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
0098-3004/

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 Coperincus (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.

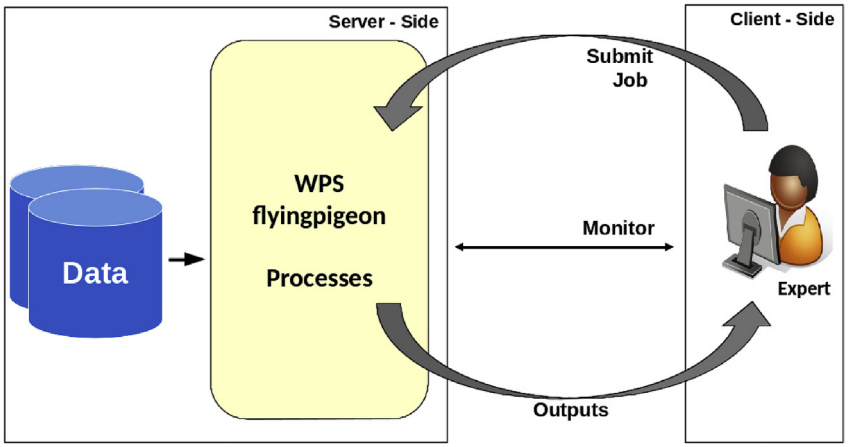

Fig. 1. Schematic of WPS operations.