Statistical Monitoring of Condemnation Rates from Swiss Slaughterhouses

---

Master Thesis in Biostatistics (STA495)

by

Sarah Thommen
05-058-656

supervised by

Prof. Dr. Leonhard Held, University of Zürich
Dr. Flavie Vial, Veterinary Public Health Institute, University of Bern

Zürich, June 2014

University of Zurich[UZH]

Master Program in Biostatistics
www.math.uzh.ch/biostat

STA495: Statistical Monitoring of Condemnation Rates from Swiss Slaughterhouses

Analysis for the Veterinary Public Health Institute of Bern

Sarah Thommen (*sarah.thommen@uzh.ch*)

Version of June 15, 2014

Contents

1 Introduction

The opportunity that made this Master Project possible came from a collaboration between the Institute of Social and Preventive Medicine (ISPM) of the University of Zürich and the Veterinary Public Health Institute (VPHI) of the University of Bern. The VPHI in Bern is a relatively new institution dedicated to animal health. It had been established in July 2009 and is based on a strategic partnership between the University of Bern and the Federal Veterinary Office of Switzerland. This collaboration between academia and veterinary services aims to contribute substantially towards animal health and food safety and thus improve public health in Switzerland. The VPHI is involved in teaching and consulting activities in the context of epidemiology, statistics and veterinary public health and has several different research projects concerning detection, control and prevention of animal diseases and animal welfare-related events. Diseases transmitted to humans directly or through products of animal origin (food) are an additional field of research relevant to the VPHI. This thesis contributes to a larger body of research being carried out at the VPHI on integrating slaughterhouse data into a national syndromic surveillance system for the early detection of emerging diseases in production animals [1].

The field of research to which the topic of this Master Thesis belongs to is called syndromic surveillance. Syndromic surveillance is defined by Triple-S (Syndromic Surveillance System in Europe) as follows: *"Syndromic Surveillance is the real-time (or near real-time) collection, analysis, interpretation and dissemination of health-related data to enable the early identification of the impact (or absence of impact) of potential human or veterinary public-health threats which require effective public health action"* [2]. Out of the two terms *surveillance* refers to the monitoring of a wide range of data to enhance the ability of public health infrastructures to increase public health. The practice of monitoring and analyzing data to detect and respond to disease outbreaks is called biosurveillance. Traditionally, data that are clearly connected to a certain disease (e.g. diagnostic results from laboratory, mortality rates) are used. Collecting such traditional data is often very time consuming and generally associated with high costs. In recent years not only the topic of outbreak detection has been of increasing interest, but especially the *early* outbreak detection of diseases in human and animals has become more and more popular. The aim is to be able to detect, investigate and respond to possible outbreaks early enough to get them under control before disease spreads and serious epidemic or even pandemic evolve. To be able to detect an outbreak at an early stage, the data to analyze are needed as early as possible. In the late 1990s a more modern type of biosurveillance evolved: syndromic biosurveillance. Modern syndromic biosurveillance uses non-specific prediagnostic data which are readily available and can be collected and processed in near real-time without any substantial additional costs. These data are called syndromic data and are not a direct measure of the cases of a specific disease. Instead, they are non-specific preclinical data that are assumed to contain an outbreak signal before diagnostic confirmation of the disease/pathogen. Syndromic data collected in the context of animal health, can be sales/prescription of antibiotic drugs, volume and quality of milk produced, number of condemned carcasses at slaughterhouses or the number of stillbirths. Although none of these data are directly linked to a specific disease, their distribution is still expected to change in the presence of a natural disease outbreak. As such, syndromic data are increasingly used by epidemiologists and public health authorities for (direct) surveillance of animal health and (indirectly) for human health [3].

It has long been recognized, that human health and animal health are tightly connected. The importance of animal health, on which humans depend for food, has been notoriously illustrated in the last two decades. However in the same time also animal welfare and the human sense of responsibility towards their natural environment (including animals) has grown. In the context of animal diseases, on one hand public health infrastructures are interested in the protection, the prevention and the well-being of animals themselves. On the other hand, animal welfare is also linked to human health to a certain degree. This is reflected clearly by the definition of veterinary public health by the World Health Organization (WHO): *"Veterinary public health is the sum of all contributions to the physical, mental and social well-being of humans through an understanding and application of veterinary science"*. One reason why animal health is investigated, is the awareness that zoonotic pathogens can be transmitted from animals to humans (in particular from production animals). 'Zoonosis' refers to an infectious disease that is transmitted into humans by another species. In the context of surveillance and warning systems, the WHO defined the expression emerging zoonosis in 2004 as *"a zoonosis that is newly recognized or newly evolved, or that has occurred previously but shows an increase in incidence or expansion in geographical, host or vector range"* [4].

The possible degree of damage that can be caused by an emerging zoonosis to humans or an emerging disease to animals and that a devastating emerging disease may be closer to reality than expected has been demonstrated several times in the last two decades. Examples include:

1. The acute respiratory syndrome (SARS), which resulted in a SARS-epidemic in 2002/2003. By July 2003, the international spread of SARS resulted in 8'098 SARS cases in 26 countries and caused 774 deaths [5].

2. The spread of the Bovine Spongiform Encephalitis (BSE), an animal disease affecting cattle. Between November 1986 and September 2002, approximately 180'900 cases of BSE were confirmed in the United Kingdom (UK). Cases have appeared in other European countries, in Israel and Japan as well, although in relatively small numbers. Cases of BSE are still occasionally reported but have decreased over the years [6].

3. The foot-and-mouth epidemic in the United Kingdom in 2001, that caused an alarming number of animal deaths (over 50 million animals were slaughtered due to the disease) and enormous economic costs [7].

4. H1N1 influenza (swine flu) in 2009/2010 affecting human as well as animal health. Between April 2009 and March 2010, 40 to 80 million cases were observed only in humans. Due to this pandemic within one year 8'000 to 18'000 people died in different countries over the world [8]. Even though the virus only spreads from human to human and not from infected pigs to human, several hundred thousand pigs were slaughtered in prevention and to contain the epidemic.

About 75% of the new diseases that have affected humans over the past 10 years, have been caused by pathogens originating from an animal or from products of animal origin. Thus emerging zoonotic diseases in food animals are important components of food safety system. In recent years detection of emerging diseases in food animals at various points along the farm-to-fork continuum has been a main interest of research [9, 10].

The world's annual meat production is projected to increase from 218 million tons in 1997-1999 to 376 million tons by 2030. The production increase is a reaction to the ongoing increase in meat consumption, which has been observed over the last 60 years. Meat consumption rose from 44 million tons in 1950 to 284 million tons in 2009 [11, 12]. Due to our concern for food safety and animal welfare the meat production process is tightly regulated, constantly inspected and documented. At present meat inspection data are collected in several countries in Europe (e.g. Finland, Sweden and Switzerland) but are not currently used for prospective syndromic surveillance [13]. There have been studies which showed, that disease outbreak patterns of production animals are reflected in the condemnation rates of slaughtered animals. These studies and several recent reports by the European Food Safety Authority [14, 15, 16] support the assumption that meat inspection data from slaughterhouses are a potential data source for the surveillance of diseases in production animals [17, 18].

In Switzerland the preparation and manufacture of food is strictly regulated by law. Regulations, inspections, obligations to report and the documentation of these tasks lead to the aggregation of data about the condition of meat products before and after entry into the production process. In Switzerland animals to be slaughtered for meat production are inspected twice at the slaughterhouse, once at their arrival and once after the slaughter. In the first inspection it is mandatory to report the number of animals that arrived sick or injured at the slaughterhouse. Those animals are separated into a slaughter group called *"emergency slaughter"* and slaughtered after the healthy animals in order to decrease the risk of pathogen transmission. After the slaughter is processed every carcass in the slaughterhouse is visually inspected by a meat inspector and classified into one of the following classes:

1. entirely fit for human consumption
2. wholly condemned (including organs and blood)
3. partially condemned (only part of the carcass is unfit for human consumption)

The outcome of the meat inspection needs to be reported to the Federal Food Safety and Veterinary Office. It is compulsory to report the number of wholly condemned carcasses (reporting of partially condemned carcasses is not compulsory). The inspectors also have to state the reason why the carcasses were condemned by choosing one out of 44 possible reasons. The reasons also have to be reported back to the producers. Reasons include for example pronounced weight loss, abscesses or lesions. All possible reasons for whole carcass condemnations are listed in the Swiss legislation [19]. Up to now it is mandatory to report those informations on a monthly basis for cattle, pigs and small ruminants (sheep and goats). These data collected in Switzerland could be a valuable syndromic indicator of national herd health. This thesis will assess the possibility and the prospects of integrating these data in a national syndromic surveillance system for early detection of emerging and re-emerging diseases in production animals.

2 Research Questions

For this master project a complete dataset of meat inspection post-mortem results in Switzerland is analyzed and investigated. One elemental aim of the thesis is to find out whether the accumulated data constitutes a convenient dataset for syndromic surveillance. The target outcome of the thesis is an algorithm that can be used during the prospective surveillance of monthly whole carcass condemnation rates for production animals in Switzerland. To reach this very specific goal, the following research questions were investigated:

1. Does the condemnation rate depend on external time related factors (seasonality, trend, autocorrelation)?

2. What statistical model describes the condemnation rates retrospectively in a reasonable way?

3. How can outbreak data for monthly condemnation rates of slaughtered production animals be simulated?

4. How good is the performance of a quasi-Poisson regression algorithm for the simulated outbreak data?

5. Is the quasi-Poisson regression algorithm convenient for prospective outbreak detection using the slaughterhouse data?

If the outcome of the project is promising and a convenient algorithm can be found, this Master Thesis could contribute to show the necessity and the importance of such reporting systems. It could contribute to the improvement of syndromic surveillance for production animals and thus improve outbreak detection and animal welfare in Switzerland. In any case this Master Thesis contributes to evaluate the use of monthly condemnation rates form Swiss slaughterhouses for early outbreak detection.

3 Methods

Statistical Methods: The analysis of the time series data was split into two parts. First the data were analyzed retrospectively before being analysed a second time based on prospective methodology.

Retrospective Analysis: To analyze the data retrospectively, the time series data were decomposed into seasonal, trend and residual components. Two different methods were used. The first method performs a seasonal-trend decomposition of time series based on LOESS. It has been proposed by Cleveland et al. [20] and is implemented in R [21] as the {stl} function in the {stats} package. The distributional assumption on which the {stl} function is based on is the normal distribution. The slaughterhouse time series data used in this study were proportional rate data (number of condemned carcasses relative to the number of slaughtered animals per month). Thus the data are assumed to be Poisson distributed. In order to stabilise the variance and achieve homoscedasticity, the data were transformed before applying the {stl} function. The transformation chosen was $\arcsin(\sqrt{x})$, which is commonly used for proportions [22]. The transformed data were then decomposed additively into three components:

$$X_t = T_t + S_t + R_t$$

$\{X_t\}$ is the monthly measure of the proportion of carcasses condemned. $\{S_t\}$ represents the seasonal component and $\{T_t\}$ the trend part. By applying the {stl} function, the seasonal and trend component are found by LOESS smoothing. In fact the seasonal component for our data was found by taking the mean of the de-trended seasonal sub-series. (The series of all January values, all February values and so on). In a second step the seasonal values are removed from the observed data and the remainder is smoothed by LOESS to find the trend. $\{R_t\}$ is the remainder component, that is the residuals from the seasonal plus trend fit. Data points with very high reminders are considered to be outliers and get replaced by locally weighted robust estimates of the observed values. Thus the decomposition is done in a robust way. To find the estimates, the procedure is iterated a few times.
The output of the decomposition based on LOESS is illustrated by a plot of the observed data and the estimated components for each time point. There is a grey bar at the right hand side of each component graph to allow a relative comparison of the magnitudes of the information coming from each component. It helps to see how much variation in the data can be attributed to each component or to the remaining part.

The second method used to decompose or model the time series data (number of condemned carcasses per month $\{y_t\}$) in a retrospective manner, is a model framework, proposed by Held et al. [23]. The function used for the decomposition is called {hhh4} and is implemented in the package {surveillance} [24, 25]. The {hhh4} function allows to model count data of uni- and multivariate time series in a flexible way. As suggested by Held et al. [23] the distributional assumption for the model is Poisson:

$$P\left(y_t\right) = \frac{e^{-\mu_t} \mu_t^{y_t}}{y_t!}$$

An alternative distributional family that is implemented is the negative binomial distribution, which also accounts for over-dispersion. By applying the {hhh4} function, the mean incidence (here the number of condemned carcasses per month) is decomposed additively into an autoregressive component and an endemic component. Assuming that the data are Poisson distributed the conditional mean is:

$$\mu_t = \lambda_t y_{t-1} + \nu_t \times e_t, \quad (\lambda_t, \nu_t > 0)$$

The autoregressive component $\{\lambda_t\}$ and the endemic component $\{\nu_t\}$ are unknown quantities that can be estimated. The autoregressive part is supposed to capture possible outbreaks and can include a seasonal pattern or a long term trend, thus $\{\lambda\}$ can be dependent on the time. If there is neither a seasonal pattern nor a trend included, an overall $\{\lambda\}$ is estimated independently of time $\{t\}$. The endemic component models the baseline amount of incidences and can include seasonality and trend. The endemic component is multiplied by the offset $\{e_t\}$ to adjust for variation in the number of total animals slaughtered per month. Using the Poisson model or the negative binomial model respectively the estimated variances of the mean incidence rates are:

$$Var(\mu_t) = \mu_t$$
$$Var(\mu_t) = \mu_t + \mu_t^2 \times \theta, \quad \text{(with estimated overdispersion parameter } \theta)$$

All the variation of seasonal patterns and long term trends are motivated by patterns observed in the time series data, in the season-trend decomposition based on LOESS and in the findings that were published by Vial and Reist [26]. The following parametric models for the endemic and autoregressive component were used:

$$\log(\nu_t) = \alpha + \beta t + S_t$$
$$\log(\lambda_t) = \tau + \omega t + A_t$$

In the endemic part a baseline condemnation rate is estimated with the intercept $\{\alpha\}$, the trend is estimated with the parameter $\{\beta\}$ and the seasonal component is estimated with different terms for $\{S_t\}$. In the autoregressive component, a baseline estimate of the impact of the observation from the previous month on the current month is estimated with $\{\tau\}$. A long term trend for the dependence of the observations on the previous ones can be estimated with the parameter $\{\omega\}$. In this model framework a seasonal pattern within the autoregressive part can be estimated with different terms for $\{A_t\}$. There were four different types of trends used to model the data:

1. no trend (t0)
2. (log-) linear trend (t1)
3. (log-) linear trend starting in 2010 (t2010)
4. no trend but a shift in the intercept in 2010 (j2010)

For the different seasonal patterns that could be included in the endemic or in the autoregressive part the following seven variations of parametric models were used (the same models were used for A_t as for S_t):

1. no seasonality (s0)
2. a seasonal impact of each month (monthly)
3. an impact of December only (so-called Christmas effect referred to as xmas seasonality)
4. seasonality with 1 to 4 harmonics per year modeled by a combination of sine and cosine functions suggested by Held et al. [23] (s1 to s4):

$$S_t = \sum_{h=1}^{H} \gamma_h \sin(\omega_h t) + \delta_h \cos(\omega_h t)$$

The estimated parameters γ_h and δ_h depend on the number of harmonics $\{H\}$ that are included and $\omega_h = 2\pi h/\text{freq}$ are Fourier frequencies (e.g. freq $= 12$ for monthly data). Figure 1 illustrates these seasonal patterns.

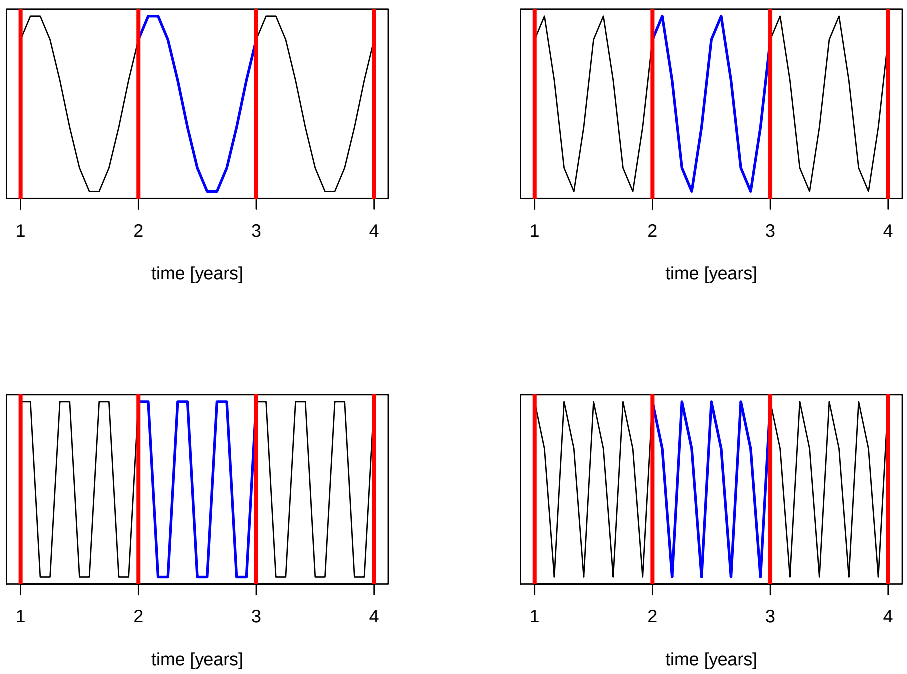


Figure 1: Seasonality with 1 to 4 harmonics per year modeled by a combination of the sine and cosine functions. (topleft: 1 harmonic, topright: 2 harmonis; buttomleft: 3 harmonics, buttomright: 4 harmonics)

To select the best fitting model, all combinations of the different seasonal patterns and long term trends within the autoregressive and endemic components were fitted. Furthermore models which exclude a whole component (either the autoregressive or endemic component) were evaluated. During preliminary analysis of a subset of the models, it was asserted that the negative binomial models fit the data generally better than the Poisson models (see Appendix). Therefore only negative binomial models were used for the final model selection. In order to analyze the different data sets in a well-structured way, the same models were fitted on all the different data sets used. In total 840 models were fitted for each data set. The criteria used for model selection was the BIC. It was shown that the BIC performs much better than AIC for time series model selection in small samples, as the AIC tends to overfit the data [27]. The best model with reliable estimates was then accepted to be used for data simulation.

After the decomposition, based on LOESS or by using the {hhh4} function, the autocorrelation function of the residuals was plotted using the {acf} function of the {stats} package that is included in R [21], in order to check visually whether the models captured the seasonal patterns and the trends as expected.

Data Simulation: Data simulations were generated by using the best model found in the retrospective analysis. First baseline data series with a length of 72 months were generated by using the best models found. In a second step outbreak cases with random size and random starting time points were generated and added to the baseline time series. To each simulated baseline time series there was exactly one outbreak added within a defined *"Outbreak-Risk-Period"* (from time point 39 to 62). The time series were structured into three periods: the *"Baseline-Period"* of a bit more than 3 years (38 months), the *"Outbreak-Risk-Period"* of 2 years (24 months) and the *"Post-Outbreak-Period"* of 10 months (from time point 63 to 72). The starting time points of the outbreaks were randomly sampled from the 24 time points within the Outbreak-Risk-Period. The outbreak data sets were simulated according to the paper of Noufaily et al. [28]. The outbreak sizes were generated by random Poisson variables. Thereby the estimated standard deviation of the simulation model was multiplied with different scaling parameters k (we chose k 2 to 10). The product was then used as the mean rate to generate random Poisson numbers. Each case of the total outbreak size was then randomly distributed to the time points from outbreak start to end according to a lognormal distribution with a mean of 0 and a standard deviation of 0.5. For each parameter k 1000 time series were simulated. For parameters in the range of 2 to 10 the resulting outbreak durations were typically between 3 to 6 months. Figure 2 helps to visualize the generation process of one simulation.

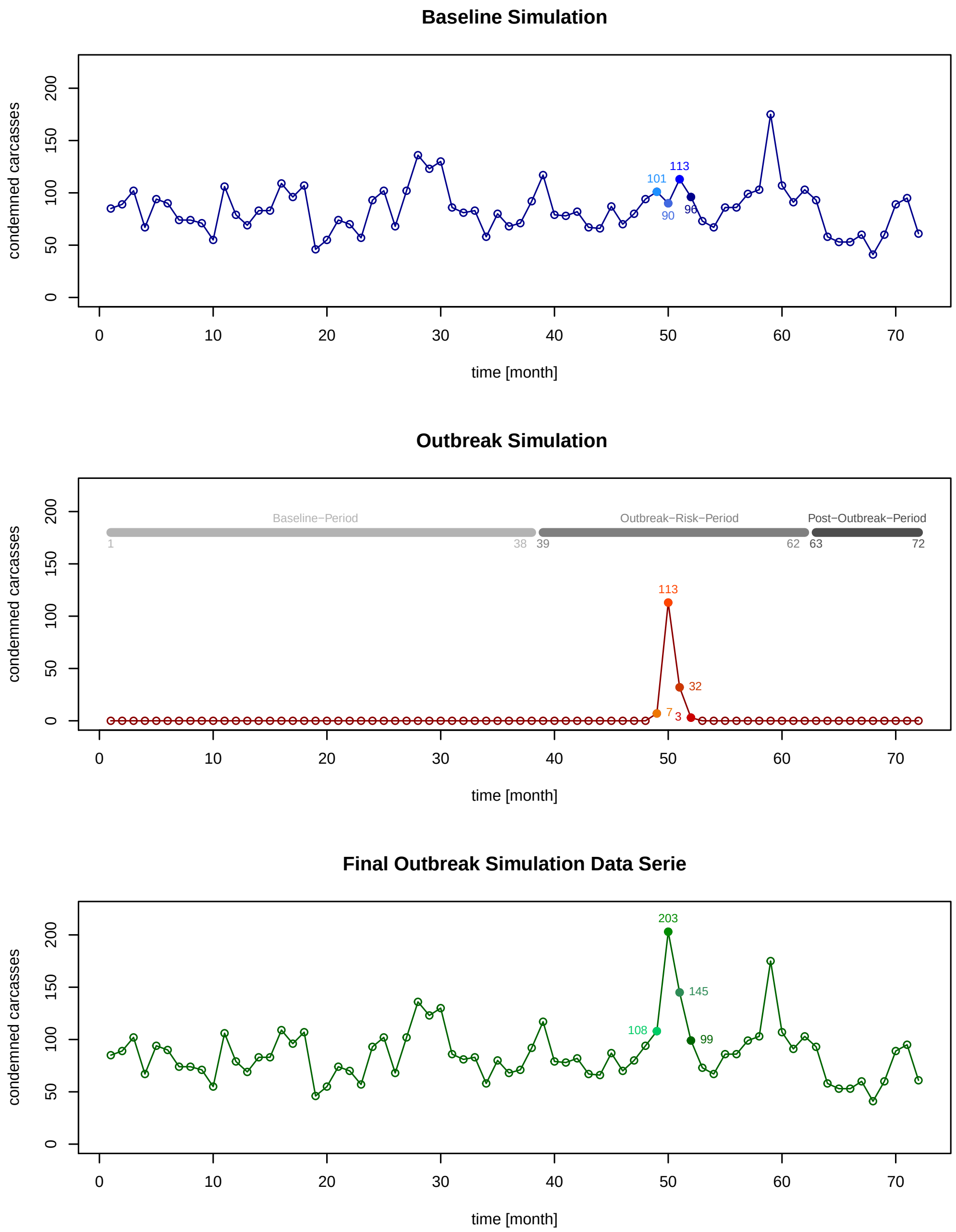


Figure 2: Illustrative example of outbreak data simulation.

Prospective Analysis: To perform outbreak detection on the simulated time series, the improved Farrington algorithm was applied. Farrington algorithm is an outbreak detection method published in 1996 by Farrington et al. [29]. The improved Farrington algorithm, as suggested by its function name, {farringtonFlexible}, is the more flexible version of the original Farrington algorithm. The improved Farrington method was described by Noufaily et al. [28]. It is implemented in the R-package {surveillance} [30]. The first step of the algorithm is to fit a log linear quasi Poisson model using the available baseline data (historic data). The amount of historic data that should be used as reference values to fit the model can be chosen (with parameter {b}) such that only recent values (within b years from current time point) are included to fit the model.

$$E[y_t] = \mu_t = (\alpha + \beta t + f_t) \times e_t, \qquad Var(y_t) = \theta \times \mu_t, \quad \text{(with } \theta > 1)$$

The model can include a baseline incidence rate $\{\alpha\}$, a trend $\{\beta\}$, a seasonal pattern $\{f_t\}$ and the population offset $\{e_t\}$. In the first algorithms tested the parameters were chosen according to the results of the retrospective part of the analysis. Thus a trend was only included if there was evidence for a trend based on the retrospective analysis otherwise no trend was fitted. Seasonality was handled either as in the original Farrington algorithm, as supposed by the improved Farrington algorithm or no seasonal component was included at all. To exclude seasonality, the parameter $\{w\}$ was set to 6. In this case all time points within the time window (2w + 1), evidently data of the whole year, are used to fit the model. Using the original algorithm, excluding data which are assumed to not lay in the current season is the only one way to account for seasonality. To still be able to estimate a trend parameter, the window size was set to 3 (w=1). With an even more restrictive window size (w=0) even with the largest possible learning period (b=3) the sample size for the model fit is only 3. Thus by estimating three parameters (a trend parameter, an intercept and the overdispersion parameter) the model is overfitting the data (the number of estimated parameters (p) is equal to the sample size (n) and thus p<n is violated). The inclusion of only a small subset of the available reference data is in general one of the limitations of the original Farrington algorithm [28]. To consider seasonality and simultaneously include the biggest part of the historic data, in FarringtonFlexible the parameter {noPeriods} is introduced. This gives the option to choose a number of seasonal periods that are modelled by a zero order spline function with $\{noPeriod + 1\}$ knots. This option allows the use of more baseline values to calculate seasonal factors (with 1 to 11 levels) for all the time points outside of the window (2w + 1) [30]. While using this option the parameter $\{w\}$ was chosen according to the retrospective result (e.g. if a seasonality with two harmonics per year should be modeled $\{w\}$ was set to 1 and $\{noPeriods\}$ was set to 3). For each data set the the result of the retrospective analysis decided if seasonality was included in the Farrington algorithm.

In a second step the estimates of the model fit are used to predict the expected observation and more important the threshold that defines the upper limit of alarm free observations. There are several options to calculate the threshold. For our analysis the option "muan" was chosen, as it is statistically the most correct one. With this method the threshold (one-sided) $(1-\alpha) \times 100\%$ prediction interval) is calculated based on the assumption that the estimated prediction parameters are asymptotically normal distributed. Thus the threshold can be derived from the $(1-\alpha) \times 100\%$ quantile of the normal distributed estimates (thresholds based on the 0.995 and 0.975 quantile were used). The uncertainty of the estimated overdisperson parameter is thereby disregarded. If the threshold is exceeded by the simulated observed value, the counts are considered as possible outbreaks. Alarms are flagged by calculating the following exceedance score {Z} for those possible outbreak counts:

$$Z = \frac{y_0 - \hat{\mu}_0}{U_0 - \hat{\mu}_0}$$

There by $\{y_0\}$ is the observed count for the current time point, $\{\hat{\mu}_0\}$ is the estimated current expected value and $\{U_0\}$ is the calculated threshold value for the current time point. An alarm is flagged if $Z \geq 1$ [28]. Time points for which an alarm is flagged should be investigated in greater detail.

After the evaluation of the algorithm with parameters chosen according to the retrospective analysis, several additional algorithms were tested. The purpose was to see if the performance could be improved by modifying the parameter settings of {b} (b=2 or b=3), {w} (w=6, w=0, w=1, w=2) and {trend} (trend=TRUE). The hypotheses were that the performance can be improved by (a) increasing the learning period of the algorithm (form b=2 to b=3) or (b) by including a trend (if not already included). As a consequence of parameter {b} modification the window size had to be adjusted (w=0, w=1 or w=2). The algorithm with the most restrictive windowsize (w=0) only includes data of the months identical to the current month for each of the {b} previous years. To include both extreme options for {w} (0 and 6) this algorithm was evaluated as well. This algorithm is less reliable because of the small sample size included. To include the biggest possible amount of historic data and at the same time keep parameter w=0, {noPeriods}=11 was included in an additional algorithm tested. The inclusion of a trend even if there was no evidence for a trend based on the retrospective analysis was based on the recommendation in the paper of Noufaily et al. (it was stated that a trend should always be included) [28]. The parameter {pastWeeksNotIncluded} was set to 0 to not exclude the highly informative most resent historic data. Therefore the adaption of the algorithm to emerging outbreaks had to be expected and consequently the reduction of sensitivity and the increase of specificity. In addition the algorithms with w=6 were also tested with the parameter {pastWeeksNotIncluded} set to 2 in order to evaluate the influence of emerging outbreaks that started shortly before the current time point on algorithm performance. To not further decrease the amount of the very recent historic data that is included, parameters bigger than 2 were not tested. As there were no outbreaks included in the baseline data none of the algorithms included any re-weighting of high values. The provided options to account for minimal outbreak sizes can be neglected for all algorithms ({limit54} was kept, as it did not have any effect on the analysis and {powertrans} was set to ”none” as no transformation was needed). For all algorithms tested the option to include an offset was applied, thereby the offset of the collected historic data was used for all simulations.

In order to evaluate the performance of the algorithms, the false positive rate (FPR) and the probability that an outbreak is detected (POD) were calculated. To calculate the FPR, the number of alarms flagged in outbreak free months within the Outbreak-Risk-Period was divided by the total number of months that were outbreak-free and tested. The POD was calculated by dividing the number of outbreaks that were detected by the total number of simulations which were tested. Thereby an outbreak was regarded as detected, if there was at least one alarm flagged during the entire outbreak duration. Thus FPR indicates a rate per month and POD a rate per outbreak. Next to the FPR and POD the mean outbreak duration for each parameter k as well as the mean outbreak size, the mean time to detection in months (TTD) and the mean number of cases added until detection (CUD) were recorded and evaluated. The mean TTD and CUD were calculated only from simulations during which an outbreak was detected. A mean TTD of 1 indicates that the mean time until an alarm was flagged is 1 month after the start of the outbreak. A mean CUD of e.g. 27 means that out of the all injected outbreak cases, 27 were added until the outbreak was detected (including the cases of the detection month).

Software: All statistical analyses were performed in R [21].

4 Data

The data used in this study were extracted from the "**Fle**isch**ko**ntrolldatenbank" (FLEKO database). FLEKO is a database owned by the Swiss Federal Food Safety and Veterinary Office (FSVO). FLEKO contains ante- and post-mortem meat inspection data from all animals slaughtered in authorised slaughterhouses of Switzerland (over 600). FLEKO database includes information about all hooved animals slaughtered in Switzerland including cattle, pigs and small ruminants. The data mandatory to report to FLEKO are the number of normal and emergency slaughtered animals (ante-mortem inspection), the number of whole carcass condemnations (post-mortem inspection) and the reason for condemnation. The official veterinarian in charge needs to report the results of the visual inspections to FLEKO in a monthly manner. This obligation started at the end of 2006. Thus full data of the period between January 2007 and December 2012 could be extracted. FLEKO was originally created for economic reason, in order to be able to keep track of the number of animals slaughtered whose by-products would need disposal. The slaughterhouses are supported financially by the government in the proper disposal of animal by-products [26, 19].

FLEKO provides also details of the slaughterhouse and canton in which each individual was processed. Preliminary analysis of the data by Dr. Flavie Vial showed that there are several factors which could have an effect on the number of animals slaughtered as well as on the number of condemned carcasses. There are for example connections between the meat price and the number of cattle and pigs sent for slaughter. The slaughterhouse size has an effect on the number of condemned carcasses for some slaughter and animal types too. There were also cantonal differences detected in the rate of carcass condemnations. However, these effects have not been further investigated. Furthermore it was seen that more than 100 slaughterhouses, all of which process more than 1'000 animals a month, did not report a single whole carcass condemnation case over the 6 years of data acquisition. Based on this observation some non-recording bias is expected in the data [26].

Only part of the available data from FLEKO was extracted for this study. Two animal types (cattle and pigs) and both slaughter groups (normal and emergency slaughter) were included in the analysis. To each individual the following information was used:

- total number of carcasses processed each month
- total number of carcass condemnations for each month
- time (month and year)

The descriptive statistics section as well as the analysis section are structured according to the different data sets analyzed. Each section contains one subsection for each data set (normal slaughtered cattle, emergency slaughtered cattle, normal slaughtered pigs and emergency slaughtered pigs).

The study includes in total 20.4 million observations that were collected over a period of 72 months. Table 1 shows the number of observations per animal type and slaughter group. Overall many more pigs (17 million) than cattle (4 million) were slaughtered within the given time frame. For the normal slaughter group the number of slaughtered pigs was approximately 4 times higher than the number of slaughtered cattle. However, it was the opposite for the emergency slaughter group. The number of emergency slaughtered cattle was approximately 2 times higher than the number of emergency slaughtered pigs. The portion of animals that were classified into the emergency slaughter group upon their arrival, was a minor part of all slaughters reported. 1.71% of cattle and 0.2 % of pigs arrived injured or sick at the slaughterhouse. Thus the general state of health prior slaughter for cattle is inferior compared to pigs. This could be explained by the fact that accidents occur more often to cattle, as they are more often kept on alpine pastures [26].

	normal slaughter	emergency slaughter
cattle	3’750’805	65’304
pigs	16’517’599	33’720
total	20’268’404	99’024

Table 1: Number of observations per animal type and slaughter group

Table 2 shows the number of condemnations per type of animal and slaughter group together with the corresponding percentage of condemned carcasses. The condemnation proportions observed for normal slaughtered pigs and cattle were, with 0.15% and 0.16% respectively, fairly low and very similar. As to be expected, the percentage of carcasses condemned after normal slaughter is much lower than after emergency slaughter. The percentage of condemnation in the emergency group was about 9.75% for pigs and approximately 20.4% for cattle. The discrepancy of the two different animal types could be a result of the widespread injuries and hematomas which are often observed for cattle that were involved in an accident on alpine pastures (leading to whole carcass condemnations). Another reason could be that abscesses on emergency slaughtered pigs are more often isolated compared to cattle (leading to partial carcass condemnation that are not reported) [26].

Taken both slaughter groups together, the mean condemnation proportions for pigs and cattle respectively are 0.17% and 0.5%. These proportions are comparable to the proportions observed for pigs and cattle respectively in Ontario (0.37% and 0.4-0.8%), for pigs in UK (0.35%) and for cattle in France (0.67%) [26].

	normal slaughter		emergency slaughter	
	#	%	#	%
cattle	5863	0.16	13322	20.4
pigs	25310	0.15	3288	9.75
total	31173	0.15	16610	16.77

Table 2: Condemnations per animal type and slaughter group

4.1 Descriptive Statistics: Normal slaughtered Cattle

The observed mean number of cattle slaughtered normally each month was 52'095. In the upper panel of Figure 3 all the observed monthly numbers are shown together with the mean (green line) and the standard deviation (dotted red line). All monthly counts range between 42'836 and 64'073 normal slaughters. The observed number of normal slaughters for cattle is highly fluctuating over time. There is a slight increase in the number of slaughtered cattle visible over the whole time period. The months with the highest and lowest number of cattle slaughtered were March and July respectively.

In the second panel of Figure 3 the number of condemned carcasses is displayed. The mean number of carcasses declared as condemned within the normal slaughtered cattle group was 81 per month with a standard deviation of 19. The minimal and maximal number of condemned carcasses reported per month was 49 and 153 respectively. This corresponds to a condemnation proportion of 0.1% to 0.26% with a mean of 0.16%. Even though the number of slaughters increased slightly over time, the number of condemned carcasses was steadily low. There is one period between June and December 2009 that shows an increased number of condemned carcasses, peaking in December 2009. This value drops again in January 2010 and does not increase noticeable anymore afterwards.

	Slaughtered cattle	Condemned cattle	Proportion [%]
min	42'836	49	0.10
mean	52'095	81	0.16
median	51'457	79	0.16
max	64'073	153	0.26
sd	5'211	19	0.03

Table 3: Summary statistics of normal slaughtered cattle

Almost the same pattern can be observed in the percentage of condemned carcasses after normal slaughter. There is neither an obvious long term trend visible over the whole time period of data acquisition nor a seasonal pattern. The observed condemnation proportions are distributed almost symmetrically around the mean. The fluctuation in the data is rather low, but the variability is considerable high. The peak in the number of condemned carcasses, seen in December 2009, is visible in the condemnation proportion as well, but it is downscaled. There are two months (December and February) that could have a seasonal impact on the data, as there is in general an increase in the percentage of condemned carcasses visible for those two months. This increase could be an effect of winter, as food and hay are limited in winter months. Cattle that are less productive or not fit enough to be sent back to the alpine pastures in summer may be sent for slaughter during those months [26]. Figure 3 shows the time series data of normal slaughtered cattle regarding the total number of animals slaughtered, the number of condemned carcasses and the condemnation proportions. The summary statistics are listed in Table 3.

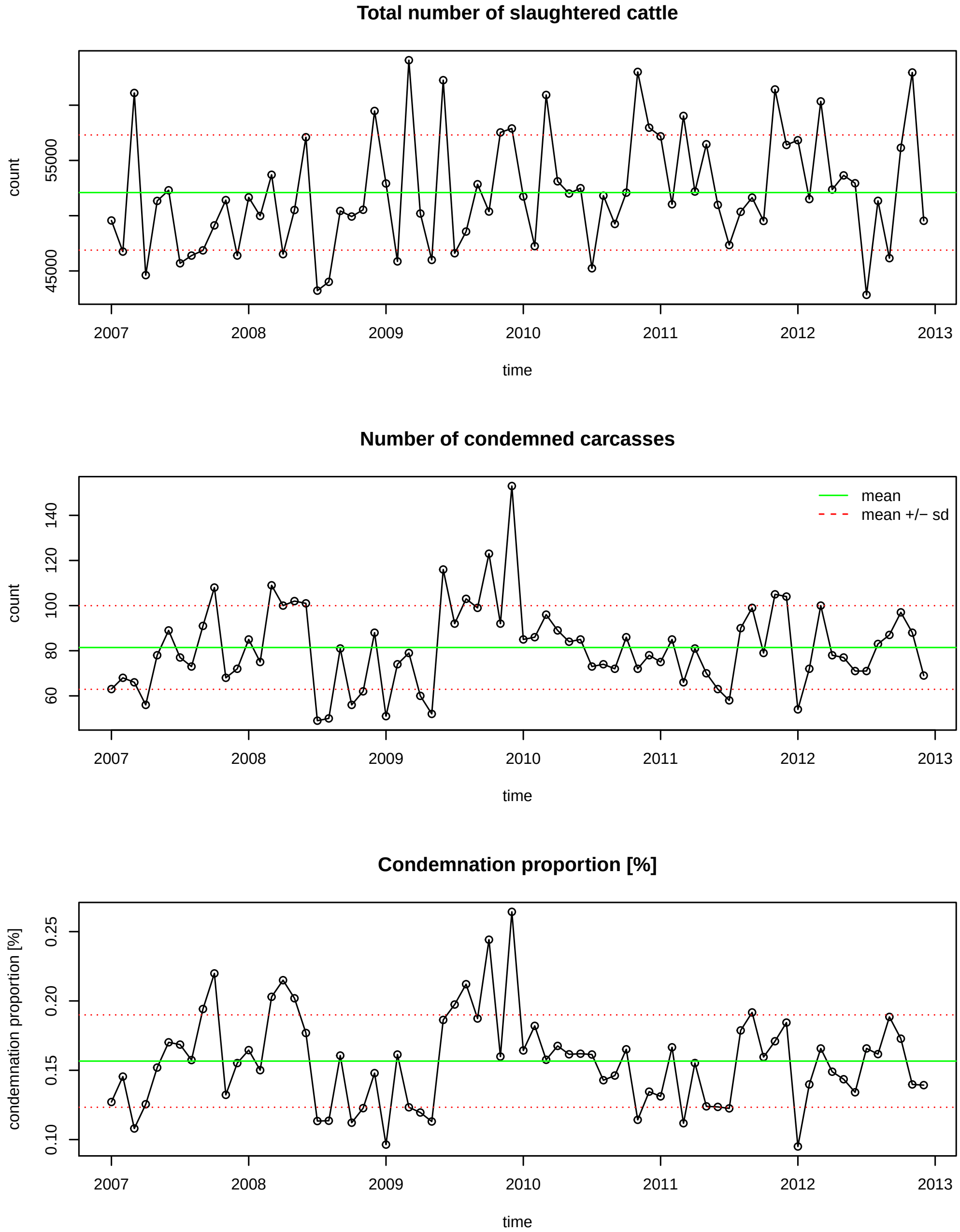


Figure 3: Time series data of normal slaughtered cattle

4.2 Descriptive Statistics: Emergency slaughtered Cattle

The mean number of emergency slaughtered cattle per month was 907 with a standard deviation of 181. The minimum number of cattle slaughtered as emergency slaughter within one month was 659, whereas the maximal number of cattle that arrived injured or sick at the slaughterhouse within one month was 1'398. In the upper panel of Figure 4 the numbers of emergency slaughtered cattle over time are visualized. The highest numbers observed clearly lay in the years 2007 to 2009. Not only the highest numbers but also very low numbers were observed within this time period. Thus the fluctuation is much higher in the first three years of the study. The BVL (Bundesamt für Lebensmittelsicherheit und Veterinärwesen) could not provide a satisfactory explanation for the change of stability in the data. Therefore it should be looked at as a reporting artefact.

The mean number of carcasses declared as wholly condemned was 185 with a standard deviation of 33. The minimal number of carcasses condemned in the emergency slaughter group within one month was 120 whereas the count of condemned carcasses reached its maximum at 257. These numbers correspond to a range in proportion between 14.17% and 28.5% of carcasses declared as wholly condemned. On average 20.78% of the carcasses from emergency slaughtered cattle were declared as wholly condemned after slaughter each month with a standard deviation of 3.72%. Thus the condemnation proportion for this group was noticeably high, as was already the percentage of emergency slaughters. Figure 4 shows the count data of emergency slaughtered cattle over time, as well as the number of carcasses condemned within the emergency slaughter cattle group and the corresponding condemnation proportions in percentage over time. The summary statistics are shown in Table 4.

	Slaughtered cattle	Condemned cattle	Proportion [%]
min	659	120	14.17
mean	907	185	20.78
median	850	186	20.32
max	1'398	257	28.50
sd	181	33	3.72

Table 4: Summary statistics of emergency slaughtered cattle

In all the three graphs of Figure 4 the variability of the time series data is high. Despite the fact that the fluctuation in the total number of emergency slaughters processed is changing between 2009 and 2010, the fluctuation in the other two data series is fairly stable. For the total number of emergency slaughtered cattle there is a strong general peak visible in March and December. These are the same peak months as observed for the normal slaughtered cattle. These peaks are visible in the second panel as well, that shows the number of condemned carcasses after slaughter but they are not reflected by the condemnation proportions. In the lowest panel, that shows the times series data of the condemnation proportion in percentage, there is a positive trend visible over time. This trend is in contrary to the negative trend observed in total numbers of slaughters in the upper panel and is thus also seen in the number of condemned carcasses (second panel). The trend does not last until the end of the acquisition period, the percentage of condemned carcasses decreases again in the middle of 2012. The most remarkable observations on the emergency slaughter group for cattle, is on one hand the change of the distribution of the total slaughters in 2009/2010 and on the other hand the positive trend for the condemnation proportion.

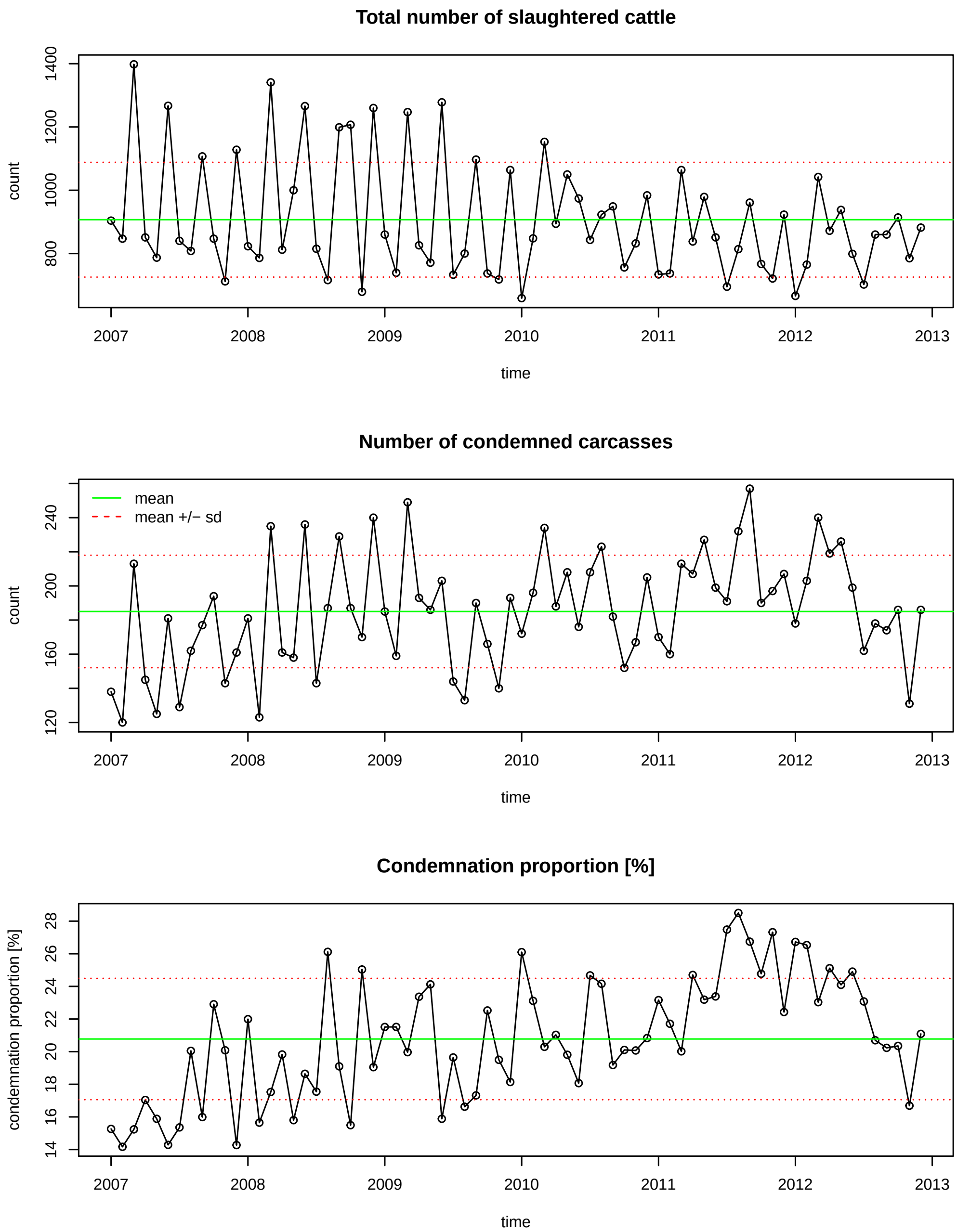


Figure 4: Time series data of emergency slaughtered cattle

4.3 Descriptive Statistics: Normal slaughtered Pigs

The number of animals slaughtered in the normal slaughtered pig group was by far the largest with a mean number of 229'411 normal slaughtered pigs per month. The lowest number of normal slaughtered pigs was observed in Mai 2008 with 186'822 slaughters. The maximal number of pigs was slaughtered in December 2010 (262'941 individuals). The amount of carcasses declared as wholly condemned after slaughter was reasonable low. On average 352 carcasses were declared as wholly condemned per month with a standard deviation of 70. The minimal (244 individuals) and maximal (496 individuals) numbers of condemned carcasses were observed in August 2008 and November 2012 respectively. The percentage of pig carcasses that were disallowed for food consumption after normal slaughter was thus between 0.11% and 0.22% with a mean of 0.15% and a standard deviation of 0.03%. The statistics are summarized in Table 5. All reported numbers of normal slaughtered pigs, as well as the amount of observed condemnations and the percentage of condemned carcasses are illustrated in Figure 5.

	Slaughtered pigs	Condemned pigs	Proportion [%]
min	186'822	244	0.11
mean	229'411	352	0.15
median	229'954	342	0.15
max	262'941	496	0.22
sd	16'858	70	0.03

Table 5: Summary statistics of normal slaughtered pigs

The amount of fluctuation and variability in the number of normal slaughtered pigs is similar to what was seen for cattle rather high. The number often exceeds or drops below the mean plus minus one standard deviation. The upper plot of Figure 5 illustrates the time series data of total slaughters. It does not show an overall trend or a noticeable seasonal pattern, contrary to the second graph in Figure 5 which shows the number of carcasses declared as condemned. There is a clear increase in the number of observed condemnations over time visible. It is not clear at which point in time this positive trend starts. One possibility is that the trend starts in December 2010. Another possibility could be a shift of the baseline number of condemned carcasses in 2010. The graph at the bottom of Figure 5 shows the condemnation proportion. The same increase as seen in the total number of condemnations can be observed again for the condemnation proportion. The trend or shift in 2010 could be a result of the unfavorable economic situation for the pig industry in the past 5 years. There had been an over-production of pigs in Switzerland which has resulted in low prices for pork and a campaign urging farmers to try to reduce the number of sows. If farmers are getting rid of less productive (injured or sick) individuals first, this may partly explain the rise in the number of condemnations in pigs sent to normal slaughter [26].

Two months show very high counts of condemned carcasses that could possibly be considered as outliers (February 2009 and February 2010). In addition there is some sort of seasonal pattern visible in the graph at the bottom of Figure 5. At the end of each year from November to December there is always an increase in the condemnation proportion. The increase of the percentage at the end of the year could simply be a result of the increase in number of animals with diseases in winter. Another reason might be the hypothesis that individuals of lower quality are sent for slaughter in winter, when hay and food are limited and a very high demand for pork is expected (due to Christmas) [26].

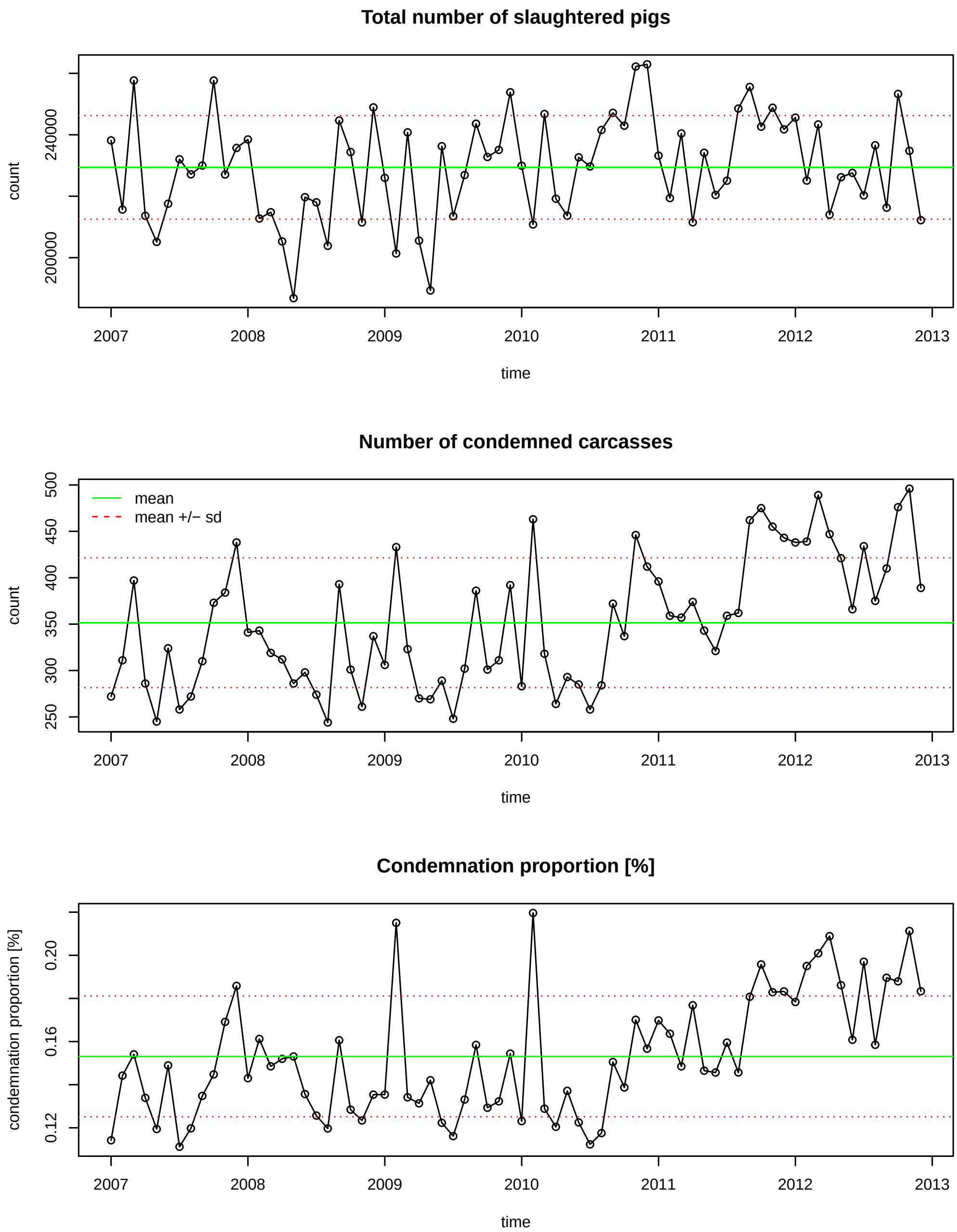


Figure 5: Time series data of normal slaughtered pigs

4.4 Descriptive Statistics: Emergency slaughtered Pigs

During the time of data acquisition the number of pigs that arrived sick or injured at the slaughterhouses was between 248 (observed in April 2007) and 911 (observed in January 2010). The numbers of emergency slaughters for pigs were fairly low (the mean number of normal slaughters was 490 times higher). The mean number of emergency slaughters per month was 468 with a standard deviation of 115. The monthly numbers of emergency slaughtered pigs are illustrated in the upper panel of Figure 6. The variability and fluctuation is as for all other data sets quite high. There is a clear peak visible in January 2010 in the total number of pigs slaughtered. Overall there is a positive trend visible in the data. This trend seems to be moderate in the time period before the peak. In general there is a lower number of emergency slaughtered pigs observed in July and increased numbers are seen in March and December.

The minimal (23 individuals) and maximal (75 individuals) numbers of condemned carcasses were observed in Mai 2009 and January 2008 respectively. The percentage of pig carcasses that were disallowed for food consumption after emergency slaughter was thus between 4.39% and 15.77%. The time series of the number of condemned carcasses (seen in the second panel of Figure 6) shows a similar trend pattern as in the upper panel. Again after 2010 there seems to be an increase in the trend. The fluctuation is higher in this time series, especially between 2007 and 2010. The peak which was observed in January 2010 is not reflected by the number of condemned carcasses. Similar to the upper panel, there is a general increase in the condemnation numbers seen in December.

	Slaughtered pigs	Condemned pigs	Proportion[%]
min	248	23	4.39
mean	468	46	9.96
median	454	44	9.79
max	911	75	15.77
sd	115	12	2.24

Table 6: Summary statistics of emergency slaughtered pigs

As to be expected and already seen in the cattle data set, also for pigs the condemnation proportion in the emergency group is much higher than in the normal slaughter group. With a mean percentage of 9.96%, the average condemnation proportion is 65 times higher than for normal slaughtered pigs (0.15%). This mean percentage value corresponds to an average of 46 carcasses that were wholly condemned with a standard deviation of 12. The panel at the bottom of Figure 6 shows the percentage of condemned carcasses. This plot contradicts the trend pattern observed in the panels above. There is in fact no trend visible at all for the condemnation proportion. Furthermore, the data is even less stable than the data in the upper panel. The peak seen in January 2010 is not reflected contrariwise the condemnation proportion even shows a trough in January 2010. Thus the number of animals that were grouped into the emergency slaughter group was exceptionally high in January 2010 but also an exceptional big part of the carcasses could be used for food consumption after slaughter (or at least parts of the carcasses, as only whole condemnations are reported).

The summary statistics is shown in Table 6. All reported numbers of emergency slaughtered pigs, the monthly numbers of observed condemnations and condemnation proportions are illustrated in Figure 6.

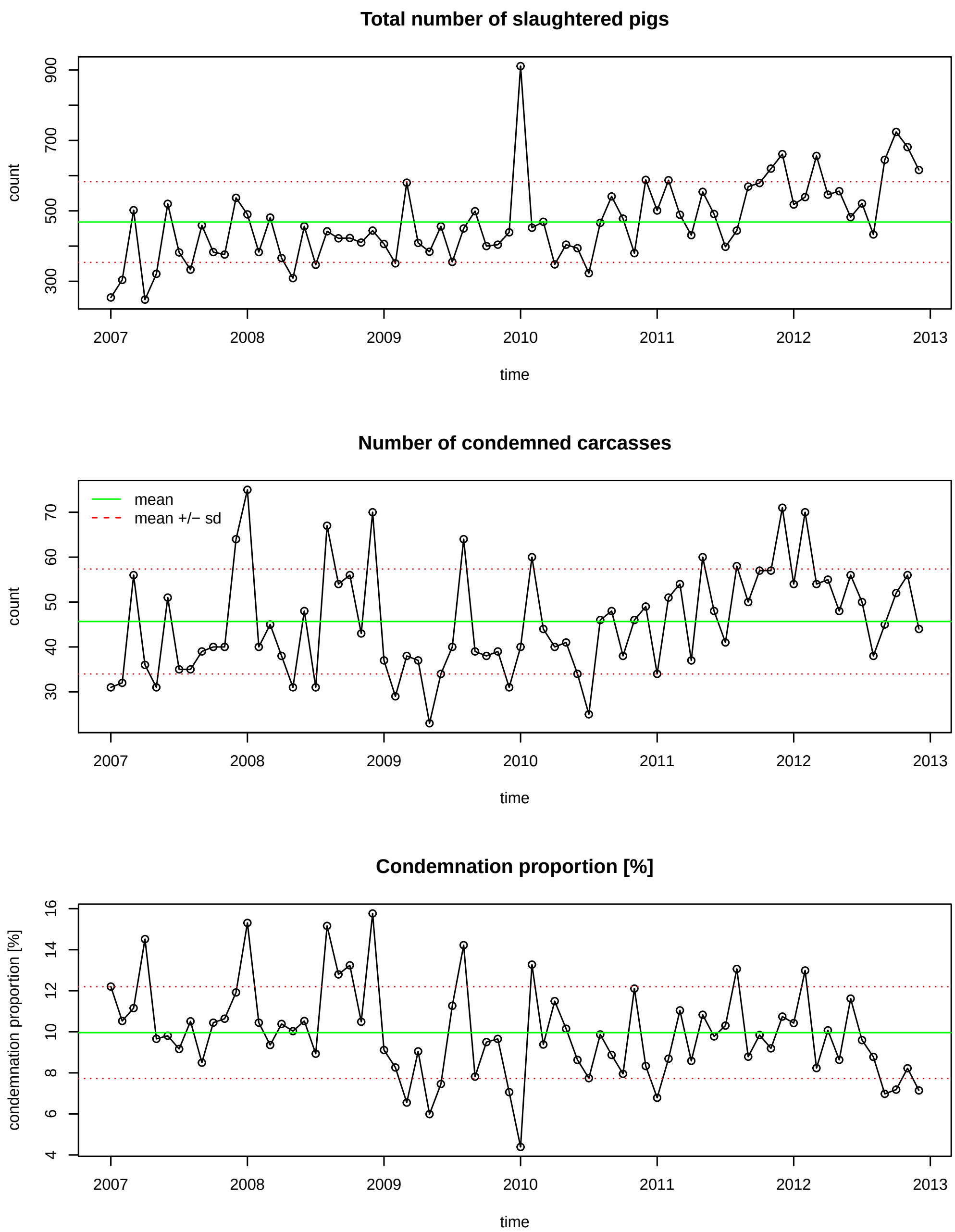


Figure 6: Time series data of emergency slaughtered pigs

4.5 Missing Data

As this FLEKO data set is a complete data set, there was no need to deal with methods for missing values.

5 Retrospective Analysis

5.1 Retrospective Analysis: Normal slaughtered Cattle

The decomposition of the transformed carcass condemnation proportion for normal slaughtered cattle (shown in Figure 7) did not give evidence for a linear trend or a strong seasonal pattern in the data. The variation in the transformed observed condemnation proportion (upper panel) was much higher compared to the minor information coming from the estimated trend and seasonal component. Additionally the estimated remaining part (lowest panel) illustrates that the variation can be attributed more to the remaining part than to the effect of season and trend (comparison of the grey bars on the right hand side of the plot). However a weak seasonal effect was modelled in autumn (peaking in October). Furthermore the seasonal impact that has been observed in the descriptive analysis at the end and in the beginning of each year can also be seen in the decomposition model. The estimated trend (third panel) is a non-monotone function, indicating that there was no long term trend influencing the data. It alternates annually between positive and negative trend phases.

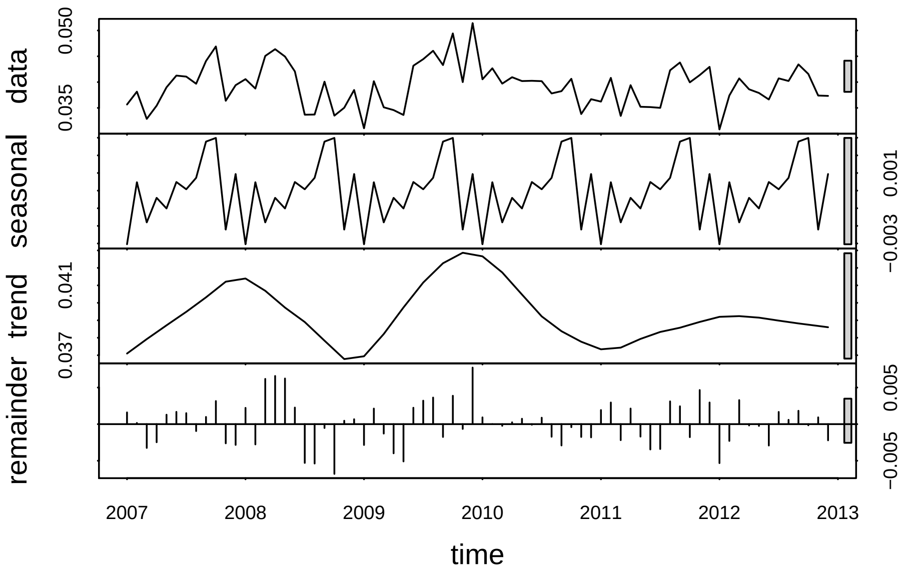


Figure 7: Season and trend decomposition of the transformed proportion of condemned carcasses for normally slaughtered cattle (based on LOESS)

Unsurprisingly the condemnation counts were modelled best with simple models (modelling done with method two (hhh4-models)). None of the top 5 models for normal slaughtered cattle (Table 7) was found to include a trend and a seasonal pattern at the same time. The fourth and fifth best models included a seasonal effect of December only (Christmas effect), that was included either in the endemic or in the autoregressive component and is consistent with what was observed in the descriptive part. The third best model included a log-linear trend starting in 2010 in the autoregressive component. The estimated seasonal or trend parameters were very small for those models, leading to very little effect on the predicted number of condemnation counts.

	trendAR	trendEND	seasonAR	seasonEND	df	BIC	lambda	overdisp.
1	t0	t0	s0	s0	3	608.35	0.26	0.028
2	c.e.	t0	c.e.	s0	2	610.59		0.031
3	t2010	t0	s0	s0	4	610.66	0.28	0.027
4	t0	t0	xmas	s0	4	610.68	0.26	0.027
5	t0	t0	s0	xmas	4	610.93	0.27	0.027

Table 7: Top five model fits for normal slaughtered cattle (t0=no trend, t2010=log-linear trend starting in 2010, c.e.=whole autoregressive component excluded, s0=no seasonality, xmas=effect of December only)

The best fit (with a BIC of 608.35) for normal slaughtered cattle was reached with a model including an offset (total number of animals slaughtered), a baseline condemnation rate and a constant autoregression parameter. This model did not include any trend or seasonal pattern and thus confirmed the result of method one (decomposition based on LOESS). Figure 8 shows the best model fit along with the observed data.

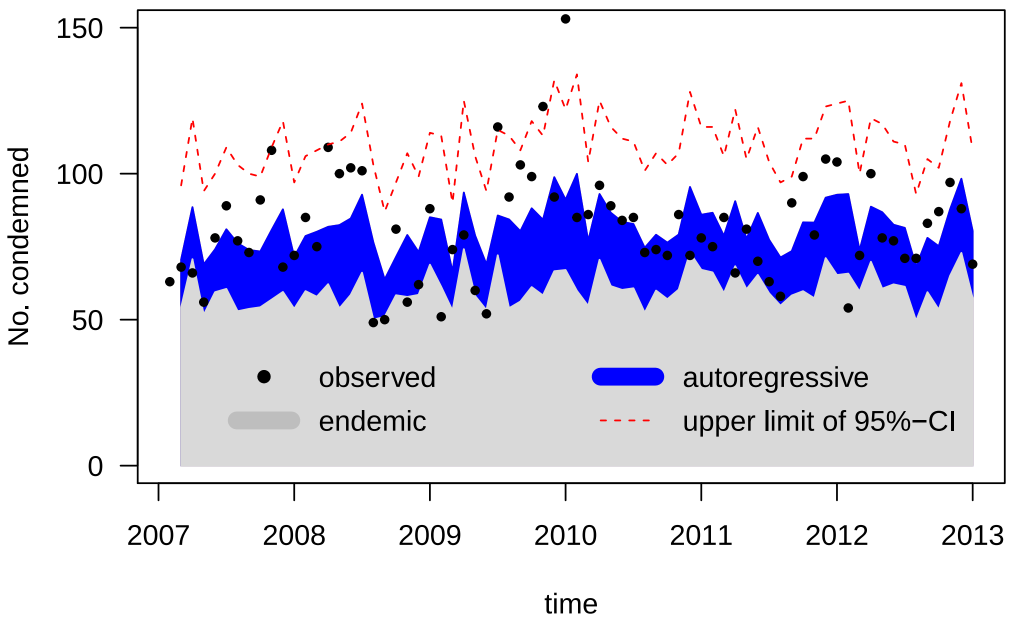


Figure 8: Best model fit according to the BIC for normal slaughtered cattle (model components endemic:t0, s0; autoregressive:t0, s0)

The estimated $\{\lambda\}$ for this model was 0.26. This means that about a quarter of the information used to calculate the expected number of condemned carcasses can be gained from the autoregressive component (previous month). It is clearly visible by comparing the blue and the grey components of the plot (grey part is approximately four times bigger than the blue part). For all the five top model fits, the estimated overdispersion parameters are very similar and rather small. It is about 0.028 for the best model. In the context of outbreak detection an upper limit of condemnation counts that can be expected to be observed without excessing an inconspicuous range is of great interest. On that account the upper limit of the 95%-confidence-interval was plotted in Figure 8 (evidentially the lower limit can be neglected for the purpose of outbreak detection).

5.2 Retrospective Analysis: Emergency slaughtered Cattle

Compared to normal slaughters for emergency slaughtered cattle the additive decomposition of the transformed proportion of condemned carcasses discloses different information about possible time related patterns. The observed positive trend (seen in the descriptive part) for emergency slaughtered cattle was reflected by the estimated trend in method one (third panel of Figure 9). Likewise the turning point of this positive trend into a negative trend in 2012 was estimated as seen in the descriptive part. Taking the grey bar at the right hand side of the plot into account, it can be concluded that the estimated trend contains approximately as much information as there is variation in the data. Hence other than for normal slaughtered cattle, for emergency slaughtered cattle it is reasonable to include a monotone trend to model the data (even though it deviates from the actual data at least for the last year observed).

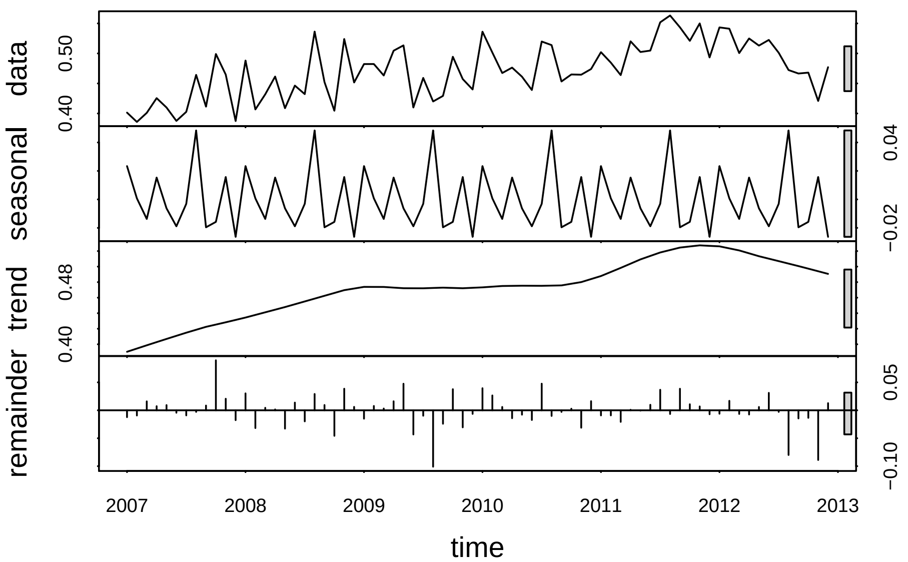


Figure 9: Season and trend decomposition of the transformed proportion of condemned carcasses for emergency slaughtered cattle (based on LOESS)

The estimated impact of a seasonal pattern does not contain as much information as the trend component. A weak seasonal pattern with the highest peak in August was estimated. Anyway, the estimated effect of seasonality is much smaller than the variation in the data and therefore it can be deduced that these data do not exhibit a seasonal pattern. Overall Figure 9 indicates that the trend and the random part dominate these data.

The result of method one (the inclusion of a trend and exclusion of a seasonal pattern) was consistent with the top five models found upon applying method two (Table 8). None of the best five models included a seasonal pattern but all of them included some kind of trend (either a shift in the intercept 2010 or a log-linear trend).

	trendAR	trendEND	seasonAR	seasonEND	df	BIC	lambda	overdisp.
1	t0	t1	s0	s0	4	659.16	0.28	0.009
2	j2010	t0	s0	s0	4	661.24	0.25	0.009
3	j2010	t1	s0	s0	5	661.69	0.25	0.009
4	j2010	j2010	s0	s0	4	661.76	0.29	0.01
5	t1	t1	s0	s0	5	662.59	0.24	0.009

Table 8: Top five model fits for emergency slaughtered cattle (t0=no trend, t1=log-linear trend, j2010=a shift in the intercept in January 2010, s0=no seasonality)

The model that best fits the data (BIC of 659.16) included a log-linear trend in the endemic part only. This model agrees well with what was observed in the descriptive part and in the analysis of method one. A shift in the intercept in 2010 would be less compatible with the estimated trend of method one (see third panel of Figure 9). The estimated autoregression parameter $\{\lambda\}$ for the best model was 0.28. A visualization of the best model fit is shown in Figure 10.

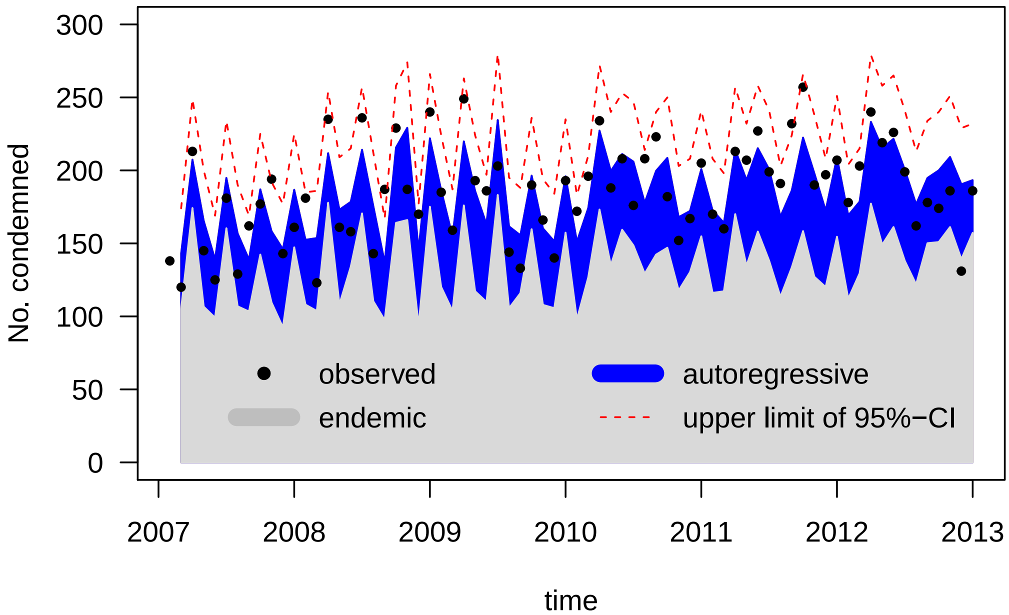


Figure 10: Best model fit according to the BIC for emergency slaughtered cattle (model components endemic:t1, s0; autoregressive:t0, s0)

The endemic part (shown in grey) contains the offsets and thus shows nicely the fluctuation seen in the total numbers of emergency slaughtered cattle (see descriptive part). On the other hand it can be seen, that the effect of the trend is very small compared to the offset (Figure 10). The best model fits the observed data quite well. The upper 95%-confidence-interval is fairly narrow compared to the one seen for the normal slaughtered cattle, due to the lower estimated overdispersion parameter (0.009).

5.3 Retrospective Analysis: Normal slaughtered Pigs

According to method one, the time series data of the normal slaughtered pigs might include some time trend but very likely no strong seasonality impacting the data. The grey bar on the trend panel is slightly larger than the one on the data panel, revealing that the trend signal is quite large relative to the variation in the data (Figure 11). The estimated seasonal component of the decomposition for the transformed proportion of condemned carcasses does not provide considerable information about the data according to Figure 11. However, the shape of this weak estimated seasonal pattern includes four harmonics with bigger effects in the period from September to February and lower impacts within the season from February to August. This pattern can be explained with the increase of the number of condemned carcasses which was observed from November to December in the descriptive part. The estimated trend strengthens what was already observed in the descriptive analysis. Until 2010 there was no (or maybe a very small non monotone) trend. After 2010 the trend is clearly positive and larger for the second part of the time series. It can be deduced that these data do not exhibit a seasonal pattern but include a trend (probably starting in 2010).

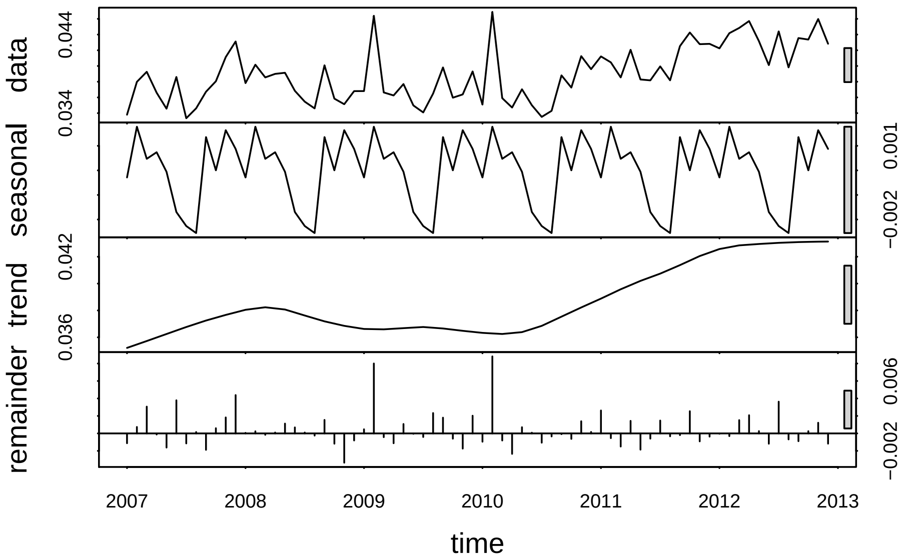


Figure 11: Season and trend decomposition of the transformed proportion of condemned carcasses for normally slaughtered pigs (based on LOESS)

The trend starting in 2010, that was observed in the descriptive statistics as well as with method one, reappeared again in the top five models of method two (Table 9). The five models with the lowest BICs all include a log-linear trend starting in 2010 in the endemic component. Despite the assumption of method one, that the data does not feature a seasonal pattern, all the five models include a seasonal component in the endemic part. But none of the estimated seasonal parameters in the endemic part lead to noticeable effects, as the estimated coefficients are all very small.

The best model fit (BIC of 749.44) was achieved with a model including a log-linear trend that starts in 2010 and a seasonal pattern with one harmonic in the endemic component. This model does not include an autoregressive component at all (consequently no $\{\lambda\}$ was estimated). As the data are time series data the exclusion of the whole autoregressive component was rather surprising. Although the data is assumed to be dependent on the previous observation, the best model fit resulted without including any information of the previous time point. The estimated overdispersion parameter of the best model is quite small (0.01) but still increases the model fit compared to the Poisson model (see Appendix).

	trendAR	trendEND	seasonAR	seasonEND	df	BIC	lambda	overdisp.
1	c.e.	t2010	c.e.	s1	5	749.44		0.011
2	t0	t2010	s0	s1	6	753.7	0	0.011
3	t2010	t2010	s2	s1	11	755.01	0	0.008
4	t0	t2010	s1	s1	8	755.49	0	0.01
5	c.e.	t2010	c.e.	s2	7	755.58		0.01

Table 9: Top five model fits for normal slaughtered pigs (t0=no trend, t2010=log-linear trend starting in 2010, c.e.=whole autoregressive component excluded, s0=no seasonality, s1=sine-cosine seasonality with 1 harmonic, s2=sine-cosine seasonality with 2 harmonics)

The visualization of the best fitting model (Figure 12) only contains the endemic component (and does not include the autoregressive component in blue). The estimated trend starting in 2010 as well as the seasonal harmonics can be seen. This very simple model, which does not even include autoregression, fits the data quite well. The upper limit of the 95%-confidence-interval (red dotted line) was fairly narrow. Also the two data points that were suspected to be outliers in the descriptive part (February 2009 and February 2010), lie close to the 95%-confidence-interval, which supports the assumption that the data are outbreak free.

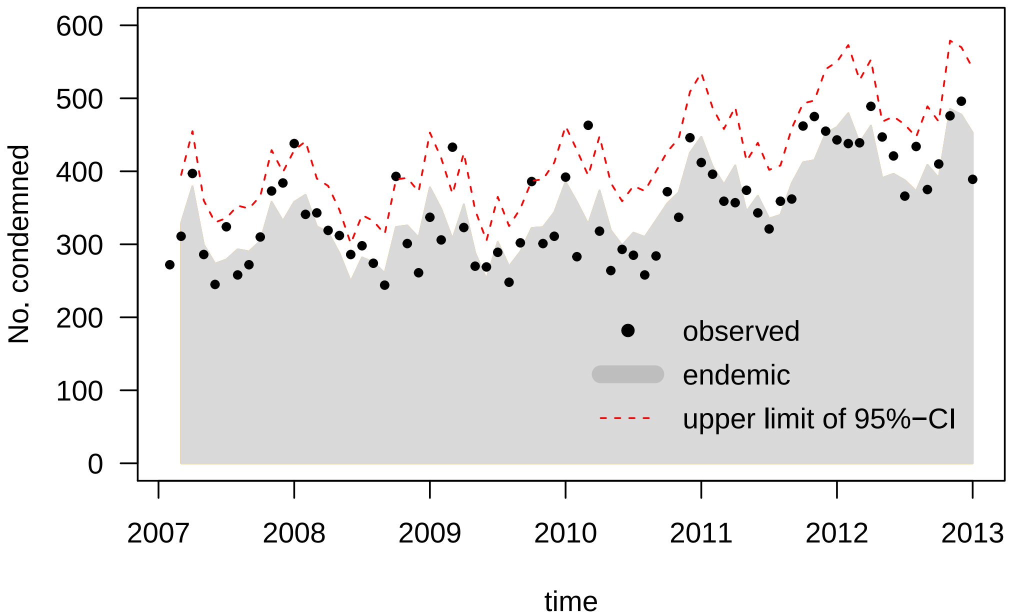


Figure 12: Best model fit according to the BIC for normal slaughtered pigs (model components endemic:t2010, s1; autoregressive:no autoregressive component)

5.4 Retrospective Analysis: Emergency slaughtered Pigs

The condemnation proportions of emergency slaughtered pigs are not affected by strong seasonal effects or a time trend. The smallest impact on the data was caused by a seasonal pattern (according to method one, second panel Figure 13). This is consistent with the lack of seasonality observed in the condemnation proportion in the descriptive part. However, this weak estimated pattern includes four harmonics with the highest peak in August. Slightly more information can be gained by the estimated trend (but again the information was not of dominating importance). Despite all previous observations of the other slaughter groups, the weak estimated trend is in general negative for emergency slaughtered pigs, indicating a slight decrease of the condemnation proportion for emergency slaughtered pigs over the last few years. This trend as well as any seasonality is not obvious and was thus not recognized in the descriptive part.

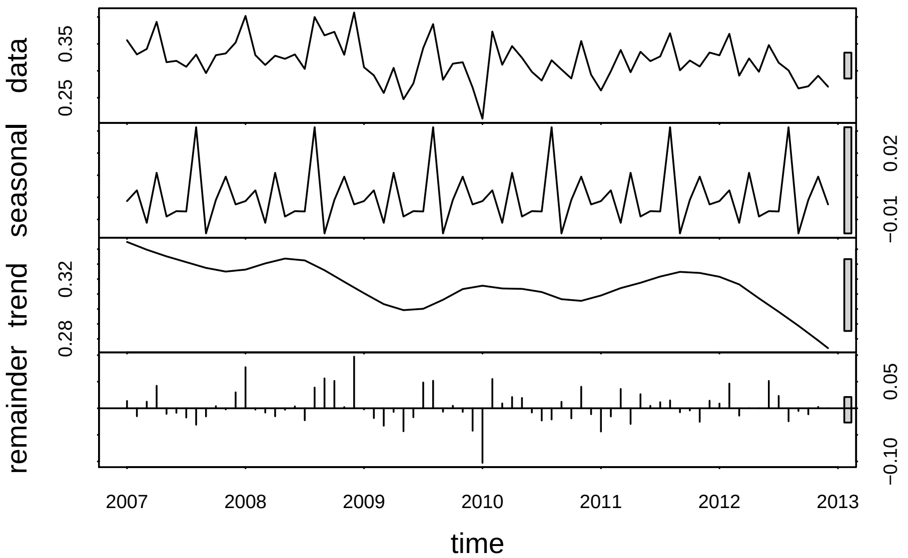


Figure 13: Season and trend decomposition of the transformed proportion of condemned carcasses for emergency slaughtered pigs (based on LOESS)

Since very simple models were found to best fit the other data sets, it is not surprising that among all the 840 different models, also for emergency slaughtered pigs the top five models appear to be very simple ones (Table 10). All of the five best fitting models do not include any seasonal pattern in the endemic and autoregressive component. The negative trend that is estimated with method one re-emerged in different forms among the top five models again.

	trendAR	trendEND	seasonAR	seasonEND	df	BIC	lambda	overdisp.
1	t0	t0	s0	s0	3	536.42	0.26	0.026
2	t0	t1	s0	s0	4	536.74	0.23	0.023
3	t1	t0	s0	s0	4	537.9	0.3	0.024
4	t0	j2010	s0	s0	4	537.97	0.24	0.024
5	t0	t2010	s0	s0	4	538.64	0.24	0.024

Table 10: Top five model fits for emergency slaughtered pigs (t0=no trend, t1=log-linear trend, t2010=log-linear trend starting in 2010, j2010=a shift in the intercept in January 2010, s0=no seasonality)

The model which best fits the data (BIC of 536.42) was the same model that was found to best fit the normal slaughtered cattle data. It includes neither a trend nor a seasonal pattern and thus only consists of the offset and a baseline condemnation rate in the endemic part and the autoregressive component. The ratio between the information coming from the previous data and the information coming from the endemic part is approximately 1 to 4 ($\{\lambda\}$ of 0.26). Besides the autoregressive component a big part of the information is contributed by the offset (thus the fluctuation that can be observed in the grey part of Figure 14 is mainly caused by the offset). The estimated overdispersion parameter for the best model fit was, same as for the other data sets, fairly low (0.03). Anyway it enhances the model fit compared to the Poisson model (see Appendix). The upper limit of the 95%-confidence-interval is, as for normal slaughtered cattle, wider than in the emergency cattle and normal pig data sets. However, also for emergency slaughtered pigs, the best model fits the data well.

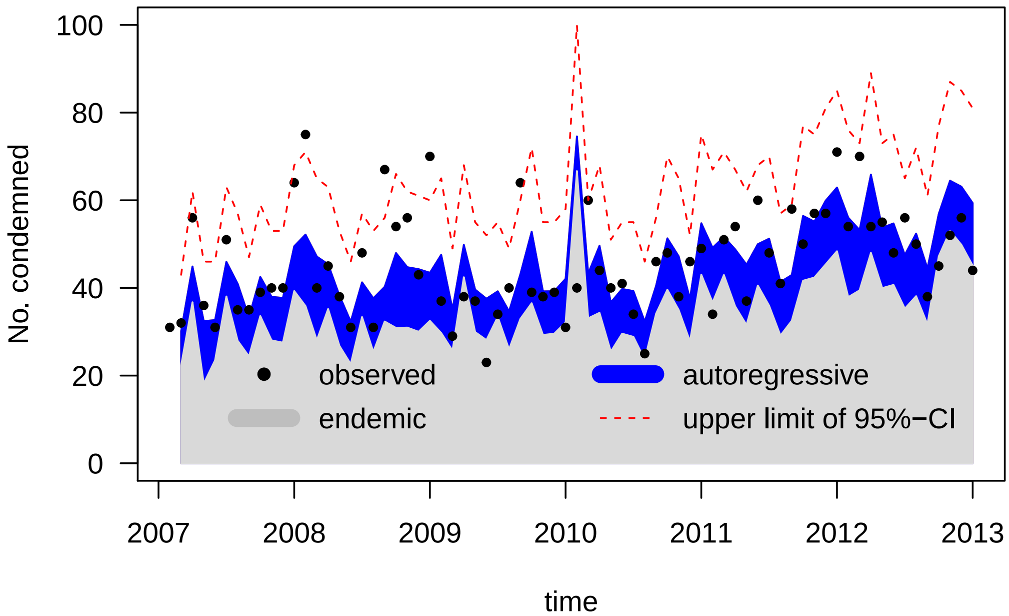


Figure 14: Best model fit according to the BIC for emergency slaughtered pigs (model components endemic:t0, s0; autoregressive:t0, s0)

The plots of the autocorrelation function of the residuals for all the datasets for method one and two are included in the Appendix. These plots show that for both methods used potential seasonal or trend patterns are disappeared in the residuals of the decompositions. Thus both methods captured the patterns well (this is not surprising as for most of the data sets there was no or only little evidence for a trend or a seasonal pattern). The majority of the sample correlations is particularly small and can therefore be attributed to random noise.

6 Prospective Analysis

For all data sets, the prospective analysis part includes the performance of the algorithm with parameters chosen according to the result of the retrospective analysis, as well as the results of the best performing algorithm in terms of POD for small outbreak sizes (k=2 to 3). The detection of small outbreaks was the main focus as big outbreaks can be detected with other sources as well. The outcomes of some of the other algorithms tested are shown in the Appendix as representative examples for all algorithms tested. In general with thresholds based on the 0.995 quantile (alpha=0.005 see Appendix) the decrease in POD for small outbreaks was unconvincing compared to the decrease in FPR. The inclusion of a trend did not improve performance of the algorithms for our data (in terms of POD). By including a trend the performance was either similar or worse. This is reasonable as the simulations did only include a very weak trend or no trend at all (see representative examples in the Appendix). The systematic exclusion of data to account for emerging outbreaks (by setting {pastWeeksNotIncluded} to 2) had no remarkable effect on the POD or FPR and the mean TTD and CUD were not shortened or decreased as well (see representative examples in the Appendix). Thus the impact of emerging outbreaks on the sensitivity and specificity can be neglected for the outbreak simulations used. This is why the performance of the algorithms with pastWeeksNotIncluded=0 are shown in the prospective results part. In general different settings of {w} and {b} only had little effect on algorithm performance. For w=1 or w=2 the performance of the algorithms was either worse or very similiar to w=6 in terms of POD, TTD and CUD. The FPR of algorithms with w=1 or w=2 were very similar or improved compared to w=6 (see representative examples in Appendix). The difference between w=0 and w=6 was more clear and will be discussed in detail in the next section part. By including a higher amount of historic data (setting noPeriods=11) while using w=0 the performance was not improved in terms of POD for small outbreak detection (see representative example for emergency slaughtered cattle in Appendix). In general for all algorithms the CUD proportion of the whole outbreak size but also the TTD as well as the FPR were fairly stable over k with a tendency to slightly decrease with increasing k. The POD clearly increased with increasing k for all the algorithms tested.

6.1 Prospective Analysis: Data Simulation

For all simulated outbreak data sets the mean outbreak duration increases with increasing scaling parameter k (from 2 to 10). The outbreak durations are very similar among all the four data sets (on average between 3.6 and 5.6 months). Differences in the outbreak sizes between the data sets are more obvious. These deviations are due to the specific estimated standard deviation used to calculate the outbreak sizes. Naturally, the outbreaks increase in size with increasing scaling parameter k. Figure 15 shows an example of the outbreak simulations for each of the four different groups and for different parameters k.

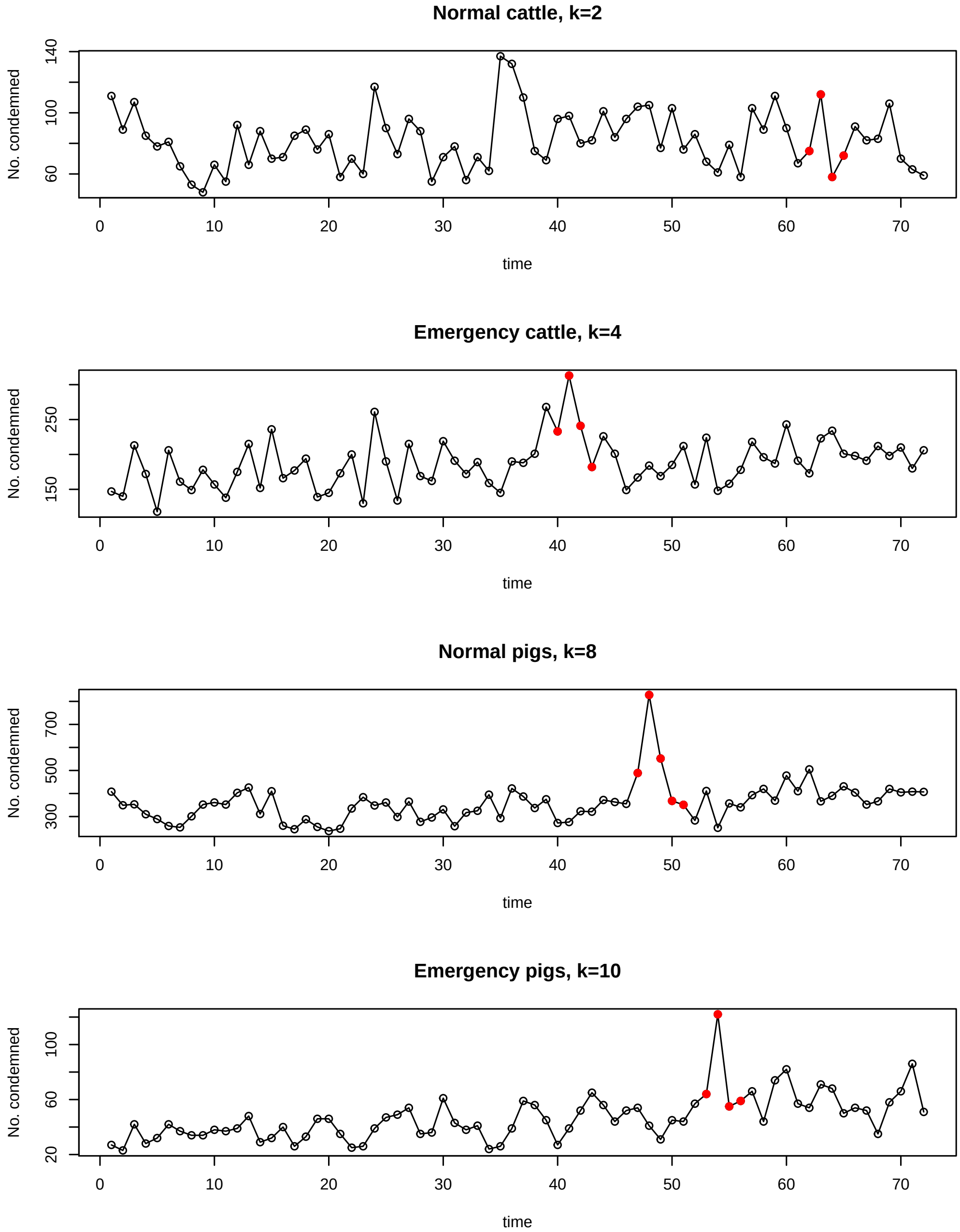


Figure 15: Illustrative example of outbreak data simulation for different parameters k and for the four different slaughter groups (number of condemnations during the outbreak are marked in red).

6.2 Prospective Analysis: Normal slaughtered Cattle

The simulations for normal slaughtered cattle resulted in outbreaks lasting 3.9 to 4.9 months on average. The mean outbreak size lie between 33.1 and 166.3 for the different scaling parameters (k=2 to 10). For outbreaks of 2 standard deviations (k=2), on average 33.1 additional cases were distributed to the 3.9 outbreak months (leading to an average of 8.5 additional cases per outbreak month). Taking into account the mean baseline number of condemned carcasses predicted by the best model ($\bar{\mu}$=81.71) and the estimated standard deviation of 16.33 this is a fairly low number of additional cases. Only by increasing the parameter k to 5 the mean number of additional cases per month (19) is above the estimated standard deviation. The outbreak size of the most extreme simulations (with k=10) is on average 166.3 with a mean outbreak duration of 4.9 months. On the other hand those outbreak sizes are perceptible high compared to the estimated mean and standard deviation of the baseline data. Mean outbreak sizes and durations for normal slaughtered cattle are summarized in Table 11.

Table 11 (a) and (b) show the outcome of two different outbreak detection algorithms used. Algorithm 1 (Table 11 (a)) includes parameters chosen according to the results of the retrospective analysis. Therefore it includes no trend and no seasonal components. Algorithm 2 was the algorithm with the best outbreak detection performance (in terms of POD) for low outbreak sizes. The resulting algorithm includes more of the historic data (b=3 instead of b=2). As a consequence the window size of the data used for the first step of the algorithm (the model fitting) had to be narrowed (w=0). Thus Algorithm 2 includes no trend but considers monthly seasonal effects (in an analogous manner as the original Farrington algorithm).
Both algorithms show low FPR values in the context of early outbreak detection. Whereas Algorithm 1 achieves slightly better FPRs (one positive alarm between every 9 and 17 years), the FPRs for Algorithm 2 are less compelling (one positive alarm between every 2.5 and 2.7 years). Both algorithms achieve FPRs which lead to less than one false positive alarm every year resulting in investigation costs which should be deemed acceptable. In terms of POD Algorithm 1 is again more convincing, at least in terms of the detection of larger outbreaks (k=5 to 10). But for small outbreak detection Algorithm 2 performs slightly better. Algorithm 1 only detects more than 30% of the outbreaks for k bigger than 4 whereas Algorithm 2 already detects more than 30% of all outbreaks if k equals 3 (Table 11).

k	Dur.	Size	TTD	CUD	POD	FPR
2	3.9	33.1	1.1	27.4	0.11	0.009
3	4.2	49.6	1.0	39.8	0.26	0.009
4	4.3	66.5	1.0	53.8	0.45	0.007
5	4.4	83.5	1.0	65.8	0.66	0.006
6	4.6	99.6	1.0	78.7	0.80	0.006
7	4.7	116.2	1.0	89.5	0.90	0.007
8	4.7	132.6	1.0	102.6	0.93	0.005
9	4.9	149.6	1.0	115.9	0.96	0.006
10	4.9	166.3	0.9	126.1	0.97	0.006

(a) Algorithm 1 - according to retrospective analysis (b=2, w=6, trend=FALSE, noPeriods=1)

k	Dur.	Size	TTD	CUD	POD	FPR
2	3.9	33.1	1.2	25.2	0.23	0.033
3	4.2	49.6	1.1	38.2	0.35	0.032
4	4.3	66.5	1.1	51.0	0.46	0.031
5	4.4	83.5	1.1	62.8	0.58	0.033
6	4.6	99.6	1.0	73.6	0.66	0.031
7	4.7	116.2	1.0	86.5	0.73	0.033
8	4.7	132.6	1.0	99.5	0.79	0.032
9	4.9	149.6	0.9	109.7	0.85	0.032
10	4.9	166.3	0.9	122.0	0.88	0.031

(b) Algorithm 2 - best POD for small outbreak detection (b=3, w=0, trend=FALSE, noPeriods=1)

Table 11: Outbreak detection performance for the simulated data sets of normal slaughtered cattle

The difference between the two algorithms in terms of TTD is very small. Algorithm 2 needs on average slightly more than 1 month to detect an outbreak for small outbreak sizes (k=2 to 5). Whereas Algorithm 1 detects outbreaks on average within the first month independent of the outbreak size. The CUD shows that the peak of the outbreaks is or was already reached at the time of detection. The number of outbreak cases which emerged until detection (CUD) is between 27.4 and 126.1 for both algorithms. Thus 73% to 83% of the outbreak cases already occurred before outbreak detection. The CUDs of Algorithm 2 are slightly smaller than the ones for Algorithm 1.

Over all the improvement for small outbreak detection from Algorithm 1 to Algorithm 2 is not striking. Its performance is only better for outbreaks smaller than 4 standard deviations and at the same time the FPR is more than 3 times higher than in Algorithm 1. Algorithm 1 outperforms Algorithm 2 in the detection of outbreaks the size of more than 4 standard deviations and reaches a POD of 97% as opposed to 88% for Algorithm 2. Based on the sample size included for threshold estimation, Algorithm 1 is doubtless the more reliable one as well.

6.3 Prospective Analysis: Emergency slaughtered Cattle

For emergency slaughtered cattle the mean of the estimated incident rates ($\bar{\mu}$=185.86) is more than twice as high as for normal slaughtered cattle and has a slightly higher estimated standard deviation of 22.27 as well. The mean simulated outbreak sizes and durations for emergency slaughtered cattle resulted to be very similar to the ones obtained for normal slaughtered cattle (compare Table 11 and 12). To add on average more cases than one estimated standard deviation to each month of the outbreak, a scaling parameter k of at least 5 was needed. For the most extreme outbreak scenario simulated (k=10) the mean outbreak size is about (230.8). With a mean outbreak duration of 5.1 this makes on average 45.25 additional cases per outbreak month. This corresponds to adding on average 24% of the mean incidence rate of the best model ($\bar{\mu}$=185.86) on top of the baseline counts for each outbreak months.

The parameters which are compatible with the best fitting model of the retrospective analysis (no seasonality but a trend included) were used for Algorithm 1. The improved algorithm for emergency slaughtered cattle is the equivalent to Algorithm 2 for normal slaughtered cattle (more historic data is included with b=3, a smaller window size is used with w=0 and no trend is included) and is thus again very similar to the original Farrington algorithm. Table 12 (a) and (b) list the performances of the Algorithms 1 and 2 for emergency slaughtered cattle.

The difference between the two algorithms for emergency slaughtered cattle is more pronounced to what was seen for the normal slaughtered cattle. Whereas Algorithm 1 detects 20% of the smallest outbreaks (k=2), Algorithm 2 is able to detect 46%. Algorithm 1 is only able to detect more than 40% of the outbreaks of size bigger than 91.4 (k=4). Both algorithms are convenient for large outbreak detection with maximal PODs of 0.96 (Algorithm 1) and 0.92 (Algorithm 2). For outbreaks of the size bigger than 5 standard deviations, Algorithm 1 is outperforming Algorithm 2 in terms of POD. The difference of the two Algorithms in terms of FPR is remarkable. Again Algorithm 1 is more convenient. The FPR for Algorithm 2 is on average 9 times higher than the one for Algorithm 1. However, Algorithm 2 might still be acceptable with an FPR between 8.4% and 9% for data that is reported in a monthly manner. This would lead to maximal one false positive alarm every 11 months. The TTD for the two algorithms are very similar but differences in the CUDs are observed. On average Algorithm 2 detects the small outbreaks (k=2 and 3) a bit later, but both algorithms detect outbreaks on average within 1 month after outbreak start. Both algorithms detect outbreaks after 64% to 80% of the outbreak cases occurred. Algorithm 2 is slightly better in terms of CUD (CUD for Algorithm 1 are between 37.2 and 171.9 and between 31.4 and 148.6 for Algorithm 2).

k	Dur.	Size	TTD	CUD	POD	FPR
2	4.1	46.4	1.0	37.2	0.20	0.013
3	4.4	69.6	1.0	55.2	0.35	0.013
4	4.5	91.4	1.0	72.1	0.51	0.011
5	4.7	114.9	1.0	89.7	0.72	0.008
6	4.8	136.6	1.0	106.5	0.83	0.011
7	4.9	161.7	1.0	125.3	0.90	0.009
8	5.0	185.0	0.9	139.9	0.95	0.009
9	5.1	207.1	0.9	154.0	0.96	0.008
10	5.1	230.8	0.9	171.9	0.96	0.008

(a) Algorithm 1 - according to retrospective analysis (b=2, w=6, trend=TRUE, noPeriods=1)

k	Dur.	Size	TTD	CUD	POD	FPR
2	4.1	46.4	1.1	31.4	0.46	0.087
3	4.4	69.6	1.1	48.4	0.56	0.089
4	4.5	91.4	1.0	60.6	0.64	0.088
5	4.7	114.9	1.0	79.3	0.73	0.086
6	4.8	136.6	0.9	93.7	0.80	0.086
7	4.9	161.7	0.9	109.3	0.84	0.084
8	5.0	185.0	0.9	123.0	0.88	0.090
9	5.1	207.1	0.8	135.6	0.91	0.088
10	5.1	230.8	0.8	148.6	0.92	0.085

(b) Algorithm 2 - best POD for small outbreak detection (b=3, w=0, trend=FALSE, noPeriods=1)

Table 12: Outbreak detection performance for the simulated data sets of emergency slaughtered cattle

6.4 Prospective Analysis: Normal slaughtered Pigs

For normal slaughtered pigs higher numbers of condemned carcasses were observed (see descriptive statistics). The estimated standard deviation of the best model (41.42) is higher for normal slaughtered pigs compared to the other groups. Consequently the simulated outbreaks are larger in size (more than double the size of for normal and emergency slaughtered cattle). Furthermore the outbreak duration is slightly longer for normal slaughtered pigs (Table 13). The mean outbreak duration is between 4.5 and 5.6 months and includes an average of between 88.4 and 444.4 additional cases for the different parameters k. These numbers are high compared to the outbreak sizes simulated for the other data sets. Still relative to the mean predicted baseline counts ($\bar{\mu}$=353.09), the numbers are alike to what was simulated for the other data sets. For the most extreme outbreak situation (k=10), the average of cases added to each outbreak month, are approximately $^1/_4$ of the mean baseline counts ($\bar{\mu}$) (which is the same than what was seen for emergency slaughtered cattle).

For normal slaughtered pigs the best fit in the retrospective part was found for a model which includes a seasonal pattern. As there are two ways used to include seasonality for outbreak detection (either according to the original or the improved Farrington algorithm) there are two Algorithms 1 for normal slaughtered pigs (1a and 1b). Table 13 (a) shows the result of the first variation of the algorithm used (Algorithm 1a). This version considers seasonality in an analogous manner compared to the original of Farrington algorithm. Algorithm 1a includes a trend, the biggest possible amount of historic data (b=3) and a narrow window size around the current observation (w=1). In terms of POD the outbreak detection is similar to what was observed for the cattle data sets. Algorithm 1a detects between 18.1% and 91.3% of the outbreaks for k form 2 to 10. The FPR of Algorithm 1a is convincing for monthly data (between 0.021 and 0.024) and would lead to up to one false positive alarm every 3.5 years.

The TTD of Algorithm 1a for normal slaughtered pigs is similar to what was observed for the cattle data sets and thus convenient. Again one drawback of the algorithm is the CUD. Between 73% and 75% of the total outbreak cases emerged before the time of detection.

k	Dur.	Size	TTD	CUD	POD	FPR
2	4.5	88.4	1.3	64.2	0.18	0.024
3	4.8	133.2	1.1	98.4	0.29	0.021
4	5.0	176.9	1.1	131.4	0.38	0.023
5	5.1	221.8	1.0	163.0	0.51	0.023
6	5.3	267.0	1.0	201.0	0.62	0.022
7	5.3	311.0	1.0	228.8	0.75	0.021
8	5.4	355.8	0.9	265.8	0.84	0.022
9	5.5	397.8	0.9	292.7	0.90	0.022
10	5.6	444.4	0.9	327.2	0.91	0.023

(a) Algorithm 1a - according to retrospective analysis (b=3, w=1, trend=TRUE, noPeriods=1)

k	Dur.	Size	TTD	CUD	POD	FPR
2	4.5	88.4	1.1	69.2	0.10	0.007
3	4.8	133.2	1.1	104.5	0.21	0.006
4	5.0	176.9	1.0	136.8	0.37	0.007
5	5.1	221.8	1.0	171.8	0.57	0.007
6	5.3	267.0	1.0	208.7	0.71	0.006
7	5.3	311.0	1.0	240.5	0.83	0.005
8	5.4	355.8	1.0	274.1	0.91	0.005
9	5.5	397.8	1.0	302.8	0.95	0.005
10	5.6	444.4	1.0	338.8	0.94	0.005

(b) Algorithm 1b - according to retrospective analysis (b=3, w=1, trend=TRUE, noPeriods=3)

k	Dur.	Size	TTD	CUD	POD	FPR
2	4.5	88.4	1.2	58.1	0.65	0.140
3	4.8	133.2	1.1	89.1	0.75	0.135
4	5.0	176.9	1.0	116.0	0.81	0.138
5	5.1	221.8	0.9	137.4	0.87	0.137
6	5.3	267.0	0.9	166.7	0.90	0.134
7	5.3	311.0	0.8	187.8	0.92	0.133
8	5.4	355.8	0.8	218.4	0.95	0.134
9	5.5	397.8	0.8	236.5	0.97	0.134
10	5.6	444.4	0.7	256.1	0.96	0.133

(c) Algorithm 2 - season only
Best POD for small outbreak detection
(b=3, w=0, trend=FALSE, noPeriods=1)

k	Dur.	Size	TTD	CUD	POD	FPR
2	4.5	88.4	1.3	67.0	0.33	0.040
3	4.8	133.2	1.1	100.1	0.49	0.037
4	5.0	176.9	1.0	129.7	0.62	0.036
5	5.1	221.8	0.9	162.7	0.74	0.031
6	5.3	267.0	0.9	192.3	0.84	0.028
7	5.3	311.0	0.9	223.1	0.91	0.026
8	5.4	355.8	0.9	250.2	0.96	0.022
9	5.5	397.8	0.9	273.3	0.97	0.021
10	5.6	444.4	0.9	301.8	0.96	0.020

(d) Algorithm 3 - trend only
Increased POD for small outbreak detection
(b=2, w=6, trend=TRUE, noPeriods=1)

Table 13: Outbreak detection performance for the simulated data sets of normal slaughtered pigs

The second variation of the algorithm used (Algorithm 1b) takes advantage of the parameter {noPeriods} of {farringtonFlexible}. The seasonality found in the retrospective analysis (a function with 2 harmonics per year) is taken into account by setting {w}=1 and {noPeriods}=3. The performance of this outbreak detection variation is summarized in Table 13 (b). The POD of Algorithm 1b is very good for large outbreaks. More than 50% of the outbreaks are detected if the outbreaks are of size 221.8 or bigger ($k \geq 5$). For outbreaks of 2 to 4 standard deviations, the performance is not as convenient in terms of POD (only 10% to 37% are detected for k=2 to 4). In terms of FPR Algorithm 1b is very satisfactory with a maximal FPR of 0.007. The TTD of Algorithm 1b is comparable with what was obtained for Algorithm 1a and the other data sets (the outbreaks are detected within one month on average). The performance of Algorithm 1b is comparable to Algorithm 1a in terms of CUD. On average Algorithm 1b detects outbreaks after 77% of the total outbreak cases occurred for Algorithm 1a it is on average after 74%.

The performance of outbreak detection for the simulated times series of normal slaughtered pigs, can be increased in terms of POD by including either only a seasonal or only a trend pattern. These algorithm variations are shown as Algorithm 2 and 3 respectively (Table 13 (c) and (d)). Algorithm 2 includes no trend, the biggest possible amount of historic data (b=3) and a narrow window size around the current observation (w=0). In terms of POD this outbreak detection is much more efficient than what was achieved so far. Algorithm 2 already detects more than 60% of the outbreaks for the smallest outbreak size simulated (k=2). The TTD of Algorithm 2 for normal slaughtered pigs is convenient and is even slightly better for larger outbreaks than what was observed for Algorithm 1a and 1b. One drawback of the algorithm is again the large CUD even though it is already improved compared to what was seen for the other algorithms. Between 58% and 67% of the total outbreak cases emerged before detection. Another drawback of Algorithm 2 is the FPR, which is remarkably high (between 0.133 and 0.14). Thus Algorithm 2 leads to up to 1.7 false positive alarms each year. This algorithm variation includes again only very little data for threshold estimation and is therefore less reliable than other algorithms. Nevertheless it is interesting that similar to what was seen previously for the cattle data sets, the performance of this algorithm is the best in terms of POD for small outbreak detection.

However, there is a fourth mentionable Algorithm (Algorithm 3 see Table 12 (d)) that can increase the POD and at the same time leads to more convenient FPRs. This rather simple algorithm includes a trend but no seasonality and thus the estimated threshold is based on a bigger sample. The algorithm is able to detect between 33.4% and 97.4% of outbreaks (for k=2 to 10). With Algorithm 3 a mean FPR of 0.029 is achieved where the mean FPR in Algorithm 2 was 0.135 (Tables 13 (c) and (d)). The improvement in terms of FPR is not striking, but at least the expected emergence of investigations are decreased from one every 7 months (Algorithm 2) to one every 25 months (Algorithm 3). The FPR of Algorithm 3 decreases with increasing outbreak size, indicating that in this case the specificity of the algorithm is influenced by emerging diseases. The TTD for Algorithm 3 (between 0.9 and 1.3) are comparable to the outcomes of the other algorithms and are not dramatically long for outbreaks that last between 4.5 and 5.6 months. The CUD for Algorithm 3 is comparable to what was seen for Algorithms 1a and 1b. Algorithm 3 detects an outbreak on average only after 72% of the total outbreaks have already occurred.

6.5 Prospective Analysis: Emergency slaughtered Pigs

The simulated time series for emergency slaughtered pigs are similar to the ones obtained for cattle. In contrast to the normal slaughtered pigs (with the highest outbreak sizes and durations simulated), for emergency slaughtered pigs the lowest and shortest outbreaks were obtained. This is due to the estimated standard deviation of 10.03 and the estimated mean incidence rate ($\bar{\mu}$=46.15) used to simulate the data. The resulting time series contain outbreaks of an average size of 20.6 to 102.7 (for k=2 to 10). The outbreak durations are on average between 3.6 and 4.6 months. To exceed the estimated standard deviation of the best model (10.03) with the mean number of cases added to each month of the outbreak, parameter k needs to be at least 4. The mean of the estimated outbreak durations and sizes for all parameters k are listed in Table 14.

In the retrospective analysis the best model for emergency slaughtered pigs was found to be the same as for normal slaughtered cattle. Therefore the same algorithms were tested for the outbreak detection of those two groups. Table 14 (a) and (b) show the outbreak detection performance of these two algorithms with emergency slaughtered pig outbreak simulations. Algorithm 1 (the very simple algorithm without seasonal and trend pattern) is convincing in terms of POD for larger outbreaks (with k=7 to 10). It is able to detect at least 85% of the simulated outbreaks of the size 72 and bigger ($k \geq 7$). The performance of this algorithm is comparable with the performance of Algorithm 1 for normal slaughtered cattle.

Algorithm 2 (which includes a longer learning period and considers seasonality but no trend) reaches slightly higher PODs for smaller outbreaks (with k=1 to 2) compared to Algorithm 1. This is what was seen for normal slaughtered cattle as well. Both algorithms are able to detect more than 50% of the outbreaks for outbreak sizes bigger than 51.2 (k=5).

k	Dur.	Size	TTD	CUD	POD	FPR
2	3.6	20.6	1.0	17.8	0.12	0.010
3	3.9	31.3	1.0	26.9	0.28	0.009
4	4.0	41.1	1.0	33.8	0.42	0.007
5	4.2	51.2	1.0	41.9	0.60	0.007
6	4.3	61.9	1.0	49.3	0.74	0.006
7	4.4	72.0	1.0	56.1	0.85	0.006
8	4.5	82.9	1.0	64.9	0.89	0.007
9	4.5	92.7	1.0	71.1	0.92	0.006
10	4.6	102.7	1.0	78.0	0.96	0.005

(a) Algorithm 1 - according to retrospective analysis (b=2, w=6, trend=FALSE, noPeriods=1)

k	Dur.	Size	TTD	CUD	POD	FPR
2	3.6	20.6	1.2	16.9	0.19	0.027
3	3.9	31.3	1.1	25.3	0.31	0.025
4	4.0	41.1	1.1	31.9	0.38	0.026
5	4.2	51.2	1.0	40.4	0.52	0.028
6	4.3	61.9	1.0	45.4	0.61	0.026
7	4.4	72.0	1.0	53.1	0.70	0.026
8	4.5	82.9	1.0	62.8	0.78	0.024
9	4.5	92.7	1.0	69.4	0.80	0.025
10	4.6	102.7	0.9	75.7	0.85	0.024

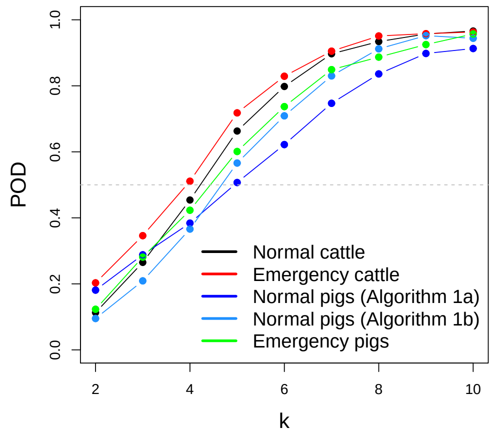
(b) Algorithm 2 - best POD for small outbreak detection (b=3, w=0, trend=FALSE, noPeriods=1)

Table 14: Outbreak detection performance for the simulated data sets of emergency slaughtered pigs

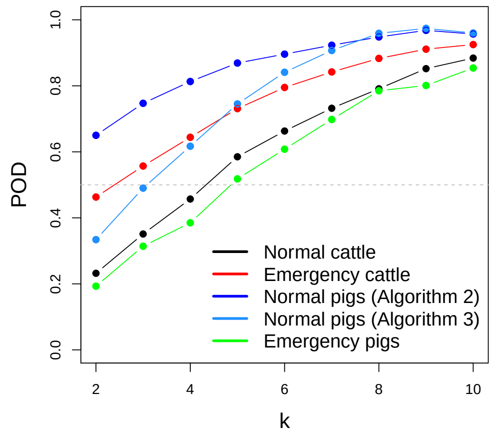
In terms of FPR, both algorithms are convenient. Even though the FPR of Algorithm 2 is on average 3.6 times higher compared to the one of Algorithm 1. With a mean FPR of 0.026 it would still lead to only one false positive alarm every 3 years. Algorithm 1 performs slightly better in terms of TTD as well. However with TTDs between 0.9 and 1.2 both algorithms can detect outbreaks within a similar time period as seen for the other groups. Also the CUDs of the two outbreaks lie in a similar range. Both algorithms detect outbreaks only after the majority of the outbreak cases have already occurred (73% to 86%). Algorithm 2 is doing slightly better in terms of CUD (81% on average for Algorithm 1 and 77% on average for Algorithm 2).

6.6 Prospective Analysis: Summary and Conclusion

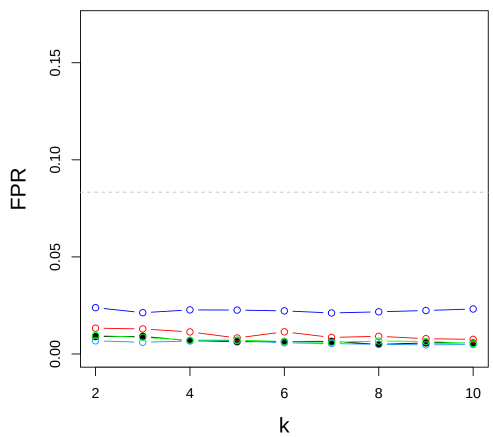
Figure 16 summarizes the performance of the algorithms for all the simulated data sets. The PODs achieved with the algorithms according to the retrospective analysis (Figure 16 (a)) are not prominent for small to medium outbreak sizes. Only outbreaks with high magnitudes (k=8 to 10) are detected reliably (POD≥80%). To detect more than 50% of the outbreaks with Algorithms 1, the number of added cases need to be more than 102%, 49% and 111% of the mean monthly baseline counts (for normal and emergency slaughtered cattle and emergency slaughtered pigs respectively). For normal slaughtered pigs the number of added cases needs to be more than 63% of the mean monthly baseline counts to detect at least 50% of the outbreaks with Algorithm 1a and more than 63% of the mean monthly baseline counts with Algorithm 1b.



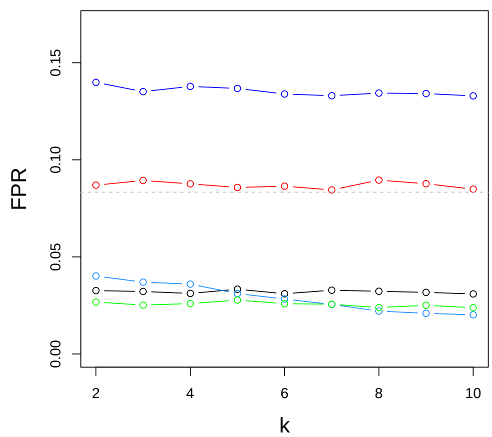
(a) Algorithm 1 - according to retrospective analysis



(b) Algorithm 2&3 - adjusted Algorithms



(c) Algorithm 1 - according to retrospective analysis



(d) Algorithm 2&3 - adjusted Algorithms

Figure 16: Outbreak detection performance for all detection algorithms and data simulations. The line in grey marks the FPR that corresponds to one false positive alarm per year and a POD of 0.5 respectively. The adjusted algorithms are the algorithms with the best PODs for small outbreak detection or in case of Algorithm 3 for normal slaughtered pigs with increased PODs compared to Algorithms 1.

The FPRs for the algorithms according to the retrospective analysis (shown in Figure 16 (c)) are far below the rate which corresponds to one false positive alarm per year (dotted grey line in the plot). For all data sets the FPRs are between 0.005 and 0.024, corresponding to one investigation every 3 to 18 years.

By modifying the outbreak detection algorithm the PODs for the small outbreaks can be improved for all data sets. Algorithms 2 and 3 achieve PODs (for k=2 or 3) which are 1.6 to 6.8 times as high as the ones for Algorithms 1. On the other hand the improved algorithms decrease the PODs for outbreaks with k larger than 6 except for the algorithms used for normal slaughered pigs. Compared to Algorithm 1a and 1b the two modified algorithms used for the normal slaughtered pigs (shown in blue in Figure 16 (d)), increase the outbreak detection proportion for all the outbreak sizes.

For all data sets the algorithms which are improving the POD for small outbreaks lead to an increase of the FPR at the same time. In some cases this increase is not dramatic. It shortens the mean expected time between two false positive alarms by several years but the resulting FPR is still acceptable. It reduces the gap from 12.4 to 2.6 years for normal slaughtered cattle and for emergency slaughtered pigs and from 4 years (Algorithm 1a) or 14.5 years (Algorithm 1b) to 2.9 years (Algorithm 3) for normal slaughtered pigs. The increase in FPR is more severe for emergency slaughtered cattle and for Algorithm 2 for normal slaughtered pigs. For emergency slaughtered cattle one false alarm can be expected every 8 years with Algorithm 1 and with Algorithm 2 once every 11 months (which is just at the edge of acceptance). The highest number of false positive alarms (on average) are expected with Algorithm 2 for normal slaughtered pigs (1 alarm every 7 months). All the outcomes (range of POD and FPR) for the simulated outbreak sizes used are comparable to what was observed in a similar study by Noufaily et al. [28].

7 Discussion

7.1 Discussion: Retrospective Analysis

The retrospective analysis with the two different methods used provides sufficient information about the data distribution of the different animal and slaughter groups. The dependence of the data on time related factors (trend, seasonality and autoregression) can be evaluated nicely by these methods. The purposes for which the two different methods were used each suited the advantages and disadvantages of both and the results gained revealed relevant information.
The first method, the additive decomposition based on LOESS, was convenient to get an idea of the seasonal and trend patterns that should be used in the second method. The main advantage of method one (compared to method two) is its robustness. The biggest disadvantage and the one reason why method two was needed, is the lack of an easily accessible regression function in method one. It does not produce a regression function which can be represented by a mathematical formula. Thus to simulate outbreak data time series or to predict expected condemnation counts, method one is inappropriate. These requirements are met by method two. Applying the {hhh4} function results in a regression function that describes the data in a convenient way. By using the estimated parameters of this function, baseline time series of the condemnation rates can be simulated. However, to be able to use method two efficiently, an idea about the seasonal and trend pattern that could influence the data are essential. Another disadvantage is the fact, that the models can quickly become complex (depending on the combinations of seasonal and trend patterns used) and several parameters possibly have to be estimated. If long term data is not available and at the same time the reporting frequency is low, it can lead to unreliable and thus useless estimates with large standard errors. In this project the time series ranged over 6 years and the data were reported in a monthly manner, which is why the sample size was rather small. E.g. for estimates of monthly seasonal effects, only 6 observations per month could be used. Another disadvantage of method two is, that it cannot be applied in a robust way. Especially for short time series (which lead to low sample sizes for monthly data), one outlier can strongly influence the estimates.

According to the retrospective analysis it can be concluded that simple models best fit the condemnation rates of Swiss slaughterhouse. Thus time related factors, such as seasonality and long term trends do not influence the data strongly in general. Autoregression on the other hand influences the data to a great extent which is reasonable for time series data.

7.2 Discussion: Prospective Analysis

The simulation of outbreaks for condemnation rates of Swiss slaughterhouses were done according to Noufaily et al. [28]. One can argue that using exactly the same method to simulate outbreaks for monthly condemnation counts from slaughterhouses as is used for weekly organism counts in laboratories might not be appropriate. The simulation of the outbreak sizes, might be feasible but the parameters used to lognormal distribute the data in time could be disproportionate. The parameters chosen lead to outbreak durations of several months, which does not fit the expected duration of aberrations reflecting (re-)emerging outbreaks for a wide range of infectious diseases. The incubation times for infectious diseases differ between several days to years (e.g. 1-5 days for foot-and-mouth disease; 1-4 weeks for Lumpy skin disease or several years for BSE). Also the duration of the disease and its transmission rates differ between diseases [31]. Thus outbreak simulations between 3 to 6 months can be too restrictive. On the other hand, for poor animal handling these outbreak durations can be appropriate, as manifestation of maltreatment of animals can be protracted and thus aberrations can be expected to last for several months. By choosing other parameters or another simulation method a wider range of outbreak dynamics can be captured. For example with the {hhh4} models outbreaks could be simulated for the data sets, by increasing the parameter $\{\lambda\}$ progressively for a randomly chosen time period at different starting points. In general only a small amount of data was used to estimate the threshold for outbreak detection (e.g. with b=3 or b=2). With larger amounts of data (e.g. b=5) the uncertainty of the estimates can be decreased and will lead to a more precise threshold. Therefore using a higher amount of historic data can potentially increase the POD or decrease the FPR. Therefore simulating longer time series (longer than 72 months) for the use of the algorithm could improve outbreak detection and might therefore be of interest. The simulation of longer time series might also require simulations of the offsets. As in this case the reuse of the offsets of the historic data is only appropriate if the offset data are not influenced by a trend or seasonal pattern. The offset contributes to a great part to the threshold estimation. Thus in any case the reuse of the historically observed offset for all the simulations might also influence the performance. As such, the investigation of offset simulation is of interest in general.

The performance of the {farringtonFlexible} algorithm for the simulated outbreaks was convenient regarding FPR. One can argue that the method of FPR calculation leads to underrated FPRs (due to the influence of the injected outbreaks in the beginning of the Outbreak-Risk-Period). As these outbreaks can impact the specificity and the FPR for those simulations might be underestimated. This concern cannot be refuted entirely, but since the FPR is fairly stable over k (for almost all algorithms) and this effect can be expected to increase with increasing outbreak size it can be neglected. Thus the FPR calculation seems appropriate for the evaluation of the algorithms on the simulations used. A way to reassure the validity of the FPRs would be to recalculate the rates by applying the algorithm to simulations which do not include any outbreaks (as done by Noufaily et al. [28]). The major part of the convenience in the FPR might be attributed to the monthly reporting system. For more frequent reports (e.g. on weekly or daily basis) the FPRs would need to be reinvestigated again and could be the limiting factor. The same is true for the quantile used to calculate the thresholds. For monthly reports a 0.975 quantile leads to sufficiently low FPRs, but for more frequently reported data the use of a 0.995 quantile might be required (as stated in the studies of Noufaily et al. [28]).

In terms of POD the performance of the algorithm used in this study was convenient for large outbreak simulations. For outbreaks of small sizes, the proportion of outbreaks detected was less satisfying. One problem which is faced is the presence of high variations in the baseline data and the low amount of data points used for parameter estimation. The low amount of data leads to uncertain estimates with potentially high standard errors which lead to increased thresholds. High variances in the baseline data lead to high estimated mean incidence rates and therefore to presumably high thresholds (relative to months with low condemnation counts). Therefore exceeding a threshold with small outbreak sizes is getting less likely. As such it is especially difficult to detect small outbreaks efficiently with the data and the settings used (Figure 15 illustrates this problematic).

Surprisingly the best performing algorithms (in terms of POD for small outbreak detection) turned out to be the algorithms which included the smallest amount of data for threshold estimation. These algorithms are not reliable and should be considered with serious concerns. However, it is still interesting to see that these algorithms performed the best for all the groups. Thus further investigations on monthly seasonal effects with more frequent or larger amount of data would be of interest. The options to account for seasonal effects are not convincing. A resonable approach for improvement might be the modelling of seasonality by sine and cosine functions instead of step functions. Even though there was little evidence for seasonal effects in the monthly data, to reinvestigate seasonality with more frequent data would be of interest.

For our study the possibility to use the algorithm in a robust way was not needed. As the data was known to be outbreak free, this option did not seem to be of interest. It might be interesting to include the option anyway to see if better results can be obtained in terms of POD by down-weighting counts which could be outliers or part of an emerging outbreak. Furthermore it would be interesting to investigate the performance of the algorithm for time series which include outbreaks already in the baseline data, as such scenarios could be faced with real data. With this option, Farrington algorithm is able to account for known as well as unknown outbreaks in the past.

In terms of TTD and especially CUD the algorithm performance for early outbreak detection using monthly data is inconvenient. Despite the fact that the majority of the outbreaks are detected within one month, more than 50% of the outbreak cases already have occurred at the time of detection. Thus as a matter of fact, outbreaks with the assumed distribution in time can hardly be detected early by using monthly data. The applicability of the algorithm for early outbreak detection needs to be investigated using more frequently reported data and cannot be evaluated based on this study.

Recommendation for preliminary test stage on real data (short-term): Overall the application of Farrington algorithm on the currently available data seems reasonable. Despite the lack of timeliness and even though the PODs are not convincingly high for all outbreak sizes, the output of the algorithm is of avail. The cost of applying the algorithms in a monthly manner and the cost of possible investigations due to false positive alarms are low and therefore acceptable. If an alarm is flagged the first investigations that are made on condemnation counts are not very time and cost consuming. First the reason for the condemnations is analysed in detail. A heterogeneous mix of reasons indicates that the alarm is possibly due to occasional events. A second step is the investigation of the geographical pattern of the condemnation counts. Again a heterogeneous distribution of the condemnation counts in terms of the location of the slaughter and the origin of the animal (farm) indicates a coincidental increase of condemnations. Only if there is a high occurrence of one specific reason or a geographical accumulation of condemnations more intensive investigations at higher costs are needed.

The interest in the detection of small outbreaks is of higher priority as bigger outbreaks might also be detected using other sources. Therefore algorithms that increase the PODs of small outbreaks and lead to less than 2 investigations per year are considered applicable. Based on the PODs and the FPRs and by disregarding the reliability of the algorithms (according to the size of historic data used) Algorithm 2 is favourable for all the data sets except for normal slaughtered pigs. For this group Algorithm 3 is of better use. However, taking into account the sample size and the reliability of the algorithms, Algorithms 1 should be applied for all the data sets (Algorithm 1b for normal slaughtered pigs).

For normal slaughtered cattle (with an estimated mean incidence rate of $\bar{\mu}$=81.71) Algorithm 1 is convenient (POD$\geq$75%) for outbreaks of sizes bigger than 122% of the average monthly count in the baseline data. At the same time it leads to far less than two investigation per year. The same applies to Algorithm 1 for emergency slaughtered cattle ($\bar{\mu}$=185.86) for outbreak sizes bigger than 73% of the average monthly count in the baseline data. For normal and emergency slaughtered pigs the outbreak size has to be at least 88% (Algorithm 1b) and 156% (Algorithm 1) of the mean predicted baseline incidents rate respectively to meet the above mentioned conditions of FPR and POD ($\bar{\mu}$=353.09 for normal slaughtered pigs and $\bar{\mu}$=46.15 for emergency slaughtered pigs). The proportion of outbreak detection for the smallest outbreak size (k=2) with the same algorithms are 11% for normal slaughtered cattle, 20% for emergency slaughtered cattle, 10% (Algorithm 1b) for normal slaughtered pigs and 12% for emergency slaughtered pigs.

With thresholds based on the 0.995 quantile (see Appendix) the decrease in POD for small outbreaks was unconvincing compared to the decrease in FPR. Therefore for outbreak detection using the mentioned algorithms and the Swiss slaughterhouse data an alpha of 0.025 should be used.

Outlook: A limiting factor of the data was that only the number of wholly condemned carcasses are reported, which limits the information that can be gained from the data. The picture of condemnation is only complete if the partially condemned carcasses are included as well. The goal of this study was to investigate if the available data are adequate for early detection of outbreaks or increasing incidence rates in order to prevent disease outbreaks or poor handling of production animals. Unfortunately the data available so far do not provide enough information for early outbreak detection. The problem of the available data is the deficit in timeliness as the data are collected in a monthly manner. By using monthly reports the detection of an outbreak will not be as early as possible. Another problem which is caused by the monthly reports is its resolution. The detection of outbreaks shorter than one month is possible but it would be earlier as well as more successful and reasonable with weekly or daily reports. One way to deal with these problems could be to neglect the offset. Currently the date on which a condemnation happens is available but the total numbers of slaughtered animals are often only reported at the end of the month. As the offset is of crucial importance it would be more convenient if the total number of slaughters is reported together with the condemnation counts in a daily or weekly manner [26].

Besides changing the reporting system, another possibility to improve outbreak detection would be to investigate the performance of other algorithms. For example the {EarsC} algorithm which is included in the package {surveillance} could be suitable for outbreak detection in the condemnation rates of Swiss slaughterhouses. This method only uses very recent data (low amount of historic data) and thus is suitable for data which does not include seasonal or trend effects but is still influenced by the previous time points [32]. Another possibility is to apply an approximated version of the {CUMSUM} algorithm that also accounts for the Poisson distribution of the data [33]. But as a matter of fact whatever algorithm is used, the problem of *early* outbreak detection based on monthly data reports will be faced and cannot be avoided. However Farrington algorithm seems to be really convenient for outbreak detection especially because there are many options that can be included e.g. trend and seasonal effects, robustness against outliers and inclusion of an offset.

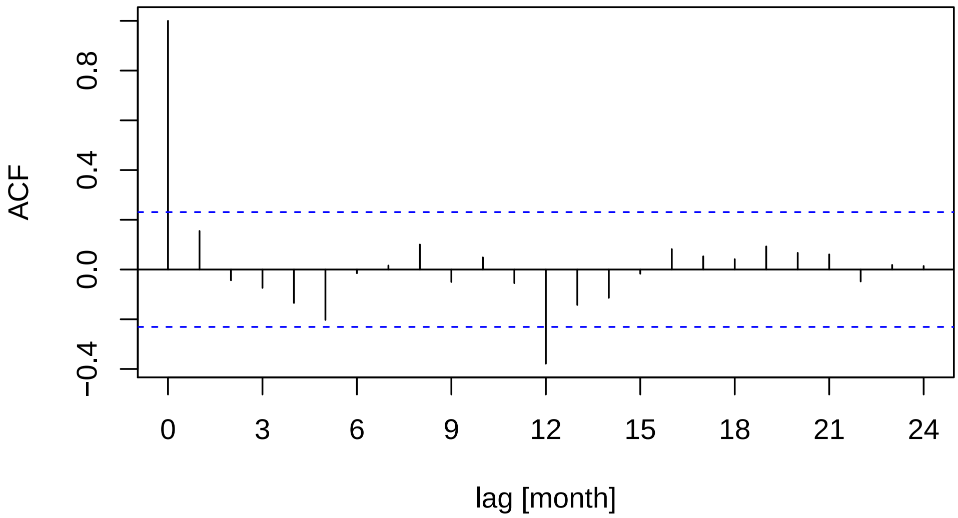
Recommendation for established use on real data (long-term): Conclusively Farrington algorithm is suitable for outbreak detection using condemnation data of Swiss slaughterhouses subject to the condition that the data contains useful information for outbreak detection. Even for monthly data the proportion of outbreaks detected seems convenient when put in relation to the costs of investigation that are caused due to false positive alarms. To establish a long-term outbreak detection system further investigations on both the algorithm and the data are still needed. Especially the inclusion of small subsets of data for the threshold estimation in algorithms (e.g. Algorithm 2) is giving cause of concern. However, it is highly recommended to investigate on the possibility to increase the reporting frequency to at least weekly reports. This would allow the algorithm to be applied in a much more reliable and effective way especially for *early* disease outbreak detection.

8 Acknowledgements

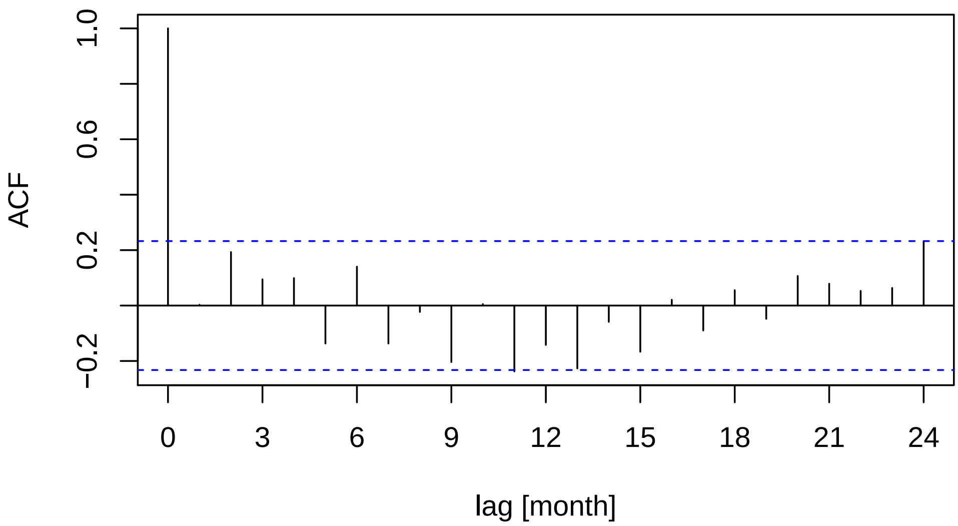
I would very much like to take this opportunity to thank Prof. Dr Leonhard Held and Dr Flavie Vial for giving me the opportunity to work on this highly interesting topic throughout my Master Thesis. I would also like to thank them for all the valuable discussions we shared and the many helpful advices they offered. I am furthermore very grateful to Sebastian Meyer for his great assistance during programming especially for the model selection and data simulations part as well as the helpful discussions throughout the past months. Many thanks also to all the proof readers. Last but by no means least I would also like to thank everyone who has continuously supported me in writing this thesis and making this time such a pleasure for me.

9 Appendix

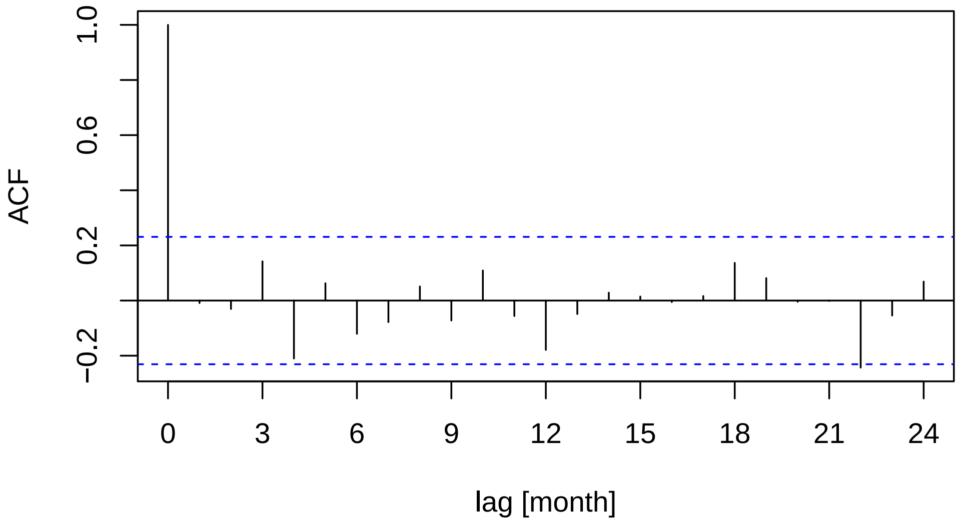
9.1 Appendix: Retrospective Analysis



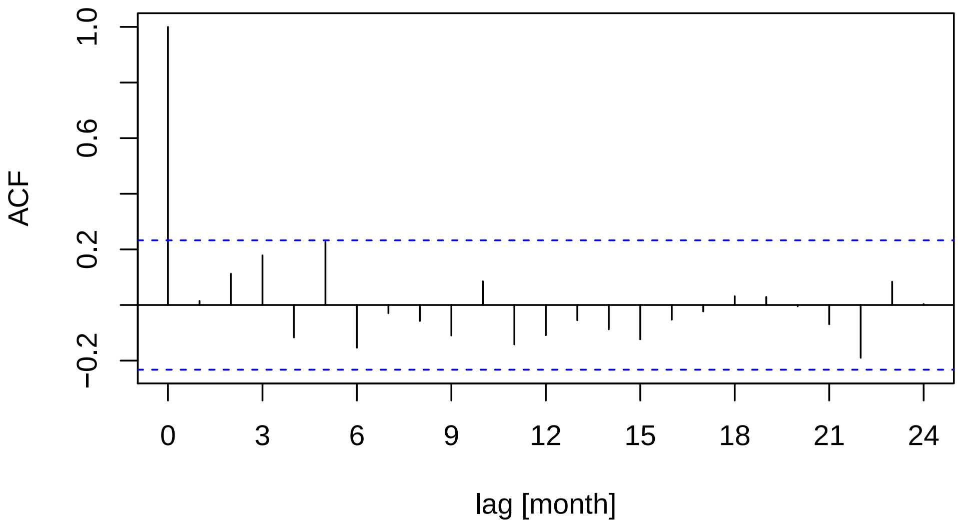
Autocorrelation function of the residuals of the decomposed transformed proportion of condemned carcasses for **normal slaughtered cattle** (based on LOESS)



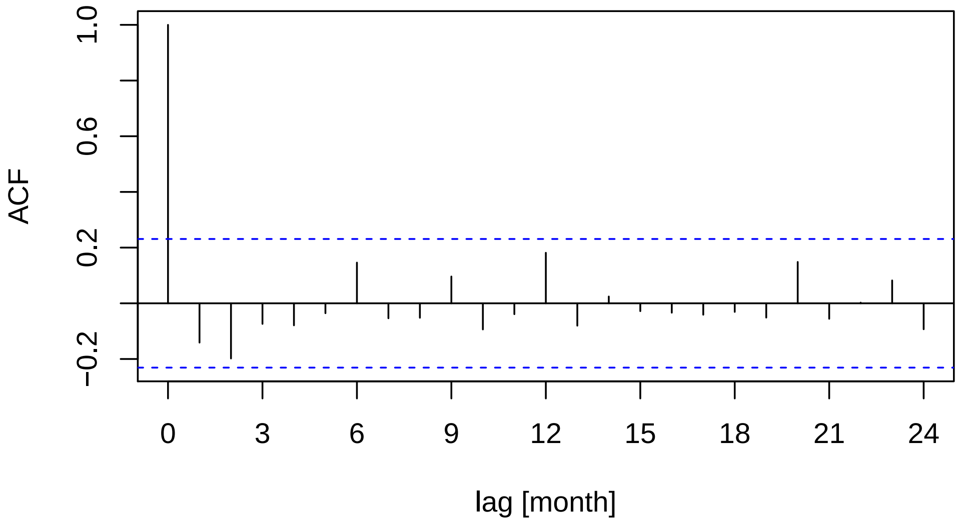
Autocorrelation function of the raw residuals of the best {hhh4}-model for the number of condemned carcasses for **normal slaughtered cattle**



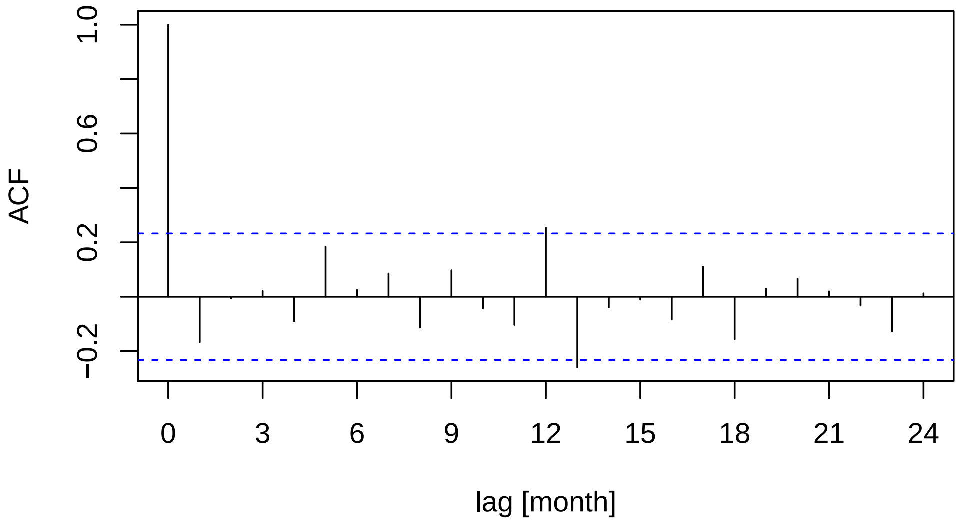
Autocorrelation function of the residuals of the decomposed transformed proportion of condemned carcasses for **emergency slaughtered cattle** (based on LOESS)



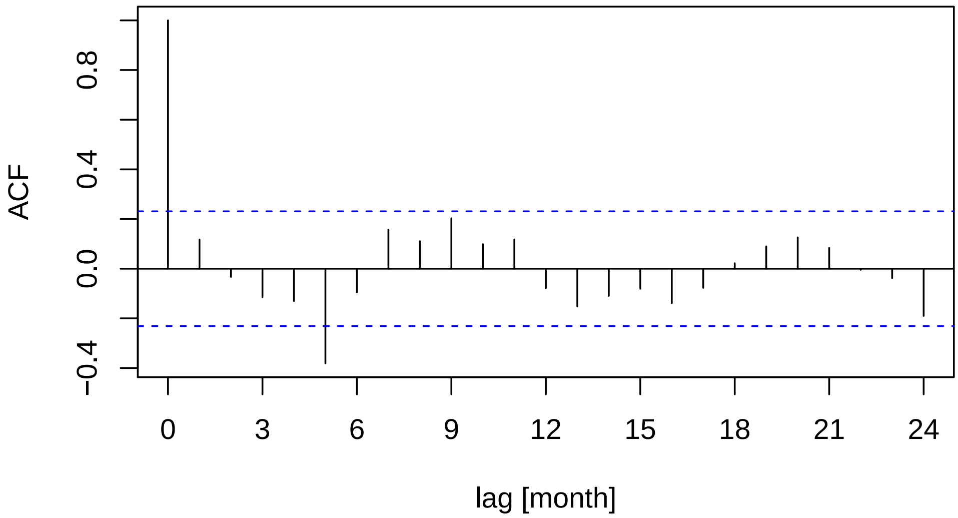
Autocorrelation function of the raw residuals of the best {hhh4}-model for the number of condemned carcasses for **emergency slaughtered cattle**



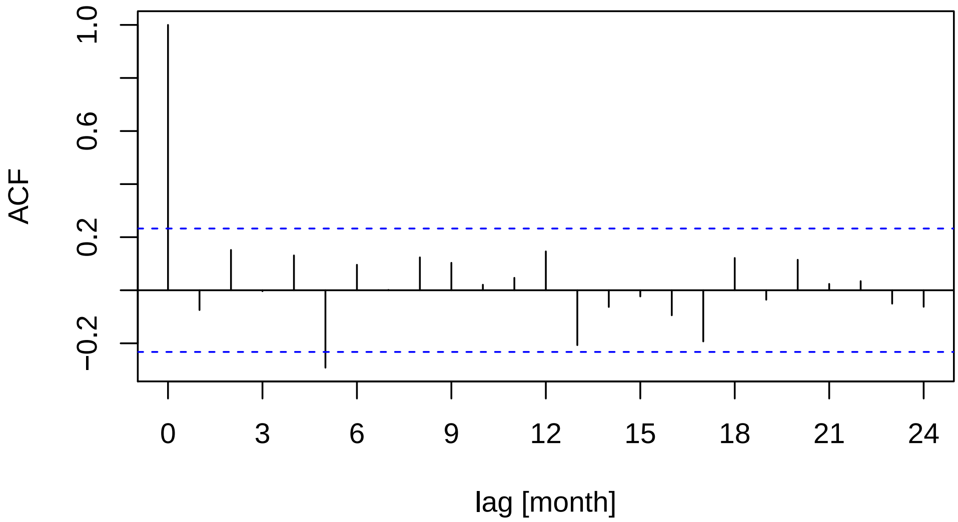
Autocorrelation function of the residuals of the decomposed transformed proportion of condemned carcasses for **normal slaughtered pigs** (based on LOESS)



Autocorrelation function of the raw residuals of the best {hhh4}-model for the number of condemned carcasses for **normal slaughtered pigs**



Autocorrelation function of the residuals of the decomposed transformed proportion of condemned carcasses for **emergency slaughtered pigs** (based on LOESS)



Autocorrelation function of the raw residuals of the best {hhh4}-model for the number of condemned carcasses for **emergency slaughtered pigs**

	trendAR	trendEND	seasonAR	seasonEND	df	BIC	lambda
1	t0	t0	s0	s0	2	684.59	0.26
2	c.e.	t0	c.e.	s0	1	701.86	
3	t2010	t0	s0	s0	3	681.28	0.28
4	t0	t0	xmas	s0	3	681.25	0.26
5	t0	t0	s0	xmas	3	682.28	0.26

According Poisson model fits to the top five negative binomial models for **normal slaughtered cattle** (t0=no trend, t2010=log-linear trend starting in 2010, c.e.=whole autoregressive component excluded, s0=no seasonality, xmas=effect of December only)

	trendAR	trendEND	seasonAR	seasonEND	df	BIC	lambda
1	t0	t1	s0	s0	3	699.97	0.27
2	j2010	t0	s0	s0	3	705.33	0.25
3	j2010	t1	s0	s0	4	699.42	0.25
4	j2010	j2010	s0	s0	3	707	0.28
5	t1	t1	s0	s0	4	701.99	0.24

According Poisson model fits to the top five negative binomial models for **emergency slaughtered cattle** (t0=no trend, t1=log-linear trend, j2010=a shift in the intercept in January 2010, s0=no seasonality)

	trendAR	trendEND	seasonAR	seasonEND	df	BIC	lambda
1	c.e.	t2010	c.e.	s1	4	904.36	
2	t0	t2010	s0	s1	5	908.62	0
3	t2010	t2010	s2	s1	10	847.87	0
4	t0	t2010	s1	s1	7	NA	NA
5	c.e.	t2010	c.e.	s2	6	900.2	

According Poisson model fits to the top five negative binomial models for **normal slaughtered pigs** (t0=no trend, t2010=log-linear trend starting in 2010, c.e.=whole autoregressive component excluded, s0=no seasonality, s1=sine-cosine seasonality with 1 harmonic, s2=sine-cosine seasonality with 2 harmonics)

	trendAR	trendEND	seasonAR	seasonEND	df	BIC	lambda
1	t0	t0	s0	s0	2	563.64	0.28
2	t0	t1	s0	s0	3	558.95	0.26
3	t1	t0	s0	s0	3	561.74	0.33
4	t0	j2010	s0	s0	3	560.93	0.26
5	t0	t2010	s0	s0	3	563.59	0.27

According Poisson model fits to the top five negative binomial models for **emergency slaughtered pigs** (t0=no trend, t1=log-linear trend, t2010=log-linear trend starting in 2010, j2010=a shift in the intercept in January 2010, s0=no seasonality)

9.2 Appendix: Prospective Analysis

k	Dur.	Size	TTD	CUD	POD	FPR
2	3.9	33.1	1.1	28.6	0.03	0.002
3	4.2	49.6	1.0	41.1	0.10	0.002
4	4.3	66.5	1.0	55.3	0.24	0.001
5	4.4	83.5	1.0	68.4	0.42	0.001
6	4.6	99.6	1.0	81.6	0.53	0.001
7	4.7	116.2	1.0	93.2	0.73	0.002
8	4.7	132.6	1.0	105.3	0.84	0.001
9	4.9	149.6	1.0	118.2	0.91	0.001
10	4.9	166.3	1.0	130.9	0.95	0.001

(a) Algorithm 1 - according to retrospective analysis (b=2, w=6, trend=FALSE, noPeriods=1, **alpha=0.005**)

k	Dur.	Size	TTD	CUD	POD	FPR
2	3.9	33.1	1.1	25.1	0.12	0.014
3	4.2	49.6	1.2	39.8	0.21	0.015
4	4.3	66.5	1.1	52.8	0.32	0.014
5	4.4	83.5	1.1	65.4	0.41	0.015
6	4.6	99.6	1.0	75.9	0.50	0.015
7	4.7	116.2	1.0	90.1	0.58	0.014
8	4.7	132.6	1.0	102.3	0.65	0.015
9	4.9	149.6	1.0	113.0	0.71	0.015
10	4.9	166.3	1.0	126.8	0.77	0.014

(b) Algorithm 2 - best POD for small outbreak detection (b=3, w=0, trend=FALSE, noPeriods=1, **alpha=0.005**)

k	Dur.	Size	TTD	CUD	POD	FPR
2	3.9	33.1	1.2	27.5	0.12	0.010
3	4.2	49.6	1.0	39.9	0.27	0.010
4	4.3	66.5	1.0	53.6	0.46	0.007
5	4.4	83.5	1.0	65.8	0.66	0.007
6	4.6	99.6	1.0	78.3	0.78	0.007
7	4.7	116.2	1.0	89.6	0.90	0.007
8	4.7	132.6	1.0	102.2	0.93	0.005
9	4.9	149.6	1.0	115.7	0.96	0.007
10	4.9	166.3	0.9	125.8	0.97	0.007

(c) Algorithm 1 with pastWeeksNotIncluded=2 (b=2, w=6, trend=FALSE, noPeriods=1, **pastWeeksNotIncluded=2)**

k	Dur.	Size	TTD	CUD	POD	FPR
2	3.9	33.1	1.1	27.4	0.11	0.009
3	4.2	49.6	1.0	39.8	0.26	0.009
4	4.3	66.5	1.0	53.8	0.45	0.007
5	4.4	83.5	1.0	65.8	0.66	0.006
6	4.6	99.6	1.0	78.7	0.80	0.006
7	4.7	116.2	1.0	89.5	0.90	0.007
8	4.7	132.6	1.0	102.6	0.93	0.005
9	4.9	149.6	1.0	115.9	0.96	0.006
10	4.9	166.3	0.9	126.1	0.97	0.006

(d) Algorithm 1 with trend included (b=2, w=6, **trend=TRUE**, noPeriods=1)

k	Dur.	Size	TTD	CUD	POD	FPR
2	3.9	33.1	1.2	26.5	0.13	0.013
3	4.2	49.6	1.1	40.2	0.24	0.012
4	4.3	66.5	1.0	52.3	0.43	0.012
5	4.4	83.5	1.0	64.7	0.56	0.013
6	4.6	99.6	1.0	76.4	0.73	0.012
7	4.7	116.2	1.0	90.4	0.82	0.012
8	4.7	132.6	1.0	102.9	0.88	0.013
9	4.9	149.6	0.9	113.6	0.94	0.011
10	4.9	166.3	0.9	125.8	0.95	0.013

(e) Algorithm 2 with w=1 (b=3, **w=1**, trend=FALSE, noPeriods=1)

k	Dur.	Size	TTD	CUD	POD	FPR
2	3.9	33.1	1.2	27.2	0.13	0.010
3	4.2	49.6	1.0	39.9	0.26	0.010
4	4.3	66.5	1.0	52.8	0.43	0.009
5	4.4	83.5	1.0	65.7	0.61	0.009
6	4.6	99.6	1.0	78.2	0.76	0.010
7	4.7	116.2	1.0	90.0	0.88	0.010
8	4.7	132.6	1.0	102.7	0.91	0.009
9	4.9	149.6	1.0	115.1	0.95	0.009
10	4.9	166.3	1.0	126.9	0.96	0.009

(f) Algorithm 2 with w=2 (b=3, **w=2**, trend=FALSE, noPeriods=1)

Outbreak detection performance for the simulated data sets of **normal slaughtered cattle**

k	Dur.	Size	TTD	CUD	POD	FPR
2	4.1	46.4	1.1	40.0	0.06	0.002
3	4.4	69.6	1.0	59.7	0.14	0.002
4	4.5	91.4	1.0	76.8	0.28	0.002
5	4.7	114.9	1.0	94.8	0.49	0.001
6	4.8	136.6	1.0	111.7	0.57	0.001
7	4.9	161.7	1.0	130.1	0.77	0.001
8	5.0	185.0	1.0	146.8	0.86	0.001
9	5.1	207.1	1.0	163.0	0.92	0.001
10	5.1	230.8	1.0	181.3	0.95	0.001

(a) Algorithm 1 - according to retrospective analysis (b=2, w=6, trend=TRUE, noPeriods=1, **alpha=0.005**)

k	Dur.	Size	TTD	CUD	POD	FPR
2	4.1	46.4	1.1	33.4	0.30	0.045
3	4.4	69.6	1.1	50.5	0.39	0.044
4	4.5	91.4	1.1	64.0	0.47	0.045
5	4.7	114.9	1.1	84.6	0.56	0.041
6	4.8	136.6	1.1	99.9	0.63	0.045
7	4.9	161.7	1.0	117.7	0.70	0.044
8	5.0	185.0	0.9	132.0	0.77	0.047
9	5.1	207.1	0.9	148.0	0.80	0.044
10	5.1	230.8	0.9	161.7	0.83	0.043

(b) Algorithm 2 - best POD for small outbreak detection (b=3, w=0, trend=FALSE, noPeriods=1, **alpha=0.005**)

k	Dur.	Size	TTD	CUD	POD	FPR
2	4.1	46.4	1.1	37.0	0.23	0.015
3	4.4	69.6	1.0	54.9	0.38	0.016
4	4.5	91.4	1.0	71.9	0.53	0.013
5	4.7	114.9	1.0	89.4	0.74	0.010
6	4.8	136.6	1.0	106.1	0.84	0.013
7	4.9	161.7	1.0	124.9	0.91	0.010
8	5.0	185.0	0.9	139.0	0.95	0.011
9	5.1	207.1	0.9	152.6	0.96	0.009
10	5.1	230.8	0.9	170.8	0.96	0.010

(c) Algorithm 1 with pastWeeksNotIncluded=2 (b=2, w=6, trend=TRUE, noPeriods=1, **pastWeeksNotIncluded=2**)

k	Dur.	Size	TTD	CUD	POD	FPR
2	4.1	46.4	1.1	38.5	0.05	0.004
3	4.4	69.6	1.0	57.8	0.09	0.005
4	4.5	91.4	1.0	76.1	0.16	0.003
5	4.7	114.9	1.0	93.1	0.29	0.002
6	4.8	136.6	1.0	110.9	0.38	0.003
7	4.9	161.7	1.0	130.4	0.56	0.003
8	5.0	185.0	1.0	147.9	0.69	0.003
9	5.1	207.1	1.0	163.3	0.80	0.002
10	5.1	230.8	1.0	180.6	0.86	0.003

(d) Algorithm 2 with noperiods=11 (b=3, w=0, trend=TRUE, **noPeriods=11**)

k	Dur.	Size	TTD	CUD	POD	FPR
2	4.1	46.4	1.1	36.9	0.05	0.004
3	4.4	69.6	1.0	56.9	0.09	0.004
4	4.5	91.4	1.0	73.6	0.17	0.004
5	4.7	114.9	1.0	95.1	0.30	0.003
6	4.8	136.6	1.0	111.3	0.39	0.004
7	4.9	161.7	1.0	130.8	0.54	0.003
8	5.0	185.0	1.0	148.2	0.63	0.003
9	5.1	207.1	1.0	165.1	0.72	0.004
10	5.1	230.8	1.0	182.1	0.81	0.003

(e) Algorithm 2 with w=1 (b=3, **w=1**, trend=TRUE, noPeriods=1)

k	Dur.	Size	TTD	CUD	POD	FPR
2	4.1	46.4	1.0	37.2	0.06	0.003
3	4.4	69.6	1.0	60.1	0.11	0.003
4	4.5	91.4	1.0	76.5	0.21	0.003
5	4.7	114.9	1.0	95.6	0.38	0.002
6	4.8	136.6	1.0	111.2	0.49	0.003
7	4.9	161.7	1.0	131.4	0.64	0.003
8	5.0	185.0	1.0	148.1	0.74	0.003
9	5.1	207.1	1.0	163.7	0.83	0.003
10	5.1	230.8	1.0	181.4	0.88	0.003

(f) Algorithm 2 with w=2 (b=3, **w=2**, trend=TRUE, noPeriods=1)

Outbreak detection performance for the simulated data sets of **emergency slaughtered cattle**

k	Dur.	Size	TTD	CUD	POD	FPR
2	4.5	88.4	1.3	69.7	0.08	0.008
3	4.8	133.2	1.1	99.1	0.12	0.007
4	5.0	176.9	1.1	134.6	0.18	0.008
5	5.1	221.8	1.0	169.6	0.28	0.008
6	5.3	267.0	1.0	207.4	0.39	0.008
7	5.3	311.0	1.0	241.0	0.51	0.007
8	5.4	355.8	1.0	273.2	0.61	0.007
9	5.5	397.8	1.0	304.1	0.71	0.007
10	5.6	444.4	1.0	341.2	0.78	0.008

(a) Algorithm 1a - according to retrospective analysis (b=3, w=0, trend=TRUE, noPeriods=1, **alpha=0.005**)

k	Dur.	Size	TTD	CUD	POD	FPR
2	4.5	88.4	1.0	74.3	0.02	0.001
3	4.8	133.2	0.9	106.7	0.06	0.001
4	5.0	176.9	1.0	141.6	0.11	0.001
5	5.1	221.8	1.0	178.9	0.26	0.001
6	5.3	267.0	1.0	218.1	0.41	0.001
7	5.3	311.0	1.0	251.5	0.58	0.001
8	5.4	355.8	1.0	281.6	0.76	0.000
9	5.5	397.8	1.0	314.5	0.87	0.000
10	5.6	444.4	1.0	347.6	0.91	0.001

(b) Algorithm 1b - according to retrospective analysis (b=3, w=1, trend=TRUE, noPeriods=3, **alpha=0.005**)

k	Dur.	Size	TTD	CUD	POD	FPR
2	4.5	88.4	1.3	61.6	0.49	0.083
3	4.8	133.2	1.2	93.0	0.57	0.082
4	5.0	176.9	1.1	123.4	0.67	0.083
5	5.1	221.8	1.0	150.8	0.74	0.082
6	5.3	267.0	1.0	181.5	0.80	0.080
7	5.3	311.0	1.0	211.7	0.83	0.078
8	5.4	355.8	0.9	241.7	0.89	0.081
9	5.5	397.8	0.9	267.0	0.92	0.080
10	5.6	444.4	0.9	286.4	0.93	0.084

(c) Algorithm 2 - season only
best POD for small outbreak detection
(b=3, w=0, trend=FALSE, noPeriods=1), **alpha=0.005**

k	Dur.	Size	TTD	CUD	POD	FPR
2	4.5	88.4	1.1	70.8	0.12	0.010
3	4.8	133.2	1.1	108.8	0.22	0.009
4	5.0	176.9	1.0	141.7	0.34	0.008
5	5.1	221.8	1.0	177.7	0.53	0.007
6	5.3	267.0	1.0	210.7	0.66	0.007
7	5.3	311.0	1.0	242.5	0.78	0.006
8	5.4	355.8	1.0	275.2	0.88	0.005
9	5.5	397.8	1.0	300.0	0.94	0.005
10	5.6	444.4	1.0	338.2	0.94	0.005

(d) Algorithm 3 - trend only
Increased POD for small outbreak detection
(b=2, w=6, trend=FALSE, noPeriods=1, **alpha=0.005**)

Outbreak detection performance for the simulated data sets of **normal slaughtered pigs**

k	Dur.	Size	TTD	CUD	POD	FPR
2	3.6	20.6	1.0	19.6	0.04	0.002
3	3.9	31.3	1.0	28.2	0.11	0.001
4	4.0	41.1	1.0	35.7	0.21	0.001
5	4.2	51.2	1.0	43.9	0.36	0.001
6	4.3	61.9	1.0	51.5	0.52	0.001
7	4.4	72.0	1.0	58.6	0.67	0.001
8	4.5	82.9	1.0	66.6	0.80	0.001
9	4.5	92.7	1.0	74.4	0.85	0.001
10	4.6	102.7	1.0	81.0	0.92	0.001

(a) Algorithm 1 - according to retrospective analysis (b=2, w=6, trend=FALSE, noPeriods=1, **alpha=0.005**)

k	Dur.	Size	TTD	CUD	POD	FPR
2	3.6	20.6	1.2	17.8	0.10	0.010
3	3.9	31.3	1.3	27.1	0.18	0.009
4	4.0	41.1	1.1	33.6	0.22	0.010
5	4.2	51.2	1.0	41.2	0.36	0.010
6	4.3	61.9	1.0	47.8	0.42	0.011
7	4.4	72.0	1.0	56.6	0.52	0.010
8	4.5	82.9	1.0	65.4	0.64	0.010
9	4.5	92.7	1.0	72.2	0.66	0.011
10	4.6	102.7	1.0	79.9	0.72	0.009

(b) Algorithm 2 - best POD for small outbreak detection (b=3, w=0, trend=FALSE, noPeriods=1, **alpha=0.005**)

k	Dur.	Size	TTD	CUD	POD	FPR
2	3.6	20.6	1.1	17.7	0.14	0.011
3	3.9	31.3	1.0	26.5	0.28	0.011
4	4.0	41.1	1.0	33.9	0.42	0.008
5	4.2	51.2	1.0	41.6	0.60	0.008
6	4.3	61.9	1.0	49.0	0.74	0.008
7	4.4	72.0	1.0	56.1	0.84	0.008
8	4.5	82.9	1.0	64.5	0.89	0.008
9	4.5	92.7	1.0	71.1	0.92	0.008
10	4.6	102.7	1.0	77.4	0.96	0.007

(c) Algorithm 1 with pastWeeksNotIncluded=2 (b=2, w=6, trend=FALSE, noPeriods=1, **pastWeeksNotIncluded=2)**

k	Dur.	Size	TTD	CUD	POD	FPR
2	3.6	20.6	1.0	17.8	0.12	0.010
3	3.9	31.3	1.0	26.9	0.28	0.009
4	4.0	41.1	1.0	33.8	0.42	0.007
5	4.2	51.2	1.0	41.9	0.60	0.007
6	4.3	61.9	1.0	49.3	0.74	0.006
7	4.4	72.0	1.0	56.1	0.85	0.006
8	4.5	82.9	1.0	64.9	0.89	0.007
9	4.5	92.7	1.0	71.1	0.92	0.006
10	4.6	102.7	1.0	78.0	0.96	0.005

(d) Algorithm 1 with trend included (b=2, w=6, **trend=TRUE**, noPeriods=1)

k	Dur.	Size	TTD	CUD	POD	FPR
2	3.6	20.6	1.1	17.1	0.13	0.012
3	3.9	31.3	1.0	27.1	0.26	0.010
4	4.0	41.1	1.0	33.6	0.36	0.011
5	4.2	51.2	1.0	41.4	0.54	0.011
6	4.3	61.9	1.0	48.8	0.67	0.011
7	4.4	72.0	1.0	56.2	0.79	0.011
8	4.5	82.9	1.0	64.4	0.85	0.010
9	4.5	92.7	1.0	71.8	0.89	0.012
10	4.6	102.7	1.0	78.0	0.93	0.011

(e) Algorithm 2 with w=1 (b=3, **w=1**, trend=FALSE, noPeriods=1)

k	Dur.	Size	TTD	CUD	POD	FPR
2	3.6	20.6	1.0	16.6	0.12	0.011
3	3.9	31.3	1.0	26.8	0.27	0.009
4	4.0	41.1	1.0	33.9	0.40	0.010
5	4.2	51.2	1.0	41.5	0.57	0.009
6	4.3	61.9	1.0	48.8	0.71	0.010
7	4.4	72.0	1.0	55.7	0.83	0.009
8	4.5	82.9	1.0	63.8	0.89	0.008
9	4.5	92.7	1.0	71.7	0.92	0.010
10	4.6	102.7	1.0	77.8	0.96	0.008

(f) Algorithm 2 with w=2 (b=3, **w=2**, trend=FALSE, noPeriods=1)

Outbreak detection performance for the simulated data sets of **emergency slaughtered pigs**

10 References

References

[1] VPHI. Veterinary Public Health Institute Bern. Available online: http://www.vetsuisse.unibe.ch/vphi.html, 2013, last accessed: June 2014.

[2] Triple S Project. Assessment of syndromic surveillance in Europe. *The Lancet*, 378(9806):1833–1834, 2011.

[3] G Shmueli and H Burkom. Statistical challenges facing early outbreak detection in biosurveillance. *Technometrics*, 52(1), 2010.

[4] World Health Organization. Emerging zoonoses. Available online: http://www.who.int/zoonoses/emerging_zoonoses, 2004, last accessed: June 2014.

[5] World Health Organization. WHO guidelines for the global surveillance of severe acute respiratory syndrome (SARS). Available online: www.who.int/csr/resources/publications/WHO_CDS_CSR_ARO_2004_1, 2004, last accessed: June 2014.

[6] World Health Organization. WHO Guidelines on BSE. Available online: http://apps.who.int/iris/bitstream/10665/68932/1/a85721.pdf, 2003, last accessed: June 2014.

[7] D Thompson, P Muriel, D Russell, P Osborne, A Bromley, M Rowland, S Creigh-Tyte, and C Brown. Economic costs of the foot and mouth disease outbreak in the United Kingdom in 2001. *Revue scientifique et technique-Office international des epizooties*, 21(3):675–685, 2002.

[8] Centers of Disease Control and Prevention. CDC Estimates of 2009 H1N1 Influenza Cases, Hospitalizations and Deaths in the United States, April 2009 to March 2013. Available online: http://www.cdc.gov/h1n1flu/estimates_2009_h1n1.htm, 2010, last accessed: June 2014.

[9] C Brown. Emerging zoonoses and pathogens of public health significance–an overview. *Revue Scientifique et Technique-Office International des Epizooties*, 23(2):435–442, 2004.

[10] GD Alton, DL Pearl, KG Bateman, WB McNab, and O Berke. Factors associated with whole carcass condemnation rates in provincially-inspected abattoirs in Ontario 2001-2007: implications for food animal syndromic surveillance. *BMC Veterinary Research*, 6:42, January 2010.

[11] World Health Organization. Global and regional food consumption patterns and trends. Available online: http://www.who.int/nutrition/topics/3_foodconsumption, 2014, last accessed: June 2014.

[12] Earth-Policy-Institute. Rising Meat Consumption Takes Big Bite out of Grain Harvest. Available online: http://www.earth-policy.org/data_highlights/2011/highlights22, 2011, last accessed: June 2014.

[13] C Dupuy, A Bronner, E Watson, L Wuyckhuise-Sjouke, M Reist, A Fouillet, D Calavas, P Hendrikx, and J-B Perrin. Inventory of veterinary syndromic surveillance initiatives in Europe (Triple-S project): Current situation and perspectives. *Preventive Veterinary Medicine*, 111(3):220–229, 2013.

[14] C Dupuy, E Morignat, X Maugey, J-L Vinard, P Hendrikx, Ch Ducrot, D Calavas, and E Gay. Defining syndromes using cattle meat inspection data for syndromic surveillance purposes: a statistical approach with the 2005–2010 data from ten French slaughterhouses. *BMC Veterinary Research*, 9(1):88, 2013.

[15] L Ellerbroek, A Mateus, K Stärk, S Alonso, and A Lindberg. Contribution of meat inspection to animal health surveillance in swine. Available online: `http://www.efsa.europa.eu/en/supporting/doc/191e.pdf`, 2012, last accessed: June 2014.

[16] J Hardstaff, A Nigsch, N Dadios, K Stärk, S Alonso, and A Lindberg. Contribution of meat inspection to animal health surveillance in sheep and goats. Available online: `http://www.efsa.europa.eu/en/search/doc/320e.pdf`, 2012, last accessed: June 2014.

[17] AL Thomas-Bachli, DL Pearl, RM Friendship, and O Berke. Suitability and limitations of portion-specific abattoir data as part of an early warning system for emerging diseases of swine in Ontario. *BMC Veterinary Research*, 8(1):3, 2012.

[18] R Amezcua, DL Pearl, A Martinez, and RM Friendship. Patterns of condemnation rates in swine from a federally inspected abattoir in relation to disease outbreak information in Ontario (2005–2007). *The Canadian Veterinary Journal*, 52(1):35, 2011.

[19] EDI. Verordnung des eidgenoessischen Departementes des Inneren über die Hygiene beim Schlachten (SR 817.190.1). Available online: `http://www.admin.ch/opc/de/classified-compilation/20051438`, 2013, last accessed: June 2014.

[20] RB Cleveland, WS Cleveland, JE McRae, and I Terpenning. STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1):3–73, 1990.

[21] R Development Core Team. *progR: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2009. ISBN 3-900051-07-0, `http://www.R-project.org`.

[22] PA Bromiley and NA Thacker. The effects of an arcsin square root transform on a binomial distributed quantity. *Tina Memo*, 7, 2002.

[23] L Held, M Höhle, and M Hofmann. A statistical framework for the analysis of multivariate infectious disease surveillance counts. *Statistical Modelling*, 5(3):187–199, 2005.

[24] Michaela Paul. The function 'hhh4' in the R-package 'surveillance'. 2013.

[25] Michaela Paul and Leonhard Held. Predictive assessment of a non-linear random effects model for multivariate time series of infectious disease counts. *Statistics in Medicine*, 30(10):1118–1136, 2011.

[26] F Vial and M Reist. Evaluation of Swiss slaughterhouse data for integration in a syndromic surveillance system. *BMC Veterinary Research*, 10(1):33, 2014.

[27] CM Hurvich and Ch-L Tsai. Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307, 1989.

[28] A Noufaily, DG Enki, P Farrington, P Garthwaite, N Andrews, and A Charlett. An improved algorithm for outbreak detection in multiple surveillance systems. *Statistics in Medicine*, 32(7):1206–1222, 2013.

[29] CP Farrington, NJ Andrews, AD Beale, and MA Catchpole. A statistical algorithm for the early detection of outbreaks of infectious disease. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, pages 547–563, 1996.

[30] M Salmon, Schumacher D, and M Höhle. Monitoring Count Time Series in R: Aberration Detection in Public Health Surveillance. Technical report, Robert Koch Institute, 2014.

[31] BVET. Tierseuchen der Schweiz. Available online: http://www.blv.admin.ch/gesundheit_tiere, 2012, last accessed: June 2014.

[32] L Hutwagner, W Thompson, GM Seeman, and T Treadwell. The bioterrorism preparedness and response early aberration reporting system (EARS). *Journal of Urban Health*, 80(1):i89–i96, 2003.

[33] G Rossi, L Lampugnani, and M Marchi. An approximate CUSUM procedure for surveillance of health events. *Statistics in Medicine*, 18(16):2111–2122, 1999.

R version and packages used to generate this report:
R version: R version 3.0.2 (2013-09-25)
Base packages: parallel, stats, graphics, grDevices, utils, datasets, methods, base
Other packages: calibrate, MASS, lattice, reporttools, RColorBrewer, surveillance, polyCub, xtable, sp, Rcpp
This document was generated on Juni 15, 2014 at 15:45.