Bridging the gap between climate models and impact studies: the FORESEE Database

L. Dobor[1], Z. Barcza[1,2]*, T. Hlásny[3,4], Á. Havasi[5], F. Horváth[2], P. Ittzés[2] and J. Bartholy[1]

[1]Department of Meteorology, Eötvös Loránd University, Budapest, Hungary
[2]Institute of Ecology and Botany, Centre for Ecological Research, Hungarian Academy of Sciences, Vácrátót, Hungary
[3]National Forest Centre – Forest Research Institute, Zvolen, Slovakia
[4]Faculty of Forestry and Wood Sciences, Czech University of Life Sciences, Prague, Czech Republic
[5]Department of Applied Analysis and Computational Mathematics, Eötvös Loránd University, Budapest, Hungary

*Correspondence: Z. Barcza, H-1117 Budapest, Pázmány P. sétány 1/A, Hungary, E-mail: zoltan.barcza@ttk.elte.hu

Studies on climate change impacts are essential for identifying vulnerabilities and developing adaptation options. However, such studies depend crucially on the availability of reliable climate data. In this study, we introduce the climatological database called FORESEE (Open Database for Climate Change Related Impact Studies in Central Europe), which was developed to support the research of and adaptation to climate change in Central and Eastern Europe: the region where knowledge of possible climate change effects is inadequate. A questionnaire-based survey was used to specify database structure and content. FORESEE contains the seamless combination of gridded daily observation-based data (1951–2013) built on the E-OBS and CRU TS datasets, and a collection of climate projections (2014–2100). The future climate is represented by bias-corrected meteorological data from 10 regional climate models (RCMs), driven by the A1B emission scenario. These latter data were developed within the frame of the ENSEMBLES FP6 project. Although FORESEE only covers a limited area of Central and Eastern Europe, the methodology of database development, the applied bias correction techniques, and the data dissemination method, can serve as a blueprint for similar initiatives.

Dataset

Identifier: doi:10.5281/zenodo.9614
Creator: Dobor, L., Barcza, Z.
Title: FORESEE database v2.0
Publisher: Eötvös Loránd University, Hungary
Publication year: 2014
Version: 2.0

Introduction

Recently, a growing body of studies has been published on the impacts of climate change on ecosystems and societies, in response to increasing interest in the anticipated economic development and options for adaptation. Such knowledge supports policymakers, diverse stakeholder groups, and individuals in mitigating the adverse effects, and taking advantage of the positive effects, of the changing climate (IPCC, 2007).

The utilization of climate model results is currently the most feasible way to estimate future meteorological conditions (Calanca and Semenov, 2012; IPCC, 2013). State-of-the-art impact analyses use ensembles of climate projections to gain an insight into the uncertainty of projections (Baigorria *et al.*, 2008; Cook *et al.*, 2010; Jönsson and Bärring, 2011; Supit *et al.*, 2012; Carter, 2013).

General circulation models (GCMs) are run at coarse spatial resolution, so they are unable to capture sub-grid cell processes induced, for example, by regional topography, which is often required for regional impact studies. For this reason, downscaling methods are applied to provide climate data in higher resolution than that of GCMs (Wood *et al.*, 2004; Castro *et al.*,

2005). In the case of dynamic downscaling, outputs of which are used in the current study, regional climate models (RCMs) are run using the initial and boundary conditions provided by GCMs (Giorgi, 1990) to generate higher resolution meteorological fields. The RCMs are used to simulate smaller scale atmospheric processes, owing to their higher resolution topography and physics (Wang *et al.*, 2004).

Direct use of RCM data is limited by systematic errors inherently present in the simulated variables as a result of uncertainties in the parameterization and model structure (Varis *et al.*, 2004; Christensen *et al.*, 2008). Such data may be unusable for impact studies which, typically, need unbiased climate data, as the models are sensitive to biases in the driving meteorological data (Baigorria *et al.*, 2007; Teutschbein and Seibert, 2010; Calanca and Semenov, 2012). The assumption that systematic errors in the past are propagated equally to the future (Maraun, 2012) allows for using various bias correction methods (Ines and Hansen, 2006; Li *et al.*, 2010; Piani *et al.*, 2010; White and Toumi, 2013). Reliable, observation-based datasets, covering a sufficiently long time period are, however, required to remove the biases.

The climate in Central Europe is projected to face substantial changes (e.g. Seneviratne *et al.*, 2006), with adverse effects on natural resources, ecosystems, and societies. Transitional economies covering large parts of this region often generate concern about the sustainability of current management practices, owing to possible amplification of the vulnerability to climate change. Central and Eastern Europe were found to be surrounded by three of the most prominent climate change hot spots in the world (Giorgi, 2006). The importance of European ecosystems (Nabuurs *et al.*, 1997) and their vulnerability also calls for extending the currently insufficient knowledge of climate change impacts.

Despite the growing recognition of need for impact studies, the availability of suitable meteorological data is limited in many regions, and their quality often does not meet the criteria of impact studies; e.g. considering spatial and temporal resolution. This situation holds true for Central Europe which means that impact studies are often hampered by the insufficient data availability. For these reasons, in this paper we introduce a new meteorological database called FORESEE (Open Database for Climate Change Related Impact Studies in Central Europe) that aims to support studies on climate change impacts in Central Europe. The main motivation for the construction FORESEE was to bridge the gap between the raw results of climate models and the end-users of climate projections, as such a link is, by experience, often not straightforward and proper. The main objectives of the paper are as follows:

— to survey the needs of potential database users on the content and structure, retrieval, and pre-processing of climate data driving diverse models
— to document the development of the FORESEE database containing daily observation-based data and bias-corrected RCM results, covering the period 1951–2100
— to document the applied bias correction technique
— to document the availability options of FORESEE.

1. Data production methods

1.1. Geographical coverage of the database

FORESEE covers Bosnia and Herzegovina, Croatia, the Czech Republic, Hungary, Slovakia, Slovenia, and parts of Albania, Austria, Bulgaria, Germany, Italy, Kosovo, Moldova, Montenegro, Poland, Romania, Ukraine, and Serbia (Figure 1). The total area covered by FORESEE is about 1 270 000 km^2. The region includes the entire Carpathian Mountains and large part of the Danube catchment, which have been recently receiving increased attention in EU coordinated actions, such as the Carpathian EcoRegion Initiative (http://www.carpates.org) and the Danube Strategy (http://www.danube-region.eu).

1.2. Survey of users' needs

To make FORESEE really supportive of climate change research in Central Europe, a questionnaire surveying users' needs from climate data was circulated within the Central European scientific community. Hydrologists, foresters, ecologists, climatologists, all of them experienced in climate change research, were questioned primarily.

Five questions addressed the type of applied meteorological variables commonly used in impact studies, and their temporal and spatial resolution. Another two questions addressed users' knowledge of climate model results, the need for addressing the uncertainty related to climate modelling through the inclusion of more than one model, and the awareness of the effect and need of bias correction. Finally, suitable file formats and software used for manipulating the climate data, and problems experienced when accessing meteorological datasets were addressed.

The 42 returned questionnaires indicated that basic daily meteorological data – i.e. maximum and minimum temperature and precipitation – are most commonly required by the users. Interest in hourly data was only marginal, while interest in all resolutions from daily to 30-year long averages was very frequent. Therefore, provision of daily data, which can be further aggregated, seems necessary. The responders were interested both in gridded data, allowing for spatial modelling, and for single-point localized data. This suggests that preparation of gridded data, which may be further restricted to specific locations, should meet users' needs. The responders indicated interest in the highest possible resolution data which are optimal for regional studies. Voting for the length of time
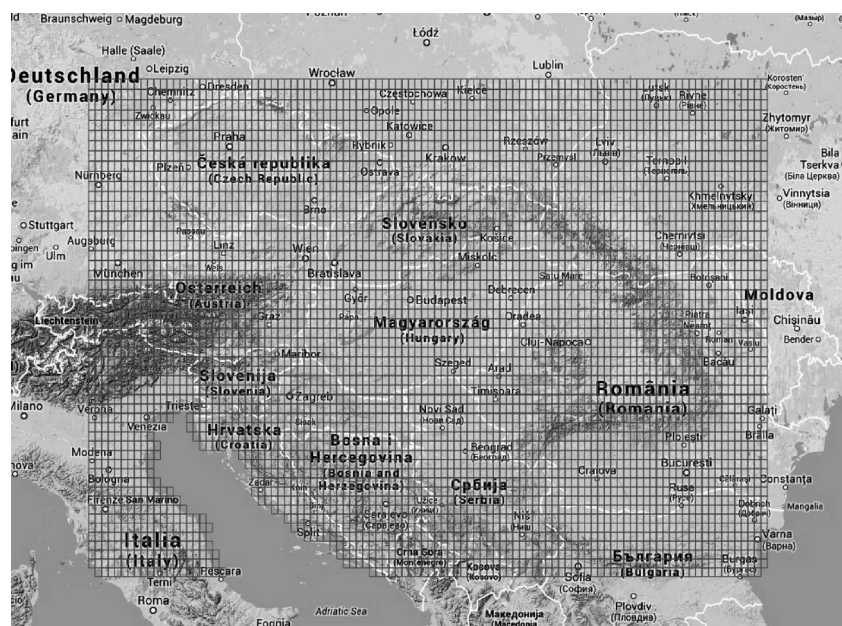
Figure 1. Target area of the FORESEE database (the location of the individual grid cells is also shown).

period for which the data are provided was quite balanced, indicating the need for the provision of as long a time series as possible; the responders denoted interest in both past and future climate data. Demand for RCM-GCM combinations varied among the responders, who basically pointed out the intention to use more than two models in their research. The group interested in one model only was, however, also large. This implies the need for the construction of a larger number of models, with guidance for selecting an appropriate subset or a single model.

Surprisingly, 20% of the answers indicated a lack of awareness about the limitations and applicability of raw climate model results (without bias correction applied), and about the potential effects of applying bias correction on the results of impact studies. As limitations of using raw model outputs are generally well recognized, the provision of corrected climate data with detailed documentation of the correction procedure and its effects seem to be an optimal choice.

Forty-two per cent of responders indicated that they can only handle ASCII files (MS Excel-compatible text), so they are unable to read formats such as NetCDF and HDF, which are commonly used for storing climate data. The most obvious problem reported in relation to the acquisition of climate data was general data availability, low spatial resolution and the lack of meta-data. Appendix S1 provides the complete statistics about the survey.

1.3. Observed meteorology in the FORESEE

Past climate within FORESEE is based on the E-OBS dataset (Haylock *et al.*, 2008), which contains observation-based gridded datasets in $\sim 25 \times 25$ km horizontal resolution, in daily time steps for the period 1951–2013, and the CRU TS 1.2 dataset (Climatic Research Unit, University of East Anglia, UK; Mitchell *et al.*, 2004), which contains monthly data in $1/6 \times 1/6°$ horizontal resolution for the period 1901–2000.

Currently, E-OBS has 10 versions. E-OBS covers the whole Europe, although the density of meteorological stations used for the construction of this dataset varies (Klok and Klein Tank, 2008). In the latest version of the FORESEE database (v2.0) the most up-to-date E-OBS version was used (v10).

It was found that the CRU TS 1.2 dataset contains very accurate meteorological data on a monthly scale, thus we sought for options for improving the E-OBS data using the CRU TS 1.2. For example, Szabó (2008) compared different data sources (ECA, ERA40, CRU TS 2.1 and CRU TS 1.2) with the Hungarian, high-resolution meteorological observations (the so-called HUGRID dataset). Szabó (2008) showed that long-term averages of CRU TS 1.2 had the best fit to HUGRID. To support the adjustment of E-OBS by CRU TS 1.2, we performed additional evaluation of the accuracy of both E-OBS and CRU TS datasets, using the CarpatClim dataset (http://www.carpatclim-eu.org/; Szalai *et al.*, 2013). The CarpatClim is a daily meteorological dataset for the Carpathian Region, based on a large number of observed meteorological data and a state-of-the-art interpolation technique, so it could represent a reasonable control data in the accuracy evaluation. (Note that as CarpatClim only covers about 38% of the FORESEE domain, it is not an adequate substitute for E-OBS or CRU TS 1.2 datasets.) A better match of the CRU TS 1.2 with the CarpatClim than with the E-OBS (see Appendix S2 for details) lends

support to the correction of daily E-OBS using monthly CRU TS 1.2 data.

Hence, for the construction of FORESEE, the E-OBS monthly means were forced to match the CRU TS 1.2 monthly means. As CRU TS 1.2 data were not available for the period 2001–2013, average correction factors for the period 1951–2000 were applied.

1.4. Model selection and post-processing of raw climate data

For the construction of FORESEE we used high resolution RCM results disseminated within the frame of the ENSEMBLES EU project (Van der Linden and Mitchell, 2009). About 31 model output sets are available for download at the ENSEMBLES database. The model selection was performed using the following logic. We selected all output sets which are available for the period 1951–2100, driven by the A1B scenario (the A1B greenhouse gas emission scenario represents a balanced emphasis on all energy sources responsible for greenhouse gas emission (IPCC, 2000); note that at the ENSEMBLES data portal 30 models are driven by the A1B greenhouse gas scenario out of the 31 transient model results. Of the 30 models 23 model output sets fulfil the criteria of sufficiently high spatial resolution (25 × 25 km), but only 14 output sets cover the entire 1951–2100 period, thus we only focused on this latter subset. We neglected two output sets because they used the same GCM-RCM combination as another one but with high- and low-sensitivity RCM settings. From the remaining 12 models one was not available at the website and another one contains too many data gaps at the end of the simulation period, which prevented us from the proper use of the output. The remaining ten models formed the basis for the construction of the FORESEE database (Table 1).

Although the selected 10 RCM outputs have been created within the same EU project, there are specific differences between the disseminated datasets. Some models use a 360-day calendar, some of them use standard calendar which contains leap years. To prepare the data for additional impact modelling

purposes all RCM results were converted to a 365-day calendar. In case of the standard calendar in leap years 31 December was removed from the database. The rationale behind removing the last day of the year was that during winter removal of a single day in every 4 years cannot have significant effect on hydrology, plant growth, human health, and other processes. When a climate model used a 360-day calendar additional days had to be created in case of January, March, May, July, August, October, and December. For these days we calculated the average of the previous and following days' temperatures, and dry days were inserted into the precipitation time series. In case of February some days had to be removed. In a few cases the last year (2100) was missing from the datasets, in this case year 2099 have been duplicated.

After the temporal standardization, the RCMs were interpolated to a uniform 1/6 × 1/6° horizontal resolution grid, using an inverse distance interpolation technique (the resulting grid is equivalent with the CRU TS 1.2 grid). FORESEE is available on this 1/6 × 1/6° grid, where the resolution represents a compromise between the highest possible resolution versus database size (trade-off between quality and applicability of the dataset).

As direct use of GCM and RCM data has been found to be inadequate, bias correction seems inevitable to produce realistic meteorological data driving the impact models. Considering the 10 selected RCMs, comparison of climate maps based on observations and the uncorrected model output clearly reveals the need for bias correction (Appendix S3). It is also well recognized that the number of wet days tends to be overestimated in the raw output of climate models (Mearns *et al.*, 1990; Goddard *et al.*, 2001; Gutowski *et al.*, 2003; Dosio and Paruolo, 2011; Themeßl *et al.*, 2012).

Assuming that systematic errors are stable in time (Maraun, 2012), the correction factors determined by the comparison of observations and model results (1951–2013) were used for the correction of model results for the future (2014–2100). The period 1951–2013 was used as the basis of the correction. We applied a bias correction based on the cumulative

Table 1. List of the regional climate models used in the database.

Model ID	Model name (RCM-GCM)	Developing institute
1	ALADIN-ARPEGE	National Centre for Meteorological Research (CNRM)
2	CLM-HadCM3Q0	Swiss Federal Institute of Technology Zürich (ETHZ)
3	HadRM3Q0-HadCM3Q0	Hadley Centre for Climate Prediction and Research (HC)
4	HIRHAM5-ARPEGE	Danish Meteorological Institute (DMI)
5	HIRHAM5-ECHAM5	Danish Meteorological Institute (DMI)
6	RACMO2-ECHAM5	Royal Netherlands Meteorological Institute (KNMI)
7	RCA-ECHAM5	Sweden's Meteorological and Hydrological Institute (SMHI)
8	RCA-HadCM3Q0	Sweden's Meteorological and Hydrological Institute (SMHI)
9	RegCM3-ECHAM5	The Abdus Salam International Centre for Theoretical Physics (ICTP)
10	REMO-ECHAM5	Max-Planck-Institute for Meteorology (MPI)

RCM, regional climate models; GCM, general circulation models

distribution function (CDF) fitting technique (also known as quantile mapping/fitting or histogram equalization), at monthly time intervals (Ines and Hansen, 2006) for each grid point in the target area.

As temperature and precipitation possess different statistical properties, specific bias correction techniques had to be applied (Hansen *et al.*, 2006). In the case of precipitation, both the amounts and frequency of precipitation were corrected. In the case of temperature, the correction was based on additive shifting, while correction factors for precipitation amounts were based on the multiplication. In the next section we describe the utilized precipitation correction method in detail (temperature correction is similar to the second phase of the precipitation amount correction).

1.5. Detailed description of the applied bias correction method in case of the precipitation

1.5.1. Precipitation frequency correction

Precipitation frequency correction is based on the comparison of the monthly number of simulated (N_{wd}^m) and observed (N_{wd}^o) wet days (i.e. when the daily precipitation amount is greater than or equal to 0.1 mm):

$$N_{wd}^o = \sum_{i=1}^{k} W_i^o, \text{ where } W_i^o = \left\{ \begin{matrix} 1, x_i^o \geq 0.1 \text{ mm} \\ 0, x_i^o < 0.1 \text{ mm} \end{matrix} \right\} \quad (1)$$
$$\text{for } i = 1, k,$$

$$N_{wd}^m = \sum_{i=1}^{k} W_i^m, \text{ where } W_i^m = \left\{ \begin{matrix} 1, x_i^m \geq 0.1 \text{mm} \\ 0, x_i^m < 0.1 \text{mm} \end{matrix} \right\} \quad (2)$$
$$\text{for } i = 1, k,$$

where k means the number of the days which belong to the given month for the whole 1951–2013 period, x_i^o and x_i^m are the observed and the simulated precipitation time series (respectively) for a given month; w_i^o and w_i^m are the daily indicators of the rainfall events.

If the model simulates more wet days than the observed ($N_{wd}^m > N_{wd}^o$), wet days are removed starting from the smallest precipitation amount. If the model simulates fewer wet days than the observation ($N_{wd}^m < N_{wd}^o$), rainfall events are generated artificially. In this way, additional days with precipitation are created in random fashion within a month. Precipitation amounts take on a random value between 0.1 mm and the 90th percentile of the simulated dataset. To determine the correction factors (f_{corr}), the number of observed wet days is divided by the number of simulated wet days:

$$f_{corr} = \frac{N_{wd}^o}{N_{wd}^m}. \quad (3)$$

The correction factors are calculated monthly for each grid point. The monthly number of wet days in the future (2014–2100) is scaled by the correction factor (f_{corr}) which is calculated on the basis of the period 1951–2013.

### 1.5.2. Precipitation amount correction

After the frequency correction, the precipitation amount correction was performed using CDF fitting technique (Figure 2; note that the same technique was used for temperature correction). The p-quantile (q_p) is a precipitation value which the random variable will be at, or below, with probability $p/1000$ ($p = 1, 2, ..., 1000$) (1000 partitions were used to split the [0,1] probability interval, this partition ensure the required continuity but keeps the computing demand in acceptable level).

$$q_p = \inf \left\{ X \in R : \frac{p}{1000} \leq F(X) \right\}, \quad (4)$$

where $F(X)$ is the CDF of the precipitation (x):

$$F(X) = p(x \leq X). \quad (5)$$

Quantile functions (or inverse cumulative density functions) using 1000 partitions were determined based on the observed and the simulated (and already
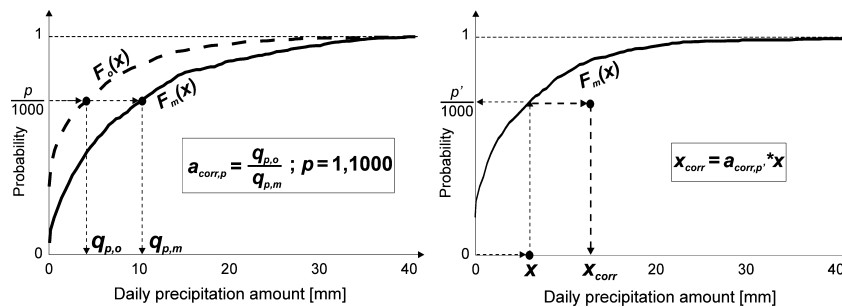
Figure 2. Demonstration of the monthly cumulative distribution fitting (CDF) technique for one pixel. $F_o(x)$ and $F_m(x)$ show the cumulative distribution functions, $q_{p,o}$ and $q_{p,m}$ mean the p-quantiles for the observed and the modelled daily precipitation amount. The left figure shows the calculation of the correction factor ($a_{corr,p}$) for each $p/1000$ probability (p means the serial number of the probability values). The right figure shows the correction of a given precipitation amount (x) using the associated $a_{corr,p}$ correction factor.
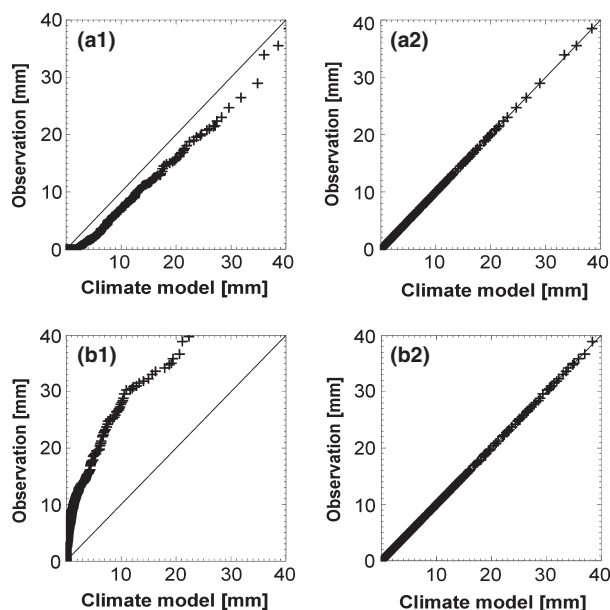
Figure 3. The q–q plots for HIRHAM5-ECHAM5 model (the quantile values based on the modelled (*x*-axis) and the observed (*y*-axis) precipitation rates) before (a/1, b/1) and after (a/2, b/2) the bias correction procedure in case of removal (a/1, February) and generation of wet days (b/1, August). After the correction the quantiles of the datasets are approximately the same, fit to the identity function (figures on the right side).

frequency corrected) precipitation data for the given month and pixel (Figure 2).

The correction factors $a_{corr,p}$ ($p = 1, 2, \ldots, 1000$) are the quotients of the *p*-quantiles based on the simulated and observed precipitation, and so form an 1000 elements array. In order to correct a precipitation value (*x*), the closest *p*-quantile was determined from the 1000 elements quantile function (*p'*). The associated 'serial number' (*p'*) of the correction factor ($a_{corr,p'}$) was selected and used as a scale factor (Figure 2, right hand side, x_{corr} is the corrected value of *x*).

The precipitation correction for the future works the same way. The closest quantile has to be found and then scaled by the relevant number of correction factor created based on the comparison in the past.

Figure 3 shows the quantile–quantile function diagrams (q–q plot) based on the observed and the simulated datasets by the HIRHAM5-ECHAM5 model before the correction. The q–q plots shows all the days for a given month in the period 1951–2013, for a given pixel. The upper left figure shows when the model simulates too many wet days (February) and the bottom left figure shows when it simulates fewer wet days (August). After the correction the quantiles of the datasets are approximately the same (figures on the right-hand side).

1.5.3. Effect of bias correction on the precipitation time series

The applied precipitation bias correction method is expected to modify the statistical properties of the

precipitation time series, which needs further consideration. An important question is whether the variance of precipitation is affected by the correction, and if it does, to what extent? Additionally, as FORESEE consists of observations for the past (not bias-corrected RCM data), and bias-corrected RCM data for the future, consistency of these two components in terms of frequency of wet days needs to be preserved.

Standard deviation of simulated precipitation was compared to observations with and without the bias correction for the 1961–1990 time period. The results indicated that precipitation variance (more precisely, square root of variance) became closer to the observed variance after the bias correction. Appendix S4 demonstrates the effect of bias correction on the standard deviation of simulated data in graphical form for each model.

Frequency of wet days was calculated both for the past and future based on the uncorrected and corrected RCM data. Appendix S5 shows the comparison of number of wet days for the FORESEE domain according to the different GCM-RCM combinations and the observations. The results demonstrate that the applied bias correction method improved the consistency between observations and projections in terms of frequency of wet days.

2. Results – structure and dissemination of the FORESEE

FORESEE data are primarily available in NetCDF format, which was designated to contain both data and metadata in an easily accessible format (http://www.unidata.ucar.edu/). The total size of the database is about 22 GB. Once a request has been sent out through the FORESEE website (http://nimbus.elte.hu/FORESEE), the requested files (one file per model per variable) are made available for download from the site. The dataset is also available via Zenodo (http://zenodo.org/record/9614?ln=en; doi: 10.5281/zenodo.9614).

Questionnaire responses indicated that NetCDF format is not usable for most responders, and MS Excel-compatible data (i.e. pure text retrieval of data) were indicated as the most suitable. To meet this demand, the web service technology was used to enable easy access to data for the users.

The REST style web service (Fielding and Taylor, 2002) of FORESEE database is provided by the 'ecos' server of the Centre for Ecological Research of the Hungarian Academy of Sciences, catalogued in 'BiodiversityCatalogue' – The Biodiversity Sciences Web Services Registry for discovering, registering, annotating, and monitoring biodiversity web services (http://www.biodiversitycatalogue.org) powered by BioCatalogue (Bhagat *et al.*, 2010). This service provides tailored access to the database at http://nimbus.elte.hu/FORESEE/map_query/index.html. A query has to be formulated according to the geographical location, requested meteorology variables, time interval in years

(from 1951 to 2100), and selection of RCM, if time range is over the observational period. The service finds the corresponding FORESEE grid cell and serves the requested dataset in comma separated ASCII text format. The base URL of the service is 'http://ecos.ok-ologia.mta.hu:5000/foresee', the REST endpoint is '/find' (GET method). Parameters can be set as an example to '/18.63854,47.37291/1,1,0/1951,2100/ala-din' –/latitude and longitude in decimal degrees/set daily maximum and minimum temperatures, precipitation on or off/the requested time interval in years and/short name of the RCM.

The GIS-based webpage has two main parts: the online map and the menu with operations. One can set up retrieval options in the 'filter' menu point. Downloading is possible for one given location for a given period between 1951 and 2100 for one, two or three meteorological variable(s) on daily time step. If the end-year of the retrieval is after 2013, climate model selection is also needed. The location could set using the online map by directly clicking on the location of interest, or set by providing longitude/latitude coordinates, or it can also be chosen by typing a settlement name under the 'find city/location' menu point. The FORESEE grid layer helps to understand the geographical coverage of the downloaded time series. Note that in the present version of the GIS tool interpolation is not implemented, which means that for a given grid cell the same dataset is retrieved which is assigned to the centre point of the grid cell.

3. Discussion

3.1. Added value of the FORESEE

The reliability of information acquired within various studies on climate change impacts depends on a number of factors, including availability and quality of driving climate data (Maurer *et al.*, 2007), which may represent an important source of uncertainty (Olesen *et al.*, 2007). The effect of climate data quality, spatial resolution, or completeness on simulation outputs is, however, poorly understood, even though it may limit the interpretability and question the validity of some findings. To address these issues, we developed the FORESEE database which provides unified meteorological data based on observations and an ensemble of RCMs, seamlessly covering a large part of Central Europe. The importance of FORESEE for this region is highlighted by the remarkably limited options for the use of national meteorological data, owing to diverse legal regulations and restrictions in all Central European countries. For these reasons, the use of freely available gridded datasets is becoming a frequent practice in this region, although lower quality and accuracy of such data in comparison with data from national meteorological networks is well recognized. Hence, the availability of the presented database may support the regional research of climate change impacts, which is presently inadequate to anticipate

climate change threats (e.g. Seneviratne *et al.*, 2006; Hlásny *et al.*, 2011).

On the basis of the questionnaire-based survey the designation of FORESEE was an important step towards database development, potentially strengthening database's effect on climate change research in the FORESEE region. Although the representativeness of the responders' sample in the returned questionnaires is unknown, this feedback helped to shape some parts of the database.

A seamless combination of observations with an ensemble of climate model results creates an improved modelling environment in comparison with, for example, the original ENSEMBLES database.

Though the effect and need of bias correction was found to be largely unrecognized among the questionnaire responders, we saw the provision of bias-corrected data as necessary, since the uncorrected data are often beyond the scope of permissible inputs to impact models. For example, Ines and Hansen (2006) found that the coupled correction of amounts and the frequency of precipitation had a positive effect on crop simulations, which were much more reliable as compared with those run with uncorrected data. This method was also applied successfully by Baigorria *et al.* (2007) to simulate crop production. The need for the inclusion of bias-corrected data in FORESEE is also promoted by the more or less lacking skills of users to perform such correction. In particular, the correction of daily precipitation data may pose difficulties because both amounts and frequency need to be corrected (Déqué, 2007; Dosio and Paruolo, 2011; Calanca and Semenov, 2012).

The scientific workflow concept is a new and emerging paradigm for the support of research in general (Goble and De Roure, 2009), and specifically in biodiversity informatics (Hardisty *et al.*, 2013). A workflow can apply online web services, manage large, distributed data sources, and perform demanding computations within a simple framework. The development of web services and the creation of workflows is a major task within the frame of the BioVeL project (http://www.biovel.eu). The construction of the GIS tool is the first major step to satisfy the needs of the end-users for climate projections in Central Europe.

3.2. Limitations and future developments

Currently, FORESEE contains only climate projections based on the A1B emission scenario, which covers approximately the central part of plausible realizations of the future climate (in terms of warming). Therefore, impact studies which will utilize the FORESEE data will consider only a limited portion of the variability in anticipated responses. At the same time, options for the exploration of effects of various CO_2 trajectories on ecosystems are also limited. For this reason, the inclusion of additional simulations, driven by emission scenarios other than A1B, is an important part of future developments. FORESEE is intended to be

extended with RCM results based on the new IPCC greenhouse gas emission scenarios, the so-called representative concentration pathways (RCPs; Moss *et al.*, 2008).

Another limitation of the current version of FORESEE is related to the applied bias correction method, which cannot retain intervariable dependencies within the RCM outputs (mainly as a result of artificial introduction and removal of wet days). As there is a functional relationship between air temperature and precipitation, the applied technique may violate the consistency between these meteorological elements. This fact should be kept in mind during the utilization of the dataset. Similarly, as the bias correction method alters the temporal correlation of the daily precipitation time series, the consecutive number of wet days, the length of dry periods, and other statistics related to precipitation are modified. This artifact might be important in some impact studies.

Therefore, methods which allow to retain the intervariable dependencies – e.g. the so-called linear correction (White and Toumi, 2013) – are expected to be implemented. Another option is to use alternative bias correction methods for RCM post-processing, e.g. the so-called 'change factor' method (Ho *et al.*, 2012).

Currently, FORESEE contains only basic meteorological data (minimum, maximum temperature and precipitation), however, the need for additional variables, such as water vapour pressure, relative humidity, solar irradiation or daylight temperature, may arise. Indeed, some variables can be derived from the currently available data. For example, the MT-CLIM (Mountain Microclimate Simulation Model; Hungerford *et al.*, 1989; Thornton and Running, 1999) was found to be an adequate (validated) tool for estimating the daytime temperature and global radiation (which are often needed for crop or biogeochemical modelling). The generated values proved to be consistent with measured temperature and precipitation data (Glassy and Running, 1994; Thornton and Running, 1999; Thornton *et al.*, 2000). MT-CLIM uses the observed relationship between daily temperature range and atmospheric transmittance for the estimation of solar radiation (Thornton and Running, 1999; Fodor and Mika, 2011). As the changing climate might alter this relationship, adjustment of the MT-CLIM internal model parameters might be necessary to avoid biased radiation estimations (tendency of the model parameters can be estimated based on the raw climate model results). In any case, by using MT-CLIM the intervariable dependencies are retained, what stresses their potential value for further research. The MT-CLIM-generated data are expected to be included in FORESEE, to provide additional daily meteorological variables.

4. Concluding remarks

FORESEE strives to fill a gap in climate data availability in the Central European region and, at the same time,

to present the concept of climatic database with a strong focus on users rather than on the conventional principles of climatic databases. The Central European region may benefit from such a database, as climate data availability has been recognized for a long time as having many weaknesses with regard to regional climate change research. The accessibility of an ensemble of climate change scenarios, and the developed web services and GIS tools makes such data available for a broad community of non-climatologists, and in general, to experts who have not been able to access the climate data in obviously available formats. This fact may enhance the value and reliability of impact studies and consequently accelerate the transfer of scientific knowledge to management of natural resources.

We believe that such a database may act as a bridge between climate modelling groups and scientists dealing with climate change impact studies. We encourage other research groups to create similar databases to support the wider scientific community in the research of the currently changing environment.

Acknowledgements

References

Baigorria GA, Jones JW, Shin DW, Mishra A, O'Brien JJ. 2007. Assessing uncertainties in crop model simulations using daily bias-corrected regional circulation model outputs. *Climate Research* **34**: 211–222. doi:10.3354/cr00703.

Baigorria GA, Jones JW, O'Brien JJ. 2008. Potential predictability of crop yield using an ensemble climate forecast by a regional circulation model. *Agricultural and Forest Meteorology* **148**: 1353–1361. doi:10.1016/j.agrformet.2008.04.002.

Bhagat J, Tanoh F, Nzuobontane E, Laurent T, Orlowski J, Roos M, Wolstencroft K, Aleksejevs S, Stevens R, Pettifer S, Lopez R, Goble C. 2010. BioCatalogue: a universal catalogue of web services for the life sciences. *Nucleic Acids Research* **38**: 689–694.

Calanca P, Semenov MA. 2012. Local-scale climate scenarios for impact studies and risk assessments: integration of early 21st century ENSEMBLES projections into the ELPIS database. *Theoretical and Applied Climatology* **113**: 445–455. doi:10.1007/s00704-012-0799-3.

Carter TR. 2013. Agricultural impacts: multi-model yield projections. *Nature Climate Change* **3**: 784–786. doi:10.1038/nclimate1995.

Castro CL, SrRA P, Leoncini G. 2005. Dynamical downscaling: assessment of value retained and added using the Regional Atmospheric Modeling System (RAMS). *Journal of Geophysical Research* **110**: D05108. doi:10.1029/2004JD004721.

Christensen JH, Boberg F, Christensen OB, Lucas-Picher P. 2008. On the need for bias correction of regional climate change projections of temperature and precipitation. *Geophysical Research Letters* **35**: L20709. doi:10.1029/2008GL035694.

Cook BI, Terando A, Steiner A. 2010. Ecological forecasting under climatic data uncertainty: a case study in phenological modeling. *Environmental Research Letters* **5**: 044014. doi:10.1088/1748-9326/5/4/044014.

Déqué M. 2007. Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: model results and statistical correction according to observed values. *Global and Planetary Change* **57**: 16–26. doi:10.1016/j.gloplacha.2006.11.030.

Dobor L, Barcza Z. 2014. FORESEE database v2.0. Eötvös Loránd University, Hungary. doi: 10.5281/zenodo.9614.

Dosio A, Paruolo P. 2011. Bias correction of the ENSEMBLES high-resolution climate change projections for use by impact models: evaluation on the present climate. *Journal of Geophysical Research* **116**: D16106. doi:10.1029/2011JD015934.

Fielding RT, Taylor RN. 2002. Principled design of the modern web architecture. *ACM Transactions on Internet Technology* **2**: 115–150.

Fodor N, Mika J. 2011. Using analogies from soil science for estimating solar radiation. *Agricultural and Forest Meteorology* **151**: 78–85. doi:10.1016/j.agrformet.2010.09.006.

Giorgi F. 1990. Simulation of regional climate using a limited area model nested in general circulation model. *Journal of Climate* **3**: 941–963. doi:10.1175/1520-0442(1990) 003.

Giorgi F. 2006. Climate change hot-spots. *Geophysical Research Letters* **33**: L08707. doi:10.1029/2006GL025734.

Glassy JM, Running SW. 1994. Validating diurnal climatology logic of the MT-CLIM model across a climatic gradient in Oregon. *Ecological Applications* **4**: 248–257.

Goble C, De Roure D. 2009. The impact of workflow tools on data-centric research. In *Data Intensive Computing: The Fourth Paradigm of Scientific Discovery*, Hey T, Tansley S, Tolle K (eds.). Microsoft Research: Redmond, WA; 137–145.

Goddard L, Mason SJ, Zebiak SE, Ropelewski CF, Basher R, Cane MA. 2001. Current approaches to seasonal to interannual climate predictions. *International Journal of Climatology* **21**: 1111–1152. doi:10.1002/joc.636.

Gutowski WJ, Decker SG, Donavon RA, Pan Z, Arritt RW, Takle ES. 2003. Temporal-spatial scales of observed and simulated precipitation in central US climate. *Journal of Climate* **16**: 3841–3847. doi:10.1175/1520-0442(2003) 016.

Hansen JW, Challinor A, Ines A, Wheeler T, Moronet V. 2006. Translating forecasts into agricultural terms: advances and challenges. *Climate Research* **33**: 27–41. doi:10.3354/cr033027.

Hardisty A, Roberts D, The Biodiversity Informatics Community. 2013. A decadal view of biodiversity informatics: challenges and priorities. *BMC Ecology* **13**: 16–23. doi:10.1186/1472-6785-13-16.

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. *Journal of Geophysical Research* **113**: D20119. doi:10.1029/2008JD010201.

Hlásny T, Barcza Z, Fabrika M, Balázs B, Churkina G, Pajtík J, Sedmák R, Turcáni M. 2011. Climate change impacts on growth and carbon balance of forests in Central Europe. *Climate Research* **47**: 219–236. doi:10.3354/cr01024.

Ho CK, Stephenson DB, Collins M, Ferro CAT, Brown SJ. 2012. Calibration strategies: a source of additional uncertainty in climate change projections. *Bulletin of the American Meteorological Society* **93**: 21–26. doi:10.1175/2011BAMS3110.1.

Hungerford RD, Nemani RR, Running SW, Coughlan JC. 1989. *MTCLIM: A Mountain Microclimate Simulation Model*. USDA Forest Service: Ogden, UT, USA.

Ines AVM, Hansen JW. 2006. Bias correction of daily GCM rainfall for crop simulation studies. *Agricultural and Forest Meteorology* **138**: 44–53. doi:10.1016/j.agrformet.2006.03.009.

IPCC. 2000. Emissions scenarios. In *Contribution of Working Group III to the Special Report of the Intergovernmental Panel on Climate Change*. Nakicenovic N and Swart R (eds). Cambridge University Press: Cambridge, United Kingdom, 570.

IPCC. 2007. Climate change 2007: impacts, adaptation and vulnerability. In *Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Parry ML, Canziani , OF , Palutikof JP, dervan Linden PJ, Hanson CE (eds.). Cambridge University Press: Cambridge, UK; 976.

IPCC. 2013. Climate change 2013: the physical science basis. *In Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds.). Cambridge University Press: Cambridge, United Kingdom and New York, NY, USA; 1535.

Jönsson AM, Bärring L. 2011. Future climate impact on spruce bark beetle life cycle in relation to uncertainties in regional climate model data ensembles. *Tellus A* **63**: 158–173. doi:10.1111/j.1600-0870.2010.00479.x.

Klok EJ, Klein Tank AMG. 2008. Updated and extended European dataset of daily climate observations. *International Journal of Climatology* **29**: 1182–1191. doi:10.1002/joc.1779.

Li H, Sheffield J, Wood EF. 2010. Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. *Journal of Geophysical Research* **115**: D10101. doi:10.1029/2009JD012882.

Maraun D. 2012. Nonstationarities of regional climate model biases in European seasonal mean temperature and precipitation sums. *Geophysical Research Letters* **39**: L06706. doi:10.1029/2012GL051210.

Maurer EP, Brekke L, Pruitt T, Duffy PB. 2007. Fine-resolution climate projections enhance regional climate change impact studies. *Eos, Transactions American Geophysical Union* **88**: 504. doi:10.1029/2007EO470006.

Mearns LO, Schneider SH, Thompson SL, McDaniel LR. 1990. Analysis of climate variability in general circulation models: comparison with observations and changes in variability in $2xCO_2$ experiments. *Journal of Geophysical Research* **95.D12**: 20469–20490. doi:10.1029/JD095iD12p20469.

Mitchell TD, Carter TR, Jones PD, Hulme M, New M. 2004. A comprehensive set of high-resolution grids of monthly climate for Europe and the globe: the observed record (1901-2000) and 16 scenarios (2001-2100). Tyndall Centre Working Paper 55; 30.

Moss R, Babiker M, Brinkman S, Calvo E, Carter T, Edmonds J, Elgizouli I, Emori S, Erda L, Hibbard K, Jones R, Kainuma M, Kelleher J, Lamarque JF, Manning M, Matthews B, Meehl J, Meyer L, Mitchell J, Nakicenovic N, O'Neill B, Pichs R, Riahi K, Rose S, Runci P, Stouffer R, van Vuuren D, Weyant J, Wilbanks T, van Ypersele JP, Zurek M. 2008. *Towards New Scenarios for Analysis of Emissions, Climate Change, Impacts, and Response Strategies*. Intergovernmental Panel on Climate Change: Geneva; 132.

Nabuurs GJ, Päivinen R, Sikkema R, Mohren GMJ. 1997. The role of European forests in the global carbon cycle – a review. *Biomass and Bioenergy* **13**: 345–358. doi:10.1016/S0961-9534(97)00036-6.

Olesen JE, Carter TR, Díaz-Ambrona C, Fronzek S, Heidmann T, Hickler T, Holt T, Minguez MI, Morales P, Palutikof JP, Quemada M, Ruiz-Ramos M, Rubak GH, Sau F, Smith B, Sykes MT. 2007. Uncertainties in projected impacts of climate change on European agriculture and terrestrial ecosystems based on scenarios from regional climate models. *Climatic Change* **81**: 123–143. doi:10.1007/s10584-006-9216-1.

Piani C, Haerter JO, Coppola E. 2010. Statistical bias correction for daily precipitation in regional climate models over Europe. *Theoretical and Applied Climatology* **99**: 187–192. doi:10.1007/s00704-009-0134-9.

Seneviratne SI, Lüthi D, Litschi M, Schär C. 2006. Land atmosphere coupling and climate change in Europe. *Nature* **443**: 14. doi:10.1038/nature05095.

Supit I, Van Diepen CA, de Wit AJW, Wolf J, Kabat P, Baruth B, Ludwig F. 2012. Assessing climate change effects on European crop yields using the Crop Growth Monitoring System and a weather generator. *Agricultural and Forest Meteorology* **164**: 96–111. doi:10.1016/j.agrformet.2012.05.005.

Szabó P. 2008 Comparison of precipitation and temperature fields in different data sets used for evaluating Regional Climate Models at the Hungarian Meteorological Service. Research Report, 10.

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. www.carpatclim-eu.org.

Teutschbein C, Seibert J. 2010. Regional climate models for hydrological impact studies at the catchment scale: a review of recent model strategies. *Geography Compass* **4**: 834–860. doi:10.1111/j.1749- 8198.2010.00357.x.

Themeßl MJ, Gobiet A, Heinrich G. 2012. Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. *Climatic Change* **112**: 449–468. doi:10.1007/s10584-011-0224-4.

Thornton PE, Running SW. 1999. An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. *Agricultural and Forest Meteorology* **93**: 211–228.

Thornton PE, Hasenauer H, White MA. 2000. Simultaneous estimation of daily solar radiation and humidity from observed temperature and precipitation: an application over complex terrain in Austria. *Agricultural and Forest Meteorology* **104**: 255–271. doi:10.1016/S0168-1923(00)00170-2.

Van der Linden P, Mitchell JFB. 2009. ENSEMBLES: Climate Change and its Impacts: Summary of research and results from the ENSEMBLES project. Office; 160.

Varis O, Kajander T, Lemmelä R. 2004. Climate and water: from climate models to water resources management and vice versa. *Climatic Change* **66**: 321–344. doi:10.1023/B:CLIM.0000044622.42657.d4.

Wang Y, Leung LR, McGregor JL, Lee D-K, Wang W-C, Ding Y, Kimura F. 2004. Regional climate modeling: progress, challenges, and prospects. *Journal of the Meteorological Society of Japan* **82**: 599–1628.

White RH, Toumi R. 2013. The limitations of bias correcting regional climate model inputs. *Geophysical Research Letters* **40**: 2907–2912. doi:10.1002/grl.50612.

Wood AW, Leung LR, Sridhar V, Lettenmaier DP. 2004. Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic Change* **62**: 189–216. doi:10.1023/B:CLIM.0000013685.99609.9e.

Supporting information

The following supporting information is available as part of the online article:

Appendix S1. Interpretation of the users' needs based on a self-made questionnaire.
Appendix S2. Comparison of the E-OBS and the CRU TS 1.2 databases with the CarpatClim reference database.

© 2015 The Authors.

Geoscience Data Journal **2**: 1–11 (2015)   *Geoscience Data Journal* published by Royal Meteorological Society and John Wiley & Sons Ltd.

Appendix S3. Differences between the non-corrected/corrected climate model results and the observation-based dataset.

Appendix S4. Relationship between simulated and observed standard deviation of precipitation time series based on the non-corrected and the corrected climate model results.

Appendix S5. Comparison of simulated (uncorrected and corrected) and observed annual number of wet days.