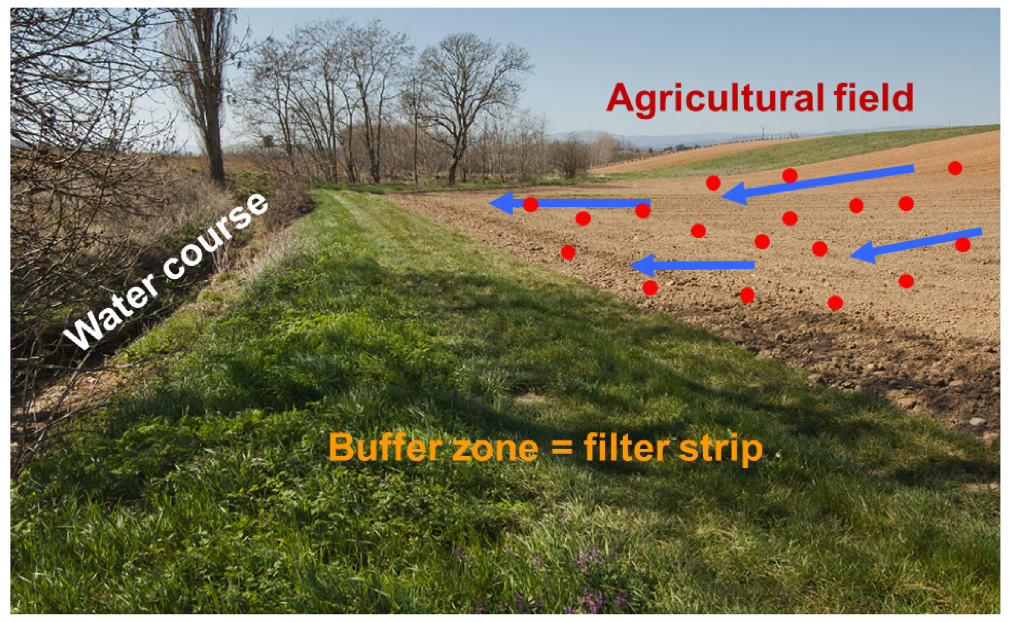
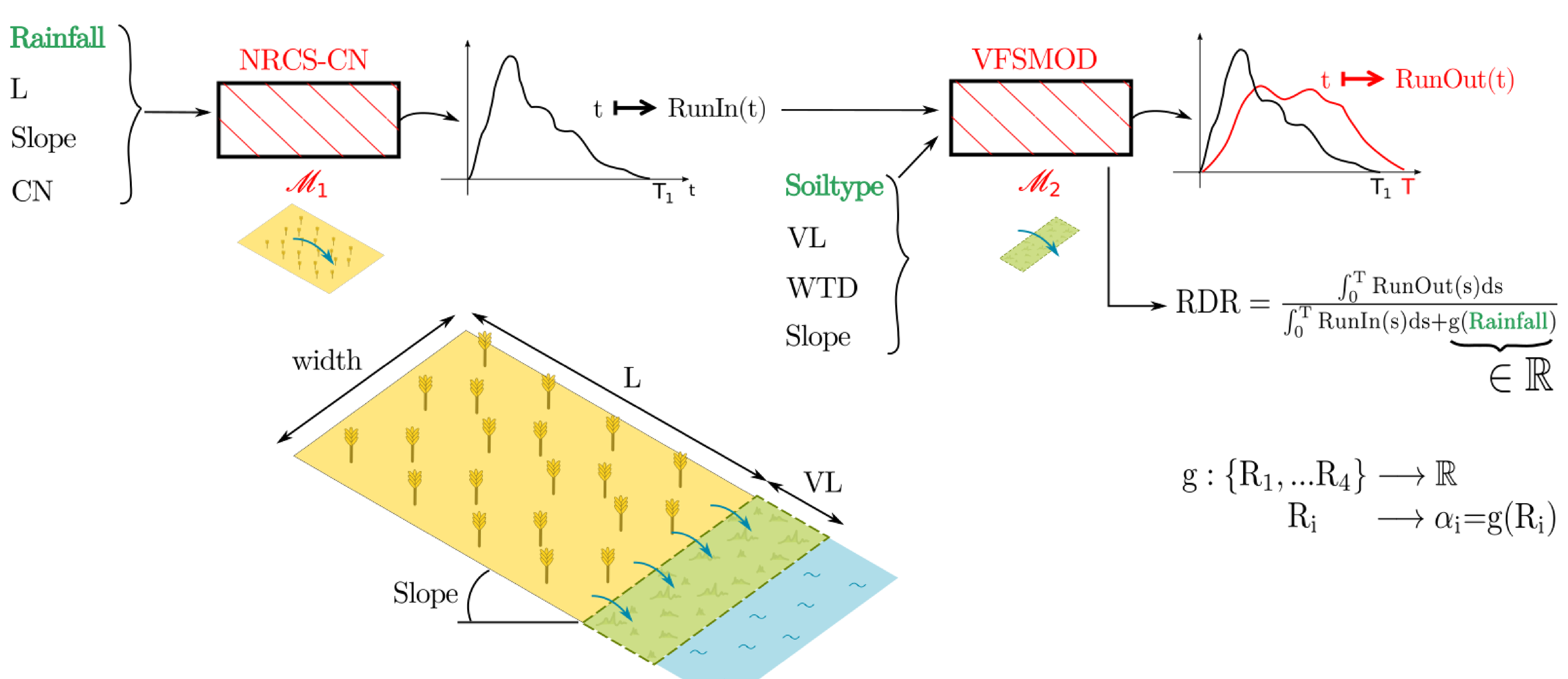



# Metamodeling methods that incorporate qualitative variables for improved design of vegetative filter strips


Claire Lauvernet, Céline Helbert, Xujia Zhu, Bruno Sudret

RIVERLY, INRAE Lyon, FR - UMR CNRS 5208, Ecole Centrale de Lyon, FR - ETH Zürich, Chair of Risk, Safety & Uncertainty Quantification, CH


## Context : The BUVARD tool




BUffer strip for runoff Attenuation and pesticides Retention Design [1]

- Physical processes are complex and interact
- based on non linear equations and/or conceptual and/or stochastic
- a large set of parameters that are difficult to measure/estimate
- a chain of several models, complex to use in practice

⇒ a high uncertainty in an operationnal context

Metamodeling BUVARD to bridge the gap between modeling and decision support

## Challenges for the surrogate of BUVARD

- a chain of several models (var. of interest = ratio)
- inputs are quantitative and qualitative (categorical)
- a huge number of zero values of Runin, Runout, and then RDR
- The output variable RDR has to range between 0 and 1

## References


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

[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] A. Damianou and N.D. Lawrence. Deep Gaussian processes. In Carlos M. Carvalho and Pradeep Ravikumar, editors, v. 31 of *Proceedings of Machine Learning Research*, 207–215, Scottsdale, Arizona, USA, 2013.

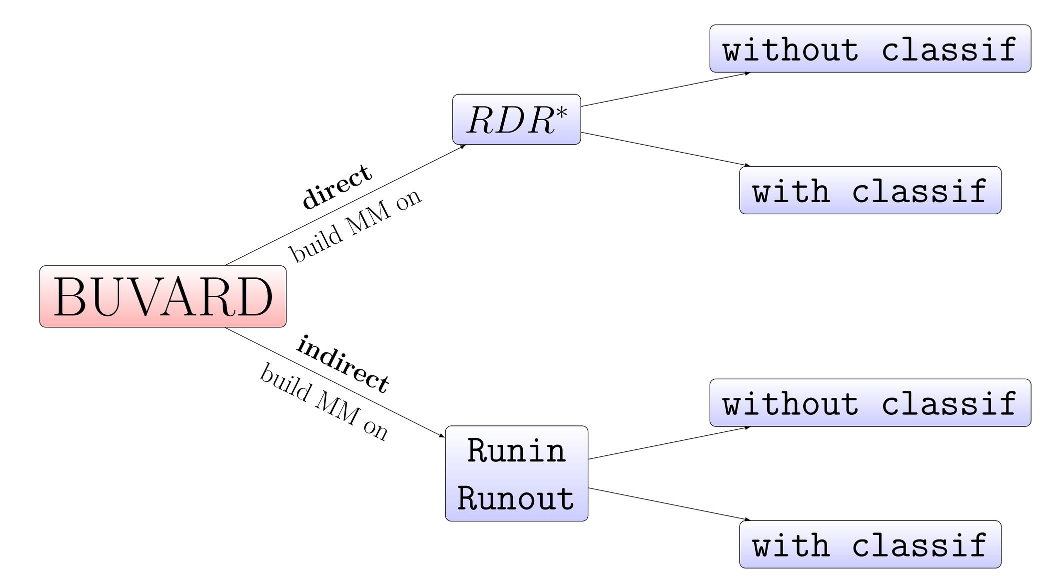
[5] X. Zhu and B. Sudret. Construction of sparse polynomial chaos surrogate model for simulators with mixed continuous and categorical variables. *Proc. 4th Int. Conf. Uncertainty Quantification in Computational Sc. and Engin. (UNCECOMP)*, Athens, 2021.


RÉPUBLIQUE FRANÇAISE

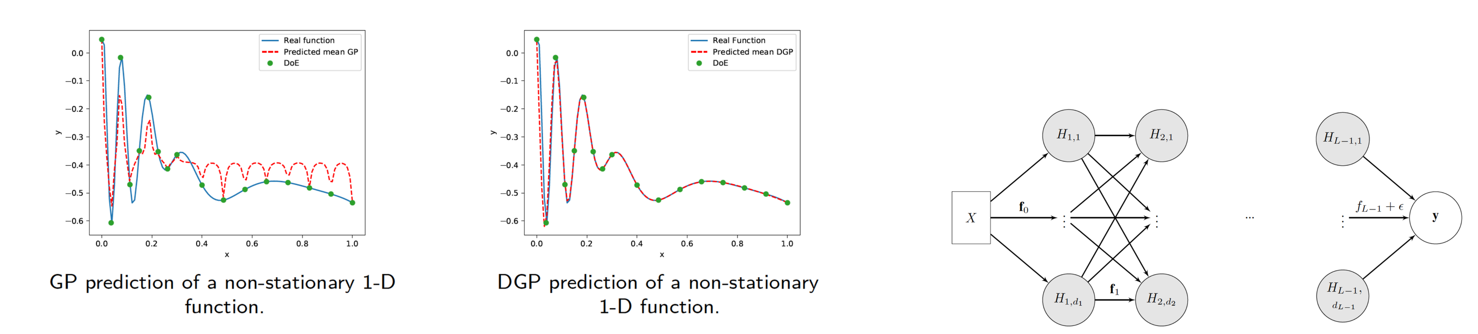
Liberté
Égalité
Fraternité

INRAE

## Metamodeling setup

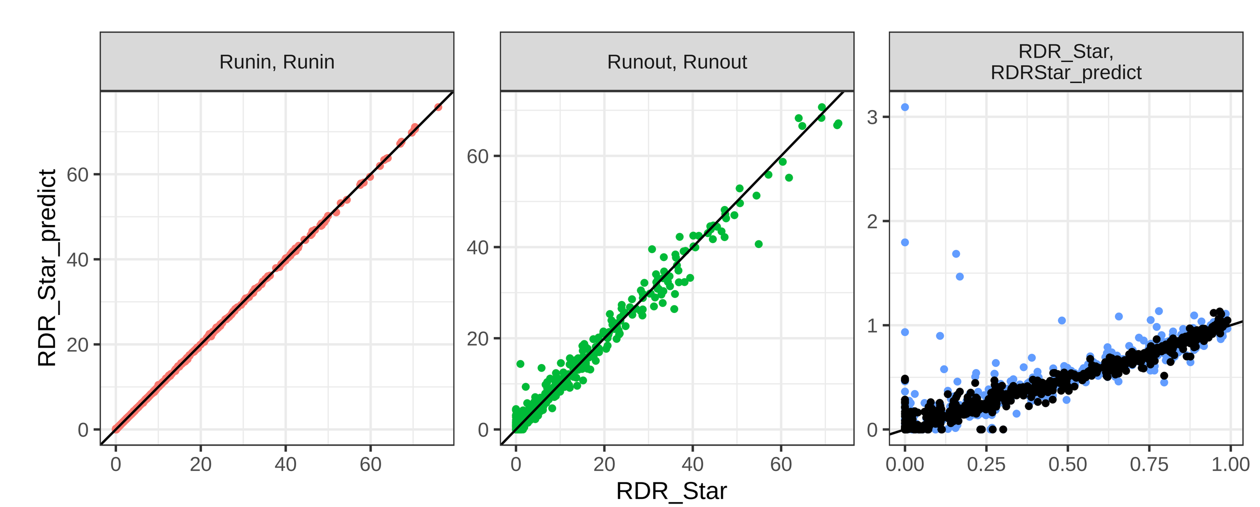



- How to deal with the categorical variables? adaptation of Gaussian processes[3] and Polynomial Chaos Expansion [5]
- How to deal with the plateau of null observations? DeepGP[4], for each couple of modalities



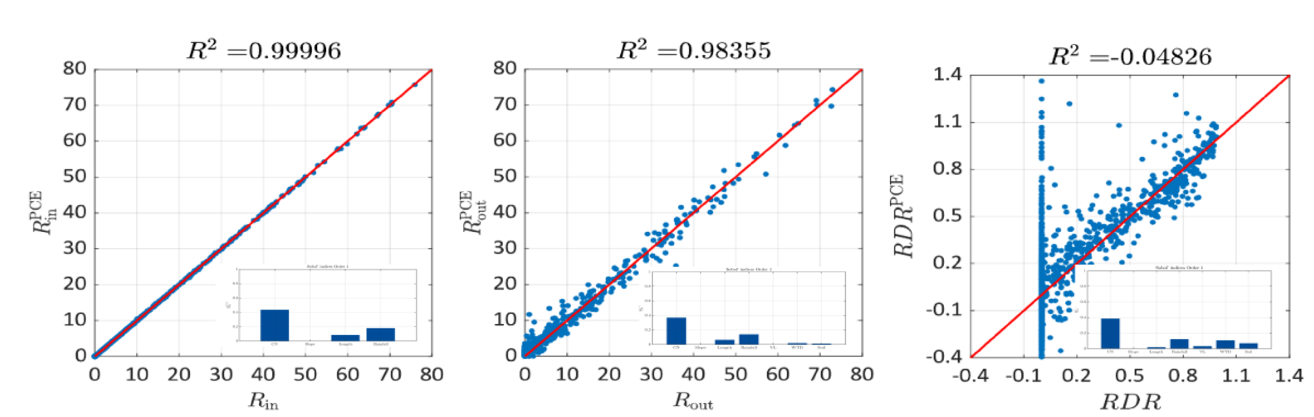

Extracted from PhD defense of Ali Hebbal

## Results : Direct MM or indirect MM?




Kriging : $R^2 = 0.999$ $R^2 = 0.985$ $R^2 = 0.753$ ⇒ 0.96

⇒ Surrogate of the ratio is much more reliable



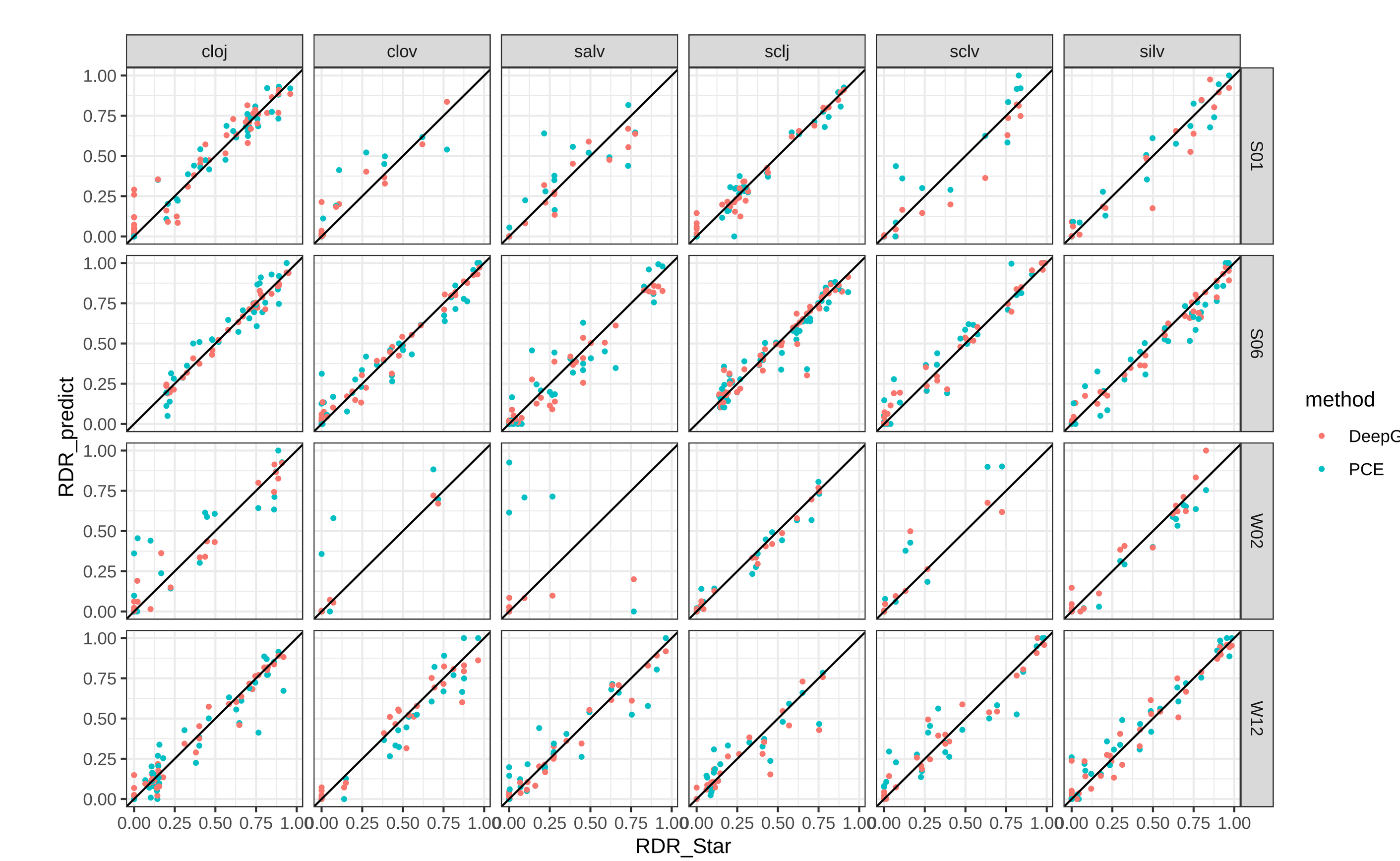

PCE: Direct, Runin Direct, Runout Indirect (ratio)

## Results: MM with classif / boundaries

| method (per cat.) | GP | PCE | DeepGP |
|---|---|---|---|
| no classif | $R^2 = 0.951$ | $R^2 = 0.903$ | $R^2 = 0.964$ |
| classif | $R^2 = 0.955$ | $R^2 = 0.911$ | $R^2 = 0.964$ |

⇒ DeepGP does not need any classification or boundaries

## Results : mixed variables or MM by category?




DeepGP and PCE MM by couple of category (Soil type x Rain type)

⇒ Both methods are in trouble with soils with a predominance of zeros

| Method | $R^2$ per category | $R^2$ for mixed var. |
|---|---|---|
| PCE | 0.916 | 0.966 |
| Kriging | 0.955 | 0.964 |
| DeepGP | 0.964 | - |

- Methods for mixed variables are more efficient and robust
- DeepGP performs well except for the worst soils, and is costly

## Conclusion

- Categorical var. were properly included into GP and PCE (+ Sobol)
- Mixed variables methods outperform the MM by category
- Classification does not improve the surrogate
- Good quality of prediction (96 % of variance is explained)

⇒ Next step : DeepGP for categorical variables