Estimating trends and seasonality in Australian monthly lightning flash counts

Bryson C. Bates

CSIRO Oceans and Atmosphere Flagship, Centre for Australian Weather and Climate Research, Wembley, Western Australia, Australia

Richard E. Chandler

Department of Statistical Science, University College London, London, United Kingdom

Andrew J. Dowdy

Bureau of Meteorology, Australia

Corresponding author: B. C. Bates, CSIRO Oceans and Atmosphere Flagship, Centre for Australian Weather and Climate Research, Private Bag No. 5, Wembley, Western Australia, Australia 6913 (bryson.bates@csiro.au)

Resubmitted to

Journal of Geophysical Research – Atmospheres, 16 March 2015

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/2014JD023011

Abstract

We present the results of a statistical analysis of lightning characteristics in mainland Australia for the period from approximately 1988 to 2012, based on monthly lightning flash count (LFC) series obtained from a network of 19 CIGRE-500 sensors. The temporal structures of the series are examined in terms of detecting and characterizing seasonal cycles, long-term trends and changes in seasonality over time. A generalized additive modeling approach is used to ensure that the estimated structures are determined by the data, rather than by the constraints of any assumed mathematical form for the trends and seasonal cycle. Results indicate strong seasonality at all sites, the presence of long-term trends at 16 sites, and interactions between trend and seasonality (corresponding to changes in seasonality over time) at 13 sites. The most systematic change corresponds to a progressive deepening of the seasonal cycle (i.e. an ongoing decline in winter lightning flash counts), and is most noticeable across southern Australia (south of 30° S). These results are consistent with previous analyses that have detected decreasing atmospheric instability during the austral winter since the mid 1970s. This is associated with increasing mean sea level pressure and declining rainfall.

Index Terms and Key Words

1986 *Informatics*: Statistical methods: inferential

1988 *Informatics*: Temporal analysis and representation

3309 *Atmospheric Processes*: Climatology

3324 *Atmospheric Processes*: Lightning

3394 *Atmospheric Processes*: Instruments and techniques

9330 *Geographic Location*: Australia

Key Points

We examine monthly lightning flash counts series from 19 sites across Australia

Generalized additive models used to analyze temporal structure of the series

Trends and changes in seasonality over time are evident at 16 sites

1. Introduction

Lightning is a weather phenomenon that is episodic and highly variable in space and time. It causes the ignition of wildfires, disruptions to electric power transmission and distribution, disruption to communication systems, infrastructure damage, injury and death to humans and livestock, and increases in airline operation costs and passenger delays [see, e.g. *Mackerras*, 1991; *Coates et al*., 1993; *Elsom*, 1993; *Curran et al*., 2000; *Uman*, 2010]. Lightning flashes can be grouped into two broad categories: those that do not strike the ground (so-called "cloud flashes") and those that do ("cloud-to-ground flashes"). Most lightning is generated in summer thunderstorms with each cloud-to-ground flash consisting of one or more return strokes. These strokes have typically short durations of some tens of microseconds and are separated by time intervals of some tens to hundreds of milliseconds [*Cummins and Murphy*, 2009]. Instruments for measuring lightning activity include space-based sensors, ground-based lightning flash counters, and lightning location systems (LLSs) [*Kuleshov et al*., 2009; *Schulz et al*., 2005; *Cecil et al*., 2014].

The analysis of lightning activity data is a broad area of research with many previous studies focusing on different aspects of lightning activity: summary statistics; mean annual, seasonal and diurnal climatologies; and relationships between lightning and severe weather, aerosols, indices of convection and modes of climate variability [see, e.g., *Sheridan et al*., 1997; *Petersen and Rutledge*, 1998; *Kuleshov and Jayaratne*, 2004; *Dai et al*., 2009; *Ranalkar and Chaudhari*, 2009; *Tinmaker et al*., 2009; *Darden et al*., 2010; *Schultz et al*., 2011; *Bovalo et al*., 2012; *Chronis*, 2012; *Yuan et al*., 2012; *Dowdy and Kuleshov*, 2014]. A subject that has received far less attention is the quantitative analysis of seasonality and trends in lightning series [*Collier et al*., 2013; *Villarini and Smith*, 2013].

Our area of interest is the detection and characterization of any long-term trends in lightning activity data and temporal changes in the seasonal cycle. The time series considered

herein are monthly lightning flash counts (LFCs). Statistically, the analysis of such data is challenging because their probability distribution is discrete rather than continuous, limited to non-negative integer values, usually positively skewed with a high number of zeros, and the variance of the data can greatly exceed the mean as is the case herein (section 4). Therefore, the use of either conventional linear and Poisson regression [*Faraway*, 2005, 2006] is not appropriate since the data violate the assumptions that underpin these methods. Moreover, there are no guarantees that: (1) the seasonal cycle will be well described by one or more sinusoids [cf. *Collier et al*., 2013]; (2) the underlying trend will be monotonic [cf. *Villarini and Smith*, 2013]; or (3) the observations are independent, a requirement of classical nonparametric tests for trend [see, e.g., *Chandler and Scott*, 2011, Sec. 2.4]. Consequently we have adopted a flexible nonparametric approach which imposes minimal assumptions on the functional forms of the seasonal cycle and trend function and hence allows the data to "speak for themselves" as far as possible. Such an approach is particularly useful for characterizing nonlinearity and non-monotonicity in time series data.

In this paper we use monthly LFC series for 19 sites across Australia for the period approximately from 1988 to 2012. In the next section we briefly describe the instruments and data used. Section 3 explains the methodology and the results of our analyses are presented in section 4. A discussion and our conclusions are given in sections 5 and 6, respectively.

2. Instruments and Data

The data used in this study were collected from a ground-based lightning detection network installed and operated by the Australian Bureau of Meteorology with the assistance of electricity supply authorities. The sensor used is the CIGRE-500 (Comité Internationale des Grands Réseaux Electriques, 500 Hz peak transmission filter circuit). Although the sensor was designed specifically to detect cloud-to-ground flashes, it can also respond to cloud flashes with about 68% of the lightning flash counts due to cloud-to-ground flashes

[*Kuleshov and Jayaratne*, 2004]. Estimates of the effective horizontal ranges of the sensor areas are 30 km for cloud-to-ground flashes and 15 km for cloud flashes. Consequently, we analyzed total lightning flash counts rather than flash densities (flashes km^{-2}) or flash rates (flashes km^{-2} mo^{-1}) since the flash count is the original unit of measurement, and the portion of cloud-to-ground flashes and estimates of the effective range of the sensor are subject to uncertainty. The number of flashes is registered by an electromechanical counter, and multiple strokes within a flash are eliminated by a one-second dead time interval after the first stroke. The counters are read manually each day between 0800 and 0900 hours local time and the records sent to the Bureau of Meteorology at the end of each month. The sensors have been maintained throughout their operational life. Observers routinely perform test counts at times when there is no lightning activity. During these tests, a test button is pressed and the counter is expected to advance by about one count per second. The daily logs for each site reveal the timing and the number of test counts. The tests are performed on a weekly to monthly basis. Any fault in a sensor is noted and dated in the records, as is the recommissioning date. Data from periods where an instrument malfunctioned have been coded as 'missing values' in our analysis, thus ensuring that the remaining record is homogeneous with respect to the ability to detect LFCs when they occur. Further details can be found in *Kuleshov and Jayaratne* [2004], *Jayaratne and Kuleshov* [2006] and *Kuleshov et al*. [2009].

While the CIGRE-500 network has consisted of about 40 sites scattered widely across the Australian mainland during its operational life, there are only 19 sites that have at least 200 complete monthly observations during the period from the late 1980s to the end of available record (Figure 1 and Table 1). For 17 out of the 19 sites, monthly LFC records are available from January 1988, whereas existing records for Townsville and Eucla commence during August 1989 and March 1995, respectively. In recent years, there has been a shift away from

maintaining the CIGRE-500 network in favor of investment in the use of lightning location systems (LLSs) due to their greater spatial coverage. Nevertheless, we prefer to use the CIGRE-500 data for reasons discussed in section 5.

< Insert Figure 1 and Table 1 about here >

The lightning climatology of Australia varies considerably in both space and time. Previous analyses of the annual distribution of lightning occurrence data obtained from the CIGRE-500 network have identified a strong seasonal cycle [*Jayaratne and Kuleshov,* 2006; *Kuleshov et al*., 2009]. The occurrence frequency of thunderstorms varies somewhat with latitude. They are most frequent in the north and generally decrease in frequency southward. In the northern half of Australia, the development of thunderstorms is enhanced by low surface pressure and high boundary-layer moisture levels in the wet season (October to April). During the dry season the sub-tropical high pressure belt lies over the continent bringing clear skies and dry, stable conditions. At higher latitudes the annual distribution of lightning activity is somewhat more uniform since it can occur in cooler months (May to September) due to frontal systems associated with Southern Ocean depressions. There is, however, a maximum of lightning activity in the Australian region during the cooler months that occurs near the central east coast [Dowdy and Kuleshov, 2014]. This local maximum in lightning activity could potentially relate to a number of phenomena, including the warm East Australian Current as a potential source of latent heat flux into the cool wintertime boundary layer, as well as the occurrence of cold fronts or extratropical cyclones in the region (known as East Coast Lows). The central east coast of Australia is a favored region for East Coast Lows to occur [*Dowdy et al*., 2013], and thunderstorm observations have been associated with East Coast Lows occurrence [*Chambers et al*., 2014].

We readily concede that the spatial coverage of the CIGRE-500 network in Australia is very sparse, particularly when it is compared to those of modern ground-based lightning

location systems (LLSs) and the gridded climatologies of total lightning flash rates observed by NASA's spaceborne Optical Transient Detector (OTD) and Lightning Imaging Sensor (LIS) [*Boccippio et al*., 2002; *Christian et al*., 2003; *Dowdy and Kuleshov*, 2014]. However there are several reasons why we chose to use the CIGRE-500 dataset. It has been found that the LFC registrations are more accurate than Australian LLS data for distances less than 30 km [*Kuleshov et al*., 2009]. Moreover, there have been long-term changes in detection rates associated with changes in the density of sensors for the LLS and changes in the methods used for signal processing. These changes could compromise the interpretation of the results of our trend analysis. This together with the relatively short length of record (about 15 years) means that the data obtained by the LLS sensors are not ideal for our type of investigation. At the time of writing (September 2014), the merged LIS–OTD gridded climatology products cover the period from May 1995 to December 2012. For time series analysis purposes, *Cecil et al*. [2014] recommend the use of the monthly flash rate dataset on a 2.5° × 2.5° grid, but note that the data prior to 1998 (i.e. the period when the LIS data were unavailable) should be treated with caution. Thus the spatial resolution of the LIS-OTD dataset is some 20 times coarser than that of the CIGRE-500 counters, and the length of reliable record (156 months) is noticeably smaller than the number of observations obtained from most of the CIGRE-500 counters considered herein (209 to 298 months, see Table 1).

3. Methods

3.1. Generalized Additive Modeling

We used generalized additive models (GAMs) [*Wood*, 2006] to disentangle the individual and joint effects of seasonality and long-term trend in monthly LFC series. GAMs permit the seasonal cycle to be irregular and not perfectly harmonic, the long-term trend to be nonlinear and non-monotonic, and any two-dimensional dependency between seasonality and long-term trend to be assayed without the imposition of a rigid functional form. In this framework, time

(*month*, *date*) is modeled as a smooth function of seasonality (*month* = 1, …, 12) and *date* = *year* + *month*/12 – 1/24 where *year* denotes the calendar year (1988, for example). The fraction 1/24 corresponds to half a month and ensures that calculated values of *date* correspond to the mid-points of the respective values of *month*. This ensures, for example, that observations and model predictions are properly aligned in time series plots. For an individual monthly LFC series $Y = Y_1, \ldots, Y_n$, the GAMs considered herein specify a distribution for Y_i with mean μ_i, linked to one or more vectors of covariates via an equation of one of the following forms

$$g(\mu_i) = \beta_0 + f_1(month_i) \tag{1}$$

$$g(\mu_i) = \beta_0 + f_1(month_i) + f_2(date_i) \tag{2}$$

$$g(\mu_i) = \beta_0 + f_3(month_i, date_i) \tag{3}$$

where: $\mu_i \equiv E(Y_i)$ in which $E(\cdot)$ is the expectation operator; $g(\cdot)$ is some monotonic function known as the link function; β_0 is the intercept term; and $f_1(\cdot)$, $f_2(\cdot)$ and $f_3(\cdot)$ are centered smooth functions of the covariates (i.e. they are constrained to sum to zero over the data). The smooth (potentially non-monotonic) functions are estimated using penalized maximum likelihood, where a penalty term is added to the usual log-likelihood criterion in order to avoid overfitting the data with functions that are too 'wiggly'. The size of the penalty is controlled via a smoothing parameter that is chosen using leave-one-out cross-validation; larger values of the smoothing parameter produce smoother estimates. Discussion of the choice of distribution and link function is deferred until later in the paper. Throughout the

study, we used a range of diagnostics to check the fit of our GAMs; for details of these diagnostics, see *Wood* [2006, Chap. 5].

The model defined by (1) presumes that the distribution of observed LFCs is dependent on the seasonal cycle alone. Model (2) is the traditional long-term trend plus seasonality model where the dependence on the seasonal cycle is assumed to remain unchanged throughout the period of observation. Model (3) is a bivariate model that allows the seasonal cycle to change through time along with the overall mean. All of these models were fitted using the mgcv package in the R programming environment [*Wood*, 2006; *R Development Core Team*, 2014]. This package represents the smooth functions in the models using flexible collections of spline bases, and also reports the estimated degrees of freedom EDF (or effective number of parameters) as a measure of model complexity. We used cyclic penalized cubic regression splines for $f_1(\cdot)$ as they connect the beginning and end points of the seasonal cycle, penalized cubic regression splines for $f_2(\cdot)$, and tensor product smooths for $f_3(\cdot)$. The tensor product smooths use bivariate splines constructed from the individual basis functions for each variable, analogous to the technique described in *Chandler* [2005, Sec. 4.3]. Further details can be found in *Wood* [2006].

3.2. Inference

A formal comparison of models (1) to (3) can be undertaken by noting that Model (1) is a special case of Model (2) in which $f_2(\cdot)$ is set to zero, and Model (2) is itself a special case of Model (3). One could also contemplate a "null" model containing only an intercept. If the models are successively fitted in order of decreasing complexity, the (penalized) log-likelihood should decrease at each stage providing the amount of smoothing is not allowed to decrease (a smaller smoothing parameter could allow a simpler model to fit the data better by being more wiggly). The magnitudes of these log-likelihood decreases provide a means of

comparing the model fits by testing, for each successive pair of models, the null hypothesis that the data conform to the simpler of them. A comparison of Models (1) and (2), for example, is a formal test of the hypothesis that the series contains no time trend; a comparison of Models (2) and (3) is a test of the hypothesis that the seasonal cycle is constant through time. The procedure, which also takes account of differences in model complexity as measured by the EDF values for the competing models, is known as the analysis of deviance. In the work reported below, Model (3) was fitted first, followed by Model (2) and then Model (1). At each stage, the estimated smoothing parameter for the more complex model was used as a lower bound in a search for the optimal smoothing parameter in the simpler model.

In the mgcv library used for our modelling, the calculation of *p*-values for deviance tests is based on large sample approximations that are most accurate for GAMs based on normal distributions, with an identity link function $g(\mu_i) = E(Y_i)$ and when the smoothing parameter is fixed rather than being estimated from the data. In the present context, the LFC series have highly non-normal distributions; moreover, in the work reported below we used a log link function to ensure that the estimated means $\{\mu_i\}$ of the LFC distributions are all positive, and we used cross-validation to select smoothing parameters. As a check on the accuracy of the approximate *p*-values for testing Model (1) against Model (2) therefore, we compared them with the *p*-values obtained from a permutation test which is a nonparametric approach for computing the sampling distribution for any test statistic [*Faraway*, 2005, Sec. 3.3]. In this study it involves copying the dataset many times and then, for each copy, randomly shuffling the observed values of *date* (without replacement), refitting Model (2) and calculating the deviance statistic (note that the fit for Model (1) is always the same as for the original dataset because the *month* variable is not shuffled). The result is a large number of simulated deviance statistics for datasets in which there is no time trend because the dates have been randomly permuted, but that are otherwise identical to the observed data. A *p*-value for the

test can then be calculated as the proportion of these that exceed the observed deviance statistic. We used 10000 realizations for each test so that the resulting *p*-value is accurate to two decimal places [*Chandler and Scott,* 2011, Sec. 3.6]. A procedure for carrying out permutation tests for the Model (3) versus Model (2) comparisons is less clear as it would require shuffling the data in such a way that *date* and *month* are in the right order taken individually but the combination of the two is not. This was not attempted in this study.

The highly skewed and overdispersed nature of the LFC count distributions suggests that a negative binomial distribution may be an appropriate choice within the GAM framework (section 4). A further complication here is that the negative binomial family requires the estimation of a shape parameter that must be held fixed for the approximate *p*-values to be valid [*Venables and Ripley,* 2002, pp. 207]. To identify an appropriate choice of shape parameter at each site, initial estimates were obtained separately for each of Models (1), (2) and (3), and then the models were refitted with the shape parameter fixed at the mean of these three initial estimates. Additional tests were carried out to assess whether the findings were robust to plausible alternative parameter values.

4. Results

When studying data consisting of counts, it is natural to consider models based on the Poisson distribution in the first instance. However, time series plots of the monthly LFC data revealed that their means are much smaller than their variances (see, e.g., Figure 2a). For each CIGRE-500 site, Table 1 lists the sample mean of the monthly LFC data ($\overline{Y}$) and the dispersion index $D = s_Y^2 / \overline{Y}$ where s_Y^2 denotes the sample variance. The index values lie in the interval $279 \le D \le 3730$ which means that the data are highly overdispersed relative to the Poisson distribution $(D = 1)$. There are two standard options for use in such situations: the first is a direct adjustment of the Poisson model to incorporate an adjustment for the

overdispersion (this is usually called ‘quasi-Poisson’ modelling – see, e.g., *Faraway* [2006, Sec. 7.4]) and the second is the use of a negative binomial distribution in place of the Poisson. A key difference between these two choices is the implied relationship between the mean and variance of the LFCs in different, roughly homogeneous, subsets of the data: a quasi-Poisson model implies that the variance is proportional to the mean, whereas a negative binomial implies that the variance is a quadratic function of the mean. By calculating, for example, the mean and variance of the counts within each calendar month therefore, it is possible to determine which of the two approaches is more suitable for the data. For 13 out of the 19 LFC series, there is a distinct quadratic relationship between mean and variance (see, e.g., Figure 2b), suggesting the use of a negative binomial distribution. The evidence of curvature was less compelling in the corresponding plots for Meekatharra, Port Headland, Tennant Creek and Ceduna. Although this suggests that quasi-Poisson regression may be justified for these sites, it does not preclude the use of a negative binomial model. For consistency across sites therefore, we used the negative binomial distribution with the log link function to model each LFC series.

Specimen results are given in Figures 2c and 2d, which display the estimated smooth functions of *date* and *month* in Model (2) for the Melbourne LFC series. The plots include so-called variability bands indicating the size of two standard errors above and below the estimated functions. Visual inspection suggests that a long-term trend is present in this instance, which is linear on the logarithmic scale of the link function; and that the seasonal cycle is particularly strong. Note that the apparent linearity of the fitted trend is determined by the data via the estimated smoothing parameter; in this case, the width of the variability band shrinks to zero in the centre of the plot because the smooth terms in GAMs are constrained to sum to zero as described above.

< Insert Figure 2 about here >

Figure 3 displays a panel of plots of the estimated smooth functions of *date* in Model (2). The plots are configured to roughly represent the geographical locations of the CIGRE-500 network sensors. While the estimated smooth functions exhibit a variety of forms (linear, nonlinear, non-monotonic), the strength of evidence against the null hypothesis of no trend appears to be weak for the Port Headland, Darwin, Tennant Creek, Coffs Harbour, Kalgoorlie, and Ceduna series. A cursory visual inspection of the remaining plots suggests there has been a general decline in the remaining LFC series. Comparison of Figures 1 and 3 reveals another interesting feature. There are several examples where the forms of the estimated smooth functions for sites that are close (in relative terms) and within roughly the same latitudinal band are reasonably similar. These include: Port Hedland, Darwin and Tennant Creek; Mount Isa and Townsville; Geraldton, Meekatharra, Moora, Perth, Kalgoorlie and Ceduna; Cobar and Coffs Harbour; Albany and Eucla; and Ballarat, Melbourne and Whitlands. Notable exceptions: are Ceduna and Port Augusta; and Eucla and Ceduna. Overall, these features suggest inter-regional differences and in many cases temporal changes in regional and local meteorological forcing variables. This will be the subject of future research.

< Insert Figure 3 about here >

Tables 2 and 3 summarize the results of fitting GAMs to the monthly LFC series. Table 2 lists the EDF, the residual deviance and the deviance explained (i.e. the proportion of the variability in the data set that is accounted for by the statistical model) when the shape parameter for the negative binomial distribution is held fixed for each data set. The deviance explained ranges from low to moderate values (4.53 to 58.5% for Model 1, 13 to 72.5% for Model 2 and 14.2 to 74% for Model 3) which reflect the variability in the underlying data (e.g. Figure 2a). Table 3 reports the analysis of deviance results. Entries are *p*-values

indicating the approximate significance of the smooth terms. It is apparent that there is overwhelming evidence against the null hypothesis of no seasonality for all datasets (approximate *p*-values range from $<2 \times 10^{-16}$ to 0.0011). Comparisons of Models (3) and (2) suggest that the additive structure can be safely rejected in eight cases (approximate *p*-value < 0.01). These results indicate the presence of a very strong interaction between *month* and *date* (i.e. that the seasonal cycle has changed over time) for Albany, Meekatharra, Port Hedland, Port Augusta, Mount Isa, Townsville, Melbourne and Whitlands. There is strong evidence against the simpler additive structure for Eucla, Moora, Perth, Darwin and Coffs Harbour (0.01 < approximate *p*-value < 0.05), and little to no evidence for Geraldton, Kalgoorlie, Ceduna, and Ballarat. Comparison of Models (1) and (2) reveals evidence against the hypothesis of no trend in all but six cases (Kalgoorlie, Port Hedland, Darwin, Tennant Creek, Ceduna, and Coffs Harbour).

< Insert Table 2 about here >

Table 3 also lists the *p*-values obtained from the permutation test (shown in parentheses) and those from the analysis of deviance to test the significance of *date* in Model (2). Recalling that the *p*-value from an analysis of deviance is approximate in our modeling context, there is only one noticeable discrepancy between the sets of *p*-values and that is for Townsville (0.08 versus <0.001). For this case, further investigation revealed that the *p*-value for the permutation test was highly sensitive to variations in the mean of the shape parameter ($\bar{\theta} = 0.234$). For example, variations from $\bar{\theta} - 0.1$ to $\bar{\theta} + 0.1$, which encompass the range of the shape parameter estimates for Models (1) to (3), led to *p*-values ranging from about 10^{-4} to 0.64. Nevertheless, the results for the 18 remaining sites indicate that the approximate *p*-values from the analysis of deviance tests are reasonably reliable and the *p*-values for the permutation tests were far less sensitive than that for the Townsville series.

< Insert Table 3 about here >

Figure 4 displays a panel of plots of the logarithms of the modeled means μ for the years 1990, 1999 and 2009, obtained from Model (3). These plots reveal four key features. First, marked changes in the amplitude and shape of the seasonal cycle of the LFC series are evident for Port Hedland, Geraldton, Meekatharra, Cobar, Perth, Port Augusta, Albany, Ballarat, Melbourne and Whitlands. For these series there is a decline in LFCs in the winter months with some exhibiting changes in the timing of the trough in the seasonal cycle (e.g., Port Hedland, Cobar, Perth, Port Augusta, Albany and Whitlands). The change in the seasonal cycle of the Albany series is intriguing in this regard because the 1990 time slice has an apparently anomalous winter maximum which is absent in the time slices for 1999 and 2009. Inspection of the station metadata did not reveal any inhomogeneities in the LFC series and there is very little change in the results if the outlying value of 2372 for June 1991 is excluded from the analysis. We are therefore confident that these results provide a faithful representation of lightning activity at Albany in the early 1990s. It is relevant to note that the changes for sites south of 30° S are consistent with major changes in the large-scale winter circulation of the Southern Hemisphere since the early 1970s. Previous analyses of changes in baroclinic instability on the hemispheric scale have revealed a 20% reduction in the strength of the subtropical jet over Australia, a sizeable warming of the atmosphere south of 30° S and an associated reduction in the likelihood of unstable synoptic disturbances [*Frederiksen and Frederiksen*, 2007, 2011]. This has been associated with increasing mean sea level pressure and declining rainfall in the austral winter. Second, LFCs at Ballarat and Whitlands have decreased over time throughout the year. Although the amplitude of the seasonal cycle for the Ballarat series has increased, the shape has remained reasonably constant. This is consistent with the analysis of deviance results for this site, which suggest that the increased complexity of Model (3) is not warranted here. The time slices of the tensor

product smooth for Whitlands indicate a marked decline in LFCs for spring as well as winter, and a relatively smaller decline in summer and autumn. Third, the changes in the seasonal cycle of the LFC series for Mount Isa and Townsville are somewhat more complex. Visual inspection suggests that their seasonal cycles changed slightly during the first half of the period of record but have returned somewhat to their previous levels during the second half. Fourth, apart from Ballarat and Whitlands the changes in the monthly LFCs for the austral summer have been relatively slight. The more pronounced changes for these two sites need further investigation, particularly in view of the fact that they seem dissimilar from the changes at nearby Melbourne which is consistent with the remaining stations. In the meantime we note that Melbourne is likely to have a different lightning climatology to that of Ballarat and Whitlands due to its greater proximity to the ocean, noting that thunderstorm occurrence can be related to strong temperature gradients that can occur in coastal regions particularly during summer.

< Insert Figure 4 about here >

5. Conclusions

We have analyzed 25 years, approximately, of monthly lightning flash count (LFC) data for 19 sites across mainland Australia. We have used the generalized additive modeling approach to detect and characterize the presence of long-term trends, seasonal cycles and their interactions. Our study has several strengths including: it adds to the small volume of literature on the quantitative analysis of seasonality and long-term trends in lightning activity series; the lengths of the LFC series used are longer than those commonly reported in the lightning literature; the models used do not impose any parametric form on the fitted curve or surface and allow the analyst to discover the appropriate shape for the covariate effects; and it involves a detailed evaluation of changes in the temporal structure of monthly LFC series. Our main findings can be summarized as follows.

1. There are marked long-term trends in 16 out of 19 cases examined. Although the functional forms of these trends vary (linear, nonlinear, non-monotonic) there is a general impression of a decline in LFCs over the common period of record (approximately 1988 to 2012).

2. There are pronounced changes in seasonality over time in 13 cases. This indicates that the seasonal cycle is non-stationary at these sites, and is expressed in terms of changes in amplitude and shape. The locations of the six remaining sites show no spatial organization that would point to common meteorological conditions.

3. For sites south of 30° S there is a distinct reduction in LFCs during the austral winter. This phenomenon is consistent with the reduction in atmospheric instability that has occurred at these latitudes since the mid 1970s [*Frederiksen and Frederiksen*, 2007, 2011].

4. There is a need to identify and understand the local and regional meteorological factors leading to the observed changes in the temporal structure of the LFC series. This will inform our interpretation of the causes of these changes, particular for the Albany case, and the apparent spatial distribution of the LFC trends. This subject will be the focus of future research.

Acknowledgments. The CIGRE-500 lightning flash counter registration data were provided by the Observations and Engineering Branch, Australian Bureau of Meteorology, with assistance from Tamika Tihema from the National Climate Centre, Australia Bureau of Meteorology. Financial support for this work was provided by the Australian Climate Change Science Program, administered by the Department of Environment. We are indebted to Lorraine Bates for her assistance with data checking and coding and Tamika Tihema from the Bureau of Meteorology for assistance with lightning data from Queensland. All data used in this work are available on request from Andrew J. Dowdy (a.dowdy@bom.gov.au). We also

thank three anonymous referees for their thoughtful and constructive comments on the original manuscript.

References

Boccippio, D. J., W. J. Koshak, and R. J. Blakeslee (2002), Performance assessment of the optical transient detector and lightning imaging sensor. Part I: Predicted diurnal variability, *J. Atmos. Oceanic Technol*., *19*, 1318–1332.

Bovalo, C., C. Barthe, and N. Bègue (2012), A lightning climatology of the South-West Indian Ocean, *Nat. Hazards Earth Syst. Sci*., *12*, 2659–2670, doi:10.5194/nhess-12-2659-2012.

Cecil, D. J., D. E. Buechler, and R. J. Blakeslee (2014), Gridded lightning climatology from TRMM-LIS and OTD: Dataset description, *Atmos. Res*., *135–136*, 404–414, doi:10.1016/j.atmosres.2012.06.028.

Chambers, C. R., G. B. Brassington, I. Simmonds, and K. Walsh (2014), Precipitation changes due to the introduction of eddy-resolved sea surface temperatures into simulations of the “Pasha Bulker” Australian east coast low of June 2007. *Meteorol. Atmos. Phys.*, *125*, 1-15, doi:10.1007/s00703-014-0318-4.

Chandler, R. E. (2005), On the use of generalized linear models for interpreting climate variability, *Environmetrics*, *16*, 699-715, doi:10.1002/env.731.

Chandler, R. E., and M. Scott (2011), *Statistical Methods for Trend Detection and Analysis in the Environmental Sciences*, 388 pp., Wiley, Chichester, United Kingdom.

Christian, H. J., R. J. Blakeslee, D. J. Boccippio, W. L. Boeck, D. E. Buechler, K. T. Driscoll, S. J. Goodman, J. M. Hall, W. J. Koshak, D. M. Mach, and M. F. Stewart (2003), Global frequency and distribution of lightning as observed from space by the Optical Transient Detector, *J. Geophys. Res.*, *108*(D1), 4005, doi: 10.1029/2002JD002347.

Chronis, T. G. (2012), Preliminary lightning observations over Greece, *J. Geophys. Res.*, *117*, D03113, doi:10.1029/2011JD017063.

Coates, L., R. Blong, and F. Siciliano (1993), Lightning fatalities in Australia, 1824–1991, *Nat. Hazards*, *8*(3), 217–233.

Collier A. B., R. E. Bürgesser, and E .E. Ávila (2013), Suitable regions for assessing long term trends in lightning activity, *J. Atmos. Sol-Terr. Phys.*, *92*, 100–104, doi:10.1016/j.jastp.2012.10.012.

Cummins, K. L., and M. J. Murphy (2009), An overview of lightning locating systems: History, techniques, and data uses, with an in-depth look at the U.S. NLDN, *IEEE Trans. Electromagn. Compat.*, *51*(3), 49 –518, doi:10.1109/TEMC.2009.2023450.

Curran, E. B., R. L. Holle, and R. E. López (2000), Lightning casualties and damages in the United States from 1959 to 1994, *J. Climate*, *13*(19), 3448–3464, doi:http://dx.doi.org/10.1175/1520-0442(2000)013<3448:LCADIT>2.0.CO;2.

Dai, J., Y. Wang, L. Chen, L. Tao, J. Gu, J. Wang, X. Xu, H. Lin, Y. Gu (2009), A comparison of lightning activity and convective indices over some monsoon-prone areas of China. *Atmos. Res.*, *91*(2-4), 438-452, doi:10.1016/j.atmosres.2008.08.002.

Darden, C.B., D. J. Nadler, B. C. Carcione, R. J. Blakeslee, G. T. Stano, and D. E. Buechler (2010), Utilizing total lightning information to diagnose convective events, *Bull. Amer. Meteor. Soc.*, *91*(2), 167–175, doi:10.1175/2009BAMS2808.1.

Dowdy, A. J., G. A. Mills, B. Timbal, and Y. Wang (2013), Changes in the risk of extratropical cyclones in eastern Australia, *J. Climate*, *26*, 1403-1417, doi:10.1175/JCLI-D-12-00192.1.

Dowdy, A. J., and Y. Kuleshov (2014), Climatology of lightning activity in Australia: spatial and seasonal variability, *Aust. Meteor. Oceanogr. J.*, *64*(2), 103-108.

Elsom, D. M. (1993), Deaths caused by lightning in England and Wales, 1852-1990, *Weather*, *48*(3), 83–90, doi:10.1002/j.1477-8696.1993.tb05846.x.

Faraway, J. J. (2005), *Linear Models with R*, 240 pp., Chapman & Hall/CRC, Boca Raton, Florida.

Faraway, J.J. (2006), *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*, 312 pp., Chapman & Hall/CRC, Boca Raton, Florida.

Frederiksen, J. S., and C. S. Frederiksen (2007), Interdecadal changes in Southern Hemisphere winter storm track modes, *Tellus*, *59A*, 599–617, doi:10.1111/j.1600-0870.2007.00264.x.

Frederiksen, J. S., and C. S. Frederiksen (2011), Twentieth century winter changes in Southern Hemisphere synoptic weather modes, *Adv*. *Meteor*., Article ID 353829, 16pp, doi:10.1155/2011/353829.

Jayaratne, E. R., and Y. Kuleshov (2006), Geographical and seasonal characteristics of the relationship between lightning ground flash density and rainfall within the continent of Australia, *Atmos*. *Res*., *79*(1), 1–14, doi:10.1016/j.atmosres.2005.03.004.

Kuleshov Y., G. de Hoedt, W. Wright, and A. Brewster (2002), Thunderstorm distribution and frequency in Australia, *Aust. Meteor. Mag.*, *51*, 145–154.

Kuleshov, Y., and E. R. Jayaratne (2004), Estimates of lightning ground flash density in Australia and its relationship to thunder-days, *Aust*. *Meteor*. *Mag*., *53*, 189–196.

Kuleshov, Y., Mackerras, D, and M. Darveniza (2009), Spatial distribution and frequency of thunderstorms and lightning in Australia, in *Lightning: Principles, Instruments and Applications*, edited by H. D. Betz, U. Schumann, and P. Laroche, pp. 187–207, Springer, Dordrecht, The Netherlands, doi:10.1007/978-1-4020-9079-0_8.

Mackerras, D. (1991), Occurrence of lightning death and injury, in *Lightning Injuries: Electrical, Medical and Legal Aspects*, edited by C. J. Andrews, M. A. Cooper, M. Darveniza, and D. Mackerras, pp. 39–46, CRC Press, Boca Raton, Florida.

Petersen, W. A., and S. A. Rutledge (1998), On the relationship between cloud-to-ground lightning and convective rainfall, *J. Geophys. Res., 103*(D12), 14025–14040.

R Development Core Team (2014), *R: A Language and Environment for Statistical Computing*, R Found. for Statist. Comput., Vienna, http://www.R-project.org/.

Ranalkar, M. R., and H. S. Chaudhari (2009), Seasonal variation of lightning activity over the Indian subcontinent, *Meteorol. Atmos. Phys.*,*104*, 125–134, doi:10.1007/s00703-009-0026-7.

Schultz, C. J., W.A. Petersen, and L. D. Carey (2011), Lightning and severe weather: A comparison between total and cloud-to-ground lightning trends, *Wea. Forecasting*, *26*(5), 744–755, doi:10.1175/WAF-D-10-05026.1.

Schulz, W., K. Cummins, G. Diendorfer, and M. Dorninger (2005), Cloud-to-ground lightning in Austria: A 10-year study using data from a lightning location system, *J. Geophys Res., 110*, D09101, doi:10.1029/2004JD005332.

Sheridan, S. C., J. F. Griffiths, and R. E. Orville (1997), Warm season cloud-to-ground lightning-precipitation relationships in the south-central United States, *Wea. Forecasting*, *12*(3), 449-458.

Tinmaker, M. I. R., Ali Kaushar, and G. Beig (2009), Relationship between lightning activity over peninsular India and sea surface temperature, *J. Appl. Meteor. Climatol., 49*, 828–835, doi:10.1175/2009JAMC2199.1.

Uman, M. A. (2010), *The Art and Science of Lightning Protection*, 256 pp., Cambridge University Press, Cambridge, United Kingdom.

Venables, W. N., and B. D. Ripley (2002), *Modern Applied Statistics with S*, 4th ed., 497 pp., Springer, New York, N. Y.

Villarini, G., and J. A. Smith (2013), Spatial and temporal variability of cloud-to-ground lightning over the continental U.S. during the period 199 –2010, *Atmos. Res.*, *124*, 137–148, doi:10.1016/j.atmosres.2012.12.017.

Wood, S. N. (2006), *Generalized Additive Models: An Introduction with R*, 392 pp., Chapman and Hall/CRC, Boca Raton, Florida.

Yuan, T., L. A. Remer, H. Bian, J. R. Ziemke, R. Albrecht, K. E. Pickering, L. Oreopoulos, S. J. Goodman, H. Yu, and D. J. Allen (2012), Aerosol indirect effect on tropospheric ozone via lightning, *J. Geophys. Res.*, *117*, D18213, doi:10.1029/2012JD017723.

Table 1. Station and data details for CIGRE-500 lightning flash counters

Site No.	Location	Altitude (m)	Number of Complete Monthly Observations	Mean Monthly LFC	Dispersion Index, *D*
1	Albany	68	280	62.8	539
2	Eucla	93	209	81.9	514
3	Geraldton	33	269	60.7	335
4	Kalgoorlie	365	282	278	770
5	Meekatharra	517	298	358	1870
6	Moora	203	280	81.8	638
7	Perth	15	298	108	565
8	Port Hedland	6	292	300	1630
9	Darwin	30	284	2100	3730
10	Tennant Creek	376	274	592	1690

Table 1. continued

Site No.	Location	Altitude (m)	Number of Observations	Mean Monthly LFC	Dispersion Index, *D*
11	Ceduna	15	282	105	343
12	Port Augusta	7	279	259	1340
13	Mount Isa	380	287	835	2130
14	Townsville	4	255	253	1520
15	Cobar	260	298	380	1740
16	Coffs Harbour	5	274	501	1350
17	Ballarat	435	287	51.8	279
18	Melbourne	113	298	265	600
19	Whitlands	816	290	375	999

Table 2. Summary of generalized additive models for monthly lightning flash count series

		Model 1			Model 2			Model 3		
Site			Residual	Deviance		Residual	Deviance		Residual	Deviance
No.	Location	EDF	Deviance	Explained (%)	EDF	Deviance	Explained (%)	EDF	Deviance	Explained (%)
1	Albany	4.914	392.71	4.53	7.338	327.74	20.3	15.53	283.66	31.0
2	Eucla	4.873	204.13	11.9	6.072	202.93	14.5	10.754	198.25	18.3
3	Geraldton	4.401	264.60	8.87	5.862	263.14	13.0	9.996	259.00	15.1
4	Kalgoorlie	5.518	276.48	32.4	6.418	275.58	32.6	10.357	271.64	33.6
5	Meekatharra	5.803	292.20	36.2	7.262	290.74	38.7	13.08	284.92	42.6
6	Moora	5.041	274.96	22.0	6.882	273.12	23.4	13.32	266.68	27.3
7	Perth	4.309	293.69	11.5	6.055	291.95	13.0	10.925	287.08	16.3
8	Port Hedland	6.448	285.55	49.4	8.087	283.91	49.8	16.71	275.29	54.9
9	Darwin	6.736	277.26	72.2	8.436	275.56	72.5	16.76	267.24	74.0
10	Tennant Creek	6.038	267.96	58.5	6.954	267.05	58.6	11.18	262.82	59.7

Table 2. continued

			Model 1			Model 2			Model 3	
Site			Residual	Deviance		Residual	Deviance		Residual	Deviance
No.	Location	EDF	Deviance	Explained (%)	EDF	Deviance	Explained (%)	EDF	Deviance	Explained (%)
11	Ceduna	5.167	276.83	13.3	6.046	275.95	13.5	8.291	272.71	14.2
12	Port Augusta	5.477	273.52	12.9	8.753	270.25	22.1	20.94	258.06	30.6
13	Mount Isa	5.99	281.01	46.1	11.838	275.16	51.0	40.07	246.93	61.7
14	Townsville	5.56	249.44	31.5	10.268	244.73	39.5	27.63	227.37	50.2
15	Cobar	5.05	292.95	30.0	6.045	291.95	31.0	9.461	288.54	32.5
16	Coffs Harbour	5.289	268.71	37.4	6.377	267.62	37.4	10.211	263.79	39.6
17	Ballarat	4.106	282.89	10.4	5.499	281.50	31.3	8.197	278.80	32.4
18	Melbourne	5.454	292.55	22.3	6.253	291.75	23.3	9.805	288.19	26.3
19	Whitlands	4.53	285.47	11.3	5.5	284.50	29.1	8.231	281.77	31.5

Table 3. Comparisons of fitted models to monthly lightning flash count series

Site		Analysis	Of	Deviance
No.	Location	Model (1) versus Null	Model (2) versus Model (1)	Model (3) versus Model (2)
1	Albany	0.0011	<0.001 (<0.01)	<0.001
2	Eucla	2.8×10^{-7}	0.009 (0.03)	0.046
3	Geraldton	3.6×10^{-7}	<0.001 (<0.01)	0.135
4	Kalgoorlie	$<2 \times 10^{-16}$	0.294 (0.30)	0.306
5	Meekatharra	$<2 \times 10^{-16}$	<0.001 (<0.01)	0.001
6	Moora	$<2 \times 10^{-16}$	0.062 (0.19)	0.043
7	Perth	5.0×10^{-11}	0.033 (0.10)	0.016
8	Port Hedland	$<2 \times 10^{-16}$	0.211 (0.64)	<0.001
9	Darwin	$<2 \times 10^{-16}$	0.136 (0.31)	0.033
10	Tennant Creek	$<2 \times 10^{-16}$	0.314 (0.49)	0.097

Table 3. continued

Site		Analysis	Of	Deviance
No.	Location	Model (1) versus Null	Model (2) versus Model (1)	Model (3) versus Model (2)
11	Ceduna	2.3×10^{-9}	0.306 (0.34)	0.508
12	Port Augusta	2.2×10^{-14}	<0.001 (<0.01)	<0.001
13	Mount Isa	$<2 \times 10^{-16}$	<0.001 (<0.01)	<0.001
14	Townsville	$<2 \times 10^{-16}$	<0.001 (0.08)	<0.001
15	Cobar	$<2 \times 10^{-16}$	0.018 (0.03)	0.062
16	Coffs Harbour	$<2 \times 10^{-16}$	0.867 (0.97)	0.020
17	Ballarat	9.4×10^{-9}	<0.001 (<0.01)	0.248
18	Melbourne	$<2 \times 10^{-16}$	0.021 (0.03)	0.005
19	Whitlands	8.5×10^{-13}	<0.001 (<0.01)	0.007

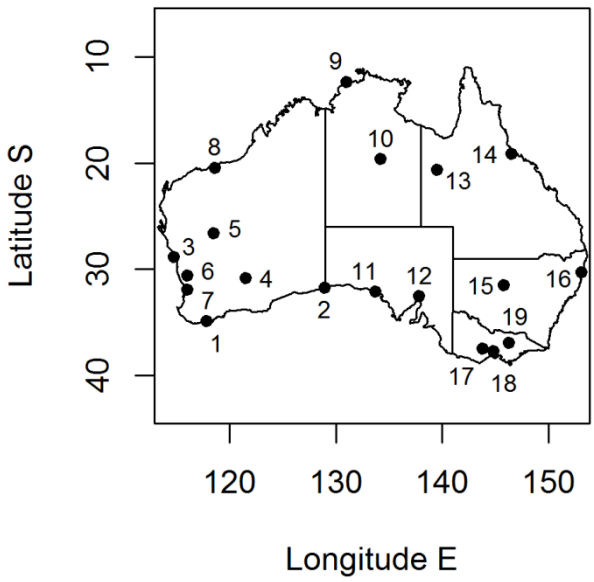

Figure 1. Locations of CIGRE-500 lightning flash counters. (Key to numerals is given in Table 1).

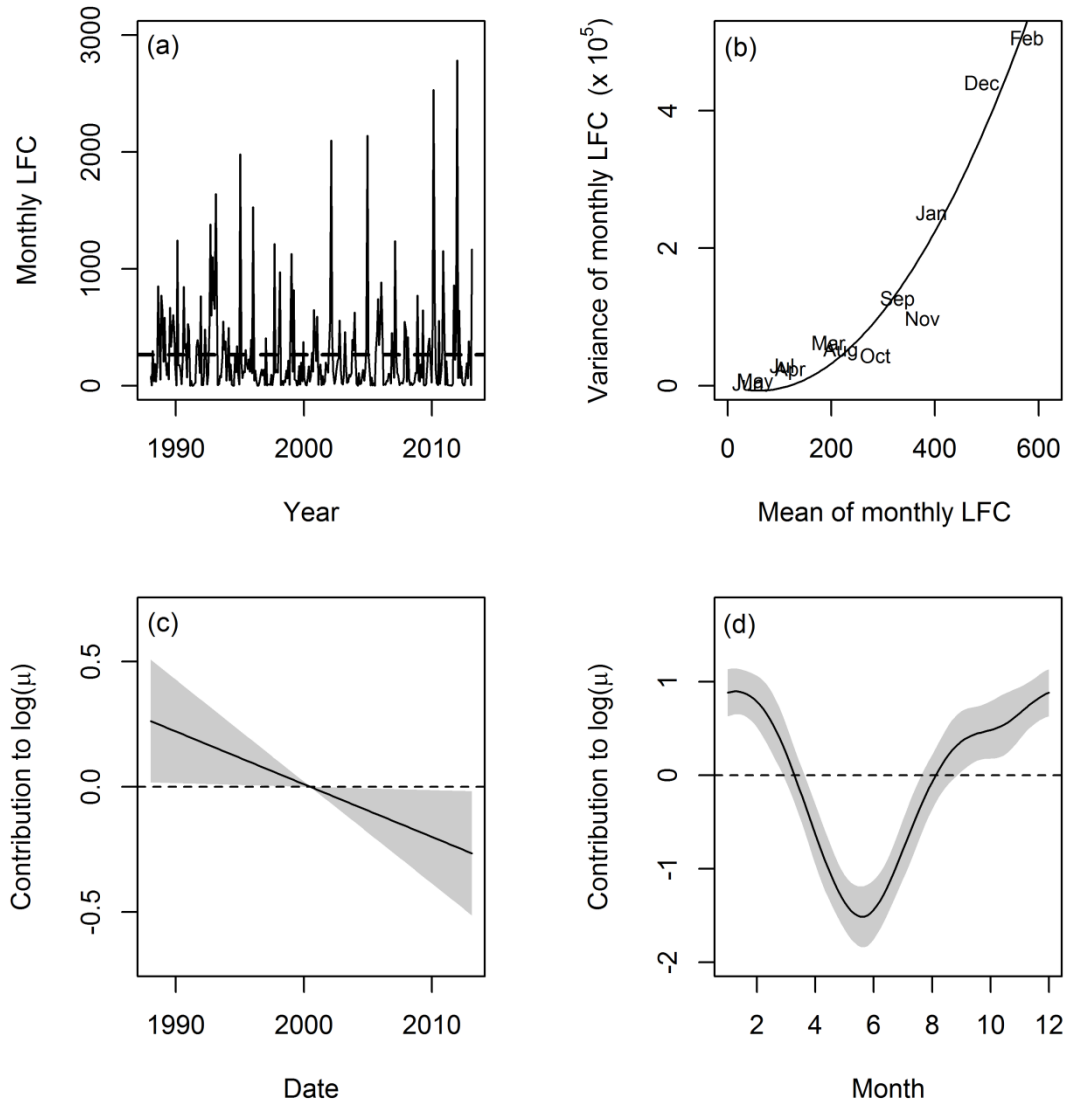

Figure 2. Sample of results from analysis of the Melbourne monthly lightning flash count (LFC) data (a) raw time series (dashed horizontal line depicts the mean) (b) variance-mean relationship by month (solid curve is a quadratic regression fit) (c) contribution to the log link function $[g(\mu) = \log(\mu)]$ by penalized cubic regression spline fit to LFC versus *date* data for Model (2) with variability band (gray shading) (d) contribution to the log link function $[g(\mu) = \log(\mu)]$ by penalized cyclic cubic regression spline fit to LFC versus *month* data for Model (2) with variability band (gray shading).

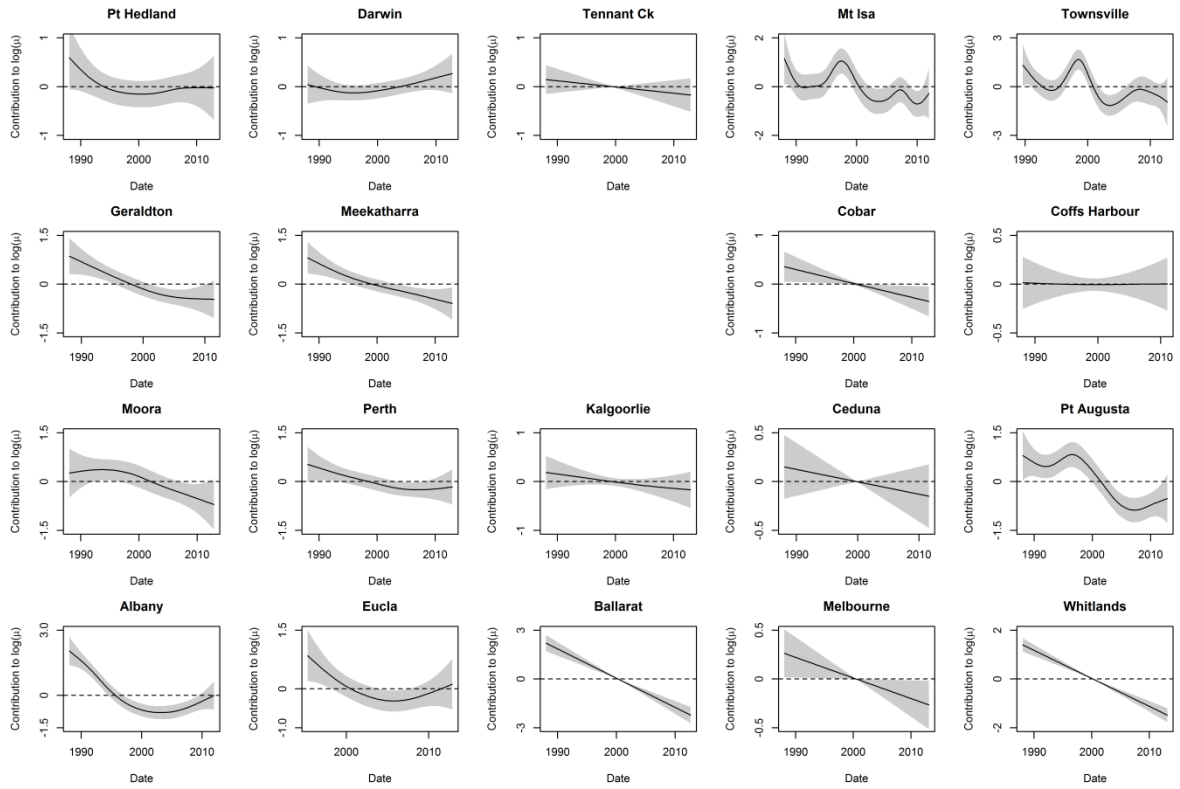

Figure 3. Contributions to log link functions $[g(\mu) = \log(\mu)]$ by penalized cubic regression spline fits to LFC versus *date* data for Model (2) with variability bands (gray shading).

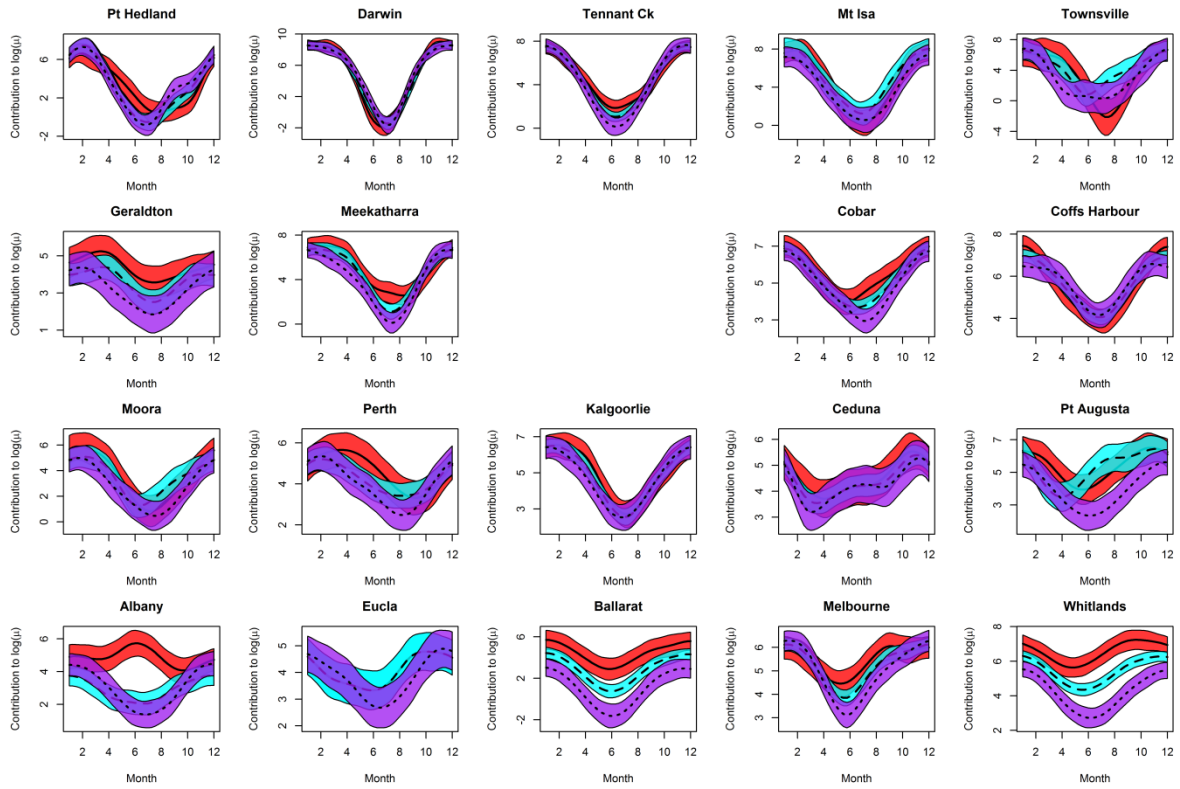

Figure 4. Logarithms of modeled means μ for the years 1990, 1999 and 2009, obtained from Model (3). Colored bands depict variability bands: red = 1990; cyan = 1999; purple = 2009.