France-wide application of the RAINSIM rainfield stochastic generator

Emmanuel PAQUET (EDF-DTG)

SWGEN 2025

Conference on Stochastic Weather Generator

Grenoble, 2-4 December 2025

Outline

1. INTRODUCTION

Context

Rainfield stochastic simulation for hydrological applications (focus on extreme floods, cf. P. Guillemin's talk)
Coupled with temperature simulation (cf. J. Blanchet's talk), provides input data for distributed rainfall-runoff model

The RAINSIM generator has been developed on hydrological domains where the stationarity of the spatial dependence is a reasonable hypothesis (<5 000 km² ?)
→ Can be questioned for bigger domains

This presentation

Application of RAINSIM to whole France (500 000 km²)
At this scale, the variety of climates and precipitation patterns is huge
→ the hypothesis of stationarity of the spatial dependence is questionable

Two models and estimation methods of the spatial dependence are assessed here:

- stationary anisotropic power-exponential (as implemented in RAINSIM)
- non-stationary power-exponential with spatial deformation (using the *deform* R-package)

Focus on areal rainfall over big catchments and inter-catchment correlation of intense events

2. METHODS

RAINSIM model framework *(Vaittinada-Ayar et al., 2020)*

Rain-field stochastic generator, simulating daily fields with either multi-sites or field modes

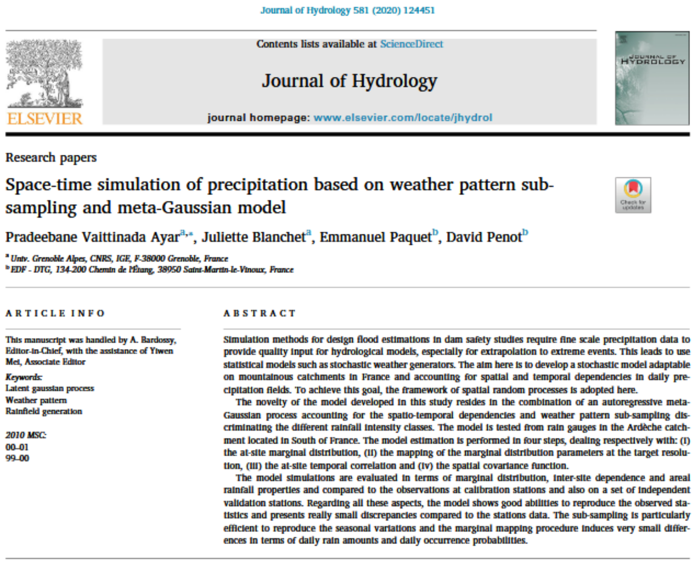
Combines a Weather Pattern sub-sampling and a meta-Gaussian process:

- Up to 12 subsets for parametrization and simulation : 4 Seasons x 3 Weather Types.
- Meta-Gaussian process: $Y(x)$ a r.v. of the precipitation for all x over a given domain $\mathcal{D}$.

Rain-field $Y(x)$ is obtained from a model based on a transformed latent censored Gaussian field

Model components conditionned by subsets have to be estimated:

1. At-site marginal distribution
2. Mapping of marginal distribution parameters on the whole domain
3. At-site temporal correlation
4. Spatial covariance function


Journal of Hydrology 581 (2020) 124451

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

ELSEVIER

Research papers

Space-time simulation of precipitation based on weather pattern sub-sampling and meta-Gaussian model

Pradeebane Vaittinada Ayar[a,*], Juliette Blanchet[a], Emmanuel Paquet[b], David Penot[b]

[a] Univ. Grenoble Alpes, CNRS, IGE, F-38000 Grenoble, France
[b] EDF - DTG, 134-200 Chemin de l'Étang, 38950 Saint-Martin-le-Vinoux, France

ARTICLE INFO

This manuscript was handled by A. Bardossy, Editor-in-Chief, with the assistance of Yiwen Mei, Associate Editor

Keywords:
Latent gaussian process
Weather pattern
Rainfield generation

2010 MSC:
00-01
99-00

ABSTRACT

Simulation methods for design flood estimations in dam safety studies require fine scale precipitation data to provide quality input for hydrological models, especially for extrapolation to extreme events. This leads to use statistical models such as stochastic weather generators. The aim here is to develop a stochastic model adaptable on mountainous catchments in France and accounting for spatial and temporal dependencies in daily precipitation fields. To achieve this goal, the framework of spatial random processes is adopted here.

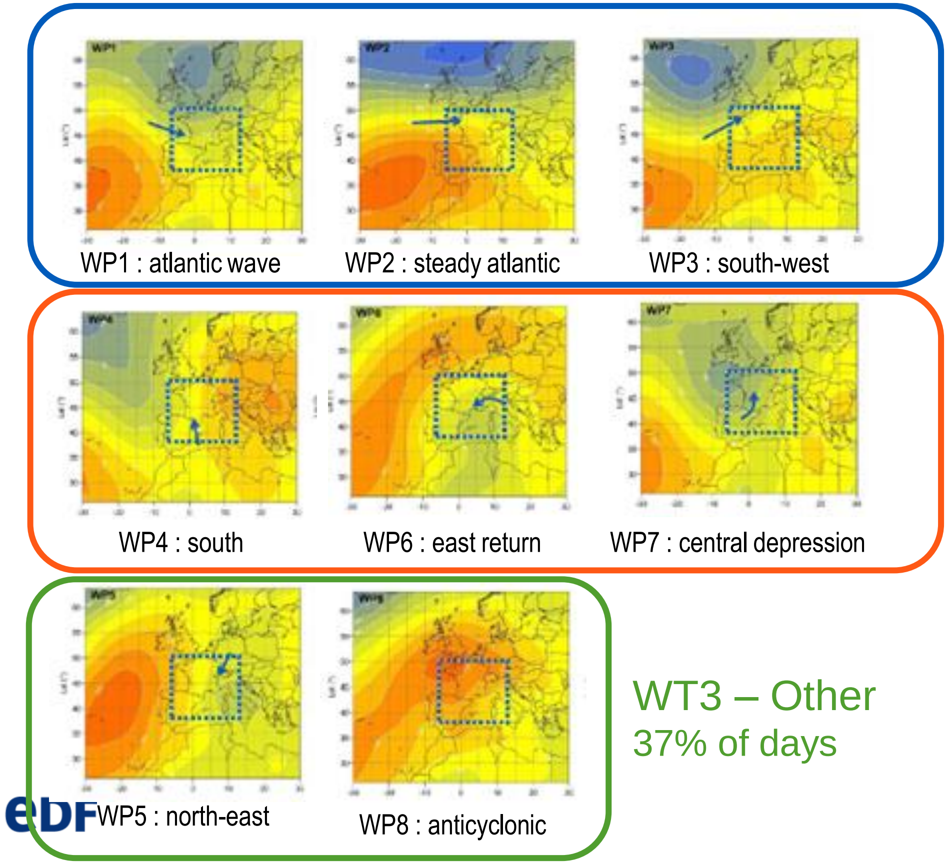
The novelty of the model developed in this study resides in the combination of an autoregressive meta-Gaussian process accounting for the spatio-temporal dependencies and weather pattern sub-sampling discriminating the different rainfall intensity classes. The model is tested from rain gauges in the Ardèche catchment located in South of France. The model estimation is performed in four steps, dealing respectively with: (i) the at-site marginal distribution, (ii) the mapping of the marginal distribution parameters at the target resolution, (iii) the at-site temporal correlation and (iv) the spatial covariance function.

The model simulations are evaluated in terms of marginal distribution, inter-site dependence and areal rainfall properties and compared to the observations at calibration stations and also on a set of independent validation stations. Regarding all these aspects, the model shows good abilities to reproduce the observed statistics and presents really small discrepancies compared to the stations data. The sub-sampling is particularly efficient to reproduce the seasonal variations and the marginal mapping procedure induces very small differences in terms of daily rain amounts and daily occurrence probabilities.

2. METHODS

Subset of data for parametrization and simulation = season + weather type

Weather types (WT) are a grouping of 8 daily Weather Patterns (WP) build for France by EDF (precipitation oriented)
Allows processing more homogeneous samples, linked to seasonal & meteorological processes

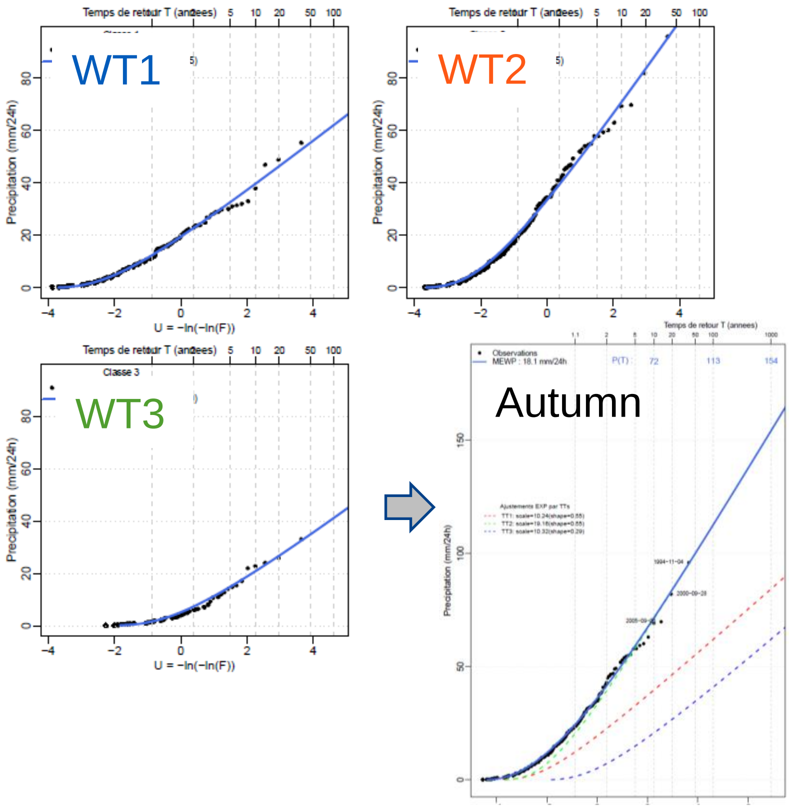


WT1- Atlantic
38% of days

WT2 – Mediterranean
25% of days

WT3 – Other
37% of days

Example :
Autumn distributions of daily rainfall at Mende



RAINSIM model components and estimation (1)

1. Marginal model *(Blanchet et al., 2019)*

- At site distribution $F_{Y(x)}$: gamma distribution fitted for each sub-set j by probability weighted moments or likelihood maximization
- Parameters mapping by bivariate thin plate spline (and altitude drift) estimated by a penalised least square method

2. Spatio-temporal model

Let be $H(x) = \Phi^{-1}(F_{Y(x)}\big(Y(x)\big)$, where Φ is the CDF of the normal $\mathcal{N}(0,1)$.

Since $Y(x) > 0 \Longrightarrow H(x) > T(x) \equiv \Phi^{-1}(F_{Y(x)}(0))$ then $H(x) = max(T(x), G(x))$

$\Longrightarrow G(\vec{x})$: latent autoregressive Gaussian process such as: $G_t(\vec{x})$ = $A_{t,t-1}G_{t-1}(\vec{x}) + E_t(\vec{x})$

$G_t(\vec{x})$ of marginal $\mathcal{N}(0,1)$ and spatial covariance Σ_t and $A_{t,t-1}$ a $\mathrm{N} \times N$ matrix of autoregressive parameters assumed to be diagonal and $E_t(\vec{x})$ has zero mean and covariance $\Sigma_t - A^2_{t,t-1}\Sigma_{t-1}$

RAINSIM model components and estimation (2)

3. Spatial dependence

Spatial covariance function : power-exponential $C_j(h) = exp\left(-\left(\frac{h}{\lambda_j}\right)^{\nu_j}\right)$

Anisotropy (with Mahalanobis distance) models a preferential dependence axis

For all t in sub-set j, $\Sigma_t = C_j(h)$ is estimated by censored likelihood

4. Temporal dependence

All the diagonal coefficient of $A_{t,t-1}$ are the at-site autocorrelation estimated by composite likelihood

for all transition from a WT to another.

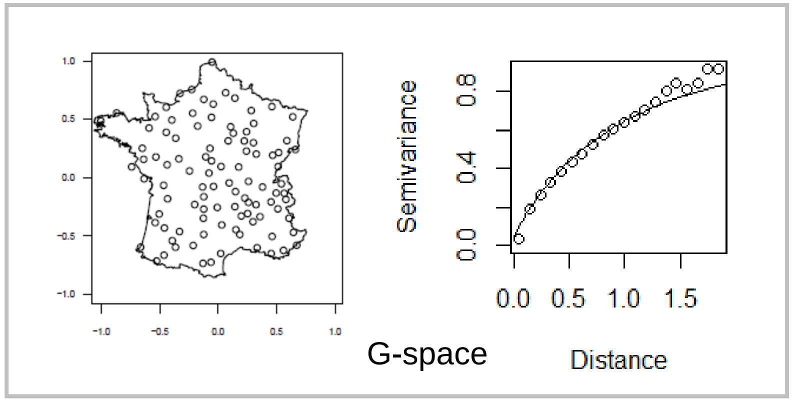
Spatial dependence model with space deformation

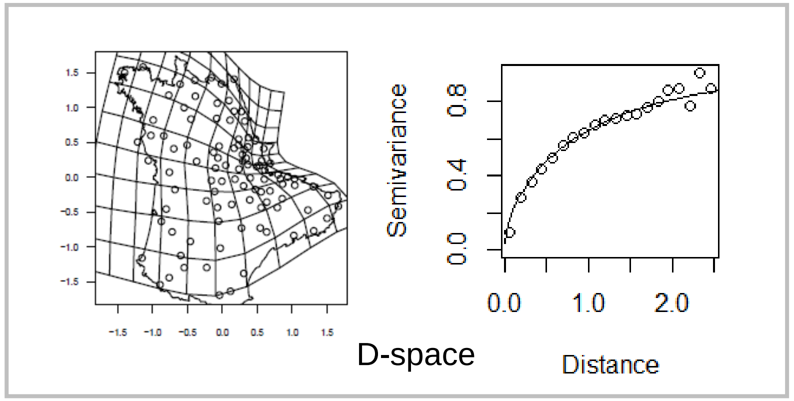
Introduced by Sampson & Guttorp (1992), with multiple developments and publications since then:

Transforms geographic space (G-space) in a dispersion space (D-space) where stationarity and isotropy of the spatial dependence model hold

Implemented in the *deform* R-package by Youngman (2023):

- Uses thin plate regression splines for topologic transformations
- Handles bijectivity constraints to avoid folding in deformations
- Powered exponential covariance function $C(h) = \kappa\, exp(-h^{\gamma})$
- Estimation of parameters (deformation, smoothing, covariance) on censored data (allowing focus on intense values)





deform: An R Package for Nonstationary Spatial Gaussian Process Models by Deformations and Dimension Expansion

Benjamin D. Youngman
Department of Mathematics and Statistics
University of Exeter

Abstract

Gaussian processes (GP) are a popular and powerful tool for spatial modeling of data, especially data that quantify environmental processes. However, in stationary form, whether covariance is isotropic or anisotropic, GPs may lack the flexibility to capture dependence across a continuous spatial process, especially across a large domain. The **deform** package aims to provide users with user-friendly R functions for the fitting and visualization of nonstationary spatial GPs. Users can choose to capture nonstationarity with either the spatial deformation approach of Sampson and Guttorp (1992) or the dimension expansion approach of Bornn, Shaddick, and Zidek (2012). Thin plate regression splines are used for both approaches to bring transformations of locations to give a new set of locations that bring isotropic covariance. Fitted models in **deform** can be used to predict these new locations and to simulate nonstationary GPs for an arbitrary set of locations.

Keywords: Gaussian process, geostatistics, spatial deformation, dimension expansion, thin plate regression spline, R.

Scoring areal rainfall (ARR) simulations (1)

Needed to complement at-site scoring of simulations for hydrological applications

ARR values are computed from multi-sites simulations with Thiessen weighting

$$\{P^{int}\} = nint\ integer\ values\ in\ the\ obs\ range$$

$$score\ U_j = \frac{1}{nint} \sum_{i=1}^{n} \left| U^*\left(P_i^{int}\right) - U\left(P_i^{int}\right) \right| \quad j \in [1, nsim]$$

$$where\ U = -log(-\log(f))$$

$$score\ U = median\left(score\ U_j\right)_{j=1,nsim}$$

$$P_0 = dry\ day\ frequency$$

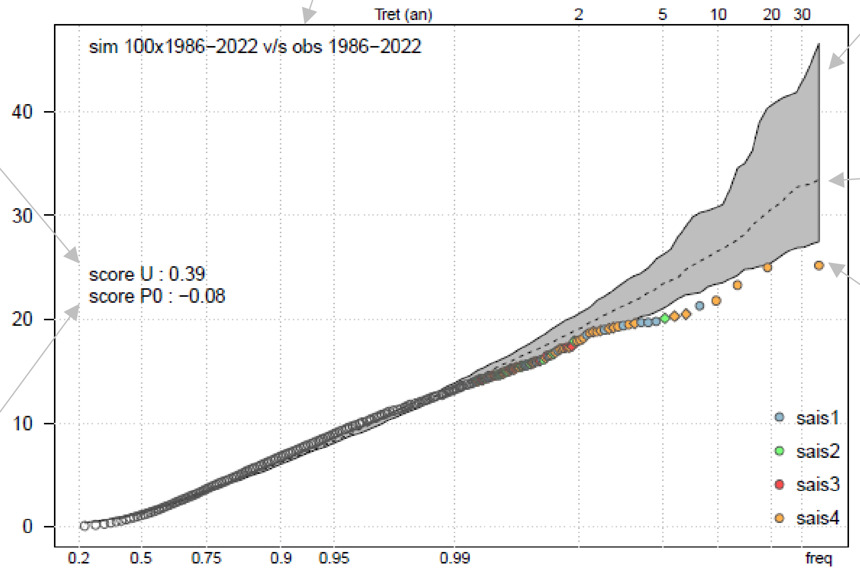
$$score\ P_0 = median\ (P_0^*[j] - P_0)_{j=1,nsim}$$

100 alternative realizations of the 1986-2022 period
≈ 3700 years

90% CI of simulated quantiles

median of simulated quantiles

observed 1986-2002 values
colored by subset



Distribution of daily simulated ARR

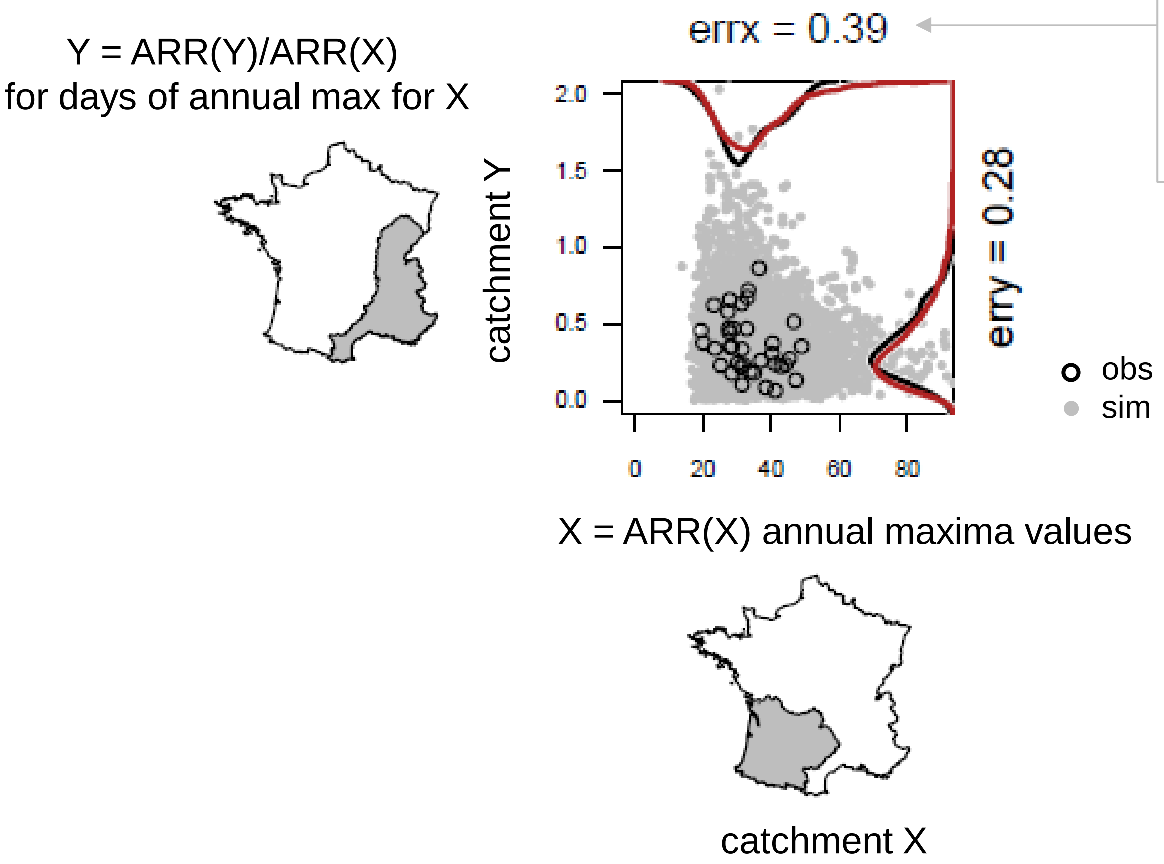
Scoring areal rainfall (ARR) simulations (2)

ARR & catchment-to-catchment correlation of intense events

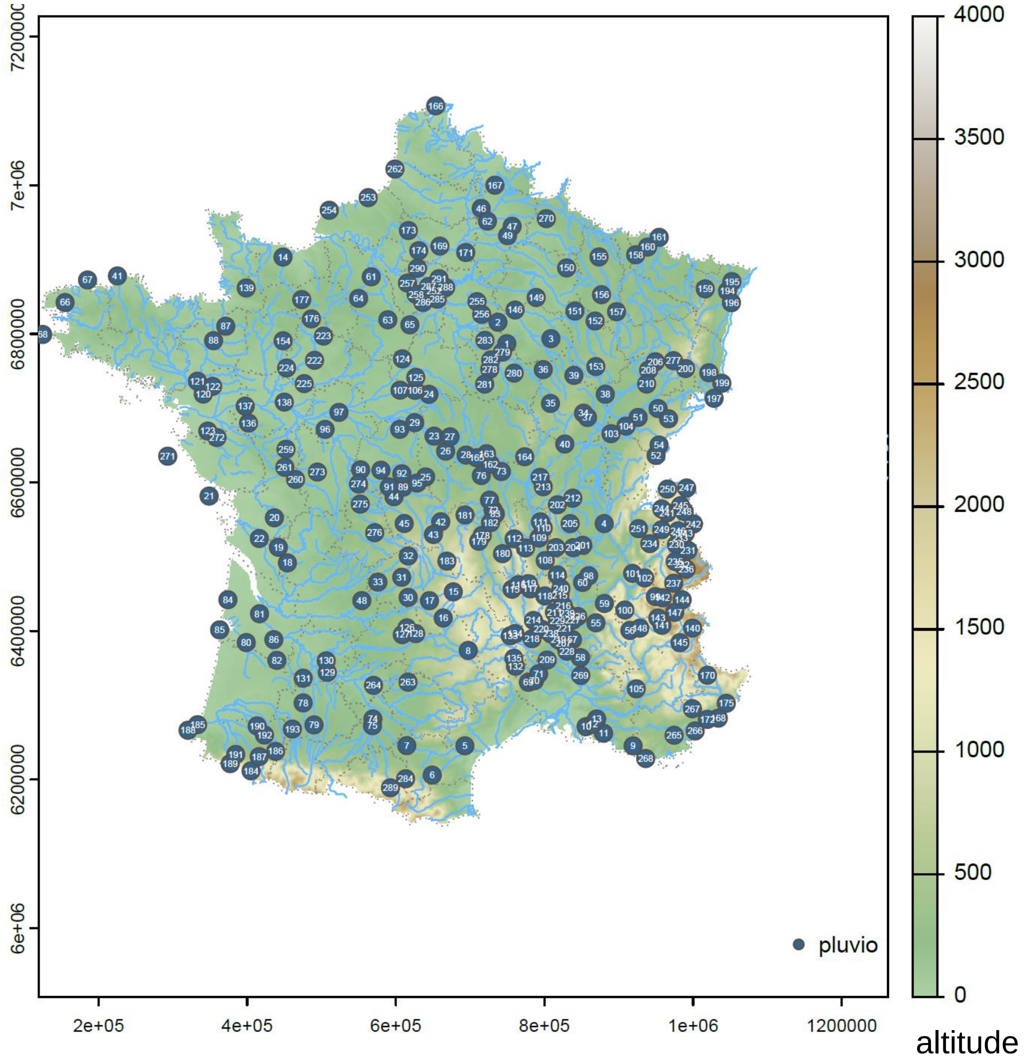
$$err(X) = \sum_{i=1,10} |f(x \in bin_i) - f(x^* \in bin_i)|$$

$$f(x \in bin_i) = relative\ frequency\ of\ values\ in\ bin_i$$

$$CI(90\%) \approx [0.10; 0.40]$$

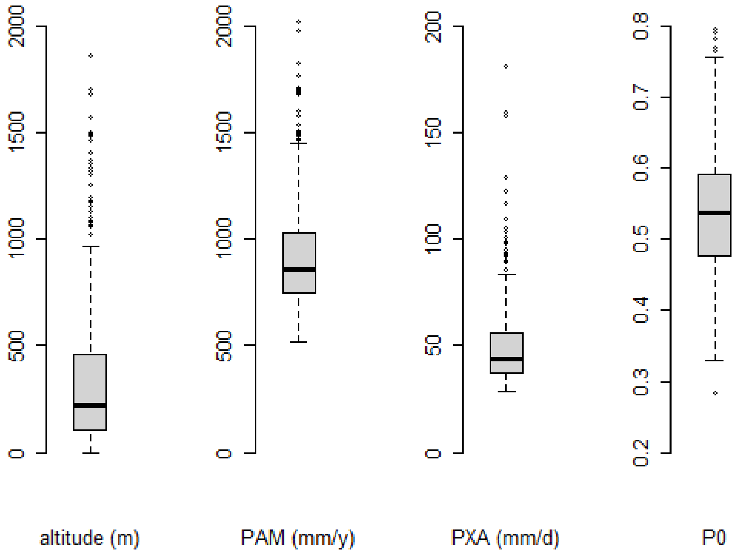


A quality-checked dataset of daily precipitation



SQR reference rainfall series from Météo-France
→ high homogeneity, good spatial coverage

291 stations covering all France
Study domain ≈ 500 000 km²
Mostly complete on 1950-2022

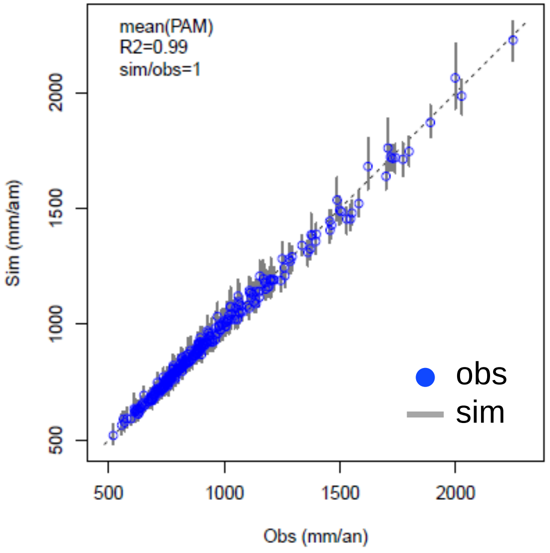


3. RESULTS

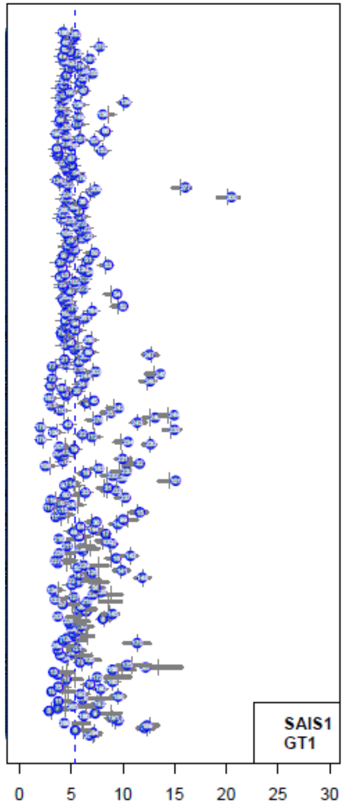
At-site simulations scoring (marginals)

Parametrizations of RAINSIM on the 1986-2022 period
Multi-sites simulation
Sim. of 100 alternative realizations of 1986-2022 (≈3700 years)

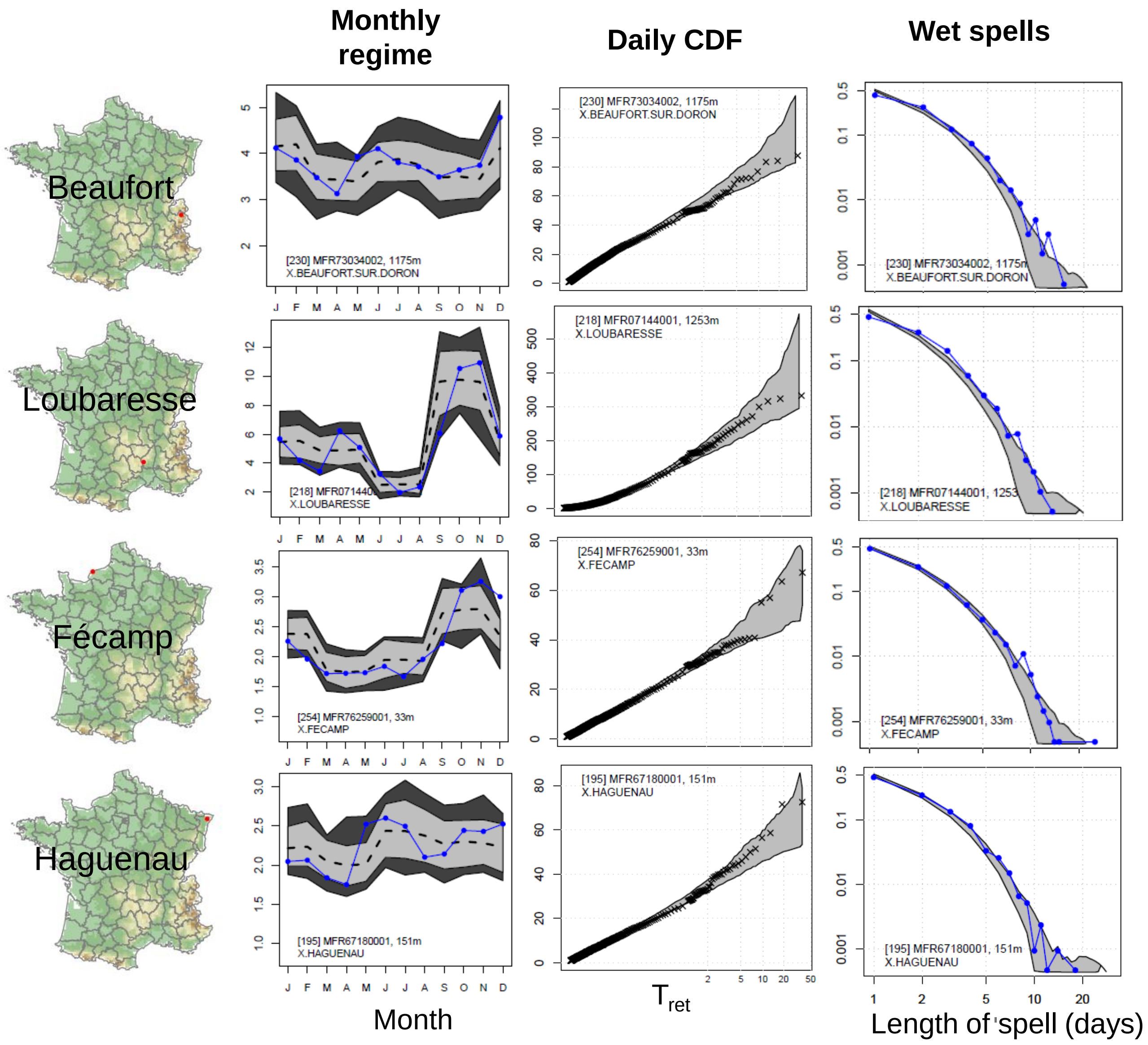
Scatter plot of PAM ↓



Mean of P>0
JFM season / WT1 →



stations
North
↓
South



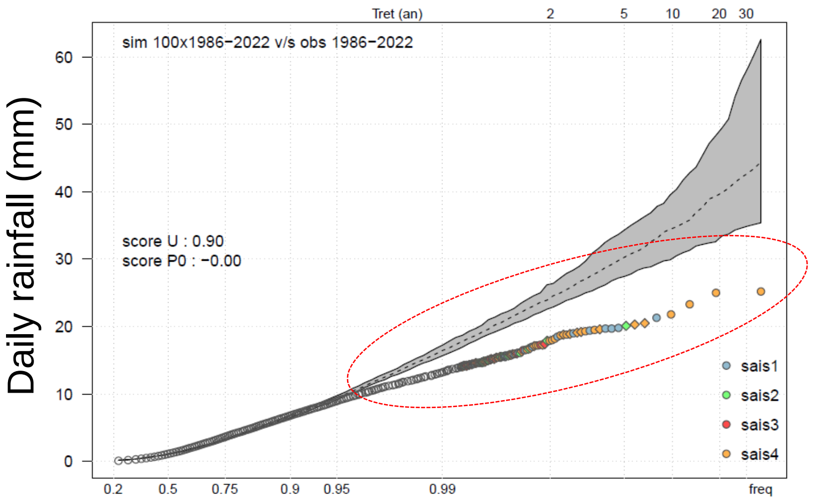
Original spatial dependence model: ARR scoring (1)

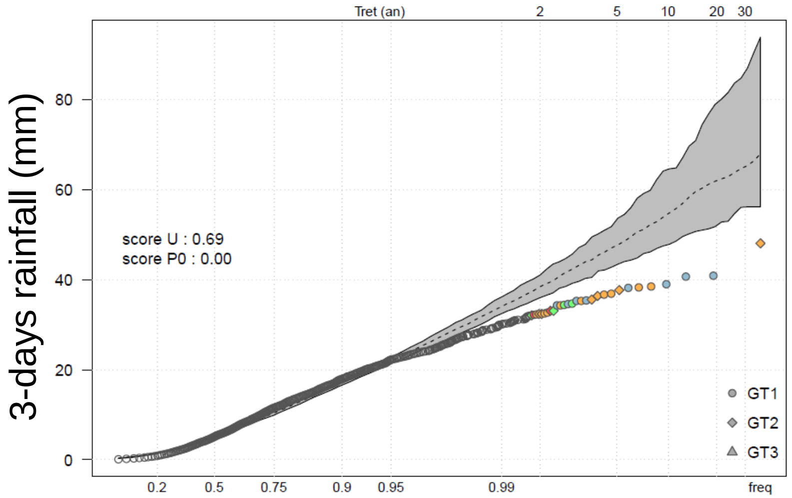
RAINSIM "original" model:
anisotropic power exponential function
+ censored likelihood optimization

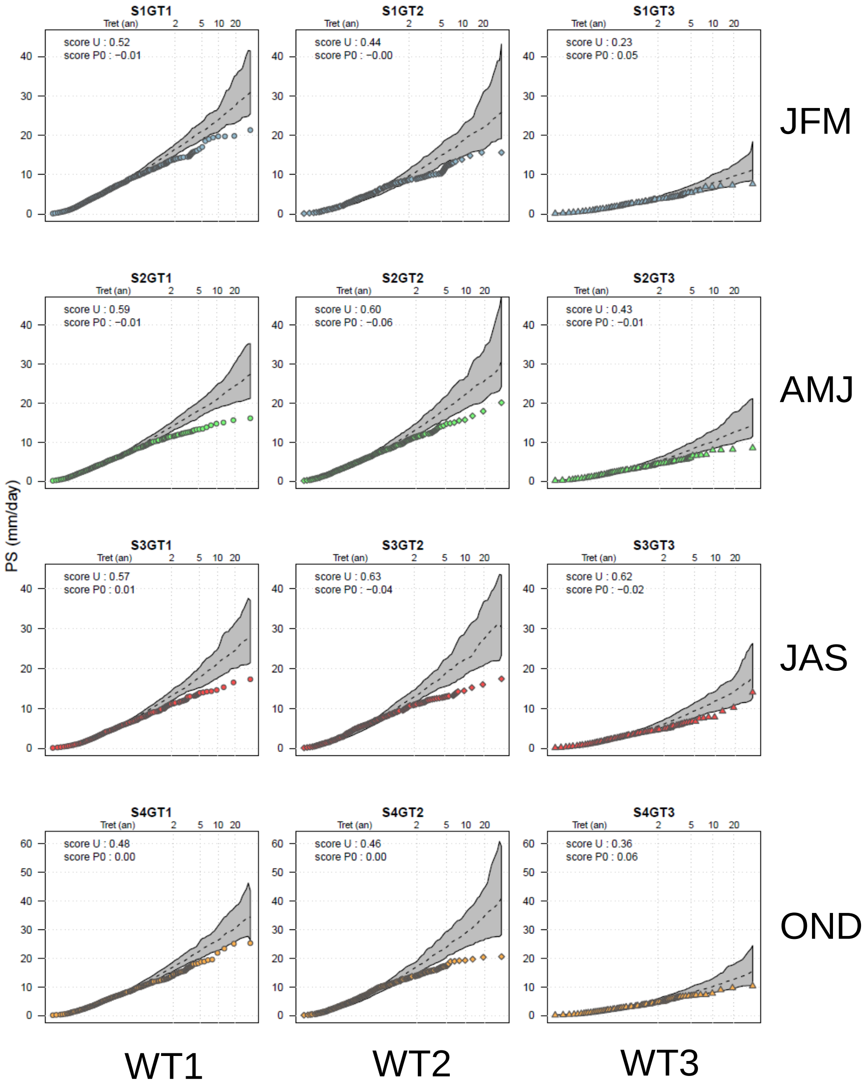
Scoring of global ARR for France

→ Overestimation of high quantiles
especially for WT1 and WT2

→ Very few biases for P0







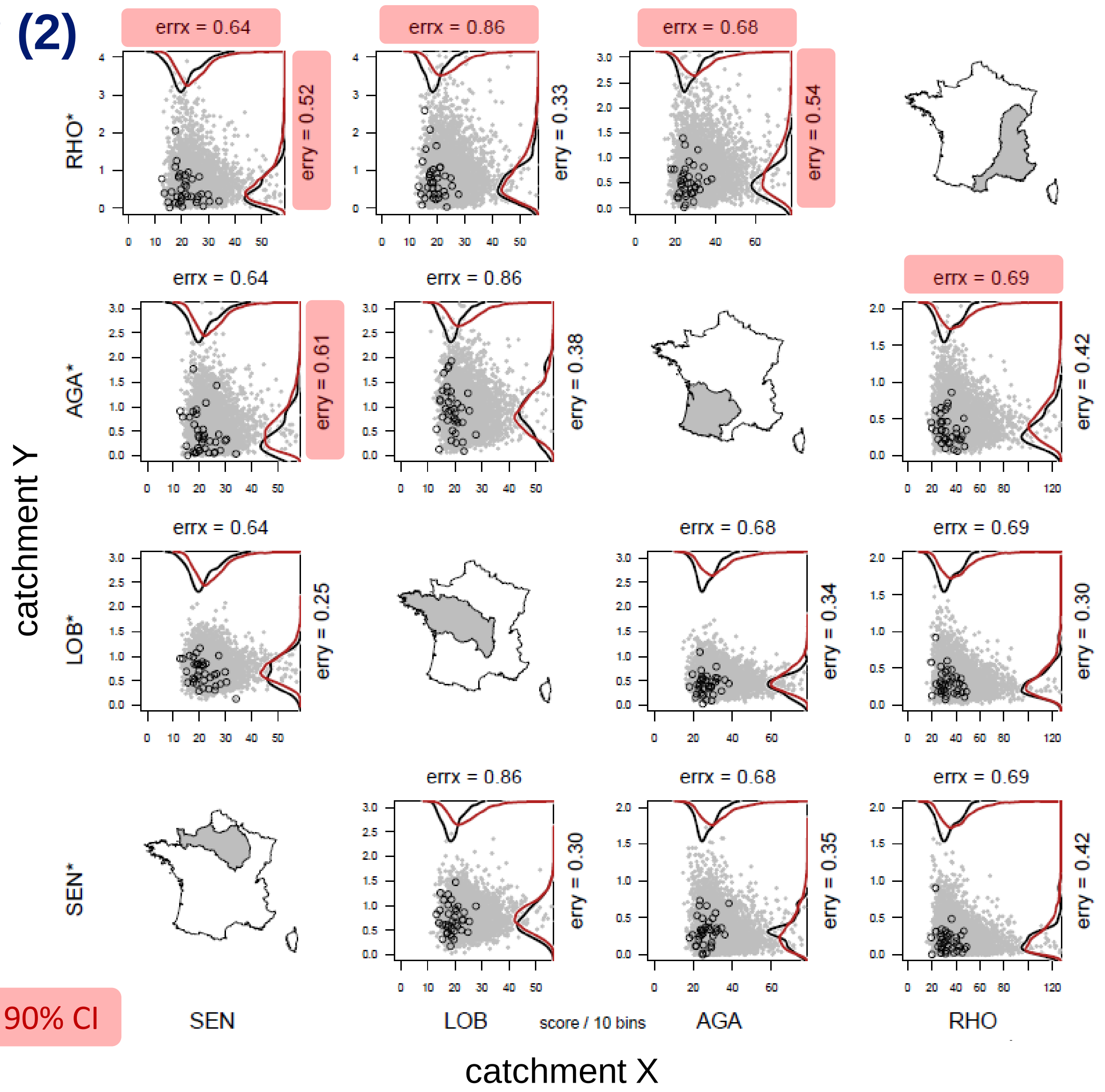
Original spatial dependence model: ARR scoring (2)

Scoring of catchment-to-catchment correlation

SEN = Seine-Normandie (95 000 km²)
LOB = Loire-Bretagne (157 000 km²)
AGA = Adour-Garonne (118 000 km²)
RHO = Rhone-Méditerrannée (122 000 km²)
→ in total, covering 89 % of the whole domain area

→ Overestimation of high ARR quantiles in all cases
(scores out of CI for all catchments)

→ Inter-catchment correlation mostly correct
(8 scores within CI out of 12 pairs, but close to upper value of CI)



3. RESULTS

deform spatial dependence model

Deformation models fitted for each subset

Computed on a sub sample of 100 stations

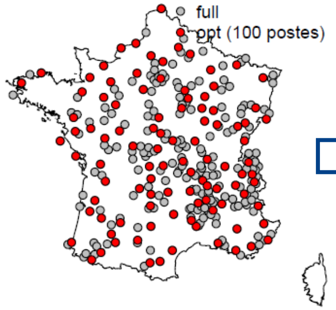
Gaussian at-sites values completed randomly below the local P0 quantile to run *deform*

Left-censoring set for each subset, from 0.75 to 0.95

→ WT1 et WT2 patterns coherent across seasons

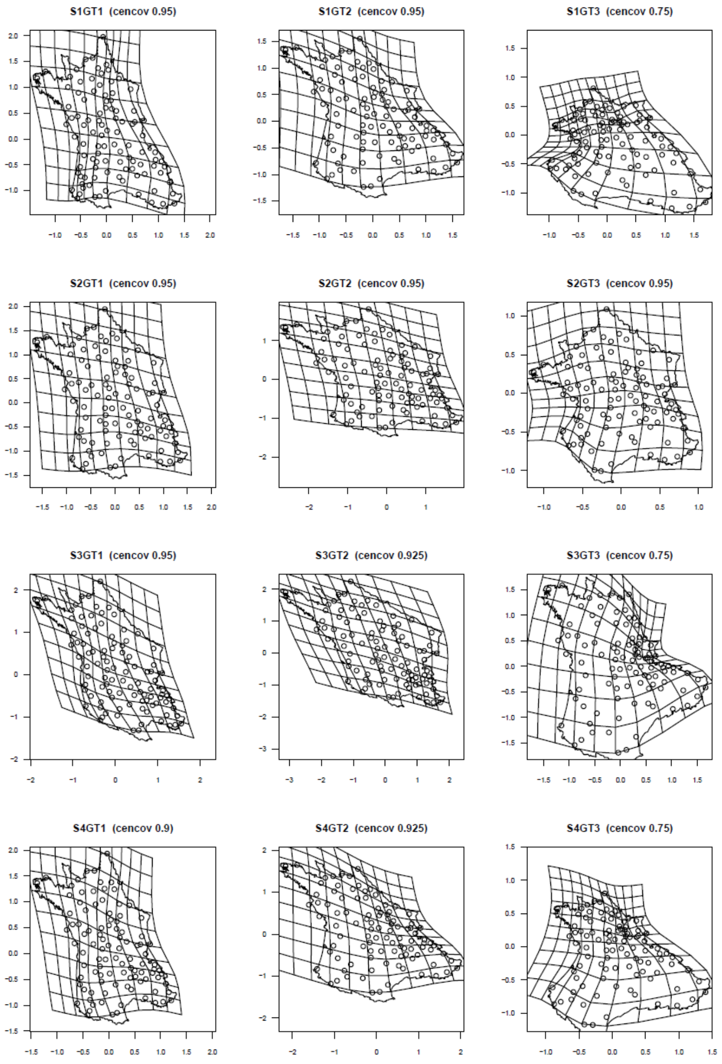
→ WT is the main influence over D space structure
WT1 (Atlantic) → N-S correlation dist > W-E
W2 (Medit.) → SE-NW correlation dist > SW-NE

→ WT3 deformations less coherent



deform
R-package

« D » space



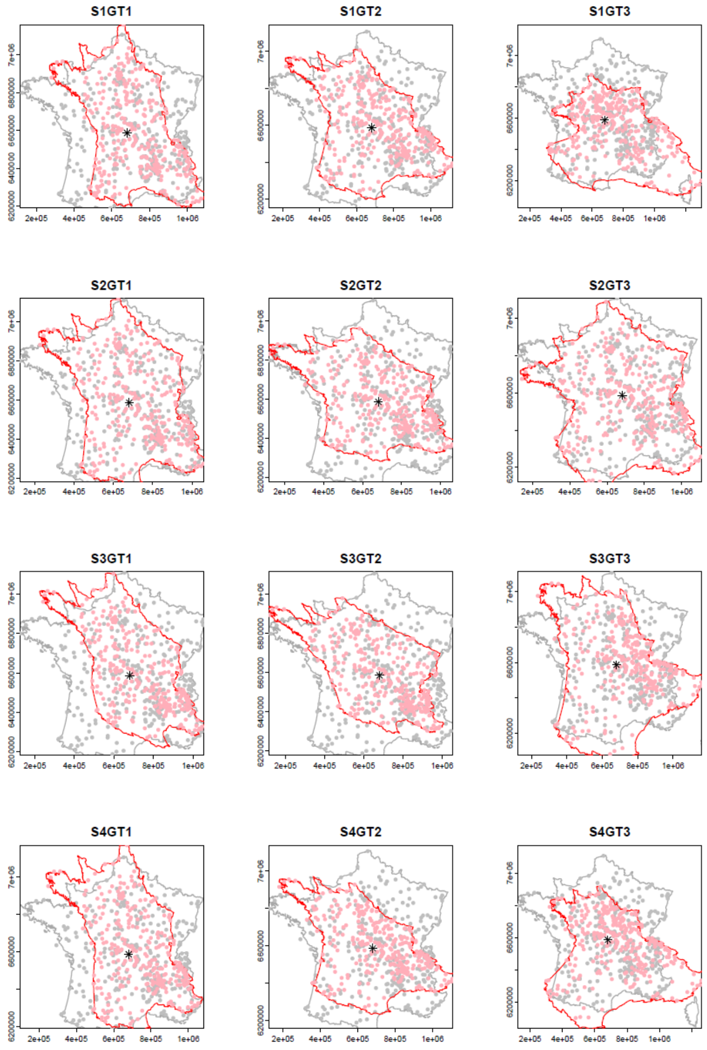
JFM

AMJ

JAS

OND

« D » space projected into geo. space



WT1 WT2 WT3 WT1 WT2 WT3

deform spatial dependence model: ARR scoring (1)

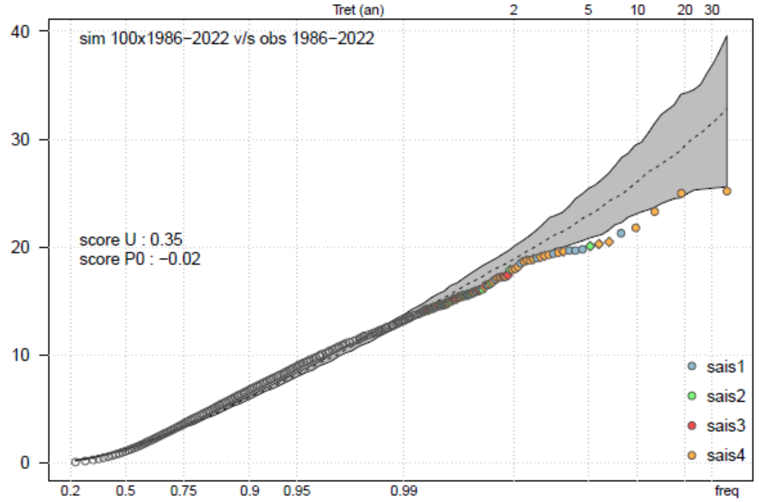
power exponential function #2
+ deformed correlation space
+ censored likelihood optimization

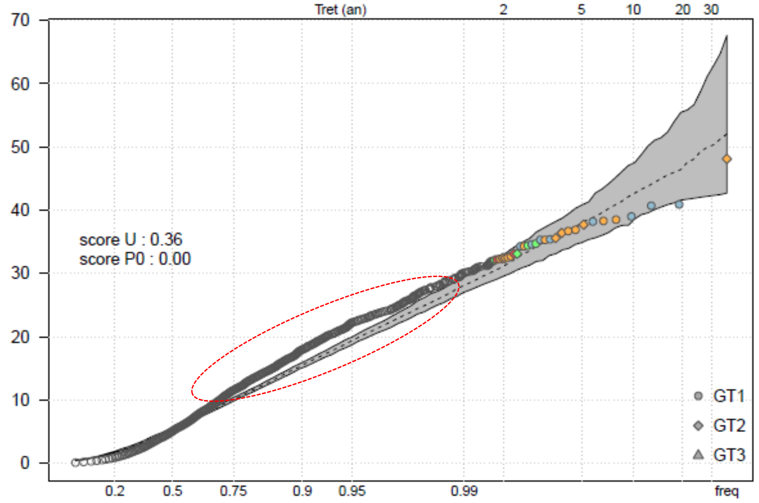
Scoring of global ARR for France

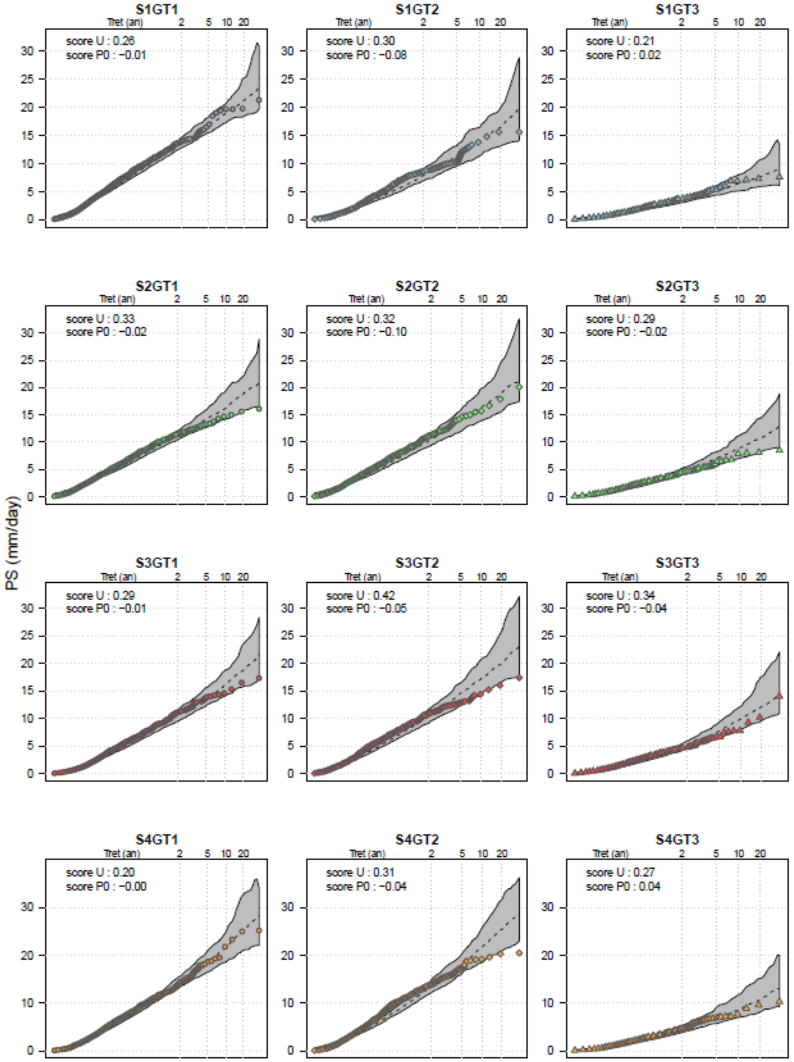
→ Score U greatly improved globally, and for almost all the subsets (≈ for S1WT3)

→ Light biases for some P0 scores

→ Bias on intermediate quantiles of 3-days amounts







3. RESULTS

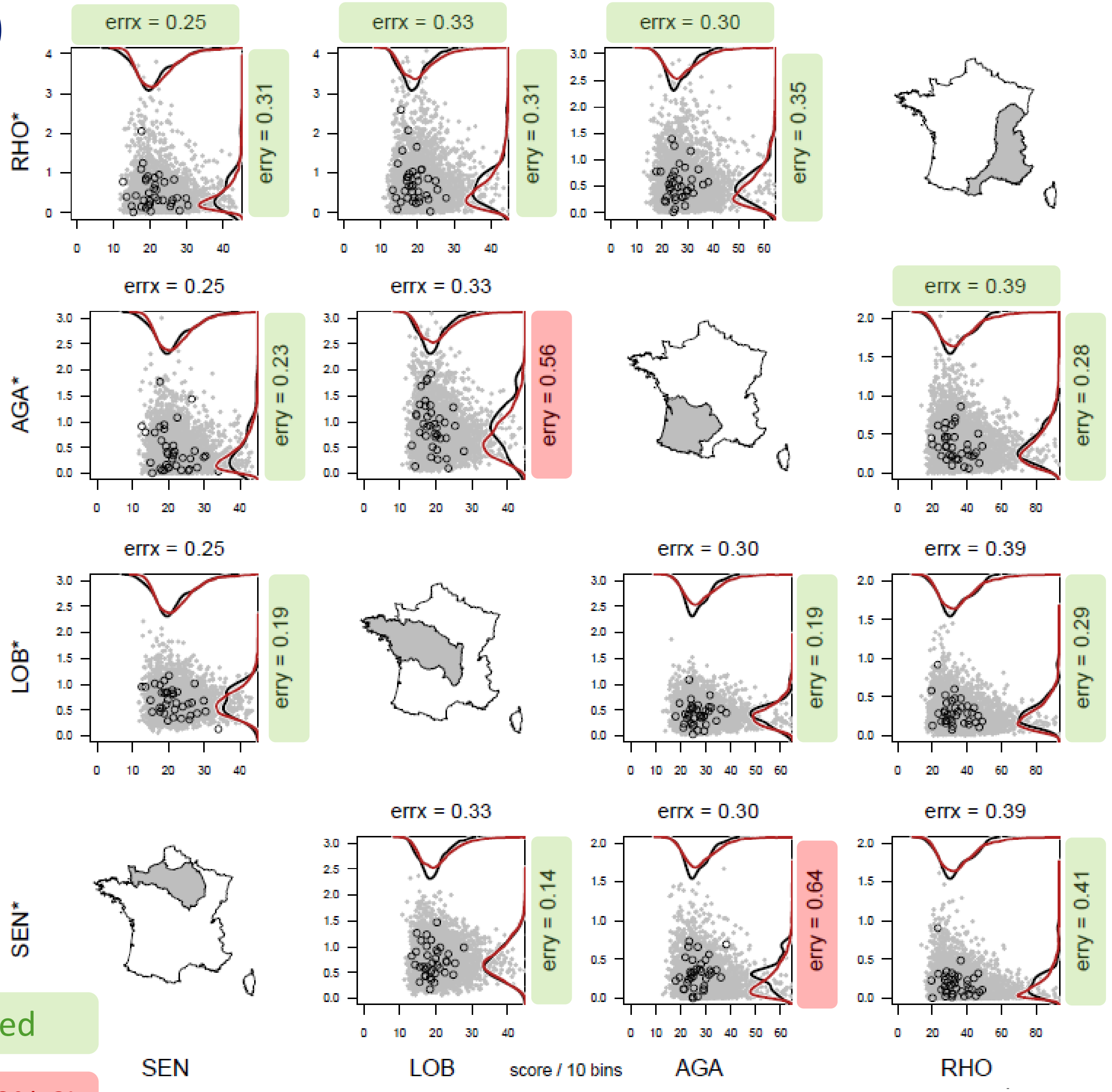
deform spatial dependence model: ARR scoring (2)

Scoring of catchment-to-catchment correlation

Compared to simulation with the original spatial model

→ All catchment ARR scores within CI and improved

→ Inter-catchment correlations globally improved (10 out of 12), but AGA v/s LOB and SEN v/s AGA now outside CI



score improved

score outside 90% CI

deform spatial dependence model: ARR scoring (3)

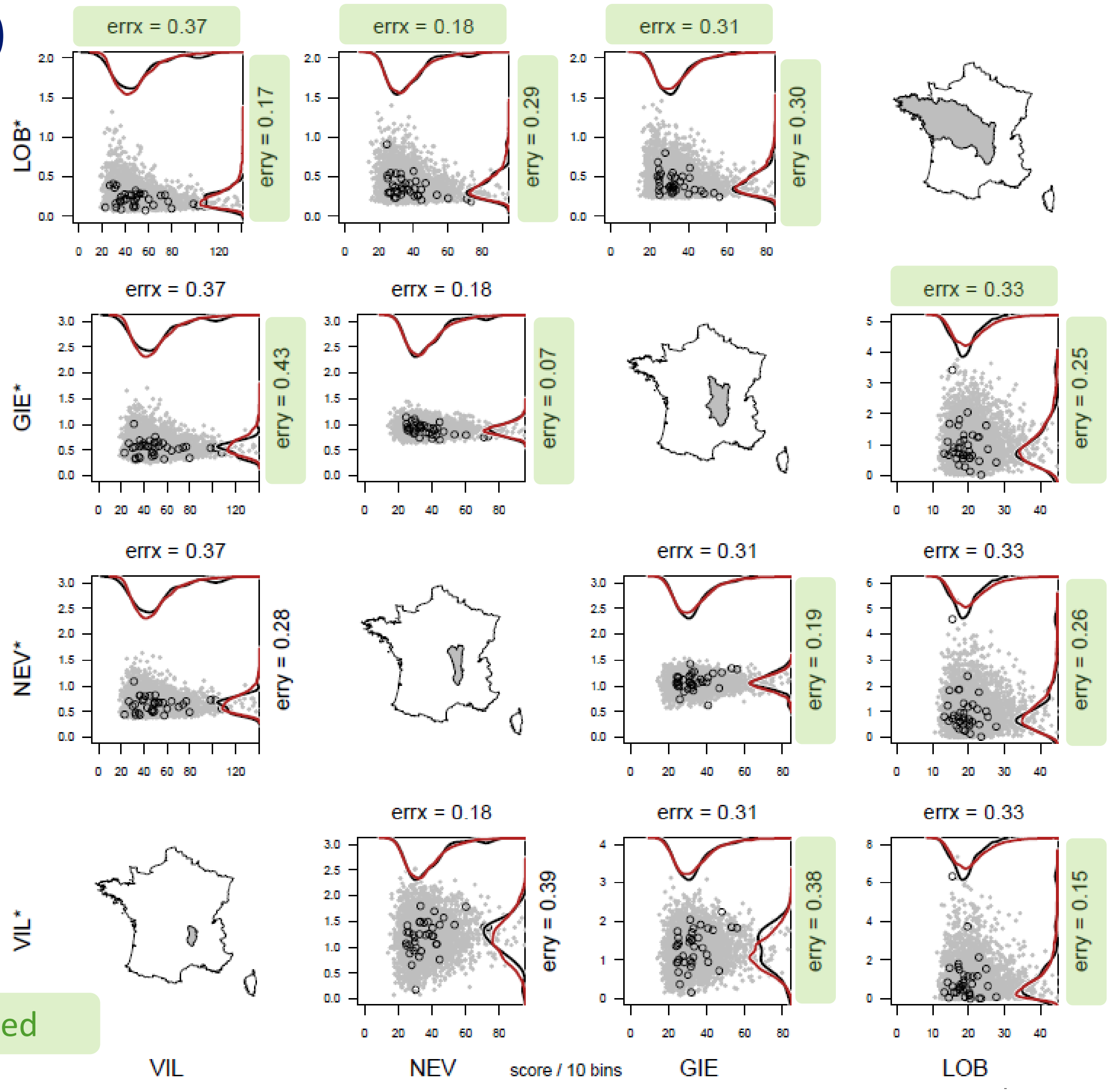
Scoring of catchment-to-catchment correlation

Nested sub-catchments of the LOB catchment (1 to 20 scale range):
VIL = Loire@Villerest (6 600 km²)
NEV = Loire@Nevers (17 600 km²)
GIE = Loire@Gien (36 000 km²)
LOB = Loire-Bretagne (157 000 km²)

Compared to simulation with the original spatial dependence model

→ All catchment ARR scores improved, and are within CI

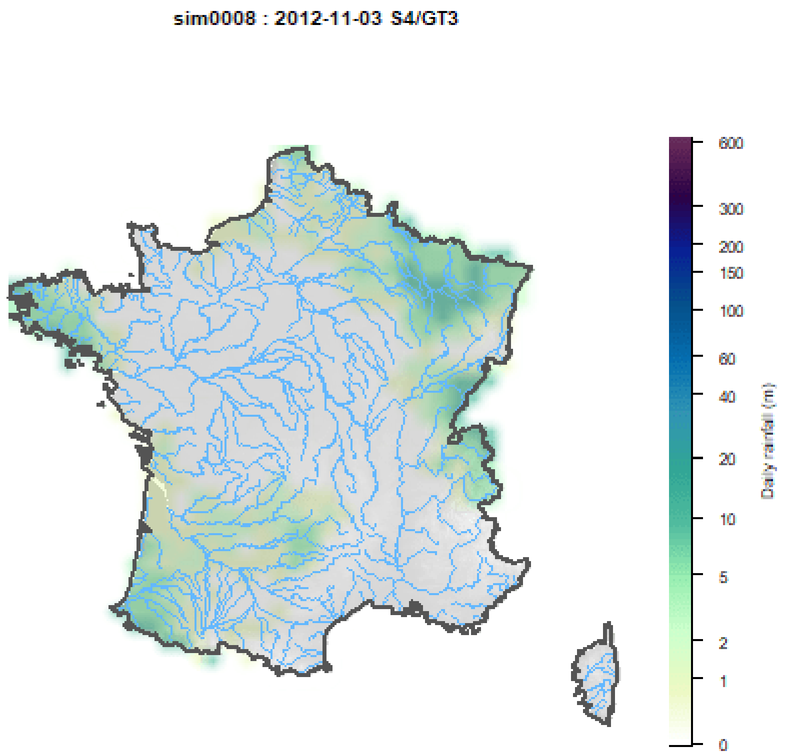
→ Inter-catchment correlations mostly improved (10 out of 12),
VIL v/s NEV not significantly different



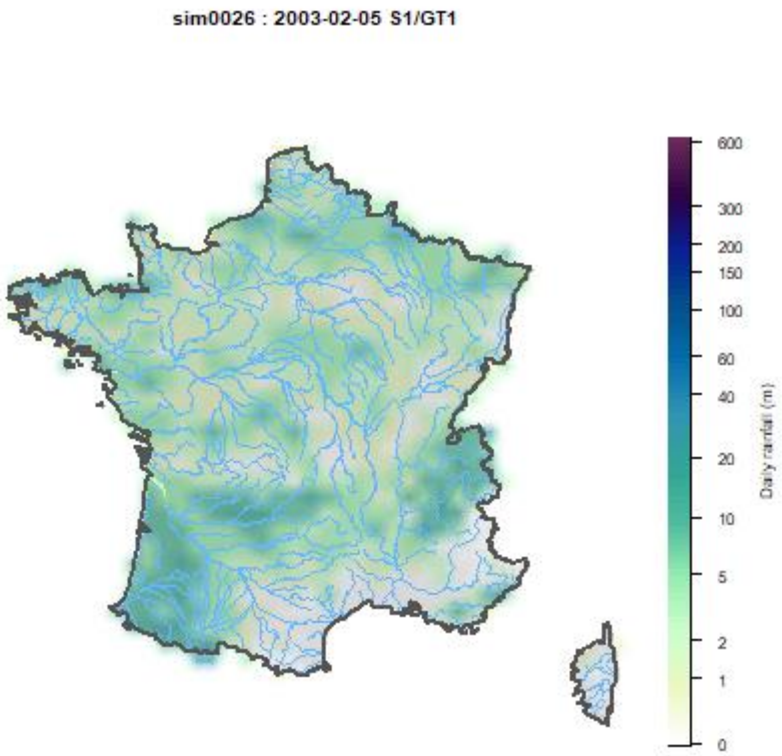
score improved

Examples of simulated rainfield sequences over France

Field simulation on a 30 x 30 km grid (791 pixels)
30-days sequences with comparable 30-days total amounts



Original RAINSIM spatial model



Deform spatial model

A final twist...

Using *deform* without space deformation

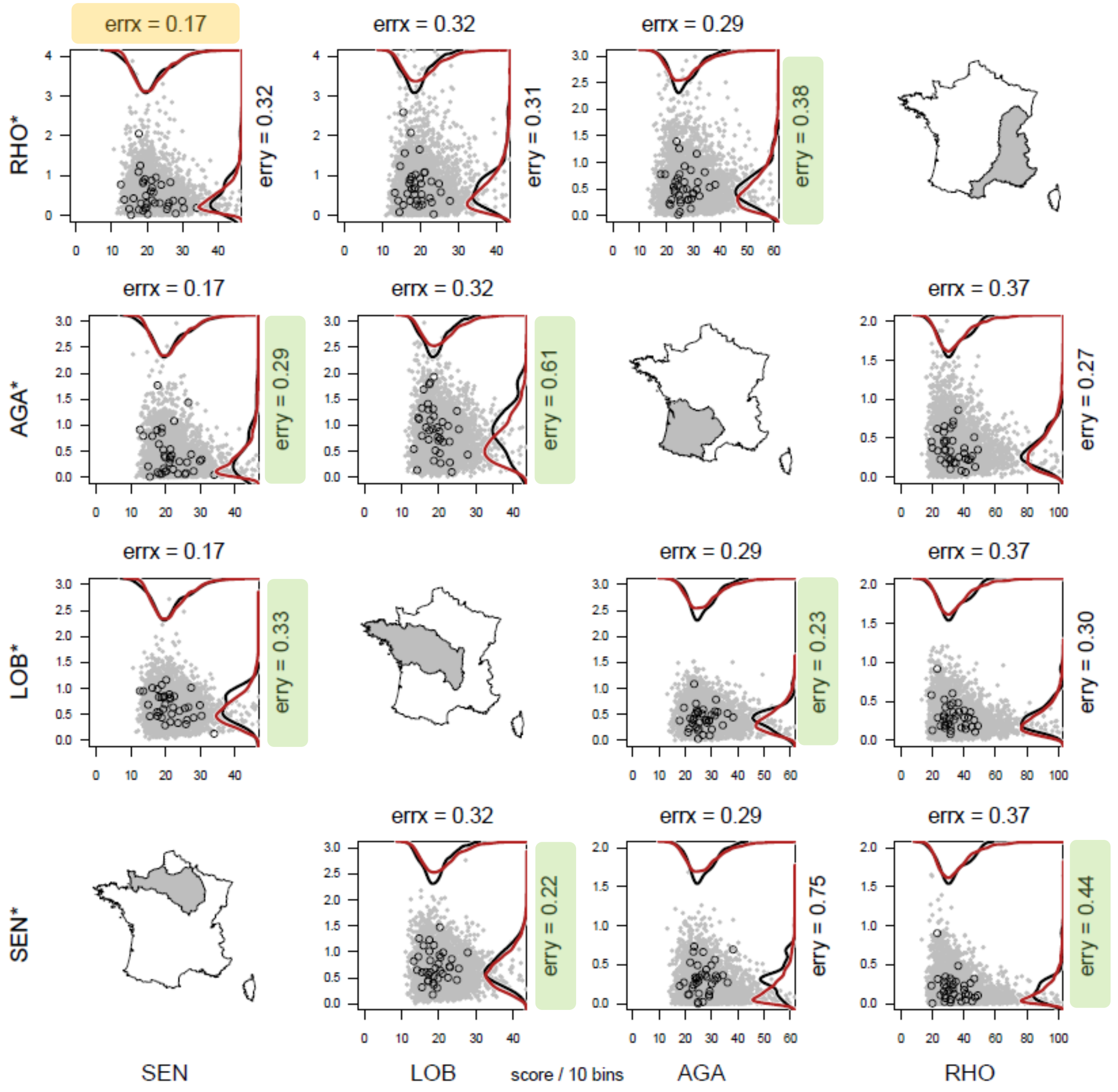
A simple anisotropic option is also available in the package:
Same covariance function and estimation of parameters,
Geographical space only scaled differently along X and Y

Compared to simulations with fully deformed space

→ France ARR scores are equivalent (not shown)

→ All but one catchment ARR scores are equivalent,
only RHO improved with simple anisotropy

→ Inter-catchment correlations lightly improved by deformation for 7 out of 12, others are equivalent



best score with deformation

best score with simple anisotropy

equivalent

Discussion

Introduction of suitable scorings for ARR on big catchments and inter-catchment correlations of intense events

The original dependence model of RAINSIM (supposed stationary) fails to capture some medium (areal rainfall on catchment) and large (whole domain and inter-catchment) scale correlations

The spatial deformation approach of Sampson & Guttorp (1992), implemented in the *deform* R-package by Youngman (2023), provides a convenient no-stationary spatial correlation model solving these issues

For France, the deformed space structures are consistent across seasons for WT1 and WT2
For WT1 and WT2, they remain almost the same being identified on a different period (1950-1984, not shown)

In this application, lot of the improvement of scores is due to the *deform* parameter estimation method,
the space deformation itself improves them to a lesser extent compared to a simple anisotropy
→ The parsimony principle would advice to actually deform the space only if needed (according scores)

Its use is flexible and numerically efficient (estimation time divided by more than 10)

Perspectives

Sensitivity study to the WT classification (different subset for correlation spaces, better scores ?)

Testing a spatial dependence estimated on COMEPHORE reanalysis

Building a France-wide modelling framework:

- Setting of a TEMPSIM generator for temperature simulation
- Running complete hydrological simulations covering most of the domain (e.g. using the EXPLORE2 hyd. models)
- Assessing the range of scales for which a global modelling is relevant for flood estimation by stochastic simulation

Driving the marginal distributions with climate indices to model climate non-stationarity (current joint work with CNRS/IGE)

References

Blanchet, J., Paquet, E., Vaittinada Ayar, P., & Penot, D. (2019). Mapping rainfall hazard based on rain gauge data: an objective cross-validation framework for model selection. *Hydrology and Earth System Sciences*, 23(2), 829-849.

Ayar, P. V., Blanchet, J., Paquet, E., & Penot, D. (2020). Space-time simulation of precipitation based on weather pattern sub-sampling and meta-Gaussian model. *Journal of Hydrology*, *581*, 124451.

Sampson, P. D., & Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, *87*(417), 108-119.

Youngman, B. D. (2023). deform: An R Package for Nonstationary Spatial Gaussian Process Models by Deformations and Dimension Expansion. *arXiv preprint arXiv:2311.05272*.

Météo-France : Séries Quotidiennes de Référence
https://www.data.gouv.fr/datasets/donnees-changement-climatique-sqr-series-quotidiennes-de-reference/

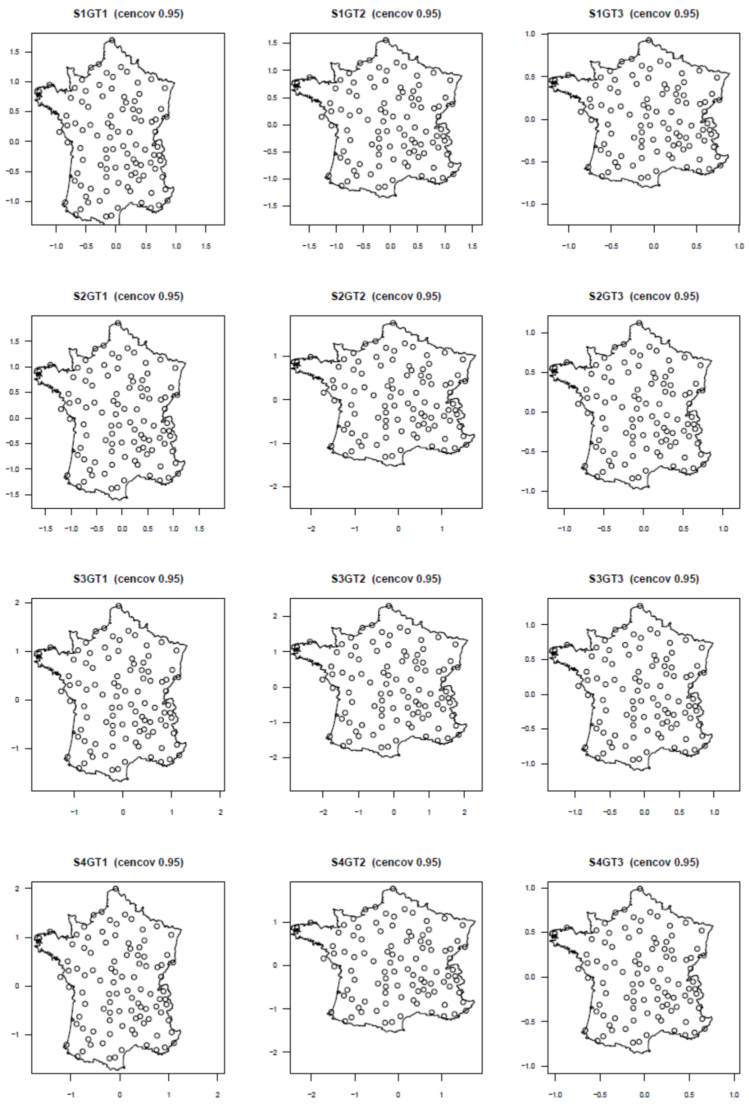
EDF

Merci !

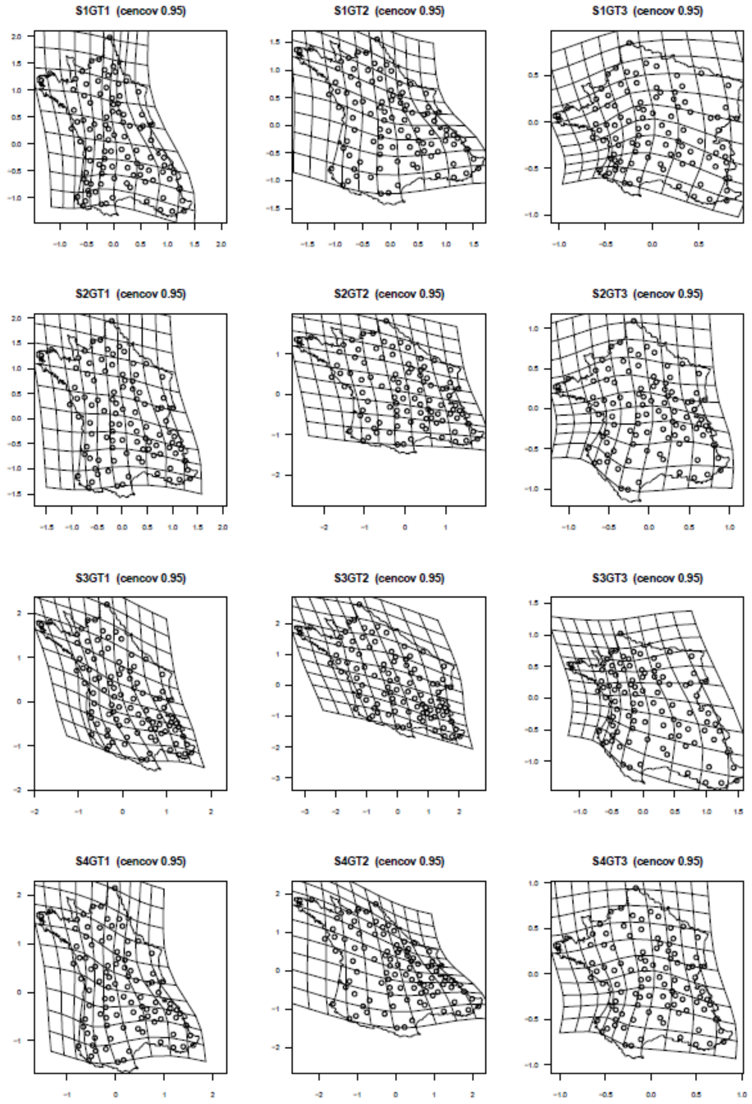
5. COMPLEMENTS

aniso v/s deform

D space - aniso



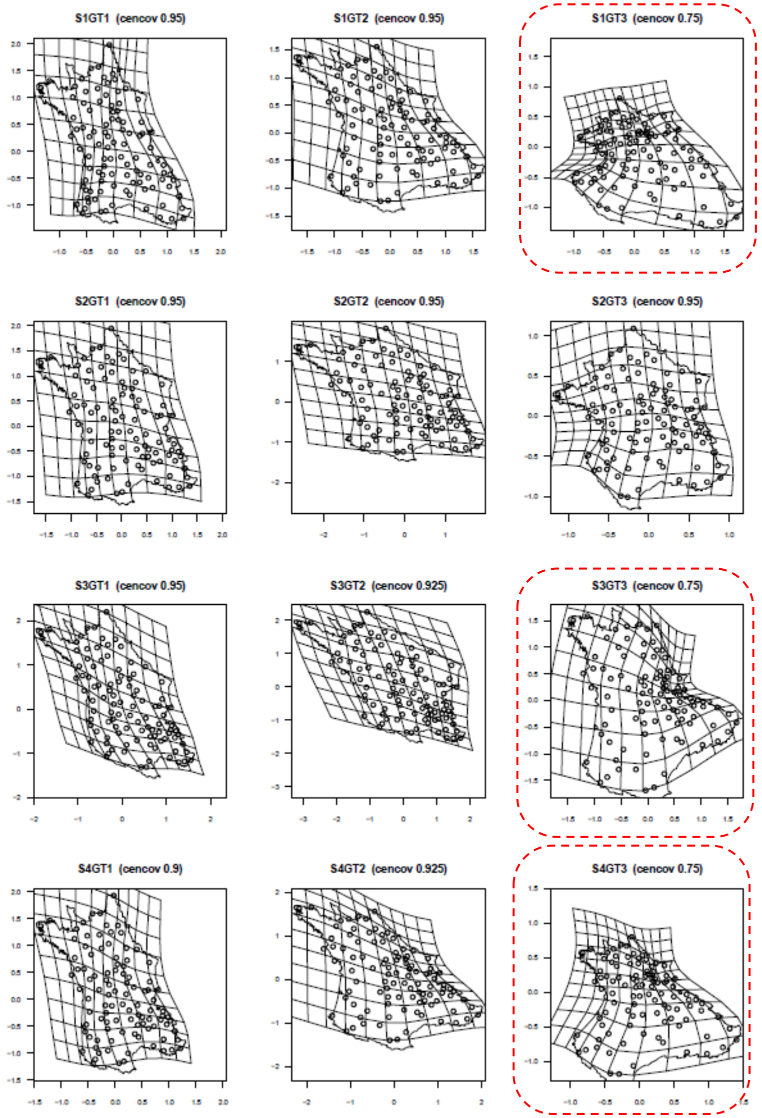
D space - deform



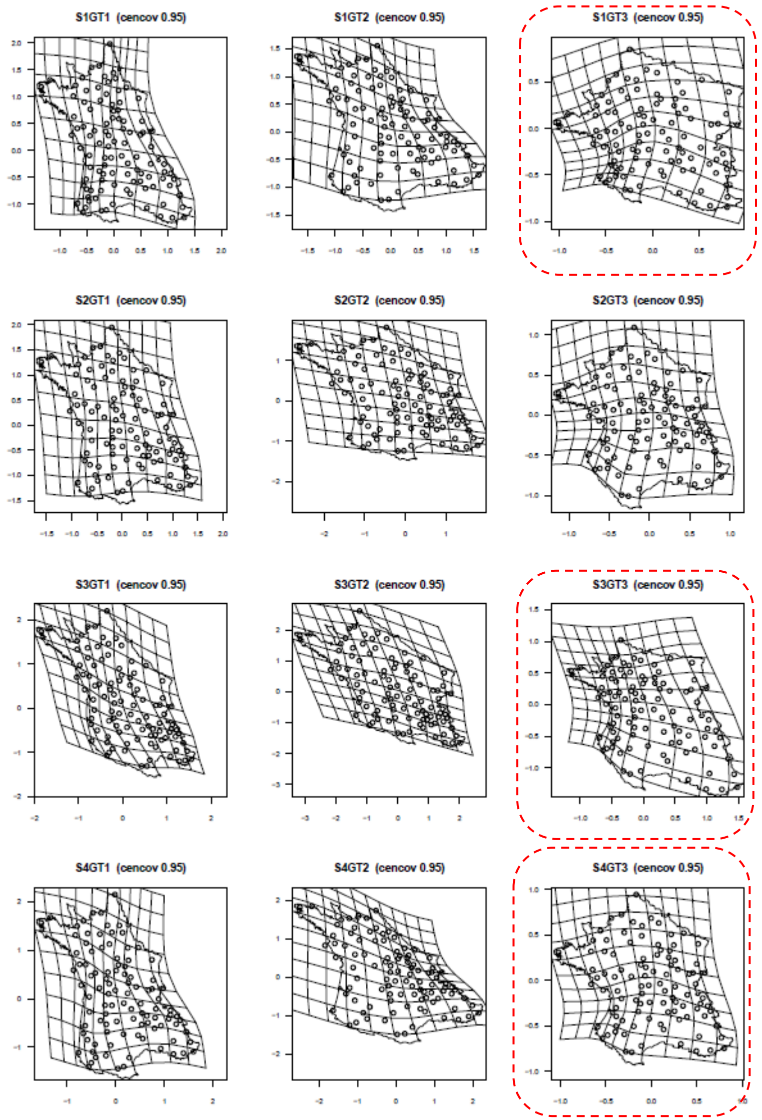
5. COMPLEMENTS

deform
1950-1984 v/s 1985-2022

D space – deform 5084



D space – deform 8522



sensible to period