



universität  
wien

# DIPLOMARBEIT

Titel der Diplomarbeit

Climatic thresholds for ecosystem stability:

The case of the Western Congolian Lowland Rainforest

Verfasser

Johannes Elias Bednar

angestrebter akademischer Grad

Magister der Naturwissenschaft (Mag.rer.nat)

Wien, 2011

Studienkennzahl lt. Studienblatt: A 411

Studienrichtung lt. Studienblatt: Physik

Betreuerin / Betreuer: Univ. Prof. Dipl.-Ing. Dr. Hubert Hasenauer



## **Abstract**

The heterogeneity in the composition of species and the mix of forest ecosystems of the present tropical flora in western Central Africa has been subject of many publications. Most of the authors agree on the idea that changing climatic conditions in the past have led to disturbances that subsequently caused different stages in plant succession in the present picture. This work's aim is to find out which climatic parameters have a significant impact on the stability of tropical forest ecosystems, such as the showcase biome of the Western Congolian Lowland Rainforest (WCLR).

Using the stochastic weather generator MarkSim, climate time series with quantified meteorological parameters, such as the amount and year-to-year variation of annual rainfall, the distribution of rainfall within the year and the quality of the cloud cover, are generated. For this reason MarkSim is adapted and validated for sites in Gabon where the WCLR-biome is native.

The mechanistic ecosystem model Biome-BGC, parametrized for the WCLR-biome, simulates the cycling of water, energy, carbon and nitrogen through different plant compartments and is applied to assess tropical forest ecosystem stability, based on the climate time series generated with MarkSim.

The methods developed in the course of this work are applicable to other forest ecosystems and can be regarded as an innovative approach to assess the impact of climatic change.

## Zusammenfassung

Die Heterogenität der Artenzusammensetzung, sowie der Waldökosystemgefüge der heutigen tropischen Flora im westlichen Zentralafrika ist Thema vieler Publikationen. Ein Großteil der Autoren gibt Veränderungen des Klimas, welche in Folge zu Störungen dieser Systeme führten, die Verantwortung für die heute bestehenden unterschiedlichen sukzessionalen Stadien tropischer Wälder. Zielsetzung dieser Arbeit ist es, anhand des Beispielbioms Westkongolesischer Tieflandregenwald (WCLR) den Einfluss klimatischer Parameter auf die Stabilität tropischer Waldökosysteme zu untersuchen.

Der stochastische Wettergenerator MarkSim wird benutzt um Klimazeitserien mit quantifizierten klimatischen Parametern, wie die Gesamtniederschlagsmenge, die jährliche Variation des Niederschlags, die Verteilung des Niederschlags über das Jahr sowie die Art der Wolkendecke, zu generieren. Zu diesem Zweck wird MarkSim für Gabun, welches das WCLR-Biom beheimatet, adaptiert und validiert.

Das für das WCLR-Biom parametrisierte mechanistische Ökosystemmodell Biome-BGC simuliert die Kreisläufe von Wasser, Energie, Kohlenstoff und Stickstoff durch unterschiedliche Bestandteile eines Waldökosystems, und wird zur Abschätzung der Stabilität solcher Systeme basierend auf den zuvor generierten Klimaten herangezogen.

Die im Zuge dieser Arbeit entwickelten Methoden lassen sich auch auf andere Typen von Waldökosystemen übertragen und können als innovativer Ansatz zur Bewertung des Einflusses klimatischer Veränderungen auf diese Systeme erachtet werden.

## Acknowledgement

I would like to show my sincere gratitude to Dr. Stephan Pietsch for his inspiration and enthusiasm and his great efforts to explain things clearly and simply which helped a lot to enter a new field of science. Throughout my thesis-writing period he provided excellent ideas and most of all, good company. It was a honor for me to travel with him and get to know his lovely family.

I am very grateful to Prof. Hubert Hasenauer who's motivation, and advice in structural questions helped me in all the time of research and writing of this thesis.

This work would not have been possible without the support of Dr. Richard Petritsch, who provided aid to solve countless computational tasks. I would have been lost without his advice. Besides my advisors at the Boku University, I would like to thank the rest of my Professors at the University of Vienna, I owe special thanks to Prof. Helmuth Horvath who agreed to join my board of examiners.

I am indepted to my fellow students and friends Martin Diermaier, who has found a research group for me, and Lukas Pichelstorfer, flatmate and close friend for many years.

I wish to thank my entire extended family for providing a loving environment for me. My girlfriend Miriam Bammer, my half-sister Katharina Urbanek, and her mother Lisa Urbanek were particularly supportive.

Most importantly I want to thank my parents, Waltraud Bednar and Hans Bednar. They bore me, raised me, taught me, and loved me. To them I dedicate this thesis.

It is a pleasure to thank the many people who made this thesis possible.

# Table of Contents

1 Introduction.....	1
1.1 Background.....	1
1.2 Availability of climate records in Gabon .....	3
2 Goals.....	4
3 Methods.....	6
3.1 Weather generators.....	6
3.1.1 Why to use weather generators?.....	6
3.1.2 MarkSim: A probabilistic weather generator.....	7
3.2 Biome-BGC: A biogeochemical ecosystem model.....	8
4 Data.....	10
5 Analysis and improvements.....	15
5.1 Validating and correcting stochastic climate for Gabon.....	15
5.1.1 Comparison of international tropical sites with sites in Gabon.....	15
5.1.2 Correction of daily solar radiation and maximum temperature.....	20
5.1.3 Validation of generated climate using data from three weather stations .....	21
5.1.4 Evaluating the distribution of annual rainfall.....	32
5.2 Customizing climate for BGC-simulations .....	33
5.2.1 Customizing dry season and annual rainfall.....	34
5.2.2 Climate for Biome-BGC simulations.....	36
5.2.3 Climatic parameters and varieties for BGC-simulations.....	42
6 Results.....	43
6.1 Testing ecosystem stability under different climatic conditions.....	43
6.1.1 Methodical approach.....	44
6.1.2 Two concepts to approach a stable state.....	45
6.1.3 Spinup simulations.....	46
6.1.4 Spindown simulations.....	64
6.1.5 Hysteresis (spinup vs. spindown).....	80
7 Discussion.....	83
8 Bibliography.....	85
APPENDIX A - MarkSim file structure.....	89
APPENDIX B - Statistical methods .....	91
B.1 Test for normality.....	91
B.2 Testing predicted vs. observed.....	91
Confidence (CI) and prediction-interval (PI) for residuals.....	91
Pearson's linear correlation coefficient & sample mean error.....	92
B.3 Box-and-whiskers plot.....	93

B.4 Logistic regression analysis.....	93
APPENDIX C - Correction of solar radiation and maximum temperature.....	95
C.1 The main cycle.....	96
Correction Factors (CF) and Truncating Factors (TF).....	97
The probability function (PF).....	98
The probability function – a different representation.....	100
Dynamic transition dates (DTD).....	101
C.2 The dry season.....	103
Estimating dry season boundaries.....	104
One application .....	107
C.3 Vapor pressure deficit (VPD).....	108
C.4 Output files.....	108
C.5 Optimization.....	109
Iterations of the main cycle.....	109
dyreg – regulating the length of the dry season.....	111
APPENDIX D - Generating paleo climate and climate change.....	113

## Table of Figures

Figure 1.1: Vegetation map of eastern Central Africa (source: Joint Research Center). 1. Monte Alén-Monts de Cristal Landscape ; 2. Gamba-Mayumba-Conkouati Landscape; 3. Lopé-Chaillu-Louesse Landscape; 4. Dja-Odzala-Minkébé (Tridom) Landscape; 5. Sangha Tri-National Landscape; 6. Léconi-Batéké-Léfini Landscape ; 7. Lake Télé-Lake Tumba Landscape; 8. Salonga-Lukenie-Sankuru Landscape; 9. Maringa-Lopori-Wamba Landscape; 10. Maiko-Tayna-Kahuzi-Biega Landscape ; 11. Ituri-Epulu-Aru Landscape; 12. Virunga Landscape. ....	1
Figure 4.1: Map of tropics in and outside of Gabon, as listed in table 4.1 and 4.2, Latin American & western African sites (blue, 16-20), Gabonese sites from Maloba Makanga 2010, p.25 (red, yellow, green; 1-14), Gabonese sites used for comparison in section 5 (yellow & green), sites where long-term weather data are available (green). ....	11
Figure 5.1: Validation of precipitation: monthly precipitation residuals [mm/month] with 95%-confidence intervals for Moanda, Mouila and Lastoursville. If the dashed line stays within the CI, the data point is unbiased.....	23
Figure 5.2: Validation of minimum temperature: monthly minimum temperature residuals [mm/month] with 95%-confidence intervals for Moanda, Mouila and Lastoursville. If the dashed line falls into the CI, the data point is unbiased.....	25
Figure 5.3: Validation of uncorrected maximum temperature: monthly maximum temperature residuals [mm/month] with 95%-confidence intervals for Moanda, Mouila and Lastoursville, before correction. If the dashed line falls into the CI, the data point is unbiased.....	27
Figure 5.4: Validation of corrected maximum temperature: monthly maximum temperature residuals [mm/month] with 95%-confidence intervals for Moanda, Mouila and Lastoursville, after correction. If the dashed line falls into the CI, the data point is unbiased.....	28
Figure 5.5: Monthly precipitation [mm/month] from weather station (black line), generated (red line) .....	31
Figure 5.6: Monthly mean tmin [°C] from weather station (black line), generated (red line).....	31
Figure 5.7: Monthly mean solar radiation [°C] uncorrected (black line), and corrected (red line)....	31
Figure 5.8: Monthly mean tmax [°C] from weather station (black line), generated uncorrected (red line), corrected (green line).....	32
Figure 5.9: Monthly mean temperature difference [°C] from weather station (black line), generated uncorrected (red line), corrected (green line).....	32
Figure 5.10: Forming a dry season: monthly precipitation [mm/month] for two to six months of dry season, site Birougou Mountains P3: 12°20' E / 1°45' S or 12.334 / -1.750, 1500 mm annual precipitation, elevation 800 masl. Climate normal (black line) and generated rainfall (dashed line). ....	35
Figure 5.11: Histograms proving information about the amount of available climate year per 100mm of rainfall for 2 to 6 DM. ....	41



Figure 6.1: Exemplary transition from 100% collapse (left side) to 100% development of stationary system (right side) based on spinup simulations with a 3DM, uncorrected climate (with respect to solar radiation and maximum temperature), and 37.5% standard deviation of the annual rainfall distribution. P is the chance (in %) to develop a stationary state, X0.5 is the point of inflection, and W the width of the transition phase (=risky phase). .....46

Figure 6.2: Point of inflection (solid line), 1% and 99% probability boundaries in mm rainfall per year (upper and lower dashed line, respectively) derived from logistic regression as marker for stability, based on simulations with uncorrected climate. The underlying rainfall distributions ranges from SD=10% to SD=25%..... 48

Figure 6.3: Point of inflection (solid line), 1% and 99% probability boundaries in mm rainfall per year (upper and lower dashed line, respectively) derived from logistic regression as marker for stability, based on simulations with uncorrected climate. The underlying rainfall distributions ranges from SD=27.5% to SD=35%..... 49

Figure 6.4: Point of inflection (solid line), 1% and 99% probability boundaries in mm rainfall per year (upper and lower dashed line, respectively) derived from logistic regression as marker for stability, based on simulations with uncorrected climate. The underlying rainfall distributions ranges from SD=37.5% to SD=40%..... 50

Figure 6.5: Exponential growth of point of inflection (black points) in mm/yr with increasing year-to-year variation of rainfall (SD of the precipitation distribution). Larger red points indicate the mean of black points, the dashed red line is derived from an exponential regression. ....51

Figure 6.6: Point of inflection (solid line), 1% and 99% probability boundaries in mm rainfall per year (upper and lower dashed line, respectively) derived from logistic regression as marker for stability, based on simulations with corrected climate. The underlying rainfall distributions ranges from SD=10% to SD=25%..... 54

Figure 6.7: Point of inflection (solid line), 1% and 99% probability boundaries in mm rainfall per year (upper and lower dashed line, respectively) derived from logistic regression as marker for stability, based on simulations with corrected climate. The underlying rainfall distributions ranges from SD=27.5% to SD=35%..... 55

Figure 6.8: Exponential growth of point of inflection (black points) in mm/yr with increasing year-to-year variation of rainfall (SD of the precipitation distribution). Larger red points indicate the mean of black points, the dashed red line is derived from an exponential regression. ....56

Figure 6.9: Exponential growth of mean point of inflection corrected (red symbols) versus uncorrected (blue symbols) in mm/yr with increasing year-to-year variation of rainfall (SD of the precipitation distribution). The dashed lines present the exponential regression. ....58

Figure 6.10: Exponential growth of mean length of the transition phase (W) corrected (red symbols) versus uncorrected (blue symbols) in mm/yr with increasing year-to-year variation of rainfall (SD of the precipitation distribution). The dashed lines present the exponential regression. ....59

Figure 6.11: Width of transition phase in percent of the according X0.5 (corrected vs. uncorrected) for 2DM to 6DM. Dots illustrate the mean, the boxplot provides information about the underlying distribution of relative W. .... 60

Figure 6.12: Above-ground living biomass carbon (= leaf carbon + dead stem carbon) from all successful simulations averaged over the last mortality cycle of each simulation. Carbon levels (in different colors) correspond to different amounts of dry months (2 to 6). .... 62

Figure 6.13: Mortality cycle of the underlying parametrization (black line), selected points where climatic change is performed (dots)..... 64

Figure 6.14: Point of inflection (X0.5) in mm/yr derived from logistic regression for different shift dates within the mortality cycle of 500 years, based on an underlying uncorrected climate with different values of annual rainfall variation (SD=10% to 30%). The numbers indicate the amount of dry months (2-6DM) of the according graph. The dashed line presents a linear regression performed on the data to illustrate a possible trend..... 66

Figure 6.15: Point of inflection (X0.5) in mm/yr derived from logistic regression for different shift dates within the mortality cycle of 500 years, based on an underlying uncorrected climate with different values of annual rainfall variation (SD=32.5% to 40.0%). The numbers indicate the amount of dry months (2-6DM) of the according graph. The dashed line presents a linear regression performed on the data to illustrate a possible trend..... 67

Figure 6.16: Point of inflection (X0.5) and upper limit of the transition phase (U) in mm/yr derived from the logistic regression, for simulations based on uncorrected climate with 2-6 dry months and annual rainfall variation from SD=10% to SD=25%. Boxplots indicate the variation resulting from different shift dates in the mortality cycle, the dashed line represents the mean values. .... 68

Figure 6.17: Point of inflection (X0.5) and upper limit of the transition phase (U) in mm/yr derived from the logistic regression, for simulations based on uncorrected climate with 2-6 dry months and annual rainfall variation from SD=27.5% to SD=35%. Boxplots indicate the variation resulting from different shift dates in the mortality cycle, the dashed line represents the mean values. .... 69

Figure 6.18: Point of inflection (X0.5) and upper limit of the transition phase (U) in mm/yr derived from the logistic regression, for simulations based on uncorrected climate with 2-6 dry months and annual rainfall variation from SD=37.5% to SD=40%. Boxplots indicate the variation resulting from different shift dates in the mortality cycle, the dashed line represents the mean values. .... 70

Figure 6.19: Point of inflection (X0.5) in mm/yr derived from logistic regression for different shift dates within the mortality cycle of 500 years, based on an underlying corrected climate with different values of annual rainfall variation (SD=10% to 30.0%). The numbers indicate the amount of dry months (2-6DM) of the according graph. The dashed line presents a linear regression performed on the data to illustrate a possible trend..... 71

Figure 6.20: Point of inflection (X0.5) in mm/yr derived from logistic regression for different shift dates within the mortality cycle of 500 years, based on an underlying corrected climate with

different values of annual rainfall variation (SD=32.5% to 40.0%). The numbers indicate the amount of dry months (2-6DM) of the according graph. The dashed line presents a linear regression performed on the data to illustrate a possible trend.....72

Figure 6.21: Point of inflection (X0.5) and upper limit of the transition phase (U) in mm/yr derived from the logistic regression, for simulations based on corrected climate with 2-6 dry months and annual rainfall variation from SD=10% to SD=25%. Boxplots indicate the variation resulting from different shift dates in the mortality cycle, the dashed line represents the mean values. ....73

Figure 6.22: Point of inflection (X0.5) and upper limit of the transition phase (U) in mm/yr derived from the logistic regression, for simulations based on corrected climate with 2-6 dry months and annual rainfall variation from SD=27.5% to SD=35%. Boxplots indicate the variation resulting from different shift dates in the mortality cycle, the dashed line represents the mean values. ....74

Figure 6.23: Point of inflection (X0.5) and upper limit of the transition phase (U) in mm/yr derived from the logistic regression, for simulations based on corrected climate with 2-6 dry months and annual rainfall variation from SD=37.5% to SD=40.0%. Boxplots indicate the variation resulting from different shift dates in the mortality cycle, the dashed line represents the mean values. ....75

Figure 6.24: Exponential growth of mean point of inflection corrected (red symbols) versus uncorrected (blue symbols) in mm/yr with increasing year-to-year variation of rainfall (SD of the precipitation distribution). The dashed lines present the exponential regression. ....76

Figure 6.25: Exponential growth of mean length of transition phase (W) corrected (red symbols) versus uncorrected (blue symbols) in mm/yr with increasing year-to-year variation of rainfall (SD of the precipitation distribution). The dashed lines present the exponential regression. ....77

Figure 6.26: Length of transition phase (W) in percent of the according point of inflection (X0.5) (corrected vs. uncorrected) for 2DM to 6DM. Dots illustrate the mean, the boxplot provides information about the underlying distribution of relative W. ....78

Figure 6.27: Length of transition phase (W) in percent of the according point of inflection (X0.5) (corrected vs. uncorrected) for different shift dates in the mortality cycle. Dots illustrate the mean, the boxplot provides information about the underlying distribution of relative W. ....79

Figure 6.28: Above-ground living biomass carbon (= leaf carbon + dead stem carbon) from all successful simulations averaged over the last mortality cycle of each simulation. Carbon levels (in different colors) correspond to different amounts of dry months (2 to 6). ....79

Figure 6.29: Hysteresis: Instability is increasing differently for spinup (solid line) and spindown (dashed line). Exponential regression of mean point of inflection (symbols) in mm/yr with increasing year-to-year variation of rainfall (SD of the precipitation distribution). ....81

Figure 6.30: Hysteresis: only minor difference between spinup (solid line) and spindown (dashed line). Exponential regression of mean width of transitions phase (symbols) in mm/yr with increasing year-to-year variation of rainfall (SD of the precipitation distribution). ....82

Figure A.1: Example of a CLX file (Moanda, Gabon), monthly climate normals (bold).....89

Figure A.2: Extract of the header and first seven entries of a WTG file (Moanda, Gabon).....	90
Figure B.1: Box-and-Whisker plot, schematically.....	93
Figure B.2: Logistic regression function, schema-tically. Probability (P), predictor (x).....	94
Figure C.1: Flow chart of corrective procedure, main cycle indicated by highlighted boxes.....	96
Figure C.2: Residuals of solar radiation (sradCLX-sradWTG; [MJ/m <sup>2</sup> /day]), exemplary for Moanda . .....	99
Figure C.3: Monthly correction probabilities ( $0 < p < 1$ ), for Moanda exemplary.....	100
Figure C.4: Daily mean values of solar radiation [MJ/m <sup>2</sup> /day] averaged over N=99 simulation years, with leaps, for Moanda exemplary.....	101
Figure C.5: Daily mean values of solar radiation [MJ/m <sup>2</sup> /day] averaged over N=99 simulation years, smoothed, for Moanda exemplary. Black line = uncorrected, red line = corrected output.....	102
Figure C.6: Shifted correction probabilities ( $0 < p < 1$ ) over the year for a shift of 0 to 30 days....	103
Figure C.7: Daily mean rainfall [mm/day] (left scale) and uncorr. solar radiation [W/m <sup>2</sup> ] (right scale), averaged over N=99 simulation years, for Batéké Plateau, exemplary. Regions with lower rainfall exhibit higher amounts of incident solar radiation and vice versa (uncorrected radiation).....	104
Figure C.8: Integrated precipitation signal (from figure C.7), the plateau in the middle of the curve indicating the dry season becomes visible.....	105
Figure C.9: Mean error development of solar radiation.....	110
Figure C.10: Mean error development of maximum temperature.....	110
Figure C.11: Development of linear correlation, solar radiation.....	110
Figure C.12: Development of linear correlation, maximum temperature.....	110
Figure C.13: Dry seasons compared. Grey drop lines: literature values, black squares: computed mean dry seasons with dyreg=15, and light grey circles: computed values with an iteratively optimized dyreg (a desired dry season has been entered).....	111
Figure D.1: Example paleo climate reconstruction from pollen records for Lake Kamalété, Gabon, as described in Ngomanda, 2005. The grey line in the upper graph presents the mean annual precipitation from 99 climate years, the red line in the lower graph shows annual precipitation of selected climate years which were chosen to stay within a +/- 5% range of the original time series (black line).....	115

## List of Tables

Table 4.1: Coordinates and elevation of weather stations in Gabon, from Maloba Makanga 2010, p.25 (*sites where long term data in monthly resolution is available).....	10
Table 4.2: Coordinates and elevation of international tropical sites that will be used in an evaluation in chapter 5.1.....	10
Table 4.3: Annual precipitation mean, median, minimum and maximum value, 25%-quartile (Q1) and 75%-quartile (Q3) derived from long term data of three weather stations. Relative minimum, maximum, Q1 and Q3 in percent of the median (min%, max%, Q1% and Q3%, respectively), all values either in mm/yr or percent. Number of recorded years (N) within recording period from 1961 to 2000. ....	11
Table 4.4: Annual precipitation tested for normality (SW-test, p-value), standard deviation in % of the mean, number of records (N) within recording period from 1961 to 2000. ....	12
Table 4.5: Literature values of annual precipitation for sites in Gabon, after Maloba Makanga (2010), p. 83 and 86. Median, minimum and maximum value, 25%-quartile (Q1) and 75%-quartile (Q3). Relative minimum, maximum, Q1 and Q3 in percent of the median (min%, max%, Q1% and Q3%, respectively), all values either in mm/yr or percent. Recording period from 1951 to 1990. . .	12
Table 4.6: Mean monthly precipitation and standard deviation (sd) derived from data of three weather stations in mm/month. For number of recorded years (N) refer to table 4.3.....	13
Table 4.7: Annual maximum temperature mean, median, minimum and maximum value, 25%-quartile (Q1) and 75%-quartile (Q3) derived from data of three weather stations. Relative minimum, maximum, Q1 and Q3 in percent of the median (min%, max%, Q1% and Q3%, respectively), all values either in °C or percent. Number of recorded years (N) within recording period from 1961 to 2000. ....	13
Table 4.8: Mean monthly maximum temperature and standard deviation (sd) derived from data of three weather stations in °C. Number of recorded years (N) = 40.....	13
Table 4.9: Annual minimum temperature mean, median, minimum and maximum value, 25%-quartile (Q1) and 75%-quartile (Q3) derived from data of three weather stations. Relative minimum, maximum, Q1 and Q3 in percent of the median (min%, max%, Q1% and Q3%, respectively), all values either in °C or percent. Number of recorded years (N) within recording period from 1961 to 2000. ....	13
Table 4.10: Mean monthly minimum temperature and standard deviation (sd) derived from data of three weather stations in °C. Number of recorded years (N) = 40.....	14
Table 4.11: Hours of sunshine per month mean, median, minimum and maximum value, 25%-quartile (Q1) and 75%-quartile (Q3) derived from data of two weather stations. Relative minimum, maximum, Q1 and Q3 in percent of the median (min%, max%, Q1% and Q3%, respectively), all values either in hrs/month or percent. Number of recorded years (N) within recording period from 1961 to 2000. ....	14

Table 4.12: Hours of sunshine per month and standard deviation (sd) derived from data of two weather stations in °C. For number of recorded years (N) refer to table 4.11.....14

Table 5.1: Comparing precipitation of Gabonese and international sites: climate normal vs. simulation output (CLX vs. WTG). Pearson correlation (r), residual (D), trimmed residual (DT), standard deviation (SD), p-value (t-test), p-value (u-test), p-value (Shapiro-Wilk-test = SW-test), 95%-confidence-interval (+/-CI), 95%-prediction-interval (+/-PI).....17

Table 5.2: Comparing minimum temperature of Gabonese and international sites: climate normal vs. simulation output (CLX vs. WTG). Pearson correlation (r), residual (D), trimmed residual (DT), standard deviation (SD), p-value (t-test), p-value (u-test), p-value (Shapiro-Wilk-test = SW-test), 95%-confidence-interval (+/-CI), 95%-prediction-interval (+/-PI).....17

Table 5.3: Comparing maximum temperature of Gabonese and international sites: climate normal vs. simulation output (CLX vs. WTG). Pearson correlation (r), residual (D), trimmed residual (DT), standard deviation (SD), p-value (t-test), p-value (u-test), p-value (Shapiro-Wilk-test = SW-test), 95%-confidence-interval (+/-CI), 95%-prediction-interval (+/-PI).....18

Table 5.4: Comparing solar radiation of Gabonese and international sites: Climate normal vs simulation output (CLX vs WTG). Pearson correlation (r), residual (D), trimmed residual (DT), standard deviation (SD), p-value (t-test), p-value (u-test), p-value (Shapiro-Wilk-test = SW-test), 95%-confidence-interval (+/-CI), 95%-prediction-interval (+/-PI).....19

Table 5.5: Evaluating corrected maximum temperature of Gabonese sites: climate normal vs. corrected simulation output (CLX vs. corrected WTG). Pearson correlation (r), residual (D), trimmed residual (DT), standard deviation (SD), p-value (t-test), p-value (u-test), p-value (Shapiro-Wilk-test = SW-test), 95%-confidence-interval (+/-CI), 95%-prediction-interval (+/-PI).....20

Table 5.6: Evaluating corrected solar radiation of Gabonese sites: climate normal vs. corrected simulation output (CLX vs. corrected WTG. Pearson correlation (r), residual (D), trimmed residual (DT), standard deviation (SD), p-value (t-test), p-value (u-test), p-value (Shapiro-Wilk-test = SW-test), 95%-confidence-interval (+/-CI), 95%-prediction-interval (+/-PI).....21

Table 5.7: Validation of precipitation: Pearson correlation (r), average precipitation residual of all months (D), standard deviation (SD), p-value (t-test), p-value (Shapiro-Wilk-test = SW-test), 95%-confidence-interval (+/-CI), 95%-prediction-interval (+/-PI).....22

Table 5.8: Validation of minimum temperature: Pearson correlation (r), annual minimum temperature residual (D), standard deviation (SD), p-value (t-test), p-value (u-test), p-value (Shapiro-Wilk-test = SW-test), 95%-confidence-interval (+/-CI), 95%-prediction-interval (+/-PI).....24

Table 5.9: Validation of maximum temperature: Pearson correlation (r), annual maximum temperature residual (D), standard deviation (SD), p-value (t-test), p-value (u-test), p-value (Shapiro-Wilk-test = SW-test), 95%-confidence-interval (+/-CI), 95%-prediction-interval (+/-PI).....26

Table 5.10: Validation of temperature difference (corrected and uncorrected): Pearson correlation (r), annual temperature range residual (D), standard deviation (SD), p-value (t-test), p-value

(Shapiro-Wilk-test = SW-test), 95%-confidence-interval (+/-CI), 95%-prediction-interval (+/-PI).....	29
Table 5.11: Pearson correlation (r) between solar radiation [MJ/m <sup>2</sup> /day] and insolation [hrs/day] before (bc) and after correction (ac), for Moanda and Lastoursville.....	30
Table 5.12: Annual precipitation from N=99 generated years for sites in Gabon. Median, Mean, minimum (min) and maximum (max) value, 25%-quartile (Q1) and 75%-quartile (Q3) of N=99 climate years, all in mm/yr. ....	33
Table 5.13: Mean residual (observed-predicted) in percent of the according observed quantity, p-value (t-test), 95%-confidence interval also in percent of the according observed quantity. Refer to table 5.12 and table 4.5.....	33
Table 5.14: Precipitation analysis for 2DM and 3DM: annual rainfall as defined in the climate normal (clx), true median and mean, residual (D), further minimum and maximum value, 25%-quartile (Q1) and 75%-quartile (Q3) and standard deviation (SD). Relative minimum, maximum, Q1 and Q3 in percent of the median (min%, max%, Q1% and Q3%, respectively) and relative standard deviation in percent of the mean (SD%). Number of climate years within a climate (N). All values are either in mm/year or percent.....	37
Table 5.15: Precipitation analysis for 4DM, 5DM and 6DM: for a description of the listed quantities refer to table 5.14.....	39
Table 5.16: Mean values of relative minimum, maximum, Q1 and Q3 in percent of the according median and standard deviations, from table 5.14 & 5.15 and table 4.5 (Lastoursville and Mouila).	40
Table 6.1: Results from exponential regression of mean points of inflection derived from simulations with uncorrected climate.....	52
Table 6.2: Results from exponential regression of mean points of inflection derived from simulations with corrected climate.....	53
Table 6.3: Corrected (stratiform) vs uncorrected (cumuliform) estimates of instability: SD = standard deviation of the rainfall distribution [%], p-value (Shapiro-Wilk-test), p-value (t-test), D = X0.5(corr.) - X0.5(uncorr.), confidence interval (CI). One data-set corresponds to one value of SD and comprises 5 data points (= 2 dry months to 6 dry months) from the corrected, and 5 points from the uncorrected version.....	59
Table 6.4: Results from the exponential regression of the transition phase (W) for simulations based on uncorrected and corrected climates.....	60
Table 6.5: Estimates for point of inflection (X0.5), transition width and upper boundary of the transition phase. U (110%) is the most probable estimate for U, U (170%) constitutes a "worst case scenario". Lastoursville*: mean annual rainfall was taken from literature (table 4.5). ....	62
Table 6.6: Corrected (stratiform) vs. uncorrected (cumuliform) estimates of instability:SD = standard deviation of the rainfall distribution [%], p-value (Shapiro-Wilk-test), p-value (t-test), D = X0.5(corr.) - X0.5(uncorr.), confidence interval (CI). One data-set corresponds to one value of SD and comprises 40 data points (= 5 different lengths of dry season x 8 different shift dates in	

mortality cycle) from the corrected, and 40 points from the uncorrected version.....76

Table 6.7: Results from the exponential regression of points of inflection ( $X_{0.5}$ ) for simulations based on uncorrected and corrected climates.....77

Table 6.8: Results from the exponential regression of length of transition phase ( $W$ ) for simulations based on uncorrected and corrected climates.....77



# 1 Introduction

## 1.1 Background

Putting our focus on African tropical regions, including Southern Cameroon, Gabon and Congo as far as the eastern Congo Basin (figure 1.1), different authors have documented the heterogeneity in the composition of species of the present tropical flora for this region. One can observe evergreen forests as well as forests of type semi-deciduous, savannas and typical for some areas: forest-savanna mosaics, with pioneer species on their lines of intersection. Most of the authors agree on the idea that changing climatic conditions in the past have led to disturbances that

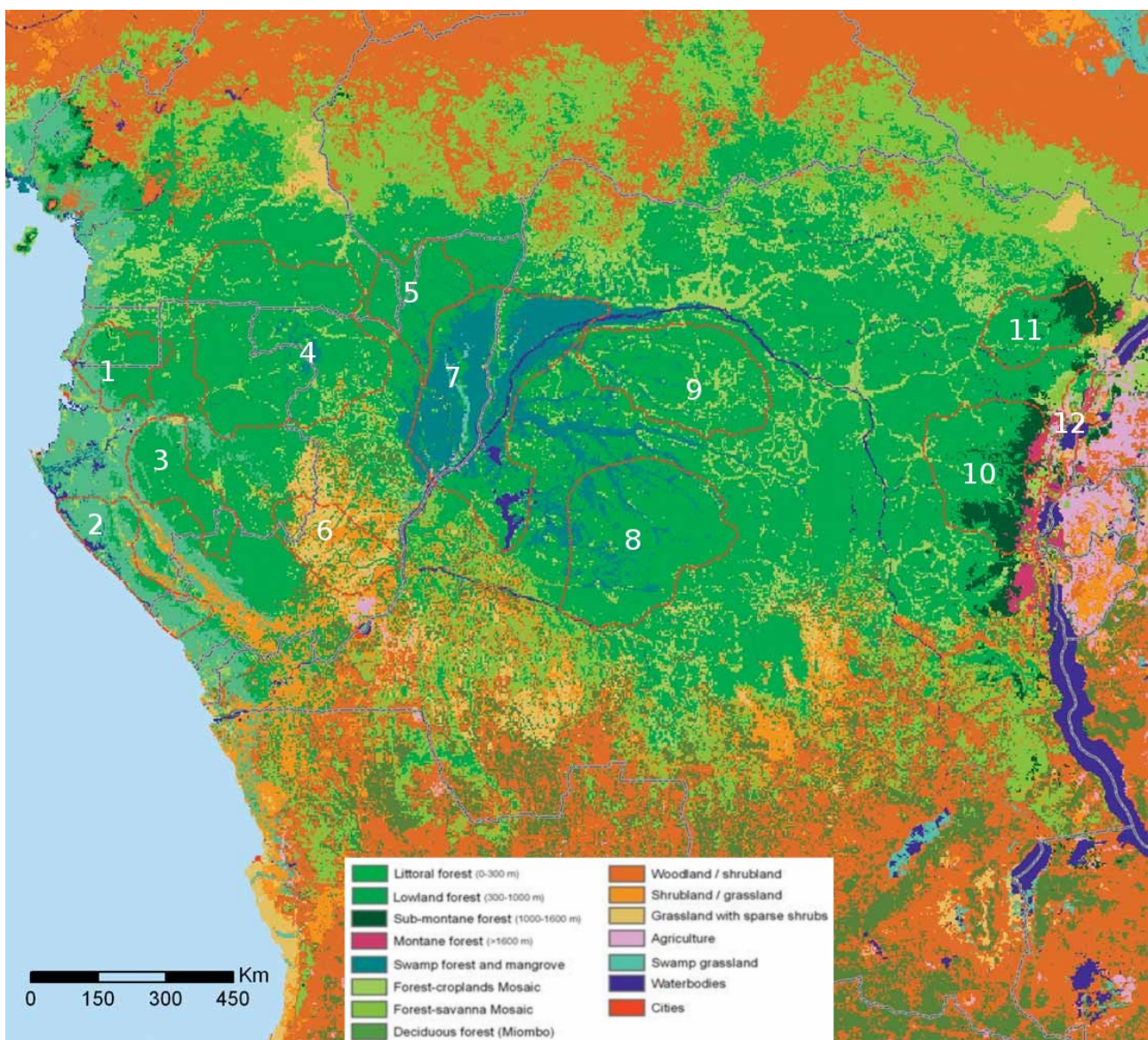


Figure 1.1: Vegetation map of eastern Central Africa (source: Joint Research Center). 1. Monte Alén-Monts de Cristal Landscape ; 2. Gamba-Mayumba-Conkouati Landscape; 3. Lopé-Chaillu-Louesse Landscape; 4. Dja-Odzala-Minkébé (Tridom) Landscape; 5. Sangha Tri-National Landscape; 6. Léconi-Batéké-Léfini Landscape ; 7. Lake Télé-Lake Tumba Landscape; 8. Salonga-Lukenie-Sankuru Landscape; 9. Maringa-Lopori-Wamba Landscape; 10. Maiko-Tayna-Kahuzi-Biega Landscape ; 11. Ituri-Epulu-Aru Landscape; 12. Virunga Landscape.

subsequently caused different stages in plant succession in the present picture (Elenga et al., 1994; Reynaud-Farrera et al., 1996; Zogning et al., 1997; Maley & Brenac, 1998a).

To trace down the effect of climate change on vegetation, interpreting the “stratification of fossil pollen records and other metabolic proxies” (Leal, 2004), which can help to determine the dominant species of the period of deposition, has become very popular (Ngomanda, 2005; Maley, 2001). Additionally lake level reconstructions based on similar methods can provide estimates of rainfall on a decadal or centennial scale. Especially J. Maley (Maley, 1997 and 2001), as well as A. Ngomanda (Ngomanda et al., 2009) have given rise to the idea, that not only a change in the total amount of rainfall, or the year-to-year variation of rainfall (which is hard to determine considering the rather rough resolution of paleo-climatic reconstructions), but also a shift in the distribution of rainfall within the year could have led to catastrophic events, such as massive forest break down in the past, as documented for about 3000-2500 BP by Maley, 2001 and Ngomanda, 2005.

The present-day climate dynamics for this region can be connected to sea surface temperatures (SST) which are controlled by the interaction of the Benguela Current (cold water) and the Gulf of Guinea (relatively warm water): If the cold water from the Benguela current is able to reach the sea surface (“upwelling”), which is usually the case during the long dry season in boreal summer, the humidity gradient is reversed as evaporation is replaced by condensation, which influences the direction of convection and as a result facilitates the formation of a stratiform cloud cover (Flohn, 1983). Stratiform clouds of this kind are usually non-precipitant and can therefore persist throughout the whole dry season, which significantly reduces the amount of incident solar radiation and thereby also temperature and evaporation. Different records of SST from 1963-1975 (Maley, 1997) prove the existence of anomalies in the Gulf of Guinea expressed as year-to-year fluctuations of SST of several degrees Celsius. Years have been recorded where the absence of the upwelling of cold water inhibited the formation of stratiform clouds, and lead to the generation of cumuliform clouds, which goes hand in hand with an increase of annual rainfall, solar radiation and also evaporation. A shift in the interaction of the two currents on a larger temporal scale could therefore have led to a change in the quality of the dry season and the distribution of rainfall during the year (Maley, 2001).

We can summarize that the climatic factors that might have disturbed forest growth in the past can be cut down to the amount of annual rainfall, the distribution of rainfall within the year, the year to year variation of precipitation and also the role of the cloud cover, influencing solar radiation input, temperature and evaporation. Ecosystem behavior of the Western Congolian Lowland Rainforest (Gabon) under different climatic setups will be subject of this work, based on generated climate and ecosystem model simulations.

## ***1.2 Availability of climate records in Gabon***

Gabon is a Central African state where, according to J.D. Maloba Makanga (Maloba Makanga, 2010), climatological or meteorological research as well as weather forecasting face two fundamental problems: Not only the bad quality of archived data, but also the low density of weather stations available. Actually, only 14 out of 97 stations in Gabon, a country with about three times the size of Austria (where more than 260 weather stations take measurements of daily precipitation, minimum and maximum temperature), can be accounted for as “synoptic stations”, where several weather variables are registered – most of the others only consist of a pluviometer or rain gauge, an instrument to capture and measure precipitation. Heterogeneity concerning the periods of observation in different stations make the recorded data in the most part unsuitable for interpolation. Furthermore 83% of all weather stations were installed between 1950 and 1959 and since the 1980ies we register a declining number of weather stations in Gabon.

Generally, considering homogeneous landscape, one precipitation station within a region of 100 km<sup>2</sup>, one thermo-hygrometric station measuring temperature and humidity for 500 km<sup>2</sup> and one station capturing air pressure, wind speed and solar radiation or insolation in about 5000 km<sup>2</sup> should be sufficient for regional studies, if the observation period covers at least 30 consecutive years (according to the World Meteorological Organization). For more complex terrain, taking Austria's mountainous regions as an example, a considerably higher density of weather stations is crucial for a detailed description of a local climate. Taking into account that 14 synoptic stations register data more or less frequently all over the country, the observation network is very weak, speaking of one station on an average of 19000 km<sup>2</sup>! Hence, according to the former director of Gabon's National Meteorological Service (“Service Nationale de la Météorologie”), the country requires an increase in the number of weather stations that at least doubles the current amount (Maloba Makanga, 2010).

## 2 Goals

This work's main focus is put on the stability of tropical forest ecosystems, like those described in the introduction. We want to uncover climatic triggers for system break down, and develop mechanisms to quantify criteria for stability or resilience. To force a shift towards unstable conditions, different climatic setups will be applied to an ecosystem model, that has been parametrized for the Western Congolian Lowland Rainforest which was shown to have existed for several thousand years (Pietsch, Tanga and Ngok-Banak, 2009), including those years where surrounding forest ecosystems faced massive break down or a change in their species composition (~2500 BP). Since tropical climate only exhibits minor differences in temperature in the course of the year, the meteorological parameters we want to investigate include the amount of annual precipitation, the distribution of rainfall within a year (which can be connected to the duration of the long dry season), and the year-to-year variation of precipitation, as well as the role of the cloud cover (cumuliform vs. stratiform). In order to customize these climatic conditions we will not use real climate, as recorded from a weather station, but need to make use of a stochastic weather generator. MarkSim is a weather generator that has been validated for various tropical regions around the world (Jones and Thornton, 2000), and that we will use for our research. First MarkSim will be tested with respect to precipitation, temperature and solar radiation for several sites in Gabon, Central Africa, where the eco-physiological parameters for Biome-BGC, the forest ecosystem model of our choice (Thornton, 1998; Thornton et al., 2002; Pietsch et al. 2003; Pietsch and Hasenauer 2006), have been measured and validated (Gautam and Pietsch, 2011 submitted). The quality of generated weather data for Gabon will then be compared with generated climate from other tropical sites in Africa and South America. Since the long dry season in Gabon is marked by a continuous stratiform cloud cover causing the lowest amount of incident solar radiation and coolest temperatures during the year, there is a possibility that MarkSim creates biased estimates of solar radiation and maximum temperature (which is derived from solar radiation) for this region. In a global context this cloud cover is a rather local phenomenon, and a bias could result from the procedure how solar radiation is calculated: The reduction of extraterrestrial radiation to global radiation, in MarkSim only depends on the appearance of precipitation. If we detect discrepancies in generated and observed maximum temperature and solar radiation, we will adapt MarkSim to fit accordingly. We will develop extensions for the weather generator that will allow us to customize climate, by defining the amount of annual precipitation, the duration of the dry season and the year-to-year variation of rainfall, without changing MarkSim's site specific parameters that determine daily precipitation patterns, and thereby patterns of temperature and solar radiation. Furthermore, we will introduce a way of generating paleo time series and climate change time series of daily weather. Generating vast amounts of climatic setups, including low precipitation regimes and climates exhibiting very long dry seasons of up to

half a year, these data will be used for Biome-BGC simulations to spot unstable regions and triggers for system break down in a statistical approach. The dependence of quantified measures of stability on well-defined climatic conditions will be investigated in the final section.

The working steps of this thesis can be summarized as follows:

- 1) a) MarkSim will be introduced as one possibility to create daily climate data for tropical Africa. The weather generator will be validated and adapted for sites in Gabon (chapter 5.1 - Validating and correcting stochastic climate for Gabon).  
b) Further methods will be introduced to customize climates with quantitatively defined meteorological parameters (chapter 5.2 - Customizing climate for BGC-simulations ).
- 2) Using the mechanistic ecosystem model Biome-BGC and applying large amounts of climate time series, ecosystem stability under different climatic conditions will be tested and according climatic thresholds will be identified (chapter 6.1 - Testing ecosystem stability under different climatic conditions).

## **3 Methods**

The computational methods utilized within this work comprise the stochastic weather generator MarkSim and the mechanistic ecosystem-model Biome-BGC. Statistical methods necessary for evaluations and validations are illustrated in APPENDIX B.

### **3.1 Weather generators**

Weather data originating from a stochastic weather generator are commonly produced in two steps: First, using the usually site specific statistical characteristics of one variable, e.g. precipitation (Richardson, 1981; Jones and Thornton, 1993) or temperature (Strandman et al., 1993) a time series of desired length on a daily scale is modeled including a stochastic element, such as a Markov chain approach. Next, all remaining variables, like daily maximum and minimum temperature, solar radiation, wind speed etc., are computed according to their correlations with the initially generated variable and the correlations with each other. A set of parameters for every month is required to reflect the seasonal trend of a generated year. The most typical approach, however, follows the Richardson model (Richardson, 1981), where missing weather variables are generated depending on the occurrence or non-occurrence of precipitation. One criticism of this approach is that solar radiation and maximum temperature strongly depend on the occurrence of significant cloud, which is not necessarily linked to precipitation (Hutchinson, 1995). Further, as described by Jones and Thornton (Jones and Thornton, 1993), both, the Richardson model (Richardson, 1981) and their own precipitation model, based on a third order Markov chain approach, fail to mimic adequately the length of dry or wet periods. Later, however, their model gave rise to the commercial weather generator MarkSim (Jones and Thornton, 2000). Examples for other weather generators are WGEN (Soltani et al., 2000) and an approach by Durban and Glasbey (Durban and Glasbey, 2001). Anyway, weather generators are not to be confused with climate interpolation tools, where geostatistical interpolation methods, such as kriging (Krige, 1951; Delfiner and Delhomme, 1975), thin plate smoothing splines (e.g. Hutchinson, 1995; Price et al., 2000) or a weighted Gaussian filter (Thornton et al., 1997, Hasenauer et al., 2003) create weather data for a specific site based on the data available in surrounding weather stations.

#### **3.1.1 Why to use weather generators?**

There are situations when the use of a weather generator remains the only option. Usually the motive to use a weather generator is limited availability of recorded data for the site of interest (Jones and Thornton, 1997). The constraints can be of different nature:

1. Limited number of weather variables in the database: While daily temperature and precipitation measurements may be available from official weather stations, humidity (VPD) or solar radiation are restricted to a small number of stations, and may have

to be estimated as functions of the available variables. In this case just some structural parts of the weather generator would be required to calculate the missing variables.

2. Limited number of years with climate records, or missing data in the record: Concerning weather variability a large number, usually a minimum of 30 recorded years (suggested by the World Meteorological Organization) is crucial in order to describe the climatic state of a site of interest including most of its variability.
3. Inappropriate temporal scale: Often weather data is only available on a monthly scale which may be insufficient for many research applications. Daily values therefore have to be generated using some sort of stochastic approach.
4. Insufficiently dense network of weather stations and lack of interpolation facilities.

For our research, however, it might be sufficient, and the only choice we have, to use a weather generator. One assumption that justifies the use of a weather generator is that for assessing the qualitative behavior of an ecosystem, *realistic*, but not necessarily *real* weather data are required. This implies, that we are not necessarily interested in the exact daily, monthly or even annual values, as long as weather patterns (duration and variability of the length, onset and ending of the dry season, number of dry vs. wet days, amount of precipitation on a wet day, etc.) are close to the real conditions, and the trend over decades is reasonable. We will even profit from the ability to customize our own climate, to test ecosystem robustness or stability under climatic stress or climate change conditions, which would not easily be possible with *real* weather data from the last 50 to 60 years. As this work goes hand in hand with the effort of a group of researchers at the Institute of Forest Growth Research, University of Natural Resources and Life Sciences Vienna, to model the Western Congolian Lowland Rainforest and savanna ecosystems and since the simulation software of choice, namely Biome-BGC (Thornton, 1998; Thornton et al., 2002; Pietsch et al. 2003; Pietsch and Hasenauer 2006) requires, among other parameters, daily weather input data and because climate records in Gabon are of insufficient quality for the analytic work performed in the course of this thesis, we chose to generate climate time series using the probabilistic weather generator MarkSim:

### **3.1.2 MarkSim: A probabilistic weather generator**

MarkSim's core algorithm is a rainfall generator based on a 3rd order Markov Chain approach to estimate daily precipitation values, extensively tested for applications within tropical regions (Jones and Thornton, 1993, 1997 and 2000). The occurrence and amount of rain on the last three days together with site specific parameter sets and a stochastic element decide upon the quality of an event on the present simulation day, while the amount of rain on a wet day is derived from a

truncated gamma distribution. Based on whether a day is dry or wet, daily maximum and minimum temperature are generated using the DSSAT weather generator (Pickering et al., 1994) based on routines of Richardson (Richardson, 1985) and Geng et al. (Geng et al., 1988). Therefore long-term monthly mean values are needed as input parameters. These climate normals are computed through an interpolation procedure, using data from about 10.000 weather stations for Latin America, 7000 for Africa and 4500 for Asia creating climate surfaces over a digital elevation map. Each set of surfaces comprises monthly precipitation, monthly average temperature, and monthly average diurnal temperature range. In order to generate solar radiation, a procedure suggested by Donatelli and Campbell (Donatelli and Campbell, 1997) is applied which reduces potential, or extra terrestrial radiation outside the earth's atmosphere by a transmissivity coefficient to an estimate of solar radiation on the earth's surface. Potential radiation is a function of the declination, the day of the year, and the latitude. Clear sky transmissivity, daily maximum and minimum air temperature and two empirical parameters define actual transmissivity, a further reduction of solar radiation is then performed on wet days.

Further information about MarkSim is provided in APPENDIX A.

### ***3.2 Biome-BGC: A biogeochemical ecosystem model***

For our studies we used the terrestrial ecosystem model Biome-BGC 4.1.1 (Thornton, 1998; Thornton et al., 2002; Pietsch et al. 2003; Pietsch and Hasenauer 2006).

Biome-BGC is a mechanistic computer model simulating pools and fluxes of mass and energy for different vegetation compartments (leaf, root, stem) and other external storages (e.g. soil, litter) on a daily basis. Biomass can therefore be seen as evenly spread over the whole environment: the unit for total carbon content, for example, is given in kg / m<sup>2</sup>.

The most important structural parts of this ecosystem model include the leaf area index (LAI), which is the ratio of the total one-sided leaf surface and the ground surface on which vegetation is growing. LAI controls canopy radiation absorption and thereby influences photosynthesis, water interception from rainfall, and litter inputs, having a direct impact on decomposition. GPP (gross primary production) is calculated via a routine suggested by Farquhar et al. (Farquhar et al., 1980), NPP (net primary production) is GPP minus autotrophic respiration. Autotrophic respiration is split into maintenance respiration and growth respiration which refers to the release of CO<sub>2</sub> for generating energy and other metabolic intermediates to maintain a healthy state or to support plant growth, respectively. Maintenance respiration depends on the concentration of tissue nitrogen (Ryan, 1991), whereas growth respiration is a function of carbon allocated to different plant compartments. Carbon not consumed by autotrophic respiration (NPP) is dynamically allocated to leaf, root and stem pools regarding the availability of and competition for nitrogen.



The model requires certain inputs, which include daily weather data, such as minimum and maximum temperature, solar radiation, precipitation and vapour pressure deficit. Further, elevation, aspect, physiological soil properties, atmospheric CO<sub>2</sub> and nitrogen deposition as well as a set of eco-physiological parameters are required to compute: “daily canopy interception, evaporation and transpiration; soil evaporation, outflow, water potential and water content; LAI; stomatal conductance and assimilation of sunlight and shaded canopy fractions; growth and maintenance respiration; GPP and NPP; allocation; litter-fall and decomposition; mineralization, denitrification, leaching and volatile losses” (Pietsch and Hasenauer, 2002).

Biome-BGC is a dissipative dynamical system with persisting energy input through solar radiation, and energy losses mainly through latent heat by plant transpiration and soil evaporation. Note, that energy lost this way is not transferred to another form of energy, as it does not have any effect on the micro climate (i.e. on vapour pressure deficit or temperature). In this study we used eco-physiological parameters for the Western Congolian Lowland Rainforest biome established by Gautam and Pietsch, 2011 submitted.

To perform a simulation using Biome-BGC the model is usually initialized with little leaf carbon on which growth is based. Biomass is accumulated depending on the quality of the climate. By definition a stationary state is reached as soon as fluctuations in the soil carbon pool, which is the slowest changing pool, are reduced to a minimum. In most of the cases measurements are taken as soon as the system is stationary. Unfavourable climate, however, can also lead to system break down characterized by a leaf carbon pool carrying so little carbon, that growth in the next season is inhibited, which immediately results in zero productivity. Leaf C is then set to zero once passing a certain threshold. Another option to initialize a simulation is offered by using a so-called “restart-file”, where all necessary information about a system state to start from are stored.

## 4 Data

This chapter lists data from literature and weather stations required for the validation of MarkSim. All data required for the BGC simulations will be provided later in chapter 5.2 (Customizing climate for BGC-simulations ).

Generated data comes from a series of 99 simulation years (the maximum number of years that can be generated during one MarkSim simulation with one simulation seed), monthly mean values are computed from daily values. MarkSim's output provides maximum temperature (tmax) [°C], minimum temperature (tmin) [°C], precipitation (prcp) [mm/month] and solar radiation (srad) [MJ/m<sup>2</sup>/day]. For the evaluation we chose 13 sites from Gabon, were longitude, latitude and

*Table 4.1: Coordinates and elevation of weather stations in Gabon, from Maloba Makanga 2010, p.25 (\*sites where long term data in monthly resolution is available)*

<b>SITE</b>	<b>LATITUDE</b>	<b>LONGITUDE</b>	<b>ELEVATION m.a.s.l</b>
1 Bitam	02°05' N	11°29' E	600
2 Cocobeach	01°00' N	09°36' E	8
3 Franceville	01°38' S	13°34' E	424
4 Lambaréné	00°43' S	10°14' E	26
5 Lastoursville*	00°50' S	12°43' E	483
6 Libreville	00°27' N	09°25' E	12
7 Makokou	00°34' N	12°52' E	513
8 Mayumba	03°25' S	10°39' E	31
9 Mékambo	01°01' N	13°56' E	499
10 Mitzic	00°41' N	11°32' E	583
11 Mouila*	01°52' S	11°01' E	88
12 Port-Gentil	00°42' S	08°45' E	3
13 Tchibanga	02°51' S	11°01' E	83
14 Moanda*	01°32' S	13°16'30" E	572

elevation correspond to weather stations found at the specific site (Maloba Makanga, 2010, p.25) and 6 international sites in tropical regions. The determination of elevation of the international sites, is based on MarkSim's digital elevation map (DEM).

*Table 4.2: Coordinates and elevation of international tropical sites that will be used in an evaluation in chapter 5.1.*

<b>SITE</b>	<b>LATITUDE [°N]</b>	<b>LONGITUDE [°W]</b>	<b>ELEVATION [m.a.s.l.]</b>
15 Mombasa, Kenia	-4,1	39,7	61
16 Nairobi, Kenia	-1,3	36,8	1661
17 Mogadishu, Somæ	2,1	45,3	12
18 Usangi, Tanzania	-3,7	37,7	300
19 Bogotá, Colombia	4,6	-74,1	2640
20 Manaus, Brazil	-3,1	-60,0	92



Figure 4.1: Map of tropics in and outside of Gabon, as listed in table 4.1 and 4.2, Latin American & western African sites (blue, 16-20), Gabonese sites from Maloba Makanga 2010, p.25 (red, yellow, green; 1-14), Gabonese sites used for comparison in section 5 (yellow & green), sites where long-term weather data are available (green).

Additionally recorded weather data from 1961 to 2000 of three weather stations in Gabon (Moanda, Lastoursville, Mouila) were available in monthly resolution. The measured variables are  $t_{max}$  [°C],  $t_{min}$  [°C],  $prcp$  [mm/month] and insolation, or sun shine hours [hrs/month], which makes a direct comparison of incident solar radiation difficult. Due to years of missing observation, some of the recorded data sets are fairly short (~25 years). Data sets that differ in size from N=40 years are: Lastoursville:  $prcp$  (N=25), Moanda:  $prcp$  (N=29), insolation (N=34), Mouila:  $prcp$  (N=35). In Mouila, however, no insolation has been recorded at all.

Monthly mean values of the measured variables as well a quartile analysis of the annual values (mean, median, minimum, 25%-quartile, 75%-quartile, maximum) for Moanda, Mouila and Lastoursville are provided in tables 4.3 and 4.6 to 4.12 below. Further a similar representation of annual precipitation of 13 sites in Gabon (table 4.5) derived from Maloba Makanga, 2010, will be utilized for the evaluation of rainfall distributions.

Table 4.3: Annual precipitation mean, median, minimum and maximum value, 25%-quartile (Q1) and 75%-quartile (Q3) derived from long term data of three weather stations. Relative minimum, maximum, Q1 and Q3 in percent of the median (min%, max%, Q1% and Q3%, respectively), all values either in mm/yr or percent. Number of recorded years (N) within recording period from 1961 to 2000.

Site	Annual precipitation [mm/yr]										
	mean	median	min	Q1	Q3	max	min%	Q1%	Q3%	max%	N
Lastoursville	1317	1172	799	975	1535	2304	68	83	131	197	25
Moanda	1940	1974	777	1792	2221	3048	39	91	113	154	29
Mouila	2069	1988	1079	1755	2352	3473	54	88	118	175	35

Table 4.4: Annual precipitation tested for normality (SW-test, p-value), standard deviation in % of the mean, number of records (N) within recording period from 1961 to 2000.

Site	Variation of annual precipitation [mm/yr]		
	SW-test (p-val)	SD%	N
Lastoursville	0,01	34,5	25
Lastoursville *		25,6	25
Moanda	0,52	24,1	29
Mouila	0,56	24,4	35

\* literature mean of 1778 mm/yr was taken to compute relative SD

Table 4.5: Literature values of annual precipitation for sites in Gabon, after Maloba Makanga (2010), p. 83 and 86. Median, minimum and maximum value, 25%-quartile (Q1) and 75%-quartile (Q3). Relative minimum, maximum, Q1 and Q3 in percent of the median (min%, max%, Q1% and Q3%, respectively), all values either in mm/yr or percent. Recording period from 1951 to 1990.

Site	Literature annual precipitation [mm/yr]								
	median	min	Q1	Q3	max	min%	Q1%	Q3%	max%
Bitam	1735	1210	1555	1831	2526	70	90	106	146
Cocobeach	3242	1668	2646	3570	4715	51	82	110	145
Franceville	1870	1193	1733	2055	2333	64	93	110	125
Lambaréné	1961	1397	1825	2148	2721	71	93	110	139
Lastoursville	1778	851	1561	1955	2422	48	88	110	136
Libreville	2840	1857	2506	3265	3981	65	88	115	140
Makokou	1693	1285	1521	1866	3209	76	90	110	190
Mayumba	1801	746	1493	2109	2875	41	83	117	160
Mékambo	1610	947	1428	1820	2184	59	89	113	136
Mitzic	1772	787	1544	1853	2412	44	87	105	136
Mouila	2205	1325	1891	2388	3006	60	86	108	136
Port-Génil	1976	1149	1683	2304	3099	58	85	117	157
Tchibanga	1465	802	1243	1577	1954	55	85	108	133

Note the difference of the median of annual precipitation for Lastoursville, derived from the weather station data (1172 mm/yr) and from literature (1778 mm/yr). The recorded data set with the lower value, however, has only 25 recorded years in the period from 1961 to 2000. Bad recording quality might also be the reason why Lastoursville is the only site where annual rainfall is not normally distributed, as suggested by the Shapiro-Wilk test at a 95% level of significance. Computing the relative standard deviation using the literature mean yields a reasonable result of SD=25.6%.

Table 4.6: Mean monthly precipitation and standard deviation (sd) derived from data of three weather stations in mm/month. For number of recorded years (N) refer to table 4.3.

	Monthly precipitation [mm/month]					
	Lastoursville		Moanda		Mouila	
	mean	sd	mean	sd	mean	sd
Jan	91	67	172	80	202	104
Feb	117	83	183	81	242	167
Mar	173	89	249	104	230	104
Apr	166	95	252	154	232	95
May	183	87	194	121	163	74
Jun	42	52	46	53	23	41
Jul	5	8	11	18	6	9
Aug	14	27	16	23	8	7
Sep	108	73	102	54	49	73
Oct	177	131	264	135	310	155
Nov	144	128	267	135	367	153
Dec	96	85	184	90	235	87

Table 4.7: Annual maximum temperature mean, median, minimum and maximum value, 25%-quartile (Q1) and 75%-quartile (Q3) derived from data of three weather stations. Relative minimum, maximum, Q1 and Q3 in percent of the median (min%, max%, Q1% and Q3%, respectively), all values either in °C or percent. Number of recorded years (N) within recording period from 1961 to 2000.

Site	Annual average maximum temperature [°C]										
	mean	median	min	Q1	Q3	max	min%	Q1%	Q3%	max%	N
Lastoursville	28,3	28,2	27,6	28,1	28,5	29,1	98	100	101	103	40
Moanda	28,4	28,3	27,3	28,1	28,8	29,5	96	99	102	104	40
Mouila	30,4	30,3	29,5	30,1	30,7	31,1	97	99	101	103	40

Table 4.8: Mean monthly maximum temperature and standard deviation (sd) derived from data of three weather stations in °C. Number of recorded years (N) = 40.

	Monthly maximum temperature [°C]					
	Lastoursville		Moanda		Mouila	
	mean	sd	mean	sd	mean	sd
Jan	28,8	0,7	28,8	0,7	31,4	0,6
Feb	29,5	0,6	29,6	0,7	32,1	0,6
Mar	29,9	0,7	30,0	0,7	32,3	1,0
Apr	29,7	0,7	29,9	0,7	32,5	0,7
May	28,8	0,8	28,9	0,6	31,0	1,1
Jun	27,1	0,6	27,2	0,7	28,3	0,8
Jul	25,6	0,8	26,3	0,6	27,2	0,9
Aug	26,1	0,7	26,9	0,5	27,5	0,6
Sep	28,3	0,6	28,2	0,6	29,4	0,6
Oct	28,6	0,5	28,6	0,6	30,9	0,6
Nov	28,5	0,5	28,5	0,9	31,0	0,6
Dec	28,2	0,6	28,1	1,1	30,7	0,9

Table 4.9: Annual minimum temperature mean, median, minimum and maximum value, 25%-quartile (Q1) and 75%-quartile (Q3) derived from data of three weather stations. Relative minimum, maximum, Q1 and Q3 in percent of the median (min%, max%, Q1% and Q3%, respectively), all values either in °C or percent. Number of recorded years (N) within recording period from 1961 to 2000.

Site	Annual average minimum temperature [°C]										
	mean	median	min	Q1	Q3	max	min%	Q1%	Q3%	max%	N
Lastoursville	20,6	20,2	19,2	19,8	21,3	23,5	95	98	105	116	40
Moanda	19,9	19,9	19,0	19,6	20,1	20,9	96	98	101	105	40
Mouila	22,1	22,1	21,4	22,0	22,3	22,7	97	99	101	103	40

Table 4.10: Mean monthly minimum temperature and standard deviation (sd) derived from data of three weather stations in °C. Number of recorded years (N) = 40.

	Monthly minimum temperature [°C]					
	Lastoursville		Moanda		Mouila	
	mean	sd	mean	sd	mean	sd
<b>Jan</b>	21,2	1,2	20,3	0,4	22,7	0,5
<b>Feb</b>	21,1	1,3	20,3	0,6	22,6	0,6
<b>Mar</b>	21,1	1,2	20,3	0,6	22,6	0,5
<b>Apr</b>	21,3	1,4	20,5	0,5	22,8	0,4
<b>May</b>	21,0	2,1	20,4	0,5	22,9	0,4
<b>Jun</b>	20,2	1,0	19,4	0,5	21,4	0,6
<b>Jul</b>	19,2	1,0	18,6	0,7	20,1	0,6
<b>Aug</b>	19,5	1,0	19,0	0,6	20,6	0,6
<b>Sep</b>	20,4	1,0	19,7	0,4	21,9	0,5
<b>Oct</b>	20,7	1,2	19,9	0,4	22,6	0,3
<b>Nov</b>	20,8	1,3	19,9	0,4	22,5	0,4
<b>Dec</b>	20,9	1,3	20,1	0,5	22,6	0,4

Table 4.11: Hours of sunshine per month mean, median, minimum and maximum value, 25%-quartile (Q1) and 75%-quartile (Q3) derived from data of two weather stations. Relative minimum, maximum, Q1 and Q3 in percent of the median (min%, max%, Q1% and Q3%, respectively), all values either in hrs/month or percent. Number of recorded years (N) within recording period from 1961 to 2000.

Site	Average hours of sunshine per month										N
	mean	median	min	Q1	Q3	max	min%	Q1%	Q3%	max%	
<b>Lastoursville</b>	131	130	107	124	137	157	82	95	106	121	40
<b>Moanda</b>	123	125	99	114	131	141	79	91	104	112	34

Table 4.12: Hours of sunshine per month and standard deviation (sd) derived from data of two weather stations in °C. For number of recorded years (N) refer to table 4.11.

	Hours of sunshine per month			
	Lastoursville		Moanda	
	mean	sd	mean	sd
<b>Jan</b>	151	26	142	19
<b>Feb</b>	154	24	138	21
<b>Mar</b>	166	20	154	20
<b>Apr</b>	160	20	144	24
<b>May</b>	146	27	131	27
<b>Jun</b>	119	28	96	30
<b>Jul</b>	82	24	94	25
<b>Aug</b>	74	23	82	23
<b>Sep</b>	108	16	97	21
<b>Oct</b>	135	16	127	17
<b>Nov</b>	138	21	139	16
<b>Dec</b>	139	25	137	20

## 5 Analysis and improvements

MarkSim has been introduced as the method of choice to produce daily weather data required for the assessment of ecosystem stability using Biome-BGC. This chapter addresses the evaluation of generated weather and the provision of long-term climate time series needed for simulations using Biome-BGC. This part of the work can be structured as follows:

- 1) First, in chapter 5.1 - Validating and correcting stochastic climate for Gabon - daily weather variables will be examined, and adapted if necessary. Results will be validated using data from three weather stations in Gabon. The distribution of annual rainfall of generated climate will then be compared to literature values.
- 2) Second, in chapter 5.2 - Customizing climate for BGC-simulations - we will provide long-term climate time series for the BGC-simulations (in chapter 6.1 - Testing ecosystem stability under different climatic conditions). Since we want to determine thresholds for ecosystem stability it is important that the exact value for mean annual precipitation as well as its variation and the distribution of rainfall within the year are parameters that we are able to define and alter in a quantified way.

### 5.1 Validating and correcting stochastic climate for Gabon

First, it will be illustrated that discrepancies in solar radiation and maximum temperature is a problem typical for Gabon, comparing MarkSim's climate normals with the actual simulation output. As we will show MarkSim produces contradictory results for solar radiation and maximum temperature, while precipitation and minimum temperature seem reasonable. A correction procedure for maximum temperature and solar radiation on a daily basis is introduced in APPENDIX C. For sites within Gabon generated weather output will be tested against the climate normals to underline need for a correction procedure.

#### 5.1.1 Comparison of international tropical sites with sites in Gabon

In this section the focus is put on the main meteorological variables (precipitation = prcp, minimum temperature = tmin, maximum temperature = tmax and incident solar radiation = srad), and how well generated weather data fits to MarkSim's intrinsic climate normals (CLX vs. WTG, APPENDIX A). It is crucial to understand that no external data are used for this comparison, all information comes from MarkSim itself. A measure for the seasonality, as described in APPENDIX B, is Pearson's correlation coefficient  $r$ . For this test we expect  $r$  values close to +1 (proper representation of seasonality), close to -1 (inverse representation of seasonality) or close to zero (no correlation at all). The deviation of monthly predictions from climate normals is given by the residual  $D$ , its standard deviation  $SD$ , 95%-confidence interval  $CI$  and 95%-prediction interval  $PI$ .  $D$ ,  $SD$ ,  $CI$  and  $PI$  all have the same unit as the variable in question. If the data set fails the Shapiro-

Wilk-test (Shapiro and Wilk, 1965:  $p < 0.05$ ), i.e. they are not normally distributed, then CI and PI will be determined using the trimmed estimates for the bias  $D^T$  according to Rauscher, 1986, and instead of performing a t-test, a non-parametric test, namely the U-test (Wilcoxon, 1945; Mann and Whitney 1947) will be applied. Monthly averages of each sample are computed from  $N=99$  generated years. For the comparison we chose six sites from tropical regions in western Africa and Latin America, and nine sites within Gabon that cover most of the country's climatic regions (Maloba Makanga, 2010; see Figure 4.1). All results are derived from uncorrected MarkSim output.

Note, that while making a comparison between simulation output and climate normals, 12 monthly residuals averaged to form one annual residual, which can be either biased or unbiased. Whether this annual residual is biased or not, does not tell us how each predicted monthly value mimics its corresponding climate normal. But we can interpret the statistics as follows: If the generated data fit well to the according climate normals, we will obtain a narrow confidence interval, and no bias. If the seasonal trend of the generated data and the climate normals diverge, we might get a wider confidence interval since the residuals contributed by each month differ highly in size, also resulting in unbiased annual means. If for example, the generated data are able to mimic the seasonal trend, but overestimate the climate normal by the same amount in each month, this results in a biased annual mean since the confidence interval becomes narrow. Therefore “unbiased” results do not always indicate good quality of prediction vs observation (or simulation output and climate normal). A “biased” output, on the other hand, does not tell us that the data are useless, since we can account for an over- or underestimation by redefining the climate normals, if the bias is the same in all months which is indicated by a narrow CI. To get an understanding for the origin of a possible bias (is it due to a bad seasonal trend or due to a constant over- or underestimation?) the correlation coefficient is introduced as an indicator of the quality of a seasonal trend.

### ***Precipitation***

Since the evaluation is based on monthly values, the residual of the annual value is also given in units of mm/month, which results from the fact that precipitation is an accumulative variable, compared to temperature and solar radiation, where monthly and annual values should stay within the same scale.

As for the minimum temperature, we cannot identify differences in the quality of predicted precipitation comparing sites in Gabon and international sites. D, PI and CI are generally of the same scale and also the linear correlation is high for all sites: [0.81,1.00] is the range for Gabon, and [0.65,1.00] the range for international sites. The rainfall generator seems to face difficulties in predicting the right amount of precipitation with the right distribution over all months in Usangi, where we observe the lowest correlation ( $r=0.65$ ) accompanied by the highest residual ( $D=18.2$



mm/month) and the widest CI of the international sites. Since our interest concerns Gabonese sites only, we wont put further attention to this.

Table 5.1: Comparing precipitation of Gabonese and international sites: climate normal vs. simulation output (CLX vs. WTG). Pearson correlation (*r*), residual (*D*), trimmed residual (*DT*), standard deviation (*SD*), *p*-value (*t*-test), *p*-value (*u*-test), *p*-value (Shapiro-Wilk-test = SW-test), 95%-confidence-interval (+/-CI), 95%-prediction-interval (+/-PI)

SITE	<i>r</i>	<i>D</i>	<i>D</i> <sup>T</sup>	PRCP [mm/month]			+/- CI	+/- PI	
				<i>SD</i>	<i>p</i> (t-test)	<i>p</i> (u-test)			
<i>INTERNATIONAL</i>									
Mombasa	0,98	-0,2		15,6	0,97		0,68	9,9	35,8
Nairobi	0,99	-3,7		9,1	0,19		0,35	5,8	21,0
Mogadishu	0,96	-10,4	-8,6	16,8		0,13	<b>0,03</b>	9,4	31,1
Bogotá	1,00	1,6		3,8	0,16		1,00	2,4	8,6
Usangi	0,65	18,2		39,5	0,14		0,76	25,1	90,6
Manaus	0,98	0,6		17,9	0,91		0,23	11,4	41,1
<i>GABON</i>									
Mitzic	0,81	16,6		64,5	0,39		0,88	41,0	147,8
Libreville	1,00	2,2		17,0	0,67		0,77	10,8	38,9
Makokou	0,99	6,3		16,4	0,21		0,14	10,4	37,6
Mékambo	0,99	-6,6		13,0	0,11		0,65	8,3	29,8
Mouila	0,98	3,1		26,4	0,70		0,45	16,8	60,4
Tchibanga	0,96	0,4		27,5	0,97		0,21	17,4	62,9
Mayumba	0,99	-2,6		15,7	0,57		0,20	9,9	35,9
Moanda	0,99	1,4		10,2	0,64		0,88	6,5	23,4
Lastoursville	0,97	-9,7		21,7	0,15		0,27	13,8	49,7

## Minimum temperature

Table 5.2: Comparing minimum temperature of Gabonese and international sites: climate normal vs. simulation output (CLX vs. WTG). Pearson correlation (*r*), residual (*D*), trimmed residual (*DT*), standard deviation (*SD*), *p*-value (*t*-test), *p*-value (*u*-test), *p*-value (Shapiro-Wilk-test = SW-test), 95%-confidence-interval (+/-CI), 95%-prediction-

SITE	<i>r</i>	<i>D</i>	<i>D</i> <sup>T</sup>	TMIN [°C]			+/- CI	+/- PI	
				<i>SD</i>	<i>p</i> (t-test)	<i>p</i> (u-test)			
<i>INTERNATIONAL</i>									
Mombasa	1,00	0,02		0,13	0,60		0,31	0,08	0,29
Nairobi	0,99	0,03	0,05	0,16		0,23	<b>0,01</b>	0,07	0,23
Mogadishu	1,00	0,01		0,15	0,78		0,12	0,10	0,34
Bogotá	0,96	0,06		0,18	0,27		0,65	0,11	0,41
Usangi	1,00	0,01		0,09	0,60		0,31	0,05	0,20
Manaus	0,93	-0,01		0,13	0,81		0,18	0,09	0,31
<i>GABON</i>									
Mitzic	0,98	0,02		0,15	0,72		0,99	0,09	0,33
Libreville	0,99	0,05		0,13	0,19		0,87	0,08	0,31
Makokou	1,00	0,02		0,09	0,41		0,87	0,06	0,20
Mékambo	0,99	-0,02		0,16	0,65		0,81	0,10	0,37
Mouila	1,00	0,02		0,23	0,73		0,42	0,15	0,54
Tchibanga	1,00	< 0,01		0,14	0,91		0,07	0,09	0,32
Mayumba	1,00	0,01		0,18	0,82		0,63	0,11	0,40
Moanda	0,99	0,01		0,13	0,78		0,31	0,09	0,31

*interval (+/-PI)*

For the minimum temperature (table 5.2) we cannot detect any major differences between Gabon

and international sites. All data are unbiased,  $|D|$  is continuously below 0.10 for all locations and the largest PI and CI appear in Lastoursville, which are still narrow compared to the intervals observed for maximum temperature (table 5.3). The accurate representation of seasonality of  $t_{min}$  is outlined by  $r$  values close to 1.

### Maximum temperature

Data in table 5.3 indicate unbiased predictions in terms of significantly high or low D-values for all of the international sites and sites within Gabon, supported with p-values (from the t-test) far away from the 0.05 threshold. Comparing the residuals Gabon's sites are better off (with its maxima in Libreville,  $D=0.07^{\circ}\text{C}$  and Lastoursville,  $D=-0.07^{\circ}\text{C}$ ) compared to Mogadishu ( $D=0.27^{\circ}\text{C}$ ) and Usangi ( $D=0.26^{\circ}\text{C}$ ). More interesting, though, is the size of SD, CI and PI as they give an idea of

Table 5.3: Comparing maximum temperature of Gabonese and international sites: climate normal vs. simulation output (CLX vs. WTG). Pearson correlation ( $r$ ), residual ( $D$ ), trimmed residual ( $D^T$ ), standard deviation ( $SD$ ), p-value (t-test), p-value (u-test), p-value (Shapiro-Wilk-test = SW-test), 95%-confidence-interval (+/-CI), 95%-prediction-interval (+/-PI)

SITE	r	D	$D^T$	TMAX [ $^{\circ}\text{C}$ ]			+/- CI	+/- PI
				SD	p (t-test)	p (u-test)		
<i>INTERNATIONAL</i>								
Mombasa	0,97	0,03		0,53	0,83		0,34	1,22
Nairobi	0,81	0,01		1,03	0,97		0,44	2,35
Mogadishu	0,71	0,27		0,86	0,30		0,74	1,98
Bogotá	0,85	0,02		0,34	0,87		0,85	0,79
Usangi	0,97	0,26		0,67	0,20		0,62	1,52
Manaus	0,98	0,03	0,07	0,32		0,30	<b>0,03</b>	0,40
<i>GABON</i>								
Mitzié	0,55	0,01		1,03	0,98		0,52	2,37
Libreville	0,81	0,07		0,74	0,75		0,56	1,70
Makokou	0,51	0,01		1,31	0,99		0,45	3,00
Mékambo	0,70	-0,01		1,10	0,96		0,53	2,52
Mouila	0,61	-0,01		0,98	0,98		0,10	2,25
Tchibanga	0,77	0,02		1,27	0,95		0,45	2,90
Mayumba	0,90	0,01		0,86	0,98		0,66	1,96
Moanda	0,53	0,01		0,94	0,99		0,45	2,15
Lastoursville	0,69	-0,07		1,33	0,85		0,30	3,04

how far monthly residuals are scattered around their annual mean. These measures indicate that predictions for sites in Gabon exhibit a higher uncertainty, i.e. higher discrepancies between  $D$ , the estimator for the bias, and the true bias, as well as broader confidence intervals. This fact is accompanied by lower correlation coefficients, suggesting that the seasonal trend is met better by the international sites, with a range for  $r$  of 0.71 to 0.98 compared to 0.51 – 0.90 for Gabon. As explained before, improper representation of seasonality leads to broader CI and PI while the  $D$  value might still stay close to zero. In Usangi we see a high correlation ( $r=0.97$ ), a CI in the lower range (+/-CI =  $0.42^{\circ}\text{C}$ ) and a large residual, compared to the other sites. This indicates a “constant underestimation scenario” (constant underestimation in all months), as the seasonality seems good, the CI is narrow but  $D$  is still high and furthermore, we observe a lower p-value ( $p=0.20$ ). The

other extreme can be seen in Moanda, where the annual means of prediction and climate normal are virtually equal ( $D=0.0027$ , not rounded) and the CI is larger than before in Usangi ( $\pm CI=0.60^{\circ}\text{C}$ ). But this time the width of the CI is caused by an improper seasonality ( $r=0.53$ ), which indicates scattered monthly residuals. These scattered monthly values lead to a behavior that is not detected by the t-test, which indicates no significant differences between predicted and observed data ( $p=0.99$ , table 5.3).

### Solar radiation

The monthly values of solar radiation seem to be the most difficult to predict, as we can read from table 5.4. Usually the range for srad values over the year is roughly between 15 and 25 MJ/m<sup>2</sup>/day, speaking of a monthly average. The difference in the quality of generated srad is reflected in by the r values: For international sites the correlation coefficient has its lowest value of  $r=0.69$  in Mogadishu and its highest of  $r=0.89$  in Mombasa, while for Gabon r is spread between  $-0.72$  in Moanda, and its highest value of 0.50 in Mekambo.  $r=-0.72$  means that the average generated trend is close to the inverse trend of the climate normals! Nevertheless the annual means are all unbiased due to a CI of  $\pm 2^{\circ}\text{C}$ , caused by the inaccurate representation of the seasonality.

Table 5.4: Comparing solar radiation of Gabonese and international sites: Climate normal vs simulation output (CLX vs WTG). Pearson correlation ( $r$ ), residual ( $D$ ), trimmed residual ( $DT$ ), standard deviation ( $SD$ ),  $p$ -value ( $t$ -test),  $p$ -value ( $u$ -test),  $p$ -value (Shapiro-Wilk-test = SW-test), 95%-confidence-interval ( $\pm CI$ ), 95%-prediction-interval ( $\pm PI$ )

SITE	r	D	SRAD [MJ/m <sup>2</sup> /day]			$\pm CI$	$\pm PI$
			SD	p (t-test)	p (SW-test)		
<i>INTERNATIONAL</i>							
Mombasa	0,89	-0,15	1,40	0,71	0,93	0,89	3,21
Nairobi	0,73	< 0,01	1,38	0,99	0,45	0,88	3,17
Mogadishu	0,69	0,65	1,43	0,15	0,98	0,91	3,28
Bogotá	0,83	-0,23	0,74	0,32	0,32	0,47	1,70
Usangi	0,92	0,80	1,26	0,05	0,53	0,80	2,88
Manaus	0,83	-0,39	1,83	0,47	0,08	1,16	4,20
<i>GABON</i>							
Mitzic	-0,31	-0,10	2,45	0,89	0,77	1,56	5,61
Libreville	-0,16	-0,21	3,86	0,85	0,17	2,45	8,85
Makokou	0,00	-0,07	3,09	0,94	0,46	1,96	7,07
Mékambo	0,50	0,08	1,44	0,85	0,42	0,91	3,29
Mouila	-0,44	0,53	2,97	0,55	0,16	1,89	6,81
Tchibanga	0,30	< 0,01	3,55	1,00	0,71	2,26	8,14
Mayumba	0,17	0,07	3,17	0,94	0,16	2,01	7,25
Moanda	-0,72	0,12	3,70	0,92	0,21	2,35	8,48
Lastoursville	-0,17	0,20	3,36	0,84	0,50	2,13	7,69

Even working with a relatively small data set of six international sites, and nine sites in Gabon only, two trends become apparent: First, generated minimum temperature and precipitation highly correlate with their according climate normals for both, sites within Gabon, and all other tropical sites. On the other hand, predicting values for tmax and srad for Gabon yield biased results or an improper seasonal trend. Tmax estimates exhibit lower correlation coefficients and broader CI and PI for Gabon than for the other African and Latin American sites (table 5.3). The same is true for predictions of solar radiation in Gabon, with high SD and broad CI and PI, and with correlation coefficients around zero, and below. We can conclude that MarkSim's difficulties to generate weather data is arise especially for sites in Gabon, and that the development of a corrective procedure for this region is appropriate, though we might question the usefulness of intrinsic climate normals, if the stochastic output does not hold on to them.

### 5.1.2 Correction of daily solar radiation and maximum temperature

The previous section has pointed out clearly the lack of accuracy when it comes to generating estimates of daily solar radiation and maximum temperature and underlines the need for a correction procedure. For the interested reader, a detailed description of a correction on a daily basis is outlined in APPENDIX C (Correction of solar radiation and maximum temperature). Here, we will go on with the evaluation of the corrected output:

#### ***Evaluating the quality of the correction***

Similar to the evaluation performed in chapter 5.1.1, now the corrected output values of solar radiation and maximum temperature are tested against the climate normals (CLX vs corrected WTG). Thus we'll put our attention on data sets from sites in Gabon only and the focus will be put only on variables that have been subject to a corrective procedure described in APPENDIX C.

Table 5.5: Evaluating corrected maximum temperature of Gabonese sites: climate normal vs. corrected simulation output (CLX vs. corrected WTG). Pearson correlation (*r*), residual (*D*), trimmed residual (*DT*), standard deviation (*SD*), *p*-value (*t*-test), *p*-value (*u*-test), *p*-value (Shapiro-Wilk-test = *SW*-test), 95%-confidence-interval (+/-*CI*), 95%-prediction-interval (+/-

SITE	CORRECTED TMAX [°C]								
	<i>r</i>	<i>D</i>	<i>D<sup>T</sup></i>	<i>SD</i>	<i>p</i> ( <i>t</i> -test)	<i>p</i> ( <i>u</i> -test)	<i>p</i> ( <i>SW</i> -test)	+/- <i>CI</i>	+/- <i>PI</i>
<i>GABON</i>									
Mitzic	0,99	0,03		0,15	0,46		0,31	0,09	0,31
Libreville	0,99	0,02		0,14	0,53		0,06	0,09	0,31
Makoko	1,00	0,03		0,15	0,48		0,23	0,09	0,33
Mekambo	0,98	< 0,01	-0,05	0,29		0,52	<b>0,03</b>	0,10	0,32
Mouila	1,00	-0,01		0,11	0,78		0,25	0,07	0,25
Tchibanga	0,99	< 0,01		0,23	0,97		0,15	0,14	0,52
Mayumba	1,00	0,03		0,17			0,07	0,11	0,39
Moanda	0,99	0,01		0,14	0,77		0,26	0,09	0,32
Lastoursville	0,99	< 0,01		0,22	0,96		0,46	0,14	0,50

Table 5.6: Evaluating corrected solar radiation of Gabonese sites: climate normal vs. corrected simulation output (CLX vs. corrected WTG. Pearson correlation (*r*), residual (*D*), trimmed residual (*DT*), standard deviation (*SD*), *p*-value (*t*-test), *p*-value (*u*-test), *p*-value (Shapiro-Wilk-test = *SW*-test), 95%-confidence-interval (+/-*CI*), 95%-prediction-interval (+/-*PI*)

SITE	CORRECTED SRAD [MJ/m <sup>2</sup> /day]								
	<i>r</i>	<i>D</i>	<i>D</i> <sup>T</sup>	<i>SD</i>	<i>p</i> (t-test)	<i>p</i> (u-test)	<i>p</i> (SW-test)	+/- <i>CI</i>	+/- <i>PI</i>
GABON									
Mitzic	0,99	0,12		0,32	0,21		0,17	0,21	0,75
Libreville	1,00	0,07		0,23	0,18		0,80	0,16	0,58
Makoko	0,99	0,12		0,30	0,19		0,08	0,19	0,69
Mekambo	0,99	0,04		0,14	0,30		0,99	0,09	0,33
Mouila	0,97	0,13		0,27	0,13		0,31	0,17	0,62
Tchibanga	0,98	0,16	0,11	0,50		0,62	<b>0,01</b>	0,27	0,88
Mayumba	0,99	0,12	0,09	0,33		0,62	<b>0,03</b>	0,20	0,67
Moanda	1,00	0,10		0,20	0,11		0,78	0,12	0,45
Lastoursville	1,00	0,14		0,29	0,11		0,53	0,18	0,66

A reduction of *SD*, narrower *CI* and *PI* as well as the convergence of *r* towards one become apparent investigating the adjusted output. The reduction of the range of the confidence and prediction interval can be connected to the fact that each simulated monthly value mimics its climate normal with higher accuracy than before the correction, which is underlined by the range of the correlation coefficients that has narrowed down and moved towards one: [0.97, 1.00] for corrected solar radiation and [0.98,1.00] for maximum temperature. All data is unbiased.

### 5.1.3 Validation of generated climate using data from three weather stations

So far we have modified solar radiation and maximum temperature in a way so that they fit to their according climate normals, while this is not necessary for minimum temperature and precipitation. Chapter 5.2 (Customizing climate for BGC-simulations ), as well as APPENDIX A tell us that climate normals can be customized by the user, and due to the correction we can be sure that these climate normals are really reflected in the stochastic output.

But until now we have only focused on MarkSim's intrinsic elements, i.e. corrected and uncorrected measures of monthly average weather data have been tested against climate normals. No statement has yet been made on whether the output makes any sense if we validate it using real climate records, which will be the topic of this chapter.

First the residuals *D* (i.e. observed – predicted values) for each available variable (*prcp*, *tmax*, *tmin*, *tdif* = *tmax*-*tmin*), using monthly means, where put together on one stack, i.e. data of one kind from all different weather stations were treated as one set of independent variables, which resulted in one set of 36 (=12 months x 3 weather stations) residuals for every variable. Performing the Shapiro-Wilk tests for normality, it turned out that sets from combined weather station data are very

unlikely to meet the normality preconditions as required to perform further statistic tests. Only prcp and tmax\_bc (bc standing for “before correction”, ac for “after correction”) residuals showed normal behavior. As a result, data was then checked for normality using several sets of 12 residuals for each weather station separately. This time most of the stets seemed to meet the normality preconditions, excluding two sets from Lastoursville that appeared to be significant at the 5% level (prcp and tmax\_ac).

First, in the manner of the previous evaluation, for each climate variable (prcp, tmin, tmax\_bc, tdif\_bc, tmax\_ac, tdif\_ac) monthly residuals were comprised to one annual residual, which was then statistically tested against the null hypothesis of being equal to zero. Further 95%-confidence and prediction intervals (CI and PI, respectively) as well an estimation of Pearson's correlation coefficient  $r$  were computed.

Since solar radiation was not measured directly for our data sets, we cannot directly determine a bias in this variable. We therefore only compute the linear correlation, and compare corrected and uncorrected versions of the output. D-values accompanied by SD, CI and PI give an idea of how strong data is biased,  $r$  can be interpreted as how well the seasonal trend has been mimicked by the weather generator. However, only both combined can give a whole picture of the quality of the generated data.

### **Precipitation**

R values are spread out between 0.9 and 1.0 which stands for MarkSim's capability of mimicking the seasonality. Lastoursville, however, is biased as it faces a significant overestimation (negative D). Here it is important to mention, that the climate record of precipitation for Lastoursville only comprises 25 years, and compared to the literature values from table 4.5 the annual amount of rainfall derived from weather stations is 461 mm/yr lower. A mean monthly overestimation of 43.1 mm (which is equal to an annual overestimation of 517 mm) corresponds to the difference between literature value and data from the weather station. Therefore the bad quality of the recorded weather data, or the rather small number of recorded years in the data base can be made responsible for the bias in annual precipitation for the site Lastoursville. The confidence and prediction intervals are of the same scale as those resulting from a comparison of generated weather data with MarkSim's climate normals.

*Table 5.7: Validation of precipitation: Pearson correlation ( $r$ ), average precipitation residual of all months (D), standard deviation (SD), p-value (t-test), p-value (Shapiro-Wilk-test = SW-test), 95%-confidence-interval (+/-CI), 95%-prediction-interval (+/-PI)*

SITE	PRCP [mm/month]						
	r	D	SD	p (t-test)	p (SW-test)	+/- CI	+/- PI
Moanda	0,98	1,1	19,4	0,85	0,18	12,3	44,4
Lastoursville	0,94	-43,1	38,5	< 0,01	0,98	24,5	88,2
Mouila	0,97	-4,4	31,5	0,63	0,35	20,0	72,2

The month-wise comparison will uncover possible biases in monthly values. As we can see in figure 5.1, all of the monthly residuals of Moanda are unbiased since the confidence intervals all include 0, the reason for the narrow annual CI. For the site Mouila, all monthly values from May to

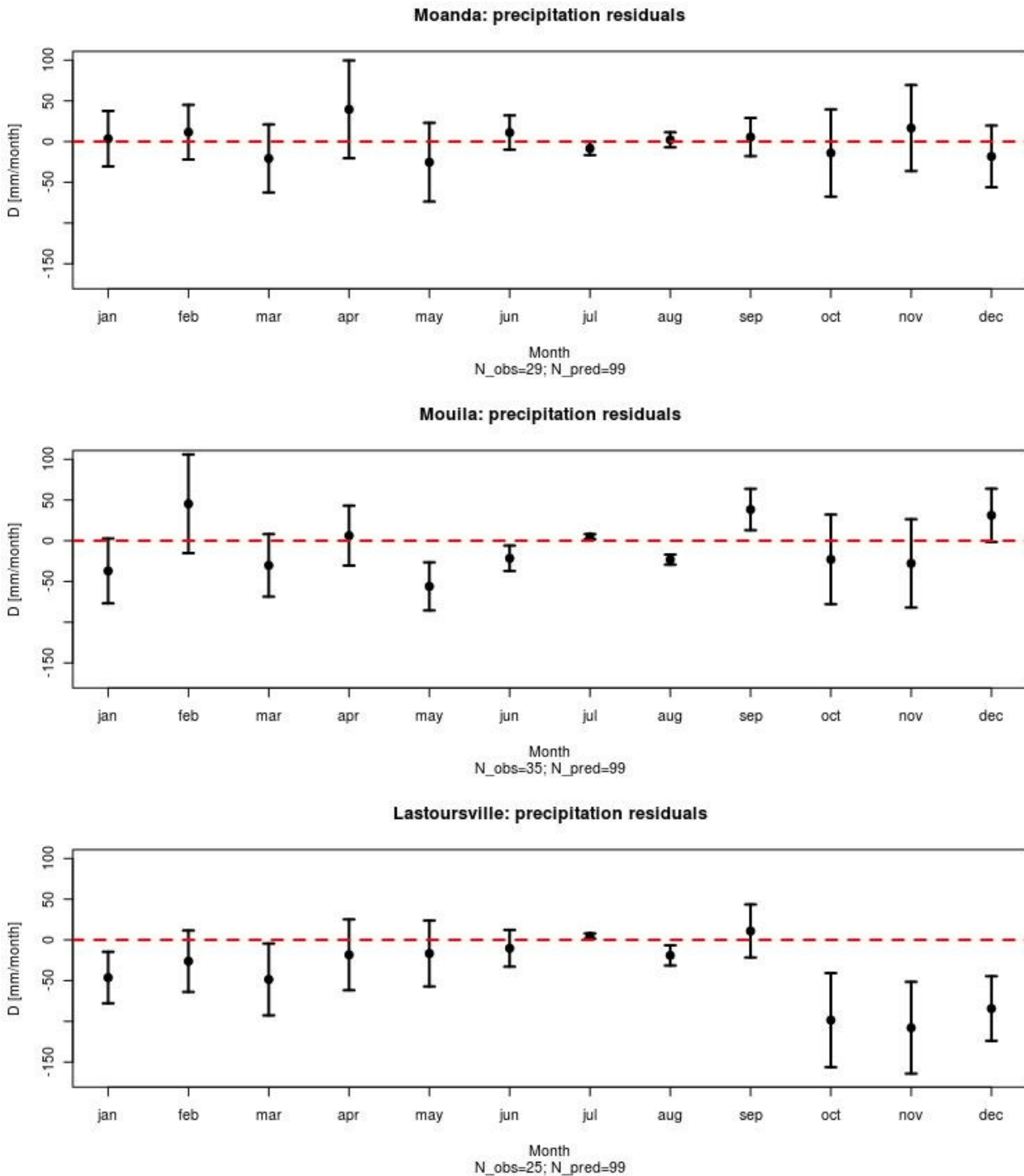


Figure 5.1: Validation of precipitation: monthly precipitation residuals [mm/month] with 95%-confidence intervals for Moanda, Mouila and Lastoursville. If the dashed line stays within the CI, the data point is unbiased.

September are biased, but the maximum absolute bias does not exceed 60mm of rainfall per month. It seems clear that biased data most frequently appears during the drier months, since little

rainfall variation leads to a narrow CI and a little over- or underestimation might hence result in a bias. Responsible for the biased annual residual of Lastoursville are most definitely not the drier months, but the wet, boreal fall and winter months October to January, with overestimation of around 100 mm/month.

### **Minimum temperature**

In the case of minimum temperature, which has not been subject of a correction procedure, CI and PI are narrow, the mean of the residual (D) is between 0 and 1, and  $r = 0.95$ . Table 5.8 suggests that  $t_{min}$  for Moanda is biased at the 5%-level of significance. As we can see, this results from its small CI and PI. We can interpret that phenomenon as follows: There is a bias in  $t_{min}$ , and since CI and PI are so narrow, the bias won't exceed  $1^{\circ}\text{C}$  in the annual mean. In Lastoursville, we detect the smallest D, but CI and PI are stretched out further than in Moanda, which can be explained by a seasonal trend which is not perfectly met by the generated data ( $r = 0.72$ ). The U-test, however, does not suggest any bias at the 5% level. Similar as for Moanda, Mouila's generated seasonality shows a high correlation to the real trend, but the data is biased at an underestimation of about  $3^{\circ}\text{C}$  (positive D), suggested by the t-test. Note that for all three sites minimum temperature is underestimated.

Table 5.8: Validation of minimum temperature: Pearson correlation ( $r$ ), annual minimum temperature residual (D), standard deviation (SD), p-value (t-test), p-value (u-test), p-value (Shapiro-Wilk-test = SW-test), 95%-confidence-interval (+/-CI), 95%-prediction-interval (+/-PI)

SITE	TMIN [ $^{\circ}\text{C}$ ]								
	r	D	D <sup>T</sup>	SD	p (t-test)	p (u-test)	p (SW-test)	+/- CI	+/- PI
Moanda	0,95	0,58		0,20	< 0,01		0,62	0,13	0,47
Lastoursville	0,72	0,35	0,57	1,00		0,05	< 0,01	0,32	1,06
Mouila	0,94	2,93		1,00	< 0,01		0,07	0,63	2,29



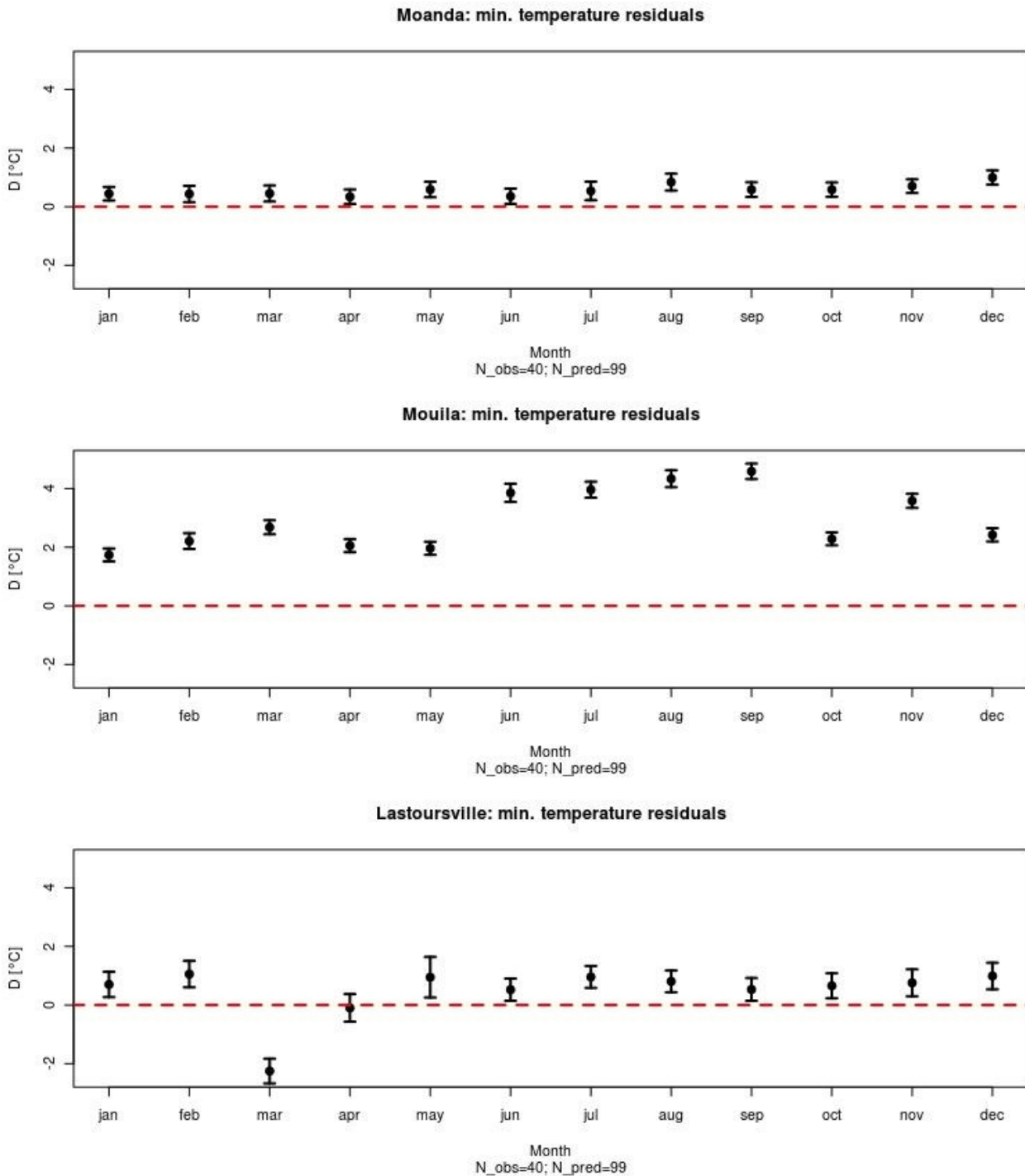


Figure 5.2: Validation of minimum temperature: monthly minimum temperature residuals [mm/month] with 95%-confidence intervals for Moanda, Mouila and Lastoursville. If the dashed line falls into the CI, the data point is unbiased. The month-wise analysis in figure 5.2 underlines what has been mentioned before in the case of site Moanda: All residuals are lined up at approximately 0.6°C, parallel to the zero-line, but none of the monthly values include zero in their CI, which means that all values can be seen as biased, but all of them at virtually the same level that does not exceed 1°C. The same is true for Lastoursville, with one exception: the minimum temperature in March, which overestimates the true value for about 2°C. Since all other months tend to underestimate the observed, this leads to an unbiased

annual residual. For the site Mouila, MarkSim fails to estimate minimum temperature, which is especially true for the drier months, June to September, with a maximum underestimation of more than 4°C in September.

### Maximum temperature

As already mentioned, a correction has been applied on daily tmax values, which of course leads to a change of its monthly mean values. One thing that we recognize immediately is that the correction narrows prediction and confidence intervals of the annual residuals, and shows an almost perfect seasonal trend, with r close to 1. D, however, does not change a lot. A contradictory result of that comparison is that after the correction annual tmax is biased in Moanda, which hasn't been the case before. We can again give the following explanation: The seasonal trend is met better by the corrected output, which means that the monthly residuals are all of the same size. This results in a narrow CI, compared to the uncorrected output where bad seasonality created a wide CI due to fluctuating residuals. Since PI and CI are narrower now, we can be sure that the bias is not above 1°C, which we couldn't be sure of before, where the CI was wider. Also in Mouila we observe the same pattern as we did in the case of tmin: a constant overestimation of the generated tmax, but still a good seasonality.

Table 5.9: Validation of maximum temperature: Pearson correlation (r), annual maximum temperature residual (D), standard deviation (SD), p-value (t-test), p-value (u-test), p-value (Shapiro-Wilk-test = SW-test), 95%-confidence-interval (+/-CI), 95%-prediction-interval (+/-PI)

SITE	TMAX [°C]								
	r	D	D <sup>T</sup>	SD	p (t-test)	p (u-test)	p (SW-test)	+/- CI	+/- PI
Moanda bc	0,54	0,16		1,00	0,59		0,18	0,63	2,28
Moanda ac	1,00	0,20		0,18	< 0,01		0,92	0,11	0,40
Lastoursville bc	0,71	-0,30		1,02	0,33		0,28	0,64	2,33
Lastoursville ac	0,93	-0,22	-0,09	0,72		0,52	< 0,01	0,18	0,60
Mouila bc	0,73	-2,92		1,36	< 0,01		0,29	0,86	3,11
Mouila ac	0,98	-2,89		0,70	< 0,01		0,10	0,45	1,61

The month-wise analysis reveals the impact of the correction procedure. In figure 5.3, which shows residuals before the correction has been applied, we observe for sites Lastoursville and Moanda that most of the estimated monthly values are biased. Residuals perform an S-shaped curve around the zero-line, which indicates that maximum temperature is overestimated in boreal winter and summer, but underestimated in spring and fall. This variation causes a broad annual CI, which leads to unbiased annual estimates, while at the same time the monthly estimates are biased. As already observed for the minimum temperature, Mouila has to be treated with special care. All months are biased, but here all tend to overestimate the observed values, with a maximum of 5°C

in July.

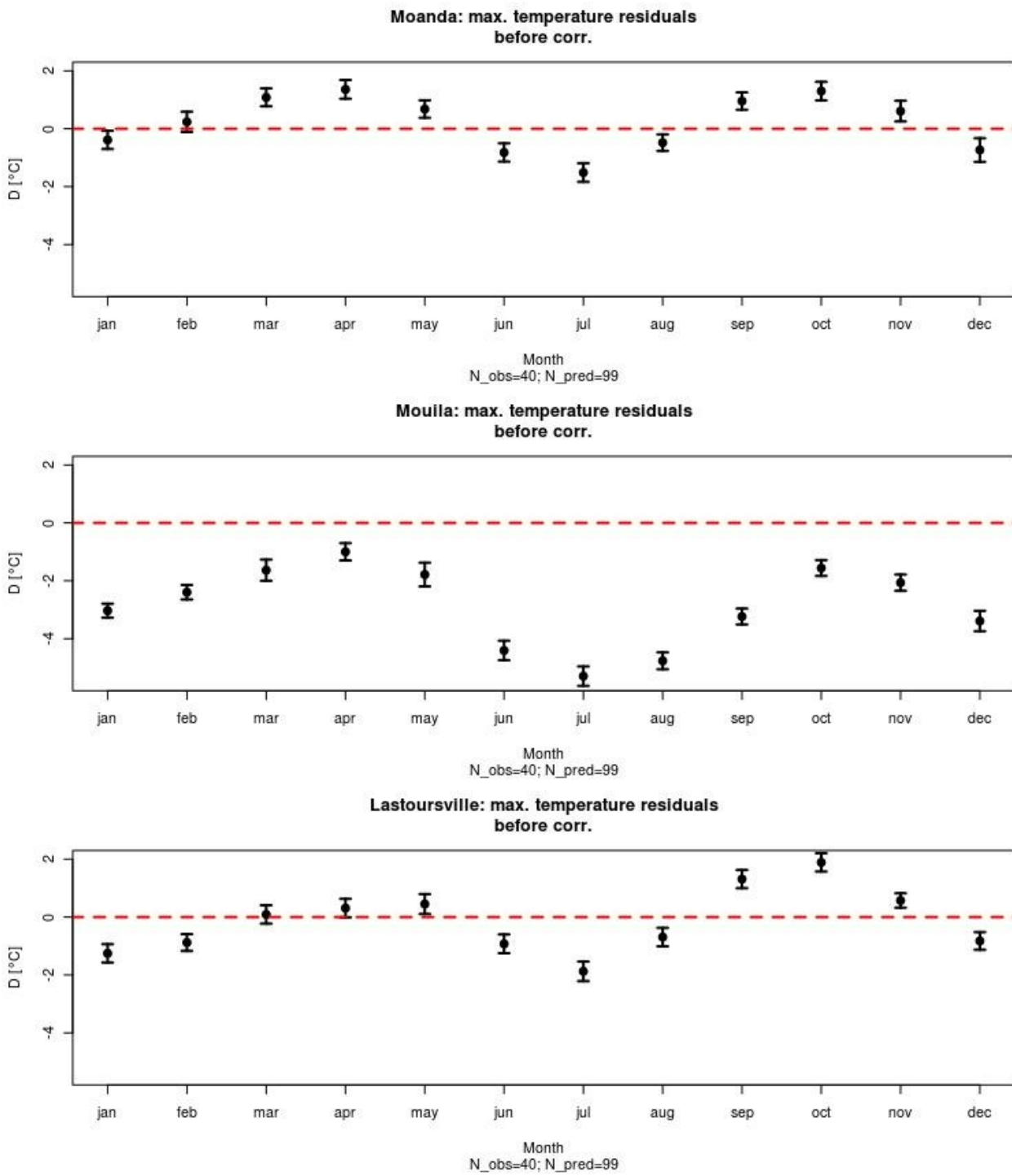


Figure 5.3: Validation of uncorrected maximum temperature: monthly maximum temperature residuals [mm/month] with 95%-confidence intervals for Moanda, Mouila and Lastoursville, before correction. If the dashed line falls into the CI, the data point is unbiased.

After the correction (figure 5.4), this bias has at least been reduced to maximum of 4°C in August. For the other two sites the outcome of the correction procedure is quite satisfactory, where most monthly values are unbiased. Nevertheless, all of the monthly estimates, except for July and August for the site Moanda tend to be lower than the real values for not more than 0.5°C, which causes a bias in the estimate of the annual mean. Note, that the pattern for corrected maximum

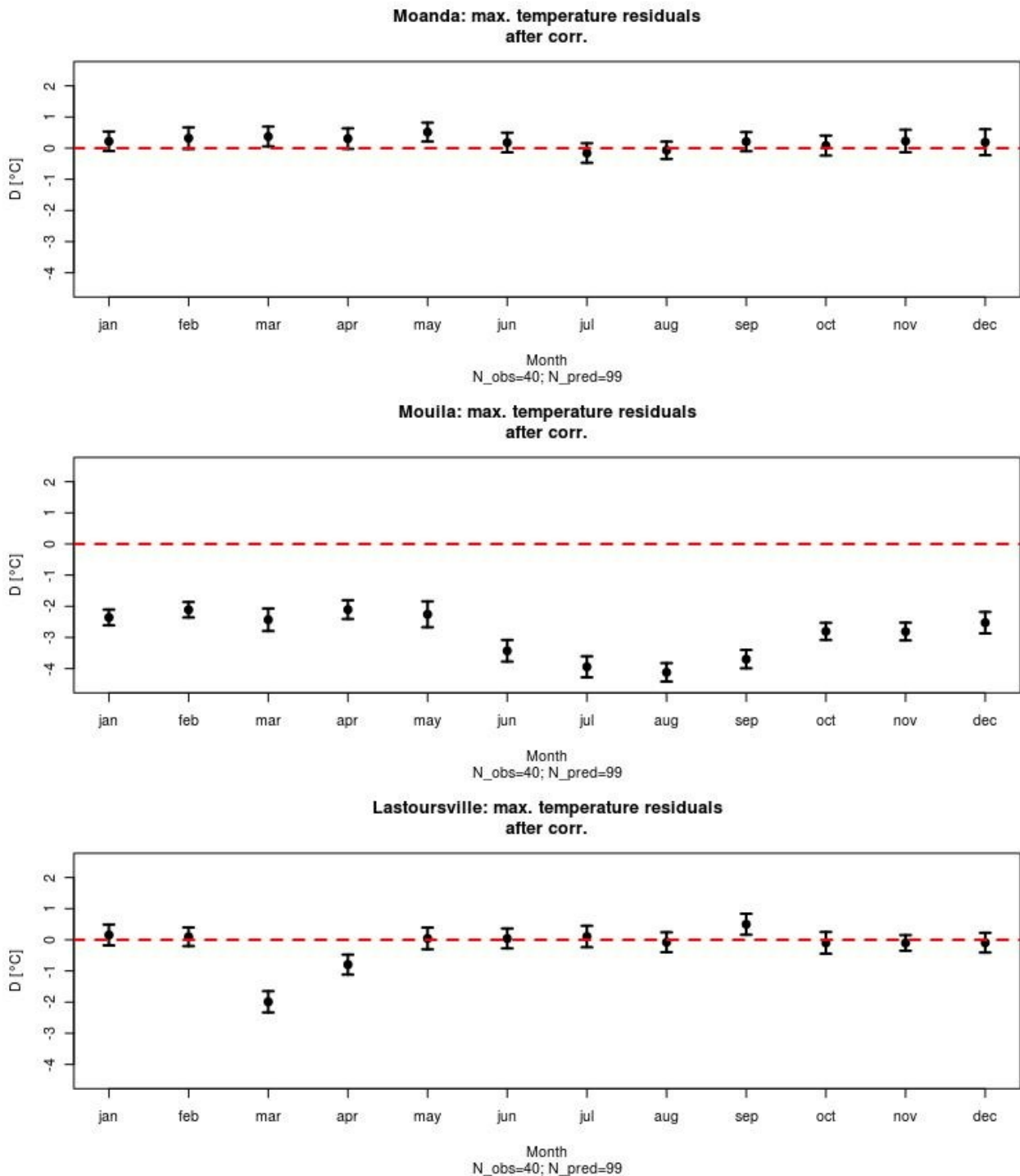


Figure 5.4: Validation of corrected maximum temperature: monthly maximum temperature residuals [mm/month] with 95%-confidence intervals for Moanda, Mouila and Lastoursville, after correction. If the dashed line falls into the CI, the data point is unbiased.

temperature and minimum temperature in Lastoursville seems to be similar, with about 2 degrees overestimation in March. This suggests an error in the climate normal for the mean temperature in March, which can be accounted for when generating climate for sites near Lastoursville. On the other hand, Mouila shows an overestimation in tmax, and an underestimation in tmin, which indicates a badly defined climate normal for temperature range which is more than 5°C higher than it should be in the annual average.

### **Temperature range (temperature difference)**

What we can observe in tdif is actually just a superposition of phenomena we have seen so far in tmax and tmin. So Mouila faces a overestimation in tdif of nearly 6°C, while Moanda and Lastoursville are better off, with an overestimation not exceeding 1°C. Similar, to what we have seen before, the correction produces a bias in the annual mean, that has not been present before the correction in Lastoursville and Moanda. The explanation is analogous: Narrow PI and CI indicate that each single monthly estimate mimics the real value with higher accuracy, but the reduction in variation of the monthly residuals results in a bias in annual tdif. But we know exactly what size of error we have to reckon with in each month. Note that the r values for all locations have been negative before a correction has been performed, which indicates a seasonal course of the generated data that is closer to the inverse than to the real trend. The correction, however, moves r for Moanda and Lastoursville closer towards one. Since Mouila is biased in tmax and tmin, we see a superposition of these effects which result in a bias in tdif, and an obviously wrong seasonal trend (r remains negative).

*Table 5.10: Validation of temperature difference (corrected and uncorrected): Pearson correlation (r), annual temperature range residual (D), standard deviation (SD), p-value (t-test), p-value (Shapiro-Wilk-test = SW-test), 95%-confidence-interval (+/-CI), 95%-prediction-interval (+/-PI)*

SITE	r	D	TDIF [°C]				+/- CI	+/- PI
			SD	p (t-test)	p (SW-test)			
Moanda bc	-0,35	-0,43	1,03	0,18	0,54	0,66	2,37	
Moanda ac	0,91	-0,38	0,28	<b>&lt; 0,01</b>	0,49	0,18	0,65	
Lastoursville bc	-0,31	-0,65	1,50	0,16	0,62	0,95	3,44	
Lastoursville ac	0,78	-0,57	0,45	<b>&lt; 0,01</b>	0,44	0,29	1,03	
Mouila bc	-0,62	-5,85	2,16	<b>&lt; 0,01</b>	0,12	1,37	4,95	
Mouila ac	-0,31	-5,81	1,66	<b>&lt; 0,01</b>	0,06	1,06	3,81	

### **Solar radiation & insolation**

Solar radiation is a special case since there were no measurements available for this work. We can only compare the seasonal course of radiation with the measured insolation (sunshine hours per month) on-hand for Moanda and Lastoursville only, and assume that they are correlated. As table 5.11 suggests the correction pushes the negative r value (Moanda) or a r value close to zero (Lastoursville), indicating no correlation at all, more towards the desired  $r=1$  mark. Unfortunately we can't make any statement on the bias of the corrected versions of the data, but at least we

*Table 5.11: Pearson correlation (r) between solar radiation [MJ/m<sup>2</sup>/day] and insolation [hrs/day] before (bc) and after correction (ac), for Moanda and Lastoursville.*

SITE	SRAD [MJ/m <sup>2</sup> /day]	r
Moanda bc		-0,68
Moanda ac		0,84
Lastoursville bc		0,10
Lastoursville ac		0,85

know that the seasonality is presented more accurately by the corrected variables under the assumption that insolation and radiation are correlated in reality.

The effort to develop a correction procedure pays off, as confirmed by an enhanced quality of the corrected output, which is reflected in higher correlation coefficients, narrower confidence intervals (in annual resolution, tables 5.7 - 5.11) and smaller number of months that are biased (figures 5.1, 5.2, 5.3 and 5.4). Biased precipitation estimates for Lastoursville can be connected to a lack of accuracy of the recorded data (literature values in table 4.5 suggest a different mean annual rainfall than the climate records bought from Lastoursville's weather station). The bias in temperature (detected in maximum and minimum temperature) for Mouila can be traced back to a wrong climate normal of temperature range, since the seasonality is of good quality, but we face over estimation of maximum temperature and underestimation of minimum temperature.

### **Graphical presentation of the seasonal trends**

Monthly mean values over the year for each variable will be shown in addition, to make statistical results more understandable. All plots are organized in the same way:

Line-1 in black represents the observed data, line-2 in red the uncorrected simulated data, and if available, line-3 in green the corrected data. Only in the case of solar radiation line-1 is uncorrected and line-2 is corrected data, since no measured values are available.

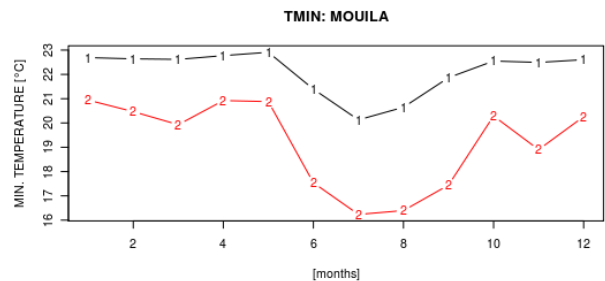
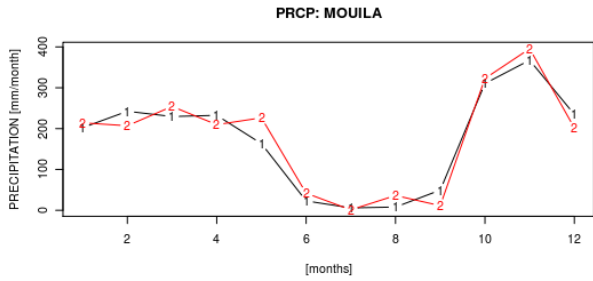
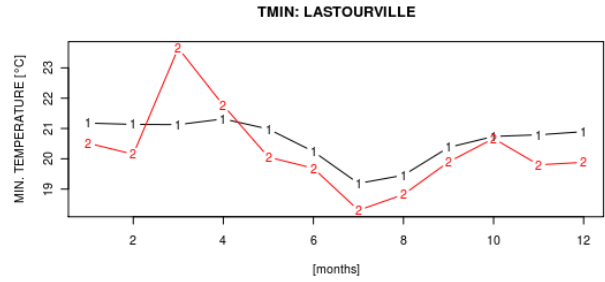
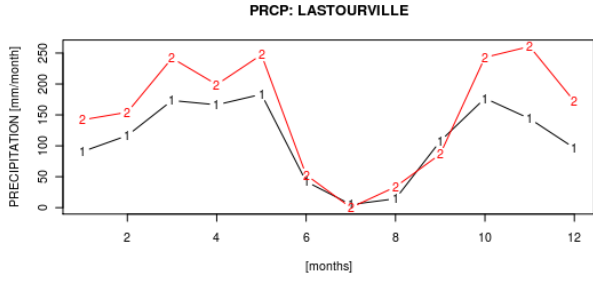
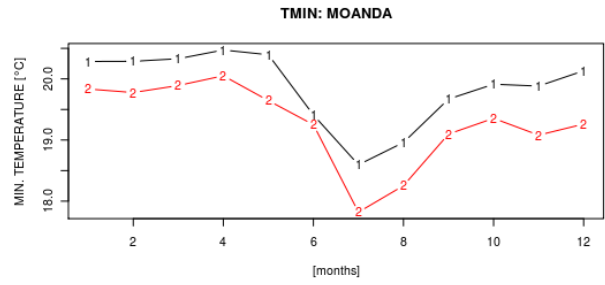
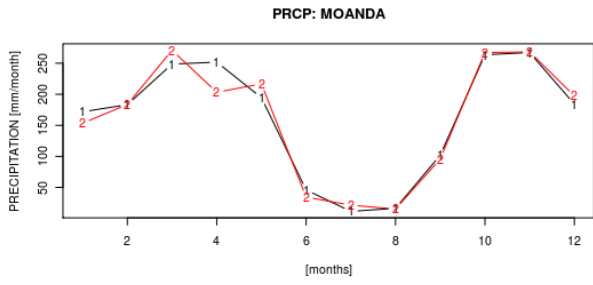


Figure 5.5: Monthly precipitation [mm/month] from weather station (black line), generated (red line)

Figure 5.6: Monthly mean tmin [°C] from weather station (black line), generated (red line)

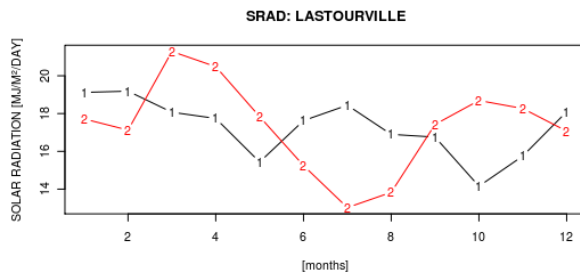
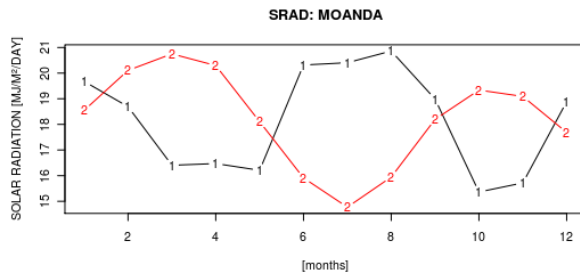


Figure 5.7: Monthly mean solar radiation [°C] uncorrected (black line), and corrected (red line)

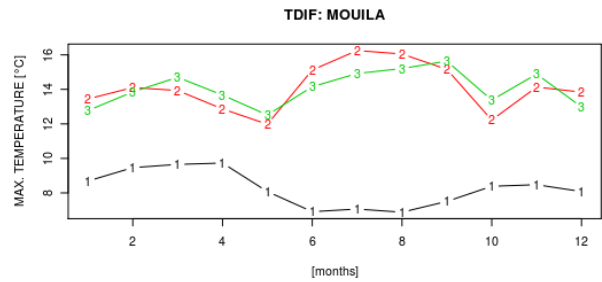
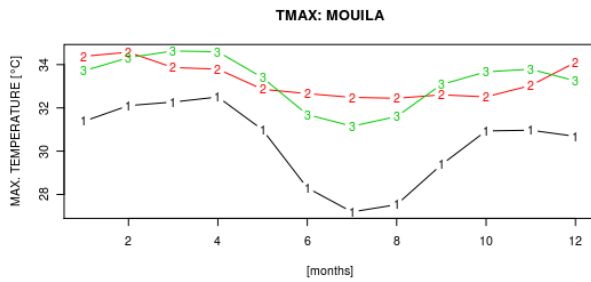
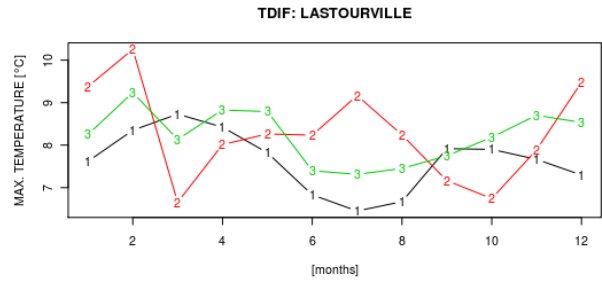
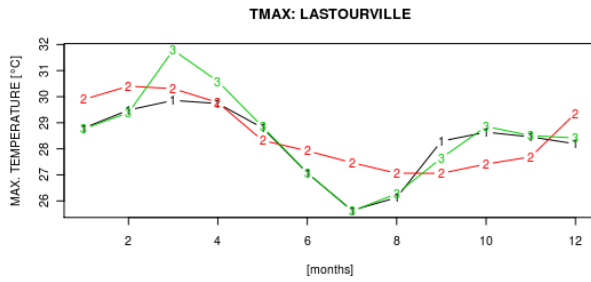
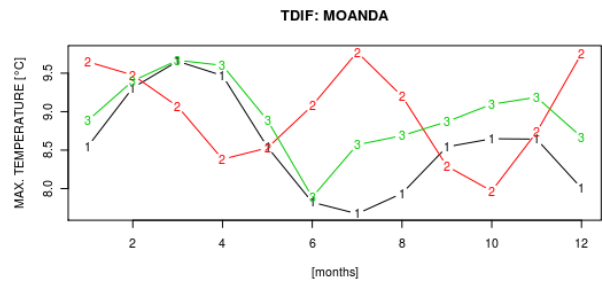
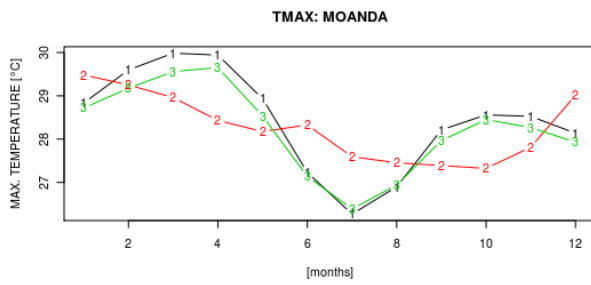


Figure 5.8: Monthly mean tmax [°C] from weather station (black line), generated uncorrected (red line), corrected (green line)

Figure 5.9: Monthly mean temperature difference [°C] from weather station (black line), generated uncorrected (red line), corrected (green line)

### 5.1.4 Evaluating the distribution of annual rainfall

The distribution of observed annual rainfall from table 4.5 (Maloba Makanga, 2010) is compared with generated precipitation for the same sites. For each quantity (min, Q1, median, Q2, max) and each site a residual (observed minus predicted) is computed, and transformed to relative values in percent of the according literature (=observed) quantity. This results in a set of 10 relative residuals for each quantity, which is tested against the null hypothesis of the sets mean residual being equal to zero, performing a t-test. The mean residual (in percent of the according median), p-value and 95%-confidence interval are presented in table 5.12. For an unknown reason, MarkSim was not able to produce weather data for Franceville, Lamabaréné and Port-Gentil.



Table 5.12: Annual precipitation from N=99 generated years for sites in Gabon. Median, Mean, minimum (min) and maximum (max) value, 25%-quartile (Q1) and 75%-quartile (Q3) of N=99 climate years, all in mm/yr.

Site	Generated annual precipitation [mm/yr]					
	median	mean	min	Q1	Q3	max
Bitam	1784	1818	1278	1576	2012	2623
Cocobeach	2995	3010	1925	2758	3232	4120
Lastoursville	1744	1761	1099	1569	1972	2498
Libreville	2920	2937	1674	2703	3192	4069
Makokou	1665	1654	810	1414	1844	2584
Mayumba	1752	1695	830	1439	1978	2572
Mitzic	1688	1672	989	1330	1916	2857
Moanda	1913	1925	1327	1741	2129	2448
Mouila	2115	2128	1586	1950	2278	2739
Tchibanga	1603	1619	932	1422	1843	2384

Table 5.13: Mean residual (observed-predicted) in percent of the according observed quantity, p-value (t-test), 95%-confidence interval also in percent of the according observed quantity. Refer to table 5.12 and table 4.5.

	D	p (t-test)	+/- CI
min	-12,4	0,18	19,3
Q1	-2,7	0,29	5,3
median	-1,0	0,58	3,9
Q3	-2,4	0,42	6,4
max	-1,2	0,81	11,1

None of the compared quantities is biased. In average MarkSim overestimates the minimum for 12.4% of the literature value for the minimum, and the maximum for only 1.2%, and we see broader confidence intervals for the extreme values (min and max), which implies that for some of the sites there is considerable over- or underestimation. The smallest CI of +/-3.9% is exhibited by the median, which is in average 1.0% higher than the literature median. Q1 and Q3 are 2.4% and 2.7% above the observed, respectively. Note that in general all quantities tend to be overestimated by MarkSim (negative residuals!).

## 5.2 Customizing climate for BGC-simulations

In the introduction of chapter 5 we have underlined the importance being able to quantitatively change and adapt meteorological parameters, so when a climate is used for an ecosystem-simulation we can separate observed effects and dedicate them to certain qualities of the underlying climate. Since fluctuations of temperature in tropical regions are small, the parameters we wish being able to adjust are: mean annual rainfall, the distribution of rainfall within the year (the length of the dry season), the year-to-year variation of annual rainfall, and also the kind of cloud cover (stratiform vs. cumuliform).

A functional implementation of MarkSim allows us to define our own climate normals (APPENDIX A - MarkSim file structure). This was originally designed to allow users to alter monthly mean values of precipitation, temperature and temperature range if better knowledge of certain climate variables than those derived from the interpolated climate surfaces (of MarkSim's database) is available. MarkSim uses the same parameters for the stochastic (Markov Chain) process but tries to achieve different monthly mean values during the simulation. We will make use of this functionality to create and customize our own climatic setups, including extremely dry and humid regimes. How this is done will be explained in this chapter, and since we cannot be sure that MarkSim produces the desired results, the generated output will be tested with respect to annual values and the distribution of annual rainfall. New methods to define the mean and standard deviation of rainfall of a climate will further be introduced.

### **5.2.1 Customizing dry season and annual rainfall**

The usefulness of “customizing” a climate might appear questionable in the first place if one is interested in an as accurate as possible approach of the real climate at a specific site, but for system analytical approaches well defined amounts of annual rainfall and the number of dry months are crucial parameters. As will be illustrated later in this work, it might be of interest to learn about the influence of the length of the dry season on the capability of an ecosystem (-model) to store carbon, or to investigate how ecosystem breakdown is connected to the amount of annual precipitation.

In order to create different climatic setups with respect to the amount of annual rainfall and the length of the dry season, rising from the same underlying parametrization (which is MarkSim intrinsic and depends on longitude, latitude and elevation) a method was designed to modify existing climate normals (monthly precipitation, mean temperature and temperature range). Note, that as for the correction procedure illustrated in APPENDIX C (Correction of solar radiation and maximum temperature), we consider climates that have their driest months in boreal summer (i.e. July and August). The aim is to keep site specific parameters, but to change the amount of annual precipitation and how it is distributed over the year (with respect to the length of the dry season). The procedure to do so is quite simple: The precipitation, temperature, temperature range and solar radiation means of dry months July and August are calculated (i.e.  $\text{mean} = (\text{aug} + \text{jul})/2$ ), and one month with these values is pasted between July and August. This shifts the monthly means of June and July to the left (towards the beginning of the year): June overwrites the old value of May, July becomes the new value June, all other values stay the same. In the next step the same is applied to the other side of the year: October is replaced by September, September by August, all other months stay the same. This procedure is repeated until the desired number of dry months is reached, with a maximum of 4 months that can be pasted in between. The original setup up is

referred to as climate with “two months of dry season”, sticking with this definition the range of dry seasons that can be produced that way lies between two and six months. Precipitation is then raised (or reduced) in all months in relative proportions to adjust the annual amount to a desired value.

The information required for this process is mainly taken from the CLX-file (APPENDIX A) that

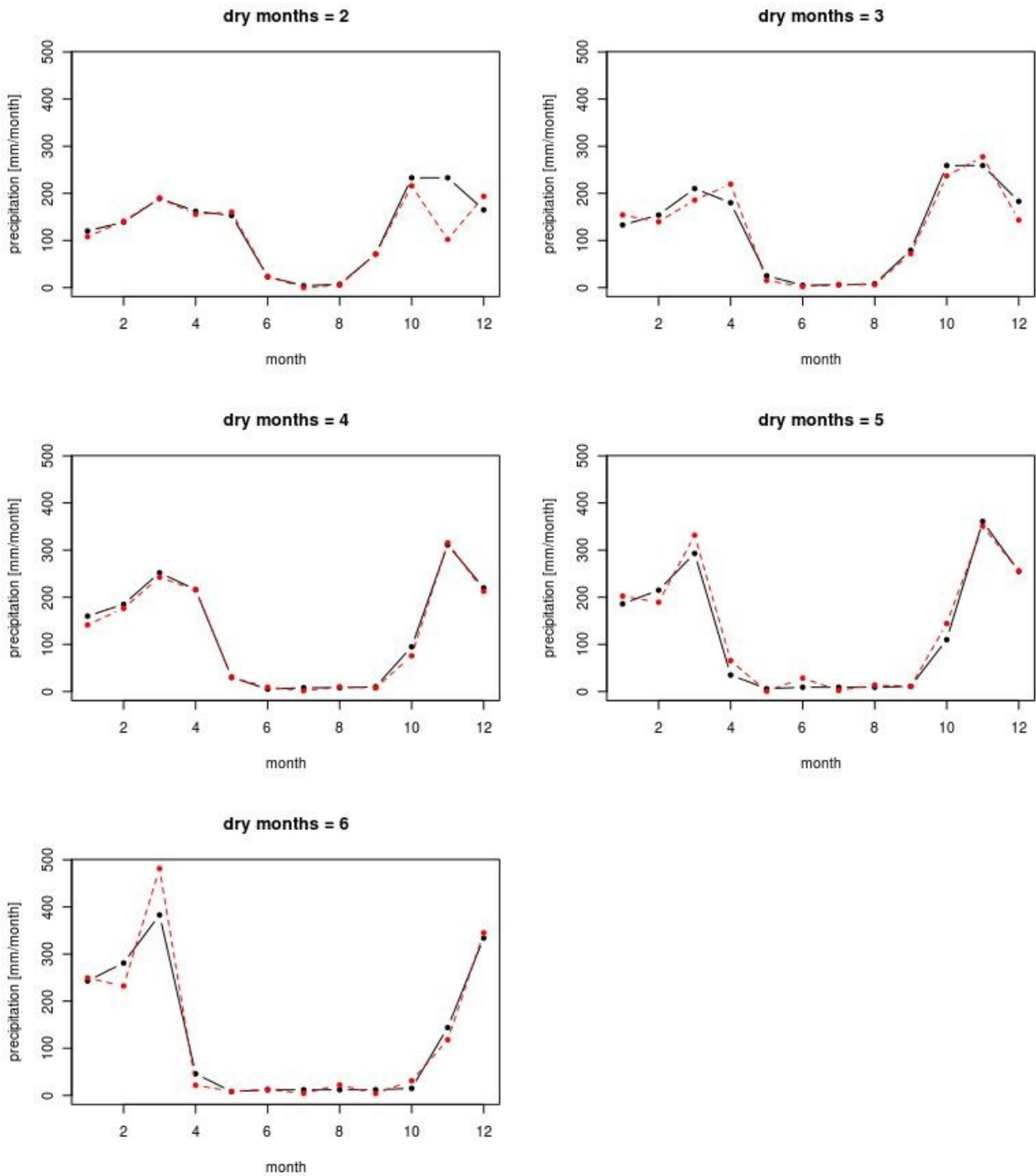


Figure 5.10: Forming a dry season: monthly precipitation [mm/month] for two to six months of dry season, site Birougou Mountains P3: 12°20' E / 1°45' S or 12.334 / -1.750, 1500 mm annual precipitation, elevation 800 masl. Climate normal (black line) and generated rainfall (dashed line).

serves as template, additionally annual rainfall and the desired number of dry months (2 to 6) have to be defined. The header of the template CLX consisting of a site identification code (maximum of 8 letters), as well as latitude, longitude and elevation (see APPENDIX A - MarkSim file structure, for closer information) is then together with the new annual distribution of precipitation, mean temperature and temperature range written into a DAT-file. This file carries the necessary information for MarkSim to generate a new CLX file, and further the daily climate. In order to perform the correction of maximum temperature and solar radiation, another CLX file is created directly that should only be used for the correction (APPENDIX C), but never to run MarkSim itself. The reason, why we do not make use of the CLX file that is generated by MarkSim for the correction procedure is that the adjusted solar radiation can not be addressed within the DAT-file, which also results in an unchanged profile of solar radiation in the original CLX file.

For an unknown reason MarkSim does not exactly hold on to the climate normals, but adds some variation. This is especially true shifting the climate towards longer dry seasons, as we can see in the example illustrated in figure 5.10.

In APPENDIX D a further method to create climate time series mimicking climate change or paleo climate is introduced.

## 5.2.2 Climate for Biome-BGC simulations

This section will focus on the climate needed for simulations performed with Biome-BGC (chapter 6.1 - Testing ecosystem stability under different climatic conditions). For our study we chose to simulate ecosystem behavior for a small region in the tropical Congo basin, the Birougou Mountains of Gabon, where a parametrization of Biome-BGC concerning eco-physiological (epc) parameters for the Western Congolian Lowland Rainforest has been performed (Gautam and Pietsch, 2011 submitted). From a variety of possible sites we chose P3 with the following coordinates at 800 m.a.s.l.

P3: 12°20' E / 1°45' S or 12.334 / -1.750

In chapter 5.1 (Validating and correcting stochastic climate for Gabon) we have seen that MarkSim faces no difficulties generating precipitation for sites in Gabon. The configurations we have compared within that section only contained “real” climates, i.e. the present climates at specific sites, based on MarkSim's intrinsic database for its climate normals. Is the same quality of the output guaranteed if we change the climate normals (especially towards low precipitation regimes, or long dry seasons)?

Annual precipitation as given in the climate normal (CLX-file) is compared with the actual value of annual rainfall after the simulation, for some records (a similar approach as in chapter 5.1). We again define the residual (D) as  $D = prcp_{clx} - prcp_{mean} = \text{climate\_normal} - \text{simulation\_mean}$ . A positive residual means underestimation (or the generated output is too small compared to the climate normal), while values below zero indicate overestimation. The analysis covers selected climates with annual precipitation ranging from 250mm/yr to 5000 mm/yr and dry seasons from 2 to 6 months. In most of the lower range records we observe overestimation of around 50 mm/yr (tables 5.14 and 5.15), an error that can be accounted for when using MarkSim, if it was the same for all climatic setups. But for unknown reason especially the 500 mm and 550 mm records are

Table 5.14: Precipitation analysis for 2DM and 3DM: annual rainfall as defined in the climate normal (clx), true median and mean, residual (D), further minimum and maximum value, 25%-quartile (Q1) and 75%-quartile (Q3) and standard deviation (SD). Relative minimum, maximum, Q1 and Q3 in percent of the median (min%, max%, Q1% and Q3%, respectively) and relative standard deviation in percent of the mean (SD%). Number of climate years within a climate (N). All values are either in mm/year or percent.

	clx	median	mean	D	min	Q1	Q3	max	SD	min%	Q1%	Q3%	max%	SD%	N
P3_800_2DM	250	231	238	12	87	191	288	461	70	38	83	125	199	29	99
	500	338	345	155	80	222	434	917	151	24	66	129	271	43	99
	550	341	365	185	89	260	453	800	150	26	76	133	235	41	99
	600	641	654	-54	347	531	768	1166	168	54	83	120	182	25	99
	650	672	680	-30	263	558	791	1177	167	39	83	118	175	24	99
	700	732	748	-48	298	620	889	1267	191	41	85	121	173	25	99
	750	780	779	-29	387	639	883	1246	200	50	82	113	160	25	99
	800	783	814	-14	358	664	954	1407	205	46	85	122	180	25	99
	850	862	862	-12	492	725	1006	1435	190	57	84	117	166	22	99
	900	1007	987	-87	506	832	1141	1527	205	50	83	113	152	20	99
	950	941	954	-4	452	821	1098	1478	213	48	87	117	157	22	99
	1000	1034	1029	-29	645	876	1179	1503	205	62	85	114	145	19	99
	1050	1082	1073	-23	564	903	1212	1683	223	52	83	112	156	20	99
	1100	1136	1164	-64	530	1002	1332	1736	242	47	88	117	153	20	99
	1150	1144	1179	-29	728	1027	1282	1925	231	64	90	112	168	19	99
	1200	1173	1186	14	630	1035	1328	1717	234	54	88	113	146	19	99
	1250	1289	1297	-47	746	1156	1450	2088	232	58	90	112	162	17	99
	1500	1356	1365	135	665	1108	1595	2193	332	49	82	118	162	24	99
	1750	1372	1382	368	566	1171	1580	2355	348	41	85	115	172	25	99
	2000	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
2500	2498	2533	-33	1744	2244	2872	3414	405	70	90	115	137	15	99	
3000	3126	3123	-123	2104	2794	3375	4193	432	67	89	108	134	13	99	
3500	3545	3512	-12	2270	3226	3849	4365	458	64	91	109	123	13	99	
5000	5190	5089	-89	2910	4503	5733	7120	834	56	87	110	137	16	99	
P3_800_3DM	250	292	315	-65	22	233	395	708	125	8	80	135	242	39	99
	500	505	531	-31	187	438	675	939	161	37	87	134	186	30	99
	550	361	388	162	113	256	467	1043	170	31	71	129	289	43	99
	600	665	674	-74	326	574	748	1192	162	49	86	112	179	24	99
	650	682	706	-56	305	593	815	1433	182	45	87	120	210	25	99
	700	719	745	-45	279	605	858	1389	203	39	84	119	193	27	99
	750	826	842	-92	292	683	993	1308	218	35	83	120	158	25	99
	800	845	854	-54	297	718	1000	1422	205	35	85	118	168	24	99
	850	895	921	-71	503	782	1036	1531	205	56	87	116	171	22	99
	900	947	955	-55	357	793	1075	1578	225	38	84	114	167	23	99
	950	994	996	-46	494	827	1165	1650	236	50	83	117	166	23	99
	1000	1029	1046	-46	630	915	1176	1830	214	61	89	114	178	20	99
	1050	1161	1143	-93	565	944	1304	1877	279	49	81	112	162	24	99
	1100	1208	1189	-89	657	1017	1326	1742	237	54	84	110	144	19	99
	1150	1181	1198	-48	583	1033	1353	1732	241	49	87	115	147	20	99
	1200	1198	1210	-10	748	1037	1370	1772	239	62	87	114	148	19	99
	1250	1204	1242	8	751	1054	1401	1937	256	62	88	116	161	20	99
	1500	1450	1459	41	1159	1327	1550	1903	166	80	92	107	131	11	99
	1750	1750	1760	-10	1336	1620	1910	2217	199	76	93	109	127	11	99
	2000	1968	1983	17	1290	1856	2146	2560	227	66	94	109	130	11	99
2500	2474	2468	32	1970	2312	2602	3024	222	80	93	105	122	8	99	
3000	3046	3029	-29	1916	2797	3313	4094	445	63	92	109	134	14	99	
3500	3414	3456	44	2255	3234	3719	4496	412	66	95	109	132	11	99	
5000	5014	5024	-24	3937	4787	5292	6151	430	79	95	106	123	8	99	

generated with more than 100 mm underestimation. We have tried to resolve that problem by changing the simulation seed in the weather generator, which unfortunately yielded the same underestimated output. In the higher precipitation range we faced problems while generating climate, where MarkSim stopped the simulation with an error concerning a gamma shape parameter. This resulted in either a reduced number of climate years for a given climate normal, or no output years at all. Noteworthy, a record close to the true climate at P3, namely the setup with 2000 mm rainfall and unchanged number of dry months couldn't be generated.

The fact that MarkSim does not actually produce an output which fits the desired input (as soon as we make use of customized climate normals), and that it is not possible to generate climate years at certain precipitation ranges at all, couples with a second effect: the heterogeneity in rainfall distributions on an annual scale. We have tested climate records with respect to the distribution of all climate years (maximum:  $N = 99$ ) that together form one climate. How are the annual values of precipitation of the climate years distributed, what are their maximum and minimum values and how far are they scattered within one climatic configuration?

Table 5.15: Precipitation analysis for 4DM, 5DM and 6DM: for a description of the listed quantities refer to table 5.14

	clx	median	mean	D	min	Q1	Q3	max	SD	min%	Q1%	Q3%	max%	SD%	N
P3_800_4DM	250	266	304	-54	55	216	382	701	128	21	81	144	263	42	99
	500	494	515	-15	218	400	622	877	149	44	81	126	178	28	99
	550	593	600	-50	210	501	708	990	164	35	85	119	167	27	99
	600	622	606	-6	276	479	688	1086	159	44	77	111	175	26	99
	650	690	693	-43	338	558	822	1127	185	49	81	119	163	26	99
	700	728	755	-55	393	635	858	1266	175	54	87	118	174	23	99
	750	795	798	-48	413	669	922	1223	184	52	84	116	154	23	99
	800	869	855	-55	412	673	997	1345	230	47	77	115	155	26	99
	850	897	925	-75	337	802	1051	1427	207	38	89	117	159	22	99
	900	987	975	-75	490	819	1100	1574	215	50	83	111	159	22	99
	950	964	983	-33	515	790	1174	1659	240	53	82	122	172	24	99
	1000	1058	1058	-58	594	896	1204	1624	226	56	85	114	153	21	99
	1050	1078	1108	-58	625	970	1242	1558	207	58	90	115	145	18	99
	1100	1109	1127	-27	655	998	1251	1706	213	59	90	113	154	18	99
	1150	1215	1175	-25	562	1027	1320	1577	206	46	85	109	130	17	99
	1200	1170	1185	15	760	1076	1283	1688	171	65	92	110	144	14	99
	1250	1196	1207	43	697	1106	1335	1511	154	58	92	112	126	12	99
	1500	1452	1439	61	1095	1298	1559	1933	183	75	89	107	133	12	99
	1750	1683	1682	68	1080	1552	1819	2386	213	64	92	108	142	12	99
	2000	1944	1956	44	1384	1744	2109	2656	261	71	90	108	137	13	99
2500	1933	1937	563	966	1743	2102	2656	284	50	90	109	137	14	6	
3000	1996	2088	912	1384	1793	2288	3755	432	69	90	115	188	20	20	
3500	3470	3475	25	2437	3236	3696	4573	367	70	93	107	132	10	99	
5000	4934	4936	64	3562	4699	5227	5848	435	72	95	106	119	8	99	
P3_800_5DM	250	309	313	-63	96	236	388	545	108	31	77	126	176	34	99
	500	558	556	-56	156	451	643	911	139	28	81	115	163	25	99
	550	596	608	-58	246	498	721	918	146	41	83	121	154	24	99
	600	656	646	-46	188	558	744	1044	149	29	85	113	159	23	99
	650	740	728	-78	311	627	831	1101	147	42	85	112	149	20	99
	700	743	734	-34	369	641	832	979	124	50	86	112	132	16	99
	750	783	810	-60	305	560	1054	1627	304	39	72	135	208	37	99
	800	860	891	-91	199	704	1111	1757	315	23	82	129	204	35	99
	850	942	951	-101	364	751	1127	1584	253	39	80	120	168	26	99
	900	932	930	-30	260	737	1104	1796	266	28	79	118	193	28	99
	950	978	982	-32	337	798	1132	1592	252	34	82	116	163	25	99
	1000	1127	1101	-101	530	857	1281	1948	280	47	76	114	173	25	99
	1050	1151	1111	-61	292	844	1388	2116	366	25	73	121	184	32	99
	1100	1164	1165	-65	495	948	1375	1825	286	43	81	118	157	24	99
	1150	1258	1258	-108	656	1014	1470	2000	314	52	81	117	159	25	99
	1200	1188	1227	-27	323	989	1458	2065	335	27	83	123	174	27	99
	1250	1258	1299	-49	451	1095	1549	2190	339	36	87	123	174	26	99
	1500	1593	1596	-96	938	1416	1795	2366	249	59	89	113	149	15	99
	1750	1805	1816	-66	1228	1621	1973	2474	265	68	90	109	137	14	99
	2000	2096	2078	-78	1408	1906	2233	2742	259	67	91	107	131	12	99
2500	2093	2068	432	1408	1874	2232	3396	304	67	90	107	162	14	22	
3000	2096	2134	866	1408	1883	2259	4354	436	67	90	108	208	20	8	
3500	2189	2453	1047	1408	1933	2621	5085	802	64	88	120	232	32	24	
5000	5140	5134	-134	3987	4751	5448	6223	532	78	92	106	121	10	99	
P3_800_6DM	250	265	270	-20	38	192	322	628	114	14	72	121	237	42	99
	500	394	396	104	98	280	508	820	149	25	71	129	208	37	99
	550	467	465	85	121	348	566	1007	161	26	75	121	216	34	99
	600	625	621	-21	232	546	732	1088	160	37	87	117	174	25	99
	650	646	666	-16	312	561	776	1035	144	48	87	120	160	21	99
	700	681	695	5	288	574	800	1258	175	42	84	117	185	25	99
	750	713	717	33	213	608	812	1172	172	30	85	114	164	24	99
	800	790	792	8	432	673	881	1215	173	55	85	112	154	21	99
	850	852	824	26	485	686	954	1272	186	57	81	112	149	22	99
	900	841	853	47	490	758	955	1302	167	58	90	114	155	19	99
	950	993	997	-47	706	894	1092	1401	127	71	90	110	141	12	99
	1000	1048	1038	-38	739	928	1118	1544	145	70	89	107	147	14	99
	1050	1106	1106	-56	685	992	1219	1658	175	62	90	110	150	15	99
	1100	1143	1144	-44	791	1031	1233	1589	172	69	90	108	139	15	99
	1150	1194	1196	-46	817	1094	1297	1584	153	68	92	109	133	12	99
	1200	1256	1234	-34	860	1136	1341	1697	160	69	90	107	135	12	99
	1250	1283	1277	-27	830	1144	1402	1908	174	65	89	109	149	13	99
	1500	1522	1532	-32	1054	1404	1670	1944	187	69	92	110	128	12	99
	1750	1938	1908	-158	1181	1686	2129	2636	311	61	87	110	136	16	99
	2000	2114	2064	-64	1093	1818	2334	2812	338	52	86	110	133	16	99
2500	2486	2529	-29	1732	2229	2844	3524	409	70	90	114	142	16	99	
3000	2872	2914	86	2058	2636	3216	3823	377	72	92	112	133	12	99	
3500	3070	3245	255	1668	2713	3678	6749	892	54	88	120	220	27	99	
5000	5048	5049	-49	3948	4613	5355	6944	500	78	91	106	138	9	99	

It seems obvious that large fluctuations of annual rainfall within a climate might be a trigger for ecosystem collapse. Equally delicate are the precipitation minima within a climate, as they might be far away from the initially desired mean value, and thereby create a stressful environment when they pop up by chance during a simulation. The problem that comes along with the use of MarkSim, besides that the desired value of annual rainfall is more often over- or underestimated than exact, is that the distribution climate years with respect to precipitation differs strongly from climate to climate.

In table 5.15, P3\_800\_5DM we see that numerous climates contain years with extremely low precipitation (columns min and min%). 800mm (clx) has a minimum of 23% of the median (860mm) which is 199mm, 1050mm (clx) at 25%, and 1200mm (clx) at 27%. The problem is not that we do have these extremes, but that they appear in some of the records, and miss in others. For our work it is crucial to have comparable distributions, maxima and minima in order to perform analytical work using Biome-BGC. As we can see from table 4.5, recorded minima for sites all over Gabon are not below 41% of the median (Mayumba), which is by far not the case for the artificial climates in table 5.14 and 5.15, for 5DM configurations in particular. Especially interesting are the sites Lastourville and Mouila with 48% and 60% minima respectively, and 136% maxima, since they are geographically closest to P3. A summary of mean relative minimum, maximum, Q1 and Q3 values in percent of the median of all generated climates as well as for the two closest sites is presented in table 5.16.

*Table 5.16: Mean values of relative minimum, maximum, Q1 and Q3 in percent of the according median and standard deviations, from table 5.14 & 5.15 and table 4.5 (Lastourville and Mouila)*

P3 (generated):	Mouila:	Lastourville:
min = (52 +/- 16)%	min = 60%	min = 48%
Q1 = (86 +/- 6)%	Q1 = 86%	Q1 = 88%
Q3 = (115 +/- 8)%	Q3 = 108%	Q3 = 110%
max = (163 +/- 32)%	max = 136%	max = 136%

As we can see, minimum and maximum, as well as Q1 and Q3-quartile of the literature values are at least within the error range (standard deviation) of the generated climate. The problem that we see is the large error of these values, which indicate differences in the underlying rainfall distributions. Hence climates are not comparable, as they might cause qualitatively diverging model behavior, caused by different characteristics of the extreme values.

We can conclude that we cannot be sure that generated climate does exhibit a mean precipitation value close to the desired value. Further we don't know how the underlying rainfall distribution of



the climate looks like, with respect to minimum and maximum, which as result might lead to wrong predictions or ambiguous observations using these climates for other software applications, such as Biome-BGC.

Thus our strategy to select climate years for certain simulation purposes has to be changed. Instead of labeling the output of one MarkSim simulation as “one climate”, a whole series of climate records using different simulation seeds and climate normals (from 50mm/yr-6500mm/yr) will be generated and put together in a pool of climate years to actively select from. We thereby level out the effect of under- or overestimation of the mean annual rainfall, we can define the distribution of

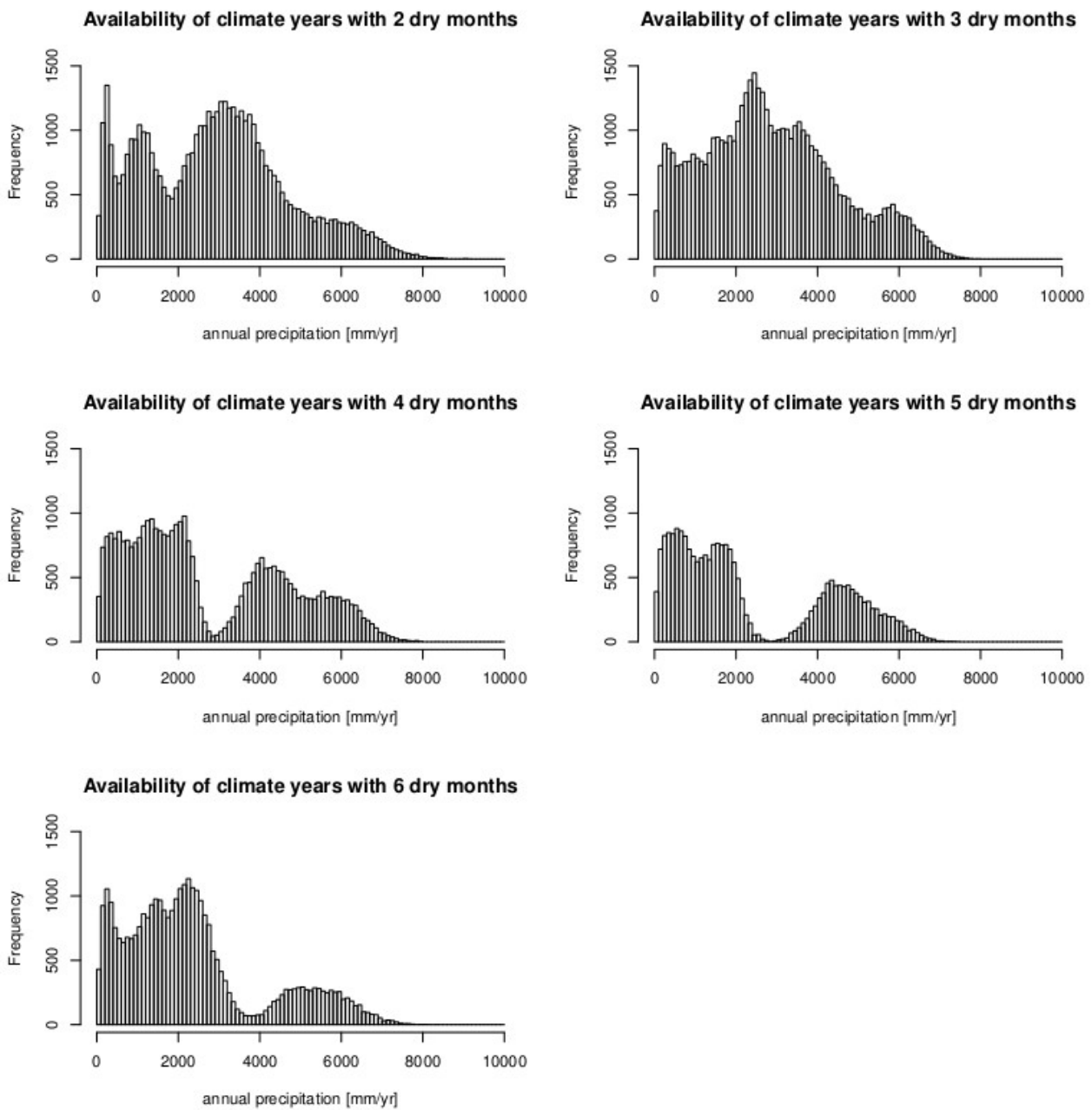


Figure 5.11: Histograms proving information about the amount of available climate year per 100mm of rainfall for 2 to 6 DM.

climate years, the size of the minimum and maximum, as well as the number of records that together form one climate. Table 4.4 consistently suggests that annual rainfall is distributed normally, with a standard deviation of around 25% (using the literature value of mean annual rainfall for Lastoursville to calculate the relative standard deviation). A procedure we use to generate normally distributed climate (with respect to annual rainfall) with defined standard deviation and mean can be described as follows: First we generate  $N=100$  random numbers to form a Gaussian distribution with a desired standard deviation and mean. Next a Shapiro-Wilk test is performed on these data and only highly significant distributions with  $p > 0.9$  are accepted. Further we only allow distributions with a mean diverging for a maximum of 5% from the desired mean. Extreme values exceeding two times the standard deviation are cut off and set to this limit (statistically nearly 96% of all random numbers should fall into the interval spanned by  $\pm 2$  times the standard deviation). Finally we pick climate records out of the pool with annual rainfall closest to the random numbers generated before, and make sure that each climate year can only appear once within each climate, though it is allowed to be present within another climates.

With this approach we can be sure that our data follows a well-defined distribution ( $p > 0.9$ ), we also know the extreme values ( $\pm 2 \cdot SD$ ) and that the mean is close to the desired mean ( $\pm 5\%$ ), but still climates are selected stochastically to some extent.

Figure 5.11 shows that the highest availability of climate data is given for 2 and 3 dry months. MarkSim stopped generating weather for regions around 3000-4000mm/yr and longer dry seasons 4-6DM), which is the reason for obvious shortages illustrated by the histograms. Even in areas with lower availability the data should be enough to create accurate distributions of rainfall.

### 5.2.3 Climatic parameters and varieties for BGC-simulations

The quantified climatic parameters on which the investigations in chapter 6.1 are based are:

- 1) Mean annual precipitation: 50mm/yr to 4500mm/yr in steps of 50mm/yr which results in 90 varieties.
- 2) Standard deviation of annual precipitation distribution: 10%, 15%, 20%, 25%, 27.5%, 30%, 32.5%, 35%, 37.5%, 40% of the mean respectively, resulting in 10 varieties.
- 3) Number of dry months (distribution of rainfall within the year): 2 to 6 resulting in 5 options.
- 4) Quality of the cloud cover: cumuliform vs. stratiform resulting in 2 options.

All in all this results in  $90 * 10 * 5 * 2 = 9000$  different climates each comprising 100 different years.

## 6 Results

### **6.1 Testing ecosystem stability under different climatic conditions**

Performing BGC simulations there is a chance that the virtual ecosystem does not develop into a stationary state, or first develop and then after some hundred or thousand years break down. In detail, a break down is characterized by a leave carbon pool carrying so little carbon, that growth in the next season is inhibited, which immediately results in zero productivity. Leave C is set to zero once passing a certain threshold. Anyway, two identical simulations, i.e. same underlying climate and parametrization, with the only exception that they use different seeds, which determine the sequence of climate years used for the simulation, can result in a qualitative difference in system state, with refer to break down. This behavior suggests weather sensitivity of the ecosystem, i.e. diverging short term fluctuations within the same climate can at one point lead to break down, and at another point lead to full development of a more ore less stable system.

In this chapter we want to detect stable and unstable regions and determine what triggers could cause a shift in stability. In the first place we expect mean annual rainfall to have major influence on whether a simulation is successfully carried out, or collapses. A rule of thumb states that below 2000 mm of precipitation per year tropical rain forests can not develop, and that the existence of such forests in Gabon at even lower amounts of rainfall is made possible only by the presence of the stratiform cloud cover. This cloud cover might therefore be another parameter causing shifts in the stability as it directly influences incident solar radiation and temperature, and thereby has an impact on evaporation and transpiration. On the other hand, as we see in table 4.7 and table 4.9 annual mean temperatures do not exhibit strong fluctuations, and wont be subject of our investigation. Ngomanda et al., 2009 and Maley, 2001 connect massive forest break down in the past (~2500 BP) to a change in the repartition of rainfall within the year, linked to an elongation of the long dry season. Finally, since climate change is often connected to growing climatic fluctuations, it might be of interest to understand what influence the year to year variation of rainfall has on the stability of the simulated ecosystem. Summarizing, we want to investigate dependence of system break down on

- 1) the amount of annual precipitation
- 2) the distribution of precipitation within the year (the length of the dry season)
- 3) the year-to-year variation of annual precipitation.
- 4) the influence of the stratiform cloud cover during the dry season (corrected versus uncorrected climate)

### 6.1.1 Methodical approach

Since simulations with the same setup but a different sequence of the same (!) climate years can result in a different final system state a statistical approach is required for this investigation, i.e. to test whether a certain climatic setup yields a stable simulation, multiple simulations based on the same climate have to be carried out in order to sum up over the dichotomous state variables (“dead” or “alive”) and receive a probabilistic image. Note, that the only difference in repeated simulations is the way single climate years are shuffled together, determined by the seed of the random number generator selecting climate years. We classify the simulation output in two groups, “dead” or “alive”, without paying attention to the actual system's productivity. Each climate group characterized by (see chapter 5.2.3 - Climatic parameters and varieties for BGC-simulations)

- 1) its mean annual rainfall (50mm to 4500mm in steps of 50mm)
- 2) the length of the dry season (2 to 6 months)
- 3) the width of the distribution of annual precipitation (As explained in chapter 5.2.2, Climate for Biome-BGC simulations, we generate normally distributed climates, with respect to rainfall, each consisting of 100 climate years. The distribution is truncated at twice the standard deviation, and the standard deviations we used for our investigation are 10%, 15%, 20%, 25%, 27.5%, 30%, 32.5%, 35%, 37.5% and 40% of the mean.)
- 4) whether or not daily maximum temperature and global radiation have been corrected or not (influence of the cloud cover).

Each group is formed by results of 20 similar simulations with different seeds. This binary approach allows us to define some sort of state probability, that tells us how stable a certain system is with refer to the applied climate. If, for instance 18 out of 20 systems collapse, the systems probability to fully develop is 10%.

How can this binary classification method be implemented in Biome-BGC? The problem that one faces right away is the the lack of clarity when it comes to the definition of the state (“dead”, or “alive”). It is quite clear that “alive” means positive productivity, and “dead” refers to zero productivity, or an empty leave carbon pool. But usually, an ecosystem is productive in the first place and accumulates some carbon and nutrients during its ongoing simulation and breaks down after that. It might even develop until reaching its upper limit of productivity, or in other words become saturated, and then collapse. Is this system “dead” or “alive”? Or at what point is the decision made whether it is “dead” or “alive”? Since any ecosystem might collapse if you just wait long enough, the classification has to be applied in combination with a certain simulation period, after which the measurement is done.

### 6.1.2 Two concepts to approach a stable state

As a representative time scale we chose the spinup interval. BGC is initiated with a small amount of leave carbon in the order of  $10^{-2}$ , still large enough to enable future growth, and converges towards a dynamical equilibrium. "Dynamical" means that fluctuations caused by weather and mortality are of course possible and desired, but they don't show trends. In other words: the system is stationary. Equilibrium is reached when soil carbon, the slowest changing pool, has reduced its fluctuations to a minimum. The time span that has elapsed between BGC initiation and the point where soil C becomes stable is referred to as *spinup* interval, which is not constant but usually around 5000 years. So state probability in this context can be interpreted as the chance that an ecosystem develops into its stationary state. Along with this approach come two conceptual errors: First, considering two systems that both "die" in 100% of all cases, but one in average reaches a higher level of development than the other one are treated as equal with refer to state probability. This error is tiny, keeping in mind that one system that develops further within the *spinup* interval also has a higher chance that one of 20 simulations will actually reach a stable state, and thereby change the state probability. It can therefore be neglected as it only affects the 0% to 1% interval of the probability (e.g. a system with real probability of 0.8% might be mistaken for a system with 0%). The second error that comes hand in hand with this approach is of the same nature, but on the other side of the probability scale: It concerns ecosystems that are almost perfectly stable. A system that has reached a stable state in terms of satisfying the requirements to finish the *spinup* can still be disturbed and break down by a "bad" combination of certain climate years. Considering a mean *spinup* time of  $5 \cdot 10^3$  years and 20 simulations per climate, set together this results in a simulated time of hundred thousand years. What about a system, that collapses only after 150.000 years? It is addressed as perfectly stable by our classification. This error, however, cannot be eliminated since it is in the nature of experiments to work with finite time series. Of course, the length of the time series, i.e. the number of simulations, could be raised easily, but this would result in enormous amounts of data, and computing capacity needed during simulation procedure and analysis. We therefore accept that systems surviving a period of 100.000 years are referred to as "100% stable".

In the preceding illustration concerning error estimation for the spinup approach, we have implicitly assumed that 20 consecutive *spinup* simulations, all of a duration of around  $5 \cdot 10^3$  years can be regarded as one single time series of hundred thousand years. In fact, it is quite obvious that non-stationary systems, or systems in development, might react to weather fluctuations in a different way than stable state systems (Yet, we don't want to make a guess whether stationary systems or systems in development are more vulnerable). To set forth this idea, starting with a low carbon ecosystem increases or decreases the probability of break down compared to initiating the system directly into a stationary state and then applying a difficult climate. At this point we have to distinguish between two issues, that only diverge in a little detail: *Do we want to learn what is*

going to happen when already stable forest ecosystems face an unfavorable change in climate, or do we want to put a statement on whether or not a forest is able to grow (from scratch) once a different climate has actually become present? We have so far only discussed the conceptual approach to define a probability of growth from scratch (which can also be understood as regrowth after a disturbance), which will be referred to as “spinup-approach”. Of course the other issue is of equal interest but it requires a slightly different systematic approach: The simulation will first be initialized with a stable state, and then the climate will be changed. Again the time it takes to reduce the fluctuations in soil carbon to a minimum will be the period of measurement. Simulations of this kind will be addressed as “spindown-approach”. Differences in the stability exhibited during these two approaches can be seen as a hysteresis effect.

### 6.1.3 Spinup simulations

We have already explained that if a simulation performs a successful *spinup*, this single simulation contributes with a “one” (=“alive”). On the other hand, a marker for break down is when the leave C pool becomes (exactly) zero. Systems that have broken down in terms of that classification contribute with a “zero” (=“dead”) to the state probability of their group (formed by 20 similar simulations of the same climate). Since we assume systems based on dryer climates, i.e. with lower mean annual precipitation to break down more frequently than those with more humid

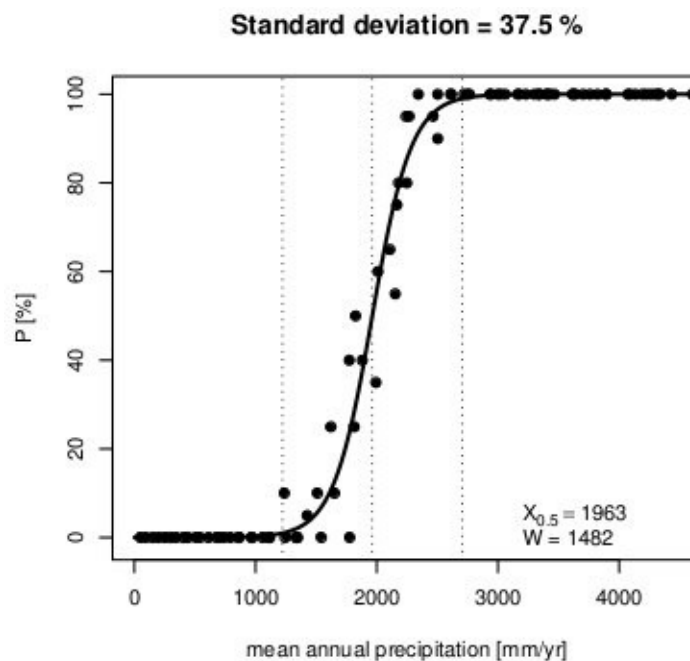


Figure 6.1: Exemplary transition from 100% collapse (left side) to 100% development of stationary system (right side) based on spinup simulations with a 3DM, uncorrected climate (with respect to solar radiation and maximum temperature), and 37.5% standard deviation of the annual rainfall distribution. P is the chance (in %) to develop a stationary state,  $X_{0.5}$  is the point of inflection, and W the width of the transition phase (=risky phase).

climates, we expect to see a transition from 0 (all simulations dead) to one (all simulations alive), with increasing precipitation. It seems appropriate to apply a logistic regression (refer to APPENDIX B - Logistic regression analysis) which is designed to describe transitions between binary states and therefore is a tool to quantitatively compare transitions resulting from different set ups.

A measure derived from the logistic regression that marks the point at which this transition occurs is the point of inflection ( $X_{0.5}$ ) at  $P=0.5$  (=50%). This point of inflection will be taken as a surrogate for the instability of system based on a certain climate (a point of inflection at a higher level of precipitation indicates a less stable state). Further the width ( $W$ ) of the transition phase is defined as the distance between the precipitation value at  $P=0.01$  (=1%) and  $P=0.99$  (=99%) and gives an idea of the size of the risky phase, as it includes all points where  $P$  is either not exactly one or zero, with a few exceptions.  $X_{0.5}$  and  $W$  will be used for further analysis, where presentations like figure 6.1 wont be illustrated.

In order to draw conclusions on the connection between the amount of biomass and stability of the ecosystem, of all successful simulations, total carbon content, leaf carbon and dead stem carbon content are recorded and averaged over the last mortality cycle (where carbon fluctuations should be at a minimum). Leaf carbon and dead stem carbon form the pool of above ground living biomass.

### ***Uncorrected climate (cumuliform)***

First the results from simulations based on climate that has not been subject of a correction procedure will be presented. As explained before, these climates exhibit tendentially higher temperatures and larger amounts of incident solar radiation during the dry months, but are generally cooler in wet months, where solar radiation is also on a lower level. The influence of the standard deviation (SD) and the number of dry months (DM) on the position of the point of inflection ( $X_{0.5}$ ) and the width of the transition phase ( $W$ ) is shown in figures 6.2 to 6.4.

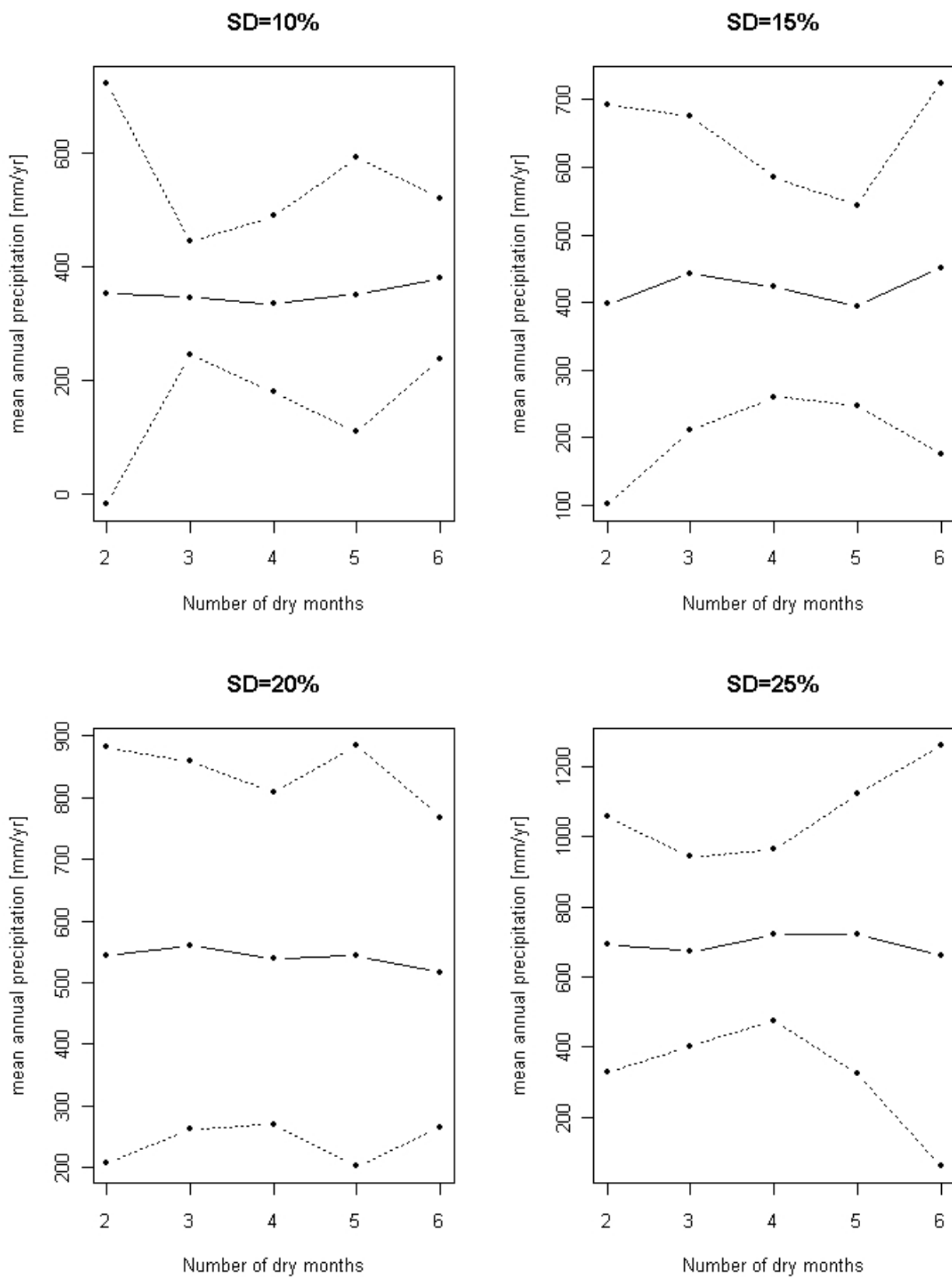


Figure 6.2: Point of inflection (solid line), 1% and 99% probability boundaries in mm rainfall per year (upper and lower dashed line, respectively) derived from logistic regression as marker for stability, based on simulations with uncorrected climate. The underlying rainfall distributions ranges from SD=10% to SD=25%.



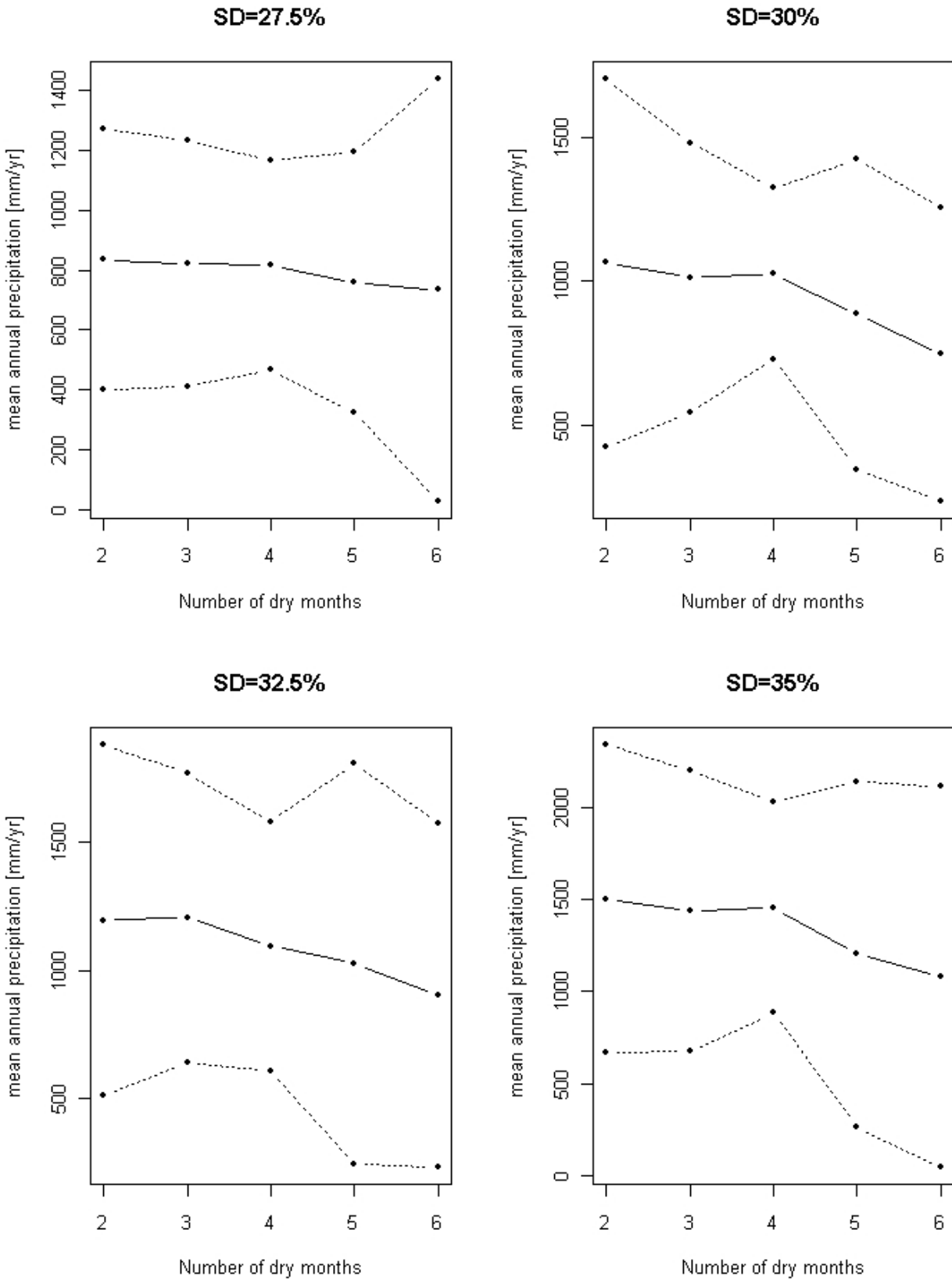


Figure 6.3: Point of inflection (solid line), 1% and 99% probability boundaries in mm rainfall per year (upper and lower dashed line, respectively) derived from logistic regression as marker for stability, based on simulations with uncorrected climate. The underlying rainfall distributions ranges from SD=27.5% to SD=35%.

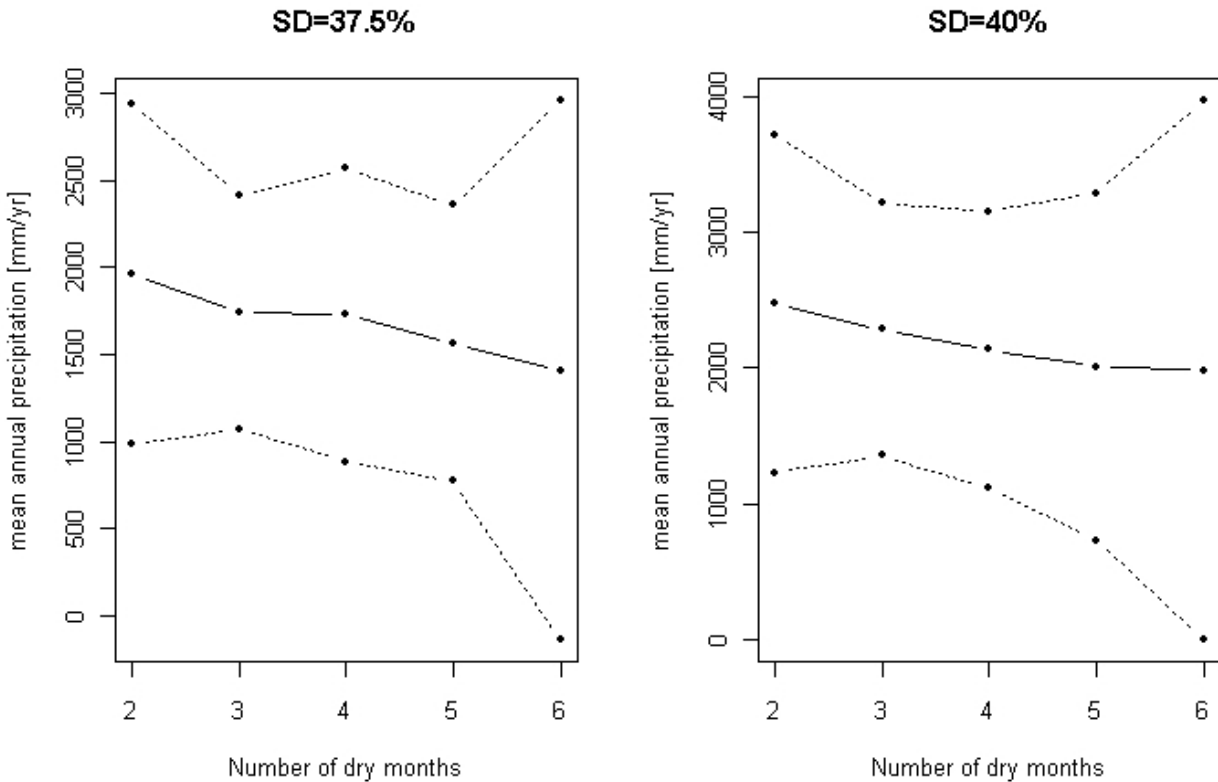


Figure 6.4: Point of inflection (solid line), 1% and 99% probability boundaries in mm rainfall per year (upper and lower dashed line, respectively) derived from logistic regression as marker for stability, based on simulations with uncorrected climate. The underlying rainfall distributions ranges from SD=37.5% to SD=40%.

An important insight from illustrations 6.19-6.21 is, that the year-to-year variation of rainfall (i.e. the standard deviation of the precipitation distribution) seems to have the largest impact on where the transition from “dead” to “alive” occurs. While the border between “more likely to break down” and “more likely to fully develop” indicated by the solid line can be drawn below 400mm/yr for SD=10%, we see the same border between 2500 and 2000 mm/yr at a variation of SD=40%. Obviously, a change like that could not be induced by the variation of the number of dry months, though we are able to see an influence caused by the length of the dry season. For SD=10%, climates with 6 dry months (as defined in chapter 5.2.1 - Customizing dry season and annual rainfall) yield the unstablist simulations, though there is little difference between simulations based on different number of dry months. But this trend seems to be gradually inverted, as variation increases, and simulations with 6 dry months become the stablest. As we have seen, standard deviations of around 25% seem to be a realistic estimate (table 4.4) for sites around P3, and for SD=25% there is also only little influence of the dry season, as the (solid) line that connects the points of inflection does not show a clear trend.

As explained before, the region enclosed by the dashed lines (W) can be interpreted as a “risky

phase”, where the ecosystem is between the completely stable state, and the state where any development is impossible. We get the impression that this phase of transition starts relatively wide with 2 dry months, narrows down towards 4 months and gets wider again approaching 6 months of dry season. It is, however, hard to conclude whether longer dry seasons do have a positive impact on system stability or not, including the length of the transition phase in our thoughts. Even if 6DM-simulations seem more stable for higher variations, we must not forget, that the transition phase is the broadest here. Since the influence of the length of the dry season remains unclear to some extent, the focus will mainly be put on the impact of the variation. In figure 6.5 all points of inflection for a given SD of the precipitation distribution are put together, neglecting the influence of the dry season. As we can observe, the amount of rainfall describing the point of inflection increases exponentially with SD!

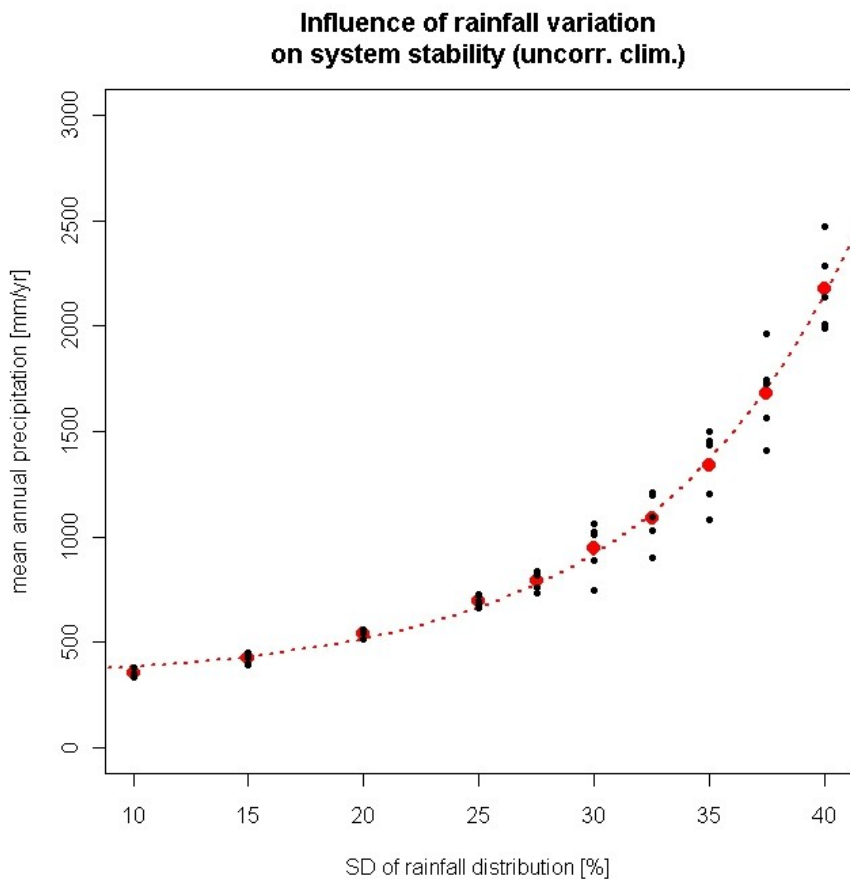


Figure 6.5: Exponential growth of point of inflection (black points) in mm/yr with increasing year-to-year variation of rainfall (SD of the precipitation distribution). Larger red points indicate the mean of black points, the dashed red line is derived from an exponential regression.

An exponential regression based on a scaled Levenberg-Marquardt algorithm (Levenberg, 1944; Marquardt, 1963) was used to fit the mean values (red points) of the points of inflection (black

points) in figure 6.5. The form of the exponential function is

$$y = y_0 + A \cdot \exp\left(\frac{x}{t}\right) \quad \text{eq. 6.1}$$

where  $y$  is the estimated mean point of inflection,  $y_0$  is the offset,  $A$  the amplitude and  $x$  corresponds to the standard deviation  $SD$ .  $r^2$  (square of the linear correlation coefficient as introduced in APPENDIX B “Testing predicted vs. observed”) and RMSE (root mean square error, definition analogous to standard deviation) have been calculated using all (black) points. Results are presented in table 6.1.

Table 6.1: Results from exponential regression of mean points of inflection derived from simulations with uncorrected climate.

$A = (21 \pm 5) \text{ mm/yr}$	$r^2 = 0.96$
$t = (9.0 \pm 0.5) [\%]^{-1}$	adj. $r^2 = 0.96$
$y_0 = (320 \pm 30) \text{ mm/yr}$	RMSE = 174 mm/yr

### **Corrected climate (stratiform)**

Before we allow a comparison between corrected and uncorrected climate with respect to stability, similar illustrations as featured in the previous section will be presented for simulations based on a climate that has undergone a correction procedure. These (corrected) climates are closer to the real climatic situation in Gabon, as we learn from Chapter 5.1.3 (Validation of generated climate using data from three weather stations ). Reduced global solar radiation and daily maximum temperature try to mimic the presence of a stratiform cloud layer during the dry season.

While simulations based on climate with little fluctuations of  $SD=10\%$  show their inflection point at 400mm/yr, and an instable phase (marked by the dashed lines) ending at about 800mm/yr for 6 dry months, simulations with  $SD=40\%$  exhibit a transition slightly below 3000mm/yr, accompanied by a transition phase reaching even 5000mm/yr. A change in variation of the underlying climate from 10% to 40% (which implies that the extreme values of these distributions should be close to 20% or 80% of the mean, due to the truncation performed before → chapter 5.2.2 - Climate for Biome-BGC simulations) causes a shift in stability of about 750%! It seems that, against our expectations, the uncorrected climate yields simulations with higher stability. Similarly we observe an inversion of the trend as also present for the uncorrected climate, where 6 dry months can be connected to the unstablest simulations for  $SD=10\%$  and with increasing variation give rise to the stablest simulations. On the other hand, the width of the transition phase for 6DM is generally larger than for other configurations, (seemingly the narrowest phase occurs at 4DM) which leads to the effect

that the unstable phase begins at similar amounts of annual rainfall, and sometimes at even higher amounts than for shorter dry seasons. Later in this chapter we will see, that this inversion of the trend can be connected to the amount of carbon stored in living biomass.

The same methods as introduced in the previous section to fit the exponential growth of instability with increasing variation have been applied in figure 6.8, results of the regression are presented in table 6.2.

*Table 6.2: Results from exponential regression of mean points of inflection derived from simulations with corrected climate.*

$A = (20 \pm 3) \text{ mm/yr}$	$R^2 = 0.98$
$t = (8.5 \pm 0.3) [\%]^{-1}$	adj. $R^2 = 0.98$
$y_0 = (340 \pm 25) \text{ mm/yr}$	RMSE = 100 mm/yr

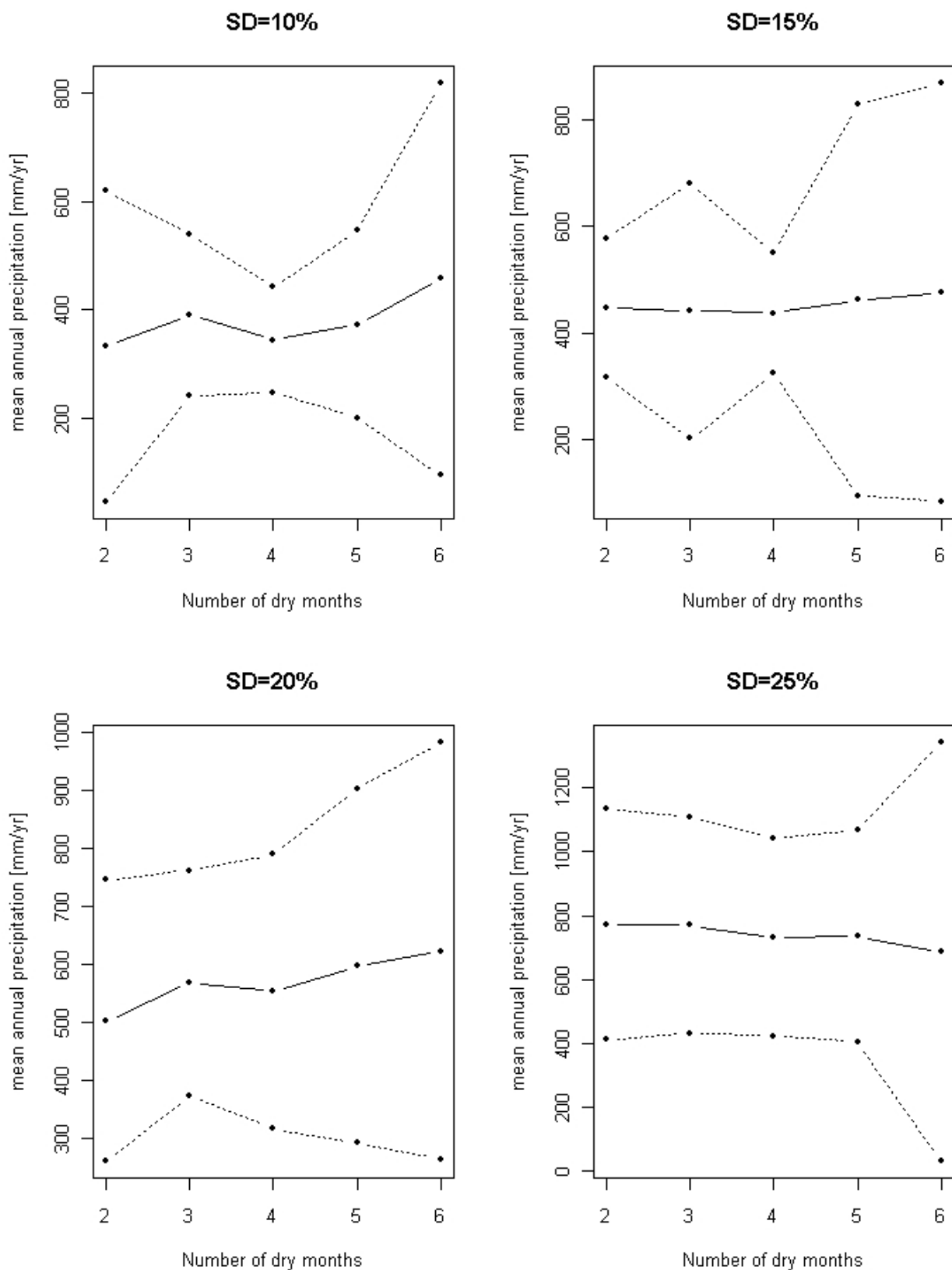


Figure 6.6: Point of inflection (solid line), 1% and 99% probability boundaries in mm rainfall per year (upper and lower dashed line, respectively) derived from logistic regression as marker for stability, based on simulations with corrected climate. The underlying rainfall distributions ranges from SD=10% to SD=25%.

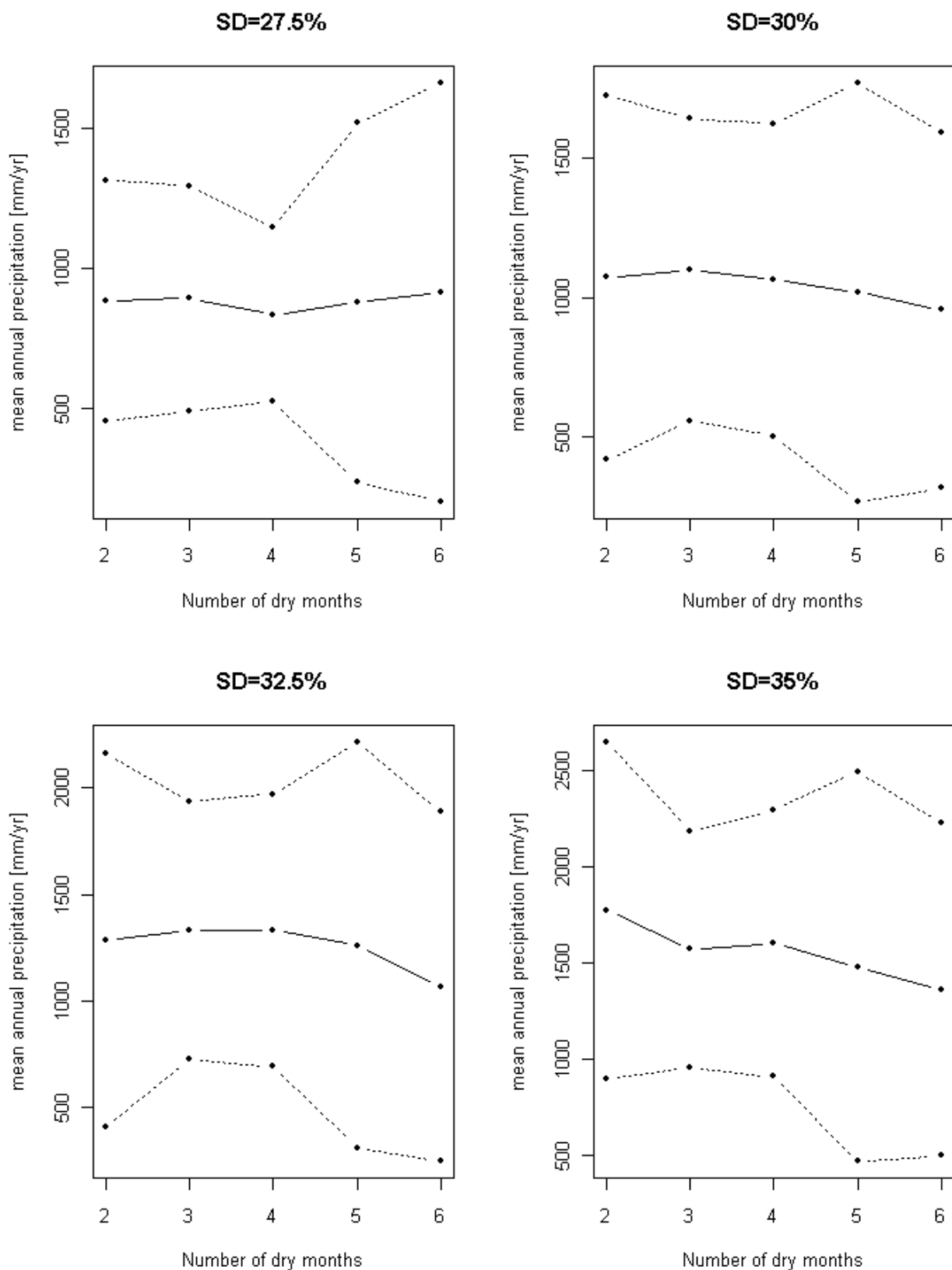


Figure 6.7: Point of inflection (solid line), 1% and 99% probability boundaries in mm rainfall per year (upper and lower dashed line, respectively) derived from logistic regression as marker for stability, based on simulations with corrected climate. The underlying rainfall distributions ranges from SD=27.5% to SD=35%.

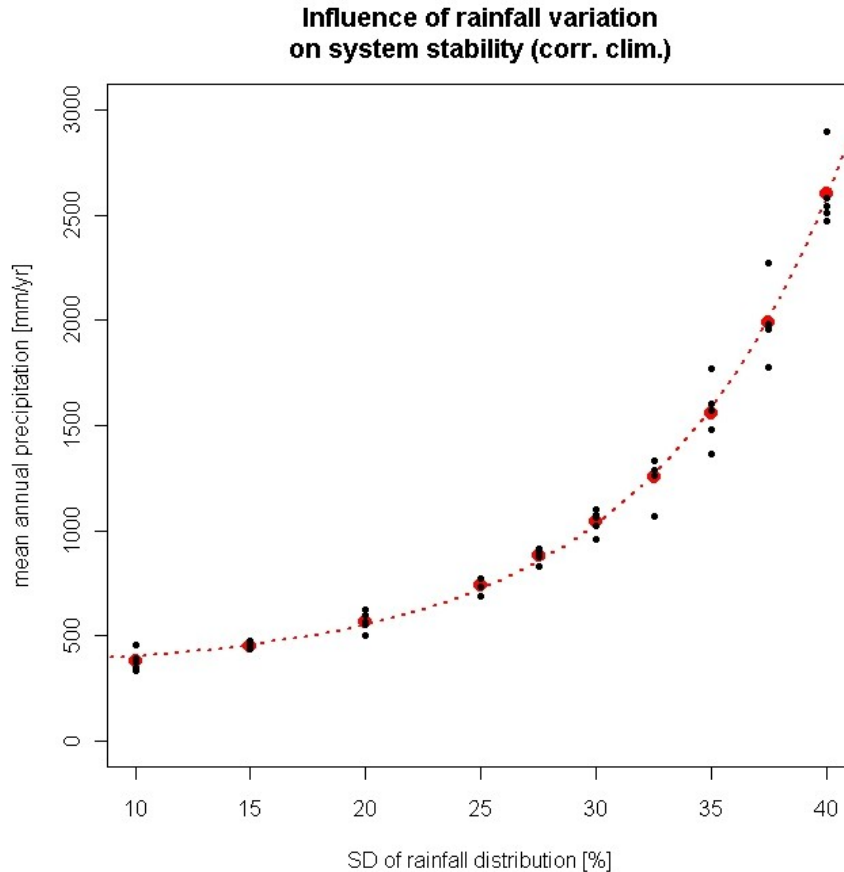


Figure 6.8: Exponential growth of point of inflection (black points) in mm/yr with increasing year-to-year variation of rainfall (SD of the precipitation distribution). Larger red points indicate the mean of black points, the dashed red line is derived from an exponential regression.

### ***Corrected and uncorrected compared***

Both, the corrected and the uncorrected variants yield exponential regression estimates that highly correlate with their underlying data ( $R^2 = 0.98$  and  $R^2 = 0.96$ , respectively). Note, that for a theoretical SD of 0%, which means that all consecutive climate years have virtually the same amount of annual rainfall, the regression predicts  $X_{0.5} = (341 \pm 35)$  mm/yr (uncorr.) and  $X_{0.5} = (360 \pm 28)$  mm/yr (corr.), which cover the same range regarding the errors. This might be interpreted as a natural threshold for water required to enable growth, neglecting all other parameters, such as variation of rainfall, and also the influence of the dry season (which causes only little difference in stability in the lower range of SD). We have, however, seen that simulations could successfully



develop a stationary state with annual rainfall down to about 100 mm/yr, but in this case we used one repeated climate year only – which reduces not only year-to-year variation of the total amount of rainfall to zero, but also the variation of the repartition of dry days versus wet days during the year. So the threshold of around 350mm/yr might be a realistic, though theoretical, estimate.

The data set based on corrected climate has been tested against the uncorrected data performing a paired t-test (table 6.3), to determine whether there is a significant difference in stability. One data-set corresponds to one value of SD and therefore comprises 5 data points (= 2 dry months to 6 dry months) from the corrected, and 5 points from the uncorrected version. The difference between  $X_{0.5}(\text{corrected})$  and  $X_{0.5}(\text{uncorrected})$  is defined as  $D = X_{0.5}(\text{corrected}) - X_{0.5}(\text{uncorrected})$ , and remains positive for all values of SD. This suggests (neglecting the impact of the width of the transition phase) that simulations with underlying corrected climate are less stable than those based on the uncorrected climate. Note that a set of D has to be distributed normally in order to perform the t-test, which is the reason why a Shapiro-Wilk test (APPENDIX B - Test for normality) has been performed in advance. Only one set (SD=35%) exhibits non-normal behavior at a level of significance of 5%. By truncating the data (as suggested by Rauscher, 1986) the number of data points would have been reduced to 3, which would reduce the validity of the t-test. We therefore apply the test with the full data set, but have to keep in mind that discrepancies could arise from the non-normal nature of the data.

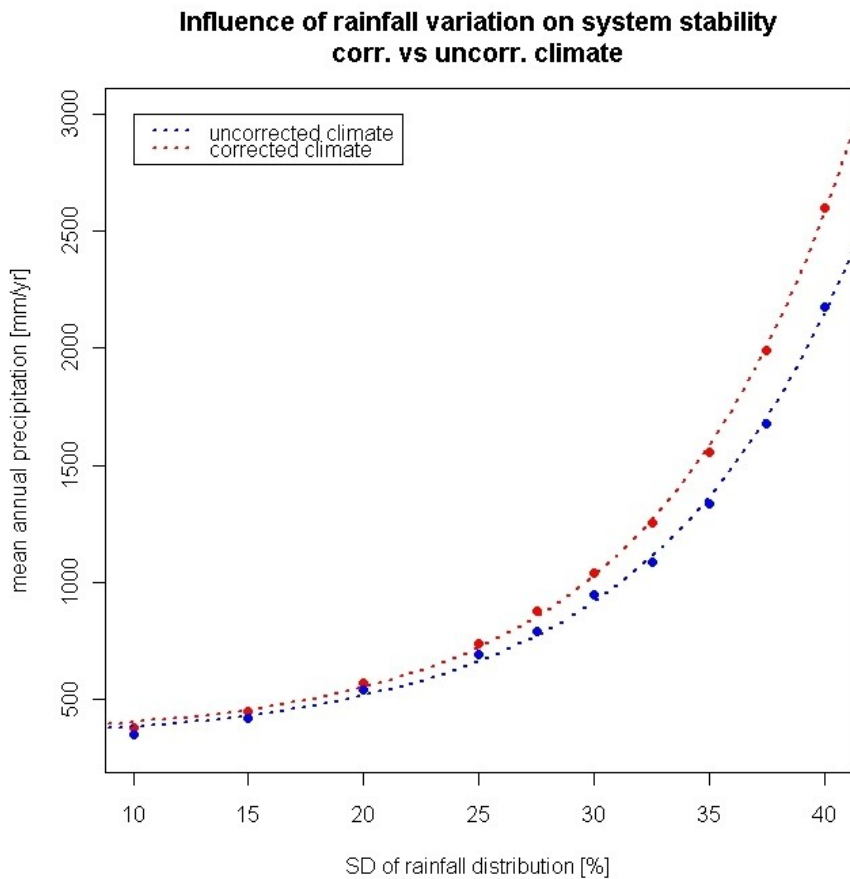


Figure 6.9: Exponential growth of mean point of inflection corrected (red symbols) versus uncorrected (blue symbols) in mm/yr with increasing year-to-year variation of rainfall (SD of the precipitation distribution). The dashed lines present the exponential regression.

Even though  $X_{0.5}(\text{uncorrected})$  lies always below  $X_{0.5}(\text{corrected})$  (compare the blue and red line in figure 6.9), the t-test suggests that this difference becomes significant at  $SD=27.5\%$  for the first time, and remains significant for variations larger than  $SD=30.0\%$  at a level of significance of 5%. This implies, that for a realistic variation of  $SD=25\%$  there is no significant difference between corrected and uncorrected, or in other words, whether the sky in the long dry season is covered or remains clear does not have an effect on the precipitation threshold where the (simulated) ecosystem becomes unstable (Note that this is only valid for a system that is able to grow “from scratch” under these conditions, what we can not be sure of in the case of the real ecosystems, which have always been confronted with changing climate)!

Table 6.3: Corrected (stratiform) vs uncorrected (cumuliform) estimates of instability: SD = standard deviation of the rainfall distribution [%], p-value (Shapiro-Wilk-test), p-value (t-test), D =  $X0.5(\text{corr.}) - X0.5(\text{uncorr.})$ , confidence interval (CI). One data-set corresponds to one value of SD and comprises 5 data points (= 2 dry months to 6 dry months) from the corrected, and 5 points from the uncorrected version.

Corrected vs. Uncorrected				
SD[%]	p-val (SW-test)	p-val (t-test)	D	+/-CI
10,0	0,99	0,17	27	46
15,0	0,91	0,06	31	34
20,0	0,92	0,32	28	69
25,0	0,13	0,06	46	50
27,5	0,83	<b>0,04</b>	87	79
30,0	0,86	0,05	96	99
32,5	0,40	<b>&lt; 0,01</b>	169	81
35,0	<b>0,04</b>	<b>&lt; 0,01</b>	222	94
37,5	0,79	<b>&lt; 0,01</b>	310	88
40,0	0,76	<b>&lt; 0,01</b>	421	165

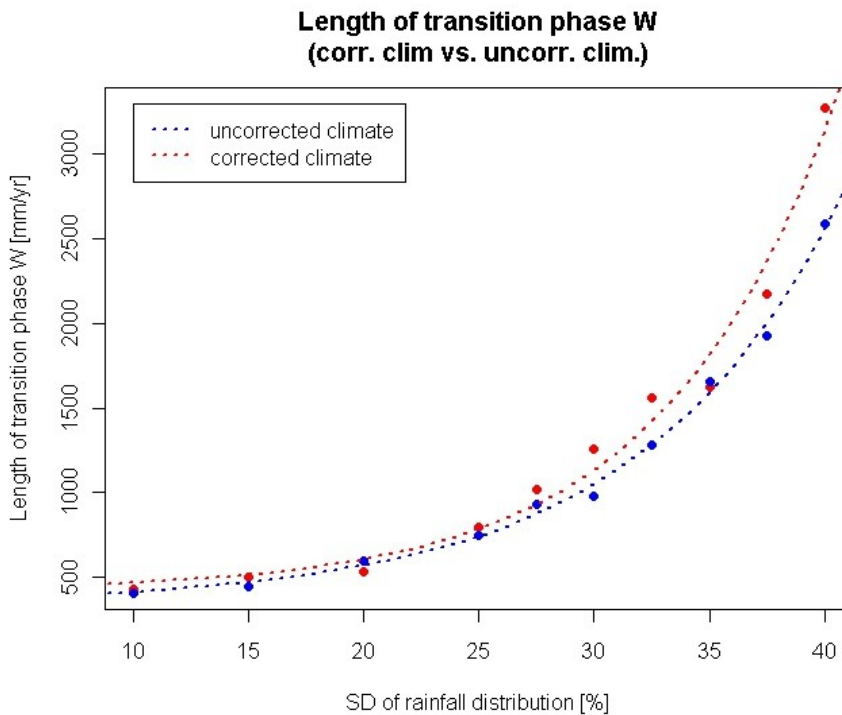


Figure 6.10: Exponential growth of mean length of the transition phase (W) corrected (red symbols) versus uncorrected (blue symbols) in mm/yr with increasing year-to-year variation of rainfall (SD of the precipitation distribution). The dashed lines present the exponential regression.

To make the picture complete, we must also include the width of the transition phase in our calculations. The steps performed are exactly the same as before (when the mean points of inflection were fitted), table 6.4 provides a comparison of estimates derived from the exponential

regression for the corrected and uncorrected version.

Table 6.4: Results from the exponential regression of the transition phase ( $W$ ) for simulations based on uncorrected and corrected climates.

Uncorr:	$A = (23 \pm 8) \text{ mm/yr}$	$r^2 = 0.77$
	$t = (8.8 \pm 0.7) [\%]^{-1}$	adj. $r^2 = 0.76$
	$y_0 = (340 \pm 60) \text{ mm/yr}$	RMSE = 370 mm/yr
Corr:	$A = (13 \pm 8) \text{ mm/yr}$	$r^2 = 0.79$
	$t = (7.5 \pm 0.2) [\%]^{-1}$	adj. $r^2 = 0.78$
	$y_0 = (420 \pm 120) \text{ mm/yr}$	RMSE = 440 mm/yr

Higher variation in  $W$  makes it harder to predict than  $X_{0.5}$ . This is expressed by lower  $r^2$  values and root mean squared errors. Still, the fact that the uncorrected climates yield stabler simulations also seems to hold in this case (compare blue and red line in figure 6.10). We have mentioned that there is a trend visible in  $W$ , as it appears that the width of the transition phase starts at a relatively high value for 2DM, narrows down towards 4DM, and that widens again approaching 6DM (figures 6.6-Error: Reference source not found). To make that trend visible, in figure 6.11 we expressed  $W$  in percent of the according point of inflection for all dry months. The distribution derived that way is

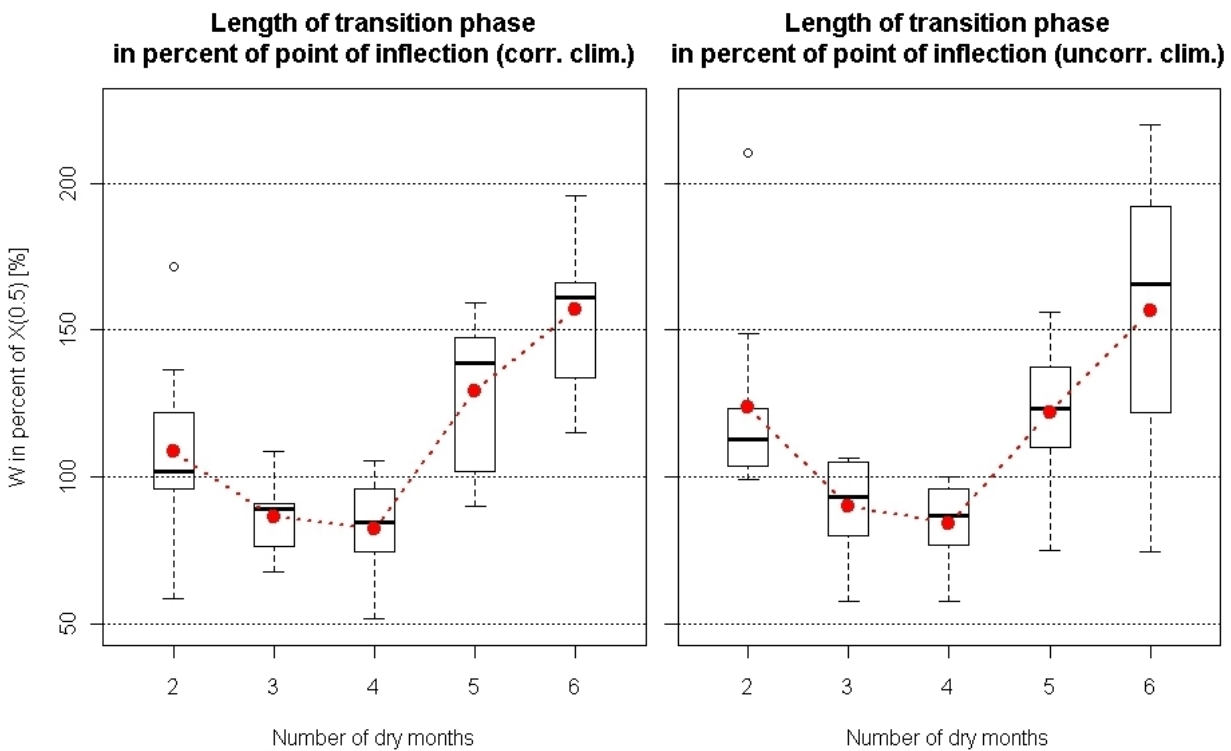


Figure 6.11: Width of transition phase in percent of the according  $X_{0.5}$  (corrected vs. uncorrected) for 2DM to 6DM. Dots illustrate the mean, the boxplot provides information about the underlying distribution of relative  $W$ .

presented in form of a boxplot, mean values of the distribution are indicated by round symbols and dashed lines.

What we have qualitatively observed is supported by figure 6.11, mean (dots) as well as median (black lines) at 4DM are the lowest, in both, the corrected and the uncorrected version, and furthermore most of the distribution stays below the 100% mark. 6DM exhibits the largest  $W$  in percent of  $X_{0.5}$ , but at least in the uncorrected case the distribution is probably too wide to make save predictions.

How does this presentation help us? To demonstrate the surplus value of the functions we have derived,  $X_{0.5}$  &  $W$  will be estimated for sites Moanda, Mouila and Lastoursville, and we will determine whether they are within a stable or unstable region, according to the simulation results.  $U$  is defined as the upper boundary of the transition phase ( $U = W/2 + X_{0.5}$ ), so it tells where the system might start to get unstable. A worst case estimation for  $U$  using figure 6.11 will be provided as well.

We consider stratiform cloud during the dry season, and therefore apply the functions derived for the corrected climate. (Note that to calculate  $X_{0.5}$ ,  $W$  and  $U$  the exact regression coefficients were used, so the results might diverge applying the rounded functions below.)

$$X_{0.5} = [340 + 20 \cdot \exp(\frac{SD}{8.5}) \pm 100] \text{ mm/yr} \quad \text{eq. 6.2}$$

$$W = [420 + 13 \cdot \exp(\frac{SD}{7.5}) \pm 440] \text{ mm/yr} \quad \text{eq. 6.3}$$

Moanda's mean annual rainfall is 1940 mm/yr, with a standard deviation of 24.1% of the mean. This results in  $X_{0.5} = (690 \pm 100)$  mm/yr and  $W = (740 \pm 440)$  mm/yr and further  $U = (1060 \pm 320)$  mm/yr or in  $U = (1060 \pm 650)$  mm/yr if we make use of the Gaussian error propagation. Taking the worst case as suggested by the error estimates, we obtain  $U = 1710$  mm/yr. If we assume a climate with 2 dry months (as originally generated by MarkSim), figure 6.11 suggests that  $W$  is in average around 110% of  $X_{0.5}$ , which would result in  $U = (1070 \pm 160)$  mm/yr (since  $W = 1.1 \cdot X_{0.5}$  and  $U = X_{0.5} + W/2 = X_{0.5} + 1.1 \cdot X_{0.5}/2$ ,  $U$  only depends on the error in  $X_{0.5}$ ). Taking the furthest point out in the distribution for 2DM in figure 6.11, we would approximately obtain  $W = 170\%$  of  $X_{0.5}$ , further using the maximum error suggested for  $X_{0.5}$  (which leads to in  $X_{0.5} = 790$  mm/yr), this results in  $U = 1460$  mm/yr, as a worst case scenario. To summarize:  $X_{0.5} = (690 \pm 100)$  mm/yr, a good estimate for  $U$  might be  $U = (1070 \pm 100)$  mm/yr, but in the worst case  $U = 1460$  mm/yr. We can conclude that Moanda's rainforest (neglecting the fact that the samples taken to parametrize the ecosystem model were not taken in Moanda exactly) should be in a stable region with  $\text{prcp} = 1940$  mm/yr  $\pm 24.1\%$ . Table 6.5 provides further estimates for the other sites, always using the functions for corrected climate, and presenting a 2DM worst case estimate for  $U$ .

Table 6.5: Estimates for point of inflection ( $X_{0.5}$ ), transition width and upper boundary of the transition phase.  $U$  (110%) is the most probable estimate for  $U$ ,  $U$  (170%) constitutes a “worst case scenario”. Lastoursville\*: mean annual rainfall was taken from literature (table 4.5).

Site	Annual rainfall [mm/yr]	Rainfall variation [%]	$X_{0.5}$ [mm/yr]	$W$ [mm/yr]	$U$ [mm/yr]	$U$ (110%) [mm/yr]	$U$ (170%) worst case [mm/yr]	Conclusion
Moanda	1940	24.1%	690+/-100	740+/-440	1060+/-650	1070+/-160	1460	stable
Mouila	2069	24.4%	700+/-100	760+/-440	1080+/-650	1090+/-160	1480	stable
Lastoursv.	1317	34.5%	1500+/-100	1720+/-440	2360+/-630	2330+/-160	2960	unstable
Lastoursv.*	1778	25.6%	750+/-100	810+/-440	1160+/-650	1170+/-160	1580	stable

We can conclude that Moanda and Mouila are far away from unstable regions. Lastoursville, including data from literature (Maloba Makanga, 2010) for our calculations is still in a stable region. If we make use of data derived from weather station measurements (which are seemingly fragmentary) the results tell us that Lastoursville is in a region below the point of inflection, and that forest should only be able to develop with little chance.

In order to explain how against our expectations the uncorrected climates (with higher temperature in the dry season and more global radiation causing larger amounts of water losses through evaporation and transpiration) could yield stabler simulations than the corrected climates, and why a dry season of 6 dry months increases stability when variation is getting higher, it might be helpful to assess the role of carbon stored in above-ground living biomass. It seems obvious, that the

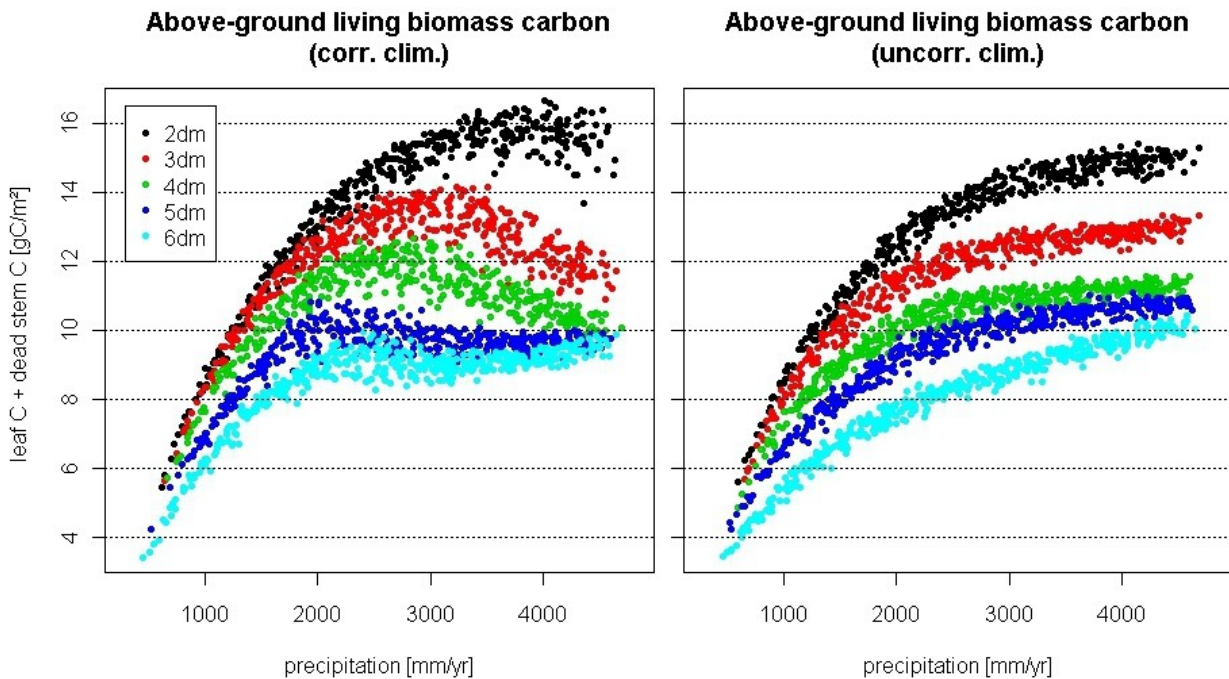


Figure 6.12: Above-ground living biomass carbon (= leaf carbon + dead stem carbon) from all successful simulations averaged over the last mortality cycle of each simulation. Carbon levels (in different colors) correspond to different amounts of dry months (2 to 6).

more carbon stored in living biomass, the higher are the costs to support the ecosystem with water and nutrients. If variation of annual rainfall is little, i.e. the supply of available water remains more or less constant over time, ecosystems carrying a lot of carbon can develop with a minimum of risk. With increasing variation of annual precipitation, the chance is getting higher that the system might have to face a situation where water supply is not enough to support the living costs of the biomass. If a stressful situation like that remains for some years, the ecosystem will probably collapse. Systems with smaller living biomass have lower "cost of living", so they are adapted better to water shortages. On the other hand, a surplus of water is not used for growth but leaves the system through runoff events. This might explain why a simulation based on a climate with 600mm rainfall per year and little variation of SD=10% (which means that the extreme values are 480mm/yr and 720mm/yr) can be perfectly stable, but based on a higher annual mean of 1200 and a variation of SD=30% (also resulting in a minimum of 480mm/yr and a maximum of 1920mm/yr) rather tends to collapse: The system developing with less water does not accumulate a lot of carbon in the first place and is prepared for water shortage. The other system might even use years with precipitation of 1920 mm/yr to extent its storage of carbon over its average level, and facing a year with little rainfall the costs to support living biomass are higher than the available supply, which causes break down. This theory is supported by figure 6.12, which suggests that the above-ground living biomass stored in a system with prcp=1200mm/yr is about 50% higher than the amount stored in an ecosystem based on prcp=600mm/yr.

Furthermore Figure 6.12 demonstrates, that systems with longer dry seasons develop less living biomass, in fact we can observe distinct levels of carbon corresponding to the amount of dry months of the underlying climate. Again the same explanation applies: Longer dry seasons inhibit the accumulation of carbon due to a reduced growing season, the surplus of water in the short wet season can not be used for growing purposes. A smaller pool of living biomass allows the system to survive shortages of water and nutrients, due to reduced "costs of living" and is thereby better prepared for higher climatic fluctuations. This explains why for little variation 2DM climates yield stabler simulations, or simulations equally stable, and as variation increases the role of carbon stored in living biomass becomes more and more important, and these simulations tendentially become the unstablest. So if we state that an ecosystem developing under a climate with a long dry season of 6DM is more stable than the according system based on a climate with a shorter dry season, we must not forget that the first system has less standing biomass in the first place, which is not desirable in common sense!

Does this explication also hold to make us understand the difference in stability between the corrected and the uncorrected versions? In figure 6.12 the corrected climate produces carbon curves that first form and then decline towards higher precipitations, while the other one seems to approach a constant level. Depending on the range of precipitation considered, either the corrected

or the uncorrected version is at a higher level, and the differences are smaller than the divergences in carbon resulting from different numbers of dry months, which cause a shift in stability at the same scale.

### 6.1.4 Spindown simulations

Until now we have considered the development of stationary ecosystems “from scratch”, probabilities computed this way can be interpreted as probabilities of regrowth or reestablishment after a disturbance. A different situation, as discussed before, is when a steady state system, established on a “well-behaved” climate with only a short dry season and enough rainfall, suddenly faces a climatic change. The methods described for the spinup-simulations are still valid, but include further considerations since break down might on one hand depend on the initial steady state system itself, and on the other hand also on where within the mortality cycle the climate is changed, since mortality might affect the system's reaction to the same climatic transition. This implies a larger number of time series to be computed considering all combinations of initial stationary systems and positions in the mortality cycle (shift dates), and keeping in mind that each simulation has to be repeated multiple times. Since we could theoretically choose of an infinite variety of climates setting up the initial stable state system, a reasonable approach that reduces the amount of simulations is to select only the original MarkSim climate of P3, where the samples for the epc-parametrization of Biome-BGC have been taken. Or in other words we use a climate based on the unchanged climate normals that served as template to create all P3 climate varieties (since this allows us to tell how the *present* ecosystem might react to a shift in climate). This climate's mean annual precipitation = 1550 mm/yr, with SD=26% (normally distributed). The stationary state variables are stored, and each simulation is initialized based on the same

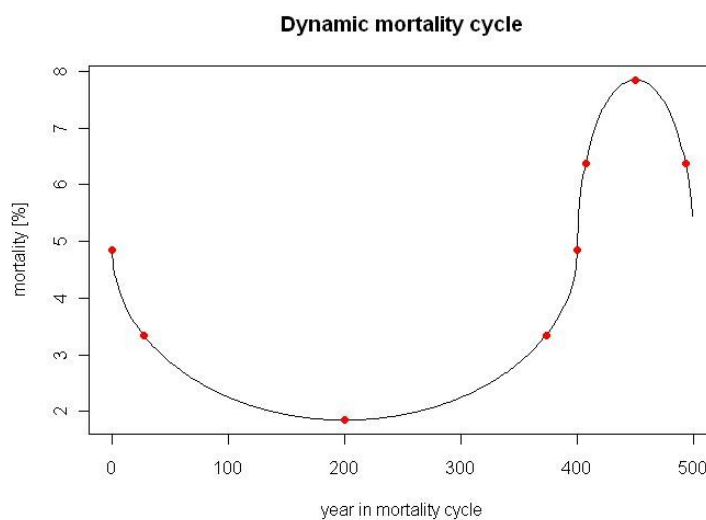


Figure 6.13: Mortality cycle of the underlying parametrization (black line), selected points where climatic change is performed (dots).



variables. Within the mortality cycle of 500 years 8 positions are defined where the climate is changed. Until having reached this point, the model continues to base its simulations on the original climate.

The maximum of the mortality cycle is at 7.85%, while the minimum is at 1.85%. Further the low and high mortality phase take 400 years and 100 years, respectively. We chose 5 equidistant mortality steps between minimum and maximum, and computed their according years of occurrence in the mortality cycle: 1.85% → 200yrs, 3.35% → 27yrs and 373yrs, 4.85% → 0yrs and 400yrs, 6.35% → 407yrs and 493yrs, 7.85% → 450yrs

This results in 8 shift dates in the mortality cycle, where the climatic change is performed (ordered: 0, 27, 200, 373, 400, 407, 450 and 493 yrs).

The analytical steps to assess the results delivered by the spindown simulations are undertaken in analogy to the methods presented before, discussing the outcome of the spinup procedure, with one exception, that is, the different shift dates in the mortality cycle. For this reason, the following chapter is based mainly on the presentation of figures, while explaining text sections are reduced, since many observations and arguments are similar to statements in the previous chapter. First simulations based on an uncorrected climate are discussed, after that the corrected version is presented. Finally we will again provide a comparison between corrected and uncorrected.

### ***Uncorrected climate (cumuliform)***

Since we chose to split the mortality cycle with respect to equidistant mortality values (instead of equidistant shift dates), some regions in the cycle appear in higher resolutions (the high mortality phase between the 401<sup>st</sup> and the 500<sup>th</sup> year in the mortality cycle) than others.

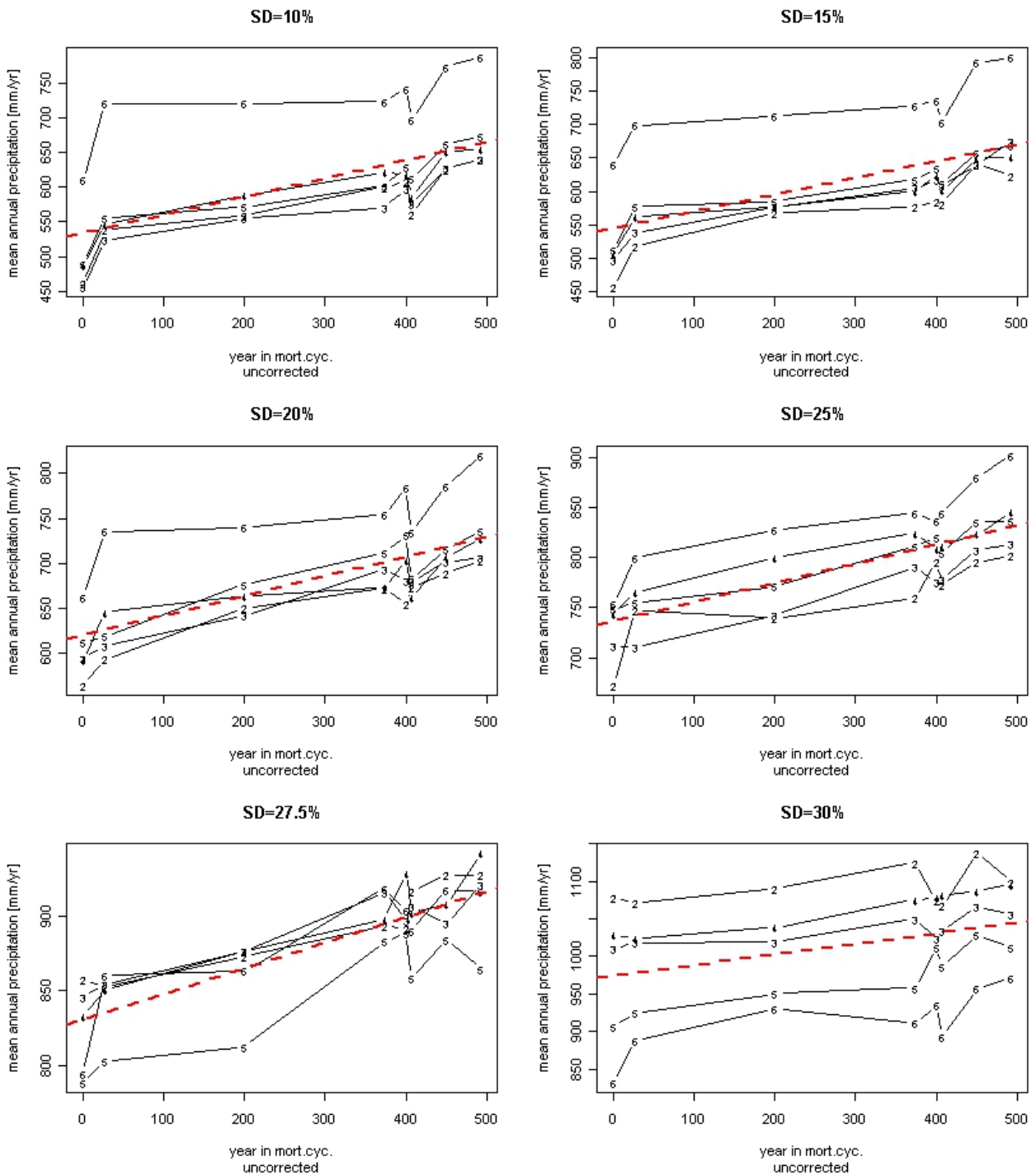


Figure 6.14: Point of inflection (X0.5) in mm/yr derived from logistic regression for different shift dates within the mortality cycle of 500 years, based on an underlying uncorrected climate with different values of annual rainfall variation (SD=10% to 30%). The numbers indicate the amount of dry months (2-6DM) of the according graph. The dashed line presents a linear regression performed on the data to illustrate a possible trend.

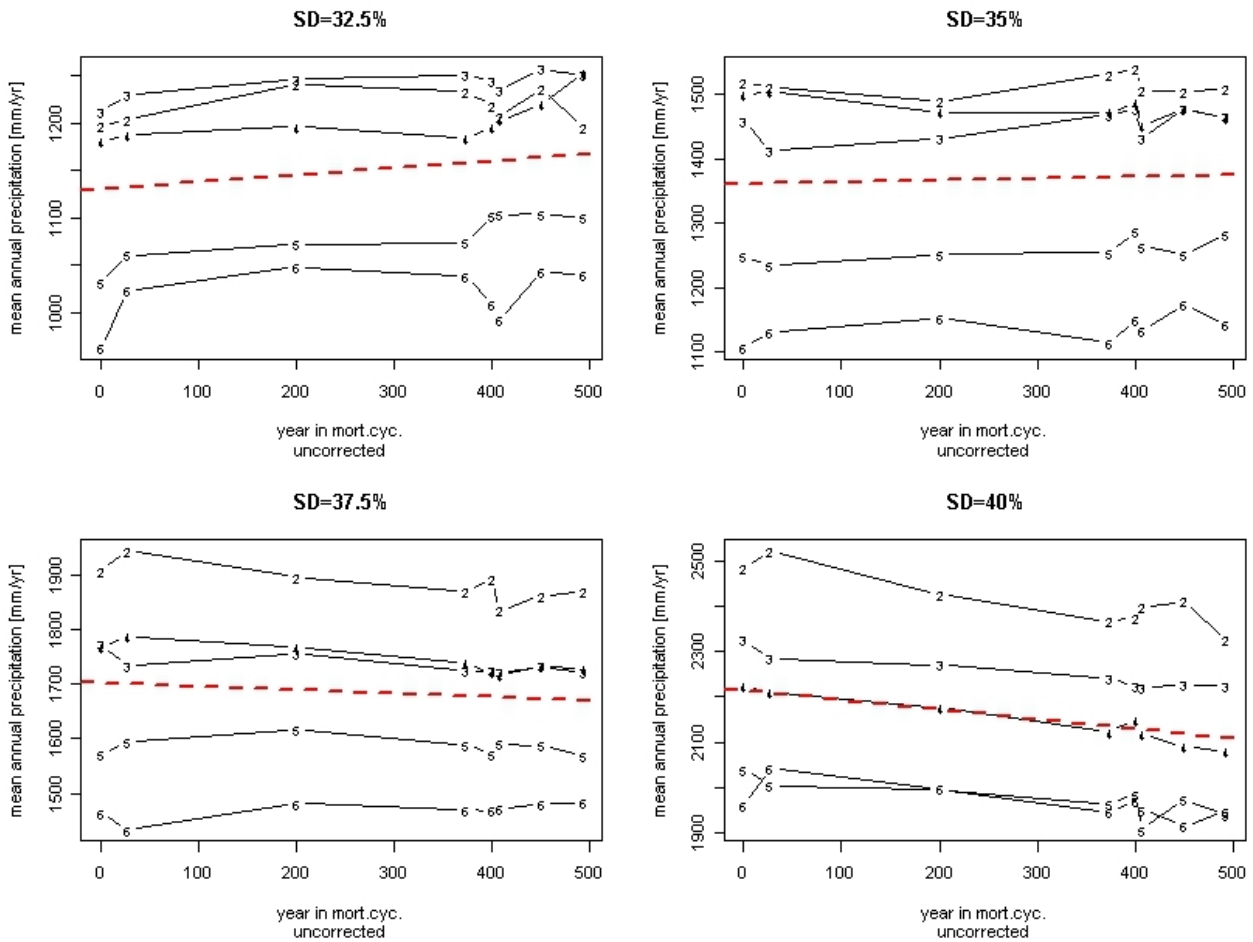


Figure 6.15: Point of inflection (X0.5) in mm/yr derived from logistic regression for different shift dates within the mortality cycle of 500 years, based on an underlying uncorrected climate with different values of annual rainfall variation (SD=32.5% to 40.0%). The numbers indicate the amount of dry months (2-6DM) of the according graph. The dashed line presents a linear regression performed on the data to illustrate a possible trend.

This approach has the disadvantage that possible fluctuations in regions with a lower resolution (the low mortality phase) might be hidden, and that it becomes difficult to put a statement on whether the fluctuations around year 450 in the cycle can be connected to the high mortality (compare with figure 6.13), or whether this variations occur all along the cycle. Nevertheless, one effect becomes apparent: A linear regression (dashed line) has been fitted to all data points of one certain variation (standard deviation of the rainfall distribution), just to qualitatively assess the trend. We can observe an increasing trend for low variations that becomes flatter as variation is rising, and seems horizontal at SD=35%, where it starts to turn around to become decreasing for SD=37.5% and SD=40.0%. This means, that a shift performed to a climate with low variation in the beginning of the mortality cycle yields stabler simulations than when the shift is done towards the end of the mortality cycle. The opposite becomes true, once the variation has exceeded SD=35.0%. An interesting fact, that results from the periodicity of the mortality cycle is, that a shift in year 0

should yield the same results as a shift in year 500. As a matter of fact, we see the biggest difference between year 0 and year 493 (or, with this assumption, between year 493 and 500). So, a small difference of 7 years in the cycle causes the largest difference in stability! (Note that  $W$  does not depend on the shift date in the mortality cycle, figure 6.27)

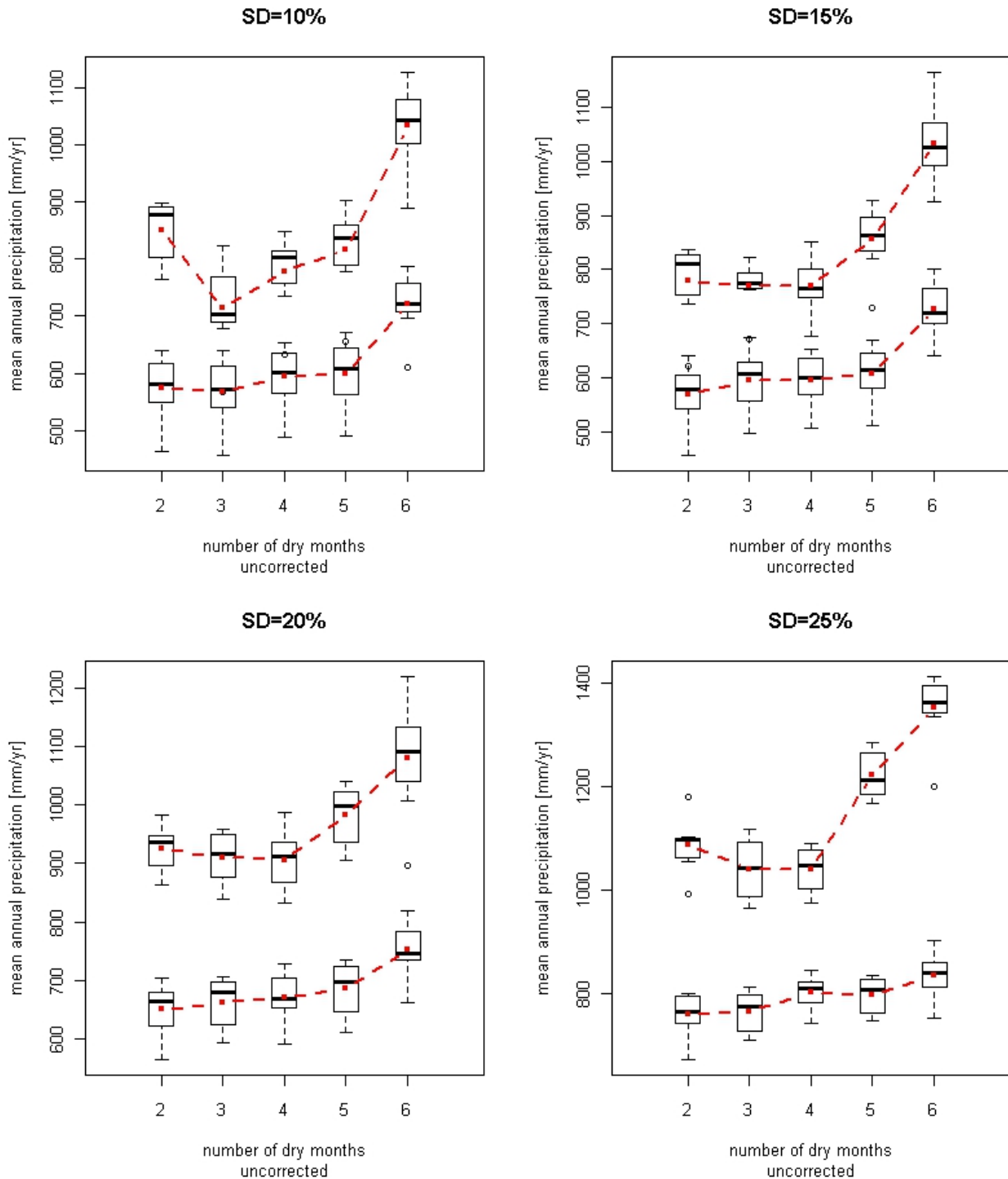


Figure 6.16: Point of inflection ( $X_{0.5}$ ) and upper limit of the transition phase ( $U$ ) in mm/yr derived from the logistic regression, for simulations based on uncorrected climate with 2-6 dry months and annual rainfall variation from  $SD=10\%$  to  $SD=25\%$ . Boxplots indicate the variation resulting from different shift dates in the mortality cycle, the dashed line represents the mean values.

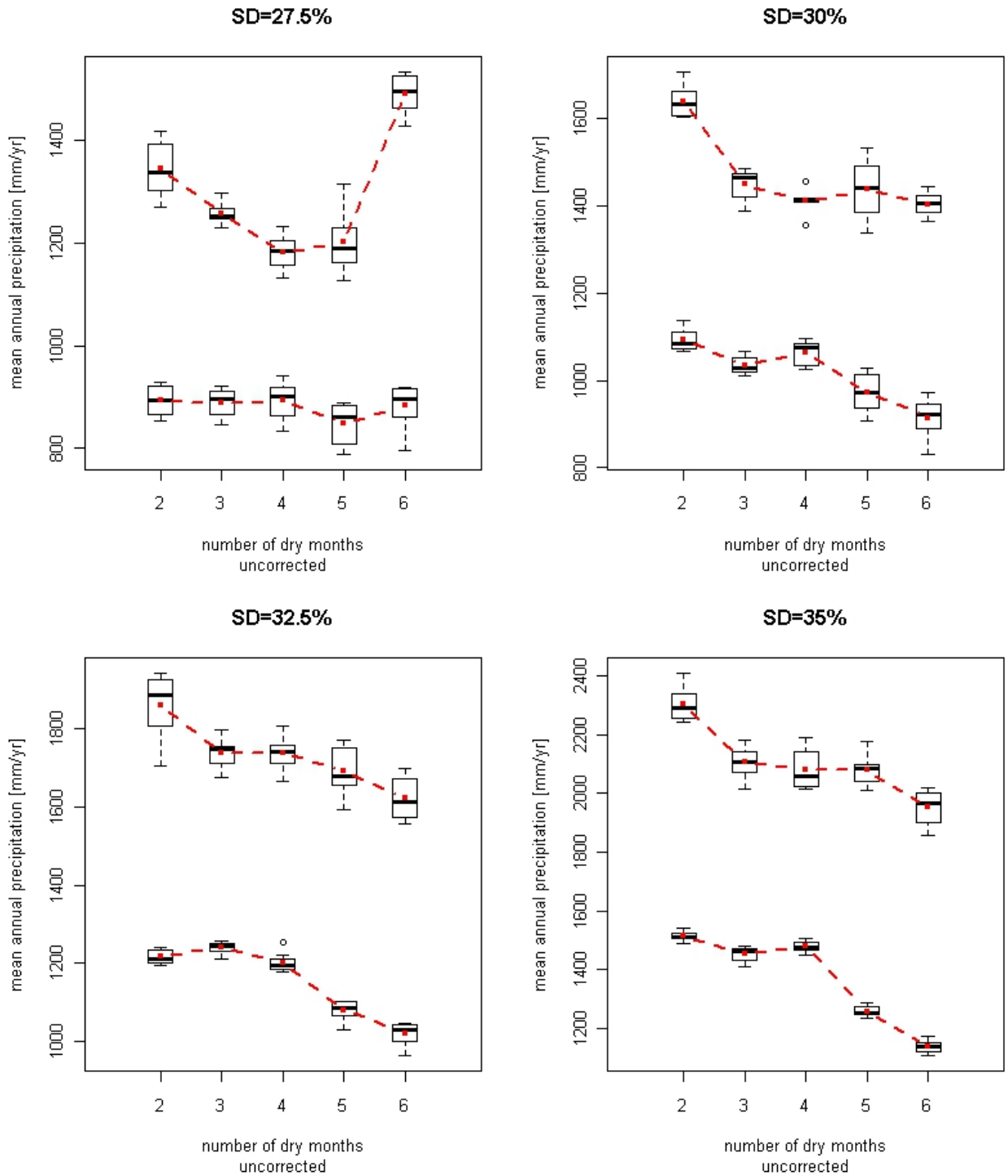


Figure 6.17: Point of inflection (X0.5) and upper limit of the transition phase (U) in mm/yr derived from the logistic regression, for simulations based on uncorrected climate with 2-6 dry months and annual rainfall variation from SD=27.5% to SD=35%. Boxplots indicate the variation resulting from different shift dates in the mortality cycle, the dashed line represents the mean values.

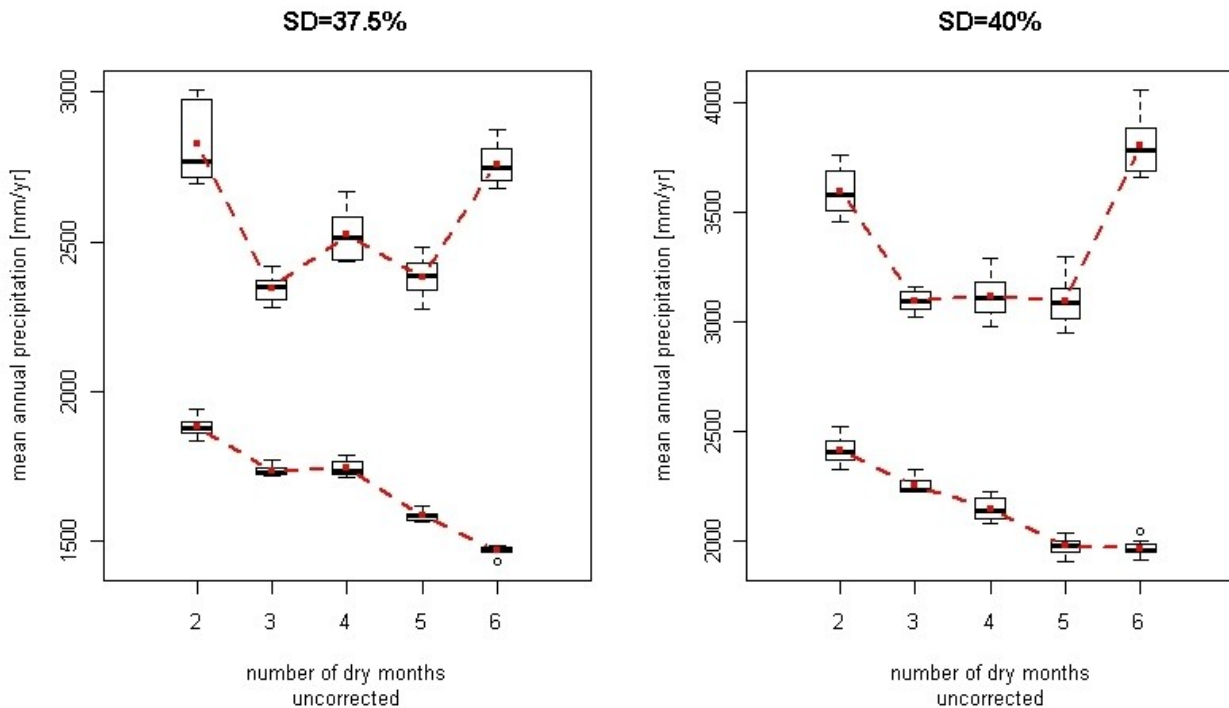


Figure 6.18: Point of inflection ( $X_{0.5}$ ) and upper limit of the transition phase (U) in mm/yr derived from the logistic regression, for simulations based on uncorrected climate with 2-6 dry months and annual rainfall variation from SD=37.5% to SD=40%. Boxplots indicate the variation resulting from different shift dates in the mortality cycle, the dashed line represents the mean values.

In figures 6.16-6.18,  $X_{0.5}$  and U (the upper boundary of the transition phase) are presented (the lower boundary is symmetric to U around  $X_{0.5}$ ), boxplots indicate the variation due to the different shift dates. As figures 6.14 and 6.15 already suggest, the influence of the different shift dates is dominant for lower variations of annual precipitation, as boxplots showing wide distributions indicate. With higher variation, this effect shrinks, and gradually the influence of the dry season becomes predominant. In analogy to results of the spinup simulations we see that for low variations simulations based on climates with 6 dry months are the unstablest, and step-by-step become the most stable as variation increases.

One might suspect similar effects behind the inversion of the trend of stability with respect to different shift dates in the mortality cycle, and with respect to the length of the dry season caused by increasing variation of the annual rainfall distribution.

### Corrected climate (stratiform)

Simulations based on corrected climates don't show any surprising behavior compared to the results illustrated before. For the sake of completeness the results are presented below without discussion.

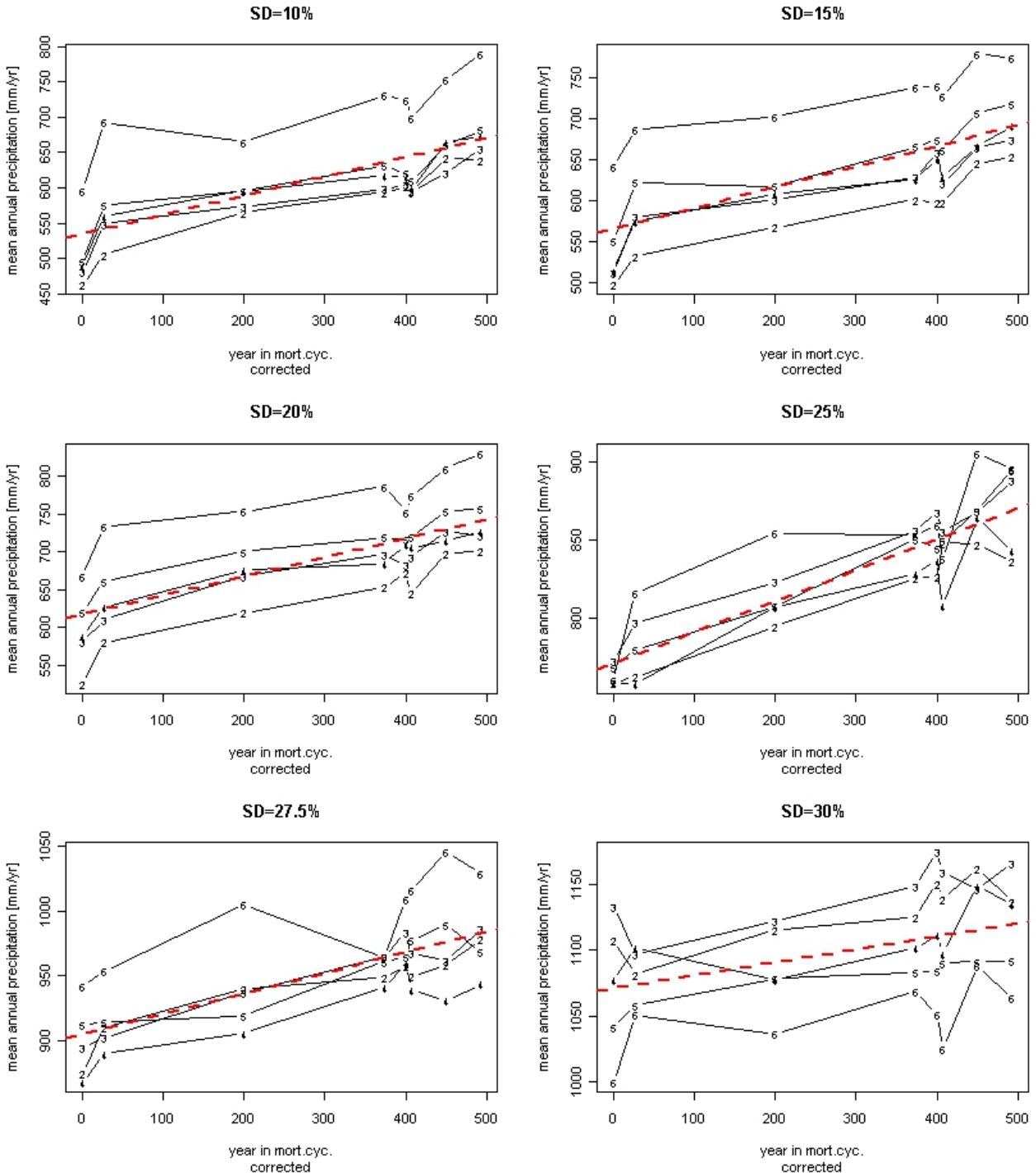


Figure 6.19: Point of inflection (X0.5) in mm/yr derived from logistic regression for different shift dates within the mortality cycle of 500 years, based on an underlying corrected climate with different values of annual rainfall variation (SD=10% to 30.0%). The numbers indicate the amount of dry months (2-6DM) of the according graph. The dashed line presents a linear regression performed on the data to illustrate a possible trend.

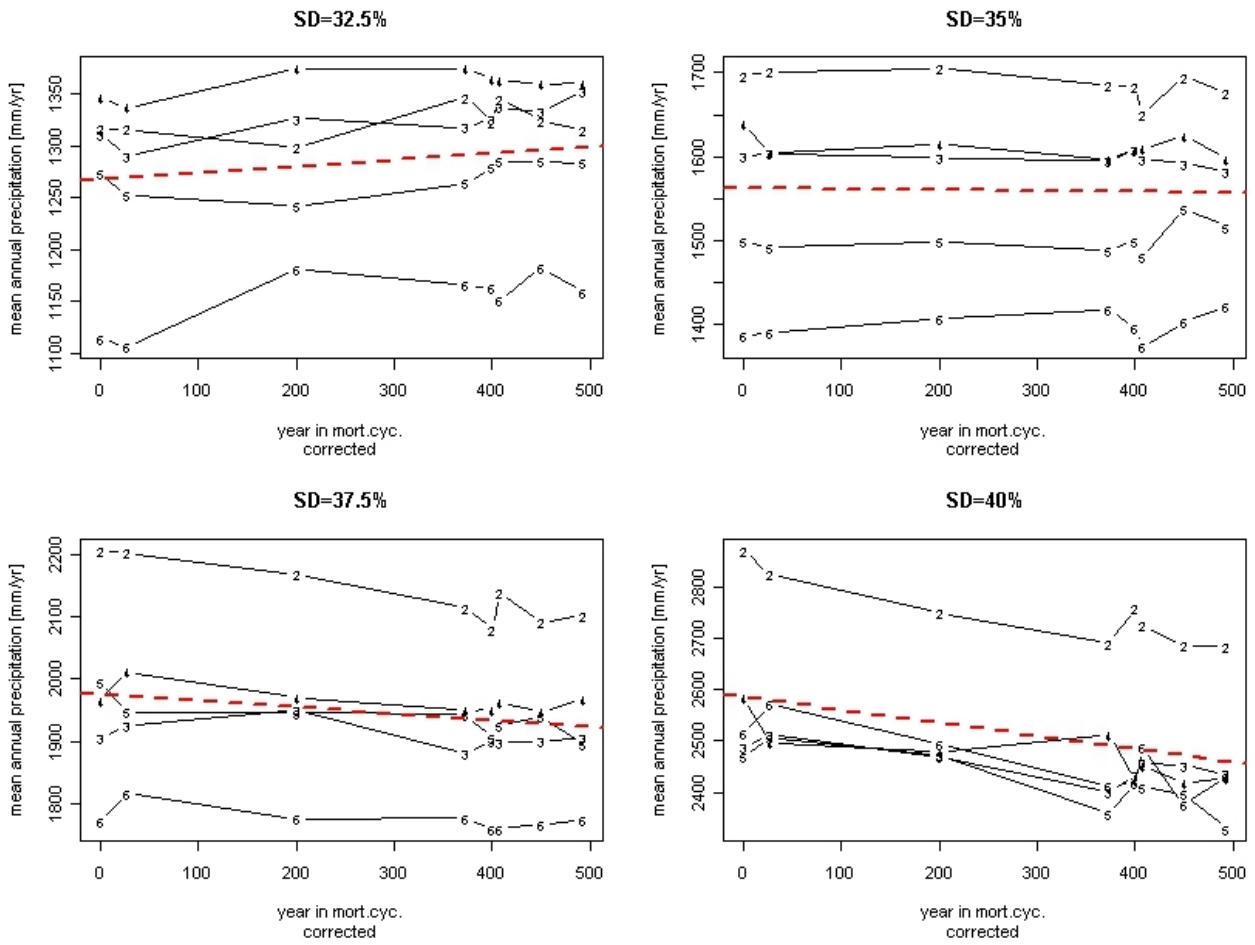


Figure 6.20: Point of inflection (X0.5) in mm/yr derived from logistic regression for different shift dates within the mortality cycle of 500 years, based on an underlying corrected climate with different values of annual rainfall variation (SD=32.5% to 40.0%). The numbers indicate the amount of dry months (2-6DM) of the according graph. The dashed line presents a linear regression performed on the data to illustrate a possible trend.



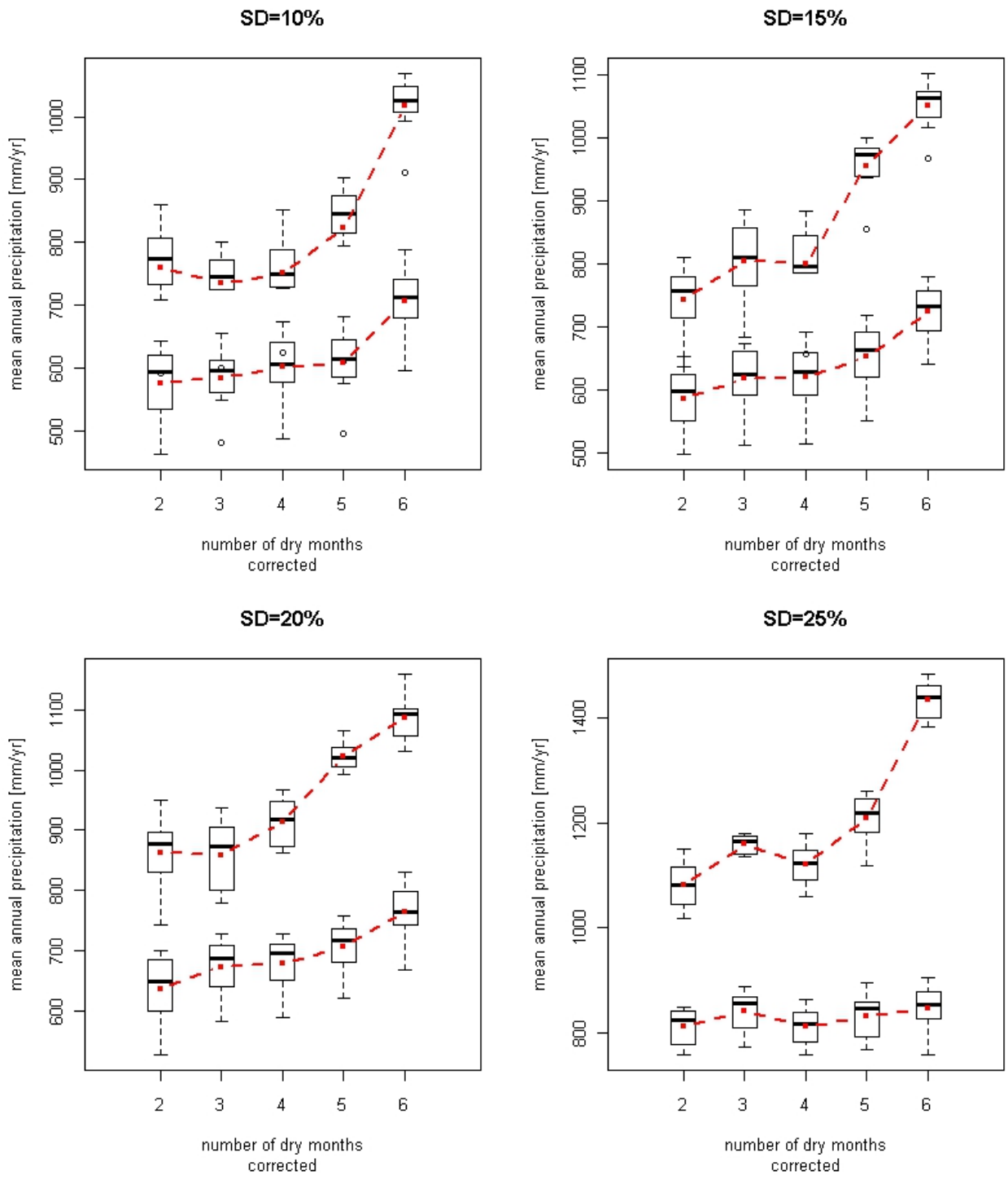


Figure 6.21: Point of inflection (X0.5) and upper limit of the transition phase (U) in mm/yr derived from the logistic regression, for simulations based on corrected climate with 2-6 dry months and annual rainfall variation from SD=10% to SD=25%. Boxplots indicate the variation resulting from different shift dates in the mortality cycle, the dashed line represents the mean values.

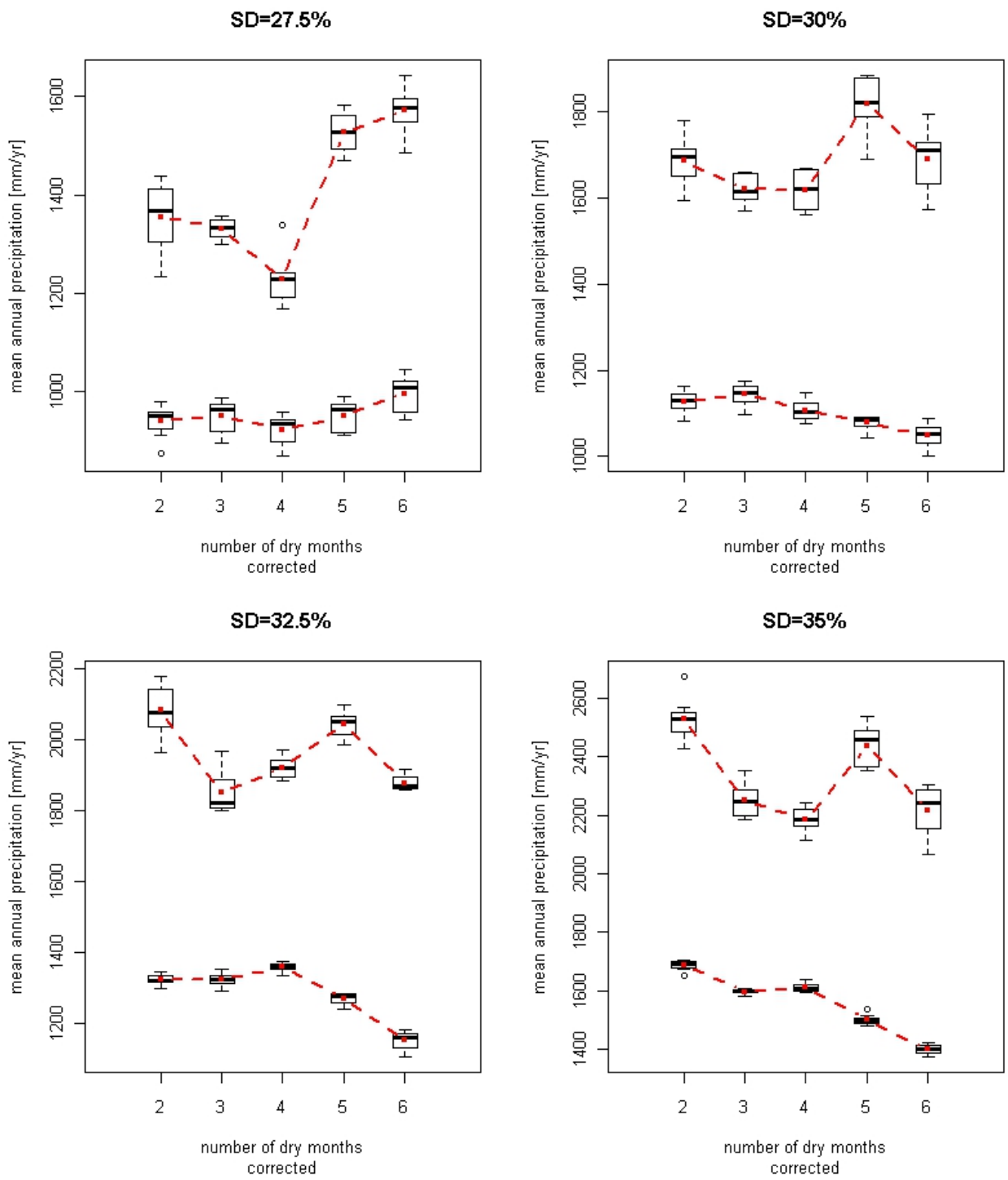


Figure 6.22: Point of inflection (X0.5) and upper limit of the transition phase (U) in mm/yr derived from the logistic regression, for simulations based on corrected climate with 2-6 dry months and annual rainfall variation from SD=27.5% to SD=35%. Boxplots indicate the variation resulting from different shift dates in the mortality cycle, the dashed line represents the mean values.

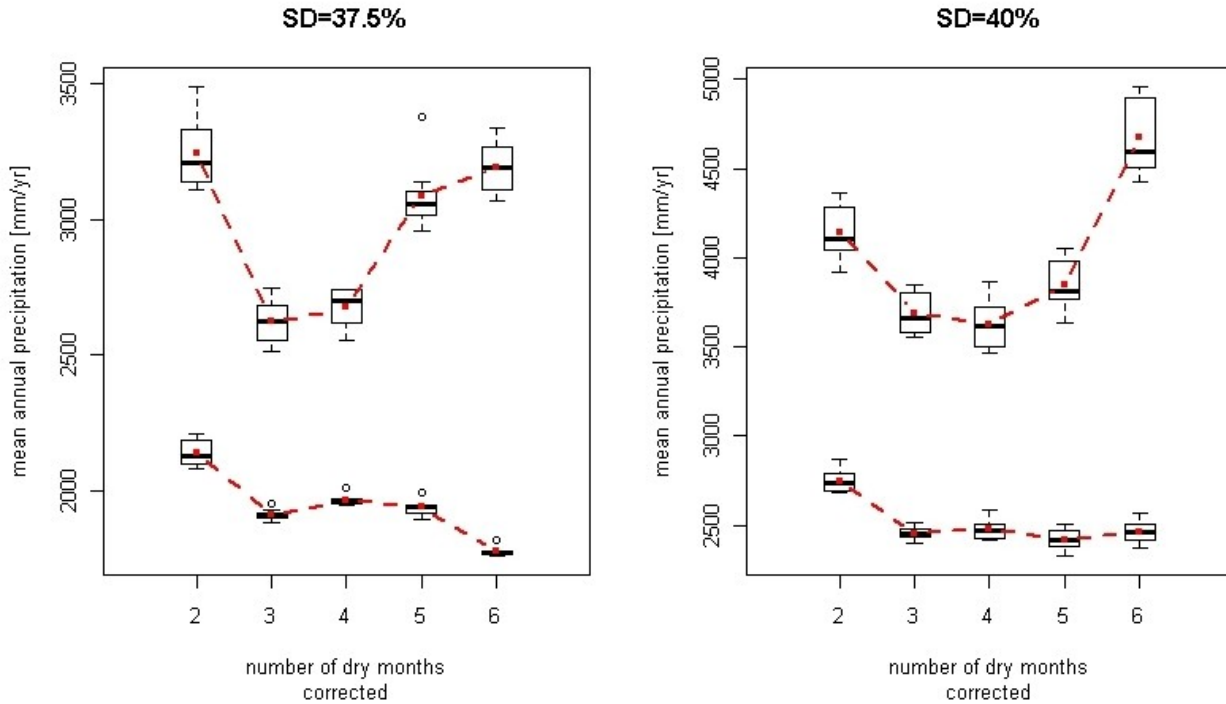


Figure 6.23: Point of inflection ( $X_{0.5}$ ) and upper limit of the transition phase ( $U$ ) in mm/yr derived from the logistic regression, for simulations based on corrected climate with 2-6 dry months and annual rainfall variation from SD=37.5% to SD=40.0%. Boxplots indicate the variation resulting from different shift dates in the mortality cycle, the dashed line represents the mean values.

### **Corrected (stratiform) and uncorrected (cumuliform) compared**

Table 6.6 shows information on whether the exponential increase in instability of simulations based on corrected climate are significantly different from the uncorrected. Compared to the spinup approach, the data set is here considerably larger due to the different shift dates (40 pairs for the spindown versus 5 pairs for the spinup approach). This larger number of pairs results in a reduction of the 95% confidence intervals, which means further that the difference between corrected and uncorrected becomes significant at lower variations, than for the spinup. Before SD=27.5% was the first rainfall variation to show a significant difference, here the level is down at the 15.0% threshold. In both situations, the spinup and the the spindown, the simulations with underlying corrected climate show less stability than their corrected counterparts.

Both, the increase in  $W$  and  $X_{0.5}$  are fitted using a exponential regression model based on a scaled Levenberg-Marquardt algorithm (Levenberg, 1944). Note that only the mean values (for each SD-value) are fitted. To put a statement on the quality of the entire fit with respect to the whole data set, the RMSE and  $r^2$  are computed with all points. Eq. 6.1 defines the form of the exponential function.

Table 6.6: Corrected (stratiform) vs. uncorrected (cumuliform) estimates of instability: SD = standard deviation of the rainfall distribution [%], p-value (Shapiro-Wilk-test), p-value (t-test),  $D = X0.5(\text{corr.}) - X0.5(\text{uncorr.})$ , confidence interval (CI). One data set corresponds to one value of SD and comprises 40 data points (= 5 different lengths of dry season x 8 different shift dates in mortality cycle) from the corrected, and 40 points from the uncorrected version.

Corrected vs. Uncorrected				
SD[%]	p-val (SW-test)	p-val (t-test)	D	+/-CI
10,0	0,06	0,17	4	6
15,0	0,29	< 0,01	21	6
20,0	0,52	0,05	7	7
25,0	0,10	< 0,01	37	9
27,5	0,12	< 0,01	71	12
30,0	0,20	< 0,01	85	15
32,5	0,44	< 0,01	134	14
35,0	<b>0,03</b>	< 0,01	191	18
37,5	0,54	< 0,01	261	22
40,0	0,36	< 0,01	362	35

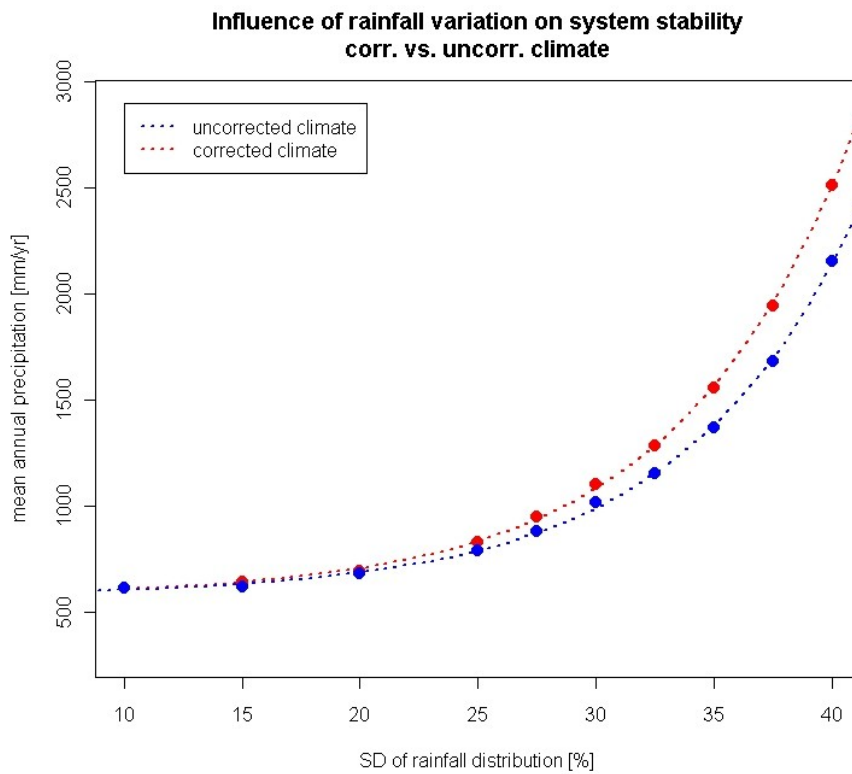


Figure 6.24: Exponential growth of mean point of inflection corrected (red symbols) versus uncorrected (blue symbols) in mm/yr with increasing year-to-year variation of rainfall (SD of the precipitation distribution). The dashed lines present the exponential regression.

Table 6.7: Results from the exponential regression of points of inflection ( $X_{0.5}$ ) for simulations based on uncorrected and corrected climates.

Uncorr:	$A = (7.7 \pm 0.2) \text{ mm/yr}$	$r^2 = 0.96$
	$t = (7.5 \pm 0.2) [\%]^{-1}$	adj. $r^2 = 0.95$
	$y_0 = (580 \pm 20) \text{ mm/yr}$	RMSE = 110 mm/yr
Corr:	$A = (9.5 \pm 0.9) \text{ mm/yr}$	$r^2 = 0.98$
	$t = (7.5 \pm 0.2) [\%]^{-1}$	adj. $r^2 = 0.98$
	$y_0 = (570 \pm 20) \text{ mm/yr}$	RMSE = 90 mm/yr

Also the width of the transition phase ( $W$ ) in figure 6.26 shows a behavior that we already know from the spinup analysis.  $W$  is increasing exponentially with SD (the standard deviation of the rainfall distribution) and we can therefore expect a linear relation between  $X_{0.5}$  and  $W$ . The regression performed on  $W$  results in a higher level of uncertainty than the fitted  $X_{0.5}$  as suggested by a higher root mean squared error (RMSE) and a lower correlation coefficient ( $r^2$ ).

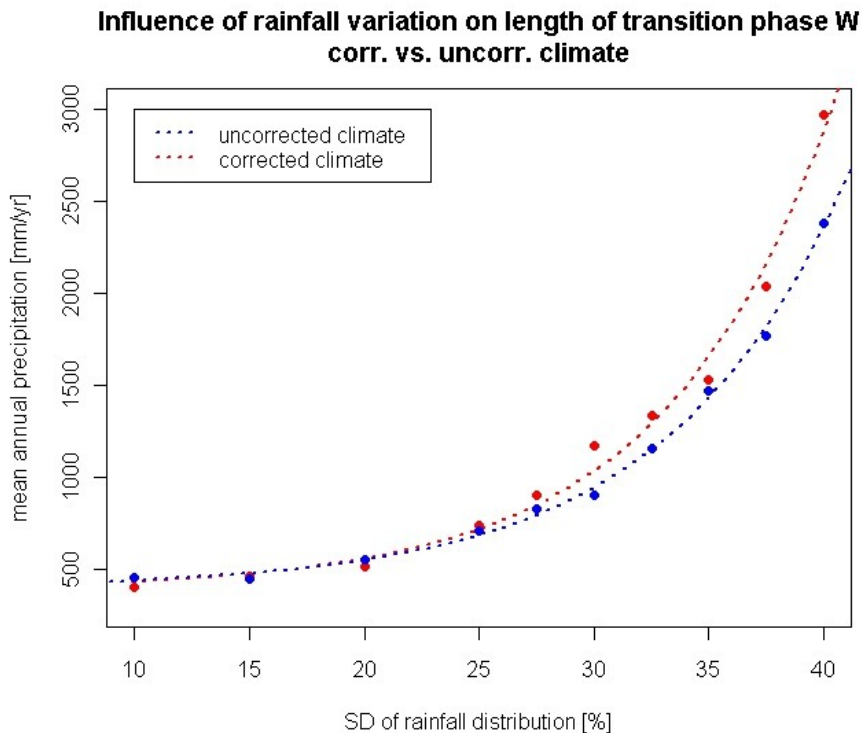


Figure 6.25: Exponential growth of mean length of transition phase ( $W$ ) corrected (red symbols) versus uncorrected (blue symbols) in mm/yr with increasing year-to-year variation of rainfall (SD of the precipitation distribution). The dashed lines present the exponential regression.

Table 6.8: Results from the exponential regression of length of transition phase ( $W$ ) for simulations based on uncorrected

and corrected climates.

Uncorr:	$A = (12 \pm 4) \text{ mm/yr}$	$r^2 = 0.79$
	$t = (7.8 \pm 0.4) [\%]^{-1}$	adj. $r^2 = 0.78$
	$y_0 = (400 \pm 10) \text{ mm/yr}$	RMSE = 320 mm/yr
Corr:	$A = (11 \pm 7) \text{ mm/yr}$	$r^2 = 0.81$
	$t = (7.4 \pm 0.8) [\%]^{-1}$	adj. $r^2 = 0.81$
	$y_0 = (400 \pm 10) \text{ mm/yr}$	RMSE = 370 mm/yr

The influence of the length of the dry season on  $W$  on percent of  $X_{0.5}$  is similar for simulations with corrected and uncorrected climate: A minimum is visible at 3-4DM, and 6DM in average exhibits the largest  $W/X_{0.5}$  ratio, which means that although 6DM simulations are seemingly getting the stablest for higher variations of annual rainfall, their unstable phases are wider than for the rest of the simulations. A similar behavior has already been observed at the spinup approach.

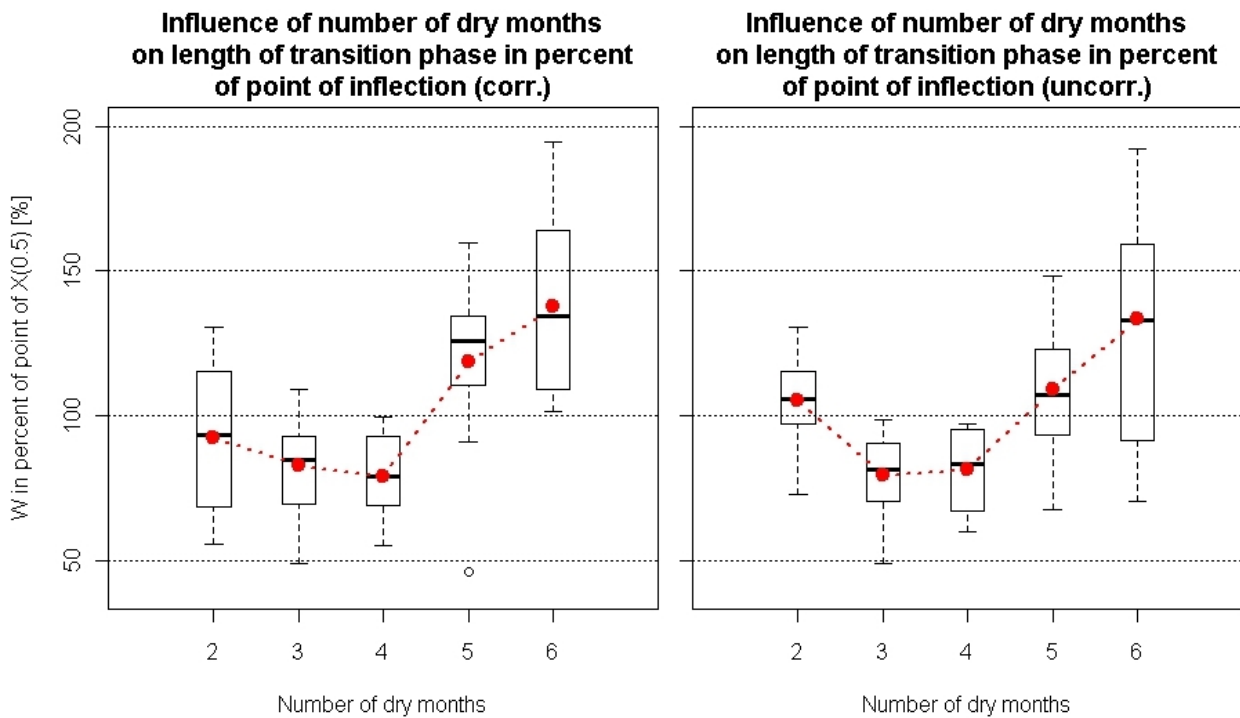


Figure 6.26: Length of transition phase ( $W$ ) in percent of the according point of inflection ( $X_{0.5}$ ) (corrected vs. uncorrected) for 2DM to 6DM. Dots illustrate the mean, the boxplot provides information about the underlying distribution of relative  $W$ .

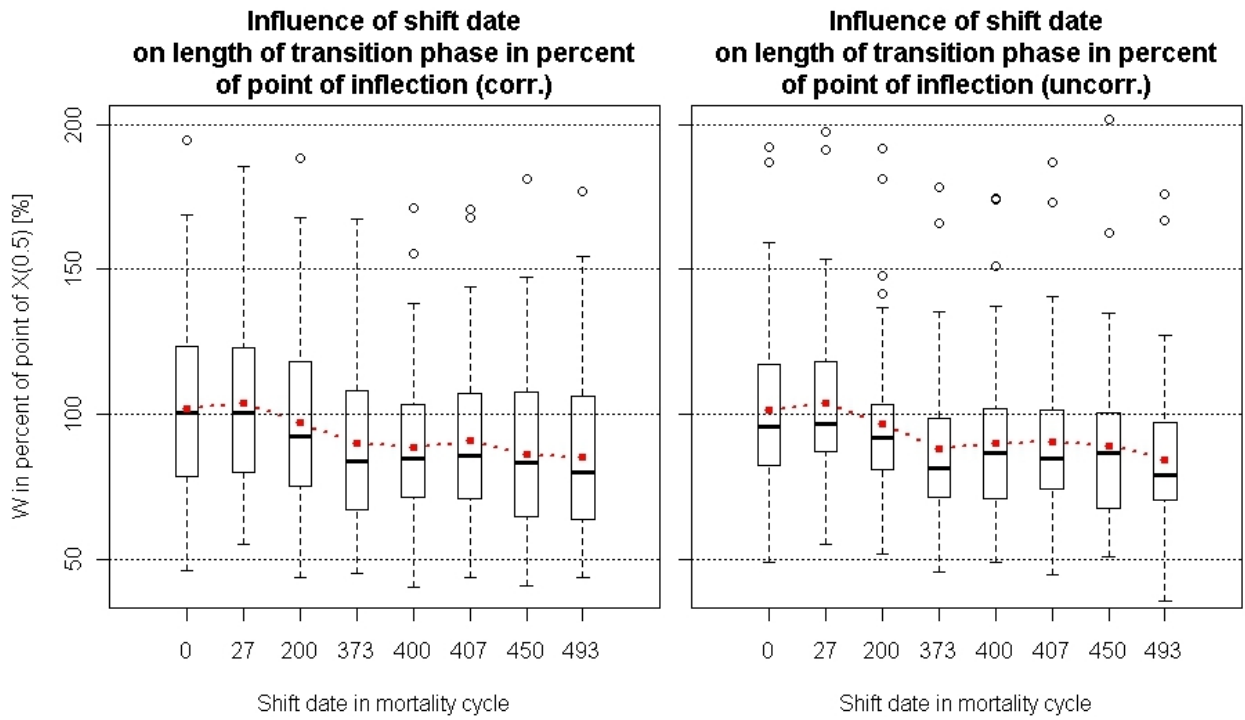


Figure 6.27: Length of transition phase ( $W$ ) in percent of the according point of inflection ( $X(0.5)$ ) (corrected vs. uncorrected) for different shift dates in the mortality cycle. Dots illustrate the mean, the boxplot provides information about the underlying distribution of relative  $W$ .

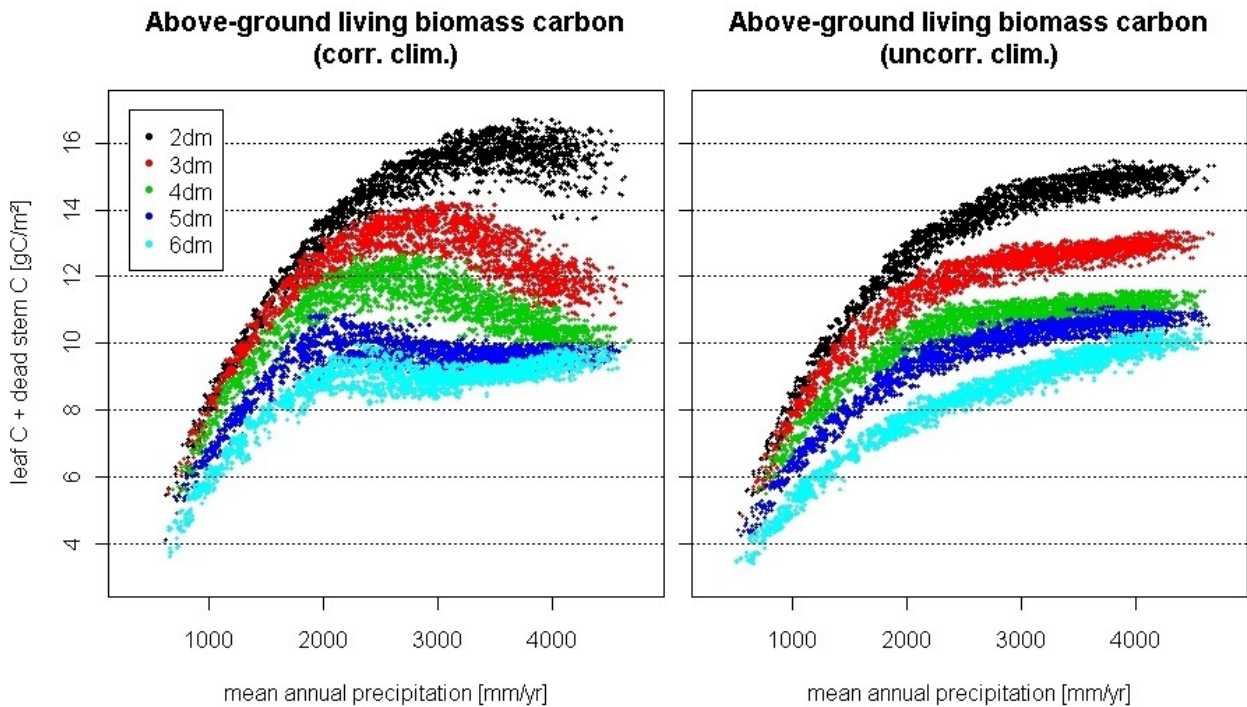


Figure 6.28: Above-ground living biomass carbon (= leaf carbon + dead stem carbon) from all successful simulations averaged over the last mortality cycle of each simulation. Carbon levels (in different colors) correspond to different amounts of dry months (2 to 6).

Although we see a decreasing trend, a clear dependence of the  $W/X_{0.5}$  ratio on the shift date in the mortality cycle can not be determined (figure 6.27) since the scale of the width of the distributions (illustrated as boxplots) is far above the scale of differences between successive means / medians of the distributions.

### 6.1.5 Hysteresis (spinup vs. spindown)

In this last section of the chapter assessing ecosystem stability, the results derived from the spinup and spindown approaches will be compared. In the introduction we have stated that spinup-simulations mimic the growth or regrowth of an ecosystem under certain climatic conditions. On the other hand, the spindown emulates a situation where an existing ecosystem that has reached a stationary state faces climate change.

In ecology the term “hysteresis” describes the path dependency of the stable state that is reached by similar ecosystems under the *same* environmental conditions. “Path” in this case refers to the past of the system, or the states that have been occupied by the system before a shift in external forces pushed the system into the current state. We can directly translate the idea of hysteresis to our experiment: On one side the a simulated ecosystem has to develop from scratch (spinup), on the other side climate change is performed on an existing stable system (spindown). We thereby have different past situations, or different paths to the current state. It is crucial to know that the change in environmental factors has to be identical in both situations, which is also true here: We have reduced the external influence to the impact of climate on which the simulations are based. The climate files used in the spinup and spindown approach *are* identical!

Although the model is kept rather simple (no alternative forms of vegetation within one parametrization) and although the ecosystem does not have a feedback effect on the micro climate in the forest, we can observe effects of hysteresis (figure 6.29), that is, in lower regions of variation of annual rainfall the spinup yields stabler simulations than the spindown, under equal conditions. In other words, a system developing under unfavorable climatic conditions (e.g. under dry conditions) has a higher chance to reach a stable state, than a functioning system that is confronted with the same climate. The difference we observe between spinup and spindown might take place at a small scale, but again, the type of ecosystem can not be altered (since it is defined by the eco-physiological paramters included in Biome-BGC) and feedback effects on climate are not implemented, and both factors are very important factors in connection with hysteresis in real ecosystems.

In figure 6.29 we see that the small difference between the fitted curves resulting from spinup and spindown simulations vanishes for variations of annual rainfall above  $SD=30\%$  to  $35\%$ . The width of the transition phase (figure 6.30) is virtually unaffected by the underlying path to the current stable state, i.e. it exhibits only minor to no hysteresis at all.



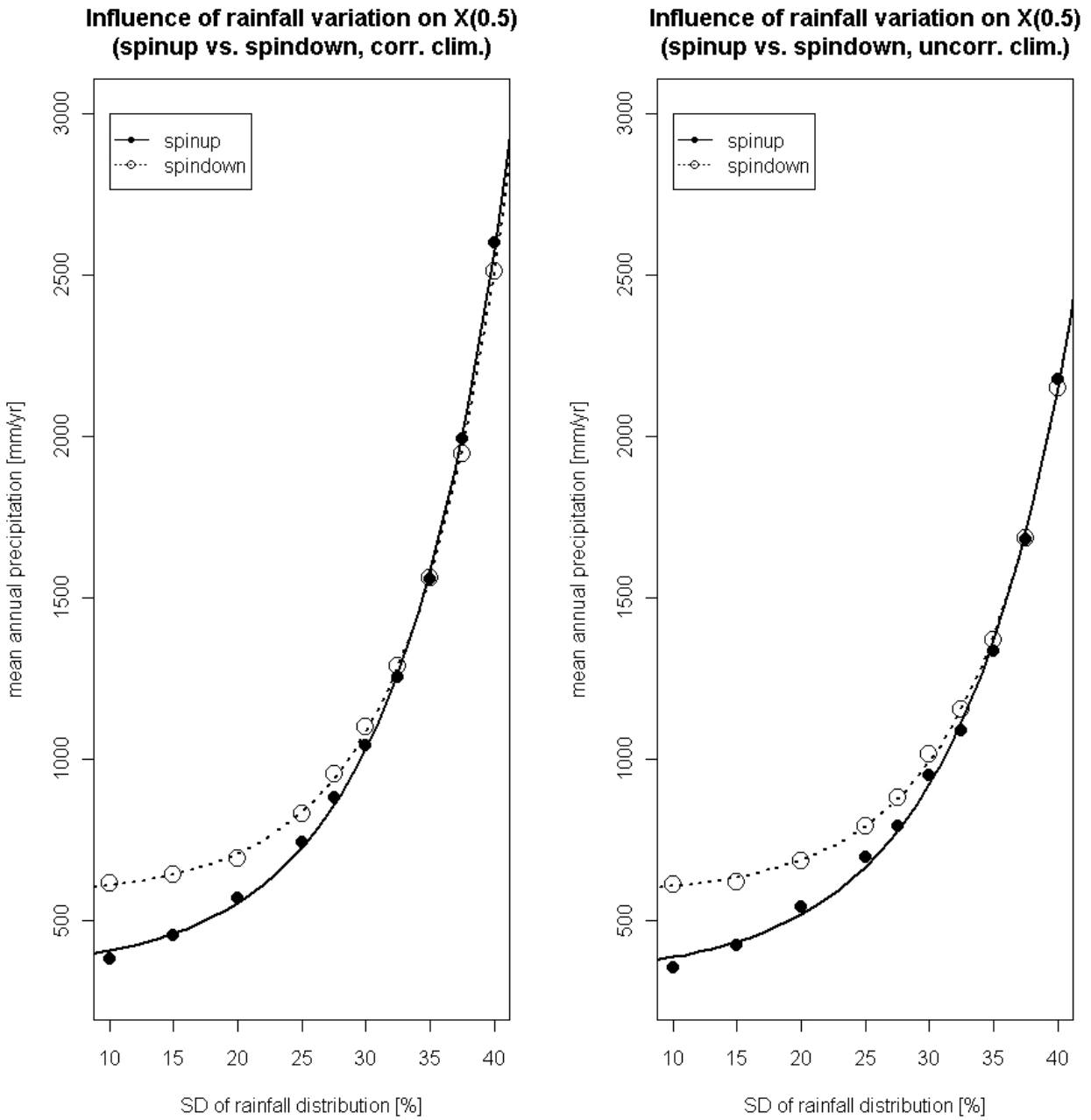


Figure 6.29: Hysteresis: Instability is increasing differently for spinup (solid line) and spindown (dashed line). Exponential regression of mean point of inflection (symbols) in mm/yr with increasing year-to-year variation of rainfall (SD of the precipitation distribution).

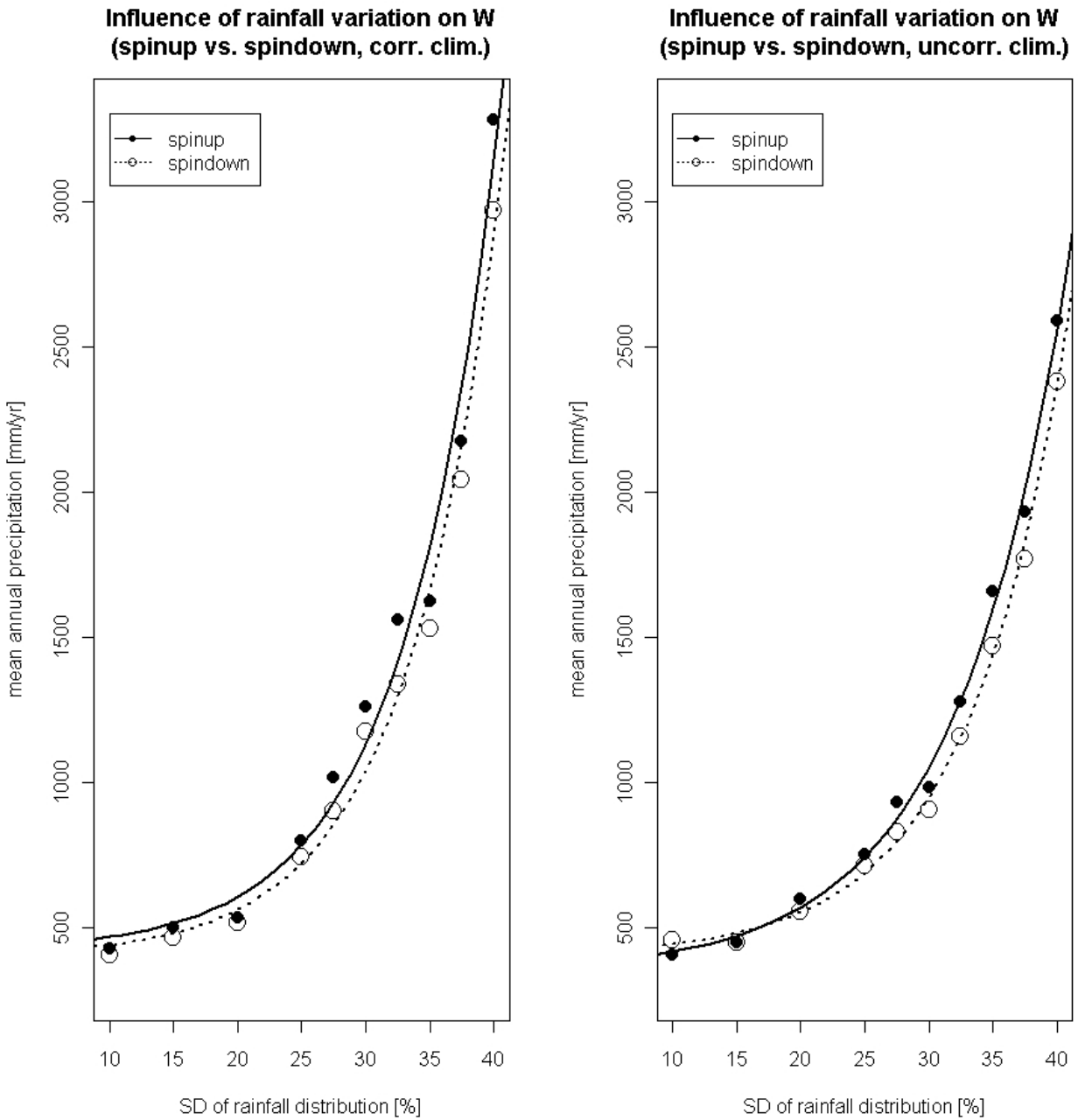


Figure 6.30: Hysteresis: only minor difference between spinup (solid line) and spindown (dashed line). Exponential regression of mean width of transitions phase (symbols) in mm/yr with increasing year-to-year variation of rainfall (SD of the precipitation distribution).

## 7 Discussion

In the course of this work, concepts and methods to assess the stability of tropical ecosystems have been developed. These included the adaption and validation of the weather generator MarkSim for sites in Gabon, where the computation of incident solar radiation and maximum temperature underwent an improper procedure and hence created biased results. After the correction and validation using data from three weather stations, MarkSim's functionality was enhanced to the effect that we could alter and define climatic parameters in a quantified way. These parameters included the amount of annual rainfall, the distribution of precipitation within the year (which connects to the length of the dry season), the quality of the cloud cover (cumuliform vs. stratiform) and the exact year-to-year variation of annual rainfall. Utilizing the ecosystem model Biome-BGC, repeated simulations for each of the 900 climates we generated, made it possible to put a statistical statement on the stability of an ecosystem based on a certain climate. We integrated these data using a logistic regression function, and used the function's point of inflection combined with the width of the transition phase (or risky phase) as a surrogate for instability. Finally we defined two ways for a simulated ecosystem to approach a stable state: The first option refers to forests that grow "from scratch" ("spinup"), the second method mimics stationary forests that have already accumulated biomass and that undergo climatic change ("spindown"). Among many results, where some require further in depth research, three points clearly stand out:

- 1) Variation of annual rainfall has proved to be the key factor influencing the stability of modeled ecosystems, which is, as climate change is on everyones lips, an interesting result. The quality of the exponential regression fitted to the data can be valued as very high, underlined by high coefficients of determination (correlation coefficients) and rather small RMSEs, which allows accurate predictions.
- 2) We observed a significant influence of the length of the dry season on the carbon level of living biomass, in a way that systems based on longer dry periods exhibit lower carbon content. Can this reduction in C be interpreted as forest break down, observed by many authors (chapter 1.1) around 2500BP? Maley, 2001, states that in some regions the climatic event that has taken place around 2500BP has not expressed itself in forest break down, but in a visible change in vegetation composition, that is, a significant advance of pioneer species. In a breath, Maley suggests that the amount of annual rainfall has staid at the same level, while the repartition of rainfall within the year might have changed. This idea results from the fact that sediment layers belonging to this epoch indicate the presence of enhanced erosive events, as they would occur if all the rain was falling within a shorter period of the year due to an extended dry season. Biome-BGC, as mentioned before, lacks the possibility to support alternative forms of vegetation within one simulation, so it cannot switch to a more successful plant type, if one type fails to adapt adequately. It is very likely

that the reduction in biomass of the modeled old grown forest due to a shift in seasonality might in a real forest ecosystem be accompanied by the propagation of pioneer species, as documented by Maley, 2001.

- 3) The third interesting result is, that we could to some extent observe hysteresis-effects, i.e. systems developing “from scratch” showed a significant difference in stability compared to established systems facing climate change. We showed that, within the lower range of rainfall variation, forests growing in undesirable climatic conditions are stabler than grown forests that undergo a shift in climate. In a real system this effect could be connected to the possibility of adaptation, i.e. a forest developing in certain climatic conditions can adapt due to alternative genetic pathways or by creating niches for other plants, for example, while the old grown forest has already adapted to conditions that are no longer present. The result is remarkable because Biome-BGC does not support adaptation of the eco-physiological-parameters nor does it support alternative forms of vegetation. Although the model is based on rather simple functional implementations, we could still observe an effect of hysteresis.

The concepts and methods developed in the course of this work are applicable to other types of forest ecosystems and can be regarded as an innovative approach to assess the impact of climate on the stability of forest ecosystems.

## 8 Bibliography

- Abbott, P.F.; Tabony, R.C., 1985. The estimation of humidity parameters. *Meteorol. Mag.* 114, 49-56.
- Dalgaard, P., 2008. *Introductory Statistics with R*, 2nd ed.. Springer. *Statistics and Computing*, 364 p.
- Delfiner, P. and J. P. Delhomme, 1975. Optimum interpolation by kriging. *Display and Analysis of Spatial Data*. J. C. Davis and M. J. McCullagh (Ed.). New York: John Wiley, 96-114.
- Donatelli, M. and Campbell, G.S., . A simple model to estimate global solar radiation. PANDA Project, Subproject 1, Series 1, Paper 26. *International Symposium on Computational Intelligence (ISCI)*, Bologna, 3 pp.
- Durban M. and Glasbey C.A., 2001. Weather modeling using a multivariate latent Gaussian model. *Agricultural and Forest Meteorology*. 109, 187-201.
- Elenga H., Schwartz D., Vincens A., 1994. Pollen evidence of late Quaternary vegetation and inferred climate changes in Congo. *Palaeogeography, Palaeoclimatology, Palaeoecology* - Elsevier. 109, 345-356.
- Farquhar G.; Von Caemmerer, S.; Berry, J., 1980. A biochemical model of photosynthetic CO<sub>2</sub> fixation in leaves of C<sub>3</sub> species. *Planta*. 149, 78-90.
- Flohn H., 1983. A climate feedback mechanism involving oceanic upwelling, atmospheric CO<sub>2</sub> and water vapour. *Variations in the global water budget*. F. A. Street-Perrot, M. Beran, and R. Ratcliffe (eds.), D. Reidel Publ. Co., Dordrecht, Holland, 405-418.
- Gautam, S.; Pietsch, S.A., 2011 (submitted). Content: EPC parametrization for Biome-BGC. WCLR-Biome. Boku-University, Vienna.
- Geng, S.; Auburn, J.; Brandstetter, E.; Li, B., 1988. A program to simulate meteorological variables: Documentation for SIMMETEO.. *Agronomy Report*. 204, Univ. of California, Crop Extension, Davis, CA.
- Hasenauer, H.;Merganicova, K.; Petritsch, R.; Pietsch S.A.; Thornton, P.E., 2003. Validating daily climate interpolation over complex terrain in Austria. *Agricultural and Forest Meteorology*. 119, 87-107.
- Hutchinson, M., 1995. Stochastic Space-Time Weather Models from Ground-Based Data. *Agricultural and Forest Meteorology*. 73, 237–265.
- Jones, P.G. and Thornton, P.K., 1993. A rainfall generator for agricultural applications in the tropics.

Agricultural and Forest Meteorology. 63, 1-19.

Jones, P.G. and Thornton, P.K., 2000. MarkSim: software to generate daily weather data for Latin America and Africa . *Agronomy Journal*. 92, 445-453.

Jones, P.G.; Thornton, P.K., 1997. Spatial and temporal variability of rainfall related to a third-order Markov model. *Agricultural and Forest Meteorology*. 86, 127-138.

Kimball, J.S.; Running, S.W.; Nemani, R.R., 1997. An improved methods for estimating surface humidity from daily minimum temperature. *Agric. Forest Meteorol.* 85, 87-98.

Krige, D.G., 1951. A statistical approach to some mine valuation problems at the Witwatersrand. *J. Chem. Metall. Mining Soc. South Afr.* 52, 119-138.

Leal, M.E., 2004. The African rain forest during the Last Glacial Maximum: an archipelago of forests in a sea of grass. PhD thesis Wageningen University, Wageningen. , .

Levenberg, K., 1944. A Method for the Solution of Certain Non-Linear Problems in Least Squares. *The Quarterly of Applied Mathematics*. 2, 164–168.

Maley J., 1997. Middle to late Holocene changes in tropical Africa and other continents : paleomonsoon and sea surface temperature variations. Berlin : Springer. Nüzhet Dalfes H. (ed.), Kukla G. (ed.), Weiss H. (ed.) *Third millennium BC climate change and old world collapse*, 611-639.

Maley J., 2001. La destruction catastrophique des forêts d'Afrique centrale survenue il y a environ 2500 ans exerce encore une influence majeure sur la répartition actuelle des formations végétales. *Syst. Geogr. Pl.* 71, 777-796.

Maley J., Brenac P., 1998a. Vegetation dynamics, paleoenvironments and climatic changes in the forest of west Cameroon during the last 28000 years. *Rev. Palaeobot. & Polyno.* 99, 157-188.

Maloba Makanga, J.D., 2010. Les précipitations au Gabon: climatologie analytique en Afrique. L'Harmattan. Paris, .

Mann, H. and Whitney, D., 1947. On a test of whether one of two random variables is stochastically larger than the other. *Annals of mathematical Statistics*. 18, 50-60.

Marquardt, D., 1963. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM Journal on Applied Mathematics*. 11 (2), 431–441.

Ngomanda A., 2005. Dynamique des écosystèmes forestiers du Gabon au cours des cinq derniers millénaires. PhD thesis, Université Montpellier II, Montpellier. , .

Ngomanda, A.; Neumann, K.; Schweizer, A.; Maley, J., 2009. Seasonality change and the third millenium BP rainforest crisis in southern Cameroon (Central Africa). *Quaternary Research*. 71,

307-318.

Pickering, N.B.; Hansen, J.W.; Wells, C.M.; Chan, V.K.; Godwin, D.C., 1994. WeatherMan: A utility for managing and generating daily weather data. *Agron. J.* . 86, 332-337.

Pietsch S.A., Tanga J.J., Ngok-Banak L. , 2009. The Birougou Mountains: Forested throughout the Holocene . European Geoscience Union (Eds.), *Geophysical Research Abstracts*. Vol. 11, EGU 2009-12550.

Pietsch, S.A; Hasenauer, H., 2006. Evaluating the self-initialization procedure for large-scale ecosystem models. *Global Change Biology*. 12, 1658–1669.

Pietsch, S.A; Hasenauer, H.; Kucera J.; Cermak J., 2003. Modeling effects of hydrological changes on the carbon and nitrogen balance of oak in floodplains. *Tree Physiology*. 23, 735-746.

Pietsch, S.A.; Hasenauer, H., 2002. Using mechanistic modeling within forest ecosystem restoration. *For. Ecol. Manage.*. 159, 111-131.

Price, D.T.; McKenney, D.W.; Nalder, I.A.; Hutchinson, M.F.; Kesteven, J.L., 2000. A comparison of two statistical methods for spatial interpolation of Canadian monthly mean climate data. *Agricultural and Forest Meteorology*. 101, 81-94.

Rauscher, H.M., 1986. The Microcomputer Scientific Software Series 4: Testing Prediction Accuracy. US Department of Agriculture. General Technical Report, NC-107.

Reynaud-Farrera I., Maley J., Wirrmann D., 1996. Végétation et climat dans les forêts du sud-est Cameroun depuis 4770 ans B.P.: analyse pollinique des sédiments du Lac Ossa. *C.R. Acad. Sci.*. 2a 322, 749-755.

Reynolds, M.R. Jr., 1984. Estimating the Error in Model Predictions. *Forest Sci.*. Vol. 30, No. 2, 454-469.

Richardson C.W., . Weather simulations for crop management models. *Trans. ASAE*. 28(5), 1602-1606.

Richardson, C.W., 1981. Stochastic simulation of daily precipitation, temperature and solar radiation. *Water Resource Research*. 17 (1), 182-190.

Ryan, M.G., 1991. Effects of climate change on plant respiration. *Ecol. Appl.*. 1, 157-167.

Shapiro, S.S; Wilk, M.B, 1965. An analysis of variance test for normality (for complete samples). *Biometrika*. Vol. 52, Nr. 3-4, 591-611.

Soltani, A., Latifi, N., Nasiri, M., 2000. Evaluating of WGEN for generating long term weather data for crop simulations. *Agricultural and Forest Meteorology*. 102, 1-12.

Strandman, H., Väisämä, H., Kellomäki, S., 1993. A procedure for generating synthetic weather

records in conjunction of climatic scenario for modeling of ecological impacts of changing climate in boreal conditions. *Ecological Modeling*. 70, 195-200.

Thornton, P.E., 1998. Description of a numerical simulation model for predicting the dynamics of energy, water, carbon and nitrogen in a terrestrial ecosystem. PhD thesis, University of Montana, Missoula. MT, 280 pp..

Thornton, P.E.; Law, B.E.; Gholz, H.L; Clark, K.L.; Falge, E.; Ellsworth, D.S.; Goldstein, A.H.; Monson, R.K.; Hollinger, D.; Falk, M.; Chen, J.; Sparks J.P., . Modeling and measuring the effects of disturbance history and climate on carbon and water budgets in evergreen needleleaf forests. *Agricultural and Forest Meteorology*. 113, 185-222.

Thornton, P.E.; Running, S.W.; White, M.A., 1997. Generating surfaces of daily meteorological variables over large regions of complex terrain. *J. Hydrol.* 190, 214-251.

Turkey, J.W., 1977. Box-and-Whiskers Plots. *Exploratory Data Analysis*. Reading, MA: Addison-Wesley, 39-43.

Wilcoxon, F., 1945. Individual Comparisons by Ranking Methods. *Biometrics Bulletin*. 1, 80-83.

Zogning A., Giresse P., Maley J., Gadel F., 1997. The late Holocene paleoenvironment in the Lake Njupi area, west Cameroon: implications regarding the history of Lake Nyos. *J. African Earth Sc.* 24, 285-300.

FAO, Food and Agriculture Organization of the United Nations:  
<http://www.fao.org/forestry/country/18310/en/gab/>; accessed: March 17, 2011

Joint Research Center, European Commission, Global Environment Monitoring Unit:  
<http://bioval.jrc.ec.europa.eu/products/glc2000/publications.php>; accessed: August 29, 2011



## APPENDIX A - MarkSim file structure

For our needs, the two most important file types are the climate parameter file (or CLX file) and the daily weather output in the WTG-file (figures A.1 and A.2).

```

moanda Interpolated -1.533 13.267 571
1.000 0.190 0.119 0.161 0.197-0.024-0.174 0.053-0.163-0.244-0.018 0.345
0.190 1.000 0.101 0.255 0.303 0.247 0.221 0.065 0.011-0.016-0.088 0.462
0.119 0.101 1.000 0.332 0.218-0.115 0.073 0.239-0.065 0.066 0.140 0.154
0.161 0.255 0.332 1.000 0.335-0.013-0.136-0.203-0.136-0.164 0.082 0.210
0.197 0.303 0.218 0.335 1.000 0.298-0.044-0.050-0.163-0.138-0.038 0.239
-0.024 0.247-0.115-0.013 0.298 1.000-0.066-0.276 0.154-0.011 0.061 0.162
-0.174 0.221 0.073-0.136-0.044-0.066 1.000 0.242 0.296 0.303-0.077-0.018
0.053 0.065 0.239-0.203-0.050-0.276 0.242 1.000 0.114 0.046 0.289 0.148
-0.163 0.011-0.065-0.136-0.163 0.154 0.296 0.114 1.000 0.306 0.184 0.037
-0.244-0.016 0.066-0.164-0.138-0.011 0.303 0.046 0.306 1.000 0.101 0.193
-0.018-0.088 0.140 0.082-0.038 0.061-0.077 0.289 0.184 0.101 1.000-0.017
0.345 0.462 0.154 0.210 0.239 0.162-0.018 0.148 0.037 0.193-0.017 1.000
MONTH AV P BETA RAINDAYS S.E.
1 8.9 0.693 -0.173 0.619 0.35800
2 9.6 0.673 0.014 0.722 0.35674
3 9.6 0.653 0.566 0.900 0.36151
4 8.8 0.647 0.236 0.800 0.36364
5 9.9 0.677 -0.016 0.697 0.35118
6 6.0 0.798 -1.265 0.154 0.41392
7 5.6 0.821 -1.384 0.091 0.40734
8 5.1 0.883 -1.449 0.083 0.42337
9 8.1 0.743 -0.641 0.402 0.36036
10 10.7 0.770 0.417 0.849 0.36281
11 10.4 0.763 0.285 0.825 0.36601
12 10.1 0.719 -0.133 0.646 0.35387
D1-3 0.4920 0.1880 0.1200 N= 2 Cluster 372 Phase 0.082
rain 174. 195. 261. 218. 204. 37. 11. 16. 100. 277. 262. 199.
temp 24.2 24.5 24.8 24.9 24.0 23.0 22.2 22.7 23.7 23.9 23.8 23.5
rang 8.9 9.5 9.8 9.7 8.8 8.0 8.1 8.5 9.1 9.1 9.1 8.6
radn 18.6 20.2 21.0 20.4 18.2 15.8 14.6 16.0 18.6 19.5 19.3 17.5

```

Figure A.1: Example of a CLX file (Moanda, Gabon), monthly climate normals (bold)

The CLX file contains latitude, longitude and elevation of a point of interest in the first line, followed by a correlation matrix for the baseline probits and a list of monthly parameters responsible for the Markov Chain rainfall estimation. Printed bold are the computed climate normals of rain [mm/month], average temperature (temp, [°C]), diurnal temperature range (rang, [°C]), and radiation (radn, [MJ/m<sup>2</sup>/day]) in monthly resolution covering the whole course of a year. Applying procedures mentioned in chapter 3.1.2, this file gives rise to the WTG output files, one single WTG file covering one year's constitutes. It is possible to generate a maximum of 99 simulation years, each year's daily weather variables represented in one WTG file. It is, however, possible to define your own climate normals using the DAT-file, which includes only longitude, latitude, elevation and the climate normals for monthly mean temperature, temperature range and precipitation. MarkSim

uses the DAT-file as a template to create a CLX file and to compute the stochastic output in the next step.

To generate multiple climates from different DAT-files at the same time, MarkSim provides CBF and XBF files where input DAT-files and output CLX-files, or input CLX-files and output WTG-files, respectively, are listed in the order that climates should be produced. These files (CBF and XBF) can be generated using MarkSim's drag and drop function.

```
*WEATHER : moan From Interpolated Surfaces
@ INSI      LAT      LONG  ELEV  TAV   AMP REFHT WNDHT
  moan     -1.533   13.267  571  23.8   9.8 -99.0 -99.0

@DATE  SRAD  TMAX  TMIN  RAIN
01001  20.6  26.6  18.7  36.4
01002  25.1  27.9  17.6   0.0
01003  20.7  29.8  18.1   0.0
01004  27.8  28.1  20.2   0.0
01005  26.4  28.5  21.1   0.0
01006  22.8  25.6  18.5   0.0
01007  24.8  26.3  18.2  11.8
...

```

Figure A.2: Extract of the header and first seven entries of a WTG file (Moanda, Gabon)

## APPENDIX B - Statistical methods

### B.1 Test for normality

In order to perform a t-test, or conventional confidence and prediction interval, the precondition is a normally distributed dataset. All data will therefore be tested for normality using the Shapiro-Wilk-test as described in Shapiro and Wilk, 1965. Tabulated p values < 0.05 lead to the rejection of the null hypothesis, that the tested data set is distributed normally.

### B.2 Testing predicted vs. observed

Either two data sets are compared using a paired test, or the residuals (i.e. observed minus predicted) of the two sets are tested using a one sample test – which yields the same exact result. Assuming a one sample test, data is tested against the null hypothesis that the mean bias (mean of residuals) is equal to zero. Again,  $p < 0.05$  leads to rejection of the null hypothesis, and the underlying data set will be considered “biased at a 5%-level of significance”. Normally distributed data sets are tested by the conventional t-test, while data failing the precondition to be normally distributed undergo a non-parametric test, the U-test (Wilcoxon, 1945; Mann and Whitney 1947).

### Confidence (CI) and prediction-interval (PI) for residuals

The precondition to apply statistics as suggested by Reynolds (Reynolds, 1984) is a normally distributed underlying data set. Prediction and confidence interval are defined as:

$$CI: \left[ D - \frac{S \cdot t_{1-\alpha/2}(n-1)}{\sqrt{n-1}}; D + \frac{S \cdot t_{1-\alpha/2}(n-1)}{\sqrt{n-1}} \right] \quad \text{eq. B.1}$$

$$PI: \left[ D - \sqrt{1 + \frac{1}{n}} \cdot S \cdot t_{1-\alpha/2}(n-1); D + \sqrt{1 + \frac{1}{n}} \cdot S \cdot t_{1-\alpha/2}(n-1) \right] \quad \text{eq. B.2}$$

where  $n$  are the degrees of freedom,  $D$  (eq. B.3) is the mean residual or the estimator for the bias, and  $S$  its standard deviation (eq. B.4):

$$D = \frac{1}{n} \cdot \sum_i^n D_i \quad \text{eq. B.3}$$

$$S^2 = \frac{1}{n-1} \cdot \sum_i^n (D_i - D)^2 \quad \text{eq. B.4}$$

$t_{1-\alpha/2}(n-1)$  is the  $1-\alpha/2$  quantile of the t distribution with  $n-1$  degrees of freedom.

The interpretation of the prediction interval in the case of the correction procedure introduced in APPEDNIX C is, that if MarkSim is repeatedly used to produce new stochastic weather data with the same underlying climate normals, the probability that a future value of  $D$  will fall in this interval

is  $1-\alpha$ . Each sample leading to a specific D value and its according PI must therefore be selected at random from the distribution of D.

On the other hand, the confidence interval gives an idea of the scale of discrepancies between D and E(D), where E(D) can be thought of as the expected value of the bias over all possible sets of generated weather data from one site which use the same underlying climate normals. In the usual terminology, assuming that  $\alpha = 0.05$ , we can be 90% confident, that the value which includes 95% of absolute errors  $|D|$  is located somewhere within the CI. Testing residuals against the null hypothesis of being equal to zero, if the confidence interval includes 0 the tested data is referred to as “unbiased”, and as “biased” if it does not.

If, however, the underlying data set is not normally distributed, the upper and lower 10% of the distribution of D are truncated, as suggested in Rauscher, 1986. The trimmed bias  $D^T$  and the trimmed standard deviation are then used to compute PI and CI.

### **Pearson's linear correlation coefficient & sample mean error**

The Pearson correlation (eq. B.5), or product momentum correlation coefficient gives an idea of how well two samples are correlated in a linear way and is within this work often taken as a surrogate for the quality of seasonality. Consider the comparison of weather data: If both samples (“observed” and “predicted”) experience higher precipitation in one part of the year, and lower values in a different time period, then r will be positive and close to one. On the other hand contrasting seasonal trends lead to negative r, close to -1. This, however, does not tell us how far two samples are apart from each other, and we therefore define the sample's mean error (eq. B.6).

$$r = \frac{\sum_i^n [(y_i - \hat{y}) \cdot (x_i - \hat{x})]}{\sqrt{\sum_i^n (y_i - \hat{y})^2 \sum_i^n (x_i - \hat{x})^2}} \quad \text{eq. B.5}$$

$$err = \frac{1}{n} \sum_i^n |x_i - y_i| \quad \text{eq. B.6}$$

$r = -1$  means complete negative correlation of the two samples, i.e. their seasonal trends contrast perfectly.  $r = 1$  on the other hand tells you that both samples feature the same seasonality. If  $r = 0$  the two samples are not correlated linearly. The subscripted variables are monthly averages of the two samples (e.g.: climate normal vs. actual value), x and y are their mean values, i.e. annual means. Note, that performing regression analysis of any kind,  $r^2$  is usually used as a measure to quote the quality of fitted data.

### B.3 Box-and-whiskers plot

A box-and-whiskers plot is, similar to a histogram, a graphical representation of a distribution, making use of the quantiles of the distribution in question. The line in the middle presents the median (or 50%-quartile, Q2), the two boxes to the left and right (or below and above the middle line) provide information about the size of the 25% (Q1) and 75%-quartile (Q3), respectively. The lines (whiskers) indicate a range for all points of the distribution that are within 1.5 times the interquartile range (Q3-Q1), all other points are referred to as “outliers”, and are displayed by a dot outside the whiskers. (Turkey, 1977; Dalgaard, 2008)

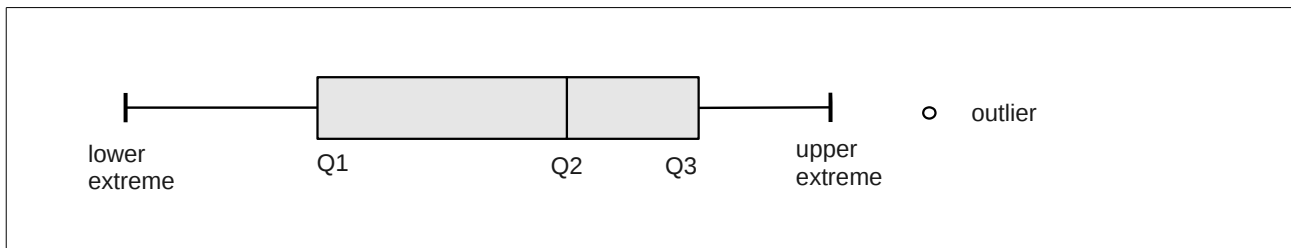


Figure B.1: Box-and-Whisker plot, schematically

### B.4 Logistic regression analysis

Later in this work techniques to assess ecosystem stability will be presented, and therefore simulation results will be categorized into two groups: “dead” and “alive”, telling us whether a simulation has successfully developed a “living” ecosystem or whether it has failed to do so. In common sense logistic regression analysis (LRA) refers to situations where a dichotomous variable is explained by the linear combination of a set of predictors and errors, which is an extension of multiple linear regression analysis. A dichotomous set can be divided into two parts, where every element must belong to either one part, or the other, but none can simultaneously belong to both parts (Dalgaard, 2008). An example is given by the system states “dead” or “alive”, as present in our approach, but here we only have one discrete predictor, which is mean annual precipitation. We will adopt LRA techniques, not to distinguish influential predictors from less important predictors, which is the usual purpose, but to quantify the transition between the two states (from low rainfall regimes which face breakdown more often, to more humid climates, which should lead to full development with a higher probability).

The logistic regression function with only one predictor is of the form:

$$P(x) = \frac{1}{1 + e^{-(a+b \cdot x)}} \quad \text{eq. B.7}$$

P is restricted to [0,1] and can be interpreted as the probability that an ecosystem simulation at a given mean annual rainfall, with an according number of dry months and variation of annual rainfall

does develop into stationary state.  $x$  is the only predictor (precipitation) and  $a$  and  $b$  are regression coefficients, intercept and slope, respectively. The regression coefficients are estimated by a maximum likelihood procedure (Dalgaard, 2008).

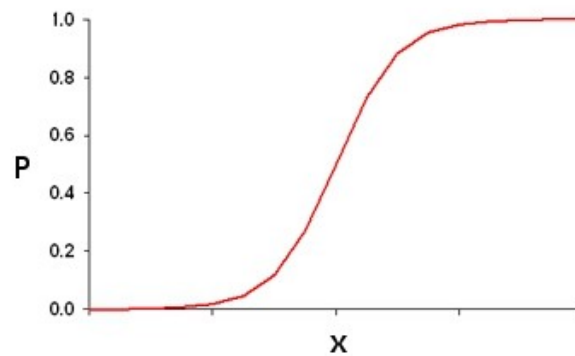


Figure B.2: Logistic regression function, schematically.  
Probability ( $P$ ), predictor ( $x$ ).

Using the logistic regression function data can be integrated and reduced to two measures, which can be interpreted as surrogates for instability:

- 1) The  $x$ -value of the function's point of inflection [mm/yr] at  $P = 0.50$ .
- 2) The width of the transition phase, from  $P = 0.01$  to  $P = 0.99$ .

## **APPENDIX C - Correction of solar radiation and maximum temperature**

Even though generating weather is based on a stochastic procedure, the monthly average of climate variables should not be too far off the original climate normals. This statement is true for many locations worldwide, but not for sites within Gabon, where we face constant overestimation of solar radiation (SRAD) and maximum temperature (TMAX) within the long dry season during boreal summer, and underestimation of the same variables during the rest of the year. Based on this fact a correction procedure has been developed, which fits the monthly averages computed from a set of WTG files to the climate normals from the CLX file (figure A.1), assuming that the interpolated climate normals are close to the real conditions. A correction is performed on SRAD and TMAX only.

The different working steps of the correction algorithm can be summarized as follows:

1. One CLX and its corresponding number of N WTG files (= N simulation years) are read into the system and daily, as well as monthly average values over N simulation years are computed.
2. A cumulative approach determines the boundaries of each year's long dry season (DS). According to the FAO (source: FAO) Gabon features a long dry season, taking place between June and October, during which the sky stays covered.
3. Monthly mean WTG variables are compared with CLX climate normals and the monthly residuals between those two are estimated. Further, correction factors (CF) are calculated that define how values of a single day are changed.
4. Residuals are computed again within each loop if the main cycle (CLX-WTG)
5. Depending on the CF and the size and sign of the residuals, a probability function is defined that is responsible for the amount of days within a month to be changed, so that the mean of the whole month reaches a different desired value.
6. A stochastic correction function chooses days according to its probabilities where daily SRAD and TMAX are adjusted by their according CFs. The results cannot exceed certain truncation factors TF to prevent from generating extremely high or low values of the given variable.
7. Optionally changes on the dry season only can be performed using the defined boundaries of each year's DS.
8. Some statistics on the correction, corrective factors and daily to monthly values before and after correction are printed to different files.

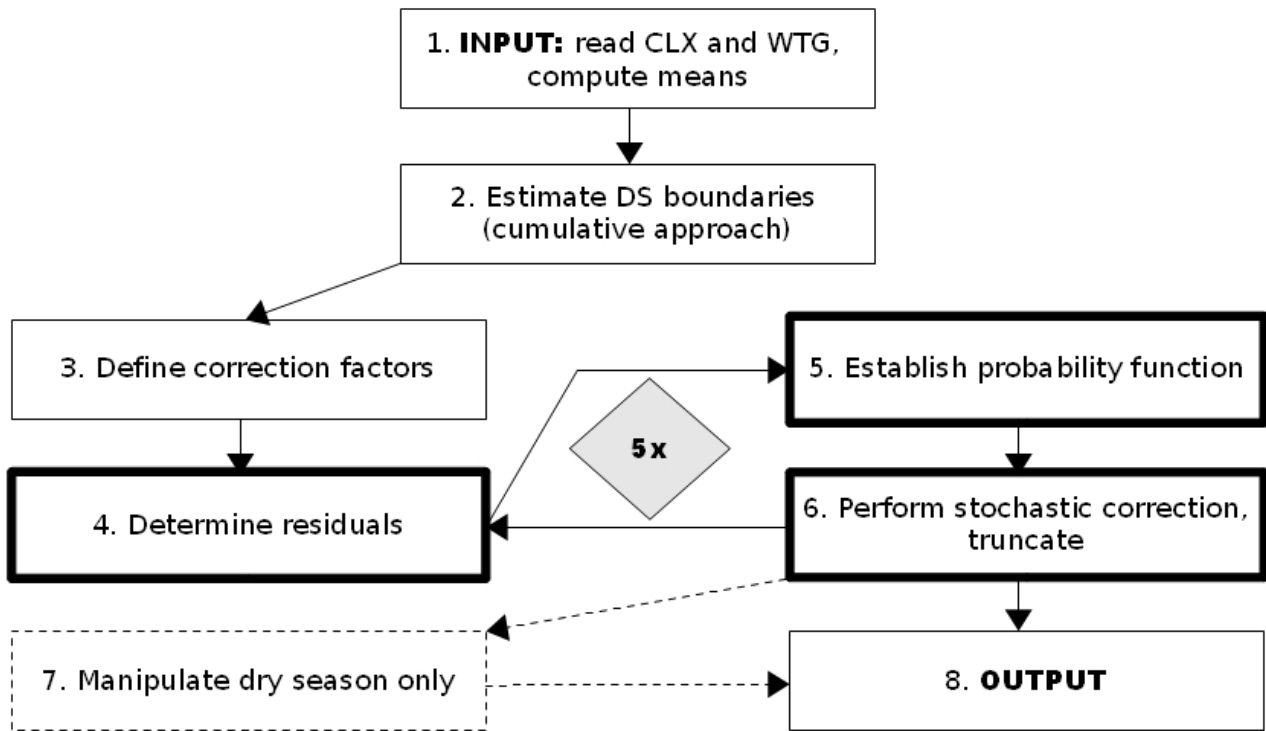


Figure C.1: Flow chart of corrective procedure, main cycle indicated by highlighted boxes.

In the following section the focus will first be put on the main cycle of the algorithm which comprises step 4, 5 and 6 (highlighted in the flow chart). After that, step 2 and 7 will be explained since 7 explicitly depends on 2.

### C.1 The main cycle

To apply a correction we can distinguish between two approaches: Either a correction is performed on a monthly, or on a daily basis.

1. The benefit of the monthly corrective approach is, that all days in a month in question are treated equally and the initial sequence of daily values produced by MarkSim stays more or less undisturbed since each daily value is multiplied by the same correction factor that leads to an either decreased or increased monthly mean value.
2. A more invasive method applies a correction on a daily basis, i.e. single selected days within a month are multiplied by a corrective factor with the result that the monthly mean is changed. The advantage of this second approach is higher flexibility: A bigger range of options to try out different CFs, and the possibility to actively select days that should undergo a correction.

All that will become more clear after we have discussed how that algorithm has been implemented in our case: First, correction factors that tell us how a certain daily value is changed, are computed. These factors can be chosen on a monthly scale, or to be valid for the whole year and they work in



both directions, i.e. there are factors slightly below one, and slightly above one, depending on whether the initial value of the meteorological variable is too high or too low, respectively. The size of a CF and the residual between climate normal and WTG average, however, define a certain probability function (PF) that is responsible for the amount of days within a month to be changed, so that the mean of the whole month reaches a different desired value. According to the PF each day has a certain probability ( $0 < p < 1$ ) to undergo a corrective procedure, where  $p$  is compared with a random number ( $0 < r < 1$ ). Now we can again emphasize the advantages of that method:

- a) The size of the CF defines the shape of the PF dynamically. If we move the CFs closer towards 1, the PF becomes higher and converges towards the case, where all days are corrected. So approach 2 in fact includes method 1, if CFs are chosen in the right way.
- b) In the present implementation probabilities are distributed equally over each month, i.e. all days within a certain month have the same chance to be “selected” by the random number generator. The PF, however, can be changed or extended easily to apply certain selective patterns. If, for example, we find out that rainy days show a too large amount of incident solar radiation in general, we can adjust the PF with the result that these days are more likely to be corrected in a specific way.

Since the correction is based on the stochastic selection of days that are adjusted, and afterwards truncated if they exceed certain defined values, the residuals will not completely vanish after one iteration loop of the main cycle. Further, we expect a certain saturation effect taking place, were the difference between desired and actual values will not get any smaller due to stochasticity and truncation. As we will discuss later, after five iterations the procedure has usually reduced the error to its minimum. The monthly probabilities, however, will get smaller with every iteration performed since they directly depend on the difference between desired and actual value of a variable in question (TMAX or SRAD), i.e. the residual.

## **Correction Factors (CF) and Truncating Factors (TF)**

As already mentioned, the choice of the corrective factors depends on the desired output and can be adapted according to the end user needs. Nevertheless a set of CFs and TFs, which is implemented in the current version is suggested here:

SRAD\_MAX +  $\epsilon$ : Mean over all N simulations of the maximum values of daily solar radiation, each year (or each WTG) contributes with its highest SRAD value. Fitting daily solar radiation, this value cannot be exceeded.

SRAD\_MIN +  $\epsilon$ : Mean over all N simulations of the minimum values of daily solar radiation, each simulation year contributes with its lowest SRAD value. Fitting daily solar radiation, the corrected value cannot fall below SRAD\_MIN +  $\epsilon$ .

TMAX\_MAX +  $\epsilon$ : Similar to SRAD\_MAX, but using the maximum values of TMAX.

TDIF\_MIN +  $\epsilon$ : Average minimum diurnal temperature range, the smallest TDIF (=TMAX-TMIN) of every simulation year contributes to this TF. A newly generated daily TMAX value must not be lower than its corresponding TMIN value plus TDIF\_MIN+  $\epsilon$ .

SRAD\_CORR\_H: The ratio between the mean solar radiation on wet versus dry days, averaging over the whole year and all WTGs. SRAD\_CORR\_H < 1 is the CF correcting the days that experience a too high value of SRAD in their monthly mean.

SRAD\_CORR\_L: The ratio SRAD\_MAX/SRAD, where SRAD\_MAX has been explained before and SRAD is the average solar radiation of all simulation years. SRAD\_CORR\_L > 1 is responsible for corrections on days where SRAD is too low in its monthly mean.

TMAX\_CORR\_H is equivalent to the relation TMIN/TMAX < 1, where TMIN is the mean annual minimum and maximum temperature, respectively, averaged over all simulation years.

TMAX\_CORR\_L is chosen similar to SRAD\_CORR\_L: TMAX\_MAX/TMAX > 1.

For simplicity only SRAD\_CORR\_H and SRAD\_CORR\_L have to be chosen explicitly. These factors determine the probability function, and since it makes sense to perform a correction of daily TMAX on the same days as SRAD is changed, TMAX\_CORR is computed depending on the CFs of solar radiation (explanation follows). TMAX\_CORR\_H and TMAX\_CORR\_L simply restrict TMAX\_CORR to the interval TMAX\_CORR\_H < TMAX\_CORR < TMAX\_CORR\_L. This approach, again, ensures the dynamic behavior of the correction procedure, since the user only has to make up a pair of CFs for the solar radiation, which then define the average amount of days that will have to undergo an adjustment (PF) and which finally leads to the definition of the CFs for the maximum temperature.  $\epsilon$  is a normally distributed random error, derived from the samples standard deviation. It is computed as follows:

$$\epsilon = \hat{m} + x \cdot \sigma \quad \text{eq. C.1}$$

$$x = \rho \cdot \cos(\phi) \quad \text{eq. C.2}$$

$$\rho = \sqrt{-2 \cdot \log(1-b)} \quad \text{eq. C.3}$$

$$\phi = 2 \cdot \pi \cdot c \quad \text{eq. C.4}$$

b and c are random numbers between zero and one, m and  $\sigma$  are the sample's mean and standard deviation, respectively.  $\epsilon$  is truncated if it leaves the interval [-2 $\sigma$ , 2 $\sigma$ ].

## The probability function (PF)

The PF in the recent implementation allocates the same probability p (0 < p < 1) to all days of a certain month:

$$days[i] = \frac{srad_{CLX}[i] - srad_{WTG}[i]}{srad_{WTG}[i]} \cdot \frac{dpm[i]}{CF - 1} \quad \text{eq. C.5}$$

$$PF : p[i] = \frac{srad_{CLX}[i] - srad_{WTG}[i]}{srad_{WTG}[i]} \cdot \frac{1}{CF - 1} \quad \text{eq. C.6}$$

days[i] is the amount of days within a month with a total number of dpm[i] days, where a correction has to be performed in order to meet the desired monthly mean value for SRAD.  $srad_{CLX}[i]$  is the monthly climate normal, or desired monthly mean value, and  $srad_{WTG}[i]$  refers to the actual mean value that has to be adjusted.  $srad_{CLX}[i] - srad_{WTG}[i]$  therefore define the residual that has to be reduced during the correction. If  $srad_{WTG}[i] > srad_{CLX}[i]$  :  $CF = SRAD\_CORR\_H$ , and if  $srad_{WTG}[i] < srad_{CLX}[i]$  :  $CF = SRAD\_CORR\_L$ , the index  $i$  runs over all months from 0 to 11 (including 0 as the first month),  $p[i]$  stays constant for all days within one month.

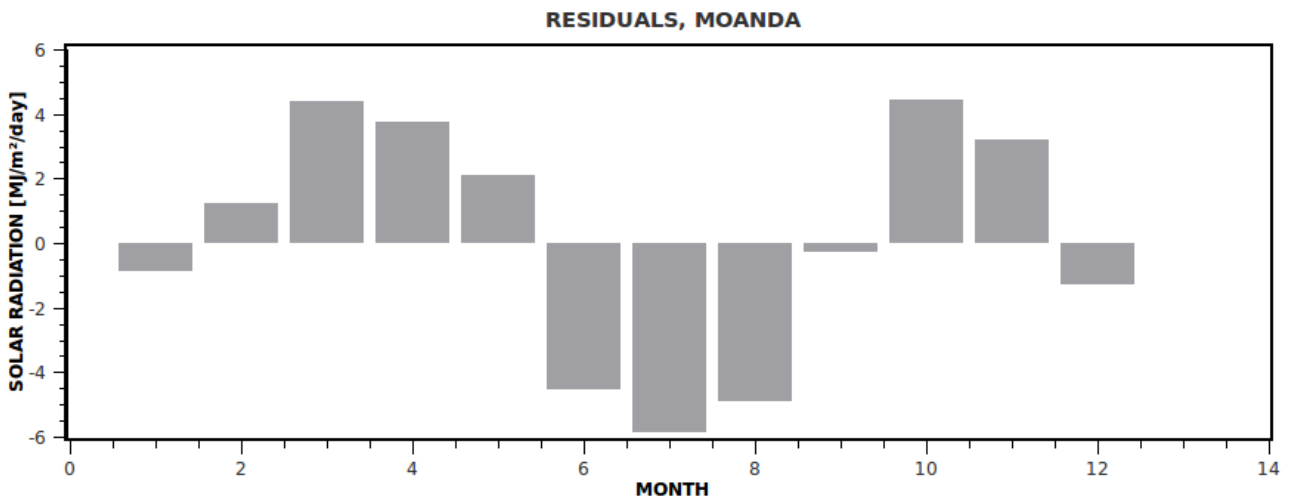


Figure C.2: Residuals of solar radiation (sradCLX-sradWTG; [MJ/m²/day]), exemplary for Moanda .

Figure C.2 presents an example for the residuals of solar radiation for the site Moanda before a correction has been performed and considering no dynamic transition dates (we will come to this soon). Negative values indicate an overestimation while values above zero stand for an underestimation of the desired value. This distribution of the residuals and assuming  $SRAD\_CORR\_H = 0.58$  and  $SRAD\_CORR\_L = 1.67$  leads to the probability distribution in figure C.3

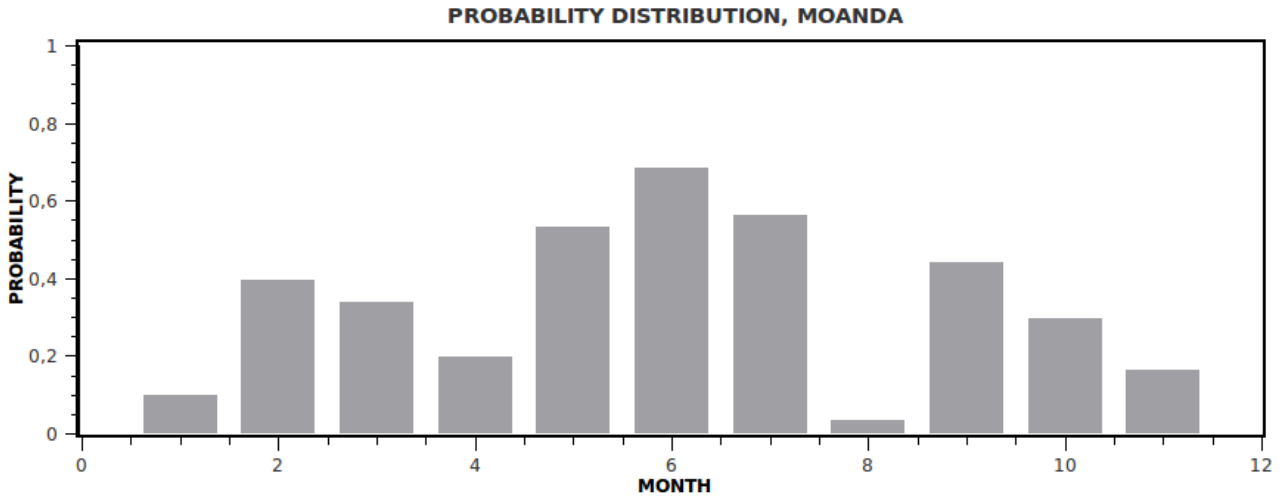


Figure C.3: Monthly correction probabilities ( $0 < p < 1$ ), for Moanda exemplary

As mentioned before, this PF is defined on a monthly scale, i.e. all days within a certain month have the same probability as indicated by the distribution. Theoretically, applying selective patterns, a non-flat PF on a daily basis could be defined, with the side condition that the over all monthly probability stays the same.

A dynamic CF responsible for TMAX adjustments can now be defined as follows:

$$TMAX\_CORR[i] = \frac{dpm[i] \cdot (tmax_{CLX}[i] - tmax_{WTG}[i])}{days[i] \cdot tmax_{WTG}[i]} + 1 \quad \text{eq. C.7}$$

$tmax_{CLX}[i]$  is the climate normal, and  $tmax_{WTG}[i]$  the actual monthly mean value, again  $tmax_{CLX}[i] - tmax_{WTG}[i]$  define the residual between current and desired situation.

To summarize: TMAX\_CORR has a different value from month to month, but the same value for the same month in different simulation years, and explicitly depends on SRAD\_CORR and the SRAD-residuals. These solar radiation CFs are fixed, i.e. they don't change between months, and neither they do so between simulation years.

Finally all necessary elements have been defined to perform the correction:

$$\text{if } rand[i] < p[i] : \quad srad_{new} = srad \cdot CF \quad \text{and} \quad tmax_{new} = tmax \cdot CF \quad \text{eq. C.8}$$

If that condition is met, an adjustment of this certain day according to the month's CF is initialized.  $rand[i]$  is a random number in  $[0, 1]$  which is newly computed for each day in each year. The correction follows the simple relation: new value = old value x CF.

## The probability function – a different representation

We have already emphasized the advantage of the dynamic handling of the probability function as it has been implemented here. Just two correction factors have to be computed, some climate

normals to be defined, the rest is taken care of by the algorithm. In the introduction of this chapter (the main cycle) we have introduced two approaches of how a correction can be applied on the weather data: Either all days in a month in question are treated equally or single days are selected (by chance, or actively). We have stated that in fact method 2 (active selection) includes method 1 (treat all days equally), if the corrective factors are chosen in the right way. We therefore define the CFs that will lead to probability = 1 on every day (i.e. each day will be corrected)

$$CF = \frac{srad_{CLX}[i] - srad_{WTG}[i]}{srad_{WTG}[i]} + 1 \quad \text{eq. C.9}$$

## Dynamic transition dates (DTD)

It has to be emphasized, that even though we adjust single daily values, the focus of the whole correction process is to be put on monthly values. But “one month” in this algorithm is somewhat different from its conventional meaning. Assuming months as we know them, i.e. January, February (...) and their fixed transition dates from one month to another, e.g. January 31 to February 1, certain artifacts have emerged concerning daily mean values. A daily mean value here is the average of all values of a certain variable on that particular day but of all different simulation years (e.g. February 1 in year 1, year 2...). These artifacts, however, express themselves as leaps of that variable on transition dates.

Looking at figure C.4 the problem becomes apparent, especially between August and September, as well as September and October. This happens, because statistically, the whole month is just

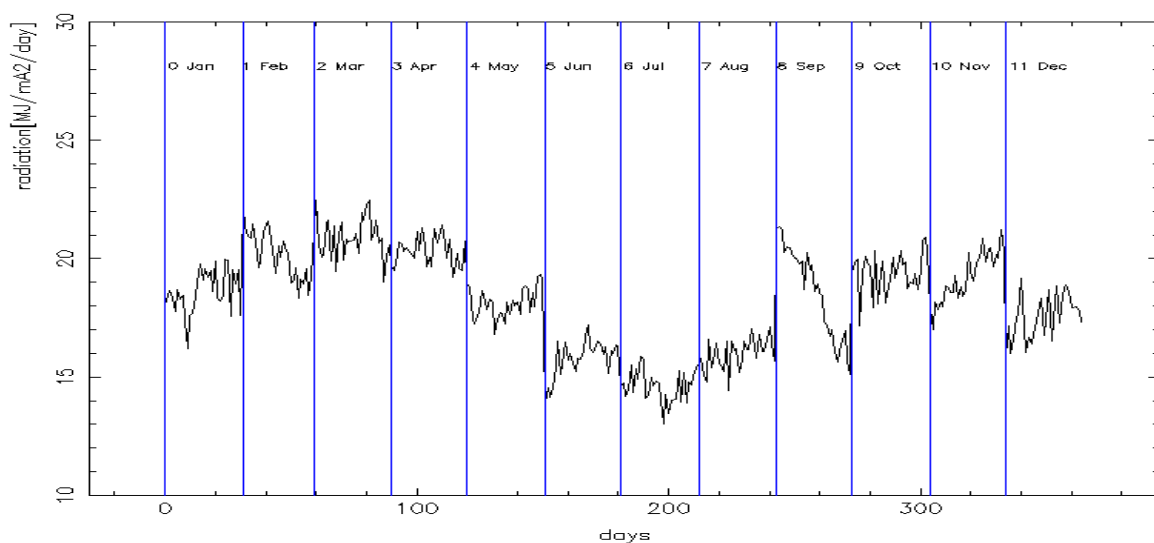


Figure C.4: Daily mean values of solar radiation [MJ/m²/day] averaged over N=99 simulation years, with leaps, for Moanda exemplary.

raised or lowered to a different mean value, each day having the same probability to take part in the corrective procedure. So, the trend within the month is not addressed directly or, roughly spoken, one month is just cut out of the original seasonal course and pasted into the new one, according to its corrected monthly averages, without changing the shape of the curve within the month. This fact, however, does not have any influence on the quality of monthly mean values, but avoiding to take a look at daily mean values can hide important trends that can lead to unpredictable behavior in succeeding applications.

The solution to get rid of those leaps is the introduction of dynamic transition dates. The idea is quite simple: If we avoid to choose the same transition dates in every simulation year, the leaps will hopefully level out and the whole course of one daily mean variable over the year will get smoother. The introduction of new boundaries of each month of course requires the definition of new climate normals, which where assumed to be located on a straight line between its neighboring, original climate normals. The procedure can be described as follows: For a given simulation year, i.e. for one WTG file, the start of the year is shifted for  $s$  days,  $0 < s < 31$ , where  $s$  is generated randomly for each simulation year. The transition dates retain the same distances between each others, but the whole set of transition days is shifted for  $s$  days. The new climate normals are computed as follows:

$$cn_{new}[i] = cn[i] + (cn[i+1] - cn[i]) \cdot \frac{s}{dpm[i]} \quad \text{eq. C.10}$$

$cn[i]$  are the monthly climate normals for SRAD or TMAX,  $dpm[i]$  is the number of days per month,  $i$  indicates the corresponding month and  $s$  is the random start day. The approach works on closed boundary conditions, which means that if the index reaches the end of the year, it starts again at the beginning.

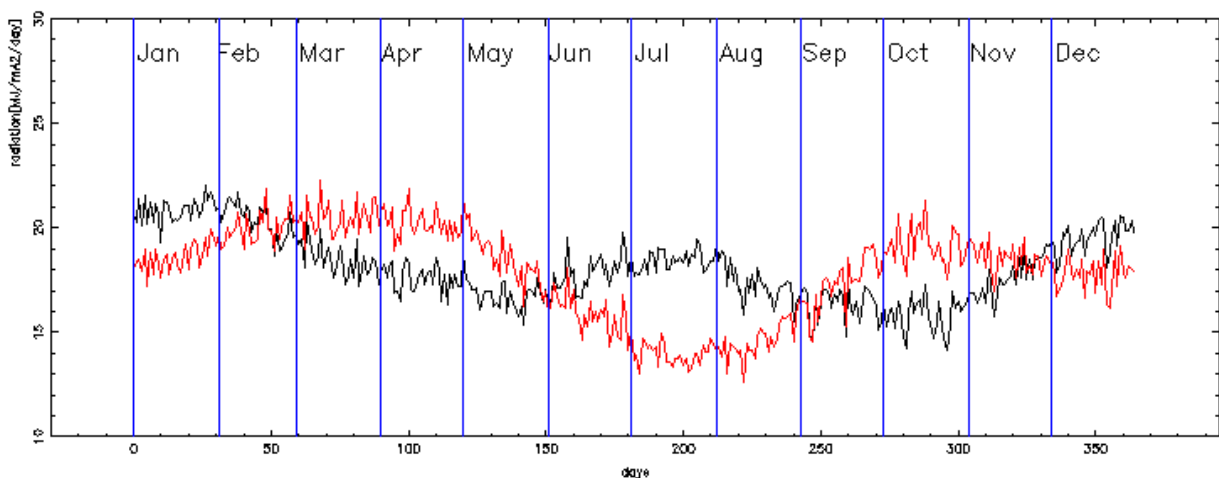


Figure C.5: Daily mean values of solar radiation [MJ/m<sup>2</sup>/day] averaged over N=99 simulation years, smoothed, for Moanda exemplary. Black line = uncorrected, red line = corrected output.

If we now have a look at the seasonal trend of the same data set of 99 years (figure C.5), that has led to the curve in figure C.4 full of leaps but now assuming dynamic transition dates, we see how technically all of the leaps have vanished. The red curve is the corrected, the black one shows the uncorrected data set. The monthly mean values stay unaffected using dynamic instead of static boundaries. Since new (shifted) climate normals have been defined, for each shifted time interval the PF changes too.

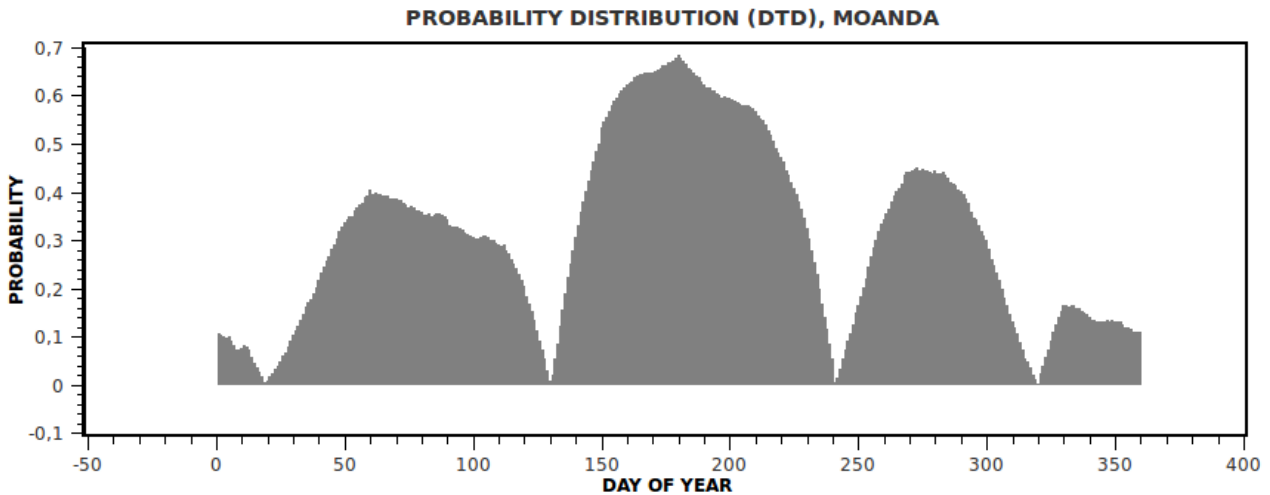


Figure C.6: Shifted correction probabilities ( $0 < p < 1$ ) over the year for a shift of 0 to 30 days.

To produce this distribution (figure C.6), the start day of the year has been iterated from 0 to 30, while the sizes of the time intervals stay the same, e.g. January has 31 days, but the beginning of January in the course of the year has been moved forward for  $0 \leq n \leq 30$  days. The same is valid for the other months too.

After CFs for solar radiation have been computed, a probability function has been established, the corresponding CFs for TMAX have been defined, and the correction has been performed, to finish the main cycle, daily values of TMAX and SRAD are truncated to their allowed range using the previously defined truncating factors TMAX\_MAX, TDIF\_MIN, SRAD\_MAX and SRAD\_MIN. The main cycle is repeated five times before further steps are performed.

## C.2 The dry season

The dry season deserves special attention since its onset and duration critically affect plant growth, as described in Ngomanda et al., 2009. It is crucial being able to treat this period independently from the rest of the year, e.g. to apply changes that only affect the dry season and to have a clear and well defined transition from rain period to dry season, especially when it comes to investigating the dynamics of a model driven by daily input weather data. It is therefore necessary to re-adapt dry season boundaries year by year, for each year of the simulation. For that reason a function

depending on precipitation patterns over the year has been implemented to determine each year's onset and end of the dry period. Additionally one application depending on the knowledge of the dry season's boundaries will be demonstrated.

## Estimating dry season boundaries

To extract the dry season from a precipitation signal over the year, it appears helpful to apply a cumulative approach, i.e. to sum up daily precipitation, setting the lower boundary of the summation  $A$  constantly to the first day of the year, and moving the upper boundary  $B$  step by step from the second day to the end of the year (the summation goes from day  $A$  to day  $B$ ). Anyway, instead of summing up we chose to integrate the precipitation curve, with the same result. Theoretically a flat precipitation curve (i.e. constant precipitation, or zero precipitation as in a "perfect" dry season) creates a flat interval, or plateau, in the integrated curve (which is a function of  $B$ ), while possible  $\delta$ -function-like precipitation peaks interrupting the dry season should lead to  $\Theta$ -function-like steps in the integrated curve. In the first of our approach we're trying to detect those plateaus.

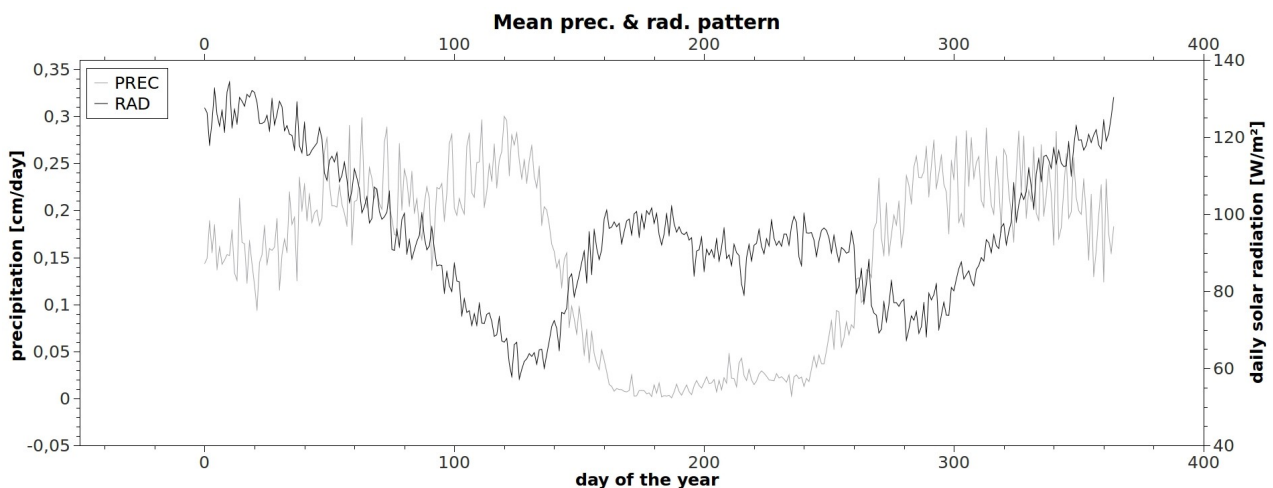


Figure C.7: Daily mean rainfall [mm/day] (left scale) and uncorr. solar radiation [W/m<sup>2</sup>] (right scale), averaged over N=99 simulation years, for Batéké Plateau, exemplary. Regions with lower rainfall exhibit higher amounts of incident solar radiation and vice versa (uncorrected radiation).

Figure C.7 shows precipitation and radiation patterns of the mean of 99 simulations from Batéké Plateau, Gabon (02°13'S, 14°02'E, 609m, elevation according to MarkSim's DEM) before correction. As in the following figures, day one of the year corresponds to Jan/01, day 365 to Dec/31.



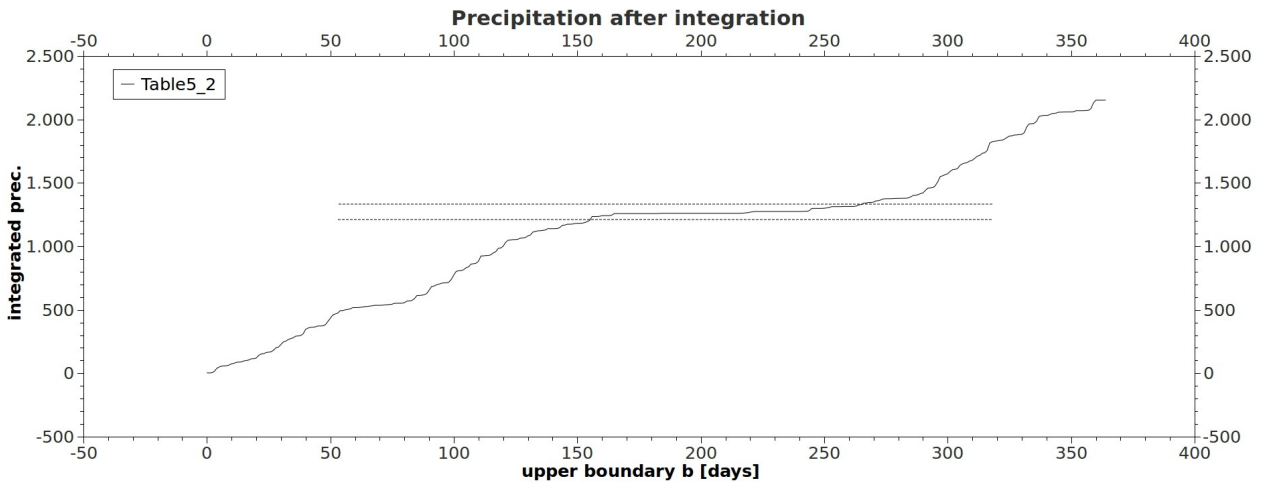


Figure C.8: Integrated precipitation signal (from figure C.7), the plateau in the middle of the curve indicating the dry season becomes visible.

1. First an integration is performed on the precipitation signal using “Simpson's Rule”, an approved method in numerical integration. The precipitation signal in daily resolution is the mean of all  $N$  simulation years. For a first estimate of the dry season, the algorithm determines a value along the ordinate (y-axis) of the integrated signal, referred to as *mean*, with a maximum number of neighbors with the same value, or values within a certain tolerance. In a next step we define a tolerance interval  $dy$  (also placed on the ordinate), which defines the length of the dry season in the first place. More precisely, the x-value of the points of intersection between the integrated precipitation signal and  $y = \text{mean} - dy$  or  $y = \text{mean} + dy$  (see figure C.8) define the onset and end of the dry season, respectively. The tolerance interval  $dy$  therefore directly regulates the size of the initial estimate of the dry season.
2. Second, the position of the whole dry season as defined above is moved for  $\pm 20$  days without changing its total length, trying to maximize the total number of dry days within the interval. The improved dry season is the interval with the highest number of dry days, i.e. were the daily precipitation = 0 mm/day.
3. Next another correction is performed, where the boundaries of that period are modified independently by  $\pm 30$  days (i.e. a month), looking for the highest value of the ratio *dry days* : *season length*, where the duration of the dry season is not kept at a constant value and *dry days* refers to days with precipitation < 0.5 mm. It has to be mentioned that the algorithm works perfectly symmetrical on both sides, i.e. at the beginning and at the end. The end will be set 30 days earlier in the year and then be shifted forward day by day, when at the same time the beginning is held at its constant date – and vice versa. The variable defining the shift of the boundaries will be called *move* in the following description. For each

DS (dry season) formed that way the *ratio* (as defined above) will be computed and finally the begin/ending with the highest *ratio* will be used. Additionally, if this boundary is moved over a certain period of dry days (*interntolerance* = 7 days custom settings) in a row (i.e. where precipitation < 0.5 mm, a certain tolerance was allowed to make the algorithm less sensitive to small precipitation events), the DS' end cannot be set to dates before (the begin cannot be set after) or in that period. Even if the possible end of the variation interval (determined by *move*) is reached, *move* can be enlarged to make sure the DS does not end between a number of dry days. The attempt of this is to include smaller rainless episodes that are separated from the main DS through a number of precipitation events, but that are still close to the main DS. The parameter *interntolerance* controls the duration of the dry season, by defining the length of these smaller rainless patches.

*move* is downscaled iteratively from 30 to a minimum of 7 days, i.e. the interval in which the boundaries can be varied is narrowed. Of course the aforementioned exceptions to extend *move* are still valid (e.g. the end of the DS must not be set if there are still dry days coming up later in the same row of dry days). Each reduced value for *move* produces dry season boundaries, some of which are equal to each other. Finally the DS with the most hits, i.e. the time interval that results from multiple reduced shifting intervals, is picked. This assures that even if we choose different ranges for the iteration (to generate the onset and ending of the DS), the algorithm will look for reasonable dates on which to set the boundaries. If the program finds the same dates for onset and ending, despite different starting days and range of the iteration procedure, we can be quite sure the detected day fulfills the required conditions. This iterative function, however, can be seen as the fine tuning of start and end of the DS and works at the same time against a bias towards too long dry seasons, which could result if *move*=30 is always the favored shift of the dry season.

Once again from the beginning: We are looking for a new value for the one boundary of the dry season, lets pick the onset for example. The interval we're looking in is [onset-30 days; onset+30 days], and we're looking for the day with the highest *ratio* (as defined above), and apply additional rules, that assure that we really pick a day which has a lot of precipitation on one side, and a dry period on the other side. These rules can also widen the interval we're looking in, to be sure we don't pick a day in the middle of the dry season, just because it has the highest ratio. Step by step we narrow the interval we're looking in, and according to our side conditions the same optimal boundary will be detected multiple times, despite a narrower searching interval. But the point will come, that the interval becomes so small that this boundary might not be included in the searching range any more, and a new optimal boundary will be detected multiple times. We now pick that one boundary, that has been detected more often.

As will be discussed later the initial tolerance interval determining the first estimate of the size of a dry season, is linked to the variability of the integrated precipitation signal around its mean. This assures that years with higher precipitation do not necessarily experience a shorter dry season.

$$error = \frac{1}{N} \cdot \sum_{n=1}^{365} |prcp_{integ}[n] - \hat{m}| \quad \text{eq. C.11}$$

$$dy = \frac{error}{dyreg} \quad \text{eq. C.12}$$

We can think of the mean (  $\hat{m}$  ) as the value that defines the plateau (figure C.8), and of  $prcp_{integ}[n]$  as all the values that together form the integrated precipitation curve.  $Dyreg = 15$  (empirical). By calling the program with a desired (different) length of the dry season, i.e. if the dry season is known from literature,  $dyreg$  is iteratively optimized to move the mean duration of the computed dry season towards the desired result.

On one hand, the newly defined boundaries for each years dry season are used within the main cycle of the correction procedure. As we have learned before, correction does take place on a daily basis, but based on monthly residuals between the desired climate normals and the actual WTG monthly averages. So, there are months where the mean value of the variable in question (srad or tmax) have to be raised and that border months where the opposite is the case, i.e. the mean has to be lowered. The transition date from one month to the other therefore represents a change in the qualitative behavior of the daily values of tmax or srad. If a transition date of that kind falls into the dry season, it is moved to the day where the dry season starts, or ends. This is put into effect by only allowing reductions of daily values (of tmax or srad) during the dry season since temperature and solar radiation are constantly overestimated during the DS at all sites tested in Gabon. If a month's average has to be increased, and this months is partly within the dry season (this could easily happen since we implemented dynamic transition dates), the correction is only performed outside the DS and therefore creates a qualitative change between daily values within and outside the dry season.

On the other hand the information, when a dry season starts and when it stops is valuable since it permits us to perform changes that concern the DS only. One application, that should only serve as inspiration but that has not been validated, will be introduced:

## One application

It is reasonable to assume that during the dry season, the solar radiation will experience less day-to-day fluctuations than during the rest of the year. In Gabon the skies in that time of the year stay mainly covered, so the variables that influence the amount of incident solar radiation per day are molecular structure of the cloud aerosol, thickness of the cloud layer and day length. Since clouds

don't produce precipitation and the continuous layer stays stable for several months, we don't expect too many fluctuations resulting from the cloud cover. Since Gabon is close to the equator, day length is of no consequence. Nevertheless MarkSim produces a radiation (and maximum temperature) signal that fluctuates during dry season just as much as it does during the rest of the year. The boundaries of the dry season can, for example, be used to smoothen the signals that they enclose:

$$srad_{new}[i] = srad_{DS} - (srad_{DS} - srad_{old}[i]) \cdot SMF \quad \text{eq. C.13}$$

$srad_{DS}$  is the mean value of the solar radiation in all estimated dry seasons,  $srad[i]$  are daily values of the solar radiation within the dry season and SMF is a smoothing factor between zero and one. The new daily values of  $srad[i]$  will now be scattered around the over all mean of the dry season. The magnitude of the variation is regulated by SMF: values close to 0 mean little variation, values near one lead to results close to the initial distribution. Again it must be said here that this is only a suggestion, that has not been validated!

### C.3 Vapor pressure deficit (VPD)

Since MarkSim does not support estimations for daily VPD in the atmosphere, although this variable is required as input for Biome-BGC, it is also generated within this process. The estimation of vapor pressure deficit is based on the assumption that VPD stays constant during the day and that the minimum temperature can be set equal to the dew point temperature (Kimball et al., 1997). VPD at a given daily mean temperature can then be calculated according to Abbott and Tabony, 1985, where vapor pressure at dew point ( $pva$ ) is subtracted from saturation vapor pressure ( $pvs$ ) at the daily mean temperature ( $tday$ ).  $Tday$  is calculated from minimum and maximum temperature:

$$tday = \frac{tmax + tmin}{2} + 0,45 \cdot (tmax - \frac{tmax + tmin}{2}) \quad \text{eq. C.14}$$

$$pva = 610,7 \cdot e^{\frac{17,38 \cdot tmin}{239 + tmin}} \quad pvs = 610,7 \cdot e^{\frac{17,38 \cdot tday}{239 + tday}} \quad \text{eq. C.15}$$

$$VPD = \begin{cases} pvs - pva & (pvs > pva) \\ 0 & (pvs \leq pva) \end{cases} \quad \text{eq. C.16}$$

### C.4 Output files

The output interface of the whole application works through text files where weather data are stored at different temporal resolutions. For most of these files there is a "beforecorr" (before the correction has been applied, i.e. unchanged output) and an "aftercorr" (after the correction has been applied) representation in the same format. We distinguish between the following output files:

*DailyMeanVal\_beforecorr.dat* & *DailyMeanVal\_aftercorr.dat*: In these files  $tmin$  (minimum

temperature in °C), tmax (maximum temperature in °C), srad (solar radiation MJ/m<sup>2</sup>/day) and prcp (precipitation in mm/day) are listed in daily resolution. Each entry refers to a day in the year and represents the average on that specific day of all N simulations, which means that there is a total of 365 entries.

*DayMetList\_beforecorr.dat & DayMetList\_aftercorr.dat*: This is probably the most interesting output concerning further use of the weather data. Tmin [°C], tmax [°C], prcp [cm/day] and srad [W/m<sup>2</sup>], as well as tday (daily mean temperature) [°C], VPD [Pa] and day length [s] are presented in daily resolution. The file has N x 365 entries, where N is (as usually) the number of simulation years.

*MonthlyMeanVal\_beforecorr.dat & MonthlyMeanVal\_aftercorr.dat*: tmax [°C], tmin [°C], prcp [mm/month] and srad [MJ/m<sup>2</sup>/day] are listed in monthly resolution, monthly mean values are the average of all days in the according month and of all N simulation years, the file therefore has 12 entries. Each variable is accompanied by its climate normal (e.g. TMAXclx) and its standard deviation (e.g. sdTMAX).

*Statistics\_beforecorr.dat & Statistics\_aftercorr.dat*: Pearson's linear correlation coefficient r & sample mean error (err) as defined in APPENDIX B (Pearson's linear correlation coefficient & sample mean error) are computed for every variable. The intention is to give the user a feeling for the quality and success of the correction that has been performed.

*Seasons.dat*: For each of all N simulation years the onset and ending, as well as the duration of the estimated dry season are listed.

*CorrFactors.dat*: Here one can observe what correction and truncation factors have been computed.

## **C.5 Optimization**

In this chapter we will briefly discuss how certain parameters used for the corrective process have been defined and tuned to achieve the most accurate results possible.

### **Iterations of the main cycle**

Every time running through the main cycle, the residuals become a bit smaller. But we expect a certain number of repeated runs of the main cycle, until the mean error between desired and current output is saturated, i.e. when it doesn't get any smaller due to random fluctuations caused by the stochastic correction and truncation process. We examine how the mean error (as defined before), and the Pearson correlation develop with the number of iterative runs of the main cycle. Since the correction contains a stochastic element too, each correction process has been performed 20 times with differing number of iterations of the main cycle, and the average r and the mean error accompanied by their standard deviations have been computed.

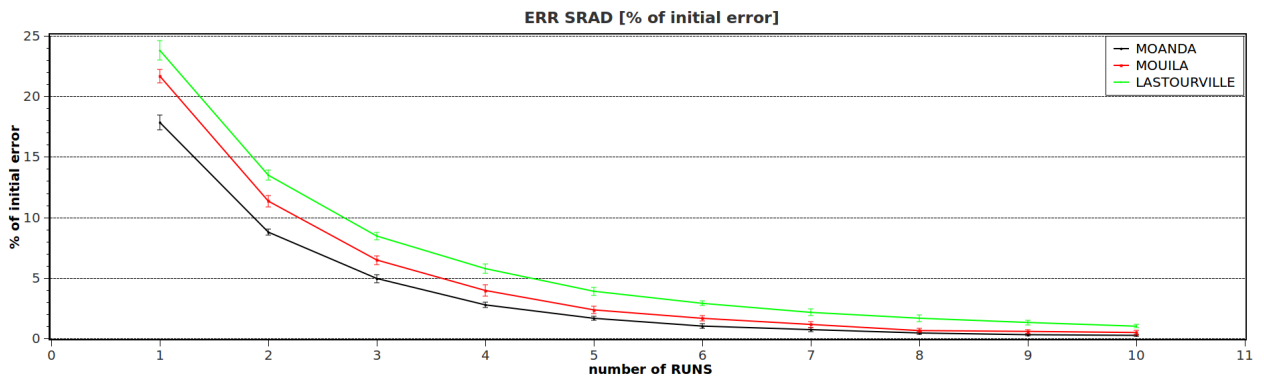


Figure C.9: Mean error development of solar radiation

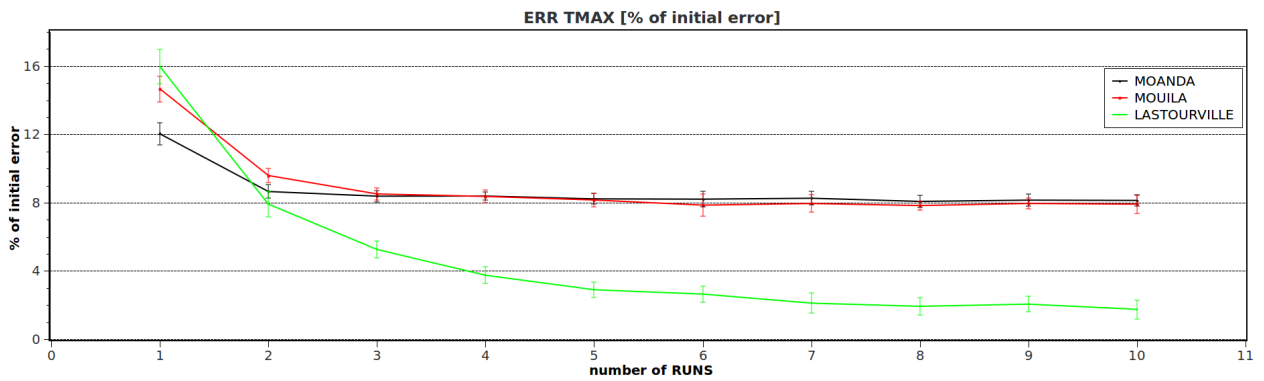


Figure C.10: Mean error development of maximum temperature

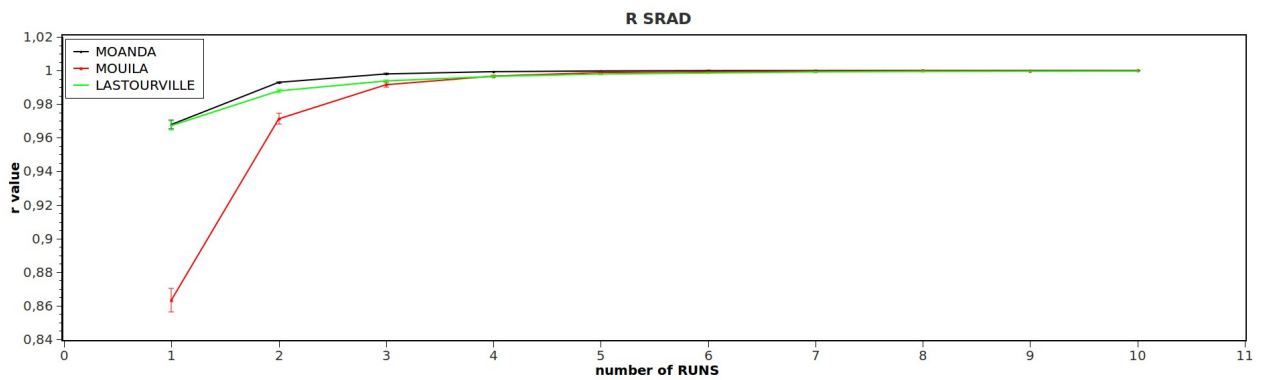


Figure C.11: Development of linear correlation, solar radiation

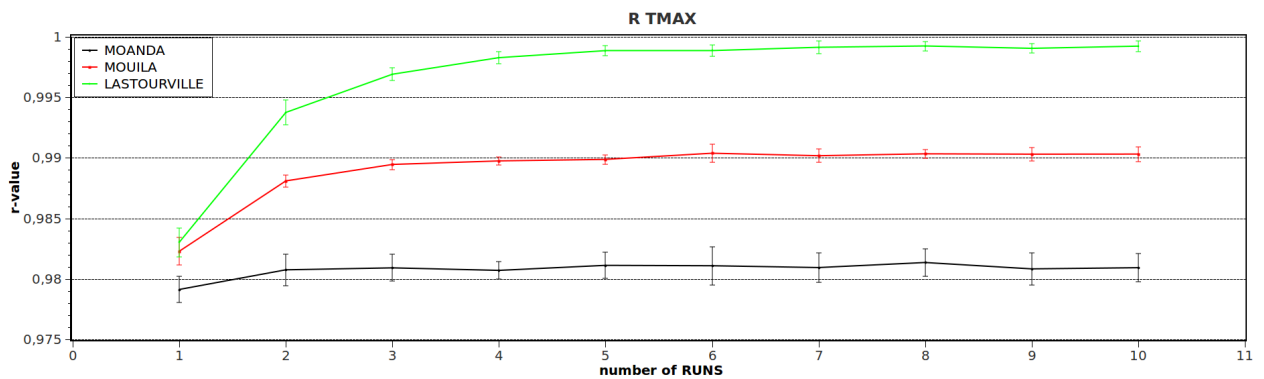


Figure C.12: Development of linear correlation, maximum temperature

As we can read from the graphs, the error of the solar radiation has shrunk to less than 5% of its initial value for all three sites after only 5 iterations. For Lastoursville, the error of the maximum temperature after 5 iterations is also below 5% of the original value, while for the other sites we can observe a saturation effect at the 8% level. The same is apparent for the trend of the linear correlations  $r$ , where after 5 iterations their final value is reached. For this reason the main cycle is repeated 5 times per corrective process.

### dyreg – regulating the length of the dry season

It can be concluded that (besides *interntolerance*) *dy* is the most important parameter within the application that determines year by year boundaries of the dry season, as it defines the size of the first estimate of the dry season. Tuning *dy* turned out tricky, since each precipitation profile (=one WTG file, i.e. one year of daily precipitation values) has to be checked by hand and it has to be decided whether the boundaries of the dry season as defined by the algorithm introduced before make sense, or not. For each site this procedure was repeated 99 times with multiple values of *dy*. It appeared helpful to link *dy* to the variation of the cumulated precipitation signal around its annual mean (i.e. the “plateau”) divided by a certain number. Different numbers have been tried, with the attempt to fit the mean duration of the dry season of certain sites in Gabon to their actual literature values. *Dy* was finally defined as  $dy = error / dyreg$ , where *error* is the mean difference between annual mean of the precipitation and its daily values, and *dyreg* = 15. Apart from this default setting, *dy* can be changed within a certain range to achieve a value for the mean duration of the dry season closer to the desired value. For this, a desired duration of the DS can be entered when calling the program, and *dy* will be iteratively changed to optimize the duration of the DS.

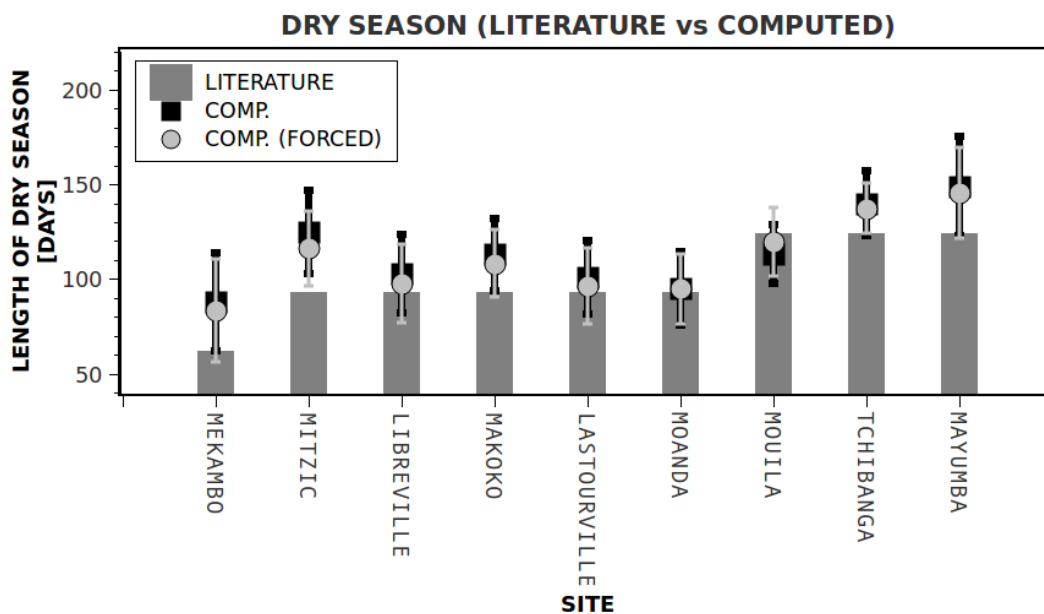


Figure C.13: Dry seasons compared. Grey drop lines: literature values, black squares: computed mean dry seasons with dyreg=15, and light grey circles: computed values with an iteratively optimized dyreg (a desired dry season has been entered).

We have to admit, that literature values are in monthly resolution and therefore quite imprecise. Taking into account that literature values are vague estimates the results are quite satisfying. Except for sites Moanda and Mouila, the duration of the dry season is overestimated by around 10 days, the computed values with optimized *dyreg* are somewhat closer to the literature values. Technically the computed estimates of the dry season with fixed *dyreg* (black squares) could be moved closer towards the desired values by increasing *dyreg* (and thereby reducing *dy*), but this results in significantly short annual dry seasons for the site Moanda and is therefore not possible.



## **APPENDIX D - Generating paleo climate and climate change**

There is growing interest in ecosystem behavior under changing climate, which comprises scenarios of future climate with increasing temperature and a possible shift in annual precipitation, and the influence of past climate on the present state of an ecosystem of interest. Climate time series containing annual mean values might stem from paleo climate reconstructions, abundantly available for many parts of the world, or be variants of climate change that might affect us in future, such as the 2°+ or the 4°+ scenario. It is, however, possible to directly translate a climatic trend, expressed in annual mean values of temperature and rainfall, into a series of daily weather files with well defined fluctuations around the desired trend.

All the information required is one time series of annual rainfall and one of annual mean temperature and the according years of occurrence in ascending order. This can either be interpreted as future events in years AC (?) (e.g. 2050 – 2100 – 2200 etc) or as past events, in years BP or years BC (8 – 50 – 130 etc.). The years of occurrence of the two time series don't have to be equal (which allows us to combine information from various climate reconstructions for instance), but of course an overlap of the time intervals is necessary. Additionally a template CLX file is required that contains the proper distribution of monthly rainfall, temperature and temperature range, as well as longitude, latitude and elevation of the site of interest.

The application interpolates linearly between two data points (one data point contains either annual precipitation [mm/yr], or annual mean temperature [°C] and the year of occurrence). For each year in the overlapping time interval of the two climate time series a DAT-file is created containing the interpolated values of rainfall and temperature, distributed like in the template CLX-file. Precipitation is increased or reduced (from the template value to the real value) in relative proportions of the monthly values to adjust the annual amount to the desired value. The same is true for monthly mean temperature, which has first to be transformed to the Kelvin scale, in order to enable calculation with relative amounts. As explained in the previous section, MarkSim features XBF and CBF files to deal with multiple DAT-files at the same time, which is used to turn the time series, which contained only annual values in the beginning, into a series in daily resolution. Each DAT file should be set to generate 99 climate years, even if only one time series of daily values is needed, and the amount of computing time and disk space is much higher. The reason for this is that on one hand MarkSim often tends to under, or overestimate its output compared to the desired annual mean (see Figure D.1, grey line for instance), and by generating 99 climate years per year in the time series, we can select only those years containing values within a certain range of the mean (Figure D.1, red line). Rather than having linear transitions from one data point to the other, year-to-year fluctuations are usually preferred. Furthermore, applying climate change or paleo climate series on ecosystem models, only repeated simulations with similar time series that exhibit different year-to-year variations, will allow us to define a range for the error of the predicted

measure. Therefore, by working with 99 climate years we can create multiple similar time series that only differ in year to year variation, but that are similar in their general trend. The other reason for choosing 99 climate years rather than any smaller number is that the correction procedure of solar radiation and maximum temperature has been developed for 99 climate years, even though it has shown to work with smaller numbers too.

Once for each year in the climate series there are 99 climate years are available, the corrective procedure of solar radiation and maximum temperature can be performed. Next we have to apply a mechanism that selects only one year out of each set of 99. One option is to randomly select a climate year, and to check whether it is within a desired range of the mean, and to reapply this procedure until one year is found that meets the conditions. Another possibility would be to chose a random precipitation value from a normal distribution with mean as defined in the original time series and a certain variance, and see which year of the set of 99 has the amount of annual rainfall closest to this random number. While the first approach only makes sure that all variation stays within a certain range of the mean, the second approach also tries to mimic the distribution of the fluctuations, which can be considered as Gaussian. By repeating these processes with a different seed of the random number generator, various time series of daily weather data can be produced.

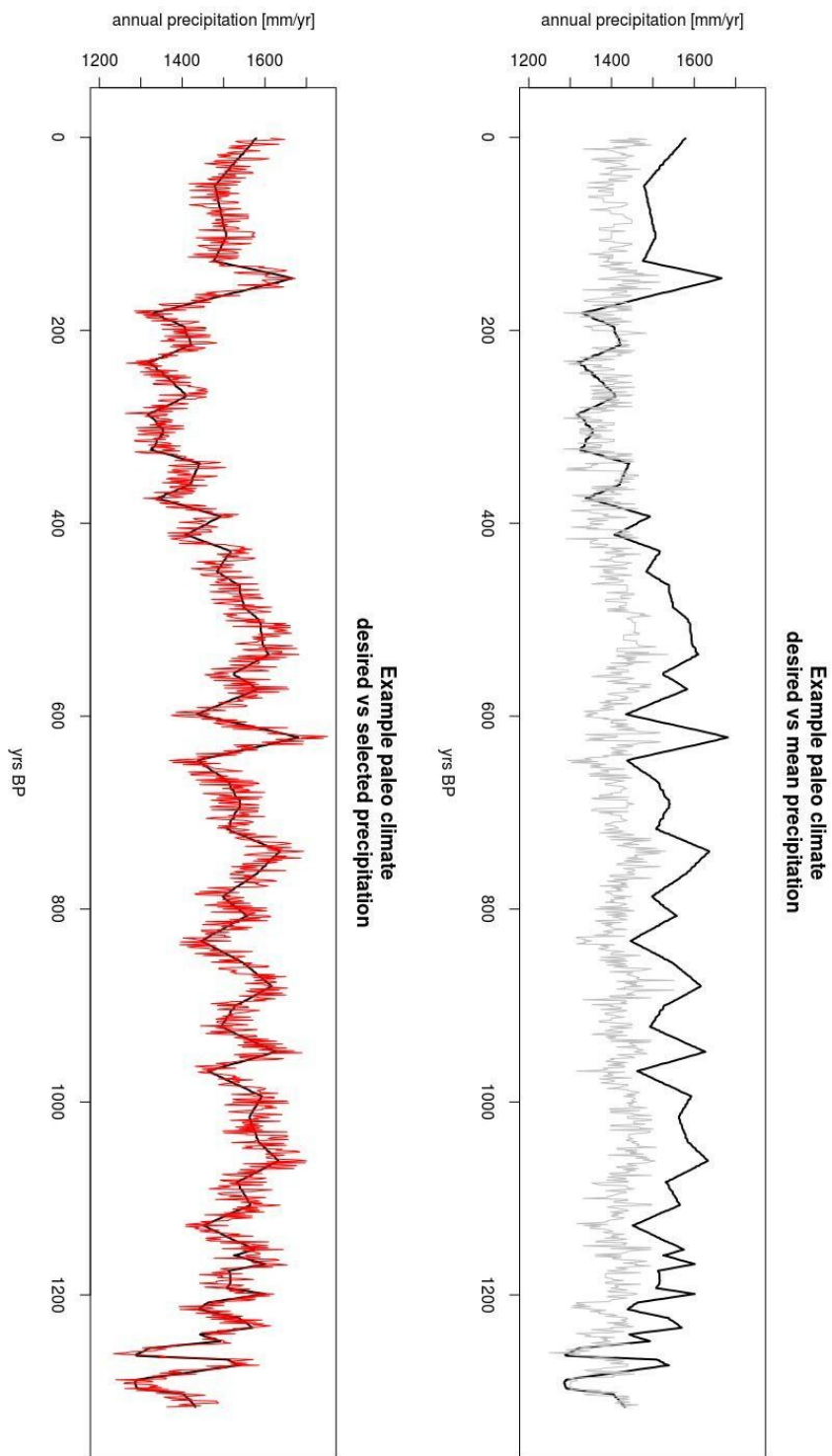


Figure D.1: Example paleo climate reconstruction from pollen records for Lake Kamalété, Gabon, as described in Ngomanda, 2005. The grey line in the upper graph presents the mean annual precipitation from 99 climate years, the red line in the lower graph shows annual precipitation of selected climate years which were chosen to stay within a +/- 5% range of the original time series (black line).

**Johannes Elias Bednar**  
Curriculum Vitae, November 2011

PERSONAL DATA

Place of birth: Korneuburg, Lower Austria  
Citizenship: Austrian  
DOB: Nov 2nd, 1985

CONTACT INFORMATION

Schallergasse 11/18  
A-1120 Vienna, Austria  
email: [johannes.bednar@chello.at](mailto:johannes.bednar@chello.at)

EDUCATION

1985-1989 PRIMARY EDUCATION,  
Maria Lanzendorf, Lower Austria

1996-2004 HIGH SCHOOL WITH EMPHASIS ON NATURAL SCIENCES  
Naturwissenschaftliches Realgymnasium  
A-1100 Vienna, Laabergstraße  
Matura (final exam) *cum laude*

2005-2011 DIPLOMA PROGRAMME IN PHYSICS,  
University of Vienna

2008-2009 JOINT STUDY SCHOLARSHIP,  
University of Melbourne, Australia

2010 ERASMUS PROGRAMME  
University of Montpellier II (UM2), France

2010-2011 DIPLOMA THESIS  
University of Natural Resources and Life Sciences, Vienna

PROFESSIONAL EXPERIENCE

2004-2005 Architecture Internship  
Architectural firm Halbritter&Halbritter  
A-7100 Neusiedl am See, Burgenland,  
Austria

2007-present Snowboarding instructor  
during winter season

2009-2011 Tutor of physics for nutritional science  
University of Vienna

LANGUAGES SPOKEN

German – Native  
English – excellent fluency, written and spoken.  
French – Good fluency, written and spoken