Introduction

Climate change is described as long-term changes in temperatures, rainfall and other phenomena, which clearly have strong impacts on society as a whole. Climate change projection data are used in a variety of areas in applications such as forecasting crop yields, water availability, and electricity generation. In agriculture, to determine the climate change impacts is a complex task and often requires detailed analysis because of the variety of effects that can occur in the field (Knox et al., 2012).

The availability of these data and the easy access to them is crucial for those who use them in their research and applications where one of the main problems is how
to provide very large amounts of data, the main issue when it comes to data that focus on climate change projections.

There are some techniques that can provide information for climate change impact studies, such as model simulations, which are important tools for evaluating the potentials of a crop productivity. From these model outputs, it is possible to draw analysis to support the management of the growth and development of a crop, according to the climatic information for any decision making. The numerical models have contributed substantially to the development of these issues, helping in the production, in the optimization on the use of water (Fortes, Platonov e Pereira, 2005) and helping to understand the impacts of the climate changes on the agriculture crops. Many researchers have used these numerical model outputs to study and understand the climate change effects on various existing crops (Jagtap e Jones, 2002).

Several studies have demonstrated the success in the combined use of crop models and geographic information systems (GIS) (Sriharan et al., 2008). GIS is a collection of tools and geographic information, which are applied in different areas, and can be used in various software and services, helping decision making (Jia et al., 2008).

On the climate change issues, the impacts on agricultural production, and its use on crop and disease simulation models faces us with the manipulation of a large amount of data and, at this point, comes to the term “Big data”.

Big Data is a paradigm applied to a dataset whose size is greater than the capacity of commonly used tools to capture, manage, and process data within an acceptable time (Wigan e Clarke, 2013). It can be manipulated with various technologies to aid in the ability to transform the business, changing the way how the data analysis is done by obtaining useful information for decision making (Yin e Kaynak, 2015). Associated with this fact, there is an increasing demand for diversified services, contents and products, which meet even more specific segments and needs. Dealing with this flow of constant information becomes a complex task for companies and institutions, which often need to make rapid decisions based on data analysis. Therefore, new approaches or technologies have been developed to provide, manipulate, collect, analyze and visualize this data, and with the technological advance, traditional forms are losing their space and new solutions are always emerging. Among them, the Internet of Things (IoT), Web Services, global positioning system (GPS), remote sensing and Big Data analysis can be highlighted (Hey e Trefethen, 2005; Wang et al., 2005; Yuan et al., 2013; Bröring et al., 2011; Manyika et al., 2011).

For the implementation of the National Adaptation Planning to Climate Change (PNA), instituted on May 10, 2016 through Ordinance Nº 150, under the responsibility of Brazil’s Ministry of Environment (MMA), several socioeconomic sector groups have been engaged in the process of assessing the impacts, vulnerability, and adaptation to climate change. Multidisciplinary groups, using various approaches have shown difficulty in manipulating this downscaling climate change dataset, due to unfamiliar format, large size, and names of the variables for example. The Brazilian Ministry of Environment with support by the German International Cooperation Agency, GIZ (Deutsche Gesellschaft für Internationale Zusammenarbeit) promoted the development of PROJETA platform (Projections of climate change for South America downscaled by the Eta model).

The main goal of the PROJETA platform is to automate access to climate change projection data generated by the Eta Model at Center for Weather Forecasting and Climate Studies (CPTEC) of the Brazilian National Institute for Space Research (INPE). The platform provides easy use and fast responses to requests made by the user and, at the same time, eliminate the manual process now performed by INPE assistants to make this data available, reducing the waiting time of users for the requested data and releasing the assistants to do more research activities. In addition, the user can choose different formats offered by the platform (CSV, JSON, XML, Geotiff, and binary), especially georeferenced formats like GEOJSON and CSV, which are the best suited for the user’s needs. In addition, the implemented API, responsible for the communication task between the platform and the users, has made easy the import of the data directly via HTTP and, therefore, its use in legacy systems of companies, governmental entities, educational institutions and research centers.

Material and methods

The Web platform PROJETA (Figure 1), which is an acronym for “Projections of climate change for South America downscaled by the Eta model”, was built to automatically access, prepare and make available to the users the dataset of the downscaling climate change scenarios. These projections are based on four global climate downsampling models carried out by the Eta Model at CPTEC/INPE. The PROJETA project is a partnership between CPTEC/INPE and the University of Passo Fundo, promoted by the Brazilian Ministry of the Environment and funded by the German agency Deutsche Gesellschaft für Internationale Zusammenarbeit.

It is important to mention that the works related to the goals and functions of the PROJETA, such as those that address: the construction of Web environments for the visualization of climate (Alder e Hostetler, 2015) and environmental data (Jones et al., 2016), the use of Web technologies for large environmental datasets (Blower et al., 2013, Vitolo et al., 2015), computational tools to access climate
data stored in repositories maintained by governmental entities or by private companies (Fraley et al., 2011; Kemp et al., 2012; Serrano-Notivoli et al., 2017; Manubes et al., 2018; Frias et al., 2018), and a virtual laboratory to biodiversity and climate change with big data (Hallgren et al., 2016).

The platform was developed in three stages: preparation of the downscaling dataset to be uploaded to the database management system what allows the georeferenced storage and access; development of an Application Programming Interface (API); and development of the Web platform, with its functionalities and interactions with the users.

The Eta is a limited area atmospheric model in grid point mesh. It is based on finite volume and uses the vertical coordinates (Mesinger, 1984) which is appropriate in regions of complex orography. The model is operational since 1996 in CPTEC/INPE (Chou, 1996) to produce weather forecasts over South America, since 2002 to produce seasonal forecasts (Chou et al., 2005; Pilotto et al., 2012) and since 2012 to support the Second National Communication for the United Nations Framework for Climate Change Convention (MCTI, 2012; Chou et al., 2012; Marengo et al., 2012). During these years, the Eta model has been modified and upgraded (Mesinger et al., 2012; Mesinger et al., 2016).

The model is applied to increase the resolution of the driver model and to capture more extreme events that is to downscale the coarse global model forecasts. The Eta model has been applied to studies of weather phenomena (Saulo et al., 2000; Seluchi et al., 2003; Seluchi et al., 2011; Dereczynski et al., 2010), of agriculture crop productivity (Vieira Junior, 2006; Vieira Junior et al., 2009), etc. The version of the model developed for very long-term integrations (Pesquero et al., 2010; Chou et al., 2012; Marengo et al., 2012), suitable for climate change studies, has modifications with respect to the weather version, such as: the ingestion of long time-series of sea surface temperature, vegetation greenness from constant to monthly variation, calendar count for more than a hundred years, and equivalent CO$_2$ concentrations increased with time according to the future greenhouse gas emission scenarios.

The downscaling dataset available over the PROJETA platform was generated initially to support Brazil’s Third National Communication (MCTI, 2016) and the Brazil 2040 project of the Secretary for Strategic Affairs (SAE, 2015). The Eta model was setup at 20-km resolution and the domain covered South America and Central America. The run from 1961 and 2005 correspond to the reference period, or also called historical or baseline periods. These runs were evaluated by Chou et al. (2014a). The projections of the future climate change for the period between 2006 and 2100 were divided into three time-slices of thirty years: 2011-2040, 2041-2070, 2071-2100. The emission scenarios available via PROJETA considered two Representative Concentration Pathways (RCP): RCP4.5 and RCP8.5 (Van Vuuren et al., 2011a, 2011b). The first scenario is relatively optimistic while the second is the most pessimistic scenario. At the end of the 21st century, the equivalent CO$_2$ concentration reaches about 650 ppm in the RCP4.5 scenario and exceeds 1000 ppm in the scenario RCP8.5. The climate change over South America using these projections is assessed in Chou et al. (2014b).

The PROJETA platform makes available the downscaling of four global climate model runs: the HadGEM2-ES model (Collins et al., 2011) of the British Service UK Met Office, the Japanese model MIROC5 (Watanabe et al., 2010) was developed by a consortium of Japanese institutions (NIES-Frontier Research Center for Global Change, Univer-
sity of Tokyo and Japan Agency for Marine-Earth Science and Technology), the CanESM2 model (Arora et al., 2011), developed by Canadian Center for Climate Modeling and Analysis (CCCMA), and the Brazilian model BESM (Nobre et al., 2013). Table 1 shows the grid resolution for atmospheric and ocean components for these models from PROJETA.

The platform also makes available the very-high resolution downsampling dataset of theEta model. This version has included the non-hydrostatic mode to operate in very high resolutions in continuous long-term simulations (Lyra et al., 2017). These high resolution runs covers part of Southeast of Brazil and are driven by the downsampling of theEta 20-km of HadGEM2-ES under RCP4.5 and RCP8.5 scenarios. This version, although more computationally demanding, is suitable for studying the climate related issues of small scale, such as climate change in metropolitan areas, hydrology of small basin or within complex topography basin, smaller patches of land cover, as examples.

The storage of climate change datasets was performed in the PostgreSQL (https://www.postgresql.org/) Database Management System (DBMS). The main characteristics that led to the choice of this DBMS were the ease of performing transactions, programming with triggers, security of access to the system, need to work with relational objects, development of functions and data integrity. PostGIS (https://postgis.net/), spatial and geographic objects for PostgreSQL, was also used for having many useful functions in queries, as well as storage and access to georeferenced data, essential features for the platform’s functionality.

The binary files in TIFF format provided by CPTEC/INPE with the climate change data sets were read and stored in PostgreSQL in a table format. Figure 2 is the class diagram that shows the structure and relations between the classes of thePROJETA database.

The variables available to download are listed in the “Variable” field on the main page of the PROJETA platform. This list was constructed after a survey sent to users and potential users.

In this first stage of the platform, the database has a size of about 35 Terabytes. With the inclusion of more hourly frequency data and other climate variables, the database size will be about 80 Terabytes.

After the database was loaded with projections of climate change, the next step was to build the web structure of the platform. This structure has three parts: the back-end that is responsible for the entire process of manipulation and data processing; the front-end that covers the entire interface of the platform and the access to API, and that was implemented using resources of HTML5, CSS and JavaScript with the help of AngularJS (https://angularjs.org/); and finally, the API that is used for the data communication throughout the platform. Conversions of the binary files with the climatic data were made, in the initial phase of the database, without the help of extra tools.

This data is made available to the user through links provided by the API. All usage tests were analyzed and designed according to the needs of the platform, following standard and specific parameters, so that the service could perform the queries correctly, providing the data according to the requests requested by the users and according to the technical guidelines the CPTEC researchers.

Figure 3 shows the flow for these queries, where the user accesses the platform through a Web browser, via Hypertext Transfer Protocol Secure (HTTPS), where an Angular application accesses the API that, in turn, accesses the NodeJS server layer (https://nodejs.org/en/). This server layer accesses the platform database management system, where GET, POST, PUT and DELETE requests are performed. The R package presented in this figure was only used early in the development of the platform to convert the projection data into binary format into the TIFF format. Currently, the binary data extracted from the projections are stored directly in the database in TIFF format, without the need for this conversion done by the script in R.

Table 1: Global climate models with downscaling by the Eta model available over PROJETA. In case of the atmospheric grid and its latitude, the tabulated resolution is only valid for the tropics. For higher latitudes, deviations may occur. In case of the ocean, if two values are given for the latitude resolution, the resolution is not constant. The first value is that for the tropics and the second for the poles (maximum for the two poles if different).

<table>
<thead>
<tr>
<th>Model</th>
<th>Atmospheric grid</th>
<th>Oceanic grid</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Latitude</td>
<td>Longitude</td>
</tr>
<tr>
<td>HadGEM2-ES</td>
<td>1.25°</td>
<td>1.875°</td>
</tr>
<tr>
<td>MIROC5</td>
<td>1.4008°</td>
<td>1.40625°</td>
</tr>
<tr>
<td>BESM</td>
<td>1.875°</td>
<td>1.875°</td>
</tr>
<tr>
<td>CANESM2</td>
<td>2.75°</td>
<td>2.8125°</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The API developed for the platform uses features that make it easy and agile to integrate with new products and services. These services or products include mobile applications, packages or libraries for the most diverse programming languages, legacy systems of public or private companies, Web applications and Cloud Computing (Figure 4). The exchange of information through API has the role in standardizing these integrations and providing some autonomy for companies and institutions to enjoy the services available that make uses of the platform data.

Results and discussion

The PROJETA platform enables users to make requests to the database containing projections of climate change in two ways: via the Web or directly via the API. In the Web
request, one can select an area on a map, a geographic point by indicating its longitude and latitude coordinates or a Brazilian city. In the API request, one can define only a specific geographic point. Examples of requests for these two forms are given below.

Web request is performed in four steps: (1) choosing the request data (Figure 5); (2) visualization of the area, geographical point or Brazilian city selected in step (1) (Figure 6); (3) registration of the data of the user who is executing the requisition; and (4) presentation of a summary with the data of the request made by the user.

After completing this last step, a message is displayed to the user informing them about the progress of the processing of their request and about sending a message with instructions for downloading the requested data to their email. In this email, there will be a link to access the data generated by the request. When accessing this link, a Web page will open on the platform where the request summary will be presented and a button to download the request data (Figure 7).

In this request example, the file with the requested data is generated in CSV (comma-separated values) format. The request consists of an area that covers the Brazilian state of Rio Grande do Sul (area bounded by the blue rectangle on the map), with the following specifications: climatic scenario (MIROC5 continental model downscaling by the Eta model, 20-km resolution, with RCP 4.5), annual frequency, variables such as the air temperature at 2 meters from the surface (TP2M) and period from 2006 to 2099.

Figure 8 shows the first 28 rows out of a total of 157,169 of the CSV file generated by the request, where each line corresponds to the temperature (TP2M) in a given year (Data) for a given latitude (Latitude) and longitude (Longitude) of the requisition. The Hour column is generated by default by the system but the annual frequency is not used and its values are zeroed (00:00:00).

For the API request it is recommended that the user first access the platform’s "API" menu (https://projeta.cptec.inpe.br/#/api) where a tutorial on how to configure the API is presented, presenting a description of all parameters used to build the HTTPS link through the API, and a service for automatic generation of the link for access via the API, similar to the main menu when the user makes a
Web request. The API link structure is as follows:

https://projeta.cptec.inpe.br/api/v1/public/ETA/id-model/freq/id-freq/im/fm/fy/var/lat/lon/

In this link to the request, the id-model parameter is the identifier of the model coupled to the Eta model, freq is the frequency of the data and id-freq is its identifier, im and iy are the initial month and year of the request, fm and fy are the final month and year of the requisition, var is the climate variable, lat is the latitude and lon is the longitude of the geographic coordinates of the point to be ordered. With this link configured the user can use it directly in a Web browser or any tool that can access APIs via HTTP. The data view, when accessed by the API, is in JavaScript Object Notation (JSON) format. The following is an example of an API request link:


This request consisted of a geographic point centered

Figure 6: The second step of Web request - in this step the area selected in the request is displayed (area represented by the blue rectangle on the map). This area comprises the rectangle with the geographical coordinates defined in (C) in the web request example shown in Figure 5.

Figure 7: Access to the data of Web request: when accessing the link notified in the email sent to the user after the conclusion of the request, a platform Web page will show the request summary (A) and a button for starts downloading the data (B). The data will be downloaded in a compressed file, containing the data in the file format choose in the request and a file in pdf format containing the request summary.
in the Brazilian city of Passo Fundo, state of Rio Grande do Sul, with the following specifications: climatic scenario (MIROC5 continental model coupled to Eta model, resolution of 20 km with RCP 4.5), frequency monthly, variable is the air temperature at 2 meters from the surface (TP2M) and period of June and July 2018. Figure 9 shows the output of the request in JSON format, where each set of data corresponds to the value of the temperature variable (TP2M) in a given month/year (date) for the latitude (lat) and longitude (lng) of the request. The time variable is generated by default by the system but in the monthly frequency is not used and its values are zeroed (00:00:00).

Conclusions

The PROJETA platform made it possible to make requests, via the Web and via API, to climate change projection data generated by CPTEC/INPE. The computational architecture used in the development of the platform allowed the opportunity to scale the use of its data through the API, the inclusion of new modules in the structure and the adaptation or inclusion of new scenarios of projections in an easy and organized way in the structure of the database.

The platform automates the process of requesting, making available and viewing the climate change data that was previously performed manually. In this way, the user can make use of the PROJETA interface and filter the data as needed. Upon completion of the request, in a few minutes, the user receives an email with the steps to download the requested data. Some restrictions were imposed to prevent the user from requesting a very large amount of data, mainly for the 3-hour frequency. To provide this restriction, limits are set to only one variable for each request, and the time interval that can be requested in a single request, according to the type of selected frequency.

The API developed to communicate the database with the PROJETA’s website allowed its use by other government agencies, companies and educational and research institutions. In this case the communication is done by URL, where any programming language that supports HTTP will be able to easily access the data of this database via the use of this API and integrate it into your applications.

As future work, some improvements in the PROJETA platform are under development, both in platform structure and in creating tools to ease access to this data more easily. Among these improvements is the creation of a new platform module to present climate change data charts and maps, offering data manipulation and visualization capabilities in a dashboard-style displayed in a graphical layout. Another advancement is the creation of an R package to reach and manipulate directly the platform data.

Figure 8: First 28 lines of the CSV file generated by the request shown in Figure 5.

Figure 9: Output of the request in JSON format made by API.

```json
{
  "value":11.9082612991333,
  "date":"2018-06-01",
  "time":"00:00:00",
  "variable":"TP2M",
  "lat":-28.12499999999998,
  "lng":-52.47499999999994
}
{
  "value":10.259069442749,
  "date":"2018-07-01",
  "time":"00:00:00",
  "variable":"TP2M",
  "lat":-28.12499999999998,
  "lng":-52.47499999999994
}
```

Acknowledgments

The PROJETA platform was developed under the Climate Change Policy Program (PoMuC), coordinated by the Ministry of the Environment (MMA), and supported by the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH, in the context of the Brazil-Germany Sustainable Development Cooperation.
PROJETA platform: accessing high resolution climate change projections over Central and South America using the Eta model.

CITATION

Plataforma PROJETA: acesso a projeções de mudança do clima em alta resolução sobre as Américas do Sul e Central usando o modelo Eta

Carlos Amaral Hölbig1(*), Angela Mazzonetto1, Felipe Borella1, Willingthon Pavan1, José Maurício Cunha Fernandes1,2, Diego José Chagas3, Jorge Luis Gomes3 e Sin Chan Chou3

1Universidade de Passo Fundo – Programa de Pós-Graduação em Computação Aplicada, Rodovia BR 285, Bairro São José, CEP 99052-900 Passo Fundo, RS, Brazil. E-mails: holbig@upf.br, angelamazzonettofw@gmail.com, borellaster@gmail.com e pavan@upf.br
2Embrapa Trigo, Rodovia BR 285, km 294, Caixa Postal 3081, CEP 99050-970 Passo Fundo, RS, Brazil. E-mail: mauricio.fernandes@embrapa.br
3Instituto Nacional de Pesquisas Espaciais (INPE), Rodovia Presidente Dutra, km 39, SP-RJ, CEP 12630-000 Cachoeira Paulista, SP, Brazil, E-mail: diegojc@gmail.com, jorge.gomes@inpe.br e chou.sinchan@inpe.br
(* Autor para correspondência.

RESUMO

A busca por dados de mudanças climáticas, por parte de pesquisadores, entidades governamentais ou empresas privadas, é uma demanda recorrente que é dificultada pela forma de acesso a este tipo de informações, devido, principalmente, à complexidade de se trabalhar e disponibilizar estes dados, que muitas vezes podem superar os terabytes de tamanho. A plataforma PROJETA visa automatizar o processo de extração e disponibilização dos dados das projeções climáticas regionalizadas para a América do Sul, por meio do modelo regional Eta, gerados pelo CPTEC/INPE. O processo de tratamento e conversão dos dados, que antes era feito de forma manual e complexa, agora possui acesso a um amplo e irrestrito conjunto de diversas variáveis climáticas, disponíveis aos diferentes usuários via Web ou API, de forma flexível quanto ao formato e ao volume de dados. Para tanto, a plataforma dispõe de uma base de dados com novas projeções que utilizam diferentes modelos climáticos globais, integrando tecnologias que possibilitam o acesso ao banco de dados de forma ágil e fácil, para posterior utilização em aplicações que necessitam destes conjuntos de dados climáticos.

Referências