



**HAL**  
open science

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

Nils Hempelmann, Carsten Ehbrecht, M Carmen Alvarez-Castro, Patrick Brockmann, Wolfgang Falk, Jörg Hoffmann, Stephan Kindermann, Ben Koziol, Cathy Nangini, Sabine Radanovics, et al.

### ► To cite this version:

Nils Hempelmann, Carsten Ehbrecht, M Carmen Alvarez-Castro, Patrick Brockmann, Wolfgang Falk, et al.. Web processing service for climate impact and extreme weather event analyses. Flyingpigeon (Version 1.0). 2016. hal-01375615v1

**HAL Id: hal-01375615**

**<https://hal.science/hal-01375615v1>**

Preprint submitted on 4 Oct 2016 (v1), last revised 21 May 2017 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

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

<sup>a</sup>*Le Laboratoire des Sciences du Climat et de l'Environnement*

<sup>b</sup>*German Climate Computing Center*

<sup>c</sup>*Bayerische Landesanstalt für Wald und Forstwirtschaft*

<sup>d</sup>*Julius Kühn-Institut - Federal Research Centre for Cultivated Plants*

<sup>e</sup>*NOAA Environmental Software Infrastructure and Interoperability Group/University of Colorado-Boulder*

---

## 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) [23], to be used by climatologists and impact modelers as a tool for analyzing large datasets remotely.

---

\*Corresponding author

*Email addresses:* [info@nilshempelmann.de](mailto:info@nilshempelmann.de) (Nils Hempelmann),  
[ehbrecht@dkrz.de](mailto:ehbrecht@dkrz.de) (Carsten Ehbrecht)

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.

*Keywords:* Web Processing Service, climate impact, extreme weather events, birdhouse, OGC

---

1     **Highlights:**

- 2     • A tool for remote processing of climate model data is presented  
3     • Reduces the difficulties of climate model data analysis  
4     • Standardized scientific methods for climate impact and extreme weather  
5     events  
6     • Able to communicate with other web services

7     **1. Introduction**

8     Processing of climate data is typically carried out by individual researchers,  
9     who create and run their own scripts in their preferred programming lan-  
10    guage, either locally or in environments internal to their institutions. Thus,  
11    there is a vast but unconnected body of knowledge that is not readily avail-  
12    able to the climate science community with the risk of being continually  
13    replicated as researchers write scripts for processes that have already been  
14    well-developed by others. Furthermore, climate data, such as the upcoming  
15    Phase 6 of the Coupled Model Intercomparison Project [CMIP6](#) and the Co-  
16    ordinated Regional Climate Downscaling Experiment [CORDEX](#), is becoming  
17    too large to download and process locally.

18

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

25  
26 Flyingpigeon (henceforth always referring to Version 1.0) is part of the  
27 open source project [birdhouse](#) (under the Apache License 2.0), a collection of  
28 Open Geospatial Consortium (OGC) WPSs that provides data processing for  
29 the climate science community. Like all other compartments in birdhouse,  
30 flyingpigeon communicates over the web using the HTTP protocol based  
31 WPS Interface Standard for geospatial processing services defined by OGC  
32 [23]. A make file that handles appropriate software dependencies makes it  
33 easy to install. Flyingpigeon and the processes we developed are freely avail-  
34 able from the [flyingpigeon repository](#) on GitHub. To run efficiently, the code  
35 should be installed on a system with appropriate resources.

36 Birdhouse evolved out of data management projects (C3Grid-INAD [19],  
37 ExArch [3]), LSDMA [17]), while the processes designed for flyingpigeon were  
38 developed within projects related to climate impact and extreme weather  
39 events analysis ([EUCLEIA](#), [A2C2](#), [Extremoscope](#)).

40 Inter-WPS communication enables operative services for international collab-  
41 orations such as the Infrastructure for the European Network of Earth System  
42 Modeling ([IS-ENES](#)) and Earth System Grid Federation ([ESGF](#)), and is in  
43 line with other WPS developments like [52° North](#) [2] (enabling standardized  
44 deployment of geo-processes on the web), the [ZOO-Project](#) [10] (able to pro-  
45 cess geospatial or non geospatial data online), [climate4impact](#), and Climate  
46 Information Portal of Copernicus ([CLIPC](#) [7]).

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

## 49 2. WPS general description

50 A WPS is a technical solution ([WPS Concepts](#)) in which processes are  
51 hosted on a server and accessed over the web (Fig. 1). These processes con-  
52 form to a standardized format, ensuring that they follow the principle of  
53 reusable design: they can be instantiated multiple times for different input  
54 arguments or data sources, customized following the same structure to handle

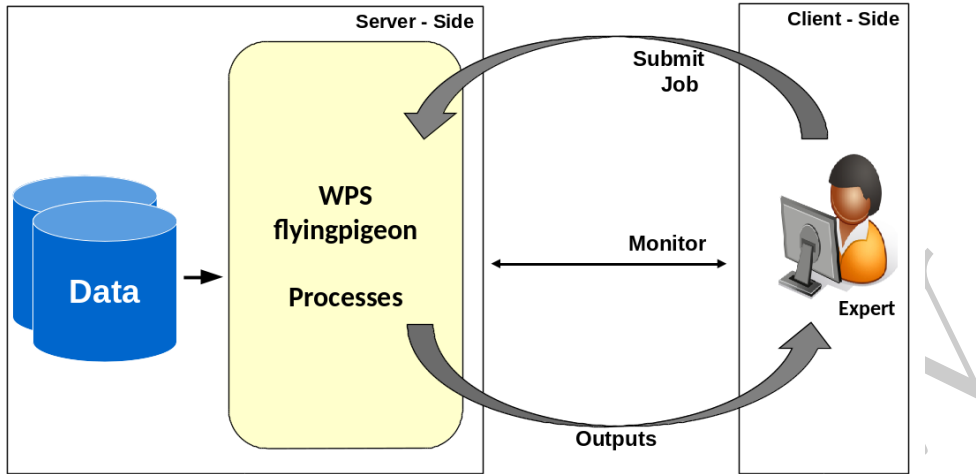


Figure 1: Schematic of WPS operations.

55 new inputs, and are modular, hence can be combined to form new processes.  
 56 In addition, a WPS can be installed close to the data to enable processing  
 57 directly out of the archive. A WPS can also be linked to a theoretically limit-  
 58 less combination of several other WPSs, or generally OpenGIS Web Services  
 59 (OWS).

60 In this paper **process** is used in the same sense as in the OGC standard: *'for*  
 61 *any algorithm, calculation or model that either generates new data or trans-*  
 62 *forms some input data into output data'* [23]. A submitted process is a **job**.  
 63 A **service** provides a collection of processes containing scientific **methods**  
 64 that focus on climate impact and extreme weather events. A combination  
 65 of processes is called a **workflow**, and a collection of WPS-related software  
 66 compartments is a **framework** (see Section 2).

67 WPS divides the operation into **server** and **client** side 2.2, with appropriate  
 68 **security** 2.3 in between to avoid misuse.

69

### 70 2.1. WPS flyingpigeon

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

74

75 One of the main software components of flyingpigeon is [OpenClimateGIS](#)  
76 (OCGIS), an open source Python package designed for geospatial manipu-  
77 lation, subsetting, computation, and translation of climate datasets stored  
78 locally in NetCDF files or served via OPeNDAP protocols. OCGIS inter-  
79 prets the climate data community’s canonical metadata standard – the [Cli-](#)  
80 [mate and Forecast \(CF\) Convention](#) – maintaining standards compliance for  
81 any derived CF-NetCDF output. OCGIS also supports numerous time-aware  
82 computations (i.e. monthly mean, seasonal maximum), multiple output for-  
83 mats (i.e. ESRI Shapefile, Comma Separated Value), and spatial interpola-  
84 tion using the [Earth System Modeling Framework](#) (ESMF). OCGIS calls can  
85 be executed on the whole dataset at once or divided into chunks to reduce  
86 the memory load. In flyingpigeon, memory availability is checked before and  
87 the call is executed accordingly.  
88 Flyingpigeon is mainly written in Python but also includes processes based  
89 on [CDO](#) commands, R scripts, and Fortran.  
90 After submitting a job to the WPS, several data preprocessing steps are run  
91 automatically. Flyingpigeon has utilities to e.g. sort files belonging to differ-  
92 ent datasets, checks the variable contained in the data files, converts units,  
93 rotates or unrotates grid coordinates, or generates file names based on the  
94 metadata or its values.  
95 An installation system simplifies the deployment of flyingpigeon on a server.  
96 It is designed for Linux distributions and fetches required dependencies with  
97 the [Conda](#) package management system. It uses [Buildout](#) to setup the ap-  
98 plication by calling a simple install command.  
99 Since the code is open source, contributions by the community are possi-  
100 ble. [Online documentation](#) is generated with Sphinx [1], including [automatic](#)  
101 [documentation](#) for the functions.

## 102 2.2. Server - Client side

103 Flyingpigeon is a server-side service embedded in the birdhouse (see Fig. 2),  
104 a framework to support the development and deployment of climate data  
105 processing services based on the OGC WPS standard. Flyingpigeon can be  
106 run in combination with other birdhouse compartments such as **phoenix**,  
107 a graphical user interface (GUI) for common web browsers that includes an  
108 OGC web mapping service (WMS), or **malleefowl**, the WPS server backend  
109 for data search in and fetching from the ESGF archive. Other compart-  
110 ments offer processes such as quality checks for technical aspects of NetCDF  
111 files ([climate forecast](#) compliance checker) or processes to execute climate

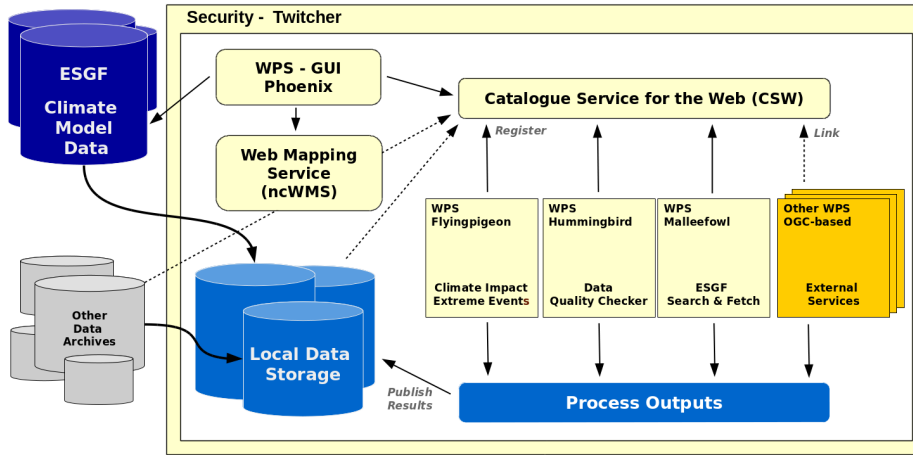


Figure 2: Main components of the birdhouse framework and connection to data archives.

112 data operations with CDO. A catalog service for the web ([CSW](#)) is used to  
 113 publish results and discover available services. However, flyingpigeon is a  
 114 self-contained service and can also be run operationally as a stand-alone.

115

116 On the client side, the user needs to connect to the server with an ap-  
 117 propriate URL to enable communication via an HTTP protocol, which is  
 118 possible in three ways:

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

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

126

### 127 2.3. Security

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

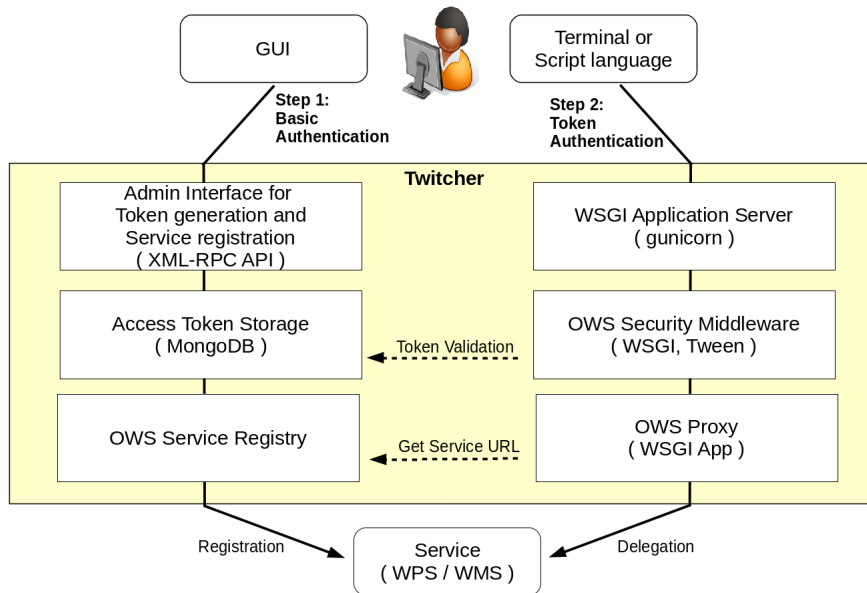


Figure 3: Twitcher security schema.

133 access to the WPS itself.

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

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

148

149 The management of the access tokens and the registered WPS is controlled by the Twitcher administration interface using the [XML-RPC](#) proto-  
 150



151 col (first layer in Fig. 3, left column). Access tokens are generated, validated  
152 and persisted by the Access Token Storage (second layer) which uses a [MongoDB](#)  
153 database. The OWS Service Registry (third layer) stores the registered  
154 WPS services with the service URL and a unique service name.

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

#### 163 2.4. Input Data

164 Climate model data are commonly stored in NetCDF file format and or-  
165 ganized in data archives. If data are not already stored on the WPS server,  
166 they have to be fetched and stored in a structured way. Besides many other  
167 datasets, the output of the CMIP5 (and upcoming CMIP6) and CORDEX  
168 are stored for public access in the ESGF data archive. Birdhouse provides a  
169 data search for ESGF as a search interface within the GUI or from the com-  
170 mand line. A second important group of datasets are reanalyses data, which  
171 are outputs from a data assimilation component of a weather forecast model.  
172 In climate impact and extreme event assessment, the use of reanalyses data  
173 is very common, thus some processes (see Section 3) in flyingpigeon provide a  
174 preselection of reanalyses datasets and variables. By selecting an appropriate  
175 variable, the data are automatically fetched within the process.  
176 In the processes for analogue circulation (3.4.2) and weather regimes (3.4.1),  
177 the preselection of variables are dataset subsets of the following global re-  
178 analysis projects:

- 179 • *NCEP/NCAR Reanalysis 1:*  
180 NCEP data [18] are available from 1948 to the present in 17 pressure  
181 levels with a spatial grid resolution of  $2.5^\circ \times 2.5^\circ$
- 182 • *20th Century Reanalysis version 2 (20CR):*  
183 The 20CR dataset [5] is based on surface pressure observations only and  
184 provides an ensemble of 56 members that contains global weather con-  
185 ditions and their uncertainty from 1871–2012, available on 24 pressure  
186 levels with a spatial grid resolution of  $2^\circ \times 2^\circ$ .

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

189 In climate impact and extreme event assessments, some analytic methods  
190 require non-climatic data. The species distribution model (SDM) process  
191 (3.3.2) in flyingpigeon to predict favorability of tree species requires geograph-  
192 ical coordinates of tree occurrences. A free and open access data base for  
193 biodiversity data is provided by the Global Biodiversity Information Facil-  
194 ity (GBIF), where observations of trees (and other species) are stored and  
195 automatically fetched by the SDM process.

### 196 3. Processes embeded in flyingpigeon

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

200

#### 201 3.1. Extracting spatial subsets

202 A standard operation is spatial subsetting using bounding boxes and  
203 arbitrarily-sided polygons. Spatial subset operations use the OCGIS Python  
204 package. Following a subset, data masks are updated on subsetted variables  
205 reflecting the coordinates' spatial relationship with the subset boundary. In  
206 flyingpigeon, the polygon subset process is possible at three levels of increas-  
207 ing spatial resolution: continents, world countries and European administra-  
208 tive regions obtained from the global administrative areas GADM database  
209 (*gadm26\_levels.gdb, v2.5*). To optimize subsetting performance, high reso-  
210 lution GADM boundaries were simplified [42] using *mapshaper* [12] with a  
211 1% point retention. The coarse resolution of target climate model inputs  
212 compared with the down-sampled resolution of GADM boundaries did not  
213 result in any significant data loss following a spatial subset operation. Very  
214 small administrative regions were merged on the country level.

215 An additional mosaic option merges multiple polygons; otherwise, a separate  
216 output file is created for each polygon. This increases the flexibility to meet  
217 specific user needs. The outputs are provided as NetCDF files wrapped in a  
218 tar archive.

219 In addition to polygons, a process is also available to extract longitude and  
220 latitude points. Point subset data products are returned in text files con-

221 taining a timeseries table for each longitude and latitude point.

222

### 223 3.2. Computation of climate indices

224 Climate indices are metrics to describe climate conditions and can be  
225 used to assess changes in climate over time (see the use case below), or as  
226 forcing data for complex impact models (e.g. the species distribution model  
227 3.3.2). Climate indices can be user-defined, but there is also a standardized  
228 set provided by the [European Climate Assessment & Dataset](#) website.

229

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

234

- 235 1. standard
- 236 2. percentile-based
- 237 3. multi-variable indices
- 238 4. user-defined

239 Standard indices are based on one daily input variable (near surface tem-  
240 perature, precipitation). The provided indices are related to the [ICCLIM](#)  
241 python package deployed in OCGIS. In the percentile-based process, a per-  
242 centile is calculated over a user defined reference period. The process outputs  
243 the number of days that exceed or fall below this calculated threshold (for  
244 percentiles  $> 50$  or  $< 50$ , respectively). When deployed, the multi-variable  
245 index process will allow users to specify a combination of indices. User-  
246 defined indices, also under development, will allow custom index calculations  
247 for single or multiple variables.

248

249 *Use case: Extremoscope.* We built an [interactive data visualization](#) to show  
250 the evolution of 10 extreme climate indices in the 13 regions of France under  
251 two climate scenarios (RCP 4.5 and 8.5) as part of the [Extremoscope](#) project.  
252 Flyingpigeon was used to calculate the climate indices using 7 climate pro-  
253 jection models for five time aggregates (four seasons and yearly).

254 The visualization displays the probability of extreme seasons or years oc-  
255 ccurring, defined as exceeding the 90th/95th percentile, or falling below the

256 5th/10th percentile, based on a reference period between 1972–2100. The  
257 data is filterable by each dimension (region, time aggregate, index, scenario,  
258 percentile, model and year) with different charts that are linked (crossfiltered)  
259 using the [crossfiltering library](#) in [dc.js](#), the Dimensional Charting JavaScript  
260 library, to allow the effects of each dimension to be observed and compared.  
261

### 262 3.3. *Climate impact related processes*

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

#### 269 3.3.1. *Segetal flora*

270 'Segetal flora' is a term used for weeds growing in crop fields. The major-  
271 ity of segetal flora species have very positive and important ecological effects  
272 as a source of pollen or nectar for several insects or as nutrition for birds [16].  
273 Based on field monitoring, a relation between the number of different species  
274 occurring in crop fields and the annual mean temperature can be shown.  
275 The relations are investigated separately for seven segetal flora groups (e.g.  
276 Mediterranean or Nordic groups) and three land use types (conventional and  
277 ecological land use, and one–two year self-greened fallow land). The species  
278 number to mean annual temperature relations are expressed as regression  
279 functions [15, 14], which are the core of the segetal flora process. With tem-  
280 perature at 2m height (tas) as input data, and arguments to select segetal  
281 flora type and land use, the number of species is predicted. A polygon subset  
282 for a specified country can be selected with an optional argument.  
283 The process returns a tar file containing NetCDF files with the appropriate  
284 number of segetal flora species per grid point for each input dataset. The  
285 calculation is only relevant for areas with agricultural land use.  
286

#### 287 3.3.2. *Species distribution model*

288 Species distribution models (SDMs) are numerical tools to describe the  
289 relation between the distribution of a species and the environmental con-  
290 ditions that are thought to lead to this distribution [11, 24]. Models are

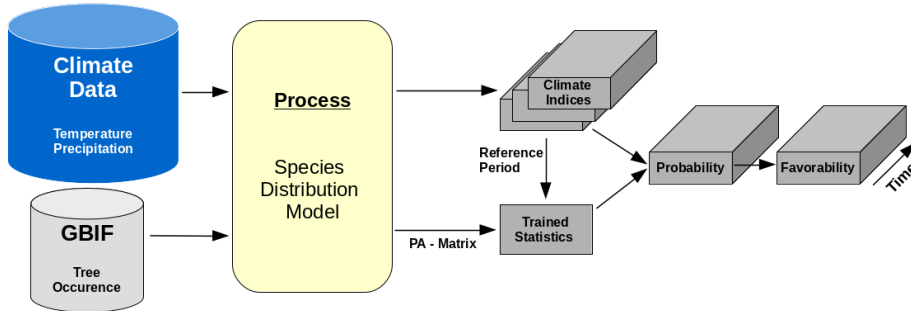


Figure 4: Schema of the Species Distribution Model process.

291 used to gain ecological insights and predict distributions across landscapes,  
 292 sometimes requiring extrapolation in space and time, e.g. for climate change  
 293 impact modeling (modified from [8]).

294 We implemented a SDM process that relates the distribution of tree species to  
 295 climate conditions to analyze the impacts of climate change [9]. The method  
 296 focuses on tree species where the distribution is independent of non-climate  
 297 factors like soil or specific site conditions. We hypothesize that the climate  
 298 conditions which limit the tree species distribution (e.g. heat, precipitation,  
 299 winter temperatures, drought [21]) can be described with climate indices (de-  
 300 fined in section 3.2).

301 Fig. 4 shows the schematic workflow of the SDM process. Climate indices are  
 302 calculated based on climate model variables (daily temperature and precipi-  
 303 tation values). Tree species data can be fetched from the global biodiversity  
 304 information facility (GBIF) database and fed to the SDM process. The  
 305 geographical coordinates of the tree occurrence are then translated into a  
 306 presence-absence (PA) matrix with the grid size of the input climate model  
 307 data. Ocean and large lake areas are excluded in the PA matrix.

308 Statistical training [36, 35] is performed based on the calculated climate  
 309 indices as a mean over a reference period and the PA matrix using general-  
 310 ized additive models (GAMs)[33, 37]. GAMs are flexible regression models  
 311 with non-parametric (smooth) additive components [34].

312 After the statistical training, the probability of occurrence is predicted over  
 313 the entire time series and transformed into a favorability [25] to compare  
 314 different models. The SDM process can used with different climate indices,  
 315 tree species, training performance values, and climate model datasets.

316

317 *3.4. Extreme weather events related processes*

318 Besides climate impact studies, the WPS flyingigeon also focuses on ex-  
319 treme weather events investigations. This section contains examples of more  
320 complex processes to calculate weather regimes (3.4.1) or analogues of circu-  
321 lation (3.4.2).

322 *3.4.1. Weather regimes*

323 Weather regimes are recurring states of the atmosphere and provide a use-  
324 ful description of atmospheric variability [22, 6, 30]. In extreme event studies,  
325 for instance, it is possible to associate some specific seasonal weather regimes  
326 with the extremes of surface variables (precipitation or temperature) by an-  
327 alyzing which regime is prevailing when an extreme is encountered [40].

328 Following the methods of [22] and [39], the weather regime process computes  
329 a given number of regimes (the  $\kappa$  value, by default  $\kappa = 4$  [22]) for a region  
330 (by default the North Atlantic region [80°W - 50°E , 20°N - 70°N]) using  
331 values of sea level pressure (SLP) or geopotential height ( $Z_g(h)$ ) anomalies  
332 of a given season.

333 The weather regimes are computed with a k-mean classification algorithm  
334 [13] on the first 10 principal components (PCs) of the SLP or geopotential  
335 height ( $Z_g(h)$ ) fields. The PCs are computed from a Empirical Orthogonal  
336 Function decomposition [32] of the field. The data are weighted by the square  
337 root of the cosine of the latitude to account for grid cell surface variations.  
338 The obtained PCs are then classified onto  $\kappa$  clusters. The weather regimes  
339 can be computed on a reference data set (e.g. reanalyses data). Then, other  
340 datasets (e.g. climate model data) can be classified according to these refer-  
341 ence weather regimes.

342 Fig. 5 shows the output for a use case where weather regimes in winter

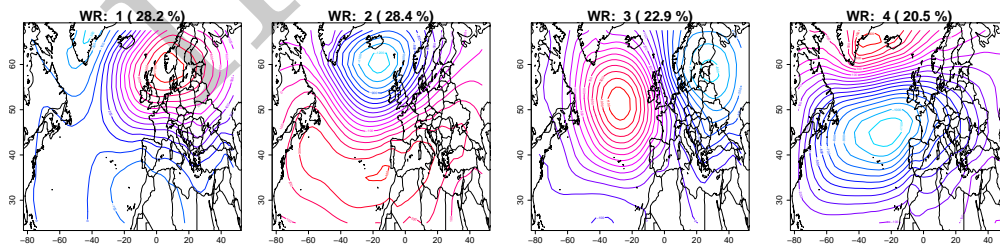


Figure 5: Weather regimes (WR) for winter months (DJF) based on NCEP sea surface pressure data (1970–2010).

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
...				

343 months were calculated based on NCEP sea surface pressure data over a re-  
 344 gion in the North Atlantic for  $\kappa = 4$ . The process also output an R workspace  
 345 containing the statistical training values as well as a text file with the PCs  
 346 and the NetCDF file with the normalized pressure values. These outputs can  
 347 be used for manual postprocessing or for the weather regime projection pro-  
 348 cess which calculates the percentage of each given weather regime in every  
 349 year. Such a projection was done in this example for a CMIP5 dataset and  
 350 Table 1 shows the calculated values.

### 351 3.4.2. Analogues of circulation

352 Analogues of circulation provide a versatile tool to investigate the rela-  
 353 tion between climate variables (such as temperature or precipitation) and  
 354 large-scale atmospheric circulation patterns (SLP or  $Z_g(\text{h})$ ).  
 355 For an SLP/ $Z_g(\text{h})$  pattern on a given day, the idea is to select days that have  
 356 a calendar proximity (i.e. within a time window around the given date in  
 357 all years except the year of the given day) and that minimize a distance be-  
 358 tween the circulation patterns. This approach has been used to infer climate  
 359 reconstructions from SLP fields [26, 41], in weather forecast predictability as-  
 360 sessment [20], downscaling of climate variables [43, 29], detection/attribution  
 361 studies [31, 27, 4] and stochastic weather generators [38].

362

363 *Analogues detection process.* The analogues from this flyingpigeon process  
 364 are mainly used to study the conditional attribution of extreme events to dy-  
 365 namics, thus describing only the thermodynamic changes in weather events.

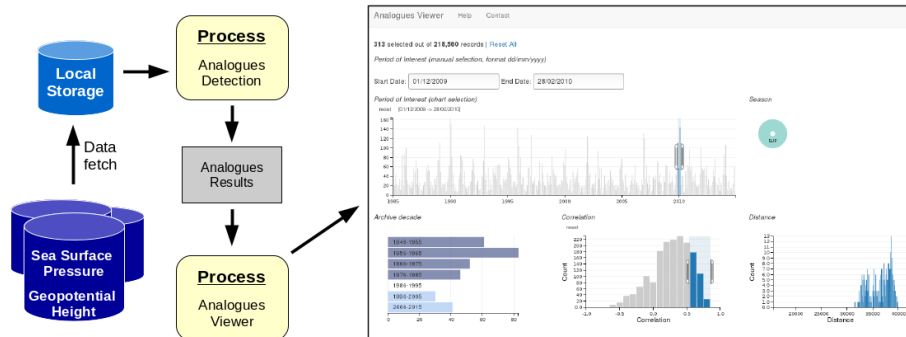


Figure 6: Schema of the analogues workflow.

366 For a given continuous extreme event period (e.g. an extreme monthly tem-  
 367 perature), the process seeks in past archives flows similar to those in the  
 368 continuous period and reconstructs a monthly temperature from these past  
 369 flow analogue days. The resulting temperature is then compared to the cur-  
 370 rent one, providing an estimate of thermodynamic changes.

371 The process uses a circulation analogue simulation FORTRAN code ([CASTf90](#))  
 372 to compute analogues of circulation. The user can choose a time period for  
 373 which daily analogues are computed and an archive period, that is, the time  
 374 period to resample from. The user can also select a distance (Euclidean,  
 375 Mahalanobis, correlation, Teweles-Wobus S1 [28], how many analogues per  
 376 day to retain, a rectangular region (in longitude–latitude), whether to work  
 377 with anomalies or not and some more options like the output file format.  
 378 The result is a list of analogue dates for each simulation day and the corre-  
 379 sponding distance values.

380

381 *Analogues viewer process.* To explore the properties of the analogues detec-  
 382 tion process, we developed an analogues viewer process that produces an  
 383 interactive data visualization output as an HTML page hosted on the server-  
 384 side. This viewer uses [dc.js](#), the Dimensional Charting JavaScript library  
 385 and a [crossfiltering library](#) to produce interactive charts that can be filtered  
 386 based on data parameters. For each filter selection, all charts are simultane-  
 387 ously updated (crossfiltered).

388 Figure 6 shows a schematic of the analogues workflow with a screenshot of  
 389 the visualisation. For this example, the winter of 2010 was selected in the  
 390 bar plot, which filtered all charts to display counts of only those analogues



391 pertaining to this time period. The display was further filtered by selecting  
392 analogues with correlation coefficients  $> 0.5$ . The decadal distribution of  
393 analogues and changes across time can be immediately seen by sliding the  
394 time window.

395 The analogues viewer demonstrates two powerful advantages of using WPS  
396 for data analysis: 1) it is straight-forward to 'chain' processes together so  
397 that the output of one process can be used as the input of another, allowing  
398 data analysis to be re-imagined as a workflow, and 2) processes can be com-  
399 pletely custom-designed. Here, we integrate data visualization capability –  
400 normally a step performed in an environment completely separate from the  
401 analysis and requiring a specialized skill that is often not within the scope  
402 of the scientific researcher – directly into flyingpigeon, accessible simply by  
403 launching the output HTML link.

404

#### 405 **4. Conclusion**

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

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

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

421

#### 422 **5. Outlook**

423 The increasing number of developers and users ensures further improv-  
424 ment of features, documentation, guidelines and tutorials. Currently, high-  
425 performance computing providers located close to the data archives are being

426 established in the climate community to keep up with the demand for WPS.  
427

## 428 Acknowledgments

429 NCEP Reanalysis data was obtained from NOAA/OAR/ESRL PSD,  
430 Boulder, Colorado, USA, from their website <http://www.esrl.noaa.gov/psd/>.  
431  
432 20th Century Reanalysis V2 data was obtained from NOAA/ OAR/ESRL  
433 PSD, Boulder, Colorado, USA, from their website <http://www.esrl.noaa.gov/psd/>.  
434  
435 CAC, SR and PY are supported by the ERC Advanced Grant No. 338965-  
436 A2C2.

## 437 References

- 438 [1] G. Brandl and Sphinx team. Sphinx. python documentation generator,  
439 2016.
- 440 [2] J. Brauner. *Formalizations for geoperators-geoprocessing in Spatial  
441 Data Infrastructures*. Dissertation, Technische Universität Dresden,  
442 Fakultät Umweltwissenschaften, 10 2015.
- 443 [3] A. M. Castronova, J. L. Goodall, and M. M. Elag. Models as web services  
444 using the open geospatial consortium (ogc) web processing service (wps)  
445 standard. *Environmental Modelling & Software*, 41(0):72–83, 2013.
- 446 [4] J. Cattiaux, R. Vautard, C. Cassou, P. Yiou, V. Masson-Delmotte, and  
447 F. Codron. Winter 2010 in europe: A cold extreme in a warming climate.  
448 *Geophysical Research Letters*, 37(20):L20704, 2010.
- 449 [5] G. P. Compo et al. The twentieth century reanalysis project. *Q. J. Roy.  
450 Meteor. Soc.*, 137(654):1–28, 2011.
- 451 [6] S. Corti, F. Molteni, and T. N. Palmer. Signature of recent climate  
452 change in frequencies of natural atmospheric circulation regimes. *Nature*,  
453 398(6730):799–802, 1999.

- 454 [7] C. Déandreis, C. Pagé, P. Braconnot, L. Barring, E. Bucchignani, W. S.  
455 de Cerff, R. Hutjes, S. Joussaume, C. Mares, S. Planton, and M. Plieger.  
456 Towards a dedicated impact portal to bridge the gap between the impact  
457 and climate communities : Lessons from use cases. *Climatic Change*,  
458 125(3):333–347, 2014.
- 459 [8] J. Elith and J. R. Leathwick. Species distribution models: Ecological  
460 explanation and prediction across space and time. *Annual Review of*  
461 *Ecology, Evolution, and Systematics*, 40(1):677–697, 2009.
- 462 [9] W. Falk and N. Hempelmann. Species favourability shift in Europe due  
463 to climate change: A case study for *Fagus sylvatica* and *Picea abies* (L.)  
464 karst. based on an ensemble of climate models. *Journal of Climatology*,  
465 2013:1–18, 2013.
- 466 [10] G. Fenoy, N. Bozon, and V. Raghavan. Zoo-project: the open wps  
467 platform. *Applied Geomatics*, 5(1):19–24, 2013.
- 468 [11] J. Franklin. *Mapping species distributions: spatial inference and predic-*  
469 *tion*. Cambridge University Press, 2010.
- 470 [12] M. Harrower and M. Bloch. Mapshaper.org: A map generalization web  
471 service. *IEEE Comput. Graph. Appl.*, 26(4):22–27, July 2006.
- 472 [13] J. A. Hartigan and M. A. Wong. Algorithm as 136: A k-means clustering  
473 algorithm. *Journal of the Royal Statistical Society. Series C (Applied*  
474 *Statistics)*, 28(1):100–108, 1979.
- 475 [14] N. Hempelmann and C. Ehbrecht. Web processing services for climate  
476 data - with examples for impact modelers. In *EGI Community Forum*  
477 *2014*, 2014.
- 478 [15] J. Hoffmann, N. Hempelmann, M. Glemnitz, L. Radics, G. Czimber, and  
479 U. Wittchen. Einfluss von temperatur und nutzung auf die floristische  
480 artenvielfalt in getreideanbaugebieten europas. 436:70-76. *Julius-Kühn-*  
481 *Archiv*, 2012.
- 482 [16] J. Hoffmann, U. Wittchen, N. Hempelmann, M. Glemnitz, and  
483 L. Radics. Wildpflanzen auf dem acker – möglichkeiten und grenzen für  
484 den artenschutz im Ökolandbau. *Forschungs Report Spezial Ökologischer*  
485 *Landbau*, page 20.21, 2013.

- 486 [17] C. Jung, M. Gasthuber, A. Giesler, M. Hardt, J. Meyer, F. Rigoll,  
487 K. Schwarz, R. Stotzka, and A. Streit. Optimization of data life cy-  
488 cles. *Journal of Physics: Conference Series*, 513(3):032047, 2014.
- 489 [18] E. Kalnay et al. The NCEP/NCAR 40-year reanalysis project. *Bulletin*  
490 *of the American Meteorological Society*, 77(3):437–471, 1996.
- 491 [19] S. Kindermann, F. Schintke, and B. Fritsch. A collaborative data man-  
492 agement infrastructure for climate data analysis. *Geophysical Research*  
493 *Abstracts*, 14(EGU201):10569, April 2012.
- 494 [20] E. N. Lorenz. Atmospheric predictability as revealed by naturally oc-  
495 ccurring analogues. *Journal of the Atmospheric Sciences*, 26(4):636–646,  
496 1969.
- 497 [21] K. H. Mellert, V. Deffner, H. Küchenhoff, and C. Kölling. Modeling  
498 sensitivity to climate change and estimating the uncertainty of its im-  
499 pact: A probabilistic concept for risk assessment in forestry. *Ecological*  
500 *Modelling*, 316:211–216, Nov 2015.
- 501 [22] P.-A. Michelangeli, R. Vautard, and B. Legras. Weather regimes: Re-  
502 currence and quasi stationarity. *J. Atmos. Sci.*, 52(8):1237–1256, 1995.
- 503 [23] M. Mueller and B. Pross. OGC® WPS 2.0 interface standard corrigendum 1, 03 2015.  
504
- 505 [24] A. T. Peterson, J. Soberón, R. G. Pearson, R. P. Anderson, E. Martínez-  
506 Meyer, M. Nakamura, and M. B. Araújo. *Ecological Niches and Geo-*  
507 *graphic Distributions (MPB-49) (Monographs in Population Biology)*.  
508 Princeton University Press, 2011.
- 509 [25] R. Real, A. M. Barbosa, and J. M. Vargas. Obtaining environmen-  
510 tal favourability functions from logistic regression. *Environ Ecol Stat*,  
511 13(2):237–245, Jun 2006.
- 512 [26] F. Schenk and E. Zorita. Reconstruction of high resolution atmospheric  
513 fields for northern europe using analog-upscaling. *Climate of the Past*,  
514 8(5):1681–1703, 2012.
- 515 [27] P. A. Stott, N. Christidis, F. E. L. Otto, Y. Sun, J.-P. Vanderlinden,  
516 G. J. van Oldenborgh, R. Vautard, H. von Storch, P. Walton, P. Yiou,

- 517 and F. W. Zwiers. Attribution of extreme weather and climate-related  
518 events. *WIREs Climate Change*, 7:23–41, 2016.
- 519 [28] S. Teweles and H. B. Wobus. Verification of prognostic charts. *Bulletin*  
520 *of the American Meteorological Society*, 35(10):455–463, 1954.
- 521 [29] P. Vaittinada Ayar, M. Vrac, S. Bastin, J. Carreau, M. Déqué, and  
522 C. Gallardo. Intercomparison of statistical and dynamical downscaling  
523 models under the euro- and med-cordex initiative framework: present  
524 climate evaluations. *Climate Dynamics*, 46(3):1301–1329, 2016.
- 525 [30] R. Vautard. Multiple weather regimes over the north atlantic: Analysis  
526 of precursors and successors. *Monthly weather review*, 118(10):2056–  
527 2081, 1990.
- 528 [31] R. Vautard and P. Yiou. Control of recent european surface climate  
529 change by atmospheric flow. *Geophysical Research Letters*, 36:L22702,  
530 2009.
- 531 [32] H. Von Storch and F. W. Zwiers. *Statistical analysis in climate research*.  
532 Cambridge university press, 2001.
- 533 [33] S. Wood. *Generalized Additive Models: An Introduction With R*, vol-  
534 ume 66. CRC Press, 2006.
- 535 [34] S. N. Wood. Modelling and smoothing parameter estimation with mul-  
536 tiple quadratic penalties. *Journal of the Royal Statistical Society (B)*,  
537 62(2):413–428, 2000.
- 538 [35] S. N. Wood. Stable and efficient multiple smoothing parameter estima-  
539 tion for generalized additive models. *Journal of the American Statistical*  
540 *Association*, 99(467):673–686, 2004.
- 541 [36] S. N. Wood. Fast stable restricted maximum likelihood and marginal  
542 likelihood estimation of semiparametric generalized linear models. *Jour-*  
543 *nal of the Royal Statistical Society (B)*, 73(1):3–36, 2011.
- 544 [37] S. N. Wood and N. H. Augustin. Gams with integrated model selec-  
545 tion using penalized regression splines and applications to environmental  
546 modelling. *Ecological Modelling*, 157(2-3):157–177, Nov 2002.

- 547 [38] P. Yiou. AnaWEGE: a weather generator based on analogues of at-  
548 mospheric circulation. *Geoscientific Model Development*, 7(2):531–543,  
549 2014.
- 550 [39] P. Yiou et al. Weather regime dependence of extreme value statistics  
551 for summer temperature and precipitation. *Nonlinear Proc. Geoph.*,  
552 15(3):365–378, 2008.
- 553 [40] P. Yiou and M. Nogaj. Extreme climatic events and weather regimes  
554 over the north atlantic: When and where? *Geophysical Research Letters*,  
555 31(7), 2004.
- 556 [41] P. Yiou, T. Salameh, P. Drobinski, L. Menut, R. Vautard, and M. Vrac.  
557 Ensemble reconstruction of the atmospheric column from surface pres-  
558 sure using analogues. *Climate Dynamics*, 41:1333–1344, 2013.
- 559 [42] S. Zhou and C. B. Jones. *Developments in Spatial Data Handling: 11th International Symposium on Spatial Data Handling*, chapter Shape-Aware Line Generalisation With Weighted Effective Area, pages 369–380. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.
- 563 [43] E. Zorita and H. von Storch. The analog method as a simple statistical  
564 downscaling technique: comparison with more complicated methods.  
565 *Journal of Climate*, 12(8):2474–2489, 1999.