Reviewer #1

This paper presents a valuable dataset that may encourage greater application of soil erosion modelling in China. The results of the cross-validation show good agreement, and an interesting comparison of ordinary and universal kriging is made. Some details of the methodology could be clarified, particularly in terms how the covariates and parameter layers are used in universal kriging. Importantly, spatial interpolation error could also be discussed more. Overall, the paper is well written and describes a valuable dataset.

Comment (1): Line 22: Word “in” is not needed in this context.

Response: It will be revised.

Comment (2): Line 23: typo: available

Response: It will be corrected.

Comment (3): Line 78: WGEN has not been defined. This is another stochastic weather generator?

Response: WGEN is the abbreviation of the Weather GENerator, which is a weather generator developed by Richardson and Wright (1984). The description and reference will be added to the manuscript. (Reference: Richardson, C.W., Wright, D.A.: WGEN: A model for generating daily weather variables, 1984.)

Comment (4): Line 152: This statement could be clarified: “Rainfall intensity is basically assumed to be ranked from high to low in CLIGEN”.

Response: In CLIGEN (Nicks et al., 1995), also Arnold and Williams (Arnold and Williams, 1989; Williams et al. 1984), it is assumed that the magnitude of precipitation intensity decreases exponentially from the maximum rate when time distribution of precipitation intensities is discarded. We will revise it to make it clearer.

Comment (5): Line 173: How was TimePk determined in Wang et al. (2018b). More information would be helpful beyond the fact there was available hourly precipitation and MX.5P values to estimate TimePk. More detail could be given about how the other intensity parameter, MX.5P was determined if it requires high-resolution data.

Response: For hourly data as we collected, the time to peak intensity, t_p, can be calculated for every storm directly using equations (9) to (11) listed in the manuscript. For example, if the storm duration is 4 hours, and the peak intensity occurs in the 2^{nd} hour, then t_p = (2-0.5)/4 = 0.375 based on equation (10). Then using equation (11), parameters TimePk can be obtained for all stations. From Wang et al. (2018b), we know that two sets of TimePk parameter values prepared using hourly rainfall and 1-min rainfall generate very similar CLIGEN outputs. Therefore, we used hourly rainfall to prepare TimePk for all stations following equations (9) to (11) for this study.

Ideally, MX.5P values should be prepared using rainfall intensity data with an observed time interval of no more than 30 minutes. Depending on the temporal resolution, I_{30} can be calculated directly from moving averages of the original data over successive 30-min.

More sentences will be added to explain how TimePk and MX.5P were calculated in the revised manuscript.

Comment (6): Line 191: typo: CLIGEN parameters.

Response: It will be revised.

Comment (7): Line 191: Clarify how twelve groups were used.

Response: Twelve groups means twelve months for each parameter, and the expression is not clear. The statement will be revised to “The longitude, latitude, elevation, and annual rainfall amount were found correlated with the parameters one for each month for CLIGEN” to make it clearer.

Comment (8): Line 195: How many iterations of leave-one-out cross-validation were there that produced the four performance metrics? Is this equal to the number of stations?

Response: Yes, this is correct. For each interpolated parameter, the leave-one-out cross-validation procedures were iterated for 2405 times, which equals to the number of stations. More descriptions will be added to the manuscript for clarification.

Comment (9): Line 197: How was the 131 input parameters number arrived at?

Response: There was a typo here. There were 13 groups of input parameters required for CLIGEN for temperature, precipitation and solar radiation, and in each group, there are 12 parameters. The total number of interpolated parameters is 155, which is equal to 12 parameters $\times$ 13 groups-1, as the 12th parameter of TimePk is always equal to 1. Therefore, there were 155 parameters involved in the interpolation, and this will be revised in the manuscript.

Comment (10): Line 210: Word "of" not needed.

Response: It will be revised.

Comment (11): Line 210: Clarify how the 156 number of parameter layers was arrived at.

Response: Please refer to the comment (9).

Comment (12): Line 234: Clarify "the value became convergent from cold season to warm".

Response: It will be revised to "spatial variance became smaller from the cold season to the warm one".

Comment (13): Line 285: Interpolation accuracy is stated here to be temporally dependent, but more discussion of how it is spatially dependent would be helpful. I would guess that in data sparse areas, interpolation error is much higher. The leave-one-out cross-validation does not account for the fact that data sparse areas could actually represent large parts of the total interpolated area, so it could be the case that error would be much higher if more observations were available to check error in data sparse areas. Is it indeed the case that error is higher in western China? Would it be possible to make an error map for MEANP and TMAX as examples? Or, consider some way of presenting spatial error.

Response: We agree with you that the interpolation error for data sparse areas is higher. We've discussed the influence of spatial distribution of stations on the interpolation accuracy in the second paragraph in Discussion. We calculated and compared the mean absolute relative errors (MAREs) derived from leave-one-out cross-validation for three regions in China (the Eastern Monsoon Area, the Northwest Arid Area and Qinghai-Tibet Plateau) with different station density. Results were listed in Table 7 and showed that the station density has an influence on the quality of the interpolation. Error in the Eastern Monsoon Area is the lowest and the highest in the Qinghai-Tibet Plateau.

Considering that the station density is quite sparse in the western region of China, and the leave-one-out cross validation can't show the interpolation error in regions without stations. We've plotted the standard error of the interpolation results for two parameters, TMAX AV and MEAN P in August as an example (Fig. 1). It can be seen from the figures that the errors are relatively high in the western part, especially in the south-western part, where is a large area without stations and characterized with the highest standard errors for both parameters. The following figure will be added to the discussion part of the manuscript to make this clearer.

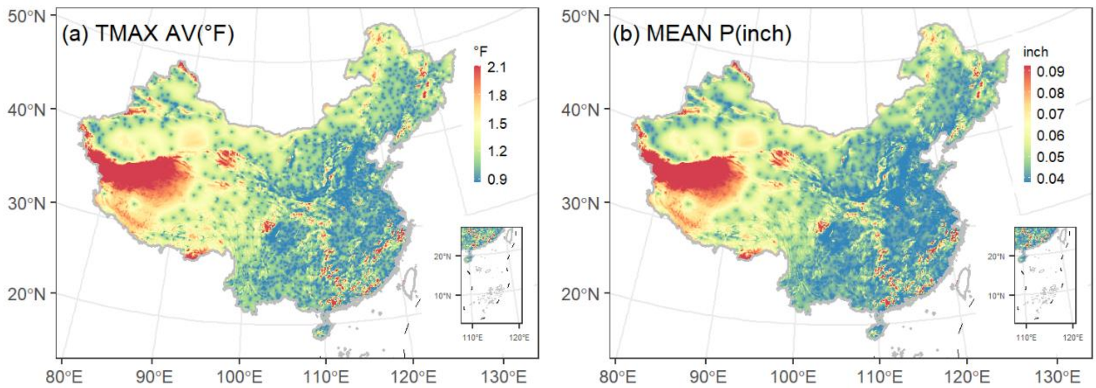

Fig. 1. Spatial distribution of the standard error for interpolation results of TMAX AV and MEAN P in August using Universal Kriging.

Comment (13): Table 3: RMSE for TimePk of 0.01 is very small considering TimePk ranges from 0-1. Why is NSE particularly low for TimePk? Is it the small numeric scale of TimePk?

Response: Thank you for your comment. We've checked the computation and data that the results are without mistake. TimePk is a group of dimensionless parameters, which represents the cumulative probability distribution of the time to peak. The equation for NSE is,

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n}(P_{ij}^{obs} - P_{ij}^{K})^2}{\sum_{i=1}^{n}(P_{ij}^{obs} - \overline{P^{obs}})^2}$$

where $i = 1, 2, \ldots 2045$ stations, $j = 1, \ldots 11$ parameters. It's only 11 parameters of TimePk need to be interpolated because the 12th parameter of TimePk is always equal to 1, which is known already. NSE for 11 parameters can be calculated, and the average value over 11 NSEs is the number listed in Table 3 in the manuscript.

As the equation for NSE represents the ratio between predicted errors and the variation of the observed values. The variation of the observed values depends on its specific shape of distribution. For each of the 11 TimePk values, the variation of observed values across the whole study area is relatively small compared with the predicted errors (as showed in Fig. 7m in the manuscript), leading to a particularly low NSE value. However, if we put all 11 parameters of TimePk together to calculate NSE, it will become much better and equal to 0.998 for both OK and UK.

Reviewer's comment reminds us that NSE may be not a good indicator for parameters in interval scales which are without true zero. Therefore, we will reorganize results listed in Table 3 that the NSEs will only be kept for parameters in ratio scales (MEAN P, S DEV P, SKEW P, SOL.RAD, and SD SOL). RMSE will be kept for all the 13 groups of parameters. The statements, figures and tables in Results will be reorganized correspondingly.