Journal of Hydrology 364 (2009) 285–297

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Describing long-term trends in precipitation using generalized additive models

Fiona M. Underwood *

Quantitative Biology and Applied Statistics, School of Biological Sciences, University of Reading, Reading RG6 6BX, UK

ARTICLE INFO

Article history:
Received 6 June 2008
Received in revised form 24 August 2008
Accepted 5 November 2008

Keywords:
Generalized additive models (GAMs)
Generalized linear models (GLMs)
Daily rainfall records
Long-term trend
Non-linear effects
Climate change

SUMMARY

With the current concern over climate change, descriptions of how rainfall patterns are changing over time can be useful. Observations of daily rainfall data over the last few decades provide information on these trends. Generalized linear models are typically used to model patterns in the occurrence and intensity of rainfall. These models describe rainfall patterns for an average year but are more limited when describing long-term trends, particularly when these are potentially non-linear. Generalized additive models (GAMs) provide a framework for modelling non-linear relationships by fitting smooth functions to the data. This paper describes how GAMs can extend the flexibility of models to describe seasonal patterns and long-term trends in the occurrence and intensity of daily rainfall using data from Mauritius from 1962 to 2001. Smoothed estimates from the models provide useful graphical descriptions of changing rainfall patterns over the last 40 years at this location. GAMs are particularly helpful when exploring non-linear relationships in the data. Care is needed to ensure the choice of smooth functions is appropriate for the data and modelling objectives.

© 2008 Elsevier B.V. All rights reserved.

Introduction

Coe and Stern (1982) demonstrated how generalized linear models (GLMs) could be used to model daily rainfall data, by separately modelling how the occurrence and intensity of rain vary through the year and treating years as replicate observations. This approach provides descriptions of rainfall through the year, enables simulation of daily rainfall for use in hydrological models (Wilks and Wilby, 1999), and can lead to the calculation of the risks of particular events, for example a long dry spell, using either simulated weather or recurrence relations (Stern and Coe, 1982).

What is needed in the context of climate change is the ability to describe long-term trends in the seasonal pattern of rainfall, rather than the average pattern. One approach is to model possible causes of change in rainfall patterns by, for example, including covariates measuring large-scale climatic processes such as the Northern Atlantic Oscillation (NAO) (Chandler and Wheatear, 2002) or the Southern Oscillation Index (SOI) (Furrer and Katz, 2007) in the GLMs. This is useful, but it does not provide a simple description of *how* the rainfall pattern is changing through time.

An alternative is to consider functions of time as candidate explanatory variables and examine which best describes the long-term trend in rainfall patterns. Chandler and Wheatear (2002) considered a number of such candidate variables including, a linear trend through time, a cyclical function of time and a linear effect of time after time t_0 only. A limitation to this approach is that the real long-term trend may be some other, possibly non-linear, function, that is either difficult to pre-specify and/or not conceived of by the modeller.

Generalized additive models (GAMs) (Hastie and Tibshirani, 1990) are GLMs in which some of the terms in the model are smooth, non-linear functions of explanatory variables. Rather than pre-specifying the form of the function, the data are permitted to influence the shape of the smooth functions. This provides a flexible modelling framework in which to explore the relationship between response and explanatory variables.

Algorithms for fitting GAMs are available in off-the-shelf statistical software packages, such as R (R Development Core Team, 2008). They have been extensively used in the environmental and ecological literature for many years, to explore and describe, for example, how the number of individuals in a population changes in time and space (Guisan et al., 2002; Augustin et al., 1998). More recently, GAMs have been applied to hydrological and climatic time series. Morton and Henderson (2008) used GAMs to estimate non-linear trends in water quality. Mestre and Hallegatte (in press) model how SOI and sea surface temperatures affect the annual number of cyclone events in a non-linear manner.

There is little evidence of the application of GAMs for modelling daily rainfall data in the scientific literature. Chandler (2005) indicates that GAMs may be useful, Beckmann and Buishand (2002) used GAMs to model daily rainfall as part of a statistical downscaling exercise, where smooth functions of a number of physical covariates were included and Hyndman and Grunwald (2000) used GAMs to model daily rainfall data from Melbourne. The emphasis of Hyndman and Grunwald was specifically on combining two

* Tel.: +44 (0)118 378 7927.
E-mail address: f.m.underwood@reading.ac.uk

0022-1694/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2008.11.003

models, occurrence and intensity, into a single model for rainfall amounts to use as a stochastic weather generator, and to include a smooth term for the SOI in addition to a smooth term for long-term trend.

In this paper the use of GAMs for modelling daily rainfall data is examined. Specifically, it demonstrates how GAMs can be used to effectively investigate (1) how rainfall varies through the year; (2) evidence for overall long-term trend in rainfall and (3) evidence for long-term trend in the pattern of rainfall through the year. The focus is on describing broad large-scale smooth patterns rather than trying to identify and model differences between individual years. The aim is to describe *how* rainfall patterns are changing rather than *why* they are changing, so attribution of long-term trends to large-scale climatic processes is not pursued.

The GAMs build on the GLM framework for modelling daily rainfall in which two separate models are fitted, one for the probability of rain and the second the intensity of rain, defined as the amount of rain on a wet day. The GLMs are described in detail in a number of papers (Stern and Coe, 1984; Chandler and Wheatear, 2002). Briefly, for a series of $t = 1,\ldots,T$ days the amount of rain on day t is recorded as x_t. If J_t is an indicator variable such that $J_t = 1$ if day t is wet, i.e. $x_t > a$ for a pre-specified threshold $a \geqslant 0$, and $J_t = 0$ otherwise then J_t has a Bernouilli distribution $J_t \sim Bern(\theta_t)$ where $\theta_t = Pr(J_t = 1)$ is the probability of the occurrence of rain. The distribution of intensity, the amount of rain on a rainy day $y_t = x_t - a$ when $J_t = 1$, is generally found to be highly skewed and modelled using a Gamma distribution so that $Y_t \sim Gamma(\mu_t,\kappa)$ where the expected value, $E(Y_t)$, is μ_t and its variance, $var(Y_t)$, is $(\mu_t\kappa)^2$.

The systematic components of the GLMs, $\eta_t = logit(\theta_t) = log(\theta_t/(1-\theta_t))$ and $\zeta_t = log(\mu_t)$, are modelled as linear functions of a number of covariates that account for variation in the occurrence and intensity of rain over time. Possible covariates include: the time of year; occurrence of rain on previous days; the amount of rain on previous days; and external variables for large-scale climate indicators such as the NAO or SOI.

To illustrate the use of GAMs, daily rainfall data from 1st October 1962 to 30th September 2001 from a sugar estate in Mauritius are modelled. For the 39 years of Mauritius data, rainfall amounts x_t were recorded to the nearest millimetre. To avoid issues of trace amounts, a wet day was defined as one in which more than 1 mm of rain were observed, $x_t > 1$. Annual summaries of the data, where a year runs from 1st October of that year to 30th September the following year, are shown in Figure 1. There is an overall increase in the number of rainy days each year (a), the total rainfall per sea-

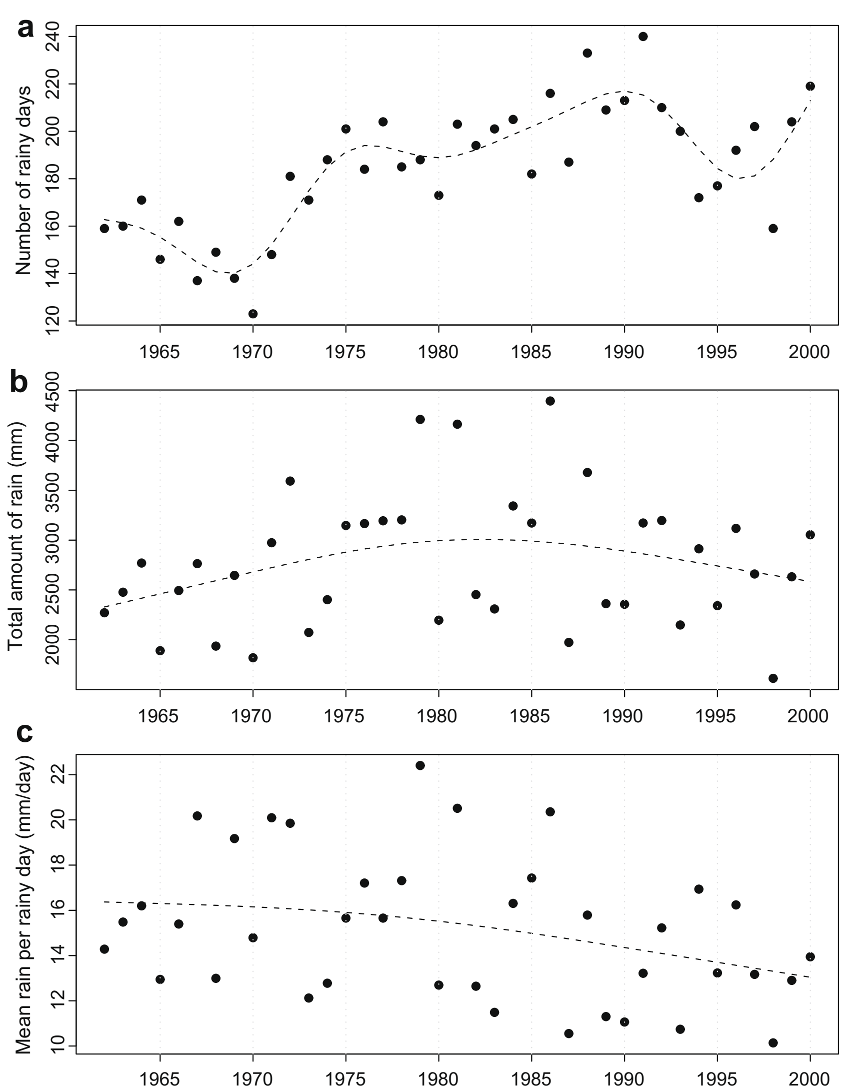

Figure 1 Summary statistics for each season: (a) number of rainy days; (b) total amount of rain and (c) mean intensity of rain per rainy day. Dashed line shows a smooth of the data.

son stays relatively constant (b) and there is some evidence of a decrease in the mean amount of rain per rainy day (c). Senapathi et al. (submitted) calculated these summary statistics for two seasons of the year (December–May and June–November) and showed that some of the variability between years can be explained by large-scale climatic processes such as the Indian Monsoon and the West Indian Sea Surface Temperatures. There is also evidence of an unexplained increase in the number of rainy days in both seasons. Their work did not investigate within-year variability in any detail. This paper demonstrates that the ability to model long-term trend in the seasonal pattern can provide further insights into changing rainfall patterns.

Generalized additive models

The theory of GAMs is complex, although in practice fitting GAMs can be relatively easy using readily available statistical software. Several algorithms for fitting GAMs have been described (Wabha, 1990; Hastie and Tibshirani, 1990; Gu, 2002; Wood, 2006). Here, the method of penalized maximum likelihood, first developed by Wood (2000), is used because of its computational efficiency and model selection and inferential capabilities (Wood and Augustin, 2002). The most recent development of this method is available in the statistical software environment *R* (R Development Core Team, 2008) within the *mgcv* package (Wood, 2008). Wood (2006) provides a detailed introduction, explanation and description of this method. Although some of the methods described below may at first appear quite complex, they are all easily executed as options in the *mgcv* software.

The essential difference between a GAM and a GLM is in the terms in the systematic component, say η_t. In a GLM the linear effect of an explanatory variable x_t is modelled so that

$$\eta_t = \beta_0 + \beta_1 x_t \tag{1}$$

where β_0 and β_1 are parameters to be estimated representing the intercept and slope, respectively. In a GAM, the linear component $\beta_1 x_t$ in Eq. (1) is replaced by a smooth function $f(x_t)$ so that

$$\eta_t = \beta_0 + f(x_t)$$

There are several ways of defining this smooth function $f(x_t)$ and Hastie and Tibshirani (1990) give a useful overview of different types of smoothers. Wood's methods are based on penalized regression smoothers. A regression smoother $f(x_t)$ is a smooth function of the form $f(x_t) = \sum_{m=1}^{M} b_m(x_t)\beta_m$, where the $b_m(x_t)$ for $m = 1,\ldots,M$, are a set of *basis* functions of known form chosen by the user, and the β_m are parameters to be estimated. The choice of basis functions depends on the nature of the covariates and the context. This is discussed in more detail in Section 'Cubic regression splines'. A simple example of a basis is the polynomial basis $b_m(x_t) = x_t^m$. For example, with $M = 3$ the systematic component is

$$\eta_t = \beta_0 + f(x_t, M = 3) = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \beta_3 x_t^3$$

and the model is now a GLM with four degrees of freedom. Maximum likelihood estimates of the parameters $\beta_0,\ldots,\beta_3$ can be obtained using iteratively re-weighted least squares (IRLS) (McCullagh and Nelder, 1989), implemented in most general purpose statistical packages including *R*. Increasing the basis dimension *M*, that is increasing the number of basis functions, allows $f(x_t)$ to track the data more closely. As *M* decreases, $f(x_t)$ becomes smoother. As in any statistical model, there is a trade-off between closely fitting the observed data by choosing a high-dimensional basis that uses a large number of degrees of freedom, and a more parsimonious approach that uses fewer degrees of freedom by selecting a low dimensional basis. The latter will capture less of the variability in the observed data, but is more able to describe smooth broad-scale patterns that may be more widely applicable.

Part of the model-fitting process is to choose the appropriate degree of smoothness. Wood's method includes a penalty term in the model likelihood. This penalty term is controlled by a smoothing parameter λ which determines the trade-off between the goodness of fit of the model and model smoothness. If $\lambda = 0$ there is no smoothing, equivalent to fitting the model using the original basis dimension, and as $\lambda \to \infty$ there is a large amount of smoothing, giving a similar effect to choosing a lower basis dimension. For this method to be effective it is assumed that the user initially chooses a high-dimensional basis. Both λ and basis parameters are then estimated by a fitting algorithm.

Estimates of the basis parameters are obtained using penalized IRLS (Wood, 2006). The smoothing parameter is estimated using one of two methods, depending on whether the scale parameter (McCullagh and Nelder, 1989), which links the expectation to the variance, in the model is known or estimated. If it is known, the smoothing parameter is selected to minimize the un-biased risk estimator (UBRE) (Craven and Wabha, 1979). When the scale parameter is unknown and must be estimated then the smoothing parameter is selected to minimize the generalized cross validation score (GCV) (Craven and Wabha, 1979).

Given the estimated smoothing parameter, λ, and the basis dimension, *M*, the *effective* degrees of freedom (e.d.f.) for the smoothed covariate $f(x_t)$ can be calculated. The e.d.f. will be in the range $(1,M)$ and, similar to degrees of freedom in a GLM, is a measure of the contribution of that term to the overall model complexity. A term, $f(x_t)$, with one e.d.f. essentially represents a straight-line relationship and could be included as a linear term, βx_t. As the e.d.f. approaches *M* the relationship is more complex. If the e.d.f. are very close to *M*, Wood (2006) suggests that the model be refitted using a higher dimensional basis for that term.

Model formulation

The systematic component of a GAM can contain several explanatory variables, just as in a GLM. Covariates (continuous variables) can be included as linear terms or smooth, non-linear functions. An example with three covariates *w*, *x* and *z* could be

$$\eta = \beta_0 + \beta_1 w + f_2(x) + f_3(z)$$

Here, covariates *x* and *z* are included as smooth non-linear terms, whilst *w* is included as a linear term. Different bases can be used for each smoothed covariate and the smoothing parameter will be estimated separately for each covariate.

Smooth functions of two or more covariates can be fitted. Two covariates, *x* and *z*, could be included in the systematic component, η, in three ways depending on the relationship between them and the response variable: (1) Both terms are additive: $f_2(x) + f_3(z)$. In this case, each term is smoothed independently of the other covariate. (2) Variable coefficient models: $z + zf(x)$. These models were first proposed by Hastie and Tibshirani (1993). One covariate, in this case *x*, varies smoothly but the effect of the smooth is modified by the value of the other covariate *z*. (3) Bivariate smooths: $f(x,z)$. Here η varies non-linearly with *x* and *z* jointly. That is, the effect of *x* on η is non-linear and this changes non-linearly with *z*. This enables the exploration of complex interactions between covariates.

In addition to covariates, factors (categorical explanatory variables) can be included in GAMs in the same way as GLMs. It may be that the effect of a covariate is expected to depend on the value of a factor. Such an interaction can be estimated by fitting a separate smooth function of the covariate for each level of the factor. For example, a model with: (1) a continuous covariate *x*, (2) a factor *G* with three levels represented by three indicator variables g_1,

g_2 and g_3 so that $g_i = 1$ when $G = i$ and $g_i = 0$ otherwise, and (3) an interaction between G and x is modelled as

$$\beta_0 + \beta_1 g_1 + \beta_2 g_2 + f_1(x)g_1 + f_2(x)g_2 + f_3(x)g_3$$

The main effect of the factor G is included, but there is no main effect for the covariate.

Model comparison

For two competing models fitted to the same response variable, the GCV or UBRE scores can be compared and the model with the lower score is considered a better fit. This comparison can be made between two models with different explanatory variables, or two models that contain the same variables but differ in whether covariates are included as linear or smooth non-linear terms. Other methods for model comparison, such as Akaike's Information Criterion (AIC) (Akaike, 1973), and tests of the significance of individual terms are approximately valid, with limiting assumptions detailed by Wood (2006). Testing of linear terms proceeds as in a GLM.

Cubic regression splines

The regression smoothers $f(x_t)$ used by Wood are regression splines. This paper only considers *cubic* regression splines, a special, but at the same time very general, case of regression splines, because they are efficient when fitting large amounts of data. In a cubic regression spline, the basis functions define a natural cubic spline. This is a continuous curve consisting of separate segments of cubic polynomial curves. The segments are joined together at points, called *knots*. For a cubic regression spline of dimension M, $M-2$ knots, $(x_1^*, \ldots, x_{M-2}^*)$ are specified. The knots are usually selected to be equally spaced over the range of x, or as quantiles of the distribution of x, but other meaningful locations are possible. Given the knot locations, a set of basis functions $b_m(x)$ for $m = 1,\ldots,M$ can be defined that combine to give a natural cubic spline. Lancaster and Šalkauskas (1986) and Gu (2002) define some of the many different sets of basis functions that describe natural cubic splines.

A special case of a cubic regression spline is a cyclic cubic regression (c.c.r.) spline. These are useful when the covariate represents a cyclic feature, such as the day of the year. The c.c.r. splines are such that $f(x_t)$ and its first and second derivatives have the same values when $x_t = min(x)$ and $x_t = max(x)$.

To model bivariate smooths, for example $f(x,z)$ when the variables are on different scales, tensor product smooths are required. The basis for these smooth functions is constructed by combining separate bases for each covariate. For example, if $f(x_1) = \sum_{m=1}^{M} b_m(x_1)\beta_m$ and $f(x_2) = \sum_{n=1}^{N} c_n(x_2)\gamma_n$ both describe cubic regression splines, the basis for $f(x_1,x_2)$ is $\sum_{m=1}^{M}\sum_{n=1}^{N}\delta_{mn} b_m(x)c_n(x_2)$. Separate smoothing penalties are calculated for the two covariates so that the degree of smoothness is not necessarily the same for each covariate. The e.d.f. for $f(x_1,x_2)$ will satisfy $1 \leqslant$ e.d.f. $\leqslant MN$.

Using GAMs to model daily rainfall data

General strategy

In a GLM a typical model for the pattern of rainfall through the year could include terms describing: (1) the seasonal pattern of rainfall through the year modelled using Fourier series, $S_{kt} = sin(kt')$ where $t' = 2\pi t/366$ and $C_{kt} = cos(kt')$ for $k = 1,\ldots,K$; (2) the effect of the occurrence of rain on the previous day, J_{t-1}, or days, i.e. the state of the previous day(s); (3) how the effect of the occurrence of rain on the previous day differs depending on the time of year, modelled by including interactions between J_{t-1} and some $K' \leqslant K$ of the Fourier series terms; and (4) in the intensity model, the effect of the amount of rain on previous days modelled as $log(x_{t-1} + 1)$ when $J_{t-1} = 1$, so that only non-zero amounts on the scale of the linear predictor are included. The systematic component with these four components is

$$\alpha_0 + \sum_{k=1}^{K}(\alpha_{sk}S_{kt} + \alpha_{ck}C_{kt}) + \gamma_0 J_{t-1} + J_{t-1}\sum_{k=1}^{K'}(\gamma_{sk}S_{kt} + \gamma_{ck}C_{kt}) + \delta J_{t-1} log(x_{t-1} + 1)$$

where the αs, γs and δ are parameters to be estimated.

Using GAMs, the seasonal pattern could be modelled using a c.c.r. spline on d_t, the day of the year at time t, as an alternative to Fourier series terms. To explore whether the effect of the previous days rainfall on the seasonal pattern differs through the year, two separate c.c.r. splines, $f_1(d_t)$ and $f_2(d_t)$, can be fitted so that $f_1(d_t)$ describes the seasonal pattern when the previous day is wet, $J_{t-1} = 1$, and $f_2(d_t)$ describes the seasonal pattern when the previous day is dry, $J_{t-1} = 0$. In the intensity model, the effect of $log(x_{t-1} + 1)$, the amount of rain on the previous wet day, can be included as a smooth function $f_3(log(x_{t-1} + 1))$ rather than constrained to be linear. The systematic component using these smooth terms is

$$\alpha + \gamma J_{t-1} + J_{t-1} f_1(d_t) + (1 - J_{t-1}) f_2(d_t) + J_{t-1} f_3(log(x_{t-1} + 1))$$

Overall evidence of long-term trend can be examined by including either a linear function of time t (Chandler and Wheatear, 2002) or a smooth function of time t using a cubic regression spline basis. To avoid overfitting, the basis dimension should be of the same order as the number of years at most.

Because the pattern of rainfall through the year may depend on both the seasonal pattern, described by Fourier series or c.c.r. splines, and the state of weather on previous days, long-term trends in both these effects can be examined. For example, it may be that there is strong evidence for long-term trend in the wettest part of the year but not in the rest of the year, i.e. there is long-term trend in the seasonal pattern. Or it may be that the probability of rain following a dry day is increasing over time but the probability following a wet day is not changing, i.e. long-term trend depends on the state of the previous day.

To explore whether long-term trend varies depending on the state of the previous day, separate cubic regression splines can be fitted for long-term trend depending on whether the previous day is wet, $J_{t-1} = 1$, or dry, $J_{t-1} = 0$.

To examine long-term trend in the seasonal pattern, Fourier series must be used to describe the seasonal pattern. The interaction between individual Fourier series terms and long-term trend can be investigated so that either: (1) $f(t)$ varies linearly with a Fourier series term, i.e. $S_{kt}f(t)$ or (2) $f(t)$ varies non-linearly with a Fourier series term, $f(t,S_{kt})$. In the latter case, tensor product smooths are required because the two sets of covariates are measured on different scales t from $1,\ldots,T$ and S_{kt} from -1 to 1. Because the total basis dimension for two-way smooths is the product of the individual basis dimensions, it is preferable to choose a small basis dimension for each term, although the dimensions do not need to be the same. Bivariate smooths are not possible using c.c.r. splines.

In the occurrence model, UBRE can be used for model comparison because the scale parameter is one; the variance $var(J_t)$ is known given θ_t because $var(J_t) = \theta_t(1 - \theta_t)$. To model intensity, GCV is used because the variance $var(Y_t)$ requires the estimation of κ from which the scale parameter is derived.

Modelling strategy for the Mauritius data

The modelling strategy, for both the models of rainfall occurrence and intensity, addressed two issues. First, the order and rules for adding terms into the model is specified. This is important because of the large number of terms that are being considered. Second, to compare GLMs to GAMs, models with linear and non-linear terms for each covariate are considered.

The ordering of terms is based on the premise that, a priori, a seasonal pattern is expected and that this will be the largest source of variability in the data. Hence, models first capture the pattern of rainfall in an average year, by describing the seasonal pattern, using Fourier series or c.c.r splines and then, the effect of the occurrence and amount of rain on previous days. Given the pattern for an average year, the effect of long-term trend in capturing any of the remaining unexplained variability in the data can then be examined.

To compare linear and non-linear functions of covariates that describe the seasonal pattern of rainfall, two different sets of candidate variables were used: the first eight Fourier series terms, that is the variables S_{kt} and C_{kt} for $k = 1,\ldots,4$; and a c.c.r. spline for day of the year, d_t, with basis dimension $M = 10$. Using this basis dimension, the e.d.f. using the c.c.r. spline will be in a similar range to the degrees of freedom in a model using at most eight Fourier series terms.

Candidate variables for the effect of the weather on at most the previous two days were included. If the occurrence of the previous day J_{t-1} was important, the effect of rainfall occurrence on the previous two days was then examined. For this a factor, JJ_{4t}, that describes the four states of rain possible over the previous two days, replaced J_{t-1} in the model. The four states of this factor are given in Table 1. The interaction between the selected factor, if either was important, and seasonal terms were then examined. If intensity depended on the state of rain on the previous one, J_{t-1}, or two, J_{4t}, days, linear and smooth terms of the non-zero amounts of rain on the previous 1 or 2 days, respectively, $log(x_{t-l} + 1)$ when $J_{t-l} = 1$ for $l = 1,2$, were examined.

After describing the average pattern of rainfall through the year, evidence for long-term trend was examined by comparing models with linear and smooth functions of time, t. The basis dimension for the cubic regression spline was 41 with knot locations on the middle day of each of the 39 years, 1962–2001.

Different patterns of long-term trend through the year were examined by: (1) the interaction between time t, and rainfall occurrence on the previous day, J_{t-1}, if there was evidence of overall long-term trend and if J_{t-1} or JJ_{4t} was already included in the model and (2) in the Fourier series models, the interaction between t and the Fourier series terms, S_{1t} and C_{1t}. Here, the basis dimension for t was set to 10. When tensor products were fitted the basis dimension for the Fourier series terms was set to 4 so that only broad changes in the seasonal pattern were examined.

According to Wood (2006), there is a danger of overfitting GAMs. To reduce the risk, the UBRE of GCV was altered so that each effective degree of freedom had a weight of more than one. Using Wood's (2006) recommendation this is implemented in the *mgcv* software by setting the parameter *gamma* = 1.4.

Results

Model fitting and selection

Table 1 shows the notation for the variables considered in the different models. Tables 2 and 3 summarize the results for each of the occurrence models where the seasonal pattern is described using Fourier series (Table 2, OF* models) or cyclic cubic regression splines (Table 3, OC* models). Tables 4 and 5 give the results for each of the intensity models using Fourier series (IF* models) and c.c.r. splines (IC* models), respectively, to describe the seasonal pattern. Each row in the tables corresponds to a specific model, as numbered on the left-hand side, and the specific term added to the model is given. The e.d.f. are given for each of the terms in the model, the total e.d.f. and the UBRE/GCV for each model.

Within each table, the order in which the models are presented follows the order in which terms were added according to the modelling strategy outlined in Section 'Modelling strategy for the Mauritius data'. In general, models with a large model number include more terms and so have a greater total e.d.f. than models with a low model number. For example in Table 2, models OF1–OF3 only include terms describing pattern of rainfall through an

Table 1
Terms used in models.

Covariate	Term	Definition	
Time	t		
Day of the year	d_t		
Fourier series – sine	S_{kt}	$sin(kt')$	$k = 1,\ldots,4$, $t' = 2\pi t/366$
Fourier series – cosine	C_{kt}	$cos(kt')$	$k = 1,\ldots,4$, $t' = 2\pi t/366$
Amount of rain	x_t		
Intensity of rain	y_t	$x_t - 1$	When $x_t \geqslant 1$
Categorical variable	**Term**	**Levels**	**Definition**
Day t wet	J_t	$J_t = 1$	$y_t > 0$
Day t dry		$J_t = 0$	$y_t = 0$
Weather states for day $t-1$ and $t-2$	JJ_{4t}	$JJ_{4t} = rr$	$J_{t-1} = 1$ and $J_{t-2} = 1$
		$JJ_{4t} = rd$	$J_{t-1} = 1$ and $J_{t-2} = 0$
		$JJ_{4t} = dr$	$J_{t-1} = 0$ and $J_{t-2} = 1$
		$JJ_{4t} = dd$	$J_{t-1} = 0$ and $J_{t-2} = 0$
Number of wet days in last two days	JJ_{3t}	$JJ_{3t} = rr$	$J_{t-1} = 1$ and $J_{t-2} = 1$
		$JJ_{3t} = rd$	$J_{t-1} = 1$ and $J_{t-2} = 0$
		$JJ_{3t} = d$	$J_{t-1} = 0$
Systematic component	η_t	$log(\theta_t/(1-\theta_t))$	$\theta_t = Pr(J_t = 1)$
	ζ_t	$log(\mu_t)$	$\mu_t = E(Y_t)$
Other terms	$\alpha, \beta, \gamma, \delta$		Parameters
	$f()$		Smooth spline

Table 2
E.d.f. and model summaries for occurrence model using Fourier Series for seasonal pattern.

No.	Terms	Effective degrees of freedom										Model summaries
		Season	PD		Trend		Trend.PD		Trend.Season		Total	
		FS[b]	JJ_{4t}	C_{2t} $J_{t-1}=1$	t $M=20$	$f(t)$ $J_{t-1}=1$	$f(t)$ $J_{t-1}=0$	$f(t)$ S_{1t}	$f(t)$	$f(t,s_{1t})$		UBRE
OF1.	$S_{1t}+S_{2t}+C_{1t}+C_{2t}+C_{4t}$	5									6	0.3545
OF2.	Prev. days (PD)[a]	5	3								9	0.3035
OF3.	Seasonal.PD	5	3	1							10	0.3031
OF4.	Trend	5	3	1	1						11	0.2980
OF5.	$f(\text{Trend}, M=20)$	5	3	1		9.5					19.5	0.2937
OF6.	$f(\text{Trend})$.PD	5	3	1			8.8	3.6			22.4	0.2936
OF7.	$f(\text{Trend}).S_{1t}$	5	3	1			8.8	3.6	1.0		23.4	0.2938
OF8.	$f(\text{Trend}, S_{1t})$	4	3	1			8.5	2.6		3.3	23.4	0.2938

[a] Effect of previous days.
[b] Fourier series.

Table 3
E.d.f. and model summaries for occurrence model using c.c.r. spline for seasonal pattern.

No.	Terms	Effective degrees of freedom									Model summary
		Season	PD	Season.PD		Trend		Trend.PD		Total	
		d_t	JJ_{4t}	d_t $J_{t-1}=1$	d_t $J_{t-1}=0$	t	$f(t)$ $M=20$	$f(t)$ $J_{t-1}=1$	$f(t)$ $J_{t-1}=0$		UBRE
OC1.	$f(\text{day})$	6.2								7.2	0.3545
OC2.	Prev. days (PD)[a]	5.4	3							9.4	0.3036
OC3.	$f(\text{day})$.PD		3	5.1	4.3					13.5	0.3035
OC4.	Trend		3	5.1	4.4	1				14.5	0.2984
OC5.	$f(\text{Trend}, M=20)$		3	5.2	4.4		9.5			23.0	0.2941
OC6.	$f(\text{Trend})$.PD		3	5.2	4.4			8.8	3.5	25.9	0.2940

[a] Effect of previous days.

Table 4
E.d.f. and model summaries for intensity model using Fourier Series for seasonal pattern.

No.	Terms	Effective degrees of freedom															Model summary
		Seasonal			Amounts				Trend			Trend.PD		Trend.FS		Total	
		FS[b]	JJ_{3t}	c_{1t} $J_{t-1}=1$	x_{t-1} $J_{t-1}=1$	$f(x_{t-1})$ $J_{t-1}=1$	x_{t-2} $JJ_{3t}=rr$	$f(x_{t-2})$ $JJ_{3t}=rr$	t	$f(t)$ $M=41$	$f(t)$ $M=10$	$f(t)$ $J_{t-1}=1$	$f(t)$ $J_{t-1}=0$	$f(t)$ S_{1t}	$f(t,S_{1t})$		GCV[c]
IF1.	$S_{1t}+S_{2t}+C_{1t}+C_{3t}$	4														5	150
IF2.	PD[a]	4	2													7	133
IF3.	FS.PD	4	2	2												9	121
IF4.	Amount_{t-1}	4	2	2	1											10	35
IF5.	$f(\text{Amount}_{t-1})$	4	2	2		2.6										11.6	33
IF6.	Amount_{t-2}	4	2	2		2.7	1									12.7	31
IF7.	$f(\text{Amount}_{t-2}))$	4	2	2		2.7		2.6								14.2	30
IF8.	Trend	4	2	2		2.7		2.6	1							15.2	24
IF9.	$f(\text{Trend}, M=41)$	4	2	2		2.6		2.4		30.1						43.9	14
IF10.	$f(\text{Trend}, M=10)$	4	2	2		2.6		2.5			3.4					17.6	23
IF11.	$f(\text{Trend})$.PD	4	2	2		2.7		2.6				1.9	2.3			18.4	23
IF12.	$f(\text{Trend}).S_{1t}$	4	2	2		2.6		2.5						6.1		20.1	29
IF13.	$f(\text{Trend}, S_{1t})$	3	2	2		2.6		2.4							18.1	33.0	20

[a] Effect of previous days.
[b] Fourier series.
[c] $100 * (\text{GCV} - 1.3)$.

average year, models OF4 and OF5 include functions of time, t, to describe long-term trend, and OF6–OF8 include terms describing long-term trend in the pattern through the year. The total e.d.f. for OF1 is 6 because of the five Fourier series terms in the model, and model OF8 with an extra five terms in the model has a total e.d.f. of 23.4. Of the eight models in Table 2, OF6 has the lowest UBRE and so, best describes how the occurrence of rain varies through time.

The same conclusions can be drawn about the pattern of rainfall occurrence using either Fourier series terms or c.c.r. splines to describe the seasonal pattern. Essentially, given the seasonal pattern, rainfall occurrence depends on whether the previous day was wet or dry, J_{t-1}, there is evidence of long-term trend and this trend differs depending on the state of the previous day.

Comparing the total e.d.f. for the best occurrence model using Fourier series (OF6) and c.c.r. splines (OC6), the model using Fourier series terms to describe the seasonal pattern is more parsimonious and a better fit; the UBRE and total e.d.f.s are lower than when using c.c.r. splines. This is because two separate c.c.r. splines are required to model the seasonal pattern of rainfall, one when the previous day is wet, $J_{t-1}=1$, that uses 5.2 e.d.f., and the other, using 4.4 e.d.f., when the previous day is dry, $J_{t-1}=0$. Each spline

Table 5
E.d.f. and model summaries for intensity model using c.c.r. spline for seasonal pattern.

No.	Term	Effective degrees of freedom															Model summary
		Seasonal					Amounts				Trend			Trend*PD		Total	
		d_t	JJ_{3t}	d_t $JJ_{3t} = rr$	d_t $JJ_{3t} = rd$	d_t $JJ_{3t} = dd$	x_{t-1} $J_{t-1} = 1$	$f(x_{t-1})$ $J_{t-1} = 1$	x_{t-2} $JJ_{3t} = rr$	$f(x_{t-2})$ $JJ_{3t} = rr$	t	$f(t)$ $M = 41$	$f(t)$ $M = 10$	$f(t)$ $J_{t-1} = 1$	$f(t)$ $J_{t-1} = 0$		GCV[b]
IC1.	f(day)	6.3														7.3	152
IC2.	Prev. days (PD)[a]	6.3	2													9.3	134
IC3.	f(day).PD		2	6.8	4.7	5.8										20.3	134
IC4.	Amount$_{t-1}$		2	2.6	2.3	4.7	1									13.6	48
IC5.	f(Amount$_{t-1}$)		2	2.6	2.3	4.7		2.6								15.3	44
IC6.	Amount$_{t-2}$		2	2.6	2.3	4.5		2.7	1							16.1	42
IC7.	f(Amount$_{t-2}$)		2	2.6	2.3	4.4		2.7		2.5						17.6	41
IC8.	Trend		2	2.5	2.3	4.9		2.7		2.5	1					18.9	35
IC9.	f(Trend, $M = 41$)		2	2.7	2.2	4.0		2.5		2.5		30.2				47.2	26
IC10.	f(Trend, $M = 10$)		2	2.5	2.3	4.7		2.6		2.5			3.7			21.3	34
IC11.	f(Trend).PD		2	2.4	2.3	4.6		2.7		2.5				2.7	2.5	22.7	34

[a] Effect of previous days.
[b] $100 * (\mathrm{GCV} - 1.3)$.

only uses part of the observed data for fitting the seasonal pattern and the two splines are not constrained to be in any way similar. Using Fourier series, the interaction between J_{t-1} and each of the Fourier series term is assessed separately. Because, in this case only one interaction, with C_{2t}, is required, most of the seasonal pattern is described using all of the data. To describe only the pattern of

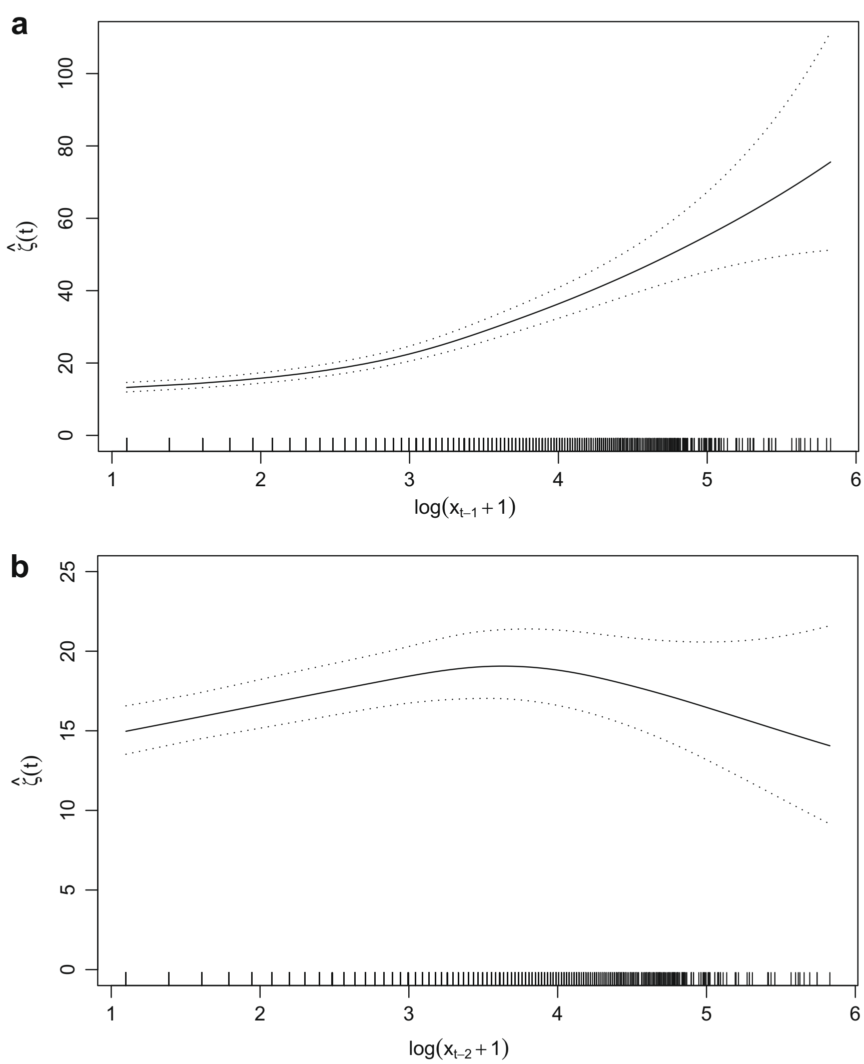

Figure 2 Effect of previous days of rainfall (a) x_{t-1} and (b) x_{t-2} on the systematic component with 95% confidence intervals (dotted lines). ζ_t is calculated using parameters from model IF7 for 15th February, day 183 ($d_t = 183$), when $JJ_{3t} = rr$ and (a) $x_{t-2} = \bar{x}_{t-2}$ when $J_{t-2} = 1$ and $d_t = 181$ and (b) $x_{t-1} = \bar{x}_{t-1}$ when $J_{t-1} = 1$ and $d_t = 182$.

rainfall occurrence through the year, a GLM using Fourier series (model OF3) would better describe the data than a GAM using c.c.r. splines (model OC3).

For intensity, the best model using Fourier series terms is model IF13; using c.c.r. splines it is model IC10. Again, Fourier series terms provide a more parsimonious model than c.c.r. splines for the seasonal pattern. Both intensity models show a seasonal pattern that is modified by the state of rain on the previous two days and smooth effects of rainfall amounts on the the previous two days. Fitting the four level factor JJ_{4t}, that described the state of rain on the last two days, it was clear that only three levels were important. These three levels describe the number of consecutive wet days and so a three level factor JJ_{3t}, described in Table 1, replaced JJ_{4t}. There was evidence of long-term trend and no evidence that this trend depended on the state of previous days. However, long-term trend in the seasonal pattern was evident using Fourier series and tensor product smooths. Because these smooths cannot be fitted using c.c.r. splines the best c.c.r. spline model, IC10 has a higher GCV than the best Fourier series model, IF13 because it does not explain as much of the variability in the data.

Non-linear effects of all of the continuous variables that were examined, provided a better fit to the data than linear terms in both the intensity and occurrence models. In the models examined in this paper, linear effects were important – they always showed an improvement over a model without that term – and this suggests that GLMs could be used to model long-term trend. However, fitting non-linear functions within a GAM provided an even better fit.

In this study, the relationship between the covariates and the response variable was unknown prior to fitting, at least to the author. One reason for using GAMS is that they provide a simple mechanism for examining these relationships. For example, using the estimated parameters from IF7, the effect of the amount of rain on the previous day, $log(x_{t-1}+1)$, on intensity can be seen. Figure 2a shows how ζ_t varies with $log(x_{t-1}+1)$ for the middle day of the season (15th February) when the previous two days are wet, $JJ_{3t} = rr$. A straight line would indicate a linear effect of $log(x_{t-1}+1)$ on ζ_t which could be refitted using a linear term. Figure 2b shows how ζ_t varies with x_{t-2} using estimated model parameters from IF7. These plots could be used to suggest a functional form for the covariates, for example a quadratic function of $log(x_{t-1}+1)$. If all smooth terms in a GAM were replaced with known functional forms then the model becomes a GLM. This may be useful, in particular if tests of significance are important, because of the stronger inferential basis for GLMs.

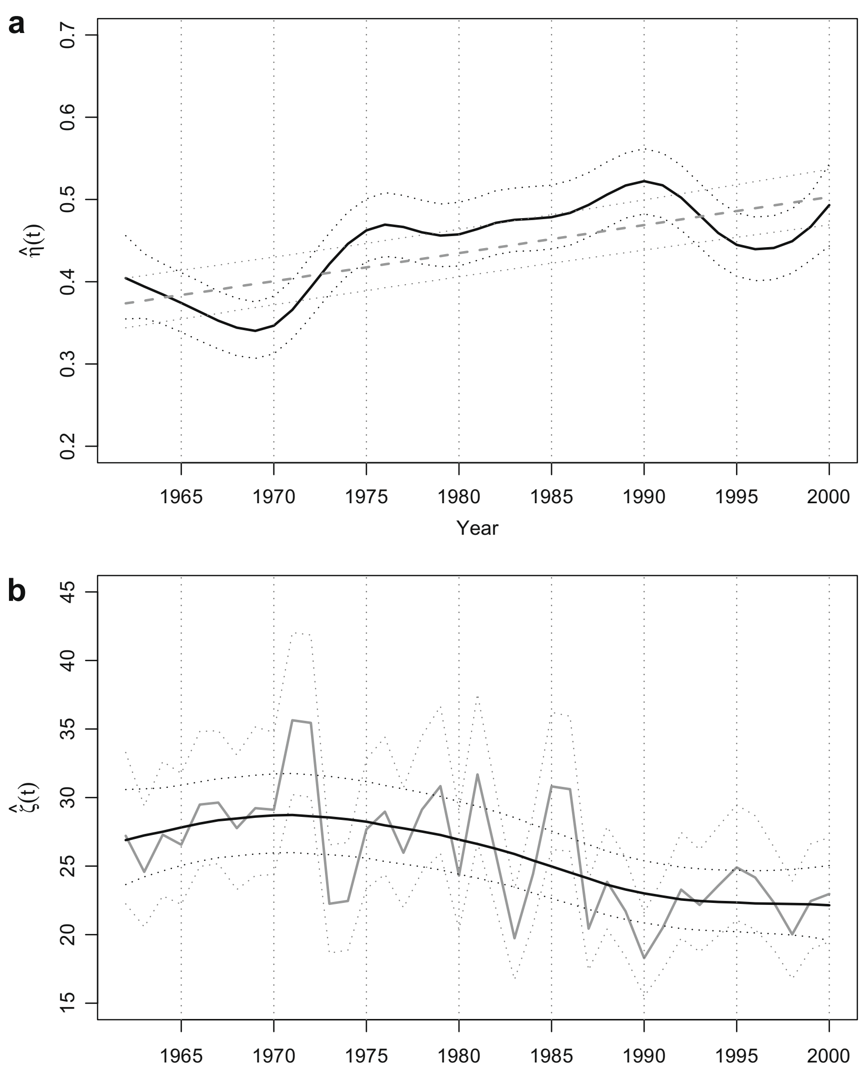

Figure 3 Effect of long-term trend on linear predictor with 95% confidence intervals (small dotted lines) for (a) occurrence model (OF5), where dashed line represents a linear trend and the solid line a smooth function of time $f(t, M=20)$; (b) intensity model, where solid gray line is $f(t, M=41)$ (IF9) and solid black line is $f(t, M=10)$ (IF10).

A difficulty with fitting smooth terms for long-term trend was the choice of basis dimension, M, originally set to 41, in both the occurrence and intensity models. In the occurrence models, the e.d.f. for this term was less than 10, before interactions with other terms were considered. Using a lower-dimensional basis for this term, so that $M = 20$ with knots located every other year, did not change the UBRE or the e.d.f. for this model (OF5 and OC5). Note that the basis dimension $M = 20$ is still much larger than the final e.d.f. for this term, suggesting that the choice of M was reasonable. Figure 3a shows the average long-term trend on η_t using paramater estimates from model OF5. The trend (black line) is smooth, different to a linear effect (grey line) and is similar to the long-term trend in the total number of wet days per year, Figure 1a.

For the intensity models, the e.d.f. was 30 for the non-linear trend term (models IF9 and IC9). Figure 3b shows this effect (grey line) using estimated parameters from model IF9. The trend overfits the data, because it is not smooth, and does not achieve the original aim of describing broad changes in rainfall patterns. Refitting the model with a smaller, 10-dimensional, basis ($f(t, M = 10)$) resulted in a non-linear trend with 3.9 e.d.f. (IF10) or 3.7 (IC10), which suggests that M was a reasonable choice if the objective is to provide a smoothed trend. The effect of $f(t, M = 10)$ is shown as the smooth black line in Figure 3b indicating evidence of a non-linear declining trend in the intensity of rain. But, this model does not fit the data so well; the GCV using $M = 41$ (IF9) is 14 compared to 23 when $M = 10$ (IF11) although this is still lower than a model with a linear trend (GCV = 24, IF8). It is possible that the extra variability apparently explained by the interaction between long-term trend and the seasonal pattern, $f(t, S_{1t})$, is capturing this unexplained variability.

Model checking, using the methods described by Chandler and Wheatear (2002), indicated that the models generally fitted well. Model IF9 for intensity, which fits $f(t, M = 41)$, still describes the variability in the data better than more complex models with $f(t, M = 10)$, for example model IF13.

Interpretation

Figures. 4–7 demonstrate how the best models for occurrence (OF6) and intensity (IF13) can provide useful visual summaries of how rainfall patterns have changed from 1962 to 2001. Using these models, the estimated probability of occurrence and rainfall inten-

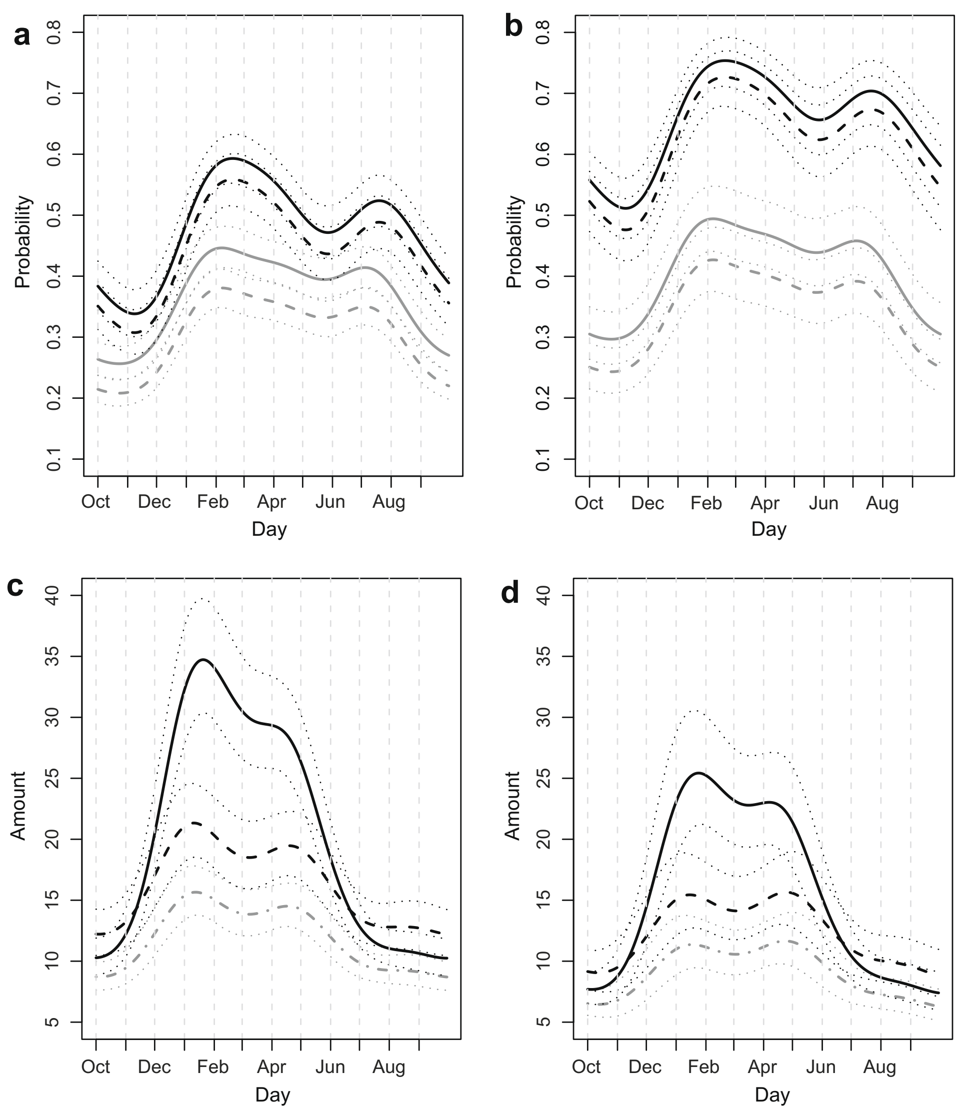

Figure 4 Predicted mean values and 95% confidence intervals (small dotted lines) for probability of rain (using model OF6) in (a) 1969 and (b) 2000 and intensity of rain (using model IF13) in (c) 1969 and (d) 2000. Solid black line is probability when $JJ_t = rr$, dashed black line when $JJ_t = rd$, solid gray line is $JJ_{4t} = dr$, dashed gray line $JJ_{4t} = dd$. Dot-dash gray line (for intensity) is $JJ_{3t} = d$. Note $JJ_t = JJ_{4t} = JJ_{3t}$.

sity can be calculated for every day in this period. Because the models include a term describing the state of weather in the previous two days (JJ_{4t} for occurrence and JJ_{3t} for intensity), the probabilities and intensities are estimated assuming each of the possible combinations of weather in the previous two days. To obtain estimates of rainfall intensity, y_t, the model requires values for x_{t-1} and x_{t-2}, the amount of rain on the previous two days when these were wet, $J_{t-1} = 1$ and $JJ_{3t} = rr$, respectively. Because smooth estimates of intensity were required, values of x_{t-1} and x_{t-2} were estimated to vary smoothly over the year and assumed constant over years, so that long-term trend in intensity can be assessed. Estimates were obtained by: (1) calculating y_{t-1} and y_{t-2}, the mean intensity of rain on day $t-1$ and day $t-2$ for each day of the year when $J_{t-1} = 1$ and $JJ_{3t} = rr$, respectively; (2) fitting a simple GAM (of the form of intensity model IC1) to these data and (3) using these models to estimate $x_{t-1} = y_{t-1} + 1$ and $x_{t-2} = y_{t-2} + 1$ for days 1–366.

Figure 4 shows the seasonal patterns of occurrence and intensity for two different years, 1969 and 2000, and how this pattern varies depending on the state of rain in the previous two days. The probability of rain, and the intensity of rain is greatest if both previous days were wet ($JJ_{4t} = rr$) and lowest when both were dry (dd), for occurrence and when the previous day was dry ($JJ_{3t} = d$) for intensity. The wettest time of year is in January and February and the driest in September. Figure 5 shows long-term trend in occurrence and intensity over the 39 years for two different days of the year, 1st January and the 1st June and how this trend differs depending on the state of rain on previous days. For rainfall occurrence, long-term trend is very different depending on the whether the previous day was wet or dry. In general, the probability that it rains is increasing over time, particularly when the previous day is wet. Both Figures 4 and 5 show how the shape of the seasonal pattern of occurrence depends on the state of the previous day only, and does not differ between years, because only the interaction between long-term trend and J_{t-1} is in the model (OF6). In comparison, long-term trend in rainfall intensity differs greatly depending on the time of year because there is trend in the seasonal pattern; the model (IF13) includes a tensor product smooth of time with Fourier series terms.

Figure 6 shows the seasonal pattern for the each year for the wettest and driest states for rainfall occurrence (a) and intensity

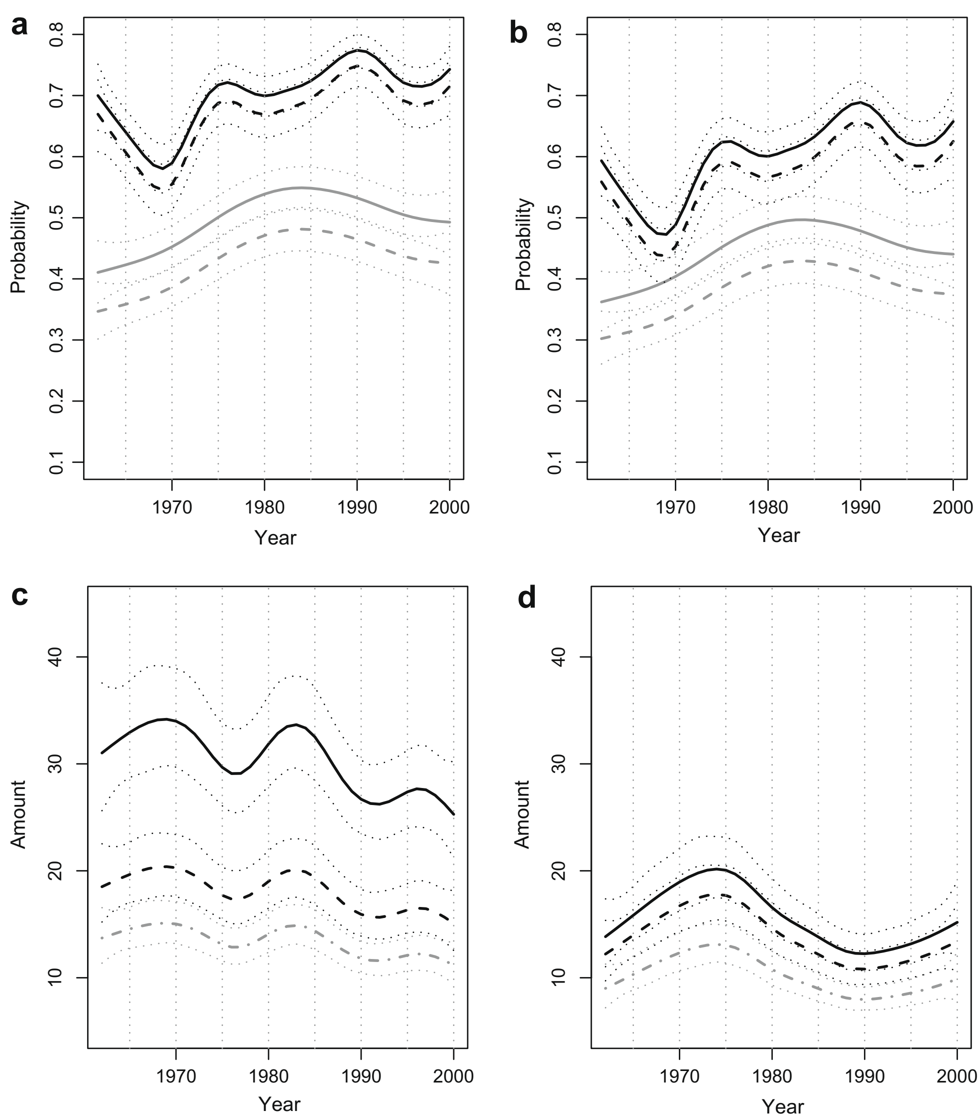

Figure 5 Predicted mean values and 95% confidence intervals for probability of rain (using model OF6) for (a) 1st January and (b) 1st June and intensity of rain (using model IF13) for (c) 1st January and (d) 1st June each year. Solid black line is probability when $JJ_t = rr$, dashed black line when $JJ_t = rd$, solid gray line is $JJ_{4t} = dr$, dashed gray line $JJ_{4t} = dd$. Dot-dash gray line (for intensity) is $JJ_{3t} = d$. Note $JJ_t = JJ_{4t} = JJ_{3t}$.

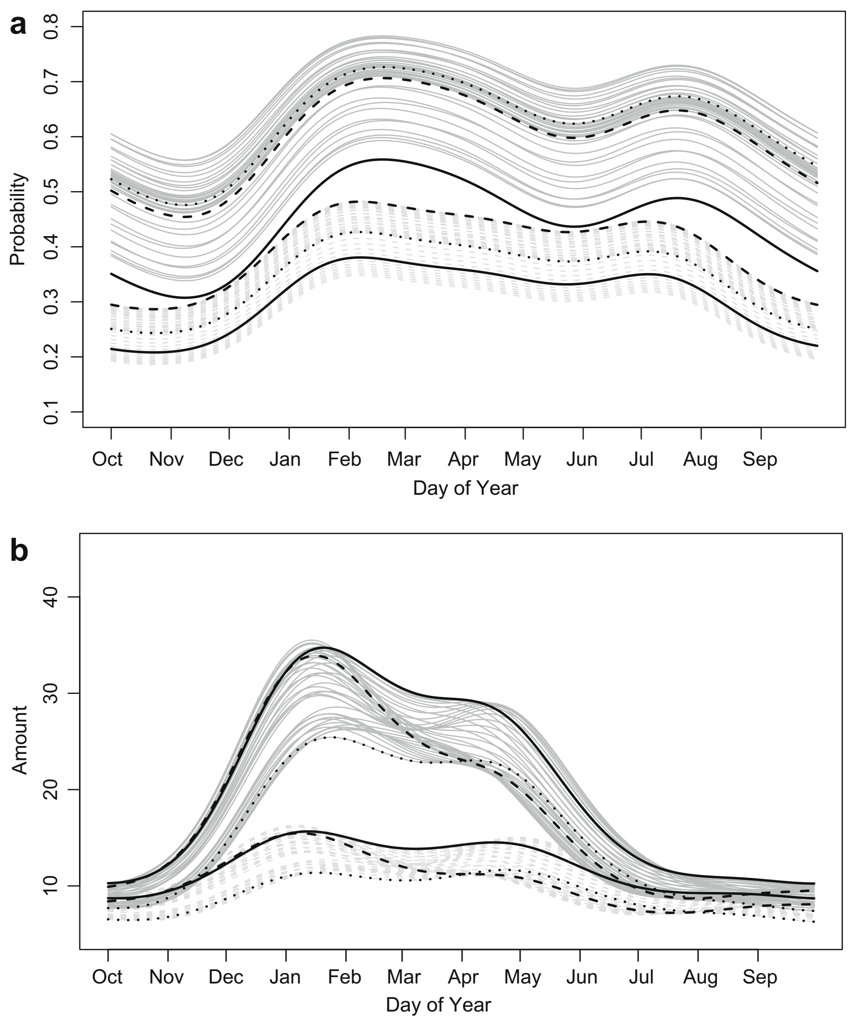

Figure 6 Predicted mean values for (a) probability of rain (using model OF6) and (b) intensity of rain (using model IF13) for each year with 3 years highlighted. 1969 – Solid line, 1985 – long dashes and 2000 – short dashes. In (a) upper darker profiles are $JJ_{4t} = rr$ and lower lighter profiles are $JJ_{4t} = dd$. In (b) upper darker profiles are $JJ_{3t} = rr$ and lower lighter profiles are $JJ_{3t} = d$.

(b). Three years, 1969, 1985 and 2000 are highlighted. The very different profile of rainfall intensity in 1985 compared to 1969 or 2000 is clear. Figure 7 expands these views as contour plots. Higher probabilities generally occur at the top of the plots of occurrence, indicating the increased probability of rain through time. The reverse is observed for intensity, although this is not uniformly true because of the changing seasonal pattern through time. The extended tail for intensity when $J_{3t} = rr$ around the year 1969 and the different shape for 1985 can be seen.

The models should not be used to extrapolate into the future because GAMs are unconstrained at the limits of the data. If future predictions were required, a more successful approach would be to fit models that attribute long-term trends to changes in atmospheric variables. Future prediction of atmospheric variables, say from a Global Climate Model, could then be used to predict future patterns of rainfall occurrence or intensity.

Discussion

The original motivation for this paper was to investigate how GAMs could be used to model long-term trends in daily rainfall data. After accounting for the pattern of rainfall through the year in models of occurrence or intensity of rainfall, GAMs can describe long-term trend by fitting smooth functions of time. Separate smooths can be fitted to examine whether there are different patterns in long-term trend depending on the occurrence of rain on the previous day and tensor product smooths can be used to explore whether there is evidence of long-term trend in the seasonal pattern itself.

Average seasonal patterns of rainfall can also be modelled using GAMs. In particular, a smooth function of the day of the year can be modelled by a c.c.r. spline basis in place of Fourier series functions. However, Fourier series functions provide more parsimonious and better fitting models if the seasonal pattern depends on the occurrence of rain on previous days. Furthermore, Fourier series functions provide greater flexibility because long-term changes in seasonal patterns can be modelled using tensor product smooths of Fourier series terms and time t.

Finally, other covariate can successfully be modelled using GAMs to explore whether effects are linear or non-linear. Here the amount of rain on previous days is included and large-scale climate effects such as SOI (Hyndman and Grunwald, 2000) could also be examined. This could be further extended to examine whether the effect of the SOI on the seasonal rainfall pattern is changing over time.

In general, GAMs provide a useful extension to GLMs because they give the ability to examine the relationship between covariates and response variables and whether they are linearly or

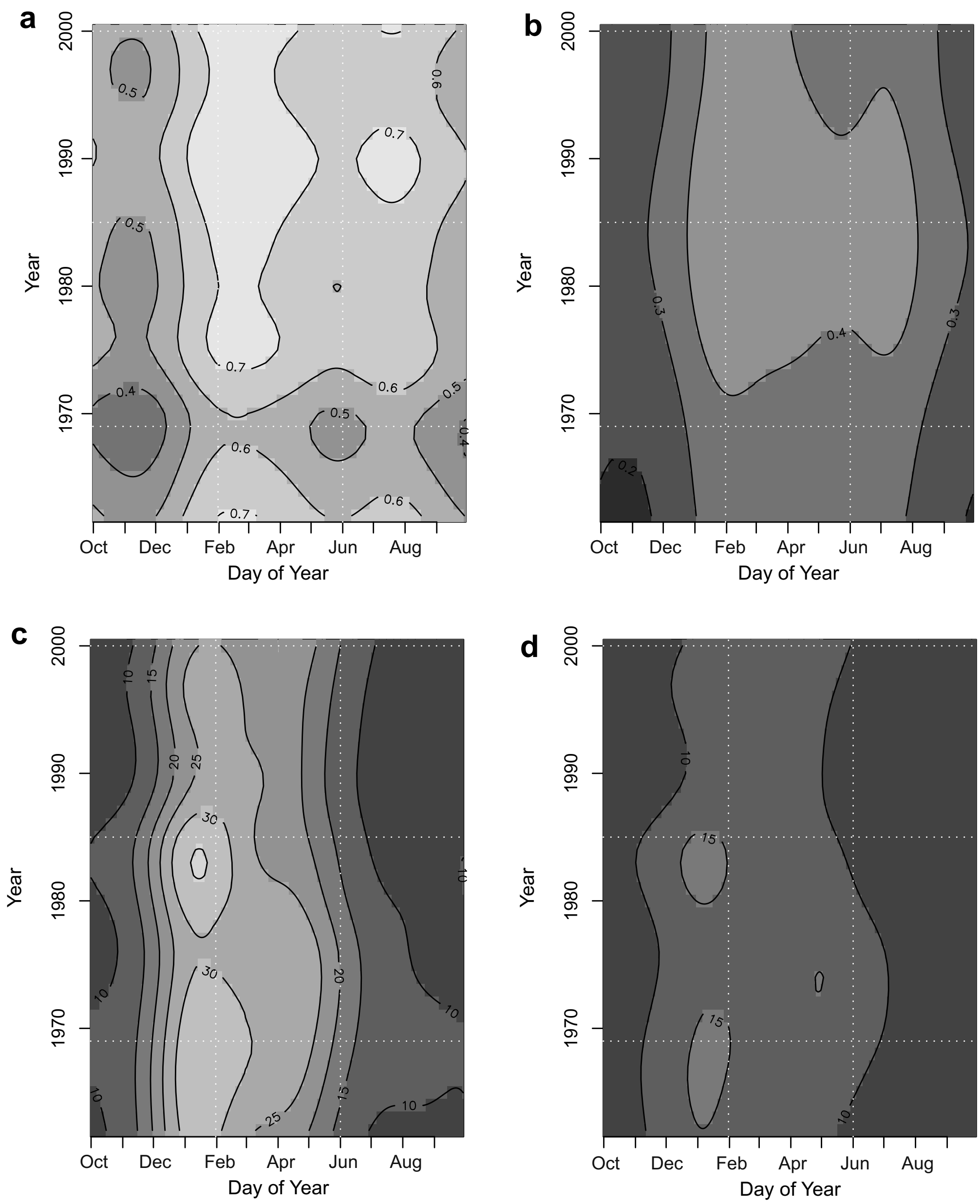

Figure 7 Predicted mean probability (using model OF6) and intensity (using model IF13) of rain for each day in each year for (a) probability given $JJ_{4t} = rr$; (b) probability given $JJ_{4t} = dd$; (c) intensity given $JJ_{3t} = rr$ and (d) intensity given $JJ_{3t} = d$. White dashed lines indicate cross-sections shown in other figures – vertical lines are profiles in Fig. 5 and horizontal lines are profiles in Fig. 6.

non-linearly related. They are therefore most helpful for describing unknown relationships between covariates and response variables when there is no prior knowledge. If the functional relationship is known prior to fitting, then it seems more appropriate to fit this functional form directly within a GLM, as is the case with modelling the seasonal pattern using Fourier series. Using GAMs as an exploratory tool to suggest functional forms, that can then be refitted explicitly, could be an appropriate strategy, especially if explanations for the functional form can be provided.

Although GAMs are useful, care is needed when choosing the basis dimension for smooth terms. The choice of basis dimension is crucial, and somewhat subjective. If data are very variable, a high-dimensional basis provides a high-degree of flexibility so that detailed patterns in the data are described. Using a lower-dimensional basis will, when the data are very variable, only describe very smooth patterns. If the variability in the data is less, so that the trend dominates the data, the choice of basis is less important, as long as the basis is sufficiently large. This accounts for the differences between the behaviour of smooths for the occurrence and intensity models. Because intensity is so very variable, the choice of a high-dimensional basis did not give smooth long-term trends. As a further example, the trend shown in Figure 1a, of the total number of wet days per year is relatively stable with respect to the basis dimension ($M = 10$ or $M = 25$ give similar shapes). In comparison in Figure 1c, the mean rain per wet day, the trend is very smooth using a basis dimension of 10, as shown and was much less smooth using a higher basis dimension ($M = 25$, not shown), because it becomes flexible enough to follow the variability in the individual data values. The key, when using GAMs is therefore to keep the objectives of the model-fitting in mind when choosing the basis dimension, possibly in an iterative manner.

Retaining a high-dimensional basis for long-term trend could be useful if the purpose was to simulate rainfall in the period 1962–2000. An alternative modelling approach however would be a generalized additive mixed model (Lin and Zhang, 1999). This models the hierarchical structure of the data so that days of the year are nested within years, and each year varies about an average profile. Wood (2006) considers a simple case of this for temperatures.

The use of GAMs could be extended to model daily rainfall data from several stations together as an alternative to the GLM models of Yang et al. (2005). The interdependence between sites could be modelled by allowing model parameters to vary smoothly over space as well as time. Spatio-temporal modelling of ecological processes using GAMs is common (see Augustin et al., 1998 for an early example). This merits further attention and application to modelling daily rainfall data.

Acknowledgements

The Darwin Initiative provided funding for an initial analysis of this data. Dr. M.A. Nicoll at the University of Reading entered the daily rainfall data from records provided from the Agonomist, Mr. D. Le Blanc, at Ferney Sugar estate. Mr. R.W. Burn provided thoughtful comments on the analysis and paper.

References

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petran, B., Csaaki, F. (Eds.), International Symposium on Information Theory. Akadeemiai Kiadi, Budapest, Hungary, pp. 267–281.
Augustin, N.H., Borchers, D.L., Clarke, E.D., Buckland, S.T., Walsh, M., 1998. Spatiotemporal modelling for the annual egg production method of stock assessment using generalized additive models. Canadian Journal of Fisheries and Aquatic Sciences 55, 2608–2621.
Beckmann, B.R., Buishand, T.A., 2002. Statistical downscaling relationships for precipitation in the Netherlands and North Germany. International Journal of Climatology 22, 15–32.
Chandler, R.E., 2005. On the use of generalized linear models for interpreting climate variability. Environmetrics 16, 699–715.
Chandler, R.E., Wheatear, H.S., 2002. Analysis of rainfall variability using generalized linear models: a case study from the west of Ireland. Water Resources Research 38 (10). doi:10.1029/2001WR000906.
Coe, R., Stern, R.D., 1982. Fitting models to daily rainfall data. Journal of Applied Meteorology 12, 1024–1031.
Craven, P., Wabha, G., 1979. Smoothing noisy data with spline functions. Numerische Mathematik 31, 377–403.
Furrer, E.M., Katz, R.W., 2007. Generalized linear modeling approach to stochastic weather generators. Climate Research 34, 129–144.
Gu, C., 2002. Cross-validation non-Gaussian data. Journal of Computational and Graphical Statistics 1, 169–179.
Guisan, A., Edwards, T.C., Hastie, T., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological Modelling 157, 89–100.
Hastie, T., Tibshirani, R., 1990. Generalized Additive Models. Chapman and Hall.
Hastie, H., Tibshirani, R., 1993. Varying-coefficient models. Journal of the Royal Statistical Society Series B 55, 757–796.
Hyndman, R.J., Grunwald, G.K., 2000. Generalized additive modelling of mixed distribution Markov models with application to Melbourne's rainfall. Australian & New Zealand Journal of Statistics 42, 145–158.
Lancaster, P., Šalkauskas, K., 1986. Curve and Surface Fitting: An Introduction. Academic Press.
Lin, X., Zhang, D., 1999. Inference in generalized additive mixed models using smoothing splines. Journal of the Royal Statistical Society Series B 61, 381–400.
McCullagh, P., Nelder, J., 1989. Generalized Linear Models. Chapman and Hall.
Mestre, O., Hallegatte, S., in press. Predictors of tropical cyclone numbers and extreme hurricane intensities over the North Atlantic using generalized additive and linear models. Journal of Climate, doi:10.1175/2008JCLI2318.1.
Morton, R., Henderson, B., 2008. Estimation of nonlinear trends in water quality: an improved approach using generalized additive models, Water Resources Research 44, W07420. doi:10.1027/2007WR006191.
R Development Core Team, 2008. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (www.R-project.org.
Senapathi, D., Underwood, F.M., Black, E., Nicoll, M., Norris, K., submitted for publication. Evidence for long-term regional changes in precipitation on the East Coast Mountains in Mauritius. International Journal of Climatology.
Stern, R.D., Coe, R., 1982. The use of rainfall models in agricultural planning. Agricultural Meteorology 26, 35–50.
Stern, R.D., Coe, R., 1984. A model fitting analysis of daily rainfall data (with discussion). Journal of the Royal Statistical Society Series A 147, 1–34.
Wabha, G., 1990. Spline Models for Observational Data. SIAM, Philadelphia.
Wilks, D.S., Wilby, R.L., 1999. The weather generation game: a review of stochastic weather models. Progress in Physical Geography 23, 329–357.
Wood, S.N., 2000. Modelling and smoothing parameter estimation with multiple quadratic penalties. Journal of the Royal Statistical Society Series B 62, 413–428.
Wood, S.N., 2006. Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC.
Wood, S.N., 2008. Fast stable direct fitting and smoothness selection for generalized additive models. Journal of the Royal Statistical Society Series B 70, 495–518.
Wood, S.N., Augustin, N.H., 2002. GAMs with integrated model selection using penalized regression splines and applications to environmental modelling. Ecological Modelling 157, 157–177.
Yang, C., Chandler, R.E., Ishamn, V., Wheater, H.S., 2005. Spatial–temporal rainfall simulation using generalized linear models. Water Resources Research 41, W11415. doi:10.1029/2004WR003739.