UPGRADE AND EVALUATION OF A LIGHTNING DETECTION SYSTEM

Hans Beekhuis
Iwan Holleman

the Netherlands, KNMI De Bilt
26 April 2004

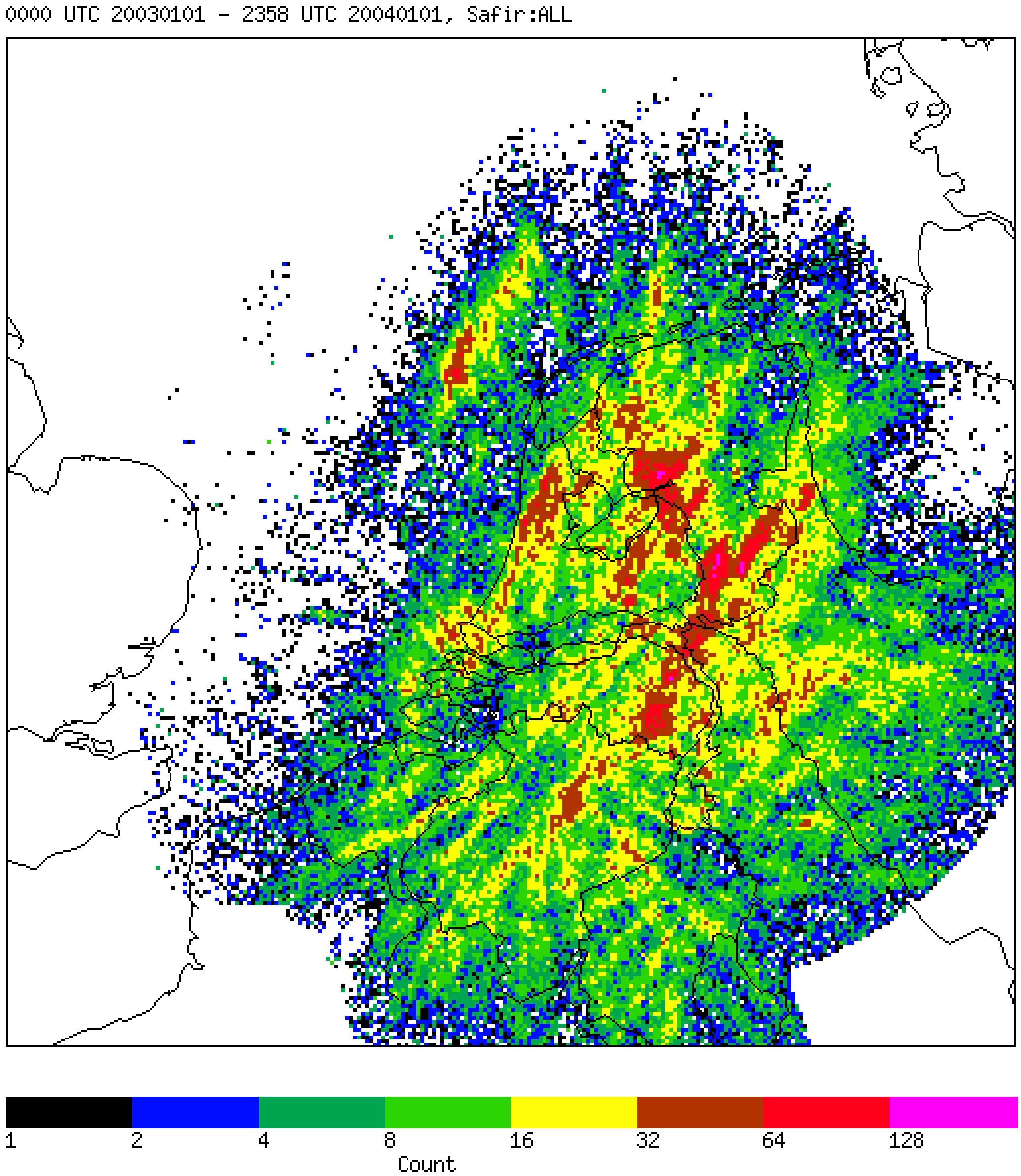

Table of contents

1 INTRODUCTION

KNMI (Royal Netherlands Meteorological Institute) has upgraded its Lightning Detection System (LDS). This new system is called FLITS (Flash Localisation by Interferometry and Time of arrival System) to distinguish it from the former system called SAFIRK.

From 1987 onward KNMI uses automated observing systems to record lightning activity for nowcasting and climatological purposes. The first system was a time of arrival system called LPATS (Lightning Tracking and Positioning System) operated by KEMA, a Dutch energy provider.
In 1995 this system was replaced by a Safir system, called SAFIRK, that was acquired in a joint effort by KNMI, The Royal Dutch Air force, and The Royal Dutch Navy. Soon it became clear that the three detection stations used by this initial system offered too little coverage. Therefore raw data from the Belgium Safir system is used from 1997 onward to enhance reliability and coverage in southern direction. In 2001 a fourth detection station of type Safir 3000, situated in the north east of the Netherlands, came into operation. Within the scope of the project LDUP that upgrades the SAFIRK system to FLITS, the three original detection stations are now upgraded to SAFIR 3000 type stations as well. In this project also the servers and all vendor software is replaced.

In this technically oriented paper the following topics are presented:

a) General system layout, including the software components used.
b) KNMI LIGHTNING HDF5 format. Some details of the HDF5 internals are presented.
c) Use of lightning data in the nowcasting messages SYNOP and METAR used in aviation meteorology.
d) A basic outline of the data processing.
e) An evaluation of the characteristics of FLITS compared to SAFIRK. Offline tests have been conducted on a dataset recorded both by FLITS and SAFIRK. Results of these tests concerning the detection of ground strokes and localisation accuracy are discussed.

1.1 Definitions

Throughout this document the following definitions apply:

Table 1. Definitions

Name	Definition
Sample	Acquisition performed by a single detection station
Burst	Grouping of series of Samples on a single detection station, originating from the same lightning event
Localisation	A triangulated Burst
Trace	Time series of Localisations, assigned to the same lightning, classified into: Type_0: Isolated Point Type_1: First Localisation in a Trace Type_2: Zero or more Localisations assigned to the same Trace Type_3: Last Localisation in a Trace Type_4: The first Localisation in a Trace discriminated to be a ground stroke. Type_5: Subsequent Localisations discriminated to be multiples of the ground stroke
Flash	Reduction of a Trace to a point. The coordinates of a Flash are taken as the mean of the locations of the Type_1 event and the corresponding Type_3 event
Ground_stroke	A Trace can contain a ground stroke, It is assigned the coordinates of the Type_4 event
Return_stroke	A Trace can contain one or more return strokes. It is assigned the coordinates of the Type_5 event

2 GENERAL LAYOUT

The LDS as operated by KNMI/KLU/KM consists of four Detection Stations situated in De Kooy, Valkenburg, Deelen and Hoogeveen. From the LDS operated by the Belgian KMI data of the Detection Stations OeLegem, Mourcourt and La Gilleppe are acquired. The locations of the Detection Stations are marked as “+ “ in Figure 2. The Detections Stations send their data over serial leased lines to the central computing module. In the near future the connections will migrated to network connections. For the data from the Belgian Detections Stations an internet connection using VPN will be used.

In Figure 1 is given an overview of the system.
The Servers, taking the data of the Detection Stations, are at the facilities of KNMI in De Bilt. They run Windows 2000 advanced server, service pack 4, as Operating System.

In order to guarantee the operational presence of the LDS a high availability configuration is used. The servers are configured to operate in a cluster, sharing their data disks. In case of a failure in hardware or software the processes running on the active cluster node are moved automatically to the other node.

To monitor and analyse data a Maintenance Terminal is placed at the technical maintenance department of KNMI.

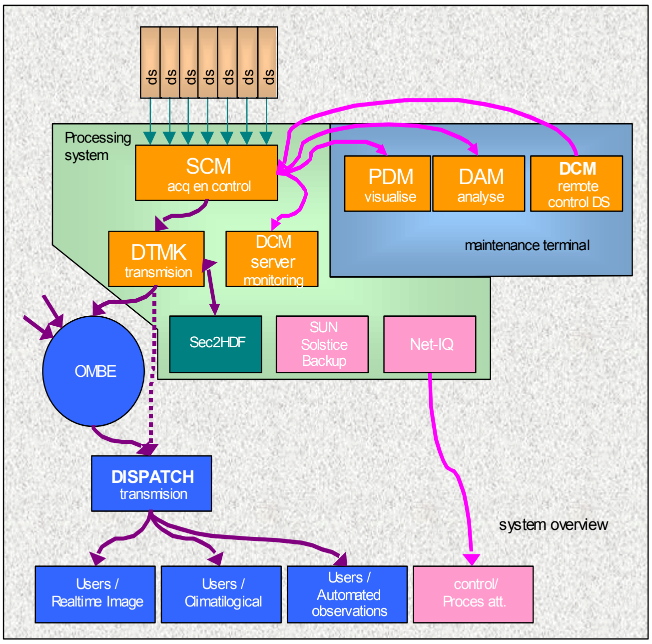

Figure 1. System overview software components

2.1 Software components

In Table 2 an overview of the software used on FLITS is presented. A short description is given and the intended user is mentioned.

Table 2: Software overview

Processing System	Purpose	User level
SCM SAFIR Control Module	Acquisition and processing of data from the Detection Stations, Primary and Secondary data storage.	Administrator
DTMK Data Transmitter Module For KNMI	This component handles the transfer of data to the OMBE database, using FTP. It handles also the calls to the HDF5 converter.	Administrator
SEC2HDF KNMI HDF Converter	The HDF5 converter is developed by KNMI. It handles the conversion of a secondary file into a HDF file.	Administrator

DCM-server Detection Control Module	Takes care of the connection to the DCM-client(s).	Administrator
SUN Solstice Backup	Client Part of the backup service used by KNMI.	System Maintenance
NET IO Application manager	Provides information on computer resources, and proactively signals problems.	System Maintenance
Service Terminal	Purpose	User level
PDM Processing & Display Module	Geographical display of data, with real-time and replay modes, based on standard GIS .	Technical Maintenance
DCM-client Detection Control Module	Remote control and supervision of Detection Stations. Until now control of a detection Station is not possible.	Technical Maintenance
DAM Data Analysis Module	Analysis of lightning data over a large period, analysis of Safir system properties.	Technical Maintenance
Detection Station		
DSS Detection Station Software	Acquisition of the HF and LF discharge data. Pre-processing and time stamping of these data. Transfer of data to SCM.	Technical Maintenance

2.2 Data distribution

One of the major goals of the upgrade was to create a manageable data distribution. This is achieved by positioning FLITS as a frontend for a database, called OMBE, used to archive and distribute image data. Other frontends that deliver data to OMBE are (inter)national weather radars and satellite reception stations. Users have access to the database, or get their data delivered by the dispatcher, that is part of OMBE. Direct user access to the frontend is not allowed.
To create a rationale distribution the only supported output format of FLITS is LIGHNTING HDF5 (see paragraph 2.3 for details). The HDF5 files are available in three chunksizes, covering an interval of 1 minute, 5 minutes or 24 hours time.

The users that get data from FLITS are summarised below:

Table 3. Data distribution to Users

User	Chunk	Description
MWS	5M	KNMI Meteorological workstation developed by 3SI is able to read HDF5 files in order to annotate radar images with lightning warnings.
KLU	5M	Royal Dutch Air Force, Assimilates HDF5 data in their workstation.
LVNL	5M	Air Traffic Control the Netherlands, Assimilates HDF5 data in their applications.
WNI	5M	Whether News International, a whether provider in the Netherlands.
BRAS	5M	KNMI does not use a third party viewer for presenting Lightning to the end user, A lightning viewer is integrated in web based Radar and Lightning viewer.
CIBIL	1M	Data aquisition, database and application software that generates automated observations. (see paragraph 2.4 for details)
KD	24H	KNMI Climatological services, public service for climatological information.

2.3 The KNMI HDF5 Lightning format

HDF5 (Hierarchical Data Format version 5) is a multi-object file format for sharing scientific data in a distributed environment. An important feature of HDF5 files is that they are self-describing. The format was developed at the National Centre for Supercomputing Applications (NCSA) at the University of Illinois. The aim of KNMI HDF5 is to make files containing earth observation data self-contained. By grouping related pieces of information together and forcing a certain structure of the file, it will be easier for other users to make use of the data in the files. Self-contained also means that all meta-data associated with the dataset and/or used in the processing of the dataset is included in the file.

2.3.1 Data description

In a KNMI lightning HDF5 file the data is available as an image as well as an indexed arrays of Localisations. The image data is accumulated over a certain period of 5 minutes or 24 hours and is processed to a geographic grid. A 1 minute interval is possible for future use. The grid used is the same grid as that what is used for presentation of radar data. In the Localisation array (see Table 4) several parameters are stored. In essence the Localisation array allows extracting lightning data with a time-resolution of 100 µs . The system status is available also in the HDF5 file, and allows for evaluating of the system status with a time resolution of a minute. As it is allowed within KNMI HDF5 to store several images within one file, two images are defined one containing the accumulation of all Flashes (/ALL), the other contains the accumulation of all Ground_strokes (/CG).

2.3.2 Localisation arrays

The Localisation arrays are used for storage of the all Localisations observed by the Lightning Detection System. Each variable resolved by the system is stored as a separate array. The date and time of a Localisation are given as a time offset in seconds against a reference data and time, which is stored as well. The time offset is stored in a double precision float to allow for storage of a high accuracy time offset over a daily period. The variables resolved are given in Table 4, but addition of other variables is possible.

Table 4. Array data stored in the KNMI Lightning HDF5

Name	Type	Unit	Description
number_discharges	Int		Number of discharges in timeseries, i.e., length of discharge datasets.
reference_datetime	String		Date and time stamp against which discharges are referenced.
time_offset	Table of Double	seconds	Dataset with time offsets of discharges with respect to reference date and time, double allows for microsec accuracy.
longitude	Table of Float	decimal degree	Dataset with geographical longitudes of discharges.
Latitude	Table of Float	decimal degree	Dataset with geographical latitudes of discharges.
event_type	Table of Char		Dataset with types of observed discharges: “o” single-point, “1” start of CC, “2” CC discharge, “3” end of CC, “4” CG stroke, “5” CG return stroke.
position_error	Table of Float	meter	Dataset with position errors of CG stroke localizations as deduced by detection system.
rise_time	Table of Float	second	Dataset with rise times of induced current for detected CG strokes..
decay_time	Table of Float	second	Dataset with decay times of induced current for detected CG strokes.
Current	Table of Float	Ampere	Dataset with estimated currents of CG strokes.

2.3.3 Image description

In Figure 2 is depicted a HDF5 image as it is generated with the current production. In these images the Trace are reduced to a Flash. For a Trace with an identified Stroke, the location of the Stroke is used to position the event. For a Trace without an identified Stroke, the mean of the position of start and endpoint is used.

The image is constructed based on the following rules:

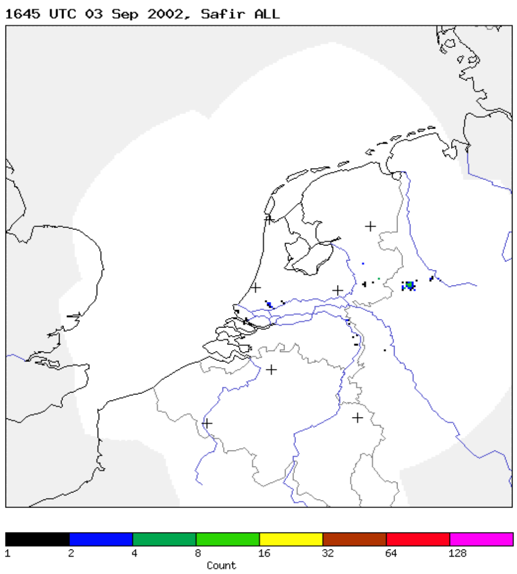

Figure 2. Image in HDF5 format showing NO_DATA area and location of Detection Stations (+)

- Isolated Points are neglected.
- The product contains the accumulated count of the detected events over a certain pixel, in the specified integration time interval.
- Geographical projection and cut are identical with the radar products of the KNMI, containing 256 x 256 pixels, in a stereographic projection. When using geographical projection formulas the use of the free available proj4 library is highly recommended.(See reference 6)
- The actual detection area is dynamically generated based on the availability of the Detection Stations. The NO_DATA value is assigned to all pixels in the image that are out of the detection area. A detection station is assumed to be unavailable when its uptime in the integration interval is 0 minutes. A pixel is assumed to be within the detection area when: (1) The angle between the lines from discharge to the contributing stations is over 15^o (parameter) but less the 150^o (parameter) AND (2) The distance from discharge to the contributing station is over 20 (parameter) but less then 250 km (parameter).
- Data generated by the system in the NO_DATA data area is blended out. This only applies for the /ALL image. The /CG image does not carry no-data. Because the Time Of Arrival (TOA) measurement offers a greater range.

2.3.4 Support of HDF5: Libraries and Test Data

There are libraries available for accessing HDF5-formatted data. There is a freeware version available at the NCSA. (See http://hdf.ncsa.uiuc.edu).
KNMI has developed a library (ansi-c) that facilitates high level of abstraction access to KNMI HDF5 formatted files. This library is to be used for free, as long as copyrights are respected and mentioned. With this library any data element is easy accessible.

A set of data covering a time span of 26 hours is available. The set starts at 2003/07/19 23:00:00 and ends 2003/07/21 01:00:00, so it covers a full day, and some overlap at both sides in order to test on the change of date at midnight.

2.4 Automated observations

KNMI has fully automated synoptic observations using sensors as Present Weather, Ceilometer, and the Lightning Detection System. Lightning synoptic messages are derived from the Lightning Detection System. In near future Autometar reports will be generated during “closing hours” of some cilvil airports and air force bases.

For the Lightning data a simple algorithm accounts for the contribution of lightning in the code.
As explained before Localisations are coded in 6 classes (Type_0… Type_5). If any Localisation, Type_0 excluded, within a 15 km radius of a station is detected, the automatic system reports lightning.

Wauben 2) performed a comparison with observed lightning events based on hourly climatological reports. In these reports is indicated whether lightning events have been detected in the previous hour. Table 5 shows the contingency matrix for stations De Bilt and Schiphol in the year 2000. The conclusion of Wauben is: “The few cases where the SAFIRK system ‘misses’ observed lightning events probably took place outside the 15 km radius around the station. On the other hand the SAFIRK system gives many ‘false’ events, but the observer is mostly located inside and surrounded by a lot of noise and can therefore easily overhear lightning events.”

Table 5. Contingency matrix DB and SPL

De Bilt		LDS	
		Yes	No
OBS	yes	87	8
	no	72	8616
Schiphol		LDS	
		Yes	No
OBS	yes	64	6
	no	102	8611

Event		LDS	
		Yes	No
OBS	yes	Hit	Miss
	no	False	None

LDS = Lightning Detection System
OBS = Observer
Probability Of Detection (POD) = 100%*H/(H+M)
False Alarm Rate (FAR) = 100%*F/(F+H)
Critical Success Index (CSI) = 100%*H/(H+M+F)
Bias = (H+F)/(H+M)

In order to reduce the false alarm rate in near future a filter will be implemented that suppresses the False hits by evaluating the total lighting activity that is registered by the FLITS. The basic assumption being that a lightning out of the blue is a rare phenomenon. Thus in order to present lightning in a synoptic report the total activity must be over a user definable threshold.

3 OPERATION PRINCIPLES

3.1 Detection station

A detection station in principle offers two types of data:

- **Localisation data recorded by a VHF sensor.** Atmospherics (electro-magnetic waves) generated by discharges are received in the 110..120 MHz band. All signals over a certain configurable threshold (typical –95 dBm) are processed using Interferometry. Samples are taken every 100 µs. Time accuracy of the samples is on the order of 1 µs. Accurate (0.5 µs) time is derived from a GPS receiver integrated in the LF sensor. It has to be noted that due to the used geometry of 5 dipoles a direction in the range 0-360 degrees is determined unambiguously. The former 4 dipole systems delivered a 0-180 degrees range. Measures are taken to avoid interference by synchronous radio waves.

Figure 3. Top showing dipole-array and LF-antenna

  Every 5 minutes the VHF sensor calibrates all internal electronics by a fully automated procedure. An equivalent procedure can be used to check the proper functioning of the antenna array by sending a calibration signal into one of the dipoles, and receiving on the others. This more stringent test can only be started by hand. As a result malfunctioning and degrading of the antenna array, which is one of the major error sources, can occur without being noticed. It is suggested that this procedure, called CNX test, should be able to run on a daily schedule, and leave a log of the calibration on the central processing system.

- **Electrical Field Changes recorded by a LF sensor.** Due to discharge currents, the electrical field changes rapidly. In case of a ground stroke the changes measured are much larger then those observed during cloud-cloud discharges. The LF sensor records the changes, in a band of 300 Hz to approximately 3 MHz. The electrical radiation field signal is digitised with an 8 MHz sample rate before the detection station processes it. Typical validation constraints imposed upon the signal processing are rise time, decay time, and applying a minimum and a maximum threshold. A signal that passes this test is analysed on the detection station to deliver the rise time, decay time, time of the peak value, and the peak value itself. For the LF sensor an auto calibration procedure is implemented.

A Detection Station is able to deliver a maximum of 100 events to the central processor each second, accompanied by at maximum 20 events recorded by the discrimination sensor. Variables delivered, excluding those generated for status, calibration results and housekeeping information, are given in Table 6.

Table 6. Variables delivered by a detection station

Sensor	Symbol	Description
Interferometer	Θ_i	azimuth
Interferometer	T_i	time of occurrence
Interferometer	A_i	Signal amplitude (VHF)
Interferometer	D_i	Signal density (VHF)
Discrimination	E_i	peak field value
Discrimination	T_{pi}	time of occurrence of the peak
Discrimination	T_{ri}	rise time
Discrimination	T_{di}	Decay time

3.2 Central processor

On the central processor, Samples received from all detection stations are combined to deliver Localisations and electrical parameters. It is beyond the scope of this report to go into details, but it is noted that it is hard to get a detailed view because the system documentation from Vaisala is far from complete in respect to the algorithms applied. For scientific use, it is of vital importance that a complete description of the data processing, including all methods that act upon the data, is available. It is strongly recommended that Vaisala will make the description of algorithms public for the lighting community, thus facilitating scientific research.

The processing by the Central Processor is outlined below:

a) Setting prerequisites and cleaning the primary dataset according to the following process steps:

   1. **Burst identification**.
   Samples are considered to belong to one Burst if they are separated in time less then 700 µs. This grouping is performed for each detection station separately.
   2. **Association of simultaneous data.**
   Signal stemming from different detection stations that are close in time are considered synchronous, when differing less then 700 µs.
   3. **Estimation of XY First stage of triangulation**.
   Once the bursts are associated to be simultaneous (in step 2) triangulation is performed on them in order to get a rough estimate of the position.
   4. **Correct for signal propagation**.
   From this preliminary position delay times due to propagation are calculated. The Samples are time shifted accordingly.
   5. **Selection of possible couples of sensors**.
   Based on geometrical criteria, such as minimum and maximum distances, and minimum and maximum angles that are allowed in the triangulation process, couples of Detection Stations are selected that are able to deliver accurate data. When more then one couple is available the couple that is able to do the triangulation in the most accurate way is selected.

b) Generating Trace's in the secondary dataset.

   In this phase Traces are created. A Trace is a recording of a set of Localisations that are considered to be stemming from the same lightning event. In a Trace, Localisations are classified according to be one of the following: an isolated event (Type_0), or a start point (Type_1), followed by zero or more channel points (Type_2), and an endpoint (Type_3). A Flash is defined as the occurrence of a Type_1 event. The coordinates of a Flash are taken as the mean of the locations of the Type_1 event and the corresponding Type_3 event.

c) Assessment of Strokes

   Finally the data of the discrimination sensor are used to classify Strokes in a Trace. The user can choose to use the Safir algorithm (referenced as Old_Discrimination) that evaluates rise time and decay time to discriminate between CC and CG, or use a TOA algorithm. As a result the Ground_stroke (Type_4) and eventually the multiple Return_strokes assumed using the same channel (Type_5) are added.

4 EXPERIMENTAL EVALUATION OF FLITS VERSUS SAFIRK

4.1 Remarks on the used dataset and runs

KNMI has been using SAFIRK lightning data for approximately a decade. With the introduction of FLITS the question arose how the performance of FLITS relates to the performance of SAFIRK. As the main new feature of the FLITS is its capability to do discriminate Strokes with a TOA algorithm, this issue required special interest.

Due to the way the upgrade is implemented data from the seven Safir-3000 detection stations is processed both by FLITS and SAFIRK. This is accomplished by splitting the signals from the detection stations. Consequently, both systems generate a complete set of primary data. Note that the detection stations have to operate in a compatibility mode, as long as SAFIRK acquires data.

The dataset used starts 16 July 2003 00:00 and extends to 29 October 2003 00:00.

This dataset starts on the day the Acceptance period of FLITS started, the weather responding on this fact with an abundant thunderstorm (over 800k intra cloud events). This storm did blow the fuses of the data link to the detection station Deelen within 6 hours of the onset of the acceptance period. In July, some more days passed with moderate storms, the rest of the year passed very quietly, with only a few minor storms in the months August until October.

The reference data set (SAF) for this comparison is generated by SAFIRK in real time. Data in T$-format is converted to HDF5-format with the same converter that has been used operationally to deliver HDF5 data to customers. In order to test the influence of changes in parameter setting, datasets from FLITS have been reprocessed by the Data Analysis Module (DAM Vaisala). All files on the system are contain one day of data, so over the period of investigation for al sets 105 daily entries are processed. All files have been converted to ascii based tables with the utility HDF2DIS. The possible processing chains are shown in Figure 4.

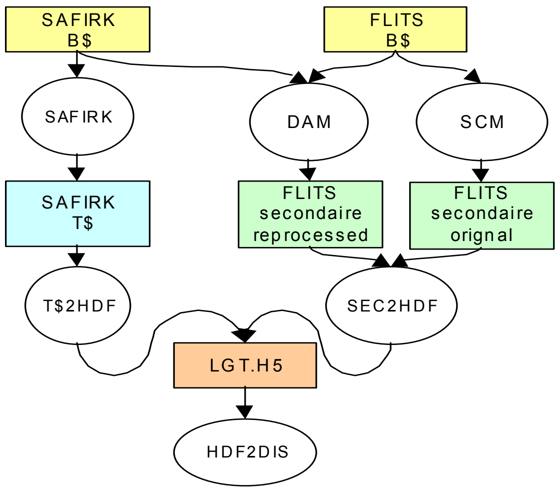

Figure 4. The possible chains of data processing

The following parameters have been subject of investigation:

a) The sensitivity setting of the LF sensors. Default value of this setting is 80 V/m, The SAFIRK uses this default value operationally. FLITS uses operationally a much lower setting of approx. 20 V/m, this being different for each detection station. We applied the default settings and the adjusted settings
b) The processing of Strokes was performed with the Old type of Processing (like in SAFIRK) as well as with TOA processing.
c) The parameter “Discri Max Time Window” was changed. This parameter influences the time that a multiple discriminated strokes in Old_discrimination mode are assigned to the same location. We applied the default value (700 ms) as well as the allowed minimum value (1ms), in order to effectively to switch off this behaviour.

The eight possible parameter-sets have been introduced in the parameter settings of DAM, resulting in eight reprocessed data sets. All datasets including the reference set are found in Table 7.

Table 7. The nine different datasets and respective settings

RUN	Discrimination	Sensitivity	Window
SAF	Old_discrimination	80 V/m sensitivity	Discri_window=0ms
TSL	Time_Of_Arrival	80 V/m sensitivity	Discri_window=700ms
TFL	Time_Of_Arrival	20 V/m sensitivity	Discri_window=700ms
OSL	Old_discrimination	80 V/m sensitivity	Discri_window=700ms
OFL	Old_discrimination	20 V/m sensitivity	Discri_window=700ms
TSS	Time_Of_Arrival	80 V/m sensitivity	Discri_window=1ms
TFS	Time_Of_Arrival	20 V/m sensitivity	Discri_window=1ms
OSS	Old_discrimination	80 V/m sensitivity	Discri_window=1ms
OFS	Old_discrimination	20 V/m sensitivity	Discri_window=1ms

All data sets contain HDF5 files, to create HD5 files from T$YYMMDD.KNM files generated by SAFIRK a converter was used. Al the sets are then analysed using the same tools.

4.2 Results & Scores from the dataset and runs

The nine datasets are analysed on the occurrence of the type of events. The table does not present all repeated point of a Trace as this is less interesting. What is presented is total number of Flashes in the dataset, and the number of Strokes (First, Multiple, All), For completeness the number of Isolated events is presented too.
The scores for the different sets are presented in Table 8. This table allows for several conclusions:

Table 8. The nine datasets, high scores

Run	Type_0 Isolated	Type_1 Flash	Type_2	Type_3	Type_4 Gnd_stroke	Type_5 Rtn_stroke	Type_4+5 All strokes
SAF	101767	109658	536437	109682	31827	37011	68838
TSL	111625	116756	511181	116756	10566	928	11494
TFL	120096	120799	528025	120799	5769	627	6396
OSL	111623	116755	511181	116755	41575	28034	69609
OFL	120096	120799	528025	120799	42158	27852	70010
TSS	111623	116755	511181	116755	10554	928	11482
TFS	120096	120799	528025	120799	5769	627	6396
OSS	111623	116755	511181	116755	16256	2979	19235
OFS	120096	120799	528025	120799	15801	2620	18421

1) Whatever parameter is changed, the values in the columns Isolated and Flash do not differ much. A small influence is found for the sensitivity setting (+11% for changing sensitivity between OSL and OFL sets, SAF scores –11% to its most resembling set OSL), and reassuring, no influence is found for the change from TOA to Old Discrimination. In addition, no influence from changes in the “Discri_window” parameter is seen. Thus proving the latter two parameters only influence the discrimination processing.
2) Because the number of events in a Trace varies much, it is preferred to compare Flashes, in which a lightning is reduced to a single event. Vaisala sometimes presents the number of IntraClouds being the sum of all events in all Traces (sum of the columns Type_0 … Type_3). This is definitely not the number of events a human observer counts observing the sky, an observer is thought to count Flashes.
   Note that SAFIRK sometimes generates an endpoint to a lightning without starting one.
3) Comparing SAF and OSL:
   These two sets are almost identical in their settings, both have Old type discrimination, 80 V/m settings. Both score 70K Strokes. Note that, although the totals of strokes are almost equal, this

does not hold for the Type_4 to Type_5 ratios: SAF4/5= 0.85, OSL4/5= 1.48). This may reflect internal changes in the processing that assigns the multiple strokes.

4) Comparing SAF and TSL: (TOA,80V/m,0.7s)
We now switch to using the TOA system. In the SAF set the total amount of Flashes seen by SAFIRK and FLITS is in good agreement (+6% for FLITS), but a dramatic change occurs to the number of Strokes seen, Ground_strokes drop to less than 30%.
5) Still comparing SAF and TSL: (TOA,20V/m,0.7s)
Apparently, in TOA mode the multiplicity of the stroke drops to a such a low value, that it allows for the conclusion that a multiple discharge is a very rare phenomenon.
6) Comparing TSL and TFL: (TOA,80V/m,0.7s with TOA,20V/m,0.7s)
Comparing two TOA modes there seems to be a small influence caused by the sensitivity settings on the number of Flashes (+3 % for TFL), however impact is noted on the number of Strokes, they drop again with some 45%.
7) Neglecting multiplicity, comparing Flashes and Strokes we find the following rations:
SAF:Strokes/Flashes=0.29; TSL:Strokes/Flashes=0.09; TFL:Strokes/Flashes=0.05. Clearly SAFIRK over estimates the number of Ground_strokes. On the other hand, the ratio 0.05, when using recommended settings, is rather low.

4.3 Day-by-Day Comparison of Flashes

To get a more detailed view on the behaviour on the Old_discrimination compared to the TOA technique a day-by-day comparison is performed. Instead of the accumulated scores as presented in Table 8, the lightning events are now compared using their score per day. Although the dataset is of limited size, some details are noticeable.

The number of Flashes/day for the SAF reference set against the Flashes/day is depicted in Figure 5 for the two Old_discrimination sets, one with 80 V/m sensitivity and the other with 20 V/m sensitivity. Note that the Axes of the figure both have a logarithmic scale. The small offset (plotting N+1 instead of N) is introduced to avoid the undefined log(0). As mentioned before the number of Flashes detected changes only slightly under change of this parameter, which is well reflected in the figure. Closer inspection reveals some overestimation of SAFIRK on days with a few Flashes. This is believed to stem from the enhanced recognition of spurious responses in the software of FLITS.

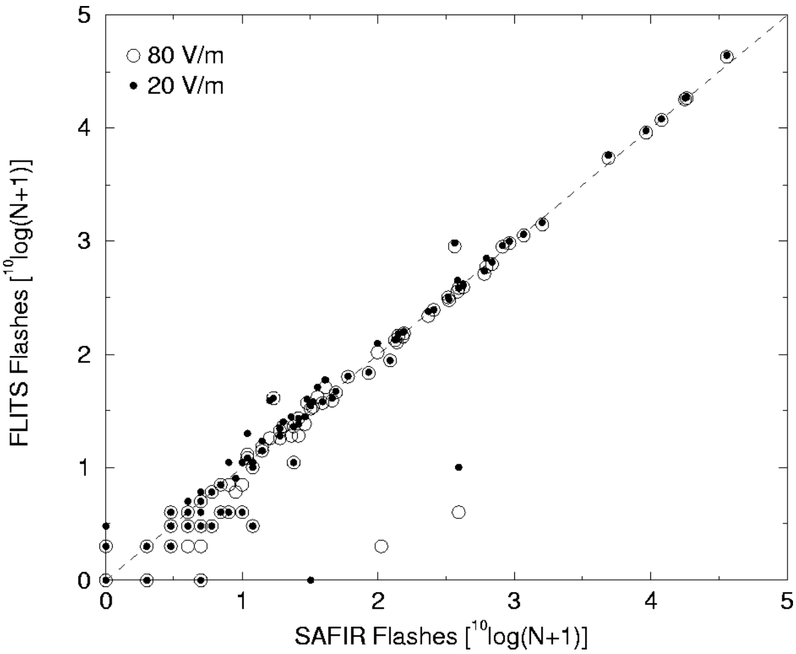

Figure 5. Number of Flashes;
FLITS against SAFIRK, Daily scores from 26 Jul to 28 Oct 2004

4.4 Day-by-Day Comparison of Strokes

When comparing the Strokes recorded by SAFIRK and FLITS in Old_discrimination mode shown in more scatter is observed. It is noted that the influence of the sensitivity parameter setting is small especially on days with substantial lightning. In days with a low number of lightning it is seen that changing the sensitivity parameter leads noticeable differences. This might reflect the changes in assignment of Type_4 en Type_5 events between SAFIRK and FLITS. Vaisala documentation reveals here that definitions of the types are changed, which may be reflecting another way of processing this data. The overestimation of SAFIRK is seen here as well. Some days FLITS did not notice any strokes while SAFIRK recorded over 30 strokes.

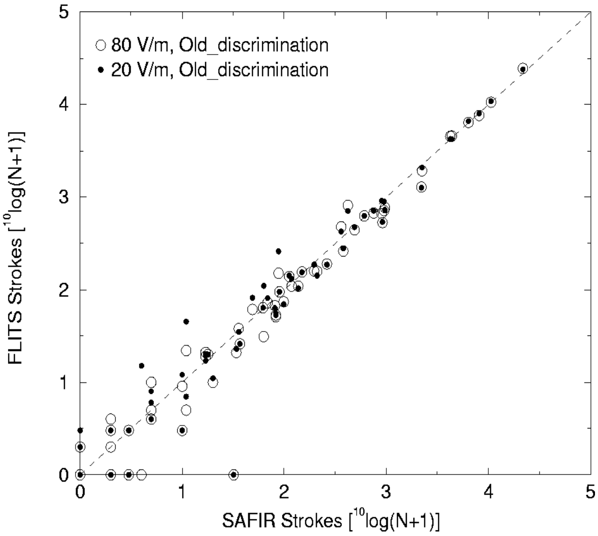

Figure 6: Number of Strokes; OSL, OFL against SAF

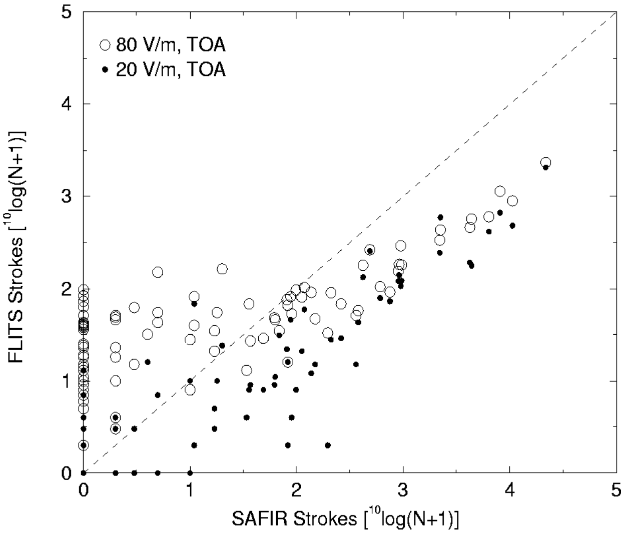

Figure 7. Number of Strokes: TSL, TFL against SAF

Finally a comparison is made of the Time of Arrival processing against the Old_discrimination of SAFIRK Figure 7 shows that both methods yield different results. With the 20 V/m sensitivity setting the total number of strokes is only weakly correlated. The total number of strokes recorded by FLITS differs an order of magnitude with the numbers seen by SAFIRK.

In the 80 V/m default setting, the correlation almost completely vanishes. Although the numbers of discharges rise to a more realistic value, several days on which FLITS scores close to hundred Strokes are seen, whilst none are seen by SAFIRK. This stresses the importance of having applied the sensitivity settings correctly when using the TOA system. This behaviour poses a serious problem to KNMI with using the TOA method. As one of the main characteristics of system changes so much with the parameter settings, the question is which value of the parameter is the “real” one,

4.5 Localisation accuracy

In the references 1) and 5) Wessels presented an estimation of the localisation accuracy of the SAFIRK system. This estimation is based on the assumption that if two Strokes occur within 0.2 second time of each other they are assumed to use the same physical discharge channel, and as a consequence they would have the same coordinates. To avoid contamination by faraway Strokes results were neglected when their relative distance is over 15 km. In Figure 8 and Figure 9 the vector between the independent results is printed. When this test was repeated for FLITS it did not deliver any results, as internally in the Vaisala software the Strokes are assigned the same coordinates. By tweaking the system parameter discriminaton_time_window on a low value the desired independence of Strokes was recovered.

It is pointed out that this is only needed in the case of using the Old_discrimination algorithm. The TOA algorithm already proved to deliver independent coordinates between separated Strokes.
In both figures, the same dot size is used for presenting a vector. It is clear that the Old_discrimination delivers far more points compared to the TOA algorithm. The scattering in Figure 8 looks like a Gausian distribution. The TOA scatter plot shows a smaller footprint. However some peculiar asymmetry is seen in the plot of Figure 9.

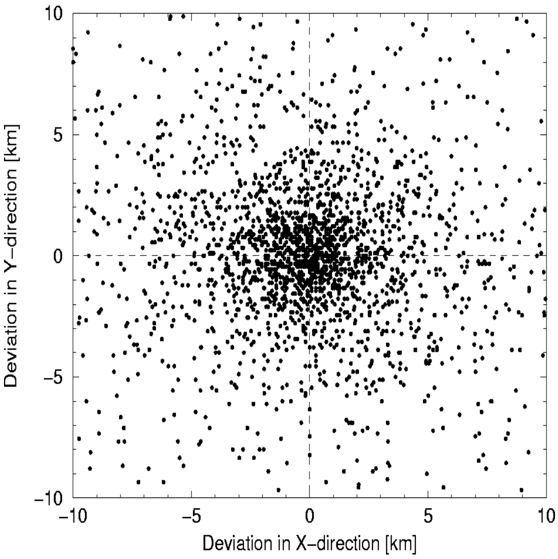

Figure 8: FLITS position deviation in Old_discrimination mode

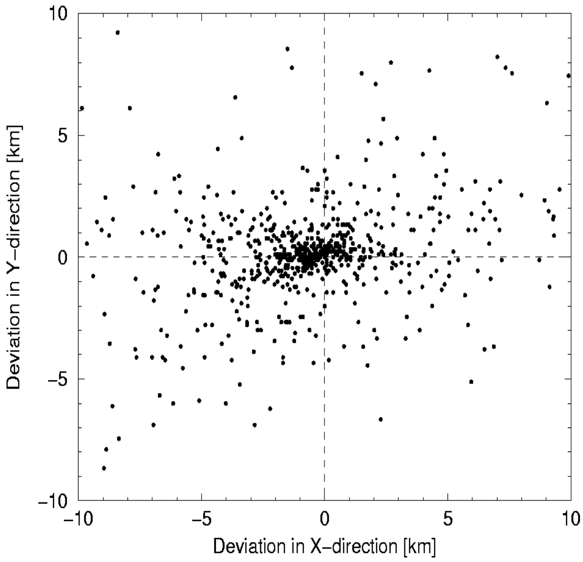

Figure 9. FLITS position deviation in TOA mode

In Figure 10 the histogram of distances is given. In this histogram the number of points are plotted that fall in the ring between R1 and R1 + 1 km, for each km. Median value for the Old_discrimination calculated is 3.7 km. For the TOA algorithm a median value of 2.0 km is found. It is remarked that the histogram Old_discrimination reveals a Raleigh distribution, which is the expectation when samples in a distribution are independent. Obviously, the histogram for the TOA algorithm does not resemble a Raleigh distribution. This may point to dependencies in the software that hide the real nature of the phenomenon under study.
Wessels reported a median value of 1.8 km. This smaller value might be obtained because Wessels used an area, that was favourable for a small median value, because of the good view by all the three detection stations used at that time.

It is remarked that once SAFIRK is taken out of operation, the KNMI Detection Stations are set in high-resolution mode. This may improve the accuracy of the localisations.

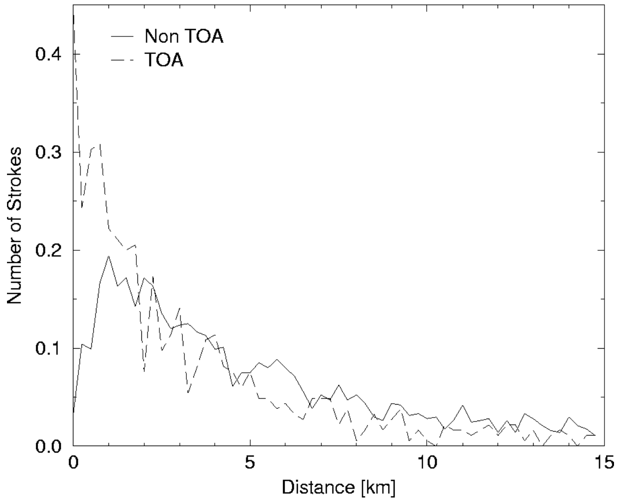

Figure 10. Histogram of distances

5 CONCLUSIONS

The conclusions to be drawn from evaluating the FLITS system are summarised here:

1) In respect to the Flashes determination, both systems agree well. There appear to be some enhancements in the software that suppress false Flashes in FLITS.

2) In respect to Stroke determination, there is good agreement between FLITS ad SAFIRK when using the Old_discrimination in FLITS. By analysing multiple Strokes, the localisation accuracy shows an overall median value of 3.7 km.

3) When using the TOA method to discriminate Strokes the agreement between both systems is poor. It is observed that the TOA method offers a better spatial resolution, an overall median value of 2.0 km is found.

4) The spatial distribution of multiples strokes does not show a Raleigh distribution. This may point to dependencies in the software that hide the real nature of the phenomenon under study. This is reflected in Figure 9, the scatter plot of multiple Strokes, showing an asymmetric behaviour.

5) The number a Strokes detected is highly dependant on the sensitivity settings of the LF sensors. This parameter does not affect the number of Traces detected.

6) When applying the sensitivity settings, as recommended by an auto analysis procedure, the ratio Strokes/Flashes drops to 1:20, which is a rather low value.

6 RECOMMENDATIONS

The quality of the Lightning detection system FLITS shall improve when the calibration procedure of a Detection Station takes the antenna array in its measuring loop. The implementation of the so called CNX test is highly recommended.

The Documentation of the Vaisala systems is insufficient in respect to the Algorithms used to derive user data. For scientific as well as meteorological use of the Vaisala Lightning Detection System it is of vital importance that a complete description of the data processing, including all methods that act upon the data, is available. It is strongly recommended that Vaisala will make the description of algorithms public for the lighting community.

The TOA method now is implemented as a separated module, the user can choose to use the Old_discrimination method, or the TOA method. Both methods should be integrated in order to use each other’s stronger points. In the meanwhile KNMI will use the Old_discrimination algorithm.

7 REFERENCES

1) Roozenkrans, H.; Holleman,I. KNMI HDF5 Data Format Specification v3.5 KNMI internal report IR 2003-05
2) Wauben, W. M. F. AUTOMATION OF VISUAL OBSERVATIONS AT KNMI; (I) COMPARISON OF PRESENT WEATHER, entry J3.1, AMS Symposium 2002, Orlando, Florida.
3) Wauben, W.M.F. SHM-056 [SDD_CIBIL], 12 April 2001, System Definition Document for the KNMI HIM project
4) Wessels, H.R.A. Experimental evaluation of an arrival time difference lightning positioning system. KNMI Scientific Report 92-01
5) Wessels, H.R.A. Evaluation of a radio interferometry lightning positioning system. KNMI Scientific Report 98-04
6) Evenden, G.: 1990, PROJ.4 Cartographic Projections Library (see http://www.remotesensing.org/proj/) developed at US Geological Survey.

8 TABLES AND FIGURES