Translated to English from paper in Japanese, originally published in: Gijutsu Houkoku, Vol.37, No.1, 2 (No.104), 58–68, 1997 with permission of the authors.

Digitization of the K-index Determination Process
— In the Case of the LRNS Method —

by

Yuji YAMADA

1. Introduction

It was approved by the IAGA Working Group V-5 that in addition to the conventionally used hand-scaling method, the following four methods using digital values are to be used as methods for obtaining the K-index from two horizontal components of geomagnetic field.

"USGS"	(Wilson, 1987)
"AS"	(Nowozynski et al., 1991)
"FMI"	(Sucksdorff et al., 1991)
"LRNS"	(Hattingh et al., 1989)

The accuracy of calculation and characteristics of each of these methods had been roughly studied in the research plan until last year (Ozima et al., 1995; Yamada et al., 1996). Ozima et al. (1995) studied the FMI method and discovered that the method is not suited for our use in routine work since the K-indices of the Kakioka Magnetic Observatory and its branches obtained by the method contained very large errors. Yamada et al. (1996) conducted preliminary tests on three other methods and demonstrated that the level of their performance is adequate and there is a possibility that we can use these methods in the routine work. They also indicated that there are cases where the scaling errors become large although this occurs less frequently.

If we use these methods in our routine work, we should select one of the above three methods except for the FMI method. Considering the work of porting and maintenance of the program, a method with a short and simple program is desired. Since we judged that there is no big difference in the performance between these three methods, we selected the LRNS method which has the simplest program (this does not necessarily mean that the working principle is simple) and conducted a detailed research of this method. This document describes the results of this reserch.

2. Determining Parameters

The Linear-phase Robust Non-linear Smoothing (LRNS) method is a method of estimating geomagnetically quiet daily patterns based on geomagnetic data of one given day. Using this method, the K-index is computed from the difference between estimated quiet patterns and original data. To compute the quiet patterns, the LRNS method uses two parameters, μ and r (here the symbol r is used tentatively and is explained only by wording in the original document). Since the amplitudes and patterns of geomagnetically quiet days are differnt by observation points, it is necessary to prepare optimum parameters that allow the most accurate K-index to be obtained based on data at a target observatory. (For example, Hattingh et al. (1989) defined μ as 0.15 and r as 0.07 to determine the K-index at the Hermanus Observatory.) In calculating the K-index at Kakioka, Memambetsu, and Kanoya, the first step is a determination of the parameters to be used for each observatory.

When determining the parameters, it is the most reliable way to calculate the K-index by changing the two parameters by small increments and then to adopt the combination of parameters that results in the most accurate K-index. To determine the parameters, we used the one-minute values of the geomagnetic fields at

Kakioka, Memambetsu, and Kanoya over the period of 12 years from 1985 to 1996. Specifically, we established a total of 70 combinations of μ and r by changing μ from 0.11 to 0.20 in increments of 0.01 and r from 0.04 to 0.10 in increments of 0.01, and took 12-year statistics of K-index errors (the difference between computed K-index values and hand-scaled K-index values) for each combination. The total number of K-index values used for comparison is 35,064 at Kakioka, 35,027 at Memambetsu, and 34,875 at Kanoya. (The slight differences in the total numbers of K-index values are the differences in the numbers of missing one-minute values between these observatories.)

In determining the most appropriate parameters as described above, we face a fundamental problem to be resolved: how we can identify a certain K-index value as being the most appropriate. For example, how the distribution of error frequency changes at Kakioka if μ is changed from 0.11 to 0.20 and r is fixed to 0.05 is shown in Table 1. It is noted from Table 1 that the difference (ΔK) between computed K-index values and hand-scaled K-index values are very small statistically. As shown on the extreme right in this table, the number of K-index values with $\Delta K \leqq 1$ accounts for 94 to 98% of the total number of K-index values. There are rare cases, however, in which the difference becomes extremely large (there are also cases where the difference from the hand-scaled K-index value is 4 depending on a certain combination of parameters). The frequency of such an extreme difference occurring can be decreased by selecting appropriate parameters. When the parameters are as in Table 1, the number of cases in which ΔK is 3 or larger can be reduced to a minimum by limiting the range of μ values to 0.18 to 0.19.

One more point to be noted in Table 1 is that the frequency with positive difference differs from the frequency with negative one. For example, if μ is 0.20, the frequency with negative difference is larger than the frequency with positive one by more than three times. This large deviation means that the computed K-index value by the method is, on average, smaller or larger than the hand-scaled value. To keep the positive and negative differences in balance, in the case of Table 1, we should select of μ values from 0.13 to 0.14.

We find from this table that it is impossible to simultaneously meet the following three requirements even if we manipulate the parameters closely and carefully:

- Maximizing the matching rate with the hand-scaled values
- Minimizing the frequency with which large errors occur
- Optimizing the balance between the number of positive errors and that of negative errors

If one requirement is met, other requirements cannot be met. Therefore, there would be no other choice but to determine the most appropriate parameters that allow each requirement to be met within certain limits.

We established the following conditions to be referenced when determining the parameters.

(Conditions for determining the parameters)

(1) The ratio of the number of cases of $\Delta K > 0$ to that of $\Delta K < 0$ does not exceed 2.

(2) The number of cases of $|\Delta K| > 3$ is zero.

(3) The number of cases of $|\Delta K| = 3$ is minimum among the combinations of parameters that satisfy the above conditions.

Table 1 Frequency distribution of ΔK of the K-index at Kakioka when μ is changed and r is fixed to 0.05. Percent figures on the extreme right are the ratios of the numbers of $|\Delta K| \leq 1$ to the total number for each case.

	⊿K											
μ	-5	-4	-3	-2	-1	0	1	2	3	4	5	%
0. 11	0	0	3	116	4211	22155	6819	1555	188	17	0	94. 6
0. 12	0	0	4	166	4710	22563	6275	1217	123	6	0	95. 7
0. 13	0	0	5	218	5318	22772	5711	965	75	0	0	96. 4
0. 14	0	0	7	248	5926	22932	5180	724	47	0	0	97. 1
0. 15	0	0	8	296	6575	22903	4682	570	30	0	0	97. 4
0. 16	0	0	13	344	7131	22921	4213	427	15	0	0	97. 7
0. 17	0	0	18	406	7714	22758	3833	325	10	0	0	97. 8
0. 18	0	0	21	474	8228	22680	3397	259	5	0	0	98. 0
0. 19	0	0	24	531	8768	22542	3002	195	2	0	0	97. 9
0. 20	0	0	29	575	9399	22240	2676	144	1	0	0	97. 9

Condition (1) is for prevention of the imbalance between positive and negative errors (although the ratio value of 2 is somewhat arbitrary, the level at which the two other requirements are met will degrade if a value more restrictive than 2 is used). Condition (2) is established as the minimum condition that must be satisfied to maintain the accuracy of routine work. The condition will also implicitly maintain the overall matching rate above a certain level.

Table 2 shows the combinations of parameters that are selected from the previously explained 70 parameter combinations and are considered to satisfy the above three conditions at Kakioka, Memambetsu, and Kanoya, as well as the distribution of ΔK value frequencies when the selected combinations are used. Figure 1 is a graph showing the frequency distribution. As shown in Table 2, the numbers of values in which $|\Delta K| \geq 3$ are 28 at Kakioka, 5 at Memambetsu, and 20 at Kanoya. That is, the average number per year is 3 or less at each observatory (1 every 2.4 years at Memambetsu), which is considered to be at a level that will not cause any inconvenience to our routine work.

It should be pointed out that the above conditions for determining the parameters are in no way imperative. It is of course permissible to select only the parameters that cause the number of $\Delta K = 0$ to reach a maximum without giving any consideration to the frequency with which ΔK becomes extremely large or the imbalance between positive and negative ΔK values. We presume that the appropriateness of conditions will be judged from the standpoint of how the K values given by a computer should be utilized effectively for the routine work.

Table 2 Parameters determined for each observatories, and the distribution of frequency of ΔK (the number of K values of $|\Delta K| > 3$ is zero at each observatory and, therefore, it is omitted from this table.)

	μ	r	ΔK -3	-2	-1	0	1	2	3
Kakioka	0.16	0.05	13	344	7131	22921	4213	427	15
Memambetsu	0.19	0.05	5	221	6263	24720	3670	148	0
Kanoya	0.18	0.05	13	362	6894	23182	4101	316	7

3. Other Researches

Figure 1 can be thought of as representing the average accuracy of computed K-index values at each observatory. We examined the errors (ΔK) in more detail. Besides percentage values of ΔK, we used the following two indicators:

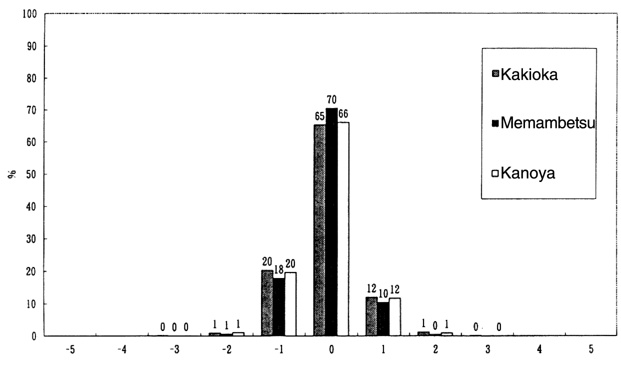
r_1: Ratio (%) of $|\Delta K| > 1$
r_2: Difference between the ratio (%) of positive ΔK and negative ΔK

The value r_1 reflects the "matching ratio," while r_2 represents the imbalance between positive and negative values described in the previous section. How the use of these indicators and the information provided by them will affect routine work has not yet been evaluated. In this section, we limit our discussion to the presentation of the results of a research conducted using these indicators.

a. Seasonal variations

Figure 2 shows how the state of errors changes every month. The solid line represents r_1 and the broken line represents r_2. It can be noted that r_1 hardly changes each month at each observatory, and that the matching rate does not change very much with the seasons. The value r_2 becomes a large negative value in winter, while it comes close to 0 from spring to fall.

The distribution was computed based on the K-indices of Kakioka, Memambetsu, and Kanoya



Errors (difference between the computed and hand-scaled values)

Fig.1 Frequency distribution of the differences between the computed K-indices by the LRNS method and the hand-scaled K-indices.

during the period of 1985 to 1996. The abscissa axis shows the differences and the ordinate axis shows the percentage (%). The number at the top of each bar graph shows a percent value. 0 means the percentage less than 0.5%. When the number of K-index values is exactly zero (0%), a percent value is not shown. The parameters used here were determined for each observatory according to the previously described conditions.

Differences are shown on the abscissa axis and the percentage (%) is shown on the ordinate axis. The number at the top of each bar graph is a percent value. 0 means the percentage less than 0.5%. When the number of ΔK is zero (0%), a percent value is not shown. The values given for N are the numbers of hand-scaled K-index belonging to each value. Dotted lines are drawn at points $K = \pm 1$ to make it easier to see the difference value.

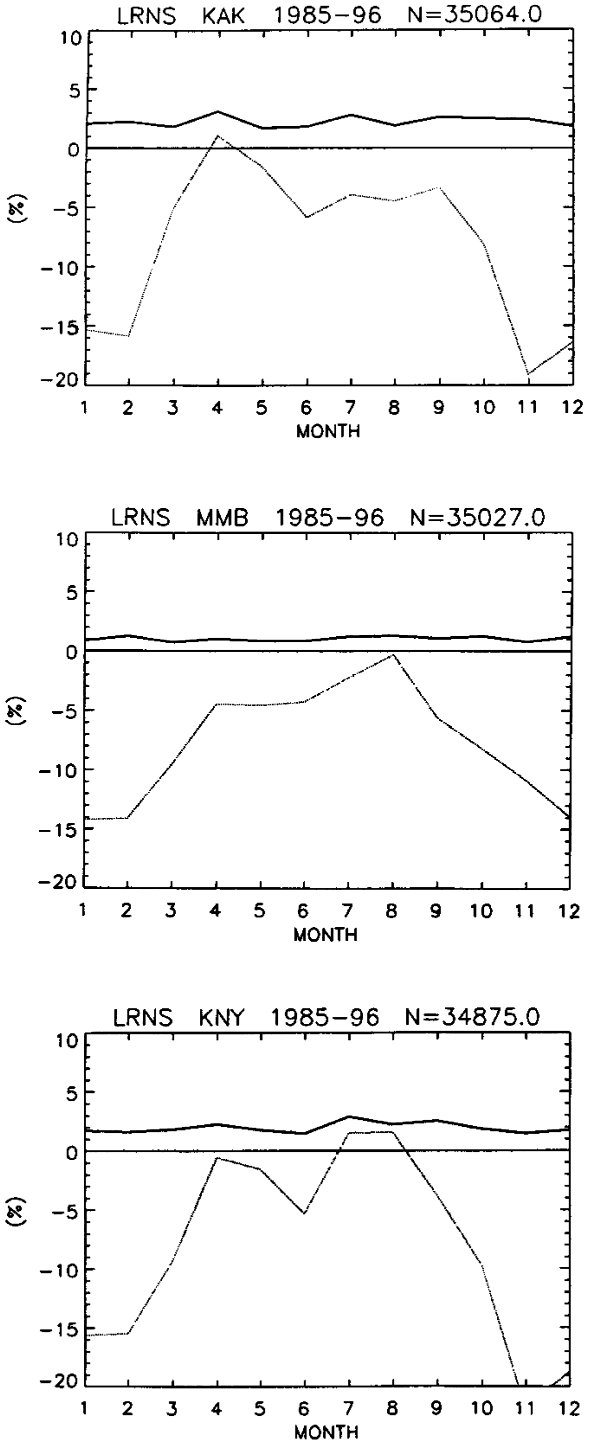


Fig. 2 Seasonal variations of computed K-index values at Kakioka (top), Memambetsu (middle), and Kanoya (bottom).
The solid line represents r_1 and the broken line represents r_2. Months are shown on the abscissa axis and percent values are shown on the ordinate axis.

b. Secular variations

Figure 3 shows how the state of errors changes every year. r_1 changes only slightly every year. Although the variation of r_2 is different at each observatory, peaks appear around 1990, 1991 and 1992. These peaks may be associated with the 11-year cyclic variation of solar activity or there is a possibility that they may have appeared due to hand-scaling errors.

c. Differences depending on the 3-hour intervals

Figure 4 shows the differences between eight 3-hour intervals (0:00 to 3:00 UT, 3:00 to 6:00 UT, ⋯ ., 21:00 to 0:00 UT) where K-index values are determined. Although there are differences between these intervals, the general tendency is that the value of r_1 becomes relatively large during the daytime (sections 1, 2, 3, 7 and 8). The value of r_2 becomes negative in intervals 3, 4, 5 and 6.

d. Differences according to the value of K-index

Figures 5 (a), (b) and (c) show the distribution of errors for each value of hand-scaled K- index (0 to 9) ((a) for Kakioka, (b) for Memambetsu, and (c) for Kanoya). It is natural that errors are distributed toward the positive side when K is 0 or 1. Conversely, they are distributed more toward the negative side when K is 3 or larger. Overall, they are distributed more toward the negative side.

We can find some common characteristics in the distribution of errors at the three observatories shown in (a) − (c), and they are thought to be associated with the calculation methods used here. The possibility that the secular variations of errors in the hand-scaled K-index over years may

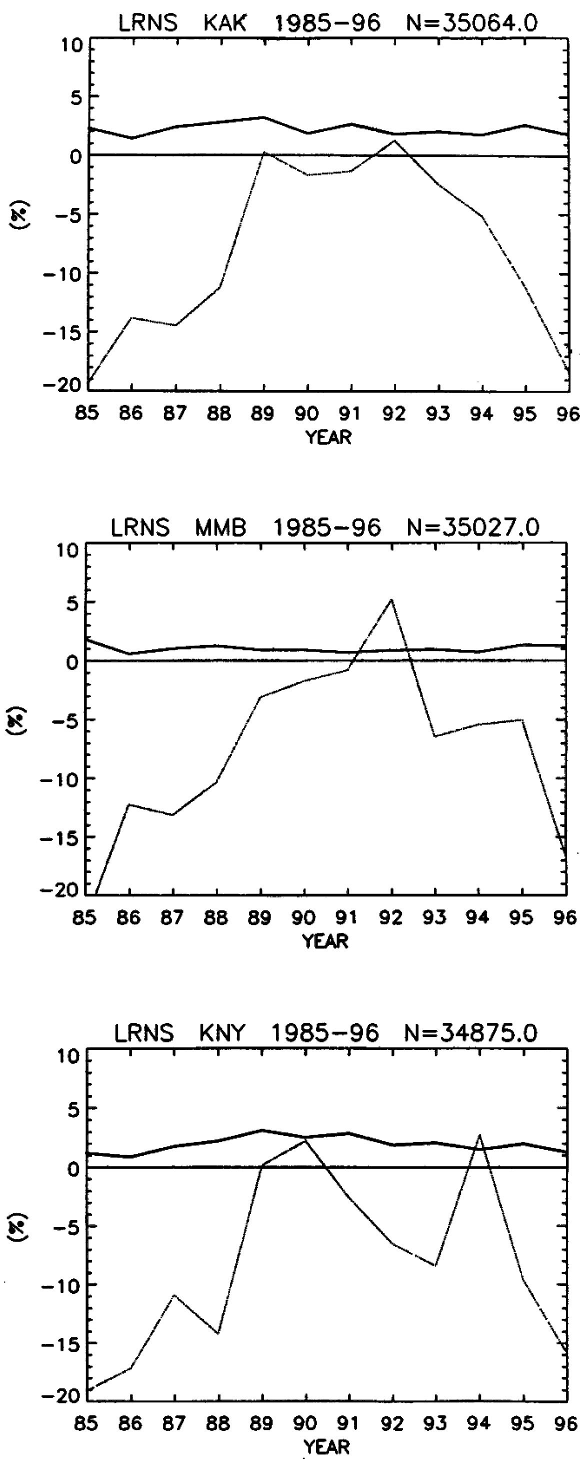


Fig. 3 Secular variations of computed K-index values at Kakioka (top), Memambetsu (middle), and Kanoya (bottom).
The solid line represents r_1 and the broken line represents r_2. Calendar years are shown on the abscissa axis and percent values are shown on the ordinate axis.

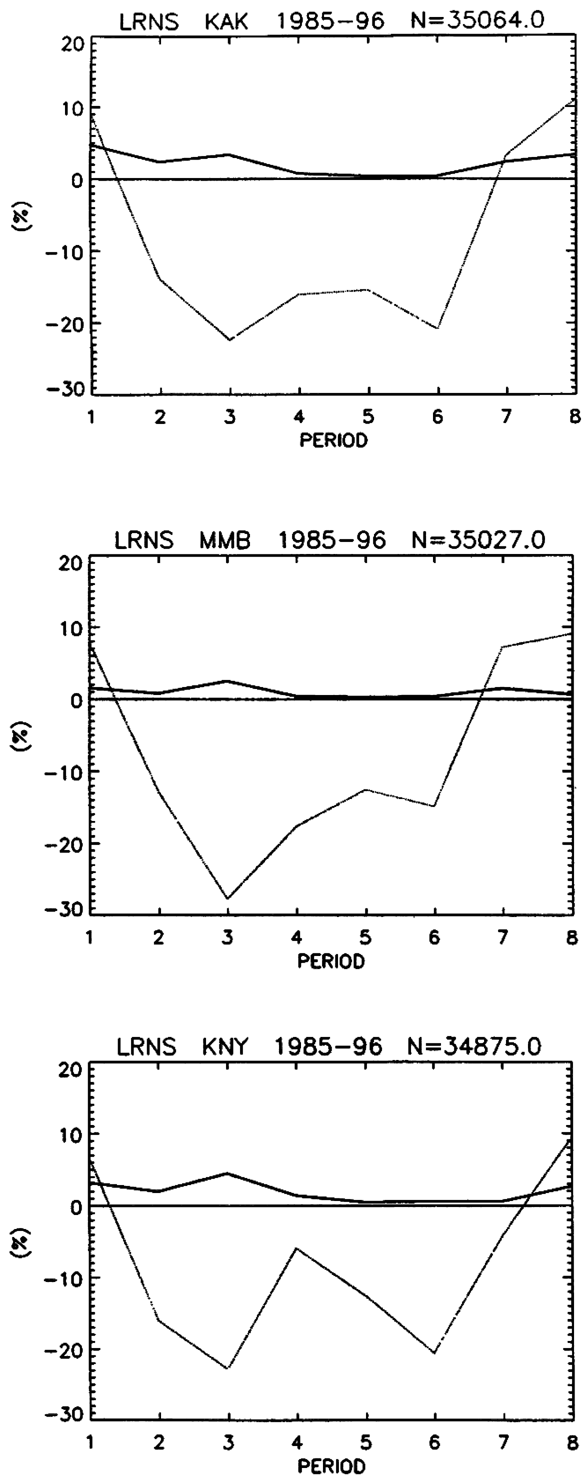


Fig. 4 Differences in computed K-index values for each 3-hour interval at Kakioka (top), Memambetsu (middle), and Kanoya (bottom).
The solid line represents r_1 and the broken line represents r_2. Intervals (1 to 8) are shown on the abscissa axis and percent values are shown on the ordinate axis.

exhibit a certain tendency cannot be denied. Errors are supposed to be included in the imbalance between positive and negative values due to the use of one-minute values of geomagnetic field (it makes K-index always smaller than a true value). But, the amount of the imbalance is yet to be verified because it is in practically impossible to conduct the same comparison using one-second values of geomagnetic field (abnormal values will be mixed in and the amount of data is enormous).

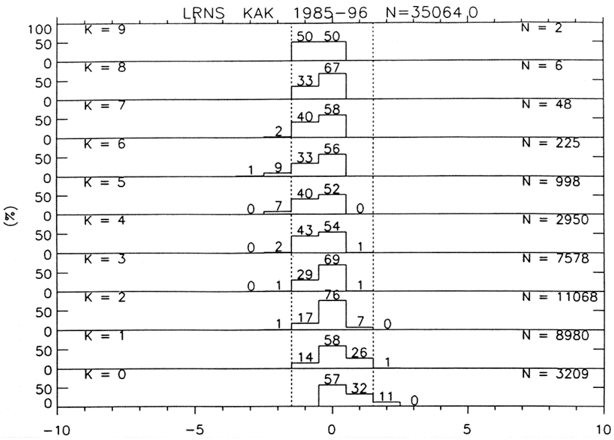


Fig. 5 (a) Frequency distribution of ΔK for each value of hand scaled K-index at Kakioka.

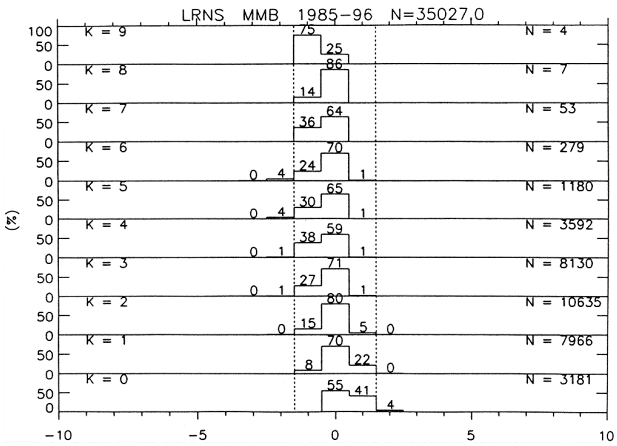


Fig. 5 (b) Same as Fig. 5(a) but for Memambetsu.

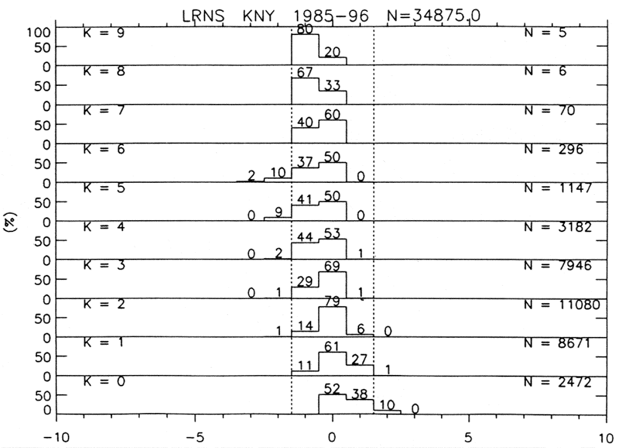


Fig. 5 (c) Same as Fig. 5(a) but for Kanoya.

4. To Use the LRNS Method for Routine Work

To use the computed K-index values obtained by the LRNS method or other methods in routine work, errors resulting from the use of one-minute values must be taken into consideration. Since the use of one-minute values results in a considerably large amount of error in the scaling of geomagnetic phenomena (Shoji et al., 1996), geomagnetic phenomena and K-index values are now hand-scaled by using a chart of one-second geomagnetic values. There are cases in practice that rapid phenomena affect K-index values greatly under disturbed geomagnetic conditions. If a noticeable geomagnetic phenomenon occurs though the frequency of its occurrence is not necessarily high, the temporal resolution of one-minute mean values will be insufficient to determine correct K-index values. Furthermore, K-index values obtained under such disturbed geomagnetic conditions are frequently used by external organizations or institutions. Although the IAGA does not necessarily specify the one-minute value as the "digital data" to be used to compute K-indices (IAGA News, 1993), programs of many of the derivation methods are premised on the use of one-minute values. Because the sampling interval of data to be used can be arbitrarily selected in the program of LRNS method, it is possible to compute the K-index using one-second values. If one-second values are to be introduced, however, another set of parameters different from those for one-minute values must be determined; the work of changing parameters and finding the most suitable parameters through trial and error consumes a much greater amount of time than when parameters are determined by one-minute values. Therefore, it is not realistic to acquire K-indices directly from one-second values.

In computing the quiet daily patterns, however, there will not be much difference between the result acquired using one-second values and that acquired using one-minute values. In determining the K-index based on one-second values, therefore, it will be effective to perform the procedure that, at first, the quiet daily pattern is derived by one-minute values of the day using the LRNS method and then the K-indices are determined as the deviation of one-second values from the quiet daily pattern. Parameters determined based on the one-minute values recorded in the past can be used in such a procedure. If the K-index is to be determined using one-second values, however, artificial noises or missing values must be taken into consideration. Therefore, if computed K-index values are to be routinely used, the both K-indices based on definitive one-minute values and on one-second values should always be computed, and when there is a difference between these two K-index values, the operator intervenes to verify whether or not the difference has resulted from noises in one-second values. This procedure is thought to be practical (see Figure 6).

It seems that the computed K-index values are useful to determine the definitive K-indices for each month. On the other hand, the Observations Division of Kakioka reports the provisional values of K-index every day (URSIGRAM). Can the computed K-index values also be applied to this work? The Observations Division presently issues the provisional K-index values (of past 24 hours) at 0:00 and 6:00 UT every day. A problem arises if we are to replace the provisional K-index values with computed K-index values: data available for the determination of computed K-index values is in half way of a daily variation. In the LRNS method, 24-hour continuous data is needed to determine the K-index value, and it is recommended that data begins at nighttime when the diurnal variation is small. Although the K-index value can be computed based on data that begins at an arbitrary time during the day, it is believed that the accuracy deteriorates. To compute the provisional K-index values, it is necessary to

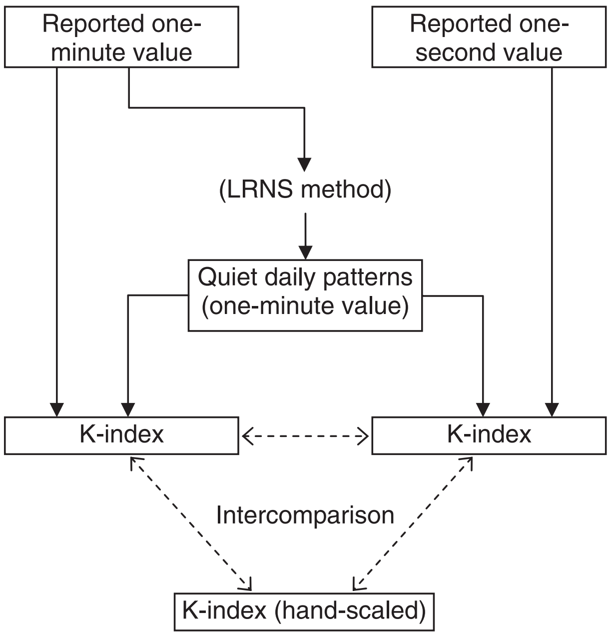


Fig. 6 Example of the K-index determination procedure using the LRNS method.
Quiet patterns are computed based on definitive one-minute values using the LRNS method and the K-index is determined from the difference from the one-minute values. The K-index is also determined based on the quiet patterns and definitive one-second values. There is a considerable amount of error with the K-index determined based on reported one-minute values during the period when rapid phenomena predominate. On the other hand, there is a possibility that artificial disturbances or missing values may have been mixed into the definitive one-second values. To cover the shortcomings of these one-minute and one-second values, both values (and hand-scaled K-index) must be compared to determine the definitive K-index.

check in advance with what level of accuracy the K-index value can be determined if the data begins at 0:00 or 6:00. This point must be further examined and clarified. (Even if a proper level of accuracy can be achieved and the K-index can be computed using digital values that day, the computed K-index values must be manually input into a terminal for data transmission by reason of the limitations of the currently used system. Therefore, there will not be much merit in terms of labor-saving.) The FMI and USGS methods are not suited to computing the provisional K-index values, because data on the days before and after the day is necessary to compute the K-index value for these two methods.

References

Hattingh, M., L. Loubser, and D. Nagtegaal, Computer K-index estimation by a new linear-phase, robust, non-linear smoothing method, *Geophys. J. Int.*, **99**, 533–547, 1989.

IAGA News, No. **32**, pp. 27–28, Dec., 1993.

Nowozynski, K., T. Ernst, and J. Jankowski, Adaptive smoothing method for computer derivation of K-indices, *Geophys. J. Int.*, **104**, 85–93, 1991.

Ozima, M., A. Yamazaki, and T. Takizawa, Research for digitizing the K-index determination process, Report of research plan of Kakioka Magnetic Observatory in 1994, 1995 (in Japanese).

Shoji, T., M. Hoshino, S. Teshima, and S. Tsunomura, Research concerning the production of the geomagnetic variation records using the KASMMER one-second values, Report of research plan of Kakioka Magnetic Observatory in 1995, 1996 (in Japanese).

Sucksdorff, C., R. Pirjola, and L. Häkkinen, Computer production of K-values based on linear elimination, *Gephys. Trans.*, **36**, 333–345, 1991.

Wilson, L. R., An evaluation of digitally derived K-indices, *J. Geomag. Geoelectr.*, **39**, 97–109, 1987.

Yamada, Y., T. Takizawa, T. Shoji, S. Teshima, and A. Yamazaki, Research for digitizing the K-index determination process, Report of research plan of Kakioka Magnetic Observatory in 1995, 1996 (in Japanese).