

# Influence of topography on monthly rainfall distribution over East Africa

P. Oetli\*, P. Camberlin

Centre de Recherches de Climatologie, FRE 2740 CNRS, Université de Bourgogne, 6 Boulevard Gabriel, 21000 Dijon, France

**ABSTRACT:** Statistical models were defined for explaining the spatial distribution of rainfall in eastern Africa (southern Kenya and NE Tanzania) on the basis of various topographical descriptors. The procedure is first presented for the month of April, and subsequently month-to-month changes in the role of topography on rainfall distribution are discussed. Predictors included quantitative estimates of slope, mean and standard deviation of elevation, and principal components that describe the topographical environments around the meteorological stations. Different window sizes were considered in order to account for processes of different scales. Linear models were defined for each mean monthly rainfall field, and cross-validated. Inclusion of interpolated residuals yield explained variabilities in the range of 53 to 89%. Results were compared to a simple interpolation of gauge rainfall. The models are particularly useful in areas with few rain gauges. Selected predictors, as well as their scales, change seasonally. North–south exposure contrasts are the main factor of rainfall variation, except in the northern summer. East-facing stations are wetter in the short rains season (October to December), and drier in the monsoon season. These variations coincide with seasonal atmospheric circulation changes above the region. Slopes also contribute to rainfall variability. At the scale of the study mean elevation had little effect; other predictors, especially exposure, actually describe the interaction between rainfall and topography more adequately.

**KEY WORDS:** Rainfall · Geostatistical modelling · Orography · Kenya · Tanzania

*Resale or republication not permitted without written consent of the publisher*

## 1. INTRODUCTION

Regionalized analyses of climate variability and change are of critical importance for the proper assessment of climate impacts. Seasonal rainfall forecasts have been improved considerably in the last decade, but their usefulness is limited by the scale at which they are delivered (i.e. large regions often covering several hundred thousand km<sup>2</sup>). Similarly, the impacts of 21st century climate changes induced by enhanced greenhouse gas concentrations cannot be properly assessed without downscaling, since general circulation models only provide large-scale projections.

For regions with complex terrain, downscaling requires in-depth knowledge of the local relationships between climate and orography. Rainfall is one of the most difficult variables to model, because it is subject

to large spatial variability, and its relation to the relief is complicated. A first modelling step is assessment of these relationships for mean climate fields. Such an analysis, through the definition of empirical models, also makes it possible to produce more accurate regional maps (Funk et al. 2003).

There are 2 methods for deriving rainfall fields from rain gauge data: (1) Precipitation is estimated from elevation data combined with atmospheric data (Basist et al. 1994, Kyriakidis et al. 2001, Daly et al. 2002), or (2) only relief data are used, but with 2 slightly different approaches. The first method uses elevation and other basic descriptors such as slope, exposure, and distance from the coast (Hevesi et al. 1992, Daly et al. 1994). The second approach involves a multidirectional parameterization of the relief by summarizing the topographical patterns, using morphometry and

\*Email: oetlip@u-bourgogne.fr

hypsoetry (Drogue et al. 2002) or principal component analysis (PCA) (Benichou & Le Breton 1987, Wotling et al. 2000), before assessing their relationship with rainfall.

While extratropical areas have been the subject of a number of studies, tropical regions have largely been ignored. In the tropics, it is the rainfall distribution which matters most, e.g. for agroclimatic applications, but its assessment is hampered by the uneven spacing of meteorological stations. This study defines statistical models for mean monthly rainfall estimation over eastern Africa, an equatorial region with complex terrain. The objectives were (1) from an applied standpoint, construction of tentative rainfall maps and (2) from a theoretical standpoint, improvement of our knowledge on the dynamic climatology of the region. In particular, we examined how various topographical factors interact with atmospheric circulation to produce seasonally different rainfall fields. The aim of the present study was to select topographical predictors, and to discuss their physical interpretation and their ability to estimate spatial rainfall distribution. The rainfall maps obtained are preliminary and will require further improvement.

We chose a method that combines the use of altitude and other simple topographical descriptors with more complex variables which synthesize the topographical environment of the meteorological stations. Such empirical methods complement the use of limited-area numerical models, an alternative solution for the reconstruction of small-scale climate variability. Sun et al. (1999a,b) have shown that the NCAR regional numerical model (RegCM2) reproduces the wind fields and rainfall variability for the short rains in East Africa reasonably well. However, these models still include several biases, and they are often too expensive to run in many developing countries.

This study is organized as follows: Section 2 delineates the area selected for study and presents the rainfall and topographical data. Section 3 describes the methodology followed to build our statistical models. Section 4 describes the model and its validation for the month of April. Section 5 presents the results for all months, and Section 6 presents the conclusions.

## 2. STUDY AREA AND DATASET

### 2.1. Study area and general climate

Our choice of study area was based on 3 requirements:

(1) Contrasted topography. In East Africa (Fig. 1), the coastal plain (0 to 200 m) extends from SW to NE along the Indian Ocean. Higher ground occurs in the central

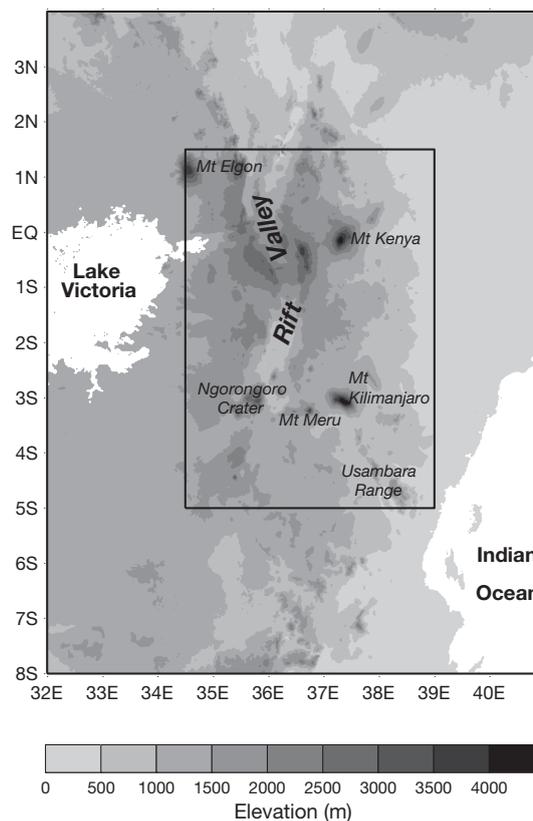


Fig. 1. Relief map of East Africa. The rectangle indicates the study area

and western parts of the region, with altitudes between 1500 and 2500 m and several isolated mountain ranges above 3000 m. In this area, the Rift Valley forms an elongated north–south depression, with minimum altitudes near 500 m.

(2) A rainfall station network which is representative of the various topographical environments, with sufficient density to resolve relatively small-scale features. In East Africa, this condition is fulfilled for central and southern Kenya, Uganda and parts of NE Tanzania. An additional condition is that the records be sufficiently homogenous. This is not fully met over Africa in general, but a number of fairly long time series are available in eastern equatorial Africa (albeit with missing values).

(3) Seasonal and interannual variations of rainfall must exhibit spatial coherence. The existence of a common forcing on regional rainfall may partly guarantee that similar rainfall mechanisms are at work in the various parts of the study area.

Regional climate in East Africa is constrained by its location on the equator and by the influence of the Indian Ocean, as the dominant (often diffluent) easterly flow advects moist air towards the area. The

regional climate is marked by 2 rainy seasons (Hills 1978, Nicholson 1993), locally known as the Long Rains (March to May) and the Short Rains (October to December). Between these seasons, the other months (June to September, January and February) are quite dry. The Long Rains typically concentrate between 35 and 50% of the mean annual rainfall. The month of April, selected for detailed discussion, is in the middle of the Long Rains, and relief is expected to have a dominant influence on rainfall patterns (early or late rainy season months could show stronger meridional gradients related to the location of the InterTropical Convergence Zone (ITCZ)).

The Short Rains exhibit strong interannual co-variations throughout East Africa, and the Long Rains are quite spatially coherent in central Kenya and NE Tanzania (Camberlin & Philippon 2002, Camberlin & Okoola 2003). Since rainfall variations are largely governed by similar mechanisms, we may expect all monthly models to have some skill for the whole region under study.

An additional constraint for the selection of the study area was that we did not want to mix terrain-induced climate features with those associated with local land/sea or land/lake contrasts. The East African coast and the shores of Lake Victoria were excluded from the analysis.

The study region therefore extends from 34.5 to 39° E and 5° S to 1.5° N. It comprises much of the Kenya

Highlands (at an average altitude of 1500 to 2500 m) on both sides of the Rift Valley, including the Mt. Kenya area to the east. To the south, it includes NE Tanzania, where more isolated mountains are found (e.g. Mt. Kilimanjaro, Mt. Meru, Usambara Range).

## 2.2. Rainfall data

The rainfall dataset is from 305 rain gauge stations that are unequally distributed between Kenya and Tanzania. Whereas the network is adequate over Kenya (Fig. 2), it is less so in Tanzania, especially in the plateau area west of Mt. Meru. Resulting biases are discussed in Sections 3.2.3 and 4.1.

Long-term monthly rainfall averages were calculated over the longest span of years available for each station, with most of the time series covering the period from 1950 to 1990. We used additional stations, for which only long-term averages are available, sometimes for an unknown period. In most cases, the average is over at least 30 yr. The homogeneity of the record is clearly not optimal, since the period over which the averages are computed may differ from station to station. It is not possible to use a standard period common to all stations, because the series often include missing values, or begin or end on different years. Keeping only those stations which have long, uninterrupted series over the same period would result in the exclusion of valuable

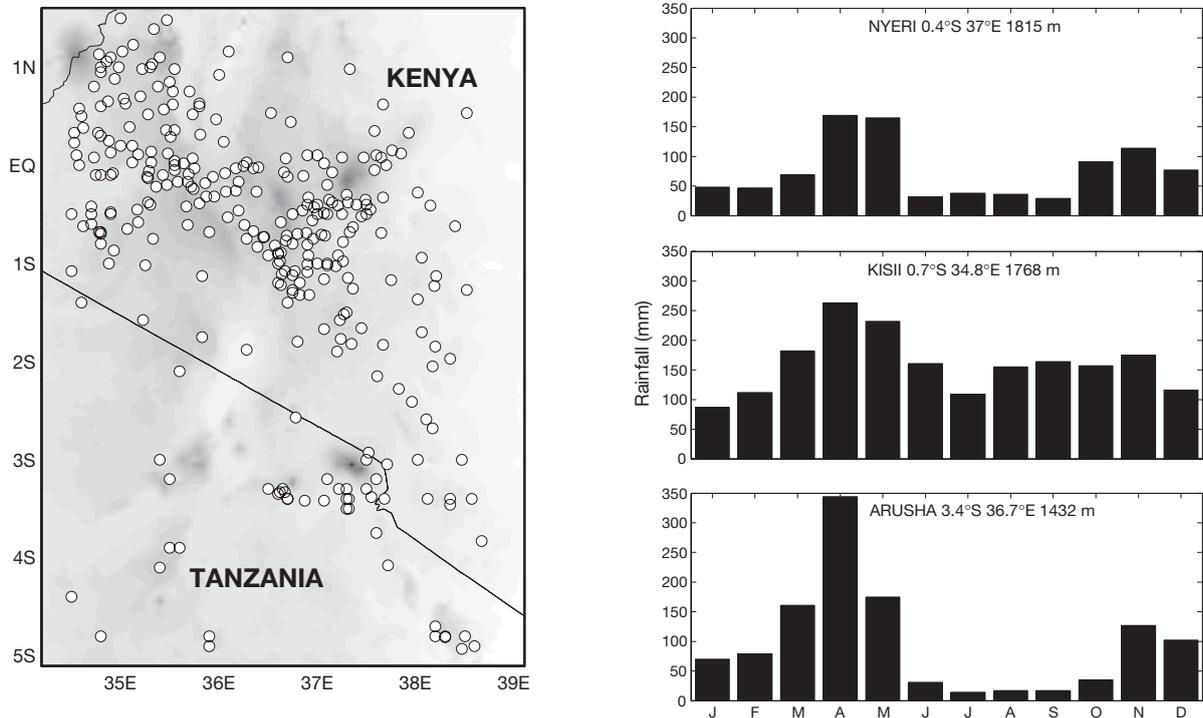


Fig. 2. Station network and political boundaries. Mean monthly rainfall for selected stations is shown in the right-hand panels

information over large poorly sampled areas. The region does not exhibit long-term rainfall trends (e.g. Hulme et al. 2001), so that the use of stations with shorter or different record periods does not seriously affect the outcome of the models.

### 2.3. Topographical data

The U.S. Geological Survey's GTOPO30 Digital Elevation Model (DEM) (Gesch & Larson 1996) with a cell side of 30" (~1 km near the equator) was used to calculate most rainfall predictors: parameterization of the topography by PCA applied to relative elevation around each location (see Section 3.2.1), altitudinal variables (mean absolute elevation and relative variations around stations), and morphometry.

## 3. MODEL DESIGN

### 3.1. Methodology

Our aim was to explain mean monthly rainfall distribution using various topographical descriptors (slope, principal components of topographic variables, geographical position), and to produce terrain-aided interpolated rainfall fields. The procedure consisted of 5 steps:

- (1) Definition of a linear regression model that gives the best estimate of mean monthly rainfall, based on data from all stations.
- (2) Test of the quality of regression using cross-validation.
- (3) Application of the model to all regular grid points, to compute rainfall estimates.
- (4) Calculation of the residuals for those grid points in which rainfall stations are located.
- (5) Kriging of the residuals and adding the results to the predicted values, to improve the rainfall estimates obtained at Step (3).

At Step (1), a multiple regression was defined in a stepwise mode (forward). Independent variables, which have the higher common variance with rainfall (i.e. dependent variable  $y$ ), were selected based on Fisher's  $F$ -statistic, which was first calculated for each independent variable. The predictor with the highest and most significant  $F$ -value ( $\mathbf{X}_1$ ) was added to the model ( $\hat{y} = a_1\mathbf{X}_1 + b_1$ ). Then, the predictor variable which brought the largest increase in the variance of  $y$  explained by the model was added. This was repeated until the variance increase obtained by the inclusion of a new predictor  $\mathbf{X}_k$  was not significant. To avoid collinearity, the quality of the regression and the fate of each of the variables already included in the model

were assessed at each step. The  $F$ -statistic makes it possible to test the contribution of a given variable to explaining the variance of the dependent variable. In our models, the predictors were maintained in the model when  $F$  was significant at the 95 % confidence level. The models included a maximum of 4 predictors.

The resulting multiple regression equation was in the form:

$$\hat{y} = a_1\mathbf{X}_1 + a_2\mathbf{X}_2 + \dots + a_k\mathbf{X}_k + b$$

where  $\hat{y}$  is the estimated value of  $y$ ,  $a_k$  is the slope for the  $k$ th predictor,  $\mathbf{X}_k$  is the  $k$ th independent variable, and  $b$  is the intercept.

The quality of the model (Step 2) was tested by a 'leave-one-out' cross-validation. Using the predictors defined in the last step, a model was fitted to the 304 stations minus the first one, and the rainfall estimate for this station was computed. The procedure was repeated by successively dropping each station, one at a time. Predicted values for the 304 stations were then compared to the observed ones. For that purpose, 5 classical criteria were used: multiple correlation coefficient ( $R$ ), r-square ( $R^2$ ), Root Mean Square Error (RMSE) (Hastenrath et al. 1995), Linear Error in Probability Space (LEPS) (Ward & Folland 1991) and average Skill (SK), an improvement of LEPS (Potts et al. 1996). The residuals of the model were considered as ratios of the mean rainfall, as follows :

$$Rs = Fs/Ps$$

where  $Ps$  is the observed value for Station  $s$ ,  $Fs$  is the estimate, and  $Rs$  is the residual.

A kriging of the residuals (Step 5) was performed by cubic interpolation. Prior to the interpolation, grid point data (estimated residuals) located >40 km from any station were set to zero, to reduce the drifts related to cubic interpolation and to account for the fact that the station residuals have mostly local significance. The resulting interpolated residual at each grid point was then added to the estimate  $Fg$  for that point, as follows:

$$Vg = Fg/(1 - Rg)$$

where  $Rg$  is the interpolated residual for grid point  $g$  and  $Vg$  is the final estimate.

### 3.2. Predictors

#### 3.2.1. Topographical principal components

The environment of each station, described by elevations extracted from the DEM, is depicted by basic patterns built using PCA (Benichou & Le Breton 1987) (Fig. 3). To capture this environment at different scales, several windows around the station were con-

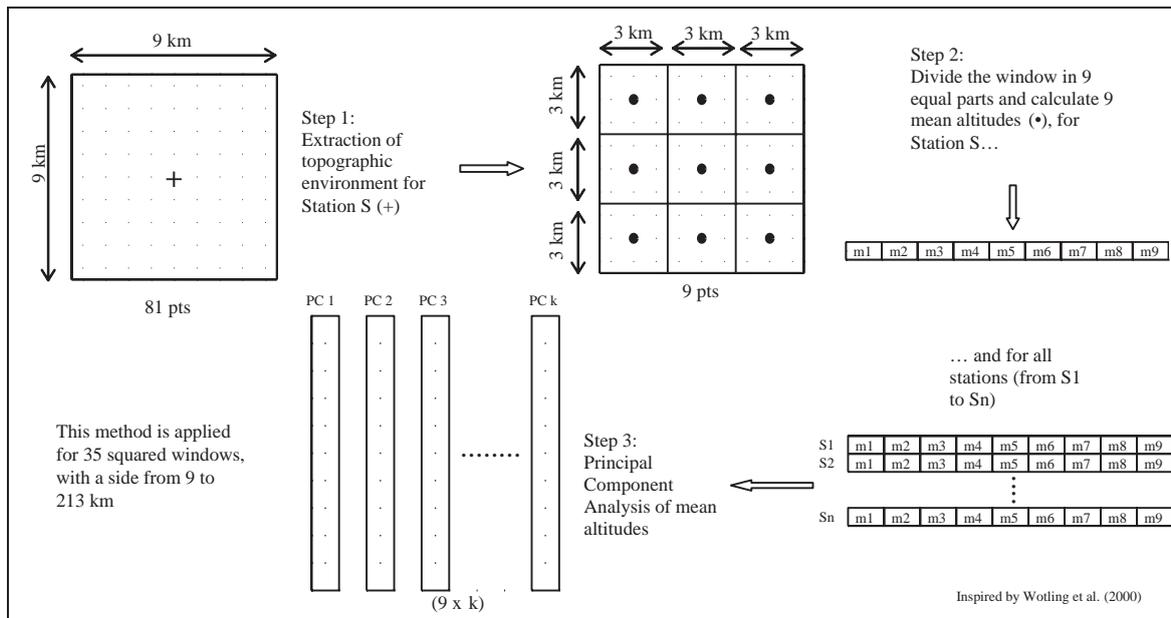


Fig. 3. Procedure followed for computing principal components of local topographical environments

sidered. These windows were square and their size ranged from approximately  $9 \times 9$  km (our example) to  $213 \times 213$  km, for a total of 35 windows. Each window was divided into 9 boxes, and the 9 corresponding mean elevations were subjected to PCA (Fig. 3). One PCA was performed for each window, and 4 principal components (PCs) were retained in each case. The first 4 PCs always summarized most of the variance (from 83.6% for the  $141 \times 141$  km window to 92.5% for the  $9 \times 9$  km window). Each PC represents a simple topographical pattern (see Fig. 4a,c,e,g), i.e. a pattern of *relative* elevations around each location, independent of the absolute elevation. Each PC displays spatial patterns which are almost the same for all windows. For the definition of the models, based on preliminary assessments, PCA results for only 3 windows were retained ( $33 \times 33$ ,  $123 \times 123$  and  $213 \times 213$  km), in order to avoid collinearity in the set of predictors.

PC 1 (Fig. 4a) has negative values (lower altitudes) to the west of the station and positive values (higher altitudes) to the east, i.e. it represents a western exposure. The projection of all grid points on PC 1 (e.g. Fig. 5a for the  $123 \times 123$  km window) discriminates west-facing locations (positive scores) and east-facing locations (negative scores). In the east, the east-facing slopes gradually rise from the Indian Ocean coastal plains to the Central Highlands. In the central part of the region, 2 narrow corridors of positive and negative PC 1 scores denote the Rift Valley. Further west, the terrain slopes down towards Lake Victoria.

PC 2 represents a southern exposure (Fig. 4c). Due to the generally meridional organization of East African relief, the corresponding map (Fig. 5b) shows a pattern that is often more localized than that pattern associated with PC 1. PC 3 (Fig. 4e) represents a col located along a NW/SE valley (positive scores), or a saddle along a mountain ridge oriented in the same direction (negative scores). The first pattern is found for instance between Mt. Kenya and the Aberdare Range ( $0.5^\circ$  S,  $37^\circ$  E; Fig. 5c), the second one on both sides of the Rift Valley (e.g. equator at  $35.5^\circ$  E). Finally, PC 4 (Fig. 4g) corresponds to isolated relief features, either a summit (positive scores) or a depression (negative scores). Mt. Kenya and Mt. Kilimanjaro stand out on the projection map (Fig. 5d), but less conspicuous features do so as well, since all PCs depict relative elevations only.

### 3.2.2. Slopes and other predictors

Absolute variations in relief and relative variations around the central point were also considered as predictors. The average (mean) and the median (med) elevations characterize absolute variations. The standard deviation (std), amplitude (ampl) and skewness and kurtosis (kurto) coefficients depict relative variations. Each predictor was calculated for each window, as described above. Slopes (slop) were also computed for the 35 windows by the maximum downhill slope method (Dunn & Hickey 1998).

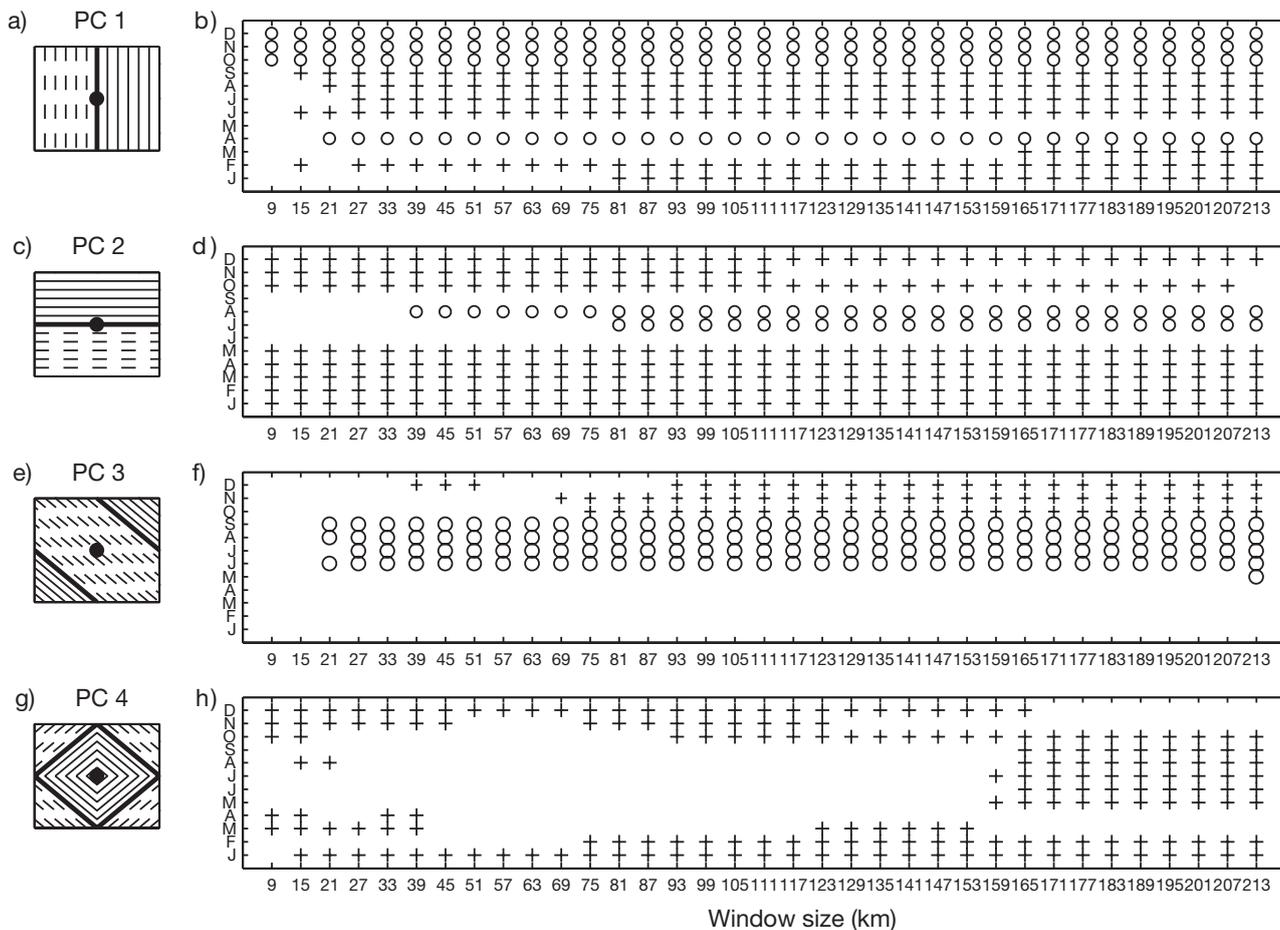


Fig. 4. (a,c,e,g) Simplified topographical patterns described by the first 4 principal components; solid (dashed) lines show higher (lower) elevations than the bold contour line. (b,d,f,h) Field correlations between mean monthly rainfall and the corresponding PC scores for the 305 stations; symbols denote correlations significant at the 95% confidence level (+: positive, O: negative)

In addition, 3 variables describing geographical location were included: latitude (lat), longitude (lon) and distance to Lake Victoria (dst). The latter affects the regional climate by a mesoscale circulation in the form of land/lake breezes. Although the lake itself is excluded from the study area, in the Western Kenya Highlands this circulation interacts with slope winds and the prevailing easterly flow, resulting in enhanced rainfall on the lower and mid slopes (Asnani & Kinuthia 1979, Okeyo 1987, Mukabana & Pielke 1996).

### 3.2.3. Representativeness of the stations

The distribution of the 305 stations was compared to that of the 105 300 grid points of the study area with respect to the topographical predictors. Histograms showing the percentage distribution of stations and grid points were plotted for each predictor, e.g. Fig. 6 for the smallest window and a selection of predictors: mean elevation, standard deviation of elevation (std),

slope and PCs 1 to 3. The distribution of grid points and stations was generally very similar, despite the fact that the sample of stations was small compared to the number of grid points. For mean elevation (Fig. 6a), however, the agreement was less good, because human population distribution (and hence station locations) are skewed towards higher altitudes. Therefore, lowlands are slightly under-represented in the station sample, but most altitudinal classes are described. With the partial exception of mean elevation, we consider that the station network adequately sampled the topographical features of the region.

## 4. MODEL FOR THE MONTH OF APRIL

### 4.1. Description of the model and validation

Using the above methodology and predictors, the final regression model for April mean rainfall (RAIN) was:

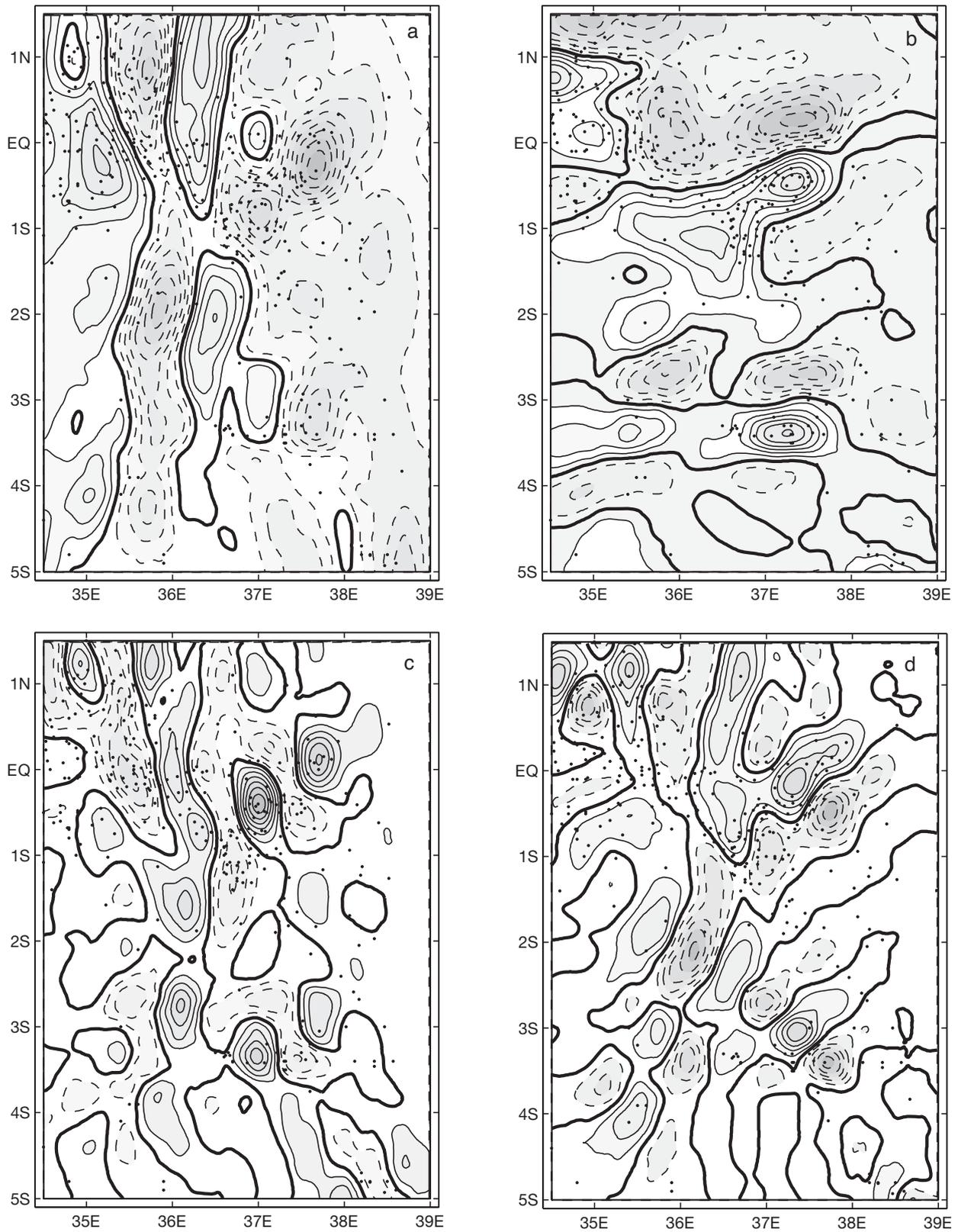


Fig. 5. Projection over the 105 300 grid points, for the  $123 \times 123$  km window size, of (a) PC 1, (b) PC 2, (c) PC 3 and (d) PC 4. Solid lines: positive values; dashed lines: negative values; bold lines: zero

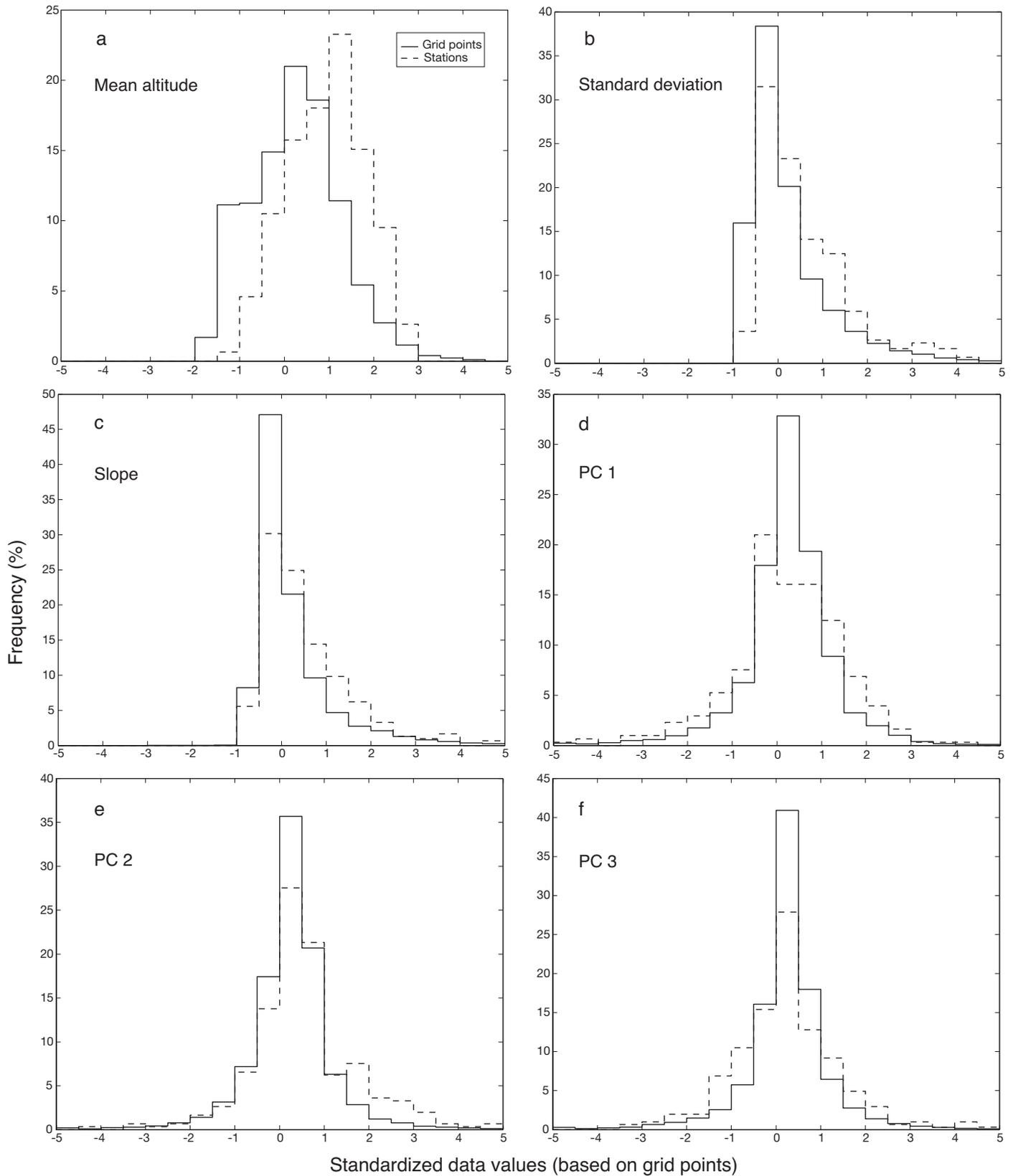


Fig. 6. Percentage distribution of stations (dashed line) and grid points (solid line) with respect to (a) mean altitude, (b) standard deviation, (c) slope, (d) PC 1, (e) PC 2 and (f) PC 3. The plots are for the smallest windows ( $11 \times 11$  km for panels a–c,  $33 \times 33$  km for panels d–f)

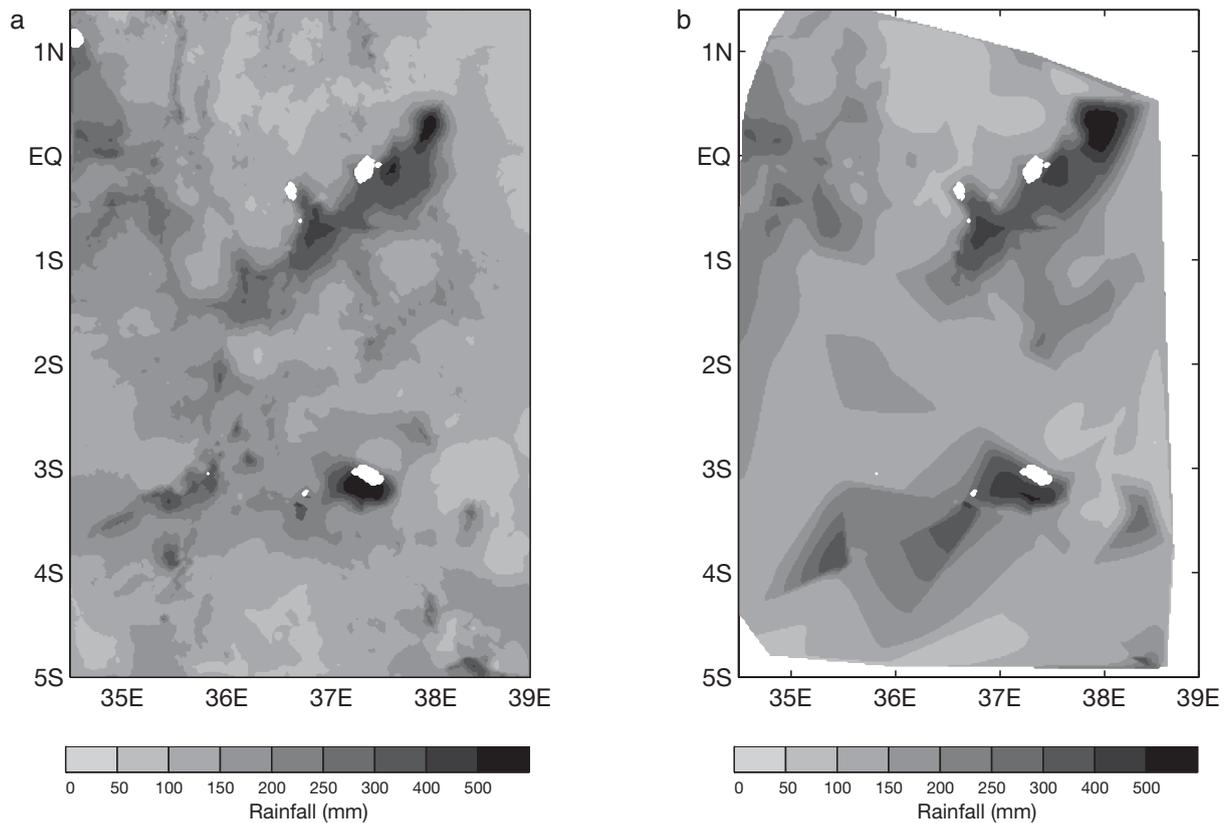


Fig. 7. Estimated mean April rainfall: (a) model and kriged residuals, (b) simple kriging. Blank areas correspond to high altitudes for which no estimate was computed

$$\text{RAIN} = 168 \times \text{PC}_{2\_33} + 0.248 \times \text{std}_{71} + 145 \times \text{PC}_{2\_213} + 9.71 \times \text{slop}_{11} + 71.1 + \varepsilon$$

where  $\text{PC}_{2\_33}$  and  $\text{PC}_{2\_213}$  are PCs computed for the  $33 \times 33$  and  $213 \times 213$  km windows, respectively,  $\text{std}_{71}$  is the standard deviation for the  $71 \times 71$  km window, and  $\text{slop}_{11}$  is the slope for the  $11 \times 11$  km window. The predicted values were mapped, including a kriging of the residuals (Fig. 7a). Areas above 3500 m remain blank, to account for the probable decrease of rainfall at high elevations (as found on Mt. Kenya and Mt. Kilimanjaro: Thompson 1966, Coutts 1969).

The robustness of the model was assessed by cross-validation. Observed data and RAIN model outputs are significantly correlated ( $R = 0.65$ ;  $p = 0.05$ ; see Table 1). Simple kriging gives a higher correlation. After the addition of kriged residuals, the correlation increased slightly. RMSE, LEPS and SK scores for the model plus kriged residuals and for simple kriging are similar. Positive LEPS and SK scores (approaching +100%) denote a good (perfect) model (Ward & Folland 1991). Note that the model skill scores, before interpolation of the residuals, are already fairly high (>40%).

It is deceptive that the results obtained for the model plus kriged residuals are marginally better than those obtained for simple kriging. However, Fig. 7 shows that the model estimates clearly outperform those based on simple kriging for areas with low station density (Fig. 2) and contrasted topography, i.e. in the southern and SW parts of the region. The fact that high skills are obtained with simple kriging is due to the

Table 1. Leave-one-out cross-validations of estimates of April rainfall in East Africa. Comparisons between observed and predicted values (RAIN), between observed and predicted values (RAIN + residuals kriging), and between observed and kriged values. Criteria: multiple correlation coefficient (R); r-square ( $R^2$ ); Root Mean Square Error (RMSE); Linear Error in Probability Space (LEPS); and average skill (SK)

	R	$R^2$ (%)	RMSE (mm)	LEPS (%)	SK (%)
Predicted (RAIN)	0.65	43	71	44	41
Predicted (RAIN + residuals kriging)	0.81	65	57	67	67
Kriged	0.78	60	60	66	66

uneven station density. Some areas with a dense station network (Central and Western Kenya Highlands, southern slopes of Mt. Meru and Mt. Kilimanjaro) belong to the wettest areas; here, simple interpolation is adequate for estimating rainfall amounts, and taking topography into account adds little. Thus, the basic contrasts between the wet and the dry zones are reasonably well reproduced by kriging. However, kriging fails to document small- and medium-scale variations of rainfall over hilly but unsampled areas, e.g. many Tanzanian mountain ranges, or the Aberdares (0.3° S, 36.6° E) and Cherangani Hills (1.2° N, 35.4° E), where the use of the statistical model is a decisive improvement. The resulting modelled rainfall distribution agrees with rainfall maps of East Africa (e.g. Tomsett 1969), with additional detail over hitherto poorly sampled areas.

As the results could be biased by the higher station density in Kenya, we further investigated the reliability of rainfall estimates in the Tanzanian sector. Mean April values for 46 additional Tanzanian stations, which were initially not available for inclusion in the model, were obtained from the FAO Climwat database. The observed values are strongly correlated (0.82) to the estimates. The RMSE (74 mm) is not much higher than that obtained for the 305 stations used in the model (57 mm). When restricting the dataset to the western stations of Tanzania, where very few stations were available in our initial sample, the correlation is slightly lower (0.73 for 16 stations) but still highly significant (99.9% confidence level), with a RMSE at 62 mm. These results show that the model is still useful for regions away from the main cluster of rainfall stations. Confirmation is obtained by comparisons between the April rainfall map, and vegetation and land cover maps. April rainfall is only a fraction of annual rainfall, but there is a good correspondence between the extent of evergreen and montane forest and the areas receiving >250 mm rainfall (Fig. 7a), e.g. in northern Tanzania, where the isolated forest areas of Mt. Hanang, Mt. Meru, and Mt. Kitumbeine (especially their southern slopes) stand out as a series of wetter patches on the estimated April rainfall map. Based on more detailed vegetation maps, the dissymmetry found in the Ngorongoro crater range (3° S, 36° E) between the forested south and SE slopes and the unforested northern slopes (and summits) is also replicated by April rainfall as simulated by the model. The map based on simple kriging fails to show these isolated wet pockets (Fig. 7b).

#### 4.2. Interpretation of the model

Among the 4 predictors included in the model, 2 are topographical patterns obtained from the PCA (PC 2,

twice), one is a topographical variable (slope) and the last one a summary statistic (standard deviation). North–south exposure contrasts (PC 2; Fig. 4c) are the main predictor, controlling rainfall at 2 different scales: small (33 km windows) and large (213 km). PC 2 has high loadings to the north and low loadings to the south. Thus, locations displaying positive (negative) scores for this PC are exposed to the south (north). The positive regression coefficients for this predictor at both scales imply that more April rainfall is linked with southern exposures. This result will be discussed in the next section.

The third predictor, *std\_71*, has a positive coefficient in the regression equation. Larger rainfall amounts are found for high *std\_71* values, i.e. when the topography in a 71 km window around a station varies greatly.

The last predictor, *slop\_11*, represents terrain organization at small scales (11 km windows). Local rainfall is higher when the maximum slope around the station is steep.

## 5. RESULTS FOR THE OTHER MONTHS

Results for the entire year are summarized in Table 2. Predictors, sign of the regression coefficients and cross-validation statistics are given for the comparison between observed and predicted (RAIN + residuals kriging) values. The fact that some predictors are selected for consecutive months (e.g. PC 1\_213 and *slop\_41* between June and September, PC 2\_33, PC 2\_213 and *std* in April and May) increases confidence in the models' veracity and suggests a consistent physical interpretation. However, there is less stability in the dry period from December to March.

### 5.1. Predictors

Fig. 4 shows the correlations between monthly rainfall (at all 305 stations) and the scores of the first 4 topographical PCs for the 35 windows described above (from 9 to 213 km). PC 1 (Fig. 4a) depicts western (positive scores) and eastern (negative scores) exposures. The month-to-month changes in the correlation between mean rainfall and this topographical pattern are shown in Fig. 4b, which shows coefficients that are significant at the 95% level (positive: +; negative: O). A positive (negative) correlation means that monthly rainfall is enhanced for western (eastern) exposures.

There are marked seasonal differences, replicating the distribution of wet and dry seasons. The dry seasons (boreal winter and summer) have mostly positive correlations (western orientations are wetter), whereas the middle of the Long Rains (April) and the Short Rains present negative correlations (eastern orienta-

Table 2. Predictors, signs of the regression coefficients (+: positive, O: negative) and cross-validation statistics for monthly models of rainfall in East Africa. See Section 3.2 'Predictors' and Table 1 for definition of abbreviations

	PC	Absolute elevation	Relative variation	Slope	Geographical location	R <sup>2</sup> (%)	LEPS (%)	SK (%)
Jan	PC 4_123 (+) PC 2_213 (+)				lat (O) lon (O)	60	59	58
Feb	PC 2_213 (+)			slop_41 (+)	lon (O) lat (O)	72	65	65
Mar	PC 2_213 (+) PC 1_213 (+)	med_71 (O)	ampl_11 (+)			61	61	61
Apr	PC 2_33 (+) PC 2_213 (+)		std_71 (+)	slop_11 (+)		65	67	67
May	PC 2_33 (+) PC 2_213 (+)		std_41 (+)		dst (O)	67	68	67
Jun	PC 1_213 (+)		std_71 (+)	slop_41 (+)	lon (O)	83	79	78
Jul	PC 1_213 (+)			slop_41 (+)	lon (O) lat (+)	86	80	80
Aug	PC 1_213 (+)			slop_41 (+)	lon (O) lat (+)	89	82	82
Sep	PC 1_213 (+)			slop_41 (+)	dst (O) lat (+)	88	81	80
Oct	PC 2_213 (+) PC 3_213 (+)		std_71 (+) kurto_41 (O)			71	65	65
Nov	PC 1_213 (O) PC 3_213 (+) PC 2_213 (+)		kurto_41 (O)			70	65	65
Dec	PC 3_213 (+)	mean_71 (O)		slop_11 (+)	lat (O)	53	60	60

tions are wetter). The relationships are the same for all windows, except for the smaller ones (from 9 to 21 km), for which the signal is lost.

These seasonal changes are explained by the marked shifts in the lower- to mid-tropospheric wind direction during the year, exemplified by the 700 hPa mean monthly vector wind maps (Fig. 8). Twice a year, the ITCZ passes above the region, between March and May (Long Rains) and between October and December (Short Rains). At this time, easterlies prevail and orographic rainfall occurs on east-facing slopes. Note that the 2 rainy seasons are not symmetrical: the rainfall enhancement along eastern slopes is more evident in the Short Rains than in the Long Rains. For the latter, it is only found in April, and it is a weak feature: PC 1 does not serve as a predictor in the model for this month (Table 2). In November, by contrast, PC 1 is a key predictor.

During the rest of the year, the meridional component of the wind is more accentuated (northerlies in the boreal winter, southerlies in summer, Fig. 8). Flohn (1965) and Anyamba & Kiangi (1985) have shown that the corresponding monsoon flows are strongly divergent over East Africa, especially in the low-level plains

to the east. However, in the Western Kenya Highlands and around Lake Victoria, mesoscale wind systems develop (Asnani 1993, Mukabana & Pielke 1996). These features explain why western exposures have relatively more rainfall at these times.

PC 2 describes southern exposure (Fig. 4c). From the beginning of the Short Rains (October) to the end of Long Rains (May), correlations are consistently positive (Fig. 4d), i.e. southward-oriented stations have more rainfall. In contrast, summer (July to August) exhibits negative correlations, but these attain the 95% significance level only with large windows. PC 2 is a major predictor, appearing in all models from October to May, except for December (Table 2). During the Long Rains (April and May), southern exposures stand out as distinct predictors associated with 2 different window sizes (33 and 213 km).

During most of the year, southern orientations are wetter, suggesting that northerly winds are drier than southerlies, although in October to December this is valid mostly at local scales (see below under PC 3). The pattern is reversed in the northern summer. In addition to the differences in the moisture content of the air mass, this seasonal change may also reflect the shift in

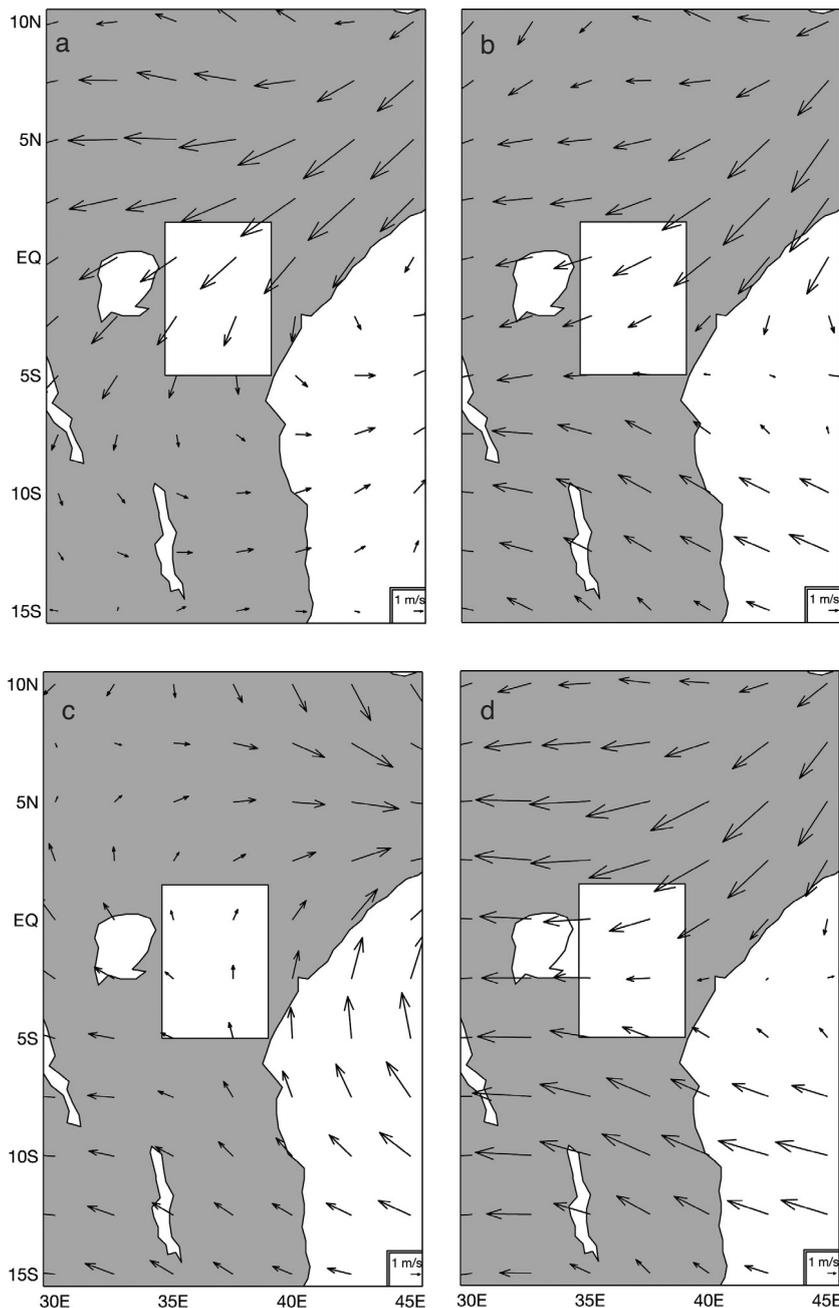


Fig. 8. Mean monthly vector winds at 700 hPa (NCEP/NCAR reanalysis data averaged over 1950 to 1990) for (a) January, (b) April, (c) July and (d) November

the latitudinal location of the ITCZ, which is further south in the boreal winter (hence wetter conditions in the south), and further north in the boreal summer (wetter conditions in the north).

The topographical features depicted by PC 3 correspond to a valley or a mountain ridge (Fig. 4e). June to September rainfall is negatively correlated with this pattern, though only on medium and large scales (Fig. 4f). This represents the fact that ridges that are

parallel to the SE flow tend to get wetter (e.g. Western Highlands, and parts of the Eastern Highlands and Usambara Range). However, it is also an artefact, as this topographical pattern is better defined over the wet high ground areas of the northwest. Therefore, this predictor does not add variance to the models for these months, and is not retained in the end (Table 2).

During the Short Rains (October to December), rainfall is positively correlated with the same topographical pattern, though only for intermediate to large windows. This represents the fact that stations located at the foot of escarpments that face NE tend to get wet, as NE winds are moister in the Short Rains than in the Long Rains. This applies to Mt. Kilimanjaro (Coutts 1969). As in June to September, the influence of this topographical pattern is only effective at medium to large scales (213 km windows in the October and November models, 123 km in the December model).

The last topographical pattern (PC 4) corresponds to isolated relief features, either depressions or summits (Fig. 4g). PC 4 and rainfall display only positive correlations, which means that an isolated high ground location (depressed area) is generally conducive to higher (lower) rainfall. However, the correlations are not always stable from one window to the next (Fig. 4h), and this pattern is a relatively minor predictor, being selected only in January (Table 2). For small windows, we find significant positive correlations for the beginning and the end of year (including the Long Rains and the Short Rains). In the summer months, there are significant correlations for very large windows only.

The seasonal opposition of PC 4 on window size can be explained as follows. In the wet seasons, a local relief feature is likely to enhance convection, since atmospheric conditions and circulation are already propitious to rainfall. In the dry seasons, it requires larger scale high ground areas to substantially alter the atmospheric circulation.

Among the other predictors included in the monthly regression models (Table 2), the mean slope and altitudinal standard deviation are the most frequently

selected. They indicate increasing rainfall with increasing hilliness. Trend surfaces (combining latitude and longitude) additionally account for a significant portion of rainfall distribution in the dry seasons (January to February and June to September). The eastern part of the study region is noticeably drier at these times, as a result of the monsoon flow divergence close to the Indian Ocean shores (Fig. 8). Note that absolute elevation itself is seldom a predictor of rainfall distribution at the scale under investigation. It is even negatively correlated with rainfall in the March and December models. In these 2 cases, elevation appears as the 4th predictor in the models, and can actually be considered as a weighing factor moderating the effect of other predictors with respect to elevation.

### 5.2. Performance of the models

Based on a maximum of 4 predictors, the predictive models explain from 53 to 89% of the rainfall variance (Table 2) in cross-validation mode. The highest  $R^2$  is for the northern summer. This is not unexpected, since rainfall distribution at this time is very contrasted, with virtually no rainfall over the eastern and southern areas, and relatively wet conditions over the NW corner. For the other months,  $R^2$  is still fairly high (60 to 70%) except for December. LEPS and SK scores show similar features. The lowest scores are found in December and January, the highest in June to September. As in April, these scores are of limited usefulness given the distribution of the station network.

## 6. CONCLUSIONS

Linear prediction models were defined for estimating rainfall at ungauged locations and for identifying the topographical features that account for the distribution of rainfall over southern Kenya and NE Tanzania. Topography was described by quantitative estimates of slope, mean and standard deviation of elevation, and other variables. Four PCs depict the basic features of the topographical environments (east–west or north–south exposures, valley/ridge patterns etc.) around the stations. All these descriptors are based on windows of different sizes (from about 10 to 200 km), in order to account for processes at different scales.

Using 4 predictors at most, between 53 and 89% of the spatial variance of the mean monthly rainfall field is explained by the models in cross-validation mode. The predictors which are selected cover the whole array of scales, and sometimes (in April for instance) a given predictor may enter twice in the same model, depicting topographical features at 2 different spatial

scales. This supports the use of nested windows for the description of topographical patterns.

There are significant month-to-month variations in the predictors retained in the models. These variations are closely associated with the seasonal atmospheric circulation changes above the region. North–south exposure contrasts are the dominant factor of rainfall variation throughout the year, except in the northern summer (June to September). South-facing stations are wetter, especially during the Long Rains (March to May) when the corresponding predictor appears twice in the regression models, i.e. for 2 different window sizes. This is because southerly winds are slightly wetter than those with a northerly component. East-facing stations only tend to be wetter in the Short Rains (October to December), and to some extent in April. In the dry seasons, especially between June and September, monsoon flows tend to be strongly divergent along the East African coast, and west-facing locations have more rainfall than east-facing locations, which are fairly dry. Other topographical patterns had less importance.

Slope magnitudes, as well as the general ‘hilliness’ of the topographical environment of the station, had a positive effect on rainfall amounts. They often contributed significantly to the variance explained by the models. In contrast, there was little relationship between rainfall and mean elevation. At first sight this is a surprising result, given that highlands are known to be wetter than lowlands, East Africa being no exception. For a smaller area in central Kenya, Basist et al. (1994) retained elevation as the only explanatory variable for annual rainfall. However, there is considerable inconsistency in the elevation–rainfall relationship, which only applies to either very small or very large space scales. In addition, other predictors more adequately describe the interaction between rainfall and topography, and elevation was therefore rarely selected as a significant predictor. For April, the correlation between rainfall and elevation (0.04, for all the 305 stations) was not significant at the 95% confidence level. Significant correlations occurred locally. For example, in a 10 000 km<sup>2</sup> region comprising 15 stations SE of Mt. Kenya, the correlation coefficient was 0.75.

The models used here may not be the best for the area under investigation. They have certain limitations, as they are linear and do not take into account possible thresholds in the relationship between rainfall and topography. In particular, a mask is applied to rainfall estimates for higher mountain areas, but the decrease in rainfall with height above a given elevation is not explicitly accounted for by the models. An integration of the interactions between terrain and low-level winds might also be useful, and is the scope of further investigation.

*Acknowledgements.* The authors thank 3 anonymous reviewers for very useful remarks on key issues of the paper.

#### LITERATURE CITED

- Anyamba EK, Kiangi PMR (1985) Mean motion field in East Africa at the level of the East African low level jet core. *Arch Meteorol Geophys Bioklimatol Ser B* 36:29–41
- Asnani GC (1993) Tropical meteorology. Indian Institute of Tropical Meteorology, Poona
- Asnani GC, Kinuthia JH (1979) Diurnal variation of precipitation in East Africa. Res Rep no. 8/79, Kenya Meteorological Department, Nairobi
- Basist A, Bell GD, Meentemeyer V (1994) Statistical relationships between topography and precipitations patterns. *J Clim* 7:1305–1315
- Benichou P, Le Breton O (1987) Prise en compte de la topographie pour la cartographie des champs pluviométriques statistiques. *Météorologie* 7(19):23–34
- Camberlin P, Okoola RE (2003) The onset and cessation of the 'Long Rains' in eastern Africa and their interannual variability. *Theor Appl Climatol* 75:43–54
- Camberlin P, Philippon N (2002) The East African March–May season: associated atmospheric dynamics and predictability over the 1967–68 period. *J Clim* 15:1002–1019
- Coutts HH (1969) Rainfall of the Kilimanjaro area. *Weather* 24:66–69
- Daly C, Neilson RP, Phillips DL (1994) A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *J Appl Meteorol* 33:140–158
- Daly C, Gibson WP, Taylor GH, Johnson GL, Pasteris P (2002) A knowledge-based approach to the statistical mapping of climate. *Clim Res* 22:99–113
- Drogue G, Humbert J, Deraisme J, Mahr N, Freslon N (2002) A statistical-topographic model using an omnidirectional parameterization of the relief for mapping orographic rainfall. *Int J Climatol* 22:599–613
- Dunn M, Hickey R (1998) The effect of slope algorithms on slope estimates within a GIS. *Cartography* 27(1):9–15
- Flohn H (1965) Contributions to a synoptic climatology of North-East Africa. *WMO Tech Notes* 69:236–244
- Funk C, Michaelsen J, Verdin J, Artan G, Husak G, Senay G, Gadain H, Magadzire T (2003) The collaborative historical African rainfall model: description and evaluation. *Int J Climatol* 23:47–66
- Gesch DB, Larson KS (1996) Techniques for development of global 1-kilometer digital elevation models. *Proc Am Soc Photogrammetry Remote Sensing*, Sioux Falls, South Dakota, August 20–22, 1996; In: Pecora Thirteen, Human Interactions with the Environment—Perspectives from Space; available at <http://edcdaac.usgs.gov/gtopo30/papers/geschtxt.asp>
- Hastenrath S, Greischar L, Van Heerden J (1995) Prediction of the summer rainfall over South Africa. *J Clim* 8:1511–1518
- Hevesi JA, Istok JD, Flint AL (1992) Precipitation estimation in mountainous terrain using multivariate geostatistics. Part I: structural analysis. *J Appl Meteorol* 31:661–676
- Hills RC (1978) The organization of rainfall in East Africa. *J Tropical Geogr* 42:40–50
- Hulme M, Doherty R, Ngara T, New M, Lister D (2001) African climate change: 1900–2100. *Clim Res* 17:145–168
- Kyriakidis PC, Kim J, Miller NL (2001) Geostatistical mapping of precipitation from rain gauge data using atmospheric and terrain characteristics. *J Appl Meteorol* 40:1855–1877
- Mukabana JR, Pielke RA (1996) Investigating the influence of synoptic-scale monsoonal winds and mesoscale circulations on diurnal weather patterns over Kenya using a mesoscale numerical model. *Mon Weather Rev* 124:224–243
- Nicholson SE (1993) An overview of African rainfall fluctuations of the last decade. *J Clim* 6:1463–1466
- Okeyo AE (1987) The influence of Lake Victoria on the convective activities over the Kenya highlands. *J Meteorol Soc Jpn* 65:689–695
- Potts JM, Folland CK, Jolliffe IT, Sexton D (1996) Revised 'LEPS' scores for assessing climate model simulations and long-range forecasts. *J Clim* 9:34–53
- Sun L, Semazzi FHM, Giorgi F, Ogallo L (1999a) Application of the NCAR regional climate model to eastern Africa. 1. Simulation of the Short Rains of 1988. *J Geophys Res* 104(D6):6529–6548
- Sun L, Semazzi FHM, Giorgi F, Ogallo L (1999b) Application of the NCAR regional climate model to eastern Africa. 2. Simulation of interannual variability of Short Rains. *J Geophys Res* 104(D6):6549–6562
- Thompson BW (1966) The mean annual rainfall of Mount Kenya. *Weather* 21:48–49
- Tomsett JE (1969) Average monthly and annual rainfall maps of East Africa. East African Meteorological Department, Tech Memo no. 14
- Ward MN, Folland CK (1991) Prediction of seasonal rainfall in the north Nordeste of Brazil using eigenvectors of sea-surface temperature. *Int J Climatol* 11:711–743
- Wotling G, Bouvier C, Danloux J, Fritsch JM (2000) Regionalization of extreme precipitation distribution using the principal components of the topographical environment. *J Hydrol* 233:86–101

*Editorial responsibility:* Otto Kinne, Oldendorf/Luhe, Germany

*Submitted:* August 9, 2004; *Accepted:* March 30, 2005  
*Proofs received from author(s):* May 13, 2005