each row corresponds to each of the 29 considured outputs of NitroScape and each column corresponds to each of the 11 main sensitivity indices and each of the 55 ($= \binom{11}{2}$) pairwise interaction indices.

R packages: *Multisensi* [Lamboni et al., 2009] and *FactoMineR* [Husson et al., 2008] to carry out these analyses.

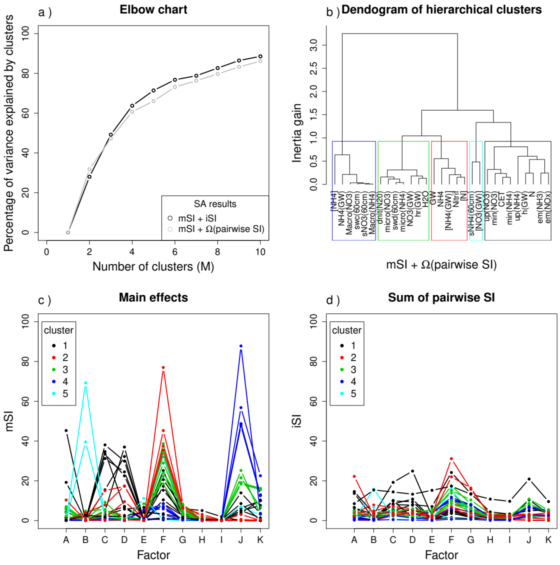
Outline

1. ESCAPADE ANR and NitroScape Computer Model
2. Design of Experiments
3. Sensitivity Analyses
4. Analyses of some Results
   - Cluster Sensitivity Analysis
   - Time-Series Analysis
   - Spatial Map Analysis

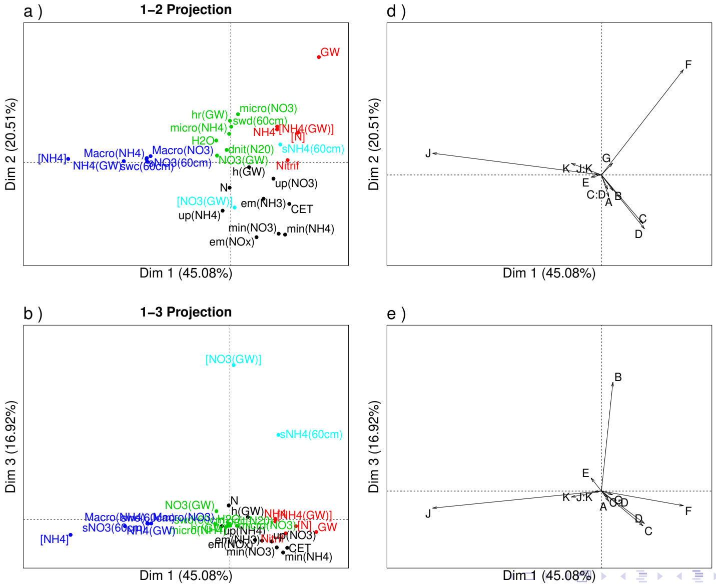
Outline

1. ESCAPADE ANR and NitroScape Computer Model
2. Design of Experiments
3. Sensitivity Analyses
4. Analyses of some Results
   - Cluster Sensitivity Analysis
   - Time-Series Analysis
   - Spatial Map Analysis

Hierarchical Clustering of Outputs according to their Sensitivity Indices



Projection of Outputs



Cluster	N	Output variables	Characteristics
K=1 / black	9	Evapotranspiration, nitrogen emission and uptake by plants, nitrogen mineralization, depth of the groundwater table and total amount of NO_3^- at the catchment outlet.	Mostly affected by soil lateral water transmissivity (factor C) and soil transmissivity decrease with depth (factor D).
K=2 / red	5	Total amount of NO_3^- and NH_4^+ discharged at the catchment outlet, nitrification, total amount of NO_3^- and NH_4^+ in groundwater.	Mostly affected by soil surface porosity (factor F).
K=3 / green	7	Surface water depths, streaming flow and water discharge at the catchment outlet, nitrogen adsorbed in soil microporosity, soil NO_3^- in groundwater.	Mostly affected by soil surface porosity (factor F), fertilization type (factor J), soil lateral transmissivity (factor C) and soil transmissivity decrease with depth (factor D). Moderate effect of interaction terms.
K=4 / dark blue	6	NH_4^+ concentration in groundwater and at the catchment outlet, nitrogen adsorbed in soil macroporosity, soil NO_3^- concentration in soil surface.	Mostly affected by fertilization type (factor J), high effect of the interaction term J:K.
K=5 / light blue	2	NH_4^+ concentration in soil surface and NO_3^- concentration in groundwater.	Mostly affected by vertical resolution (factor B).

Table: Description of the five clusters of spatially- and temporally-aggregated outputs found by a hierarchical clustering on their corresponding main and second order sensitivity indices. The cluster columns provides the number of each cluster and its related color on Figures 6 and 7. The *N* column provides the number of output variables included in each cluster.

Outline

Visualization and clustering

- **Visualization** Time series of the highest densities of outputs obtained from a functional boxplot of the highest density region (HDR) [Hyndman and Shang, 2010, Hyndman, 1996].
  (HDR boxplots were defined by computing a bivariate kernel density estimate on the first two principal components of the time-series and then applying the bivariate HDR boxplot).
  Middle region that contained half of the time series was also plotted (adaptations of functions of the *Rainbow* R-package).
- **Clustering** 243 time series simulated from the different configurations of the FFD were split into three clusters that grouped curves with similar features (*e.g*. slope, range of variation).
  R package *KML* [Genolini et al., 2015] based on a k-means algorithm applied to the features of the curves.
  Number of clusters set to three $=$ number of levels for each input factor.

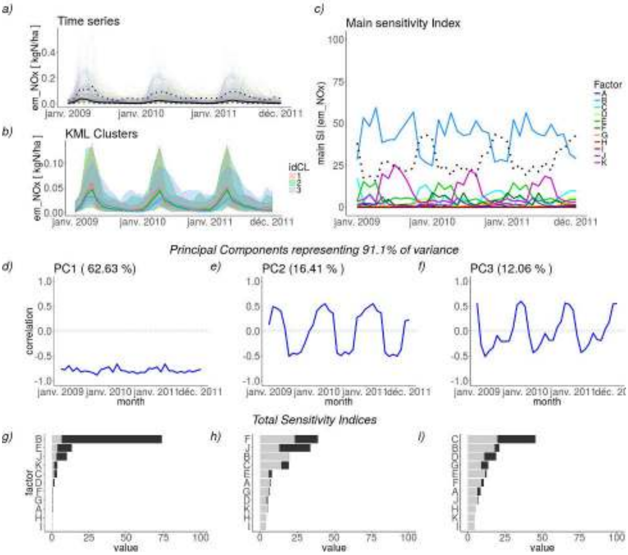


Figure: Temporal sensitivity analysis of NO_x emissions simulated for the whole landscape and averaged by area unit;

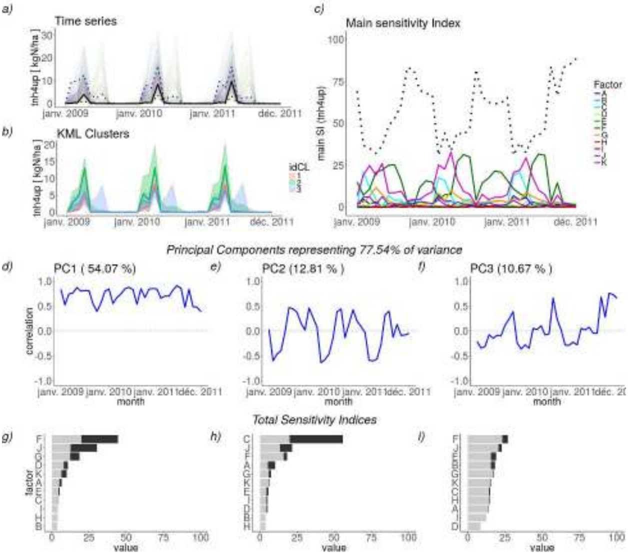


Figure: Temporal sensitivity analysis of NH_4^+ uptake by plants simulated for the whole landscape and averaged by area unit;

Outline

- **Visualization** Central map obtained from a functional boxplot of the highest density region (HDR).

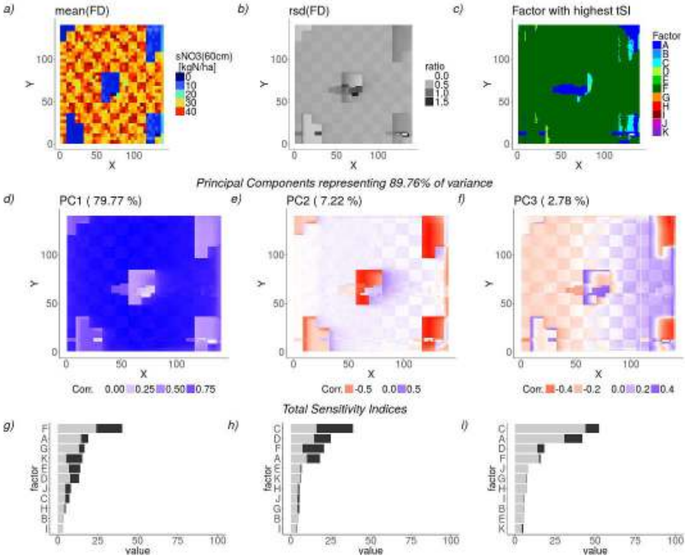


Figure: Spatial sensitivity analysis of soil NO_3^- amount between 0 and 60 cm depth cumulated on the three-year period of interest in each grid cell of the landscape and averaged by area unit;

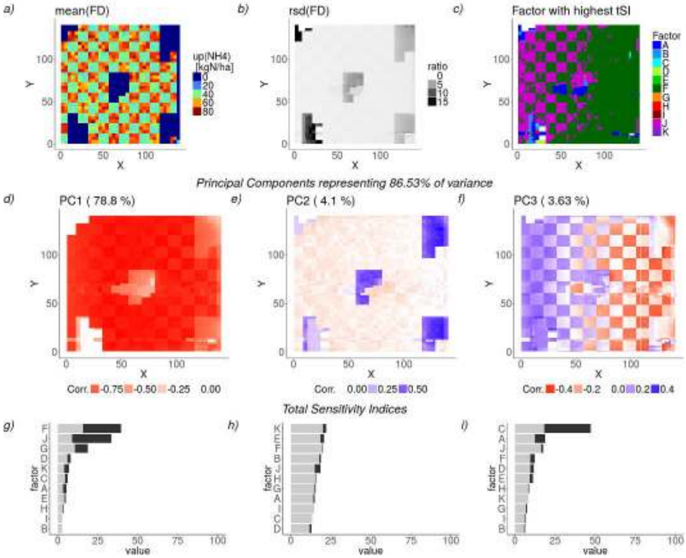


Figure: Spatial sensitivity analysis of NH_4^+ uptake by plants cumulated on the three-year period of interest in each grid cell of the landscape and averaged by area unit;

Concluding Remarks

First analysis of NitroScape

- Clustering methods for a global SA with several outputs,
- Adapted plots for time-series and map outputs.
- SA also applied to real landscape.

Further works:

- Other types of data aggregation: e.g. Sensitivity analyses w.r.t. land managements (type of crop, farms, grassland), meteorological inputs ...
- Challenging emulation of 4 daily-coupled computer models.
- Calibration on the basis of available field data.

Box, G. E. and Draper, N. R. (1987).
Empirical model-building and response surfaces.
John Wiley & Sons.

Duretz, S., Drouet, J.-L., Durand, P., Hutchings, N. J., Theobald, M., Salmon-Monviola, J., Dragosits, U., Maury, O., Sutton, M., and Cellier, P. (2011).
Nitroscape: a model to integrate nitrogen transfers and transformations in rural landscapes.
Environmental Pollution, 159(11):3162–3170.

Genolini, C., Alacoque, X., Sentenac, M., and Arnaud, C. (2015).
kml and kml3d: R packages to cluster longitudinal data.
Journal of Statistical Software, Articles, 65(4):1–34.

Husson, F., Josse, J., Lê, S., and Mazet, J. (2008).
Factominer: an r package for multivariate analysis.
Journal of statistical software, 25(1):1–18.

Hyndman, R. J. (1996).
Computing and graphing highest density regions.
The American Statistician, 50(2):120–126.

Hyndman, R. J. and Shang, H. L. (2010).
Rainbow plots, bagplots, and boxplots for functional data.
Journal of Computational and Graphical Statistics, 19(1):29–45.

Kobilinsky, A., Bouvier, A., and Monod, H. (2012).
Planor: an r package for the automatic generation of regular fractional factorial designs.
Technical report.

Lamboni, M., Makowski, D., Lehuger, S., Gabrielle, B., and Monod, H. (2009).
Multivariate global sensitivity analysis for dynamic crop models.
Field Crops Research, 113(3):312–320.

Saltelli, A., Chan, K., and Scott, E. M. (2000).
Sensitivity analysis, volume 1.
Wiley New York.

Savall, J. F., Franqueville, D., Barbillon, P., Benhamou, C., Durand, P., Taupin, M.-L., Monod, H., and Drouet, J.-L. (2019).
Sensitivity analysis of spatio-temporal models describing nitrogen transfers, transformations and losses at the landscape scale.
Environmental modelling & software, 111:356–367.