KNN-WG Tutorial

The KNN-WG tool, or K-Nearest Neighbor Weather Generator, is a software tool used in the field of meteorology and hydrology. It is designed to generate synthetic weather data for specific locations or regions based on historical weather data. The tool utilizes the concept of K-Nearest Neighbors (KNN) to simulate weather variables for future time periods by identifying similar historical weather patterns.

1- Installation and Registration

You can obtain the installer by following this link. Once you've clicked on the installer, the installation process will be straightforward, and the tool will be automatically installed on your computer. After installation, you can easily access the tool either by clicking on the desktop shortcut or by searching for "KNN-WG" in your computer's program list.

When you purchase a license through our online store, FastSpring will promptly send you a registration key via email. To activate the tool, please follow these steps:

1. Open the tool by launching it from your desktop or the installation directory.

2. Once the tool is running, click on the key icon on the topbar. This action will trigger the registration window to appear.

3. In the registration window, you'll need to input your registration key. Make sure to enter the key accurately.

4. After entering the key, proceed by clicking the "Register" button.

5. The tool will then validate your registration key, and if it's valid, the registration process will be successfully completed.

6. A confirmation message will be displayed, indicating that your registration was successful.

Should you encounter any difficulties during the registration process or if you have any inquiries regarding your registration key, please feel free to reach out to our support team for assistance.

2. Input Data

This tool operates by assessing the similarity of a given day's weather conditions to those of previous days. Therefore, it requires access to daily historical weather data for a specific location, a single station, or an average of data from multiple stations. It is advisable to have a dataset that includes variables such as Rainfall, Maximum Temperature (Tmax), Minimum Temperature (Tmin), Evapotranspiration (ET0), Solar Radiation (Srad),

Humidity, and Wind Speed (Windspd). However, you can also consider incorporating other pertinent meteorological variables into your dataset for a more comprehensive analysis.

Please note that the variable names in the tool are for informational purposes, and you can use any variables of your choice. For instance, you can designate Tmin as variable1, Tmax as variable2, and so on, according to your specific dataset and preferences.

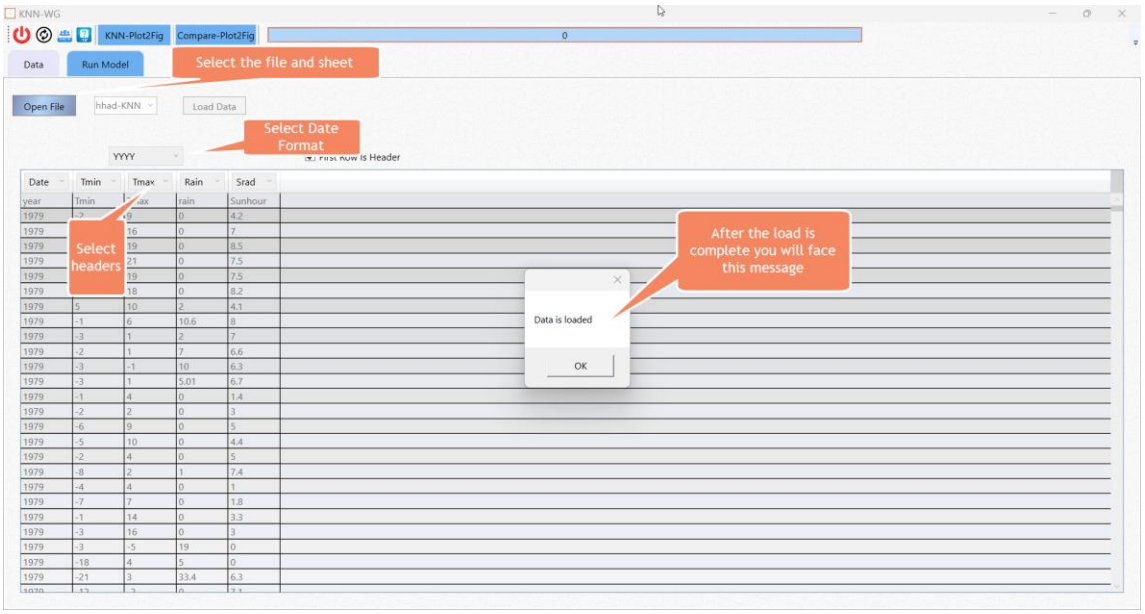

Fig1 – Input Data

If you have historical data spanning from 1980 to 2020, it's essential to ensure that your data columns in the Excel file are sorted chronologically from January 1, 1980, to December 31, 2020. Please note that incomplete years are not supported. Additionally, your data should include a date

column with a specific format. It is recommended to use the YYYY format for dates, and you can easily convert any date in Excel to this format by utilizing the Year function.

Loading input data in the first tab is a straightforward process. You can select your file and sheet, specify the date format, and identify headers. Afterward, when you click the "Load Data" button, a message box will appear with the message "Data is Loaded" to confirm successful data loading.

3. Run Model

You can effortlessly choose both the base period and the future period. The base period corresponds to your historical data, while the future period is the timeframe for which you intend to generate data. The calibration period, which automatically aligns with the overlap between these two periods, will be selected by the tool.

You should opt for a temporal window, usually set at 14 days, which covers one week before and one week after the current day.

In this tool, you can run the model multiple times and then ensemble the results. The ensemble method is described in the "Formula.PDF" file.

You should select a set of variables that you loaded and then click on the "Generate" button to initiate the data generation process. The progress bar in the top bar will activate for the number of runs.

The calibration period in KNN-WG (K-Nearest Neighbor Weather Generator) is the timeframe within your historical weather data that you use to train and calibrate the model. During this period, the model learns the relationships between various weather variables and their patterns. It allows the model to understand how weather conditions change over time and how they relate to each other. This knowledge is then used to generate synthetic weather data for the future.

In simpler terms, the calibration period is like a training period for the KNN-WG model, where it learns from historical data before it can accurately simulate or generate data for future periods.

You can easily create linear or box plots for any variable in the generated data, either on a daily or monthly scale. Additionally, you have the option to overlay historical data on the graphs for comparison. A button in the top bar allows you to save the generated graphs to a file for your convenience.

In the Efficiency Criteria section, you can calculate any of the six indices for either daily or monthly data. These indices comprise Root Mean Square Error

(RMSE), Mean Bias Error (MBE), Pearson index, Nash-Sutcliffe Efficiency (NSE), Spearman index, and Index of Agreement (d).

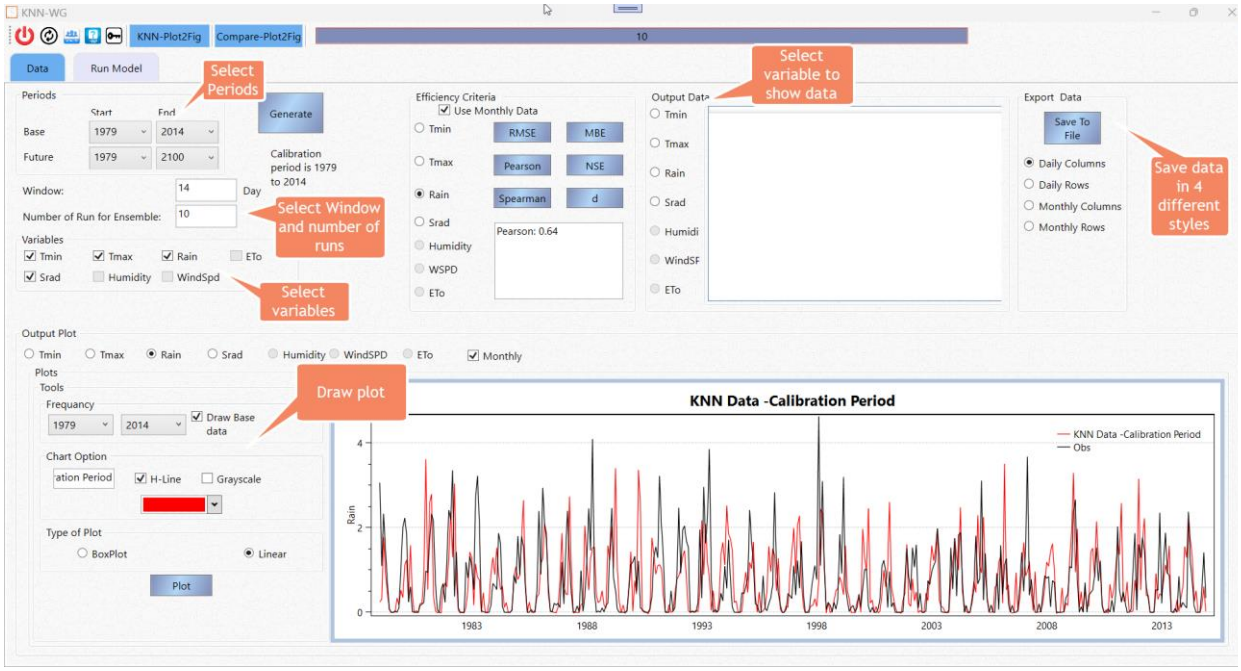

Fig2 – Run Model

You can view the generated data for any variable by clicking on its corresponding radio button in the Output Data section. Additionally, you have the option to save the data to Excel in either monthly or daily formats, and you can choose to arrange it in rows or columns.