Weather Analogue: A Tool for Predicting Realizations of Daily Weather Data

Mohammad Bannayan and Gerrit Hoogenboom

Department of Biological and Agricultural Engineering, The University of Georgia, Griffin, GA 30223, USA

Bannayan@uga.edu

Introduction

• High quality weather data are quite crucial for the accurate simulation of the underlying crop, soil and atmospheric processes of agricultural, ecological and hydrological simulation models.

• The *k*-nearest neighbors (*k*-NN) is an analogue approach and has its origin as a non-parametric statistical pattern recognition procedure to distinguish between different patterns according to the selection criteria.

• A tool was developed (Fig. 1) that predicts daily weather data consisting of solar radiation, maximum and minimum temperature and precipitation based on the *k*-NN approach.

Materials & Methods

• In this approach (Fig. 2) a target year consisting of observed weather data including radiation, precipitation, maximum and minimum temperature is constructed.

• All days within a moving window of width "ww" centered on day t are selected as potential candidates for day t+1.

• Data were converted to standardized variables and Euclidean distance, d_j is computed between the target year of the current day's weather and the vector of observed data for each of "ww" days of each year in historical years as the measure of similarity.

$$d_j = \sqrt{\left[\sum_{j=1}^{d} W_j (V_{ij} - V_{mj})^2\right]}$$

where d_j, is Euclidean distances, V_j is the jth component of either of vectors (feature and historical) and W_j are weights.

• The *k*-NN approach selects from Euclidian distances and assigns probability weights to a subset of *K* distances with smallest to largest. Euclidean distances, d_j, are sorted in ascending order and the first *k*-NN are retained. Then a uniform random number U (0, 1) is generated and if $u \geq W_1$ then the day corresponding to distance d_1 is selected. If $u \leq W_k$ then day corresponding to distance d_k is selected. For $W_1 < u < W_k$ the day t corresponding to d_j is selected for which u is closer to W_j. At this stage weather data for day t are predicted.

• For evaluation our tool we used daily weather data from 10 representative sites located in Georgia, USA (Fig.3) and 16 sites located in USA, UK, Africa, and Asia (Fig. 4).

Results

• The *k*-NN approach successfully reproduced the number of wet days for both the monthly and the entire study period (Fig. 3). There was no significant difference (P > 0.05) for both accumulated radiation per month and total accumulated radiation across all sites (Fig. 3).

• The *k*-NN approach for the prediction of the number of days when the minimum temperature is at or below freezing (0˚C) and maximum temperature was greater than 35˚C across all sites were shown close similarity with observation in the target year (Fig. 3).

Fig. 1. Weather Analogue interface

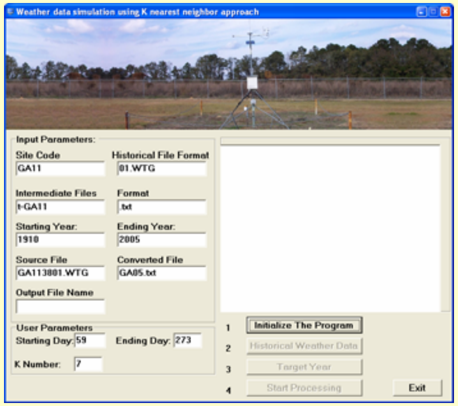


Fig. 2. Sequence of steps of k-NN methodology.

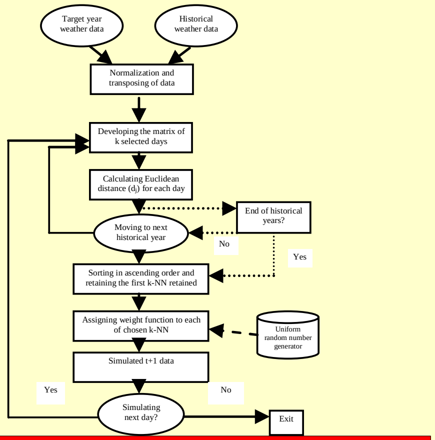


Fig. 3. A comparison of predicted and observed number of wet days, radiation, number of freezing events, and number of days above 35˚C across all sites.

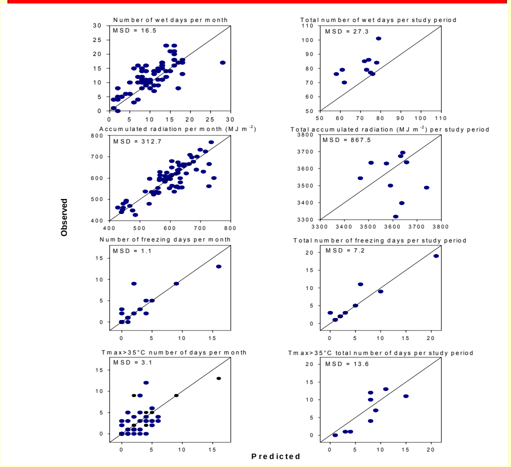
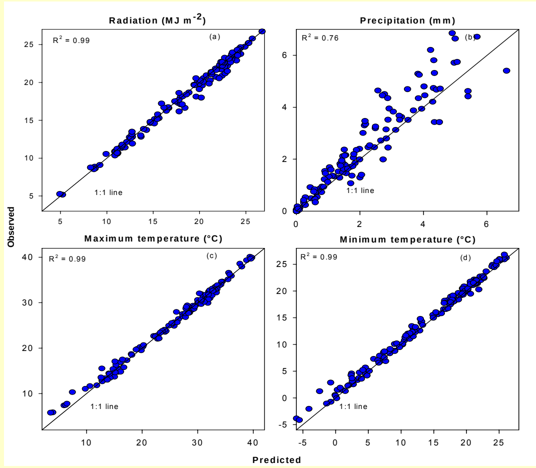


Fig. 4. A comparison of predicted and observed radiation, total precipitation and maximum and minimum temperature over study period across all sites.



• The mean square difference (MSD) for all weather variables ranged from 0.30 for maximum temperature to 1.51 for precipitation across all study sites (Fig. 4).

Conclusion

The *k*-NN method successfully predicted the weather sequences for multiple sites and was able to reproduce similar pattern of the target year. *k*-NN is a simple and reliable approach for predicting daily weather data.