

✓ *What is the basis of KNN method for implementation?*

To simulate weather variables for a new day ( $t+1$ ), days with similar characteristics as those simulated for day  $t$  are selected from the historical record. One of these nearest neighbors is then selected according to a defined probability distribution or kernel and the observed values for the day subsequent to that nearest neighbor are adopted as the simulated values for day  $t+1$  (Sharif et al., 2007). In this software following steps were followed (For further details, refer to Sharif and Burn, 2006):

Step 1: Compute regional means of the aim variables across the  $S$  stations for each day of the historical record.

Step 2: According to Yates et al. (2003) we should use a temporal window of 14 days that the window are considered as potential candidates to the current feature vector.

Step 3: Compute mean vector of the station for each day.

Step 4: Compute the covariance matrix,  $C_t$  for the current day  $t$  using the data block of size  $L \times p$ .

Step 5: Specify the number of first  $K$  nearest neighbors. According to the study of Yates et al. (2003), the use of a heuristic method for choosing  $K$  according to which  $K = \sqrt{L}$

Step 6: Select one of the nearest neighbors to represent the weather for day  $t+1$  of the simulation period.

Step 7: Compute Mahalanobis distances.

Step 8: Sort the Mahalanobis distances in ascending order and retain.

Step 9: Determine the nearest neighbor of the current day by using the cumulative probability metric.

Step 10: For each station and each variable, a nonparametric distribution is fitted to the  $K$  nearest neighbor's identified in step 8.

Steps 6–10 are repeated to generate as many years of synthetic data as required.

✓ *What is the basis of ensemble?*

In KNN-WG Version 1.1 we've added ensemble of N times run. In this method we calculate weight of each run and multiply this weight to each run.

$$V^{ensemble} = \sum_{i=1}^N w_i \times V_i^{knn}$$

$$w_i = \frac{1/\Delta V_i}{\sum_{i=1}^N 1/\Delta V_i}$$

$$\Delta V_i = |\bar{V}_i^{obs} - \bar{V}_i^{knn}|$$

N: times of run

i: number of run

$V^{ensemble}$ : Final values of variable in ensemble

$V_i^{knn}$ : Values of KNN variable in i<sup>th</sup> run

$w_i$ : Weight of i<sup>th</sup> run

$\bar{V}_i^{obs}$ : Mean of observation variable in evaluation period

$\bar{V}_i^{knn}$ : Mean of KNN variable in evaluation period

**How can we run KNN-WG Tool?**

### **Step 1. Input Data**

In this software application, we have 4 tabs. In the first tab “Data”, the user can load the Excel file in his/her system. Fig. 1 show this tab. In this step by clicking on the “Open File”, the Excel file can be indicated.

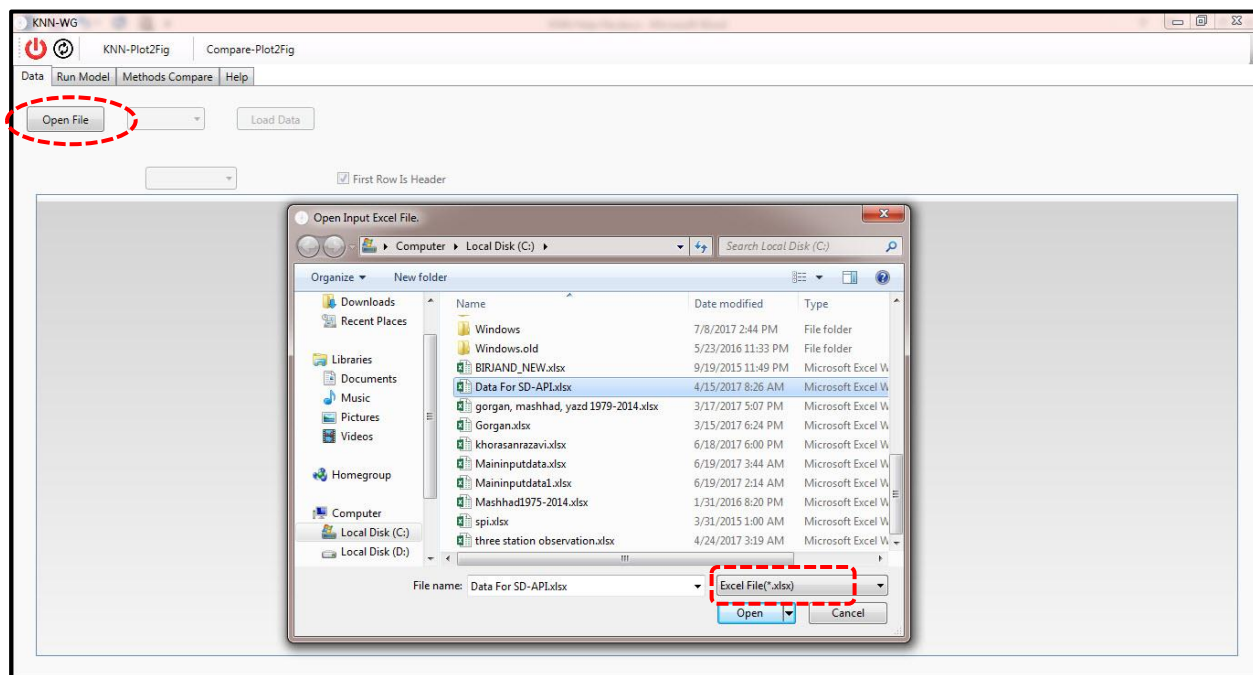


Figure 1

By selecting the Input Excel file, then the user can assign the intent sheet. The Fig. 2 show that the sheets in the input file. After selecting the aim sheet, the user can observe all the content of the excel file (Fig. 3).

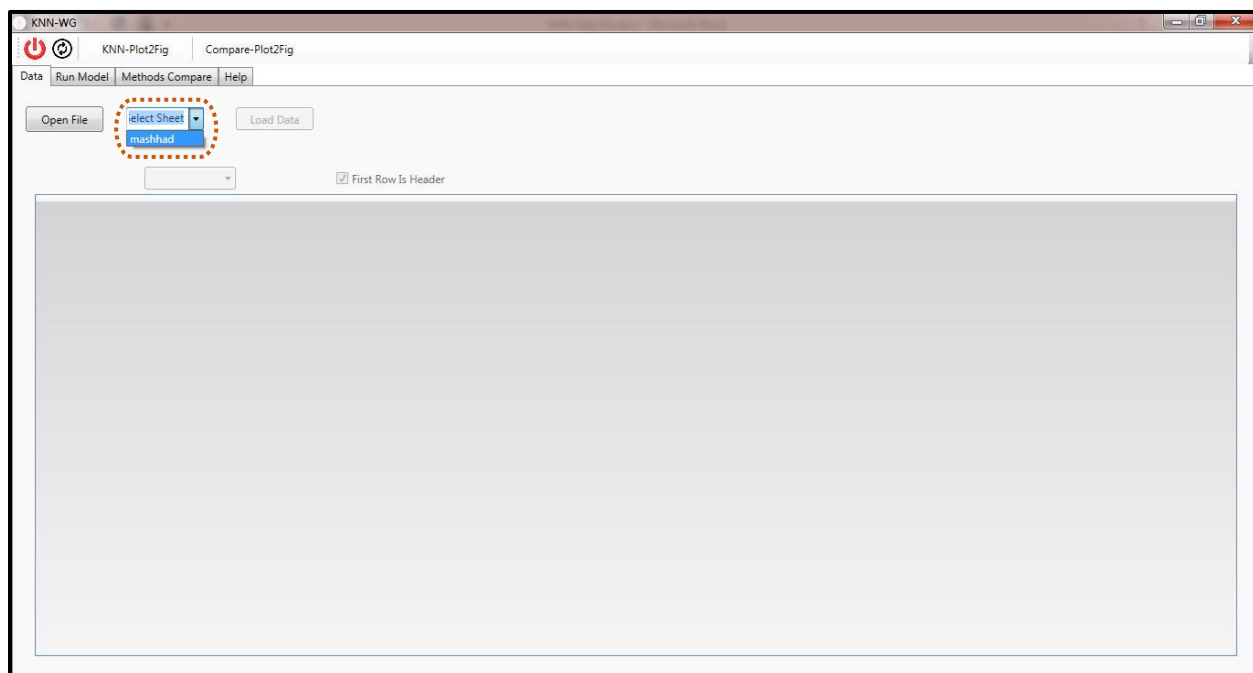


Figure 2

First point: If the input Excel file has header, check the “First Row Is Header”.

Second point: Note to the format of Date, if the input date has just the year with 4 digits (Foe example: 1989), select “YYYY”. If the format of the year like 99 instead of 1999, then select the format of “yy” (Fig. 4 and Fig. 5).

Third point: When the user wants to assign the name of each column, if the user made mistake to select the right one, first he/she select the none data for the header then select the proper name of variable of that column. For example, the user select “Rain” instead of “Tmin”, so he/she should to correct this fault, therefore, first the user should select “none” for that column and then assign the “Tmin” to that column’s header.

This tool can modify all empty cells or non-numeric cells and substitute it with mean of data up to that day.

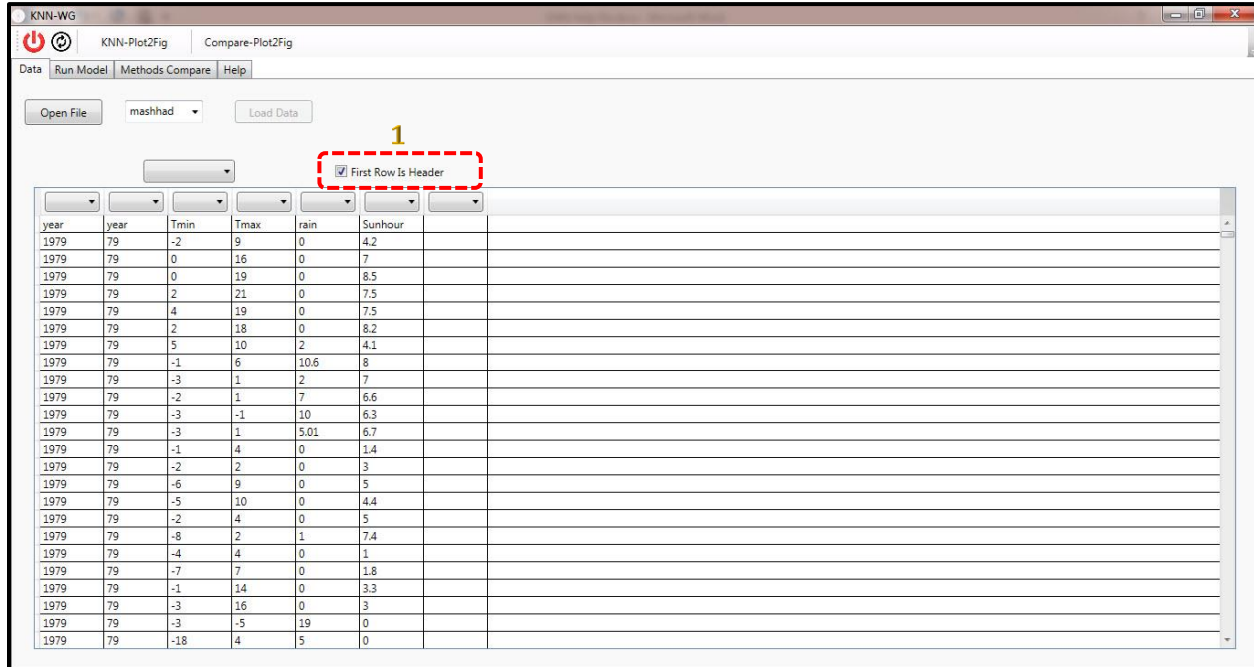


Figure 3

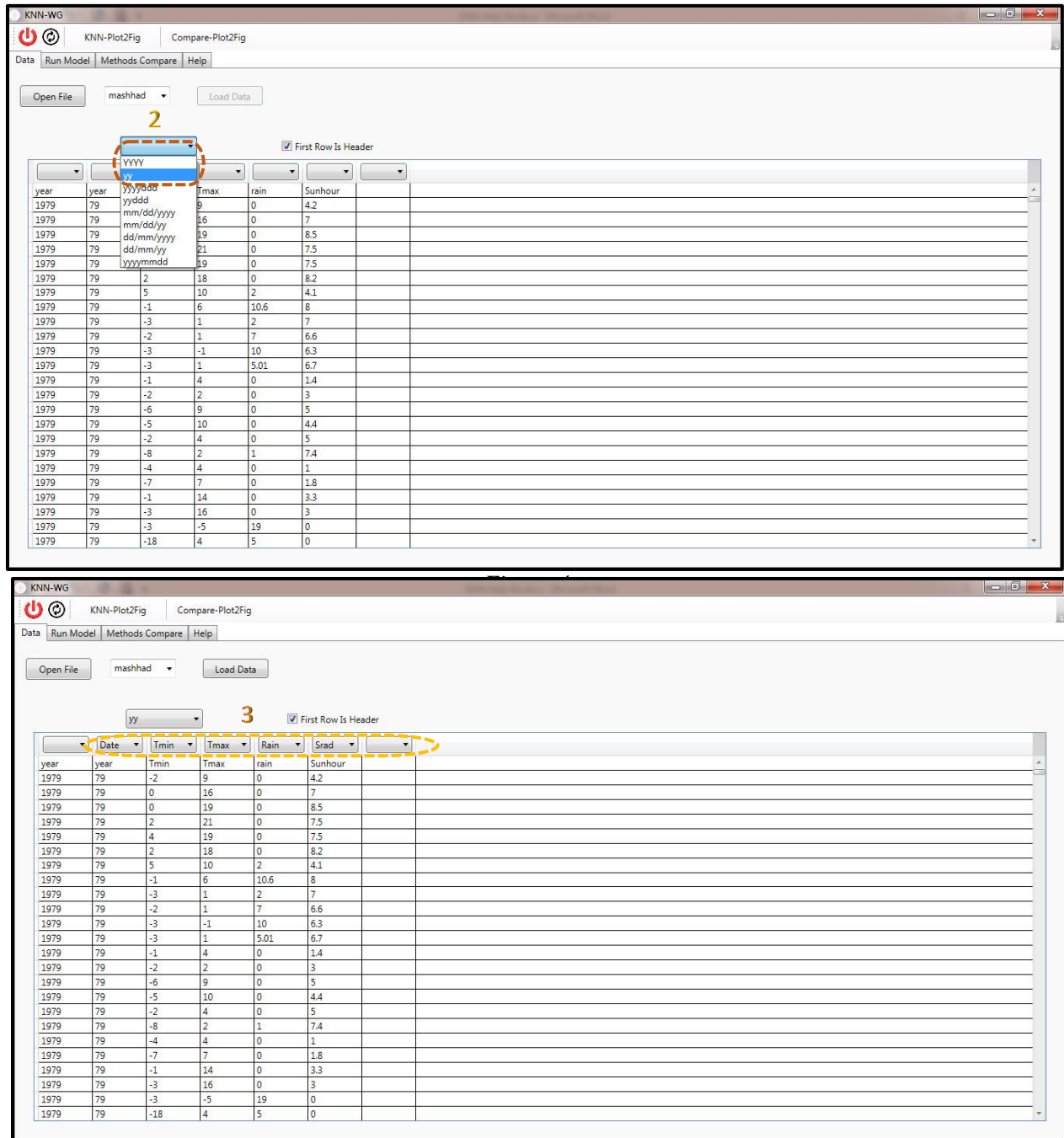


Figure 5

After all variables, has been assigned to the correct header, then click on the “Load Data” button. By clicking this button, if all things is OK, then the user can see the message “Data loaded” (Fig. 6).

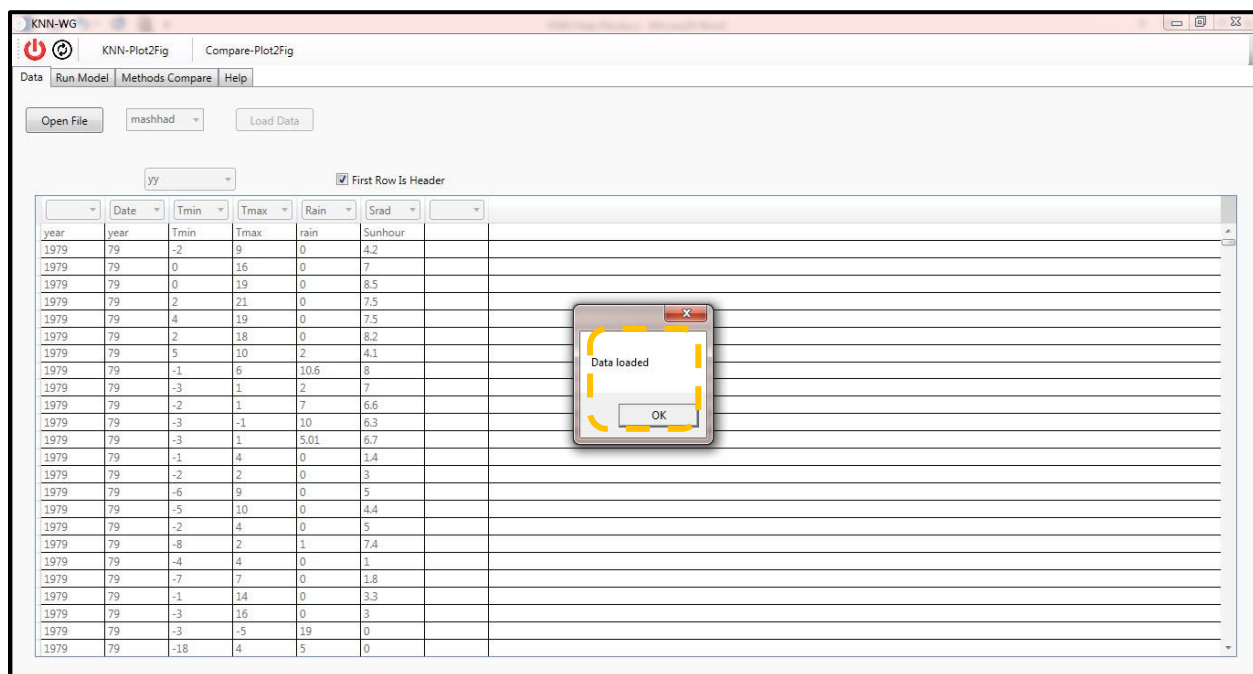


Figure 6

## **Step 2. Run Model**

In this step and in the “Run Model” tab, the user should assign the Base period. In this tab firstly, set favorite future period and base period (the proper base period should be more than 30 years). You can see this step in Fig. 7. For the future period, we need a common base with observation (base) period to simulate data. Whatever the size of common period is longer, it will be better for simulation and decrease the uncertainty of the generation data. The user can select future data up to 2100. In this example, we select 2030. So, the common base will be 1980-2014 (Fig. 8).

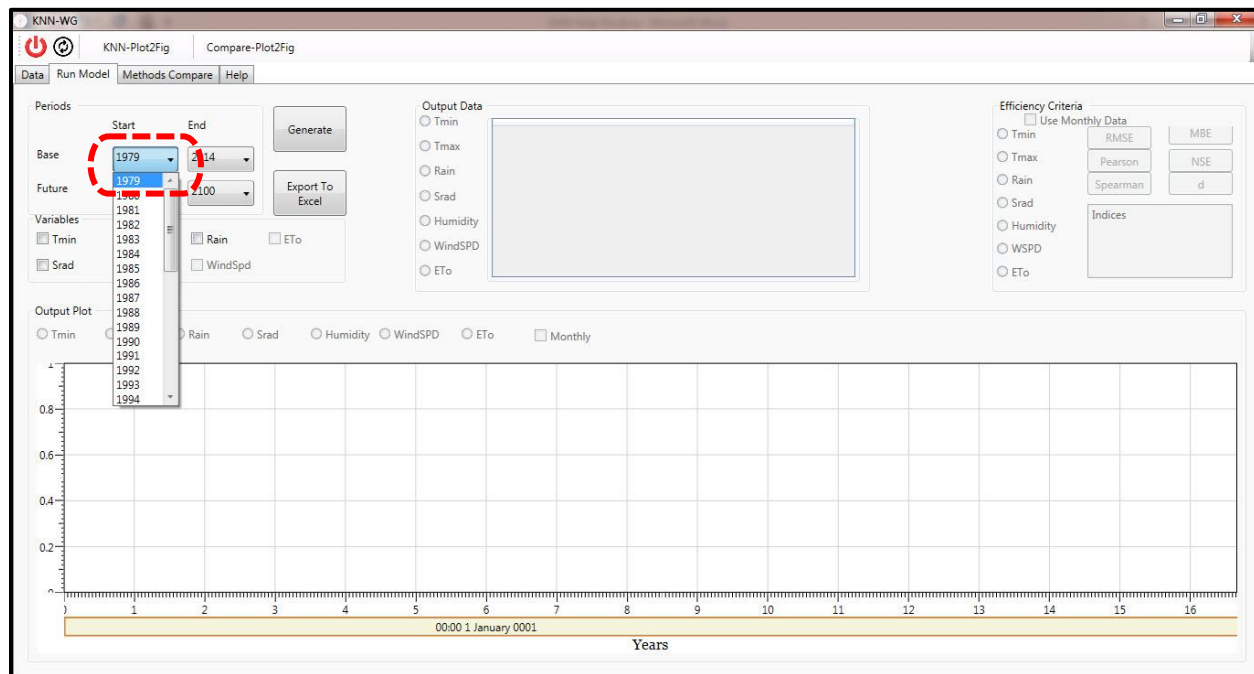


Figure 7

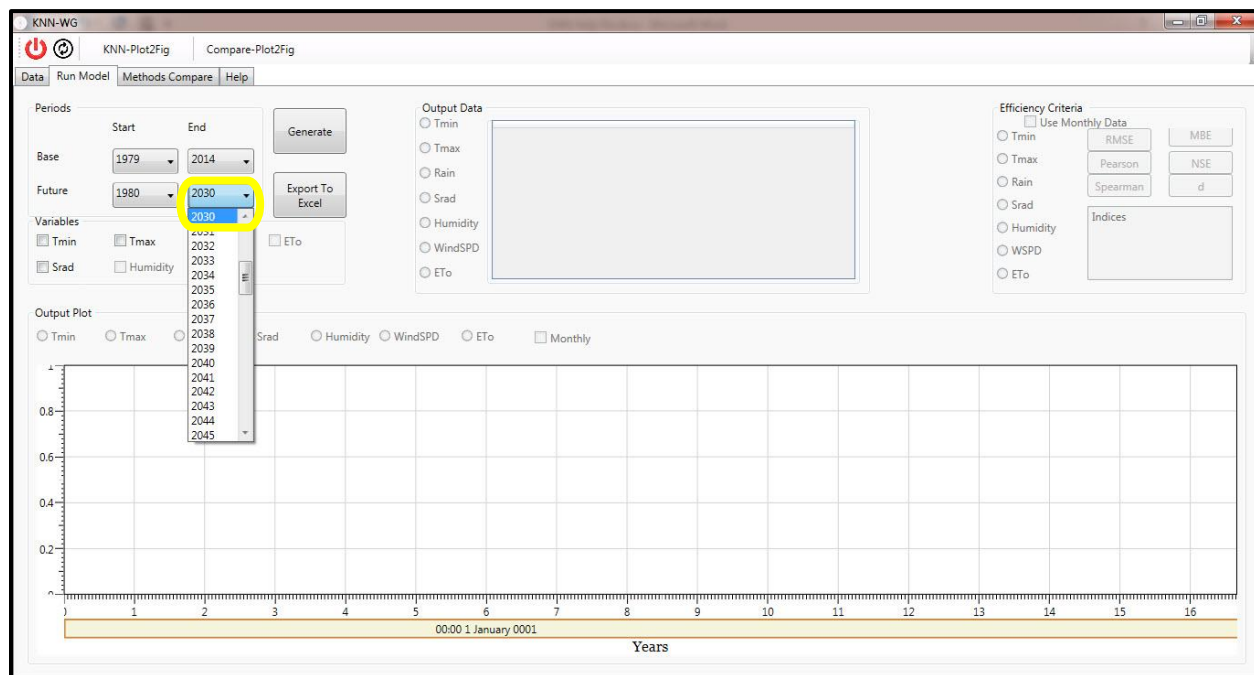


Figure 8

After the user select both periods (the observation and future period), then the user can check the intent variables from the available radio buttons. Then select arbitrary set of variables for generate data. The user can select at least three variables up to 7

of them (Fig. 9, Num. 1). With this variables and their similarity with the select window days, KNN can generate future data. Click on the “Generate” button (Fig. 9, Num. 2). If all things and selections are correct, the user face to the message of “Generate Complete” (Fig. 9, Num. 3).

When the data has generated with KNN-WG, the user can see outputs for each variable and observe the corresponding graph of every variable. According to Fig. 10, you can see the plot of Tmin as an example, the common base is started from 1980-2014, and 2015-2030 is the future generate data. The user can export the output data to an Excel file, by clicking the “Export to Excel” button.

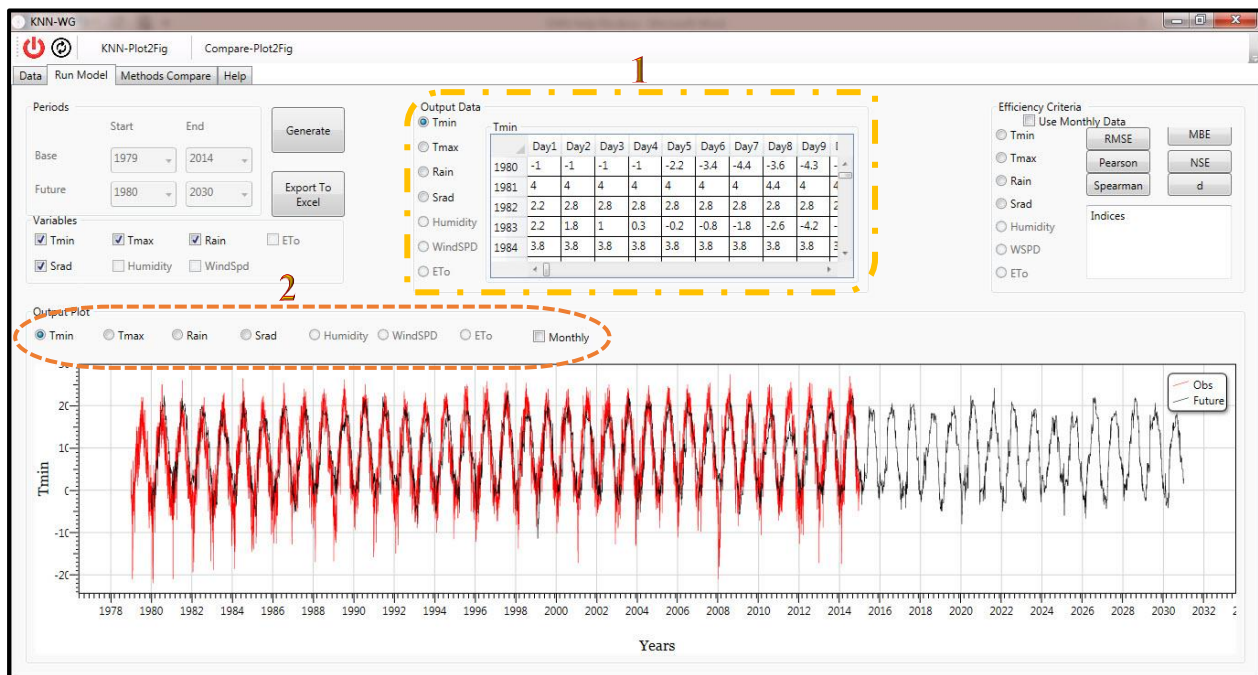


Figure 10

The user can observe the monthly plot of a generated variable, by clicking on the “Monthly” option, the user can see the monthly plot. See the Fig. 11. By selecting monthly option, the monthly generate data has plotted by KNN-WG.



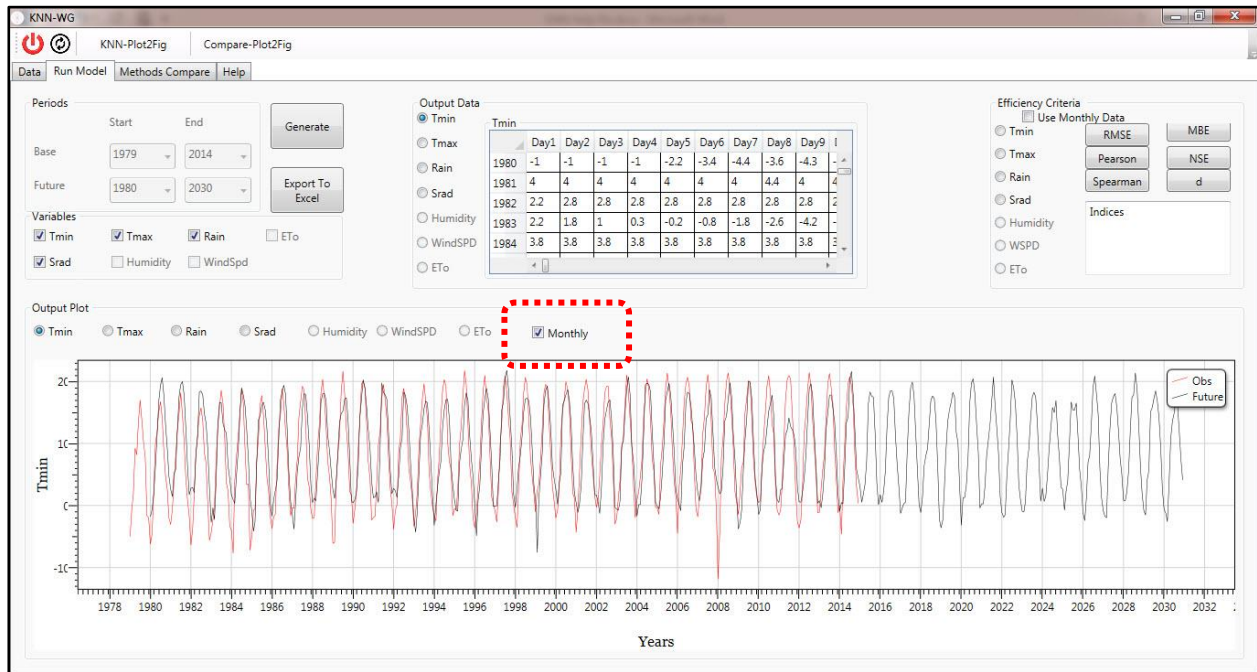


Figure 11

For assessing the generate data, KNN-WG has a section for calculating “Efficiency Criteria”. In this section, the user can achieve six important criteria that it can be calculated in daily or monthly time scale. In Fig. 12, by selecting “Tmin”, and Pearson, the user can observe the Pearson correlation value of Tmin, between observation and generation data.

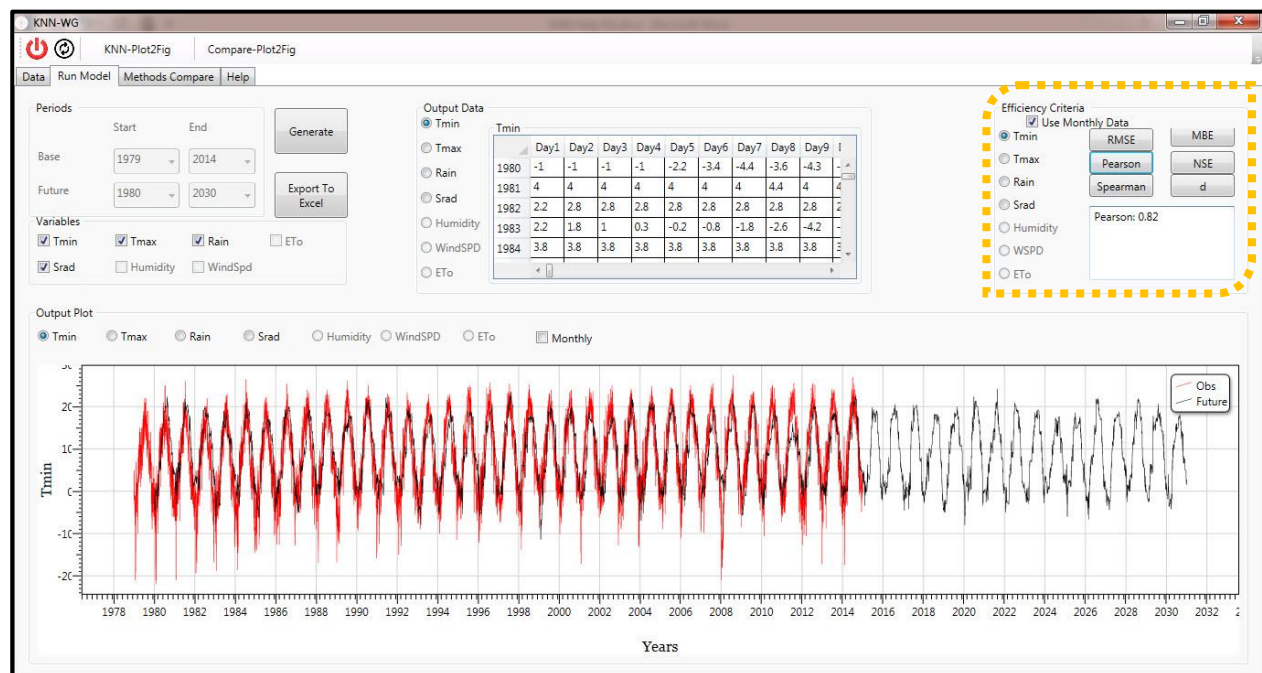
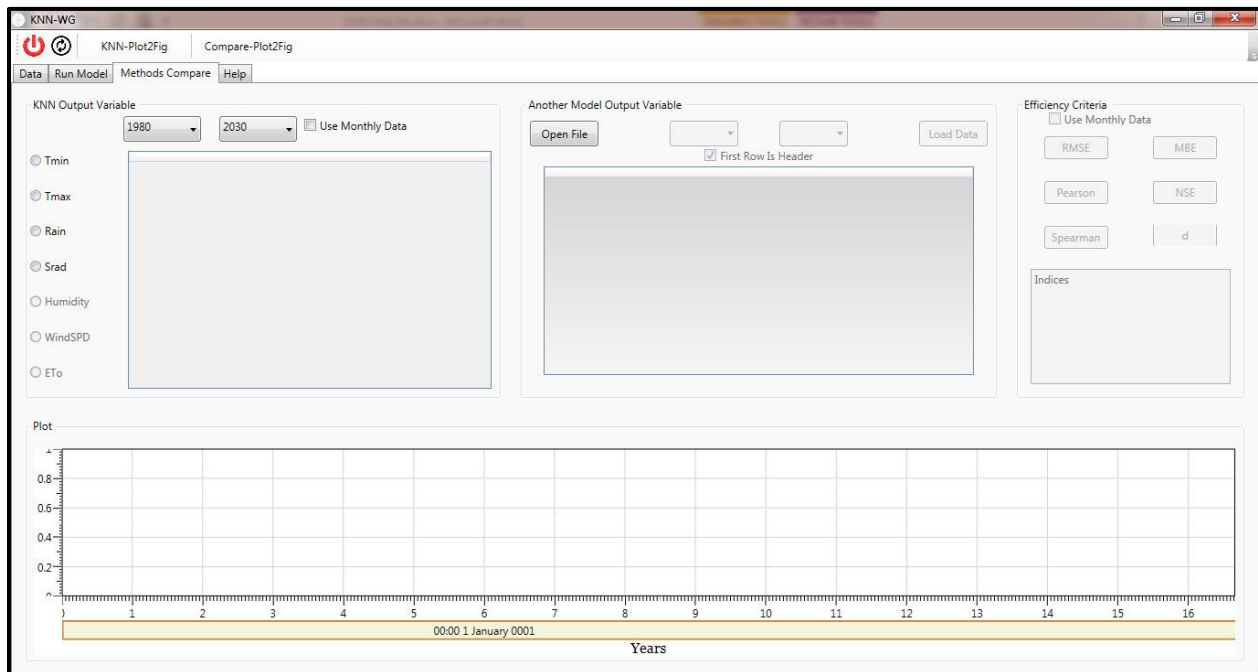


Figure 12

### **Step 3. Models comparison**

In comparison step, pay attention that the length's period of KNN and another models should be the same. Simply, you can load KNN data by selecting radio button of variables. However, for another model the user should load data from excel file (same as step 1). One of the best advantages of the KNN-WG is comparing the result of KNN approach to another model, such as LARS-WG, SDSM, CMIP5 model's output, ASD, and etc.

Fig. 13, present this step. The user can select the intent years for comparing data. The period that located in the KNN Output Variable is corresponding to the last selection in step 2, for future data period. In this step user can limit this period, according to another model that the user wants to compare it to KNN result.



*Figure 13*

In this example, we select the outputs of LARS-WG5, for comparing it to the result of KNN-WG. In Fig. 14, you can see the period that KNN has been run for it.

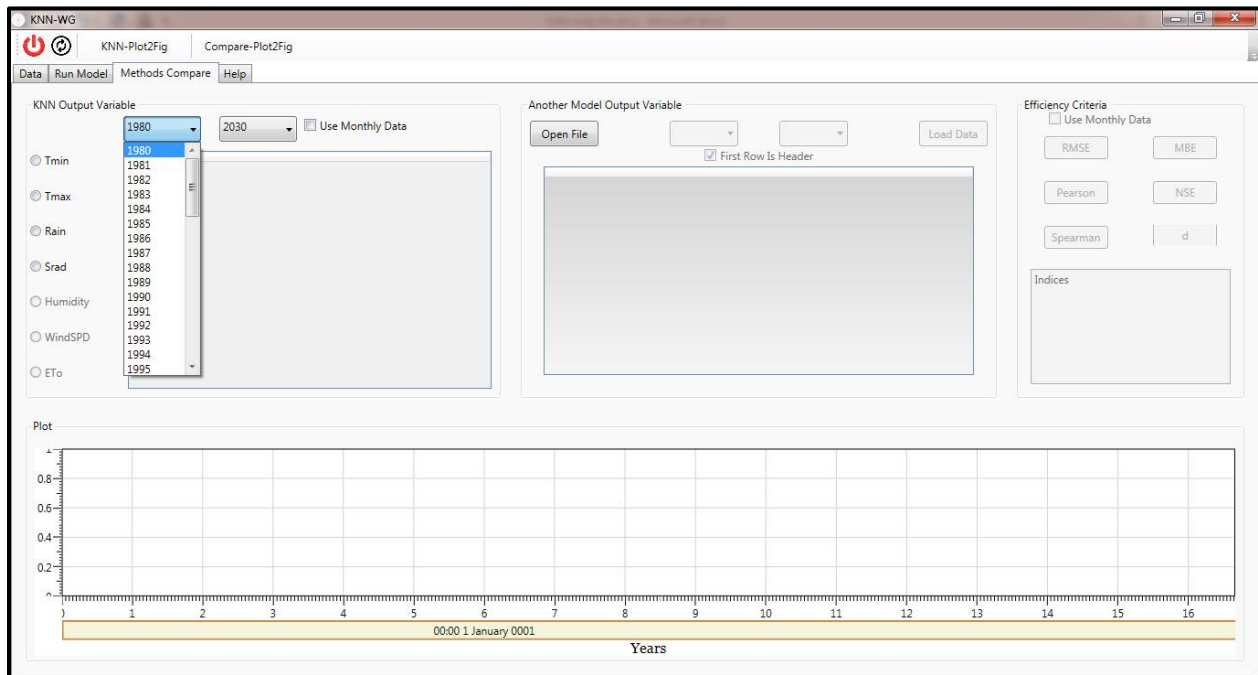


Figure 14

By selecting the arbitrary variable, the user can see the result of KNN in the table. If the user wants to see the monthly result, click on the “Use Monthly Data” button. In this example we select “Tmin”, after clicking the “Tmin” the results are appeared in the table (Fig.15).

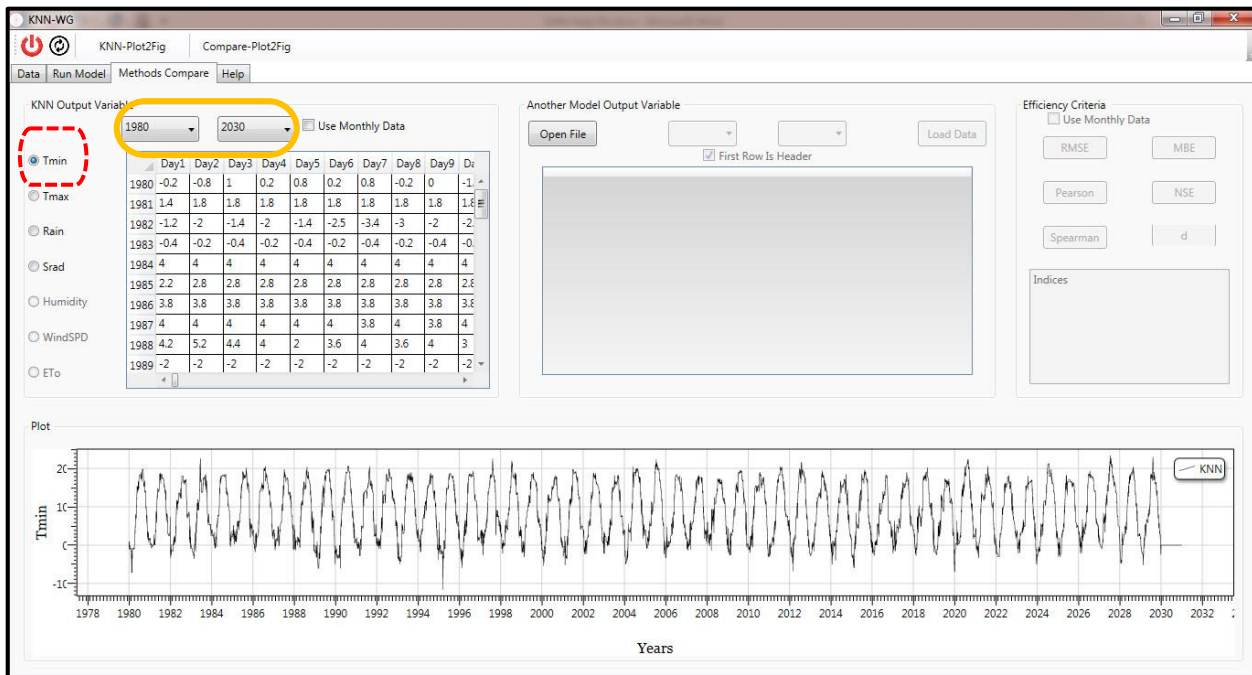


Figure 15

For the period, note that the second model (in this sample Lars-WG5) should be the same as KNN period of data. Check the period of second model and according to that data, change the period and may be limit the period of KNN like it. For example, our period of LARS-WG5 output data is 2011-2030 (Fig. 16 Num. 1), so we change the KNN period according to this range, as Fig. 16. Then Click “Open File” (Fig. 16 Num. 2), browsing the appropriate file and then click it to open the data file (Fig. 16 Num. 3). The user can select the input file of the intent model with .xls or .xlsx format file.

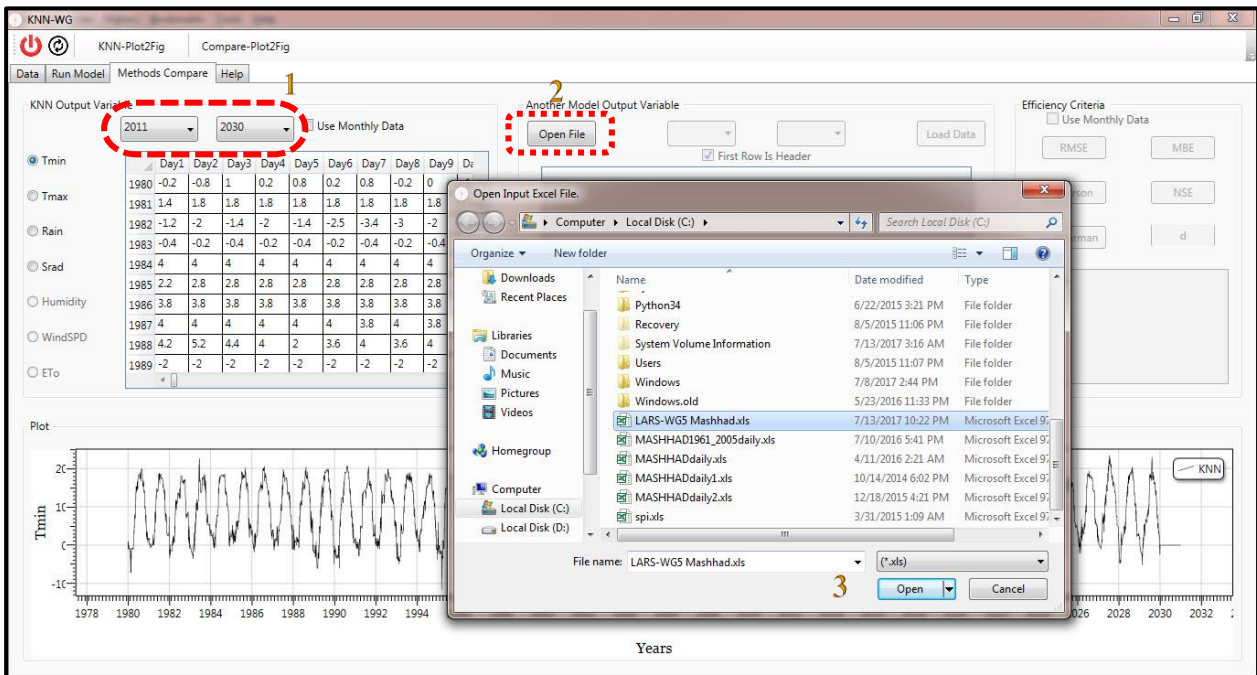


Figure 16

According to Fig. 17, as the first step, the user should assign the data format and name of the arbitrary variable with respect to the points that we mentioned in step 2. In this step (for the second model), one variable would be assign, in this example we select Tmin of LARS-WG5 model.

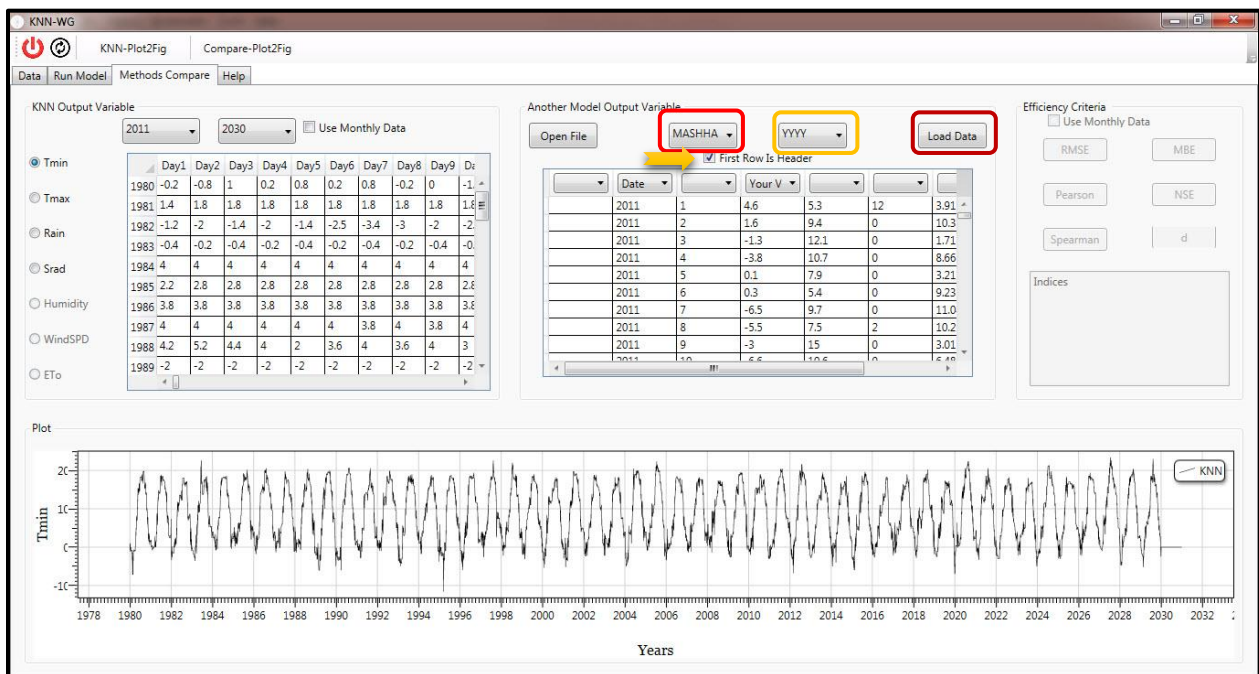


Figure 17

After the period and variable set, then click the “Load Data” button. If everything is OK, then the user see a message box of “Succeed” (Fig. 18). By clicking “OK” button, the user can see the graph of data in Fig. 19.

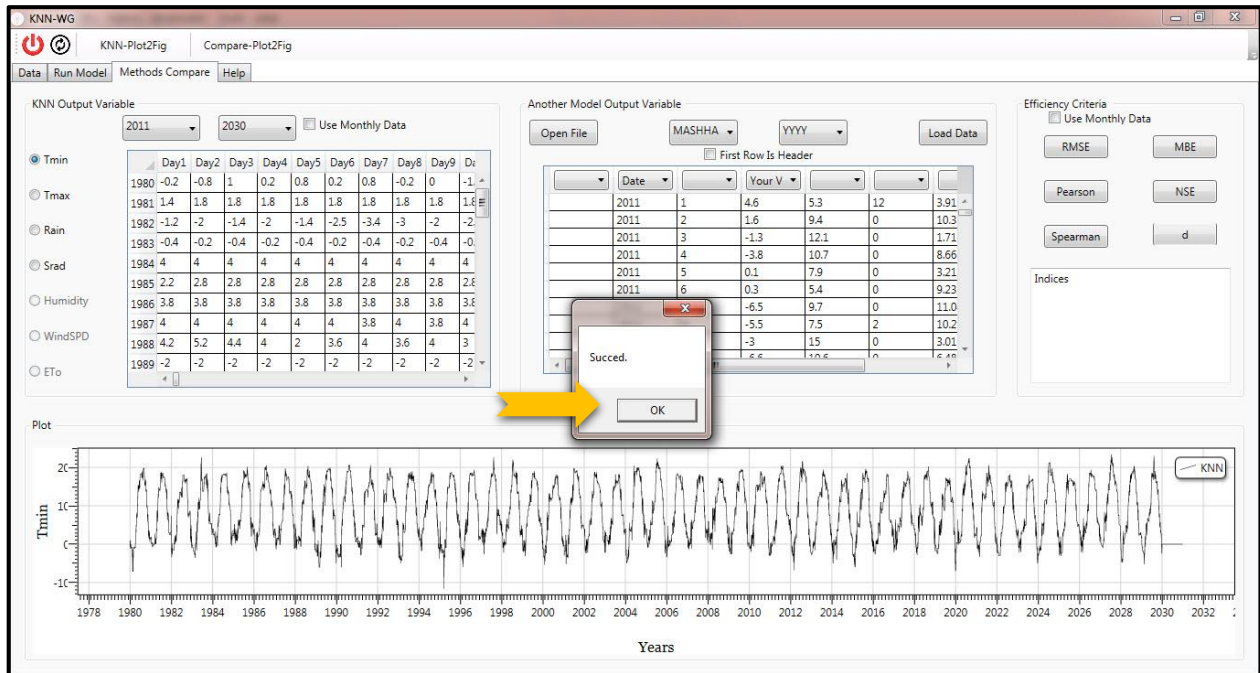


Figure 18

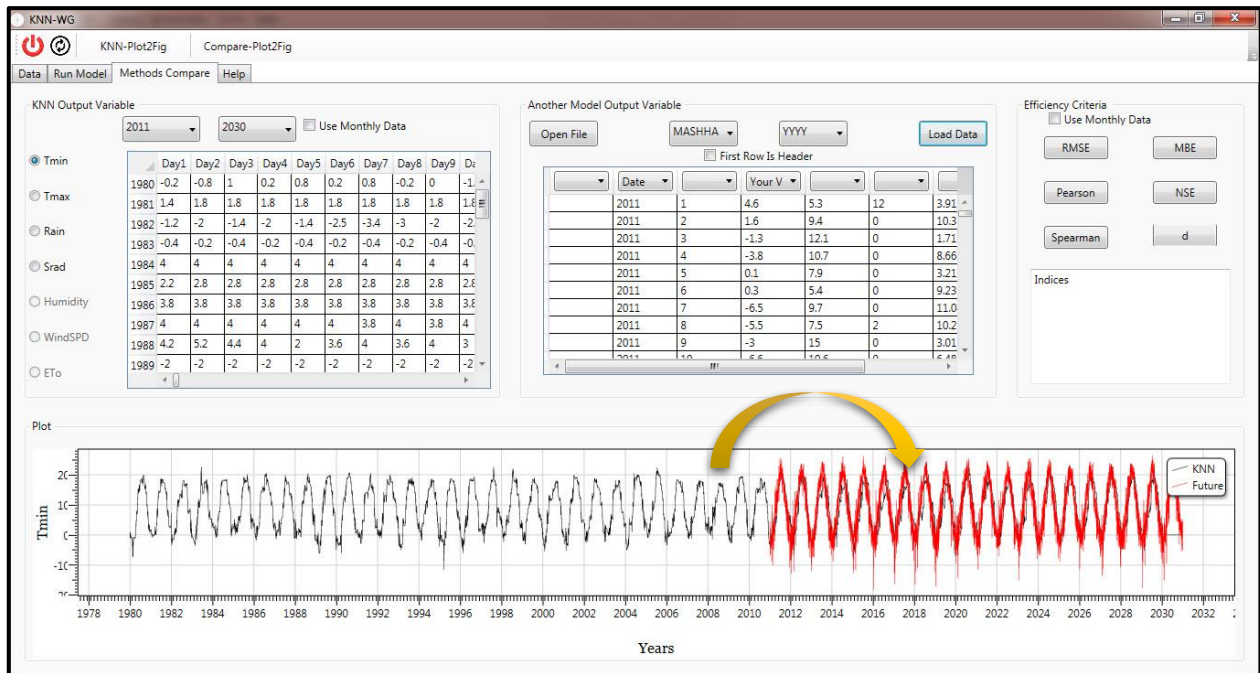


Figure 19

Finally, if the user wants to assess the efficiency criteria, the KNN-WG presents a section to the user at this step. By clicking every button in the Fig. 20 the user can observe this values. If necessary, check “Use Monthly Data”.

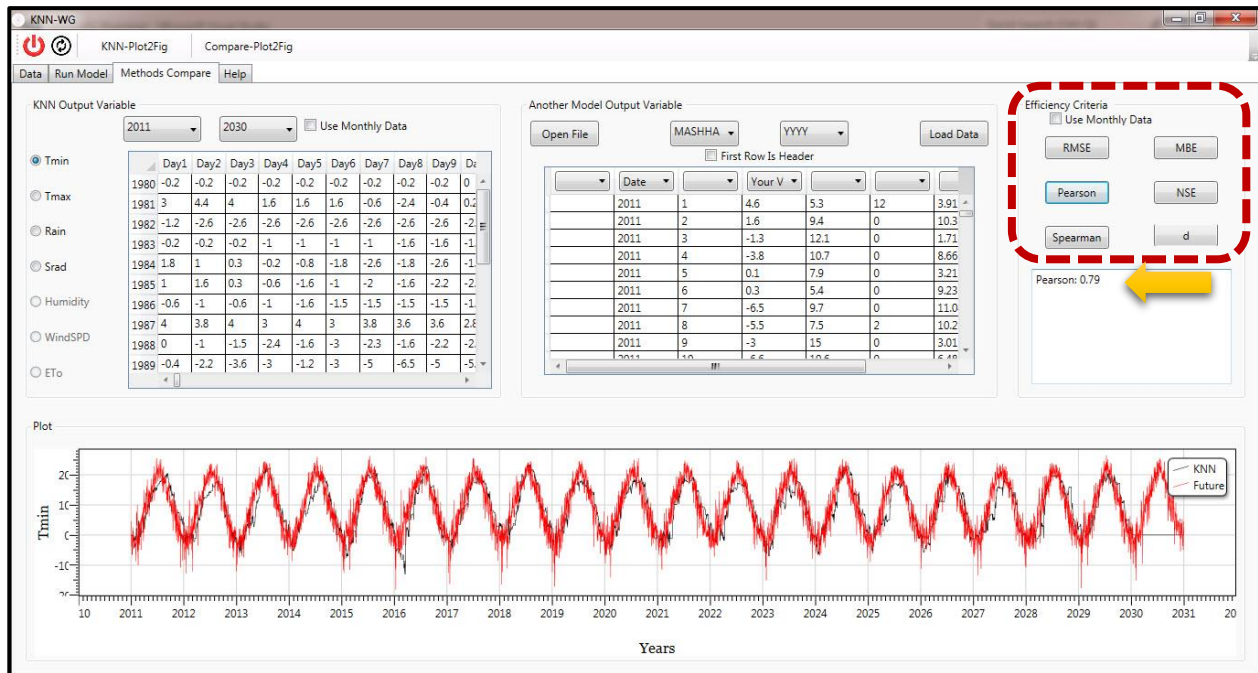


Figure 20

In this tab, click on "Help" button after you can read the instructions and methodology of this tool. By clicking on "Help Movie" button, the user can watch a typical sample of running tool.

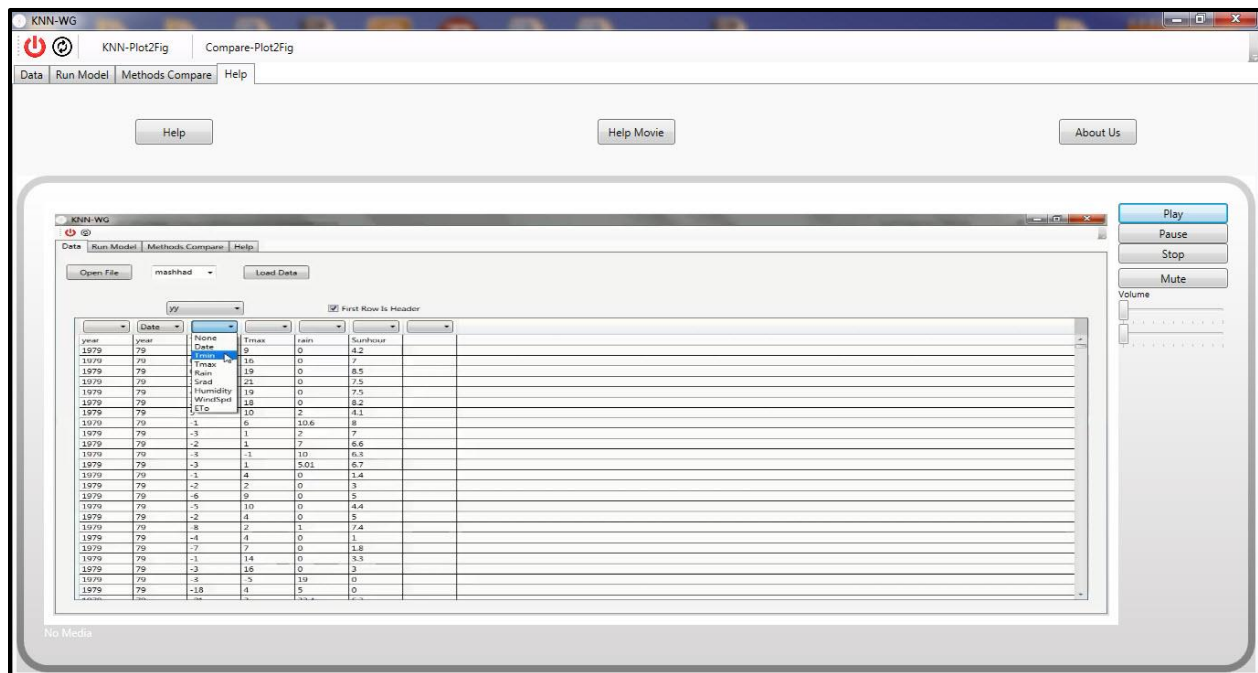


Figure 21