# A spatio-temporal weather generator for the temperature over France

Caroline Cognot [1&2] & Liliane Bel [2] & Sylvie Parey [1] & David Métivier [3]

[1] *EDF R&D, 91120, Palaiseau, France - caroline.cognot@edf.fr*
[2] *Université Paris-Saclay, AgroParisTech, INRAE, UMR MIA Paris-Saclay, 91120, Palaiseau, France*
[3] *UMR MISTEA, 34000, Montpellier, France*

**Résumé.** Les générateurs de temps sont des simulateurs qui permettent de reproduire la variabilité climatique et générer un grand nombre de situations pour des variables météorologiques selon un modèle statistique ajusté sur les observations. La plupart des générateurs de temps fournissent des simulations pour une ou plusieurs variables météorologiques site par site. Dans ce travail, nous nous proposons de concevoir un générateur spatialisé pour une seule variable météorologique. Nous nous focalisons sur la température. La moyenne et la variance sont décomposées chacune en une tendance et une saisonnalité. La partie stochastique est modélisée par un champ gaussien avec une fonction de corrélation spatio-temporelle non séparable. Ces deux étapes permettent de simuler sur une grille, sur n'importe quelle période de temps. La validation du modèle a été réalisée à la fois sur le jeu de données des stations et sur la grille. En calculant plusieurs indicateurs liés à la structure spatiale, à la structure temporelle ou aux extrêmes, on constate que le générateur fournit des simulations adéquates et permet la génération de séries de températures spatialement cohérentes, y compris dans des extrêmes élevés.
**Mots-clés.** Générateur de temps, processus spatio-temporel, processus Gaussiens

**Abstract.** Weather generators are simulators useful to reproduce climate variability and generate a great number of situations for meteorological variables according to a statistical model fitted on observations. Most weather generators provide simulations for one or several meteorological variables site by site, we aim in this work to design a spatial weather generator for one meteorological variable. We focus on temperature. The mean and variance are each modeled as a trend and a seasonality function. The stochastic part is modeled as a Gaussian field with a non-separable spatio-temporal correlation function. These two steps allow for simulation on a grid, over any time period. The validation of the model was conducted both on the stations dataset and the grid. By computing several indicators related to the spatial structure, the temporal structure or extremes, it is found that the generator provides adequate simulations and allows the generation of spatially coherent temperature series, including high extremes.
**Keywords.** Weather generator, Spatio-temporal process, Gaussian Processes

## 1 Introduction

Weather generators consider the observations of a meteorological variable to be one of the infinite possible trajectories of a stochastic model. The main objective is to allow the simulation of many plausible sequences of a meteorological variable, that share common statistical properties with the observations. The parameters are estimated on the observations. Simulations are fast and allow for a more complete sampling of the climate variables than physical models. Since the first weather generator of Richardson [1981] many others have been proposed using various statistical tools, for univariate and multivariate series. In the case of multisite temperature-focused models, using regular multivariate tools leads to a high number of parameters (for example in the auto-regressive model of Dubrovsky et al. [2020]). Strategies to reduce this amount include Empirical Orthogonal Functions [Sparks et al., 2018], and Gaussian fields [Kleiber et al., 2012]. The latter allows the model to include spatial information. In this work, we focused on building a spatio-temporal Stochastic Weather Generator (SWG) for the temperature using a Gaussian field. The model is built in two steps: first, the mean and variance are represented as a sum of trend and seasonality functions. Trends are depicted as non-parametric functions, whereas the seasonality components are expressed using trigonometric polynomials. Then, a Gaussian field model is fitted on the residuals.

# 2 Methodology

## 2.1 Model

Let $X(s,t)$ be the temperature at location $s$ and time $t$. Following Hoang [2010], $X(s,t)$ is written as a combination of deterministic and stochastic terms :

$$X(s,t) = T_m(s,t) + S_m(s,t) + \sqrt{T_{\sigma^2}(s,t)}\sqrt{S_{\sigma^2}(s,t)}Z(s,t), \tag{1}$$

$T_m$ and $S_m$ are the trend and the seasonality in the mean, $T_{\sigma^2}$ and $S_{\sigma^2}$ are the trend and seasonality in the variance. $Z$ is the stochastic part called residuals. The deterministic terms are interpreted as representing long-term climate variations while the stochastic part represents the intra-annual climate variability.

In order to better capture the long-term dynamics, we use a LOESS (LOcally Estimated Scatterplot Smoothing, Cleveland [1979]) regression which fits linear regression lines locally to the data according to a suitably chosen smoothing parameter. The seasonality terms are modeled as trigonometric polynomials of low order, not exceeding 5, and obtained through linear regression.

We assume the spatio-temporal residuals $Z(.)$ is a Gaussian 0 mean, stationary, isotropic random field. Under these assumptions, it is entirely described by its covariance function $C(h,u)$ where $h$ is a space lag and $u$ is a time lag. Simplest choices of covariance functions do not perform well, we turn to non-separable models and use the Gneiting-Matérn class of Gneiting [2002], popular for meteorological variables.

$$C(h,u) = \frac{\sigma^2}{\left(\left(\frac{u}{a}\right)^{2\alpha}+1\right)^{\tau}}\mathcal{M}\left(\frac{h}{\sqrt{\left(\left(\frac{u}{a}\right)^{2\alpha}+1\right)}^{b}};r;\nu\right), \tag{2}$$

where $\mathcal{M}(d,r,\nu)$ is the value of the Matern covariance function at distance $d$, with smoothness parameter $\nu$ and scale parameter $r$. A nugget effect is added to the covariance models, representing small scale variability and measurement error. The parameters of this model are estimated using a composite pairwise likelihood approach with a grid search for the separability parameters.

## 2.2 Simulation

Simulations are performed over a grid covering the country, according to (1), where the deterministic components are evaluated on each grid point, and the residuals are simulated as a spatio-temporal Gaussian field with covariance function determined by (2).

The seasonality coefficients $\boldsymbol{\beta}_m(s), \boldsymbol{\beta}_\sigma(s)$ are obtained by ordinary kriging [Cressie, 1991] of the coefficients calculated at locations $\{s_1, ..., s_{n_S}\}$ using a Gaussian variogram. The annual cycles are then obtained multiplying these predicted coefficients by the corresponding sines and cosines.

To provide the trends at each grid point, we first correct the trend in mean to account for elevation and then use kriging with a linear variogram.

Simulation of the gridded Gaussian vector have been done in two ways. The first possibility is to build an auto-regressive model of order $\ell$ using the conditionnal distribution of the Gaussian process. The other choice was to use the spectral method of Allard et al. [2020].

### 2.3 Validation

To assess the validity of the weather generator, several indicators are computed on the simulations and compared to those obtained from the observations.

For spatially-focused models, easy to compute indicators are cross-correlation between pairs of stations $(s_i, s_j)$ both in the temperature $Cor(X_i, X_j)$ and in the residuals $Cor(Z_i, Z_j)$. They can be compared with the covariance function.

Another crucial point of interest is the behavior of the generator in high or low temperatures compared to the observations. This can be evaluated by looking at pairwise conditional threshold exceedances probabilities for high or low quantiles, the proportion of locations exceeding a quantile at any given time (spatial hot days), or the distribution of the length of spatial hot days of a given size (heat events).

## 3 Results

The model was fitted on data from 41 weather stations from the ECA&D dataset [Klein Tank, 2002]). These stations are regularly distributed in space and located in areas of low elevation (less than 500m).

For the covariance model, parameters were estimated by extended season (October, November, December, January, February, March - ONDJFM, April, May, June, July, August, September - AMJJAS) to account for possible remaining seasonality using the R package GeoModels [Bevilacqua et al., 2024].

Temperature series were simulated 100 times at each of the 41 fitting data points by first simulating 31 summers and 31 winters from the fitted spatio-temporal covariance model (2) and then adding the deterministic parts in (1).

The Gneiting-Matérn model was simulated using an iterative method and a spectral method. The iterative method was also used to simulate from two additional models. A purely temporal version of this covariance model (same values when $h = 0$, else 0) represents a spatially uncorrelated model. Its objective is to understand how much of the spatial information is carried by the deterministic trend and seasonality as opposed to the spatial model in the stochastic part. A separable exponential model is used to highlight the shortcomings of a too simple model.

Figure 1 (left) shows all of the simulated pairwise correlation coefficients between stations, on the right for the residuals and on the left for the temperature. It shows that the model without spatial dependence (green) produces simulation that greatly underestimate the correlation in the temperature, while the model with spatial dependence (all other colors) reproduce well the observed values.The separable model performs as well as the non-separable model, except that it slightly overestimates low correlations.

Figure 1 (right) shows the pairwise conditional probabilities of threshold exceedance in the observations compared to the simulations. For all models except the spatially uncorrelated model, the values in simulation are in the range of the observations. For the highest quantile (0.99), the small probabilities are over-estimated, which means the model over-estimates the extremal dependance.

This model can also be used on the grid corresponding to the E-OBS dataset. We generated 40 simulation over the same 31 years, and found that it is able to simulate sequences of spatial hot days with similar spatial size or length in time as the 2003 heat wave.

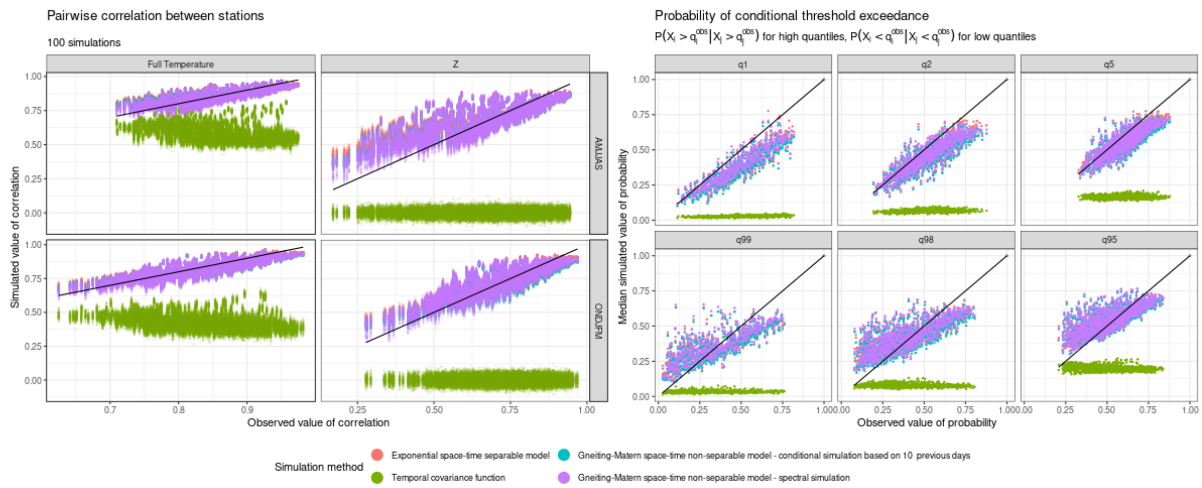


Figure 1: Left : Pairwise correlation coefficient for all stations, first in the full temperature and in the reduced series. Top line represents the result in the extended summer, second line in the winter.
Right : Pairwise conditional probabilities of exceeding a given quantile given another station does. q1 denotes the 1% quantile, q2 the 2%, etc.

# 4 Conclusion

The spatio-temporal stochastic weather generator presented in this study is a tool to generate realistic temperature series across a large area. Making use of spatial statistics, it is able to reproduce the spatial correlation, which makes it possible to use it to simulate plausible large-scale events such as heat waves across the study area. This feature makes it possible to use the model to generate large sample for use in impact studies. Future work will combine this model with other variables, such as precipitation.

# References

D. Allard, X. Emery, C. Lacaux, and C. Lantuéjoul. Simulating space-time random fields with nonseparable Gneiting-type covariance functions. Statistics and Computing, 30(5):1479–1495, Sept. 2020. ISSN 0960-3174, 1573-1375. doi: 10.1007/s11222-020-09956-4.

M. Bevilacqua, V. Morales-Oñate, and C. Caamaño-Carrillo. GeoModels, May 2024. URL https://CRAN.R-project.org/package=GeoModels.

W. S. Cleveland. Robust Locally Weighted Regression and Smoothing Scatterplots. Journal of the American Statistical Association, 1979.

N. A. C. Cressie. Statistics for spatial data. Wiley series in probability and mathematical statistics. Wiley, New York, 1991. ISBN 978-0-471-84336-8.

M. Dubrovsky, R. Huth, H. Dabhi, and M. W. Rotach. Parametric gridded weather generator for use in present and future climates: focus on spatial temperature characteristics. Theoretical and Applied Climatology, 139(3):1031–1044, Feb. 2020. ISSN 1434-4483. doi: 10.1007/s00704-019-03027-z.

T. Gneiting. Nonseparable, Stationary Covariance Functions for Space–Time Data. Journal of the American Statistical Association, 97(458):590–600, June 2002. ISSN 0162-1459, 1537-274X. doi: 10.1198/016214502760047113.

T. T. H. Hoang. Modeling of non-stationary, non-linear time series. Application to the definition of trends on mean, variability and extremes of air temperatures in Europe. phdthesis, Université Paris Sud - Paris XI, Jan. 2010. URL https://theses.hal.science/tel-00531549.

W. Kleiber, R. W. Katz, and B. Rajagopalan. Daily spatiotemporal precipitation simulation using latent and transformed Gaussian processes: PRECIPITATION SIMULATION USING GAUSSIAN PROCESSES. Water Resources Research, 48(1), Jan. 2012. ISSN 00431397. doi: 10.1029/2011WR011105. URL http://doi.wiley.com/10.1029/2011WR011105.

A. a. C. Klein Tank. Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. Int. J. of Climatol., 22, 1441-1453, 2002.

C. W. Richardson. Stochastic simulation of daily precipitation, temperature, and solar radiation. Water Resources Research, 17(1):182–190, 1981. ISSN 1944-7973. doi: 10.1029/WR017i001p00182. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/WR017i001p00182.

N. J. Sparks, S. R. Hardwick, M. Schmid, and R. Toumi. IMAGE: a multivariate multi-site stochastic weather generator for European weather and climate. Stochastic Environmental Research and Risk Assessment, 32(3):771–784, Mar. 2018. ISSN 1436-3259. doi: 10.1007/s00477-017-1433-9.