

# **A lightning threat index for South Africa using numerical weather prediction data**

by

**Morné Gijben**

Submitted in partial fulfilment of the requirements for the degree of

**Master of Science (Meteorology)**

In the Faculty of Natural & Agricultural Sciences

University of Pretoria

Pretoria

April 2016

# A lightning threat index for South Africa using numerical weather prediction data

Student: Morné Gijben  
Supervisor: Dr Liesl L. Dyson  
Department: Geography, Geoinformatics and Meteorology  
Faculty: Natural and Agricultural Sciences  
University: University of Pretoria  
Degree: Master of Science (Meteorology)

## ABSTRACT


Lightning is a phenomenon that can cause death or injury to humans and animals, damage to infrastructures, and can be a hazard to various sectors like the aviation and forestry industries. There is a need for prediction techniques to ensure the protection of people and property. In this dissertation, a new lightning threat index (LTI) is proposed for southern Africa. The aim of the LTI is to identify the areas where lightning is likely to occur during the day. Before the LTI could be developed, it was necessary to identify candidate model predictors capable of predicting the occurrence of lightning. In total 25 predictors were selected from literature that showed promising results to forecast the occurrence of lightning. The selected predictors are different variations from the following six groups of parameters; convective available potential energy, lifted index, precipitable water, equivalent potential temperature, relative humidity and air temperature. This study identifies the parameter from each of the six groups capable of predicting the occurrence of lightning over southern Africa the best during spring and summer by means of stepwise logistic regression techniques. The six parameters identified in this study for spring are; the most unstable convective available potential energy in the 1 - 6 km above ground level range, surface lifted index, mean precipitable water in the 850 to 300 hPa layer, minimum relative humidity in the 3-6 km above ground level layer, equivalent potential temperature lapse rate between 700 and 500 hPa and mean temperature in the 850 – 700 hPa layer. During summer, the same parameters were identified, except that the average relative humidity in the 3-6 km above ground level layer and equivalent potential temperature

lapse rate between 850 and 400 hPa were identified. After the most appropriate parameters, capable of predicting the occurrence of lightning, were identified, the development of the new LTI could commence. Since the goal was to develop a single index that utilises the different model predictors to forecast the binary outcome of lightning occurrence (yes or no), attention was given to binary logistic regression techniques. In this study a rare-event binary logistic regression technique is used to develop equations for the LTI that utilise NWP model output early in the morning to provide a probability forecast of where lightning is expected to occur during the day between 07:00 and 21:00 UTC. The new LTI is evaluated over an entire independent spring and summer season. Results show that the LTI forecasts have a high sensitivity and specificity for both the spring and summer seasons. The LTI is not so reliable during the spring season, since it over-forecasts the occurrence of lightning, but during the summer season, the LTI forecast is reliable, only slightly over-forecasting the lightning activity. The LTI produces sharp forecasts during both the spring and summer seasons. The LTI will be a useful tool to operational weather forecasters or sectors interested in lightning forecasts, to provide guidance early in the morning on the areas of interest where lightning can be expected during the day, and can ultimately contribute to society by aiding with timely warnings of lightning or thunderstorms to protect humans, animals and property.

## DECLARATION

---

I, Morné Gijben declare that the dissertation, which I hereby submit for the degree Master of Science (Meteorology) at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

SIGNATURE: .....

DATE: .....29 April 2016.....

## ACKNOWLEDGEMENTS

---

I would like to express my deepest gratitude to:

- Dr. Liesl Dyson, my supervisor, for all her assistance, guidance and support during the course of the study and for the many coffees to discuss the work.
- Mr. M.T. Loots for his statistical advice and for running some of the programming with SAS statistical software.
- Dr. E. de Coning, my work supervisor, for her encouragement and support during this study, and also for making the funds available from her Water Research Commission projects.
- The South African Weather Service for providing the technology, the extremely valuable data and the time to complete the study.
- The Water Research Commission for the awarding of projects that encourages capacity building, from which this study could be funded.
- The Almighty God for providing me with the health, strength and ability to complete this study.
- My parents, Robert and Ursula Gijben, for all their support and encouragement, not only during the course of this study, but through my entire academic career.

## PREVIOUS PUBLICATIONS BY AUTHOR

---

- Gijben M. 2013. Verification of a Unified Model based Lightning Risk Indicator for Southern Africa. *Proceedings of the 29<sup>th</sup> Annual Conference of South African Society for Atmospheric Sciences*. Durban, South Africa, September 2013, pp. 58-61. ISBN 978-0-620-56626-1
- Gijben, M. 2014. Investigating the plausibility of a model based lightning risk indicator for South Africa. *Proceedings of the 15<sup>th</sup> International Conference on Atmospheric Electricity*. Norman Oklahoma. June 2014.
- Gijben, M., Dyson, L.L. and Loots, M.T. 2015. A lightning threat index for South Africa using numerical weather prediction data. *Proceedings of the 31st Conference of the South African Society for Atmospheric Science*. Hennops River Valley, South Africa, September 2015, pp. 136-139. ISBN 978-0-620-67825-4

## TABLE OF CONTENTS

---

ABSTRACT.....	i
DECLARATION .....	iii
ACKNOWLEDGEMENTS .....	iv
PREVIOUS PUBLICATIONS BY AUTHOR.....	v
TABLE OF CONTENTS.....	vi
LIST OF SYMBOLS AND ABBREVIATIONS .....	x
LIST OF FIGURES .....	xv
LIST OF TABLES.....	xviii
<b>CHAPTER 1 .....</b>	<b>1</b>
1.1 BACKGROUND.....	1
1.2 MOTIVATION .....	3
1.3 AIMS AND OBJECTIVES .....	4
1.4 DISSERTATION OUTLINE .....	5
<b>CHAPTER 2 .....</b>	<b>6</b>
2.1 INTRODUCTION .....	6
2.2 LIGHTNING FORMATION .....	6
2.2.1 Electrical structure of a thundercloud .....	6
2.2.2 Electrification of a thundercloud .....	8
2.2.3 Electrical Discharge .....	10
2.3 LIGHTNING DETECTION.....	12
2.4 DISTRIBUTION OF LIGHTNING .....	14
2.4.1 Globally.....	14
2.4.2 South Africa .....	15
2.5 PREDICTION OF LIGHTNING.....	17
2.5.1 Remote Sensing Instruments.....	18
2.5.2 Numerical Weather Prediction Models .....	18
2.5.3 Parameters useful for lightning prediction .....	19
2.5.3.1 Convective Available Potential Energy .....	19
2.5.3.2 Lifted Index.....	22
2.5.3.3 Equivalent Potential Temperature .....	24
2.5.3.4 Precipitable Water .....	26



2.5.3.5	Relative Humidity .....	27
2.5.3.6	Air temperature .....	28
2.6	NWS LIGHTNING POTENTIAL INDEX.....	29
2.7	SUMMARY .....	31
<b>CHAPTER 3</b>	.....	<b>32</b>
3.1	INTRODUCTION .....	32
3.2	STUDY AREA AND PERIOD.....	32
3.2.1	Study area .....	32
3.2.2	Study period .....	33
3.3	LIGHTNING DATA.....	35
3.3.1	Southern African Lightning Detection Network .....	35
3.3.1.1	The network .....	35
3.3.1.2	The sensors .....	37
3.3.1.3	Advantages and limitations.....	38
3.3.2	Lightning data utilised in this study.....	38
3.4	NUMERICAL WEATHER PREDICTION MODEL DATA .....	39
3.4.1	The Unified Model.....	39
3.4.1.1	The model.....	39
3.4.1.2	Advantages and limitations.....	40
3.4.2	Model data utilised in this study .....	41
3.5	DEVELOPMENT OF A NEW LIGHTNING THREAT INDEX (LTI).....	44
3.5.1	Parameter Selection .....	44
3.5.1.1	Stepwise regression with SAS.....	44
3.5.1.2	Stepwise regression with R .....	45
3.5.2	Development of LTI with logistic regression .....	46
3.6	VERIFICATION OF THE LTI.....	48
3.6.1	Probabilistic verification .....	48
3.6.1.1	ROC curve .....	48
3.6.1.2	Reliability diagram .....	50
3.6.1.3	Sharpness diagram .....	51
3.6.2	Dichotomous verification .....	52
3.6.2.1	Probability of Detection .....	53
3.6.2.2	Probability of false detection.....	53
3.6.2.3	False alarm ratio .....	54
3.6.2.4	Hanssen and Kuipers Discriminant.....	54





3.6.2.5	Frequency Bias .....	54
3.6.3	Eyeball verification .....	54
3.7	SUMMARY .....	56
<b>CHAPTER 4</b>	.....	<b>57</b>
4.1	INTRODUCTION .....	57
4.2	PARAMETER SELECTION .....	57
4.2.1	Convective Available Potential Energy (CAPE) .....	59
4.2.2	Lifted Index (LI) .....	62
4.2.3	Precipitable Water (PW) .....	64
4.2.4	Relative Humidity (RH) .....	67
4.2.5	Equivalent Potential Temperature ( $\Theta_e$ ) .....	70
4.2.6	Temperature .....	73
4.2.7	Summary of selected parameters .....	75
4.3	DEVELOPMENT OF THE NEW LTI .....	76
4.3.1	SON period .....	76
4.3.2	DJF period .....	78
4.4	SUMMARY .....	80
<b>CHAPTER 5</b>	.....	<b>81</b>
5.1	INTRODUCTION .....	81
5.2	PROBABILISTIC VERIFICATION .....	81
5.2.1	ROC curve .....	82
5.2.2	Reliability .....	83
5.2.3	Sharpness .....	84
5.3	DICHOTOMOUS VERIFICATION .....	84
5.4	EYEBALL VERIFICATION .....	89
5.4.1	Case 1: 19 January 2013 .....	90
5.4.2	Case 2: 10 October 2013 .....	92
5.4.3	Case 3: 7 November 2013 .....	93
5.4.4	Case 4: 20 December 2013 .....	95
5.4.5	Case 5: 15 January 2014 .....	97
5.4.6	Case 6: 6 February 2014 .....	98
5.5	SUMMARY .....	100



<b>CHAPTER 6</b> .....	101
6.1 INTRODUCTION .....	101
6.2 SUMMARY .....	101
6.2.1 Objective 1: Identification of candidate NWP model predictors .....	102
6.2.2 Objective 2: Selection of the most appropriate NWP model parameters for lightning prediction. ....	103
6.2.3 Objective 3: Development of a new lightning prediction index .....	106
6.2.4 Objective 4: Quantitative evaluation of the new lightning prediction index....	107
6.2.5 Objective 5: Qualitative evaluation of the new index against existing products	108
6.3 RECOMMENDATIONS.....	109
6.4 CONTRIBUTION TO SCIENCE .....	110
REFERENCES .....	111

## LIST OF SYMBOLS AND ABBREVIATIONS

---

AGL	-	Above ground level
AIC	-	Akaike Information Criterion
AUC	-	Area under the ROC curve
aveRH <sub>3,6km_AGL</sub>	-	Average relative humidity between 3 – 6 km AGL
BLI	-	Best lifted index
CAPE	-	Convective available potential energy
CAPE <sub>surf</sub>	-	Surface CAPE
CG	-	Cloud-to-ground
cm	-	Centimetres
c <sub>p</sub>	-	Specific heat of dry air at constant pressure
Coef	-	Regression coefficient
DE	-	Detection efficiency
DJF	-	December, January and February
e	-	Vapour pressure
e <sub>s</sub>	-	Saturation vapour pressure
FAR	-	False alarm ratio
g	-	Gravity
GPS	-	Global positioning systems
HK	-	Hanssen and Kuipers Discriminant
hPa	-	Hectopascal
Hz	-	Hertz
IC	-	Intra-cloud
J kg <sup>-1</sup>	-	Joules per kilogram
K	-	Kelvin
kA	-	Kilo ampere
km	-	Kilometre

km <sup>2</sup>	-	Square kilometre
km <sup>-2</sup>	-	Per square kilometre
L <sub>c</sub>	-	Latent heat of condensation
LA	-	Location accuracy
LDN	-	Lightning detection network
LF	-	Low frequency
LFC	-	Level of free convection
LI	-	Lifted index
LIS	-	Lightning imaging sensor
LMA	-	Lightning mapping array
LNB	-	Level of neutral buoyancy
LPI	-	Lightning Potential Index
LTl	-	Lightning Threat Index
m	-	Meter
maxRH <sub>3,6km_AGL</sub>	-	Maximum relative humidity between 3 – 6 km AGL
mb	-	Millibar
MDF	-	Magnetic direction finding
MHz	-	Megahertz
minRH <sub>3,6km_AGL</sub>	-	Minimum relative humidity between 3 – 6 km AGL
mCAPE	-	Mean layer CAPE
MPI	-	Measure of parameter importance
MPV	-	Maximum parameter value
ms <sup>-1</sup>	-	Meters per second
muCAPE	-	Most unstable CAPE
muCAPE <sub>0,3km_AGL</sub>	-	Most unstable CAPE between 0 – 3 km AGL
muCAPE <sub>1,6km_AGL</sub>	-	Most unstable CAPE between 1 – 6 km AGL
muCAPE <sub>lowest,300</sub>	-	Most unstable CAPE in the lowest 300 hPa
NALDN	-	National Lightning Detection Network

NASA	-	National Aeronautics and Space Administration
NCAR	-	National Centre for Atmospheric Research
NPEP	-	Network Performance Evaluation Program
NWP	-	Numerical weather prediction
NWS	-	National Weather Service
Odds	-	Odds ratio
OTD	-	Optical Transient Detector
P	-	P-value
$p$	-	Starting pressure level
$p_0$	-	Stopping pressure level
POD	-	Probability of detection
POFD	-	Probability of false detection
PW	-	Precipitable water
$PW_{850,300}$	-	Mean precipitable water between 850 and 300 hPa
$PW_{700,400}$	-	Mean precipitable water between 700 and 400 hPa
$PW_{surf,100}$	-	Mean precipitable water from surface to 100 hPa
$q_s$	-	Saturation mixing ratio
RH	-	Relative humidity
$RH_{m10}$	-	Mean relative humidity at the $-10^{\circ}\text{C}$ level
$RH_{m12,m18}$	-	Mean relative humidity between $-12^{\circ}\text{C}$ and $-18^{\circ}\text{C}$ levels
ROC	-	Receiver Operating Characteristic
SALDN	-	Southern African Lightning Detection Network
SAS	-	Statistical and Analysis Software
SAST	-	South Africa Standard Time
SAWS	-	South African Weather Service
SE	-	Standard error on regression coefficient
SLI	-	Surface lifted index
SON	-	September, October and November

T	-	Temperature
$T_{1p5m}$	-	Temperature at 1.5 m above the ground
$T_{850}$	-	Temperature at 850 hPa
$T_{700}$	-	Temperature at 700 hPa
$T_{850,700}$	-	Mean T of pressure levels between 850 and 700 hPa
$T_{500,300}$	-	Mean T of pressure levels between 500 and 300 hPa
$T_{500}$	-	Temperature of the environment at 500 hPa
$T_{p500}$	-	Temperature of an air parcel at 500 hPa
$T_{V_e}$	-	Virtual temperature of the environment
$T_{V_p}$	-	Virtual temperature of an air parcel
TOA	-	Time of arrival
TRMM	-	Tropical Rainfall Measuring Mission
UKMO	-	United Kingdom Meteorological Office
UM	-	Unified Model
USA	-	United States of America
UTC	-	Coordinated Universal Time
VHF	-	Very high frequency
VLF	-	Very low frequency
$yr^{-1}$	-	Per year
Z	-	Z-value
$\Theta$	-	Potential temperature
$\Theta_e$	-	Equivalent potential temperature
$\Theta_e\Gamma$	-	Equivalent potential temperature lapse rate
$\Theta_e\Gamma_{surf}$	-	Surface $\Theta_e$ lapse rate
$\Theta_e\Gamma_{600}$	-	$\Theta_e$ lapse rate at 600 hPa
$\Theta_e\Gamma_{850,400}$	-	$\Theta_e$ lapse rate between 850 and 400 hPa
$\Theta_e\Gamma_{850,500}$	-	$\Theta_e$ lapse rate between 850 and 500 hPa
$\Theta_e\Gamma_{1,6km\_AGL}$	-	$\Theta_e$ lapse rate between 1 and 6 km AGL

$\Theta_e \Gamma_{m10,m20}$	-	$\Theta_e$ lapse rate between -10°C and -20°C levels
$\Theta_e \Gamma_{700,500}$	-	$\Theta_e$ lapse rate between 700 and 500 hPa
°	-	Degree
°C	-	Degrees Celsius
<	-	Less than
≤	-	Less than and equal
>	-	Greater than
≥	-	Great than and equal
%	-	Percentage
μs	-	Microsecond

## LIST OF FIGURES

---

Figure 2-1: Most likely charge structure of a South African thundercloud [Cited from Uman (2012), Source: Malan (1963)] .....	7
Figure 2-2: Cloud electrification by means of charge transfer because of collisions between graupel and ice particles in the presence of water droplets. The assumption is made that the charge reversal temperature is $-15^{\circ}\text{C}$ and occurs at a height of 6 km. [Cited from Rakov and Uman, 2003].....	9
Figure 2-3: The main types of lightning (cloud-to-ground, intra-cloud, cloud-to-cloud, and cloud-to-air) [Adapted from Jaffer et. al. (2011), Source: Rakov and Uman (2003)] .	10
Figure 2-4: The processes involved in a typical negative cloud-to-ground lightning flash. P indicates the main positive charge region, N the main negative charge region and LP the lower positive charge region [Cited from Rakov and Uman (2003), Source: Uman (1987; 2001)].....	12
Figure 2-5: Electromagnetic pulse signatures from lightning flashes in the different frequency ranges [Cited from Vaisala (2004), Source: Malan (1963)] .....	13
Figure 2-6: Global average annual lightning flash density on a $0.5^{\circ} \times 0.5^{\circ}$ resolution from the combined observations of the NASA OTD and TRMM LIS instruments [Source: Information Technology and Systems Centre, University of Alabama, 2015] .....	15
Figure 2-7: The average annual lightning ground flash densities (flashes $\text{km}^{-2} \text{yr}^{-1}$ ) for the 9-year period from 2006 to 2014 over South Africa [Source: South African Weather Service] .....	16
Figure 2-8: The average annual lightning ground flash densities for (a) Spring (September, October, November), (b) Summer (December, January, February), (c) Autumn (March, April, May) and (d) Winter (June, July, August) over South Africa from 2006-2014 [Source: South African Weather Service]. .....	17
Figure 2-9: LPI outlook maps for the State of Colorado together with the colour key and explanation [Source: National Weather Service, 2013] .....	29
Figure 3-1: The study domain. Grey areas are excluded.....	33
Figure 3-2: The average diurnal cycle of lightning over South Africa for the period 2006-2013, expressed as a percentage of the daily total. ....	34
Figure 3-3: The location of the 24 lightning detection sensors of the SALDN network operated by SAWS [Map created from information from Ngwato (2014)] .....	35
Figure 3-4: The projected (a) detection efficiencies and (b) location accuracies of the SALDN from the Vaisala NPEP .....	36



Figure 3-5: Absolute maximum range of the SALDN, where the blue area represents the areas where lightning can be detected. Outside the borders of SA, the accuracy of lightning detection reduces rapidly..... 37

Figure 3-6: The domain of the 12 km horizontal resolution Unified Model at SAWS ..... 40

Figure 3-7: An explanation of a ROC curve. Line A represents a perfect forecast (AUC=1), line B a typical ROC curve (AUC=0.85) and line C represents no skill or random chance (AUC=0.5) [Adapted from Zou et. al., 2007]..... 49

Figure 3-8: An example of a reliability diagram. The blue line represents a hypothetical forecast, the black diagonal line will be a forecast with perfect reliability, and the horizontal dotted line will be a forecast with no resolution. .... 50

Figure 3-9: Example of a sharpness diagram [Adapted from Cintineo et al., 2010] ..... 51

Figure 3-10: A standard 2 x 2 contingency table ..... 52

Figure 4-1: Box and whisker plots of  $\mu\text{CAPE}_{1,6\text{km\_AGL}}$  between 07:00 – 21:00 UTC for lightning and no Lightning. SON is shown on the left and DJF on the right. CAPE units are in  $\text{J kg}^{-1}$ . The thick horizontal bar indicates the median value; the boxes denote the 25<sup>th</sup> –75<sup>th</sup> percentiles, and the whiskers show the full range of values. The circles indicate outliers..... 61

Figure 4-2: Box and whisker plots of SLI between 07:00 – 21:00 UTC for lightning and no lightning. SON is shown on the left and DJF on the right. SLI units are in  $^{\circ}\text{C}$ . The thick horizontal bar indicates the median value; the boxes denote the 25<sup>th</sup> –75<sup>th</sup> percentiles, and the whiskers show the full range of values. The circles indicate outliers..... 64

Figure 4-3: Box and whisker plots of  $\text{PW}_{850,300}$  between 07:00 – 21:00 UTC for lightning and no Lightning. SON is shown on the left and DJF on the right. PW units are in cm. The thick horizontal bar indicates the median value; the boxes denote the 25<sup>th</sup> –75<sup>th</sup> percentiles, and the whiskers show the full range of values. The circles indicate outliers..... 66

Figure 4-4: Box and whisker plots of  $\text{minRH}_{3,6\text{km\_AGL}}$  during SON (left) and  $\text{aveRH}_{3,6\text{km\_AGL}}$  during DJF (right) between 07:00 – 21:00 UTC for lightning and no Lightning. RH units are in %. The thick horizontal bar indicates the median value; the boxes denote the 25<sup>th</sup> –75<sup>th</sup> percentiles, and the whiskers show the full range of values. The circles indicate outliers..... 69

Figure 4-5: Box and whisker plots of  $\Theta_e\Gamma_{700,500}$  during SON (left) and  $\Theta_e\Gamma_{850,400}$  during DJF (right) between 07:00 – 21:00 UTC for lightning and no Lightning.  $\Theta_e\Gamma$  units are in K. The thick horizontal bar indicates the median value; the boxes denote the 25<sup>th</sup> –75<sup>th</sup> percentiles, and the whiskers show the full range of values. The circles indicate outliers..... 72

Figure 4-6: Box and whisker plots of $T_{850,700}$ during SON (left) and DJF (right) between 07:00 – 21:00 UTC for lightning and no Lightning. Temperature units are in K. The thick horizontal bar indicates the median value; the boxes denote the 25th –75th percentiles, and the whiskers show the full range of values. The circles indicate outliers.....	74
Figure 5-1: ROC curves and AUC values for the UM LTI forecasts during (a) SON of 2013 and (b) DJF of 2013/14.....	82
Figure 5-2: Reliability diagrams for the UM LTI forecasts during (a) SON of 2013 and (b) DJF of 2013/14.....	83
Figure 5-3: Sharpness diagrams for the UM LTI forecasts during (a) SON of 2013 and (b) DJF of 2013/14.....	84
Figure 5-4: Statistical scores for the UM LTI forecasts during SON of 2013.....	86
Figure 5-5: Frequency bias for the UM LTI forecasts during SON of 2013.....	87
Figure 5-6: Statistical scores for the UM LTI forecasts during DJF of 2013/14.....	88
Figure 5-7: Frequency bias for the LTI forecasts during the DJF period.....	89
Figure 5-8: LPI (top left), LTI (top right), UM convective rainfall in cm (bottom left), and occurrence of lightning (bottom right) for 19 September 2013 between 07:00 and 22:00 UTC.....	91
Figure 5-9: LPI (top left), LTI (top right), UM convective rainfall in cm (bottom left), and occurrence of lightning (bottom right) for 10 October 2013 between 07:00 and 22:00 UTC.....	93
Figure 5-10: LPI (top left), LTI (top right), UM convective rainfall in cm (bottom left), and occurrence of lightning (bottom right) for 7 November 2013 between 07:00 and 22:00 UTC.....	94
Figure 5-11: LPI (top left), LTI (top right), UM convective rainfall in cm (bottom left), and occurrence of lightning (bottom right) for 20 December 2013 between 07:00 and 22:00 UTC.....	96
Figure 5-12: LTI on UM 12 km resolution (top left), LTI on 0.5° X 0.5° resolution (top right), UM convective rainfall in cm on UM 12 km resolution (bottom left), and UM convective rainfall on 0.5° X 0.5° resolution (bottom right) for 20 December 2013 between 07:00 and 22:00 UTC.....	97
Figure 5-13: LPI (top left), LTI (top right), UM convective rainfall in cm (bottom left), and occurrence of lightning (bottom right) for 15 January 2014 between 07:00 and 22:00 UTC.....	98
Figure 5-14: LPI (top left), LTI (top right), UM convective rainfall in cm (bottom left), and occurrence of lightning (bottom right) for 6 February 2014 between 07:00 and 22:00 UTC.....	99

## LIST OF TABLES

---

Table 2-1: Possible thresholds for CAPE [Adapted from Haby, Severe Weather Indices Page2015] .....	20
Table 2-2: Possible thresholds for LI [Adapted from Haby, Severe Weather Indices Page, 2015].....	23
Table 3-1: Candidate predictors for inclusion in the lightning prediction model.....	42
Table 3-1: Candidate predictors for inclusion in the lightning prediction model (continued) .	43
Table 3-2: Example of dataset used as input to the regression analysis to build the LTI .....	47
Table 3-3: Case studies selected for the eyeball verification of the LPI, LTI and convective rainfall against the occurrence of lightning .....	55
Table 4-1: SAS output from the stepwise logistic regression model for all CAPE variables. 60	
Table 4-2: R output of the stepwise logistic regression model for all CAPE variables.....	60
Table 4-3: SAS output of the stepwise logistic regression models for all LI variables. ....	62
Table 4-4: R output of stepwise logistic regression for all LI variables. ....	63
Table 4-5: SAS output of stepwise logistic regression for all PW variables.....	65
Table 4-6: R output of stepwise logistic regression for all PW variables. ....	65
Table 4-7: SAS output of stepwise logistic regression for all RH variables. ....	67
Table 4-8: R output of stepwise logistic regression for all RH variables.....	68
Table 4-9: SAS output of stepwise logistic regression for all $\Theta_e$ variables. ....	70
Table 4-10: R output of stepwise logistic regression for all $\Theta_e$ variables. ....	71
Table 4-11: SAS output of stepwise logistic regression for all temperature variables. ....	73
Table 4-12: R output of stepwise logistic regression for all temperature variables.....	74
Table 4-13: Summary of most appropriate parameters selected for the SON and DJF periods.....	75
Table 4-14: Output from the rare event logistic regression for SON, providing the regression coefficients (Coef), p-value (P), standard error on Coef (SE), Odds ratio (Odds), z-value (Z), maximum parameter value (MPV) and measure of parameter importance (MPI) for the intercept term and the six model parameters. ....	77
Table 4-15: Results from the new rare event logistic regression for the SON period with $\mu\text{CAPE}_{1-6\text{km}}$ removed from the analysis. ....	77

Table 4-16: Output from the rare event logistic regression for DJF.....	79
Table 5-1: Contingency table for all the probability thresholds of the UM LTI during SON of 2013.....	85
Table 5-2: Contingency table for all the probability thresholds of the UM LTI during DJF of 2013/14.....	87

---

# CHAPTER 1

## INTRODUCTION

---

### 1.1 BACKGROUND

At any given moment, there are approximately 2000 thunderstorms present around the world (Blumenthal et al., 2012). Severe thunderstorms are of paramount concern to the weather community as well as the public due to their ability to cause death, injury and damage (Lang et al., 2004). Lightning, tornadoes, strong wind, heavy rainfall and hail are some of the phenomenon associated with severe thunderstorms (Kohn et al., 2011). Lightning by itself is a phenomenon that can cause injury or death to humans and animals (Blumenthal et al., 2012), damage to infrastructures (Lynn and Yair, 2010), and can be a hazard to various sectors like the aviation and forestry industries (Price, 2013). It is estimated that lightning strikes the earth about 100 times every second (Ahrens, 2012; Blumenthal et al., 2012; Newcott, 1993). This equates to more than 8 million strikes a day.

Lightning is one of the leading causes of death from natural disasters, causing approximately 24,000 deaths annually around the globe as well as approximately 240,000 injuries (Blumenthal et al., 2012). The annual death rate due to lightning in South Africa is estimated to be between 1.5 (urban areas) and 8.8 (rural areas) per million of the population (Blumenthal et. al., 2012, Holle, 2008). These statistics are likely to be an underestimation of the actual death rate since lightning deaths are often not reported, especially in rural areas (Blumenthal et. al., 2012; Trengrove and Jandrell, 2011), and the pathology of damage to the human body due to lightning is not so well understood (Gill, 2008). Bhavika (2007) stated that the number of deaths in South Africa is about four times higher than the global average.

Severe damages to infrastructures can occur because of lightning. Just in the USA alone, the annual financial losses due to lightning damage are higher than \$1 billion (McCaul, et al., 2009). In South Africa, insurance claims due to electrical equipment failure or fire caused by lightning amounts to more than R 500 million annually (Gill, 2008; Evert and Schulze, 2005). The Climate Information section at the South African Weather Service (SAWS) received 2,103 queries from the insurance sector during the 2010/11 financial year, and most of these queries were lightning related. (Gijben, 2012; de Jager, 2011). During 1993 - 1999, about 30% of transmission line faults from Eskom, the power utility in South Africa, were due to lightning

(Bhavika, 2007; Bologna et. al., 2001). More recently, Peter and Mokhonoana (2010) reported that 24% of the transmission line faults from Eskom were due to lightning.

The reason why lightning is so dangerous is due to the extremely large electrical currents and heat generated in the lightning stroke (Price, 2013). Typical currents in a lightning stroke are 20 kA and the heat generated by a lightning stroke can reach temperatures of 30,000°C (Soul et al., 2002). The formation of lightning is related to the dynamical as well as microphysical processes found in clouds (Price, 2013), and will be presented in Chapter 2. A lightning stroke is the result of the neutralisation of the electrical charge in cloud causing the rapid expansion of the channel along which the charge flows. This causes a shock wave that produces the sound of thunder, while the flash is produced from the electrical discharge (Soul et al., 2002). Due to the hazardous nature of lightning, the need for prediction techniques to safeguard people and property is clear (McCaul et al. 2009; Lynn and Yair, 2010).

Forecasting methods exist to aid with the nowcasting (0-2 hours), very short-range (2-12 hours) and short-range (12-72 hours) forecasting of thunderstorms (Kohn et al. 2011, McCaul et al. 2009, Lynn & Yair, 2010) but remains a challenge. Past studies have utilised lightning data from lightning detection networks (LDN), parameters from atmospheric soundings and numerical weather prediction (NWP) models to aid with lightning forecasts.

LDN are capable of detecting lightning strokes in real-time and the data measured by these networks could be utilised to aid in the nowcasting of thunderstorms. The data becomes available within seconds permitting the continuous monitoring of the most intense parts of storms. The nearly instantaneous data availability enables the monitoring of thunderstorms in those periods when data from other remote sensing instruments, like radar and satellite, are still being processed (Finke & Kreyer, 2002). Many of these LDN networks however are designed to detect only cloud-to-ground (CG) lightning (lightning that strikes the earth), and it has been shown that inconsistent relationships exist between CG lightning trends and thunderstorm nowcasting (Schultz et al, 2011). This means that CG lightning might be useful to monitor thunderstorms in real-time but have a much lower capability to aid with the nowcasting of thunderstorms with sufficient lead-times (minutes to hours ahead). In the past few years Very High Frequency (VHF) total lightning sensors, which detects lightning inside the cloud as well CG lightning, has become more readily available (Schultz et al, 2011). These sensors have been found to be useful in the nowcasting of CG lightning strikes since lightning in the clouds mostly precedes CG lightning on the ground. MacGorman et al. (2011) showed that cloud lightning can precede the first CG lightning flash by up to an hour. As a result, LDN which measures all lightning (cloud + CG) can be used to forecast the occurrence of CG lightning with a lead time of up to an hour since the cloud lightning occurs before the CG

lightning. LDN that only measures CG lightning is more useful for the monitoring of thunderstorms and lightning verification studies. Both types of LDN are however not designed to forecast the threat of lightning for the very short-range and short-range forecasting time scales but rather to monitor lightning occurrence.

Statistical techniques have been used extensively to predict thunderstorms and lightning (Shafer and Fuelberg, 2008). These prediction schemes identify the threat of lightning by relying on connections between lightning occurrence and parameters of the pre-storm environment (Rajeevan et al., 2012; McCaul et al., 2009). These parameters are often derived from information of atmospheric soundings (Shafer and Fuelberg, 2008), which are typically performed only twice daily and at a limited number of locations. This limits the spatial and temporal representations of forecasts made with these accents. (de Coning et al., 2011).

NWP models nowadays offer an alternative for thunderstorm predictions. They are capable of providing accurate forecasts, mimicking atmospheric soundings, with high spatial and temporal resolutions (Shafer and Fuelberg, 2008). Many lightning prediction techniques also now rely on NWP models. The advantage of these methods are that lightning can be predicted on both the very short-range as well as short-range forecast time scales due to NWP models being capable of providing accurate parameters related to lightning formation for several hours ahead (McCaul et al., 2009).

## 1.2 MOTIVATION

Lightning is a common occurrence in South Africa. The summer rainfall regions, particularly the Highveld (region over the interior of South Africa with altitudes exceeding ~1500 m above sea level) and north-eastern parts of the country, receive most of its precipitation through convective storms. Each year, these storms produce large amounts of lightning which results in numerous deaths, injuries and damages (Gijben, 2012; Blumenthal et al., 2012; Gill, 2008). In order to facilitate early warning of lightning there is a need to develop methods that can aid operational weather forecasters in issuing timeous lightning warnings.

To the author's knowledge, no product currently exists at SAWS or in South Africa that attempts to forecast the potential lightning threat directly. Different types of instability indices exist and are used to aid with the prediction of thunderstorms. Although lightning is a by-product of all thunderstorms and consequently these indices can be used in an indirect manner to gain insight into lightning formation, no product exist that attempts to specifically assess the risk of lightning over South Africa in short range weather forecasts.

The atmospheric variables associated with lightning are well known, and parameters available from NWP models are used to aid with assessing the possible lightning risk (Reap, 1994; Burrows et al., 2005; Bothwell, 2008; Shafer and Fuelberg, 2008; Rajeevan et al., 2012). The work done by Frisbie et al. (2009) that motivated this study combined NWP parameters into an index, called the Lightning Potential Index (LPI), that provides an outlook map of the possible lightning risk. The exact same model parameters and methodology used by Frisbie et al. (2009) was applied over South Africa and showed the potential to provide lightning forecasts for the country (Gijben, 2013). However, the geographical location and height above sea level of South Africa requires the development of a bespoke Lightning Threat Index (LTI). In this dissertation, an LTI is proposed for South Africa. The NWP parameters that are able to predict the occurrence of lightning over South Africa are identified first and then optimised into a new lightning prediction index designed specifically for South Africa.

### **1.3 AIMS AND OBJECTIVES**

There are three main aims in this study:

1. Identify the NWP parameters that are best suited to predict lightning over South Africa.
2. Integrate the selected parameters into an LTI over South Africa for spring and summer.
3. Evaluate the skill of the LTI over South Africa.

To achieve these aims the following objectives will be applied:

1. Identify numerous candidate NWP model parameters from the literature that have previously been found useful for lightning prediction.
2. Make use of statistical techniques to identify the model parameters best capable of predicting the occurrence of lightning from the list of candidate predictors.
3. Utilise the selected model parameters to develop a new LTI by means of rare-event logistic regression techniques.
4. Evaluate the predictive capability of the LTI against observations of lightning by means of standard forecast verification techniques.
5. Compare the newly developed LTI against the standard convective precipitation forecast from the Unified Model (UM) as well as against the Frisbie et al. (2009)



methodology of the LPI developed in the United States of America (USA). This will be achieved by means of case studies in order to determine if the new index adds value to lightning forecasts over South Africa.

## **1.4 DISSERTATION OUTLINE**

In Chapter 2, an overview of lightning is presented. This includes a discussion of the processes involved in the formation of lightning, a summary of lightning detection systems, the distribution of lightning over South Africa and the globe, and lightning prediction techniques.

Chapter 3 deals with the data and methodologies used in this study, and includes a discussion of the study area and period, the NWP and lightning data, the methods utilised to select the most appropriate model parameters for lightning prediction, how the new LTI was developed, and finally the verification techniques considered.

Chapter 4 covers the development of the new LTI for South Africa. This chapter shows the results from the statistical techniques utilised to select the most appropriate model parameters for lightning prediction, followed by the results of the development of the LTI.

In Chapter 5, the LTI is quantitatively evaluated against observations of lightning over an entire spring and summer season. A qualitative comparison between the LTI, the UM convective precipitation forecast, the Frisbie et al. (2009) LPI developed in the USA and observations of lightning is presented by means of six case studies.

Chapter 6 provides the final summary, conclusions, contribution and recommendations of the material covered in this study.

# CHAPTER 2

## LIGHTNING FORMATION, DETECTION AND PREDICTION

---

### 2.1 INTRODUCTION

In this chapter, the formation of lightning is presented. The discussion begins with an overview of the typical electrical structure of a thundercloud, the processes involved in the electrification of a thunderstorm (of which the ice-graupel mechanism is thought to be the most likely mechanism), and finally an explanation of the lightning discharge which acts to neutralise the charge in a thundercloud.

The discussion on the formation of lightning is followed by a summary of lightning detection systems used to detect electromagnetic waves emitted by lightning discharges. These systems make it possible to provide information on the distribution of lightning around the globe and over South Africa.

The chapter ends with a discussion on lightning prediction techniques. NWP model parameters are discussed as well as their usefulness for lightning prediction. The NWP parameters discussed are; Convective Available Potential Energy (CAPE), Lifted Index (LI), Precipitable Water (PW), Relative Humidity (RH), Equivalent Potential Temperature ( $\Theta_e$ ), and air temperature. The last section provides an overview of the LPI presented by Frisbie et al. (2009), which inspired the work in this thesis.

### 2.2 LIGHTNING FORMATION

#### 2.2.1 Electrical structure of a thundercloud

Wilson (1916) was the first to assume a dipole structure in a thundercloud. He concluded, by using ground-based observations, that positive charge regions are typically found above negative charge regions. This result was confirmed by extensive measurements of changes in electrical fields in New Mexico thunderstorms (Krehbiel et al., 1979). From observations, we now know that the charge structures in thunderclouds can be more complicated than a simple dipole, but most thunderclouds exhibit this feature where a main

positive charge region is found at the top of the cloud while a main negative charge region is found lower down (Lang et al., 2004). A small region of positive charge is also commonly observed near the base of the thundercloud (Krehbiel, 1986). When considering this region of lower positive charge, the thundercloud has a tripole structure (Lang et al., 2004). Studies have shown that the main positive charge region in a thundercloud is typically found at heights where the temperatures are lower than  $-20^{\circ}\text{C}$ . The main negative charge region usually occurs between the  $-10^{\circ}\text{C}$  and  $-20^{\circ}\text{C}$  temperature level, and the small region of positive charge is found close to the  $0^{\circ}\text{C}$  level (Simpson and Scrase, 1937; Simpson and Robinson, 1941; Takahashi, 1978; Lang et al., 2004; Krehbiel, 1986; Koshak and Krider, 1989; Stolzenberg et al. 1998a,b). Malan (1963) investigated thunderstorms in South Africa and produced Figure 2-1 that shows the most likely distribution of charge in a thundercloud.

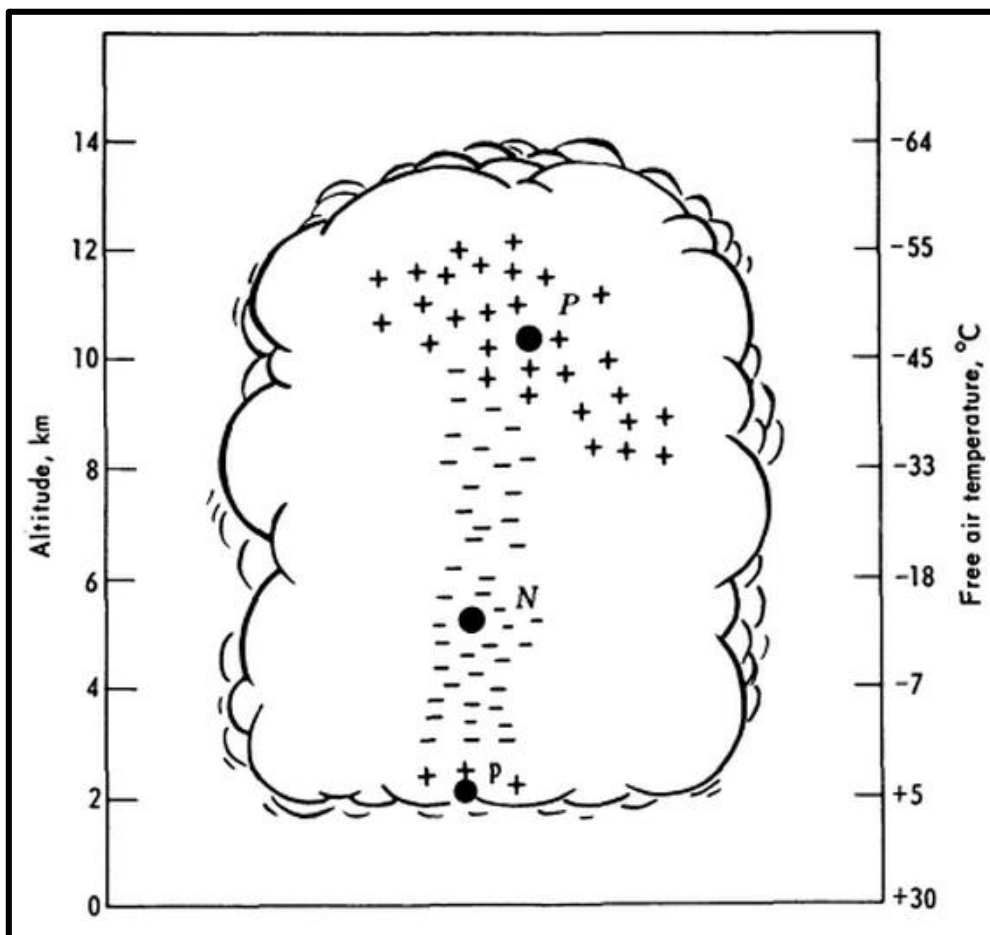


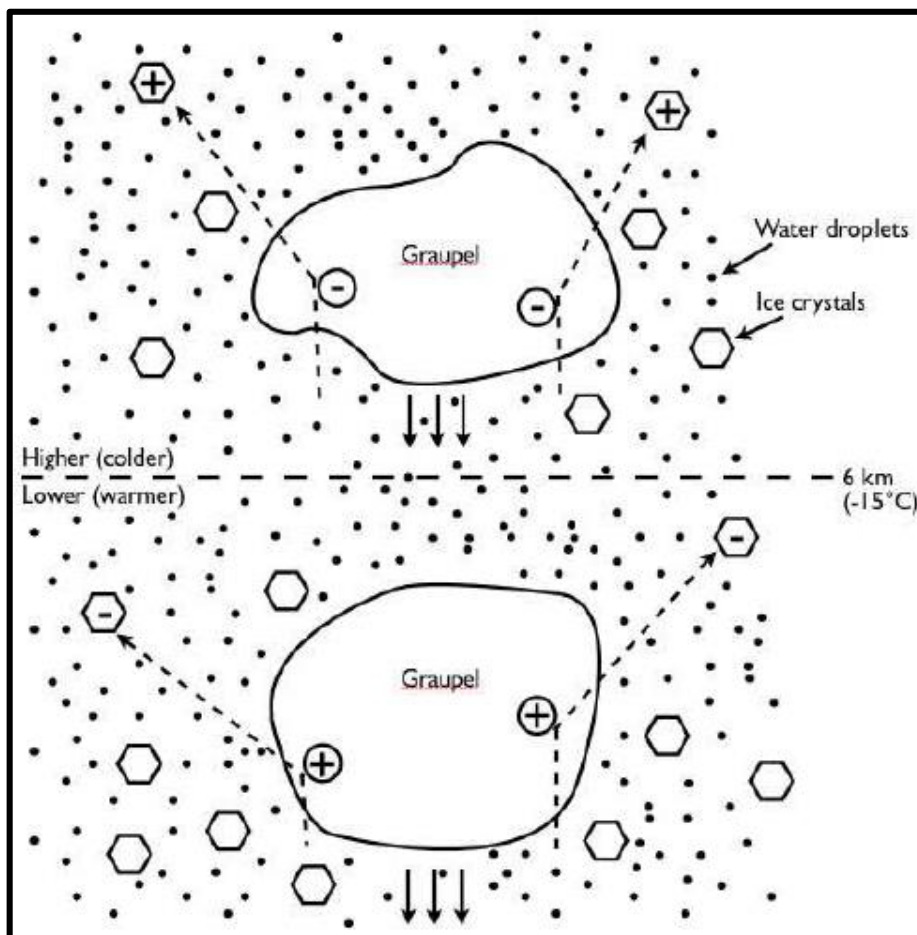
Figure 2-1: Most likely charge structure of a South African thundercloud [Cited from Uman (2012), Source: Malan (1963)]

## 2.2.2 Electrification of a thundercloud

The electrification of a thundercloud involves a process that electrifies individual cloud-particles as well as a process that separates these cloud-particles according to their polarity in different levels of the thundercloud (Rakov and Uman, 2003). Many theories have been proposed to explain the charge separation mechanisms of a thundercloud (Takahashi, 1984). Saunders (2008) provides a detailed description on possible mechanisms of thundercloud electrification that includes, drop break-up, ion charging, the convection mechanism, inductive charging, ice particle mechanism and ice-crystal/graupel mechanism. There is a growing amount of evidence that the non-inductive collisional graupel-ice mechanism is the main cause of electrification in a thundercloud (Rakov and Uman, 2003). This mechanism is called “non-inductive” since the ambient electric field is not required in the electrification of the cloud-particles (Saunders, 2008; Rakov and Uman, 2003). The non-inductive charge-transfer model is considered the most likely cause of electrification in a thundercloud, thus only this graupel-ice mechanism will be discussed.

In the non-inductive graupel-ice mechanism, electrification of a thundercloud occurs because of collisions between precipitation particles (graupel) and cloud particles (ice-crystals). Precipitation particles are larger particles with fall speeds greater than  $0.3 \text{ ms}^{-1}$  while cloud particles are smaller particles with lower fall speeds (Rakov and Uman, 2003). When graupel and ice-crystals collide, charge is transferred between them without any influence from the local electric field (non-inductive) (Saunders, 1993). Laboratory studies have shown that the collisions between graupel and ice particles should occur in the presence of supercooled water droplets in order for significant charge transfer to occur (Rakov and Uman, 2003; Reynolds et al., 1957; Takahashi, 1978). Figure 2-2 shows how charge is transferred. The larger and heavier graupel particles fall through ice-crystals and supercooled water droplets which are suspended in the cloud. The water droplets remain supercooled until they collide with the graupel or ice-crystals from where they freeze onto the surface in a process called riming. When the falling graupel particles collide with the suspended ice-crystals, charge is transferred between them. The sign and magnitude of charge transfer seems to be dependent on; the temperature where these collisions occur, cloud water content, rate of rime accretion, and droplet sizes. The temperature at which a reversal of the sign of the charge on graupel particles occur is called the reversal temperature. At heights above the reversal temperature (colder temperatures), collisions between the graupel and ice-crystals results in the graupel particles becoming negatively charged while the ice-crystals becomes positively charged. This can be due to the subliming graupel surfaces. When the temperature is below the reversal temperature (lower heights), graupel particles become positively charged while

the ice-crystals become negatively charged and can be due to the graupel particles growing by vapour diffusion. The exact cause of the reversal of charge is still debateable (Rakov and Uman, 2003; Saunders, 1993; Mansell et al., 2005). Studies have shown that reversal temperature usually occurs between  $-10^{\circ}\text{C}$  and  $-20^{\circ}\text{C}$  (Rakov and Uman, 2003; Saunders, 1993; Mansell et al., 2005; Church, 1966; Takahashi, 1978; Gaskell and Illingworth, 1980). In Figure 2-2, the reversal temperature is assumed to be at  $-15^{\circ}\text{C}$ . Jayratne et al. (1983) showed that the sign of charge reversal is dependent on the reversal temperature and that this reversal temperature height is dependent on the amount of liquid water content in a cloud. What is important to note is that the most commonly observed height of the reversal temperature (between  $-10^{\circ}\text{C}$  and  $-20^{\circ}\text{C}$ ) corresponds with the height of the main negatively charged region of a thundercloud (Rakov and Uman, 2003; Saunders et al., 2006).



**Figure 2-2: Cloud electrification by means of charge transfer because of collisions between graupel and ice particles in the presence of water droplets. The assumption is made that the charge reversal temperature is  $-15^{\circ}\text{C}$  and occurs at a height of 6 km. [Source: Rakov and Uman, 2003].**

### 2.2.3 Electrical Discharge

Lightning is an electrical discharge that occurs between oppositely charged regions in the cloud when the accumulation of charge builds up sufficiently (Krehbiel, 1986). The primary forms of discharges are CG, intra-cloud (between oppositely charged regions of the same cloud), inter-cloud (cloud-to-cloud) and air discharge (cloud-to-air) (Figure 2-3). The three types of cloud lightning (intra-cloud, inter-cloud and air discharge) are usually collectively referred to as intra-cloud (IC) lightning. CG lightning is electrical discharges that transfers charge to the ground (Rakov and Uman, 2003). This is the most familiar type of lightning since it strikes the earth and affects humans and infrastructures. CG lightning mostly occurs between the main negatively charged region of a cloud and the ground (Krehbiel, 1986). Approximately 90% of all CG lightning flashes transfer negative charge from the cloud down to the ground, and is called downward negative lightning. The remaining 10% or less, of CG lightning flashes, lowers positive charge to the ground and is called downward positive lightning. Upward positive and negative lightning also occurs but only from tall structures or structures on mountains (Rakov and Uman, 2003).

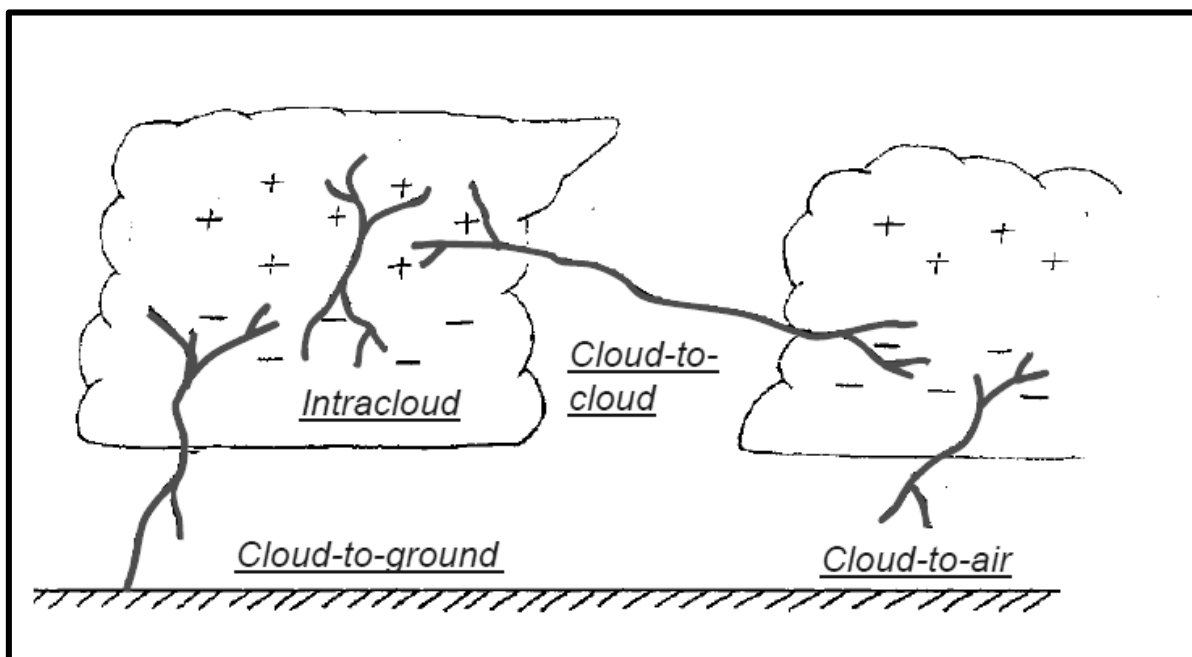


Figure 2-3: The main types of lightning (cloud-to-ground, intra-cloud, cloud-to-cloud, and cloud-to-air) [Adapted from Jaffer et. al. (2011), Source: Rakov and Uman (2003)]

Figure 2-4 shows the typical sequence of events in a downward moving, negatively charged CG lightning flash and is best described by Rakov and Uman (2003):

The sequence starts with the distribution of charges in the cloud according to the main positive (P), main negative (N), and lower positive (LP) regions. From here, a process called the initial breakdown occurs inside the cloud prior to the occurrence of the stepped-leader. There is no agreed upon explanation for the initial breakdown, but it provides conditions favourable for the formation of the stepped-leader. After the initial breakdown, the stepped-leader forms. The stepped-leader is a negatively charged plasma channel extending from the cloud to the ground. It descends towards the ground in a series of steps. Each step takes about  $1\mu\text{s}$ . The stepped-leader serves to create a conducting path between the cloud and the ground, removes negative charge from the cloud, and transfers it downwards along the stepped-leader. As the downward moving stepped-leader approaches the ground, the electric field on the ground starts to increase. When the electric field on the ground exceeds a certain critical value, one or more upward-connecting leaders, called streamers, develop. The development of these streamers, because of the descending return-stroke, indicates the start of the attachment process. The attachment process ends when the downward moving stepped-leader connects with an upward moving streamer. When the two leaders connect, the first return stroke is initiated. A fully formed return-stroke occurs which neutralises the cloud charge stored in the stepped-leader by transporting the negative charge in the stepped-leader to the ground. The return stroke rapidly heats up the channel to  $\pm 30\,000^\circ\text{C}$ , resulting in the channel to expand to  $\pm 10$  atmospheres, due to the high-current wave. This causes rapid expansion, optical radiation and a shockwave, which is the lightning we hear and see. In some cases, the lightning flash may end after the first return-stroke, but in most cases, a dart leader moves down the same channel from the first return-stroke. After the first return-stroke and before the initiation of this dart leader, J and K processes occur. J-processes are the redistribution of charges in the cloud after the return-stroke and are a slow moving leader from the origin of the flash into the negatively charged region. K-processes last for only a short period during the slower J-processes and are a streamer that starts at the tip of the positive leader and propagates to the origin of the flash. K-processes are often described as attempted dart-leaders. After the J and K processes, a dart leader moves down the channel from the preceding return-stroke. As the dart leader gets close to the ground, an attachment process similar to that of the first return-stroke takes place. When the dart leader connects to the ground, the second return-stroke takes place.

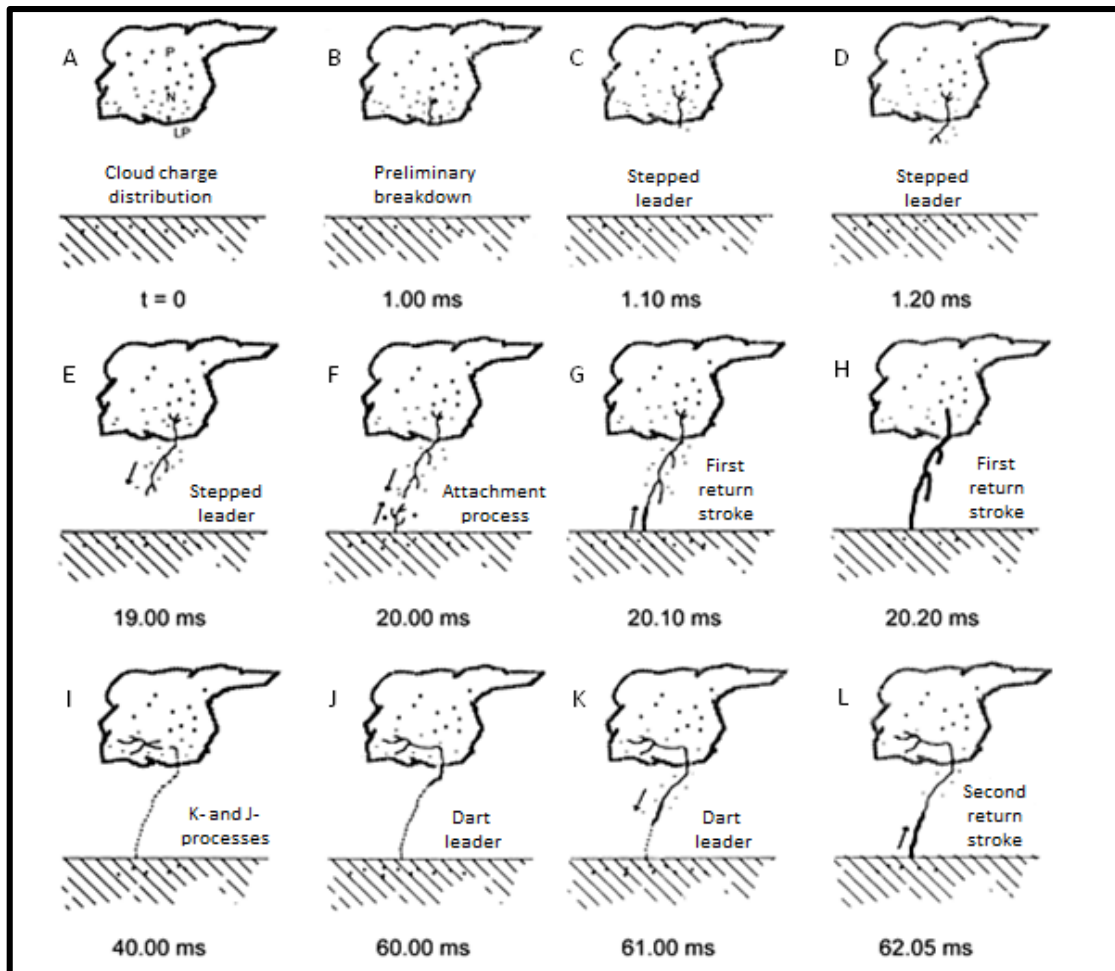


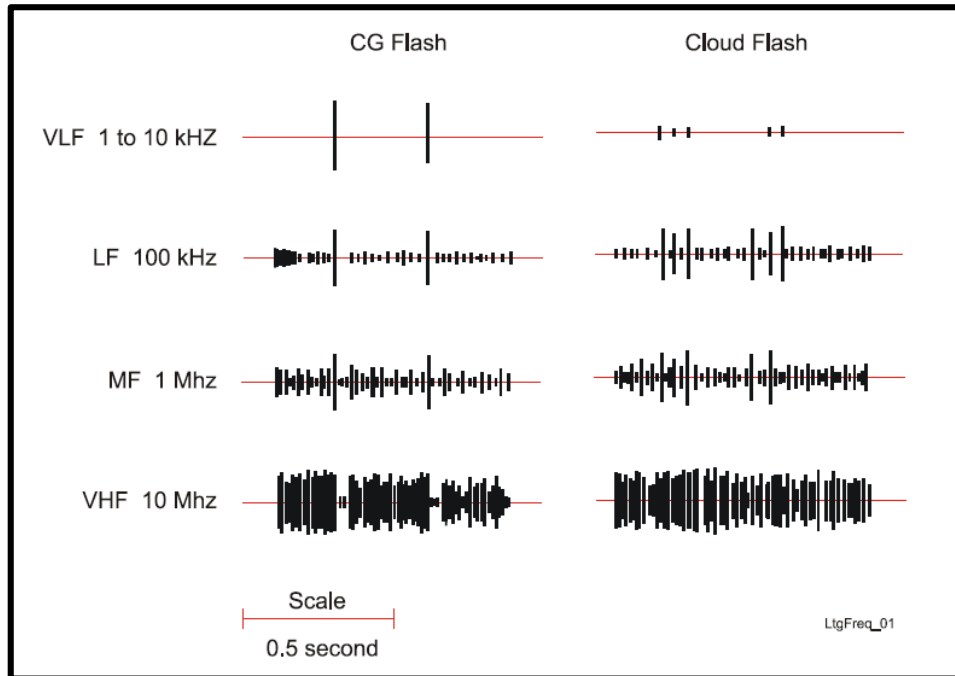
Figure 2-4: The processes involved in a typical negative cloud-to-ground lightning flash. P indicates the main positive charge region, N the main negative charge region and LP the lower positive charge region [Cited from Rakov and Uman (2003), Source: Uman (1987; 2001)]

### 2.3 LIGHTNING DETECTION

Lightning can be detected in real-time and over great distances due to lightning flashes producing electromagnetic radiation that propagate across the globe (Price, 2008). Electromagnetic radiation from lightning is emitted in a wide range of frequencies, from below 1 Hz up to almost 300 MHz (Rakov and Uman, 2003). Higher frequency waves attenuate much quicker than lower frequency waves, and as a result, the lower frequency waves can be detected at much greater distances (Price, 2008). Lightning discharges predominantly emits electromagnetic waves in the low frequency (LF)/very low frequency (VLF) (3 – 300 kHz) up to VHF (3 – 300 MHz) frequency ranges (Betz et al., 2008). CG lightning produces a small amount of large amplitude electromagnetic pulses in the VLF range, while IC lightning mostly produces a couple of small amplitude pulses in the VLF range (Figure 2-5). CG and IC lightning



produces many pulses in the VHF range and both with similar amplitude (Cummins et al. 2000).



**Figure 2-5: Electromagnetic pulse signatures from lightning flashes in the different frequency ranges [Cited from Vaisala (2004), Source: Malan (1963)]**

Most lightning locating systems that detect electromagnetic radiation make use of the time of arrival (TOA) and magnetic direction finding (MDF) methods (Rakov and Uman, 2003). For a detailed discussion of these techniques, the reader is referred to Gill (2008) and Vaisala (2004). Some sensors make use of a combination of TOA and MDF techniques to detect lightning, which provides more accuracy (Price, 2008)

Most regional LDN's utilise ground-based lightning sensors, which detect electromagnetic pulses in the VLF/LF range, since they can detect lightning over greater distances. These networks however detect mostly CG lightning and only a small fraction of IC lightning (Price, 2008). Some examples of regional LDN's include the National Lightning Detection Network in the USA (Orville et al., 2011) and the Southern African Lightning Detection Network (SALDN) in South Africa (Gijben, 2012). Regions typically operate networks of sensors spaced relatively close to each other (e.g. the SALDN consists of 24 sensors over South Africa). A number of global LDN's also exist. Examples include, the World Wide Lightning Location Network (Virts et al., 2013) and the Global Lightning Dataset

(GLD360) (Mallick, 2014). These sensors operate in the VLF frequency and consist of sensors separated by large distances. The sensors mostly detect CG lightning, with large amplitude pulses and with lower detection efficiencies due to the increased distances between sensors (Rudlosky, 2014).

Lightning sensors, which detects pulses in the VHF spectrum, can detect both CG and IC lightning with great accuracy but at small spatial scales (Price, 2008). Total lightning sensors, which detect CG and IC lightning, often make use of VHF waves. Examples include the Vaisala total lightning sensor, which also utilises VHF pulses to detect IC lightning (Murphy et al., 2013) and Lightning Mapping Arrays (LMA) such as the north Alabama Mapping Array (Koshak et al., 2004). These LMA's consists of sensors grouped closely together (tens of kilometres) and can detect lightning in both two and three dimensions (Price, 2008).

Lightning can also be detected from space by means of earth-orbiting satellites which either detects the light or electromagnetic waves from lightning discharges (Rakov and Uman, 2003). The most famous lightning detectors on satellites include the National Aeronautics and Space Administration (NASA) Optical Transient Detector (OTD) and the Lightning Imaging Sensor (LIS) on board the Tropical Rainfall Measuring Mission (TRMM) satellite (Christian et al., 2003).

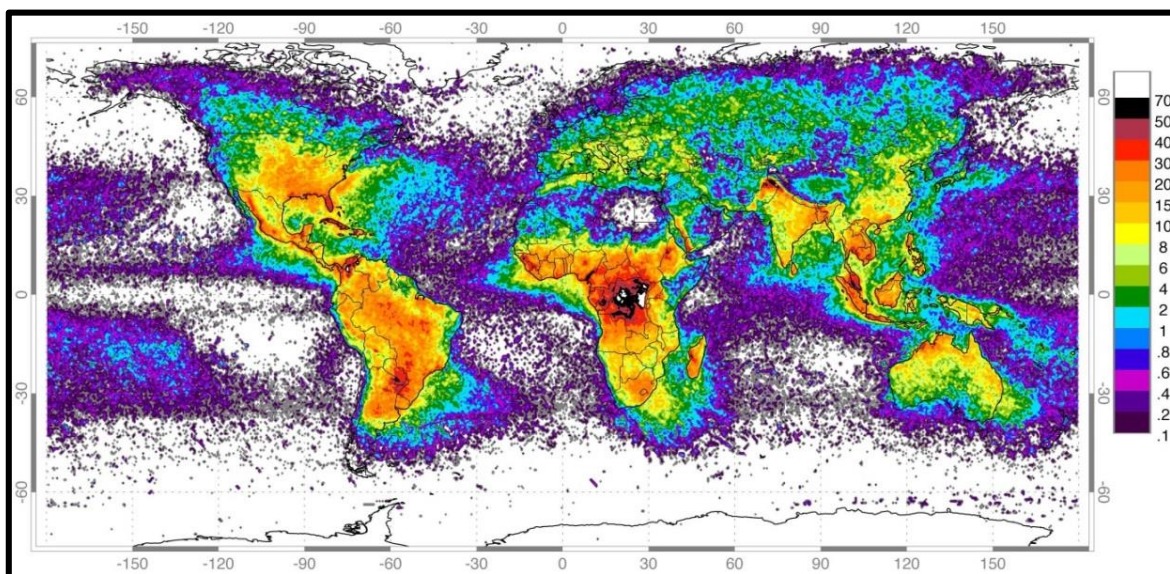
## **2.4 DISTRIBUTION OF LIGHTNING**

### **2.4.1 Globally**

Lightning occurrence varies across the globe. It is influenced by the atmospheric circulation patterns as well as the geographical locations of the continents and oceans (Price, 2013; Price, 2006). Solar heating and water vapour is particularly useful to explain the distribution of lightning across the globe, since water vapour concentrations can increase with an increase in temperature, and water vapour plays an important role in the development of thunderstorms (Price, 2013). Approximately 78% of the global lightning occurs in the tropical regions between latitudes 30S and 30N (Christian et al., 2003) where the temperatures and water vapour concentrations are higher compared to extra-tropical and polar regions. Christian et al. (2003) also showed that the vast majority of lightning occurs over the continents.

Figure 2-6 shows a map of the average annual lightning flash densities (flashes  $\text{km}^{-2} \text{yr}^{-1}$ ) across the globe and is produced by observations from the NASA OTD and TRMM LIS instruments (Information Technology and Systems Centre, University of Alabama, 2015). The instruments on board these satellites measure all types of lightning (CG + IC) (Thompson et

al., 2014). The area that stands out the most is the Congo in central Africa, where flash densities exceed  $70 \text{ flashes km}^{-2} \text{ yr}^{-1}$  in certain areas. Christian et al. (2003) stated that in his analysis of the LIS observations, flash densities over the Congo basin exceed  $30 \text{ flashes km}^{-2} \text{ yr}^{-1}$  over an area larger than 3 million  $\text{km}^2$  and that approximately 90% of lightning occurs over the continental regions while about 10% occurs over the oceans. The largest flash densities are seen in the tropics. The central interior of South Africa record flash densities between 20 and 30  $\text{flashes km}^{-2} \text{ yr}^{-1}$  and is comparable with the lightning hotspot areas of North and South America, Asia and Northern Australia, but is much lower than over Equatorial Africa.

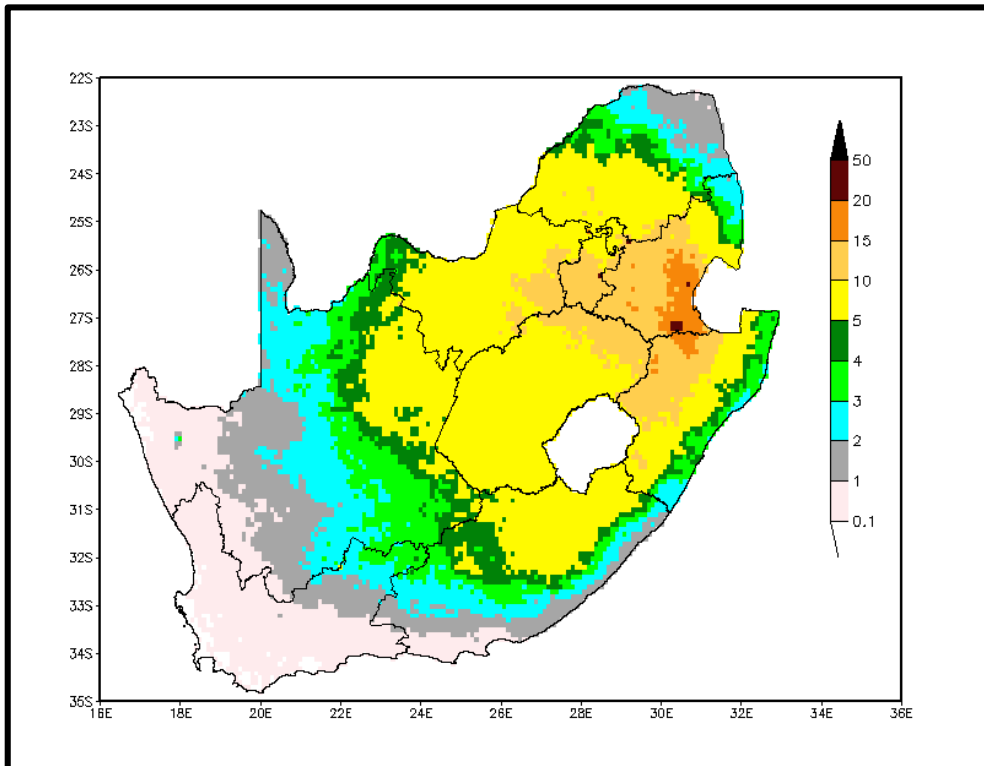


**Figure 2-6: Global average annual lightning flash density on a  $0.5^\circ \times 0.5^\circ$  resolution from the combined observations of the NASA OTD and TRMM LIS instruments [Source: Information Technology and Systems Centre, University of Alabama, 2015]**

## 2.4.2 South Africa

Lightning occurrence in South Africa is measured by the SALDN operated by the SAWS. This network detects only CG lightning, but is able to pinpoint the locations of lightning that strikes the earth with great accuracy. A detailed description of this network is provided in Chapter 3. Gill (2008) developed an initial lightning climatology of South Africa with the 2006 lightning data from the SALDN. This was updated using data for 5-years (2006-2010) by Gijben (2012). The lightning climatology has consequently been updated annually. Over South Africa, the highest flash densities occur along the eastern escarpment of the country with values exceeding  $15 \text{ flashes km}^{-2} \text{ yr}^{-1}$  over most areas and values higher than  $20 \text{ flashes km}^{-2} \text{ yr}^{-1}$  over some areas.

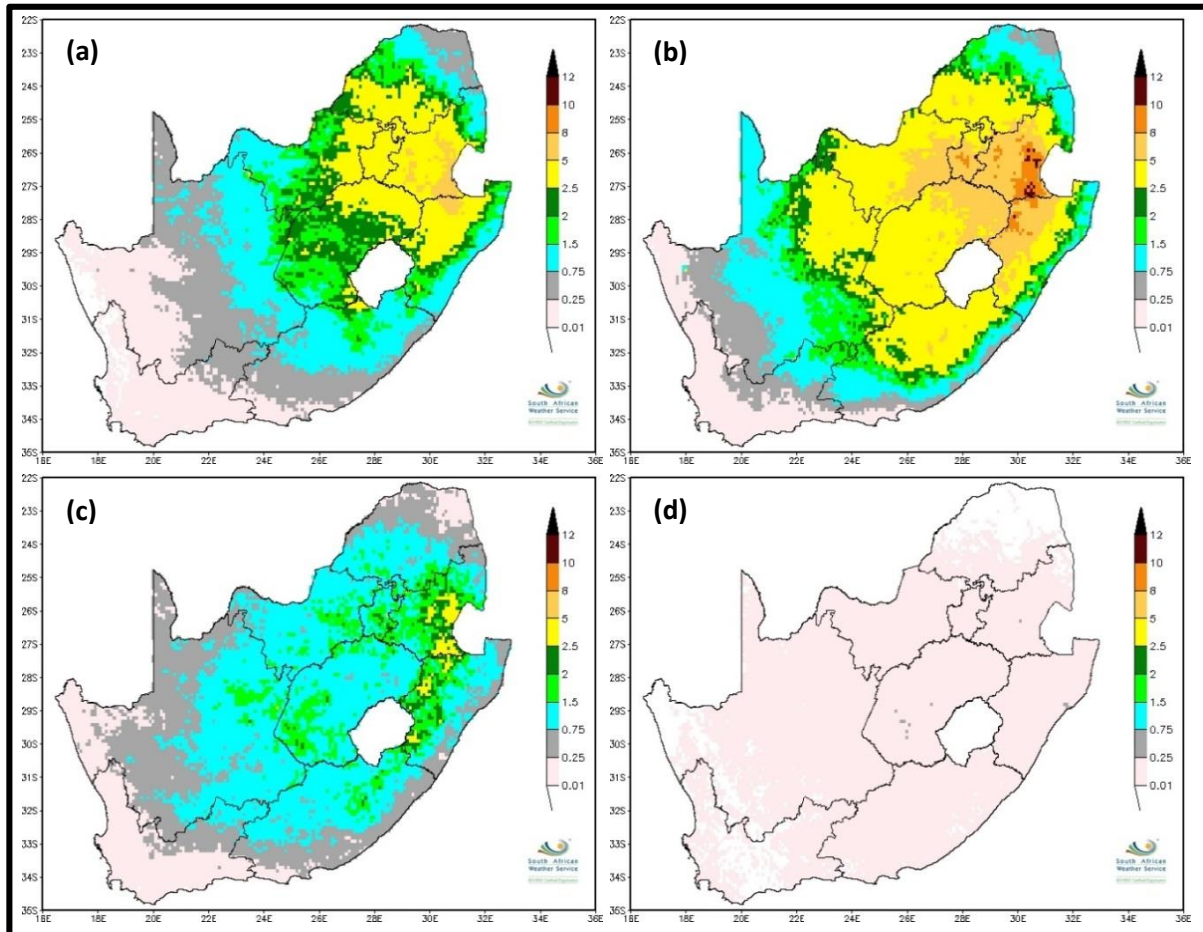
$\text{yr}^{-1}$  in a few isolated areas. A large section of the Highveld region sees flash densities exceeding  $10 \text{ flashes km}^{-2} \text{ yr}^{-1}$ , while most of the central interior of the country experience more than  $5 \text{ flashes km}^{-2} \text{ yr}^{-1}$ . Flash densities decrease towards the west, the northern and north-eastern parts, as well as to the east and south coasts of the country.



**Figure 2-7: The average annual lightning ground flash densities ( $\text{flashes km}^{-2} \text{ yr}^{-1}$ ) for the 9-year period from 2006 to 2014 over South Africa [Source: South African Weather Service]**

South Africa consists mainly of two rainfall seasons, the summer and winter rainfall seasons (Taljaard, 1996). Most of the central to northern interior of the country falls within the summer rainfall region and receive most of its rainfall from convective thunderstorms (de Coning and Poolman, 2011; Kruger, 2007; Landman et al., 2012; Tyson, 1986; Dyson et al., 2015). During the months October to December (early summer), the weather systems are extra-tropical in nature while during January to March (late summer) the circulation is tropical in nature (Dyson et al., 2015). During late summer, tropical rain showers are more common than heat generated storms. For the winter rainfall regions of the south-western and coastal parts of the country, most rainfall is stratiform in nature and results from ridging high-pressure systems and cold fronts (de Coning and Poolman, 2011). Over South Africa, most lightning occurs during the spring/summer months of September to November (SON) and December

to February (DJF) (Figure 2-8a&b). Lower and very low amounts of lightning are detected during the autumn/winter months between March to May and June to August (Figure 2-8c&d). This is due to the convective nature of storms for the summer rainfall regions and the stratiform nature of rainfall over the winter rainfall region.



**Figure 2-8: The average annual lightning ground flash densities for (a) Spring (September, October, November), (b) Summer (December, January, February), (c) Autumn (March, April, May) and (d) Winter (June, July, August) over South Africa from 2006-2014 [Source: South African Weather Service].**

## 2.5 PREDICTION OF LIGHTNING

Forecasting thunderstorms remains a challenge due to their small spatial and temporal scales as well as variability in the processes governing thunderstorm development (Rajeevan et al., 2012). Thunderstorms require three key ingredients: instability, moisture and a trigger mechanism. Many thermodynamic and kinematic techniques have been developed to predict these ingredients (Kunz, 2007). To predict lightning in a thunderstorm remains an even bigger

challenge since the processes that govern the electrification of a thundercloud are still poorly understood (Shafer and Fuelberg, 2008). Many techniques have been developed to forecast lightning, ranging from the nowcasting up to short-range forecasting scales.

### **2.5.1 Remote Sensing Instruments**

To address the nowcasting of lightning, techniques have been developed to forecast lightning initiation by means of weather radar (e.g. Woodard et al., 2012), and sophisticated LDN's (e.g. Kohn et al., 2011). Statistical techniques have been used extensively to predict thunderstorms and lightning, with multiple linear regression and binary logistic regression techniques being the most common (Shafer and Fuelberg, 2008). These prediction schemes forecast the threat of lightning by relying on connections between lightning occurrence and parameters of the pre-storm environment (Rajeevan et al., 2012; McCaul et al., 2009). Many examples of such lightning prediction schemes exist (Livingston et al., 1996; Mazany et al., 2002; Benson, 2005; Lambert et al., 2005; Shafer and Fuelberg, 2006). Parameters are often derived from atmospheric soundings to predict lightning (Shafer and Fuelberg, 2008), however soundings are typically only performed twice daily and at a limited amount of locations (de Coning et al., 2011). As a result, morning soundings are typically used to predict thunderstorms or lightning later in the day, which may result in inaccurate forecasts due to changes in atmospheric conditions later in the day or the site-specific sounding not being able to represent a large forecast domain (Shafer and Fuelberg, 2008).

### **2.5.2 Numerical Weather Prediction Models**

NWP models offer an opportunity for lightning prediction since they are capable of providing sounding forecasts with high spatial and temporal resolution (Shafer and Fuelberg, 2008). Many institutions that predict lightning are moving away from empirical lightning prediction methods to NWP model based techniques (Goodman, 2012). The advantage of these methods are that lightning can be predicted on both the very short-range as well as short-range forecast scales due to NWP models being capable of providing accurate parameters related to lightning formation for several hours ahead (McCaul et al., 2009). Statistical prediction schemes that forecast the threat of lightning by relying on connections between lightning occurrence and parameters of the pre-storm environment has also been developed by making use of NWP data (Reap, 1994 ; Burrows et al., 2005 ; Bothwell, 2008 ; Shafer and Fuelberg, 2008 ; Rajeevan et al., 2012)

Many of the statistical lightning prediction techniques make use of thermodynamic parameters to forecast lightning (Lynn et al., 2012 ; Lynn and Yair, 2010) and typically perform quite well in identifying the lightning risk but they have a much lower capability to forecast the amount of lightning (Soul et al., 2002 ; McCaul et al., 2009). NWP models are now capable of forecasting various microphysical parameters related to charge separation in thunderstorms, like the mixing ratio of hydrometeors, which is correlated with lightning flash rates (McCaul et al., 2009). As a result, lightning prediction techniques are starting to focus on utilising the NWP microphysical parameters to improve the skill of current lightning forecasting techniques as well as to predict the amount of lightning that will occur. The capability of NWP models to resolve microphysical parameters is a new development and many older NWP models operated by some institutions currently do not have these capabilities. Due to the limitation where microphysical parameters are not available in some older NWP models, attempts to forecast the amount of lightning should not be considered with the available thermodynamic parameters from NWP models, however useful guidance forecasts about the location of where lightning may occur can be successfully developed (Shafer and Fuelberg, 2005).

### **2.5.3 Parameters useful for lightning prediction**

A large number of parameters have been found to be useful predictors to forecast the occurrence of lightning. Numerous studies utilised CAPE, LI,  $\Theta_e$ , PW, RH, and air temperature, in one form or the other, as predictors to forecast lightning (Livingston et al. 1996; Burrows et al., 2005; Lambert et al., 2005; Shafer and Fuelberg, 2006; Shafer and Fuelberg, 2008; Frisbie et al., 2009; Rajeevan et al., 2012; Zepka, 2014). A description of these parameters follows below:

#### **2.5.3.1 Convective Available Potential Energy**

CAPE is the total amount of energy available to an air parcel when lifted from the level of free convection (LFC) to the level of neutral buoyancy (LNB), and is calculated by a vertical integration of the parcel and environment virtual temperatures between these levels (Doswell and Rasmussen, 1994; Blanchard, 1998; Qie et al., 2003). CAPE can be expressed by Equation 2-1:

$$CAPE = g \int_{LFC}^{LNB} \left( \frac{T_{vp} - T_{ve}}{T_{ve}} \right) dz \quad (2-1)$$

- Where:
- $T_{vp}$  - Virtual temperature of the parcel
  - $T_{ve}$  - Virtual temperature of the environment
  - $LFC$  - Level of free convection
  - $LNB$  - Level of neutral buoyancy
  - $g$  - Acceleration due to gravity

The virtual temperature is defined as the temperature that dry air must have in order to have the same density as moist air at the same pressure (Stull, 2012). Many computations of CAPE makes use of normal temperatures, but Doswell and Rasmussen (1994) showed the importance of using virtual temperatures in the calculations of CAPE in order to reduce errors, especially for lower values of CAPE. The LFC is the level at which the temperature of a rising parcel starts to exceed the temperature of the environment and becomes unstable, while the LNB is defined as the level where the temperature of the environment again starts to exceed the temperature of the parcel and becomes stable (Blanchard, 1998).

CAPE is an important parameter used to predict thunderstorm development (Jayaratne and Kuleshov, 2006; Qie et al., 2003), and is a measure of the conditional instability in the atmosphere (Dyson et al., 2015). Conditional instability is defined as when the temperature lapse rate of a column of air is lower than the dry-adiabatic lapse rate but higher than the moist-adiabatic lapse rate (Peppler, 1988). Table 2-1 shows meaningful values of CAPE used in the United States.

**Table 2-1: Possible thresholds for CAPE [Adapted from Haby, Severe Weather Indices Page2015]**

Convective Available Potential Energy	
1 – 1500	Marginal Instability
1500 – 2500	Large Instability
≥ 2500	Extreme Instability



CAPE can however be computed in a variety of ways. Usually CAPE is calculated by lifting an air parcel from the surface however, this surface-based CAPE can be zero while a parcel lifted from higher up in the atmosphere may produce positive values of CAPE (Tuduri and Ramus, 1997). As a result, the most unstable CAPE (muCAPE) and mean layer CAPE (mlCAPE) are often used. The muCAPE is calculated by lifting the most unstable air parcel, within a defined section of the atmosphere, from the level with the highest  $\Theta_e$  (Blanchard et al., 2002; Benson, 2005). The mlCAPE is computed by lifting a parcel with average properties of the  $\Theta_e$  and mixing ratio within a section of the atmosphere (Benson, 2005; Bunkers et al., 2002; Rasmussen and Blanchard, 1998; Doswell and Rasmussen, 1994). The muCAPE provides a better representation for elevated convection when the surface layer is stable, but instability exists above the surface layer. This makes muCAPE useful during the evening when the surface layer cools down and becomes stable, while the layer above remains unstable. On the other hand, the surface CAPE provides the best estimate for surface-based convection, but is often similar to the muCAPE during the afternoon and early evening (Bunkers et al., 2002).

One of the variables that may be used to describe lightning occurrence is CAPE. (Burrows et al., 2005). The size and shape of CAPE in the vertical is strongly related to the updraft velocities in thunderstorms (Murugavel et al., 2014). The velocities of updrafts above the LFC are a function of CAPE, when CAPE is perfectly converted into vertical kinetic energy (Bluestein et al., 1988; Lucas et al., 1994). Updrafts play an important role in the distribution of hydrometeors, which in turn plays an important role in the electrification of a thundercloud (Singh and O’Gorman, 2015). Sufficient values of CAPE is necessary in the  $0^\circ\text{C}$  to  $-20^\circ\text{C}$  level in order to ensure that the updraft provides sufficient hydrometeors for charge separation (Bright et al., 2005). As was discussed in section 2.2.1, the main positive charge region in a thundercloud occurs at heights where the temperature is lower than  $-20^\circ\text{C}$ , the main negative charge region is found between the  $-10^\circ\text{C}$  and  $-20^\circ\text{C}$  temperature level and the small region of positive charge is found near the  $0^\circ\text{C}$  level. In section 2.2.2, it was also discussed that the charge reversal zone is found in the  $-10^\circ\text{C}$  and  $-20^\circ\text{C}$  temperature range. From Figure 2-1, it can be deduced that sufficient values of CAPE is thus required in about the 3-6 km above ground level range ( $0^\circ\text{C}$  to  $-20^\circ\text{C}$ ) to supply hydrometeors for charge separation. Sufficient values of CAPE above 6 km may also be important since updrafts need to transport the positively charged ice crystals to the higher parts of a cloud.

It has been shown in numerous studies that CAPE is useful for lightning prediction. In a study done by Solomon and Baker (1994), they showed that CAPE, calculated by lifting a parcel from cloud-base to cloud-top, is useful to predict lightning. They also showed that CAPE is a good predictor for the electrification of thunderstorms, but was not useful for forecasting

the amount of lightning. Livingston et al. (1996) found that mCAPE in the lowest 2 km above ground level (AGL) performed well in their method to forecast active and inactive lightning days. Burrows et al. (2005), in the development of a warm season lightning prediction method, found that muCAPE in the lowest 200 hPa, was not a top-ranked predictor, but rather a middle-ranked predictor for forecasting lightning, and concluded that CAPE must be used in conjunction with other predictors such as PW. Shafer and Fuelberg (2005; 2006) utilised surface CAPE as well as a modified CAPE in the development of a statistical procedure to forecast lightning. In another lightning forecast scheme by Shafer and Fuelberg (2008), various different variations of CAPE was considered. They utilised the top three performing CAPE's, muCAPE from cloud base to -20°C, muCAPE between -10°C and -25°C, and muCAPE between -15°C and -20°C in the development of their scheme. Frisbie et al. (2009) utilised muCAPE in the 0 to 3 km AGL range as input to his model, while Frisbie et al. (2013) made use of muCAPE in the 1 to 6 km AGL level. In a study by Zepka et al. (2014), it was showed that surface-based CAPE is useful for lightning prediction.

In this dissertation, four different variations of CAPE were considered. The first CAPE to be considered was the surface CAPE since it was previously used in lightning studies (Shafer and Fuelberg, 2005; Zepka et al., 2014) and is calculated when a parcel is lifted from the surface. The remaining three CAPE parameters considered were the muCAPE between different sections of the atmosphere. Firstly, the muCAPE when a parcel is lifted between the surface and the lowest 300 hPa was calculated, secondly the muCAPE when a parcel is lifted between the 1-6 km AGL range and lastly the muCAPE when a parcel is lifted between the surface and 3 km AGL was considered.

### 2.5.3.2 Lifted Index

The LI is calculated by taking the temperature an air parcel will have when the parcel is lifted adiabatically from the surface layer of the atmosphere up to the 500 mb level, and then subtracting this temperature value from the temperature of the environment at 500 mb (Peppler, 1988; Galway, 1956). The LI can be expressed by Equation 2-2 (Peppler, 1988):

$$LI = T_{500} - T_{p_{500}} \quad (2-2)$$

Where:  $T_{500}$  - Temperature of the environment at 500mb

$T_{p_{500}}$  - Temperature of parcel at 500mb

Many variations of the thickness of the surface layer from where an air parcel is lifted have been proposed. The original LI developed by Galway (1956) considered the mid-level of the lowest 3000 feet of the atmosphere as the starting point for an air parcel to be lifted (Peppler, 1988). In some cases, the surface layer has been selected as the lowest 100 hPa (Tuduri and Ramus, 1997; Solomon and Baker, 1994; Goodman, 1990; Kunz, 2007), while the lowest 50 hPa has also been considered. Other variations for the calculation of the LI exist. One type of variation is calculated by lifting the air parcel from the surface to the 500 hPa level, and is often called the surface LI (SLI) (Shafer et al., 2005; Fuelberg and Biggar, 1994; Kunz, 2007; Barlow, 1993). Another method uses the most unstable LI obtained when air parcels are lifted at levels between the surface and the 700 hPa level, and is called the best LI (BLI) (Shafer and Fuelberg, 2008; Stano et al., 2010; Frisbie et al., 2013).

The LI measures the latent instability in the atmosphere. Latent instability is a measure of the stability of the section of a conditionally unstable column of air above the LFC, and is used to assess whether an air parcel is negative or positively buoyant (Peppler, 1988). Table 2-2 shows typical threshold values for the LI used in the USA.

**Table 2-2: Possible thresholds for LI [Adapted from Haby, Severe Weather Indices Page, 2015]**

Lifted Index	
-1 to -4	Marginal Instability
-4 to -7	Large Instability
Less than -8	Extreme Instability

The LI is a frequently used instability index to forecast lightning (Harats et al., 2010). As with CAPE, the LI measures the buoyancy of an air parcel in the atmosphere, and as such the potential for strong updrafts, which supplies hydrometeors to a thundercloud (Virts and Houze, 2015). As was stated in section 2.5.2.1, updrafts play an important role in the distribution of hydrometeors, which in turn plays an important role in the electrification of a thundercloud (Singh and O’Gormon, 2015). The LI provides the difference between the actual temperature of the environment at 500 hPa and the temperature a parcel will have when it was lifted adiabatically to the 500 hPa level (Haklander and Van Delden, 2003), and the 500

hPa level corresponds with the charge separation zone in a thundercloud. As such, it can be deduced that the LI can be a useful measure of the potential updraft that feeds the charging zone with a sufficient supply of hydrometeors for electrification.

Many studies have utilised the LI to forecast lightning. Livingston et al. (1996) found that the LI is useful to forecast inactive lightning days. Shafer (2004) used a modified LI, which is similar to the standard LI, except that the maximum surface temperature is considered, in the development of statistical guidance to forecast warm season lightning over South Florida. Folsom (2004) developed a lightning forecast tool for the North Central and North Eastern USA and found that the LI was one of the two top predictors of lightning in many of the locations he investigated. Burrows et al. (2005) made use of the LI in their scheme to forecast warm season lightning over Canada and the Northern USA. Shafer and Fuelberg (2005; 2006) found that the LI is an important predictor of warm season lightning over South Florida. Another study by Shafer and Fuelberg (2008) considered the BLI in their final list of predictors to forecast lightning over South Florida. Keller (2006) utilised the LI in his lightning forecast scheme. The LI was used by Stano et al. (2010) to forecast lightning cessation at the Cape Canaveral Air Force Station and Kennedy Space Centre. Farukh et al. (2011) used the LI to assess lightning conditions in Alaska. Frisbie et al. (2013) utilised the BLI in his lightning forecasting scheme. Santos et al. (2013) found that CG lightning is related to five forcing factors, with the LI being one of them. Zepka et al. (2014) considered the LI as one of the parameters in her scheme to forecast lightning over south-eastern Brazil.

In this dissertation both the SLI and BLI were considered as possible predictors of lightning. The SLI was calculated by lifting an air parcel adiabatically from the surface to the 500 hPa level and subtracting this temperature from the temperature of the environment. The BLI was calculated by obtaining the most unstable LI when an air parcel was lifted adiabatically, between the surface and 700 hPa level in increments of 50 hPa, up to the 500 hPa level, from where the parcels temperature was extracted from the temperature of the environment.

### **2.5.3.3 Equivalent Potential Temperature**

The  $\Theta_e$  is the potential temperature that an air parcel will have when lifted to the lifting condensation level dry adiabatically, then wet adiabatically to a height where all of the water vapour has condensed out of the parcel, and finally back to the surface dry adiabatically (Houze, 1993; Bolton, 1980). A good approximation for  $\Theta_e$  for a saturated parcel is described by Holton (2013) with Equation 2-3:

$$\theta_e = \theta \exp\left(\frac{L_c q_s}{c_p T}\right) \quad (2-3)$$

Where:	$\theta$	-	Potential Temperature
	$L_c$	-	Latent heat of condensation
	$q_s$	-	Saturation Mixing Ratio
	$c_p$	-	Specific Heat of dry air at constant pressure
	$T$	-	Temperature

The  $\Theta_e$  measures the potential (or convective) instability of a layer in the atmosphere. When  $\Theta_e$  decreases with height ( $\frac{d\theta_e}{dz} < 0$ ), an unsaturated layer in the atmosphere is said to be potentially unstable (Kunz, 2007; Emanuel, 1994; Schultz et al., 2000; Madhulatha et al., 2013). The  $\Theta_e$  lapse rate ( $\Theta_e \Gamma$ ) is useful to assess instabilities necessary for thunderstorms development (Madhulatha et al., 2013).  $\Theta_e$  is also useful to assess changes in air masses (Dyson et al., 2015; Houston and Wilhelmson, 2012; Cummings, 2013; Huntrieser et al., 2007). Kuo (1966) showed that  $\Theta_e \Gamma$  is related to vertical velocities, while Smith et al. (2000) also states that surface  $\Theta_e$  can be used to estimate updraft velocities. As such,  $\Theta_e$  can be useful to describe the potential updrafts in thunderstorms, which supplies hydrometeors for charge separation in a thunderstorm. Cohen et al. (2007), however states that  $\Theta_e \Gamma$  and CAPE represents similar processes in the atmosphere. Dyson et al. (2015) also showed that CAPE and  $\Theta_e \Gamma$  were closely related. As such, it would seem reasonable to suggest that both CAPE and  $\Theta_e \Gamma$  should not be considered together when forecasting lightning. However,  $\Theta_e \Gamma$  can provide information of elevated convection that can be missed by CAPE (Frisbie et al., 2009).

Some studies have utilised  $\Theta_e$  to predict lightning. Livingston et al. (1996) considered surface  $\Theta_e$  in the development of his forecasting scheme and found a reasonable correlation between  $\Theta_e$  and lightning. However, he did not consider  $\Theta_e$  in his final prediction scheme since other parameters proved more useful. Rajeevan et al. (2012) considered mean  $\Theta_e$  advection between 925 and 900 hPa for the development of lightning forecasting scheme. Frisbie et al. (2009) considered the  $\Theta_e \Gamma$  at 600 mb, Frisbie (2009) the  $\Theta_e \Gamma$  between 1 and 6 km AGL, and Frisbie et al. (2013)  $\Theta_e \Gamma$  between 2 and 5 km AGL as well as at the  $-10^\circ\text{C}$  level, for the LPI. Zepka et al. (2014) did consider the surface  $\Theta_e$  in the development of a lightning forecasting methodology, but found that the CAPE and  $\Theta_e$  provided similar results when compared.

Seven different variation of the  $\Theta_e$  was considered in this dissertation. The first to be considered was the surface  $\Theta_e$ . The remaining six considered was  $\Theta_e\Gamma$  between different levels. The lapse rates were calculated at 600 hPa, between 850 and 400 hPa, between 850 and 500 hPa, between 1 and 6 km AGL, between the  $-10^\circ\text{C}$  and  $-20^\circ\text{C}$  levels, and lastly between the 700 and 500 hPa levels.

#### 2.5.3.4 Precipitable Water

PW is a parameter that gives an indication of the amount of water vapour available in the atmosphere, and is the mass of water vapour in a column of unit cross-sectional area between two layers in the atmosphere (Dupilka and Reuter, 2006). The PW can be calculated by Equation 2-4 from Tuduri and Ramus (1997):

$$PW = \frac{1}{g} \int_p^{p_0} q dp \quad (2-4)$$

Where:

$g$	-	Acceleration due to gravity
$q$	-	Specific humidity
$p$	-	Starting pressure level
$p_0$	-	Stopping pressure level

One of the requirements for lightning formation (or thunderstorm formation) is moisture (Burrows et al., 2005). PW can be utilised to estimate the amount of moisture available in the atmosphere (Dupilka and Reuter, 2006). Colson (1960) found that lightning flash rates were generally greater on days with large PW values. Thunderstorms will also produce large amounts of lightning when CAPE and PW values are high (Haby, Skew-T: A Look at PW, 2015). This can be due to large updrafts transporting more water vapour to heights where charge separation occurs in thunderclouds (Rose, 2008).

Mazany et al. (2002) found that PW derived from Global Positioning Systems (GPS) measurements proved useful to get an idea of where lightning will occur. Burrows et al. (2005) found that PW ranked third overall to predict warm-season lightning over Canada and the

Northern USA. Lambert et al. (2005) considered PW in his lightning scheme, but did not include PW in his final calculations since it was not one of the top five predictors. Shafer & Fuelberg (2005; 2006) considered PW as a possible predictor but did not include it in their final scheme. Another study by Shafer and Fuelberg (2008) utilised the PW to forecast warm-season lightning over Florida. Kehrer et al. (2008) found that GPS PW showed favourable results to forecast lightning. Frisbie et al. (2009) considered PW in the development of his LPI. Stano et al. (2010) considered PW to forecast lightning but did not utilise it in his final prediction. PW was one of the 12 short listed parameters in the lightning forecast scheme of Rajeevan et al. (2012). In this study, PW was calculated between the surface and 100 hPa, between 700 and 400 hPa, and between 850 and 300 hPa.

### 2.5.3.5 Relative Humidity

The RH is described as the ratio between the vapour pressure and the saturation vapour pressure of water that gives an indication of how close to saturation the air is. RH values of 100% indicate that the atmosphere is saturated (Houze, 1993). The RH can be expressed by Equation 2-5 (Houze, 1993):

$$RH = \frac{e}{e_s} \quad (2-5)$$

Where:  $e$  - Vapour Pressure  
 $e_s$  - Saturation vapour pressure

One of the requirements for lightning formation (or thunderstorm formation) is moisture (Burrows et al., 2005). RH provides information on how saturated the air is. It has been shown that RH can have an influence on lightning formation. Xiong et al. (2006) found that lightning corresponds well with RH in longitudinal belts where RH values are greater than 74% and negatively with RH in longitudinal belts where the RH is less than 72%. They also found a similar result for latitudinal belts. This means that higher RH values in dry regions causes more lightning activity, while high RH values in wet regions lowers the levels of lightning activity (Xiong et al., 2006). This same study also showed that when the RH is too high (greater than 72%) in wet regions, lightning can be suppressed. Berdeklis and List (2001) performed an

experiment with a new Triple Interaction Facility that allows for the analysis of cloud particles in a simulated cloud. In their experiments, they found that there is a good correlation between charge transfer and RH. This is most pronounced in the -12 to -18 degrees Celsius level where charge transfers occurs when ice and graupel particles in the presence of supercooled water collides. They stated that increases in RH values results in greater negative charges while lower values of RH is related closer to smaller negative charges and higher positive charges.

Livingston et al. (1996) considered RH in his final scheme to forecast inactive lightning days. Lambert et al. (2005), Shafer and Fuelberg (2006; 2008) and Stano et al. (2010) considered RH in their schemes to forecast lightning, but did not include RH in their final product since RH was not one of the top predictors. Frisbie et al. (2009) considered RH at -10°C in their LPI, while Frisbie et al. (2013) considered the mean RH, maximum RH and minimum RH in the 3-6 km AGL level. Rajeevan et al. (2012) found that the RH was one of the 12 top ranked predictors in his scheme to forecast lightning.

Five different RH parameters were considered in this study, which includes the RH at the -10°C level, the mean RH between the -12°C and -18°C levels, and the minimum, maximum and average RH between the 3-6 km AGL range.

### **2.5.3.6 Air temperature**

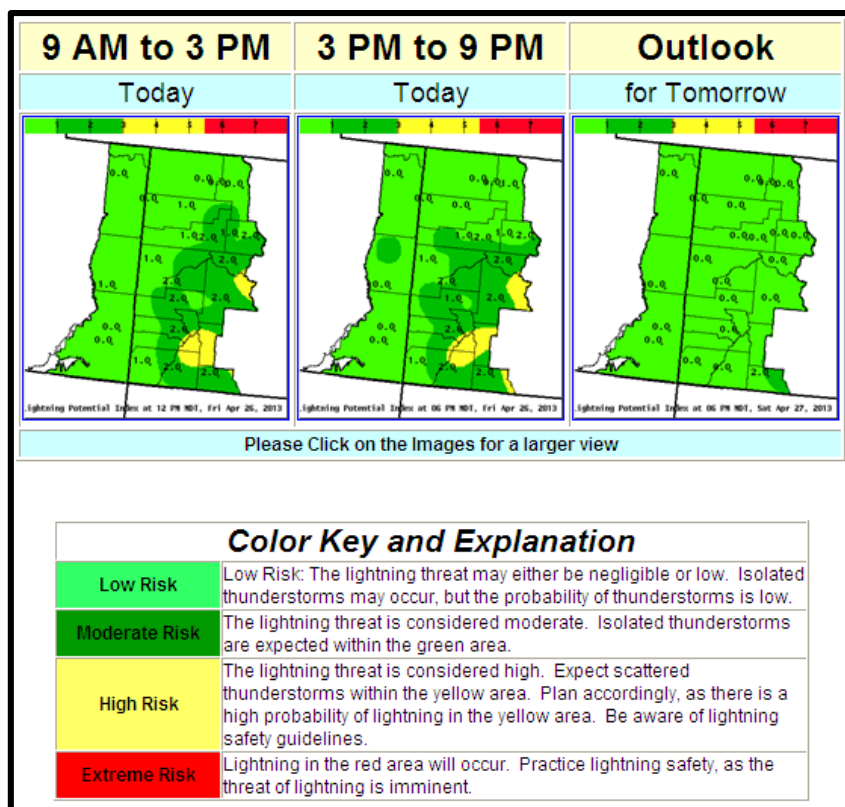
The atmosphere can be seen as a heat engine, where the sun heats up the earth's surface, which in turn causes our weather systems that transfers heat and moisture to other parts of the globe (Laliberté et al., 2015). Surface heating from the sun is responsible for the convective processes that result in atmospheric instabilities (Bharatdwaj, 2006). This means that air temperature plays a big role in the development of thunderstorms. Price (2013) lists many studies (e.g. Williams, 1992; Williams, 1994; Williams, 2009; Reeve and Toumi, 1999; Markson and Price, 1999; Price, 1993; Price and Asfur, 2006; Markson, 2007), which have shown that lightning is closely related to the surface temperature. Due to this important link between temperature and lightning, temperature should be investigated as a possible predictor for lightning occurrence.

In this dissertation four different air temperature parameters were consider. This include the temperature at 1.5 m AGL, temperature at the 700 hPa level, the mean temperature of all pressure levels between the 850 and 700 hPa level, and the mean temperature between the 500 and 300 hPa levels.



## 2.6 NWS LIGHTNING POTENTIAL INDEX

This study was inspired by a paper presented by Frisbie et al. (2009). The LPI, uses the NWP prognosis of CAPE, LI,  $\Theta_e \Gamma$ , air temperature at 850 mb ( $T_{850}$ ), PW and RH at  $-10^\circ\text{C}$  to calculate the potential of lightning (Frisbie et al., 2009). This version of the LPI differs from the LPI proposed by Lynn and Yair (2010) in that the Lynn and Yair (2010) methodology utilises model microphysical fields. The capability of NWP models to resolve microphysical parameters is a relatively new development and many older NWP models operated by some institutions currently do not have these capabilities or at least the specific microphysical parameters needed for the Lynn and Yair (2010) LPI. As such the Lynn and Yair (2010) approach was not considered. The Frisbie et al. (2009) version of the LPI is still in the developing phase, but has shown the ability to be useful in improving lightning predictions. It was developed at The National Weather Service (NWS) Forecast Office in Grand Junction, Colorado. The operational index provides three lightning outlook maps for each day. Figure 2-9 contains three outlook maps for the State of Colorado, which displays the level of risk that can be expected during the given time periods and the legend at the bottom of the maps explains what the risk categories means.



**Figure 2-9: LPI outlook maps for the State of Colorado together with the colour key and explanation [Source: National Weather Service, 2013]**

Frisbie et al. (2009) developed his methodology for calculating the LPI by combining the model parameters into a single index using Equations 2-6 to 2-8:

$$A = (RH)^2 \times (\theta e \Gamma) \times (LI)^2 \times (-1) \quad (2-6)$$

$$B = (\mu CAPE) \times (PW) \times (RH) \times 0.001 \quad (2-7)$$

$$LPI = (A + B) \times (T_{850} - 272.15) \quad (2-8)$$

Where:	RH	-	Relative Humidity at -10°C
	$\Theta e \Gamma$	-	Equivalent Potential Temperature Lapse Rate at 600mb
	LI	-	Lifted Index
	muCAPE	-	Most Unstable CAPE in the 0-3 km above ground level range
	PW	-	Precipitable Water
	$T_{850}$	-	850mb Temperature in Kelvin

With the following conditions:

LI=0 if the LI>0

LPI=0 if the LPI<0

LPI=20000 if LPI>20000

The same methodology was tested over South Africa and showed promising results on certain days (Gijben, 2013). Atmospheric conditions in South Africa are different to the conditions in Colorado, due to for example the altitude, latitude, distance from the oceans and general circulation patterns (Nelson, 2007; Taljaard, 1994). The traits of the different NWP models used in Colorado and South Africa can also affect the performance of the LPI. In this dissertation, the LPI was redeveloped for South African conditions.

## 2.7 SUMMARY

Thunderstorms require three key ingredients to develop: instability, moisture and a trigger mechanism, while the electrification of a thundercloud involves a process that electrifies individual cloud-particles as well as a process that separates these cloud-particles according to their polarity in different levels of the thundercloud. Electrification of a thundercloud occurs because of collisions between cloud-particles in the presence of supercooled water droplets, and updrafts play an important role to distribute these particles to different heights in a thundercloud. The updrafts transport positively charged ice particles to heights where the temperature is lower than  $-20^{\circ}\text{C}$  (main positive charge region), negatively charged graupel particles to heights between  $-10^{\circ}\text{C}$  and  $-20^{\circ}\text{C}$  (main negative charge region), and positively charged ice crystals to a small region found near the  $0^{\circ}\text{C}$  level. Statistical prediction schemes forecast the threat of lightning by relying on connections between lightning occurrence and parameters of the pre-storm environment to model the processes involved for thunderstorm development and electrification of a thundercloud. The CAPE, LI and  $\Theta_e$ , which are measures of conditional, latent, and potential instability respectively, provides useful insights into the buoyancy of air parcels in the atmosphere. These parameters are linked to potential updrafts that supply thunderclouds with the hydrometeors necessary for electrification, as well as to distribute these charged particles in a thundercloud. In order for thunderstorms to develop and to ensure that sufficient amounts of hydrometeors are present for the electrification of a thundercloud, the availability of moisture is vital. PW and RH are useful parameters to monitor the amount moisture available in the atmosphere. Solar heating plays an important role in the distribution of lightning across the global since increases in temperatures results in the increase of water vapour concentrations as well the instabilities that drive thunderstorm and lightning development. These parameters are usually derived from atmospheric soundings, however with the advent of accurate numerical weather prediction models, many of the limitations found with soundings can be overcome in order to provide accurate forecasts of the potential areas where lightning can occur.

---

# CHAPTER 3

## DATA AND METHODOLOGY

---

### 3.1 INTRODUCTION

This chapter discusses the data and methodologies used in this study. The first section (Section 3.2) gives an overview of the study area and period considered. The next section (Section 3.3) deals with the lightning and NWP data. An overview is provided of the SALDN, and the NWP data is presented with an overview of the UM. In Section 3.4, a description of the development of the LTI is provided. This includes the techniques used to select the most appropriate parameters for lightning prediction as well as the development of the new LTI by means of rare-event logistic regression. The chapter concludes (Section 3.5) with the methods used to evaluate the LTI. This includes a discussion on probabilistic and deterministic statistical scores, as well as eyeball verification techniques.

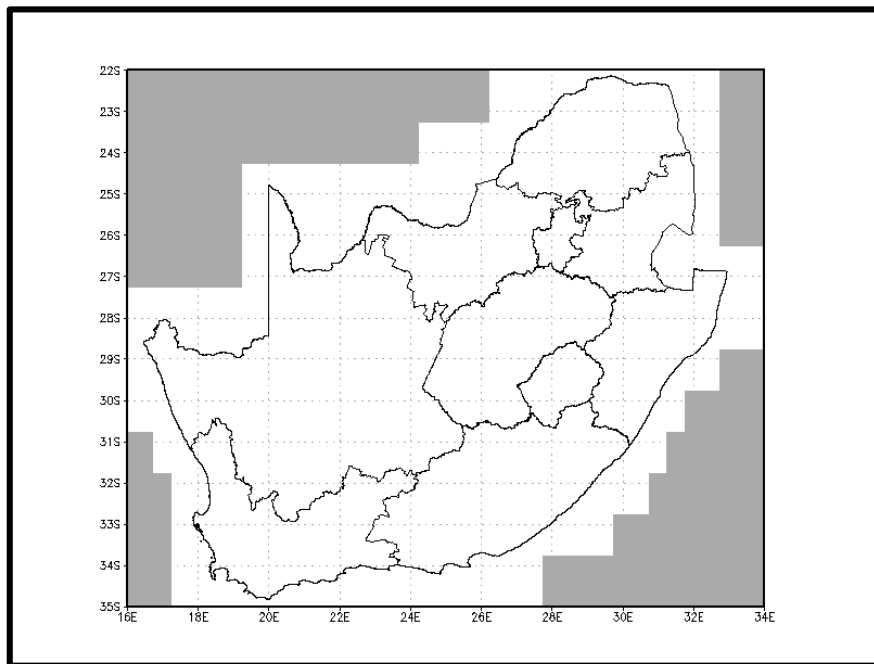
### 3.2 STUDY AREA AND PERIOD

#### 3.2.1 Study area

This study was conducted over southern Africa in the area depicted in Figure 3-1. The grey in Figure 3-1 depicts areas where the SALDN cannot provide accurate information. This is due to the detection efficiency (DE) of LDN's decreasing rapidly outside the confines of a network (Cummins, 1998a).

The study domain was divided into a  $0.5^\circ \times 0.5^\circ$  grid resulting in 795 grid points. The model data, (Section 3.4), is available on a higher resolution ( $0.11^\circ \times 0.11^\circ$ ) grid. Due to the extensive computational time, the amount of computing resources available and the long periods considered (Section 3.2.2), the coarser grid had to be implemented. If the model grid resolution was utilised in this study, there would have been a 22 times increase in computation time of all the parameter extractions and calculations. Apart from this increase in the extractions and calculations of model data, the regression procedures would also have taken much longer due to the increase in data points. It was thus decided that utilising the model grid would have been unfeasible. A higher resolution would have provided more detail in the

LTI especially with the smaller forecast scales of thunderstorms. The coarser grid was however used only for the development of the LTI, and the equations can still be applied to a higher resolution model when the product becomes operationally available. A visual comparison between the LTI on the  $0.11^\circ \times 0.11^\circ$  UM resolution and the LTI on the  $0.5^\circ \times 0.5^\circ$  resolution utilised in this study is shown in Case 4 of Section 5.4.4.

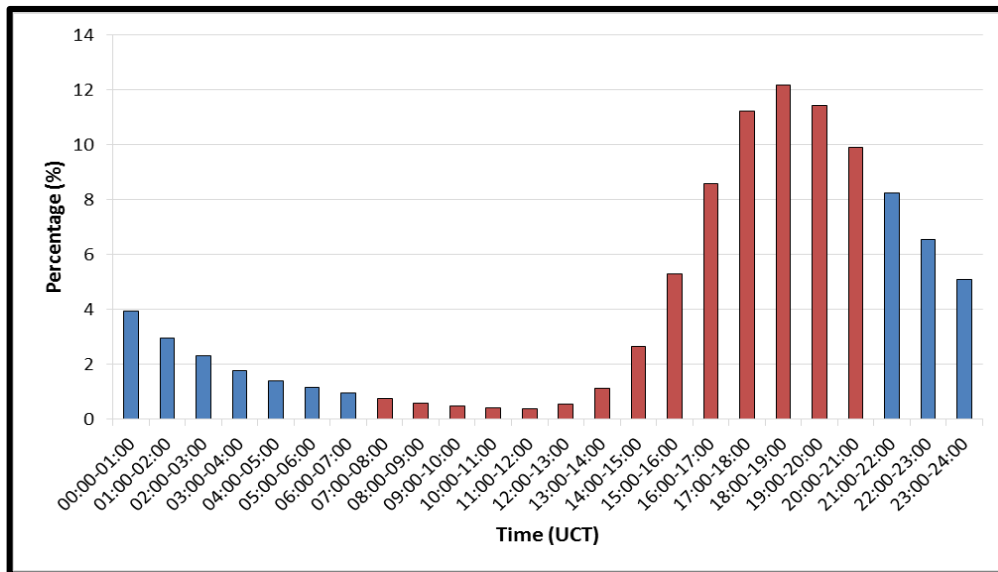


**Figure 3-1: The study domain. Grey areas are excluded.**

### 3.2.2 Study period

One of the aims of this study was to produce a lightning outlook map for the day, and was created by utilising hourly NWP prognosis for the period 07:00 - 21:00 Coordinated Universal Time (UTC). Most thunderstorm activity occurs during the afternoon and evening in South Africa (de Coning et al., 2011), but convection also occurs in the morning (Rouault et al., 2013). On average approximately 65% of the daily lightning over South Africa occurs between 07:00 - 21:00 UTC. Figure 3-2 shows the average 8-year diurnal cycle of lightning over South Africa, produced with lightning data for the period 2006 – 2013. The figure shows the percentage of the total daily lightning amounts for each hour of the day, where the red bars indicate the lightning within the period 07:00 – 21:00 UTC. Figure 3-2 shows that the largest percentages of lightning occur during late afternoon and early evening, but lightning is also seen in the early hours of the morning. Ideally one would want to produce a lightning outlook map for the entire day but the NWP model output and the processing of the lightning outlook

maps with the NWP fields would only be available operationally early in the morning. As a result, the outlook map was created with NWP prognosis for the period 07:00 – 21:00 UTC which is 09:00 - 23:00 South Africa Standard Time (SAST). Lightning data for the period 06:30 – 21:30 UTC (the reasoning for this period will be discussed in Section 3.3.2) was considered in this study, and essentially the NWP prognosis will forecast lightning activity during this 15-hour period, ending just before midnight SAST.



**Figure 3-2: The average diurnal cycle of lightning over South Africa for the period 2006-2013, expressed as a percentage of the daily total.**

The NWP parameters best suited to predict lightning over South Africa were identified by extracting the appropriate data from the NWP model for the austral summer days of 2011/12 and 2012/13. The same period were used to select the most appropriate parameters to predict lightning and to train the LTI statistical model (Section 3.5). However, austral summer was divided into spring months (September to November) and summer months (December to February) and a different LTI was developed for these seasons separately. The atmospheric conditions in South Africa are different between the seasons. This is because in early summer, the atmospheric circulation is generally extra-tropical with a conditionally unstable atmosphere over certain parts of South Africa, while in late summer the circulation is tropical with a convectively unstable atmosphere (Dyson et al., 2015). The spring and summer LTI's were verified for the austral summer of 2013/14. For the training period, 17 out of the 363 days considered could not be utilised, while for the evaluation period, 6 out of the 181 days could not be used. This was due to the NWP model that did not run on these days.

### 3.3 LIGHTNING DATA

#### 3.3.1 Southern African Lightning Detection Network

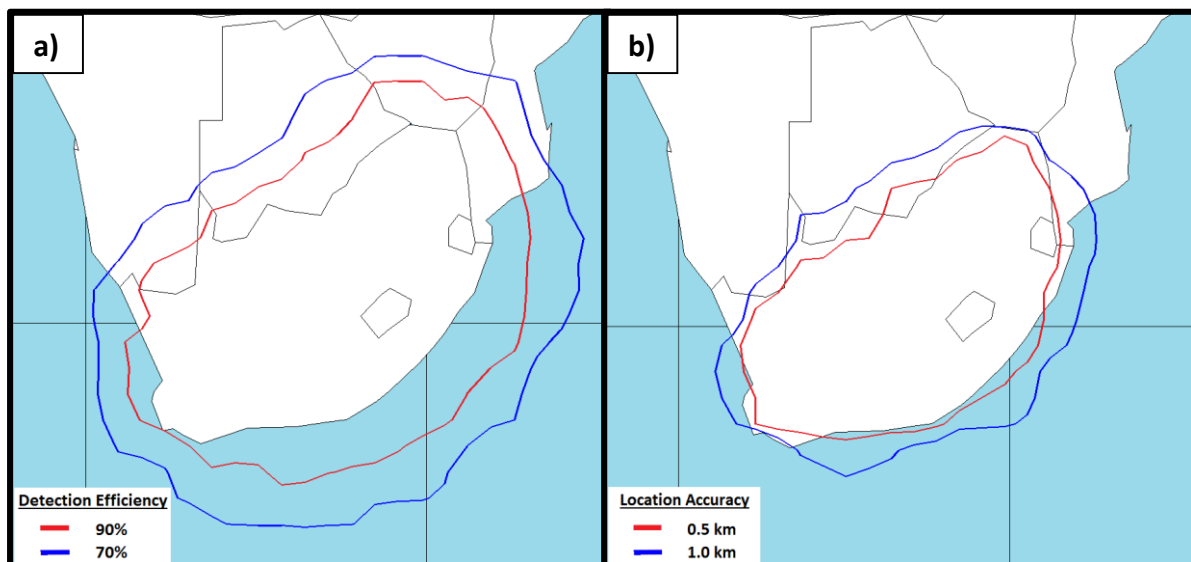
##### 3.3.1.1 The network

Towards the end of 2005, SAWS installed its first LDN that utilises Vaisala sensors. This original network comprised of 19 sensors distributed across the country (van de Groenendaal, 2007). Over the years, the network has undergone a series of upgrades where sensors have been relocated and new sensors installed (Gijben, 2012). During the 2009/10 upgrade, 4 new sensors were added to the network (one of the new sensors replaced the old sensor at Springbok), which resulted in a network of 22 sensors. One of these sensors were installed in Swaziland, after which the network became known as the SALDN. Another upgrade was performed on the SALDN in 2011, where 4 sensors were relocated and 2 new sensors were added to the network. This resulted in a network consisting of 24 sensors, which was completed in August 2011 (Ngwato, 2014). These upgrades improved the performance of the SALDN (Hunt et al., 2014). The 24 sensors shown in Figure 3-3 made up the SALDN network for the entire study period.



**Figure 3-3: The location of the 24 lightning detection sensors of the SALDN network operated by SAWS [Map created from information from Ngwato (2014)]**

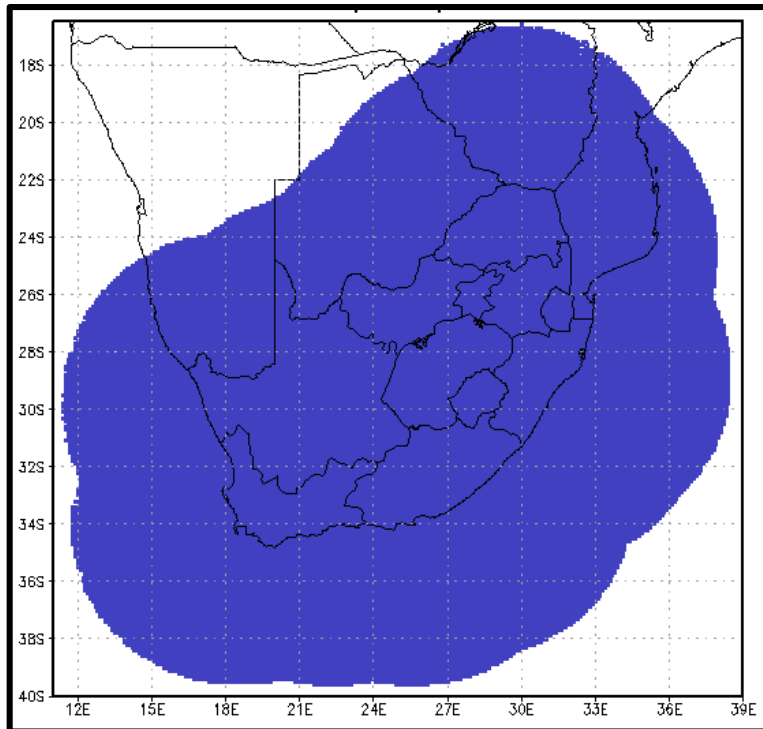
The SALDN can detect lightning flashes with a predicted DE of 90% or more over most of the country (van de Groenendaal; Gill, 2008; Gijben, 2012). This means that at least 90% of CG lightning flashes are detected over South Africa. Lightning strokes can also be detected with a median location accuracy (LA) of 0.5 km over most of the country, which means that a CG lightning stroke can be accurately positioned within 500 meters of the actual stroke (Gijben, 2012). These estimated values are derived from the Vaisala Network Performance Evaluation Program (NPEP) (Vaisala, 2016). In order to get exact network performance statistics would require either the controlled measurements of rocket triggered lightning or photographs of lightning striking objects (Gill, 2008). Figure 3-4a shows the projected DE and Figure 3-4b the median LA. The red lines show the 90% DE and 0.5 km LA, while the blue lines show the 70% DE and 1.0 km LA range rings.



**Figure 3-4: The projected (a) detection efficiencies and (b) location accuracies of the SALDN from the Vaisala NPEP**

The maximum range of the SALDN is shown in Figure 3-5 and extends far outside the borders of South Africa. The sensors that make up the SALDN is located within the borders of South Africa and the DE and LA of LDN networks decreases rapidly outside the confines of a network (Cummins, 1998a). In order to utilise accurate lightning data, the areas inside the range rings of the DE and LA from Figure 3-4 were used to identify the study domain (Figure 3-1).





**Figure 3-5: Absolute maximum range of the SALDN, where the blue area represents the areas where lightning can be detected. Outside the borders of SA, the accuracy of lightning detection reduces rapidly.**

### 3.3.1.2 The sensors

The SALDN consists of Vaisala lightning detection sensors, which makes use of a combination of LS7000 and LS7001 sensors (Gijben 2012). Most lightning detection sensors utilise either MDF or TOA techniques to detect lightning (Cummins et al., 1998b), however the latest lightning detection sensors, including those in the SALDN, makes use of a combination of MDF and TOA principles. This combined technology sensors are able to detect lightning more accurately since the combined intersection of the MDF vectors and TOA circles will provide a more accurate measurement than using one of these techniques individually (Gill, 2008; Cummins et al., 2000). By using only the MDF or TOA technique, at least three sensors need to participate in a network to accurately detect a lightning flash. With the combined technology, a lightning flash can be accurately detected by using only two sensor (Vaisala, 2004; Gill, 2008; Cummins et al., 1998b; Cummins et al., 2000). The SALDN sensors are designed to detect the electromagnetic waves emitted by CG lightning in the VLF and LF range (Gijben, 2012). As was discussed in Section 2.3 of Chapter 2, this enables the sensors to detect CG lightning over greater distances, but only a small amount of IC lightning can be detected.

### 3.3.1.3 Advantages and limitations

The biggest advantage of lightning measurements from the SALDN is the ability to measure the occurrence of CG lightning over the entire South Africa with great accuracy. Each lightning stroke or flash that strikes the earth is recorded with a date, time, latitude and longitude coordinate, strength and many more parameters. Lightning data from the SALDN is extremely reliable and contains no gaps. Even if a lightning sensor in the network fails, the other sensors in the network are able to detect lightning near the offline sensor.

A limitation of the SALDN is the inability to detect most of the IC lightning. Intra and inter-cloud lightning (IC) is the lightning that occurs within a cloud and between two different clouds respectively and makes up approximately 70% of all lightning (Price, 2008). The SALDN is designed to detect CG lightning that contributes to about 30% of all lightning. One can argue that CG lightning is the type of lightning that affects humans and infrastructures, however cloud lightning can provide valuable information on thunderstorms.

### 3.3.2 Lightning data utilised in this study

Lightning data served three purposes in this study: 1) to determine which model parameters are the best to predict lightning 2) to train the new LTI and 3) to evaluate the performance of the newly developed LTI.

Lightning flash and stroke data are available from the SALDN. A lightning flash is the entire electrical discharge while the components of a flash that strikes the ground are called strokes (Gjiben, 2012; Rodger and Russel, 2002; Uman, 1987). A lightning flash can produce more than one stroke. The SALDN uses an algorithm to cluster strokes into a flash, and a description on this algorithm can be found in Gjiben (2012) and Gill (2008). In this study lightning stroke data from the SALDN was utilised.

The study domain was divided into grid boxes of size  $0.5^{\circ} \times 0.5^{\circ}$ . Lightning data was accumulated between 06:30 and 21:30 UTC and was spatially assigned to the correct grid box. Model data is available hourly and represents a forecast on the hour. This means that for example a 12:00 UTC model forecast is valid for 12:00 UTC and does not necessarily mean a forecast for 12:00 – 13:00 UTC. For this reason lightning data surrounding the model time step (in the 30-minutes before and after) was considered. Since one outlook map for the day, between 07:00 UTC and 21:00 UTC, was created from model data, the corresponding accumulation of lightning data was between 06:30 and 21:30 UTC.

The purpose of the LTI is to predict whether lightning will occur or not and it does not aim to predict the amount of lightning. Statistical techniques that make use of thermodynamic parameters to forecast lightning perform well in predicting the risk of lightning, but have a lower capability to predict the amount of lightning (Soul et al., 2002; McCaul et al., 2009). As a result, the lightning accumulations in each grid box were converted to a binary outcome. If one or more lightning strokes occurred in a grid box during the forecast period of the day, the grid box was given a value of 1. If no lightning occurred the grid box was given a value of 0.

### **3.4 NUMERICAL WEATHER PREDICTION MODEL DATA**

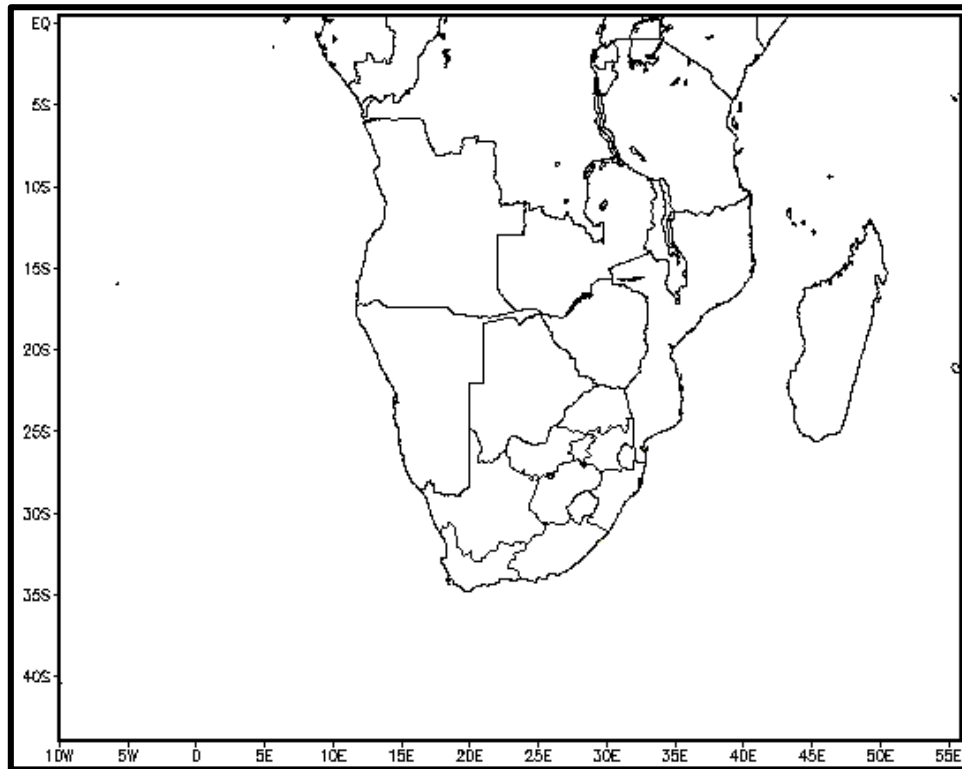
#### **3.4.1 The Unified Model**

##### **3.4.1.1 The model**

The UM is the NWP model developed at the United Kingdom Meteorological Office (UKMO) (de Coning et al., 2011). This UM was first made operational during 1991/2 at the UKMO (Cullen et al., 1997), but is regularly updated to improve its forecasting capabilities (Davies et al., 2005). In 2002, a series of upgrades was performed on the UM, which resulted, among others, in the model becoming a non-hydrostatic model that follows the terrain and resolves many layers in the atmosphere with height based vertical coordinates (Davies et al., 2005). The UKMO runs the UM on various resolutions, but also offers a global model which runs 4 times daily on a 40 km resolution (Landman et al., 2012).

The SAWS runs a local version of the UM, which has been operational at SAWS since 2006. Different version of the UM runs with different configurations (parametrization schemes, horizontal resolutions and with or without data assimilation) at SAWS and makes use of initial and boundary conditions from the global model from the UKMO (Landman et al., 2012). Among the different configurations are three model runs, called xaana, xaang and xaant, all with a 12 km horizontal resolution. The xaana and xaang model runs utilise version 6.1 of the UM, while the xaant model run makes use of the newer version 7.3 of the UM. All three of these models run once daily and produce hourly forecasts on 38 vertical levels and for 48 hours ahead (Landman et al., 2012). Both the xaana and newer xaant model have no data assimilation, while the xaang model has continuous data assimilation. Data assimilation is the process where observational data is assimilated into a model in order to achieve a more accurate representation of reality (Constantinescu et al., 2007). Observational data is often assimilated into numerical weather prediction models to provide the initial conditions of the

model (Whitaker et al., 2004). In this study the newer xaant model (version 7.3 of UM) was utilised. The local version of the UM covers the domain as depicted in Figure 3-6.



**Figure 3-6: The domain of the 12 km horizontal resolution Unified Model at SAWS**

### **3.4.1.2 Advantages and limitations**

NWP models have improved significantly during the last few years, due to improved dynamical and physical representations of reality, increases in the horizontal and vertical resolutions and improved parametrization schemes (Schulze, 2007). As a result, accurate and reliable forecasts are now possible from NWP models. The UM has also undergone upgrades and provides accurate forecasts of various parameters for 48 hours ahead, on a 12 km horizontal resolution for 38 levels in the vertical (Landman et al., 2012). Using UM derived parameters in the creation of the LTI means that data for the entire South Africa can be utilised with great accuracy. A limitation of the UM is that the parameters remain a model representation or proxy of reality, and can sometimes perform poorly.

### 3.4.2 Model data utilised in this study

Frisbie et al. (2009) utilised a form of CAPE, LI,  $\Theta_e$ , PW, RH and air temperature in the development of the LPI. In this study, the same type of parameters were considered. The atmospheric conditions as well as the NWP model utilised in South Africa are different to Colorado in the USA and as such, different variations of Frisbie et al.'s parameters were considered for the development of the LTI. Table 3-1 lists the six main groups of parameters as well as the individual predictors in each of the main groups considered in this study. The predictors were selected from literature (listed in Table 3-1) where they were found useful in either lightning prediction studies or thunderstorm/rainfall development. Most of the parameters listed in Table 3-1 were not directly available from the UM and had to be calculated. Only the  $T_{1p5m}$ , and  $T_{700}$  parameters in Table 3-1 was available. A description on how the different parameters were calculated is provided in the table.

All of the model predictors were interpolated to a  $0.5^\circ \times 0.5^\circ$  grid in order to correspond to the lightning data. Hourly model data between 07:00 and 21:00 UTC were utilised. To group the hourly model data into a single daily forecast, the maximum absolute value of the hourly forecasts was used.

**Table 3-1: Candidate predictors for inclusion in the lightning prediction model**

Group	Abbreviation	Name	Equation	Reference
1	$\text{muCAPE}_{0,3\text{km\_AGL}}$ [J kg <sup>-1</sup> ]	Most unstable CAPE between various levels	Largest CAPE obtained when each parcel between the surface and 3 km above the ground is lifted from the level with the highest $\Theta_e$	Frisbie et al. (2009) Groenemeijer and van Delden (2007)
	$\text{muCAPE}_{1,6\text{km\_AGL}}$ [J kg <sup>-1</sup> ]	Most unstable CAPE between various levels	Largest CAPE obtained when each parcel between 1 km and 6 km above the ground is lifted from the level with the highest $\Theta_e$	Frisbie et al. (2013)
	$\text{muCAPE}_{\text{lowest\_300}}$ [J kg <sup>-1</sup> ]	Most unstable CAPE between various levels	Largest CAPE obtained when each parcel between the surface and lowest 300 hPa above the ground is lifted from the level with the highest $\Theta_e$	Craven & Brooks (2004)
	$\text{CAPE}_{\text{surf}}$ [J kg <sup>-1</sup> ]	Surface CAPE	CAPE obtained when a parcel is lifted from the surface	Zepka et al. (2014)
2	$\theta_e$ [K]	Surface equivalent potential temperature	$\theta_e = \theta \exp\left(\frac{L_c q_s}{c_p T}\right)$	Livingston et al. (1996) Zepka et al. (2014)
	$\theta_e \Gamma_{600}$ [K]	Equivalent potential temperature lapse rate	$\Theta_e$ lapse rate at 600 hPa	Frisbie et al. (2009)
	$\theta_e \Gamma_{850,400}$ [K]	Equivalent potential temperature lapse rate	$\Theta_e$ lapse rate between 850 and 400 hPa	Dyson et al. (2015)
	$\theta_e \Gamma_{850,500}$ [K]	Equivalent potential temperature lapse rate	$\Theta_e$ lapse rate between 850 and 500 hPa	Dyson et al. (2015)
	$\theta_e \Gamma_{1,6\text{km\_AGL}}$ [K]	Equivalent potential temperature lapse rate	$\Theta_e$ lapse rate between 1 and 6 km AGL	Frisbie (2009)
	$\theta_e \Gamma_{m10,m20}$ [K]	Equivalent potential temperature lapse rate	$\Theta_e$ lapse rate between -10°C and -20°C levels	Frisbie et al. (2013)
	$\theta_e \Gamma_{700,500}$ [K]	Equivalent potential temperature lapse rate	$\Theta_e$ lapse rate between 700 and 500 hPa	Zepka et al. (2013)

**Table 3-1: Candidate predictors for inclusion in the lightning prediction model (continued)**

Group	Abbreviation/units	Name	Equation	Reference
3	SLI [°C]	Surface lifted index	$SLI = T_{500} - T_{p_{500}}$ (Lifting the parcel from the surface)	Garreaud et al. (2014) Haklander & van Delden (2003)
	BLI [°C]	Best lifted index	The most unstable LI when each parcel is lifted between the surface and 700 hPa	Frisbie et al. (2013) Shafer and Fuelberg (2008)
4	PW <sub>850,300</sub> [cm]	Precipitable water	Mean precipitable water in the 850-300 hPa level	Dyson et al. (2015)
	PW <sub>700,400</sub> [cm]	Precipitable water	Mean precipitable water in the 700-400 hPa level	Burrows et al. (2005)
	PW <sub>surf,100</sub> [cm]	Precipitable water	Mean precipitable water from the surface to 100 hPa level	Burrows et al. (2005) Shafer and Fuelberg (2008)
5	RH <sub>m10</sub> [%]	Relative humidity at the -10°C level	Relative humidity at the -10°C level ( $RH = \frac{e}{e_s}$ )	Frisbie et al. (2009)
	RH <sub>m12,m18</sub> [%]	Mean relative humidity	Mean relative humidity between the -12°C to -18°C level	Frisbie et al. (2013)
	aveRH <sub>3,6km_AGL</sub> [%]	Mean relative humidity	Mean relative humidity between 3-6 km above ground level	Frisbie et al. (2013)
	maxRH <sub>3,6km_AGL</sub> [%]	Maximum relative humidity	Maximum relative humidity between 3-6 km above ground level	Frisbie et al. (2013)
	minRH <sub>3,6km_AGL</sub> [%]	Minimum relative humidity	Minimum relative humidity between 3-6 km above ground level	Frisbie et al. (2013)
6	T <sub>1p5m</sub> [K]	Temperature	Temperature at 1.5 meters above the ground	Mazany et al. (2002)
	T <sub>700</sub> [K]	Temperature	Temperature at the 700 hPa level	Burrows et al. (2005)
	T <sub>850,700</sub> [K]	Mean temperature	Mean temperature of all pressure levels between 850 and 700 hPa	Dyson et al. (2015)
	T <sub>500,300</sub> [K]	Mean temperature	Mean temperature of all pressure levels between 500 and 300 hPa	Dyson et al. (2015)

## 3.5 DEVELOPMENT OF A NEW LIGHTNING THREAT INDEX (LTI)

### 3.5.1 Parameter Selection

An aim of this study was to select the model parameters that work best to predict lightning in South Africa. The objective is to select one parameter out of each of the six main groups (CAPE, LI,  $\Theta_e$ , PW, RH and air temperature) that will be used in the development of the LTI. Both the Statistical and Analysis Software (SAS) and R software were utilised to select the best model predictors.

#### 3.5.1.1 Stepwise regression with SAS

In order to select the best parameters to predict lightning, SAS statistical software was used to perform a full (backwards and forwards) stepwise logistic regression by using Firth's Penalised Likelihood method. Firth's methodology (Firth, 1993) solves the issue of bias for maximum likelihood estimates (due to separation) in logistic regression (Shen and Gao, 2008). The PROC LOGISTIC function in SAS was used to perform the stepwise logistic regression by making use of the option SELECTION=STEPWISE. In order to make use of Firth's Penalised Likelihood Method, the FIRTH option was used in the PROC LOGISTIC function (Robin et al., 2011; SAS Institute Inc., 2010).

When SAS performs a stepwise regression, it first estimates parameters based on the intercepts and explanatory variables added to the model. The score chi-square statistic is then calculated for each variable. The parameter with the highest chi-square score is added to the model if it is significant at the specified significance level. With a stepwise regression, parameters that were already added to the model can still be removed. The parameters are added and removed in such a way that every forward selection can be followed by backward removals of parameters. Once all of the variables were added to the model, or if the model is the same as the previous model, the stepwise regression stops (SAS Institute Inc., 2010).

Not all 25 parameters listed in Table 3-1 were added to the stepwise regression simultaneously. The parameters from each of the 6 main groups were added in turn to the stepwise regression models, in order to identify the parameter that predicted lightning occurrence the best. As an example, all the CAPE parameters were added to the stepwise regression model from where the CAPE that performed the best were selected. The most appropriate parameter was selected by making use of the chi-square value from the score chi-square statistic. The following procedure was followed in this study:



- 1) Add the parameters from each main group to the logistic procedure.
- 2) Calculate the score chi-square statistic for each variable. The parameter with the highest chi-square statistic is considered the best and is added to the model.
- 3) Add/remove parameters to the one with the highest chi-square value in step 2. Once again, the best model that contains two variables is selected based on the highest chi-square value.
- 4) Step 3 is repeated until all the variables were added to the model or if the model is the same as the previous.

Since the aim in this study was to select one parameter out of the six main groups that predicts lightning the best, only the first two steps were of interest. As such, the variables out of a main group was added to the stepwise regression, and the one with the highest chi-square value was considered the best in that group.

### 3.5.1.2 Stepwise regression with R

A stepwise regression with R-software was performed in order to select the best parameters to predict lightning. Not all 25 parameters listed in Table 3-1 were added to the stepwise regression simultaneously. The parameters from each of the 6 main groups were added in turn to the stepwise regression models, in order to achieve the goal of selecting one parameter out of each main groups that performs the best for lightning prediction. As an example, all the CAPE parameters were added to the stepwise regression model individually from where the CAPE that performed the best were selected. The most appropriate parameter was selected by making use of the Akaike Information Criterion (AIC).

The AIC is one of the most common model or variable selection techniques used in many statistical software (Chaurasia and Harel, 2012). The AIC utilises the Kullback-Leibler divergence, which is a measure of how close a model is to reality, to determine how much information is lost when using a model to predict reality (Posada and Buckley, 2004). It is a simple measure that is easily applied, but is based on sophisticated statistical techniques (Burnham and Anderson, 2004). The model having the lowest AIC value represent the model, which is closest to reality (Snipes and Taylor, 2014). The AIC can be computed by Equation 3-1:

$$AIC = 2K - 2L \quad (3-1)$$

where  $K$  is the number of estimable parameters and  $L$  is the maximised log-likelihood (Posada and Buckley, 2004). A single value of the AIC has no meaning and needs to be used together with other AIC values. AIC values are also affected by the size of the samples used in the model and can take on a range of values. As such, the AIC values are often re-written with Equation 3-2:

$$\Delta_i = AIC_i - AIC_{min} \quad (3-2)$$

where  $AIC_{min}$  is the lowest AIC value from all the models with the best model will have a  $\Delta$  value of 0 (Burnham and Anderson, 2004).

In this study, the “glm” and “step” functions from the R-package “stats” were utilised to perform the stepwise logistic regression for selecting the best parameters (R Development Core Team, 2015). The “glm” package performed the binary logistic regression between lightning occurrence and the parameters, while the initial model output from the “glm” package was used in the “step” function. The AIC values from the “step” function were written to file and could be used to identify the parameter that performed the best to predict lightning.

### 3.5.2 Development of LTI with logistic regression

Logistic regression techniques were used to develop the new LTI. Logistic regression is often used to predict the probability of an event by means of a set of predictors (Kiezun et al., 2009) and can be expressed by Equation 3-3, where  $p_i$  is the probability of the event as a function of  $m$  independent variables  $X$ , when  $i$  ranges from 1 to  $m$ . The regression coefficients,  $\hat{\alpha}$  and  $\hat{\beta}$ , are estimated from the dataset by means of the maximum likelihood method (Guns and Vanacker, 2012; Kleinbaum and Klein, 2010).

$$p_i = \frac{1}{1 + e^{-(\hat{\alpha} + \sum \hat{\beta}_i X_i)}} \quad (3-3)$$

In the development of the LTI by means of the logistic regression technique, lightning was the dependent variable, while the six model parameters selected in Section 3.5.1 were

the independent variables. The daily lightning and associated model data were concatenated into a single text file. Table 3-2 shows an example of the format of the text file. All of the daily data for the 2011 and 2012 SON period (in the format of Table 3-2) were then concatenated into one text file. Similarly, all the daily data for the 2011/12 and 2012/13 DJF period were added into one text file. This produced two large text files for the regression analysis to train the LTI model.

**Table 3-2: Example of dataset used as input to the regression analysis to build the LTI**

Lightning (yes/no)	Model Parameters					
	CAPE	PW	LI	$\Theta_e$	RH	T
1	1654.4	2.8	-4.3	-20.2	47.4	291.5
1	1610.5	2.9	-4	-19.9	52	291.1
0	1370	2.9	-4.6	-18	55.2	290.9

Initial tests in the development of the LTI by means of logistic regression revealed that probabilities of lightning occurrence are extremely low. Upon further investigation it was discovered that a study by King and Zeng (2001) showed that ordinary logistic regression, as indicated by Equation 3-3, often underestimates the probabilities of rare events (Guns and Vanacker, 2012). This underestimation is due to the logistic regression favouring the larger amount of non-events (0's) compared to the smaller amount of events (1's) when developing a model. King and Zeng (2001) states that rare events in a dataset are classified as dozens to thousands of times more non-events compared to events, while Yap et al. (2014) considers a rare event to be when the events make up 5% or less of the data. In the datasets considered in this study, there were approximately 20 times more non-events than events for the SON dataset, and approximately 34 times more non-events than events for the DJF dataset. In both datasets, the non-events made up less than 5% of data.

King and Zeng (2001) developed an approach to perform a rare event logistic regression. This 3-step approach is summarised by Guns and Vanacker (2012) as follows: 1) Make use of endogenous stratified sampling by taking all events together with a random sample of non-events. 2) Apply a correction to the intercept term that may be significantly biased due to sampling. 3) The underestimation of probabilities is taken into account by adding the correction term to the estimated probabilities. An R software package, called 'Zelig' exists which performs the above-mentioned corrections to the probabilities (Imai et al., 2008; 2009).

The LTI was developed by means of the following approach:

1. Take all the events (1's or lightning occurrences) in the dataset and select a random sample of non-events (0's or no lightning occurrences) with equal size from the data.
2. Run the 'Zelig' package in R to perform a rare event logistic regression with the bias correction and addition of the correction term to the estimated probabilities.
3. Repeat the above steps 1000 times by selecting a new sample of random non-events (0's or no lightning occurrences). The random samples of non-events are taken with repetition where the non-events of the previous sample are added back to the dataset and have the chance to be chosen again.

The 1000 models produced by the procedure above were combined by averaging their output. This means that the average of the intercept term and regression coefficients of the 1000 models were calculated. This process is similar to the bootstrap aggregating technique that aims to improve any instability found in the estimation of the regression output (Kotsiantis et al., 2006). The average of the intercept term and regression coefficients could then be added to Equation 3-3. Separate equations for the SON and DJF periods were developed with this approach by performing the rare event logistic regression on each of the SON and DJF datasets.

## **3.6 VERIFICATION OF THE LTI**

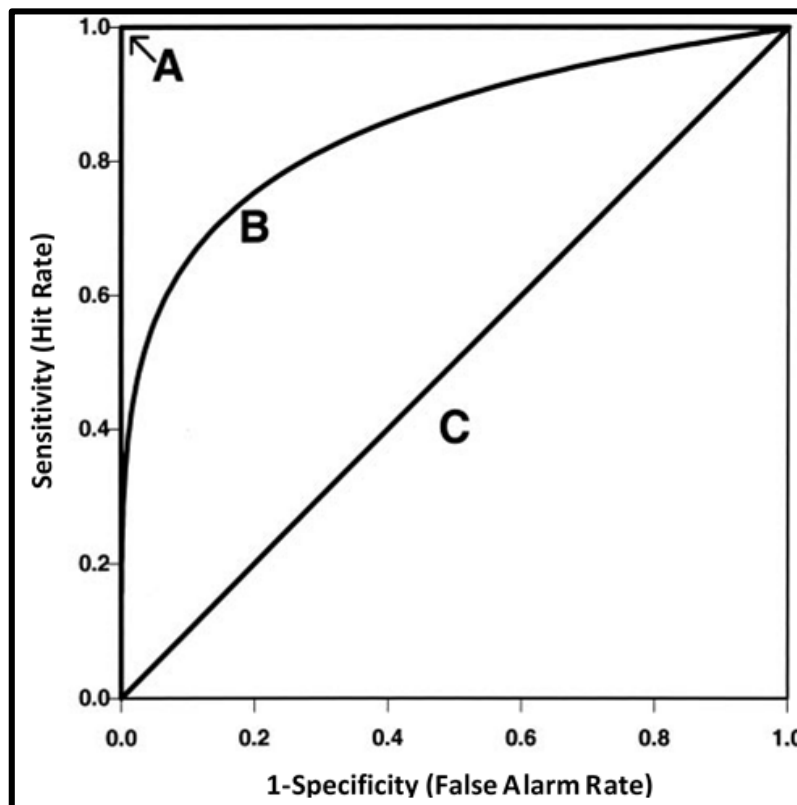
### **3.6.1 Probabilistic verification**

#### **3.6.1.1 ROC curve**

The Receiver Operating Characteristic (ROC) curve is a useful tool to determine the performance of a probabilistic forecast (Mason and Graham, 2002). ROC curves are used to compare the sensitivity and specificity of a forecast over the range of all possible values (Florkowski, 2008). Sensitivity is the ability of a forecast to predict events, while specificity is the ability of the forecast to predict the non-events (Robin et al., 2011). The ROC curve is created by using corresponding forecasts and observations to plot the sensitivity (hit rate) against 1-specificity (false alarm rate) (Mason and Graham, 2002). The ideal forecast would have a sensitivity and specificity of 1, which will be a curve starting at 0 on the x and y axis, moving to a value of 1 on the y-axis and a value of 0 on the x-axis, and ending at 1 on the x and y axis (Bewick et al., 2004) (A on Figure 3-7). It is however unlikely that this ideal forecast can be achieved (Fan et al., 2006). Forecasts with good predictive skill approach the top left

corner of the ROC plot (Curve B in Figure 3-7), while the diagonal line (Curve C in Figure 3-7) indicates a forecast with no skill (Zou et al., 2007). A ROC curve is often accompanied by the area under the ROC curve (AUC), which is a single value that gives a representation of the overall performance of the forecast (Fawcett, 2006). For curve A in Figure 3-7, the AUC will be 1.0 and for curve B it will be 0.85 (Zou et al., 2007). The diagonal line C in Figure 3-7 will have an AUC of 0.5, but since this line represents a random guess with no skill, no realistic AUC value should be 0.5 or less (Fawcett, 2006).

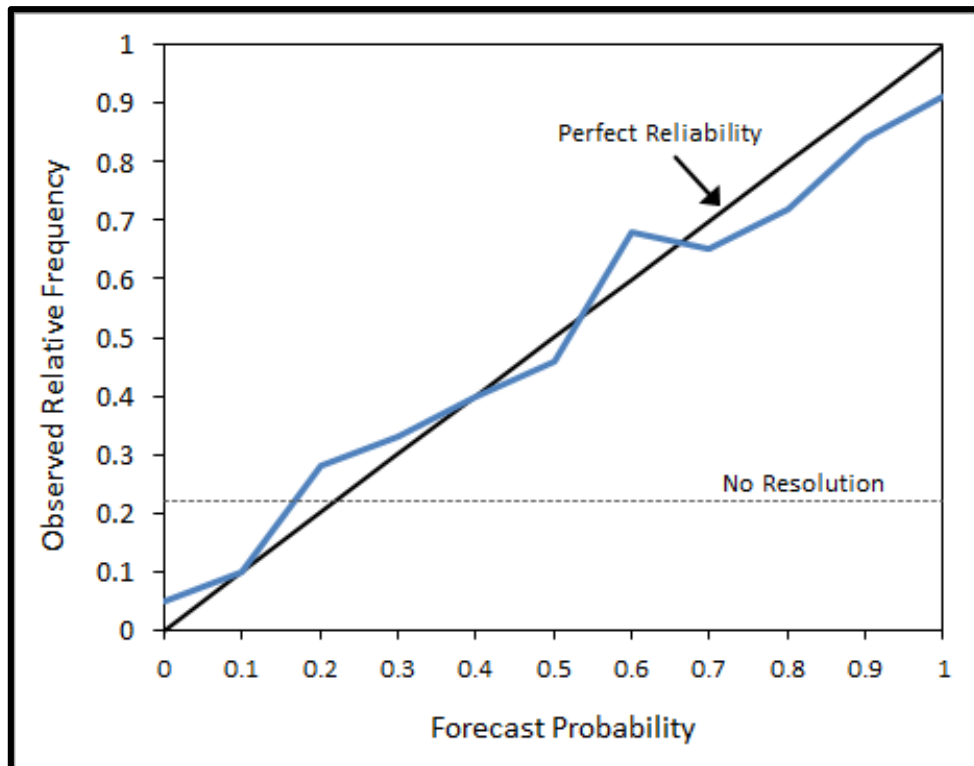
In this study, the “plot.roc” function in the R-package “pROC” was utilised to plot the ROC curves and AUC values for the validation of the LTI (Robin et al., 2011; R Development Core Team, 2015). The lightning observations (1’s for lightning occurrence and 0’s for no lightning occurrence) and LTI forecast probabilities (probability values represented by values between 0 and 1) were incorporated into the “plot.roc” function. The daily forecasts and observations were combined into one dataset for the entire evaluation periods. A ROC plot was created for both the SON and DJF periods.



**Figure 3-7: An explanation of a ROC curve. Line A represents a perfect forecast (AUC=1), line B a typical ROC curve (AUC=0.85) and line C represents no skill or random chance (AUC=0.5) [Adapted from Zou et. al., 2007]**

### 3.6.1.2 Reliability diagram

A reliability diagram is often used to determine the reliability of a probabilistic forecast by showing how well forecasted probabilities correspond to their observed frequency of occurrence (Weisheimer and Palmer, 2014). As such, the reliability diagram is created by plotting the observed relative frequencies against forecast probabilities, where the forecast probabilities are divided into bins (Bröcker and Smith, 2007). A perfectly reliable forecast will follow the diagonal line as indicated in Figure 3-8 (Lott et al., 2014). The closer a curve lies to the diagonal line, the more reliable a forecast. If a curve lies below the diagonal line it indicates over-forecasting, and if the curve is above the diagonal line, it indicates under-forecasting. A curve that follows the horizontal line will have no resolution since the forecast cannot discriminate from one probability bin to next (Holliday et al., 2012).



**Figure 3-8: An example of a reliability diagram. The blue line represents a hypothetical forecast, the black diagonal line will be a forecast with perfect reliability, and the horizontal dotted line will be a forecast with no resolution.**

In this study the “verify” and “reliability.plot” functions in the R-package “verification” were utilised to plot the reliability diagrams for the validation of the LTI (NCAR - Research Applications Laboratory, 2015; R Core Team, 2015). The lightning observations (1’s for

lightning occurrence and 0's for no lightning occurrence) and LTI forecast probabilities (probability values represented by values between 0 and 1) were incorporated into the “verify” function to create a class object to be used in the “reliability.plot” function (NCAR - Research Applications Laboratory, 2015). The daily forecasts and observations were combined into one dataset for the entire evaluation periods. A reliability plot was created for both the SON and DJF periods.

### 3.6.1.3 Sharpness diagram

The sharpness of a forecast is a measure of how forecast probabilities varies and is often presented on a sharpness diagram or sharpness histogram which displays the relative frequencies of occurrence for probability intervals (bins). Sharpness diagrams often accompany reliability plots (Callado et al., 2013). A probabilistic forecast has perfect sharpness when only probabilities of 0 and 1 are forecasted, good sharpness with a U-shaped distribution (values close to 0 and 1), and a lack of sharpness when probabilities are uniformly distributed (Murphy and Wilks, 1998). Figure 3-9 shows an example of a typical sharpness diagram.

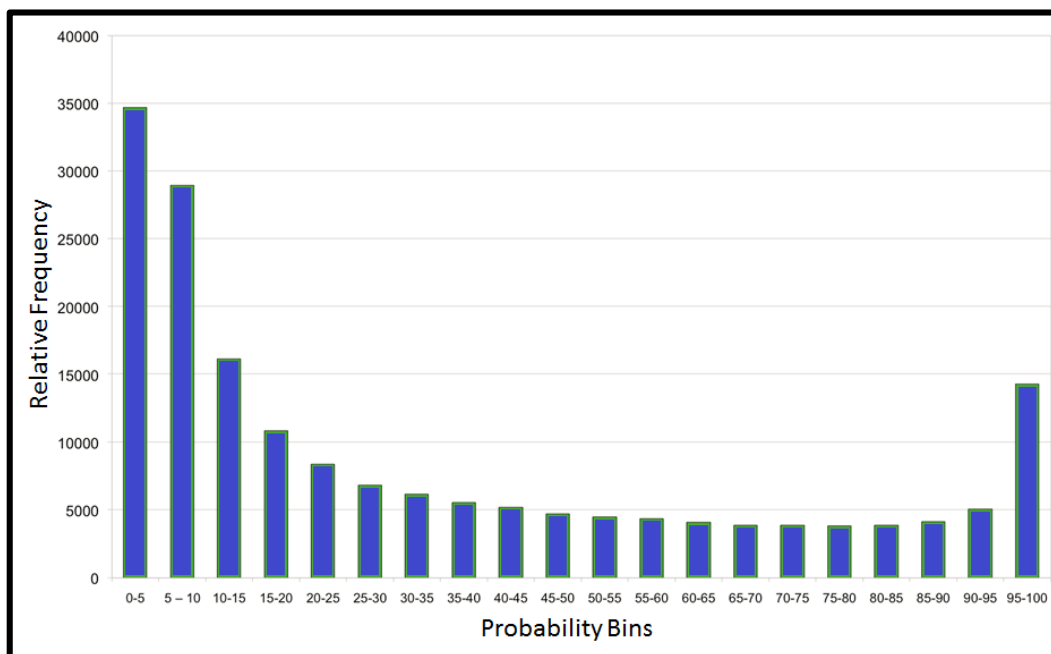


Figure 3-9: Example of a sharpness diagram [Adapted from Cintineo et al., 2010]

In this study the “verify” and “reliability.plot” functions in the R-package “verification” were modified to plot the sharpness diagrams separately from the reliability plots for the validation of the LTI (NCAR - Research Applications Laboratory, 2015; R Development Core Team, 2015). The lightning observations (1’s for lightning occurrence and 0’s for no lightning occurrence) and LTI forecast probabilities (probability values represented by values between 0 and 1) were incorporated into the “verify” function to create a class object to be used in the “reliability.plot” function (NCAR - Research Applications Laboratory, 2015). The daily forecasts and observations were combined into one dataset for the entire evaluation periods. A sharpness plot was created for both the SON and DJF periods.

### 3.6.2 Dichotomous verification

In addition to the standard probabilistic verification techniques, the LTI was also evaluated by means of dichotomous verification techniques. The performance of a yes/no-dichotomous forecast is evaluated by means of a contingency table approach that contains hits, misses, false alarms and correct negatives (Tartaglione, 2010). Figure 3-10 shows a standard 2 X 2 contingency table. The performance of a forecast is measured by the frequency of the ‘yes’ and ‘no’ forecasts compared to the ‘yes’ and ‘no’ observations (Landman et al., 2012). From the contingency table verification scores like the Probability of Detection (POD), Probability of False Detection (POFD), False Alarm Ratio (FAR), Hanssen and Kuipers Discriminant (HK) and Frequency bias can be calculate. These skills scores are discussed in section 3.6.2.1 to 3.6.2.5.

		OBSERVED	
		YES	NO
FORECAST	YES	HITS	FALSE ALARMS
	NO	MISSES	CORRECT NEGATIVES

Figure 3-10: A standard 2 x 2 contingency table



Since the LTI is a probabilistic forecast, the dichotomous verification had to be performed over the entire range of probabilities. A similar technique was used by de Coning et al. (2011). In this approach, probability thresholds in intervals of 10% were considered for the calculation of skill scores. The first threshold will be 0% where all areas exceeding 0% probability is considered the forecast. This will be the entire domain where all observed lightning would occur in the forecasted area. As such, the POD will be 100% but the FAR will also be high. The next threshold will be all areas exceeding 10% probability, then 20% and so on. By the time, the probability threshold is for example 80%, the POD will be low but the FAR will be low as well, since lightning will also occur in areas where the probability is below 80%. The ultimate aim of this evaluation approach will be to find the areas where the POD is high and the FAR is low in order to establish at which probabilities of the LTI it is the mostly likely to see lightning.

### 3.6.2.1 Probability of Detection

The POD or hit rate measures the percentage of observed events that were correctly forecasted. The POD can range between 0 and 1, where 1 is a perfect score (Jolliffe and Stephenson, 2003). The POD can be expressed by Equation 3-4:

$$POD = \frac{HITS}{HITS+MISSES} \quad (3-4)$$

### 3.6.2.2 Probability of false detection

The POFD or false alarm rate indicates what percentage of the observed non-events were incorrectly forecasted. The POFD can range between 0 and 1 where 0 is a perfect score (Jolliffe and Stephenson, 2003). The POFD is given by Equation 3-5:

$$POFD = \frac{FALSE ALARMS}{CORRECT NEGATIVES+FALSE ALARMS} \quad (3-5)$$

### 3.6.2.3 False alarm ratio

The FAR is the percentage of forecasted events that did not occur. The FAR can range between 0 and 1, where 0 is as a perfect score (Jolliffe and Stephenson, 2003). The FAR can be expressed by Equation 3-6:

$$FAR = \frac{FALSE\ ALARMS}{HITS + FALSE\ ALARMS} \quad (3-6)$$

### 3.6.2.4 Hanssen and Kuipers Discriminant

The HK or true skill score is a measure of how well a forecast distinguishes between the yes and no events. The HK can range from -1 to 1, where 0 indicates no skill and 1 is a perfect score (Weusthoff and Ament, 2010). The HK can be calculated with Equation 3-7:

$$HK = POD - POFD = \frac{HITS}{HITS + MISSES} - \frac{FALSE\ ALARMS}{CORRECT\ NEGATIVES + FALSE\ ALARMS} \quad (3-7)$$

### 3.6.2.5 Frequency Bias

The Frequency bias is the ratio between the forecasted events and the observed events. A perfect forecast has a bias of 1 (Jolliffe and Stephenson, 2003). A bias less than 1 indicates under-forecasting, while a bias greater than 1 shows over-forecasting. The bias can be calculated with Equation 3-8:

$$BIAS = \frac{HITS + FALSE\ ALARMS}{HITS + MISSES} \quad (3-8)$$

## 3.6.3 Eyeball verification

Sections 3.6.1 and 3.6.2 deals with the statistical evaluations of the LTI against lightning and gives a quantitative evaluation of the overall performance of the model. Another

verification method is the old-fashioned eyeball method that visually compares a forecast against an observation (WWRP/WGNE Joint Working Group on Verification, 2015).

The methodology of the LPI developed by Frisbie et al. (2009), which inspired this work, has been tested over South Africa and proved useful in certain conditions (Gijben, 2013). As was discussed in Chapter 2, the atmospheric conditions as well as the NWP models in South Africa and Colorado are different, which necessitated the development of the LTI for local conditions. It seems reasonable then to compare the Frisbie et al. (2009) LPI with the newly developed LTI for South Africa to determine if an improved lightning prediction model was developed. The LPI forecast assesses the lightning risk (low, moderate, severe and extreme) for the day while the newly developed LTI gives the probability of lightning occurrence. This makes a statistical comparison between the two difficult since they both provide different types of forecasts.

The UM model produces daily forecasts of convective rainfall. Just like lightning, convective rainfall is associated with thunderstorm cloud (cumulonimbus) dynamics and microphysics (Petersen and Rutledge, 1998). It also seems reasonable to compare the newly developed LTI against the convective rainfall from the UM. This is to determine if a superior product to the standard convective rainfall from the UM was developed that can predict the areas where lightning will occur. The convective rainfall product forecasts the amount of rainfall in cm that can be expected, and as such provides a different type of forecast than both the LPI and LTI.

By considering the differences in the types of forecasts discussed above, eyeball verification between the LPI, LTI, convective rainfall, and lightning occurrence was conducted by means of case studies to determine which of the three products performs the best. In total six cases were considered covering the SON months of 2013 and DJF months of 2013/14. One case from each of the months was randomly selected where significant amounts of lightning were observed. Table 3-3 shows the cases considered:

**Table 3-3: Case studies selected for the eyeball verification of the LPI, LTI and convective rainfall against the occurrence of lightning**

19 September 2013	20 December 2013
10 October 2013	15 January 2014
7 November 2013	6 February 2014

### 3.7 SUMMARY

This chapter presented the data and methodologies used in this study. It started with an overview of the study period and domain, followed by a description of the lightning and NWP data utilised. The methods applied in the development of the new LTI were presented, which includes a description of the techniques used for the selection of the most appropriate parameters for lightning prediction as well as the development of the equations for the new LTI. The chapter ended with the evaluation methods used for the verification of the newly developed LTI.

---

# CHAPTER 4

## DEVELOPMENT OF A NEW LIGHTNING THREAT INDEX

---

### 4.1 INTRODUCTION

This chapter deals with the development of the new LTI. The LTI was developed with CG lightning data from the SALDN as well as NWP data from the UM. Data from the 2011 and 2012 SON and the 2011/12 and 2012/13 DJF seasons was utilised for the development of the LTI forecast over the daily period between 07:00 – 21:00 UTC. Before the new index could be developed, the most appropriate model parameters had to be selected from Table 3-1 that will predict the occurrence of lightning over South Africa the best. This was achieved by means of stepwise logistic regression techniques. Six model parameters were selected for both SON and DJF. These selected parameters could then be used in the development of the LTI, where a rare event logistic regression technique was utilised to produce 1,000 regression models. The models were combined to produce a single lightning prediction model from where the regression output was used to develop the equation of the new LTI. The LTI predicts the probability of lightning occurrence and an equation was produced for SON as well as DJF.

### 4.2 PARAMETER SELECTION

Before the new LTI could be developed, it was necessary to select the most appropriate NWP model parameters to include in the LTI model. Frisbie et al. (2009), whose work formed the basis of this study, utilised a form of CAPE, LI, PW, RH,  $\Theta_e$  and air temperature in the development of his LPI for Colorado in the USA. More specifically they utilised the  $\mu\text{CAPE}_{0,3\text{km\_AGL}}$ , LI, total column PW,  $\text{RH}_{m10}$ ,  $\Theta_e\Gamma_{600}$ , and the  $T_{850}$ . However, the most appropriate combination of variables for South African conditions and operational NWP model had to be identified. The parameters utilised in this study are listed in Table 3-1 and are similar to those used by Frisbie et al. (2009), but different variations of these parameters were considered.

In this study, the goal was to select the best performing CAPE, LI, PW, RH,  $\Theta_e$  and air temperature capable of predicting the occurrence of lightning over South Africa that could ultimately be used in the development of the new LTI. To achieve this goal, both SAS and R

software's were used to select the most appropriate parameters for lightning prediction by making use of stepwise logistic regression techniques. The SAS and R methodologies are similar, and serve the same purpose of selecting the most appropriate model parameters to include in the LTI. Both methodologies were considered to serve as an additional confirmation that correct parameters are chosen for the development of the LTI.

In both the SAS and R procedures, separate stepwise regression models were considered for each main group of parameters. As an example, a stepwise regression was performed on all the CAPE parameters. The CAPE parameter that predicted lightning occurrence the best in the regression model was selected to be the most appropriate CAPE that could be used in the LTI. This was repeated for all the main groups. The candidate predictors from each main group of NWP parameters were added one-by-one to the regression model. After all of the parameters in a main group were added individually to the regression model, a second parameter was added to the model by means of forward and backwards addition/elimination, to produce two-parameter models. This continued until all the parameters from main groups were added to the regression model. Only the results from the one-parameter model of the regression procedure will be shown since only these results are of interest to select the top performing parameter from each main group.

Lightning occurrence, represented by the binary outcome of 1 (lightning) and 0 (no lightning), was the dependent variable in both the SAS and R stepwise regression procedures, while the NWP model parameters from the UM were the independent variables. The regression procedures were performed on SON and DJF separately in order to determine the most appropriate parameters to predict lightning for the different seasons. This was necessary since the atmospheric conditions in South Africa are different between the seasons. In early summer, the atmospheric circulation is generally extra-tropical with a conditionally unstable atmosphere over certain parts of South Africa, while in late summer the circulation is tropical with a convectively unstable atmosphere (Dyson et al., 2015). These differences in the atmospheric circulation between seasons can result in the selection of different parameters to include in the LTI.

In order to select the most appropriate NWP parameter to predict the occurrence of lightning in the SAS regression procedure, the score chi-square statistic was calculated for each NWP parameter added to the regression model. The parameter with the highest chi-square statistic was selected to be the most appropriate parameter to predict lightning occurrence. A comparison between the chi-square value from the best one-parameter model and the chi-square value of the model with all the parameters from a main group added will be discussed. The output from the SAS regression models will be shown in tables that are

sorted numerically, where the parameter with the highest chi-square score is located at the top and the one with the lowest chi-square score at the bottom of the table i.e.: Table 4-1

Just like with the SAS procedure, R software was also used select the most appropriate NWP model parameters to predict the occurrence lightning and to confirm the results found with the SAS regression procedure With the R regression procedure, the parameter with the lowest AIC score was selected to be the most appropriate parameter to predict lightning occurrence. Unlike with SAS, the stepwise regression in R did not continue with the addition of more than one parameter from a main group to the model. Only the first step of the stepwise regression was of interest since the aim was to select the one-parameter model that predicted lightning occurrence the best. The output from the R regression model will also be shown in tables that are sorted numerically. In these tables, the parameter with the lowest AIC value is located at the top and the one with the highest AIC at the bottom of the table i.e.: Table 4-2

The output from the SAS and R stepwise regression models will now be discussed and results presented according to parameter group. At the end of Section 4.2, a summary will be provided on the six parameters selected for the SON and DJF seasons.

#### 4.2.1 Convective Available Potential Energy (CAPE)

The first stepwise regression analysis was performed on all the CAPE parameters. Four different CAPE parameters were added together with lightning occurrence to the stepwise logistic regression models. The CAPE parameters considered in the analysis were the most unstable CAPE in the 0-3 km AGL range ( $\text{muCAPE}_{0,3\text{km\_AGL}}$ ), most unstable CAPE in the 1-6 km AGL range ( $\text{muCAPE}_{1,6\text{km\_AGL}}$ ), most unstable CAPE between the surface and 300 hPa level ( $\text{muCAPE}_{\text{lowest},300}$ ), and finally the surface CAPE ( $\text{CAPE}_{\text{surf}}$ ).

From the SAS stepwise regression analysis,  $\text{muCAPE}_{1,6\text{km\_AGL}}$  had the highest score chi-square statistic for both SON and DJF (Table 4.1). This was followed by  $\text{muCAPE}_{\text{lowest},300}$  and  $\text{muCAPE}_{0,3\text{km\_AGL}}$ , which had a chi-square score very similar to  $\text{muCAPE}_{1,6\text{km\_AGL}}$ .  $\text{CAPE}_{\text{surf}}$  had the lowest score chi-square statistic and this value was much lower compared to the other CAPE variables.  $\text{muCAPE}_{1,6\text{km\_AGL}}$  had a chi-square score of 5,927.18 for SON and 5,958.34 for DJF. When all four of the CAPE parameters were added simultaneously to the regression model, the chi-square score was 6,005.20 for SON and 5,975.28 for DJF. As such, the chi-square score increased by 78.02 for SON and by 16.94 for DJF. These small improvements to the chi-square scores show that the inclusion of all four of the CAPE parameters into the model did not add additional predictive value to the model compared to when only

muCAPE<sub>1,6km\_AGL</sub> was used to predict the occurrence of lightning. Based on these results, the muCAPE<sub>1,6km\_AGL</sub> parameter performed the best with the SAS regression model process and was selected to be the most appropriate parameter to predict lightning occurrence.

**Table 4-1: SAS output from the stepwise logistic regression model for all CAPE variables.**

Regression Models Selected by Score Criterion			
SON		DJF	
Chi-Square	Variables Included in Model	Chi-Square	Variables Included in Model
5,927.18	muCAPE <sub>1,6km_AGL</sub>	5,958.34	muCAPE <sub>1,6km_AGL</sub>
5,926.20	muCAPE <sub>lowest,300</sub>	5,957.69	muCAPE <sub>lowest,300</sub>
5,926.19	muCAPE <sub>0,3km_AGL</sub>	5,957.69	muCAPE <sub>0,3km_AGL</sub>
5,404.34	CAPE <sub>surf</sub>	5,569.30	CAPE <sub>surf</sub>

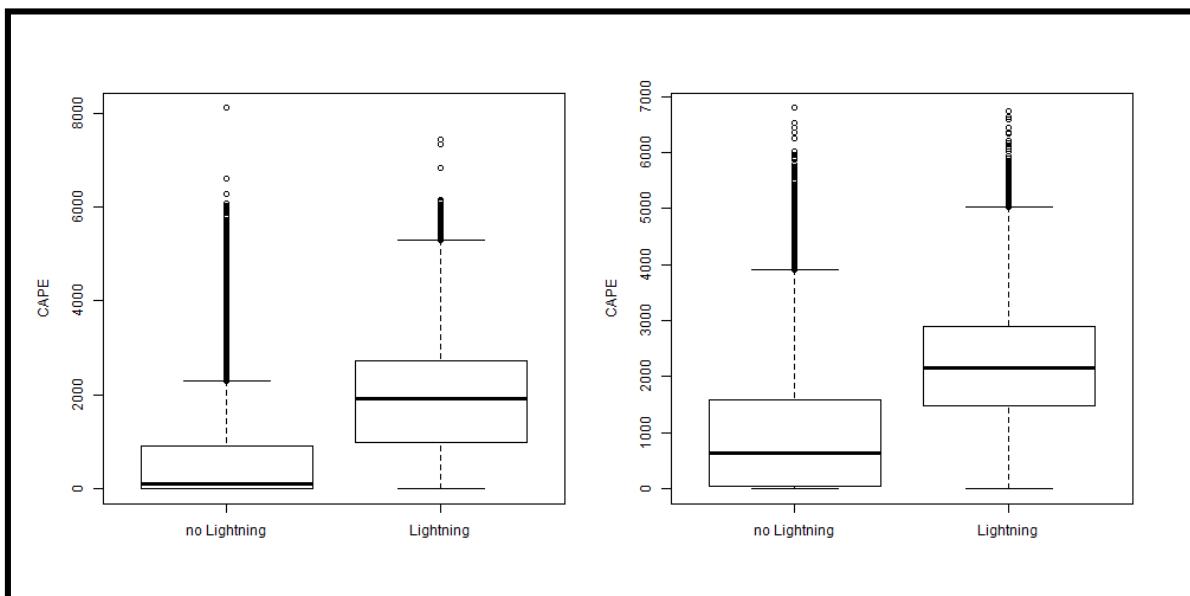
In the R regression analysis, the muCAPE<sub>1,6km\_AGL</sub> parameter once again performed the best since it had the lowest AIC score for both SON and DJF (Table 4-2). This was followed by muCAPE<sub>lowest,300</sub> and muCAPE<sub>0,3km\_AGL</sub>, which had an AIC value very similar to muCAPE<sub>1,6km\_AGL</sub>. CAPE<sub>surf</sub> had the highest AIC value and was higher compared to the other CAPE parameters. The muCAPE<sub>1,6km\_AGL</sub> had an AIC score of 88,602.42 for SON and 126,330.31 for DJF, while CAPE<sub>surf</sub> had an AIC value of 5,404.31 and 5,569.30 for the SON and DJF periods respectively.

**Table 4-2: R output of the stepwise logistic regression model for all CAPE variables.**

Regression Models Selected by Score Criterion			
SON		DJF	
Score AIC	Variables Included in Model	Score AIC	Variables Included in Model
88,602.42	muCAPE <sub>1,6km_AGL</sub>	126,330.31	muCAPE <sub>1,6km_AGL</sub>
88,608.10	muCAPE <sub>lowest,300</sub>	126,332.75	muCAPE <sub>0,3km_AGL</sub>
88,608.13	muCAPE <sub>0,3km_AGL</sub>	126,332.76	muCAPE <sub>lowest,300</sub>
90,488.28	CAPE <sub>surf</sub>	128,122.26	CAPE <sub>surf</sub>



Based on the results from the stepwise regression with SAS and R,  $\mu\text{CAPE}_{1,6\text{km\_AGL}}$  was selected to be the best CAPE to be used when predicting lightning occurrence since it had the highest chi-square score and the lowest AIC value. As was discussed in Section 2.5.3.1 of Chapter 2, CAPE is strongly related to the updraft velocities in thunderstorms and plays an important role in the distribution of hydrometeors responsible for lightning formation (Murugavel et al., 2014; Singh and O’Gormon, 2015). Sufficient values of CAPE are required in the  $0^{\circ}\text{C}$  to  $-20^{\circ}\text{C}$  level of a thunderstorm to ensure that the updraft provides the hydrometeors necessary for electrification to occur (Bright et al., 2005). In Figure 2-1, it was shown that the  $0^{\circ}\text{C}$  to  $-20^{\circ}\text{C}$  level in a South African thundercloud is found approximately 3-6 km above ground level. Since CAPE is closely related to updraft velocities in thunderclouds, sufficient CAPE is required as low as 1 - 3 km AGL to feed the charge separation zone of the storm (3 - 6 km AGL) with hydrometeors for electrification to occur. Updrafts are also important inside the separation zone to transport the positively charged ice particles to the top of the cloud. This can explain why the  $\mu\text{CAPE}_{1,6\text{km\_AGL}}$  parameter performed the best since strong updrafts are required below and inside the charge separation zone of a thundercloud.



**Figure 4-1: Box and whisker plots of  $\mu\text{CAPE}_{1,6\text{km\_AGL}}$  between 07:00 – 21:00 UTC for lightning and no Lightning. SON is shown on the left and DJF on the right. CAPE units are in  $\text{J kg}^{-1}$ . The thick horizontal bar indicates the median value; the boxes denote the 25<sup>th</sup> –75<sup>th</sup> percentiles, and the whiskers show the full range of values. The circles indicate outliers.**

During SON, 75% of the  $\mu\text{CAPE}_{1,6\text{km\_AGL}}$  values were  $\leq 1,000 \text{ J kg}^{-1}$  when no lightning occurred, while 75% of the  $\mu\text{CAPE}_{1,6\text{km\_AGL}}$  values were  $\geq 1000 \text{ J kg}^{-1}$  when lightning did occur

(Figure 4-1). For the DJF period, 75% of the  $\text{muCAPE}_{1,6\text{km\_AGL}}$  values were  $\leq 1,750 \text{ J kg}^{-1}$  or less when no lightning occurred, while 75% of the  $\text{muCAPE}_{1,6\text{km\_AGL}}$  values were  $\geq 1,750 \text{ J kg}^{-1}$  or more when lightning occurred. The  $\text{muCAPE}_{1,6\text{km\_AGL}}$  is useful to distinguish between the occurrence of lightning and no lightning. Figure 4-1 provides additional confirmation that this parameter is a good choice for lightning prediction over South Africa. During the SON period,  $\text{muCAPE}_{1,6\text{km\_AGL}}$  values above  $1,000 \text{ J kg}^{-1}$  can be a good indicator that lightning might occur, while  $\text{muCAPE}_{1,6\text{km\_AGL}}$  values of  $1,750 \text{ J kg}^{-1}$  or more during the DJF season can be indicative of lightning formation.

#### 4.2.2 Lifted Index (LI)

The next parameter to be considered was the LI. Two different types of LI parameters were added together with lightning occurrence to the stepwise logistic regression models, and were the surface lifted index (SLI) and the best lifted index (BLI) (Table 3-1).

The two LI parameters were added to the SAS model from where the score chi-square statistic was calculated. SLI had the highest score chi-square statistic during SON and DJF, while the score for BLI was considerably lower than that of the SLI (Table 4-3). The chi-square value of the SLI was 4,067.59 for SON and 4,470.85 for DJF. The BLI values of 2,088.78 and 1,873.81 for SON and DJF were much lower. When both the SLI and BLI were added to the regression model, the score chi-square statistic was 4,516.43 for SON and 4,509.48 for DJF. As such, the chi-square score increased by 448.84 for SON period and by 38.63 for DJF. This shows that the inclusion of both forms of the LI into the model does not add additional predictive value, especially for DJF. The SLI was selected to be the best LI to predict lightning occurrence.

**Table 4-3: SAS output of the stepwise logistic regression models for all LI variables.**

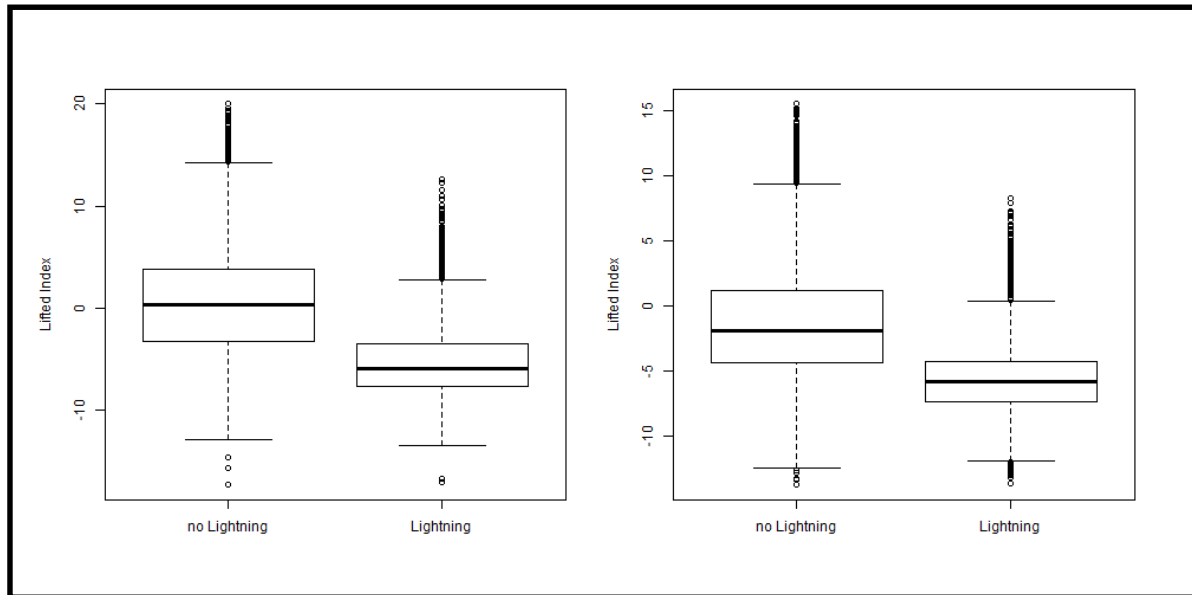
Regression Models Selected by Score Criterion			
SON		DJF	
Chi-Square	Variables Included in Model	Chi-Square	Variables Included in Model
4,067.59	SLI	4,470.82	SLI
2,088.78	BLI	1,873.81	BLI

The two LI parameters were also added to the R model from where the AIC score was calculated. The SLI parameter had the lowest AIC score for both SON and DJF, while the BLI had the highest AIC value. The AIC value of BLI was considerably higher when compared to the SLI (Table 4.4). The SLI had an AIC score of 84,375.93 for SON, which is 12,498.69 lower than that of the BLI. For DJF, the SLI had an AIC value of 119,939.39, which is 19,323.13 lower than that of the BLI parameter. The AIC scores show that the SLI was the parameter that performed the best.

**Table 4-4: R output of stepwise logistic regression for all LI variables.**

<b>Regression Models Selected by Score Criterion</b>			
<b>SON</b>		<b>DJF</b>	
<b>Score AIC</b>	<b>Variables Included in Model</b>	<b>Score AIC</b>	<b>Variables Included in Model</b>
84,375.93	SLI	119,939.39	SLI
96,874.62	BLI	139,262.52	BLI

The SLI was selected to be the most appropriate form of the LI to predict lightning occurrence. As with CAPE, the LI is a good indicator of updraft strength and supplies the hydrometeors responsible for electrification to occur in a thundercloud (Singh and O’Gormon, 2015). As was discussed in Section 2.5.3.2, the SLI provides the difference between the actual temperature of the environment at 500 hPa and the temperature a parcel will have when it was lifted adiabatically from the surface to the 500 hPa level (Haklander and Van Delden, 2003). The 500 hPa level corresponds with the charge separation zone in a thundercloud. The BLI, used by Frisbie et al. (2013), is the most unstable LI obtained when a parcel is lifted from the levels between 850 hPa and 700 hPa up to the 500 hPa level. The reason why the BLI could have performed significantly worse than the SLI in this study can be due to some areas of South Africa being lower in pressure than the 850 hPa pressure level. As such, the 850 to 700 hPa layer will be higher up in the atmosphere over these areas resulting in the instability near the surface to be missed. The SLI will capture the instability in the lower levels of the atmosphere missed by the BLI over the areas of South Africa lower in pressure than 850 hPa level.



**Figure 4-2: Box and whisker plots of SLI between 07:00 – 21:00 UTC for lightning and no lightning. SON is shown on the left and DJF on the right. SLI units are in °C. The thick horizontal bar indicates the median value; the boxes denote the 25th –75th percentiles, and the whiskers show the full range of values. The circles indicate outliers.**

During SON, 75% of the SLI values were  $\geq -3^{\circ}\text{C}$  when no lightning occurred, while 75% of the SLI values were  $\leq -3^{\circ}\text{C}$  when lightning did occur (Figure 4-2). For DJF the same applied, except that the SLI value was  $-5^{\circ}\text{C}$ . This shows that the SLI parameter can in most cases be useful to distinguish between the occurrence of lightning, and gives additional confirmation that this parameter is a good choice for lightning prediction over South Africa. A SLI value of  $-3^{\circ}\text{C}$  or less during SON, and a SLI value of  $-5^{\circ}\text{C}$  or less during DJF may be used as thresholds for lightning to occur.

### 4.2.3 Precipitable Water (PW)

Three different PW parameters were added together with lightning occurrence to the stepwise logistic regression models. The PW parameters considered were the mean PW in the 850-300 hPa level ( $PW_{850,300}$ ), mean PW in the 700-400 hPa level ( $PW_{700,400}$ ), and mean PW from the surface to 100 hPa level ( $PW_{\text{surf},100}$ ).

In the SAS regression procedure, the three PW parameters were added individually to the model from where the score chi-square statistic was calculated. The analysis shows that  $PW_{850,300}$  had the highest score chi-square statistic for both SON and DJF. During SON and DJF, the chi-square score for  $PW_{700,400}$  and  $PW_{\text{surf},100}$  was considerably lower than that of

$PW_{850,300}$  (Table 4.5). This was especially true for  $PW_{surf,100}$ . The chi-square value of  $PW_{850,300}$  was 8,701.11 for SON and 5,680.31 for DJF, while the  $PW_{surf,100}$  had a chi-square score of 6,820.04 and 2,574.49 for SON and DJF respectively. If all three of the different PW parameters were added together in the regression model, the score chi-square statistic was 8,721.76 for SON and 5,940.14 for DJF. This shows that the chi-square score increase by 20.65 for SON and by 259.83 for DJF. When all three forms of PW are entered simultaneously into the model, no additional predictive value is achieved to the lightning prediction model compared to when only  $PW_{850,300}$  is utilised. This is true especially for SON. As such,  $PW_{850,300}$  was selected to be the most appropriate PW parameter to predict the occurrence of lightning.

**Table 4-5: SAS output of stepwise logistic regression for all PW variables.**

Regression Models Selected by Score Criterion			
SON		DJF	
Chi-Square	Variables Included in Model	Chi-Square	Variables Included in Model
8,701.11	$PW_{850,300}$	5,680.31	$PW_{850,300}$
7,559.84	$PW_{700,400}$	4,781.82	$PW_{700,400}$
6,820.04	$PW_{surf,100}$	2,574.49	$PW_{surf,100}$

**Table 4-6: R output of stepwise logistic regression for all PW variables.**

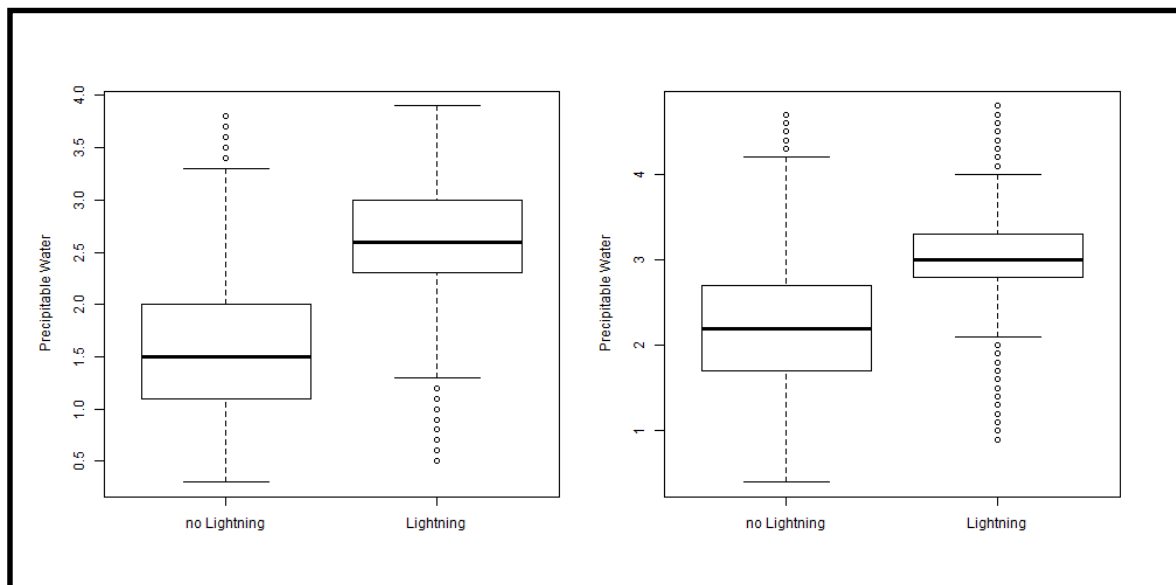
Regression Models Selected by Score Criterion			
SON		DJF	
Score AIC	Variables Included in Model	Score AIC	Variables Included in Model
71,881.75	$PW_{850,300}$	113,031.51	$PW_{850,300}$
77,236.62	$PW_{700,400}$	122,863.53	$PW_{700,400}$
89,262.38	$PW_{surf,100}$	145,142.50	$PW_{surf,100}$

$PW_{850,300}$  had the lowest AIC score for both SON and DJF and was the parameter that performed the best (Table 4-6).  $PW_{surf,100}$  had the highest AIC value and was considerably higher than the AIC value for  $PW_{850,300}$ .  $PW_{850,300}$  had an AIC score of 71,881.75 (17,380.63 lower than the  $PW_{surf,100}$ ) for SON and 113,031.51 (32,110.99 lower than the  $PW_{surf,100}$ ) for

DJF. Based on these results,  $PW_{850,300}$  was again the PW parameter that performed the best during SON and DJF and was selected to be the most appropriate PW parameter to use when predicting lightning occurrence.

Based on the result from the SAS and R regression models,  $PW_{850,300}$  was selected to be the most appropriate form of the PW to predict lightning occurrence. As was discussed in Chapter 2, moisture is one of the requirements for lightning formation (or thunderstorm formation) and PW provided the liquid water content within a layer in the atmosphere (Burrows et al., 2005; Duplika and Reuter, 2006).  $PW_{850,300}$  corresponds well with the charging zone in a thundercloud and shows that sufficient moisture is required between the 850 hPa and 300 hPa levels in the atmosphere for electrification to occur in storm.

During SON, 75% of the  $PW_{850,300}$  values were  $\leq 2$  cm when no lightning occurred, while 75% of the  $PW_{850,300}$  values were  $\geq 2.3$  cm when lightning did occur (Figure 4-3). For DJF, 75% of the  $PW_{850,300}$  values were  $\leq 2.8$  cm when no lightning occurred, while the  $PW_{850,300}$  values were  $\geq 2.8$  cm when lightning did occur. This shows that the  $PW_{850,300}$  parameter can in most cases be useful to distinguish between the occurrence of lightning, and gives additional confirmation that this parameter is a good choice for lightning prediction over South Africa. A  $PW_{850,300}$  value of  $\geq 2.3$  cm during SON season and  $\geq 2.8$  cm during DJF may be useful to determine if lightning will occur.



**Figure 4-3: Box and whisker plots of  $PW_{850,300}$  between 07:00 – 21:00 UTC for lightning and no Lightning. SON is shown on the left and DJF on the right. PW units are in cm. The thick horizontal bar indicates the median value; the boxes denote the 25th –75th percentiles, and the whiskers show the full range of values. The circles indicate outliers.**

#### 4.2.4 Relative Humidity (RH)

Five different RH parameters were added together with lightning occurrence to the stepwise regression models. The different variations of RH considered in this analysis were the RH at the  $-10^{\circ}\text{C}$  level ( $\text{RH}_{m10}$ ), mean RH between the  $-12^{\circ}\text{C}$  and  $-18^{\circ}\text{C}$  level ( $\text{RH}_{m12,m18}$ ), mean RH between 3-6 km AGL ( $\text{aveRH}_{3,6\text{km\_AGL}}$ ), maximum RH between 3-6 km AGL ( $\text{maxRH}_{3,6\text{km\_AGL}}$ ), and minimum RH between 3-6 km AGL ( $\text{minRH}_{3,6\text{km\_AGL}}$ ).

In the SAS regression procedure,  $\text{minRH}_{3,6\text{km\_AGL}}$  had the highest score chi-square statistic for SON, while  $\text{aveRH}_{3,6\text{km\_AGL}}$  had the highest score for DJF (Table 4-7). For both SON and DJF,  $\text{RH}_{m10}$ ,  $\text{RH}_{m12,m18}$  and  $\text{maxRH}_{3,6\text{km\_AGL}}$  performed considerably lower. In SON,  $\text{aveRH}_{3,6\text{km\_AGL}}$ , which was the top performing parameter in DJF, also performed considerably lower than  $\text{minRH}_{3,6\text{km\_AGL}}$ . During DJF however, the  $\text{aveRH}_{3,6\text{km\_AGL}}$  and  $\text{minRH}_{3,6\text{km\_AGL}}$  parameters were very similar. When all five of the RH parameters were added to the regression model, the score chi-square statistic was 7,025.88 for SON and 4,663.71 for DJF period. As such, the chi-square score increased by 154.44 for SON and by 279.65 for DJF. This shows that the inclusion of all five forms of the RH into the model does not add additional predictive value to the prediction of lightning occurrence compared to when only  $\text{minRH}_{3,6\text{km\_AGL}}$  and  $\text{aveRH}_{3,6\text{km\_AGL}}$  is used for SON and DJF. Based on these results from the analysis, the  $\text{minRH}_{3,6\text{km\_AGL}}$  parameter was selected for SON as the best form of RH to predict lightning occurrence, while during the DJF season the  $\text{aveRH}_{3,6\text{km\_AGL}}$  parameter was selected.

Table 4-7: SAS output of stepwise logistic regression for all RH variables.

Regression Models Selected by Score Criterion			
SON		DJF	
Chi-Square	Variables Included in Model	Chi-Square	Variables Included in Model
6,871.44	$\text{minRH}_{3,6\text{km\_AGL}}$	4,384.06	$\text{aveRH}_{3,6\text{km\_AGL}}$
5,789.37	$\text{aveRH}_{3,6\text{km\_AGL}}$	4,363.11	$\text{minRH}_{3,6\text{km\_AGL}}$
4,556.92	$\text{RH}_{m10}$	3,398.36	$\text{RH}_{m10}$
4,556.47	$\text{RH}_{m12,m18}$	3,353.86	$\text{maxRH}_{3,6\text{km\_AGL}}$
3,865.23	$\text{maxRH}_{3,6\text{km\_AGL}}$	3,062.44	$\text{RH}_{m12,m18}$

From Table 4-8,  $\text{minRH}_{3,6\text{km\_AGL}}$  had the lowest AIC score for SON, while  $\text{aveRH}_{3,6\text{km\_AGL}}$  had the lowest AIC score during DJF. On the other end of the table,  $\text{RH}_{m12,m18}$  had the highest AIC value for both SON and DJF. During SON,  $\text{minRH}_{3,6\text{km\_AGL}}$  had an AIC score of 82,909.95, which is 7,118.28 lower than that of  $\text{RH}_{m12,m18}$ . For DJF,  $\text{aveRH}_{3,6\text{km\_AGL}}$  had an AIC score of 128,754.57. This is 10,338.89 lower than the AIC score of  $\text{RH}_{m12,m18}$ . Just like in the SAS analysis, the results from the R regression model confirm that the  $\text{minRH}_{3,6\text{km\_AGL}}$  parameter performed the best during SON and the  $\text{aveRH}_{3,6\text{km\_AGL}}$  parameter performed the best during DJF.

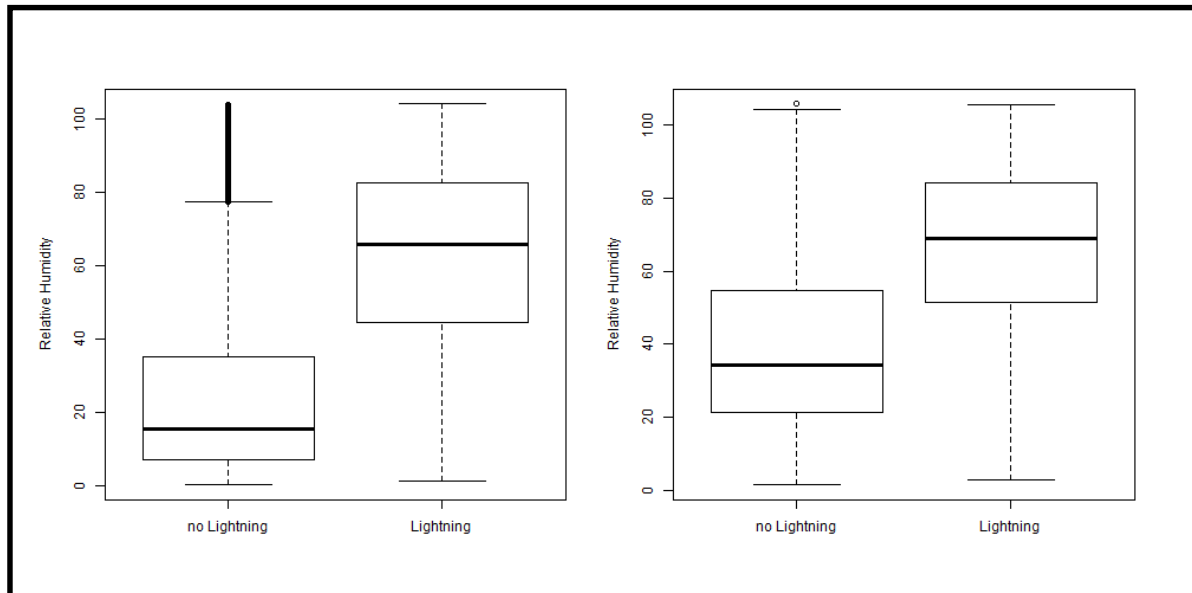
**Table 4-8: R output of stepwise logistic regression for all RH variables.**

Regression Models Selected by Score Criterion			
SON		DJF	
Score AIC	Variables Included in Model	Score AIC	Variables Included in Model
82,909.95	$\text{minRH}_{3,6\text{km\_AGL}}$	128,754.57	$\text{aveRH}_{3,6\text{km\_AGL}}$
83,385.71	$\text{aveRH}_{3,6\text{km\_AGL}}$	130,608.02	$\text{minRH}_{3,6\text{km\_AGL}}$
87,377.42	$\text{maxRH}_{3,6\text{km\_AGL}}$	131,387.89	$\text{maxRH}_{3,6\text{km\_AGL}}$
88,584.35	$\text{RH}_{m10}$	136,558.43	$\text{RH}_{m10}$
90,028.23	$\text{RH}_{m12,m18}$	139,093.46	$\text{RH}_{m12,m18}$

The  $\text{minRH}_{3,6\text{km\_AGL}}$  parameter was selected to be the most appropriate parameter for lightning prediction during SON, while  $\text{aveRH}_{3,6\text{km\_AGL}}$  was selected for DJF. RH provides information on how close to saturation the atmosphere is. Studies by Berdeklis and List (2001) have shown that there is a good correlation between RH and the charge transfer in the electrification zone of a thundercloud. Xiong et al. (2006) found that lightning corresponds well with RH in longitudinal belts where RH values are greater than 74% and negatively with RH in longitudinal belts where the RH is less than 72%. They also found a similar result for latitudinal belts. This means that higher RH values in dry regions causes more lightning activity, while high RH values in wet regions lowers the levels of lightning activity. This same study also showed that when the RH is too high (greater than 72%) in wet regions, lightning can be suppressed. When considering the typical charge structure of a thundercloud (Figure 2-1, the 3-6km AGL corresponds well with the charge separation zone of the cloud. The output from the regression analysis shows that  $\text{minRH}_{3,6\text{km\_AGL}}$  is the most appropriate RH parameter during SON while  $\text{aveRH}_{3,6\text{km\_AGL}}$  is the most appropriate for DJF to predict lightning. In early



summer, the atmospheric circulation is generally extra-tropical with a conditionally unstable atmosphere over certain parts of SA and the atmosphere is drier compared to late summer. In late summer, the circulation is tropical with a convectively unstable atmosphere and more moisture is available (Dyson et al., 2015).



**Figure 4-4: Box and whisker plots of  $\text{minRH}_{3,6\text{km\_AGL}}$  during SON (left) and  $\text{aveRH}_{3,6\text{km\_AGL}}$  during DJF (right) between 07:00 – 21:00 UTC for lightning and no Lightning. RH units are in %. The thick horizontal bar indicates the median value; the boxes denote the 25th –75th percentiles, and the whiskers show the full range of values. The circles indicate outliers.**

During SON, 75% of the  $\text{minRH}_{3,6\text{km\_AGL}}$  values were ~38% or less when no lightning occurred, while 75% of the  $\text{minRH}_{3,6\text{km\_AGL}}$  values were ~45% or more when lightning did occur. For the DJF period, 75% of the  $\text{aveRH}_{3,6\text{km\_AGL}}$  values were ~58% or less when lightning did not occur, while 75% of the  $\text{aveRH}_{3,6\text{km\_AGL}}$  values were ~56% or more when lightning did occur. This shows that the  $\text{minRH}_{3,6\text{km\_AGL}}$  and  $\text{aveRH}_{3,6\text{km\_AGL}}$  parameters can in most cases be useful to distinguish between the occurrence of lightning, and gives additional confirmation that these parameters are good choices for lightning prediction over South Africa. A threshold value of 45% or more for  $\text{minRH}_{3,6\text{km\_AGL}}$  during the SON season and 58% or more for  $\text{aveRH}_{3,6\text{km\_AGL}}$  during the DJF season can be a good indicator that lightning will occur.

#### 4.2.5 Equivalent Potential Temperature ( $\Theta_e$ )

Seven different  $\Theta_e$  parameters were added together with lightning occurrence to the stepwise logistic regression models. The  $\Theta_e$  parameters considered were the  $\Theta_e\Gamma$  between 700 and 500 hPa ( $\Theta_e\Gamma_{700,500}$ ),  $\Theta_e\Gamma$  at 600 hPa ( $\Theta_e\Gamma_{600}$ ),  $\Theta_e\Gamma$  between 850 and 400 hPa ( $\Theta_e\Gamma_{850,400}$ ),  $\Theta_e\Gamma$  between 850 and 500 hPa ( $\Theta_e\Gamma_{850,500}$ ),  $\Theta_e\Gamma$  between 1 and 6 km AGL ( $\Theta_e\Gamma_{1,6km\_AGL}$ ),  $\Theta_e\Gamma$  between the  $-10^\circ\text{C}$  and  $-20^\circ\text{C}$  level ( $\Theta_e\Gamma_{m10,m20}$ ), and the surface  $\Theta_e$  ( $\Theta_e\Gamma_{surf}$ ).

In the SAS stepwise logistic regression,  $\Theta_e\Gamma_{700,500}$  had the highest score chi-square statistic for SON, while  $\Theta_e\Gamma_{850,400}$  had the highest chi-square value for DJF (Table 4-9). For both SON and DJF,  $\Theta_e\Gamma_{m10,m20}$  had the lowest score chi-square statistic. Interestingly,  $\Theta_e\Gamma_{700,500}$ , the best performing parameter during SON, did not perform well during DJF. The best performing parameter during DJF,  $\Theta_e\Gamma_{850,400}$ , was the third best performing parameter during SON. When all seven of the  $\Theta_e$  parameters were added to the regression model, the score chi-square statistic was 5,617.41 for SON and 5,401.02 for DJF. As such, the chi-square score increased by 915.10 for SON and by 895.10 for DJF. This shows that the inclusion of all seven forms of the  $\Theta_e$  into the model does not add additional value to the lightning prediction. A comparison between the  $\Theta_e$  parameters for SON and DJF show that differences exist and that  $\Theta_e$  plays different roles in the prediction of lightning occurrence between the two seasons.

Table 4-9: SAS output of stepwise logistic regression for all  $\Theta_e$  variables.

Regression Models Selected by Score Criterion			
SON		DJF	
Chi-Square	Variables Included in Model	Chi-Square	Variables Included in Model
4,702.31	$\Theta_e\Gamma_{700,500}$	4,505.92	$\Theta_e\Gamma_{850,400}$
4,602.85	$\Theta_e\Gamma_{1,6km\_AGL}$	3,853.54	$\Theta_e\Gamma_{surf}$
4,257.70	$\Theta_e\Gamma_{850,400}$	3,727.38	$\Theta_e\Gamma_{850,500}$
3,771.14	$\Theta_e\Gamma_{850,500}$	3,420.59	$\Theta_e\Gamma_{1,6km\_AGL}$
3,713.37	$\Theta_e\Gamma_{surf}$	2,813.12	$\Theta_e\Gamma_{700,500}$
3,665.83	$\Theta_e\Gamma_{600}$	1,618.77	$\Theta_e\Gamma_{600}$
1,083.31	$\Theta_e\Gamma_{m10,m20}$	327.76	$\Theta_e\Gamma_{m10,m20}$

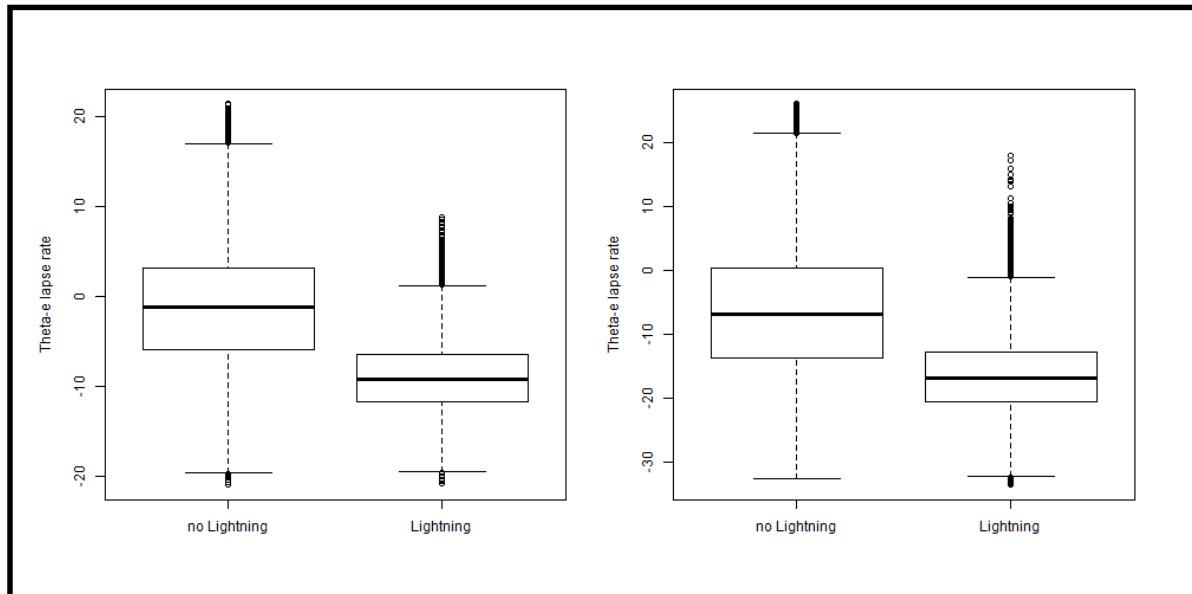
In the R regression procedure, the seven  $\Theta_e$  parameters were added individually to the model from where the AIC score was calculated.  $\Theta_e\Gamma_{700,500}$  had the lowest AIC score for SON, while  $\Theta_e\Gamma_{850,400}$  had the lowest AIC score during DJF. The  $\Theta_e\Gamma_{m10,m20}$  parameter had the highest AIC value during SON and DJF. During SON,  $\Theta_e\Gamma_{700,500}$  had an AIC score of 83,527.00 (21,356.01 lower than the  $\Theta_e\Gamma_{m10,m20}$ ) and  $\Theta_e\Gamma_{850,400}$  had an AIC score of 123,469.63 (33,228.45 lower than the  $\Theta_e\Gamma_{m10,m20}$ ) for DJF. Based on these results, the  $\Theta_e\Gamma_{700,500}$  performed the best during SON and  $\Theta_e\Gamma_{850,400}$  performed the best during DJF.

**Table 4-10: R output of stepwise logistic regression for all  $\Theta_e$  variables.**

Regression Models Selected by Score Criterion			
SON		DJF	
Score AIC	Variables Included in Model	Score AIC	Variables Included in Model
83,527.00	$\Theta_e\Gamma_{700,500}$	123,469.63	$\Theta_e\Gamma_{850,400}$
85,267.96	$\Theta_e\Gamma_{850,400}$	128,390.15	$\Theta_e\Gamma_{surf}$
87,293.15	$\Theta_e\Gamma_{850,500}$	129,355.69	$\Theta_e\Gamma_{850,500}$
87,899.68	$\Theta_e\Gamma_{1,6km\_AGL}$	136,539.83	$\Theta_e\Gamma_{1,6km\_AGL}$
90,102.71	$\Theta_e\Gamma_{600}$	138,085.18	$\Theta_e\Gamma_{700,500}$
92,081.29	$\Theta_e\Gamma_{surf}$	146,943.98	$\Theta_e\Gamma_{600}$
104,883.01	$\Theta_e\Gamma_{m10,m20}$	156,698.08	$\Theta_e\Gamma_{m10,m20}$

Based on the results above,  $\Theta_e\Gamma_{700,500}$  was selected as the most appropriate parameter to predict lightning occurrence during SON, while  $\Theta_e\Gamma_{850,400}$  was selected for DJF. The lapse rate of  $\Theta_e$  is useful to assess instabilities necessary for thunderstorms development, can be useful to assess changes in air masses, and can be used to estimate updraft velocities (Madhulatha et al., 2013; Dyson et al., 2015; Houston and Wilhelmson, 2012; Cummings, 2013; Huntrieser et al., 2007; Kuo, 1966). This means that  $\Theta_e\Gamma$  can be useful to describe the potential updrafts in thunderstorms, which supplies hydrometeors for charge separation in a thunderstorm. The differences in the parameters selected for SON and DJF can be due to the atmospheric conditions in South Africa being different between the seasons. In early summer, the atmospheric circulation is generally extra-tropical with a conditionally unstable atmosphere over SA, while in late summer the circulation is tropical with a convectively unstable atmosphere (Dyson et al., 2015).  $\Theta_e\Gamma$  measures the convective instability of a layer in the

atmosphere.  $\Theta_e \Gamma_{850,400}$  fares well in both seasons but from these results, a thinner layer of convective instability in SON (between 700 and 500 hPa) is better associated with lightning.



**Figure 4-5: Box and whisker plots of  $\Theta_e \Gamma_{700,500}$  during SON (left) and  $\Theta_e \Gamma_{850,400}$  during DJF (right) between 07:00 – 21:00 UTC for lightning and no Lightning.  $\Theta_e \Gamma$  units are in K. The thick horizontal bar indicates the median value; the boxes denote the 25th –75th percentiles, and the whiskers show the full range of values. The circles indicate outliers.**

During SON, 75% of the  $\Theta_e \Gamma_{700,500}$  values were  $\geq -5$  K when no lightning occurred, while 75% of the  $\Theta_e \Gamma_{700,500}$  values were  $\leq -5$  K when lightning did occur. For DJF, 75% of the  $\Theta_e \Gamma_{850,400}$  values were  $\geq -12$  K when lightning did not occur, while 75% of the  $\Theta_e \Gamma_{850,400}$  values were  $\leq -12$  K when lightning did occur. This shows that the  $\Theta_e \Gamma_{700,500}$  and  $\Theta_e \Gamma_{850,400}$  parameters can in most cases be useful to distinguish between the occurrence of lightning, and gives additional confirmation that these parameter is a good choice for lightning prediction over South Africa. A threshold value of  $\leq -5$  K for  $\Theta_e \Gamma_{700,500}$  during SON and  $\leq -12$  K for  $\Theta_e \Gamma_{850,400}$  during DJF can be a good indicator that lightning might occur. It is interesting to note that during DJF the atmosphere is mostly convectively unstable even on occasions when no Lightning occurs.

## 4.2.6 Temperature

The air temperature parameters considered were the temperature at 1.5 m above the ground ( $T_{1p5m}$ ), the temperature at 700 hPa ( $T_{700}$ ), the mean temperature between 850 hPa and 700 hPa ( $T_{850,700}$ ), and the mean temperature between 500 hPa and 300 hPa ( $T_{500,300}$ ).

The SAS output in Table 4-11 shows that  $T_{850,700}$  had the highest score chi-square statistic during SON and DJF.  $T_{850,700}$  had a chi-square score of 1,161.85 for SON and a chi-square score of 1,674.74 for DJF.  $T_{1p5m}$  was the temperature parameter that performed the worst during SON, while  $T_{500,300}$  had the lowest chi-square score during DJF. When all four of the temperature parameters were added to the regression model, the score chi-square statistic was 1,271.96 for SON and 2,087.32 for DJF. As such, the chi-square score increased by 110.11 for SON and by 412.58 for DJF. This shows that the inclusion of all four forms of temperature into the model does not add additional value to the lightning prediction, especially for SON.

**Table 4-11: SAS output of stepwise logistic regression for all temperature variables.**

Regression Models Selected by Score Criterion			
SON		DJF	
Chi-Square	Variables Included in Model	Chi-Square	Variables Included in Model
1,161.85	$T_{850,700}$	1,674.74	$T_{850,700}$
1,114.29	$T_{700}$	957.20	$T_{700}$
738.46	$T_{500,300}$	725.57	$T_{1p5m}$
551.75	$T_{1p5m}$	406.79	$T_{500,300}$

For the analysis with R (Table 4-12),  $T_{850,700}$  had the lowest AIC score for both SON and DJF. The  $T_{1p5m}$  parameter had the highest AIC value during SON, while  $T_{500,300}$  had the highest AIC value during DJF. During SON,  $T_{850,700}$  had an AIC score of 102,927.19 (4,638.58 lower than the  $T_{1p5m}$ ) and an AIC score of 147,184.06 (7,876.73 lower than the  $T_{500,300}$ ) for DJF. Based on these results, the  $T_{850,700}$  parameter performed the best during SON and DJF.

Table 4-12: R output of stepwise logistic regression for all temperature variables.

Regression Models Selected by Score Criterion			
SON		DJF	
Score AIC	Variables Included in Model	Score AIC	Variables Included in Model
102,927.19	$T_{850,700}$	147,184.06	$T_{850,700}$
104,828.31	$T_{700}$	153,294.85	$T_{700}$
107,561.45	$T_{500,300}$	154,742.13	$T_{1p5m}$
107,565.77	$T_{1p5m}$	155,060.79	$T_{500,300}$

Based on the results above,  $T_{850,700}$  was selected to be the most appropriate parameter to predict lightning occurrence. Surface heating from the sun is responsible for the convective processes that result in atmospheric instabilities (Bharatdwaj, 2006) and this means that temperature plays a big role in the development of thunderstorms. Price (2013) listed many studies (e.g. Williams, 1992; Williams, 1994; Williams, 2009; Reeve and Toumi, 1999; Markson and Price, 1999; Price, 1993; Price and Asfur, 2006; Markson, 2007), which have shown that lightning is closely related to the surface temperature. The regression analysis above shows that the average temperatures in the lowest 1500 m AGL (850 hPa to 700 hPa) level play an important role in lightning formation.

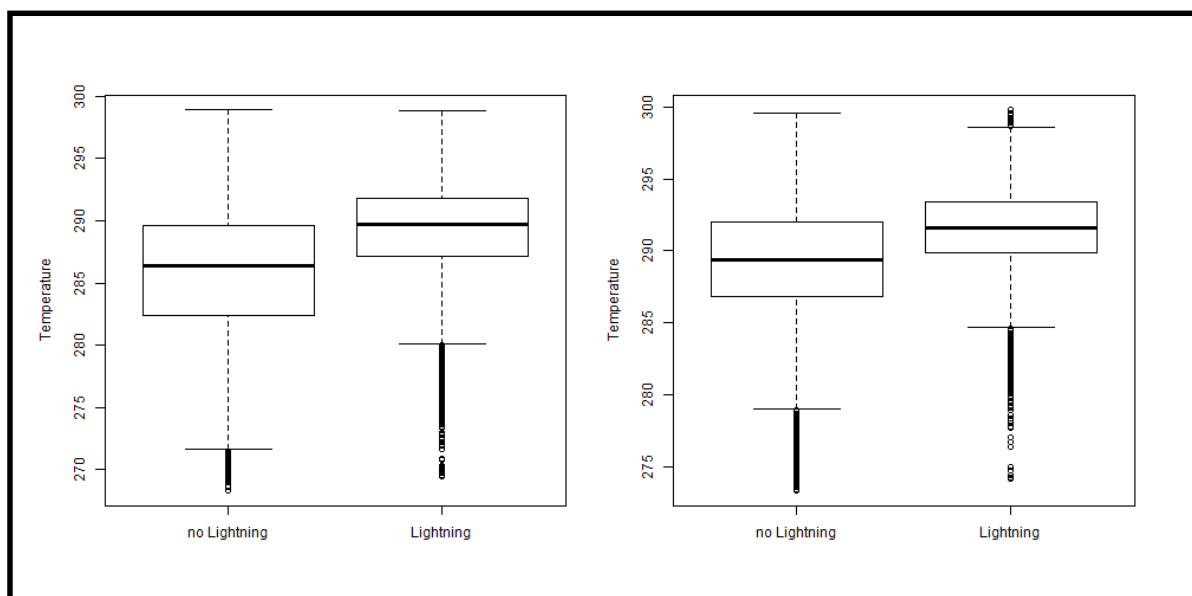


Figure 4-6: Box and whisker plots of  $T_{850,700}$  during SON (left) and DJF (right) between 07:00 – 21:00 UTC for lightning and no Lightning. Temperature units are in K. The thick horizontal bar indicates the median value; the boxes denote the 25th –75th percentiles, and the whiskers show the full range of values. The circles indicate outliers.

During SON, 75% of the  $T_{850,700}$  values were  $\leq 290$  K when no lightning occurred, while 75% of the  $T_{850,700}$  values were  $\geq 287$  K when lightning did occur (Figure 4-6). For DJF, 75% of the  $T_{850,700}$  values were  $\leq 292$  K when no lightning occurred, while 75% of the  $T_{850,700}$  values were  $\geq 290$  K when lightning did occur. This shows that the  $T_{850,700}$  parameter can in most cases be useful to distinguish between the occurrence of lightning, and gives additional confirmation that this parameter is a good choice for lightning prediction over South Africa. There was a slight overlap between the temperature values when lightning did and did not occur. For SON, a temperature  $\geq 290$  K might be a good threshold value for lightning to occur and a value of  $\geq 293$  K for DJF.

#### 4.2.7 Summary of selected parameters

The same variant of CAPE, LI, PW and temperature were selected for SON and DJF while different RH and  $\Theta_e\Gamma$  parameters were identified (Table 4-13). All of the parameters, except for LI, differ from the ones used by Frisbie et al. (2009) in their LPI. This highlights the importance of selecting the most appropriate parameters for different climatic regions and for the utilisation of a different NWP model when developing a lightning prediction model.

**Table 4-13: Summary of most appropriate parameters selected for the SON and DJF periods.**

<u>SON</u>	<u>DJF</u>
muCAPE <sub>1,6km_AGL</sub>	muCAPE <sub>1,6km_AGL</sub>
SLI	SLI
PW <sub>850,300</sub>	PW <sub>850,300</sub>
minRH <sub>3,6km_AGL</sub>	aveRH <sub>3,6km_AGL</sub>
$\Theta_e\Gamma_{700,500}$	$\Theta_e\Gamma_{850,400}$
$T_{850,700}$	$T_{850,700}$

### 4.3 DEVELOPMENT OF THE NEW LTI

The new LTI was developed by making use of rare event logistic regression techniques. In this approach, the most appropriate parameters selected in Section 4.2 and listed in Table 4-13 were used as the independent variables in the regression analysis, while the binary outcome of lightning occurrence was the dependent variable. The rare event logistic regression procedure, discussed in Section 3.5.2, was followed. The aim was to develop an equation to predict lightning occurrence for SON and DJF separately.

#### 4.3.1 SON period

The dataset for SON consisted of 137,864 observations, from which 3,945 were events and 133,919 were non-events. This essentially means that 137,864 grid boxes were considered during the SON period, where 3,945 grid boxes contained at least one lightning stroke (events) and 133,919 grid boxes did not contain any lightning strokes (non-events). Events were given a value of 1, while the non-events were given a value of 0. The six NWP model parameters listed in Table 4-13 had a value corresponding to the 137,864 grid boxes. The SON dataset used in the regression analysis looked similar to the example dataset shown in Table 3-2 and consisted of 137,864 lines.

The rare event logistic regression procedure discussed in Section 3.5.2 was followed. All of the 3,945 events were taken together with a random sample of 3,945 non-events to form a new dataset. This new dataset was used to perform the rare event logistic regression that provided the output for the first regression model. From here the random sample of non-events were added back to the original dataset and had an equal probability of being selected again. This entire process was repeated 1,000 times, where a new random sample of non-events was selected for every model. In the end, there were 1,000 models, each with its own regression analysis output. The 1,000 models were combined by averaging the intercept term, regression coefficients and statistics output.

Table 4-14 shows the output of the rare-event logistic regression when the 1,000 models were combined. The intercept term does not have a MPV and MPI value. The MPV is simply the largest value of a parameter (or smallest for parameters where a negative value is of importance) in the dataset. The MPI is the MPV multiplied by Coef and is a measure to determine the most important variables in the analysis (Guns and Vanacker, 2012; Vanwalleghem et al., 2008). Most of the values in Table 4-14 were rounded for presentation purposes



**Table 4-14: Output from the rare event logistic regression for SON, providing the regression coefficients (Coef), p-value (P), standard error on Coef (SE), Odds ratio (Odds), z-value (Z), maximum parameter value (MPV) and measure of parameter importance (MPI) for the intercept term and the six model parameters.**

	Coef	P	SE	Odds	Z	MPV	MPI
<b>Intercept</b>	81.256445	<0.05	1.8435	4.96E+35	44.08		
<b>muCAPE<sub>1,6km_AGL</sub></b>	-0.000034	0.35	3.15E-5	1.000	-1.09	7545.15	-0.26
<b>PW<sub>850,300</sub></b>	1.871132	<0.05	0.0549	6.5031	34.10	3.90	7.30
<b>SLI</b>	-0.307571	<0.05	0.0092	0.7352	-33.36	-17.05	5.24
<b><math>\Theta_e \Gamma_{700,500}</math></b>	-0.200460	<0.05	0.0059	0.8184	-34.11	-20.83	4.18
<b>minRH<sub>3,6km_AGL</sub></b>	0.021858	< 0.05	0.0008	1.0221	26.18	100.00	2.19
<b>T<sub>850,700</sub></b>	-0.307643	< 0.05	0.0066	0.7352	-46.36	298.81	-91.93

The intercept term as well as five out of the six model parameters are significant at the 5% significance level (P less than 0.05 and Z not being between -1.96 and +1.96) (Table 4.14).  $\mu\text{CAPE}_{1-6\text{km}}$  is the only exception. The MPI values also confirm that  $\mu\text{CAPE}_{1-6\text{km}}$  is not as important when predicting lightning occurrence during SON since it has a very low MPI value compared to the other parameters. Since the  $\mu\text{CAPE}_{1-6\text{km}}$  parameter was not significant during SON, it was discarded from the list of candidate predictors in Table 4-13. The rare-event logistic regression procedure for SON, discussed above, was again repeated, but this time with  $\mu\text{CAPE}_{1-6\text{km}}$  removed from the analysis. Only the remaining five parameters were considered in the new regression analysis. Table 4-15 shows the output of the new rare-event regression for SON.

**Table 4-15: Results from the new rare event logistic regression for the SON period with  $\mu\text{CAPE}_{1-6\text{km}}$  removed from the analysis.**

	Coef	P	SE	Odds	Z	MPV	MPI
<b>Intercept</b>	81.597771	<0.05	1.8112	6.612E+35	45.05		
<b>PW<sub>850,300</sub></b>	1.858302	<0.05	0.0536	6.4196	34.70	3.90	7.25
<b>SLI</b>	-0.301027	<0.05	0.0070	0.7401	-42.72	-17.04	5.13
<b><math>\Theta_e \Gamma_{700,500}</math></b>	-0.199863	<0.05	0.0059	0.8189	-34.16	-20.83	4.16
<b>minRH<sub>3,6km_AGL</sub></b>	0.021953	<0.05	0.0008	1.0222	26.41	104.10	2.29
<b>T<sub>850,700</sub></b>	-0.308811	<0.05	0.0065	0.7343	-47.29	298.81	-92.28

The intercept term and all five of the model parameters are significant at the 5% significance level (P less than 0.05 and Z not being between -1.96 and +1.96). All of the parameters have a very low SE value, which means that the model fits the data well.  $T_{850-700}$  plays the biggest role in the regression model as can be seen from the large MPI value.  $PW_{850-300}$  is the second most important parameter, followed by the SLI,  $\Theta_e\Gamma_{700-500}$  and  $minRH_{3-6km}$ .

The intercept term and the regression coefficients listed in Table 4-15 could then be added to Equation 3-3 to produce the new LTI. The LTI for SON is given by Equation 4-1 that provides the probability of lightning occurrence. The output will give a value between 0 and 1, which can be multiplied by 100 to provide a probability between 0% and 100%.

$$LTI = \frac{1}{1 + e^{-[\hat{\alpha} + \hat{\beta}_1(PW_{850,300}) + \hat{\beta}_2(SLI) + \hat{\beta}_3(\Theta_e\Gamma_{700,500}) + \hat{\beta}_4(minRH_{3,6km\_AGL}) + \hat{\beta}_5(T_{850,700})]}}$$

(4-1)

Where:

$$\hat{\alpha} = 81.597771 \quad \hat{\beta}_1 = 1.858302 \quad \hat{\beta}_2 = -0.301027$$

$$\hat{\beta}_3 = -0.199863 \quad \hat{\beta}_4 = 0.021953 \quad \hat{\beta}_5 = -0.308811$$

### 4.3.2 DJF period

The dataset for the DJF period consisted of 128,562 observations. From the 128,562 grid boxes considered during DJF 6,118 grid boxes contained at least one lightning stroke (events) and 122,444 grid boxes contained no lightning strokes (non-events). The same techniques used for SON was applied to DJF and the output of the rare-event logistic regression when the 1,000 models were combined are shown in Table 4-16

The intercept term and all six of the model parameters are significant at the 5% significance level (P less than 0.05). The Z confirms that all of the parameters are significant since Z does not fall between -1.96 and +1.96. All of the parameters have a very low SE value, which means that the model fits the data well.  $T_{850-700}$  is by far the most important parameter in the regression model as can be seen from the large MPI value.  $\Theta_e\Gamma_{850-400}$  is the second most important parameter in the model, followed by the  $PW_{850-300}$ , SLI,  $\mu CAPE_{1-6km}$  and  $aveRH_{3-6km}$ .

Table 4-16: Output from the rare event logistic regression for DJF.

	Coef	P	SE	Odds	Z	MPV	MPI
<b>Intercept</b>	96.37400	<0.05	1.8712	1.985E+42	51.50		
<b>muCAPE<sub>1,6km_AGL</sub></b>	-0.00036	<0.05	2.378E-5	0.9996	-15.23	6757.06	-2.45
<b>PW<sub>850-300</sub></b>	1.64985	<0.05	0.0349	5.2081	47.25	4.80	7.92
<b>SLI</b>	-0.27441	<0.05	0.0089	0.7600	-30.94	-13.63	3.74
<b>Θ<sub>eΓ</sub><sub>850,400</sub></b>	-0.25384	<0.05	0.0039	0.7758	-64.75	-33.60	8.53
<b>aveRH<sub>3,6km_AGL</sub></b>	0.02284	<0.05	0.0007	1.0231	30.69	105.60	2.41
<b>T<sub>850-700</sub></b>	-0.36468	<0.05	0.0066	0.6944	-55.19	299.80	-109.33

The intercept term and the regression coefficients listed in Table 4-16 could then be added to Equation 3-3 to produce the new lightning forecast. The LTI for DJF is given by Equation 4-2 that provides the probability of lightning occurrence. The output will give a value between 0 and 1, which can be multiplied by 100 to provide a probability between 0% and 100%.

$$LTI = \frac{1}{1 + e^{-[\hat{\alpha} + \hat{\beta}_1(\text{muCAPE}_{1,6\text{km\_AGL}}) + \hat{\beta}_2(\text{PW}_{850,300}) + \hat{\beta}_3(\text{SLI}) + \hat{\beta}_4(\Theta_{e\Gamma_{850,400}}) + \hat{\beta}_5(\text{aveRH}_{3,6\text{km\_AGL}}) + \hat{\beta}_6(\text{T}_{850,700})]}}$$

(4-2)

Where:

$$\hat{\alpha} = 96.373973 \quad \hat{\beta}_1 = -0.000362 \quad \hat{\beta}_2 = 1.649851$$

$$\hat{\beta}_3 = -0.274410 \quad \hat{\beta}_4 = -0.253844 \quad \hat{\beta}_5 = 0.022844$$

$$\hat{\beta}_6 = -0.364683$$

#### 4.4 SUMMARY

Chapter 3 deals with the development of the new LTI. Stepwise logistic regression techniques were utilised to select the most appropriate NWP model parameters to predict the occurrence of lightning over South Africa. During SON, the most unstable CAPE in the 1-6 km above ground level, surface Lifted Index, mean Precipitable Water between the 850 and 300 hPa levels, minimum Relative Humidity in the 3-6 km above ground level, Equivalent Potential Temperature lapse rate between 700 and 500 hPa, and mean Temperature between 850 and 700 hPa proved to be the most appropriate parameters. For the DJF season, the same parameters were selected except that the minimum Relative Humidity in the 3-6 km above ground level was replaced with the average Relative Humidity in the 3-6 km above ground level, and the Equivalent Potential Temperature lapse rate between 700 and 500 hPa was replaced with the Equivalent Potential Temperature lapse rate between 850 and 400 hPa.

The six parameters selected for each season were utilised to develop the new LTI by means of rare event logistic regression techniques. Two equations were produced, one for SON and one for DJF. The  $\mu\text{CAPE}_{1,6\text{km\_AGL}}$  parameter was omitted from the LTI equation during SON, since it did not add value to the forecast. As such, five parameters were utilised. For DJF, all six of the selected parameters were used. Both equations make use of the selected NWP model parameters to predict the probability of lightning occurrence. This new LTI will utilise UM model output early in the morning to provide a probability forecast of where lightning is expected to occur between 07:00 and 21:00 UTC.

---

# CHAPTER 5

## VERIFICATION OF THE LIGHTNING THREAT INDEX

---

### 5.1 INTRODUCTION

In the previous chapter the statistical techniques used to develop of the new LTI were described. Two LTI's are proposed for the SON and DJF seasons separately. This chapter will evaluate the performance of the newly developed LTI over independent SON and DJF seasons to determine how well the index performs when predicting the occurrence of lightning. The new LTI was developed over the 2011 and 2012 SON and 2011/12 and 2012/13 DJF seasons, while the evaluation was performed on the 2013 SON and 2013/14 DJF seasons.

In section 5.2 and 5.3, a quantitative evaluation of the new LTI against the occurrence of lightning will be performed over SON and DJF. Section 5.2 will deal with the probabilistic verification of the LTI, while dichotomous verification statistics will be shown in section 5.3. The chapter ends with a visual comparison, by means of case studies, of the original LPI, newly developed LTI and UM convective rainfall forecast against the occurrence of observed lightning. The case studies will demonstrate the output of the product that will be used in an operational environment.

### 5.2 PROBABILISTIC VERIFICATION

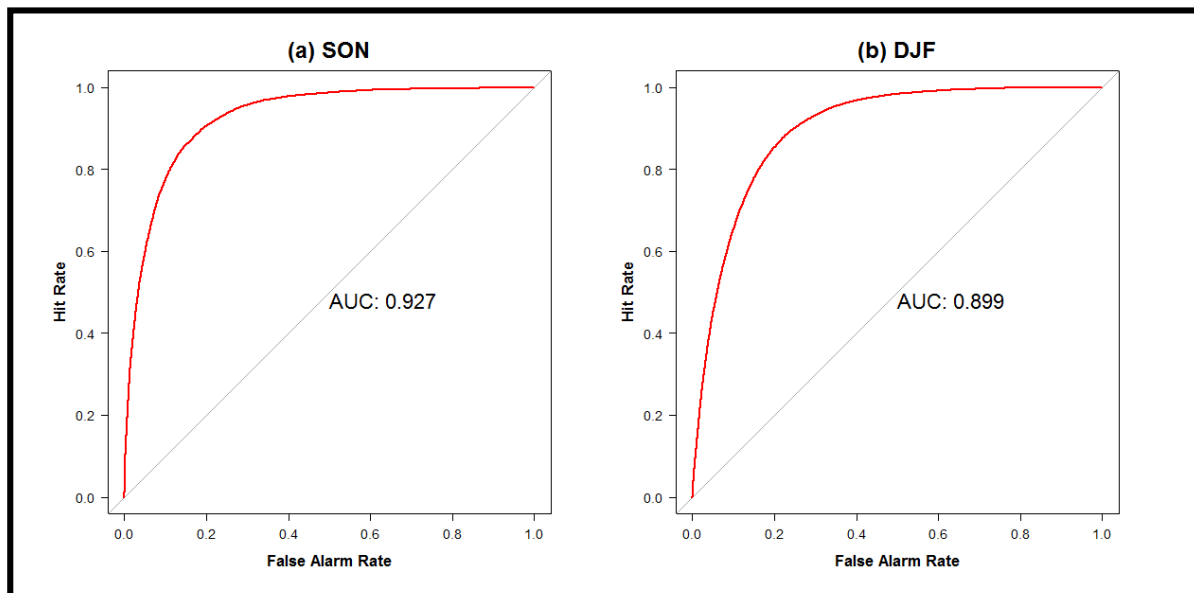
This section deals with the probabilistic evaluation of the new LTI against observed lightning. The LTI is a probabilistic forecast, and as such, this section provides the main evaluation results of this chapter. From the probabilistic evaluation scores the sensitivity, reliability and sharpness of the LTI will be examined. For this a ROC analysis, reliability diagram, and sharpness diagram will be considered.

## 5.2.1 ROC curve

The ROC analysis performed on the LTI (Figure 5-1) shows how well the LTI can distinguish between the occurrences and non-occurrences of lightning over all possible probability values.

During both the SON and DJF seasons, the ROC curves in Figure 5-1 approaches the top left corner of the diagram and falls way above the no skill diagonal line. The curves show that the LTI has a high sensitivity (or high hit rate) and high specificity (or low false alarm rate) across all of the possible probability ranges. The high sensitivity indicates that the LTI correctly predicts the lightning events, while the high specificity shows that the LTI correctly predicts the lightning non-events. One can conclude from the ROC curves in Figure 5-1 that the LTI discriminates well between lightning occurrences and non-occurrences and that the LTI forecasts are accurate.

The ROC curves in Figure 5-1 are also accompanied by an AUC value that gives a representation of the overall performance of the LTI. During SON, the AUC value was 0.927, and for DJF it was 0.899. Since an AUC value of 1.0 represents a perfect forecast, and a value of 0.5 or less represent a worthless forecast (Fawcett, 2006), the LTI performed really well, since the AUC values are close to 1.0. The AUC was slightly higher for SON than for DJF.



**Figure 5-1: ROC curves and AUC values for the UM LTI forecasts during (a) SON of 2013 and (b) DJF of 2013/14.**

## 5.2.2 Reliability

The reliability plots of the probabilistic LTI forecasts are depicted in Figure 5-2. These diagrams show how well the forecasted probabilities of the LTI correspond with the observed frequency of lightning occurrence (Weisheimer and Palmer, 2014). The probabilities in Figure 5-2 are divided into 10 probability bins.

During SON (Figure 5-2a), the LTI over-forecasts the observed frequency of lightning occurrence. This is evident from the curve falling quite a bit under the diagonal line. For the first bin, the forecast is reliable, but becomes increasingly more unreliable towards the 8<sup>th</sup> bin, and then starts moving back to the diagonal line. The LTI forecast during SON generally over-forecasts the lightning.

The reliability diagram for DJF (Figure 5-2b) looks much better than that of the SON. The LTI only slightly over-forecasts the observed frequency of lightning occurrence, which is evident from the curve falling just under the diagonal line. The forecast starts out reliable in the first bin, then slightly moves away from the diagonal line up to about the 6<sup>th</sup> bin, from where it gradually moves back to the diagonal line. This shows that the LTI forecast is reliable during the DJF season and much more reliable than the LTI forecasts for the SON period.

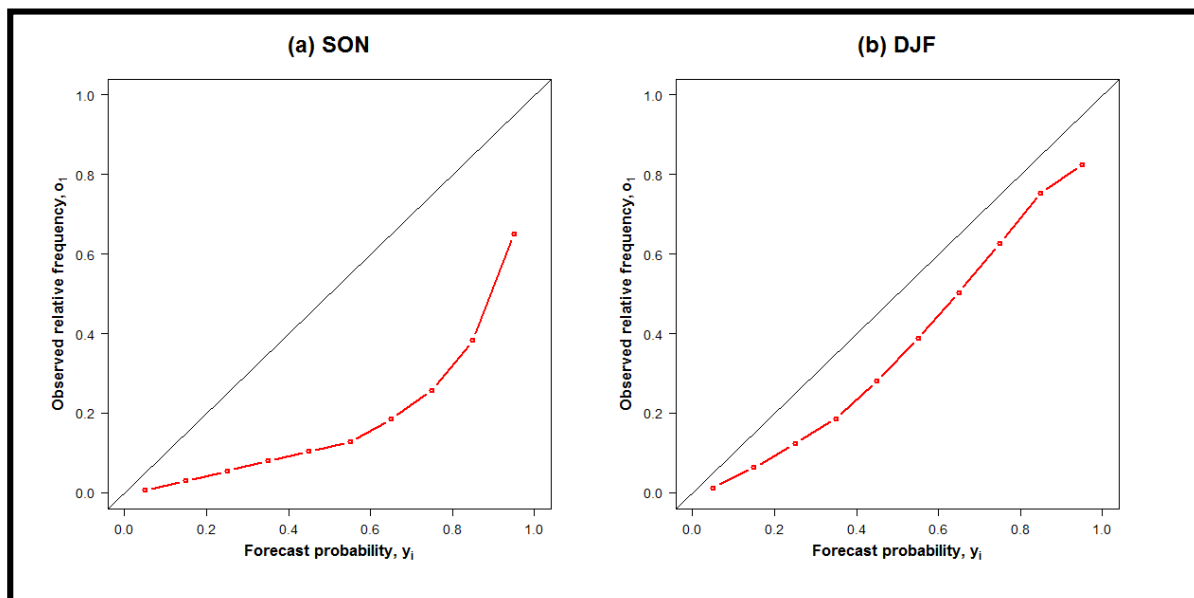
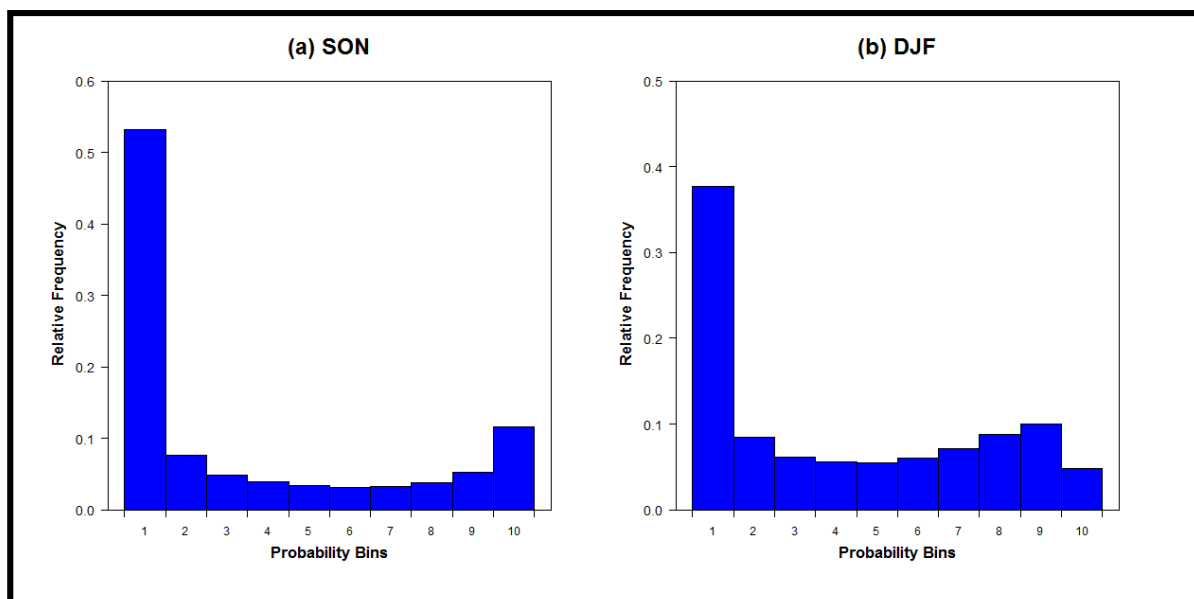


Figure 5-2: Reliability diagrams for the UM LTI forecasts during (a) SON of 2013 and (b) DJF of 2013/14

### 5.2.3 Sharpness

The sharpness diagrams in Figure 5-3 measures, as the name suggests, the sharpness of a forecast, which is a measure of how the relative frequency of lightning occurrence varies across all the probabilities of the LTI forecast (Callado et al., 2013). The probabilities of the sharpness diagrams in Figure 5-3 are divided into 10 bins.

The LTI forecast has good sharpness during SON and DJF since the sharpness diagrams have a U-shape distribution. The most number of forecasts are made in the first probability bin for both the SON and DJF seasons, from where it decreases to the sixth probability bin for SON and fifth bin for DJF. From here, it then starts to increase again. During DJF, the number of forecasts in the ninth probability bin is higher compared to the tenth probability bin.



**Figure 5-3: Sharpness diagrams for the UM LTI forecasts during (a) SON of 2013 and (b) DJF of 2013/14**

### 5.3 DICHOTOMOUS VERIFICATION

This section provides the results from the contingency table evaluation as discussed in Section 3.6.2. Since the LTI forecast provides the probability of lightning occurrence as output, this dichotomous evaluation had to be performed over the entire range of probabilities. Probability thresholds in intervals of 10% were considered to calculate the evaluation statistics for areas of the domain where the LTI probabilities exceeded certain threshold values. At first,



the statistics are calculated for the area of the LTI forecast where the probabilities exceed 0% (the entire forecast domain). Then the scores are calculated for the areas where the LTI forecast probabilities are greater than 10%, and so forth. This continues in intervals of 10% until the last threshold is reached where the probabilities of the LTI are greater than 90%. In the first interval, the POD is high, but so is the POFD and FAR. From here, the POD decreases, but so does the POFD and FAR. This approach evaluates the performance of the LTI over all of the probabilities of the forecast and is useful to determine at what probability threshold of the LTI one is likely to see lightning.

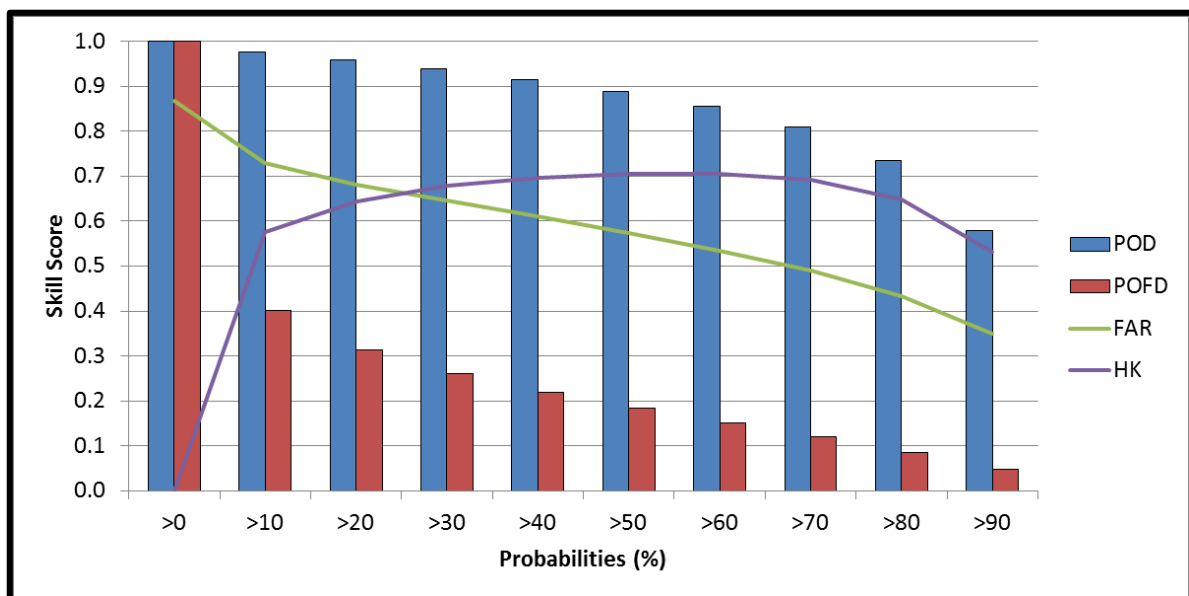
**Table 5-1: Contingency table for all the probability thresholds of the UM LTI during SON of 2013.**

Probabilities (%)	Hits	Misses	False Alarms	Correct Negatives
>0	9,099	0	59,574	0
>10	8,887	212	23,889	35,685
>20	8,727	372	18,721	40,853
>30	8,542	557	15,531	44,043
>40	8,326	773	13,038	46,536
>50	8,078	1,021	10,901	48,673
>60	7,794	1,305	8,963	50,611
>70	7,373	1,726	7,110	52,464
>80	6,682	2,417	5,116	54,458
>90	5,266	3,833	2,839	56,735

The statistical scores shown in Figure 5-4 were calculated from the contingency table shown in Table 5-1. For probabilities > 0%, the POD is 1.0, but the POFD (1.0) and FAR (0.87) is also high. This is due to the area of probabilities > 0% consisting of the entire study domain, so all lightning activity will be captured (no misses and correct negatives), but there will be a large number of false alarms where no lightning occurred. The HK score is also 0 for this threshold. For the next threshold (probabilities greater than 10%), the POD remains high (0.98), the FAR decreases to 0.73, the POFD drops to 0.4, and the HK score increases to 0.58. From here, as the probability threshold increases, the POD gradually decreases, reaching a minimum of 0.58 when the probability threshold is 90%. The same applies for the POFD and FAR, reaching its lowest value of 0.05 and 0.35 respectively when the probability

threshold is 90%. The HK score increases up to the 60% probability threshold, reaching a maximum value of 0.706, and then starts decreasing again.

The POD values are high for all probability thresholds while the POFD values are low. This shows that the LTI performs well since it is able to identify the lightning events while at the same time also identify the non-events. The FAR is high, which indicates that the LTI does somewhat over-forecast lightning occurrence during SON. This confirms the result obtained from the reliability plot in the previous section. Since the HK score measures the overall performance of how well the LTI forecast can distinguish between the lightning events and non-events, the maximum HK value of 0.706 when the probabilities exceed 60%, is a good threshold value to predict areas where lightning is likely to occur.



**Figure 5-4: Statistical scores for the UM LTI forecasts during SON of 2013.**

Another measure of how the LTI is over or under-forecasting lightning activity during the SON season is the frequency bias shown in Figure 5-5. A perfect bias is 1.0. For the first interval, the bias is very large (7.55). From here, the bias decreases as the probability threshold increases. By the time the threshold reaches the 60% probability interval, where the HK was the highest, the bias is 1.84, indicating that the LTI does still over-forecast the lightning occurrence. The LTI bias is 0.89 when the probability threshold is 90%, indicating slight under-forecasting. It can be concluded from Figure 5-5 that the LTI over-forecasts for all probabilities, except for the 90% threshold, but the bias does improve with higher probabilities.

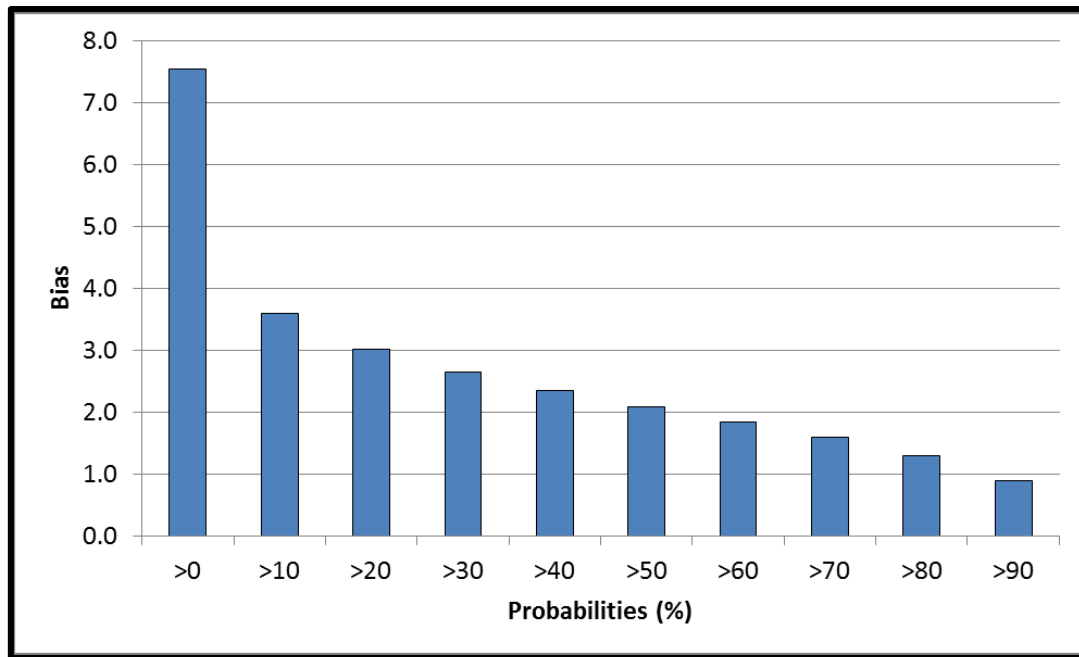


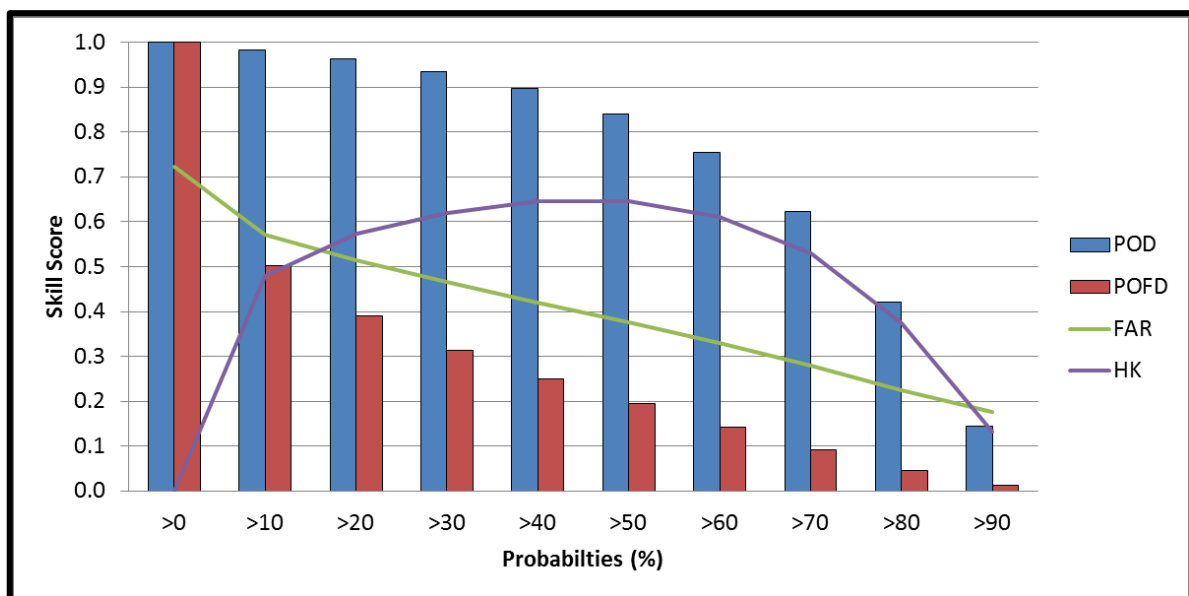
Figure 5-5: Frequency bias for the UM LTI forecasts during SON of 2013.

Table 5-2: Contingency table for all the probability thresholds of the UM LTI during DJF of 2013/14.

Probabilities (%)	Hits	Misses	False Alarms	Correct Negatives
>0	18,393	0	47,963	0
>10	18,077	316	24,120	23,843
>20	17,712	681	18,728	29,235
>30	17,196	1,197	15,071	32,892
>40	16,488	1,905	11,986	35,977
>50	15,457	2,936	9,336	38,627
>60	13,877	4,516	6,842	41,121
>70	11,455	6,938	4,452	43,511
>80	7,734	10,659	2234	45,729
>90	2,656	15,737	570	47,393

The statistical scores for DJF shown in Figure 5-6 was calculated from the contingency table shown in Table 5-2. Once again, in the first threshold, when the probabilities are > 0%, the POD is 1.0, but the POFD (1.0) and FAR (0.72) is high. This is due to the area of

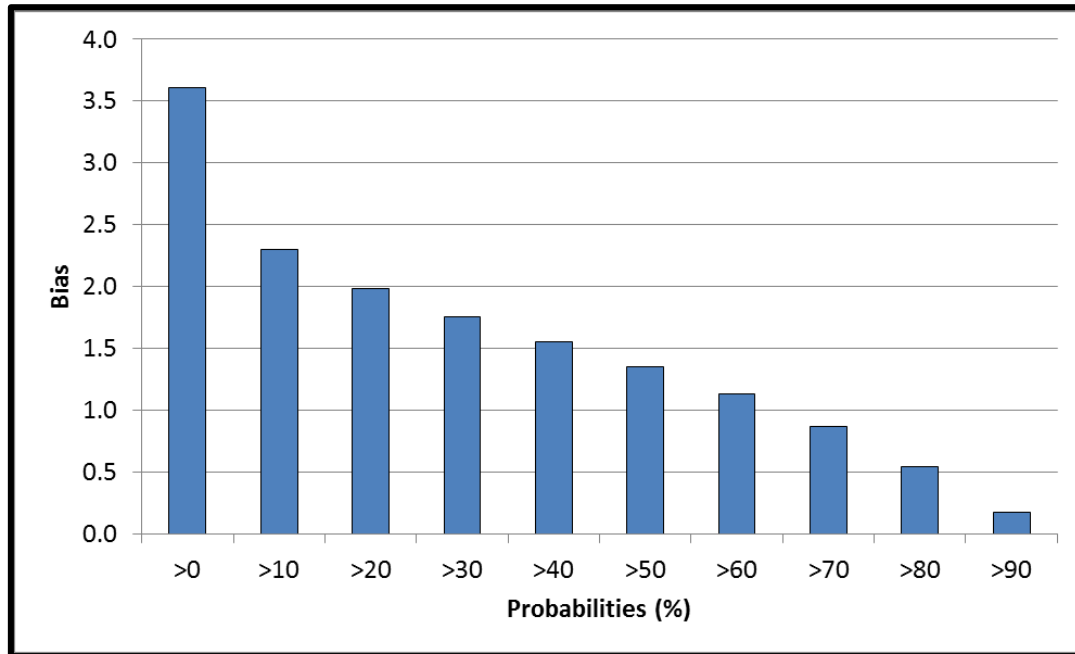
probabilities > 0% consisting of the entire study domain, so all lightning activity will be captured (no misses and correct negatives), but there will be a large amount of false alarms where no lightning occurred. The HK score is also 0 for this threshold. For the next threshold (probabilities greater than 10%), the POD remains high (0.98), the FAR decreases to 0.57, the POFD drops to 0.5, and the HK score increases to 0.48. From here, as the probability threshold increases, the POD gradually decreases, reaching a minimum of 0.13 when the probability threshold is 90%. The same applies for the POFD and FAR, reaching its lowest value of 0.012 and 0.177 respectively when the probability threshold is 90%. The HK score increases up to the 40% probability threshold, reaching a maximum value of 0.64, and then starts decreasing again.



**Figure 5-6: Statistical scores for the UM LTI forecasts during DJF of 2013/14.**

The POD values are high for all probability thresholds, up to 70%, from where it decreases to below 0.6. The POFD values are low, especially for larger probability thresholds. This shows that the LTI performs fairly well since it is able to correctly identify the lightning events while at the same time also identify the non-events. The FAR is high, which indicates that the LTI does somewhat over-forecast lightning occurrence during the DJF season, but is lower compared to the FAR values of the SON period. This confirms the result obtained from the reliability plot in the previous section where the LTI slightly over-forecasts the occurrence of lightning. Since the HK score measures the overall performance of how well the LTI forecast can distinguish between the lightning events and non-events, the maximum HK value of 0.64

when the probabilities exceed 40%, is a good threshold value for DJF to look for areas where lightning is likely to occur.



**Figure 5-7: Frequency bias for the LTI forecasts during the DJF period.**

In the first probability interval of Figure 5-7, the bias is 3.6. This shows considerable over-forecasting but can be expected due to the area of probabilities > 0% consisting of the entire study domain. From here, the bias decreases, showing that probabilities up to 70% over-forecasts, while probabilities exceeding 70% under-forecasts the occurrence of lightning. By the time the threshold reaches 40%, where the HK is the highest, the bias is 1.55, indicating that the LTI does still slightly over-forecast the lightning occurrence.

#### 5.4 EYEBALL VERIFICATION

Section 5.2 and 5.3 dealt with the quantitative evaluation of the newly developed LTI against the occurrence of lightning in the 2013 SON and 2013/14 DJF seasons. In order to evaluate how the newly developed LTI forecast compares with the original Frisbie et al., (2009) LPI and the convective rainfall forecast from the UM, a qualitative evaluation was performed against lightning occurrence. The new LTI was compared qualitatively with the old LPI, which uses the Frisbie et al., (2009) methodology, in order to determine if a superior product was

developed. Similarly, the UM model produces a convective rainfall product that forecast the amount of precipitation from convective storms on the ground, which can be used as a proxy for the occurrence of lightning. As a result the new LTI was compared to the convective rainfall forecast of the UM to determine whether the LTI improves upon the convective rainfall product. Since the output of the original LPI was given by four risk categories ranging from low to extreme risk, the output from the new LTI is in probabilities, and the convective rainfall is measured in centimetres, a quantitative evaluation between the different forecasts would have been difficult to perform. As such, a visual comparison between the three different forecasts and the occurrence of lightning are shown. This also provides examples of how the final output of the new LTI will look in an operational environment.

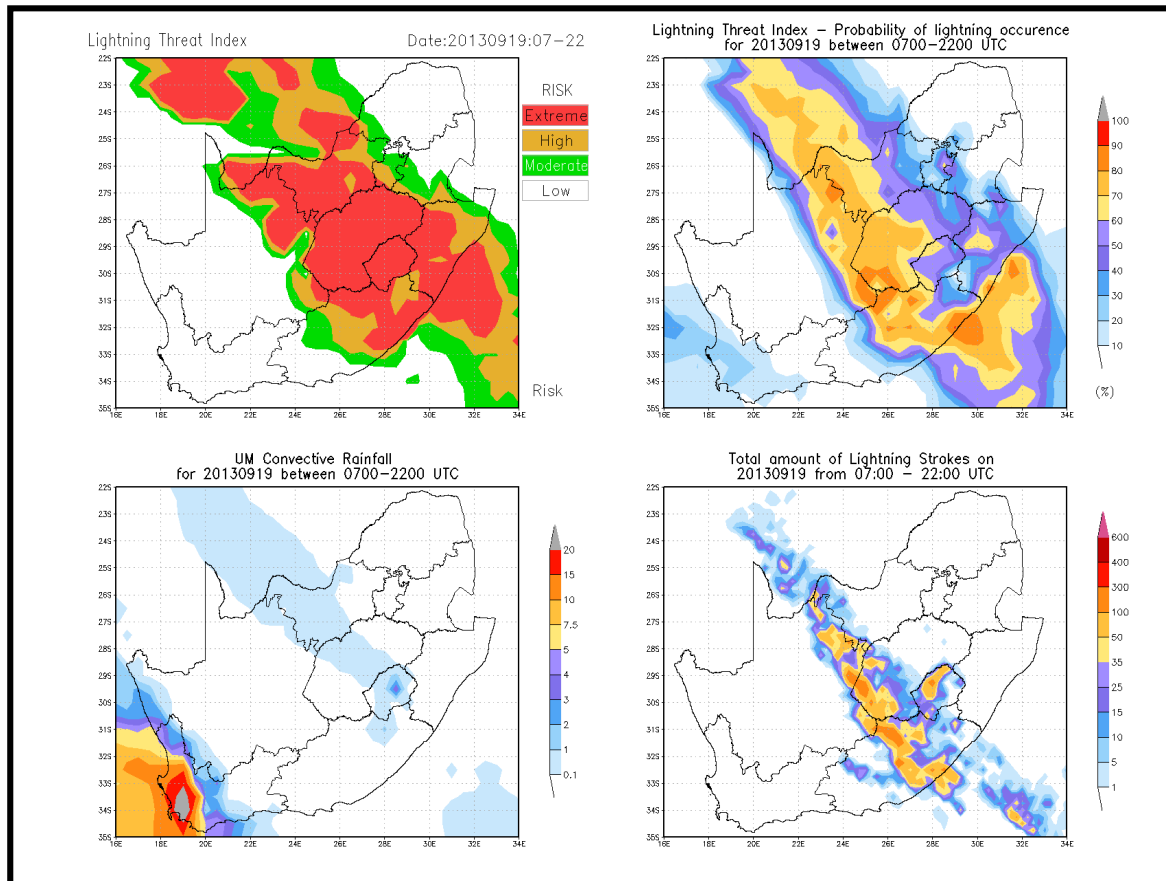
The original LPI, new LTI and convective rainfall forecasts were for the period between 07:00 and 22:00 UTC and were compared with the sum of observed lightning strokes during this period. The six case studies considered were selected randomly as days where the size of the daily lightning data file indicated that a fair amount of lightning occurred. One day from each month of the evaluation period was selected.

#### **5.4.1 Case 1: 19 January 2013**

On 19 September 2013, a narrow band of lightning activity occurred over South Africa (Figure 5-8). The band of lightning activity extends from Botswana in the north-western parts of the domain, towards the Eastern Cape and Indian Ocean in the south-eastern parts of the domain.

The UM convective rainfall forecast did identify a band of convective rainfall in the general area where the lightning occurred. Areas over Botswana were detected quite well but not the areas over the north-eastern parts of the Northern Cape and south-western parts of the Free State. Over these regions, the forecast was displaced from where the lightning activity occurred. Only light convective rain (0.1-10 cm) were predicted. The lightning activity over the Eastern Cape and off the coast was mostly missed by the UM convective rainfall forecast, but it did forecast the area over the ocean in the south-eastern most parts of the domain. Over the Western Cape and along the west coast of the country large amounts of convective rainfall were predicted in association with a cold frontal system. No lightning was observed in this area but convective precipitation, which occurs in association with cold fronts over these areas, are from shallow convective cloud (cold air cumulus) which seldom produce any lightning (Taljaard, 1995).

The LPI did forecast lightning for most of the areas where lightning occurred but also identified extensive areas where no lightning occurred. Over the central to northern parts of the Free State, almost no lightning occurred, but the LPI predicted that there is a high risk of lightning activity. Similarly, over most of the North West Province and KwaZulu-Natal, the LPI forecasted a high and extreme risk of lightning, yet no lightning occurred.



**Figure 5-8: LPI (top left), LTI (top right), UM convective rainfall in cm (bottom left), and occurrence of lightning (bottom right) for 19 September 2013 between 07:00 and 22:00 UTC.**

A visual comparison between the LTI and the lightning occurrence show a good agreement. In the previous section it was stated that probabilities exceeding 60% is a good threshold to use to predict lightning. When the areas with probabilities exceeding 60% are considered for the LTI (orange shades on Figure 5-8 – top left), there is good agreement with the observed lightning. However, even then the LTI predicted lightning to far east over the Northwest and Free State provinces and no lightning was observed over Kwa-Zulu Natal. Almost all of the noteworthy areas where lightning occurred can be seen in the areas where the probabilities are 70% or more. Not a lot of lightning activity occurred in the areas where

the probabilities were lower than 60%, but in the few isolated areas where it did, the LTI was able to identify it. Also, as with the UM convective rainfall, the LTI forecasted an area of low probabilities over the Western Cape and along the west coast of the country where no lightning occurred. From all the forecasts, the new LTI seemed the most able to capture the occurrence of lightning on this day.

#### **5.4.2 Case 2: 10 October 2013**

The 10 October 2013 case, shown in Figure 5-9, is an example of where none of the forecast products performed well. This example illustrates that the products derived from the UM model can only be as good as the UM prognosis for a specific day. Nevertheless, the value of the LTI is illustrated in this case study, as it was capable of identifying areas of lightning risk where the convective rainfall predicted no rainfall. On this day lightning occurred over the north-eastern parts of the country, extending into Botswana.

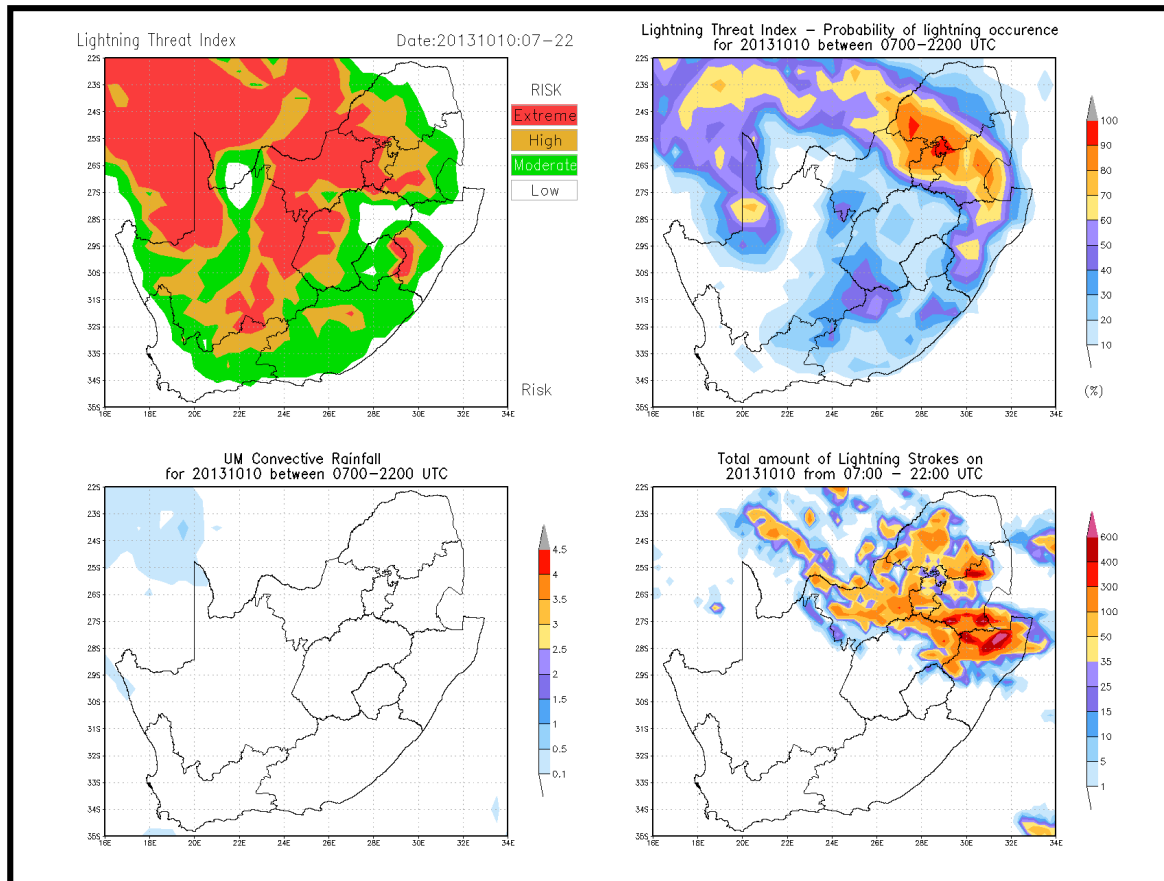
The UM convective rainfall forecast did not perform well at all on this day. Rainfall of less than 1.0 cm was predicted and none of this over South Africa. No rainfall was predicted over the extensive area where lightning was observed.

Most of South Africa, Namibia and Botswana were under moderate to extreme risk of lightning according to the LPI. The high and extreme risk areas over the northern parts of the North West Province, Gauteng, south-western parts of Limpopo, eastern parts of Mpumalanga, and Botswana did at least predict the lightning that occurred over these areas. No lightning occurred over the Western Cape, Eastern Cape, central to southern Free State, and most of the Northern Cape, yet the LPI predicted moderate and even high and extreme risk of lightning in those areas. The area over northern KwaZulu-Natal, where the most lightning activity occurred, was not predicted by the LPI.

Lightning did occur in most of the areas where the LTI probabilities exceeded 60%, over South Africa, except for the area in the northern parts of the Northern Cape. The LTI did identify the area that received the most lightning activity over the northern parts of KwaZulu-Natal and this area was not forecasted by the LPI. Areas over the North West Province and northern to north-eastern parts of the Free State, where lightning did occur, only showed low probabilities of 10% and 30% of lightning risk. The LTI performed better than the LPI over the Free State and the Northern, Western and Eastern Cape, since the forecasted areas were smaller and low probabilities can be seen. The LTI was capable to identify areas of lightning risk over the north-eastern parts of the country and outperformed the other two products as a



lightning prediction tool. However, one should consider that the convective rainfall forecast of the UM endeavours to predict rainfall on the ground and not lightning. It may be that very little rainfall was observed on the ground in the north-eastern parts of South Africa.

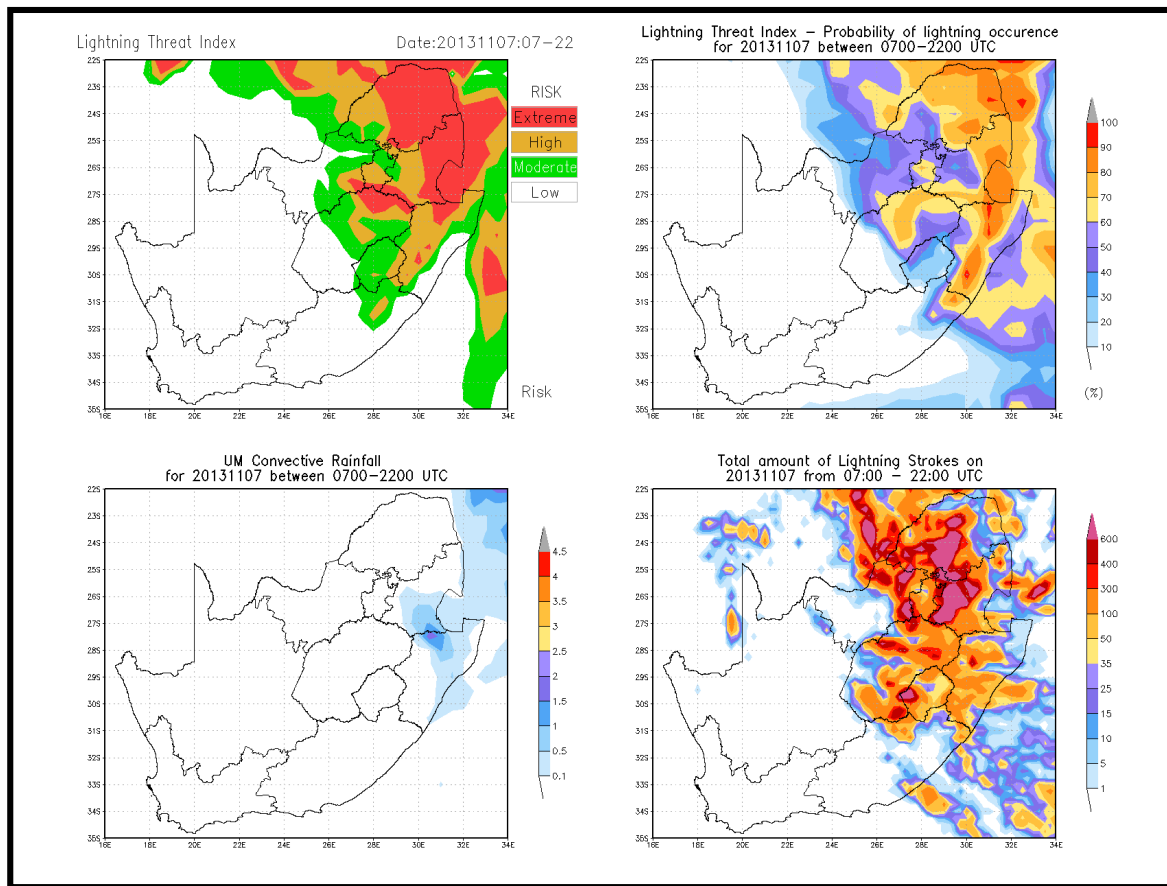


**Figure 5-9: LPI (top left), LTI (top right), UM convective rainfall in cm (bottom left), and occurrence of lightning (bottom right) for 10 October 2013 between 07:00 and 22:00 UTC.**

### 5.4.3 Case 3: 7 November 2013

On this day, the LPI and the LTI predicted similar areas of lightning over the northern provinces of South Africa and KwaZulu-Natal (Figure 5-10). Lightning was observed in a similar area although slightly west of the predicted areas. In this case, the LTI again outperformed the UM convective rainfall forecast.

The UM convective rainfall forecast, predicted rainfall over southern Mpumalanga and KwaZulu-Natal. Lightning did occur over these areas, but the forecast completely missed the lightning over the Limpopo Province, Gauteng, north-eastern parts of the North West Province, Free State, eastern parts of KwaZulu-Natal and Lesotho.



**Figure 5-10: LPI (top left), LTI (top right), UM convective rainfall in cm (bottom left), and occurrence of lightning (bottom right) for 7 November 2013 between 07:00 and 22:00 UTC.**

Most of the areas where the LPI predicted a moderate to extreme risk of lightning matched up with the observed lightning. The LPI failed to identify the lightning over the south-western parts of the Free State and along the coast of KwaZulu-Natal. It also somewhat struggled to predict the lightning over the North West Province. In general, it predicted the lightning further east than the actual occurrence.

The LTI produced a similar forecast compared to the LPI, where the areas of lightning probability exceeding 60% were similar to the high and extreme risk areas of the LPI, and probabilities greater than 40% matching up with the moderate risk areas. The LTI forecasted the lightning along the coast and off the shore of KwaZulu-Natal better than the LPI, while areas of observed lightning over the North West Province were also captured more accurately by the LTI.

On 7 November 2013, the convective rainfall forecast missed most of the observed lightning, while the LPI and LTI were very similar. The moderate to extreme risk areas of the

LPI forecast, and LTI probabilities exceeding 40% predicted the occurrence of lightning fairly well. Some areas were missed by both the LPI and LTI, but the LTI performed slightly better than the LPI.

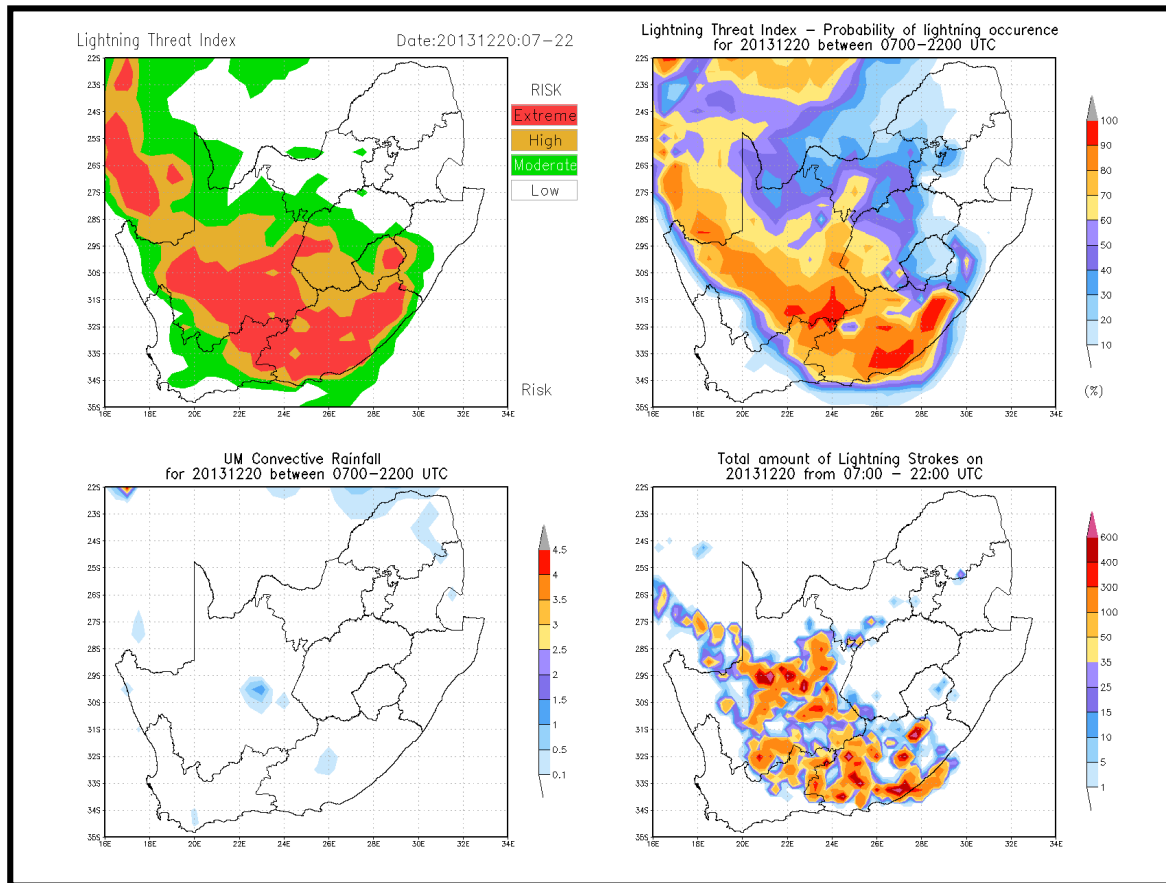
#### **5.4.4 Case 4: 20 December 2013**

On 20 December 2013, lightning occurred over the southern to north-western parts of South Africa (Figure 5-11). Most lightning occurred over the Northern Cape, north-western parts of the Western Cape, the Eastern Cape, and southern Namibia. A few isolated storms can be seen over the North West Province, Free State, Lesotho, KwaZulu-Natal, Gauteng, Limpopo and Mpumalanga.

The UM convective rainfall forecast did not perform at all, where only small areas over the Northern Cape, Eastern Cape and Namibia predicted convective activity. The areas over Botswana, Zimbabwe, Swaziland, Limpopo Province and Western Cape did not match up with lightning occurrence. Most of the areas where lightning occurred were missed by the convective rainfall forecast.

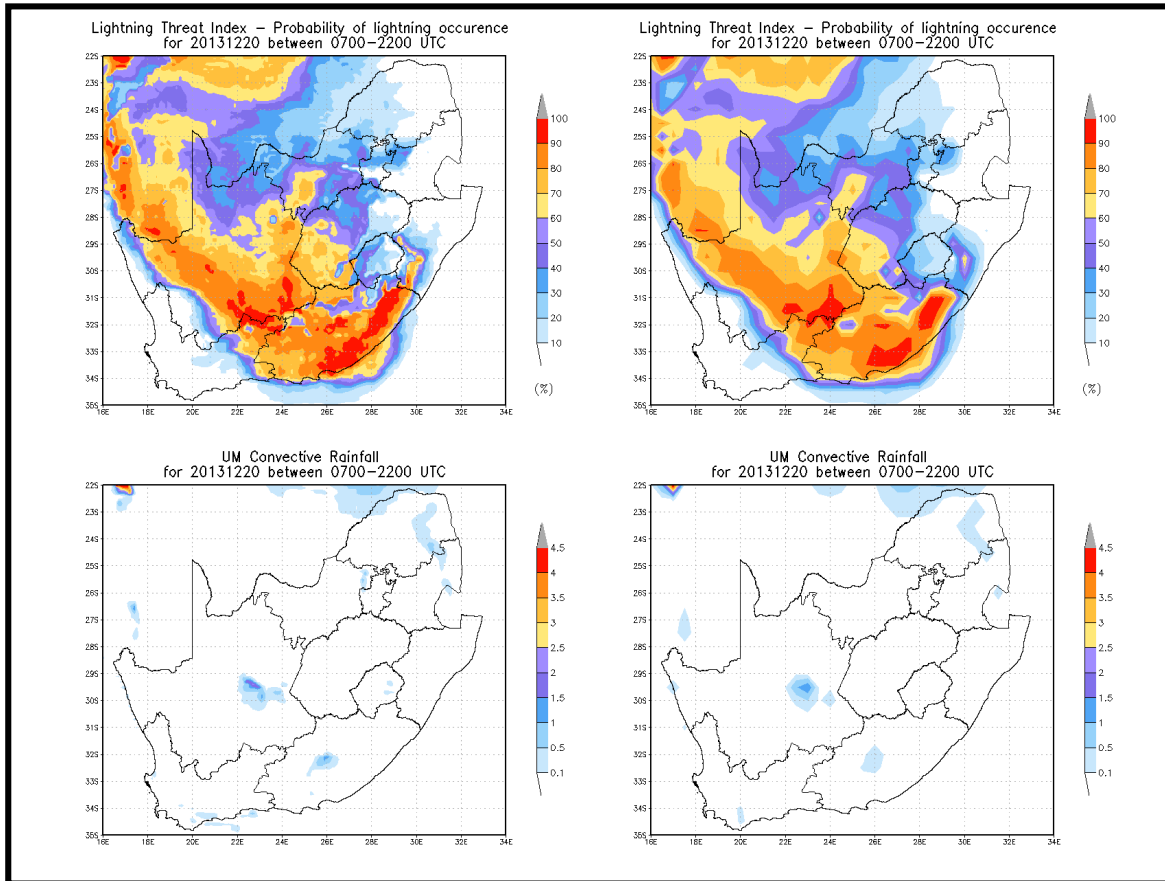
The LPI performed well on this day. Areas of high to extreme risk compared well with the observed lightning. There was some over-forecasting over the Free State, while the lightning off the coast of the Eastern Cape and over the North West Province fell in the moderate risk category. The isolated storms over Limpopo, Mpumalanga, Gauteng and western parts of the North West Province were not forecasted by the LPI.

Like the LPI, the LTI also performed well. Most of the lightning activity occurred in the areas where the LTI probabilities exceeded 60%. The remaining lightning activity was captured by the lower probabilities. As with the LPI, the LTI over-forecasted the lightning over the Free State, but was able to forecast the lightning off the coast of the Eastern Cape and also the isolated storms over Limpopo, Mpumalanga, Gauteng and western parts of the North West Province. Both the LPI and LTI outperformed the convective rainfall as a lightning prediction tool.



**Figure 5-11: LPI (top left), LTI (top right), UM convective rainfall in cm (bottom left), and occurrence of lightning (bottom right) for 20 December 2013 between 07:00 and 22:00 UTC.**

Figure 5-12 shows a comparison between the LTI and convective rainfall forecast on the actual UM resolution against the LTI and convective rainfall forecast on the  $0.5^\circ \times 0.5^\circ$  resolution utilised in this study. Due to the extensive computational time, the amount of computing resources available and the long periods considered, the coarser grid had to be implemented in this study. If the model grid resolution was utilised in this study, there would have been a 22 times increase in computation time of all the parameter extractions, calculations and the regression analysis. It was thus decided that utilising the model grid would have been unfeasible. This comparison shows that the increase in resolution provides much more detail in the forecasts, but the general patterns of the forecasts are similar. The similarities between the higher and lower resolution forecasts also demonstrate that the equations developed from the coarser resolution can be directly applied to a higher resolution model. As such, the coarser grid was used only for the development of the LTI, and the equations can still be applied to a higher resolution model when the product becomes operationally available.



**Figure 5-12: LTI on UM 12 km resolution (top left), LTI on 0.5° X 0.5° resolution (top right), UM convective rainfall in cm on UM 12 km resolution (bottom left), and UM convective rainfall on 0.5° X 0.5° resolution (bottom right) for 20 December 2013 between 07:00 and 22:00 UTC.**

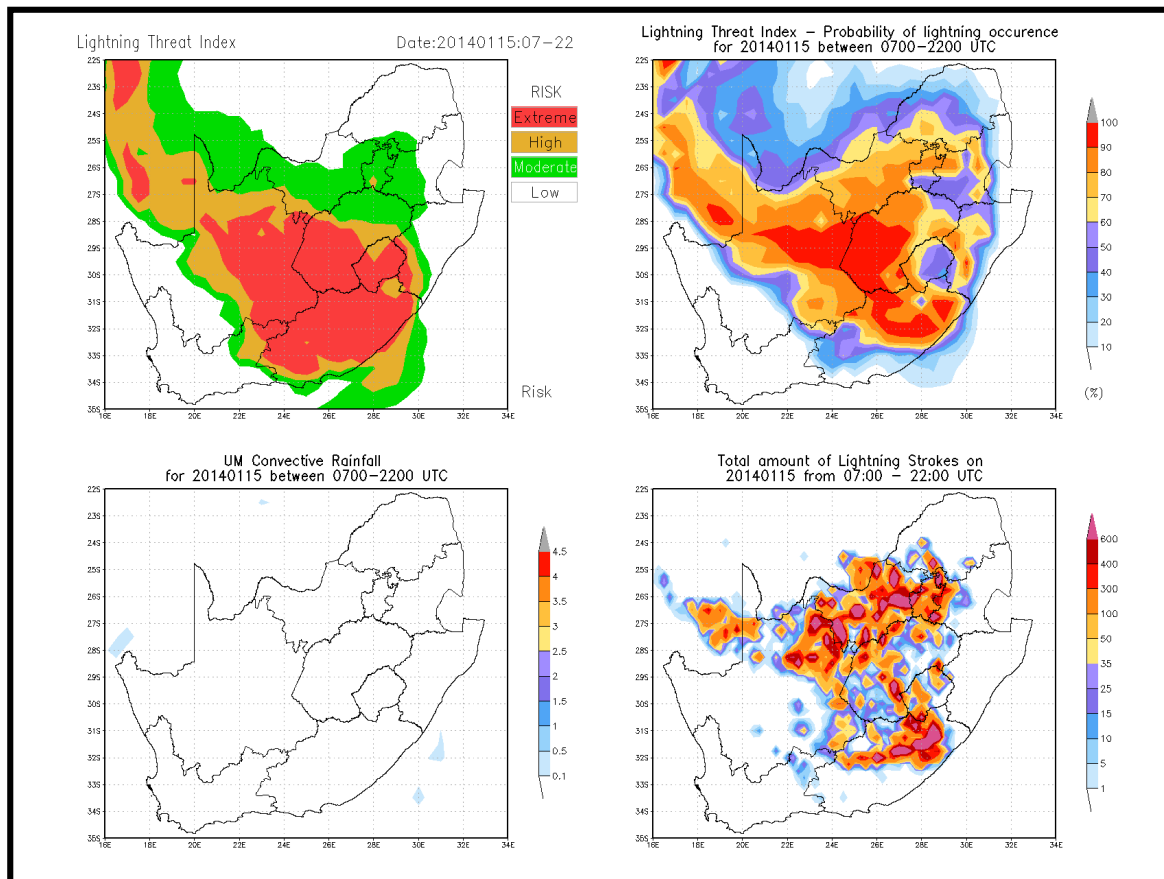
#### 5.4.5 Case 5: 15 January 2014

Figure 5-13 shows that on the 15<sup>th</sup> of January 2014, lightning occurred over the central parts of the country, extending into Namibia. The UM convective rainfall forecast failed to predict any of the lightning that occurred over the country.

The LPI did not perform too well on this day. Lightning risk were in the high and extreme risk categories over the northern parts of the Eastern Cape, the Free State, Northern Cape and Namibia however, certain parts were over-forecasted. The lightning over the northern parts of the Free State, Gauteng, eastern Mpumalanga, and the North West Province were predicted as being at moderate risk of lightning. Areas over the northern parts of the North West Province, Limpopo Province and Northern Cape were not predicted by the LPI.

The LTI performed really well on this day. When comparing the area of lightning activity with the forecast area of the LTI, one can see that there is an excellent agreement between the two patterns. Most of the lightning activity was observed in the areas where the

probabilities  $\geq 60\%$ . The remaining lightning activity occurred in areas with lower probabilities, especially when the probabilities were  $\geq 40\%$ . Some areas over Mpumalanga and KwaZulu-Natal were over-forecasted, while the small isolated storms over the southern parts of the Northern Cape were missed.

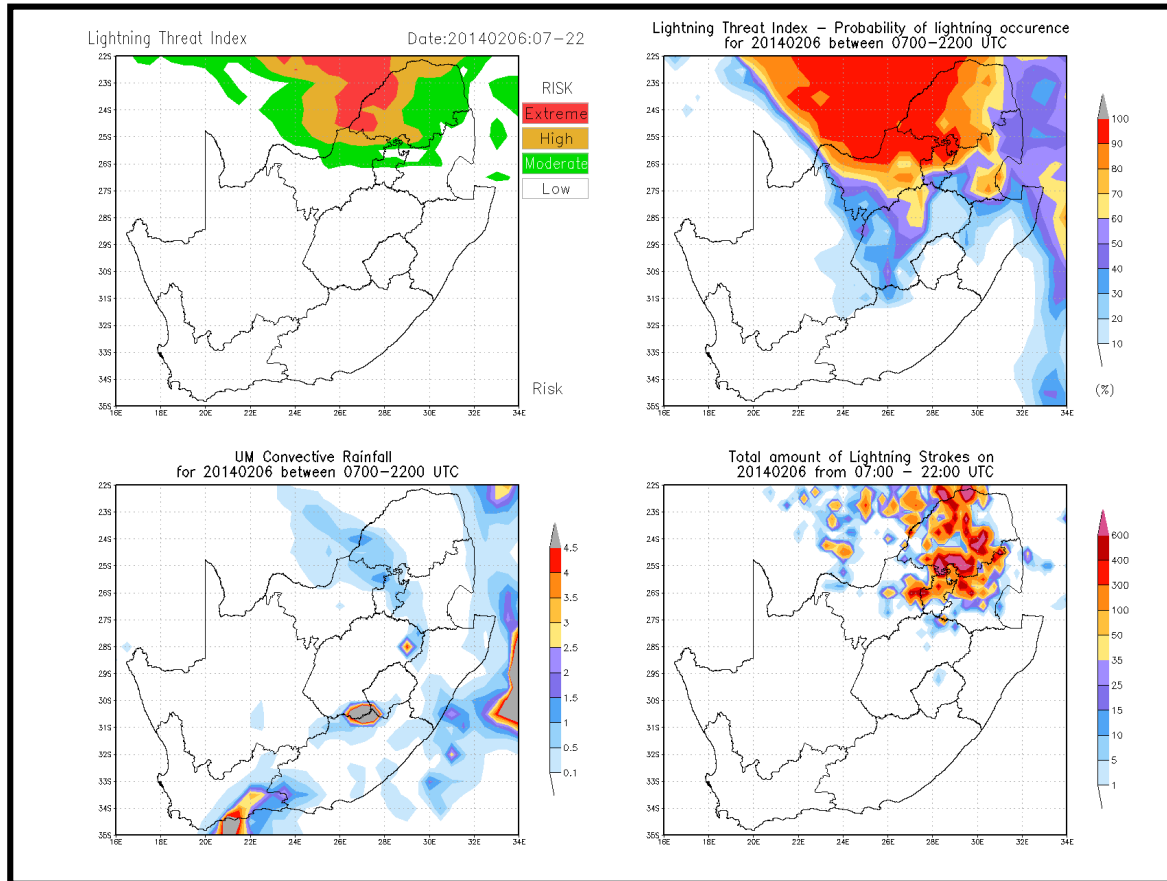


**Figure 5-13: LPI (top left), LTI (top right), UM convective rainfall in cm (bottom left), and occurrence of lightning (bottom right) for 15 January 2014 between 07:00 and 22:00 UTC.**

#### 5.4.6 Case 6: 6 February 2014

The last case study considered, 6 February 2014, is shown in Figure 5-14. During this day, lightning occurred over the northern parts of South Africa and Botswana. The UM convective rainfall forecast expected rainfall along the Western Cape coast, southern Free State and northern parts of the Eastern Cape, as well as over the Indian Ocean. No lightning was however recorded in these areas. Rainfall was forecasted over parts of Gauteng, Mpumalanga, North West Province, Limpopo Province and Botswana, where lightning did

occur. Most of the lightning over the Limpopo and Mpumalanga provinces were not identified by the convective rainfall forecast.



**Figure 5-14: LPI (top left), LTI (top right), UM convective rainfall in cm (bottom left), and occurrence of lightning (bottom right) for 6 February 2014 between 07:00 and 22:00 UTC.**

The LPI under forecasted the occurrence of lightning. Only a small area of high and extreme risk was seen over eastern Botswana, the northern parts of the North West Province and western parts of the Limpopo province. The LPI did not perform well with predicting the lightning over Mpumalanga, Gauteng, and some parts of Botswana. It also failed to predict the lightning over the Free State and some parts of the North West Province.

The LTI performed well. Most lightning activity were seen when the probabilities exceeded 60%, while some lightning also occurred at lower probabilities. Some over-forecasting by the LPI occurred, especially over southern Botswana where there were gaps between the storms. The LTI forecast however outperformed the LPI and convective rainfall forecast on this day.

## 5.5 SUMMARY

The evaluation of the newly developed LTI was presented in this chapter. A quantitative evaluation was performed over an entire SON and DJF season, which was independent from the two seasons utilised in the development of the LTI. Both probabilistic as well as dichotomous evaluation techniques were utilised to perform the quantitative evaluation of the LTI. The new LTI, together with the LPI and UM convective rainfall forecast, were compared with lightning observations by means of six case studies.

A probabilistic evaluation showed that the LTI forecasts have a high sensitivity and specificity for both the SON and DJF seasons. The LTI is not as reliable during the SON season, since it over-forecasts the occurrence of lightning. During the DJF season, the LTI forecast is reliable, only slightly over-forecasting the lightning. Lastly, the results also show that the LTI produces sharp forecasts during both the SON and DJF seasons.

The evaluation statistics from the dichotomous verification show that for both the SON and DJF seasons the LTI can successfully identify the lightning events while at the same time also identify the non-events. This is seen in the high POD values for all probability thresholds while the POFD values are low. The FAR is high, which indicates that the LTI does over-forecast lightning occurrence, especially during the SON season. This was confirmed by the results from the frequency bias. HK scores show that LTI probabilities exceeding 60% for the SON period, and 40% for the DJF period, can be good threshold values to anticipate lightning activity.

Six case studies were presented, where the new LTI, original LPI and UM convective rainfall forecasts were compared to observations of lightning. From all three of the products, the LTI provided the most accurate forecast for all six of the days. The convective rainfall forecast performed poorly as a lightning prediction tool, while the LPI did reasonably well. The LTI provided an accurate forecast in five of the cases, while the LPI did well in three of them. All three of the products underperformed on the day when the LTI did not perform too well, and shows that there will be cases when the forecasts will struggle. This is due to the NWP model data used as input to the forecasts not being able to provide an accurate representation of reality on particular days. On the three days when the LPI performed well, the forecasts from the LPI and LTI were similar, but the LTI provided the superior forecast on these days.



---

# CHAPTER 6

## SUMMARY AND CONCLUSIONS

---

### 6.1 INTRODUCTION

Lightning is a phenomenon that can cause injury or death to humans and animals (Blumenthal et al., 2012), damage to infrastructures (Lynn and Yair, 2010), and can be a hazard to various sectors like the aviation and forestry industries (Price, 2013). South Africa is a country that experiences large amounts of lightning every year that results in numerous deaths, injuries and damages. In order to reduce the risk there is a need for products that can provide operational weather forecasters and the public with lightning forecasts. Methods exist to aid in the forecasting of thunderstorms and the associated lightning (Kohn et al. 2011, McCaul et al. 2009, Lynn & Yair, 2010), but accurate forecasts remains a challenge. To the author's knowledge, no product currently exists in South Africa that attempts to directly forecast the potential lightning threat.

Past studies have shown many NWP model parameters to be related to lightning formation and have been utilised internationally to produce lightning forecasts. The work done by Frisbie et al., (2009), that motivated this study, is an example of such a lightning forecast. The parameters and methodology utilised by Frisbie et al. (2009) for Colorado in the USA was tested over South Africa and showed promising results to assess the lightning risk over the country. However, atmospheric conditions in South Africa are different from those in the USA, and as a result, the parameters utilised to produce the lightning forecast will be different. The NWP model utilised in Colorado also differ from the one used at the South African Weather Service and therefore the LPI is not ideal for South African conditions and NWP models.

### 6.2 SUMMARY

This section of the chapter provides a summary of all the objectives in this study and highlights the main results.

### 6.2.1 Objective 1: Identification of candidate NWP model predictors

Frisbie et al. (2009) developed a Lightning Potential Index for the State of Colorado in the USA. In their index, a form of Convective Available Potential Energy (CAPE), Lifted Index (LI), Precipitable Water (PW), Equivalent Potential Temperature ( $\Theta_e$ ), Relative Humidity (RH) and air temperature was used. The CAPE, LI and  $\Theta_e$ , which are measures of conditional, latent, and potential instability respectively, provides useful insights into the buoyancy of air parcels in the atmosphere (Peppler, 1988; Kunz, 2007; Dyson et al., 2015). They are also linked to potential updrafts that supply thunderclouds with the hydrometeors necessary for electrification, as well as to distribute these charged particles in a thundercloud (Kuo, 1966; Smith et al., 2000; Bright et al., 2005; Murugavel et al., 2014; Singh and O’Gormon, 2015). In order for thunderstorms to develop and to ensure that sufficient amounts of hydrometeors are present for the electrification of a thundercloud, the availability of moisture is vital (Colson, 1960; Berdeklis and List, 2001; Burrows et al., 2005; Xiong et al., 2006). PW and RH are useful parameters to monitor the amount of moisture available and how close to saturation the atmosphere is. Solar heating plays an important role in the distribution of lightning across the globe since air temperature is related to water vapour concentrations as well the instabilities that drive thunderstorm and lightning development (Bharatdwaj, 2006; Price, 2013; Laliberté et al., 2015).

The first aim of this study was to select the most appropriate NWP parameters capable of predicting the occurrence of lightning over South Africa. Since the six parameters mentioned above are closely related to the processes necessary for the development and electrification of thunderclouds, it seemed reasonable to consider these parameters as possible predictors of lightning over South Africa. Many different forms of these parameters exist that are calculated between different levels in the atmosphere. Frisbie et al. (2009) utilised a specific form of CAPE, RH, LI, PW,  $\Theta_e$ , and air temperature in their LPI that was found useful for the prediction of lightning in Colorado, USA. The atmospheric conditions in South Africa differ from those in the USA, and as a result, it was necessary to identify parameters suitable for local conditions. In this study, different variations of the six parameters were selected from the literature. The selected parameters were found useful in other studies to predict lightning or thunderstorms.

In total, 25 parameters were selected from literature. This include four different CAPE parameters, two LI parameters, seven types of  $\Theta_e$ , three PW parameters, five RH’s, and four different temperature parameters. These parameters are listed in Table 3-1 of Chapter 3. Among them are the six parameters utilised by Frisbie et al. (2009).

## 6.2.2 Objective 2: Selection of the most appropriate NWP model parameters for lightning prediction.

In total 25 NWP model parameters were selected as potential predictors of lightning. It is likely that not all of these parameters will perform equally when predicting the occurrence of lightning. Utilising all 25 parameters to develop an index to provide a lightning forecast is also not desirable. As such, the goal was to identify the most appropriate parameters capable of predicting the occurrence of lightning over South Africa. There were six groups of parameters: CAPE, PW, RH, LI,  $\Theta_e$ , and temperature. In each group, there are different variations of the parameters. These variations of the parameters in every group provide similar information, but are calculated between different levels in the atmosphere. Only one parameter out of each group had to be selected that best describes the conditions favourable for the electrification of a thunderstorm.

In order to select the best parameters to predict lightning, SAS statistical software was used to perform a full (backwards and forwards) stepwise logistic regression by using Firth's Penalised Likelihood method. The parameters from each of the 6 main groups were added in turn to the stepwise regression models. The most appropriate parameter was selected by making use of the chi-square value from the score chi-square statistic. A stepwise regression with R-software was also performed to confirm the results from the SAS analysis. Once again, the parameters from each of the 6 main groups were added in turn to the stepwise regression models, but instead the most appropriate parameter was selected by making use of the AIC. These regression procedures were performed separately for SON and DJF, since the weather patterns differ between the two seasons, which resulted in the selection of different parameters.

The results from the stepwise regression with SAS and R software yielded the same results. For SON the following six parameters were selected:

- Most unstable Convective Available Potential Energy in the 1-6 km above ground level range
- Surface Lifted Index
- Mean Precipitable Water in the 850 to 300 hPa layer
- Minimum Relative Humidity in the 3-6 km above ground level layer
- Equivalent Potential Temperature lapse rate between 700 and 500 hPa

- Mean Temperature in the 850-700 hPa layer

Four out of the six selected parameters for SON were the same for DJF, where only the RH and  $\Theta_e$  parameters differed. The parameters selected for DJF is:

- Most unstable Convective Available Potential Energy in the 1-6 km above ground level range
- Surface Lifted Index
- Mean Precipitable Water in the 850 to 300 hPa layer
- Average Relative Humidity in the 3-6 km above ground level layer
- Equivalent Potential Temperature lapse rate between 850 and 400 hPa
- Mean Temperature in the 850-700 hPa layer

The above-mentioned parameters were found to describe the processes involved in the formation of lightning over South Africa the best, since they were able to distinguish between the events and non-events of lightning occurrence. The selection of these parameters leads to the following conclusions:

1. CAPE is strongly related to the updraft velocities in thunderstorms and plays an important role in the distribution of hydrometeors responsible for lightning formation (Murugavel et al., 2014; Singh and O’Gormon, 2015). Sufficient values of CAPE are required in the 0°C to -20°C level of a thunderstorm to ensure that the updraft provides the hydrometeors necessary for electrification to occur (Bright et al., 2005). The 0°C to -20°C level in a South African thundercloud is found at about 3-6 km above ground level. Since CAPE is closely related to updraft velocities in thunderclouds, sufficient CAPE is required as low as 1-3 km AGL to feed the charge separation zone of the storm (3-6 km AGL) with hydrometeors for electrification to occur. Updrafts are also important inside the separation zone to transport the positively charged ice particles to the top of the cloud. This can explain why the most unstable convective available potential energy in the 1 - 6 km above ground level range parameter performed the best since strong updrafts are required below and inside the charge separation zone of a thundercloud.

2. As with CAPE, the surface lifted index is a good indicator of updraft strength and indicates the availability of the hydrometeors responsible for electrification to occur in a thundercloud (Singh and O’Gormon, 2015). The surface lifted index provides the difference between the actual temperature of the environment at 500 hPa and the temperature a parcel will have when it was lifted adiabatically from the surface to the 500 hPa level (Haklander and Van Delden, 2003). This 500 hPa level corresponds with the charge separation zone in a thundercloud, where sufficient hydrometeors are required for electrification to occur.
3. Moisture is one of the requirements for lightning formation (or thunderstorm formation) and precipitable water is used to estimate the amount of moisture available in the atmosphere (Burrows et al., 2005; Duplika and Reuter, 2006). The mean precipitable water in the 850 to 300 hPa layer corresponds well with the charging zone in a thundercloud and shows that sufficient moisture is required between the 850 hPa and 300 hPa levels in the atmosphere for electrification to occur in storm.
4. RH provides information on how close to saturation the atmosphere is and this influences lightning formation. Studies by Berdeklis and List (2001) have shown that there is a good correlation between RH and the charge transfer in the electrification zone of a thundercloud. Xiong et al. (2006) found that lightning corresponds well with RH in longitudinal belts where RH values are greater than 74% and negatively with RH in longitudinal belts where the RH is less than 72%. They also found a similar result for latitudinal belts. This means that higher RH values in dry regions causes more lightning activity, while high RH values in wet regions lowers the levels of lightning activity. This same study also showed that when the RH is too high (greater than 72%) in wet regions, lightning can be suppressed. When considering the typical charge structure of a thundercloud, 3-6 km above the ground corresponds well with the charge separation zone of the cloud. The output from the regression analysis shows that the minimum RH in the 3 - 6 km above ground level layer is the most appropriate RH parameter during SON and the average RH in the 3 - 6 km above ground level layer is the most appropriate for DJF to predict the occurrence of lightning.
5. The lapse rate of  $\Theta_e$  is useful to assess instabilities necessary for thunderstorms development, can be useful to assess changes in air masses, and can be used to estimate updraft velocities (Madhulatha et al., 2013; Dyson et al., 2015; Houston and Wilhelmson, 2012; Cummings, 2013; Huntrieser et al., 2007; Kuo, 1966). This means that  $\Theta_e\Gamma$  is useful to describe the potential updrafts in thunderstorms, which supplies hydrometeors for charge separation in a thunderstorm. The differences in the

parameters selected for SON and DJF can be due to the atmospheric conditions in South Africa being different between the seasons. In early summer, the atmospheric circulation is generally extra-tropical with a conditionally unstable atmosphere over SA, while in late summer the circulation is tropical with a convectively unstable atmosphere (Dyson et al., 2015).  $\Theta_e\Gamma$  measures the convective instability of a layer in the atmosphere.  $\Theta_e\Gamma$  between 850 and 400 hPa fares well in both seasons but from these results, a thinner layer of convective instability in SON (between 700 and 500 hPa) is better associated with lightning.

6. Surface heating from the sun is responsible for the convective processes, which result in atmospheric instabilities (Bharatdwaj, 2006), and therefore plays a big role in the development of thunderstorms. Many studies have shown that lightning is closely related to the surface temperature. The regression analysis above shows that temperatures between the 850 hPa to 700 hPa levels play an important role in lightning formation.

### 6.2.3 Objective 3: Development of a new lightning prediction index

After the most appropriate parameters, capable of predicting the occurrence of lightning, were identified, the development of the new LTI could commence. Since the goal was to develop a single index that utilises the different model predictors to forecast the binary outcome of lightning occurrence (yes or no), attention was given to binary logistic regression techniques. Logistic regression is often used to predict the probability of an event by means of a set of predictors (Kiezun et al., 2009). This technique fits the goal, and will provide a probability forecast of lightning occurrence. Initial tests in the development of the LTI by means of ordinary logistic regression revealed that probabilities of lightning occurrence are extremely low. Upon further investigation it was discovered that a study by King and Zeng (2001) showed that ordinary logistic regression often underestimates the probabilities of rare events (Guns and Vanacker, 2012). This underestimation is due to the logistic regression favouring the larger amount of non-events (0's) compared to the smaller amount of events (1's) when developing a model. This proved to be true in the development of the LTI, and it was decided to make use of an approach developed by King and Zeng (2001) to perform a rare event logistic regression. In the development of the LTI the following approach was followed:

1. Take all the events (1's or lightning occurrences) in the dataset and select a random sample of non-events (0's or no lightning occurrences) with equal size from the data.

2. Run the 'Zelig' package in R to perform a rare event logistic regression with the bias correction and addition of the correction term to the estimated probabilities.
3. Repeat the above steps 1000 times by selecting a new sample of random non-events (0's or no lightning occurrences). The random samples of non-events are taken with repetition where the non-events of the previous sample are added back to the dataset and have the chance to be chosen again.

The 1000 models produced by the procedure above were combined by averaging their output. This means that the average of the intercept term and regression coefficients of the 1000 models was calculated. This process is similar to the bootstrap aggregating technique that aims to improve any instability found in the estimation of the regression output (Kotsiantis et al., 2006).

This process was performed on both SON and DJF. Lightning and model data for two SON and DJF seasons were used in the development of the LTI. The six parameters selected for both seasons in the previous section were utilised in the steps described above. During the analysis of the output from the regression procedure, it was discovered that for SON the  $\mu\text{CAPE}_{1,6\text{km\_AGL}}$  parameter was not significant since it had a large P-value. The  $\mu\text{CAPE}_{1,6\text{km\_AGL}}$  parameter was discarded from the SON list of predictors, and the regression analysis was repeated with the remaining five parameters. During DJF all six of the selected parameters, including  $\mu\text{CAPE}_{1,6\text{km\_AGL}}$  were used.

The regression coefficients obtained from the analysis was then used to set up the equations for the LTI. The equation of the LTI for SON contains five parameters (Equation 4-1 in Chapter 4), while the equation for DJF contains six parameters (Equation 4-2 in Chapter 4). Both equations make use of the selected NWP model parameters to predict the probability of lightning occurrence. This new LTI utilises NWP model output early in the morning to provide a probability forecast of where lightning is expected to occur during the day between 07:00 and 21:00 UTC.

#### **6.2.4 Objective 4: Quantitative evaluation of the new lightning prediction index**

A quantitative evaluation of the new LTI was performed over an entire SON and DJF season, which was independent from the two seasons utilised in the development of the LTI.

Both probabilistic as well as dichotomous evaluation techniques were considered in the quantitative evaluation of the LTI.

A probabilistic evaluation showed that the LTI forecasts have a high sensitivity and specificity for both SON and DJF. The LTI is not as reliable during SON as DJF, since it over-forecasts the occurrence of lightning quite significantly, but during DJF, the LTI forecast is reliable, only slightly over-forecasting the lightning activity. Lastly, the results also show that the LTI produces sharp forecasts during both SON and DJF.

The evaluation statistics from the dichotomous verification show that for both SON and DJF the LTI can successfully identify the lightning events while at the same time also identify the non-events. This is because the POD values are high for all probability thresholds while the POFD values are low. The FAR is high, which indicates that the LTI does over-forecast lightning occurrence, especially during SON. This was confirmed by the results from the frequency bias. HK scores show that LTI probabilities exceeding 60% for SON, and 40% for DJF, can be good threshold values to anticipate lightning activity.

### **6.2.5 Objective 5: Qualitative evaluation of the new index against existing products**

The new LTI was also evaluated qualitatively by means of six case studies, where the forecasts from the LTI, the Frisbie et al. (2009) LPI, and the Unified Model convective precipitation forecast, were compared with lightning observations. From all three of the products, the LTI provided the most accurate forecast for all six of the days. The convective rainfall forecast performed poorly, while the LPI did reasonably well. The LTI provided an accurate forecast in five of the cases, while the LPI did well in three of them. All three of the products underperformed on the day when the LTI did not perform well, and shows that there will be cases when the forecasts will be wrong. This is due to the NWP model data used as input to the forecasts not being able to provide an accurate representation of reality on particular days. On the three days when the LPI performed well, the forecasts from the LPI and LTI were similar, but the LTI provided the superior forecast on these days.



### 6.3 RECOMMENDATIONS

The following recommendations can be made from this study:

1. Six types of NWP model parameters were utilised in this study and showed the ability to distinguish between lightning events and non-events. Many other types of parameters not considered in this study can also be obtained from NWP models. These parameters should be investigated and might add additional value to the lightning forecast. As an example, vertical velocities from the model might also be useful in a lightning forecast since it can provide information on the updrafts that feeds a storm with the hydrometeors necessary for lightning formation.
2. Many of the lightning prediction techniques, including the LTI developed in this study, make use of thermodynamic parameters to forecast lightning. These techniques perform well to forecast the occurrence of lightning, but have a lower capability to forecast the amount of lightning. Many NWP models are capable of forecasting various microphysical parameters related to the charge separation in thunderstorms, and has proved to be very capable in the prediction of lightning amounts. Studies are needed to investigate the plausibility of using cloud microphysical fields from NWP models to predict the occurrence of lightning in South Africa.
3. Due to the extensive computation times involved in this study because of the long periods considered, the LTI was developed on a  $0.5^\circ \times 0.5^\circ$  grid over South Africa. The NWP model utilised in this study produced its output on a higher resolution than considered in this study, and was recently upgraded to produce forecasts on an even higher resolution. The LTI should be tested on the higher resolution models to determine if the forecasts will provide more detail.
4. This new LTI product should be made available to operational weather forecaster or users interested in lightning forecasts, since it can be useful to give guidance early in the morning on the areas of interest where lightning can be expected during the day and hopefully aid in the warning of end users. Users of the LTI should however be aware of the uncertainty associated with any weather forecast and especially when using NWP data to generate the forecast.
5. The LTI forecasts have high sensitivity, specificity and sharpness but are not so reliable, especially during SON, since it over-forecasts the occurrence of lightning quite significantly. For DJF, the LTI forecast is reliable, only slightly over-forecasting the

lightning activity. More research is required to identify the causes of the overestimation in order to improve the forecast.

6. The Unified Model, which is the operational NWP model at the South African Weather Service, covers the entire SADC domain, and as such, this product can be extended to cover other African countries south of the equator. A lightning forecast over these developing countries, where limited technology is available, can add tremendous value.

## 6.4 CONTRIBUTION TO SCIENCE

In this study several NWP model parameters, calculated from the UM, were identified to be useful to predict the occurrence of lightning over South Africa. Not only could these model parameters be used in the development of a new LTI for South Africa, but they can also be used in future lightning studies since they have been proved to be closely related to the conditions that govern the formation of lightning. The LTI developed in this study is the first of its kind in South Africa. Although many similar lightning prediction products have been developed in other countries, to the author's knowledge no other products exist in South Africa that directly attempts to predict the occurrence of lightning. Many lightning prediction schemes exist in other countries that used ordinary logistic regression techniques for the development of the products. This study proposes a rare-event logistic regression approach to predict the occurrence of lightning, which is an under-utilised technique in the field of lightning prediction. The approach followed in this study is not restricted to the prediction of lightning and can be applied to similar studies across various disciplines. This work also demonstrates that historical lightning data from the sophisticated SALDN operated by SAWS can be utilised in many different ways to improve our understanding of thunderstorms and their associated lightning. The LTI will be a useful tool to operational weather forecasters or sectors interested in lightning forecasts, to provide guidance early in the morning on the areas of interest where lightning can be expected during the day, and can ultimately contribute to society by aiding with timely warnings of lightning or thunderstorms to protect humans, animals and property.

## REFERENCES

---

- AHRENS, C.D., JACKSON, P.L. and JACKSON, C.E.J., 2012. *Meteorology Today: An Introduction to Weather, Climate, and the Environment*. Cengage Learning.
- BARLOW, W., 1993. A new index for the prediction of deep convection, *Proceedings of the 17<sup>th</sup> conference on severe local storms*. St. Louis, MO, 1993. Amer. Meteorol Soc, pp. 129-132.
- BENSON, R.P., 2005. Predicting Lightning Strikes for the Enhancement of Fire Weather Forecasts, *Proceedings of the 6<sup>th</sup> Symposium on Fire and Forest Meteorology*. Canmore, Alberta, 2005. Amer. Meteorol Soc,
- BERDEKLIS, P. and LIST, R., 2001. The Ice Crystal–Graupel Collision Charging Mechanism of Thunderstorm Electrification. *Journal of the Atmospheric Sciences*, **58**(18), pp. 2751-2770.
- BETZ, H.-D., MARSHALL, T.C., STOLZENBURG, M., SCHMIDT, K., OETTINGER, W.P., DEFER, E., KONARSKI, J., LAROCHE, P. and DOMBAI, F., 2008. Detection of in-cloud lightning with VLF/LF and VHF networks for studies of the initial discharge phase. *Geophysical Research Letters*, **35**(23).
- BEWICK, V., CHEEK, L. and BALL, J., 2004. Statistics review 13: Receiver operating characteristic curves. *Critical Care*, **8**(6), pp. 508-512.
- BHARATDWAJ, K., 2006. *Physical Geography: Hydrosphere*. New Delhi: Discovery Publishing House.
- BHAVIKA, B., 2007. *The influence of terrain elevation on lightning density in South Africa*. MSc thesis, University of Johannesburg. [viewed 25 April 2016]. Available from: <http://hdl.handle.net/10210/3066>
- BLANCHARD, D.O., 1998. Assessing the Vertical Distribution of Convective Available Potential Energy. *Weather and Forecasting*, **13**(3), pp. 870-877.
- BLUESTEIN, H.B., MCCAUL JR, E.W., BYRD, G.P. and WOODALL, G.R., 1988. Mobile Sounding Observations of a Tornadic Storm near the Dryline: The Canadian, Texas Storm of 7 May 1986. *Monthly Weather Review*, **116**(9), pp. 1790-1804.
- BLUMENTHAL, R., TRENGROVE, E., JANDRELL, I.R., SAAYMAN, G., 2012. Lightning medicine in South Africa. *South African Medical Journal*, **102**(7), pp. 625-626.
- BOLOGNA, F., BRITTEN, A. and VOSLOO, H., 2001. Current research into the reduction of the number of transmission line faults on the Eskom MTS, *Proceedings of the 2<sup>nd</sup> South African Electric Power Research Conference*. South Africa, 2001.
- BOLTON, D., 1980. The Computation of Equivalent Potential Temperature. *Monthly Weather Review*, **108**(7), pp. 1046-1053.

- BOTHWELL, P.D., 2008. Predicting the location and intensity of lightning using an experimental automated statistical methods. *Proceedings of the 3<sup>rd</sup> Conference on Meteorological Applications of Lightning Data*. New Orleans, LA, 2008. Amer. Meteor. Soc.
- BRIGHT, D.R., WANDISHIN, M.S., JEWELL, R.E. and WEISS, S.J., 2005. A physically based parameter for lightning prediction and its calibration in ensemble forecasts. *Proceedings of the Conference on Meteorological Applications of Lightning Data*. San Diego, CA, 2005. Amer. Meteor. Soc.
- BRÖCKER, J. and SMITH, L.A., 2007. Increasing the Reliability of Reliability Diagrams. *Weather and Forecasting*, **22**(3), pp. 651-661.
- BUNKERS, M.J., KLIMOWSKI, B.A. and ZEITLER, J.W., 2002. The importance of parcel choice and the measure of vertical wind shear in evaluating the convective environment. *Proceedings of the 21<sup>st</sup> Conference on Severe Local Storms*. San Antonio, TX 2002. Amer. Meteor. Soc., pp. 379-382.
- BURNHAM, K.P. and ANDERSON, D.R., 2004. Multimodel Inference Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, **33**(2), pp. 261-304.
- BURROWS, W.R., PRICE, C. and WILSON, L.J., 2005. Warm Season Lightning Probability Prediction for Canada and the Northern United States. *Weather and Forecasting*, **20**(6), pp. 971-988.
- CALLADO, A., ESCRIBÀ, P., GARCÍA-MOYA, J.A., MONTERO, J., SANTOS, C., SANTOS-MUÑOZ, D. and SIMARRO, J., 2013. Ensemble Forecasting. In: ZHANG, Y. and RAY, P., ed. *Climate change and regional/local responses*. Intech, pp. 3-57.
- CHAURASIA, A. and HAREL, O., 2012. Using AIC in multiple linear regression framework with multiply imputed data. *Health Services and Outcomes Research Methodology*, **12**(2-3), pp. 219-233.
- CHRISTIAN, H.J., BLAKESLEE, R.J., BOCCIPPIO, D.J., BOECK, W.L., BUECHLER, D.E., DRISCOLL, K.T., GOODMAN, S.J., HALL, J.M., KOSHAK, W.J., MACH, D.M. and STEWART, M.F., 2003. Global frequency and distribution of lightning as observed from space by the Optical Transient Detector. *Journal of Geophysical Research: Atmospheres*, **108**(D1), pp. ACL 4-1-ACL 4-15.
- CHURCH, C.R., 1966. *The electrification of hail*, PhD thesis, Durham University. [viewed 28 September 2015]. Available from: <http://etheses.dur.ac.uk/8779/>
- CINTINEO, J., LAKSHMANAN, V. and SMITH, T., 2010. Performance of a Probabilistic Cloud-to-Ground Lightning Prediction Algorithm, *Proceedings of the 20th Conference on Probability and Statistics in the Atmospheric Sciences*. Atlanta, GA, 2010. Amer. Meteor. Soc.
- COHEN, A.E., CONIGLIO, M.C., CORFIDI, S.F. and CORFIDI, S.J., 2007. Discrimination of Mesoscale Convective System Environments Using Sounding Observations. *Weather and Forecasting*, **22**(5), pp. 1045-1062.
- COLSON, D., 1960. HIGH LEVEL THUNDERSTORMS OF JULY 31–AUGUST 1, 1959. *Monthly Weather Review*, **88**(8), pp. 279-285.

- CONSTANTINESCU, E.M., SANDU, A., CHAI, T. and CARMICHAEL, G.R., 2007. Ensemble-based chemical data assimilation. II: Covariance localization. *Quarterly Journal of the Royal Meteorological Society*, **133**(626), pp. 1245-1256.
- CRAVEN, J.P. and BROOKS, H.E., 2004. Baseline climatology of sounding derived parameters associated with deep, moist convection. *Natl. Wea. Dig.*, **28**(1), pp. 13-24.
- CULLEN, M.J.P., DAVIES, T., MAWSON, M.H., JAMES, J.A., COULTER, S.C. and MALCOLM, A., 1997. An Overview of Numerical Methods for the Next Generation U.K. NWP and Climate Model. *Atmosphere-Ocean*, **35**(sup1), pp. 425-444.
- CUMMINGS, K.A., 2013. *Lightning Flash Rate and Chemistry Simulation of Tropical Island Convection Using a Cloud-resolved Model*. MSc dissertation, University of Maryland. [viewed 19 November 2015]. Available from: [https://dods.atmos.umd.edu/theses\\_archive/2013/kristin/UMd\\_MS\\_ScholarlyPaper\\_Cummings.pdf](https://dods.atmos.umd.edu/theses_archive/2013/kristin/UMd_MS_ScholarlyPaper_Cummings.pdf)
- CUMMINS, K.L., KRIDER, E.P. and MALONE, M.D., 1998a. The US National Lightning Detection Network™ and Applications of Cloud-to-Ground Lightning Data by Electric Power Utilities. *IEEE Transactions on Electromagnetic Compatibility*, **40**(4), pp. 465-480.
- CUMMINS, K.L., MURPHY, M.J. and TUEL, J.V., 2000. Lightning detection methods and meteorological applications, Proceedings of the IV International Symposium on Military Meteorology. Malbork, Poland, 2000, pp. 26-28.
- CUMMINS, K.L., MURPHY, M.J., BARDO, E.A., HISCOX, W.L., PYLE, R.B. and PIFER, A.E., 1998b. A Combined TOA/MDF Technology Upgrade of the U.S. National Lightning Detection Network. *Journal of Geophysical Research: Atmospheres*, **103**(D8), pp. 9035-9044.
- DAVIES, T., CULLEN, M.J.P., MALCOLM, A.J., MAWSON, M.H., STANFORTH, A., WHITE, A.A. and WOOD, N., 2005. A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Quarterly Journal of the Royal Meteorological Society*, **131**(608), pp. 1759-1782.
- DE CONING, E. and POOLMAN, E., 2011. South African Weather Service operational satellite based precipitation estimation technique: applications and improvements. *Hydrology and Earth System Sciences*, **15**(4), pp. 1131-1145.
- DE CONING, E., KOENIG, M. and OLIVIER, J., 2011. The combined instability index: a new very-short range convection forecasting technique for southern Africa. *Meteorological Applications*, **18**(4), pp. 421-439.
- DOSWELL, C.A. and RASMUSSEN, E.N., 1994. The Effect of Neglecting the Virtual Temperature Correction on CAPE Calculations. *Weather and Forecasting*, **9**(4), pp. 625-629.
- DUPIILKA, M.L. and REUTER, G.W., 2006. Forecasting Tornadic Thunderstorm Potential in Alberta Using Environmental Sounding Data. Part II: Helicity, Precipitable Water, and Storm Convergence. *Weather and Forecasting*, **21**(3), pp. 336-346.

- DYSON, L.L., VAN HEERDEN, J. and SUMNER, P.D., 2015. A baseline climatology of sounding-derived parameters associated with heavy rainfall over Gauteng, South Africa. *International Journal of Climatology*, **35**(1), pp. 114-127.
- EMANUEL, K.A., 1994. *Atmospheric Convection*. New York: Oxford University Press.
- EVERT, R. and SCHULZE, G., 2005. Impact of a new lightning detection and location system in South Africa, *Proceedings of the Inaugural IEEE Power Engineering Society Conference and Exposition in Africa*, Durban, South Africa, 2005. IEEE, pp. 356-363.
- FAN, J., UPADHYE, S. and WORSTER, A., 2006. Understanding receiver operating characteristic (ROC) curves. *Can.J.Emerg.Med*, **8**(1), pp. 19-20.
- FARUKH, M.A., HAYASAKA, H. and KIMURA, K., 2011. Characterization of Lightning Occurrence in Alaska Using Various Weather Indices for Lightning Forecasting. *Journal of Disaster Research*, **6**(3), pp. 343-355.
- FAWCETT, T., 2006. An introduction to ROC analysis. *Pattern Recogn.Lett.*, **27**(8), pp. 861-874.
- FINKE, U. and KREYER, O., 2002. *Detect and Locate Lightning Events from Geostationary Satellite Observations*. Report. Institute für Meteorologie und Klimatologie Universität Hannover. Germany.
- FIRTH, D., 1993. Bias reduction of maximum likelihood estimates. *Biometrika*, **80**(1), pp. 27-38.
- FLORKOWSKI, C.M., 2008. Sensitivity, Specificity, Receiver-Operating Characteristic (ROC) Curves and Likelihood Ratios: Communicating the Performance of Diagnostic Tests. *The Clinical Biochemist Reviews*, **29**(Suppl 1), pp. S83-S87.
- FOLSOM JR, M.I., 2004. *Developing a Forecast Tool for Cloud-to-Ground Lightning in the North Central and Northeastern United States*, MSc thesis, Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio [viewed 25 April 2016]. Available from:  
<http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA422988>
- FRISBIE, P., COLTON, J., PRINGLE, J., DANIELS, J. and MEYERS, M., 2013. *A Forecasting Methodology that Uses Moisture Parameters to Pinpoint Locations of Potential Lightning*. Central Region Technical Attachment Number 13-01. Grand Junction, CO: NOAA/National Weather Service.
- FRISBIE, P.R., COLTON, J., PRINGLE, J., DANIELS, J., RAMEY JR, J. and MEYERS, M., 2009. Lightning Prediction by WFO Grand Junction using Model Data and Graphical Forecast Editor Smart Tools, *Proceedings of the 4th Conference on the Meteorological Applications of Lightning Data*. Phoenix, AZ, 2009. American Meteorological Society.
- FUELBERG, H.E. and BIGGAR, D.G., 1994. The Preconvective Environment of Summer Thunderstorms over the Florida Panhandle. *Weather and Forecasting*, **9**(3), pp. 316-326.

- GALWAY, J.G., 1956. The lifted index as a predictor of latent instability. *Bull.Amer.Meteor.Soc*, **37**, pp. 528-529.
- GARREAUD, R.D., NICORA, M.G., BÜRGESESSER, R.E. and ÁVILA, E.E., 2014. Lightning in Western Patagonia. *Journal of Geophysical Research: Atmospheres*, **119**(8), pp. 4471-4485.
- GASKELL, W. and ILLINGWORTH, A.J., 1980. Charge transfer accompanying individual collisions between ice particles and its role in thunderstorm electrification. *Quarterly Journal of the Royal Meteorological Society*, **106**(450), pp. 841-854.
- GIJBEN, M., 2012. The lightning climatology of South Africa. *South African Journal of Science*, **108**(3-4), pp. 44-53.
- GIJBEN, M., 2013. Verification of a Unified Model based lightning risk indicator for Southern Africa, *Proceedings of the 29th Annual conference of South African Society for Atmospheric Sciences*, Durban, South Africa, 2013. ISBN 978-0-620-56626-1, pp. 58-61.
- GILL, T., 2008. *Initial steps in the development of a comprehensive lightning climatology of South Africa*. MSc Dissertation, University of the Witwatersrand.
- GOODMAN, S.J., 1990. *Predicting thunderstorm evolution using ground-based lightning detection networks*. NASA Technical Memo 103521. NASA.
- GROENEMEIJER, P.H. and VAN DELDEN, A., 2007. Sounding-derived parameters associated with large hail and tornadoes in the Netherlands. *Atmospheric Research*, **83**(2-4), pp. 473-487.
- GUNS, M. and VANACKER, V., 2012. Logistic regression applied to natural hazards: rare event logistic regression with replications. *Natural Hazards and Earth System Sciences*, **12**(6), pp. 1937-1947.
- HABY, J., *Severe Weather Indices Page* [online]. The Weather Prediction Webpage. [viewed 3 August 2015]. Available from: <http://www.theweatherprediction.com/severe/indices/>
- HABY, J., *Skew-T: A Look at PW* [online]. The Weather Prediction Webpage. [viewed 10 August 2015]. Available from: <http://www.theweatherprediction.com/habyhints/294>
- HAKLANDER, A.J. and VAN DELDEN, A., 2003. Thunderstorm predictors and their forecast skill for the Netherlands. *Atmospheric Research*, **67**, pp. 273-299.
- HARATS, N., ZIV, B., YAIR, Y., KOTRONI, V. and DAYAN, U., 2010. Lightning and rain dynamic indices as predictors for flash floods events in the Mediterranean. *Advances in Geosciences*, **23**, pp. 57-64.
- HOLLE, R.L., 2008. Annual rates of lightning fatalities by country. *Proceedings of the 20th International Lightning Detection Conference, Tuscon, AZ, 2008*. Vaisala Inc., pp. 21-23.
- HOLLIDAY, J.R., GAGGIOLI, W.J. and KNOX, L.E., 2012. Testing earthquake forecasts using reliability diagrams. *Geophysical Journal International*, **188**(1), pp. 336-342.

- HOLTON, J.R., 2013. *An Introduction to Dynamic Meteorology*. 3<sup>rd</sup> ed. San Diego, CA: Academic Press.
- HOUSTON, A.L. and WILHELMSON, R.B., 2012. The Impact of Airmass Boundaries on the Propagation of Deep Convection: A Modeling-Based Study in a High-CAPE, Low-Shear Environment. *Monthly Weather Review*, **140**(1), pp. 167-183.
- HOUZE, R.A., 1993. *Cloud Dynamics*. Vol 53. San Diego, CA: Academic Press.
- HUNT, H.G.P., NIXON, K.J. and JANDRELL, I.R., 2014. Establishing a methodology to investigate LDN median error ellipses used as corroborating evidence for a lightning event at a specific geographic location. *Electric Power Systems Research*, **113**, pp. 104-114.
- HUNTRIESER, H., SCHLAGER, H., ROIGER, A., LICHTENSTERN, M., SCHUMANN, U., KURZ, C., BRUNNER, D., SCHWIERZ, C., RICHTER, A. and STOHL, A., 2007. Lightning-produced NO<sub>x</sub> over Brazil during TROCCINOX: airborne measurements in tropical and subtropical thunderstorms and the importance of mesoscale convective systems. *Atmospheric Chemistry and Physics*, **7**(11), pp. 2987-3013.
- IMAI, K., KING, G. and LAU, O., 2008. Toward A Common Framework for Statistical Analysis and Development. *Journal of Computational and Graphical Statistics*, **17**(4), pp. 892-913
- IMAI, K., KING, G. and LAU, O., 2009. Zelig: Everyone's statistical software. *R package version*, **3**(5), <http://gking.harvard.edu/zelig>
- INFORMATION TECHNOLOGY AND SYSTEMS CENTRE. *Lightning Imaging Sensor* [online]. The University of Alabama, Huntsville, AL. [viewed August 16, 2015]. Available from: <https://www.itsc.uah.edu/home/projects/lightning-imaging-sensor>.
- JAFFER, G., EICHELBERGER, H.U., SCHWINGENSCHUH, K. and KOUDELKA, O., 2011. A LEO nano-satellite mission for the detection of lightning VHF sferics. In: GARCIA, M., ed. *Adaptive Filtering Applications*. Intech Open Access.
- JAYARATNE, E.R. and KULESHOV, Y., 2006. The relationship between lightning activity and surface wet bulb temperature and its variation with latitude in Australia. *Meteorology and Atmospheric Physics*, **91**(1), pp. 17-24.
- JAYARATNE, E.R., SAUNDERS, C.P.R. and HALLETT, J., 1983. Laboratory studies of the charging of soft-hail during ice crystal interactions. *Quarterly Journal of the Royal Meteorological Society*, **109**(461), pp. 609-630.
- JOLLIFFE, I.T. and STEPHENSON, D.B., 2003. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. West Sussex, England: Wiley.
- KEHRER, K., GRAF, B. and ROEDER, W.P., 2008. Global Positioning System (GPS) Precipitable Water in Forecasting Lightning at Spaceport Canaveral. *Weather and Forecasting*, **23**(2), pp. 219-232.



- KELLER, D.L., 2006. An Algorithm to forecast lightning using only AFWA MM5 model output: "Bolt of lightning technique" (BOLT), *Proceedings of the 19<sup>th</sup> International Lightning Detection Conference and 1<sup>st</sup> International Lightning Meteorology Conference*. Tucson, AZ, 2006. Vaisala Inc.
- KIEZUN, A., LEE, I.T.A. and SHOMRON, N., 2009. Evaluation of optimization techniques for variable selection in logistic regression applied to diagnosis of myocardial infarction. *Bioinformatics*, **3**(7), pp. 311-313.
- KING, G. and ZENG, L., 2001. Logistic Regression in Rare Events Data. *Political Analysis*, **9**, pp. 137-163.
- KLEINBAUM, D.G. and KLEIN, M., 2010. *Logistic Regression: A Self-Learning Text*. 3<sup>rd</sup> ed. New York: Springer.
- KOHN, M., GALANTI, E., PRICE, C., LAGOUVARDOS, K. and KOTRONI, V., 2011. Nowcasting thunderstorms in the Mediterranean region using lightning data. *Atmospheric Research*, **100**(4), pp. 489-502.
- KOSHAK, W.J. and KRIDER, E.P., 1989. Analysis of lightning field changes during active Florida thunderstorms. *Journal of Geophysical Research: Atmospheres*, **94**(D1), pp. 1165-1186.
- KOSHAK, W.J., SOLAKIEWICZ, R.J., BLAKESLEE, R.J., GOODMAN, S.J., CHRISTIAN, H.J., HALL, J.M., BAILEY, J.C., KRIDER, E.P., BATEMAN, M.G., BOCCIPPIO, D.J., MACH, D.M., MCCAUL, E.W., STEWART, M.F., BUECHLER, D.E., PETERSEN, W.A. and CECIL, D.J., 2004. North Alabama Lightning Mapping Array (LMA): VHF Source Retrieval Algorithm and Error Analyses. *Journal of Atmospheric and Oceanic Technology*, **21**(4), pp. 543-558.
- KOTSIANTIS, S.B., KANELLOPOULOS, D. and ZAHARAKIS, I.D., 2006. Bagged Averaging of Regression Models. In: MAGLOGIANNIS, I., KARPUZIS, K. and BRAMER, M., ed. *Artificial Intelligence Applications and Innovations*. Vol 204 of IFIP International Federation for Information Processing. Springer US, pp. 53-60.
- KREHBIEL, P.R., 1986. The Electrical Structure of Thunderstorms. In: Geophysics Study Committee, ed. *The Earth's Electrical Environment*, Washington, D.C.: National Academy Press, pp. 90-113.
- KREHBIEL, P.R., BROOK, M. and MCCRORY, R.A., 1979. An analysis of the charge structure of lightning discharges to ground. *Journal of Geophysical Research: Oceans*, **84**(C5), pp. 2432-2456.
- KRUGER, A.C., 2007. *Climate of South Africa, Precipitation*. Report No. WS47. Pretoria, South Africa: South African Weather Service.
- KUNZ, M., 2007. The skill of convective parameters and indices to predict isolated and severe thunderstorms. *Nat. Hazards Earth Syst. Sci.*, **7**, pp. 327-342.
- KUO, H.L., 1966. On the Dynamics of Convective Atmospheric Vortices. *Journal of the Atmospheric Sciences*, **23**(1), pp. 25-42.

- LALIBERTÉ, F., ZIKA, J., MUDRYK, L., KUSHNER, P.J., KJELLSSON, J. and DÖÖS, K., 2015. Atmospheric dynamics. Constrained work output of the moist atmospheric heat engine in a warming climate. *Science*, **347**(6221), pp. 540-543.
- LAMBERT, W., WHEELER, M. and ROEDER, W., 2005. Objective lightning forecasting at Kennedy space center and Cape Canaveral air force station using cloud-to-ground lightning surveillance system data, *Proceedings of the Conference on Meteorological Applications of Lightning Data*. San Diego, CA, 2005. Amer. Meteor. Soc.
- LANDMAN, S., ENGELBRECHT, F.A., ENGELBRECHT, C.J., DYSON, L.L. and LANDMAN, W.A., 2012. A short-range weather prediction system for South Africa based on a multi-model approach. *Water SA*, **38**(5), pp. 765-774.
- LANG, T.J., MILLER, L.J., WEISMAN, M., RUTLEDGE, S.A., BARKER, L.J., BRINGI, V.N., CHANDRASEKAR, V., DETWILER, A., DOESKEN, N., HELSDON, J., KNIGHT, C., KREHBIEL, P., LYONS, W.A., MACGORMAN, D., RASMUSSEN, E., RISON, W., RUST, W.D. and THOMAS, R.J., 2004. The Severe Thunderstorm Electrification and Precipitation Study. *Bulletin of the American Meteorological Society*, **85**(8), pp. 1107-1125.
- LIVINGSTON, E.S., NIELSEN-GAMMON, J. and ORVILLE, R.E., 1996. A Climatology, Synoptic Assessment, and Thermodynamic Evaluation for Cloud-to-Ground Lightning in Georgia: A Study for the 1996 Summer Olympics. *Bulletin of the American Meteorological Society*, **77**(7), pp. 1483-1495.
- LOTT, F.C., GORDON, M., GRAHAM, R.J., SCAIFE, A.A. and VELLINGA, M., 2014. Reliability of African climate prediction and attribution across timescales. *Environ.Res.Lett.*, **9**(10), pp. 104017-104023.
- LUCAS, C., ZIPSER, E.J. and LEMONE, M.A., 1994. Convective Available Potential Energy in the Environment of Oceanic and Continental Clouds: Correction and Comments. *Journal of the Atmospheric Sciences*, **51**(24), pp. 3829-3830.
- LYNN, B. and YAIR, Y., 2010. Prediction of lightning flash density with the WRF model. *Advances in Geosciences*, **23**, pp. 11-16.
- LYNN, B.H., YAIR, Y., PRICE, C., KELMAN, G. and CLARK, A.J., 2012. Predicting Cloud-to-Ground and Intracloud Lightning in Weather Forecast Models. *Weather and Forecasting*, **27**(6), pp. 1470-1488.
- MACGORMAN, D.R., APOSTOLAKOPOULOS, I.R., LUND, N.R., DEMETRIADES, N.W., MURPHY, M.J. and KREHBIEL, P.R., 2011. The Timing of Cloud-to-Ground Lightning Relative to Total Lightning Activity. *Monthly Weather Review*, **139**(12), pp. 3871-3886.
- MADHULATHA, A., RAJEEVAN, M., VENKAT RATNAM, M., BHATE, J. and NAIDU, C.V., 2013. Nowcasting severe convective activity over southeast India using ground-based microwave radiometer observations. *Journal of Geophysical Research: Atmospheres*, **118**(1), pp. 1-13.
- MALAN, D., 1963. *Physics of Lightning*. London: Eng.Univ.Press.

- MALLICK, S., RAKOV, V.A., NGIN, T., GAMEROTA, W.R., PILKEY, J.T., HILL, J.D., UMAN, M.A., JORDAN, D.M., NAG, A. and SAID, R.K., 2014. Evaluation of the GLD360 performance characteristics using rocket-and-wire triggered lightning data. *Geophysical Research Letters*, **41**(10), pp. 3636-3642.
- MANSELL, E.R., MACGORMAN, D.R., ZIEGLER, C.L. and STRAKA, J.M., 2005. Charge structure and lightning sensitivity in a simulated multicell thunderstorm. *Journal of Geophysical Research: Atmospheres*, **110**(D12101).
- MARKSON, R., 2007. The Global Circuit Intensity: Its Measurement and Variation over the Last 50 Years. *Bulletin of the American Meteorological Society*, **88**(2), pp. 223-241.
- MARKSON, R. and PRICE, C., 1999. Ionospheric potential as a proxy index for global temperature. *Atmospheric Research*, **51**(3), pp. 309-314.
- MASON, S.J. and GRAHAM, N.E., 2002. Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Quarterly Journal of the Royal Meteorological Society*, **128**(584), pp. 2145-2166.
- MAZANY, R.A., BUSINGER, S., GUTMAN, S.I. and ROEDER, W., 2002. A Lightning Prediction Index that Utilises GPS Integrated Precipitable Water Vapor. *Weather and Forecasting*, **17**(5), pp. 1034-1047.
- McCAUL JR, E.W., GOODMAN, S.J., LACASSE, K.M. and CECIL, D.J., 2009. Forecasting Lightning Threat Using Cloud-Resolving Model Simulations. *Weather and Forecasting*, **24**(3), pp. 709-729.
- MURPHY, A.H. and WILKS, D.S., 1998. A Case Study of the Use of Statistical Models in Forecast Verification: Precipitation Probability Forecasts. *Weather and Forecasting*, **13**(3), pp. 795-810.
- MURPHY, M.J., NAG, A., LOJOU, J. and SAID, R.K., 2013. Preliminary analysis of the Vaisala TLS-200 network deployed during the CHUVA campaign in Brazil, *Proceedings of the 6<sup>th</sup> Conference on the Meteorological Applications of Lightning Data*. Austin, TX, 2013. American Meteorological Society.
- MURUGAVEL, P., PAWAR, S.D. and GOPALAKRISHAN, V., 2014. Climatology of lightning over Indian region and its relationship with convective available potential energy. *International Journal of Climatology*, **34**(11), pp. 3179-3187.
- NATIONAL WEATHER SERVICE. *Grand Junction's Lightning Potential Index* [online]. National Weather Service. [viewed 26 April 2013]. Available from: <http://www.weather.gov/gjt/lightningpotentialindex>
- NCAR - Research Applications Laboratory, 2015. Verification: Weather Forecast Verification Utilities. R package version 1.41, <http://CRAN.R-project.org/package=verification>.
- NELSON, M., 2007. *Colorado Weather Almanac*. Boulder, CO: Big Earth Publishing.
- NEWCOTT, W.R. and MENZEL, P., 1993. Lightning, nature's high-voltage spectacle. *National Geographic*, **184**, pp. 83-83.

- ORVILLE, R.E., HUFFINES, G.R., BURROWS, W.R. and CUMMINS, K.L., 2011. The North American Lightning Detection Network (NALDN)—Analysis of Flash Data: 2001–09. *Monthly Weather Review*, **139**(5), pp. 1305-1322.
- PEPLER, R.A., 1988. *A review of static stability indices and related thermodynamic parameters*. Illinois State Water Survey Division, Climate and Meteorology Section. SWS Misc. Publ. 104
- PETER, L. and MOKHONOANA, F., 2010. Lightning detection improvement FALLS brought to Eskom's transmission line design and fault analysis, *Proceedings of the 21<sup>st</sup> International Lightning Detection Conference and 3<sup>rd</sup> Lightning Meteorology Conference*. Orlando, FL, 2010. Vaisala Inc., pp. 19-22.
- PETERSEN, W.A. and RUTLEDGE, S.A., 1998. On the relationship between cloud-to-ground lightning and convective rainfall. *Journal of Geophysical Research: Atmospheres*, **103**(D12), pp. 14025-14040.
- POSADA, D. and BUCKLEY, T.R., 2004. Model Selection and Model Averaging in Phylogenetics: Advantages of Akaike Information Criterion and Bayesian Approaches Over Likelihood Ratio Tests. *Systematic Biology*, **53**(5), pp. 793-808.
- PRICE, C., 1993. Global surface temperatures and the atmospheric electrical circuit. *Geophysical Research Letters*, **20**(13), pp. 1363-1366.
- PRICE, C., 2006. Global Thunderstorm Activity. In: FÜLLEKRUG, M., MAREEV, E.A. and RYCROFT, M.J., ed. *Sprites, Elves and Intense Lightning Discharges*. Vol 225. The Netherlands: Springer Science and Business Media, pp. 85-99.
- PRICE, C., 2008. Lightning Sensors for Observing, Tracking and Nowcasting Severe Weather. *Sensors*, **8**(1), pp. 157-170.
- PRICE, C.G., 2013. Lightning Applications in Weather and Climate Research. *Surveys in Geophysics*, **34**(6), pp. 755-767.
- PRICE, C. and ASFUR, M., 2006. Can lightning observations be used as an indicator of upper-tropospheric water vapor variability?. *Bulletin of the American Meteorological Society*, **87**(3), pp. 291-298.
- QIE, X., TOUMI, R. and ZHOU, Y., 2003. Lightning activity on the central Tibetan Plateau and its response to convective available potential energy. *Chinese Science Bulletin*, **48**(3), pp. 296-299.
- R DEVELOPMENT CORE TEAM, 2015. *R: A language and environment for statistical computing*. ISBN 3-900051-07-0. Vienna, Austria: R Foundation for Statistical Computing.
- RAJEEVAN, M., MADHULATHA, A., RAJASEKHAR, M., BHATE, J., KESARKAR, A. and RAO, B.V.A., 2012. Development of a perfect prognosis probabilistic model for prediction of lightning over south-east India. *Journal of Earth System Science*, **121**(2), pp. 355-371.
- RAKOV, V.A. and UMAN, M.A., 2003. *Lightning: Physics and Effects*. Cambridge, UK: Cambridge University Press.

- RASMUSSEN, E.N. and BLANCHARD, D.O., 1998. A Baseline Climatology of Sounding-Derived Supercell and Tornado Forecast Parameters. *Weather and Forecasting*, **13**(4), pp. 1148-1164.
- REAP, R.M., 1994. Analysis and Prediction of Lightning Strike Distributions Associated with Synoptic Map Types over Florida. *Monthly Weather Review*, **122**(8), pp. 1698-1715.
- REEVE, N. and TOUMI, R., 1999. Lightning activity as an indicator of climate change. *Quarterly Journal of the Royal Meteorological Society*, **125**(555), pp. 893-903.
- REYNOLDS, S.E., BROOK, M. and GOURLEY, M.F., 1957. THUNDERSTORM CHARGE SEPARATION. *Journal of Meteorology*, **14**(5), pp. 426-436.
- ROBIN, X., TURCK, N., HAINARD, A., TIBERTI, N., LISACEK, F., SANCHEZ, J. and MÜLLER, M., 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, **12**(1), pp. 1-8.
- RODGER, C.J. and RUSSEL, N.A., 2002. Lightning Flash Multiplicity Measurements by the U.S. National Lightning Detection Network, *Proceedings of the 27<sup>th</sup> general assembly of the International Union of Radio Science (URSI)*. Maastricht, The Netherlands, 2002, International Union of Radio Science.
- ROSE, L.S., 2008. *A Spatial Analysis of Lightning Strikes and Precipitation in the Greater Atlanta, Georgia (USA) Region*. MSc dissertation, Florida State University, Tallahassee.
- ROUAULT, M., ROY, S.S. and BALLING, R.C., 2013. The diurnal cycle of rainfall in South Africa in the austral summer. *International Journal of Climatology*. **33**(3), pp. 770-777.
- RUDLOSKY, S.D., 2014. Evaluating Ground-Based Lightning Detection Networks Using TRMM/LIS Observations, *Proceedings of the 23<sup>rd</sup> International Lightning Detection Conference and 5<sup>th</sup> International Lightning Meteorology Conference*. Tucson, AZ, 2014. Vaisala Inc.
- SANTOS, J.A., REIS, M.A., DE PABLO, F., RIVAS-SORIANO, L. and LEITE, S.M., 2013. Forcing factors of cloud-to-ground lightning over Iberia: regional-scale assessments. *Natural Hazards and Earth System Sciences*, **13**(7), pp. 1745-1758.
- SAS INSTITUTE INC, 2010. *SAS/STAT 9.22 User's Guide*. Cary, NC: SAS Institute Inc.
- SAUNDERS, C.P.R., 1993. A Review of Thunderstorm Electrification Processes. *Journal of Applied Meteorology*, **32**(4), pp. 642-655.
- SAUNDERS, C., 2008. Charge Separation Mechanisms in Clouds. *Space Sci.Rev.*, **137**, pp. 335-353.
- SAUNDERS, C.P.R., BAX-NORMAN, H., EMERSIC, C., AVILA, E.E. and CASTELLANO, N.E., 2006. Laboratory studies of the effect of cloud conditions on graupel/crystal charge transfer in thunderstorm electrification. *Quarterly Journal of the Royal Meteorological Society*, **132**(621), pp. 2653-2673.

- SCHULTZ, C.J., PETERSEN, W.A. and CAREY, L.D., 2011. Lightning and Severe Weather: A Comparison between Total and Cloud-to-Ground Lightning Trends. *Weather and forecasting*, **26**(5), pp.744-755.
- SCHULTZ, D.M., SCHUMACHER, P.N. and DOSWELL, C.A., 2000. The Intricacies of Instabilities. *Monthly Weather Review*, **128**(12), pp. 4143-4148.
- SCHULZE, G.C., 2007. Atmospheric observations and numerical weather prediction. *South African Journal of Science*, **103**(7-8), pp. 318-323.
- SHAFER, P.E., 2004. *Developing Statistical Guidance for Forecasting the Amount of Warm Season Afternoon and Evening Lightning in South Florida*. MSc thesis, Florida State University, Tallahassee. [viewed 25 April 2016]. Available from: [https://www.researchgate.net/profile/Phillip\\_Shafer/publication/254671165\\_Developing\\_Statistical\\_Guidance\\_for\\_Forecasting\\_the\\_Amount\\_of\\_Warm\\_Season\\_Afternoon\\_and\\_Evening\\_Lightning\\_in\\_South\\_Florida/links/560497ec08ae8e08c08aa859.pdf](https://www.researchgate.net/profile/Phillip_Shafer/publication/254671165_Developing_Statistical_Guidance_for_Forecasting_the_Amount_of_Warm_Season_Afternoon_and_Evening_Lightning_in_South_Florida/links/560497ec08ae8e08c08aa859.pdf).
- SHAFER, P.E. and FUELBERG, H.E., 2005. A Statistical Procedure to Forecast the Daily Amount of Warm Season Lightning in South Florida, *Proceedings of the Conference on Meteorological Applications of Lightning Data*, San Diego, CA, 2005. Amer. Meteor. Soc.
- SHAFER, P.E. and FUELBERG, H.E., 2006. A Statistical Procedure to Forecast Warm Season Lightning over Portions of the Florida Peninsula. *Weather and Forecasting*, **21**(5), pp. 851-868.
- SHAFER, P.E. and FUELBERG, H.E., 2008. A Perfect Prognosis Scheme for Forecasting Warm-Season Lightning over Florida. *Monthly Weather Review*, **136**(6), pp. 1817-1846.
- SHEN, J. and GAO, S., 2008. A Solution to Separation and Multicollinearity in Multiple Logistic Regression. *Journal of Data Science*, **6**(4), pp. 515-531.
- SIMPSON, G. and ROBINSON, G.D., 1941. The Distribution of Electricity in Thunderclouds, II. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, **177**(970), pp. 281-329.
- SIMPSON, G. and SCRASE, F.J., 1937. The Distribution of Electricity in Thunderclouds. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, **161**(906), pp. 309-352.
- SINGH, M.S. and O'GORMAN, P.A., 2015. Increases in moist-convective updraught velocities with warming in radiative-convective equilibrium. *Quarterly Journal of the Royal Meteorological Society*, **141**(692), pp. 2828-2838.
- SMITH, S.B., LADUE, J.G. and MACGORMAN, D.R., 2000. The Relationship between Cloud-to-Ground Lightning Polarity and Surface Equivalent Potential Temperature during Three Tornadic Outbreaks. *Monthly Weather Review*, **128**(9), pp. 3320-3328.
- SNIPES, M. and TAYLOR, D.C., 2014. Model selection and Akaike Information Criteria: An example from wine ratings and prices. *Wine Economics and Policy*, **3**(1), pp. 3-9.

- SOLOMON, R. and BAKER, M., 1994. Electrification of New Mexico Thunderstorms. *Monthly Weather Review*, **122**(8), pp. 1878-1886.
- SOUL, K.M., ARCHIBALD, E.J., HARDAKER, P.J. and HOUNSELL, A., 2002. Using the GANDOLF system as a tool to aid the forecasting of lightning strikes. *Meteorological Applications*, **9**(2), pp. 229-238.
- STANO, G.T., FUELBERG, H.E. and ROEDER, W.P., 2010. Developing empirical lightning cessation forecast guidance for the Cape Canaveral Air Force Station and Kennedy Space Center. *Journal of Geophysical Research: Atmospheres*, **115**(D9).
- STOLZENBURG, M., RUST, W.D. and MARSHALL, T.C., 1998a. Electrical structure in thunderstorm convective regions: 2. Isolated storms. *Journal of Geophysical Research: Atmospheres*, **103**(D12), pp. 14079-14096.
- STOLZENBURG, M., RUST, W.D. and MARSHALL, T.C., 1998b. Electrical structure in thunderstorm convective regions: 3. Synthesis. *Journal of Geophysical Research: Atmospheres*, **103**(D12), pp. 14097-14108.
- STULL, R.B., 2012. *An Introduction to Boundary Layer Meteorology*. Vol 13. Netherlands: Springer Science and Business Media.
- TAKAHASHI, T., 1978. Riming Electrification as a Charge Generation Mechanism in Thunderstorms. *Journal of the Atmospheric Sciences*, **35**(8), pp. 1536-1548.
- TAKAHASHI, T., 1984. Thunderstorm Electrification—A Numerical Study. *Journal of the Atmospheric Sciences*, **41**(17), pp. 2541-2558.
- TALJAARD, J.J., 1994. *Atmospheric Circulation Systems, Synoptic Climatology and Weather Phenomena of South Africa. Part 1. Controls of the weather and climate of South Africa*. South African Weather Bureau Technical Paper No 27, 45.
- TALJAARD, J.J., 1995. *Atmospheric Circulation Systems, Synoptic Climatology and Weather Phenomena of South Africa. Part 3. Synoptic Climatology in South Africa in January and July*. South African Weather Bureau Technical Paper No 29
- TALJAARD, J.J., 1996. *Atmospheric Circulation Systems, Synoptic Climatology and Weather Phenomena of South Africa. Part 6. Rainfall in South Africa*. South African Weather Bureau Technical Paper No. 32, 53.
- TARTAGLIONE, N., 2010. Relationship between Precipitation Forecast Errors and Skill Scores of Dichotomous Forecasts. *Weather and Forecasting*, **25**(1), pp. 355-365.
- THOMPSON, K.B., BATEMAN, M.G. and CAREY, L.D., 2014. A Comparison of Two Ground-Based Lightning Detection Networks against the Satellite-Based Lightning Imaging Sensor (LIS). *Journal of Atmospheric and Oceanic Technology*, **31**(10), pp. 2191-2205.
- TRENGROVE, E. and JANDRELL, I.R., 2011. Strategies for Understanding Lightning Myths and Beliefs. *International Journal of Research and Reviews in Applied Sciences*, **7**(3), pp. 287-294.
- TUDURÍ, E. and RAMIS, C., 1997. The Environments of Significant Convective Events in the Western Mediterranean. *Weather and Forecasting*, **12**(2), pp. 294-306.

- TYSON, P.D., 1986. *Climatic change and variability in southern Africa*. Cape Town: Oxford University Press.
- UMAN, M.A., 1987. *The Lightning Discharge*. Orlando, FL: Academic Press Inc.
- UMAN, M.A., 2001. *The Lightning Discharge*. 2<sup>nd</sup> ed. Mineola, New York: Dover Publications.
- UMAN, M.A., 2012. *The Lightning Discharge*. 3<sup>rd</sup> ed. Mineola, New York: Dover Publications.
- VAISALA, 2004. *CP Series: CP7000 CP8000 User's Guide*. Helsinki, Finland: Vaisala Oyj, M210557EN-A.
- VAISALA, 2008. *Vaisala Thunderstorm Lightning Network Performance Evaluation Program (NPEP)* [online]. Vaisala Inc. [viewed 12 April 2016]. Available from: <http://www.vaisala.com/Vaisala%20Documents/Brochures%20and%20Datasheets/NPEP-Datasheet-B210775EN-A.pdf>
- VANWALLEGHEM, T., VAN DEN EECKHAUT, M., POESEN, J., GOVERS, G. and DECKERS, J., 2008. Spatial analysis of factors controlling the presence of closed depressions and gullies under forest: Application of rare event logistic regression. *Geomorphology*, **95**(3), pp. 504-517.
- VAN DE GROENENDAAL, H., 2007. SA Weather Service introduces real-time display and warning system. *VECTOR*, **5**, pp. 68-69.
- VIRTS, K.S. and HOUZE JR, R.A., 2015. Variation of Lightning and Convective Rain Fraction in Mesoscale Convective Systems of the MJO. *Journal of the Atmospheric Sciences*, **72**(5), pp. 1932-1944.
- VIRTS, K.S., WALLACE, J.M., HUTCHINS, M.L. and HOLZWORTH, R.H., 2013. Highlights of a New Ground-Based, Hourly Global Lightning Climatology. *Bulletin of the American Meteorological Society*, **94**(9), pp. 1381-1391.
- WEISHEIMER, A. and PALMER, T.N., 2014. On the reliability of seasonal climate forecasts. *Journal of The Royal Society Interface*, **11**(96), p.20131162.
- WEUSTHOFF, T., AMENT, F., ARPAGAUS, M. and ROTACH, M.W., 2010. Assessing the Benefits of Convection-Permitting Models by Neighborhood Verification: Examples from MAP D-PHASE. *Monthly Weather Review*, **138**(9), pp. 3418-3433.
- WHITAKER, J.S., COMPO, G.P., WEI, X. and HAMILL, T.M., 2004. Reanalysis without Radiosondes Using Ensemble Data Assimilation. *Monthly Weather Review*, **132**(5), pp. 1190-1200.
- WILLIAMS, E.R., 1992. The Schumann Resonance: A Global Tropical Thermometer. *Science*, **256**(5060), pp. 1184-1187.
- WILLIAMS, E.R., 1994. Global Circuit Response to Seasonal Variations in Global Surface Air Temperature. *Monthly Weather Review*, **122**(8), pp. 1917-1929.
- WILLIAMS, E.R., 2009. The global electrical circuit: A review. *Atmospheric Research*, **91**(2-4), pp. 140-152.



- WILSON, C.T.R., 1916. On Some Determinations of the Sign and Magnitude of Electric Discharges in Lightning Flashes. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, **92**(644), pp. 555-574.
- WOODARD, C., CAREY, L., PETERSEN, W. and ROEDER, W.P., 2012. Operational Utility of Dual-Polarization Variables in Lightning Initiation Forecasting. *Electronic J. Operational Meteor.*, **13**(6), pp. 79-102.
- WWRP/WGNE Joint Working Group on Verification, 2015. *Forecast Verification—Issues, Methods and FAQ* [online]. WWRP/WGNE Joint Working Group on Verification. [viewed 20 August 2015]. Available from: <http://www.cawcr.gov.au/projects/verification/>
- XIONG, Y., QIE, X., ZHOU, Y., YUAN, T. and ZHANG, T., 2006. Regional Responses of Lightning Activities to Relative Humidity of the Surface. *Chinese Journal of Geophysics*, **49**(2), pp. 311-318.
- YAP, B.W., RANI, K.A., RAHMAN, H.A.A., FONG, S., KHAIRUDIN, Z. and ABDULLAH, N.N., 2014. An Application of Oversampling, Undersampling, Bagging and Boosting in Handling Imbalanced Datasets. *Proceedings of the 1<sup>st</sup> International Conference on Advanced Data and Information Engineering (DaEng-2013)*. 285, Kuala Lumpur, Malaysia, 2014, Springer Singapore, pp. 13-22.
- ZEPKA, G.S., PINTO JR., O. and SARAIVA, A.C.V., 2014. Lightning forecasting in southeastern Brazil using the WRF model. *Atmospheric Research*, **135–136**, pp. 344-362.
- ZOU, K.H., O'MALLEY, A.J. and MAURI, L., 2007. Receiver-Operating Characteristic Analysis for Evaluating Diagnostic Tests and Predictive Models. *Circulation*, **115**(5), pp. 654-657.