



Ministério da
**Ciência, Tecnologia
e Inovação**



sid.inpe.br/mtc-m21b/2014/03.14.20.54-TDI

RAINFALL ESTIMATION FROM CLOUD-TO-GROUND LIGHTNING DATA

João Victor Cal Garcia

Doctorate Thesis Course Graduate
in Applied Computing, guided by
Dr. Stephan Stephany, approved in
March 31, 2014.

URL of the original document:

<<http://urlib.net/8JMKD3MGP5W34M/3FTKT5L>>

INPE
São José dos Campos
2014

PUBLISHED BY:

Instituto Nacional de Pesquisas Espaciais - INPE

Gabinete do Diretor (GB)

Serviço de Informação e Documentação (SID)

Caixa Postal 515 - CEP 12.245-970

São José dos Campos - SP - Brasil

Tel.:(012) 3208-6923/6921

Fax: (012) 3208-6919

E-mail: pubtc@sid.inpe.br

BOARD OF PUBLISHING AND PRESERVATION OF INPE INTELLECTUAL PRODUCTION (RE/DIR-204):**Chairperson:**

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Members:

Dr. Antonio Fernando Bertachini de Almeida Prado - Coordenação Engenharia e Tecnologia Espacial (ETE)

Dr^a Inez Staciarini Batista - Coordenação Ciências Espaciais e Atmosféricas (CEA)

Dr. Gerald Jean Francis Banon - Coordenação Observação da Terra (OBT)

Dr. Germano de Souza Kienbaum - Centro de Tecnologias Especiais (CTE)

Dr. Manoel Alonso Gan - Centro de Previsão de Tempo e Estudos Climáticos (CPT)

Dr^a Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação

Dr. Plínio Carlos Alvalá - Centro de Ciência do Sistema Terrestre (CST)

DIGITAL LIBRARY:

Dr. Gerald Jean Francis Banon - Coordenação de Observação da Terra (OBT)

DOCUMENT REVIEW:

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SID)

ELECTRONIC EDITING:

Maria Tereza Smith de Brito - Serviço de Informação e Documentação (SID)

Luciana Manacero - Serviço de Informação e Documentação (SID)



Ministério da
**Ciência, Tecnologia
e Inovação**



sid.inpe.br/mtc-m21b/2014/03.14.20.54-TDI

RAINFALL ESTIMATION FROM CLOUD-TO-GROUND LIGHTNING DATA

João Victor Cal Garcia

Doctorate Thesis Course Graduate
in Applied Computing, guided by
Dr. Stephan Stephany, approved in
March 31, 2014.

URL of the original document:

<<http://urlib.net/8JMKD3MGP5W34M/3FTKT5L>>

INPE
São José dos Campos
2014

Cataloging in Publication Data

Garcia, João Victor Cal.

G165r Rainfall Estimation from cloud-to-ground lightning data / João Victor Cal Garcia. – São José dos Campos : INPE, 2014.

xxxii + 139 p. ; (sid.inpe.br/mtc-m21b/2014/03.14.20.54-TDI)

Thesis (Doctorate in Applied Computing) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2014.

Guiding : Dr. Stephan Stephany.

1. lightning. 2. weather radar. 3. rainfall estimation. 4. Python.
I.Título.

CDU 551.578

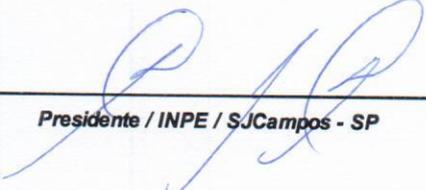


Esta obra foi licenciada sob uma Licença [Creative Commons Atribuição-NãoComercial 3.0 Não Adaptada](https://creativecommons.org/licenses/by-nc/3.0/).

This work is licensed under a [Creative Commons Attribution-NonCommercial 3.0 Unported License](https://creativecommons.org/licenses/by-nc/3.0/).

Aprovado (a) pela Banca Examinadora
em cumprimento ao requisito exigido para
obtenção do Título de **Doutor(a)** em
Computação Aplicada

Dra. Sandra Aparecida Sandri



Presidente / INPE / SJC Campos - SP

Dr. Stephan Stephany



Orientador(a) / INPE / SJC Campos - SP

Dr. Walter Abrahão dos Santos



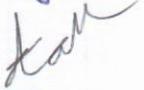
Membro da Banca / INPE / São José dos Campos - SP

Dr. Nelson Jesus Ferreira



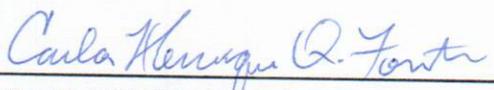
Membro da Banca / INPE / SJC Campos - SP

Dr. Roberto Vicente Calheiros



Convidado(a) / UNESP / Bauru - SP

Dr. Carlos Henrique Quartucci Forster



Convidado(a) / ITA / São José dos Campos - SP

Este trabalho foi aprovado por:

() maioria simples

unanimidade

Aluno (a): **João Victor Cal Garcia**

São José dos Campos, 31 de Março de 2014

“We plan for days and days and when the time comes, we proceed to improvise”.

PIETRO MAXIMOFF
in “Amazing X-Men #2”, 1995

*A meus pais Antônio e Cylla, e à minha irmã
Maria Angélica, por seu suporte e compreensão*

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Stephan Stephany, for without his guidance and persistence, this thesis would not have been possible.

I would like to thank the co-authors and contributors of this work for their support and insights.

I would like to thank the CEMADEN for their interest in this work.

I would like to thank CEMADEN, CPTEC and IPMet/UNESP for the data used in this work.

I would like to thank the CNPq for the financial support as a PhD scholarship (process 140983/2010-4).

ABSTRACT

The correlation between convective activity and the occurrence of lightning has been exploited in several works. The monitoring and tracking of convective activity is conventionally achieved by the use of weather radars, but many countries, like Brazil, lack an extensive coverage of such radars. On the other hand, Brazil has a network of lightning detectors that covers most of its territory and such detectors are quite less expensive than weather radars. Assuming that convective activity and lightning are correlated to some degree, lightning data is currently employed to monitor and track convective activity by means of the EDDA software in the Center for Natural Disasters Monitoring and Alert in Brazil. Besides this correlation, some works also address a quantitative relationship between the number of lightning and the precipitated mass. While some approaches were proposed to estimate rainfall using lightning data, they seem to have poor accuracy and to be limited to the same region and conditions where their specific parameters were obtained, limiting the application of lightning-estimated rainfall. This thesis proposes a new and more reliable approach to estimate the accumulated rainfall mass and distribution from lightning occurred in a preceding interval of time for a defined grid. The estimation of the rainfall mass is accomplished by a function inferred using grid points covered by weather radar data. This function is then mapped to the remaining grid points based on Brazilian rainfall statistics of recent years, making it applicable outside the range of weather radars. Rainfall spatial distribution is also estimated based on the density of occurrence of lightning. Rainfall estimation in this thesis is implemented by the Edda-chuva software that is being evaluated in the same center. Besides convective rainfall amount and distribution, the software presented in this thesis also allows to estimate the amount and distribution of the total rainfall, including thus the stratiform rainfall. Test results show an improvement of the accumulated rainfall monitoring over most of the Brazilian territory covered by the network of electrical discharge detectors.

ESTIMAÇÃO DE PRECIPITAÇÃO ATRAVÉS DE DADOS DE DESCARGAS ELÉTRICAS ATMOSFÉRICAS NUVEM-SOLO

RESUMO

A correlação entre a atividade convectiva e a ocorrência de descargas elétricas atmosféricas foi explorada em vários trabalhos. O monitoramento e rastreamento de atividade convectiva é convencionalmente feita por meio de radares meteorológicos, mas muitos países, como o Brasil, não têm uma extensa cobertura de tais radares. Por outro lado, o Brasil tem uma rede de detectores de descargas elétricas atmosféricas que cobre a maior parte de seu território e esses detectores são muito menos caros do que radares meteorológicos. Supondo-se que a atividade convectiva e descargas elétricas atmosféricas estejam correlacionadas em algum grau, dados de descarga elétrica são atualmente empregados para monitorar e rastrear a atividade convectiva por meio do software EDDA no Centro Nacional de Monitoramento e Alertas de Desastres Naturais (Cemaden), no Brasil. Além dessa correlação, alguns trabalhos também abordam uma relação quantitativa entre o número de descargas elétricas atmosféricas e a massa precipitada. Alguns trabalhos propuseram estimar precipitação a partir de dados de descargas, mas seus resultados foram imprecisos e limitados a mesma região e condições na qual foram obtidos, limitando o uso de precipitação estimada por descargas. Esta tese propõe uma nova metodologia para estimar a precipitação acumulada a partir do número de descargas ocorridas em um intervalo de tempo anterior para uma grade definida. Esta estimativa é realizada por uma função inferida usando pontos de grade cobertos pelos dados de radar meteorológico e foi mapeada para os pontos de grade restantes com base em estatísticas de precipitação no Brasil para anos recentes, sendo utilizável fora do alcance de radares meteorológicos. A distribuição espacial da precipitação também é estimada com base na densidade de ocorrência das descargas. A estimativa de precipitação é implementada neste tese pelo software EDDA-chuva que está sendo avaliado no mesmo centro. Além da quantidade e distribuição da precipitação convectiva, o software permite estimar a quantidade e distribuição da precipitação total, incluindo portanto a precipitação estratiforme. Os resultados do teste mostram uma melhoria do monitoramento precipitação acumulada sobre a maior parte do território brasileiro coberto pela rede de detectores de descargas elétricas atmosféricas.

LIST OF FIGURES

	<u>Page</u>
2.1 Coverages of the Brazilian territory for the different meteorological data employed in this work: (i) quarterly rainfall averages (entire area), (ii) RINDAT lightning detection network (blue rectangle), and (iii) weather radars (red circles).	6
2.2 Schematic view of a thunderstorm cloud, with the convective (right) and stratiform (left) rainfall.	7
2.3 Schematic view of radar pulses in a vertical plane for different elevation angles various, with the gray band highlighting the 3 km altitude CAPPI.	9
2.4 Altitude distortion due to the curvature of the Earth.	10
2.5 Coverage of the 150 km range of the weather radars used in this work.	12
2.6 Location of the current 25 detectors of RINDAT (two sensors are superposed, in central Paraná). The blue dashed line represents the range of the RINDAT with probability of detection above 80% (RINDAT, 2011).	16
2.7 Average yearly rainfall totals (mm) per season for Brazil during 1977–2006.	18
2.8 Spatial distribution of rain gauges of different sources/agencies over South America in 2010.	19
3.1 The intensity-dependent convective radius as a function of the average background reflectivity Z_{bg}	26
4.1 Temporal series of accumulated 30 min values of the number of CG strokes and precipitated mass (rainfall) for the year of 2009 and for the square area near the city of Bauru (curves smoothed by 1D Gaussian filter).	29
4.2 Temporal series of accumulated 30 min values of the number of CG strokes and precipitated mass (rainfall) for the month of September of 2009 and for the square area near the city of Bauru (curves smoothed by 1D Gaussian filter).	30
4.3 Temporal series of accumulated 30 min values of the number of CG strokes and precipitated mass (rainfall) for some days of September 2009 and for the square area near the city of Bauru during a particular thunderstorm (curves smoothed by 1D Gaussian filter).	30
4.4 Location of the squared area of 200 km side centered at the Bauru radar location (the circle shows the 150 km radar range).	35
4.5 Sliding time-window with size $\Delta t = 4$ units and advance $\delta t = 1$ unit.	36

4.6	The square area of side 200 km divided into (a) several Q_j sampling squares with no displacement, (b) with a latitudinal shift δx , (c) with a longitudinal shift δx , and (d) with both latitudinal and longitudinal shifts δx	37
4.7	The scatter-plot of (n_{ij}, r_{ij}) tuples (in black) and of the rainfall-averaged (N, r_N) tuples (in red).	39
4.8	True target spatial adjust factors (black dots) compared to the proposed adjust factors given by the square of the ratio between areas (dashed line).	42
4.9	The functions derived from the tuples are shown in solid lines WRLR _{20 km} (black) and WRLR _{50 km} (blue), while the approximated function $[X_{20 \text{ km}} \cdot \text{WRLR}_{50 \text{ km}}]$ is shown by the red dashed line.	43
4.10	True target temporal adjust factors (black dots) compared to the proposed adjust factors given by the ratio between durations (dashed line).	44
4.11	The functions derived from the tuples are shown in solid lines WRLR _{15 min} (black) and WRLR _{30 min} (blue), while the approximated function $[T_{15 \text{ min}} \cdot \text{WRLR}_{30 \text{ min}}]$ is shown by the red dashed line.	45
4.12	Values of the elements of the K remapping matrix for the four seasons and considered grid.	48
4.13	Schematic view of the spatial distribution of rainfall for Tapia's Model (left) and EDDA-chuva (right). The rainfall is showed in low intensity (blue) and high intensity (red), for the lightning occurrences (yellow).	51
5.1	Scatter plot of the convective rainfall mass and the number of CG flashes. Each point corresponds to the individual RLR of given thunderstorm, while the dashed line corresponds to overall RLR.	55
5.2	Scatter plot of the tuples of convective rainfall mass and number of CG strokes.	56
5.3	Histogram for the number of tuples with a given number N of CG strokes.	57
5.4	Hexagonal binning for the tuples of convective rainfall and number of CG strokes. Each hexagon shows the averaged number of windows for each tuple.	58
5.5	Scatter plots of the averaged tuples (average rainfall for each number of strokes N) for each season. The WRLR functions obtained for all tuples of the dataset are plotted in black for each season.	59
5.6	The WRLR function for Summer, Autumn, Winter and Spring.	60
5.7	Summer WRLR functions obtained for different Δx (10, 20, 25, 40, and 100 km, in black) compared to the standard WRLR function (in blue) mapped by the corresponding spatial adjust factors (red dashed line).	62

5.8	Summer WRLR functions obtained for different Δt (7.5, 15 and 60 min, in black) compared to the standard WRLR function (in blue) mapped by the corresponding temporal adjust factors (red dashed line).	63
5.9	The months of January and February for the Bauru radar during the Summer of 2013. The top image shows the daily accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), of the RLR constant (in dashed green) and of the WRLR function (in dashed red) for the convective rainfall mass.	65
5.10	BRU-1 thunderstorm occurred in 1st January 2013. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), of the RLR constant (in dashed green) and of the WRLR function (in dashed red) for the convective rainfall mass.	67
5.11	BRU-2 thunderstorm occurred in 15th January 2013. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), of the RLR constant (in dashed green) and of the WRLR function (in dashed red) for the convective rainfall mass.	68
5.12	BRU-3 thunderstorm occurred in 25th January 2013. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), of the RLR constant (in dashed green) and of the WRLR function (in dashed red) for the convective rainfall mass.	69
5.13	BRU-4 thunderstorm occurred in 23rd February 2013. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), of the RLR constant (in dashed green) and of the WRLR function (in dashed red) for the convective rainfall mass.	70
5.14	BRU-5 thunderstorm occurred in 19th January 2013. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), of the RLR constant (in dashed green) and of the WRLR function (in dashed red) for the convective rainfall mass.	71

5.15	BRU-6 thunderstorm occurred in 16th January 2013. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), of the RLR constant (in dashed green) and of the WRLR function (in dashed red) for the convective rainfall mass.	72
5.16	BRU-7 thunderstorm occurred in 26th February 2013. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), of the RLR constant (in dashed green) and of the WRLR function (in dashed red) for the convective rainfall mass.	73
5.17	Rainfall for the Summer of 2009 in Presidente Prudente. The top image shows the daily accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.	76
5.18	Rainfall for the Autumn of 2009 in Presidente Prudente. The top image shows the daily accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.	77
5.19	Rainfall for the Winter of 2009 in Presidente Prudente. The top image shows the daily accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.	78
5.20	Rainfall for the Spring of 2009 in Presidente Prudente. The top image shows the daily accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.	79
5.21	Rainfall for the Summer of 2010 in Presidente Prudente. The top image shows the daily accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.	80

5.22	Rainfall for the Autumn of 2010 in Presidente Prudente. The top image shows the daily accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.	81
5.23	Rainfall for the Winter of 2010 in Presidente Prudente. The top image shows the daily accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.	82
5.24	PPR-1S thunderstorm occurred in 11th January 2009. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.	84
5.25	PPR-2S thunderstorm occurred in 13th March 2010. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.	85
5.26	PPR-3S thunderstorm occurred in 6th January 2010. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.	86
5.27	PPR-4A thunderstorm occurred in 22nd April 2010. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.	88
5.28	PPR-5A thunderstorm occurred in 1st April 2010. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.	89

5.29	PPR-6A thunderstorm occurred in 14th May 2009. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.	90
5.30	PPR-7W thunderstorm occurred in 17th August 2009. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.	91
5.31	PPR-8W thunderstorm occurred in 22nd July 2009. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.	92
5.32	PPR-9W thunderstorm occurred in 27th September 2010. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.	93
5.33	PPR-10R thunderstorm occurred in 20th November 2009. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.	94
5.34	PPR-11R thunderstorm occurred in 11th December 2009. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.	95
5.35	PPR-12R thunderstorm occurred in 15th October 2009. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.	96

5.36	Rainfall for the months of January to March of 2013 in São Roque. The top image shows the daily accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.	98
5.37	RQ-1 thunderstorm occurred in 25th January 2013. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.	99
5.38	RQ-2 thunderstorm occurred in 15th February 2013. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.	100
5.39	RQ-3 thunderstorm occurred in 26th February 2013. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.	101
5.40	Rainfall for the months of January to March of 2013 in Pico do Couto. The top image shows the daily accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.	103
5.41	PI-1 thunderstorm starting in 14th January 2013. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.	104
5.42	PI-2 thunderstorm starting in 3rd March 2013. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR adjusted function (in dashed red) for the convective rainfall mass.	105

5.43	PI-3 thunderstorm starting in 26th February 2013. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR adjusted function (in dashed red) for the convective rainfall mass.	106
5.44	Rainfall spatial distributions accumulated for the first hour for thunderstorm BRU-1 given by (i) weather radar for convective rainfall (“filtered radar”), (ii) EDDA-chuva for convective rainfall (“EDDA-chuva (C)”), (iii) Tapia’s model for convective rainfall, (iv) weather radar for total rainfall (“unfiltered radar”), and (v) EDDA-chuva for total rainfall (“EDDA-chuva (T)”).	109
5.45	Rainfall spatial distributions accumulated for the first hour for thunderstorm PPR-1S given by (i) weather radar for convective rainfall (“filtered radar”), (ii) EDDA-chuva for convective rainfall (“EDDA-chuva (C)”), (iii) weather radar for total rainfall (“unfiltered radar”), and (iv) EDDA-chuva for total rainfall (“EDDA-chuva (T)”).	110
5.46	Rainfall spatial distributions accumulated for the first hour for thunderstorm RQ-1 given by (i) weather radar for convective rainfall (“filtered radar”), (ii) EDDA-chuva for convective rainfall (“EDDA-chuva (C)”), (iii) weather radar for total rainfall (“unfiltered radar”), and (iv) EDDA-chuva for total rainfall (“EDDA-chuva (T)”).	111
5.47	Rainfall spatial distributions accumulated for the first hour for thunderstorm PI-1 given by (i) weather radar for convective rainfall (“filtered radar”), (ii) EDDA-chuva for convective rainfall (“EDDA-chuva (C)”), (iii) weather radar for total rainfall (“unfiltered radar”), and (iv) EDDA-chuva for total rainfall (“EDDA-chuva (T)”).	112
5.48	Rainfall spatial distributions accumulated for the first six hours for thunderstorm BRU-1 given by (i) weather radar for convective rainfall (“filtered radar”), (ii) EDDA-chuva for convective rainfall (“EDDA-chuva (C)”), (iii) Tapia’s model for convective rainfall, (iv) weather radar for total rainfall (“unfiltered radar”), and (v) EDDA-chuva for total rainfall (“EDDA-chuva (T)”).	113
5.49	Rainfall spatial distributions accumulated for the first six hours for thunderstorm PPR-1S given by (i) weather radar for convective rainfall (“filtered radar”), (ii) EDDA-chuva for convective rainfall (“EDDA-chuva (C)”), (iii) weather radar for total rainfall (“unfiltered radar”), and (iv) EDDA-chuva for total rainfall (“EDDA-chuva (T)”).	114

5.50	Rainfall spatial distributions accumulated for the first six hours for thunderstorm RQ-1 given by (i) weather radar for convective rainfall (“filtered radar”), (ii) EDDA-chuva for convective rainfall (“EDDA-chuva (C)”), (iii) weather radar for total rainfall (“unfiltered radar”), and (iv) EDDA-chuva for total rainfall (“EDDA-chuva (T)”).	115
5.51	Rainfall spatial distributions accumulated for the first six hours for thunderstorm PI-1 given by (i) weather radar for convective rainfall (“filtered radar”), (ii) EDDA-chuva for convective rainfall (“EDDA-chuva (C)”), (iii) weather radar for total rainfall (“unfiltered radar”), and (iv) EDDA-chuva for total rainfall (“EDDA-chuva (T)”).	116
5.52	Daily rainfall spatial distributions accumulated for thunderstorm BRU-1 given by (i) weather radar for convective rainfall (“filtered radar”), (ii) EDDA-chuva for convective rainfall (“EDDA-chuva (C)”), (iii) Tapia’s model for convective rainfall, (iv) weather radar for total rainfall (“unfiltered radar”), (v) EDDA-chuva for total rainfall (“EDDA-chuva (T)”), and (vi) MERGE data.	118
5.53	Daily rainfall spatial distributions accumulated for thunderstorm PPR-1S given by (i) weather radar for convective rainfall (“filtered radar”), (ii) EDDA-chuva for convective rainfall (“EDDA-chuva (C)”), (iii) weather radar for total rainfall (“unfiltered radar”), (iv) EDDA-chuva for total rainfall (“EDDA-chuva (T)”), and (v) MERGE data.	119
5.54	Daily rainfall spatial distributions accumulated for thunderstorm RQ-1 given by (i) weather radar for convective rainfall (“filtered radar”), (ii) EDDA-chuva for convective rainfall (“EDDA-chuva (C)”), (iii) weather radar for total rainfall (“unfiltered radar”), (iv) EDDA-chuva for total rainfall (“EDDA-chuva (T)”), and (v) MERGE data.	120
5.55	Daily rainfall spatial distributions accumulated for thunderstorm PI-1 given by (i) weather radar for convective rainfall (“filtered radar”), (ii) EDDA-chuva for convective rainfall (“EDDA-chuva (C)”), (iii) weather radar for total rainfall (“unfiltered radar”), (iv) EDDA-chuva for total rainfall (“EDDA-chuva (T)”), and (v) MERGE data.	121
5.56	Daily rainfall spatial distribution accumulated for January 1st, 2013 as given by the EDDA-chuva software (left) and from MERGE data (right). Thunderstorm BRU-1 occurred during this day and inside the depicted area.	122

5.57	Daily rainfall spatial distribution accumulated for January 11th, 2009 as given by the EDDA-chuva software (left) and from MERGE data (right). Thunderstorm PPR-1S occurred during this day and inside the depicted area.	123
5.58	Daily rainfall spatial distribution accumulated for January 25th, 2013 as given by the EDDA-chuva software (left) and from MERGE data (right). Thunderstorm RQ-1 occurred during this day and inside the depicted area.	123
5.59	Daily rainfall spatial distribution accumulated for January 14th, 2013 as given by the EDDA-chuva software (left) and from MERGE data (right). Thunderstorm PI-1 occurred during this day and inside the depicted area.	124

LIST OF TABLES

	<u>Page</u>
2.1 Location of the weather radars used in this work.	11
2.2 Months/years of weather radar data in ASCII files for the radars of Bauru and Presidente Prudente.	13
2.3 Description of the columns of the Universal ASCII Lightning Format (UALF).	15
2.4 Months/years of RINDAT data employed in this work.	15
4.1 Influence of the parameters Δx_G and Δt_G in the resulting thunderstorms identified by Algorithm 4.1	34
5.1 Real constants a , b and c that describe the WRLR power function (Equation 4.2) for each season.	60
5.2 Convective rainfall mass in 10^9 kg estimated by weather radar, by the WRLR function and by the RLR constant for the 7 selected thunderstorms. The corresponding number of CG lightning strokes and flashes, the duration and the reference to the specific figures also appear.	65
5.3 Percentage error (<i>error</i>) and correlation (<i>corr</i>) of the WRLR and RLR estimatives for the 7 selected thunderstorms.	66
5.4 Estimation performance of the mapped WRLR function for the considered seasons/quarters for the area under the coverage of the Presidente Prudente radar. Convective rainfall mass is given in 10^9 kg and the percentage error is taken in relation to the radar estimation.	75
5.5 Comparison of the estimations for the convective rainfall mass given by the Presidente Prudente weather radar and by the mapped WRLR function for the 12 selected thunderstorms. Percentage error is calculated in relation to the radar estimation, while the correlation considers 30-min accumulated values for each thunderstorm.	83
5.6 Comparison of the estimations for the convective rainfall mass given by the São Roque weather radar and by the mapped WRLR function for the three selected thunderstorms. Percentage <i>error</i> is calculated in relation to the radar estimation, while the correlation (<i>corr</i>) considers 30 min accumulated values for each thunderstorm.	98

5.7	Comparison of the estimations for the convective rainfall mass given by the Pico do Couto weather radar and by the mapped WRLR function for the three selected thunderstorms. Percentage <i>error</i> is calculated in relation to the radar estimation, while the correlation (<i>corr</i>) considers 30 min accumulated values for each thunderstorm.	103
5.8	Metadata of the four thunderstorms employed in the rainfall spatial distribution tests: starting time and extension of time for the cumulative rainfall images.	108
5.9	Boundaries of the four thunderstorms employed in the rainfall spatial distribution tests.	108

LIST OF ABBREVIATIONS

BRASILDAT	–	Brazilian Lightning Detection Network
BRU	–	Weather Radar of Bauru
CAPPI	–	Constant Altitude Plan Position Indicator
CC	–	Cloud-to-Cloud
CEMADEN	–	Center for Natural Disasters Monitoring and Alert
CEMIG	–	Energy Company of Minas Gerais
CG	–	Cloud-to-Ground
CPRM	–	Brazilian Geological Service
EDDA	–	Estimator of Density of Atmospheric Electrical Discharges
FURNAS	–	Furnas Electric Centers
GIS	–	Geographic Information Systems
GIS	–	Geographical Information Services
GOES	–	Geostationary Operational Environmental Satellite
GPS	–	Global Position System
IC	–	Intracloud
IPMET	–	Institute for Meteorological Research
INPE	–	National Institute For Space Research
LPATS	–	Lightning Position and Tracking System
LtraX	–	Real-time Lightning Tracking software
MOS	–	Model Output Statistics
NOAA	–	National Oceanic Atmospheric Administration
PI	–	Weather Radar of Pico do Couto
PPI	–	Plan Position Indicator
PPR	–	Weather Radar of Presidente Prudente
RINDAT	–	Brazilian Integrated Lightning Detection Network
RLR	–	Rainfall-Lightning Ratio
RQ	–	Weather Radar of São Roque
SALVAR	–	Platform for Monitoring and Emission of Alerts of Natural Disasters
SIMEPAR	–	Meteorological System of Paraná
TRMM	–	Tropical Rainfall Measuring Mission
UALF	–	Universal ASCII Lightning Format
UNESP	–	São Paulo State University
WRLR	–	Windowed Rainfall-to-Lightning function

LIST OF SYMBOLS

H	– radar altitude correction
a_e	– Earth radius at the radar location
h_a	– target height relative to the radar
k_e	– refraction coefficient of the air
d	– radar distance to the target
Θ_e	– radar elevation angle
$R(t, x)$	– Rainfall rate at time t and position x for Tapia’s method
t	– instant in time
x	– position in space
C	– Unit conversion factor for Tapia’s method
N_f	– Counter for the number of flashes for Tapia’s method
RLR	– Constant rainfall-lightning ratio
T_i	– Time of occurrence of the i -th flash for Tapia’s method
X_i	– Location of occurrence of i -th flash for Tapia’s method
Δt_G	– maximum allowed difference in the time of occurrence of two flashes
Δx_G	– maximum allowed Euclidean distance of two flashes
$G(E, V)$	– graph of related lightning occurrences
F	– graph of lightning clusters
a	– real coefficient of the WRLR power function
b	– real exponent of the WRLR power function
c	– independent term of the WRLR power function
$WRLR(N)$	– Windowed Rainfall-to-Lightning function
N	– number of lightning strokes occurrences
n_{ij}	– number of lightning occurrences for the (n_{ij}, r_{ij}) tuple
r_{ij}	– convective rainfall mass for the (n_{ij}, r_{ij}) tuple
r_N	– convective rainfall mass for the (N, r_N) tuple

CONTENTS

	<u>Page</u>
1 INTRODUCTION	1
2 METEOROLOGICAL DATA	5
2.1 Rainfall	6
2.2 Lightning	7
2.3 Weather radar data	8
2.3.1 Weather radar range limitation	9
2.3.2 Weather radar metadata	11
2.4 Lightning data	13
2.4.1 The RINDAT lightning detection network	14
2.5 Isohyetal map of quarterly rainfall for Brazil	17
2.6 MERGE precipitation data	18
3 RELATED WORK	21
3.1 Tapia's model	21
3.2 The EDDA software	22
3.3 The Steiner criteria	25
4 PROPOSED METHODOLOGY	27
4.1 A temporal series approach to check the correlation between CG lightning and convective activity	27
4.2 Automatic selection of thunderstorms	31
4.2.1 Conversion of lightning stroke to lightning flash data	33
4.2.2 Influence of some parameters in the thunderstorm selection algorithm .	33
4.3 Temporal sliding-window for lightning and weather radar data	34
4.4 Estimation of the WRLR function	37
4.4.1 Optimization of the parameters for the sliding-window approach	40
4.4.2 Mapping of the WRLR function different sampling area sizes and win- dow durations	41
4.5 Mapping of the WRLR function for areas outside the reference radar range	45
4.6 Inference of the total rainfall	48
4.7 EDDA-chuva software	49

5	VALIDATION TESTS	53
5.1	RLR computation for Bauru	54
5.2	WRLR function estimation for the four seasons of the year	55
5.3	Mapping of the Summer WRLR function for different area sizes and window durations	61
5.4	Validation tests of the rainfall estimation for the WRLR function	63
5.4.1	Validations tests using the Bauru weather radar	64
5.4.2	Validations tests using the Presidente Prudente weather radar	74
5.4.3	Validations tests using the weather radar of São Roque	97
5.4.4	Validations tests using the weather radar of Pico do Couto	102
5.5	Spatial distribution of convective/total rainfall and estimation of total rainfall	107
5.5.1	Spatial distribution tests including MERGE data	117
6	CONCLUSIONS	125
6.1	Further works	127
	REFERENCES	129
	ANNEX A - PUBLISHED ARTICLES RELATED TO THIS THESIS	137

1 INTRODUCTION

The monitoring and tracking of convective activity is conventionally achieved by the use of weather radars, but many countries of America, Africa and Asia, as is the case of Brazil, lacks an extensive coverage of such radars. On the other hand, Brazil has a ground-based lightning detector network called RINDAT (Brazilian Integrated Lightning Detection Network), one of the largest in the world (PINTO et al., 2006). This network is currently being expanded in order to cover most of the Brazilian territory. Lightning detectors have a lower cost in comparison to weather radars. Assuming that convective activity and cloud-to-ground (CG) lightning are correlated to some degree, a software that generates fields of density of occurrence of lightning named EDDA (STRAUSS et al., 2013) is currently being employed to monitor and track convective activity at the Center for Natural Disasters Monitoring and Alert (CEMADEN) in Brazil. The EDDA name stands, in Portuguese, for estimation of density of atmospheric electrical discharges.

Convective activity is related to the convection suffered by a parcel of air that is at a higher temperature than the surrounding atmosphere. This implies in an unstable condition that may lead to the raising of the parcel. As it rises it may reach the condensation level and generate a cumulonimbus cloud that has a high vertical extension and presents intense updrafts and downdrafts favoring the occurrence of lightning and presenting heavy rainfall. Such phenomenon is called a thunderstorm (AHRENS, 1988). Thunderstorms may cause flash floods, damage urban infrastructure and properties, and cause landslides, resulting in deaths or injuries to the inhabitants of the affected area (KATZ; MURPHY, 1997). The real time or near real time monitoring of thunderstorms is critical for civil defense in order to take precautionary measures. This monitoring is typically performed using weather radar and meteorological satellites. Another thread is weather forecasting using numerical models that simulate the state of the atmosphere. Weather forecasts are issued by meteorologists using model outputs along with current meteorological data and images trying to predict the occurrence of thunderstorms and to identify the associated risks (AFANDI et al., 2013).

The focus of this work is the estimation of the amount and distribution of rainfall, which may cause floods and landslides. Rainfall estimation in near real time is typically performed from weather radar data. However, Brazil is a huge country that has over 8.5 million km² with less than 15% of its area is covered by weather radar. Alternatively, rainfall estimations can be obtained from meteorological satellites

like those of the Tropical Rainfall Measuring Mission (TRMM), National Oceanic Atmospheric Administration (NOAA) or Geostationary Operational Environmental Satellite (GOES) satellite series, but the accuracy of the corresponding rainfall estimations can be poor (RAMIREZ-BELTRAN et al., 2008; LIAO; MENECHINI, 2009).

However, assuming that convective rainfall can be correlated to cloud-to-ground lightning, some approaches proposed rainfall estimation from lightning data for areas without weather radar coverage. The most common approach is the computation of a constant-valued rainfall–lightning ratio (RLR) that expresses the convective rainfall mass per CG lightning flash (TAPIA et al., 1998). Nevertheless, such ratio may depend heavily on seasonal and geographical factors, local climatology, convective regime, storm type, lightning patterns or intensity, dominant lightning polarity of CG lightning, intracloud to CG ratio and thunderstorm life cycle (BUECHLER; GOODMAN, 1990; SOULA; CHAUZY, 2001; LANG; RUTLEDGE, 2002). Such dependency may explain why some former approaches failed to provide values of RLR with low variability (SIST et al., 2010).

A number of studies were performed since 1998 to estimate the rainfall mass directly from CG lightning observations. Petersen and Rutledge (1998) used the total rainfall mass and the density of CG lightning to examine their relationship on a number of spatial and temporal scales for different parts of the world. The lightning flash incidence is more intense in clouds associated to high-level precipitation, as the electrification increases with altitude as in the case of tall cumulonimbus (SINGH et al., 2010). The most known work of Tapia et al. (1998) computed the RLR as being the median of a set RLR values obtained for particular thunderstorms. Assuming a circular distribution of rainfall around each lightning flash, the overall spatial distribution of rainfall is assumed as the summation of the distribution of the flashes, which was checked against weather radar data. In addition, Kempf and Krider (2003) presented a compilation of RLR values including some obtained from other similar works, and found values ranging from 38×10^6 kg to 72×10^6 kg per flash for isolated thunderstorms in Florida, Spain and France, and values as high as 5000×10^6 kg per flash for mesoscale thunderstorms in Australia and Central United States. Molinie et al. (1999) found values as low as 3×10^6 kg per flash for the Pyrenees, while Williams et al. (1992) found value composed of up to 500×10^6 kg per flash for Australia.

The correlation between convective activity and CG lightning has been exploited in several works (BATTAN, 1965; KINZER, 1974; PIEPGRASS et al., 1982; BUECHLER; GOODMAN, 1990; WILLIAMS et al., 1992; WILLIAMS; RENNO, 1993; ZIPSER, 1994;

CHÈZE; SAUVAGEOT, 1997; TAPIA et al., 1998; PETERSEN; RUTLEDGE, 1998; MOLINIE et al., 1999; SOULA; CHAUZY, 2001; LANG; RUTLEDGE, 2002; KEMPF; KRIDER, 2003; SIST et al., 2010; SIINGH et al., 2010). An algorithm to select thunderstorms to obtain such correlation is also described in this thesis (GARCIA et al., 2012).

This methodology presented in this work may overcome the shortcomings of the former approach to estimate rainfall from lightning data. The current work proposes a simpler approach in order to estimate the rainfall mass from lightning data, that seems to be more accurate, the windowed-RLR function (WRLR), which employs a temporal sliding-window to scan both CG lightning and rainfall data. This was presented in Garcia et al. (2013), as part of this thesis, being based on the assumption that convective activity is correlated to electrically active cells that correspond to areas with high density of CG strokes. It is a way to tackle the high variability of the RLR, to extend rainfall estimation for different regions and conditions and to map the rainfall-lightning correlation for different windows in space and time.

As already mentioned, such density is calculated by the EDDA software that implements standard kernel estimation (STRAUSS et al., 2013).

A set of thunderstorms that occurred in the year of 2009 to 2013 in the Southeastern Brazil was selected from weather radar data to estimate a WRLR function. Typically, the radar and lightning data are partitioned into training and test sets. The former are used to infer the WRLR function, while the latter, to test it. For instance, data corresponding to the coverage of one weather radar was used as training data, and data of a different radar, as test data. Evaluation tests demonstrated that a 50 km \times 50 km sliding window with a temporal resolution of 30 min approximately presents the lower error for rainfall estimation. Therefore, a 50 km spatial grid was defined to cover the part of Brazil under lightning detector coverage. The WRLR function was inferred for a region covered by two specific weather radars, composed of 49 squares with 50 km edges in a superposition scheme detailed in Chapter 4. This function is mapped to the remaining of the grid using Brazilian rainfall statistics of recent years. Rainfall spatial distribution is derived from the density of occurrence of cloud-to-ground lightning. This approach for rainfall estimation is already implemented by the Edda-chuva software (“chuva” stands for rain, in Portuguese) that is being currently evaluated at CEMADEN. In addition to the convective rainfall amount and distribution, the software allows to estimate the amount and distribution of the total rainfall, i.e. convective plus stratiform rainfall. Test results show an improvement of the accumulated rainfall monitoring over the part of the Brazilian territory covered

by the network of lightning detectors.

Lightning data is acquired and processed, being available at CEMADEN in near real time. This assures that Edda-chuva have a reasonable monitoring capability of thunderstorms. Additionally, nowcasting using numerical models can be improved for forecasting purposes, since the meteorologist have additional information to issue his/her forecast. It is worth to note that the approach proposed in this work is not related to the prediction of thunderstorms. For this matter, there are standard approaches like Model Output Statistics (MOS) that constitute the backbone of the current rainfall prediction. MOS techniques provide objective forecasting of rainfall by establishing a statistical relationship between a predicate (occurrence of rainfall) and the variables outputted by a numerical model forecast (GLAHN; LOWRY, 1972). MOS techniques typically employ an extensive historical time series of weather data using similar atmospheric physical states to statistically correct a current forecast (HAMILL; WHITAKER, 2006). Rainfall estimations obtained from weather radar and meteorological satellites images are used to supplement the rain gauge data acquired in ground meteorological stations. However, rainfall estimations obtained from meteorological satellites like TRMM, NOAA or GOES, can be inaccurate (RAMIREZ-BELTRAN et al., 2008; LIAO; MENEGHINI, 2009).

Therefore, the main objective of this work is to show the proposed approach to estimate rainfall from CG lightning data, to present the EDDA-chuva software that implements this approach and to analyze its performance for rainfall estimation being operationally employed at CEMADEN.

The main contributions of this work can be viewed as the use of a temporal sliding-window to compute rainfall averages and the corresponding lightning data, the optimization of window size and temporal resolution to minimize rainfall estimation errors, the proposal of a function of the number of CG lightning occurrences to estimate rainfall, the spatial mapping of this function to grid points outside weather radar coverage using available rainfall statistics, and the inference of the total rainfall (convective plus stratiform) amount and spatial distribution from the convective ones.

Chapter 2 exposes the meteorological data used in this work. Chapter 3 is about techniques and methodologies that are related to the current work, but were not developed by the author. Chapter 4 presents the techniques and methodologies proposed for this research, while Chapter 5, the results of the numerical tests. Finally, Chapter 6 shows the conclusions and final remarks.

2 METEOROLOGICAL DATA

The meteorological data employed in this work is composed of weather radar data, lightning data acquired by a network of ground-based lightning detectors, and quarterly rainfall averages provided by a network of rain gauges and presented in a isohyets map. Isohyets are lines that have the same amount of accumulated rainfall. Rainfall estimations are inferred from the number of cloud-to-ground lightning occurrences for different periods and locations using the WRLR function. These estimations are then compared to the corresponding estimations obtained from weather radar data, which are assumed as references. For this purpose, a grid of 50 km resolution was adopted, covering most of the Brazilian territory. Since most of this territory is not covered by weather radars, the WRLR function is mapped to the other grid points using the quarterly rainfall averages. Weather radar are for the years of 2009, 2010 and 2013, lightning data are for the years 2007 to 2010 and 2013, and quarterly rainfall averages are for the years of 1977 to 2006.

Figure 2.1 shows the superposed areas of the Brazilian territory that correspond to the different data. Quarterly rainfall averages are available for the entire territory, while the blue rectangle refers to the standard coverage of the RINDAT lightning detector network, and the red circles, to the standard coverage of the four weather radars. In the case of RINDAT, the standard coverage is defined by a probability of detection above 80%. This coverage is detailed in Figure 2.6. In the case of the weather radars, by a range of 150 km considered as a limit for rainfall detection.

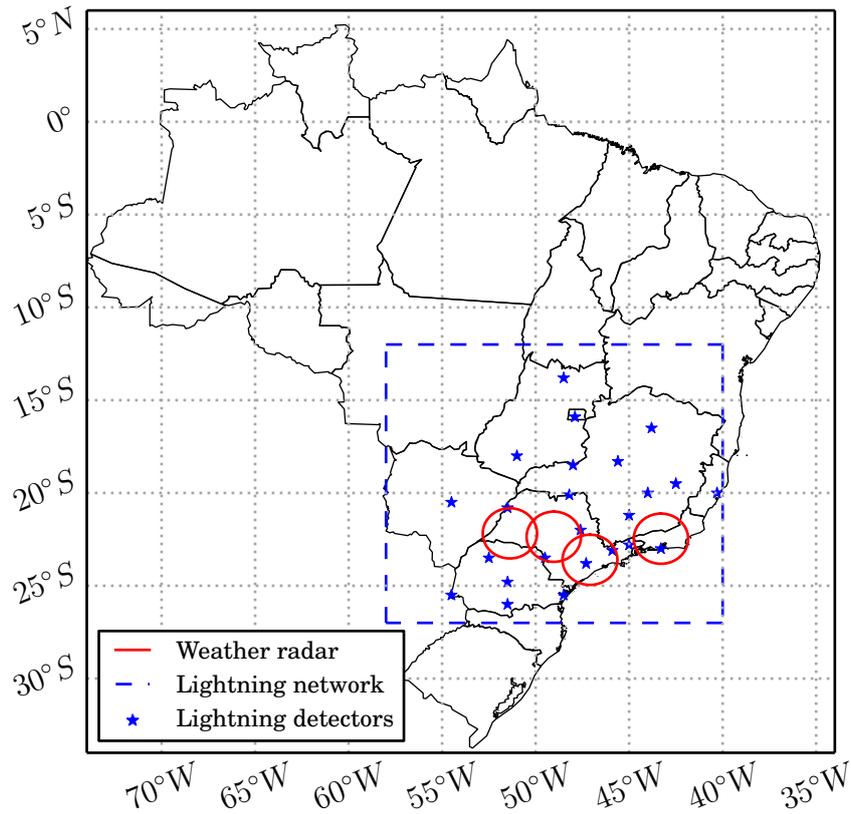


Figure 2.1 - Coverages of the Brazilian territory for the different meteorological data employed in this work: (i) quarterly rainfall averages (entire area), (ii) RINDAT lightning detection network (blue rectangle), and (iii) weather radars (red circles).

2.1 Rainfall

Rainfall is liquid water in the form of droplets that have condensed from atmospheric water vapor and then precipitated. This work focuses on the convective and stratiform rainfall associated with thunderstorms. Convective rainfall occurs from convective clouds, i.e., cumulonimbus or cumulus congestus. It falls as showers with rapidly changing intensity, and occurs over a certain area for a relatively short time during the earlier stages of a thunderstorm, as convective clouds have limited horizontal extension. Stratiform rainfall falls out of nimbostratus clouds and, in the cases studied in this work, occurs in later stages of a thunderstorm. It occurs over a larger area than convective rainfall, over a longer period and has lower intensity.

Figure 2.2 shows a lateral schematic view of a thunderstorm. The advancing front side of the cloud (right) has high vertical development and produces convective rainfall, while the tail part (left) presents weaker stratiform rainfall over a larger

area. The associated cloud is commonly referred as anvil, due to its shape.

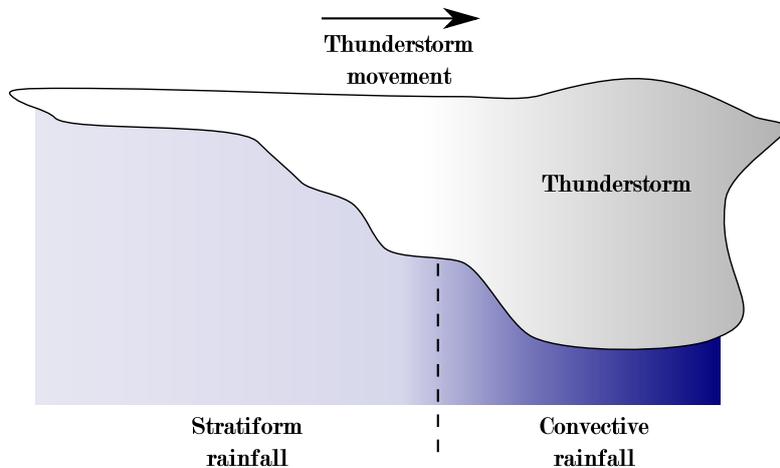


Figure 2.2 - Schematic view of a thunderstorm cloud, with the convective (right) and stratiform (left) rainfall.

2.2 Lightning

Lightning is a massive electrostatic discharge that occurs between the electrically charged regions within clouds or between a cloud and the surface of the Earth. The charged regions within the atmosphere temporarily equalize themselves through a lightning flash. There are three primary types of lightning, that occurs inside a cloud (IC), from one cloud to another (CC) and between a cloud and the ground (CG) (AHRENS, 1988).

Many factors affect the frequency, distribution, strength, and physical properties of a “typical” lightning flash in a particular region of the world. These factors include ground elevation, latitude, prevailing wind, relative humidity, proximity to warm and cold bodies of water, etc. To a certain degree, the ratio between IC, CC and CG lightning occurrence may also vary by season or region.

A typical CG lightning flash culminates in the formation of an electrically conducting plasma channel through the air from within the cloud to the ground surface. The resulting discharge is then the final stage of a very complex process. A typical thunderstorm produces, at its peak, three or more CG lightning per minute. CG lightning is more frequent near the convective portion of a thunderstorm.

2.3 Weather radar data

Radar is an acronym for *R*ADio *D*etection *A*nd *R*anging, which was coined during the Second World War. Such devices emit electromagnetic pulses in a given frequency, typically on the S band, and acquire the signal that is backscattered by the target. The corresponding delay is used to estimate target distance, altitude, size, relative speed and direction. In Meteorology, such targets are the hydrometeors like raindrops, snow or hail. The larger the raindrops and their number, the higher is the backscattered energy that allows to estimate the rainfall amount and distribution over the covered range and to track and characterize thunderstorms. Weather radar data amount is typically huge, and a single day of data from raw readings can go over 4 GB for each radar in a network (NATIONAL WEATHER SERVICE, 2007). Some low-altitude obstacles may generate false echoes, like buildings or even airplanes, birds or insects. Other limitation is a nearby intense thunderstorm blocking a more distant rainfall.

The energy backscattered by hydrometeors is measured by the reflectivity factor Z (in dBZ). The corresponding rainfall rate (in mm h^{-1}) is then estimated using the Z - R relationship, presented in Equation 2.1, whose parameters A and b are derived using rain gauge data as a reference of the actual rainfall.

$$Z = A \times R^b \quad (2.1)$$

This work adopted the Z - R relationship as described by Calheiros and Gomes (2010) for the radars of Bauru and Presidente Prudente located in the Brazilian state of São Paulo, as it was obtained for those radars, with $A = 32$ and $b = 1.65$.

However, for the other weather radars, Pico do Couto and São Roque, this work uses the pair of Z - R relationships adopted by CEMADEN for convective and stratiform rainfall, respectively, the Woodley (MARSHALL; PALMER, 1948) and the Marshall-Palmer (MARSHALL; PALMER, 1948) relationships. Considering Equation 2.1, the first one is given by $A = 300$ and $b = 1.4$ for $Z > 35$, while the second, by $A = 200$ and $b = 1.6$ for $Z \leq 35$. The limit reflectivity of $Z = 35$ to distinguish between convective and stratiform rainfall is a specific value adopted by CEMADEN.

Weather radar data employed in this work correspond to Constant Altitude Plan Position Indicator (CAPPI) images for an altitude of 3 km. These images are obtained by successive sweeps of the radar for different elevations, in a mode called PPI

(Plan Position Indicator). Each sweep corresponds to a fixed elevation angle, but varying the azimuthal angle from 0° to 360° , resulting in a particular PPI image. The composition of the set of PPI images for a given altitude gives the corresponding CAPPI image (ATLAS, 1990). Therefore, a CAPPI image corresponds to a horizontal cut of the atmosphere at a given altitude and allow to employ standard coordinates, instead of the cylindrical coordinates composed of the elevation and azimuthal angles. Figure 2.3 shows a vertical cut of the atmosphere highlighting the CAPPI at the 3 km altitude. It is worth to note that a weather radar has the capability of sweeps varying both the elevation and azimuthal angles. Therefore, it is possible to acquire a set of volumetric data and some weather forecast centers provide post processing that allows 3D visualizations.

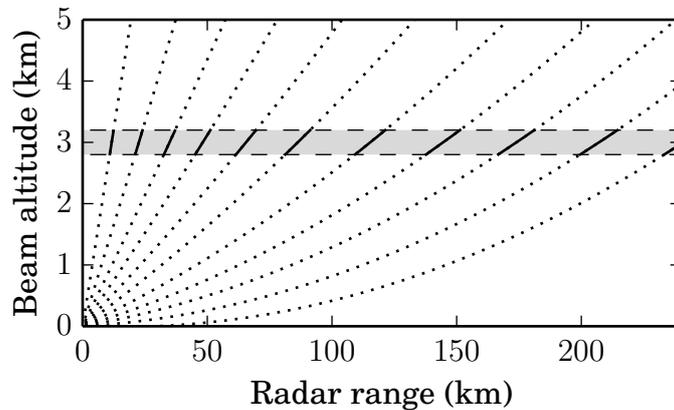


Figure 2.3 - Schematic view of radar pulses in a vertical plane for different elevation angles various, with the gray band highlighting the 3 km altitude CAPPI.

Weather radar networks are used worldwide, and operation examples include NEXRAD in the United States, OPERA¹ in Europe, and the Canadian weather radar network. However, many countries of continents like Central and South America, Africa and Asia generally lack a weather radar coverage, if any.

2.3.1 Weather radar range limitation

One of the weather radar limitations is an immediate consequence of Earth's curvature. A radar pulse may be considered as emitted tangent to the Earth surface at the radar location. The altitude of the pulse increases rapidly as the pulse moves away from the radar. Figure 2.4 shows this altitude distortion. Equation 2.2 allows

¹OPERational program for weather RAdar networking

to correct the altitude of the target (RINEHART, 1997) as

$$H = h_a - a_e k_e + \sqrt{r^2 + (a_e k_e)^2 + 2da_e k_e \sin \Theta_e}, \quad (2.2)$$

where,

- H : altitude correction,
- a_e : Earth radius at the radar location,
- h_a : target height relative to the radar,
- k_e : refraction coefficient of the air,
- d : distance to the target,
- Θ_e : elevation angle.

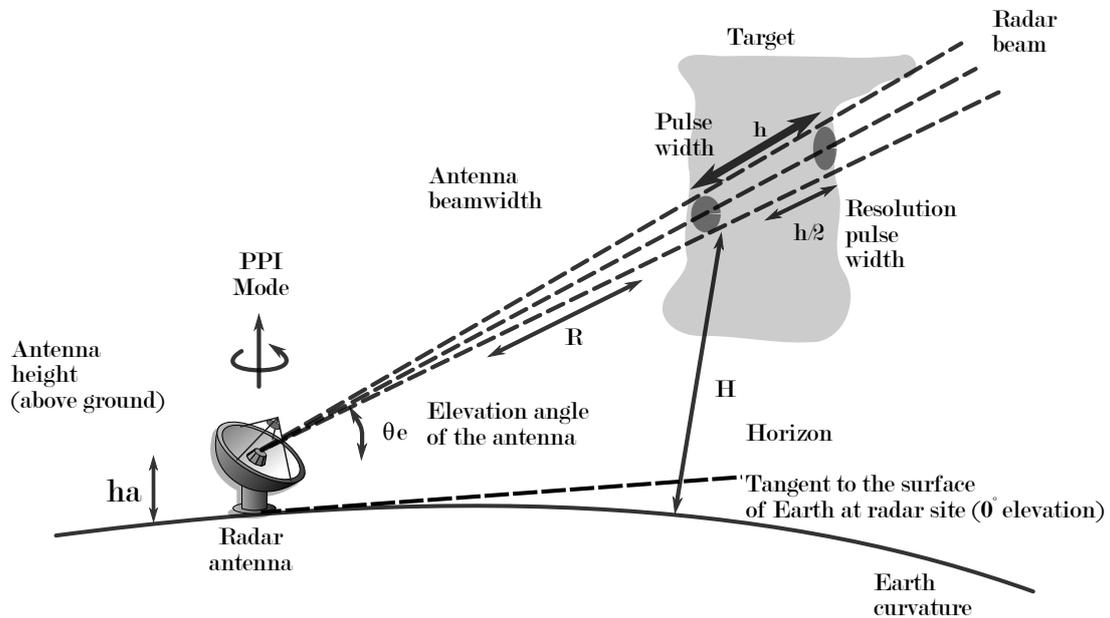


Figure 2.4 - Altitude distortion due to the curvature of the Earth.

The farthest measurable radar readings are typically over 450 km from the radar. At this distance, radar pulses are above the height of rain clouds, which cannot be detected. The typical useful radar range for S band radars is about 240 km, but for rainfall detection the range is limited here to 150 km (CALHEIROS; D'OLIVEIRA, 2007).

A radar pulse may detect a nearby target, but also a more distant target in the same direction, since part of the energy of the pulse is backscattered or absorbed by the nearby target, but the remaining part may reach the other target being backscattered later. However, a nearby intense thunderstorm may block all target behind

it (AHRENS, 1988). In some extreme cases of strong thunderstorms occurring just above the radar, all readings may be impaired. Similarly, variations on atmospheric density may cause refraction. In the case of super-refraction, the pulse is deflected downwards to the ground. In the case of under-refraction, it is deflected upwards. The pulse may also be reflected multiple times, causing artifacts (LEMON, 1998). In addition, some low-altitude obstacles may generate false echoes, as in the case of buildings or even airplanes, birds or insects.

2.3.2 Weather radar metadata

This work uses data acquired by four weather radars located in the Southwest of Brazil, in Bauru, Presidente Prudente and São Roque, in the state of São Paulo, and Pico do Couto, in the state of Rio de Janeiro. Their coordinates and altitudes are shown in Table 2.1.

Table 2.1 - Location of the weather radars used in this work.

Code	Radar	Latitude	Longitude	Altitude
BRU	Bauru	22°21'30"S	49°01'42"W	640 m
PI	Pico do Couto	22°27'51"S	43°17'50"W	1771 m
PPR	Presidente Prudente	22°10'30"S	51°22'30"W	493 m
RQ	São Roque	23°36'07"S	47°05'39"W	1148 m

All four used radars are Doppler radars and emit RF pulses in the S-band that is defined by frequencies from 2.0 GHz to 4.0 GHz, but considered All weather radar data employed in this work were acquired by four Doppler radars that emit RF pulses in the S-band. This band is defined by frequencies from 2.0 GHz to 4.0 GHz, but these particular four radars have frequencies ranging from 2.7 GHz to 3.0 GHz. Figure 2.5 shows the boundaries of the 150 km range for the four weather radars.

In general, the related CAPPI 3 km data files are available in two formats, binary files with single precision matrices or ASCII files containing sparse matrices. The spatial resolution for all radar data is 1 km, and maximum radar range is 240 km for the Bauru and Presidente Prudente radars and 250 km for the Pico do Couto and São Roque radars. The temporal resolution for the Bauru and Presidente Prudente radars is 7.5 min, but due to particularities of the data processing system such data is presented alternating intervals of 7 min and 8 min. The radars of Pico do Couto and São Roque have a temporal resolution of 15 min.

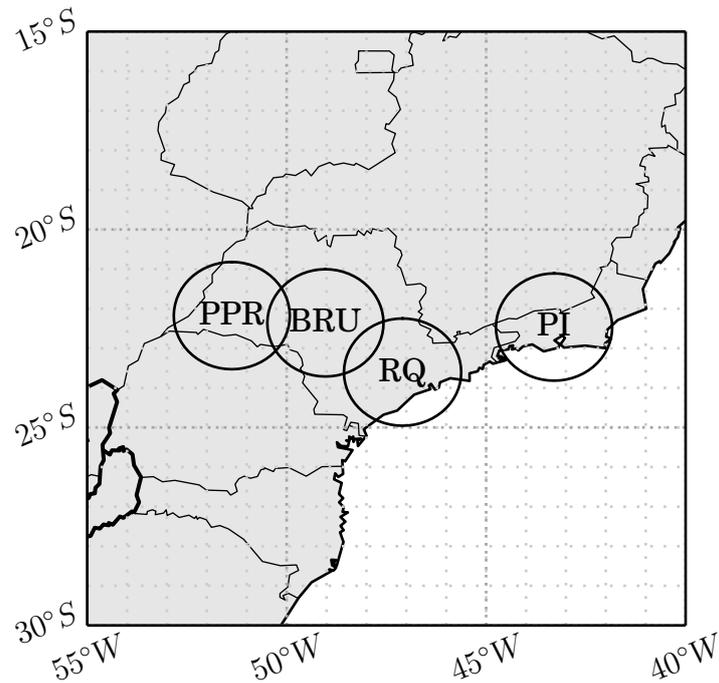


Figure 2.5 - Coverage of the 150 km range of the weather radars used in this work.

The ASCII files correspond to data of the radars of Bauru and Presidente Prudente for the years 2009 and 2010. Data is stored in compressed files using three columns per line. The first and second columns contain line and column number, respectively, and the third, the corresponding reflectivity value in dBZ, if not null. A sparse matrix results, since most of the grid points have zero reflectivity. These files follow the standard notation for images, with the first pixel being the uppermost-leftmost one. Additional ASCII files contain the latitude and longitudes corresponding to the line and column numbers. The corresponding months and years of data employed in this work are shown in Table 2.3.2.

The remaining data for the radars of Bauru and Presidente Prudente and the data of the radars of Pico do Couto and São Roque are presented in binary files. All this data is from the first three months of 2013. Each radar dataset requires a world file, an ASCII data file used by Geographic Information Systems (GIS) to georeference raster map images. Those files are stored with the bottommost-leftmost pixel being the first data point. A simple coordinate transformation allows to match the coordinates of CAPPI images of different radars.

Table 2.2 - Months/years of weather radar data in ASCII files for the radars of Bauru and Presidente Prudente.

	2009	2010
January	•	•
February	•	•
March	•	•
April	•	•
May	•	•
June	•	•
July	•	•
August	•	•
September	•	•
October	•	
November	•	
December	•	

2.4 Lightning data

Lightning is a massive electrostatic discharge between electrically charged regions of the atmosphere that occurs within clouds, between clouds or between a cloud and the Earth’s surface. A lightning flash occurs when the electric potential difference between particles of opposite charges exceeds the insulating capacity of the surrounding air. Each lightning flash is defined as one or more transient high-current discharges known as strokes, which may have an extent of kilometers (UMAN; KRIDER, 1989). The strokes that compose a flash usually last less than a millisecond and typically repeat themselves so rapidly that an observer cannot detect the multiple events. Sometimes, such repetition is slow enough in a way that the lightning appears to flicker. A lightning stroke acts as an electromagnetic emitter, and its RF signal can be measured even at great distances (NACCARATO et al., 2006). The number of strokes composing a single flash is called multiplicity.

Lightning flashes are classified into three classes, cloud-to-ground (CG), intracloud (IC) and cloud-to-cloud (CC). While CC flashes correspond to most of the lightning occurrences (UMAN; KRIDER, 1989), the CG flashes are better correlated with convective activity, and therefore are much more monitored due to the potential harm that they may cause to persons and properties (CUMMINS et al., 1998).

Each flash may also be classified according to the negative or positive polarity of its electrical charge. Negative CG flashes are more frequent as thunderclouds tend

to be negatively charged at the bottom, near the ground. Since thunderclouds are positively charged at their tops, positive CG flashes are usually more intense as the distance to the ground is larger.

Lightning data is presented in the Universal ASCII Lightning Format (UALF), a format developed by the Finnish company Vaisala that sells lightning detectors and weather radars, among other products. The information about each flash/stroke is stored in a separate line of the compressed ASCII file that is generated for a given time interval. Each attribute of the stroke is stored in one of the 25 separate columns. Columns include attributes such as time and location of occurrence, polarity and type (CG, IC or CC) that were employed in this work. Other columns include the estimated peak current and rise/decay time of the corresponding waveform that allow to estimate the electrical charge of the stroke. Table 2.4 lists the 25 columns of the UALF format. Columns 2 to 8 refer to the time of occurrence of the lightning flash/stroke, while columns 9 to 10, to the location of its occurrence. In this work, lightning data was provided by the RINDAT network, described in the next section. In this case, each line corresponds to a stroke and thus the column 12 (flash multiplicity) is always 1.

2.4.1 The RINDAT lightning detection network

RINDAT stands for National Integrated Network for Atmospheric Discharge Detection, or Rede Integrada Nacional de Detecção de Descargas Atmosféricas (in Portuguese). Its is an integrated network of lightning detectors and processing centers that allow to detect in real time lightning occurrences in part of the Brazilian territory. This network is due to a cooperation between two Brazilian electric system operators, CEMIG (Companhia Energética de Minas Gerais) and FURNAS (Furnas Centrais Elétricas) and two research institutions, INPE and SIMEPAR (Sistema Meteorológico do Paraná). Lightning data used in this work was acquired by RINDAT for the months/years shown in Table 2.4.1.

RINDAT has 5 processing centers, in the cities of Belém, Belo Horizonte, Curitiba, Rio de Janeiro and São José dos Campos. It employs a proprietary relational database (Sybase) and the LtraX (Real-time Lightning Tracking software) visualization and analysis environment from Vaisala. Additionally, RINDAT develops integrated visualization and analysis of lightning data (RINDAT, 2011). RINDAT is a hybrid sensor network composed of 25 sensors, from which 8 are IMPACT type and 17 are LPATS (Lightning Position and Tracking System) type, also from Vaisala. LPATS sensors are designed to detect CG flashes/strokes, but not CC ones, while IMPACT

Table 2.3 - Description of the columns of the Universal ASCII Lightning Format (UALF).

Column	Description
01	Version
02	Year
03	Month
04	Day
05	Hour
06	Minute
07	Second
08	Nanosecond
09	Latitude
10	Longitude
11	Estimated peak current
12	Flash multiplicity
13	Number of employed lightning detectors
14	Degrees of freedom
15	Ellipse angle
16	Length of semi-major axis of ellipse
17	Length of semi-minor axis of ellipse
18	Chi-square from location optimization
19	Rise time of the waveform
20	Peak-to-zero time of the waveform
21	Maximum rate-of-rise of the waveform
22	Type of lightning (CG or CC)
23	Angle indicator
24	Signal indicator
25	Timing indicator

Table 2.4 - Months/years of RINDAT data employed in this work.

	2007	2008	2009	2010	2013
January	•	•	•	•	•
February	•	•	•	•	•
March	•	•	•	•	•
April	•	•	•	•	•
May	•	•	•	•	•
June	•	•	•	•	
July	•	•	•	•	
August	•	•	•	•	
September	•	•	•	•	
October	•	•	•		
November	•	•	•		
December	•	•	•		

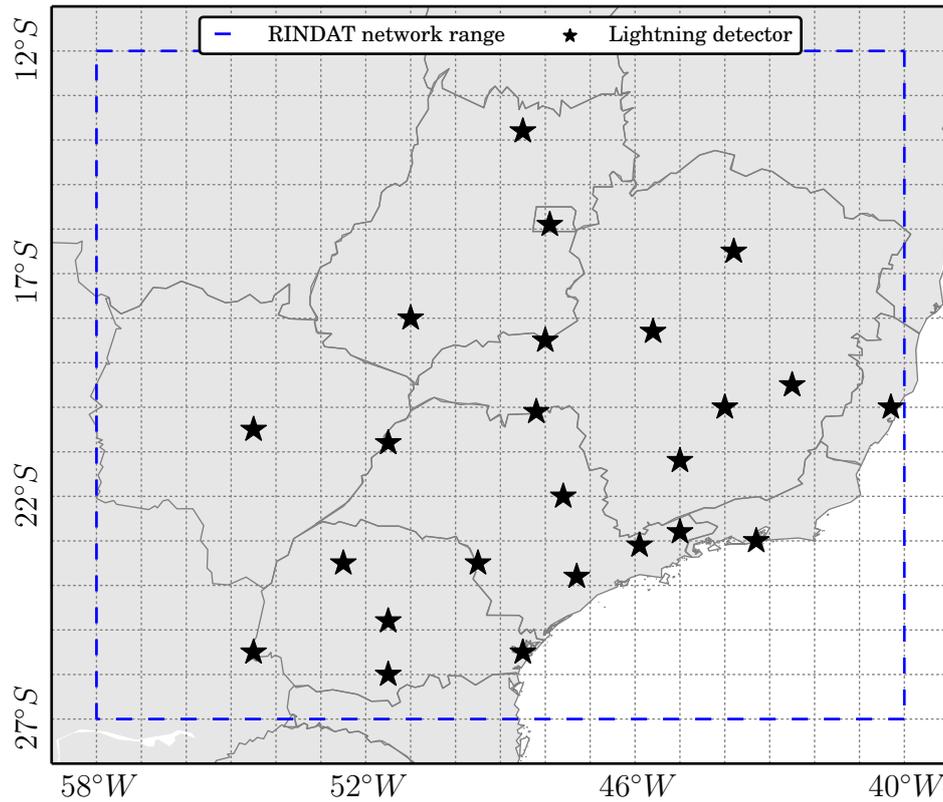


Figure 2.6 - Location of the current 25 detectors of RINDAT (two sensors are superposed, in central Paraná). The blue dashed line represents the range of the RINDAT with probability of detection above 80% (RINDAT, 2011).
SOURCE: Naccarato et al. (2006)

sensors are able to detect both CG and CC flashes/strokes. CC lightning occur three times higher than CG lightning (UMAN, 2001), but the latter is important in the scope of this work, since its occurrence is correlated to convective activity. The location of the current RINDAT lightning detectors and the approximate coverage above 80% detection efficiency for the same network appears in Figure 2.6.

In the case of the LPATS sensors, a processing center receives data of a set of lightning detectors synchronized by the Global Position System (GPS), supplying time information with a 100 nanosecond resolution as well as its location, polarity and maximum return stroke current. An algorithm based on the time of arrival of the RF signal emitted by the lightning allows to define its location, but requires at least 3 detectors. On the other hand, IMPACT sensors also employ magnetic direction finding and require only 2 detectors. If a flash/stroke is detected by a suitable number of lightning detectors, RINDAT can provide its location with an

error up to 500 m and can differentiate two different strokes that occurred with a 300 ns delay.

2.5 Isohyetal map of quarterly rainfall for Brazil

Another source for rainfall data for this work are the quarterly rainfall averages obtained from the Pluviometric Atlas of the Brazilian Geological Service (CPRM - Serviço Geológico do Brasil, in Portuguese). Rainfall averages are obtained from a network of rain gauges (pluviometers) that covers most of the Brazilian territory. This data is presented as shapefiles containing isohyets of rainfall averages from January 1977 to December 2006. It is possible to display yearly, quarterly, or monthly rainfall averages as well as to check the more rainy or dry months or quarters for a given year. A shapefile is a digital vector storage format for storing geometric location and associated attribute information. Shapefile was created to be used with Geographical Information Services (GIS), but are now used in a variety of softwares.

A isohyet is a line on a map connecting points that have the same amount of rainfall in a given period, being similar to contour lines. The shapefiles containing the isohyets have to be converted to raster images files with the same grid resolution of the weather radar data (1 km).

The overall area covered by the isohyetal maps is bounded by a square of 40° of latitude by 40° of longitude. The latitude is limited northernmost at $06^\circ 00' 00''$ N and southernmost at $34^\circ 00' 00''$ S, while the longitude is limited westernmost at $74^\circ 00' 00''$ W and easternmost at $34^\circ 00' 00''$ W. This square encompasses all the continental area of Brazil. However, only a small part of this area has coverage of weather radars and the lightning detector network does not cover the entire area.

This work uses two sets of isohyetal maps. The first one refers to the rainfall averages for the 1977–2006 period, in order to give a general view of the rainfall distribution in Brazil. The second set is composed of the quarterly rainfall average for all quarters of these years (1977 to 2006), specifically January–March, April–June, July–September and October–December, as a way to approximately describe the rainfall distribution for each of the four seasons. The cumulative rainfall for each season (approximated by the nearest quarter) is presented in Figure 2.7, with Summer corresponding to January–March, Autumn to April–June, Winter to July–September and Spring to October–December. It is important to notice the isohyets represent total rainfall and, therefore, do not allow to distinguish between convective and stratiform rainfall. However, if a constant ratio between the amounts of convective and stratiform

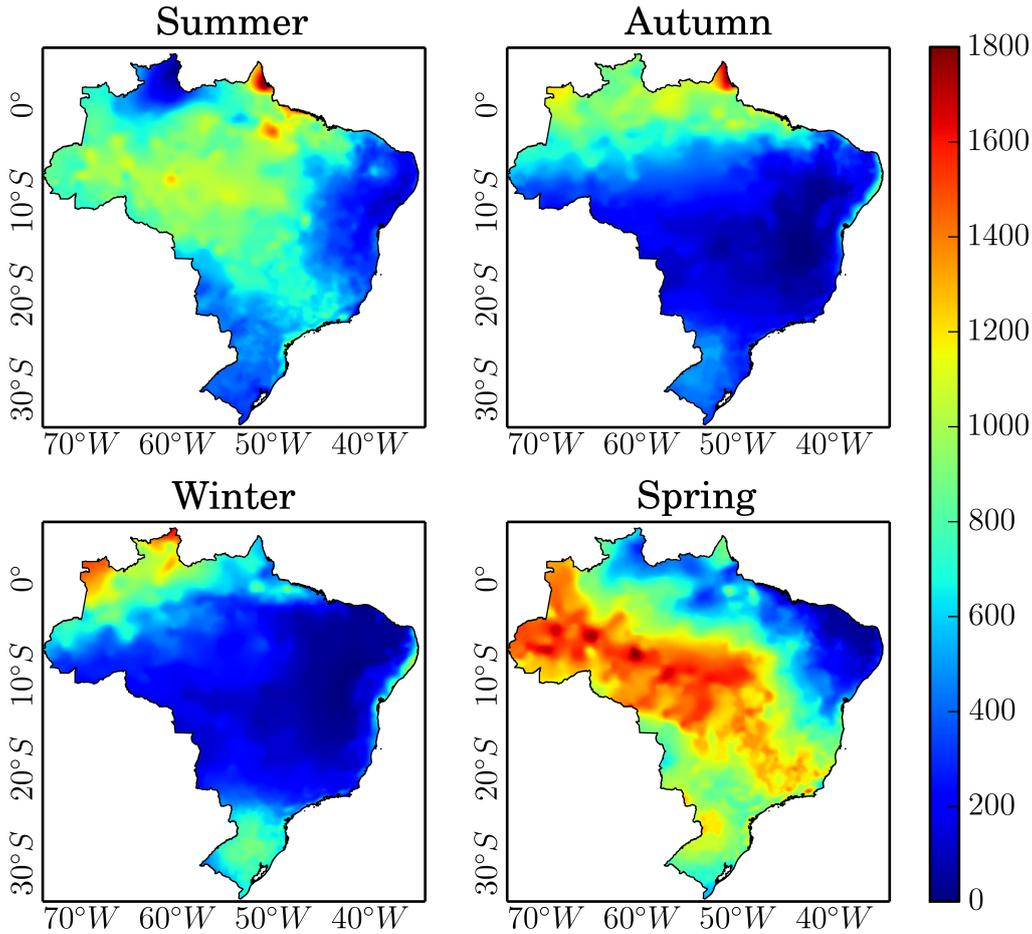


Figure 2.7 - Average yearly rainfall totals (mm) per season for Brazil during 1977–2006.

rainfall is assumed, then the isohyetal maps can be employed for the convective rainfall. Alternatively, if an estimative of such ratio is available for the grid points, more accurate convective rainfall amounts would be obtained. These issues are discussed ahead, in Chapter 4.

2.6 MERGE precipitation data

This work employed the isoyethal maps of quarterly rainfall averages, described in the previous section. This data is acquired by a network of rain gauges of different agencies/sources. However, as Figure 2.8 shows for the year of 2010, the spatial distribution of the rain gauges is very uneven over Brazil, presenting a higher density along the Atlantic coast. On the other hand, estimations of rainfall from the TRMM satellites may be innacurate, since systematic errors were observed, for instance in the coast of Northeast Brazil due to warm clouds (ROZANTE et al., 2010). in order to try to improve rainfall estimations, some researchers of CPTEC/INPE and

from the University of Maryland proposed a new approach, implemented in the projet called MERGE, that combines TRMM rainfall estimations and rain gauge data over South America (ROZANTE et al., 2010). This approach considers either rain gauge observations or TRMM rainfall estimations depending on the presence or neighbourhood of rain gauges in a 0.25° grid. Test results show that MERGE performance for rainfall estimation was equivalent to the rain gauge network over regions with high density of observations, but was better over regions with sparse observations, based on climatological expected values of rainfall. MERGE data is currently considered by CPTEC meteorologists as the most reliable rainfall averages for Brazil.

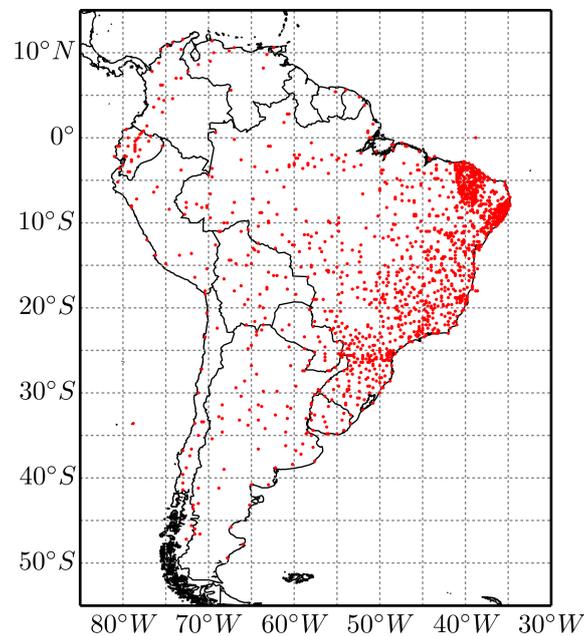


Figure 2.8 - Spatial distribution of rain gauges of different sources/agencies over South America in 2010.

SOURCE: Rozante et al. (2010)

In this work, MERGE data was considered for the month of January 2013, for the rectangular area bounded by latitudes $27^\circ 00' 00'' S$ to $12^\circ 00' 00'' S$ and longitudes $58^\circ 00' 00'' W$ to $40^\circ 00' 00'' W$, which corresponds to the area of RINDAT network with probability of detection above 80%, shown in Figure 2.6.

3 RELATED WORK

The methods and techniques described in this Chapter, while important for this work, were not developed by the author. Section 4.1 provides some evidence about the correlation of CG lightning occurrence and convective activity. Section 3.1 presents the standard approach proposed by Tapia to estimate the amount and distribution of convective rainfall from the number of occurrences of CG lightning. Section 3.2 is about the EDDA software that generates fields of density of occurrence of lightning, while Section 3.3, about the Steiner criteria to filter out convective precipitation from the reflectivity values in weather radar images.

3.1 Tapia's model

The model proposed by Tapia et al. (1998) allows to estimate rainfall spatial and temporal distribution using a previously estimated RLR value. A uniform rainfall distribution is assumed in a circle of $\Delta x = 5$ km radius centered at each CG lightning flash and in a $\Delta t = 5$ min interval centered at its time of occurrence. The model presented by Tapia et al. (1998) to estimate the spatial and temporal distribution of rainfall intensity is represented by Equation 3.1.

$$R(t, x) = C \sum_{i=1}^{N_f} \text{RLR} \cdot f(t, T_i) \cdot g(x, X_i), \quad (3.1)$$

where,

- $R(t, x)$: Rainfall rate in mm/h at time t and position x ,
- C : Unit conversion factor,
- N_f : Counter for the number of flashes up to the time $t + \frac{\Delta t}{2}$,
- RLR: Constant rainfall-lightning ratio (kg/CG flash),
- T_i : Time of occurrence of the i -th flash,
- X_i : Location of occurrence of i -th flash.

The temporal distribution $f(t, T_i)$ is presented in Equation 3.2. According to Tapia et al. (1998) the rainfall intensity is considered to be constant for a Δt time interval.

$$f(t, T_i) = \begin{cases} 1 & \text{if } |t - T_i| < t + \frac{\Delta t}{2} \\ 0 & \text{otherwise} \end{cases} \quad (3.2)$$

Rainfall is assumed as uniformly distributed inside a Δx circle centered at the light-

ning flash, as described by Equation 3.3.

$$g(x, X_i) = \begin{cases} 1 & \text{if } |x - X_i| < \Delta x \\ 0 & \text{otherwise} \end{cases} \quad (3.3)$$

3.2 The EDDA software

The EDDA software estimates the density of lightning occurrence for a given geographic region and time interval (STRAUSS et al., 2010; STRAUSS et al., 2013). The density field of occurrences is smooth, defining more clearly the region of convective activity than the lightning occurrences, which are very sparse in space and time. This approach was originally proposed in Politi et al. (2006) in order to track the electrically active convective activity using CG lightning data.

This software is currently being operationally evaluated at CEMADEN since October 2012. It outputs binary files containing the density of lightning stroke occurrence for every 15 min for the area with the RINDAT coverage. These files are visualized by the SALVAR environment that provides selection, handling and visualization of geospatial data of multiple sources like numerical models, meteorological satellites, land use type, political and topographic maps, etc. Point-related information can also be displayed, like rain gage readings. The name SALVAR stands for “to save” in Portuguese and means Platform for Monitoring and Emission of Alerts of Natural Disasters. These alerts are typically issued due to high amounts of accumulated rainfall in a given area or excessive flow rate of rivers. These events may cause floods or landslides and is obviously important to issue alerts to the Civil Defense.

The EDDA software implements the Gaussian kernel estimator with fixed (Equation 3.9). The Earth is considered a sphere and the distance $d(A, B)$ over the surface between two points A and B is approximated by the corresponding Euclidean distance calculated using geographical coordinates. However, the component given by the latitude difference is corrected by the cosine of the mean latitude of the two points, since this component is not along a great-circle of Earth’s surface. In the equation below, “lat” and “lon” mean latitude and longitude, respectively, while “lat_{max}” and “lat_{min}” are the maximum and minimum latitude of the considered grid points.

$$d(A, B)^2 = (\text{lat}_A - \text{lat}_B)^2 + (\text{lon}_A - \text{lon}_B)^2 \times \cos\left(\frac{\text{lat}_{max} + \text{lat}_{min}}{2}\right) \quad (3.4)$$

The EDDA software uses a configuration file to select the time interval, the geographical area, the grid size, as well as a filter to select the polarity and type of the lightning. It is possible to select the output format as ASCII tables containing longitude, latitude and the calculated density for each grid point, in order to perform meteorological data mining. The software also generates animations of the lightning density field in order to follow the temporal evolution of convective structures. The corresponding frames are generated using a temporal sliding-window. Since lightning data are processed in near real time, such animations allow to monitor the current status of the weather, as well as its temporal evolution.

Additionally, the EDDA software has been used in some studies concerning data mining of weather data, in this case, using high lightning density values as a decision attribute indicative of convective activity. The CG lightning density field of occurrences was used in the training phase of algorithms which identify patterns on forecasts of numerical models (LIMA; STEPHANY, 2013; PESSOA et al., 2012).

Density estimation allows to generate a density field of occurrence from a set of sampled events. In the case of this study, such events are lightning occurrences, resulting in a fictitious density field of lightning occurrences for the considered time interval. The classic two-dimensional kernel estimator (symmetric in relation to both dimensions) (SILVERMAN, 1986) can be written as:

$$\hat{f}(X) = \frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{d(X, X_i)}{h}\right), \quad (3.5)$$

where,

- X : coordinate pair (x, y) for a two-dimensional grid,
- $\hat{f}(X)$: density estimation at a grid point X ,
- n : number of samples,
- $d(X, X_i)$: Euclidean distance of a grid point X to the event X_i ,
- $K(r)$: two-dimensional kernel function,
- r : Euclidean distance normalized by h ,
- h : smoothing window width at the grid point X .

For the two-dimensional symmetric Gaussian kernel:

$$\hat{f}(X) = \frac{1}{2\pi nh^2} \sum_{i=1}^n \exp\left[-\frac{(x-x_i)^2 + (y-y_i)^2}{2h^2}\right] \quad (3.6)$$

The parameter h controls the smoothness of the field and can be chosen manually according to the desired visualization of the field. The higher its the value, the smoother the generated field.

However, an automatic scheme for adjustment h was proposed in [Silverman \(1986\)](#), in which the value of h is calculated to minimize the Mean Integrated Squared Error (MISE):

$$\text{MISE}(\hat{f}) = E \int \{\hat{f}(X) - f(X)\}^2 dx \quad (3.7)$$

Assuming a Gaussian distribution, the optimal value of h that minimizes the MISE is expressed in terms of the standard deviation σ of the Gaussian as:

$$h_{\text{opt}} = n^{-1/6} \sigma \quad (3.8)$$

A possible choice for σ is the root of the average of the variances of the coordinates of events sampled in each dimension:

$$\sigma^2 = \frac{1}{2}(\sigma_x^2 + \sigma_y^2), \quad (3.9)$$

where,

$$\sigma_x^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2$$

$$\sigma_y^2 = \frac{1}{n} \sum_{i=1}^n y_i^2 - \left(\frac{1}{n} \sum_{i=1}^n y_i \right)^2$$

A similar scheme ([SCOTT, 2009](#)) considers the non-symmetric case, calculating the window width h and the standard deviation σ separately for each dimension (in this case the kernel function is one-dimensional):

$$\hat{f}(X) = \frac{1}{nh_x h_y} \sum_{i=1}^n K_1 \left(\frac{x - x_i}{h_x} \right) K_1 \left(\frac{y - y_i}{h_y} \right), \quad (3.10)$$

where,

$$K_1(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$$

$$h_x = n^{-1/6} \sigma_x$$

$$h_y = n^{-1/6} \sigma_y$$

The schemes described above calculate a constant value of h for all grid points. Silverman (1986) proposes that better results can be obtained by varying the value of h as the density at the grid point considered:

$$\hat{f}(X) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_i^2} K \left(\frac{d(X, X_i)}{h_i} \right), \quad (3.11)$$

where,

$$h_i = h \lambda_i$$

$$\lambda_i = (\tilde{f}(X_i)/g)^{-\alpha}$$

Here h is the optimal global window, h_i is the local window width, λ_i is the local window width factor, $\tilde{f}(x_i)$ is a prior estimate of the local density using fixed window and α is a parameter between 0 and 1 which defines the influence degree of the local density on the local window. The λ_i are normalized by g , which is the geometric mean of $\tilde{f}(x_i)$. This approach is employed in Section 4.7 (EDDA-chuva software) for the spatial distribution of the convective rainfall.

3.3 The Steiner criteria

Steiner et al. (1995) proposed a technique to identify convective rainfall in 2D weather radar images of reflectivity in dBZ, specifically CAPPI 3 km images. According to this criteria, convective precipitation is supposed to present a high contrast with the background stratiform precipitation. In these images, a grid point is classified as presenting convective rainfall if it satisfies any of the following criteria:

- a) **Intensity:** the grid point has reflectivity of at least 40 dBZ;
- b) **Peakedness:** the grid point exceeds the average background reflectivity Z_{bg} (in dBZ) of the neighbouring grid points by at least the difference ΔZ (in dB) defined in Equation 3.12. This background intensity is given by the average of non-zero radar echoes within a radius of 11 km around the grid point;

$$\Delta Z = \begin{cases} 10, & Z_{bg} < 0 \\ 10 - \frac{Z_{bg}^2}{180}, & 0 \leq Z_{bg} < 42.43 \\ 0, & Z_{bg} \geq 42.43 \end{cases} \quad (3.12)$$

- c) **Surrounding area:** the grid point is on the convective neighborhood of other grid point that meets the criteria “a” or “b”; this convective neighborhood is defined by an intensity-dependent convective radius defined in Figure 3.1.

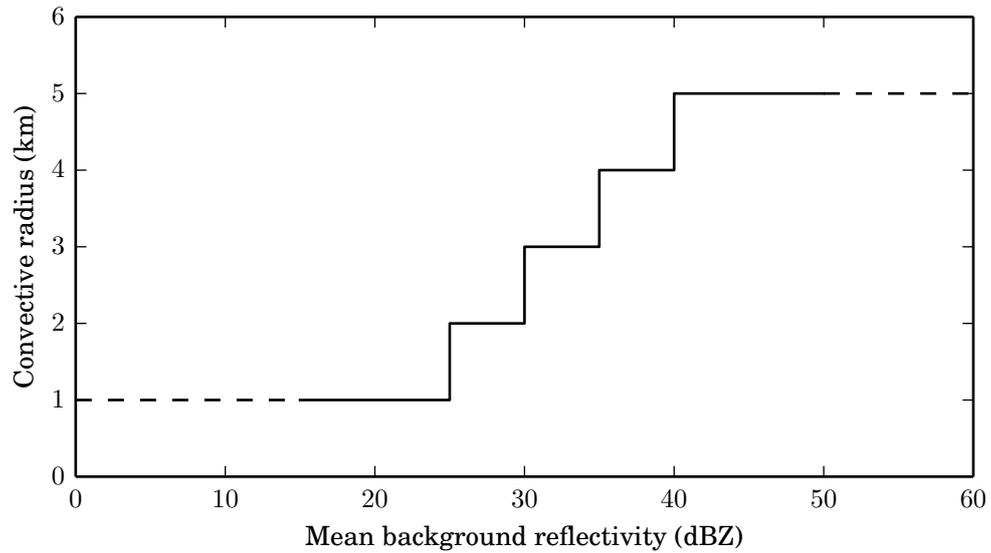


Figure 3.1 - The intensity-dependent convective radius as a function of the average background reflectivity Z_{bg} .

4 PROPOSED METHODOLOGY

The main objective of this work is to estimate rainfall from the number of CG lightning occurrences using a windowed-RLR function (WRLR) that was proposed by the author in (GARCIA et al., 2013). Such approach, based on a function, seems more suitable for this purpose, since a fixed-value RLR would present too much variability, as already discussed in Chapter 1. The WRLR function is derived employing a temporal sliding window to scan both CG lightning and convective rainfall data considering a set of sampling squares of 50 km side. Convective rainfall data is itself estimated from weather radar data, filtered by the Steiner criteria. Part of the lightning and rainfall data is used to derive the function, during the “training phase”, while the remaining part of the data is used to test the accuracy of the WRLR function, in the “test phase”.

The proposed approach allows to derive a specific WRLR function using rainfall data of a particular weather radar. In the case of this work, the Bauru weather radar was chosen to derive the “reference” WRLR function. However, the goal is to estimate convective rainfall using only lightning data and outside the coverage of any weather radar, since obviously a weather radar does not require any other way to estimate rainfall inside its coverage. Therefore a new question arises, how to map the WRLR function in order to get rainfall estimates from lightning data outside radar coverage? The answer is to use quarterly averages of rainfall for a specific grid to map the reference function from the grid point corresponding to the location of the radar to the remaining grid points.

4.1 A temporal series approach to check the correlation between CG lightning and convective activity

Mattos and Machado (2011) analyzed the life cycle of convective mesoscale systems over the state of São Paulo, using satellite data in the infrared and microwave bands, and also lightning data. Adopting the Pearson correlation coefficient, they determined that CG lightning occurrence was well correlated to the size of the convective system (correlation of 0.96), to the cloud top height (correlation of 0.84), to the integrated ice content (correlation of 0.86) and to the size of precipitating particles (correlation of 0.90). They concluded that thunderstorms, i.e. storms with high occurrence of lightning are, in average, longer and cover a larger area than standard storms. The maximum density of lightning occurs at the beginning of the convective mesoscale system, while lightning occurrence rate reaches a maximum during the growth phase, near the maturation of the convective mesoscale system. Oliveira

and Mattos (2011) studied such correlation for the city of São Paulo considering a square area of 1° side. They found that in Summer there is a higher amount of rainfall and high number of occurrences of lightning, but CG lightning is less intense, while in the Winter, there is less rainfall and less lightning occurrences, but lightning is more intense. They also concluded that correlation between rainfall and lightning was low, but it can be due to the delay between rainfall and lightning. However, in another work, Beneti et al. (2012) found a good correlation between rainfall estimated from weather radar and CG lightning occurrences for the diurnal cycle of convective mesoscale systems in the Brazilian Southeast.

In order to check the correlation between CG lightning and convective activity, weather radar and lightning data of 2009 were compiled for a square area of $50 \text{ km} \times 50 \text{ km}$ in the nearby of the city of Bauru, in a way that Bauru is exactly on the southeast vertex of the square. Weather radar data provided by radars of Bauru and Presidente Prudente was employed to estimate convective precipitation by means of a suitable Z-R relationship (CALHEIROS; GOMES, 2010). The corresponding lightning data was obtained from RINDAT in order to generate temporal series. Each point of the series is given by a pair of 30 min accumulated values of the CG strokes and of the convective precipitation. Three different series are shown for the same square area: the year the 2009, the month of September 2009, and 54 hours of September 8-10th 2009 that cover the extent of a particular thunderstorm. The curves that correspond to the temporal series of these periods of time, appear in Figures 4.1, 4.2 and 4.3. These curves were smoothed by a one-dimensional Gaussian filter. Filter width was 24 hours (48 data points) for the first set of curves (year of 2009), 2 hours (4 data points) for the second (September of 2009) and one hour for the third (thunderstorm of September 8-10th 2009).

Correlation values between CG stroke number and precipitated mass were then calculated for the three sets of curves, but using raw data before smoothing. However, calculating correlation using the complete temporal series would be useless, since most data points present no rainfall or no strokes. Therefore, we identified 259 thunderstorms for the year of 2009 and 31 thunderstorms for the month of September 2009, always for the same square area. An event was considered as a thunderstorm if it presented rainfall above a threshold, if the rainfall persisted for at least two 30 min interval and if it also complied to the Steiner criteria for convective precipitation. The correlation is then calculated between the number of CG strokes and the precipitated mass for each thunderstorm, always considering 30 min accumulated values of the number of strokes and rainfall. We also performed some cross correlation checks

using the same data, but maximum correlation was attained with zero delay. The study of such correlation is important, but would be out of the scope of the current work.

The correlation results and the three figures follow. The first case refers to the 259 thunderstorms included in the temporal series of the 30 min accumulated number of CG strokes and of precipitated mass for the entire year of 2009 (see Figure 4.1). Correlation values were calculated for each one of the 259 thunderstorms and presented mean of 0.77, median of 0.78, standard deviation of 0.15 and range [0.29, 0.98].

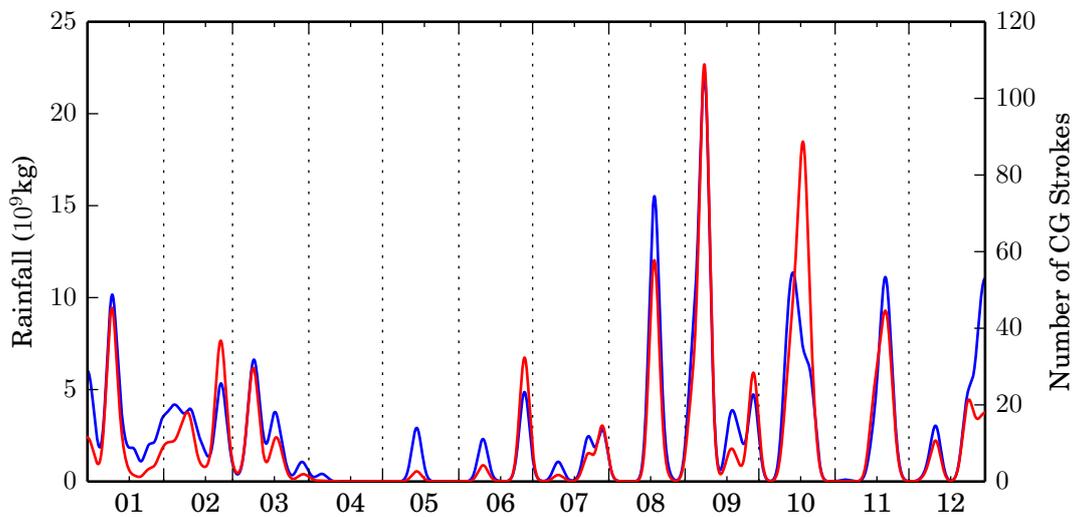


Figure 4.1 - Temporal series of accumulated 30 min values of the number of CG strokes and precipitated mass (rainfall) for the year of 2009 and for the square area near the city of Bauru (curves smoothed by 1D Gaussian filter).

The second case refers to the 31 thunderstorms included in the temporal series of the 30 min accumulated number of CG strokes and of precipitated mass for the month of September 2009 (see Figure 4.2). Correlation values were calculated for each one of the 31 thunderstorms and presented mean and median of 0.82, standard deviation of 0.12 and range [0.53, 1.00]. Finally, the third case refers to a sole thunderstorm that appeared in the temporal series of the 30 min accumulated number of CG strokes and of precipitated mass for the days of 8-10th of September 2009 (see Figure 4.3). The correlation calculated for this particular thunderstorm is 0.78.

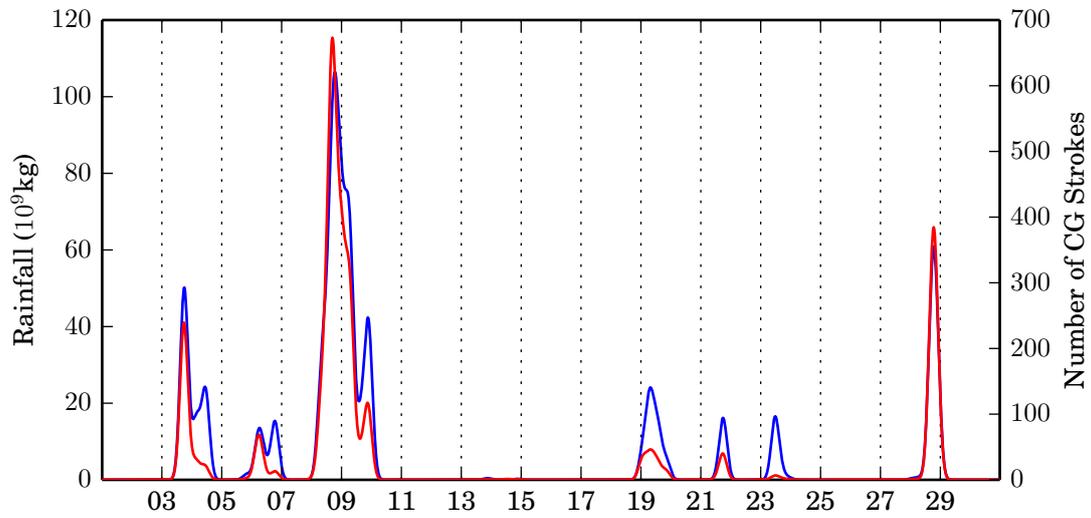


Figure 4.2 - Temporal series of accumulated 30 min values of the number of CG strokes and precipitated mass (rainfall) for the month of September of 2009 and for the square area near the city of Bauru (curves smoothed by 1D Gaussian filter).

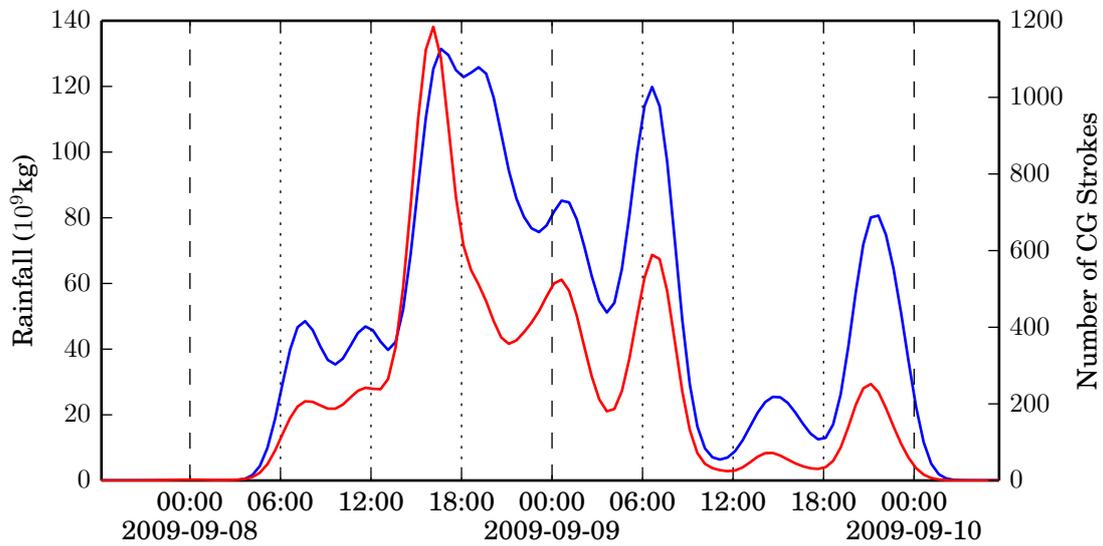


Figure 4.3 - Temporal series of accumulated 30 min values of the number of CG strokes and precipitated mass (rainfall) for some days of September 2009 and for the square area near the city of Bauru during a particular thunderstorm (curves smoothed by 1D Gaussian filter).

4.2 Automatic selection of thunderstorms

An approach that was adopted in the beginning of this work was the automatic selection of thunderstorms. Thunderstorm selection by hand would be an exhaustive task due to the high volume of lightning and rainfall data and the non-uniformity of the corresponding databases. As a consequence, previous works that selected thunderstorms in this fashion employed a relatively small number of events and found RLR values with high variability. In this work, a specific algorithm was developed for the automatic selection of thunderstorms (GARCIA et al., 2012), which is described as follows. The use of this algorithm was later abandoned since the sliding-window scheme was adopted. Nevertheless, the automatic selection of thunderstorms was useful for applying the Tapia’s model for lightning data in the Southeast Brazil.

The proposed algorithm for automatic thunderstorm selection performs a search for CG lightning clusters in the lightning data and checks if there is a minimum amount of convective rainfall associated to each cluster. Considering a list of lightning occurrences $V = \{v_1, v_2, \dots, v_n\}$, a graph $G = (V, E)$ is generated with each node of v_i of the graph representing a lightning occurrence with corresponding time of occurrence t_i and location x_i , i.e. $v_i = (t_i, x_i)$. Each edge e_{ij} connects two nodes v_i and v_j that do not exceed minimum time and distance differences of the corresponding lightning occurrences. These differences are expressed by the parameters Δt_G and Δx_G , defined in a suitable manner in Section 4.2.2. Thus, $E(G)$ is obtained by:

$$E(G) = \{e_{ij} \mid v_i, v_j \in V(G), (|t_i - t_j| \leq \Delta t_G) \wedge (|x_i - x_j| \leq \Delta x_G)\}, \quad (4.1)$$

where,

- Δt_G : maximum allowed difference in the time of occurrence of two flashes,
- Δx_G : maximum allowed difference in the Euclidean distance of two flashes.

As a result, $G(E, V)$ is an undirected, non weighted, simple and disconnected graph. It is possible to determine a subset w_i of nodes v_j that are connected to a given node v_i by edges e_{ij} , i.e. $w_i = \{v_j \mid e_{ij} \in E(G)\}$. The resulting set $W = \{w_1, w_2, \dots, w_n\}$ does not represent a set of lightning clusters since their elements may not overlap (for instance, w_1 may be connected to w_4 , while w_2 may also be connected to w_4 , but not to w_1). A spanning tree algorithm (Algorithm 4.1) was then developed in order to traverse the set $W(G)$ in an element-wise manner merging nodes that are connected to obtain a subgraph of G corresponding to a tree f_k . Each tree is a representation of a lightning cluster, while the union of all these disjoint trees constitutes the forest $F = \{f_1, f_2, \dots\}$.

Algorithm 4.1 Spanning tree algorithm for the generation of the set of lightning clusters F .

Require: The set $W(G)$ derived from the graph $G(V, E)$.

```

1: function FORESTGENERATOR( $W$ )
2:    $F \leftarrow \emptyset$ 
3:   for all  $w_i \in W$  do
4:      $f_i \leftarrow \text{JOIN}(W, w_i)$ 
5:     if  $f_i \neq \emptyset$  then
6:        $F \leftarrow F \cup \{f_i\}$  ▷ Adds the sub-graph  $f_i$  to  $F$ 
7:   return  $F$ 

8: function JOIN( $W, w_i$ )
9:   for all  $w_j \in w_i$  do
10:     $m \leftarrow \text{JOIN}(W, w_j)$ 
11:     $w_i \leftarrow w_i \cup m$  ▷ Merges all nodes from  $w_j$  to  $w_i$ 
12:     $W \leftarrow W \setminus \{w_j\}$  ▷ Removes the subset  $w_j$  from  $W$ 
13:  return  $w_i$ 

```

The lightning database is composed of files in the UALF format. Each record/line corresponds to an individual flash/stroke following a chronological order and therefore the records are already sorted in time, reducing the computational cost of the above algorithm by half, since only subsequent flashes must be checked for a particular flash.

Once obtained the set of lightning clusters F , the next step is to check if there is a minimum rainfall R_k associated to each cluster f_k . The amount R_k is computed adding up the rainfall r_i associated to each node $v_i = (t_i, x_i)$ of the cluster from the weather radar images. In order to perform such computation, a region growth algorithm (GONZALEZ; RICHARD, 2002) is applied. Considering a lightning occurrence v_i , every neighboring pixel of the radar image that presents rainfall is used as a seed for the region growth algorithm. Once identified these rainfall pixels, the corresponding pixels are checked for rainfall in the preceding and the subsequent radar images. In the affirmative case, any pixel is also considered as a seed. In both cases, the neighborhood is given by a space-time window defined by $(\Delta t_G, \Delta x_G)$.

All the rainfall corresponding to the pixels determined by the region growth algorithm (in space and in time) is then computed for the lightning occurrence v_i . A two-step procedure is repeated alternating spatial growth and temporal growth until no more rainfall is added to v_i . The total amount of rainfall associated to a lightning

cluster f_k is then given by the summation of the rainfall of all nodes v_i that compose it. If this total is above a defined threshold, the lightning cluster is then considered a thunderstorm.

Finally, a consistency check is performed: if any two thunderstorms corresponding to lightning clusters f_a and f_b have overlapping rainfalls, i.e. $R_{f_a} \cap R_{f_b} \neq \emptyset$, they are merged into a new thunderstorm that corresponds to a new lightning cluster $f_c = f_a \cup f_b$, and has a total rainfall given by $R_{f_c} = R_{f_a} \cup R_{f_b}$.

4.2.1 Conversion of lightning stroke to lightning flash data

Despite the temporal sliding-window approach adopted in this work, the pioneering application of the Tapia's model to estimate a RLR for the Southeast Brazil (GARCIA et al., 2012) required a selection of individual storms that was performed by the proposed algorithm. However, once Tapia's model requires lightning flash data and RINDAT provides lightning stroke data, some pre-processing was required to convert strokes to flashes. A lightning flash is composed of one or more strokes that are usually less of 1 km apart and occur with a delay of few milliseconds. Goodman et al. (2012) presented a clustering scheme to derive flashes from stroke data, but the algorithm described in this section to perform the clustering of lightning was also employed to group strokes into flashes by a convenient adjustment of Δx_G and Δt_G . Here, these parameters were adopted as 16.5 km and 300 ms, respectively, as also suggested in Goodman et al. (2012) to identify two different strokes as part of the same flash. Therefore, lightning stroke data was converted to lightning flash data, obtaining an average multiplicity of 2.8 that is similar to the value presented in Pinto et al. (2003) for the same Brazilian region.

4.2.2 Influence of some parameters in the thunderstorm selection algorithm

The output of the algorithm proposed for thunderstorm selection depends on two parameters, Δx_G and Δt_G that were already described. These parameters must to be adjusted in such a way that the resulting thunderstorms be a good representation of the corresponding convective activity in the spatial scale convenient for the RLR calculation. However, in general, distinct thunderstorms should be detected as different events, while a single large storm should be detected as one single event. Another point is the different rainfall estimation errors are obtained for different spatial scales. Thus, an empirical adjustment of these parameters is required using weather radar images as a reference.

Preliminary tests showed that values of Δx_G lower than 5 km are bad choices since a single large thunderstorm will often be split into several smaller ones. Conversely, values of Δx_G higher than 30 km will cause distinct thunderstorms merged into a single one. A similar behavior is observed with the parameter Δt_G . Values lower than 2 min split a larger thunderstorm, while values higher than 20 min tend to merge several smaller thunderstorms. Four combination of values of Δx_G and Δt_G were tested considering two values of Δx_G (10 and 20 km) and two values of Δt_G (5 and 10 min). Table 4.1 shows the number of flashes per thunderstorm obtained for different values of Δt_G and Δx_G . The values $\Delta x_G = 10$ km and $\Delta t_G = 5$ min were adopted for the RLR calculation since they allow to identify thunderstorms that are more similar than those observed in the corresponding weather radar images.

Table 4.1 - Influence of the parameters Δx_G and Δt_G in the resulting thunderstorms identified by Algorithm 4.1

Test number	1	2	3	4
Δt_G	5 min	10 min	5 min	10 min
Δx_G	10 km	10 km	20 km	20 km
Maximum number of flashes per storm	796	4164	7410	8793
Average number of flashes per storm	91.18	121.04	144.67	176.84
Standard deviation	40.42	108.67	191.43	282.26
Longest storm duration	3.51 h	6.47 h	6.67 h	8.83 h
Average storm duration	0.16 h	0.26 h	0.19 h	0.35 h
Standard Deviation	0.20 h	0.35 h	0.28 h	0.56 h

4.3 Temporal sliding-window for lightning and weather radar data

The use of overlapping temporal sliding windows to compute the rainfall and the lightning data is the one of the cornerstones of the proposed approach to estimate convective rainfall from the number of lightning strokes using a WRLR function. Such computation smooths out the data and allow to estimate a suitable WRLR function. Additional smoothing is provided by the temporal overlapping of the windows, that precludes outliers from misleading the results. However, this overlapping is employed only in the training phase to compute data and estimate the function, not in the test phase.

The Bauru weather radar was chosen to generate the reference WRLR function, due to the availability of corresponding data and its temporal resolution. In Garcia et

al. (2013) both the Bauru and Presidente Prudente radars were taken as reference radars. In this work, data used to derive the WRLR function was computed by temporal sliding windows considering a square region of 200 km side centered at the Bauru radar, that is completely inside the 150 km radar range, as shown in Figure 4.4. However, this 200 km square area was divided into a set of smaller sampling squares. The temporal sliding window is then employed separately for each sampling square.

The training data set is from the period of January 2009 to September 2010. The corresponding convective rainfall was obtained from the CAPPI radar images using the criteria proposed by Steiner et al. (1995), described in Section 3.3, and the Z-R relationship presented in Section 2.3. The lightning stroke data was obtained from the RINDAT data files in UALF format, while the lightning flash data was derived from the stroke data by using a flash clustering algorithm, as proposed in Section 4.2.1.

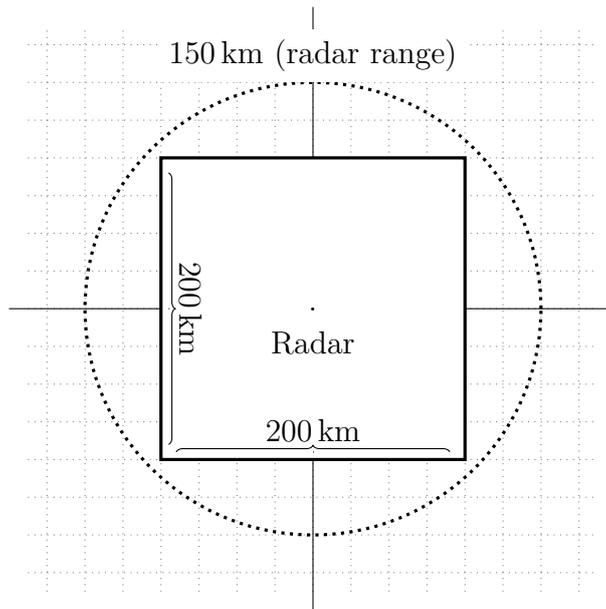


Figure 4.4 - Location of the squared area of 200 km side centered at the Bauru radar location (the circle shows the 150 km radar range).

The temporal sliding-window requires the definition of the window duration Δt and window advance δt . Thus $\delta t = 0$ means no overlap, and it is required that $\Delta t > \delta t$ to avoid gaps. The size of the equal-size sampling squares must also be defined. These choice of these three parameters is discussed in the following sections. The considered time extension is divided with Δt and δt values that are multiples of the

temporal resolution of the radar that is 7.5 min. Figure 4.5 shows a temporal sliding window advancing by δt assuming an hypothetical time unit.

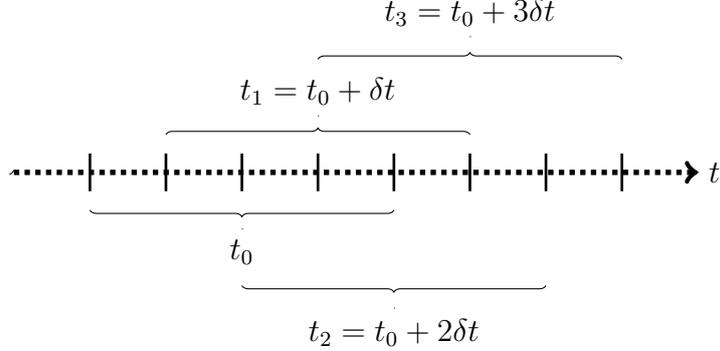


Figure 4.5 - Sliding time-window with size $\Delta t = 4$ units and advance $\delta t = 1$ unit.

The 200 km square area is divided into several sampling squares Q_j of side Δx . Since these squares do not overlap, Δx must subdivide the size of the square area into integer parts, i.e. 200 km must be multiple of Δx . However, in order to check if the position of this grid has any influence in the proposed computation (for instance, strong rainfall or lightning activity at the borders of a grid square), and also to increase the number of obtained windows to smooth-out local variations, the region study was re-sampled using different grids of the same dimensions, but displaced along the North-South direction, the East-West direction and even both directions. Figure 4.6 shows the sampling grid and re-sampling grid variations. These displacements are defined by a constant shift δx , with $\delta x < \Delta x$. The adopted values for Δx and δx are always multiple of the spatial resolution of the radar data that is 1 km.

The computation of rainfall and lightning data is performed by the application of the temporal sliding window to the sampling squares. Therefore, considering each square Q_j and each interval of time starting at t_i , it is possible to define a tuple (n_{ij}, r_{ij}) of the corresponding number of CG lightning strokes occurrences n_{ij} and convective rainfall r_{ij} .

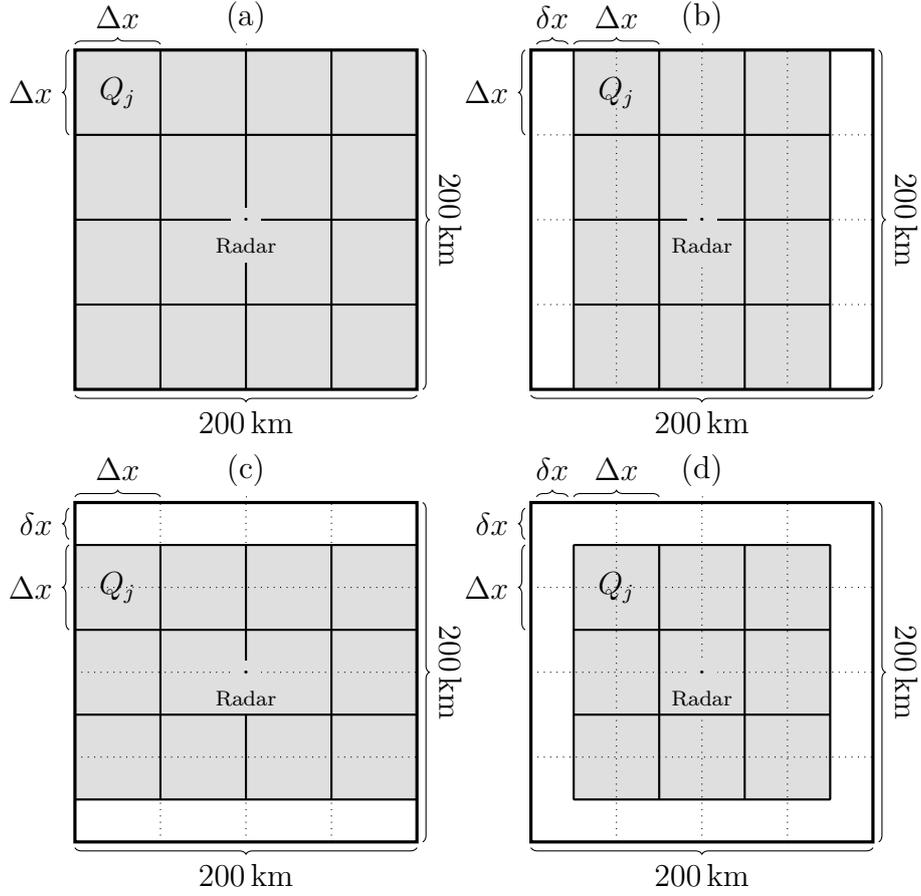


Figure 4.6 - The square area of side 200 km divided into (a) several Q_j sampling squares with no displacement, (b) with a latitudinal shift δx , (c) with a longitudinal shift δx , and (d) with both latitudinal and longitudinal shifts δx .

4.4 Estimation of the WRLR function

The computation of rainfall and lightning data described in the previous section allows to obtain a set of tuples. These tuples are used for the estimation of a WRLR function, which yields the amount of convective rainfall R (in 10^6 kg) from the number of CG lightning strokes N , i.e. $\text{WRLR}(N) = R$. In order to fit the function values to the set of tuples (n_{ij}, r_{ij}) , a power function seemed to be the most convenient choice,

$$\text{WRLR}(N) = a \cdot N^b + c, \quad (4.2)$$

where a , b and c are real constants, with $0 < b < 1$ and $N \in \mathbb{N}_{>0}$.

Once a function is estimated from the set of tuples, it is possible to calculate the

corresponding rainfall estimation error. A number of tests were performed with different values for the temporal window duration Δt and for the side Δx of the sampling squares, resulting in different sets of tuples (n_{ij}, r_{ij}) and the corresponding WRLR functions. In general, higher values of Δt and Δx result in tuples with higher rainfall and higher number of strokes, but tend to excessively smooth out the data masking eventual local variations. On the other hand, lower values of Δt and Δx imply in tuples with higher variability. The choice of the parameters Δt and Δx is discussed as follows. A set of feasible values for Δt and Δx were found, in a sense that rainfall estimation errors were acceptable, but it was not possible to determine optimal values for them.

The vast majority of the tuples presented no rainfall or lightning since thunderstorm occurrence is very sparse in time and space. CG lightning and convective rainfall do not necessarily match in space and time, even considering that the corresponding data is computed using a temporal sliding window for each sampling square Q_j . Besides this eventual mismatch, lightning or rainfall may occur outside the considered Q_j , but in a nearby point. These occurrences near the border are treated by the (spatial) re-sampling, also described in the preceding section.

In order to compute the WRLR (N) function, tuples with $n_{ij} = 0$ (no lightning) are discarded, even if presenting rainfall $r_{ij} \neq 0$. In the same way, tuples with $r_{ij} = 0$ (no rainfall) are also discarded, even presenting lightning $n_{ij} \neq 0$. There is also the presence of outliers, which are tuples with extreme or inconsistent values when compared to the rest of the tuples, i.e. very high values of n_{ij} for low values of r_{ij} or vice-versa. These outliers were removed using the Tukey-Kramer method (TUKEY, 1977), as described in Equation 4.3: A given tuple with $n_{ij} = N$ strokes can be written as $w_{ij} = (n_{ij}, r_{ij}) = (N, r_{Nij})$ is checked against all other tuples with the same number of strokes. It is considered as an outlier and removed from the dataset if it does not comply to the following condition:

$$Q_1 - 1.5 \cdot IQR \leq r_{Nij} \leq Q_3 + 1.5 \cdot IQR, \quad (4.3)$$

where,

- W_N : the set of all tuples for a given by $(n_{ij} = N, r_{ij})$,
- r_{Nij} : rainfall of a given tuple w_{ij} with $n_{ij} = N$ strokes,
- F_N : current subset of W_N without outliers,
- Q_1 : the lower rainfall quartile of the r_{Nij} of the tuples of W_N ,
- Q_3 : the upper rainfall quartile of the r_{Nij} of the tuples of W_N ,
- IQR : the inter-quartile range, defined by $(Q_3 - Q_1)$.

As already mentioned, the high variability of the sets of tuples (n_{ij}, r_{ij}) makes difficult to estimate a suitable WRLR function. A scatter plot of the tuples was made to check if the tuples present some tendency. For each N , an average rainfall value r_N was calculated by the average of the rainfall values of all tuples with $n_{ij} = N$. The scatter plot is shown along with these averages in Figure 4.7. These averages form a curve up to $N = 100$ approximately. On the other hand, the scatter plot shows less tuples for increasing values of N . It can be concluded that this poor sampling for high values of N precludes the estimation of a function that fits the tuples with low errors. Therefore, the value $N = 100$ was then chosen as a limit for the domain of the WRLR function: $[1, 100]$. It is important to note that lightning stroke data were employed as provided by RINDAT, but some tests were performed using lightning flash data, obtained as described in Section 4.2.1. The next section describes the optimization of parameters Δt and Δx .

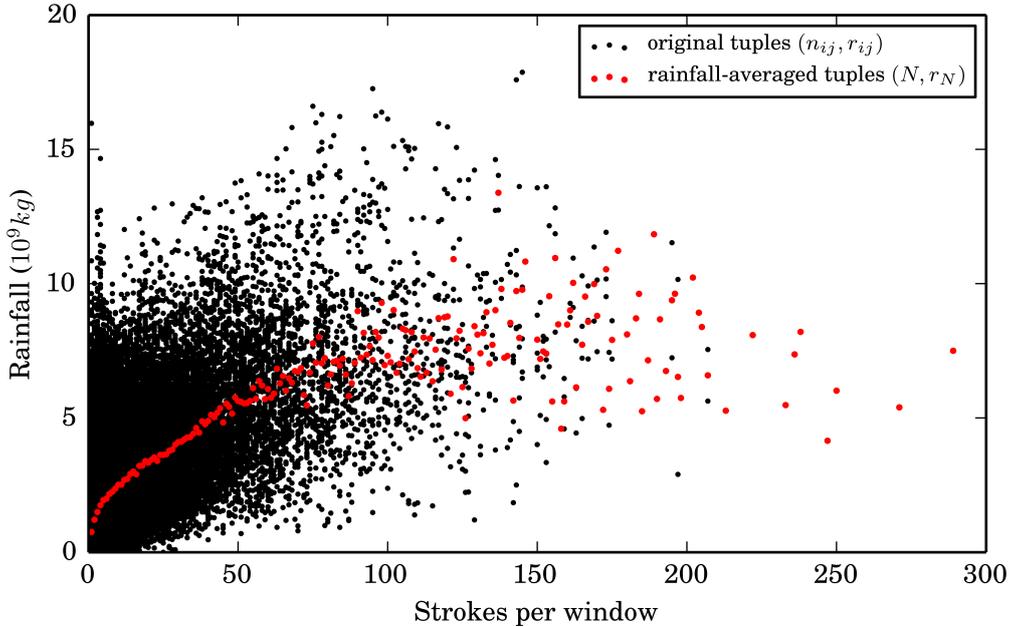


Figure 4.7 - The scatter-plot of (n_{ij}, r_{ij}) tuples (in black) and of the rainfall-averaged (N, r_N) tuples (in red).

The WRLR function was obtained using a particular non-linear least squares fitting, that fits a set of m observations with a model that is non-linear in n unknown parameters ($m > n$). The basis of such method is to approximate the model by a linear one and to refine the parameters by successive iterations. The adopted computational routine was the *curve_fit* function from the SciPy package for Python (MCKINNEY, 2012).

Some values of N are related to more tuples ($n_{ij} = N, r_{ij}$) than others, and these more frequent N values are related to a more accurate average rainfall r_N , as a larger number of samples tend to more closely represent the behavior of a phenomena (MEYER, 2000). Thus, more frequent values of N are more important to the WRLR than less frequent ones, as those in the latter could jeopardize the fitting of the function due to poorer sampling. To avoid this problem, weighted least squares are used in the fitting of the WRLR. The *curve_fit* routine employs weights. The main advantage that weighted least squares enjoys over other methods is the ability to handle regression situations in which the data points are of varying quality (RYAN, 2008). Less weight is given to the less frequent values of N , which are considered less accurate, while more weight is given to more frequent values of N . This means the rainfall r_N for more frequent values of N will be more meaningful for obtaining the WRLR function, as the function will fit those points more closely, which will also limit the impact of under sampled values. The chosen weight is the frequency of quantity of tuples ($n_{ij} = N, r_{ij}$) for a given N .

4.4.1 Optimization of the parameters for the sliding-window approach

As commented above, the sliding-window for the computation of rainfall and lightning data requires the definition of the parameters Δt (window duration) and δt (window advance). Values that are multiple of the radar resolution of 7.5 min were tried for Δt : 7.5, 15, 22.5, 30, 45, 60, 120 and 360 min, while δt was selected as $\delta t = 7.5$ min to increase the number of windows and to smooth-out local effects. Additionally, the edge Δx of the sampling squares and the spatial re-sampling displacement δx must also be defined. Test values of Δx were of 10, 20, 25, 40, 50, 100 and 200 km, while δx was selected as $\Delta x/2$ for each test case.

Preliminary tests provided some guidance in choosing values for the parameters Δt and Δx . Higher values of these parameters tend to merge different thunderstorms. Additionally, “clear weather” areas and time intervals are mixed up with “stormy weather” ones. Such local variations imply in estimating a WRLR function prone to errors. For example, the tuples obtained assuming $\Delta x > 200$ km and $\Delta t > 360$ min

imply in WRLR functions that do vary too much within N and have estimation errors similar to the adoption of a constant-valued RLR. On the contrary, the adoption of low values for these parameters reflect the temporal and spatial mismatching between the occurrence of CG lightning and convective rainfall, since lightning and rainfall data are not smoothed out sufficiently and the consequent high variability of the tuples preclude the estimation of a suitable WRLR function.

As already mentioned, it would be difficult to obtain optimal values for Δt and Δx or to assure that such values are optimal. The performed tests allowed to optimize these values up to a certain degree, according to the test estimation error, and to adopt a pair of values that is convenient considering the temporal and spatial resolution of the data. In this way, this work adopted $\Delta x = 50$ km and $\Delta t = 30$ min. The corresponding values of $\delta x = 25$ km and $\delta t = 7.5$ min were then defined.

An estimation test was performed for a WRLR function derived using the same parameters, but assuming $\delta x = 10$ km. The estimation error was similar, but processing time increased too much, in proportion to the number of tuples.

4.4.2 Mapping of the WRLR function different sampling area sizes and window durations

The adopted value for the sampling square area size is $\Delta x = 50$ km, as described in the previous section. However, the several tests performed using different values of Δx show that is possible to estimate an linear relationship between two WRLR functions obtained with different values of Δx , considering a range from 10 to 100 km. This relationship is approximated by the root square of the ratio between the corresponding square areas, as shown in Equation 4.4. In this case, this relationship corresponds to the ratio between both values of Δx . However, for convenience, the root square of the ratio between the corresponding square areas is adopted, since there are cases with trapezoidal areas. The mapping presented in this equation considers always a window duration $\Delta t_{ref} = 30$ min.

$$\text{WRLR}_{tgt} \approx X_{tgt} \cdot \text{WRLR}_{ref}, \quad (4.4)$$

where,

- WRLR_{tgt}: target WRLR function,
- WRLR_{ref}: reference WRLR function,
- A_{Δx_{tgt}}: area for WRLR_{tgt} (km²),
- A_{Δx_{ref}}: reference square area of 50 km edge (2500 km²),
- X_{tgt}: target spatial adjust factor, given by:

$$X_{tgt} = \sqrt{\frac{A_{\Delta x_{tgt}}}{A_{\Delta x_{ref}}}}. \quad (4.5)$$

Different WRLR functions were computed using temporal windows of different sizes, with Δx of 10, 20, 25, 40, 50 (reference) and 100 km, all with the same duration $\Delta t = 30$ min. A least square approach that minimizes the errors was then employed to derive the corresponding “true” spatial target adjust factors that allow to map the reference WRLR to each one of the WRLR functions with different size.

Figure 4.8 presents these “true” six spatial adjust factors compared to the proposed X_{tgt} that are along the diagonal line. Values of X_{tgt} vary linearly with Δx , being similar to the “true” adjust factors for each Δx , except for the case of $\Delta x = 100$ km.

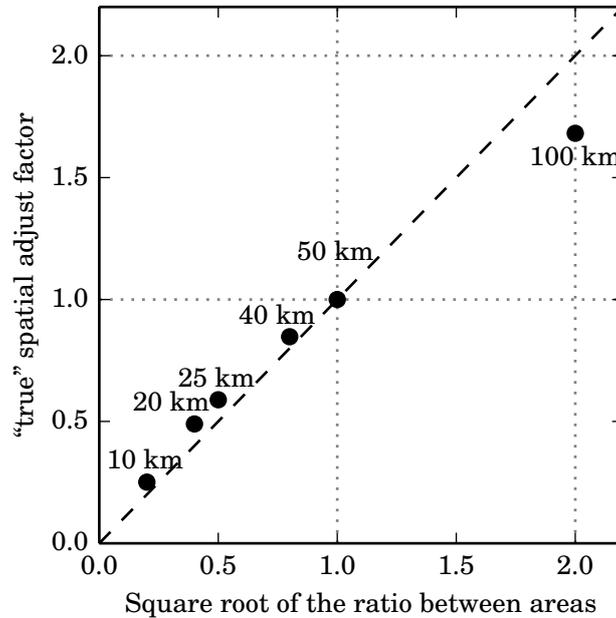


Figure 4.8 - True target spatial adjust factors (black dots) compared to the proposed adjust factors given by the square of the ratio between areas (dashed line).

The target spatial adjust factor X_{tgt} results in a new WRLR function that is very similar to the WRLR function derived directly from the tuples with different window size. An example is given in Figure 4.9 by comparing the WRLR function derived from tuples with window edge of $\Delta x_{tgt} = 20$ km to the function given by $X_{tgt} \cdot WRLR_{ref}$. It can be observed that both functions are very similar. Function $WRLR_{ref}$ is also shown in the figure. Functions $WRLR_{ref}$ and $WRLR_{tgt}$ were derived using $\Delta t_{ref} = 30$ min.

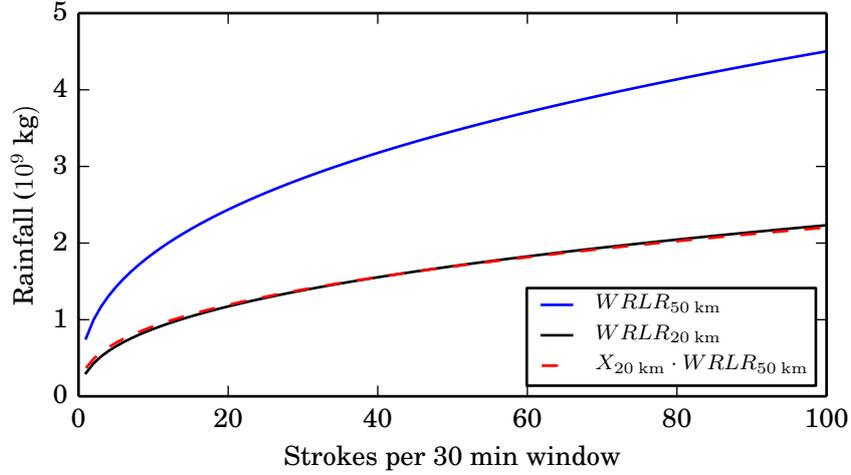


Figure 4.9 - The functions derived from the tuples are shown in solid lines $WRLR_{20 \text{ km}}$ (black) and $WRLR_{50 \text{ km}}$ (blue), while the approximated function $[X_{20 \text{ km}} \cdot WRLR_{50 \text{ km}}]$ is shown by the red dashed line.

The WRLR function was derived using data obtained with a sliding window of $\Delta x_{ref} = 50$ km and $\Delta t_{ref} = 30$ min. Similarly to the mapping presented above for different sizes of the sampling square area, it is also possible to linearly map the WRLR function for different window durations Δt , using 7.5, 15 and 60 min, as expressed in Equation 4.6, but maintaining $\Delta x = 50$ km.

$$WRLR_{tgt} \approx T_{tgt} \cdot WRLR_{ref}, \quad (4.6)$$

where,

WRLR: target WRLR function,
 WRLR_{ref}: reference WRLR function,
 Δt_{tgt} : target window duration (min),
 Δt_{ref} : reference window duration (30 min),
 T_{tgt} : target temporal adjust factor, given by:

$$T_{tgt} = \frac{\Delta t_{tgt}}{\Delta t_{ref}}. \quad (4.7)$$

Different WRLR functions were computed for the same window size, but for different window durations Δt of 7.5, 15, 30 (reference) and 60 min. Again, a least square approach that minimizes the errors was then employed to derive the corresponding “true” temporal target adjust factors that allow to map the reference WRLR to each one of the WRLR functions with different duration.

Figure 4.10 presents these “true” four temporal adjust factors compared to the proposed T_{tgt} that are along the diagonal line. Values of T_{tgt} vary linearly with Δt , being similar to the “true” adjust factors for each Δt , except for the case of $\Delta t = 60$ min.

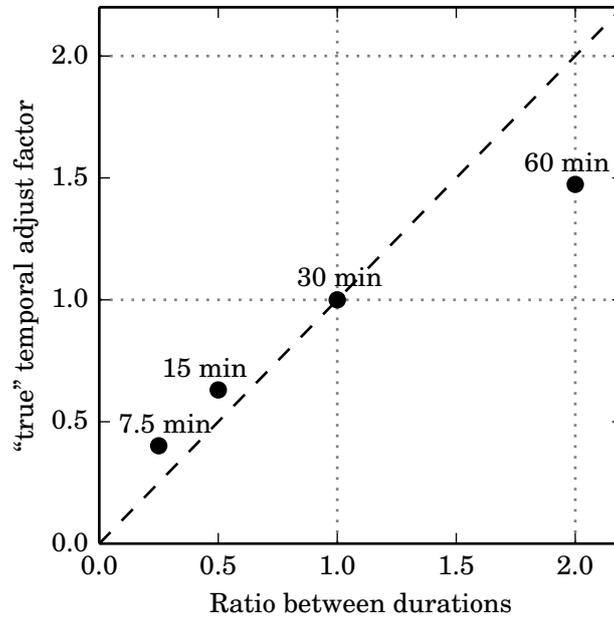


Figure 4.10 - True target temporal adjust factors (black dots) compared to the proposed adjust factors given by the ratio between durations (dashed line).

The target temporal adjust factor T_{tgt} results in a new WRLR function that is very similar to the WRLR function derived directly from the tuples with a different window duration. An example is given in Figure 4.11 by comparing the WRLR function derived from tuples with window duration $\Delta t_{tgt} = 15$ min to the function given by $T \cdot WRLR_{ref}$. It can be observed that both functions are very similar. Function $WRLR_{ref}$ is also shown in the figure. Functions $WRLR_{ref}$ and $WRLR_{tgt}$ were derived using $\Delta x_{ref} = 50$ km.

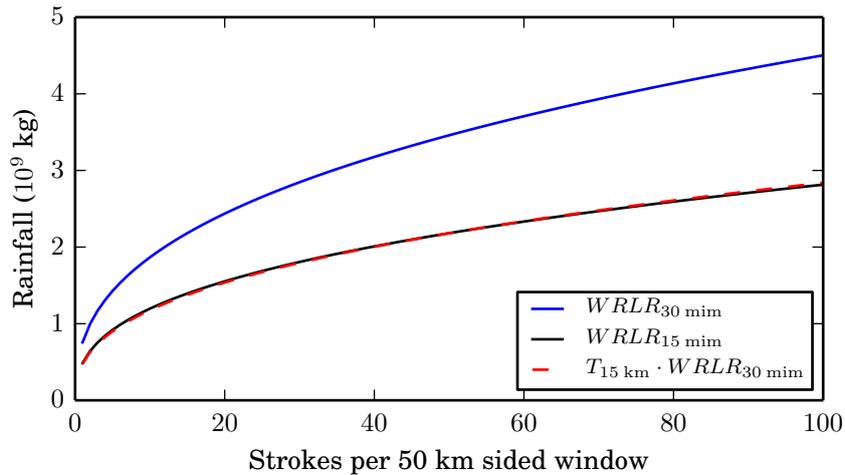


Figure 4.11 - The functions derived from the tuples are shown in solid lines $WRLR_{15 \text{ min}}$ (black) and $WRLR_{30 \text{ min}}$ (blue), while the approximated function $[T_{15 \text{ min}} \cdot WRLR_{30 \text{ min}}]$ is shown by the red dashed line.

4.5 Mapping of the WRLR function for areas outside the reference radar range

A cornerstone of the proposed rainfall estimation WRLR function is how to map the reference WRLR function derived using lightning and rainfall data for a specific reference radar (in this case, the Bauru radar) for other areas of the Brazilian territory. Such areas are typically outside the coverage of any radar, but must be within range of the RINDAT lightning detection network. The works that employed a constant-valued RLR, based on the Tapia's model (TAPIA et al., 1998), presented values of RLR that may heavily depend on the location, like Williams et al. (1992), Molinie et al. (1999) or Kempf and Krider (2003). Therefore, a regional variation of the WRLR function was also expected. Besides the Bauru weather radar, data from three other radars were available for this study (Presidente Prudente, São Roque and Pico do Couto radars). It was possible to derive the WRLR function using the

Bauru radar as reference and to test it in areas covered by the remaining radars. Such tests confirmed that the WRLR function must be somehow adjusted to be employed in other areas.

A specific grid with resolution 0.5° was defined for a rectangular area bounded by latitudes $12^\circ 00' 00''\text{S}$ to $27^\circ 00' 00''\text{S}$, and longitudes $58^\circ 00' 00''\text{W}$ to $40^\circ 00' 00''\text{S}$. This resolution is convenient for the EDDA-chuva software, described in the next section. It approximately corresponds to the coverage of the RINDAT lightning detector network. Grid point (m, n) is associated to a square area with 0.5° edge centered at this point. A quarterly-averaged rainfall-to-lightning ratio QRLR was calculated for each season of the year and each grid point (m, n) . Quarterly amounts of rainfall were obtained from the isohyets described in Section 2.5, while the quarterly accumulated number of CG lightning strokes, from the RINDAT data described in Section 4.4. Reference quarterly-averages $QRLR_{ref}$ were also computed for each season considering the square area of 200 km edge shown in Figure 4.4 that is centered at the reference weather radar of Bauru.

Assuming, for simplicity, that the reference WRLR function (Equation 4.2) can be linearly mapped to other grid points by a single coefficient, its value is proposed as the ratio between quarterly-averaged rainfall values at the desired location and at the location of the reference radar, as expressed by Equation 4.9. The new function is limited to values $N < 100$ to avoid high estimation errors. Function values are also given in 10^6 kg.

$$\text{WRLR}_{mn}(N) \approx \left(\frac{QRLR_{mn}}{QRLR_{ref}} \right)^b \cdot X_{mn} \cdot \text{WRLR}_{ref}(N), \quad (4.8)$$

where,

- WRLR_{mn} : WRLR function adjusted for grid point (m, n) ,
- WRLR_{ref} : reference WRLR function,
- $QRLR_{mn}$: quarterly-averaged RLR at grid point (m, n) ,
- $QRLR_{ref}$: reference quarterly-averaged RLR,
- b : power factor for WRLR function (Equation 4.2),
- X_{mn} : spatial adjust factor (Equation 4.4) from 50 km to 0.5° .

The spatial adjust factor $X_{0.5^\circ}$ allows to map the reference WRLR function obtained using 50 km edge squares to the 0.5° grid resolution that is equal to 55.56 km, if not considered the latitude-distortion. Along a circle of (constant) latitude, i.e. an Earth's parallel, 0.5° is equivalent to a distance that varies from 46.14 to 55.56 km

for the studied area due to the Earth’s shape. Therefore, the square area associated to each point of the 0.5° resolution grid is actually a trapezoidal area.

In order to simplify the computation, a matrix $K = \{k_{mn}\}$ can be defined to embed the ratio of the quarterly-averaged RLR’s and the spatial adjust factor. K elements are dimensionless. Thus, Equation 4.9 can be rewritten in a simpler form as:

$$\text{WRLR}_{mn}(N) \approx K_{mn} \cdot \text{WRLR}_{ref}(N), \quad (4.9)$$

where,

$$K_{mn} = \left(\frac{QRLR_{mn}}{QRLR_{ref}} \right)^b \cdot X_{mn}$$

K can be defined as the matrix of remapping factors for the reference WRLR function. The precomputed values of the elements of K are shown for each season and for the considered grid in Figure 4.12.

The K matrix was obtained using a relatively low number of years for lightning activity (2007 to 2010). This poor sampling resulted in some grid points that presented a quarterly amount of rainfall, but low occurrence or absence of CG lightning, with the related K_{mn} values being very high. In order to filter out such values considered as outliers, the upper limit of the K elements was set as 5.0. This limit was chosen based on the highest values of RLR compiled in Petersen and Rutledge (1998). It is also important to notice that a K_{mn} is only valid for regions above the land, as there is no isohyetal data available corresponding to rainfall over the ocean. For simplification, K_{mn} values corresponding to $0.5^\circ \times 0.5^\circ$ trapezoidal areas that are partially over the ocean are computed considering only lightning occurrences and rainfall over land.

The isohyetal maps of rainfall presented in Section 2.5 refer to the total rainfall, but the WRLR function only estimates convective rainfall. In this work, it is assumed a constant ratio p between the convective and the total rainfall for all points of the considered grid. Using this simplification, the term p is canceled in the fraction between $QRLR_{mn}$ and $QRLR_{ref}$

Henceforth, the function WRLR multiplied by the specific K element will be referred as “mapped WRLR”.

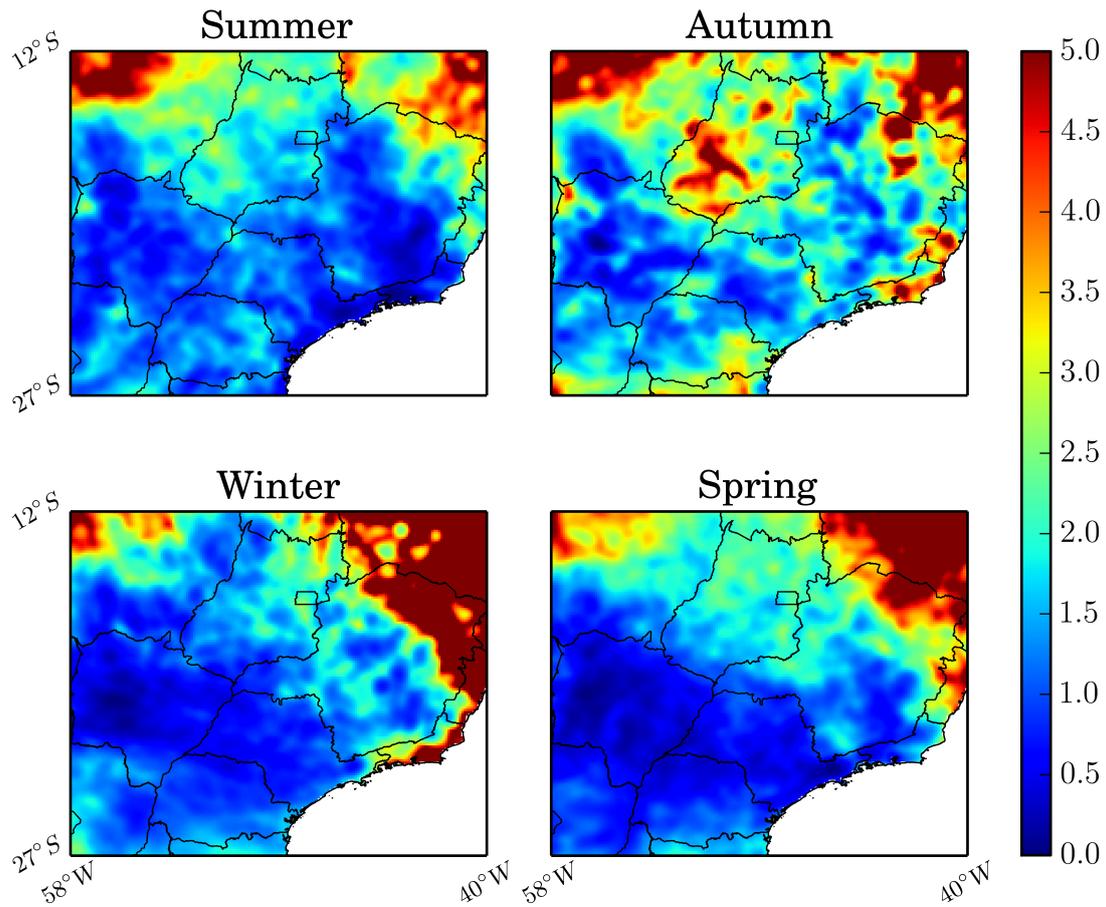


Figure 4.12 - Values of the elements of the K remapping matrix for the four seasons and considered grid.

4.6 Inference of the total rainfall

The WRLR function and its mapping as presented in the previous sections applies for the estimation of the convective rainfall and, furthermore, only for the convective rainfall associated to lightning occurrence. Meteorologists are typically interested in the estimation of the total rainfall, i.e. convective plus stratiform rainfall. Some works addressed this issue considering TRMM rainfall estimates, that are performed by its precipitation radar. Schumacher and Houze (2003) and Takayabu (2003) investigated characteristics of convective and stratiform rainfall in the tropics. More recently, Liu et al. (2012) analysed climatic characteristics of convective and stratiform rainfall over tropical and subtropical areas based on TRMM data. In particular, Schumacher and Houze (2003) estimated the convective fraction of the rainfall for different seasons, regions and synoptic regimes. Such estimation is restricted up to latitudes 20°S and 20°N and employs a grid with poor resolution. Average convective

fraction is 60% in mass, and 30% in area (SCHUMACHER; HOUZE, 2003). In particular, the convective fraction in mass is about 43% in mass for the Summer months. Typically, for convective events, the stratiform part occurs after the convective one and tends to be less intense.

This work assumes a fraction of convective rainfall of 50% in mass, due to the lack of better estimates. Considering the spatial distribution of rainfall, it assumes that the convective fraction covers 30% of the area of the total rainfall. Some tests employing weather radar data are discussed in the next chapter and confirm the suitability of the 50% fraction in mass: in average, one half of the precipitated mass is due to convective rainfall with lightning, while the other half corresponds to convective rainfall without lightning and stratiform rainfall. It is worth to stress that only the stratiform rainfall associated to convective events is included here.

4.7 EDDA-chuva software

The research described here allowed to add a new functionality to the software EDDA, described in Section 3.2. Besides the generation of fields of density of occurrence of lightning for the near real time monitoring of convective activity, it would be possible to estimate the amount of convective rainfall accumulated in a given interval of time using lightning data. Therefore, a new software was developed, the EDDA-Chuva (“chuva” stands for rainfall, in Portuguese), with the same functionalities of the former software EDDA, but including the convective rainfall estimation.

In the new software, that started to be evaluated operationally at CEMADEN in January, 2014, lightning stroke occurrences are stored every 30 min and distributed over a grid of 0.5° resolution that covers all the part of Brazil with RINDAT coverage, and even the larger coverage of the new BRASILDAT lightning detector network.

The part of the software that performs standard density estimation employs the same configuration parameters of the previous version (EDDA software). Such parameters include grid resolution of 0.01° and interval of time of 15 min for the integration of lightning occurrences, values that were defined as convenient for near real time monitoring of convective weather. On the other hand, the part of the software that performs rainfall estimation integrates CG lightning strokes for the 0.5° grid every 30 min, obtaining a number of occurrences N_{mn} that is mapped by the $WRLR_{mn}$ function to an amount of rainfall R_{mn} . The software outputs hourly-accumulated amounts of convective rainfall for the 0.5° grid. These outputs are further employed for the generation of 6, 12, 24, 48 and 72 h accumulated values of convective rainfall

for visualization in the SALVAR environment of CEMADEN.

Another point in the proposed methodology is the spatial rainfall distribution. This work proposes the estimation of the convective rainfall amount as given by the WRLR function (Equation 4.9), while the corresponding spatial distribution is assumed as being the same of the field of density of occurrence of CG lightning strokes, as proposed for the EDDA software (STRAUSS et al., 2013). Such density is obtained using standard Gaussian kernel density estimation for the considered area and time interval, as described in Section 3.2. Therefore, the normalized density of CG lightning stroke occurrences is mapped to the corresponding density of precipitated mass by multiplying by the accumulated mass of rainfall yielded by the WRLR function for the area around the considered point of the 0.5° grid and for the considered 30 min window duration.

However this spatial rainfall distribution is output for a 0.05° grid. The 0.01° resolution of the lightning density grid would be too refined for the rainfall distribution since there is a spatial mismatch between lightning and rainfall – the temporal mismatch can be partially attenuated considering the 30 min integration of lightning and rainfall data. A constant-valued smoothing window width h (Section 3.2) is also assumed as 0.05° for the convective rainfall.

Tapia’s model spatial distribution of the rainfall (Equation 3.1) is based on circles of constant rainfall intensity, resulting in a rainfall distribution with rough transitions where rainfall intensity level changes. Such rainfall distribution is rough and oversimplified. In the case of EDDA-chuva, the use of Gaussian kernel density estimation allows for smoother transitions, resulting in spatial distributions more similar to those observed from weather radars. Thus, EDDA-chuva produces a more realistic distribution than Tapia’s model. Figure 4.13 illustrates this difference, showing the estimated rainfall distribution given from the two approaches for the same two lightning occurrences.

Besides the estimation of convective rainfall, the EDDA software also estimates the total rainfall, assuming that half the precipitated mass is due to the convective rainfall with lightning, while the other half, to the convective precipitation without lightning and also the stratiform precipitation, as discussed in the previous section. Hourly-accumulated amounts of total rainfall are also output for the 0.05° grid. The spatial distribution of the total rainfall employs the same 0.05° grid of the convective rainfall, but with a value of window width h that results in an area that is about 3.33 times larger than the area covered by convective rainfall.

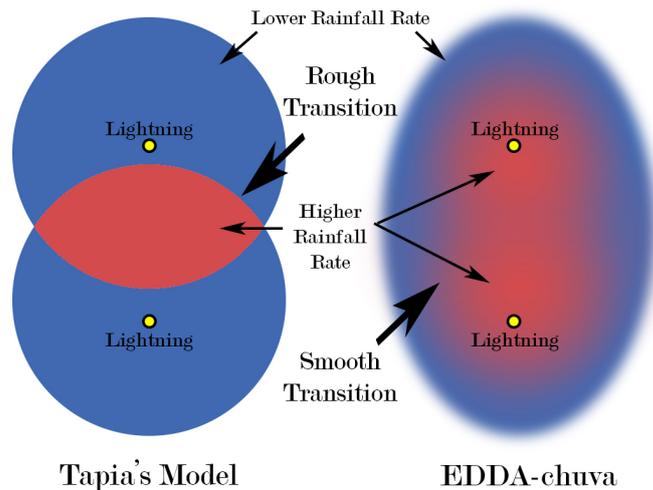


Figure 4.13 - Schematic view of the spatial distribution of rainfall for Tapia's Model (left) and EDDA-chuva (right). The rainfall is showed in low intensity (blue) and high intensity (red), for the lightning occurrences (yellow).

Initially, rainfall spatial and temporal distributions were obtained using a Gaussian filter applied on the CG lightning stroke occurrences weighted by the distance. This approach was abandoned since the EDDA software is more user friendly and accurate than the application of the Gaussian filtering. Besides, EDDA takes into account temporal variations since it integrates the occurrences for a time interval. Therefore, EDDA provided images of density of occurrence of CG lightning stroke that correlate better to the corresponding precipitating systems observed in the weather radar images.

5 VALIDATION TESTS

This chapter presents the results of the tests performed using the methodologies shown in the preceding chapters. The estimation of a RLR value for Bauru by the application of the Tapia’s model is described in the next section, followed by the major results of this work, the estimation of a WRLR function, its mapping for areas outside the reference radar range, and the test results concerning estimation of rainfall from lightning data.

All algorithms and methods developed in this work have been implemented using Python 2.7 on an SciPy based IPython Notebook environment. Python is a general-purpose high-level programming language that emphasizes code readability, and which syntax allows programmers to express concepts in fewer lines of code than would be possible in languages such as C, and supports multiple programming paradigms, including object-oriented, imperative, and functional programming or procedural styles (ROSSUM; DRAKE, 2011). This environment includes a set of scientific computing libraries that are part of the SciPy package (MCKINNEY, 2012), like:

SciPy a library of algorithms and mathematical tools containing modules for optimization, linear algebra, integration, interpolation, special functions, FFT, signal and image processing, ODE solvers and other algorithms common in science and engineering;

NumPy an extension to the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large library of high-level mathematical functions to operate on these arrays;

matplotlib a plotting library containing a procedural *pylab* interface based on a state machine, designed to closely resemble that of MATLAB;

IPython a command shell for interactive computing in multiple programming languages, especially focused on the Python programming language, that offers enhanced introspection, “rich” media, additional shell syntax, tab completion, and “rich” history;

Pandas a software library for data manipulation and analysis, offering data structures and operations for manipulating numerical tables and time series.

All tests were performed in a Supermicro server with two quad-core processors Intel® Xeon® CPU E5530 of 2.40 GHz.

5.1 RLR computation for Bauru

This section presents the standard RLR calculation for an area around the weather radar of Bauru using the methodology proposed by [Tapia et al. \(1998\)](#). This was the first RLR calculation for the Southeast Brazil and was performed using data of the year of 2009 for the square area of side 200 km inside the range of the weather radar of Bauru that was defined in [Figure 4.4](#). A set of 491 thunderstorms was selected using the algorithm described in [Section 4.2.2](#), with $\Delta t_G = 5$ min and $\Delta x_G = 10$ km. The total convective rainfall mass and the accumulated number of CG flashes were accounted for all thunderstorms.

Following the [Tapia's](#) methodology, RLR values were calculated for each thunderstorm of the set, resulting in RLR values ranging from 0.4×10^6 kg to 1094×10^6 kg per flash. The overall RLR is given by the median of these ratios, 219×10^6 kg per flash. [Figure 5.1](#) shows the scatter plot for the set of thunderstorms. Each point corresponds to the pair convective rainfall mass and total number of CG flashes of a given thunderstorm, i.e. to its RLR. A dashed line shows the estimation provided by the overall RLR, which is the median of the individual RLRs. It can be seen that the individual RLRs are hardly approximated by the overall RLR making difficult to obtain a good estimation of rainfall from the corresponding number of CG flashes. The value of the overall RLR is certainly influenced by the many thunderstorms with low rainfall and low number of CG flashes. The resulting overall RLR seems to overestimate the amount of rainfall for thunderstorms with high number of CG flashes.

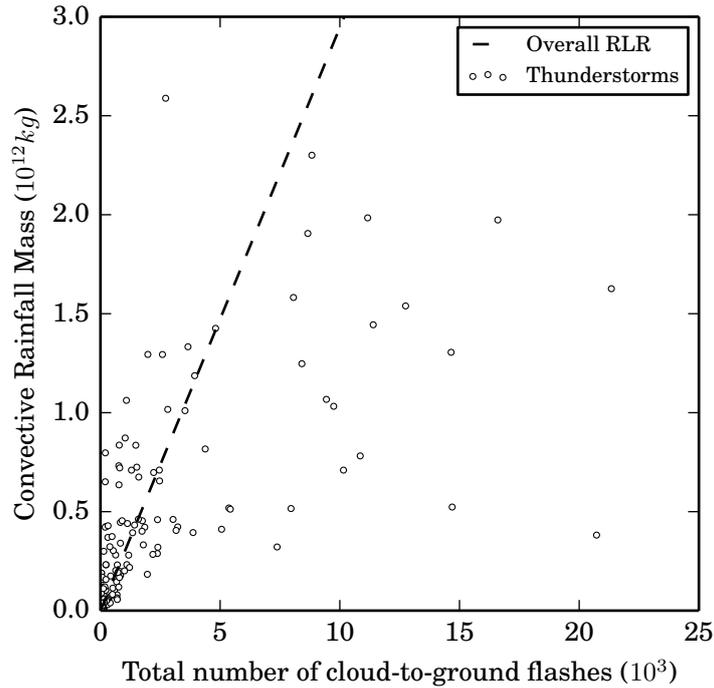


Figure 5.1 - Scatter plot of the convective rainfall mass and the number of CG flashes. Each point corresponds to the individual RLR of given thunderstorm, while the dashed line corresponds to overall RLR.

5.2 WRLR function estimation for the four seasons of the year

The proposed methodology for estimating the WRLR function was described in Sections 4.3 and 4.4. This section presents the estimation of WRLR functions for the four seasons of the year, obtained using data of the corresponding quarters of the year. These functions are reference WRLR functions estimated for the 200 km-edge square area centered at the Bauru weather radar (Figure 4.4). The tuples (n_{ij}, r_{ij}) of number of CG strokes and convective rainfall mass (in kg) were derived using sliding-windows of duration $\Delta t = 30$ min over sampling squares with edge $\Delta x = 50$ km. The temporal sliding-window advances $\delta t = 7.5$ min each time, i.e. an overlap of 22.5 min. The spatial re-sampling employs a displacement $\delta x = 25$ km.

Lightning and radar data spans from January 2009 to September 2010, divided by season/quarter, covering two Summers (January–March of 2009 and 2010), two Autumns (April–June of 2009 and 2010), two Winters (July–September of 2009 and 2010) and one Spring (October–December of 2009). However, only sampling squares that presented rainfall were considered, since there is no generation of radar data in the absence of rainfall. The same way, only sampling squares that presented lightning

strokes were considered. A total of 147,812 tuples resulted, being 79,906 for Summer, 10,339 for Autumn, 32,846 for Winter, and 24,721 for Spring. Figure 5.2 shows a scatter plot of the tuples for each season.

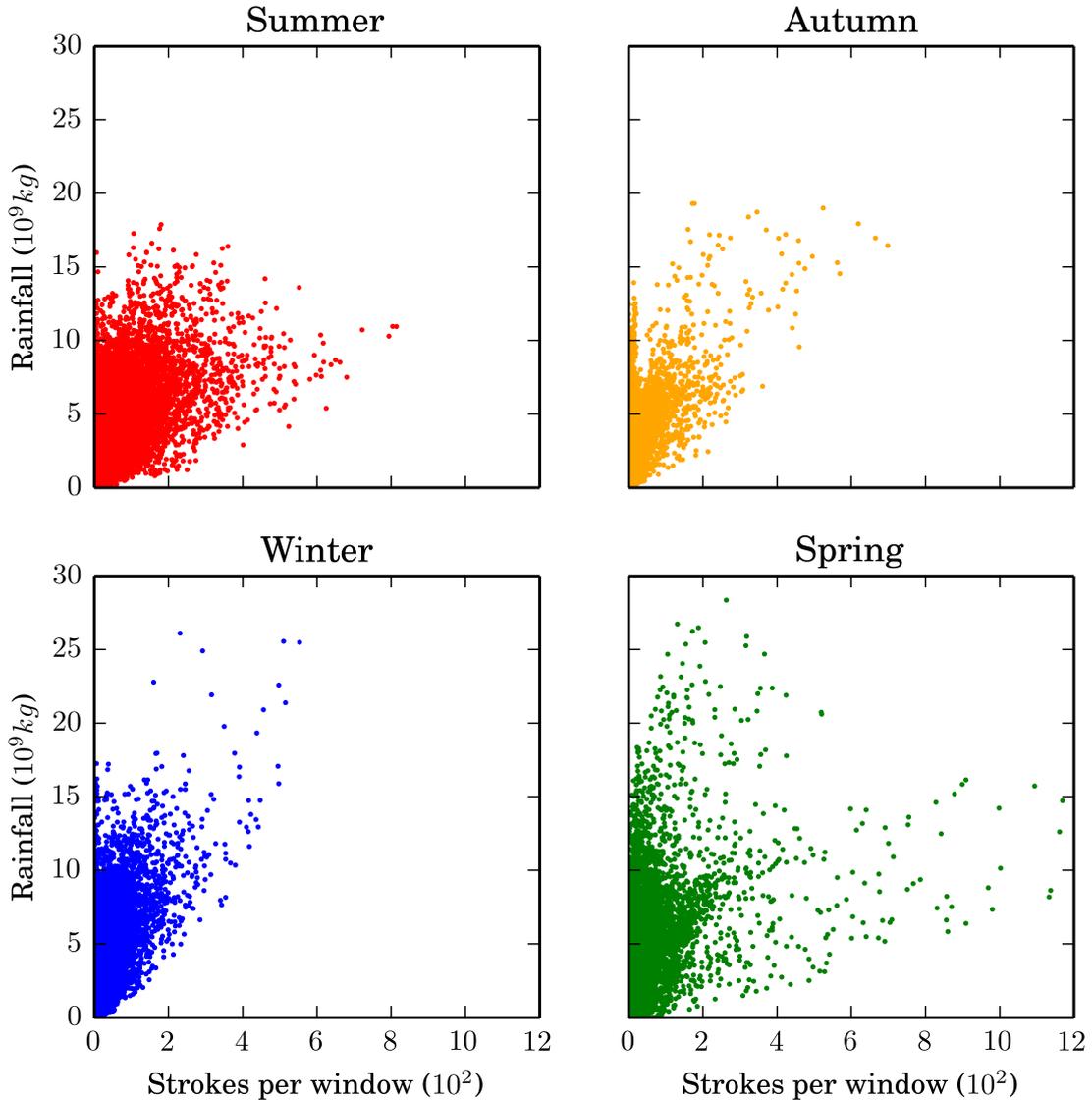


Figure 5.2 - Scatter plot of the tuples of convective rainfall mass and number of CG strokes.

The scatter plots above show the high variability of the tuples. Over 96% of these tuples presented a number of strokes $N \leq 100$ and the majority of the tuples present low number of strokes. On the other hand, tuples with high number of strokes are less common. Figure 5.3 shows the histogram for the quantity of tuples for a given value of N .

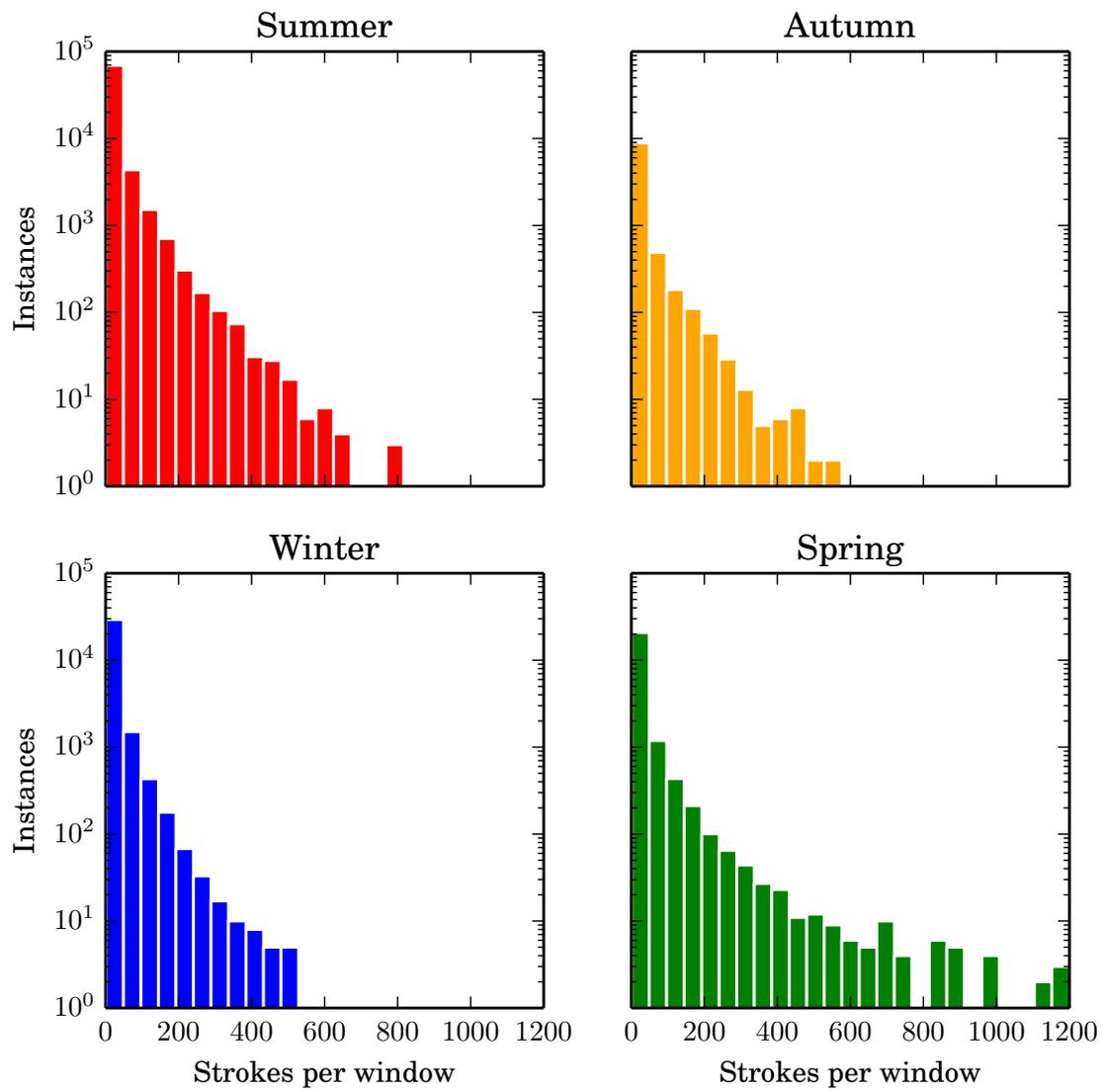


Figure 5.3 - Histogram for the number of tuples with a given number N of CG strokes.

In a first step, outliers were filtered out for each value of N using the Tukey-Kramer rule (TUKEY, 1977), described in Equation 4.3. Figure 5.4 shows the result of the hexagonal binning applied for the same tuples for each season. Darker hexagons denote tuples that occur with higher frequency.

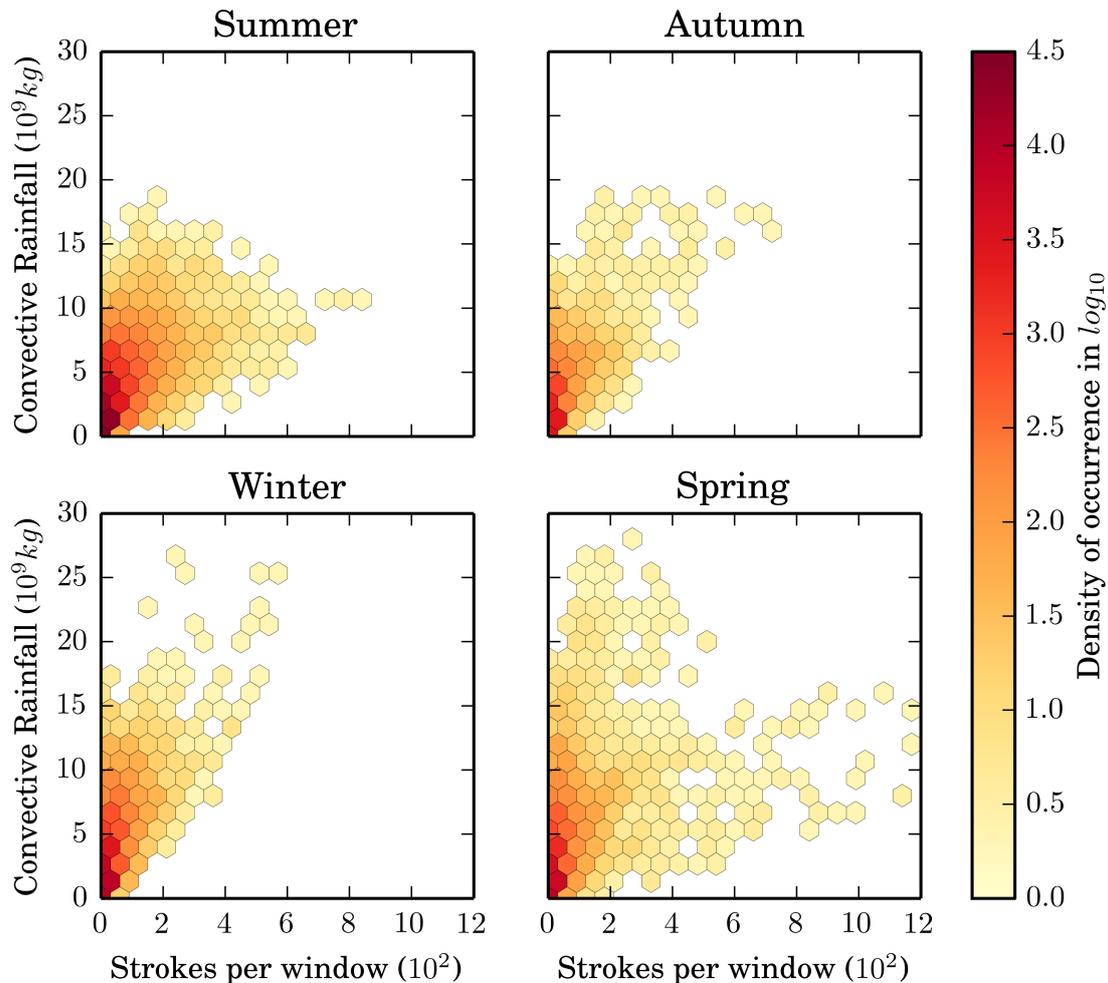


Figure 5.4 - Hexagonal binning for the tuples of convective rainfall and number of CG strokes. Each hexagon shows the averaged number of windows for each tuple.

New scatter plots were made averaging the convective rainfall for each value of N , showing a good tendency between the number of strokes and the convective rainfall for each season, as seen in Figure 5.5 for the averaged tuples. The same figure shows the corresponding WRLR functions obtained for each season assuming power functions, as stated in Section 4.4, and based on the scatter plots. These functions were obtained using a frequency-weighted least squares fitting for all the tuples, which defines weights to redundant tuples. A tuple that appears twice in the

database has a weight that is the double than a tuple that appears once. This fitting yields the real constants a , b and c that allows to define the power function stated in Equation 4.2. Figure 5.5 also shows that these WRLR functions for each season fit better the averaged tuples when $N < 50$ that correspond to more than 90% of all tuples of the dataset.

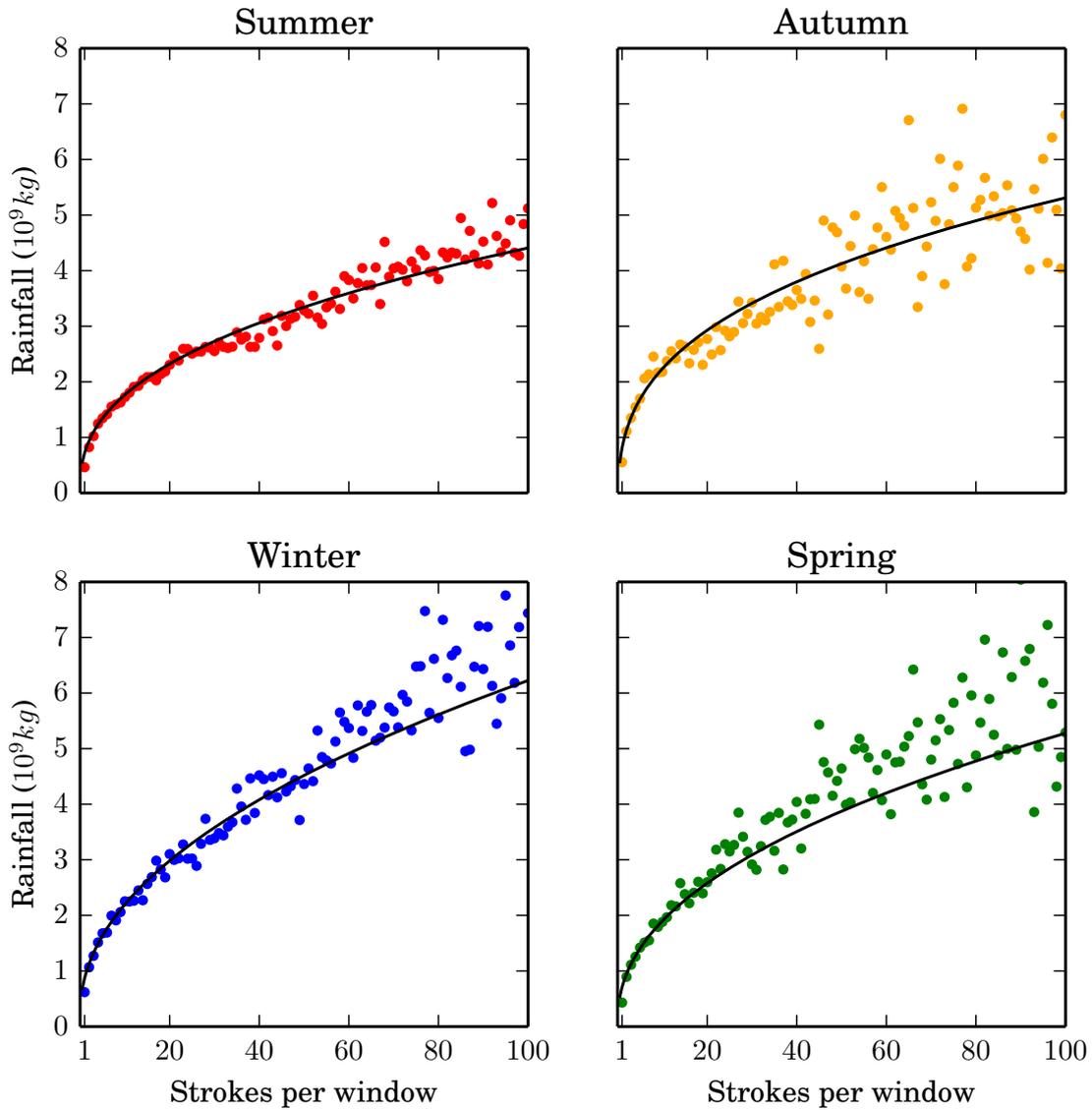


Figure 5.5 - Scatter plots of the averaged tuples (average rainfall for each number of strokes N) for each season. The WRLR functions obtained for all tuples of the dataset are plotted in black for each season.

The same WRLR functions for each season of Figure 5.5 are shown together in Figure 5.6. The convective rainfall mass per number of strokes tends to be higher in

Winter than in Summer, while Autumn and Spring have similar in-between values. The plots were limited to $N = 100$, but these functions provide an “reasonable” estimative of rainfall for the rare occurrences of $N > 100$.

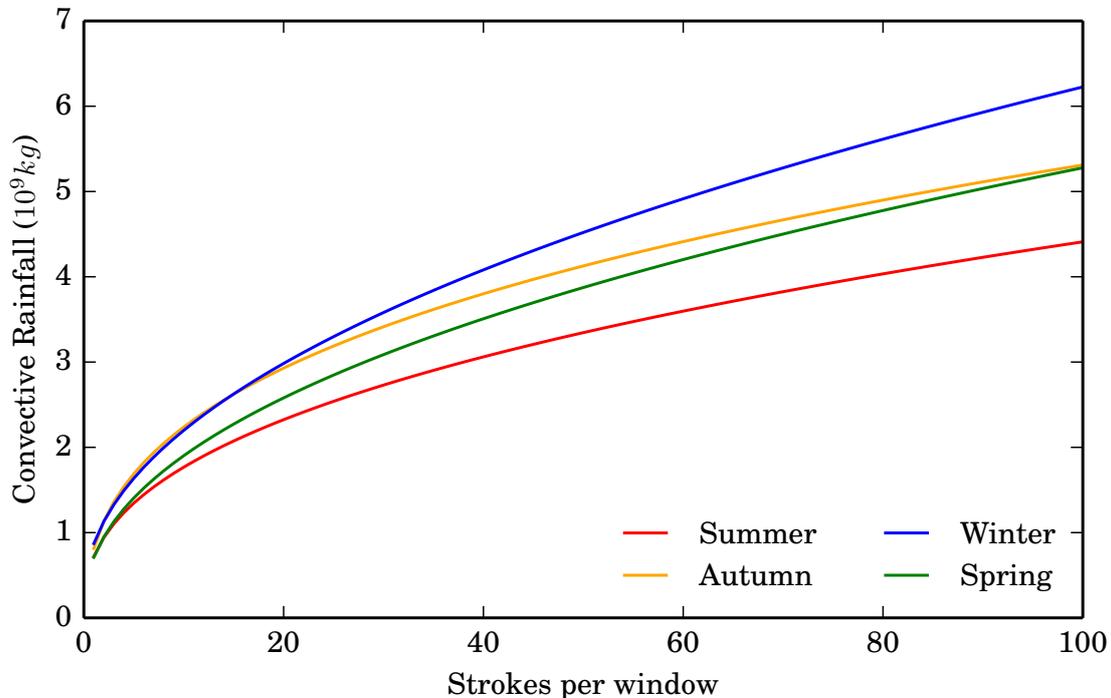


Figure 5.6 - The WRLR function for Summer, Autumn, Winter and Spring.

Considering that the database has about 54% of the tuples for Summer, 22% for Winter, 17% for Spring, and only 7% for Autumn, it is expected that the Summer WRLR function would be more accurate than the WRLR functions for the other seasons. Table 5.1 shows the real constants a , b , and c for these WRLR power functions.

Table 5.1 - Real constants a , b and c that describe the WRLR power function (Equation 4.2) for each season.

Season	a	b	c
Summer	685.70	0.40	35.18
Autumn	1227.00	0.33	-412.50
Winter	651.60	0.48	220.10
Spring	650.70	0.45	58.88

5.3 Mapping of the Summer WRLR function for different area sizes and window durations

In order to minimize the estimation error of the convective rainfall mass, the sliding-window time interval and area size were optimized for the generation of tuples that result in a WRLR function. These parameters were set as 50 km (edge of the square area) and 30 min (window duration). However, Section 4.4.2 presents the spatial adjust factor X (Equation 4.4) and the temporal adjust factor T (Equation 4.6) that allows to map the standard WRLR function to other area sizes and time intervals, respectively. This feature is not employed in this work, but may be useful for operational purposes. This sections illustrates the mapping of the Summer WRLR function described in the previous section, but the same scheme can be applied for the functions of the remaining seasons.

A comparison between $WRLR_{\Delta x}$ functions obtained for different edges Δx and the standard function $WRLR_{50 \text{ km}}$ mapped by the spatial adjust factor $X_{\Delta x}$, which is $[X_{\Delta x} \cdot WRLR_{50 \text{ km}}]$ is shown in Figure 5.7, always considering the standard window duration of 30 min. Assuming the specific $WRLR_{\Delta x}$ function as reference, the mean relative errors of the approximations $[X_{\Delta x} \cdot WRLR_{50 \text{ km}}]$ are relatively low, being 0.0131 for $\Delta x = 10$ km, 0.0206 for $\Delta x = 20$ km, 0.0193 for $\Delta x = 25$ km, 0.0081 for $\Delta x = 40$ km, and 0.0201 for $\Delta x = 100$ km. Thus, the spatial adjust factor X may be eventually used without significantly affecting the rainfall estimation.

Similarly, a comparison between $WRLR_{\Delta T}$ functions obtained for different window durations Δt and the standard function $WRLR_{30 \text{ min}}$ mapped by the temporal adjust factor $T_{\Delta t}$, which is $[T_{\Delta t} \cdot WRLR_{30 \text{ min}}]$ is shown in Figure 5.8, always considering the standard edge of 50 km. Assuming the specific $WRLR_{\Delta T}$ function as reference, the mean relative errors of the approximations $[T_{\Delta t} \cdot WRLR_{30 \text{ min}}]$ are relatively low, being 0.0049 for $\Delta t = 7.5$ min, 0.0067 for $\Delta t = 15$ min, and 0.0041 for $\Delta t = 60$ min. Also similarly, the temporal adjust factor T may be eventually used without significantly affecting the rainfall estimation.

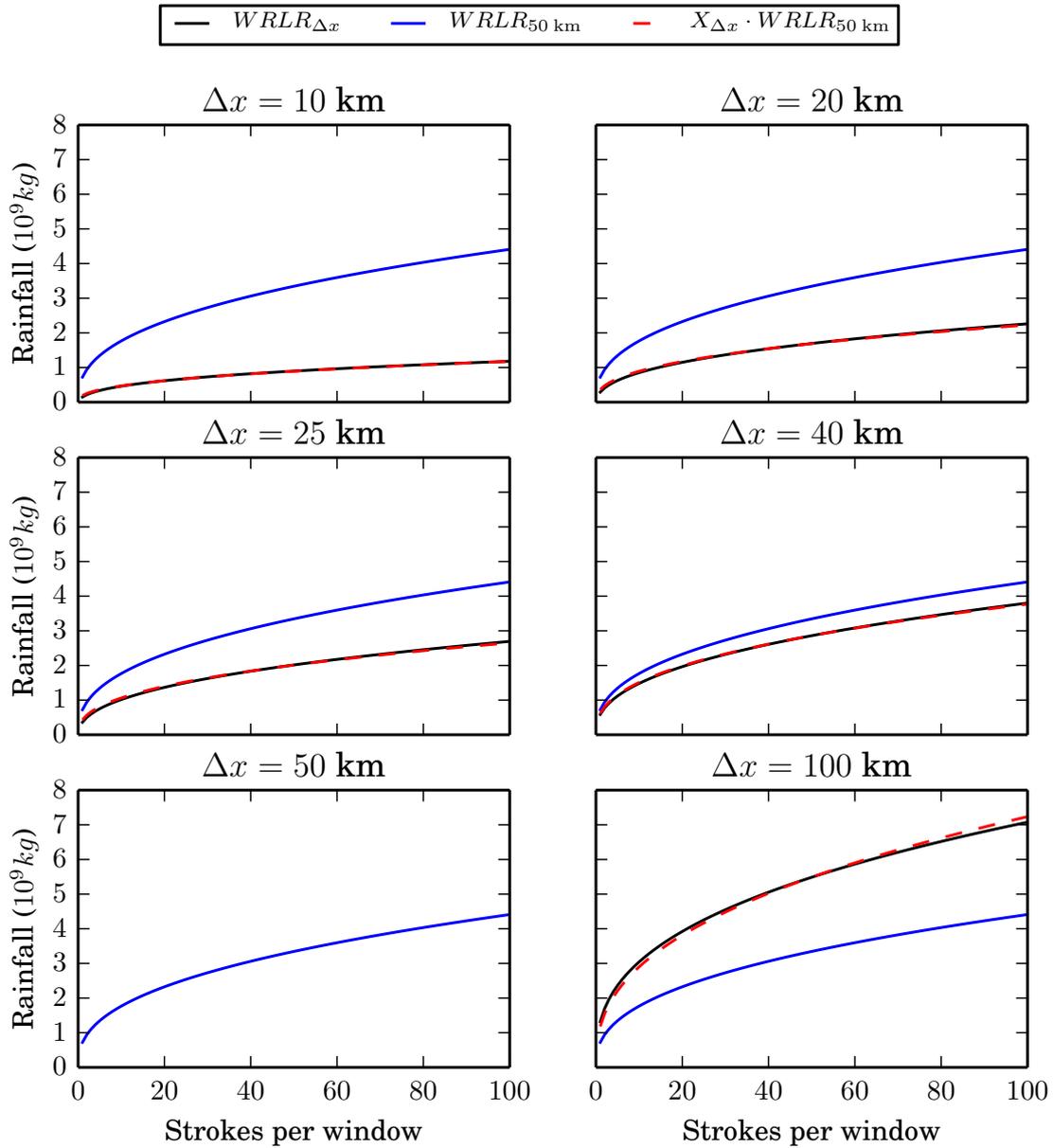


Figure 5.7 - Summer WRLR functions obtained for different Δx (10, 20, 25, 40, and 100 km, in black) compared to the standard WRLR function (in blue) mapped by the corresponding spatial adjust factors (red dashed line).

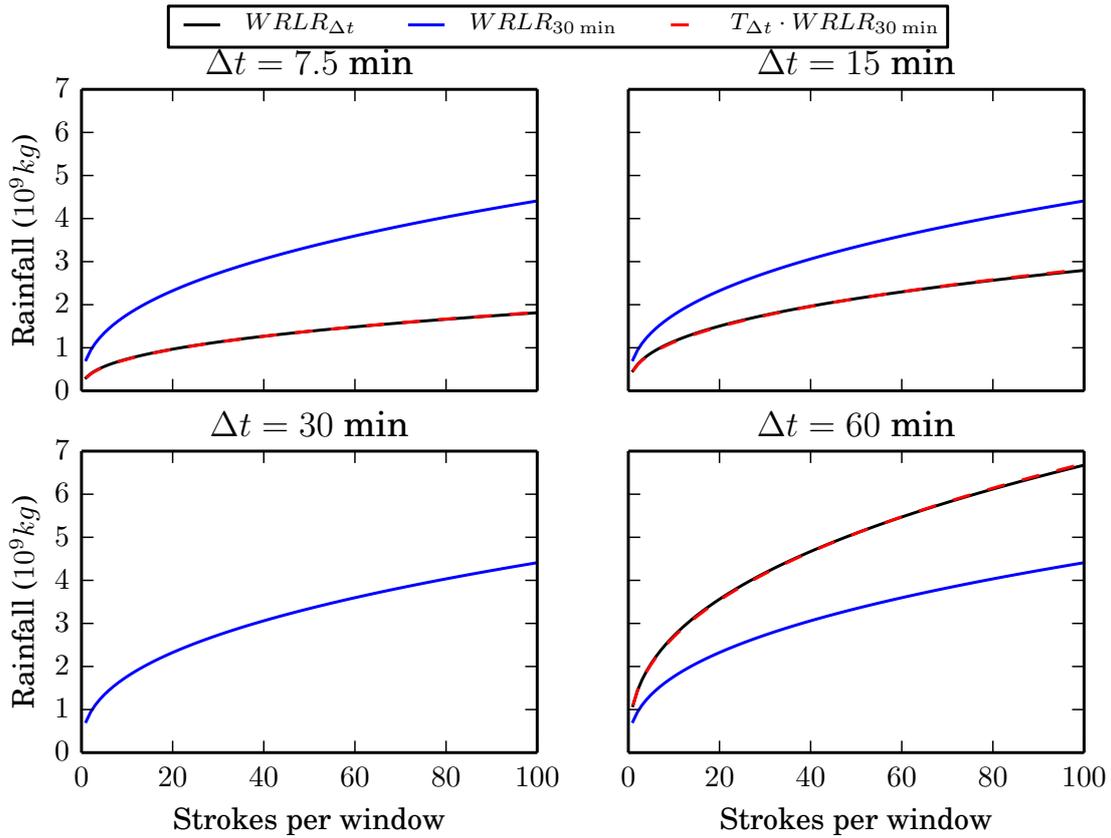


Figure 5.8 - Summer WRLR functions obtained for different Δt (7.5, 15 and 60 min, in black) compared to the standard WRLR function (in blue) mapped by the corresponding temporal adjust factors (red dashed line).

5.4 Validation tests of the rainfall estimation for the WRLR function

This work proposes the estimation of convective rainfall from CG lightning stroke data using a temporal sliding-window approach (Section 4.3) to estimate a WRLR function (Section 4.4) using reference weather radar data, in this case of the Bauru radar, and mapping this function to other areas by the K matrix (Section 4.5) that was derived using isohyetal maps of quarterly cumulative rainfall. Since the real rainfall is assumed as the one estimated from weather radar data using a suitable Z–R relationship, the only way to validate such estimation outside the coverage of the reference radar is by means of another weather radar. Therefore, the validation tests concerning the convective rainfall estimation were performed for areas under the coverage of the weather radars described in Section 2.3.2. For each weather radar, a set of thunderstorms was selected. Since 54% of the tuples of the database, tuples of number of CG strokes and convective rainfall, refer to Summer days, these validation tests were only performed for Summer days.

The next section (Section 5.4.1) shows the validation tests using the Bauru weather radar, followed by the validation tests using the Presidente Prudente weather radar (Section 5.4.2), the São Roque weather radar (Section 5.4.3), and the Pico do Couto weather radar (Section 5.4.4).

5.4.1 Validations tests using the Bauru weather radar

The validation of the WRLR function for convective rainfall estimation was performed using data of two months of the Summer of 2013, January and February. The estimations of convective rainfall mass given by the WRLR function and the Bauru weather radar are compared, using the latter as a reference. Additionally, the estimation given by the constant-valued RLR of the Tapia’s model was also included in the comparison. Test area is the square of 200 km edge inside the 150 km range of the Bauru radar, as shown in Figure 4.4. This area is approximately bounded by latitudes $23^{\circ}15'00''\text{S}$ to $21^{\circ}27'00''\text{S}$ and longitudes $50^{\circ}00'00''\text{W}$ to $48^{\circ}04'00''\text{W}$. This area was divided into 16 squares of edge $\Delta x = 50$ km corresponding to the standard sampling square size, as shown in Figure 4.5 (a), while the window duration was $\Delta t = 30$ min with window advance of $\delta t = 30$ min.

During these two months, the total convective rainfall estimated by weather radar was 3461×10^9 kg, against the RLR estimative of 4794×10^9 kg, and the WRLR estimative of 3407×10^9 kg. Figure 5.9 shows the daily estimative of convective rainfall mass and the cumulative value along the two months. It can be observed that the RLR tends to overestimate the rainfall by 38%, while the WRLR function, to slightly underestimate it by 2%, for the entire period. Considering the error for the daily accumulated rainfall, the mean of the RLR error was an overestimation of 30%, while the median was an overestimation of 16%. For the WRLR, the mean error for the daily accumulated rainfall was an overestimation of 6%, with the median presenting an underestimation of 1%.

A set of 7 thunderstorms was selected from this data in order to analyze more closely the performance of both estimators, as shown in Table 5.2. Thunderstorms are denoted by identifiers “BRU-#”. This table presents the input data for these thunderstorms: the number of CG lightning strokes for the WRLR function and the number of CG lightning flashes for the RLR constant, as well as the corresponding estimations of convective rainfall mass, compared to the value obtained from the Bauru weather radar. Thunderstorm duration is also included in the table. Table 5.3 shows the corresponding percentage errors for the total mass of the thunderstorms and the correlation for the hourly mass estimations in relation to the radar-inferred

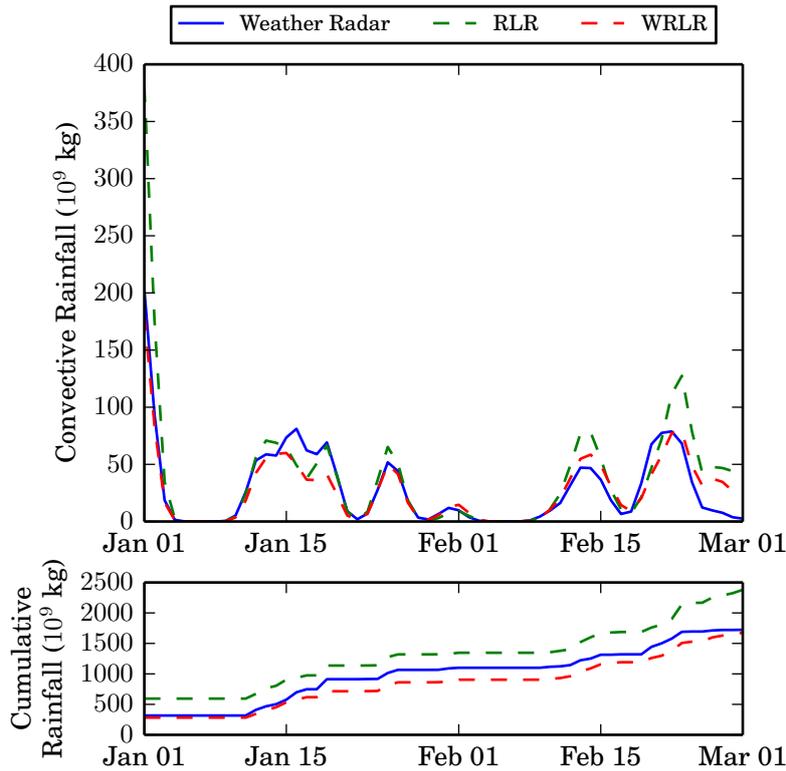


Figure 5.9 - The months of January and February for the Bauru radar during the Summer of 2013. The top image shows the daily accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), of the RLR constant (in dashed green) and of the WRLR function (in dashed red) for the convective rainfall mass.

values. The results of the WRLR function outperformed the RLR ones.

Table 5.2 - Convective rainfall mass in 10^9 kg estimated by weather radar, by the WRLR function and by the RLR constant for the 7 selected thunderstorms. The corresponding number of CG lightning strokes and flashes, the duration and the reference to the specific figures also appear.

Storm	Figure	CG Lightning		Rainfall Estimation			Duration (h)
		Strokes	Flashes	Radar	WRLR	RLR	
BRU-1	5.10	6210	5428	633	564	1188	9
BRU-2	5.11	1034	912	141	157	199	9
BRU-3	5.12	1370	1224	194	191	268	10
BRU-4	5.13	2588	2202	223	254	482	10
BRU-5	5.14	1574	1472	330	193	322	9
BRU-6	5.15	282	270	241	88	59	14
BRU-7	5.16	950	788	35	132	172	7

Table 5.3 - Percentage error (*error*) and correlation (*corr*) of the WRLR and RLR estimates for the 7 selected thunderstorms.

Storm	WRLR		RLR	
	<i>error</i>	<i>corr</i>	<i>error</i>	<i>corr</i>
BRU-1	-11%	0.93	+88%	0.81
BRU-2	+11%	0.89	+41%	0.87
BRU-3	-1%	0.99	+38%	0.96
BRU-4	+14%	0.97	+116%	0.94
BRU-5	-41%	0.80	-2%	0.83
BRU-6	-63%	0.86	-75%	0.75
BRU-7	+274%	0.84	+389%	0.95

It can be expected a mismatch between the convective rainfall mass estimations estimated from lightning data and the ones obtained from weather radar. The results show that, considering each thunderstorm, the RLR constant tends to overestimate the rainfall mass, while the WRLR function, to underestimate it, reproducing the same behavior already observed for the totals for the two months. However, this a general tendency, and some thunderstorms were selected as exceptions. Hourly convective rainfall mass estimation are shown for thunderstorms BRU-1 to BRU-7 in Figures 5.10 to 5.16, respectively. The RLR constant grossly overestimated mass values for the first four thunderstorms, while the WRLR function resulted in better estimations. The three remaining thunderstorms, selected as worst cases for the WRLR function, resulted in poorer estimations. In BRU-6, the rainfall was underestimated by the RLR and the WRLR, while in BRU-7, both approaches overestimated the rainfall. Thunderstorms BRU-6 and BRU-7 are weaker than the others shown, which suggests the rainfall-lightning correlation may be linked to the thunderstorm strength.

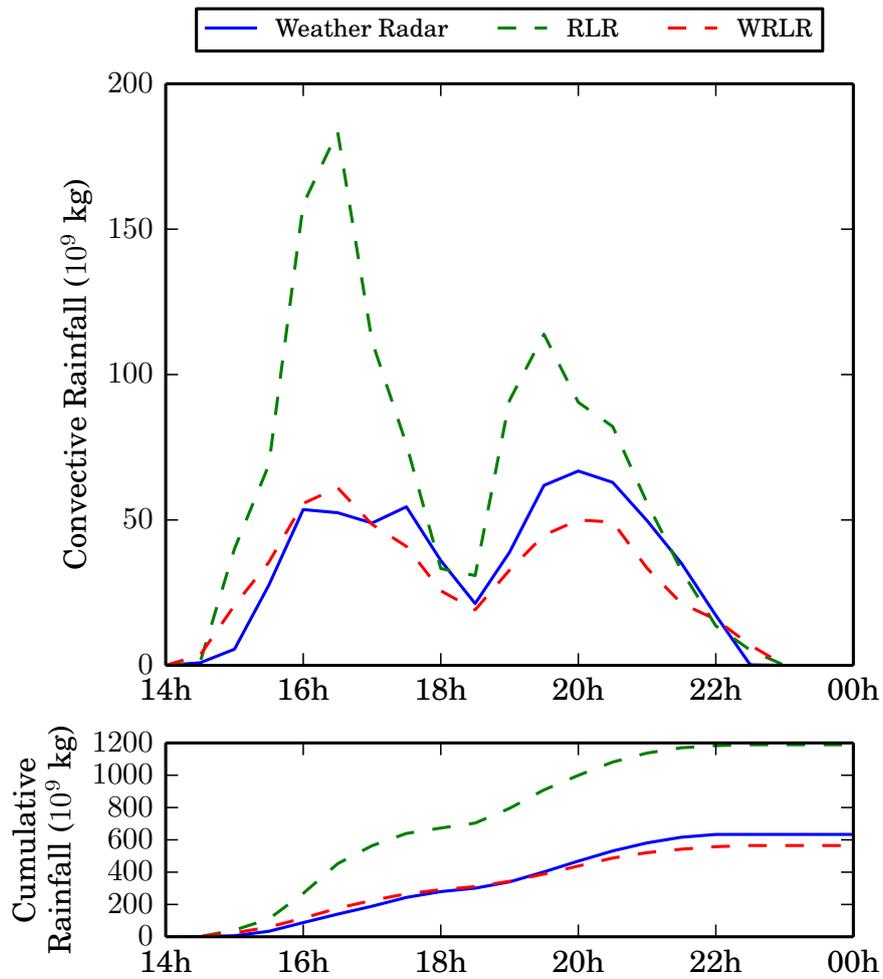


Figure 5.10 - BRU-1 thunderstorm occurred in 1st January 2013. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), of the RLR constant (in dashed green) and of the WRLR function (in dashed red) for the convective rainfall mass.

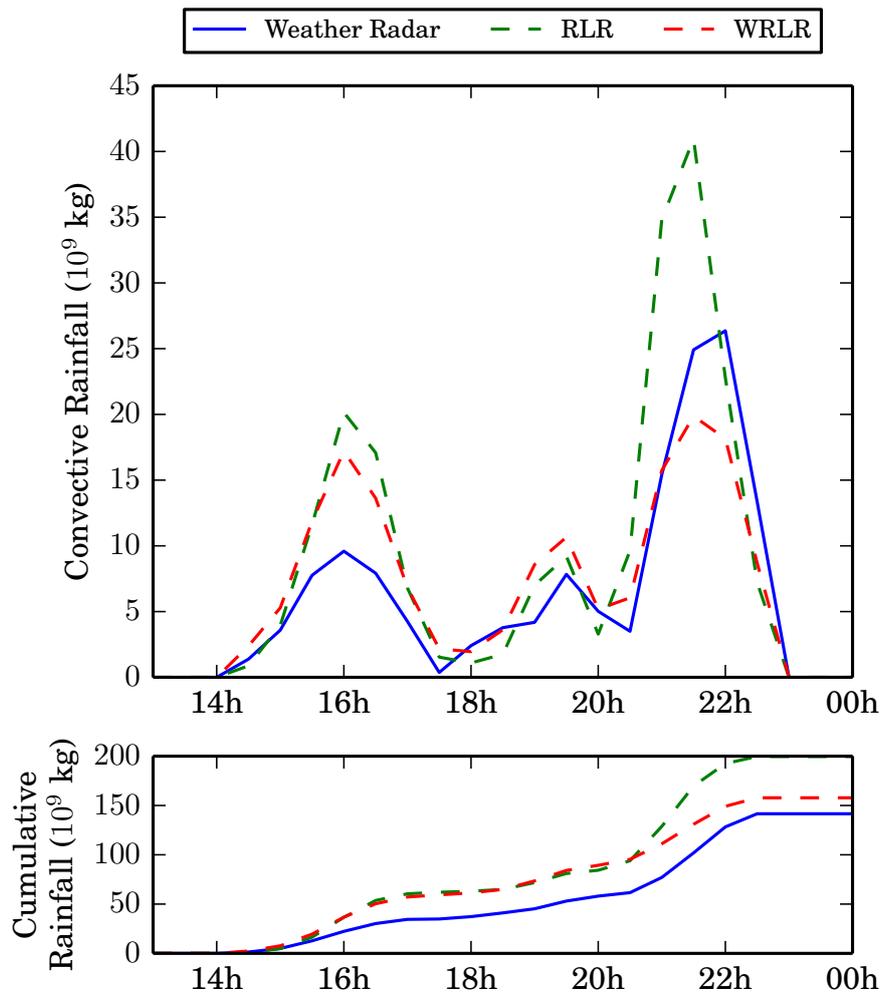


Figure 5.11 - BRU-2 thunderstorm occurred in 15th January 2013. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), of the RLR constant (in dashed green) and of the WRLR function (in dashed red) for the convective rainfall mass.

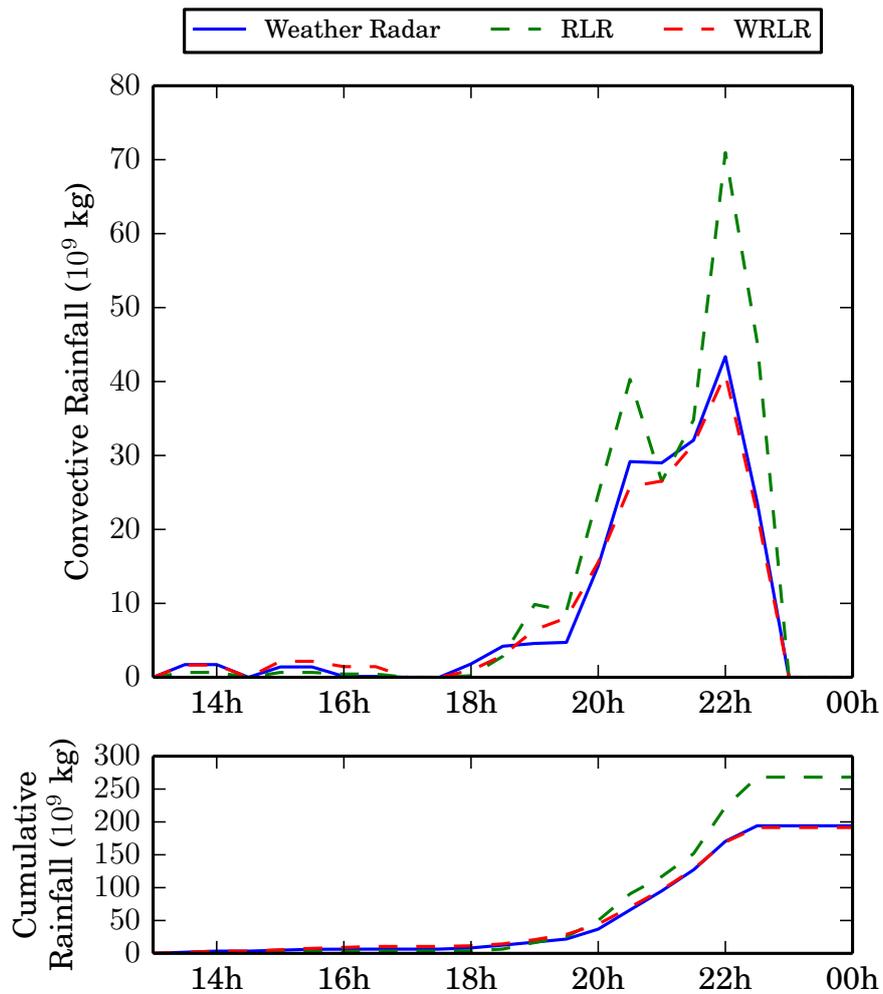


Figure 5.12 - BRU-3 thunderstorm occurred in 25th January 2013. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), of the RLR constant (in dashed green) and of the WRLR function (in dashed red) for the convective rainfall mass.

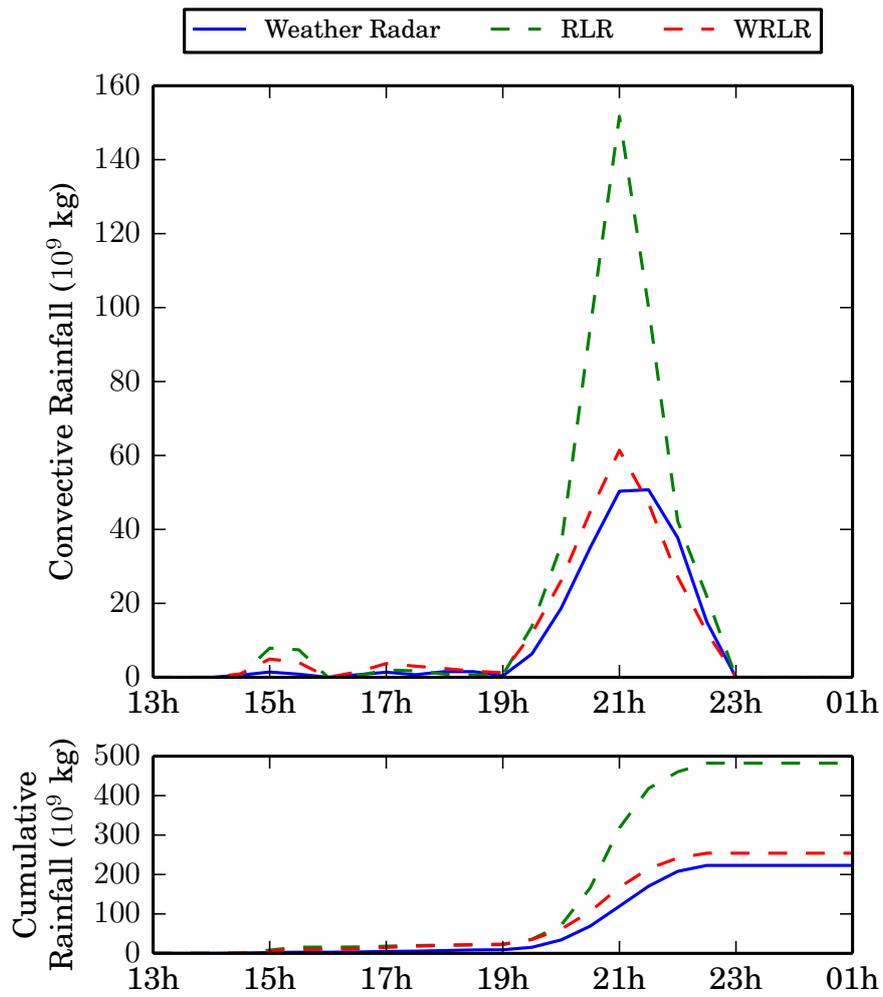


Figure 5.13 - BRU-4 thunderstorm occurred in 23rd February 2013. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), of the RLR constant (in dashed green) and of the WRLR function (in dashed red) for the convective rainfall mass.

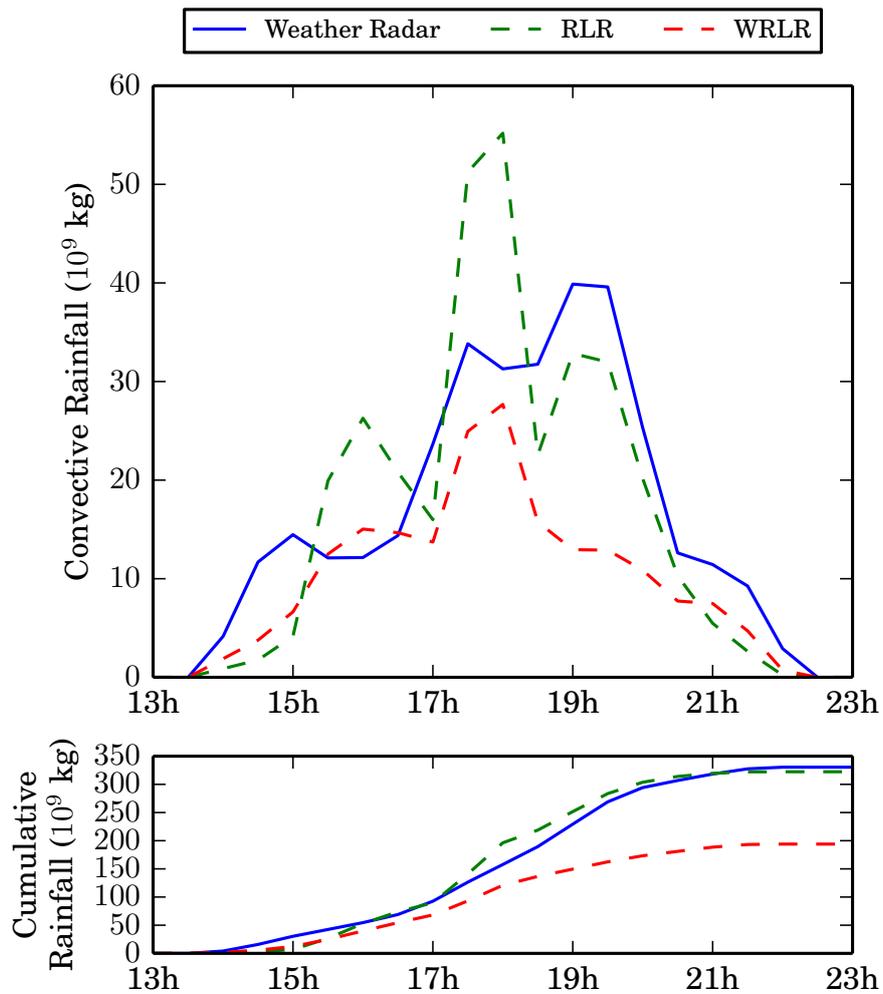


Figure 5.14 - BRU-5 thunderstorm occurred in 19th January 2013. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), of the RLR constant (in dashed green) and of the WRLR function (in dashed red) for the convective rainfall mass.

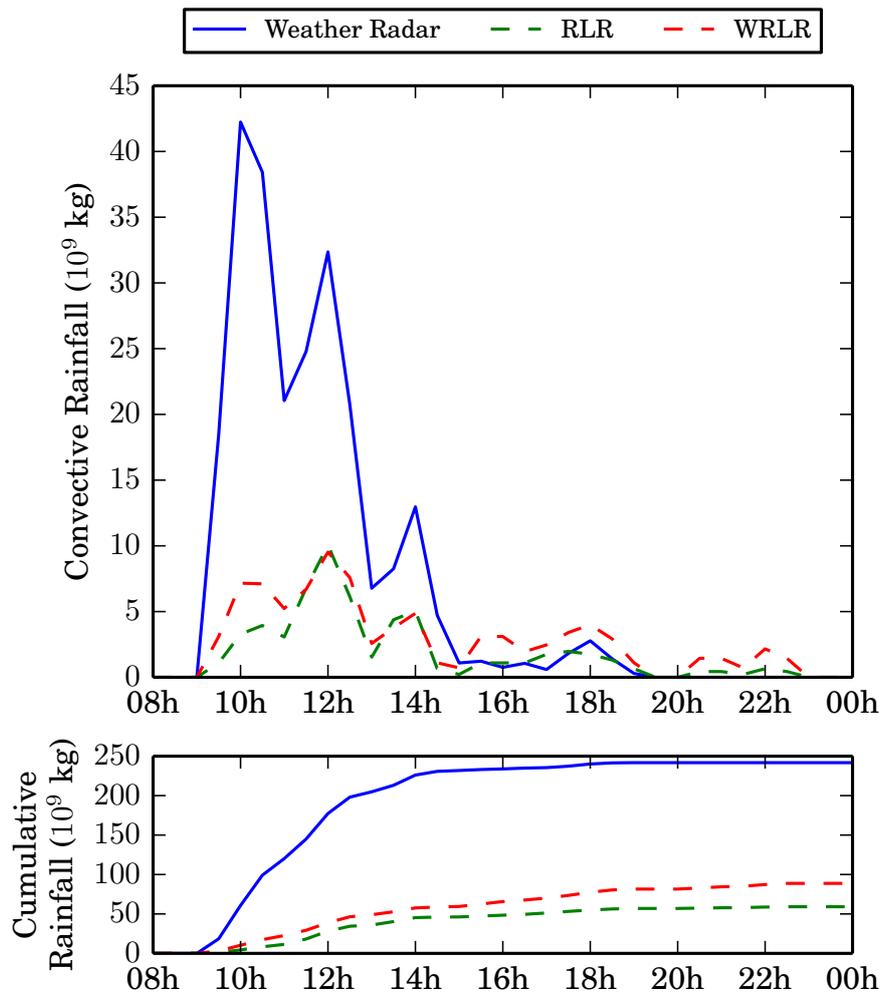


Figure 5.15 - BRU-6 thunderstorm occurred in 16th January 2013. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), of the RLR constant (in dashed green) and of the WRLR function (in dashed red) for the convective rainfall mass.

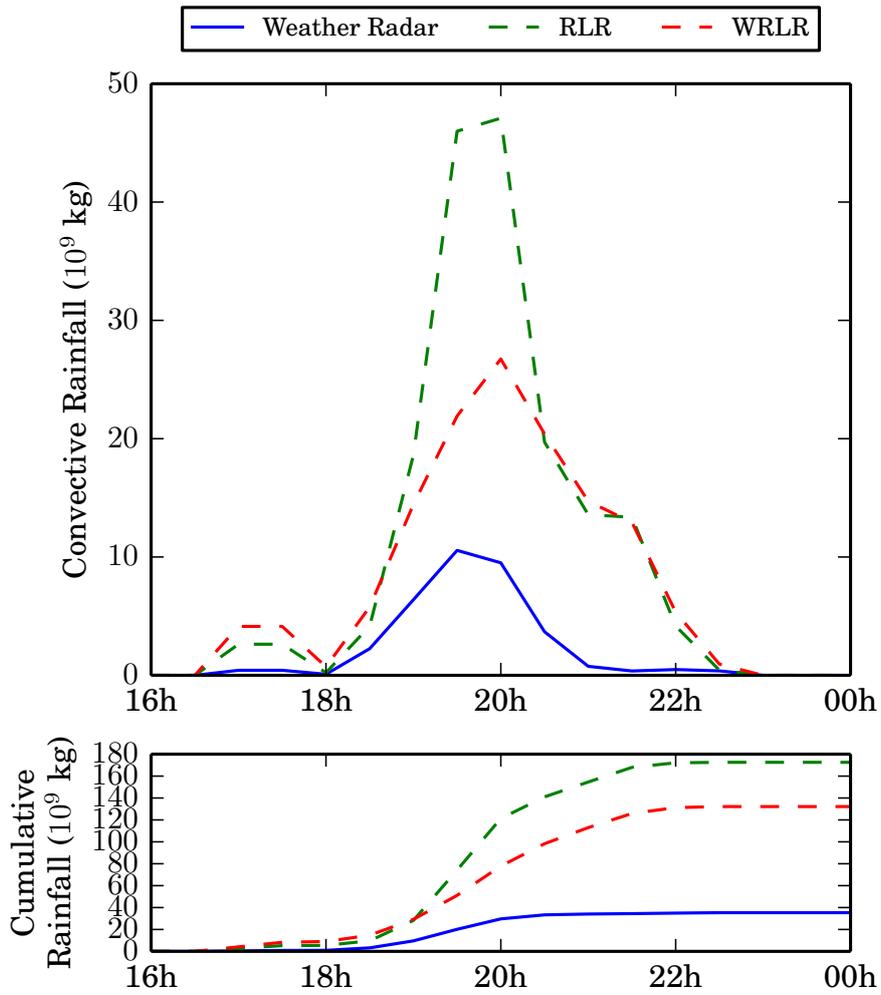


Figure 5.16 - BRU-7 thunderstorm occurred in 26th February 2013. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), of the RLR constant (in dashed green) and of the WRLR function (in dashed red) for the convective rainfall mass.

5.4.2 Validations tests using the Presidente Prudente weather radar

The previous section presented the validation tests of the WRLR function and of the Tapia's model (RLR constant) using the Bauru weather radar. The WRLR function was inferred using the same Bauru radar. Rainfall estimations for any place outside the coverage of the Bauru radar demands the spatial mapping of the function, described in Section 4.5, by means of the specific elements of the K matrix. This section presents validation tests using the Presidente Prudente weather radar, but including the four seasons. Each season is approximately represented by the correspondent quarter: Summer (January, February and March), Autumn (April, May and June), Winter (July, August and September) and Spring (October, November and December).

The data for the four seasons were only available for the radars of Bauru and Presidente Prudente. The data for the radars of São Roque and Pico do Couto included only Summer months. The corresponding WRLR functions are very similar due to the overlap of their coverages. Since the Bauru radar was taken as the reference radar and employed to derive the WRLR function, it was chosen to use to perform validation tests for the four seasons employing the Presidente Prudente radar. The results for the four seasons using the Bauru radar itself would be probably too similar.

The validation tests for the four seasons were performed considering a square with edge of 200 km centered at the Presidente Prudente radar, as described in Figure 4.4, bounded approximately by latitudes $23^{\circ}04'00''\text{S}$ to $21^{\circ}16'00''\text{S}$ and longitudes $52^{\circ}21'00''\text{W}$ to $50^{\circ}24'00''\text{W}$. This area was divided into 16 squares of edge $\Delta x = 50$ km corresponding to the standard sampling square size, as shown in 4.5 (a), while the window duration was $\Delta t = 30$ min with window advance of $\delta t = 30$ min.

Data of seven seasons/quarters of the years of 2009 and 2010 was available for this square area in order to analyze the performance of the mapped WRLR function for the estimation the cumulative convective rainfall mass, assuming the rainfall estimated using the Presidente Prudente weather radars as a reference. Table 5.4 describes the seasons/quarters, the cumulative number of CG lightning strokes that serves as input to the WRLR function as well as the cumulative convective rainfall mass of the function estimated by the function and by the radar. The estimation error for the quarterly accumulated convective rainfall as well as the mean daily estimation error for the period are also shown.

Table 5.4 - Estimation performance of the mapped WRLR function for the considered seasons/quarters for the area under the coverage of the Presidente Prudente radar. Convective rainfall mass is given in 10^9 kg and the percentage error is taken in relation to the radar estimation.

Season	Figure	CG Strokes	Rainfall Estimation Radar	WRLR	Quarterly error	Daily error mean	median
Summer 2009	5.17	28265	3960	3364	-15%	-4%	-5%
Autumn 2009	5.18	5189	595	643	+8%	+11%	-8%
Winter 2009	5.19	47329	5936	5736	-3%	-3%	-9%
Spring 2009	5.20	44834	6326	5070	-20%	+4%	-10%
Summer 2010	5.21	36707	5375	4628	-14%	-4%	-7%
Autumn 2010	5.22	3549	577	578	+0.2%	+7%	+4%
Winter 2010	5.23	12008	3146	2175	-31%	+13%	+6%

Figures 5.17 to 5.23 show the estimations for the daily accumulated convective rainfall mass given by radar and by the mapped WRLR function for the seasons/quarters listed in Table 5.4. The same figures also show the corresponding cumulative rainfall estimations. In general, this function underestimates rainfall by about 10%. For the Summer quarters, this underestimation was 14-15%, with a mean daily underestimation of 4%. The amount of data can be grossly inferred from the number of CG strokes thus both Autumn quarters would have poor estimations since there is less data. However, specifically for the 2010 Autumn quarter, the thunderstorms were few and weak, but the related convective rainfall was well-correlated to the lightning strokes, resulting in good estimations for the daily accumulates. As seen in Figure 5.6, the Winter WRLR function yields more convective rainfall mass per lightning stroke, but the 2010 Winter quarter had the worst estimation error (-31%), with a daily mean overestimation of 13%.

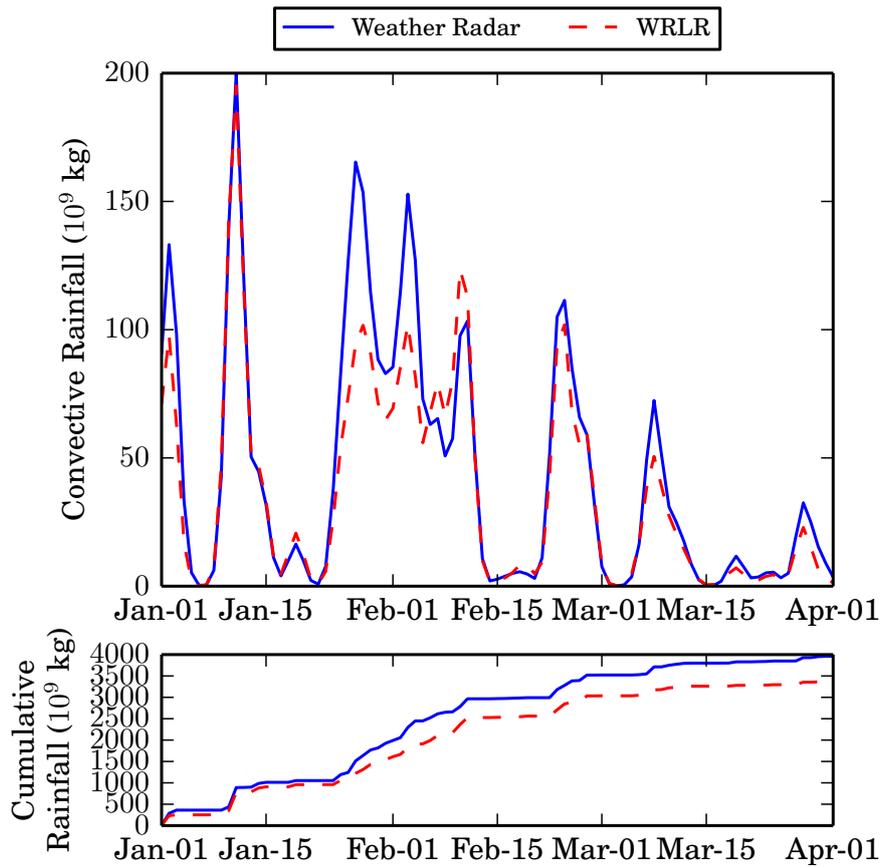


Figure 5.17 - Rainfall for the Summer of 2009 in Presidente Prudente. The top image shows the daily accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.

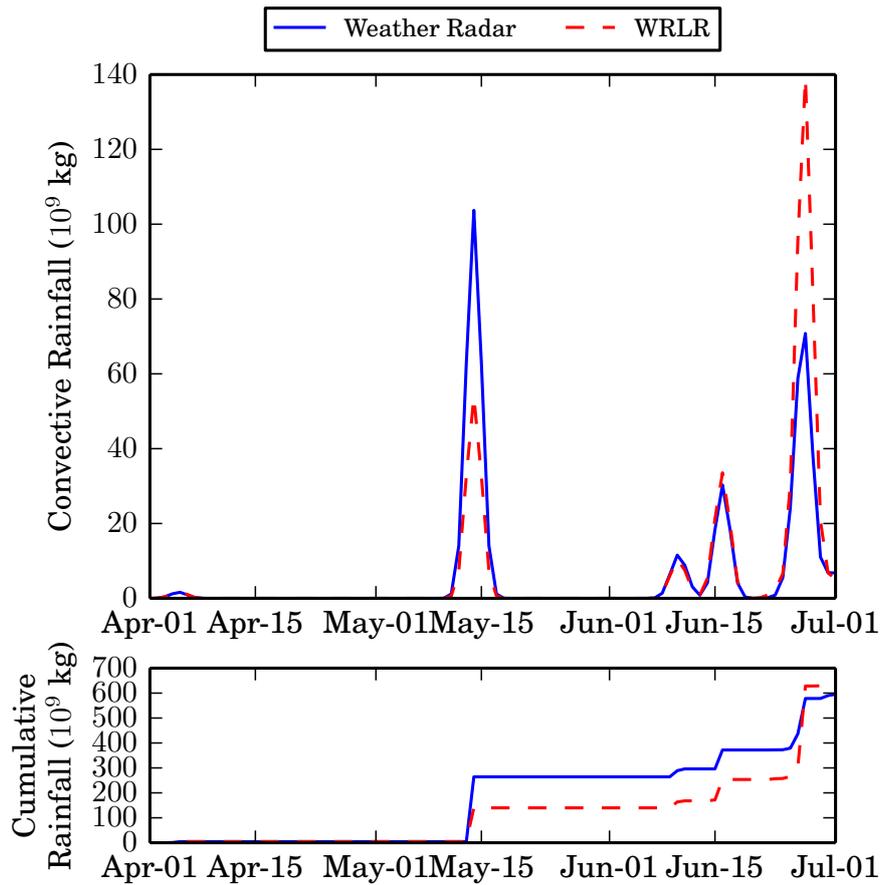


Figure 5.18 - Rainfall for the Autumn of 2009 in Presidente Prudente. The top image shows the daily accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.

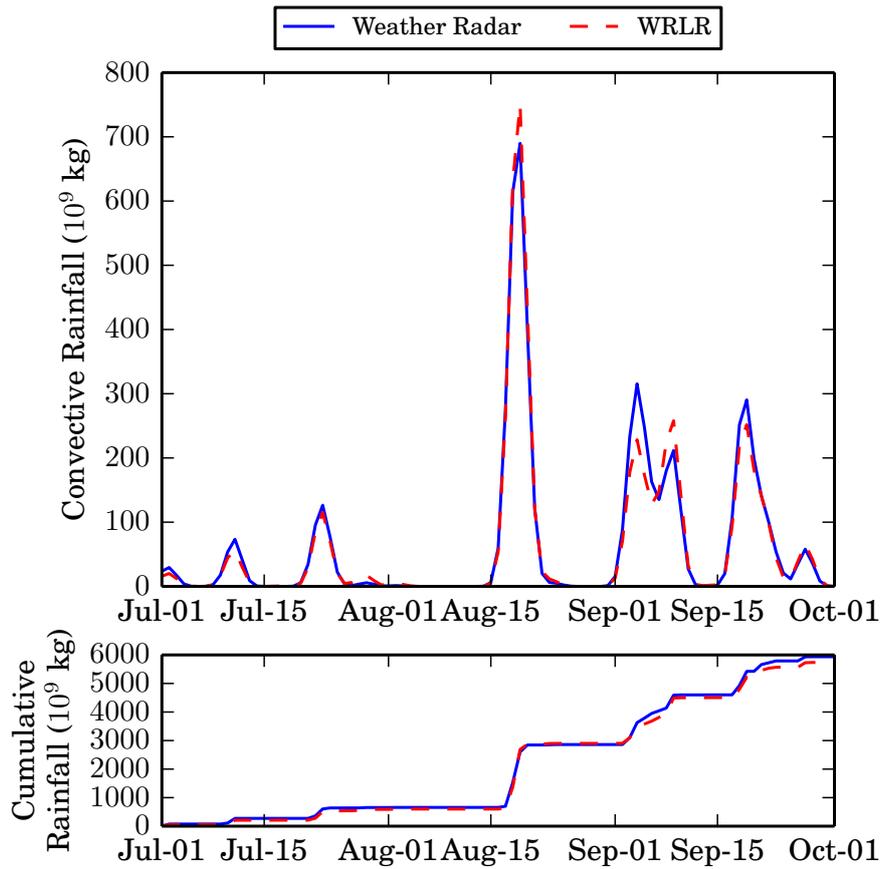


Figure 5.19 - Rainfall for the Winter of 2009 in Presidente Prudente. The top image shows the daily accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.

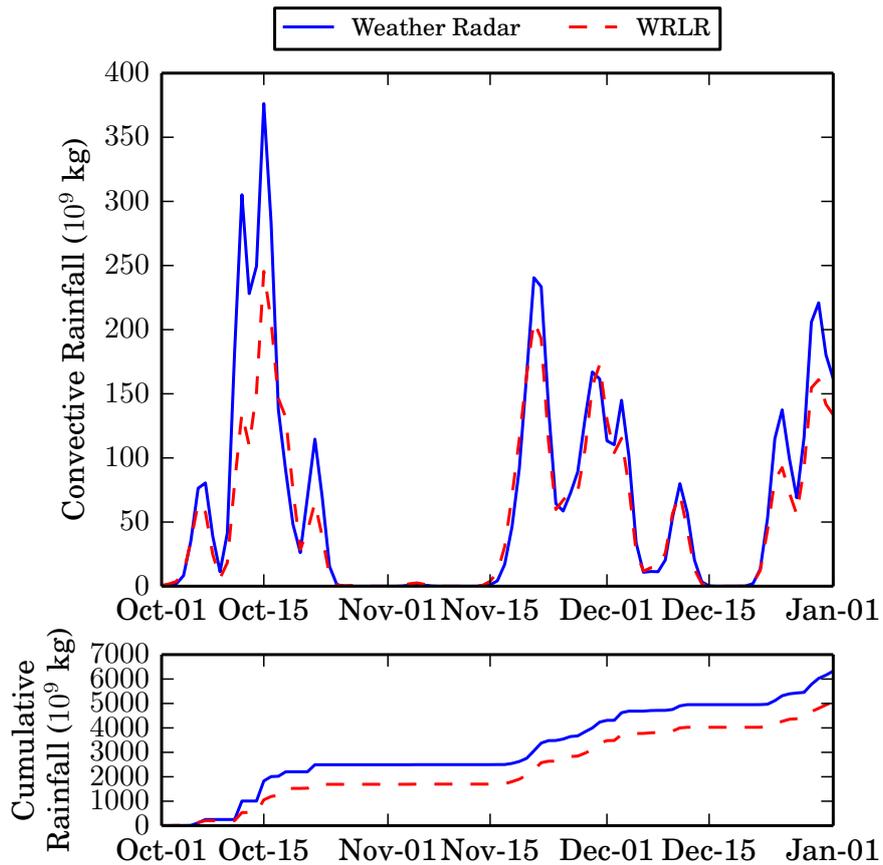


Figure 5.20 - Rainfall for the Spring of 2009 in Presidente Prudente. The top image shows the daily accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.

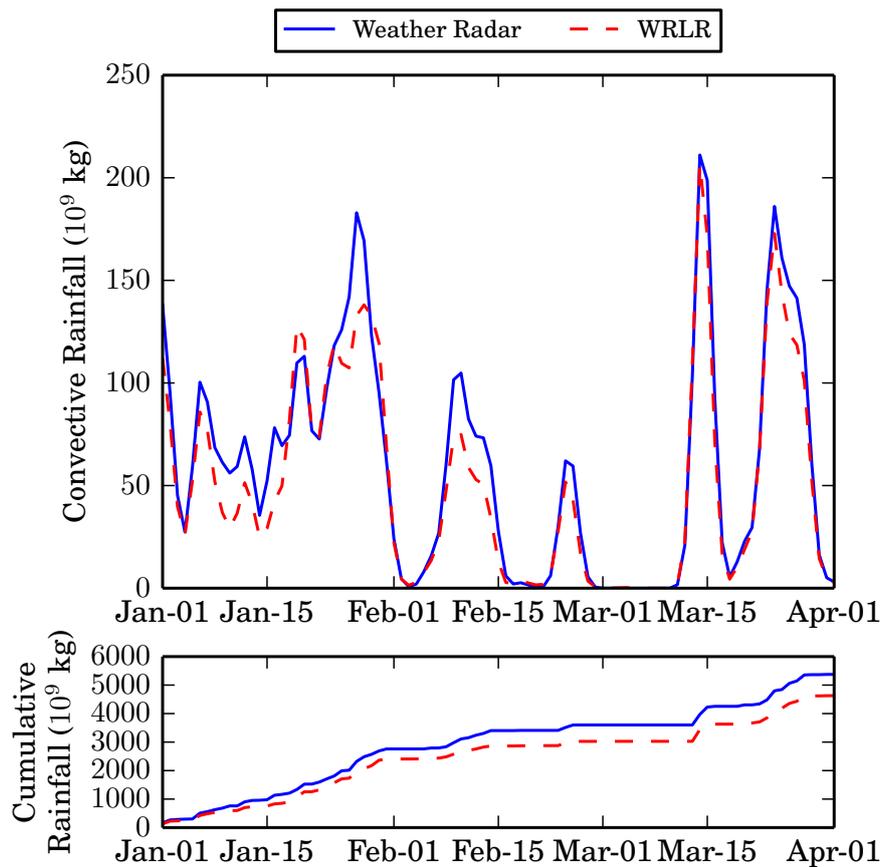


Figure 5.21 - Rainfall for the Summer of 2010 in Presidente Prudente. The top image shows the daily accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.

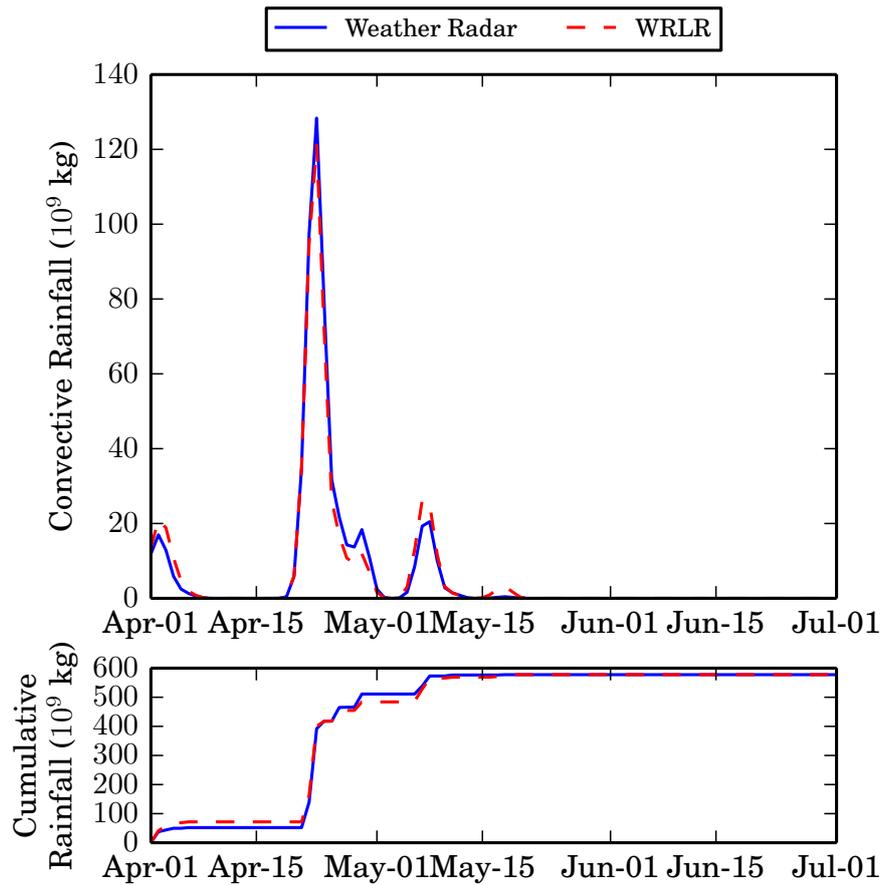


Figure 5.22 - Rainfall for the Autumn of 2010 in Presidente Prudente. The top image shows the daily accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.

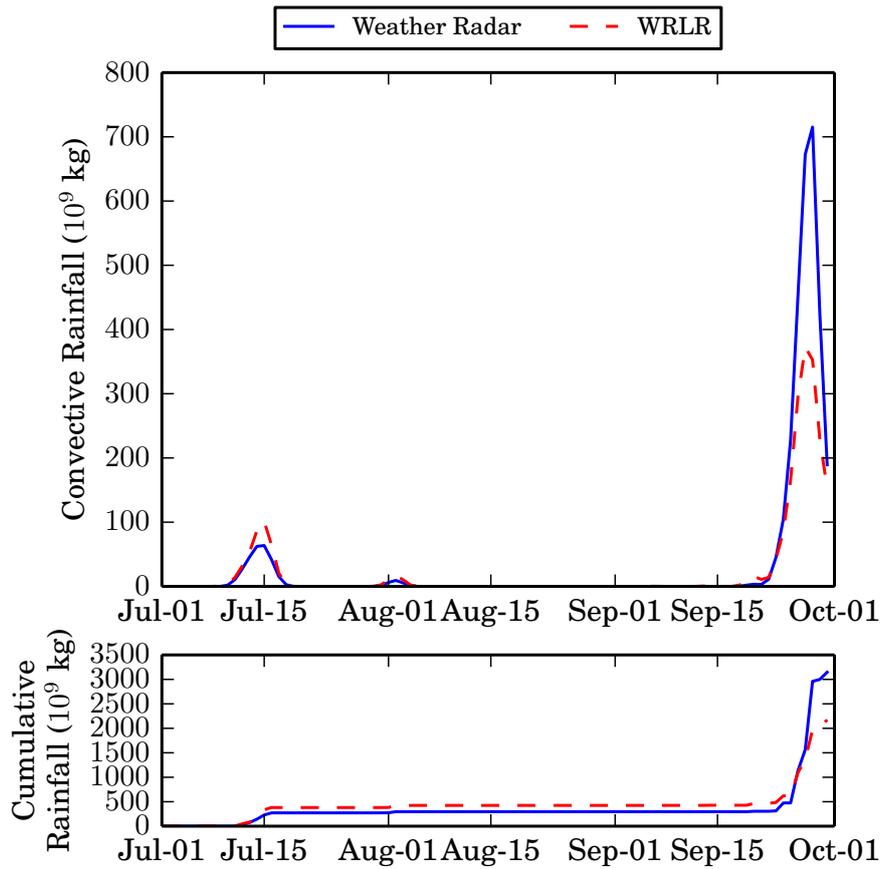


Figure 5.23 - Rainfall for the Winter of 2010 in Presidente Prudente. The top image shows the daily accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.

Besides the analysis of the mapped WRLR function for estimation of the daily accumulates of convective rainfall, a set of 12 thunderstorms was selected, which occurred inside the coverage of the Presidente Prudente weather radar, three for each season/quarter. The selected thunderstorms for the Summer were denoted as PPR-1S, PPR-2S and PPR-3S, for the Autumn, as PPR-4A, PPR-5A and PPR-6A, for the Winter, PPR-7W, PPR-8W and PPR-9W, and for the Spring, PPR-10R, PPR-11R and PPR-12R, as shown in Table 5.5. The same table shows the estimation error for the total cumulative convective rainfall mass in relation to the radar estimation for each thunderstorm, and the correlation between the mapped WRLR function and the radar values for 30-min accumulated convective rainfall mass, also for each thunderstorm.

Table 5.5 - Comparison of the estimations for the convective rainfall mass given by the Presidente Prudente weather radar and by the mapped WRLR function for the 12 selected thunderstorms. Percentage error is calculated in relation to the radar estimation, while the correlation considers 30-min accumulated values for each thunderstorm.

Storm	Figure	CG Convective Rainfall		Duration	<i>error</i>	<i>corr</i>	
		Strokes	Radar				WRLR
PPR-1S	5.24	5143	451	423	24 h	-6%	0.92
PPR-2S	5.25	4991	593	559	45 h	-6%	0.89
PPR-3S	5.26	1153	209	163	12 h	-22%	0.96
PPR-4A	5.27	2419	331	316	32 h	-4%	0.94
PPR-5A	5.28	161	37	40	15 h	+10%	0.92
PPR-6A	5.29	735	259	135	12 h	-48%	0.99
PPR-7W	5.30	20138	1947	2072	58 h	+6%	0.85
PPR-8W	5.31	2203	364	320	40 h	-12%	0.87
PPR-9W	5.32	3915	1478	666	21 h	-55%	0.92
PPR-10R	5.33	2542	403	338	20 h	-16%	0.92
PPR-11R	5.34	978	144	123	14 h	-15%	0.87
PPR-12R	5.35	7637	819	514	15 h	-37%	0.88

Estimation performance for the three Summer thunderstorms of Presidente Prudente:

The three selected Summer thunderstorms, PPR-1S (Figure 5.24), PPR-2S (Figure 5.25), and PPR-3S (Figure 5.26), are shown bellow. PPR-1S and PPR-2S are very strong thunderstorm systems, with rainfall and CG lightning activity lasting for 24h and the latter lasting for 45h in the studied region. The convective rainfall mass

for these two thunderstorms is well estimated by the WRLR function, as shown in the cumulative rainfall for each figure. PPR-3S is a relatively shorter and weaker thunderstorm compared to the former two thunderstorms, and it underestimates the total rainfall.

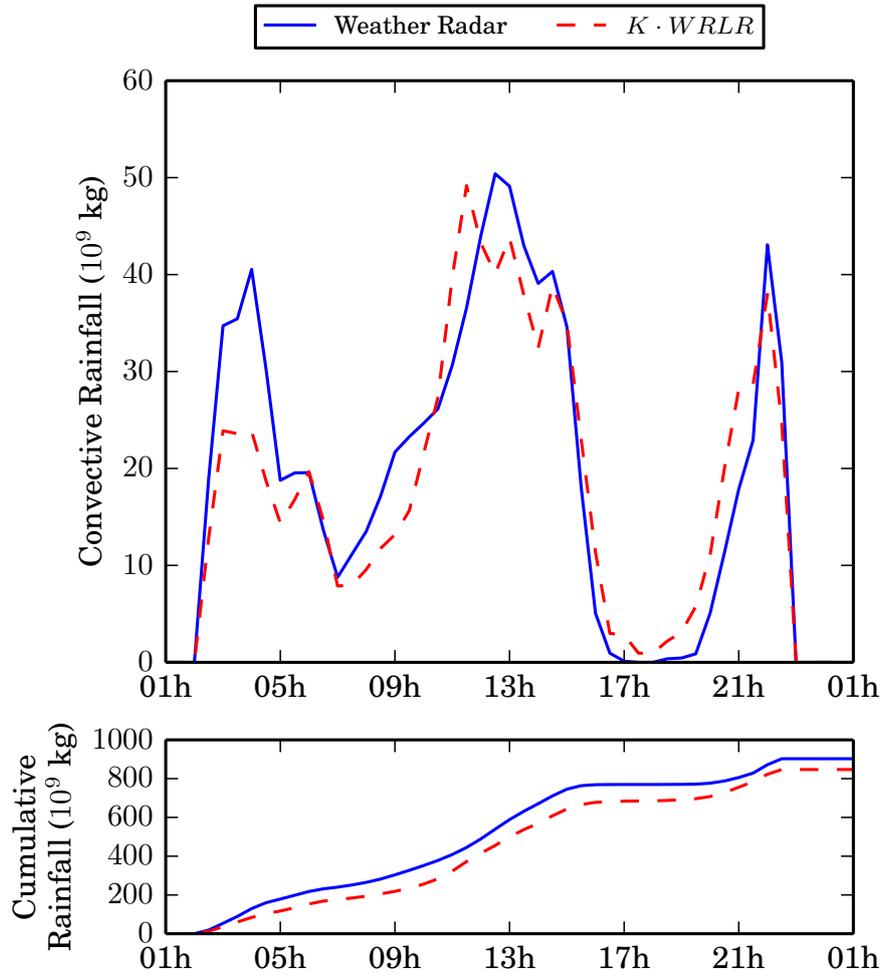


Figure 5.24 - PPR-1S thunderstorm occurred in 11th January 2009. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.

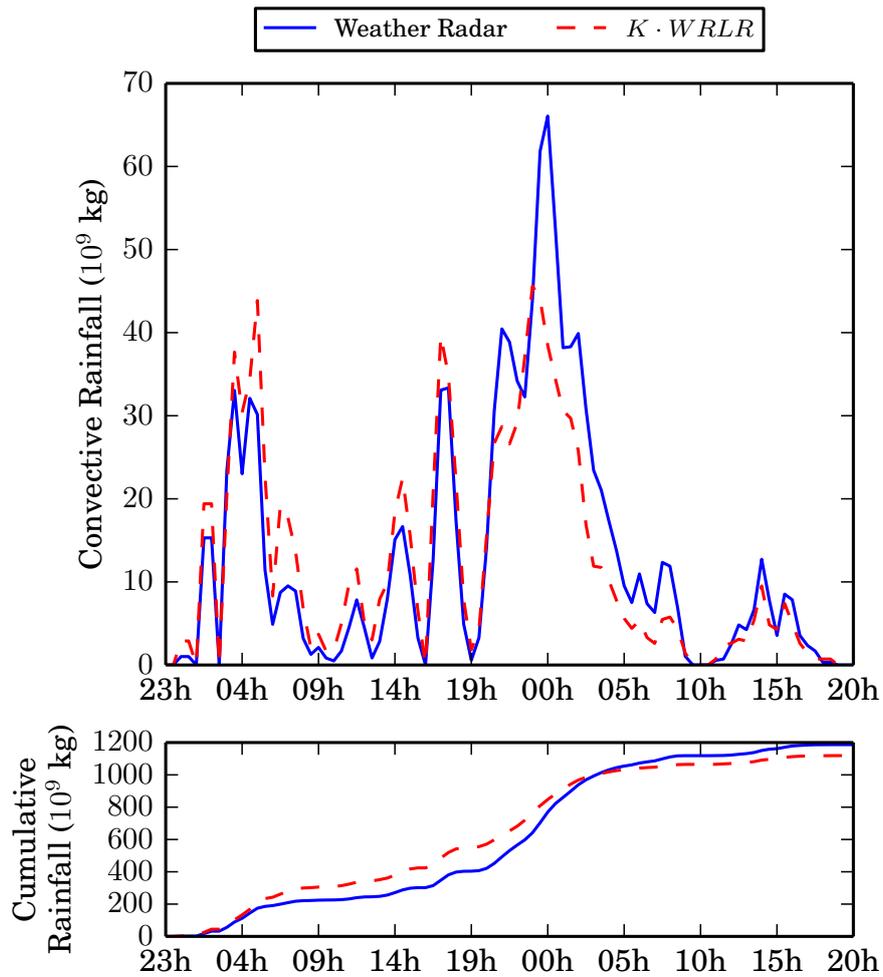


Figure 5.25 - PPR-2S thunderstorm occurred in 13th March 2010. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.

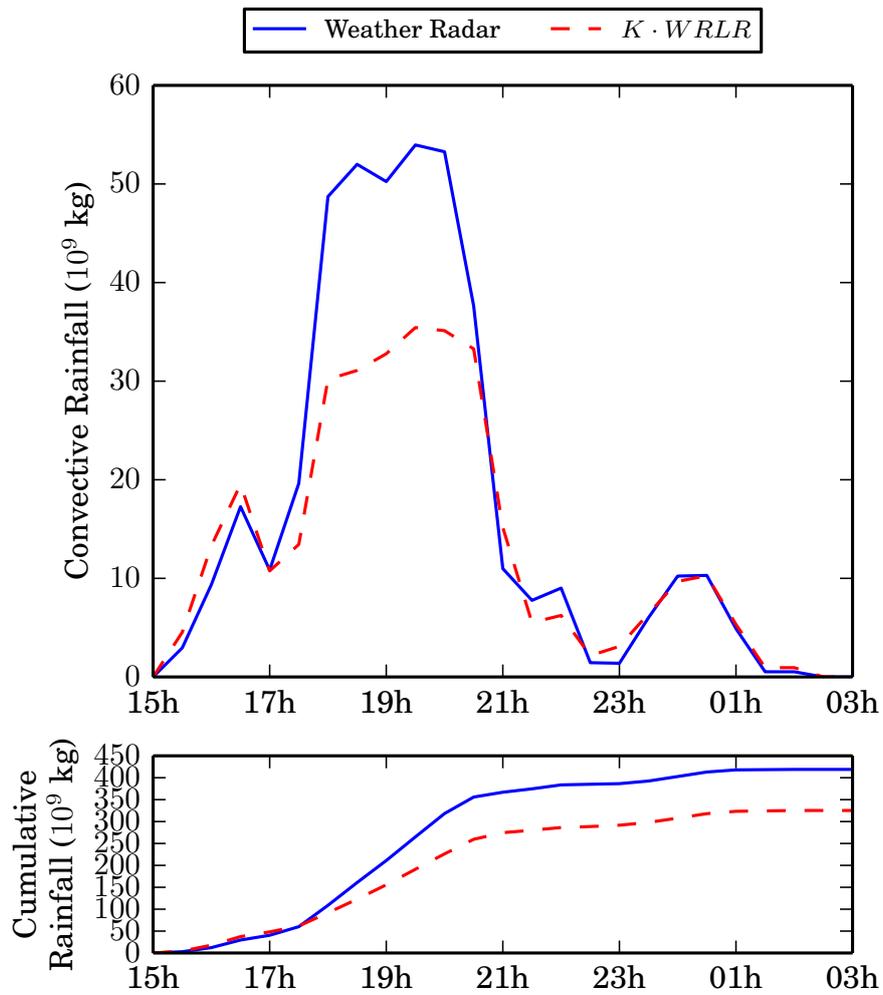


Figure 5.26 - PPR-3S thunderstorm occurred in 6th January 2010. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.

Estimation performance for the three Autumn thunderstorms of Presidente Prudente:

The three selected Autumn thunderstorms, PPR-4A (Figure 5.27), PPR-5A (Figure 5.28), and PPR-6A (Figure 5.29), are shown bellow. These are weaker thunderstorms than those from the Summer, and tend to last less. In PPR-4A, the WRLR estimation presents a very good fit, in PPR-5A, there is a small overestimation by the WRLR in the final phase of the studied thunderstorm, while in PPR-6A there is a more significant underestimation of the total rainfall by the mapped WRLR function. Thunderstorms were less frequent in Autumn than in any other season, and they were also significantly weaker than those occurring in other seasons.

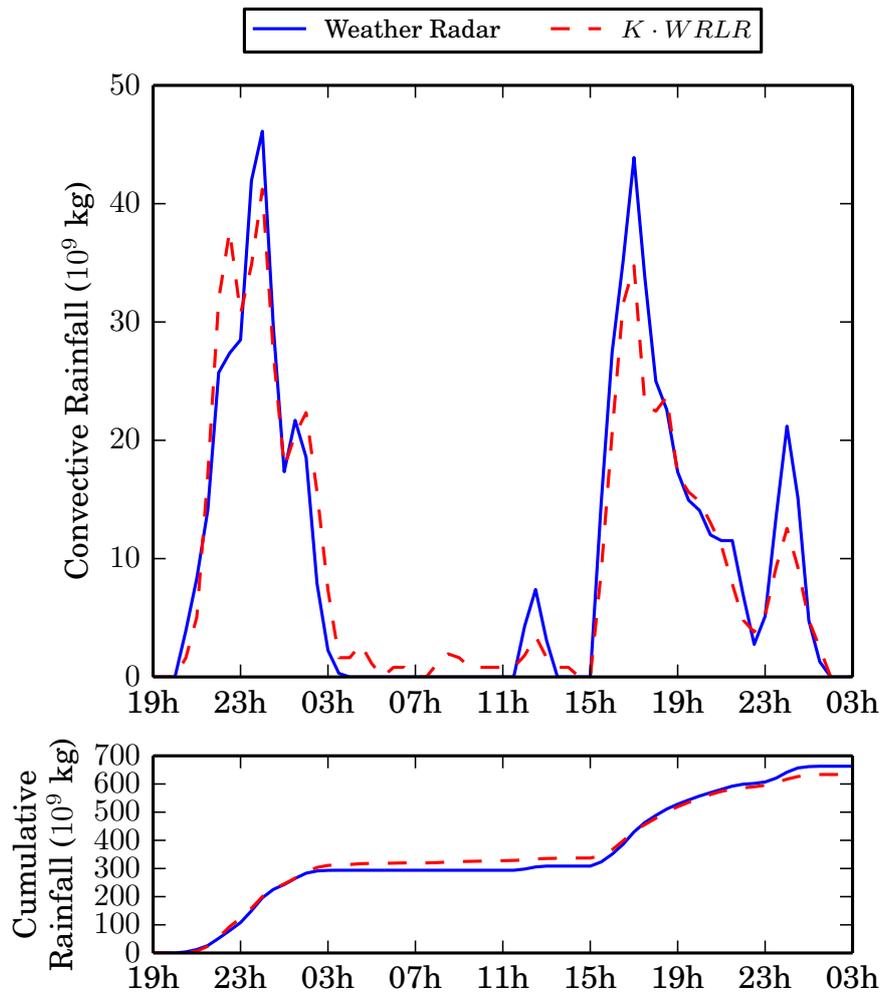


Figure 5.27 - PPR-4A thunderstorm occurred in 22nd April 2010. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.

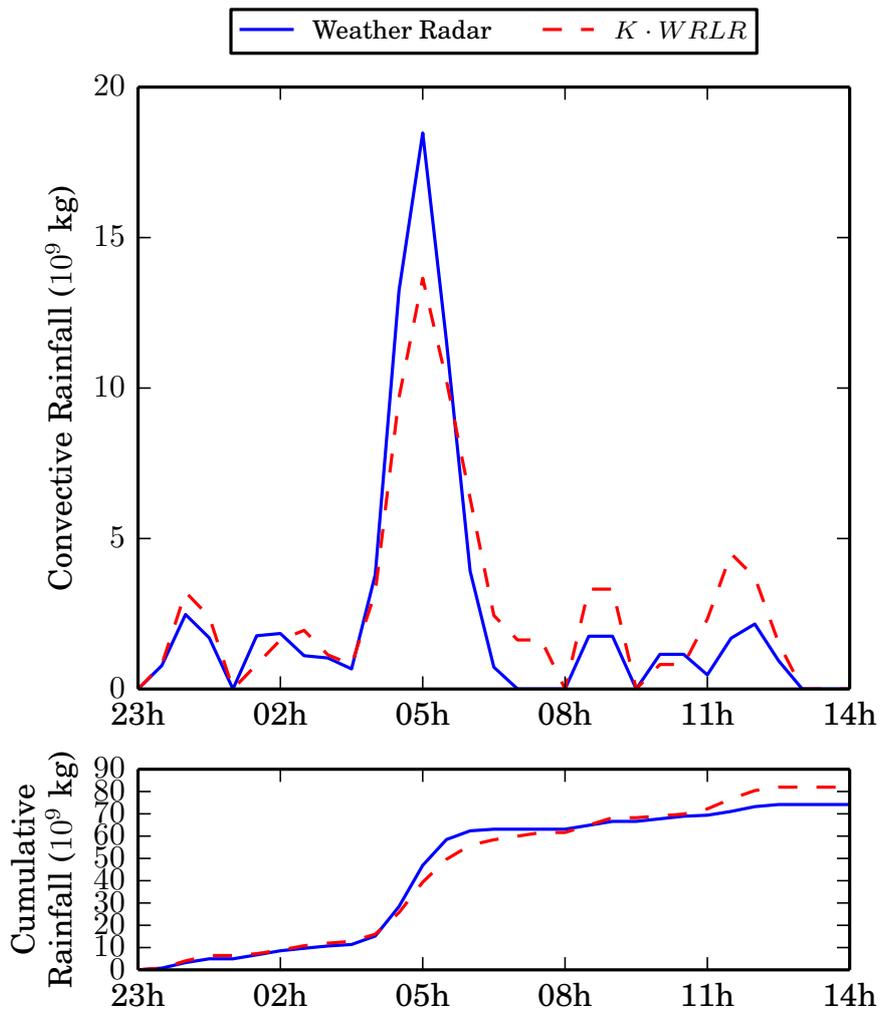


Figure 5.28 - PPR-5A thunderstorm occurred in 1st April 2010. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.

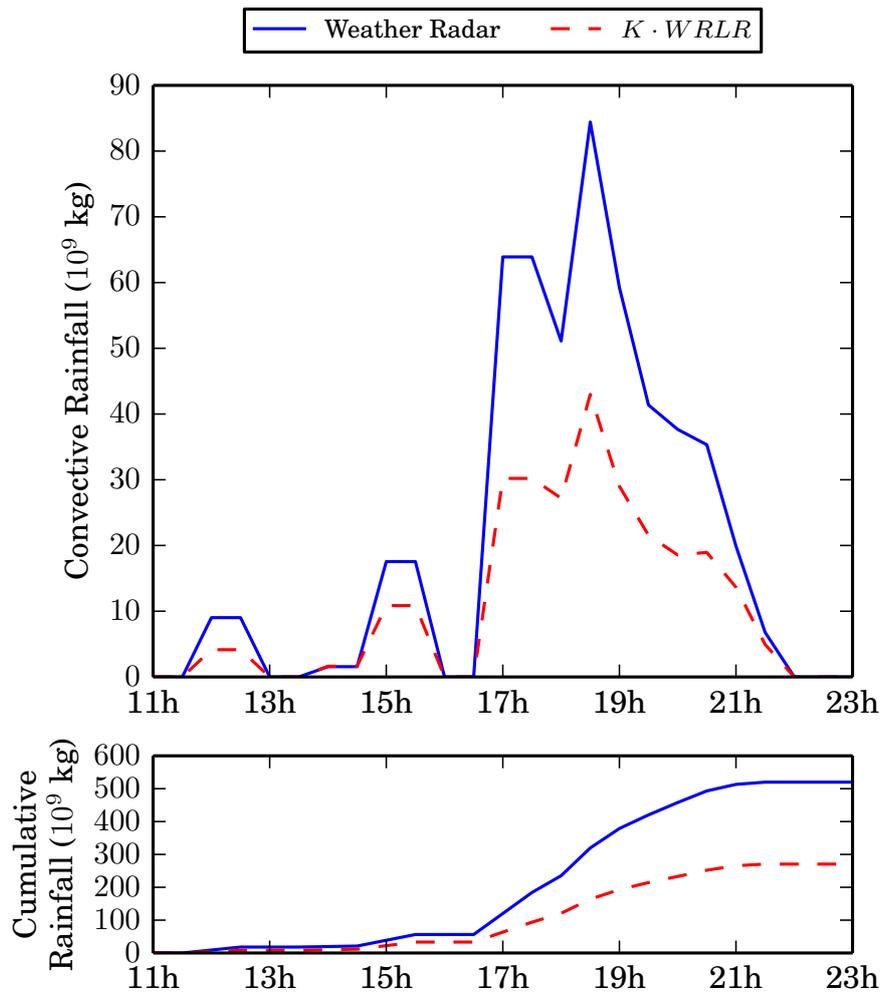


Figure 5.29 - PPR-6A thunderstorm occurred in 14th May 2009. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.

Estimation performance for the three Winter thunderstorms of Presidente Prudente:

The three selected Winter thunderstorms, PPR-7W (Figure 5.30), PPR-8W (Figure 5.31), and PPR-9W (Figure 5.32), are shown bellow. While WRLR heavily underestimates the PPR-9W thunderstorm, it presents a good match for PPR-7W and PPR-8W, which are relatively weaker than PPR-9W. Thunderstorms in Winter tended do last longer and to have significantly less electrical activity compared thunderstorms with similar total rainfall in other seasons.

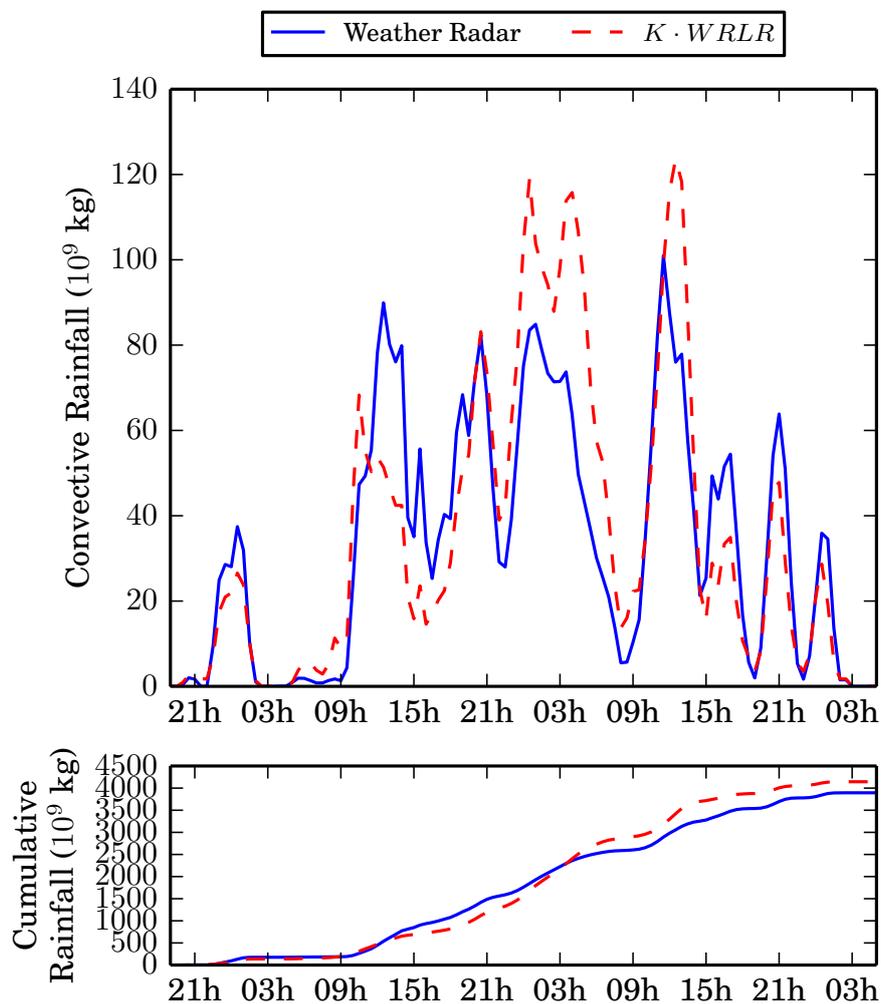


Figure 5.30 - PPR-7W thunderstorm occurred in 17th August 2009. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.

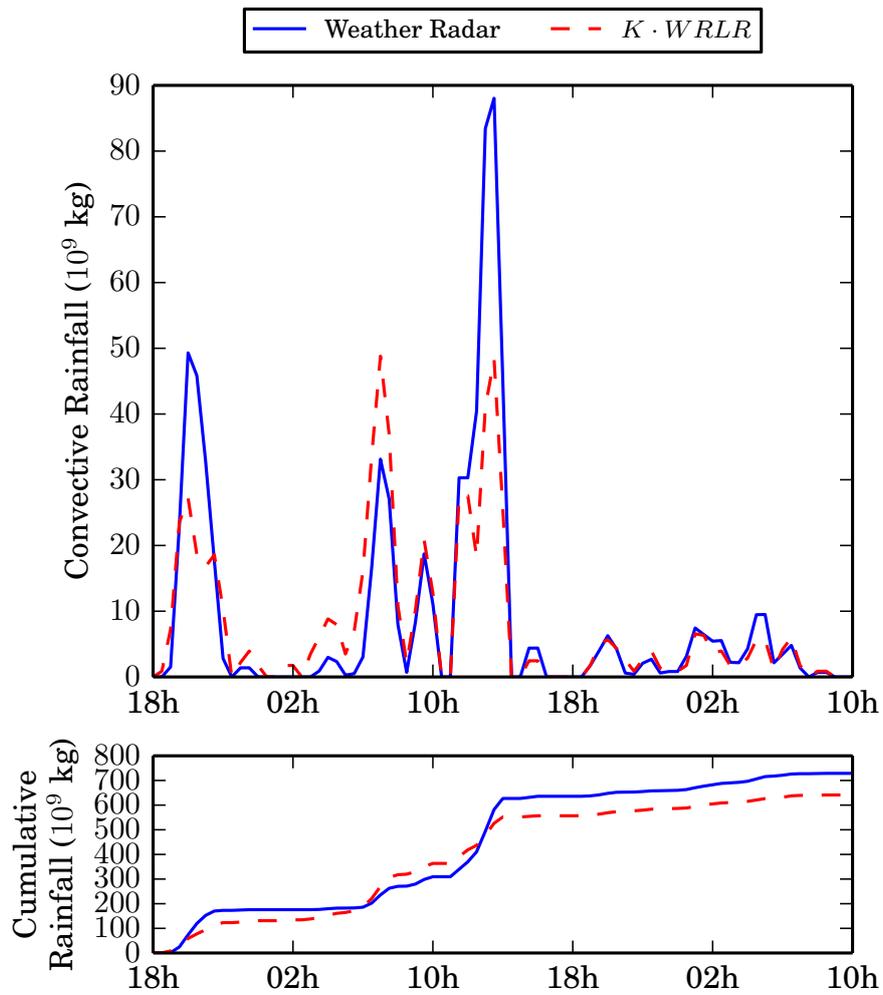


Figure 5.31 - PPR-8W thunderstorm occurred in 22nd July 2009. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.

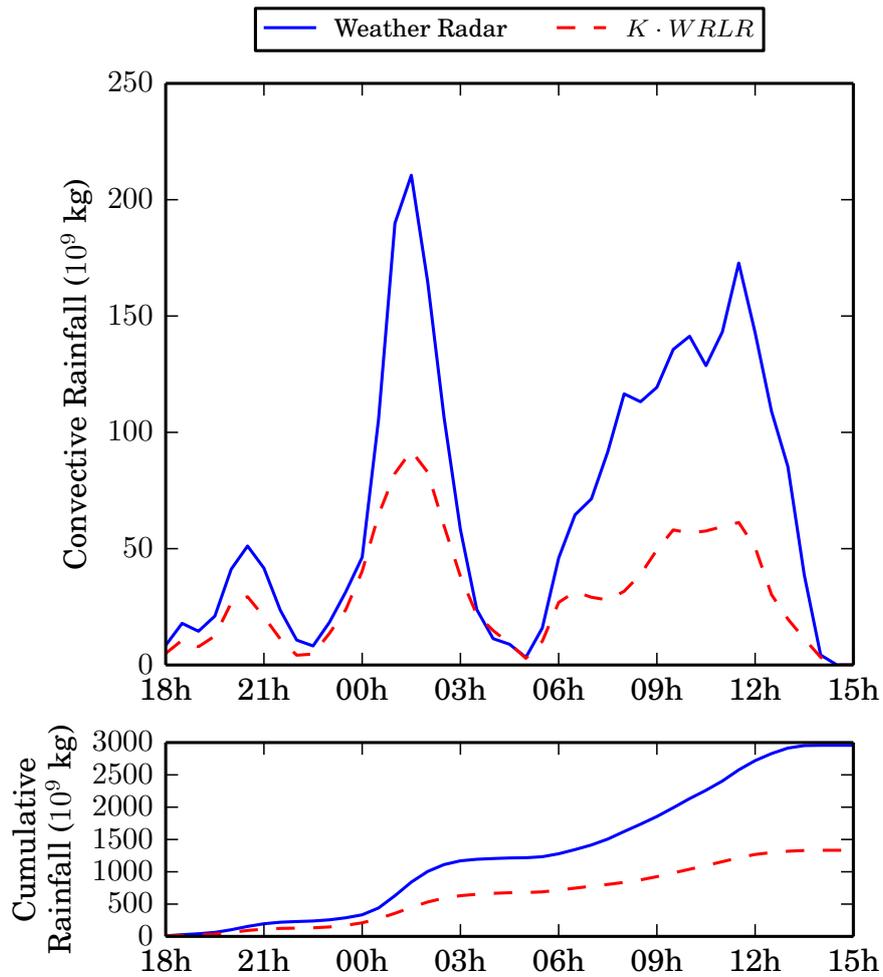


Figure 5.32 - PPR-9W thunderstorm occurred in 27th September 2010. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.

Estimation performance for the three Spring thunderstorms of Presidente Prudente:

The three selected Spring thunderstorms, PPR-10R (Figure 5.33), PPR-11R (Figure 5.34), and PPR-12R (Figure 5.35), are shown bellow. The WRLR underestimates PPR-10R and PPR-11R slightly, while PPR-12R is more severely underestimated. It is important to notice that only three months of training data were available for the Spring, which resulted in a less accurate WRLR function for this season.

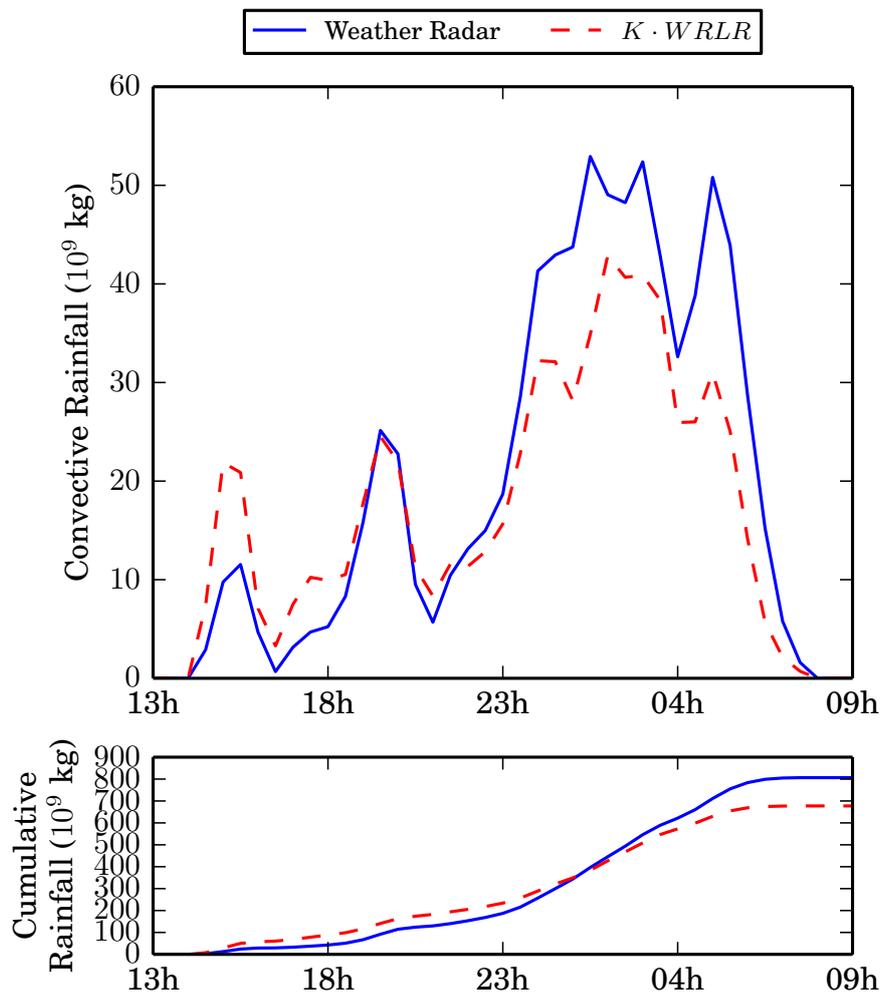


Figure 5.33 - PPR-10R thunderstorm occurred in 20th November 2009. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.

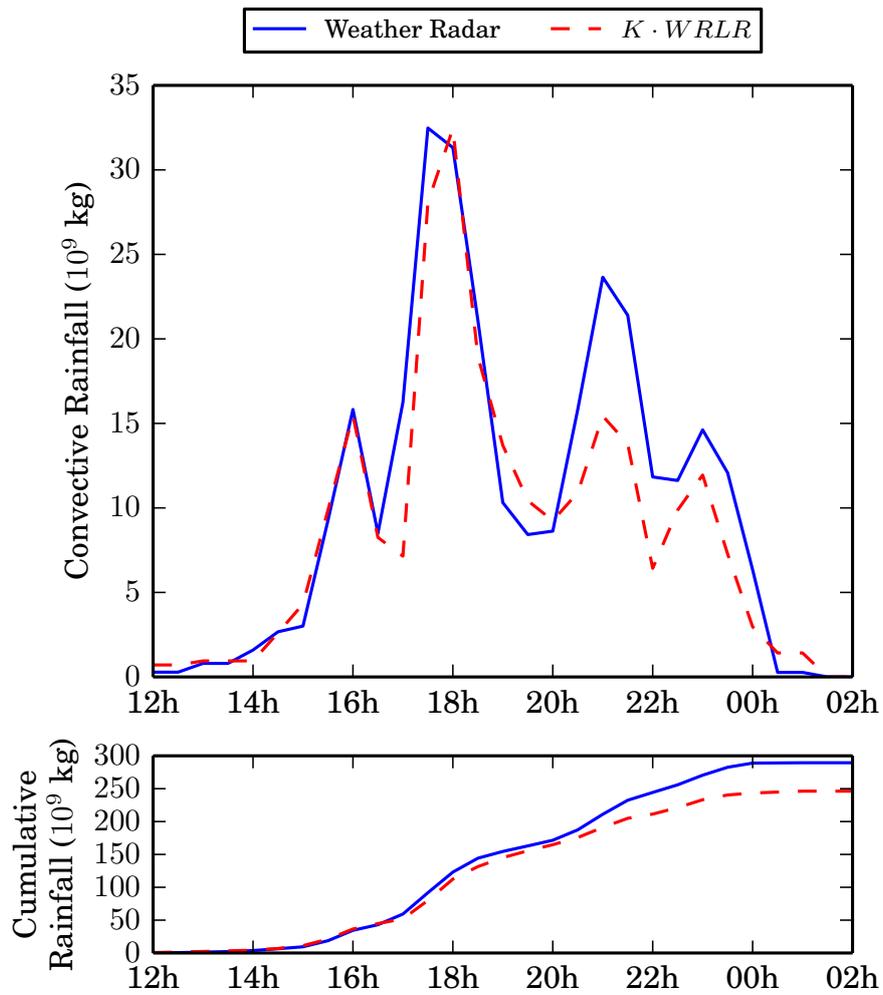


Figure 5.34 - PPR-11R thunderstorm occurred in 11th December 2009. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.

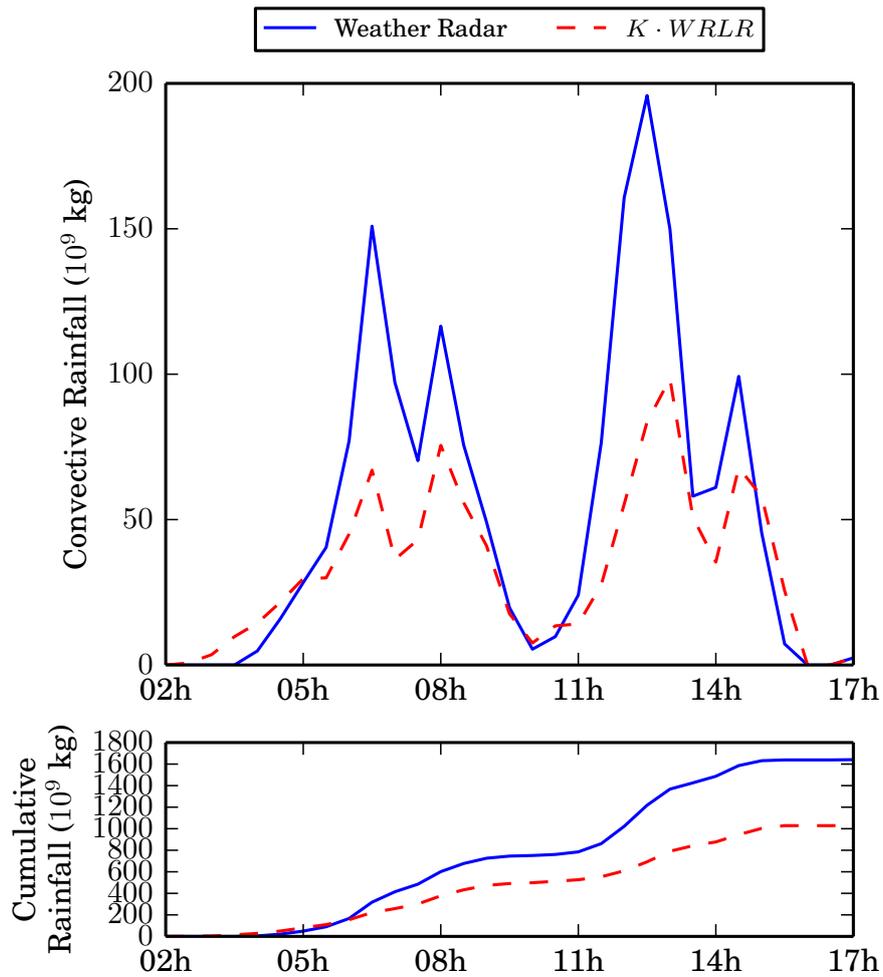


Figure 5.35 - PPR-12R thunderstorm occurred in 15th October 2009. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.

5.4.3 Validations tests using the weather radar of São Roque

This section presents the validation tests were performed considering a squared area with edge of 2.0° centered at the São Roque radar, as described in Figure 4.4, bounded by latitudes $24^\circ 30' 00''\text{S}$ to $22^\circ 30' 00''\text{S}$ and longitudes $48^\circ 00' 00''\text{W}$ to $46^\circ 00' 00''\text{W}$, for three months of the Summer of 2013, January, February and March. This area was divided into 16 squares of edge $\Delta x = 0.5^\circ$, as shown in Figure 4.5 (a), while the window duration was $\Delta t = 30$ min with window advance of $\delta t = 30$ min. The WRLR function was mapped as described in Section 4.5, by means of the specific elements of the K matrix.

A total of 75,071 lightning strokes occurred in this area for the considered period, radar estimation for the cumulative convective rainfall mass was 3288×10^9 kg, while the estimation given by the mapped WRLR function estimation was about 5% higher, 3462×10^9 kg. The mean error for the daily accumulated rainfall was an overestimation of 25%, with the median presenting an underestimation of 3%. These estimation errors were higher than those of the validation tests performed using data from the Bauru and Presidente Prudente weather radars, in the previous sections. Figure 5.6 compares the radar and mapped WRLR function estimations for the daily accumulated rainfall mass over these three Summer months (top). The cumulative rainfall mass is also shown (bottom).

Besides the estimation for the daily accumulated convective rainfall mass, three particular thunderstorms were selected, being denoted as RQ-1, RQ-2 and RQ-3 and shown, respectively, in Figures 5.37, 5.38 and 5.39. Table 5.6 presents for each thunderstorm the number of CG lightning strokes, the corresponding values of convective rainfall mass estimated by the weather radar and by the mapped WRLR function, the estimation errors. It also includes the correlation between the mapped WRLR function and the radar values for 30-min accumulated convective rainfall mass, also for each thunderstorm. Thunderstorm duration is also included in the table.

The current method of estimating the K matrix (Section 4.5) is not accurate for regions outside the continental Brazil. Thus, estimations over the ocean are not reliable. A large part of thunderstorm RQ-3 occurred over the ocean, resulting in a largely overestimated rainfall. On the contrary, thunderstorms RQ-1 and RQ-2 that occurred overland resulted in better rainfall estimations.

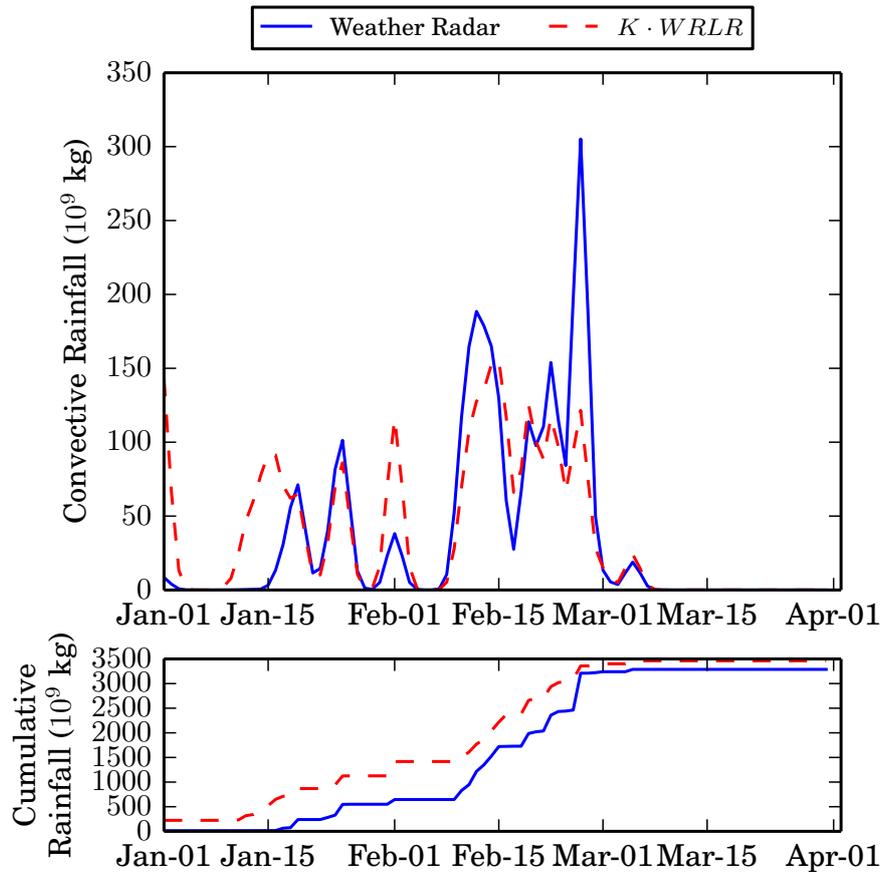


Figure 5.36 - Rainfall for the months of January to March of 2013 in São Roque. The top image shows the daily accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.

Table 5.6 - Comparison of the estimations for the convective rainfall mass given by the São Roque weather radar and by the mapped WRLR function for the three selected thunderstorms. Percentage *error* is calculated in relation to the radar estimation, while the correlation (*corr*) considers 30 min accumulated values for each thunderstorm.

Storm	Figure	CG Strokes	Rainfall in 10^9 kg	WRLR in 10^9 kg	<i>error</i>	<i>corr</i>	Duration
RQ-1	5.37	5228	437	363	+17%	0.77	10h
RQ-2	5.38	6422	391	354	-9%	0.91	10h
RQ-3	5.39	11798	1497	541	-64%	0.79	12h

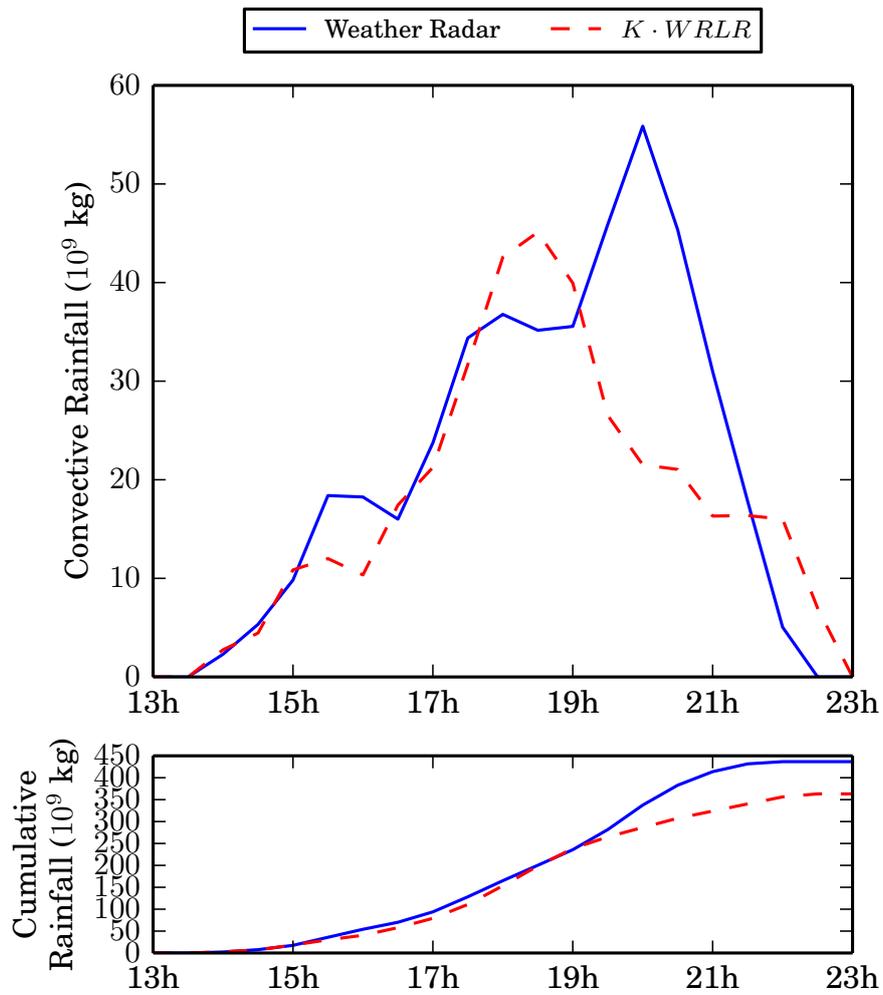


Figure 5.37 - RQ-1 thunderstorm occurred in 25th January 2013. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.

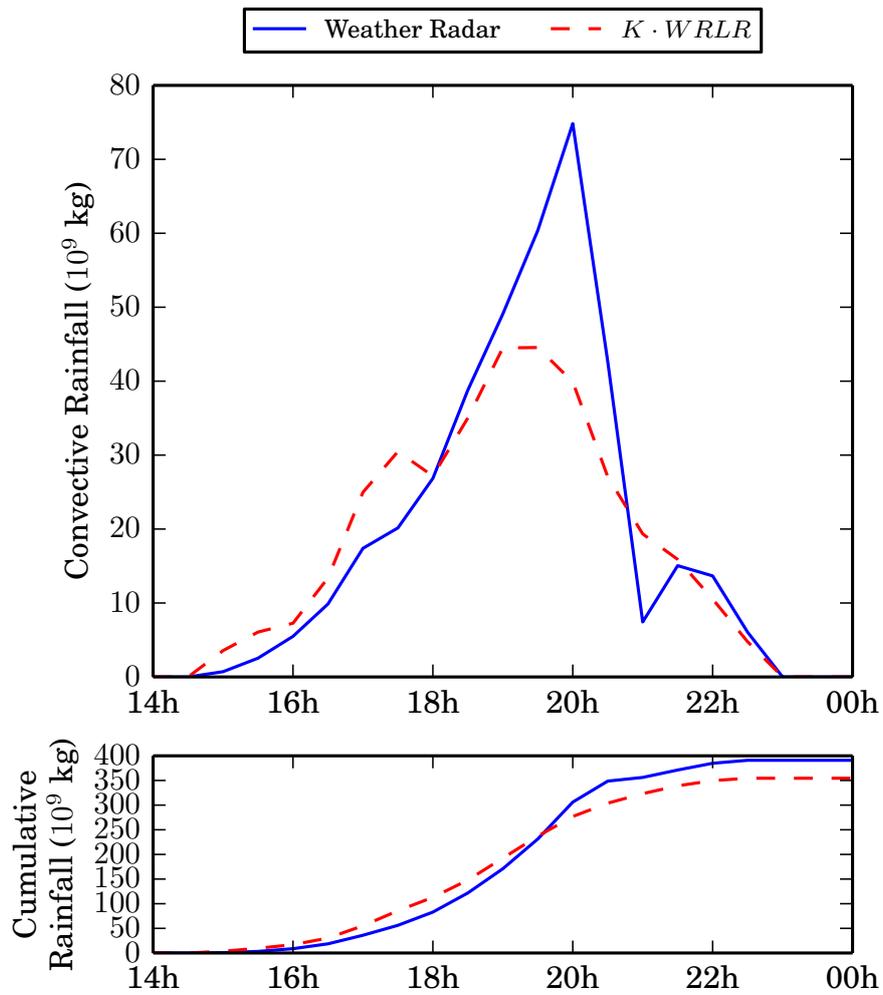


Figure 5.38 - RQ-2 thunderstorm occurred in 15th February 2013. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.

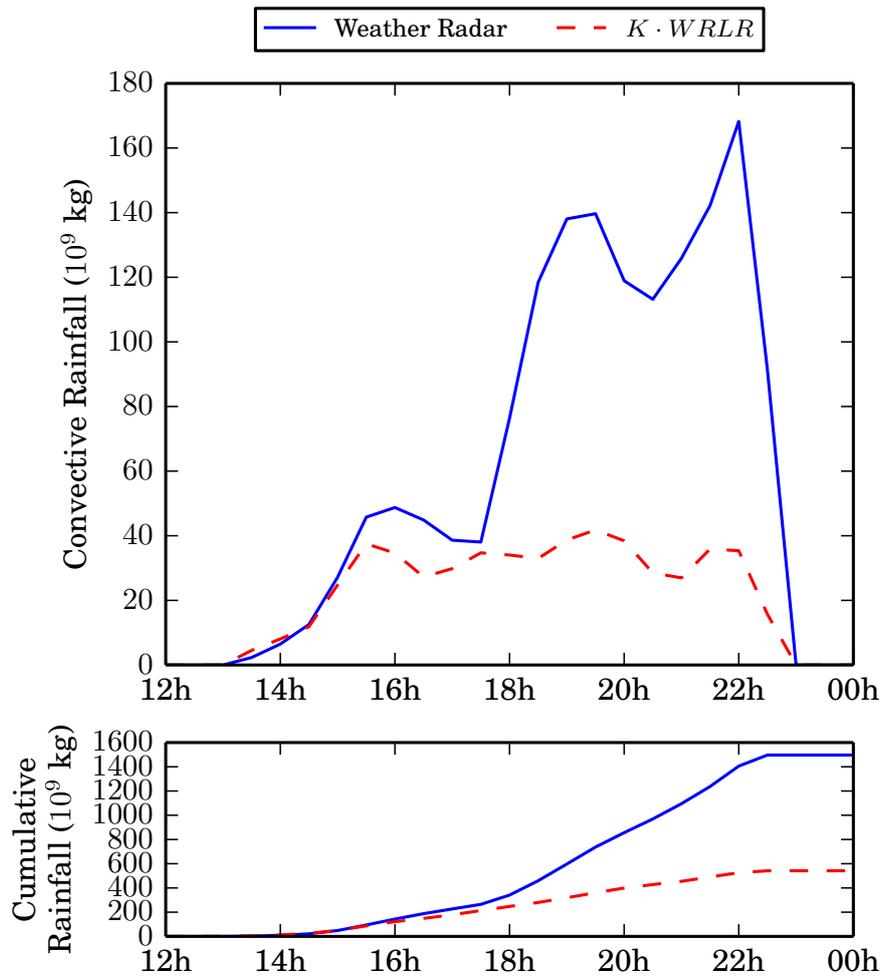


Figure 5.39 - RQ-3 thunderstorm occurred in 26th February 2013. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.

5.4.4 Validations tests using the weather radar of Pico do Couto

This section presents the validation tests were performed considering a squared area with edge of 2.0° centered at the Pico do Couto radar, as described in Figure 4.4, bounded by latitudes $23^\circ 30' 00''\text{W}$ to $21^\circ 30' 00''\text{W}$ and longitudes $44^\circ 30' 00''\text{S}$ to $42^\circ 30' 00''\text{S}$, for three months of the Summer of 2013, January, February and March. This area was divided into 16 squares of edge $\Delta x = 0.5^\circ$, as shown in Figure 4.5 (a), while the window duration was $\Delta t = 30$ min with window advance of $\delta t = 30$ min. The WRLR function was mapped as described in Section 4.5, by means of the specific elements of the K matrix.

A total of 45,634 lightning strokes occurred in this area for the considered period, radar estimation for the cumulative convective rainfall mass was 1949×10^9 kg, while the estimation given by the mapped WRLR function estimation was about 43% higher, 2795×10^9 kg. The mean error for the daily accumulated rainfall was an overestimation of 67%, with the median presenting an overestimation of 56%. However, in January the mean is an overestimation of 25%, and the median an overestimation of 11%. These estimation errors were higher than those of the validation tests performed using data from the Bauru and Presidente Prudente weather radars, in the previous sections, and even worst than those obtained in the previous section, using the São Roque weather radar. This lower estimation performance can be attributed to the ocean-land mixed coverage of the Pico do Couto radar, since isohyetal data only applies over land. Therefore, the corresponding elements of the K matrix cannot be properly calculated. This difference can be attributed to the fact the K matrix can't be properly mapped for oceanic areas as the isohyetal data only applies to land. Figure 5.40 compares the radar and mapped WRLR function estimations for the daily accumulated rainfall mass over these three Summer months (top). The cumulative rainfall mass is also shown (bottom).

Besides the estimation for the daily accumulated convective rainfall mass, three particular thunderstorms were selected, being denoted as PI-1, PI-2 and PI-3 and shown, respectively, in Figures 5.41, 5.42 and 5.43. Table 5.7 presents for each thunderstorm the number of CG lightning strokes, the corresponding values of convective rainfall mass estimated by the weather radar and by the mapped WRLR function, the estimation errors. It also includes the correlation between the mapped WRLR function and the radar values for 30 min accumulated convective rainfall mass, also for each thunderstorm. Thunderstorm duration is also included in the table.

A significant part of the range of the Pico do Couto radar is over the ocean, which

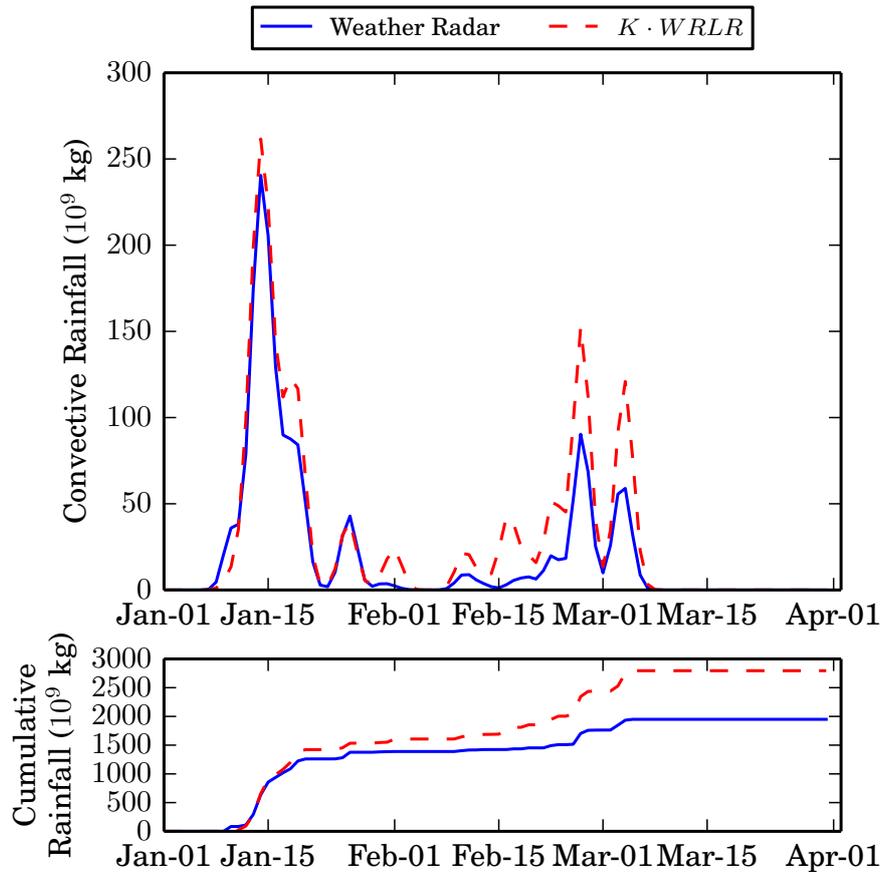


Figure 5.40 - Rainfall for the months of January to March of 2013 in Pico do Couto. The top image shows the daily accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.

Table 5.7 - Comparison of the estimations for the convective rainfall mass given by the Pico do Couto weather radar and by the mapped WRLR function for the three selected thunderstorms. Percentage *error* is calculated in relation to the radar estimation, while the correlation (*corr*) considers 30 min accumulated values for each thunderstorm.

Storm	Figure	CG Strokes	Rainfall in 10^9 kg	WRLR in 10^9 kg	RE	<i>corr</i>	Duration
PI-1	5.41	20282	683	723	+6%	0.91	10h
PI-2	5.42	2644	168	176	+5%	0.99	6h
PI-3	5.42	12016	378	603	+60%	0.99	10h

is affected by poorly estimated elements in the K (Section 4.5). This is the case of thunderstorm PI-3, that presents a large part over the ocean, and resulted in a poor rainfall estimation.

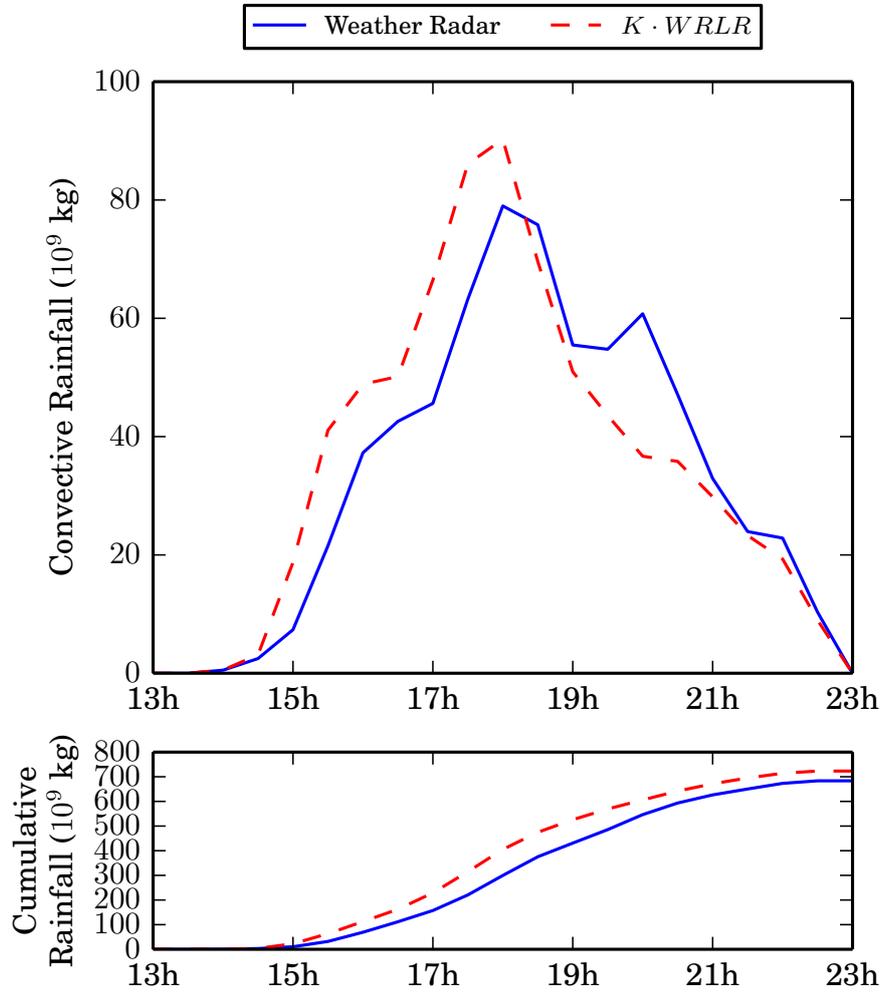


Figure 5.41 - PI-1 thunderstorm starting in 14th January 2013. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR function (in dashed red) for the convective rainfall mass.

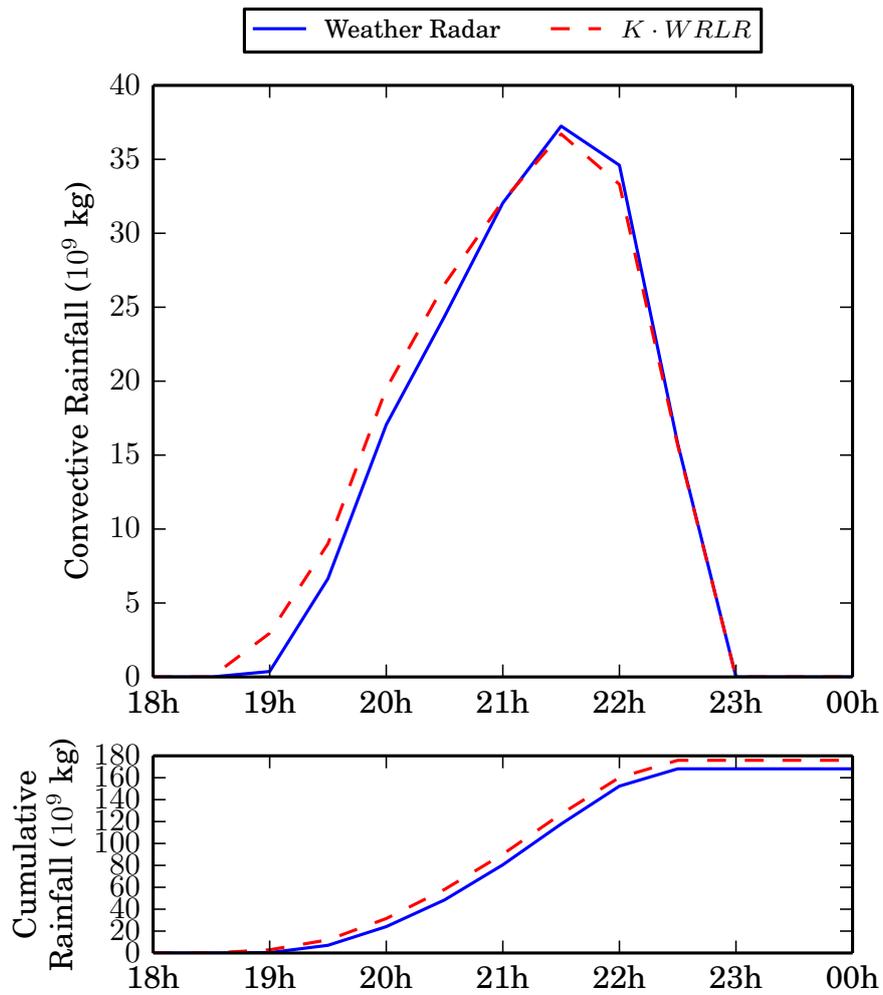


Figure 5.42 - PI-2 thunderstorm starting in 3rd March 2013. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR adjusted function (in dashed red) for the convective rainfall mass.

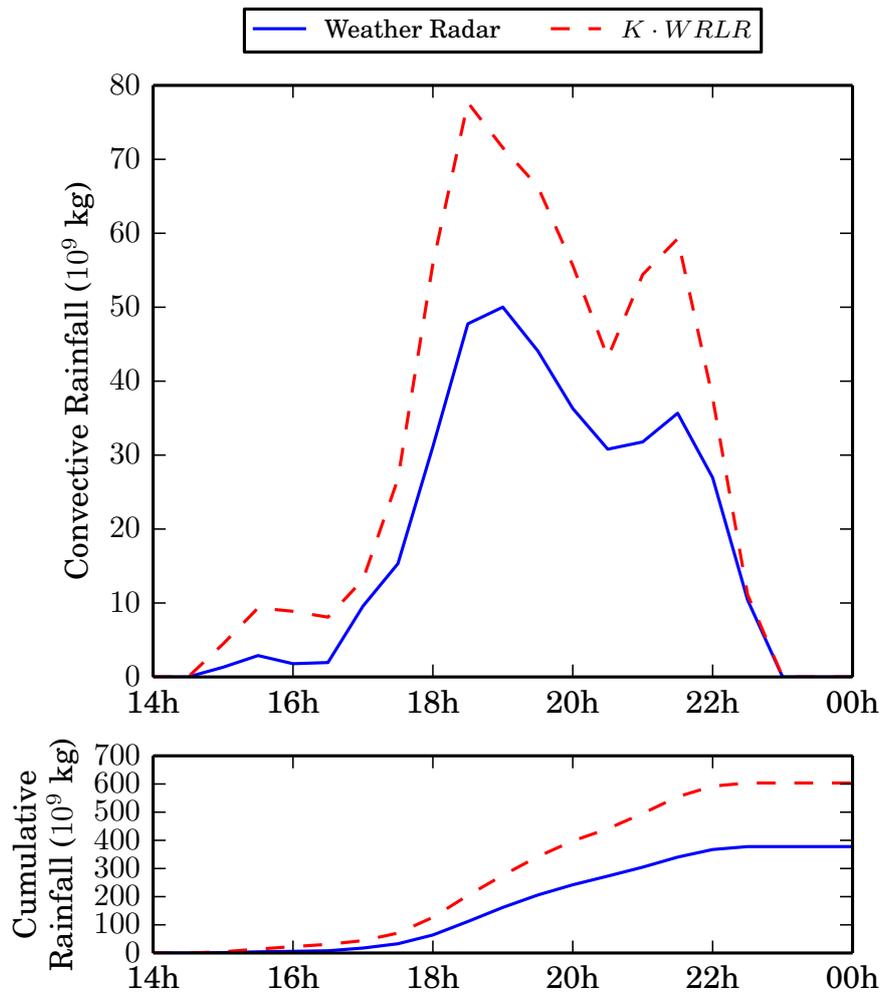


Figure 5.43 - PI-3 thunderstorm starting in 26th February 2013. The top image shows the 30 min accumulated rainfall mass while the bottom image shows the cumulative rainfall, comparing the estimations of the weather radar (in blue), and of the mapped WRLR adjusted function (in dashed red) for the convective rainfall mass.

5.5 Spatial distribution of convective/total rainfall and estimation of total rainfall

The previous sections presented validation tests concerning the estimation of the convective rainfall mass using different weather radars. The spatial distribution of the convective rainfall for these tests is assumed as given by the normalized field of density of occurrence of CG lightning strokes, as explained in Section 4.7 (EDDA-chuva software). The intensity of the normalized field is adjusted/multiplied by the amount of estimated convective rainfall for each $0.5^\circ \times 0.5^\circ$ square, yielding a field of “density of rainfall” expressed in cumulative mm of convective rain for the considered period of time. It would be no suitable to generate such field with a high resolution, since convective rainfall and CG lightning strokes do not necessarily match in space.

Another issue is the estimation of the total rainfall mass and spatial distribution from the corresponding convective rainfall mass and distribution, as presented in Section 4.6, and already implemented in the current version of the EDDA-chuva software. The spatial distribution of the estimated convective rainfall could have been discussed in Section 5.4, but it seemed more suitable to show such distribution in this section, jointly with the total rainfall distribution.

The approach for estimating the total rainfall is to estimate the convective rainfall mass from the number of occurrences of CG lightning strokes for the considered area and duration and, to assume that the total rainfall mass would be the double, corresponding to the associated stratiform rainfall and the associated convective rainfall without lightning. Spatial distribution of convective or total rainfall were also discussed in Section 4.7, assuming that the area covered by the convective rainfall with lightning corresponds to about 30% of the total rainfall area. Although CG lightning and convective rainfall are not completely correlated in space and time, the proposed approach provide a good approximation for the spatial distribution of convective rainfall, as discussed in the following results.

The EDDA-chuva outputs hourly-accumulated images for the WRLR-estimated rainfall obtained by integrating two successive 30 min temporal windows, and may also generate accumulated rainfall for longer periods. The tests related to the estimation of the total rainfall and its spatial distribution (as well as such distribution for the convective rainfall) were performed for four selected thunderstorms whose data was employed in the validation tests of Section 5.4. These storms are BRU-1, PPR-1S, RQ-1, and PI-1, described in Table 5.8, while their considered spatial boundaries are described in Table 5.9. The cumulative rainfall for those thunderstorms

was previously shown on Figures 5.10, 5.24, 5.37 and 5.41, respectively.

Table 5.8 - Metadata of the four thunderstorms employed in the rainfall spatial distribution tests: starting time and extension of time for the cumulative rainfall images.

Storm	Starting Date	Time	Images	
			1h	6h
BRU-1	1st January 2013	15:00	5.44	5.48
PPR-1S	11th January 2009	03:00	5.45	5.49
RQ-1	25th January 2013	14:00	5.46	5.50
PI-1	14th January 2013	15:00	5.47	5.51

Table 5.9 - Boundaries of the four thunderstorms employed in the rainfall spatial distribution tests.

Storm	Westernmost Longitude	Easternmost Longitude	Southernmost Latitude	Northern Latitude
BRU-1	50°00'00"W	48°04'00"W	23°15'00"S	21°27'00"S
PPR-1S	52°21'00"W	50°24'00"W	23°04'00"S	21°16'00"S
RQ-1	48°00'00"W	46°00'00"W	24°30'00"S	22°30'00"S
PI-1	44°30'00"S	42°30'00"S	23°30'00"W	21°30'00"W

The following spatial distributions were generated for each one of the four selected thunderstorms in order to evaluate the distributions estimated by the EDDA-chuva software: (i) total rainfall estimated from weather radar data (“unfiltered radar”), (ii) convective rainfall using Steiner criteria from radar data (“filtered radar”), (iii) EDDA-chuva convective rainfall (“EDDA-chuva (C)”), and (iv) EDDA-chuva total rainfall (“EDDA-chuva (T)”) In the case of the BRU-1 thunderstorm, the rainfall spatial distribution obtained by the Tapia’s model is also shown.

Figures 5.44, 5.45, 5.46, and 5.47 show the corresponding rainfall distributions for the four selected thunderstorms BRU-1, PPR-1S, RQ-1 and PI-1 respectively, considering the first 1 hour of accumulated rainfall. Figures 5.48, 5.49, 5.50, and 5.51 are similar, but considering the first 6 hours of accumulated rainfall. In general, the rainfall distributions generated by the EDDA-chuva software for the convective rainfall are more similar to the corresponding distributions generated from radar data than those obtained by the Tapia’s model, since the spatial distributions generated

by the software from the fields of density to CG lightning strokes occurrences are smoother. Even in the case of isolated lightning occurrences, the software outperforms the Tapia’s model. It is worth to mention that some distinct CG lightning strokes that serve of input to the EDDA-chuva software may be part of a single CG lightning flash that serves of input to the Tapia’s model, as previously discussed in Section 4.2.1. The spatial distributions obtained for 6 h of accumulated rainfall tend to be more similar to the corresponding weather radar images than those obtained for only 1 h of accumulated rainfall. Finally, the spatial distributions of the total rainfall generated by the EDDA-chuva software, not provided by the Tapia’s model, were less accurate, i.e. less similar than the corresponding radar images in comparison to the ones related to the convective rainfall, as it would be expected.

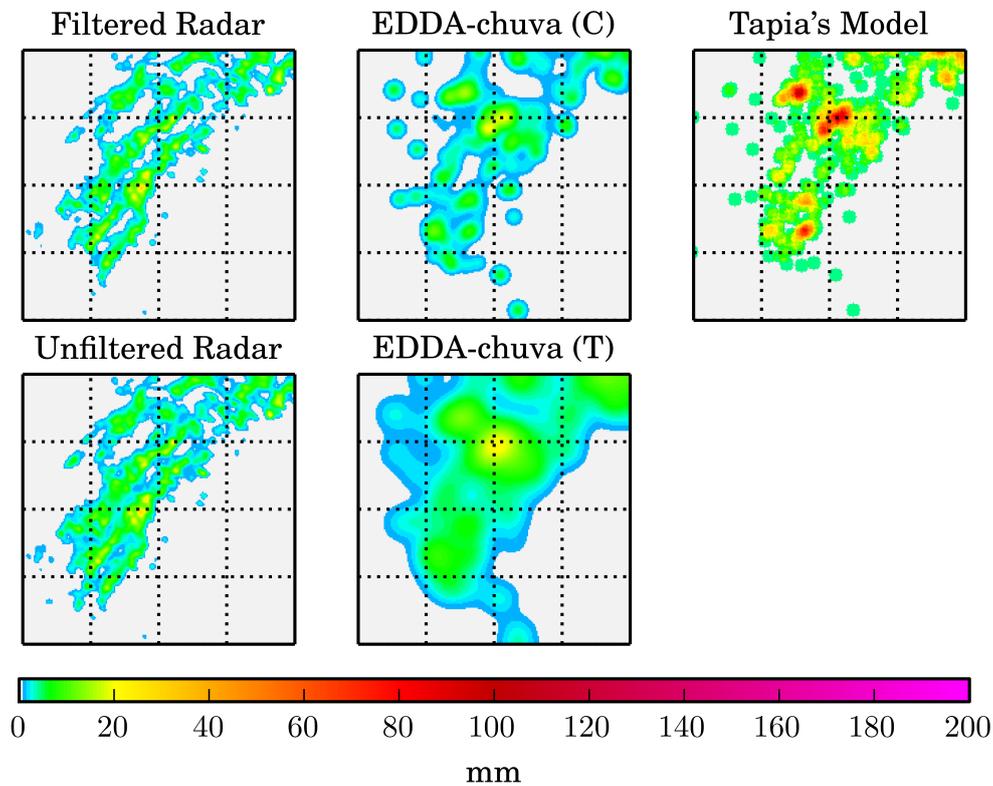


Figure 5.44 - Rainfall spatial distributions accumulated for the first hour for thunderstorm BRU-1 given by (i) weather radar for convective rainfall (“filtered radar”), (ii) EDDA-chuva for convective rainfall (“EDDA-chuva (C)”), (iii) Tapia’s model for convective rainfall, (iv) weather radar for total rainfall (“unfiltered radar”), and (v) EDDA-chuva for total rainfall (“EDDA-chuva (T)”).

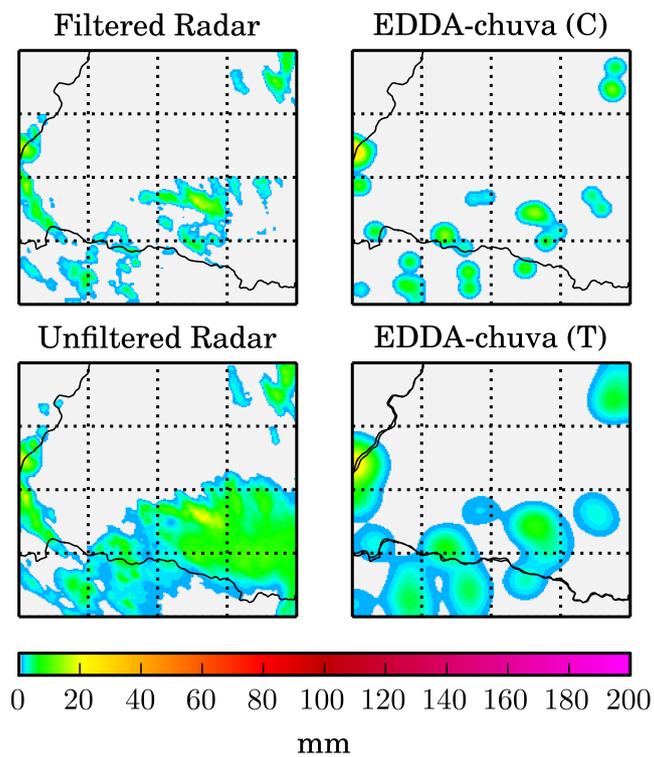


Figure 5.45 - Rainfall spatial distributions accumulated for the first hour for thunderstorm PPR-1S given by (i) weather radar for convective rainfall (“filtered radar”), (ii) EDDA-chuva for convective rainfall (“EDDA-chuva (C)”), (iii) weather radar for total rainfall (“unfiltered radar”), and (iv) EDDA-chuva for total rainfall (“EDDA-chuva (T)”).

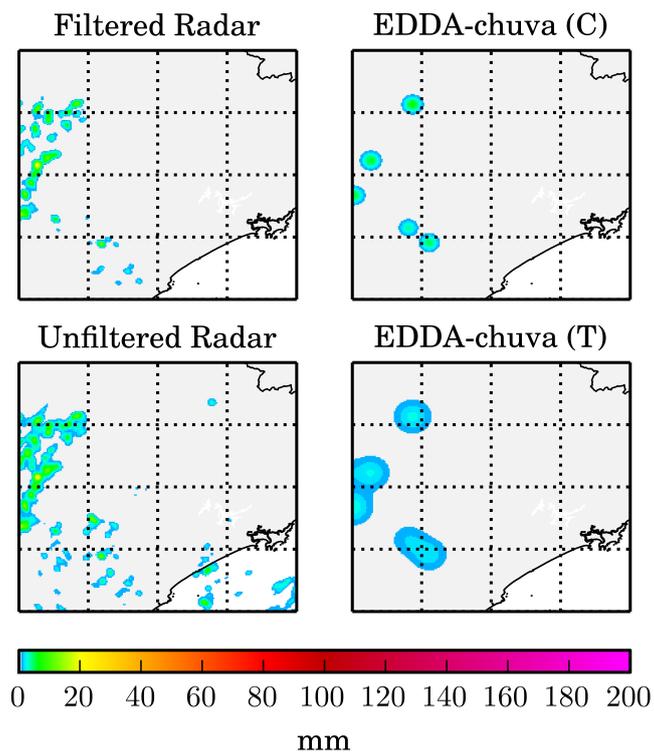


Figure 5.46 - Rainfall spatial distributions accumulated for the first hour for thunderstorm RQ-1 given by (i) weather radar for convective rainfall (“filtered radar”), (ii) EDDA-chuva for convective rainfall (“EDDA-chuva (C)”), (iii) weather radar for total rainfall (“unfiltered radar”), and (iv) EDDA-chuva for total rainfall (“EDDA-chuva (T)”).

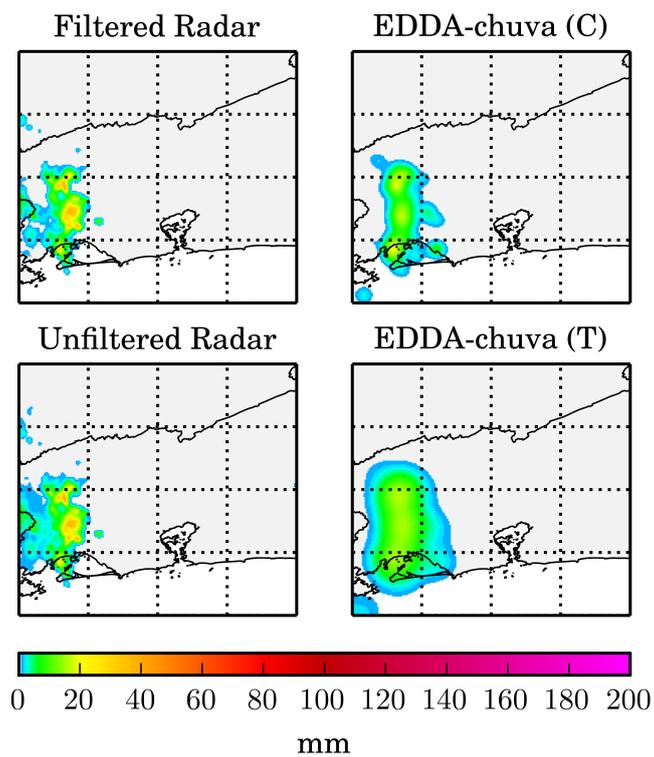


Figure 5.47 - Rainfall spatial distributions accumulated for the first hour for thunderstorm PI-1 given by (i) weather radar for convective rainfall (“filtered radar”), (ii) EDDA-chuva for convective rainfall (“EDDA-chuva (C)”), (iii) weather radar for total rainfall (“unfiltered radar”), and (iv) EDDA-chuva for total rainfall (“EDDA-chuva (T)”).

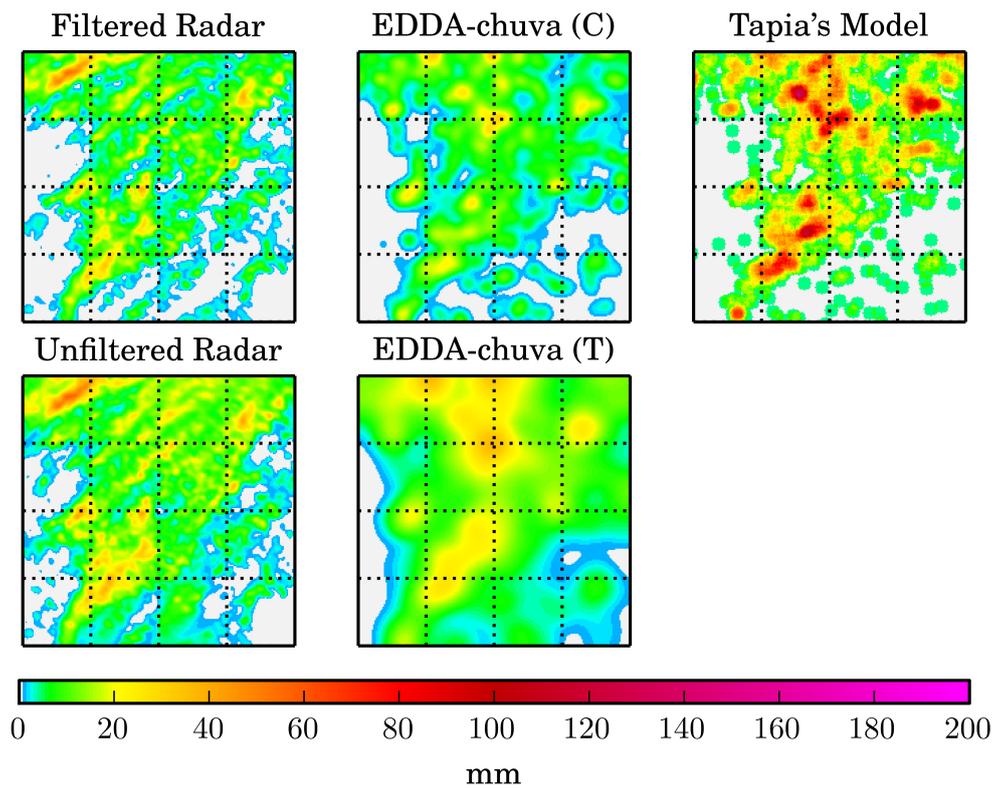


Figure 5.48 - Rainfall spatial distributions accumulated for the first six hours for thunderstorm BRU-1 given by (i) weather radar for convective rainfall (“filtered radar”), (ii) EDDA-chuva for convective rainfall (“EDDA-chuva (C)”), (iii) Tapia’s model for convective rainfall, (iv) weather radar for total rainfall (“unfiltered radar”), and (v) EDDA-chuva for total rainfall (“EDDA-chuva (T)”).

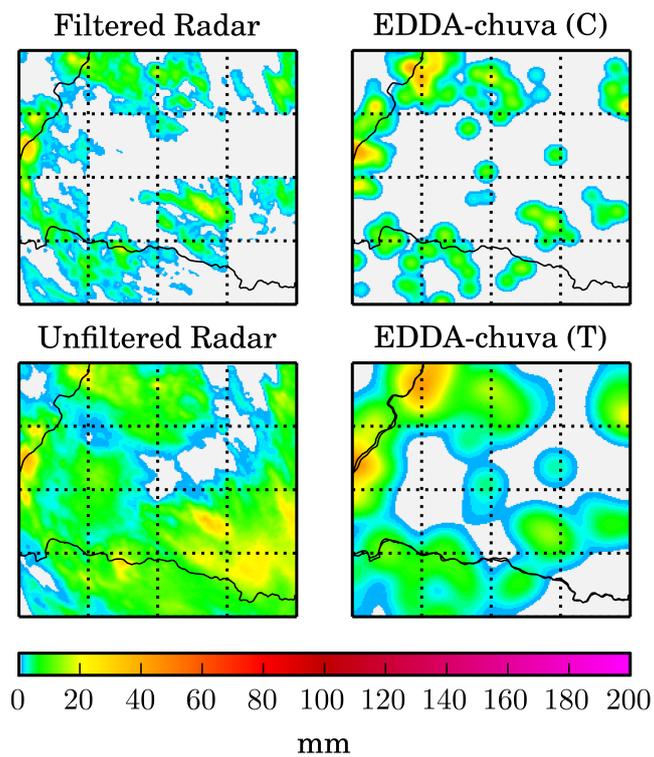


Figure 5.49 - Rainfall spatial distributions accumulated for the first six hours for thunderstorm PPR-1S given by (i) weather radar for convective rainfall (“filtered radar”), (ii) EDDA-chuva for convective rainfall (“EDDA-chuva (C)”), (iii) weather radar for total rainfall (“unfiltered radar”), and (iv) EDDA-chuva for total rainfall (“EDDA-chuva (T)”).

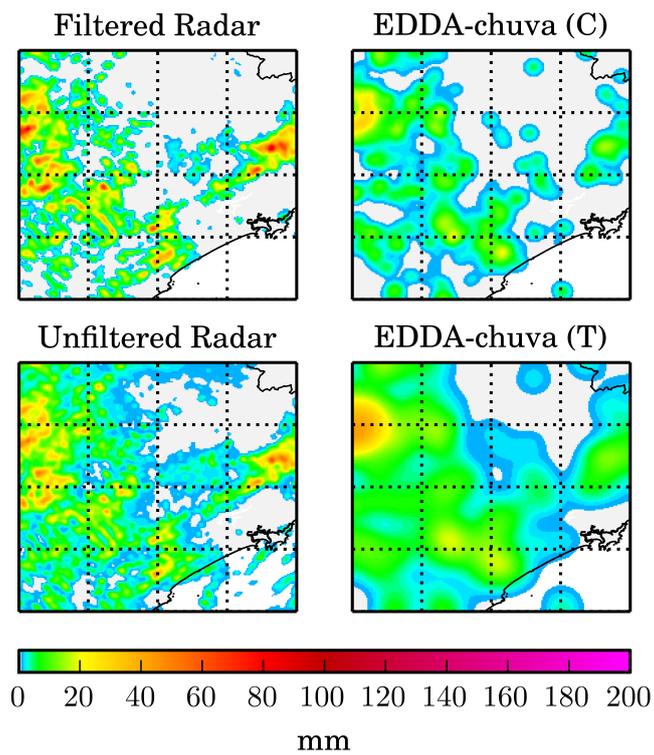


Figure 5.50 - Rainfall spatial distributions accumulated for the first six hours for thunderstorm RQ-1 given by (i) weather radar for convective rainfall (“filtered radar”), (ii) EDDA-chuva for convective rainfall (“EDDA-chuva (C)”), (iii) weather radar for total rainfall (“unfiltered radar”), and (iv) EDDA-chuva for total rainfall (“EDDA-chuva (T)”).

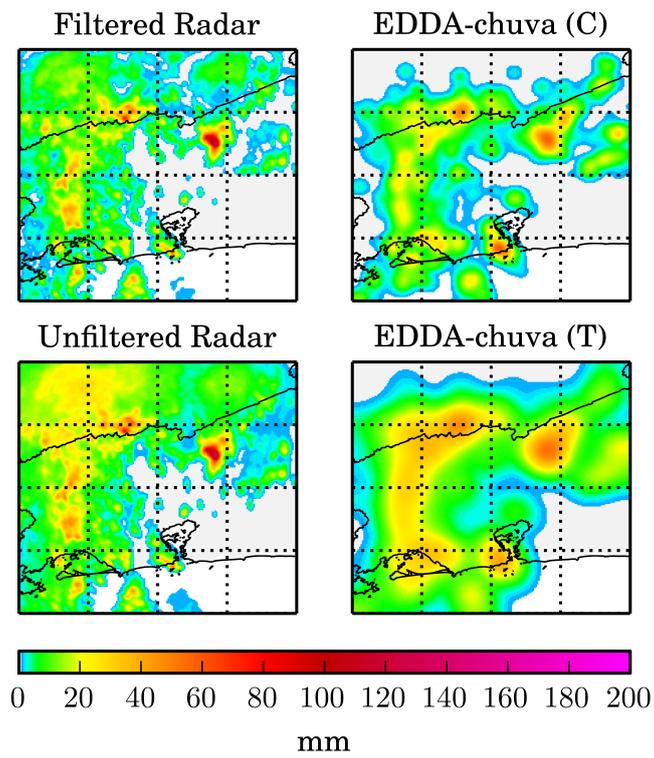


Figure 5.51 - Rainfall spatial distributions accumulated for the first six hours for thunderstorm PI-1 given by (i) weather radar for convective rainfall (“filtered radar”), (ii) EDDA-chuva for convective rainfall (“EDDA-chuva (C)”), (iii) weather radar for total rainfall (“unfiltered radar”), and (iv) EDDA-chuva for total rainfall (“EDDA-chuva (T)”).

5.5.1 Spatial distribution tests including MERGE data

The proposed approach for rainfall estimation from lightning data is based on the WRLR function, that was derived using quarterly averages of rainfall from the isohyetal data described in Section 2.5. Since the MERGE rainfall data, as exposed in Section 2.6 is considered more accurate than the isohyetal data, this section compares the 24 h accumulated rainfall spatial distribution for the same four selected thunderstorms, in order to compare to the distribution given by MERGE for the same day. The rainfall spatial distributions obtained by the EDDA-chuva software above (Section 5.5) referred to 1 h and 6 h accumulated rainfall, but considering the starting time of each thunderstorms. On the other hand, the following distributions obtained by the same software for 24 hours, start at 0:00 UTC and finish at 23:59 UTC, in order to conform to MERGE data for the same day.

The 24 h accumulated rainfall spatial distributions for the storms BRU-1, PPR-1S, RQ-1, and PI-1, are respectively shown in Figures 5.52, 5.53, 5.54, and 5.55, comparing the radar and EDDA-chuva rainfall distributions for the convective rain, and the radar, EDDA-chuva and MERGE rainfall distributions for the total rainfall. The EDDA-chuva spatial rainfall distributions are similar to the weather radar distributions considering the convective or the total rainfall, in different degrees, considering a “visual analysis”. EDDA-chuva distributions are more similar to the weather radar ones than the MERGE images. This would be expected, since the mapped WRLR function was estimated using weather radar data. Another issue is the poorer spatial resolution of the MERGE images.

This work proposes rainfall estimation based on the use of weather radar data. The rainfall estimated by radar corresponds, in this case, to the rainfall as seen at the altitude of the CAPPI image. Obviously, rainfall is distributed at a layer that encompasses a range of altitudes. In addition, not the entire rainfall “seen” by the radar reaches the ground. The use of radar volumetric data (3D data), when available, would demand much more processing time, but would probably provide a more accurate WRLR function. In the same way, the use of MERGE data, that include rain gauge data would allow to improve the accuracy of the K matrix that maps the WRLR function.

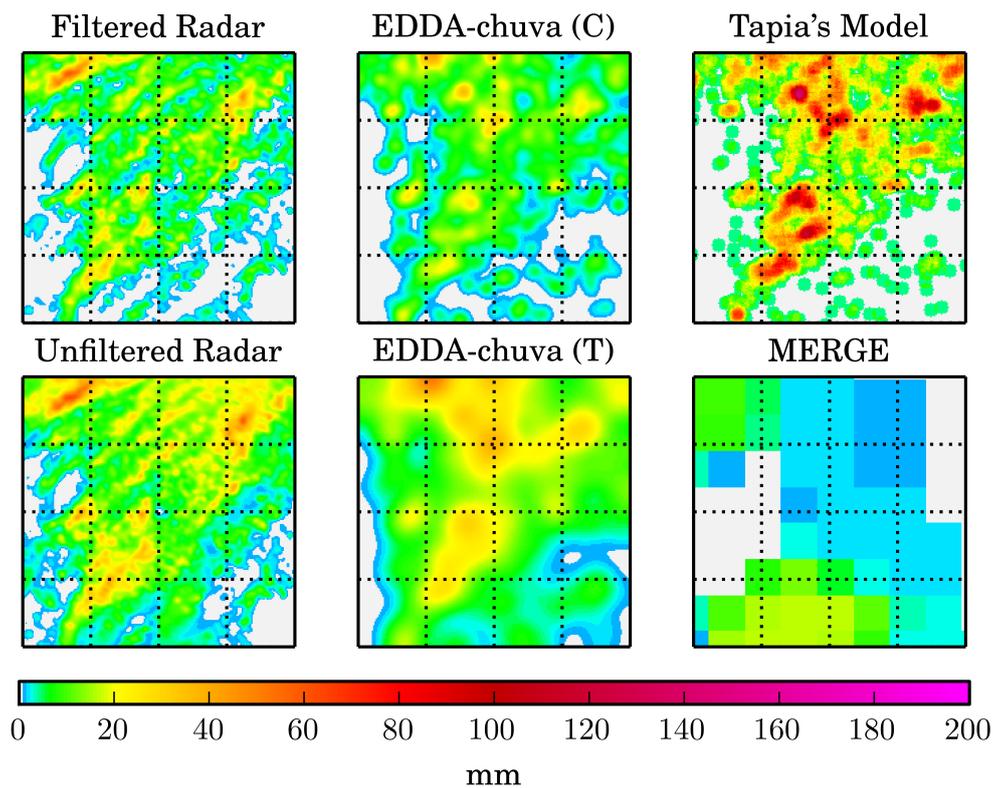


Figure 5.52 - Daily rainfall spatial distributions accumulated for thunderstorm BRU-1 given by (i) weather radar for convective rainfall (“filtered radar”), (ii) EDDA-chuva for convective rainfall (“EDDA-chuva (C)”), (iii) Tapia’s model for convective rainfall, (iv) weather radar for total rainfall (“unfiltered radar”), (v) EDDA-chuva for total rainfall (“EDDA-chuva (T)”), and (vi) MERGE data.

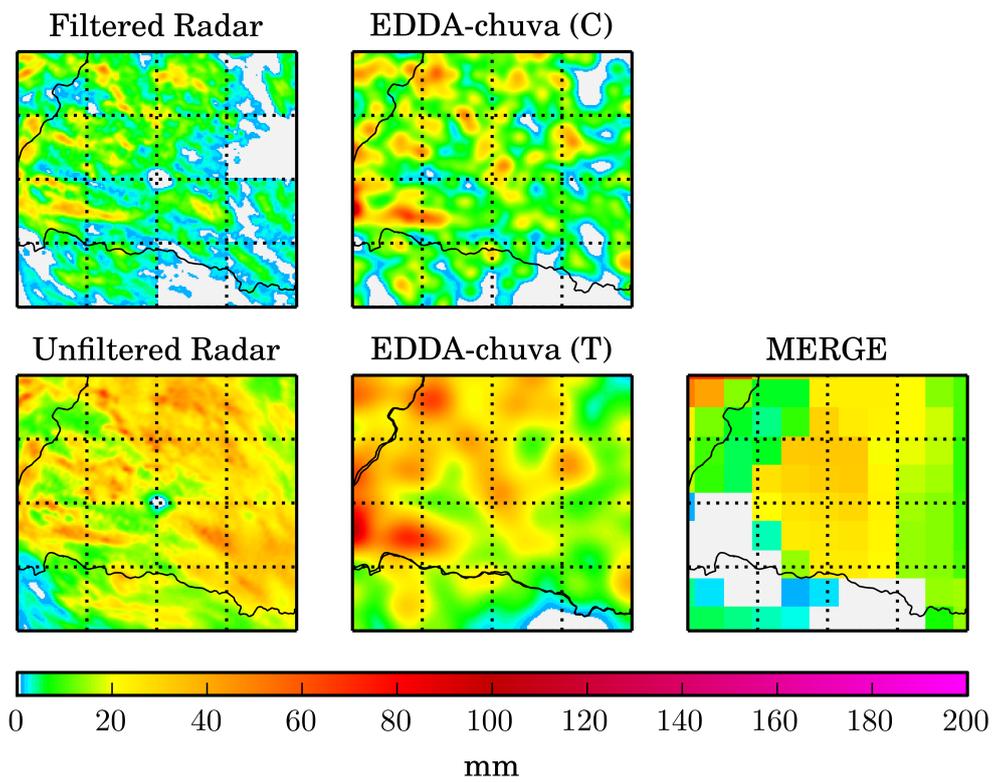


Figure 5.53 - Daily rainfall spatial distributions accumulated for thunderstorm PPR-1S given by (i) weather radar for convective rainfall (“filtered radar”), (ii) EDDA-chuva for convective rainfall (“EDDA-chuva (C)”), (iii) weather radar for total rainfall (“unfiltered radar”), (iv) EDDA-chuva for total rainfall (“EDDA-chuva (T)”), and (v) MERGE data.

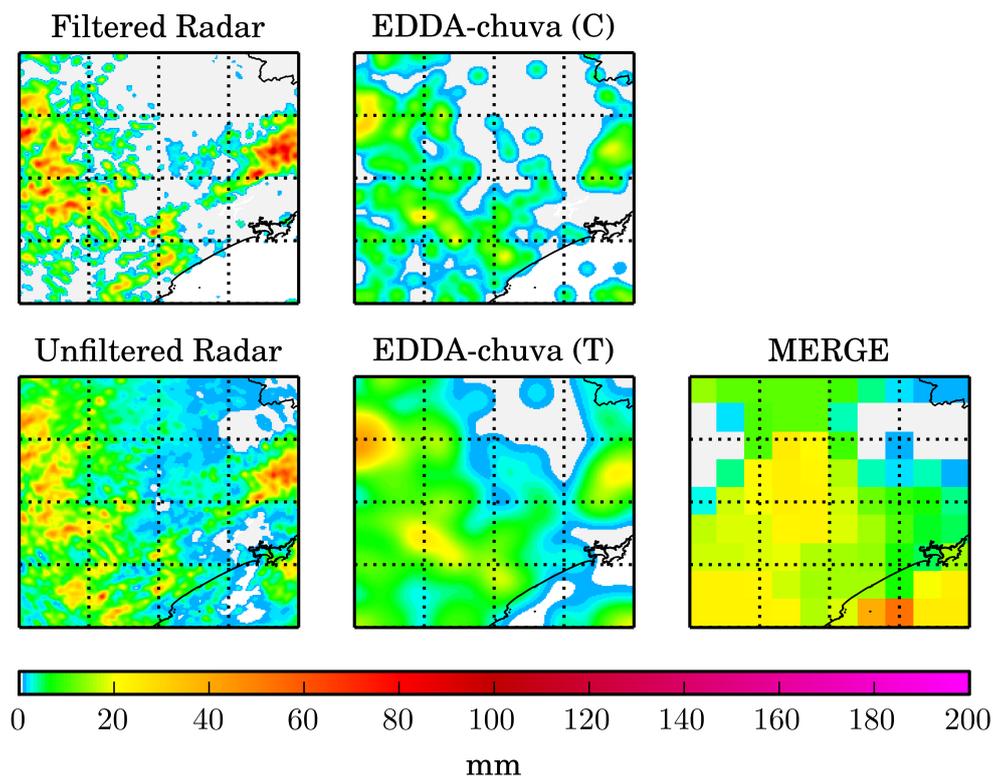


Figure 5.54 - Daily rainfall spatial distributions accumulated for thunderstorm RQ-1 given by (i) weather radar for convective rainfall (“filtered radar”), (ii) EDDA-chuva for convective rainfall (“EDDA-chuva (C)”), (iii) weather radar for total rainfall (“unfiltered radar”), (iv) EDDA-chuva for total rainfall (“EDDA-chuva (T)”), and (v) MERGE data.

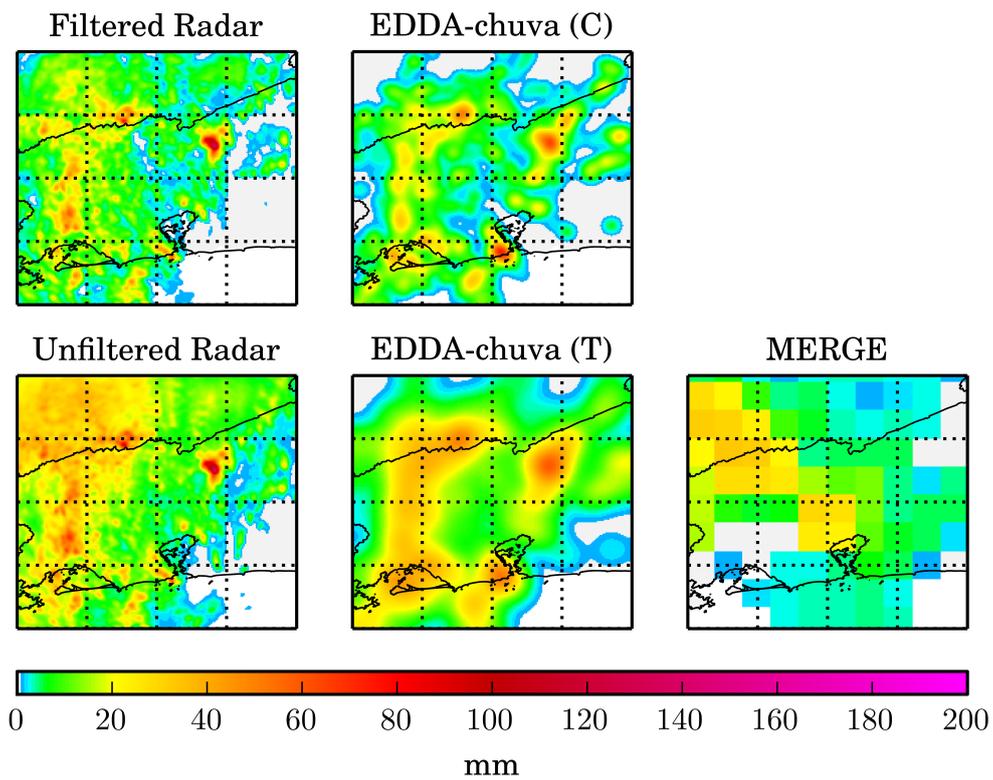


Figure 5.55 - Daily rainfall spatial distributions accumulated for thunderstorm PI-1 given by (i) weather radar for convective rainfall (“filtered radar”), (ii) EDDA-chuva for convective rainfall (“EDDA-chuva (C)”), (iii) weather radar for total rainfall (“unfiltered radar”), (iv) EDDA-chuva for total rainfall (“EDDA-chuva (T)”), and (v) MERGE data.

Finally, since the EDDA-chuva was designed to estimate rainfall mass and distribution outside any weather radar coverage, some validation tests of total rainfall distribution were performed using MERGE data as a reference. Test cases refer to the days of occurrence of the same thunderstorms previously analyzed, i.e. BRU-1, PPR-1S, RQ-1 and PI-1. For these test cases, besides the 24 h period (0:00 to 23:59 UTC), compatible to MERGE data, a larger area was considered, defined by longitudes 40° to 58° W and latitudes 12° to 27° S, that correspond to the RINDAT coverage defined in Figure 2.6. All the four thunderstorms occurred in this area. As expected, correlation of total rainfall spatial distribution as given by the EDDA-chuva software to the MERGE images was poorer than the correlation to weather radar images. The distributions estimated by the software over the ocean must be ignored, since the isohyetal data employed to derive the K matrix elements only includes data over land.

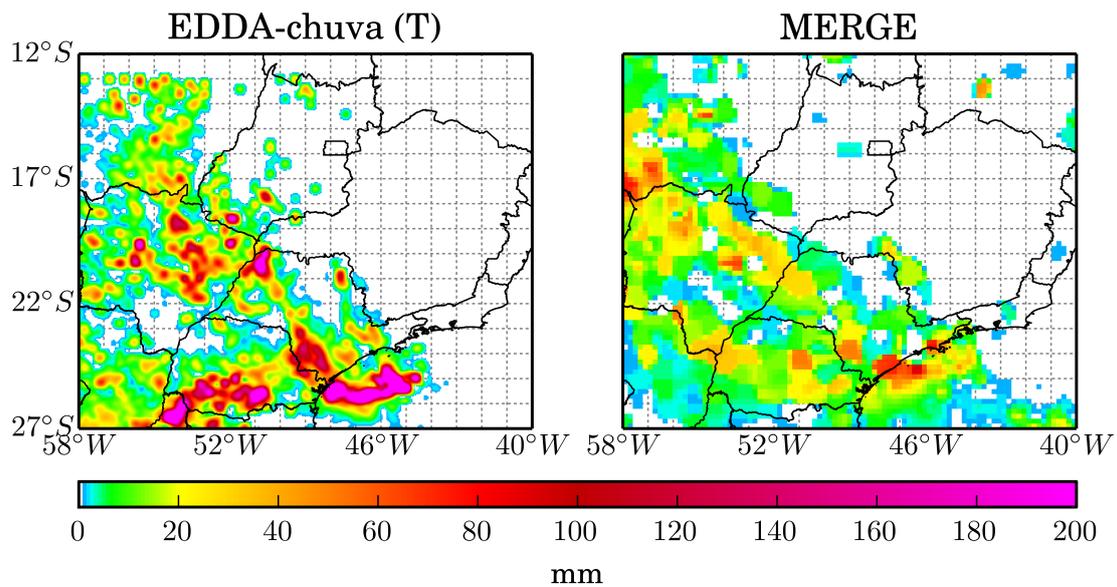


Figure 5.56 - Daily rainfall spatial distribution accumulated for January 1st, 2013 as given by the EDDA-chuva software (left) and from MERGE data (right). Thunderstorm BRU-1 occurred during this day and inside the depicted area.

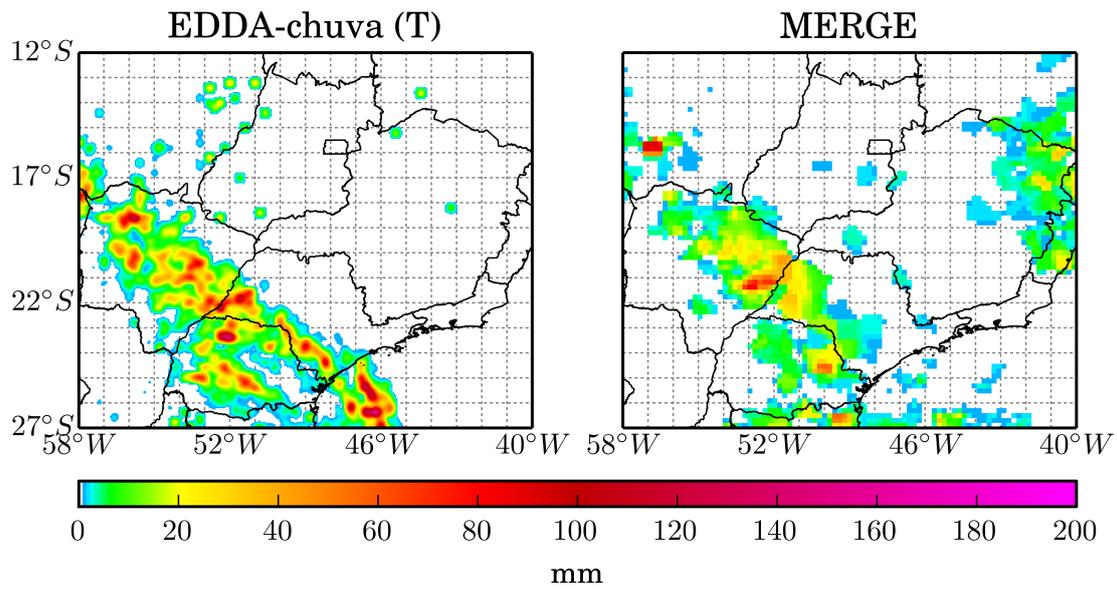


Figure 5.57 - Daily rainfall spatial distribution accumulated for January 11th, 2009 as given by the EDDA-chuva software (left) and from MERGE data (right). Thunderstorm PPR-1S occurred during this day and inside the depicted area.

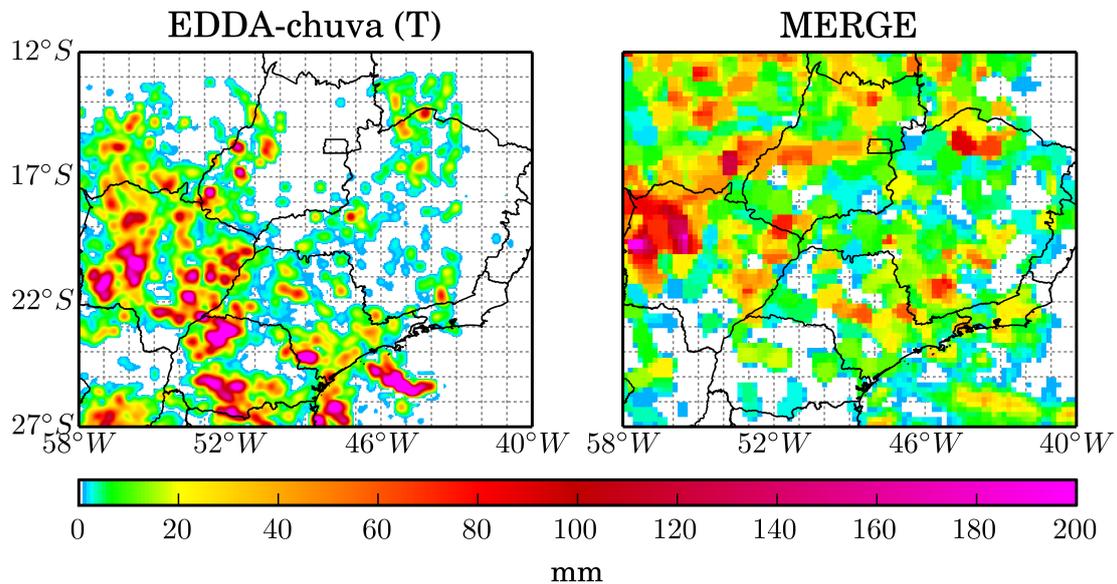


Figure 5.58 - Daily rainfall spatial distribution accumulated for January 25th, 2013 as given by the EDDA-chuva software (left) and from MERGE data (right). Thunderstorm RQ-1 occurred during this day and inside the depicted area.

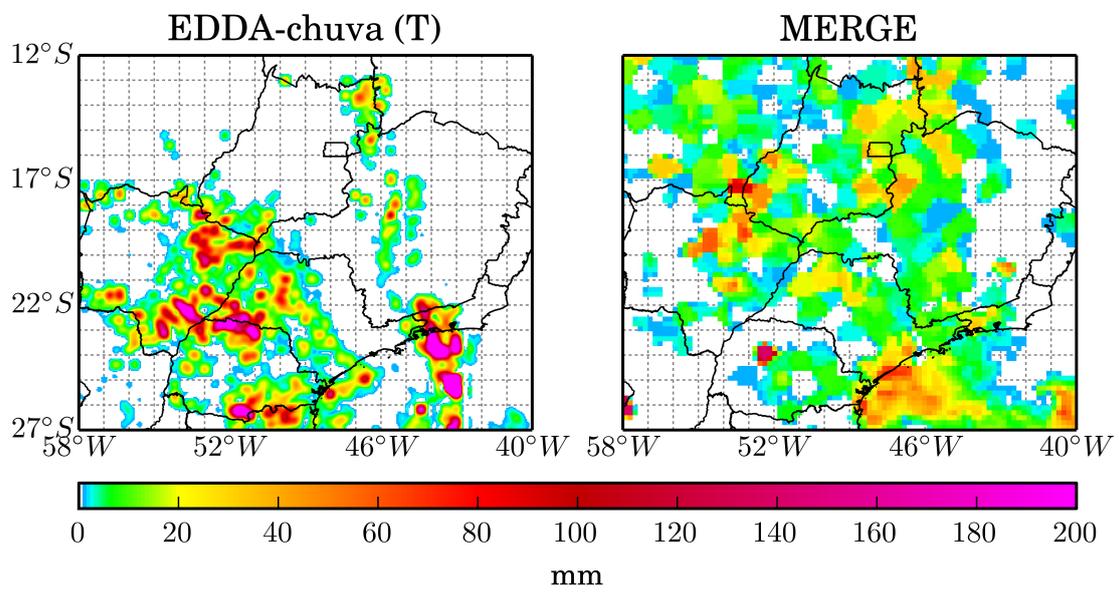


Figure 5.59 - Daily rainfall spatial distribution accumulated for January 14th, 2013 as given by the EDDA-chuva software (left) and from MERGE data (right). Thunderstorm PI-1 occurred during this day and inside the depicted area.

6 CONCLUSIONS

A new approach for estimating convective rainfall mass and spatial distribution from CG lightning stroke data was proposed in this thesis. It may present a significant improvement over the current state of the art on the subject. The rationale behind this work is that countries like Brazil lack a significant coverage of weather radars, and even the images generated by meteorological satellites are limited due to design choices of orbit that imply in specific coverage, duration and periodicity of each pass, etc. Therefore, the use of lightning data to monitor and track convective activity, and even to estimate rainfall, as proposed, seems to represent a suitable option in operational weather forecasting considering the availability of the current lightning detector network (RINDAT).

The contributions of this work in the area of Computer Science are a graph-based algorithm for detecting and tracking thunderstorms by clustering lightning occurrences, and the WRLR/K-mapping approach to estimate rainfall from lightning data that was employed in the EDDA-chuva software.

In the proposed approach, sampling square areas and temporal sliding windows were employed to compute lightning and weather radar data in order to generate a dataset of convective rainfall masses and number of occurrences of strokes that allow to infer a WRLR function for each season for the reference weather radar of Bauru. A mapping matrix K allows to adjust the WRLR function for different grid points, based on isohyetal maps estimated from data of a network of rain gauges over Brazil. Besides the estimation of convective rainfall, the total rainfall mass, which includes the associated stratiform rainfall and convective rainfall without lightning, can also be estimated as the double of the convective mass. This assumption was based on some TRMM estimations and on the available weather radar data. The spatial distribution of the total rainfall is estimated considering that convective rainfall covers 30% of the total rainfall area.

The proposed method assumes that convective activity and lightning are strongly correlated. It may represent an enhancement over the former approach, the Tapia's model, that proposed a constant rainfall to lightning ratio, instead of a spatially mapped function to estimate rainfall mass and that is also based on the same correlation. Another difference between the two approaches is the spatial distribution of the convective derived from the density of occurrence of lightning strokes, instead of considering circles of rainfall around lightning flash occurrences of the Tapia's model. Density fields of occurrence are obtained by Gaussian kernel density estimation

and provide smoother distributions.

Validation tests were performed for Summer months under the coverage of the Bauru weather radar, for different seasons/quarters, under the coverage of the Presidente Prudente weather radar, and also for Summer months under the São Roque and Pico do Couto radars. These tests evaluated the estimated convective rainfall mass considering daily cumulative values, using weather radar data as reference. Additionally, a set of thunderstorms was selected for each weather radar, always including some thunderstorms that presented poor estimations. Convective rainfall mass estimation error was typically lower than 20%, but some variability can be observed for different seasons and locations. Such variability may be attributed to WRLR functions that were inferred using small datasets, since the amount of weather radar data was restricted for some years or radars, and even the lightning data is related to less than ten years. However, even considering such limitations, the mapped WRLR function seems to have outperformed the Tapia's model.

Validation tests were also performed for the estimated rainfall spatial distributions considering the same radars and show that these distributions are correlated to the ones observed in radar images. However, the WRLR function was estimated using weather radar data and certainly tends to generate distributions that are more correlated to radar images than to those corresponding to rain gauge data. As it would be expected, the spatial distributions for the total rainfall were less accurate than the ones obtained for convective rainfall.

Another issue was the spatial mapping of the WRLR function to different grid points, which was based on quarterly isohyetal maps derived from rain gauge data. Another possible source of data is the MERGE project that supplies daily averages of rainfall using both rain gauges and TRMM data. The latter is employed for areas with a sparse distribution of rain gauges. Another advantage of MERGE data is to supply rainfall averages over sea/ocean using TRMM data. The use of MERGE data may represent an improvement over the isohyetal maps, but would employ the same proposed methodology.

The mapped WRLR function was recently implemented by the EDDA-chuva. The EDDA software that generates a density field of occurrence of lightning strokes to monitor convective activity in quasi real time is being employed operationally at CEMADEN since the end of 2012. The recently-developed EDDA-chuva software have the same functionalities of the EDDA software, but also implements the mapped WRLR function in order to generate estimations of convective and total rainfall

masses and spatial distributions. This software is being currently evaluated at the same institution.

6.1 Further works

In face of some issues discussed above, some further works can be devised.

While the present work uses lightning data provided by the RINDAT network, the results could be improved by using data from the Brazilian Lightning Detection Network (BRASILDAT), which was not available at the time. This new network have more lightning detectors spread over a larger area of Brazil, including those that compose the RINDAT network. However, the BRASILDAT network adds the capability of detecting intracloud (IC) discharges, besides the obviously wider coverage and better probability of detection provided by the new detectors (NACCARATO; PINTO, 2009; NACCARATO et al., 2013).

As mentioned above, the adoption of lightning data from BRASILDAT and rainfall data from MERGE will allow to derive a more accurate K matrix that would map better grid points outside Southeastern Brazil. As a consequence, an ever-increasing dataset will result. This a first step to improve the WRLR function and to extend it to a wider area with better accuracy for rainfall estimation than in this work. It is expected that an extended dataset would allow to treat undersampled thunderstorms that have a high number of lightning strokes, but low rainfall mass, or vice-versa.

The WRLR function was obtained using the Bauru weather radar as reference radar, since it was the radar with the larger amount of available data. However, the accuracy of the proposed rainfall estimation can be improved deriving WRLR functions with the use of different weather radars spread over Brazil, in addition to the São Roque and Pico do Couto radars. Therefore, besides an improved mapping of the K matrix, different WRLR could be applied to different regions of Brazil.

In this work, the WRLR function was obtained using temporal windows that are overlapped in time and also a re-sampling using spatial shifts along latitudinal and longitudinal directions. Such schemes were employed to smoothen out lightning and rainfall data and also to reduce the effect of outliers. They also allow to treat the spatial and temporal mismatching between related occurrences of lightning and rainfall. However, the spatial re-sampling is applied only inside the area of the sampling square, i.e. does not consider outer lightning and rainfall. It would be possible to include these kind of “border effect” by defining an extended re-sampling that take

into account neighbor sampling squares. In the same way the estimation would also take into account the “electrical” neighborhood of the considered square in the application of the WRLR function. This would be straightforward, since the lightning stroke density of occurrence is available for the entire considered area in the EDDA-chuva software. The current approach can be thus improved, leading to better rainfall estimations.

Additional improvements can be accomplished by taking advantage of the extra information provided by IC lightning data, as well as by analyzing lightning polarity or lightning electrical charge. The current approach employs only time and space related to the CG lightning occurrences of any polarity. Another approach would be to derive WRLR functions derived with different spatial and temporal resolutions that are specific to the scale of the convective event. Functions derived using smaller values of Δx and/or Δt may be applicable to small events like localized thunderstorms, while larger values, to synoptic scale events.

Despite its relative simplicity, the novelty of this research leaves a wide range of further works open, but only the operational use of the EDDA-chuva software in weather monitoring would guide the completion of such works.

REFERENCES

AFANDI, G. E.; MORSY, M.; HUSSIENY, F. E. Heavy Rainfall Simulation over Sinai Peninsula Using the Weather Research and Forecasting Model.

International Journal of Atmospheric Sciences, v. 2013, p. 1–11, 2013. ISSN 2314-4122. Available from:

<<http://www.hindawi.com/journals/ijas/2013/241050/>>. 1

AHRENS, C. D. **Meteorology today**: an introduction to weather, climate, and the environment. Minneapolis: West Publishing Company, 1988. ISBN 0314624775. 1, 7, 11

ATLAS, D. E. **Radar in meteorology**. Boston: American Meteorological Society, 1990. 806 p. ISBN 9780933876866. 9

BATTAN, L. J. Some factors governing precipitation and lightning from convective clouds. **Journal of Atmospheric Sciences**, v. 22, p. 79–84, 1965. 2, 3

BENETI, C.; FILHO, A. J. P.; DAMIAN, E.; CALVETTI, L. Weather radar and lightning observations of mesoscale systems in the south of Brazil. In: INTERNATIONAL SYMPOSIUM ON NOWCASTING AND VERY SHORT RANGE FORECASTING, 3., 2012, Houston, Texas. **Proceedings...** Rio de Janeiro: World Meteorological Organization, 2012. v. 3, n. August, p. 3–4. ISBN . 28

BUECHLER, D. E.; GOODMAN, S. J. Echo Size and Asymmetry: Impact on NEXRAD Storm Identification. **Journal of Applied Meteorology**, v. 29, n. 9, p. 962–969, sep. 1990. ISSN 0894-8763. Available from:

<<http://journals.ametsoc.org/doi/abs/10.1175/1520-0450%281990%29029%3C0962%3AESAAI0%3E2.0.CO%3B2>>. 2, 3

CALHEIROS, R. V.; D'OLIVEIRA, A. B. Radar-satellite retrieval of cell structure: impact of Tb-Z relationships. In: RICHTER, A. (Ed.). **Proceedings...** Göttingen: Copernicus Publications, 2007. v. 9, p. 10621. ISBN . 10

CALHEIROS, R. V.; GOMES, A. M. Flow forecasting in the Corumbataí River basin: radar rainfall stratification and runoff-rainfall relations. In: EUROPEAN CONFERENCE ON RADAR IN METEOROLOGY AND HYDROLOGY, 6., 2010, Sibiu, Romania. **Proceedings...** Sibiu: National Meteorological Administration, 2010. p. . ISBN 978-973-0-09100-7. Available from:

<http://www.erad2010.org/pdf/POSTER/Thursday/01_hydro/02_ERAD2010_0360_modificat.pdf>. 8, 28

CHÈZE, J.-L.; SAUVAGEOT, H. Area-average rainfall and lightning activity. **Journal of Geophysical Research**, American Geophysical Union, v. 102, n. D2, p. 1707, jan. 1997. ISSN 0148-0227. Available from: <<http://doi.wiley.com/10.1029/96JD02972>>. 2, 3

CUMMINS, K. L.; MURPHY, M. J.; BARDO, E. A.; HISCOX, W. L.; PYLE, R. B.; PIFER, A. E. A combined TOA/MDF technology upgrade of the US National Lightning Detection Network. **Journal of Geophysical Research**, v. 103, n. D8, p. 9035–9044, 1998. 13

GARCIA, J. a. V. C.; STEPHANY, S.; D'OLIVEIRA, A. B. Estimação da razão entre a massa precipitada e número de descargas elétricas atmosféricas para o sudeste brasileiro. In: CONGRESSO NACIONAL DE MATEMÁTICA APLICADA E COMPUTACIONAL, 34., 2012, Águas de Lindóia. **Anais... Águas de Lindóia**, Brasil: Sociedade Brasileira de Matemática, 2012. p. 1271–1277. ISBN . Available from: <<http://plutao.dpi.inpe.br/rep/dpi.inpe.br/plutao/2012/11.28.19.33.32?languagebutton=pt-BR>>. 3, 31, 33

_____. Estimation of convective precipitation mass from lightning data using a temporal sliding-window for a series of thunderstorms in Southeastern Brazil. **Atmospheric Science Letters**, v. 14, n. 4, p. 281–286, oct. 2013. ISSN 1530261X. Available from: <<http://doi.wiley.com/10.1002/as12.453>>. 3, 27, 35

GLAHN, H. R.; LOWRY, D. A. The Use of Model Output Statistics (MOS) in Objective Weather Forecasting. **Journal of Applied Meteorology**, v. 11, n. 8, p. 1203–1211, dec. 1972. ISSN 0021-8952. Available from: <<http://journals.ametsoc.org/doi/abs/10.1175/1520-0450%281972%29011%3C1203%3ATUOMOS%3E2.0.CO%3B2>>. 4

GONZALEZ, R. C.; RICHARD, E. **Digital image processing**. [S.l.]: Prentice Hall Press, 2002. ISBN 0-201-18075-8. 32

GOODMAN, S.; NESDIS, N.; KOSHAK, W.; BLAKESLEE, R. **GLM Lightning Cluster-Filter Algorithm**. Washington, DC: NOAA NESDIS Center for Satellite Applications and Research, 2012. Available from: <<http://www.star.nesdis.noaa.gov/goesr/docs/ATBD/LCFA.pdf>>. 33

HAMILL, T.; WHITAKER, J. Probabilistic quantitative precipitation forecasts based on reforecast analogs: Theory and application. **Monthly weather review**, v. 134, n. 11, p. 3209–3229, nov. 2006. ISSN 0027-0644. Available from: <http://journals.ametsoc.org/doi/abs/10.1175/MWR3237.1>. 4

KATZ, R. W.; MURPHY, A. H. **Economic value of weather and climate forecasts**. Cambridge: Cambridge University Press, 1997. 240 p. ISBN 9780521435710. 1

KEMPF, N. M.; KRIDER, E. P. Cloud-to-Ground Lightning and Surface Rainfall during the Great Flood of 1993. **Monthly Weather Review**, v. 131, n. 6, p. 1140–1149, jun. 2003. ISSN 0027-0644. Available from: [http://journals.ametsoc.org/doi/abs/10.1175/1520-0493\(2003\)131%3C1140:CLASRD%3E2.0.CO%3B2http://journals.ametsoc.org/doi/abs/10.1175/1520-0493%282003%29131%3C1140%3ACLASRD%3E2.0.CO%3B2](http://journals.ametsoc.org/doi/abs/10.1175/1520-0493(2003)131%3C1140:CLASRD%3E2.0.CO%3B2http://journals.ametsoc.org/doi/abs/10.1175/1520-0493%282003%29131%3C1140%3ACLASRD%3E2.0.CO%3B2). 2, 3, 45

KINZER, G. D. Cloud-to-Ground Lightning Versus Radar Reflectivity in Oklahoma Thunderstorms. **Journal of Atmospheric Sciences**, v. 31, p. 787–799, 1974. 2, 3

LANG, T. J.; RUTLEDGE, S. A. Relationship between convective storm kinematics, microphysics and lightning. **Montly Weather Review**, v. 130, p. 2492–2506, 2002. 2, 3

LEMON, L. R. The Radar “Three-Body Scatter Spike”: An Operational Large-Hail Signature. **Weather and Forecasting**, v. 13, n. 2, p. 327–340, jun. 1998. ISSN 0882-8156. Available from: [http://journals.ametsoc.org/doi/abs/10.1175/1520-0434\(1998\)013%3C0327:TRTBSS%3E2.0.CO;2http://journals.ametsoc.org/doi/abs/10.1175/1520-0434%281998%29013%3C0327%3ATRTBSS%3E2.0.CO%3B2](http://journals.ametsoc.org/doi/abs/10.1175/1520-0434(1998)013%3C0327:TRTBSS%3E2.0.CO;2http://journals.ametsoc.org/doi/abs/10.1175/1520-0434%281998%29013%3C0327%3ATRTBSS%3E2.0.CO%3B2). 11

LIAO, L.; MENEGHINI, R. Validation of TRMM precipitation radar through comparison of its multiyear measurements with ground-based radar. **Journal of Applied Meteorology and Climatology**, v. 48, n. 4, p. 804–817, 2009. 2, 4

LIMA, G. R. T. de; STEPHANY, S. A new classification approach for detecting severe weather patterns. **Computers & Geosciences**, v. 57, p. 158–165, aug. 2013. ISSN 00983004. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0098300413001167>. 23

LIU, P.; LI, C.; WANG, Y.; FU, Y. Climatic characteristics of convective and stratiform precipitation over the Tropical and Subtropical areas as derived from

TRMM PR. **Science China Earth Sciences**, v. 56, n. 3, p. 375–385, jul. 2012. ISSN 1674-7313. Available from:

<<http://link.springer.com/10.1007/s11430-012-4474-4>>. 48

MARSHALL, J. S.; PALMER, W. M. K. The distribution of raindrops with size. **Journal of Meteorology**, v. 5, n. 4, p. 165–166, aug. 1948. ISSN 0095-9634.

Available from: <<http://journals.ametsoc.org/doi/abs/10.1175/1520-0469%281948%29005%3C0165%3ATDORWS%3E2.0.CO%3B2>>. 8

MATTOS, E. V.; MACHADO, L. a.T. Cloud-to-ground lightning and Mesoscale Convective Systems. **Atmospheric Research**, Elsevier B.V., v. 99, n. 3-4, p. 377–390, mar. 2011. ISSN 01698095. Available from:

<<http://linkinghub.elsevier.com/retrieve/pii/S0169809510003108>>. 27

MCKINNEY, W. **Python for data analysis: data wrangling with Pandas, NumPy, and IPython**. Sebastopol: O'Reilly Media, 2012. ISBN 1449319793. 40, 53

MEYER, P. L. **Probabilidade aplicada à estatística**. [S.l.]: Editora LTC SA, 2000. 40

MOLINIE, G.; SOULA, S.; CHAUZY, S. Cloud-to-ground lightning activity and radar observations of storms in the pyrénées range area. **Quarterly Journal of the Royal Meteorological Society**, Wiley Online Library, v. 125, n. 560, p. 3103–3122, 1999. 2, 3, 45

NACCARATO, K.; PINTO, O. Improvements in the detection efficiency model for the Brazilian lightning detection network (BrasilDAT). **Atmospheric Research**, v. 91, n. 2-4, p. 546–563, feb. 2009. ISSN 01698095. Available from:

<<http://linkinghub.elsevier.com/retrieve/pii/S0169809508002299>>. 127

NACCARATO, K. P.; Pinto Jr, O.; PINTO, I. A detection efficiency model for the Brazilian Lightning Detection Network (RINDAT). In: INTERNATIONAL LIGHTNING DETECTION CONFERENCE, 19., 2006, Helsinki, Finland. **Proceedings...** Tucson: Vaisala, 2006. p. . ISBN ILDC/ILMC. 13, 16

NACCARATO, K. P.; SABA, M. M. F.; SCHUMANN, C.; PINTO, O.; MEDEIROS, C.; HECKMAN, S. Waveform analysis of cloud-to-ground flashes as detected by fast e-field antennas and lightning location systems: On the way to precisely estimate the stroke peak current. In: INTERNATIONAL SYMPOSIUM ON LIGHTNING PROTECTION, 12., 2013, Washington, USA. **Proceedings...** Belo Horizonte: IEEE, 2013. p. 57–61. ISBN 978-1-4799-1344-2. Available from:

<[http:](http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6729196)

[//ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6729196](http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6729196)>.

127

NATIONAL WEATHER SERVICE. NEXRAD Radar Operations Center, WSR-88D. **Web Site:** <http://www.roc.noaa.gov/>, 2007. 8

OLIVEIRA, R.; MATTOS, E. The Spatial-Temporal Relationship between Cloud-to-Ground Lightning and Precipitation Distributions in the State of São Paulo. In: INTERNATIONAL CONFERENCE ON ATMOSPHERIC ELECTRICITY, 14., 2011, Norman, USA. **Proceedings...** Rio de Janeiro: ICAE, 2011. v. 2, p. 1–4. ISBN . Available from:

<http://mtc-m19.sid.inpe.br/col/sid.inpe.br/mtc-m19/2011/07.07.16.51/doc/Mattos_thespatial_temporal.pdf>. 28

PESSOA, A. S. A.; LIMA, G. R. T. de; DEMÍSIO, J.; SILVA, S. o. da; STEPHANY, S.; STRAUSS, C.; Nelson, Caetano, M. E.; FERREIRA, J. Mineração de dados meteorológicos para previsão de eventos severos. **Revista Brasileira de Meteorologia**, v. 27, n. 1, p. 61—74, 2012. 23

PETERSEN, W. A.; RUTLEDGE, S. A. On the relationship between cloud-to-ground lightning and convective rainfall. **Journal of Geophysical Research**, American Geophysical Union, v. 103, n. D12, p. 14025, jun. 1998. ISSN 0148-0227. Available from: <<http://doi.wiley.com/10.1029/97JD02064>>. 2, 3, 47

PIEPGRASS, M. V.; KRIDER, E. P.; MOORE, C. B. Lightning and surface rainfall during Florida thunderstorms. **Journal of Geophysical Research**, American Geophysical Union, v. 87, n. C13, p. 11111–11193, 1982. 2, 3

PINTO, J. O.; NACCARATO, K. P.; SABA, M. M. F.; PINTO, I.; GARCIA, S. A. M. d. M.; FILHO, A. C.; JR, O. P.; ABDO, R. Recent upgrades to the Brazilian Integrated Lightning Detection Network. In: INTERNATIONAL LIGHTNING DETECTION CONFERENCE, 19., 2006, Helsinki, Finland. **Proceedings...** Tucson: Vaisala, 2006. p. . ISBN ILDC/ILMC. Available from: <http://www.vaisala.com/VaisalaDocuments/Scientificpapers/Recent_upgrades_to_the_Brazilian_Integrated_Lightning_Detection_Network.pdf>. 1

PINTO, O.; Regina C.A. Pinto, I.; H. Diniz, J.; Cazetta Filho, A.; C.L. Cherchiglia, L.; M. Carvalho, A. A seven-year study about the negative cloud-to-ground lightning flash characteristics in Southeastern Brazil. **Journal of**

Atmospheric and Solar-Terrestrial Physics, v. 65, n. 6, p. 739–748, apr. 2003. ISSN 13646826. Available from:

<<http://linkinghub.elsevier.com/retrieve/pii/S1364682603000774>>. 33

POLITI, J.; STEPHANY, S.; DOMINGUES, M.; JUNIOR, O. M. Mineração de dados meteorológicos associados à atividade convectiva empregando dados de descargas elétricas atmosféricas. **Revista Brasileira de Meteorologia**, v. 21, n. 2, p. 232—244, 2006. 22

RAMIREZ-BELTRAN, N. D.; KULIGOWSKI, R. J.; HARMSSEN, E. W.; CASTRO, J. M.; CRUZ-POL, S.; CARDONA, M. J. Rainfall estimation from convective storms using the hydro-estimator and NEXRAD. **WSEAS Transaction on Systems**, v. 7, n. 10, p. 1016–1027, 2008. 2, 4

RINDAT. **Rede Integrada de Detecção de Descargas Atmosféricas**. 2011. Available from: <<http://http://www.rindat.com.br/>>. xv, 14, 16

RINEHART, R. E. **Radar for meteorologists**. [S.l.]: Rinehart Publications Grand Forks, ND, 1997. 10

ROSSUM, G. V.; DRAKE, F. L. **Python language reference manual**. Ingram: Network Theory, 2011. ISBN 1906966141. 53

ROZANTE, J. R.; MOREIRA, D. S.; GONCALVES, L. G. G. de; VILA, D. A. Combining TRMM and Surface Observations of Precipitation: Technique and Validation over South America. **Weather and Forecasting**, v. 25, n. 3, p. 885–894, jun. 2010. ISSN 0882-8156. Available from: <<http://journals.ametsoc.org/doi/abs/10.1175/2010WAF2222325.1>>. 18, 19

RYAN, T. P. **Modern regression methods**. 2nd. ed. New York: Wiley-Interscience, 2008. 672 p. ISBN 978-0-470-08186-0. 40

SCHUMACHER, C.; HOUZE, R. A. Stratiform Rain in the Tropics as Seen by the TRMM Precipitation Radar*. **Journal of Climate**, v. 16, n. 11, p. 1739–1756, jun. 2003. ISSN 0894-8755. Available from: <<http://journals.ametsoc.org/doi/abs/10.1175/1520-0442%282003%29016%3C1739%3ASRITTA%3E2.0.CO%3B2>>. 48, 49

SCOTT, D. W. **Multivariate density estimation: theory, practice and visualization**. [S.l.]: Wiley. com, 2009. 383 p. 24

SIINGH, D.; KUMAR, S.; SINGH, A. K. Thunderstorms/Lightning Generated Sprite and Associated Phenomena. **Earth Science India**, v. 3, n. II, p. 124–145, 2010. 2, 3

SILVERMAN, B. **Density estimation for statistics and data analysis**. [S.l.: s.n.], 1986. 1–22 p. 23, 24, 25

SIST, M.; ZAULI, F.; MELFI, D.; BIRON, D. A study about the correlation link between lightning data and meteorological data. In: METEOROLOGICAL SATELLITE CONFERENCE, ., 2010, Darmstadt, Germany. **Proceedings...** Córdoba, Spain: EUMETSAT, 2010. p. . ISBN 978-92-9110-089-7. 2, 3

SOULA, S.; CHAUZY, S. Some aspects of the correlation between lightning and rain activities in thunderstorms. **Atmospheric research**, Elsevier, v. 56, n. 1-4, p. 355–373, 2001. Available from:

<<http://www.dvfu.ru/meteo/library/43980058.pdf>>. 2, 3

STEINER, M.; Houze Jr, R. A.; YUTER, S. E. Climatological characterization of three-dimensional storm structure from operational radar and rain gauge data. **Journal of Applied Meteorology**, [Boston, Mass.]: The Society, c1988-c2005., v. 34, n. 9, p. 1978–2007, 1995. 25, 35

STRAUSS, C.; ROSA, M. B.; STEPHANY, S. Spatio-temporal clustering and density estimation of lightning data for the tracking of convective events.

Atmospheric Research, Elsevier B.V., v. 134, p. 87–99, dec. 2013. ISSN 01698095. Available from:

<<http://linkinghub.elsevier.com/retrieve/pii/S0169809513001993>>. 1, 3, 22, 50

STRAUSS, C.; STEPHANY, S.; CAETANO, M. A ferramenta EDDA de geração de campos de densidade de descargas atmosféricas para mineração de dados meteorológicos. In: CONGRESSO NACIONAL DE MATEMÁTICA APLICADA E COMPUTACIONAL, 33., 2010, Rio de Janeiro, Brasil. **Anais...** Águas de Lindóia: Sociedade Brasileira de Matemática, 2010. p. . ISBN . 22

TAKAYABU, Y. N. Characteristics of Precipitation as Observed by TRMM PR. In: GEWEX-GPCP WORKSHOP ON PRECIPITATION ANALYSIS, ., , Reading, UK. **Proceedings...** Reading, UK: European Centre for Medium-Range Weather Forecasts, 2003. p. 84. ISBN . 48

TAPIA, A.; SMITH, J. A.; DIXON, M. Estimation of convective rainfall from lightning observations. **Journal of Applied Meteorology**, v. 37, n. 11, p. 1497–1509, 1998. ISSN 0894-8763. 2, 3, 21, 45, 54

TUKEY, J. W. **Exploratory data analysis**. [S.l.: s.n.], 1977. 38, 58

UMAN, M. A. **The lightning discharge**. [S.l.]: Dover Pubns, 2001. 16

UMAN, M. A.; KRIDER, E. P. Natural and artificially initiated lightning. **Science**, American Association for the Advancement of Science, v. 246, n. 4929, p. 457, 1989. ISSN 0036-8075. 13

WILLIAMS, E.; RENNO, N. An Analysis of the Conditional Instability of the Tropical Atmosphere. **Monthly Weather Review**, American Meteorological Society, v. 121, n. 1, p. 21–36, jan. 1993. ISSN 0027-0644. Available from: <http://journals.ametsoc.org/doi/abs/10.1175/1520-0493%281993%29121%3C0021%3AAAOTCI%3E2.0.CO%3B2>>. 2, 3

WILLIAMS, E. R.; GEOTIS, S. G.; RENNO, N.; RUTLEDGE, S. A.; RASMUSSEN, E.; RICKENBACH, T. A Radar and Electrical Study of Tropical “Hot Towers”. **Journal of the Atmospheric Sciences**, American Meteorological Society, v. 49, n. 15, p. 1386–1395, aug. 1992. ISSN 0022-4928. Available from: <http://journals.ametsoc.org/doi/abs/10.1175/1520-0469%281992%29049%3C1386%3AARAES0%3E2.0.CO%3B2>>. 2, 3, 45

ZIPSER, E. J. Deep Cumulonimbus Cloud Systems in the Tropics with and without Lightning. **Monthly Weather Review**, v. 122, n. 8, p. 1837–1851, aug. 1994. ISSN 0027-0644. Available from: [http://journals.ametsoc.org/doi/abs/10.1175/1520-0493\(1994\)122%3C1837:DCCSIT%3E2.0.CO;2http://journals.ametsoc.org/doi/abs/10.1175/1520-0493%281994%29122%3C1837%3ADCCSIT%3E2.0.CO%3B2](http://journals.ametsoc.org/doi/abs/10.1175/1520-0493(1994)122%3C1837:DCCSIT%3E2.0.CO;2http://journals.ametsoc.org/doi/abs/10.1175/1520-0493%281994%29122%3C1837%3ADCCSIT%3E2.0.CO%3B2)>. 2, 3

ANNEX A - PUBLISHED ARTICLES RELATED TO THIS THESIS

GARCIA, J. V. C.; STEPHANY, S.; D'OLIVEIRA, A. B. Estimation of convective precipitation mass from lightning data using a temporal sliding-window for a series of thunderstorms in Southeastern Brazil. **Atmospheric Science Letters**, v. 14, n. 4, p. 281–286, oct. 2013. ISSN 1530261X. Available from: <<http://doi.wiley.com/10.1002/as12.453>>

GARCIA, J. V. C.; STEPHANY, S.; D'OLIVEIRA, A. B. Estimaco da razo entre a massa precipitada e nmero de descargas eltricas atmosfricas para o sudeste brasileiro. In: **34º Congresso Nacional De Matemtica Aplicada e Computacional (CNMAC-2012)**. [s.n.], 2012. p. 1271–1277. Available from: <<http://plutao.dpi.inpe.br/rep/dpi.inpe.br/plutao/2012/11.28.19.33.32?languagebutton=pt-BR>>

Estimation of convective precipitation mass from lightning data using a temporal sliding-window for a series of thunderstorms in Southeastern Brazil

João V. C. Garcia,^{1*} Stephan Stephany¹ and Augusto B. d'Oliveira²

¹National Institute for Space Research (INPE), Sao Jose dos Campos, Brazil

²Center for Monitoring and Warnings of Natural Disasters (CEMADEN-MCTI), Cachoeira Paulista, Brazil

*Correspondence to:

J. V. C. Garcia, Programa de Pós-graduação Computação Aplicada (CAP/INPE), Caixa postal 515, CEP 12245-970, Sao Jose dos Campos, Brazil.
E-mail: sawamano@gmail.com

Received: 1 February 2013
Revised: 29 May 2013
Accepted: 2 July 2013

Abstract

Some studies have proposed the estimation of convective rainfall from lightning observations by the computation of the rainfall–lightning ratio (RLR). However, as such ratio may depend on season, convective regime and other factors, known approaches failed to provide values of RLR with low variability. An accurate RLR would allow estimating rainfall from lightning data in areas that lack weather radar coverage. This work proposes a straightforward approach for the computation of RLR, based on a temporal sliding-window and a fitting function. It was tested for thunderstorms observed in the Southeastern Brazil with good results.

Keywords: weather radar; lightning; rainfall estimation

1. Introduction

Rainfall estimation is typically performed from weather radar data. However, assuming that convective rainfall can be correlated to lightning, some approaches propose rainfall estimation from lightning data for areas without weather radar coverage, supporting nowcasting. The most common approach is the computation of the rainfall–lightning ratio (RLR), given by the convective rainfall mass per cloud-to-ground (CG) lightning flash. Nevertheless, such ratio may depend heavily on seasonal and geographical factors, local climatology, convective regime, storm type, lightning patterns or intensity, dominant lightning polarity of CG lightning, intracloud to CG ratio and thunderstorm life cycle (Buechler and Goodman, 1990; Soula and Chauzy, 2001; Lang and Rutledge, 2002). Therefore, known approaches may fail to provide values of RLR with low variability (Sist *et al.*, 2010).

A number of studies were performed to estimate the rainfall mass directly from CG lightning observations. Petersen and Rutledge (1998) used the total rainfall mass and the density of CG lightning to examine their relationship on a number of spatial and temporal scales for different parts of the world. The lightning flash incidence is more intense in clouds associated to high-level precipitation, as the electrification increases with altitude as in the case of tall cumulonimbus (Siingh *et al.*, 2010). Tapia *et al.* (1998) computed the RLR by dividing the total convective rainfall mass by the number of CG flashes in a thunderstorm, and proposed a model to reconstruct the spatial and temporal distribution of the rainfall. The summation of the rainfall

distribution of the flashes yields the overall rainfall distribution, which was checked against weather radar data. Kempf and Krider (2003) presented a compilation of RLR values including some obtained from other works, and found values ranging from 38 to 72×10^6 kg per flash for isolated thunderstorms in Florida, Spain and France, and values as high as 5000×10^6 kg per flash for mesoscale thunderstorms in Australia and Central United States. Molinie *et al.* (1999) found values as low as 3×10^6 kg per flash for the Pyrenees, while Williams *et al.* (1992) found values up to 500×10^6 kg per flash for Australia.

The current work proposes a simpler and more accurate approach, the function windowed RLR (WRLR), which employs a temporal sliding-window. This approach is based on the assumption that convective activity is correlated to electrically active cells that correspond to areas with high density of CG strokes. Such density is calculated by the EDDA software that implements standard kernel estimation (Strauss *et al.*, 2010). This software is being evaluated for operational use in order to detect convective precipitation in the recently established Center for Natural Disasters Monitoring and Alert (CEMADEN) in Brazil.

A set of thunderstorms that occurred in 2009 in the Southeastern Brazil was selected from weather radar data to obtain a WRLR function, while another set of January 2010 was employed to test this WRLR function as rainfall estimator. It is expected to include this function as a new module of the EDDA software. This may provide rainfall estimation in parts of Brazil, a huge country that has over 8.5 million km², but less than 15% of its area is covered by weather radar. Rainfall estimations can be obtained from

Estimação da razão entre a massa precipitada e número de descargas elétricas atmosféricas para o sudeste brasileiro

João V. C. Garcia

INPE - Programa de Pós-Graduação em Computação Aplicada (CAP),
12227-010, São José dos Campos, SP
E-mail: sawamano@gmail.com,

Stephan Stephany

INPE - Laboratório Associado de Computação e Matemática Aplicada (LAC),
12227-010, São José dos Campos, SP
E-mail: stephan@lac.inpe.br,

Augusto B. d'Oliveira

UMC - Programa de Pós-Graduação em Engenharia Biomédica,
08780-210, Mogi das Cruzes, SP, e
CPTEC/INPE - Divisão de Satélites Ambientais
15630-000, Cachoeira Paulista, SP
E-mail: abdoliveira@yahoo.com.br.

Resumo: *Vários estudos tentam relacionar estatisticamente ocorrências de descargas elétricas atmosféricas com a taxa de precipitação em eventos convectivos. Uma abordagem empregada é a razão entre a massa precipitada e o número de descargas elétricas atmosféricas, a qual permite estimativas da taxa de precipitação a partir de dados de descargas, conforme expresso pelo modelo de Tapia. O presente trabalho estima essa razão para eventos convectivos observados no sudeste do Brasil utilizando dados de radares meteorológicos e de uma rede de detecção de descargas elétricas atmosféricas. Entretanto, propõe uma abordagem nova para seu cálculo e automatiza a seleção de eventos. Os resultados foram condizentes com aqueles encontrados na literatura da área.*

Palavras-chave: *descargas elétricas atmosféricas, radar meteorológico, atividade convectiva*

1 Introdução

O interesse em se melhor entender o funcionamento de sistemas de precipitação convectiva levou vários autores a procurar o relacionamento entre descargas elétricas atmosféricas (DEA) e pluviosidade. Uma melhor compreensão do funcionamento de sistemas de atividade convectiva severa pode levar a uma melhor estimativa da pluviosidade em regiões onde a cobertura de radar é insuficiente, e pode ser útil para se obter dados para tomada de decisão e alimentar modelos meteorológicos.

A atividade convectiva severa, caracterizada pela presença de ventos fortes, intensa atividade elétrica atmosférica e precipitação elevada, é um tipo de sistema de precipitação convectiva de particular interesse. As DEA são mais intensas em nuvens de alta convectividade; conforme a eletrificação aumenta com a altitude das nuvens convectivas, as cumulonimbus mais altas podem produzir atividade elétrica mais intensa [8].

Estudos foram realizados com o intuito de estimar o volume precipitado diretamente a partir da observação de descargas nuvem-solo. Zipser [12] usou a razão da precipitação mensal pelo

PUBLICAÇÕES TÉCNICO-CIENTÍFICAS EDITADAS PELO INPE

Teses e Dissertações (TDI)

Teses e Dissertações apresentadas nos Cursos de Pós-Graduação do INPE.

Manuais Técnicos (MAN)

São publicações de caráter técnico que incluem normas, procedimentos, instruções e orientações.

Notas Técnico-Científicas (NTC)

Incluem resultados preliminares de pesquisa, descrição de equipamentos, descrição e ou documentação de programas de computador, descrição de sistemas e experimentos, apresentação de testes, dados, atlas, e documentação de projetos de engenharia.

Relatórios de Pesquisa (RPQ)

Reportam resultados ou progressos de pesquisas tanto de natureza técnica quanto científica, cujo nível seja compatível com o de uma publicação em periódico nacional ou internacional.

Propostas e Relatórios de Projetos (PRP)

São propostas de projetos técnico-científicos e relatórios de acompanhamento de projetos, atividades e convênios.

Publicações Didáticas (PUD)

Incluem apostilas, notas de aula e manuais didáticos.

Publicações Seriadas

São os seriados técnico-científicos: boletins, periódicos, anuários e anais de eventos (simpósios e congressos). Contam destas publicações o Internacional Standard Serial Number (ISSN), que é um código único e definitivo para identificação de títulos de seriados.

Programas de Computador (PDC)

São a seqüência de instruções ou códigos, expressos em uma linguagem de programação compilada ou interpretada, a ser executada por um computador para alcançar um determinado objetivo. Aceitam-se tanto programas fonte quanto os executáveis.

Pré-publicações (PRE)

Todos os artigos publicados em periódicos, anais e como capítulos de livros.