An automated adaptative framework for weather data interpolation

Cédric Magain[1*], Guillaume Renard[1*], Philippe Orban[2], Aurore Degré[1], Jeroen Meersmans[1], Caroline De Clerck[1], Serge Brouyère[2], Joost Wellens[1]

1 ULiège, Water-Soil-Plants Exchanges TERRA teaching and Research Centre, Gembloux Agro-Bio Tech, 5030 Gembloux, Belgique;2 ULiège, Urban and Environmental Engineering, 4000 Liège, Belgique

Corresponding authors: cedric.magain@uliege.be, guillaume.renard@uliege.be

Highlights

- A low user effort tool for selecting spatial interpolation methods was built using error metrics.
- The approach was validated with real and synthetic cases ensuring its robustness.
- No single interpolation method is universally preferable for all the variables tested.
- Effectiveness was retained with sparse data, matching or exceeding single methods.

Keywords

spatial interpolation, automatic selection, combined error metric, water balance, evapotranspiration, rainfall, optimum station network, cross-location validation.

The final published version is available at:10.1016/j.jhydrol.2026.136419
© 2026. This manuscript version is made available under the CC BY-NC-ND 4.0 license

AUTHOR ACCEPTED MANUSCRIPT

Abstract

Spatially continuous estimates of reference evapotranspiration and rainfall are crucial for environmental management applications. However, their estimation often relies on weather station measurements characterized by incomplete temporal records. Spatial interpolation methods (SIMs) address this issue but are variable and data-specific, making their selection challenging due to dynamic, unharmonized networks and missing data. Advances in computing capabilities enable automation of selection routines, allowing development of automatic SIMs selection process. Several authors proposed employing combined metrics to optimize model selection. Accordingly, this paper developed a new composite metric (CM), based on random and systematic error, integrated into an automated process, called JW*grid* (Judicious Widget for Geospatial Robust Interpolation Design or JW), to identify optimal interpolation methods across various contexts. Using Leave-One-Out Cross-Validation (LOOCV), the SIM with the lowest CM value is selected from popular interpolation techniques. The methodology was assessed on weather station networks in Luxembourg, evaluating rainfall and reference evapotranspiration related variables. Results showed low error values with SIM selection frequencies being variable dependent. The methodology's robustness was assessed by (1) reducing data to 25, 50, and 75% of the original network size, by (2) using synthetic datasets with varying data variability, network configurations, and presence or absence of an external drift and (3) applying to a mountainous study area. Network density minimally affected overall accuracy, though SIM selection frequencies varied with no universally preferable method. In synthetic tests and the mountainous area, the JW methodology performed at least equally to the best single method by selecting the most

AUTHOR ACCEPTED MANUSCRIPT

appropriate method at every timestep. Those results highlight the proposed methodology's advantages, producing reliable gridded datasets across diverse climatic conditions. The approach enables users with limited SIMs experience to easily generate gridded data, providing an effective solution for spatial interpolation challenges in environmental applications.

1. Introduction

The Food and Agriculture Organisation of the United Nations (FAO) recognised the need to develop a common and accurate method for estimating reference evapotranspiration (ET_0) that could be applied across a wide range of climates or in cases where data are limited. After a series of comparative studies, FAO recommended the modified Penman-Monteith method (Allen et al., 1998). Various sensitivity analyses and regional studies confirmed its applicability to a large variety of environments (Pereira et al., 2015). ET_0 is a complex and non-linear variable since it depends on the interaction of several weather variables (temperature, humidity, wind speed and radiation) (Hodam et al., 2017)(Hodam et al., 2017). It describes the phase transition of water from the liquid to the gaseous state at the surface and the subsequent transfer to the atmosphere. It is a variable made up of three components: evaporation from the soil or water bodies, transpiration from vegetation and evaporation from precipitation intercepted by vegetated surfaces(Yang et al., 2023) (Yang et al., 2023). It is one of the most basic components of the hydrological cycle (Kumar et al., 2013; Ning et al., 2016; Hodam et al., 2017). It therefore plays a fundamental role in studies in agriculture, hydrology and environmental sciences more generally (Ning et al., 2016). Accurate estimates of reference evapotranspiration, alongside rainfall, are therefore essential for the management and planning of water

AUTHOR ACCEPTED MANUSCRIPT

resources (R. Kumar et al., 2013; Hodam et al., 2017) and irrigation (Maruyama, 2007; Verstraeten et al., 2008; Łabędzki et al., 2014; Ning et al., 2016).

Due to the importance of reliable measurements of weather variables, including ET_0 and rainfall, meteorological networks have grown considerably. However, despite denser networks, uncertainties continue to hamper appropriate water management. Indeed, stations may not be harmonised (i.e. they may measure different meteorological parameters) and evolve over time, with the addition and removal of stations. Not every station has all the required weather variables to compute ET_0 and missing data further impede ET_0 estimations. FAO's Irrigation and Drainage Paper no. 56 recommended estimating missing data rather than modifying the basic method itself (Allen et al., 1998); as the intent of FAO was to adopt a single, physically based method (Pereira et al., 2015). Research has focused on solutions to the lack of complete data sets such as the estimation of proxies of missing variables (e.g. Allen et al., 1998) or the development of replacement equations based on machine learning algorithms (e.g. Alizamir et al., 2020; El-Shafie et al., 2013). However, using proxies can introduce temporal discontinuities or spatial uncertainties, and empirical methods may lack transferability across time and space, as there is no replacement for basic physics.

Another approach is using spatial interpolation methods (SIMs) (Li and Heap, 2014; Mardikis et al., 2005). SIMs can be categorised as: geostatistical, non-geostatistical or combined methods (Li and Heap, 2014). SIMs have been extensively studied for generating continuous weather variable maps (e.g. Hofstra et al., 2008; Luo et al., 2008; Ly et al., 2011; Rosillon et al., 2024). However, SIMs are often data and variable specific, with many factors affecting the predicting performance. Furthermore, the most

AUTHOR ACCEPTED MANUSCRIPT

appropriate SIM can even vary from day to day for the same weather variable. For example, in hilly and mountainous areas, the orographic impact on spatial rainfall patterns depends on the associated weather pattern (Gottardi et al., 2012). These specificities hinder the selection of the most suitable method. In that regard, numerous authors have investigated optimal SIMs for specific areas, finding that the most appropriate SIM varies according to study-specific conditions (e.g. Antal et al., 2021; Borges et al., 2016; Delbari et al., 2013; Pellicone et al., 2018; Workneh et al., 2024)

A potential overcoming is the automation of methods' selection. Even while SIMs have been extensively compared, to our knowledge, few studies have attempted to implement automation for optimizing SIMs parameters and even fewer have automated the SIM selection. For example, Betzek et al. (2019) automated the selection of the best parameters for variogram fitting for kriging techniques and the power factor of Inverse Distance Weight method (IDW). Hiemstra et al., (2009) developed an automatic variogram fitting procedure using iterative reweighted least squares for real-time radiation monitoring data mapping.

To select the most appropriate SIM, the choice of adequate performance metrics is crucial. Commonly used error metrics include Bias, also called Mean Error (ME), mean absolute error (MAE) and root mean squared error (RMSE). Both MAE and RMSE are considered among the most reliable measures of model performance (Li and Heap, 2011). Since MAE and RMSE both estimate average error (Li and Heap, 2011) but RMSE gives more penalty to extreme errors, RMSE is retained as the primary metric. Several authors employed composite metrics, combining multiple metrics into a single score, for model evaluation. For example, Vogt et al. (2018) proposed a sum of four equally weighted

AUTHOR ACCEPTED MANUSCRIPT

statistical indices, while Ahmed et al. (2019) and Chen et al. (2011) used ranking-based metrics.

Hence, this paper hypothesized that SIM selection procedures can be automated to generate daily maps of the weather variables constituting ET_0 and rainfall, relying on performance metrics that effectively address the varying density and location of the weather network. As this paper aims to propose a simple approach that ensures straightforward interpretation and effective comparison of different interpolation methods, it is essential to consider both random and systematic errors to prevent cumulative inaccuracies over time. This paper expects that incorporating Bias alongside RMSE in the evaluation process enhances the reliability of interpolation selection. This study develops a novel methodology for interpolating weather variables in an environmental and hydrological context. The approach is designed to automatically select the best suited SIM at the desired time step thanks to a composite performance metric and is tailored to adapt to multiple weather variables and networks configurations. The performance and robustness of the methodology is assessed on an actual case study and on synthetic cases.

2. Materials

2.1 Study area

The study area corresponds to the Grand Duchy of Luxemburg; a landlocked country, that covers 2 586 km^2 bordered by Belgium, Germany and France. The country is divided into two natural regions, the Oesling in the north, a hilly region that forms part of the Ardennes-Eiffel massif, and the Gutland, a less topographically complex region with lower altitude.

AUTHOR ACCEPTED MANUSCRIPT

The country has a temperate oceanic climate, with an average temperature around 10 °C and 830 mm of annual rainfall, evenly distributed through the year (Erpicum et al., 2018). Furthermore, the spatial distribution of annual rainfall is determined to a longer extend by the topographical barrier of the Ardennes-Eiffel massif (Meersmans et al., 2016; Erpicum et al., 2018). Due to its geographical enclosure among multiple countries, which complicates the weather data network, and its diverse topography, the region provides an ideal setting for testing the proposed methodology in a complex environment.

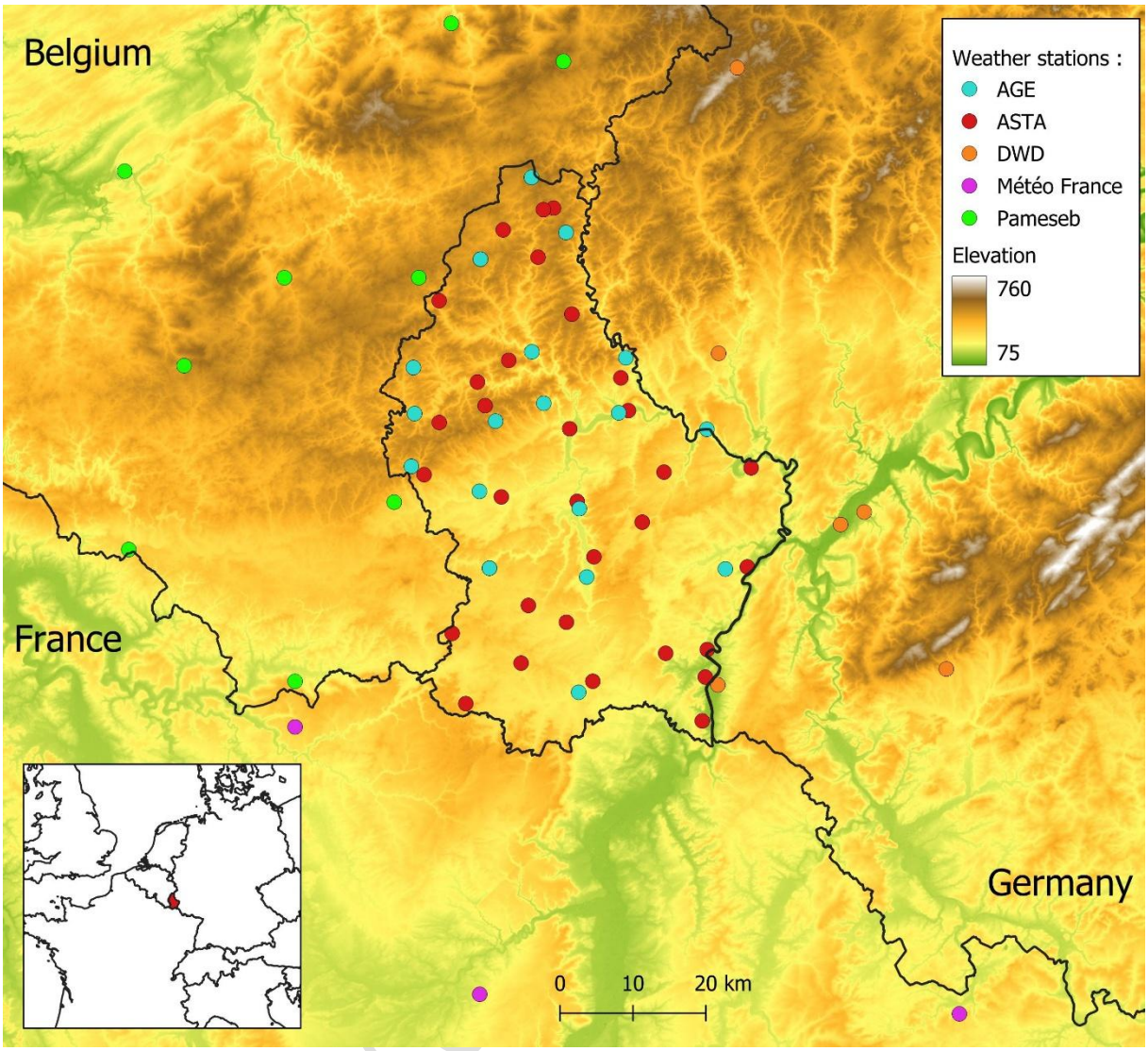
2.2 Weather stations

Interpolations are based on observed weather data from the *Administration des services techniques de l'agriculture* (ASTA) and the *Administration de la gestion de l'eau* (AGE); Luxemburgish administrations covering agriculture and water resources management. To avoid extrapolation and enhance estimates at the outer boundaries of Luxembourg, data from border stations from Météo France, Pameseb (Belgium) and Deutscher Wetterdienst (DWD, Germany) networks were also added to the database (Figure 1). The stations IDs, names, locations, elevation and recorded weather variables are listed in Appendix A. Considered weather variables are maximum and minimum air temperature (TMAX, TMIN), relative humidity (RH), wind speed (WS), solar radiation (SR) and precipitation (P). The data covers the 2010-2020 included period. Due to the lack of harmonization among weather networks and missing data, the number of stations with available observations varies over time. When integrating stations from different organizations, meteorological parameters are not consistently recorded across sites. For instance, the AGE network primarily measures rainfall, while the ASTA network focuses on ET_0 and rainfall. These differences in network priorities lead to variations in data frequency and availability. Most

AUTHOR ACCEPTED MANUSCRIPT

159 weather variables exhibit an increasing trend in the number of observations (Figure A.2 in
160 appendix). This is due to a densification of the meteorological network. Moreover, not all
161 the required climatic variables are available at all stations, and their availability can vary
162 over time within the same station (Figure A.2 and A.3 in appendix). A total of 72 monitoring
163 stations are included in the dataset. Coverage varies across stations: 11% record only one
164 variable, 8% record three variables, 40% record four variables, 11% record five variables,
165 and 30% record all six variables considered in this study. As a result, only 21 out of 70
166 stations can compute a water balance (i.e. P-ET_0). Moreover, even among these 21
167 stations, it is not possible to perform accurate estimates of ET_0 or to calculate water
168 balances for all days at each station (Figure 2). Indeed, at least 29% of the stations have
169 0 to 1% of data missing (Figure A.2 in appendix) This results in a sub-optimal network
170 configuration, which is suited for evaluating the proposed methodology.

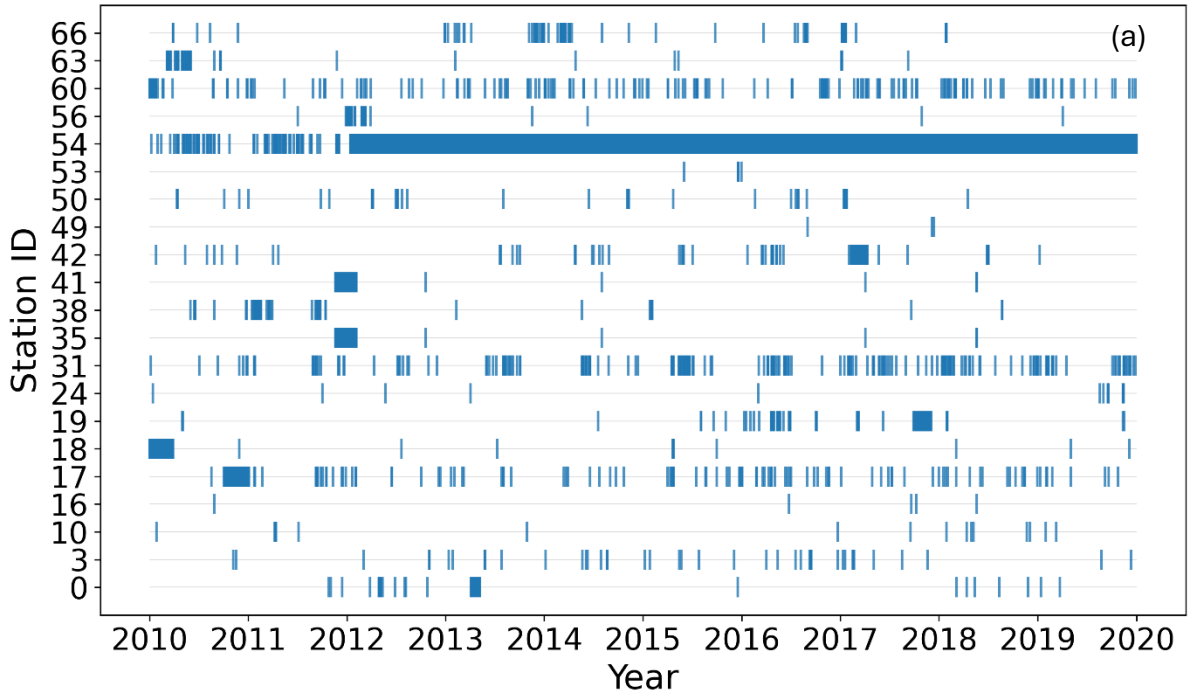
AUTHOR ACCEPTED MANUSCRIPT



171

172 Figure 1. Locations of weather stations included in the database used to introduce the JW*grid*
173 methodology for spatialization and estimation of reference evapotranspiration. Location of
174 Luxembourg in Europe is indicated by a red shape on the map in the lower left corner.

AUTHOR AC



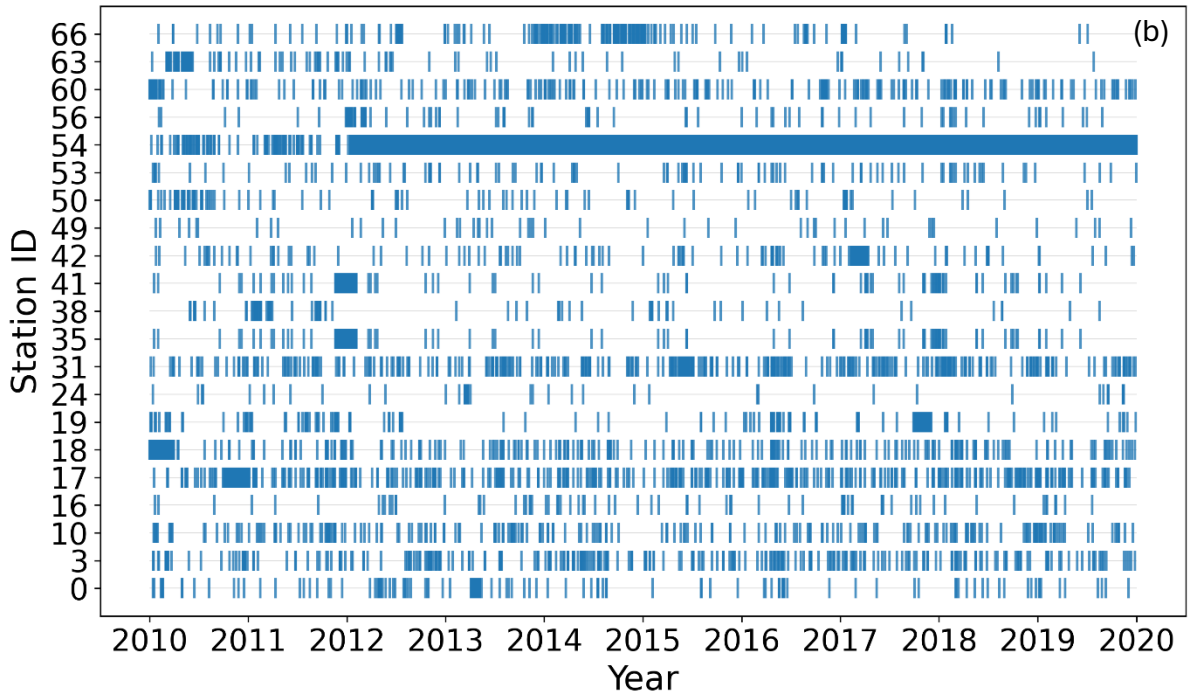


Figure 2. Days (in blue) on which it is not possible to compute (a) reference evapotranspiration or (b) the water balance (i.e. P-ET_0), despite metadata indicating that the stations (21 out of 72) are recording all required climatic variables.

3. Methods

When computing ET_0, two strategies can be identified: (1) interpolate weather factors and subsequently compute ET_0 (Interpolate then Compute - IC), and (2) calculate ET_0 first for each station and then interpolate (Compute then Interpolate - CI) (Ning et al., 2016). The IC methodology takes into account that ET_0 is a composite variable and that each weather variable has different geographical drivers. In contrast, the CI strategy prevents the buildup of errors resulting from the interpolation of various weather variables. For both ET_0 spatialization strategies (CI and IC), the overall error is the cumulative outcome of the measurement error, the estimation equation error and the interpolation error (Mardikis et al., 2005). While the CI methodology performs spatialization only once, potentially resulting in lower spatialization errors; the IC methodology accounts for the topographical effects of each climate factors composing ET_0. In this study, the IC strategy is preferred for several reasons. First, ET_0 is computed using the non-linear Penman-Monteith equation, and Bechini et al. (2000) recommended IC when spatial dependencies are identifiable in the input data (e.g. temperature). Second, each weather variable contributing to ET_0 responds to distinct geographical drivers; interpolating them separately may better capture these patterns and improve accuracy. Consequently, the flexibility of IC offers an advantage, as the possibility of estimating reference evapotranspiration or performing a water balance varies greatly over time and space (Figure 2).

3.1 Datasets cleaning

Albeit data had already been checked by their emitters, a simple Z-score procedure for a few remaining manually detected outliers is performed.

AUTHOR ACCEPTED MANUSCRIPT

$$Z = \frac{x - \mu}{\sigma} \quad (1)$$

With Z: the score Z , x: the observed value, μ, the mean , σ, the standard deviation.

Outliers were excluded on the basis of Z scores, which is the number of standard deviations separating the value in question from the mean. Values with an absolute Z-score of more than three were discarded (Mellenbergh, 2019).

3.2 Interpolation methods

The interpolation methods used in this study are among the most frequently used in environmental sciences (Mardikis et al., 2005 ; Li & Heap, 2014) and are classified as either geostatistical or non-geostatistical. The non-geostatistical methods used are *Inverse Distance Weight* (IDW), *Natural Neighbours* (NatN) and *Thin Plate Spline* (TPS). While the geostatistical methods used are *ordinary kriging* (OK) and *kriging with external drift*) (KED). The considered drift here being elevation. For the ordinary kriging and external drift methods, two semi-variograms are calculated: a so-called 'static' one (abbreviated S) on the mean values per station for the entire study period (KEDS and OKS) and a so-called 'dynamic' one (abbreviated D) for the selected day (KEDD and OKD). The rationale for testing the two variograms lies in the potential day-to-day variability of spatial dependences (Gottardi et al., 2012; Rosillon et al., 2024). For instance, temperature distribution might be influenced by the altimetric profile on certain days, while on others, it could be shaped by atmospheric convection, or other factors. The most suitable variogram model was automatically estimated thanks to the *automap* package (Hiemstra et al., 2009). For KED, the SRTM Version 4 (Jarvis et al., 2008), downloaded through Google Earth Engine (GEE), rescaled to a 1 km^2 resolution was used as the external drift. The SIMs

AUTHOR ACCEPTED MANUSCRIPT

224 used in this study, their strengths and weaknesses, and the associated R packages (R core

225 team, 2023) are presented in Table 1.

226 Table 1. Strengths and weaknesses of tested SIMs and their packages

Method	Advantages	Disadvantages	Package used
Inverse distance weighting (IDW)	Can estimate extreme changes; Can alter the amount of sample points and weight factor	Risk of pit cells effects; Cannot estimate above maximum or below minimum values	Gstat (Pebesma and Graeler, 2003)
Natural neighbours (NaN)	Less pit cells effects (as compared to IDW)	Ridge effects; Cannot estimate above maximum or below minimum values	*Whitebox* (Lindsay, 2016)
Thin Spline plates (TPS)	Passes through all the inputs points; Can estimate above maximum or below minimal values	May oversmooth; Tens to generate extreme values when extrapolating	Fields (Nychka et al., 2001)
Ordinary Kriging with static variogram (OKS) or dynamic variogram (OKD)	Considers spatial dependency; Produces smooth surfaces; Allows to estimate precision at a given location	Requires spatial structure of data; Requires sufficient sampling density and distribution for variogram	Gstat (Pebesma and Graeler, 2003) ; (Hiemstra et al., 2009)
Kriging with external drift with static variogram (KEDS) or dynamic variogram (KEDD)	Considers spatial dependency; Produces smooth surfaces; Allows to estimate precision at a given location; Considers external trend; Allows to estimate precision at a given location	Requires spatial structure of data. Requires sufficient sampling density and distribution for variogram; Requires external trend	Gstat (Pebesma and Graeler, 2003)

227

AUTHOR ACCEPTED MANUSCRIPT

3.3 Interpolation method evaluation and selection

This paper proposes a novel approach that allows a straightforward interpretation and effective comparison of different interpolation methods by focusing on random and systematic errors. When modelling hydrological processes at a (sub)daily timestep, avoiding the accumulation of Bias over time is of utmost importance. To do so, a composite metric (CM), i.e., a metric that combines at least two metrics to produce a single result (Botchkarev, 2018), is proposed. In this study, the Euclidean norm of RMSE and Bias is used as the criterion for selecting the optimal interpolation method. This composite metric represents the shortest geometric distance to perfect model performance, providing an intuitive and balanced assessment of both random errors (RMSE) and systematic deviations (Bias). Integrating these two complementary components ensures that the method accounts not only for the magnitude of errors but also for their directionality.(Jackson et al., 2019; Samantaray et al., 2024).

The Euclidean distance offers a straightforward and interpretable mean of quantifying the separation between an observed performance point and an idealized reference. When performance indicators are conceptualized as orthogonal dimensions, this framework enables their integration into a single composite score (De Liz Arcari et al., 2025). Several established hydrological and environmental performance metrics adopt this principle. For example, the Kling–Gupta Efficiency (KGE) and its variants compute error magnitude from Euclidean distances among correlation, variability, and Bias (Gupta et al., 2009), while Hu et al. (2019) use a similar formulation to construct the DISO combining correlation, absolute error, and RMSE. More broadly, Euclidean-based approaches are increasingly applied to unify multiple evaluation indicators (e.g. Samantaray et al., 2024).

AUTHOR ACCEPTED MANUSCRIPT

The popularity of the Euclidean distance stems from its conceptual simplicity, interpretability and robustness , with applications in model evaluation, parameter optimization, and classification problems (Samantaray et al., 2024).

Despite these considerations, maintaining a low-dimensional Euclidean space, RMSE and Bias, enhances robustness, preserves interpretability and avoids the degradations associated with high dimensionality (Samantaray et al., 2024). Additionally, Using equal weights for RMSE and Bias ensures balanced penalization across components and avoids disproportionate sensitivity to a single indicator score (De Liz Arcari et al., 2025).

$$Composite\ metric = \sqrt{RMSE^2 + bias^2} \quad (2)$$

With RMSE: Root Mean Square Error, Bias: bias error.

Errors were calculated thanks to a leave-one-out cross validation (LOOCV). The LOOCV assumes that each observation is independent and drawn from the same distribution (Wong, 2015). For weather data, some temporal or spatial correlations may exist, which can affect the estimated errors (Daly, 2006). However, because the observation networks cover the diversity of the country's topoclimates, the LOOCV still provides relevant estimates of model performance. The error was computed for each day and each station available as the difference between observed and predicted values. Daily RMSE and Bias were then computed based on the errors of the different stations (Equation 3 and 4).

$$RMSE_j = \sqrt{\sum_{i=1}^{n_i} \frac{(y_i - \hat{y}_i)^2}{n_i}} \quad (3)$$

With j: the day considered, y_i: the observed value at station i, $\hat{y}_i$,: the predicted value at station i, n: the number of available stations for day j.

AUTHOR ACCEPTED MANUSCRIPT

$$Bias_j = \sum_{i=1}^{n_i} \frac{y_i - \widehat{y_i}}{n_i} \quad (4)$$

With j: the day considered, y_i: the observed value at station i, $\widehat{y_i}$,: the predicted value at station i, n: the number of available stations for day j.

In order to compare the performances of the JW method between different variables, an average relative composite metric is computed after the simulations. It consists of the spatial and temporal (2010-2020 period) average composite metric divided by the interquartile range of the observed data.

$$relative\ CM = \frac{\frac{\sum_{j=1}^{n_j} CM_j}{n_j}}{Q3 - Q1}$$

With j: the day considered, CM_j: the CM computed for day j, n_j: the number of days, Q3 and Q1: the third and first quartiles of observed data.

3.4 Time step interpolation scheme

Each timestep, here daily, for each weather variable, SIMs are tested on the available data set using the 1km^2 SRTM grid and accuracies are assessed thanks to the LOOCV. The interpolation with the lowest composite metric value is selected. This process is repeated daily, enabling the selection of the most suitable interpolation method for each specific day. The workflow, constituting the JW method, developed in R (R core team, 2023), is repeated in Figure 3. Safeguards are also implemented; for example, interpolated outputs for precipitation or solar radiation are automatically set to zero when the computed values are negative. Furthermore, no SIMs are performed on days for which only a single input value is specified. Cumulative maps can then be produced to estimate annual or

AUTHOR ACCEPTED MANUSCRIPT

seasonal balances. In addition, the proposed methodology is independent to the time step.

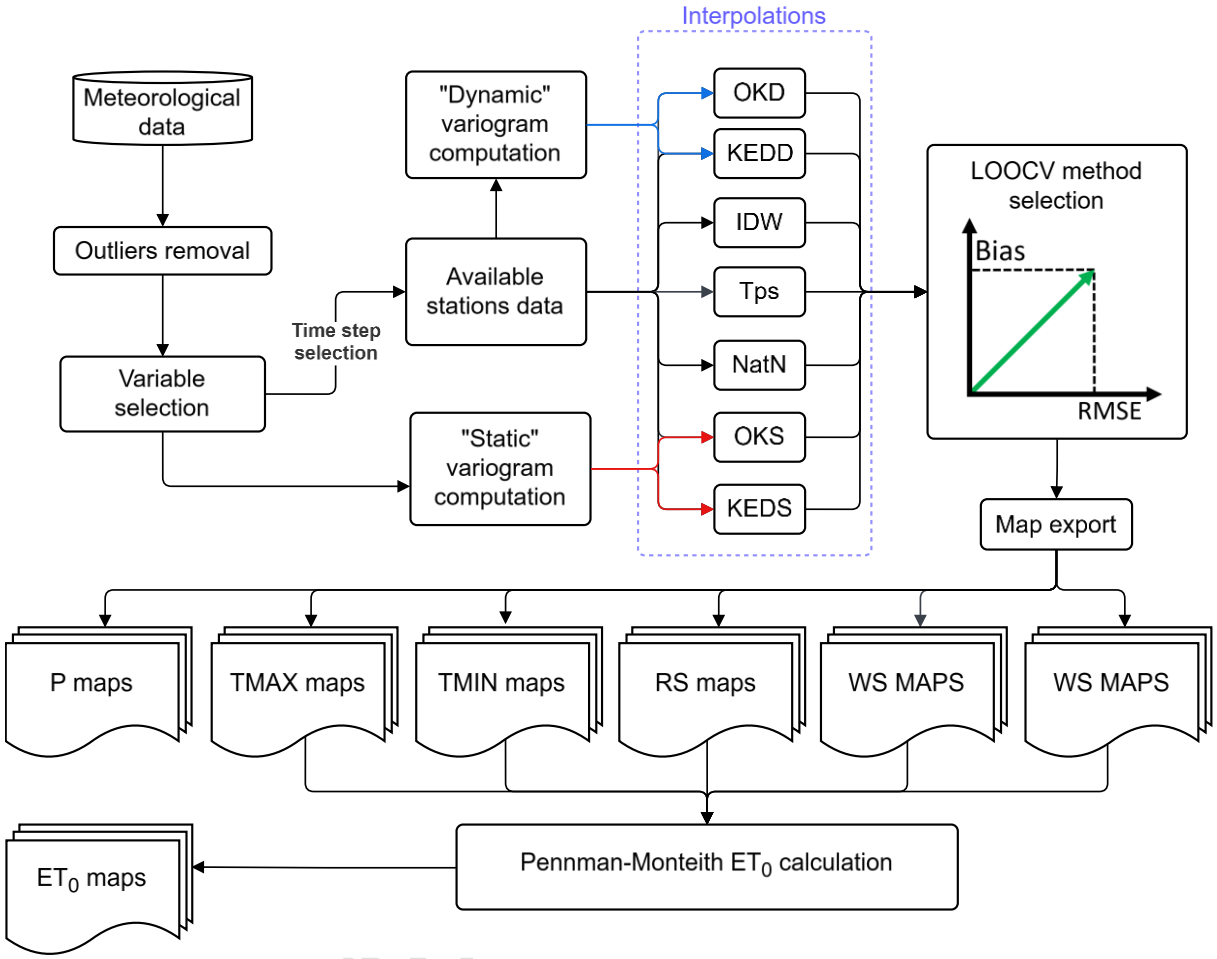


Figure 3. Computational workflow of the interpolation and selection of spatial interpolation methods, constituting the JW method.

3.5 Estimation of reference evapotranspiration

ET_0 is calculated using the Penman-Monteith equation proposed by the Food and Agriculture Organization of the United Nations (FAO) in Irrigation and drainage paper 56 (Allen et al., 1998). The variables required are TMAX, TMIN, RH, RS and WS.

$$ET_0 = \frac{0.408\,\Delta\,(\mathrm{RS} - G) + \gamma\,\frac{900}{T+273}\,u_s\,(e_s - e_a)}{\Delta + \gamma\,(1 + 0{,}34\,\mathrm{WS})} \quad (5)$$

*With ET_0: the reference evapotranspiration [mm/day], RS: the net radiation at the crop surface [MJ/m^2*day], G: the soil heat flux density [MJ/m^2*day], WS: the wind speed at a height of 2 meters [m/s], T: mean value of TMAX and TMIN [°C], e_s: the saturated vapor pressure [kPa], e_a: the actual vapor pressure [kPa], e_s-e_a: the saturation deficit [kPa], Δ: the slope of the vapor pressure curve [kPa/C], γ: the psychrometric constant [kPa/C]*

3.6 Robustness analysis

SIMs' performances can be limited by spatial density and distribution, temporal variation, surface type, landscape characteristics and their combined effect (Li and Heap, 2014). To demonstrate that our methodology can leverage individual strengths of SIMs, two approaches are considered. First, the effect of station density and distribution is analyzed through an occurrence-based approach on the case study. This involves selecting a percentage of the total available stations. The developed methodology is applied to temperatures and precipitation using 25%, 50%, and 75% of the weather stations randomly selected from those available on the considered day. Second, to challenge our methodology, specific artificial cases are constructed. Several authors employed artificial datasets to evaluate their methods. For example, Williams (2025) assesses the suitability of NSE and KGE for hydrological model performance by generating synthetic discharge data. In spatial interpolation cases, Mushagalusa et al. (2024) generates artificial species-abundance data to compare Random Forest performance under standard versus spatial cross-validation . Similarly, Sekulić et al. (2020) complements real case studies with a dataset generated through sequential simulation. The key

AUTHOR ACCEPTED MANUSCRIPT

advantage of such synthetic datasets is the full control of influencing factors (Mushagalusa et al., 2024).

In this study, artificial cases are situations unrelated to the case study in terms of input weather data and station locations. The only features retained from the real case study are topography and countries' boundaries. Artificial cases consist in creating particular situations by combining spatial distribution of artificial station locations, synthetic dataset and spatial correlation. In practice, for each computed situation, a normally distributed random sample of 30 values to be interpolated is generated. Its mean value is 50 and its standard deviation is either 10 or 20 (column input data distribution in Table *2*). This allows for the analysis of both low and high variance input distributions. Secondly, those 30-artificial interpolation input can be assigned on a uniform or random grid (Network configuration in Table 2). On a random spatial distribution, the spatial configuration varies with each analyzed step. To achieve this, a grid of 1024 possible point locations is generated over the study area. The 30 interpolated values are then placed either following a regular grid pattern (first row in Table 3) or randomly assigned to positions on the pre-existing grid (second row in Table 3). To ensure a consistent loss of points, they are distributed such that one-sixth of them are located in the adjacent countries Thirdly, in order to mimic the loss of stations, stations from adjacent countries are retrieved (columns in Table 3). This reflects the possibility that certain station networks may experience temporary failures or that their data may no longer be available through open access. Finally, the combination of those cases is repeated for i) randomly spatial distributed points, ii) spatially correlated points with high-altitude stations and iii) without high-altitude stations, to challenge SIMs correlated to altitude. This ensures artificial cases where the relationship of synthetic stations values with elevation doesn't

AUTHOR ACCEPTED MANUSCRIPT

exist (Appendix B). For spatial correlation without high-altitude stations, the latter are removed, resulting in a loss of additional stations. Both cases (ii) and (iii) are based on the digital elevation model (DEM) of the case study area.

The artificial case consists of 48 scenarios, derived from 16 combinations of artificial input data distribution and artificial network configurations, each applied across three topographical drivers (Table 2). Each scenario is repeated 100 times (Sekulić et al., 2020), resulting in a total of 4800 simulations processed through the selection method. As an example, the HS3 configuration involves applying 100 times the JW method. They are a combination of a high standard deviation (H) as artificial input data distribution with a systematic grid (S) as artificial network configuration and three station networks being considered (3). Those combinations are repeated for each geographical driver. By combining input data distribution, station spatial location, and the presence or absence of correlation with altitude, this paper aims to demonstrate both the strengths and limitations of the JW method.

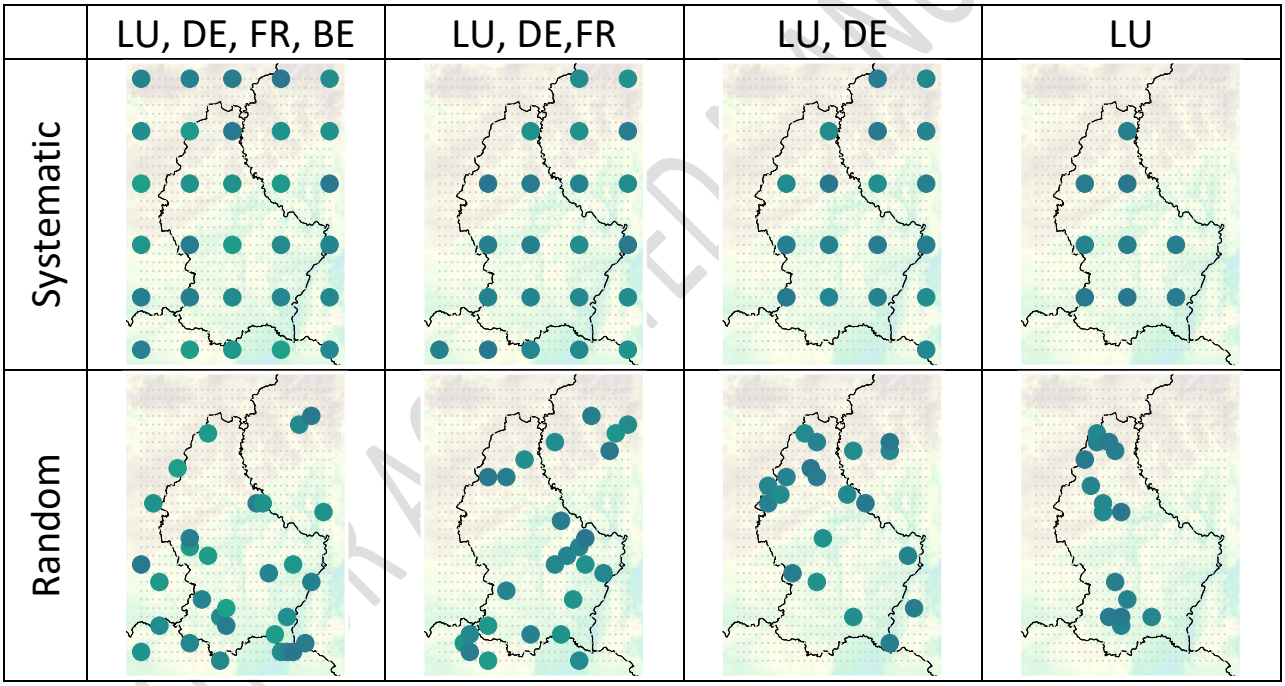
Table 2. Creation of artificial cases to challenge the JW methodology, artificial cases are labelled using a code where the first letter indicates the standard deviation of input values: 'L' for low and 'H' for high. The second letter denotes the network configuration: 'S' for systematic and 'R' for random. The accompanying number represents the number of networks. Each configuration is applied to three scenarios: (i) randomly distributed points, (ii) spatially correlated points including high-altitude stations, and (iii) spatially correlated points excluding high-altitude stations.

Input data distribution (standard deviation)	Network configuration	Stations networks considered			
		LU, DE, FR, BE	LU, DE, FR	LU, DE	LU

AUTHOR ACCEPTED MANUSCRIPT

Low (10)	Systematic	LS4	LS3	LS2	LS1
	Random	LR4	LR3	LR2	LR1
High (20)	Systematic	HS4	HS3	HS2	HS1
	Random	HR4	HR3	HR2	HR1

Table 3. Station location configurations, with LU (Luxembourgish), DE (German), FR (French), and BE (Belgian) boundaries used to define the different artificial network setups. The systematic distribution of the station network remains identical across all repeated artificial test cases, whereas the random distribution changes with each computation.



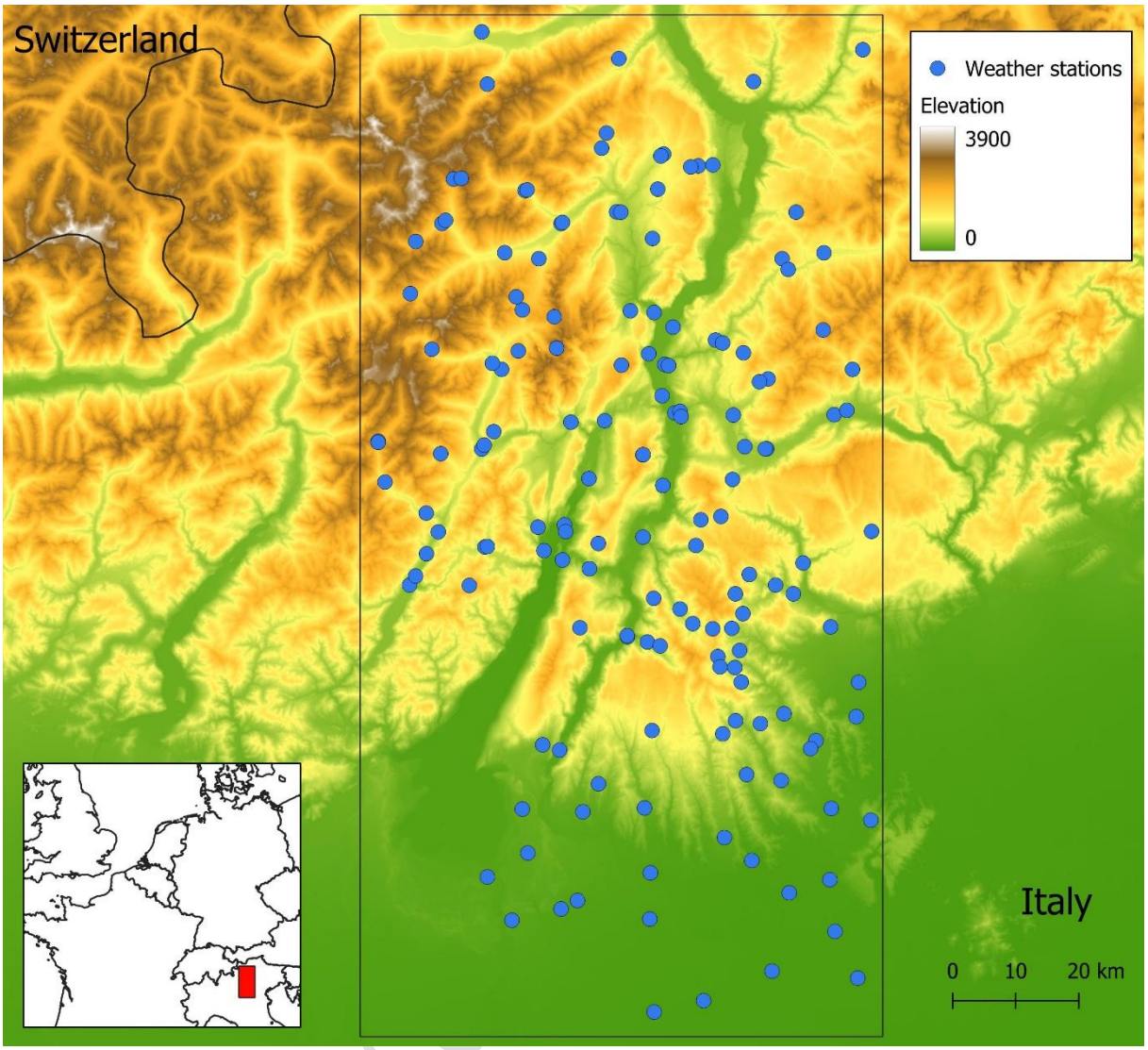
3.7 Mountainous case

To further evaluate the transferability of the JW method, a new study area is introduced in this paper. This area, located in the Alpine region (Figure 4), was chosen to test the methodology in a setting characterized by a remarkably higher degree of topographical complexity and, consequently, greater variability in meteorological variables. Mapping

AUTHOR ACCEPTED MANUSCRIPT

precipitation in mountainous or high-altitude catchments is difficult due to strong spatial variability and limited observational data (Gottardi et al., 2012). The dataset is sourced from the European Climate Assessment & Dataset (ECA&D) (Klein Tank et al., 2002), which compiles thirteen climatic variables from over 19 000 stations across Europe and the Mediterranean region. Two interpolation grids were tested: 1 km and 5 km resolutions. Compared to the artificial test cases developed to evaluate methodological robustness, this dataset contains more data points but less year of records, as it represents a real-case scenario to test the limits of the JW methodology. The dataset spans one year of observations. The study area covers a 13 456 km^2 region, which is approximately five times larger than the Luxembourg case study. Within this area, 151 stations record precipitation during the study period. However, the observed precipitation values recorded at different stations vary over time. More than 80% of the stations have less than 1% missing data. Stations terminate their records during 2020, while others exhibit large gaps or occasional missing values. Additional information regarding missing data is provided in Appendix A.

AUTHOR ACCEPTED MANUSCRIPT



391

392 Figure 4. Locations of weather stations included for the mountainous case study. Location of the

393 study are in Europe is indicated by a red shape on the map in the lower left corner.

394

AUTHOR ACC

4. Results and discussions

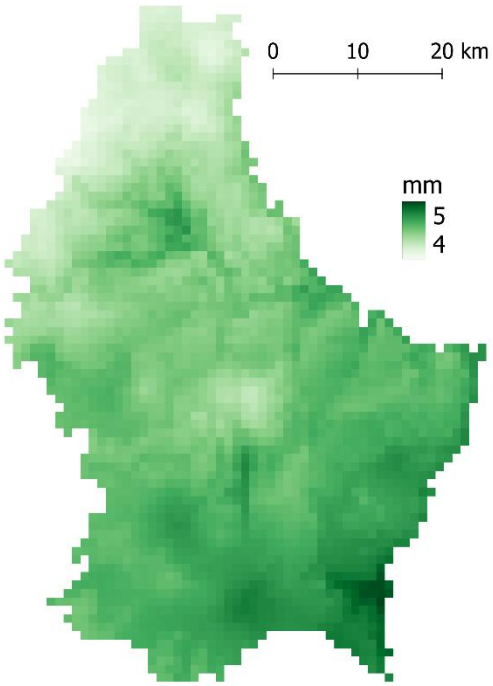
4.1 Case study: the Grand Duchy of Luxemburg

4.1.1 Resulting maps

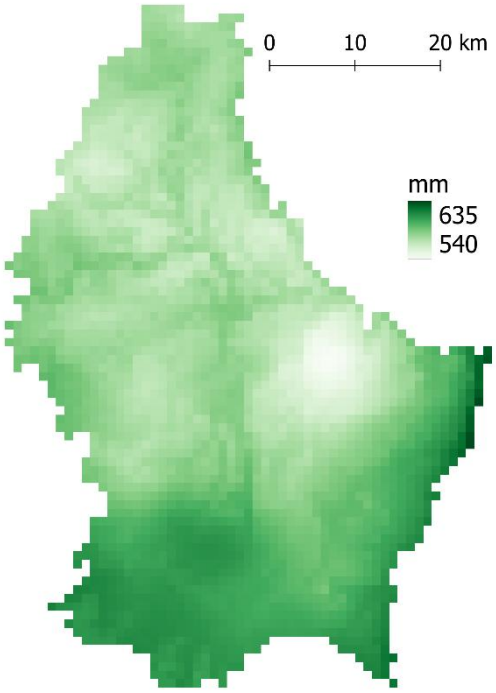
Since the developed methodology is time-step independent, it enables the computation of both timestep and cumulated maps. Initially, interpolated maps (in this study, daily) are generated for each weather variable (TMAX, TMIN, RS, RH, WS, and P). Subsequently, ET_0 is calculated using the Penman-Monteith equation for each cell (Figure 5 (a) and (b)). For the specific day analyzed, ET_0 appears to correlate with elevation, exhibiting a spatial contrast of 1mm across the country.

Timestep maps can be aggregated over specific periods to create cumulative and mean maps, revealing regional trends. Mean annual cumulative ET_0 patterns align with altitude gradients (Figure 5). While precipitation maps highlight localized pits and peaks consistent with mean annual precipitation patterns presented in Erpicum et al. (2018) (Figure 5 (c) and (d)).

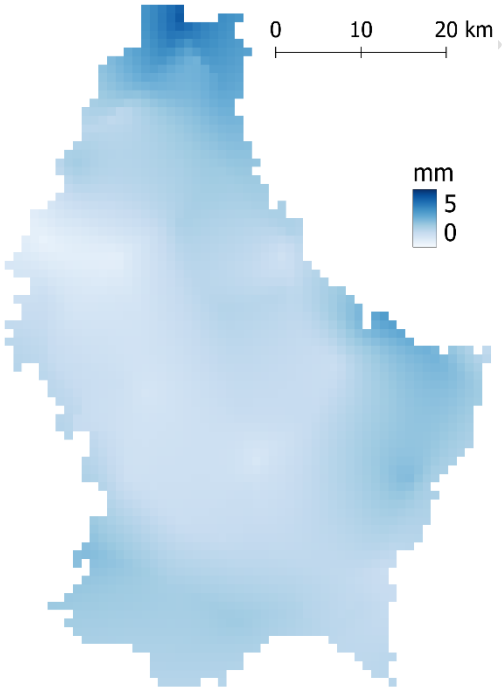
AUTHOR ACCEPTED MANUSCRIPT

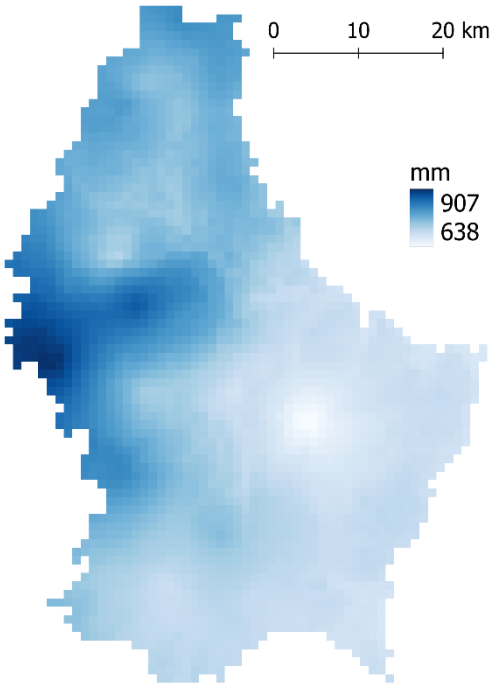


(a) spatialized reference evapotranspiration values for one specific day (17-07-2019)



(b) Mean annual cumulated reference evapotranspiration values (2010-2020)





(c): Daily spatialized precipitation values for a specific day (10-04-2018)

(d): Mean annual cumulated precipitation (2010-2020)

Figure 5. Outputs interpolation map from the JW method on a 1 km^2 grid for daily outputs (a-c) and mean annual cumulated (c-d), spatialized reference evapotranspiration being computed through CI method.

4.1.2 SIMs' selection occurrences

In Figure 6, the SIMs' selection occurrences are presented by weather variables. The prevalence of KED for TMAX, TMIN, WS and RH is remarkable, indicating the interest of using elevation as a predictor for those variables. It is also remarkable that for SR, non-geostatistical methods have been selected predominantly, those results contrast with Antonanzas et al. (2015) who identified OK as the most suited method for interpolating SR over Spain. Nevertheless, the network settings differ and the performance differences among the different methods in our study are minimal (Figure *7*). Jeong et al (2017) also found that geostatistical interpolations were more performant than stochastics one in dense networks, while stochastic ones had the same performance in low and dense networks.

Concerning WS, Luo et al. (2008) yielded the best results using altitude as an external drift for modeling the variables over England and Wales, which agrees with the presented selections.

Regarding P, Ly et al., (2011) also found OK and IDW to be the best methods to interpolate daily precipitation in Belgium; which are also amongst our most selected, with IDW being selected more when density was high. KED techniques are often used because of their ability to represent orographic barriers (Meersmans et al., 2016). In the daily case, these

AUTHOR ACCEPTED MANUSCRIPT

trends are however less pronounced; probably causing the low selection of these models. The work conducted by Wagner et al., 2012 supports this trend. They did not emphasize the importance of KED as no relationship was found between altitude and precipitation.

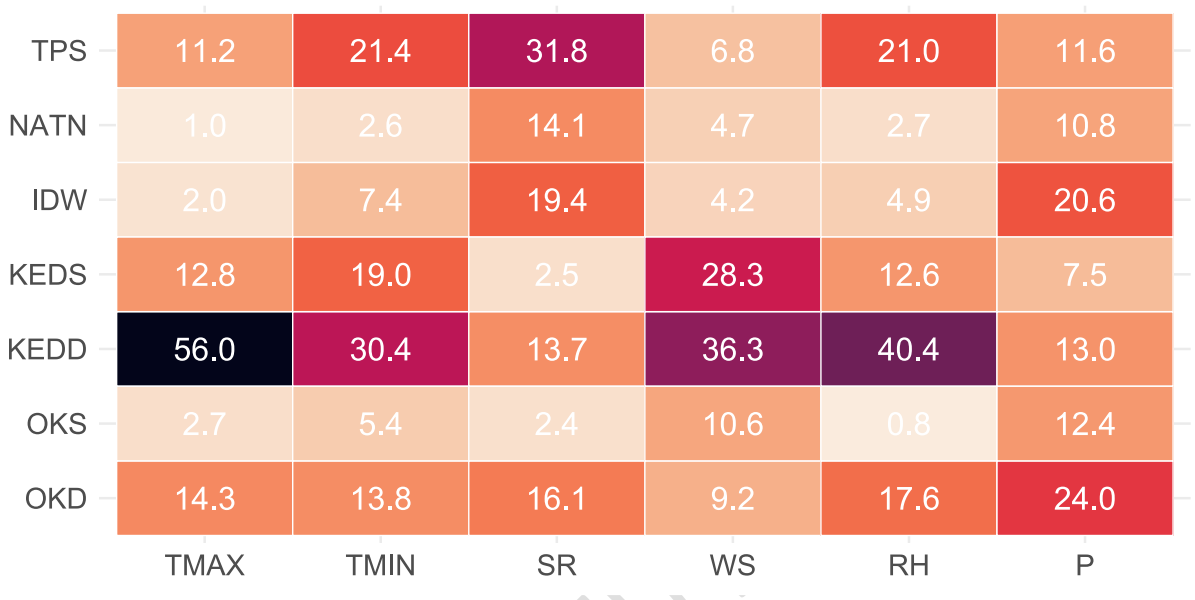


Figure 6. Relative frequencies (in percent) of model selection (y-axis) for the weather variables studied (x-axis) across the Luxemburg country based on JW method for the 2010-2020 period. Red indicates lower selection of the interpolation method, whereas black denotes the method most frequently selected, with TMAX: maximum daily air temperature, TMIN: minimum daily air temperature, RS: solar radiation; WS: wind speed, RH: relative air humidity, P: precipitation, TPS: thin spline interpolation method, NATN: Natural Neighbours, IDW: Inverse Distance Weight, KEDS: kriging with external drift using a static semi-variogram (computed from mean values per station for the entire study period), KEDD: kriging with external drift using a dynamic semi-variogram (computed from available values of different stations for the selected day), OKS: ordinary kriging using a static semi-variogram (computed from mean values per station for the entire study period), OKD: ordinary kriging using a dynamic semi-variogram (computed from available

AUTHOR ACCEPTED

values of different stations for the selected day),), JW: the developed methodology (Judicious Widget for Geospatial Robust Interpolation Design)

Overall accuracies and comparison with SIMs

Overall accuracies

The first and third quartiles (Q1-Q3) of observed data, along with RMSE, Bias, the composite metric and the relative composite metric (composite metric divided by the interquartile range) are presented as country averages for 2010-2020 period in Table 4. Relative error metrics were not computed at the daily time step due to the occurrence of infinite values in some cases. The RMSEs and biases used for this calculation are those associated with the selected SIMs. Compared to their observed interquartile range, the RMSE results are satisfactory for TMAX, TMIN, RH, and P. For RS and WS, the errors appear relatively larger; but mainly due to seasonal variations. Bias values remain relatively low for all variables compared to their Q1 and Q3, which is encouraging for the use of daily time steps and the cumulative error over time. When comparing RMSE, Bias and the composite metric, the values of the latter are closer to those of RMSE primarily due to the higher magnitude of RMSE. However, for variables like RS, the influence of Bias is noticeable. Finally, TMAX, TMIN and WS exhibit relatively low σ_{CM} values, indicating more consistent daily CM values. In contrast, SR exhibits a σ_{CM} value close to its mean CM value, and P exhibits a σ_{CM} value even higher than its mean CM value, possibly reflecting greater day-to-day variability in model accuracy.

Table 4. Summary statistics and error metrics for weather variables across the country (2010-2020), with Q1-Q3 values of observed data, average RMSE, Bias, Composite Metric (CM), standard deviation of the composite metric and relative Composite Metric (divided

AUTHOR ACCEPTED MANUSCRIPT

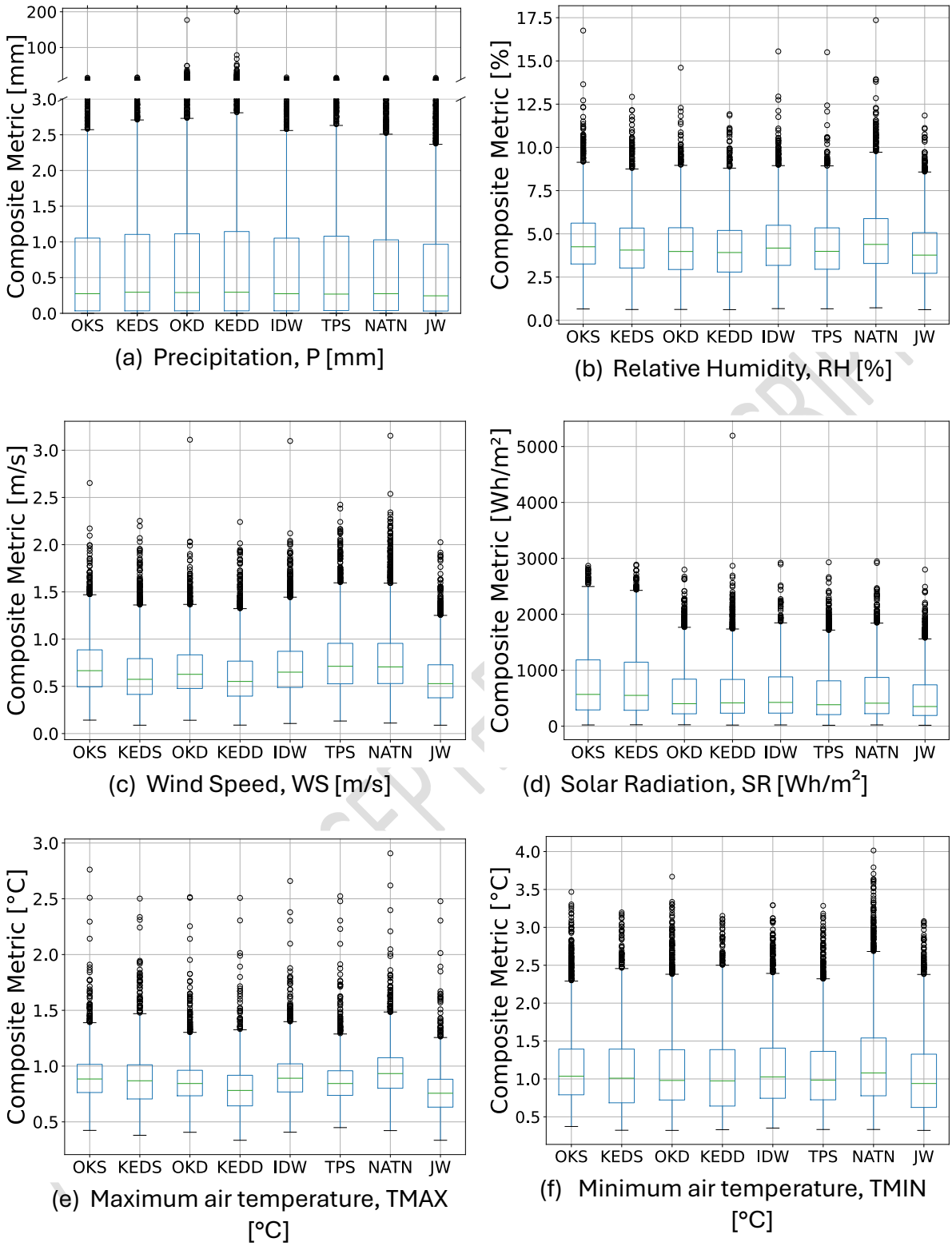
by the interquartile range) values of the different weather variables for the country, TMAX: maximum daily air temperature, TMIN: minimum daily air temperature, RS: solar radiation; WS: wind speed , RH: relative air humidity and P: precipitation computed based on JW method for the 2010-2020 period.

Variable	Q1-Q3	RMSE	Bias	CM	σ_{CM}	Relative CM
TMAX (°C)	7.66-20.85	0.76	0.07	0.77	0.18	0.0583
TMIN (°C)	0.87-9.73	1.03	-0.07	1.04	0.50	0.1174
SR (wh/m^2)	763.09-4482.12	487.29	-49.43	502.06	419.09	0.1350
WS (m/s)	1.30-2.32	0.57	-0.01	0.58	0.27	0.5688
RH (%)	73.62-92.53	3.98	0.41	4.02	1.64	0.2125
P (mm)	0.02-2.86	0.69	0.0006	0.69	1.07	0.2448

JW method against individual SIMs computation

The accuracy of the JW method compared to individual methods is evaluated using boxplots of CM values for each weather variable (Figure 10). Plots of the RMSE and Bias values are available in the Appendix D. It is noteworthy that the median CM of the JW method is equal to or slightly better than those of the single methods for all variables. Additionally, the JW method allows to avoid extreme CM values, which is particularly advantageous when computing annual cumulates or using spatialized data as temporal input data for models. For instance, in the case of precipitation (P), both OKD and KEDD show satisfactory medians and interquartile range. Yet they exhibit CM values approaching 200 mm, highlighting the benefit of JW in such cases.

AUTHOR ACCEPTED MANUSCRIPT



(a) Precipitation, P [mm]

(b) Relative Humidity, RH [%]

(c) Wind Speed, WS [m/s]

(d) Solar Radiation, SR [Wh/m^2]

(e) Maximum air temperature, TMAX [°C]

(f) Minimum air temperature, TMIN [°C]

485 Figure 7. Boxplots of country averages Composite Metric (y-axis, in the unit of the
486 presented variable) for the different weather variables and the different SIMS tested
487 through the JW method for the case study (2010-2020), with TPS: thin spline interpolation
488 method, NATN: Natural Neighbours, IDW: Inverse Distance Weight, KEDS: kriging with

external drift using a static semi-variogram (computed from mean values per station for the entire study period), KEDD: kriging with external drift using a dynamic semi-variogram (computed from available values of different stations for the selected day), OKS: ordinary kriging using a static semi-variogram (computed from mean values per station for the entire study period), OKD: ordinary kriging using a dynamic semi-variogram (computed from available values of different stations for the selected day)), JW: the developed methodology (Judicious Widget for Geospatial Robust Interpolation Design)

4.2 Robustness analysis

4.2.1 Case study: station density effects

The first part of the robustness analysis focuses on the effect of station density through an occurrence-based approach on the case study. The percentage of each method selected for TMAX, TMIN and P for the four density sampling (25%, 50%, 75% and 100% of the weather stations randomly selected from those available on the considered day dataset) are presented in Figure 8.

Not all variables exhibit the same general trend concerning (1) the applied interpolations and (2) the changes in interpolation as the weather network density increases. No single interpolation method is universally preferable for all variables. For TMIN and TMAX, the frequency of selection of KEDD models increases with increasing station density. The frequency of use of KEDD for TMAX is also around twice that of TMIN regardless of the station density used. For temperature, it is logical that an external drift with altitude is preferred, given their direct relationship. Additionally, selection of the daily variogram suggests that spatial relationship for temperatures is more correlated on a daily basis

AUTHOR ACCEPTED MANUSCRIPT

than when using the whole period mean, signifying that it allows to take into account the varying relationship between temperature and altitude impacted by climate phenomena. Precipitation showed limited variability across the different station densities. OK methods are always the most selected, followed by IDW.

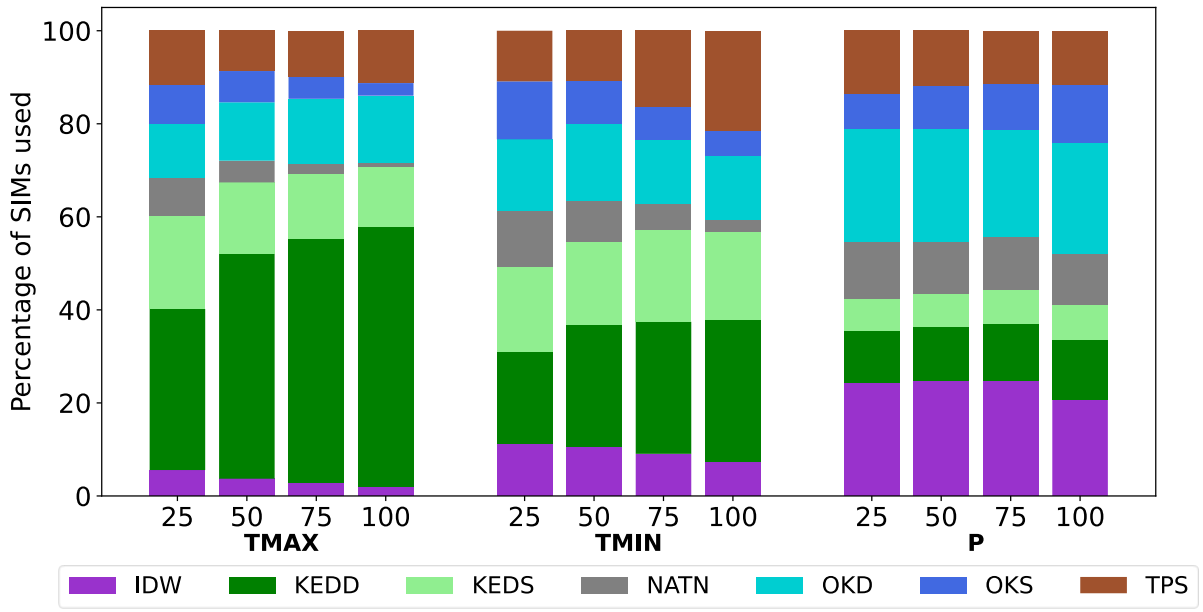


Figure 8. Interpolation methods selected by weather variable applied on different percentages of the dataset, with TMAX: maximum daily air temperature, TMIN: minimum daily air temperature and P: precipitation; TPS: thin spline interpolation method, NATN: Natural Neighbours, IDW: Inverse Distance Weight, KEDS: kriging with external drift using a static semi-variogram (computed from mean values per station for the entire study period), KEDD: kriging with external drift using a dynamic semi-variogram (computed from available values of different stations for the selected day), OKS: ordinary kriging using a static semi-variogram (computed from mean values per station for the entire study period), OKD: ordinary kriging using a dynamic semi-variogram (computed from available values of different stations for the selected day) .

AUTHOR ACCEPTED

The RMSEs slightly increase for both temperatures. Nevertheless, the difference between using 25% and 100% of the dataset is negligible: less than 0.05 °C for TMIN and of 0.01°C for TMAX (Figure 9). Regarding precipitation, the RMSE diminishes of 0.18mm. Focusing on Bias, the same analysis can be done. Those results, i.e. the stable error metrics and the varying frequency of model selections, seem to indicate the ability of the proposed methodology to handle the different variables for varying network densities. This allows to properly handle the evolution of network when reconstructing gridded data from time series. Values presented in Figure 9 are available in Appendix C.

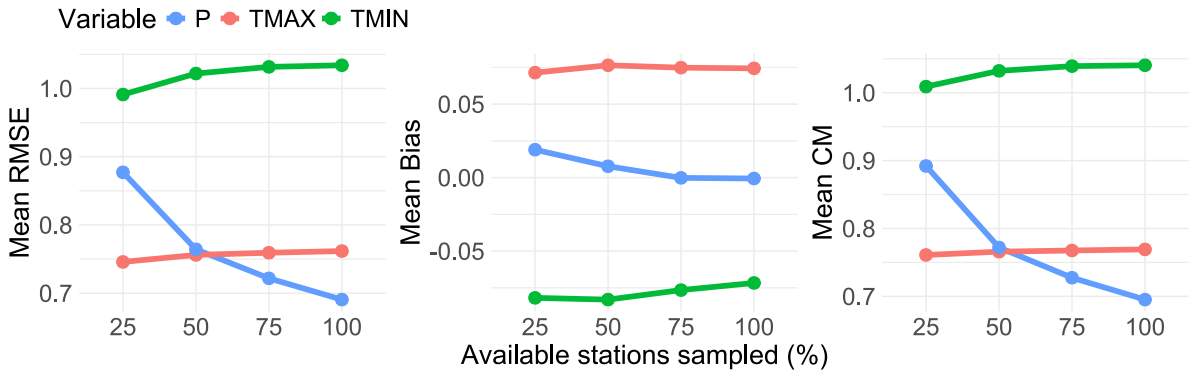
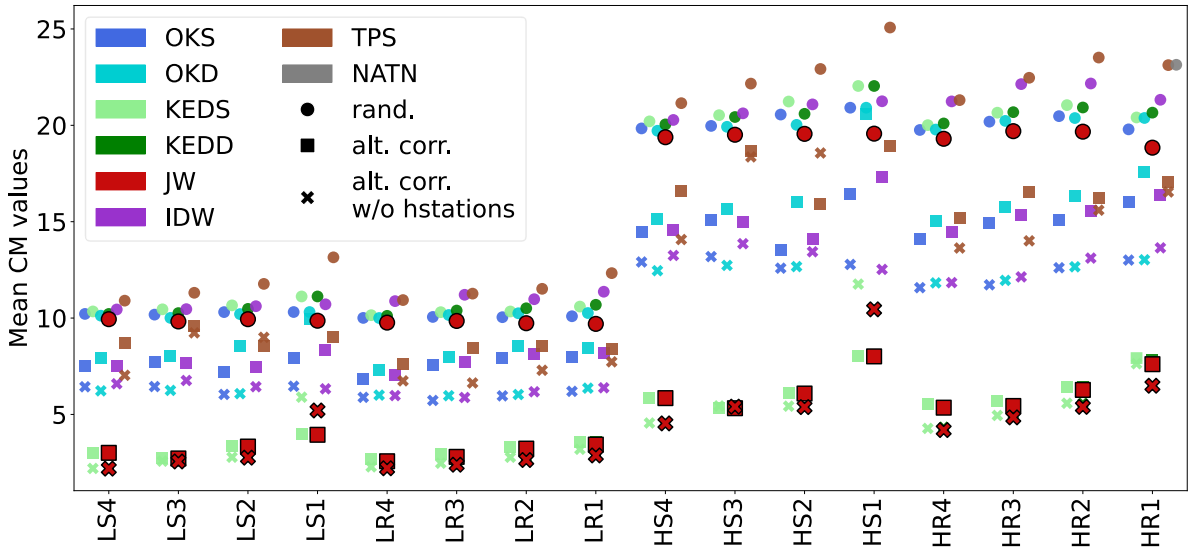


Figure 9. Evolution of the country mean RMSE and Bias for precipitation (P), maximum temperature (TMAX) and minimum temperature (TMIN) using subsets of 25, 50, 75 and 100 percent of the weather network for the 2010-2020 period.

4.2.2 Artificial cases study

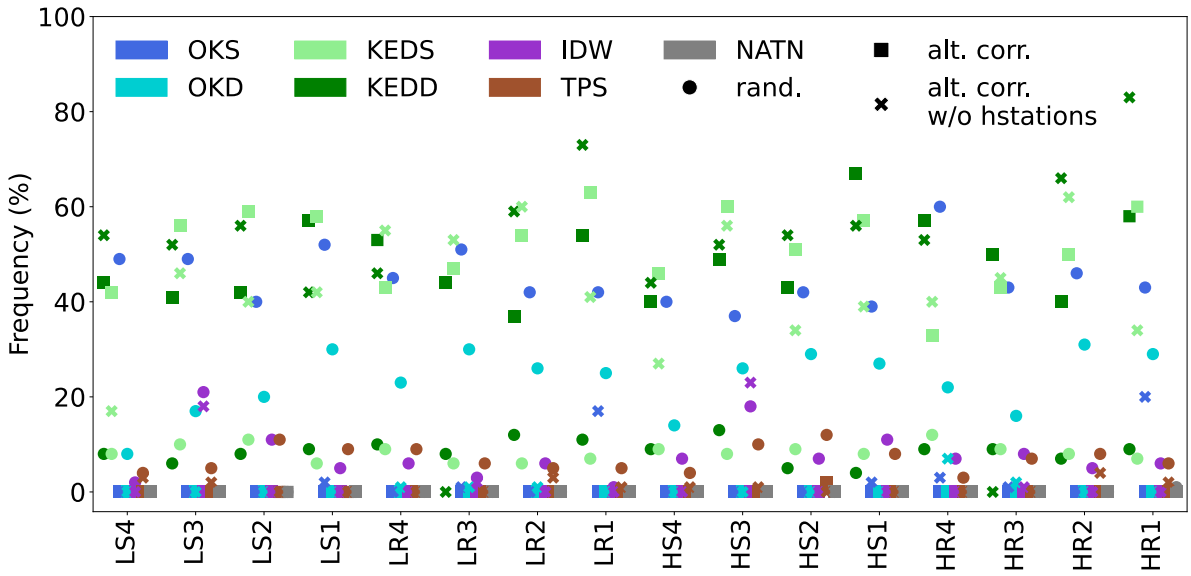
Figure *10* presents the mean CM values for specific cases. Firstly, The JW method consistently shows the lowest error or an error equal to that of the best-performing individual method for each configuration. Furthermore, it allows to diminish the interquartile range (Appendix E). This indicates that, across the diverse scenarios, the JW method successfully adapts to the data configuration and selects the most appropriate

544 interpolation technique, confirming its flexibility. Secondly, it can be seen that for most
545 cases the NATN is absent. While it performs fairly well for the case study dataset, the
546 extreme cases encountered here caused the method to fail. Thirdly, the different specific
547 cases studied show that when going from random spatial distribution to spatially
548 correlated one with altitude, the JW method is able to adapt by selecting the SIM
549 minimising the CM metric. Indeed, as shown in Figure *10* (b), the frequency of SIMs
550 selection changes. Finally, as already previewed in the case study, the JW ensured to
551 combine the best part of each SIMs, ensuring robust interpolated maps. More detailed
552 results on frequency selection and CM distribution are in Appendix E.



(a) Mean Composite Metric values error

AUTHOR MANUSCRIPT



(b) Frequency selection of SIMs

553 Figure 10. Response of the JW method to the developed artificial cases. Each point (dot,
554 square, cross) represents the mean of the composite metric (euclidian distance of RMSE
555 and bias) (a) or frequency (b) for each tested artificial case (x axis). With TPS: thin spline
556 interpolation method, NATN: Natural Neighbours, IDW: Inverse Distance Weight, KEDS:
557 kriging with external drift using a static semi-variogram (computed from mean values per
558 station for the entire study period), KEDD: kriging with external drift using a dynamic semi-
559 variogram (computed from available values of different stations for the selected day),
560 OKS: ordinary kriging using a static semi-variogram (computed from mean values per
561 station for the entire study period), OKD: ordinary kriging using a dynamic semi-variogram
562 (computed from available values of different stations for the selected day),), JW: the
563 developed methodology (Judicious Widget for Geospatial Robust Interpolation Design),
564 artificial cases are labelled using a code where the first letter indicates the standard
565 deviation of input values: 'L' for low and 'H' for high. The second letter denotes the network
566 configuration: 'S' for systematic and 'R' for random. The accompanying number
567 represents the number of networks being considered. Each configuration is applied to

AUTHOR ACCEPTED MANUSCRIPT

three scenarios: (i) randomly distributed points (rand.), (ii) spatially correlated points including high-altitude stations (alt. corr.), and (iii) spatially correlated points excluding high-altitude stations (alt. corr. w/o hstations), dots representing random distribution of the weather network, square representing correlated distribution of the weather station network with altitude and crosses correlated distribution of the network with altitude and high station location being removed. The dots, squares, and crosses outlined in black represent the JW method.

4.2.3 Mountainous case study

While applied to the mountainous case, the JW method was able to reduce the value of the composite metric. This independent test region demonstrates that the methodology preserves its characteristic behaviour, namely maintaining the lowest combined RMSE and bias values compared to other interpolation methods (Figure 12). Moreover, the JW method mitigates extreme cases in which the error could not be adequately addressed by a single interpolation technique, yielding extremes that are less distant from the median distribution and a reduced interquartile range. The frequency of selection for the different SIMS is also pronounced: OKS was selected 102 times, IDW 121 times, KEDS 33 times, NATN 55 times and TPS 17 times. These results may indicate JWgrid's ability to adapt in selecting the most appropriate SIM. Indeed, Gottardi et al. (2012) highlights that the spatial structure of precipitation in the French Alps varies according to atmospheric flow. Overall, this analysis suggests that the JW approach not only improves overall interpolation accuracy but also provides a robust framework for handling challenging spatial variability in heterogeneous environments. Additionnal information on RMSE and boxplot are in appendix F.

AUTHOR ACCEPTED MANUSCRIPT

591 Regarding computational performance, the script required approximately 15.6 hours for
592 single-core processing on a 1 km grid and 4.9 hours for the 5 km grid. Computations were
593 performed on a workstation with an Intel® i9-9900X (3.5 GHz, 10 cores) and 128 GB RAM.
594 While this is substantial, the study area was particularly challenging and there are
595 possibilities of parallel processing. Comparison to previous works on the study area
596 confirms patterns similarities (e.g. Lee et al., 2024; Mazzoglio et al., 2020).

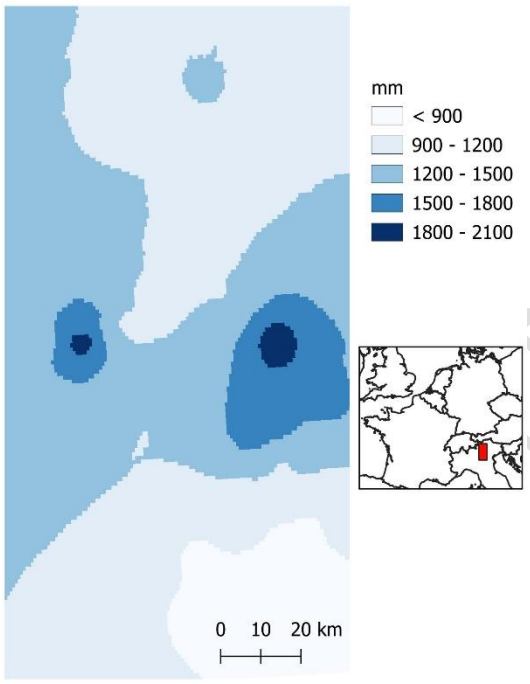


Figure 11 : Interpolated precipitation map generated by the JW method on a 1 km^2 grid, showing annual cumulated precipitations (mm) for the year 2020 in the mountainous case study area.

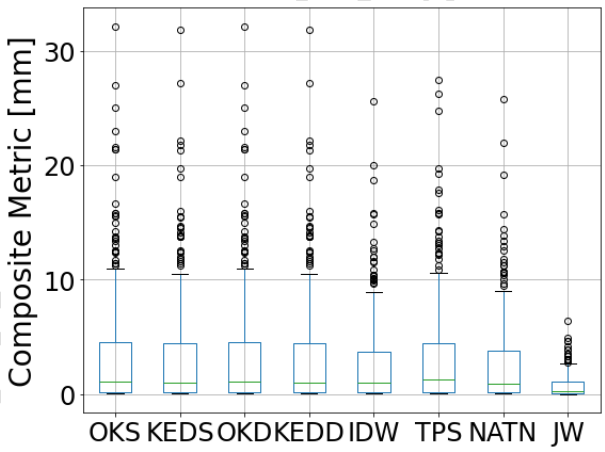


Figure 12. Boxplots of the composite metric for the mountainous case for one year of precipitation dataset, with TPS: thin spline interpolation method, NATN: Natural Neighbours, IDW: Inverse Distance Weight, KEDS: kriging with external drift using a static semi-variogram (computed from mean values per station for the entire study period), KEDD: kriging with external drift using a dynamic semi-variogram (computed

from available values of different stations for the selected day), OKS: ordinary kriging using a static semi-variogram (computed from mean values per station for the entire study period), OKD: ordinary kriging using a dynamic semi-variogram (computed from available values of different stations for the selected day),), JW: the developed methodology (Judicious Widget for Geospatial Robust Interpolation Design)

4.3 Comparison with existing automation frameworks

While comparing SIMs application is abundant in the literature (e.g. Hofstra et al., 2008; Luo et al., 2008; Ly et al., 2011; Rosillon et al., 2024), only few studies encountered the automation of SIMs parametrization (i.e., defining or optimizing the parameters of interpolation models) or selection (Table 5). Existing automation frameworks for spatial interpolation generally fall into two categories: (i) parameter tuning within a single geostatistical method, and (ii) parameters tuning alongside method selection among a limited set of SIMs. Firstly, Pardo-Igúzquiza et al. 2005) and Hiemstra et al. (2009) automatically fit variogram parameters for universal kriging, optimizing the variogram model using the Akaike Information Criterion (AIC) and RMSE. Secondly, Betzek et al. (2019) used the Interpolation Selection Index (ISI) proposed by Bier and de Souza (2017)

AUTHOR ACCEPTED MANUSCRIPT

609 to choose between IDW and OK and tune their parameters. Building on this, Sobjak et al.
610 (2023) introduced supplementary criteria to ensure spatial dependence in
611 semivariograms during the interpolator selection process. In contrast, the JWgrid
612 framework integrates a broader set of deterministic and geostatistical SIMs, considers
613 external drift, and employs the proposed composite metric (CM). This design enables
614 JWgrid to adapt dynamically to changing network densities and spatial dependence,
615 offering a versatile solution for environmental applications.

616 Table 5. Comparison of JWgrid against existing automation frameworks for spatial
617 interpolation automation

Study	Automation scope	SIMs	Objective function	Context
JWgrid	Method + parameter selection	IDW, Natural Neighbours, TPS, OK (S/D), KED (S/D)	CM	High frequency weather data
Pardo-Igúzquiza et al. (2005)	Parameter tuning	KED (UK)	AIC	Rainfall data
Hiemstra et al. (2009)	Parameter tuning	KED (UK)	Variogram fit/RMSE	Real-time radiation mapping
Betzek et al. (2019)	Method + parameter selection	IDW, OK	ISI	Agricultural management zones
Sobjak et al. (2023)	Method selection	IDW, OK	ISI/%ESDI/%γ/ %SMEI	Agricultural management zones

618

619 4.4 Limitations and perspectives

620 This study aimed to present an approach for automating the interpolation of climate data
621 based on the selection of SIMs. Overall, the comparison of the errors boxplots indicates
622 that the JWgrid method reduces extreme errors values. Rosillon et al. (2024) note that

AUTHOR ACCEPTED MANUSCRIPT

623 although a KED model was selected for temperature due to lower prediction errors,
624 substantial inaccuracies can occur during thermal inversions, an issue that JWgrid could
625 potentially mitigate. Moreover, the selection frequencies reveal logical trends in model
626 choice (e.g., KED for temperature, OK and IDW for precipitation). However, kriging models
627 may sometimes be chosen even when no real spatial dependency exists. To address this,
628 Sobjet et al. (2023) proposed indices to quantify spatial dependence and thresholds for
629 the application of such models. Refining selection metrics could be an avenue.

630 Another limitation is the reliance on LOOCV. While widely used, this technique increases
631 computational demands and becomes impractical for large datasets. Implementing k-
632 fold cross-validation could streamline the process without compromising accuracy.
633 Additionally, the usage of multiple software packages (e.g., WhiteboxTools) for
634 interpolation introduces workflow disruptions and inconsistencies due to updates. A
635 standardized framework or unified toolset would enhance reproducibility and stability
636 across environments.

637 Both the case study and the specific cases show the added value of using relevant
638 auxiliary data (e.g. elevation). The use of climate models outputs, radar data, or even
639 remote sensing products could enhance the results. Such data could be integrated
640 through the external drift in the KEDD (e.g. Rosillon et al., 2024; Sideris et al., 2014;
641 Viggiano et al., 2019). Equally, other further avenues could be explored. Inclusion of
642 crowd-sourced observations as input data could enhance the spatial and temporal
643 resolution of current networks (Gharesifard et al., 2017). For example, around 850
644 farmers' stations are estimated in Wallonia (Rosillon et al., 2024), or the French semi-
645 professional StatIC network that accounts for 880 working stations (Garcelon et al.,

AUTHOR ACCEPTED MANUSCRIPT

2023). Nevertheless, such inclusions would need robust automatic quality control (e.g. Dandrifosse et al. 2024). Such inclusions could make the framework even more versatile. Another perspective is the reconstruction of missing time series at weather stations. Since JWgrid produces daily gridded estimates, these values could be used as first-guess values for the missing data. Finally, although the JW methodology is applied to weather data, it could be transferable to other types of spatial data, such as soil mapping (Robinson and Metternicht, 2006), streamflow distribution (Eslamian et al., 2023) or even atmospheric components.

5. Conclusion

The use of regional gridded datasets is of uttermost importance for environmental applications but are not always available. This study showed that meteorological data can be accurately interpolated using the proposed JW framework. This novel approach allows a straightforward interpretation and effective comparison of different interpolation methods by focusing on random and systematic errors. It effectively manages different variables and varying network structures and densities while relying on a simple performance metric. Regarding the case study, results showed extremely low overall random and systematic errors relative to their range of values and allowed to produce maps of several weather variables using the same methodology. Finally, the increase of the weather network density assessed for P, TMAX and TMIN did not show changes in overall bias and RMSEs showing that the methodology correctly handles the variation in spatial and temporal data variations while the SIMs' selection occurrences changed. The simulations with the synthetic datasets proved the JW method' flexibility, while the mountainous case underlined its transferability even to regions characterized by a

AUTHOR ACCEPTED MANUSCRIPT

669 remarkably higher degree of topographical complexity and consequently a greater
670 variability in meteorological variables. Hence, this paper demonstrated that no method is
671 universally preferred and is context dependent, and that the JW method can, with minimal
672 user labor, select the optimal interpolation scheme. Given the limited amount of data
673 required to apply the methodology framework it seems easily replicable to other regions.
674 Hence, combined to the ease of use, the methodology has a large potential to improve
675 spatial distribution of weather variables in a context of changing weather network. The
676 method is well-suited for generating gridded environmental data to support management,
677 decision making and modelling. Further avenues, such as the inclusion of crowd-sourced
678 observations, the inclusion of other explanatory variables or the testing of other selection
679 metrics, seem promising.

680 6. Acknowledgment

681 The authors are grateful to the meteorological services of the “Administration des services
682 techniques de l’agriculture” and of the “Administration de la gestion de l’eau” for the
683 sharing of data. Also, they are grateful to the CRA-W, to Météo France and to the DWD for
684 the sharing of their meteorological data.

685 7. CRediT authorship contribution statement

686 **Cédric Magain**: Writing – original draft, Methodology, Data curation, Conceptualization.
687 **Guillaume Renard**: Writing – original draft, Methodology, Data curation,
688 Conceptualization. **Philippe Orban**: Writing – review and editing, Funding acquisition.
689 **Aurore Degré**: Writing – review and editing, Methodology. **Jeroen Meersmans**: Writing –
690 review and editing, Methodology. **Caroline De Clerck**: Writing – review and editing. **Serge**

AUTHOR ACCEPTED MANUSCRIPT

Brouyère: Writing – review and editing. **Joost Wellens**: Writing – review and editing, Funding acquisition, supervising, Project administration.

8. Funding

This project was funded by the *fond de la gestion de l'eau* of the *Administration de la gestion de l'eau* (AGE).

9. Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

10. Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author used ChatGPT 5.0 (OpenAI, USA) in order to check grammar, paraphrase and translate. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

AUTHOR ACCEPTED MANUSCRIPT

11. References

La mise à jour automatique des citations est désactivée. Pour voir la bibliographie, cliquez sur Actualiser dans l'onglet Zotero.

AUTHOR ACCEPTED MANUSCRIPT

12. Appendices

A. Weather stations

Table A.1. Latitude, longitude, elevation and identifiers of the stations constituting the tested dataset

					Observed variables						
Station name	Latitude	Longitude	Elevation	ID	P	RH	RS	TMAX	TMIN	WS	TOTAL
Amberloup	5.551	50.036	410	0	1	1	1	1	1	1	6
Arsdorf	5.849	49.859	412	1	1	1		1	1		4
Asselborn	5.970	50.097	479	2	1	1		1	1		4
Bergeval	5.869	50.352	411	3	1	1	1	1	1	1	6
Beringen	6.112	49.762	219	4	1	1		1	1		4
Bettendorf	6.209	49.874	187	5	1	1		1	1		4
Bigonville	5.801	49.870	331	6	1						1
Bleesbreck	6.191	49.871	191	7	1	1		1	1		4
Bollendorf	6.359	49.851	168	8	1						1
Breidfeld	6.066	50.124	440	9	1	1		1	1		4
Chassepierre	5.259	49.699	347	10	1	1	1	1	1	1	6
Clemency	5.875	49.599	334	11	1	1		1	1		4
Dahl	5.981	49.936	480	12	1	1	1	1	1		5
Dellen	5.956	49.861	489	13	1						1
Diekirch	6.151	49.864	192	14	1	1	1	1	1		5
Doncourt-les-Conflans	5.929	49.153	235	15	1	1		1	1		4
Echternach	6.443	49.803	243	16	1	1	1	1	1	1	6
Elsenborn	6.244	50.444	625	17	1	1	1	1	1	1	6
Emmels	6.085	50.305	536	18	1	1	1	1	1	1	6
Eschdorf	5.936	49.880	512	19	1	1	1	1	1	1	6
Esch-sur-Sure	5.921	49.909	339	20	1	1		1	1		4
Ettelbruck	6.097	49.852	202	21	1	1		1	1		4
Fouhren	6.195	49.914	300	22	1	1		1	1		4
Godbrange	6.236	49.736	328	23	1	1		1	1		4
Grevenmacher	6.435	49.681	191	24	1	1	1	1	1	1	6
Harlange	5.799	49.927	479	25	1	1	1	1	1		5
Heinerscheid	6.090	50.094	526	26	1			1	1		3
Holler	6.047	50.122	464	27	1	1		1	1		4
Holtz	5.796	49.805	465	28	1			1	1		3
Hosingen	6.101	49.993	510	29	1	1	1	1	1		5
Huldange	6.024	50.162	560	30	1	1	1	1	1		5
Jemelle	5.243	50.166	234	31	1	1	1	1	1	1	6
Kautenbach	6.025	49.947	261	32	1			1	1		3
Koerich (AGE)	5.945	49.679	272	33	1						1

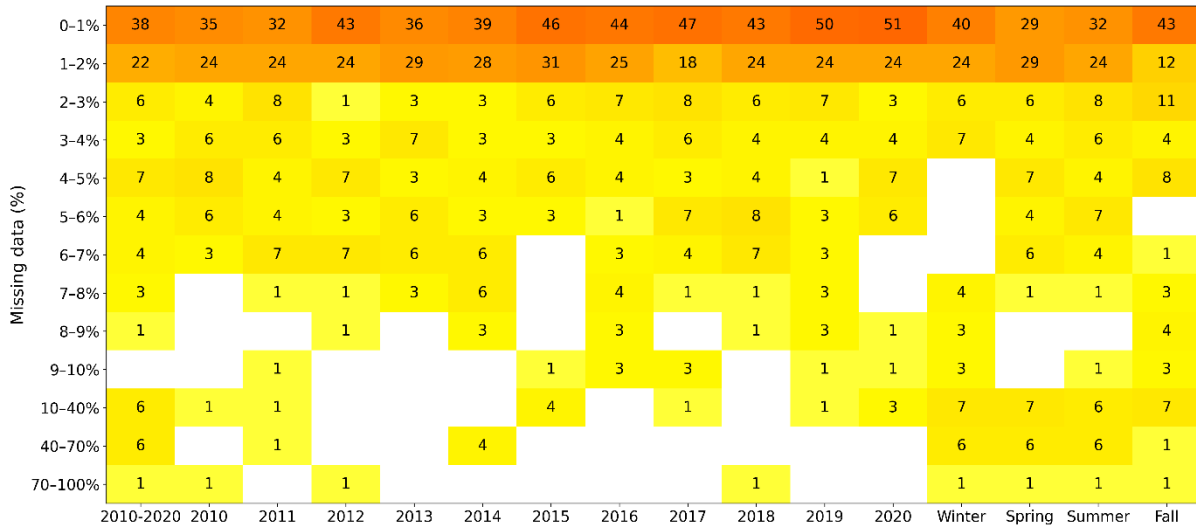
AUTHOR ACCEPTED MANUSCRIPT

Koerich (ASTA)	5.945	49.679	272	34	1	1		1	1		4
Libramont	5.361	49.927	472	35	1	1	1	1	1	1	6
Livange	6.115	49.526	269	36	1			1	1		3
Mamer	6.019	49.633	315	37	1	1		1	1		4
Merl	6.092	49.613	307	38	1	1	1	1	1	1	6
Mersch	6.116	49.753	221	39	1	1	1	1	1		5
Metz-Frescaty	6.125	49.069	192	40	1	1		1	1		4
Michamps	5.808	50.038	498	41	1	1	1	1	1	1	6
Obercorn	5.901	49.512	377	42	1	1	1	1	1	1	6
Olsdorf	6.382	49.944	321	43	1	1		1	1		4
Perl-Nennig	6.379	49.535	149	44	1	1		1	1		4
Potaschberg	6.394	49.678	261	45	1			1	1		3
Reckange	6.006	49.562	296	46	1	1		1	1		4
Reichlange	5.926	49.774	258	47	1			1	1		3
Remerschen	6.349	49.491	187	48	1	1		1	1		4
Remich	6.355	49.545	206	49	1	1	1	1	1	1	6
Reuler	6.037	50.063	489	50	1	1	1	1	1	1	6
Roeser	6.142	49.540	265	51	1	1		1	1		4
Roodt	5.820	49.794	441	52	1	1		1	1		4
Ruette	5.577	49.539	315	53	1	1	1	1	1	1	6
Schimpach	5.847	50.009	390	54	1	1	1	1	1	1	6
Schneifelforsthaus	6.419	50.297	657	55	1	1		1	1		4
Schockville	5.763	49.760	349	56	1	1	1	1	1	1	6
Seingbouse	6.831	49.128	345	57	1	1	1	1	1		5
Stadtbredimus	6.359	49.579	222	58	1	1		1	1		4
Steinsel	6.144	49.694	240	59	1	1		1	1		4
Trier-Petrisberg	6.658	49.748	261	60	1	1	1	1	1	1	6
Trier-Zewen	6.613	49.733	133	61	1	1		1	1		4
Trintange	6.280	49.575	200	62	1	1		1	1		4
Useldange	5.967	49.767	281	63	1	1	1	1	1	1	6
Vianden	6.205	49.939	213	64	1						1
Villette	5.578	49.482	344	65	1	1		1	1		4
Waldbillig	6.277	49.798	314	66	1	1	1	1	1	1	6
Walferdange	6.130	49.669	228	67	1						1
Weiskirchen-Saar	6.812	49.553	382	68	1	1		1	1		4
Welscheid	6.048	49.883	264	69	1						1
Wincrange (AGE)	5.926	50.061	501	70	1						1
Wincrange (ASTA)	5.926	50.061	501	71	1	1		1	1		4
TOTAL					72	58	28	64	64	21	

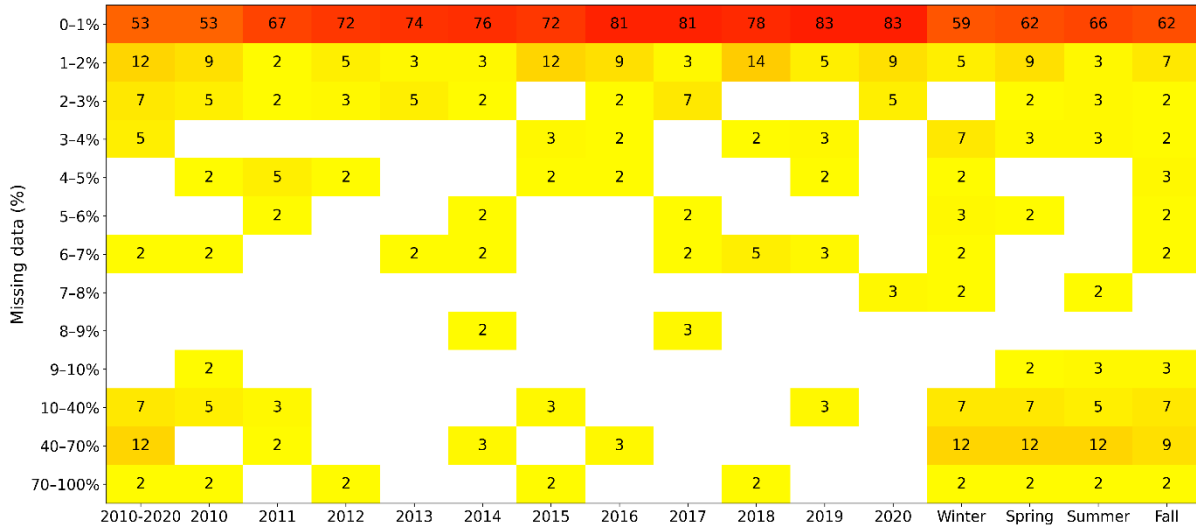
713

714

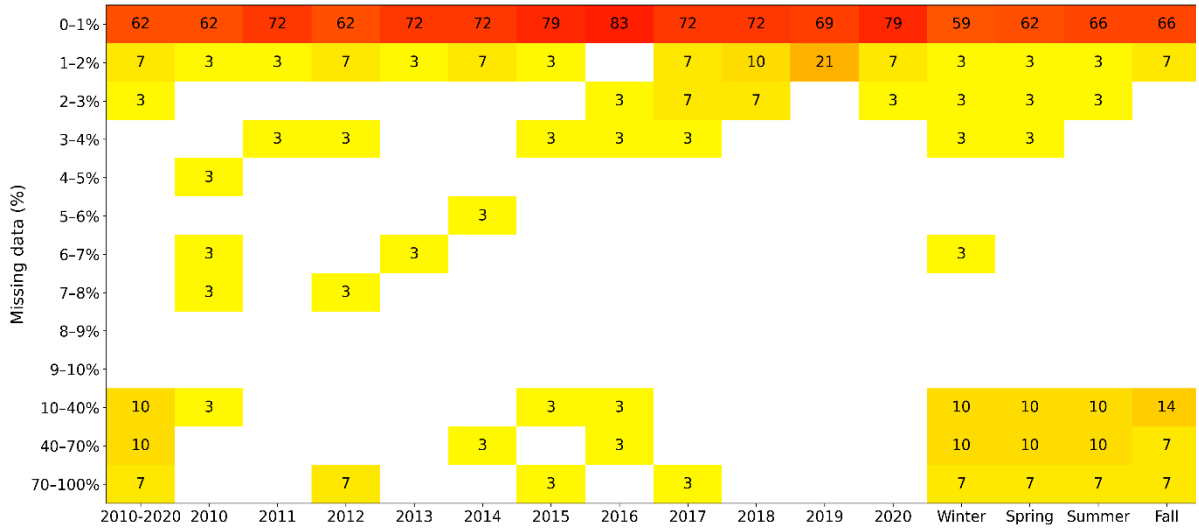
AUTHOR ACCEPTED MANUSCRIPT



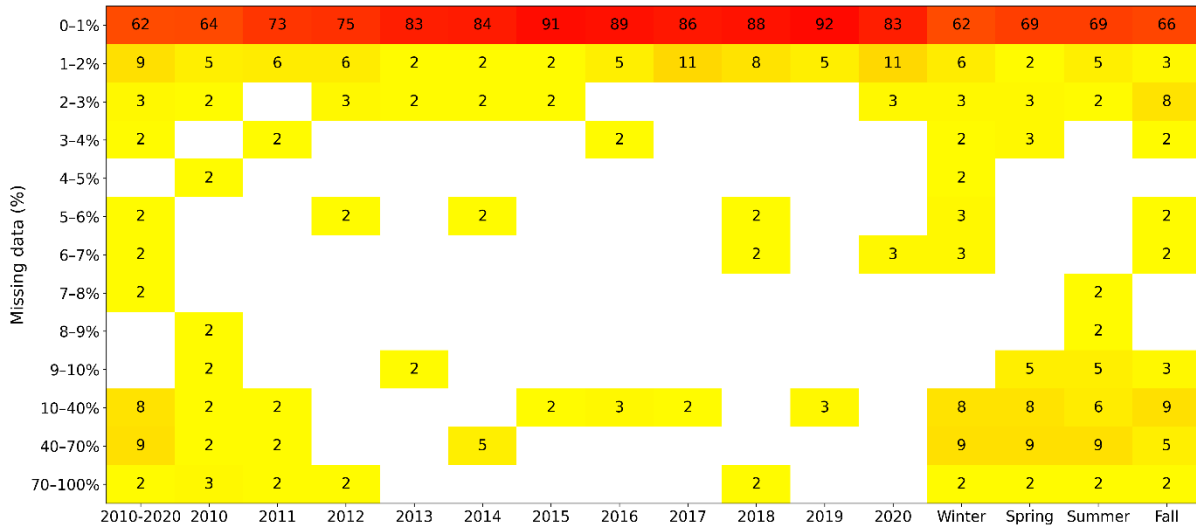
(a) Precipitation



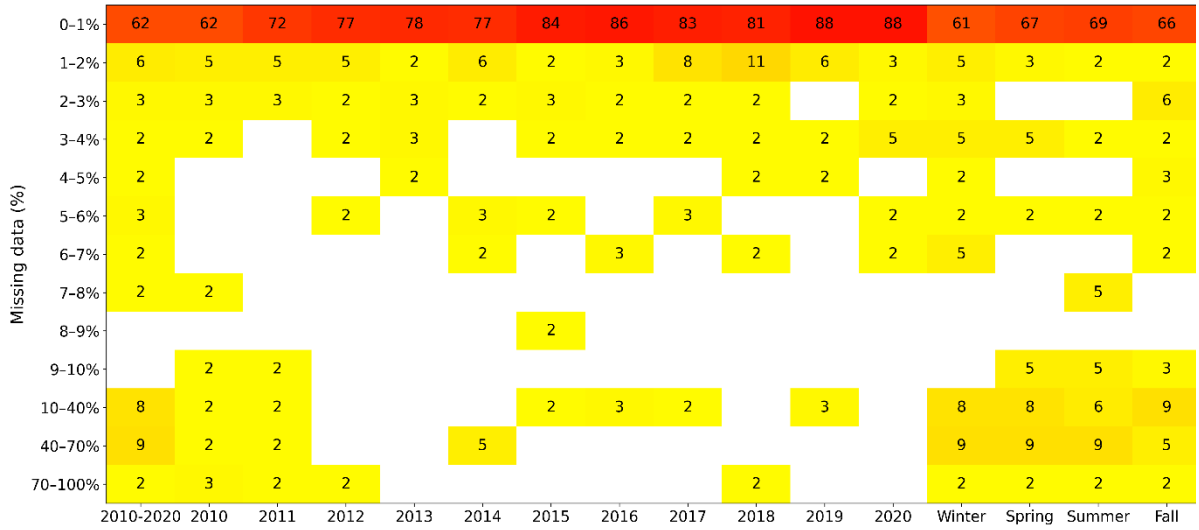
(b) Relative humidity



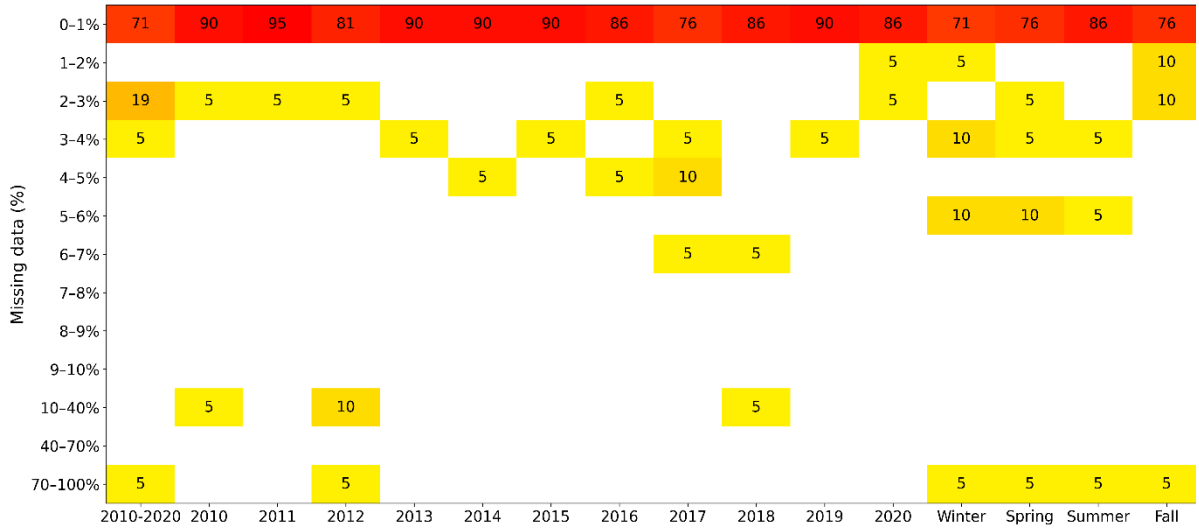
(c) Solar radiation



(d) Maximum air temperature



(e) Minimum air temperature



(f) Wind speed

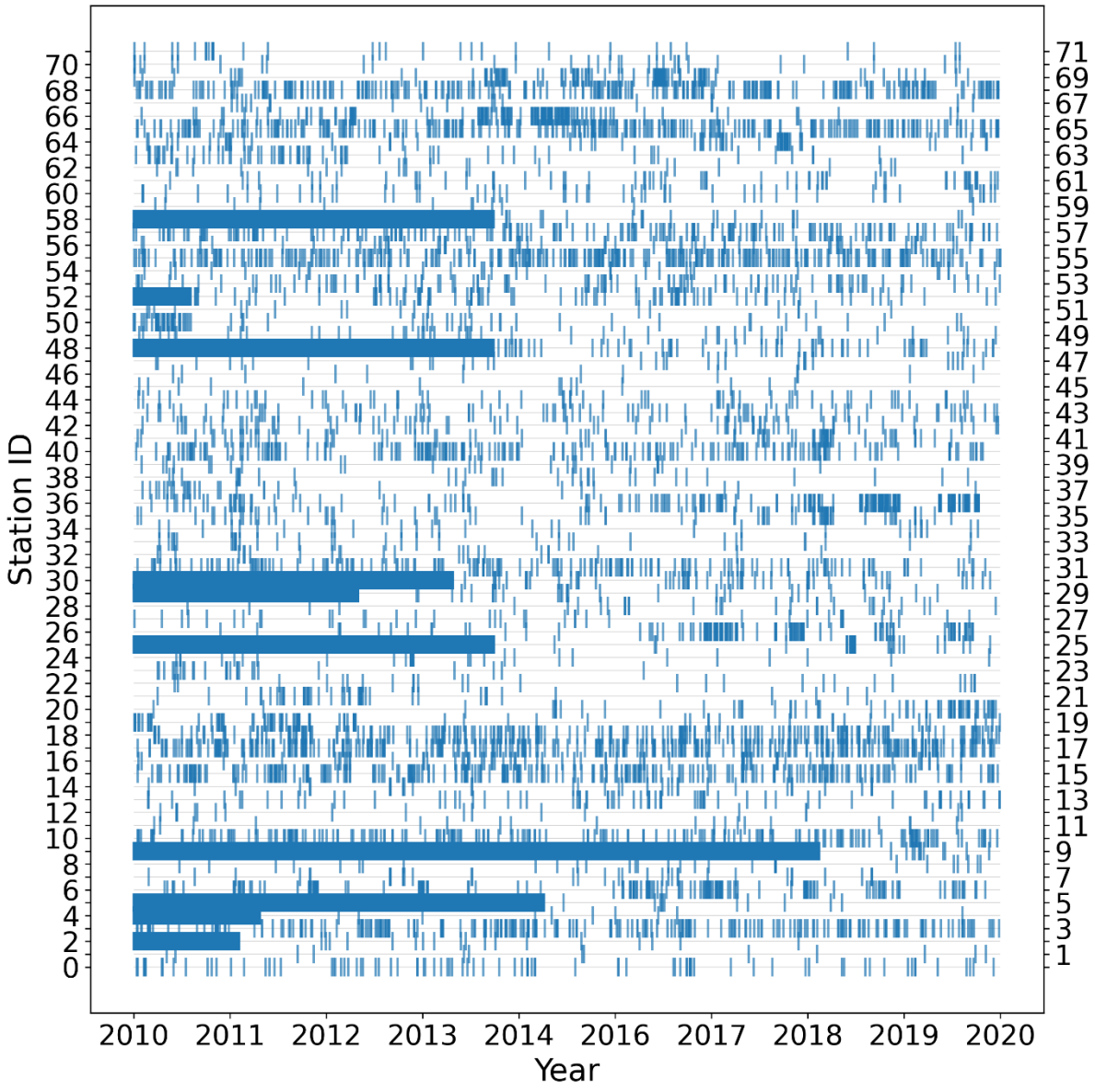
715 Figure A.2. Heatmaps of the percentage of missing stations classified by percentage of

716 missing data, with y axis indicating the percentage of missing data, x axis the timestep

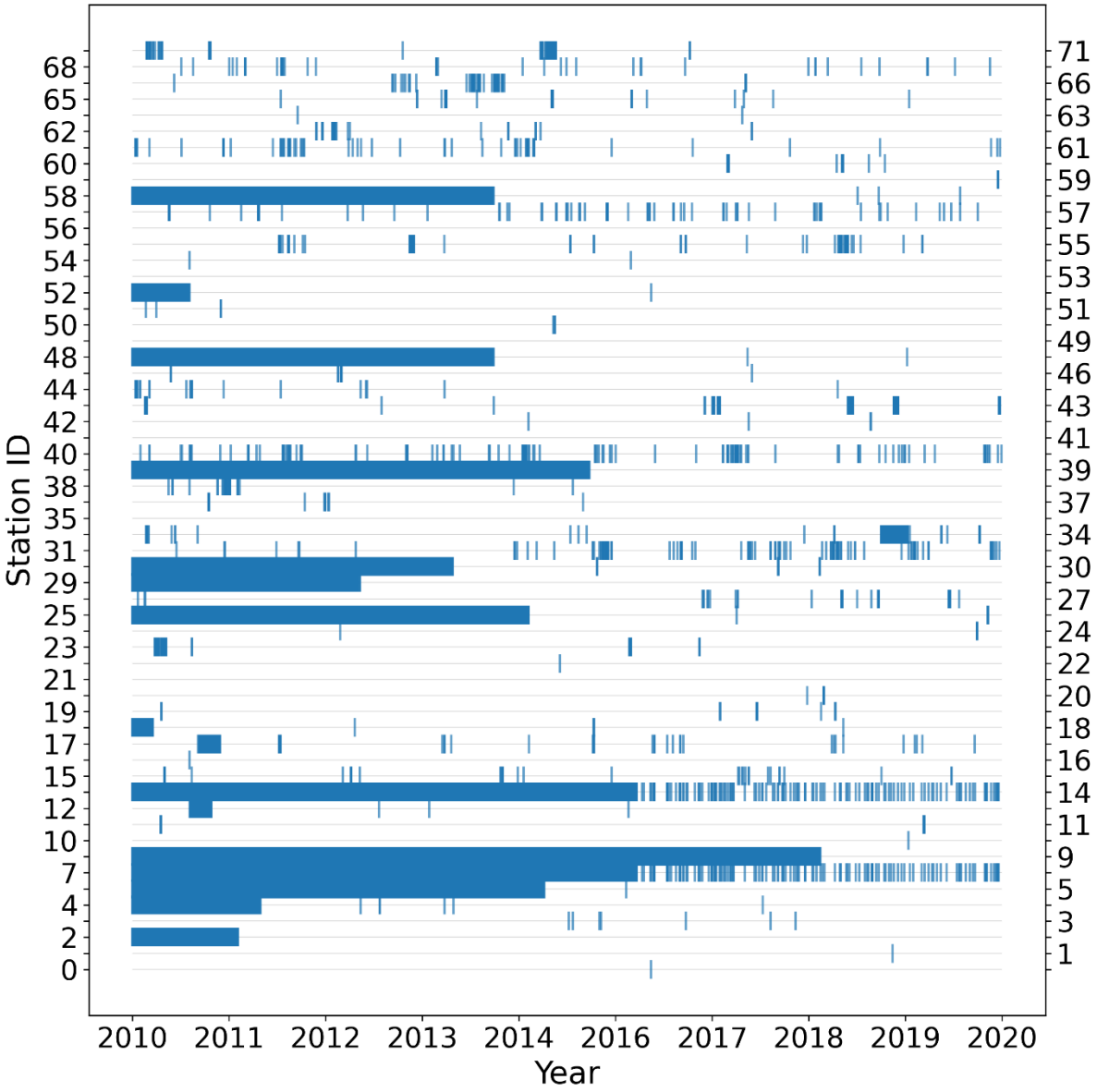
717 considered. The redder the value is, the highest the percentage of missing station for the

718 given percentage of missing station is.

719

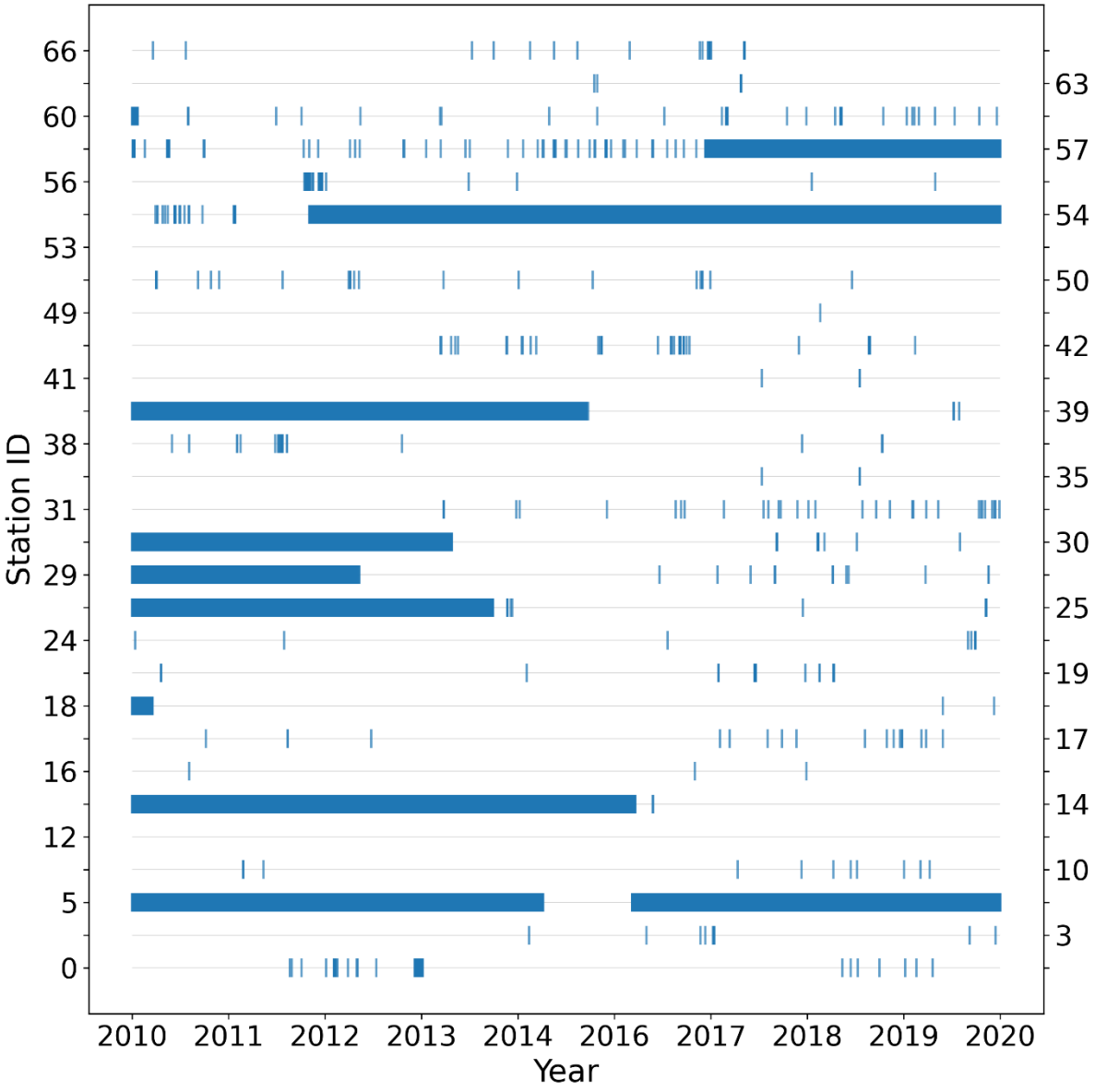


(a) Precipitation



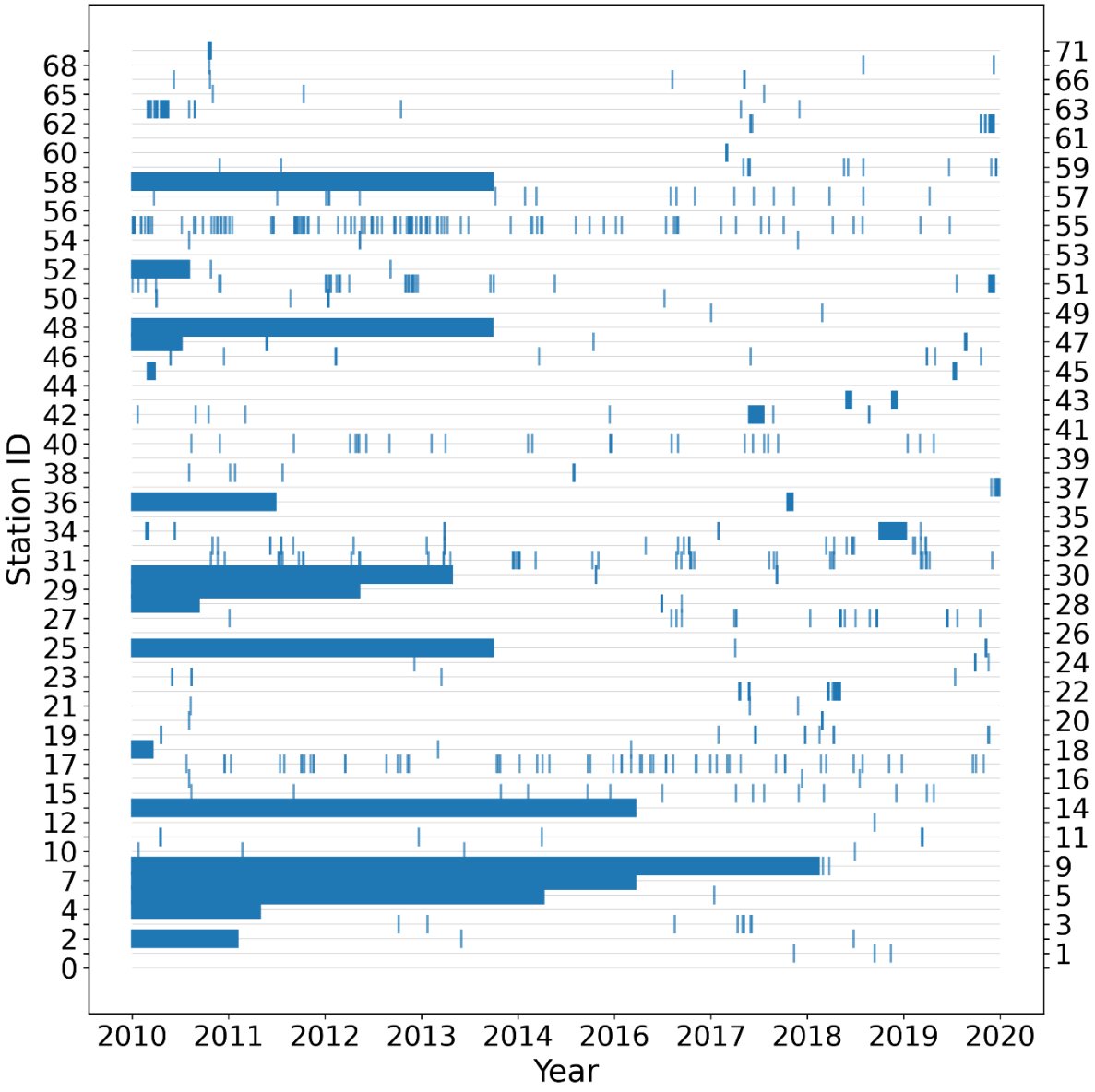
(b) Relative air humidity

AUTHOR



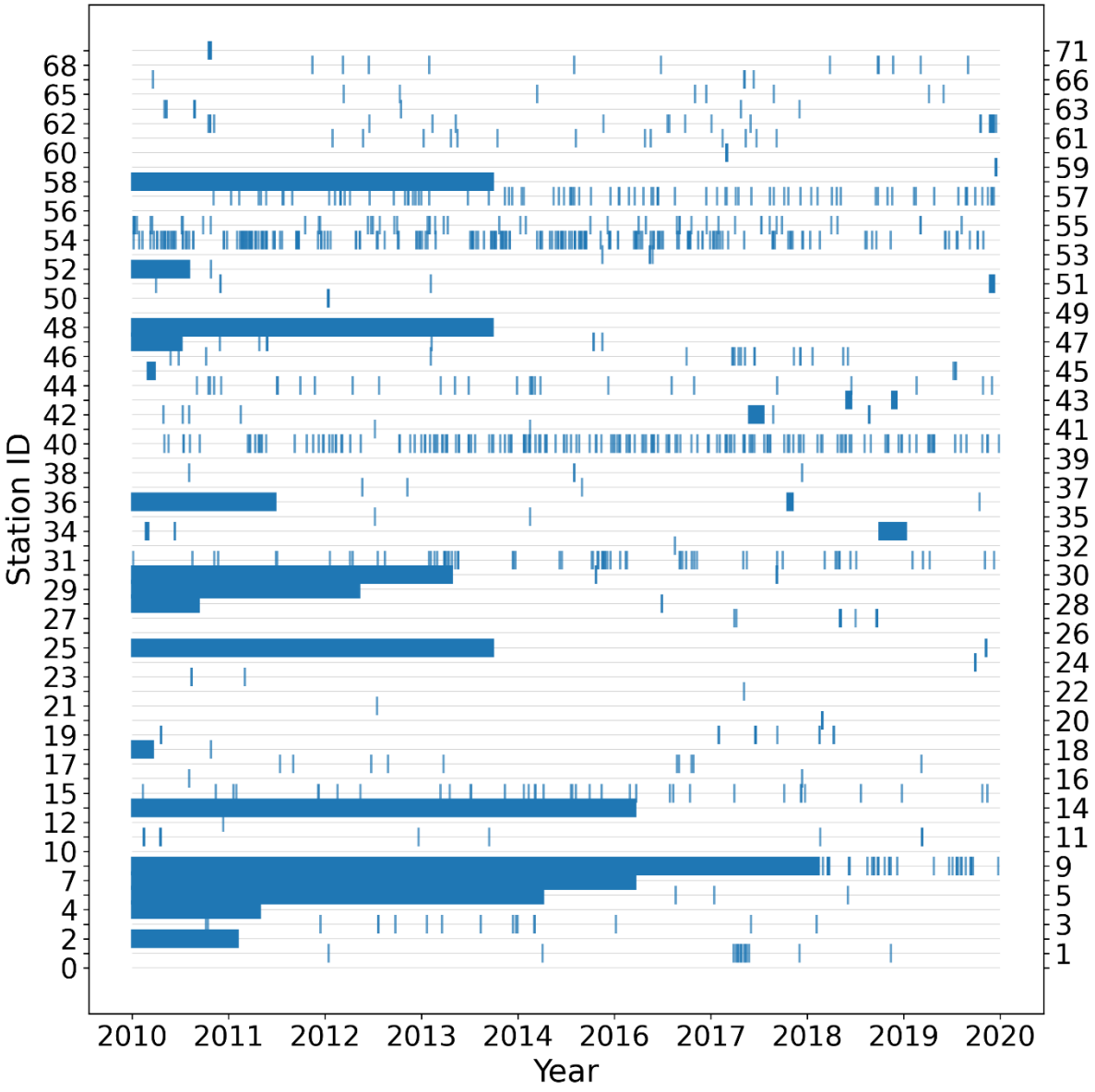
(c) Solar radiation

AUTHOR



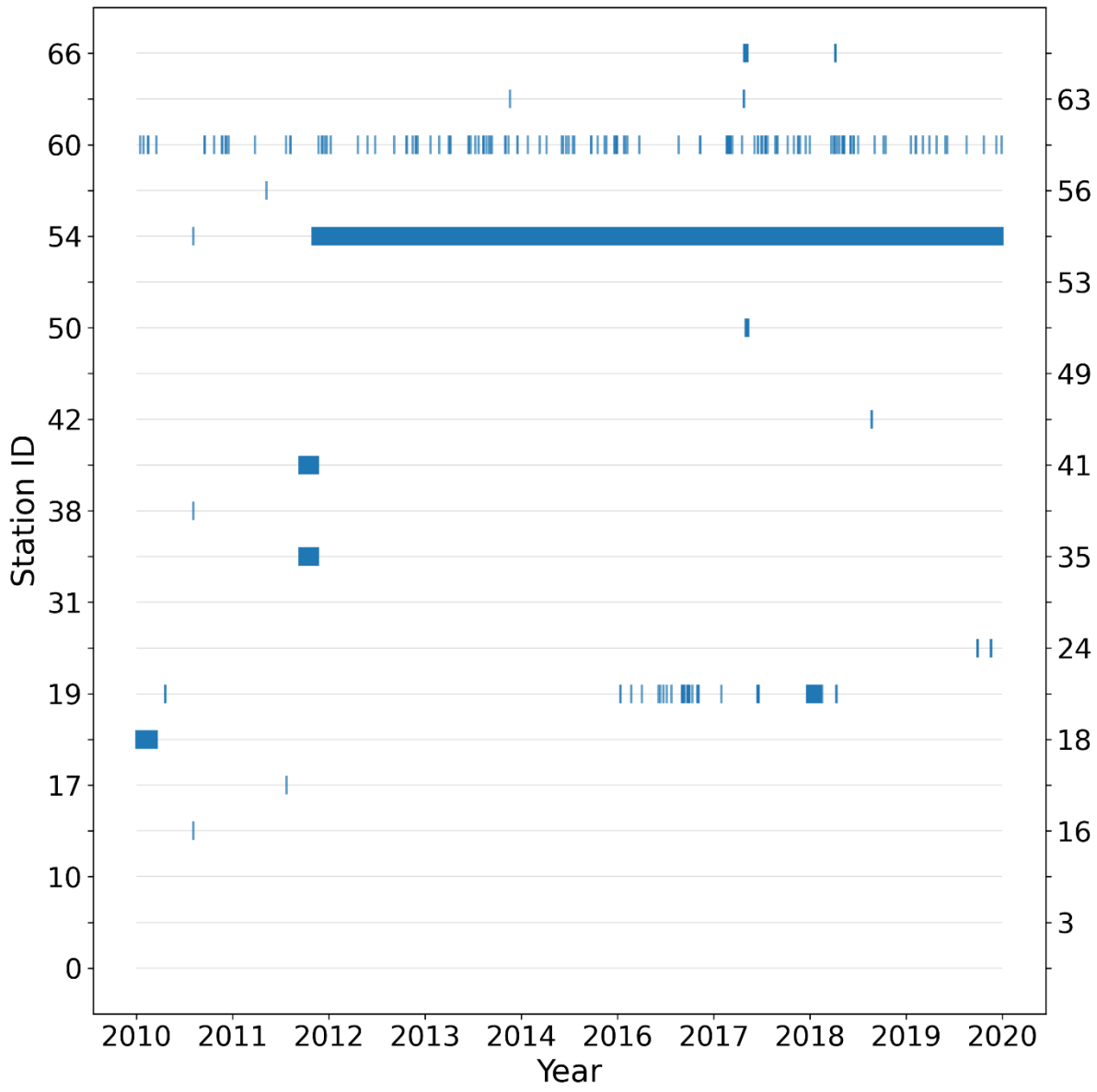
(d) Maximum air temperature

AUTHOR



(e) Minimum air temperature

AUTHOR



(f) Wind speed

720 Figure A.3. Temporal visualisation of missing data per station. The blue dot indicates when
721 the variables is not recorded. Only the station recording the considered variables are
722 plotted.

723

AUTHOR

B. Relationship of mean station value with altitude of the synthetic dataset

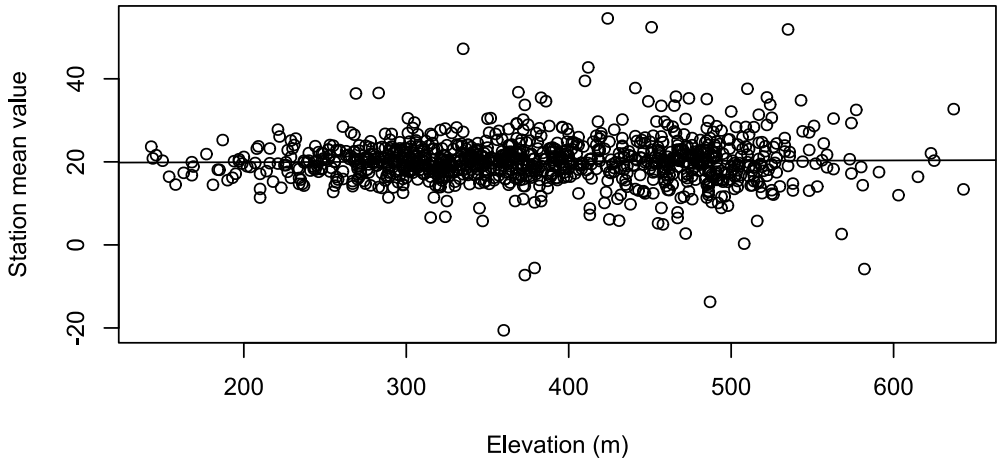
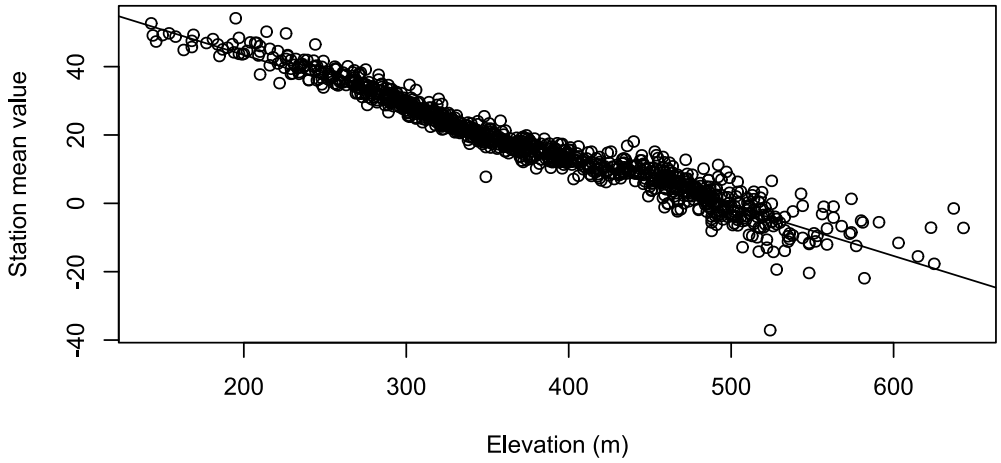


Figure B.1. Artificial case stations mean values function of elevation values for the dataset with the geographical driver “random values”.

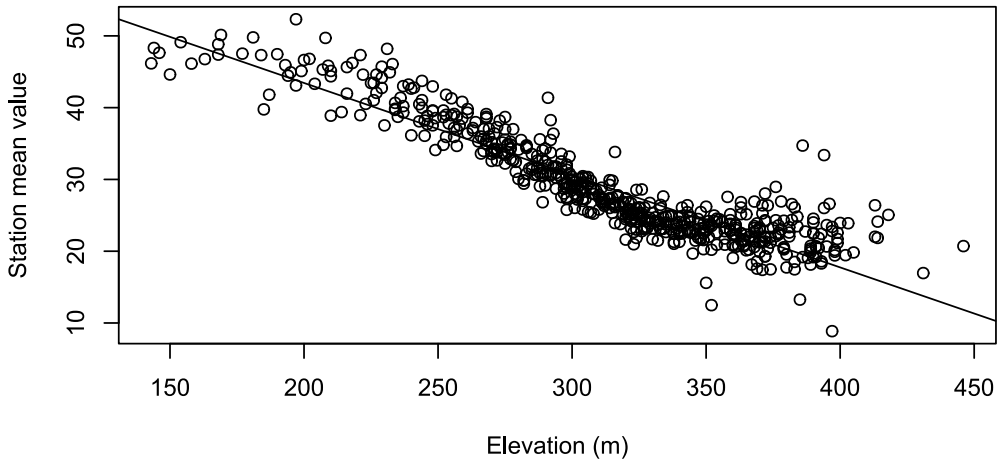
AUTHOR ACCEPTED



729

730 Figure B.2. Artificial case stations mean values function of elevation values for the dataset

731 with the geographical driver “altitude correlation”.



732

733 Figure B.3. Artificial case stations mean values function of elevation values for the dataset

734 with the geographical driver “altitude correlation without high stations”.

C.Accuracy metrics and variable network density tables

Tables C.1, C.2 and C.3 respectively present the RMSE, Bias and Composite metric values four the four networks densities studied and the three variables.

Table C.1. RMSE country mean values for four network density sampling for P, TMAX and TMIN for the 2010-2020 period

Density	P	TMAX	TMIN
25	0.877	0.746	0.991
50	0.764	0.756	1.022
75	0.722	0.759	1.032
100	0.691	0.762	1.034

Table C.2. Bias country mean values for four network density sampling for P, TMAX and TMIN for the 2010-2020 period

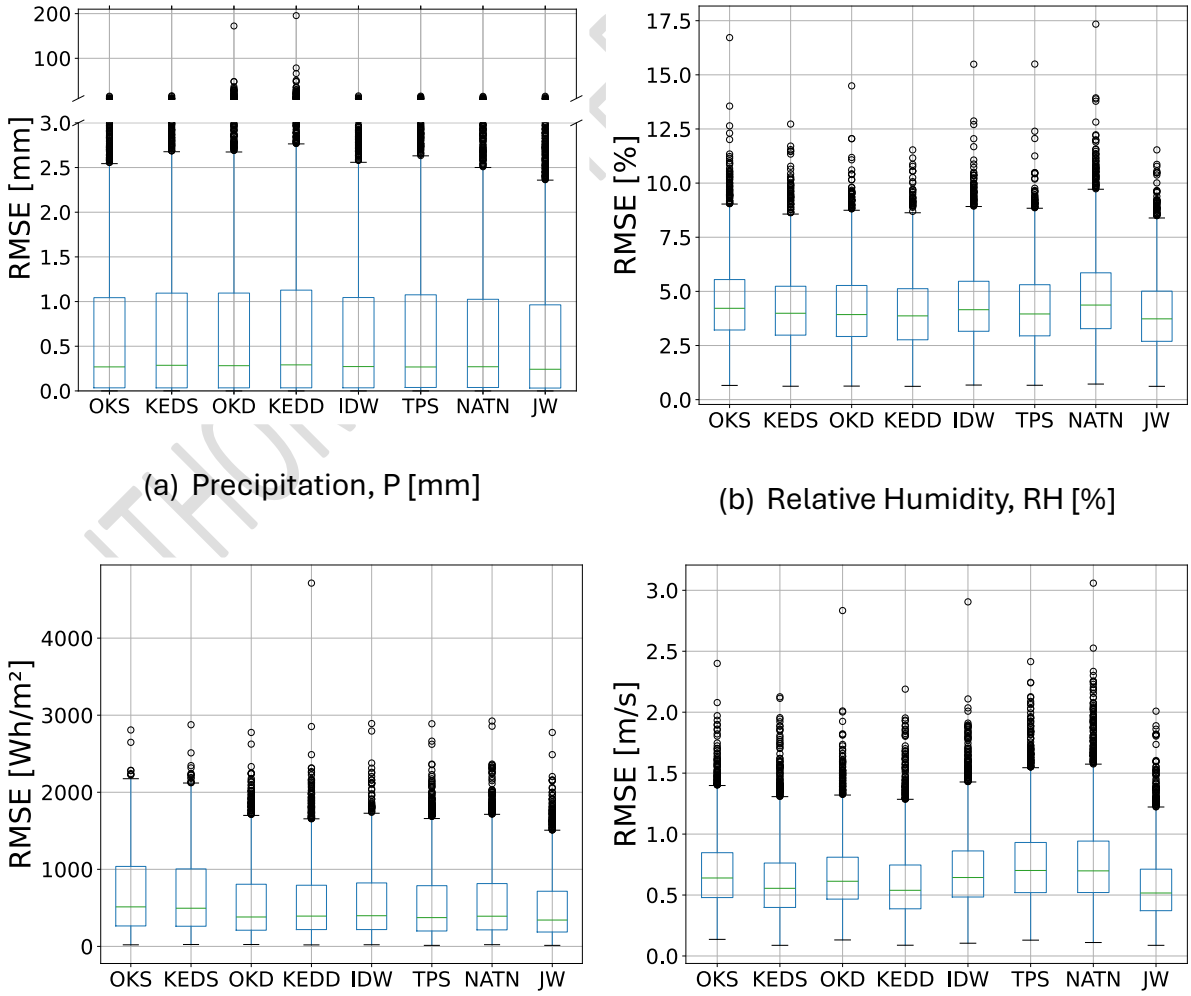
Density	P	TMAX	TMIN
25	0.019	0.071	-0.082
50	0.008	0.076	-0.083
75	0.000	0.075	-0.077
100	-0.001	0.074	-0.072

AUTHOR ACCEPTED MANUSCRIPT

Table C.3. Composite metric mean values for four network density sampling for P, TMAX and TMIN for the 2010-2020 period

Density	P	TMAX	TMIN
25	0.892	0.761	1.009
50	0.772	0.766	1.032
75	0.727	0.768	1.039
100	0.695	0.769	1.040

D. Bar plots for all the interpolations methods

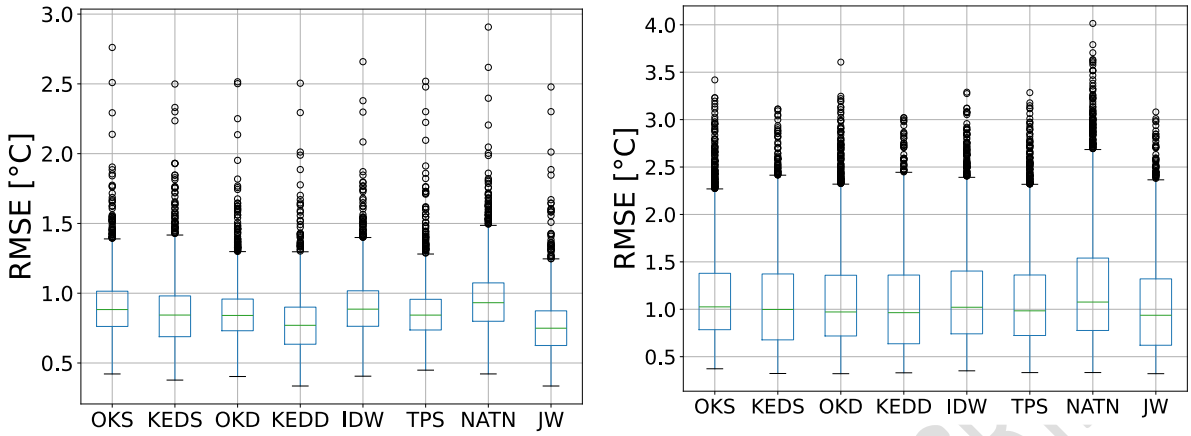


(a) Precipitation, P [mm]

(b) Relative Humidity, RH [%]

(c) Solar Radiation, SR [Wh/m^2]

(d) Wind Speed, WS [m/s]

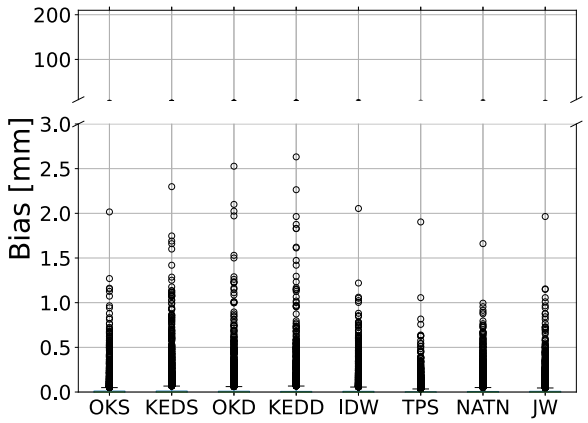


(e) Maximum air temperature, TMAX [°C]

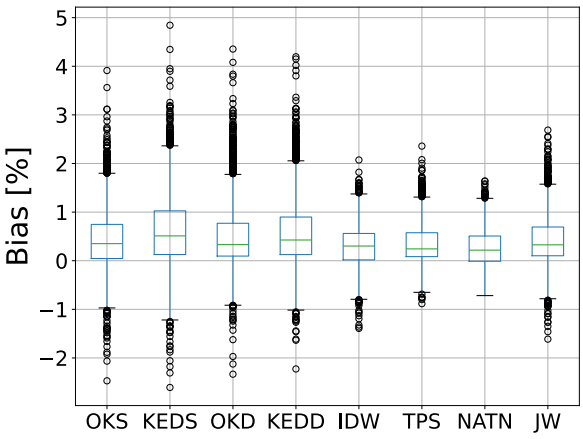
(f) Minimum air temperature, TMIN [°C]

749 Figure D.1. Boxplots of country averages RMSE (y-axis, in the unit of the presented
750 variable) for the different weather variables and the different SIMS tested through the JW
751 method for the case study (2010-2020), , with TPS: thin spline interpolation method,
752 NATN: Natural Neighbours, IDW: Inverse Distance Weight, KEDS: kriging with external drift
753 using a static semi-variogram (computed from mean values per station for the entire
754 study period), KEDD: kriging with external drift using a dynamic semi-variogram
755 (computed from available values of different stations for the selected day), OKS: ordinary
756 kriging using a static semi-variogram (computed from mean values per station for the
757 entire study period), OKD: ordinary kriging using a dynamic semi-variogram (computed
758 from available values of different stations for the selected day),), JW: the developed
759 methodology (Judicious Widget for Geospatial Robust Interpolation Design)

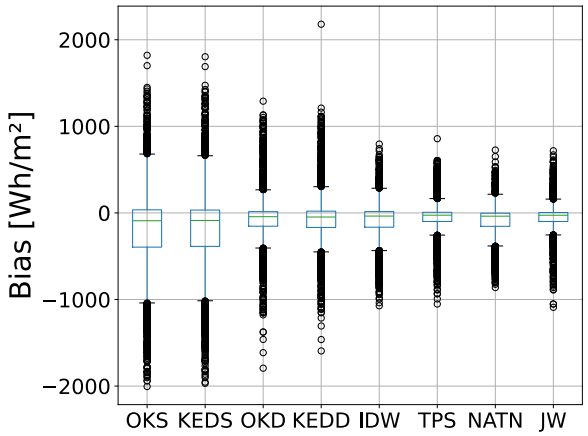
AUTHOR ACCEPTED MANUSCRIPT



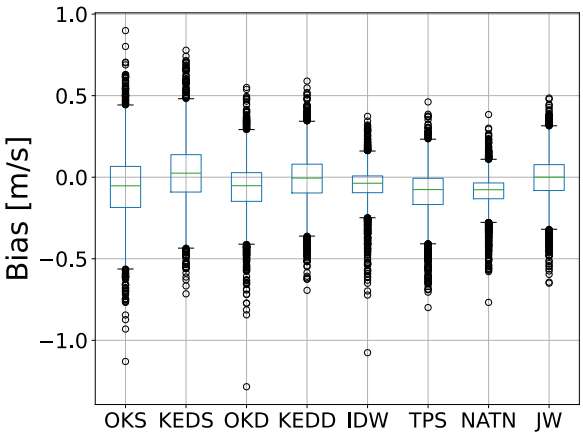
(a) Precipitation, P [mm]



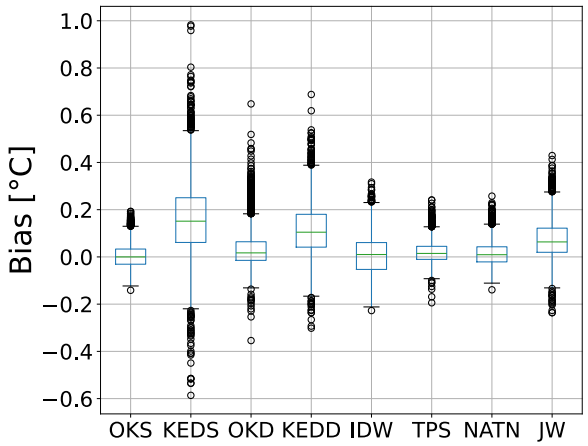
(b) Relative Humidity, RH [%]



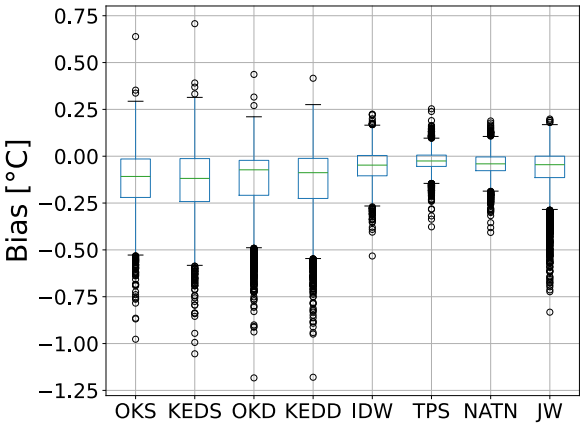
(c) Solar Radiation, SR [Wh/m^2]



(d) Wind Speed, WS [m/s]



(e) Maximum air temperature, TMAX [°C]

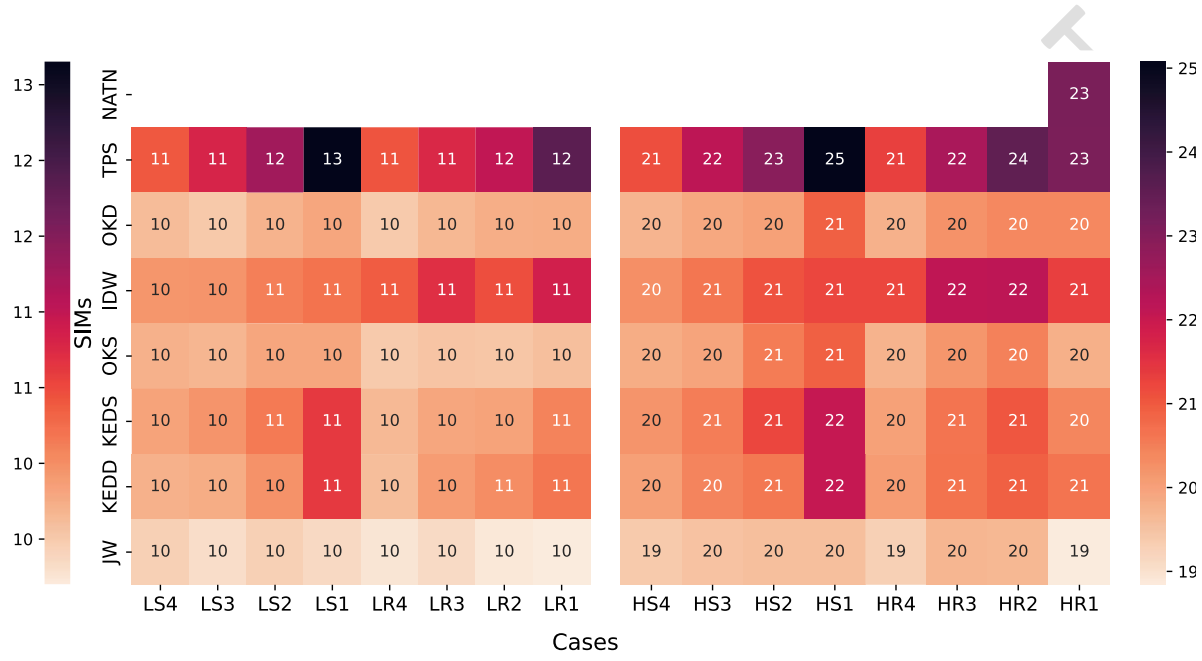


(f) Minimum air temperature, TMIN [°C]

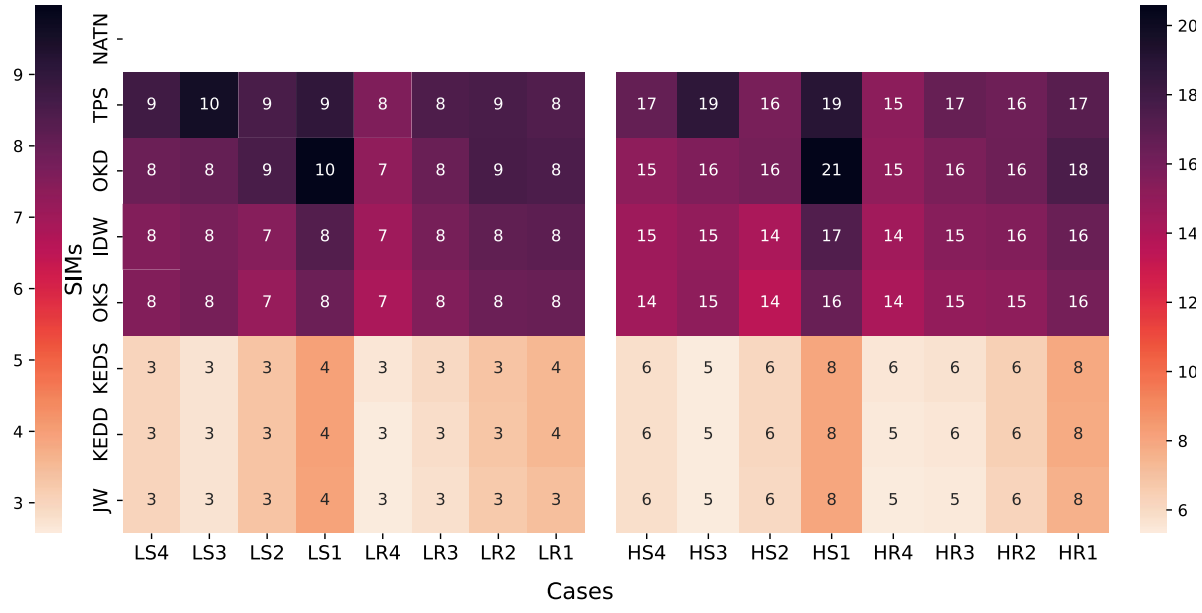
Figure D.2. Boxplots of country averages bias(y-axis, in the unit of the presented variable) for the different weather variables and the different SIMS tested through the JW method for the case study (2010-2020), with TPS: thin spline interpolation method, NATN: Natural Neighbours, IDW: Inverse Distance Weight, KEDS: kriging with external drift using a static semi-variogram (computed from mean values per station for the entire study period), KEDD: kriging with external drift using a dynamic semi-variogram (computed from available values of different stations for the selected day), OKS: ordinary kriging using a static semi-variogram (computed from mean values per station for the entire study period), OKD: ordinary kriging using a dynamic semi-variogram (computed from available values of different stations for the selected day), JW: the developed methodology (Judicious Widget for Geospatial Robust Interpolation Design)

AUTHOR ACCEPTED MANUSCRIPT

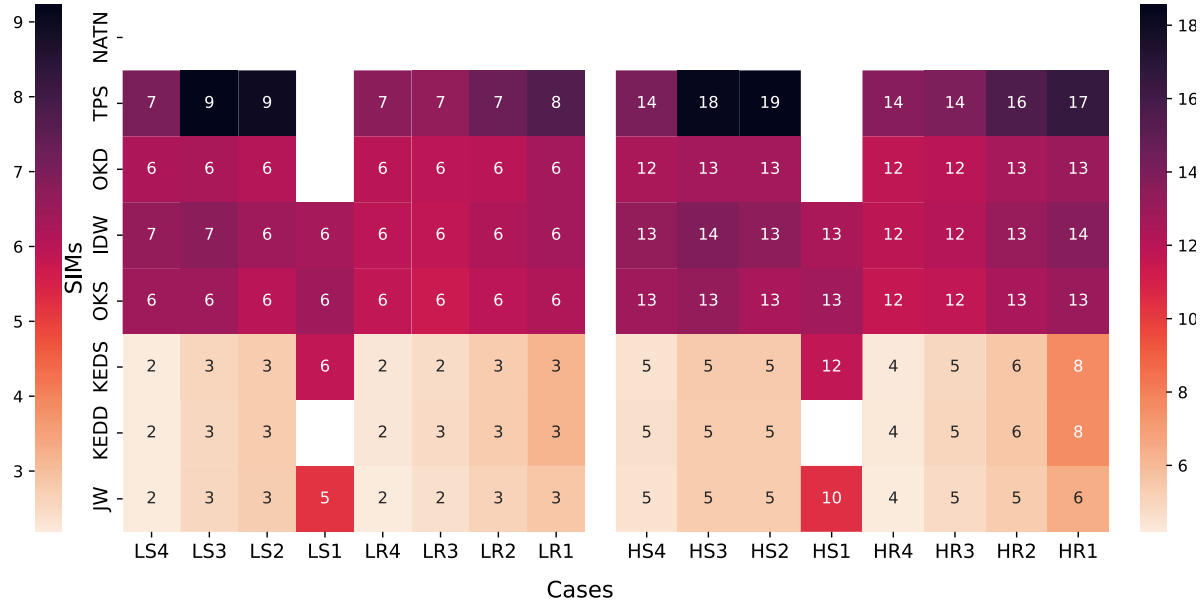
E. Artificial cases: heatmaps and boxplots of mean CM values of the synthetic dataset methodology test and frequency selection of SIMs



(a) Random spatial distribution



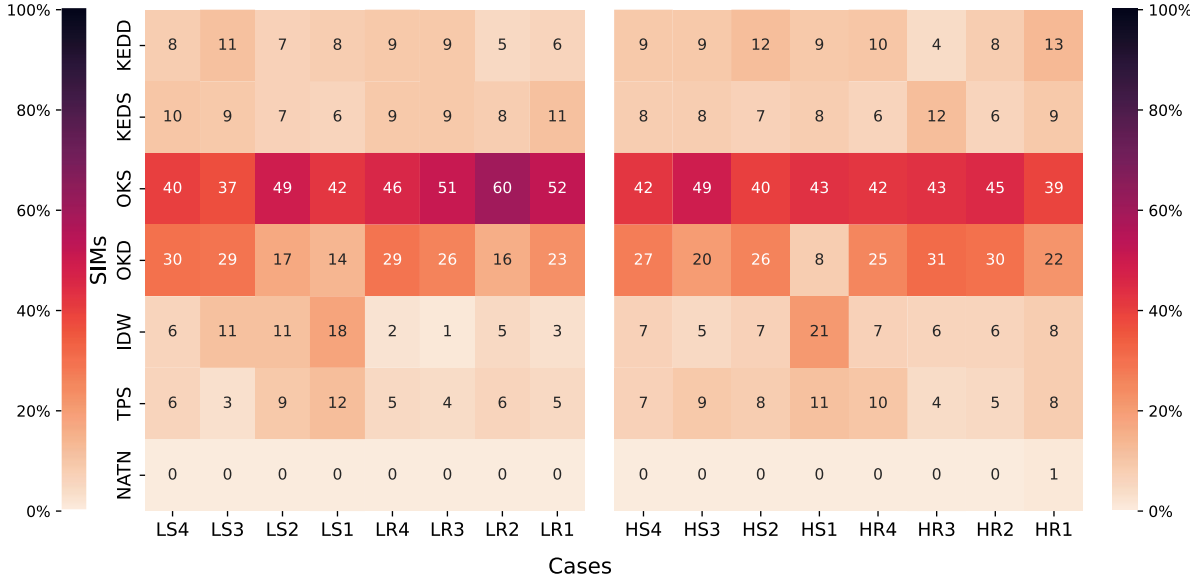
(b) Spatial correlation with altitude



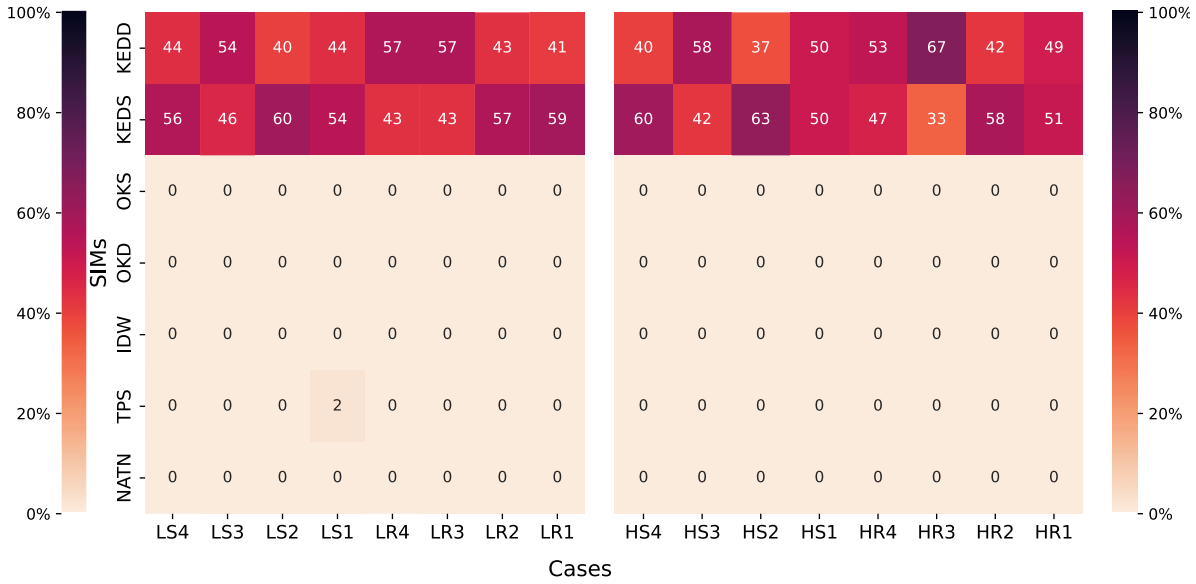
(c) Spatial correlation with altitude without high elevated stations

Figure E.1. Heatmap of country median CM values obtained using the LOOCVprocedure for the case study (2010-2020), with TPS: thin spline interpolation method, NATN: Natural Neighbours, IDW: Inverse Distance Weight, KEDS: kriging with external drift using a static semi-variogram (computed from mean values per station for the entire study period), KEDD: kriging with external drift using a dynamic semi-variogram (computed from available values of different stations for the selected day), OKS: ordinary kriging using a static semi-variogram (computed from mean values per station for the entire study period), OKD: ordinary kriging using a dynamic semi-variogram (computed from available values of different stations for the selected day), JW: the developed methodology (Judicious Widget for Geospatial Robust Interpolation Design), artificial cases are labelled using a code where the first letter indicates the standard deviation of input values: 'L' for low and 'H' for high. The second letter denotes the network configuration: 'S' for systematic and 'R' for random. The accompanying number represents the number of networks being considered.

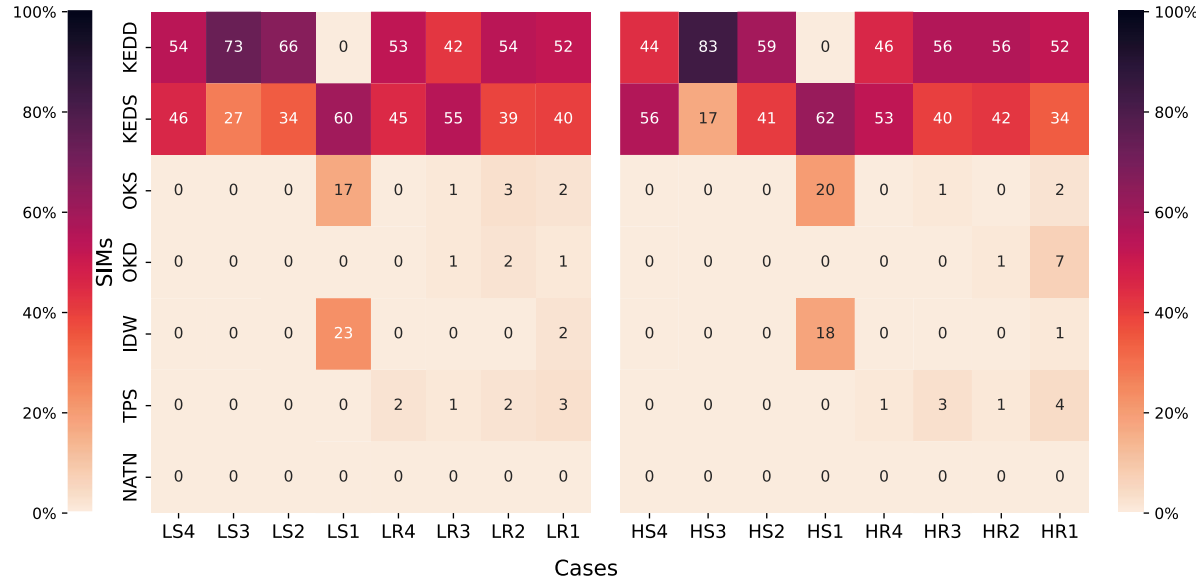
AUTHOR ACCEPTED MANUSCRIPT



(a) Random spatial distribution



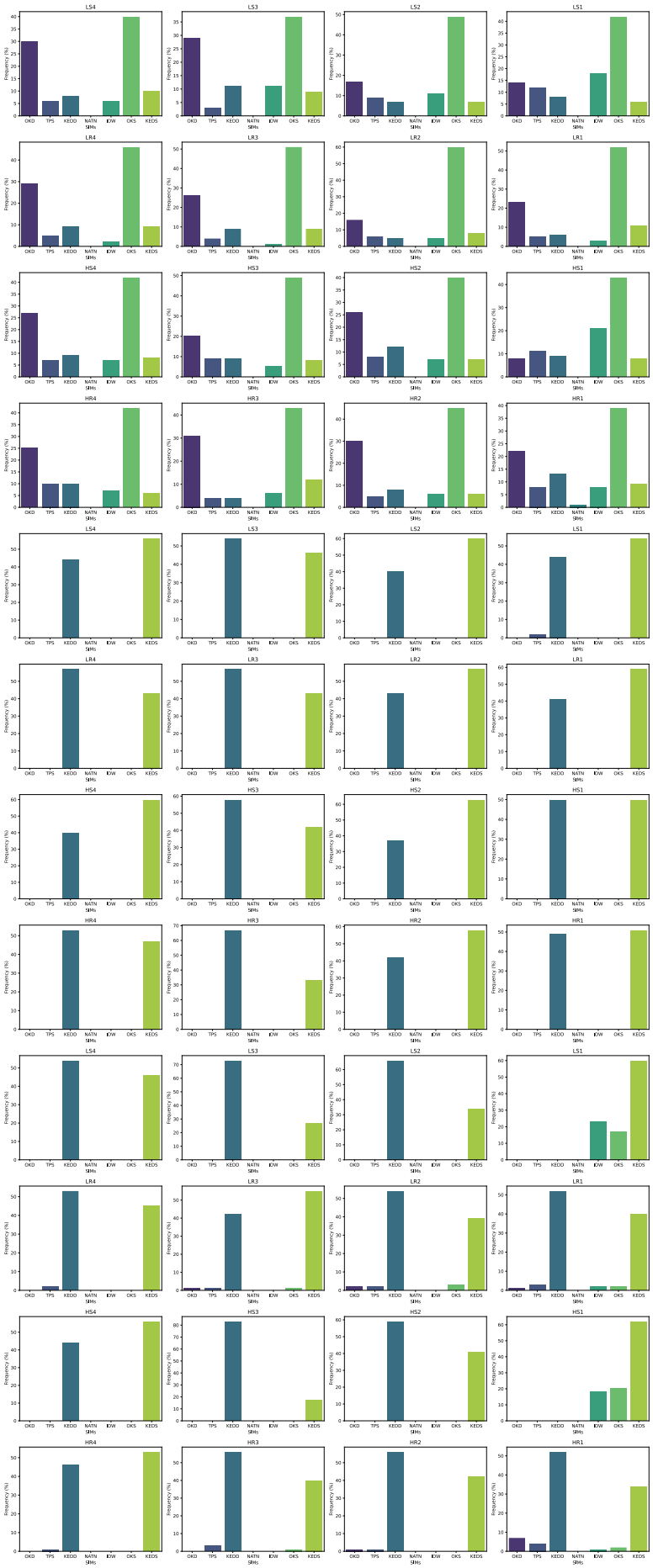
(b) Spatial correlation with altitude



(c) Spatial correlation with altitude without high elevated stations

Figure E.2. Heatmap of frequencies selection for the different SIMs, with TPS: thin spline interpolation method, NATN: Natural Neighbours, IDW: Inverse Distance Weight, KEDS: kriging with external drift using a static semi-variogram (computed from mean values per station for the entire study period), KEDD: kriging with external drift using a dynamic semi-variogram (computed from available values of different stations for the selected day), OKS: ordinary kriging using a static semi-variogram (computed from mean values per station for the entire study period), OKD: ordinary kriging using a dynamic semi-variogram (computed from available values of different stations for the selected day)), JW: the developed methodology (Judicious Widget for Geospatial Robust Interpolation Design), artificial cases are labelled using a code where the first letter indicates the standard deviation of input values: 'L' for low and 'H' for high. The second letter denotes the network configuration: 'S' for systematic and 'R' for random. The accompanying number represents the number of networks being considered.

AUTHOR ACCEPTED MANUSCRIPT


LS4
LS3
LS2
LS1
LR4
LR3
LR2
LR1
HS4
HS3
HS2
HS1
HR4
HR3
HR2
HR1
Frequency (%)
OKD
TPS
KEDD
NATN SMs
IDW
OKS
KEDS

MANUSCRIPT

804 Figure E.3. Frequencies of SIMs' selection for all specific cases tested, with the sixteen
805 first plot: the randomly distributed values, the folowing sixteen: the spatially correlated
806 ones with altitude and the last sixteen: the spatial distribution correlated with altitude
807 without the highest stations, with TPS: thin spline interpolation method, NATN: Natural
808 Neighbours, IDW: Inverse Distance Weight, KEDS: kriging with external drift using a static
809 semi-variogram (computed from mean values per station for the entire study period),
810 KEDD: kriging with external drift using a dynamic semi-variogram (computed from
811 available values of different stations for the selected day), OKS: ordinary kriging using a
812 static semi-variogram (computed from mean values per station for the entire study
813 period), OKD: ordinary kriging using a dynamic semi-variogram (computed from available
814 values of different stations for the selected day), JW: the developed methodology
815 (Judicious Widget for Geospatial Robust Interpolation Design) artificial cases are labelled
816 using a code where the first letter indicates the standard deviation of input values: 'L' for
817 low and 'H' for high. The second letter denotes the network configuration: 'S' for
818 systematic and 'R' for random. The accompanying number represents the number of
819 networks being considered.

AUTHOR ACCEPTED MANUSCRIPT


LS4
LS3
LS2
LS1
LR4
LR3
LR2
LR1
HS4
HS3
HS2
HS1
HR4
HR3
HR2
HR1
OKS KEDS OKD KEDD IDW TPS NATN JW

MANUSCRIPT

Figure E.4. Boxplots of CM for all SIMs on all specific cases tested, with the sixteen first plot: the randomly distributed values, the flowing sixteen: the spatially correlated ones with altitude and the last sixteen: the spatial distribution correlated with altitude without the highest stations, with TPS: thin spline interpolation method, NATN: Natural Neighbours, IDW: Inverse Distance Weight, KEDS: kriging with external drift using a static semi-variogram (computed from mean values per station for the entire study period), KEDD: kriging with external drift using a dynamic semi-variogram (computed from available values of different stations for the selected day), OKS: ordinary kriging using a static semi-variogram (computed from mean values per station for the entire study period), OKD: ordinary kriging using a dynamic semi-variogram (computed from available values of different stations for the selected day), JW: the developed methodology (Judicious Widget for Geospatial Robust Interpolation Design), artificial cases are labelled using a code where the first letter indicates the standard deviation of input values: 'L' for low and 'H' for high. The second letter denotes the network configuration: 'S' for systematic and 'R' for random. The accompanying number represents the number of networks being considered.

AUTHOR ACCEPTED MANUSCRIPT

F. Mountainous case

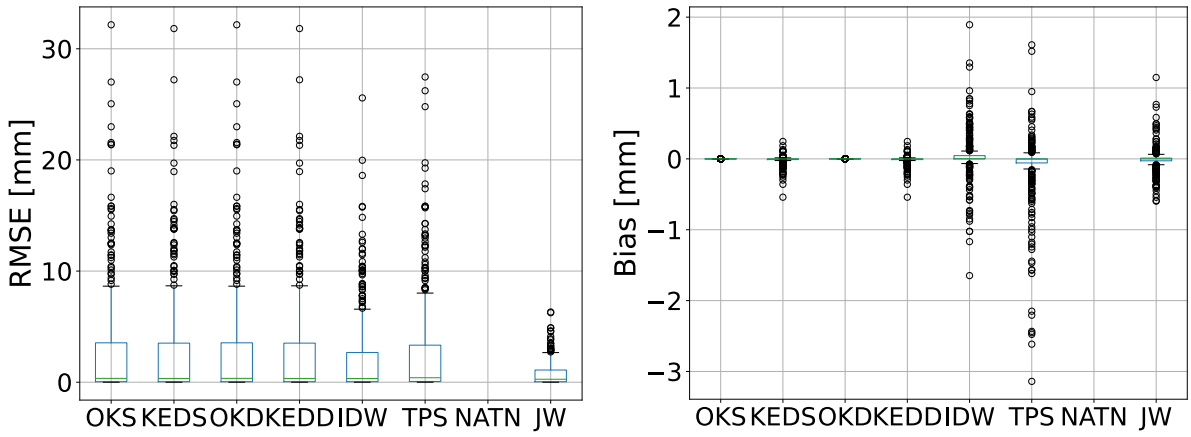


Figure F.1. Boxplots of country averages RMSE (a) and bias (b) (y-axis, in the unit of the presented variable) for the precipitation in the mountainous case and the different SIMS tested through the JW method for the case study (2010-2020), with TPS: thin spline interpolation method, NATN: Natural Neighbours, IDW: Inverse Distance Weight, KEDS: kriging with external drift using a static semi-variogram (computed from mean values per station for the entire study period), KEDD: kriging with external drift using a dynamic semi-variogram (computed from available values of different stations for the selected day), OKS: ordinary kriging using a static semi-variogram (computed from mean values per station for the entire study period), OKD: ordinary kriging using a dynamic semi-variogram (computed from available values of different stations for the selected day), JW: the developed methodology (Judicious Widget for Geospatial Robust Interpolation Design)

AUTHOR ACCEPTED MANUSCRIPT