INRAØ

Annals of
Forest Science

DATA PAPER

Interpolated daily temperature and precipitation data for Level II ICP Forests plots in Germany

Shah Rukh[1,2]* , Tobias Schad[1], Maximilian Strer[1], Marco Natkhin[1], Inken Krüger[1] , Stephan Raspe[3], Nadine Eickenscheidt[4], Rainer Hentschel[5], Andrea Hölscher[6], Philipp Reiter[7] and Tanja G. M. Sanders[1,8]

Abstract

Key message: A harmonized, comprehensive meteorological time series for 78 German intensive forest monitoring plots (Level II) has been made available from 1961 to 2019. The used hybrid spatial interpolation routine using simple linear regression and inverse distance weighting allows for gap filling of missing data and also for extrapolation outside measurement period to analyze long-term effects of climate on forest ecosystems. The dataset is available at https://www.openagrar.de/receive/openagrar_mods_00079174. The associated metadata are available at: https://metadata-afs.nancy.inra.fr/geonetwork/srv/fre/catalog.search#/metadata/433a028f-dfc8-4a7c-82af-b8d7efafd724.

Keywords: Climate data, Intensive forest monitoring, Linear regression, Inverse distance weighting, Spatial interpolation, ICP Forests Level II, Germany

1 Background

The intensive forest monitoring (Level II) is part of the International Co-operative Programme on Assessment and Monitoring of Air Pollution Effects on Forests (ICP Forests, http://icp-forests.net) under the umbrella of the Convention on Long-Range Transboundary Air Pollution of the United Nations Economic Commission for Europe (CLRTAP/UNECE). Across Europe, parameters such as meteorology, deposition, tree growth, and crown conditions are assessed on nearly 620 Level II plots following harmonized methods to study cause-effect relationships in forest ecosystems (Ferretti 2021). In Germany, data is available for a maximum of 100 Level II plots with some of them sharing open-field meteorological stations; 68 plots are mandatory under the German forest law (BMJ

1975) and operated by the forest research institutions of the federal states (Seidling 2005; Sanders et al. 2020).

In all surveys, data gaps can occur for various reasons (Sanders and Seidling 2012). However, daily meteorological observations such as air temperature and precipitation are key to assessing changes in forest ecosystems as they affect tree growth and vitality, nutrient cycles, and phenology (de Vries et al. 2014; Ruiz-Benito et al. 2020; Ziche and Seidling 2010).

To fill these gaps within the measured data, spatial interpolation procedures using a hybrid approach of linear regression and inverse distance weighting (Müller-Westermeier 1995) are used to interpolate daily temperature and precipitation utilizing the data from the German Weather Service (DWD) stations. The resulting interpolated data is validated against the available measured Level II data, and corrected for bias. Gaps within the measured data are eventually filled with the bias-corrected data.

---

*Correspondence: shah.rukh@thuenen.de

[1] Thünen Institute, Institute of Forest Ecosystems, Alfred-Möller-Straße 01, H. 41/42, 16225 Eberswalde, Germany
Full list of author information is available at the end of the article

BMC

2 Material and methods

2.1 Description of Level II plots

Meteorological data for 100 German Level II plots is available in the ICP Forests database. Of these plots, data from 78 plots covering at least ~20% of the 1996 to 2019 measurement period (about 5 years; number of days $N >= 1746$ out of 8766) was deemed suitable for validation and plot-specific bias correction after interpolation. The rest of the plots had sparse data (about 3 years; $N <= 1093$ out of 8766), with half-yearly measurements and completely missing months in-between. Therefore, they were not considered for validation and bias correction to avoid any implausibility. The plots are located along different environmental gradients across Germany. Climate data is recorded on open-field areas generally less than 2000 m from the main forest plots (Raspe et al. 2016). It is measured on a quasi-continuous basis and then aggregated as daily sums and means with a required degree of completeness (Raspe et al. 2016). The aim of this interpolation is to gain ready-to-use datasets which can subsequently be continued without changing older data series (Rukh et al. 2022).

2.2 Variables description

We interpolated the following variables and filled the missing gaps in the time series from 1996 until 2019 (see Table 1):

a. Daily average of air temperature in °C, denoted by T_{mean}
b. Daily minimum air temperature in °C denoted by T_{min}
c. Daily maximum air temperature in °C denoted by T_{max}
d. Daily sum of precipitation in mm denoted by P

Note that the complete dataset consists of two subsets. First dataset covers the measurement period from the start of 1996 until the end of 2019 on Level II plots. 1996 is the earliest year when the measurements of climatic parameters started. However, for some Level II plots, measurements of climatic parameters started later than 1996. In those cases, the missing values were also gap-filled after bias correction (see Section 2.4 and technical validation).

The second subset covers the timeline from the start of 1961 until the end of 1995. In this time period, no empirical measurements of climatic parameters exist on Level II plots. This subset contains only interpolated data, which was also corrected for bias.

After interpolation and bias correction, checks for goodness of fit were performed (Table 2).

2.3 DWD stations

From the German Weather Service (DWD), a total of 1213 climate stations are available across Germany (DWD CDCa 2021). Not all of them were necessarily active continuously throughout the observation period. We used them to interpolate T_{mean}, T_{min}, and T_{max}. To interpolate daily precipitation P, we used the data from the larger in number and therefore denser precipitation network of the German Weather Service (DWD CDCb 2021). It includes the observations from the abovementioned 1213 stations and from additional precipitation

Table 1 Descriptive summary of the dataset for the four interpolated variables. Shown here is the information on the measurement period, when empirical measurements were carried out, information on before the measurement period, quality control criteria, and total number of plots

	Variable	Daily mean air temperature	Daily min. air temperature	Daily max. air temperature	Daily precipitation sum
	Abbreviation	T_{mean}	T_{min}	T_{max}	P
Within measurement period	Time period	1996–2019	1996–2019	1996–2019	1996–2019
	Number of days within this time period	8766	8766	8766	8766
	Average number of daily measurements per plot within this time period which are missing[a]	2674	2696	2700	2838
	Percentage of missing measurements	31%	31%	31%	32%
Outside measurement period	Time period	1961–1995	1961–1995	1961–1995	1961–1995
	Number of days within this time period	12783	12783	12783	12783
Quality control	Min. plausible value	−50 °C	−50 °C	−50 °C	0 mm
	Max. plausible value	+50 °C	+50 °C	+50 °C	150 mm
	Min. daily completeness	95%	95%	95%	100%
Summary	Total number of plots	78	78	78	78

[a] Values on the missing days were filled with the interpolated values after bias correction

Table 2 Statistical summary of the climate data validation. Bias on 78 plots was calculated as a difference between the daily empirical and interpolated values and aggregated into mean for each plot. Mean bias in the table below depicts the mean value of the climatic variables on the plots, so do the SD, mean correlation, mean R^2, and mean RMSE

	Variable	T_{mean}	T_{min}	T_{max}	P
Before bias correction	Mean bias	0.52 °C	0.41 °C	0.44 °C	0.1 mm
	SD	1.05 °C	1.58 °C	1.27 °C	3.60 mm
	Mean correlation	0.989	0.969	0.988	0.752
	Mean R^2	0.978	0.940	0.977	0.57
	RMSE	1.26 °C	1.90 °C	1.51 °C	3.61 mm
After bias correction	Mean bias	0 °C	0 °C	0 °C	0 mm
	SD	1.02 °C	1.54 °C	1.21 °C	3.54 mm
	Mean correlation	0.989	0.97	0.989	0.754
	Mean R^2	0.978	0.941	0.978	0.57
	RMSE	1.02 °C	1.54 °C	1.21 °C	3.54 mm

Explanation of statistic: (1) *SD*, standard deviation of the mean bias; (2) mean correlation, mean Pearson's correlation coefficient of all the plots; (3) mean R^2, mean regression coefficient; (4) *RMSE*, root-mean-square error. 2, 3, and 4 are the agreement between the interpolated (before and after the bias correction) and the measured Level II data

monitoring network stations. This results in a total use of 5619 stations to interpolate precipitation. These extra precipitation monitoring stations were not necessarily active continuously throughout the observation period.

The geographical locations of the Level II plots as well as the DWD climate and precipitation stations within Germany are presented in Fig. 1.

2.4 Hybrid interpolation approach and bias correction

Based on the approach of Müller-Westermeier (1995), we performed the interpolation of the climate data in four steps (also see Fig. 2):

1. Parameterize a general linear relationship between any given climate variable k and the elevation h above the sea level in meters by linear regression and using all DWD stations:

$$k_{DWD} = b_{DWD} + a \cdot h_{DWD} \qquad (1)$$

where a and b_{DWD} denote the slope and the intercept of the regression, respectively. Equation 1 was run for each day between 1961 and 2019. For each day, only the active DWD stations during that time were used in the equation. For practical reason and to build a linear relationship with many points on *xy*-scale possible, we calculated the slope a using all at that time active DWD

stations within Germany, irrespective of the surrounding radius.

2. With knowledge of the slope a, it is possible to reduce the climatic parameter at each DWD station to sea level (Eq. 2):

$$b_{i,DWD} = k_{i,DWD} - a \cdot h_{i,DWD} \qquad (2)$$

3. After that, we applied inverse distance weighting (IDW) using $b_{i,\,DWD}$ from Eq. 2, to interpolate the climatic parameter reduced to sea level at the Level II plot (Eq. 3):

$$b_{LII} = \frac{\sum_{i=1}^{n} \frac{b_{i,DWD}}{d_i^2}}{\sum_{i=1}^{n} \frac{1}{d_i^2}} \qquad (3)$$

The interpolated value b_{LII} is the climatic parameter reduced to sea level at the corresponding Level II plot. The subscript i denotes the DWD stations present within the radius of 50 km of the respective Level II plot; n is the total number of DWD stations present within this radius and used for interpolation. The d_i denotes the distance of a DWD station i, maximum 50 km, from the respective Level II plot. This radius was selected to allow maximum "active" DWD stations for IDW within our timeline of interest.

4. In order to calculate the value at the actual elevation h_{LII} of the Level II plot, Eq. 4 has been used. Here, climatic variable k_{LII} denotes the value at a Level II plot.

$$k_{LII} = b_{LII} + a \cdot h_{LII} \qquad (4)$$

During the interpolation, biases arise due to systematic errors in the models (Luo et al. 2018), such as due to model parametrization to determine a climate variable, especially in case of precipitation which is inherently heterogenous (Herrera et al. 2010; Pan et al. 2001). Also, the statistical distributions of the measured and the interpolated data (modeled data, per se) might differ, and a correction should be applied (Ayar et al. 2021; Ivanov et al. 2018).

To correct for bias — the difference between the daily measured and interpolated climate value — we applied the method of linear scaling outlined by Luo et al. (2018); also see Lenderink et al. (2007). For daily temperature T_{mean}, T_{min}, and T_{max}, a correction factor was calculated as a difference between the monthly mean values of the
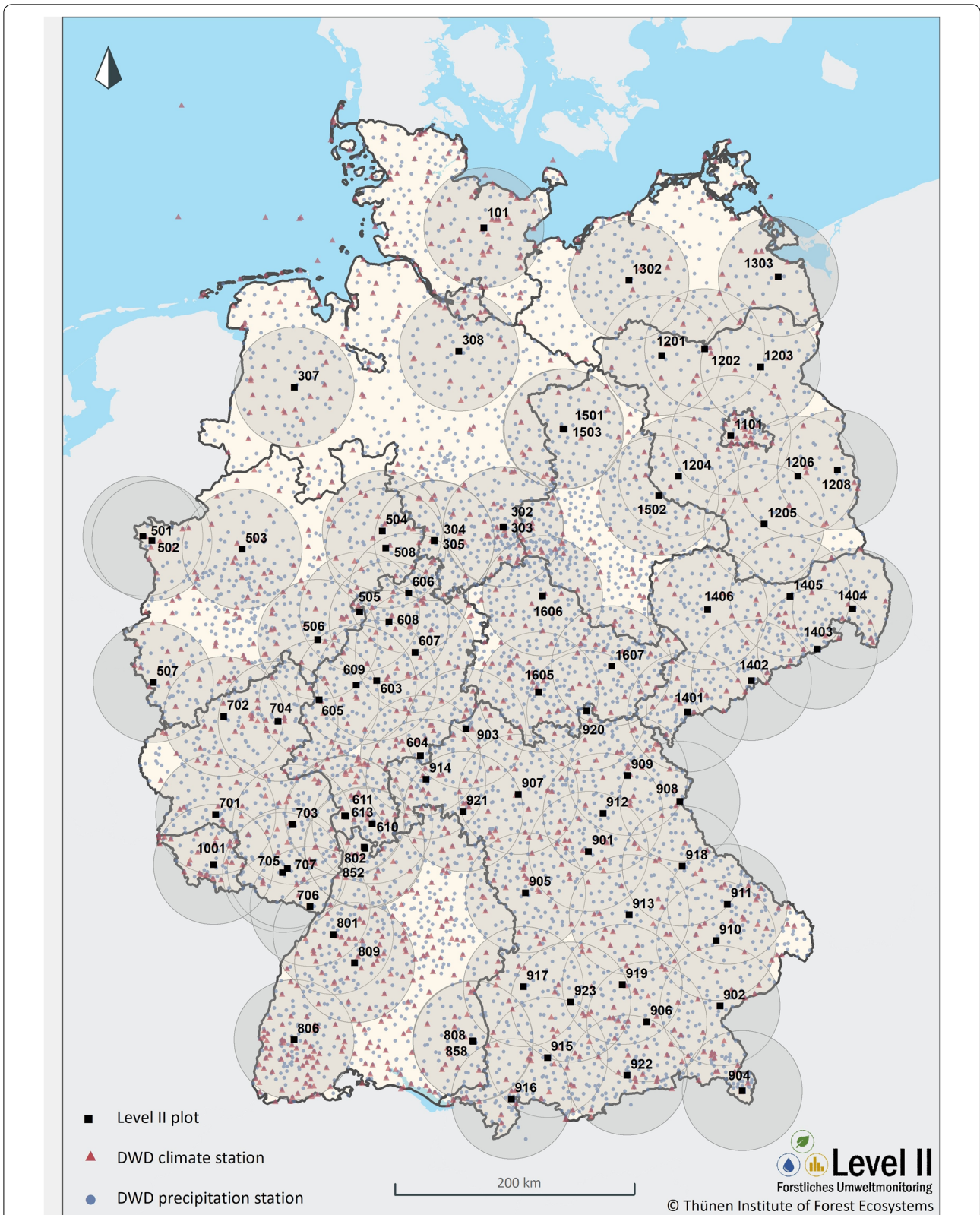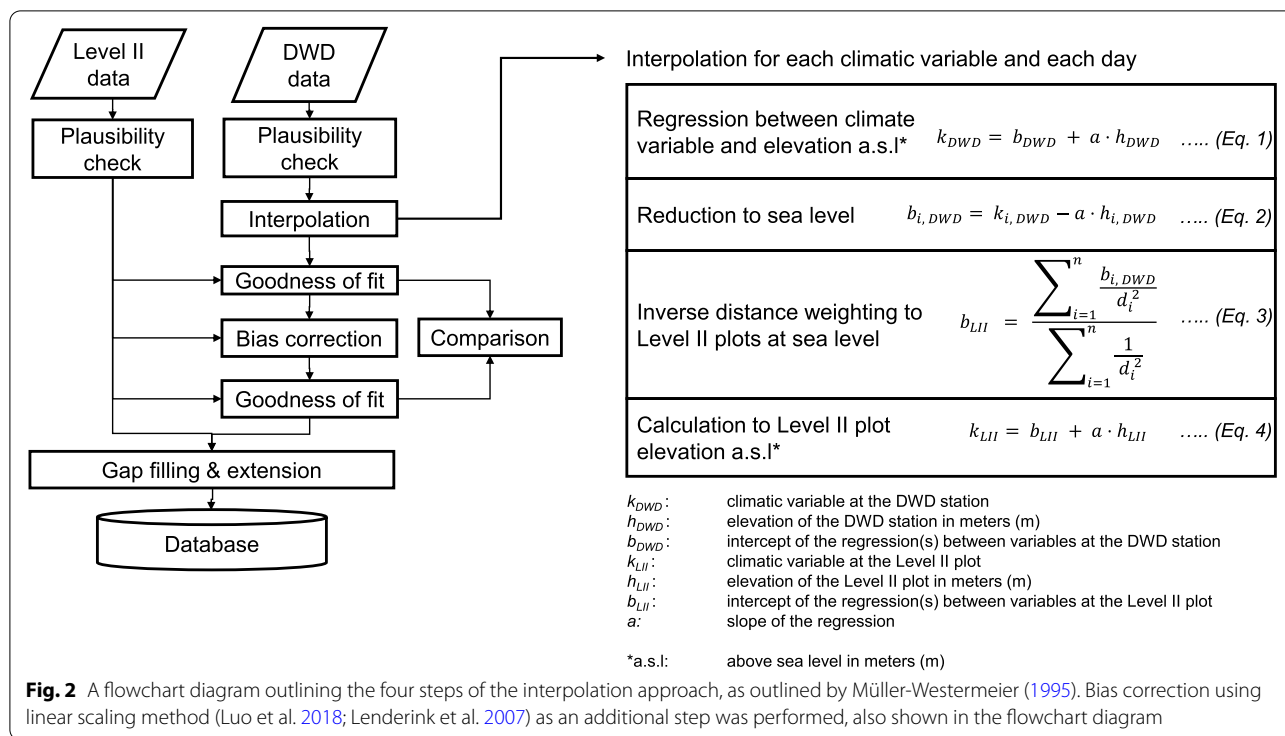
Fig. 1 Map showing the geographical locations of the ICP Forests Level II plots and DWD stations across Germany (source: DWD CDCa 2021; DWD CDCb 2021). The gray circles represent 50 km radius around the Level II plots

Rukh *et al. Annals of Forest Science*   (2022) 79:47

Page 5 of 7

Fig. 2 A flowchart diagram outlining the four steps of the interpolation approach, as outlined by Müller-Westermeier (1995). Bias correction using linear scaling method (Luo et al. 2018; Lenderink et al. 2007) as an additional step was performed, also shown in the flowchart diagram

daily measured temperature and the monthly mean values of the daily interpolated temperature. This was added to the daily interpolated temperature itself to correct for its bias.

$$T_{corr, daily} = T_{int, daily} + \left[T_{obs, monthly\ mean} - T_{int, monthly\ mean} \right] \quad (5)$$

$T_{corr, daily}$, $T_{int, daily}$, $T_{obs, monthly\ mean}$, and $T_{int, monthly\ mean}$ are the corrected daily temperature, interpolated daily temperature, monthly mean values of the daily measured temperature, and the monthly mean values of the daily interpolated temperature on Level II plots, respectively. The difference within the square brackets is the correction factor. T in Eq. 5 is valid for all three temperature variables T_{mean}, T_{min}, and T_{max}.

In case of precipitation P, the correction factor was calculated as a ratio between the monthly mean values of the daily measured precipitation and the monthly mean values of the daily interpolated precipitation. This factor was multiplied with the daily interpolated precipitation itself to correct for its bias.

$$P_{corr, daily} = P_{int, daily} \cdot \left[\frac{P_{obs, monthly\ mean}}{P_{int, monthly\ mean}} \right] \quad (6)$$

$P_{corr, daily}$, $P_{int, daily}$, $P_{obs, monthly\ mean}$, and $P_{int, monthly\ mean}$ are the corrected daily precipitation, interpolated daily precipitation, monthly mean values of the daily measured precipitation, and the monthly mean values of the daily interpolated precipitation, respectively. The ratio within the square brackets is the correction factor.

We would like to point out that this linear scaling method corrects for bias by calculating the correction factor which is specific to each month within the same year and then corrects the daily interpolated climate values of that month within that year. In our case, since we also interpolated the data from 1961 to 1995, applying this correction method was not possible since no measured data was available to calculate the correction factor. To deal with this, we calculated over the available timeline of the measured data from 1996 until 2019 a universal correction factor which was specific to each month but not specific within the same year and was also applicable to the interpolated data outside the measured timeline. $T_{obs, monthly\ mean}$, $T_{int, monthly\ mean}$, $P_{obs, monthly\ mean}$, and $P_{int, monthly\ mean}$ were calculated specific to each month from 1996 until 2019, irrespective of the year. We then applied the universal correction factor, calculated in the square brackets of Eqs. 5 and 6 for the measured time period, in the same manner to the daily interpolated data from 1961 to 1995 outside the measured time period.

3 Technical validation

Before the interpolation (Section 2.4), we performed plausibility checks on the Level II climate data as per the quality control guidelines listed in the ICP Forests manual (Raspe et al. 2016). We performed the same

Rukh *et al. Annals of Forest Science* (2022) 79:47

Page 6 of 7

plausibility checks on the DWD temperature and precipitation data. The criteria of the minimum daily completeness (%) of the data, as well as minimum and maximum plausible values for each of the variables to be interpolated, was used (see Table 1). The data, which did not fulfil these criteria, was discarded.

After interpolation, daily bias values (difference between daily interpolated and measured Level II data) were aggregated for each plot to reflect its mean bias. Standard deviation of the mean bias for each plot was also calculated. Pearson's correlation coefficient and coefficient of determination (R^2) depict agreement between the interpolated and the measured Level II data. Root-mean-square error (RMSE) qualifies the model performance for each variable in this case. Correcting for bias significantly improved the model performance for all the interpolated variables (Table 2) at $p < 0.01$ and brought the mean bias for each plot to zero. Moreover, we provide visual assessment files in .pdf format to depict the performance of the measured data against the bias-corrected data.

4 Reuse potential and limits

The daily gap-filled and extended time series of climatic variables can be aggregated to a chosen temporal scale. It offers opportunities to characterize climatic conditions on Level II plots based on, for example, climatic water balance. The complemented time series of temperature and precipitation also allows for calculation of drought indices such as the standardized precipitation evapotranspiration index (SPEI, Vicente-Serrano et al. 2010).

Our interpolation routine shows a flexible implementation of the used method. However, the routine does not cover checks for homogeneity. These checks along the temperature and precipitation time series could be performed in addition to correct for any structural breaks in the time series, which may arise due to spatial variability in the climate variables at different weather stations, especially for precipitation, but also by changes in measuring devices and plot surroundings. Nevertheless, our performed validation checks in Table 2 suggest a reasonably good performance of the interpolation routine and usability of the data. Based on this performance on a large dataset, the routine is suitable for the raw interpolated data as well, if options for bias correction are limited due to sparseness of measured data. For users, we also make raw interpolated data without bias correction available for the plots where measured data was sparse.

Additionally, we noted a few measured data points as anomalous. We provide their visual information under the folder "possible anomalies." We do not rule them out as measurement error. It is hence subject to user, if they want to replace those data points with the provided bias-corrected data.

5 Access to the data and metadata description

The initial, untreated meteorological Level II data of air temperature and precipitation is archived by Programme Co-ordinating Centre (PCC) of ICP Forests in Eberswalde, Germany. For use beyond 2019 in the future, the data is available on request at http://icp-forests.net via the official data request form. Requests are evaluated for the scientific purpose, and access is usually granted within 2 weeks.

The DWD climate and precipitation data used in interpolation procedure can be found freely available at DWD (CDCa 2021; CDCb 2021), respectively.

The processed data — gap-filled, bias-corrected, and the statistical evaluations — are archived in the repository found at https://www.openagrar.de/receive/openagrar_mods_00079174. It contains comma separated value tables (*.csv*) for daily mean, min, max air temperature, and daily precipitation. Be aware that each time series is split into two parts covering either the period from 1961 to 1995 or the period from 1996 to 2019, and only the later period includes the gap-filled time series. Dataset has been further categorized into bias-corrected and raw interpolated time series, specific to the plots. For transparency purposes, we also publish the untreated meteorological data from 1996 until 2019, for the users to have an insight into the whole dataset and not into the gap-filled and bias-corrected data only. The metadata file https://metadata-afs.nancy.inra.fr/geonetwork/srv/fre/catalog.search#/metadata/433a028f-dfc8-4a7c-82af-b8d7efafd724 provides comprehensive information on the available datasets and data structure within the repository, including the following:

1. A basic description of Level II plots including the plot code, plot coordinates, plot names, and elevation
2. Technical variable descriptions and its location within the data structure of the repository, i.e., names of the *.csv* files that contain the prepared data
3. Contact information of the authors

Rukh *et al. Annals of Forest Science*   (2022) 79:47

Page 7 of 7

provided as data supplement in OpenAgrar repository under the link https://www.openagrar.de/receive/openagrar_mods_00079174.

Author details
[1]Thünen Institute, Institute of Forest Ecosystems, Alfred-Möller-Straße 01, H. 41/42, 16225 Eberswalde, Germany. [2]Silviculture and Forest Ecology of the Temperate Zones, Faculty of Forest Sciences and Forest Ecology, Georg-August-University of Göttingen, Büsgenweg 01, 37077 Göttingen, Germany. [3]Department of Soil and Climate, Bavarian State Institute of Forestry (LWF), Hans-Carl-von-Carlowitz-Platz 01, 85354 Freising, Germany. [4]Department of Monitoring and Efficiency Control in Nature and Landscape Conservation, State Office for Nature, Environment and Consumer Protection North Rhine-Westphalia, Leibnizstraße 10, 45659 Recklinghausen, Germany. [5]State Centre of Excellence Eberswalde (LFE), Brandenburg State Office of Forests, Alfred-Möller-Straße 01, 16225 Eberswalde, Germany. [6]Department of Soil and Environment, Baden-Württemberg State Institute of Forestry, Wonnhaldestraße 4, 79100 Freiburg, Germany. [7]Research Institute of Forest Ecology and Forestry Rhineland-Palatinate, Hauptstraße 16, 67705 Trippstadt, Germany. [8]Center of Disturbance Ecology, University of Bayreuth, Universitätsstraße 30, 95447 Bayreuth, Germany.

References
Ayar PV, Vrac M, Mailhot A (2021) Ensemble bias correction of climate simulations: preserving internal variability. Sci Rep 11:3098. https://doi.org/10.1038/s41598-021-82715-1
Bundesministerium der Justiz BMJ (1975) Gesetz zur Erhaltung des Waldes und zur Förderung der Forstwirtschaft (BWaldG). Zuletzt geändert durch Art 112 G v. 10.8.2021 I 3436
De Vries W, Dobbertin MH, Solberg S, van Dobben HF, Schaub M (2014) Impacts of acid deposition, ozone exposure and weather conditions on forest ecosystems in Europe: an overview. Plant Soil 380:1–45. https://doi.org/10.1007/s11104-014-2056-2
DWD CDCa (2021) Historical daily station observations (temperature, pressure, precipitation, sunshine duration, etc.) for Germany. V006 2018. https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/daily/kl/historical/. Accessed 7 Feb 2021
DWD CDCb (2021) Historical daily precipitation observations for Germany. V007 2019. https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/daily/more_precip/historical/. Accessed 16 Feb 2021
Ferretti M (2021) New appetite for the monitoring of European forests. Ann For Sci 78:94. https://doi.org/10.1007/s13595-021-01112-w
Herrera SL, Fita L, Fernandez JM, Gutierrez (2010) Evaluation of the mean and extreme precipitation regimes from ENSEMBLES regional climate multimodel simulations over Spain. J Geo Res 115:117. https://doi.org/10.1029/2010JD013936
Ivanov MA, Luterbacher J, Kotlarski S (2018) Climate model biases and modification of the climate change signal by intensity-dependent bias correction. J Clim 31:6591–6610. https://doi.org/10.1175/JCLI-D-17-0765.1
Lenderink G, Buishand A, van Deursen W (2007) Estimates of future discharges of the river Rhine using two scenario methodologies: direct versus delta approach. Hydrol Earth Syst Sci 11:1145–1159. https://doi.org/10.5194/hess-11-1145-2007
Luo M, Liu T, Meng F, Duan Y, Frankl A, Bao A, De Maeyer P (2018) Comparing bias correction methods used in downscaling precipitation and temperature from regional climate models: a case study from the Kaidu river basin in Western China. Water 10:1046. https://doi.org/10.3390/w10081046
Müller-Westermeier G (1995) Numerisches Verfahren zur Erstellung klimatologischer Karten. Selbstverlag des Deutschen Wetterdienstes (Berichte des Deutschen Wetterdienstes, Offenbach am Main, p 193
Pan Z, Christensen JH, Arritt RW, Gutowski WJ Jr, Takle ES, Otieno F (2001) Evaluation of uncertainties in regional climate change simulations. J Geophys Res 106:17735–17751. https://doi.org/10.1029/2001JD900193
R Core Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna https://www.R-project.org/
Raspe S, Beuker E, Preuhsler T, Bastrup-Birk A (2016) Meteorological measurements. In: UNECE ICP Forests, Programme Co-ordinating Centre (ed.): Manual on methods and criteria for harmonized sampling, assessment, monitoring and analysis of the effects of air pollution on forests. http://www.icp-forests.org/manual.htm
Ruiz-Benito P, Vacchiano G, Lines ER, Reyer CPO, Ratcliffe S, Morin X, Hartig F, Mäkelä A, Yousefpour R, Chaves JE, Palacios-Orueta A, Benito-Garzón M, Morales-Molino C, Camarero JJ, Jump AS, Kattge J, Lehtonen A, Ibrom A, Owen HJF, Zavala MA (2020) Available and missing data to model impact of climate change on European forests. Ecol Modell 416:108870. https://doi.org/10.1016/j.ecolmodel.2019.108870
Rukh S, Schad T, Strer M et al (2022) Interpolated daily temperature and precipitation data for Level II ICP Forests plots in Germany. [dataset], vol V1. Open Agrar Repository https://www.openagrar.de/receive/openagrar_mods_00079174
Sanders T, Krüger I, Holzhausen M (2020) Das intensive Forstliche monitoring – Level II. Thünen-Institut, Bundesforschungsinstitut für Ländliche Räume, Wald und Fischerei. Project Brief 25, Braunschweig. https://doi.org/10.3220/PB1608106763000
Sanders TGM, Seidling W (2012) Quality aspects in intensive forest monitoring. GI Edition Proc 194:271–274
Seidling W (2005) Outline and examples for integrated evaluations of data from the intensive (Level II) monitoring of forest ecosystems in Germany. Eur J For Res 124:273–287. https://doi.org/10.1007/s10342-005-0083-5
Vicente-Serrano SM, Beguería S, López-Moreno JI (2010) A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. J Clim 23:1696–1718. https://doi.org/10.1175/2009JCLI2909.1
Ziche D, Seidling W (2010) Homogenisation of climate time series from ICP Forests Level II monitoring sites in Germany based on interpolated climate data. Ann For Sci 67:804. https://doi.org/10.1051/forest/2010051

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.