Computers and Geosciences 110 (2018) 65–72

Contents lists available at ScienceDirect

Computers and Geosciences

journal homepage: www.elsevier.com/locate/cageo

Research paper

Web processing service for climate impact and extreme weather event analyses. Flyingpigeon (Version 1.0)

Nils Hempelmann[a,*], Carsten Ehbrecht[b,**], Carmen Alvarez-Castro[a], Patrick Brockmann[a], Wolfgang Falk[c], Jörg Hoffmann[d], Stephan Kindermann[b], Ben Koziol[e], Cathy Nangini[a], Sabine Radanovics[a], Robert Vautard[a], Pascal Yiou[a]

[a] Le Laboratoire des Sciences du Climat et de l'Environnement, France
[b] German Climate Computing Center, Germany
[c] Bayerische Landesanstalt für Wald und Forstwirtschaft, Germany
[d] Julius Kühn-Institut - Federal Research Centre for Cultivated Plants, Germany
[e] NOAA Environmental Software Infrastructure and Interoperability Group/University of Colorado-Boulder, USA

ARTICLE INFO

Keywords:
Web processing service
Climate impact
Extreme weather events
Birdhouse
OGC

ABSTRACT

Analyses of extreme weather events and their impacts often requires big data processing of ensembles of climate model simulations. Researchers generally proceed by downloading the data from the providers and processing the data files "at home" with their own analysis processes. However, the growing amount of available climate model and observation data makes this procedure quite awkward. In addition, data processing knowledge is kept local, instead of being consolidated into a common resource of reusable code. These drawbacks can be mitigated by using a web processing service (WPS). A WPS hosts services such as data analysis processes that are accessible over the web, and can be installed close to the data archives.

We developed a WPS named 'flyingpigeon' that communicates over an HTTP network protocol based on standards defined by the Open Geospatial Consortium (OGC), to be used by climatologists and impact modelers as a tool for analyzing large datasets remotely.

Here, we present the current processes we developed in flyingpigeon relating to commonly-used processes (preprocessing steps, spatial subsets at continent, country or region level, and climate indices) as well as methods for specific climate data analysis (weather regimes, analogues of circulation, segetal flora distribution, and species distribution models). We also developed a novel, browser-based interactive data visualization for circulation analogues, illustrating the flexibility of WPS in designing custom outputs.

Bringing the software to the data instead of transferring the data to the code is becoming increasingly necessary, especially with the upcoming massive climate datasets.

1. Introduction

Processing of climate data is typically carried out by individual researchers, who create and run their own scripts in their preferred programming language, either locally or in environments internal to their institutions. Thus, there is a vast but unconnected body of knowledge that is not readily available to the climate science community with the risk of being continually replicated as researchers write scripts for processes that have already been well-developed by others. Furthermore, climate data, such as the upcoming Phase 6 of the Coupled Model Intercomparison Project CMIP6 and the Coordinated Regional Climate Downscaling Experiment CORDEX, is becoming too large to download and process locally.

Here, we present a web processing service (WPS) named 'flyingpigeon' (Version 1.0) containing processes written for and by climatologists and impact modelers for climate impact and extreme weather events analyses. These users are experts in their scientific fields with a good knowledge of climate model data usage, including the uncertainties associated with the data and the methods implemented in the processes.

Flyingpigeon (henceforth always referring to Version 1.0) is part of

* Corresponding author.
** Corresponding author.
E-mail addresses: info@nilshempelmann.de (N. Hempelmann), ehbrecht@dkrz.de (C. Ehbrecht).

https://doi.org/10.1016/j.cageo.2017.10.004
Received 20 August 2016; Received in revised form 8 May 2017; Accepted 6 October 2017
Available online 9 October 2017

the open source project birdhouse (under the Apache License 2.0), a collection of Open Geospatial Consortium (OGC) WPSs that provides data processing for the climate science community. Like all other compartments in birdhouse, flyingpigeon communicates over the web using the HTTP protocol based WPS Interface Standard for geospatial processing services defined by OGC (Mueller and Pross, 2015). A make file that handles appropriate software dependencies makes it easy to install. Flyingpigeon and the processes we developed are freely available from the flyingpigeon repository on GitHub. To run efficiently, the code should be installed on a system with appropriate resources.

Birdhouse evolved out of data management projects (C3Grid-INAD (Kindermann et al., 2012), ExArch (Castronova et al., 2013), LSDMA (Jung et al., 2014)), while the processes designed for flyingpigeon were developed within projects related to climate impact and extreme weather events analysis (EUCLEIA, A2C2, Extremoscope).

Inter-WPS communication enables operative services for international collaborations such as the Infrastructure for the European Network of Earth System Modeling (IS-ENES) and Earth System Grid Federation (ESGF), and is in line with other WPS developments like 52° North (Brauner, 2015) (enabling standardized deployment of geo-processes on the web), the ZOO-Project (Fenoy et al., 2013) (able to process geospatial or non geospatial data online), climate4impact, and Climate Information Portal of Copernicus (CLIPC (Déandreis et al., 2014)).

The goal of this paper is to introduce flyingpigeon as a WPS for climatologists and impact modelers.

2. WPS general description

A WPS is a technical solution (WPS Concepts) in which processes are hosted on a server and accessed over the web (Fig. 1). These processes conform to a standardized format, ensuring that they follow the principle of reusable design: they can be instantiated multiple times for different input arguments or data sources, customized following the same structure to handle new inputs, and are modular, hence can be combined to form new processes. In addition, a WPS can be installed close to the data to enable processing directly out of the archive. A WPS can also be linked to a theoretically limitless combination of several other WPSs, or generally OpenGIS Web Services (OWS).

In this paper **process** is used in the same sense as in the OGC standard: *'for any algorithm, calculation or model that either generates new data or transforms some input data into output data'* (Mueller and Pross, 2015). A submitted process is a **job**. A **service** provides a collection of processes containing scientific **methods** that focus on climate impact and extreme weather events. A combination of processes is called a **workflow**, and a collection of WPS-related software compartments is a **framework** (see Section 2).

WPS divides the operation into **server** and **client** side (see Section 2.3), with appropriate **security** (see Section 2.4) in between to avoid misuse.

2.1. Birdhouse

Birdhouse is a collection of WPS-related Python components to support data processing in the climate science community according to their own needs and use cases. In birdhouse, we currently use the Python implementation of WPS, PyWPS, but birdhouse is not restricted to a single WPS implementation. Birdhouse is not "yet another" processing framework, instead it provides the "glue" and the missing parts to successfully run WPS for climate data processing.

Birdhouse consists of several components like Flyingpigeon and Emu (see Fig. 2). Each of them can be installed individually. The installation is done using the Python-based build system Buildout and Ansible. Most of the dependencies are maintained in the Anaconda Python distribution. For convenience, each birdhouse component has a Makefile to ease the installation so you don't need to know how to call the build tools.

For managing and interacting with processing services, Birdhouse uniformly exposes OGC WPS standard based interfaces. The OGC WPS interface descriptions can be registered in an OGC Web Catalog Service supporting standards-based service discovery. Processing results can be published in the same Catalog Service.

Birdhouse has a web-client "Phoenix" to interact with web processing services and to feed them with data from climate data archives.

To control the user access to WPS services (and other OGC services), birdhouse has an OGC Web Service (OWS) security proxy "Twitcher" which can be placed in front of any WPS service.

Birdhouse has several web processing services which combine processes related to different aspects of climate data processing. Currently these are:

- **Flyingpigeon** contains a variety of processes ranging from simple polygon subsetting to complex data analysis methods and workflows used in climate impact or extreme weather event studies. Flyingpigeon is the main focus of this paper.
- **Hummingbird** provides processes to check conformance to climate metadata standards. These standards are the NetCDF-CF (Climate and Forecast conventions) and metadata conventions of climate data simulation projects like CORDEX and CMIP6.
- **Malleefowl** has processes to access climate data archives like the Earth System Grid Federation (ESGF) and Thredds data catalogs. It includes a workflow process to fetch climate data from a selected archive and provides this data to a selected analysis process. If the requested climate data files are not already locally available on disk, they will be downloaded and cached on the file-system.

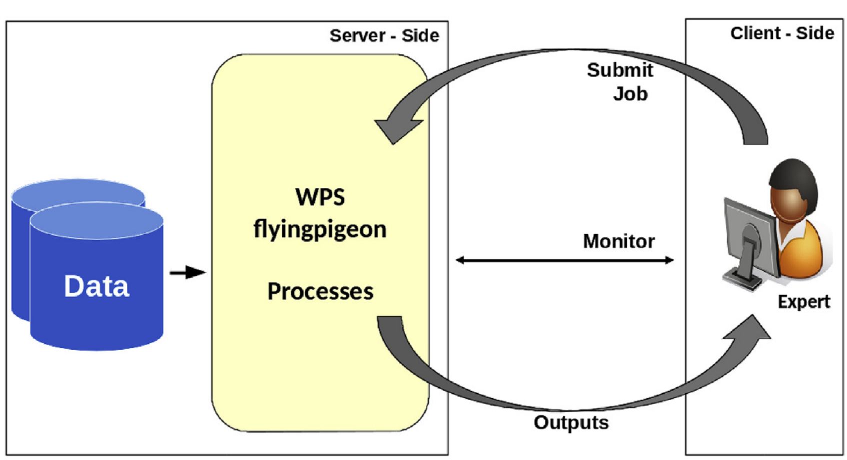

Fig. 1. Schematic of WPS operations.

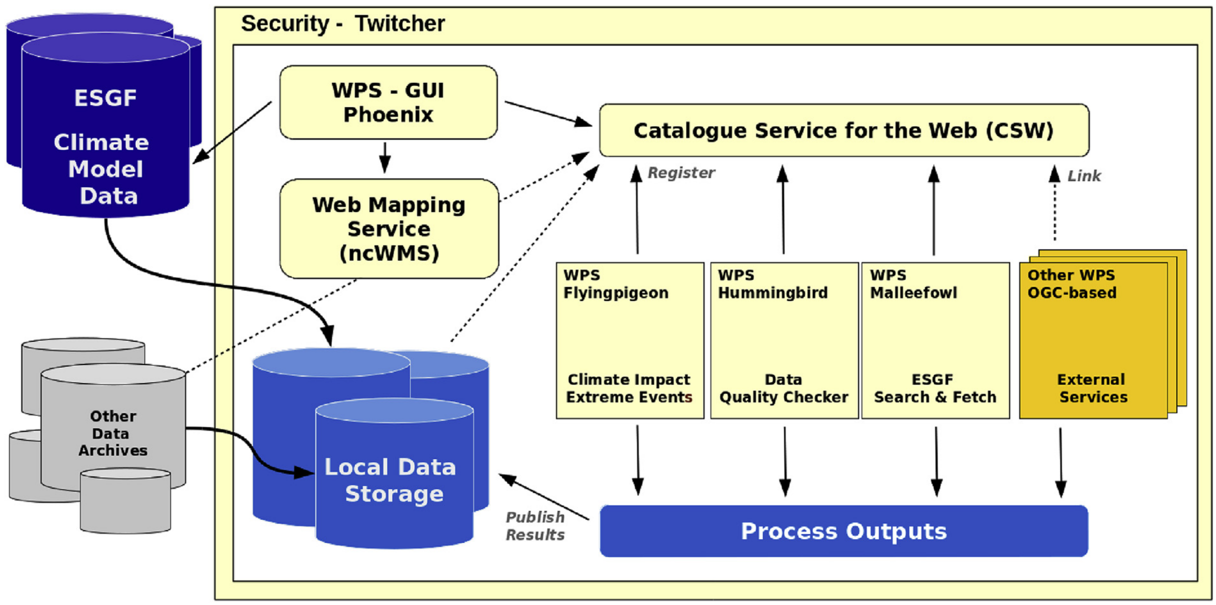

Fig. 2. Main components of the birdhouse framework and connection to data archives.

- **Emu** has some lightweight processes to show which input and output parameters are supported by WPS and provides examples for writing your own processes.

2.2. WPS flyingpigeon

Flyingpigeon is based on pywps4 and contains a variety of processes ranging from simple polygon subsetting to complex data analysis methods and workflows used in climate impact or extreme weather event studies (see Section 3).

One of the main software components of flyingpigeon is OpenClimateGIS (OCGIS), an open source Python package designed for geospatial manipulation, subsetting, computation, and translation of climate datasets stored locally in NetCDF files or served via OPeNDAP protocols. OCGIS interprets the climate data community's canonical metadata standard – the Climate and Forecast (CF) Convention – maintaining standards compliance for any derived CF-NetCDF output. OCGIS also supports numerous time-aware computations (i.e. monthly mean, seasonal maximum), multiple output formats (i.e. ESRI Shapefile, Comma Separated Value), and spatial interpolation using the Earth System Modeling Framework (ESMF). OCGIS calls can be executed on the whole dataset at once or divided into chunks to reduce the memory load. In flyingpigeon, memory availability is checked before and the call is executed accordingly.

Flyingpigeon is mainly written in Python but also includes code parts based on CDO commands, R scripts or Fortran and are called out of the python code. The birds hummingbird, malleefowl and emu currently implemented in birdhouse are also based on python but with respect of the OGC standard it is also possible to include WPS services (other birds) completely written in other languages than python.

After submitting a job to the WPS, several data preprocessing steps are run automatically. Flyingpigeon has utilities to e.g. sort files belonging to different datasets, checks the variable contained in the data files, converts units, rotates or unrotates grid coordinates, or generates file names based on the metadata or its values.

An installation system simplifies the deployment of flyingpigeon on a server. It is designed for Linux distributions and fetches required dependencies with the Conda package management system. It uses Buildout to setup the application by calling a simple install command.

Since the code is open source, contributions by the community are possible. Online documentation is generated with Sphinx (Brandl and Sphinx Team, 2016), including automatic documentation for the functions.

2.3. Server - client side

Flyingpigeon is a server-side service embedded in the birdhouse (see Fig. 2), a framework to support the development and deployment of climate data processing services based on the OGC WPS standard. Flyingpigeon can be run in combination with other birdhouse compartments such as **phoenix**, a graphical user interface (GUI) for common web browsers that includes an OGC web mapping service (WMS), or **malleefowl**, the WPS server backend for data search in and fetching from the ESGF archive. Other compartments offer processes such as quality checks for technical aspects of NetCDF files (climate forecast compliance checker) or processes to execute climate data operations with CDO. A catalog service for the web (CSW) is used to publish results and discover available services. However, flyingpigeon is a self-contained service and can also be run operationally as a stand-alone.

On the client side, the user needs to connect to the server with an appropriate URL to enable communication via an HTTP protocol, which is possible in three ways:

- with a terminal command
- within a script language (e.g. Python)
- with a browser-based GUI (e.g. Phoenix)

Besides connecting to the server via a terminal or a script language, a job can be submitted via the Phoenix GUI, directly or through a user-friendly wizard that guides through the steps from data search to literal argument input, and finally monitors the status of the job.

2.4. Security

Web Processing Services are HTTP services and access to them can be secured by using HTTPS and firewalls. One can enable basic authentication on the HTTPS service by requiring a username and password to use the WPS service. Most WPSs are used in an internal network and their processing capabilities are used by web portals. External users usually don't have direct access to the WPS itself.

Birdhouse uses Twitcher, a security proxy we developed for WPSs that provides an easy-to-use mechanism to access WPS directly by users in a secure way. The concept is not restricted to WPS and can be extended to other services like Web Mapping Services (WMS).

The Twitcher security proxy protects registered WPSs and allows the execution of processes over a terminal or within a script only if a valid access token is provided. A basic authentication via a GUI with username and password (Step 1 in Fig. 3) allows access to the Twitcher administration interface and triggers a token (unique string) generation. This token can be used to access the registered WPS services on the command line or in a script (Step 2 in Fig. 3). Tokens are valid only for a short period of time and can be regenerated by repeating Step 1. Without a valid token, it is not possible to execute a process, but it is still possible to retrieve information about the service to explore provided processes and their descriptions.

The management of the access tokens and the registered WPS is controlled by the Twitcher administration interface using the XML-RPC protocol (first layer in Fig. 3, left column). Access tokens are generated, validated and persisted by the Access Token Storage (second layer) which uses a MongoDB database. The OWS Service Registry (third layer) stores the registered WPS services with the service URL and a unique service name.

In Step 2, the Twitcher proxy service uses a WSGI application service (first layer in Fig. 3, right column), a specification for communication between web servers and web applications. The OWS Security Middleware (second layer) gets the provided token from the request (generated in Step 1 and embedded in the URL or an HTTP header variable) and validates it using the Access Token Storage. If the token is valid, the OWS Proxy (third layer) retrieves the service URL for the given service name and delegates the request to the registered service (WPS, WMS, etc.).

2.5. Input data

Climate model data are commonly stored in NetCDF file format and organized in data archives. If data are not already stored on the WPS server, they have to be fetched and stored in a structured way. Besides many other datasets, the output of the CMIP5 (and upcoming CMIP6) and CORDEX are stored for public access in the ESGF data archive. Birdhouse provides a data search for ESGF as a search interface within the GUI or from the command line. A second important group of datasets are reanalyses data, which are outputs from a data assimilation component of a weather forecast model. In climate impact and extreme event assessment, the use of reanalyses data is very common, thus some processes (see Section 3) in flyingpigeon provide a preselection of reanalyses datasets and variables. By selecting an appropriate variable, the data are automatically fetched within the process.

In the processes for analogue circulation (see Section 3.4.2) and weather regimes (see Section 3.4.1), the preselection of variables are dataset subsets of the following global reanalysis projects:

- NCEP/NCAR Reanalysis 1:
  NCEP data (Kalnay et al., 1996) are available from 1948 to the present in 17 pressure levels with a spatial grid resolution of $2.5^\circ \times 2.5^\circ$
- *20th Century Reanalysis version 2 (20CR):*
  The 20CR dataset (Compo et al., 2011) is based on surface pressure observations only and provides an ensemble of 56 members that contains global weather conditions and their uncertainty from 1871 to 2012, available on 24 pressure levels with a spatial grid resolution of $2^\circ \times 2^\circ$.

Local data stored on the server side can also be directly provided as input data for these two processes.

In climate impact and extreme event assessments, some analytic methods require non-climatic data. The species distribution model (SDM) process (see Section 3.3.2) in flyingigeon to predict favorability of tree species requires geographical coordinates of tree occurrences. A free and open access data base for biodiversity data is provided by the Global Biodiversity Information Facility (GBIF), where observations of trees (and other species) are stored and automatically fetched by the SDM process.

3. Processes embedded in flyingpigeon

Flyingpigeon is a service targeting climate researchers and experts in general with a focus on climate impact models and extreme event analyses. This section describes the processes that we developed so far.

3.1. Extracting spatial subsets

A standard operation in climate analysis is the extraction of data from a specific spatial region, a process known as polygon spatial subsetting. In flyingpigeon, the polygon subset process is possible at three levels of increasing spatial resolution: continents, world countries and European administrative regions obtained from the global administrative areas GADM database (*gadm26_levels.gdb*, *v*2.5). To optimize subsetting performance, high resolution GADM boundaries were simplified (Zhou and

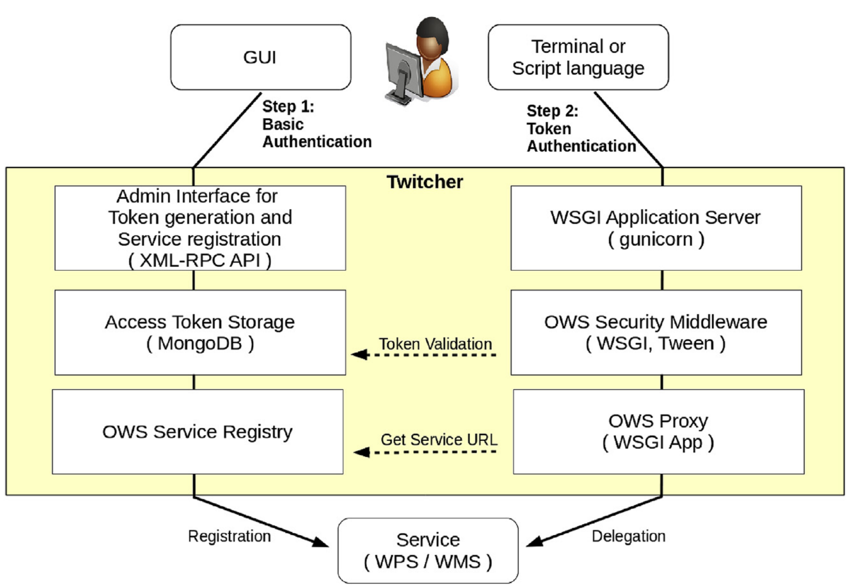

Fig. 3. Twitcher security schema.

Jones, 2005) using mapshaper (Harrower and Bloch, 2006) with a 1% point retention. The coarse resolution of target climate model inputs compared with the down-sampled resolution of GADM boundaries did not result in any significant data loss following a spatial subset operation. Very small administrative regions (e.g. cantons of Switzerland) were merged on the country level. For further details, please see the online documentation.

If more than one polygon is selected, a seperate netCDF file will be provided for each selected dataset. However, if the mosaic option is checked, the selected polygons are merged into one polygon. This increases the flexibility to meet specific user needs. The provided NetCDF files are wrapped in a tar archive.

In addition to polygons, a process is also available to extract longitude and latitude points. These point inspection data products are returned in text files containing a timeseries table for each longitude and latitude point.

3.2. Computation of climate indices

Climate indices are metrics to describe climate conditions and can be used to assess changes in climate over time (see the use case below), or as forcing data for complex impact models (e.g. the species distribution model see Section 3.3.2). Climate indices can be user-defined, but there is also a standardized set provided by the European Climate Assessment & Dataset website.

The climate indices are calculated based on a time aggregation (e.g. year, month, season) and include world countries polygon subsetting as an optional feature. The processes fall into four categories (the last two are still under development):

1. Standard
2. Percentile-based
3. Multi-variable indices
4. User-defined

Standard indices are based on one daily input variable (near surface temperature, precipitation). The provided indices are related to the ICCLIM Python package deployed in OCGIS. In the percentile-based process, a percentile is calculated over a user defined reference period. The process outputs the number of days that exceed or fall below this calculated threshold (for percentiles > 50 or < 50, respectively). When deployed, the multi-variable index process will allow users to specify a combination of indices. User-defined indices, also under development, will allow custom index calculations for single or multiple variables.

Use case: Extremoscope. We built an interactive data visualization to show the evolution of 10 extreme climate indices in the 13 regions of France under two climate scenarios (RCP 4.5 and 8.5) as part of the Extremoscope project. Flyingpigeon was used to calculate the climate indices using 7 climate projection models for five time aggregates (four seasons and yearly, see Fig. 4).

The visualization displays the probability of extreme seasons or years occurring, defined as exceeding the 90th/95th percentile, or falling below the 5th/10th percentile, based on a reference period between 1972 and 2100. The data is filterable by each dimension (region, time aggregate, index, scenario, percentile, model and year) with different charts that are linked (crossfiltered) using the crossfiltering library in dc.js, the Dimensional Charting JavaScript library, to allow the effects of each dimension to be observed and compared.

3.3. Climate impact related processes

One main target user group of the flyingpigeon is the climate impact community. In this section, the currently embedded processes for general or specific climate impact studies are presented. More processes like population dynamics for *anopheles gambia*, the vector transmitting *Plasmodium falciparum* to assess risks of malaria infection are under development and will be implemented in upcoming versions.

3.3.1. Segetal flora

'Segetal flora' is a term used for weeds growing in crop fields. The

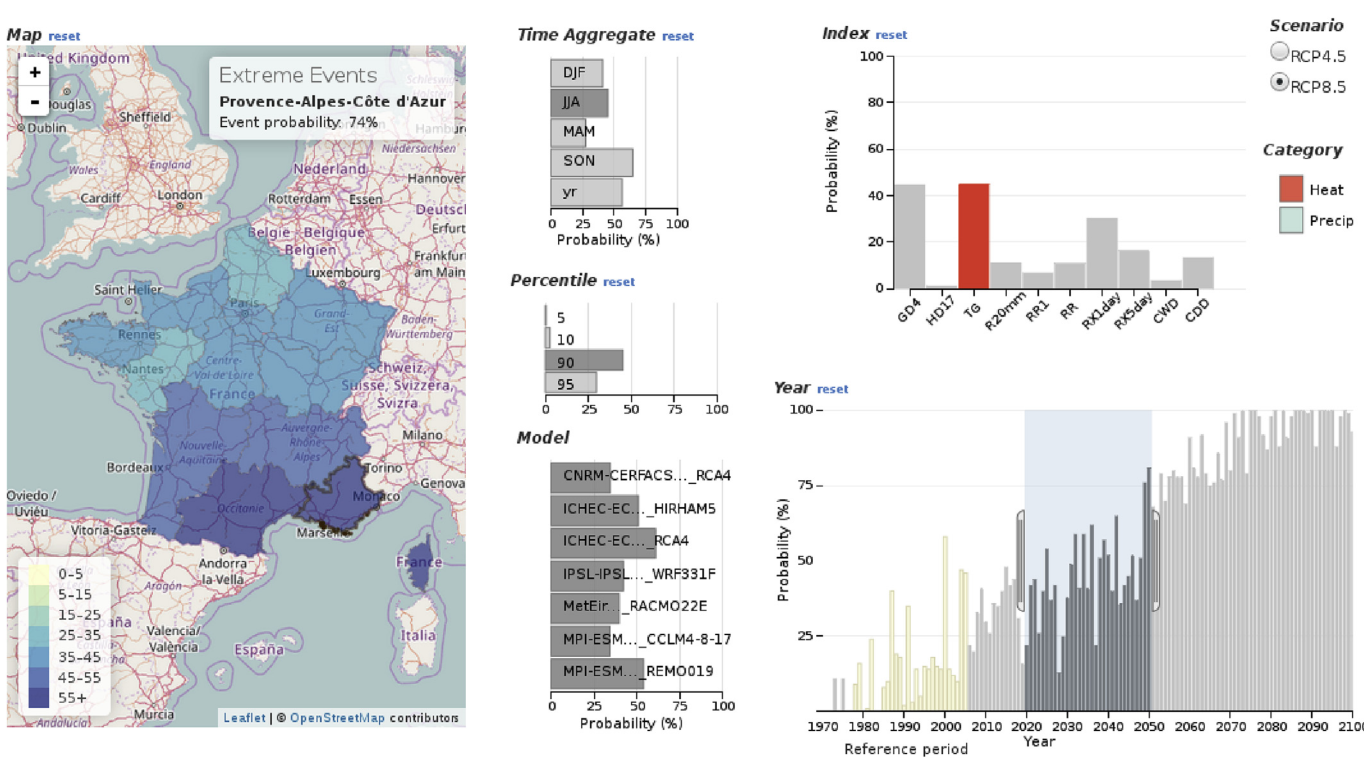

Fig. 4. Screenshot of the Extremoscope visualization showing the probability averaged over all models of extreme (>90th percentile) summer (JJA) temperatures over France from 2020 to 2050 under climate scenario RCP8.5.

majority of segetal flora species have very positive and important ecological effects as a source of pollen or nectar for several insects or as nutrition for birds (Hoffmann et al., 2013).

Based on field monitoring, a relation between the number of different species occurring in crop fields and the annual mean temperature can be shown. The relations are investigated separately for seven segetal flora groups (e.g. Mediterranean or Nordic groups) and three land use types (conventional and ecological land use, and one–two year self-greened fallow land). The species number to mean annual temperature relations are expressed as regression functions (Hoffmann et al., 2012; Hempelmann and Ehbrecht, 2014), which are the core of the segetal flora process. With temperature at 2 m height (tas) as input data, and arguments to select segetal flora type and land use, the number of species is predicted. A polygon subset for a specified country can be selected with an optional argument.

The process returns a tar file containing NetCDF files with the appropriate number of segetal flora species per grid point for each input dataset. The calculation is only relevant for areas with agricultural land use.

3.3.2. Species distribution model

Species distribution models (SDMs) are numerical tools to describe the relation between the distribution of a species and the environmental conditions that are thought to lead to this distribution (Franklin, 2010; Peterson et al., 2011). Models are used to gain ecological insights and predict distributions across landscapes, sometimes requiring extrapolation in space and time, e.g. for climate change impact modeling (modified from (Elith and Leathwick, 2009)).

We implemented a SDM process that relates the distribution of tree species to climate conditions to analyze the impacts of climate change (Falk and Hempelmann, 2013). The method focuses on tree species where the distribution is independent of non-climate factors like soil or specific site conditions. We hypothesize that the climate conditions which limit the tree species distribution (e.g. heat, precipitation, winter temperatures, drought (Mellert et al., 2015)) can be described with climate indices (defined in Section 3.2).

Fig. 5 shows the schematic workflow of the SDM process. Climate indices are calculated based on climate model variables (daily temperature and precipitation values). Tree species data can be fetched from the global biodiversity information facility (GBIF) database and fed to the SDM process. The geographical coordinates of the tree occurrence are then translated into a presence-absence (PA) matrix with the grid size of the input climate model data. Ocean and large lake areas are excluded in the PA matrix.

Statistical training (Wood, 2004, 2011) is performed based on the calculated climate indices as a mean over a reference period and the PA matrix using generalized additive models (GAMs) (Wood, 2006; Wood and Augustin, 2002). GAMs are flexible regression models with non-parametric (smooth) additive components (Wood, 2000).

After the statistical training, the probability of occurrence is predicted over the entire time series and transformed into a favorability (Real et al., 2006) to compare different models. The SDM process can be used with different climate indices, tree species, training performance values, and climate model datasets.

3.4. Extreme weather events related processes

Besides climate impact studies, the WPS flyingigeon also focuses on extreme weather events investigations. This section contains examples of more complex processes to calculate weather regimes (see Section 3.4.1) or analogues of circulation (see Section 3.4.2).

3.4.1. Weather regimes

Weather regimes are recurring states of the atmosphere and provide a useful description of atmospheric variability (Michelangeli et al., 1995; Corti et al., 1999; Vautard, 1990). In extreme event studies, for instance, it is possible to associate some specific seasonal weather regimes with the extremes of surface variables (precipitation or temperature) by analyzing which regime is prevailing when an extreme is encountered (Yiou and Nogaj, 2004).

Following the methods of (Michelangeli et al., 1995) and (Yiou et al., 2008), the weather regime process computes a given number of regimes (the κ value, by default $\kappa = 4$ (Michelangeli et al., 1995)) for a region (by default the North Atlantic region [80°W - 50°E, 20°N - 70°N]) using values of sea level pressure (SLP) or geopotential height (Z_g(h)) anomalies of a given season.

The weather regimes are computed with a k-mean classification algorithm (Hartigan and Wong, 1979) on the first 10 principal components (PCs) of the SLP or geopotential height (Z_g(h)) fields. The PCs are computed from an Empirical Orthogonal Function decomposition (Von Storch and Zwiers, 2001) of the field. The data are weighted by the square root of the cosine of the latitude to account for grid cell surface variations. The obtained PCs are then classified onto κ clusters. The weather regimes can be computed on a reference data set (e.g. reanalyses data). Then, other datasets (e.g. climate model data) can be classified according to these reference weather regimes.

Fig. 6 shows the output for a use case where weather regimes in winter months were calculated based on NCEP sea surface pressure data over a region in the North Atlantic for $\kappa = 4$. The process also output an R workspace containing the statistical training values as well as a text file with the PCs and the NetCDF file with the normalized pressure values. These outputs can be used for manual postprocessing or for the weather regime projection process which calculates the percentage of each given weather regime in every year. Such a projection was done in this example for a CMIP5 dataset and Table 1 shows the calculated values.

3.4.2. Analogues of circulation

Analogues of circulation provide a versatile tool to investigate the relation between climate variables (such as temperature or precipitation) and large-scale atmospheric circulation patterns (SLP or Z_g(h)).

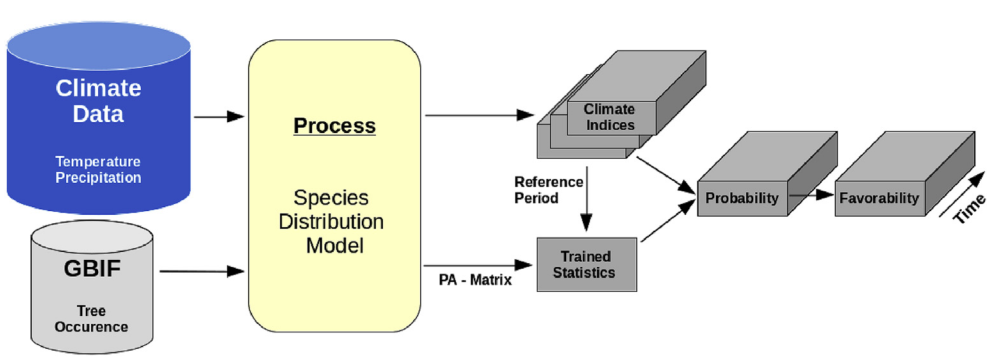

Fig. 5. Schema of the Species Distribution Model process.

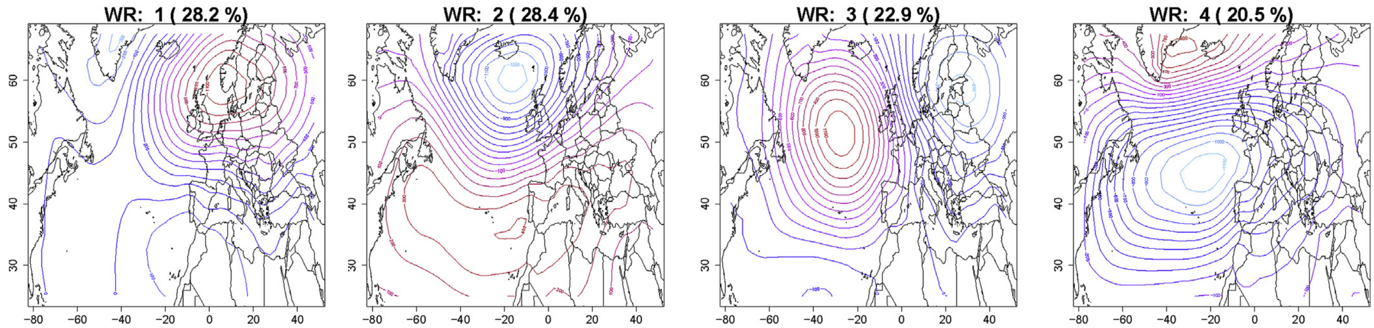

Fig. 6. Weather regimes (WR) for winter months (DJF) based on NCEP sea surface pressure data (1970–2010).

Table 1
Percentual occurrence of weather regimes in winter months (DJF) trained on NCEP sea surface pressure data projected on a CMIP5 global dataset.

Year	WR 1	WR 2	WR 3	WR 4
2084	35.16	28.57	30.77	5.49
2085	33.33	24.44	36.67	5.56
2086	21.11	28.89	40.00	10.00
2087	37.78	11.11	10.00	41.11
2088	18.68	19.78	37.36	24.17
2089	34.44	44.44	17.78	3.33

For an SLP/Z_g(h) pattern on a given day, the idea is to select days that have a calendar proximity (i.e. within a time window around the given date in all years except the year of the given day) and that minimize a distance between the circulation patterns. This approach has been used to infer climate reconstructions from SLP fields (Schenk and Zorita, 2012; Yiou et al., 2013), in weather forecast predictability assessment (Lorenz, 1969), downscaling of climate variables (Zorita and von Storch, 1999; Vaittinada Ayar et al., 2016), detection/attribution studies (Vautard and Yiou, 2009; Stott et al., 2016; Cattiaux et al., 2010) and stochastic weather generators (Yiou, 2014).

Analogues detection process. The analogues from this flyingpigeon process are mainly used to study the conditional attribution of extreme events to dynamics, thus describing only the thermodynamic changes in weather events. For a given continuous extreme event period (e.g. an extreme monthly temperature), the process seeks in past archives flows similar to those in the continuous period and reconstructs a monthly temperature from these past flow analogue days. The resulting temperature is then compared to the current one, providing an estimate of thermodynamic changes.

The process uses a circulation analogue simulation FORTRAN code (CASTf90) to compute analogues of circulation. The user can choose a time period for which daily analogues are computed and an archive period, that is, the time period to resample from. The user can also select a distance (Euclidean, Mahalanobis, correlation, Teweles-Wobus S1 (Teweles and Wobus, 1954), how many analogues per day to retain, a rectangular region (in longitude–latitude), whether to work with anomalies or not and some more options like the output file format. The result is a list of analogue dates for each simulation day and the corresponding distance values.

Analogues viewer process. To explore the properties of the analogues detection process, we developed an analogues viewer process that produces an interactive data visualization output as an HTML page hosted on the server-side. This viewer uses dc. js, the Dimensional Charting JavaScript library and a crossfiltering library to produce interactive charts that can be filtered based on data parameters. For each filter selection, all charts are simultaneously updated (crossfiltered).

Fig. 7 shows a schematic of the analogues workflow with a screenshot of the visualization. For this example, the winter of 2010 was selected in the bar plot, which filtered all charts to display counts of only those analogues pertaining to this time period. The display was further filtered by selecting analogues with correlation coefficients >0.5. The decadal distribution of analogues and changes across time can be immediately seen by sliding the time window.

The analogues viewer demonstrates two powerful advantages of using WPS for data analysis: 1) it is straight-forward to 'chain' processes together so that the output of one process can be used as the input of another, allowing data analysis to be re-imagined as a workflow, and 2) processes can be completely custom-designed. Here, we integrate data visualization capability – normally a step performed in an environment

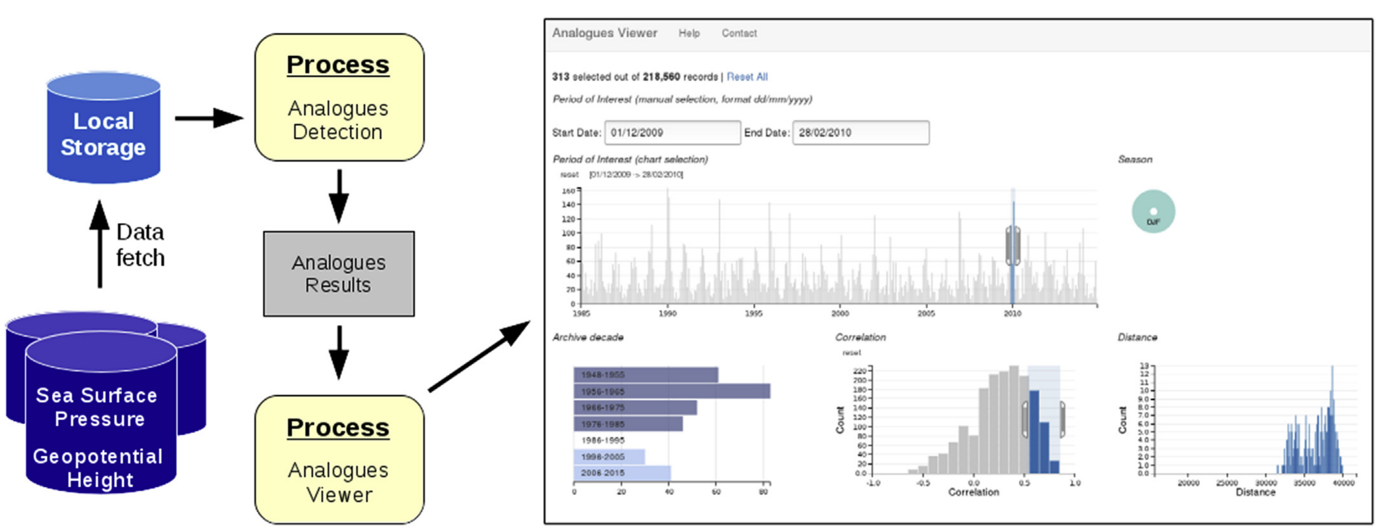

Fig. 7. Schema of the analogues workflow.

completely separate from the analysis and requiring a specialized skill that is often not within the scope of the scientific researcher – directly into flyingpigeon, accessible simply by launching the output HTML link.

4. Conclusion

Flyingpigeon, as a compartment of birdhouse, is an advanced toolbox that can be used to process the increasingly large amount of climate model data in a standardized and secure way. This current version is an initial selection of processes commonly used for climate impact and extreme weather event analysis. Flyingpigeon, connected to other OWSs, has access to various convenient features such as ESGF data search, visualization via the WMS, and catalog services. It allows collaborative data and methods sharing and access to several data archives. Process outputs can be recycled in other processes in a workflow. All these advantages are possible regardless of the script language of the methods.

The transparency of the source code as an open source project enables quick and easy exchange of developer knowledge, good quality control and frequent updates of the analysis methods and performance improvements.

Flyingpigeon reduces the difficulty of data processing and is a solution to facilitate the daily work for the climate community.

5. Outlook

The increasing number of developers and users ensures further improvement of features, documentation, guidelines and tutorials. Currently, high-performance computing providers located close to the data archives are being established in the climate community to keep up with the demand for WPS.

Acknowledgments

NCEP Reanalysis data was obtained from NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their website http://www.esrl.noaa.gov/psd/.

20th Century Reanalysis V2 data was obtained from NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their website http://www.esrl.noaa.gov/psd/.

CAC, SR and PY are supported by the ERC Advanced Grant No. 338965-A2C2.

References

Brandl, G., Sphinx Team, 2016. Sphinx. Python Documentation Generator.

Brauner, J., 2015. Formalizations for Geooperators-geoprocessing in Spatial Data Infrastructures (Dissertation). In: Fakultät Umweltwissenschaften, vol 10. Technische Universität Dresden.

Castronova, A.M., Goodall, J.L., Elag, M.M., 2013. Models as web services using the open geospatial consortium (ogc) web processing service (wps) standard. Environ. Model. Softw. 41 (0), 72–83.

Cattiaux, J., Vautard, R., Cassou, C., Yiou, P., Masson-Delmotte, V., Codron, F., 2010. Winter 2010 in europe: a cold extreme in a warming climate. Geophys. Res. Lett. 37 (20), L20704.

Compo, G.P., et al., 2011. The twentieth century reanalysis project. Q. J. Roy. Meteor. Soc. 137 (654), 1–28.

Corti, S., Molteni, F., Palmer, T.N., 1999. Signature of recent climate change in frequencies of natural atmospheric circulation regimes. Nature 398 (6730), 799–802.

Déandreis, C., Pagé, C., Braconnot, P., Bärring, L., Bucchignani, E., de Cerff, W.S., Hutjes, R., Joussaume, S., Mares, C., Planton, S., Plieger, M., 2014. Towards a dedicated impact portal to bridge the gap between the impact and climate communities : lessons from use cases. Clim. Change 125 (3), 333–347.

Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. Annu. Rev. Ecol. Evol. Syst. 40 (1), 677–697.

Falk, W., Hempelmann, N., 2013. Species favourability shift in Europe due to climate change: a case study forfagus sylvatical. andpicea abies(l.) karst. based on an ensemble of climate models. J. Climatol. 1–18, 2013.

Fenoy, G., Bozon, N., Raghavan, V., 2013. Zoo-project: the open wps platform. Appl. Geomatics 5 (1), 19–24.

Franklin, J., 2010. Mapping Species Distributions: Spatial Inference and Prediction. Cambridge University Press.

Harrower, M., Bloch, M., July 2006. Mapshaper.org: a map generalization web service. IEEE Comput. Graph. Appl. 26 (4), 22–27.

Hartigan, J.A., Wong, M.A., 1979. Algorithm as 136: a k-means clustering algorithm. J. R. Stat. Soc. Ser. C Appl. Stat. 28 (1), 100–108.

Hempelmann, N., Ehbrecht, C., 2014. Web processing services for climate data - with examples for impact modelers. EGI Community Forum 2014.

Hoffmann, J., Hempelmann, N., Glemnitz, M., Radics, L., Czimber, G., Wittchen, U., 2012. Einfluss von temperatur und nutzung auf die floristische artenvielfalt in getreideanbaugebieten europas. Julius-Kühn-Archiv 436, 70–76.

Hoffmann, J., Wittchen, U., Hempelmann, N., Glemnitz, M., Radics, L., 2013. Wildpflanzen auf dem acker - möglichkeiten und grenzen für den artenschutz im Ökolandbau. Forschungs Report Spezial Ökologischer Landbau, pp. 20–21.

Jung, C., Gasthuber, M., Giesler, A., Hardt, M., Meyer, J., Rigoll, F., Schwarz, K., Stotzka, R., Streit, A., 2014. Optimization of data life cycles. J. Phys. Conf. Ser. 513 (3), 032047.

Kalnay, E., et al., 1996. The NCEP/NCAR 40-year reanalysis project. Bull. Am. Meteorol. Soc. 77 (3), 437–471.

Kindermann, S., Schintke, F., Fritzsch, B., April 2012. A collaborative data management infrastructure for climate data analysis. Geophys. Res. Abstr. 14 (EGU201), 10569.

Lorenz, E.N., 1969. Atmospheric predictability as revealed by naturally occurring analogues. J. Atmos. Sci. 26 (4), 636–646.

Mellert, K.H., Deffner, V., Küchenhoff, H., Kölling, C., Nov 2015. Modeling sensitivity to climate change and estimating the uncertainty of its impact: a probabilistic concept for risk assessment in forestry. Ecol. Model. 316, 211–216.

Michelangeli, P.-A., Vautard, R., Legras, B., 1995. Weather regimes: recurrence and quasi stationarity. J. Atmos. Sci. 52 (8), 1237–1256.

Mueller, M., Pross, B., 2015. OGC® WPS 2.0 Interface Stand. Corrigendum 1 (03).

Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M., Araújo, M.B., 2011. Ecological Niches and Geographic Distributions (MPB-49) (Monographs in Population Biology). Princeton University Press.

Real, R., Barbosa, A.M., Vargas, J.M., Jun 2006. Obtaining environmental favourability functions from logistic regression. Environ. Ecol. Stat. 13 (2), 237–245.

Schenk, F., Zorita, E., 2012. Reconstruction of high resolution atmospheric fields for northern europe using analog-upscaling. Clim. Past 8 (5), 1681–1703.

Stott, P.A., Christidis, N., Otto, F.E.L., Sun, Y., Vanderlinden, J.-P., van Oldenborgh, G.J., Vautard, R., von Storch, H., Walton, P., Yiou, P., Zwiers, F.W., 2016. Attribution of extreme weather and climate-related events. WIREs Clim. Change 7, 23–41.

Teweles, S., Wobus, H.B., 1954. Verification of prognostic charts. Bull. Am. Meteorol. Soc. 35 (10), 455–463.

Vaittinada Ayar, P., Vrac, M., Bastin, S., Carreau, J., Déqué, M., Gallardo, C., 2016. Intercomparison of statistical and dynamical downscaling models under the euro- and med-cordex initiative framework: present climate evaluations. Clim. Dyn. 46 (3), 1301–1329.

Vautard, R., 1990. Multiple weather regimes over the north atlantic: analysis of precursors and successors. Mon. Weather Rev. 118 (10), 2056–2081.

Vautard, R., Yiou, P., 2009. Control of recent european surface climate change by atmospheric flow. Geophys. Res. Lett. 36, L22702.

Von Storch, H., Zwiers, F.W., 2001. Statistical Analysis in Climate Research. Cambridge university press.

Wood, S., 2006. Generalized Additive Models: an Introduction with R, vol. 66. CRC Press.

Wood, S.N., 2000. Modelling and smoothing parameter estimation with multiple quadratic penalties. J. R. Stat. Soc. (B) 62 (2), 413–428.

Wood, S.N., 2004. Stable and efficient multiple smoothing parameter estimation for generalized additive models. J. Am. Stat. Assoc. 99 (467), 673–686.

Wood, S.N., 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. J. R. Stat. Soc. (B) 73 (1), 3–36.

Wood, S.N., Augustin, N.H., Nov 2002. Gams with integrated model selection using penalized regression splines and applications to environmental modelling. Ecol. Model. 157 (2–3), 157–177.

Yiou, P., 2014. AnaWEGE: a weather generator based on analogues of atmospheric circulation. Geosci. Model Dev. 7 (2), 531–543.

Yiou, P., et al., 2008. Weather regime dependence of extreme value statistics for summer temperature and precipitation. Nonlinear Proc. Geophys. 15 (3), 365–378.

Yiou, P., Nogaj, M., 2004. Extreme climatic events and weather regimes over the north atlantic: when and where? Geophys. Res. Lett. 31 (7).

Yiou, P., Salameh, T., Drobinski, P., Menut, L., Vautard, R., Vrac, M., 2013. Ensemble reconstruction of the atmospheric column from surface pressure using analogues. Clim. Dyn. 41, 1333–1344.

Zhou, S., Jones, C.B., 2005. In: Developments in Spatial Data Handling: 11th International Symposium on Spatial Data Handling, Chapter Shape-aware Line Generalisation with Weighted Effective Area. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 369–380.

Zorita, E., von Storch, H., 1999. The analog method as a simple statistical downscaling technique: comparison with more complicated methods. J. Clim. 12 (8), 2474–2489.