Spatial Correlation Between Lightning Strikes and Whistler Observations

Jonas Öster

Master of Science Thesis
Stockholm, Sweden 2008
Spatial Correlation Between Lightning Strikes and Whistler Observations

Jonas Öster

May 20, 2008

Advisors:
Dr Andrew Collier, University of KwaZulu-Natal
Prof. Lars Blomberg, Royal Institute of Technology
Abstract

A whistler wave is a Very Low Frequency (VLF) trace that obtains its characteristics from dispersive propagation in the magnetosphere. Field aligned ducts of enhanced plasma density ensure the propagation from one hemisphere to the other. The origin of these signals is lightning strikes that emit radiation which spans the entire spectrum with the bulk being in the VLF band. The VLF portion can travel great distances within the Earth-ionosphere waveguide (EIWG) before penetrating through the ionosphere, and exciting a duct. The relative location, compared to the duct, of those strikes that cause whistlers is unknown.

It is of interest to examine where the whistlers that have been observed at Tihany, Hungary, and Dunedin, New Zealand, originate. This is one tool to gain further understanding of the properties, especially the plasma density structure, of the ionosphere and the magnetosphere. Therefore time series with observed whistlers from these stations has been correlated with lightning data obtained from the World Wide Lightning Location Network (WWLLN). The results show that whistlers observed at Tihany mainly originate from lightning in an area surrounding the magnetic conjugate point which is situated in the ocean just off East London, South Africa. This area, called the source region, has a radius slightly less than 1000 km. Whistlers also originate from lightning activity over the rest of Southern Africa and the northern parts of South America. A clear diurnal distinction is seen in that the correlation is maximized when the whistler station and the source region are covered in darkness. This is believed to relate to the diurnal variation of the ionospheric profile, which becomes more transparent to VLF waves at night. A similar diurnal correlation pattern for Dunedin was also obtained. The general correlation results for Dunedin were very sporadic.

Whistler statistics for the two stations and lightning statistics for the Tihany’s magnetic conjugate point are also presented. It reveals a general diurnal maximum in received whistlers in dark hours for Tihany with absolute maximum at 1 UTC and for Dunedin, the maximum occurs in the afternoon with absolute maximum at 15 UTC. It also reveals a seasonal maximum when the conjugate point is in the summer season. The lightning statistics for Tihany’s magnetic conjugate point reveals a diurnal maximum ranging from the afternoon until a couple of hours after midnight. Something worth noting is the delay between the peaks of lightning activity and whistler registration at Tihany. The lightning activity peaks around 18 UTC. The explanation is once again believed to relate to the behavior of the ionosphere in darkness.
Acknowledgements

I would like to thank the following people for their help and contribution to the success of this project. First of all I really want to thank Dr Andrew Collier at the University of KwaZulu-Natal, Durban, South Africa. Dr Collier was the main supervisor over the project. Without his professional advice, help, input, and support this project would have consumed considerably more time. I want to thank Prof. Lars Blomberg at the Royal Institute of Technology, Stockholm, Sweden, who has been my Swedish supervisor. Prof. Blomberg was the initiator of the project and has offered input during its course. I am indebted to Doc. Stephan Buchert at the Swedish Institute of Space Physics, Uppsala, Sweden, who has contributed by proofreading the report as well as offering input during the project. Prof. A. Hughes at the University of KwaZulu-Natal, Durban, South Africa; he was a part of the research group in Durban and provided ideas for this project. I would like to thank the various WWLLN hosts for keeping this global network functional. Also, the two whistler stations are acknowledged for their contribution of data. Mikael Lundberg for his help getting me started with $\LaTeX$. Other colleagues in Durban: Andreas, Charles, Judy, Mikael, Remmy, and Thomas. I want to thank you all for being such great friends during my time in South Africa! My family, who always provides me with food boxes and other help in every way possible, should also be mentioned. And last, but by no means least, Jessica Lister, my girlfriend, who always supports me and in every way made my stay in South Africa so much more fun, interesting, and enjoyable.
Preface

The reader of this report is assumed to have a basic understanding and knowledge in physics. The reader is also assumed to understand some programming, such as the basics in any of the most common languages C++, Matlab, Java or, as used in this report, R.

The chapters in this report can briefly be described as follows:

**Chapter 1** introduces the reader to lightning in general and it describes the theory behind the World Wide Lightning Location Network (WWLLN) as well as introduces the reader to what whistler waves really are.

**Chapter 2** goes more in depth on what the goal of this project was. It describes some more specific settings in the WWLLN and whistler receiver stations. It also displays the methodology used when deriving the data and results.

**Chapter 3** describes pure lightning statistics derived from the WWLLN concerning the Tihany magnetic conjugate point. Graphs representing seasonal and diurnal variations are displayed.

**Chapter 4** presents whistler statistics from the two whistler stations used in this report, Tihany, Hungary, and Dunedin, New Zealand. Graphs representing seasonal and diurnal variations for both stations are presented.

**Chapter 5** gives a detailed description of the correlation results obtained. Correlation is defined mathematically and correlation methodology is discussed. The results are presented with extensive graphs and a discussion concerning the results and the validity of the results is made.

**Chapter 6** presents the conclusions made from this project. It also outlines some future research suggestions.
Abbreviations

ATD = Arrival Time Difference
CC = Cloud to Cloud
CG = Cloud to ground
EIWG = Earth Ionosphere Waveguide
FS = Frame Size
IR = Infra Red
LIS = Lightning Imaging Sensor
LT = Local Time
MF = Medium Frequency (approximately 0.3-3 MHz)
NLDN = National Lightning Detection Network
NOAA = National Oceanic and Atmospheric Administration
OTD = Optical Transient Detector
SANAE = South African National Antarctic Expedition
TOA = Time of Arrival
TOGA = Time of Group Arrival
TRMM = Tropical Rainfall Measuring Mission
TS = Time Slice
TW = Time Window
UTC = Universal Time Coordinated
VLF = Very Low Frequency (approximately 3-30 kHz)
WWLLN = World Wide Lightning Location Network
Contents

Acknowledgements iii
Preface v
Abbreviations vii

1 Introduction 1
  1.1 Lightning 1
  1.2 World Wide Lightning Location Network 3
  1.3 Whistler Waves 5

2 The Project 9
  2.1 Goal 9
  2.2 Input Raw Data Structure 9
  2.3 Methodology 11

3 WWLLN Statistics 13
  3.1 WWLLN Settings 13
  3.2 Tihany Conjugate Point 14

4 Whistler Statistics 19
  4.1 Tihany, Hungary 19
  4.2 Dunedin, New Zealand 22

5 Correlation 27
  5.1 Correlation Definition 27
  5.2 Correlation Methodology 28
  5.3 Settings and Abbreviations 29
  5.4 Correlation Results for Tihany, Hungary 30
  5.5 Correlation Results for Dunedin, New Zealand 33
  5.6 Discussion and Validity of Correlation Approach 37

6 Conclusions 41
  6.1 Conclusion of the Project 41
Contents

6.2 Future Research Suggestions ........................................... 41

A Appendix, Lightning Statistics ........................................... 43
B Appendix, Whistler Statistics ........................................... 47
C Appendix, Correlation .................................................... 53
D Appendix, Source Code ................................................... 57

Bibliography ................................................................. 83
List of Figures

1.1 Cloud-to-cloud lightning .......................... 2
1.2 Global lightning distribution ....................... 3
1.3 Whistler creation and propagation ................. 6
1.4 Sample whistler wave ............................. 7

2.1 Whistler database .................................. 10
2.2 Lightning database ................................. 11

3.1 Lightning divided into year of registration at Tihany, Hungary, magnetic conjugate point .................... 15
3.2 Total amount of registered lightning at Tihany, Hungary, magnetic conjugate point ......................... 15
3.3 Seasonal distr of lightning at Tihany, Hungary, magnetic conjugate point .................................... 16
3.4 Diurnal distr of lightning at Tihany, Hungary, magnetic conjugate point .................................... 17
3.5 Diurnal distr of lightning divided into months at Tihany, Hungary, magnetic conjugate point ........... 18

4.1 Whistlers divided into year of registration at Tihany, Hungary .................................................. 19
4.2 Total amount of registered whistlers at Tihany, Hungary ............................................................... 20
4.3 Seasonal distr of whistlers at Tihany, Hungary ................................................................. 21
4.4 Diurnal distr of whistlers at Tihany, Hungary ................................................................. 21
4.5 Diurnal distr of whistlers divided into months at Tihany, Hungary ................................................ 22
4.6 Total amount of registered whistlers at Dunedin, New Zealand .................................................... 23
4.7 Seasonal distr of whistlers at Dunedin, New Zealand ................................................................. 24
4.8 Diurnal distr of whistlers at Dunedin, New Zealand ................................................................. 25
4.9 Diurnal distr of whistlers divided into months at Dunedin, New Zealand ........................................ 25

5.1 Correlation Tihany, Hungary, FS=3° TW=30s TS=0–24 UTC ......................................................... 30
5.2 Correlation Tihany, Hungary, FS=1° TW=30s TS=0–24 UTC ......................................................... 31
5.3 Correlation Tihany, Hungary, FS=0.5° TW=30s TS=0–24 UTC ......................................................... 32
5.4 Correlation Tihany, Hungary, FS=1° and 0.5° TW=15s TS=0–24 UTC ............................................. 32
List of Figures

5.5 Correlation Tihany, Hungary, FS=1° TW=30s TS=3h ........ 34
5.6 Correlation Dunedin, New Zealand, FS=1° TW=30s TS=0 – 24 UTC ............................................. 35
5.7 Correlation Dunedin, New Zealand, FS=1° TW=30s TS=3h .......................... 36
5.8 Days to include or exclude in the correlation analysis for Tihany, Hungary ........................................... 37

A.1 Diurnal distr of lightning for Jan and Feb at Tihany, Hungary, magnetic conjugate point ........................................... 44
A.2 Diurnal distr of lightning for Mar and Apr at Tihany, Hungary, magnetic conjugate point ........................................... 44
A.3 Diurnal distr of lightning for May and Jun at Tihany, Hungary, magnetic conjugate point ........................................... 44
A.4 Diurnal distr of lightning for Jul and Aug at Tihany, Hungary, magnetic conjugate point ........................................... 45
A.5 Diurnal distr of lightning for Sep and Oct at Tihany, Hungary, magnetic conjugate point ........................................... 45
A.6 Diurnal distr of lightning for Nov and Dec at Tihany, Hungary, magnetic conjugate point ........................................... 45

B.1 Diurnal distr of whistlers for Jan and Feb at Tihany, Hungary ...... 48
B.2 Diurnal distr of whistlers for Mar and Apr at Tihany, Hungary ...... 48
B.3 Diurnal distr of whistlers for May and Jun at Tihany, Hungary ...... 48
B.4 Diurnal distr of whistlers for Jul and Aug at Tihany, Hungary ...... 49
B.5 Diurnal distr of whistlers for Sep and Oct at Tihany, Hungary ...... 49
B.6 Diurnal distr of whistlers for Nov and Dec at Tihany, Hungary ...... 49
B.7 Diurnal distr of whistlers for Jan and Feb at Dunedin, New Zealand 50
B.8 Diurnal distr of whistlers for Mar and Apr at Dunedin, New Zealand 50
B.9 Diurnal distr of whistlers for May and Jun at Dunedin, New Zealand 50
B.10 Diurnal distr of whistlers for Jul and Aug at Dunedin, New Zealand 51
B.11 Diurnal distr of whistlers for Sep and Oct at Dunedin, New Zealand 51
B.12 Diurnal distr of whistlers for Nov and Dec at Dunedin, New Zealand 51

C.1 Correlation Dunedin, New Zealand, FS=1° TW=30s TS=1h ........ 54
C.2 Correlation Dunedin, New Zealand, FS=1° TW=30s TS=1h ........ 55
1 Introduction

The best way out is always through.

-Robert Frost

1.1 Lightning

The occurrence of lightning and thunderstorms are due to a number of factors. The Sun plays the most significant role along with additional geographical features that strongly affect whether or not a storm is to be created in a region. The Sun is heating air over ground, this results in the warm air rising due to decreased density compared to colder air above and around, for example over surrounding water such as an ocean. This warm air contains moisture that is turned into ice with decreased temperatures. When this takes place there is a charge separation due to the relative motion of water and ice within the forming cloud. Finally the electrical field exceeds the resistivity in the surrounding medium (air) and a channel of current i.e. lightning is formed to neutralize the field.

The charge separation within the cloud takes the form of a negative base and a positive top [6]. This in turn affects the surrounding environment with, for example, a slight charge separation in the very ground, which then turns slightly positive around the surface and more negative further down. Having this complex separation of charges means that there are many types of lightning that can occur. The most common to public knowledge are the so-called cloud-to-ground flashes or CG's. The most common lightning in terms of numbers, however, are cloud-to-cloud lightning or CC's. A snapshot of a CC lightning flash is presented in Figure 1.1. CC's outnumber CG's in general with a ratio of approximately 6:1 in a thunderstorm [4]. Close to the equator this number might be even greater, however it drops off closer to the poles [4]. The reason for this is because the cumulonimbus clouds can extend to greater altitudes close to the equator due to the higher altitude (approximately 15 km) where the tropopause is found [6]. The altitude of the tropopause decreases the closer to the poles you go. In general, cumulonimbus clouds cannot extend above the tropopause which acts as a roof for the clouds [6]. This means that lightning is actually more dangerous the further away from the equator you
go since there is a greater possibility that a lightning flash will actually strike down to the ground instead of staying within the cloud. However, lightning as a general phenomenon is far more common around the equator due to temperature and moisture conditions.

Figure 1.1: Most lightning flashes are cloud-to-cloud, CC’s. Picture obtained from NOAA

A lightning flash consists of several strikes. When a CG lightning flash occurs there is first a build-up of the lightning within the cloud due to small local differences in charge distribution in the area where the lightning originates. This initial strike will then grow and extend down toward the ground due to the major electrical field that it “feels”. After reaching a significant distance, another strike is originated from the ground itself. When these two strikes meet we have a complete strike that carries current within itself. It is the strike that originates from the ground that is of most interest when it comes to locating the lightning flash in terms of position. Several strikes will then use the same channel. Different strikes can have different strengths of the current within the same flash, and the individual strikes can last different lengths of time. The current in a strike is in the kA-range, and the length is in the μs-range. Different lightning location networks use different strikes to register a lightning flash. This is due to the fact that different strikes do not radiate identical electromagnetic waves. This is of no significance as long as the complete system of receivers is operating on the same band [1]. Lightning strikes are in general generating electromagnetic waves ranging from a few Hz all the way up to, and well beyond, optical frequencies [2].

There are, of course, many different kinds of lightning. Positively charged lightning from the top layer of a cumulonimbus cloud can reach the ground far aside from the base of the cloud, and this lightning is generally very strong [6]. Sprites are an example of an event related with ordinary lightning. A
sprite is a discharge that originates from the top positively charged layer of a cloud and reaches all the way up to the ionosphere. They can be 100 km long and 10 km wide, i.e. they are shaped as a cone [6].

Lightning exists all over the world. Figure 1.2 presents the global distribution of lightning. Lightning is most prevalent in the equatorial region which is due to the high insolation. The lightning density is also higher over land then over ocean, which depends on land being heated faster than the ocean which in turn heats the air over land leading to convection, cloud formation, and consequently, lightning, as discussed above. Furthermore, lightning is in general most prevalent in the late afternoons on a diurnal basis since the heating of the ground reaches its maximum when the insolation equals the outgoing Infra-Red (IR) radiation. The same argument can be applied on a seasonal level which makes lightning most prevalent in late summer and the beginning of fall.

Figure 1.2: Distribution of global lightning. Image derived from data generated by LIS during 1997-2002 and OTD during 1995-2002, LIS is a part of TRMM. Image obtained from [7]

All these factors combined make lightning strikes being a very complex phenomenon. It is therefore very hard to construct a lightning sensor system that can observe each and every flash. The World Wide Lightning Location Network (WWLLN) described below has a goal of catching 50% of all CG flashes in the future [3].

1.2 World Wide Lightning Location Network

The World Wide Lightning Location Network (pronounced “woollen”) is a system developed during the last decade to detect lightning strikes on a global coverage basis. Currently the system consists of 30 stations spread around
the world [5]. The system is designed to detect the signature of strikes in the Very Low Frequency (VLF) band, which suffer from very low attenuation when propagating from the point of origin. This is due to the fact that VLF signals propagate in the Earth-ionosphere wave guide (EIWG), which is found between the ionosphere and the ground. This allows the signals to travel thousands of kilometers.

However, a drawback using the VLF portion of the wave and one reason that this has not been implemented earlier is that the wave-front is smeared out into a wave-train, or spheric, with no clear front or “starting-point” when propagating in the EIWG. This creates a systematic error if using a threshold amplitude for registration since the amplitude is inversely dependent on distance from origin [1]. That makes it hard to detect the beginning of the wave and thus the Time of Arrival (TOA) which gives the Arrival Time Difference (ATD) in combination with other station’s measurements, which in turn gives the location of the lightning strike.

Therefore other, mostly commercial, lightning detection systems have historically used much higher frequencies, i.e. the Medium Frequency (MF) band (0.3-3 MHz) [2]. This eliminates the EIWG wave and one is only left with the ground-wave with a clear and sharp wave-front, i.e. a pulse. The drawback with this system is that these pulses suffer from relatively high attenuation when spreading, so one needs a large number of stations to collect the data. This is expensive and sometimes unfeasible, for example over oceans. As a reference one can take the National Lightning Detection Network (NLDN), located in the continental USA, which used a total of 106 stations in 1996 to cover an area of approximately the size of 107km², i.e. the USA [2]. This is a very large concentration compared to what WWLLN has.

There are two solutions to the problem of registering the ATD with high accuracy in the VLF band. They are both based upon recording the whole wave-train [1]. The first one calculates the ATD by comparing the complete wave-trains to each other. This works well but turns out to be very bulky since one ends up with a relatively large amount of data to manage. This can be very time consuming and costly. The second solution is to use the Time of Group Arrival (TOGA). Here one is concerned with the rate of change of the phase of the wave-train compared to its frequency at the trigger time in the receiver [1]. Dowden et al. [1] defines the TOGA as “…that instant when the regression line of phase versus frequency over a specific band has zero slope”. In other words, one can describe the TOGA method as a measurement of where exactly within the wave-train the change of phase with respect to frequency is zero. This point is defined as the TOGA for a wave-train or spheric. This method has the advantage of being able to register lightning on a world coverage basis with a relatively small number of stations that the VLF band provides while at the same time remain accurate in the measurements which only MF band measurements could previously provide. The WWLLN is based upon the TOGA method.
1.3. Whistler Waves

In order to classify a lightning discharge as registered within the WWLLN it needs to be registered by at least four stations independently. The reason is pretty clear. Considering a flat world, i.e. two dimensions, one needs three stations in order to locate the point of origin of a discharge. An easy way to look at this is to reverse the situation and assume one has three stations which at the same time emits a radio pulse with the speed of light. In a two dimensional world, these three pulses would meet simultaneously at only one point. However, if the surface of a spherical body is considered, i.e. the Earth, one finds that even with three stations one gets two points where the waves interact simultaneously. One point is correct but there will also be another point situated on the other side of the sphere where the waves interact simultaneously. Therefore, to eliminate this false solution, a fourth station has to be added to obtain one and only one point of origin.

The individual WWLLN stations are thus only recording the TOGA of a lightning pulse. This data is then transmitted to a central data processor where the longitude, latitude, and time of the lightning discharge are calculated with high precision.

Today there is no other system that can compete with the WWLLN in terms of global coverage. The only realistic alternative is to use satellite measurements, which also has been done on a number of occasions. However, satellites can only cover a limited part of the planet during a specific time unless formations are used. No such system exists today and it is very costly to implement. Therefore the WWLLN is the only feasible tool for global lightning coverage today.

1.3 Whistler Waves

As previously described, a lightning strike emits electromagnetic radiation that spans from a few Hz all the way up to and beyond optical frequencies with the bulk of emitted energy in the VLF range [3]. As also described above this VLF wave follows the EIWG. However, fractions of the wave can actually, during favorable conditions, get caught by the Earth’s magnetic field. The wave might then penetrate through the wave guide and the ionosphere and enter the magnetosphere following the magnetic field lines, see Figure 1.3. Since the magnetosphere is occupied by plasma, the wave will suffer dispersion. This means that the wave will be smeared out forming a characteristic whistling tone. Hence the name: whistler wave.

When the wave has penetrated through the ionosphere into the magnetosphere there are two options: the wave can follow a duct of enhanced plasma density which is aligned with the magnetic field lines, or take another curved path that does not perfectly follow the field lines [9]. In the latter case the wave will not likely be able to penetrate back through the ionosphere in the opposite hemisphere [9]. In the first case, the enhanced plasma density along
Introduction

Figure 1.3: Whistlers are generated when VLF signatures penetrate through the ionosphere into the magnetosphere following a duct of enhanced plasma density. Image obtained from [8]

the magnetic field lines defines a duct. Ducted whistlers can re-enter the EIWG through the ionosphere.

Considering a ducted whistler, the magnetosphere will, due to its field lines, lock in the wave in the duct and the wave will follow the magnetic field lines and therefore re-enter the ionosphere and the atmosphere and reach down to the ground in the magnetic conjugate point as seen from the origin. That is, if there, for example, is a lightning strike at 60°S and 20°W, the wave will re-enter at approximately 60°N and 20°W, magnetic coordinates assumed.

At very low latitudes the inclination of the magnetic field is not favorable for trapping whistlers. So even though lightning is most prevalent here, whistlers are very rare. At medium latitudes whistlers become far more common. Whistlers recorded in this region have the general characteristics of higher frequencies arriving slightly before the lower ditto. At higher latitudes the whistlers will have a distinct nose-frequency. This means that after the initial frequency a tone of both rising and descending frequencies will be recorded.

What determines the specific shape of a whistler is the time spent in the dispersive magnetosphere together with other factors such as plasma density in the duct and strength of the magnetic field. Whistlers generated at higher latitudes spend more time in the duct thus experiencing greater dispersion and the nose-characteristic shape is formed. However, due to the sparse occurrences of lightning at high latitudes whistlers are fairly uncommon compared to middle latitudes.

Once the whistler has entered into the magnetosphere it can also mirror back and forth between the hemispheres, which can make it hard to distinguish where the whistler originates from. This is more prevalent for whistlers generated at higher latitudes [4]. Figure 1.4 shows a typical nose-whistler recorded at SANAEB-IV in Antarctica.

Far from all lightning strikes generate detectable whistlers. The creation of
1.3. Whistler Waves

Figure 1.4: Example of a whistler recorded at SANAEP-IV, Antarctica. Since it is recorded at relative high latitude it shows a clear nose-structure.

natural existing whistlers, i.e. not manmade, are due to VLF signatures from lightning strikes. However, the properties of the ionosphere [4] play a significant role, whether a wave will be able to penetrate through following a duct of enhanced plasma density and thus form a whistler or not. This justifies the interest in understanding the connection of lightning discharges and whistler waves. It is a tool to gain further understanding of the properties, especially the plasma density structure, of the ionosphere and the magnetosphere.
2 The Project

2.1 Goal

It is of interest to examine the spatial relationship between the regions where whistlers are generated and where they are observed. That is, where are all the whistlers that are being registered coming from?

Far from all lightning strikes generate a detectable whistler wave. An understanding of where the whistlers originate in combination with other approaches such as, for example, appropriate modeling and other theories regarding the ionosphere and the magnetosphere, can lead to a better understanding of how whistlers are generated and how they propagate in the ionosphere and the inner magnetosphere. This, in turn, might then offer a new method for obtaining information about the plasma density structure in the topside ionosphere and the magnetosphere which is otherwise difficult to measure. The plasma density structure is of great interest and importance, for example in regards to satellite based navigation. Summarizing this one could say that the approach used in this study of linking received whistlers with its causative lightning strikes can contribute toward a better understanding of the plasma density structure in the topside ionosphere and the magnetosphere.

In accordance with this, the goal of this project was to generate a statistical image of the region where the whistlers are generated for a particular whistler station. The method of obtaining this is to analyze the correlation of recorded lightning strikes and observed time-series of whistler waves.

The second goal after the image was complete was to examine whether or not times exist where the correlation is stronger than other times, i.e. to search for diurnal and seasonal differences and also, if possible, to isolate unique pairs of whistler waves and lightning strikes.

2.2 Input Raw Data Structure

The whistler data used in this report comes from two whistler stations. The first station is situated in Tihany, Hungary, 46.89°N, 17.89°E and the second station is located in Dunedin, New Zealand, 45.52°S, 170.3°E. The lightning data has been obtained from the WWLLN.
The whistler data consists of large arrays of information. It is basically a very large text file containing a list of entries which have the form of YYYYMMDDHHMMSS.SS where Y=year, M=month and so forth. This indicates when a whistler was recorded at the station concerned. For example, the set of data from Tihany, Hungary, consists of over 681000 whistler events that were used in this project. There is one recording per row in the file where the data is listed. A small snapshot of the whistler file from Tihany, Hungary, is listed in Figure 2.1.

```
20020227194331.14
20020227201513.11
20020227201513.12
20020228005948.16
20020228011020.14
20020228011020.20
20020228011712.12
20020228011712.49
20020228013247.16
20020228014442.37
```

Figure 2.1: Snapshot of events in the whistler database from Tihany, Hungary

The whistler stations that collect this data have a number of tunable parameters that makes it possible to distinguish a whistler. Examples include trigger levels i.e. what amplitude (strength) the signal must have to be considered an authentic whistler and not just noise. Also time separation, i.e. how large the time gap must be between two whistlers to make sure that they don’t originate from different strikes within the same lightning flash, is of importance. If these settings are not adhered to the signal will be filtered away by the algorithm at the whistler station. The efficiency of these stations is constantly improving and the algorithm updated. When the data that is used in this survey was collected the stations were still considered to be operating in an experimental mode [4]. One must bear these things in mind when considering the results.

The lightning data in this project is consistently taken from the WWLLN. The input files consist of large arrays of data just as in the whistler case. However, there are three fields per row in the data file. The first field represents longitude, the second field, latitude, and the third field gives the time in decimal days from epoch, which is defined as midnight on Jan 1st 1970. Figure 2.2 gives an example of the WWLLN data structure.

The sensitivity in the VLF receivers collecting lightning data is controlled in a similar pattern as the whistler stations. No further investigation in to how the individual WWLLN stations have been set is made in this report but
2.3 Methodology

The whistler data used in this report has, as mentioned, been collected at two separate stations. Tihany, Hungary, is considered the primary station because additional data that is of importance to the analysis is found and which has been used in this report. Examples are accurate terminator times (i.e. sunrise and sunset times) and a log file containing data from which the efficiency of the station can be extracted (see Section 5.6). Not to mention it is the largest database hence giving the largest statistical accuracy. As for Dunedin, New Zealand, no additional data has been used. Considerations were made for acquiring data but were turned down due to lack of time. In accordance these lightning statistics have been obtained only from Tihany's magnetic conjugate point but not from Dunedin's respective point. The effect of this is that the analysis in this report is more extensive for Tihany compared to Dunedin.

The approach of this project is based on three steps. In order to get useful results, and also to understand the final output, it was decided to generate general statistics for both whistlers and, for Tihany, also lightning before the actual correlation was made. This approach will give a deeper understanding of the final correlation results since the statistics will provide backup and help when analyzing the final result.

Figure 2.2: Snapshot of events in the WWLLN database. Longitude, latitude, and decimal days since epoch, Jan 1 1970

\[-167.4499 -78.7965 12082.4705709295\]
\[-166.7422 -76.2714 12196.2055423695\]
\[-174.8861 -77.2213 12298.7462467956\]
\[-178.8819 -75.9237 12306.9921309887\]
\[-772.6348 -82.3575 12495.7162462807\]
\[-772.6879 -78.8873 12502.96942767\]
\[-169.9489 -84.1538 12506.8604680261\]
\[-166.9075 -80.7991 12507.0727792883\]

the interested reader can find information on this matter in [1, 2, 3].

It is important to mention that all registered events, both for whistlers and lightning, in the data bases are treated as equally authentic. That is, every event is being considered to be a correct and accurate registration by the system and all events in the data bases are used in the analysis.
Software

In order to handle the raw data in a smooth way, and get useful statistics from it, a tool called “R” was used. R is a software similar to Matlab but is free to download and use. The main difference compared to Matlab thought is that R was created for being a particularly powerful tool in terms of statistics and graphics. This made R especially useful in this project. The construction of R was heavily influenced by S which is a program for data analysis and graphics. During this project R was run on a Linux system. R is, just as C++, Java, and also Matlab, a generic language which means that the user has extreme freedom in using variables and setting parameters. For the generation of the statistical maps a tool called Generic Mapping Tools (GMT), was used. Just like R it is a free software and is mainly used for the generation of different types of maps in various contexts. All source codes derived for, and used in this project, can be found in Appendix D.
3 WWLLN Statistics

3.1 WWLLN Settings

The efficiency of the WWLLN is low. To reach a high efficiency it is desirable to have around 500 stations evenly distributed around the world [5]. The WWLLN has gradually been expanded to reach its size of 30 stations in operation today. In 2004 the network consisted of 18 stations [3] and in 2003 only 11 stations were active [2]. It is worth to mention the fact that 7 out of these 11 stations were situated in Japan, Taiwan, Singapore, Australia, and New Zealand [2]. Furthermore, before March 2003 these seven stations were the only active stations in the whole network [2]. This strongly affects the overall efficiency of the system and creates a discrepancy with a significant weight in lightning registered in this part of the world, especially in the equator region where lightning is most prevalent. However, this discrepancy has gradually been removed with the expansion of the WWLLN. During the time that the whistler statistics used in this report were collected, WWLLN had an efficiency of approximately 2% [3], i.e. approximately 2% of the total global lightning was being detected. One must have these facts in mind when considering the results presented.

Despite these limitations the WWLLN registers an immense amount of lightning. The number of registrations in the data base used here is in the order of ~ one hundred million. The amount is sufficient to produce valid and accurate results from a statistical point of view.

The location accuracy for those lightning strikes that do get registered is fairly high. The TOGA method was fully implemented on August 1, 2003 in WWLLN. Before this date the location accuracy was estimated to have an error ranging between 7.5-100 km with the global mean being 30 km and the median 15 km [2]. After the TOGA implementation the accuracy increased and it was estimated to be between 1.9-19 km with a mean of 3.4 km and a median of 2.9 km on a global basis [3]. As the number of WWLLN stations increase it is also reasonable to assume that the location accuracy increases correspondingly. This means that the later part of the data has a significantly higher accuracy compared to the data collected in the beginning of the period. In the equatorial region one degree longitude or latitude corresponds to around 110 km. This means that the spatial resolution can be selected fairly high when
the correlation is performed in Chapter 5.

3.2 Tihany Conjugate Point

The magnetic conjugate point of Tihany, Hungary, is, as of 2002, located at 33.45°S, 28.34°E, which is in the Indian Ocean just off East London, South Africa. One can therefore assume that most whistlers registered in Tihany originate from around this area. However, it is impossible to state a certain whistler source region including physical boundaries for any whistler station, including Tihany. The coordinates of a discharge play a major role whether a whistler will be detected or not as well as the properties of the ionosphere and the magnetosphere which also has an affect.

However, in order to extract lightning statistics for the conjugate point a source region has to be defined. It must be large enough so that it will contain enough registrations to form a statistical base, but it must also be small enough not to contain strikes that originate from a location which is typically not a part of the source region due to its location. In this report a source region with the center in the conjugate point and with a radius of 500 km has been selected. This distance was selected fairly arbitrarily but it contains 57385 strikes (see Figure 3.2) which is considered a large enough number of registrations for obtaining good statistical accuracy. Furthermore, the radius is also small enough to fit into that region that the correlation results (see Chapter 5) later relieved being the major source region for whistlers observed in Tihany, Hungary. Attempts have been made earlier to define a source region and Collier et al. [4] defined a source region with a radius of 600 km surrounding the magnetic conjugate point, mainly to fit in with data obtained from the LIS satellite. There are no set rules and one has to consider what the goal of defining such a source region is. Here, the purpose is to be able to extract lightning statistics used for comparison with whistler statistics.

When analyzing the graphical figures presenting lightning statistics below it must be taken into consideration that it is the accumulated numbers being discussed. That is, if, for example, the lightning density for a particular hour is presented it is the total amount of lightning, between the dates given in the figure, that were registered within that hour that is presented. It is not any mean of amount/hour and day but the accumulated amount during the whole period for this particular hour. Therefore the total sum of all lightning discharges presented will be the same in all figures.

Figure 3.1 shows the distribution of lightning strikes divided into the year that the registration took place. It is clear that 2004 was by far the most active year with over 30000 registrations. This is more than the other three years combined.

There are two interesting conclusions that can be drawn from analyzing the extracted WWLLN statistics. First of all a seasonal dependence in lightning
3.2. Tihany Conjugate Point

Figure 3.1: Lightning divided into year of registration at Tihany, Hungary, magnetic conjugate point. 2003-01-11 through 2006-04-23. Selected circular area surrounds Lat=$-33.45^\circ$ Long=$28.34^\circ$ with R=500km

Figure 3.2: Total amount of registered lightning at Tihany, Hungary, magnetic conjugate point. 2003-01-11 through 2006-04-23. Selected circular area surrounds Lat=$-33.45^\circ$ Long=$28.34^\circ$ with R=500km. Each bin has a width of one day
occurrence can clearly be observed. Figure 3.2 presents all lightning strikes that the WWLLN has detected from Jan 11, 2003 to April 23, 2006 within the defined source region. For 2003, 2004, and 2005 a clear seasonal pattern can be seen with peaks in the detection rate during and after New Year's. For 2006 the peak is less significant. The picture gets even more clear when Figure 3.3 is analyzed. The occurrence of lightning during the southern hemisphere summer (Dec-Feb) and the beginning of the fall is by far larger than during the winter (Jun-Aug) and spring (Sep-Nov). This is also expected considering that lightning activity increases with a slight seasonal delay relative increased insolation, as discussed in Chapter 1.

![Seasonal distribution of lightning](image)

Figure 3.3: Seasonal distribution of lightning at Tihany, Hungary, magnetic conjugate point. 2003-01-11 through 2006-04-23. Selected circular area surrounds Lat=−33.45° Long=28.34° with R=500km. Data is not normalized, thus is February and Mars presented for 4 years, January for 4 years minus 11 days, April for 4 years minus 7 days. All other months are presented for 3 years.

The other interesting feature concerns the diurnal variation. This is shown in Figure 3.4. The occurrence of lightning peaks around 20 local time with over 6000 flashes per hour. However, the pickup begins already in the early afternoon and the overall peak does not drop off until around 4 local time. From 5 up until noon local time the lightning activity is, in general, less than 500 flashes per hour. Basically one can therefore divide the day into two distinct sections: one starting from around 4 stretching for approximately 10 hours during the day with very low lightning activity and the other period, with relative high lightning activity, stretching from early afternoon through the night. This result is very logical considering that lightning is heavily
dependent on temperature gradients which are being built up during the day in the clouds. This energy is then released when the clouds have formed in the afternoon.

Figure 3.4: Diurnal distribution of lightning at Tihany, Hungary, magnetic conjugate point. 2003-01-11 through 2006-04-23. Selected circular area surrounds Lat=−33.45° Long=28.34° with R=500km

Figure 3.5 shows both the diurnal and the seasonal variation. The absolute maximum amount of lightning activity is during February and March from approximately 19 up until midnight. Consequently the absolute minimum takes place in June and July from around 6 up until noon. Something worth mentioning is that the diurnal difference between maximum and minimum is around 12 hours whereas the seasonal difference is only a couple of months. However, the number of lightning discharges observed in the period of low activity, both seasonal and diurnal, is so low that just a couple lightning discharges will change the look of the figure. Therefore, this difference is most likely due to random effects. A more clear two dimensional representation of Figure 3.5 is presented in Appendix A.
Figure 3.5: Diurnal distribution of lightning divided into months at Tihany, Hungary, magnetic conjugate point. 2003-01-11 through 2006-04-23. Selected circular area surrounds Lat=$-33.45^\circ$ Long=$28.34^\circ$ with R=500km
4 Whistler Statistics

4.1 Tihany, Hungary

This data base consists of 681107 registrations collected between February 27, 2002 and May 18, 2005. On a yearly basis most whistlers were collected during 2004 with almost 300000 registrations on its account. 2003 and 2005 account for two more years with heavy activity with approximately 170000 and 160000 registrations respectively. Considering that the data ends in May 2005 the amount collected this year is impressive. However, it is likely explained by the fact that the first months of a year are the high season which accounts for the major part of the whistlers received on a total basis for the year.

![Whistlers divided into year of registration at Tihany, Hungary. 2002-02-27 through 2005-05-18](image)

Figure 4.1: Whistlers divided into year of registration at Tihany, Hungary. 2002-02-27 through 2005-05-18

This is indeed clarified in Figure 4.2 and 4.3 which presents the total amount of whistlers received and the monthly distribution of whistlers during the period. Just as in the WWLLN case, a clear seasonal dependence can be observed by analyzing these graphs. November through March can
Figure 4.2: Total amount of registered whistlers at Tihany, Hungary. 2002-02-27 through 2005-05-18. Each bin has a width of one day.

be considered “high season” whereas April through October is “low season”. Analyzing Figure 4.3 the seasonal pattern becomes clear. February is by far the most active month. However, Collier et al. [4] found that Feb 14 and 15, 2003 and Feb 26 and 27, 2004 accounted for approximately 65000 whistlers. This suggests that these days were extraordinarily active in lightning activity which in combination with other favorable properties allowed whistler transfer in the magnetosphere and reception at the station on an abnormally high scale. If those whistlers are removed from the plot a sinusoidal pattern is more distinguishable on a seasonal basis.

The other identified dependence is, just as with the lightning case, the diurnal variation. Figure 4.4 is displaying this dependence. A maximum in whistler reception occurs between approximately 17 local time and lasting until 5. During this time the received accumulated number of whistlers per hour varies between 35000, received in the very beginning and the very end of the interval, and 60000 around midnight. Between 5 and 17 the accumulated number of received whistlers per hour is altering around the level of 1000. This is quite a significant diurnal variation.

Comparing Figure 3.4 and Figure 4.4 reveals an interesting feature. Lightning activity in the conjugate point peaks around 18 UTC whereas the maximum amount of observed whistlers peaks after midnight UTC. This indicates that the properties of the ionosphere play a very significant role in the generation of whistlers. When the level of ionization in the ionosphere decreases at night due to decreased insolation it seems to become more transparent for the VLF traces from lightning thus increasing the numbers of whistlers even
4.1. Tihany, Hungary

Figure 4.3: Seasonal distribution of whistlers at Tihany, Hungary. 2002-02-27 through 2005-05-18. Data is not normalized, thus Mars and April presented for 4 years, May for 4 years minus 13 days, and all other months for 3 years.

Figure 4.4: Diurnal distribution of whistlers at Tihany, Hungary. 2002-02-27 through 2005-05-18.
though the peak time for lightning activity has already past.

Figure 4.5 gives the diurnal and seasonal variation combined, including terminator times. The motivation for including the terminator times (i.e. sunrise and sunset time) is that the solar rays strongly affect the level of ionization in the ionosphere. Indeed most whistlers are received when Tihany and its conjugate point is in darkness. However, Figure 4.5 is not detailed enough to distinguish whether the receiving station, the conjugate point, or both should be in darkness to enhance whistler transmission. A more detailed investigation of whistler activity coupled to daylight would be desired here. A two dimensional representation of Figure 4.5 is presented in the first part of Appendix B.

The overall peak of whistler detection is around midnight in February and March. The minimum can be located to the morning hours of August, September, and October as seen in Figure 4.5.

![Figure 4.5: Diurnal distribution of whistlers divided into months at Tihany, Hungary. 2002-02-27 through 2005-05-18. Terminator times given in black and blue lines and at an altitude of 100 km](image)

4.2 Dunedin, New Zealand

This data base consists of 236019 registrations recorded between May 20, 2005 and October 30, 2006. It is hard to draw any conclusions on the distribution of whistlers on a yearly basis considering only two years are present in the data, and none of them are complete. 2005 accounted for just above 140000 registrations and 2006 counted just above 90000 registrations which mean that roughly 50% more data was collected during 2005 compared to 2006. This is
4.2. Dunedin, New Zealand

a bit odd since the peak lightning time in the conjugate point is during the northern hemisphere summer which is included in both years. Furthermore, the station was operating during a larger fraction, 10 months, during 2006 compared to only 6 months 2005. The most reasonable conclusion must thus be that the northern hemisphere summer of 2005 hosted more thunderstorms that generated detectable whistlers than the summer of 2006 in the source region for this station.

The total amount of received whistlers is displayed in Figure 4.6. This graph is very indistinct in that no major seasonal dependence can be observed. Even though a seasonal dependence is clear for 2006, the activity during the northern autumn of 2005 is relatively high and similar to that of the northern summer of 2005 and 2006.

![Figure 4.6: Total amount of registered whistlers at Dunedin, New Zealand. 2005-05-20 through 2006-10-30. Each bin has a width of one day](image)

When analyzing Figure 4.7 the seasonal dependence becomes clearer. The relatively large jump of received whistlers from May to June can partly be explained by the fact that two summers and only one spring are part of the data. This will bias the output. The high season for receiving whistlers can never be set to June through October roughly. Even if we doubled the number of whistlers for the spring months they would not even be close to the numbers received during the summer. The accumulated account for the summer months are roughly 30000 to 40000 whistlers per month and during the winter this number is only around 5000 whistlers per month.

The diurnal variation, as shown in Figure 4.8, has its peak in the early afternoon with approximately 25000 accumulated recordings around 15 local time. The minimum occur around 9 local time. This pattern is significantly
Figure 4.7: Seasonal distribution of whistlers at Dunedin, New Zealand. 2005-05-20 through 2006-10-30. Data is not normalized, thus is June through October presented for two years whereas November through April is presented for one year. May is presented for two months minus 11 days different to the pattern observed at Tihany. This might indicate some difference concerning lightning activity and/or whistler generation and transmission.

Figure 4.9 gives the combined diurnal and seasonal variation of whistlers at Dunedin, New Zealand. A two dimensional representation of Figure 4.5 is presented in the second part of Appendix B.
4.2. Dunedin, New Zealand

Figure 4.8: Diurnal distribution of whistlers at Dunedin, New Zealand. 2005-05-20 through 2006-10-30

Figure 4.9: Diurnal distribution of whistlers divided into months at Dunedin, New Zealand. 2005-05-20 through 2006-10-30
5 Correlation

With the results of the two previous chapters concerning whistler and lightning statistics it is of high interest to investigate whether or not any kind of relation between these features can be established. Since the whistler and WWLLN data consists of large arrays of observations, a good way to obtain this possible connection is to calculate the linear correlation coefficient $r$ between these data sets.

5.1 Correlation Definition

The linear correlation coefficient $r$ is mathematically defined as

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

(5.1)

where $x_i$ and $y_i$ represent the two series to be correlated with each other respectively. $\bar{x}$ is defined as

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

(5.2)

and the same for $\bar{y}$ respectively. It is also worthy to note that $n$ must be the same for both $x_i$ and $y_i$, i.e. the length of the series must be the same. The denominator in Equation 5.1 normalizes the output and $r$ will thus always obtain a value between $-1$ and $+1$. The latter corresponds to a perfect positive correlation. This implies that when the data in $x$ is “moving” in any direction, the data in $y$ will “move” in the exact same way. This would correspond to that if, for example, the rate of lightning strikes are increasing in a region so would the received whistler rate increase in precisely the same way. An $r = 0$ would correspond to no correlation at all. That is, if the lightning rate rose the whistler rate would remain completely independent of that. Negative $r$ corresponds to the two data sets moving in opposite directions. This means that if the number of lightning strikes are picking up in the region then, consequently, the number of received whistlers are decreasing. In this project the result of the correlation is interesting from both an absolute value and also on a relative basis. That is, we are also concerned for the relative difference in correlation between selected areas of the world.
5.2 Correlation Methodology

The method used when obtaining the correlation results was as follows. As described in section 2.2 the input raw data consists of large arrays of time [UTC] for the whistlers and similar arrays containing longitude, latitude, and time [UTC] for lightning.

First the time from the first day with a received whistler to the last day is sorted into bins or time intervals. Typically such an interval is in the order of 30 seconds. Since the first day was on Feb 27, 2002 and the last on May 18, 2005 for Tihany, this result in a vector containing just above 3.3 million time intervals. In this particular report time intervals of 15 seconds and 30 seconds were used. The whistler counts are then placed in the respective interval that suits them on behalf of their time of registration creating very long histogram. The data file from Tihany, Hungary, consist of roughly 681000 registrations and since those whistlers are divided into roughly 3.3 million intervals it means that a whistler on average is recorded in every 5th interval when 30 second intervals are used.

A similar approach is taken with the lightning data. However, first the spatial dependence in this data is removed by sorting the strikes into cells, or frames, depending on where on the Earth the lightning occurred. A typical cell size is having an area of approximately 3° squared down to 1° squared. In this report, cell sizes of 3°, 1°, and 0.5° were used. Then, for each particular cell, the lightning is divided into the same time intervals as the whistler data was. Now we have two time series that could be correlated using equation 5.1 above. We then obtain the correlation coefficient for that particular cell on the globe. This calculation is then repeated for all cells covering the globe and we are left with a map displaying the various correlation coefficients for all cells covering the Earth.

However, due to the low efficiency of the WWLLN in terms of amount of registered lightning (see Chapter 3), the data in the time intervals was normed to either one or zero. That is, if there was activity in an interval, regardless of how much, the number of registered events was set to one. If there was no activity that interval was set to zero. It makes no sense today to perform correlation with the absolute numbers of whistler and lightning registrations in every time interval toward each other. This approach implies that our results will display correlation based on similarities in activity, but not on the base of how much activity. This approach, at least partly, circumvents the WWLLN inefficiency.

When looking at the individual time intervals this simplification means that only four possibilities of matching between the two series can occur. If there are both whistler and lightning registrations for a specific time interval this interval will contribute toward positive correlation. Further, if there are neither whistler nor any lightning registration for a time interval this will also contribute toward positive correlation. It is important to remember that inac-
tivity of lightning in a source region gives less activity of registered whistlers; this must be taken into account. If there is registered lightning but no whistlers or vice versa for respective time interval this will contribute toward negative correlation.

The size of the time intervals is based upon a number of factors. The exact time gap from when lightning occurs until the generated whistler wave is being detected can vary. The time spent in the magnetosphere is ~ 1 seconds. The size of the time interval must thus be large enough to minimize the risk of a whistler being detected in a different interval to its lightning discharge. However, the interval must also be small enough to avoid correlation between lightning strikes and whistlers generated by other lightning strikes. This is a tricky balance and a compromise must be made. As of now the problem is partly resolved by the norming of the data, but with increased WWLLN efficiency this norming could be removed and then this issue becomes critical.

The selection of cell size is based upon the fact that a grid selection that is too coarse won’t produce good enough resolution in the resulting correlation map. However, selecting too high a resolution will result in a lower correlation value due to the decreased number of lightning events in each cell, while also increasing the risk of lightning ending up in the wrong grid cell. The lower correlation values that follow from a decreased cell size might threaten the statistical significance. A compromise is also desirable here. The values described above were selected to fulfill these criterions.

5.3 Settings and Abbreviations

Based on the discussion above, a number of parameters have to be introduced and taken into consideration when analyzing the correlation results. The Frame Size (FS) is defined as the length in degrees of the sides of the cells serving as spatial dividers for lightning strikes. The time series for each cell is then compared with the whistler time series. Time Window (TW) represents the length of the time interval in seconds, selected when sorting the events depending on the time of occurrence. Time Slice (TS) represent a fractional part of all days, in this report one or three hours, which the correlation is being made for. This is to determine if there are any specific diurnal dependencies of the correlation. An example here would be that a TS of 0-1 extracts those time intervals that are situated between just midnight and 1 am for every day in the total interval and then the correlation is calculated for only this limited time series.

Another feature to be aware of when analyzing the results below is the L-value which is plotted on all graphs. The definition of L-value is the height in relative Earth radii that a magnetic field line has when it crosses the magnetic equator. For example, the magnetic field line that takes a relatively narrow path and just scratches the surface of the Earth at the equator has an L-value
equal to one. The closer to the poles that the field line is emerging the higher its L-value becomes due to its higher relative altitude when it crosses the magnetic equator. One must also remember that the closer to the poles that a field line emerges the higher its inclination, i.e. angle compared to the surface, becomes. This has an effect in the ability of a field line to catch a whistler. The field lines from the Earth’s magnetic field passing through Tihany, Hungary, have an L-value of 1.81. The corresponding value for Dunedin, New Zealand, is 2.75. Both values are for 2002.

5.4 Correlation Results for Tihany, Hungary

A number of figures that contain different settings are presented here. It is motivated to present all figures since no figure alone can give a total overlook and understanding as all figures in combination can do. This applies mainly to Figures 5.1 through 5.4 which all present the results for a TS of 0-24 UTC, i.e. without considering diurnal differences.

Figure 5.1: Correlation results for Tihany, Hungary. FS=3° TW=30s TS=0–24 UTC

Figure 5.1 presents a relative coarse grid with a Frame Size of 3°. The correlation clearly has its maximum around the magnetic conjugate point. As can be expected, a weaker positive correlation is also obtained from the rest of southern Africa. Negative correlation is obtained north of the equator in Africa, which is expected due to the seasonal difference in lightning compared to southern Africa. Most whistlers are received during the southern hemisphere summer. This is when lightning activity is minimized north of the equator thus yielding a negative correlation on a seasonal basis. In Asia the
negative correlation north of the equator is even greater. This is most likely due to the increased efficiency of WWLLN in this area. The most surprising result regards the relative high positive correlation obtained over equatorial South America. The most likely explanation is that VLF waves originating from these lightning discharges can travel for a long distance in the EIWG before entering a magnetic duct around the conjugate point.

Figure 5.2: Correlation results for Tihany, Hungary. FS=1° TW=30s TS=0–24 UTC

Figure 5.2 strengthens the initial view that Figure 5.1 provided. The resolution is higher though and a pattern of correlation is starting to emerge south of the equator in South East Asia. Positive correlation is obtained over the ocean and negative correlation is obtained over land. An attempt to explain this behavior would be very speculative at this point in time. A further investigation would be interesting. Another pattern which was also visible in figure 5.1 is the fact that there seems to be a “tail” attached to the source region surrounding the conjugate point, pointing toward the southeast. A likely explanation is that this tail is actually parallel to the horizontal direction of the magnetic field. This suggests that the energy from a lightning strike has a higher probability of exciting a duct when aligned in this way. The lack of such a tail pointing northwest from the source region might suggest that the inclination of the magnetic field also plays a role. That is, it seems to be easier for a wave to excite a duct if it approaches the footprint from the direction of the magnetic pole.

Figure 5.3 gives an even more detailed view in terms of spatial resolution. The overall image provided by Figure 5.1 and 5.2 is maintained. However, the problem with higher resolution becomes more evident in terms of smaller absolute values of the correlation coefficient. This can reduce the statistical
security of the result.

Figure 5.3: Correlation results for Tihany, Hungary. FS=0.5° TW=30s TS=0 – 24 UTC

Figure 5.4 is provided to give an indication of what happens with the correlation with a Time Window of 15 seconds. The overall picture is maintained. However, the inefficiency in the WWLLN and the combination of small Frame Sizes and a small Time Window results in very small absolute values in the correlation coefficients. Table 5.1 gives a representation of the maximum and minimum correlation values obtained in Figure 5.1 through 5.4. When analyzing the minimum correlation values in Table 5.1 the absolute value is clearly approaching zero with decreased Frame Size and decreased Time Window. On the positive correlation side that tendency is far from clear. The mean value of $r$ and the locations of the obtained max and min values were not obtained.

Figure 5.4: Correlation results for Tihany, Hungary, with TW=15s. Left plot has FS=1.0° and right plot has FS=0.5°. TS=0 – 24 UTC
5.5. Correlation Results for Dunedin, New Zealand

Table 5.1: Maximum and minimum correlation coefficients obtained for different figures concerning Tihany, Hungary. TS=0 – 24 UTC for all figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Min-correlation</th>
<th>Max-correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 5.1</td>
<td>-0.027729</td>
<td>0.049027</td>
</tr>
<tr>
<td>Figure 5.2</td>
<td>-0.014914</td>
<td>0.038783</td>
</tr>
<tr>
<td>Figure 5.3</td>
<td>-0.009518</td>
<td>0.031640</td>
</tr>
<tr>
<td>Figure 5.4 left</td>
<td>-0.009299</td>
<td>0.041082</td>
</tr>
<tr>
<td>Figure 5.4 right</td>
<td>-0.005980</td>
<td>0.035647</td>
</tr>
</tbody>
</table>

One conclusion that can be drawn mainly from Figure 5.1, 5.2, and 5.3 is that the source region surrounding the magnetic conjugate point for Tihany has a radius of approximately up to 1000 km or just slightly less.

Tihany Diurnal Variation

Figure 5.5 is provided to give an indication of what happens with the correlation coefficient on a diurnal basis, presenting correlation calculated for three hour intervals. When analyzing the graphs it is evident that there is a maximum in the correlation during the hours when Tihany and its conjugate point are in darkness. The positive correlation for the magnetic conjugate point starts to pick up between 15 and 18 UTC, reaching a maximum around 21 to 24 and then decreases in the early morning. During day time the correlation is almost nonexistent. This result also relates back to the issues discussed in previous chapters on WWLLN and whistler statistics for Tihany, Hungary. Figure 4.5 clearly shows that most whistlers are indeed received in darkness. Since there is a strong correlation with the conjugate region during the same time it seems reasonable to draw the conclusion that most of these whistlers indeed originate from here.

The positive correlation from the equatorial part of South America follows a similar diurnal pattern. This supports the theory that this correlation is obtained after the VLF signatures from these lightning strikes have traveled in the EIWG and are captured by the magnetic field around the conjugate point. If they were caught by the magnetic field at some other place on Earth and the whister then traveled a long distance after re-entering the atmosphere a different time pattern would be expected on this correlation due to the time difference in the emerging and disappearing of the ionospheric layers. However, further investigation on this field would be desirable.

5.5 Correlation Results for Dunedin, New Zealand

When analyzing the results from Dunedin, New Zealand, the correlation is not so clear. Figure 5.6 gives the overall correlation with a Frame Size of $1^\circ$
Figure 5.5: Correlation results for Tihany, Hungary, with TS=3 hour intervals. FS=1° TW=30s
and a Time Window of 30s. A trace of positive correlation can be seen in the pacific northwest of the Hawaiian Islands. The magnetic conjugate point for Dunedin is situated just south of the Aleutian Islands to the north at 55.85°N, 195.34°E. One reason a clear correlation is not obtained here could be due to the relatively small amount of lightning that occurs so far north. Furthermore, the lightning that does occur might also be generally weaker, yielding that it is harder for WWLLN to detect these strikes.

![Correlation Results for Dunedin, New Zealand](image)

**Figure 5.6:** Correlation results for Dunedin, New Zealand. FS=1° TW=30s TS=0 – 24 UTC

**Dunedin Diurnal Variation**

However, when investigating the diurnal difference the correlation pattern becomes somewhat clearer. Figure 5.7 represents a split of the day in three hour intervals. Since the diurnal differences seem to be more significant for Dunedin compared to Tihany in terms of interpreting the correlation, a one hour split of the data from Dunedin, New Zealand, is presented in Appendix C. There seems to be a maximum correlation for the source region from around 15 to 21 UTC. For the rest of the day small traces can be seen but they seem to be somewhat sporadic. "Islands" of positive correlation can also be observed around the globe, especially in the East Asian region but also in the Caribbean. The positive correlation for these two regions seems to have some sort of maximum between 9 through 21 UTC. One must remember that the local time in Dunedin has a 12 hour shift compared to UTC. The traces of correlation obtained from Asia and the Caribbean region could therefore be explained in a similar way as for the equatorial part of South America in the Tihany case. This correlation takes place when Dunedin and its magnetic
conjugate point are mostly in darkness, the ionosphere might thus be more transparent for these whistlers. However, why the correlation for the conjugate point at the same time is low is difficult to explain. It must be stated though that these results are very unsecure. Figure 5.7 presents very sporadic and irrational behavior and no real conclusions can be drawn from this data. Further work with a more thorough investigation of lightning activity in the magnetic conjugate point and possibly more whistler recordings is desirable.

Figure 5.7: Correlation results for Dunedin, New Zealand, with TS=3 hour intervals. FS=1 ° TW=30s

It is of interest that a large part of the cells containing negative, as well as positive, correlation in Figure 5.7 are confined within the lines defining L=2.75. This is pretty clear when observing the North American part of the map for example. If this is simply due to the low lightning activity north of
this region or if there is another explanation remains unanswered. However, lightning activity so far north is indeed pretty sparse all around the globe.

5.6 Discussion and Validity of Correlation Approach

There are a number of issues that need attention in order to justify the results. The first issue concerns the effectiveness of the receiver stations. The WWLLN is assumed to more or less function at all times since there are numerous stations in this network. If one station ceases to function for a limited time there are other stations that will continue recording, thus maintaining the overall functionality. It is a different situation for the whistler stations. Days when the station is down for some reason must be identified and all data removed for these days in order to avoid false correlation. This procedure was implemented on this project. The approach on how to identify these days was simply to check for days when no whistlers were recorded. However, it can be risky to just state that the station is broken if there were no whistlers received. Therefore a log file monitoring the functionality of the station in Tihany was acquired. This log file was then compared to the days when no whistlers were received. The result of this comparison is presented in Figure 5.8.

![Figure 5.8: Days to include or exclude in the correlation analysis for Tihany, Hungary. Blue days correspond to days with received whistlers and the log file saying the station is functioning. Green days correspond to days without whistler registrations, hence these days are excluded, and the log file says the station is down. Red days correspond to excluded days but the log file says the station is functioning. Blue and green days make up just above 80% of the total days. Black days don’t exist or no data was available.](image)

As can be seen there is a very good agreement between what the log file
reveals and the empirical method. The blue days correspond to days with received whistlers and the log file indicating that the station is functioning. The green days correspond to no received whistlers and the log file reporting that the station is not functioning. The red days, however, display days when there were no whistlers in the data base but the log file says that the station is functioning. Yellow days would report that whistlers were received and we would include those days in the analysis but the log file indicates that the station was down. Luckily there were no such instances of days registered.

However, an analysis of the red days is needed. This log file is far from perfect in itself. The detector system basically consists of an antenna, some amplifiers and wires, and a computer. The incoming whistlers are registered on an initial memory and every hour these registrations are transferred to the hard drive. This second transfer is what is contained in the log file. That means that the log file can report the station as functional, since the data transfer system works, but in reality the physical antenna might be down for some reason, or some other component. That means that the log data is not 100% accurate. However, it is the best that can be accomplished at this point in time.

The agreement between the log file and the empirical method is good. The green and blue days in Figure 5.8 account for just above 80% of the total days. This must be considered a very good match. Taking into account that the average number of received whistlers is 577 per day in Tihany, for our data set, it is very unlikely that there would be days without any whistlers. Even in the winter months when the average rate drops down to around 100 whistlers per day this is still pretty high. Therefore, those days totally lacking registrations are excluded from the analysis.

A second interesting section to discuss in terms of validity of the empirical method concerns the spatial resolution of the correlation. Even though WWLLN has a low efficiency of the amount of lightning events being registered it has a relatively high resolution in terms of time and coordinates of the strikes that do get registered as discussed in Chapter 3. Therefore the decision was taken to run the correlation for cells with a smaller Frame Size than was considered initially, thus improving the resolution of the result. As discussed above, going to smaller cells can be risky since the risk increases that lightning strikes might end up in the wrong cell. There is indeed a lower limit in spatial resolution. However, this limit can be pushed quite significantly if the efficiency of WWLLN is improved. However, if there is a need for such a push, that is a different matter. The results during this study, with a source region having an approximate 1000 km radius, suggest that higher spatial resolution than achieved here might not be necessary, at least not for the purpose of correlating lightning and whistlers on a statistical basis.

A third issue is in regards to the resolution in selecting the Time Window. With increased WWLLN efficiency the normalization could maybe be removed, as discussed earlier. The problem of selecting the right Time Win-
5.6. Discussion and Validity of Correlation Approach

dow can then become critical. Different whistlers need different time to travel to the receiver. This is mainly due to which path they take, i.e. which L-value the magnetic field line that they follow has. A solution to this might be to introduce a time lag, i.e. a "delay" in the whistler series compared to the WWLLN series in Equation 5.1. If the Time Window is too short this time lag could be a way to address that problem.

Fourth, one important issue that has not been considered in this report is the statistical significance of the correlation results. However, this was not part of this project but work on this field would bring about a more stable foundation for the results. Such analysis is currently under development.
6 Conclusions

6.1 Conclusion of the Project

In this report observed whistler series from Tihany, Hungary, and Dunedin, New Zealand, have been correlated with data from the WWLLN. The data from Tihany generates a relatively clear correlation image which shows that the source region, surrounding the conjugate point, for whistlers observed at Tihany has a radius of up to approximately 1000 km. Somewhat weaker positive correlation is also obtained from the rest of Southern Africa and equatorial South America. For Dunedin the correlation was more sporadic but a diurnal variance was observed. The whistler and WWLLN statistics obtained do not contradict with the achieved correlation results. The results achieved do however conform very well to the general theory of how whistlers are generated and propagate.

The results of this report cannot be taken as evidence that all whistlers observed at a whistler station were picked up around the magnetic conjugate point. However, it indicates rather strongly that at least a major portion originate from that region.

6.2 Future Research Suggestions

There are several interesting aspects in this field that still have not been examined. The work on correlation and detailed lightning and whistler statistics has just but begun.

An interesting approach that was initiated but not completed was to generate a code extracting a “movie” displaying an increase in lightning activity in the source region on an individual lightning strike basis. This code would then also generate a similar movie for the whistler registrations during the same time period. Then one can examine if an increase in the lightning rate generates a pickup in whistler rates. Identification of individual lightning/whistler pairs might then be possible. A draft of this not completed movie-code is presented in Appendix D.

A suggestion related to the above mentioned movie would be to investigate the possibility of taking the latitude where the whistlers was observed into
A Appendix, Lightning Statistics

This appendix includes plots of lightning statistics from Tihany’s magnetic conjugate point. The plots are based on months and hours of registration. This gives a good overview of both seasonal and diurnal differences in lightning detection in this region. The plots here present the same material as Figure 3.5.
Figure A.1: Diurnal distribution of lightning, based on hours, for January (left) and February (right) at Tihany, Hungary, magnetic conjugate point. 2003-01-11 through 2006-04-23. Selected circular area surrounds $\text{Lat}= -33.45^\circ$, $\text{Long}=28.34^\circ$ with $R=500\text{km}$.

Figure A.2: Diurnal distribution of lightning, based on hours, for Mars (left) and April (right) at Tihany, Hungary, magnetic conjugate point. 2003-01-11 through 2006-04-23. Selected circular area surrounds $\text{Lat}= -33.45^\circ$, $\text{Long}=28.34^\circ$ with $R=500\text{km}$.

Figure A.3: Diurnal distribution of lightning, based on hours, for May (left) and June (right) at Tihany, Hungary, magnetic conjugate point. 2003-01-11 through 2006-04-23. Selected circular area surrounds $\text{Lat}= -33.45^\circ$, $\text{Long}=28.34^\circ$ with $R=500\text{km}$.
Figure A.4: Diurnal distribution of lightning, based on hours, for July (left) and August (right) at Tihany, Hungary, magnetic conjugate point. 2003-01-11 through 2006-04-23. Selected circular area surrounds Lat=-33.45° Long=28.34° with R=500km

Figure A.5: Diurnal distribution of lightning, based on hours, for September (left) and October (right) at Tihany, Hungary, magnetic conjugate point. 2003-01-11 through 2006-04-23. Selected circular area surrounds Lat=-33.45° Long=28.34° with R=500km

Figure A.6: Diurnal distribution of lightning, based on hours, for November (left) and December (right) at Tihany, Hungary, magnetic conjugate point. 2003-01-11 through 2006-04-23. Selected circular area surrounds Lat=-33.45° Long=28.34° with R=500km
B  Appendix, Whistler Statistics

This appendix includes plots of whistler statistics based on months and hours of registration. This gives a good overview of both seasonal and diurnal differences of whistler detection rates in Tihany, Hungary, and Dunedin, New Zealand. The plots here present the same material as Figure 4.5 and Figure 4.9.
Figure B.1: Diurnal distribution of whistlers, based on hours, for January (left) and February (right) at Tihany, Hungary. 2002-02-27 through 2005-05-18

Figure B.2: Diurnal distribution of whistlers, based on hours, for Mars (left) and April (right) at Tihany, Hungary. 2002-02-27 through 2005-05-18

Figure B.3: Diurnal distribution of whistlers, based on hours, for May (left) and June (right) at Tihany, Hungary. 2002-02-27 through 2005-05-18
Figure B.4: Diurnal distribution of whistlers, based on hours, for July (left) and August (right) at Tihany, Hungary. 2002-02-27 through 2005-05-18

Figure B.5: Diurnal distribution of whistlers, based on hours, for September (left) and October (right) at Tihany, Hungary. 2002-02-27 through 2005-05-18

Figure B.6: Diurnal distribution of whistlers, based on hours, for November (left) and December (right) at Tihany, Hungary. 2002-02-27 through 2005-05-18
Appendix, Whistler Statistics

Figure B.10: Diurnal distribution of whistlers, based on hours, for July (left) and August (right) at Dunedin, New Zealand. 2005-05-20 through 2006-10-30

Figure B.11: Diurnal distribution of whistlers, based on hours, for September (left) and October (right) at Dunedin, New Zealand. 2005-05-20 through 2006-10-30

Figure B.12: Diurnal distribution of whistlers, based on hours, for November (left) and December (right) at Dunedin, New Zealand. 2005-05-20 through 2006-10-30
Appendix, Correlation

Additional correlation results based on TS intervals of one hour for Dunedin, New Zealand.
Figure C.1: Correlation results for Dunedin, New Zealand, with TS=1 hour intervals. FS=1° TW=30s
Figure C.2: Correlation results for Dunedin, New Zealand, with TS=1 hour intervals. FS=1° TW=30s
D Appendix, Source Code

This appendix contains all source code derived in R and GMT. The source code is divided into seven different R programs plus one GMT program. In order these are:

- Whistlercode.R (whistler statistics)
- Wwllncode.R (WWLLN statistics)
- Correlationcode.R (correlation calculation)
- CorrelationcodeTS.R (correlationTS derivation)
- Corrmodelcontrol.R (derivation of days to include or exclude in correlation analysis)
- Wlidentificationcode.R (relating whistlers with corresponding lightning strikes)
- Correlationcode.gmt (code to derive the graphical output from .xyz file produced by correlation- and correlationTS-code in R)

Some codes have received a slight modification in layout to fit in this report.
```r
#!/ Whistlercode

#-Variables that needs to be given by user prior to running code
wfile = sprintf("/home/jonas/work/correlation/data/Tihany.txt") #which file should be scanned
station = sprintf("Tihany, Hungary") #give the name of the whistler station (include country)
sn = sprintf("TTI") #give short name of station, i.e. the first letters in name and country
timediff = 1 #give time difference in hours from GMT for station
timeteststart = ISOdate(2004,03,19,15,00,00, tz="GMT") #selecting start time for test
timeteststop = ISOdate(2004,03,19,18,00,00, tz="GMT") #selecting stop time for test
timeinterval = 30 #choose time interval [s] between frames in specific wreg plot
daytimewindow = 1 #[days] select accuracy when plotting complete distribution of whistlers
movieinterval = 30 #[minutes] select range of one movie frame (suggested is 30 min)
output = sprintf("/home/jonas/work/whistlers/output/TTI/",) #select output directory for output files

#!/ Whistlerdata that need to be scanned in
while = scan(wfile)

#reading in and creating dataframes for whistlers
register_time = function(a)
{
  tlist = list()
  tlist$year = as.integer(substr(a, 1, 4))
  tlist$month = as.integer(substr(a, 5, 6))
  tlist$day = as.integer(substr(a, 7, 9))
  tlist$hour = as.integer(substr(a, 9, 10))
  tlist.minute = as.integer(substr(a, 11, 13))
  tlist$second = as.double(substring(a, 13, 16))

  #--checking diurnal differences
  tlist$hour = tlist$hour + tlist$minute/60 + tlist$second/3600
  tlist$min = tlist$hour*60
  tlist$sec = tlist$min*60

  #--checking general dates
  tlist$date = as.Date(ISOdate(tlist$year, tlist$month, tlist$day)) #full date, ex: 2002-01-01
  tlist$idate = as.integer(tlist$date) #integer days since epoch (Jan 1 1970)
  tlist$day = tlist$date + tlist$hour/24 #days with decimal hours since epoch

  #--checking seasonal differences (not in use)
  tlist$month = tlist$month + tlist$day/30 #decimal months
  tlist$nrmonth = as.integer(format(tlist$date, "%m")) #month number from jan 1st that year
  tlist$nrday = as.integer(format(tlist$date, "%d")) #number of day during the year

  tlist
}

wdata = as.data.frame(register.time(wfile))

#!/Creating factors in to which you can sort data
typer = as.integer(levels(factor(wdata$year))) #a list with years represented
fmonth = factor(wdata$month)
#sorting data on a specific factor
wdata$hour = split(wdata$hour, fmonth) #hours compared to month, i.e all hours during specific month

#!/ Adjusting local time at station
clockmarks = c(0,2,4,6,8,10,12,14,16,18,20,22,24) #clockmarks in GMT
localclocks = clockmarks + timediff #unadjusted local clockmarks

#!/ Adjusting local clockmarks
for(i in 1:13)
{
  if(clockmarks[i]>=24)
  {
  
```
# Variables needed in several plots

```r
period = sprintf("%s to %s", wdata$date[1], wdata$date[length(wdata$date)]) # showing period of measurements, used in plot1-6
ylab = sprintf("Nr of observed whistlers") # ylab text used in plot1-5,7,8
```
\begin{verbatim}
60
Appendix, Source Code

at = c(2.5,4.5,6.5,8.5,10.5,12.5)
postscript(filename) #recording to .pdf
hist = hist(wdata$month, breaks=1:13, col="blue", right=FALSE, main=main,
          xlab="", ylab=ylab, sub=sub, axes=FALSE)
axis(side=1, labels=labels, at=at, tick=FALSE, line=-1)
axis(side=2)
dev.off() #recording off

# Plot4 Creating Histogram over total amount of observed whistlers based on
days
filename = sprintf("%s%04dwhours.eps", output, sn)
main = sprintf("Whistlers divided into days of registration (\%s \%s (%s),
               station, period)
sub = sprintf("LT=Local Time, UTC=Universal Time Coordinated \%s\%sروم
        consideration taken for days when receiver did not operate")
postscript(filename) #recording to .pdf
hist = hist(wdata$day, breaks=1:32, col="blue", right=FALSE, main=main,
          xlab="Day number in month", ylab=ylab, sub=sub, axes=FALSE)
axis(side=1, labels=labels, at=at)
axis(side=2)
dev.off() #recording off

# Plot5 Creating Histogram over total amount of observed whistlers based on
hours
filename = sprintf("%s%05whours.eps", output, sn)
main = sprintf("Whistlers divided into hours of registration (\%s \%s (%s),
               station, period)
sub = sprintf("Local Time of Headline")
postscript(filename) #recording to .pdf
hist = hist(wdata$hour, breaks=0:24, col="blue", right=FALSE, main=main,
          xlab="", ylab=ylab, sub=sub, axes=FALSE)
axis(side=1, labels=clockmarks, at=clockmarks, line=0)
mtext(text="LT", side=1, line=1, at=25.5)
axis(side=1, labels=clockmarks, at=clockmarks, line=1, tick=FALSE)
mtext(text="UTC", side=1, line=2, at=25.5)
axis(side=2)
dev.off() #recording off

# Plot6-17 Plotting histogram with diurnal differences during all months for
the year during the whole period
z = matrix(ncol=24, nrow=12) #defining matrix
x = 1:24
r = 6
for(y in 1:12)
  mname = month.abb[y] #selecting the correct name of month
  filename = sprintf("%s%dwhours%05s.%s", output, sn, nr, mname) #giving name
to "filename".pdf
  main = sprintf("Diurnal Dist of Whistlers in (\%s \%s (%s),
                 station, period)")
  sub = sprintf("LT=Local Time (%s), UTC=Universal Time Coordinated \%s\%sروم
            consideration taken for days when receiver did not operate")
  postscript(filename) #recording to .pdf
  hist = hist(wdata$hour[y], breaks=0:24, right=FALSE, main=main,
             xlab="", ylab=ylab, sub=sub, axes=FALSE)
  axis(side=1, labels=clockmarks, at=clockmarks, line=0)
mtext(text="LT", side=1, line=1, at=25.5)
  axis(side=1, labels=clockmarks, at=clockmarks, line=1, tick=FALSE)
mtext(text="UTC", side=1, line=2, at=25.5)
  axis(side=2)
dev.off() #recording off

z[y,x] = hist(\[1\]) #extracting the counts for every round
nr = nr+1

# Plot18 3D-image containing above diurnal distr plots
filename = sprintf("%s%03Dimage.%s", output, sn)
main = sprintf("Diurnal Dist of Whistlers Divided into Months (\%s \%s (%s),
               station, period)
sub = sprintf("LT=Local Time (%s), UTC=Universal Time Coordinated \%s\%sروم\%s\%sروم
            consideration taken for days when receiver did not operate")
if(sn="TH")
{

\end{verbatim}
Appendix, Source Code

sub = sprintf("LT=Local Time (%s), UTC=Universal Time Coordinated \nNo consideration taken for days when receiver did not operate \nSunrise and sunset times at an altitude of 100km", station)

breaks = c(1.5,10,50,100,500,1000,5000,10000,50000,100000)

text = paste(breaks[1:10], breaks[2:11], sep="-",) #text showing values of bar colors

labels = c("Feb", "Apr", "Jun", "Aug", "Oct", "Dec")

at = c(2.5,4.5,6.5,8.5,10.5,12.5)

postscript(filename) #recording to .pdf

par(mar=c(5,4,4,7.5), for pdf:mar=c(4.15,3.2,3.2,6.8), ps:mar=c(3.7,3.4,3.6,6.95))

image(z=z, x=1:13, y=1:24, col=rev(heat.colors(10)), breaks=breaks, main="main",

xlab="", ylab="", sub="sub", axes=FALSE)

axis(side=2, labels=clockmarks, at=clockmarks, line=0)

mtext(text="LT", side=2, line=-2)

par(mar=c(3.7,3.4,3.6,6.95), new=TRUE)

plot(daynr, THsunrise, type='l', axcs=FALSE, xlab="", ylab="", xlim=c(0,365), ylim=c(0,24))
text(x=180, y=1.1, labels="Tihany Sunrise")

par(mar=c(3.7,3.4,3.6,6.95), new=TRUE)

plot(daynr, THsunset, type='l', axcs=FALSE, xlab="", ylab="", xlim=c(0,365), ylim=c(0,24))
text(x=180, y=20.5, labels="Conjugate Sunset")

par(mar=c(3.7,3.4,3.6,6.95), new=TRUE)

plot(daynr, DURsunrise, type='l', axcs=FALSE, xlab="", ylab="", xlim=c(0,365), ylim=c(0,24))
text(x=180, y=15.3, labels="Conjugate Sunrise", col="blue")

dev.off()
movieframe = seq(from=timeteststart, to=timeteststop, by=movieinterval*60) # limits for longitude in l-data, size of square

wtime = cut(wtimetest, breaks=movieframe) # factor for ltimetest, i.e. sorting which movieframe they fall in

movieframe = seq(from=timeteststart, to=timeteststop, by=movieinterval*60) # limits for longitude in l-data, size of square

wtime = cut(wtimetest, breaks=movieframe) # factor for ltimetest, i.e. sorting which movieframe they fall in

movieframe = seq(from=timeteststart, to=timeteststop, by=movieinterval*60) # limits for longitude in l-data, size of square

# Plot 20 movie with whistler registrations
filename = sprintf("%s20wmoviehist.eps", output, sn)

main = sprintf("Observed whistlers at %s", station)
timemovie = timeteststart # setting start time for movie

postscript(filename = "Observed whistlers at %s", station)

for(i in 1:length(wtime)){
  sub = sprintf("Timewindow for each frame is %d seconds \n Total amount of
whistlers are %d, time in UTC", timeinterval, length(wtime[[i]]))
  breaks = seq(from=as.integer(timemovie), to=as.integer(timemovie+movieinterval*60), by=movieinterval*60) # dividing lightningstrikes into frames

  labels = c(as.character(timemovie), as.character(timemovie+movieinterval*60)) # selecting which labels to show
  at = c(as.integer(timemovie), as.integer(timemovie+movieinterval*60)) # setting tickmarks at selected labels
  timemovie = timemovie + movieinterval*60 # setting interval for next frame

  hist = hist(wtime[[i]], breaks=breaks, col="blue", right=TRUE, main=main, xlab="", ylab=ylab, sub=sub, axes=FALSE)
  axis(side=1, labels=labels, at=at)
  axis(side=2)
}

dev.off() # recording off

#
Appendix, Source Code

1 // Lightening Code
2 #/ Variables that need to be given by user prior to running code
3 longref = 28.34 #degrees] Selecting lightning longitude to compare with
4 latref = -33.45 #degrees] Selecting lightning latitude to compare with
5 reference = 200 #km] Selecting max distance from refpoints allowed
6 timeteststart = ISOdate(2004,03,19,15,00,00, tz="GMT"] #selecting starttime
7 timeteststop = ISOdate(2004,03,19,18,00,00, tz="GMT"] #selecting stoptime for test
8 timeintervall = 30 #s] chose timeintervall between frames in specific regulare
9 movieintervall = 30 #minutes] select range of one movie frame (suggested is 30 min)
10 output = sprintf("/home/ionas/work/wwlln/output/TH/R=200/"") #select output
directory for output files
11 #// Calculating which lightningfile to read in
12 for(i in 1:12) {checking latitudes
13 for(j in 1:24) {checking longitudes
14 latselect = -90+15*(i-1)+7.5 #midpoint in each square
15 longselect = -180+15*(j-1)+7.52 distance from midpoin to selected refpoint
16 latdiff = latselect - latref
17 longdiff = longselect - longref
18 if(londiff <=7.5 & latdiff <=7.5) #checking if refpoints is within this square
19 k = i #index of file to read in
20 l = j
21 }
22 }
23 #// Selecting which file to scan
24 filename = sprintf("/home/ionas/data/wwlln/wwlln-%02d-%02d.txt.bz2",k,l)
25 }
26 }
27 }
28 #// Scanning in lightningfile
29 connection = bzfile(filename) #defining connection to file
30 ifile = scan(connection) #scan in from .bz2-file
31 close(connection) #closing connection (to avoid reach max connections allowed)
32 #// Sorting out data from file
33 x = (1:length(ifile)-1)/3-1 #index vector from 0 to 1/3-1 of length(ifile)
34 longitude = ifile[1+3*x] #takes every third value of file with index 1,4,7...
35 latitude = ifile[2+3*x] #takes every third value of file with index 2,5,8...
36 lightningtimedays = ifile[3+3*x] #takes every third value of file with index 3,6,9...
37 lindays = as.integer(lightningtimedays)
38 #// Calculating distance between refpoint and points of lightningstrikes [km]
39 R = 6378.135 #earthis equatorial radius [km]
40 rad = pi/180 #defining radians radians
41 latrefrad = latref*rad #latref in radians
42 longrefrad = longref*rad
43 latrad = latitude*rad #lat in radians
44 longrad = longitude*rad
45 dlong = longrad - longrefrad #difference in latitude
46 dangle = 2*asin((sin(dlong) + longrad*tan(latrefrad)*cos(latrad)*sin(dlong)/2) / 2)
47 #// Only using measurements within that circle
48 assigncircle = Sc reference #setting TRUE for elements in "lfile" if Sc= reference, FALSE otherwise
49 lat = latitude[assigncircle] #only elements with TRUE
50 long = longitude[assigncircle]
Itimcdays = lightningtimedays[assigncircle]

// Converting time into YYYY-MM-DD HH:MM:SS.SS from ...DDD.DDD...
ltsec = lightningtimedays*24*3600 # converting to floating seconds
ltime = ISOdatetime(1970,1,1,0,0,0,0,tz="GMT") + ltsec # setting dates
ltime$test = as.integer(ltime) # used for testing timewindow selected above

# Dividing timedata into a frame called data$"time", ex: data$year
register.time = function(b)
{
  tlist = list()
  tlist$year = as.integer(substr(b, 1, 4)) # extracting years
  tlist$day = as.integer(substr(b, 6, 7))
  tlist$hour = as.integer(substr(b, 9, 10))
  tlist$minute = as.integer(substr(b, 12, 13))
  tlist$second = as.double(substring(b, 15, 16))
  # Checking diurnal differences
  tlist$hour = tlist$hour + tlist$minute/60 + tlist$second/3600 # column with decimal hours, i.e. when during the day did whistler occur in hours
  tlist$min = tlist$hour/60 # column with decimal minutes, i.e. when during the day did lightning occur in minutes
  tlist$sec = tlist$min/60 # column with decimal seconds, i.e. when during the day did lightning occur in seconds

  # Checking general dates
  tlist$date = as.Date(ISOdate(tlist$year, tlist$month, tlist$day)) # full date, ex: (2002-01-01)
  tlist$idate = as.integer(tlist$date) # integer days since epoch (Jan 1, 1970)
  tlist$fdate = tlist$idate + tlist$hour/24 # floating dates, i.e. days with decimal hours since epoch

  # Checking seasonal differences
  tlist$nrday = as.integer(format(tlist$date, "%j")) # day number, i.e. which day did a lightning occur counted from Jan 1st that year
  tlist$nrmonth = as.integer(format(tlist$date, "%m")) # month number, i.e. which month did a lightning occur during the year

  tlist
}
data = as.data.frame(register.time(ltime)) # defining call to function

# Creating some factors needed
fyear = factor(data$year)
ffmonth = factor(data$month)
ffday = factor(data$day)

data$hour = split(data$hour, fffmonth) # hours compared to month, i.e. diurnal distributions during a specific month
idata$month = split(data$month, fyear) # months compared to years, i.e. very long term changes
fyear = as.integer(levels(fyear)) # only using selected years once

# Setting local time at refpoints
lowlongref = -172.5 # setting longrefpoints to compare with
highlongref = -157.5
for(j in 1:24) # checking longitudes
{
  if(lowlongref<j & j<highlongref) # if longref is within this square
    timediff = -12+j # setting correct timedifference at refpoint from UTC
lowlongref = highlongref + 15
}
clockmarks = c(0,2,4,6,8,10,12,14,16,18,20,22,24) # clockmarks in UTC in plots
lclockmarks = clockmarks + timediff # unadjusted local clockmarks

for(i in 1:13)
if(lclockmarks[i] >=24)
Appendix, Source Code

```r
# Variables needed in plots
period = sprintf("%s to %s", ldata$date[1], ldata$date[length(ldata$date)]) # showing period of measurements
refpoint = sprintf("%s", longref, latref) # longref and latref as a string
refdist = sprintf("%s", reference) # reference as a string
ylab = sprintf("%s", ylab)

# Histograms

//Plot1 Complete distribution of lightning
breaks = seq(from=min(ldata$date), to=max(ldata$date)+1, by=daytimewindow) # setting breaks for dividing data
timewindow = as.integer(period) # not using Jan 1 in first year, data for this date probably don't exist
labels = as.Date(sprintf("%s-01-01", years)) # first day of each year with measurements
at = as.integer(labels) # setting tickmarks at these dates
data = divit lightning into frames defined by "breaks" and plotting a histogram including all the lightning
filename = sprintf("%s01ltot.eps", output)
main = sprintf("Observed lightning strikes \n %s \n %s", refpoint, refdist, period)
sub = sprintf("Timewindow for each frame is %d day(s) \n Total nr of observed lightning are %s", daytimewindow, length(ldata$year))
postscript(filename) # recording to eps
hist = hist(ldata$year, breaks=breaks, main=main, sub=sub, xlab="", ylab=ylab, axis=FALSE)
axis(side=1, labels=years, at=years+5, tick=FALSE, line=-1)
axis(side=2)
dev.off()

//Plot2 Creating Histogram over total amount of observed lightning based on years
filename = sprintf("%s02lyears.eps", output)
main = sprintf("Lightning divided into year of registration \n %s \n %s", refpoint, refdist, period)
sub = sprintf("")
breaks = fyear[1]:(fyear[length(fyear)]+1)
postscript(filename, filename) # recording to eps
hist = hist(ldata$year, breaks=breaks, col="blue", right=FALSE, main=main, xlab="", ylab="", sub=sub, axes=FALSE)
axis(side=1, labels=years, at=fyear+0, tick=FALSE, line=-1)
axis(side=2)
dev.off() # recording off

//Plot3 Creating Histogram over total amount of observed lightning based on months
filename = sprintf("%s03months.eps", output)
main = sprintf("Lightning divided into month of registration \n %s \n %s", refpoint, refdist, period)
postscript(filename) # recording to eps
hist = hist(ldata$month, breaks=1:13, col="blue", right=FALSE, main=main, xlab="", ylab="", sub=sub, axes=FALSE)
axis(side=1, labels=labels, at=at, tick=FALSE, line=-1)
axis(side=2)
```

dev.off() # recording off

### Plot 4 Creating Histogram over total amount of observed lightning based on days

filename = sprintf("%s04days.eps", output)
main = sprintf("Lightning divided into days of registration \n %s %s \n", reftime, refdist, period)
labels = c(5,10,15,20,25,30)
at = c(5.5,10.5,15.5,20.5,25.5,30.5)
hist = hist(ldata$day, breaks=1:32, col="blue", main=main, xlab="Day number in month", ylab=ylab, sub=sub, axes=FALSE)
axis(side=1, labels=labels, at=at)
axis(side=2)
dev.off() # recording off

### Plot5 Creating Histogram over total amount of observed lightning based on hours

filename = sprintf("%s05hours.eps", output)
main = sprintf("Lightning divided into hours of refpoint, refdist, period")
postscript(filename) $recording to .eps
hist = hist(ldata$hour, breaks=0:24, col="blue", main=main, xlab="", ylab=ylab, sub=sub, axes=FALSE)
axis(side=1, labels=lclockmarks, at=clockmarks, line=0)
xaxis(side=1, labels=clockmarks, at=clockmarks, tick=FALSE)
mtext(text="LT", side=1, line=1, at=25.5)
mtext(text="UTC", side=1, line=2, at=25.5)
dev.off() # recording off

### Plot 6-17 Plotting histogram with diurnal differences during all months for the the year during the whole period

z = matrix(ncol=24, nrow=12) # defining matrix
tax = 7:24
nr = 6
for(y in 1:12)
  mname = month.name[y] # selecting the correct name of month
  mn = month.abb[y] # short name of month
  filename = sprintf("%s%02dlhours%s.eps", output, Dr, mn) # giving name to "filename".pdf
  main = sprintf("Diurnal distribution of lightning for %s %s %s", reftime, refdist, period)
  postscript(filename) $recording to .eps
  hist = hist(ldatahour[y,], breaks=0:24, right=FALSE, col="blue", main=main, xlab="", ylab=ylab, sub=sub, axes=FALSE)
  axis(side=1, labels=lclockmarks, at=clockmarks, line=0)
xaxis(side=1, labels=clockmarks, at=clockmarks, tick=FALSE)
mtext(text="LT", side=1, line=1, at=25.5)
mtext(text="UTC", side=1, line=2, at=25.5)
  dev.off() # recording off
  z[y,x] = hist([2]) # extracting the counts for every round
  nr = nr + 1

### Plot 18 3D-image containing above diurnal distr plots

filename = sprintf("%s18image.eps", output)
main = sprintf("Diurnal distr of lightning divided into months \n %s %s %s", reftime, refdist, period)
breaks = c(1,5,10,50,100,500,1000,5000,10000,50000,100000)
text = paste(breaks[1:10], breaks[2:11], sep="-") # text showing values of bar colors
labelsc = c("Feb", "Apr", "Jun", "Aug", "Oct", "Dec")
axis(side=1, labels=labelsc, at=marc, line=0)
postscript(filename) $recording to .eps
par(mar=c(5.5,5,8.5)) # setting margin parameters with par(), for pdf: mar=c(5,4,4,7.5) ps: mar=c(5.5,5,8.5)
image(x=x, y=y, col=rev(heat.colors(10)), breaks=breaks, main=main, xlab="", ylab="", sub=sub, axes=FALSE)
ggrid(nx=12, ny=24, col="black")
axis(side=2, labels=clockmarks, at=marc, line=0)
Appendix, Source Code

244 mtext(text="LT", side=2, line=1, at=-2)
245 axis(side=2, labels=clockmarks, at=clockmarks, line=1.5, tick=FALSE)
246 mtext(text="UTC", side=2, line=2.5, at=-2)
247 axis(side=1, labels=labels, at=at, line=-1, tick=FALSE)
248 par(mar=c(10,4.5,10,5.7), new=TRUE) # defining shape for next plot and pasting
249 it on top of previous plot, for pdf: mar=c(10,23,10,6) ps: mar=(
250 10,4.5,10,5.7)
251 barplot(height=rep(1.1, 10), axes=FALSE, horiz=TRUE, col=rev(heat.colors(10)),
252 width=1, space=0.5) # plot showing colors
253 mtext(text="Nr of lightning", side=4, at=c(1, 2.5, 4.5, 5.7, 8.5, 10, 11.5, 13, 14.5), las=2, line =.5) # text showing values
254 mtext(text="Nr of lightning", side=4, at=16, las=2, line =.5) # headline of
255 dev.off()
256 #/ Plot 20 3D-persp containing above diurnal distr plots (quite bad plot)
257 filename = sprintf("%s19persp.eps", output)
258 main = sprintf("Diurnal distribution of lightning for all months \n %s %s %s",
259 reposition, refdist, period)
260 postscript(filename) # recording to .eps
261 persp(z=x, x=1:12, y=1:24, col="blue", main=main, xlab="Months", ylab="Hours"
262 , shade=0.5, sub="", tick=FALSE)
263 mtext(text=text, side=4, at=c(1, 2.5, 4.5, 5.7, 8.5, 10, 11.5, 13, 14.5), las=2, line =.5) # text showing values
264 mtext(text="UTC", side=2, linc=2.5, at=-2)
265 mtext(text="LT", side=2, linc=1, at=-2)
266 dev.off()
267 #/ Preparing for plot 20 and 21, sorting lightning occasions into into
268 correct movie frames
269 movieframe = seq(from=startmovie, to=timeteststop, by=
270 movieinterval*60) # limits for longitude in l-data, size of square
271 timetest = cut(timetest, breaks=movieframe) # factor for timetest, i.e.
272 sorting which movieframe they fall in
273 time = split(timetest, timetest) # splitting upp l-data depending on which
274 movieframe they belong to
275 latmovie = split(lat, timetest)
276 longmovie = split(long, timetest)
277 #
278 #/ Plot 20 movie with lightning strikes
279 filename = sprintf("%s20movielightningplot. eps", output) # name of file
280 timemovie = timeteststart # starting time for movie
281 postscript(filename) # recording to .pdf
282 for(i in 1:length(time)) # looping through movie frames
283 {
284 main = sprintf("Observed lightning strikes %s %s \n Time frame = %s to %s UTC",
285 reposition, refdist, as.character(time[i]),
286 as.character(timemovie[i])) # printing time intervall in each frame
287 timemovie = timemovie + movieinterval*60 # setting intervall for next frame
288 screen(i) # plotting each frame on a new screen
289 plot(longmovie, timemovie, main=main[i], xlab="Longitude", ylab="Latitude", sub=
290 sub, type="p", col="blue") # plotting lightning
291 }
292 dev.off()
293 #/ Plot 21 movie with hist of lightning
294 filename = sprintf("%s21movielightninghist. eps", output)
295 timemovie = timeteststart # setting start time for movie
296 postscript(filename) # recording to .pdf
297 for(i in 1:length(time))
298 {
299 main = sprintf("Total amount of lightning strikes are %d\n Length(longmovie[i])")
300 print(time intervall in each frame)
301 timemovie = timemovie + movieinterval*60 # setting intervall for next frame
302 screen(i) # plotting each frame on a new screen
303 plot(longmovie, timemovie, main=main[i], xlab="Longitude", ylab="Latitude", sub=
304 sub, type="p", col="blue") # plotting lightning
305 }
306 dev.off()
296 screen(i)
297 hist = hist(ltime[11]); breaks=breaks, col="blue", right=TRUE, main=main,
      xlab="", ylab=mylab, sub=sub, axes=FALSE)
298 axis(side=1, labels=labels, at=at)
299 axis(side=2)
300 }
301 dev.off() #recording off
302 #
Appendix, Source Code

1 // Correlation Code
2 #/ Variables that need to be given by user prior to running code
3 framesize = 0.5 # [degrees] Selecting size of square to cover the earth
4 and hence to run correlation for, must be a factor of 15, i.e. 1, 3, 5 or 15,
5 due to sourcedata sorted into frames of 15 degrees of width, less than
6 1 is ok, for example 0.5. Warning, program will run slowly for small
7 frames
8 timewindow = 30 # [s] Select slice ot time for running correlation
9 between
10 whistlers and lightning. Not too large due to bad correlation, but also
11 not too small due to risk of whistler and its lightningstroke ending up
12 in different frames. Good suggestion is around 30 seconds
13 wfile = sprintf("/home/jonas/work/correlation/data/Tihany.txt") # Select which
14 whistler file you want to scan, full path to file
15 sn = sprintf("TH") # give short name
16 of whistler station, ex: Tihany Hungary = TH, Dunedin New Zealand = DN...
17 used for naming output file
18 output : print(sprintf("/home/jonas/work/correlation/data/Tihany.txt", wfile)) # Select output
19 directory for output file, full path needed
20 #// Name of files containing final result
21 resultfile = sprintf("%s%scorrFS%3.1fW5d.xyz", output, timewindow)
22 rangefile = sprintf("%scorr-rangefile.txt", output)
23 #// Whistler data that need to be scanned in
24 wfile = scan(wfile)
25 #/ Reading in and creating a list with whistler data
26 register.wtime = function(a)
27 { #-- Checking diurnal differences --
28   tlist = list()
29   tlist$year = as.integer(substr(a, 1, 4))
30   tlist$month = as.integer(substr(a, 5, 6))
31   tlist$day = as.integer(substr(a, 7, 8))
32   tlist$hour = as.integer(substr(a, 9, 10))
33   tlist$minute = as.integer(substr(a, 11, 12))
34   tlist$second = as.double(substring(a, 13, 17))
35   tlist$dhour = tlist$hour * tlist$minute/60 + tlist$second/3600 # decimal hours, i.e. when during the day did
36   whistler occur in hours
37   #-- Checking general dates --
38   tlist$daydate = as.Date(ISOdate(tlist$year, tlist$month, tlist$day)) # full date, ex: 2002-01-01
39   tlist$daydate = as.integer(tlist$daydate) # integer days since epoch (Jan 1 1970)
40   tlist$daydate = tlist$daydate + tlist$hour/24 # floating dates, i.e. days
41   with decimal hours since epoch
42   tlist
43 }
44 wdata = as.data.frame(register.wtime(wfile)) # wdata name to call for variable
45 #/ Preparing comparison by defining time-frames (breaks) in which to sort
46 whistler and lightning data
47 start = wdata$daydate[1] # startdate in time for comparison
48 stop = wdata$daydate[length(wdata$daydate)] # stopdate for comparison
49 datetimewindow = timewindow/86400 # converting timewindow from seconds to days
50 breaks = seq(from=start, to=stop+1, by=datetimewindow) # setting breaks for
51 dividing data in between
52 breaks = sort(breaks) # sorting breaks
53 nB = length(breaks) # number of breaks
54 # fuzz to handle cases where points are "effectively on" the boundaries
55 # between breaks
56 fuzz = 0.5 + median(diff(breaks)) # small number based on median difference
57 # to move breaks slightly to include points on limit
58 breaks = breaks + fuzz # setting new breaks values
59 storage.model(breaks) = "double" # storing breaks as double since C below
60 calls C-code
61 #/ Hist function (don’t use this function in other programs, all security
62 and options removed for increased speed!!!)
63 newhist = function(x, breaks)
64 {
n = as.integer(length(x))  # length of data-vector to be sorted
storage.mode(x) = "double"  # storing data as double since .C below calls C-code

# with the fuzz adjustment above, the "right" and "include" arguments are often irrelevant (not with integer data)
counts = .C("bincount", x, n, breaks, as.integer(nB), right=TRUE, include=TRUE, naok=FALSE, naOK=FALSE, DUP=FALSE, PACKAGE="base") $counts

wcounts = newhist(wdata$ffdate, breaks=breaks)
wcounts[wcounts>1] = 1

// Dividing whistlers into time-frames defined by breaks above and assigning 1 to values>1

wdayspan = start:stop # all days between start and stop days for w-measurements
wdays = as.integer(levels(factor(wdata$ffdate))) # w-days with whistlers

wassign = is.element(wdayspan, wdays) TRUE for days including w-measurements
wassign = rep(wassign, each=daycountindex) # repeating wassign for all cells for these particular days
wcounts = wcounts[wassign] only cells for days with w-measurements selected

wdiffs = wdiff - mean(wcounts) # extracting difference for use further down

# Dividing w-measurements selected into frames defined by breaks above and assigning 1 to values>1

rvdiffs = mean(wdiffs) 

rm(timewindow, filelist, output, wdata, start, stop, daytimewindow, diddle, fuzz, wcounts, daycountindex, wdayspan) # removing all unnecessary variables

// Reading in the files containing lightning data, data divided into files covering 15 x 15 degrees, selecting correct file

nrofsquares = 15/frame-size # nr of squares in each dimension per file
halfframe = frame-size/2 # used in arguments below to determine distance from ref-point to end of square

for(i in 1:12) # index of files with lightning data in latitude

for(j in 1:24) # index of files with lightning data in longitude

latref = -90+15*(i-1) # start value for latref in selected file
longref = -180+15*(j-1) # start value for longref in selected file

// Scanning in lightning file

filename = sprintf("/home/ionas/data/wwlln/wwlln-O/OO2d-VoA2d.Lxt", i, j)
fileconnection = bzfile(filename) # defining connection to file
logfile = scan(fileconnection) # scan in from .bz2-file
close(fileconnection) # closing connection (to avoid max connections allowed)

// Sorting out data from file

x = 0:(length(lfile)/3)-1 # index vector from 0 to 1/3-1 of length(lfile)

longitude = lfile[1+3*x] # takes every third value of file with index 1, 4, 7, 10, 13, 16...
latitude = lfile[2+3*x] # takes every third value of file with index 2, 5, 8, 11, 14, 17...

lightningtimedays = lfile[3+3*x] # takes every third value of file with index 3, 6, 9, 12,...

lintdays = as.integer(lightningtimedays)

// Sorting out lightning dates for when whistler measurements exist, no interest in lightning statistics where there are no w-measurements. At deeper inspection this section might look ok to remove (similar sorting takes place in inner loop below), however, speed in loops below is increased plus a security feature when measurements are out of bounds for receiver are catered for. Keep this part of code!

assigndays = is.element(lintdays, wdays) TRUE for days with whistlers, length(assigndays)=length(lintdays)
Appendix, Source Code

103 lightningtimedays = lightningtimedays[assigndays] #only using days with
104 whistlers
105 longitude = longitude[assigndays]
106 latitude = latitude[assigndays]
107 #/Sorting lightning occasions into correct squares
108 longlim = seq(from=longref , to=longref+15, by=framesize) #limits for
109 longitude in data, size of square set by framesize
110 latlim = seq(from=latref , to=latref+15, by=framesize)
111 flong = cut(longitude , longlim) #factor for longitude
112 flat = cut(latitude , latlim)
113 lightningtimedays = split(lightningtimedays, list(flong, flat)) #splitting upp
114 l-data depending on long and lat
115
116 longref = longlim[-1]-halfframe #using all elements minus first one,
117 subtracting half frame to get midpoint
118 latref = latlim[-1]-halfframe
119 longref = rep(longref, times=nrofsquares) #repeating longref to fill squares
120 of file (ex:123123123)
121 latref = rep(latref, each=nrofsquares) #ex:111222333)
122
123 #/Looping for every square in file
124 for(k in lrnrofsquares^2) fnrofsquares defined just before first loop
125 above
126 {
127 ltimedays = lightningtimedays[[k]] #selecting lightning for every
128 square
129 longselect = longref[k] #selecting repoints
130 latselect = latref[k]
131 #/Dividing lightning into time-frames defined by breaks above
132 lcounts = newhist(ltimedays , breaks=breaks)
133 lcounts = lcounts[wassign] #only select days with w-measurements
134 to avoid false
135 from start to stop defined by breaks)
136 lcounts = lcounts>1 #assigning logical values (i.e.1) to counts>
137 1 #/Correlating logical results and extracting value
138 ldiff = lcounts-mean(lcounts) #used in corr function below
139 corr = sum(ldiff*wdiff)/(sum(ldiff ^2)^.5*wdiffsq) #same as ccf but much
140 faster
141 #/Writing result to resultfile
142 result = sprintf("%6.1f %5.1f %12.9f", longselect, latselect, corr)
143 write(result, resultfile, append=TRUE) #writing result to resultfile ,
144 appending result behind earlier results
145 #/Closing loops
146 }
147 }
148
149 #/Extracting the range (min and max) of corr and writing to rangefile
150 result = scan(resultfile) #scanning in resultfile
151 x = 0:((length(result)/3)-1) #index vector from 0 to 1/3-1 of length(result)
152 corr = result[x+3*x] #extracting corr
153 range = range(corr , na.rm=TRUE) #extracting max and min of corr
154 text = sprintf("%6.1f Min-correlation=%6.1f Max-correlation=%6.1f", resultfile, range
155 [1], range[2]) #creating text to be printed
156 write(text, rangefile, append=TRUE) #writing text to rangefile, appending
157 result behind earlier results
158 #/Removing all remaining variables to clean up after running program
159 rm(framesize, resultfile, rangefile, breaks, nB, wdays, wassign, wdiff,
160 wdiffsq, nrofsquares, halfframe, longref, latref, filename, lfile , x,
161 longitude, latitude, lightningtimedays, ltimedays, assigndays, longlim,
162 latlim, flong, flat, ltimedays, longselect, latselect, lcounts, ldiff,
163 corr, result, text)
// Correlation Code Time Slize

Variables that needs to be given by user prior to running code

frameSize = frameSizeSelect size of square to cover the earth and
hence to run correlation for, must be a factor of 15, i.e 1,3,5 or 15.
due to sourcedata sorted into frames of 15 degrees of width, less than
1 is ok, for example 0.5. Warning, program will run slowly for small
frames

timeWindow = 30 Select slize of time for running correlation between
whistlers and lightning. Not to large due to bad correlation, but also
not to small due to risk of whistler and its lightningstroke ending up
in different frames. Good suggestion is around 30 seconds

while = wfileSelect whistler file you want to scan, full path to file needed

sn = sprintf("%s", output) #select output diretory for output file, full path needed

sizeOfTimeIntervals = 3 Select size of time slice to sort
lightningdata into, must be factor of 24, select 24 if no slicing
desired (slows down code)

Name of files containing final result
rangeFile = "temp-rangefile.txt", output

#Whistlerdata that need to be scanned in
wfile = scan(wfile)

#Reading in and creating a list with whistler data
register.wtime = function(a)

{...

tlist = list()
tlist$year = as.integer(substr(a, 1, 4))
tlist$month = as.integer(substr(a, 5, 6))
tlist$day = as.integer(substr(a, 7, 8))
tlist$hour = as.integer(substr(a, 9, 10))
tlist$minute = as.integer(substr(a, 11, 12))
tlist$second = as.double(substring(a, 13, 17))

#-Checking diurnal differences

tlist$hour = tlist$hour * tlist$minute/60 + tlist$second/3600

decimal hours, i.e. when during the day did whistler occur in hours

#-Checking general dates

tlist$date = as.Date(ISOdate(tlist$year, tlist$month, tlist$day))

#full date, ex: (2002-01-01)
tlist$idate = as.integer(tlist$date) #integer days since epoch (Jan 1 1970)
tlist$fdate = tlist$idate + tlist$hour/24
tlist$idate = tlist$idate + tlist$hour/24

#-Preparing comparison by defining time-frames (breaks) in which to sort
whistler and lightning data

start = wdata$idate[1] #startdate in time for comparison
stop = wdata$idate[length(wdata$idate)] #stopdate for comparison
timewindow = (stop - start) #setting time window from seconds to days

breaks = seq(from=start , to=stop+1, by=timewindow) #setting breaks for
dividing data in between

#-fuzz to handle cases where points are "effectively on" the boundaries

diddle = 1e-7 * median(diff(breaks)) #small number based on median difference
betweens breaks

fuzz = c(-diddle, rep.int(diddle, length(breaks) - 1)) #vector lenght(breaks)
to move breaks slightly to include points on limit

breaks = breaks + fuzz #setting new breaks values

storage.mode(breaks) = "double" #storing breaks as double since .C below
calls C-code

#-Histfunction (don't use this function in other programs, all security
and options removed for increased speed!!)

newhist = function(x, breaks)
Appendix, Source Code

```r
52 [ 53 n = as.integer(length(x))  # length of data-vector to be sorted 54 storage.mode(x) = "double"  # storing data as double since C below calls C-code 55 # with the fuzz adjustment above, the "right" and "include" arguments are 56 # often irrelevant (not with integer data!) 57 counts = C("bincount", n, breaks = as.integer(nB), counts = integer(nB - 1), 58     right = TRUE, include = TRUE, naok = FALSE, DUP = FALSE, PACKAGE = "base") 59 ) 60 ) 61 # // Dividing whistlers into time-frames defined by breaks above and assigning 62 # to values >1 63 wcounts = newhist(wdata$fdate, breaks = breaks) 64 wcounts[wcounts > 1] = 1 65 # // Sorting out and only using cells for days with whistlers received, (wwln 66 # is assumed to always function). Due to that newhist returns zeros for 67 # days with receiver not functioning. These zeros must not contribute 68 # to correlation. Also preparing result for use when calculating corr 69 daycountindex = 1/daytimewindow  # nr of cells assigned per day to sort counts into 70 wdays = start:stop  # all days between start and stopdays for w-measurements 71 wdays = as.integer(levels(factor(wdata$sdate)))  # days with whistlers 72 wassign = is.element(wdays, wdays)  # TRUE for days including w-measurements 73 wassign = rep(wassign, each = daycountindex)  # repeating wassign for all cells 74 for these particular days 75 wcounts = wcounts[wassign]  # only cells for days with w-measurements selected 76 # // Fixing timeslices 77 nrofintervalsperday = 24 / sizeoftimeintervals  # nr of time intervals that 24 78 h period is split into 79 nrofbinsperintervalslice = 3600 / sizeoftimeintervals / daytimewindow  # nr of bins per 80 selected time slice 81 nrofbinsperday = nrofbinsperintervalslice * daytimewindow  # nr of bins per day 82 nrofbins = sum(wassign)  # nr of counts used 83 ref = 0:((nrofbinsperintervalslice - 1))  # ref vector to add on to index 84 startindex = seq(from = 1, to = nrofbins, by = nrofbinsperday)  # vector containing 85 the first index elements for the first time slice 86 # // Cleaning up to increase speed of program 87 rm(timezone, while, sn, output, wdata, start, stop, daytimewindow, diddle, 88 fuzz, daycountindex, wdays)  # removing all unnecessary variables 89 # // Reading in the files containing lightningdata, data divided into files 90 # covering 15*15 degrees, selecting correct file 91 nrofsquares = 15 / framesize  # nr of squares in each dimension per file 92 halfframe = framesize / 2  # used in arguments below to determine distance from 93 # refpoint to end of square 94 for(i in 1:12)  # index of files with lightningdata in latitude 95 for(j in 1:24)  # index of files with lightningdata in longitude 96 [ 97 [ 98 latref = -90 + 15 * (i - 1)  # start value for latref in selected file 99 longref = -180 + 15 * (j - 1)  # start value for longref in selected file 100 ] 101 # // Scanning in lightningfile 102 filename = sprintf("/home/ionas/data/wwln/wwln-ToO2d-ToO2d.txt.bz2", i, j) 103 fileconnection = bzfile(filename)  # defining connection to file 104 ifile = scan(fileconnection)  # scan in from .bz2-file 105 close(fileconnection)  # closing connection (to avoid reach max connections allowed) 106 # // Sorting out data from file 107 k = 0:((length(ifile) / 3) - 1)  # index vector from 0 to 1/3-1 of length(ifile) 108 longitude = ifile[1 + 3 * k]  # takes every third value of file with index 1, 4, 7, 109 latitude = ifile[2 + 3 * k]  # takes every third value of file with index 2, 5, 8, .. 110 lightningtimedays = ifile[3 + 3 * k]  # takes every third value of file with index 111 3, 6, 9 .. 112 lindays = as.integer(lightingtimedays) 113 #
```
Appendix, Source Code

107 // Sorting out lightning dates for when whistler measurements exist, no interest in lightning statistics where there are no w-statistics. At deeper inspection this section might look ok to remove (similar sorting takes place in inner loop below), however, speed in loops below is increased plus a security feature when measurements are out of bounds for breaks in newhist are catered for. Keep this part of code!

108 assigndays = is.element(lintdays, wdays) # TRUE for days with whistlers, length(assigndays)=length(lintdays)
109 lightningtimedays = lightningtimedays[assigndays] # only using days with whistlers
110 longitude = longitude[assigndays]
111 latitude = latitude[assigndays]
112 #
113 // Sorting lightning occasions into correct squares
114 longlim = seq(from=longref, to=longref+15, by=framesize) # limits for longitude in l-data, size of square set by framesize
115 latlim = seq(from=latref, to=latref+15, by=framesize)
116 flong = cut(longitude, longlim) # factor for longitude
117 flat = cut(latitude, latlim)
118 lightningtimedays = split(lightningtimedays, list(flong, flat)) # splitting upp l-data depending on long and lat
119
120 longref = longlim[-1]-halfframe # using all elements minus first one, subtracting halfframe to get midpoints
121 latref = latlim[-1]-halfframe
122 longref = rep(longref, times=nrofsquares) # repeating longref to fill squares of file (ex:123123123)
123 latref = rep(latref, each=nrofsquares) # ex(117222333)
124 #
125 # // Looping for every square in file
126 for(k in 1:nrofsquares^2) { # nrofsquares defined just before first loop above
127  timedays = lightningtimedays[[k]] # selecting lightning for every square
128  longselect = longref[k]
129  latselect = latref[k]
130  #
131  # // Dividing lightning into time-frames defined by breaks above
132  lcounts = newhist(timedays, breaks=breaks)
133  #
134  # // Selecting days with w-measurements to avoid false correlation (all days from start to stop defined by breaks)
135  lcounts = lcounts[lcounts>1] # only cells for days with w-measurements (removing zeros for other days)
136  #
137  # // Dividing whistlers (and lightning) into time intervalls during 24h period
138  tstart = 0
139  tstop = sizeoftimeintervals
140  for(l in 1:nrofintervalsperday) {
141    tstart = tstop * nrofbinspertimeslice # setting startindex for selecting next time slice
142    tstop = tstop + sizeoftimeintervals
143    indexref = startindex
144    tstart = tstart # tstart for next file
145    tstop = tstop + sizeoftimeintervals # tstop for next file
146    #
147    binindex = rep(indexref, each=nrofbinspertimeslice) # attaining right length of index with correct "base" value
148    binindex = binindex + ref # adding on to get correct index values
149    wcountsuse = wcounts[binindex] # only using selected bins
150    lcountsuse = lcounts[binindex]
151    indexref = indexref + nrofbinspertimeslice # setting startindex for selecting next time slice
152    #
153    tstart = tstop # tstart for next file
154    tstop = tstop + sizeoftimeintervals # tstop for next file
155    #
156    # // Correlating logical results and extracting value
157    wdiffsq = sum(wdiff^2) .5 # extracting factor for denominator in calculating corr
158    ldiff = lcountsuse-mean(lcountsuse) # used in corr function below
159    ldiffsq = sum(ldiff^2) .5
160    corr = sum(ldiff*wdiff)/(ldiffsq*wdiffsq) # same as ccf but much faster
161    
162}
Appendix, Source Code

```c
163 // Writing result to resultfile
164 result = sprintf("%6.1f %5.1f %12.9f", longselect, latselect, corr)
165 write(result, resultfile, append=TRUE) # writing result to resultfile,
166 appending result behind earlier results
167
168 // Closing loops
169 }
170 }
171 }
172
173 // Extracting the range (min and max) of corr and writing to rangefile
174 tstart = 0
175 tstop = sizeoftimeintervals
176 for(m in 1:nrofintervalsperday)
177 {
178  resultfile = sprintf("%s%scorrFSTdTrvlP/d\%d.xyz", output, sn, framesize,
179     timewindow, tstart, tstop)
180  result = scan(resultfile) # scanning in resultfile
181  x = 0:(length(result)/3-1) # index vector from 0 to 1/3-1 of length(result)
182  corr = result[3*x] # extracting corr
183  range = range(corr, na.rm=TRUE) # extracting max and min of corr
184  text = sprintf("%. Min-correlation=%f Max-correlation=%f", resultfile, range
185     [1], range[2]) # creating text to be printed
186  write(text, rangefile, append=TRUE) # writing text to rangefile, appending
187  result behind earlier results
188  tstart = tstop # tstart for next file
189  tstop = tstop + sizeoftimeintervals # tstop for next file
190 }
191
192 // Removing all remaining variables to clean up after running program
193 rm(framesize, resultfile, rangefile, breaks, nB, wdays, wassign, wdiff,
194     wdiffsq, nrofsquares, halfframe, latref, longref, filename, lfile, x,
195     longitude, latitude, lightningtimedays, lintdays, assigndays, longlim,
196     latlim, flong, flat, litimdays, longselect, latselect, lcounts, ldiff,
197     corr, result, text)
```
### Correlation Model Control Code

Variables that need to be given by user prior to running code:
- `wfile = sprintf("/home/jonas/data/whistlers/whd.cs0.20020227-20050518.txt")` # Select which whistler file you want to scan, full path to file needed
- `efile = sprintf("/home/jonas/work/correlation/data/whistler-logs-tihany.log")` #efile data

Output = `output = sprintf("/home/jonas/work/correlation/output/")` #select output directory for output files

Station = `station = sprintf("Tihany, Hungary")` #give name of whistler station (include country)

### Reading in and creating a list with whistler data

```r
register.wtime = function(a)
  tlist = list()
  tlist$year = as.integer(substr(a, 1, 4))
  tlist$month = as.integer(substr(a, 5, 6))
  tlist$day = as.integer(substr(a, 7, 8))
  tlist$hour = as.integer(substr(a, 9, 10))
  tlist$minute = as.integer(substr(a, 11, 12))
  tlist$second = as.double(substr(a, 13, 24))
  tlist$dhour = tlist$hour + tlist$minute/60 + tlist$second/3600 # decimal hours, i.e. when during the day did whistler occur in hours
  tlist$date = as.Date(ISOdate(tlist$year, tlist$month, tlist$day)) #full date, ex: (2002-01-01)
  tlist$date = as.integer(tlist$date) #integer days since epoch (Jan 1 1970)
  tlist$edate = as.data.frame(tlist)
```

### Reading in and creating a list with reference data

```r
register.etime = function(b)
  tlist = list()
  tlist$year = as.integer(substr(b, 18, 21))
  tlist$month = as.integer(substr(b, 23, 24))
  tlist$day = as.integer(substr(b, 26, 27))
  tlist$date = as.Date(ISOdate(tlist$year, tlist$month, tlist$day)) #full date, ex: (2002-01-01)
  tlist$edate = as.data.frame(tlist)
```

### Preparing comparison by defining variables

```r
start = max(wdata$idate[1], edata$idate[1]) #start date for comparison
stop = min(wdata$idate[length(wdata$idate)], edata$idate[length(edata$idate)])
```

```r
wedaysspan = start:stop #all days between start and stopdate for w/e-measurements
wdays = as.integer(levels(factor(wdata$idate))) #days with whistlers
edays = as.integer(levels(factor(edata$idate))) #days with ref measurements
wassign = is.element(wedaysspan, wdays) #TRUE for days including w-measurements
```
Appendix, Source Code

eassign = is.element(wedaysspan, edays)
compare = wassign==eassign #TRUE when eassign and wassign match
compare = as.integer(compare) #TRUE=1, FALSE=0

// Creating reference vector. 1=match, 2=no whistlers system running, 3=
// whistlers, system not running
reference = 1:length(wedaysspan)
for( i in 1:length(wedaysspan))
  if( wassign[i]==1 & eassign[i]==1) reference[i] = 1
  if( wassign[i]==0 & eassign[i]==0) reference[i] = 2
  if( wassign[i]==0 & eassign[i]==1) reference[i] = 3
  if( wassign[i]==1 & eassign[i]==0) reference[i] = 4

# Converting time into YYYY-MM-DD from DDDD...
wefile = ISOdate(1970,1,1, tz="GMT") + wedaysspan*86400 #setting dates

#Reading in and creating a list with this data
register.wetime = function(c)
  tlist : list otlist
  tlist$year : as.integer(substr(c, 1, 4))
  tlist$month : as.integer(substr(c, 6, 7))
  tlist$day : as.integer(substr(c, 9, 10))
  #--Checking general dates--
  tlist$year = as.Date(ISOdate(tlist$year, tlist$month, tlist$day)) #
  full date, ex: (2002-01-01)
  tlist$day = as.integer(tlist$day) #integer days since epoch (Jan 1
  1970)
  tlist

wedata = as.data.frame( register.wetime(wefile)) #edataSname to call for
    variable

# Creating some factors and start and stop month for plots
fyear = factor(wedata$year) #factor of years
fmonth = factor(wedata$month)
wedata$split = split(wedata, list(fmonth, fyear), drop=TRUE) #containing
hits divided by month and year, skipping months without measurement
(drop=TRUE)
startmonth = sprintf("%2s-%2s", wedata$year, wedata$month) #first year
  and month
stopmonth = sprintf("%2s-%2s", wedata$year[length(wedata$year)], wedata$month
(length(wedata$month))) #last year and month

#

//Plots

filename = sprintf("%s%scmchist. eps", output, sn)
period = sprintf("%s to %s", startdate, stopdate) #period of measurement
main = sprintf("Match ratio of model used in correlation code compared to
reference file for defining days when station not working")
sub = sprintf("%s %s \ Match ratio is \%2f percent\", station, period, ratio
)

hist(filename)

axis(side=1, labels=c("Mismatches", "Matches"), at=c(25, .75), line=-1, tick=
FALSE)
axis(side=2)
dev.off()

#//Plot2 Comparing days with image
ncol = length(wedata$split) #nr of columns in matrix
z = matrix(ncol=ncol, nrow=31) #defining matrix
Appendix, Source Code

```r
x=1:31 # refvector for matrix
null = rep(0,31) # refvector setting 0 to days in first month
ref = wedata$day[1]:length(wedatasplit[1])+wedatas$day[1]-1 # days (index)
with measurements during first month
wedatasplit[1] = replace(null,ref,wedatasplit[1]) # setting zeros before
first measurement in first month and after last date to fill to 31 if
necessary
null = rep(0,31-length(wedatasplit[[length(wedatasplit)]])) # refvector
setting zeros for last month without measurements
wedatasplit[length(wedatasplit)] = append(wedatasplit[length(wedatasplit)]
null,after=length(wedatasplit[length(wedatasplit)]))

for(y in 1:ncol) # filling matrix with values
if(identical(length(wedatasplit[[y]]),as.integer(30))) # filling up gaps in
months with zeros to reach length=31 of every month, ex 30 days gives
day 31=0
wedatasplit[[y]][31]=0
if(identical(length(wedatasplit[[y]]),as.integer(29))
wedatasplit[[y]] = append(wedatasplit[[y]],rep(0,2),after=29)
if(identical(length(wedatasplit[[y]]),as.integer(28))
wedatasplit[[y]] = append(wedatasplit[[y]],rep(0,3),after=28)
z[x,y] = wedatasplit[[y]] # extracting the values for every month to matrix

filename = sprintf("%s7ccmcimage.eps", output, sn)
main = sprintf("Days with matches of moder
used in correlation code compared to reference file")
#text = paste(breaks[1:10], breaks[2:11], sep="--") # text showing values of
bar colors
postscript(filename) # recording to .eps
image(x=x, x=1:31, y=1:ncol, col=colr("black","blue","green","red","yellow"),
breaks=breaks, main=main, xlab="", ylab="", sub="Blue=Match station down, Green=Match station functions, Green=Match station down, \nRed=Model says station broken, reference says station functions \nYellow=Model says functions, reference says broken, Black=No data ", axes=FALSE)
grid(nx=31, ny=ncol, col="black", lty="solid")
axis(side=1, labels=s10, at=c(1.5,10,15,20,25,30), tick=FALSE, line=1)
axis(side=2, labels=s10(startmonth, stopmonth), at=c(1,ncol), las=2)
mtext(side=1, line=1, text="Day number in month")
dev.off()
```
Appendix, Source Code

1 // WL Identification Code (code for helping in identifying prospective
2 # whistler/lightning pairs)
3 # Variables that needs to be given by user prior to running code
4 wfile = sprintf("/home/jonas/work/correlation/data/Tihany.txt") #which file
5 should be scanned
6 station = sprintf("Tihany, Hungary") #give the name of the whistler station (include country)
7 longref = 28.34 #[degrees] Selecting lightning longitude to compare with
8 latref = -33.45 #[degrees] Selecting lightning latitude to compare with
9 reference = 500 #[km] selecting max distance from refpoints allowed
10 timedifference = 10 #[s] selecting maximum time difference between events for
11 reporting eventual match
12 timeteststart = ISOdate(2004,12,01,00,00,00, tz="GMT") #selecting start time
13 for test
14 timeteststop : ISOdate(2004,12,31,00,00,00, tz="GMT") #selecting stop time for
15 test
16 # Whistler data that need to be scanned in
17 wfile = scan(wfile)
18 # Reading in and creating a list with whistler data
19 register.wtime = function(a)
20 {
21 tlist = list()
22 tlist$year = as.integer(substr(a, 1, 4))
23 tlist$month = as.integer(substr(a, 5, 6))
24 tlist$day = as.integer(substr(a, 7, 8))
25 tlist$hour = as.integer(substr(a, 9, 10))
26 tlist$minute = as.integer(substr(a, 11, 12))
27 tlist$second = as.double(substring(a, 13, 17))
28 #--- Checking diurnal differences ---
29 tlist$dhour = tlist$hour + tlist$minute/60 + tlist$second/3600 #
30 # decimal hours, i.e. when during the day did whistler occur in
31 # Hours
32 #--- Checking general dates ---
33 tlist$day = as.Date(ISOdate(tlist$year, tlist$month, tlist$day)) #
34 # Full date, ex. (2002-01-01)
35 tlist$day = as.integer(tlist$day) # Integer days since epoch (Jan 1
36 # 1970)
37 tlist$day = tlist$day + tlist$hour/24 # Floaing dates, i.e.
38 # Days
39 # with decimal hours since epoch
40 tlist
41 }
42 wdata = as.data.frame(register.wtime(wfile)) # wdata is name to call for variable
43 # Converting time into YYYY-MM-DD-HH-MM-SS.SS in wfile
44 wtsec = wdata$day * 24 * 3600 # Converting to floating seconds
45 wtime = ISOdatetime(1970,1,0,0,0,0, tz="GMT") + wtsec # setting dates
46 wtimestart = as.integer(wtime) used for testing time window selected above
47 # Calculating which lightning file to read in
48 for(i in 1:12) # Checking latitudes
49 {
50 latselect = -90+15*(i-1)+7.5 # Midpoint in each square
51 longselect = -180+15*(i-1)+7.5
52 longdiff = longselect-longref # Distance from midpoint to selected refpoint
53 if(longdiff < 7.5 & latselect < 7.5) # Checking if refpoint is within this square
54 {
55 l = i
56 }
57 }
58 # Selecting which file to scan
59 filename = sprintf("/home/jonas/data/wwlln/wwlln-%02d-%02d.txt.bz2",k,1)
60 if(longref > 25 & longref < 35 & latref > -35 & latref < -25)
61 {
62 }
Appendix, Source Code

# GMT Code
1 #/ Variables that needs to be given by user prior to running code
2 TS="-0-24" # give Time slice of code
3 INPUT=/home/jonas/work/correlation/output/THcorrfileFS0.5TW30.xyz # file to read in
4 OUTPUT="/home/jonas/work/gmt/output/THcorrfileFS0.5TW30.out" # output directory and output name
5 MAIN="Correlation , Thianny Hungary" # headline
6 REGION="-80,-180/-50,50" # region on earth for map to cover
7 GRIDSIZE="0.5/0.5" # size of squares x/y (use nr from FS in R, i.e 1 degree FS gives x/y=1/1, 3 degrees FS gives x/y=3/3)
8 COLORSCALE="-Cjet" # Colors to fill colorscale
9 Goritage="-T-.03/0.04/0.01" # Tmin/max/value steps (see min-min-
10 Correlation output in correlationcode.R)
11 MAPRANGE="0-2/5c" # central meridian of plot in degrees/width of plot c/cm
12 RESOLUTION="-Dc" # append f=full, h=high, intermediate, l=low, crude (for
13 world map c is suggested)
14 ANNOTATIONSIZE=14p # size of annotation
text
15 LPATH="/home/jonas/work/gmt/data" # path to folder with L-value files
16 LVALUES="" #/file with L-values (just
give filename)
17 LNAME="-LPATH/THnames.txt" # reading in text to print at given coordinates (see
18 #/names.txt) ##TH/DD...
19 INFO="#/LPATH/pstwts/TH/FS0.5TW30TS/TS.txt" # give path and name of info file to use
20 #/Setting initial parameters
21 OPTIONS="-JQMBAPRANGE -S$OPTIONS=map projection
22 GRIDFILE=gridfile.grd # name of grid-file (deleted with remove below
23 CPTFILE=cptfile cpt= name of cpt-file (deleted with remove below
24 GMT gmtset ANNOT_FONT_SIZE = $ANNOTATIONSIZE
25 #/Converting data from .xyz to .grd
26 GMT xyz2grd $INPUT -OGRID -I$GRIDSIZE $REGION =F
27 $NAME of output gridfile, -l=grid spacing, -F=force pixel registration
28 #/Creating color palette file
29 GMT makecpt $COLORSCALE $COLORRANGE =Z >$CPT
30 #/Continuous color palette
31 #/Adding grid and data
32 GMT grdimage $GRIDFILE $OPTIONS -C$CPT -B;"$MAIN": -Ts -K >OUTPUT
33 $OPTIONS=map projection
34 GMT gmtset ANNOT_FONT_SIZE =$ANNOTATIONSIZE
35 #/Creating basic map with tickmarks etc
36 GMT psbasemap $OPTIONS REGION -Bx0y300s30/0x0y300s30SW >O -K >OUTPUT
37 -$N=political boundaries/thickness of line, -W=draw coastlines
38 #/Plotting L-values on map
39 GMT pssy $LVALUES $OPTIONS $REGION =H2 -S-0.01 -O -K >OUTPUT
40 -$H2=remove first two lines in file (header), -S=symbol to plot with size of
41 symbol appended
42 #/Plotting names of L-values in map
43 GMT ptext $LNAME $OPTIONS $REGION =H1 -O -K >OUTPUT
44 #/Plotting info text in map
45 GMT ptext $INFO $OPTIONS $REGION =H1 -O -K >OUTPUT
46 #/Adding scalebar to map
47 GMT psscale =-12.5c/-1.5c/15c/0.25c/h -OCPT -E-B:correlation: -O >OUTPUT
48 -$D=position of scalebars x/y/length/width/horizontal, -E=triangles at end of
49 scale, -l= name of bar
50 #/Removing unnecessary files
51 rm $GRDJ $CPT # (shell command)
52 #/General explanations
# GMT program executes selected gmt program
# -K=append more data to file
# -O=append this data to pre-existing file
# >=store output on given file (shell command)
# >>=store output in given pre-existing file (shell command)
# $=call file/command with given name (shell command)
Bibliography


[6] Israelsson Sven, Lecture notes from the course 'Atmospheric Science with Climatology', Uppsala University, Uppsala, Sweden 2006.


