# Comparison and evaluation of metamodeling methods with qualitative variables for the design of vegetative filter strips

Claire Lauvernet, Céline Helbert, Zhu Xujia, Bruno Sudret

Speaker: Claire Lauvernet

# Metamodeling methods that incorporate qualitative variables for improved design of vegetative filter strips.

*Claire Lauvernet*
RIVERLY, INRAE Lyon-Villeurbanne, FR, claire.lauvernet@inrae.fr
*Céline Helbert*
UMR CNRS 5208, Ecole Centrale de Lyon, FR, celine.helbert@ec-lyon.fr
*Zhu Xujia*
ETH Zürich, Chair of Risk, Safety & Uncertainty Quantification, CH, zhu@ibk.baug.ethz.ch
*Bruno Sudret*
ETH Zürich, Chair of Risk, Safety & Uncertainty Quantification, CH, sudret@ethz.ch

Significant amounts of pollutant are measured in surface water, their presence due in part to the use of pesticides in agriculture. One solution to limit pesticide transfer by surface runoff is to implement vegetative filter strips (VFS) along rivers. These buffer zones are identified as the best management practices of choice for runoff mitigation to prevent and limit the transfer of pollutants from agricultural fields to water resources. They are mandatory or highly advised depending on the country and conditions. Since their location is part of the farmer's field, the sizing of these strips is a major issue. However, to be efficient, they need to be properly designed, depending on the specific context in which they are implanted (climate, soil, water table, etc.).

The BUVARD modeling toolkit was developped to design VFSs throughout France according to all these local influencing factors [1]. Processes that drive the pesticide fate are complex and interact : infiltration, surface runoff, sediment trapping, pesticide transfer, etc., and are summarized through nonlinear equations and/or conceptual and/or stochastic modeling. To represent most of them, BUVARD is composed of several models centered around the numerical model VFSMOD [4], which quantifies dynamic effects of VFS site-specific pesticide mitigation efficiency (see figure 1).

The way BUVARD is built (i.e., a chain of several models) implies a large set of parameters that are difficult to measure (for the physical modeling) or to calibrate (for the conceptual modeling). Furthermore, inputs and outputs are dynamic (for example, rainfall, surface runoff, etc.), and inputs are either quantitative or qualitative variables. For all these reasons, we get an expensive tool to use, and a high uncertainty, which has to be quantified, particularly in the case of an operationnal tool. Metamodeling BUVARD is *a priori* a relevant solution to decrease the cost and the complexity of the model, to help users design VFSs that are adapted to specific contexts.

Added to the mixed qualitative and quantitative variables, that is not often taken into account in surrogate methods, we have to deal with a huge number of zero values of

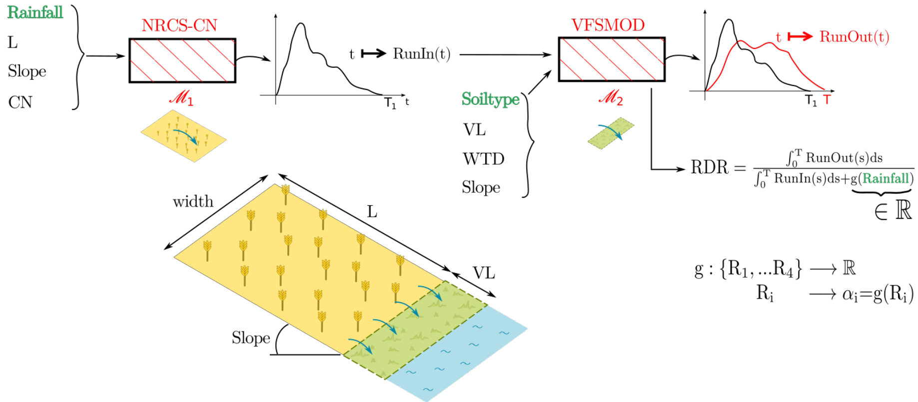


Figure 1: BUVARD toolchain, with inputs and outputs used to build the surrogate.

inputs and outputs, and to boundaries in which the main output has to range. In this study, different methods are tested: (i) Mixed-Kriging, a Kriging method that was implemented with a covariance kernel for a mixture of qualitative and quantitative inputs [3] (ii) PCE, that was also adapted to qualitative variables, encoding categorical inputs as quantitative dummy variables, thus allowing for transforming the problem into the standard regression setup [5] (iii) DeepGP (Deep Gaussian Processes), that is particularly suited for non-stationnary models [2]. We show that categorical variables are properly taken into account by the Kriging and by the PCE adaptations, and that mixed variables methods outperform the same methods applied per category, and even more with smaller samplings. DeepGP, that was not adapted to qualitative variables, does not need any classification or boundaries, and reaches the performance of the adapted methods. However, it needs repetitions for the most complex soils, with a much higher numerical cost, that is multiplied by the number of categorical variables. Finally, we perform a global sensitivity analysis with the help of the two surrogate models with the best accuracy. The results show that they give the same ranking of the importance of the input parameters.


[1] N. Carluer, C. Lauvernet, D. Noll, and R. Muñoz-Carpena. Defining context-specific scenarios to design vegetated buffer zones that limit pesticide transfer via surface runoff. *Science of The Total Environment*, 575:701–712, 2017.

[2] A. Damianou and N. D. Lawrence. Deep Gaussian processes. volume 31 of *Proceedings of Machine Learning Research*, pages 207–215, 2013. PMLR.

[3] C. Lauvernet and C. Helbert. Metamodeling methods that incorporate qualitative variables for improved design of vegetative filter strips. *Reliability Engineering & System Safety*, 204:107083, 2020.

[4] R. Muñoz-Carpena, J.E. Parsons, J.W. Gilliam. Modeling hydrology and sediment transport in vegetative filter strips. *Journal of Hydrology*, 214(1-4):111–129, 1999.

[5] X. Zhu and B. Sudret. Sparse polynomial chaos expansions for mixed categorical / continuous inputs. (in preparation).