Infectious Disease Prediction with Kernel Conditional Density Estimation

1 Introduction

Accurate prediction of infectious disease incidence is important for public health officials planning resource allocations such as the use of vector control measures, assignments of medical personnel, and implementation of potentially costly personal protective equipment policies for those personnel. Predictive distributions are preferred to point predictions because they communicate uncertainty in the predictions and give decision makers more information, particularly in cases where the predictive distribution may be skewed or have multiple modes.

In this work, we obtain predictive distributions of disease incidence using kernel conditional density estimation (KCDE), a non-parametric approach that has not been applied to predicting infectious disease incidence previously. Our contributions include a novel kernel function that handles both continuous and discrete count data by partially discretizing an underlying Gaussian kernel, and consideration of two types of conditioning variables: periodic functions of the observation time that allow us to capture seasonality in disease incidence, and smoothed observations of past incidence that mitigate the effects of noise that can obscure short to medium term trends in incidence.

KCDE is a method for estimating the conditional distribution of a random vector Y given observations of another vector X. In our work, Y is a measure of disease incidence at some future date or dates (the prediction target) and X is a vector of predictive variables that we condition on in order to make our prediction. For example, X may include observations of incidence over the most recent few time points, weather covariates, or variables indicating the time of year at which we are making a prediction. KCDE has not previously been applied to obtain predictive distributions in the context of infectious disease, but it has been successfully used for prediction in other settings such as survival time of lung cancer patients [[Hall et al.(2004)Hall, Racine, and Li]], female labor force participation [[Hall et al.(2004)Hall, Racine, and Li]], bond yields and value at risk in financial markets [[Fan and Yim(2004)]], and wind power [citejeon2012KCDEWindPower] among others.

Although KCDE has not previously been applied to predicting infectious disease, closely related methods for obtaining point predictions have been employed for diseases such as measles [[Sugihara and May(1990)]] and influenza [[Viboud et al.(2003)Viboud, Boëlle, Carrat, Valleron, and Flahault]]. In the infectious disease literature these methods have been referred to as state space reconstruction and the method of analogues, but they amount to an application of nearest neighbors regression methods. The point predictions obtained from nearest neighbors regression are equal to the expected value of the predictive distribution that are obtained from KCDE if a particular data-dependent kernel function is used in the formulation of KCDE [[Hastie et al.(2009)Hastie, Tibshirani, and Friedman]]. However, KCDE offers the advantage of providing a complete predictive distribution rather than only a point prediction. There is also a long history of using other modeling approaches, such as compartmental models, for infectious disease prediction. A full discussion of those methods is beyond the scope of this article; see *** for a recent review.

Double check citation for sugihara – right paper??

Need to find a review of prediction methods for infectious disease.

There is an extensive literature on KCDE, focusing mainly on estimation of continuous conditional densities. Here we offer a brief overview focusing on multivariate density estimation, possibly with mixed continuous and discrete variables. [Li and Racine(2006)] offer a detailed discussion of this case. Throughout this article we use the term density to refer to the Radon-Nikodym derivative of the cumulative distribution function with respect to an appropriately defined measure. In the case of random vectors where some components are continuous random variables and other are discrete, we take this measure to be a product of Lebesgue and counting measures for the corresponding random variables.

Given observations $\{(\mathbf{x}_t, \mathbf{y}_t), t = 1, \ldots, T\}$ the KCDE estimate of the conditional density of $\mathbf{Y}|\mathbf{X}$ is given by

$$\widehat{f}_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \frac{\sum_{t \in \boldsymbol{\tau}} K^{\mathbf{X},\mathbf{Y}}\left\{(\mathbf{x}', \mathbf{y}')', (\mathbf{x}_t', \mathbf{y}_t')'; \mathbf{H}^{\mathbf{X},\mathbf{Y}}\right\}}{\sum_{t \in \boldsymbol{\tau}} K^{\mathbf{X}}(\mathbf{x}, \mathbf{x}_t; \mathbf{H}^{\mathbf{X}})}. \tag{1}$$

Here, $'$ is the transpose operator and $\boldsymbol{\tau} \subseteq \{1, \ldots, T\}$ indexes the subset of observations used in obtaining the conditional density estimate. In the final density estimate, $\boldsymbol{\tau}$ is typically equal to $\{1, \ldots, T\}$, but proper subsets may be used in the estimation procedures we discuss later.

We will work with a slightly restricted specification of Equation (3) in which the kernel function $K^{\mathbf{X},\mathbf{Y}}$ can be written as the product of $K^{\mathbf{X}}$ and a "conditional kernel" $K^{\mathbf{Y}|\mathbf{X}}$:

$$K^{\mathbf{X},\mathbf{Y}}\left\{(\mathbf{x}',\mathbf{y}')',(\mathbf{x}_t',\mathbf{y}_t')';\mathbf{H}^{\mathbf{X},\mathbf{Y}}\right\} = K^{\mathbf{X}}\left\{\mathbf{x},\mathbf{x}_t;\mathbf{H}^{\mathbf{X}}\right\} K^{\mathbf{Y}|\mathbf{X}}\left\{\mathbf{y},\mathbf{y}_t|\mathbf{x},\mathbf{x}_t;\mathbf{H}^{\mathbf{X},\mathbf{Y}}\right\}. \tag{2}$$

With this restriction, we can rearrange Equation (3) to obtain

$$\widehat{f}_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \sum_{t\in\boldsymbol{\tau}} w_t K^{\mathbf{Y}|\mathbf{X}}\left\{\mathbf{y},\mathbf{y}_t|\mathbf{x},\mathbf{x}_t;\mathbf{H}^{\mathbf{X},\mathbf{Y}}\right\}, \text{ where} \tag{3}$$

$$w_t = \frac{K^{\mathbf{X}}\left\{\mathbf{x},\mathbf{x}_t;\mathbf{H}^{\mathbf{X}}\right\}}{\sum_{t^*\in\boldsymbol{\tau}} K^{\mathbf{X}}\left\{\mathbf{x},\mathbf{x}_{t^*};\mathbf{H}^{\mathbf{X}}\right\}}. \tag{4}$$

We can now interpret $K^{\mathbf{X}}$ as a weighting function, determining the contribution of each observation to our final density estimate according to its similarity with the value $\mathbf{x}$ that we are conditioning on. $K^{\mathbf{Y}|\mathbf{X}}$ is a density function that integrates to 1; intuitively, it gives a contribution to the mass of the final density estimate near the observed value $\mathbf{y}_t$. This is illustrated in Figure ***.

Intuitive description of what this does with pictures.

In order to complete the specification of the estimator given in Equation (3), we must specify the kernel functions $K^{\mathbf{X},\mathbf{Y}}$ and $K^{\mathbf{X}}$ and their parameterization in terms of the bandwidth matrices $\mathbf{H}^{\mathbf{X},\mathbf{Y}}$ and $\mathbf{H}^{\mathbf{X}}$.

Restrictions on kernel functions – something has to integrate to 1.

In settings with mixed continuous and discrete variables, to our knowledge all previous authors have taken these kernel functions to be products of univariate kernels.

Intuitive interpretation of what using diagonal/nondiagonal kernel functions does. In the multivariate setting, the bandwidth determines the orientation of the kernel function.

Functional forms

Estimation. Two main strategies: cross validation and rule-based. Targets for optimization in cross-validation. *** have shown that with dependent data, but small gains in the mean integrated square error of the density estimate relative to the true conditional density can be achieved by

The remainder of this article is organized as follows. We: - describe how kernel density estimation with a non-diagonal bandwidth can be achieved using a partially discretized multivariate normal distribution for the kernel functions.

- prove asymptotic efficiency for marginal and conditional density estimation for a non-product kernel function

- simulation study comparing product and non-product formulations for marginal and conditional density estimation

- applications

2 Method Description

Suppose we observe $\mathbf{z}_t = \{z_{t,1},\ldots,z_{t,D}\} \in \mathbb{R}^D$ at each point in time $t = 1,\ldots,T$. Our goal is to obtain a predictive distribution for one of the observed variables, with index $d_{pred} \in \{1,\ldots,D\}$, over a range of prediction horizons contained in the set $\mathcal{P}$. For example, if we have weekly data and we are interested in obtaining predictions for a range between 4 and 6 weeks after the most recent observation then $\mathcal{P} = \{4,5,6\}$. Let P be the largest element of the set $\mathcal{P}$ of prediction horizons.

For each time $t \in \boldsymbol{\tau}$, we form the vectors $\mathbf{y}_t$ and $\mathbf{x}_t$ representing the prediction target and predictive variables respectively.

In order to perform prediction, we will use lagged observations. Let $\mathbf{l}^{max} = (l_1^{max},\ldots,l_D^{max})$ specify the maximum number of lags for each observed variable that may be used for prediction, and let $L = \max dl_d^{max}$ be the overall largest lag that may be used across all variables. In the estimation procedure we describe in Section 3, we will select a subset of these lags to actually use in the predictions. We capture which lags are actually used in the vector

$$\mathbf{u} = (u_{1,0},\ldots,u_{1,l_1^{max}},\ldots,u_{D,0},\ldots,u_{D,l_D^{max}}), \text{ where}$$

$$u_{d,l} = \begin{cases} 0 \text{ if lag } l \text{ of variable } d \text{ is not used in forming predictions} \\ 1 \text{ if lag } l \text{ of variable } d \text{ is used in forming predictions.} \end{cases}$$

By analogy with the standard notation in autoregressive models, we define

$$\mathbf{y}_t = (z_{t,d_{pred}}, \ldots, B^{(P-1)} z_{t,d_{pred}}) \text{ and}$$
$$\mathbf{x}_t = (B^{(P)} z_{t,1}, \ldots, B^{(P+l_1^{max}-1)} z_{t,1}, \ldots, B^{(P)} z_{t,D}, \ldots, B^{(P+l_D^{max}-1)} z_{t,D})$$

Here, $B^{(a)}$ is the backshift operator defined by $B^{(a)} z_{t,d} = z_{t-a,d}$. Note that the lengths of $\mathbf{y}_t$ and $\mathbf{x}_t$, as well as exactly which lags are used to form them, depend on $\mathcal{P}$ and $\mathbf{l}^{max}$; we suppress this dependence in the notation for the sake of clarity. The vector $\mathbf{y}_t$ represents the prediction target when our most recent observation was made at time $t - P$: the vector of observed values at each prediction horizon $p \in \mathcal{P}$. The variable $\mathbf{x}_t$ represents the vector of all lagged covariates that are available for use in performing prediction.

To make the notation concrete, suppose that $\mathbf{z}_t$ contains the observed case count for week t in San Juan, the observed case count for week t in Iquitos, and the date on Monday of week t, and our goal is to predict the weekly case count in San Juan. Then $D = 3$ and $d_{pred} = 1$. If we want to predict the weekly case counts for the two weeks after the most recently observation, then $p = 2$. If we specify that the model may include the two most recent observations for the case counts in San Juan and Iquitos, but only the time index at the most recent observation then $\mathbf{l}^{max} = (1, 1, 0)$. If our current model uses only the most recently observed case counts for San Juan and Iquitos then $\mathbf{u} = (1, 0, 1, 0, 0)$, where the 1's are in the positions of the $\mathbf{u}$ vector representing lag 0 of the counts for San Juan and lag 0 of the counts for Iquitos. The variable $y_t^{(P)}$ is a vector containing the observed case counts for San Juan in weeks $t + 1$ and $t + 2$; $\mathbf{x}_t^{(\mathbf{l}^{max})}$ contains the observed case counts for San Juan and Iquitos in weeks t and $t - 1$ as well as the time index variable in week t.

In order to perform prediction, we regard $\{(\mathbf{y}_t, \mathbf{x}_t), t = 1 + P + L, \ldots, T\}$ as a sample from the joint distribution of $(\mathbf{Y}, \mathbf{X})$. We wish to estimate the conditional distribution of $\mathbf{Y}|\mathbf{X}$. In order to do this, we employ kernel density estimation. Let $K^{\mathbf{Y}}(\mathbf{y}, \mathbf{y}^*, H^{\mathbf{Y}})$ and $K^{\mathbf{X}}(\mathbf{x}, \mathbf{x}^*, H^{\mathbf{X}})$ be kernel functions centered at $\mathbf{y}^*$ and $\mathbf{x}^*$ respectively and with bandwidth matrices $H^{\mathbf{Y}}$ and $H^{\mathbf{X}}$. We estimate the conditional distribution of $\mathbf{Y}|\mathbf{X}$ as follows:

$$\widehat{f}_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{X} = \mathbf{x}) = \frac{\widehat{f}_{\mathbf{Y},\mathbf{X}}(\mathbf{y}, \mathbf{x})}{\widehat{f}_{\mathbf{X}}(\mathbf{x})} \tag{5}$$

$$= \frac{\sum_{t \in \tau} K^{\mathbf{Y},\mathbf{X}}\{(\mathbf{y}, \mathbf{x}), (\mathbf{y}_t, \mathbf{x}_t), H^{\mathbf{Y},\mathbf{X}}\}}{\sum_{t \in \tau} K^{\mathbf{X}}(\mathbf{x}, \mathbf{x}_t, H^{\mathbf{X}})} \tag{6}$$

$$= \frac{\sum_{t \in \tau} K^{\mathbf{Y}|\mathbf{X}}(\mathbf{y}, \mathbf{y}_t | \mathbf{x}, \mathbf{x}_t, H^{\mathbf{Y},\mathbf{X}}) K^{\mathbf{X}}(\mathbf{x}, \mathbf{x}_t, H^{\mathbf{X}})}{\sum_{t \in \tau} K^{\mathbf{X}}(\mathbf{x}, \mathbf{x}_t, H^{\mathbf{X}})} \tag{7}$$

$$= \sum_{t \in \tau} w_t K^{\mathbf{Y}|\mathbf{X}}(\mathbf{y}, \mathbf{y}_t | \mathbf{x}, \mathbf{x}_t, H^{\mathbf{Y},\mathbf{X}}), \text{ where} \tag{8}$$

$$w_t = \frac{K^{\mathbf{X}}(\mathbf{x}, \mathbf{x}_t, H^{\mathbf{X}})}{\sum_{t^* \in \tau} K^{\mathbf{X}}(\mathbf{x}, \mathbf{x}_{t^*}, H^{\mathbf{X}})} \tag{9}$$

In Equation (5), we are making use of the fact that the conditional density for $\mathbf{Y}|\mathbf{X}$ can be written as the quotient of the joint density for $(\mathbf{Y}, \mathbf{X})$ and the marginal density for $\mathbf{X}$. In Equation (7), we obtain separate kernel density estimates for the joint and marginal densities in this quotient. In Equation (8), we rewrite this quotient by passing the denominator of Equation (7) into the summation in the numerator. We can interpret the result as a weighted kernel density estimate, where each observation $t \in \tau$ contributes a different amount to the final conditional density estimate. The amount of the contribution from observation t is given by the weight w_t, which effectively measures how similar $\mathbf{x}_t$ is to the point $\mathbf{x}$ at which we are estimating the conditional density. If $\mathbf{x}_t^{(\mathbf{l}^{max})}$ is similar to $\mathbf{x}_{t^*}^{(\mathbf{l}^{max})}$, a large weight is assigned to t; if $\mathbf{x}_t^{(\mathbf{l}^{max})}$ is different from $\mathbf{x}_{t^*}^{(\mathbf{l}^{max})}$, a small weight is assigned to t.

In kernel density estimation, it is generally required that the kernel functions integrate to 1 in order to obtain valid density estimates. However, after conditioning on $\mathbf{X}$, it is no longer necessary that $K^{\mathbf{X}}(\mathbf{x}, \mathbf{x}_t, H^{\mathbf{X}})$ integrate to 1. In fact, as can be seen from Equation (9), any multiplicative constants of proportionality will cancel out when we form the observation weights. We can therefore regard $K^{\mathbf{X}}(\mathbf{x}, \mathbf{x}_t, H^{\mathbf{X}})$ as a more general weighting function that measures the similarity between $\mathbf{x}$ and $\mathbf{x}_t$. As we will see, eliminating the constraint that $K^{\mathbf{X}}$ integrates to 1 is a useful expansion the space of functions that can be used in calculating the observation weights. However, we still require that $K^{\mathbf{Y}}$ integrates to 1.

In Equations (5) through (9), τ is an index set of time points used in obtaining the density estimate. In most settings, we can take $\tau = \{1+P+L, \ldots, T\}$. These are the time points for which we can form the lagged observation vector $\mathbf{x}_t$ and the prediction target vector $\mathbf{y}_t$. However, we will place additional restrictions on the time points included in τ in the cross-validation procedure discussed in Section 3.

If we wish to obtain point predictions, we can use a summary of the predictive density. For example, if we take the expected value, we obtain kernel regression:

$$(\widehat{\mathbf{Y}}|\mathbf{X} = \mathbf{x}) = \mathbb{E}_{\widehat{f}_{\mathbf{Y}|\mathbf{X}}}\{\mathbf{Y}|\mathbf{X} = \mathbf{x}\} \tag{10}$$

$$= \int \sum_{t\in\tau} w_t K^{\mathbf{Y}}(\mathbf{y}, \mathbf{y}_t, H^{\mathbf{Y}})\mathbf{y}\, d\mathbf{y} \tag{11}$$

$$= \sum_{t\in\tau} w_t \mathbf{y}_t \tag{12}$$

The equality in Equation (12) holds if the kernel function $K^{\mathbf{Y}}(\mathbf{y}, \mathbf{y}_t, H^{\mathbf{Y}})$ is symmetric about $\mathbf{y}_t$, or more generally if it is the pdf of a random variable with expected value $\mathbf{y}_t$.

Another alternative that we pursue is the use of smoothed observations in forming the lagged observation vectors. We use smoothed case counts on a log scale for the weighting kernels, and the unsmoothed case counts on the original scale for the prediction kernels.

3 Parameter Estimation

We use cross-validation to select the variables that are used in the model and estimate the corresponding bandwidth parameters by (approximately) minimizing a cross-validation measure of the quality of the predictions obtained from the model. Formally,

$$(\widehat{\mathbf{u}}, \widehat{H}^{\mathbf{X}}, \widehat{H}^{\mathbf{Y}}) \approx \underset{(\mathbf{u}, H^{\mathbf{X}}, H^{\mathbf{Y}})}{\operatorname{argmin}} \sum_{t^*=1+P+L}^{T} Q[\mathbf{y}_{t^*}, \widehat{f}(\mathbf{y}|\mathbf{X} = \mathbf{x}_{t^*}; \mathbf{u}, H^{\mathbf{X}}, H^{\mathbf{Y}}, \{(\mathbf{y}_t, \mathbf{x}_t) : t \in \tau_{t^*}\})] \tag{13}$$

Here, Q is a loss function that measures the quality of the estimated density $\widehat{f}$ given an observation $\mathbf{y}_{t^*}$. We have made the dependence of this estimated density on the the parameters $\mathbf{u}$, $H^{\mathbf{x}}$, and $H^{\mathbf{Y}}$, as well as on the data $\{(\mathbf{y}_t, \mathbf{x}_t) : t \in \tau_{t^*}\}$, explicit in the notation. In order to reduce the potential for our parameter estimates to be affected by local correlation in the time series, we eliminate all time points that fall within one year of t^* from the index set τ_{t^*} used to form the conditional density estimate $\widehat{f}(\mathbf{y}|\mathbf{X} = \mathbf{x}_{t^*}; \mathbf{u}, H^{\mathbf{X}}, H^{\mathbf{Y}}, \{(\mathbf{y}_t, \mathbf{x}_t) : t \in \tau_{t^*}\})$.

Talk about proper scoring rules and our particular choice of Q.

We use a forward/backward stagewise procedure to obtain the set of combinations of variables and lags that are included in the final model (represented by $\mathbf{u}$). For each candidate model, we use the limited memory box constrained optimization procedure of [**?**] to estimate the bandwidth parameters. The approximation in Equation (13) is due to the fact that this optimization procedure may not find a global minimum.

4 Examples

In this Section, we illustrate the methods through applications to prediction in examples with several real time series data sets.

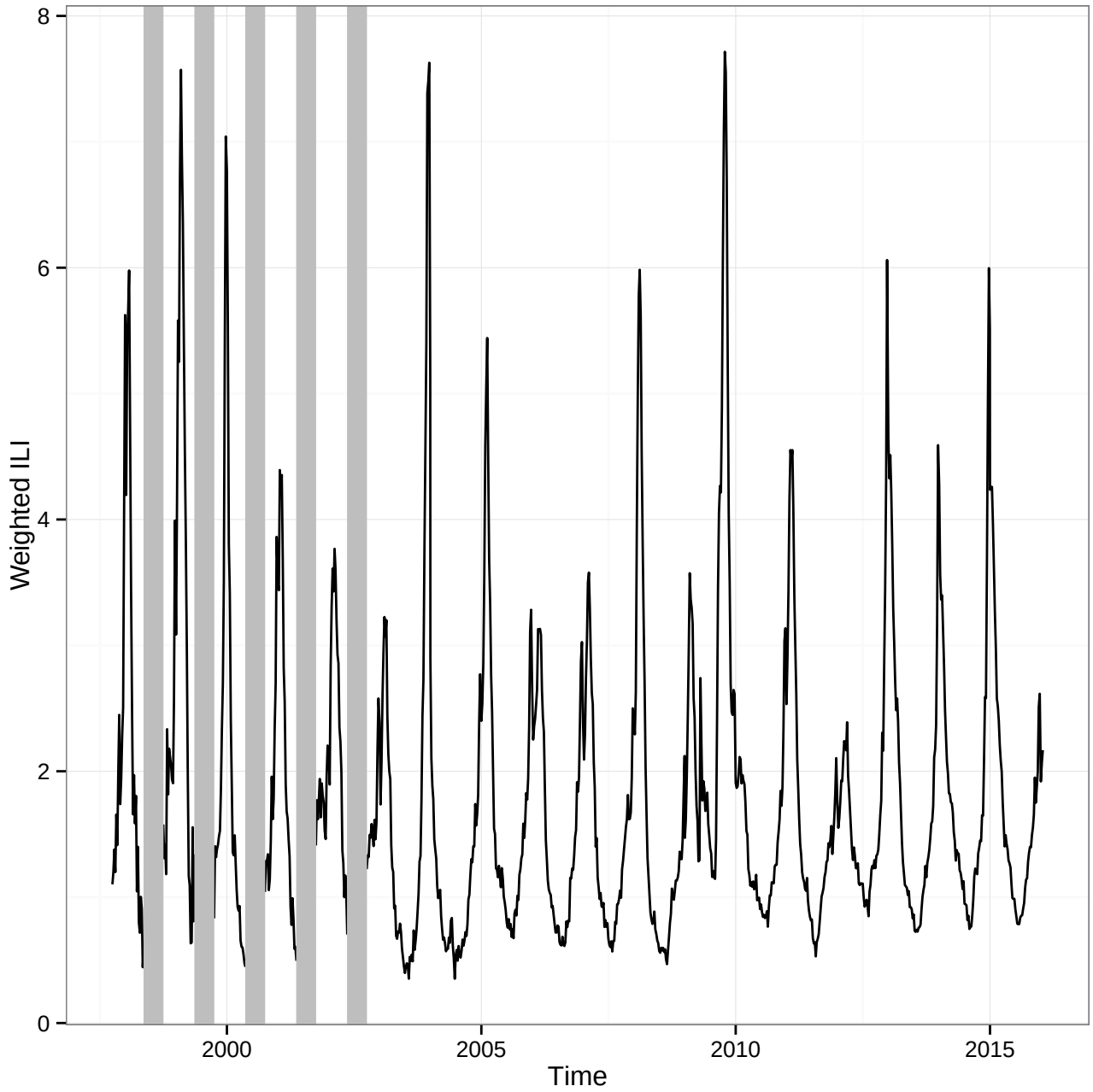
4.1 Example 1: Influenza Prediction

In our first and simplest example, we apply the method for prediction of influenza with prediction horizons of 1 through 4 weeks. Data on influenza incidence are available through `R`'s `cdcfluview` package. Here we create a data set with a nationally aggregated measure of flu incidence

```
library(cdcfluview)
library(plyr)
library(dplyr)
library(lubridate)
library(ggplot2)
library(grid)
library(kcde)

usflu<-get_flu_data("national", "ilinet", years=1997:2015)

##
  |
  |                                                      |   0%
  |
  |++++++++++++++++++++++++++++++++++++++++++++++++++++++| 100%

ili_national <- transmute(usflu,
    region.type = REGION.TYPE,
    region = REGION,
    year = YEAR,
    week = WEEK,
    weighted_ili = as.numeric(X..WEIGHTED.ILI))

## Warning in eval(substitute(expr), envir, enclos):  NAs introduced by coercion

ili_national$time <- ymd(paste(ili_national$year, "01", "01", sep = "-"))
week(ili_national$time) <- ili_national$week
ili_national$time_index <- seq_len(nrow(ili_national))

str(ili_national)

## 'data.frame': 956 obs. of  7 variables:
##  $ region.type : chr  "National" "National" "National" "National" ...
##  $ region      : chr  "X" "X" "X" "X" ...
##  $ year        : int  1997 1997 1997 1997 1997 1997 1997 1997 1997 1997 ...
##  $ week        : int  40 41 42 43 44 45 46 47 48 49 ...
##  $ weighted_ili: num  1.1 1.2 1.38 1.2 1.66 ...
##  $ time        : POSIXct, format: "1997-10-01" "1997-10-08" ...
##  $ time_index  : int  1 2 3 4 5 6 7 8 9 10 ...
```

We plot the `total_cases` measure over time, representing missing values with vertical grey lines. The low season was not measured in the first few years.

```
ggplot() +
    geom_line(aes(x = as.Date(time), y = weighted_ili), data =
ili_national) +
    geom_vline(aes(xintercept = as.numeric(as.Date(time))),
        colour = "grey",
        data = ili_national[is.na(ili_national$weighted_ili), ]) +
    scale_x_date() +
    xlab("Time") +
    ylab("Weighted ILI") +
    theme_bw()
```



There are several methods that we could employ to handle these missing data:

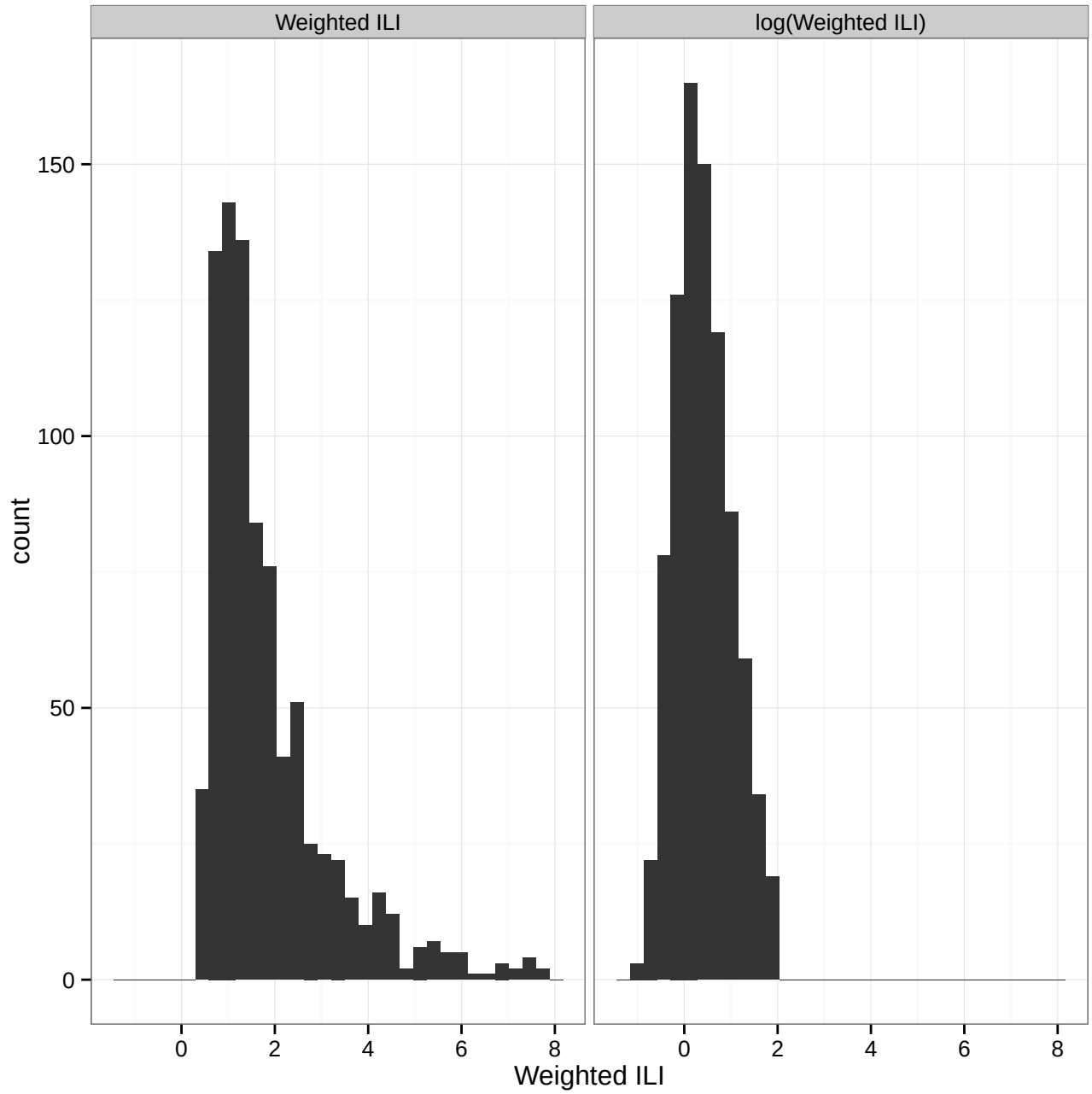
1. Impute the missing values. They are all in the low season, so this should be relatively easy to do.
2. Drop all data up through the last NA.
3. Use the data that are available.

Of these approaches, the first is probably preferred. The concern with the second is that we are not making use of all of the available data. The potential concern with the third is that in the data used in estimation, there will be more examples of prediction of values in the high season using values in the high season and middle of the season than of prediction of values in the high season using values in the low season. This could potentially affect our inference. However, we do not expect this effect to be large, so we proceed with this option for the purposes of this example.

We also plot histograms of the observed total cases on the original scale and on the log scale.

```r
hist_df <- rbind(
        data.frame(value = ili_national$weighted_ili,
        variable = "Weighted ILI"),
    data.frame(value = log(ili_national$weighted_ili),
        variable = "log(Weighted ILI)")
)

ggplot(aes(x = value), data = hist_df) +
    geom_histogram() +
    facet_wrap( ~ variable, ncol = 2) +
    xlab("Weighted ILI") +
    theme_bw()
```

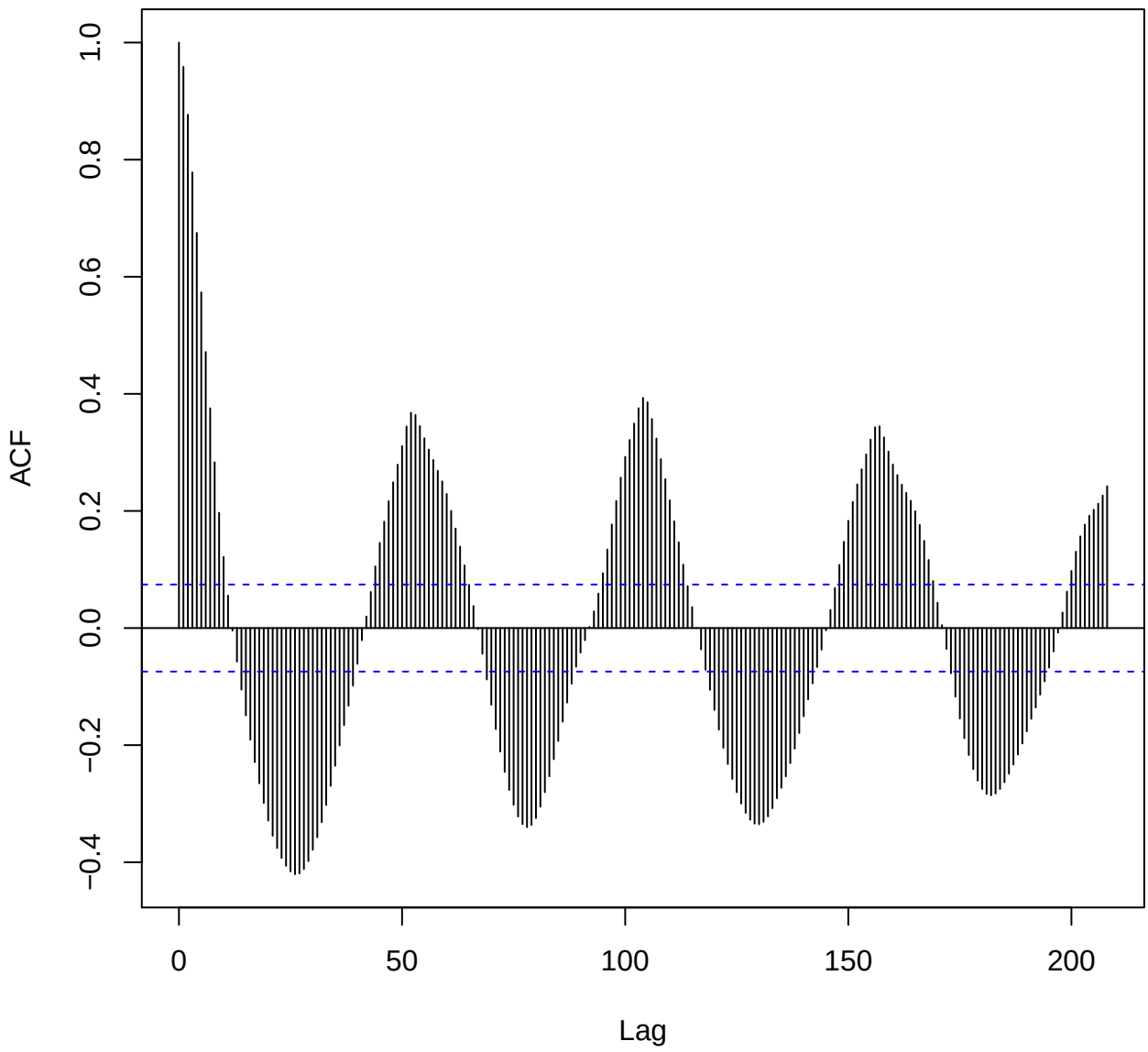
```
## stat_bin:  binwidth defaulted to range/30.  Use 'binwidth = x' to adjust this.
## stat_bin:  binwidth defaulted to range/30.  Use 'binwidth = x' to adjust this.
```



These plots demonstrate that total cases follows an approximately log-normal distribution. In the application below, we will consider modeling these data on both the original scale and the log scale. Intuitively, since we are using a kernel that is obtained from a Gaussian, modeling the data on the log scale should yield better performance. On the other hand, the performance gain may be negligible if we have enough data.

Finally, we plot the autocorrelation function:

```
last_na_ind <- max(which(is.na(ili_national$weighted_ili)))
non_na_inds <- seq(from = last_na_ind + 1, to=nrow(ili_national))
acf(ili_national$weighted_ili[non_na_inds],
  lag.max = 52 * 4)
```

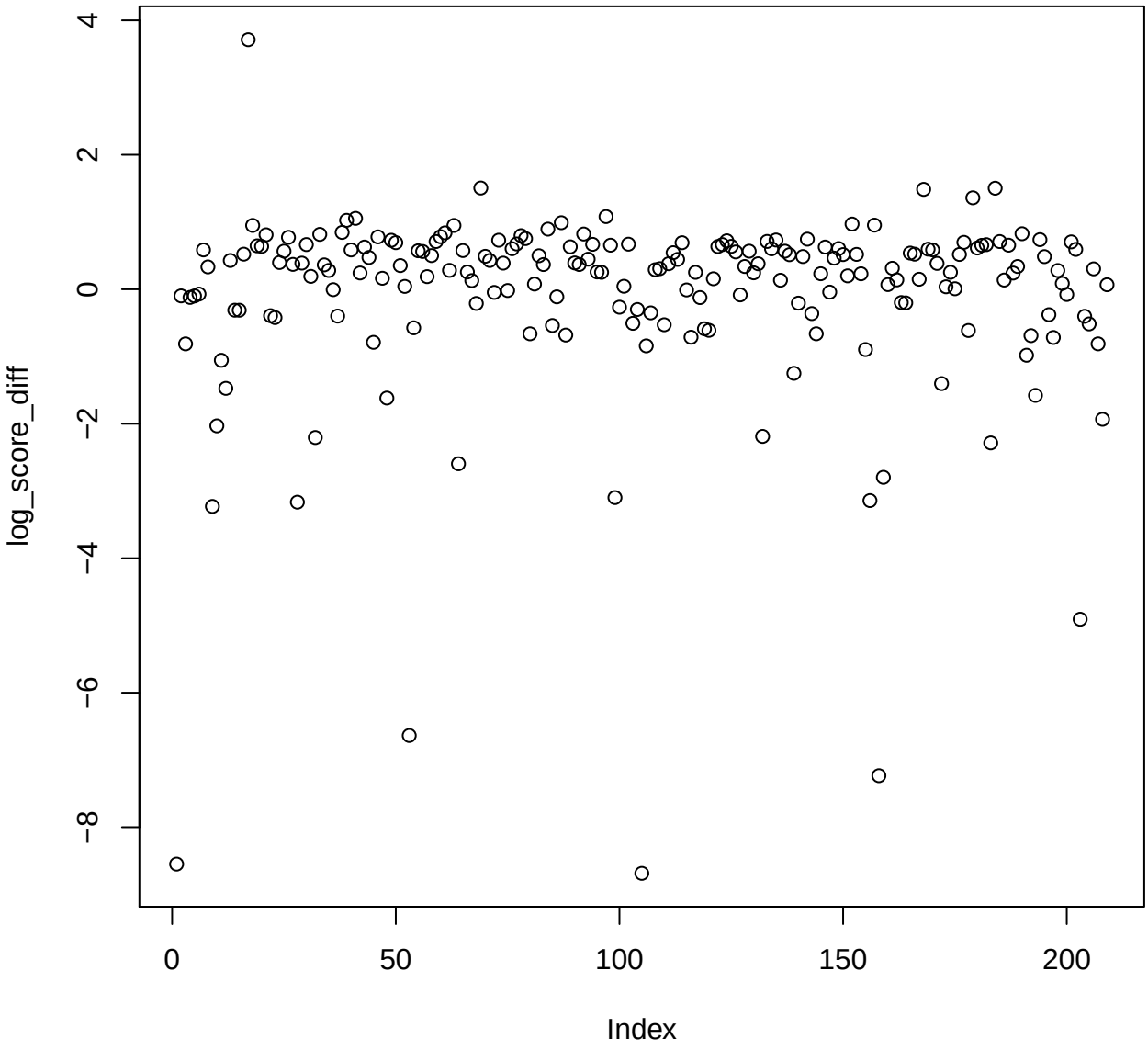


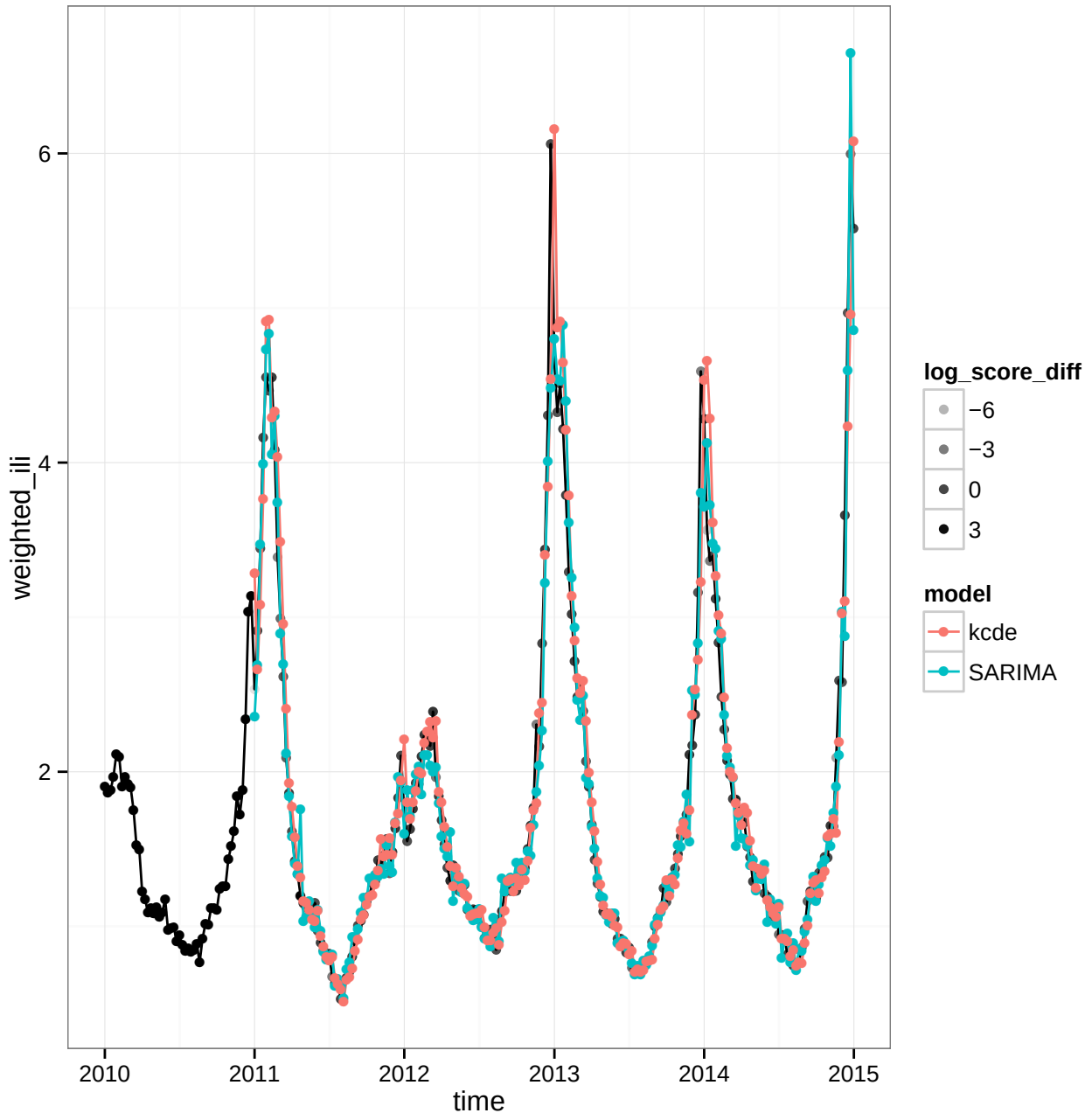
This plot illustrates the annual periodicity that was also visible in the initial data plot above. There is no apparent evidence of longer term annual cycles. We therefore include a periodic kernel acting on the time index with a period of 52.2 weeks (the length of the period is motivated by the fact that in our data, there is a year with 53 weeks once every 5 or 6 years).

We now do some set up for estimation and prediction with kcde. First, we create a list with parameters that specify the kernel function components.

```
## Error in function_list[[i]](value):  could not find function "gather"
## Error in do.call("layer", list(mapping = mapping, data = data, stat = stat, :  object 'ribbons_df'
not found
## Error in do.call("layer", list(mapping = mapping, data = data, stat = stat, :  object 'ribbons_df'
not found
## Error in do.call("layer", list(mapping = mapping, data = data, stat = stat, :  object 'ribbons_df'
```

```
not found
```





5 References

[Fan and Yim(2004)] Jianqing Fan and Tsz Ho Yim. A crossvalidation method for estimating conditional densities. *Biometrika*, 91(4):819–834, 2004.

[Hall et al.(2004)Hall, Racine, and Li] Peter Hall, Jeff Racine, and Qi Li. Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association*, 99(468):1015–1026, 2004.

[Hastie et al.(2009)Hastie, Tibshirani, and Friedman] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Science + Business Media, 2 edition, 2009.

[Li and Racine(2006)] Qi Li and Jeffrey Scott Racine. *Nonparametric econometrics: theory and practice*. Princeton University Press, Princeton, NJ, 2006.

[Sugihara and May(1990)] George Sugihara and Robert M. May. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series, April 1990.

[Viboud et al.(2003)Viboud, Boëlle, Carrat, Valleron, and Flahault] Cécile Viboud, Pierre-Yves Boëlle, Fabrice Carrat, Alain-Jacques Valleron, and Antoine Flahault. Prediction of the spread of influenza epidemics by the method of analogues. *American Journal of Epidemiology*, 158(10):996–1006, 2003.