



**HAL**  
open science

# Mapping Weather-Type Influence on Senegal Precipitation Based on a Spatial–Temporal Statistical Model

Henning W. Rust, Mathieu Vrac, Benjamin Sultan, Matthieu Lengaigne

► **To cite this version:**

Henning W. Rust, Mathieu Vrac, Benjamin Sultan, Matthieu Lengaigne. Mapping Weather-Type Influence on Senegal Precipitation Based on a Spatial–Temporal Statistical Model. *Journal of Climate*, 2013, 26 (20), pp.8189-8209. 10.1175/jcli-d-12-00302.1 . hal-01495317

**HAL Id: hal-01495317**

**<https://hal.science/hal-01495317>**

Submitted on 11 Jun 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Mapping Weather-Type Influence on Senegal Precipitation Based on a Spatial–Temporal Statistical Model\*

HENNING W. RUST

*Institut für Meteorologie, Freie Universität Berlin, Berlin, Germany*

MATHIEU VRAC

*Laboratoire des Sciences du Climat et de l'Environnement, Gif-sur-Yvette, France*

BENJAMIN SULTAN AND MATTHIEU LENGAINNE

*Laboratoire d'Océanographie et de Climatologie par Experimentation et Approche Numérique, Université Pierre et Marie Curie, Paris, France*

(Manuscript received 31 May 2012, in final form 14 March 2013)

## ABSTRACT

Senegal is particularly vulnerable to precipitation variability. To investigate the influence of large-scale circulation on local-scale precipitation, a full spatial–statistical description of precipitation occurrence and amount for Senegal is developed. These regression-type models have been built on the basis of daily records at 137 locations and were developed in two stages: (i) a baseline model describing the expected daily occurrence probability and precipitation amount as spatial fields from monsoon onset to offset, and (ii) the inclusion of weather types defined from the NCEP–NCAR reanalysis 850-hPa winds and 925-hPa relative humidity establishing the link to the synoptic-scale atmospheric circulation. During peak phase, the resulting types appear in two main cycles that can be linked to passing African easterly waves. The models allow the investigation of the spatial response of precipitation occurrence and amount to a discrete set of preferred states of the atmospheric circulation. As such, they can be used for drought risk mapping and the downscaling of climate change projections.

Necessary choices, such as filtering and scaling of the atmospheric data (as well as the number of weather types to be used), have been made on the basis of the precipitation models' performance instead of relying on external criteria. It could be demonstrated that the inclusion of the synoptic-scale weather types lead to skill on the local and daily scale. On the interannual scale, the models for precipitation occurrence and amount capture 26% and 38% of the interannual spatially averaged variability, corresponding to Pearson correlation coefficients of  $r_O = 0.52$  and  $r_I = 0.65$ , respectively.

## 1. Introduction

Climate variability and extreme events such as droughts, excessive rains, and floods affect agricultural productivity and hence rural household food security

(Haile 2005). Since the 1970s, the largest food crises in Africa that required large-scale external food aid (1974, 1984/85, 1992, and 2002) have been attributed fully or partially to extreme events (Dilley et al. 2005). Furthermore, Africa is considered particularly vulnerable to climate change due to a combination of naturally high levels of climate variability, high reliance on climate sensitive activities such as rainfed agriculture, and limited economic and institutional capacity to cope with and adapt to climate variability and change (Roudier et al. 2011).

Senegal, as part of the semiarid Sahel, witnessed a period of drought during the second half of the twentieth century, with precipitation anomalies slowly recovering

---

\* Supplemental information related to this paper is available at the Journals Online website: <http://dx.doi.org/10.1175/JCLI-D-12-00302.s1>.

---

*Corresponding author address:* Henning W. Rust, Institut für Meteorologie, Freie Universität Berlin, Carl-Heinrich-Becker-Weg 6-10, D-12165 Berlin, Germany.  
E-mail: [henning.rust@met.fu-berlin.de](mailto:henning.rust@met.fu-berlin.de)

toward the end of the century (e.g., Nicholson 2001; Biasutti et al. 2008, and references therein). This interannual variability is generally attributed to changes in sea surface temperature (e.g., Folland et al. 1986; Paeth and Hense 2006; Biasutti et al. 2008, and references therein). The resulting changes in precipitation patterns with potentially less precipitation and an increasing risk for droughts pose a threat to water resources and agricultural yields in Senegal. In addition, excessive rains are responsible for severe flooding, with serious property damage and deaths in the region as illustrated by the recent floods of August 2012 in Dakar. Precipitation in this region is dominated by the West African monsoon (WAM) being active from May to October with intraseasonal variations (Sultan and Janicot 2003). The WAM is the dominating mechanism in this region and is in the focus of atmospheric research (Lafore et al. 2011).

Strategies to adapt to changing precipitation patterns require localized information about the expected changes. To inform stakeholders from hydrology or agriculture, researchers build watershed or crop models based on meteorological variables on a daily and subregional scale (e.g., Sultan et al. 2005). Providing valuable information to develop adaptation strategies therefore requires the following question to be addressed: how does local precipitation respond to large-scale atmospheric variability in Senegal?

As these values are typically not, or not reliably, projected by general circulation models (GCMs), downscaling approaches are required to translate GCM projections to the local scale. Downscaling for this region has been undertaken using two generic approaches: dynamical downscaling using regional climate models with a horizontal resolution of about 50 km (e.g., Paeth et al. 2005) as well as statistical approaches (e.g., Moron et al. 2008b); the latter typically yield gauge-based data and not spatial fields. Statistical approaches for downscaling are manifold; recent comprehensive reviews on precipitation downscaling have been given by Fowler et al. (2007) and Maraun et al. (2010).

Here, we built a set of generalized linear models (McCullagh and Nelder 1989; Chandler 2005) similar to a study for South Africa by Ambrosino et al. (2011). These models aim at a spatial description of rainfall in Senegal, not restricted to station locations. We use season, location, and temporal dependence as predictors for baseline descriptions of rainfall occurrence and amount. In a second step, we add a discrete set of weather types as predictors to these baseline models to establish the link between local precipitation and atmospheric conditions. This allows the assessment of local precipitation response to large-scale atmospheric variability, thus addressing the question mentioned above. It furthermore

allows for downscaling in a perfect prognosis sense (cf. Maraun et al. 2010) and thus can be used to obtain gauge-based scenarios of daily precipitation from GCM climate change projections.

The crucial step for any statistical downscaling approach is the selection of predictors. They are required to be informative for precipitation and should capture the variation on the relevant time scales (e.g., the climate change signal). To compress the information of high-dimensional gridpoint-based predictors from GCMs, suitable transformations are needed. These can be, for example, statistically motivated approaches for dimensionality reduction such as principal component analysis (PCA; e.g., Preisendorfer 1988; Hannachi et al. 2007), assigning a large fraction of the variability to a small number of principal components that can be readily used as predictors for downscaling (e.g., Benestad 2001). Weather types or circulation patterns are a meteorologically motivated predictor transformation. The daily synoptic situation is mapped on a small set of discrete categories by means of an objective or subjective criterion of similarity (e.g., Michelangeli et al. 1995; Philipp et al. 2007; Huth 2000). The use of weather types straightforwardly allows the capture of some nonlinear aspects of the link between predictor and precipitation. It is thus a popular approach in statistical downscaling (e.g., Conway and Jones 1998; Wilby et al. 1998; Boé et al. 2006; Vrac et al. 2007b; Moron et al. 2008b).

The choice of relevant variables from atmospheric dynamics to be included for downscaling precipitation is critical and certainly more important than the choice of a classification method to detect weather types (Beck and Philipp 2010). They must capture the main features of the weather over the studied region, which in the monsoon areas will be those which have significant impacts on the rainfall regime (Gueye et al. 2011). Over the Senegal region, Deme et al. (2003) have selected 64 atmospheric variables to predict daily rainfall amounts around Dakar in August and found that the most suitable variables are lifting condensation level, vorticity at 700 hPa, humidity at 925 hPa, the total water vapor flux in the monsoon layer, and the meridional flux of water vapor in the layer between 600 and 300 hPa. More recently, Gueye et al. (2011) used both mean sea level pressure (SLP) and the 850-hPa wind to define synoptic regimes that are relevant for understanding the daily variability of rainfall during the monsoon over Senegal. The 850-hPa wind field is an intermediate level between the low-level monsoon winds that brings moisture inland and the level of the African easterly waves (AEW) that modulate convection within the intertropical convergence zone (Diedhiou et al. 1999). According to these previous studies, we chose the 850-hPa wind field and

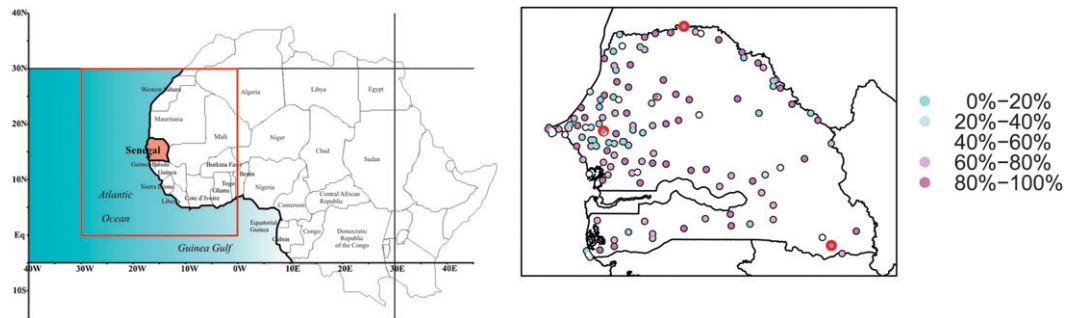


FIG. 1. (left) Senegal and the box (red) where atmospheric data is used in this study. (right) Location of rain gauge stations in Senegal. The color depicts the data availability in the percent of days in the wet season (May–Oct) from 1961 to 2000 (7360 points). The red circles mark the location of example stations used in the later sections.

the relative humidity at 925 hPa to define weather types in Senegal. Additionally, SLP was used in the beginning of this study but was left out as it did not notably improve the results.

The subsequent section 2 describes the precipitation and atmospheric data and section 3 depicts the approach to define weather types and summarizes the basics of generalized linear modeling for rainfall. The baseline precipitation models are developed in section 4 and the influence of the weather types is presented in section 5. The article closes with a summary (section 7a), as well as discussion and outlook in section 7b.

## 2. Data

### a. Precipitation data

Precipitation in Senegal occurs almost exclusively from May to October (wet season). It is dominated by the WAM, leading to a peak in precipitation in August. The northern part of the country belongs to the semiarid region south of the Sahara, the Sahel, and is typically dryer than the south (Nicholson 2001).

Daily total precipitation is available for a set of 137 stations with a nominal precision of 0.1 mm. The data come from the measurement network of the National Meteorological Service from Senegal, compiled by the Agrometeorology–Hydrology–Meteorology (AGRHYMET) Regional Centre. Stations are predominantly located in the western part of the country and along the border with Mauritania where most of the cities are, see Fig. 1 (right). The quality of the data greatly varies in terms of missing values and the time period covered. No distinct regions dominated by low-quality stations can be identified. A good overlap of precipitation observations at all stations with the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis

(see the following section 2b) is given for the time period from 1 January 1961 to 31 December 2000. Summing the number of observations for all stations in the wet season of the given time period yields more than 600 000 data points.

In the following, we separately consider the occurrence and amount of precipitation. For a day to be considered as a rainy or wet day, and thus as a day with rainfall occurrence, we set a 1-mm threshold of precipitation per day. A value of less than 1 mm is generally considered as not being useful for agricultural applications. For instance, Sivakumar (1988) considered dry periods that are critical for the crops by counting the number of consecutive days with rainfall values weaker than 1 mm. The rainfall occurrence time series is therefore a daily series of 1 and 0 depending on rainfall being greater than 1 mm or not, respectively. A rainy day conditional on rainfall amount is thus a series when the daily rainfall amount for the day has been classified as wet ( $>1$  mm). The number of observations of rainfall amount is thus smaller than for the occurrence time series, since only wet days are considered. The sum of rainy days for all stations yields more than 116 000 observations ( $\approx 20\%$ ) in the observation period.

### b. Atmospheric data

For the description of the atmospheric circulation, we use data from the NCEP–NCAR reanalysis project on a  $2.5^\circ \times 2.5^\circ$  grid (Kalnay et al. 1996). The domain has been chosen between  $0^\circ$ – $30^\circ$ N and  $30^\circ$ W– $0^\circ$  (red box in Fig. 1, left) in order to be centered over Senegal and to cover a space consistent with the AEW wavelength (about 3000 km) in coherence with previous studies (Moron et al. 2008a; Gueye et al. 2011). It contains  $13 \times 13 = 169$  grid points. We use longitudinal ( $U$ ) and meridional ( $V$ ) wind at 850 hPa and relative humidity (RHUM) at 925 hPa. The latter is the ratio of partial pressure obtained from the specific humidity and the

TABLE 1. Number of principal components capturing 90% of the total variability.

	Gridpoint scaling	Variable scaling
Raw	40	30
Anomalies	48	40

saturation vapor pressure of water ( $p_{\text{sat}}$ ) calculated using 925-hPa temperature ( $T$ ) and the Goff–Gratch equation (Goff and Gratch 1946; the appendix section a).

Variables are weighted with  $\sqrt{\cos(2\pi\text{latitude}/360)}$  at their grid box to account for slightly different areas. The resulting set of variables will be referred to in the following as the “raw” data. Additionally, we use a set of “anomalies” defined as the difference with the climatological mean seasonal cycle obtained from 1961 to 2000.

### 3. Methods

#### a. Defining weather types

Cluster analysis yields a discrete set of weather types defined on  $U$  and  $V$  winds and relative humidity for the given time period. We have thus three variables at 169 grid points yielding 507 dimensions. To reduce dimensionality and avoid the overrepresentation of highly correlated values at neighboring grid points, we use PCA (Huth 1996; Jolliffe 2002; Hannachi et al. 2007) and continue with the first principal components accounting for 90% of the total variability.

It is common to define the weather types on variable anomalies (e.g., Rust et al. 2010; Gueye et al. 2011); the alternative, working with so called raw values is particularly interesting for resolving the seasonal cycle (e.g., Moron et al. 2008b; Vrac et al. 2013). Another popular choice is to normalize the input dimension prior to the PCA such that the three variables at each grid point have unit variance. We refer to this procedure as “gridpoint scaling.” As an alternative, we suggest normalizing only with respect to the variable type (i.e., horizontal or meridional wind or relative humidity) such that one variable over all grid points has unit variance. This approach is referred to as “variable scaling.” It allows the per grid-point variance of, for example, relative humidity to be different at different grid points. Grid points with a larger variance are thus weighted stronger than those with smaller variance.

As it is not evident a priori which scaling approach to favor for downscaling or whether to use raw values or anomalies, we compare all four combinations and measure their performance using a goodness-of-fit criterion. The dimensions of these sets (i.e., the number of principal components retained) are given in Table 1.

For clustering, we use the Gaussian mixture model (GMM) approach (Fraley and Raftery 2002). This directly implements the idea that the probability density of atmospheric states is a multimodal probability distribution function (PDF), which can be approximated with superpositions of Gaussian PDFs (Branstator and Selten 2009). GMMs have been used for atmospheric circulation clustering in previous works (Haines and Hannachi 1995; Hannachi 1997, 2007; Smyth et al. 1999; Rust et al. 2010). They showed more consistent circulation patterns across various levels and are more sensitive to day-to-day variations in pattern frequencies over the eastern United States in comparison to hierarchical ascending clustering (HAC; Vrac et al. 2007a); they also provided useful weather types for precipitation downscaling (Vrac et al. 2007b).

GMMs are a likelihood-based approach and a reasonable strategy to define the cluster number is the Bayesian Information Criterion (BIC; Fraley and Raftery 2007). In the present setting, however, we suggest to choose the number of clusters so that it best suits the application in mind: a statistical description of rainfall. We thus systematically evaluate the precipitation model for several cluster numbers and use model performance on daily scales and at station level to guide the choice of the appropriate cluster number.

#### b. Generalized linear modeling of rainfall

We use generalized linear models (GLMs; McCullagh and Nelder 1989) to describe the seasonal and spatial variations of precipitation, as well as to analyze the influence of weather types. GLMs allow a flexible and easily interpretable description, while being computationally inexpensive. They are widely used in statistics and have been proven to be useful in climate related questions (Chandler 2005), in particular in modeling daily rainfall (Coe and Stern 1982; Stern and Coe 1984) or exploring climate factors driving monthly rainfall (Ambrosino et al. 2011). The underlying idea is to consider the response  $\mathbf{y}$  (here daily precipitation) as a random variable from an adequately chosen probability distribution. It is linked to a linear predictor  $\boldsymbol{\eta}$  via a link function  $g$ :  $g(\mathbf{y}) = \boldsymbol{\eta}$ . The linear predictor is a sum of  $p$  covariates  $\mathbf{x}_i$  and related coefficients  $\beta_i$ :  $\boldsymbol{\eta} = \sum_{i=1}^p \mathbf{x}_i \beta_i$ . In the case the response can be assumed to be normally distributed and the link function is the identity [ $g(\mathbf{y}) = \mathbf{y}$ ], the GLM is equivalent to multiple linear regression. However, in the following, we assume a gamma-distributed response and a logarithmic link function for rainfall amounts and a binomial response with a logit link function for occurrence probabilities. With iteratively reweighted least squares, maximum likelihood estimates for the coefficients  $\beta_i$  are obtained (McCullagh



and Nelder 1989).<sup>1</sup> As observations enter the likelihood for parameter estimation with their associated temporal and spatial information, irregularly spaced observations do not pose a problem; in particular, missing values in the data records can be ignored for parameter estimation. Covariates  $\mathbf{x}_i$  are defined for every observation and are either continuous variables, such as the position in the annual cycle or location in space, or incident vectors for qualitative factors, such as the prevailing weather type or previous day rainfall occurrence. Besides these direct effects, interactions between covariates can be included (e.g., two-way interactions  $\beta_{12}\mathbf{x}_1\mathbf{x}_2$ ). They modulate the influence of covariate  $\mathbf{x}_1$  by the value of covariate  $\mathbf{x}_2$  or vice versa.

The probability of rainfall occurrences  $\pi_i$  at day  $i$  can be described using logistic regression; that is, a binomial distribution and a logit link function  $g_\pi(\pi_i) = \ln[\pi_i/(1 - \pi_i)]$  (Coe and Stern 1982; Stern and Coe 1984; Ambrosino et al. 2011). On the right hand side of the regression model, we include five types of predictors including a seasonal component describing the course of the monsoon, a spatial component, a dependence on the previous day precipitation, interactions between those, and finally a weather type–dependent component:

$$\ln\left(\frac{\pi_i}{1 - \pi_i}\right) = \eta_{i,\text{season}} + \eta_{i,\text{space}} + \eta_{i,\text{previous}} + \eta_{i,\text{Interactions}} + \eta_{i,\text{WT}} = \eta_{i,\text{baseline}} + \eta_{i,\text{WT}} \tag{1}$$

The first four terms are pooled in a baseline predictor  $\eta_{\text{baseline}}$  developed in section 4. It contains no information on the large-scale atmospheric situation. The weather type–dependent predictor  $\eta_{\text{WT}}$  is added in section 5.

The amounts  $r_i$  on a wet day  $i$  are described with a gamma distribution with a scale parameter  $\theta$  and a constant shape parameter  $k$ . The distribution has mean  $k\theta$  and variance  $k\theta^2$ , and thus a constant coefficient of variation  $k\theta/\sqrt{k\theta^2}$ , leading to a variance increasing with the mean. This hypothesis, as well as the associated logarithmic link function  $g_r(r_i) = \ln(r_i)$ , is a common choice for rainfall amounts (Coe and Stern 1982; McCullagh and Nelder 1989; Ambrosino et al. 2011). Analogously to the occurrence model, the amount model reads

$$\ln(r_i) = \eta_{i,\text{season}} + \eta_{i,\text{space}} + \eta_{i,\text{previous}} + \eta_{i,\text{Interactions}} + \eta_{i,\text{WT}} = \eta_{i,\text{baseline}} + \eta_{i,\text{WT}} \tag{2}$$

Predictors involve the same categories (seasonal, spatial, previous day precipitation, interactions, and weather types) but they do not contain the same terms as in Eq. (1). The baseline part is developed in section 4, and the weather-type influence is added in section 5.

As it is not evident a priori which covariates are important, different covariates are tested and model improvement is assessed using two criteria: (i) a modified Akaike Information Criterion (AIC) and (ii) the Brier score for occurrence probabilities and mean-square prediction error for amounts obtained from cross validation (see the appendix, section b).

#### 4. Modeling seasonal and regional variation

At first, a model is constructed to describe the seasonal and regional variations of daily occurrence probabilities and amounts for Senegal. This model does not yet include information from atmospheric variables and will be called the baseline model in the following. The predictor  $\eta_{\text{baseline}}$  used to model these quantities involves covariates  $\mathbf{x}_i$  constructed from seasonal  $\eta_{\text{season}}$  and spatial components  $\eta_{\text{spatial}}$ , as well as from previous day precipitation characteristics  $\eta_{\text{previous}}$ , as detailed below. The associated coefficients  $\beta_i$  are estimated on the basis of all available daily occurrence and amount records. The resulting baseline model gives a spatially continuous estimate of the expected rainfall occurrence probability and expected amount for any given day in the wet season as it assumes that these expected values are varying smoothly in space and time. As such, it provides an average monsoon cycle as a spatial field for the whole country (also at unobserved sites). In section 5, we address the much weaker interannual signal, which is modulating the strong seasonal cycle.

##### a. Construction of covariates

A series of harmonic functions is used to describe the seasonal variation in  $\eta_{\text{season}}$  (Rust et al. 2009; the appendix, section c), and Legendre polynomials of longitude, latitude, and altitude (Ambrosino et al. 2011; the appendix, section d) are used for spatial variation. Periods of persistent rainfall are accounted for by including the previous day rainfall explicitly as a covariate. This is straightforward for the occurrence model using the previous day rainfall occurrence (0 or 1); the previous day precipitation amount  $r_{-1}$  instead enters as  $\log(1 + r_{-1})$ , ensuring a zero value for a previous day being dry and a slow (logarithmic) increase of the predictor with the increasing precipitation amount (Ambrosino et al. 2011).

<sup>1</sup> Estimates are obtained using the environment for statistical computing R (R Development Core Team 2004) and the function glm.

TABLE 2. Abbreviations of predictors with their associated transformation.

Abbreviation	Predictor	Transformation
$\beta_0$	Constant offset	—
$\eta^0$	Basis predictor for stepwise search	—
sink	Day of the year [doy( <i>t</i> )]	$\sin[2\pi k \text{ doy}(t)/365.25]$
cosk	Day of the year [doy( <i>t</i> )]	$\cos[2\pi k \text{ doy}(t)/365.25]$
lonn	Rescaled longitude <i>x</i>	$P(x, n)$
latn	Rescaled latitude <i>y</i>	$P(y, n)$
altn	Rescaled altitude <i>z</i>	$P(z, n)$
pr1	Previous day occurrence $\pi_{-1}$	$\pi_{-1}$
lobs1	Previous day amount $r_{-1}$	$\log(1 + r_{-1})$
WTn	Weather type <i>n</i>	Categorical variable

We start the search for a parsimonious model with  $\eta^0_{\text{baseline}}$  including first- and second-order harmonic functions [the appendix Eq. (A5), 4 terms] to describe seasonality, first- and second-order Legendre polynomials of longitude and latitude [the appendix Eq. (A6), 4 terms], first-order Legendre polynomial of altitude [the appendix Eq. (A6), 1 term], and previous day precipitation occurrence (1 term). We stepwise add higher orders of these terms until third-order harmonic functions and fifth-order Legendre polynomials. Factors are added according to their potential to improve the model. After each step, factors are checked for being obsolete; if so, they are deleted. Model improvement is measured using the modified AIC, Eq. (A1) of the appendix.

Besides direct effects, interactions (products between covariates) are taken into account. For example, an interaction between seasonal and spatial terms allows seasonal variation to change with location. Analogously, with an interaction between seasonality/location and previous day rainfall, the models take seasonal/regional variations of dependence into account. The abbreviations of predictors used in the occurrence and amount models developed in the following sections are given in Table 2.

### b. Occurrence model results

Figure 2 shows the AIC (open symbols) and the Brier score from cross validation (closed symbols) for the stepwise search for factors in the occurrence model.

The stepwise search stops at 55 factors, when the AIC cannot be further reduced. However, the minimum Brier score from cross validation is attained for a model with 35 factors (more lead to overfitting). The minimum Brier score model includes a constant offset, 13 main effects and 21 interaction terms. Major model improvements are

obtained by adding interactions between seasonal and regional factors (e.g., sin1:lon1, cos2:lon1, and sin1:lat1),<sup>2</sup> implying that the variation of the seasonal cycle in space is a major source of variability. Similarly, the dependence on the previous day occurrence strongly varies with longitude (lon1:pr1). The occurrence of precipitation two or more days prior does not contribute significantly to the model.

A comparison of the seasonal cycle of model predictions and observations is depicted in Fig. 3. We show exemplarily the result for Podor (northernmost station), Bambey Meteo (central western station), and Kedougou (southeastern station). As the model includes the previous day rainfall, model predictions have been obtained using three hypotheses for the previous day precipitation: a dry previous day (red line), a wet previous day (blue line), and a previous day precipitation occurrence probability estimated by the spline smoothing the observations (green line). The shape of the seasonal cycle is reasonably well recovered, in particular the asymmetric onset and offset of the monsoon. Figure 3 visually corroborates that seasonality is different at different locations, as indicated already by the high importance of the corresponding interaction terms (sin1:lon1, cos2:lon1, sin1:lat1, etc.) in Fig. 2. Predictions assuming a previous wet (dry) day have a tendency to lie above (below) the spline-smoothed observations (black line). Predictions based on a smoothed average for the previous day precipitation occurrence (green line) closely follow the smoothed observations (black). For Podor and Bambey Meteo, occurrence probability is larger (smaller) if the previous day was wet (dry). This difference is particularly small for stations in the southeast during the peak phase of the monsoon where occurrence probability is generally high. Toward the east, previous day precipitation occurrence becomes less important. This regional dependence is in line with the longitude–previous-day-dependence interaction term (lon1:pr1), leading to a large improvement of the model (Fig. 2). A potential explanation for subsequent rainy days in the west is based on the mechanisms of convective system generation: they typically develop east of Senegal during the afternoon and move westward (Laing and Fritsch 1993). The systems arriving at night or early morning in the

<sup>2</sup>The interaction term sin1:lon1 denotes an interaction between the first-order sine wave describing seasonality and first-order Legendre polynomial describing zonal variation; other interactions are denoted correspondingly using the abbreviations in Table 2.

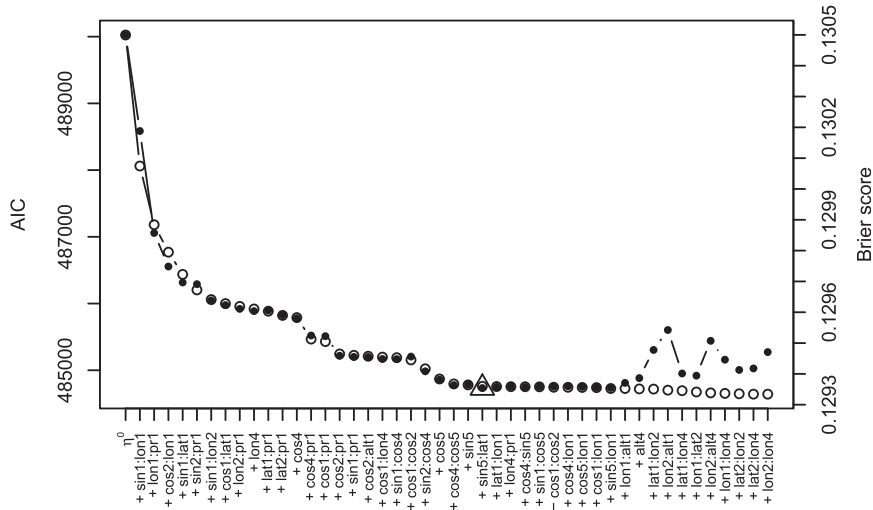


FIG. 2. Modified AIC ( $k = 6.63$ ; open symbols; left axis) and Brier score from cross validation (closed symbols; right axis; triangle depicts minimum) for the occurrence baseline model as effects are added (+) or subtracted (-). The abbreviations of predictors on the abscissa are listed in Table 2.

west of Senegal distribute rainfall over two consecutive days of measurement.

We want to assess to what extent the baseline model is able to reconstruct daily occurrence probabilities at unobserved sites given daily records at observed sites. Therefore, we perform a 10-fold cross-validation experiment. As before, stations are randomly taken out, the model parameters are estimated for the remaining stations, and the Brier score is obtained at the daily level for unobserved stations. If the quality of prediction at unobserved sites is comparable to a climatological

forecast, the baseline model has proven beneficial. With the stations' climatology (spline-smoothed occurrence frequency) as reference (obtained from the full record), an average Brier skill score (BSS) = 1.6% is obtained. A BSS > 0 implies that the baseline model has positive skill reconstructing daily precipitation occurrence at unobserved sites. As it uses the previous day precipitation occurrence at the observed sites, it has information on the daily weather conditions and thus can perform even better than the climatology at that site.

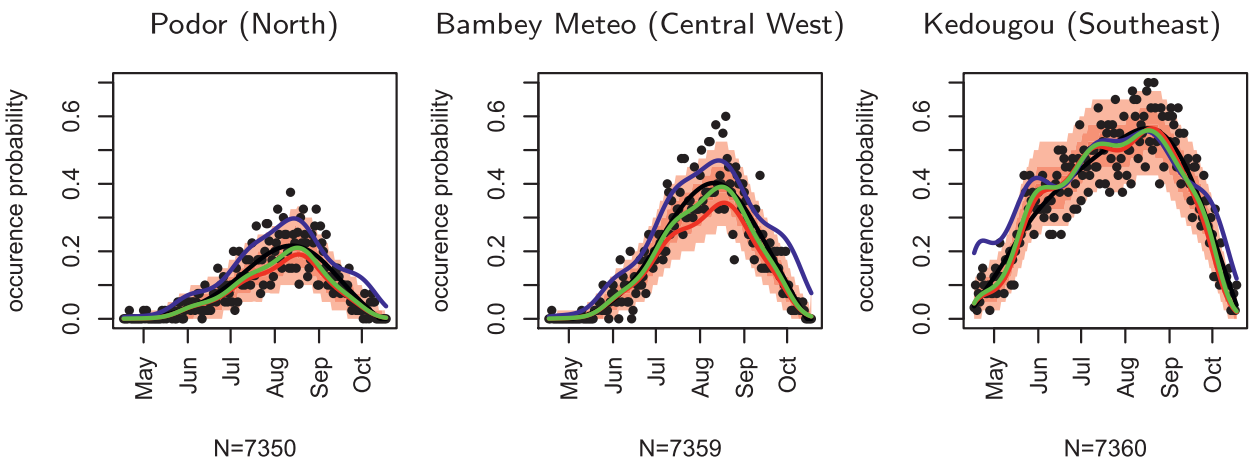


FIG. 3. Mean seasonal cycle (May–Oct) of rainfall occurrence probability as estimated from the station data (black dots), spline-smoothed observations (black lines), and predictions from the baseline model using three different hypotheses for the previous day occurrence probabilities:  $\pi_{-1} = 0$  (red line),  $\pi_{-1} = 1$  (blue line), and the occurrence probability from the spline-smoothed observations (green line). Red shades give the mass of the binomial distribution for the inner 90%, 60%, and 20%. The location of the three example stations is depicted as red circles in Fig. 1.



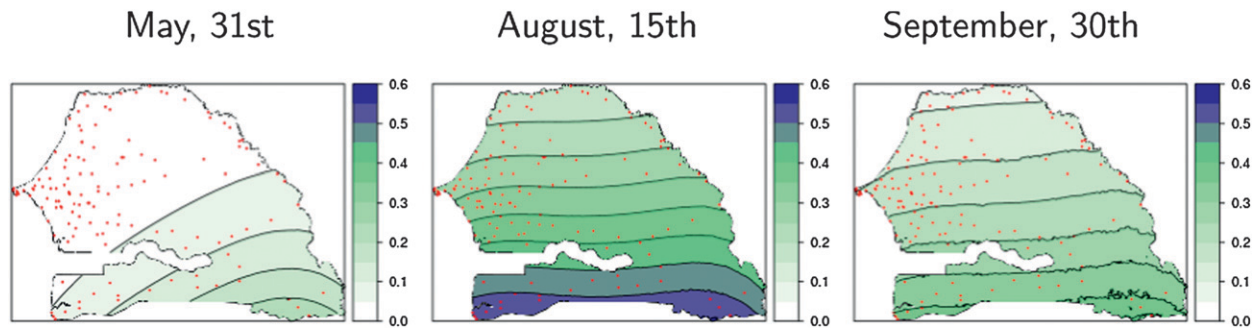


FIG. 4. Spatial predictions for the baseline rainfall occurrence model for (left) late May, (center) mid-August, and (right) late September; colors denote daily precipitation occurrence probability with black lines separating different color shadings. Red dots represent station locations.

Together with the DIVA-GIS altitude profile (<http://www.diva-gis.org/Data>), the baseline model can be used to reconstruct a full spatial field of daily precipitation occurrence probability from station information. Furthermore, given station climatologies, a full spatial climatology can be constructed. Precipitation is a process with high spatial variability on scales smaller than the typical distance between stations; it is furthermore influenced by other factors than longitude, latitude and altitude. The interpretation of the regional predictions in Fig. 4 on small scales, in particular on the scale of the DIVA-GIS altitude profile (about 1-km spatial resolution), is thus not meaningful. The figures do, however, clearly illustrate the zonal and meridional variations of precipitation occurrence on a subcountry scale. With the onset of the monsoon in May, precipitation occurrence probability increases first in the southeast and shows

a northwest–southeast gradient (Fig. 4, left). During the course of the monsoon, this gradient turns north–south (Fig. 4, middle). An animated sequence of plots showing the development of precipitation occurrence probability with the onset of the monsoon is available in the supplemental material.

### c. Amount model results

Similar to the occurrence model, we build the baseline model for rainfall amounts on wet days. With 116700 data points, the number of observations is much smaller. Consequently, model building starts with only the first-order harmonics and Legendre polynomial for longitude, latitude, and altitude.

Figure 5 shows the AIC and the mean-square prediction error from cross validation for the amount model. The stepwise search guided by the AIC stops for a model

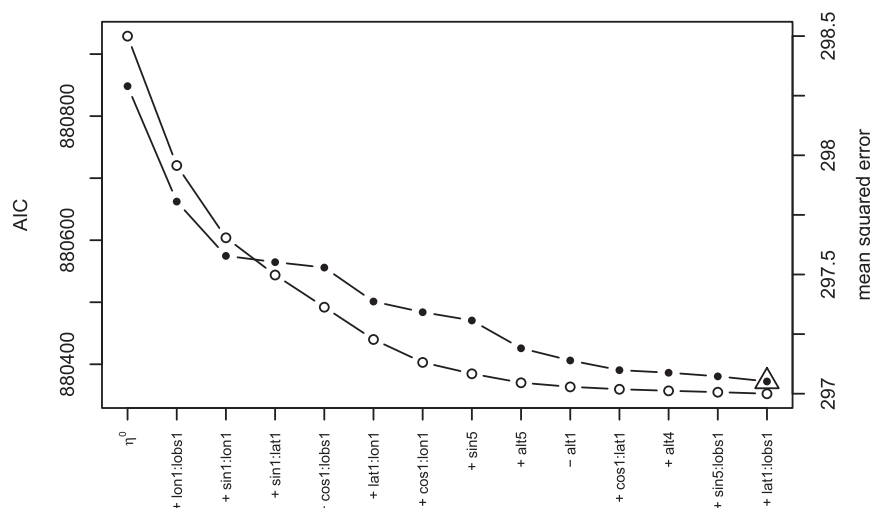


FIG. 5. Modified AIC ( $k = 6.63$ ; open symbols; left axis) and the mean-square prediction error obtained from cross validation (closed symbols; right axis; triangle depicts minimum) for the amount baseline model as effects are added (+) or subtracted (-). The abbreviations of predictors on the abscissa are listed in Table 2.

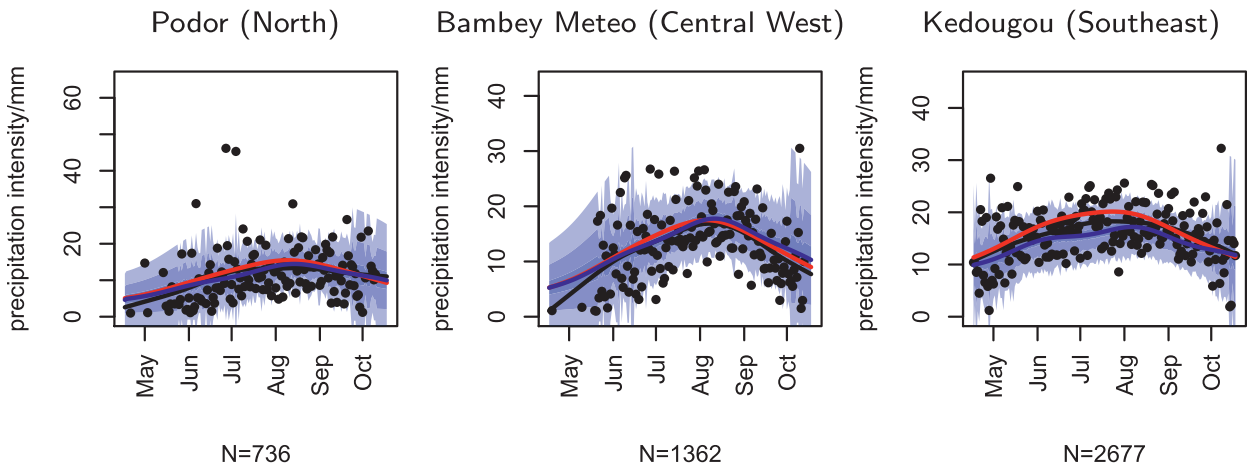


FIG. 6. Mean seasonal cycle (May–Oct) of rainfall amount ( $\text{mm day}^{-1}$ ) as estimated from the station data (black dots), spline-smoothed observations (black lines), and predictions from the baseline model setting the previous day precipitation amount to the value of the spline-smoothed observations (blue line) and to 0 (red line). Blue shades give the mass of a normal distribution for the inner 90%, 60%, and 20%.

including 18 components. Within this set, cross validation does not indicate a smaller model to fit better. Besides a constant offset, there are eight main effects and nine interactions. The largest model improvements result from an interaction between longitude and previous day precipitation ( $\text{lon1}:\text{lobs1}$ ), implying that the dependence on the previous day shows a strong zonal variation. The strong dependence of the seasonal cycle on location is reflected in the interaction terms of the harmonic cycle with location ( $\text{sin1}:\text{lon1}$  and  $\text{sin1}:\text{lat1}$ ). The constant shape parameter  $k$  of the gamma distribution is estimated a posteriori by maximum likelihood to  $\hat{k} = 1.129(0.004)$ .

Figure 6 compares the observed mean seasonal cycle for precipitation amounts to the model. The red line depicts predictions assuming a previous dry day; the blue line assumes average previous day precipitation (black line), obtained by the spline smoothing the observations (black dots). To compare daily averages to their model counterpart, Fig. 6 shows quantiles of a normal distribution around the predicted values as blue shades. Its variance equals the variance of the fitted gamma distribution at the given day scaled by  $1/n_{\text{day}}$ , the number of observations available for the day of the year. If no observations are present,  $n_{\text{day}}$  is set to one. This representation assumes convergence to the normal distribution when averaging  $n_{\text{day}}$  gamma-distributed observations; it becomes more accurate with increasing  $n_{\text{day}}$ . The easternmost station (right panel) shows a smaller amplitude than the other stations but remains on a higher level of amounts; thus, rainfall amounts conditional on a wet day do not change as much with the season as in the western parts of the country.

Again, we want to assess the baseline model's potential to reconstruct expected amounts for unobserved

sites. We perform a 10-fold cross validation similar to the occurrence model but with the mean-square prediction error as cost function. The stations' climatology (spline-smoothed amounts) is the reference (obtained from the full record). The average mean-square prediction error skill score (MSPESS) for a daily values is 5.9% and thus positive.

Analogously, we use the DIVA-GIS altitude profile to obtain a full spatial picture of the rainfall amounts (Fig. 7). Also, the amounts show a north–south gradient, tilted toward the east at the beginning of the monsoon phase and toward the west at the peak. The strongest rainfall amounts can be observed in the southwest in July. Figures 6 and 7 corroborate the impression that the seasonal variation in the east is smaller than in the west. The animated sequence of plots showing the development of daily precipitation amounts for the whole country during the wet season is available as supporting online material.

These two baseline models reconstruct spatial fields of expected daily occurrence probabilities and daily rainfall amounts from gauge-based data assuming that the expected values vary smoothly in space and time. Given the stations' climatology, the baseline models give spatiotemporal climatologies for daily precipitation occurrence probabilities and daily rainfall amounts for the whole country.

## 5. Modeling the effect of weather types

Weather types are now added as a categorical variable to the predictor  $\boldsymbol{\eta} = \boldsymbol{\eta}_{\text{baseline}} + \boldsymbol{\eta}_{\text{WT}}$ , with  $\boldsymbol{\eta}_{\text{WT}} = \sum \beta_{\text{WT},j} \boldsymbol{\Theta}(\text{WT} = \text{WT}_j)$  having essentially one coefficient  $\beta_{\text{WT},j}$  per weather type  $j$ ;  $\boldsymbol{\Theta}_i(\text{WT} = \text{WT}_j) = 1$  if the

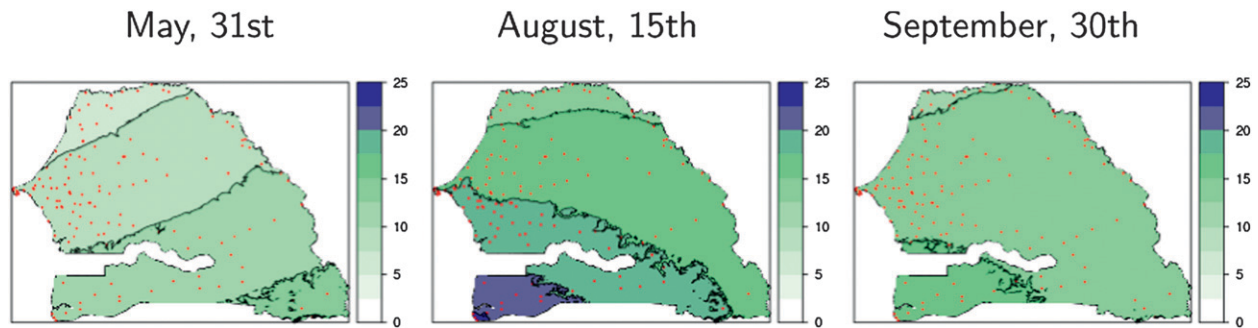


FIG. 7. Spatial predictions for the baseline rainfall amount model for (left) late May, (center) mid-August, and (right) late September; colors depict rainfall amount ( $\text{mm day}^{-1}$ ) with black lines separating different color shadings.

weather type  $j$  is present at day  $i$  and zero otherwise. We use all four definitions of weather types (raw/grid point, raw/variable, anomalies/grid point, and anomalies/variable) resulting from the different treatment prior to clustering. The aim of this section is to determine the optimal number of types to be used and to show that skill at the local and daily scale is obtained by the inclusion of large-scale weather types. We try various numbers of types and choose the optimal number a posteriori according to model performance. Performance is measured using the AIC and Brier score or mean-square error (MSE), as in the previous section. The associated skill scores with the mean seasonal cycle as reference show that positive skill is obtained at the local daily scale.

#### a. Occurrence model

The improvement of the occurrence model dependent on the number of weather types used is shown in Fig. 8, using again the modified AIC and the Brier score obtained from cross validation. Anomaly weather types are shown in the top row, raw values are shown in the bottom row, gridpoint normalization is in the left column, and variable-based normalization is in the right column. The straight lines at the top represent the baseline model. For gridpoint-scaled anomalies (top left) model improvement is large, using four instead of three types, and is only minor thereafter until another larger improvement occurs from 16 to 17 types. The variable-scaled anomalies (top right) show a similar behavior when adding 4 instead of 3 types and another strong improvement from 7 to 10 types. Improvement is minor for more than 10 types. Adding weather types based on raw values (bottom row) has a somewhat different effect: adding gridpoint-scaled types (left) leads to an improvement until about 14 types, stagnating thereafter. For variable-scaled types, the performance gets worse for more than 14 types. For the raw variable-scaled types, an even number seems to be preferable, leading to an alternation between decline and increase of the curve

in the range of 5–15 types. According to the Brier score, the overall model improvement is largest for variable-scaled raw values (bottom right) and 14 types. A Brier skill score calculated on the basis of a new independent 10-fold cross-validation experiment in space and time (i.e., taking out arbitrary observations at arbitrary stations) with a mean annual cycle as reference yields a 14-type skill score ( $SS_{14} \approx 3.4\%$ ). Including large-scale weather types in the model leads thus to skill at the small scale (daily values at station level). The small absolute value indicates that the seasonal cycle is the dominant signal on the intraannual (i.e., daily) scale. In section 6, we investigate this improvement on the interannual time scale where the seasonal cycle does not mask the interannual variability anymore. The model performance on the interannual time scale is more readily visible there.

#### b. Amount model

Analogously, weather types are added to the predictor  $\boldsymbol{\eta}$  of the rainfall amount model. The AIC and the mean prediction error based on cross validation are shown in Fig. 9. Both show very similar tendencies. Anomaly based types (top row) lead to an improvement for small numbers ( $<7$ ), basically stagnating thereafter. For weather types based on raw values, model improvement can be observed until about 16 types, although not in a monotonous way. Again, an even number of weather types seem to be preferable for the variable-scaled raw values in the range of 5–10 types. The minimum mean-square prediction error is obtained for variable-scaled raw values and 16 types and yields a skill score based on an independent 10-fold cross validation in space and time, with reference to the mean seasonal cycle of  $SS_{16} \approx 6.3\%$ . Again, including the large-scale weather types leads to skill on the local scale for daily observations. The small absolute values are a consequence of the dominating seasonal cycle on the intraannual time scale.

To obtain an optimal combined model for rainfall occurrence and amount, the mean-square prediction errors

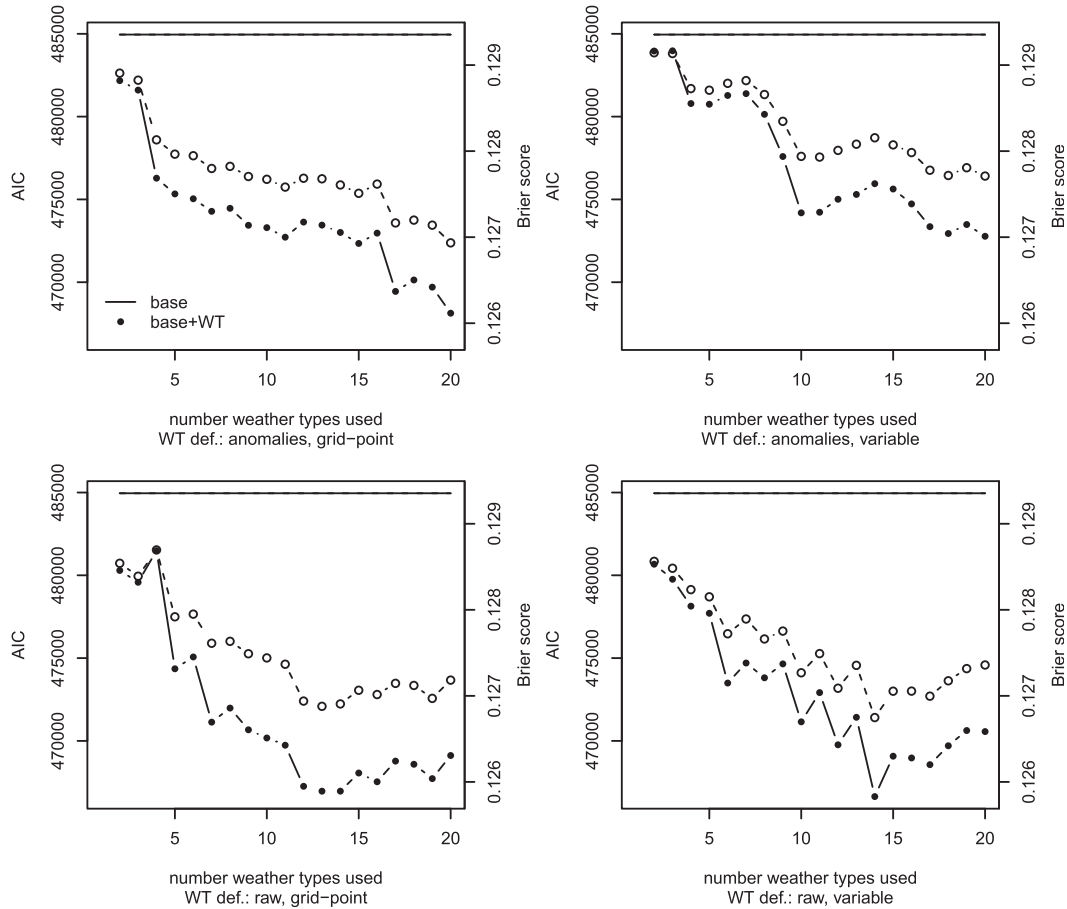


FIG. 8. Rainfall occurrence model improvement compared to the baseline model (straight lines at the top) for an increasing number of weather types (abscissa). Weather types are defined by (top) anomalies or (bottom) raw values and (left) gridpoint-based and (right) variable-based normalization. Open symbols refer to the AIC scale (left axis) and closed symbols to the Brier score (right axis).

(MPEs) from the occurrence and amount models have been combined into an overall skill score

$$SS_{\text{comb}} = 1 - \left[ \frac{MPE_{\text{Occ}}(\text{model}) \quad MPE_{\text{Int}}(\text{model})}{MPE_{\text{Occ}}(\text{baseline}) \quad MPE_{\text{Int}}(\text{baseline})} \right]. \quad (3)$$

Figure 10 shows the skill scores for the four variants of defining weather types. Types based on variable-scaled anomalies (filled squares) show the lowest combined skill score for all numbers of types. Raw value-based types (circles) are superior in the range from 10 to 18 types. The choice of normalization has little effect as both curves increase in a quite similar manner. For anomalies, gridpoint normalization yields larger skill scores for all numbers of types, thus gridpoint-based normalization is advantageous here. The absolute maximum combined skill score in the given range is attained for 14 types defined on variable-scaled raw values (closed

circles;  $SS_{\text{comb},14} \approx 4.0\%$ ), and it is chosen as model for the subsequent analysis.

*c. Weather-type effect*

For a set of weather types to be useful as predictors for precipitation, the types must be able to separate local conditions with different precipitation characteristics. Figure 11 shows the types ranked according to their spatially averaged anomaly (taken with respect to the baseline model) occurrence probability (left) and amount (right). For each weather type, this effect is evaluated at the day in the wet season where the respective type is most likely to occur, see also section 5d. The association of colors to the types and the order of types on the abscissa, as defined in Fig. 11 (left), is kept as follows. Types associated with an increase in precipitation occurrence do not generally lead to an increase in amount. Two of the three types with the largest increase in occurrence probability exhibit a negative anomaly for

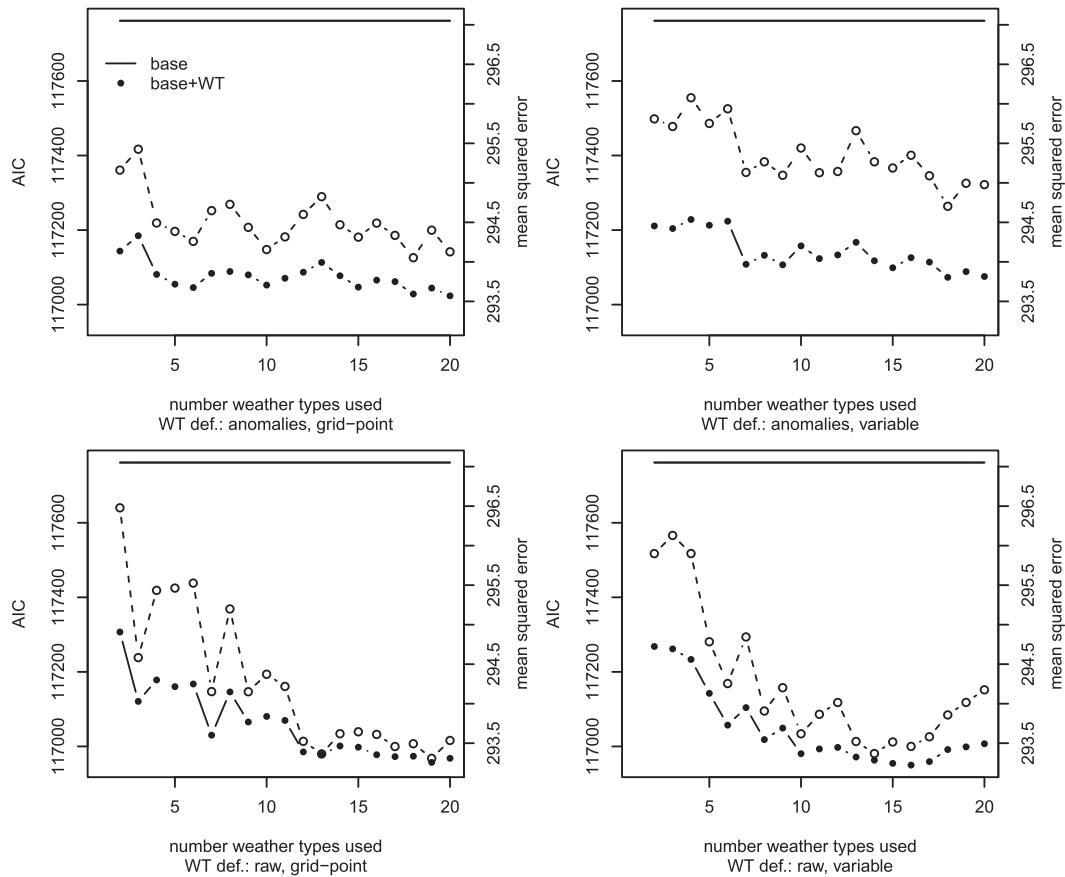


FIG. 9. Precipitation amount model improvement compared to the baseline model (straight lines) for an increasing number of weather types (abscissa). Weather types are defined by (top) anomalies or (bottom) raw values and (left) gridpoint-based and (right) variable-based normalization. Open symbols refer to the AIC scale (left axis) and closed symbols to the mean-square prediction error (right axis).

amount (types 6 and 2); one shows a strong positive anomaly (type 4). Types 11 and 9 show very similar occurrence anomalies, while their amount anomalies have the opposite sign.

Figure 12 depicts the spatially resolved differences to the baseline model for occurrences (left) and amounts (right) at the day the particular weather type is most likely to occur. For occurrence probability, the wettest types (types 4, 2, 1, 6, 11, 9, and 14) show a north–south gradient, which is steeper for the former wetter types and getting flatter toward the dryer ones. For types 10, 7, and 12, the gradient is tilted toward a northwest–southeast gradient, and it is inverted for type 8. Types 3, 13, and 5 have very little or no spatial structure. Similar to rainfall amounts, the north–south gradient (types 4, 2, 11, 9, 14, and 3), as well as the tilted gradient (types 1, 6, 10, 3, 7, and 8), are found. Very fine structures can be observed for the amounts of types 5 and 12; they stem from the high-resolution altitude profile and depict small-scale changes in altitude, such as river beds. Although this

precipitation model is linked to altitude, it is not designed to resolve precipitation on these small scales. These structures must not be overinterpreted.

The (tilted) north–south gradient is a result of the mesoscale systems passing more frequently in the south (east) than in the north (west) (Mathon and Laurent 2001). Certain types (toward the top of Fig. 12) are related to an increased generation of these systems, other types (toward the bottom of Fig. 12) are linked to the inhibition of their formation.

#### d. Weather-type occurrence frequencies and transitions

The types associated with the presented response pattern have different occurrence frequencies and dominate at their specific period of the year. Figure 13 depicts the average number of days per wet season when a type is present. Two dry (types 8 and 13) and one intermediate type (type 7) dominate in frequency with more than 20 days per wet season. Other types are present on average



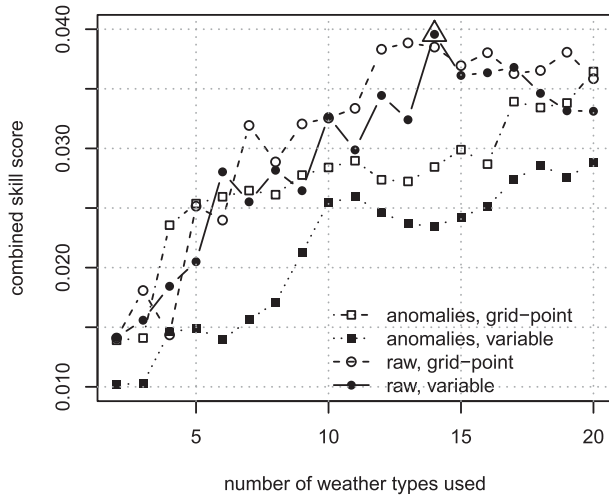


FIG. 10. Combined forecast skill score for occurrence and amount model for the four variants of defining weather types. The triangle depicts the maximum combined skill score.

10–14 days while types 14 and 10 occur only about 5 days per season.

Logistic regression can be straightforwardly used to describe the temporal course of type occurrence as shown in Fig. 14. The occurrence of types based on raw values shows the expected pronounced intraseasonal variability (cf. Moron et al. 2008a, their Fig. 4). Types 12, 3, 5, and 7 occur typically at the onset of the monsoon, while types 10, 8, 1, 6, 4, 9, and 2 occur

during the peak phase and types 13, 11, and 14 occur during the offset.

The estimated transition matrix between types allows a flowchart of type transitions to be drawn. Figure 15 shows those transitions, which are more likely than the unconditional occurrence frequency of the following types. Types occurring frequently during the onset of the monsoon are depicted at the top of Fig. 15, peak monsoon types are in the center, and offset types are at the bottom. Except for the rarely occurring type 10, three disjoint transition systems form: one between onset types, another involving peak monsoon types, and a third for offset types.

*e. Weather-type composites*

Composites of the weather types are given in Fig. 16 as absolute values (columns one and three) and the relative seasonal mean (columns two and four). Peak monsoon types are arranged roughly in order of appearance in the seasonal cycle in columns one and two. Types related to the onset and offset are in columns three and four.

*f. Discussion*

At the beginning of the wet season, the dry Harmatan winds manifest in type 12 with a strong negative humidity anomaly over Senegal and generally dry precipitation conditions. In May, this type is frequently interrupted by type 3, with northwesterly winds over the Sahara, a weakening of the northeasterlies over Senegal,

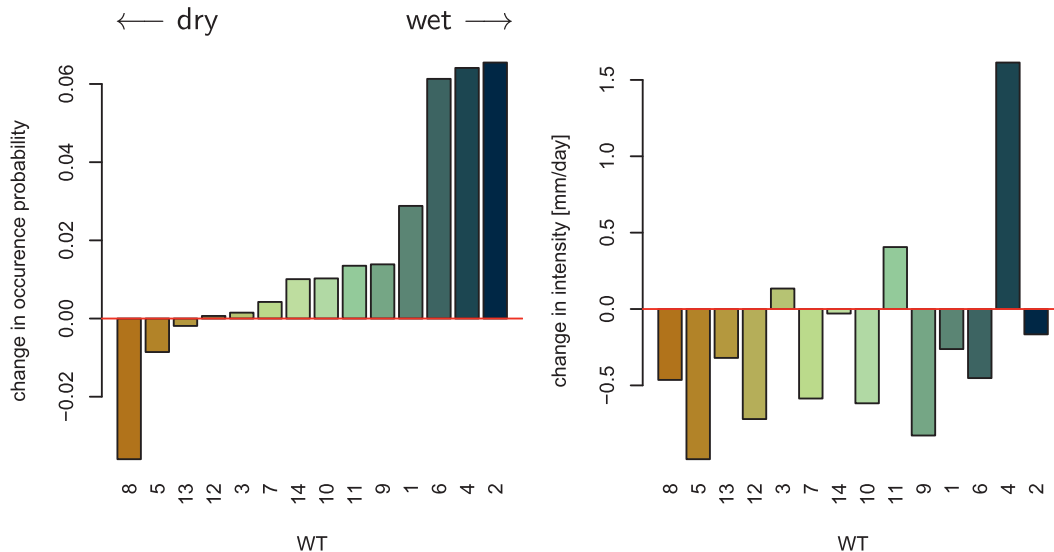


FIG. 11. Weather type–induced change in (left) precipitation occurrence (probability) and (right) amount (mm day<sup>-1</sup>). Given are the spatial averages of the anomalies for the occurrence and amount model to their baseline models. Models are evaluated at the day the respective weather types show their highest probability of occurrence. In the left panel, the color correlates with the bar heights and is hereby associated to the weather types. This color type relation is maintained in the right panel; bar height now indicates change in rainfall amount.

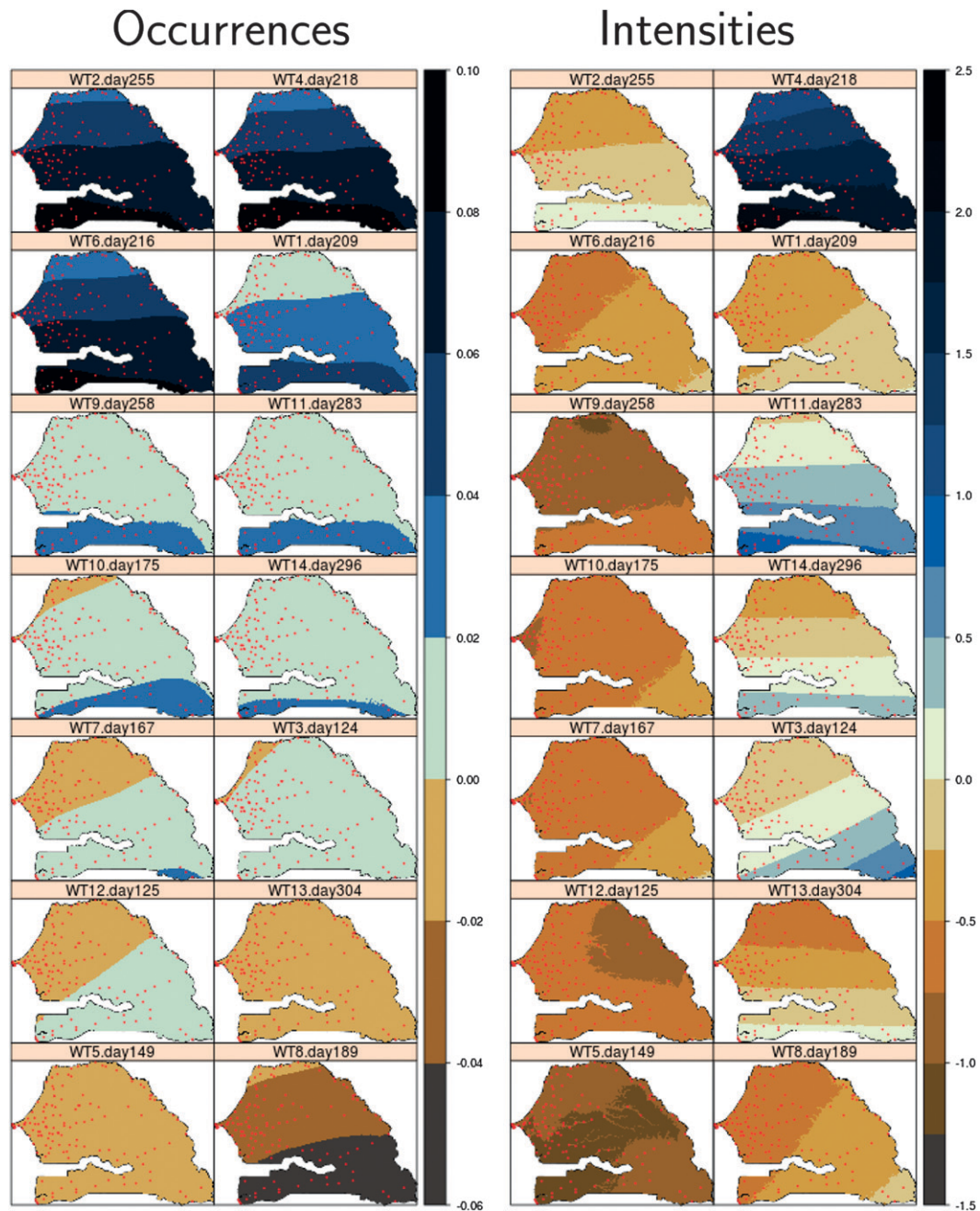


FIG. 12. Spatially resolved differences to the baseline model for (left) precipitation occurrence (probability) and (right) amount ( $\text{mm day}^{-1}$ ). As a given weather type can occur at various stages of the monsoon cycle, its difference to the baseline model is obtained for the day the given weather type is most likely to occur (cf. Fig. 14). Types are sorted from top to bottom and left to right according to their spatially averaged occurrence anomalies. Station locations are indicated by red dots.

and a slight increase in precipitation probability. Later in May the very dry type 5 appears, again strengthening the dry northeasterlies over Senegal. Type 7 dominates in June with wind patterns similar to those of type 5 but with less dry air over Senegal; it brings along an increase in occurrence probability particular in the southeast.

In late June, with a notable increase in precipitation occurrence in the south, type 10 marks the transition to peak monsoon patterns. The latter occurs in two typical cycles, which are interconnected (cf. Fig. 15): in late July–beginning of August, types 8, 1, 4, and 6 form a dominant cycle and later in August types 4 and 1 are

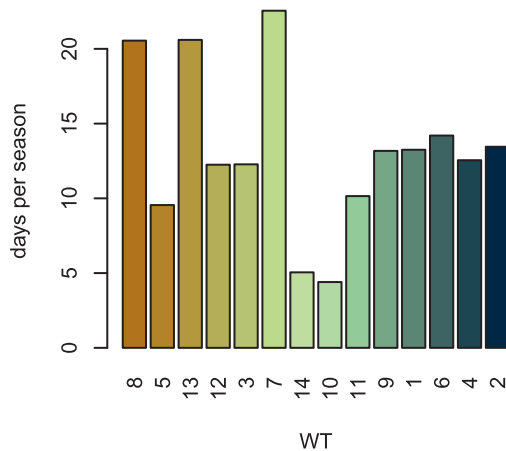


FIG. 13. Seasonal frequencies of weather types. Color coding from Fig. 11.

replaced by types 2 and 9 leading to a new cycle. Toward the end of the monsoon phase, type 6 occurs less frequently and thus is less frequently part of the cycle. Both cycles start with relatively dry air and northeasterly winds north of Senegal (type 8) resulting in one of the dry patterns found. In July and August, a subsequent sudden rise in humidity is related to the succession of type 1 and 4 and associated with a strong increase in occurrence probability and rainfall amounts (only type 4). The subsequent southeasterlies north of the country (type 6) indicate the backside of a westward moving trough and thus a passing easterly wave. This cycle quickly repeats with a transition from type 6 to 1 or decays toward the dry type 8 via type 9. Later in August and September, types 4, 6, and 1 are replaced by types 9

and 2 with lower occurrence probabilities and amounts. While both cycles show typical characteristics of westward AEWs, the anomalies depicted in Fig. 16 show that the characteristic features of types 1 and 4 (positive anomaly of humidity and westerly winds over Senegal) are less pronounced in the later types 2 and 9. The monsoon phase ends with a final cycle involving types 11, 14, and 13. While types 11 and 14 are associated with increased occurrence probability in the south and a pronounced north–south gradient for amounts, type 13 is dominating in October and brings back the dry Harmatan winds and heralds the dry season.

**6. Interannual variability**

Figure 17 depicts the observed annual-mean rainfall occurrence and amount over all available stations in the country for all days in the monsoon season as well as the corresponding model predictions. As observations are missing more or less randomly at different locations in space and time in the period of investigation, a fraction of interannual variability is due to this incompleteness. To reduce this effect, observations and model predictions are given relative to the baseline models at the respective spatial and temporal locations. From 1965 to 2000, observed mean precipitation occurrence varies between 70% and 130% of the 40-yr mean. The interannual variability of observed rainfall amount is smaller (80%–115%) and is—apart from a few years—almost flat in the time period 1975–95. Figure 17 further shows a tendency for observed occurrence and amount to be larger in the 1960s and lower toward the 1980s, thus reproducing the well-known behavior of Sahelian

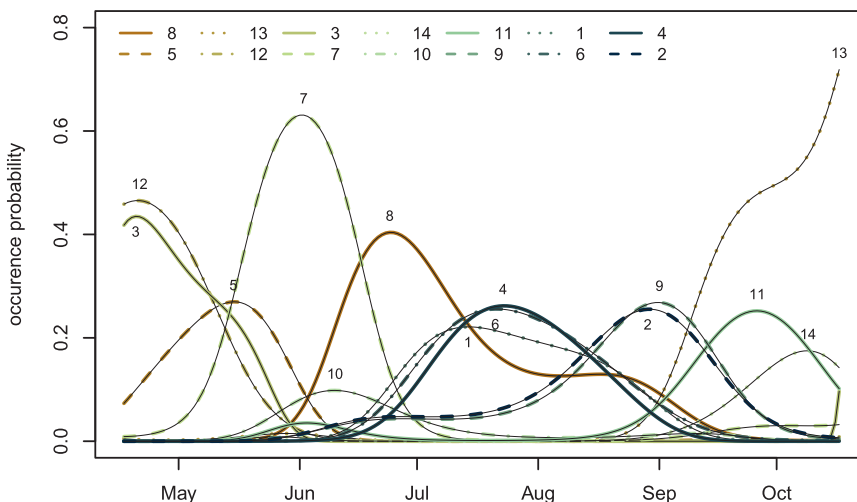


FIG. 14. Estimated weather-type occurrence probability during the monsoon season. Type numbers are noted just above or below the corresponding peaks in occurrence probability. The color code follows Fig. 11.

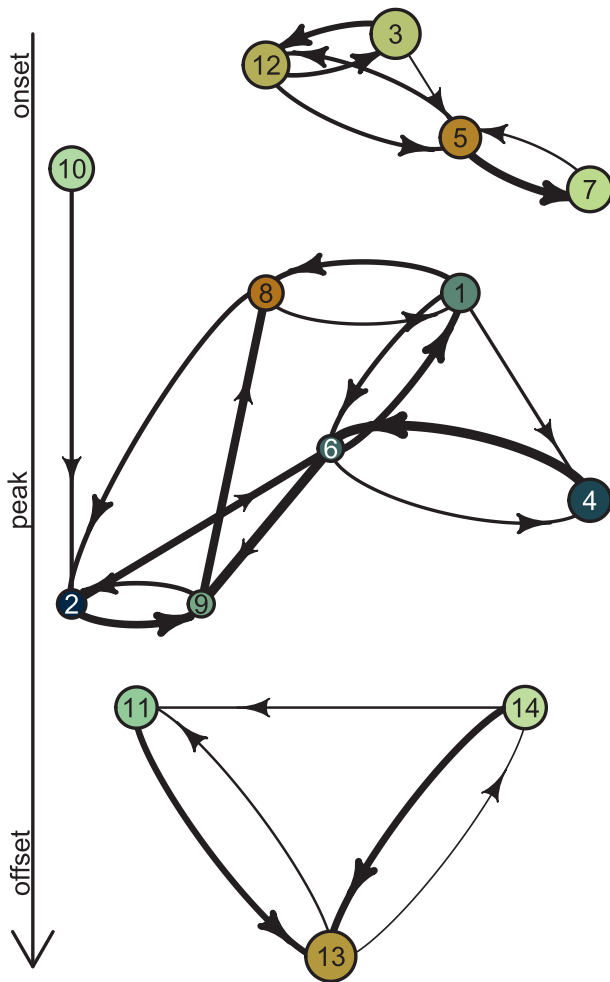


FIG. 15. Graph of the weather-type transition matrix. The size of the circles is proportional to the persistence; the width of the arrows is proportional to the transition probability. Arrows are drawn only if the transition probability exceeds the unconditional occurrence probability of the target type.

precipitation as reported by Nicholson (2001), a tendency reproduced by the model. This tendency is, however, more pronounced for occurrences. The occurrence and amount models capture 26% and 38% of the total interannual variability and estimates of their Pearson correlation coefficients yield  $r_O = 0.52$  and  $r_I = 0.65$ , respectively. For the latter, a significance test leads to  $p$  values of  $p < 10^{-2}$  based on 10 000 replicates obtained via the iteratively amplitude-adjusted Fourier transform (Schreiber and Schmitz 1996; Venema et al. 2006). Thus, the interannual variability of the occurrences is not captured by the model as well as the amount signal. The latter is flatter on the decadal scale and a lot smoother on the interannual scale.

Depicting annually resolved shares of occurrences, Fig. 18 gives a visual impression of the interannual

variability in the occurrence of types. Drier types are relatively rare in the first half of the 1960s, peaking in the beginning of the 1970s and in the mid-1980s. The wet types 2 and 4 have a pronounced share in the beginning of the 1960s declining later and recovering toward the end of the twentieth century. By that time the intermediate types 6 and 9 hold a notable share. Those types have not been occurring at that rate in the 1960s. The interannual variation in precipitation occurrence and frequency as depicted in Fig. 17 is plausible given the dynamics of weather types. This corroborates the idea that there is a link between the interannual variability in weather-type occurrence and precipitation for Senegal.

## 7. Summary and discussion

### a. Summary

We give a spatiotemporal statistical description of daily precipitation occurrence and amount for Senegal. Two variants of generalized linear models were used: logistic regression describing precipitation occurrence and regression with a gamma-distributed response for precipitation amounts. In a first step, baseline models were built using the phase of the seasonal cycle, the spatial location, and the previous day observations as input. A selection of relevant input factors was accomplished by comparing the performance measures from a cross-validation experiment. The resulting baseline models describe the spatially resolved intraannual variability and serve as spatiotemporal climatologies. As the locations of the gauges directly enter the model, spatial interpolation is possible on the predictor level. The temporal evolution of the spatial occurrence and amount patterns during the wet season can be visualized. A general north–south gradient of both occurrence and amounts is the dominant signal. This gradient is tilted toward the east at the onset of the monsoon and toward the west for the late monsoon phase. The tilt is a lot more pronounced for amounts than for occurrences.

A link to atmospheric dynamics was established using a set of weather types defined from the NCEP–NCAR reanalysis 850-hPa winds and 925-hPa relative humidity within a  $30^\circ \times 30^\circ$  box centered over Senegal. These weather types are an efficient summary of the possible states of the local atmospheric dynamics and are expected to describe a notable fraction of the interannual variation of precipitation. The optimal number of 14 types, as well as the preferred filtering and normalization, was determined by the performance of the precipitation model.

The spatial averages of the occurrence responses to the various weather types lead to a ranking of the types



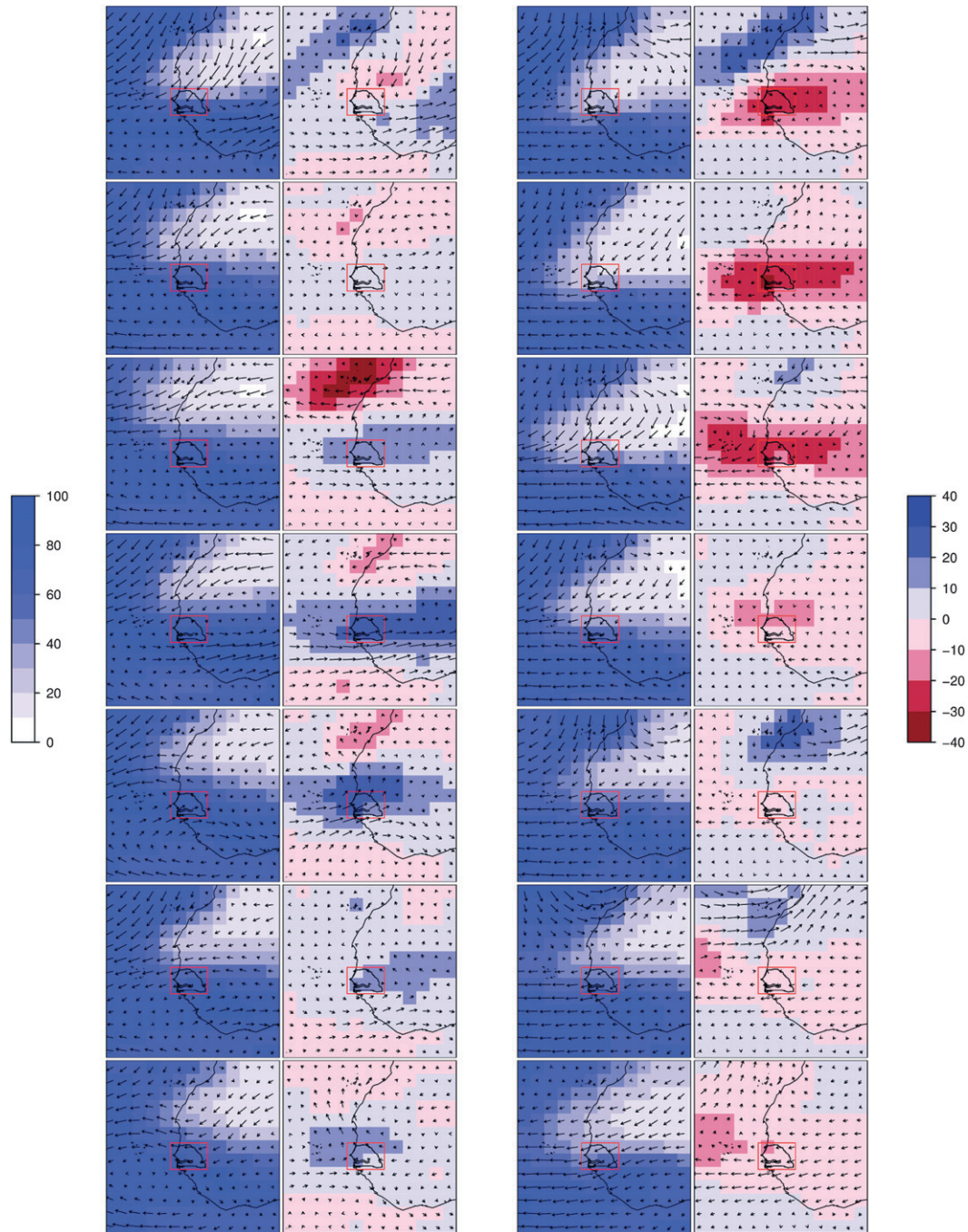


FIG. 16. Weather-type composites of absolute values (first and third column) and associated difference to the seasonal mean fields (second and fourth column) of relative humidity. Vectors depict the 850-hPa wind field; relative humidity is color coded. (top)–(bottom) The peak monsoon types 10, 8, 1, 4, 6, 2, and 9 are in the first column; onset 3, 12, 5, and 7 and offset types 11, 14, and 13 are in the second column.

from dry to wet. The amount response does not follow this ranking, which suggests that occurrence and amounts do not respond in the same way to the weather types: a posteriori justifying a separate modeling. Weather types were further characterized by composites of their

absolute values of wind and relative humidity, as well as by the differences of the composites to the respective seasonal mean fields. Together with the time of their occurrence, the transition between types, and the associated response in precipitation occurrence and amounts,



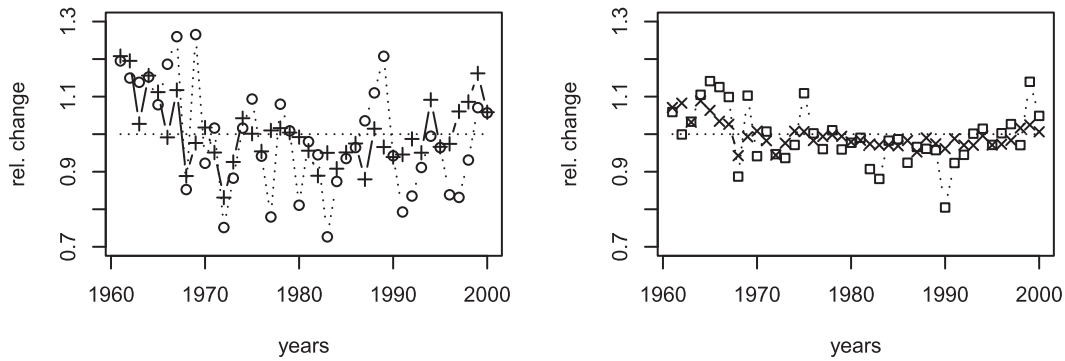


FIG. 17. Interannual variability of (left) spatially and temporally averaged rainfall occurrence and (right) amount relative to the baseline model. Observations for a given year are averaged (dotted line; open symbols) and compared to the predictions (solid line; closed symbols) of the occurrence and amount. The range of the ordinate is the same in both plots to illustrate the lower overall variability of precipitation amount.

a consistent picture of the types' dynamics and the associated response in rainfall could be drawn. Besides having a four-type dynamic during the onset and three types for the offset, we could identify two connected cycles in the peak monsoon phase: one dominating end of July and August involving types related to the largest occurrence and strongest amount patterns. Toward September, this cycle smoothly translates into a second cycle involving types with less rain. Both cycles can be related to the African easterly waves.

If an intraseasonal change in occurrence frequencies and/or transition frequencies of types leads to a shift in the monsoon, the model developed will be able to track this shift to what extent it reproduced the weather type associated rainfall patterns. The occurrence frequency of the weather types based on the 850-hPa wind field and 925-hPa relative humidity exhibits a pronounced interannual variability, expected to propagate to the

precipitation patterns. This modeling approach quantifies the effect of these types from precipitation: they capture 26% ( $r_O = 0.52$ ) and 38% ( $r_I = 0.65$ ) of the spatially averaged interannual variability of precipitation, respectively. Thus, this set of weather types together with their associated spatially resolved precipitation responses contribute to resolving the question of how local precipitation responds to large-scale atmospheric dynamics.

#### b. Discussion

It is a priori not evident whether types defined on absolute (raw) values or anomalies are to be preferred for downscaling precipitation. Comparing the two variants results in a clear advantage for types defined on absolute values for the present study, whereas the differences in normalization do not lead to major improvements.

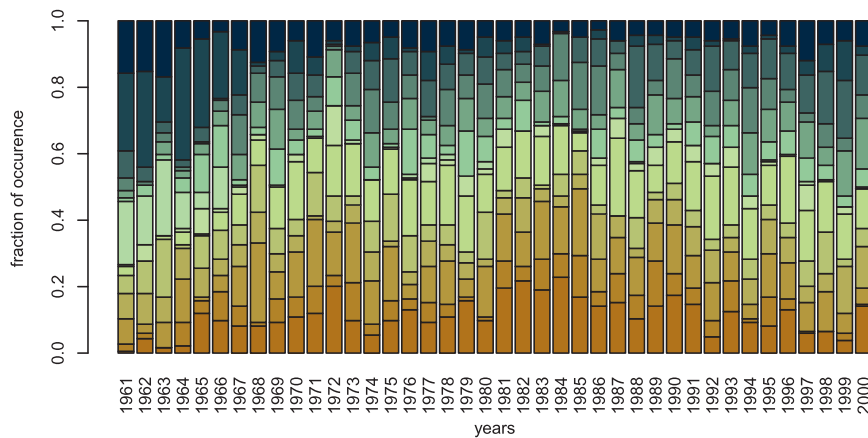


FIG. 18. Interannual variability of weather-type occurrence. The colors represent precipitation occurrence probabilities associated with the types (color coding taken from Fig. 11).

Frequently, the choice of the number of types to be used is made before developing the statistical downscaling model. However, Moron et al. (2008a) chose the number of types partially according to an information criterion involving the discrimination of rainfall by the weather types and thus already with the application in mind. We suggest a more consequent implementation of this idea and choose the number of types according to the performance of the precipitation model. Among the range of the number of types tested, 14 types yield the optimal model performance. There is, however, a considerable fraction of remaining variability for both occurrences and amounts. Precipitation amounts on wet days are better described than precipitation occurrences on daily as well as interannual time scales. For Senegal, with mesoscale convective systems being the most important rainfall generating mechanisms, this implies that it is easier to predict mean daily amounts given that a mesoscale system is passing a specific location than predicting whether a system is arriving at all at that location. The lower described variability of occurrence probability on the interannual scale points toward the importance of influences not captured by the weather types. The amount signal is instead smoother on that scale and captured to a larger extent by the weather types. This underlines the difference in mechanisms for precipitation occurrences and amounts in the sense that they do not respond in the same way to large-scale atmospheric drivers. It supports the separate modeling approach pursued here.

A potential candidate for driving precipitation—and for future model refinement—is certainly sea surface temperature (SST), as it might influence precipitation in a different way than via winds and relative humidity captured by the weather types. According to the previous paragraph, the predictors for occurrence and amounts do not necessarily have to be constructed from the same variables as

it is the case in this study. Initially, we included also sea level pressure for the definition of types; it was left out as it did not improve the results.

We expect an improvement in model performance when including the previous day weather type as a predictor. Currently, antecedent information is only indirectly available through the previous day precipitation.

The GLM approach with spatial location explicitly entering the predictor allows a spatially resolved modeling of precipitation occurrence probability and amount. A pointwise realization of precipitation scenarios is straightforwardly possible (e.g., to be used for agricultural models). In the context of climate change projections, the model can be used for downscaling under the common assumption that the weather types remain the same but their frequency changes. Furthermore, as the result is a probability distribution for occurrence probability and amounts varying in space and time, this model allows, for example, a mapping of the drought risks or risk of the exceedance of a certain amount of rain in a given time span. As these maps can be generated from GCM climate change projections, they are a valuable tool for impact assessment.

*Acknowledgments.* This work was financially supported by the GIS REGYNA project. We thank A. Fink and V. Ermert for inspiring discussions. The authors are furthermore much obliged to three anonymous referees for valuable and constructive comments, which substantially influenced this work.

## APPENDIX

### Technical Supplements

#### a. Goff–Gratch equation

$$\log_{10} p_{\text{sat}} = -7.90298 \left( \frac{373.16}{T} - 1 \right) + 5.02808 \log_{10} \left( \frac{373.16}{T} \right) - 1.3816 \times 10^{-7} \{ 10^{11.344[1-(T/373.16)]} - 1 \} + 8.1328 \times 10^{-3} \{ 10^{-3.49149[(373.16/T)-1]} - 1 \} + \log_{10}(1013.246).$$

#### b. Model performance

Model performance is assessed in two ways: (i) with an information criterion (modified AIC) and (ii) with a prediction error estimated with cross validation. The modified AIC is given by

$$\text{AIC} = -2 \log L(\theta | \mathbf{x}_1 \dots \mathbf{x}_N) + km, \quad (\text{A1})$$

with  $L$  as the likelihood,  $m$  as the number of parameters used, and  $k = 6.63$  (standard AIC:  $k = 2$ ). This corresponds roughly to a likelihood-ratio test against the  $\chi^2$  distribution on the 1% level.

Furthermore, we use cross validation to calculate a mean prediction error. For the latter, the set of stations is randomly divided into three disjunctive sets of roughly equal size and roughly uniform spatial distribution.

Model parameters are estimated for all stations in two of the three sets and a mean-square prediction error  $\Phi$  is calculated for all stations and all days in the third set. This is repeated, exchanging the sets until a mean prediction over all stations is obtained. For the amount model, the mean-square prediction error is calculated as

$$\Phi_r = 1/N \sum_i [r_i - g_r^{-1}(\eta_i)]^2, \quad (\text{A2})$$

with the inverse link function  $g^{-1}$  and observations of amounts  $r_i$ . As precipitation occurrence is described with a binomial variable, the equivalent measure is derived from the Brier score (Brier 1950; Friederichs and Hense 2008)

$$\Phi_\pi = 1/N \sum_i [\pi_i - g_\pi^{-1}(\eta_i)]^2. \quad (\text{A3})$$

A forecast skill score with respect to a reference model can be calculated as (Wilks 2006)

$$\text{SS} = 1 - \frac{\Phi(\text{model})}{\Phi(\text{reference})}. \quad (\text{A4})$$

### c. Harmonic functions

Similar to Rust et al. (2009) and Maraun et al. (2009), we use harmonic functions to describe seasonal variability:

$$\eta_{\text{season}} = \sum_i \left[ \beta_{i,\sin} \sin\left(2\pi \frac{\text{idoy}(t)}{365.25}\right) + \beta_{i,\cos} \cos\left(2\pi \frac{\text{idoy}(t)}{365.25}\right) \right], \quad (\text{A5})$$

where  $\text{idoy}(t)$  gives the day of the year for a given date  $t$  [e.g., for  $\text{idoy}(1 \text{ May } 2000) = 122$ ].

### d. Legendre polynomials

Legendre polynomials are a series of orthogonal polynomials on the subset  $[-1, 1]$  of real numbers and are used here to describe spatial variability in the baseline model. We use

$$\eta_{\text{spatial}} = \sum_j \beta_{j,P} P(x, j) + \sum_k \beta_{k,P} P(y, k) + \sum_l \beta_{l,P} P(z, l), \quad (\text{A6})$$

with  $P(x, j)$  being the Legendre polynomial for order  $j$  (Abramowitz and Stegun 1965). Consequently, we shifted and rescaled the spatial coordinates in such a way that they obtain values between  $-1$  and  $1$  for longitude ( $x$ ) and latitude ( $y$ ) or  $0$  and  $1$  for altitude ( $z$ ) in the cube

defined by  $18^\circ\text{--}11^\circ\text{W} \times 12^\circ\text{--}17^\circ\text{N} \times 0\text{--}400\text{ m}$ . The longitudinal and latitudinal extension covers the whole country; altitude is restricted to  $400\text{ m}$ , which includes the highest rain gauge used here (Fongolimby at  $396\text{ m}$ ). Precipitation data for higher elevations, like Nepen Diakha ( $584\text{ m}$ ), are not available. Similar to harmonic functions, Legendre polynomials form an orthogonal basis and ensure the independence of successively included factors of higher order.

## REFERENCES

- Abramowitz, M., and I. Stegun, Eds., 1965: *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*. Dover, 1046 pp.
- Ambrosino, C., R. Chandler, and M. C. Todd, 2011: Southern African monthly rainfall variability: An analysis based on generalized linear models. *J. Climate*, **24**, 4600–4617.
- Beck, C., and A. Philipp, 2010: Evaluation and comparison of circulation type classifications for the European domain. *Phys. Chem. Earth*, **35**, 374–387.
- Benestad, R. E., 2001: A comparison between two empirical downscaling strategies. *Int. J. Climatol.*, **21**, 1645–1668, doi:10.1002/joc.703.
- Biasutti, M., I. M. Held, A. H. Sobel, and A. Giannini, 2008: SST forcings and Sahel rainfall variability in simulations of the twentieth and twenty-first centuries. *J. Climate*, **21**, 3471–3486.
- Boé, J., L. Terray, F. Habets, and E. Martin, 2006: A simple statistical-dynamical downscaling scheme based on weather types and conditional resampling. *J. Geophys. Res.*, **111**, D23106, doi:10.1029/2005JD006889.
- Branstator, G., and F. Selten, 2009: “Modes of variability” and climate change. *J. Climate*, **22**, 2639–2658.
- Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, 1–3.
- Chandler, R. E., 2005: On the use of generalized linear models for interpreting climate variability. *Environmetrics*, **16**, 699–715.
- Coe, R., and R. D. Stern, 1982: Fitting models to daily rainfall data. *J. Appl. Meteor.*, **21**, 1024–1031.
- Conway, D., and P. D. Jones, 1998: The use of weather types and air flow indices for GCM downscaling. *J. Hydrol.*, **212–213**, 348–361.
- Deme, A., A. Viltard, and P. de Felice, 2003: Daily precipitation forecasting in Dakar using the NCEP–NCAR reanalyses. *Wea. Forecasting*, **18**, 93–105.
- Diedhiou, A., S. Janicot, and A. Viltard, P. de Felice, and H. Laurent, 1999: Easterly wave regimes and associated convection over West Africa and tropical Atlantic: Results from NCEP/NCAR and ECMWF reanalysis. *Climate Dyn.*, **15**, 795–822.
- Dilley, M., R. S. Chen, U. Deichmann, A. L. Lerner-Lam, and M. Arnold, 2005: Natural disaster hotspots: A global risk analysis. Disaster Risk Management Series 5, World Bank Tech. Rep., 132 pp.
- Folland, C. K., T. N. Palmer, and D. E. Parker, 1986: Sahel rainfall and worldwide sea temperatures, 1901–85. *Nature*, **320**, 602–607, doi:10.1038/320602a0.
- Fowler, H. J., S. Blenkinsop, and C. Tebaldib, 2007: Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modelling. *Int. J. Climatol.*, **27**, 1547–1578.

- Fraley, C., and A. E. Raftery, 2002: Model-based clustering, discriminant analysis, and density estimation. *J. Amer. Stat. Assoc.*, **97**, 611–631.
- , and —, 2007: Model-based methods of classification: Using the *mlcst* software in chemometrics. *J. Stat. Software*, **18** (6), 1–13.
- Friederichs, P., and A. Hense, 2008: A probabilistic forecast approach for daily precipitation totals. *Wea. Forecasting*, **23**, 659–673.
- Goff, J. A., and S. Gratch, 1946: Low-pressure properties of water from  $-160$  to  $212^{\circ}\text{F}$ . *Trans. Amer. Soc. Heat. Vent. Eng.*, **51**, 125–164.
- Gueye, A. K., S. Janicot, A. Niang, S. Swadogo, B. Sultan, A. Diongue-Niang, and S. Thiria, 2011: Weather regimes over Senegal during the summer monsoon season with self-organizing maps. Part I: Synoptic time scale. *Climate Dyn.*, **36**, 1–18.
- Haile, M., 2005: Weather patterns, food security and humanitarian response in sub-Saharan Africa. *Phil. Trans. Roy. Soc.*, **260**, 2169–2182.
- Haines, K., and A. Hannachi, 1995: Weather regimes in the Pacific from a GCM. *J. Atmos. Sci.*, **52**, 2444–2462.
- Hannachi, A., 1997: Low-frequency variability in a GCM: Three-dimensional flow regimes and their dynamics. *J. Climate*, **10**, 1357–1379.
- , 2007: Tropospheric planetary wave and mixture modeling: Two preferred regimes and a regime shift. *J. Atmos. Sci.*, **64**, 3521–3541.
- , I. T. Jolliffe, and D. B. Stephenson, 2007: Empirical orthogonal functions and related techniques in atmospheric science: A review. *J. Climatol.*, **27**, 1119–1152.
- Huth, R., 1996: An intercomparison of computer-assisted circulation classification methods. *Int. J. Climatol.*, **16**, 893–922.
- , 2000: A circulation classification scheme applicable in GCM studies. *Theor. Appl. Climatol.*, **67**, 1–18.
- Jolliffe, I. T., 2002: *Principal Component Analysis*. 3rd ed. Springer, 518 pp.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.
- Lafore, J.-P., and Coauthors, 2011: Progress in understanding of weather systems in West Africa. *Atmos. Sci. Lett.*, **12**, 7–12.
- Laing, A. G., and J. M. Fritsch, 1993: Mesoscale convective complexes in Africa. *Mon. Wea. Rev.*, **121**, 2254–2263.
- Maraun, D., H. W. Rust, and T. J. Osborn, 2009: The annual cycle of heavy precipitation across the UK: A model based on extreme value statistics. *J. Climatol.*, **29**, 1731–1744.
- , and Coauthors, 2010: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Rev. Geophys.*, **48**, RG3003, doi:10.1029/2009RG000314.
- Mathon, V., and H. Laurent, 2001: Life cycle of Sahelian mesoscale convective cloud systems. *Quart. J. Roy. Meteor. Soc.*, **127**, 377–406.
- McCullagh, P., and J. Nelder, 1989: *Generalized Linear Models*. 2nd ed. CRC Press, 532 pp.
- Michelangeli, P. A., R. Vautard, and B. Legras, 1995: Weather regimes: Recurrence and quasi-stationarity. *J. Atmos. Sci.*, **52**, 1237–1256.
- Moron, V., A. W. Robertson, M. N. Ward, and O. Ndiaye, 2008a: Weather types and rainfall over Senegal. Part I: Observational analysis. *J. Climate*, **21**, 266–287.
- , —, —, and —, 2008b: Weather types and rainfall over Senegal. Part II: Downscaling GCM simulations. *J. Climate*, **21**, 288–307.
- Nicholson, S. E., 2001: Climatic and environmental change in Africa during the last two centuries. *Climate Res.*, **17**, 123–144.
- Paeth, H., and A. Hense, 2006: On the linear response of tropical African climate to SST changes deduced from regional climate model simulations. *Theor. Appl. Meteor.*, **83**, 1–19.
- , K. Born, R. Podzun, and D. Jacob, 2005: Regional dynamical downscaling over West Africa: Model evaluation and comparison of wet and dry years. *Meteor. Z.*, **14**, 349–367.
- Philipp, A., P. M. Della-Marta, J. Jacobeit, D. R. Fereday, P. D. Jones, A. Moberg, and H. Wanner, 2007: Long-term variability of daily North Atlantic–European pressure patterns since 1850 classified by simulated annealing clustering. *J. Climate*, **20**, 4065–4095.
- Preisendorfer, R. W., 1988: *Principal Component Analysis in Meteorology and Oceanography*. Elsevier, 425 pp.
- R Development Core Team, cited 2004: A language and environment for statistical computing. R Foundation for Statistical Computing Rep. [Available online at <http://www.R-project.org/>]
- Roudier, P., B. Sultan, P. Quirion, and A. Berg, 2011: The impact of future climate change on West African crop yields: What does the recent literature say? *Global Environ. Change*, **21**, 1073–1083, doi:10.1016/j.gloenvcha.2011.04.007.
- Rust, H. W., D. Maraun, and T. J. Osborn, 2009: Modelling seasonality in extreme rainfall: A UK case study. *Eur. Phys. J. Spec. Top.*, **174**, 99–111.
- , M. Vrac, M. Lengaigne, and B. Sultan, 2010: Quantifying differences in circulation patterns based on probabilistic models: AR4 multimodel comparison for the North Atlantic. *J. Climate*, **23**, 6573–6589.
- Schreiber, T., and A. Schmitz, 1996: Improved surrogate data for nonlinearity tests. *Phys. Rev. Lett.*, **77**, 635–638.
- Sivakumar, M. V. K., 1988: Predicting rainy season potential from the onset of rains in southern Sahelian and Sudanian climatic zones of West Africa. *Agric. For. Meteorol.*, **42**, 295–305.
- Smyth, P., K. Ide, and M. Ghil, 1999: Multiple regimes in Northern Hemisphere height fields via mixture model clustering. *J. Atmos. Sci.*, **56**, 3704–3723.
- Stern, R. D., and R. Coe, 1984: A model fitting analysis of rainfall data (with discussion). *J. Roy. Stat. Soc.*, **147A**, 1–34.
- Sultan, B., and S. Janicot, 2003: The West African monsoon dynamics. Part I: Documentation of intraseasonal variability. *J. Climate*, **16**, 3389–3406.
- , C. Baron, M. Dingkuhn, B. Sarr, and S. Janicot, 2005: Agricultural impacts of large-scale variability of the West African monsoon. *Agric. For. Meteorol.*, **128**, 98–110.
- Venema, V., S. Bachner, H. W. Rust, and C. Simmer, 2006: Statistical characteristics of surrogate data based on geophysical measurements. *Nonlinear Processes Geophys.*, **13**, 449–466.
- Vrac, M., K. Hayhoe, and M. Stein, 2007a: Identification and intermodel comparison of seasonal circulation patterns over North America. *Int. J. Climatol.*, **27**, 603–620.
- , M. Stein, and K. Hayhoe, 2007b: Statistical downscaling of precipitation through nonhomogeneous stochastic weather typing. *Climate Res.*, **34**, 169–184.
- , P. Yiou, and P. Vaittinada Ayar, 2013: Trends and variability of seasonal weather regimes. *Int. J. Climatol.*, doi:10.1002/joc.3700, in press.
- Wilby, R. L., T. M. L. Wigley, D. Conway, P. D. Jones, B. C. Hewitson, J. Main, and D. S. Wilks, 1998: Statistical downscaling of general circulation model output: A comparison of methods. *Water Resour. Res.*, **34**, 2995–3008.
- Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. Academic Press, 648 pp.