Statistical Models for Infectious Disease Surveillance Counts

Mathias Hofmann

Dissertation
an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig–Maximilians–Universität
München

München, 16. November 2006

Statistical Models for Infectious Disease Surveillance Counts

Mathias Hofmann

Dissertation
an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Mathias Hofmann
aus München

München, den 16. November 2006

Erstgutachter: Prof. Dr Leonhard Held
Zweitgutachter: Prof. Dr. Ludwig Fahrmeir
Drittgutachter: Prof. Dr. Peter J. Diggle
Rigorosum: 30. Januar 2007

Vorwort

Diese Arbeit entstand während meiner Tätigkeit als Mitarbeiter im Sonderforschungsbereich 386 "Statistische Analyse diskreter Strukturen" am Institut für Statistik an der LMU München und wurde somit durch Mittel der Deutschen Forschungsgemeinschaft gefördert.

Zu allererst möchte ich meinem Doktorvater Leonhard Held für seine hervorragende Betreuung der Arbeit danken. Ohne sein Interesse und Engagement wäre diese Arbeit sicher nie entstanden.

Mein besonderer Dank gilt außerdem Michael Höhle, der mich während meiner gesamtem Arbeit mit Rat und Tat unterstützt hat und damit entscheidend zum Gelingen dieser Arbeit beigetragen hat.

Des weiteren möchte ich mich bei Volker Schmid für die gute Zusammenarbeit und meinen Gutachtern Ludwig Fahrmeir und Peter Diggle bedanken.

Ein großer Dank gilt meiner Familie und meinen Freunden für die vielen hilfreichen Ratschläge und Ermutigungen wenn es gelegentlich mal nicht so lief.

München, Februar 2007 Mathias Hofmann

Zusammenfassung

Modelle für Infektionskrankheitszahlen aus der Krankheitsüberwachung müssen die spezifischen Charakteristika dieser Daten berücksichtigen. Während sie über einen längeren Zeitraum ein reguläres, häufig saisonales Muster aufweisen, gibt es gelegentliche Unregelmäßigkeiten oder Ausbrüche.

Ein Modell, das ein Kompromiss zwischen einem mechanistischen und einem empirischen Modell ist, wird vorgeschlagen. Ein entscheidendes Konzept ist die Unterscheidung zwischen einer endemischen und einer epidemischen Komponente, was es ermöglicht das reguläre Muster von den Unregelmäßigkeiten und Ausbrüchen zu trennen. Das ist von besonderem Vorteil für die Ausbruchserkennung im Rahmen der Krankheitsüberwachung des öffentlichen Gesundheitswesens. Während die endemische Komponente parametergesteuert ist, basiert die epidemische Komponente auf einem beobachtungsgesteuerten Ansatz, einschließlich einer Autoregression auf vergangene Beobachtungen.

Eine besondere Herausforderung von Infektionskrankheitszahlen ist die Modellierung der Ausbrüche und Unregelmäßigkeiten in den Daten. Wir modellieren den Autoregressionsparameter der epidemischen Komponente durch ein bayesianisches Bruchpunktmodell, welches ein adaptives Maß an Glättung aufweist und in der Lage ist sowohl die Sprünge und schnellen Anstiege als auch die langsamen Rückgänge der Fälle zu modellieren. Während sich das Modell als allgemeiner Ansatz zur Modellierung von Infektionskrankheitszahlen verwenden lässt, ist es insbesondere geeignet für die Ausbruchserkennung im Rahmen der Krankheitsüberwachung des öffentlichen Gesundheitswesens. Des weiteren ermöglichen die Vorhersageeigenschaften des bayesianischen Bruchpunktmodells kurzfristige Vorhersagen der Kranheitsfälle, die von besonderem Interesse im öffentlichen Gesundheitswesen sind.

Eine sequentielle Schätzmethode des Modells durch einen ”particle filter” wird bereitgestellt, die für eine prospektive Analyse des Bruchpunktmodells, bedingt auf feste Werte der übrigen Parameter, verwendet werden kann, was von besonderem Vorteil für die Krankheitsüberwachung im öffentlichen Gesundheitswesen ist.

Eine geeignete multivariate Erweiterung wird bereitgestellt, die in der Lage ist die Interaktionen zwischen den Einheiten, z.B. Altersgruppen oder räumlichen Regionen, zu erklären. Eine Anwendung auf Influenza- und Meningokokkendaten zeigt, dass die gelegentlichen Meningokokkenausbrüche weitgehend durch den Einfluß von Influenza auf Meningokokken erklärt werden können. Das Risiko eines durch Influenza bedingten Meningokokkenausbruchs kann vorhergesagt werden. Der Vergleich der verschiedenen Modelle, einschließlich eines auf Gauss Markov Zufallsfeldern basierten Modells, zeigt, dass sowohl die Einbezie-

hung einer epidemischen Komponente, als auch eines zeitlich variierenden epidemischen Parameters die Modellanpassung und die Vorhersagefähigkeit des Modells verbessert.

Abstract

Models for infectious disease surveillance counts have to take into account the specific characteristics of this type of data. While showing a regular, often seasonal, pattern over long time periods, there are occasional irregularities or outbreaks.

A model which is a compromise between mechanistic models and empirical models is proposed. A key idea is to distinguish between an endemic and an epidemic component, which allows to separate the regular pattern from the irregularities and outbreaks. This is of particular advantage for outbreak detection in public health surveillance. While the endemic component is parameter-driven, the epidemic component is based on observation-driven approaches, including an autoregression on past observations.

A particular challenge of infectious disease counts is the modelling of the outbreaks and irregularities in the data. We model the autoregressive parameter of the epidemic component by a Bayesian changepoint model, which shows an adaptive amount of smoothing, and is able to model the jumps and fast increases as well as the smooth decreases in the data. While the model can be used as a generic approach for infectious disease counts, it is particularly suited for outbreak detection in public health surveillance. Furthermore, the predictive qualities of the Bayesian changepoint model allow for short term predictions of the number of disease cases, which are of particular public health interest.

A sequential update using a particle filter is provided, that can be used for a prospective analysis of the changepoint model conditioning on fixed values for the other parameters, which is of particular advantage for public health surveillance.

A suitable multivariate extension is provided, that is able to explain the interactions between units, e.g. age groups or spatial regions. An application to influenza and meningococcal disease data shows that the occasional outbreaks of meningococcal disease can largely be explained by the influence of influenza on meningococcal disease. The risk of a future meningococcal disease outbreak caused by influenza can be predicted. The comparison of the different models, including a model based on Gaussian Markov random fields shows that the inclusion of the epidemic component as well as a time varying epidemic parameter improves the fit and the predictive qualities of the model.

Contents

Chapter 1

Introduction

The history of infectious diseases is characterized by an everlasting fight between humans and the different types of viruses. While some diseases are widely under control, as polio or measles, new threats arise as HIV or new types of the consistently mutating influenza virus, e.g. the currently impending avian influenza.

Statistical models can help to control and prevent diseases. Outbreak detection systems, that are based on statistical models, can lead to an early detection of outbreaks. Control measures as vaccination or isolation and treatment of infected can then be initiated at an early stage and help to prevent further cases and to get the disease under control.

While bigger outbreaks will sooner or later be recognized, this is not necessarily the case for smaller outbreaks in less frequently observed diseases as hepatitis or meningitis, especially if cases spread over a wide area. Statistical models can detect such outbreaks and give rise to further investigations to find the reason of the disease, which can help to prevent further cases by eliminating the source.

Aside from these features, statistical models can help to understand the nature of the disease, e.g. the velocity of spread and ascertain factors, that have an influence on the disease, e.g. other types of diseases. Based on these models, the further spread of the disease can be predicted, which can give an idea of the magnitude of an outbreak.

We will propose a range of models, that partly address all considered features of infectious disease modelling.

1.1 Data

Counts of infectious diseases have particular characteristics, that have to be taken into account, when modeling this specific type of data. While these data often show a regular pattern over large periods, e.g. seasonality or trends, they have occasional irregularities or outbreaks. These two characteristics of infectious diseases data are related with two characteristics of infectious diseases: endemics and epidemics. Infectious diseases typically show both of them, when observed over a longer time period. It is therefore meaningful to distinguish between endemic and epidemic periods.

Outbreaks show different characteristics depending on the type of infectious disease. These characteristics depend mainly on the mode of transmission, where direct transmission, i.e. from person to person, and indirect transmission, e.g. via food, water or a vector as insects, can be distinguished. While directly transmitted diseases, as influenza, usually show outbreaks with an exploding number of cases, predominantly indirectly transmitted diseases, as the different types of hepatitis, usually show short or longer periods of a slightly increased number of cases, typically caused by a point source.

Another common feature of infectious disease counts is overdispersion with respect to the usual Poisson assumption. In directly transmitted diseases, this may be caused by the fact that the generation time of the disease does not equal the observation time. In indirectly transmitted diseases the regular pattern is usually given by a cumulation of smaller outbreaks and does therefore often not follow the Poisson assumption.

Beside these features, infectious disease counts, that usually arise in surveillance systems often show underreporting and reporting delays.

1.2 Models

1.2.1 Models for chronic diseases

Models for chronic diseases have found a brought interest in the last two decades. Although chronic diseases do not show the same characteristics as infectious diseases, in particular they do not show seasonality and epidemic outbreaks, these kind of models can be used for infectious diseases in the rare case that the data do not show seasonality and outbreaks, and are therefore of interest as a basis for more general infectious disease models.

The first models for spatial chronic disease data were developed for the purpose of disease mapping, which consists in smoothing disease rates and finding disease clusters,

which can be used to draw disease atlases and be helpful e.g. for etiological or health service research. One can distinguish two approaches for disease mapping, those based on Gaussian Markov random fields (GMRF) (e.g. Rue and Held, 2005) and those based on cluster models estimated via the reversible jump algorithm of Green (1995).

Model based disease mapping goes back to Clayton and Kaldor (1987) who smoothed the mortality rates of lip cancer observed in Scotland by a spatially structured effect using a GMRF prior. This model was extended by Besag et al. (1991) who introduced an additional unstructured effect, to adjust for unobserved heterogeneity in the data. This purely spatial model was extended to a space-time model assuming a linear (Bernardinelli et al., 1995; Xia and Carlin, 1998), and a structured (Knorr-Held and Besag, 1998) time trend. Additionally different types of space-time interactions were proposed, assuming linear time and unstructured space (Bernardinelli et al., 1995), linear time and structured space (Assunção, 2003), structured time and unstructured space (Xia and Carlin, 1998) unstructured/linear time and unstructured space (Sun et al., 2000) unstructured time and un-/structured space (Waller et al., 1997). Knorr-Held (2000) compares the four combinations of un-/structured time and space interactions.

Another purpose for modeling space-time data is the prediction of space-time trends (Berzuini et al., 1993; Besag et al., 1995; Knorr-Held and Rainer, 2001; Assunção, 2003; Schmid and Held, 2004). Particularly suited for this task are age period cohort (APC) models (Berzuini et al., 1993; Besag et al., 1995; Knorr-Held and Rainer, 2001; Schmid and Held, 2004), which will be described later in this section.

While models based on GMRFs show a constant amount of smoothing, given by the variance of the GMRF prior, cluster models allow the correlation structure to vary over space or time (Ferreira et al., 2002). A purely spatial cluster model is e.g. proposed by Knorr-Held and Raßer (2000). A space-time cluster model has recently been proposed by Yan and Clayton (2006).

A different purpose of models for chronic diseases is ecological regression, which aims to investigate the influence of risk factors, and the assessment of environmental justice by the introduction of covariates. Clayton et al. (1993) investigate the influence of urbanization and industrialization on cancer in Sardinia. Further models are proposed by Waller et al. (1997); Xia and Carlin (1998); Knorr-Held and Besag (1998); Natário and Knorr-Held (2003). A special type of models including covariate information are the age period cohort (APC) models which include these three effects, where period is the time of infection or death and cohort is the time of birth (Berzuini et al., 1993; Besag et al., 1995; Lagazio

et al., 2003; Knorr-Held and Rainer, 2001; Schmid and Held, 2004).

1.2.2 Models for infectious diseases

Models for infectious diseases can be divided into mechanistic and empirical models. Mechanistic models are not applicable to infectious disease count data, since they require to observe the complete infection process, including the exact infection time and duration and the number of susceptibles, which are only available in very special cases. Beside this, they are only applicable to directly transmitted diseases. However, given this information, mechanistic models provide insights about characteristics of diseases which may be helpful for building empirical models.

Mechanistic Models

Mechanistic models try to model the infection mechanism of a disease, based on data of a completely observed infection process. Given this information these models can give answer to questions as "How fast does the disease grow?" "How many people will be affected?" and "How can vaccination affect the spread?". We will introduce mechanistic models by the example of the SIR model, which assumes that a person is first susceptible (S), is infectious (I) for a while, if he or she gets infected, and is then recovered (R), i.e. immune or dead, and plays no further role in the spread of the disease. One of the central insights of the SIR model is that the speed of the spread of a disease depends of the number of infected and the number of susceptibles. Only if both numbers are high the disease can spread fast. Since the number of susceptibles in a closed population decreases with the spread of the disease, the lack of susceptibles stops the disease usually before the whole population has been infected. A key quantity in the SIR model is the basic reproduction number, which is the rate with which one infectious person infects other persons in a 'virgin' population, i.e. a population consisting entirely of susceptibles. This rate, together with the number of infectious and susceptibles, gives information about the number of infected by one infectious. Based on this rate, the vaccination coverage necessary to stop the disease and the final size of the disease can be estimated.

The SIR model assumes that a susceptible person gets immediately infected by the contact with an infectious person. Additionally it assumes that after having been infected a person can not become susceptible anymore. There is a variety of models assuming different characteristics of disease, e.g. the SEIR model, that includes a state exposed (E) or the SIS and SIRS models that allow a person to become susceptible after an infectious

period. Further models include heterogeneity of the population or spatial spread of the disease. For an overview of mechanistic modes see e.g. Diekmann and Heesterbeek (2000); Daley and Gani (1999); Becker (1989); Anderson and Britton (2000).

Empirical Models

Empirical models for infectious disease counts were mainly proposed for one or more of the following reasons: surveillance, ecological regression, description and prediction of the disease. These models have to deal with the particular characteristics of infectious disease counts, the most challenging being the epidemic periods.

One of the most popular approaches for the surveillance of infectious diseases based on count data was proposed by Stroup et al. (1989). They try to find outbreaks in the data by comparing the number of cases of the current month with a confidence interval based on the observed counts of the same month of the last 5 years and their surrounding month. This avoids the deal with the seasonality in the data. However, this approach does not account for outbreaks in the past data. This has the effect that past outbreaks make the system less sensitive for the detection of future outbreaks. Farrington et al. (1996) improved this model by allowing for a time trend and tackled the problem of past outbreaks by downweighting counts from possible past outbreaks. Watier and Richardson (1991) and Williamson and Hudson (1999) based surveillance systems on ARIMA models under the assumption of normal distributed data. While the inclusion of an autoregression and thereby dependence between the counts seems to be a sensible extension, the normality assumption does often not hold, since many diseases show small counts. Kleinman et al. (2004) proposed a generalized linear mixed model (GLMM) approach for the surveillance of space-time counts. The model assumes a binomial distribution for the counts and includes independent seasonal and spatial effects. However, it does not assess the problem of past outbreaks. A review and discussion of prospective statistical surveillance in public health is given in Sonesson and Bock (2003), however, the authors do not discuss the typical features of infectious disease counts, and most models considered do not take into account seasonality or past outbreaks in the data.

The most sophisticated models for infectious disease counts were proposed for the purpose of describing or predicting the disease. Zeger (1988) used a GLMM to analyse count data of polio infections in the USA from 1970-1983. These data were reanalysed be several authors using a GLMM, see Nelson and Leroux (2005) for an overview. Jørgensen et al. (1999) assumed a Poisson state space model where the mean depends on a Gamma Markov

process, which can be seen as a random walk with drift. While these models are able to model the seasonality and overdispersion in the data by seasonal covariates and random effects, the latent process may not be able to capture outbreaks in the data adequately. Lindbäck and Svensson (2001) proposed a log linear Poisson model for campylobacter data of Sweden, including a stepwise linear trend and a seasonal term with varying amplitude, peak and form from year to year. Counts from putative outbreak periods were excluded from the analysis. Mugglin et al. (2002) proposed a log linear Poisson model for space-time influenza data of Scotland assuming the logarithm of the mean to depend on a multivariate Gaussian AR(1) process, where the innovation can switch between 3 levels, an endemic level an epidemic level, in case of an outbreak, an a third level for the decline of the cases after the outbreak. Knorr-Held and Richardson (2003) propose a model for space-time meningococcal disease data distinguishing an endemic pattern for periods of no outbreaks and a "hyperendemic" pattern that models possible outbreaks in the data. While the endemic pattern is build in the spirit of chronic diseases models including structured time, space and seasonal effects, the hyperendemic pattern allows for an autoregression on functions of counts of the same and neighboring regions which can be switched on and off according to a two-stage hidden Markov model. Morton and Finkenstädt (2005) proposed a discrete time version of the SIR model allowing for immigration of cases from outside the considered region and underreporting in the data, see also Finkenstädt et al. (2002); Finkenstädt and Grenfell (2000). The influence of meningococcal disease on influenza is analysed by Hubert et al. (1992) and Jensen et al. (2004).

1.3 Scope of thesis

The thesis is organized as follows: Chapter 2 gives a short introduction to different simulation techniques for Bayesian inference that will be used in the following chapters: Markov chain Monte Carlo (MCMC) methods, including reversible jump MCMC and two sequential Monte Carlo (SMC) methods: a particle filter and the forward-backward algorithm.

Chapter 3 describes the Bayesian changepoint model, that will be used in Chapter 5-7, and the estimation using MCMC and SMC methods.

Chapter 4 is based on the model proposed in Held et al. (2005). This model is a compromise between a mechanistic and an empirical approach. While a mechanistic modelling of the data is not possible the approach is based on a branching process, which is an approximation of the mechanistic chain binomial model (see e.g. Anderson and Britton, 2000).

This allows to capture the characteristics of an infectious disease. A key idea of the model is the distinction between an endemic and an epidemic component. A Bayesian version of the model is established and compared with a model based on GMRF including space and time effects and seasonal covariates. Furthermore, a new approach for the epidemic component is proposed, showing the best fit, compared to the other approaches.

In Chapter 5 the time constant epidemic parameter of the univariate model considered in Chapter 4 is replaced by a time varying parameter following a Bayesian changepoint model. This allows the model to capture the epidemic characteristics of infectious diseases clearly better and makes the model particularly suited for outbreak detection in public health surveillance.

In Chapter 6 two types of sequential Monte Carlo methods for the estimation of the model proposed in Chapter 5 are considered: The forward-backward algorithm and the particle filter. The forward-backward algorithm can be used as an alternative to the reversible jump algorithm, applied in Chapter 5, for the update of the changepoint model within the MCMC algorithm. The particle filter can be used for a prospective analysis of the changepoint model conditioning on fixed values for the other parameters, which is of particular advantage for public health surveillance, where data arise sequentially.

In Chapter 7 a multivariate version of the model described in Chapter 5 is established, making use of the dependence structures in the epidemic component proposed in Chapter 4. We apply the model to study the influence of influenza on meningococcal disease and the spatial spread of influenza in the districts of Germany and compare it to the models considered in Chapter 4.

Chapter 2

Bayesian inference

In this chapter some basic ideas of Bayesian inference are introduced, that will be used throughout this thesis.

Bayesian inference is based on the subjectivist view of probability. From the subjectivist view the only relevant thing is uncertainty. "The actual fact of whether or not the events considered are in some sense determined, or known by other people, and so on, is of no consequence" (de Finetti, 1974). The uncertainty may be expressed by probability. In parametric models there is not only uncertainty about the data D but also about the parameters θ, which can both be expressed by probability. The starting point of Bayesian inference consists of setting up a joint distribution over all random quantities $P(D, \theta)$. We will refer to P as a probability as well as a density. This joint distribution is given by

$$P(D, \theta) \quad = \quad P(D|\theta)P(\theta),$$

where $P(D|\theta)$ is the likelihood and $P(\theta)$ is the *prior* distribution. Thus in Bayesian inference not only a distribution for the data, but also for the parameters, has to be found. Having observed the data D, the interest in Bayesian inference is the distribution of the parameters given the data $P(\theta|D)$, which is called the *posterior* distribution. This can be found by Bayes theorem,

$$P(\theta|D) \quad = \quad \frac{P(D|\theta)P(\theta)}{P(D)} = \frac{P(D|\theta)P(\theta)}{\int P(D|\theta)P(\theta)d\theta}.$$

The posterior distribution is, up to the marginal distribution of the data in the denominator, also called the marginal likelihood, proportional to the joint distribution of data and

parameters, or their representation by the likelihood and the prior distribution,

$$P(\theta|D) \propto P(D|\theta)P(\theta).$$

The marginal likelihood does not depend on the parameters and is therefore also called normalization constant. Any features of the posterior distribution that may be of interest, e.g. moments, quantiles or highest posterior density regions can be calculated by the posterior expectation of functions of θ,

$$E(f(\theta)|D) = \frac{\int P(D|\theta)P(\theta)f(\theta)d\theta}{\int P(D|\theta)P(\theta)d\theta}.$$

For further details on Bayesian inference see Bernardo and Smith (1994).

2.1 Markov chain Monte Carlo

In most practical applications neither the posterior distribution nor the posterior expectation of functions of θ can be calculated analytically, since one has to solve the integrals in the expressions. Probably the most universal way to solve this problem are Markov chain Monte Carlo (MCMC) methods. This allows to approximate the posterior distribution by a sample $X_t, t = 1, \ldots, n$ from the posterior distribution $P(\theta|D)$, where the sample size n can be determined by the user. Based on this sample features of the posterior distribution can be approximated by

$$E(f(\theta)|D) \approx \frac{1}{n}\sum_{t=1}^{n} f(X_t).$$

One can usually not sample from the posterior distribution directly, since it is not possible to evaluate the integral of the normalization constant. The idea of MCMC is to construct a Markov chain that converges to a stationary distribution which coincides with the posterior distribution. Once converged the samples of the Markov chain come from the posterior distribution. Such a Markov chain can be constructed without knowledge about the normalization constant using the Metropolis-Hastings (MH) algorithm (Metropolis et al., 1953; Hastings, 1970).

2.1.1 Metropolis-Hastings algorithm

Starting from some staring value X_0, at each time t, a candidate X^* is sampled from a proposal distribution $q(X^*|X_t)$, which may depend of the current state X_t. This candidate is then accepted with probability

$$\alpha(X_t, X^*) = \min\left(1, \frac{P(D|X^*)P(X^*)}{P(D|X_t)P(X_t)}\frac{q(X_t|X^*)}{q(X^*|X_t)}\right).$$

If the candidate is accepted the next state becomes $X_{t+1} = X^*$. If the candidate is rejected, the chain does not move, $X_{t+1} = X_t$. The second term in the minimum function is the posterior ratio of the candidate and the current state multiplied by the proposal ratio. The normalization constant of the posterior distribution cancels out in the acceptance rate. The so constructed Markov chain converges to the posterior distribution.

2.1.2 Single-component Metropolis-Hastings

Instead of updating all parameters θ en block, it is possible to divide θ to components $\theta_1, \ldots, \theta_h$ and update the components one by one. This is often computationally more efficient. A single-component Metropolis-Hastings step consists of h single steps, i.e. one for each component. Assume that all components up to i are already updated. A candidate X_i^* is sampled from a proposal distribution $q_i(X_i^*|X_{t,i}, X_{t,-i})$, that may depend on the current state of component i, $X_{t,i}$ and the current state of the other components $X_{t,-i} = (X_{t+1,1}, \ldots, X_{t+1,i-1}, X_{t,i+1}, \ldots, X_{t,h})$. The candidate is accepted with probability

$$\alpha(X_{t,i}, X_{t,-i}, X_i^*) = \min\left(1, \frac{P(X_i^*, X_{t,-i}|D)q(X_{t,i}|X_i^*, X_{t,-i})}{P(X_{t,i}, X_{t,-i}|D)q(X_i^*|X_{t,i}, X_{t,-i})}\right).$$

If the candidate is accepted the next state becomes $X_{t+1,i} = X_i^*$. If the candidate is rejected, the chain does not move, $X_{t+1,i} = X_{t,i}$. The posterior ratio reduces to the ratio of the full conditional distributions $P(\theta_i|\theta_{-i}, D)$.

2.1.3 Gibbs sampler

If the full conditional of component i of the single-component Metropolis-Hastings algorithm, $P(\theta_i|D, \theta_{-i}) \propto P(D|\theta_i)P(\theta_i)$, is known, the update of this component can be simplified. Using this distribution as proposal distribution, the terms in the acceptance probability cancel down and the acceptance probability is equal to unity. The step then

just consists of sampling from the full conditional distribution.

2.1.4 The reversible jump algorithm

The reversible jump algorithm proposed by Green (1995) is a generalization of the Metropolis-Hastings algorithm to the case that the dimension of the parameters is not fixed. The algorithm works similar to the MH algorithm. Let $P(\theta|D)$ be the posterior distribution of the parameter vector θ. The dimension of the parameters k is part of the parameter vector θ. A candidate X^* is proposed from a proposal distribution $q(X^*|X)$. The time index t of the Markov chain is omitted here. In contrast to the MH algorithm this candidate may be from a parameter vector θ^*, that is of different dimension than the parameter vector θ of the current state X. The candidate is then accepted with probability

$$\alpha(X, X^*) = \min\left(1, \frac{P(X^*|D)q(X|X^*)}{P(X|D)q(X^*|X)}\right),$$

or otherwise the current state is retained. A common way to propose a candidate of different dimension is to first propose a new dimension k^* from a proposal $q(k^*|k)$ and then the candidate for the other parameters X^*_{-k} conditioned on the proposed dimension $q(X^*_{-k}|X_{-k}, k^*)$. The proposal distribution has then the form $q(X^*_{-k}|X_{-k}, k^*)q(k^*|k)$.

It is often helpful to let the proposal depend on the current state. However, in the case that the candidate is of different dimension, it is not straightforward how the proposal distribution should depend on the current state. A way has to be defined how the proposal distribution should depend on the current state and this has to be done in a way that detailed balance is retained. Beside this, the acceptance probability has to be adjusted by an additional term.

However, it is not always necessary or helpful to let the proposal depend on the current state. Especially when the full conditional is a known distribution one can use this as proposal distribution as in the Gibbs sampler.

2.1.5 Effective sample size

Since MCMC samples are usually correlated, it is useful to estimate the effective sample size (ESS) (see Kass et al., 1998). The effective sample size is an estimate of the number of independent samples needed to obtain a parameter estimate with the same precision as the MCMC estimate considered based on M dependent samples. For one parameter this is

calculated as

$$ESS \;=\; M/(1+2\sum_{j=1}^{k}\rho(k)),$$

where M is the sample size and $\sum_{j=1}^{k}\rho(k)$ is the sum of the first k sample autocorrelations, where k is chosen based on the initial monotone sequence estimator (Geyer, 1992).

2.1.6 Deviance information criterion

To assess the model fit the deviance can be used, which is based on the log-likelihood,

$$D = -2\log(L).$$

If the data $\boldsymbol{Y}$ depend on some index r, e.g. time or space, and are assumed to be independent given the parameters $\boldsymbol{\theta}$, the deviance can be factorized,

$$D = -2\log(L) = -2\sum_{r} d_r^2,$$

with the squared deviance residuals defined as

$$d_r^2 \;=\; \log(P(Y_r|\boldsymbol{\theta})).$$

Model comparison can be based on the deviance information criterion (DIC) proposed in Spiegelhalter et al. (2002), which combines the posterior mean deviance $\bar{D}$ with a measure of complexity p_D, that penalizes overfitting. This is defined as

$$p_D \;=\; \bar{D} - D(\bar{\boldsymbol{\theta}}),$$

where $\boldsymbol{\theta}$ are the parameters of interest, that should be chosen with respect to the purpose of the investigation. The DIC can then be calculated as the sum of the posterior mean of the Deviance $\bar{D}$ and p_D,

$$DIC \;=\; \bar{D} + p_D.$$

For further details on MCMC see Gilks et al. (1996) or Denison et al. (2002).

2.2 Sequential Monte Carlo

An alternative to MCMC methods, in some special cases, e.g. if the model is a Markov state-space or a hidden Markov model, are sequential Monte Carlo methods. These methods estimate the posterior distribution sequentially over time. We will consider two types of sequential Monte Carlo methods: *particle filters* and the *forward-backward algorithm*.

2.2.1 Particle filter

If the model is a Markov state-space model, particle filters can be used to update the model sequentially in time. This is especially of advantage if the data are observed sequentially in time. It is then not necessary to estimate the complete model each time a new observation is obtained, as it is necessary e.g. in the case of the MCMC algorithm. Instead, the estimate of the model up to the current time point can be based on the estimate up to the last time point. We will outline some basic ideas of particle filters that can be used to estimate a Markov state-space model sequentially in time.

Markov state-space models

A Markov state-space model assumes that the observations y_t, $t = 1, \ldots, n$ depend on latent parameters x_t, $t = 1, \ldots, n$. The latent parameters or hidden states build a Markov process, i.e. $P(x_t|x_{1:t-1}) = P(x_t|x_{t-1})$, $t = 2, \ldots, n$, with initial distribution $P(x_1)$, where $x_{1:t}$ stands for $(x_1, \ldots, x_t)$. The observations y_t, $t = 1, \ldots, n$ are assumed to be independent conditioned on the hidden states x_t, $t = 1, \ldots, n$, where the full conditional distribution of y_t is $P(y_t|y_{1:n}, x_{1:n}) = P(y_t|x_t)$. The model has then the following components

$$\begin{aligned} P(x_1) & \\ P(x_t|x_{t-1}) & \quad t = 2, \ldots, n \\ P(y_t|x_t) & \quad t = 1, \ldots, n, \end{aligned}$$

which determine the joint distribution

$$P(x_{1:n}, y_{1:n}) = P(x_1) \prod_{t=2}^{n} P(x_t|x_{t-1}) \prod_{t=1}^{n} P(y_t|x_t),$$

where the latent parameters x_t, $t = 1, \ldots, n$ build the signal process and the observations y_t, $t = 1, \ldots, n$ build the observation process. The structure of a Markov state-space

model can be visualized by the graphical model shown in Figure 2.1. For more details on graphical models see e.g. Cowell et al. (1999). The full conditional of any node just depends on the nodes which it is connected to, e.g. $P(y_t|y_{1:n}, x_{1:n}) = P(y_t|x_t)$ and $P(x_t|y_{1:n}, x_{1:n}) = P(x_t|y_t, x_{t-1}, x_{t+1})$. Additionally the graph implies that $P(x_t|x_{t+1}, y_{1:n}) = P(x_t|x_{t+1}, y_{1:t})$ and $P(x_t|x_{t-1}, y_{1:n}) = P(x_t|x_{t-1}, y_{t:n})$.

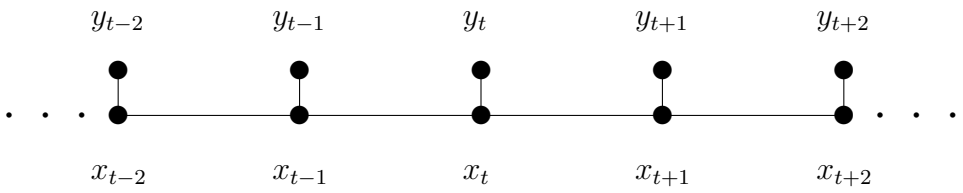


Figure 2.1: Graphical model of the Markov state-space models. The circles represent the nodes and the lines between the nodes represent undirected links.

Importance sampling

A weighted sample of the posterior distribution $P(x_{1:t}|y_{1:t})$ can be obtained by importance sampling. A sample of size N, $x_{1:n}^{(i)}$, $i = 1, \ldots, N$, is obtained from a importance distribution, also called proposal distribution, $q(x_{1:n}|y_{1:n})$, which may or may not depend on $y_{1:n}$. The N samples or *particles* are then weighted by the normalized importance weights

$$\tilde{w}(x_{1:n}^{(i)}) = \frac{w(x_{1:n}^{(i)})}{\sum_{j=1}^{N} w(x_{1:n}^{(j)})}$$

which are obtained from the importance weights

$$w(x_{1:n}^{(i)}) = \frac{P(x_{1:n}^{(i)}|y_{1:n})}{q(x_{1:n}^{(i)}|y_{1:n})}.$$

Features of the posterior distribution can then be approximated by a weighted average of functions of the samples,

$$E(f(x_{1:n})|y_{1:n}) \approx \sum_{i=1}^{N} \tilde{w}(x_{1:n}^{(i)}) f(x_{1:n}^{(i)}).$$

Sequential importance sampling

By using a proposal distribution with the following property

$$q(x_{1:t}|y_{1:t}) = q(x_{1:t-1}|y_{1:t-1}) q(x_t|x_{1:t-1}, y_{1:t}),$$

it is possible to calculate the importance weights recursively in time,

$$\tilde{w}(x_{1:t}^{(i)}) = \tilde{w}(x_{1:t-1}^{(i)}) \frac{P(y_t|x_t^{(i)}) P(x_t^{(i)}|x_{t-1}^{(i)})}{q(x_t^{(i)}|x_{1:t-1}^{(i)}, y_{1:t})}. \tag{2.1}$$

One can then estimate the posterior sequentially in time. A weighted sample $x_1^{(i)}$, $i = 1, \ldots, N$ of the posterior up to time 1, $P(x_1|y_1)$, is obtained using a proposal $q(x_1|y_1)$. Given that a weighted sample $x_{1:t-1}^{(i)}$, $i = 1, \ldots, N$ of the posterior up to time $t-1$, $P(x_{1:t-1}|y_{1:t-1})$, has already been calculated, a weighted sample $x_{1:t}^{(i)}$, $i = 1, \ldots, N$ of the posterior up to time t, $P(x_{1:t}|y_{1:t})$, can be obtained by extending the sample up to time $t-1$, $x_{1:t-1}^{(i)}$, $i = 1, \ldots, N$, by a sample $x_t^{(i)}$, $i = 1, \ldots, N$ from $q(x_t|x_{1:t-1}, y_{1:t})$. The weights can then be calculated using Equation (2.1).

Resampling step

A problem that arises, when the weights are calculated recursively is that the weights become more and more skewed. For the sample most of the probability mass will then concentrate on few samples, while the other samples have weights close to 0. A way to overcome this problem is to include a resampling step. This creates unweighted samples from the posterior by resampling from the weighted samples according to the importance weights. The weights $\tilde{w}(x_{1:t}^{(i)})$, $i = 1, \ldots, N$ are replaced by integer valued $N(x_{1:t}^{(i)})$, $i = 1, \ldots, N$, where $\sum_{i=1}^{N} N(x_{1:t}^{(i)}) = N$. Then the samples are replicated according to $N(x_{1:t}^{(i)})$. If $N(x_{1:t}^{(i)}) = 0$ the sample dies out. There are several ways to obtain the $N(x_{1:t}^{(i)})$, $i = 1, \ldots, N$. The most popular is the one proposed by Gordon et al. (1993), which corresponds to sampling the $N(x_{1:t}^{(i)})$, $i = 1, \ldots, N$ from a multinomial distribution with parameters N

and $\tilde{w}(x_{1:t}^{(i)})$, $i = 1, \ldots, N$.

The particle filter including an importance sampling step and a resampling step is called the bootstrap filter.

Markov transition step

While in sequential importance sampling the distribution of the weights becomes more skewed, the further one gets from the starting time point of the model, using the bootstrap filter one has the reverse problem: the number of distinct samples gets smaller with increasing distance to the end time point, since in every step some of the samples die out. To overcome this problem Andrieu et al. (2001) propose to include a Markov transition step, e.g. a MCMC step. For every of the N samples, a new sample is obtained e.g. by the MH algorithm, where the proposal is conditioned on the current sample. Since the distribution of the old sample is from the posterior distribution, which is the stationary distribution of the MCMC algorithm, the new sample is from the posterior distribution as well. However, the new samples are likely to be distinct after the Markov transition step.

The particle filter algorithm is summarized in Figure 2.2. The weights $\tilde{w}(x_{1:t-1}^{(i)})$ do not appear in (2.2) since the weights of the samples are uniform after the resampling step. Figure 2.3 illustrates the procedure of the considered particle filter. For further details on particle filters see e.g. Doucet et al. (2001b).

2.2.2 Forward-backward algorithm

If the model is a hidden Markov model, the forward-backward algorithm can be used to update the model sequentially. A hidden Markov model is a Markov state space model, where the parameters build a Markov chain with finite state space $\{1, \ldots, S\}$. Since a hidden Markov model is a special case of a Markov state space model, it has the same graph, shown in Figure 2.1.

The forward-backward algorithm consists of two steps: a forward and a backward step. In the forward step the distributions that are required for the backward step are calculated sequentially in time. In the backward step, the distributions that were calculated in the forward step are used to calculate the posterior distribution of the parameters sequentially in time.

Forward step

In the forward step the distributions $P(x_t, y_{1:t})$, $t = 1, \ldots, n$ are calculated. Note, that these are discrete distributions with finite sample space $\{1, \ldots, S\}$. Due to the structure of the hidden Markov model the following distributions are already known: The prior distributions given by $P(x_1)$ and $P(x_t|x_{t-1})$, $t = 2, \ldots, n$, and the distributions of the observations $P(y_t|x_t)$, $t = 1, \ldots, n$. Using these distributions, and given that $P(x_{t-1}, y_{1:t-1})$ has already been calculated, we can calculate

$$\begin{aligned} P(x_t = s, x_{t-1} = r, y_{1:t}) &= P(y_t|x_t = s, x_{t-1} = r, y_{1:t-1})P(x_t = s|x_{t-1} = r, y_{1:t-1}) \\ &\quad \cdot P(x_{t-1} = r, y_{1:t-1}) \\ &= P(y_t|x_t = s)P(x_t = s|x_{t-1} = r)P(x_{t-1} = r, y_{1:t-1}), \end{aligned}$$

for $r, s = 1, \ldots, S$. Since the state space is finite this distribution can easily be marginalized with respect to x_{t-1} and we get

$$P(x_t = s, y_{1:t}) = \sum_{r=1}^{S} P(y_t|x_t = s)P(x_t = s|x_{t-1} = r)P(x_{t-1} = r, y_{1:t-1}),$$

for $s = 1, \ldots, S$. So, the distributions given by $P(x_t, y_{1:t})$, $t = 2, \ldots, n$ can be calculated recursively, starting with

$$P(x_1 = s, y_1) = P(y_1|x_1 = s)P(x_t = s).$$

As a byproduct the marginal likelihood of the observations can be calculated as

$$P(y_{1:n}) = \sum_{r=1}^{S} P(x_n = r, y_{1:n}).$$

Backward step

In the backward step, the distributions that were calculated in the forward step are used to calculate the posterior distribution $P(x_{1:n}|y_{1:n})$ of the parameters sequentially in time. First, the distributions $P(x_n = r|y_{1:n})$ and $P(x_{t-1} = r|x_t = s, y_{1:n})$, $t = n-1, \ldots, 1$, are calculated. The distribution for $t = n$ can be calculated as

$$P(x_n = r|y_{1:n}) = \frac{P(x_n = r, y_{1:n})}{P(y_{1:n})}.$$

The rest of the distributions are calculated by

$$\begin{aligned} P(x_{t-1} = r|x_t = s, y_{1:n}) &= P(x_{t-1} = r|x_t = s, y_{1:t}) \\ &= \frac{P(x_t = s, x_{t-1} = r, y_{1:t})}{P(x_t = s, y_{1:t})} \\ &= \frac{P(y_t|x_t = s)P(x_t = s|x_{t-1} = r)P(x_{t-1} = r, y_{1:t-1})}{P(x_t = s, y_{1:t})}, \end{aligned}$$

for $t = n-1, \ldots, 1$. The backward step makes thereby use of the following property of the hidden Markov model:

$$P(x_{t-1} = r|x_t = s, y_{1:n}) = P(x_{t-1} = r|x_t = s, y_{1:t}).$$

The posterior distribution can then be calculated as

$$P(x_{1:n} = s_{1:n}|y_{1:n}) = P(x_n = s_n|y_{1:n}) \prod_{t=1}^{n-1} P(x_{t-1} = s_{t-1}|x_t = s_t, y_{1:n})$$

for $s_t = 1, \ldots, S$, $t = 1, \ldots, n$. If the algorithm is part of a MCMC algorithm, samples of the posterior of the parameters x_t, $t = 1, \ldots, n$ can be obtained by recursively sampling from the distributions $P(x_n|y_{1:n})$ and $P(x_{t-1}|x_t, y_{1:n})$, $t = n-1, \ldots, 1$. Due to the symmetry of the dependence structure of the model one could also start the algorithm at the end, i.e. at $t = n$, instead of the beginning, $t = 1$. For further details on the forward-backward algorithm see e.g. Scott (2002).

1. Initialisation
   - Start at $t = 1$.
2. Importance sampling step
   - If $t = 1$: For $i = 1, \ldots, N$, sample $\tilde{x}_1^{(i)} \sim q(x_1|y_1)$.
   - If $t > 1$: For $i = 1, \ldots, N$, sample $\tilde{x}_t^{(i)} \sim q(x_t|x_{1:t-1}, y_{1:t})$ and set $\tilde{x}_{1:t}^{(i)} = (x_{1:t-1}^{(i)}, \tilde{x}_t^{(i)})$.
   - For $i = 1, \ldots, N$, calculate the importance weights

$$\tilde{w}(x_{1:t}^{(i)}) = \frac{P(y_t|x_t^{(i)})P(x_t^{(i)}|x_{t-1}^{(i)})}{q(x_t^{(i)}|x_{1:t-1}^{(i)}, y_{1:t})}. \tag{2.2}$$

   - Normalize the importance weights.
3. Resampling step
   - Sample N integers $(N(x_{1:t}^{(1)}), \ldots, N(x_{1:t}^{(N)})) \sim \mathrm{Mu}(N, \tilde{w}_{1:t}^{(i)})$.
   - Obtain $\ddot{x}_{1:t}^{(i)}$, $i = 1, \ldots, N$ by replicating the particles $\tilde{x}_{1:t}^{(i)}$, $i = 1, \ldots, N$ according to the integers $N(x_{1:t}^{(i)})$, $i = 1, \ldots, N$.
4. Markov transition step
   - For $i = 1, \ldots, N$, sample $x_{1:t}^{(i)} \sim \Xi(x_{1:t}|\ddot{x}_{1:t}^{(i)})$, where $\Xi(\cdot|\cdot)$ is a transition kernel of invariant distribution $P(x_{1:t}|y_{1:t})$, as the transition kernel of the MCMC algorithm.
   - Set $t \leftarrow t + 1$ and go to step 2.

Figure 2.2: The considered particle filter.

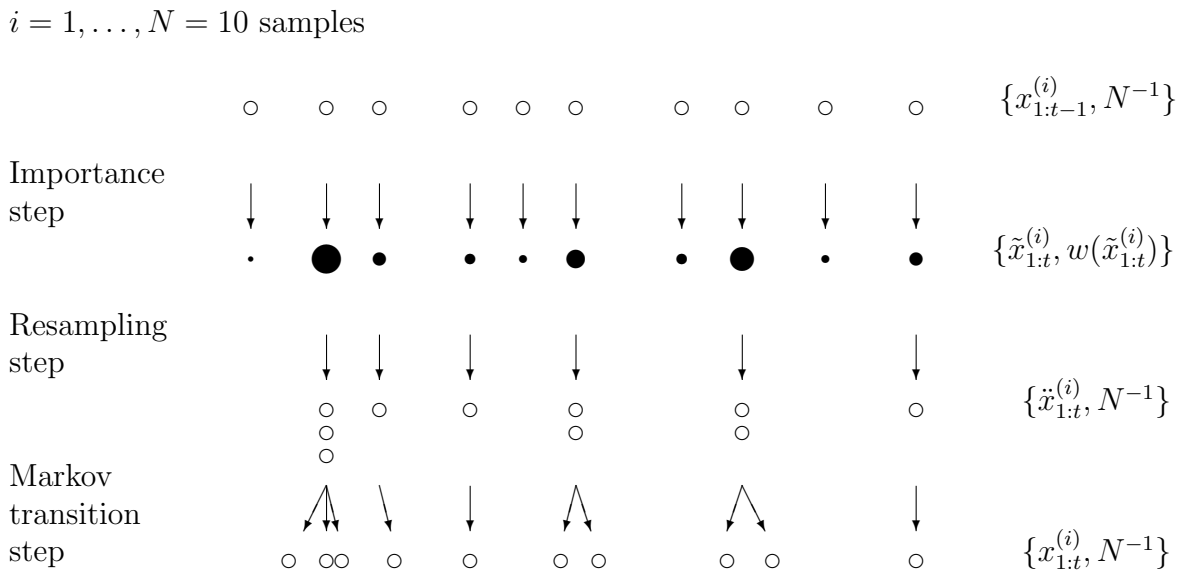


Figure 2.3: Starting with $N = 10$ unweighted and distinct samples $x_{1:t-1}^{(i)}$, $i = 1, \ldots, N$ of the posterior distribution $P(x_{1:t-1}|y_{1:t-1})$ at time $1 : t-1$, $N = 10$ unweighted and distinct sample $x_{1:t}^{(i)}$, $i = 1, \ldots, N$ of the posterior distribution $P(x_{1:t}|y_{1:t})$ at time $1 : t$ are obtained after applying the three steps of the particle filter: the importance step, the resampling step and the Markov transition step. The importance step gives weighted samples $\tilde{x}_{1:t}^{(i)}$, $i = 1, \ldots, N$ of the posterior distribution $P(x_{1:t}|y_{1:t})$ at time $1 : t$. The resampling step gives unweighted, but usually not distinct samples $\ddot{x}_{1:t}^{(i)}$, $i = 1, \ldots, N$ of the posterior distribution $P(x_{1:t}|y_{1:t})$ at time $1 : t$. The Markov transition step gives unweighted and distinct samples $x_{1:t}^{(i)}$, $i = 1, \ldots, N$ of the posterior distribution $P(x_{1:t}|y_{1:t})$ at time $1 : t$.

Chapter 3

Bayesian changepoint models

The analysis of Bayesian changepoint models, assuming an unknown number of changepoints, goes back to Yao (1984) who estimates a discrete-time changepoint model with Gaussian noise by a sequential algorithm. Barry and Hartigan (1993) propose a sequential algorithm for the estimation of a discrete-time changepoint model based on the product partition model proposed by Barry and Hartigan (1992). A continuous-time changepoint model was analysed by Green (1995) using the reversible-jump algorithm. Fearnhead (2006) extends the sequential algorithm of Barry and Hartigan (1993) to the case of prior that can be factorized to a prior on the number of changepoints and a prior on the location of the changepoints given the number of changepoints, as used in Green (1995).

Changepoint models can be used to estimate effects that vary over time. In contrast to random walks or splines, that assume a constant correlation structure, changepoint models allow the correlation structure to vary over time (Ferreira et al., 2002). In other words, the amount of smoothing is adaptive rather than constant. An additional advantage of changepoint models is that no prior for the correlation structure has to be assumed. For more details on changepoint models see e.g. Denison et al. (2002).

We will consider the following model for the data $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$,

$$Y_t \sim \text{Po}(\lambda_t X_t), \qquad t = 1, \ldots, n,$$

where X_t is a positive valued covariate and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$ is assumed to follow a changepoint model with unknown number K and locations $\theta^{(1)} < \ldots < \theta^{(K)}$ of changepoints. For convenience we define $\theta^{(0)} = 0$, $\theta^{(K+1)} = n$ and $\boldsymbol{\theta} = (\theta^{(0)}, \ldots, \theta^{(K+1)})$. The parameters $\boldsymbol{\lambda}$

are then defined as

$$\lambda_t \quad = \quad \lambda^{(k)} \quad \text{if} \quad t = \theta^{(k-1)} + 1, \ldots, \theta^{(k)}$$

One value of the changepoint effect is then a step function, as shown in Figure 3.1. We will call the data vectors $S^{(k)} = (Y_{\theta^{(k-1)}+1}, \ldots, Y_{\theta^{(k)}})$, that lie between two successive changepoints, segments, and the rates of the segments $\lambda^{(k)}$ segment means or segment rates. The segments are assumed to be independent conditioned on the changepoints. Estimating a changepoint model is a model selection problem, since the number $2K+1$ of changepoints $\theta^{(1)}, \ldots, \theta^{(K)}$ and rates $\lambda^{(1)}, \ldots, \lambda^{(K+1)}$, i.e. the number of parameters, is assumed to be unknown. However, instead of a decision for one model or number of changepoints, as usually done in model selection, one can get a distribution of the different models or number of changepoints. More precisely one gets the posterior distribution for the number of changepoints K, the changepoints $\boldsymbol{\theta}$ and the rates $\boldsymbol{\lambda}$. The effect is then the marginal posterior distribution of the rates $\boldsymbol{\lambda}$, which is a mixture of different step functions. However, it may also be interesting to look at the marginal distribution of the number of changepoints K or the location of the changepoints $\boldsymbol{\theta}$, or even on the conditional distributions, e.g. the distribution of the rates given a concrete number of changepoints, e.g. the number of changepoints with the highest posterior probability.

For $\lambda^{(k)}$, $k = 1, \ldots, K+1$ we assume independent Gamma priors, which are conjugate to the Poisson distribution of the data,

$$\lambda^{(k)} \quad \sim \quad \text{Ga}(\alpha, \beta), \quad k = 1, \ldots, K+1$$

For the number of changepoints and for the location given the number of changepoints we assume uniform priors,

$$P(K) \quad = \quad \frac{1}{n}, \quad K = 0, 1, \ldots, n-1 \tag{3.1}$$

$$P(\boldsymbol{\theta}|K) \quad = \quad \binom{n-1}{K}^{-1}. \tag{3.2}$$

The prior probability of a changepoint at a certain location t conditioned on a certain number of changepoints is then

$$P(t \in \boldsymbol{\theta}|K) = \frac{K}{n-1} \tag{3.3}$$

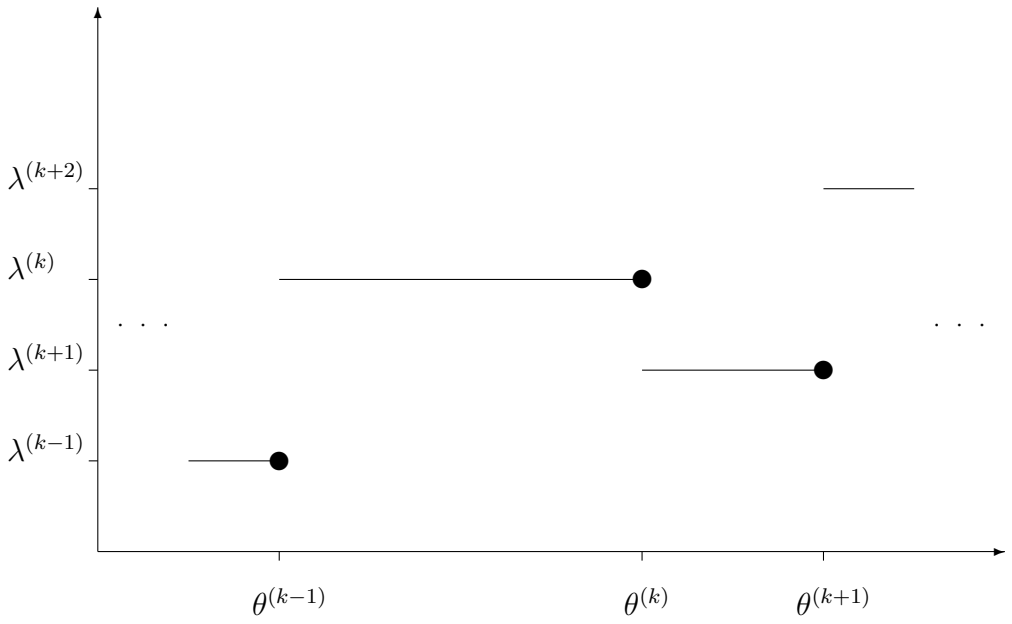


Figure 3.1: Structure of the changepoint effect.

and the unconditioned prior probability of a changepoint at a certain location t is hence

$$P(t \in \boldsymbol{\theta}) \;=\; \sum_{K=0}^{n-1} \frac{K}{n-1} \cdot \frac{1}{n} = \frac{1}{2},$$

(Held et al., 2006).

3.1 Model extensions

Additionally we will consider two extensions of the model. First we introduce a hyperprior on the second parameter of the prior of $\lambda^{(k)}$,

$$\beta \sim \text{Ga}(a, b),$$

resulting in a Gamma-Gamma prior for $\lambda^{(k)}$ (see Bernardo and Smith, 1994, p. 120). This gives a more robust estimate of the rates $\lambda^{(k)}$.
Secondly, we adjust for overdispersion by introducing time-dependent random effects ω_t,

$t = 1, \ldots, n$, that act multiplicative on the the mean $\mu_t = \lambda_t X_t$,

$$Y_t \sim \text{Po}(\omega_t \lambda_t X_t), \qquad t = 1, \ldots, n.$$

The random effects are assumed to be Gamma distributed $\omega_t \sim \text{Ga}(\psi, \psi)$ with ψ as shape and rate parameter. The marginal distribution of $Y_t|\mu_t$, integrating out ω_t, is then a Negative binomial distribution with mean μ_t and dispersion parameter ψ, $Z_t|\mu_t \sim \text{Nb}(\mu_t, \psi)$, which has variance $\sigma_t^2 = \mu_t(1-\mu_t/\psi)$. A smaller value of ψ corresponds to a higher amount of overdispersion. For $\psi \to \infty$ the variance converges to the mean $\sigma_t^2 \to \mu_t$ and we get back to the Poisson case. Figure 3.2 shows the relation of the mean μ_t and the standard deviation σ_t for different values of ψ. The relation is not linear, however, for a smaller value of ψ, there is more overdispersion independent of the value of μ_t.

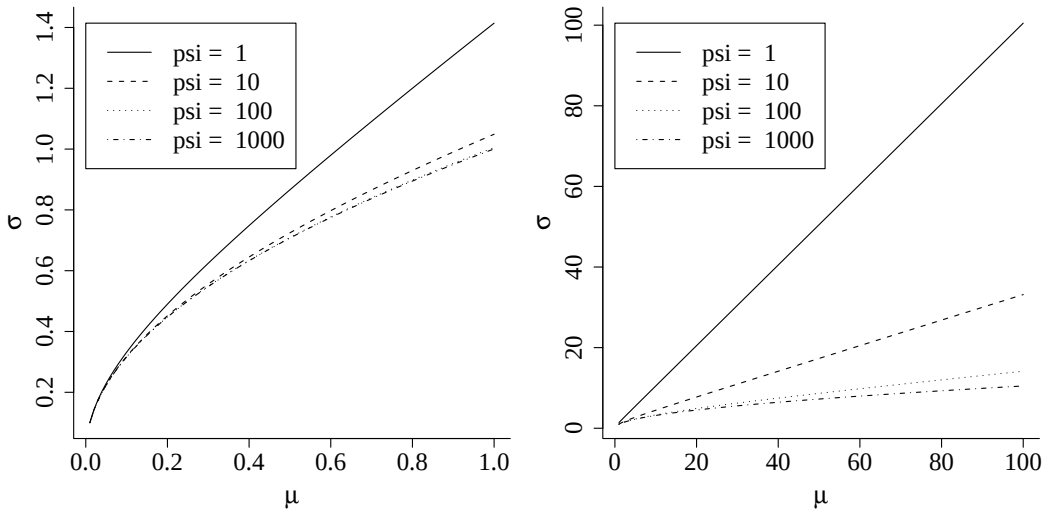
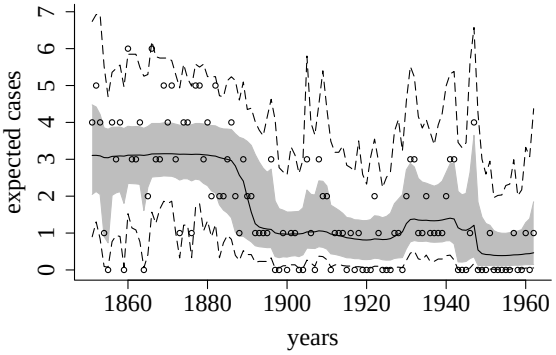


Figure 3.2: Relation of μ_t and σ_t for different values of ψ.

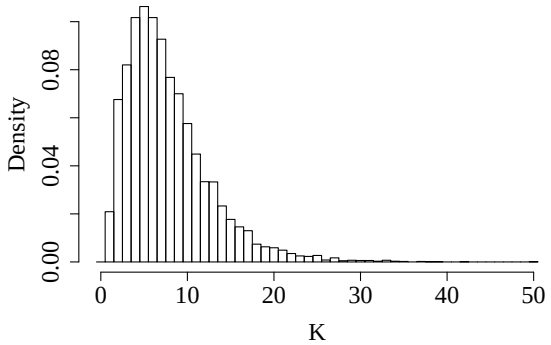
3.2 Application to coal-mining disaster data

Figure 3.3 shows an application of the changepoint model to the yearly number of observed coal-mining disasters in Britain from 1851 to 1962, that have been previously studied by various authors (e.g. Green, 1995; Denison et al., 2002), assuming X_t to be simply an intercept. For λ_k we used a $\text{Ga}(1.705, 1)$ prior with mean and variance 1.705, which is equal to the data mean. Figure 3.4 shows the results for the extended model, where a

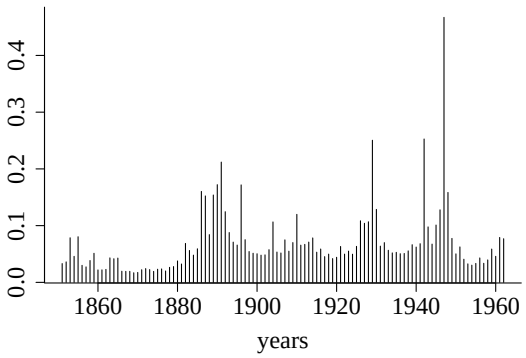
Ga(1, 1) prior has been used for β and a Ga(1, 0.1) prior for ψ. The mean and the variance of the prior for λ_k do not exist for this choice. As we will see later it is possible to estimate one step ahead predictions of the changepoint model. The Figures therefore include the predicted probability of a changepoint at time $t = 1962$ (which means that there is a change between the years 1962 and 1963) in Figure 3.3(c) and 3.4(c) and the predicted number of disasters in the year $t = 1963$ (Figure 3.3(d) and 3.4(d)). The estimation of $\boldsymbol{\lambda}$ in the extended model is a bit smoother, and the number of changepoints is slightly lower. The predictive distribution of the number of disasters is a little more dispersed showing a higher probability for very small and very high values.



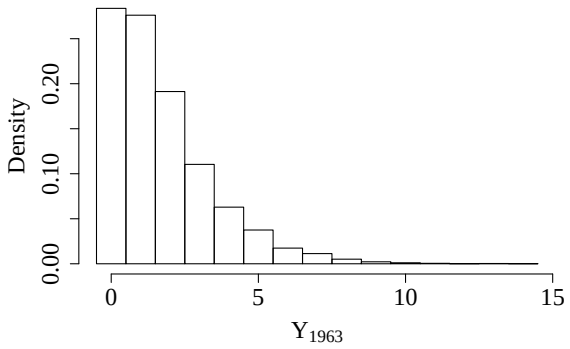
(a) Expected cases $\boldsymbol{\lambda}$ vs. data (points). Posterior mean (solid line) Pointwise 95% credibility interval (grey area) and simultaneous (Held, 2004) 95% credibility interval (dashed lines)



(b) Number of changepoints K

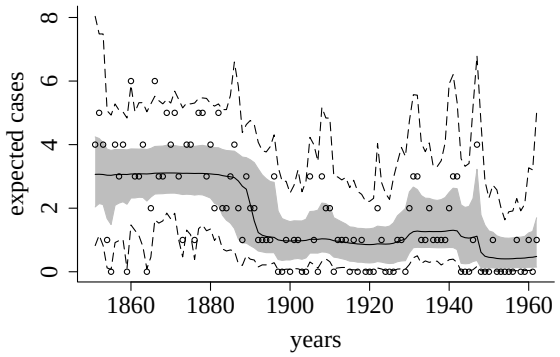


(c) Post. prob. for a changepoint at location t including the predicted θ_{1962}.

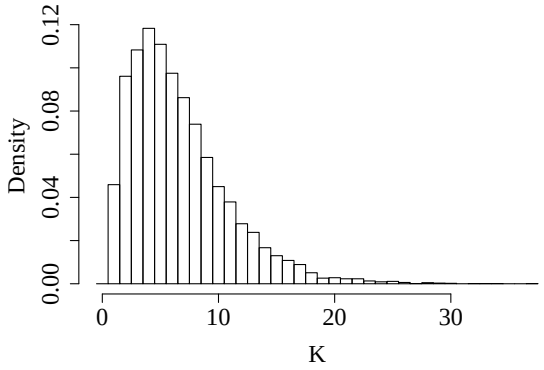


(d) Prediction of the number of disasters for 1963

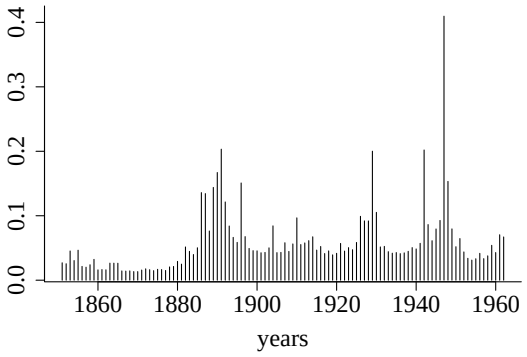
Figure 3.3: The changepoint model applied to the coal-mining disaster data.



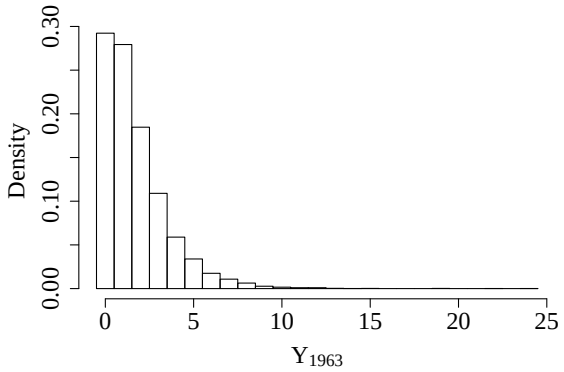
(a) Expected cases $\boldsymbol{\lambda}$ vs. data (points). Posterior mean (solid line) Pointwise 95% credibility interval (grey area) and simultaneous (Held, 2004) 95% credibility interval (dashed lines)



(b) Number of changepoints K



(c) Post. prob. for a changepoint at location t including the predicted θ_{1962}.



(d) Prediction of the number of disasters for 1963

Figure 3.4: The changepoint model with extensions applied to the coal-mining disaster data.

3.3 Posterior distribution

The posterior distribution of the model has the following form,

$$P(\boldsymbol{\lambda}, \boldsymbol{\theta}, K|\boldsymbol{Y}) \;=\; \frac{P(\boldsymbol{Y}|\boldsymbol{\lambda})P(\boldsymbol{\lambda}|\boldsymbol{\theta})P(\boldsymbol{\theta}|K)P(K)}{P(\boldsymbol{Y})}.$$

If a conjugate prior is used for the segment rates $\boldsymbol{\lambda}$, it is possible to get a analytic expression for the distribution of the data conditioned on the changepoints $P(\boldsymbol{Y}|\boldsymbol{\theta}, K)$ by marginalizing the joint distribution of the data and the segment rates conditioned on the changepoints $P(\boldsymbol{Y}, \boldsymbol{\lambda}|\boldsymbol{\theta}, K) = P(\boldsymbol{Y}|\boldsymbol{\lambda})P(\boldsymbol{\lambda}|\boldsymbol{\theta}, K)$ with respect to the segment rates $\boldsymbol{\lambda}$. This simplifies

the update of the changepoints, as we will see later. We get the following expression,

$$P(\boldsymbol{Y}|\boldsymbol{\theta},K) = \int P(\boldsymbol{Y},\boldsymbol{\lambda}|\boldsymbol{\theta},K)d\boldsymbol{\lambda} = \prod_{k=1}^{K+1} \frac{\prod_{t=\theta^{(k-1)}+1}^{\theta^{(k)}}(X_t)^{Y_t}\beta_\lambda^{\alpha_\lambda}\Gamma(\alpha_{\lambda,k-1,k})}{\prod_{t=\theta^{(k-1)}+1}^{\theta^{(k)}} Y_t!\Gamma(\alpha_\lambda)(\beta_{\lambda,k-1,k})^{\alpha_{\lambda,k-1,k}}}, \quad (3.4)$$

where

$$\begin{aligned}\alpha_{\lambda,k-1,k} &= \alpha_\lambda + \sum_{t=\theta^{(k-1)}+1}^{\theta^{(k)}} Y_t,\\ \beta_{\lambda,k-1,k} &= \beta_\lambda + \sum_{t=\theta^{(k-1)}+1}^{\theta^{(k)}} X_t.\end{aligned}$$

The distribution $P(\boldsymbol{Y}|\boldsymbol{\theta},K)$ is factorized with respect to the segments, i.e. the segments are independent conditioned on the changepoints even after averaging over $\boldsymbol{\lambda}$.

3.4 Markov state space form of the changepoint model

The considered changepoint model is a Markov state space model. This is of particular advantage for the calculation of the predictive distribution and allows for a sequential update of the model.

Let K_t be the number of changepoints within time $1:t$, i.e. $K = K_{n-1}$, and $\theta_t = 1$ if there is a changepoint at time t and $\theta_t = 0$ otherwise. For convenience we define $\theta_0 = 1$ and $\theta_n = 1$. The distribution $P(\boldsymbol{\lambda}|\boldsymbol{\theta})$ can then be written as

$$\lambda_t|(\theta_{t-1}=0) = \lambda_{t-1} \quad (3.5)$$

$$\lambda_t|(\theta_{t-1}=1) \sim \mathrm{Ga}(\alpha_\lambda,\beta_\lambda). \quad (3.6)$$

We set $K_0 = 0$. This representation is in analogy to Yao (1984). The following probabilities are already known from (3.1) and (3.3),

$$P(K_t = l) = \frac{1}{t+1}. \quad (3.7)$$

and

$$P(\theta_t = 1|K_t = l) = \frac{l}{t}. \quad (3.8)$$

The probability in (3.7) can only be derived from (3.1) in this way if the marginal prior distribution of K_t in the model for time $1:n$, is the same, as the prior for K in the model for time $1:(t+1)$. The probability $P(K_t = l)$ does then not depend on n. This is the case for a uniform distribution. Using these probabilities we can calculate

$$\begin{aligned}
P(\theta_t = 1|K_{t-1} = k) &= \sum_{j=0}^{t} P(\theta_t = 1, K_t = j|K_{t-1} = k) \\
&= P(\theta_t = 1, K_t = k+1|K_{t-1} = k) \qquad (3.9) \\
&= \frac{P(\theta_t = 1, K_t = k+1, K_{t-1} = k)}{P(K_{t-1} = k)} \\
&= \frac{P(\theta_t = 1, K_t = k+1)}{P(K_{t-1} = k)} \\
&= \frac{P(\theta_t = 1|K_t = k+1)P(K_t = k+1)}{P(K_{t-1} = k)}, \qquad (3.10) \\
&= \frac{k+1}{t+1}. \qquad (3.11)
\end{aligned}$$

The model has then the following Markov representation. Defining the parameter vector $x_t = (\lambda_t, \theta_{t-1}, K_{t-1})$ at time t we get

$$\begin{aligned}
P(x_t|x_{1:t}) &= P(\lambda_t, \theta_{t-1}, K_{t-1}|\lambda_{1:(t-1)}, \theta_{1:(t-2)}, K_{1:(t-2)}) \\
&= P(\lambda_t, \theta_{t-1}, K_{t-1}|\lambda_{t-1}, \theta_{t-2}, K_{t-2}) \\
&= P(x_t|x_{t-1}).
\end{aligned}$$

This can be further decomposed,

$$\begin{aligned}
&= P(\lambda_t|\theta_{t-1}, \lambda_{t-1})P(\theta_{t-1}, K_{t-1}|K_{t-2}) \\
&= P(\lambda_t|\theta_{t-1}, \lambda_{t-1})P(\theta_{t-1}|K_{t-2}).
\end{aligned}$$

3.5 Markov structure of the changepoints

The changepoints build a priori a Markov chain, $P(\theta^{(k+1)} = s|\theta^{(k)} = t, \theta^{(1:k-1)}) = P(\theta^{(k+1)} = s|\theta^{(k)} = t)$, with finite state space $\{0, \ldots, n-1\}$. Using (3.11) we get

$$P(\theta^{(k+1)} = s|\theta^{(k)} = t) = \frac{k+1}{s+1}, \quad s = t+1, \tag{3.12}$$

$$\begin{aligned} P(\theta^{(k+1)} = s|\theta^{(k)} = t) &= \prod_{i=t+1}^{s-1} \left(1 - \frac{k+1}{i+1}\right) \frac{k+1}{s+1} \\ &= \prod_{i=t+1}^{s-1} \left(\frac{i-k}{i+1}\right) \frac{k+1}{s+1}, \quad s > t+1. \end{aligned} \tag{3.13}$$

As mentioned before, the segments $S^{(k)}$, $k = 1, \ldots, K+1$ are independent given the changepoints, and for their distribution conditioned on the changepoints applies

$$P(S^{(k)}|\boldsymbol{\theta}) = P(S^{(k)}|\theta^{(k-1)}, \theta^{(k)}).$$

The model of the data and the changepoints marginalized by the segment means can therefore be represented by the graphical model shown in Figure 3.5.

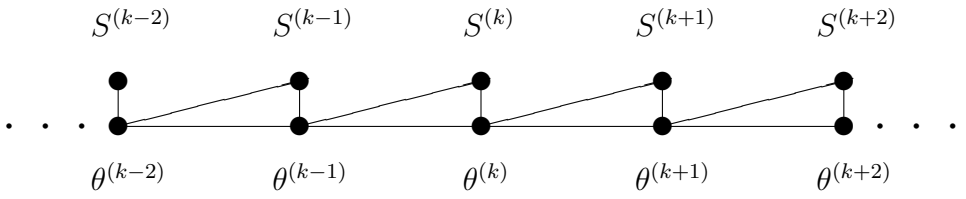


Figure 3.5: Graphical model of the changepoint model marginalized by the segment means $\boldsymbol{\lambda}$. The circles represent the nodes and the lines between the nodes represent undirected links.

3.6 Estimation of the changepoint model

We consider three ways to estimate the changepoint model: the reversible jump algorithm (2.1.4), the particle filter (2.2.1) and the forward-backward algorithm (2.2.2).

3.6.1 Reversible jump MCMC

We make use of the decomposed proposal by first proposing a new dimension, by proposing a new number of changepoints K^* from the proposal $q(K^*|K)$, and then a new parameter vector $(\boldsymbol{\lambda}^*, \boldsymbol{\theta}^*)$ from the proposal $q(\boldsymbol{\lambda}^*|\boldsymbol{\theta}^*)q(\boldsymbol{\theta}^*|\boldsymbol{\theta}, K^*)$, which is decomposed to a proposal for the location of the changepoints $\boldsymbol{\theta}^*$, that depends on the current location and the proposed number of changepoints $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}, K^*)$ and a proposal for the segment rates $\boldsymbol{\lambda}^*$, depending on the proposed locations of the changepoints, $q(\boldsymbol{\lambda}^*|\boldsymbol{\theta}^*)$. For the number of changepoints we use a random walk proposal, where a *split*, $K^* = K + 1$, is proposed with probability 0.5 or otherwise a *merge*, $K^* = K - 1$, is proposed. If $K = 0$ a split is always proposed and if $K = n - 1$ a merge is always proposed. The proposal of the locations of the changepoints has then the following form, depending on the proposed candidate for the number of changepoints:

split: Add a uniformly chosen new changepoint.

merge: Remove a uniformly chosen existing changepoint.

Since the full conditional of the segment rate $P(\lambda^{(k)}|\boldsymbol{\theta}, K, \boldsymbol{Y})$ is a Gamma distribution, i.e. know, we use this as proposal distribution. This does not depend on the current state of the segment rates and, as we will see, this is enough to avoid the additional difficulties that may arise by letting the proposal depend on the current states.

The acceptance probability $\alpha((\boldsymbol{\lambda}, \boldsymbol{\theta}, K), (\boldsymbol{\lambda}^*, \boldsymbol{\theta}^*, K^*))$ is then

$$\min\left(1, \frac{P(\boldsymbol{Y}|\boldsymbol{\lambda}^*)P(\boldsymbol{\lambda}^*|\boldsymbol{\theta}^*)P(\boldsymbol{\theta}^*|K^*)P(K^*)q(\boldsymbol{\lambda}^*|\boldsymbol{\theta}^*)q(\boldsymbol{\theta}^*|\boldsymbol{\theta}, K^*)q(K^*|K)}{P(\boldsymbol{Y}|\boldsymbol{\lambda})P(\boldsymbol{\lambda}|\boldsymbol{\theta})P(\boldsymbol{\theta}|K)P(K)q(\boldsymbol{\lambda}|\boldsymbol{\theta})q(\boldsymbol{\theta}|\boldsymbol{\theta}^*, K)q(K|K^*)}\right).$$

The full conditional of $\boldsymbol{\lambda}$ can be expressed as

$$P(\boldsymbol{\lambda}|\boldsymbol{\theta}, K, \boldsymbol{Y}) \;=\; \frac{P(\boldsymbol{Y}|\boldsymbol{\lambda})P(\boldsymbol{\lambda}|\boldsymbol{\theta})}{P(\boldsymbol{Y}|\boldsymbol{\theta}, K)}.$$

Substituting this into the acceptance probability, as proposed in (Holmes and Mallick, 2000), and cancelling leaves

$$\alpha((\boldsymbol{\lambda},\boldsymbol{\theta},K),(\boldsymbol{\lambda}^*,\boldsymbol{\theta}^*,K^*)) \;=\; \min\left(1,\frac{P(\boldsymbol{Y}|\boldsymbol{\theta}^*,K^*)P(\boldsymbol{\theta}^*|K^*)P(K^*)q(\boldsymbol{\theta}^*|\boldsymbol{\theta},K^*)q(K^*|K)}{P(\boldsymbol{Y}|\boldsymbol{\theta},K)P(\boldsymbol{\theta}|K)P(K)q(\boldsymbol{\theta}|\boldsymbol{\theta}^*,K)q(K|K^*)}\right).$$

The acceptance probability is now independent of the rates $\boldsymbol{\lambda}$. The probability $P(\boldsymbol{Y}|\boldsymbol{\theta},K)$ is given in (3.4). The prior ratio of the number of changepoints, and the prior ratio and the proposal ratio of the location of the changepoints reduce to 1. The proposal ratio of the number of changepoints, that we call c, is

$$c \;=\; \begin{cases} 0.5 & \text{for} \quad K=0 \quad \text{and} \quad K=n-1 \\ 1 & \text{for} \quad 2\le k\le n-3 \\ 2 & \text{for} \quad (K=1,K^*=0) \quad \text{and} \quad (K=n-2,K^*=n-1) \end{cases}$$

The acceptance probability $\alpha((\boldsymbol{\lambda},\boldsymbol{\theta},K),(\boldsymbol{\lambda}^*,\boldsymbol{\theta}^*,K^*))$ for a new changepoint m is then

$$\min\left(1,\frac{c\beta^{\alpha}\Gamma(\alpha_{\lambda,m-1,m})\Gamma(\alpha_{\lambda,m,m+1})(\beta_{m-1,m+1})^{\alpha_{\lambda,m-1,m+1}}}{\Gamma(\alpha)\Gamma(\alpha_{\lambda,m-1,m+1})(\beta_{\lambda,m-1,m})^{\alpha_{\lambda,m-1,m}}(\beta_{\lambda,m,m+1})^{\alpha_{\lambda,m,m+1}}}\right). \tag{3.14}$$

If a changepoint m is removed, the second term in the minimum function has to be replaced by its inverse.

Since the acceptance probability does not depend on the segment means $\lambda^{(k)}$ $k=1,\ldots,K+1$, the update of the changepoint model can be decomposed to two steps. First the changepoints are updated using a reversible jump algorithm with the acceptance probability in (3.14). Then the parameters $\lambda^{(k)}$ $k=1,\ldots,K+1$ are updated in a Gibbs step. The full conditional of $\lambda^{(k)}$, $k=1,\ldots,K+1$ is

$$\lambda^{(k)}|\ldots \;\sim\; \mathrm{Ga}\left(\alpha_{\lambda,k-1,k},\beta_{\lambda,k-1,k}\right).$$

Predictive distribution

Due to the Markov structure of the changepoint model it is possible to get the predictive distribution $P(\lambda_{t+1},\theta_t|Y_{1:t})$ by sampling in every iteration from the predictive distribution conditioned on the current sample of the parameters $P(\lambda_{t+1},\theta_t|Y_{1:t},\lambda_t,K_{t-1}) =$

$P(\lambda_{t+1}|\theta_t, \lambda_t)P(\theta_t|K_{t-1})$, which can be obtained from (3.11), (3.5) and (3.6),

$$\begin{aligned} P(\theta_t = 1|K_{t-1} = k) &= \frac{k+1}{t+1}, & (3.15)\\ \lambda_{t+1}|\theta_t = 0, \lambda_t &= \lambda_t, & (3.16)\\ \lambda_{t+1}|\theta_t = 1, \lambda_t &\sim \text{Ga}(\alpha_\lambda, \beta_\lambda). & (3.17) \end{aligned}$$

This corresponds to averaging $P(\lambda_{t+1}, \theta_t|Y_{1:t}, \lambda_t, K_{t-1})$ over λ_t and K_{t-1}.

Model extensions

The two model extension considered for the changepoint model at the beginning of this chapter can be easily handled within an MCMC algorithm. The hyperparameter β can be drawn from its full conditional

$$\beta|\ldots \sim \text{Ga}\left(a + K + 1, b + \sum_{k=1}^{K+1} \lambda^{(k)}\right), \quad (3.18)$$

as well as the random effects of the overdispersion

$$\omega_t|\ldots \sim \text{Ga}(\psi + Y_t, \psi + \lambda_t X_t). \quad (3.19)$$

For ψ a MH step has to be designed, since the full conditional is no known distribution. Since $\psi > 0$ we prefer to update $\tilde{\psi} = \log(\psi)$ with a simple Metropolis-Hastings Gaussian random walk proposal. The full conditional of ψ is

$$p(\psi|\ldots) \propto p(\psi)\prod_{t=1}^{n} p(\omega_t|\psi)$$

and the corresponding full conditional of $\tilde{\psi}$ can be obtained through a change of variables. The variance of the random walk proposal is tuned automatically within the algorithm in order to obtain a suitable acceptance rate between 30 and 50% (Gelman et al., 1996). The

changepoints can be updated in the same way as in the simple cases, but with

$$\begin{aligned}\alpha_{\lambda,k-1,k} &= \alpha_\lambda + \sum_{t=\theta^{(k-1)}+1}^{\theta^{(k)}} Y_t, \\ \beta_{\lambda,k-1,k} &= \beta_\lambda + \sum_{t=\theta^{(k-1)}+1}^{\theta^{(k)}} \omega_t X_t.\end{aligned}$$

The reversible jump algorithm is very fast compared to the update of the other parameters. However, due to the simple design of the steps, the algorithm mixes very slow, which leads to very long runs of the MCMC algorithm. We will therefore update the changepoints and steps 10 times per iteration. This clearly improves the mixing and leads to shorter runs and run times.

3.6.2 Particle filter

Since the considered changepoint model is a Markov state space model, the particle filter described in Section 2.2.1 can be used to estimate the model sequentially in time. To establish the algorithm for the considered changepoint model we have to define the importance distribution and the Markov transition step. Taking the predictive distribution (3.15)-(3.17) as importance distribution we get the following importance weights for $t > 1$, using the Markov state space form (3.4) of the model,

$$\begin{aligned}\tilde{w}(x_{1:t}^{(i)}) &= \frac{P(Y_t|x_t^{(i)})P(x_t^{(i)}|x_{t-1}^{(i)})}{q(x_t^{(i)}|x_{1:t-1}^{(i)}, Y_{1:t})} \\ &= \frac{P(Y_t|\lambda_t^{(i)})P(\lambda_t^{(i)}|\theta_{t-1}^{(i)}, \lambda_{t-1}^{(i)})P(\theta_{t-1}^{(i)}|K_{t-2}^{(i)})}{P(\lambda_t^{(i)}|\theta_{t-1}^{(i)}, \lambda_{t-1}^{(i)})P(\theta_{t-1}^{(i)}|K_{t-2}^{(i)})} \\ &= P(Y_t|\lambda_t^{(i)}).\end{aligned}$$

For $t = 1$ the prior distribution is used as proposal distribution. This leads to the same importance weights as in the case $t > 1$,

$$\begin{aligned}\tilde{w}(x_1^{(i)}) &= \frac{P(Y_1|x_1^{(i)})P(x_1^{(i)})}{q(x_1^{(i)}|Y_1)} \\ &= \frac{P(Y_1|\lambda_1^{(i)})P(\lambda_1^{(i)})}{P(\lambda_1^{(i)})} \\ &= P(Y_1|\lambda_1^{(i)}).\end{aligned}$$

As Markov transition step a reversible jump step is used to generate a new value for every particle.

Model extensions

By the inclusion of the hyperparameter and the adjustment for overdispersion, time constant parameters, namely β and ψ, are introduced to the model. However, the update of time constant parameters causes difficulties in the update using particle filters. The algorithm proposed in Andrieu et al. (2001) can be used, where the time constant parameters are replaced by time varying parameters. However, the number of samples needed increases over time, eventually exponentially (Crisan and Doucet, 2002).

To avoid these problems on can alternatively fix the time constant parameters. We will use the following proceeding. First the MCMC algorithm is run until time n_{MCMC}. A point estimate of the time constant parameters can thereby be obtained, e.g. the posterior mean. From time $n_{MCMC} + 1$ until time n the model is estimated sequentially using the described particle filter, where the time constant parameters are fixed, taking the values obtained from the MCMC run.

The algorithm changes as follows. The Poisson distribution for the data $P(Y_t|\lambda_t^{(i)})$, used in the importance step, is replaced by a Negative Binomial distribution. This avoids the update of the random effects ω_t in this step. In the Markov transition step, however, these random effects are updated, since they allow to use the reversible jump step for the changepoints.

3.6.3 Forward-backward algorithm

Due to their Markov property (3.5), the forward-backward algorithm, described in Section 2.2.2, can be used to estimate the posterior distribution of the changepoints. To update

the complete model, the update of the changepoints in the reversible jump MCMC algorithm can be replaced by sampling one time from the posterior of the changepoints using the forward-backward algorithm. Since the forward-backward algorithm allows to sample directly from the posterior, the resulting algorithm is a simple Monte Carlo algorithm, giving independent samples from the posterior of the complete model. The algorithm is computational inexpensive, since the forward step has to be calculate just once.

Forward step

For the update of the changepoints we will start the algorithm at the end $t = n$ instead of the beginning $t = 1$. We will calculate the distributions $P(Y_{t:n}|\theta^{(k)} = t-1)$, for $k = 1, \ldots, n-1$ and $t = k+1, \ldots, n$. The following probabilities are already known: the probability of the segments $S^{(k)} = (Y_{\theta^{(k-1)}+1}, \ldots, Y_{\theta^{(k)}})$, $k = 1, \ldots, K+1$ given the changepoints form (3.4),

$$\begin{aligned} P(S^{(k)}|\theta^{(k-1)}, \theta^{(k)}) &= \frac{\prod_{t=\theta^{(k-1)}+1}^{\theta^{(k)}} (X_t)^{Y_t} \beta_\lambda^{\alpha_\lambda} \Gamma(\alpha_{\lambda,k-1,k})}{\prod_{t=\theta^{(k-1)}+1}^{\theta^{(k)}} Y_t! \Gamma(\alpha_\lambda) (\beta_{\lambda,k-1,k})^{\alpha_{\lambda,k-1,k}}}, \\ \alpha_{\lambda,k-1,k} &= \alpha_\lambda + \sum_{t=\theta^{(k-1)}+1}^{\theta^{(k)}} Y_t, \\ \beta_{\lambda,k-1,k} &= \beta_\lambda + \sum_{t=\theta^{(k-1)}+1}^{\theta^{(k)}} X_t, \end{aligned}$$

and the prior probabilities of the changepoints from (3.20) and (3.13),

$$\begin{aligned} P(\theta^{(k+1)} = s|\theta^{(k)} = t) &= \frac{k+1}{s+1} \quad s = t+1, \\ P(\theta^{(k+1)} = s|\theta^{(k)} = t) &= \prod_{i=1}^{s-t-1} \left(\frac{t+i-k}{t+1+i}\right) \frac{k+1}{s+1} \quad s > t+1. \end{aligned}$$

The probabilities of no further changepoint is given by

$$P(\theta^{(k+1)} = n|\theta^{(k)} = t) = \prod_{i=1}^{n-t-1} \left(\frac{t+i-k}{t+1+i}\right).$$

The desired probabilities can be calculate sequentially by first calculating the probabilities

$$\begin{aligned} P(Y_{t:n}, \theta^{(k+1)} = s|\theta^{(k)} = t-1) &= P(Y_{t:n}|\theta^{(k+1)} = s, \theta^{(k)} = t-1)P(\theta^{(k+1)} = s|\theta^{(k)} = t-1) \\ &= P(S^{(k)}|\theta^{(k+1)} = s, \theta^{(k)} = t-1)P(Y_{s+1:n}|\theta^{(k+1)} = s) \\ &\quad \cdot P(\theta^{(k+1)} = s|\theta^{(k)} = t-1), \end{aligned}$$

starting with the known probability $P(Y_{n:n}|\theta^{(n-1)} = n-1) = P(S^{(K+1)}|\theta^{(n-1)} = n-1, \theta^{(n)} = n)$, and then marginalizing by $\theta^{(k+1)}$,

$$\begin{aligned} P(Y_{t:n}|\theta^{(k)} = t-1) &= \sum_{s=t}^{n-1} P(S^{(k)}|\theta^{(k+1)} = s, \theta^{(k)} = t-1)P(Y_{s+1:n}|\theta^{(k+1)} = s) \\ &\quad \cdot P(\theta^{(k+1)} = s|\theta^{(k)} = t-1), \end{aligned}$$

for $k = n-1, \ldots, 1$ and $t = n, \ldots, k+1$. The marginal likelihood is given by $P(Y_{1:n}|\theta^{(0)} = 0)$.

Backward step

Using the following probabilities,

$$\begin{aligned} & P(\theta^{(k+1)} = s|\theta^{(k)} = t-1, Y_{1:n}) \\ = \; & P(\theta^{(k+1)} = s|\theta^{(k)} = t-1, Y_{t:n}) \\ = \; & \frac{P(\theta^{(k+1)} = s, Y_{t:n}|\theta^{(k)} = t-1)}{P(Y_{t:n}|\theta^{(k)} = t-1)} \\ = \; & \frac{P(Y_{t:n}|\theta^{(k+1)} = s, \theta^{(k)} = t-1)P(\theta^{(k+1)} = s|\theta^{(k)} = t-1)}{P(Y_{t:n}|\theta^{(k)} = t-1)} \\ = \; & \frac{P(Y_{t:s}|\theta^{(k+1)} = s, \theta^{(k)} = t-1)P(Y_{s+1:n}|\theta^{(k+1)} = s)P(\theta^{(k+1)} = s|\theta^{(k)} = t-1)}{P(Y_{1:n}|\theta^{(k)} = t-1)}, \end{aligned}$$

we can sample from the posterior distribution of the changepoints as follows:

- Sample the first changepoint from $P(\theta_1 = t|Y_{1:n})$ using the efficient algorithm of Carpenter et al. (1999).
- If the k-th changepoint is at time $t-1$, sample the next changepoint $k+1$ from $P(\theta^{(k+1)} = s|\theta^{(k)} = t-1, Y_{1:n})$.

The algorithm is analog to the algorithms presented in Fearnhead (2006). However, in contrast to the algorithm of Fearnhead (2006) for this prior type, we perform the simulation

of the changepoints without conditioning the recursions on the number of changepoints. This leads to a faster estimate, since the computational complexity of the forward step is of order $O(n^3)$, due to the dependence of the probabilities on $k = n-1, \ldots, 1$ and $t = n, \ldots, k+1$ and the summation over $s = t, \ldots, n-1$, instead of $O(n^4)$ for the additional dependence on the number of changepoints $K = 0, \ldots, n-1$.

To further reduce the computational complexity of our algorithm, we follow Fearnhead (2006) by truncating the sums from the marginalization in the forward step at term $l+1$ when

$$\frac{P(S^{(k)}|\theta^{(k-1)}, \theta^{(k)})P(Y_{l+1:n}|\theta^{(k+1)} = l)P(\theta^{(k+1)} = l|\theta^{(k)} = t-1)}{\sum_{s=t}^{l} P(S^{(k)}|\theta^{(k-1)}, \theta^{(k)})P(Y_{s+1:n}|\theta^{(k+1)} = s)P(\theta^{(k+1)} = s|\theta^{(k)} = t-1)}$$

is less then 10^{-20}. If the number of changepoints is proportional to the observation time n, this reduces the computational cost to $O(n^2)$, since the length of the sums do not increase with n. Figure 3.6 shows 10 independent samples of the changepoints sampled by the forward-backward algorithm.

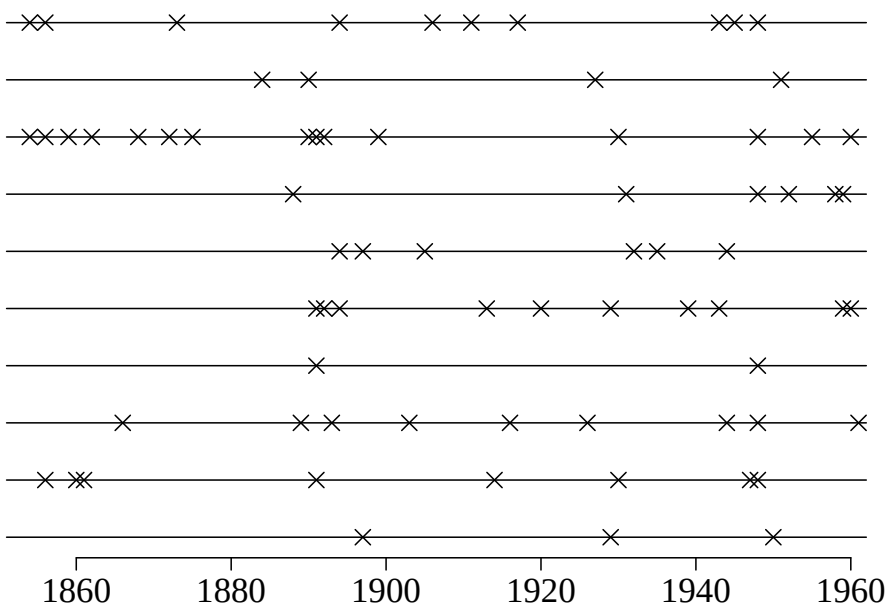


Figure 3.6: Ten independent samples of the changepoints.

Model extensions

In case of the considered model extensions the forward-backward algorithm is not applicable directly, since the changepoint problem depends on the hyperparameter β and the random effects ω_t, $t = 1, \ldots, n$. However, the algorithm can be included in a MCMC algorithm. The changepoints are then sampled from their full conditional, which is conditioned on the hyperparameter β and the random effects ω_t, $t = 1, \ldots, n$. This means that the changepoint model is updated in a Gibbs step. Since the forward step has to be run every iteration, the update of the changepoints by forward-backward algorithm is much more time consuming than by the reversible jump algorithm. However, the use of the forward-backward algorithm leads to a better mixing of the MCMC algorithm and therefore to shorter runs of the Markov chain.

Chapter 4

A model for multivariate infectious disease counts based on a branching process with immigration

4.1 Introduction

Modelling disease counts has been frequently tackled by the use of generalize linear mixed models (GLMM) (e.g. Zeger, 1988; Kleinman et al., 2004), as used for chronic diseases (e.g. Waller et al., 1997; Knorr-Held and Besag, 1998), which are based on latent parameters assuming Gaussian Markov random field (GMRF) priors. While these kind of models are well suited under the assumption of a purely endemic disease, the smooth effects, assuming a constant amount of smoothing, may not be able to capture the epidemic characteristics that are present in infectious disease data.

A promising approach to this problem is given in Held et al. (2005) and Toschke and Held (2006), who base their model on a branching process, which is an approximation of a mechanistic model, namely the chain binomial model (see e.g. Becker, 1989). A key idea of the model is the distinction between an endemic component, which models the regular pattern of the disease and an epidemic component, which explains for the outbreaks in the data. The epidemic component includes an autoregression directly on the disease counts and interactions between the units, e.g. regions or age groups, based on past disease counts of related units, e.g. neighbours in a spatial setting. This observation-driven approach should be better suited for modelling outbreaks in the data than the purely parameter-driven GLMM approaches. However, the model is relatively simple in the way of modelling

the endemic cases, consisting of independent spatial random effects, a linear trend and a seasonal term based on independent sine and cosine curves of different frequencies.

The aim here is to bring together the branching process based model proposed by Held et al. (2005) and GLMMs. While the epidemic part is based on the branching process as in Held et al. (2005) the endemic part is now based on GLMMs, including structured time effects. Since the dependence between units should be explained by the epidemic component we do not introduce dependence between the spatial random effects of the endemic component. We adjust for overdispersion by introducing random effects that act multiplicatively on the mean. The marginal distribution of the disease counts, integrating out the random effects, is then a negative binomial distribution.

We will compare seven models for the interactions between the units in the epidemic component, where four are similar to Held et al. (2005) and Toschke and Held (2006). We then examine the benefits of the inclusion of a random walk and an epidemic component, by comparing the different models including a random walk, an epidemic component and both, a random walk and an epidemic component.

In Section 2 and 3 we will present the univariate model formulation and the implementation using MCMC algorithms. In Section 4 the univariate version of the model is applied to five disease time series and compared to the two models without random walk and epidemic component. In Section 5 and 6 the model is extended to the multivariate case. In Section 7 the multivariate model is applied to data of the 2001/02 outbreak of measles, observed in the seven districts of Bavaria and the seven epidemic model formulations are compared to each other and with and without the inclusion of a random walk. The chapter concludes with a discussion and an outline of possible extensions.

4.2 The univariate case

Let Z_t, $t = 1, \ldots, n$ be the observed number of counts at time t. We assume the data to be Poisson distributed, conditioned on the auxiliary variables ω_t,

$$Z_t|\omega_t \quad \sim \quad \text{Po}(\omega_t \mu_t),$$

where the auxiliary variables are Gamma distributed with shape and rate parameter ψ, $\omega_t \sim \text{Ga}(\psi, \psi)$. As described in Chapter 3 this leads to a negative binomial distribution for Z_t integrating out ω_t, with mean μ_t and dispersion parameter ψ, $Z_t \sim \text{Nb}(\mu_t, \psi)$. A Gamma prior is assumed for the dispersion parameter, $\psi \sim \text{Ga}(\alpha_\psi, \beta_\psi)$. The mean μ_t is

assumed to be the sum of an endemic part ν_t, that explains the regular amount of cases and can be interpreted as a baseline and an epidemic part η_t, that explains for epidemic activity or irregularities in the data, and the interaction between units,

$$\mu_t = \eta_t + \nu_t.$$

4.2.1 The endemic component

The logarithm of endemic part of the mean ν_t, also referred to as endemic mean, is assumed to be the sum of a time trend β_t and a seasonal part ζ_t,

$$\log \nu_t = \beta_t + \zeta_t.$$

For the time trend $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)$ we assume a random walk prior of second order. This assumes that conditional on the past the parameter β_t has a normal distribution depending only on the two preceding parameters β_{t-1} and β_{t-2},

$$\beta_t \sim N(2\beta_{t-1} - \beta_{t-2}, \sigma_\beta^2), \quad t = 3, \ldots, n,$$

where the parameters of the first two time points are assumed to have a diffuse distribution, $\beta_1 \propto$ const and $\beta_2 \propto$ const. The density of the joint distribution of the parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)$ is then

$$\begin{aligned} p(\boldsymbol{\beta}|\sigma_\beta^2) &= (2\pi\sigma_\beta^2)^{(n-2)/2} \exp\left(-\frac{1}{2\sigma_\beta^2}\sum_{t=3}^{n}(\beta_t - 2\beta_{t-1} + \beta_{t-2})^2\right) \\ &= (2\pi\sigma_\beta^2)^{(n-2)/2} \exp\left(-\frac{1}{2\sigma_\beta^2}\boldsymbol{\beta}'\boldsymbol{R}\boldsymbol{\beta}\right) \end{aligned}$$

with penalty matrix

$$R = \begin{pmatrix} 1 & -2 & 1 & & & & & \\ -2 & 5 & -4 & 1 & & & & \\ 1 & -4 & 6 & -4 & 1 & & & \\ & 1 & -4 & 6 & -4 & 1 & & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & & 1 & -4 & 6 & -4 & 1 \\ & & & & 1 & -4 & 5 & -2 \\ & & & & & 1 & -2 & 1 \end{pmatrix}.$$

The variance of the random walk is assumed to be inverse Gamma distributed,

$$\sigma_\beta^2 \sim IG(\alpha_\sigma, \beta_\sigma).$$

We follow Natário and Knorr-Held (2003) by using $\alpha_\sigma = 1$ and $\beta_\sigma = 0.00005$.

The seasonal part ζ_t is assumed to be the sum of L harmonic waves, an intercept and, in case of no random walk, a centered linear time trend. This can be represented in the following form (e.g. Diggle, 1990),

$$\zeta_t = \sum_{j=0}^{J} \gamma_j s_{tj}, \tag{4.1}$$

with $J = 2L + 1$ and

$$\begin{aligned} s_{t0} &= 1, \\ s_{t1} &= t - (n+1)/2, \\ s_{tj} &= \begin{cases} \sin(\frac{\rho t(j+1)}{2}) & \text{for } j = 2, 4, \ldots, 2L, \\ \cos(\frac{\rho tj}{2}) & \text{for } j = 3, 5, \ldots, 2L+1, \end{cases} \end{aligned}$$

where ρ is the base frequency, e.g. $\rho = \frac{52}{2\pi}$ for weekly observed data. Independent normal priors are assumed for the parameters γ_j, $j = 0, \ldots, J$,

$$\boldsymbol{\gamma} = (\gamma_0, \ldots, \gamma_J) \sim \mathrm{N}(0, \sigma_\gamma^2 \boldsymbol{I}), \quad \sigma_\gamma^2 = 10^6.$$

4.2.2 The epidemic component

The epidemic component is modeled by an autoregression on the number of cases Z_{t-1} at the last time point,

$$\eta_t = \lambda Z_{t-1}.$$

For the rate parameters λ we assume a Gamma prior with a Gamma hyperprior on the second parameter,

$$\begin{aligned} \lambda &\sim \text{Ga}(\alpha_\lambda, \beta_\lambda), \\ \beta_\lambda &\sim \text{Ga}(a, b). \end{aligned}$$

The model is a branching process with immigration, which is ergodic in the case $\lambda < 1$ and increases exponentially for $\lambda > 1$. In the ergodic case $\lambda < 1$ the mean number of cases at time t is $\nu_t/(1-\lambda)$ (Guttorp, 1995), and the mean number of epidemic cases is $\nu_t/(1-\lambda) - \nu_t = \lambda\nu_t/(1-\lambda)$. The parameter λ can therefore be interpreted as the proportion of epidemic cases from the total number of cases.

4.3 Estimation by MCMC

The key to a successful application of MCMC methods to the specified models lies in a suitable decomposition of the model using auxiliary variables.

4.3.1 Alternative representation using auxiliary variables

We therefore introduce auxiliary variables for the endemic component $X_t \sim \text{Po}(\omega_t\nu_t)$ and epidemic component $Y_t \sim \text{Po}(\omega_t\eta_t)$, where $Z_t = X_t + Y_t$. These can be interpreted as the number of endemic cases X_t and the number of epidemic cases Y_t.

4.3.2 Update of the parameters

Conditioned on X_t the estimation of the endemic parameters β_t and γ_j is equivalent to the estimation in a Bayesian log-linear Poisson regression model with response variable X_t. This can be done in a Metropolis-Hastings (MH) step, where a normal approximation of the full conditional by a Taylor expansion of second order is used as proposal distribution,

see for example Rue and Held (2005, Section 4.4). While the parameters of the time trend $\boldsymbol{\beta}$ and σ_β can be updated jointly, the parameters γ_j are updated individually, since a joint update leads to very poor acceptance rates.

The full conditional of the parameter λ is, conditioned on the corresponding auxiliary variable Y_t, a Gamma distribution, which makes it possible to update it in a simple Gibbs step, as the hyperparameter β_λ,

$$\begin{aligned} \lambda|Y_t, \beta_\lambda, Z_t, \ldots &\sim \text{Ga}(\alpha_\lambda + \sum_{t=1}^{n} Y_t, \beta_\lambda + \sum_{t=1}^{n} \omega_t Z_t), \\ \beta_\lambda|\lambda, Z_t, \ldots &\sim \text{Ga}(\alpha_\lambda(a+1), b+\lambda). \end{aligned}$$

Since $\psi > 0$ we prefer to update $\tilde{\psi} = \log(\psi)$ with a simple Metropolis-Hastings Gaussian random walk proposal. The full conditional of ψ is

$$p(\psi|\ldots) \propto p(\psi) \prod_{t=1}^{n} p(\omega_t|\psi)$$

and the corresponding full conditional of $\tilde{\psi}$ can be obtained through a change of variables. The variance of the random walk proposal is tuned automatically within the algorithm in order to obtain a suitable acceptance rate between 30 and 50% (Gelman et al., 1996). The full conditionals of the parameters ω_{it} are Gamma distributions $\omega_{it}|\ldots \sim \text{Ga}(\psi+Z_t, \psi+\mu_t)$ allowing for an update in a Gibbs step.

The full conditional distribution of the auxiliary variables can be factorised as

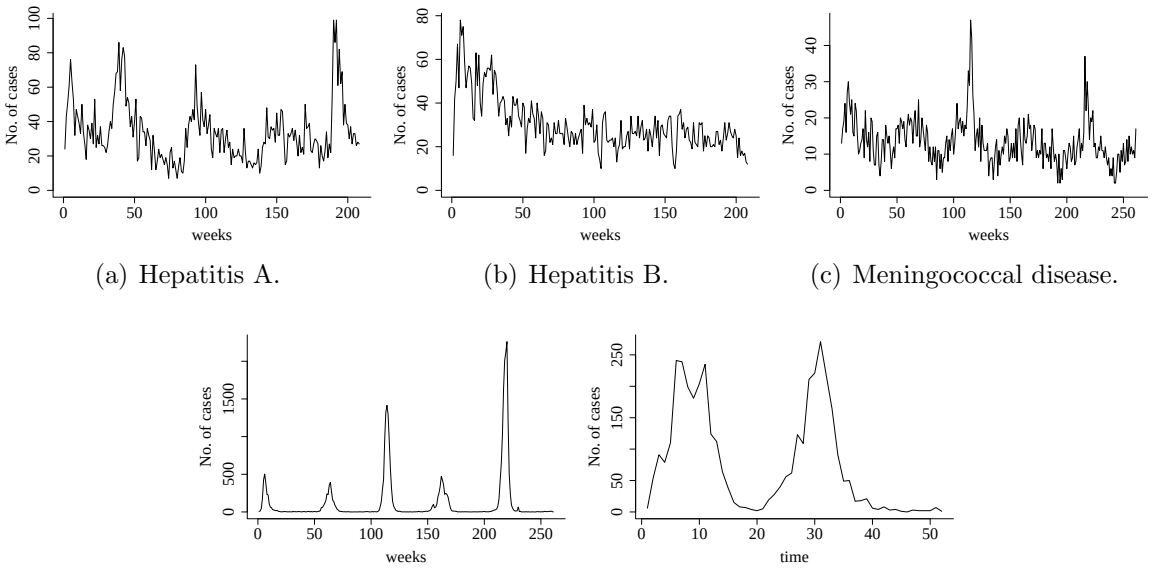
$$P(X_t, Y_t|Z_t, \ldots) = P(Y_t|X_t, Z_t, \ldots)P(X_t|Z_t, \ldots),$$

where the first term $P(Y_t|X_t, Z_t, \ldots)$ is deterministic: $Y_t = Z_t - X_t$. Due to the Poisson assumption the second term is binomial,

$$X_t|\ldots \sim \text{Bin}\left(Z_t, \frac{\nu_t}{\mu_t}\right).$$

4.4 Application to univariate disease time series

We will apply the three models, a model with epidemic component, a model with random walk and a model with both, epidemic component and random walk, to five disease time series shown in Figure 4.1, that are collected in the German infectious disease sur-



(a) Hepatitis A. (b) Hepatitis B. (c) Meningococcal disease.

(d) Influenza. (e) Measles.

Figure 4.1: Five disease counts time series.

veillance system, administrated by the Robert Koch Institute in Berlin (obtained from SurvStat@RKI, `http://www3.rki.de/SurvStat`): the weekly observed disease counts in Germany from 2001 until 2004 of hepatitis A and B, and from 2001 until 2005 of meningococcal disease and influenza, and the disease counts of the measles outbreak of 2001 and 2002 in Bavaria, that are aggregated to a basis of two weeks, which is closer to the generation time.

Hepatitis A and B do not show considerable person to person transmission, however, the hepatitis A data may show some outbreaks. Meningococcal disease shows both characteristics. While usually cases are not caused by person to person transmission due to health care measures, there is always a risk of a bigger outbreak caused by person to person transmission. Influenza and measles are almost entirely transmitted from person to person causing bigger outbreaks in the absence of health care measures. We will compare the three models, the model with random walk, with epidemic component and with both, random walk and epidemic component, for the five time series. The deviance summaries together with the median and 95% credibility interval of the estimates of the epidemic parameter λ and the dispersion parameter ψ for the five time series are shown in Table 4.1-4.5. For hepatitis A, hepatitis B and measles the model with random walk and the model with

	$\hat{\lambda}$(SD)	$\hat{\psi}$(SD)	$\bar{D}$	p_D	DIC
epid. comp.	0.59(0.08)	14.4(2.2)	1538	13	1550
random walk	–	22.5(4.3)	1473	22	1494
epid. comp. & random walk	0.12(0.1)	22.3(4.1)	1475	22	1497

Table 4.1: Hepatitis A

	$\hat{\lambda}$(SD)	$\hat{\psi}$(SD)	$\bar{D}$	p_D	DIC
epid. comp.	0.48(0.06)	36.6(7.6)	1396	6	1402
random walk	–	46.8(11.5)	1368	13	1381
epid. comp. & random walk	0.26(0.09)	50.3(11.8)	1360	13	1373

Table 4.2: Hepatitis B

	$\hat{\lambda}$(SD)	$\hat{\psi}$(SD)	$\bar{D}$	p_D	DIC
epid. comp.	0.17(0.07)	26.7(7)	1486	6	1493
random walk	–	27.9(7.4)	1483	14	1497
epid. comp. & random walk	0.11(0.08)	28.2(7.9)	1481	15	1496

Table 4.3: Meningococcal disease

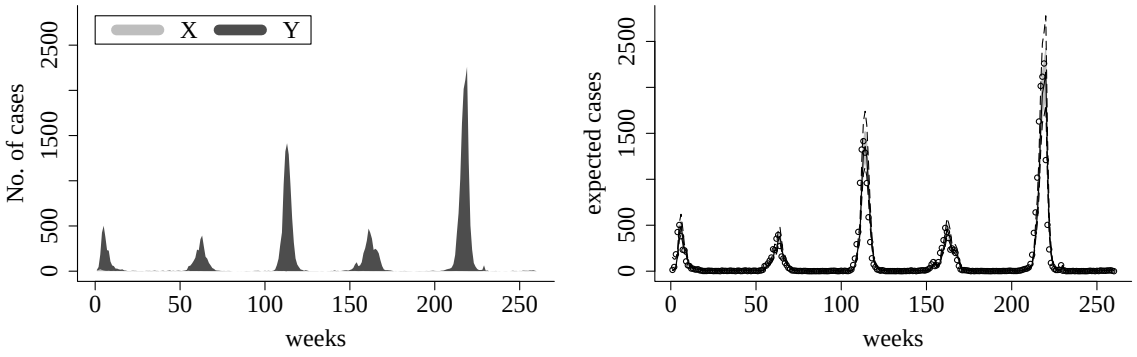
	$\hat{\lambda}$(SD)	$\hat{\psi}$(SD)	$\bar{D}$	p_D	DIC
epid. comp.	0.97(0.06)	2.4(0.3)	1838	4	1842
random walk	–	36.9(14.3)	1504	81	1585
epid. comp. & random walk	0.96(0.06)	2.5(0.3)	1833	9	1842

Table 4.4: Influenza

	$\hat{\lambda}$(SD)	$\hat{\psi}$(SD)	$\bar{D}$	p_D	DIC
epid. comp.	0.86(0.09)	4.1(1.1)	433	5	438
random walk	–	24(10.5)	364	18	382
epid. comp. & random walk	0.01(0.03)	24.5(10.5)	366	18	383

Table 4.5: Measles

both, random walk and epidemic component show almost the same DIC, while both are better in terms of DIC than the model with epidemic component. For meningococcal disease there is no considerable difference between the three models. For influenza the model with random walk is the best, while the model with epidemic component, and the model with both, random walk and epidemic component, show approximately the same DIC. It is surprising that although both components are included in the third model the model including just a random walk shows a much better fit. This is due to the fact that model 3 explains the outbreaks in the data by the epidemic component rather than the random walk (Figure 4.2), although the random fits the data better. An explanation may be that the autoregressive structure is better suited for modelling outbreaks of influenza. Due to the absence of an explosion of the number of cases, the value of λ can not take values beyond 1. The temporary fast increases of the cases, however, can just be explained by a bigger λ, clearly exceeding the value 1. To allow for a time varying λ_t would therefore be a promising extension. Figure 4.3 shows the results for meningococcal disease. The model with epidemic component shows a permanently increased number of epidemic cases, which does not coincide with our intuition of the epidemic component, since we expect outbreaks to be occasional events. This can also be observed for other time series. The



(a) Post. mean of X_t and Y_t with known Z_t, $t = 1, \ldots, n$

(b) Post. mean, pointwise (grey area) and simultaneous (dashed line) and 95% credibility interval of μ_t plotted against the data Z_t (points).

Figure 4.2: Results for influenza for the model with random walk and epidemic component.

model including both, random walk and epidemic component, in contrast shows just two little increases of the epidemic cases, where an outbreak is visible in the data. However,

many of the cases that we would expect to be part of the outbreak are explained by the random walk. It seems that the fact that the epidemic component is close to 0 most of the time, leads to a too low estimate of λ to explain the fast increase of cases at the beginning of the outbreaks. This reinforces the benefit of allowing for a time depending epidemic parameter. The expected cases of the model without epidemic component do not explain for the two outbreaks, that can be seen in the data, while the expected cases of the two model with epidemic component show at least a little increase during the outbreaks. This can also be seen in the predicted cases, where prediction is based on the whole time series $(Z_1, \ldots, Z_n)$.

In Figure 4.4 the deviance residuals based on a negative binomial assumption for the data are considered, i.e. the squared deviance residuals are defined as

$$\begin{aligned} d_t^2 &= \log(\Gamma(Z_t + \psi)) - \log(\Gamma(\psi)) - \log(Z_t!) - (Z_t + \psi)\log(\mu_t + \psi) \\ &\quad + \psi \log(\psi) + Z_t \log(\mu_t). \end{aligned}$$

The deviance residuals of all three models show little autocorrelation, where the model with random walk and epidemic component shows slightly less autocorrelated residuals.

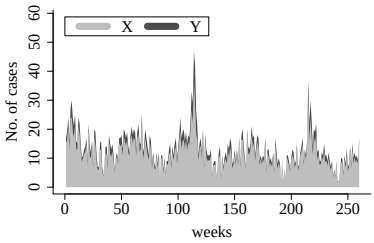
4.5 The multivariate case

Let Z_{it}, $i = 1, \ldots, I$, $t = 1, \ldots, n$ be the observed number of counts in unit i at time t. We assume the data to be Poisson distributed, conditioned on the auxiliary variables ω_{it},

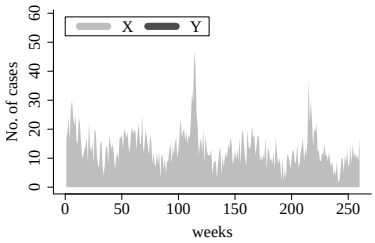
$$Z_{it}|\omega_{it} \sim \text{Po}(\omega_{it}\mu_{it}),$$

where the auxiliary variables are Gamma distributed with shape and rate parameter ψ, i.e. $\omega_{it} \sim \text{Ga}(\psi, \psi)$. As described in Chapter 3 this leads to a negative binomial distribution for Z_{it} integrating out ω_{it}, with mean μ_{it} and dispersion parameter ψ, $Z_{it} \sim \text{Nb}(\mu_{it}, \psi)$. A Gamma prior is assumed for the dispersion parameter, $\psi \sim \text{Ga}(\alpha_\psi, \beta_\psi)$. The mean μ_{it} is, as in the univariate case, assumed to be the sum of an endemic part ν_{it}, that explains for the regular amount of cases and an epidemic part η_{it}, that explains for epidemic activity or irregularities in the data,

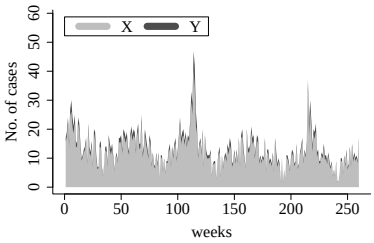
$$\mu_{it} = \eta_{it} + \nu_{it}.$$



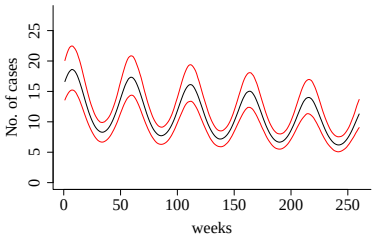
(a) Post. mean of X_t and Y_t with known Z_t, $t = 1, \ldots, n$



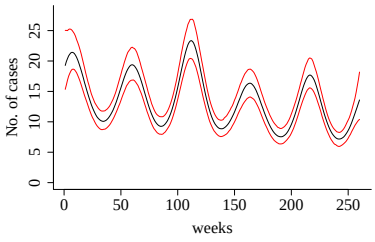
(b) Post. mean of X_t and Y_t with known Z_t, $t = 1, \ldots, n$



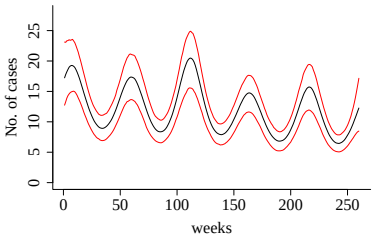
(c) Post. mean of X_t and Y_t with known Z_t, $t = 1, \ldots, n$



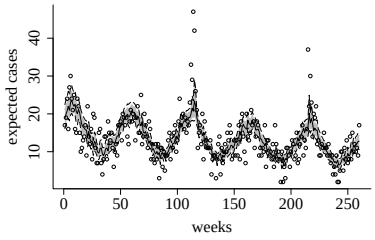
(d) Post. median and pointwise 95% credibility interval of ν_t, $t = 1, \ldots, n$.



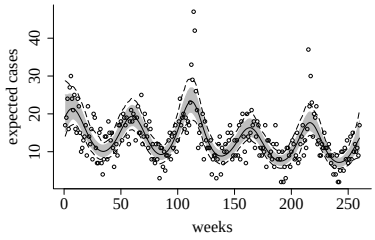
(e) Post. median and pointwise 95% credibility interval of ν_t, $t = 1, \ldots, n$.



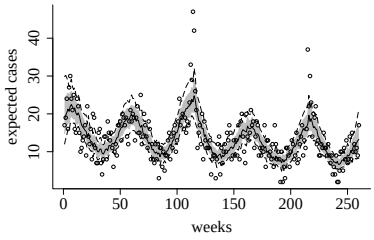
(f) Post. median and pointwise 95% credibility interval of ν_t, $t = 1, \ldots, n$.



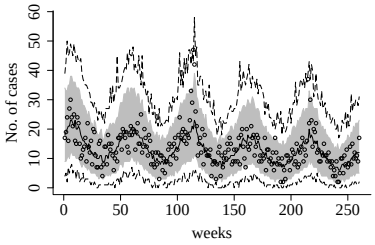
(g) Post. mean, ptw. (grey area) and simult. (dashed line) 95% CI of μ_t vs. Z_t (points).



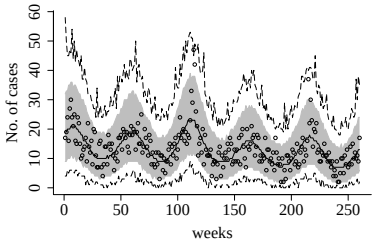
(h) Post. mean, ptw. (grey area) and simult. (dashed line) 95% CI of μ_t vs. Z_t (points).



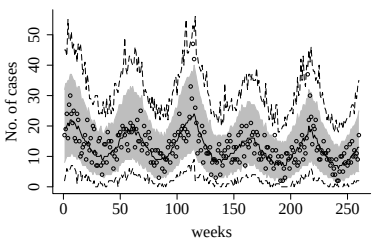
(i) Post. mean, ptw. (grey area) and simult. (dashed line) 95% CI of μ_t vs. Z_t (points).



(j) Post. mean, ptw. (grey area) and simult. (dashed line) 95% CI of the pred. cases vs. Z_t (points).

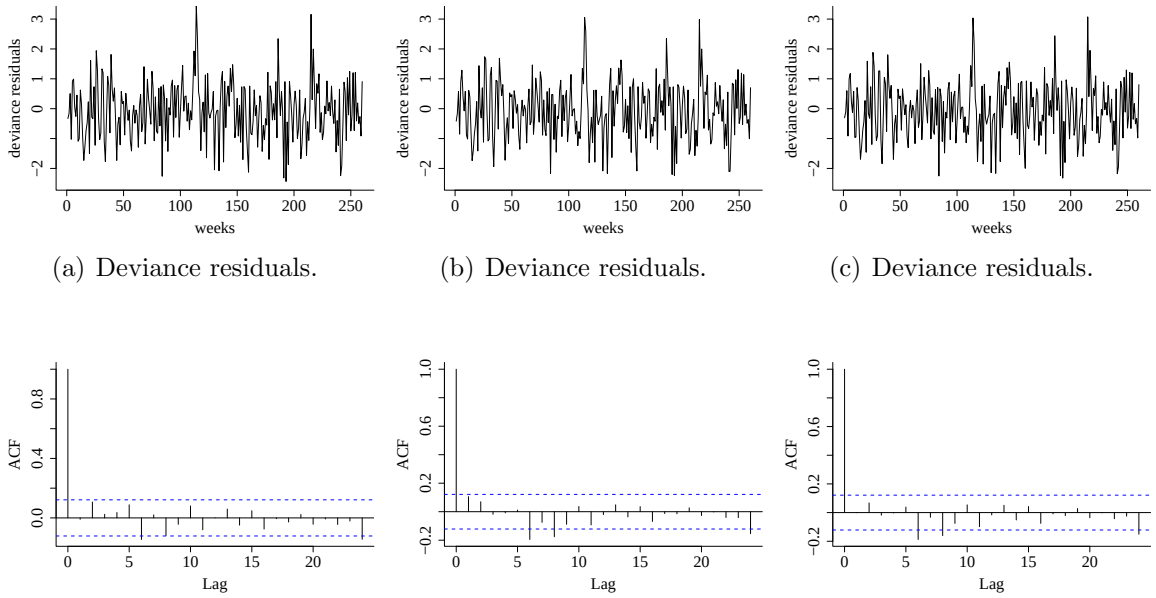


(k) Post. mean, ptw. (grey area) and simult. (dashed line) 95% CI of the pred. cases vs. Z_t (points).



(l) Post. mean, ptw. (grey area) and simult. (dashed line) 95% CI of the pred. cases vs. Z_t (points).

Figure 4.3: Results for meningococcal disease for the model with epid. comp.(left), random walk(middle), epid. comp. & random walk (right).



(a) Deviance residuals. (b) Deviance residuals. (c) Deviance residuals.

(d) Autocorrelation of the deviance residuals. (e) Autocorrelation of the deviance residuals. (f) Autocorrelation of the deviance residuals.

Figure 4.4: Deviance residuals for meningococcal disease for the model with epid. comp.(left), random walk(middle), epid. comp. & random walk (right).

4.5.1 The endemic component

The logarithm of endemic part of the mean ν_{it} is assumed to be the sum of a unit dependent part α_i, a time trend β_t and a seasonal part ζ_t,

$$\log \nu_{it} = \alpha_i + \beta_t + \zeta_t,$$

The unit dependent effects α_i are assumed to be independent normal distributed with a large variance

$$\boldsymbol{\alpha} \;=\; (\alpha_1, \ldots, \alpha_I) \sim \mathrm{N}(0, \sigma_\alpha^2 \boldsymbol{I}), \quad \sigma_\alpha^2 = 10^6.$$

The time trend β_t and the seasonal part ζ_t are defined as in the univariate case, where ζ_t does not include an intercept to avoid identification problems. Instead, α_i is now the intercept of unit i.

4.5.2 The epidemic component

For the epidemic part η_{it} we consider the following models shown in Table 4.6. Model 1 is the simplest model assuming an autoregression on the number of cases in the same unit, where the rate is the same in every unit. There is no interaction between the units. Model 2 assumes an additional regression on the sum of the cases of related units at the last time point, where the rate is the same in each unit. Model 3 and 4 are generalisations of model 1 and 2, respectively, allowing for a different rate for each unit. Model 5 allows for an individual effect from each unit on each unit and can be seen as a generalisation of model 4. Model 6 assumes that the cases of the last time point of each unit have an influence on all units, including the own, with total rate λ, which is the same for all units. The influence is distributed on the related units, including the own, according to the proportions π_{ji}, where $\sum_{(i=j)\vee(i\sim j)} \pi_{ji} = 1$. Model 7 is a generalisation of model 6 allowing for individual rates λ_j.

Model	η_{it}
1	$\lambda Z_{i,t-1}$
2	$\lambda Z_{i,t-1} + \phi \sum_{j\sim i} Z_{j,t-1}$
3	$\lambda_i Z_{i,t-1}$
4	$\lambda_i Z_{i,t-1} + \phi_i \sum_{j\sim i} Z_{j,t-1}$
5	$\sum_{(j=i)\vee(j\sim i)} \lambda_{ji} Z_{j,t-1}$
6	$\sum_{(j=i)\vee(j\sim i)} \lambda \pi_{ji} Z_{j,t-1}$
7	$\sum_{(j=i)\vee(j\sim i)} \lambda_j \pi_{ji} Z_{j,t-1}$

Table 4.6: The epidemic component.

For the rate parameter λ of model 1, as well as for all other rate parameters, we assume a Gamma prior with a Gamma hyperprior on the second parameter,

$$\begin{aligned} \lambda &\sim \text{Ga}(\alpha_\lambda, \beta_\lambda), \\ \beta_\lambda &\sim \text{Ga}(a, b). \end{aligned}$$

We will use $\alpha_\lambda = 1$, $a = 1$ and $b = 0.01$. For the proportions $\boldsymbol{\pi}_j = (\pi_{j1}, \ldots, \pi_{jI})$ we assume a Dirichlet prior (Denison et al., 2002, p. 243)

$$\boldsymbol{\pi}_j \sim \text{Di}(\alpha_{\pi j1}, \ldots, \alpha_{\pi jI}),$$

where $\alpha_{\pi jj} = 99$, $\alpha_{\pi ji} = 1/I_j$ if $i \sim j$, where I_j is the number of related units and $\alpha_{\pi ji} = 0$

otherwise. This means that we expect a proportion of 0.99 of the cases to be caused in the same region on average, with a standard deviation of around 0.01 and, in case of $I_j = 4$ related units, a two-sided 95% credibility interval of $(0.97, 1)$.

4.6 Estimation by MCMC

The key to a successful application of MCMC methods to the specified models lies, as in the univariate case, in a suitable decomposition of the model using auxiliary variables.

4.6.1 Alternative representation using auxiliary variables

We again introduce auxiliary variables for the endemic component $X_{it} \sim \text{Po}(\omega_{it}\nu_{it})$ and epidemic component $Y_{it} \sim \text{Po}(\omega_{it}\eta_{it})$ where $Z_{it} = X_{it}+Y_{it}$. Additionally we define auxiliary variables for the components of the epidemic component, related to the different units j that cause the cases, $Y_{jit} \sim \text{Po}(\omega_{it}\eta_{jit})$ where for example in model 3, $\eta_{iit} = \lambda Z_{i,t-1}$, and $\eta_{jit} = 0$ if $j \neq i$, and for model 4, $\eta_{iit} = \lambda_i Z_{i,t-1}$, $\eta_{jit} = \phi_i Z_{j,t-1}$ if $j \sim i$ and $\eta_{jit} = 0$ else. For the rest of the models η_{jit} is defined analogous. We get $Y_{it} = \sum_{j=1}^{I} Y_{jit}$. For the number of observed cases we get the representation $Z_{it} = X_{it} + \sum_{j=1}^{I} Y_{jit}$. We use the index i for the unit where the cases arise and j for the unit that causes the cases.

4.7 Update of the parameters

Conditioned on X_{it} the estimation of the endemic parameters α_i, β_t and γ_j is equivalent to the estimation in a Bayesian log-linear Poisson regression model with response variable X_{it}, and can be updated as in the univariate case. The parameters α_i are, as the parameters γ_j, updated individually, since a joint update leads to poor acceptance rates.

The full conditional of the parameter λ of model 1, as well as all other rates of model 2 to 7, is, conditioned on the corresponding auxiliary variable, a Gamma distribution, which makes it possible to update it in a simple Gibbs step, as the hyperparameter β_λ,

$$\begin{aligned} \lambda|Y_{it}, \beta_\lambda, Z_{it}, \ldots &\sim \text{Ga}(\alpha_\lambda + \sum_{i=1}^{I}\sum_{t=1}^{n} Y_{it}, \beta_\lambda + \sum_{i=1}^{I}\sum_{t=1}^{n} \omega_{it} Z_{it}), \\ \beta_\lambda|\lambda, Z_{it}, \ldots &\sim \text{Ga}(\alpha_\lambda(a+1), b+\lambda). \end{aligned}$$

The parameters $\boldsymbol{\pi}_j$ of model 6 and 7 are updated in a MH step. Their full conditional is

$$p(\boldsymbol{\pi}_j|\ldots) \quad \sim \quad \prod_{i=1}^{I}(\pi_{ji}^{\alpha_{\pi ji}+\sum_{t=1}^{n}Y_{jit}-1})\exp(-\sum_{i=1}^{I}\sum_{t=1}^{n}\omega_{it}\pi_{ji}\lambda_j Z_{j,t-1}).$$

We use the following Dirichlet distribution (Denison et al., 2002, p. 243) as proposal

$$\boldsymbol{\pi}_j^* \quad \sim \quad \mathrm{Di}(\alpha_{\pi j1}^*,\ldots,\alpha_{\pi jI}^*), \quad j=1,\ldots,I,$$

where $\alpha_{\pi ji}^* = \alpha_{\pi ji} + \sum_{t=1}^{n} Y_{jit}$, which has the density

$$p(\boldsymbol{\pi}_j^*) \quad = \quad \prod_{i=1}^{I}\pi_{ji}^{\alpha_{\pi ji}+\sum_{t=1}^{n}Y_{jit}-1}.$$

This leads to a relatively simple acceptance probability, since the terms of the proposal cancel with the corresponding terms in the full conditional,

$$acc \quad = \quad \max\left(1, \frac{\exp\left(-\sum_{i=1}^{I}\sum_{t=1}^{n}\omega_{it}\pi_{ji}^*\lambda_j Z_{j,t-1}\right)}{\exp\left(-\sum_{i=1}^{I}\sum_{t=1}^{n}\omega_{it}\pi_{ji}\lambda_j Z_{j,t-1}\right)}\right).$$

Since $\psi > 0$ we prefer to update $\tilde{\psi} = \log(\psi)$ with a simple Metropolis-Hastings Gaussian random walk proposal. The full conditional of ψ is

$$p(\psi|\ldots) \quad \propto \quad p(\psi)\prod_{i=1}^{I}\prod_{t=1}^{n}p(\omega_{it}|\psi)$$

and the corresponding full conditional of $\tilde{\psi}$ can be obtained through a change of variables. The variance of the random walk proposal is tuned automatically within the algorithm in order to obtain a suitable acceptance rate between 30 and 50% (Gelman et al., 1996). The full conditionals of the parameters ω_{it} are Gamma distributions $\omega_{it}|\ldots \sim \mathrm{Ga}(\psi+Z_{it}, \psi+\mu_{it})$ allowing for an update in a Gibbs step. The full conditional distribution of the auxiliary variables can be factorised as

$$\begin{aligned} P(X_{it}, Y_{it}, Y_{1it},\ldots,Y_{Iit}|Z_{it},\ldots) \quad &= \quad P(Y_{1it},\ldots,Y_{Iit}|Y_{it},X_{it},Z_{it},\ldots)P(Y_{it}|X_{it},Z_{it},\ldots) \\ &\quad \cdot P(X_{it}|Z_{it},\ldots), \end{aligned}$$

where the second term $P(Y_{it}|X_{it}, Z_{it}, \ldots)$ is deterministic: $Y_{it} = Z_{it} - X_{it}$. Due to the Poisson assumption the first and third term are multinomial and binomial respectively:

$$\begin{aligned}(Y_{1it}, \ldots, Y_{1it})|\ldots &\sim \text{Mu}\left(Y_{it}, \left(\frac{\eta_{1it}}{\eta_{it}}, \ldots, \frac{\eta_{Iit}}{\eta_{it}}\right)\right), \\ X_{it}|\ldots &\sim \text{Bin}\left(Z_{it}, \frac{\nu_{it}}{\mu_{it}}\right).\end{aligned}$$

4.8 Measles in the districts of Bavaria

We will now apply the seven models and a model without an epidemic component to the data of the measles outbreak of Bavaria (Figure 4.5), observed in the years 2001/02 in the seven districts of Bavaria. Every model is estimated with and without a random walk. Table 4.7 shows the deviance summaries of the different models. All models with an

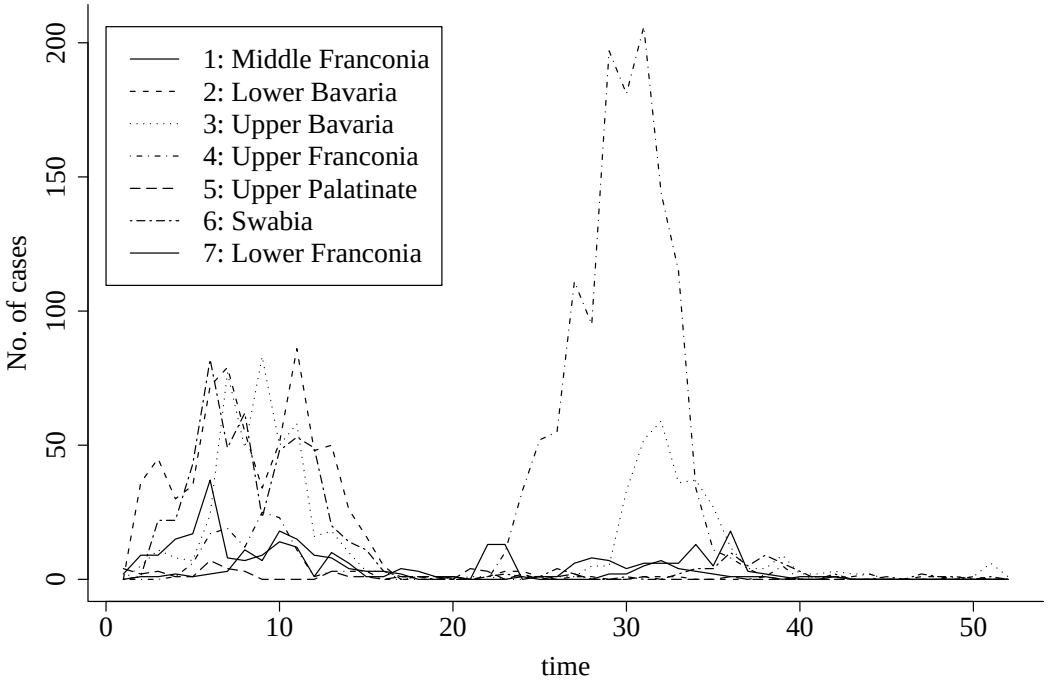


Figure 4.5: Measles in the seven districts of Bavaria.

epidemic component show a clearly better fit than the models without epidemic component with model 6 being the best model in terms of DIC. While the inclusion of the random walk

	$\bar{D}$	p_D	DIC
linear trend	1853	8	1861
random walk	1742	15	1757
model 1	1587	10	1597
model 2	1595	15	1610
model 3	1576	16	1592
model 4	1565	23	1589
model 5	1557	23	1580
model 6	1565	12	1577
model 7	1561	20	1580
model 1 with random walk	1585	11	1596
model 2 with random walk	1592	16	1609
model 3 with random walk	1574	17	1591
model 4 with random walk	1559	23	1582
model 5 with random walk	1556	27	1582
model 6 with random walk	1562	13	1575
model 7 with random walk	1558	21	1579

Table 4.7: Deviance summaries

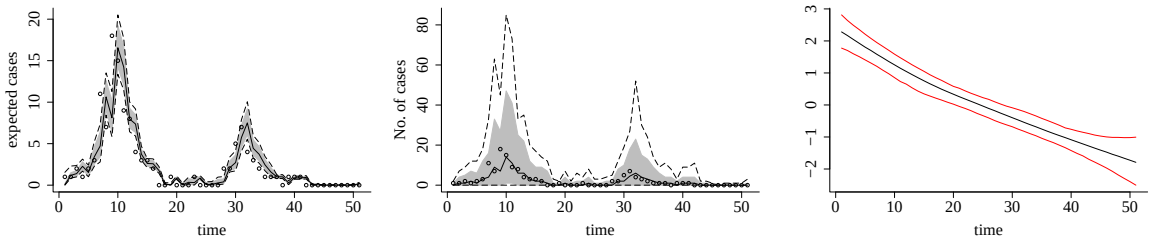
leads to a clearly better fit in the model without epidemic component, it has little effect on the DIC in the models with epidemic component. The posterior median of the parameters π_{ji} of model 6 with random walk are shown in Table 4.8. While for all regions the highest proportion of the cases is caused within the region, there is considerable interaction between the regions. The posterior median of the epidemic parameter λ is 0.84 and the two-sided credibility interval is $(0.72, 0.95)$. Most of the cases are therefore explained by the epidemic component. Figure 4.6 shows the results for model 6 with random walk. The estimated

	1	2	3	4	5	6	7
1	0.99	–	0	0	0	0	0
2	–	1	0	–	0	–	–
3	0.03	0	0.96	–	0	0.01	–
4	0	–	–	1	0	–	0
5	0	0	0	0	0.98	–	–
6	0	–	0.1	–	–	0.9	–
7	0.03	–	–	0.18	–	–	0.79

Table 4.8: Matrix of the posterior median of π_{ji} of model 6 with random walk.

time trend is actually linear, which is possibly the reason for the little effect of the random

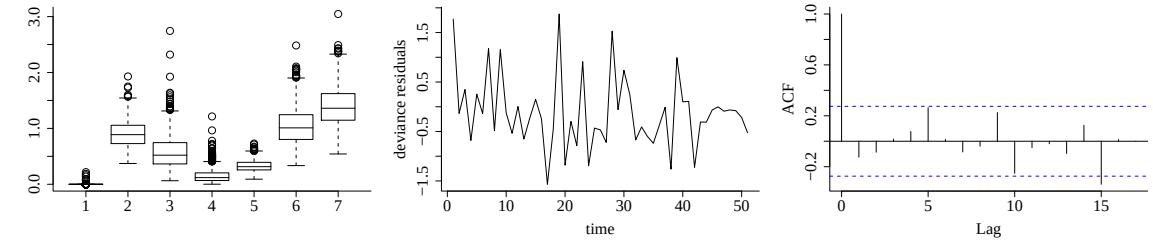
walk on the fit. The estimated cases fit the data quite well, which can also be seen by the predicted cases, where prediction is again based on the whole time series, and by the deviance residuals, which show little autocorrelation for region 1.



(a) Post. mean, pointwise (grey area) and simultaneous (dashed line) 95% credibility interval of μ_{1t} vs. Z_{1t} (points).

(b) Post. mean, pointwise (grey area) and simultaneous (dashed line) 95% credibility interval of the predicted cases of region 1 $1, \ldots, n$. vs. Z_{1t} (points).

(c) Post. mean and pointwise 95% credibility interval of structured time trend, β_t, $t = 1, \ldots, n$.



(d) Spatial effects $\exp(\alpha_i)$, $i = 1, \ldots, I$.

(e) Deviance residuals of region 1.

(f) Autocorrelation of the deviance residuals of region 1.

Figure 4.6: Results of model 6 with random walk.

4.9 Discussion

In the multivariate cases the inclusion of an epidemic component leads to a clearly better fit, which is reached by the inclusion of an autoregression as well as interactions between the units. In the univariate case it can be seen that the inclusion of an autoregression is not flexible enough to apparently improve the fit. A promising extension of the model would be to allow the parameter of the epidemic component to vary over time.

Chapter 5

A two-component model for counts of infectious diseases

5.1 Introduction

The distinction between an endemic and an epidemic component introduced in Chapter 4 has shown to be a promising concept. However, the time constant parameter of the epidemic component λ shows some limitations: (a) if the observed number of cases Z_t does not explode, what is rarely the case, the epidemic parameter is restricted to $\lambda < 1$; (b) in the regular case $\lambda \leq 1$ the epidemic parameters is just the proportion of the epidemic cases Y_t from the total number of cases Z_t. The latter means that the proportion of epidemic cases is constant over time, i.e. the same in outbreaks and periods of no outbreak. However, the observation-driven form of the epidemic component, with an autoregression on the number of cases of the previous time point, gives the model some flexibility to explain for outbreaks.

A promising extension of the model described in Chapter 4 is to allow the epidemic parameter to vary over time, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$, as proposed in Held et al. (2006). This leads to the following improvements of the model: (a) the epidemic parameter $\boldsymbol{\lambda}$ is no longer restricted to be lower or equal unity, but can exceed unity for some time period, where the number of cases indeed explodes, and then go back to a value below or equal to unity; (b) the proportion of the epidemic cases Y_t from the total number of cases Z_t is no longer forced to be constant over time, but can switch between different levels, even to values greater than unity.

These improvements allow to capture the epidemic characteristics of an infectious disease clearly better. Direct, i.e. person to person transmitted diseases, on the one hand,

show big outbreaks with an indeed exponentially increasing number of cases in the beginning and a slow decrease of the cases due to the reduction of susceptible individuals towards the end of the outbreak. Additionally, control measure as vaccination of susceptible individuals and the isolation of infected individuals, can lead to a reduction of the number of cases. Between the outbreaks there are often long periods with no or just some sporadic cases, until a new outbreak arises. At the beginning of the outbreak, the epidemic parameter jumps from a value below unity to a value that usually clearly exceeds unity. The reduction of susceptible individuals and control measures lead to a usually slow reduction of the epidemic parameter to a value below unity. Indirect transmitted infectious disease, as food or water borne infectious diseases, on the other hand, usually show a pronounced endemic component. Outbreaks are usually caused by point sources as contaminated food from a hotel or a food distributor. Outbreaks are often characterized by a short or permanently slightly increased number of cases. In case of an outbreak, the epidemic parameter jumps upwards to a value, that is usually below unity.

To capture the characteristics of the epidemic parameter $\boldsymbol{\lambda}$, which are sudden changes at the beginning of an outbreak as well as smooth decreases during the outbreak, it is useful to allow for an adaptive amount of smoothing. A state-space or dynamic model (e.g. Jørgensen et al., 1999; Fahrmeir and Knorr-Held, 2000) with an autoregressive or random walk prior on $\boldsymbol{\lambda}$ shows a constant amount of smoothing due to the time-constant correlation structure and is therefore not appropriate, since it does not allow for sudden changes. A Bayesian changepoint model with unknown number of changepoints (e.g. Denison et al., 2002) described in Chapter 3 is better suited for this setting, since it allows the correlation to vary over time. Through Bayesian model averaging, the estimated epidemic parameter $\boldsymbol{\lambda}$ may still be smooth, because it is obtained through averaging over different changepoint models of variable dimension with different location of the changepoints (Green, 1995; Clyde, 1999).

This chapter is based on the work in Held et al. (2006). In Section 2 we will present the model and the estimation using MCMC methods. In Section 3 the behavior of the model is examined by an application to simulated data and in Section 4 the model is applied to the hepatitis A, hepatitis B and meningococcal disease data, introduced in Chapter 4, and possibilities to build a outbreak detection system on the model are considered. The chapter concludes with a discussion of the results and an outline of possible extensions in Section 5.

5.2 Model

Let Z_t, $t = 1, \ldots, n$ be the observed number of counts at time t. We assume the data to be Poisson distributed, conditioned on the auxiliary variables ω_t,

$$Z_t|\omega_t \sim \text{Po}(\omega_t \mu_t),$$

where the auxiliary variables are Gamma distributed with shape and rate parameter ψ, $\omega_t \sim \text{Ga}(\psi, \psi)$. As described in Chapter 3 this leads to a negative binomial distribution for Z_t integrating out ω_t, with mean μ_t and dispersion parameter ψ, $Z_t \sim \text{Nb}(\mu_t, \psi)$. A Gamma prior is assumed for the dispersion parameter, $\psi \sim \text{Ga}(\alpha_\psi, \beta_\psi)$. We will use $\psi \sim \text{Ga}(1, 0.1)$ assuming ψ to have a mean of 10 and a large variance of 100. The mean μ_t is assumed to be the sum of an endemic part ν_t, that explains the regular amount of cases and can be interpreted as a baseline and an epidemic part η_t, that explains for epidemic activity as outbreaks or irregularities in the data,

$$\mu_t = \eta_t + \nu_t.$$

As in Chapter 4 we define an endemic component X_t and an epidemic component Y_t giving the number of endemic and epidemic cases, where $X_t|\omega_t \sim \text{Po}(\omega_t \nu_t)$ and $Y_t|\omega_t \sim \text{Po}(\omega_t \eta_t)$.

5.2.1 The endemic component

The logarithm of the endemic mean ν_t is defined, analog to Chapter 4, as the sum of L harmonic waves of different frequencies an intercept and a linear time trend. Due to the complexity of the epidemic component, we will not include a random walk. Even the linear time trend will not be included in the applications in Section 3 and 4.

5.2.2 The epidemic component

The mean of the epidemic component η_t is defined by an autoregression on the number of cases at the previous time point, $\eta_t = \lambda_t Z_{t-1}$, where the epidemic parameter $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$ is now assumed to follow a Bayesian changepoint model (see Chapter 3) with unknown number K and locations $\theta^{(1)} < \ldots < \theta^{(K)}$ of changepoints. For convenience we define $\theta^{(0)} = 0$, $\theta^{(K+1)} = n$ and $\boldsymbol{\theta} = (\theta^{(0)}, \ldots, \theta^{(K+1)})$. The parameters $\boldsymbol{\lambda}$ are then defined

as

$$\lambda_t \quad = \quad \lambda^{(k)} \quad \text{if} \quad t = \theta^{(k-1)} + 1, \ldots, \theta^{(k)}.$$

We can use the interpretation of the time constant epidemic parameter λ of Chapter 4, to obtain insight to the characteristics of the time-varying epidemic parameter $\boldsymbol{\lambda}$. In the time constant case, we can distinguish two cases, $\lambda < 1$ and $\lambda > 1$. In the first case, $\lambda < 1$, the epidemic parameter can be interpreted as the proportion of the epidemic cases Y_t from the total number of cases Z_t. In the second case, $\lambda > 1$, the number of cases increases exponentially. These properties can be transferred to the time-varying case. If $\lambda_t > 1$ the number of cases increases exponentially. If $\lambda_t < 1$ the process converges to a state where λ_t is the proportion of the epidemic cases Y_t from the total number of cases Z_t. However, due to the autoregression on Z_{t-1} the number of epidemic cases Y_t need some time to converge, especially if the process has grown exponentially for a long time, before reaching a state with $\lambda_t < 1$.

For $\lambda^{(k)}$, $k = 1, \ldots, K+1$ we assume independent Gamma priors, with a Gamma hyperprior on the second parameter,

$$\begin{aligned} \lambda^{(k)} &\sim \text{Ga}(\alpha_\lambda, \beta_\lambda), \quad k = 1, \ldots, K+1 \\ \beta_\lambda &\sim \text{Ga}(a, b) \end{aligned}$$

This choice implies that the marginal prior distribution of $\lambda^{(k)}$ is Gamma-Gamma (see Bernardo and Smith, 1994, page 120). In our applications we use $\alpha_\lambda = 1$, $a = 10$, $b = 10$ where the Gamma-Gamma marginal of $\lambda^{(k)}$ turns out to be simply an F-distribution with degrees of freedom equal to 2 and 20. This choice gives a marginal prior probability of 0.39 to the event $\lambda^{(k)} \geq 1$, while always favouring smaller values of $\lambda^{(k)}$ (the density function has a unique mode at zero and is monotonically decreasing).

The number of changepoints and the location given the number of changepoints are assumed to be uniformly distributed,

$$\begin{aligned} P(K = k) &= 1/n, \quad k = 0, 1, \ldots, n-1, \\ P(\boldsymbol{\theta}|K) &= \binom{n-1}{K}^{-1}. \end{aligned}$$

5.2.3 Statistical analysis via MCMC

The key to a successful application of MCMC methods to the specified model lies again in the decomposition of Z_t into X_t and Y_t. The parameters of the endemic component the overdispersion parameters and the auxiliary variables are updated similar to Chapter 4.

The parameters of the epidemic component are updated using the reversible jump algorithm of Green (1995) applied to the changepoint model in Chapter 3. The acceptance probability $\alpha((\boldsymbol{\lambda}, \boldsymbol{\theta}, K), (\boldsymbol{\lambda}^*, \boldsymbol{\theta}^*, K^*))$ of a new changepoint m is

$$\min\left(1, \frac{c\beta_\lambda^{\alpha_\lambda}\Gamma(\alpha_{\lambda,m-1,m})\Gamma(\alpha_{\lambda,m,m+1})(\beta_{\lambda,m-1,m+1})^{\alpha_{\lambda,m-1,m+1}}}{\Gamma(\alpha_\lambda)\Gamma(\alpha_{\lambda,m-1,m+1})(\beta_{\lambda,m-1,m})^{\alpha_{\lambda,m-1,m}}(\beta_{\lambda,m,m+1})^{\alpha_{\lambda,m,m+1}}}\right),$$

where

$$\begin{aligned}\alpha_{\lambda,k-1,k} &= \alpha_\lambda + \sum_{t=\theta^{(k-1)}+1}^{\theta^{(k)}} Y_t,\\ \beta_{\lambda,k-1,k} &= \beta_\lambda + \sum_{t=\theta^{(k-1)}+1}^{\theta^{(k)}} \omega_t Z_{t-1},\end{aligned}$$

and c is the proposal ratio defined in Chapter 3. The full conditional of the parameters $\lambda^{(k)}$, $k = 1, \ldots, K+1$ and β_λ are $\lambda^{(k)}|\ldots \sim \text{Ga}\left(\alpha_{\lambda,k-1,k}, \beta_{\lambda,k-1,k}\right)$ and $\beta_\lambda|\ldots \sim \text{Ga}\left(a + K + 1, b + \sum_{k=1}^{K+1}\lambda^{(k)}\right)$.

The reversible jump algorithm is very fast compared to the update of the other parameters. However, due to the simple design of the steps, the algorithm mixes very slowly, which leads to very long runs of the MCMC algorithm. We will therefore update the changepoints and steps 10 times per iteration. This clearly improves the mixing and leads to shorter runs and run times.

5.2.4 One-step ahead prediction

Of particular interest in infectious disease surveillance are *short-term* predictions, in particular one-step-ahead predictions. Our model is well suited for this setting, since it is based on the entire available time series and does not assume that there are no outbreaks in the past. While outbreak detection could be based e.g. on the posterior probability $P(\lambda_n \geq 1)$, the predictive distribution of the number of new cases Z_{n+1} is perhaps of more direct public health importance.

We omit the technical details here, but note only that with obvious modifications, the model can be written down for data $Z_1, \ldots, Z_{n+1}$ where the counts Z_{n+1} are missing. This allows us to simulate from the posterior predictive distribution of ν_{n+1} and λ_{n+1} and subsequently of $Z_{n+1}|Z_n \sim \text{Po}(\nu_{n+1} + \lambda_{n+1}Z_n)$ in the Poisson case. If we include overdispersion, samples from $\omega_{n+1} \sim \text{Ga}(\psi, \psi)$ based on the posterior samples of ψ are generated and subsequently $Z_{n+1}|Z_n \sim \text{Po}(\omega_{n+1}(\nu_{n+1} + \lambda_{n+1}Z_n))$ is simulated.

However, due to the Markov structure of the changepoint model (see Chapter 3) there is a simpler way to obtain the posterior predictive distribution of λ_{n+1} and Z_{n+1} based on a model for $Z_1, \ldots, Z_n$ only. One step ahead predictions of the epidemic parameter λ_{n+1} can easily be obtained using the predictive distribution (3.15)-(3.17).

Note, that the predictive distribution of λ_{n+1} is a mixture of two components. One component (which corresponds to the case that there is a changepoint between Z_n and Z_{n+1}) is, due to the independence of $\lambda^{(k)}$, $k = 1, 2, \ldots, K+2$, the conditional prior distribution $\lambda^{(K+2)}|\beta_\lambda \sim \text{Ga}(1, \beta_\lambda)$. Here the posterior samples of β_λ enter. The other component, which corresponds to the case of no changepoint between Z_n and Z_{n+1} is the posterior of $\lambda^{(K+1)}$. The mixing weights are determined by the probability p, say, for a changepoint between Z_n and Z_{n+1}.

For fixed number of changepoints $K = k$ among $n-1$ possible locations, the probability p is just $(K+1)/(n+1)$ (c.f. (3.11)). In each iteration of the algorithm we hence simulate the posterior predictive distribution of λ_{n+1} with probability $(K+1)/(n+1)$ from the conditional prior distribution $\lambda^{(K+2)}|\beta_\lambda \sim \text{Ga}(1, \beta_\lambda)$, otherwise we set $\lambda_{n+1} = \lambda^{(K+1)}$. Note how nicely the posterior distribution of K determines the probability for a changepoint in the future in the sense that the more changepoints there are in the past, the more likely is it that there will be a changepoint in the future.

We finally note that m-step predictions, if required, may be obtained by sequentially repeating this process given the current number of breakpoints up to time $n + m - 1$. At this point it is worth noting that for *long-term* predictions, eventually only the posterior of the endemic part $\boldsymbol{\nu}$ will enter, while the epidemic part will reduce to the conditional prior distribution with large probability.

5.3 Analysis of simulated data

To study the flexibility of the changepoint model we first present an analysis of simulated data ($n = 200$, $\rho = 2\pi/52$). While the simulation in Held et al. (2006) is based on a

sufficient number of simulated cases and without simulating overdispersion, we will here consider a simulation of the more adverse case of a low number of cases, implying little information in the data, that are additionally overdispersed. It seems natural to look at this adverse case, since it is usually the case in real data. The true $\boldsymbol{\lambda}$ sequence (Figure 5.2(e)) is piecewise constant with two changepoints at $\theta_1 = 60$ and $\theta_2 = 70$. The parameter λ switches from $\lambda^{(1)} = 0.1$ to $\lambda^{(2)} = 0.7$ and then back to $\lambda^{(3)} = 0.1$. This comes close to the epidemic structure of a food or water borne disease, with little epidemic activity for longer periods and an outbreak that equals one caused by a point source. The endemic parameters are chosen to reflect the behaviour of a more common infectious disease with approximately 20 weekly endemic cases on average with a seasonal minimum of approximately 10 and a seasonal maximum of approximately 41 cases: $\gamma_0 = 3$, $\gamma_1 = 0.5$ and $\gamma_2 = 0.5$. We allowed for overdispersion in the simulation, using $\psi = 10$. The data are analysed with the proposed model and the results are shown in Figure 5.1 and 5.2.

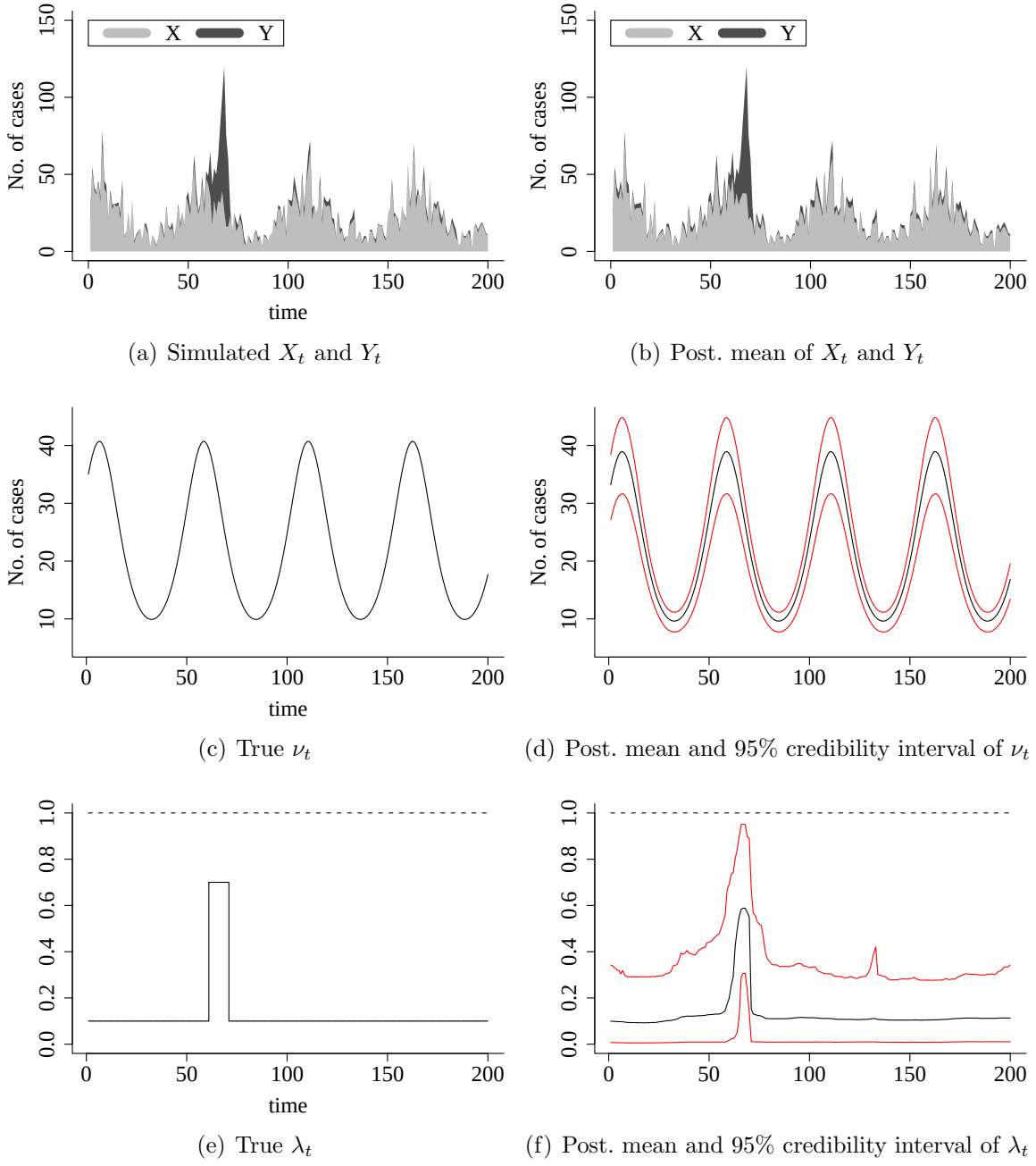
Although there is not much information in the data, due to the relatively low number cases, and the data are simulated with overdispersion, the model is able to detect the changepoint structure quite well. There is some uncertainty about the number and location of the changepoints. However, with a bit more information in the data the number and locations of the changepoints can be identified very well (see Held et al., 2006). The second changepoint can be detected better compared to the first. This can be explained by the low disease incidence at the first changepoint, so the model has more information to precisely determine the location of the second than the first changepoint. Consequently, the estimated $\boldsymbol{\lambda}$ sequence is smooth around the first changepoint, but abrupt at the second. Note also that the seasonal structure in the data has been estimated quite well, see Figure 5.1 (c) and (d).

5.4 Analysis of real data

We analyse weekly surveillance data on hepatitis A and B from Germany from the years 2001 to 2004 (208 weeks so $n = 207$) and meningococcal disease from Germany from the years 2001 to 2005 (261 weeks so $n = 260$) (Figure 5.3) as introduced in Chapter 4.

5.4.1 Hepatitis A

Figure 5.4 displays the results from our model applied to the hepatitis A time series. We have used $L = 4$ seasonal terms. One can see that a strong seasonal pattern has been



(a) Simulated X_t and Y_t

(b) Post. mean of X_t and Y_t

(c) True ν_t

(d) Post. mean and 95% credibility interval of ν_t

(e) True λ_t

(f) Post. mean and 95% credibility interval of λ_t

Figure 5.1: Simulated data for known ν_t and λ_t (left panel) and posterior estimates (right panel).

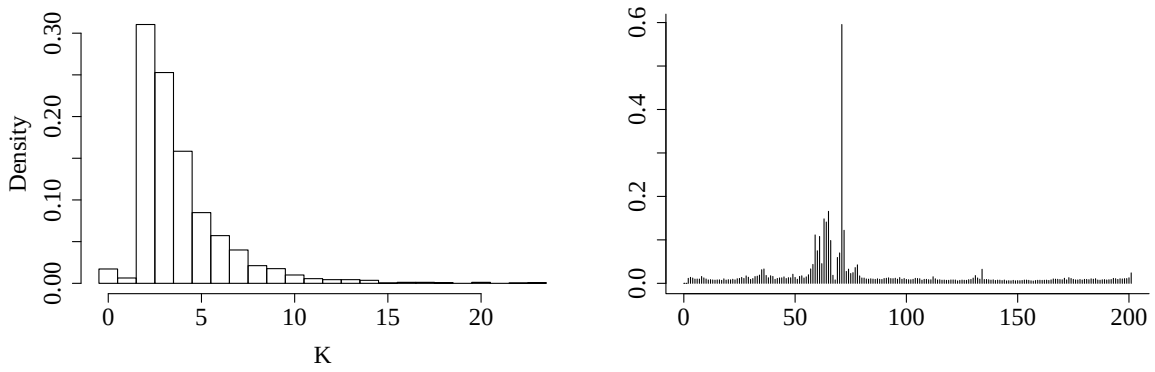
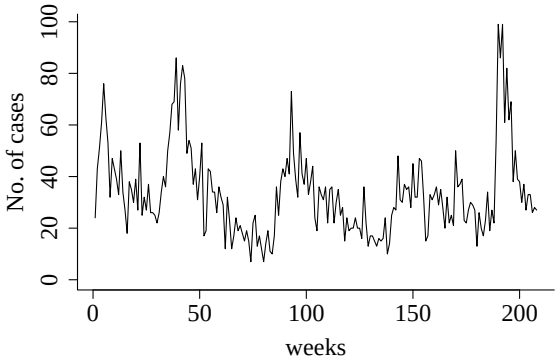


Figure 5.2: Post. distribution of K (left) and posterior probability for a changepoint at time t (right)

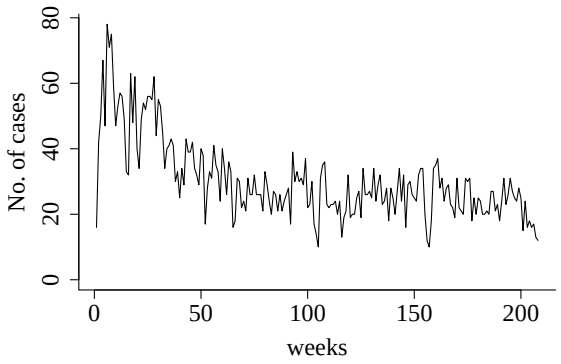
estimated which peaks in September. This peak may be caused by travel-associated cases. Between 13 (in July) and 38 (in September) cases per week can be attributed to the regular endemic incidence pattern, see Figure 5.4(b).

The epidemic parameter 5.4(c) shows an increased level of around 0.5 on average, during the first year, which can be interpreted as the result of a lower vaccination coverage. In November 2001 the epidemic parameter decreases to values close to 0 and stays there for a long period. In 2004 a temporary increase of the epidemic parameter can be observed, starting with a slight increase in March to 0.4 on average followed by a second increase in August with values up to 0.7 on average. This second increase in the high holiday season (August) is discussed further in Anonymous (2004), and can be linked to holiday-makers in a certain hotel in Egypt. Outbreaks occurred also in other European countries. The increase of the epidemic cases can also be seen in Figure 5.4(a).

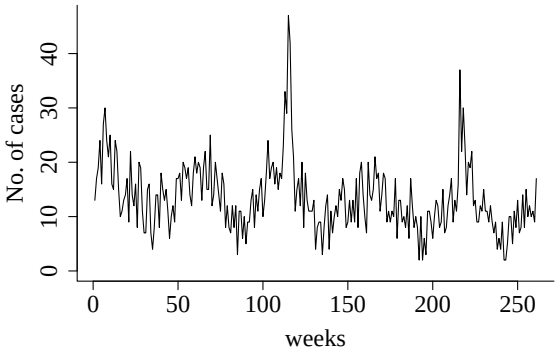
To illustrate the predictive capabilities of the model we show in Figure 5.5 the one-step-ahead predictive distribution of number of counts for the weeks $t = 188$ and $t = 189$, based on the data up to week $t = 187$ and $t = 188$, respectively. These two weeks represent the beginning of the documented (Anonymous, 2004) outbreak: the actually observed counts are 22 in week 187, 54 in week 188, and 99 in week 189. It is interesting to see that the predictive distribution for week $t = 188$ is fairly symmetric, the actually observed value has some borderline support with $P(Z_{188} \geq 54|Z_1, \ldots, Z_{187}) = 0.01$. However, the model immediately reacts to this unusually high value and consequently, the predictive distribution for $t = 189$ has a long tail towards larger values. The actually observed



(a) Hepatitis A.



(b) Hepatitis B.

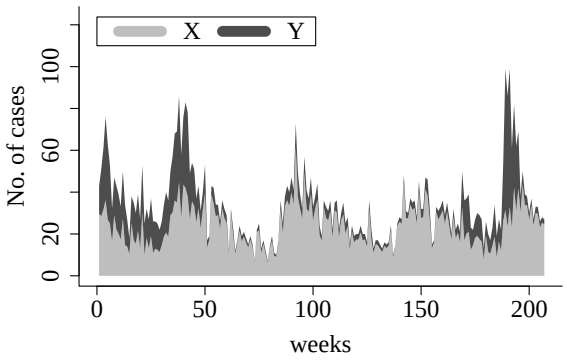


(c) Meningococcal disease.

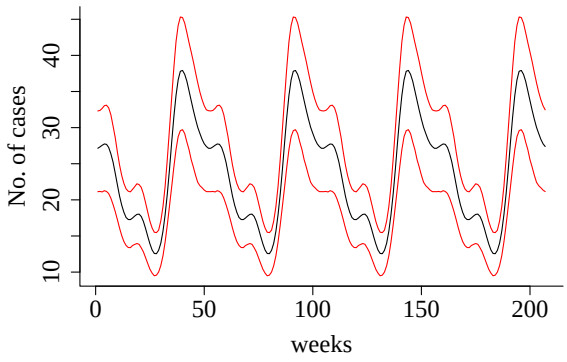
Figure 5.3: The three considered infectious disease counts time series.

number of counts is now well supported by the predictive distribution with $P(Z_{189} \geq 99|Z_1, \ldots, Z_{188}) = 0.13$.

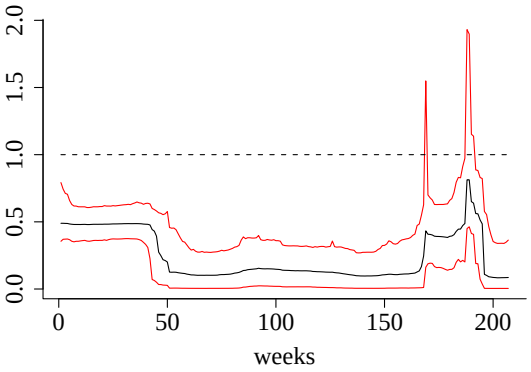
Considering the documented Anonymous (2004) outbreak in August 2004, we will discuss 3 approaches for an outbreak detection system, based on: 1) the probability $P(\lambda_n \geq 1)$; 2) the posterior median of the epidemic parameter λ_n; 3) the upper 99.9% credibility limit of the predictive endemic cases X_n. We will therefore use the data up to week 187 and 188 according to a decision based on a prediction 3) or no prediction 1), 2), respectively. For approach 1) an outbreak will be flagged if $P(\lambda_{189} \geq 1)$ is greater than 1%. Also other choices could be made. Using approach 2) an outbreak is flagged if the mean of the epidemic parameter is above 0.2. This is a useful choice with respect to the uncertainty about the mean of the parameter. For approach 3) an outbreak is flagged if the observed value



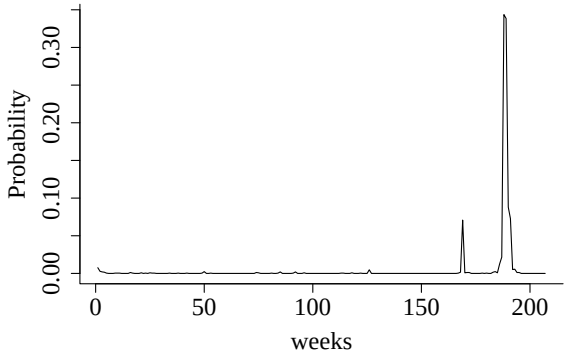
(a) Post. mean of X_t and Y_t



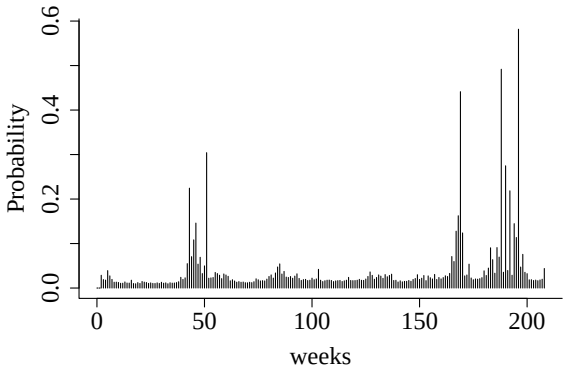
(b) Post. mean and 95% credibility interval of ν_t



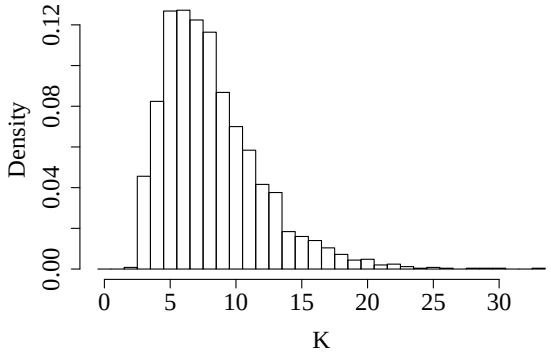
(c) Post. mean and 95% credibility interval of λ_t



(d) Post. probability $P(\lambda_t \geq 1|\boldsymbol{Z})$



(e) Posterior probability for a changepoint at location t



(f) Posterior distribution of K.

Figure 5.4: Results for hepatitis A

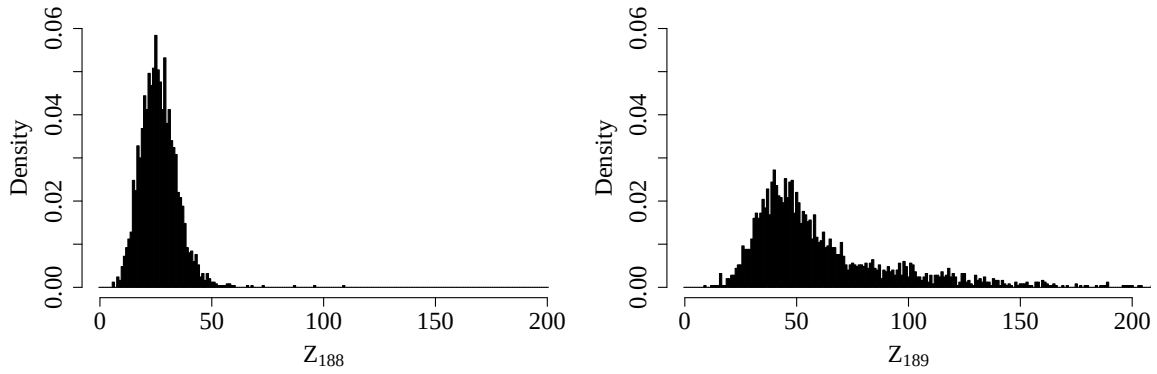


Figure 5.5: Predictive distribution of $Z_{188}|Z_1, \ldots, Z_{187}$ (left) and $Z_{189}|Z_1, \ldots, Z_{188}$ (right).

Z_{188} is above the one-sided 99.9% credibility interval of the predicted endemic cases X_{188} based on the data up to week 187. A false alarm is then flagged every 1000-th week on average, i.e. approximately once in 3 years. On could also use the estimation of X_{188} based on the data up to time 188. We use the one step ahead prediction of X_{188}, since a decision about an outbreak is then directly available as the value Z_{188} is observed. For comparison, we have applied the Farrington et al. (1996) algorithm for outbreak detection in week $t = 188$. We used the implementation available in the package `surveillance` (Höhle and Riebler, 2005). An eleven week window has been chosen with three years of historical data (2001-2003). More data are not available, as the German surveillance system has been set up in 2001. An outbreak is flagged if the actually observed number of counts is larger than an upper threshold, defined as the 99.9% quantile of the predictive distribution (based on a normal approximation of the transformed counts). The results depend on whether or not a linear time trend is included in the model. If included, the observed number of cases $Z_{188} = 54$ are flagged as an outbreak, as the upper threshold is 50.6. However, this is mainly due to the estimated (decreasing) time trend, since the reference values in 2001 are unusually high, with up to 70 cases, see Figure (5.3). The algorithm does downweight these observations, but only to a certain degree, so the decreasing time trend remains significant. If we apply the algorithm without a time trend, the upper threshold is 77.2 so no alarm is flagged. Of course, these results also depend heavily on the nominal false positive rate.

All 3 approaches considered for our model clearly detect the outbreak at time 188. In the case of the considered hepatitis A outbreak, we get $P(\lambda_{188} \geq 1) = 0.24$ which is clearly above the threshold value 0.01. The posterior median of the epidemic parameter is 0.48 which is also clearly above the threshold value 0.2. The upper 99.9% credibility limit of

the predictive distribution of X_{188} is 41, i.e. the observed value 54 is clearly above the limit. The latter approach is similar to the one in Farrington et al. (1996). However, we eliminate the effect of past outbreaks by the estimation of an epidemic component instead of an arbitrary downweight of to high values. This should give an unbiased estimation of the endemic component. In the case of the considered hepatitis A outbreak, this leads to a better outbreak detection.

5.4.2 Hepatitis B

Figure 5.6 now displays the results from our model applied to the hepatitis B time series. One can see that there is virtually no seasonality present, so the sinusoidal terms could as well have been omitted in the model. The autoregressive parameter λ_t decreases smoothly from values around 0.65 to values well close to 0.1, which can be interpreted as a consequence of the increasing vaccine coverage. The posterior mode of the number of changepoints is just three; however, the possible locations of the changepoints are more dispersed than in the hepatitis A example, so the estimated λ_t's are smoother than for hepatitis A. This nicely illustrates the smoothing capabilities of the model through Bayesian model averaging.

5.4.3 Model comparison

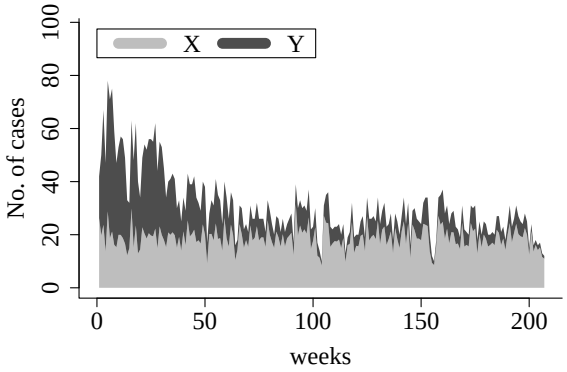
Table 5.1 shows the mean deviance $\bar{D}$, the estimated number of parameters p_D and the deviance information criterion (DIC) (Spiegelhalter et al., 2002) for the hepatitis A, hepatitis B and additional for the meningococcal disease data, that were also analyzed in Chapter 4. There is a clear improvement compared to the model with time-constant

	$\bar{D}$	p_D	DIC
Hepatitis A	1472	30	1501
Hepatitis B	1376	16	1392
Meningococcal disease	1473	16	1488

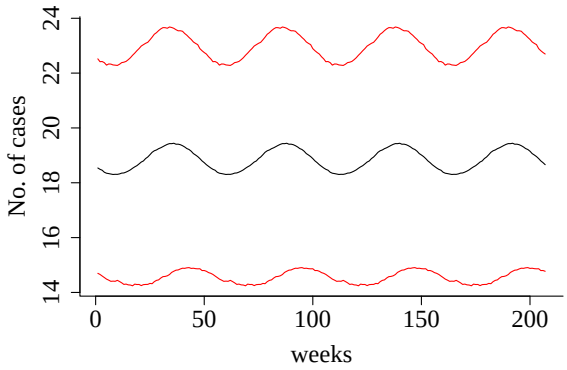
Table 5.1: Deviance summaries

epidemic parameter of Chapter 4, especially for the hepatitis A data. Compared to the model with random walk of Chapter 4, there is no considerable difference in terms of DIC.

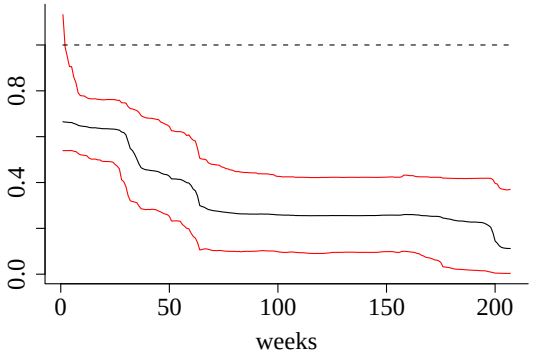
To further compare the model with the one in Chapter 4 we consider the expected and predicted cases for the meningococcal disease data (Figure 5.7 (a),(b)), where prediction is



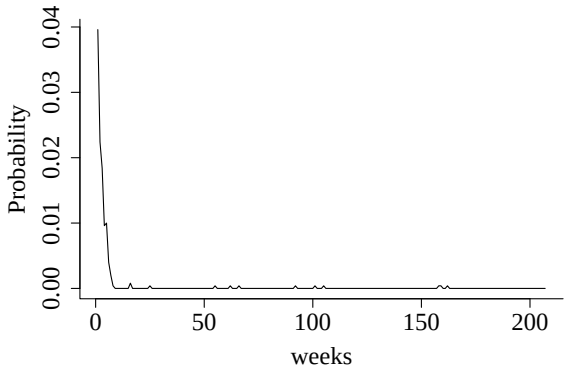
(a) Post. mean of X_t and Y_t



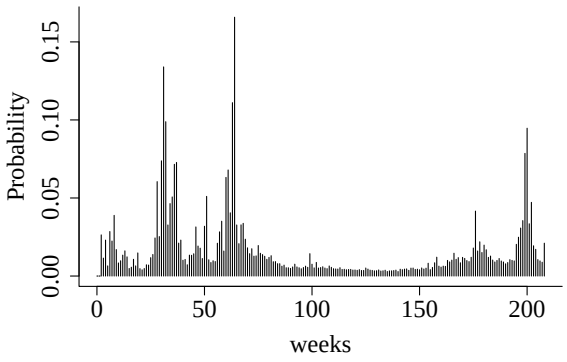
(b) Post. mean and 95% credibility interval of ν_t



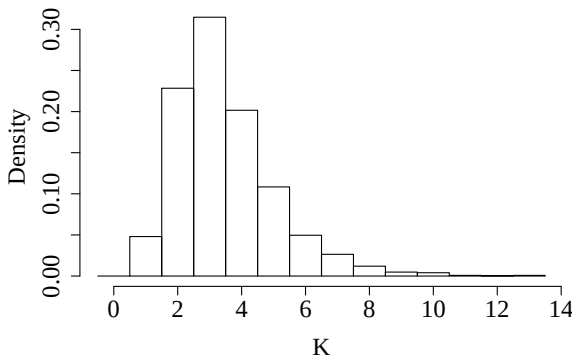
(c) Post. mean and 95% credibility interval of λ_t



(d) Post. probability $P(\lambda_t \geq 1|\boldsymbol{Z})$



(e) Posterior probability for a changepoint at location t



(f) Posterior distribution of K.

Figure 5.6: Results for hepatitis B

based on all observations. The flexibility of the time-varying epidemic parameter modeled by a changepoint model allows for a clearly better prediction of the outbreaks, especially of the 2003 outbreak, compared to the model with time constant epidemic parameter and random walk (Figure 4.3). The Pearson residuals (Figure 5.7 (c),(d)) show little autocorrelation.

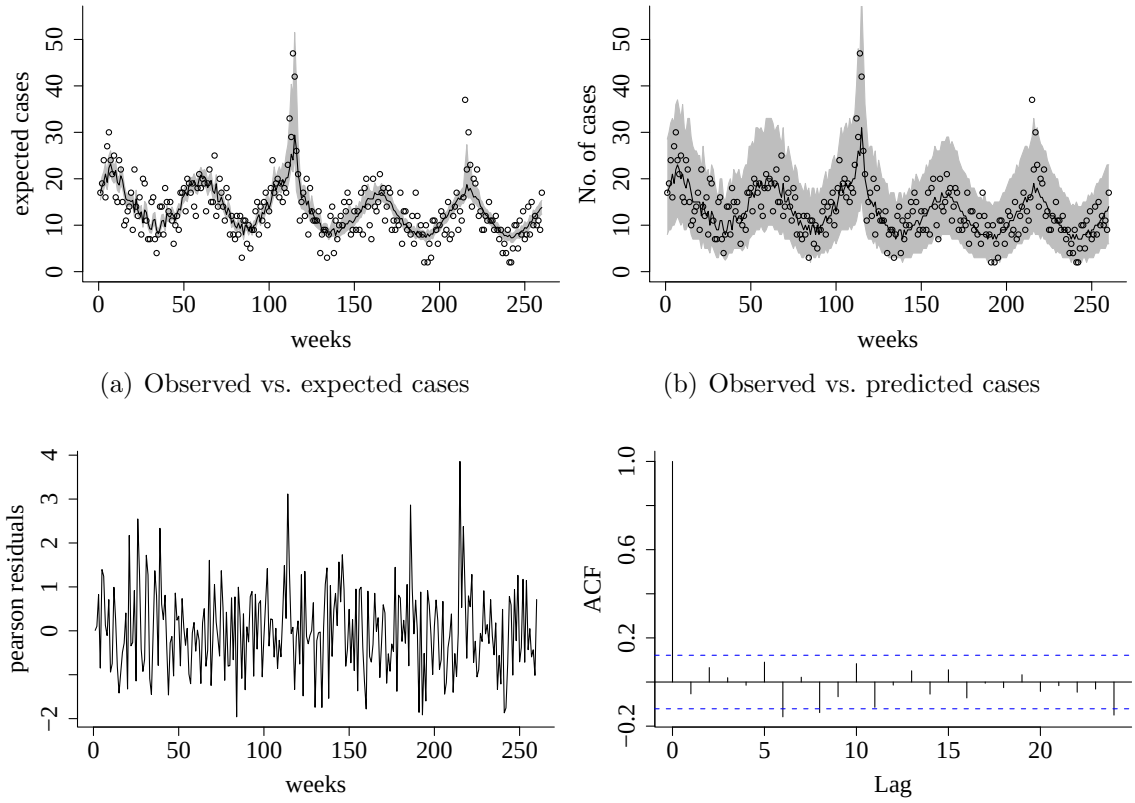
5.5 Discussion

In this chapter we have introduced a generic model for time series of infectious disease counts. The central assumption of the model is, like in Chapter 4, that the disease counts can be viewed as the sum of an endemic and an epidemic component. The proportion of the epidemic component λ_t is now allowed to vary over time according to a Bayesian multiple changepoint model with an adaptive amount of smoothing. This allows for better modelling of the outbreaks in the data. Analyses of simulated and real data have illustrated how the model can be used for modelling diseases with increasing vaccination coverage (hepatitis B) and for detecting and predicting outbreaks (hepatitis A). Furthermore, the comparison with the models considered in Chapter 4 shows that the model fits the data quite well, especially the outbreaks in the data. While the model can be seen as a general approach to model infectious diseases, it is also particularly suited for the challenging task of outbreak detection.

We now comment on some other extensions. For routine use in prospective disease surveillance, a *sequential* algorithm for inference will be helpful (Sonesson and Bock, 2003). Note in this context that the changepoint model used here has indeed such a sequential representation, based on the Markov structure of the changepoint problem (Chapter 3), whereas our current implementation in `twins` is based on a retrospective analysis, given a fixed amount of data. Sequential updating of the parameter estimates could be based, for example, on particle filtering (e.g. Berzuini and Gilks, 2003; Andrieu et al., 2001; Doucet et al., 2001a) or the forward-backward algorithm (e.g. Scott, 2002; Fearnhead, 2006) and we will consider such an algorithmic modification in Chapter 6. For example, we might simply fix the estimates of the global model parameters $\boldsymbol{\nu}$ and update only $\boldsymbol{\lambda}$. A similar approach has been advocated in Brix and Diggle (2001) and Diggle et al. (2005) for spatiotemporal prediction, see also Diggle et al. (2003). However, we note that all (retrospective) analyses in this chapter take only little time compared to the weekly resolution in which surveillance data are typically collected. Nevertheless, a fast sequential algorithm will be useful for a

detailed study of the predictive qualities of our model.

A multivariate or perhaps even spatial extension of our model is the other area with a lot of potential in applications. For example in ecological regression one might be interested to relate the endemic incidence ν or the epidemic parameter λ to area-level covariates. Also the area of monitoring disease outcomes across multiple units is of great interest in practice (Marshall et al., 2004; Kleinman et al., 2004). We will consider a multivariate extension of the model in Chapter 7.



(a) Observed vs. expected cases

(b) Observed vs. predicted cases

(c) Pearson residuals

(d) Autocorrelation of the Pearson residuals

Figure 5.7: Posterior median and pointwise 95% credibility interval (grey area) of the data mean $\mu_{i,t}$, $t = 1, \ldots, n$ (top, left) and the predicted cases (top, right) plotted against the data (points), Pearson residuals (bottom, left), and autocorrelation of the Pearson residuals (bottom, right), for meningococcal disease.

Chapter 6

Sequential Monte Carlo methods for the estimation of the two component model

We will consider two types of sequential Monte Carlo methods for the estimation of the two component model proposed in Chapter 5: the forward-backward algorithm and the particle filter. Both methods are described in Chapter 2 and applied to a Bayesian changepoint model in Chapter 3. While the forward-backward algorithm can be used as an alternative to the reversible jump algorithm for the update of the changepoint model for the retrospective analysis within the MCMC algorithm, the particle filter can be used for a prospective analysis of the changepoint model conditioning on fixed values for the other parameters.

6.1 Retrospective analysis using the forward-backward algorithm

The changepoint problem of the two component model proposed in Chapter 5 is updated using a reversible jump algorithm. This algorithm is very fast. However, due to the simple design of the steps, the algorithm mixes very slow, which leads to very long runs of the MCMC algorithm. In Chapter 5 this problem is tackled by updating the changepoint problem 10 times per iteration, which leads to shorter runs and run times.

However, instead of sampling from the full conditional of the changepoints using a slow mixing reversible jump algorithm, one can sample directly from the full conditional using

the forward-backward algorithm described in Chapter 3. The reversible jump step is then replaced by a simple Gibbs step. This clearly improves the mixing of the algorithm.

6.1.1 Update of the changepoints using the forward-backward algorithm

The algorithm proceeds in analogy to the algorithm described in Chapter 3. The probabilities of the segments $S^{(k)} = (Y_{\theta^{(k-1)}+1}, \ldots, Y_{\theta^{(k)}})$, $k = 1, \ldots, K+1$ given the changepoints are now

$$\begin{aligned} P(S^{(k)}|\theta^{(k-1)}, \theta^{(k)}) &= \frac{\prod_{t=\theta^{(k-1)}+1}^{\theta^{(k)}} (Z_{t-1})^{Y_t} \beta_\lambda^{\alpha_\lambda} \Gamma(\alpha_{\lambda,k-1,k})}{\prod_{t=\theta^{(k-1)}+1}^{\theta^{(k)}} Y_t! \Gamma(\alpha_\lambda) (\beta_{\lambda,k-1,k})^{\alpha_{\lambda,k-1,k}}}, \\ \alpha_{\lambda,k-1,k} &= \alpha_\lambda + \sum_{t=\theta^{(k-1)}+1}^{\theta^{(k)}} Y_t, \\ \beta_{\lambda,k-1,k} &= \beta_\lambda + \sum_{t=\theta^{(k-1)}+1}^{\theta^{(k)}} Z_{t-1}. \end{aligned}$$

The probability of the changepoints are the same as in Chapter 3. The full conditional of the changepoints depends now additionally on the auxiliary variables Y_t, $t = 1, \ldots, n$.

6.1.2 Performance of the forward-backward algorithm

We will now compare the forward-backward with the reversible jump algorithm. We have estimated the model for the hepatitis A data (Figure 5.3) using the two algorithms. A sample of size 2500 has been obtained by taking every 10-th iteration after a burn in of 2000 samples. Table 6.1 shows the the effective sample size (see Section 2.1.5) of K and the calculation time in seconds. Calculations were done on a 2.80 GHz PC.

reversible jump		forward-backward		
ESS	CPU (s)	ESS	CPU (s)	Rel. ESS
105	116	380	12928	0.026

Table 6.1: Effective sample size of K and estimation time in seconds for the reversible jump algorithm (column 1 and 2) and the forward-backward algorithm (column 3 and 4), and the relative effective sample size standardized by the CPU time of the two algorithms (ESS(rj)/CPU(rj))/(ESS(fb)/CPU(fb)) (column 5), for the hepatitis A data.

Additionally, the relative effective sample size of the two algorithms has been calculated. This is the ratio of the effective sample sizes of the two algorithms, each standardized by the CPU time, i.e. (ESS(1)/CPU(1))/(ESS(2)/CPU(2)).

The forward-backward algorithm shows a better mixing than the reversible jump algorithm, which can be seen in a higher effective sample size. However, the reversible jump algorithm is clearly faster, resulting in a clear advantage in terms of relative effective sample size.

Figure 6.1 shows the trace and the autocorrelation of the parameter K and the overdispersion parameter ψ. It can be seen that the improved mixing of the changepoints also affects the mixing of the other parameters, since also the overdispersion parameter ψ shows a slightly improved mixing.

6.2 Particle filter for prospective surveillance

The two component model described in Chapter 5 has proved to be a promising approach to tackle the problem of infectious disease surveillance. However, as mentioned in Chapter 5, one disadvantage of the model with respect to the surveillance of a disease is that the estimation using Markov chain Monte Carlo (MCMC) methods is retrospective. In other words, the problem is treated as fixed sample situation. Since the data collected in public health systems are observed sequentially in time, the model has to be estimated for the complete time series, as a new observation arrives. For the use in prospective disease surveillance, a sequential algorithm for inference would therefore be helpful (Sonesson and Bock, 2003). As proposed in Chapter 5 the model could be estimated for the currently observed time series, and then the time-constant parameter of the endemic component could be fixed for a prospective analysis. The posterior mean of this parameter could for example be used. This procedure can also be an advantage in terms of outbreak detection, as mentioned in Diggle et al. (2003), who proceed in a similar way by fixing the baseline of their model, since fixing this parameter reduces the variation in the model and may lead to more sensitivity in the outbreak detection. We will show that by fixing the parameter of the endemic component and two further parameters, the overdispersion parameter and the hyperparameter of the epidemic component, a sequential analysis of the epidemic parameter, that is used for the outbreak detection in Chapter 5, can be realized using particle filters, as described in Andrieu et al. (2001). This is further simplified by the fact that the signal process of our model, that is the changepoint model, has the Markov

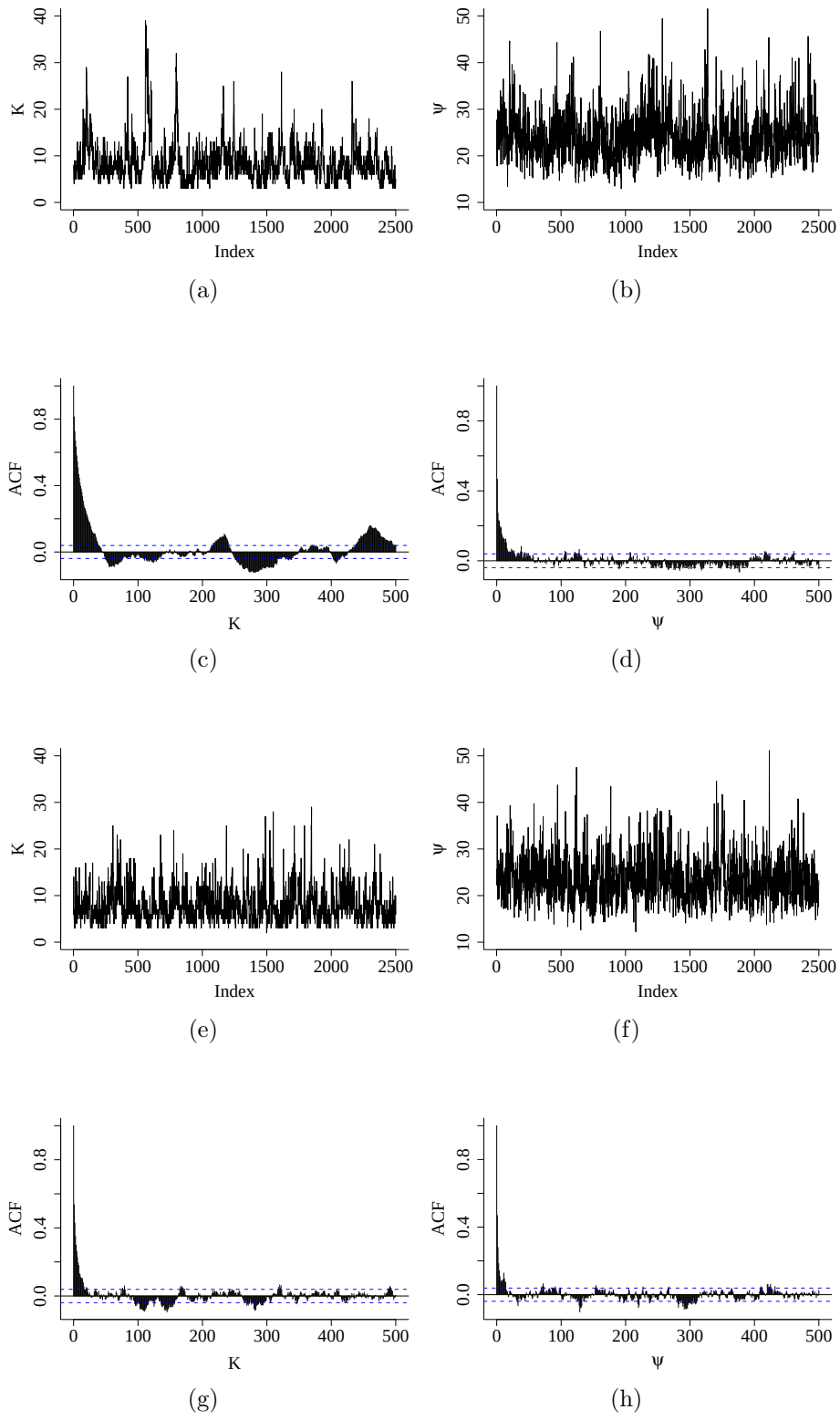


Figure 6.1: Trace and autocorrelation of K and ψ for the reversible jump algorithm (top) and the forward-backward algorithm (bottom).

property, as shown in Chapter 3.

6.2.1 Sequential update using the particle filter

We will use the particle filter described in Chapter 3. As proposed in that chapter the model can be first estimated up to the current time n_{MCMC} using the MCMC algorithm described in Chapter 5 to get point estimates of the parameters to fix, $\boldsymbol{\gamma}$, ψ and β_λ. As new observations arrive, these parameters are fixed to the posterior mean, and the model is estimated sequentially using the particle filter.

6.2.2 Application to the hepatitis A data

We will now apply the particle filter to the hepatitis A data introduced in Chapter 4. The model is first estimated up to time $n_{MCMC} = 107$ using the MCMC algorithm, to obtain point estimates of the parameters, that have to be fixed. We then fix these parameters and estimate the models up to time $n = 108, \ldots, 207$ using the particle filter. It took approximately 6.5 minutes on a 2.80 GHz PC to obtain 2500 samples of each of the 100 models, which is 4 seconds per model on average. For comparison, the MCMC algorithm for the model up to time 107 needed approximately 13.5 minutes.

It is now possible to perform model validation using the predictive distribution, which has been estimated for each of the 100 models. We will illustrate this by calculating the probability integral transform (PIT) values u_t (Frühwirth-Schnatter, 1996) which are obtained from the predictive probabilities $P(Z_t^{pred} \leq Z_t)$ of the observed value Z_t. By a randomization of these probabilities one obtains PIT values, that are uniformly distributed on $[0, 1]$ if the model is true, $u_t = (1 - \alpha_t)P(Z_t^{pred} \leq (Z_t - 1)) + \alpha_t P(Z_t^{pred} \leq Z_t)$, where α_t is uniformly distributed on $[0, 1]$. Corresponding histograms can be "de-randomized" by adding rectangles between $P(Z_t^{pred} \leq (Z_t - 1))$ and $P(Z_t^{pred} \leq Z_t)$ with height $1/(P(Z_t^{pred} \leq Z_t) - P(Z_t^{pred} \leq (Z_t - 1)))$ (Held, 2006).

Figure 6.2 shows the "de-randomized" histogram of the PIT values for $t = 108, \ldots, 207$. The distribution of the PIT values is neither concave nor convex, i.e. the variance of the predictive distribution has been estimated well. A slight decrease of the PIT values can be noted, i.e. the predictive distribution overestimates the number of cases. However, 100 predictive distributions are relatively few to get insights about the predictive performance of the model.

As supposed, the model is now more sensitive in terms of outbreak detection, which can

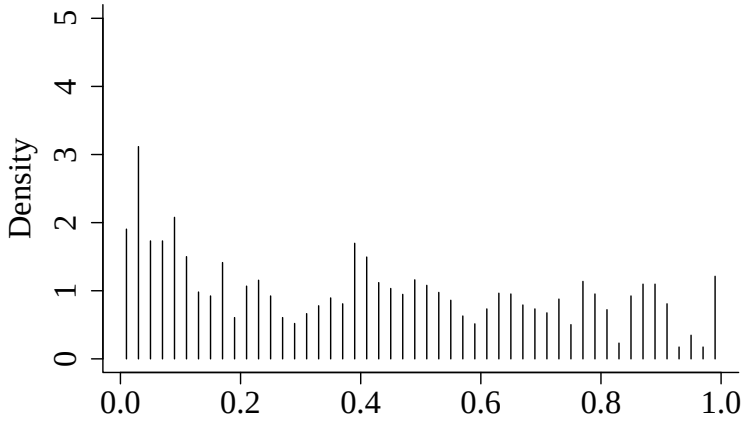


Figure 6.2: Histogram of the PIT values u_t, $t = 108, \ldots, 207$.

be seen by the probability $P(\lambda_{188} \geq 1)$ which is 0.42 now, compared to 0.28 in Chapter 5.

6.3 Discussion

Using the particle filter for the update of the model drastically reduces the calculation time, if a prospective estimation of the model is the aim of the analysis. This is of particular advantage for the surveillance of infectious diseases, where new observations arrive sequentially. Another advantage is that model validation based on the predictive distribution is now possible within moderate time. However, this is at the cost of fixing the time constant parameters.

Chapter 7

A model for multivariate time series of infectious disease counts

7.1 Introduction

In this chapter we propose a model for the analysis of multivariate count data of infectious diseases, that are observed in units, e.g. spatial regions, age groups or pathogenes, and over time. The model is a multivariate extension of the model proposed by Held et al. (2006), which is described in Chapter 5. The characteristics of infectious disease counts differ substantially depending on the type of infectious disease. While indirect transmitted diseases, as food or water-borne diseases often show a strong seasonality but only short outbreaks, also called 'hyperendemic' periods (Knorr-Held and Richardson, 2003) or longer periods of a slightly increased number of cases, caused e.g. by contaminated food from a food manufacturer, direct transmitted diseases that spread from person to person show big outbreaks with a rapidly increasing number of infected, that we call epidemic periods. However, infectious diseases do not necessarily fit in one of these two groups, e.g. meningococcal disease shows seasonality and small outbreaks as well as some bigger outbreaks where the disease spreads from person to person.

Our aim is to develop a model that can handle the different types of infectious diseases described above. As in the other chapters, we therefore include two components in our model, an endemic and an epidemic component, where the endemic component explains a possible baseline rate of cases, that may include seasonality and differences between units and the epidemic component that should be able to explain for all kinds of outbreaks. The counts are then the sum of these two components, which may, may not or may to a certain

amount be part of a disease. A central role in our model plays the design of the epidemic component, since it should be able to capture the characteristics of a disease that spreads from person to person across units and over time. The most realistic models for the spread of an infectious disease over time are mechanistic models as the chain binomial or the SIR model (Anderson and Britton, 2000) that directly model the infection process of the spread from person to person on an individual level. However, these models require information that is only available in very special cases, when it was possible to observe the infection process directly. Additionally, information on the number of susceptible individuals has to be available. However, there is a branching process approximation of the chain binomial model (Anderson and Britton, 2000), and the idea of a branching process (e.g. Guttorp, 1995) is used in Held et al. (2006) to model the epidemic part of a disease. While the branching process approximation of the chain binomial model is based on the generation time of the disease, the model proposed in Held et al. (2006) is based on the observation time of the counts, which makes the parameters not directly comparable. However, the so defined model should capture the characteristics of an outbreak.

The branching process approximation of the chain binomial model is based on the assumption of an unlimited amount of susceptibles, and is especially a good approximation if the disease is at an early stage. At a later stage the decreasing number of susceptibles slows down the growth of the disease, that can not be explained by a branching process with unchanged offspring distribution. To overcome this problem Held et al. (2006) let the parameter of the offspring distribution, which is the parameter of the epidemic component, vary over time, assuming a Bayesian changepoint model (Denison et al., 2002), that provides an adaptive amount of smoothing. The epidemic component can then explain the decreasing growth due to a decreasing number of susceptibles, but also other factors that influence the growth as control measures.

Another problem of a simple branching process is that it dies out if the number of cases goes down to zero. The same applies for all mechanistic models, which are therefore only used to explain one outbreak. Since the aim in Held et al. (2006) is to explain all outbreaks that arise over a longer time, there are periods without cases and Held et al. (2006) therefore base their model on a branching process with immigration. If there is no outbreak, the parameter of the epidemic component should be close to zero and jump to a higher level if there is a new outbreak.

The model proposed in Held et al. (2006) has shown to be able to model the incidence over time of all kinds of infectious diseases, including those where person to person trans-

mission is present. However, infectious disease counts are often available as multivariate time series observed in different age groups or as space-time data in different regions. To be able to make use of this additional information, a model that can explain the spread of a disease over multiple units is needed. Another scenario, where such a model could be useful is the influence between different disease types, as influenza and meningococcal disease, where an increased number of meningococcal disease cases is related to an influenza outbreak.

To provide such an approach, we extend the model proposed in Held et al. (2006) to a multivariate version, that should be suitable to explain the spread of a disease across units. The extension is based on the multivariate model of Chapter 4.

The most challenging part of our model is the spread of the disease across units. We propose seven different models, of different complexity, and compare them as regards content and using the DIC (Spiegelhalter et al., 2002). The models are all based on the idea of a multivariate branching process with immigration (Mode, 1971).

There is a vast amount of models for infectious diseases counts present in the literature. Most of them model the disease either over time (e.g. Stroup et al., 1989; Farrington et al., 1996) or space (e.g. Besag et al., 1991; Clayton et al., 1993; Knorr-Held and Raßer, 2000).

Most approaches for space-time disease counts assume an "endemic" setting, i.e. that there are no outbreaks in the data. Kleinman et al. (2004) propose a generalized linear mixed model approach including seasonality and spatial random effects under the assumption of no outbreak. Knorr-Held and Besag (1998) apply a Bayesian hierarchical dynamic model, including components for time, space and age groups, amongst others, to lung cancer data, that is suitable for "endemic" disease counts. Yan and Clayton (2006) attempt to find disease cluster in time and space. Jørgensen et al. (1999) propose a state space model for longitudinal count data assuming the incidence rate to be the product of a time dependent latent Markov process and time-spatial covariates. Schmid and Held (2004) apply a Bayesian age-period-cohort model with an additional spatial component to cancer registry data for the extrapolation of space-time trends.

Knorr-Held and Richardson (2003) propose a Bayesian hierarchical dynamic model including components for time, season and space with an additional autoregressive component, that allows for dependence between regions, for small outbreaks without person to person transmission, so called 'hyperendemic' periods, that can be switched on and off.

A model for the analysis of space-time counts of influenza, a disease where person to person transmission plays the major role, is proposed by Mugglin et al. (2002). The

spread of the disease over time and space is modeled by a latent multivariate Gaussian AR(1) process, which has a similar structure as a multivariate branching process with immigration. In contrast with the branching process with immigration, the multivariate Gaussian AR(1) process has a continuous state space and allows for correlation, whereas the branching process with immigration assumes independent offspring distributions for different individuals. However, instead of allowing for outbreaks by an increase of the parameters of the autoregressive coefficient matrix, which corresponds to the offspring matrix, leading to non stationary periods, which is the natural characteristic of an outbreak with regard to the branching process approximation of the chain binomial model, they explain the outbreaks by different levels of the so called 'epidemic forcing term', which corresponds to the branching process immigration, which explains the outbreaks by switches between different possibly stationary distributions. The incidence can switch between three levels, where the periods of the levels have to be predetermined. The exponent of the latent multivariate Gaussian AR(1) process enters multiplicative in the incidence rate, which seems unnatural with regard to the branching process approximation of the chain binomial model.

The influence of influenza outbreaks on meningococcal disease is studied e.g. in Jensen et al. (2004) and Hubert et al. (1992).

7.2 Model

Let Z_{it}, $i = 1, \ldots, I$, $t = 1, \ldots, n$ be the observed number of counts in unit i at time t. We assume the data to be Poisson distributed, conditioned on the auxiliary variables ω_{it},

$$Z_{it}|\omega_{it} \quad \sim \quad \text{Po}(\omega_{it}\mu_{it}).$$

where the auxiliary variables ω_{it} are again introduced to adjust for overdispersion. We consider two types of overdispersion. On the one hand, overdispersion could be seen identical in every unit, assuming the random effects to be identical distributed as

$$\omega_{it} \quad \sim \quad \text{Ga}(\psi, \psi), \quad \psi > 0.$$

When the units are different types of diseases this assumption is unlikely to hold. If the units are age-groups or regions, this may be a realistic assumption, and we may prefer the simpler model.

On the other hand, overdispersion could be seen to differ between units. In this case,

we assume every unit to have an individual amount of overdispersion,

$$\omega_{it} \sim \text{Ga}(\psi_i, \psi_i), \quad \psi_i > 0, \quad i = 1, \ldots, I.$$

The marginal distribution of Z_{it} integrating out ω_{it} is then a negative binomial distribution

$$Z_{it} \sim \text{NegBin}\left(\mu_{it}, \psi_i\right),$$

where $\text{NegBin}(a, b)$ is a negative binomial distribution with mean a and dispersion parameter b, with the same mean as in the Poisson case but a larger variance:

$$V[Z_{it}] = E[Z_{it}]\left(1 + \frac{E[Z_{it}]}{\psi_i}\right).$$

For $\psi \to \infty$ it can be seen that $V[Z_{it}] \to E[Z_{it}]$.

For the parameters ψ_i, $i = 1, \ldots, I$, Gamma priors are assumed,

$$\psi_i \sim Ga(\alpha_\psi, \beta_\psi).$$

For the first type of overdispersion we get a negative binomial distribution with dispersion parameter ψ instead of ψ_i for Z_{it}.

The mean μ_{it} is assumed to be the sum of an endemic part ν_{it}, that explains the regular amount of cases and can be interpreted as a baseline and an epidemic part η_{it}, that explains for epidemic activity as outbreaks or irregularities in the data, and the interaction of epidemic cases between units,

$$\mu_{it} = \eta_{it} + \nu_{it}.$$

The multivariate extension of the two components is based on the multivariate design of the model considered in Chapter 4.

7.2.1 The endemic component

The logarithm of the endemic part of the mean ν_{it}, also referred to as endemic mean, is assumed to be decomposed to a unit dependent part α_i and a time dependent part ζ_t,

$$\log \nu_{it} = \alpha_i + \zeta_t,$$

where the time dependent part ζ_t is assumed to be the sum of L harmonic waves and an intercept. This can be represented in the following form (e.g. Diggle, 1990),

$$\zeta_t = \sum_{j=0}^{J} \gamma_j s_{tj}, \tag{7.1}$$

with $J = 2L$ and

$$\begin{aligned} s_{t0} &= 1, \\ s_{tj} &= \begin{cases} \sin(\frac{\rho t(j+1)}{2}) & \text{for } j = 1, 3, \ldots, 2L-1, \\ \cos(\frac{\rho tj}{2}) & \text{for } j = 2, 4, \ldots, 2L, \end{cases} \end{aligned}$$

where ρ is the base frequency, e.g. $\rho = \frac{52}{2\pi}$ for weekly observed data. This assumes that the endemic components of two units i_1 and i_2 are proportional and just differ from each other by the factor $\exp(\alpha_{i_1})/\exp(\alpha_{i_2})$. Independent normal priors are assumed for the parameters γ_j, $j = 0, \ldots, J$ and α_i, $i = 0, \ldots, I$,

$$\begin{aligned} \boldsymbol{\gamma} &= (\gamma_0, \ldots, \gamma_J) \sim \mathrm{N}(0, \sigma_\gamma^2 \boldsymbol{I}), \quad \sigma_\gamma^2 = 10^6, \\ \boldsymbol{\alpha} &= (\alpha_1, \ldots, \alpha_I) \sim \mathrm{N}(0, \sigma_\alpha^2 \boldsymbol{I}), \quad \sigma_\alpha^2 = 10^6. \end{aligned}$$

Alternatively we could assume the seasonality of the regions to differ also by a phase shift. In this case we assume the following form of ν_{it},

$$\log \nu_{it} = \alpha_i + \zeta_{it},$$

where ζ_{it} now depends on the unit i,

$$\zeta_{it} = \sum_{j=1}^{J} \gamma_{ij} s_{tj}. \tag{7.2}$$

Note that ζ_{it} does not include an intercept.

An additional modification could be to consider the different population sizes of the regions n_i by assuming

$$\mu_{it} = \eta_{it} + \frac{n_i}{n} \nu_{it}.$$

where $n = \sum_{i=1}^{I} n_i$ is the total population. The proportion $\frac{n_i}{n}$ is included in $\exp(\alpha_i)$ in the original model.

7.2.2 The epidemic component

For the epidemic part we consider seven different models shown in Table 7.1, that mainly differ from each other by the assumptions that are made on interactions between the units. The models are similar to the models for the epidemic component of Chapter 4

Model	η_{it}
1	$\lambda_t Z_{i,t-1}$
2	$\lambda_t Z_{i,t-1} + \phi_t \sum_{j\sim i} Z_{j,t-1}$
3	$\lambda_{it} Z_{i,t-1}$
4	$\lambda_{it} Z_{i,t-1} + \phi_{it} \sum_{j\sim i} Z_{j,t-1}$
5	$\sum_{(j=i)\vee(j\sim i)} \lambda_{jit} Z_{j,t-1}$
6	$\sum_{(j=i)\vee(j\sim i)} \lambda_t \pi_{ji} Z_{j,t-1}$
7	$\sum_{(j=i)\vee(j\sim i)} \lambda_{jt} \pi_{ji} Z_{j,t-1}$

Table 7.1: The epidemic component.

but allow for time-varying rate parameters, that follow a changepoint model, as the rate parameter $\boldsymbol{\lambda}$ in Chapter 5. Since the complexity of the model is now primarily determined by the number of time-varying parameters, model 6 is now comparatively simple with just one time-varying parameter. We will introduce the structure of the rate parameters $\boldsymbol{\lambda}_{ji} = (\lambda_{ji1}, \ldots, \lambda_{jin})$, $j = 1, \ldots, n$, $i = 1, \ldots, n$ for model 5, the rate parameters of the other models being defined in analogy. That is, the parameters $\boldsymbol{\lambda}_{ji} = (\lambda_{ji1}, \ldots, \lambda_{jin})$, $j = 1, \ldots, n$, $i = 1, \ldots, n$ are piecewise constant over time with unknown number of changepoints K_{ji} and unknown location of the changepoints $\theta_{ji}^{(1)} < \ldots < \theta_{ji}^{(K_{ji})}$. For convenience we define $\theta_{ji}^{(0)} = 0$, $\theta_{ji}^{(K_{ji}+1)} = n$ and $\boldsymbol{\theta}_{ji} = (\theta_{ji}^{(0)}, \ldots, \theta_{ji}^{(K_{ji}+1)})$. The parameter vectors $\boldsymbol{\lambda}_{ji}$, $j = 1, \ldots, n$, $i = 1, \ldots, n$ are then defined as:

$$\lambda_{ijt} \quad = \quad \lambda_{ji}^{(k)} \quad \text{if} \quad t = \theta_{ji}^{(k-1)} + 1, \ldots, \theta_{ji}^{(k)}.$$

For $\lambda_{ji}^{(k)}$, $i = 1, \ldots, I$, $j = 1, \ldots, I$, $k = 1, \ldots, K_{ji} + 1$ we assume independent Gamma

priors with a Gamma hyperprior on the second parameter,

$$\begin{aligned} \lambda_{ji}^{(k)} &\sim \text{Ga}(\alpha_\lambda, \beta_{\lambda ji}), \quad k = 1, \ldots, K+1 \\ \beta_{\lambda ji} &\sim \text{Ga}(a, b). \end{aligned}$$

We used the same specification as in Chapter 5, $\alpha_\lambda = 1$, $a = b = 10$. For the number of changepoints and for the location $\boldsymbol{\theta}_{ji} = (\theta_{ji}^{(1)}, \ldots, \theta_{ji}^{(k_{ji})})$ given the number of changepoints we assumed a uniform prior,

$$\begin{aligned} P(K_{ji}) &= 1/n, \quad k = 0, 1, \ldots, n-1 \\ P(\boldsymbol{\theta}_{ji} | K_{ji}) &= \binom{n-1}{K_{ji}}^{-1}. \end{aligned}$$

For the proportions $\boldsymbol{\pi}_j = (\pi_{j1}, \ldots, \pi_{jI})$ of model 6 and 7 we assume a Dirichlet prior (Denison et al., 2002, p. 243),

$$\boldsymbol{\pi}_j \sim \text{Di}(\alpha_{\pi j1}, \ldots, \alpha_{\pi jI}),$$

where $\alpha_{\pi jj} = 99$, $\alpha_{\pi ji} = 1/I_j$ if $i \sim j$, where I_j is the number of related units, and $\alpha_{\pi ji} = 0$ otherwise. This means that we expect a proportion of 0.99 of the cases to be caused in the same region on average, with a standard deviation of around 0.01 and, in case of $I_j = 4$ related units, a two-sided 95% credibility interval of $(0.97, 1)$.

7.3 Statistical analysis by MCMC

The model is estimated in analogy to Held et al. (2006). To design efficient update steps, suitable auxiliary variables are introduced analog to Chapter 4.

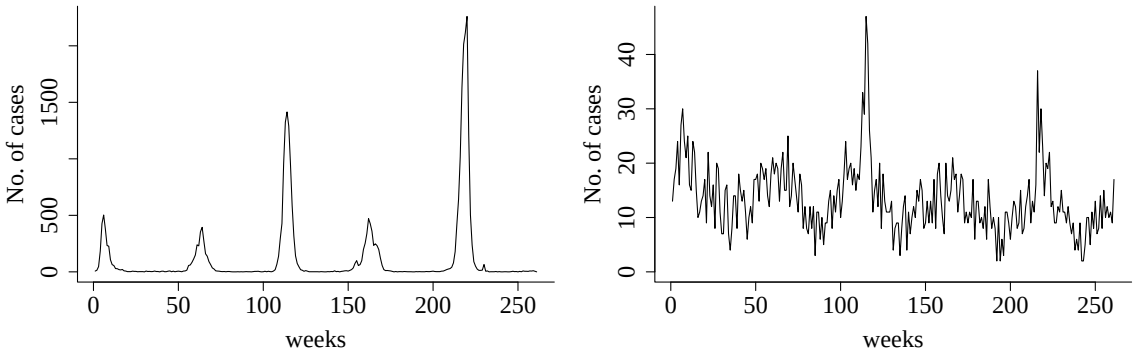
The overdispersion parameters, the endemic parameters and the changepoint parameters are updated similar to Held et al. (2006), the latter two conditioned on the auxiliary variables. The parameters $\boldsymbol{\pi}_j$ of model 6 and 7 and the auxiliary variables are updated in analogy to Chapter 4.

7.4 The influence of influenza on meningococcal diseases

Meningococcal diseases are of major interest for public health surveillance due to their severity and epidemic nature. Meningococcal infections are caused by the Neisseria meningitidis bacterium. Neisseria meningitidis is transmitted to other persons as airborne infection, e.g. by couching or sneezing. Most infected do not get a disease. These persons build immunity by developing protective antibodies and become a healthy carrier (Knorr-Held and Richardson, 2003). Screening studies showed a colonisation of the mucous membranes in the rhinopharynx of more than 30%, depending on the age group, in healthy persons (Claus et al., 2005). Under certain conditions, e.g. low immunity, damage of the mucous membranes, viral infections or dry air, the bacterium can infiltrate through the mucous membranes and cause a severe and perilous disease, meningococcal meningitis, meningococcal sepsis or the Waterhouse-Friderichsen syndrome (Anonymous, 2005). At a macro level, the incident cases can be viewed as resulting from a sporadic component linked to the carriage, on which is superimposed from time to time a small unexpected increase in the incidence, so called hyperendemic periods (Knorr-Held and Richardson, 2003). However, Knorr-Held and Richardson (2003) base their analysis on data, where there is no person to person transmission recorded, whereas in the data, that are analysed here, clusters of cases, that were caused by person to person transmission could be found (Anonymous, 2005). There is a strong evidence that some of these hyperendemic or epidemic periods can be attributed to the influence of influenza, since they regularly occur at the end of influenza outbreaks. We will analyse the influence of influenza on meningococcal diseases based on the weekly observed number of cases of both disease types in Germany, collected in the German infectious disease surveillance system.

Figure 7.1 shows the weekly observed number of influenza and meningococcal disease cases in Germany from January 2001 until December 2005 that are collected in the German infectious disease surveillance system, administrated by the Robert Koch Institute in Berlin (obtained from SurvStat@RKI, `http://www3.rki.de/SurvStat`). The influenza data show yearly outbreaks of different severity during the winter. At a first look on the data one is tempted to interpret this as seasonality. However, the outbreaks do have different start and end points. Although there is some intersection, this would probably disappear with an increasing number of observed outbreaks. The meningococcal disease data, however, show a strong seasonality and two or three little outbreaks during the win-

ters of 2001, 2003 and 2005. These outbreaks coincide with the three biggest outbreaks of influenza. The influence of influenza on meningococcal diseases has previously studied by



(a) Influenza (b) Meningococcal disease

Figure 7.1: Weekly observed number of cases in Germany from January 2001 until December 2005.

Hubert et al. (1992) and Jensen et al. (2004).

The meningococcal disease data show strong seasonality. We therefore assumed seasonality with one frequency, which should be enough with regard to the form of the seasonality.

The influenza data seem to show some seasonality due to the fact that the outbreaks always happen during the winter. However, the start and the end of the outbreak vary from outbreak to outbreak. We have fitted a model with four frequencies for influenza, with respect to the short period of the outbreaks. The estimated endemic parameter of influenza, which includes the seasonality, is shown in Figure 7.2. The estimated seasonality might, however, be an overfitting effect, i.e. the model is better fitting the seasonality in the data than the real seasonality of influenza. If the number of observed outbreaks increases, we expect the estimated seasonality to disappear. We therefore choose no frequency for influenza for all models. The two different disease types are likely to show a different amount of overdispersion. Type 2 overdispersion has therefore be chosen.

The samples of the number of changepoints showed a higher correlation than the samples of the other parameters. On the other hand the update via reversible jump is extremely fast compared to the update steps of the other parameters. We therefore decided to update only the changepoint parameters every iteration and all other parameters every 10th iteration. We obtained 2,500 samples from the posterior distribution after a burn in of

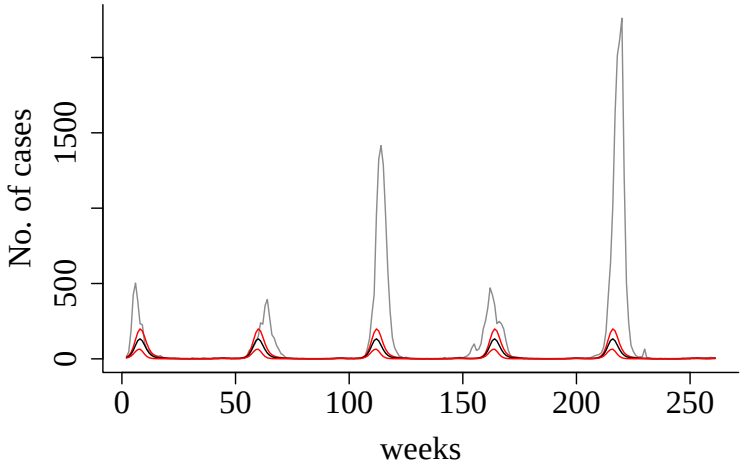


Figure 7.2: Posterior median and pointwise 95% credibility interval of the endemic parameter $\nu_{inf,t}$, $t = 1, \ldots, n$ plotted against the influenza data (light grey).

1,000,000, taking every 10,000 sample. This lead to a total length of 26,000,000 iterations. Despite this extremely hight number of samples the calculation was done in a moderate time of around 4 hours on a 2.80 GHz PC, leading to almost uncorrelated samples. Table 7.2 shows the effective sample size (see Section 2.1.5) of $K_{inf,inf}$ and the calculation time in seconds for the estimation of model 3. We estimate model 4 and model 7 for the

ESS	CPU (s)	ESS/s
2334.15	17007.40	0.14

Table 7.2: Model 3: Effective sample size of $K_{inf,inf}$, estimation time in seconds and effective samples per second.

data, since these assume an individual epidemic rate parameter for the influence within the units. Note that model 5 is equal to model 4 in the case of two units. For both models we assume an influence from influence on meningococcal disease. The parameters for the influence of meningococcal disease on influenza are fixed, that is $\pi_{men,inf} = 0$ for model 7 and $\phi_{inf,t} = 0$, $t = 1, \ldots, n$ for model 4. We compare these two models with model 3, that does not allow for interactions, which is equal to an univariate analysis of the two time series. Table 7.3 shows the mean deviance $\bar{D}$, the estimated number of parameters p_D and the deviance information criterion (DIC) for the three models. Model 4 shows a slightly better fit than model 7. Both models show a better fit than model 3 without interactions,

while there is little difference in the estimated complexity, which is surprising since model 4 has an additional changepoint parameter, that has a relatively complex structure.

	$\bar{D}$	p_D	DIC
model 3	3116	80	3197
model 4	3089	85	3174
model 7	3102	80	3182

Table 7.3: Deviance summaries

There is considerable difference between the posterior distributions of the dispersion parameters of the two time series, shown in Table 7.4, that show disjunct 95% credibility intervals. This confirms our supposition that the two disease types show a different amount of overdispersion. The mean of the parameter for the influence of influenza on

	mean	SD	2.5%	97.5%
ψ_1	12.33	3.43	7.35	20.64
ψ_2	46.33	14.08	25.71	78.43

Table 7.4: Model 3: Posterior mean, standard deviation and 95% credibility interval of the dispersion parameters.

meningococcal disease of model 4, shown in Figure 7.3, has a value around 0.006. The posterior mean of $\pi_{inf,men}$ of model 7 is 0.00724 with a two-sided 95% credibility interval of $(0.00345, 0.01096)$. The mean is slightly below the prior mean 0.01. The 95% credibility interval is smaller compared to the prior, $(0.00026, 0.03658)$, which shows that there is some information about the parameter in the data.

The model also gives us the posterior distribution of the number of meningococcal disease cases caused by influenza, given by the auxiliary variables $Y_{inf,men,t}$, $t = 1, \ldots, 260$, shown in Figure 7.4. There is an increase of the mean number of meningococcal disease cases caused by influenza every year that is largest in the three years with the biggest outbreaks of influenza, explaining a maximum of 22 meningococcal disease cases in the biggest meningococcal disease outbreak in 2003, in the estimation of model 4. The lower 97.5% credibility interval rises above 0 in the two biggest meningococcal disease outbreaks in the winters of 2003 and 2005, reaching a maximum value of 9, for model 4, in the 2003 outbreak.

The number of epidemic meningococcal disease cases cased by meningococcal disease for model 4, Figure 7.5(e), decrease from a mean value of around 3 to 0. This decrease can

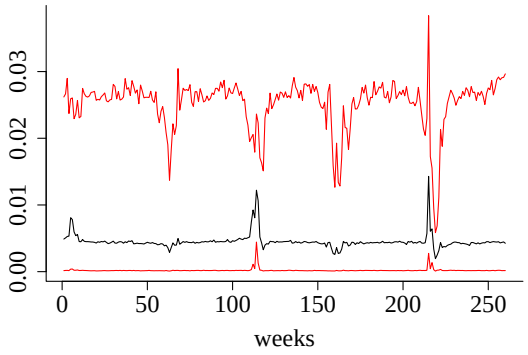
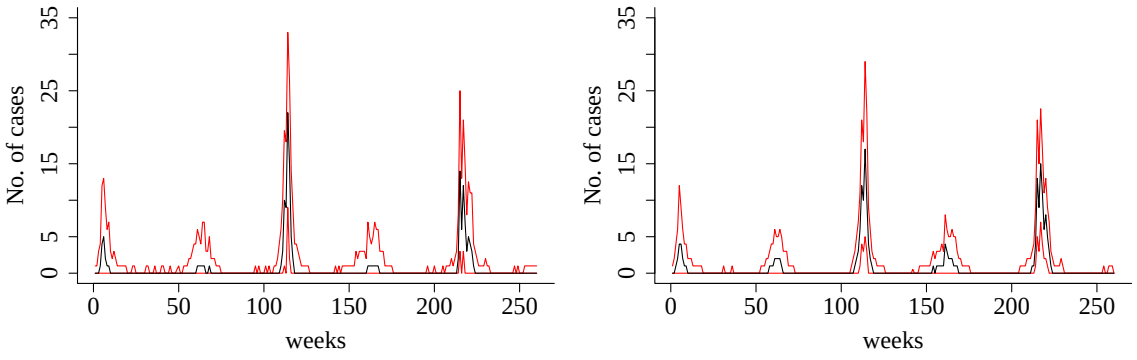


Figure 7.3: Post. median and pointwise 95% credibility interval of $\phi_{men,t}$, $t = 1, \ldots, n$ for model 4.

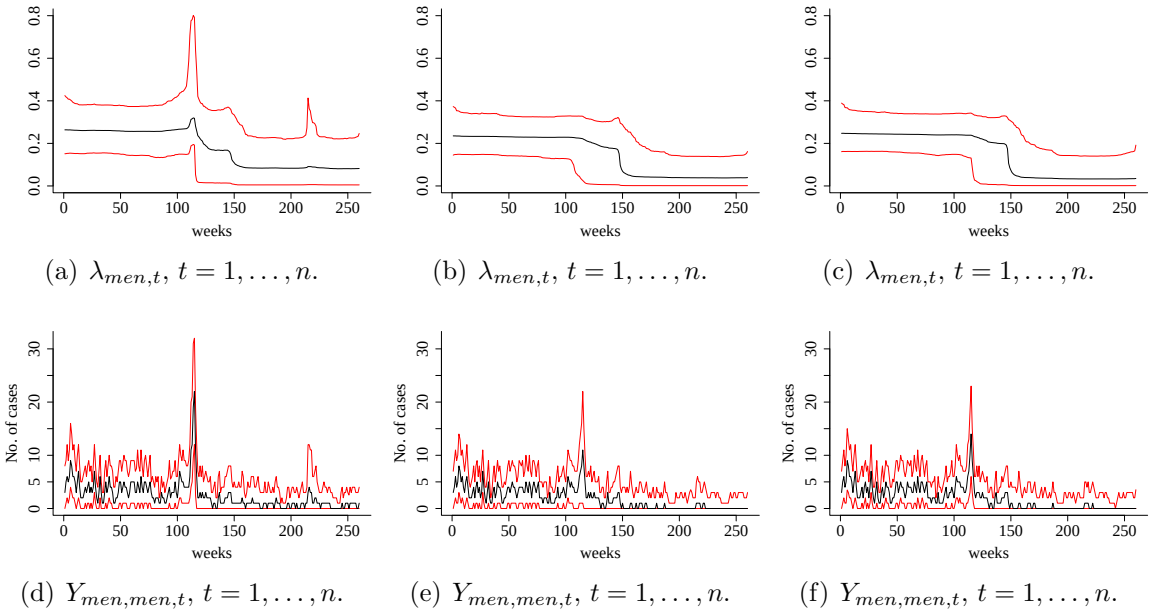


(a) $Y_{inf,men,t}$, $t = 1, \ldots, n$.

(b) $Y_{inf,men,t}$, $t = 1, \ldots, n$.

Figure 7.4: Post. median and pointwise 95% credibility interval of $Y_{inf,men,t}$, $t = 1, \ldots, n$ for model 4 (left) and 7 (right).

also be seen in the posterior of the epidemic parameter of meningococcal disease, Figure 7.5(b). The two peaks in the posterior of the epidemic parameter of meningococcal disease at the times of the outbreaks of 2003 and 2005 in model 3 disappear in the estimation of model 4 and 7. While model 3 explains the 2003 meningococcal disease outbreak by $\boldsymbol{\lambda}_{men}$ and $Y_{men,men}$ (Figure 7.5), model 4 and 7 explain the outbreak primarily by the influence of influenza.



(a) $\lambda_{men,t}$, $t = 1, \ldots, n$. (b) $\lambda_{men,t}$, $t = 1, \ldots, n$. (c) $\lambda_{men,t}$, $t = 1, \ldots, n$.

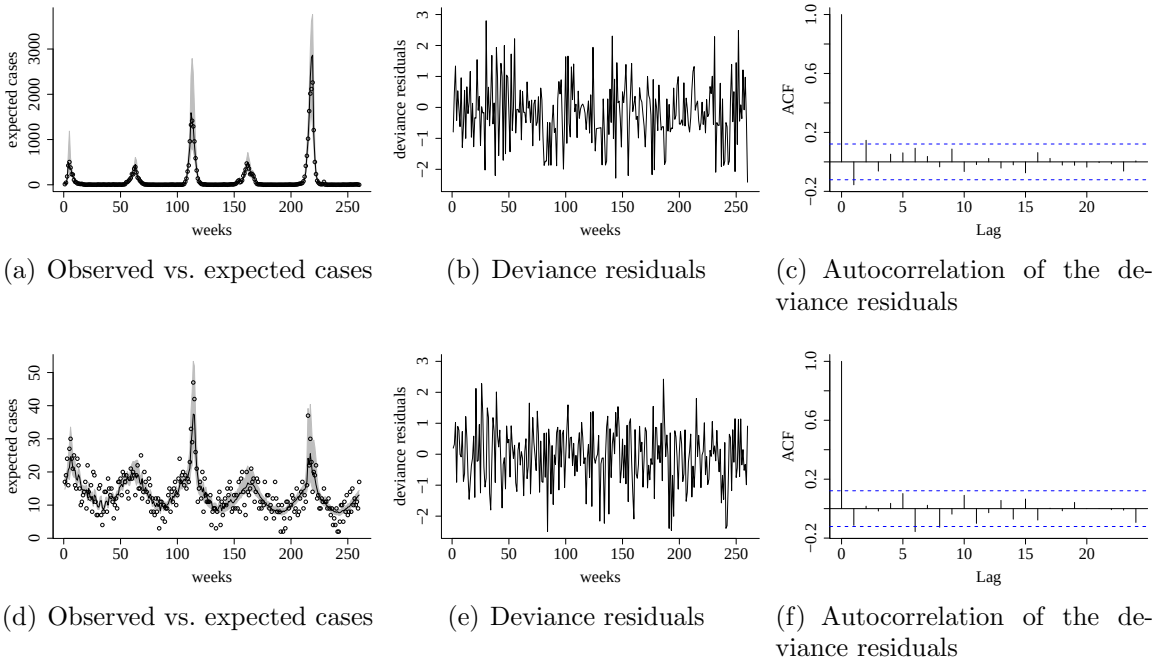
(d) $Y_{men,men,t}$, $t = 1, \ldots, n$. (e) $Y_{men,men,t}$, $t = 1, \ldots, n$. (f) $Y_{men,men,t}$, $t = 1, \ldots, n$.

Figure 7.5: Post. median and pointwise 95% credibility interval of $\lambda_{men,t}$ (top) and $Y_{men,men,t}$ (bottom), $t = 1, \ldots, n$ for model 3 (left), 4 (middle) and 7 (right).

The results of model 7 are similar to those of model 4, with a slightly higher number of epidemic meningococcal disease cases caused by meningococcal disease in the outbreak of 2003 and a slightly lower number of meningococcal disease cases caused by influenza in the 2005 outbreak. This may be a result of the lower flexibility of model 7, that assumes the epidemic rate of the influenza and meningococcal disease cases caused by influenza to be proportional.

Figure 7.6 shows the expected cases and the deviance residuals for model 4. The model gives a good fit to the data and is able to explain the seasonality and the three outbreaks in the meningococcal disease data, as well as the different severity of the influenza outbreaks. The deviance residuals support this, with a slightly better fit for the meningococcal disease

data.



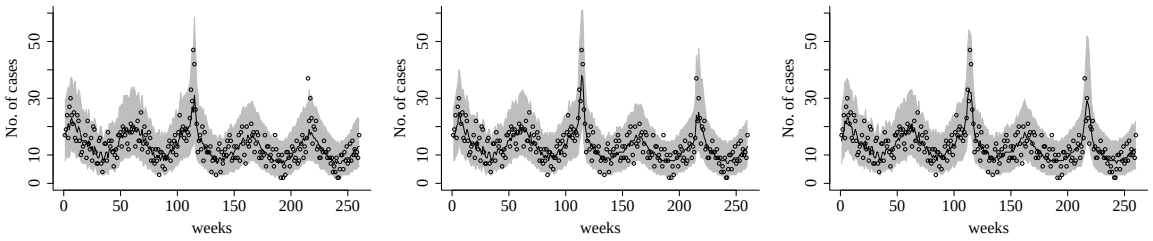
(a) Observed vs. expected cases (b) Deviance residuals (c) Autocorrelation of the deviance residuals

(d) Observed vs. expected cases (e) Deviance residuals (f) Autocorrelation of the deviance residuals

Figure 7.6: Model 4. Posterior median and pointwise 95% credibility interval of the data mean $\mu_{i,t}$, $t = 1, \ldots, n$ plotted against the data (left), deviance residuals (middle), and autocorrelation of the deviance residuals (right), for influenza (top) and meningococcal disease (bottom).

Figure 7.7 show the predicted number of cases for model 3, 4, and 7, where the prediction is based on all observations. Model 4 and 7 provide a clearly better prediction of the three meningococcal disease outbreaks, especially at the beginning. Model 3 predicts the outbreaks with delay. The short outbreak in 2005 is therefore not predicted.

We have estimated the three models for the data up to time 109, 110, 111 and 112, which is at the beginning of the 2003 meningococcal disease outbreak, to compare the prediction of this outbreak for the three models. The predictive distributions for the three models, now based on the past observations, are shown in Figure 7.8. For model 4 and 7, the predictive distribution becomes a longer tail towards larger values. The increase in the predicted number of cases shows that there is an increased risk for a meningococcal disease outbreak already before the real outbreak is observed. The observed values lie clearly inside the 95% credibility interval. In model 3 there is no increase in the predictive

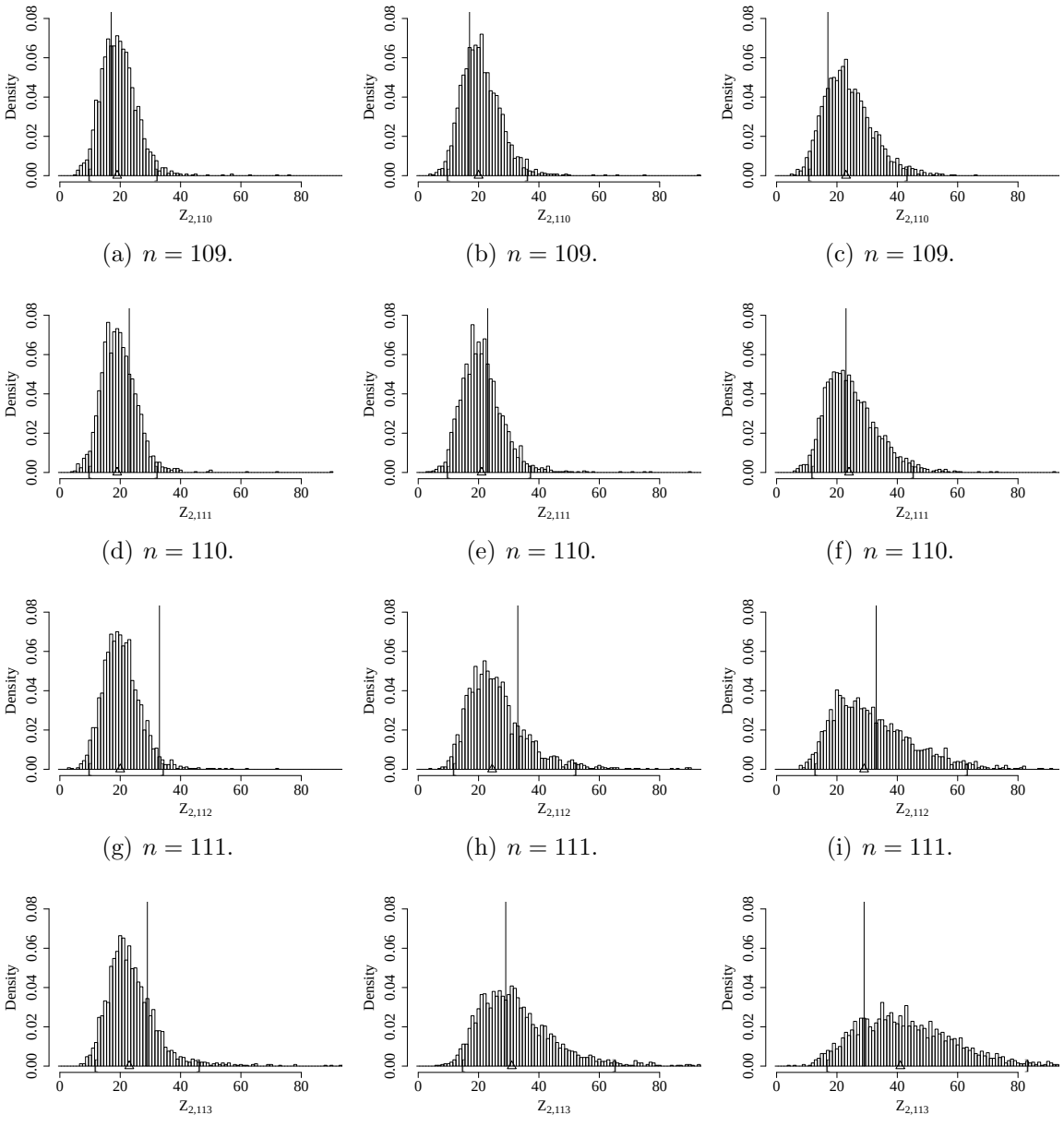


(a) model 3 (b) model 4 (c) model 7

Figure 7.7: Posterior median and pointwise 95% credibility interval of the predicted number of cases plotted against the observed number of cases for meningococcal disease.

distribution based on the data up to time 109, 110 and 111. The observed number of cases therefore trend towards the upper limit of the predictive distribution, where for time 112 the observed value lies at the upper bound of the 95% credibility interval. At time 113 the predictive distribution shows a slight increase, due to the high number of cases observed at time 112, with the observed value inside the 95% credibility interval.

The additional influence of influenza on meningococcal disease in model 4 and 7 allows for a better prediction of the meningococcal disease outbreaks. The two models also suggest that the risk of a meningococcal disease outbreak is present for a longer time period, even if there is no outbreak observed at a certain time point.



(a) $n = 109$. (b) $n = 109$. (c) $n = 109$.

(d) $n = 110$. (e) $n = 110$. (f) $n = 110$.

(g) $n = 111$. (h) $n = 111$. (i) $n = 111$.

(j) $n = 112$. (k) $n = 112$. (l) $n = 112$.

Figure 7.8: Predictive distribution ($\triangle$ = median, [ ] = 95% credibility interval and observed value (vertical line)) for meningococcal disease based on the past observations $Z_{men,n+1}|Z_{men,1}, \ldots, Z_{men,n}$, $n = 109, \ldots, 112$ for the model 3 (left), model 4 (middle) and model 7 (right).

7.5 Influenza in the federal states of Germany

We now analyse the influenza data observed in the 16 federal states of Germany from January 2001 until December 2005 that are again collected in the German infectious disease surveillance system, administrated by the Robert Koch Institute in Berlin (obtained from SurvStat@RKI, `http://www3.rki.de/SurvStat`). Model 3, 4, 5 and 7 are too complex with respect to the number of regions. We therefore consider model 1, 2 and 6, where model 2 and 6 allow for interactions between the regions. Two states are defined to be neighbours, i.e related $i \sim j$, if they share a common border. The numbering of the states that we used is shown in Table 7.8. We obtained 1,000 samples from the posterior distribution after a burn in of 1,000,000, taking every 10,000 sample. This lead to a total length of 11,000,000 iterations. The calculation took around 20 hours. Table 7.5 shows the effective sample size (see 2.1.5) of K and the calculation time in seconds for the estimation of model 6. Table

ESS	CPU (s)	ESS/s
1002.01	80038.40	0.01

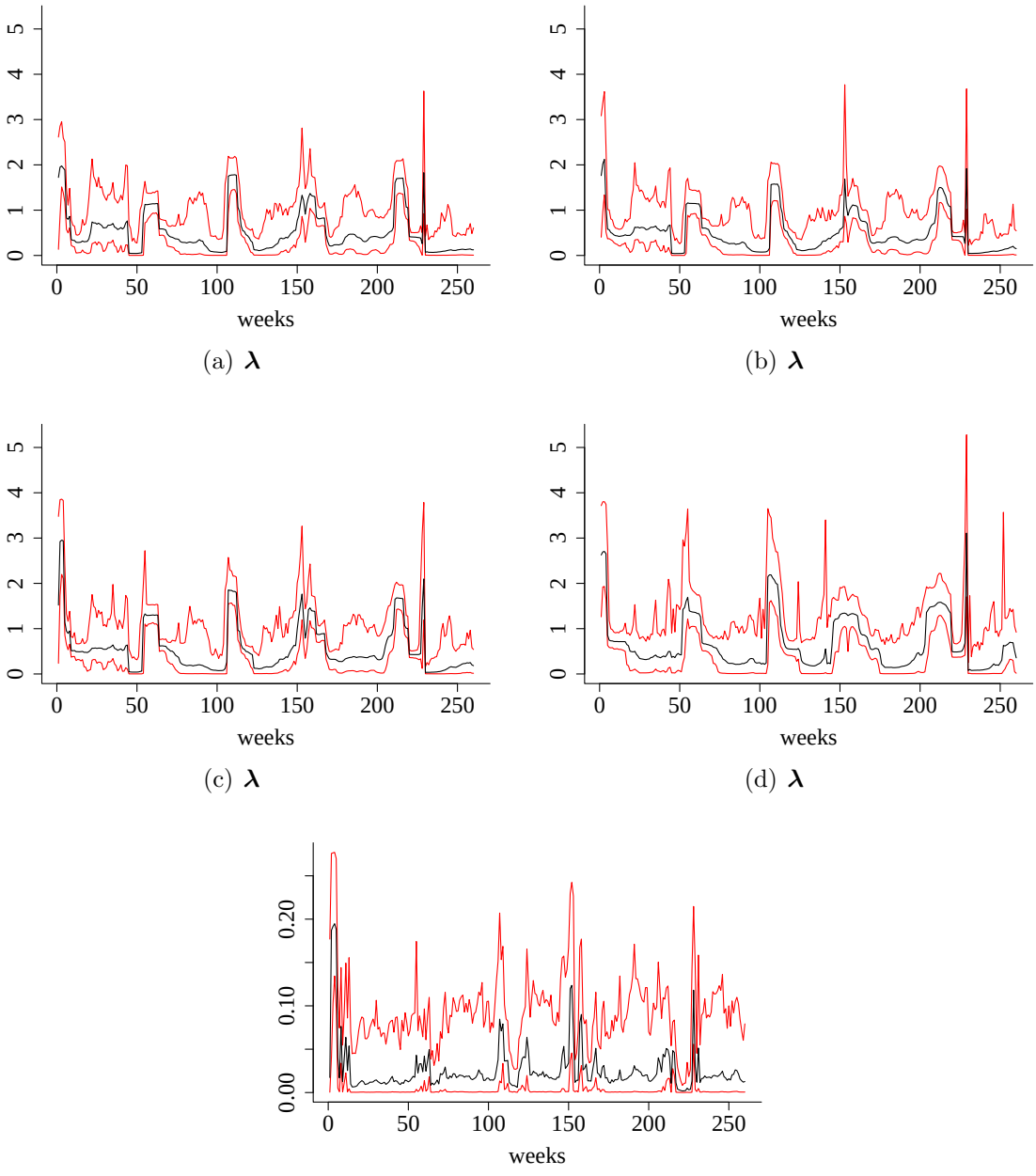
Table 7.5: Model 5: Effective sample size of K, estimation time in seconds and effective samples per second.

7.6 shows the mean deviance $\bar{D}$, the estimated number of parameters p_D and the deviance information criterion (DIC) for the three models. Model 6 is the best in terms of DIC, however, both models with interactions, model 2 and 6, are better than model 1 without interactions. The estimation of the parameter $\boldsymbol{\lambda}$, shown in Figure 7.9, is now based on a

	$\bar{D}$	p_D	DIC
model 1	10668	83	10751
model 2	10355	137	10492
model 6	10380	98	10478

Table 7.6: Deviance summaries

higher amount of data compared to the univariate case (Figure 7.9(d)). This additional information leads to a better estimate of the parameter, with smaller credibility intervals especially during the outbreak periods. Table 7.7 shows the results for the interaction parameters π_{ji} of model 6. Although most states cause the highest proportion of cases within the own state, there is considerable interaction between neighbours. Figure 7.10 and 7.11 show the space-time spread of influenza in the outbreak of 2002 estimated by



(a) $\boldsymbol{\lambda}$ (b) $\boldsymbol{\lambda}$

(c) $\boldsymbol{\lambda}$ (d) $\boldsymbol{\lambda}$

(e) $\boldsymbol{\phi}$

Figure 7.9: Posterior median and pointwise 95% credibility interval of the parameter $\boldsymbol{\lambda}$ of model 1 (a), 2 (b), 6 (c) and in the univariate model (d) and $\boldsymbol{\phi}$ of model 2 (e).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.75	0.12	–	–	–	–	0	–	–	–	0.12	–	–	–	–	–
2	0	0.81	–	–	–	–	0.01	–	–	–	–	–	0.11	–	–	0.06
3	–	–	0.89	0.11	–	–	–	–	–	–	–	–	–	–	–	–
4	–	–	0.25	0.61	–	–	–	0	0	–	–	–	0	0.13	–	–
5	–	–	–	–	0.99	–	–	–	0.01	–	–	–	–	–	–	–
6	–	–	–	–	–	0.96	–	–	0.02	–	–	–	–	–	0.01	–
7	0	0.16	–	–	–	–	0.77	–	0	0.05	0	–	–	–	–	0
8	–	–	–	0.11	–	–	–	0.85	0	–	–	–	–	–	0.03	–
9	–	–	–	0.02	0	0.01	0	0.02	0.9	0	–	–	–	0.02	0.03	0
10	–	–	–	–	–	–	0.07	–	0.12	0.78	0.02	–	–	–	–	–
11	0.18	–	–	–	–	–	0	–	–	0.08	0.72	0.01	–	–	–	–
12	–	–	–	–	–	–	–	–	–	–	0.02	0.98	–	–	–	–
13	–	0	–	0	–	–	–	–	–	–	–	–	0.9	0.1	–	0
14	–	–	–	0	–	–	–	–	0	–	–	–	0.22	0.78	–	0
15	–	–	–	–	–	0	–	0.08	0.05	–	–	–	–	–	0.87	–
16	–	0	–	–	–	–	0	–	0	–	–	–	0.23	0	–	0.75

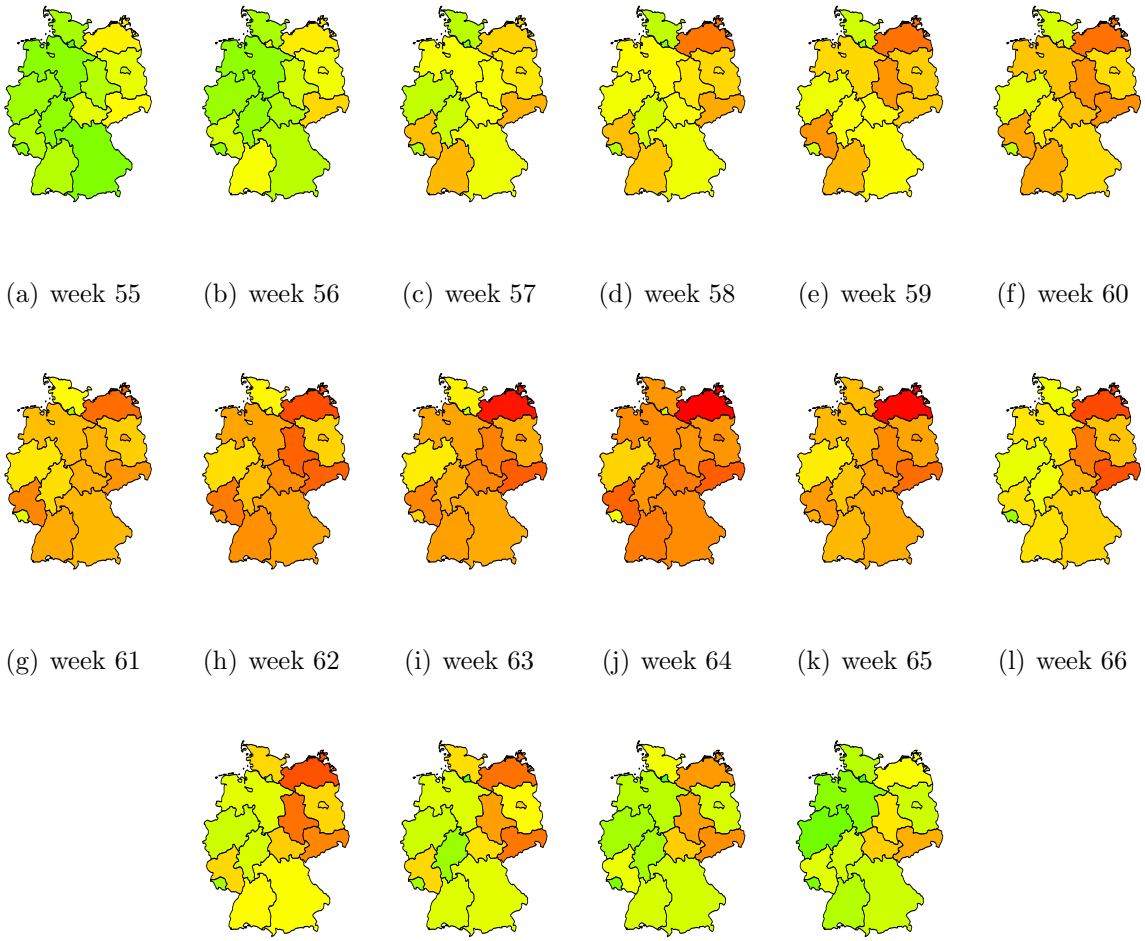
Table 7.7: Matrix of the posterior median of π_{ji}.

	Federal states of Germany
1	Baden-Württemberg
2	Bayern
3	Berlin
4	Brandenburg
5	Bremen
6	Hamburg
7	Hessen
8	Mecklenburg-Vorpommern
9	Niedersachsen
10	Nordrhein-Westfalen
11	Rheinland-Pfalz
12	Saarland
13	Sachsen
14	Sachsen-Anhalt
15	Schleswig-Holstein
16	Thüringen

Table 7.8: Federal states of Germany.

model 6. Since the growth of the disease is exponential we show the logarithm of the mean of μ_{it}, which is standardized by the population size. The scale goes from green(low number of cases) to red(high number of cases). Besides this, the velocity of growth, calculated in analogy to Mugglin et al. (2002) as $\log(\mu_{i,t+1}) - \log(\mu_{i,t})$, is shown, where red means a positive growth and green a negative. If the disease is neither growing nor shrinking, the color is yellow. The disease starts in the north east and south west of Germany. In the first weeks of the outbreak, the number of cases are still moderate, however, the disease spreads with considerable velocity. While the velocity is generally high in the first weeks of the outbreak, there is considerable difference in the states. The peak of the disease is reached in most states in week 64. This can be seen by the velocity of growth, that changes from positive in week 64 to negative in week 65 in most states. The states where the disease started, reach the highest number of cases compared to the other states. While in most states the number of cases shrinks after week 64, some states of the north east and south

west show another growth. This local outbreak spreads further, reaching again the center of Germany in week 69.

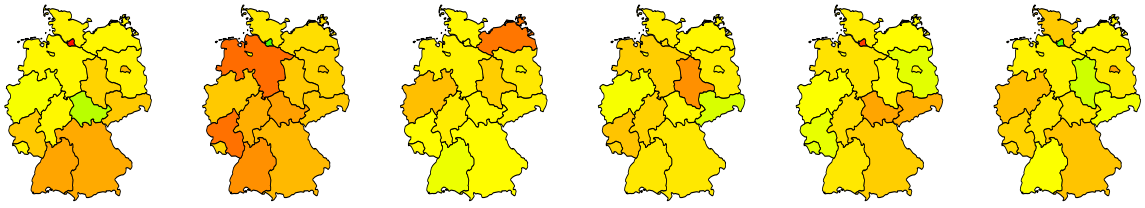


(a) week 55 (b) week 56 (c) week 57 (d) week 58 (e) week 59 (f) week 60

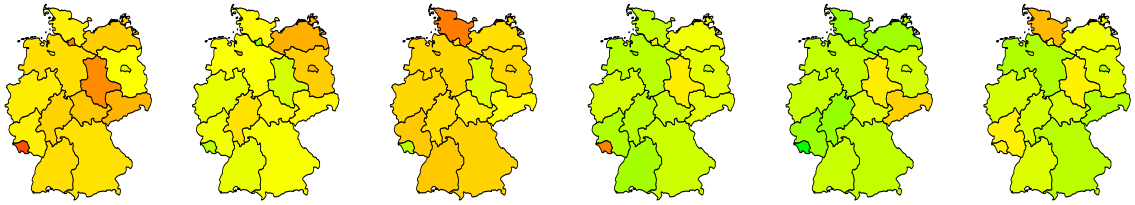
(g) week 61 (h) week 62 (i) week 63 (j) week 64 (k) week 65 (l) week 66

(m) week 67 (n) week 68 (o) week 69 (p) week 70

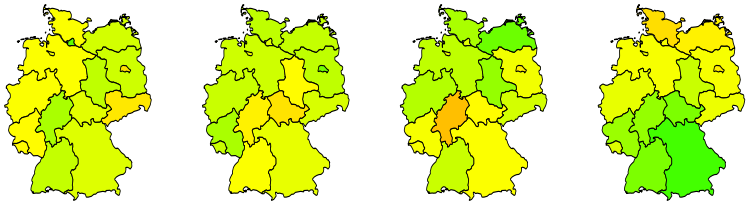
Figure 7.10: Model 6: Expected number of cases.



(a) week 55 (b) week 56 (c) week 57 (d) week 58 (e) week 59 (f) week 60



(g) week 61 (h) week 62 (i) week 63 (j) week 64 (k) week 65 (l) week 66



(m) week 67 (n) week 68 (o) week 69 (p) week 70

Figure 7.11: Model 6: Velocity of growth.

7.6 Model comparison

We will compare model 6 with the model based on a GMRF including unstructured spatial effects, structured time effects and seasonal covariates, the model with time constant epidemic parameter and the model with an endemic component based on a GMRF and a time constant epidemic parameter proposed in Chapter 4. Table 7.9 and 7.10 show the deviance summaries of the three models for the influenza data and the measles data analysed in Chapter 4 and model 6 of this chapter. The models with epidemic component are all

	$\bar{D}$	p_D	DIC
model with random walk	10953	84	11037
model 6 with time const. epid. parameters	10975	27	11002
model 6 with time const. epid. parameters and random walk	10412	79	10491
model 6 with time varying epid. parameters	10380	98	10478

Table 7.9: Deviance summaries for influenza

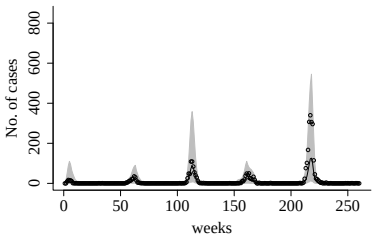
	$\bar{D}$	p_D	DIC
model with random walk	1742	15	1757
model 6 with time const. epid. parameters	1565	12	1577
model 6 with time const. epid. parameters and random walk	1562	13	1575
model 6 with time varying epid. parameters	1539	29	1568

Table 7.10: Deviance summaries for measles

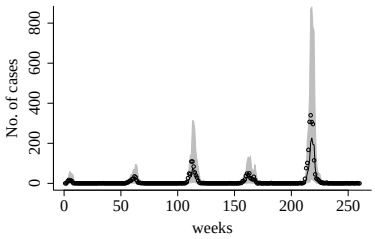
better in terms of DIC than the model purely based on a GMRF. However, the model with time varying epidemic parameter is the best model, even if the model with time constant epidemic parameter includes an additional structured time effect in the endemic component. Figure 7.12 show the predicted influenza cases of region 9 for the four models, where prediction is based on all data. The model with time varying epidemic parameter shows clearly the best prediction of the four models. As the estimate of the epidemic parameter of influenza (Figure 7.9 (b)), the estimate of the epidemic parameter of measles (Figure 7.13) is clearly not constant.

7.7 Discussion

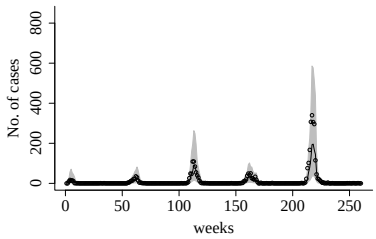
In this chapter we have introduced a multivariate version of the model proposed in Chapter 5. The basic concept is again the distinction of an endemic and an epidemic component.



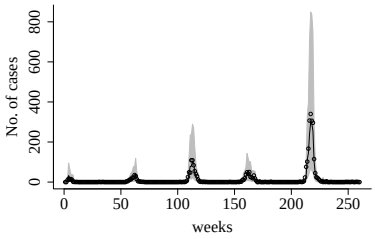
(a) model with random walk



(b) model 6 with time const. epid. parameters



(c) model 6 with time const. epid. parameters and random walk



(d) model 6 with time varying epid. parameters

Figure 7.12: Posterior median and pointwise 95% credibility interval of the predicted number of cases plotted against the observed number of influenza cases of region 9.

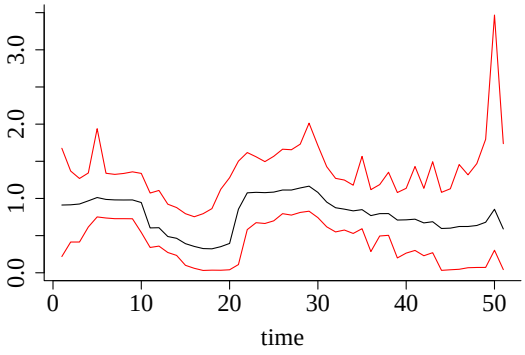


Figure 7.13: Posterior median and pointwise 95% credibility interval of the epidemic parameter $\boldsymbol{\lambda}$ for measles of model 6.

The endemic component is extended by independent unit effects. The seasonality and the amount of overdispersion may be assumed either dependent or independent of the unit. The epidemic component includes now dependencies between the units to model the interactions between the units.

The model has shown to be applicable to a wide range of problems. The application to influenza and meningococcal disease data shows that the occasional outbreaks of meningococcal disease can largely be explained and predicted by the influence of influenza on meningococcal disease. The risk of a future meningococcal disease outbreak caused by influenza can be predicted.

The application to the spatial influenza and measles data shows that the model fits the data quite well and gives an idea of the extent of interaction between different regions. The comparison with a model based on a GMRF and the model proposed in Chapter 4 with a constant epidemic parameter shows that the inclusion of the epidemic component as well as a time varying epidemic parameter improves the fit and the prediction of the model.

It would be interesting to compare the model with a more sophisticated GMRF model, including space-time interactions.

A possible alternative to the current approaches would be to model the epidemic parameter by a Bayesian cluster model e.g. the one proposed by Yan and Clayton (2006).

Chapter 8

Conclusion

Models for infectious disease surveillance counts have to take into account the specific characteristics of this type of data. While showing a regular, often seasonal, pattern over long time periods, there are occasional irregularities or outbreaks. The wide range of mechanistic models is not applicable to this kind of data. Standard empirical models as log-linear Poisson regression models or Gaussian Markov random fields (GMRF) models, as used for chronic diseases, on the other hand, are not able to capture the infectious disease specific characteristics of the data. A compromise between mechanistic models and empirical model is needed. We therefore based the models on a branching process, which is an approximation to the mechanistic chain binomial model. A further key idea is to distinguish between an endemic and an epidemic component, which allows us to separate the regular pattern from the irregularities and outbreaks. This is of particular advantage for outbreak detection in public health surveillance. While the endemic component is modeled using standard empirical models, including spatial, time and seasonal components, which can be seen as parameter-driven models, the epidemic component is based on an observation-driven approach, including an autoregression on past observation, which is more appropriate for this setting.

A particular challenge of infectious disease counts is the modelling of the outbreaks and irregularities in the data. These often show jumps or fast increases at the beginning on the one hand and smooth decreases on the other hand. Models assuming constant correlation structures as GMRF are not appropriate for this task. We therefore modeled the autoregressive parameter of the epidemic component by a Bayesian changepoint model, which shows an adaptive amount of smoothing, and is able to model the jumps and fast increases as well as the smooth decreases in the data. While the model can be used as

a generic approach for infectious disease surveillance counts, it is particularly suited for outbreak detection in public health surveillance. Besides, the predictive qualities of the Bayesian changepoint model allow for short term predictions of the number of disease cases, which are of particular public health interest.

The Markov state space form of the changepoint model and the Markov structure of the changepoints allow for the use of sequential Monte Carlo (SMC) methods. We considered two types of sequential Monte Carlo methods: the forward-backward algorithm and a particle filter. While the forward-backward algorithm can be used as an alternative to the reversible jump algorithm, for the update of the changepoint model within the MCMC algorithm, the particle filter can be used for a prospective analysis of the changepoint model conditioning on fixed values for the other parameters, which is of particular advantage for public health surveillance, where data arise sequentially.

Infectious disease surveillance data are often available as longitudinal data, observed in units, e.g. age groups or spatial regions. These data provide more information than time series data. To make the model applicable to these kind of data, multivariate versions of the model were considered. Special attention has been paid to the modelling of the dependencies and interactions between the units. These models have shown to be applicable to a wide range of problems. The application to influenza and meningococcal disease data showed that the occasional outbreaks of meningococcal disease can largely be explained by the influence of influenza on meningococcal disease. The risk of a future meningococcal disease outbreak caused by influenza can be predicted. The application to spatial influenza and measles data showed, that the model fits the data quiet well and gives an idea of the extent of interaction between different regions. The comparison of the different models, including a model based on GMRF, showed that the inclusion of the epidemic component as well as a time varying epidemic parameter improves the fit and the predictive qualities of the model.

Appendix A

Twins

The software *twins* was developed within the scope of this work. *Twins* is a software to estimate the model described in Chapter 5 and is available at

`http://www.stat.uni-muenchen.de/~mhofmann`.

On the following pages you find the *twins*-manual.

twins

A two-component model for counts of infectious diseases

Leonhard Held, Mathias Hofmann, Michael Höhle, Volker Schmid
Department of Statistics, Ludwig-Maximilians-Universität München,
Ludwigstr. 33, 80539 München, Germany

Version 1.0

twins is available at `http://www.stat.uni-muenchen.de/~mhofmann`

Licence agreement

A.1 General information

Twins is a software for estimating the stochastic model for time series count data of infectious diseases proposed in Held et al. (2006). The model is based on a Poisson or negative binomial observation model with two components: A parameter-driven component that relates the disease incidence to latent parameters describing endemic seasonal patterns, which are typical for infectious disease surveillance data, and an observation-driven or epidemic component that explains for possible outbreaks.

A.2 Starting *twins*

After you have installed twins, edit the twins.ini file or write your own ini-file. Then start the program by typing

```
twins ini-file
```

or (if you use twins.ini) just type

```
twins
```

in the command line. On window-based OS you can also start twins with a double click on the twins-icon. Twins then will use the `twins.ini` file.

A.3 Input files

A.3.1 The data

The data file must containing the number of observations followed by the observations that have to be integers. The entries have to be separated by the new line command.

A.3.2 The parameters

All parameters for the algorithm are specified in an ini-file, by default called twins.ini, which contains the following information

datafile	path of the data file
logfile	path of the first output file; the estimation results are written to the two output files as explained in Section A.4.
logfile2	path of the second output file
burnin	burnin; total number of iterations = burnin + filter*sampleSize.
filter	filter
sampleSize	sample size
seed	seed
alpha_xi	first parameter of the prior of ξ
beta_xi	second parameter of the prior of ξ
season	season of the endemic component
frequencies	number of frequencies of the endemic component
psiRWSigma	starting value for the tuned standard deviation of the proposal of ψ
alpha_psi	first parameter of the prior of ψ
beta_psi	second parameter of the prior of ψ

The ini-file must have 14 lines. The two columns have to be separated by a colon.

A.4 Output files

The output is written to three output files:

- The first output file, contains the samples of the posterior distributions of ψ, γ, K, ξ, λ, Z_{n+1} and the Deviance.
- The second output file contains the posterior means of $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{\omega}$ and the posterior probabilities of the changepoints.
- The acc file, "output file name".acc, contains the acceptance rates.

For a better handling with R, the time index starts at $t = 1$ instead of $t = 0$.

A.5 Figures

The R-program figures.R reads the output files of the estimation and creates some figures.

xyz.pdf	Posterior means of the components $\boldsymbol{X}$ and $\boldsymbol{Y}$
tms-lambda.pdf	Posterior mean and pointwise 95% credibility interval of λ
tms-nu.pdf	Posterior mean and pointwise 95% credibility interval of ν
theta.pdf	Posterior probabilities of the changepoints
lambdage1.pdf	Posterior probability of $\lambda > 1$
histogram-K.pdf	Posterior probability of K
histogram-psi.pdf	Posterior probability of ψ
histogram-Znp1.pdf	Posterior predictive probability of Z_{n+1}
traj-gamma-i.pdf	Trajectory of γ_i
traj-K.pdf	Trajectory of K
traj-psi.pdf	Trajectory of ψ
traj-xi.pdf	Trajectory of ξ
autocorrelation.pdf	Autocorrelation of K and ψ

Bibliography

Anderson, H. and T. Britton (2000). *Stochastic Epidemic Models and their Statistical Analysis*, Volume 151 of *Lecture Notes in Statistics*. New York: Springer.

Andrieu, C., A. Doucet, and E. Punskaya (2001). Sequential Monte Carlo methods for optimal filtering. In *Sequential Monte Carlo Methods in Practice*. Springer.

Anonymous (2004, 8. October). Zu einer Häufung reiseassoziierter Hepatitis A unter Ägypten-Urlaubern. Epidemiologisches Bulletin 41, Robert Koch Institute, Berlin.

Anonymous (2005, 26. August). Invasive Meningokokken-Erkrankungen im Jahr 2004. Epidemiologisches Bulletin 34, Robert Koch Institute, Berlin.

Assunção, R. M. (2003). Space varying coefficient models for small area data. *Environmetrics 14*, 453–473.

Barry, D. and J. A. Hartigan (1992). Product partition models for change point problems. *The Annals of Statistics 20*(1), 260–279.

Barry, D. and J. A. Hartigan (1993). A Bayesian analysis for change point problems. *Journal of the American Statistical Association 88*(421), 309–319.

Becker, N. G. (1989). *Analysis of infectious disease data*. London: Chapman & Hall.

Bernardinelli, L., D. Clayton, C. Pascutto, C. Montomoli, M. Ghislandi, and M. Songini (1995). Bayesian analysis of space-time variation in disease risk. *Statistics in Medicine 14*, 2433–2443.

Bernardo, J. and A. Smith (1994). *Bayesian Theory*. Chichester: Wiley.

Berzuini, C., D. Clayton, and L. Bernardinelli (1993). Bayesian inference on the Lexis diagram. *Bulletin of the International Statistical Institute 50*, 149–164.

Berzuini, C. and W. R. Gilks (2003). Particle filtering methods for dynamic and static Bayesian problems. In P. Green, N. L. Hjort, and S. Richardson (Eds.), *Highly Structured Stochastic Systems*, pp. 207–227. Oxford: Oxford University Press.

Besag, J., P. Green, D. Higdon, and K. Mengersen (1995). Bayesian computation and stochastic systems. *Statistical Science 10*(1), 3–41.

Besag, J., J. York, and A. Mollié (1991). Bayesian image restoration with applications in spatial statistics (with discussion). *Annals of the Institute of Mathematical Statistics 43*, 1–59.

Brix, A. and P. J. Diggle (2001). Spatiotemporal prediction for log-Gaussian Cox processes. *Journal of the Royal Statistical Society, Series B 63*, 823–841.

Carpenter, J., P. Clifford, and P. Fearnhead (1999). Improved particle filter for nonlinear problems. *IEE Proceedings-Radar, Sonar Navig. 146*(1).

Claus, H., M. C. J. Maiden, D. J. Wilson, N. D. McCarthy, K. A. Jolley, R. Urwin, F. Hessler, M. Frosch, and U. Vogel (2005). Genetic analysis of meningococci carried by children and young adults. *The Journal of Infectious Diseases 191*, 1263–1271.

Clayton, D. and J. Kaldor (1987). Empirical Bayes estimates of age-standardized relative risk for use in disease mapping. *Biometrics 43*(3), 671–681.

Clayton, D. G., L. Bernardinelli, and C. Montomoli (1993). Spatial correlation in ecological analysis. *International Journal of Epidemiology 22*(6), 1193–1202.

Clyde, M. (1999). Bayesian model averaging and model search strategies (with discussion). In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 6*. Oxford: Clarendon Press.

Cowell, R. G., A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter (1999). *Statistics for Engineering and Information Science*. New York: Springer.

Crisan, D. and A. Doucet (2002). A survey of convergence results on particle filtering methods for practitioners. *IEEE Transactions of signal processing 50*(3), 736–746.

Daley, D. J. and J. Gani (1999). *Epidemic Modelling; An Introduction*. Cambridge: Cambridge University Press.

de Finetti, B. (1970/1974). *Teoria delle Probabilità. Reprinted as 'Theory of Probability'*, Volume 1. Chichester: Wiley.

Denison, D., C. Holmes, B. Mallick, and A. Smith (2002). *Bayesian Methods for Nonlinear Classification and Regression*. Chichester: Wiley.

Diekmann, O. and A. P. Heesterbeek (2000). *Mathematical Epidemiology of Infectious Diseases*. Chichester: Wiley.

Diggle, P., L. Knorr-Held, B. Rowlingson, T.-L. Su, P. Hawtin, and T. Bryant (2003). Online monitoring of public health surveillance data. In R. Brookmeyer and D. Stroup (Eds.), *Monitoring the Health of Populations: Statistical Principles and Methods for Public Health Surveillance*, pp. 233–266. Oxford: Oxford University Press.

Diggle, P., B. Rowlingson, and T. Su (2005). Point process methodology for on-line spatio-temporal disease surveillance. *Environmetrics 16*, 423–434.

Diggle, P. J. (1990). *Time Series. A Biostatistical Introduction*. Oxford: Oxford University Press.

Doucet, A., N. de Freitas, and N. Gordon (2001a). An introduction to sequential Monte Carlo methods. In *Sequential Monte Carlo Methods in Practice*. Springer.

Doucet, A., N. de Freitas, and N. Gordon (2001b). *Sequential Monte Carlo Methods in Practice*. Springer.

Fahrmeir, L. and L. Knorr-Held (2000). Dynamic and semiparametric models. In M. Schimek (Ed.), *Smoothing and Regression: Approaches, Computation and Application*. New York: John Wiley & Sons.

Farrington, C. P., N. Andrews, A. D. Beale, and M. A. Catchpole (1996). A statistical algorithm for the early detection of outbreaks of infectious disease. *Journal of the Royal Statistical Society, Series A 159*, 547–563.

Fearnhead, P. (2006). Exact and efficient Bayesian inference for multiple changepoint problems. *Statistics and Computing 16*(2), 203–213.

Ferreira, J., D. Denison, and C. Holmes (2002). Partition modelling. In *Spatial Cluster Modelling*, Chapter 7. Chapman & Hall/CRC.

Finkenstädt, B., O. N. Bjørnstad, and B. T. Grenfell (2002). A stochastic model for extinction and recurrence of epidemics: estimation and inference for measles outbreaks. *Biostatistics 3*(4), 493–510.

Finkenstädt, B. and B. T. Grenfell (2000). Time series modelling of childhood diseases: a dynamical systems approach. *Appl. Statist. 49*(2), 1–19.

Frühwirth-Schnatter, S. (1996). Recursive residuals and model diagnostics for normal and non-normal state space models. *Environmetrical and Ecological Statistics 3*, 291–309.

Gelman, A., G. O. Roberts, and W. R. Gilks (1996). Efficient Metropolis jumping rules. In J. Bernardo, J. Berger, A. Dawid, and A. Smith (Eds.), *Bayesian Statistics 5*, pp. 599–607. Oxford: Oxford University Press.

Geyer, C. J. (1992). Practical Markov chain Monte Carlo. *Statistical Science 7*(4), 473–483.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter (1996). *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall.

Gordon, N. J., D. J. Salmond, and A. F. M. Smith (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F 140*(2), 107–113.

Green, P. J. (1995). Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika 82*, 711–732.

Guttorp, P. (1995). *Stochastic Modelling of Scientific Data*. London: Chapman and Hall.

Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika 57*(1), 97–109.

Held, L. (2004). Simultaneous posterior probability statements from Monte Carlo output. *Journal of Computational and Graphical Statistics 13*(1), 20–35.

Held, L. (2006). Quantitative assessment of probabilistic forecasts with applications in epidemiology. In J. Hinde, J. Einbeck, and J. Newell (Eds.), *Proceedings of the 21st International Workshop on Statistical Modelling*, pp. 38–49.

Held, L., M. Hofmann, M. Höhle, and V. Schmid (2006). A two-component model for counts of infectious diseases. *Biostatistics 7*, 422–437.

Held, L., M. Höhle, and M. Hofmann (2005). A statistical framework for the analysis of multivariate infectious disease surveillance counts. *Statistical Modelling 5*, 187–199.

Höhle, M. and A. Riebler (2005). The `R`-Package 'surveillance'. SFB 386 Discussion Paper 422, Ludwig-Maximilians-University, Department of Statistics, Munich. Software available at `http://www.statistik.lmu.de/~hoehle/software/surveillance`.

Holmes, C. C. and B. K. Mallick (2000). Bayesian wavelet networks for nonparametric regression. *IEEE Transactions on neural networks 11*(1), 27–35.

Hubert, B., L. Watier, and P. G. ans S. Richardson (1992). Meningococcal disease and influenza like syndrome: a new approach to an old question. *The Journal of Infectious Diseases 166*, 542–545.

Jensen, E. S., S. Lundbye-Christensen, S. Samuelsson, H. T. Sørensen, and H. C. Schønheyder (2004). A 20-year ecological study of the temporal association between influenza and meningococcal disease. *European Journal of Epidemiology 19*, 181–187.

Jørgensen, B., S. Lundbye-Christensen, P. X.-K. Song, and L. Sun (1999). A state space model for multivariate longitudinal count data. *Biometrica 86*(1), 169–181.

Kass, R. E., B. P. Carlin, A. Gelman, and R. M. Neal (1998). Markov chain Monte Carlo in practice: A roundtable discussion. *The American Statistician 52*(2), 93–100.

Kleinman, K., R. Lazarus, and R. Platt (2004). A generalized linear mixed model approach for detecting incident clusters of disease in small areas, with an application to biological terrorism. *American Journal of Epidemiology 159*(3), 217–224.

Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine 19*, 2555–2567.

Knorr-Held, L. and J. Besag (1998). Modeling risk from a disease in time and space. *Statistics in Medicine 17*, 2045–2060.

Knorr-Held, L. and E. Rainer (2001). Projections of lung cancer mortality in West Germany: a case study in Bayesian prediction. *Biostatistics 2*(1), 109–129.

Knorr-Held, L. and G. Raßer (2000). Bayesian detection of clusters and discontinuities in disease maps. *Biometrics 56*, 13–21.

Knorr-Held, L. and S. Richardson (2003). A hierarchical model for space-time surveillance data on meningococcal disease incidence. *Appl. Statist. 52*(2), 169–183.

Lagazio, C., A. Biggeri, and E. Dreassi (2003). Age-period-cohort models and disease mapping. *Environmetrics 14*, 475–490.

Lindbäck, J. and Å. Svensson (2001). Campylobacter infections in Sweden - A statistical analysis of temporal and spatial distributions of notified sporadic campylobacter infections. Research Report 4, Mathematical Statistics, Stockholm University.

Marshall, C., N. Best, A. Bottle, and P. Aylin (2004). Statistical issues in the prospective monitoring of health outcomes across multiple units. *Journal of the Royal Statistical Society, Series A 167*, 541–559.

Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics 21*(6), 1087–91.

Mode, C. J. (1971). *Multitype Branching Processes - Theory and Applications*. New York: Elsevier.

Morton, A. and B. Finkenstädt (2005). Discrete time modelling of disease incidence time series by using Markov chain Monte Carlo methods. *Appl. Statist. 54*(3), 575–594.

Mugglin, A. S., N. Cressie, and I. Gemmell (2002). Hirarchical statistical modelling of influenza epidemic dynamics in space and time. *Statistics in Medicine 21*, 2703–2721.

Natário, I. and L. Knorr-Held (2003). Non-parametric ecological regression and spatial variation. *Biometrical Journal 45*(6), 670–688.

Nelson, K. P. and B. G. Leroux (2005). Statistical models for autocorrelated count data. *Statistics in Medicine 25*(8), 1413–1430.

Rue, H. and L. Held (2005). *Gaussian Markov Random Fields*. Chapman & Hall/CRC.

Schmid, V. and L. Held (2004). Bayesian extrapolation of space-time trends in cancer registry data. *Biometrics 60*, 1034–1042.

Scott, S. L. (2002). Bayesian methods for hidden Markov models: Recursive computing in the 21st century. *Journal of the American Statistical Association 97*(457), 337–351.

Sonesson, C. and D. Bock (2003). A review and discussion of prospective statistical surveillance in public health. *Journal of the Royal Statistical Society Series A 166*(1), 5–21.

Spiegelhalter, D., N. Best, B. Carlin, and A. van der Linde (2002). Bayesian measures of model complexity and fit(with discussion). *Journal of the Royal Statistical Society Series B 64*, 583–639.

Stroup, D. F., G. D. Williamson, and J. L. Herndon (1989). Detection of aberrations in the occurrence of notifiable diseases surveillance data. *Statistics in Medicine 8*, 323–329.

Sun, D., R. K. Tsutakawa, H. Kim, and Z. He (2000). Spatio-temporal interaction with disease mapping. *Statistics in Medicine 19*, 2015–2035.

Toschke, M. A. and L. Held (2006). Multivariate modelling of infectious disease surveillance data. Technical report, LMU Munich.

Waller, L. A., B. P. Carlin, H. Xia, and A. E. Gelfand (1997). Hierachical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association 92*(438), 607–617.

Watier, L. and S. Richardson (1991). A time series construction of an alert threshold with application to s. bovismorbificans in France. *Statistics in Medicine 10*, 1493–1509.

Williamson, G. D. and G. W. Hudson (1999). A monitoring system for detecting aberrations in public health surveillance reports. *Statistics in Medicine 18*, 3283–3298.

Xia, H. and B. P. Carlin (1998). Spatio-temporal models with errors in covariates: mapping Ohio long cancer mortality. *Statistics in Medicine 17*, 2025–2043.

Yan, P. and M. K. Clayton (2006). A cluster model for space-time disease counts. *Statistics in Medicine 25*, 867–881.

Yao, Y.-C. (1984). Estimation of a noisy discrete-time step function: Bayes and empirical Bayes approaches. *The Annals of Statistics 12*(4), 1434–1447.

Zeger, S. L. (1988). A regression model for time series of counts. *Biometrika 75*(4), 621–629.

Lebenslauf

Mathias Hofmann
geboren am 29.9.1974 in München

Schulbildung

1981 - 1985	Grundschule Pullach
1985 - 1986	Hauptschule des Katholischen Familienwerks, Pullach
1986 - 1995	Pater-Rupert-Mayer Gymnasium, Pullach

Zivildienst

Okt. 1995 - Okt. 1996	Altenheim "Haus am Wiesenweg", Pullach

Studium

Nov. 1996 - Mai 2003	Studium der Statistik an der LMU München
Nov. 1998	Vordiplomsprüfung Statistik
Mai 2003	Diplomprüfung Statistik

Auslandsaufenthalt

Sep. 1999 - Apr. 2000	Erasmus Auslandssemester an der Universitat de Barcelona, Spanien

Beruf

Okt. 2003 - Dez. 2006	Wissenschaftlicher Mitarbeiter bei Prof. Dr. L. Held im Sonderforschungsbereich 386 "Statistische Analyse diskreter Strukturen", Teilprojekt B9 "Statistische Methodik zur Surveillance von infektiösen Krankheiten"