

Model Replication Guide: Analytic Modeling of Rainwater Harvesting in the Brazilian Semi-Arid Northeast

James Doss-Gollin Francisco de Assis de Souza Filho
Francisco Osny Enéas da Silva

September 6, 2015

Abstract

This is a guide to the model used in the paper Analytic Modeling of Rainwater Harvesting in the Brazilian Semi-Arid Northeast. This document will explain the steps needed to replicate the findings of the paper Analytic Modeling of Rainwater Harvesting in the Brazilian Semi-Arid Northeast, including accessing and downloading the data, cleaning and formatting the data, and running the model under default assumptions or sensitivity analysis. This document is intended to accompany the paper, which explains the choices and assumptions that drive this model.

Contents

1	Quick Start	3
2	System Requirements	3
3	Downloading the Raw Data	3
3.1	Data Use Policy	3
3.2	Registering with the INMET	3
3.3	Downloading the Data	4
3.3.1	Method 1: Manual Download	4
3.3.2	Download Using RCur1	5
4	Running the Model	6
4.1	Running with Default Parameters	6
4.2	Modifying Parameters	6
4.3	Sensitivity Analyses	7
5	Description of Files	7
5.1	The code Folder	7
5.2	The data Folder	8
5.3	The doc Folder	8
A	Acronyms Used	9
	References	9

1 Quick Start

There are some steps you must perform in order to replicate this model. Each step is described in greater detail below.

1. Ensure that you have all appropriate software. See section 2 for help.
2. Register with the Instituto Nacional de Meteorologia – National Meteorology Institution (INMET) and download the data. We apologize that we are not able to provide you with our raw data, due to the data use policy of the INMET which owns the data. Downloading their data is straightforward, except that the web site is in Portuguese. For step-by-step instructions please see section 3.
3. Run the model under normal conditions (section 4.1) and/or run a sensitivity analysis (section 4.2).

2 System Requirements

Access to Matlab proprietary software is required. This model was created and has been tested on Matlab 2014a and 2015a, and its compatibility with other versions is not guaranteed. Some toolkits beyond the core may be required, and specifically the Mapping Toolkit is required to draw figures (this feature is not required in order to run the program). We request that anyone attempting to run this model on other versions of Matlab advise us of the compatibility so that we can update this document. The model should function independently and should work on most standard operating systems: it has been tested successfully on Linux Mint 17, Microsoft Windows 8, and OS X Yosemite.

3 Downloading the Raw Data

3.1 Data Use Policy

We understand that the most convenient and efficient way for you to use this model is for us to include our data along with our code. Unfortunately, doing so would violate the data use policy of the INMET, which owns the data. They require that each person who uses the data register with their online site. Therefore, all we can do is provide the best possible instructions for accessing their data.

3.2 Registering with the INMET

In order to download the data from the INMET you need to create an account with them at <http://www.inmet.gov.br/projetos/rede/pesquisa/>. If you have difficulty with this step, please email james.doss-gollin@columbia.edu.

3.3 Downloading the Data

There are two methods that you can use to download the data. The first is tedious but straightforward; the other is more efficient but requires greater programming ability.

3.3.1 Method 1: Manual Download

1. Go to the INMET Banco de Dados Meteorológicos para Ensino e Pesquisa – Meteorological Database for Research and Teaching (BDMEP) website at <http://www.inmet.gov.br/projetos/rede/pesquisa/inicio.php>.
2. Enter your user name (your email address) and password
3. Under the *Pesquisa – Estação Convencional* tab, click on the link that says *Série Histórica – Dados Diários*
4. Fill out the new dialogue box as follows:

Período - Data início (dd/mm/aaaa) Start date (dd/mm/yyyy format). Analytic Modeling of Rainwater Harvesting in the Brazilian Semi-Arid Northeast uses 01/01/1950.

fim End date (dd/mm/yyyy format). Analytic Modeling of Rainwater Harvesting in the Brazilian Semi-Arid Northeast uses 31/12/2013.

Drop-Down Menu the quickest option is to select the drop-down menu on the left (*Região*) and select *Nordeste*. However, this may omit one or more of the stations used in the paper. For full replication, you must select Ceará from the drop-down menu on the right. The following steps will then need to be repeated for the states of Rio Grande do Norte, Paraíba, Pernambuco, Alagoas, Sergipe, Bahia, Minas Gerais, Piauí, and Maranhão.

checkbox Click the checkbox corresponding to *precipitação (mm)*.

submit Click the button at the bottom that says *pesquisa*. Confirm if you are asked to do so.

5. A list of stations will appear on the right. For each one, you must:
 - (a) Click on the station
 - (b) Open the resulting tab
 - (c) Save the data to a text file with the name #####.txt, where ##### is the station code (you will see (OMM: #####)). Please note that a typo in this file name should not affect the code in any way.
6. When you have downloaded all files, move all the text files into the data/INMET folder.

3.3.2 Download Using RCurl

The manual download process, while effective, is undeniably cumbersome. We have written a script using the R language and the RCurl package to automate the download process. The script is written to substitute for downloading files individually, and writes to file the same .txt files that manual download produces. This process runs slowly, but you do not need to remain at your computer while it runs. We also cannot guarantee that it will work perfectly, but we are committed to resolving any bugs with this script that arise.

To use the RCurl method, you must first have the R language installed on your computer. You can also ensure that you have the RCurl package installed with the command `install.packages("RCurl")`. If you are not familiar with R or the instructions below, it is probably easier to download the data files manually.

To use the script to automatically download data files from the BDMEP:

1. If you have not already done so, follow the steps outlined at section 3.2 and create an account with the INMET.
2. Go to <https://github.com/jdossdollin/inmet> and select the *download as .zip* option
3. Extract the zip file and open the folder
4. Enter your email and BDMEP password into the corresponding spaces of the R script.
5. If you wish, modify the start and/or end date in the script
6. Run the code as source
7. Move the data files generated to the /data/INMET directory of the Matlab program described in this document
8. For more questions about the R script to download these files, please use the features of the Github web site.

4 Running the Model

Before you can run the model, you need to download the raw data from the INMET and move the raw text files to the data/INMET folder. If you have not already done that, see section 3. Otherwise, you are ready to run the model: it is designed to work as soon as it is downloaded.

4.1 Running with Default Parameters

All you need to do to run the model is to run the script `main.m`. It is important not to modify the directory structure or move files out of the directory, because they call each other using relative paths. This script prints results to screen. By default, map generation is turned off; to change this option, change the `const.opt.figure toggle` option in `declarations.m` to 1 (see section 4.2 for details on modifying parameters). You should see results printed to screen giving average calculated reliabilities for different roof sizes.

4.2 Modifying Parameters

All parameters should be modified exclusively in the `declarations.m` file. Parameters and options are designed to be passed through the struct called `const` and are not hard-coded into individual functions. `declarations.m` describes each parameter and its role in the model. If modifying `declarations.m` it is recommended to leave the default parameters commented, rather than deleting them, so that it is easier to return to the default configuration. It is also important to note that the `getData.m` script looks for two files: `data.mat` and `data_unfiltered.mat`. If `data.mat` exists, then `getData.m` will simply load `data.mat` and will not attempt to filter the data. If `data.mat` does not exist but `data_unfiltered.mat` does exist, it will load and filter this data but will not attempt to re-read files added to the INMET folder. For this reason, after changing the criteria by which the data should be filtered it is imperative to delete `data.mat` and after altering the raw data from the INMET it is imperative to delete both `data.mat` and `data_unfiltered.mat`.

4.3 Sensitivity Analyses

To carry out sensitivity analyses, use the `mainForSensitivityAnalysis.m` script. The script is divided into sections, each of which lays out sensitivity analyses for a different model parameter. To run all analyses, run the script; to run analysis for a single variable, position the cursor within that section and click the Run Section command. To modify the values of one or more of the parameters manually, change it in this section (not in `declarations.m`). That is because this file calls `declarations.m` and over-rides the parameter for which the sensitivity analysis is being carried out. The `mainForSensitivityAnalysis.m` script displays results to the screen and creates a new folder `Results` in which results for each variable are saved as `.csv` files. Note that only numbers – no header rows or roof sizes – are printed to the `.csv` files.

5 Description of Files

The folder containing this model is organized as follows. There are three folders, called `data`, `doc`, and `code`. The contents of each are briefly summarized in this section.

5.1 The code Folder

The following Matlab files are included in the code directory. Documentation of each can be accessed by typing `help fname`, where `fname` is the name of the file. A brief summary of each file is provided here:

`main.m` calls all functions required to run the model. To run the model, just run this file.

`mainForSensitivityAnalysis.m` calls all functions required to run sensitivity analysis.

`declarations.m` defines model options and parameters; saves them to a single struct which is passed as an argument to many other functions.

`getData.m` determines what data has already been loaded and calls functions to read raw data from text file and filter data if filtered data does not already exist.

`readINMET.m` function to read data files in .txt format downloaded from the INMET website. This script reads post number, coordinates, and first word of post name; saves precipitation data as .mat file.

`makePrecipMatrix.m` reads precipitation .mat files generated by `readINMET.m` and creates a single matrix for all specified posts.

`filterData.m` filters data by location and quality.

`loadRoofData.m` reads roof data from one of the three files specified in section 5.2.

`cConsumption.m` calculates consumption as a function of the consumption rule specified in `declarations.m` and day of the year.

`cBalance.m` carries out water balance for set of meteorological posts.

`drawMaps.m` draws maps and plots results of analysis.

5.2 The data Folder

The data folder contains three child directories which contain separate data sets:

`INMET` is currently empty. Once you have downloaded the data from the INMET website, move the text files to this folder.

`Roofs` contains four text files containing roof data:

`default.txt` contains the default set of roofs used for the simulation as presented in the article.

`Milha.txt` contains the full distribution of roof sizes published available as a supplemental file to the online version of the article.

`shapefile_brasil` contains a an ESRI-format shapefile, generated by the Instituto Brasileiro de Geografia e Estatística – Brazilian Institute of Geography and Statistics (IBGE), which is used to overlay the political divisions of Brazil onto maps.

It is worth noting that after running the model, several data files in `.mat` format will be generated and saved directly in the data folder.

5.3 The doc Folder

This doc folder contains the documentation required for you to run the program. A `.pdf` file is compatible on virtually all machines. For those who are interested, we also provide the `.tex` code used to generate the `.pdf` document. The `.bib` file containing references used in this document is also provided.

Appendix

A Acronyms Used

INMET Instituto Nacional de Meteorologia – National Meteorology Institution

IBGE Instituto Brasileiro de Geografia e Estatística – Brazilian Institute of Geography and Statistics

BDMEP Banco de Dados Meteorológicos para Ensino e Pesquisa – Meteorological Database for Research and Teaching