LARS-WG: A STOCHASTIC WEATHER GENERATOR FOR CLIMATE CHANGE IMPACT ASSESSMENTS

Mikhail A. SEMENOV

LARS-WG 6.0 stochastic weather generator is a computationally inexpensive downscaling tool to generate local scale climate scenarios based on global or regional climate models for impact assessments of climate change. LARS-WG has been used in more than 75 countries for research and education. The current version 6.0 of LARS-WG incorporates climate projections from the CMIP5 ensemble used in the IPCC 5th Assessment Report. LARS-WG has been well validated in diverse climates around the world

TECHNICAL INFORMATION

LARS-WG runs on a Windows PC. The latest version of LARS-WG can be downloaded from:

https://sites.google.com /view/lars-wg/

OUTLINE OF THE WEATHER GENERATION PROCESS

The process of generating a local-scale climate scenario can be divided into two steps:

ANALYSIS: Observed daily weather for a site is analysed to compute site parameters. This information is stored in three files: a *.wgx file contains the site parameters file, a *.stx-file contains some additional statistics and a *.tst file contains the results of statistical tests of LARS-WG performance by comparing the observed and generated data. These files must not be modified. The structure of these files is described in Appendix 2-4.

GENERATOR: The site parameter file, *.wgx file, is used to generate synthetic daily weather (a baseline scenario) which is statistically similar to observed weather at a site. By applying changes in climate, Δ-changes, derived from a global or regional climate model to baseline site parameters, the user is able to generate a local-scale future scenario matching projected future climate. LARS-WG 6 incorporates climate projections from 19 global climate models (GCMs) from the CMIP5 ensembles used in the IPCC 5th Assessment Report (more details in [1,2]).

REGIONS: For some regions, where LARS-WG site parameters have been already estimated, LARS-WG can be run without observed daily weather. To generate a future scenario, the user should spesify longitude, latitude and altitude of the site of interest in the region and a climate projection defined by one of the available GCMs. Currently, several regional datasets are under development, including the ELPIS dataset constructed for Europe [3].

BATCH RUN: This option allows to estimate site parameters or generate climate scenarios as a batch run for many site × scenario combinations. You have to prepare a batch file which specifies sites and scenarios you would like to run. The following options are available: (1) generate site parameters for many sites, (2) generate climate scenarios for many combinations of site × GCM × RCP × period, (3) generate climate scenarios for many combinations of site × (user-defined scenario), (4) generate climate scenarios for many combinations of (lon, lat, alt) × GCM × RCP × period from ELPIS dataset. For more details see Appendix 5.

DATA PREPARATION

Each file with observed daily weather should be accompanied by a *.st file containing information about the site (name, location and CO2), the path to the weather files followed by 'tags' describing the data format in the weather file. The format of *.st file is described in Appendix 1. LARS-WG can use as little as a single year of observed daily weather data. However, since the generated weather will be based on probability distributions derived from this observed data, the use of longer daily records guarantees better approximation of observed climate. The use of 20-30 years of daily weather is recommended. In order to capture extreme climate events, a longer observed records might be needed. Missing value is "-99".

If LARS-WG suspects errors in the weather file, then it will display an error message and log this error into the "Error.txt" file. LARS-WG will treat potential errors as 'missing values' and they will be excluded from analysis.

TESTING PERFORMANCE

When the site parameters are calculated, LARS-WG automatically computes a *.tst file with the results of statistical tests comparing generated and observed weather. Statistical tests include the Kolmogorov-Smirnov (K-S) test to compare the probability distributions, t-test to compare means and F-test to compare standard deviations. The structure of a *.tst file is described in Appendix 4.

CONTACT DETAILS

Dr Mikhail Semenov
Rothamsted Research
e: mikhail.semenov@gmail.com

REFERENCES

[1] Semenov MA, Stratonovitch P (2010) The use of multi-model ensembles from global climate models for impact assessments of climate change. Climate Research 41:1-14

[2] Semenov MA, Stratonovitch P (2015) Adapting wheat ideotypes for climate change: accounting for uncertainties in CMIP5 climate projections. Climate Research 65: 123–139

[3] Semenov MA, Donatelli M, Stratonovitch P, Chatzidaki E, Baruth B (2010) ELPIS: a dataset of local-scale daily climate scenarios for Europe. Climate Research 44:3-15

APPENDIX 1: Structure of the st-file.

[SITE]
SiteName as a single word without spaces
[LAT, LON and ALT]
latitude, longitude and altitude (in meters) separated by space or tab
[CO2]
CO2 concentration for baseline period (in ppm), e.g. 368
[WEATHER FILES]
A list of names of weather files (one file per line). You do not need to specify a full path if all your files are in the same directory as a corresponding *.st files. The records in weather files must be in chronological order.
[FORMAT]
A FORMAT line describes the structure of your weather files. The following tags can be used: YEAR - year, MONTH - month, DAY – day of the month, JDAY – day of the year, MIN – minimum temperature, MAX – maximum temperature, RAIN – precipitation, RAD – radiation, SUN – sun hours, SKIP – skip this column
[END]

A comment line can be place anywhere in a file and should start with "// ". In your weather file the values of variables for each day should be separated by space or tab. The data should not contain blank lines, comment lines or headers. Missing data values is "-99".

APPENDIX 2: Structure of a *.wgx file

For each month a *.wgx file contains following information: semi-empirical distributions for: length of dry and wet series, precipitation, minimum and maximum temperature and radiation calculated separately for dry and wet series (wet day is defined as a day with precipitation > 0.1), and correlation and auto-correlation coefficients.

A semi-empirical distribution is described by a block of 3 lines. The 1st line consists the sample size, number of bins used, observed mean and standard deviation, generated mean and standard deviation (based on the sample of 1000). The 2nd and 3rd lines describe empirical cumulative probability function: 2rd line shows cumulative probabilities and 3rd line shows values of the climatic variable.

APPENDIX 3: Structure of the *.stx file

This file contains semi-empirical distributions for various climatic variables, including precipitation, minimum and maximum temperature and radiation, length of wet and dry spells, length of period of frost and heat waves. It also contains monthly means and standard deviation and information of minima and maxima of climatic variables for each month and its percentiles. The following is an example for precipitation.

[RAIN distributions]
12 blocks (one for each month) of 3 lines each. Each block describes semi-empirical distribution of daily precipitation.

[RAIN monthly statistics: monthly total max and min, N of observations, monthly mean and sd]
Each line contains 12 values (one value for each month). 1st line is maximum of monthly total precipitation, 2nd line is minimum of monthly total precipitation, 3rd line is the number of years, 4th line is monthly total means and 5th line is monthly total standard deviation.

[RAIN daily maxima: median, 95 percentile and maximum]
Each line contains 12 values (one value for each month) of maxima of daily precipitation for each month. 1st line is median of maxima; 2nd line is 95 percentile and 3rd line is maximum of maxima.

APPENDIX 4: Structure of the *.tst file

The *.tst file contains the results of statistical tests comparing the observed and generated data.

[KS-test for seasonal wet/dry SERIES distributions: Effective N, KS statistic and p-value]
The quarterly probability distributions for the length of wet and dry series are compared using the K-S test.

[KS-test for daily RAIN distributions: Effective N, KS statistic and p-value]
The probability distributions of daily precipitation for each month are compared using K-S test.

[RAIN monthly mean & sd: obs mean & sd, gen mean & sd, t- and f- statistics with p-values]
Block of 8 lines of 12 columns (12 months) of data. Lines 1 and 2 are the monthly mean precipitation totals and standard deviations calculated from the observed data. Lines 3 and 4 are monthly mean totals and standard deviations of the synthetic data. Lines 5 and 6 are results of comparing mean precipitation and include t-statistics and corresponding p-values. Lines 7 and 8 are the results of the F-test to compare variance and include f-statistics and corresponding p-values.

[RAIN daily maxima: obs and gen median, 95 percentile and maximum]
Block of 6 lines of 12 columns (12 month) of data. First 3 lines are median 95 percentile and maximum of maxima of observed daily precipitation for each month. The next 3 line are median 95 percentile and maximum of maxima of generated daily precipitation for each month.

Similar blocks are presented for maximum and minimum temperature and radiation.

[KS-test for seasonal frost/heat SPELLS distributions: Effective N, KS statistic and p-value]
The quarterly probability distributions for the length of frost spells (minimum temperature < 0C) and heat spells (maximum temperature > 30C) are compared using the K-S test.

INTERPRETATION of p-values

The statistical tests assume that the observed weather is a random sample from some existing distribution, which represents the ‘true’ climate at the site. The statistical tests look for differences in distributions derived from the generated and observed weather. Each of the tests computes a test statistic and a corresponding p-value, which indicate how likely generated and observed data came from the same distribution. If p-value is very low, below the significance level (set to 0.01 or 0.05), then the generated climate is unlikely to be the same as the observed climate. With the value of 0.05, a common significance level used in statistical tests, on average 1 in 20 tests show significant result even when there is no difference in distributions.

APPENDIX 5: Batch run

1. *.clp file is used by LARSWG6 in a batch run to create site parameter *.wgx files from observed data. All results are saved in "Site Analysis" directory defined in "Options". Each line contains only 1 parameter: PathToSiteFile (no "space" symbols are allowed), a full path to the *.st file with observed data. Any comment line in any batch files must begin with "//".
   This is an example of a *.clp file:

```
Z:\obs\CF.st
Z:\obs\RR.st
// my comment
```

2. A *.clg file used by LARSWG6 in batch run to generate daily climate scenarios for multiple sites. All results are save in "Output" directory defined in "Options". Each line contains 6 parameters separated by space. No "space" symbols are allowed inside parameters.

```
Path2wgxFile GCMName Emission Period NumYears RandSeed
```

Path2wgxFile is a full path to the LARS-WG parameter *.wgx file.
GCMName is the name of GCM, e.g. HadGEM2-ES for the Hadley Cnetre, UK,
Emission is the name of emission scenario, e.g. RCP45.
Period is a time period for simulation, e.g. “baseline”, “2041-2060”
NumYears is a number of generated years
RandSeed is a random seed (a prime number between 500 and 5000)

This is an example of a *.clg file:

```
//HadGEM2-ES RCP45 2021-2040
Z:\EU\wgx\CF.wgx  HadGEM2-ES    RCP45   2021-2040    100    2777
Z:\EU\wgx\DC.wgx  HadGEM2-ES    RCP45   2021-2040    100    2777
```

3. A *.cls file used by LARSWG6 in batch run to generate climate scenarios for multiple sites and multiple user-defined scenarios. All results are be saved in "Output" directory defined in "Options". Each line contains 4 parameters separated by space. No "space" symbols are allowed inside parameters.

```
Site.wgx   Scenario.sce    NumYears    RandSeed
```

Site.wgx is a full path to a *.wgx file
Scenario.sce is a full path to a user defined *.sce file
NumYears is a number of generated years
RandSeed is a random seed

This is an example of a *.cls file:

```
Z:\EU\wgx\CF.wgx myscenarios1.sce 100 2777
Z:\EU\wgx\DC.wgx myscenarios2.sce 100 3777
```

4. A *.cle file used by LARSWG6 in a batch run to generate daily climate scenarios for multiple sites in Europe using the ELPIS dataset. Each line contains 9 parameters separated by space.

```
SiteName   Lon   Lat   Alt   GCMName   Emission   Period   NumYears   RandSeed
```

SiteName, Lon, Lat, Alt are site name and geographical coordinates
GCMName is the name of GCM, e.g. HadGEM2-ES for the Hadley Centre, UK,
Emission is the name of emission scenario, e.g. RCP45.
Period is a time period for simulation, e.g. baseline, 2011-2030...
NumYears is a number of generated years
RandSeed is a random seed

This is an example a *.cle file:

```
//HadGEM2-ES RCP45 2021-2040
RRT    0.00    51.00   305    HadGEM2-ES    RCP45   2021-2040    100    2777
SLT    -5.88   37.42   79     HadGEM2-ES    RCP45   2021-2040    100    2777
```