Hydrological Sciences Journal

ISSN: 0262-6667 (Print) 2150-3435 (Online) Journal homepage: http://www.tandfonline.com/loi/thsj20

Rainfall characteristics and their implications for rain-fed agriculture: a case study in the Upper Zambezi River Basin

M. Beyer, M. Wallner, L. Bahlmann, V. Thiemig, J. Dietrich & M. Billib

To cite this article: M. Beyer, M. Wallner, L. Bahlmann, V. Thiemig, J. Dietrich & M. Billib (2016) Rainfall characteristics and their implications for rain-fed agriculture: a case study in the Upper Zambezi River Basin, Hydrological Sciences Journal, 61:2, 321-343, DOI: 10.1080/02626667.2014.983519

To link to this article: http://dx.doi.org/10.1080/02626667.2014.983519

Accepted author version posted online: 10 Nov 2014.
Published online: 17 Dec 2015.

HYDROLOGICAL SCIENCES JOURNAL – JOURNAL DES SCIENCES HYDROLOGIQUES, 2016
VOL. 61, NO. 2, 321–343
http://dx.doi.org/10.1080/02626667.2014.983519

Rainfall characteristics and their implications for rain-fed agriculture: a case study in the Upper Zambezi River Basin

M. Beyer[a,b], M. Wallner[a], L. Bahlmann[a], V. Thiemig[c,d], J. Dietrich[a] and M. Billib[a]

[a]Institute of Water Resources Management, Hydrology and Agricultural Hydraulic Engineering, University of Hannover, Hannover, Germany; [b]Federal Institute for Geosciences and Natural Resources (BGR), Hannover, Germany; [c]Climate Risk Management Unit, Institute for Environment and Sustainability, Joint Research Centre, European Commission, Ispra, Italy; [d]Faculty of Geosciences, Utrecht University, Utrecht, The Netherlands

ABSTRACT
This study investigates rainfall characteristics in the Upper Zambezi River Basin and implications for rain-fed agriculture. Seventeen indices describing the character of each rainy season were calculated using a bias-corrected version of TRMM-B42 v6 rainfall estimate for 1998–2010. These were correlated with maize yields obtained by applying a SVATmodel. Finally, a self-organizing map (SOM) was trained to examine multivariate relationships. The results reveal a significant spatio-temporal variability of rainfall indices and yields, with a gradient from north to south. Yields greater than 1 t/ha are found to be only achievable with rainy seasons longer than 160 days. For shorter durations, the interplay of total rainfall, dry spell frequency and maximum dry/wet spell durations determines agricultural success. Using total rainfall alone or wet day frequency as estimators for yields is insufficient. Alternating wet and dry spells affect yields most negatively. The results have significance in the context of agricultural planning under changing climatic conditions and agricultural planning, as well as for the development of forecasting mechanisms.

ARTICLE HISTORY
Received 31 March 2014
Accepted 23 October 2014

EDITOR
Z.W. Kundzewicz

ASSOCIATE EDITOR
A. Efstratiadis

KEY WORDS
rainfall characteristics; rain-fed agriculture; maize (*Zea mays*); Zambezi River Basin; self-organizing map (SOM); TRMM 3B42 v6

1 Introduction

The foremost farming practice over most parts of Sub-Saharan Africa (SSA) is rain-fed agriculture. In fact, 97% of the total cropland in SSA is dominated by rain-fed farming (Calzadilla *et al.* 2013) and this agricultural practice is expected to remain the major source of staple food production for the majority of people in rural areas (Cooper *et al.* 2008). In these areas severe challenges are associated with the successful cultivation of the main staple crops (in particular, maize and millet/sorghum). Milgroom and Giller (2013) identify in their study four main issues related to the extremely low yields in rural SSA: (a) non-accessibility to necessary inputs at the right time (i.e. availability of seeds, man-power or land); (b) rainfall variability; (c) losses in the field (i.e. elephant- or hippo-related losses); and (d) post-harvest losses (pests). Despite each of these factors being equally important, numerous studies emphasize the crucial role of rainfall and its variability (Reason *et al.* 2004, Tadross *et al.* 2005, Cooper *et al.* 2008, Crespo *et al.* 2011, Mupangwa *et al.* 2011, Calzadilla *et al.* 2013, Milgroom and Giller 2013). In particular, over large parts of SSA, an extremely high spatio-temporal variability (Boko *et al.* 2007) of rainfalls on an inter-annual as well as periodic basis (Tyson *et al.* 1975, Reason *et al.* 2004, Tadross *et al.* 2007) is reported. Induced impacts of climatic changes, such as a decrease of early-rainy-season rainfall and a rising number of extreme events (Christensen *et al.* 2007, Kotir 2011), further increase the risk of low yields or even yield losses. Although a seasonal forecast system exists to support the agricultural sector (Famine Early Warning System, http://www.fews.net/), this information is not precise enough (only delivering approximate rainfall season start dates) and is often not accessible in rural areas. As a consequence, rain-fed agriculture in rural areas is still based on a "trial-and-error" principle: crops such as maize or millet/sorghum are commonly planted after the first showers of the wet season, in the hope that the subsequent weeks will deliver enough rainfall to allow germination and growth. If rains are not sufficient after sowing (i.e. the occurrence of a dry spell), seeds often need to be replanted causing enormous financial damage to the farmers which might even threaten their existence.

To address these issues, several studies explicitly point out the need for a more detailed analysis of the spatio-

CONTACT M. Beyer beyer.matthias@gmail.com
This article was originally published with errors. This version has been corrected. Please see Erratum (http://dx.doi.org/10.1080/02626667.2016.1140484).

temporal variability of rainfall (Usman and Reason 2004, Tadross *et al.* 2005, Hachigonta and Reason 2006, Hachigonta *et al.* 2008). Before the development and installation of any kind of forecasting system (i.e. prediction of likely onset and cessation dates), the most important characteristics still need to be identified and quantified on a smaller spatio-temporal scale.

The understanding of rainfall patterns in SSA and particularly the interconnection between rainfall characteristics and agricultural yields have been the focus of past research. Numerous studies were executed that investigated the association of rainfall characteristics with synoptic patterns (e.g. the El Niño-Southern Oscillation phenomena). In particular, spatial and temporal patterns of dry spells within the rainy season over southern Africa as a whole (Usman and Reason 2004), dry and wet spell frequencies and rainy season onset dates over the Limpopo region, South Africa (Reason *et al.* 2005), and the variability of dry and wet spell frequency in the rainy season over Zambia (Hachigonta and Reason 2006) have been associated with El Niño and La Niña. An effect of these phenomena on rainfall distribution was found in all of these studies, which might be a starting point for a possible forecasting method. Additionally, Tadross *et al.* (2005) analysed the onset of the rainy season over South Africa and Zimbabwe. Using a self-organizing map (SOM) approach, typical rainfall patterns associated with an early/late onset were identified.

Studies relating rainfall indices to maize growth were published by Tadross *et al.* (2007) and Crespo *et al.* (2011). The former authors used daily station rainfall data and downscaled GCM scenarios to calculate a wide range of rainfall characteristics and maize growth indices over southern Africa. Their results show that there have been weak trends for a later onset and earlier offset of the rainy season in the north and thus shorter rainy season durations. Crespo *et al.* (2011) applied a crop modelling approach and future climate scenarios based on 10-day input data for precipitation and evapotranspiration, showing that maize yields are most sensitive to rainfall in the 10-day sowing period.

However, the representativeness of the above-mentioned research is partially limited by the characteristics of their input data. Some of the studies utilize day station data, others use either spatially and/or temporally coarse data (e.g. 2.5° × 2.5°; 5-day), which are not able to reproduce sufficiently the high spatio-temporal variability of rainfall patterns that determine the agricultural yields (Usman and Reason 2004, Tadross *et al.* 2005). This further limits the range and accuracy of rainfall indices that can be determined (e.g. in regard to the precision of dry/wet spell durations or extreme events). However, several studies have been carried out using station data (Camberlin and Diop 2003, Raes *et al.* 2004, New *et al.* 2006, Tadross *et al.* 2007, Camberlin *et al.* 2009, Mupangwa *et al.* 2011); hence daily data could be applied. Using station data, the spatial pattern of rainfall characteristics cannot be represented sufficiently, especially in areas with low station density. In essence, examining a new source of input data might be the crucial step forward in the understanding of the interaction between rainfall patterns and agricultural yields in semi-arid regions.

Due to a scarce network of meteorological stations and inaccessibility of information (i.e. caused by the civil wars), sufficient data allowing such analyses in a detailed manner were not available until the beginning of this century. With the emergence of satellite-based rainfall estimates (SRFE), data of adequate spatial and temporal resolution are becoming accessible with a nearly global coverage. This enables researchers to carry out hydrological studies on catchments across different scales, regardless of whether ground observational data are available or not.

Nevertheless, an in-depth validation of several SRFE is required before application, to assure the selection of the most suitable SRFE for the particular target area (Hong *et al.* 2004, Dinku *et al.* 2007). The accuracy of each SRFE differs massively with the used product and study area (Thiemig *et al.* 2012). For the target area of the present study, i.e. the Upper Zambezi River Basin (UZRB), the quality of numerous SRFE has been investigated in previous studies by Thiemig *et al.* (2012) and Cohen Liechti *et al.* (2012). Thiemig *et al.* (2012) found the RFE 2.0 (NOAA 2002) and TRMM 3B42 v6 (Huffman *et al.* 2007) to be the most accurate SRFE for the Zambezi River Basin (considering also CMORPH, GPROF 6.0, PERSIANN, GSMap MVK and ERA-Interim), showing a good capture of timing of heavy rainfall events, intra-seasonal variability, distribution pattern and annual rainfall amounts. In addition, the reanalysis dataset ERA-Interim (Dee *et al.* 2011) showed good correspondence with observed values for intraseasonal variability and spatial distribution, whereas the number of rainy days was overestimated. Timing of heavy events was captured very well within a lag time of 0 to +1 days (point-to-pixel analysis). All SRFE showed a high uncertainty in estimating the amount of heavy rainfall events in terms of general underestimation. Cohen Liechti *et al.* (2012) stated that the reliability of all SRFE increases with the temporal scale; hence, the chosen time step has a major influence on the quality of all products. In contrast to Thiemig *et al.*

(2012), they found that both RFE2.0 and TRMM overestimate rainfall. This contradiction might be explained by the fact that Cohen Liechti *et al.* (2012) did not use the official data from stations provided by the Zambia Meteorological Department which were accessible in the study of Thiemig *et al.* (2012).

Summing up, there is a gap in existing research in a sense that until now either accurate point data *or* distributed but temporally inaccurate (i.e. accumulated 5-day) data have been used to characterize rainfall patterns across SSA. Further, none of the previous studies gives a complete breakdown of the most relevant rainfall characteristics and how they affect agricultural success of rain-fed farming.

The objective of this study is a high-resolution analysis of rainfall characteristics and their impact on yields for rain-fed maize throughout a major river basin in southern Africa. As information on measured yields is only available in certain areas, an agricultural model is applied to increase the availability of data. For our research, we address the following points in order to contribute to filling gaps in the existing literature:

a. How are important rainfall characteristics distributed in time and space?
b. How and in which way are rainfall characteristics affecting rain-fed agricultural yields?
c. Which rainfall characteristics or combinations of characteristics are the determining factors for the success of rain-fed farming? Can single indicators be used as predictors?
d. Which regions within the study area are the most/least suitable for rain-fed agriculture?

This study provides the first complete investigation of rainfall characteristics, their spatio-temporal distribution, variability and implications for agricultural yields of maize, based on a daily time step and high spatial resolution. Regional patterns, as well as trends or shifts of them, are revealed and an identification of suitable or less suitable areas for rain-fed farming is presented. The results thus provide quantified information for farmers and local decision-makers, based on the best available data from the combination of better spatial representation through SRFE and better point information provided by station data. We further present a methodological framework based on freely available data that combines various types of datasets and tools in order to tackle a large-scale water management problem.

2 Materials and methods

The methodological approach comprises three parts:

a. a high spatial resolution (0.1° × 0.1°) analysis of multiple rainfall characteristics using bias-corrected SRFE data on a daily basis in the UZRB for the years 1998–2010;
b. calibration and validation of a SVAT (Soil–Vegetation–Atmosphere Transfer) model on a daily basis to simulate rain-fed agricultural yields for maize basin-wide; and
c. both univariate and multivariate analysis of rainfall characteristics *vs.* modelled agricultural yields in order to identify key characteristics triggering rain-fed farming.

This study combines multiple datasets, which are presented in Section 2.1, followed by a description of the above methodological approaches, in sections 2.2, 2.3 and 2.4, respectively. The complete methodological framework is summarized in Fig. 1 which is arranged so that data sources are presented on the left-hand side of the scheme, whereas applied methodologies and tools are on the right-hand side.

2.1 Meteorological, soil and agricultural data

2.1.1 Bias-corrected TRMM 3B42 v6

TRMM 3B42 v6 is a daily, near-global (50°N to 50°S) satellite-based rainfall estimate provided by NASA uninterruptedly since 01/01/1998. Estimations take advantage of various data sources, by combining calibrated MW-based estimates (TCI, SSM/I and AMSU) with IR-based estimates (Geostationary IR), which are then rescaled using monthly ground observations (CAMS and GPCP). The final product holds a spatial and temporal resolution of 0.25° and 3 h, respectively (Huffman *et al.* 2007, 2010).

For this study, a bias-corrected version of the TRMM 3B42 v6 was used as reference data, since a previous validation study by Thiemig *et al.* (2012) has shown that TRMM 3B42 tends to underestimate precipitation over the target area (see Fig. 2(a) and (b)). Therefore, in a post-processing step, the 'histogram equalization' (HE) method—a recent bias correction method used to correct precipitation estimates from climate models (Krajewski and Smith 1991, Piani *et al.* 2010)—has been applied to the original TRMM-3B42 data.

The meteorological ground observation network used as input to create the areal precipitation fields (which

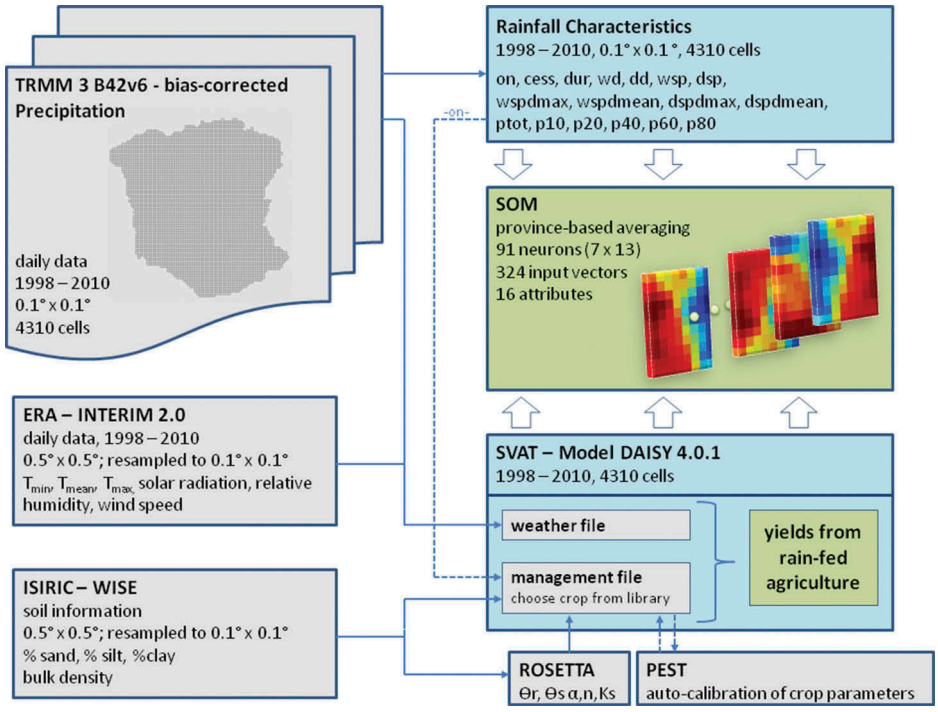

Figure 1. Schematic view of the research framework. The left side indicates data sources, and applied methodologies are depicted on the right.

were then used as reference for the bias correction) consists of 96 stations, within the Zambezi River Basin, with 77% of the data provided by the Zambia Meteorological Department and only 23% originating from the WMO GTS stations. Figure 2(c) shows the annual average of the final bias-corrected TRMM-3B42, which resembles the annual average KED-interpolated observations closely and also agrees with existing literature (Shahin 2002).

2.1.2 Raster-based soil data

Besides bias-corrected TRMM 3B42 v6 rainfall, two other datasets were used for this research, i.e. ISIRIC-WISE and ERA-Interim 2.0 (see Fig. 1). Both are freely available online and have been utilized widely in scientific studies (Batjes 2005).

The ISIRIC-WISE global dataset of derived soil properties is a harmonized, gridded (0.5° × 0.5°) dataset containing soil parameter estimates for the component soil units of each terrestrial grid cell (Batjes 2005). Each cell of the dataset can contain up to 10 soil units that were derived from the analysis of about 9600 soil profiles at the World Soil Information Centre, Wageningen. In total, the dataset compiles 22 soil variables for topsoil (0–30 cm) and subsoil (30–100 cm) enabling, for example, the use for agro-ecological zoning, crop growth simulation or analysis of global environmental change (Batjes 2005). For the study, information on the contents of sand, silt and clay as well as bulk density was extracted from ISIRIC-WISE. Due to missing data for the subsoil (30–100 cm) and generally marginal differences between top- and subsoil, only one layer (0–100 cm) was defined within the SVAT model DAISY (see Section 2.3). Further, the dataset was resampled to obtain the same spatial resolution as the bias-corrected TRMM 3B42 v6 data (0.1° × 0.1°).

2.1.3 Other raster-based meteorological data

Meteorological data other than rainfall were obtained from ERA-Interim 2.0 (Dee *et al.* 2011), a reanalysis of global atmospheric data provided by the European Centre for Medium-Range Forecasts. It might be argued that the mixing of two datasets, TRMM and ERA-Interim, introduces uncertainty because different timing of rain events might be captured by the two datasets, which could create inconsistency with the other meteorological parameters. However, this fact had to be accepted because our focus is to apply the best available rainfall estimate; hence either interpolated station data or a product had to be chosen. ERA-Interim 2.0 contains a large number of atmospheric variables (>30 parameters), has a spatial resolution of 0.5° × 0.5° and is available up to three-hourly time steps. For the purpose of this study, the following variables were extracted on a daily basis:

- minimum, mean and maximum temperature
- solar radiation
- dew point

The set-up file within DAISY requires relative humidity data; hence dew point values were converted into relative humidity. Wind speed at 2 metres was not available from ERA-Interim 2.0; therefore a wind speed of 2 m/s (model standard value if no data available) was assigned in order to avoid involving another different dataset. As for ISIRIC-WISE, the data were resampled to a spatial resolution of 0.1° × 0.1° in order to homogenize the spatial resolution with the rainfall data.

2.1.4 Data of measured yields

For calibration and validation of the agricultural model, observed maize yield data for rain-fed agriculture were necessary. Throughout the basin, only two reliable sources were found: (i) data for the Eastern Caprivi region located in the southwestern part of the UZRB that were extracted from the "Agricultural Statistics Bulletin 2000–2007" of the Department of Water Affairs, Namibia (DWA 2009), and (ii) detailed panel survey data for the years 2001, 2004 and 2008 that were provided by the Agricultural, Food and Resource Economics Department, Michigan State University (Jayne 2013).

2.2 Evaluation of rainfall characteristics

Rainfall characteristics were determined on a cell-by-cell basis for each cell and year separately. In particular, each time series was analysed and the defined rainfall characteristics were calculated based on the criteria summarized in Table 1.

In total, 17 rainfall characteristics were selected for evaluation. This study uses satellite-derived rainfall and station data (i.e. bias-corrected satellite-derived data) to combine the high spatial resolution of the former with the more precise point information of the latter dataset (refer Section 3). Criteria for the investigated rainfall characteristics were taken from existing literature and adopted whenever no general agreement on the criteria for several rainfall characteristics could be achieved. For the detection of the onset of the rainy season, for example, the criterion suggested by Tadross *et al.* (2007) was used: a sum of 25 mm in the first 10 days not followed by 10 or more consecutive dry days in the next 20 days. Through the latter condition, detection of a 'false' start of the rainy season is avoided (Mupangwa *et al.* 2011). From the same study, a threshold of 2 mm was selected for the separation of wet and dry days. Other authors use higher (e.g. Mupangwa *et al.* 2011) or lower (e.g. Camberlin *et al.* 2009) thresholds. In general, it has to be stated that the selected threshold depends on the purpose of the particular study. If looking at groundwater recharge rates, a rainfall of 1, 2 or even 5 mm might not be considered a wet day because effectively none of this water will reach the groundwater table in a region with a daily pan evaporation of 5–8 mm (Woltering 2005). Since this study is focusing on crop growth, a threshold of 2 mm was considered to be suitable for supplying crops with water (Tadross *et al.* 2007).

A slight modification is applied to the commonly used *wet spell* criterion: additional to a five-daysum of greater

Table 1. Definition of criteria for identification rainfall characteristics.

Description	Variable	Criteria for characteristic	References
Onset of rainy season	on	First 10 days > 25 mm, no 10 consecutive dry days in next 20 days	Tadross *et al.* (2007), Mupangwa *et al.* (2011)
Cessation of rainy season	cess	3 consecutive 10-day periods each with less than 20 mm (after Feb 1st)	Tadross *et al.* (2007)
Duration of rainy season	dur	Cessation—onset date	–
Wet day	wd	Rain > 2 mm	Tadross *et al.* (2007)
Dry day	dd	Rain < 2 mm	Tadross *et al.* (2007)
Wet spell	wsp	5-day sum > 10 mm and less than 3 days with no rain	Usman and Reason (2004), Reason *et al.* (2005), *adapted*
Dry spell	dsp	5-day sum < 5 mm	Usman and Reason (2004), Tadross *et al.* (2007)
Max./mean wet/dry spell duration	$dspd_{max}/dspd_{mean}$ $wspd_{max}/wspd_{mean}$	No. of consecutive days with wet/dry spell criteria satisfied	–
Total rain in rainy season	p_{tot}	Rain sum from onset to cessation	–
No. of events above 10/20/40/60/80 mm	$p_{10}/p_{20}/p_{40}/p_{60}/p_{80}$	Rain > threshold	–

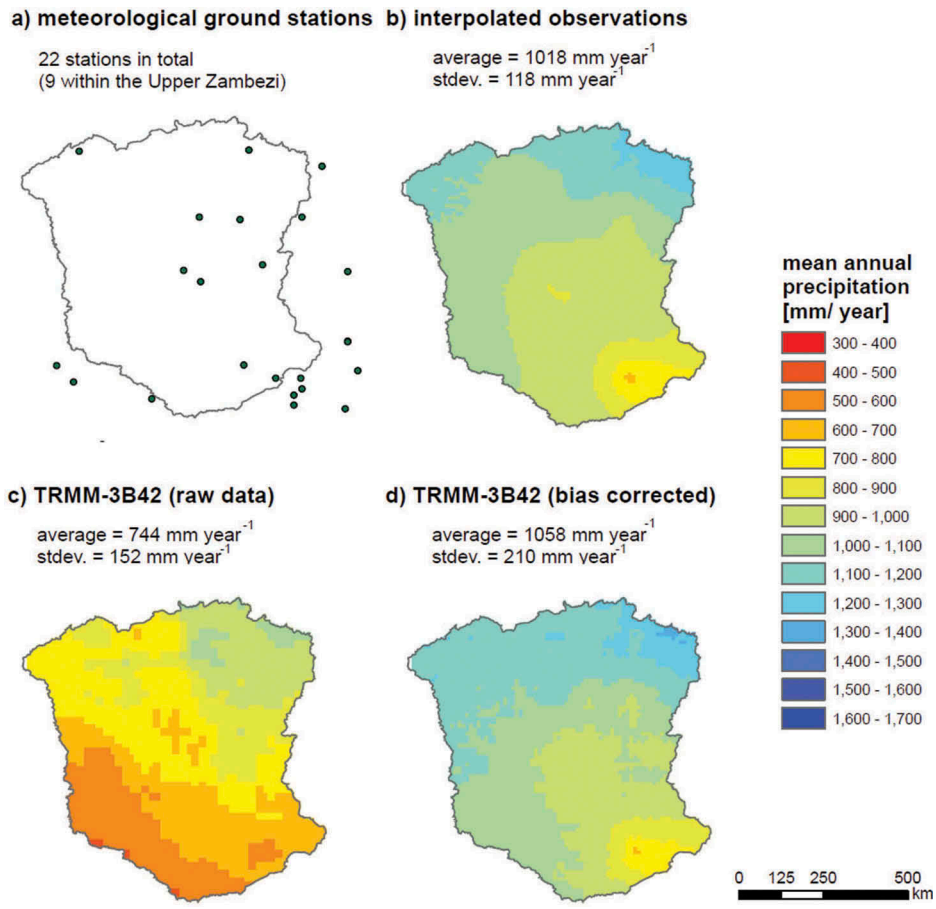

Figure 2. Station distribution (a) and mean annual precipitation (1998–2006), supplemented with information about average basin total and average standard deviation for the (b) KED interpolated observations, (c) original TRMM-3B42 and (d) the bias-corrected TRMM-3B42.

than 10 mm, a condition that three or more days out of these had to experience rainfall is introduced. This modification ensures that a wet spell is not created by only one or two single rain events. Durations for dry and wet spells are determined using a moving window with a length of five days. If the dry/wet spell criterion is fulfilled a duration of five is assigned to this particular spell. As long as this criterion remains satisfied when moving the window, one day is added to the dry/wet spell duration. The mean/maximum dry or wet spell can then be calculated for the rainy season under investigation.

The remaining rainfall characteristics will not be explained in detail here (refer to Table 1) since either previous authors are in agreement in regard to these criteria or there is no room for different interpretations.

Analysis of rainfall characteristics is performed for each year separately using the statistics programming software package R (R Core Team 2013). As a first step, onset and cessation dates are determined and thus the time frame for the rainy season is obtained. Subsequent characteristics are only analysed within this time frame (referred to as rainy season).

2.3 SVAT calibration and validation

Agricultural yields were obtained using the SVAT model DAISY (Abrahamsen and Hansen 2000, Hansen *et al.* 2012), which has been employed extensively for the simulation of physical and biological processes in agriculture. In the present work the one-dimensional version DAISY 4.01 was applied.

As input, the model requires: (a) a set-up file containing information on soils (soil hydraulic parameters), crops (variety, sowing/harvesting dates) and management practices (irrigation, tillage, fertilization); and (b) a weather file containing meteorological data (precipitation, global radiation and temperature).

Figure 1 visualizes how the different sources of input data were utilized within the research framework and which information was necessary for the above mentioned requirements for the DAISY model. Since the methodological approach is based on freely-available raster datasets, input files had to be created for each cell separately. The spatial resolution was chosen according to the resolution of the bias-corrected TRMM 3B42 v6 dataset (0.1° × 0.1°); hence a total of 4310 pairs of weather/management files

Table 2. Parameters chosen for calibration of DAISY and their corresponding default and calibrated values for maize.

Module	Parameter	Description (unit)	Default	Calibrated
Devel (phenology and development)	EmrTSum	Soil temp. sum at emergence (°C)	300.0	327.3
	DSRate1	Development in vegetative stage (d^{-1})	0.024	0.015
	DSRate2	Development in reproductive stage (d^{-1})	0.015	0.010
Leaf_phot (photosynthesis module)	Fm	Maximum assimilation rate (g CO_2 × $(m^2\ h)^{-1}$)	6.0	3.5
	Qeff	Quantum efficiency at low light ((g CO_2 × $(m^2\ h)^{-1}$)/(W × m^{-2}))	0.040	0.055
Canopy (canopy info)	DSLAI05	Development stage at crop area index = 0.5 (-)	0.250	0.251
	SpLAI	Specific leaf weight ((m^2 × m^{-2})/(g DM × m^{-2}))	0.0100	0.0083
	PARext	Photosynthetic active radiation extinction coefficient (-)	0.800	0.852
Root (root system)	maxPen	Penetration at emergence (cm)	120	95
Water_stress (effect of water stress)	y_half	Effect of water stress (-)	0.200	0.001

had to be prepared. Creation of these files for each cell was automated using precipitation data from the bias-corrected version of TRMM 3B42 v6, remaining meteorological data from ERA-Interim, the calculated rainfall onset date for each cell and year (defined as sowing date), and soil hydraulic properties, i.e. water retention parameters, saturated hydraulic conductivity and unsaturated hydraulic conductivity (Fig. 1). To obtain the latter, the software ROSETTA (Department of Soil, Water and Environmetal Science 2002) was applied using sand, silt and clay percentages and bulk density provided by ISIRIC-WISE. Rainfall onset dates were simply taken from the prior analysis of rainfall characteristics.

Since no information on the parameterization for the local maize variety was available, the initial simulation was carried out using the standard *Maize* from the library within DAISY.

Prior to the explanation of the calibration and validation procedure itself, important assumptions for the modelling approach are summarized as follows:

(1) Simulated yields for each year are calculated with a sowing date equal to the onset date that was calculated before. This implies that the modelled yields are based on an optimum decision of a farmer in terms of when to start sowing. As for other external influences on yields, such as elephants destroying the harvest, flooding, losses induced by pests or socio-economic aspects, which are infeasible to implement in a SVAT model, this introduces a great challenge for calibration of DAISY.
(2) The maize variety calibrated throughout all catchments is traditional maize (*Zea mays*), although in certain parts of SSA genetically modified varieties are grown.
(3) It is possible that the modelled maize is not actually grown in all areas (J. Mendelsohn, personal comment). The current study focuses on the main question "How and by what magnitude do different rainfall characteristics affect the successful growth of maize in the UZRB?" Therefore aspects such as flooding, wetlands or elevation are neglected.

The relevant parameters of the model were calibrated automatically using PEST (Parameter ESTimation, Doherty 2005), a software for parameter estimation and uncertainty analysis that is suitable for automatic calibration for a wide range of models. As PEST itself is not explained here in greater detail, the reader is referred to the cited literature for further information. Table 2 summarizes the parameters used for calibration in PEST.

In Table 2, the parameters of the main sub-modules of the vegetation module are represented (for details refer to Abrahamsen and Hansen 2000). These have been proven suitable as calibration parameters in related studies (Hansen *et al.* 2012, Seidel 2012). Especially the inclusion of the parameter WSE, an optional component regulating the effect of water stress on the crop was seen as crucial, since an effect of water stress on maize growth was expected in this semi-arid climate. Five out of the eight years with available data on measured yields from the Caprivi region were selected as calibration parameters, whereas the remaining data comprised a basis for validation.

2.4 Investigation of relationships between rainfall characteristics and modelled agricultural yields

For univariate analysis, simple regression was used. However multivariate relationships between yields and rainfall characteristics were investigated by applying a self-organizing map (SOM) approach.

The SOM method is an unsupervised artificial neural network developed by Teuvo Kohonen (1982). In brief, the

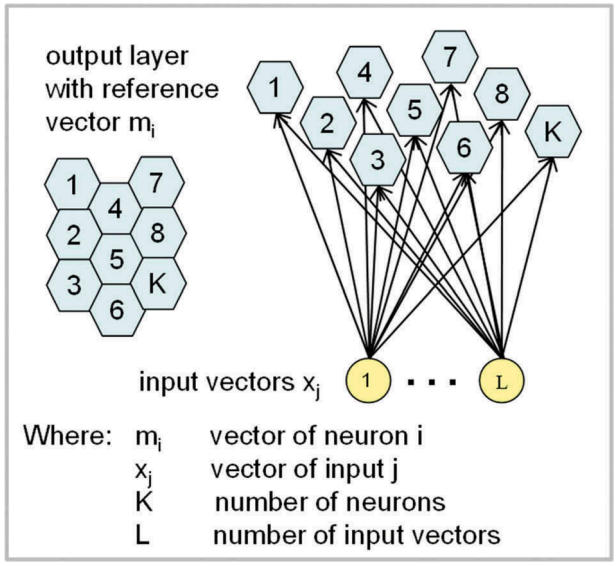

Figure 3. Structure of a self-organizing map (Wallner et al. 2013).

algorithm projects high dimensional input data into a lower, generally two-dimensional, output lattice (for details refer to Kohonen 1982, 1990, Wallner *et al.* 2013). Input data with similar characteristics are grouped in neighbouring regions of this output space. This output space is arranged in a hexagonal structure, as depicted in Fig. 3.

The main advantage of SOM is the ability to provide a user-friendly representation of large, multi-dimensional datasets; hence, similarities and relationships can directly be derived by interpreting the two-dimensional output space and the associated component planes.

In the present research, the trained SOM is visualized in three different ways:

a. component planes, illustrating the distribution of the particular variables on the map;
b. the unified distance matrix (U-matrix) showing the neuron population based on the input dataset and distances between neighbouring neurons (Rivera *et al.* 2012), and
c. k-mean-clustered U-matrix dividing the input into six groups (based on the identifier 'yield' from marginal to very high).

The size of the SOM is mainly set based on the character of its input data. According to López García and Machón González (2004), the number of output neurons is equal to five times the square root of the input data. This is implemented and automated within the SOM toolbox in Matlab (Kohonen *et al.* 1995, Vesanto *et al.* 2000). For the purpose of this research, the algorithm generated a SOM with 91 neurons (7 × 13) after external averaging of yields and rainfall characteristics for each of the regions within the study area.

3 Study area

With a total area of 1.37 million km^2, the Zambezi River Basin is the fourth largest river basin in Africa and the largest in the Southern African Development Community (WorldBank 2010, Thiemig *et al.* 2013) . Its location is in southern Africa between 9°–20°S and 18°–36°E. The basin is shared by eight countries (Angola, Zambia, Namibia, Botswana, Zimbabwe, Malawi, Tanzania and Mozambique) and hence has a high transboundary importance in regard to water resource management, not only because of the different contributions of these sub basins to the overall runoff of the Zambezi River, but also because of the very high variability of climate in space and time and major differences in sub-basin characteristics. The Upper Zambezi River Basin, on which this research focuses, compromises the basin upstream of Victoria Falls (Fig. 4) and has a total size of 514 000 km^2 (WorldBank 2010).

Annual rainfall varies between 500 mm in the southern part of the catchment to up to 1400 mm in the northern Zambian and Angolan regions (Euroconsult MottMacDonald 2007). The rainy season shows a very high seasonal, inter-annual and 10-day variability, which implies a challenge for water resources management (Tyson *et al.* 1975, New *et al.* 2006, Conway *et al.* 2009, Goulden *et al.* 2009). In the past, extended drought and flood periods affected the whole basin: the years 1907–1945 are characterized as a dry period, whereas 1946–1964 showed an above normal runoff behaviour with extensive flooding (WorldBank 2010). From 1981–1997 the basin experienced a severe drought that caused the drying

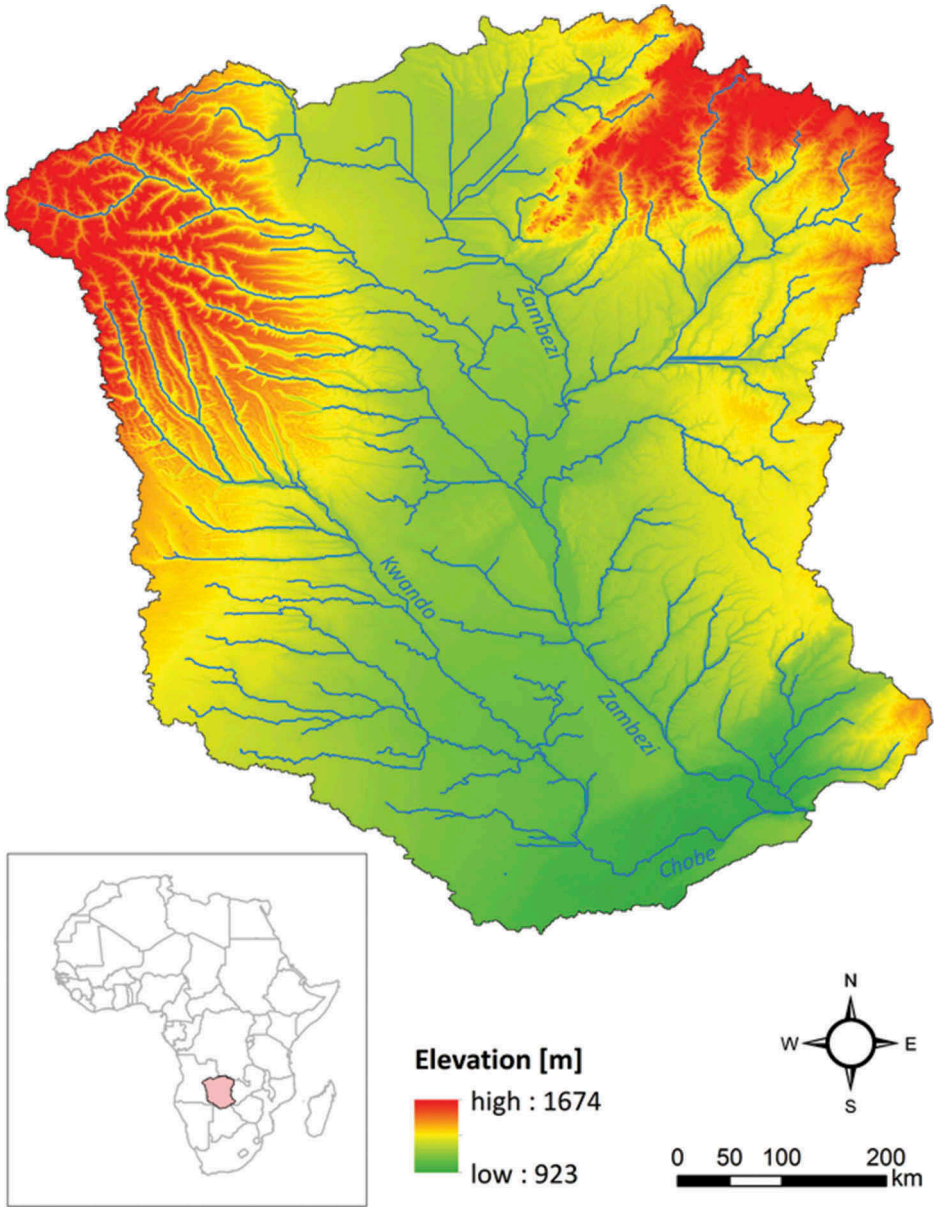

Figure 4. The Upper Zambezi River Basin (UZRB).

up of wetlands and floodplains (which are usually flooded yearly). From 1998 up to now, major extended flooding has occurred, indicating the presence of a new wet period. For agricultural outputs and livestock production, rainfall reliability is the over-riding issue. In the UZRB, there is almost no irrigated agriculture developed yet, which stresses the dependence on rainfall even more.

Temperature across the basin varies highly with elevation (ranging between 800 m in the southern parts to 1500 m in northern Zambia) although there are slight differences between different latitudes as well (Euroconsult MottMacDonald 2007). Finally, from a geological perspective, Kalahari sands (deposited and non-deposited) are present throughout the whole basin (WorldBank 2010). It is believed that in most areas these are underlain by basalt stone (BGR 2005).

4 Results

4.1 Rainfall characterization

Statistics of the rainfall characteristics are presented here in two ways: first, as mean maps for administrative regions over the whole period of investigation, and second, for each year and cell separately. Local

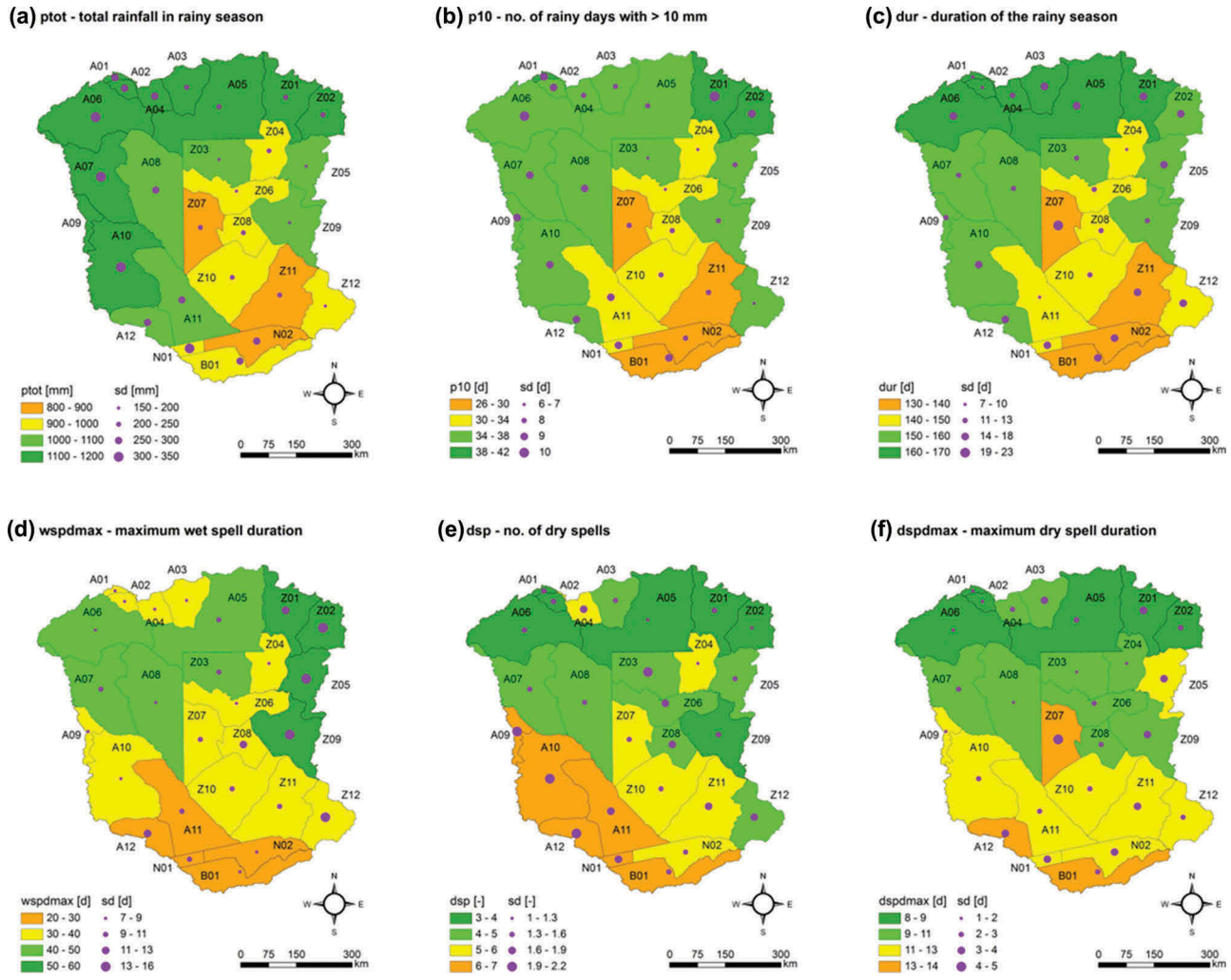

Figure 5. Mean and standard deviation (sd) of selected rainfall characteristics averaged over administrative regions for the years 1998–2010; (a) p_{tot}—total rainfall in rainy season, (b) p_{10}—no. of rainy days with >10 mm, (c) dur—duration of the rainy season, (d) $wspd_{max}$—maximum wet spell duration, (e) dsp—no. of dry spells, $dspd_{max}$—maximum dry spell duration.

decisions in SSA are often made based on the administrative region or even smaller units; a presentation in this way therefore makes great sense. The second way of analysis is more suitable for an in-depth analysis of regional patterns or distinctive features of a particular rainy season. The taxonomy used in the interpretation of results is ordered as a letter representing the country a region belongs to (A—Angola, B—Botswana, N—Namibia and Z—Zambia) followed by a region identifier.

4.1.1 Mean over regions

As shown in Fig. 5, most rainfall characteristics reveal a northwest to southeast gradient. Rainy season totals are the highest in Angola and the north of Zambia (see Fig. 5(a)). However, the western Angolan part of the catchment is also the one with the greatest variability. The northern areas of Angola and Zambia (A03—Lucano; A05—Alto Zambeze; Z01—Mwinilunga and Z02—Solwezi) are the most stable in terms of total rainfall, with high annual amounts and low variability. The central part of the UZRB, in contrast, displays less favourable conditions. Along the flat plains of Barotse, the Namibian parts of the basin and also Botswana, total rains are notably lower, while the variability continues to increase the further one looks towards the south. Despite this, mean rain amounts from the detected onset to cessation would allow the successful cultivation of maize throughout the basin (according to FAO, maize needs 500–800 mm of water depending on the climate; http://www.fao.org/nr/water/cropinfo_maize.html). However, the high coefficients of variation (cv) of up to 0.40 (i.e. administrative region N01) and the fact that the recent period has been particularly wet should be noted. Nevertheless, looking at

Table 3. Mean values and standard deviations (1998–2010) of selected rainfall characteristics averaged over administrative regions.

ID	Region	on	cess	p_{tot}	p_{10}	p_{40}	dsp	$dspd_{max}$	wsp	$wspd_{max}$
A01	Camanongue	297 ± 12	96 ± 9	1156 ± 277	38 ± 9	3 ± 3	4 ± 1	8 ± 1	10 ± 3	39 ± 9
A02	Leua	297 ± 10	94 ± 9	1131 ± 286	36 ± 9	3 ± 3	4 ± 2	8 ± 2	11 ± 2	38 ± 8
A03	Lucano	295 ± 10	91 ± 14	1103 ± 231	35 ± 7	4 ± 2	5 ± 2	10 ± 3	12 ± 2	37 ± 8
A04	Cameia	295 ± 10	92 ± 11	1129 ± 254	35 ± 7	5 ± 3	5 ± 2	11 ± 3	12 ± 2	36 ± 8
A05	Alto Zambeze	297 ± 9	91 ± 16	1103 ± 225	37 ± 8	3 ± 1	4 ± 1	9 ± 2	10 ± 1	47 ± 10
A06	Moxico	298 ± 10	93 ± 11	1139 ± 305	37 ± 10	3 ± 2	4 ± 1	8 ± 2	9 ± 2	44 ± 8
A07	Luchazes	299 ± 9	93 ± 10	1117 ± 305	36 ± 9	3 ± 3	5 ± 2	10 ± 2	9 ± 2	43 ± 10
A08	Lumbala Nguimbo	302 ± 6	90 ± 11	1038 ± 268	34 ± 9	3 ± 2	5 ± 1	10 ± 2	10 ± 1	40 ± 8
A09	Cuito Cuanavale	301 ± 10	93 ± 11	1155 ± 389	36 ± 9	4 ± 4	6 ± 2	11 ± 2	10 ± 2	36 ± 8
A10	Mavinga	303 ± 7	90 ± 12	1102 ± 317	36 ± 9	4 ± 3	6 ± 2	12 ± 3	10 ± 2	33 ± 9
A11	Rivungo	307 ± 7	87 ± 10	1016 ± 293	33 ± 9	4 ± 3	6 ± 2	12 ± 2	10 ± 2	30 ± 9
A12	Dirico	305 ± 8	91 ± 14	1063 ± 290	34 ± 9	4 ± 3	7 ± 2	14 ± 3	11 ± 2	29 ± 12
B01	Botswana	311 ± 8	83 ± 17	911 ± 280	29 ± 9	4 ± 2	7 ± 1	13 ± 3	10 ± 2	25 ± 7
N01	Kavango	309 ± 11	88 ± 19	981 ± 302	31 ± 9	5 ± 3	7 ± 2	12 ± 3	10 ± 3	26 ± 10
N02	Caprivi	312 ± 8	77 ± 18	808 ± 256	27 ± 8	3 ± 2	6 ± 1	13 ± 3	9 ± 1	27 ± 8
Z01	Mwinilunga	298 ± 12	92 ± 14	1146 ± 234	40 ± 10	2 ± 1	3 ± 1	8 ± 3	8 ± 2	56 ± 12
Z02	Solwezi	301 ± 9	92 ± 10	1199 ± 238	41 ± 9	3 ± 2	3 ± 1	9 ± 3	7 ± 2	60 ± 15
Z03	Zambezi	301 ± 9	87 ± 10	1051 ± 191	35 ± 6	3 ± 2	4 ± 2	10 ± 2	10 ± 2	41 ± 10
Z04	Kabompo	308 ± 7	87 ± 5	1049 ± 198	36 ± 8	2 ± 1	4 ± 2	10 ± 3	8 ± 2	52 ± 15
Z05	Mufumbwe	305 ± 9	93 ± 7	1070 ± 191	37 ± 8	3 ± 1	4 ± 1	11 ± 3	8 ± 2	51 ± 16
Z06	Lukulu	306 ± 6	88 ± 7	994 ± 177	33 ± 6	3 ± 1	5 ± 2	10 ± 3	10 ± 1	39 ± 9
Z07	Kalabo	308 ± 11	77 ± 19	953 ± 237	31 ± 7	3 ± 2	5 ± 1	10 ± 2	10 ± 1	35 ± 7
Z08	Mongu	307 ± 7	86 ± 7	976 ± 233	33 ± 8	3 ± 2	5 ± 2	10 ± 2	9 ± 2	39 ± 12
Z09	Kaoma	304 ± 7	92 ± 7	992 ± 185	34 ± 7	3 ± 1	5 ± 2	11 ± 3	9 ± 2	40 ± 14
Z10	Senanga	308 ± 8	87 ± 8	967 ± 240	32 ± 8	3 ± 2	5 ± 2	11 ± 3	9 ± 2	34 ± 10
Z11	Sesheke	308 ± 7	80 ± 14	856 ± 223	29 ± 8	2 ± 1	5 ± 2	12 ± 3	9 ± 2	33 ± 10
Z12	Kalomo	307 ± 11	88 ± 8	807 ± 245	27 ± 8	2 ± 2	5 ± 1	13 ± 5	9 ± 2	32 ± 11
UZRB		303 ± 9	89 ± 11	1037 ± 254	34 ± 8	3 ± 2	5 ± 2	11 ± 3	10 ± 2	39 ± 10

rainfall totals throughout the rainy season alone does not explain the low yields in most parts of the catchment. Maize is described as sensitive to water logging and, in the flowering and, yield formation period, also prone to water stress (FAO, http://www.fao.org/nr/water/cropinfo_maize.html); thus, the intraseasonal distribution of rainfall is a crucial aspect and is focused on subsequently.

The regional pattern of the number of days with more than 10 mm accumulated rainfall (p_{10}; Fig. 5(b)) and duration (dur; Fig. 5(c)) of the rainy season appears quite similar to that of total rainfall (p_{tot}). The former characteristic is commonly referred to as 'productive rainfall' and therefore of great importance for plant growth (Mendelsohn *et al.* 2013). As for p_{tot}, the northern and western parts of the catchment experience a mean of at least 34 days with more than 10 mm and duration of at least 150 days. These are the highest means throughout the basin. However, variability is considerable even in these areas (Fig. 5(a)–(c)). The central region displays a pattern similar to that of p_{tot}—shorter duration of the rainy season and a lower number of p_{10} compared to the north and northwest. The southernmost parts of the basin (N01, N02 and B01), as well as parts of Zambia (Z07 and Z11), show the least favourable conditions with short growing seasons and numbers of productive rains. At the same time, the variability in these areas is the highest for all rainfall indices.

Characteristics related to the consistency of the rainy season (i.e. number and length of wet and dry spells), as shown in Fig. 5(d)–(f), display different spatial patterns. A clear north–south gradient is apparent dividing the catchment into two, nearly equal-sized portions with the exception of the regions A01 to A04. The northern part of the basin is characterized by a high number of wet spells (wsp) and long maximum wet spell durations ($wspd_{max}$; Table 3 and Fig. 5(d)). This implies more favourable conditions for rain-fed crops than in the southern part, where the opposite is the case. At the same time, however, the regions with a very high $wspd_{max}$ are most likely those where the big floods such as in 2009, 2010 and 2011 were created. An interesting pattern is revealed if one looks at the variability of $wspd_{max}$: whereas the northern Angolan areas are more consistent, variability of this particular characteristic is very high (sd of 13–16 days) in northeastern Zambia. These particular areas (Z01, Z02, Z05, Z09) in the head-catchments of the Zambezi River might have a major impact on the magnitude and frequency of flooding downstream (i.e. in Caprivi). In terms of rain-fed agriculture, the long wet spells in the north could also be harmful for those crops being sensitive to water logging. In regard to dry spell characteristics (dsp and $dspd_{max}$; Fig. 5(e), (f), the northern half of the UZRB is less prone to droughts within the rainy season. A dry spell duration of 10 days, as identified even in the most consistent regions, could still be

sufficient to cause significant losses of harvest. In the south of the catchment, preconditions for rain-fed agriculture appear to be less favourable (Table 3 and Fig. 5(d)–(e)). As for the previously discussed indicators, a "dry corridor" in the central part of the UZRB is clearly visible. Additionally to the Namibian, Botswanan and southern Zambian provinces, parts of Angola (A10 to A12) also experience short (<40 days) maximum wet spells, a high number (>5 per rainy season) of dry spells and a long maximum dry spell duration (>11 days) based on average numbers. Most of these areas display a high variability at the same time, making them the most vulnerable for rain-fed farming.

4.1.2 Region-specific analysis

For a detailed analysis of a particular year or region, the annual maps of the rainfall onset dates (on), number of rainy days above 20 mm (p_{20}) and maximum wet spell duration (wspd_{max}) for an average (2001/02), dry (2004/05) and wet (2007/08) rainy season were chosen for in-depth analysis (Fig. 6). The statements from the analysis of regional means might be true for characterizing the average behaviour over one region; however, all characteristics vary greatly in both time and space if investigating on smaller scales. It becomes clear that variability is the overriding issue within most areas of the basin. The rainy season 2001/02 (average in terms of total rainfall), for instance, displays a homogeneous pattern of the onset of rains over the whole UZRB, whereas in 2004/05 (dry) and 2007/08 (wet) the spatial distribution is highly heterogeneous (Fig. 6(a)). Similarly, the difference of p_{20} in time and space is enormous (Fig. 6(b)). In 2001/02, major disparity between northern and southern parts of the catchment is visible. In contrast, the spatial pattern in 2004/05 and 2007/08 is more or less consistent, but the amount of p_{20} differs significantly. The pattern of wspd_{max} (Fig. 6(c)) reveals a gradient from east to west in this characteristic rather than from north to south, as is the case for most other indices. The illustration (Fig. 6) points out how challenging agricultural planning is for local farmers. An early onset of the rains does not necessarily coincide with a good rainy season in terms of total rainfall. This stresses the importance of a prediction system for rainy season onset and related characteristics within the UZRB.

4.2 Spatio-temporal analysis of yields

The scatter-plot in Fig. 7 shows the results for calibration and validation of DAISY.

With the exception of one year (2008), modelled yields are generally in agreement with the measured values. Considering uncertainty introduced by the remotely sensed information that was used, calibration results are satisfactory. A higher availability of data for calibration and validation perhaps could have improved the results. The model was calibrated mostly in a region with generally low yields (compare Fig. 8). In order to calibrate these low yielding years (i.e. the rainy seasons of 2002 and 2003), the automated calibration resulted in a combination of parameters reacting very sensitively to dryness. This might have caused the low modelled yields over certain regions of Zambia, particularly in 2008. Due to these facts, uncertainty in the results has to be considered and is acknowledged here explicitly. In terms of suitability for simulation of plant-growth in semi-arid areas, DAISY showed a good performance. Losses induced through drought stress were captured by the model accurately and growth of maize as well as harvesting dates was reproduced realistically.

Generally, yields throughout the UZRB are very low and variability is extremely high. Mean modelled yields over the basin, as depicted in Fig. 8, reveal a clearly decreasing gradient from north to south, which is in agreement with the pattern for many rainfall characteristics. The area with the highest yields is Solwezi (Z02) on the northeastern border of the UZRB, reaching mean yields of 1.1–1.2 t/ha. Analysis of rainfall characteristics shows that Solwezi is the region with the highest mean rainfall during the rainy season (p_{tot}), highest "productive" rain events (p_{10}), least dry spells (dsp) and the longest maximum wet spell durations (wspd_{max}). However, variability in this region is notable with a standard deviation greater than 0.3 t/ha and, thus, a cv of slightly above 0.25. Lowest harvests, on the other hand, were modelled for regions B01 (Botswana), N02 (Caprivi) and Z11 (Sesheke), with mean yields not exceeding 0.6 t/ha. Absolute variability in these areas (Fig. 8) is low, but taking into account the cv, variability is in ranges up to 0.5 (i.e. region B01). Regions B01, N02 and Z11 can thus be categorized as the most vulnerable regions; hence, these are least suitable for rain-fed agriculture. Often, high sd and cv of harvests are related to a high degree of variability in one or more of the rainfall characteristics. For example, Z12 (Kalomo) has the most unreliable rainy season in terms of the characteristic dur, with a mean of 135 days and sd of 23 days, which might cause the high cv of 0.38 for modelled yields. Clearly noticeable are also two small areas in the north of Angola, A01 (Camanongue) and A02 (Leua). Surrounded by regions with relatively high yields, mean harvests in these regions are significantly lower. Analysing the rainfall characteristics, it becomes obvious that in A01 and A02 the dd (number of dry days) is very high; at the same time, the mean dsp

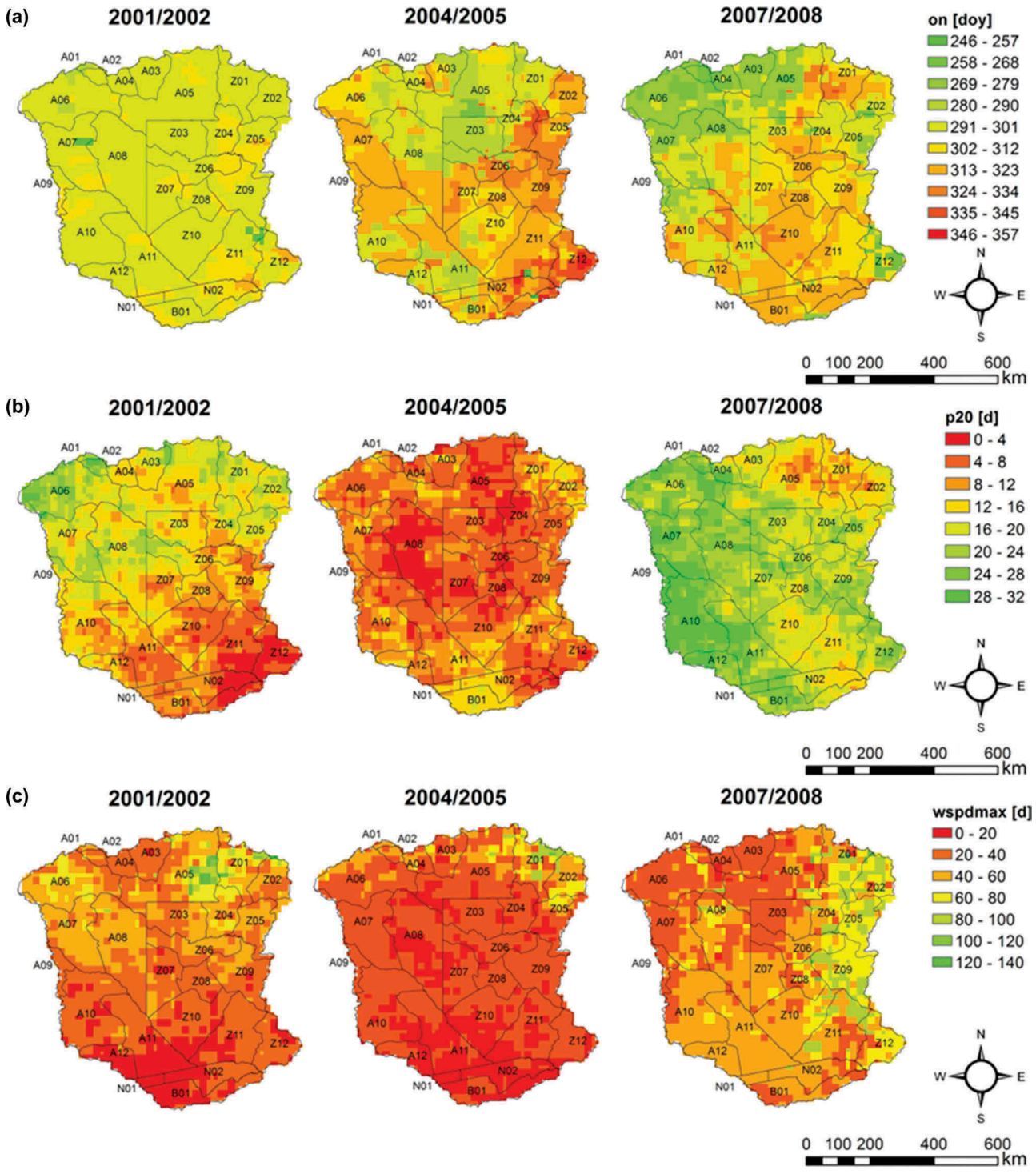

Figure 6. Detailed maps of (a) rainfall onset, (b) number of rainy days with more than 20 mm and (c) maximum wet spell duration for the rainy seasons of 2001/02 (P_{tot} = 1009 mm), 2004/05 (P_{tot} = 718 mm) and 2007/08 (P_{tot} = 1388 mm).

(number of dry spells) is low. This indicates that rains in both regions are more erratic, i.e. wet and dry days alternate, extensively affecting yields negatively. Most stable yields are obtained in the northern part of the catchment: A04 (Cameia), A07 (Luchazes) and Z01 (Mwinilunga) achieve the lowest coefficients of variation with less than 0.25.

In Fig. 9, a detailed map for region N02 (Caprivi) over the study period is shown. Here, the interplay between rainfall, flooding and agriculture is particularly complex and challenges are vast. Both the Zambezi and Chobe rivers can leave up to one third of the Caprivi flooded bringing both a blessing (fishing and recession agriculture) and a curse (destructive

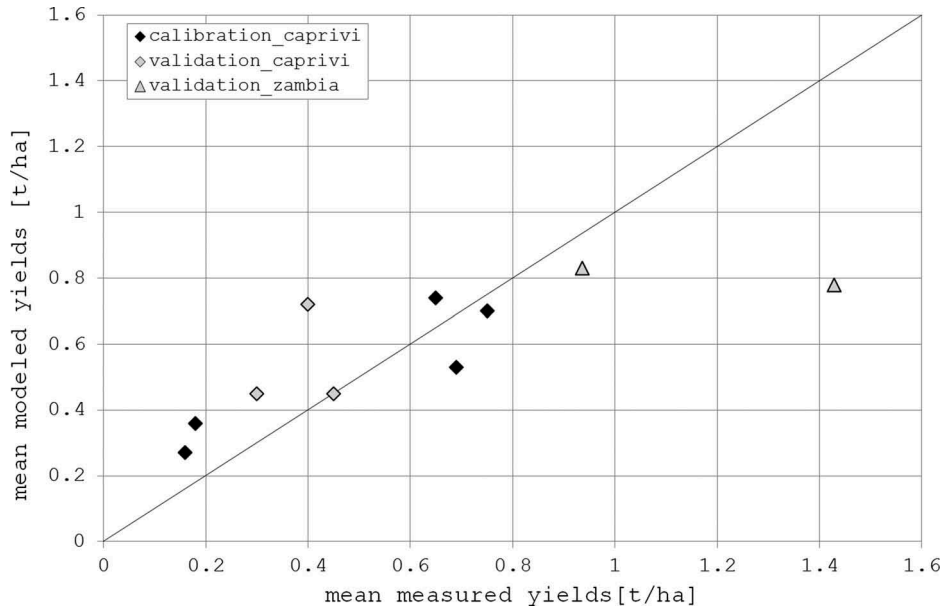

Figure 7. Mean measured *vs* modelled maize yields.

character of floods) to the region, depending on the rainfall characteristics in their head catchments. Furthermore, the elephant density here is significant, creating an additional problem for farmers. As revealed by the image, even in good years the maize yield does not exceed 1 t/ha. In recent years, yields were slightly higher due to a series of good rainy seasons. However, both spatial and temporal variability is great and needs to be considered additionally. No temporal trends or spatial consistencies can be identified which leads to the conclusion that both rainfall characteristics and yields can be considered as somewhat random. Increased climatic variability (Christensen *et al.* 2007) is very likely to worsen this issue. In Mendolsohn and Roberts (1997), soils in the Caprivi are defined as "moderate to good". However, if not considering additional factors such as the ones mentioned above, this could lead to wrong decision-making. The results further reveal the controversy that in a region with generally high availability of water people are not able to cultivate their crops successfully. It is in these areas, where adaptation and improvement are achievable with relatively low cost (i.e. rain/flood water harvesting, channel systems).

4.3 Univariate analysis of yields vs rainfall characteristics

The univariate analysis was carried out in order to identify single rainfall characteristics affecting yields most, without taking into account interactions with others. For the regression analysis, mean modelled yields over each administrative region for the whole period of investigation were plotted against all characteristics. In a previous attempt (not shown here), all 4310 cells were used; however, the scattering did not result in any significant correlation due to the high amount of single outliers (i.e. data errors in the TRMM data for single cells) or local anomalies. In Fig. 10, a scatter-plot of selected rainfall characteristics is illustrated. For further validation of modelled yields, the analysis was also conducted using measured data from the regions where data were available. These results, together with the correlation coefficients of rainfall characteristics, are summarized in Table 4.

In Fig. 10 it is clearly visible that the three indicators in the lower part of the graphic (p_{tot}, $dspd_{max}$ and $wspd_{max}$) are less spread; therefore, these have a higher impact on yields than the ones in the upper part (dur, dsp and wsp). In general, characteristics describing "duration" show higher correlation than those describing "frequency". The highest correlation of modelled yields are with $dspd_{max}/dspd_{mean}$ (−0.56), p_{10} (0.51), $wspd_{max}/wspd_{mean}$ and p_{tot} (0.48 for all three). Nevertheless, no investigated relationship between the criteria and crop yield reached a correlation coefficient greater than 0.80; hence, these results indicate that any unique characteristic can potentially be used as a predictor for the outcomes of rain-fed agriculture. It further strengthens the statement that the total rainfall amounts per rainy season do not determine the success

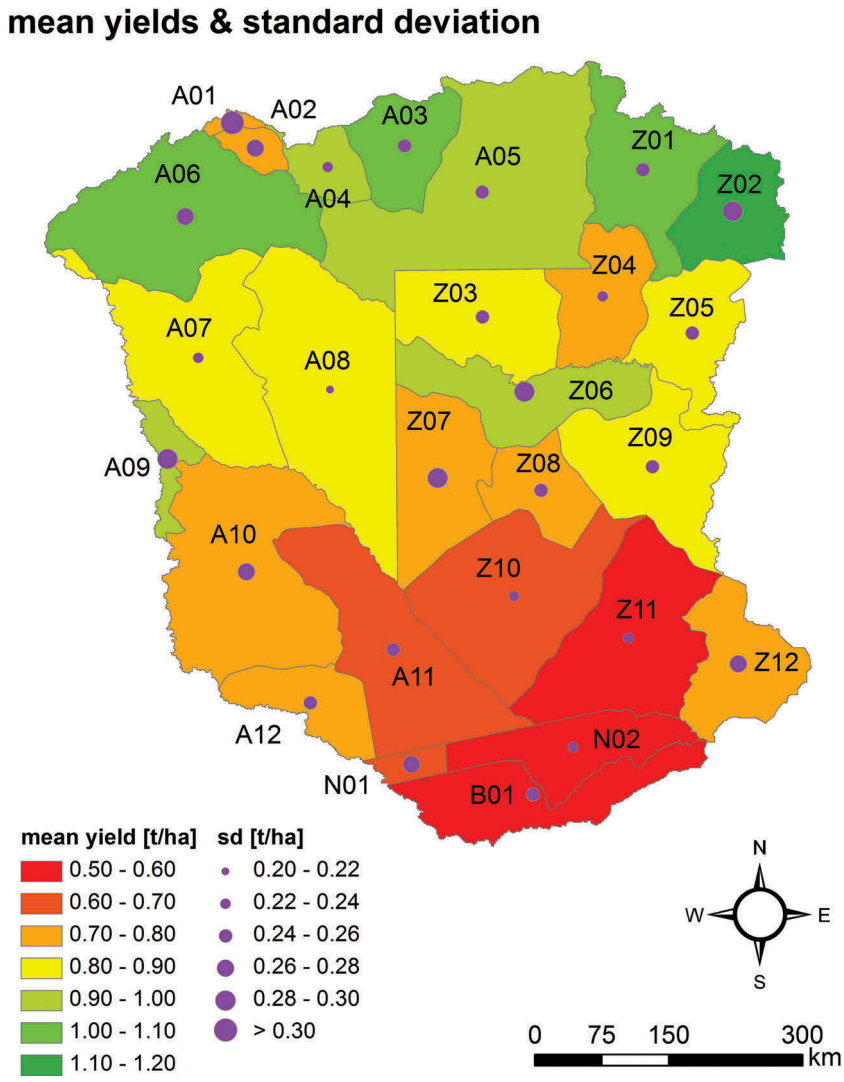

Figure 8. Mean and standard deviation (sd) of modelled yields for the studied regions for the period 1998–2010.

of rain-fed farming, but an interplay of several indicators. Unexpected, but easily explained can be the high correlation for dry days (0.36): characteristic dd is strongly related with dur; for this reason, a longer rainy season automatically means more dry and wet days, spells and total rainfall. Unlike dry spells, single dry days seem to have no negative effect on yields. Characteristics such as wd, wsp and extreme events (p_{40}, p_{60} and p_{80}) show no strong interconnection with yields. In particular, the low, even negative correlations for wd and wsp are surprising. If one compares the correlations of wet days (–0.01) and the number of rainy days above 10 mm (0.51), it becomes obvious that the selected threshold of 2 mm for a wet day is not sufficient in relation to agricultural yields. This result strengthens the assumption of defining rainy days above 10 mm as "productive rainfall events" (Mendelsohn *et al.* 2013). However, it might be interesting to investigate this criterion further, since it is possible that a strong correlation appears by using 4, 6 or 8 mm as the threshold. The low relationship of wsp indicates that it is not the number of wet spells, but rather their duration that has a strong impact. For the extreme events, a rather negative impact on yields would be expected if their number increases. However, this is not the case, indicating either there is no significant effect or that the model does not react to it.

The comparison of rainfall characteristics with measured yields resulted in similar findings for most indicators (Table 4). Whereas correlations of the "wet" characteristics with yield show similar magnitudes, stronger connections to agricultural outputs are found for p_{tot}, p_{10}, p_{20}, p_{40}, dur and dd. Major differences between modelled and measured correlations with harvest are encountered in $dspd_{max}/dspd_{mean}$. Whereas yields in the simulation react sensitively to

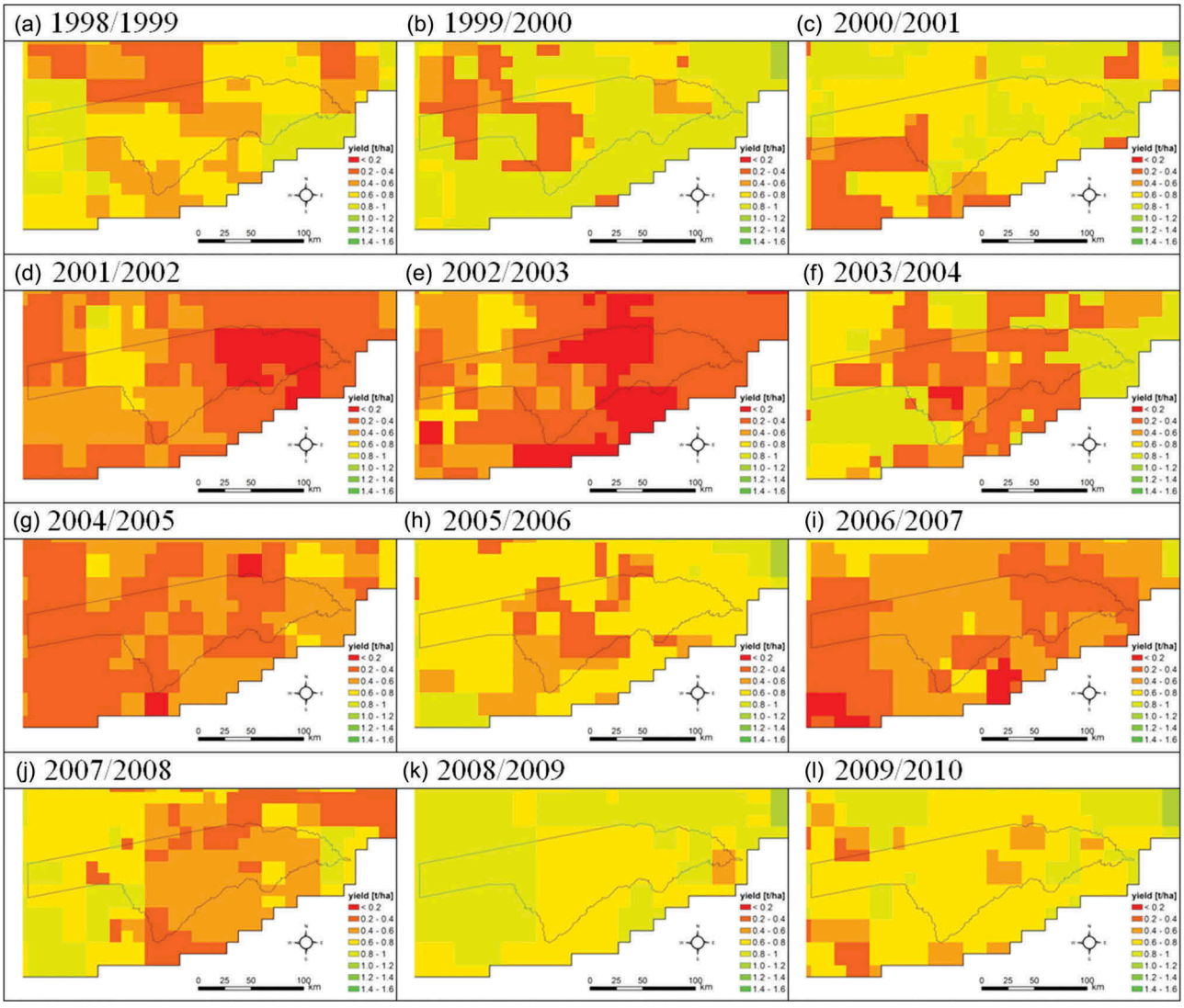

Figure 9. Distribution of modelled yields for the years 1998–2010 over Caprivi (N01).

these indicators, the relationship is much lower, in reality. A possible explanation for this could be that, through the parameterization of the model, results are strongly affected by water stress and, therefore, longer dry spells have a greater effect. It seems also possible that some kind of manual irrigation is carried out on the fields during long dry spells, or soils in some parts of the study area are moister than modelled through flooding.

Table 4. Comparison of measured and simulated yields for calibration and validation of DAISY. rxy is the correlation coefficient.

Rainfall characteristic	rxy modelled	rxy measured
p_{tot}	0.48	0.64
dur	0.36	0.57
wd	−0.01	0.14
dd	0.34	0.56
dsp	−0.45	−0.44
$dspd_{max}$	−0.56	−0.16
$dspd_{mean}$	−0.56	−0.08
wsp	−0.08	−0.07
$wspd_{max}$	0.48	0.54
$wspd_{mean}$	0.48	0.50
p_{10}	0.51	0.66
p_{20}	0.33	0.61
p_{40}	0.20	0.39
p_{60}	0.15	0.17
p_{80}	0.13	0.08

4.4 Multivariate analysis of yields vs rainfall characteristics

In this study, the input data comprises of 16 selected rainfall characteristics plus the modelled yield for each cell and year. During the training of the SOM, similar input data would then be grouped into common neighbourhoods in the output space. The component planes obtained from the SOM algorithm and the results of the subsequent clustering are visualized in Fig. 11. Six

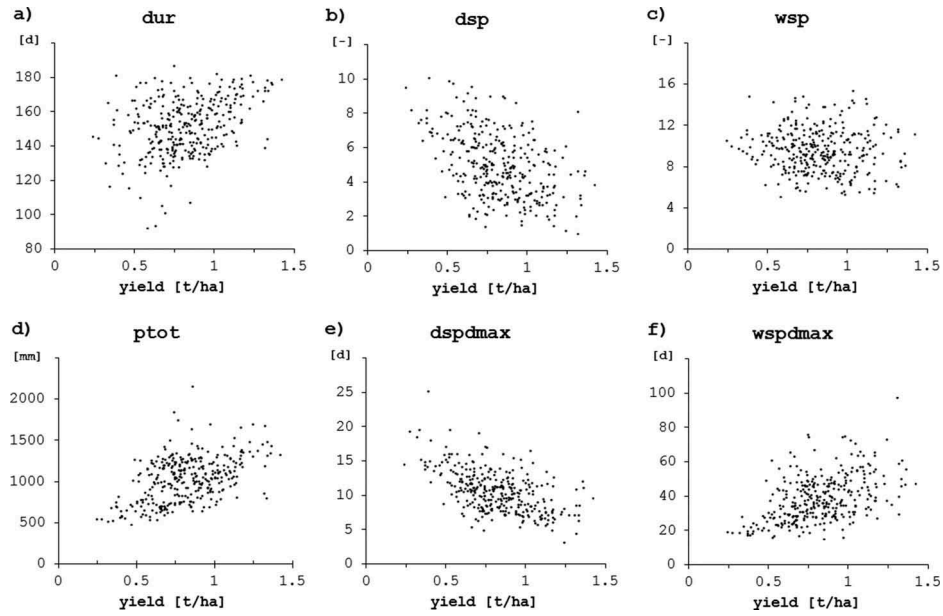

Figure 10. Scatter-plots of selected rainfall characteristics *vs* yield.

clusters were identified by the k-means algorithm. Combining the information of the component planes, these classes are characterized as follows:

(1) *Marginal yield*: short rainy season; relatively high number of wet and dry days; very high number of wet and dry spells; very long wet spell duration; short wet spell duration; very low number of rainy days above 10 mm; very low number of extreme events; very low total rainfall;

(2) *Very low yield*: short rainy season; low number of dry days; high number of wet and dry spells; dry spell duration in medium range; low number of rainy days above 10 mm; low number of extreme events; short wet spells; low total rainfall;

(3) *Low yield*: short rainy season; low number of dry days; low number of dry days; medium number of dry spells; low number of rainy days above 10 mm; high number of extreme events; medium wet spell duration; total rainfall in medium range;

(4) *Normal yield*: short rainy season; low number of dry days; low number of dry spells; medium number of wet spells; short dry spells; medium number of rainy days above 10 mm; medium number of extreme events; medium duration of wet spells;

(5) *High yield*: mid-range length of rainy season; low number of wet days; low number of dry spells; short dry spells; long wet spells; high number of rainy days above 10 mm; high number of rainy days above 20 mm; medium number of extreme events above 40 mm/d; high total rainfall; long wet spells and

(6) *Very high yield*: long rainy season; high number of wet and dry days; very low number of dry spells; very short dry spells; very long wet spells; very high numbers of rainy days above 10 mm; high number of events above 60 mm; high total rainfall.

Although purely qualitative, this classification allows an understanding of the combined effects of rainfall characteristics on yields to be gained. Quantitative information can be obtained by studying the component planes in detail.

One aspect revealed by the component planes is the great importance of dur (Fig. 11(b)). According to the image, "good" maize yields (classes 5 and 6) can only be achieved with a rainy season longer than 160 days. In all other classification units, dur varies only slightly; hence, differences in yields are then determined by differences in other characteristics. In particular, differences in dsp, $dspd_{max}$, $wspd_{max}$ as well as, to a certain degree, p_{tot} and p_{10}, define the maize yield (Fig. 11). In addition to a short rainy season, the two lowest yielding groups (clusters 1 and 2) are further described by many but short wet spells, and many but long dry spells. Surprisingly, the number of wet days (wd) is relatively high in the low yielding clusters; hence (as shown in Section 4.3), it does not correlate at all with

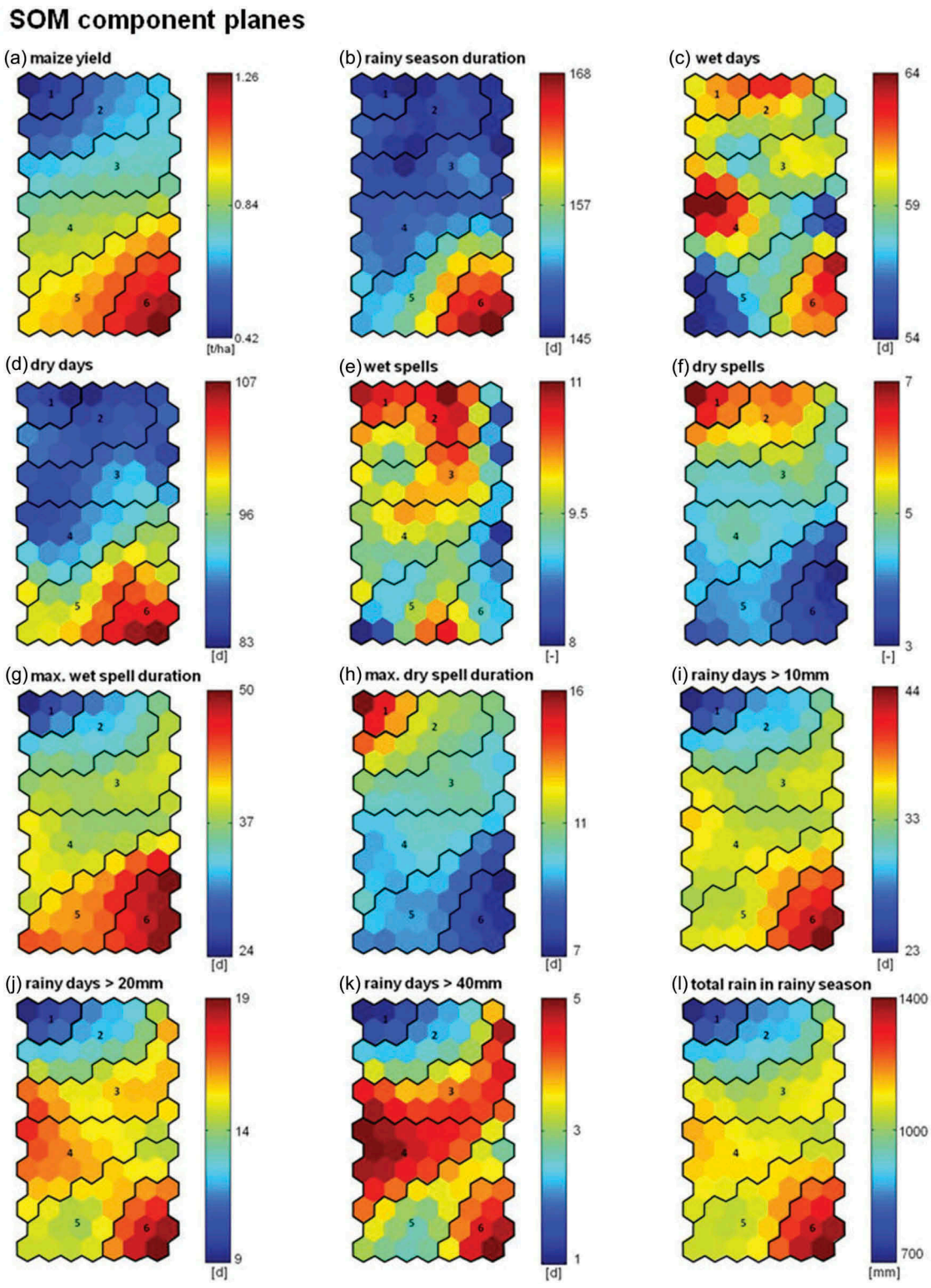

Figure 11. Component planes for yields and selected rainfall characteristics derived by the SOM algorithm. Clusters are indicated as (1) marginal yield, (2) very low yield, (3) low yield, (4) normal yield, (5) high yield and (6) very high yield.

yields. In comparison, the component plane for p_{10} (Fig. 11(i)) shows a distribution that is in greater agreement with maize yields. Again, this indicates a threshold of 2 mm for a rainy day is not sufficient in relation to rain-fed agriculture. The distribution of extreme events (only p_{20} and p_{40} shown here) reveals an interesting pattern. High and very low yielding clusters show a clear distribution; in contrast, a high number

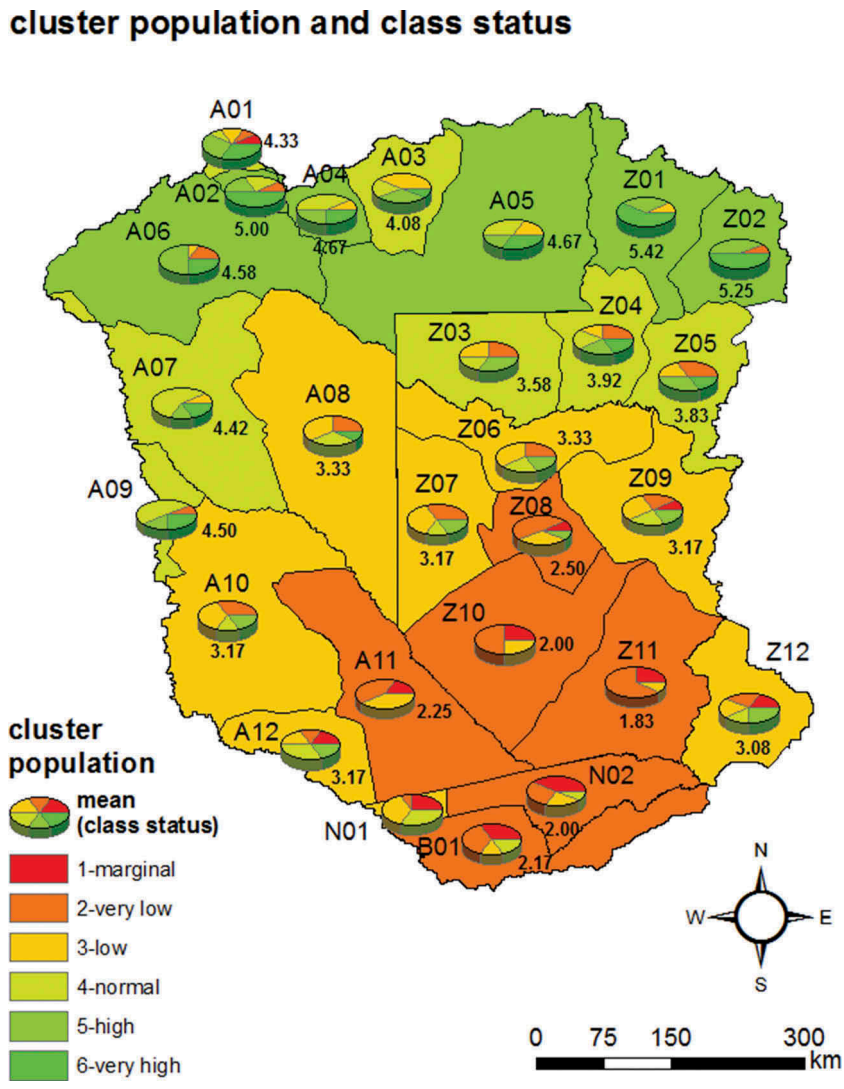

Figure 12. Class population for each cluster per region and class status (mean class population). The pie chart visualizes the class population distribution for each region between 1998 and 2010. The background colour and the numbers refer to the class status.

of extreme events is also encountered in clusters 3 and 4. Although a clear explanation for this remains difficult, there might be a relation with the higher magnitudes of wsp and dsp in combination with shorter wet spells (Fig. 11(g)) in these clusters. This would be in agreement with the previous finding that many changes of wet and dry spells affect yields negatively. Another possibility could be the high number of extreme events directly impacting yields.

In the component plane for wd, one area in the middle with a very high number of wet days stands out clearly (Fig. 11(c)). Interestingly, more than 60% of the input vectors in this area are from the western part of the catchment (Angola). This region within the component planes also experiences more extreme events and a higher rainfall total. Again, the high number of extreme events could be the reason for the lower yields. Also, specifics in soil type or local climatology might be an explanation.

Finally, a measure for the evaluation of especially suitable (or unsuitable) regions for rain-fed agriculture within the studied catchment was introduced. This was achieved by simply calculating a "class status" (CS) as the weighted mean of the class to which one region is belongs in accordance with the classification presented previously. For example, if the input vectors of a region are classified in Class 1 in three years and Class 2 for nine years, the CS will be $(3 \times 1 + 9 \times 2)/12 = 1.75$. The higher the class status of a region, the higher is its suitability for rain-fed agriculture. This allows one to directly derive the dominant combination of rainfall characteristics present in one region and also enables quantitative analysis using the component planes. The final results are summarized in Fig. 12. Additionally to CS, Fig. 12 also

includes information about the number of input vectors from a region in each of the classes.

The most suitable regions for rain-fed agriculture are, in general, the north-Zambian and Angolan regions, such as Mwinilunga (Z01; CS = 5.42), Solwezi (Z02; CS = 5.25) and Leua (A02; CS = 5.00); however, worse conditions are present in Sesheke (Z11; CS = 1.83), Senanga (Z10; CS = 2.00) and Caprivi (N01; CS = 2.00). It is in these regions where vulnerability and, thus, the need for adaptation is the highest in order to deal with future climatic changes which might worsen the situation.

5 Conclusions and Discussion

This study investigated the rainfall characteristics over the Upper Zambezi River Basin (UZRB) and their implications for rain-fed agriculture using the example of *Zea mays*. Measured yields were reproduced successfully using the SVAT model DAISY and mapped for the period 1998–2010. Finally, the relationship between rainfall characteristics and maize yields was analysed utilizing both a univariate and a multivariate approach (SOM).

It is found that yields and all investigated rainfall characteristics exhibit significant spatio-temporal variability over the UZRB. A clear gradient from north to south is identified in most rainfall characteristics: the northern parts of the basin receive more total rainfall over the wet period, and experience longer rainy seasons, more productive rainfall events, longer wet spells and shorter and fewer dry spells. Mean yields from rain-fed maize farming range from 1.2 t/ha in the northern part to only 0.5 t/ha in the southern part of the basin. The southernmost areas, in particular the Namibian, Botswanan and southern Zambian regions, display the least favourable preconditions for rain-fed farming.

In terms of connection between yields and rainfall characteristics, it is found that using total rainfall over the rainy season alone as an indicator for good yields is not sufficient. The highest correlation found is −0.56 between yields and the maximum dry spell duration ($dspd_{max}$), followed by the number of days with more than 10 mm of rain (p_{10}; 0.51), total rainfall and maximum/mean wet spell duration (p_{tot}, $wspd_{max}$, $wspd_{mean}$; 0.48) and the number of dry spells (dsp; −0.45). The number of wet days and spells (wd and wsp) as well as the frequency of extreme rain events greater higher than 20 mm do not affect yields directly. The multivariate analysis using a SOM revealed the importance of the indicator "duration of the rainy season" (dur): Yields higher than 1 t/ha can only be achieved with rainy season duration longer than ~160 days. This is an important finding, having in mind that only a few of the investigated regions reach that threshold regularly. For durations shorter than ~160 days, the interplay of p_{tot}, dsp, $dspd_{max}$ and $wspd_{max}$ (the indicators with the highest correlation coefficient) determines the success of rain-fed agriculture in a particular year. An inconsistent rainy season (such as many shifts of wet and dry spells) was found to affect yields negatively. In terms of adaptation, a key could be to cultivate faster growing varieties or to introduce small-scale water storage systems (i.e. rainwater harvesting) to extend water availability or bridge dry periods. Furthermore, it was identified that using dd and wd as well as wsp as indicators for agricultural success is not appropriate. As stated before, the wd criteria (> 2mm of rain) showed no correlation with yields, whilst one of the highest correlations was found for p_{10}. This clearly shows that small rains do not provide sufficient moisture for maize to enhance yields significantly. The threshold when interested in farming yields should thus be further investigated (i.e. by analysing from which threshold a high correlation is found). In terms of succession potential for rain-fed farming of maize, the regions Mwinilunga, Alto Zambeze, Solwezi and Cameia have the highest potential, whilst Caprivi, Botswana, Mongu, Kavango, Rivungo, Sesheke and Senanga are the least suitable areas.

This investigation using the best available dataset for the region in terms of spatial and temporal representativeness enables decision-makers, water resource managers, researchers and farmers to carry out region-specific analyses of rainfall characteristics and to develop adaptation measures, i.e. those identified and summarized by Milgroom and Giller (2013). Further interpretation by local experts who are more familiar with the particular area of interest than any outside researcher is recommended. This study and the examination of patterns solely included rainfall. It should be stated at this point that maize is not grown primarily all over the UZRB (personal comment in Mendelsohn *et al.* 2013) and was rather chosen to have a unified "scale" to evaluate and compare climatic suitability. However, over large parts of the basin maize remains the number one staple crop having also cultural importance for the people.

Even though the parameterization and calibration of the SVAT model could be improved with increased data and information on measured yields, the

usefulness of intensified research on rainfall characteristics is pointed out by this study. Rain-fed agriculture is, and will remain, challenging even in the areas identified as most suitable. However, with knowledge of length of typical dry periods, for example, a first step towards adaptation can be made. The results also reveal that rainfall variability is the overriding issue for local farmers. The effect of intra-seasonal and seasonal variability is discussed in the present research in-depth. For an investigation of 10-day period variability, however, the time period of 12 years is not sufficient. There is evidence that 10-day cycles of climate are present and should be further investigated (Tyson *et al.* 1975, Conway *et al.* 2009, WorldBank 2010). In addition, this problem further complicates the identification of trends and impacts of climatic changes. Consequently, using the criteria and indicators suggested by this study for the future, trends might be detectable. In a recent investigation, Ines and Hansen (2006) observed that wet day frequency and intensity distributions were used to improve GCM predictions, which generally predict too many rain events of too low intensity. Integrating such innovative approaches into the present study might allow the design of more precise forecasting systems.

The great variability also hinders many efforts to predict probable rainy season onset dates. Still, there is a pressing demand for such (Reason *et al.* 2006). Using neural network approaches such as SOM, which are capable of identifying relationships in highly complex and large datasets with a high number of parameters, allows advancement in this regard (as shown by Rivera *et al.* 2012 in their study in central Chile). Connecting the spatial patterns of rainy season characteristics with synoptic patterns might lead towards a forecasting system. Extended examination of the connection between the studied rainfall characteristics (i.e. how does an early/late onset of rains affect the remaining indicators?), an analysis over different stages of maize growth as well as scenario studies with other varieties or even crops are only a few possibilities for future applications. Especially investigating the effect of rainfall characteristics on phenological aspects, such as the sensitivity of maize in its different development stages, is worth examining further. Finally, the results of this investigation are not only limited to rain-fed agriculture, but can be very useful for studying flood creation and groundwater recharge, as well as providing additional information for hydrological models.

Acknowledgements

We would like to express our gratitude to the Agricultural, Food and Resource Economics Department of Michigan State University for providing their detailed data on agricultural yields over Zambia from their panel survey and the Zambia Meteorological Department. Matthias Beyer wants to thank J. Mendelsohn for sharing valuable insights and local knowledge over the past three years and D. Rivera for fruitful discussions on data mining and SOMs. Finally, we thank the reviewers for helping to improve this paper. We appreciate their effort to provide many useful comments and recommendations.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was partly funded by the German Federal Ministry for Education and Research (BMBF) within the CERPA project under contract number 01UN1007.

References

Abrahamsen, P. and Hansen, S., 2000. Daisy: an open soil-crop-atmosphere system model. *Environmental Modelling and Software*, 15 (3), 313–330. doi:10.1016/S1364-8152(00)00003-7

Batjes, N.H., 2005. ISRIC-WISE global data set of derived soil properties on a 0.5 by 0.5 degree grid (Version 3.0). Wageningen.

BGR, *2005. Groundwater* Investigations in the Eastern Caprivi Region: Main Hydrogeological Report.

Boko, M., *et al.*, 2007. Africa. *In*: M.L. Parry, *et al.*, eds. *Climate change 2007: impacts, adaptation and vulnerability.* Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press, 173–210.

Calzadilla, A., *et al.*, 2013. Economywide impacts of climate change on agriculture in Sub-Saharan Africa. *Ecological Economics*, 93, 150–165. doi:10.1016/j.ecolecon.2013.05.006

Camberlin, P. and Diop, M., 2003. Application of daily rainfall principal component analysis to the assessment of the rainy season characteristics in Senegal. *Climate Research*, 23 (2), 159–169. doi:10.3354/cr023159

Camberlin, P., *et al.*, 2009. Components of rainy seasons' variability in Equatorial East Africa: onset, cessation, rainfall frequency and intensity. *Theoretical and Applied Climatology*, 98 (3–4), 237–249. doi:10.1007/s00704-009-0113-1

Christensen, J.H., *et al.* 2007. Regional climate projections. *In*: *Climate change 2007: the physical science basis.* Contribution of Working Group I to the Fourth

Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press.

Cohen Liechti, T., *et al.*, 2012. Comparison and evaluation of satellite derived precipitation products for hydrological modeling of the Zambezi River Basin. *Hydrology and Earth System Sciences*, 16 (2), 489–500. doi:10.5194/hess-16-489-2012

Conway, D., *et al.*, 2009. Rainfall and water resources variability in Sub-Saharan Africa during the twentieth century. *Journal of Hydrometeorology*, 10 (1), 41–59. doi:10.1175/2008JHM1004.1

Cooper, P.J.M., *et al.*, 2008. Coping better with current climatic variability in the rain-fed farming systems of sub-Saharan Africa: an essential first step in adapting to future climate change? *Agriculture, Ecosystems & Environment*, 126 (1–2), 24–35. doi:10.1016/j.agee.2008.01.007

Crespo, O., Hachigonta, S., and Tadross, M., 2011. Sensitivity of southern African maize yields to the definition of sowing dekad in a changing climate. *Climatic Change*, 106 (2), 267–283. doi:10.1007/s10584-010-9924-4

Dee, D.P., *et al.*, 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137 (656), 553–597. doi:10.1002/qj.828

Department of Soil, Water and Environmetal Science, 2002. *ROSETTA: a computer program for estimating soil hydraulic parameters with hierarchical pedotrasfer functions*. Riverside, CA: US Salinity Laboratory.

Dinku, T., *et al.*, 2007. Validation of satellite rainfall products over East Africa's complex topography. *International Journal of Remote Sensing*, 28, 1503–1526. doi:10.1080/01431160600954688

Doherty, J.E., 2005. *PEST, Model-independent parameter estimation, user manual*. 5th ed. Brisbane: Watermark Numerical Computing.

DWA, 2009. *Agricultural statistics bulletin (2000-2007)*. Windhoek: Department of Water Affairs. Ministry of Agriculture,Water and Forestry.

Euroconsult MottMacDonald., 2007. *Rapid assessment report Zambezi River Basin*. Lusaka: Zambezi River Authority & Norwegian Embassy Lusaka.

Goulden, M., Conway, D., and Persechino, A., 2009. Adaptation to climate change in international river basins in Africa: a review. *Hydrological Sciences Journal*, 54 (5), 805–828. doi:10.1623/hysj.54.5.805

Hachigonta, S. and Reason, C.J.C., 2006. Interannual variability in dry and wet spell characteristics over Zambia. *Climate Research*, [online] 32 (1), 49–62. Available from: http://www.int-res.com/abstracts/cr/v32/n1/p49-62/ doi:10.3354/cr032049

Hachigonta, S., Reason, C.J.C., and Tadross, M., 2008. An analysis of onset date and rainy season duration over Zambia. *Theoretical and Applied Climatology*, 91, 229–243. doi:10.1007/s00704-007-0306-4

Hansen, S., *et al.*, 2012. Daisy: model use, calibration, and validation. *Transactions of the ASABE*, 55 (4), 1317–1335. doi:10.13031/2013.42244

Hong, Y., *et al.*, 2004. Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system. *Journal of Applied Meteorology*, 43, 1834–1853. doi:10.1175/JAM2173.1

Huffman, G.J., *et al.*, 2007. The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*, 8 (1), 38–55. doi:10.1175/JHM560.1

Huffman, G.J., *et al.*, 2010. The TRMM multi-satellite precipitation analysis. *In*: F. Hossain and M. Gebremichael, eds. *Satellite rainfall applications for surface hydrology*. Netherlands: Springer, 3–22.

Ines, A.V.M. and Hansen, J.W., 2006. Bias correction of daily GCM rainfall for crop simulation studies. *Agricultural and Forest Meteorology*, 138 (1–4), 44–53. doi:10.1016/j.agrformet.2006.03.009

Jayne, T.S., 2013. *Post harvest surveys for small and medium-sized holdings* [online]. Zambia food security research project. Michigan State University. Available from: http://fsg.afre.msu.edu/zambia/survey.htm [Accessed 23 October 2013].

Kohonen, T., 1982. Analysis of a simple self-organizing process. *Biological Cybernetics*, 44 (2), 135–140. doi:10.1007/BF00317973

Kohonen, T., 1990. The self-organizing map. *Proceedings of the IEEE*, 78 (9), 1464–1480. doi:10.1109/5.58325

Kohonen, T., *et al.*, 1995. *Self-Organizing Maps*. Springer Series in Information Science. doi:10.1007/978-3-642-97610-0

Kotir, J., 2011. Climate change and variability in Sub-Saharan Africa: a review of current and future trends and impacts on agriculture and food security. *Environment, Development and Sustainability*, 13 (3), 587–605. doi:10.1007/s10668-010-9278-0

Krajewski, W.F. and Smith, J.A., 1991. On the estimation of climatological Z-R relationships. *Journal of Applied Meteorology*, 30 (10), 1436–1445. doi:10.1175/1520-0450(1991)030<1436:OTEOCR>2.0.CO;2

López García, H. and Machón González, I., 2004. Self-organizing map and clustering for wastewater treatment monitoring. *Engineering Applications of Artificial Intelligence*, 17 (3), 215–225. doi:10.1016/j.engappai.2004.03.004

Mendolsohn, J. and Roberts, C., 1997. *An environmental profile and atlas of Caprivi*. Windhoek: Ministry of Environment and Tourism.

Mendelsohn, J.M., Jarvis, A., and Robertson, T., 2013. *A profile and Atlas of the Cuvelai-Etosha basin*. Windhoek: RAISON & Gondwana Collection, 170.

Milgroom, J. and Giller, K.E., 2013. Courting the rain: rethinking seasonality and adaptation to recurrent drought in semi-arid southern Africa. *Agricultural Systems*, 118, 91–104. doi:10.1016/j.agsy.2013.03.002

Mupangwa, W., Walker, S., and Twomlow, S., 2011. Start, end and dry spells of the growing season in semi-arid southern Zimbabwe. *Journal of Arid Environments*, 75 (11), 1097–1104. doi:10.1016/j.jaridenv.2011.05.011

New, M., *et al.*, 2006. Evidence of trends in daily climate extremes over southern and west Africa. *Journal of Geophysical Research*, 111 (D14), 1–11. doi:10.1029/2005JD006289

NOAA Climate Prediction Center, cited 2002: African rainfall estimation algorithm [online]. Version 2.0. Available from: http://www.cpc.ncep.noaa.gov/products/fews/RFE2.0_tech.pdf

Piani, C., *et al.*, 2010. Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. *Journal of Hydrology*, 395 (3–4), 199–215. doi:10.1016/j.jhydrol.2010.10.024

R Core Team, 2013. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Raes, D., *et al.*, 2004. Evaluation of first planting dates recommended by criteria currently used in Zimbabwe. *Agricultural and Forest Meteorology*, 125 (3–4), 177–185. doi:10.1016/j.agrformet.2004.05.001

Reason, C.J.C., Hachigonta, S., and Phaladi, R.F., 2005. Interannual variability in rainy season characteristics over the Limpopo region of southern Africa. *International Journal of Climatology*, 25 (14), 1835–1853. doi:10.1002/joc.1228

Reason, C.J.C., Landman, W., and Tennant, W., 2006. Seasonal to decadal prediction of Southern African climate and its links with variability of the Atlantic Ocean. *Bulletin of the American Meteorological Society*, 87 (7), 941–955. doi:10.1175/BAMS-87-7-941

Reason, C.J.C., *et al.*, 2004. *Seasonal to decadal predictability and prediction of southern African climate*. OceanDocs. Available from: http://hdl, 41 ST – Seasonal to decadal predictability and prediction

Rivera, D., *et al.*, 2012. Forecasting monthly precipitation in Central Chile: a self-organizing map approach using filtered sea surface temperature. *Theoretical and Applied Climatology*, 107 (1–2), 1–13. doi:10.1007/s00704-011-0453-5

Seidel, S., 2012. *Optimales simulationsbasiertes Design von Defizitbewässerungsexperimenten*. Dissertation. Dresden: University of Technology.

Shahin, D.M., 2002. *Hydrology and water resources of Africa*, 41. Springer/Book Series: Water science and technology library. Dordrecht: Springer.

Tadross, M., *et al.*, 2007. Changes in growing-season rainfall characteristics and downscaled scenarios of change over southern Africa: implications for growing maize. *In*: *Regional expert meeting*, 20–22 June, Nadi, 193–204.

Tadross, M.A., Hewitson, B.C., and Usman, M.T., 2005. The interannual variability of the onset of the maize growing season over South Africa and Zimbabwe. *Journal of Climate*, 18 (16), 3356–3372. doi:10.1175/JCLI3423.1

Thiemig, V., *et al.*, 2012. Validation of satellite-based precipitation products over sparsely gauged African River Basins. *Journal of Hydrometeorology*, 13 (6), 1760–1783. doi:10.1175/JHM-D-12-032.1

Thiemig, V., *et al.*, 2013. Hydrological evaluation of satellite-based rainfall estimates over the Volta and Baro-Akobo Basin. *Journal of Hydrology*, 499, 324–338. doi:10.1016/j.jhydrol.2013.07.012

Tyson, P.D., Dyer, T.G.J., and Mametse, M.N., 1975. Secular changes in South African rainfall: 1880 to 1972. *Quarterly Journal of the Royal Meteorological Society*, 101 (430), 817–833. doi:10.1002/qj.49710143008

Usman, M. and Reason, C., 2004. Dry spell frequencies and their variability over southern Africa. *Climate Research*, 26, 199–211. doi:10.3354/cr026199

Vesanto, J., *et al.*, 2000. *SOM Toolbox for Matlab 5*. Espoo: Helsinki University of Technology.

Wallner, M., Haberlandt, U., and Dietrich, J., 2013. A one-step similarity approach for the regionalization of hydrological model parameters based on self-organizing maps. *Journal of Hydrology*, 494, 59–71. doi:10.1016/j.jhydrol.2013.04.022

Woltering, L., 2005. *Estimating the influence of on-farm conservation practices on the water balance: case of the Mzinyathini catchment in Zimbabwe*. Delft: Delft University of Technology.

WorldBank, 2010. *The Zambezi River Basin a multi-sector investment opportunities analysis*. Washington, DC: World Bank Publishing.