

# **Weather Data Processor**

## **Version 2.01**

This tool allows the processing of timeseries of weather variables from NCEI Integrated Surface Datasets (ISD), Global Historical Climate Network (GHCN), COOP Hourly Precipitation, National Land Dataset Assimilation System (NLDAS), Global Land Data Assimilation System (GLDAS) datasets and the Tropical Rainfall Measuring Mission (TRMM). The tool uses webservices to query available stations and data. The processed weather timeseries are uploaded to a user selected WDM file and tool allows export to SQLite database.

WDM Files can be viewed and managed with the following tools:

### **BASINS Ver 4.5**

(<https://www.epa.gov/ceam/better-assessment-science-integrating-point-and-non-point-sources-basins>).

**SARA** (<https://www.respec.com/product/modeling-optimization/sara-timeseries-utility/>).

### **Missing data are estimated by the following methods:**

1. Use of data from nearby stations (spatial analysis using inverse distance weighting method)
2. Use of linear time series models (stochastic model)

### **Missing Rainfall Data**

Missing rainfall data are estimated using a two-state first order Markov chain to model if the current hour is dry or wet. The approach uses transitional probabilities of a dry hour following a dry ( $PD|D$ ) or wet hour ( $PD|W$ ) and a wet hour following a dry ( $PW|D$ ) or wet hour ( $PW|W$ ). Probabilities are computed by month. If the hour to be estimated is wet, the rainfall amount is estimated by a two-parameter gamma distribution. The distribution parameters are estimated by month. Approach is similar to the WGEN and CLIGEN weather generators (use of Markov chain for rainfall is well documented in the WGEN stochastic timeseries generator).

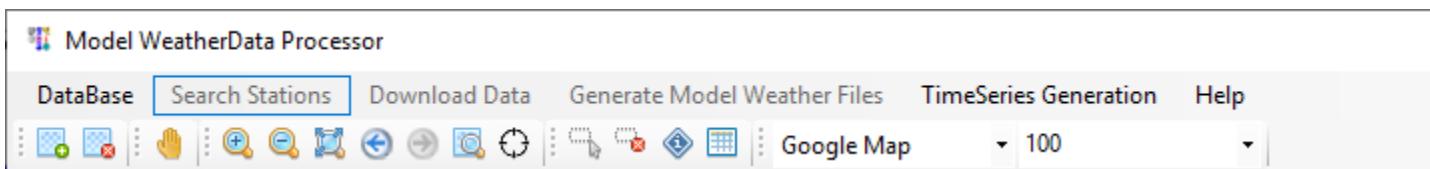
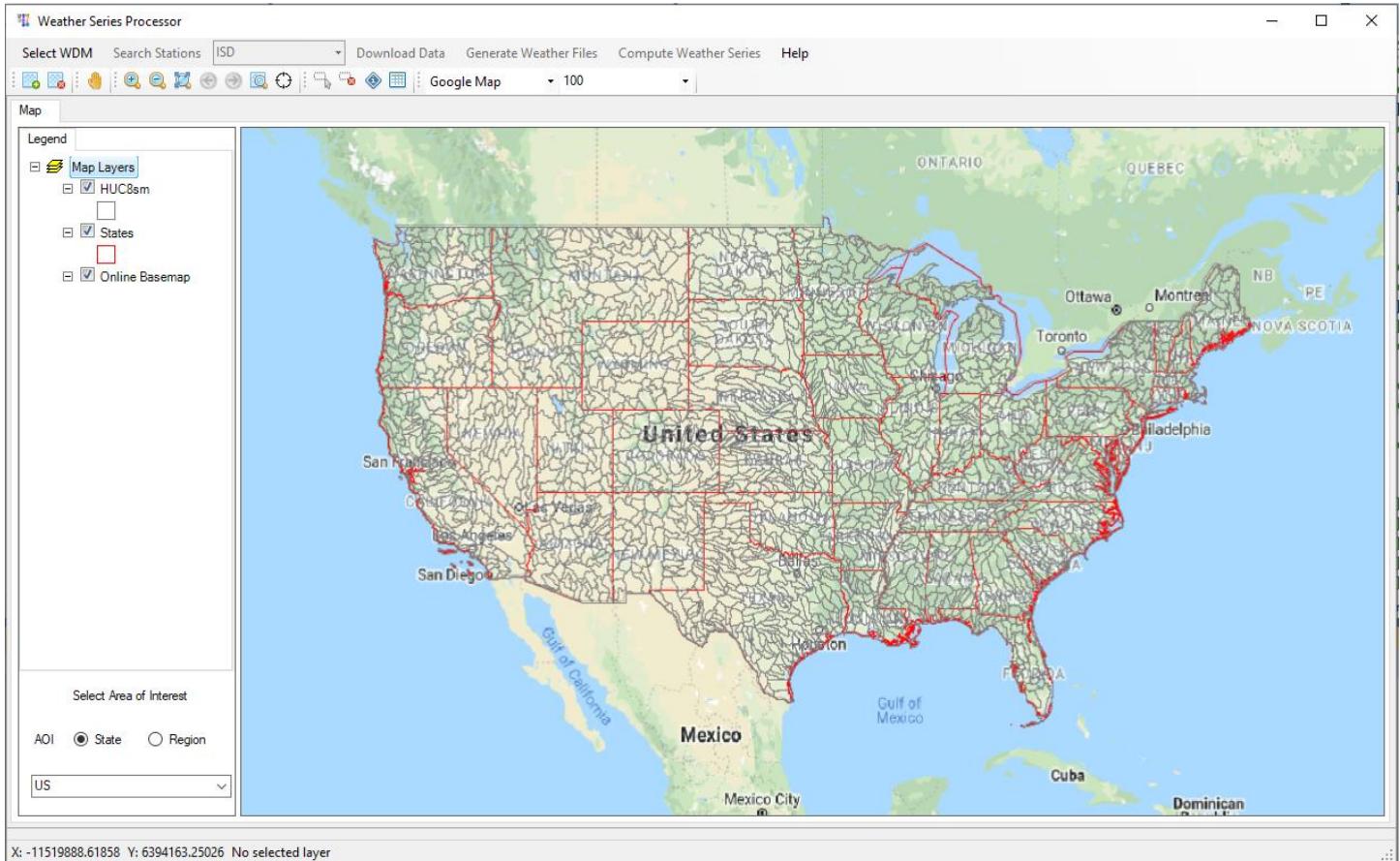
### **Missing Temperature, Dew Point, Cloud Cover and Wind**

Missing hourly weather data are estimated using a linear lag-1 autoregressive model (AR1). Hourly means and standard deviations of the weather series were calculated by month and fitted with a Fourier series. Seasonal parameters of the AR1 model were also calculated to parameterize the AR models for the weather variable. The workflow to estimate a missing data, is as follows: (see WGEN for application for daily series).

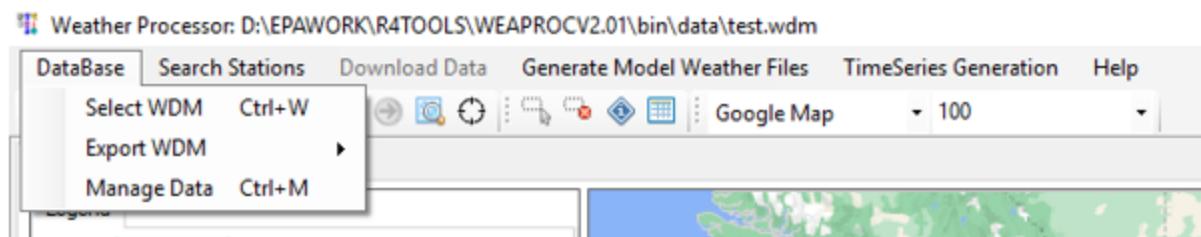
1. Standardize raw series with the periodic means and standard deviation.
2. Calculate the lag-1 serial correlation of the standardized series and the corresponding moments; the lag-1 serial correlation is an estimate of the parameter of the lag-1 AR model.
3. Fit a lag-1 AR model and use model to estimate missing hourly data.

Results of model fit and validation (using 50% random sample of the series) are saved in the **Data** folder.

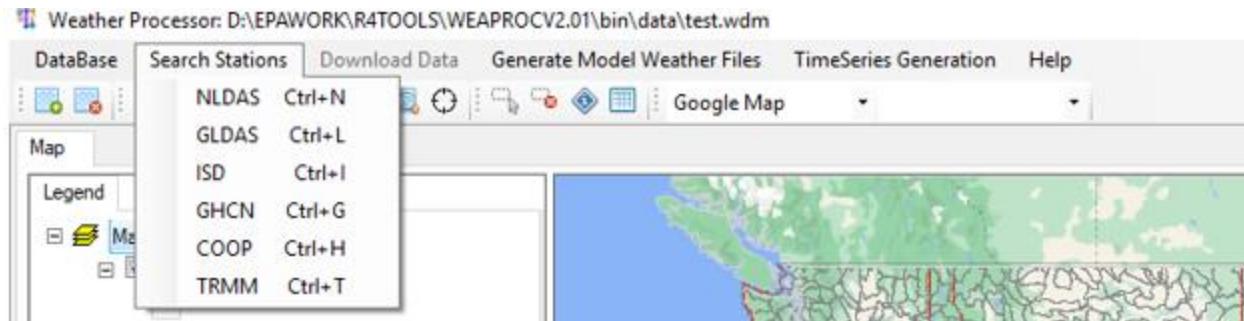
## HOW-TO



**Database:** Click to select a WDM file where downloaded and processed series are saved; export WDM to SQLite database or Managed WDM database using the SARA Time Series Utility (available from <https://www.respec.com/product/modeling-optimization/sara-timeseries-utility/>)



**Search Stations:** Click to draw an area of interest (AOI) on the map and downloads stations for the AOI. User needs to select the data source in the combo box. Options for data source are ISD, GHCN, COOP and NLDAS.



**Download Data:** Click to download data for the selected stations. The download screen is shown where users specifies the date range, variables to download, minimum years of record and threshold percent missing. If percent of missing record for downloaded series is greater than the specified threshold, the series is ignored for further processing. The date range should be more than the specified minimum years. If downloading NLDAS, GLDAS and TRMM series, the series are automatically converted from UTC time to local time.

Click on **Download** to download the selected dataseries.

Download Options

Specify range of dates for series and variables to download.

From: 12/31/2009 To: 12/31/2019 Automatic UTC shift, see WorldTZ.shp

Variables:

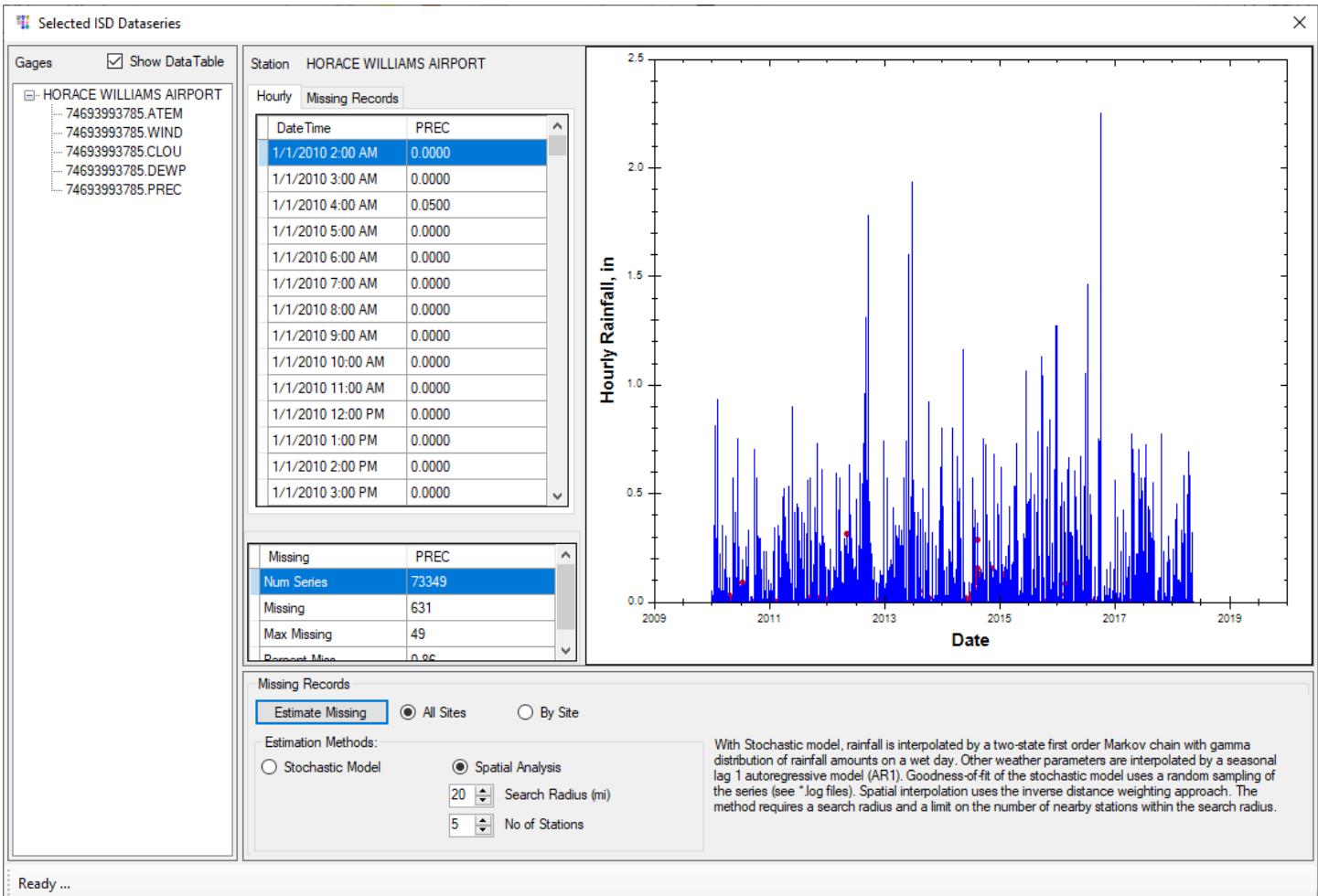
Rain    Temp    Wind    Solar    DewPt

Minimum Years of Record: 5 Timeseries with less than the specified minimum years of record are ignored.

Ignore timeseries with 50 percent of hourly/daily data are missing.

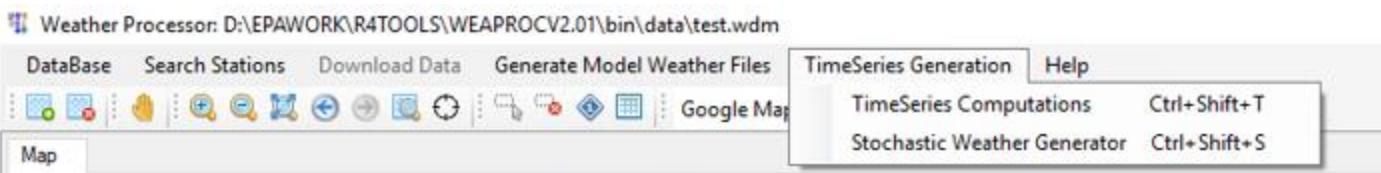
Model DataBase: D:\EPAWORK\R4TOOLS\WEAPROCV2.01\bin\data\test.mdl

After all the series are downloaded, the data screen is shown where users can view a table and plots of the data series. The screen also allows estimation of missing records using either a stochastic model or spatial interpolation using nearby stations (*Missing records of NLDAS, GLDAS and TRMM are not estimated*)



**TimeSeries Generation:** Click on *TimeSeries Computations* to estimate timeseries for potential evapotranspiration (PEVT) using Hamon or Penman methods and to compute Solar Radiation from Cloud Cover. The selected WDM file is read and the various series in the WDM are shown in the data table.

Click on Stochastic Weather Generator to generate a timeseries using a Markov chain model for rainfall or a



linear AR model for other weather variables.

## TimeSeries Computations

Weather Series Generation X

Generate Series

Routine calculates PEVT using Hamon and Penman method from input timeseries in a WDM.  
Also calculates solar radiation from cloud cover.

98 series In D:\EPAWORK\R4TOOLS\WeaProcV6\bin\data\test.wdm

DSN	Location	Scenario	Constituent	StartDate	EndDate
1	X303Y089	NLDAS	PREC	12/31/2009 1:00:00 AM	12/30/2019 1:00:00 AM
2	X303Y089	NLDAS	ATEM	12/31/2009 1:00:00 AM	12/30/2019 1:00:00 AM
3	X303Y089	NLDAS	WINDV	12/31/2009 1:00:00 AM	12/30/2019 1:00:00 AM
4	X303Y089	NLDAS	WINDU	12/31/2009 1:00:00 AM	12/30/2019 1:00:00 AM
5	X303Y089	NLDAS	WIND	12/31/2009 1:00:00 AM	12/30/2019 1:00:00 AM
6	X303Y089	NLDAS	SOLR	12/31/2009 1:00:00 AM	12/30/2019 1:00:00 AM
7	X303Y089	NLDAS	CLOU	12/31/2009 1:00:00 AM	12/30/2019 1:00:00 AM
8	X303Y089	NLDAS	DEWP	12/31/2009 1:00:00 AM	12/30/2019 1:00:00 AM
9	X303Y090	NLDAS	PREC	12/31/2009 1:00:00 AM	12/30/2019 1:00:00 AM
10	X303Y090	NLDAS	ATEM	12/31/2009 1:00:00 AM	12/30/2019 1:00:00 AM

Select All Select Clear Selection Clear Station(s)

## Stochastic Weather Generator

Stochastic Weather Generator X

Please select from list of station-parameter to generate stochastic series at the selected site. Need to specify output SQLite database.

Generate Series      From:  To:

Parameter Database: D:\EPAWORK\R4TOOLS\WEAPROCV2.01\bin\data\test.mdl

Output Database: D:\EPAWORK\R4TOOLS\WEAPROCV2.01\bin\data\test.sdb

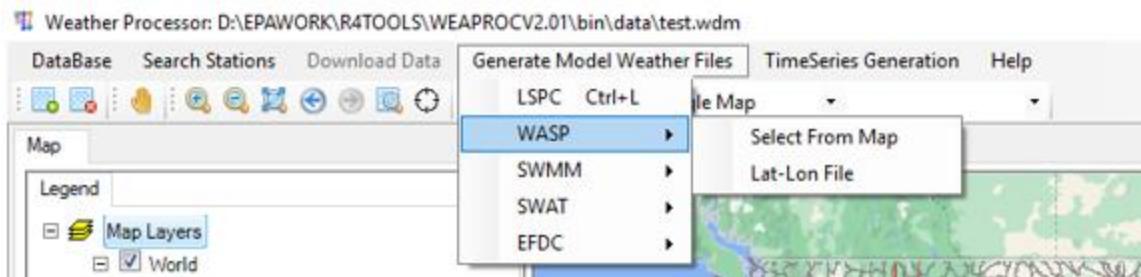
Station_ID	MetVar	Station_Name
G0404384	ATEM	G0404384
G0404384	ATMP	G0404384
G0404384	DEWP	G0404384
G0404384	LRAD	G0404384
G0404384	PREC	G0404384
G0404384	SOLR	G0404384

Select All Clear Selection

Export Generated Stochastic TimeSeries to csv files.

Ready ...

**Generate Model Weather Files:** Click to generate model specific weather files. Weather files can be generated for the following models: LSPC, WASP, SWMM, SWAT and EFDC. To generate a weather file, the user can select a location on the map or provide a lat-lon file (\*.csv) which contains coordinates of points. The LSPC model requires a sub-basin shapefile.



After selecting from map or reading points from a lat-lon file, the form below (e.g. for SWAT model) is shown where the user can set the simulation period, assign stations and write the model specific weather files.

SWAT Weather-test.wdm

Routine writes weather files for SWAT Model. User inputs a simulation period but tool adjust depending on the common period of the weather series in the selected WDM file.

Save SWAT Files in: D:\EPAWORK\R4TOOLS\WEAPROCV7\bin\data

Simulation Period	Common Period
Write SWAT Weather File(s)	Begin Date: 5/1/2013 End Date: 12/30/2019
Begin Date: 5/1/2013	End Date: 12/30/2019

SWAT Weather Locations (3 stations)

Location	PREC	PEVT	ATEM	SOLR	WIND	DEWP	CLOU
1	X302Y080:348	72103100:2	72103100:21	72103100:125	72103100:22	72103100:24	72103100:23
2	X302Y080:348	72103100:2	72103100:21	72103100:125	72103100:22	72103100:24	72103100:23
3	X302Y079:339	72097400:1	72097400:16	72097400:117	72097400:17	72097400:19	72097400:18

Ready ... Close

#### Sample lat-lon File Format

```
LOCATION,X,Y
1,-87.62651,35.03136
2,-86.57257,35.22081
3,-86.44517,34.68923
4,-87.36013,34.52718
5,-86.66523,34.34569
```

## **References:**

<https://disc.gsfc.nasa.gov/information/tools?title=Hydrology%20Data%20Rods>

Nicks, A.D. and J.F. Harp. 1980. Stochastic generation of temperature and solar radiation data. Journal of Hydrology 48:1-17.(see also CLIGEN documentation in BASINS 4.5)

Richardson, C.W. 1981. Stochastic simulation of daily precipitation, temperature and solar radiation. Water Resources Research 17(1):182-190. (see also WGEN documentation).

## **Problems/Bugs/Comments**

Email [fernandez.glen@epa.gov](mailto:fernandez.glen@epa.gov) for comments, bugs and suggested improvements.

## **DISCLAIMER**

This software is provided by USEPA Region 4 ‘as is’ and any express or implied warranties, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose are disclaimed. In no event shall USEPA Region 4 and the developers be liable for any direct, indirect, incidental, special, exemplary, or consequential damages (including, but not limited to, procurement of substitute goods or services; loss of use, data, or profits; or business interruption) however caused and on any theory of liability, whether in contract, strict liability, or tort (including negligence or otherwise) arising in any way out of the use of this software, even if advised of the possibility of such damage.

Application is powered by DotSpatial (<https://github.com/DotSpatial/DotSpatial>) and BASINS 4.5 WDM Utilities.