Stochastic generators for spatial fields of ecoclimatic agricultural indicators in France

Thomas Opitz

Biostatistics and Spatial Processes, INRAE, Avignon, France

joint with

Maël Aubry (Agroclim, INRAE)
Carina Furusho Percot (Agroclim, INRAE)

Focus project *CLIMATHS* (PEPR Maths VivES)
Project *LOST OXIGEN* (INRAE-CLIMAE)
Chair of Geolearning

SWGEN 2025, Grenoble
December 4, 2025

Save-the-date (for another workshop)

Extreme events in agriculture and forestry: challenges, methods and applications

9-11 June 2026

Avignon University

Organized by:
Focus project *CLIMATHS* (PEPR Maths VivES)
Project *LOST OXIGEN* (INRAE-CLIMAE)

Focus topics (in alphabetical order):

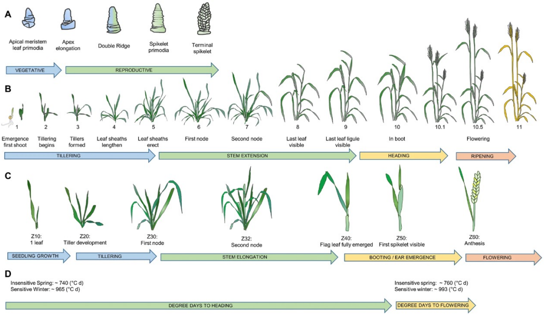
- Climate indicators
- Compound events and risks
- Extreme value theory
- Heatwaves and droughts
- Interactions between biotic and abiotic risks
- Physical processes and climate dynamics of extreme events
- Phenology and ecophysiology of plants
- Rare-event algorithms
- Stochastic weather generators for impact models

Programme and website are under under construction.
Soon available at
`https://extremes2026.sciencesconf.org/`

- **Weather impacts** on crop development and yield depend on **timing** in **life cycle**
- Modeling based on **"crop time"**, not calendar time, to consider crop vulnerability
- **Ecoclimatic indicator:** A meaningful weather summary for a crop risk during a specific phase of the life cycle

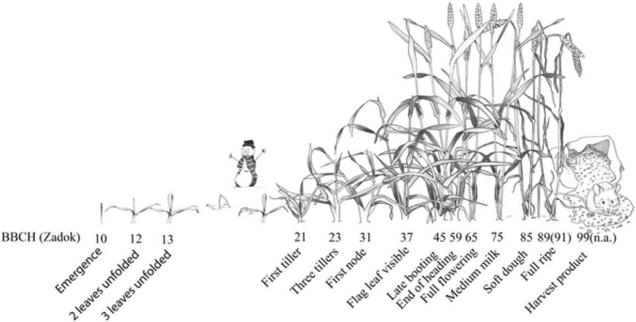
Goal: Development of a low-cost stochastic multivariate spatial ecoclimatic generator

Different representations of wheat phenology events



Source: [Hyles et al., 2020]

- Typically sowing in autumn of year $Y-1$, harvest in summer of year Y
- BBCH scale with 10 development stages from *Germination* (0) to *Senescence* (9)
- **In this work:** Focus on **drought risk** during **spring growing stages** 2–5



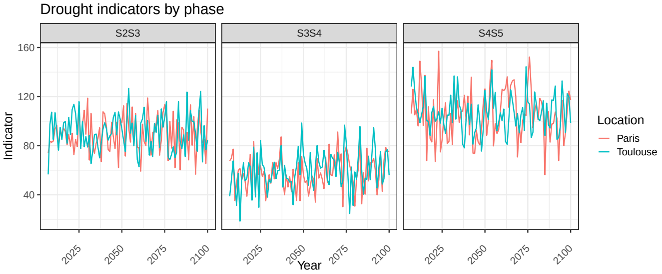
Source: Wikipedia; Group of Crop Science, ETH Zürich

One value per combination of

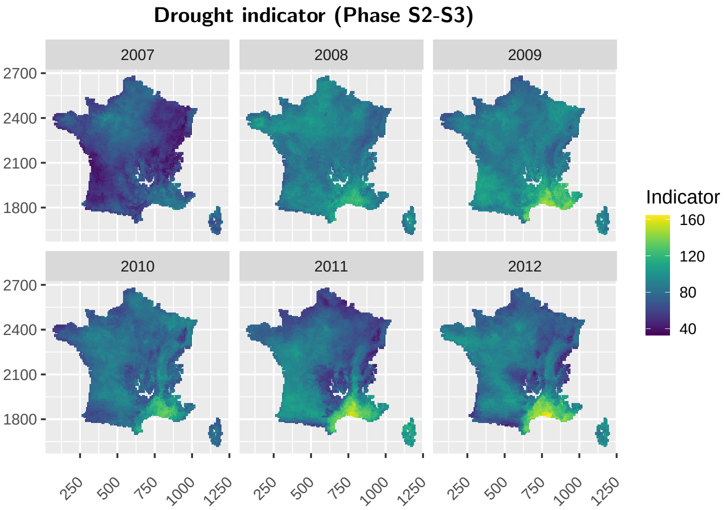
Location + Year + Phase + Indicator + Wheat ideotype

- **Gridded (8km) training data from DRIAS platform:**
  - Scenario RCP8.5, 2007–2100, GCM-RCM model of CNRM, bias-corrected
  - Provides plausible simulations of variability in present and future ecoclimate
- Data based on PhD work of Maël Aubry using 31 indicators for 14 risk categories [Aubry et al., 2025]
- Fixed wheat ideotype common in France
- Phenological phases determined using the STICS crop model
- **Drought indicators** based on Precipitation and Potential Evapotranspiration

Three drought indicator series for Paris and Toulouse pixels
(BBCH phases S2-S3, S3-S4, S5-S5)



Ecoclimate has spatial structure



↪ Stochastic structures with spatial trends and correlation

Some **desirata** (and **solutions**) for a stochastic ecoclimate generator

- **Spatial correlation** for capturing spatial compound events
  $\hookrightarrow$ **Spatial Gaussian copula**
- **Space-varying correlation range** for **large-scale modeling** (France)
  $\hookrightarrow$ **SPDE approach** for Gaussian fields with Matérn-like covariance
  $\hookrightarrow$ **Pre-estimated local range coefficient** as covariate
- **Multivariate modeling** (compound events) with many indicators
  $\hookrightarrow$ **Spatial Blind Source Separation**
  $\hookrightarrow$ Estimate weight matrix W and independent zero-mean Gaussian fields ε_j with

$$\begin{pmatrix} \mathrm{Indicator}_1(s,t) \\ \vdots \\ \mathrm{Indicator}_m(s,t) \end{pmatrix} = W \begin{pmatrix} \varepsilon_1(s,t) \\ \vdots \\ \varepsilon_m(s,t) \end{pmatrix} \quad \text{jointly for all pixels } s \text{ and years } t$$

- **Large numbers of spatial locations** (~ 8000)
  $\hookrightarrow$ **Gauss–Markov Random Fields** (GRMFs)
  $\hookrightarrow$ **INLA-based estimation**
- **Extrapolation** for generating larger extremes than in training data
  $\hookrightarrow$ **Extreme-Value Theory** (Peaks-over-Threshold)
  $\hookrightarrow$ **Conditional GMRF simulation**

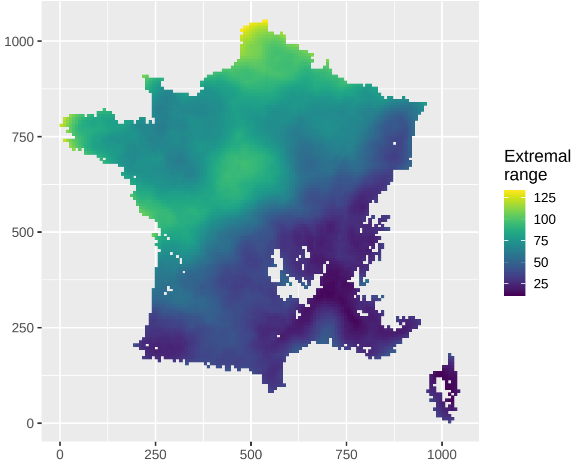
Towards a modular, fast and robust estimation procedure

1. **Estimate pixel-wise marginal model** for each indicator
   ↪ Hybrid distribution with continuous density: kde bulk, Generalized Pareto tail [Opitz et al., 2021]
2. **Transform each indicator to standard Gaussian** using distributions from Step 1
3. **Run Spatial Blind Source Separation** (`stbss` package) on multivariate space-time indicator data
4. **Calculate local range covariate** for each univariate space-time source from Step 3
5. **Fit second-order nonstationary INLA-SPDE models** separately to each source field using the covariate calculated in Step 4 (`R-INLA` package)
6. **Simulate new source field data** (GMRF simulation in `R-INLA` package) and **backtransform** to original fields and marginal distributions
   ↪ GMRF simulation allows conditioning on extreme weighted spatial averages

Example: Constructed covariate for the correlation range

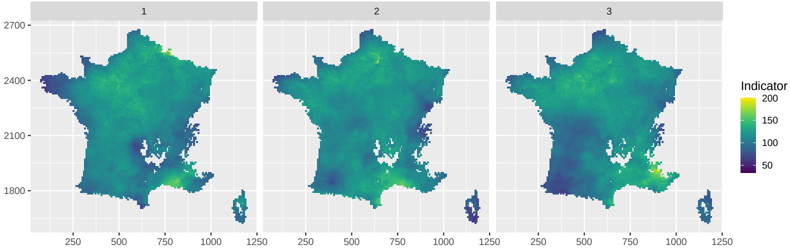
- Covariate = Estimated **median exceedance range** at each pixel [Cotsakis et al., 2024]
- Exceedance range = Distance to nearest non exceedance of a high marginal quantile (given that it is exceeded at the pixel)
- **Nonstationary SPDE Matérn range** estimated as $\rho(s) = \rho_0 \times \text{Covariate}^{\,\rho_1}$

Covariate for Drought indicator (Phase S2-S3) with $\hat{\rho}_1 = 0.7$

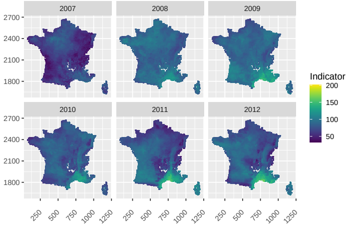


- **Proof-of-concept for a univariate model:** Drought indicator (Phase S2-S3)
- Simulation conditional on **spatial average of two Gaussian standard deviations**

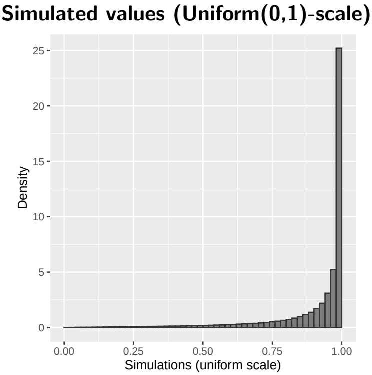
Three new simulations of the Drought indicator (Phase S2-S3)

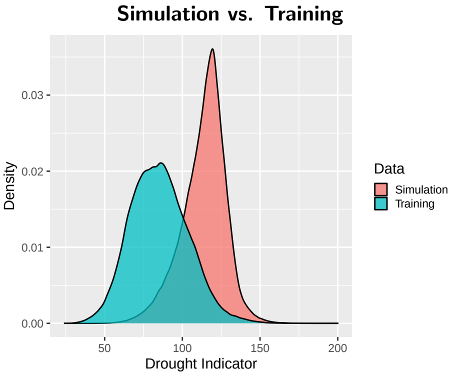


Training examples (same color scale)



New conditional simulations are much more extreme





- **Proof-of-concept for a spatial multivariate ecoclimatic generator:**
  - Construct a **modular and interpretable model** with separately estimable components
  - Enable **scalability** to many components and many locations
  - **Extrapolate** towards new extremes
- **Next steps:**
  - Estimate the **full multivariate spatial model**
  - **Validate** the simulation model
  - Transform indicators into yield distribution with a **statistical yield prediction model**
  - Assess extremes of **aggregated yield losses** at **large spatial scales**

Literature

Aubry, M., Renard, B., Launay, M., García de Cortázar-Atauri, I., and Furusho-Percot, C. (2025).
Sustainability of bread wheat in France: A GAMLSS based risk and return period assessment.
In *EGU General Assembly Conference Abstracts*, pages EGU25–10880.

Bachoc, F., Genton, M. G., Nordhausen, K., Ruiz-Gazen, A., and Virta, J. (2020).
Spatial blind source separation.
Biometrika, 107(3):627–646.

Cappello, C., De Iaco, S., Palma, M., and Nordhausen, K. (2025).
Space-time blind source separation and linear coregionalization modeling: a critical comparison.
Stochastic Environmental Research and Risk Assessment, pages 1–25.

Cotsakis, R., Di Bernardino, E., and Opitz, T. (2024).
On the spatial extent of extreme threshold exceedances.
arXiv preprint arXiv:2411.02399.

Hyles, J., Bloomfield, M. T., Hunt, J. R., Trethowan, R. M., and Trevaskis, B. (2020).
Phenology and related traits for wheat adaptation.
Heredity, 125(6):417–430.

Opitz, T. (2017).
Latent Gaussian modeling and INLA: A review with focus on space-time applications.
Journal de la société française de statistique, 158(3):62–85.

Opitz, T., Allard, D., and Mariethoz, G. (2021).
Semi-parametric resampling with extremes.
Spatial Statistics, 42:100445.

Renard, B., Barbero, R., Goukouni, I., Vidal, J.-P., Mimeau, L., Furusho-Percot, C., de Cortázar-Atauri, I. G., Aubry, M., Opitz, T., and Allard, D. (2025).
Probabilistic modeling of multiple spatial hazards: application to agricultural droughts, hydrological droughts and fire weather.
Technical report, Copernicus Meetings.