LARS-WG 5: a stochastic weather generator for climate change impact assessments

LICENCE & TECHNICAL INFORMATION

LARS-WG is protected by a licence agreement. It may be used by individuals for academic research. To use LARS-WG within a research project, a licence is required. LARS-WG cannot be used for any commercial purposes. By using LARS-WG you agree to share derivatives from LARS-WG, such as site parameters and diagnostics, with academic community.

LARS-WG 5 can be run on a PC under any version of Windows. LARS-WG requires a little processing time in most situations. LARS-WG can be downloaded from http://www.rothamsted.ac.uk/mas-models/larswg.php. You have to register and obtain a key if you are planning to use LARS-WG for more than 30 days.

OUTLINE OF THE WEATHER GENERATION PROCESS

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

ANALYSIS: Observed daily weather for a site is analysed to compute site parameters. This information is stored in two files: a wgx-file is the site parameters file and a stx-file is the file with some additional statistics. Both files must not be modified. The structure of these files is described in Appendix 1 and 2.

GENERATOR: The site parameter file derived from observed weather data is used to generate synthetic daily weather for a site which statistically resembles observed weather. By applying changes in climate derived from a global or regional climate model (a sce-file must be created), the user is able to generate site-specific daily weather with characteristics specified in the sce-file. LARS-WG 5 incorporates climate projections from the multi-model ensembles of GCMs used in the IPCC 4th Assessment Report (more details in [1]).

REGIONS: For some regions, where LARS-WG site parameters have been already estimated, the 1st step, i.e. **ANALYSIS**, could be made redundant as well as a need for observed daily weather for a site. By providing longitude, latitude and altitude for the location of interest and specifying a climate projection from one of available GCMs, the user is able to generate a local-scale climate scenario for the future. Currently, several regional datasets are under development.

DATA PREPARATION

Each file with observed daily weather should be accompanied by a st-file containing information about the site (name and location), the directory path and name of the weather file, followed by a number of 'tags' describing the format of the data file. The format of st-file is described in Appendix 3.

LARS-WG will be able to generate synthetic weather data based on as little as a single year of observed daily weather data. However, since the generated weather data will be based on probability distributions derived from this observed data, so the more data observed used the closer LARS-WG be able to match the climate for the site in question. The use of at least 20-30 years of daily weather data is recommended. In order to be able to capture some of the less frequent climate events (e.g., extreme events) a longer observed records should be used.

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. Typical errors are when minimum temperature is greater than maximum temperature, or radiation values are too high. LARS-WG will ignore these errors (i.e., treat them as missing values) and they will not be included in the parameter calculation process.

TESTING PERFORMANCE

When the site parameters are calculated, LARS-WG automatically creates a tst-file with results of statistical tests, which assess the ability of LARS-WG to reproduce variety of weather statistics accurately. 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 the tst-file is described in Appendix 4.

CONTACT DETAILS

Dr Mikhail Semenov

Centre for Mathematical and Computational Biology
Rothamsted Research, Harpenden, Herts, AL5 2JQ UK
e: mikhail.semenov@bbsrc.ac.uk

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

Appendix 1: Structure of the wgx-file

The files contain following information for each month: semi-empirical distributions for length of dry and wet series, precipitation, minimum and maximum temperature and radiation calculated separately for dry and wet days (wet day is defined as a day with precipitation > 0) and correlation and auto-correlation coefficients.

A semi-empirical distribution is described by a block of 3 lines. The 1st line consists of: 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 with the 3rd line with cumulative probabilities and 3rd line values of the variable.

Appendix 2: 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 is also contains 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 (from Jan to Dec) 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 (from Jan to Dec). 1 line is maximum of monthly total precipitation, 2 line is minimum of monthly total precipitation, 3 line is the number of years, 4 line is monthly total means and 5 line is monthly total standard deviation.

[RAIN daily maxima: median, 95 percentile and maximum]
Each line contains 12 values (from Jan to Dec) of maxima of daily precipitation for each month. 1 line is median of maxima; 2 line is 95 percentile and 3 line is maximum of maxima.

Appendix 3: Structure of the st-file.

[SITE]
SiteName as a single word
[LAT, LON and ALT]
3 numbers: latitude Longitude and altitude (in meters) separated by space or tab
WEATHER FILES]
A list of names of weather files (one file name per line). You do not need specify a full path if all your files are in the same directory as corresponding st-files. The records in the files must be in chronological order starting with the earliest record.
[FORMAT]
A number of tags which describe the format of you data files. The following tags can be used: YEAR, MONTH, DAY – day of the month, JDAY – day of the year, MIN – minimum temperature, MAX – maximum temperature, RAIN – precipitation, RAD – radiation, SUN – sun hours.
[END]

If you wish to add any explanatory notes to the st-file, you can use "//" at the beginning of lines containing comments.

In your weather data file the values of variables for each day should be separated by spaces or tabs. The data should not contain blank lines, comments lines or headers. Missing data values should be coded -99.

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 month) 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 statistics and a corresponding p-value, which indicate how likely that generated and observed data are coming from the same distribution. If p-value is very low and below the significance level, set to 0.01 or 0.05, then the generated simulated climate is unlikely to be the same as the 'true' climate. You have to note, that the 0.05 significance value is a common significance level used in statistical tests, on average 1 in 20 tests will show significant result even when there is no difference in distributions.