**DATA FILES**

Folder **Input Data**

**Data\_Input.RData** contains datosvC2, coords, Ucoords, borderS, nombres

**Grid\_Spat\_Cov.RData** contains elements related to the prediction grid: grid (sf object), mgrid, coordsGrid, tGrid, lGrid, sGrid, zGrid (climate zone of each point in the grid), CovGrid (6 cov: logt, (logt)\*\*2, log (Alt, logCoast,logtlogCoast and IxlagSp; it does not include the lag occurrence Ix and its interaction with logCoast), background,spain\_coords, nT, nL, nS, nP (number of grid points). It was created with Create\_Pred\_Grid. R

**Grid\_Spat\_Isim.parquet** contains objects Isim, Isimlag, with the simulations and the lag simulations of the record occurrence indicators in format *parquet.* They are arrays with dimensions 500 x 92 x 61 x 844

Due to the large size this file is available in the Releases folder

To read this file in R:

tab\_read <- read\_parquet("Grid\_Spat\_Isim.parquet ")   
Isim <- tab\_read$Isim  
Isimlag <- tab\_read$Isimlag

rm(tab\_read)

Folder **Created Data**

Files with created data in the provided code, but available in the repository: Coverages. RData, Cond\_Mean\_Inc.RData, AInc\_Marg\_Summaries.RData, and AInc\_Joint\_Summaries.RData

Other data files created in the modelling and analysis process are too large to make them available, but they can be created using the provided code: **Selected\_Model.RData**, Sim\_Samples\_Obs.RData, Sim\_Data\_Cond.RData, Sim\_Data\_Marg.RData, and Sim\_Data\_Joint.RData

**SCRIPT FILES**

**Fitting and selecting models**

**Fit\_Models.R**: contains all the functions to fit the main models that have been tried. It also includes the function *GoFC* to compute the DIC and RMSE of the models and the code to run the function with the fitted models. It also contains the function *GOFPlots* to compute the standardized residuals and plot them and the code to run the function with the selected model.

Input File: Data\_Input.RData

Output file: **Selected\_Model.RData** contains MSel (the selected model), MSelT, the same model but fitted with specifications to be used to generate samples, datosvC2, nT, nL, nS, mesh. It is a too large file, not available at the repository

**Sel\_Rmse\_OOS.R**: contains the functions *Sel\_Rmse\_OOS* and *Sel\_Rmse\_OOS\_S* to compute the RSME out of sample (cross-validation: ten steps omitting 4 locations at each step) in non-spatial and spatial models respectively. It also includes the code to run those functions with some selected models (M0, M1, MS0, MS1, MS2, MS3; see Table 1 in the paper).

Input File: Data\_Input.RData

**Validating and making inference**

**Val\_Coverage\_OOS.R**: contains the function *Val\_Coverage\_OOS* to compute coverage, out of sample, eliminating four locations at a time and the code to run this function with the selected model and compute the plots of these coverages by year, day within the year and sites (plots in the supplementary material).

Input File: Data\_Input.RData

Output File: Coverages.RData

**Val\_Femp\_OOS.R**: contains the function *Val\_Femp\_OOS* to compute the empirical distribution function at a grid of points and also the mean and CI of the distribution function from the model obtained by simulation; it also includes the auxiliary functions *EmpF*, *estF* and *genaux*. It includes the code to run Val\_Femp\_OOS with 4 stations (Bilbao, Daroca, Huelva and Vitoria) and to make plots; it gives plots in the Supplementary material.

Input File: Data\_Input.RData

**Inference\_SelModel.R**: contains the functions *mapSpatialE* and *plot\_coef* to plot the map of the spatial random effects fitted in a model, and the mean and CI of the fixed effects coefficients fitted in the model. It includes the code to run those functions with the selected mode (resulting plots in Figure 3 in the paper).

Input File: Selected\_Model.RData

**Gof\_Spatialaverages.R**: contains the function *Gof\_Spatialaverages* and auxiliary function *IMY* to compute plots comparing empirical spatial averages (across the observed series) of the increments with the corresponding posterior mean and CI obtained by simulation from the model, by groups (years, days, …). It also includes the code to run Gof\_Spatialaverages with the selected model; it gives the plots in Figure 5 in the paper, that is the spatial averages by year and by day within the year.

To obtain the values by simulation is necessary to generate samples from the model of the response (Increments) at the days where an increment has been observed in the observed series. The function *genParam* generates samples of the parameters of the fitted Gamma parameters, and function *genY* generates the samples of the response using the previous parameters (auxiliary function *genaux*).

Input File: Selected\_Model.RData

Output file: **Sim\_Samples\_Obs.RData** contains sampleP and sampY, objects containing simulated samples of 1,000 observations of the model parameters and the Gamma response, at the observed points and times. It is a too large file, not available at the repository

**Prediction and post model tools**

*The file Create\_Pred\_Grid. R (not available) contains the code to create the prediction grid and Spanish map background, and the covariates at the prediction grid. I t also contains the code for reading the simultions of the record indicators. The result are the files* ***Grid\_Spat\_Cov.RData*** *and* ***Grid\_Spat\_Isim.RData****. Both are available at the repository, but the last one is provided in parquet format since it is too large.*

**SpatialPred\_Inc\_Cond.R**: contains the function *SimCond* (auxiliary function *genaux2*) to generate nSim samples of the response at any day of the observed period at any point in the prediction grid, given the occurrence of the previous day (record or not). The result is sampYP a list of length nSim where each element has nT\*nL\*nP observations. The code to run the function and obtain sampYY0 and sampYP1, samples given that the previous day was not a record and it was.

Input files: Grid\_Spat\_Cov.RData and Selected\_Model.RData

Output file: **Sim\_Data\_Cond.RData** contains the lists sampYY0 and sampYP1, with the samples given that the previous day was not a record and it was. It is a too large file, not available at the repository.

**PostModel\_Inc\_Cond.R**: includes the function *CondMeanInc* that computes the posterior mean of the average increment across JJA in a specific period of time, given the occurrence of the previous day, at each point of the spatial grid. It also includes the code to compute the averages in D3, D4, D5 and D6 and to map the differences between the previous means given that the previous day was a record and it wasn’t.

Input file: Sim\_Data\_Cond.RData and Grid\_Spat\_Cov.RData

Output file: Cond\_Mean\_Inc.RData contains a matrix where each of the 8 columns is a vector of the two conditional means (given the previous day was a record or not) at each decade.

**SpatialPred\_Inc\_Marg.R:** contains the function *Sim\_Marg* (auxiliary function *genaux2*) to generate nSim marginal samples of the response at any day of the observed period at any point in the prediction grid,). The result is sampYP a list of length nSim where each element has nT\*nL\*nP observations. The code to run the function is applied to obtain sampYP.

Input data: Grid\_Spat\_Cov.RData, Grid\_Spat\_Isim.RData and Selected\_Model.RData.

Output file: **Sim\_Data\_Marg.RData** contains the lists sampYP, with samples from the marginal distribution. It is a too large file, not available at the repository.

**PostModel\_Inc\_Marg.R**: contains the function *Marg\_Dist\_Averages* to calculate posterior marginal means and percentiles of the average increments over a time period defined by timeGrid for each point of the spatial grid; *Marg\_Dist\_Dif\_Averages* to calculate posterior marginal means and percentiles of the difference of the average increments over a time period defined by timeGrid for each point of the spatial grid*; Prob\_Comp\_Month\_Averages*  to compute posterior probabilities of the average increments over one month being higher than in another at a spatial grid; *Marg\_Samp\_Averages* to calculate samples of the average increments levels defined in indGrid; *Marg\_Summaries\_Averages* to calculate mean and percentiles of the average increments levels defined in indGrid; *Marg\_F\_Averages* to calculate P(J>ref) of the average increments levels defined in indGrid.

Input file: Sim\_Data\_Marg.RData and Grid\_Spat\_Cov.RData

Output file: AInc\_Marg\_Summaries.RData contains summaries of measures involving the increments and their marginal distribution: MargDistAvD1, MargDistAvD2, MargDistAvD3, MargDistAvD4, MargDistAvD5, MargDistAvD6, MargDistAvTot, MargDistDifAvD6D2, MargDistDifAvD6D3, MargDistDifAvD6D4, MargDistDifAvD6D5, probCompMonthD1, probCompMonthD2, probCompMonthD3, probCompMonthD4, probCompMonthD5, probCompMonthD6, probCompMonthTot, meanTYsamp, meanTSYsamp, summTAve, mPD,

**SpatialPred\_PostModel\_Inc\_Marg.R**: first part contains the function *Sim\_Joint* to generate samples from the joint distribution using samples from the value of the increments (from the marginal distribution) and samples of the occurrence of records. The code to run the function is applied to obtain sampYPJ.

Second part contains the function *Mean\_Averages* to compute the samples and the corresponding posterior mean of the average (over days within the year) of the cumulative increments during the last 30 years at each point of the spatial grid.

Input file: first part: Sim\_Data\_Marg.RData and Grid\_Spat\_Isim.RData; second part: Sim\_Data\_Joint.RData and Grid\_Spat\_Cov.RData.

Output file: First part: **Sim\_Data\_Joint.RData** contains the lists sampYPJ, with samples from the joint distribution. It is a too large file, not available at the repository.

Second part: AInc\_Joint\_Summaries.RData contains a list Avesum30J that includes the sample AveJsamp and the posterior mean AveJmean of the average (over days within the years) of the cumulative increments during the last 30 years at each point of the spatial grid.