An R package to access climate change data for South America regionalized by the Eta Model of CPTEC/INPE

Marina Lopes Dezordi[1], Carlos Amaral Hölbig[1],
Willingthon Pavan[1], José Maurício Cunha Fernandes[1]

[1]Graduate Program in Applied Computing (PPGCA)
University of Passo Fundo (UPF)
BR 285, Campus 1, CEP 99052-900 – Passo Fundo – RS – Brazil

```
marinadezordi@gmail.com,
{holbig,pavan}@upf.br, mauricio.fernandes@embrapa.br
```

Abstract. *This paper introduce an R package developed with the aim of facilitating the access of climate change data from South America generated by CPTEC/INPE. It presents related works that has been used as the basis of package development. This package has functions to climate data access and functions to help users to operate package functions. With this package users and developers of R language have an easy tool to access and manipulate climate change data generated by CPTEC/INPE.*

Resumo. *Este artigo apresenta um pacote R desenvolvido com o objetivo de facilitar o acesso de dados de mudanças climáticas para a América do Sul gerados pelo CPTEC/INPE. São apresentados alguns trabalhos relacionados que foram utilizados como base no desenvolvimento do pacote. O pacote possui funções para acesso a dados climáticos e funções que auxiliam na utilização das funções do próprio pacote. Com este pacote, desenvolvedores e usuários da linguagem R podem contar com uma ferramenta de fácil utilização para o acesso e a manipulação destes dados.*

1. Introduction

The National Institute for Space Research (INPE), since 1997, has provided numerical weather forecasts with greater detail for much of South America and the climate change scenarios were generated from the Eta regional climatic model [Chou 2014]. The model was also used to produce climate change surveys [Pesquero 2010; Chou 2012; Marengo 2012] and support the development of Third National Communication of Brazil to the United Nations Framework Convention On Climate Change[1]. New emission scenarios were proposed by Intergovernmental Panel on Climate Change (IPCC) that has generated new global climate change projections. Estimate of HadGEM2-ES and MIROC5 models, in scenarios of greenhouse gas emissions RCP4.5 and RCP8.5, were regionalized by Eta model of INPE [Chou 2014; Chou 2014a].

An important point about the data generated by these models is that they are generally available via FTP, in text or binary format. Some rounds of climate change data can reach terabytes of data, for this reason it ends up being available directly by CPTEC by other physical means, for example external hard disk. Climate change data, which the package described in this work has access, comprises a historical period

---

[1] Available at http://unfccc.int/resource/docs/natc/branc3es.pdf

(observed data) from 1961 to 2005 and a projections period (simulated data) from 2006 to 2099. The Eta model was configured at 20 km resolution in the area covering South America, Central America, and the Caribbean.

The EtaModelCC package was created aiming to facilitate access to the climate change data generated by CPTEC/INPE, in order to enable R language users operate this data in their systems and simulations. There are some R packages with similar features compared to EtaModelCC such as **rNOMADS**, **rnoaa**, **darksky**, **RNCEP** and **rwunderground**, but using data generated by global climate change models from international research centers. Other works about this subject are described in [Fraley, Raftery, and Gneiting 2011] and [Serrano-Notivoli, de Luis and Beguería 2017]

2. EtaModelCC package

EtaModelCC package was developed in order to provide climate change data from South America generated by CPTEC/INPE using the regional climate model Eta. Functions have been developed to return informations as required by the user. Being a package it is possible to other systems to use its functions without making complex implementations to access data, in this way the development becomes less complex and faster.

Functions to get the climate change data generate a list (a generic vector containing other R objects) as output. The last record in the list is the dataframe (structure used for storing data tables in R) called `$data` which contains the result requested by the user. Other items in the list are: information about the models drivers, data frequency, coordinates (longitude and latitude), variables and others. The global climate models drivers were the HadGEM2-ES model of the British UK Met Office, the Japanese model MIROC5, the CanESM2 model and the Brazilian BESM model. To achieve these goals, the package includes, in its first version, four functions:

- `getClimateData()`: returns the climate change data;
- `verifyLonLat()`: verifies if the coordinates are in the data range of model (internal function);
- `adjustCoordinate()`: sets the point to the nearest coordinate (internal function);
- `getInfoClimate()`: shows the details about the parameters that will be used by `getClimateData()` function.

The package is currently available on the GitHub source code platform. To perform the installation from R it is necessary to use **devtools** R package, and `intall_github()` function. The command is as follows: `install_github("holbig/EtamodelCC")`.

EtaModelCC package has a function called `getInfoClimate()` to guide the use of climate model functions. This function contains the necessary information about the correct use of the `getClimateData()` function. The ids of the existing global climate models drivers are listed with their descriptions and date periods, so that the user can decide the climate scenario required, beyond start and end date range for a valid return. Available frequencies are reported every three hours (HOURLY), daily (DAILY), monthly (MONTHLY) or annually (YEARLY). Finally, the `getInfoClimate()` function lists all possible variables (about 60 variables) to

return values on climate change models, combined with a brief explanation (see Figure 1).

```
> getInfoClimate()
Avaiable Models: model has to be accessed by the value of
field <id>. Check the models's coverage period:
inicial date (iMonth, iYear) e final date (fMonth, fYear)

 id name                                     model couple      scenario    resolution start_year end_year
  1 20 km, RCP4.5, continental, MIROC5.       ETA   MIROC5      RCP4.5      20         2006       2099
  2 20 km, RCP8.5, continental, MIROC5.       ETA   MIROC5      RCP8.5      20         2006       2099
  3 20 km, RCP4.5, continental, HADGEM2-ES.   ETA   HADGEM2-ES  RCP4.5      20         2006       2099
  4 05 km, RCP4.5, sudeste-BR, HADGEM2-ES.    ETA   HADGEM2-ES  RCP4.5      5          2006       2099
  5 20 km, RCP8.5, continental, HADGEM2-ES.   ETA   HADGEM2-ES  RCP8.5      20         2006       2099
  6 05 km, RCP8.5, sudeste-BR, HADGEM2-ES.    ETA   HADGEM2-ES  RCP8.5      5          2006       2099
  7 20 km, RCP4.5, continental, CANESM2.      ETA   CANESM2     RCP4.5      20         2006       2099
  8 20 km, RCP8.5, continental, CANESM2.      ETA   CANESM2     RCP8.5      20         2006       2099
  9 20 km, RCP4.5, continental, BESM.         ETA   BESM        RCP4.5      20         2006       2099
 10 20 km, RCP8.5, continental, BESM.         ETA   BESM        RCP8.5      20         2006       2099
 12 20 km, Histórico,  continental, BESM.     ETA   BESM        Historical  20         1961       2005
 13 20 km, Histórico, continental, CANESM2.   ETA   CANESM2     Historical  20         1961       2005
 14 20 km, Histórico, continental, HADGEM2-ES. ETA  HADGEM2-ES  Historical  20         1961       2005
 15 20 km, Histórico, continental, MIROC5.    ETA   MIROC5      Historical  20         1961       2005
 16 05 km, Histórico, sudeste-BR, HADGEM2-ES. ETA   HADGEM2-ES  Historical  5          1961       2005

modelFrequency - Avaiable frequencies:
HOURLY : every 3 hours
DAILY  : daily
MONTHLY: monthly
YEARLY : yearly

Avaiable Variables - variable (modelVar) has to be accessed
by its short name <variable>
 variable description
 PSLM      Pressão ao Nível Médio do Mar
 PSLC      Pressão à Superfície
```

Figure 1. Part of output of `getInfoClimate()` function that describes the variables of climate change models of CPTEC/INPE.

2.2. Accessing model data

The function that access climate data is called `getClimateData()` and its access is developed through API access[2] from PROJETA Project website[3]. The original data are in JSON format and the EtaModelCC package read this data and convert then to a list of R. The users need to enter some parameters: template id (global model), frequency, variable, longitude, latitude, initial month and year, final month and year[4]. For the definition of the climate scenario it is important to consult the models available in `getInfoClimate()` function, and their respective ids, in addition to checking the coverage period of each one. Longitude and latitude has to be reported in decimal geodetic format, with integer degrees and minutes converted to fractional degrees. API transforms the coordinates to the coverage point closest to the one entered by user.

---

[2] Data generated by CPTEC/INPE and available on the PROJETA website.

[3] Available at https://projeta.cptec.inpe.br/.

[4] Example of function call: getClimateData('1', 'YEARLY', 'TP2M', '-12', '-49', iYear=2006, fYear=2010)

3. Discussion and applications

From this package R users can easily access functions to obtain climate change data. Coordinates, model, date range, frequency and variables can be chosen to obtain output data summarized as users needs. These data can be used in the execution of simulation models of applications that deal, for example, with water availability and its impact on food production and on the generation of energy and simulation of the growth of crops.

4. Conclusions

EtaModelCC package presented in this paper offers a simplified way to access climate change data generated by CPTEC/INPE. It complements the APIs development, which provide climate change data in JSON format, besides the PROJETA project website, designed to facilitate access of API links. In addition to being a more potential alternative than the availability of data by CPTEC, which often needs to use physical media by reason of the data size. The development of functions that generate graphics/maps based on these data is being developed for future versions of the package.

References

Chou, S. C., et al. (2012) "Downscaling of South America Present Climate Driven by 4-Member HadCM3 Runs", Climate Dynamics, 38: 635-653. doi: 10.1007/s00382-011-1002-8.

Chou, S. C., et al. (2014) "Evaluation of the Eta Simulations Nested in Three Global Climate Models", American Journal of Climate Change, 3, 438-454. doi:10.4236/ajcc.2014.35039.

Chou, S. C., et al. (2014a) "Assessment of Climate Change over South America under RCP 4.5 and 8.5 Downscaling Scenarios", American Journal of Climate Change, 3, 512-527. doi: 10.4236/ajcc.2014.35043.

Fraley, C., Raftery, A. E. and Gneiting, T. (2011) "Probabilistic weather forecasting in R", The R Journal, 3(June), 55–63. doi: 10.1198/jasa.2009.ap07184.

Lyra, A., et al. (2017). "Climate change projections over three metropolitan regions in Southeast Brazil using the non-hydrostatic Eta regional climate model at 5-km resolution", Theor Appl Climatol. doi:10.1007/s00704-017-2067-z.

Marengo, J. A., et al. (2012) "Development of regional future climate change scenarios in South America using the Eta CPTEC/HadCM3 climate change projections: climatology and regional analyses for the Amazon, São Francisco and the Parana River Basins", Climate Dynamics, 38: 1829-1848. doi: 10.1007/s00382-011-1155-5.

Pesquero, J. F., et al. (2010). "Climate downscaling over South America for 1961-1970 using the Eta Model", Theor Appl Climatol, 99: 75-93. doi: 10.1007/s00704-009-0123-z.

Serrano-Notivoli, R., de Luis, M. and Beguería, S. (2017), "An R package for daily precipitation climate series reconstruction", Environmental Modelling and Software, 89, 190–195. doi: 10.1016/j.envsoft.2016.11.005.