Thanks for the valuable comments by the Referee #2. Here we present responses to five main comments.

1. The authors present in the introduction a number of methodologies for assessing soil erosion: a) factorial scoring b) plot measurements c) field-based approach d) Modelling (RUSLE). Then they analyse more in detail the application of RUSLE as 3 different options: 1) sample survey 2) raster multiplication 3) sample survey and geostatistics. The authors have followed the third option. Find below the most important remarks and issues that authors should address in their revision: First remark: I would appreciate if the authors have compared their results with the second option. This would give much more advanced knowledge in the manuscript. You mentioned that you have available K-factor, R-factor maps at 250m resolution plus a land use map at 100000 scale. So, it would have been excellent to compare your results with an estimated Soil loss by water erosion (simply multiplying the above mentioned high resolution grids).

   Response: It is not difficult to conduct raster layer multiplication technically, however, the multiplication of R and K factors (assuming L=1, S=1, B=1, E=1, T=1) reflect the potential of soil erosion, which is different from the soil erosion estimated in this study.

2. As the 1st reviewer said (and I agree), the authors have presented an interpolation method which takes into account 5 different group of parameters. It is logical (and obvious that the IV and V would perform much better than the I. In a recent research (to be online soon), we identified cover management factor as the most sensitive for estimating soil loss by water erosion. The manuscript could be even more worthy if the authors have compared their findings with alternative methods (plot measurements, expert knowledge, field-based approach).

   Response: In the manuscript, we have compared with the plot measurements (Guo et al., 2015). We could at least compare the result with the third national census for soil erosion, which is based on factorial scoring method (Wang et al., 2016). Thanks for your suggestion.

3. The findings regarding the forests are much too high. Erosion of > 3 t ha-1 in forest is not at all acceptable. Even there can be very steep slopes, the forestland experience erosion of much less than 1 t ha-1 annually. Their comparison with the findings of Guo (2015) and the findings in Europe (2015) show that erosion in forests
is much less. The same applies for grasslands. Please consider also a comparison of your findings with the paper of Wang et al (2016) "Assessment of soil erosion change and its relationships with land use/cover change in China from the end of the 1980s to 2010".

   Response: Wang et al. (2016) used a factorial scoring method to assess soil erosion risk and change in China from the end of the 1980s to 2010. As it was discussed in the introduction of this study, the resulting map by the factorial scoring method depicts classes ranging from very low to very high erosion or erosion risk. However, it can't generate erosion rates. We are also concerned about the relatively higher erosion rates of forest and grassland in the Shaanxi province comparing with some previous research. Preliminary analysis showed that they came from the Primary Sampling Units, and not introduced by the spatial interpolation process. The reason for this may be due to the different definitions of forest and grassland, the unique topography in Loess plateau and intensive human activities. The minimum canopy density (crown cover) threshold for the forest in this study is 10%, which may suggest a lower cover coverage and higher B factor. The grassland includes the native and artificial grassland, with more intensive livestock and human activities. But more analysis is required and we will do it before we make the revision.

4. Authors should explain and justify the selection of their statistical model BPST and not the selection of Cubist or GPR or regression krigining? Moreover, In your geo-statistical model, the topography is ignored. Why?

   Response: In spatial data analysis, there are mainly two approaches to make the

prediction of a target variable. One approach (e.g., kriging) treats the value of a target variable at each location as a random variable and uses the covariance function between these random variables or a variogram to represent the correlation; another approach (e.g., spline or wavelet smoothing) uses a deterministic smooth surface function to describe the variations and connections among values at different locations. Our work takes the second approach. The relationship between the traditional spatial statistics, and splines have been discussed in the literature, e.g. Matheron (1981) and Wahba (1990). A brief comment is presented in the following. Specifically, as discussed in Mitas and Mitasova (1999), "Kriging assumes that the spatial distribution of a geographical phenomenon can be modeled by a realization of a random function and uses statistical techniques to analyze the data and statistical criteria for predictions. However, subjective decisions are necessary such as judgement about stationarity, choice of function for theoretical variogram, etc. In addition, often the data simply lack information about important features of the modelled phenomenon, such as surface analytical properties or physically acceptable local geometries."
In contrast, "Splines rely on a physical model with flexibility provided by change of elastic properties of the interpolation function. Often, physical phenomena result from processes which minimize energy, with a typical example of terrain with its balance between gravitation force, soil cohesion, and impact of climate. For these cases, splines have proven to be rather successful." For our problem, we also pay special attention to the following two practical issues: (1) the data are not necessarily evenly distributed; observations can be dense at some locations while sparse at others. (2) the domain for the data can take non-rectangular shapes.

In this work we introduce bivariate splines on triangulations to handle irregular domains and propose to extend the idea of univariate penalized splines (Eilers and Marx, 1996) to the two-dimensional case. The BPST method we consider have several advantages. First, it provides good approximations of smooth functions over complicated domains. Second, the computational cost for spline evaluation and parameter estimation are manageable. Third, the BPST doesn't require the data to be evenly distributed or on regular-spaced grid.

Topography factors based on smaller scale of topography map in the mountainous area have large uncertainties. A recent research by our group showed that the slope steepness based on the 30 m ASTER GDEM V1 is about 64% lower and the slope length on the other hand was increased by 265%, compared with the reference value based on the topography map with a scale of 1:2000 for a mountainous watershed in Northern China. If larger scale of topography map can be collected and it is not difficult to incorporate topography factors into our model by adding L and S factors in the equations (8) and (9).

References:

Mitas, L. and Mitasova, H. (1999). Spatial interpolation. Geographical information systems: principles, 1, 481-492.

Matheron, G. (1981) Splines and Kriging: their formal equivalence. Syracuse University Geological Contributions: 77–95.

Wahba, G. (1990). Spline models for observational data. CNMS-NSF Regional conference series in applied mathematics 59. Philadelphia, SIAM.

5. The filed survey (section 2.1) indicates that the sampling of erosion points was not so dense. Please give some levels of uncertainty taking into account that you sampled on PSU every 25 km2 even less. Moreover, you mentioned that "PSU points were surveyed" : you don't describe how you estimate the R, K, LS, B, E, T factors in each point? Did you sample and analyse the soil for estimating K-factor? Did you install a high temporal resolution rainfall station for measuring R-factor? Etc. Maybe this is somehow written in section 2.3 but it is not clear as you don't provide detailed information on how the R-factor, K-factor was calculated. In the same way that you criticize the non-availability of all input layers when multiplying the grids (factors), somebody can criticize your methodology that non all information (K-factor, R-factor, ect) is available at point

level. How you respond to this?

Response: The density of sample units in our survey depends on the level of uncertainty and the budget of the survey. We sampled a density of 4% in four experimental counties in different regions over China and found a density of 1% was acceptable given the current financial condition.

Lai and Wang (2013) provided the asymptotic properties of the BPST method, For example, they investigated how the bias and variance of the BPST estimator change with respect to the sample size and the number of the triangulations. Since our data are a little sparse in some area, we employed the roughness penalties to regularize the spline fit; see the energy functional defined in equation (12). When the sampling is sparse in certain area, the direct BPST method will not be effective since the results of the smoothing stage may have high variability due to the small sample size. The penalized BPST is more suitable for this type of data because it can help to regularize the fit.

We didn't install a rainfall station or collect soil samples for measuring R or K factor for each PSU. Instead, we collected 2678 weather and hydrologic stations with erosive daily rainfall from 1981 through 2010 and generated the R factor raster map over the entire China (Xie et al., 2016). And for the K factor, soil surface attributes for 7764 soil species from the Second National Soil Survey and more than 950 soil samples newly collected were used to generate the K factor for the entire country (Liu et al., 2013). The R and K factors for each PSU were clipped from the map of the entire country. A topography contour map with a scale of 1:10000 for each PSU was collected to derive the slope length and slope degree and to calculate the slope length factor and slope steepness factor (Fu et al., 2013). The calculation of B, E and T was based on the field survey of each PSU. As we know that R factor in USLE requires the breakpoint rainfall data more than 20 years, which is not feasible if a rainfall station was set up in each PSU. We assumed that the variation of R factor could be captured by more than 2000 stations over China, which were the most stations we could collect at present. Soil maps with scales of 1:500,000 to 1:200,000 (for different provinces) generated more than 0.18 million polygons of soil attributes over mainland China, which was the most accurate soil information we could collect at present. We assumed the result of the soil survey could be used as the information of K factor in our soil erosion survey. R factor and K factor in the sample point of NRI was also from the interpolation result of weather stations and soil survey map, respectively.

References:

Lai, M. J. and Wang, L. (2013). Bivariate penalized splines for regression. Statistica

Sinica 23 1399–1417.

Xie, Y., Yin, S. Q., Liu, B. Y., Nearing M., and Zhao, Y.: Models for estimating daily rainfall erosivity in China. J Hydrol, 535, 547–558, 2016.

Liu, B.Y., Guo, S. Y., Li, Z. G., Xie, Y., Zhang, K. L., and Liu, X. C.: Sample survey on water erosion in China. Soil and Water Conservation in China, 10, 26–34, 2013 (in Chinese with English abstract).

Fu, S. H., Wu, Z. P., Liu, B. Y., and Cao, L. X.: Comparison of the effects of the different methods for computing the slope length factor at a watershed scale. Int Soil Water Conserv Res, 1(2), 64–71, 2013.