Replies to Reviewer #3

The Authors thank the reviewer for her/his helpful comments which increased the quality of our work. Here below we reported the reviewer's comments (in black) and our replies (in red).

Major questions

a) The paper presents the situation as if each observatory has to choose its own k9. But, according to the International Service of Geomagnetic Indices (ISGI), the ISGI headquarters are in charge of the computation of k9 value for each magnetic observatory. ISG is a service of the International Association of Geomagnetism and Aeronomy (IAGA) which recognized "the unique role of the International Service of Geomagnetic Indices (ISGI) in the derivation, publication, and dissemination of these indices" (IAGA, Resolution No. 9 (1989). So, this paper should present its conclusions as an academic effort to check the assigned k9 values rather than to propose a new operational value.

We tried to determine L9 values (note: In the whole manuscript we substituted "K9" with "L9" as suggested by #2 reviewer) by means of a statistical approach and comparison with historical reference observatories. The values we found are in good agreement with the ones directly obtained from Mayaud's method (personal communication, by ISGI members). We are convinced that the contribution of ISGI in providing the L9 value is fundamental and that L9 values of observatories which have been providing the K index in the past (often rounded in multiple of 50 nT) cannot be changed since homogeneity of the series is of primary importance. However, before a new observatory starts providing K index values, it is worth evaluating K indices with a provisional L9 value, assigned by ISGI, and then refine it with a procedure like the one we have shown, based on the comparison with reference, well set, historical observatories. The authors believe that discussions derived from this deal help to reason about the suitability of past procedures with respect to the modern capability of automatic computations. In the revised version of the manuscript these considerations are included in lines 264-266 and 288-292.

b) K index is a coarse indicator of magnetic activity which simplifies the environmental situation into 10 digits for the sake of having a simple way to classify a magnetic disturbance. The scale is not linear nor logarithmic but a sort of personal choice of Bartels to have some events in each interval. Specially, in the high numbers, very diverse disturbances are assigned to the same figure (e.g. k=8 in Nimeck could be 331 nT or 500 nT) Moreover, Sq estimation, necessary because it should be removed before k computation, is rather subjective and involve a large uncertainty. In each algorithm, Sq is interpreted differently (Menvielle et al., 1995). The authors claim the necessity of giving k9 values in units of nT or even with tenths of nT (Ln 268-273) but this it would be misinterpreted as if they were very precise when they are not!

The authors agree with the question arose by the referee on the pertinence of having a precise (even on the order of unit of nT) value of K9 when the procedure of obtaining such value has a relatively large margin of subjectivity. Nevertheless, in the era of automatic computation, if a determination of a number must be done, the authors think that the final number should be used as it results from computation or at least rounded at the closest tenth of nT. Indeed, in our statistical analysis a 10 nT resolution for L9 is chosen.

Minor questions

c) Niemek (NGK) was the reference observatory where this scale was defined by Bartels. The rest of other observatories where assimilate to this to create distributions similar to that one. So, the comparison of the Italian observatories with this observatory has more sense than with other German observatory (WNG). In fact, this observatory although being located very close to NGK, has not a perfect correspondence in distribution with (19% deltaK=-1, ln 130).

The referee correctly interpreted the reason why the authors made use of Niemek (NGK) observatory for comparison with the Italian geomagnetic observatories. Nevertheless, for a further comparison, we also used WNG demonstrating that the best correlation is obtained with NGK.

d) However, as K index measures the effect of auroral activity, it seems more reasonable to compare Italian observatories' distribution with other observatories with similar latitude. Moreover, it is well known that k9 limit does not follow a regular law with the angular distance to the auroral zone (Mayaud, 1980).

Once again the referee well interpreted the limits highlighted by the authors when comparisons among observatories located at different latitudes are used. On the other hand, the amplitude of magnetic disturbances has a dependence on (magnetic) local time which affects the K index values (Chambodut et al. (2013)). Since there are not historical observatories located at a similar latitude and, at the same time, not too far in local time, we preferred to use NGK.

e) Although it is true that digital algorithms grant reproducibility (Pg 2 ln 49), this does not mean that they are more certain. In the past, K index was "estimated" for manual procedures; but, now, an automatic algorithm also produces "estimated" values. And, of course, different algorithms would produce different values (pg9 ln255).

f) Comparing the distribution of Italian indices generated by KASM algorithm with NGK and WNG indices generated by FMI algorithm (Ln 137) is a rough way to do this because these distributions change with the algorithm and even with the year being considered (Figure 7).

Points e) and f)

We modified the manuscript introduction at lines 50-57, also according to #2 reviewer's suggestions. Different algorithms for the k index "estimation" produce different values. That's exactly what comes from the ms, simply comparing the distribution of K indices generated by KASM and FMI algorithms, which is probably a granted result but not so universally known. However, our results (Fig.7) show that FMI and KASM are very similar. Indeed, both algorithms are among the 4 endorsed and recommended by IAGA (through ISGI). Maybe an open discussion at geomagnetic community level could be useful to establish new procedures or to revise the old ones, with more critical and fruitful approach.

h) The correlation analysis used to obtain the best value of k9 (figure 4) presents a flat and asymmetric shape in an interval ranking for more than 50 nT. There, any change, as the use of a different algorithm, would produce a different maximum. So, I would not take the new values as a step forward. In fact, final results (those choosing a new value of k9 for DUR (fig. 5) implies a variation of 10% of population in deltaK=-/+1, in the limit of the precision of the method.

The authors show that the correlation analysis allowed to obtain a new L9 value for Duronia observatory, slightly improving the global performance of its K index which depends on the appropriate L9 values. For this reason the authors believe that, even in the limit of the precision of the method and of the comparison proposed in the ms, the final results could be a useful improvement for DUR observatory, as well as for LMP which has to start the K index computation, and a stimulating hint for many scientists involved in the work of geomagnetic observatories.