RESEARCH

Quantifying uncertainties related to observational datasets used as reference for regional climate model evaluation over complex topography — a case study for the wettest year 2010 in the Carpathian region

Tímea Kalmár[1,2] · Erzsébet Kristóf[1] · Roland Hollós[1,3,4] · Ildikó Pieczka[1] · Rita Pongrácz[1]

Abstract

Gridded observational datasets are often used for the evaluation of regional climate model (RCM) simulations. However, the uncertainty of observations affects the evaluation. This work introduces a novel method to quantify the uncertainties in the observational datasets and how these uncertainties affect the evaluation of RCM simulations. Besides precipitation and temperature, our method uses geographic variables (e.g. elevation, variability of elevation, effect of station), which are considered as uncertainty sources. To assess these uncertainties, a complex analysis based on various statistical tools, e.g. correlation analysis and permutation test, was carried out. Furthermore, we used a special metric, the reduction of error (*RE*) to identify where the RCM shows improvement compared to the lateral boundary conditions (LBCs). We focused on the Carpathian region, because of its unique orographic and climatic conditions. The method is applied to two observational datasets (CarpatClim and E-OBS) and to RegCM simulations for 2010, the wettest year in this region since 1901.

The results show that CarpatClim is wetter than E-OBS, while temperature is similar over the lowland; however, E-OBS is significantly warmer than CarpatClim over the mountains. By the RE metric, RegCM has improvement against the LBCs over mountains for temperature and areas with dense station network for precipitation. Nevertheless, there are significant differences in the results depending on which observational dataset was used concerning precipitation. The evaluation method can be applied to other datasets, different time periods and areas. It is also suitable to find dataset errors, which is also exemplified in this paper.

Keywords Carpathian region · Gridded observational dataset · CarpatClim · E-OBS · RCM evaluation · RegCM · Uncertainty · Reduction of error

1 Introduction

Climate researchers use general circulation models (GCMs) and regional climate models (RCMs) to improve our understanding of the climate system. Observations are used during the model development phase, but model calibration and initialisation also often heavily rely on observational datasets (e.g. Bellprat et al. 2012; Hazeleger et al. 2013). Furthermore, the availability of reliable high-quality observational data is important for model evaluation (Kotlarski et al. 2019), for which temperature and precipitation are most often used (Perkins et al. 2007; Kotlarski et al. 2014, 2019; Kalmár et al. 2021). However, measuring precipitation is challenging because of its high variability in space and time and the existence of measurement errors (Bacchi

✉ Tímea Kalmár
timea.kalmar@ttk.elte.hu

1 Department of Meteorology, Institute of Geography and Earth Sciences, Eötvös Loránd University, Pázmány Péter sétány 1/A, H-1117 Budapest, Hungary

2 Doctoral School of Earth Sciences, Eötvös Loránd University, Pázmány P. sétány 1/A, H-1117 Budapest, Hungary

3 Agricultural Institute, Centre for Agricultural Research, Brunszvik u. 2, H-2462 Budapest, Hungary

4 Doctoral School of Environmental Sciences, Eötvös Loránd University, Pázmány Péter sétány 1/A, H-1117 Budapest, Hungary

and Kottegoda, 1995; Frei 2014; Prein and Gobiet 2017; Kotlarski et al. 2019).

Several studies quantify the observation-related uncertainty by comparing observation-based gridded datasets for specific variables (mostly precipitation or temperature) and different regions (e.g. Hofstra et al. 2009; Kyselý and Plavcová 2010; Palazzi et al. 2013; Rauthe et al. 2013; Gervais et al. 2014; Schneider et al. 2014; Isotta et al. 2015; Berg et al. 2016; Bandhauer et al. 2022). Most studies focusing on Europe use the E-OBS dataset, which covers the whole continent. Hofstra et al. (2009) tested precipitation and temperature from the E-OBS dataset and found inhomogeneities in the underlying station data and underestimation of extremes within the data. Kyselý and Plavcová (2010) highlighted the facts that stations may not be representative for a wider area and insufficient density of information from station observations used for the interpolation lead to bias in E-OBS, which affect the evaluation of RCMs. Interpolation tends to introduce other bias, such as the excessive smoothing of spatial variability, and may thus lead to an underestimation of extremes (Hofstra et al. 2009, Maraun et al. 2012; Bandhauer et al. 2022). The observations used for creating the dataset differ from each other in many characteristics, including spatiotemporal resolution, length and homogeneity of the measurement time series. Other differences may include quality checking and error correction procedures. In addition, some observational datasets (e.g. E-OBS) are regularly updated. All those factors lead to notable differences in the quality of the observational datasets (Tanarhte et al. 2012). Therefore, a comprehensive analysis of observational datasets is required, where less commonly used variables are also included, such as elevation and station density.

In the field of RCM evaluation, several papers examined different observational data and their effects on the evaluation (e.g. Prein and Gobiet, 2017; Beck et al., 2017; Fantini et al. 2018; Kotlarski et al. 2019). Prein and Gobiet (2017) focused on precipitation and compared many gridded observational datasets over selected parts of Europe and used this observational ensemble to evaluate RCMs. They found that observational uncertainty may be at a similar magnitude as RCM biases, particularly in regions with low station density. The magnitude of the observational uncertainty increases with increasing spatiotemporal resolution. Beck et al. (2017) showed that the uncertainty in long-term precipitation means among the datasets was generally the largest in topographically complex and arid regions. Fantini et al. (2018) mentioned that the observational datasets (e.g. CarpatClim (Spinoni et al. 2015), E-OBS (Haylock et al. 2008; Cornes et al. 2018), SAFRAN (Vidal et al. 2010) and Spain02 (Herrera et al. 2016)) are influenced by widely different station densities and methodological approaches regarding their construction, which make RCM evaluation rather difficult (e.g. some observational datasets are based only on station data, while others use additional high-resolution reanalysis data).

Kotlarski et al. (2019) employed a simple ranking method on RCM evaluation and noted that results can depend on the reference dataset used. This dependency is more important for precipitation than for temperature due to its higher variability.

Using and testing RCMs for specific time periods (e.g. when heatwaves or flash flood occur) and for specific areas — i.e. for regions with complex topography — is beneficial because the modelling of the climate conditions of these regions is quite difficult as it was pointed out by Ceglar et al. (2018) in case of the Carpathian region (located in East-Central Europe). Pall et al. (2011) focused on flood risk in the UK in 2000, and they generated several thousand GCM simulations to show that global anthropogenic greenhouse gas emissions substantially increased the risk of flood occurrence in the UK. Mitchell et al. (2016) tested the capability of RCM for capturing the synoptic conditions of the European heatwave in 2003. Varga and Breuer (2020) evaluated the performance of WRF model, which was used as an RCM, and they analysed its sensitivity to different physical and dynamical settings for the year 2013 over Central Europe.

Previous studies indicate that comprehensive examinations of RCM-simulations require a quantification of observational uncertainty in the first place. In this paper, we introduce a novel method to assess observational uncertainty, namely how the selection of the observational datasets (in this paper the CarpatClim and the E-OBS) affects the evaluation of RCM simulations (in this paper the RegCM) with respect to the Carpathian region. Section 2 describes the datasets and methods used in the study. Then, the results of the uncertainty regarding the observational datasets and evaluation results based on the RCM simulations are presented and discussed in Section 3. Finally, we present our main conclusions in Section 4.

2 Data and methods

2.1 Short description of the study area and the target period

The study region extends between 17–27°E and 44–50°N (due to the spatial range of the CarpatClim), covering the area of the Carpathian region, which consists of the Carpathian Basin and the Carpathians). The Carpathian region is characterised by unique orography and climate conditions, namely, it is a transition area between Mediterranean, oceanic, and continental climates. The Carpathian Basin is bordered by the Alps in the west, by the Dinaric Alps in the southwest and by the Carpathians in the north and east. The dominant wind direction over the basin is western, north-western (Bartholy et al. 2003), resulting in a west to east spatial gradient of precipitation modulated by local topography. As the air mass from the Atlantic region crosses the

Alps, it loses humidity resulting in a precipitation decrease towards the east. The annual mean precipitation is 700–800 mm in the western part of the Basin, while the lowest annual precipitation totals occur in eastern Hungary with 550 mm (UNEP 2007; Spinoni et al. 2015).

The Carpathian Mountains function as an important obstacle to the circulation of air masses over Europe. The Carpathian Mountains have a temperate climate, with a basic continental regime, increasingly intensive eastwards (Cheval et al. 2014). The altitude, the compact arrangement and the shape of the Carpathians introduce important disturbances in the climatic zonality and in the general atmospheric circulation (UNEP, 2007). The interaction between the mountains and the atmospheric flow is particularly complex, mountains playing a significant perturbation role in the large-scale processes with to the overall dimension and orientation of the ranges and finally resulting in the prevailing airflows. Precipitation totals rise with altitude and decrease from west to east. The average annual precipitation amounts varies from 600 to 1600 mm and is mostly between 900 and 1200 mm, depending on altitude and local conditions (UNEP 2007; Ptácek et al. 2011; Repel et al. 2021).

The year 2010 was selected as the target period, when extremely heavy and persistent rain caused severe flooding in East-Central Europe (Poland, Czechia, Slovakia, Serbia, Hungary), especially in May and June (Bissolli et al. 2011). Due to the large amount of precipitation over the whole year, 2010 was the wettest year in this region since the beginning of coordinated measurements (WMO, 2011). Figure 1 also shows the annual precipitation for the Carpathian region during 1961–2010 in the observational datasets used in this study, namely CarpatClim and E-OBS. The annual precipitation in 2010 was 978 mm/year, while the second wettest year was in 2005 in the past 50 years

with 856 mm/year in CarpatClim. This value is still 100 mm less than the value in the wettest year. For E-OBS, the annual precipitation in 2010 was 863 mm/year, and the second wettest year was 1970 with 782 mm/year.

2.2 Observational datasets and climate simulations used in the study

2.2.1 The CarpatClim dataset

The CarpatClim is a high-resolution interpolated gridded dataset for the Carpathian region with 0.1° by 0.1° horizontal resolution, covering the 1961–2010 period, containing 11 major surface meteorological variables and several derived variables for daily basis (Szalai et al. 2013; Spinoni et al. 2015). The CarpatClim is based on the observations of precipitation and temperature stations. Quality control, gap filling and homogenization were conducted by the MASH software (Szentimrey, 2007). Spatial interpolation was made following a regression kriging concept using the MISH software (Szentimrey and Bihari, 2006). The daily mean temperature is calculated as the average of the daily minimum and maximum temperature, while the altitude was also considered during the interpolation (Spinoni et al. 2015).

2.2.2 E-OBS dataset

The gridded E-OBS dataset (Haylock et al. 2008; Cornes et al. 2018) covers the entire European land surface. It spans the period 1950–2022. The E-OBS is based on the ECA&D station data and more than 2000 further stations from additional archives, and station density is increasing over the years (Kotlarski et al. 2019). The E-OBS contains

Fig. 1 Time series (1961–2010) of annual precipitation for CarpatClim and E-OBS, averaged over Carpathian region
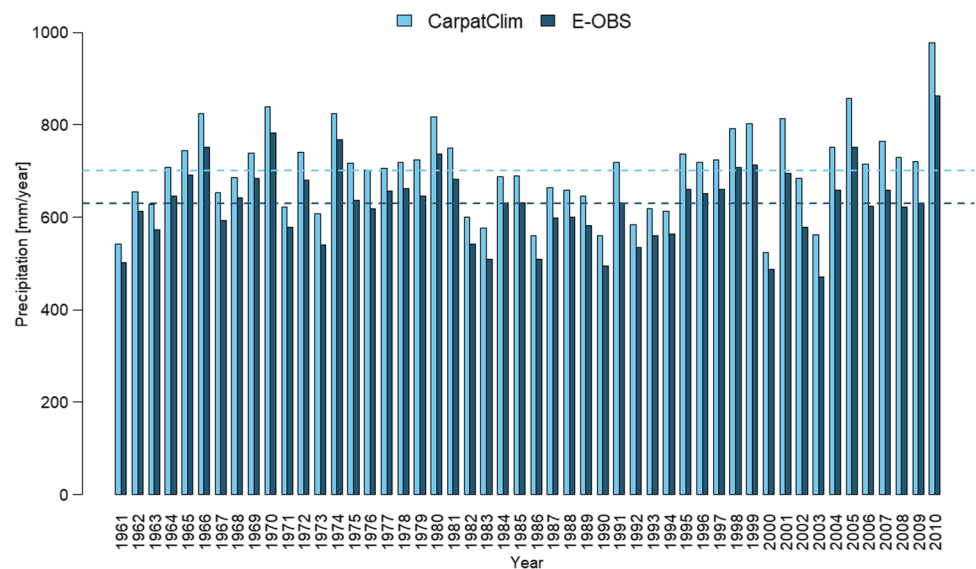
Table 1 The main features of the two observational datasets

	CarpatClim	E-OBS
Spatial range	44°N–50°N	25°N–71.5°N
	17°E–27°E	25°W–45°E
Timeframe	1961–2010	1950–2022
Temporal resolution	1-day	1-day
Spatial resolution	0.1°×0.1°	0.1°×0.1°
Number of precipitation stations in the study region in 2010	560	79
Number of temperature stations in the study region in 2010	224	69

eight meteorological variables. The version used here is 22.0e (ensemble mean for daily precipitation sum and daily mean temperature), and the horizontal resolution is 0.1° by 0.1° grid. The spatial coverage is heterogeneous, with a dense network in Czechia, and a sparse network in Ukraine. E-OBS uses ordinary kriging interpolation method (Klein Tank et al. 2002), and elevation is also considered to calculate the temperature (Wood, 2003, 2006; Cornes et al. 2018). Note that, E-OBS has advanced interpolation which captures better the influence of topography on the analysed climatic parameters (Cornes et al. 2018; Sidău et al. 2021).

The main characteristics of the datasets can be found in Table 1.

Figure 2 shows the measuring stations and elevation for both datasets, as it can be seen the stations are not distributed homogeneously over the target area. Table 2 characterises the station density for precipitation and temperature defined as the ratio of the number of grid cells containing stations relative to the total number of grid cells for each country. In general, CarpatClim uses more stations within the Carpathian region than E-OBS, and the total number

of precipitation stations is higher than the number of temperature stations. The average distance between the stations of CarpatClim is ~25 km for precipitation and ~50 km for temperature (Spinoni et al. 2015). It is worth to mention that these two observational datasets do not contain all observational station data of the national meteorological services.

The standard deviation (*sd*) of elevation can be used to characterize the target area, and it shows that the elevation field of CarpatClim is more detailed than that of E-OBS (Table 2). The *sd* of elevation of CarpatClim is higher in every country than the *sd* values of E-OBS. Romania — which is a mountainous area — has the highest *sd* of the elevation with 401 m in CarpatClim and 378 m in E-OBS. The country with the smallest *sd* is Hungary (CarpatClim: 80 m, E-OBS: 75 m), which occupies the largest plain area within the domain.

2.2.3 RegCM simulations

In this study, we used regional climate model version 4.7 (RegCM4.7, Giorgi, 1989; Giorgi et al. 2012). This study focuses on our simulations for the target period 2010 (2009 was the spin-up year) with initial and lateral boundary conditions (LBC) from the 0.75° horizontal resolution data of the ERA-Interim reanalysis (Dee et al. 2011), which is a commonly used LBC for regional climate simulations (e.g. Giorgi 2019). The horizontal resolution of our simulations is 10 km to represent the fine topography of the target area (Gao et al. 2006), and the integration timestep is 30 s, while the temporal resolution of the RegCM output is 1 day. The integration domain is over 6°–29°E and 43.8°–50.6°N after removing the buffer zones, but we analyse the simulations only over the CarpatClim domain. A great number of sensitivity analyses have been completed with the RegCM
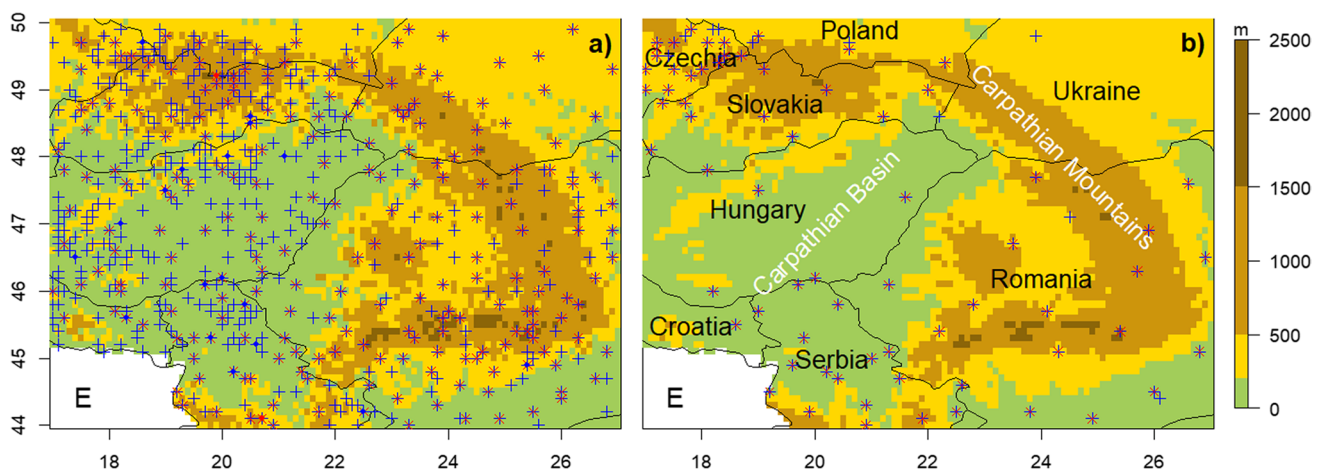

Fig. 2 The location of precipitation and temperature stations (indicated by blue + and red ×, respectively, if a grid cell contains two stations, it is marked by blue or red dot) for CarpatClim (**a**) and E-OBS (**b**) (data

source: www.carpatclim-eu.org and www.ecad.eu). The white area indicates no data is available. The topography is shown by mean grid cell elevation.

Table 2 Topographic characteristics (total number of grid cell, standard deviation of elevation) and station density by country listed in CarpatClim and E-OBS

Country	Total number of grid cells	Percentage of the grid cell from CarpatClim (%)	Precipitation station density (%)		Temperature station density (%)		Standard deviation of elevation (m)	
			CarpatClim	E-OBS	CarpatClim	E-OBS	CarpatClim	E-OBS
Croatia	169	2.9	13.6	0.6	3.6	0.6	149	102
Czechia	168	2.8	13.7	11.9	3.6	10.7	178	163
Hungary	1045	17.7	15.3	0.4	3.4	0.4	80	75
Poland	303	5.1	11.2	2.3	3	1	226	188
Romania	2192	37.2	6	0.9	3.9	0.8	401	378
Serbia	507	8.6	15	3.2	5.7	3.2	193	169
Slovakia	599	10.2	14.2	1.5	4.2	1.5	322	285
Ukraine	912	15.5	3.1	0.2	3.1	0	272	242

regarding the selection of a suitable integration domain, an adequate horizontal resolution, potential driving models, applied physics schemes and adaptation tools for Central and Eastern Europe (Torma et al. 2011; Güttler et al. 2014; Pieczka et al. 2017; Kalmár et al. 2021). These recommendations were taken into account, when we chose the physics schemes. Twenty-four simulations were carried out by using all combinations of the physics schemes (2 land surface schemes, 2 microphysics schemes, 3 cumulus schemes and 2 planetary boundary layer schemes) listed in Table 3. For the RegCM evaluation, we used daily precipitation sum and daily temperature values.

For further calculations, the E-OBS observational dataset, the ERA-Interim reanalysis and the RegCM simulations required regridding on the CarpatClim grid. We used the nearest-neighbour method to avoid the oversmoothing of the fields. A disadvantage of this interpolation method is the penalisation of low-resolution datasets, resulting in a set of large pixels with the same values on maps (Di Luca et al. 2016).

Based on the above description, we analysed 5895 grid cells over the Carpathian region (domain size is 101×61 grid cells, but Bosnia Herzegovina with 266 grid points is excluded due to lack of data). Furthermore, each grid cell is associated with a daily time series (365 elements).

2.3 Evaluation method

2.3.1 Variables used for the study

For the first step of the analysis, the two observational datasets (CarpatClim and E-OBS) are compared based on their daily precipitation and temperature values. To examine the relationships between the variables, we use the annual sum of precipitation (*PR*) and the annual mean temperature (*TAS*) for 2010. Furthermore, we examine the spatial variation of less commonly used variables: the effect of station density on precipitation (*PR_ST*) and effect of station density on temperature (*TAS_ST*), elevation (*E*), and the variability of elevation (*VE*).

To determine the *PR_ST* and *TAS_ST*, we apply a moving window filter to the data with the window size of 5×5 grid cells, which covers approximately 50 km×50 km, which is the average distance between the stations for temperature in CarpatClim (Spinoni et al. 2015). Then we count how many stations are located within the window, and the number is assigned to the central cell of the window.

To calculate the *VE*, we use the moving window method similarly to determine *PR_ST* and *TAS_ST*. We compute the

Table 3 The applied physics schemes in the 24 different RegCM simulations for 2010

Physics scheme		References
Land-surface	BATS	Dickinson et al. (1993)
	CLM4.5	Oleson et al. (2004)
Microphysics	Modified SUBEX	Pal et al. (2000); Torma et al. (2011)
	WSM5	Hong et al. (2004)
Cumulus	Grell (land)/Emanuel (ocean)	Grell (1993); Emanuel and Živković-Rothman (1999)
	Kain-Fritch	Kain (2004); Kain and Fritsch (1990)
	Tiedtke	Tiedtke (1996)
Boundary layer	Holtslag	Holtslag et al. (1990)
	UW PBL	Grenier and Bretherton (2001); Bretherton and Park (2009)

difference between the highest and the lowest elevation values within the window, and the difference is assigned to the central cell of the window.

2.3.2 Statistical analysis comparing observational datasets and pairs of variables for each observational dataset

Firstly, temporal relationships between the E-OBS and the CarpatClim were examined in each grid cell by using daily precipitation and temperature time series. In the second step, we analyse the spatial relationships between all possible pairs of variables (*E, VE, PR, PR_ST, TAS, TAS_ST*) for each observational dataset to gain deeper insight into their relationships.

In case of the analysis of temporal relationships, the CarpatClim and E-OBS were compared by calculating average difference (*DIFF*) for temperature and relative difference for precipitation (*DIFF*$_{rel}$), root-mean-square error (*RMSE*) and temporal Pearson correlation coefficient (r_t) for precipitation and temperature. These metrics are calculated for each grid cell as follows:

$$DIFF = \frac{\sum_{t=1}^{N}\left(EOBS_t - CarpatClim_t\right)}{N} \qquad (2.1)$$

$$DIFF_{rel} = \sum_{t=1}^{N}\left(\frac{EOBS_t - CarpatClim_t}{CarpatClim_t}\right) \times 100 \qquad (2.2)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{N}\left(EOBS_t - CarpatClim_t\right)^2}{N}} \qquad (2.3)$$

$$r_t = \frac{\sum_{i=1}^{N}\left(EOBS_t - \overline{EOBS}\right)\left(CarpatClim_t - \overline{CarpatClim}\right)}{\sqrt{\sum_{t=1}^{N}\left(EOBS_t - \overline{EOBS}\right)^2}\sqrt{\sum_{t=1}^{N}\left(CarpatClim_t - \overline{CarpatClim}\right)^2}} \qquad (2.4)$$

where N is the length of the time series (N=365), t is the timestep, and overline indicates the average of the corresponding time series.

In case of the analysis of spatial relationships, the spatial correlation coefficient (r_s) is computed between all possible pairs of variables for each observational dataset.

For example, in case of *VE* and *PR*:

$$r_s = \frac{\sum_{i=1}^{N}\left(VE_i - \overline{VE}\right)\left(PR_i - \overline{PR}\right)}{\sqrt{\sum_{i=1}^{N}\left(VE_i - \overline{VE}\right)^2}\sqrt{\sum_{i=1}^{N}\left(PR_i - \overline{PR}\right)^2}} \qquad (2.5)$$

where N is the total number of the grid cells (N=5895) and i denotes the *ith* grid cell.

Due to the dependency of grid cells, the significance of the r_s values obtained from Eq. 2.5 cannot be determined by

commonly used hypothesis testing method like *t*-test. All grid cells also cannot be used because the large number of the grid cells would result in significant r_s values even if it is close to zero (Maxwell et al. 2008). Due to this reason, random sampling is used first, as follows. One hundred elements of the grid cells of each pair of variables are randomly selected, and the correlation coefficients are calculated between the random samples (hereinafter called as the original correlation coefficients, $r_{s,original}$). This process is repeated 10,000 times for all examined pairs of variables. As a result, a total of 10,000 $r_{s,original}$ values are produced in each examination. Note that the number of elements used here does not have any effect on the results of the permutation test due to the high number of random sequences.

Then, permutation test (Pitman, 1937) was carried out in order to determine whether the $r_{s,original}$ values are considered as significant or they are produced by random processes. For that purpose, the 100 elements for one variable from each pair of variables are randomly shuffled. After that, correlation coefficients are calculated between the reshuffled random sample and the original elements of the other variable 10,000 times (the obtained correlations are henceforth called as random correlation coefficients, $r_{s,random}$).

Finally, significance of the $r_{s,original}$ is determined in two ways. At first, we compute the percentage when the $r_{s,random}$ values are stronger than the median of $r_{s,original}$ values. The significance level is set to 5%. Therefore, if the above mentioned percentage exceed 5%, then there is at least 5% chance that the median of $r_{s,original}$ values is produced by randomness. Thus, it is considered as not significant, indicating no significant linear relationship between them. Secondly, to gain further information about the quality of the linear relationship between the variables, empirical distributions of the $r_{s,random}$ values are compared against $r_{s,original}$ values. For that purpose, a metric was constructed hereinafter referred to as uncertainty (*U*). *U* is defined as the overlapping area of the probability density functions (PDFs) fitted on the histograms of the $r_{s,original}$ and $r_{s,random}$ values, respectively. For this, we used kernel density estimation (KDE, Rosenblatt 1956; Davis et al. 2011), which is a nonparametric method to estimate the PDF of a continuous random variable (Härdle et al. 1990). As kernel function the Gaussian kernel was used, and the kernel bandwidth was estimated by the Sheather and Jones (1991) method. The total area under the curve for any PDF is always equal to 1, as it represents the total probability. *U* is associated with the quality of the linear relationship between the variables as follows. *U* increases with increasing overlapping area that means less reliable linear relationship between the original variables. To distinguish the pairs of variables objectively, *k*-mean clustering algorithm (Lloyd 1957; MacQueen, 1967) was applied on their median of $r_{s,original}$ values and *U* values.

For the permutation test and the calculation of uncertainty (U), we constructed the required scripts using the R programming language. For the kernel density estimation (KDE) we used R package, namely stats package (R Core Team 2013).

2.3.3 Estimating the effects of the uncertainty of observational datasets on the evaluation of RCMs

Difference between the observational datasets can cause differences in the results on the evaluation of RCM simulations. To quantify this, we used the metric reduction of error (RE, Prömmel et al. 2010). The RE is applied because it is important to know not only how well they perform compared to the reference datasets, but also how much improvement they show compared to the LBC (e.g. reanalysis or GCM; Diaconescu and Laprise 2013; Xue et al. 2014). By using RE, we identify where the RCM simulation has an improvement compared to the LBC and to assess how these potential improvements depend on the selection of the reference datasets or variables.

RE is calculated as follows:

$$RE = 1 - \frac{RMSE[SIM, OBS]}{RMSE[LBC, OBS]} \qquad (2.6)$$

where SIM indicates the specific RCM simulation which is evaluated. In our study, it is the RegCM with 24 different simulations. OBS is the reference dataset for the evaluation, in our case the CarpatClim or the E-OBS, while the ERA-Interim is used as LBC.

The range of the RE is ($-\infty$,1]. Negative RE values means that the RCM simulations and the observational dataset are less similar than the LBC and the observational dataset. Therefore, RE indicates no improvement of the RCM simulation relative to the LBC. When RE value is 0, it means the same performance for the RCM simulations and for the LBC relative to the observational dataset. Positive RE values mean that the RCM simulations and the observational dataset are more similar than the LBC and the observational dataset, which express an improvement of RCM simulations compared to the LBC. When RE value is 1, it means the RCM simulation reproduces the observational data perfectly.

RE was calculated for daily precipitation and temperature as well in each grid cell by using the E-OBS and CarpatClim as reference datasets for all 24 RegCM simulations. To overview the RE values obtained from the 24 RegCM simulations, we chose the maximum of the 24 RE values over every grid cell for both observational datasets and both variables. Note that, before calculating RE, days with below 1 mm precipitation were omitted. This threshold corresponds to standard recommendations for station data (Hofstra et al. 2009), and it is also necessary because it rains too lightly and too frequently in many climate models (Stephens et al. 2010; Maraun 2013).

The dependency of the RE on the chosen reference dataset was examined by calculating the correlation values ($r_{s,original}$ and $r_{s,random}$) between the RE and the variables (E, VE, PR, TAS, PR_ST, TAS_ST) based on Eq. 2.5. Significance test was carried out similarly as described in Section 2.3.2 (calculating U too). Finally, k-means clustering based on the median of $r_{s,original}$ values and U values was used to distinguish between RE and the variables.

The entire complex method defined in this study is summarised in Fig. 3 indicating the step-by-step procedures of the detailed analysis applicable to the comparison of different datasets.

2.4 Correction of E-OBS dataset with respect to precipitation

In order to carry out the analysis described in Section 2.3, a dataset correction was necessary. When the r_t and the $RMSE$ values were calculated between the gridded time series of the CarpatClim and the E-OBS, a major discrepancy (r_t>0.6 and $RMSE$>3-4 mm/day) was detected between Serbia and its neighbouring countries (Fig. 4c and e). This is caused by the fact that in the case of Serbian stations, the precipitation time series are shifted by 1 day backward compared to other domains in E-OBS, which is probably due to the different date assigning rule applied to daily precipitation totals. This time shift was missing for Serbian precipitation data from 2009 onward (it was applied prior to this date). To fix the problem in this paper, the E-OBS data were shifted forward 1 day in the grid cells of a masked area. We defined the mask based on two correlation fields. The first field contains r_t values that are calculated between the CarpatClim and the non-shifted E-OBS. The second field contains correlation coefficients that are calculated between the CarpatClim and the shifted E-OBS datasets. The two correlation fields are compared to each other in every grid cell. The mask is created from grid cells in which the second r_t is larger than the first r_t. We used the corrected E-OBS precipitation time series for the further analysis. Figure 4d and f clearly indicates that the 1-day forward shift of the E-OBS time series results in better similarity between the CarpatClim and the E-OBS. Note that shifting the time series would not improve $RMSE$ and r_t values in other regions in the examined domain (e.g. in Ukraine).
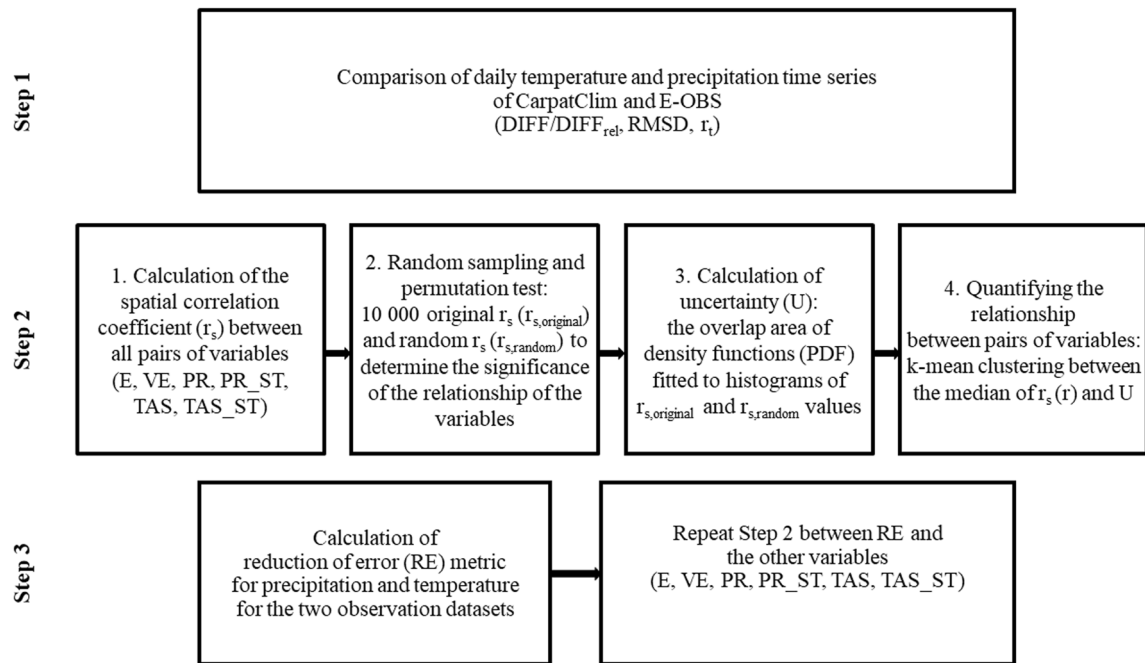
Fig. 3 The main steps of the method

3 Results and discussion

3.1 Comparing observational datasets: the examination of the temporal and the spatial distribution of variables

3.1.1 Comparison of the E-OBS and CarpatClim observational datasets

In the followings, the largest similarities and dissimilarities between the E-OBS and CarpatClim datasets are presented with respect to the precipitation ($DIFF_{rel}$, $RMSE$, r_t) and temperature ($DIFF$, $RMSE$, r_t). Both datasets (CarpatClim and E-OBS) are based on observations. CarpatClim dataset contains much more observations than E-OBS on the one hand. On the other hand, E-OBS covers a longer period, and it is continuously updated unlike CarpatClim. Furthermore, the interpolation technique of CarpatClim has been developed specifically for the climates and sampling conditions in the Carpathian region (Spinoni et al. 2015). Therefore, Bandhauer et al. (2022) attributed a higher reliability to CarpatClim and evaluated daily precipitation in E-OBS (v19.0e) against the CarpatClim as the reference dataset for the Carpathian region. In case of the precipitation, we found that the $RMSE$ and r_t values over Serbia are similar to the neighbouring domains (<3 mm/day and >0.8, respectively) after the data correction described in Section 2.4 was carried out (Fig. 4c–e). This result proves that $RMSE$ and r_t values are important to detect dataset errors in observational data before using them as a

reference for RCM evaluation or in case of testing a newly developed meteorological dataset. For example, Sekulić et al. (2021) developed a meteorological dataset at a 1-km spatial resolution across Serbia (MeteoSerbia1km), and when they compared the daily precipitation data to E-OBS, they found similar differences between the two datasets as those can be seen in Fig. 4e. We assume that their result is affected by data shift in E-OBS.

$DIFF_{rel}$ values are small (mainly between –10% and 10%) over Serbia suggesting that the amount of precipitation obtained from the E-OBS is similar to the amount of precipitation obtained from the CarpatClim in this region (Fig. 4a–b). We found similar results over Czechia, over the Carpathians in Romania, over the south-western part of Slovakia, and over the north-western part of Hungary.

The largest $RMSE$ values (~8 mm/day) and the weakest r_t values (<0.6) appear over the Ukrainian Carpathians in the analysed domain. The largest $DIFF_{rel}$ values appear also in this region, where the underestimation of precipitation in case of the E-OBS reaches 50% compared to the CarpatClim. Since the variation in topography is much higher in mountainous areas, the above-mentioned features can be considered as a consequence, which is enforced by the lack of appropriate number of stations resulting in a less detailed precipitation field in E-OBS. It is worth mentioning that despite the smaller $DIFF_{rel}$ (between –10% and 10%) and $RMSE$ values (~4 mm/day) over the eastern part of Ukraine compared to the Ukrainian Carpathians, the r_t values are relatively weak over the whole country (0.5–0.8).
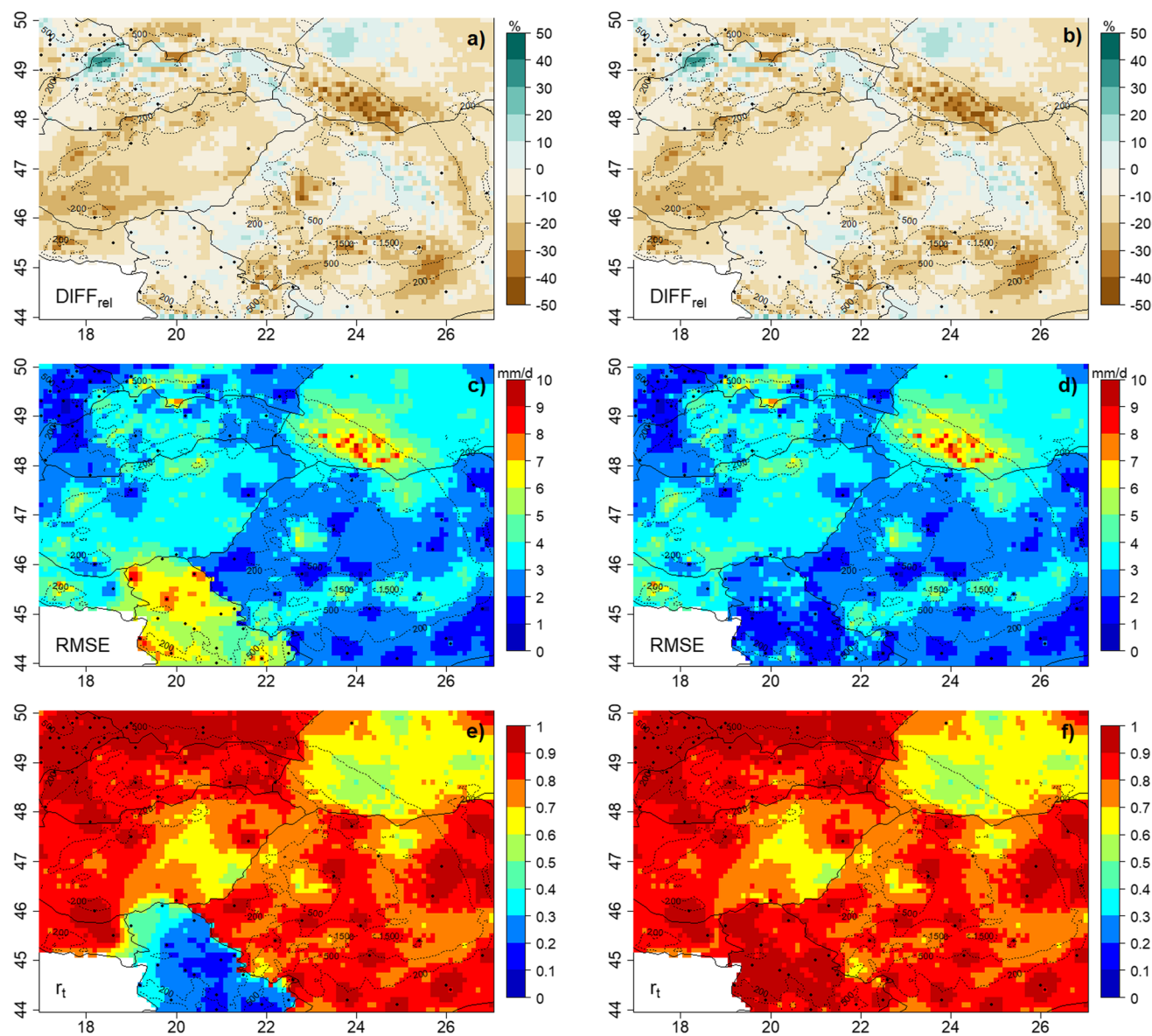
Fig. 4 Relative difference (DIFF$_{rel}$, **a**–**b**), RMSE (**c**–**d**) and temporal correlation (r_t, **e**–**f**) between CarpatClim and E-OBS for precipitation in 2010. Maps of the left (right) column are based on the stations from E-OBS before (after) the dataset correction. The black dots indicate the precipitation stations from E-OBS dataset. (Isolines represent topography.)

Over most of the territory of Hungary, *RMSE* and *DIFF$_{rel}$* values are as large as in Ukraine (~4 mm/day and between −10% and −30%, respectively) compared to Czechia, Serbia, and Romania. The r_t values over the south-eastern parts of Hungary are similarly weak as over western Ukraine (0.5–0.7). This is caused by the sparse station density over these regions in E-OBS dataset, which implies that E-OBS represents the temporal and spatial distribution of the precipitation much worse compared to CarpatClim.

Differences between the E-OBS and CarpatClim datasets often follow country borders, namely, the r_t and *RMSE* values follow the Ukrainian-Romanian border and the

Hungarian-Romanian border. Furthermore, the *RMSE* values follow the Serbian-Croatian and the Serbian-Hungarian border even after the data correction described in Chapter 2.4. The main reason behind these discrepancies is probably the unequal distribution of measuring stations (Fig. 2) and data policy. Each participating country of the CarpatClim project exchanged data only with neighbouring countries in case of stations within a belt of 50 km from their borders (Spinoni et al. 2015), which could affect data homogenization along the borders. This issue with the borders appeared in other studies, which focus on the CarpatClim dataset (e.g. Kis et al. 2015; Ács et al. 2021; Bandhauer et al. 2022).

Furthermore, the topography near the borders often changes. Consequently, the reduced density of stations is not able to capture the influence of topography on the climatic parameters (Sidău et al. 2021) which results in relatively large *RMSE* values.

Note that, the *RMSE* values between the E-OBS and CarpatClim are evidently close to zero in areas where both datasets contain the same precipitation stations (Fig. 4). This result is similar to Ly et al. (2011), who also examined *RMSE* over Belgium with different observational datasets. They found that the values of points close to the sample points were more likely to be similar than those that are further apart

Unlike in case of precipitation, major differences between the E-OBS and CarpatClim cannot be detected when temperature is analysed (results vary between –1 °C and 1 °C over the plains, Fig. 5a). More specifically, the results do not follow country borders. The distribution of stations with temperature measurements is more uniform than the distribution of stations with precipitation measurements, although it does not imply enhanced station density. However, temperature varies in a smaller extent than precipitation.

The absolute value of *DIFF* is large over the mountains, but both positive and negative *DIFF* values (–5 °C and 5 °C) are presented. It could be caused by the difference between the datasets, namely E-OBS contains area-mean temperature over the grid cell, while CarpatClim contains point value, which causes the bigger differences over a complex topography due to the high elevation variability within the grid cell. In general, the *RMSE* values are smaller than 2 °C and the r_t values between the two datasets are close to 1 due to the overall dominance of the annual course in temperature, but some dependencies on topography is observed (Fig. 5c). The *RMSE* values are larger (~4 °C), and the r_t values are slightly weaker (~0.95) over the mountains than over the plains. These can be explained by the same characteristics as in the case of *DIFF*.

3.1.2 Spatial distribution of the variables

The spatial distribution of the variables (*VE, PR, PR_ST, TAS, TAS_ST*) for CarpatClim and E-OBS is shown in Fig. 6 (for the variable *E*, Fig. 2).

According to Fig. 6, the two observational datasets show similar spatial distribution for *PR* and *TAS*. In case of *PR*, the higher values occur over the mountains, and the precipitation decreases from west to east. The effects of the altitude (orographic enhancement) and the distance from the Mediterranean Sea and Atlantic Ocean influence the precipitation amount. Among those, the Atlantic Ocean exerts the largest effect on precipitation (Bihari et al. 2018). The lowest precipitation values
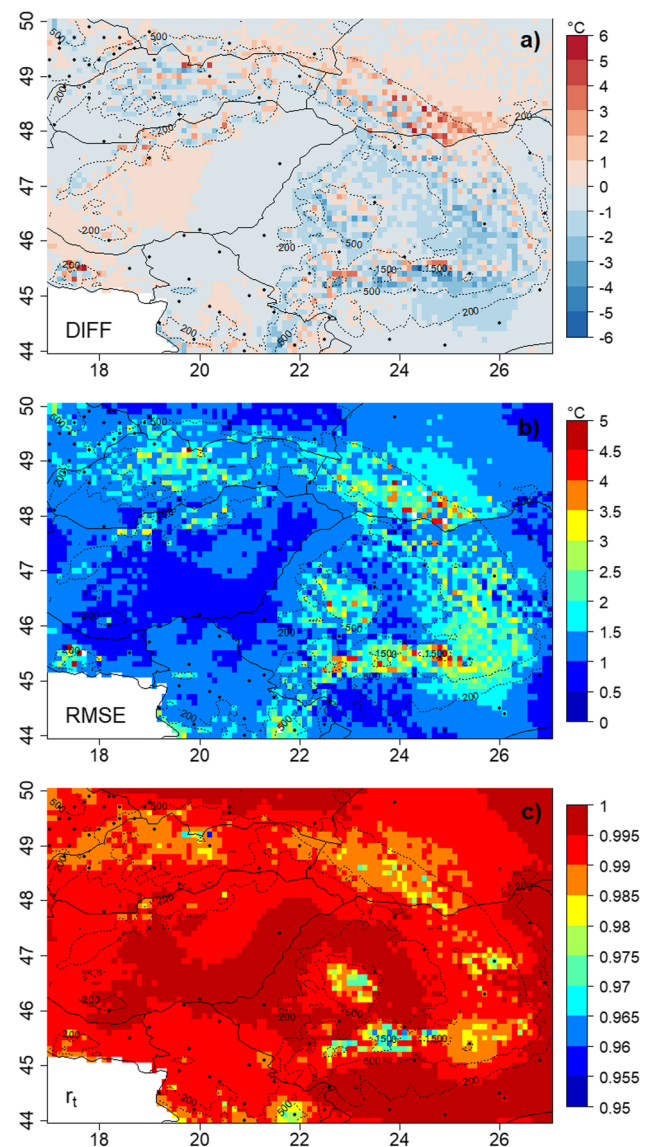

Fig. 5 The difference (DIFF, **a**), RMSE (**b**) and temporal correlation coefficient (r_t, **c**) between E-OBS and CarpatClim for daily temperature in 2010. The black dots indicate the temperature stations from E-OBS (Isolines represent topography.)

occur in the eastern part of the domain with ~600 mm in CarpatClim and ~500 mm in E-OBS.

In the case of *TAS*, the topographical features can be clearly seen (Fig. 6). Lower mean values appear at higher elevations in both datasets. The warmest area (13 °C) has a larger extent over the lowlands in CarpatClim than in E-OBS. The western part of the domain and Ukraine are slightly warmer (by ~1 °C) in CarpatClim than in E-OBS, but CarpatClim is colder over the mountains.

Maxima of *PR* at Carpathians are underestimated and oversmoothed, while TAS is oversmoothed in E-OBS compared to CarpatClim, which is probably a direct consequence

Fig. 6 The spatial distribution of variability of elevation (VE), annual total precipitation (PR), effect of station density for precipitation (PR_ST), annual mean temperature (TAS) and effect of station density for temperature (TAS_ST) in Carpat-Clim (left) and E-OBS (right) in 2010. (Isolines represent topography)
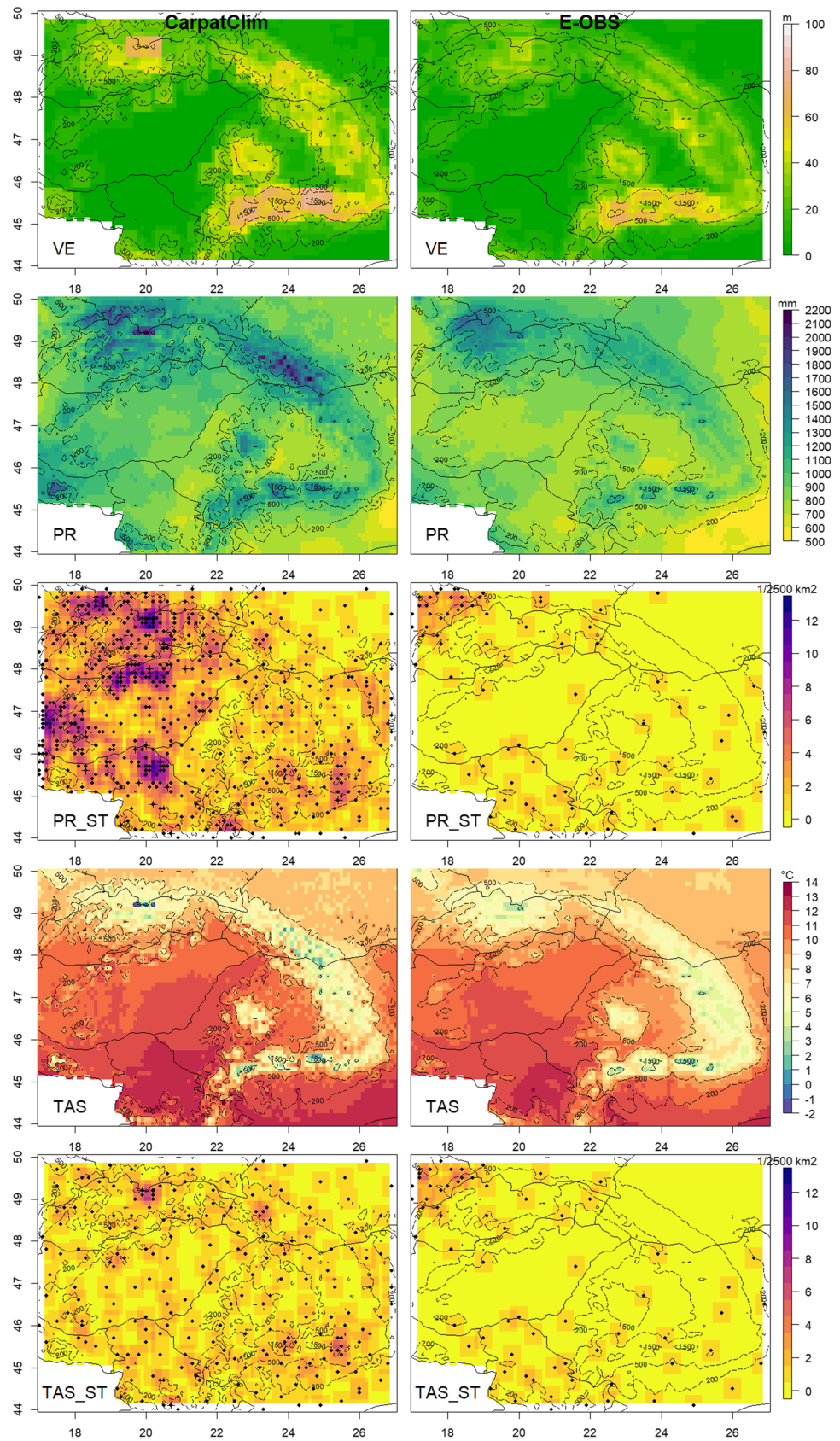
of the lower underlying network density at high elevation areas in E-OBS (Kotlarski et al. 2019; Bandhauer et al. 2022). In areas where E-OBS relies on dense observations, the agreement with CarpatClim is much better (e.g. High Tatras in the northern part of Slovakia).

VE is higher in CarpatClim than in E-OBS, especially over the Southern Carpathians and Western Carpathians (~80 m in CarpatClim and ~60 m in E-OBS). These differences are caused by the fact that the elevation field of CarpatClim is more detailed than that of E-OBS (Fig. 2 and Table 2).

According to Fig. 6, *PR_ST* is larger in CarpatClim than in E-OBS, mainly in the western part of the domain, where the number of stations is ~11/2500 km^2 in CarpatClim and ~1/2500 km^2 in E-OBS. The station coverage of the Eastern Carpathians is relatively homogeneous in CarpatClim, while E-OBS contains only a few stations over this region and there are no stations in the Ukraine, which could cause uncertainty in the results. In E-OBS dataset, the largest station density occurs over Czechia with ~4/2500 km^2.

The *TAS_ST* derived from CarpatClim covers the whole area homogeneously. The pattern of E-OBS for *TAS_ST* is very similar to *PR_ST*. The densest part in E-OBS is over Czechia with 3/2500 km^2.

3.2 Comparing observational datasets: analysis of the relationships between the variables

The strength and reliability of the linear relationship between the pairs of variables with respect to precipitation and temperature are assessed in the followings.

Based on the distributions of the correlation values ($r_{s,original}$ and $r_{s,random}$) presented in Fig. 7, there are significant relationships between *PR* and *E*, *PR* and *VE* and *PR* and *PR_ST*, respectively. It can be noted that the median of $r_{s,original}$ values is close (<0.01) to the r_s values between the non-sampled pairs of variables (where all the 5895 grid cells are considered, hereafter median of $r_{s,original}$ values is referred to as r_s).

The strongest correlation is detected between *PR* and *E* (0.57 in CarpatClim and 0.52 in E-OBS) which indicates that the mountains affect precipitation, as the orographic lifting of air masses favours condensation and cloud formation (Smith 1979). The *U* value associated to *PR* and *E* is less than 1% in both datasets indicating reliable relationship between the variables. The relationship between *PR* and *VE* is weaker (r_s=0.48 in CarpatClim and r_s=0.36 in E-OBS) and the associated *U* is still under 5%. The relatively large difference (0.12) between the correlations obtained from CarpatClim and E-OBS and stronger correlations in CarpatClim than in E-OBS are explained by the fact that CarpatClim contains more stations over the mountains. Therefore, *PR* is more realistic in CarpatClim

than in E-OBS (Table 2 and Fig. 6). This highlights the need for large number of stations if regional and local scale precipitation features are of interest, especially in mountainous regions. The r_s value between *PR* and *PR_ST* is only significant (0.23) in CarpatClim with increasing *U* (~28%). The sparse station density in E-OBS causes non-significant r_s values and larger *U* values. In general, interpolation accuracy decreases as the station density decreases, and it is less accurate for variables with greater spatial variability (e.g. precipitation) and over complex topography (Hofstra et al. 2009).

No significant linear relationships were detected between E and *PR_ST* and between *VE* and *PR_ST*. These results depict that the locations of the stations do not depend on orography but instead on historically existing settlements.

Large negative r_s values appear between *E* and *TAS* (r_s=–0.86 in CarpatClim and r_s=–0.9 in E-OBS, Fig. 8), because air temperature decreases with elevation. The associated *U* values are close to 0%. Significant, but weaker correlations are detected between *VE* and *TAS* (r_s=–0.57 in CarpatClim and r_s=–0.61 in E-OBS) with the associated *U* values ~0.1%. There are slight differences between the r_s values obtained from the two observational datasets, which may come from the different algorithms used in CarpatClim and E-OBS to derive temperature (Sections 2.2.1-2.2.2).

The relationship between *TAS_ST* and *VE* is remarkably different if CarpatClim or E-OBS is analysed. The r_s value obtained from E-OBS is not considered as significant (r_s≈0), while weak but significant r_s value is detected in case of CarpatClim (r_s=0.2). The associated *U* values are ~88% and 37%, respectively. No significant relationships were identified in cases of other pairs of variables (*TAS_ST-E*, *TAS_ST-TAS*).

In summary, the relationship between *PR* and *VE* is slightly different, and the relationship between *PR* and *PR_ST* is significantly different if CarpatClim or E-OBS is examined. It is clear that station density affects the spatial distribution of temperature to a lesser extent than elevation. Significant difference between the two datasets is detectable only in case of one pair of variables (*TAS_ST-VE*). These differences are further examined by k-mean clustering, which results are presented in the scatterplots in Fig. 9.

Two clusters are detected in all the four cases (for precipitation and temperature obtained from CarpatClim and E-OBS). The first cluster contains pairs of variables, which relationships are considered reliable, i.e. strong correlations (r_s>0.4 and r_s<–0.4) and small *U* values (*U*<30%). The second cluster contains pairs of variables, which relationships are considered less reliable, i.e. weak correlations (–0.2<r_s<0.2) and large *U* values (*U*>30%).

Figure 9 shows that the pair of variables *PR-PR_ST* belong to different cluster depending on the analysed observational datasets, namely to the first cluster in
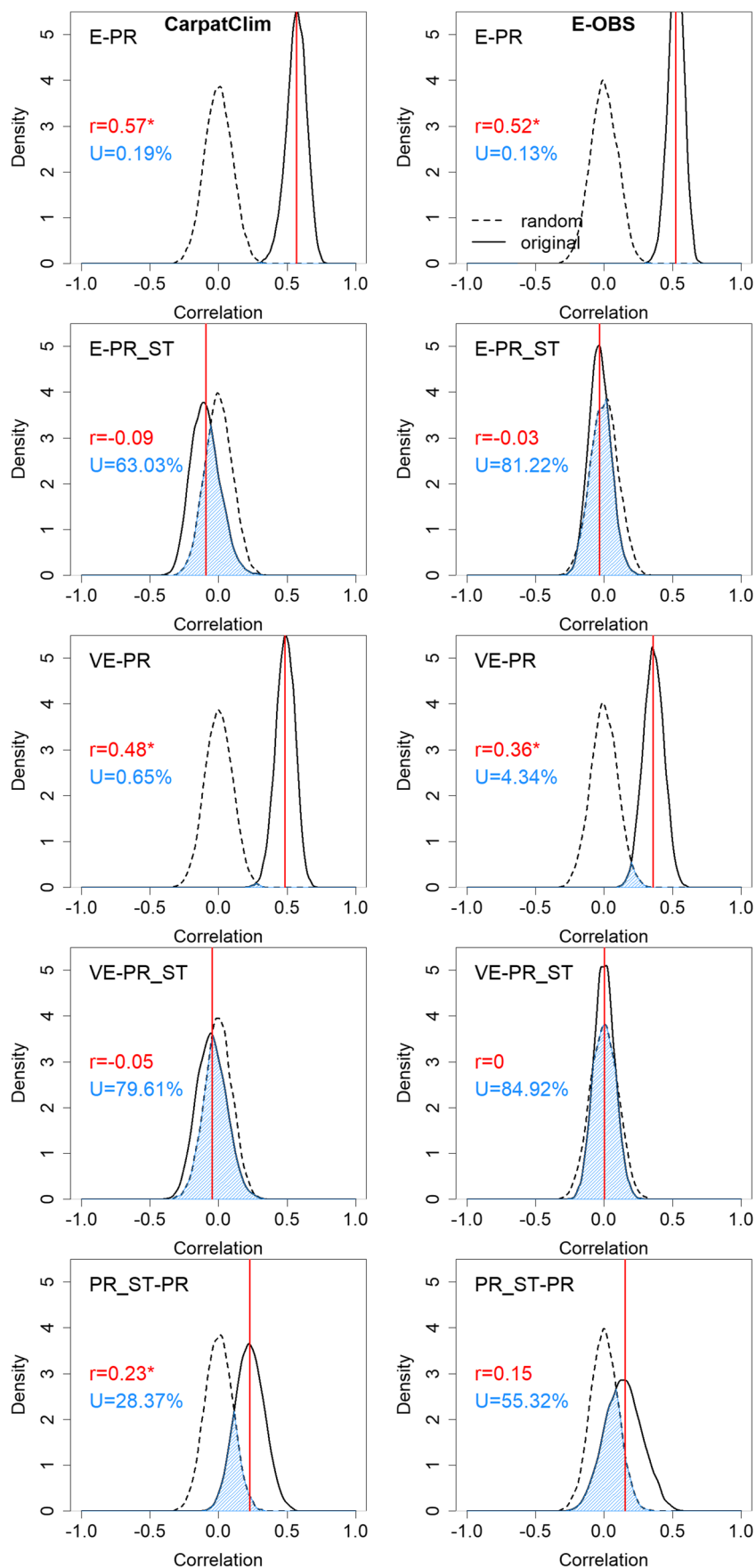
Fig. 7 Probability density functions (PDFs) of the original and random correlation coefficients ($r_{s,original}$ and $r_{s,random}$) based on sampled datasets for precipitation in CarpatClim (left) and E-OBS (right). The red vertical lines and r denote the median of the $r_{s,original}$ values. The asterisks indicate significant correlations at the significance level of 0.05. The blue shaded area and uncertainty (U) indicate the overlapping area of the PDFs, and it is expressed in percentage. The method with its interpretation is described in detail in Section 2.3.2.

Fig. 8 Probability density functions (PDFs) of the original and random correlation coefficients ($r_{s,original}$ and $r_{s,random}$) based on sampled datasets for temperature in CarpatClim (left) and E-OBS (right). The red vertical lines and r denote the median of the $r_{s,original}$ values. The asterisks indicate significant correlations at the significance level of 0.05. The blue shaded area and uncertainty (U) indicate the overlapping area of the PDFs, and it is expressed in percentage. The method with its interpretation is described in detail in Section 2.3.2.
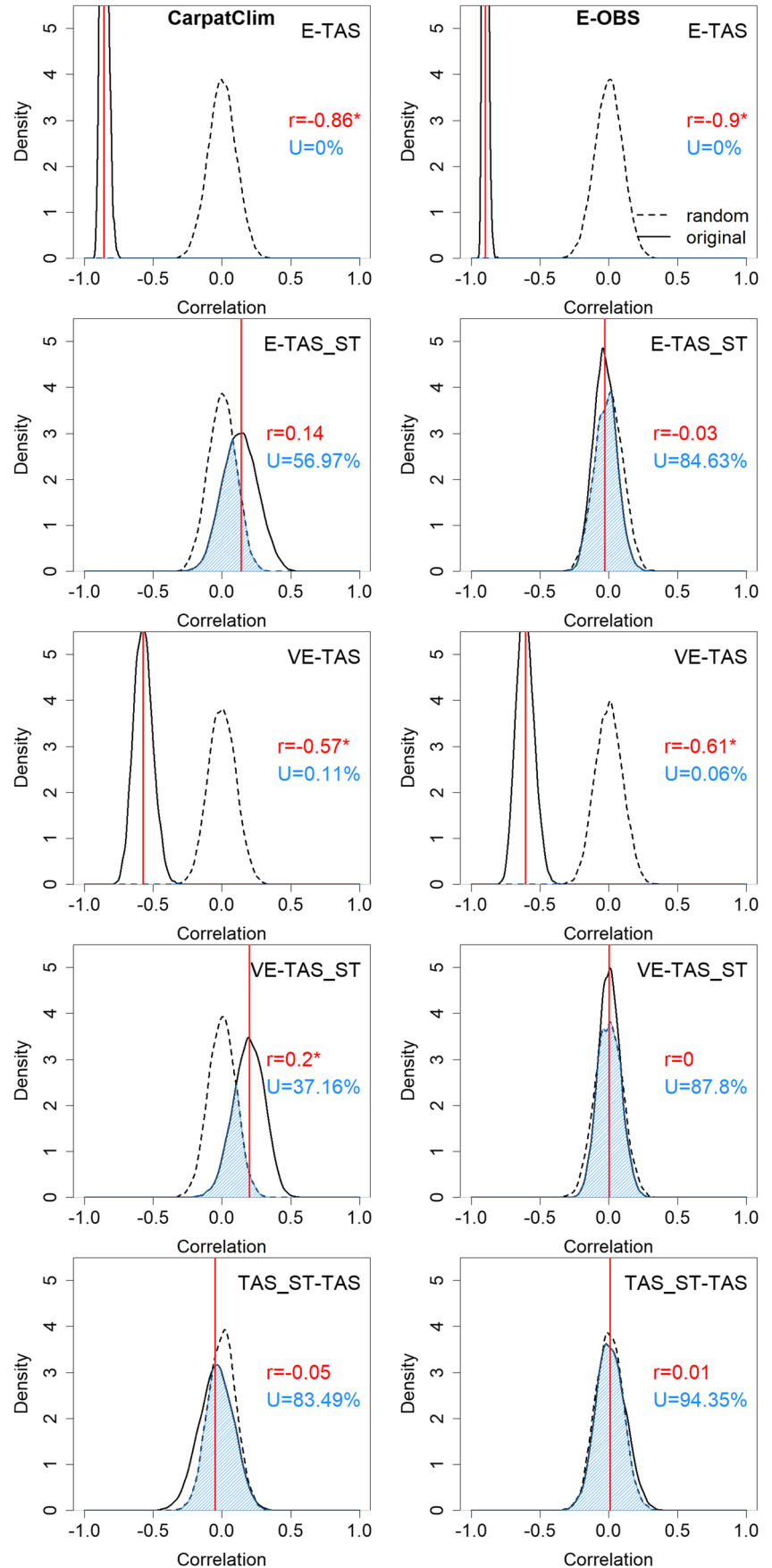
Fig. 9 Scatterplots for uncertainty (U) and the median of the $r_{s,original}$ values (r) obtained from CarpatClim (left) and E-OBS (right) for precipitation (**a–b**) and temperature (**c–d**). Two clusters of the pairs of variables are detected in all four cases: the first cluster (which contains pairs of variables with reliable relationship) is distinguished from the second cluster (which contains pairs of variables with less reliable relationship) with underlined variables. The green (red) dots show the pairs of variables with significant (non-significant) r values at a level of 0.05
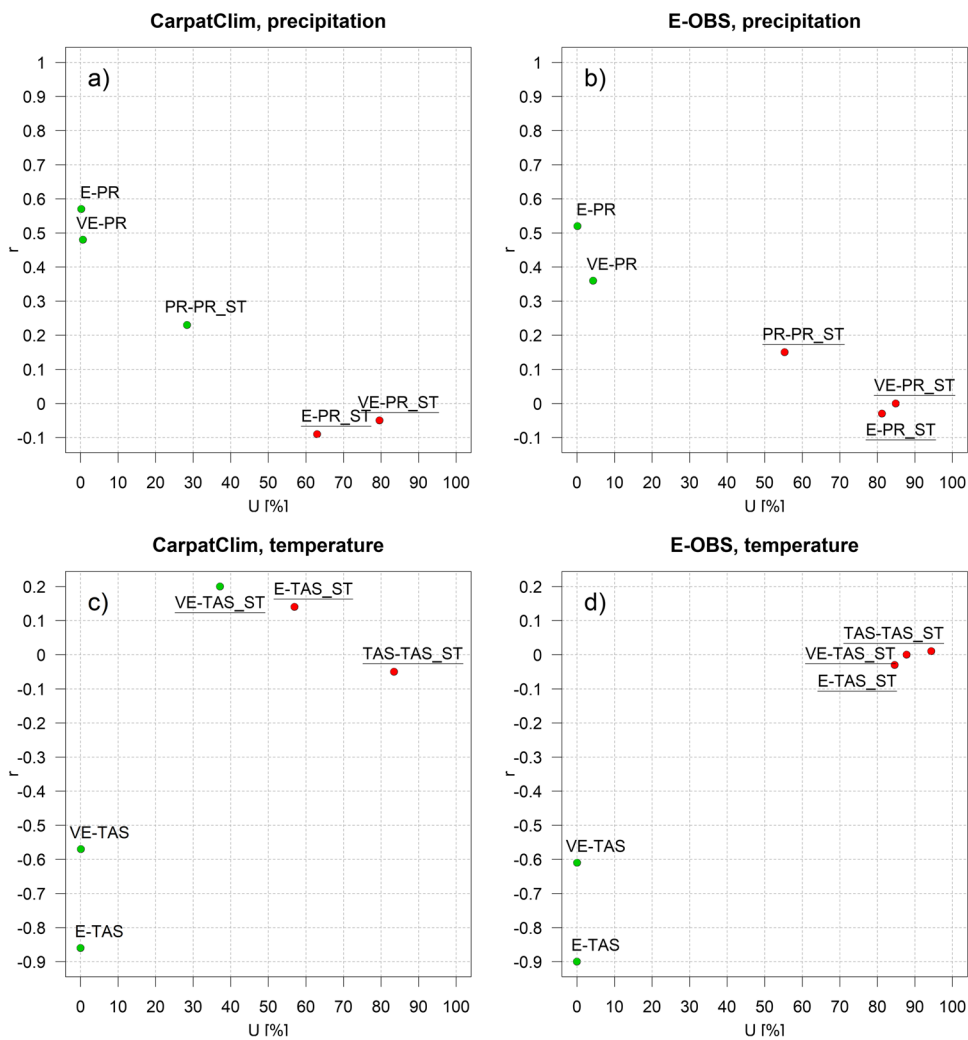
3.3 Results obtained from the estimation of the effects of uncertainty of observational datasets on the evaluation of RCMs

As already shown above, the relationships between the variables are affected by the observational datasets. The effect of the selected observational dataset on the assessment of the climate simulations of the RegCM was quantified by the metric *RE* which is shown in Fig. 10.

Possible improvements of RegCM simulations compared to ERA-Interim can be detected for precipitation (Fig. 10a–b) and for temperature (Fig. 10c–d) in cases of the two observational datasets (CarpatClim and E-OBS). The simulations show improvements (positive *RE* values) over areas that are more densely covered with stations and lowlands in case of precipitation. However, if we compare the *RE* values obtained from E-OBS and CarpatClim for precipitation, significant differences can be found. The area with positive *RE* is larger in CarpatClim than E-OBS, because the former contains more stations than the latter. Larger negative *RE* values appear over mountain ranges and peaks in E-OBS (Fig. 10b), which indicates that the evaluation over those regions shows large uncertainty because of the sparse observational network in E-OBS. The greatest difference between the *RE* fields for the two observational datasets also appears over the mountains in Ukraine, where the results show an improvement against ERA-Interim in CarpatClim, but not if we compare the RegCM simulation results to E-OBS (Fig. 10a–b). According to this result, climate simulations must be evaluated carefully over the mountains, because this uncertainty in observational data could lead to significant differences. This outcome is confirmed by the significant positive r_s values between *RE* and *PR_ST* in case of CarpatClim (~0.3) and by significant negative r_s values between *RE-E* and *RE-VE* in E-OBS (~–0.3), according

CarpatClim and to the second cluster in E-OBS. Although, the correlation is significant between *TAS_ST-VE* only in CarpatClim, it belongs to the second cluster in both observational datasets.
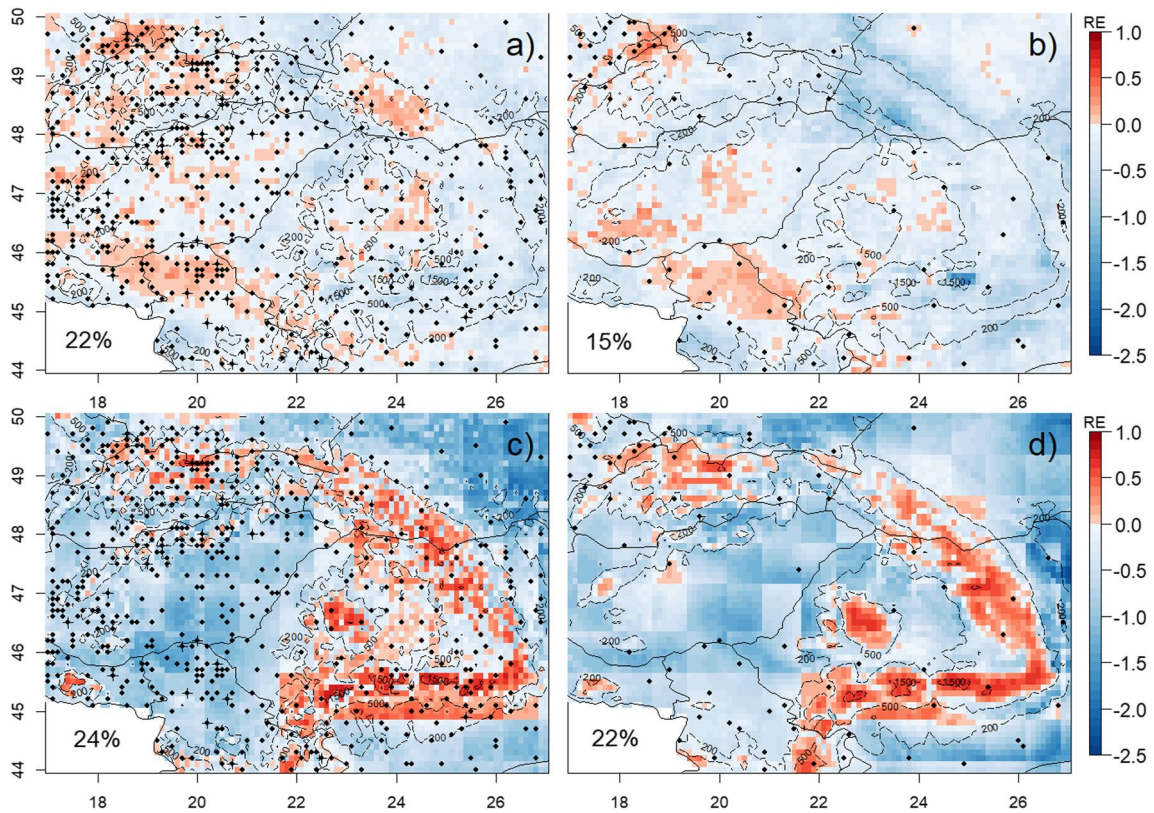
Fig. 10 Reduction of error (RE) for precipitation (**a–b**) and temperature (**c–d**) for CarpatClim (left) and E-OBS (right). The dots (one station on a grid cell) and crosses (two stations on a grid cell) indicate the stations. The percentage shows the rate of the positive RE value. (Isolines represent topography.)

to Table 4, part a. We assume that this is the effect of the sparse observational network over these regions (especially over mountainous area), which cannot represent the precipitation adequately. Uncertainties in observational datasets tend to decrease in regions where all datasets have a high station density (approximately 22% for CarpatClim and 15%

Table 4 The median of the $r_{s,original}$ values and uncertainties in % (in parenthesis) between RE and the variables E, VE, PR and PR_ST in section (a) and E, VE, TAS and TAS_ST in section (b). The numbers with asterisk indicate significant correlation coefficients at the level of 0.05

(a)	Precipitation		(b)	Temperature	
	CarpatClim	E-OBS		CarpatClim	E-OBS
E	−0.07	−0.26[*]	E	0.62[*]	0.59[*]
	(76.80%)	(22%)		(0.10%)	(0.10%)
VE	−0.14	−0.31[*]	VE	0.6[*]	0.64[*]
	(51.50%)	(12.80%)		(~0%)	(~0%)
PR	0.19	−0.09	TAS	−0.44[*]	−0.43[*]
	(51.50%)	(67.60%)		(2.90%)	(2.60%)
PR_ST	0.32[*]	0.19	TAS_ST	0.19	0.02
	(9.50%)	(31.70%)		(36.90%)	(90.40%)

for E-OBS from the total Carpathian region). This highlights the need for high station densities if regional and local scale precipitation features are of interest, especially in mountainous regions. Without such high station density behind gridded reference datasets, one cannot be certain whether RCM simulations have bias, or they represent reality, but the reference dataset is not detailed and accurate enough.

For temperature, the largest positive values of *RE* are found over the mountains in case of both datasets: the higher resolution of the simulations has improvement compared to the ERA-Interim, especially in regions with the most complex orography (Fig. 10c–d) in accordance with the results of Prömmel et al. (2010).

Differences between *RE* values obtained from E-OBS and CarpatClim for temperature are smaller than in case of precipitation which is confirmed by similar r_s values in Table 4, part b (r_s values for *RE-E* and *RE-VE* are ~0.6 and for *RE-TAS* are ~−0.4 for both datasets). Negative r_s values between *RE* and *TAS* imply that *RE* values are high in the mountains. However, differences can be visually detected if the area covered by positive *RE* values are examined, namely, this area is larger in CarpatClim than in E-OBS (24% and 22%, respectively).

Fig. 11 Scatterplots for uncertainty (U) and the median of the $r_{s,original}$ values (**r**) obtained from CarpatClim (left) and E-OBS (right) for precipitation between RE and the variables E, VE, PR and PR_ST (**a–b**) and for temperature between RE and the variables E, VE, TAS and TAS_ST (**c–d**). The 1st cluster is distinguished from the 2nd cluster with underlined variables and from the 3rd cluster with dotted underlined variables. The green (red) dots show the pairs of variables with significant (non-significant) r values at a level of 0.05
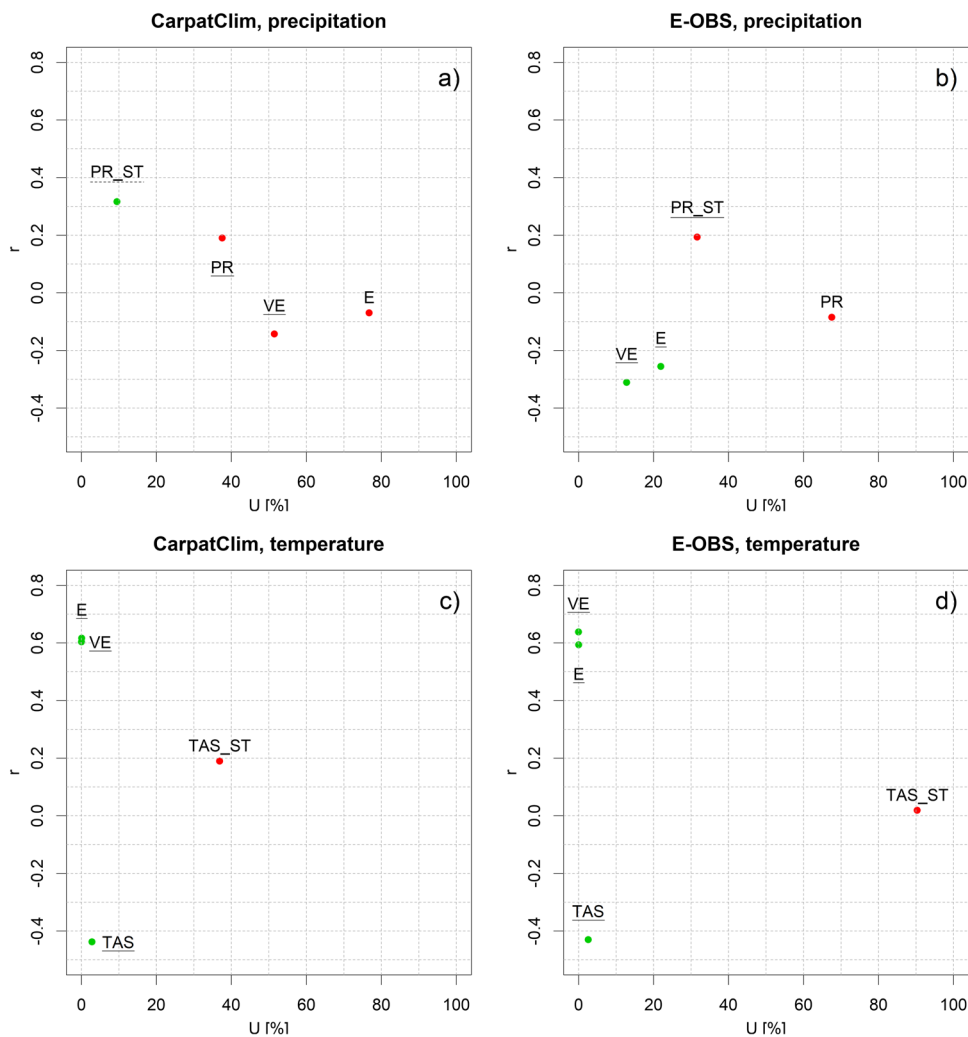
The scatterplots in Fig. 11 between r_s and U values (Table 4) shows that the strengths of the relationships between the *RE* and variables are different in the two datasets.

The strongest relationship (the smallest *U* and the largest r_s) for precipitation appears between *RE-PR_ST* in CarpatClim and *RE-VE* in E-OBS (in both cases *U* is ~10%). Meanwhile, the weakest relationship (the largest *U* and the smallest r_s) can be found between *RE-E* in CarpatClim and *RE-PR* in E-OBS (the *U* is 76% in CarpatClim and 67.6% in E-OBS).

For the temperature, the *U* is very small (<1%) between *RE* and *E*, between *RE* and *VE* and between *RE* and *TAS* in both datasets. While the weakest relationship occurs between *RE-TAS_ST* pair in both datasets, but the difference between the *U* values is ~42% in favour of CarpatClim. The low station density in E-OBS can cause this high uncertainty compared to CarpatClim. According to *RE*, the key variable is different for precipitation and temperature: *PR_ST* is crucial for precipitation, while *E* has the greatest effect on temperature.

The results of *k*-means clustering show different number of clusters for the two datasets for precipitation, namely three clusters for CarpatClim and two clusters for E-OBS (Fig. 11a–b). *RE* has considerable relationship only with *PR_ST* in CarpatClim, which is the first cluster. The second cluster contains *PR* and *VE*, while *E* is in the third cluster. The relationship between *RE* and other variables in E-OBS is less obvious than in CarpatClim. The first cluster contains significant and non-significant pairs of variables as well concerning the r_s values (significant: *RE-VE*, *RE-E*; non-significant: *RE-PR_ST*) in E-OBS. Only *PR* belongs to the second cluster, and this relationship is not reliable at all.

For temperature, there are two clusters in both observational datasets, and the members of the groups are the same. *RE-E, RE-VE, RE-TAS* pairs are in the first group and all of the r_s between the *RE* and the variables are significant (Fig. 11c–d). This result proves that the dynamical downscaling is important over complex topography. The second cluster contains only *TAS_ST*, which means that the location of stations does not have direct effect on *RE*.

Our results show that these improvements depend on the climate variable, topography, reference dataset, and applications using the RCM output. Prömmel et al. (2010) published similar results over the Alps with the REMO RCM, but analysed only temperature. Our study extends previous analyses because it focuses on the Carpathian region including both the mountainous and plain areas and contains additional variables besides temperature and precipitation.

Our results exhibit the potential improvements of RegCM simulations against the driving data according to *RE* values. As it can be seen in Fig. 11, the dependency of the *RE* on the chosen reference dataset can be clearly determined by calculating r_s between *RE* and the variables and the associated *U* values.

Kotlarski et al. (2019) mentioned the uncertainties in the observational reference datasets directly translate into uncertainties in model evaluation results. Our results confirm the importance to assess the relationships between all available variables for quantifying the uncertainties in the datasets, as using different observational datasets can lead to different evaluation results especially in case of precipitation.

4 Conclusions

A specific novel evaluation method was introduced in this study which combines widely known metrics and statistical techniques (e.g. comparison of spatiotemporal distributions by *DIFF*, *RMSE* and r_t, applying the metric *RE* and *k*-means clustering) to quantify the uncertainties in the observational datasets and how these uncertainties affect the evaluation of RCM simulations. Besides precipitation and temperature, our method uses geographic variables (e.g. elevation, variability of elevation, effect of station) that are considered as uncertainty sources. The method was applied to the observational datasets CarpatClim and E-OBS and to the RegCM simulations driven by ERA-Interim based on 2010 in the Carpathian region. 2010 was the wettest year in this area since the beginning of regular measurements, and thus, the climate simulations for such extreme conditions can provide important validation results to be used in impact studies later. Through our comprehensive analysis, we pointed out that the analysis of the time series of the variables from the observational datasets is useful for error detection as well.

Significant differences were found between the observational datasets. The spatial distribution of the examined climatic and geographical variables shows that CarpatClim is wetter over the whole region (mostly over the mountains, where the difference could be up to 50%) than E-OBS, because of the much lower number of stations in E-OBS. The temperature fields are similar in the two datasets; however, E-OBS is a little warmer than CarpatClim over the mountains; and the representation of orography is more detailed

in CarpatClim than in E-OBS. However, a shortcoming in CarpatClim is the appearance of the borders between some countries (e.g. between Hungary and Romania, Hungary and Ukraine), which may result from the inhomogeneities in the data and lower station density in Romania and Ukraine. The higher differences between the datasets in the mountainous areas can be associated with the different grid representation of the datasets, namely, E-OBS uses area-mean values for the grid cells, whereas CarpatClim contains point-values.

In accordance with previous studies, we found that the influence of observational uncertainty is larger for precipitation than for temperature. However, CarpatClim is certainly more reliable compared to E-OBS in case of precipitation, because it is based on a greater number of stations than E-OBS. The difference between the two datasets was not as remarkable for temperature as for precipitation, but the altitude dependence of temperature is a little stronger in E-OBS than in CarpatClim (0.9 vs. −0.86, respectively).

The joint investigation of spatial correlations between the pairs of variables and the associated uncertainties was useful to distinguish the pairs of variables based on reliability of their relationships. We found that the topography is important in case of precipitation and in case of temperature as well, but the effect of station density has stronger relationship with precipitation and in case of CarpatClim. This difference may be caused by the reduced number of stations in the E-OBS.

This is the first time, where RE metric has been used for a detailed evaluation of observational datasets. Using *RE* metrics, we have showed that the choice of observation dataset has a substantial effect on the evaluation of RegCM simulations. For precipitation, RegCM has improvement compared to ERA-Interim where the station network is dense over mountains (e.g. in the Carpathians in Ukraine resulted in the increase of *RE* in case of CarpatClim compared to E-OBS), while the station density over lowlands is less important. Overall, 22% of the Carpathian region show improvement, when using RegCM simulations and validating against CarpatClim. For temperature, we found that in regions with the most complex orography, the high-resolution RegCM simulations clearly improve the representation of temperature compared to the ERA-Interim in both datasets (over 22-24% of the total Carpathian region).

The main conclusion of the paper is that even small differences between the reference datasets can cause significant differences in the RCM evaluation, which can be captured by the analysis of the relationships between the *RE* and the variables. Concerning the RegCM evaluation, the main differences based on the reference datasets are detected in case of precipitation where *PR_ST* has a stronger relationship with *RE* in CarpatClim. This indicates that a sufficiently large number of stations represent the spatial variability of precipitation and extreme values more accurately, which are crucial for RCM evaluation. The *E* and *VE* are in strong negative correlation with *RE* in E-OBS

indicating the sparse station network over the mountainous area cannot represent local scale precipitation features.

Following our results, we can evaluate RCM simulations properly, if observational uncertainties are considered, especially in a year with extreme precipitation events. We strongly encourage to use reference data sets with a high station density background. The higher the station density behind the reference data, the more reliable the validation procedure. Our method is beneficial not only for comprehensive comparison of observational datasets, but also for quantifying the differences and for error detection. We illustrated the use of the complex method on a special case study; however, the main message of the study is that it can be applied to other datasets, different time periods (even much longer) and areas with complex topography.

Code availability Not applicable.

Declarations

Ethics approval Not applicable

Consent to participate Not applicable

Consent for publication All authors consent to publish the study in a journal article.

Conflict of interest The authors declare no competing interests.

References

Ács F, Zsákai A, Kristóf E, Szabó AI, Feddema J, Breuer H (2021) Clothing resistance and potential evapotranspiration as thermal climate indicators—the example of the Carpathian region. Int J Climatol 41:3107–3120. https://doi.org/10.1002/joc.7008

Bacchi B, Kottegoda NT (1995) Identification and calibration of spatial correlation patterns of rainfall. J Hydrol 165:311–348. https://doi.org/10.1016/0022-1694(94)02590-8

Bandhauer M, Isotta F, Lakatos M, Lussana C, Båserud L, Izsák B, Szentes O, Tveito OE, Frei C (2022) Evaluation of daily precipitation analyses in E-OBS (v19.0e) and ERA5 by comparison to regional high-resolution datasets in European regions. Int J Climatol 42:727–747. https://doi.org/10.1002/joc.7269

Bartholy J, Radics K, Bohoczky F (2003) Present state of wind energy utilisation in Hungary: policy, wind climate, and modelling studies. Renewable Sustainable Energy Rev 7:175–186. https://doi.org/10.1016/S1364-0321(03)00003-0

Beck HE, Vergopolan N, Pan M, Levizzani V, van Dijk AIJM, Weedon GP, Brocca L, Pappenberger F, Huffman GJ, Wood EF (2017) Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. Hydrol Earth Syst Sci 21:6201–6217. https://doi.org/10.5194/hess-21-6201-2017

Bellprat O, Kotlarski S, Lüthi D, Schär C (2012) Objective calibration of regional climate models. J Geophys Res: Atmos D23:117. https://doi.org/10.1029/2012JD018262

Berg P, Norin L, Olsson J (2016) Creation of a high resolution precipitation data set by merging gridded gauge data and radar observations for Sweden. J Hydrol 541:6–13. https://doi.org/10.1016/j.jhydrol.2015.11.031

Bihari Z, Babolcsai G, Bartholy J, Ferenczi Z, Gerhátné Kerényi J, Haszpra L, Homoki-Ujváry K, Kovács T, Lakatos M, Németh Á, Pongrácz R, Putsay M, Szabó P, Szépszó G (2018) Climate. In: Kocsis K (ed) National Atlas of Hungary: Natural Environment. MTA CSFK Geographical Institute, Budapest, pp 58–69

Bissolli P, Friedrich K, Rapp J, Ziese M (2011) Flooding in eastern central Europe in May 2010 - reasons, evolution and climatological assessment. Weather 66:147–153. https://doi.org/10.1002/wea.759

Bretherton CS, Park S (2009) A new moist turbulence parameterization in the community atmosphere model. J Clim 22:3422–3448. https://doi.org/10.1175/2008JCLI2556.1

Ceglar A, Croitoru AE, Cuxart J, Djurdjevic V, Güttler I, Ivančan-Picek B, Jug D, Lakatos M, Weidinger T (2018) PannEx: the Pannonian basin experiment. Clim Serv 11:78–85. https://doi.org/10.1016/j.cliser.2018.05.002

Cheval S, Birsan MV, Dumitrescu A (2014) Climate variability in the Carpathian Mountains region over 1961–2010. Glob Planet Change 118:85–96. https://doi.org/10.1016/j.gloplacha.2014.04.005

Cornes RC, van der Schrier G, van den Besselaar EJM, Jones PD (2018) An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets. J Geophys Res Atmos 123:9391–9409. https://doi.org/10.1029/2017JD028200

Davis RA, Lii KS, Politis DN (2011) Remarks on some nonparametric estimates of a density function. In: Davis RA, Lii K-S, Politis DN (eds) Selected Works of Murray Rosenblatt. Springer, New York, New York, NY, pp 95–100. https://doi.org/10.1007/978-1-4419-8339-8_13

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ et al (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q J Meteorol Soc 137:553–597. https://doi.org/10.1002/qj.828

Dickinson RE, Henderson-Sellers A, Kennedy PJ (1993) Biosphere-atmosphere Transfer Scheme (BATS) version 1e as coupled to the NCAR community climate model (No. NCAR/TN-387+STR). University Corporation for Atmospheric Research. https://doi.org/10.5065/D67W6959

Di Luca A, Argüeso D, Evans JP, Elía R, Laprise R (2016) Quantifying the overall added value of dynamical downscaling and the contribution from different spatial scales. J Geophys Res Atmos 121:1575–1590. https://doi.org/10.1002/2015JD024009

Diaconescu EP, Laprise R (2013) Can added value be expected in RCM-simulated large scales? Clim Dyn 41:1769–1800. https://doi.org/10.1007/s00382-012-1649-9

Emanuel KA, Živković-Rothman M (1999) Development and evaluation of a convection scheme for use in climate models. J Atmos Sci 56:1766–1782. https://doi.org/10.1175/1520-0469(1999)056<1766:DAEOAC>2.0.CO;2

Fantini A, Raffaele F, Torma C, Bacer S, Coppola E, Giorgi F, Ahrens B, Dubois C, Sanchez E, Verdecchia M (2018) Assessment of multiple daily precipitation statistics in ERA-Interim driven Med-CORDEX and EURO-CORDEX experiments against high resolution observations. Clim Dyn 51:877–900. https://doi.org/10.1007/s00382-016-3453-4

Frei C (2014) Interpolation of temperature in a mountainous region using nonlinear profiles and non-Euclidean distances. Int J Climatol 34:1585–1605. https://doi.org/10.1002/joc.3786

Gao X, Xu Y, Zhao Z, Pal JS, Giorgi F (2006) On the role of resolution and topography in the simulation of East Asia precipitation. Theor Appl Climatol 86:173–185. https://doi.org/10.1007/s00704-005-0214-4

Gervais M, Tremblay LB, Gyakum JR, Atallah E (2014) representing extremes in a daily gridded precipitation analysis over the United States: impacts of station density, resolution, and gridding methods. J Clim 27:5201–5218. https://doi.org/10.1175/JCLI-D-13-00319.1

Giorgi F (1989) Two-dimensional simulations of possible mesoscale effects of nuclear war fires: 1. Model description. J Geophys Res Atmos 94:1127. https://doi.org/10.1029/JD094iD01p01127

Giorgi F (2019) Thirty years of regional climate modeling: where are we and where are we going next? J Geophys Res Atmos 24:5696–5723. https://doi.org/10.1029/2018JD030094

Giorgi F, Coppola E, Solmon F, Mariotti L, Sylla M, Bi X, Elguindi N, Diro G, Nair V, Giuliani G, Turuncoglu U, Cozzini S, Güttler I, O'Brien T, Tawfik A, Shalaby A, Zakey A, Steiner A, Stordal F et al (2012) RegCM4: model description and preliminary tests over multiple CORDEX domains. Clim Res 52:7–29. https://doi.org/10.3354/cr01018

Grell GA (1993) Prognostic evaluation of assumptions used by cumulus parameterizations. Mon Weather Rev 121:764–787. https://doi.org/10.1175/1520-0493(1993)121<0764:PEOAUB>2.0.CO;2

Grenier H, Bretherton CS (2001) A moist pbl parameterization for large-scale models and its application to subtropical cloud-topped marine boundary layers. Mon Weather Rev 129:357–377. https://doi.org/10.1175/1520-0493(2001)129<0357:AMPPFL>2.0.CO;2

Güttler I, Branković Č, O'Brien TA, Coppola E, Grisogono B, Giorgi F (2014) Sensitivity of the regional climate model RegCM4.2 to planetary boundary layer parameterisation. Clim Dyn 43:1753–1772. https://doi.org/10.1007/s00382-013-2003-6

Härdle W, Marron JS, Wand MP (1990) Bandwidth choice for density derivatives. J R Stat Soc Series B Stat Methodol 52:223–232. https://doi.org/10.1111/j.2517-6161.1990.tb01783.x

Haylock MR, Hofstra N, Klein Tank AMG, Klok EJ, Jones PD, New M (2008) A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. J Geophys Res Atmos 113:D20119. https://doi.org/10.1029/2008JD010201

Hazeleger W, Guemas V, Wouters B, Corti S, Andreu-Burillo I, Doblas-Reyes FJ, Wyser K, Caian M (2013) Multiyear climate predictions using two initialization strategies. Geophys Res Lett 40:1794–1798. https://doi.org/10.1002/grl.50355

Herrera S, Fernández J, Gutiérrez JM (2016) Update of the Spain02 gridded observational dataset for EURO-CORDEX evaluation: assessing the effect of the interpolation methodology. Int J Climatol 36:900–908. https://doi.org/10.1002/joc.4391

Hofstra N, Haylock M, New M, Jones PD (2009) Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature. J Geophys Res Atmos 114:D21101. https://doi.org/10.1029/2009JD011799

Holtslag AAM, De Bruijn EIF, Pan H-L (1990) A high resolution air mass transformation model for short-range weather forecasting. Mon Weather Rev 118:1561–1575. https://doi.org/10.1175/1520-0493(1990)118<1561:AHRAMT>2.0.CO;2

Hong SY, Dudhia J, Chen SH (2004) A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. Mon Weather Rev 132:103–120. https://doi.org/10.1175/1520-0493(2004)132<0103:ARATIM>2.0.CO;2

Isotta FA, Vogel R, Frei C (2015) Evaluation of European regional reanalyses and downscalings for precipitation in the Alpine region. Meteorologische Zeitschrift 24:15–37. https://doi.org/10.1127/metz/2014/0584

Kain JS (2004) The Kain–Fritsch convective parameterization: an update. J Appl Meteorol Climatol 43:170–181. https://doi.org/10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2

Kain JS, Fritsch JM (1990) A one-dimensional entraining/detraining plume model and its application in convective parameterization. J Atmos Sci 47:2784–2802. https://doi.org/10.1175/1520-0469(1990)047<2784:AODEPM>2.0.CO;2

Kalmár T, Pieczka I, Pongrácz R (2021) A sensitivity analysis of the different setups of the RegCM4.5 model for the Carpathian region. Int J Climatol 41:E1180–E1201. https://doi.org/10.1002/joc.6761

Kis A, Pongrácz R, Bartholy J (2015) Analysis of projected frequency and intensity changes of precipitation in the Carpathian Region. Aerul si Apa. Componente ale Mediului 2015:125–132. https://doi.org/10.17378/AWC2015_17

Klein Tank AMG, Wijngaard JB, Können GP, Böhm R, Demarée G, Gocheva A, Mileta M, Pashiardis S, Hejkrlik L, Kern-Hansen C, Heino R, Bessemoulin P, Müller-Westermeier G, Tzanakou M, Szalai S, Pálsdóttir T, Fitzgerald D, Rubin S, Capaldo M et al (2002) Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. Int J Climatol 22:1441–1453. https://doi.org/10.1002/joc.773

Kotlarski S, Keuler K, Christensen OB, Colette A, Déqué M, Gobiet A, Goergen K, Jacob D, Lüthi D, van Meijgaard E, Nikulin G, Schär C, Teichmann C, Vautard R, Warrach-Sagi K, Wulfmeyer V (2014) Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble. Geosci Model Dev 7:1297–1333. https://doi.org/10.5194/gmd-7-1297-2014

Kotlarski S, Szabó P, Herrera S, Räty O, Keuler K, Soares PM, Cardoso RM, Bosshard T, Pagé C, Boberg F, Gutiérrez JM, Isotta FA, Jaczewski A, Kreienkamp F, Liniger MA, Lussana C, Pianko-Kluczyńska K (2019) Observational uncertainty and regional

climate model evaluation: a pan-European perspective. Int J Climatol 39:3730–3749. https://doi.org/10.1002/joc.5249

Kyselý J, Plavcová E (2010) A critical remark on the applicability of E-OBS European gridded temperature data set for validating control climate simulations. J Geophys Res Atmos 115:D23118. https://doi.org/10.1029/2010JD014123

Lloyd SP (1957) Least square quantization in PCM. Bell Telephone Laboratories Paper. Published in journal much later: Lloyd, SP: Least squares quantization in PCM. IEEE Trans Inform Theor (1957/1982) 28:129–137. https://doi.org/10.1109/TIT.1982.1056489

Ly S, Charles C, Degré A (2011) Geostatistical interpolation of daily rainfall at catchment scale: the use of several variogram models in the Ourthe and Amblève catchments, Belgium. Hydrol Earth Syst Sci 15:2259–2274. https://doi.org/10.5194/hess-15-2259-2011

MacQueen JB (1967) Some methods for classification and analysis of multivariate observations. In: Le Cam LM, Neyman J (eds) Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, vol 1. University of California Press, California, pp 281–297 http://projecteuclid.org/euclid.bsmsp/1200512992

Maraun D (2013) Bias correction, quantile mapping, and downscaling: revisiting the inflation issue. J Clim 26:2137–2143. https://doi.org/10.1175/JCLI-D-12-00821.1

Maraun D, Osborn T, Rust H (2012) The influence of synoptic airflow on UK daily precipitation extremes. Part II: regional climate model and E-OBS data validation. Clim Dyn 39:287–301. https://doi.org/10.1007/s00382-011-1176-0

Maxwell SE, Kelley K, Rausch JR (2008) Sample size planning for statistical power and accuracy in parameter estimation. Annu Rev Psychol 59:537–563. https://doi.org/10.1146/annurev.psych.59.103006.093735

Mitchell D, Heaviside C, Vardoulakis S, Huntingford C, Masato G, Guillod BP, Frumhoff P, Bowery A, Wallom D, Allen M (2016) Attributing human mortality during extreme heat waves to anthropogenic climate change. Environ Res Lett 11:074006. https://doi.org/10.1088/1748-9326/11/7/074006

Oleson KW, Dai Y, Bonan G, Bosilovich M, Dickinson R, Dirmeyer P, Hoffman F, Houser P, Levis S, Niu G-Y, Thornton P, Vertenstein M, Yang Z-L, Zeng X (2004) Technical description of the Community Land Model (CLM). Technical note NCAR/TN-461+STR. NCAR:Boulder, CO. pp 173. http://www.cgd.ucar.edu/tss/clm/distribution/clm3.0/TechNote/CLMTechNote.pdf

Pal JS, Small EE, Eltahir EAB (2000) Simulation of regional-scale water and energy budgets: Representation of subgrid cloud and precipitation processes within RegCM. J Geophys Res Atmos 105:29579–29594. https://doi.org/10.1029/2000JD900415

Palazzi E, von Hardenberg J, Provenzale A (2013) Precipitation in the Hindu-Kush Karakoram Himalaya: Observations and future scenarios. J Geophys Res Atmos 118:85–100. https://doi.org/10.1029/2012JD018697

Pall P, Aina T, Stone DA, Stott PA, Nozawa T, Hilberts AGJ, Lohmann D, Allen MR (2011) Anthropogenic greenhouse gas contribution to flood risk in England and Wales in autumn 2000. Nature 470:382–385. https://doi.org/10.1038/nature09762

Perkins SE, Pitman AJ, Holbrook NJ, McAneney J (2007) Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. J Clim 20:4356–4376. https://doi.org/10.1175/JCLI4253.1

Pieczka I, Pongrácz R, Szabóné André K, Kelemen FD, Bartholy J (2017) Sensitivity analysis of different parameterization schemes using RegCM4.3 for the Carpathian region. Theor Appl Climatol 130:1175–1188. https://doi.org/10.1007/s00704-016-1941-4

Pitman EJG (1937) Significance tests which may be applied to samples from any populations. Supplement to the J R Stat Soc 4:119. https://doi.org/10.2307/2984124

Prein AF, Gobiet A (2017) Impacts of uncertainties in European gridded precipitation observations on regional climate analysis. Int J Climatol 37:305–327. https://doi.org/10.1002/joc.4706

Prömmel K, Geyer B, Jones JM, Widmann M (2010) Evaluation of the skill and added value of a reanalysis-driven regional simulation for Alpine temperature. Int J Climatol 30:760–773. https://doi.org/10.1002/joc.1916

Ptáček P, Létal A, Ruffini FV, Renner K (2011) Atlas of the Carpathian macroregion. Europa Regional 17:108–122 https://nbn-resolving.org/urn:nbn:de:0168-ssoar-47996-2

R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria http://www.R-project.org/

Rauthe M, Steiner H, Riediger U, Mazurkiewicz A, Gratzki A (2013) A Central European precipitation climatology? Part I: generation and validation of a high-resolution gridded daily data set (HYRAS). Meteorologische Zeitschrift 22:235–256. https://doi.org/10.1127/0941-2948/2013/0436

Repel A, Zeleňáková M, Jothiprakash V, Hlavatá H, Blišťan P, Gargar I, Purcz P (2021) Long-term analysis of precipitation in Slovakia. Water 13:952. https://doi.org/10.3390/w13070952

Rosenblatt M (1956) Remarks on some nonparametric estimates of a density function. Ann Math Stat 27:832–837. https://doi.org/10.1214/aoms/1177728190

Schneider U, Becker A, Finger P, Meyer-Christoffer A, Ziese M, Rudolf B (2014) GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. Theor Appl Climatol 115:15–40. https://doi.org/10.1007/s00704-013-0860-x

Sekulić A, Kilibarda M, Protić D, Bajat B (2021) A high-resolution daily gridded meteorological dataset for Serbia made by Random Forest Spatial Interpolation. Scientific Data 8:123. https://doi.org/10.1038/s41597-021-00901-2

Sheather SJ, Jones MC (1991) A reliable data-based bandwidth selection method for kernel density estimation. J R Stat Soc Series B Stat Methodol 53:683–690. https://doi.org/10.1111/j.2517-6161.1991.tb01857.x

Sidău MR, Croitoru A-E, Alexandru D-E (2021) Comparative analysis between daily extreme temperature and precipitation values derived from observations and gridded datasets in North-Western Romania. Atmosphere 12:361. https://doi.org/10.3390/atmos12030361

Smith RB (1979) The influence of mountains on the atmosphere. Adv Geophys 21:87–230. https://doi.org/10.1016/S0065-2687(08)60262-9

Spinoni J, Szalai S, Szentimrey T, Lakatos M, Bihari Z, Nagy A, Németh Á, Kovács T, Mihic D, Dacic M, Petrovic P, Kržič A, Hiebl J, Auer I, Milkovic J, Štepánek P, Zahradnícek P, Kilar P, Limanowka D et al (2015) Climate of the Carpathian Region in the period 1961-2010: climatologies and trends of 10 variables. Int J Climatol 35:1322–1341. https://doi.org/10.1002/joc.4059

Stephens GL, L'Ecuyer T, Forbes R, Gettelmen A, Golaz J-C, Bodas-Salcedo A, Suzuki K, Gabriel P, Haynes J (2010) Dreary state of precipitation in global models. J Geophys Res Atmos 115:D24. https://doi.org/10.1029/2010JD014532

Szalai S, Auer I, Hiebl J, Milkovich J, Radim T, Stepanek P, Zahradnicek P, Bihari, Z, Lakatos M, Szentimrey T, Limanowka D, Kilar P, Cheval S, Deak Gy, Mihic D, Antolovic I, Mihajlovic V, Nejedlik P, Stastny P, Mikulova K, Nabyvanets I, Skyryk O, Krakovskaya S, Vogt J, Antofie T, Spinoni J (2013) Climate of the greater Carpathian region. Final technical report. http://www.carpatclim-eu.org

Szentimrey T (2007) Manual of homogenization software MASHv3.02. Hungarian Meteorological Service, Budapest, p 65p

Szentimrey T, Bihari Z (2006) MISH (Meteorological interpolation based on surface homogenized data basis). In: Tveito OE, Wegehenkel M, van der Wel F, Dobesch H (eds) COST action 719 final report, the use of GIS in climatology and meteorology, pp 54–56

Tanarhte M, Hadjinicolaou P, Lelieveld J (2012) Intercomparison of temperature and precipitation data sets based on observations in the Mediterranean and the Middle East. Geophys Res Atmos 117:D12. https://doi.org/10.1029/2011JD017293

Tiedtke M (1996) An extension of cloud-radiation parameterization in the ECMWF model: the representation of subgrid-scale variations of optical depth. Mon Weather Rev 124:745–750. https://doi.org/10.1175/1520-0493(1996)124<0745:AEOCRP>2.0.CO;2

Torma C, Coppola E, Giorgi F, Bartholy J, Pongrácz R (2011) Validation of a high-resolution version of the regional climate model RegCM3 over the Carpathian basin. J Hydrometeorol 12:84–100. https://doi.org/10.1175/2010JHM1234.1

UNEP (2007) Carpathian Environment Outlook. United Nations Environment Programme (UNEP). Division of Early Warning and Assessment (DEWA), Switzerland, Geneva, p 236. https://wedocs.unep.org/20.500.11822/8596

Varga ÁJ, Breuer H (2020) Sensitivity of simulated temperature, precipitation, and global radiation to different WRF configurations over the Carpathian Basin for regional climate applications. Clim Dyn 55:2849–2866. https://doi.org/10.1007/s00382-020-05416-x

Vidal JP, Martin E, Franchistéguy L, Baillon M, Soubeyroux JM (2010) A 50-year high-resolution atmospheric reanalysis over France with the Safran system. Int J Climatol 30:1627–1644. https://doi.org/10.1002/joc.2003

Wood SN (2003) Thin plate regression splines: thin plate regression splines. J R Stat Soc Series B Stat Methodol 65:95–114. https://doi.org/10.1111/1467-9868.00374

Wood SN (2006) Generalized additive models: an introduction with R. Chapman and Hall/CRC, Boca Raton, London, New York

World Meteorological Organization (2011) WMO statement on the status of the global climate in 2010. World Meteorological Organization, Geneva

Xue Y, Janjic Z, Dudhia J, Vasic R, De Sales F (2014) A review on regional dynamical downscaling in intraseasonal to seasonal simulation/prediction and major factors that affect downscaling ability. Atmos Res 147–148:68–85. https://doi.org/10.1016/j.atmosres.2014.05.001