

DANISH METEOROLOGICAL INSTITUTE
SCIENTIFIC REPORT

00-13

**Stability Indices as Indicators of
Lightning and Thunder**

September 2000

M. Steffensen

ISSN 1399-1949
ISBN 87-7478-424-2



Copenhagen 2000

1. Introduction

Severe weather often has a large impact on human activity in many different areas, for instance transport and agriculture, to mention a few. The implementation of new automatic observation systems with a much denser data coverage and the increasing resolution and improvements of the numerical limited area models provide a basis for the development of new automatic or partly automatic methods for forecasting severe weather. A project (PVV, Projekt voldsomt vejr) to improve on the forecasts of severe weather and to investigate the extent to which these can be automatically produced has been initiated at DMI.

A preliminary study in this project has been to investigate the potential use of stability indices to indicate the risk of thunder and lightning. The indices are used as measures of the necessary atmospheric conditions for thunderstorms to occur. Other studies on the use of indices as a measure of the necessary atmospheric condition for thunderstorms in Europe can be found in (H.Huntrieser 1996) and (H.Haakenstad 1999).

2. Data

The stability indices have been calculated from radio sounding from Schleswig 12 UTC. For each radio sounding the various stability indices have been calculated and stored together with the number of lightning strokes recorded in an area around Schleswig during a certain time window.

The data sample covers the period from 02-11-95 to 30-08-1999. This data sample has been divided into two sub periods, one from 02-11-1995 to 02-11-1997 and a second from 03-11-1997 to 30-08-1999. Data from the first period are used to determine the parameters in a probabilistic indication model and data from the second period are used to verify the probabilistic model.

Data from an automatic lightning stroke detection system have been used to identify days with lightning. The system can detect the occurrence of a lightning stroke and then calculate the position and strength. The number of lightning strokes occurring in an area around Schleswig during a time window of 3 hours from 12 to 15 UTC are counted and stored together with the value of the indices on that particular day.

To investigate the dependence of the area, the number of lightning strokes occurring in five areas of different size has been counted. The sizes of the areas are listed in table 1.

Table 1

| Area A0 | Area A1 | Area A2 | Area A3 | Area A4 |
|------------------------------|----------------------------|----------------------------|----------------------------|----------------------------|
| 133 * 129 [km ²] | 83 * 96 [km ²] | 55 * 61 [km ²] | 28 * 33 [km ²] | 11 * 13 [km ²] |

To investigate the dependence on time a 6 hour window from 12 to 18 UTC have been used together with area A0 and A1.

Supplementary to the automatic lightning detection system, conventional synoptic reports have been used to identify days with thunder. One data set is based on the station Schleswig alone for times 12, 13, 14 and 15 UTC. A second set is based on stations in an area corresponding in size to area A1 in the automatic data set for the same times. Thunder is indicated in this data set if at least one of the stations reports thunder. The stations included are listed in table 2.

Table 2

| | | | |
|-------|------------------|-------|-----------------------|
| 6110 | Vojens | 10042 | Olpenitz |
| 6119 | Kegnæs | 10046 | Kiel |
| 6124 | Tåsinge | 10131 | Cuxhaven |
| 10020 | List | 10124 | Leuchtturm alte Weser |
| 10022 | Leck | 10146 | Quickborn |
| 10028 | St. Peder-Ording | 10150 | Dorisk |
| 10035 | Schleswig | 10156 | Lübeck |

3. Comparison of automatic lightning detection and conventional synop reports

Simultaneous data from the automatic lightning detection system and from the synoptic reports have been compared to investigate the correspondence between the two methods of detecting lightning and thunder.

The synoptic data from Schleswig are compared to the areas A1, A2, A3 and A4 in the automatic data set. The results are summarised in table 3, where column one shows the number of days with both lightning detected by the automatic system and thunder in the synoptic reports, column two shows the number of days with thunder in the synoptic reports but no lightning detected by the automatic system, column three shows the number of days with lightning detected by the automatic system but without thunder in the synoptic reports, and finally column four shows the number of days where neither systems indicates either lightning or thunder. The last column shows the total number of days with lightning detected by the automatic system. The number of days with observed thunder in the synoptic reports at Schleswig is a constant number of 18.

Table 3

| | Both lightning (automatic) and thunder (synop) | Thunder (synop) | Lightning (automatic) | No thunder/lightning | Total |
|----------------|--|-----------------|-----------------------|----------------------|-------|
| Schleswig | | | | | |
| A4 one or more | 11 | 7 | 10 | 690 | 21 |
| A4 two or more | 7 | 11 | 7 | 693 | 14 |
| A3 one or more | 17 | 1 | 31 | 669 | 48 |
| A3 two or more | 14 | 4 | 16 | 684 | 38 |
| A2 one or more | 16 | 2 | 40 | 660 | 56 |
| A1 two or more | 18 | 0 | 83 | 617 | 101 |
| All stations | | | | | |
| A1 two or more | 65 | 17 | 60 | 836 | 125 |

First we may notice the obvious fact that the number of automatic detected lightning occurrences is increasing as the area of detection is expanded.

In the first row in table 3 *one or more* lightning strokes detected by the automatic system in area A4 defines a day with lightning and in the second row *two or more* lightning strokes detected by the automatic system defines a day with lightning. That is, a day with only one lightning stroke detected is disregarded as a day with lightning.

For days with one or more lightning strokes detected in area A4 (row one) there are eleven days with both lightning strokes detected and thunder in the synoptic reports, and there are seven days with thunder in the synoptic reports but no lightning strokes detected. The latter may be because the thunder occurs outside the small area A4. The number of days with lightning strokes detected but without thunder in the synoptic reports is ten. These could be days with only one lightning stroke detected by the automatic system.

To look into this, row two in table 3 shows the number of days with two or more lightning strokes detected by the automatic system. Now only seven days have lightning strokes detected but no thunder in the synoptic reports, thus the number of days with only one lightning stroke detected is three.

On the other hand it turns out that the number of days with both lightning detected and thunder in the synoptic reports is reduced from eleven to seven, thus four of the eleven days with both lightning detected and thunder in the synoptic reports are also days with only one lightning stroke.

Another explanation could be that only the synoptic code “ww” has been used to identify days with thunder. Including “WW” for the past weather would increase the number of days with thunder in the synoptic reports.

The third row in table 3 shows the values for area A3 for days with one or more lightning strokes detected by the automatic system. Here 17 days have both lightning strokes detected and thunder in the synoptic reports and only one day with thunder in the synoptic reports is not detected by the automatic system. This indicates that the area A3 is large enough for the automatic system to capture all thunder in the synoptic reports. 31 days with lightning strokes detected are not in the synoptic reports.

Row four in table 3 shows the values for A3 and two or more lightning strokes detected. Now the number days with both lightning strokes detected and thunder in the synoptic reports is reduced only from 17 to 14, whereas the number of days with lightning strokes detected but without thunder in the synoptic reports is reduced significantly from 31 to 16. The remaining 16 days may be further reduced in number if “WW” is used to identify days with thunder in the synoptic reports.

It seems therefore reasonable to conclude that the thunder recorded by a human observer is restricted to thunder occurring within a distance of 30 km, and that days with single lightning at large distances are not recorded by a human observer. However still approximately half of the days with lightning strokes detected have no thunder in the synoptic reports.

Increasing the area further (rows five and six in table 3), only the numbers of days with lightning strokes detected and not recorded by a human observer increases significantly, supporting the conclusion that the thunder recorded by a human observer is restricted to 30 km. This agrees well with the distance calculated from the refraction of sound waves at the interface of adjacent thin layers of air with different velocities of sound, ($v_{\text{sound}} \sim T^{1/2}$).

For area A1 the number of days with lightning strokes detected but without thunder in the synoptic reports is 83 (row five in table 3). The previous conclusion suggests then that a number of these days only have thunder more than 30 km away from Schleswig. If the stations in table 2 covering the area A1 are used to identify days with thunder from the synoptic reports when at least one of the stations reports thunder, row six in table 3 shows, that 65 days have lightning strokes detected by the automatic system and thunder in the synoptic reports, and that again approximately half of the days equal to 60 with lightning strokes detected are not found in the synoptic reports. This further supports the previous conclusion.

Again some of the days with lightning strokes detected but without thunder in the synoptic reports may be due to the used the code “ww” in identifying thunder in the synoptic reports.

However, 17 days with thunder in the synoptic reports are not detected by the automatic system. This may be due to the difficulty in matching the number of stations and the area of automatic detection so that some of the thunder in the synoptic reports have lightning strokes outside the area A1.

It may not be possible to extract some decisive conclusion about the correspondence of observations recorded by a human observer and automatic observation, but it seems to be possible to establish a relation between the thunder in the synoptic reports and automatically detected lightning strokes if one uses appropriate areas.

4. Indices

In this preliminary study a number of different indices related to thunder and lightning have been investigated. The indices are listed in table 4 with their definition and a short description.

Table 4

| | | |
|-------------------------|--|-----------------|
| K | $(T_{850} - T_{500}) + TD_{850} - (T_{700} - TD_{700})$ | Air mass index |
| T | $(T_{850} + TD_{850}) - 2 * T_{500}$ | Air mass index |
| SHO | $T_{500} - T'_{850 \rightarrow 500}$ | Lifting index |
| SI ₇₀₀ | $T_{500} - T'_{700 \rightarrow 500}$ | Lifting index |
| SLI | $T_{500} - T'_{sur \rightarrow 500}$ | Lifting index |
| DCI | $(T_{850} + TD_{850}) - SLI$ | Deep Convection |
| HI | $(T_{850} - TD_{850}) + (T_{700} - TD_{700}) + (T_{500} - TD_{500})$ | Humidity |
| CAPE _{surface} | Integral of $(T_{lift} - T_{env})/T_{env}$ when $T_{env} < T_{lift}$ | |
| CAPE ₈₅₀ | Integral of $(T_{lift} - T_{env})/T_{env}$ when $T_{env} < T_{lift}$ | |

Here T_{ppp} is the temperature at pressure level PPP and TD_{ppp} is the dew point temperature at pressure level PPP. $T'_{ppp \rightarrow 500}$ is the temperature of a parcel lifted dry adiabatically from pressure level PPP (or the surface in case PPP is equal to sur) to its condensation level and moist adiabatically thereafter to the 500 hPa pressure level.

CAPE_{surface} is the Convective Available Potential Energy where T_{lift} is the potential temperature of a parcel lifted dry adiabatically from the surface to its condensation level and moist adiabatically thereafter. T_{env} is the potential temperature of the environment. CAPE₈₅₀ is the same as CAPE_{surface} except that the parcel is lifted from 850 hPa.

The temperatures are from measured radiosoundings

5. Statistics of lightning and indices

5.1 The number of lightning occurrences as a function of the daily value of the index

The area A1 83*96 [km²] and the 3 hour window have been used as the base case and all results refer to this area and time window unless something else is explicitly stated.

Figure 1 shows the number of lightning strokes detected versus the index value of K (left hand side of figure 1) and the index value of SLI (right hand side of figure 1) for all days in the period from 02-11-95 to 30-08-1999. The red dots represent days with lightning and the blue dots represent days without lightning (The number of lightning strokes detected is 0 for such a day). Similar plots have been made for the other indices.

There is no significant mathematical correlation between K values and the number of lightning strokes because there are many days with no lightning strokes and K values similar to the K values for days with lightning strokes. On the other hand, it is seen by visual inspection that days with lightning tend to have a large K value and, though not quite as pronounced, that days with many lightning strokes detected tend to have the largest K values.

A large K indicates a large temperature gradient and a large humidity in the lower atmosphere, both favourable conditions for convection and thunder. This is also the case for the index TT, which has the same pattern as the K index (not shown).

The opposite is seen for the SLI index, where days with lightning tend to have small values and days with many lightning strokes detected tend to have the smallest SLI value. A large temperature $T'_{sur->500}$ of the air parcel lifted indicates favourable conditions for convection and results in a small value of the SLI index. This holds for all the lifting indices, which have a similar pattern in the number of lightning strokes detected versus the index value (not shown).

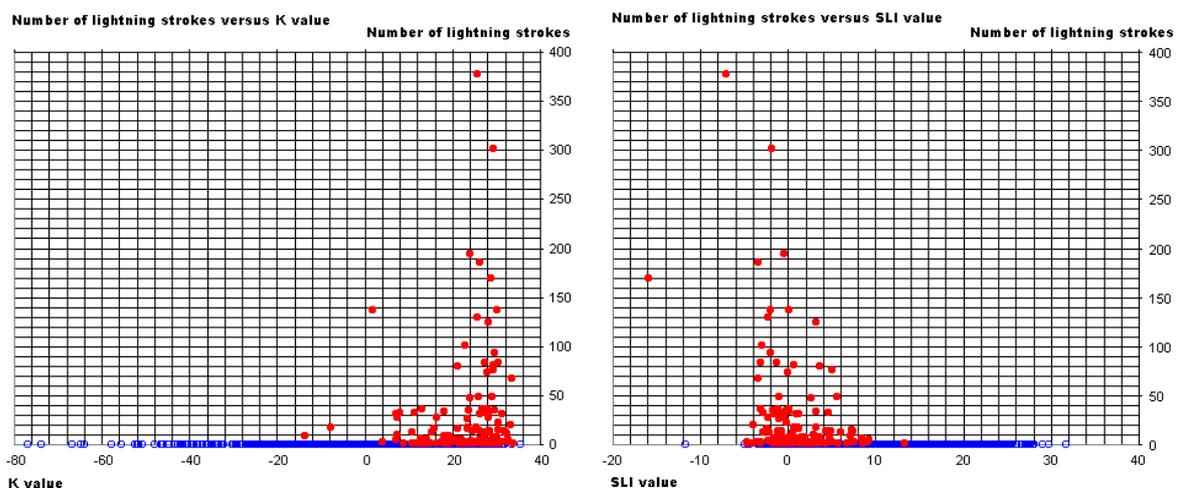


Figure 1. The number of lightning strokes detected in area A1 during 3 hours as a function of the index value. Left hand side as a function of the K index. Right hand side as a function of the SLI index.

DCI is a combination of humidity in the lower atmosphere and a lifting index. The favourable conditions for convection and thunder is large humidity together with a small lifting index, which then combine to a large value of DCI for days with lightning, which is also confirmed in the data (not shown)

The humidity index HI is a measure of the dew point depression in the atmosphere below 500 hPa. A small HI is favourable for thunder.

5.2 Frequency of days with lightning as a function of index value.

The relation between the number of days with lightning and the index value is more clearly demonstrated if the frequency of days with lightning is plotted against the index value. Each column in figure 2 shows the number of days with lightning having K and SLI values within the interval indicated under the column divided by the total number of days with lightning. This and all subsequent analyses are based on the first set of data from the period 02-11-1995 to 02-11-1997.

Notice that the lengths of the intervals are unequal, since the first interval for the K index runs from -80 to -20, and the second interval runs from -20 to 0. This is because of the low number of days with lightning in these intervals. Similarly for the SLI index the length of the intervals are enlarged at large index values.

It is clearly seen that days with lightning tend to have large K values, with a relative maximum of 33% of all days with lightning having K values in the interval from 25 to 30.

For the SLI index days with lightning tend to have small values, with almost 35% of all days with lightning having SLI values in the interval from -2 to 0.

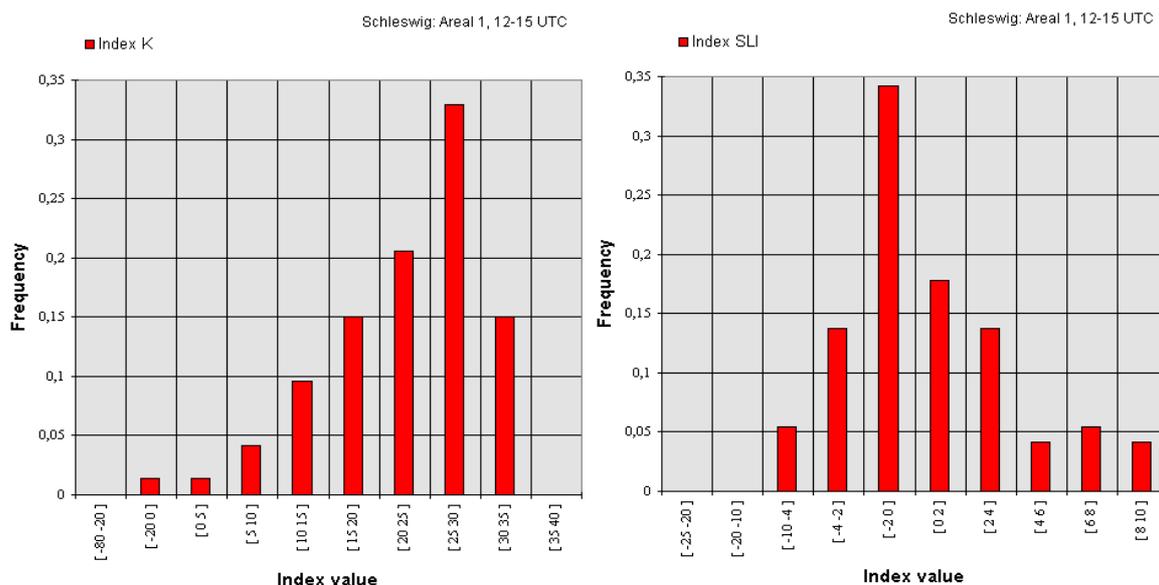


Figure 2. The fraction of days with lightning as a function of the index value. Left hand side as a function of the K index. Right hand side as a function of the SLI index.

Furthermore both frequencies are relatively narrow, which is necessary if the index is used as indicator of lightning, see later.

The other indices show similar patterns, but with a more or less broadened distribution.

5.3 The relative frequency of days as a function of index values.

In the previous section the frequency of days with lightning as a function of the index value were analysed. It is, however, also interesting to know the relative frequency of days as a function of the index value. In this section both the frequency of days and the frequency of days with lightning relative to the total number of days is studied as a function of the index value.

In figure 3 the blue columns show the number of all days with the index values within the interval below the column divided by the total number of days in the data sample. The red columns show the corresponding number of days with lightning again divided by the total number of days in the data sample. Note that this fraction is different from the frequency presented in the previous section since this fraction is now relative to the total number of all days in the data sample whereas the previous frequency was relative to the total number of days with lightning.

Notice that the remark about the interval length given in the previous section is also valid in this figure, which more pronounced affects the blue column representing all days, since a large fraction of days have index values in these intervals. K values between -20 and 0 occur in 24% of all days and the SLI values between 10 and 50 occur in more than 37% of all days. However, in both cases lightning seldom occurs on days with index values in these intervals, which shows that the occurrence of lightning is not an everyday phenomenon.

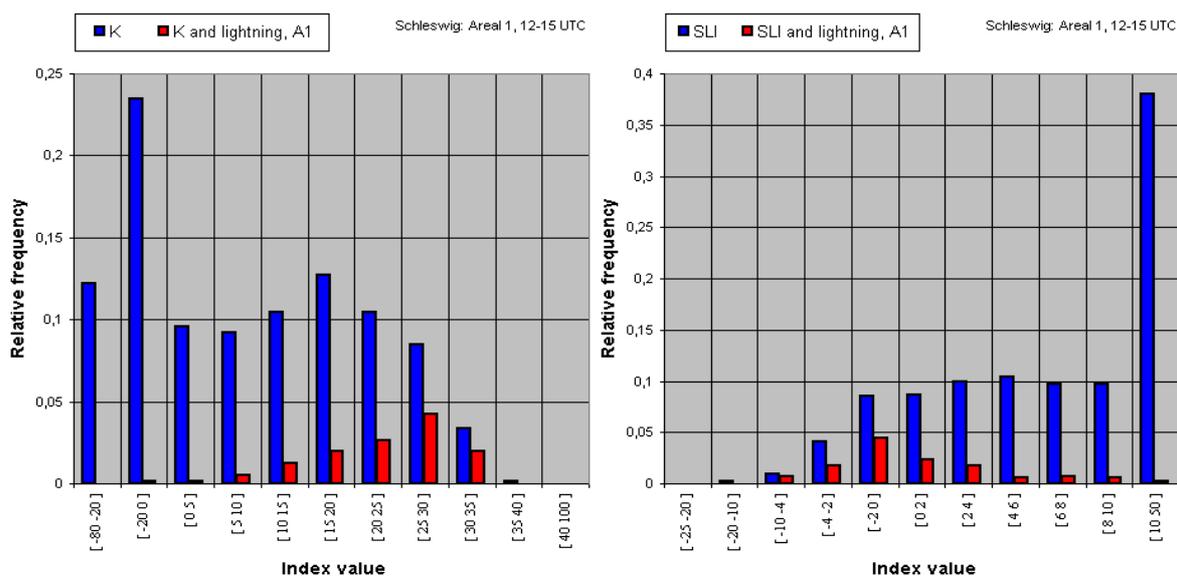


Figure 3. The frequency of days as a function of the index value. Blue columns represent all days and red columns represent days with lightning. Left hand side as function of the K index. Right hand side as a function of the SLI index.

In fact, the area of the red and the area of the blue columns are proportional to the occurrence of days with and without lightning respectively. The area of the red columns is significantly smaller than the area of the blue columns. The fraction of the area of the red column to the area of the blue column is approximately 0.12 which thus corresponds to the climatological frequency of lightning in the data sample.

The interesting feature, however, is that the relative frequency of all days in many of the intervals also having a significant relative frequency of days with lightning, are a great deal larger than the frequency of days with lightning for that interval, though varying from interval to interval. The importance of this feature is that although days with lightning tend to have index values in these intervals the opposite statement, that a day with an index value in these intervals will result in a day with lightning, is unfortunately not valid. In other words, some days without lightning also have index values in these intervals.

The same analyses have been made for other indices. For these the distributions of the days with lightning are more flat with no pronounced maximum.

5.4 Threshold value as an indicator of lightning.

The relative frequencies of all days and days with lightning as a function of the value of an index can be used to examine an indicator for the occurrence of lightning a given day defined by a threshold value for that index.

Figure 4 is a summary of the performance of the K index as a function of the threshold values indicated under the columns.

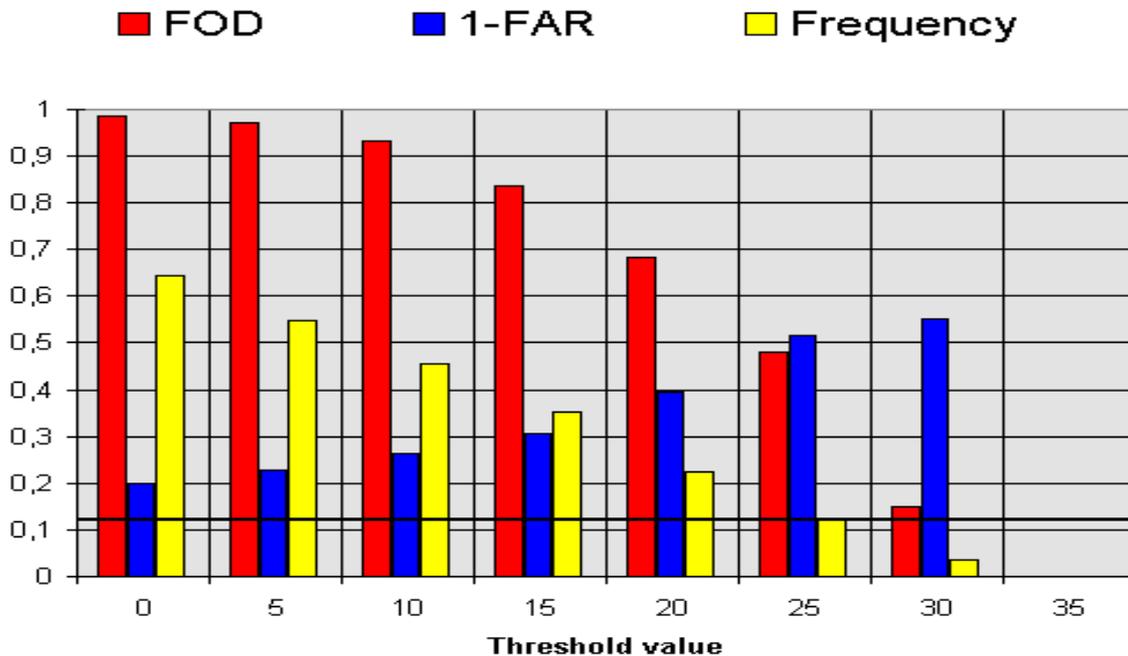


Figure 4. Summary of the performance of the K index as an indicator of the occurrence of lightning defined by the threshold value indicated under the columns, see the text for further explanation.

The curves are read as follows: Assume for instance that K greater than 15 is chosen as threshold value. The red curve then shows that more than 80% of all days with lightning are indicated, which on the other hand means that around 20% of all days with lightning is not indicated and thus referred to as surprises (or misses). The blue curve shows that in only 30% of the days indicated as days with lightning, lightning actually occur or, equivalently, that 70% of the days are falsely indicated (false alarm). The yellow curve shows that a little less than half (35 %) of all days are indicated as days with lightning. This should be compared to the climatological frequency of only 12 %.

If the threshold value is increased to 25 the yellow curve reads about 12% so that days with lightning are indicated with the correct frequency. Now in about 50% of the days indicated as days with lightning, lightning actually occur, or, equivalently, only 50% of the days are indicated wrong (false alarm) which is an improvement, but then only 50% of all days with lightning are indicated in stead of 80%, which means that the frequency of detection has gone down.

5.5 Probability of lightning as a function of index value

A more fruitful description of the relation between days with and without lightning may be defined as the relative frequency of days with lightning divided by the corresponding relative frequency of all days. For each interval this fraction can be interpreted as the probability for the occurrence of lightning given that the particular day has an index value in the interval. Figure 5 shows the probability for the K and the SLI indices.

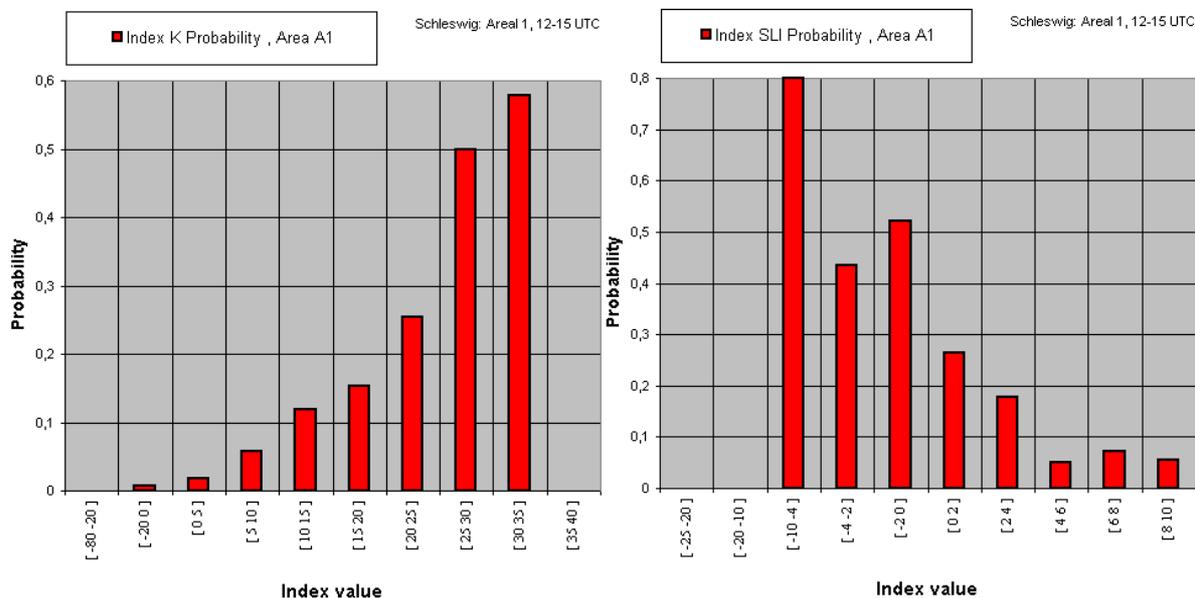


Figure 5. The probability as a function of the index value. Left hand side as a function of the K index. Right hand side as a function of the SLI index.

For the K index the probability is an increasing function of K values with a maximum just below 60% for K values in the interval from 30 to 35. The probability decreases as a function of the SLI index with a maximum of 0.8 for SLI values in the interval from -10 to -4.

This method for estimating the probability as a function of the index value has been applied to the first half of the data covering the period from 02-11-1995 to 04-11-1997 with 567 days having useful data for the index K. The probability $p(K)$ determined from these data has then been approximated by an exponential function of the form $e^{(a_0 + a_1 * K)}$ limited by a maximum probability of 1, where the coefficients a_0 and a_1 are found by minimising the sum of $(I_i - \min(e^{(a_0 + a_1 * K_i)}, 1))^2$ with I_i equal to 1 if it is a day with lightning and 0 if not. The sum is proportional to the Brier score and the procedure is similar to logistic regression.

The probability function thus found have subsequently been applied to the second half of the data covering the period from 04-11-1997 to 30-08-1999 with 549 days having useful data.

The probability function has then been verified through the calculation of the reliability curve and the sharpness. The results are shown in figure 6.

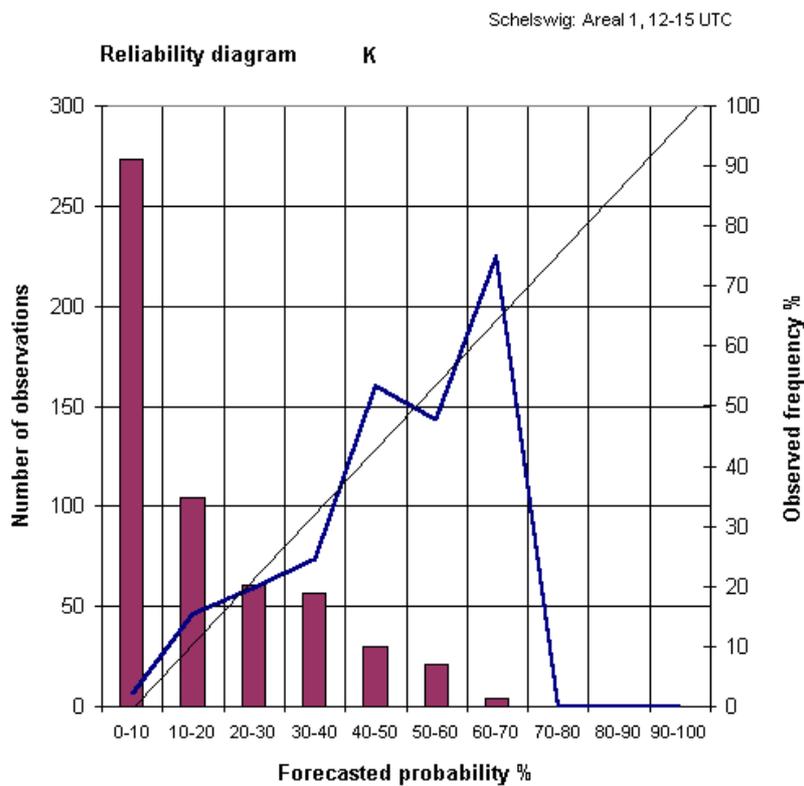


Figure 6. Reliability curve for the probability based on the K index.

Since the highest probability is less than 70% and since the occurrence of this value is rare, the model is not very efficient at indicating thunder/lightning. A better model should have more cases with high and reliable probability, that is a U shape of the columns, since this is important seen from a users point of view.

The Brier score for the model (BS_{model}) and for the climatology of the sample (BS_{climate}) and for persistence ($BS_{\text{persistence}}$) defined such that the probability of lightning today is equal to 1 if lightning occurred yesterday and 0 if not, have been calculated. The Brier skill scores with the sample climatology as reference for the model ($SS_{\text{model vs. climate}}$) and the persistence ($SS_{\text{persistence vs. climate}}$) have also been calculated and the results are shown in table 5.

Table 5.

| | | | |
|-----------------------------|-------|---|-------|
| (BS_{model}) | 0.096 | $(SS_{\text{model vs. Climate}})$ | 0.20 |
| (BS_{climate}) | 0.120 | | |
| $(BS_{\text{persistence}})$ | 0.175 | $(SS_{\text{persistence vs. Climate}})$ | -0.45 |

It is clear from table 5 that the probability model is superior to both climate and persistence. Persistence is in this case not a very useful approach in indicating lightning

5.6 The dependency of the relative frequency of days on the season.

The relative frequency of days as a function of the K index value has been calculated for winter and summer months separately. Figure 7 shows the frequencies for the winter months October to April and for the summer months May to September.

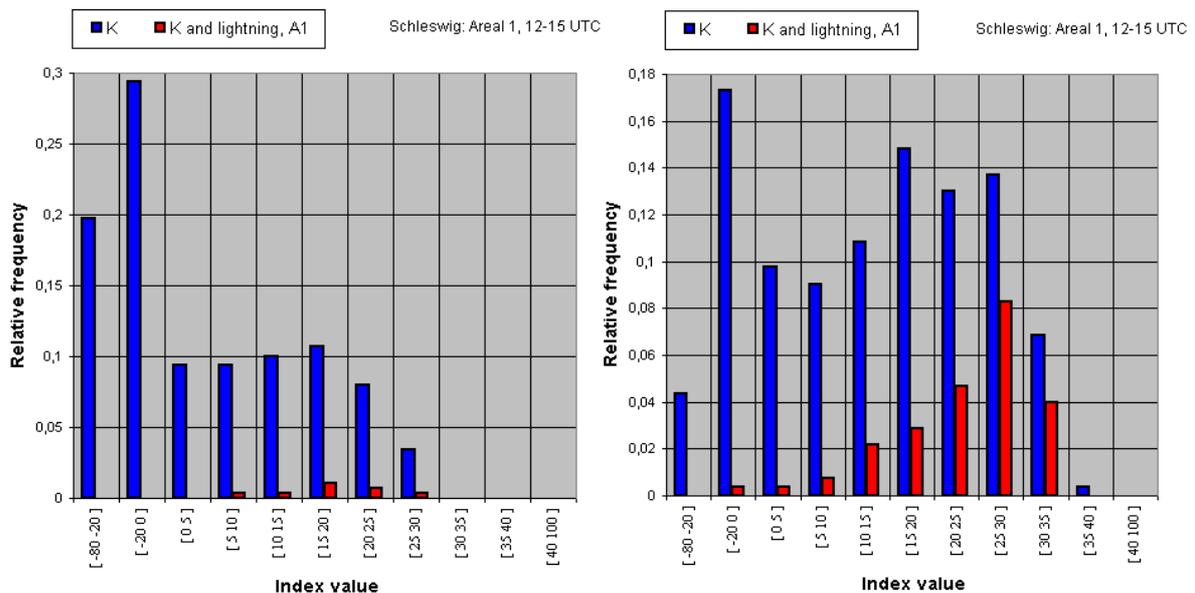


Figure 7. Frequencies for winter months October to April and for the summer months May to September.

The most important information in this figure is that the relative frequency of days with lightning is extremely low in the winter months, so that the most of the days with lightning occurs in the summer months. Table 6 shows the number of days with lightning for each month for data in the period 02-11-1995 to 04-11-1997.

Table 6

| Month | Total number of days in data | Number of days with lightning | Month | Total number of days in data | Number of days with lightning |
|----------|------------------------------|-------------------------------|-----------|------------------------------|-------------------------------|
| January | 41 | 0 | July | 58 | 16 |
| February | 35 | 2 | August | 58 | 14 |
| March | 29 | 0 | September | 49 | 5 |
| April | 30 | 1 | October | 49 | 2 |
| May | 56 | 13 | November | 51 | 3 |
| June | 56 | 17 | December | 45 | 0 |

5.7 The dependency of the relative frequency of days on the size of the area of detection

The influence of the size of the area in which lightning is detected has been studied by using five different areas. If the area is increased, the total number of days with lightning will increase, and for each interval of the index values the number of days with lightning will be constant or increase.

The following analysis is carried out for the K index, and it is assumed that large K values correspond to more favourable conditions for lightning.

For large values of K, that is favourable conditions for lightning, the number of days with lightning is expected to increase fast when going from very small areas to larger areas and then become more constant during further increase.

To outline this a little further, assume that K is large and that in fact lightning occurs. Assume further for simplicity of the arguments that the lightning strokes are equally spaced. When a small area is placed randomly, lightning is likely to lie outside the area. Increasing the area increases the likelihood that the lightning is inside the area. When the size is large enough, lightning will occur inside the area. Further increase will not result in more days with lightning since lightning is already detected, however the number of lightning strokes will increase. If the area is increased even further, the air conditions in the newly covered regions may be different from the K value at the centre, resulting in a smaller K value, and thus no lightning expected here.

For small values of K, that is unfavourable conditions for lightning, the number of days with lightning for a small area is small. When the size of the area is increased, new regions with a possibly larger K value and thus a possibility of lightning may increase the number of days with lightning, no matter how big the area becomes.

The influence of increasing the size of the area on the probabilities is the same as the influence on the number of days with lightning, because the probability is just this number divided by the total number of days with K values in the interval which is a constant for all

areas. The considerations given for the number of days with lightning holds for the probabilities as well.

Figure 8 shows the probability for K values for all five areas with A4 being the smallest and A0 being the largest. For the intervals going from 30 to 35 and from 25 to 30, that is large K values, the probability rises fast when going from A4 to A1, whereas the probability for A1 is just a little less than for A0. For the intervals from 10 to 15 and from 15 to 20, smaller K values, the probability still rises fast when going from A4 to A1, and here the probability continues to rise when going from A1 to A0.

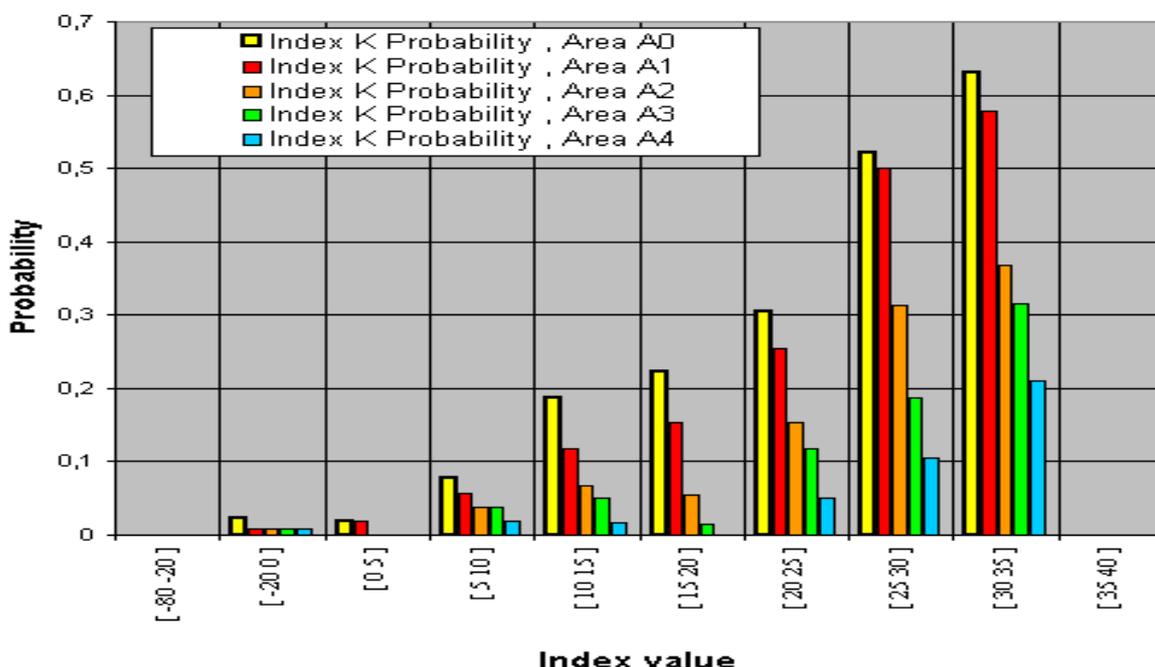


Figure 8. Probability as a function of the K index for areas A0 to A4.

It is not clear how to choose the size of the area. In this study the philosophy has been to obtain the smallest area and the largest maximum probability. The probability can be 1 if the area is chosen large enough (for instance the whole globe). A1 has been used as base case because further increase of area will not result in significantly larger maximum probability, whereas the area A2 will decrease the maximum probability.

The area A1 has linear dimension about 100 km, which corresponds to regional forecast areas.

5.8 A combined index.

The previous analysis has shown that the K value may be large without the occurrence of lightning on that particular day. Now two or more indices may be combined to give a new index which separate the days with and without lightning better. Figure 9 shows the SLI

index plotted against the K index. As known from the previous analysis, K should be large and SLI should be small when lightning occurs. Figure 9 shows however that some days with large K value simultaneously have a large SLI value when no lightning is detected. These days lie in the upper right quarter of the grey rotated co-ordinate system. If a new index is defined as $\text{new} = (K - 4 * \text{SLI})$, then these days will reduce the index value and thus the probability. The relative frequency of days has been estimated for this new index and is shown in the left hand side of the figure 10. The corresponding probability is shown in the right hand side of the figure 10.

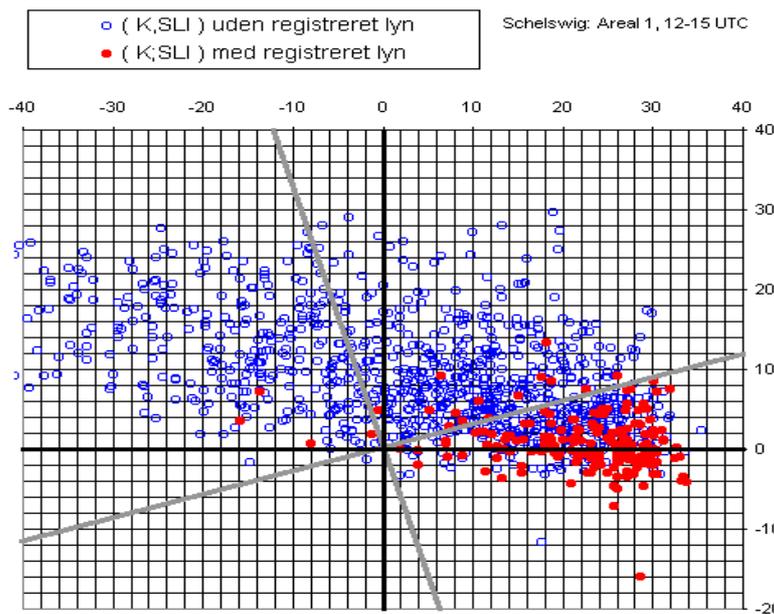


Figure 9. The SLI index plotted against the K index. The grey lines represent a rotated co-ordinate system.

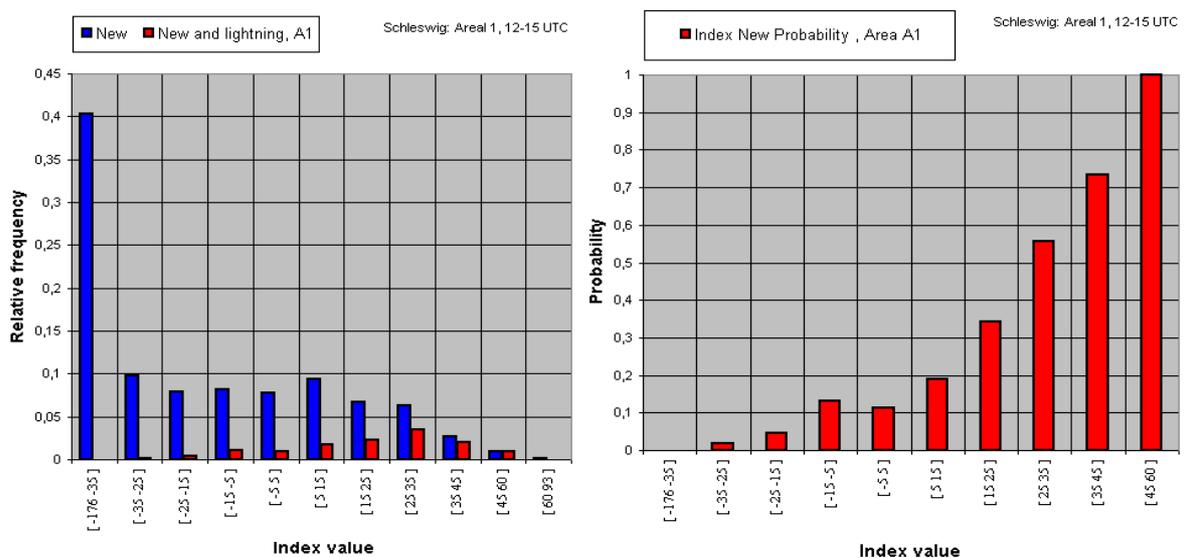


Figure 10. Left hand side: The frequency of days as a function of the new index. Right hand side: The probability based on the new index as a function of the index value.

The relative frequency of days with lightning is much more equal to the relative frequency of all days in the intervals with large index values. This results in a much larger probability for lightning for these intervals. The maximum probability is now 100% compared to the K index alone which gives a maximum less than 70%.

Figure 11 shows the reliability curve based on the probability of the combined index.

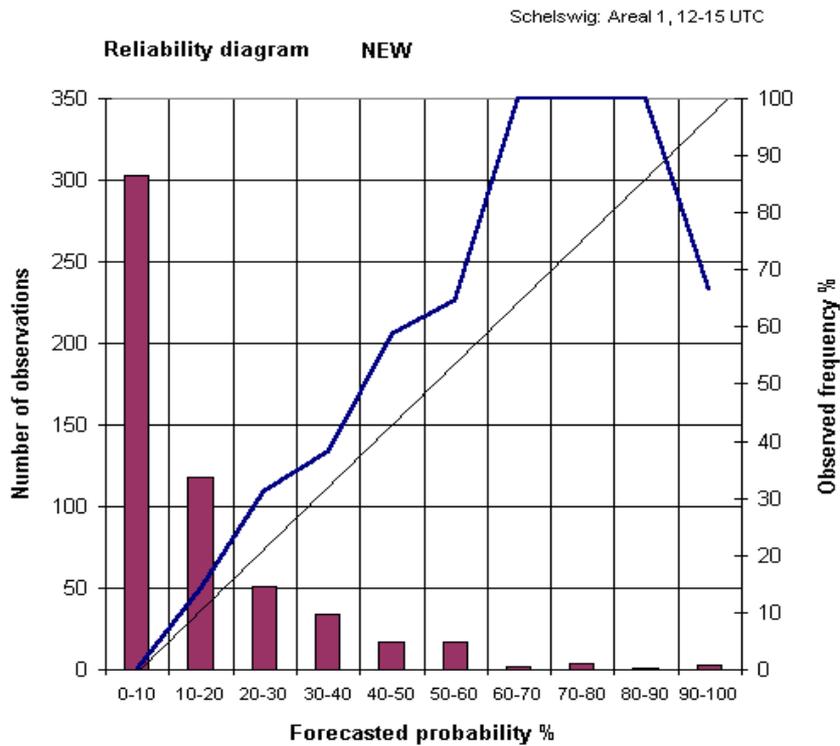


Figure 11. The reliability curve based on the probability of the combined index.

The highest probability indeed has increased, but still occurs very seldom, and the overall reliability show some under prediction of the probability . The Brier score is now equal to 0.081 and less than 0.098 for the K index alone and the Brier skill score is 0.33, one and a half times larger than 0.20 for the K index alone, which is a substantial improvement.

The under prediction may at least to some extent be due to relative small number of cases with high probability.

6. Logistic regression and Kalman filter

The problem of finding the probability as a function of the index value may be solved by logistic regression and kalman filtering. Logistic regression is similar to normal regression but with the difference, that the dependent variable can take only a discrete set of values. In the present case the values are binary data equal to 1 for days with lightning and 0 for days without.

The main difference is that the problem is transformed to a probability formulation. In normal regression of one dependent variable Y from a set of independent variables $X_i, i=1, \dots, N$ the equation is $Y^{\wedge} = a_0 + a_1 * X_1 + \dots + a_N * X_N$ for the estimated value of Y . The a 's are determined by minimising the sum of $(Y^{\wedge} - Y)^2$ over a set of known values. Using a Kalman filter technique the value of the a 's may be estimated recursively as the observed Y 's becomes available.

In logistic regression the equation $Y^{\wedge} = a_0 + a_1 * X_1 + \dots + a_N * X_N$ is transformed into a probability by a logistic function $p = p(a_0 + a_1 * X_1 + \dots + a_N * X_N)$ which take values between 0 and 1. The logistic function is usually defined by $p = e^{(a_0 + a_1 * X_1 + \dots + a_N * X_N)} / (1 + e^{(a_0 + a_1 * X_1 + \dots + a_N * X_N)})$, but in the present study the following function $p = \min(1, e^{(a_0 + a_1 * X_1 + \dots + a_N * X_N)})$ has proved to be more useful and is equal to the approximation of the probability discussed in section 5.5.

Instead of the error function $(Y^{\wedge} - Y)^2$ the error probability is calculated as $2 * \text{Arctanh}(I - p)$, where p is the estimated probability and I is 1 for days with lightning and 0 for days without.

Changing to the probability function and using the error probability function in the kalman filter the a 's and thus the probability are estimated recursively as the observed Y 's becomes available.

The logistic kalman filter for the K index has a mean value of a_0 around -2.937, and a mean value of a_1 around 0.065, which correspond well with the values of -3.366 and 0.091 used in section 5.5.

However, one of the advantages of using Kalman technique is that the coefficients are updated recursively when new data becomes available and as such not constant. The Kalman filter is therefore able to take a time dependent variation into account by changing the coefficients properly. One of the time dependent variations is the seasonal changes in the relative frequency of occurrence of lightning. The recursive estimates of a_0 and a_1 are shown in figure 12 where the variations with time is clearly seen.

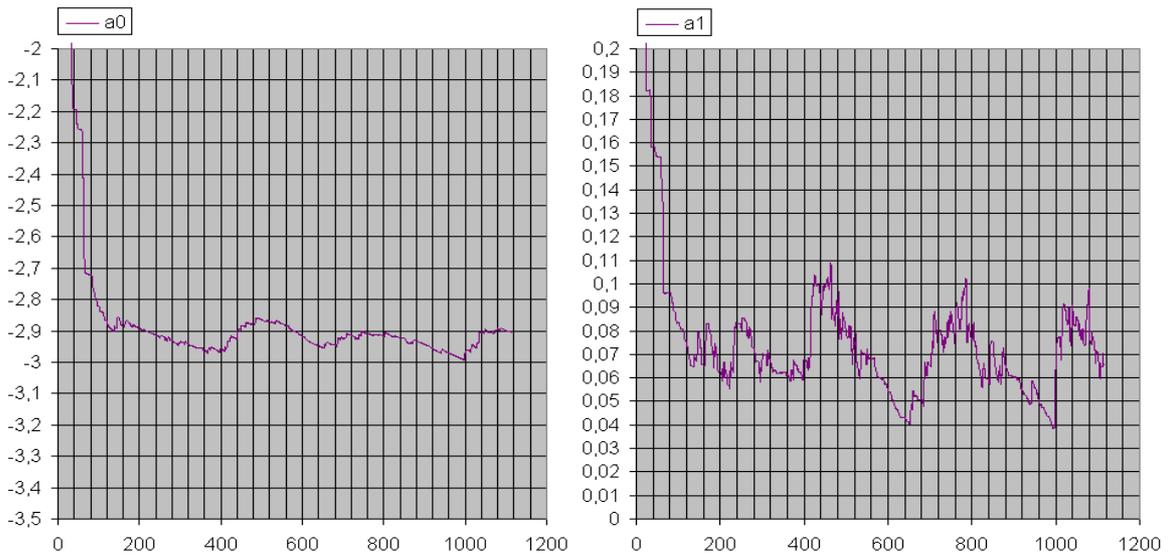


Figure 12. Left hand side: The Kalman filter estimate of a_0 for the index K. Right hand side: The Kalman filter estimate of a_1 for index K.

The long waves are due to the seasonal variations.

Figure 13 left hand side shows the estimated probability day by day together with the I function as function of the K value. This should be compared with figure 5. The broadening of the values of p for larger values is a result of the variation in the estimate for the a_0 and the a_1 values and may to some extent account for seasonal variation.

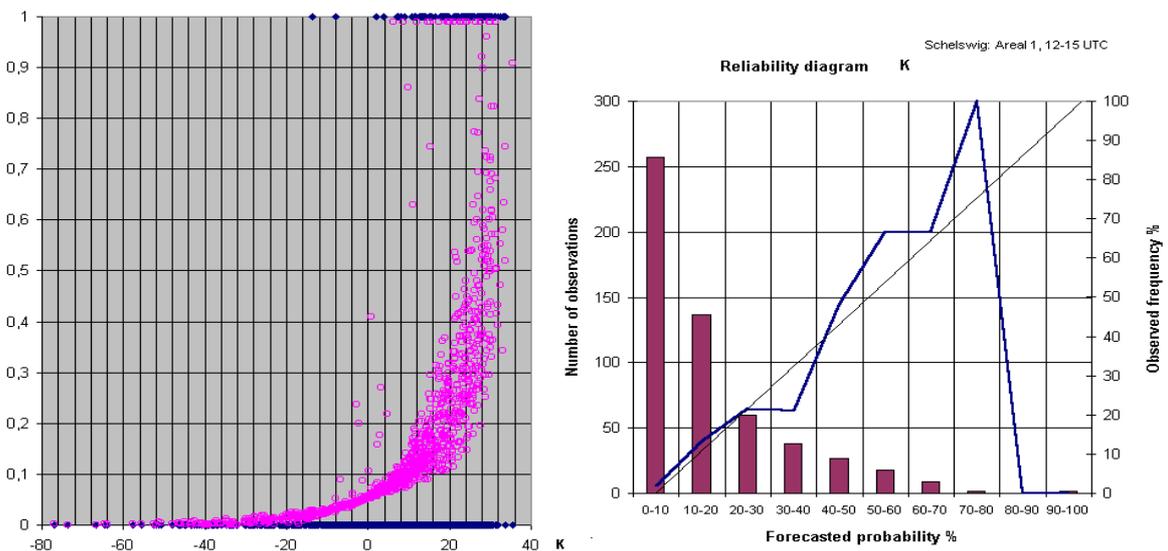


Figure 13. Left hand side: The Kalman filtered probability and the I function for each day as a function of K value. Right hand side: The corresponding reliability diagram.

Figure 13 right hand side shows the corresponding reliability diagram which should be compared to figure 6. Probabilities between 70 and 80%, which could not be estimated using the method in section 5.5 are estimated by the Kalman filteret model, so the Kalman filtered

model in that respect is an improvement, though the occurrence of this large probability is still rare. This may also be part of the explanation of the under estimate of the probability especially for the 70-80% interval, which contain only two days.

The calculated Brier score for the Kalman filtered probability is 0.096 and the Brier skill score is 0.20, which is equal to that found in section 5.5.

Also the new index described in section 5.8 can be estimated with a two-variable Kalman filter. The mean value of the coefficient a_0 is estimated to around -2.017 and the mean value of a_1 is estimated to around 0.041. The estimates used in section 5.8 of a_0 and a_1 are -1.97 and 0.044. a_2/a_1 is the coefficient to SLI used in section 5.8 and is with the Kalman filter estimated to a mean around -4.5. The recursive estimate of a_0 and a_1 and the coefficient to the SLI index are shown in figure 14, where again the seasonal time variation is clearly seen.

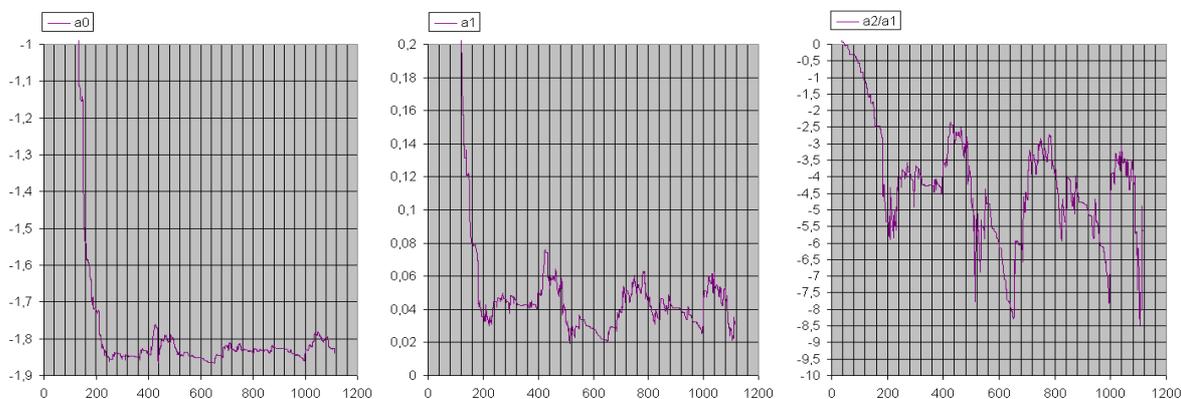


Figure 14 : Left hand side: The Kalman filter estimate of a_0 for the two indices K and SLI. The centre: The Kalman filter estimate of a_1 for the two indices K and SLI. The right hand side: The Kalman filter estimate of a_2/a_1 coefficient to SLI .

The reliability diagram for the combined index is shown in figure 15.

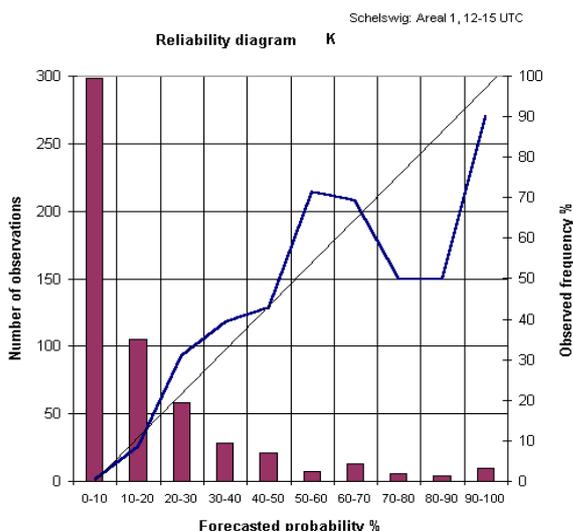


Figure 15. The reliability diagram for the Kalman filter estimate of probability for the two indices K and SLI

The largest probability is now increased to the interval 90-100% with actually also a small increase in the occurrence though still rather rare. The Brier score is calculated to 0.081 and the brier skill score is calculated to 0.33, which is similar to the values found in section 5.8.

The Kalman filtering technique has thus proved to be an efficient and powerful tool to estimate the probabilities and capture the time dependence from a set of indices.

7. Conclusion and perspectives

It has been shown in this preliminary study that it is possible to develop a useful automatic model for a probabilistic indicator of thunder/lightning based on very simple stability indices, which perform significantly better than climatology and persistence.

The model is unfortunately not very good in the region of high probability, which of course is very important seen from a users point of view.

It should though be pointed out that the present study assumes, that the atmospheric conditions at some time determines the possibility of the occurrence of lightning during the next three hours. Whether or not lightning actually occurs depends on whether a trigger mechanism (forcing) initiates the possible development necessary for lightning to occur.

On the other hand due to the simplicity of the indices used in the present model it is expected that the use of more sophisticated information from HIRLAM such as cloud water, precipitation, liquid water/ice, vorticity, wind shear, advection etc. and the numerical convection scheme should make it possible to improve the method significantly, hopefully in the region of high probabilities where the present model needs to be improved the most.

It is also shown that there is a seasonal variation, which should somehow be taken into account, either by adding a seasonal variable or having two distinct models or perhaps by incorporating it into the Kalman filter.

A spatial variation is also expected, but this has not been analysed yet. This is related to the size of the area used, which still remains to be solved. Also the relation between two areas of different size needs to be examined further.

The logistic regression using a Kalman filter has proven to be an efficient and powerful tool in developing and maintaining an automatic model for a probabilistic indicator of thunder/lightning. It is expected that this method can be used in other areas such as accumulated precipitation, development of Cb, etc.

The results are of course limited to the time 12 UTC and the location Schleswig used in the study. Different times of the day such as night time and other locations will presumably give different results.

An operational system is therefore expected to for instance contain a number of different Kalman filters for the different times of the day and locations.

8. References

- H. Haakenstad and M. H. jensen, 1999: Sammenlikning av beregnet k-indeks i HIRLAM10 med lynobservasjoner. Norwegian Meteorological Institute, research Report No. 89, 11 pp.
- H. Huntreiser, H. H. Schiesser, W. Schmid and A. Waldvogel, 1997: Comparison of traditional and Newly Developed Thunderstorm Indices for Switzerland. *Bull. Amer. Meteor. Soc.*, **Vol. 12**, 108-125.

DANISH METEOROLOGICAL INSTITUTE
Scientific Reports

Scientific reports from the Danish Meteorological Institute cover a variety of geophysical fields, i.e. meteorology (including climatology), oceanography, subjects on air and sea pollution, geomagnetism, solar-terrestrial physics, and physics of the middle and upper atmosphere.

Reports in the series within the last five years:

No. 95-1

Peter Stauning and T.J. Rosenberg:
High-Latitude, day-time absorption spike events
1. morphology and occurrence statistics
Not published

No. 95-2

Niels Larsen: Modelling of changes in stratospheric ozone and other trace gases due to the emission changes : CEC Environment Program Contract No. EV5V-CT92-0079. Contribution to the final report

No. 95-3

Niels Larsen, Bjørn Knudsen, Paul Eriksen, Ib Steen Mikkelsen, Signe Bech Andersen and Torben Stockflet Jørgensen: Investigations of ozone, aerosols, and clouds in the arctic stratosphere : CEC Environment Program Contract No. EV5V-CT92-0074. Contribution to the final report

No. 95-4

Per Høeg and Stig Syndergaard: Study of the derivation of atmospheric properties using radio-occultation technique

No. 95-5

Xiao-Ding Yu, **Xiang-Yu Huang** and **Leif Laurssen** and Erik Rasmussen: Application of the HIRLAM system in China: heavy rain forecast experiments in Yangtze River Region

No. 95-6

Bent Hansen Sass: A numerical forecasting system for the prediction of slippery roads

No. 95-7

Per Høeg: Proceeding of URSI International Conference, Working Group AFG1 Copenhagen, June 1995. Atmospheric research and applications using observations based on the GPS/GLONASS System
Not published

No. 95-8

Julie D. Pietrzak: A comparison of advection schemes for ocean modelling

No. 96-1

Poul Frich (co-ordinator), H. Alexandersson, J. Ashcroft, B. Dahlström, G.R. Demarée, A. Drebs, A.F.V. van Engelen, E.J. Førland, I. Hanssen-Bauer, R. Heino, T. Jónsson, K. Jonasson, L. Keegan, P.Ø. Nordli, **T. Schmih, P. Steffensen**, H. Tuomenvirta, O.E. Tveito: North Atlantic Climatological Dataset (NACD Version 1) - Final report

No. 96-2

Georg Kjærgaard Andreasen: Daily response of high-latitude current systems to solar wind variations: application of robust multiple regression. Methods on Godhavn magnetometer data

No. 96-3

Jacob Woge Nielsen, Karsten Bolding Kristensen, Lonny Hansen: Extreme sea level highs: a statistical tide gauge data study

No. 96-4

Jens Hesselbjerg Christensen, Ole Bøssing Christensen, Philippe Lopez, Erik van Meijgaard, Michael Botzet: The HIRLAM4 Regional Atmospheric Climate Model

No. 96-5

Xiang-Yu Huang: Horizontal diffusion and filtering in a mesoscale numerical weather prediction model

No. 96-6

Henrik Svensmark and Eigil Friis-Christensen: Variation of cosmic ray flux and global cloud coverage - a missing link in solar-climate relationships

No. 96-7

Jens Havskov Sørensen and Christian Ødum Jensen: A computer system for the management of epidemiological data and prediction of risk and economic consequences during outbreaks of foot-and-mouth disease. CEC AIR Programme. Contract No. AIR3 - CT92-0652

No. 96-8

Jens Havskov Sørensen: Quasi-automatic of input for LINCOM and RIMPUFF, and output conversion. CEC AIR Programme. Contract No. AIR3 - CT92-0652

No. 96-9

Rashpal S. Gill and Hans H. Valeur: Evaluation of the radarsat imagery for the operational mapping of sea ice around Greenland

No. 96-10

Jens Hesselbjerg Christensen, Bennert Machenhauer, Richard G. Jones, Christoph Schär, Paolo Michele Ruti, Manuel Castro and Guido Visconti: Validation of present-day regional climate simulations over Europe: LAM simulations with observed boundary conditions

No. 96-11

Niels Larsen, Bjørn Knudsen, Paul Eriksen, Ib Steen Mikkelsen, Signe Bech Andersen and Torben Stockflet Jørgensen: European Stratospheric Monitoring Stations in the Arctic: An European contribution to the Network for Detection of Stratospheric Change (NDSC): CEC Environment Programme Contract EV5V-CT93-0333: DMI contribution to the final report

No. 96-12

Niels Larsen: Effects of heterogeneous chemistry on the composition of the stratosphere: CEC Environment Programme Contract EV5V-CT93-0349: DMI contribution to the final report

No. 97-1

E. Friis Christensen og C. Skøtt: Contributions from the International Science Team. The Ørsted Mission - a pre-launch compendium

No. 97-2

Alix Rasmussen, Sissi Kiilsholm, Jens Havskov Sørensen, Ib Steen Mikkelsen: Analysis of tropospheric ozone measurements in Greenland: Contract No. EV5V-CT93-0318 (DG 12 DTEE): DMI's contribution to CEC Final Report Arctic Tropospheric Ozone Chemistry ARCTOC

No. 97-3

Peter Thejll: A search for effects of external events on terrestrial atmospheric pressure: cosmic rays

No. 97-4

Peter Thejll: A search for effects of external events on terrestrial atmospheric pressure: sector boundary crossings

No. 97-5

Knud Lassen: Twentieth century retreat of sea-ice in the Greenland Sea

No. 98-1

Niels Woetman Nielsen, Bjarne Amstrup, Jess U. Jørgensen: HIRLAM 2.5 parallel tests at DMI: sensitivity to type of schemes for turbulence, moist processes and advection

No. 98-2

Per Høeg, Georg Bergeton Larsen, Hans-Henrik Benzon, Stig Syndergaard, Mette Dahl Mortensen: The GPSOS project
Algorithm functional design and analysis of ionosphere, stratosphere and troposphere observations

No. 98-3

Mette Dahl Mortensen, Per Høeg: Satellite atmosphere profiling retrieval in a nonlinear troposphere
Previously entitled: Limitations induced by Multipath

No. 98-4

Mette Dahl Mortensen, Per Høeg: Resolution properties in atmospheric profiling with GPS

No. 98-5

R.S. Gill and M. K. Rosengren
Evaluation of the Radarsat imagery for the operational mapping of sea ice around Greenland in 1997

No. 98-6

R.S. Gill, H.H. Valeur, P. Nielsen and K.Q. Hansen: Using ERS SAR images in the operational mapping of sea ice in the Greenland waters: final report for ESA-ESRIN's: pilot projekt no. PP2.PP2.DK2 and 2nd announcement of opportunity for the exploitation of ERS data projekt No. AO2..DK 102

No. 98-7

Per Høeg et al.: GPS Atmosphere profiling methods and error assessments

No. 98-8

H. Svensmark, N. Woetmann Nielsen and A.M. Sempreviva: Large scale soft and hard turbulent states of the atmosphere

No. 98-9

Philippe Lopez, Eigil Kaas and Annette Guldborg: The full particle-in-cell advection scheme in spherical geometry

No. 98-10

H. Svensmark: Influence of cosmic rays on earth's climate

No. 98-11

Peter Thejll and Henrik Svensmark: Notes on the method of normalized multivariate regression

No. 98-12

K. Lassen: Extent of sea ice in the Greenland Sea 1877-1997: an extension of DMI Scientific Report 97-5

No. 98-13

Niels Larsen, Alberto Adriani and Guido DiDonfrancesco: Microphysical analysis of polar stratospheric clouds observed by lidar at McMurdo, Antarctica

No.98-14

Mette Dahl Mortensen: The back-propagation method for inversion of radio occultation data

No. 98-15

Xiang-Yu Huang: Variational analysis using spatial filters

No. 99-1

Henrik Feddersen: Project on prediction of climate variations on seasonal to interannual timescales (PROVOST) EU contract ENVA4-CT95-0109: DMI contribution to the final report: Statistical analysis and post-processing of uncoupled PROVOST simulations

No. 99-2

Wilhelm May: A time-slice experiment with the ECHAM4 A-GCM at high resolution: the experimental design and the assessment of climate change as compared to a greenhouse gas experiment with ECHAM4/OPYC at low resolution

No. 99-3

Niels Larsen et al.: European stratospheric monitoring stations in the Arctic II: CEC Environment and Climate Programme Contract ENV4-CT95-0136. DMI Contributions to the project

No. 99-4

Alexander Baklanov: Parameterisation of the deposition processes and radioactive decay: a review and some preliminary results with the DERMA model

No. 99-5

Mette Dahl Mortensen: Non-linear high resolution inversion of radio occultation data

No. 99-6

Stig Syndergaard: Retrieval analysis and methodologies in atmospheric limb sounding using the GNSS radio occultation technique

No. 99-7

Jun She, Jacob Woge Nielsen: Operational wave forecasts over the Baltic and North Sea

No. 99-8

Henrik Feddersen: Monthly temperature forecasts for Denmark - statistical or dynamical?

No. 99-9

P. Thejll, K. Lassen: Solar forcing of the Northern hemisphere air temperature: new data

No. 99-10

Torben Stockflet Jørgensen, Aksel Walløe Hansen: Comment on "Variation of cosmic ray flux and global coverage - a missing link in solar-climate relationships" by Henrik Svensmark and Eigil Friis-Christensen

No. 99-11

Mette Dahl Meincke: Inversion methods for atmospheric profiling with GPS occultations

No. 99-12

Benzon, Hans-Henrik; Olsen, Laust: Simulations of current density measurements with a Faraday Current Meter and a magnetometer

No. 00-01

Høeg, P.; Leppelmeier, G: ACE: Atmosphere Climate Experiment: proposers of the mission

No. 00-02

Høeg, P.: FACE-IT: Field-Aligned Current Experiment in the Ionosphere and Thermosphere

No. 00-03

Allan Gross: Surface ozone and tropospheric chemistry with applications to regional air quality modeling. PhD thesis

No. 00-04

Henrik Vedel: Conversion of WGS84 geometric heights to NWP model HIRLAM geopotential heights

No. 00-05

Jérôme Chenevez: Advection experiments with DMI-Hirlam-Tracer
(In Press)

No. 00-06

Niels Larsen: Polar stratospheric clouds micro-physical and optical models

No. 00-07

Alix Rasmussen: "Uncertainty of meteorological parameters from DMI-HIRLAM"
(In Press)

No. 00-08

A.L. Morozova: Solar activity and Earth's weather. Effect of the forced atmospheric transparency changes on the troposphere temperature profile studied with atmospheric models

No. 00-09

Niels Larsen, Bjørn M. Knudsen, Michael Gauss, Giovanni Pitari: Effects from high-speed civil traffic aircraft emissions on polar stratospheric clouds

No. 00-10

Søren Andersen: Evaluation of SSM/I sea ice algorithms for use in the SAF on ocean and sea ice, July 2000
(In Press)

No. 00-11

Claus Petersen, Niels Woetmann Nielsen: Diagnosis of visibility in DMI-HIRLAM
(In Press)

No. 00-12

Erik Buch: A monograph on the physical oceanography of the Greenland waters