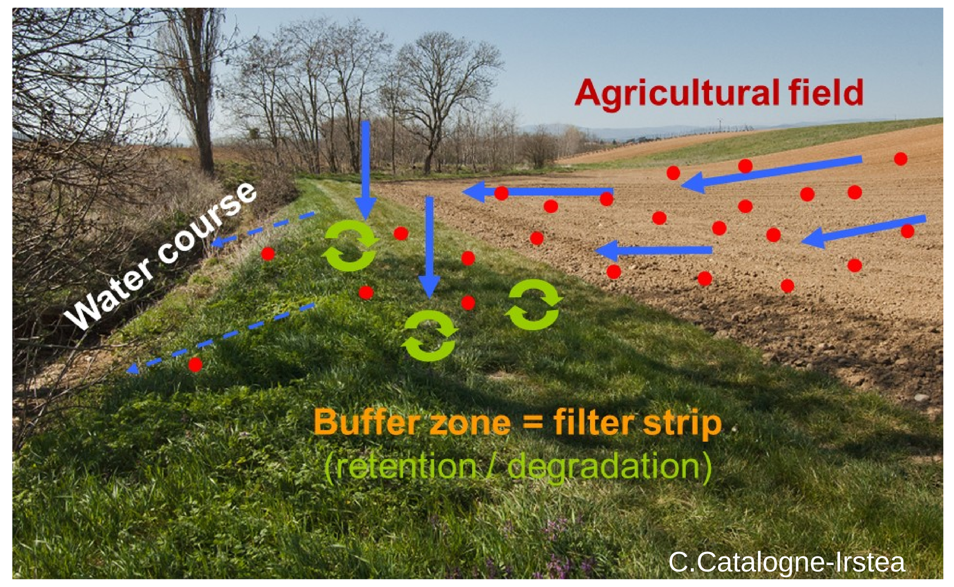
Development of metamodeling methods considering qualitative variables to evaluate a decision-making tool of pesticide transfers

Claire Lauvernet, C. Helbert



Development of metamodeling methods considering qualitative variables to evaluate a decision-making tool of pesticide transfers.

Lauvernet Claire[1], Helbert Céline[2]

[1] RIVERLY, Irstea centre de Lyon-Villeurbanne, France.

[2] Univ Lyon, UMR CNRS 5208, École Centrale de Lyon, France.

Contact : claire.lauvernet@irstea.fr

Objectives

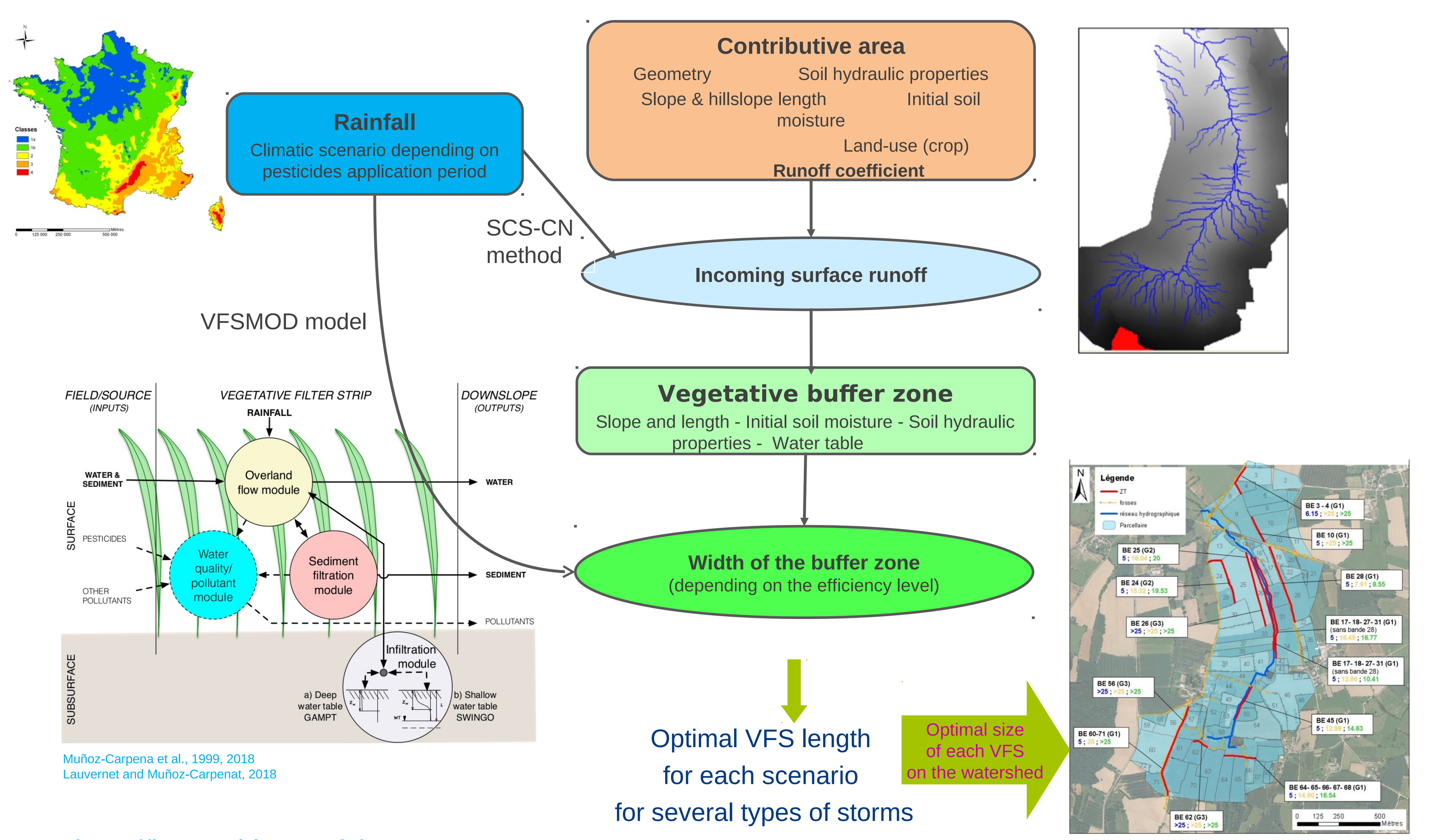
- Vegetative filter strips are identified as the BMP of Choice for Runoff Mitigation to limit contamination of surface water by pesticides.
- Their efficiency strongly depends on soil, agronomic and climatic conditions and they need to be optimized by considering appropriate sizing.
- ➔ Irstea developed a complete/complex toolkit to design site-specific VFS by simulating their efficiency to limit runoff transfers : BUVARD (Carluer et al., 2017).
- This tool is based on quantitative *and* qualitative variables
- Need for a simpler and efficient tool for end-users
- ➔ A metamodel of BUVARD = methodological challenges for operational purposes

The toolkit BUVARD*

BUffer strip for runoff Attenuation and pesticides Retention Design tool

BUVARD online

Hypothesis : buffer zone efficiency = ability to retain surface runoff



* http://buvard.irstea.fr/

Ref. Carluer, N et al. 2017. *Defining context-specific scenarios to design vegetated buffer zones that limit pesticide transfer via surface runoff*. Sci. Of Total Env. , 575. // Chen, Wang & Yang, 2013, *Stochastic Kriging with Qualitative factors.* WSC '13 Proceedings. // Muñoz-Carpena, R., 1999; Parsons, J. E. & Gilliam, J. W. Modeling hydrology and sediment transport in vegetative filter strips J.Hydrol., 214, 111-129. // Munoz-Carpena, R.; Lauvernet, C. & Carluer, N., 2018. *Shallow water table effects on water, sediment, and pesticide transport in vegetative filter strips -- Part 1: nonuniform infiltration and soil water redistribution.* Hydrol. Earth System Sci., 22. // Lauvernet, C. & Munoz-Carpena, R. 2018. *Part 2: model coupling, application, factor importance, and uncertainty.* Hydrol. Earth System Sci., 22. // Panodou & Roustant, 2017 *mixgp R package: Kriging models for mixed data. // Roustant, O., Padonou, E., Deville, Y., Clement, A., Perrin, G., Giorla, J., Wynn, H., 02 2019. Group kernels for gaussian process metamodels with categorical inputs. https://hal.archives-ouvertes.fr/hal-01702607.*

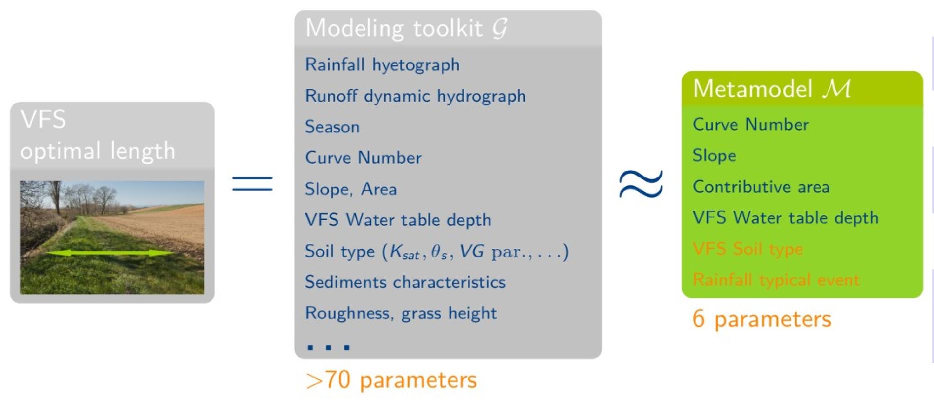
www.irstea.fr

National Research Institute of Science and Technology for Environment and Agriculture

Methods for mixed variables

Why metamodeling BUVARD?

- Simple to use but still based on physics
- Able to evaluate an output of the toolchain at any point of the domain
- Allows evaluating sensitivity indices at smaller numerical cost
- Can be easily coupled/integrated in hydrological modeling frameworks



Sampling range	Type of variable
[63 , 99]	quantitative
[2% , 20%]	quantitative
[25 , 300] m	quantitative
[0.5 , 4] m	quantitative
Summer/Winter,short/long: S01,S06,W02,W12	qualitative
Clay-loam to sandy-loam: clo, scl, SIL, CLO , SCL, SAL	qualititative

Design of experiments of the most influent and the most accessible inputs: maximin LHS composed of 100 points in the quantitative variable hypercube space per couple of (qualitative) modalities.

➔ training sample = 100 x 24 different pairs of modalities Soil Type × Rainfall test sample (independent LHS) = 40 × 24 points.

How to deal with qualitative/categorical variables (type of soil, type of rain)?

Gaussian processes /Kriging

The relation between points is expressed by a covariance structure between the obs.

➔ need for an adapted cov. kernel

GAM (Generalized Additive model)

The relation between points is assumed to be deterministic and explicit.

➔ ok with quali. var.

Gaussian Processes with mixed variables:

- Hyp. = the deterministic output of the model is the realization of a GP
- The GP *Z* is conditioned by points from the model simulations (still a GP)
- Several options to deal with quali. var.:
  - one GP per couple of modalities : => does not take any advantage of information available from other modalities
  - adapting the covariance kernel by progressive complexity, based on Roustant et al. (2019), Chen et al (2013)

Kernel	4 modalities	Corr. par.	Q2
cov-quali-isotropic $r(z_j - z'_j) = exp\left(-\rho \mathbf{1}_{z_j \neq z'_j}\right)$	$\begin{pmatrix} 1 & \rho & \rho & \rho \\ & 1 & \rho & \rho \\ & & 1 & \rho \\ & & & 1 \end{pmatrix}$	1	0.9636
cov-quali-product $r(z_j - z'_j) = exp\left(-(\rho_{z_j} + \rho_{z'_j})\mathbf{1}_{z_j \neq z'_j}\right)$	$\begin{pmatrix} 1 & \rho_1\rho_2 & \rho_1\rho_3 & \rho_1\rho_4 \\ & 1 & \rho_2\rho_3 & \rho_2\rho_4 \\ & & 1 & \rho_3\rho_4 \\ & & & 1 \end{pmatrix}$	4	0.9641
cov-quali-anisotropic $r(z_j - z'_j) = exp\left(-\rho_{z_j,z'_j} \mathbf{1}_{z_j \neq z'_j}\right)$	$\begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \rho_{14} \\ & 1 & \rho_{23} & \rho_{24} \\ & & 1 & \rho_{34} \\ & & & 1 \end{pmatrix}$	6	0.9416

$$r((x,w,z) - (x',w',z')) = r_{quanti}(x - x')r_{ordi}(w - w')r_{quali}(z - z')$$

corr. fun. of quanti. factor

$$r_{quanti}(x - x') = \prod_{j=1}^{p_x} \left(1 + \frac{\sqrt{5}|x_j - x'_j|}{\theta_j} + \frac{5(x_j - x'_j)^2}{3\theta_j^2}\right) \exp\left(-\frac{\sqrt{5}|x_j - x'_j|}{\theta_j}\right)$$

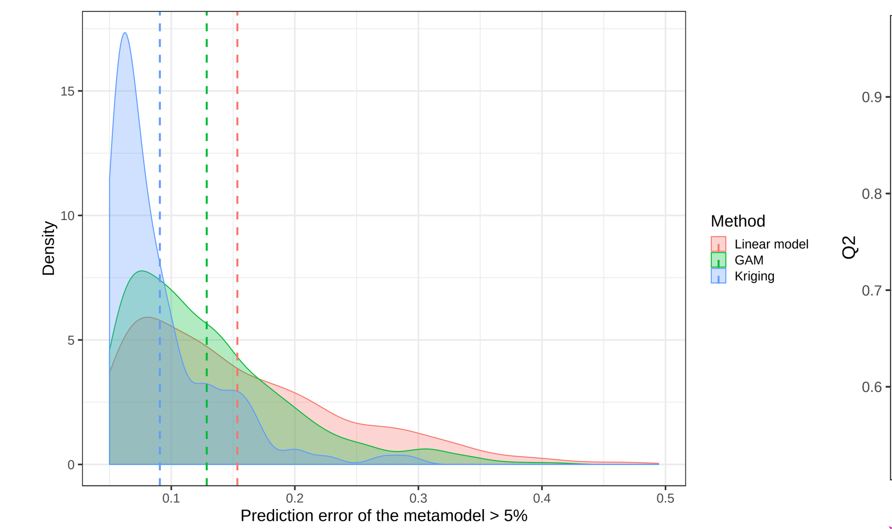
ordinality between soil types

corr. fun. of quali factor

➔ very high Q2 with the 3 kernels
cov-quali-product is selected for the study

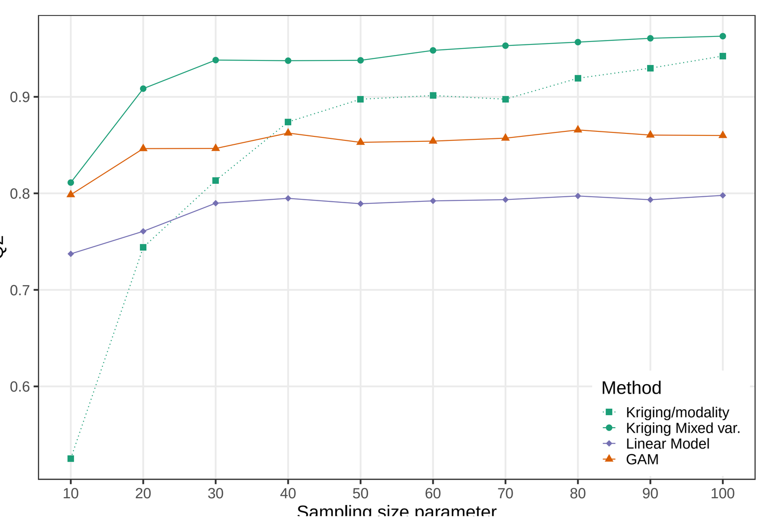
Evaluation of the methods

Performance of metamodeling methods



- ➢ Still some prediction error due to the plateau of null values: errors of kriging are concentrated next to zero.
- ➢ Average prediction errors are 0.15 for the linear model, 0.13 for GAM , and 0.09 for Gaussian process modeling. Kriging is a semi-parametric model, very smooth, as opposed to the linear model : it fits data very well without hard assumptions.

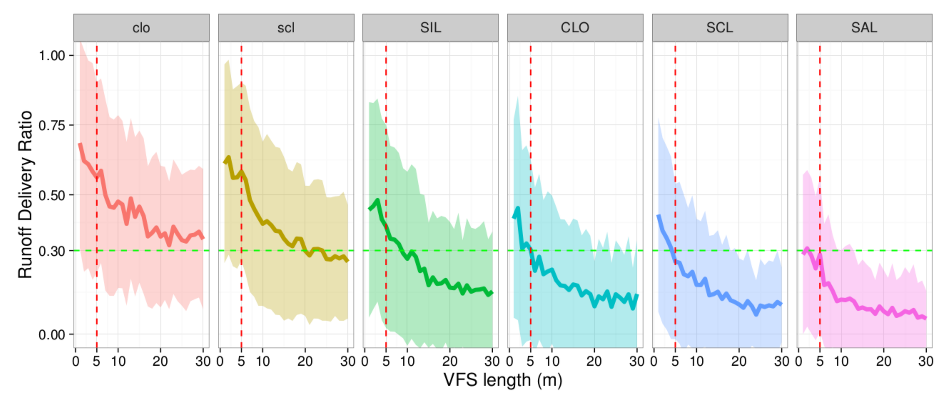
Effect of sampling size



- ➢ GAM is giving satisfying results but unstable for size<30
- ➢ Kriging with mixed var. is very robust whatever the sampling size
- ➢ For very small sampling size, mixed kriging improves drastically compared to by pair of modalities

Applications for Risk Analysis and Management

- Uncertainty analysis on large sample simulations



Output variable :

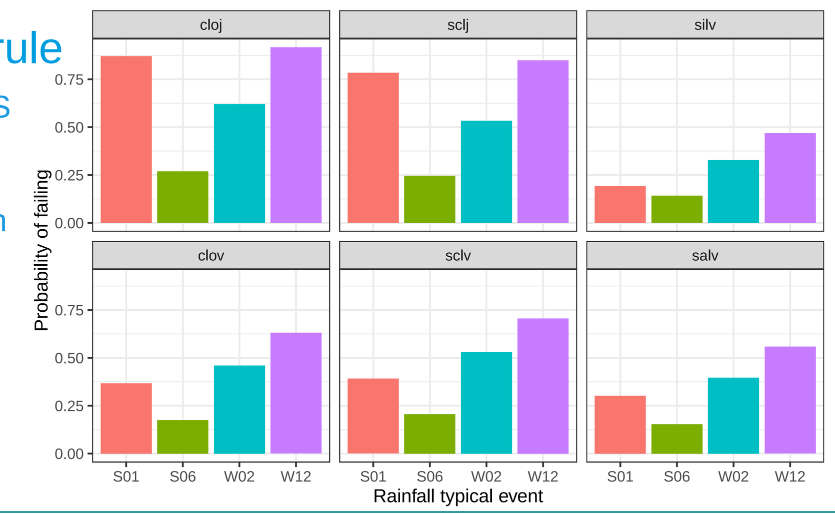
Runoff Ratio = Run_out/Run_in

RDR = 0 ⇔ efficiency = 100%

RDR = 1 ⇔ efficiency = 0%

- ➢ 24 000 sim. at a very low cost
- ➢ a very large variability of optimal length / type of soil
- ➢ possibility to quantify uncertainty of the results

- Probability of failing with a one-fits-all rule
  - ➢ Suppose the rule (in most european countries) is a VFS of 5m length along any classified river
  - ➢ It is particularly risky for short summer (S01h) and long winter (W12) events and for newly implanted VFSs with clay loam and sandy clay soils.
  - ➢ On the contrary, with established VFSs with clay loam and sandy clay soils, it can be considered reasonable
  - ➢ Globally, the risk of failing is more than 25% with this decision.
  - ➢ Depending on the type of soil and rainfall event, this is not an adapted rule. The MM is a way to test it.



Conclusion*

- Qualitative variables were properly taken into account by the GP adaptation
- MM is a promising tool to make the toolkit more operational and to perform UA at low cost
- The current MM is not satisfying on extremes, due to a large plateau of null values of the output variable. In the future :
  - global sensitivity analysis (Sobol)
  - test new methods : Chaos polynomials, random forests

Acknowledgments: Many thanks to Robert Faivre, Esperan Padonou, Oivier Roustant, Nicolas Forquet, for our fruitful discussions on metamodeling, and to Clotaire Catalogne, Nadia Carluer, Guy Le Hénaff, Etienne Leblois, Rafael Munoz-Carpena, for BUVARD development. These models have been implemented on logiciel R with package *kergp*. Thanks to LEFE-MANU that supported this study.

Lauvernet C. and Helbert C., Metamodeling methods including qualitative variables to design vegetative filter strips and reduce surface runoff pollution.Under revision in RESS.