The authors sincerely appreciate the referee for acknowledging our manuscript and providing valuable comments and/or suggestions that benefit our manuscript. After having carefully studied the referee's comments and/or suggestions, the responses to the comments and/or suggestions are addressed as follows, and the relevant responses and corrections in consideration of the comments/suggestions are revised in the marked (in blue color) manuscript accordingly. In addition, the responses to Referee #1 (**RC C5184**) are revised in the marked (underlined) manuscript as well, for Editor and Referees' convenience.

Referee #2 **(RC C5538):**

Major Comments:

1. In Table 3, the authors provide performance statistics (RMSE and CE) for the testing set only. Performance statistics for both the training and the validation set should also be provided, as it helps readers to understand the "generalization" achieved by the proposed hybrid model.

Reply: Thanks for the referee's valuable comment. Regarding the model generalization capability, the independent testing subset is commonly used for achieving such a purpose. Therefore, the performance obtained in the testing phase can be regarded as the generalization capability achieved by the proposed model. The authors greatly agree to the referee's comment that performance statistics for both training and validation subsets should be provided. Table 1 (newly added) shows the complete results, which has been added in the revised manuscript. Results displayed in the table shown below indicate that performance obtained in the testing phase is comparable to that of the training and validation phases, demonstrating the generalization was well achieved by the proposed model.

We would like to notice that the procedure of the hybrid AK modeling is that the AK model extends its estimation to ungauged sites based on the outputs provided by ANFIS, without using meteorological information but the spatial information. This is the reason why the table only shows the results obtained from ANFIS, not from AK.

Table 1 Performance of the ANFIS model in the training, validation and testing subsets

(**Table 3** of the revised manuscript)

	RMSE (mm day^{-1})	NDEI [1]	CE
Training	1.02	0.53	0.71
Validation	1.13	0.57	0.67
Testing	0.98	0.54	0.70

[1] NDEI is defined as the ratio of the root mean square error to the standard deviation of the target time series (Jang, 1993).

Reference:

Jang, J. S. R.: ANFIS - Adaptive-network-based fyzzy inference system, IEEE T. Syst. Man Cybernet., 23, 665-685, 1993.

2. Testing is done based on only 3 stations. This in my opinion is not a robust testing. If the authors disagree, then they should explain in the paper why considering only 3 stations would still be valid for testing the proposed model.

Reply: In this case study, there are 19 meteorological gauging stations which appropriately cover the whole of Taiwan (for the purpose of this study, these 19 stations are not too many but sufficient enough for representing the spatial variability of the whole Taiwan island). Many previous studies have demonstrated that the ratio of data arranged in both training and validation subsets to data arranged in the testing subset could be 5:1 or 6:1 (Pan et al., 2007; Chiang et al., 2004). Therefore, it should be acceptable to use 16 stations for model training and validation while 3 stations as the testing targets.

As for the selection of stations, the determination of station Nos. 17-19 is dependent on the spatial locations (see Fig. 1 below). The reasons are: (1) these three stations are separately located in northern, central and southern Taiwan; (2) these three sites have relatively few meteorological stations around them as compared with the other sites, and therefore enhance the practicability of choosing these three stations; and (3) these three sites are surrounded by other sites. Cressie (1990) and Schiltz et al. (1998) presented that most errors in the simple kriging are attributable to the spatial extrapolation.

Based on the description above, the authors are confident that the use of 3 stations is sufficient enough for testing the proposed model.

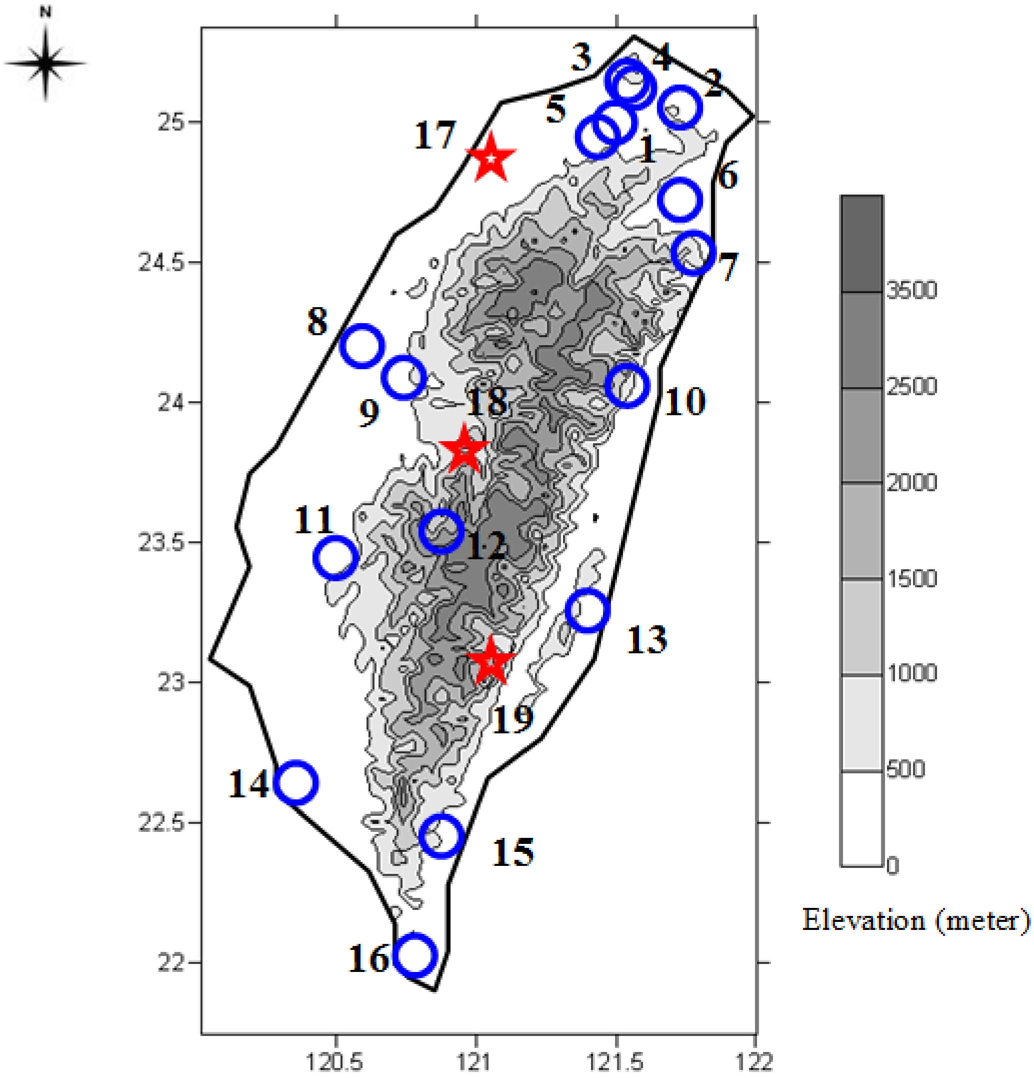

Fig. 1 Location of meteorological stations in Taiwan.
(**Fig. 3** of the revised manuscript)

References:

Chiang, Y. M., Chang, L. C., Chang, F. J.: Comparison of static-feedforward and dynamic-feedback neural networks for rainfall-runoff modeling, Journal of Hydrology, 290, 297-311. 2004.

Cressie, N.: The originas of Kriging, Mathematical Geology, 22, 239-252. 1990.

Pan, Y., Jiang, J. C., Wang, Z.: Quantitative structure–property relationship studies for predicting flash points of alkanes using group bond contribution method with back-propagation neural network, Journal of Hazardous Materials, 147, 424-430. 2007.

Schultz, C., S. Myers, J. Hipp, and C. Young .: Nonstationary Bayesian kriging: a predictive technique to generate corrections for detection, location, and discrimination, Bull. Seism. Soc. Am, 88, 1275–1288. 1998.

3. For the three stations (No. 17, No. 18, No. 19), what would be the values obtained by single ANN models? Providing this comparison would help readers gauge the merits and demerits of adopting the proposed model in a diverse topographic and climatic area.

Reply: Thanks for the referee's valuable and essential comment. The major difference between the AK model and the single ANN model is that the AK model is capable of estimating the spatial distribution of pan evaporation at ungauged sites without using their meteorological measurements, whereas the single ANN model is not able to provide estimations at ungauged sites where meteorological measurements are not available. It is worth noticing that the input information for the AK model is different from that of the single ANN model. In other words, the comparison between the AK and the single ANN models is not fair. However, the comparison of the AK, PM and the single ANFIS models is still provided in Table 2 in order to conform to the referee's comment. Even though the estimation of the AK model is not as accurate as that of the ANFIS model for stations No. 17 & 18, results produced by the AK model are satisfactory and close (even better for No. 19) to those of ANFIS when taking the difference of input information into consideration. Moreover, the AK model is significantly superior to the Penman-Monteith (PM) empirical formula in terms of smaller RMSE and higher CE values at these three sites. In sum, the established AK model in this study is capable of estimating pan evaporation at ungauged sites.

Table 2 Performance of the AK, PM and ANFIS models at individual meteorological station

	RMSE (mm day $^{-1}$)			CE		
	No.17	No.18	No.19	No.17	No.18	No.19
AK	1.17	1.02	1.08	0.64	0.12	0.59
PM	1.59	1.25	1.14	0.35	-0.32	0.54
ANFIS	1.03	0.75	1.12	0.72	0.52	0.56

<u>Minor Comments:</u>

Some of my minor comments have already been noted by the other reviewer. Here are some additional comments.

1. P 9682, L 12 – Is it radiation or "net radiation"?

Reply: Thanks! It is radiation. For a more precise definition, we revise the term as “global solar radiation”.

2. P 9682, L 11-13 – add more details about the collected data (e.g. weather station instrumentation). At what height(s) are wind speed, temperature, humidity, and radiation measured? How is precipitation handled? If any correction for precipitation has been made, please explain it in the manuscript.

Reply: Thanks for the referee’s suggestion. The data were collected from Taiwan’s Central Weather Bureau. Several sentences for weather station instrumentation were given as follow.

In general, each weather station in Taiwan has installed a piston mercury barometer, a sheathed thermometer, a propeller anemometer, a tipping-bucket rain gauge, a class A pan, a hair hygrometer, a pyranometer, a solar-cell sunshine recorder and a psychrometer for measuring pressure, temperature, wind speed, rainfall, pan evaporation, humidity, global solar radiation, sunshine hour and humidity, respectively. More detail can be found at the web site (http://www.cwb.gov.tw/V7e/index_home.htm) of the Central Weather Bureau.

For the referee’s reference, Table 3 provides the heights of the instrumentations for measuring wind speed, temperature, humidity and radiation at weather stations. The ranges of the heights for the instrumentations related to wind, temperature, humidity and radiation are 【7, 34】, 【1.1, 1.53】, 【1.1, 1.53】and 【1.1, 27.4】(unit: meter), respectively. Moreover, no corrections were made to precipitation in this study because precipitation has small influence, which has been commonly adapted by previous researches.

Table 3. Height of instrumentation for measuring meteorological variables at all 19 weather stations.

Station No. \ Variable	Wind speed (m)	Temperature (m)	Humidity (m)	Radiation (m)
1	33	1.1	1.1	22
2	34	1.2	1.2	1
3	7	1.2	1.2	1
4	8	1.2	1.2	1
5	14	1.1	1.1	22
6	26	1.2	1.2	26
7	34	1.3	1.3	1
8	33	1.5	1.5	27
9	17	1.3	1.3	14

10	30	1.3	1.3	23
11	14	1.2	1.2	9
12	15	1.3	1.3	13
13	12	1.5	1.5	5
14	14	1.2	1.2	13
15	12	1.2	1.2	13
16	14	1.5	1.5	12
17	15	1.3	1.3	13
18	8	1.3	1.3	5
19	11	1.4	1.4	9

3. P 9684, L 15-16 – State the values adopted for "slope vapor pressure curve" and "psychrometric constant"

Reply: OK. The slope vapor curve (△) and psychrometric constant (γ) are referred from (Allen et al., 1998), and the equations for calculating those values are given in Equations (1), (2) and (3).

$$\Delta = \frac{4098 \times e_s}{(T + 237.2)^2} \quad (1)$$

$$e_s = 0.611 \exp\left(\frac{17.27T}{T + 237.3}\right) \quad (2)$$

$$\gamma = 6.65 \times 10^{-4} \times P \quad (3)$$

where e_s, T and P are saturation vapor pressure, mean daily air temperature and mean atmospheric pressure, respectively.

Reference:

Allen, R.G., Pereira, S, L., Raes, D. and Smith, M.: Crop evapotranspiration, Guidelines for computing crop water requirements, Irrigation and Drainage Paper No. 56, FAO, Rome, 1998.

4. P 9686, L 3 – How are parameters "C1" and "a" estimated?

Reply: In this study, the simple kriging is selected for analyzing the spatial error of ANFIS. The exponential function (shown as Equation (4)) is introduced for the spatial covariance. It is obvious that the maximum covariance falls approximately between 0.25 and 0.3 (Fig. 2) and can be used as a reference when searching parameter "c_1". The parameters "c_1" and "a" can be obtained by the trial-and-error procedure.

$$C(r) = c_1 \exp(-\frac{3r}{a}) \quad (4)$$

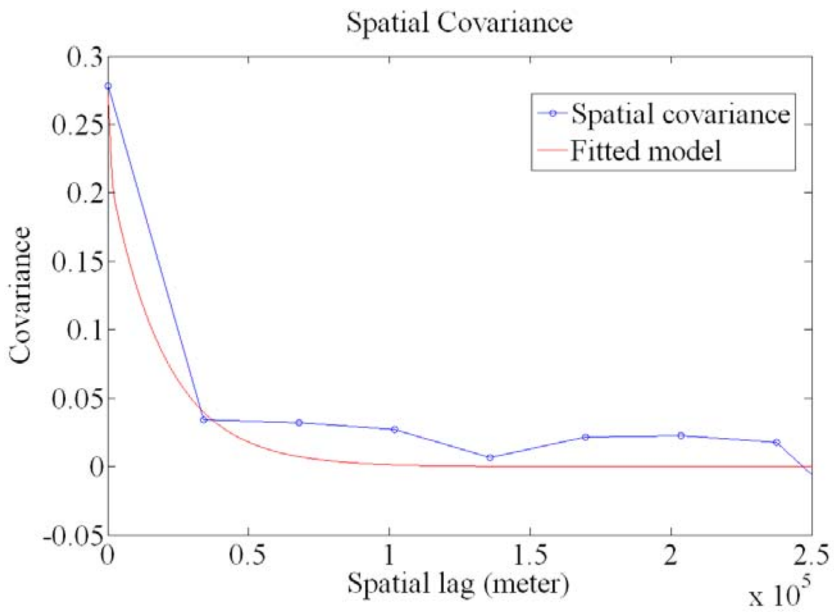

Fig. 2 Covariance values obtained from the fitted and the experimental models based on the residuals from ANFIS.

(**Fig. 5** of the revised manuscript)