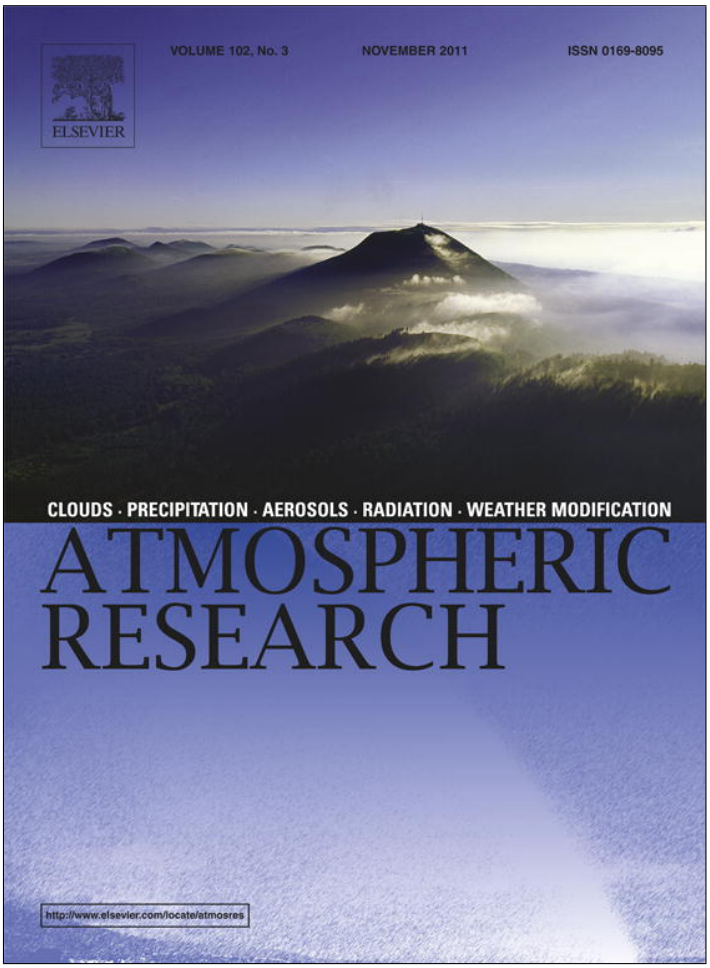

Atmospheric Research 102 (2011) 243–262

Contents lists available at ScienceDirect

Atmospheric Research

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

Bayesian analysis for extreme climatic events: A review

Pao-Shin Chu [a,*], Xin Zhao [b]

[a] Department of Meteorology, School of Ocean and Earth Science and Technology, University of Hawaii, Honolulu, Hawaii 96822, United States
[b] Sanjole Inc., Honolulu, Hawaii 96822, United States

ARTICLE INFO

Article history:
Received 8 October 2010
Received in revised form 4 July 2011
Accepted 11 July 2011

Keywords:
Bayesian analysis
Extreme events
Change-point
Tropical cyclone path clustering
Tropical cyclone prediction

ABSTRACT

This article reviews Bayesian analysis methods applied to extreme climatic data. We particularly focus on applications to three different problems related to extreme climatic events including detection of abrupt regime shifts, clustering tropical cyclone tracks, and statistical forecasting for seasonal tropical cyclone activity. For identifying potential change points in an extreme event count series, a hierarchical Bayesian framework involving three layers – data, parameter, and hypothesis – is formulated to demonstrate the posterior probability of the shifts throughout the time. For the data layer, a Poisson process with a gamma distributed rate is presumed. For the hypothesis layer, multiple candidate hypotheses with different change-points are considered. To calculate the posterior probability for each hypothesis and its associated parameters we developed an exact analytical formula, a Markov Chain Monte Carlo (MCMC) algorithm, and a more sophisticated reversible jump Markov Chain Monte Carlo (RJMCMC) algorithm. The algorithms are applied to several rare event series: the annual tropical cyclone or typhoon counts over the central, eastern, and western North Pacific; the annual extremely heavy rainfall event counts at Manoa, Hawaii; and the annual heat wave frequency in France.

Using an Expectation-Maximization (EM) algorithm, a Bayesian clustering method built on a mixture Gaussian model is applied to objectively classify historical, spaghetti-like tropical cyclone tracks (1945–2007) over the western North Pacific and the South China Sea into eight distinct track types. A regression based approach to forecasting seasonal tropical cyclone frequency in a region is developed. Specifically, by adopting large-scale environmental conditions prior to the tropical cyclone season, a Poisson regression model is built for predicting seasonal tropical cyclone counts, and a probit regression model is alternatively developed toward a binary classification problem. With a non-informative prior assumption for the model parameters, a Bayesian inference for the Poisson regression model and the probit regression model are derived in parallel. A Gibbs sampler is further designed to integrate the posterior predictive distribution. The resulting Bayesian Poisson regression algorithm is applied to predicting the seasonal tropical cyclone activity.

Contents

* Corresponding author.
E-mail address: chu@hawaii.edu (P.-S. Chu).

0169-8095/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.atmosres.2011.07.001

1. Introduction

1.1. Concept of Bayesian inference

In principle, Bayesian analysis is grounded on a probabilistic generative model of a process. With the generative model, Bayes' theorem provides an approach to inferring one or more parameters in a process from the observed data, where the parameters are supposed to characterize the process of interest. In the Bayesian viewpoint, probability can be used to quantify degrees of belief of inference with given assumptions. Thus, Bayesian inference deals with uncertainty of unknown parameters or hypotheses of interest in probabilistic forms. Under a Bayesian framework, the unknown quantities are modeled as random variables, instead of constants or fixed values. This feature fits well with climate research because climate should not be considered stationary but rather as something that is always changing. When new information is obtained, prior knowledge about the unknown quantities of interest is revised accordingly. In this regard, Bayes' theorem provides a formal mechanism to revise or update prior beliefs in light of new data to yield posterior probability statements about the unknown parameters or hypotheses.

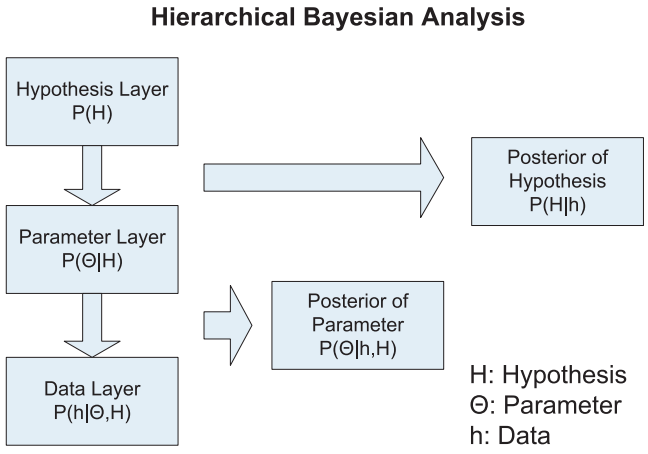

Fig. 1. A 3-layer hierarchical Bayesian analysis model.

The number of inference problems that can be tracked by Bayesian analysis is enormous and across a lot of research fields (e.g., Berger, 1985; MacKay, 2003; Trotta, 2008). The general paradigm of a Bayesian inference analysis can be sketched as the hierarchical flow chart displayed in Fig. 1. On the top of the Bayesian network is the hypothesis or model layer, which defines a hypothesis or model with its associated parameter set. Presumably, the observed data is sampled from this generative model. Before observing the data, there is some subjective belief of the hypothesis or model along with its associated parameter set, which is termed as "prior". Through the Bayes' theorem, one can update the degree of belief of a hypothesis and its relative parameters, yielding the "posterior" probability of the hypothesis and parameter set of interest. Assuming a hypothesis H is given and we denote the observed data by $\mathbf{h}$, the Bayesian formula to infer the parameter set $\boldsymbol{\theta}$ is given by:

$$P(\boldsymbol{\theta}|\mathbf{h},H) = \frac{P(\mathbf{h}|\boldsymbol{\theta},H)P(\boldsymbol{\theta}|H)}{P(\mathbf{h}|H)} = \frac{P(\mathbf{h}|\boldsymbol{\theta},H)P(\boldsymbol{\theta}|H)}{\int_{\boldsymbol{\theta}} P(\mathbf{h}|\boldsymbol{\theta},H)P(\boldsymbol{\theta}|H)d\boldsymbol{\theta}} \propto P(\mathbf{h}|\boldsymbol{\theta},H)P(\boldsymbol{\theta}|H). \quad (1)$$

If the hypothesis or model H is unanimously accepted, Eq. (1) provides the full solution for the inference problem. Note that, as the denominator $P(\mathbf{h}|H)$ does not contain any information on parameter $\boldsymbol{\theta}$, the likelihood term $P(\mathbf{h}|\boldsymbol{\theta},H)$ in Eq. (1) conveys all the "new" information obtained from the data.

Table 1
Raftery's guideline for interpreting the Bayes factor.

2lnB	Evidence for Bayesian model
0–2	Not worth more than a bare mention
2–6	Positive
6–10	Strong

In many real applications, such as a change-point or a signal detection problem, the choice of hypothesis (or model) may not be unique. The posterior probability of hypothesis is thus needed, and the Bayesian inference for the hypothesis layer is governed by

$$P(H|\mathbf{h}) = \frac{P(H)P(\mathbf{h}|H)}{P(\mathbf{h})} \propto P(H)\int_{\boldsymbol{\theta}} P(\mathbf{h}|\boldsymbol{\theta},H)P(\boldsymbol{\theta}|H)d\boldsymbol{\theta} \quad (2)$$

where the "new" information for a hypothesis or model yielded from the observed data is represented by $P(\mathbf{h}|H) = \int_{\boldsymbol{\theta}} P(\mathbf{h}|\boldsymbol{\theta},H)P(\boldsymbol{\theta}|H)d\boldsymbol{\theta}$, which is often termed as "evidence". Generally speaking, in a model or hypothesis selection problem, with increasing hypothesis or model complexity (i.e., with introducing more model parameters), the likelihood term $P(\mathbf{h}|\boldsymbol{\theta},H)$ increases, whereas the density of model parameter prior $P(\boldsymbol{\theta}|H)$ becomes thinner, which is referred to as the "Occam's razor" from a Bayesian perspective (MacKay, 2003). This tradeoff in evidence between likelihood and prior hence lays the core concept for a hypothesis or model selection problem in the Bayesian analysis. In many hypothesis selection problems, in order to compare a hypothesis H_1 against its complementary hypothesis H_0, "Bayes factor" is often referred:

$$B = \left(\frac{P(H_1|\mathbf{h})}{P(H_0|\mathbf{h})}\right)\Big/\left(\frac{P(H_1)}{P(H_0)}\right) = \frac{P(\mathbf{h}|H_1)}{P(\mathbf{h}|H_0)}. \quad (3)$$

Bayes factor can be seen as the likelihood ratio of the two hypotheses or models. In essence, it is a measure of whether the observation data $\mathbf{h}$ have increased or decreased the odds on the hypothesis H_1 relative to the hypothesis H_0. Raftery (1996) suggested a conservative guideline to interpret Bayes factor (Table 1), which favorably mitigates the effects of hypothesis prior bias.

In general, a generative model and prior is usually presumed for a particular problem. In the context of modeling climate data, for instance, a popular choice of generative model for continuous real data is normal distribution. Gamma distribution or log-normal distribution is commonly chosen for modeling non-negative data. Regarding the modeling of extreme climate events, the Poisson distribution is widely used for quantifying their occurrence counts. In order to define occurrence of an extreme event, a sufficiently high threshold value should be provided, such that the occurrence of any extreme event is independent of each other. Intensity wise, for the peak over threshold (POT) data, an exponential distribution is often selected when sample size is small or moderate, with which the associated annual or seasonal maxima shall follow a Gumbel distribution. If sample size is sufficiently large, one can choose the Generalized Pareto distribution as the POT model, and the block maxima shall follow the Generalized Extreme Value (GEV) distribution (e.g., Lang et al., 1999).

Besides the generic nature of choice of generative data model, a target learning problem can be either non-supervised, such as a clustering problem, or supervised, such as a classification or regression problem. Hence, although the Bayes' formula in Eqs. (1) and/or (2) is straightforward, it's worth emphasizing the fact that very few integrating problems embedded in a Bayesian inference analysis have analytical solutions. For this reason, researchers have developed many approximation techniques toward solving a Bayesian inference problem, which can be mainly categorized as "deterministic approximations" or "Monte Carlo methods". The deterministic approximations include such as maximum likelihood estimation and Laplace's approximation. In comparison, random numbers play an integral part in Monte Carlo methods, the main stream of which includes the Markov chain Monte Carlo (MCMC) method to integrate Eq. (1) and using the reversible jump Markov chain Monte Carlo (RJMCMC) method to integrate Eq. (2). The MCMC method is an efficient algorithm for calculating Bayesian inference and provides an unbiased estimate for the posterior expectation of a parameter of interest (e.g., Lavielle and Labarbier, 2001; Gelman et al., 2004). Essentially, a general Bayesian analysis involves calculating the posterior expectation $E[a|\mathbf{h}] = \int_{\boldsymbol{\theta}} a(\boldsymbol{\theta})P(\boldsymbol{\theta}|\mathbf{h})d\boldsymbol{\theta}$, where $P(\boldsymbol{\theta}|\mathbf{h})$ is the posterior probability of the model parameters defined in Eq. (1) and $a(\boldsymbol{\theta})$ can be any function of the parameter $\boldsymbol{\theta}$ (such as the 90th percentile value of $\boldsymbol{\theta}$). With the MCMC method, an alternative way to numerically calculate this expectation is by $E[a|\mathbf{h}] \approx \frac{1}{N}\sum_{i=1}^{N} a\left(\boldsymbol{\theta}^{[i]}\right)$, where N is a sufficiently large number and values $\boldsymbol{\theta}^{[1]}, \boldsymbol{\theta}^{[2]}, \ldots, \boldsymbol{\theta}^{[N]}$ are sampled from a Markov chain (after convergence) having $P(\boldsymbol{\theta}|\mathbf{h})$ as its stationary distribution (e.g., Neal, 1996; Ripley, 1987).

1.2. Bayesian analysis applied to climate data

There has been a wealth of Bayesian climate papers published during the past 25 years or so. Epstein (1985) was the first one who wrote a comprehensive book introducing the Bayesian paradigm and its applications to climate data. Since then, many research works related to Bayesian analysis have been done on subjects such as climate change analysis, climate data uncertainty analysis, climate model combination, climate signal detection, extreme event forecast, etc.

Bayesian analysis can be used to detect climate signals and subsequently to rate climate models based on the predictions. For example, Leroy (1998) showed the analytical solution for the climate signal detection problem under a Gaussian generative model with Gaussian prior, and the method was used to detect a signal introduced by the solar circle in surface temperature. For the change-point problem, using the normal distribution for energy inflow sequence, Perreaulta et al. (2000a) developed a Bayesian framework for detecting abrupt shift in mean level or variability of hydrometeorological time series. They further generalized the framework via incorporating the hypothesis selection into the Bayesian analysis (Perreaulta et al., 2000b).

For data uncertainty analysis, Bayes' rule is an efficient way to provide a coherent and rational framework for reducing uncertainties by incorporating diverse information sources such as subject beliefs, historical records, model simulations, and new information. For example, Rosbjerg and Madsen (1996) resorted to the partial duration series (PDS) model for individual rainfall observations and they proposed a MCMC method based Bayesian procedure to estimate the PDS-parameters. Based on this framework, they suggested an approach to adopting regional information to help improve the at-site estimation of extreme point rainfalls and provide more reliable estimates for the non-monitored site rainfalls. In another example in terms of enhancing the reliability of extreme climate event records, by combining the less reliable

historical accounts of hurricanes in the nineteenth century with the more reliable records from the twentieth century, Elsner and Bossak (2001) showed the exact solution for the Bayesian inference with a Poisson process modeling the hurricane occurrence, and demonstrated a probabilistic estimate for the annual U.S. hurricane rates through a simple Bayesian analysis. In the research line of robust analysis, based on a climate model of intermediate complexity, Tomassini et al. (2007) performed a robust Bayesian uncertainty analysis for climate sensitivity and other parameters. Instead of relying on single prior distributions or single likelihood functions, they used a robust analysis that considered sets of prior distributions or likelihood functions so it was possible to construct upper and lower bounds of quantiles of interest. In this research, the authors built an MCMC algorithm to solve the inference problem.

In the context of climate change analysis, Solow (1988) applied a Bayesian method for inferences about climate change based on the two-phase regression model. Rooted on a two-phase abrupt shift model, Chu and Zhao (2004) provided the exact solution for a hierarchical Bayesian change-point analysis to the central North Pacific tropical cyclone series. Elsner et al. (2004) proposed a MCMC approach to solving the inference problem using a similar hierarchical generative model. Bayesian analysis has also been used for detecting changes in phenological data. For example, three different models (constant, linear, and one change point) with Gaussian observation error are employed to the cherry blossom time series (1896–2002) near Frankfurt, Germany, to examine the impacts of global temperature change on regional plant phenology (Dose and Menzel, 2004). The authors gave an analytical solution to this problem and the results from Bayesian model comparison revealed that the one change-point process is most preferred.

More and more interests have been raised in the Bayesian model combination (BMC) problem during the past decade or so. For example, numerical experiments based on atmosphere–ocean general circulation models (AOGCMs) are one of the major tools used in deriving projections for future climate change. Min and Hense (2006) applied a Bayesian decision method to compare observed and simulated surface air temperatures for both single and multimodel ensembles with four forcing scenarios. When tested against varying prior probabilities, the Bayesian decision method was found to be "largely robust." Tebaldi et al. (2004, 2005) found that their Bayesian model was a useful platform for synthesizing uncertainties in regional temperature and precipitation changes from a multimodel ensemble of AOGCM simulations. Bayesian analysis yields posterior probability distributions of all the uncertain quantities of interest. These posterior distributions for regional temperature and precipitation change and for a suite of other parameters provide a wealth of information about AOGCM reliability and temporal correlations. In Tebaldi et al. (2004), the Gaussian distribution was tacitly assumed to represent seasonal precipitation amounts. Because precipitation distributions are bounded by zero on the left and positively skewed, it is perhaps more meaningful to use a gamma distribution for the likelihood functions for representing precipitation variations. In this case, the conjugate prior would also be a gamma distribution.

Berliner and Kim (2008) developed a two stage hierarchical Bayesian model to combine ensembles from multiple climate models to quantify the uncertainties of Northern and Southern Hemispheric monthly averaged surface temperature. By choosing Gaussian distribution as the generative model, the inference for this problem was derived and the authors used the MCMC method to obtain the solution. Kallache et al. (2010) followed a similar approach and recently suggested a hierarchical Bayesian model to combine several general circulation model (GCM) outputs, in which a Kalman filter was integrated to account for the potential time dependence. They also resorted to the MCMC method to solve the inference problem. Jagger and Elsner (2010) applied the Bayesian model averaging (BMA) procedure to develop a statistical consensus model for seasonal hurricane forecasting. In their study, the authors adopted the Poisson distribution for the hurricane counts and the consensus model is essentially built on a generalized linear model (GLM). The authors used the Bayesian information criteria (BIC) to solve the inference integral, which is based on the Laplace approximation method. The authors showed that the proposed BMC approach delivered more accurate forecasts than only using one "best" model.

1.3. Bayesian analysis applied to extreme climate events

Recently, interest in extreme climate events has grown rapidly. Extreme events are commonly perceived as events that depart pronouncedly from mean conditions. Examples of extreme events include hurricanes (or typhoons in the western North Pacific), heavy rainfall and associated floods, summer heat waves, sea level extreme surge, wildfires, freezing spells, and unusually strong wind gusts, to name a few. Katrina in 2005 is a grim example of the devastation a hurricane can bring to the United States. Typhoons are one of the most destructive natural catastrophes that recur frequently in the western North Pacific and on the eastern Asian coasts. Strong winds, torrential rain, mudslides, and coastal storm surges often lead to loss of life and enormous property damage.

Bayesian analysis to extreme climate events is a new frontier in research. Understanding, modeling, and predicting variations and changes of extreme events is a topic of profound societal significance because of their potential to cause severe damage. As the climate is changing, it is of great interest to explore whether there is any noticeable shift in the frequency and intensity of extreme events from past history. In terms of intensity, grounded on a Gumbel distribution model, Lima and Lall (2010) developed a hierarchical Bayesian model to assess nonstationarity in the scaling of annual maximum flood series, aiding in reducing parameter and model uncertainties while analyzing extreme streamflow time series. In a similar work on the subject of extreme intensity shifting analysis, Renard et al. (2006) presented a more comprehensive, GEV distribution model based, Bayesian framework to analyze extreme time series while considering the probabilistic models including "stationary", "step-change", and "linear trend". They suggested using regional prior knowledge of quantiles as prior distribution and applied MCMC methods to solve the Bayesian inferring problem.

In this review article, we shall particularly focus on detecting abrupt shifts in frequency of extreme climate events, as such abrupt shifts are common characteristics of climate systems. Specifically, researchers need to know when the abrupt shift occurred and what is the likelihood of its occurrence? This knowledge is not only important in its own

right but may also be extremely useful to others, such as government agencies, for better long-term planning and mitigations. In addition, many diagnostic studies are rooted on the basis of comparing active and inactive (or positive and negative) phases of climate states conditioned on the prior knowledge of a regime shift (e.g., Chu, 2002; Deser et al., 2004). In the following, we will review three applications related to the identification of abrupt shifts in climate systems — tropical cyclone (TC) series, clustering of TC tracks, and forecasting of seasonal TC activity. Elsner and Bossak (2001) proposed a fundamental Bayesian approach to modeling the U.S. hurricane climate. Subsequently, Chu and Zhao (2004) applied a hierarchical Bayesian change-point analysis to the central North Pacific tropical cyclone series. In the fundamental Bayesian framework, only two layers – a data layer and a parameter layer – were considered for deriving the posterior distribution and obtaining the optimum predictive distribution. In this framework, no change points were assumed. Expanding from this two-layer approach, Chu and Zhao (2004) introduced a new layer called the hypothesis layer (Fig. 1). For this layer, a "no change in the cyclone intensity" and a "single change in the intensity" hypotheses were considered. For the data layer, a Poisson process with gamma distributed intensity is presumed. In this three-layer paradigm, both the data layer and parameter layer are conditional on hypothesis selection. The results of this study indicate that there is a great likelihood of interdecadal variability with a shift on tropical cyclone rates around 1982. Bayesian analysis is also applied to detect shifts in the time series of seasonal typhoon counts in the vicinity of Taiwan (Tu et al., 2009). An abrupt shift occurs in 2000, with a lower annual rate prior to 2000 and a higher rate since 2000.

For a regime shift analysis problem, strictly speaking, the method employed in Chu and Zhao (2004) and Elsner et al. (2004) was only applicable to detecting a single change-point in an extreme event count series. Because of decadal climate variability and abrupt climate change due to external forcings (e.g., volcanic eruptions), multiple shifts in the climate series are not uncommon. Accordingly, Zhao and Chu (2006) extended the two-hypothesis model in Chu and Zhao (2004) to a multi-hypothesis hierarchical model and applied a MCMC approach to detecting shifts in the annual hurricane counts in the eastern North Pacific.

Although the MCMC algorithm was shown to be viable for calculating the posterior probability for a multiple hypothesis model, it suffers from a shortcoming. Because parameter spaces within each hypothesis are typically different from each other, a simulation has to be run independently for each of the candidate hypotheses. If the hypotheses involved have a high dimension, this strategy is not efficient and is computationally prohibitive. Simply put, a standard MCMC approach is not appropriate for a model selection problem because different candidate hypotheses usually do not share the same parameter sets. To overcome this problem, Zhao and Chu (2010) recently introduced a reversible jump Markov Chain Monte Carlo (RJMCMC) algorithm to calculating posterior probability for each hypothesis and its within-hypothesis parameters.

Bayesian analysis has also been applied to objective clustering historical tropical cyclone (TC) path tracks in a basin. For example, based on a mixture Gaussian model, the TCs over the western North Pacific are categorized into a definite number of path patterns (Camargo et al., 2007; Chu et al., 2010a). In the proposed numerical clustering technique therein, the membership probability of a TC track belonging to any of the clusters is virtually the Bayes' posterior probability of each track type, given all model parameter sets. The membership probability can be thought of as a kind of probability of how strongly a typhoon track belongs to a certain cluster. The Expectation-Maximization (EM) algorithm is employed to solve the model parameters. It is worth noting that other clustering methods, such as the fuzzy *c*-mean method (Kim et al., 2011), can also be applied to objectively classify TC tracks over a basin.

Besides for analyzing an unsupervised learning problem, such as making inferences of abrupt shifts in extreme event series and objectively classifying TC tracks into a few patterns, the Bayes' theorem has also been applied in many supervised learning real world applications, which include such examples as building a probabilistic predictive model for seasonal hurricane (or typhoon) rates for the Atlantic, the central North Pacific, the east China Sea, and a region near Fiji and Taiwan (Elsner and Jagger, 2006; Chu and Zhao, 2007; Ho et al., 2009; Chand et al., 2010; Lu et al., 2010). In these studies, researchers attempted to either forecast seasonal TC frequency for an entire ocean basin or for a specific region within a basin.

For the rest of this article we will briefly review for some case studies to illustrate how a Bayesian model is built for a specific problem, how the inference problem is solved, and some important concepts related to Bayesian analysis. We shall first address the issue related to the detection of abrupt shifts in three kinds of extreme events, namely, hurricanes and typhoons, heavy rainfall, and heat waves. We will demonstrate the analytical approach, MCMC approach, and RJMCMC approach to solve the inference in Eq. (2) with a hierarchical Poisson generative model. Subsequently, we will show how the Bayes' theorem can be applied to clustering spaghetti-like shapes of historical TC tracks into several distinct patterns over the western North Pacific with a mixture Gaussian generative model. The last application of Bayesian thinking is to predict the frequency of seasonal tropical cyclones in a region of interest to illuminate the Bayesian regression approach to this problem.

2. Data

Tropical cyclone data for the western North Pacific comes from the Joint Typhoon Warning Center in Honolulu, Hawaii. For the frequency data used in the change-point analysis, the period is 1960–2006 and only super-typhoons are used, which are defined as having maximum sustained 1-min wind speed equal to or greater than 130 kt (66.9 m s^{-1}). For the clustering analysis, the dataset contains measurements of the TC center location in latitude, longitude, one-minute sustained maximum wind speed, and central pressure at 6-h intervals for all TCs in the western North Pacific (WNP) and the South China Sea. Here, tropical cyclones refer to tropical storms and typhoons. Tropical storms are defined as maximum sustained wind speeds between 17.5 and 33 m s^{-1}, and typhoons are defined as wind speeds at least 33 m s^{-1}. For cluster analysis, the data cover the period 1945–2007 and a total of 1621 cases are analyzed.

For the central North Pacific (CNP), the tropical cyclone data come from the National Hurricane Center's best tracks dataset and the Central Pacific Hurricane Center, an entity of the National Weather Service Forecast Office in Honolulu,

Hawaii; the data period is 1966 to 2002. To illustrate the non-MCMC approach for detecting abrupt shifts in TC series (Section 3.2), prior information of TCs before 1966 is needed. Data prior to 1966 in the CNP can be found in Shaw (1981). We also use the annual major hurricane counts over the eastern North Pacific (ENP). Whitney and Hobgood (1997) suggested that reliable hurricane statistics over the ENP began in the early 1970s, when the Dvorak scheme for estimating the intensity of TCs was operationally implemented. Collins and Mason (2000) chose 1972 as the starting year for their investigation on interannual variations of local environmental conditions for TC activity over the ENP. Thus, the data for the ENP from 1972 to 2003 are used.

For heavy rainfall data, historical daily records for Manoa in Oahu, Hawaii, were obtained from the National Climatic Data Center web page and the archives from the Hawaii State Climate Office. The data period for Hawaii extends from 1920 to 2009. To define extreme rainfall events, we chose a specific threshold (i.e., 95th percentile) of precipitation days as heavy events. Here we ignore days of no precipitation or days with precipitation less than 0.01 in. (2.5 mm). For each year, we count the number of events when the threshold is crossed.

For heat waves, we adopt the World Meteorological Organization definition. That is, daily maximum temperature with more than five consecutive days exceeds the maximum temperature normal by 5 °C, where the normal is the period 1961–90. The infamous 2003 heat wave was one of the hottest summers in Europe, especially in France. More than 40,000 Europeans died as a result of the heat wave in 2003. The countries mostly affected by the heat waves included France, Portugal, Spain, Netherlands, Switzerland, United Kingdom, and Germany. Because France bore the major brunt of this wave with 14,802 heat-related deaths, we selected one station (Mont) from France for this study; the data set is from 1949 to 2009. As with the heavy rainfall series, for heat waves, we count the number of events when the threshold is crossed for each year.

3. Bayesian change-point analysis for extreme climate events

3.1. Hierarchical Bayesian model for extreme event count series

Because a change-point problem is essentially a hypothesis selection problem, we adopt a hierarchical Bayesian model in this study. The occurrence of independent, rare events is commonly assumed to follow a Poisson process (e.g., Elsner and Bossak, 2001; Wilks, 2006; Khaliq et al., 2007; Briggs, 2008). Given the rate parameter λ, the probability mass function (PMF) of h counts occurring in T unit seasons is (Epstein, 1985)

$$P(h|\lambda,T)=\exp(-\lambda T)\frac{(\lambda T)^h}{h!},\quad \text{where}\quad h=0,1,2,\ldots \text{ and } \lambda>0, T>0. \tag{4}$$

The Poisson mean is the product of λ and T, so is its variance. Throughout this section all of our cases are annual or seasonal count series; therefore we always set $T=1$ in Eq. (4). In Eq. (4), the rate parameter λ is treated as a random variable to construct a hierarchical framework (e.g., Chu and Zhao, 2004; Elsner and Jagger, 2004). We choose the conjugate prior, the gamma distribution, for rate λ, which is formulated by:

$$P(\lambda;h',T')=\frac{T'^{h'}\lambda^{h'-1}}{\Gamma(h')}e^{-\lambda T'},\lambda>0,h'>0,T'>0 \tag{5}$$

where the gamma function is defined by $\Gamma(x)=\int_0^\infty t^{x-1}e^{-t}dt$. The advantage to choosing a conjugate prior is that, given h counts occurring in T years, if the prior density for λ is gamma distributed with parameters h' and T', the resulting posterior density for λ is also gamma distributed with parameters $h+h'$ and $T+T'$. With regard to Eq. (5), the prior expectation of λ is $E[\lambda]=h'/T'$. It should be noted that under the statistical model defined by Eqs. (4) and (5), the marginal PMF of h counts occurring in T years when the intensity is gamma distributed with prior parameters h' and T' is a negative binomial distribution (Epstein, 1985).

To further define a generic hierarchical model for the change-point problem, we assume that there are K possible candidate hypotheses $\{H_0,H_1,\ldots,H_{K-1}\}$. Under each hypothesis H_k, $k=0,1,\ldots,K-1$, there presumably exists exactly k change-points in this period and we denote the parameter set under hypothesis H_k by $\boldsymbol{\theta}_k$. Note that the annual typhoon counts, $\mathbf{h}=[h_1,h_2,\ldots,h_n]'$, are assumed to be a series of independent random variables. The model of hypothesis H_k, $k=0,1,\ldots,K-1$, is postulated as follows.

Hypothesis H_k: "A k-changes of the rate" of the annual typhoon series:

Assuming that $\boldsymbol{\tau}=[\tau_{k1},\tau_{k2},\ldots,\tau_{kk}]'$ denotes the vector of all k change-points that occurred in the time series $\mathbf{h}$ under hypothesis H_k, where the element $\tau_{k,j}$ represents the first year of the j-th epoch. We have the model:

$$\begin{aligned}
&h_i\sim\text{Poisson}(h_i|\lambda_{k1},1),\text{if } i=\tau_{k0},\tau_{k0}+1,\ldots,\tau_{k1}-1\\
&h_i\sim\text{Poisson}(h_i|\lambda_{k2},1),\text{when } i=\tau_{k1},\tau_{k1}+1,\ldots,\tau_{k2}-1\\
&\ldots\\
&h_i\sim\text{Poisson}\left(h_i|\lambda_{k,k+1},1\right),\text{when } i=\tau_{kk},\tau_{kk}+1,\ldots,\tau_{k,k+1}-1
\end{aligned} \tag{6}$$

where $\tau_{k0}<\tau_{k1}<\tau_{k2}<\ldots<\tau_{kk}<\tau_{k,k+1}$ and

$$\begin{aligned}
&\lambda_{k1}\sim\text{gamma}\left(h'_{k1},T'_{k1}\right)\\
&\lambda_{k2}\sim\text{gamma}\left(h'_{k2},T'_{k2}\right)\\
&\cdots\\
&\lambda_{k,k+1}\sim\text{gamma}\left(h'_{k,k+1},T'_{k,k+1}\right).
\end{aligned}$$

In Eq. (6), $\tau_{k0}=1$ and $\tau_{k,k+1}=n+1$ for all k; and the prior knowledge of the hyper-parameter set $\{h'_{kj},T'_{kj}|j=1,2,\ldots,k+1;$ and $k=0,1,\ldots,K-1\}$ must be given a priori. Note that under hypothesis H_k, there are $k+1$ epochs separated by k change-points τ_{kj}, $j=1,\ldots,k$. The jth change-point is defined as the first year of the $(j+1)$-th epoch. Although the prior for change-point can be any discrete distribution, throughout this study we adopt its non-informative prior form; that is, it is uniformly distributed.

3.2. Exact Bayesian inference under the hypothesis with only one change-point

In this section we will illustrate the exact analytical Bayesian inference under the one change-point hypothesis H_1. Subject to the assumption that the number of occurrence of annual TCs is independent from year to year, and dropping the notations for the prior parameters h'_{11}, T'_{11}, h'_{12}, T'_{12} for the sake of simplicity, with the likelihood function yielded from Eq. (6), the vector form of the likelihood function with a given change-point is $P(\mathbf{h}|\tau,H_1)=\prod_{i=1}^{n}P(h_i|\tau,H_1)$, where $\mathbf{h}=[h_1,h_2,\ldots,h_n]'$ is the vector form of the observation data (Chu and Zhao, 2004). With the Bayes formula, Eq. (1), the posterior distribution function of the change-point τ under the hypothesis H_1 is given by

$$P(\tau|\mathbf{h},H_1)=\frac{P(\mathbf{h}|\tau,H_1)P(\tau|H_1)}{\sum_{\tau=2}^{n}P(\mathbf{h}|\tau,H_1)P(\tau|H_1)}\propto\prod_{i=1}^{n}P(h_i|\tau,H_1),\ \tau=2,\ldots,n. \tag{7}$$

Using Eq. (5) and keeping in mind of the conjugate property of gamma distribution, the conditional posterior distribution (with a given change-point τ) of the intensity before and after the change-point, say λ_1 and λ_2, is also a gamma distribution.

The posterior PDF for intensity λ_{11} and λ_{12} under H_1 hypothesis is

$$f\left(\lambda_{1i}|h'_{1i},T'_{1i},H_1,\mathbf{h}\right)=\sum_{\tau=2}^{n}f\left(\lambda_{1i}|h'_{1i},T'_{1i},\tau,H_1,\mathbf{h}\right)P(\tau|\mathbf{h},H_1),i=1,2. \tag{8}$$

In the Bayesian inference formula, Eq. (8), $f(\lambda_{1i}|h'_{1i},T'_{1i},\tau,H_1,\mathbf{h})$ is the conditional posterior distribution of the intensity given τ, and $P(\tau|\mathbf{h},H_1)$ is calculated by Eq. (7).

3.3. Bayesian inference under the hypothesis H_1 using MCMC method

In Section 3.2, we described the analytical formula for the change-point hypothesis model H_1. Even with this very simple hypothesis model, it can be seen that the analytical inference formulae for change-point parameter τ in Eq. (7) and rate parameters λ_1 and λ_2 in Eq. (8) are not standard and fairly complicated. To solve for the posterior distribution of the hypothesis using Eq. (2) is even harder, implying that the analytical approach to the multiple change-points analysis problem is not viable in practice.

Using a Gibbs sampler, one of the most widely used MCMC methods, as the core, Elsner et al. (2004) applied an algorithm to calculate the Bayesian inference for the change-point model under the H_1 hypothesis defined in Eq. (6). In an earlier study, Elsner et al. (2004) applied a hierarchical hyper-prior model for gamma prior parameters. Later, Zhao and Chu (2006) extended this algorithm to calculating Bayesian inference under a more general multiple change-point hypothesis H_k.

If there are p parameters, $\boldsymbol{\theta}=[\theta_1,\theta_2,\ldots,\theta_p]'$ in the model then presumably direct sampling from the posterior distribution $P(\boldsymbol{\theta}|\mathbf{h})$ will be very difficult. In a Gibbs sampler, a sample is drawn from the conditional distribution for any component of $\boldsymbol{\theta}$, given values for the rest of the other components of $\boldsymbol{\theta}$; this sampling method involves successive sampling from the complete conditional posterior densities $P(\theta_k|\mathbf{h},\theta_1,\ldots,\theta_{k-1},\theta_{k+1},\ldots,\theta_p)$, where k is from 1 to p. Referring to the model under hypothesis H_1 in Eq. (6), there are three model parameters — two rate parameters (λ_1 and λ_2) and one change-point parameter (τ). As an alternative approach to the theoretical formulae given in Eqs. (7) and (8), the algorithm for calculating the Bayesian inference under H_1 is provided as follows.

1. Initialize λ_{11}, λ_{12}, and $\tau=\tau_{11}$ with any allowable value, then proceed.
2. Draw λ_{11} from $\lambda_{11}|\mathbf{h},\tau,\lambda_{12},H_1\sim\text{gamma}\left(h_{11}'+\sum_{i=1}^{\tau-1}h_i,T_{11}'+\tau-1\right)$.
3. Draw λ_{12} from $\lambda_{12}|\mathbf{h},\tau,\lambda_{11},H_1\sim\text{gamma}\left(h_{12}'+\sum_{i=\tau}^{n}h_i,T_{12}'+n-\tau+1\right)$.
4. Draw τ from $\tau|\mathbf{h},\lambda_{11},\lambda_{12},H_1\propto e^{-(\tau-1)(\lambda_{11}-\lambda_{12})}(\lambda_{11}/\lambda_{12})^{\sum_{i=1}^{\tau-1}}$.
5. If the required iteration number is not met, go to step 2 to start a new iteration. (9)

In algorithm (9), the posterior probability of two rate parameters are gamma distributed due to the conjugate property. The detailed derivation for the conditional posterior distribution for the change-point given the Poisson rate before and after the change-point for Step 4 is found in Elsner et al. (2004). The extension of algorithm (9) to multiple change-point hypothesis H_k, $k>1$, is derived in Zhao and Chu (2006). After a proper burn-in period, the output of algorithm (9) within each iteration is equivalently drawn from the joint posterior distribution $P(\lambda_{11},\lambda_{12},\tau|\mathbf{h},h_{11}',T_{11}',h_{12}',T_{12}',H_1)$. One can use these posterior samples for any further analysis under hypothesis H_1.

3.4. Bayesian change-point analysis using RJMCMC method

In Section 3.2 and 3.3, we provide a solution for the Bayesian inference calculation under hypothesis H_1. However, the Bayesian inference for the hypothesis layer, as formulated by Eq. (2), remains untouched. The Bayesian modeling selection problem has been well studied in the literature (i.e., Congdon, 2007). In the context of Monte Carlo methods, via the MCMC output samples from algorithm (9), Zhao and Chu (2006) suggested a Monte Carlo integration approach to calculating the posterior probability for different hypotheses. However, due to the fact that parameter spaces within different hypotheses are typically different from each other, a simulation has to be independently run for each of the candidate hypotheses. Furthermore, if a candidate hypothesis has a large dimension, this strategy is even computationally prohibitive. In essence, a standard MCMC algorithm is not appropriate for a model selection problem because different candidate models or hypotheses usually do not share the same parameter sets.

To overcome this problem, Green (1995) first introduced the reversible jump Markov chain Monte Carlo (RJMCMC) algorithm as a simultaneous integrative approach to deal with the model or hypothesis selection problem. The RJMCMC algorithm is so termed to maintain the detailed balance of an irreducible and aperiodic chain that converges to the correct target measure. With the introduction of an extra pseudo

random variable for each investigated model, the dimension-mismatching for a different model is well contained. Since Green (1995), many others have successfully applied and extended the RJMCMC to a broad range of problems such as variable selection, curve fitting, neural network, etc. (e.g., Godsill, 2001). One of its main real-world applications is on the detection of multiple structure or regime changes in a physical process (e.g., Richardson and Green, 1997; Rotondi, 2002).

In a general setting, if the candidate hypotheses are enumerable and represented by the set $H=\{H_0,H_1,\ldots,H_{K-1}\}$, we denote the joint set of parameters under hypothesis H_k by $\boldsymbol{\theta}_k$, $k=0,1,\ldots K-1$. We further introduce a random vector $\mathbf{u}_k$, $k=0,1,\ldots K-1$, such that for any $k'=0,1,\ldots K-1$, $k'\neq k$, the dimension of $\{\boldsymbol{\theta}_k,\mathbf{u}_k\}$ and $\{\boldsymbol{\theta}_{k'},\mathbf{u}_{k'}\}$ can be matched. We then set $\boldsymbol{\theta}_{k'}$ to be a deterministic function of $\boldsymbol{\theta}_k$ and $\mathbf{u}_k$. For the reversible move, we propose a vector $\mathbf{u}_{k'}$ and set $\boldsymbol{\theta}_k$ to be a deterministic function of $\boldsymbol{\theta}_{k'}$ and $\mathbf{u}_{k'}$. Thus, there must be a bijection between $\{\boldsymbol{\theta}_k,\mathbf{u}_k\}$ and $\{\boldsymbol{\theta}_{k'},\mathbf{u}_{k'}\}$, which is defined by $(\boldsymbol{\theta}_{k'},\mathbf{u}_{k'})=g_{k,k'}(\boldsymbol{\theta}_k,\mathbf{u}_k)$. The iteration for a general RJMCMC algorithm (with given observation dataset $\mathbf{h}$) can be sketched as below (after initialization and assuming that the current accepted hypothesis is H_k).

1. Propose a visit to hypothesis $H_{k'}$, $k'\neq k$, with probability $J(H_k\rightarrow H_{k'})$.
2. Sample $\mathbf{u}_k$ from a proposal density $Q(\mathbf{u}_k|\boldsymbol{\theta}_k,H_k,H_{k'})$.
3. Set $(\boldsymbol{\theta}_{k'},\mathbf{u}_{k'})=g_{k,k'}(\boldsymbol{\theta}_k,\mathbf{u}_k)$.
4. Calculate the odds ratio r and accept $H_{k'}$ as the hypothesis in the next iteration with the probability "$\min(1,r)$", where

$$r=\frac{P(\mathbf{h}|\boldsymbol{\theta}_{k'},H_{k'})P(\boldsymbol{\theta}_{k'}|H_{k'})P(H_{k'})J(H_{k'}\rightarrow H_k)Q(\mathbf{u}_{k'}|\boldsymbol{\theta}_{k'},H_{k'},H_k)}{P(\mathbf{h}|\boldsymbol{\theta}_k,H_k)P(\boldsymbol{\theta}_k|H_k)P(H_k)J(H_k\rightarrow H_{k'})Q(\mathbf{u}_k|\boldsymbol{\theta}_k,H_k,H_{k'})}\left|\frac{\partial g_{k,k'}(\boldsymbol{\theta}_k,\mathbf{u}_k)}{\partial(\boldsymbol{\theta}_k,\mathbf{u}_k)}\right|$$

If $H_{k'}$ is rejected, we maintain the current hypothesis H_k.

5. Return to the step 1 until the required number of iterations is reached.

(10)

With the general RJMCMC algorithm (10), after a burn-in period the number of times that hypothesis H_k is accepted in the simulation (after Step 4) divided by the total number of iterations gives a good estimation for the posterior probability of H_k, $P(H_k|\mathbf{h})$. Also, the samples from each iteration within the hypothesis H_k will be equivalently drawn from the posterior joint PDF $P(\boldsymbol{\theta}_k|\mathbf{h},H_k)$.

In algorithm (10), the choice of an appropriate proposal probability function $Q(\mathbf{u}_k|\boldsymbol{\theta}_k,H_k,H_{k'})$ is critical to the efficiency of the algorithm, which is generally preferred to take a closer form to the posterior probability function of the proposed parameter set under the new hypothesis. For this purpose, Zhao and Chu (2010) proposed that the hypothesis space defined in the change-point model (Eq. 6) followed a "nested structure." Specifically, referring to Step 2 of the general RJMCMC algorithm (10), to move from the hypothesis H_k to the hypothesis H_{k+1}, all change-points and most rates under H_k are kept unchanged and only one new change-point is introduced, for example, τ (without losing generality, assuming $\tau_{kj}<\tau<\tau_{k,j+1}$, $0\le j\le k$). The researcher then proposes two new rate parameters for the period $[\tau_{kj},\tau-1]$ and $[\tau,\tau_{k,j+1}-1]$, respectively, under the hypothesis H_{k+1} and discards the old rate parameter $\lambda_{k,j+1}$. Conversely, to move from hypothesis H_{k+1} to hypothesis H_k, one change-point is randomly selected under H_{k+1}, say $\tau_{k+1,j}$, $1\le j\le k+1$, and then the two phases adjacent to this change-point are merged, while introducing a new rate $\lambda_{k,j}$ for the new phase $[\tau_{k+1,j-1},\tau_{k+1,j+1}-1]$ under the hypothesis H_k. Simultaneously, the two old rate parameters and the referred change-point parameter are discarded. With this nested structure model, any two adjacent hypotheses in model (6) share most of the parameters.

Adopting this nested model, Zhao and Chu (2010) developed a RJMCMC algorithm to analyze the inference for detecting multiple abrupt shifts in an extreme climate event count series. The detailed derivation and explicit formulae for the algorithm are covered in that paper. Specifically, the algorithm uses the same structure of the general RJMCMC algorithm given in algorithm (10). For the bijection function in Step 3, it is the identity mapping. For the hypothesis transition function $J(H_k\rightarrow H_{k'})$ in Step 1, the allowable moves for the current hypothesis are constrained only to its adjacent hypotheses. That is, there are only two possible moves in Step 2: from H_k to H_{k+1}, which can be viewed as the "birth" of a new change-point; and from H_{k+1} to H_k, which can be viewed as the "death" of an existent change-point. The proposal density function Q in Step 2 of algorithm (10), as defined in Zhao and Chu (2010), was motivated by and derived from the conditional posterior distribution given in the relative MCMC algorithm (9). With the proposal density function Q and hypothesis transition function, the acceptance ratio in Step 4 of algorithm (10) is available for both the "birth" move $(H_k\rightarrow H_{k+1})$ and the "death" move $(H_{k+1}\rightarrow H_k)$. Thus this RJMCMC based algorithm dedicated to detecting multiple abrupt regime shifts for an extreme event count series is completed.

3.5. Prior specification

There must always be a prior assumption for any Bayesian inference. For the prior distribution of the parameter set $\boldsymbol{\theta}_k$ under hypothesis H_k defined in Eq. (6), those change-point parameters can be assumed to be uniformly distributed. It is however not appropriate to use the non-informative prior for the Poisson rate parameters. In a hypothesis selection problem, a flat non-informative prior usually does not work well because it would almost always favor the simplest hypothesis due to the extremely huge normalization term for each rate parameter. As discussed in detail by MacKay (2003), to fit a set of data using two different models the posterior probability of the complicated model is penalized by a stronger Occam's factor, which conceptually is a ratio of its parameters' posterior and prior widths. In Eq. (6), a non-informative prior for the rate (λ) will lead to infinite prior width or, in other words, an infinitely small Occam's factor. In order to make sound model or hypothesis selection, there needs a reasonable informative prior for model parameter. For this purpose, Zhao and Chu (2010) simplified the model (6) by assuming that all hyper parameters h'_{kj} and T'_{kj} are equal to constant h' and T', respectively. They then proposed a procedure to obtain the hyper-prior parameters h' and T', which is briefly described below.

With time series $\mathbf{h}=[h_1,h_2,\ldots,h_n]'$, the L independent iterations are run first. Within the j-th iteration, $1\le j\le L$, two different points from 1 to n, i.e., k_0 and k_1 ($k_0<k_1$), are randomly chosen. Then the sample mean of this batch of samples $\{h_i,\ k_0\le i\le k_1\}$ is calculated and a realization of the Poisson rate of

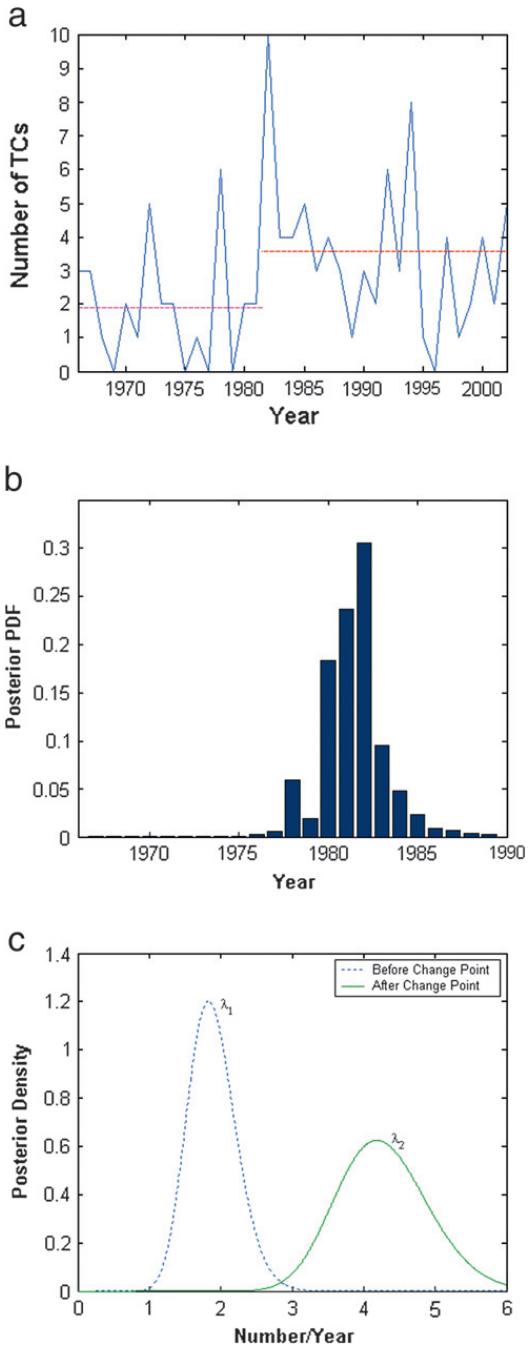

Fig. 2. Bayesian change-point analysis result for the annual tropical cyclone counts over the central North Pacific. (a) The time series of annual tropical cyclone counts over the central North Pacific from 1966 to 2002. Broken lines denote the means for the period 1966–1981 and 1982–2002 respectively. (b) The posterior probability distribution of the change-point $P(\tau|\mathbf{h},H_1)$ for the time series. (c) The smoothed posterior density function with a Gaussian kernel window of annual TC intensity before the shift, $P(\lambda_1|\mathbf{h},H_1,\tau=1982)$, and after the shift, $P(\lambda_2|\mathbf{h},H_1,\tau=1982)$ for the time series (adapted from Chu and Zhao, 2004).

this iteration, $\lambda^{[j]} = \frac{1}{k_1-k_0+1}\sum_{i=k_0}^{k_1} h_i$, is obtained. In the end, this process yields a series of samples, $\{\lambda^{[j]}, 1\le j\le L\}$. Empirically, it is assumed that this Poisson rate is gamma distributed with parameters h' and T'. Using a moment estimation approach, an approximation of the hyper-parameters h' and T' can be calculated from these samples (c.f. Zhao and Chu, 2010).

Table 2
Results of the Bayesian analysis on change-point of annual TC counts over the central North Pacific. τ stands for the change-point year, B is the Bayes factor, λ_1 and λ_2 represent the TC intensity before and after the change-point under H_1 hypothesis, respectively, and $P(H_1|\mathbf{h})$ is the posterior probability of hypothesis H_1.

$\hat{\tau}$	1982
$P(\hat{\tau}\,|\mathbf{h},H_1)$	0.31
$\overline{\lambda}_1$	1.88
$\overline{\lambda}_2$	3.57
$2\ln(B)$	2.22
$P(H_1|\mathbf{h})$	0.75

3.6. Case study for the Bayesian change-point analysis

3.6.1. Exact Bayesian inference

Example 1. Tropical cyclone activity in the central North Pacific

In this example, the exact Bayesian inference method as detailed in Section 3.2 is applied to the annual tropical cyclone count series in the CNP (Chu and Zhao, 2004). Fig. 2(a) shows the time series of annual TC rates over the CNP since 1966. The average rate from 1966 to 1981 is about 1.9 TCs per year, and it increases to almost 3.6 TCs per year from 1982 to 2002. The result of the Bayesian analysis on the shift year of the annual TC counts in CNP is listed in Table 2. From this table, we can see that the measure of Bayes factor ($2\ln(B)$) for the annual TC counts during the 1966–1989 period is 2.22, which positively favors H_1 over the H_0 hypothesis (refer to Table 1). The posterior probability that a change has occurred is rather high, reaching 0.75. Fig. 2(b) shows the posterior probability of the change-point of TC activity, plotted as a function of time. The maximum probability of 0.31 occurs in 1982. This suggests that the most likely year of the new epoch is 1982, although other change-point years such as 1981 and 1980 are plausible candidates. The posterior PDFs of TC intensity before and after the change-point, λ_1 and λ_2, are plotted in Fig. 2(c). The posterior distribution represents a combination of the prior distribution and likelihood function. In this plot, the change-point year is fixed in 1982. Fig. 2(c) shows very little overlapping in the tail areas between these two posterior distributions, and the p-value for the TC intensity difference before and after the shift, $(P(\lambda_2-\lambda_1<0|H_1,\mathbf{h}))$, is very small ($<0.01$), strongly supporting the contention of a shift toward a higher rate of annual TC intensity since 1982.

3.6.2. Bayesian inference using MCMC method

Example 2. Hurricane frequency in the eastern North Pacific

In this example, the MCMC method described in Section 3.3 is applied to the annual major hurricane count series in the eastern North Pacific (Zhao and Chu, 2006). Fig. 3(a) shows the time series of annual major hurricane counts over the eastern North Pacific from 1972 to 2003. Using the uniform prior for the hypothesis space, with a Monte Carlo integration method over

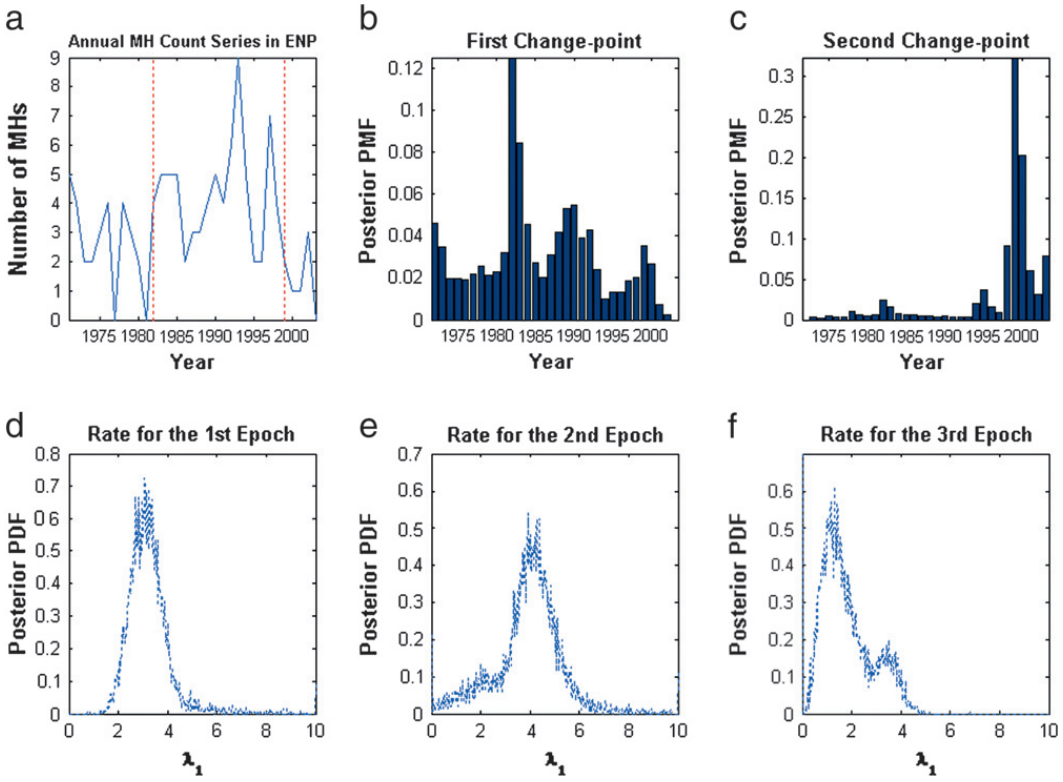

Fig. 3. Bayesian change-point analysis result for the annual major hurricane counts in the eastern North Pacific. (a) The time series of annual major hurricane counts over the ENP from 1972 to 2003. The posterior PMF for both change-points are shown in (b) and (c). The posterior PDF for the rate of each epoch is plotted in (d), (e), and (f), respectively (adapted from Zhao and Chu, 2006).

the MCMC simulation, it yields a posterior probability for each hypothesis: $P(H_0|\mathbf{h}) = 0.021$; $P(H_1|\mathbf{h}) = 0.195$; $P(H_2|\mathbf{h}) = 0.784$. This indicates that the H_2 hypothesis is by far the most likely choice. Under the H_2 hypothesis, the posterior probability mass functions for both change-points are shown in Fig. 3(b) and (c), through which it appears that the best choice for the first change-point is 1982 and the second change-point appears to be 1999. The posterior probability density function for the rate of each epoch is plotted in Fig. 3(d), (e) and (f), respectively. The average rate prior to 1982 is about 2.6 major hurricanes per year, and increases to almost 4.4 major hurricanes per year from 1982 to 1998, and drops back to 1.4 major hurricanes per year thereafter (Table 3). The p-values for the rate shift from the first epoch to the second epoch ($\lambda_2 - \lambda_1$), and the rate shift from second to the third epoch ($\lambda_2 - \lambda_3$) are all very small, 0.006 for the first and 0.004 for the second respectively. This strongly implies the existence of two change-points in this major hurricane time series.

Table 3
Bayesian analysis results on change-point of annual major hurricane counts in East North Pacific from 1972 to 2003. Here, under the uniform prior assumption in the hypothesis space, $P(H_0|\mathbf{h})$, $P(H_1|\mathbf{h})$ and $P(H_2|\mathbf{h})$ denote the posterior probability of the H_0, H_1 and H_2 hypotheses, respectively. $\hat{\tau}_1$ and $\hat{\tau}_2$ are the maximum likelihood estimates for the first and second change-point under the H_2 hypothesis, respectively. $\overline{\lambda}_1|\hat{\tau}_1, \hat{\tau}_2$, $\overline{\lambda}_2|\hat{\tau}_1, \hat{\tau}_2$ and $\overline{\lambda}_3|\hat{\tau}_1, \hat{\tau}_2$ denote the average rate, under the H_2 hypothesis, in three consecutive epochs (i.e., 1972–1981, 1982–1998, and 1999–2003), respectively, given change-points in 1982 and 1999.

Term	Value
$P(H_0\|\mathbf{h})$	0.021
$P(H_1\|\mathbf{h})$	0.195
$P(H_2\|\mathbf{h})$	0.784
$\hat{\tau}_1$	1982
$\hat{\tau}_2$	1999
$\overline{\lambda}_1\|\hat{\tau}_1, \hat{\tau}_2$	2.64
$\overline{\lambda}_2\|\hat{\tau}_1, \hat{\tau}_2$	4.41
$\overline{\lambda}_3\|\hat{\tau}_1, \hat{\tau}_2$	1.40

3.6.3. Bayesian inference via using RJMCMC method

In this section, the RJMCMC based generic Bayesian extreme event regime shift analysis method as elaborated in Section 3.4 is applied to several real-world cases. For all examples, the maximum number of change-points is set as nine.

Example 3. Typhoon frequency in the western North Pacific

In this example, the RJMCMC method is applied to the annual super typhoon count series over the WNP from 1960 to 2006 (Zhao and Chu, 2010). Fig. 4(a) shows the count time series. After running the proposed RJMCMC simulation, the output for the posterior probability of ten candidate hypotheses are plotted in Fig. 4(e). The probability for hypothesis H_2 is as high as 0.39, which is higher than that for hypothesis H_3 (0.24). Therefore H_2 is chosen as the winning hypothesis. Under hypothesis H_2, the marginal posterior PMF for the two change-points are depicted in Fig. 4(c) and (d), respectively, through which one can see that by far the most likely choice for the first change-point is 1972 and for the second change-point is 1989.

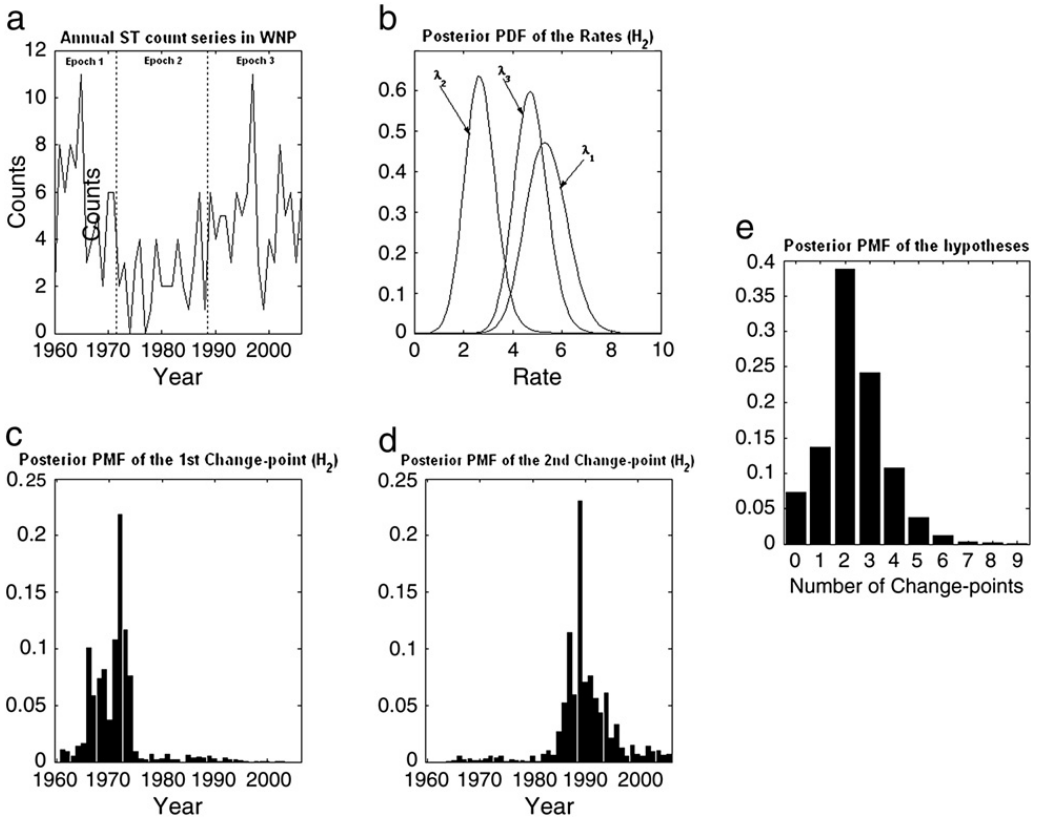

Fig. 4. Plots for the change-point analysis for the annual super typhoon count series over the western North Pacific from 1960 to 2006. (a) Time series. (b) Posterior probability density functions for the rates of each epoch. (c) Posterior probability mass function for the first change-point. (d) Posterior probability mass function for the second change-point. (e) Posterior probability mass function for the candidate hypotheses. In (b), λ_1 is for the period 1960–71, λ_2 for the period 1972–88, and λ_3 for the period 1989–2006 (adapted from Zhao and Chu, 2010).

The posterior PDF for the rates of each epoch, along with the universal prior, are collaboratively demonstrated in Fig. 4(b). The average rate prior to 1972 (λ_1) is about 5.7 super typhoons per year; decreases to about 2.4 super typhoons per year from 1972 to 1988 (λ_2); then increases to 5.0 super typhoons per year thereafter (λ_3). The p-values for the both shifts are all very small (0.054 for the first and 0.059 for the second), which strongly implies the existence of these two change-points in this storm time series. The Bayes factor between H_2 and H_3 is 1.61, which shows only slight evidence in favor of H_2 over H_3. However, under hypothesis H_3, based on the obtained change-point samples, besides the two shifts (1972 and 1989) identified under H_2, the third most probable change-point is 1987, which is very close to 1989. In summary, it is plausible to suspect two regime shifts that occurred within this series; one was around early 1970s and the other was around late 1980s.

Example 4. Heavy rainfall records in Hawaii

In this example, the RJMCMC method is applied to the annual heavy rainfall count series at Manoa in Oahu, Hawaii from 1920 to 2009. Fig. 5(a) shows the relative count series. The output for the posterior probability for ten candidate hypotheses is plotted in Fig. 5(c). The posterior probability for hypothesis H_3 is as high as 0.37, which is almost the double of that for the next highest possible choice, hypothesis H_0 or H_4 (both with 0.20). Therefore H_3 is chosen as the winning hypothesis. Under hypothesis H_3, the marginal posterior PMF for the three change-points are depicted in Fig. 5(d), (e) and (f), respectively, through which one can see that by far the most likely choice for the first change-point is 1963, 1993 for the second change-point, and 2004 for the most recent change-point. The posterior PDF for the rates of each epoch are collaboratively demonstrated in Fig. 5(b). The average rate from 1920 to 1962 is about 8.8 heavy rainfalls per year (λ_1); drastically increases to about 13.1 heavy rainfalls per year from 1963 to 1992 (λ_2); decreases to 7.0 heavy rainfalls per year from 1993 to 2003 (λ_3); and thereafter the average heavy rainfall count goes to 15.0 (λ_4). The p-values for the three shifts are all very small (0.043, 0.053, and 0.066 for the three change-points, respectively), which strongly implies the existence of these three change-points in this time series.

Example 5. Heat wave records in France

In this example, the RJMCMC method is applied to the annual heat wave count series in French from 1949 to 2009 (Fig. 6(a)). The output for the posterior probability for ten candidate hypotheses is plotted in Fig. 6(b). The posterior probability for hypothesis H_1 is as high as 0.44, which is much higher than H_2 (0.33). Therefore H_1 is chosen as the winning hypothesis. Under hypothesis H_1, the marginal posterior PMF for the only change-point is depicted in Fig. 6(c), through which one can see that the most likely choice for this change-point is 1983. The posterior PDF for the rates of both epochs are collaboratively demonstrated in Fig. 6(d). The average heat wave rate from 1949 to 1982 is about 0.6 heat wave events per year, which is equivalent to an expectation of one heat wave

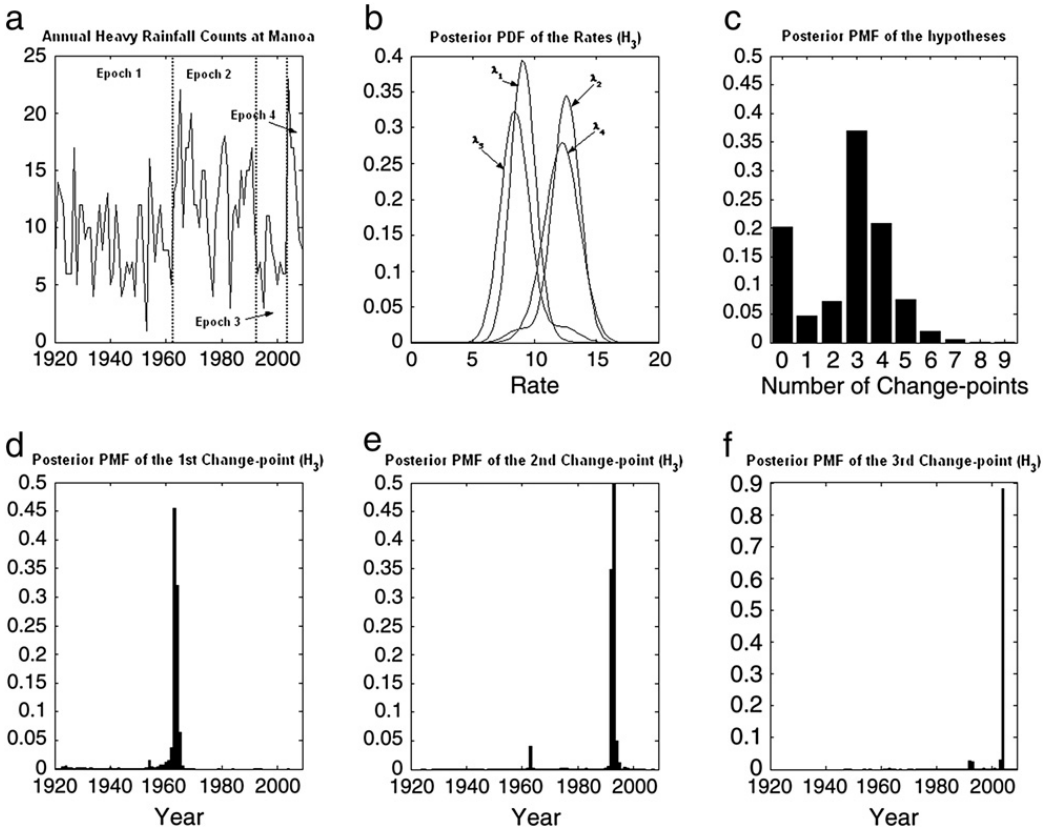

Fig. 5. Plots for the change-point analysis for the annual heavy rainfall count series at Manoa in Hawaii from 1920 to 2009. (a) Time series. (b) Posterior probability density functions for the rates of each epoch. λ_1 is for the period 1920–62, λ_2 for the period 1963–92, λ_3 for the period 1993–2003, and λ_4 for the period 2004–09. (c) Posterior probability mass function for the candidate hypotheses. (d) Posterior probability mass function for the first change-point. (e) Posterior probability mass function for the second change-point. (e) Posterior probability mass function for the third change-point.

event every other year. This rate dramatically increases to about 1.8 heat waves per year starting from 1983, almost to an expectation of 2 heat waves per year. This rate hike is significant as it has almost quadrupled since the shift. The p-values for this shift is extremely small as well (0.001), strongly suggesting the existence of the significant mean shift in this time series.

4. Bayesian tropical cyclone track pattern Clustering

4.1. Finite mixture Gaussian model for tropical cyclone track pattern clustering

Camargo et al. (2007) proposed a finite mixture Gaussian model to solve the tropical cyclone track clustering problem. More recently, Chu et al. (2010a) applied the same method to the same basin but with longer record lengths and focused on the climate change aspects for each track type. In this section, we shall elaborate this clustering model in the Bayesian context.

A key feature of the mixture Gaussian model is its ability to model multimodal densities while adopting a small set of basic component densities. Finite mixture models have been used for clustering data in a variety of areas such as large-scale atmospheric circulation. Based on the assumption that there are a few distinct types characterizing TC tracks in a basin of interest, we model each TC track path as a second-order polynomial function of the lifetime of this TC. Mathematically, for each specific track type, the set of coefficient of this polynomial function is presumably jointly Gaussian distributed. Each TC track type thus has a unique distribution parameter. The space spanned by the parameters of this track type model is a linear combination of a set of distinct Gaussian distributions.

Assuming there are n observed track records for a given TC. For each record, there are three features reported — latitude, longitude, and the time. We denote the path record of a TC and its relative observed time vector for the second order polynomial function, respectively, by

$$\mathbf{z} = \left[\mathbf{z}_{lat}, \mathbf{z}_{long}\right] = \begin{bmatrix} z_{1,lat} & z_{1,long} \\ \cdots & \cdots \\ z_{n,lat} & z_{n,long} \end{bmatrix}, \mathbf{T} = \begin{bmatrix} 1 & t_1 & t_1^2 \\ \cdots & \cdots & \cdots \\ 1 & t_n & t_n^2 \end{bmatrix} \tag{11a}$$

where $z_{i,lat}$ and $z_{i,long}$ for $i = 1,..,n$ represent the i-th latitude and longitude record; and t_i represents the time for the i-th records of this TC relative to the first record for $i = 1,..,n$. We further assume that there are K distinct TC track types in the basin of interest, where K is assumed to be a constant in a given hypothesis or model. With the model defined in Eq. (11a), if a TC is categorized as type k, $1 \le k \le K$, the link function between the TC track path and relative time is governed by the following formula

$$\mathbf{z} = \mathbf{T}\boldsymbol{\beta}^{\mathbf{k}} + \boldsymbol{\varepsilon}, \text{where} \boldsymbol{\beta}^{\mathbf{k}} = \begin{bmatrix} \beta_{0,lat}^k & \beta_{0,long}^k \\ \beta_{1,lat}^k & \beta_{1,long}^k \\ \beta_{2,lat}^k & \beta_{2,long}^k \end{bmatrix} \text{and} \boldsymbol{\varepsilon} \sim N\left(0, \boldsymbol{\Sigma}^{\mathbf{k}}\right), 1 \le k \le K . \tag{11b}$$

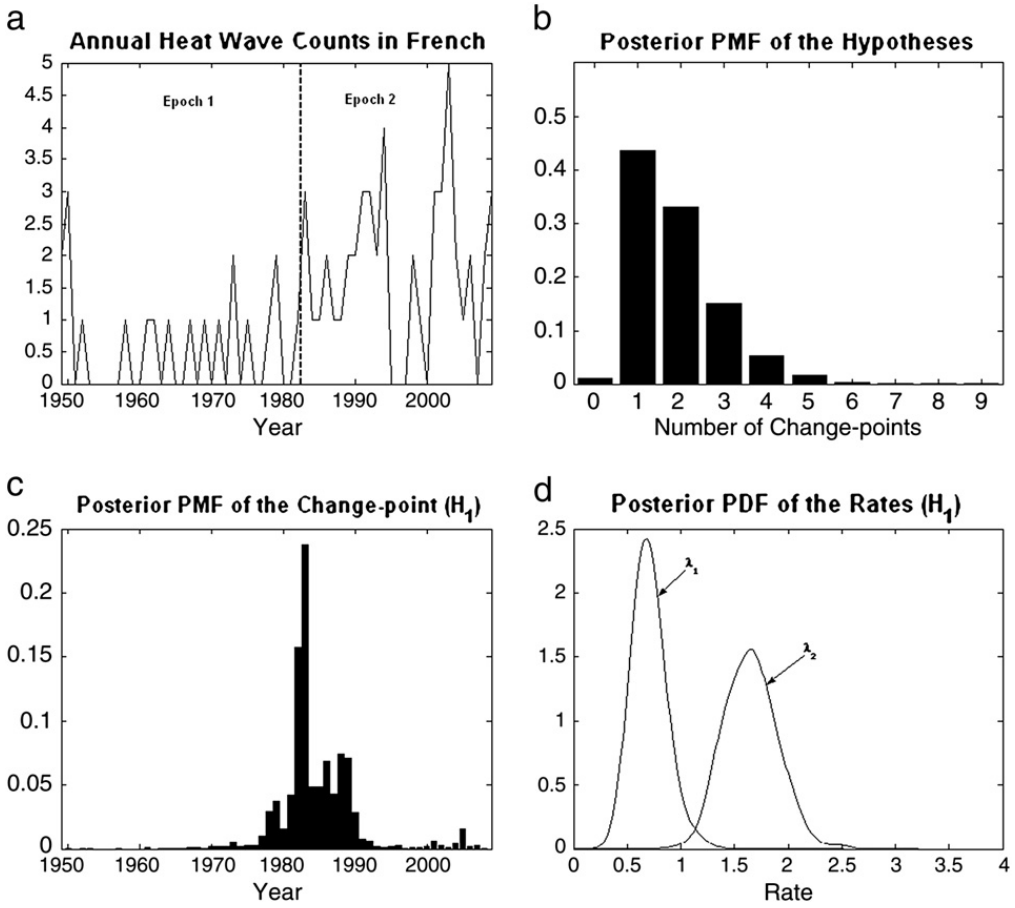

Fig. 6. Plots for the change-point analysis for the annual heat wave count series in Mont, France from 1949 to 2009. (a) Time series. (b) Posterior probability mass function for the candidate hypotheses. (c) Posterior probability mass function for the change-point. (d) Posterior probability density functions for the rates of each epoch.

In model (11b), the parameter set $\boldsymbol{\beta}^{\mathbf{k}}$ is distinct for each TC clustering types and $N(\bullet,\bullet)$ denotes the normal distribution. With this model, intuitively one can see that the zero-order coefficient dual provides the mean genesis location of this clustering type; the first-order term features the characteristic linear direction of this path type; the second-order term determines the recurving shape of the typical path of this type; and the covariance matrix (Σ) in determines the spread of a particular type. The noise term in model (11b), $\boldsymbol{\varepsilon}_i$, is assumed multivariate Gaussian with zero mean and a 2 by 2 covariance matrix, $\boldsymbol{\Sigma}_k$.

The conditional density for the i-th cyclone, conditioned on membership in the cluster type k, is therefore defined as

$$P(\mathbf{z}_i|\mathbf{T}_i,\boldsymbol{\theta}_k)=(2\pi)^{-n_i}|\boldsymbol{\Sigma}_k|^{-n_i/2}\exp\left\{-\mathrm{tr}\left[(\mathbf{z}_i-\mathbf{T}_i\boldsymbol{\beta}_k)\boldsymbol{\Sigma}_k^{-1}(\mathbf{z}_i-\mathbf{T}_i\boldsymbol{\beta}_k)'\right]/2\right\}. \tag{12a}$$

In Eq. (12a), operator exp{•} denotes an exponential function with a natural base; we adopt the notation $\boldsymbol{\theta}_k=\{\boldsymbol{\beta}_k,\boldsymbol{\Sigma}_k\}$, which is referenced in model (11b); and operator tr(.) denotes the matrix operation function "trace." By the definition of a mixture Gaussian model, Eq. (12a) leads to the marginal mixture model

$$P(\mathbf{z}_i|\mathbf{T}_i)=\sum_{k=1}^{K}\alpha_k P(\mathbf{z}_i|\mathbf{T}_i,\boldsymbol{\theta}_k) \tag{12b}$$

where, $P(\mathbf{z}_i|\mathbf{T}_i,\boldsymbol{\theta}_k)$ is given by (12a), and α_k is the posterior probability of cluster k, which implies $\sum_{k=1}^{K}\alpha_k=1$. If we let $\mathbf{Z}'=\left[\mathbf{z}_1',\mathbf{z}_2',\ldots,\mathbf{z}_N'\right]$ be the complete set of all observed TC trajectories, and $\mathbf{T}'=\left[\mathbf{T}_1',\mathbf{T}_2',\ldots,\mathbf{T}_N'\right]$ be the associated measurement times, then the full probability density of $\mathbf{Z}$ given $\mathbf{T}$, the conditional likelihood, is formulated by

$$P(\mathbf{Z}|\mathbf{T})=\prod_{i=1}^{N}\sum_{k=1}^{K}\alpha_k P(\mathbf{z}_i|\mathbf{T}_i,\boldsymbol{\theta}_k) \tag{13}$$

where $P(\mathbf{z}_i|\mathbf{T}_i,\boldsymbol{\theta}_k)$ is defined in Eq. (12a).

4.2. Bayesian inference for tropical cyclone track pattern clustering

Throughout this section, we assume that the number of cluster type, K, is given. For a real application, we can refer to the literature to choose the proper number for this parameter (e.g., Camargo et al., 2007; Kim et al., 2011). Because hypothesis selection is not the focus of this section, it is proper to choose a non-informative prior for those model coefficients; that is, $P(\boldsymbol{\theta}_k,\alpha_k)\propto 1$ for model (13). With this non-informative prior assumption, and following the basic Bayes formula given in Eq. (1), the posterior distribution for $\{\boldsymbol{\theta}_k,\alpha_k\}$ is proportional to the conditional likelihood given in Eq. (13). It is important to

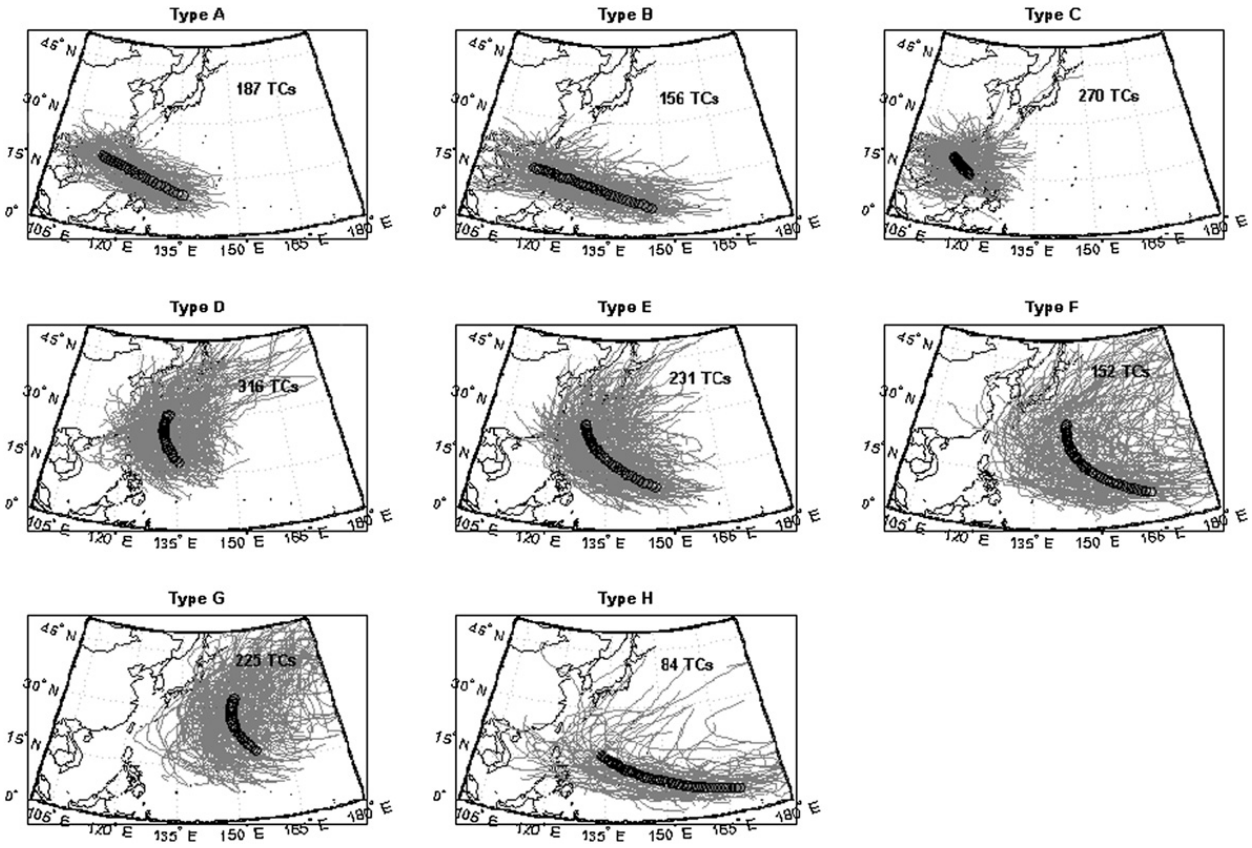

Fig. 7. Eight tropical cyclone track types over the western North Pacific identified by the mixture Gaussian model. The number in each panel indicates the number of cases in each type. Black circles denote the mean track for each type (adapted from Chu et al., 2010a).

note that a non-informative prior is not proper for a hypothesis selection problem. To include cluster number K in a hierarchical Bayesian model, it is necessary to either resort to a hierarchical hyper-prior (e.g., Richardson and Green, 1997) or make a proper informative prior (e.g., Zhao and Chu, 2010) for parameter $\boldsymbol{\theta}_k$.

In many real-world applications, only the peak areas of the posterior distribution may be of interest. An efficient approach to estimating the mode of the posterior distribution is the Expectation-Maximization (EM) algorithm. Given the likelihood model (13), in the E-step the membership probability of a TC categorized to each clustering type is calculated. In the M-step, the optimization estimation for the model parameter set of each type is calculated. These include regression parameters, the posterior probability of cluster k, and the covariance matrix. The maximization formula for coefficient parameter $\hat{\boldsymbol{\beta}}^{\mathbf{k}}$ and variance parameter $\boldsymbol{\Sigma}^{\mathbf{k}}$ are derived from a linear Bayesian regression model (Gelman et al., 2004). The details of the formula for the EM algorithm are provided in Camargo et al. (2007) or Chu et al. (2010a).

Given the number of clusters and an initial setting of the model parameters, after a few iterations, the proposed EM algorithm will converge to a fixed set of parameter estimation. Usually, the convergence of an EM algorithm is determined when the difference between two iterations is less than a sufficiently small value. Note that these convergent values are not necessarily the global optimum estimation and are determined by the initial starting values. Therefore, multiple different initial values should be selected and the set of estimation with the maximum likelihood of the observation chosen.

4.3. Typhoon track types over the WNP via using the Bayesian track path clustering

In this case study, we apply the mixture Gaussian clustering method to objectively cluster TC tracks over the WNP. Fig. 7 shows the eight major track patterns over the WNP and the South China Sea from the proposed clustering analysis, with three straight movers (types A, B and C), four recurving ones (types D, E, F, and G), and one mixed pattern of both straight moving and recurving (type H). The type A and B clusters are similar in nature in that they both move more or less straight across the Philippines to the South China Sea and/or Hong Kong, Hainan, and Vietnam. The major difference is that type B storms tend to form farther eastward and southward than type A storms. As a result, the mean track for type B storms is longer than that of Type A. Type C cyclones form in the South China Sea and are landlocked by the Indochina peninsula and southern China coast, with a very short path. Similar to types A and B, type D and E systems form in the Philippine Sea but they follow a northward path and many of them made landfall on Taiwan, the eastern China coast, Japan, and Korea. Type F storms tend to form in low-latitudes and away from Asia. Type G storms also form far away from the Asian continent but at higher latitudes (~15°N). They move northwestward and then northward to the east of Japan over the open ocean. Storms associated with type H are generally formed near the equator and to the east of 165°E, and have a long track path.

On a side note, the output results from the track path clustering can serve as the platform for further analyses. For

example, based on the results provided herein, Chu et al. (2010a) examined temporal changes of TC activity (e.g., lifetime, intensity) for each type of typhoon over the last 60 years and interesting results were reported therein.

5. Bayesian prediction for seasonal tropical cyclone activity

In Section 3 and 4, we present two different types of non-supervised Bayesian learning problems. For both types there is no response or target variable involved in the analysis. In this section, we will discuss the Bayesian analysis applied to supervised learning problems. Specifically, we first describe two statistical regression models for seasonal tropical cyclone activity — the generalized Poisson regression model and the probit regression model. We then discuss the predictor selection procedure followed by introducing the overall Bayesian forecast scheme. With a regression model, the relationship between the target response variable, seasonal typhoon counts, and selected predictors can be mathematically built; the details are discussed in this section.

5.1. Bayesian regression via the generalized Poisson regression model

If we assume that there are *N*observations that are conditional on *K* predictors, we define a latent random *N*-vector $\mathbf{Z}$ such that for each observation h_i, $i=1,2,\ldots,N$, $Z_i=\log\lambda_i$, and λ_i is the Poisson rate for the *i*-th observation. The link function between the latent variable and its associated predictors is expressed as $Z_i=\mathbf{X}_i\boldsymbol{\beta}+\varepsilon_i$, where $\boldsymbol{\beta}=[\beta_0,\beta_1,\beta_2,\ldots,\beta_K]'$ is a random vector; noise ε_i is assumed to be identical, independently distributed, and normally distributed with zero mean and σ^2 variance; $\mathbf{X}_i=[1,X_{i1},X_{i2},\ldots,X_{iK}]$ denotes the predictor vector. In vector form, the general Poisson linear regression model is formulated as follows

$$P(\mathbf{h}|\mathbf{Z})=\prod_{i=1}^{N}P(h_i|Z_i),\ \text{where}\ h_i|Z_i\sim Poisson\left(h_i|e^{Z_i}\right),$$
$$\mathbf{Z}|\boldsymbol{\beta},\sigma^2,\mathbf{X}\sim Normal\left(\mathbf{Z}|\mathbf{X}\boldsymbol{\beta},\sigma^2\mathbf{I}_N\right),\ \text{where, specifically}$$
$$\mathbf{X}'=\left[\mathbf{X}_1',\mathbf{X}_2',\ldots,\mathbf{X}_N'\right],\mathbf{I}_N\ \text{is the}\ N\times N\ \text{identity matrix, and}$$
$$\mathbf{X}_i=[1,X_{i1},X_{i2},\ldots,X_{iK}]\ \text{is the predictor vector for}\ h_i, i=1,2,\ldots,N,$$
$$\boldsymbol{\beta}=[\beta_0,\beta_1,\beta_2,\ldots,\beta_K]'. \tag{14}$$

Here, *Normal*(•) and *Poisson*(•) stand for the normal distribution and the Poisson distribution, respectively. The probability mass function of the Poisson distribution is defined in Eq. (4). In model (14), β_0 is referred to as the intercept.

To complete the Bayesian inference model for the generalized regression model, a prior distribution for the model parameters is needed. Because we do not have any credible prior information for the coefficient vector $\boldsymbol{\beta}$ and the variance σ^2, it is reasonable to choose the non-informative prior, $P(\boldsymbol{\beta},\sigma^2)\propto\sigma^{-2}$, which is not a proper probability distribution function; however, it leads to a proper posterior distribution.

The posterior distribution for the hidden variable *Z*, which is conditionally independent from each other given the model parameters $\boldsymbol{\beta}$ and σ^2, is derived in Chu and Zhao (2007). With the newly observed predictor set $\tilde{\mathbf{X}}=\left[1,\tilde{X}_{i1},\tilde{X}_{i2},\ldots,\tilde{X}_{iK}\right]$, the predictive distribution for the new latent variable $\tilde{Z}$ and TC counts $\tilde{h}$ is governed by

$$P\left(\tilde{Z}|\tilde{X},\mathbf{X},\mathbf{h}\right)=\iint_{\boldsymbol{\beta},\sigma^2}P\left(\tilde{Z}|\tilde{X},\boldsymbol{\beta},\sigma^2\right)P\left(\boldsymbol{\beta},\sigma^2|\mathbf{X},\mathbf{h}\right)d\boldsymbol{\beta}d\sigma^2 \tag{15a}$$

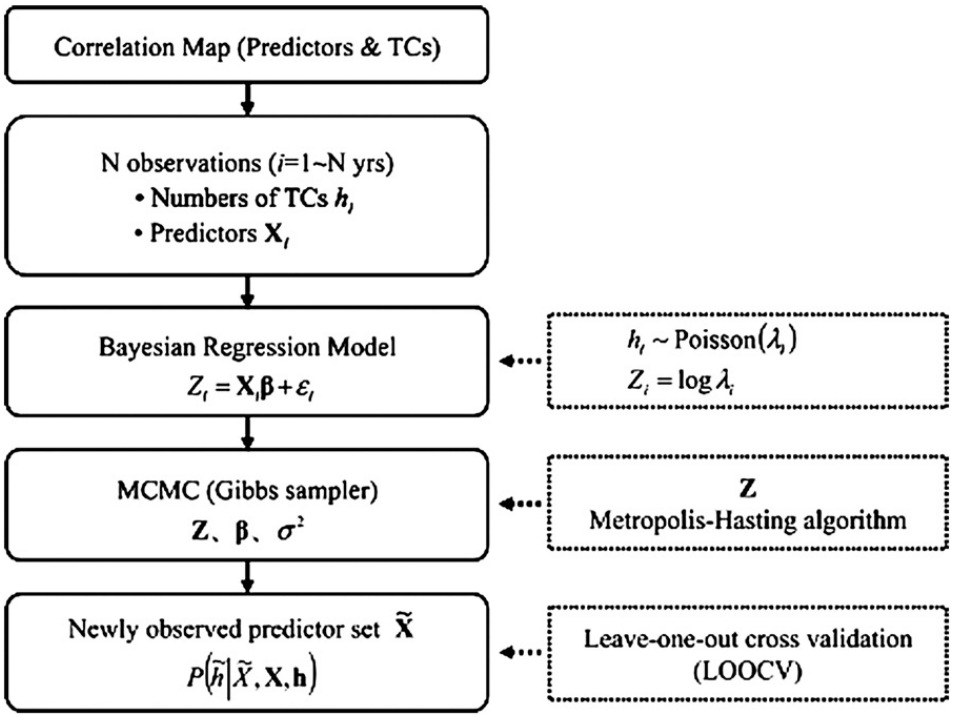

Fig. 8. Schematics of the Bayesian regression forecast model, after Lu et al. (2010).

$$P\left(\tilde{h}\,|\,\tilde{X},\mathbf{X},\mathbf{h}\right)=\int_{\tilde{Z}}\frac{\exp\left(-e^{\tilde{Z}}+\tilde{Z}\tilde{h}\right)}{\tilde{h}!}P\left(\tilde{Z}\,|\,\tilde{X},\mathbf{X},\mathbf{h}\right)d\tilde{Z}. \quad (15b)$$

In (15b), operator exp{•} denotes an exponential function with the natural base; the "new" variables refer to those that are not involved in the model construction and only used for prediction. Fig. 8 illustrates a conceptual framework of the Bayesian regression forecast model (Lu et al., 2010).

With the non-informative prior, the posterior distribution for the model parameter set $(\boldsymbol{\beta},\sigma^2)$ is not standard and sampling directly from it is difficult. Alternatively, Chu and Zhao (2007) designed a Gibbs sampler to solve this Bayesian inference problem. This inference problem is grounded on the ordinary linear regression model, which can be found in most classic Bayesian literatures (e.g., Gelman et al., 2004). The algorithm is listed in Eq. (16) and a detailed derivation for this algorithm is available in Chu and Zhao (2007).

1. Select proper initial value for $\mathbf{Z}^{[0]},\boldsymbol{\beta}^{[0]},\sigma^{2[0]}$ and set $i=1$.
2. Draw $Z_j^{[i]}$ from $Z_j^{[i]}\,|\,\mathbf{h},\boldsymbol{\beta}^{[i-1]},\sigma^{2[i-1]}$ for $j=1,2,\ldots,N$ via (17), the conditional distribution for the hidden variable Z.
3. Draw $\boldsymbol{\beta}^{[i]}$ from $\boldsymbol{\beta}^{[i]}\,|\,\mathbf{h},\mathbf{Z}^{[i]},\sigma^{2[i-1]}$ via (18), a multivariate Gaussian distribution. (16)
4. Draw $\sigma^{2[i]}$ from $\sigma^{2[i]}\,|\,\mathbf{h},\mathbf{Z}^{[i]},\boldsymbol{\beta}^{[i]}$ via (19), an inverse χ^2 distribution.
5. Set $i=i+1$ then go back to step 2 until the required number of iterations are met.

In algorithm (16), the conditional posterior distribution for each hidden variable Z is given by

$$P\left(Z_i\,|\,\mathbf{h},\boldsymbol{\beta},\sigma^2\right)\propto\exp\left\{-e^{Z_i}+Z_ih_i-\frac{1}{2\sigma^2}(Z_i-\mathbf{X}_i\boldsymbol{\beta})^2\right\},i=1,2,\ldots N. \quad (17)$$

In Eq. (17), the operator exp{•} is defined in Eq. (15b). To sample Z_i from Eq. (17), one can choose Metropolis–Hasting algorithm (Hastings, 1970) or Laplace approximation method. The conditional posterior distributions for the coefficient vector and noise variance term are provided by Eqs. (18) and (19), respectively, both of which are derived from the ordinary linear regression model.

$$\boldsymbol{\beta}\,|\,\mathbf{Z},\mathbf{h},\sigma^2\sim Normal\left(\boldsymbol{\beta}\,|\,\hat{\boldsymbol{\beta}},(\mathbf{X}'\mathbf{X})^{-1}\sigma^2\right),\ \text{where}\ \hat{\boldsymbol{\beta}}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}. \quad (18)$$

$$\sigma^2\,|\,\mathbf{Z},\mathbf{h},\boldsymbol{\beta}\sim\text{Inv}-\chi^2\left(\sigma^2\,|\,N,s^2\right),\text{where}\ s^2=\frac{1}{N}(\mathbf{Z}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{Z}-\mathbf{X}\boldsymbol{\beta}). \quad (19)$$

In (19), Inv $-\chi^2$ denotes the scaled-inverse-χ^2 distribution.

5.2. Bayesian classification via the probit regression model

The generalized Poisson regression model introduced in Section 5.1 has been proved very effective for most rare event count series (e.g., Chu et al., 2007; Lu et al., 2010). However, if the underlying rate is significantly below one then this model may introduce significant bias. For this kind of application, it is more effective instead to adopt a binary classification model. That is, the response variable here is a binary class label, which is termed by "Y." For each observation period, we define a class "Y = 1" if one or more TCs are observed and "Y = 0" otherwise.

The probit regression model (e.g., Albert and Chib, 1993) assumes that there are Nobservations conditional on K selected predictors. A latent random N-vector $\mathbf{Z}$ is defined, such that for each observation y_i, $i=1,2,\ldots,N$, $y_i=1$ if $Z_i\geq 0$ and $y_i=0$otherwise. The link function between the latent variable $\mathbf{Z}$ and its associated predictors is also linear, $Z_i=\mathbf{X}_i\boldsymbol{\beta}+\varepsilon_i$, where $\boldsymbol{\beta}=[\beta_0,\beta_1,\beta_2,\ldots,\beta_K]'$ is a random vector; noise ε_i is assumed to be identical, independently distributed, and normally distributed with zero mean and σ^2 variance; $\mathbf{X}_i=[1,X_{i1},X_{i2},\ldots,X_{iK}]$ denotes the predictor vector. In vector form, the probit regression model is described by (Chu et al., 2010b):

$$P(\mathbf{y}\,|\,\mathbf{Z})=\prod_{i=1}^{N}P(y_i\,|\,Z_i),\ \text{where}\ y_i=\begin{cases}1 & \text{if}\ Z_i\geq 0\\ 0 & \text{if}\ Z_i<0\end{cases}. \quad (20)$$

The classification model (20) is very similar to the Poisson regression model (14). In this case study, the probability of class Y = 1 can be viewed as the rate of TC counts.

In model (20), the posterior distribution of hidden variable Z is conditionally independent from each other given the model parameters $\boldsymbol{\beta}$ and σ^2. Similar to the Poisson regression model, with the newly observed predictor set $\tilde{\mathbf{X}}=\left[1,\tilde{X}_{i1},\tilde{X}_{i2},\ldots,\tilde{X}_{iK}\right]$, the predictive distribution for the latent variable $\tilde{Z}$ and TC counts $\tilde{h}$ will be

$$P\left(\tilde{Z}\,|\,\tilde{X},\mathbf{X},\mathbf{h}\right)=\iint_{\boldsymbol{\beta},\sigma^2}Normal\left(\tilde{Z}\,|\,\tilde{X}\boldsymbol{\beta},\sigma^2\right)P\left(\boldsymbol{\beta},\sigma^2\,|\,\mathbf{X},\mathbf{h}\right)d\boldsymbol{\beta}d\sigma^2 \quad (21a)$$

$$P\left(\tilde{h}\,|\,\tilde{X},\mathbf{X},\mathbf{h}\right)=\int_{\tilde{Z}\geq 0}P\left(\tilde{Z}\,|\,\tilde{X},\mathbf{X},\mathbf{h}\right)d\tilde{Z}. \quad (21b)$$

The posterior distribution for the model parameter set $(\boldsymbol{\beta},\sigma^2)$ in Eq. (21a) is not standard with a non-informative prior. Similar to Section 5.1, a Gibbs sampler is alternatively designed to solve this Bayesian inference problem; then the Monte Carlo method can be used to integrate the prediction in Eq. (21b).

Because the hierarchical probit regression model in Eq. (20) is very similar to the Poisson regression model in Eq. (14), it is proper to adopt most of the formulae provided in the Bayesian inference for a generalized regression model. Algorithm (16) can also be used to solve this model, except that the conditional posterior distribution for the latent variable Z (Step 2 in algorithm (16)) should be drawn from a truncated Gaussian distribution. That is:

$$\begin{aligned}&Z_i\,|\,\mathbf{X}_i,\beta_0,\sigma^2,y_i=1\propto N\left(\mathbf{X}_i\boldsymbol{\beta},\sigma^2\right)\ \text{truncated at the left by 0}\\ &Z_i\,|\,\mathbf{X}_i,\beta_0,\sigma^2,y_i=0\propto N\left(\mathbf{X}_i\boldsymbol{\beta},\sigma^2\right)\ \text{truncated at the right by 0}\\ &i=1,2,\ldots,N,\ \mathbf{X}_i\ \text{represents the}\ i-\text{th row the predictor matrix}\ \mathbf{X}.\end{aligned} \quad (22)$$

Note that the probit regression based Bayesian classification model described in this subsection is neither necessarily limited to a binary classification problem, nor to a generalized linear link function. Via a multinomial probit regression model, the probabilistic model (20) can be extended to solve a multi-classification problem. With some straightforward

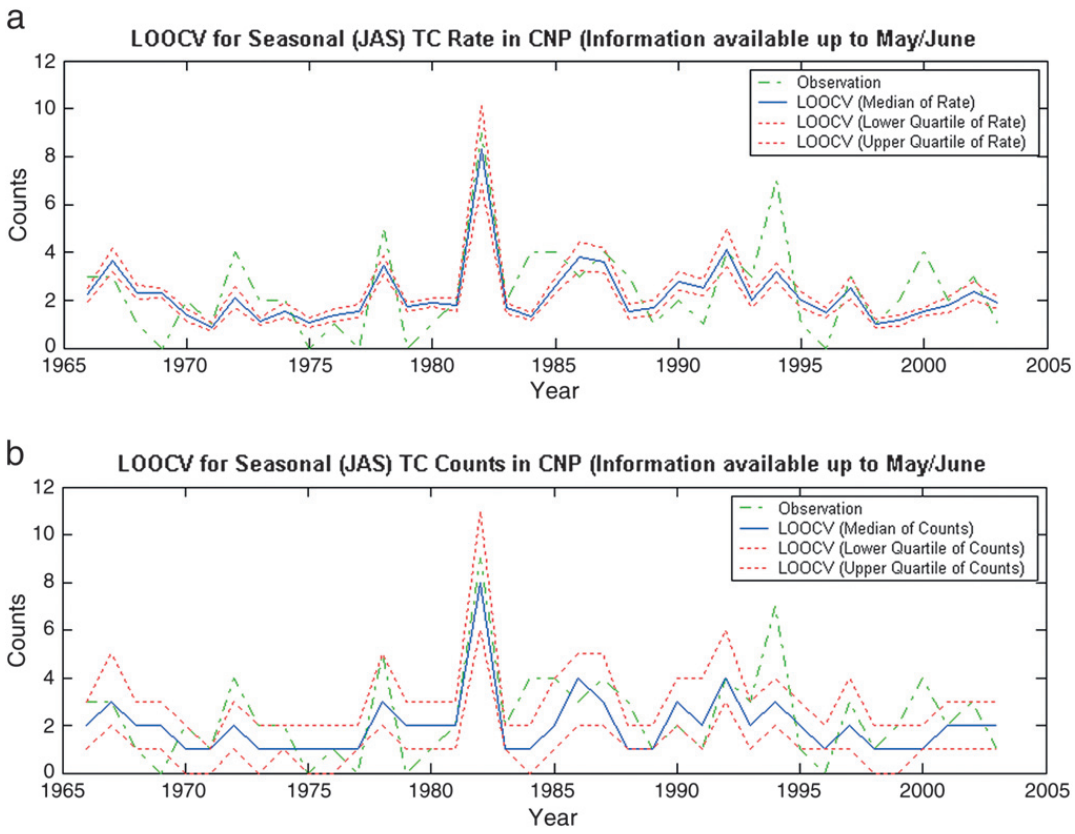

Fig. 9. (a) The median (solid line), and upper and lower quartiles (dotted line) of the LOOCV predicted TC rate are plotted together with the actual observed TC counts (dash–dotted line) over the CNP from 1966 to 2003. (b) The median (solid line), and upper and lower quartiles (dotted line) of the LOOCV predicted TC counts are plotted together with the actual observations (dash–dotted line) from 1966 to 2003 (adapted from Chu and Zhao, 2007).

revision, algorithm (16) (the version for the probit regression model) can be used to solve multi-classification problems. It's worth noting that, the linear link function in model (20) can be replaced by other more generic functions such as Gaussian processes (GP). A relevant application involving this non-trivial extension is found in Zhao and Cheung (2007, 2011).

5.3. Case study

The following example demonstrates the general procedure of the Bayesian analysis framework elaborated in this section. The example only adopts a Poisson regression model, which is detailed in Section 5.1. However, the procedure for using the probit regression based classification model given in Section 5.2 is very similar to that for a Poisson regression model. One application for using a Bayesian probit regression model can be found in Chu et al. (2010b); the probit regression model applied for tropical cyclone formation forecast can be found in Chand and Walsh (2011).

Regarding the regression model verification, a general way to verify the effectiveness of a regression method is to apply a strict cross-validation (CV) test for the relevant dataset. Because the TC variation is approximately independent from year to year, it is proper to apply the leave-one-out cross-validation (LOOCV). That is, given there are N years of observations, a target year is chosen and a model is developed using the remaining data as the training set. The observations of the selected predictors for the target year are then used as inputs to forecast the missing year. This process is repeated successively until all N forecasts are made. This LOOCV verification procedure shall be applied to the example discussed in this section.

In this example, there are a total of 38 years (1966–2003) of TC counts in the central North Pacific. After applying the simple correlation based predictor selection procedure proposed in Chu and Zhao (2007) to this count series, five predictors are chosen at a confidence level of 99% for this regression problem. These five predictors include sea surface temperature (SST), sea level pressure (SLP), vertical wind shear, low-level relative vorticity, and precipitable water. With these predictors, after applying the Gibbs sampler algorithm (16) to this data set under a LOOCV verification procedure, as detailed above, the yielded results are provided in Fig. 9. The figure shows the predicted TC rate and counts when using the median, upper, and lower quartiles (the upper 75% and lower 25%) of model parameter set through a LOOCV. These are plotted together with the actual observation for each year. The distance between the upper quartile and lower quartile locates the central 50% of the predicted TC variations. The Pearson correlation between the median of the predictive rate and independent observations is 0.63. Out of a total of 38 years, there are only 9 years in which the actual TC counts lie outside the predictive central 50% boundaries. It's worth noting that although the correlation based predictor selection approach referred herein is simple and practically robust, it is a non-Bayesian approach. To embed the predictor selection into the Bayesian regression or classification model is one of the active research subjects in many scientific fields. Also, as the predictor selection is not included in the LOOCV, the out-of-sample skill estimate tends to be slightly overrated.

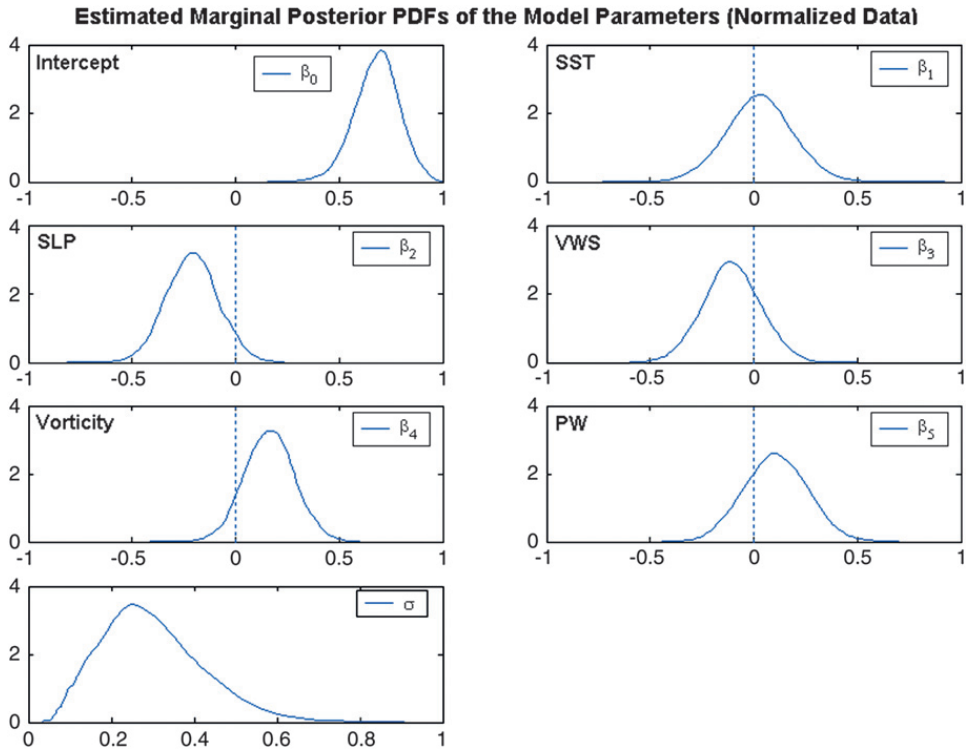

Fig. 10. Estimated marginal posterior PDFs for model parameter set (β, σ) given the peak season TC counts in the CNP and the selected predictors from 1966 to 2003 (adapted from Chu and Zhao, 2007).

With all the samples, the marginal probability density function (PDF) for the parameter set, β, and σ can also be estimated. The marginal posterior PDF for each model parameter for this example is shown in Fig. 10. The relative contribution of each regression coefficient in the Bayesian strategy can be judged approximately by the so-called p-value. This can be evaluated by the ratio of the number of samples that lie to the left of zero to the total number of iterations if the predictor is expected to have a positively orientated impact on the forecast quantity (e.g., SST). Conversely, if the predictor is to have a negatively orientated impact (e.g., SLP), the ratio of the number of samples that lie to the right of zero to the total number of iterations is of concern. Graphically, the smaller the area to the left (positively oriented predictors) or right (negatively oriented predictors) of zero in the PDF plots, the more important this predictor is in the regression model. Fig. 10 indicates that SLP and, to a lesser extent, relative vorticity are key predictors.

6. Summary

In this review article, we present three important applications of the Bayesian paradigm to extreme climatic events. The applications include: (1) identifying abrupt shifts for an extreme event count series; (2) objective clustering of the large quantity and seemingly complex historical typhoon tracks over the WNP into several distinct track types; and (3) prediction of seasonal tropical cyclone frequency in a region. For the task outlined in (1), the extreme events encompass the tropical cyclone count series in the central North Pacific, major hurricane count series in the eastern North Pacific, super typhoon count series in the western North Pacific, heavy rainfall count series at Manoa in Oahu, Hawaii, and heat wave count series in France. For task (2), the tropical cyclones in the western North Pacific are categorized into eight distinct types based on their track pathways and genesis locations. And for task (3), the seasonal tropical cyclone activity is predicted using the antecedent environmental conditions as predictors.

As the Earth's climate is changing, the frequency of extreme events is expected to change accordingly. Hence, developing a state-of-the-art method to objectively identify the turnaround of such changes is an initial vital step for a more comprehensive scientific analysis. If we know when a regime shift occurred this enables researchers to compare active and inactive epochs of climate states for future diagnostic and modeling studies. Traditionally, rates for extreme event counts have been modeled in a data-parameter two-layer hierarchical Bayesian framework. In this view, the rates are assumed to be invariant throughout the time. A few studies provided different Bayesian approaches to detecting and quantifying potential abrupt shifts in an extreme event series (e.g., Elsner et al., 2004; Zhao and Chu, 2006) by treating the Poisson rate as a random variable. However, as discussed in the "Introduction" section, each of them has some limitations.

A general 3-layer hierarchical Bayesian model – which includes data, parameter, and hypothesis – with a nested hypothesis space is built in this study. A seasonal extreme event count series is modeled as a Poisson process with a gamma distributed rate. We consider multiple candidate hypotheses, within each of which there presumably exists a certain number of abrupt shifts of the Poisson rate. We started with a simple, non-MCMC approach for detecting abrupt shift in the tropical cyclone series over the central North Pacific. The result shows that the likelihood of an abrupt shift on

tropical cyclone rates is around 1982. We then introduced a MCMC based detection method for the major hurricane series over the eastern North Pacific. The result suggests that there are two significant rate shifts for this series, one is an upshift in 1982 and the other is a downshift in 1999 with lower activity since then.

Due to the inability of a regular MCMC algorithm to effectively deal with the model selection problem, we resort to its extended version, the RJMCMC algorithm, and design an algorithm to automatically calculate the Bayesian inference of the 3-layer Bayesian model to efficiently solve the hypothesis competition problem. The RJMCMC algorithm was applied to three examples of extreme event series: the annual super typhoon counts over the WNP; the extremely heavy rainfall counts at Manoa in Oahu, Hawaii; and the annual heat wave counts in French. The results indicate that typhoon activity over the WNP is very likely to have undergone a decadal variation with two change-points occurring around 1972 and 1989; the average super-typhoon rate is 5.7 per year during the active 1960–1971 epoch, drops to 2.4 super-typhoons per year during the inactive 1972–1988 epoch, and then goes up to 5.0 super-typhoons per year from 1989 to 2006. The extreme rainfall occurrence frequency at Manoa in Hawaii has had three significant shifts between 1920 and 2009: before 1963, the heavy rainfall frequency for this site is about 8.8 counts per year; thereafter this rate increases to 13.1; from 1993 to 2003, the heavy rainfall frequency decreases to 7.0 counts a year; and starting from 2004, Manoa enters another heavy rainfall active period, recording 15.0 heavy events a year until the present. For the heat wave counts in French, there has been an abrupt jump since 1983, from 0.6 counts per year jumping to 1.8 counts per year.

This paper also shows how the complex individual, historical typhoon tracks over the vast western North Pacific and South China Sea can be reasonably clustered into a few distinct patterns, which may yield physical insights. In simple terms, cluster analysis is a statistical technique that objectively separates data into groups whose identities are not known in advance. It is the degree of similarity and difference among each individual track that is used to define the group and to assign the membership. We use a regression mixture model to separate the data into groups. Under this model, we consider the genesis location, pathway (i.e., straight or recurving), and shape of the tracks as key elements. Our results show that eight track patterns with three straight moving, four recurving, and one mixed straight moving/recurving appear to be plausible over the last 60 years. This is somewhat different from those of Camargo et al. (2007) and Kim et al. (2011), in which seven types were adopted as an optimal number. This difference may result from the different time periods used in each study.

The third application of the Bayesian paradigm to extreme events is the probabilistic forecasting of seasonal tropical cyclone frequency over the central North Pacific. By considering large-scale environmental variables prior to the hurricane season, a Poisson regression model is used to illustrate the mechanics of the forecasting approach. Specifically, the predictor variables include sea-level pressures, sea surface temperatures, vertical wind shear, low-level relative vorticity, and precipitable water. A Gibbs sampler based on the MCMC method is designed to integrate the desired posterior predictive distribution. Results from cross-validation suggest that the Bayesian, Poisson regression model is skillful in predicting seasonal TC frequency, with a correlation coefficient of 0.63 for 1966–2003. Moreover, the relative importance of each environmental variable in the forecasting strategy can be judged by the so-called Bayesian p-value. In this example, the sea-level pressure and, to a lesser extent, the relative vorticity, are identified as key predictors.

Along the TC forecasting based on the Poisson regression method, a new approach called the probit regression is demonstrated as a binary classification problem. The probit regression is adopted when the underlying TC rate is significantly below one, which is a very rare event. Following the framework for the conditional posterior distribution for a Poisson regression model, we developed a similar posterior distribution for a probit regression model explicitly.

References

Albert, J., Chib, S., 1993. Bayesian analysis of binary and polychotomous response data. J. Am. Stat. Assoc. 88, 669–679.

Berger, J.O., 1985. Statistical Decision Theory and Bayesian Analysis. Springer-Verlag. (617 pp.).

Berliner, L.M., Kim, Y., 2008. Bayesian design and analysis for super-ensemble based climate forecasting. J. Climate 21, 1891–1910.

Briggs, W.M., 2008. On the changes in number and intensity of North Atlantic tropical cyclones. J. Climate 21, 1387–1402.

Camargo, S.J., Robertson, A.W., Gaffney, S.J., Smyth, P., Ghil, M., 2007. Cluster analysis of typhoon tracks. Part I: general properties. J. Climate 20, 3635–3653.

Chand, S.S., Walsh, K.J.E., 2011. Forecasting tropical cyclone formation in the Fiji region: A probit regression approach using Bayesian fitting. Wea. Forecast 26 (2), 150–165.

Chand, S.S., Walsh, K.J.E., Chan, J.C.L., 2010. A Bayesian regression approach to seasonal prediction of tropical cyclones affecting the Fiji region. J. Climate 23, 3425–3445.

Chu, P.-S., 2002. Large-scale circulation features associated with decadal variations of tropical cyclone activity over the central North Pacific. J. Climate 15, 2678–2689.

Chu, P.-S., Zhao, X., 2004. Bayesian change-point analysis of tropical cyclone activity: the central North Pacific Case. J. Climate 17, 4893–4901.

Chu, P.-S., Zhao, X., 2007. A Bayesian regression approach for predicting seasonal tropical cyclone activity over the central North Pacific. J. Climate 20, 4002–4013.

Chu, P.-S., Zhao, X., Lee, C.-T., Lu, M.-M., 2007. Climate prediction of tropical cyclone activity in the vicinity of Taiwan using the multivariate least absolute deviation regression method. Terr. Atmos. Ocean. Sci. 18, 805–825.

Chu, P.-S., Zhao, X., Kim, J.-H., 2010a. Regional typhoon activity as revealed by track patterns and climate change. In: Elsner, J., et al. (Ed.), Hurricanes and Climate Change, vol. 2. Springer, pp. 137–148 (Chapter 8).

Chu, P.-S., Zhao, X., Ho, C.-H., Kim, H.-S., Lu, M.-M., Kim, J.-H., 2010b. Bayesian forecasting of seasonal typhoon activity: a track-pattern-oriented categorization approach. J. Climate 23, 6654–6668.

Collins, J.M., Mason, I.M., 2000. Local environmental conditions related to seasonal tropical cyclone activity in the northeastern Pacific basin. Geophys. Res. Lett. 27, 3881–3884.

Congdon, P., 2007. Bayesian Statistical Modeling, 2nd ed. Wiley Series in Probability and Statistics. (596 pp.).

Deser, C.A., Phillips, S., Hurrell, J.W., 2004. Pacific interdecadal climate variability: linkages between the tropics and the North Pacific during boreal winter since 1900. J. Climate 17, 3109–3124.

Dose, V., Menzel, A., 2004. Bayesian analysis of climate change impacts in phenology. Glob. Change Bio. 10, 259–272.

Elsner, J.B., Bossak, B.H., 2001. Bayesian analysis of U.S. hurricane climate. J. Climate 14, 4341–4350.

Elsner, J.B., Jagger, T.H., 2004. A hierarchical Bayesian approach to seasonal hurricane modeling. J. Climate 17, 2813–2827.

Elsner, J.B., Jagger, T.H., 2006. Prediction models for annual U.S. hurricane counts. J. Climate 19, 2813–2827.

Elsner, J.B., Niu, X., Jagger, T.H., 2004. Detecting shifts in hurricane rates using a Markov Chain Monte Carlo approach. J. Climate 17, 2652–2666.

Epstein, E.S., 1985. Statistical Inference and Prediction in Climatology: A Bayesian Approach. Meteor. Monogr., No. 42. Amer. Meteor. Soc (199 pp.).

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. Bayesian Data Analysis, 2nd ed. Chapman & Hall/CRC. (668 pp.).

Godsill, S.J., 2001. On the relationship between Markov Chain Monte Carlo methods for model uncertainty. J. Comput. Graph. Stat. 10, 1–19.
Green, P., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82, 711–732.
Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97–109.
Ho, C.-H., Kim, H.S., Chu, P.-S., Kim, J.-H., 2009. Seasonal prediction of tropical cyclone frequency over the East China Sea through a Bayesian Poisson-regression method. Asia-Pacific J. Atmos. Sci. 45, 45–54.
Jagger, T.H., Elsner, J.B., 2010. A consensus model for seasonal hurricane prediction. J. Climate 23, 6090–6099.
Kallache, M., Maksimovich, E., Michelangeli, P.-A., Naveau, P., 2010. Multi-model combination by a Bayesian hierarchical model: assessment of ice accumulation over the oceanic Arctic region. J. Climate 23, 5421–5436.
Khaliq, M.N., Ouarda, T.B.M.J., St.-Hilaire, A., Gachon, P., 2007. Bayesian changepoint analysis of heat spell occurrences in Montreal. Canada. Int. J. Climatol. 27, 805–818.
Kim, H.-S., Ho, C.-H., Kim, J.-H., Chu, P.-S., 2011. Pattern classification of typhoon tracks using the fuzzy *c*-means clustering method. J. Climate 24, 488–508.
Lang, M., Ouarda, T.B.M.J., Bobe'e, B., 1999. Towards operational guidelines for over-threshold modeling. J. Hydrol. 225, 103–117.
Lavielle, M., Labarbier, M., 2001. An application of MCMC methods for multiple change-points problem. Signal Process. 81, 39–53.
Leroy, S., 1998. Detecting climate signals: some Bayesian aspects. J. Climate 4, 640–651.
Lima, C.H.R., Lall, U., 2010. Spatial scaling in a changing climate: a hierarchical Bayesian model for non-stationary multi-site annual maximum and monthly streamflow. J. Hydrol. 383, 307–318.
Lu, M.-M., Chu, P.-S., Lin, Y.-C., 2010. Seasonal prediction of tropical cyclone activity in the vicinity of Taiwan using the Bayesian multivariate regression method. Wea. Forecasting 25, 1780–1795.
MacKay, D.J.C., 2003. Information Theory, Inference, and Learning Algorithms. Cambridge University Press. (628 pp.).
Min, S.-K., Hense, A., 2006. A Bayesian assessment of climate change using multimodel ensembles. Part I: global mean surface temperature. J. Climate 19, 3237–3256.
Neal, R.M., 1996. Bayesian Learning for Neural Networks. Springer, New York. (183 pp.).
Perreaulta, L., Berniera, J., Bobe'e, B., Parent, E., 2000a. Bayesian change-point analysis in hydrometeorological time series. Part 1. The normal model revisited. J. Hydrol. 235, 221–241.
Perreaulta, L., Berniera, J., Bobe'e, B., Parent, E., 2000b. Bayesian change-point analysis in hydrometeorological time series. Part 2. Comparison of change-point models and forecasting. J. Hydrol. 235, 242–263.
Raftery, A.E., 1996. Approximate Bayes factors and accounting for model uncertainty in generalized linear models. Biometrica 83, 251–266.
Renard, B., Lang, M., Bois, P., 2006. Statistical analysis of extreme events in a non-stationary context via a Bayesian framework: case study with peak-over-threshold data. Stoch Environ Res Risk Assess 21, 97–112.
Richardson, S., Green, P.J., 1997. On Bayesian analysis of mixtures with an unknown number of components. J. Royal Stat. Soc., Ser. B 59 (4), 731–792.
Ripley, B.D., 1987. Stochastic Simulation. John Wiley, New York. (237 pp.).
Rosbjerg, D., Madsen, H., 1996. The role of regional information in estimation of extreme point rainfalls. Atmospheric Research 42, 113–122.
Rotondi, R., 2002. On the influence of the proposal distributions on a reversible jump MCMC algorithm applied to the detection of multiple change-points. Comput. Stat. Data Anal. 40 (3), 633–653.
Shaw, S.L., 1981. A History of Tropical Cyclones in the Central North Pacific and the Hawaiian Islands: 1832–1979. Central Pacific Hurricane Center, NOAA/National Weather Service Forecast Office, Honolulu, HI. (137 pp.).
Solow, A.R., 1988. A Bayesian approach to statistical inference about climate change. J. Climate 1, 512–521.
Tebaldi, C., Mearns, L.O., Nychka, D., Smith, R.L., 2004. Regional probabilities of precipitation change: a Bayesian analysis of multimodel simulations. Geophys. Res. Lett. 31, L24213. doi:10.1029/2004GL021276.
Tebaldi, C., Smith, R.L., Nychka, D., Mearns, L.O., 2005. Quantifying uncertainty in projections of regional climate change: a Bayesian approach to the analysis of multimodel ensembles. J. Climate 18, 1524–1540.
Tomassini, L., Reichert, P., Knutti, R., Stocker, T.F., Borsuk, M.E., 2007. Robust Bayesian uncertainty analysis of climate system properties using Markov Chain Monte Carlo methods. J. Climate 20, 1239–1254.
Trotta, R., 2008. Bayes in the sky: Bayesian inference and model selection in cosmology. Contemp. Phys. 49, 71–104.
Tu, J.-Y., Chou, C., Chu, P.-S., 2009. The abrupt shift of typhoon activity in the vicinity of Taiwan and its association with Western North Pacific–East Asian climate change. J. Climate 22, 3617–3628.
Whitney, L.D., Hobgood, J.S., 1997. The relationship between sea surface temperatures and maximum intensities of tropical cyclones in the eastern North Pacific. J. Climate 10, 2921–2930.
Wilks, D.S., 2006. Statistical Methods in the Atmospheric Sciences, 2nd ed. Academic Press. (627 pp.).
Zhao, X., Cheung, L.W.K., 2007. Kernel-imbedded Gaussian processes for disease classification using microarray gene expression data. BMC Bioinforma. 8, 67.
Zhao, X., Cheung, L.W.K., 2011. Multi-class kernel-imbedded Gaussian processes for microarray data analysis. IEEE/ACM Trans. on Comput. Biol. & Bioinformatics 8, 1041–1053.
Zhao, X., Chu, P.-S., 2006. Bayesian multiple changepoint analysis of hurricane activity in the eastern North Pacific: a Markov Chain Monte Carlo approach. J. Climate 19, 564–578.
Zhao, X., Chu, P.-S., 2010. Bayesian changepoint analysis for extreme events (typhoons, heavy rainfall, and heat waves): a RJMCMC approach. J. Climate 23, 1034–1046.