IOP Publishing Environ. Res. Lett. 17 (2022) 024005 https://doi.org/10.1088/1748-9326/ac48b3

ENVIRONMENTAL RESEARCH
LETTERS

CrossMark

OPEN ACCESS

RECEIVED
15 September 2021

REVISED
15 December 2021

ACCEPTED FOR PUBLICATION
6 January 2022

PUBLISHED
21 January 2022

LETTER

KrigR—a tool for downloading and statistically downscaling climate reanalysis data

Erik Kusch[1,*] and Richard Davy[2]

[1] Center for Biodiversity Dynamics in a Changing World (BIOCHANGE), Section for Ecoinformatics and Biodiversity, Department of Biology, Arhus University, Ny Munkegade 116, Aarhus, Denmark
[2] Nansen Environmental and Remote Sensing Center, Jahnebakken 3, Bergen, Norway
[*] Author to whom any correspondence should be addressed.

E-mail: erik.kusch@bio.au.dk

Keywords: climate change, high resolution climate data, kriging, R, statistical downscaling

Abstract
Advances in climate science have rendered obsolete the gridded observation data widely used in downstream applications. Novel climate reanalysis products outperform legacy data products in accuracy, temporal resolution, and provision of uncertainty metrics. Consequently, there is an urgent need to develop a workflow through which to integrate these improved data into biological analyses. The ERA5 product family (ERA5 and ERA5-Land) are the latest and most advanced global reanalysis products created by the European Center for Medium-range Weather Forecasting. These data products offer up to 83 essential climate variables at hourly intervals for the time-period of 1981 to today with preliminary back-extensions being available for 1950–1981. Spatial resolutions range from 30 × 30 km (ERA5) to 11 × 11 km (ERA5-Land) and can be statistically downscaled to study-requirements at finer spatial resolutions. Kriging is one such method to interpolate data to finer resolutions and has the advantages that one can leverage additional covariate information and obtain the uncertainty associated with the downscaling. The KrigR R-package enables users to (a) download ERA5(-Land) climate reanalysis data for a user-specified region, and time-period, (b) aggregate these climate products to desired temporal resolutions and metrics, (c) acquire topographical co-variates, and (d) statistically downscale spatial data to a user-specified resolution using co-variate data via kriging. KrigR can execute all these tasks in a single function call, thus enabling the user to obtain any of 83 (ERA5)/50 (ERA5-Land) climate variables at high spatial and temporal resolution with a single R-command. Additionally, KrigR contains functionality for computation of bioclimatic variables and aggregate metrics from the variables offered by ERA5(-Land). This R-package provides an easy-to-implement workflow for implementation of state-of-the-art climate data while avoiding issues of storage limitations at high temporal and spatial resolutions by providing data according to user-needs rather than in global data sets. Consequently, KrigR provides a toolbox to obtain a wide range of tailored climate data at unprecedented combinations of high temporal and spatial resolutions thus enabling the use of world-leading climate data through the R-interface and beyond.

1. Introduction

1.1. Climate data needs in the 21st century
With the onset of the Anthropocene, the numerous fields of study that investigate the effects of climate change require spatially and temporally consistent climate data at high spatial and temporal resolutions (Hewitt *et al* 2017, Bjorkman *et al* 2018, Trisos *et al* 2020). In response to this need, an ever-growing number of climate datasets have been created (Fick and Hijmans 2017, Karger *et al* 2017, Abatzoglou *et al* 2018, Beyer *et al* 2020, Navarro-Racines *et al* 2020) making use of observations, reanalysis products, climate model outputs, or some combination thereof. Historically, a vast majority of efforts of climate data product creation have prioritised spatial resolution over temporal resolution in-line with the widely accepted notion of

Table 1. Contemporary climate data sets. A comparison of contemporary high spatial resolution climate data sets which are widely used in analyses of climate impacts. Notes for data availability: [1]… 19 of these are bioclimatic variables which are derivatives of temperature and precipitation data; [2]… 1 of these is elevation data. See Davy and Kusch (2021) for a comparison of these datasets against products generated using KrigR.

Name	Time-period	Resolution: Spatial	Resolution: Temporal	Number of variables available	Reference
WorldClim 2.1 climatologies	1960–2018	1 km	59 years	26[1,2]	(Fick and Hijmans 2017)
WorldClim historical monthly weather data	1960–2018	21 km	1 month	3	
TerraClimate	1958–2019	16 km	1 month	14	(Abatzoglou *et al* 2018)
CHELSA	1979–2013	1 km	1 month	46[1]	(Karger *et al* 2017)
ERA5	1950—today	30 km	1 h	83	(Hersbach *et al* 2020)
ERA5-Land	1981—today	11 km	1 h	50	(Sabater 2017)

small-scale processes affecting large-scale patterns (Briscoe *et al* 2019, Rapacciuolo and Blois 2019). This has resulted in climate products at spatial resolutions of up to 30 arcseconds (∼900 m) which are typically available at monthly or climatological-mean temporal resolutions. In view of climate-change effects on microclimatic processes as well as the changing frequency and intensity of climatic extremes, this emphasis on spatial rather than temporal resolution, has led to a decreased ability to identify extreme events and their consequences (Maclean 2019). Consequently, there is a pressing necessity for the development and dissemination of climate data products that offer data at high spatial and temporal resolutions.

Accurate representation of environmental conditions is facilitated not just through high spatial and temporal resolutions of climate data, but also through a wide range of climate variables. For example, contemporary studies of environmental drivers of biological patterns and processes have focused on a wide range of environmental variables including (a) water-availability and temperature (de Keersmaecker *et al* 2015), (b) compound metrics such as drought indices (Seddon *et al* 2016), (c) bioclimatic variables (Bruelheide *et al* 2018), and (d) combinations of the former (Kling *et al* 2020). To the detriment of research efforts, rarely are all necessary climatic variables for a given study available from a single data product thus necessitating the combination of climate information from several data sources. However, each of the available and widely used high resolution climate data sets offer a unique configuration of variables, period covered, methodology and data background, and spatial and temporal resolution, which makes the combination of data from different sources difficult. See table 1 for an overview of a selection of contemporary climate data sets and their combination of spatial and temporal specifications as well as the number of climatic variables offered by each data product. Accordingly, the study of climatic processes and patterns would be better served by obtaining climate data from a single, internally consistent data source rather than a patchwork of data sets of varying quality and specification.

1.2. Climate reanalyses meet demands

Climate reanalysis products represent a major achievement of climate science (Buizza *et al* 2018). They meet the demand for high temporal resolutions and abundance of self-consistent climatic information criteria (see table 1). These products optimally combine a wide range of surface and satellite observations with a dynamical model in order to produce a self-consistent dataset which includes all essential climate variables (ECVs) (Sabater 2017, Hersbach *et al* 2020) effectively eliminating the need for retrieval of data from a multitude of climate products for a full picture of environmental conditions. Reanalyses therefore avoid many of the issues of purely observational products (e.g. WorldClim, CRU). The best reanalyses are often taken as a substitute for observations when studying climate processes and change (Parker 2016).

Two of the most recent, and the most advanced global climate reanalyses have been created by the European Centre for Medium Range Weather Forecasting (ECMWF): ERA5 (Hersbach *et al* 2020) and ERA5-Land (Sabater 2017). The ERA5 reanalysis uses a vast array of observations of the Earth system to constrain a numerical model of the ocean, sea ice, land, and atmosphere using an ensemble data assimilation framework. ERA5 has been demonstrated to improve on data accuracy compared to previously published climate data products (Tang *et al* 2021). The ERA5 dataset is also the first reanalysis product to make available the uncertainty information of its ten-member ensemble used to create the analysis. This uncertainty is a measure of both the observational uncertainty (which is included in the data assimilation framework) and the stochastic uncertainty. However, this does not account for uncertainty associated with the choice of model physics, which can also be important (Banks *et al* 2016). ERA5-Land (Sabater 2017) is a global land-surface reanalysis that dynamically downscales ERA5 to a resolution of 0.1° (11 km). See table 1 for an overview of ERA5(-Land) data product parameters.

Due to increased data accuracy, temporal resolution, provision of data uncertainty metrics, and number of climate variables provided, ERA5(-Land) products are arguably the most appropriate climate data products for contemporary studies which are informed by climatic envelopes of the last five to seven decades.

1.3. Limitations of climate reanalysis products

Despite the advantages of reanalyses, these products have not been widely adopted outside climate science. This is likely a consequence of their relatively coarse spatial resolution (Sabater 2017, Hersbach *et al* 2020). This limitation has motivated several groups to downscale reanalyses to create finer resolution data products (Karger *et al* 2017, Abatzoglou *et al* 2018). However, none of the existing high-resolution climate products account for the uncertainty in the underlying climate data, or in the downscaling technique effectively biasing user perceptions of their validity in local applications. These products also provide variables which are challenging to robustly statistically downscale to high (∼900 m) spatial resolution, such as precipitation (Gutmann *et al* 2012, Hewitson *et al* 2014), or have otherwise violated the assumptions behind statistical downscaling (Chilès and Delfiner 2012). In addition to spatial resolution mismatches with pre-existing data products, climate reanalysis data can prove challenging to retrieve for potential users. Rather than downloading pre-prepared data files, the user needs to make use of an application programming interface or a webform for retrieval of ERA5(-Land) data. Therefore, to make use of climate reanalysis data effectively, one needs to overcome the two limitations of (a) spatial resolution, and (b) data accessibility.

Here, we present the R package KrigR, which has been developed to address these limitations and create an R-integrated workflow toolbox for handling climate reanalysis data. KrigR can automatically acquire and statistically downscale climate variables using kriging — a Gaussian process regression technique for interpolation (Chilès and Delfiner 2012). This package can be used to obtain high spatial (∼900 m) and temporal resolution (hourly) climate data, together with the associated uncertainty. Higher spatial resolutions can be achieved when providing third-party covariate data.

2. Statistical downscaling

Studies relying on climatic information, such as macroecological studies, often make use of climate products at spatial resolutions of 30 arcsecond (∼900 m). This spatial resolution is roughly one order of magnitude finer than the highest spatial resolution available via ERA5-products (see table 1). This mismatch of spatial resolutions can be overcome through statistical interpolation methodologies such as kriging.

2.1. Statistical interpolation with kriging

Kriging is a two-step process that requires training data that we wish to downscale, and co-variate data both at the resolution of the training data and at our target spatial resolution (Chilès and Delfiner 2012). In the first step, we fit variograms to our training data and establish covariance functions with our co-variate data at the training resolution. This gives us functions which describe the spatial autocorrelation of our training data, and its relationship with our chosen co-variate(s). During the second step we predict the value of our variable at new locations, in this case at a higher spatial resolution, using co-variate data at the target resolution.

2.2. Accuracy of kriging and implications for biological studies

Kriging is a powerful statistical interpolation method capable of accurately interpolating a multitude of climate variables to high spatial resolution with consistent performance across temporal resolutions (Davy and Kusch 2021). A recent study of vegetation memory patterns across global drylands demonstrated no difference in biological interpretation of spatio-temporal model results when using climate data at native resolution or interpolated from coarser spatial resolution thus proving kriging to be a robust downscaling technique fit for use in time-series analyses in geospatial studies (Kusch *et al* 2021). Consequently, kriging products are expected to perform well in static geospatial approaches such as species distribution modelling efforts.

One major advantage to kriging over other statistical interpolation methods is that it preserves the uncertainty obtained when fitting the variogram, which gives us an uncertainty associated with the downscaled data. In KrigR this uncertainty is given as a standard deviation of the estimate (Hiemstra *et al* 2009). This statistical uncertainty can be combined with dynamical uncertainty which is calculated by taking the standard deviation of the ten-member ensemble from ERA5 data for a measure of total data uncertainty which can be used to highlight and investigate differences between climatic data sets (Davy and Kusch 2021) and should be propagated into downstream analyses.

3. Using the KrigR toolbox

We have prepared a comprehensive overview of how the KrigR package works which can be reached via the KrigR GitHub (https://github.com/ErikKusch/KrigR) page.

```
DE_Raw <- download_ERA(
  Variable = '2m_temperature',
  DataSet = 'era5-land',
  DateStart = '1995-01-02',
  DateStop = '1995-01-02',
  TResolution = 'day',
  TStep = 1,
  Extent = DE_shp,
  FileName = "DE_Raw",
  API_User = API_User,
  API_Key = API_Key)
```

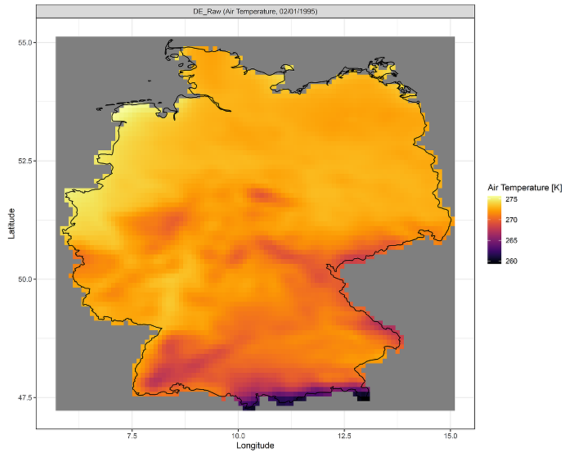

Figure 1. Obtaining ERA5(-Land) data. The download_ERA() function in KrigR has been designed to download and aggregate ERA5(-Land) to user-specifications. Here, we demonstrate the download of air temperature (2m_temperature) from the ERA5-Land data set for the 2nd of January 1995 across Germany (DE_shp, a shapefile loaded from Natural Earth 2021). Setting TResolution to 'day' and TStep to one results in aggregation of hourly data to a 1 d interval. By default, KrigR computes the mean for aggregated layers, however other functions for aggregation are supported (e.g. see figure 4). API_user and API_key are personal identifiers for the CDS webAPI.

3.1. Workflow with the KrigR package

The goal of the KrigR package is to make available state-of-the-art climate reanalysis data to R-users at user-specified spatial and temporal resolutions. KrigR does so via two routes: (a) the three-step process and (b) the pipeline.

The three-step process commences by (a) obtaining ERA5(-Land) data with calls to the ecmwfr R package (Hufkens *et al* 2019) and subsequent pre-processing to user specifications of either a rectangular area, a shape (e.g. a country border shapefile), or point-location data. See figure 1 for an example of this step.

The user specifies the target variable, climate dataset (ERA5 or ERA5-Land), geographic area, time-period and temporal resolution, and optional aggregate metric for the given period (e.g. minimum, maximum, mean, or sum). The download function can stage download requests either sequentially or in parallelised fashion (depending on user-specification) to circumvent bottle-necks in download times. For shorter time-series, the user may force the download as a singular call by using the SingularDL argument.

The second step (b) is obtaining and pre-processing the co-variate data. By default, KrigR provides GMTED2010 (Danielson and Gesch 2011) — a digital elevation model (DEM) output which provides global elevation data at a 30 arcsecond (∼900 m) resolution — to be used as a co-variate due to the demonstrated close relationship between elevation and a wide range of widely studied climate parameters (Daly *et al* 2002). The KrigR package downloads the DEM data, masks them to the area/shape/point-location-buffer the user specified and then aggregates the raw DEM data to the user's target resolution and the resolution of the training data. See figure 2 for an example.

Lastly, KrigR carries out (c) kriging as made available in R via the automap R package (Hiemstra *et al* 2009) of the raw ERA5(-Land) data obtained in step (a) (see figure 1) using the co-variates obtained in step (b) (see figure 2) which results in the output of (A) downscaled ERA5(-Land) data as well as the corresponding (B) statistical uncertainty of the downscaled data given as a standard error (Hiemstra *et al* 2009). See figure 3 for an example.

The KrigR package performs additional sanity-checks before kriging commences, allows for multi-core kriging using the cores argument, and stores temporary files so that the operation can be terminated and resumed without losing much progress (this functionality can be toggled off by setting Keep_Temporary = FALSE in the function call). The user can also specify the degree of localisation (using the nmax argument) used in the kriging which affects the estimate, uncertainty, and computational resources used. The computational cost of kriging using the KrigR package is discussed in further detail in Davy and Kusch (2021).

By default, the downloading and pre-processing functions in the KrigR package handle ERA5(-Land) and GMTED2010 data. However, the kriging function of the KrigR package is not limited to the use of these data sets. Third-party data can easily be introduced to the workflow to use the functionality of KrigR on any spatial product with any co-variate as supplied by the user. This is important for two reasons: (a) applicability of kriging has been demonstrated for non-climate spatial products (Bruelheide *et al* 2018) and users might also want to downscale other climatic data sets than ERA5(-Land),

```r
Covs_ls <- download_DEM(Train_ras = DE_Raw,
                        Target_res = .02,
                        Shape = DE_shp,
                        Keep_Temporary = TRUE)
```

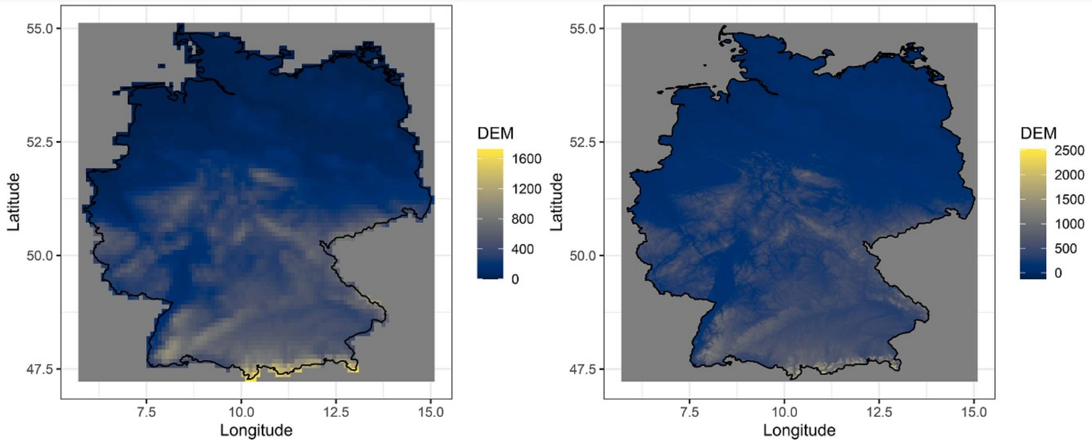

Figure 2. Obtaining covariate data. The download_DEM() function in KrigR downloads and prepares GMTED2010 data to user demands. Here, we demonstrate preparation of covariate data to be used for kriging of the data obtained in figure 1. KrigR downloads and resampled the GMTED2010 data set to match the raster data which will be kriged (DE_raw). This data need not be obtained via the download_ERA() function, but can be third-party spatial products in raster format. Next, the user specifies target resolution (0.02) or a raster at target resolution. A shapefile can be used to mask the covariate data to relevant regions. download_DEM() allows for storage of temporary files to circumvent repeated download of GMTED2010 data. This argument is shared between all KrigR functions.

```r
DE_Krig <- krigR(Data = DE_Raw,
                 Covariates_coarse = Covs_ls[[1]],
                 Covariates_fine = Covs_ls[[2]],
                 Keep_Temporary = TRUE,
                 Cores = 1,
                 FileName = "DE_Krig.nc",
                 nmax = 80)
```

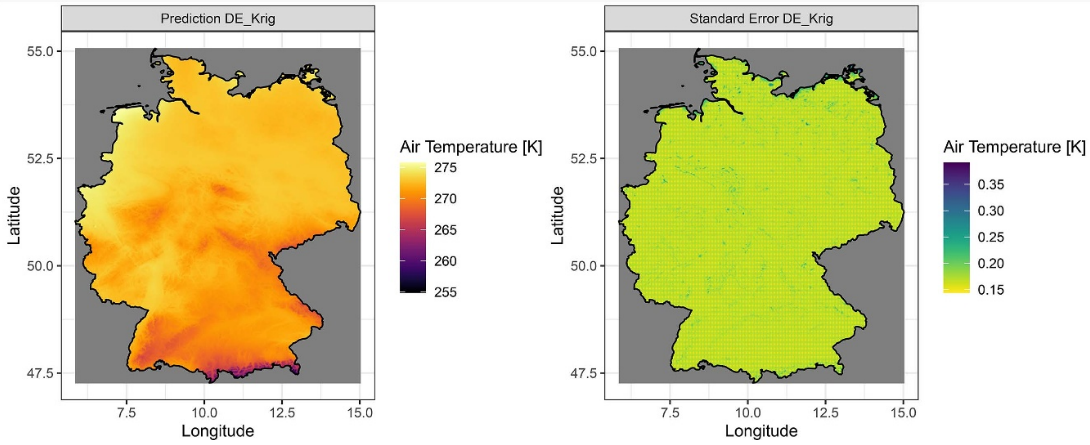

Figure 3. Kriging Using The Krigr R Package. The krigR() function in KrigR statistically interpolates spatial products to user demands. Here, we present the kriging of the data obtained in figures 1 and 2. The krigR() function produces a list containing three elements: (a) the statistical interpolation predictions pictured above, (b) the interpolation uncertainty reported as standard error of the kriging as depicted above, and (c) a summary for the function call not shown here.

(b) flexibility in choice of co-variates allows for accurate downscaling of a variety of climate variables and other spatial products. This flexibility in choice of covariates when using KrigR is demonstrated in Davy and Kusch (2021).

The second route to obtaining high-resolution climate data through KrigR is the pipeline. This involves a single function call that will automatically carry out all three steps explained above. Doing so does not allow for the use of third-party climate products

```
DE_Ens <- download_ERA(
  Variable = '2m_temperature',
  DataSet = 'era5',
  Type = "ensemble_members",
  DateStart = '1995-01-02',
  DateStop = '1995-01-02',
  TResolution = 'day',
  TStep = 1,
  FUN = sd,
  Extent = DE_shp,
  API_User = API_User,
  API_Key = API_Key)
```

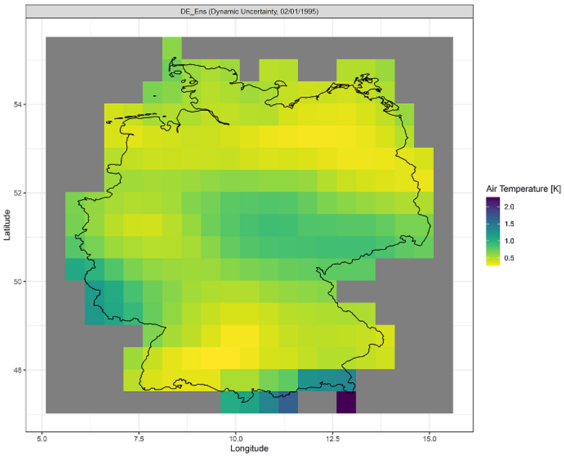

Figure 4. Dynamic uncertainty. The download_ERA() function can be used to obtain dynamical uncertainty by specifying the download of ensemble members from ERA5 and calculating their standard deviation in the aggregation step.

or co-variate data, effectively limiting the user to ERA5(-Land) data and the GMTED2010 co-variate data. Using the pipeline, a single function call can be used to run the entire process of data downloading, handling, and downscaling.

3.2. Data uncertainty

In addition to obtaining uncertainty pertaining to the statistical interpolation (i.e. statistical uncertainty), KrigR also enables users to obtain the uncertainty of the underlying data stored in ERA5. This is also referred to as dynamic uncertainty. See figure 4 for an example of how to use KrigR to obtain dynamic uncertainty.

As demonstrated by Davy and Kusch (2021), the magnitude of dynamic uncertainty is non-trivial when compared to statistical uncertainty using KrigR particularly at high temporal resolutions.

3.3. Bioclimatic variables

Due to the demonstrated usefulness of bioclimatic variables in biological studies, we have developed the BioClim() function for KrigR which automatically downloads all necessary data and carries out computation of bioclimatic variables as described by Fick and Hijmans (2017). The BioClim function can make use of functionality in KrigR thus allowing for (a) limitation of retrieved data to rectangular extents, shapes, or point-location data, (b) multi-core processing of data, (c) storing of temporary files for interruption of the computational process, and (d) full control over where to store temporary files and whether to delete them upon completion of the calculation of bioclimatic variables. Additionally, the BioClim() function allows users to specify for which temporal aggregate to identify extremes thus offering unmatched potential in the quantification of exposure to extreme events. This is of particular importance to bioclimatic variables which record extreme values such as BIO5 and BIO6 (maximum and minimum temperature, respectively) as well as variables reporting climate variability or ranges such as BIO7 (annual temperature range). Finally, the function has been conceptualised in such a way that water availability within the computed bioclimatic variables may be derived from precipitation values (as is the status quo in bioclimatic variables offered by other data products) or any other variable contained within the ERA5(-Land) data products such as soil moisture.

3.4. Kriging recommendations

For some variables, such as precipitation, the processes that determine their spatial pattern at finer resolutions than the training data are largely determined by atmospheric dynamics. Therefore, no combination of topographical co-variates is going to enable us to statistically downscale precipitation with high accuracy. We therefore do not recommend statistically downscaling precipitation data. However, there can be alternatives which also tell us about the water availability at high resolution, such as soil moisture, that we can successfully statistically downscale by using the soil properties and topographical properties as co-variates as has been demonstrated in Davy and Kusch (2021). To communicate this effectively to users, we are working on a list of recommendations of kriging specifications for individual climate variables to be implemented in the KrigR package in the near future.

3.5. Kriged climate projections

Using KrigR, it is also possible to create high-resolution climate projection products. Due to the high computational cost of kriging at large spatial scales, we recommend doing so at sub-global scales. Here, we demonstrate the KrigR workflow for projection kriging using the Coupled Model Inter-comparison Project Phase 6 (CMIP6) (Eyring *et al* 2016) simulations of historical climate and projections for the 21st century to create high

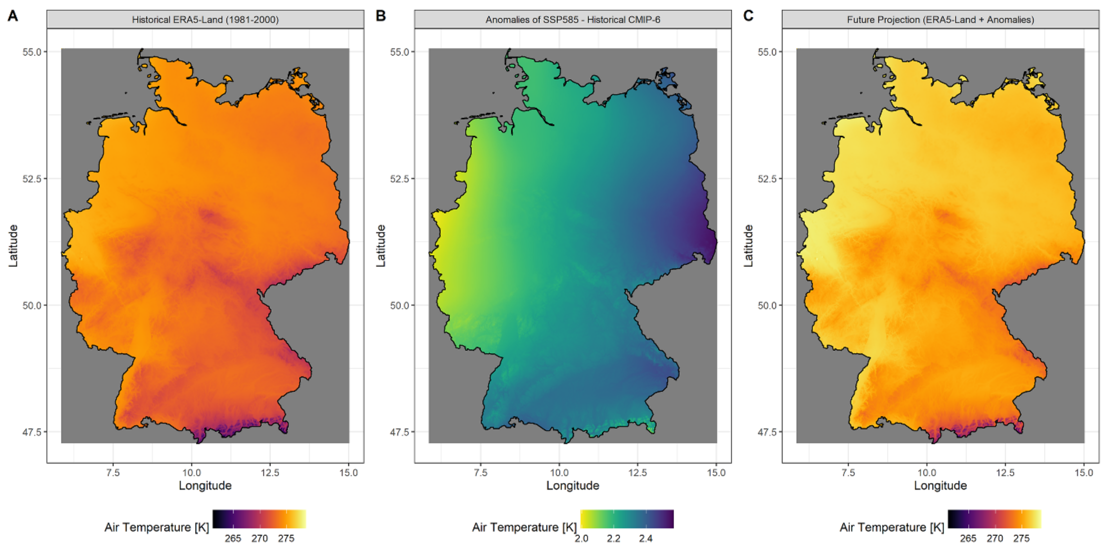

Figure 5. Creating downscaled projections. Using downscaled CMIP-6 historical (1981–2000) and SSP585 (2041–2060), we determined the anomalies between time-periods (B). Adding these to the downscaled ERA5-Land climatology for 1981–2000 (A), we obtain a climate projection for 2041–2060 according to the SSP585 scenario (C). This figure is displaying data for monthly climatologies of January.

resolution climatologies for future climate scenarios following the shared socioeconomic pathways derived from Integrate Assessment Models (O'Neill *et al* 2016). We have made a demonstration of how this can be done using the surface air temperature over Germany as an example (figure 5). First, we acquired surface air temperature data at monthly resolution from the full set of 36 available CMIP6 models for the historical simulations and the SSP585 scenario for the 21st century. We created monthly climatologies for the period 1981–2000 of the historical scenario, and for the period 2041–2060 using the SSP585 21st century scenario. Next, we downscaled each of these monthly climatologies to a 30 arcsecond (900 m) resolution using elevation as a co-variate. We then subtracted the monthly future climatology from the historical period to create monthly climatologies of temperature anomalies (see figure 5B). This was done to remove model biases in regional temperatures. Finally, we added these monthly climatologies of temperature anomalies to a downscaled ERA5-Land monthly climatology for the period 1981–2000 (see figure 5A) created using KrigR. In this way we make use of state-of-the-art CMIP6 projections for temperature changes in the 21st century under multiple realistic scenarios, while retaining the realistic spatial and seasonal variability in temperature obtained from the ERA5-Land reanalysis. The ensemble spread from the CMIP6 projections for a given scenario can also be used to obtain the projection-uncertainty of these future scenarios. However, integration of projection uncertainty, statistical uncertainty, and historical ERA5-Land uncertainty requires further research.

4. Conclusions

KrigR is a powerful, intuitive, and easy-to-use tool for acquiring and statistically downscaling state-of-the-art climate data. We have integrated the use of the ERA5 family of reanalysis products into KrigR. Currently, these are the most advanced reanalyses. KrigR offers a significant advantage in the field of high-resolution climate datasets by (a) leveraging the important advances behind the creation of the ERA5 reanalyses in terms of observations assimilated, the underlying dynamical model, and the data assimilation methodology; (b) offering access to the full range of ECVs from a single, consistent source at high temporal resolution; (c) providing the dynamical and statistical uncertainty associated with the high-resolution data, which allows for uncertainty propagation in downstream modelling efforts as well as a better understanding of data reliability; (d) offers great flexibility to tailor the data and study domain to user needs.

The ability in KrigR to pre-define spatial extent, timescale, period prior to data acquisition and carry out downloads/data processing through parallelised computation helps users overcome an important limitation of conventional workflows. With the rapid growth of climate datasets, the traditional workflow of downloading global data sets and subsequently cropping these to the required areas becomes unmanageable. This data management workflow issue has been an important topic in climate science for a decade (Overpeck *et al* 2011), but is extending to other domains where climate data is used. Thereby, KrigR is a tool with great capability to efficiently provide researchers and other users with climate data tailored to the needs of individual projects while being executed with just a few lines of code in a widely used open-source programming environment.

Acknowledgement

We thank the Nansen Environmental and Remote Sensing Center for covering the publication charges for this project.

Authors' contributions

R D and E K developed the framework for the KrigR package. E K developed the R Package KrigR. All authors contributed critically to the drafts and gave final approval for publication.

Conflict of interest

The authors declare no conflict of interest associated with this work.

Data availability

The data that support the findings of this study are openly available at the following URL/DOI: https://github.com/ErikKusch/KrigR. All data used here are freely and publicly available. ERA5(-Land) data come from the European Center for Medium range Weather Forecasting (cds.climate.copernicus.eu). The DEM data is available at the United States Geological Survey website (usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-global-multi-resolution-terrain-elevation). The CMIP6 data are made available by the Earth System Grid Federation and the data can be acquired from their data nodes (e.g. https://esgf-data.dkrz.de/projects/esgf-dkrz/). The code for production of downscaled climate projections is available at: https://github.com/ErikKusch/KrigRMS.

ORCID iDs

Erik Kusch https://orcid.org/0000-0002-4984-7646
Richard Davy https://orcid.org/0000-0001-9639-5980

References

Abatzoglou J T, Dobrowski S Z, Parks S A and Hegewisch K C 2018 TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015 *Sci. Data* **5** 1–12

Banks R F, Tiana-Alsina J, Baldasano J M, Rocadenbosch F, Papayannis A, Solomos S and Tzanis C G 2016 Sensitivity of boundary-layer variables to PBL schemes in the WRF model based on surface meteorological observations, lidar, and radiosondes during the HygrA-CD campaign *Atmos. Res.* **176–177** 185–201

Beyer R M, Krapp M and Manica A 2020 High-resolution terrestrial climate, bioclimate and vegetation for the last 120 000 years *Sci. Data* **7** 1–9

Bjorkman A D *et al* 2018 Plant functional trait change across a warming tundra biome *Nature* **562** 57–62

Briscoe N J *et al* 2019 Forecasting species range dynamics with process-explicit models: matching methods to applications *Ecol. Lett.* **22** 1940–56

Bruelheide H *et al* 2018 Global trait–environment relationships of plant communities *Nat. Ecol. Evol.* **2** 1906–17

Buizza R *et al* 2018 The EU-FP7 ERA-CLIM2 project contribution to advancing science and production of earth system climate reanalyses *Bull. Am. Meteorol. Soc.* **99** 1003–14

Chilès J P and Delfiner P 2012 *Geostatistics: Modeling Spatial Uncertainty* 2nd edn (New York: Wiley)

Daly C, Gibson W P, Taylor G H, Johnson G L and Pasteris P 2002 A knowledge-based approach to the statistical mapping of climate *Clim. Res.* **22** 99–113

Danielson J J and Gesch D B 2011 *Global Multi-Resolution Terrain Elevation Data 2010 (GMTED2010)* vol 2010 (Reston, VA: US Geological Survey) p 26 (available at: http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/GMTED2010) (Accesed 31 March 2021)

Davy R and Kusch E 2021 Reconciling high resolution climate datasets using KrigR *Environ. Res. Lett.* **16** 124040

de Keersmaecker W, Lhermitte S, Tits L, Honnay O, Somers B and Coppin P 2015 A model quantifying global vegetation resistance and resilience to short-term climate anomalies and their relationship with vegetation cover *Glob. Ecol. Biogeogr.* **24** 539–48

Eyring V, Bony S, Meehl G A, Senior C A, Stevens B, Stouffer R J and Taylor K E 2016 Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization *Geosci. Model Dev.* **9** 1937–58

Fick S E and Hijmans R J 2017 WorldClim 2: new 1 km spatial resolution climate surfaces for global land areas *Int. J. Climatol.* **37** 4302–15

Gutmann E D, Rasmussen R M, Liu C, Ikeda K, Gochis D J, Clark M P, Dudhia J and Thompson G 2012 A comparison of statistical and dynamical downscaling of winter precipitation over complex terrain *J. Clim.* **25** 262–81

Hersbach H *et al* 2020 The ERA5 global reanalysis *Q. J. R. Meteorol. Soc.* **146** 1999–2049

Hewitson B C, Daron J, Crane R G, Zermoglio M F and Jack C 2014 Interrogating empirical-statistical downscaling *Clim. Change* **122** 539–54

Hewitt C D, Stone R C and Tait A B 2017 Improving the use of climate information in decision-making *Nat. Clim. Change* **7** 614–6

Hiemstra P H, Pebesma E J, Twenhöfel C J W and Heuvelink G B M 2009 Real-time automatic interpolation of ambient gamma dose rates from the Dutch radioactivity monitoring network *Comput. Geosci.* **35** 1711–21

Hufkens K, Stauffer R and Campitelli E 2019 The ecwmfr package: an interface to ECMWF API endpointsTitle (available at: https://bluegreen-labs.github.io/ecmwfr/)

Karger D N, Conrad O, Böhner J, Kawohl T, Kreft H, Soria-Auza R W, Zimmermann N E, Linder H P and Kessler M 2017 Climatologies at high resolution for the earth's land surface areas *Sci. Data* **4** 1–20

Kling M M, Auer S L, Comer P J, Ackerly D D and Hamilton H 2020 Multiple axes of ecological vulnerability to climate change *Glob. Change Biol.* **26** 1–16

Kusch E, Davy R and Seddon A W R 2021 Vegetation memory effects and their association with vegetation resilience in global drylands (https://doi.org/10.1101/2021.08.22.457255)

Maclean I M D 2019 Predicting future climate at high spatial and temporal resolution *Glob. Change Biol.* **26** 1003–11

Natural Earth 2021 Natural earth data (available at: www.naturalearthdata.com/)

Navarro-Racines C, Tarapues J, Thornton P, Jarvis A and Ramirez-Villegas J 2020 High-resolution and bias-corrected

CMIP5 projections for climate change impact assessments *Sci. Data* **7** 1–14
O'Neill B C *et al* 2016 The scenario model intercomparison project (scenarioMIP) for CMIP6 *Geosci. Model Dev.* **9** 3461–82
Overpeck J T, Meehl G A, Bony S and Easterling D R 2011 Climate data challenges in the 21st century *Science* **331** 700–2
Parker W S 2016 Reanalyses and observations: what's the difference? *Bull. Am. Meteorol. Soc.* **97** 1565–72
Rapacciuolo G and Blois J L 2019 Understanding ecological change across large spatial, temporal and taxonomic scales: integrating data and methods in light of theory *Ecography* **42** 1247–66
Sabater J M *et al* 2017 ERA5-Land: A new state-of- the-art Global Land Surface Reanalysis Dataset *Earth Syst. Sci. Data* **13** 4349–83
Seddon A W R, Macias-Fauria M, Long P R, Benz D and Willis K J 2016 Sensitivity of global terrestrial ecosystems to climate variability *Nature* **531** 229–32
Tang W, Qin J, Yang K, Zhu F and Zhou X 2021 Does ERA5 outperform satellite products in estimating atmospheric downward longwave radiation at the surface? *Atmos. Res.* **252** 105453
Trisos C H, Merow C and Pigot A L 2020 The projected timing of abrupt ecological disruption from climate change *Nature* **580** 496–501