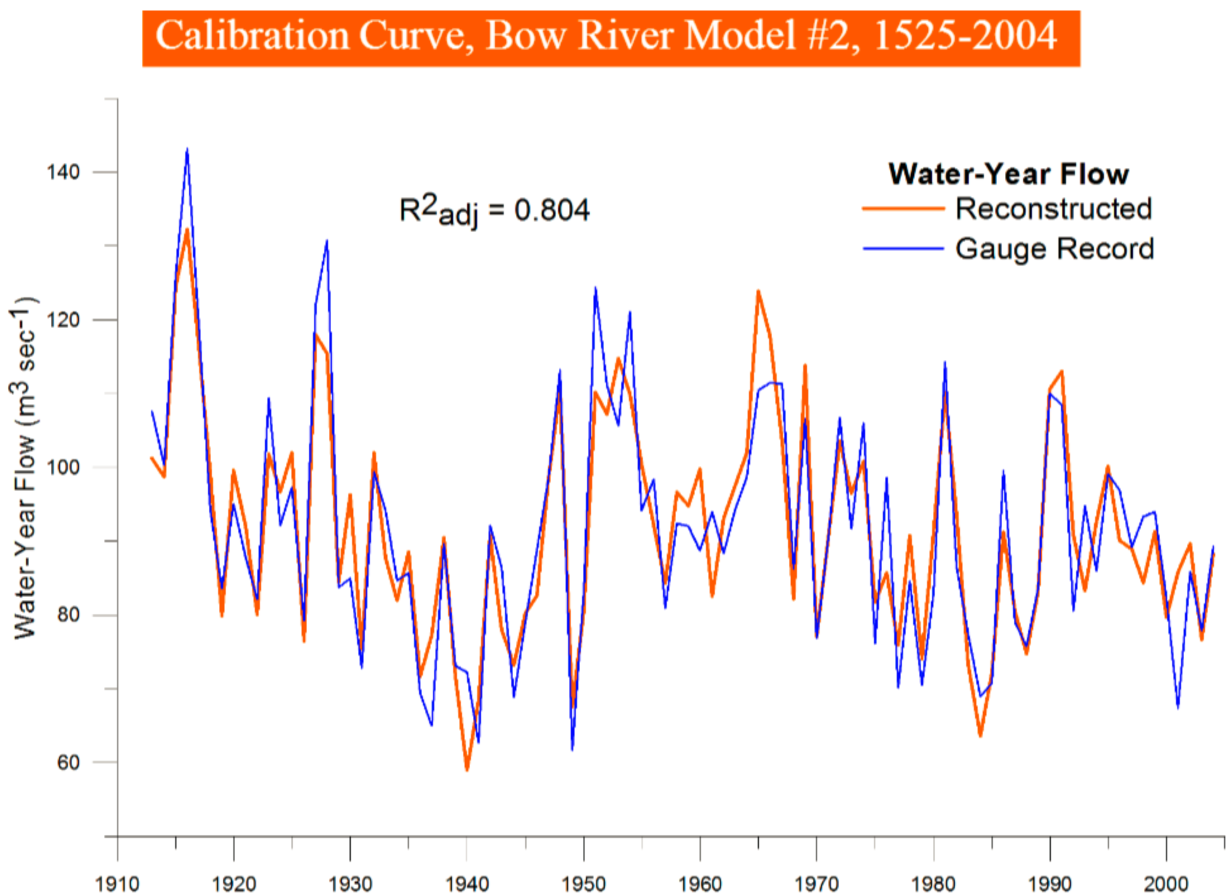
Stochastic Weekly Flow Generator with Paleoclimate Data Interface

1. Introduction

This application generates stochastic weekly flows which have all relevant statistics of historic natural weekly flows for a given site, while their annual means closely match the estimates obtained from tree rings data. In this way, historic flow estimates for several hundred years (depending on the age of the sampled tree) are linked to a proxy record, which is an advanced step compared to most stochastic models in hydrology that rely exclusively on the available historic flow record that is typically less than 100 years long.

As an example, the following figure shows the basis of developing a statistical relationship between the tree rings signal and the estimates of annual natural flows for the Bow River at Calgary. The blue line in the above graph shows historic annual natural flows obtained from naturalizing recorded flows using the project depletion method (i.e. adjustment of recorded flows to account for the effects of storage and diversions). The orange line shows the estimates of annual natural flows obtained using the tree rings data, which exhibit high correlation with the historic natural flows with a correlation coefficient of 0.804.

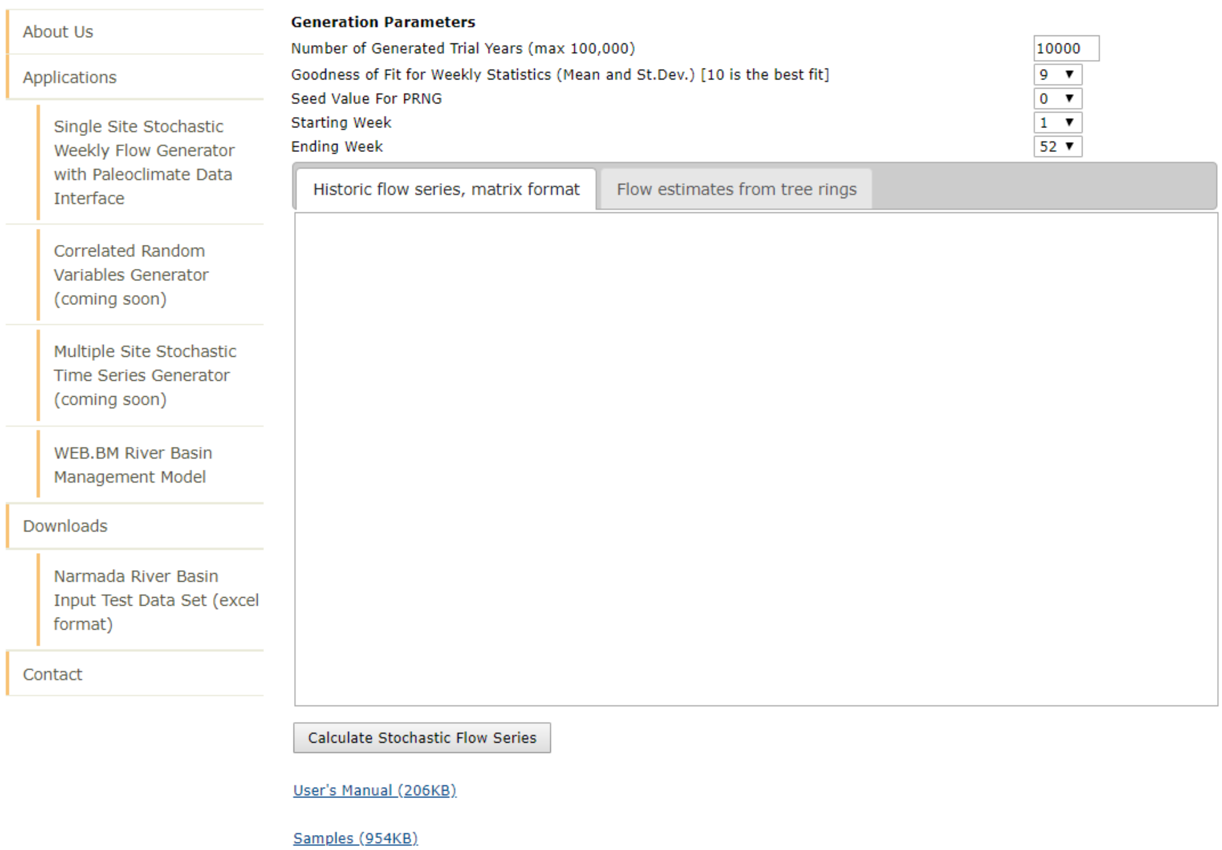


This correlation was used to develop estimates of annual flows on the Bow River from the 1111 to 1911 period when there were no historic flow records. Such annual flow estimates are then used as one of the input data files in this stochastic model. This input data file is provided to the web application by clicking on the appropriate tab in the entry form and by using the paste command to transfer data from the clipboard.

The other input is a time series of historic natural flows, which are typically monthly or weekly series. This file is provided in a tabular format, where columns represent weeks while each row contains data for a single historic year. The years should be provided in chronological (historic) sequence. The model does not handle files with missing data, hence all historic natural weekly flows should be provided for the available period of record.

2. Instructions for Use of Web Interface

Users can access www.optimal-solutions-ltd.com web site and select the first application option (Single Site Stochastic Weekly Flow Generator with Paleoclimate Data Interface). The model interface that will appear will look like this:



Number of Generated Trial Years is set to 10,000 by default. This is typically sufficient for most successful runs. The minimum should be higher than the length of the tree rings series. The maximum cannot exceed 100,000.

Goodness of Fit for Weekly Statistics determines how close the match between the historic and generated weekly flow statistics will be. The scale is from 1 to 10, with the 10 giving the closest match. Default is set to 9. This match refers to all generated data, and the final results may not be very sensitive to this parameter, since the parameter refers to the statistical match for all 10,000 generated data, while only a random fraction of that sample is selected for the final solution subject to the selection criteria.

Random Number Seed can provide alternative solutions. It is specified as a single integer value. The default is set to zero, and in this interface it can vary from 0 to 20.

The Starting and the Ending Week define a season which was used to generated the tree rings estimates. The default is set to 1 and 52, which defines a special case when the "season" is the same as the entire calendar year. This has been the only option in the past. With the latest update, the model can be run by using the tree rings data that were estimated based on the seasonal growth. Typically, in the Canadian Prairies, the warm season is assumed to begin on the 18th week (calculated as 7 day periods starting from January 1st of each year) and ending with the end of the 35th week, which has to be specified as the beginning of week 36 in the drop down box on the entry form, since the "beginning of the week" is taken as a convention that determines the exact time for start and end of a selected season. Similarly, the cold season begins at the start of week 49 and ends at the start of week 18. Users have to make sure they are using the appropriate seasonal flows estimates based on the tree rings data for each season, since the wet and dry seasonal flows can differ by up to an order of magnitude. It should be noted that the model prints out the weeks from week 1 to week 52 for any configuration of input data parameters, which makes it easier to study the sensitivity of the model results as a function of the selected seasons. However, the model is flexible to take any week as the beginning and the end of the season.

Users should copy historic weekly flows that were previously arranged in a matrix format with column 1 containing flows in week 1, column 2 containing flows in week 2, and so one until the 52nd column is filled with data. The columns can be delimited by tabs or blank fields. The data is entered by clicking on the blank field underneath the tab and right-clicking on the mouse button to invoke the paste command, which then allows users to paste the entire matrix into this form. Similarly, the other tab (Flow Estimates from Tree Rings) expects a single column of data, representing flow estimates based on the tree rings data.

<u>Users should make sure that the uploaded data originates from text files (i.e. it is not copied from excel)</u> and that the data are delimited by spaces or tabs. The historic flow data file has a tabular format without any missing data. The number of entries on each row should be the same, and it should be less or equal to 52. Typically, there will be no more than 100 years of historic weekly flows, although the model is capable of handling up to 200. The tree rings data file is a single column text file containing the seasonal (or annual of the season goes from week 1 to week 52) flow estimates that were derived using the regression fit with the historic flows for the period when both the tree rings data and historic flow data are available.

To execute the application, click on the *Calculate* button. The model will automatically download the results file in .zip format in the browser which was used to initially run the application. The output data file is also an ASCII file which can be imported to Excel spreadsheet for additional analyses. The final data series should have similar statistics (mean, standard deviation and probability distribution for each week) as the historic data series, and it should also have a similar correlation structure.

3. The Algorithm

Statistical fit of the weekly data was obtained by combining the empirical Kernel distribution which uses a combination of the moving average through the sorted historic data points and a small normalized random variate generated on the basis of standard deviation of localized points derived locally for various segments of the curve. Gumbel (extreme value) distribution is used to fit the tail ends of the statistical curve. A uniform random number [0,1] is used as a Probability axis to read the value of flows from the empirical curve. The model can generate up to 100,000 variables (although it is usually sufficient to generate only 10,000) for each individual week in this way. Following this step, the weekly variables use an algorithm that permutes them such that the target correlations for all weeks from 1 - 52 obtained from the historic series are induced among them. This is the second step of the algorithm, and it results in 52 weekly flows for each year which are mutually correlated in the same way as historic flows. There are several algorithms that had been developed to provide efficient permutation of variables generated in step 1 so as to induce desired correlation structure (Ilich, 2009 a), b)).

The final step involves selection from all generated years such that two criteria are met simultaneously:

a) The mean seasonal flow of the sampled year is close to a given mean seasonal flow estimate from the tree rings; and,
b) The selected year has correlations between the December weeks of the previously selected year which are on target with respect to the January weekly flows of the currently selected year.

The outcome of the model is a series of more than 900 years (or less if the sampled tree is younger) of generated weekly flows that are in the agreement with annual flows obtained from tree rings while all relevant weekly statistics are preserved.

3. Sample Input Data Files

Users can download three sample input data runs. They include the two input data files for each of the runs which were done on river basins in Alberta (The Bow River at Calgary, The Oldman River above the Oldman Dam, and the North Saskatchewan River near Edmonton). The first two files are produced by using the default parameters (10,000 generated years and the Goodness of Fit Statistic set to 5), and they provide an excellent fit in terms of expected mean annual flows based on tree rings. On the North Saskatchewan River the selected parameters are 90,000 generated years and the Goodness of Fit Statistic set to 7. There are some extremely dry years in the input data for North Sask. River which could not be reproduced by the weekly stochastic model. Stochastic flows follow weekly statistical curves which more or less guarantee similar weekly means as the historic means. Deviations from the mean which are larger than two standard deviations on the dry side are rare, and the probability that such deviations will find themselves in the final configuration for all 52 weeks are even less likely. For example, while the mean annual flow at North Sask. River is 215 m^3/s while the standard deviation is 45.3, the model still managed to produce some very dry years which had annual mean of around 130 m^3/s, roughly two standard deviations below the long term mean of 215 m^3/s. However, the tree rings data demand that the model create some years which are more than four standard deviations below the mean (i.e. the target annual mean is 30 m^3/s). Such low annual flows cannot be produced with a model that meets the statistical weekly distributions based on the last 100 years of historic flow data. Hence, the model picks the best available matches for the extreme years, recognizing that perfect match with such extremes is impossible with the current design of the algorithm. If desired, weekly values for a handful of extreme years which were not fitted perfectly can be adjusted manually such that the desired global statistics are still preserved, however this is not recommended practice.

4. References

David J. Sauchyn, Jennifer Stroich and Antoine Beriault. June 2003. A paleoclimatic context for the drought of 1999–2001 in the northern Great Plains of North America. DOI: 10.1111/1475-4959.05003

Jeannine-Marie St. Jacques, David J. Sauchyn and Yang Zhao. March 2010. Northern Rocky Mountain streamflow records: Global warming trends, human impacts or natural variability? DOI: 10.1029/2009GL042045View/save citation

Nesa Ilich. 2013. *An Effective Three-Step Algorithm for Multi-Site Generation of Weekly Stochastic Hydrologic Time Series*. Hydrological Sciences Journal, 59 (1 – 2) 2014. http:/dx/doi.org/10.1080/02626667.2013.822643

Nesa Ilich. 2009. *A Matching Algorithm for Generation of Correlated Random Variables with Arbitrary Distribution Functions.* European Journal of Operational Research, Vol.192 (2), pp. 468-478, doi:10.1016/j.ejor.2007.09.024.