# Real Time Surveillance for Abnormal Events: the Case of Influenza Outbreaks

Yao Rao and Brendan McCabe
Management School, The University of Liverpool, L69 7ZH, UK

October 25, 2015

**Abstract**

This paper introduces a method of surveillance using deviations from probabilistic forecasts. Realised observations are compared with probabilistic forecasts and the "deviation" metric is based on low probability events. If an alert is declared, the algorithm continues to monitor until an all-clear is announced. Specifically, this article address the problem of syndromic surveillance for influenza (flu) with the intention of detecting outbreaks, due to new strains of viruses, over and above the normal seasonal pattern. The syndrome is hospital admissions for flu like illness and hence the data are low counts. In accordance with the count properties of the observations, an integer valued autoregressive process is used to model flu occurrences. Monte Carlo evidence suggests the method works well in stylised but somewhat realistic situations. An application to real flu data indicates that the ideas may have promise. The model estimated on a short run of training data, did not declare false alarms, when used with new observations deemed in control, ex post. The model easily detected the 2009 $H1N1$ outbreak.

**Keywords**: real-time surveillance, early event detection, probability forecasts, Markov process, integer autoregressive model

## 1 Introduction

Seasonal influenza (flu) is a major cause of death among human populations and places great demands on health care resources and infrastructure. The economic costs of flu are also substantial. Potentially more destructive and costly are illnesses caused by new strains of the flu virus which have the potential to create worldwide pandemics. Hence influenza surveillance has been an important issue in public health practice and as a result significant efforts have been devoted to the development of statistical algorithms to monitor flu data. The objective of this paper is to propose a methodology for syndromic surveillance that detects outbreaks of flu that are in excess of the customary seasonal cycle and that continues to monitor until the (transient) outbreak terminates, at which point

an all-clear is declared. Thus, we seek to use *real-time* leading indicators of the flu to detect the presence of an outbreak sooner than would normally be the case, i.e. we focus on early event detection (EED). *Real-time* early detection is important as any significant time delay between the disease outbreak and the awareness of it, may mean that public health interventions cannot be effectively implemented to slow the spread of the disease. The hope here is that the economic, social and health care costs associated with an epidemic would be significantly mitigated by early preventive action should the proposed EED method prove effective. The literature on biosurveillance is vast and a large variety of statistical algorithms has been proposed; for a recent taxonomy and an overview of the statistical methods used, see Fricker [1], Unkel *et al.* [2] and the collection of essays in Zeng *et al.* [3]. We use hospital presentations with flu like illnesses as the syndrome and hence our data consists of dependent sequences of low counts; the surveillance literature on this topic is much more sparse but see Section 3.2 of [2] for related references.

The outer layer of the surveillance methodology suggested here is really quite generic and applies to a wide variety of surveillance situations. We present a general algorithm which requires only a forecast distribution for future values of the phenomenon under study. This forecast distribution may be model based, Bayesian or frequentist, or even a judgemental forecast from a panel of experts. From the forecast distribution, a rule is constructed to determine when an Alert has taken place due to an abnormal outbreak and, if there has been an alert, the rule also determines when the outbreak terminates. The decision rule and the dynamics of the algorithm as time progresses are specified at this stage. A particular feature of the decision rule, is that the statistic used is based on *probabilistic forecasts* and not on forecasts of future values. This means that the rule is automatically coherent with the observed values under study, i.e. low counts in the case of flu surveillance. The rule uses a very natural metric to assess the deviance of observed values from the forecast distributions, by calculating probability, $p$, of seeing values as or more extreme than the ones observed. Using $p$, based on the probability distribution, as a monitoring statistic is also helpful when the support of the observations is restricted, as in the case of low counts. This is because changes in the mean of the distribution automatically induce changes in the variance and changes in the variance will induce changes in other higher order moments and so on. Hence, using the full probability distribution is much less problematic than using approximations based on the Gaussian distribution where changes in the mean do not affect the variance. This part of the methodology is described intuitively in Section 2 and in detail in Appendix A. The methodology requires the ability to make forecasts of the probability distribution of observations, perhaps several steps into the future. Provided the phenomenon under study may be modelled as a, possibly nonstationary, Markov process, it turns out that all the probabilities that are needed may be computed using only the one-step transition matrix of the chain. These formulae are given in Appendix B. Given the generic surveillance methodology, a strategy for producing forecast distributions for the phenomenon under study needs to be developed. In this paper, we focus on the detection of flu epidemics where

the syndromic data utilised are hospital admissions, collected at relatively high frequencies, potentially daily, weekly or monthly. We use a frequentist model i.e. an integer valued autoregressive ($INAR$) process, a model class which is specifically designed to model dependence in the data but also to preserve the integer structure of low counts. This means that coherent probability forecast distributions, as required by the methodology, are readily available as are the one-step transition probabilities. In this class, it is straightforward to accommodate covariates and hence nonstationarity. In summary then, the method we present has the following features

- the basic decision rule is intuitive, easy to understand and uses a natural deviance metric
- transient outbreaks are seamlessly dealt with
- it may be implemented in a variety of diverse situations and all the formulae necessary for the calculations are explicitly presented. The formulae are valid under quite general conditions.
- low counts which may be dependent, when the data are collected at high frequencies, are specifically accommodated
- it requires no asymptotic justification
- nonstationarity is explicitly incorporated into the modelling process
- the deviance metric, which calculates the probability of extreme observations relative to a probability distribution, is also attractive when the support of the observations is restricted, as the flu case is to the nonnegative integers.

In Section 4 the performance of the EED procedure is assessed via Monte Carlo simulations. Since the full set of outcomes is available to a researcher when using simulations, any desired measure of performance may be computed. We use variants of traditional run lengths to assess false alarm rates when the system is deemed in control. A new method for simulating flu epidemics, based on combining traditional SIR ideas and the $INAR$ model, is introduced to evaluate how well the algorithm performs in detecting flu outbreaks when they occur in excess of the normal pattern.

The use of a training model as part of the methodology is not without its difficulties especially in the context of detecting flu outbreaks over and above the usual annual cycle. One of the difficulties is that there has to be adequate data to estimate a sufficiently complex model to produce accurate forecast distributions in the presence of seasonality but at the same time the data span has to be reasonably short so that no outbreak or other structural change has taken place. The $INAR$ process models the syndrome, counts of hospital admissions with flu like illness, as the dependant variable and uses a discrete valued arrivals distribution to generate new flu incidences. One method of incorporating

nonstationarity to accommodate seasonal flu is to model the mean of the arrivals distribution using monthly, say, dummy variables as covariates so that the average number of new flu cases per month may vary. This method has the advantage of simplicity and ease of interpretation. In addition, the covariates are completely predictable and hence may be projected into the future to produce, in turn, forecast distributions for the syndromic counts via the $INAR$ model. The main drawback of the approach is that the timing and duration of seasonal flu cycles may vary year by year, rendering fixed effect season dummies inappropriate. An effective model for forecasting the distribution of the syndromic counts requires covariates that can account for time-varying flu seasonality and at the same time, such covariates, must be predictable themselves to serve as inputs to the $INAR$ model. Failure to correctly predict the timing of seasonal flu will have a serious impact on the ability of any methodology to identify outbreaks defined as an excess of cases over and above the normal pattern. Therefore, while the syndrome is deemed to be a leading indicator of disease incidence, we need a further predictor of the syndrome itself. Finding covariates to serve as syndrome predictors is an additional challenge. These issues are explored in Section 5 which analyses a publicly available time series of flu counts, where the data include the year 2009 when the $H1N1$ virus was active.

# 2 The Monitoring Algorithm

The outer layer of the methodology is a generic algorithm for constructing a rule to react to deviance, to deal with transience and to update the decision making process as time progresses. The basic monitoring idea is the standard one, i.e. produce a forecast distribution and evaluate discrepancies. The current time is $t = 0$ and we make a prediction for one period $t = 1$ in the future. The type of prediction we make is a forecast for the entire future *distribution* of $Y_1$, sometimes called a density forecast but for discrete data like counts it is a forecast of the probability mass function

$$P\left[Y_1 = j_1 | \mathcal{F}_0\right] \tag{1}$$

for *all* values $j_1$ in the support. We condition on the information available at time $t$ , $\mathcal{F}_t$, initialised $t = 0$. For ease of exposition, we shall assume throughout the paper that the observation process is strictly Markov, i.e. has only one lag. Hence $\mathcal{F}_0$ just consists of the current event $Y_0 = y_0$. Unlike conventional forecasts which predict future values, probabilistic forecasts predict future probabilities, one for each specified possible outcome. Such forecasts are automatically coherent with the integer valued support of count observations.

The procedure has two sub-procedures: one to monitor for alerts ($AM$) and the other to monitor for the all-clear ($CM$). Algorithmic details are given in Appendix A, but the basic ideas are outlined here. We assume that we are currently in control, a forecast distribution has been produced and we start the procedure to monitor for an alert. The initial step evaluates (1). Next, we roll

time forward to $t = 1$ and a new observation $y_1$ becomes available. To decide if the $y_1$ is aberrant (out of control), we compute the probability of observing values as or more extreme than it, i.e.

$$p = \sum_{j_1=y_1}^{\infty} P\left[Y_1 = j_1 | \mathcal{F}_0\right].$$

The investigator then constructs a decision rule based on the value of $p$ to declare an alert. We choose lower and upper values $p_l \leq p_u$ and declare an alert (RED, R) when $p < p_l$, declare a warning when $p_l \leq p \leq p_u$ (ORANGE, O) and continue monitoring when $p > p_u$ (GREEN, G); these values may be chosen by the investigator to suit the application. The alert region described here only screens for extreme observations which are large, in accordance with the practice in syndromic surveillance. The alert monitor then continues as follows:

| | |
|---|---|
| If R | Declare an alert. Start $CM$ |
| If O | Construct $P\left[Y_2 = j_2 \mid \mathcal{F}_0\right]$ |
| If G | Construct $P\left[Y_2 = j_2 \mid \mathcal{F}_1\right]$ |

When an Alert is declared, the algorithm switches to checking for the all-clear (details below). If a warning is issued, a 2-step-ahead forecast distribution conditioning on the last known good information $\mathcal{F}_0$ is constructed. If the decision is G, then a *one*-step ahead forecast is constructed using the new information in $\mathcal{F}_1$, since the system is deemed in control. A new $p$ is calculated when the next observation $y_2$ arrives and the decision rule is re-applied. The rule either declares an alert or continues to monitor indefinitely, updating the O and G steps as required; see Appendix A.

The procedure, $CM$, that monitors for the all-clear is as follows. For the sake of concreteness, assume that at period $t = 3$, the $AM$ procedure issued an Alert after an Orange at $t = 2$ and a Green at $t = 1$. We require a rule that would declare the Alert over, i.e. we would like to know whether the observations that follow the alert have returned to being consistent with the original model. Noting that $\mathcal{F}_1$, delivered by the $AM$, is the latest known good information, we construct the 3-step-ahead probability mass function

$$P\left[Y_4 = j_4 | \mathcal{F}_1\right] \tag{2}$$

an alert having been raised at $t = 3$. We then compute the $p$ as usual when $y_4$ arrives and apply the ROG decision rule. Action is taken as described in the following table:

| | |
|---|---|
| If R | Continue alert, Construct $P\left[Y_5 = j_5 \mid \mathcal{F}_1\right]$ |
| If O | Continue alert, Construct $P\left[Y_5 = j_5 \mid \mathcal{F}_1\right]$ |
| If G | Declare all-clear and restart $AM$ with $\mathcal{F}_4$ |

The rule, in this illustration, is the same for R and O but refinements are possible where the decision will be different (see below). If the decision is R, a prediction for four steps ahead, conditional on the last good information $\mathcal{F}_1$, is produced. If an all-clear is declared, the algorithm re-enters the alert monitoring state using $\mathcal{F}_4$ as the current information. When $y_5$ is observed, we compute $p$ in the usual way. The Clear procedure continues to monitor new observations until the alert is over.

It is possible that two or more successive values of $p$ may fall marginally above $p_l$ and thus not trigger an alert but taken together might indicate the onset of an outbreak. In this situation, an enhanced procedure would be to compute joint probabilities for several observed values that are suspect. The details of the EED algorithm, incorporating this refinement and the updating steps, are given in Appendix A.

# 3 Model based Approaches

In many situations forecast distributions will be produced using a statistical model. There is no restriction on the type of model that may be used in the algorithm: it may be a pure time series model, a regression model, a regime switching model or even a Bayesian construction; all that is required is that forecasts are available from the model.

One of the key differences between the present algorithm and other methods is that we do not use the model to forecast future *values* of the observations but rather future *probabilities* for those values. Existing model based monitoring techniques (see, for example, Serfling [4] for a regression method or Choi [5] for an ARMA approach) typically use a threshold for issuing an alert i.e. if the new observed value is larger than some prespecified amount, an alert is issued. Thresholds are usually constructed by using the distribution of the forecast error, i.e. the difference between the new observation and its conditional mean. For observations which are low counts (small positive integers) and a conditional mean which is a real number this does not seem entirely natural. In addition, there is usually an implicit normality assumption used to calculate the size of the threshold; see the summary in Section 2.3 and also Section 3 of Unkel *et al.* [2]. Algorithms specifically designed to deal with count data based on, for example, a Poisson assumption, e.g. Farrington *et al.* [6] and Noufaily *et al.* [7], recognise the fragility of simple Gaussian approximations in determining the size of a threshold. Accordingly, they use power transformations and Taylor series approximations in an attempt to induce symmetry and near normality. Probability forecasts avoid these difficulties and automatically deal with restrictions on the support. In addition, we explicitly model the autocorrelation structure of the data, by exploiting the Markov structure to improve forecasting performance.

At some point it will be necessary to re-estimate the control model. A useful strategy is to use a rolling window and re-estimate once a year, say, deleting the first year's observations and adding the current year's; this would allow for time-

varying data generating processes to be accommodated. For example, it may be that with the advent of new vaccines, normal seasonal patterns may slowly change as they also might do as known viruses mutate. Obviously, we do not want to use any "alert" observations when the control model is being updated. Consider the simple case where the model was estimated using observations from $n$ years of monthly data $\left\{Y_{-(n*12)}, Y_{-(n*12)+1}, ..., Y_{-1}, Y_0\right\}$, following which observations for the first year from $Y_1, ..., Y_{12}$ become available. For example, if all the decisions were G until an alert was declared at $t = a \in (1, 12)$ with a value $y_a$ while $y_{a+1}$ and subsequent observations till the year end were also deemed clear, then we replace $y_a$ by an estimate of it under the control regime. A simple suggestion is to use the mode, written $\hat{y}_a$, of the predictive distribution for $y_a$ conditional on $\mathcal{F}_{a-1}$, the last known good information. This value is available as a by-product of the monitoring procedure. Then we re-estimate the model based on $\left\{y_{-(n-1)*12}, ..., y_0, y_1, ..., \hat{y}_a, y_{a+1}, ..., y_{12}\right\}$ where the first years observations have been deleted and the current years added. Obviously, if the alert lasted $k$ periods, then we would use $\hat{y}_a, \hat{y}_{a+1}, ..., \hat{y}_{a+k-1}$ to impute as many missing points as were deemed part of the outbreak episode.

Another advantage of the present approach occurs when the phenomenon under study may be modelled as a Markov chain. In the Markov case, all the forecast probability distributions required to implement the monitoring algorithm may be computed from the first-order transition matrix of the chain. For simplicity, we confine attention to the first-order case although higher order processes may be accommodated by the usual device of treating chains with multiple lags as a first-order vector system. To deal with the seasonal effects of flu, we need to consider the nonstationary case. Define the matrix $\mathbf{P}_t$ to contain the individual one-step-ahead conditional forecast probabilities, $\mathbf{P}_t[i, j] = P_t\left[i|j\right]$, where $i, j$ range over the support and $P_t\left[i|j\right]$ is the probability of moving to state $i$ at time $t$ given the system is in state $j$ at time $t-1$. Then, setting products of probabilities of the form $\prod_{j=1}^{0}$ to unity, joint probability forecast distributions $k$ steps into the future are given by

$$\begin{aligned} & P\left[Y_{t+k} = s_{t+k}, ..., Y_{t+k-i} = s_{t+k-i} | Y_t = y_t\right] \\ = & \prod_{j=1}^{i} \mathbf{P}_{t+j}[s_{t+k-j+1}, s_{t+k-j}].\mathbf{P}_{t+k-i}....\mathbf{P}_{t+1}[s_{t+k-i}, y_t] \\ = & \prod_{j=1}^{i} \mathbf{P}[s_{t+k-j+1}, s_{t+k-j}].\mathbf{P}^{k-i}[s_{t+k-i}, y_t] \qquad \text{if stationary} \end{aligned}$$

for $i = 0, ..., (k-1)$. Here $\mathbf{P}_{t+k-i}....\mathbf{P}_{t+1}$ means $(k-1)$ matrix multiplications and we extract the $[i, j]$ element of the product, written $\mathbf{P}_{t+k-i}....\mathbf{P}_{t+1}[i, j]$. The details are given in Appendix B. The next sub-section considers the $INAR$ class of parameter based models which may also be thought of as a Markov chain.

### 3.1 Integer Autoregressive ($INAR$) Models

In this subsection, we customise the algorithm and the probability calculations to deal with the details of the phenomenon we wish to investigate i.e. syndromic counts. The $INAR$ class of models was first introduced by Al-Osh and Alzaid [8] and McKenzie [9], while McKenzie [10] provides a review of the model class. It was also used by Moriña *et al.* [11] to study flu behaviour at the level of the individual hospital.

Let $Y_1, \cdots, Y_n$ be a series of dependent counts generated according to the following first order model

$$Y_t = \alpha \circ Y_{t-1} + \varepsilon_t, \tag{3}$$

where $\{\varepsilon_t\}_{t=1}^{\infty}$ is a series of independently distributed integer-valued random variables. The thinning operator "$\circ$" is defined as follows: given $Y_{t-1}$, $\alpha \circ Y_{t-1} = \sum_{i=1}^{Y_{t-1}} \mathrm{B}_{it}$, where $\mathrm{B}_{1t}, \mathrm{B}_{2t}, \ldots, \mathrm{B}_{Y_{t-1}t}$ are independent and *iid* Bernoulli random variables with $P(\mathrm{B}_{it} = 1) = 1 - P(\mathrm{B}_{it} = 0) = \alpha$. Since $\alpha \circ Y_{t-1}$ given $Y_{t-1}$ is a sum of *iid* Bernoulli random variables, it follows that it has (conditionally) a binomial distribution with parameters $\alpha$ and $Y_{t-1}$. It is further assumed that $\mathrm{B}_{jt}$ and $\varepsilon_t$ are independent. Notice that in this model, $Y_t$ is composed of two random components: the complement of the death (i.e. the survivorship) component $\alpha \circ Y_{t-1}|Y_{t-1}$ and the arrivals component $\varepsilon_t$. Neither of these two components are observed. It also follows that (3) has a Markov chain representation and that $E[Y_t|\mathcal{F}_t] = \alpha Y_{t-1} + E[\varepsilon_t]$. Hence the autocorrelation function for $Y_t$ is given by $\alpha^k$ at lag $k$ and the $INAR$ class is suited to modelling dependent integer-valued data with short memory. The transition probabilities are also available with

$$\mathbf{P}[i,j] = \sum_{s=0}^{\min(i,j)} \binom{j}{s} \alpha^s (1-\alpha)^{j-s} P[\varepsilon_t = i - s]$$

which is the convolution of the binomial and the disturbance distribution. Hence, for example, the 1-step ahead forecast distribution at time $t$ is given by

$$p(x|Y_t = y_t) = \sum_{s=0}^{\min(x,y_t)} \binom{y_t}{s} (\alpha)^s (1-\alpha)^{y_t - s} P[\varepsilon_{t+1} = x - s]$$

and this is used to construct the model probabilistic forecasts. Incidentally, if asymptotic considerations were relevant, this forecast distribution is asymptotically efficient (nonparametrically) for any arrivals distribution when ML estimators are employed, as in McCabe *et al.* [12]. This would be added justification for its use as a statistic in a surveillance algorithm.

There are many choices available for modelling the $\varepsilon_t$, e.g**.** Poisson, negative binomial, double Poisson and so on. They may even be treated nonparametrically as in Drost *et al.* [13] and McCabe *et al.* [12]. In the current context, we need the ability to handle covariates in a straightforward fashion and hence the

Poisson is chosen. This gives

$$\mathbf{P}_t[i,j] = \sum_{s=0}^{\min(i,j)} \binom{j}{s} \alpha_t^s \left(1-\alpha_t\right)^{j-s} \frac{e^{-\lambda_t}\lambda_t^{i-s}}{(i-s)!}$$

where we allow $\alpha_t$ and $\lambda_t$ to be functions of covariates. Specifically, we use

$$\alpha_t = \frac{1}{1+e^{-a_t}}, \;\; \lambda_t = e^{l_t} \tag{4}$$

where $a_t = z_t'a$ and $l_t = x_t'l$. The covariate vectors $z_t$ and $x_t$ are mapped into $(0,1)$ and $(0,\infty)$ using the logit and exponential transformations respectively. The likelihood is then straightforward with

$$L\left(a,l\right) = \prod_{t=1}^{n} \mathbf{P}_t[y_t, y_{t-1}]$$

from which maximum likelihood estimators (MLEs) $\hat{a}$ and $\hat{l}$ may be obtained. Equation (4) then gives the MLEs $\hat{\alpha}$ and $\hat{\lambda}$. Simpler estimates of $(a,l)$ may be obtained by conditional (nonlinear) least squares applied to

$$E\left[Y_t|\mathcal{F}_t\right] = \alpha_t Y_{t-1} + \lambda_t$$

and while these estimates are robust to distributional assumptions, the subsequent probabilistic forecasts are not. In what follows, we usually express models in this conditional mean format both for convenience and computational simplicity. Using iterated expectations, it follows that

$$E\left[Y_t\right] = \alpha_t E\left[Y_{t-1}\right] + \lambda_t \tag{5}$$

giving an expression for the marginal means.

# 4 Assessment of Performance

In this section, we attempt to assess the performance of the EED flu algorithm by Monte Carlo simulations. Many techniques have been proposed to evaluate the performance of EED methods and it appears they are not without controversy especially for dependent data (see, for example, Chapter 6 of Fricker [1]). The data here are additionally nonstationary. Ideally, any surveillance system would have the property that the number of times a false alarm is declared, is very low (1 - specificity), that the probability of identifying an outbreak is high (sensitivity) and the response time of the system to an outbreak is very fast. In addition, we would like to sound the all-clear as soon as possible after the outbreak finishes.

In many cases and for flu in particular, incidences are seasonal. So the first step is to construct a model using "regular" seasonal data but excluding those

years with exceptional outbreaks, e.g. due to the $H1N1$ virus. This has been done by Moriña *et al.* [11] using monthly data based on a single hospital and we use the estimated seasonal patterns reported in their paper (Table 1) as the arrivals means $\lambda_t$ in (3). This defines a somewhat realistic simulated benchmark of seasonal behaviour and a challenging environment for the algorithm to assess deviations. In the next subsection, we evaluate the performance when the model operates under "normal" circumstances (in control) while Subsection 4.2 deals with outbreaks.

## 4.1 Normal Performance

The basic assessment devices are variants of traditional run lengths. We calculate the average number of clear time periods (ACL) that elapses between (falsely) declared alerts under the condition that no outbreak occurred. We also look at the average length of time that the system stays on alert after a (false) alarm has been declared (AAL). In this calculation, we consider only two outcomes: alert (R) or no alert (G). To be concrete, imagine a sequence of observations during a *normal* period where the algorithm made the following decisions:

$$\checkmark\checkmark\checkmark\checkmark \otimes \otimes \checkmark\checkmark\checkmark\checkmark\checkmark \otimes \checkmark\checkmark\checkmark \otimes \otimes \otimes \checkmark\checkmark$$

where ✓ means that the decision was G and ⊗ means that it was R, indicating that a false alarm has been declared. Thus the $ACL = (4+5+3+2)/4 = 3.5$ and $AAL = (2+1+3)/3 = 2$. Note once an Alert has been declared, that state must stay for a minimum of one period.

We define the GREEN and RED zones using $p_l = 0.01$ except for the peak months, December to April, where we set $p_l = 0.05$. Starting in July, the seasonal means of the Poisson arrivals were, as in Table 1 of Moriña *et al.* [11],

$$\lambda_t = (0.14, 0.32, 0.61, 1.37, 0.78, \mathbf{2.68}, \mathbf{6.90}, \mathbf{3.44}, \mathbf{2.50}, \mathbf{1.56}, 0.83, 0.61)$$

where the peak flu season (December to April) is noted in bold. This is now considered to be the "normal" seasonal cycle and the task of the surveillance algorithm is to detect low probability occurrences relative to this pattern, declare alerts and subsequently all-clears. We simulated the model

$$Y_t = \alpha \circ Y_{t-1} + \varepsilon_t$$

with $\varepsilon_t \sim Pois(\lambda_t)$ for $t = 1, 2, ..., 12$ by setting $y_0 = 0$ resulting in $y_1$ being a drawing from a Poisson with parameter $\lambda_1 = 0.14$. Then, $y_2$ is the sum of a drawing from a Binomial with $n = y_1 > 0$ and $p = \alpha$ and a drawing from $\varepsilon_2$, a Poisson with parameter $\lambda_2 = 0.32$. Should $y_1 = 0$ we set the Binomial value to 0. The recursion is repeated until 12 draws are obtained and then we replicated these yearly trajectories $10,000$ times obtaining $120,000$ consecutive observations. Then our algorithm was applied to this observation set to see the decision for each observation, remembering that the algorithm is assumed to know the true model and the parameters. Table 1 gives the run lengths for a

couple of values of $\alpha$ based on the decisions obtained for all the observations as above.

Table 1. Run lengths results

| | Run Lengths (Months) | |
|---|---|---|
| | $\alpha = 0.2$ | $\alpha = 0.4$ |
| $ACL$ | 72.2 | 85.4 |
| $AAL$ | 2.2 | 1.97 |

The average number of months between false alarms was 72 ($\alpha = 0.2$) and 85 ($\alpha = 0.4$). If a false alarm was declared, the average numbers of months in which the alert was falsely maintained was 2.2 and 1.97 for $\alpha = 0.2$ and $\alpha = 0.4$ respectively. In both cases, of the $120,000$ months monitored, about 0.013% were given as false alarms.

## 4.2 Outbreak Performance

Motivated by the standard SIR model of epidemics, we simulate a flu outbreak by allowing the number of new flu cases to depend on the current level of infection, i.e. we modify the basic $INAR$ model

$$Y_t = \alpha \circ Y_{t-1} + \varepsilon_t \tag{6}$$

by making

$$\varepsilon_t \sim Pois\left(\lambda + \beta_t Y_{t-1}\right).$$

When $\beta_t = 0$, this is the usual model but for $\beta_t > 0$ the arrivals rate increases with the number of cases in the previous period $Y_{t-1}$. Assuming $\beta_t$ is a non stochastic parameter sequence (or $\mathcal{F}_{t-1}$ measurable), we also have

$$E\left[Y_t|\mathcal{F}_{t-1}\right] = \left(\alpha + \beta_t\right) Y_{t-1} + \lambda$$

which is to be compared with $E\left[Y_t|\mathcal{F}_{t-1}\right] = \alpha Y_{t-1} + \lambda$ when no outbreak is present. When $(\alpha + \beta_t) < 1$, there are shifts in the conditional mean (and marginal mean) of the process but it remains non explosive; when $(\alpha + \beta_t) \geq 1$, the conditional expectation is an explosive function of $Y_{t-1}$. Thus we can turn the epidemic on and off with $\beta_t$ while $(\alpha + \beta_t)$ can be used as a measure of the size of the outbreak. For the purposes of the simulation, we need a process for $\beta_t$.

Assume that the training model has been estimated (or has known parameters) and that we wish to simulate a flu outbreak. The rough pattern we have in mind is that arrivals increase in numbers, peak after some period and then decrease back to the baseline level. Thus, to model the outbreak period we use

$$Y_r = \alpha \circ Y_{r-1} + \varepsilon_r$$

where $\varepsilon_r \sim Pois\left(\lambda + \beta_r Y_{r-1}\right)$ and

$$\begin{aligned} \beta_r &= k.x_r^p \left(1 - x_r\right)^q \\ x_r &= \frac{r}{D}; \quad r = 0, ..., D. \end{aligned}$$

Qualitatively, $\beta_r$ follows the path, from left to right, of a discrete version of the density of a beta distribution on $[0,1]$, i.e. has an inverted $U$ shape. The constant $k$ controls the size of the peak, for example, $k = 1$ means $\beta_r$ peaks at 0.5 when $p = q = 1$. A large $q \geq 1$ (relative to $p$) means that $\beta_r$ is positively skewed so that the outbreak starts explosively, peaks and then tapers off. Setting $p \geq 1$ to be relatively large means that the outbreak gathers pace slowly, peaks and then disappears quickly. In all cases, the first and last $\beta_r = 0$ and the outbreak terminates when $r = D$, $D - 1$ being the duration. In simulations, we can control $k$, $p$, $q$ and $D$ which allows for quite a lot of flexibility in generating the outbreak pattern.

In the outbreak simulations, we continue with the monthly setup of the last section. We set $D - 1 = 5$, $p = 1$ and $q = 4$ and superimpose the outbreak (starting in December and ceasing in April) on the monthly pattern, where the flu year starts in July and ends in June. We generate $1,000$ replications of 12 monthly trajectories with the arrivals means generated by $\lambda_t + \beta_t Y_{t-1}$ with $\lambda_t$ chosen as before. Thus we have $12,000$ consecutive observations. The number $k$ is used to control the size of the outbreak and we look at two cases by making the value of $(\alpha + \max \beta_t)$ equal to 1.5 and 2.0 so that both processes are explosive.

The in-control model and its parameters are presumed known. Table 2 illustrates the timing response (in months) of the algorithm to an outbreak, and its cessation, that was superimposed on the normal seasonal pattern. We look at the cases when $\alpha = 0.2$ and $\alpha = 0.4$.

Table 2. Timing response of the Alert and Clear Monitors

| | | $\alpha + \max \beta$=1.5 | | | |
|---|---|---|---|---|---|
| | | *AM* (78.6%) | | *CM* (76.1%) | |
| | | No. of months | percentage | No. of months | percentage |
| | Lag | 2.46 | 62.6% | 1 | 6.0% |
| $\alpha = 0.2$ | Exact | 0 | 14.5% | 0 | 13.4% |
| | Lead | 2.00 | 1.5% | 1.54 | 56.8% |
| | | $\alpha + \max \beta$=2 | | | |
| | | *AM* (91%) | | *CM* (88.2%) | |
| | | No. of months | percentage | No. of months | percentage |
| | Lag | 1.94 | 66.6% | 1 | 9.7% |
| | Exact | 0 | 23.9% | 0 | 23.4% |
| | Lead | 1.57 | 1.4% | 1.28 | 55.1% |

| | | $\alpha + \max \beta$=1.5 | | | |
|---|---|---|---|---|---|
| | | *AM* (73.7%) | | *CM* (63.5%) | |
| | | No. of months | percentage | No. of months | percentage |
| | Lag | 2.27 | 54.2% | 1 | 8.7% |
| $\alpha = 0.4$ | Exact | 0 | 18.5% | 0 | 16.0% |
| | Lead | 2.30 | 1.0% | 1.5 | 37.8% |
| | | $\alpha + \max \beta$=2 | | | |
| | | *AM* (89.3%) | | *CM* (65.4%) | |
| | | No. of months | percentage | No. of months | percentage |
| | Lag | 1.79 | 61.0% | 1 | 17.5% |
| | Exact | 0 | 26.6% | 0 | 19.7% |
| | Lead | 2.29 | 1.7% | 1.25 | 28.2% |

From Table 2 we can see that when $\alpha = 0.2$ and $(\alpha + \max \beta_t) = 1.5$, in 78.6% of the 1,000 replications of the yearly trajectories an alert was detected over the 12 months by $AM$ procedure. The $AM$ got the timing of the outbreak exactly right 14.5% of the time. In 62.6% of cases that the outbreak was detected with a lag and average number of months it was late was 2.46. In 1.5% of trajectories, an outbreak was declared prematurely and the average lead time was 2 months.

Having declared an alert, the Clear procedure gave the all-clear, in exactly the right month, 13.5% of the time; 6% of the time it was late by a 1 month on average; The $CM$ anticipated the all-clear 56.8% of the time by an average of 1.54 months. Overall, in 76.1% of cases the all-clear was given by the end of the flu year, June.

When the size of the outbreak is increased to 2.0, the performance of the Alert monitor $AM$ improves with a greater detection rate (91%) and it gets the timing exactly right 23.9% of the time. The average length of the leads and lags also reduces. The performance of the $CM$ also improves. As $(\alpha + \max \beta_t)$ changes from 1.5 to 2.0 when $\alpha = 0.4$, the same qualitative features are evident, i.e. the larger the size of the outbreak, the better the performances of the $AM$ and $CM$ monitors. Increasing $\alpha$ from 0.2 to 0.4, has only a minor effect on the

$AM$ monitor but does reduce the performance of $CM$. To see this, consider Figure 1 which gives the output of the ROG decision rule (with $p_u = 0.2$) for each of the 12 months over the $1,000$ iterations of the yearly pattern. The smaller rate of thinning means that flu cases take longer to exit the system before the end of the flu year and the monitor tends to remain in the Orange state.

# 5 Analysis of Flu Data

The data in Figure 2 are weekly numbers for those diagnosed with flu (106 observations) in the Region of Catalonia (Spain) between 2010 week 23 and 2013 week 17.[1] The seasonal peaks, which reach around 100 cases per month for the region, are clearly evident. These data are post the 2009 $H1N1$ outbreak and we consider these observations to be representative of the current normal seasonal pattern.

## 5.1 Preliminary Analysis

As mentioned earlier, one of the difficulties with the idea of a training model is that there has to be adequate data to estimate a sufficiently complex model but at the same time the data span has to be reasonably short so that no outbreak or other structural change has taken place. While the span of the flu data in Section 5 is reasonably short (2 years of weekly data) it is not feasible to estimate the (52) parameters that weekly dummy variables would require. Accordingly, we use 12 monthly dummies, $D_{t,r}$, $r = 1, ..., 12$.

The conditional mean of the $INAR$ model is $\alpha_t Y_{t-1} + \lambda_t$ and a crude dummy variable linear regression of $Y_t$ on $Y_{t-1}, D_{t,1}, ..., D_{t,12}$ shows that the coefficient of $D_{t,3}$ (March dummy) is negative which is not possible in the context of the model. Attempting to force the positivity restrictions by $NLS$ using $\exp\{l_1 D_{t,1} + ... + l_{12} D_{t,12}\}$ leads to convergence problems. The reason is that there is no suitable model with constant $\alpha$ that is capable of fitting this sort of data. Essentially in conditional mean format, we have $Y_{Mar} = \alpha * Y_{Feb} + \lambda_{Mar}$ being (approximately) of the order $50 = 0.8 * 100 + \lambda_{Mar}$ and there is no positive $\lambda$ that can satisfy this. Models with time varying $\alpha_t$ such as

$$Y_t = Y_{t-1} * \left(1 + e^{a_0 + a_1 D_{t,3} + a_2 D_{t,S}}\right)^{-1} + \exp\{l_1 D_{t,1} + ... + l_{12} D_{t,12}\} \qquad (7)$$

will fit the data adequately and we can estimate by $NLS$. In (7), $D_{t,S}$ is a dummy variable for the Summer months June, July and August where there are very few suspected flu cases. This model has 3 $\alpha_t$'s: the base line $\alpha_t$, using $a_0$, only which is estimated to be 0.78, an $\alpha_t$ for March, using $a_0$ and $a_1$, estimated to be $\hat{\alpha}_{Mar} = 0.4$ and an $\alpha_t$ for Summer, using $a_0$ and $a_2$ which was $\hat{\alpha}_{Sum} = 0.6$. Further details, additional parameter estimates etc. are available on request.

[1] Unlike much hospital data, these are publically available and were kindly supplied by David Moriña.

Figure 3 shows the data as well as fitted values from the model (7). The solid line represents the data, the dots are the estimated conditional expectations $E[Y_t|\mathcal{F}_{t-1}]$ (predicted values) and the dashed lines are the estimated unconditional expectations $E[Y_t]$. The unconditional expectations are calculated using (5) with estimated parameter values and $E[Y_0] = 0$, since $Y_0$ corresponds to a Summer month with almost no flu cases. The fitted values are reasonable but the estimated $E[Y_t]$ badly under estimate the peaks. In addition, flu season seems to vary in its timing each year (coming early in 2011 in comparison to 2012) and hence models using fixed seasonal dummies may struggle to serve as useful syndromic predictors. The time-varying nature of flu seasonality is a very serious problem when attempting to detect outbreaks that are deviant from a normal dynamic pattern. If the control model mis-times the seasonality, it will confound normal seasonal behaviour with outbreak behaviour leading to an excess of false alarms in periods where there are no outbreaks.

## 5.2 An Analysis with Covariates

It is clear that some leading indicator of the timing of seasonal flu would be helpful in constructing a control model with an ability to forecast well. It is well known that cold weather is somehow associated with high flu rates but the phenomenon does not seem to be fully understood. Nevertheless, we constructed a proxy variable $W_t = (\min[0, Temp_t - Mean(Temp_t)])^2$ to look at extreme Winter deviations from the average temperature[2]. In using covariates in prediction models, it is important that the covariate is at least as predictable as the response variable itself. Poor extrapolation of the covariate will lead to poor predictions of the response variable. Fortunately, accurate short term (daily, weekly, say) forecasts of many weather variables are increasingly available for use in forecasting the syndrome.

The simple model

$$Y_t = Y_{t-1} * (1 + e^{a_0})^{-1} + \exp\{l_0 + l_1 W_t\}$$

was fitted with the coefficient estimates of $a_0$ and $l_1$ being highly significant. The estimated value for $\alpha$ was $\hat{\alpha} = 0.6$. Figure 4 shows the data and fitted values. In comparison with Figure 3, we can see that the $W_t$ variable does much better in capturing the timing of the flu outbreaks and the $E[Y_t]$ match the data more closely.

## 5.3 Detecting the 2009 H1N1 Outbreak

Having seemed to obtain a reasonable control model for seasonal flu, it is of interest to check how the EED procedure, using the parameters of the estimated

[2] Again the data were supplied by David Moriña. Only average monthly temperatures were available so weekly proxies were constructed, to illustrate the ideas, by assuming the monthly averages applied to each week of the month. Obviously different variants of temperature could be investigated and indeed other covariates (e.g. social media indicators) be employed as well. These issues are being looked at elsewhere.

control model, would perform when applied to new data thought to be in control and to other new data known to be from an outbreak. We therefore ran the EED procedure (trained on $20011-2013$) on data extracted from the years 2008 (control), 2009 ($H1N1$ outbreak) and 2010. We used $p_l$ and $p_u$ as in Section 4. Figure 5 plots the data. The year 2008 was a particularly mild year for flu and cases peaked at about 50 which is half of the benchmark figure of 100. In contrast, 2009, with the arrival of $H1N1$, peaks at about 175. In addition, the flu seems to have arrived early with extremely high values in late October and early November. Weeks where an EED alert was issued are marked with red squares, the warning weeks with orange triangles and all-clear with green circles. In the top panel of Figure 5, the performance of the weather driven model is shown while that of the dummy variable based procedure is shown in the lower panel.

We can see that the weather based model declares no abnormal behaviour for the flu year 2008/2009 but the Dummy variable model overreacts on several occasions and declares false alarms. The $H1N1$ outbreak in 2009 is easily detected early by both methods. Severely abnormal outbreaks will be detected by most surveillance systems and the major problem is controlling the rate of false alarms; in this regard, the use of, even a fairly crude, model with weather covariates seems to be beneficial.

# 6 Conclusions and Suggestions for Further Work

This paper has introduced a method of surveillance using deviations from probabilistic forecasts. Deviations are measured by assessing the probability of observed values relative to their forecasted distribution. It is shown that, when the process under study may be modelled by a discrete Markov chain, only the one-step ahead transition probabilities are required to implement the algorithm. For low count dependent time series, we suggest using the $INAR$ model class as a control model and we note that the first-order transition matrices are readily available. In a simulated environment, incorporating reasonably realistic seasonality, the algorithm works well. It is not prone to excessive false alarms and can detect easily deviations from the normal flu pattern.

In many practical applications such as the flu incidence data studied here, a nonstationary version of the method needs to be implemented. To detect outbreaks of new strains of the flu virus, only deviations from the normal seasonal cycle are relevant. Thus, the search for a control model is further complicated by the fact that the form of the nonstationarity (seasonality of flu patterns) changes over time. This means that, in order to construct a successful control model which is a useful syndromic predictor, relevant covariates that are themselves predictable need to be found. In the case of flu incidence, constructing a successful control model is a challenging task in its own right and failure to do so will result in excessive false alarms. One suggestion is to use temperature data as a predictor of the timing of the seasonal flu pattern, and this idea seems to be helpful in that weather patterns are themselves predictable and reasonable

good predictors of normal seasonal flu. Clearly there is scope for a much more detailed analysis of this problem.

An analysis of some regional data from Catalonia allowed control models to be estimated using weekly observations on flu incidence and temperature, from $2010 - 2013$. The estimated models were then used to monitor some earlier data that included the 2009 $H1N1$ outbreak. Two forms of control model were used to account for seasonality, one using the weather variable and the other using fixed seasonal dummy variables. We found that both models, as expected, identified the $H1N1$ observations as an outbreak. For non outbreak weeks, the dummy variable model had a tendency to declare false alarms more frequently than the weather based one.

In the future, it would be interesting to investigate how such a monitoring procedure might be integrated into existing real-time data bases, might operate at the level of individual hospitals, say, as well as incorporating multivariate and spatial dimensions.

# 7 References


1. Fricker RD. *Introduction to Statistical Methods for Biosurveillance:With an Emphasis on Syndromic Surveillance* 2013. Cambridge University Press.

2. Unkel S, Farrington CP, Garthwaite PH, Robertson C and Andrews N. Statistical methods for the prospective detection of infectious disease outbreaks: a review. *Journal of the Royal Statistical Society: Series A* 2012; **175(1)**; 49–82.

3. Zeng D, Chen H, Castillo-Chavez C, Lober WB. and Thurmond M *Infectious Disease Informatics and Biosurveillance*. Integrated series in information system 27. 2011. Springer. DOI: 10.1007/978-1-4419-6892-0.

4. Serfling, R. Methods for current statistical analysis of excess pneumonia-influenza deaths. *Public Health Report* 1963; **78;** 494–506.

5. Choi, K. An evaluation of influenza mortality surveillance, 1962-1979: 1, Time series forecasts of expected pneumonia and influenza deaths. *American Journal of Epidemiology* 1981; **113;** 215–226.

6. Farrington, CP., Andrews, NJ., Beale, AD. and Catchpole, MA. A statistical algorithm for the early detection of outbreaks of infectious disease. *Journal of the Royal Statistical Society: Series* A 1996; **159;** 547–563; DOI: 10.2307/2983331.

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

8. Al-Osh MA and Alzaid AA. First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis* 1987; **8**; 261-75; DOI: 10.1111/j.1467-9892.1987.tb00438.x.

9. McKenzie E. Some simple models for discrete variate time series. *Journal of the American Water Resources Association* 1985; **21(4)**; 645-650;DOI: 10.1111/j.1752-1688.1985.tb05379.x.

10. McKenzie E. Discrete variate time series. *Handbook of Statistics.* Shanbhag DN and Rao CR (eds); Elsevier; 2003; **21;** 573-606; DOI:10.1016/S0169-7161(03)21018-X.

11. Moriña D, Puig P, Ríos J, Vilella, A and Trilla A. A statistical model for hospital admissions caused by seasonal diseases. *Statistics in Medicine* 2011; **30**; 3125-3136; DOI:10.1002/sim.4336.

12. McCabe BPM, Martin GM and Harris D. Efficient probabilistic forecasts for counts, *Journal of the Royal Statistical Society: Series B* 2011; **73(2)**; 253-272; DOI: 10.1111/j.1467-9868.2010.00762.x.

13. Drost FC, Van den Akker R and Werker BJM. Efficient estimation of autoregression parameters and innovation distributions for semiparametric integer-valued AR(p) model. *Journal of the Royal Statistical Society: Series B* 2009; **71;** 467-485; DOI: 10.1111/j.1467-9868.2008.00687.x

# 8 Appendix A

The algorithm needs to keep track of absolute time and time relative to the decisions taken by the ROG rule. We use superscripts for relative time and subscripts for absolute time. So, define $P_t^i\left[\cdot|\mathcal{F}_j\right]$ to be the joint $i$-step ahead forecast distribution at time $t$ conditional on the information at time $j$, $\mathcal{F}_j$, i.e.

$$P_t^i\left[\cdot|\mathcal{F}_j\right] = P\left[Y_t^i = k_i, ..., Y_{t-i+1}^1 = k_1|\mathcal{F}_j\right]. \tag{8}$$

This notation allows for any number of joint distributions, $i$, to be evaluated at time $t$, relative to conditioning information at any time $j$. Hence the associated value of $p$ of the observations $y_t^i, ..., y_{t-i+1}^1$ is

$$p = \sum_{k_i=y_t^i}^{\infty} ... \sum_{k_1=y_{t-i+1}^1}^{\infty} P\left[Y_t^i = k_i, ..., Y_{t-i+1}^1 = k_1|\mathcal{F}_j\right]$$

Since the joint probability tends to zero as $i$ gets large we need to adjust $p$. For example, we could just use a simple Bonferonni style modification and redefine $p = p * i$ to compare with $p_l$ and $p_u$. Assume we have an ROG rule.

The numerical requirements for the algorithm are values for $i$, $j$ and $t$. To initialise the $AM$ procedure, use the information in $\mathcal{F}_j$, $j = 0$, set $i = s = 1$ and $t = 0$. The Alert monitor is then:

AM 1 Time moves on $t = t + 1$ and a new observation $y_t^s$ arrives

AM 2 Evaluate $P_t^s[\cdot|\mathcal{F}_{t-s}]$, the joint $s$-step ahead forecast distribution conditional on $\mathcal{F}_{t-s}$ using (8). Compute $p$ using $P_t^s[\cdot|\mathcal{F}_{t-s}]$ for the known observations $y_{t-s+1}^1$ up to $y_t^s$.

AM 3 Determine the outcome of the ROG rule using $p$. Then

| | |
|---|---|
| If R | Declare an alert. Go to $CM$ 1 using $r = 1$ and the current $t$, $s$ |
| If O | Set $s = s + 1$ and go back to $AM$ 1 |
| If G | Set $s = 1$ and go back to $AM$ 1 |

The Clear Monitor receives the alert time $t$, the current good information set $\mathcal{F}_{t-s}$ and $r = 1$ from the Alert Monitor. The Clear monitor is then:

CM 1 Time moves on $t = t + 1$ and a new observation $y_t^r$ arrives.

CM 2 Evaluate $P_t^r[\cdot|\mathcal{F}_{t-s}]$, the joint $r$-step ahead forecast distribution after the alert at $t$ conditional on the $\mathcal{F}_{t-s}$ as in (8). Compute $p$ using $P_t^r[\cdot|\mathcal{F}_{t-s}]$ for the known values $y_{t-r+1}^1$ up to $y_t^r$.

CM 3 Determine the outcome of the ROG rule using the $p$. Then

| | |
|---|---|
| If R | Continue alert, set $r = 1$ and go to $CM$ 1 |
| If O | Set $r = r + 1$ and go to $CM$ 1 |
| If G | Declare all-clear and go to $AM$ 1 using $s = 1$ and the current time $t$. |

The Clear Monitor passes back the absolute time $t$, current good information $\mathcal{F}_t$ and $s = 1$ to the Alert Monitor. The enhanced algorithm monitors indefinitely and switches between states as required.

# 9 Appendix B

We implicitly use the Markov property of the model with just one lag throughout.

To fix ideas set $k = 3$. Using the Markov property and summing out the 2 intervening variables, the usual 3 step ahead forecast is given by

$$\begin{aligned} P[Y_{t+3} &= s_{t+3}|Y_t = y_t] \\ &= \sum_{s_2=0}^{\infty} P_{t+3}[Y_{t+3} = s_{t+3}|Y_{t+2} = s_2] \sum_{s_1=0}^{\infty} P_{t+2}[Y_{t+2} = s_2|Y_{t+1} = s_1] P_{t+1}[Y_{t+1} = s_1|Y_t = y_t] \\ &= \mathbf{P}_{t+3}\mathbf{P}_{t+2}\mathbf{P}_{t+1}[s_{t+3}, y_t] \\ &= \mathbf{P}^3[s_{t+3}, y_t] \qquad \text{if stationary} \end{aligned} \tag{9}$$

Similarly,

$$\begin{aligned}
P[Y_{t+3} &= s_{t+3}, Y_{t+2} = s_{t+2}|Y_t = y_t] \\
&= P_{t+3}[Y_{t+3} = s_{t+3}|Y_{t+2} = s_{t+2}] \sum_{s_1=0}^{\infty} P_{t+2}[Y_{t+2} = s_{t+2}|Y_{t+1} = s_1] P_{t+1}[Y_{t+1} = s_1|Y_t = y_t] \\
&= \mathbf{P}_{t+3}[s_{t+3}, s_{t+2}]\mathbf{P}_{t+2}\mathbf{P}_{t+1}[s_{t+2}, y_t] \\
&= \mathbf{P}[s_{t+3}, s_{t+2}]\mathbf{P}^2[s_{t+2}, y_t] \qquad \text{if stationary}
\end{aligned}$$

Finally,

$$\begin{aligned}
P[Y_{t+3} &= s_{t+3}, Y_{t+2} = s_{t+2}, Y_{t+1} = s_{t+1}|Y_t = y_t] \\
&= P_{t+3}[Y_{t+3} = s_{t+3}|Y_{t+2} = s_{t+2}] P_{t+2}[Y_{t+2} = s_{t+2}|Y_{t+1} = s_{t+1}] P_{t+1}[Y_{t+1} = s_{t+1}|Y_t = y_t] \\
&= \mathbf{P}_{t+3}[s_{t+3}, s_{t+2}]\mathbf{P}_{t+2}[s_{t+2}, s_{t+1}]\mathbf{P}_{t+1}[s_{t+1}, y_t] \\
&= \mathbf{P}[s_{t+3}, s_{t+2}]\mathbf{P}[s_{t+2}, s_{t+1}]\mathbf{P}[s_{t+1}, y_t] \qquad \text{if stationary}
\end{aligned}$$

The generalisation to $k$-steps is immediate, setting products of probabilities of the form $\prod_{j=1}^{0}$ to unity, we get

$$\begin{aligned}
&P\left[Y_{t+k} = s_{t+k}, ..., Y_{t+k-i} = s_{t+k-i}|Y_t = y_t\right] \\
&= \prod_{j=1}^{i} \mathbf{P}_{t+j}[s_{t+k-j+1}, s_{t+k-j}].\mathbf{P}_{t+k-i}....\mathbf{P}_{t+1}[s_{t+k-i}, y_t] \\
&= \prod_{j=1}^{i} \mathbf{P}[s_{t+k-j+1}, s_{t+k-j}].\mathbf{P}^{k-i}[s_{t+k-i}, y_t] \qquad \text{if stationary}
\end{aligned}$$

for $i = 0, ..., (k-1)$. Setting $i = (k-1)$ gives the usual joint $k$-step ahead forecast distributions

$$\begin{aligned}
&P\left[Y_{t+k} = s_{t+k}, ..., Y_{t+1} = s_{t+1}|Y_t = y_t\right] \\
&= \prod_{j=0}^{k-1} P_{t+j}\left[Y_{t+k-j} = s_{t+k-j}|Y_{t+k-j-1} = s_{t+k-j-1}\right] \\
&= \prod_{j=0}^{k-1} P\left[Y_{t+k-j} = s_{t+k-j}|Y_{t+k-j-1} = s_{t+k-j-1}\right] \qquad \text{if stationary}
\end{aligned}$$

using the definition of $\mathbf{P}_t[i,j] = P_t\left[i|j\right]$. Similarly, setting $i = 0$ gives the marginal $k$-step-ahead forecast distributions

$$\begin{aligned}
P\left[Y_{t+k} = s_{t+k}|Y_t = y_t\right] &= \mathbf{P}_{t+k}.....\mathbf{P}_{t+1}[s_{t+k}, y_t] \\
&= \mathbf{P}^k[s_{t+k}, y_t] \qquad \text{if stationary}
\end{aligned}$$

Given a joint forecast distribution and "future" observations $\{y_{t+1}, ..., y_{t+k}\}$,

we can compute $p$ as usual, e.g.

$$\begin{aligned} p &= P\left[Y_{t+k} \geq y_{t+k}, ..., Y_{t+1} \geq y_{t+1} | Y_t = y_t\right] \\ &= \sum_{s_3=y_{t+k}}^{\infty} ... \sum_{s_1=y_{t+1}}^{\infty} P\left[Y_{t+k} = s_{t+k}, ..., Y_{t+1} = s_{t+1} | Y_t = y_t\right] \\ &= \sum_{s_3=y_{t+k}}^{\infty} ... \sum_{s_1=y_{t+1}}^{\infty} \prod_{j=0}^{k-1} P\left[Y_{t+k-j} = s_{t+k-j} | Y_{t+k-j-1} = s_{t+k-j-1}\right] \end{aligned}$$

These formulae show that, for Markov chains, all the calculations required by the monitoring algorithm may be computed for any system where the matrix of first-order transition probabilities $\mathbf{P}_t$ is available.

# 10 Diagrams

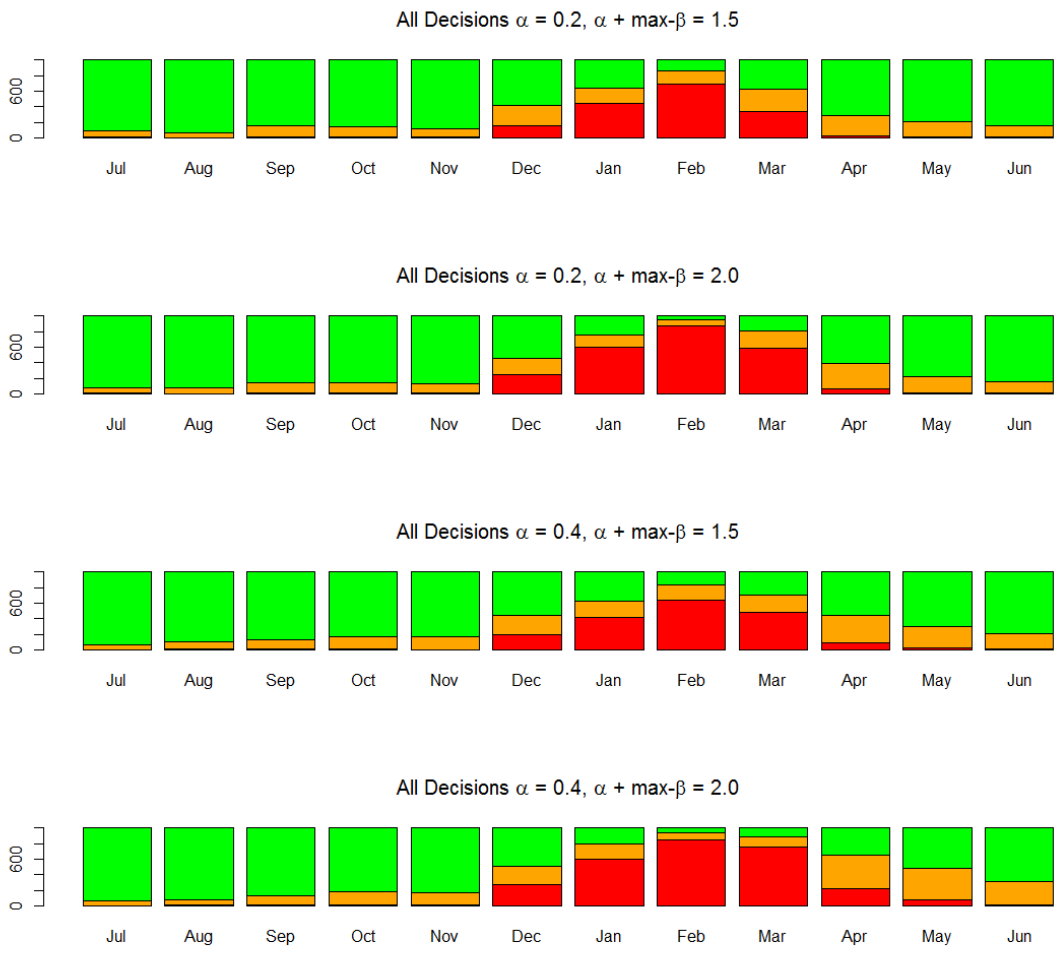


Figure 1: ROG decision output

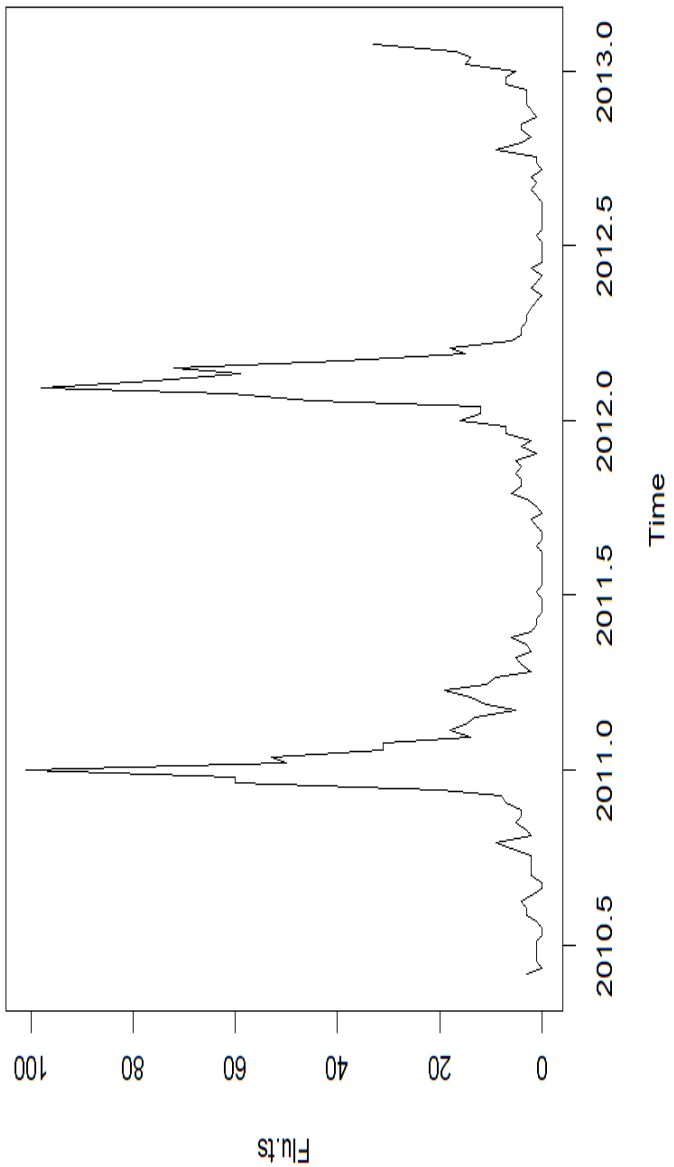


Figure 2: Plot of Flu data from 2010 to the beginning of 2013

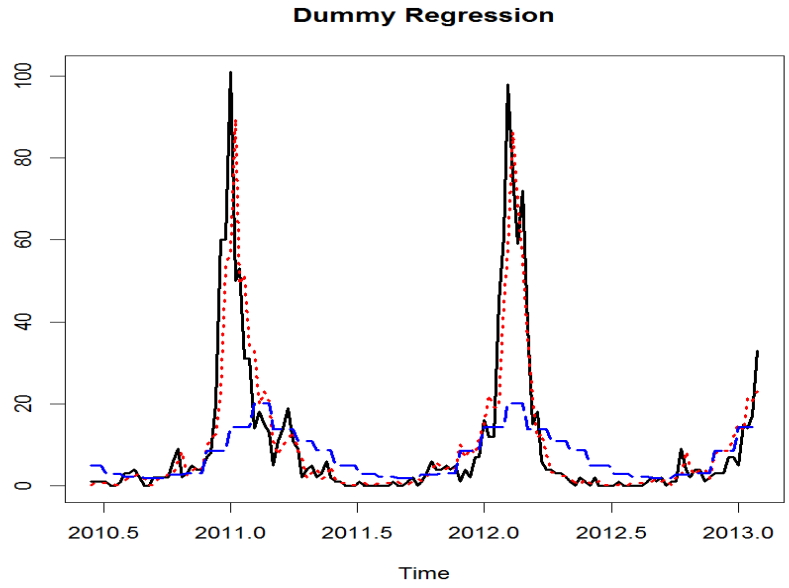


Figure 3: Plot of Flu data (black smooth line), fitted values (red dotted line) and unconditional mean (blue broken line) using Dummy variable model

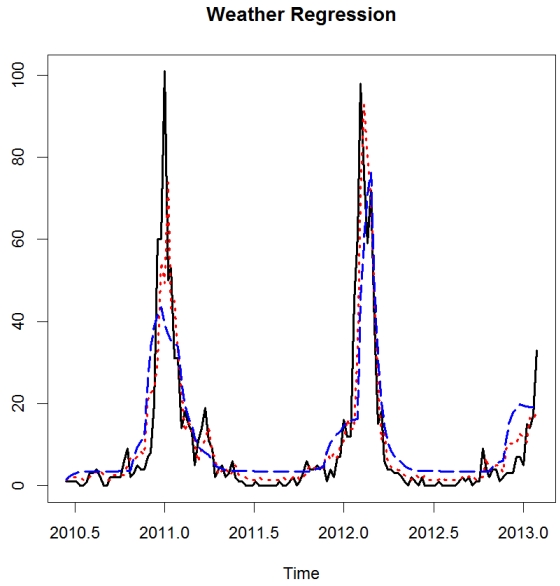


Figure 4: Plot of Flu data (black smooth line), fitted values (red dotted line) and unconditional mean (blue broken line) for the weather variable model

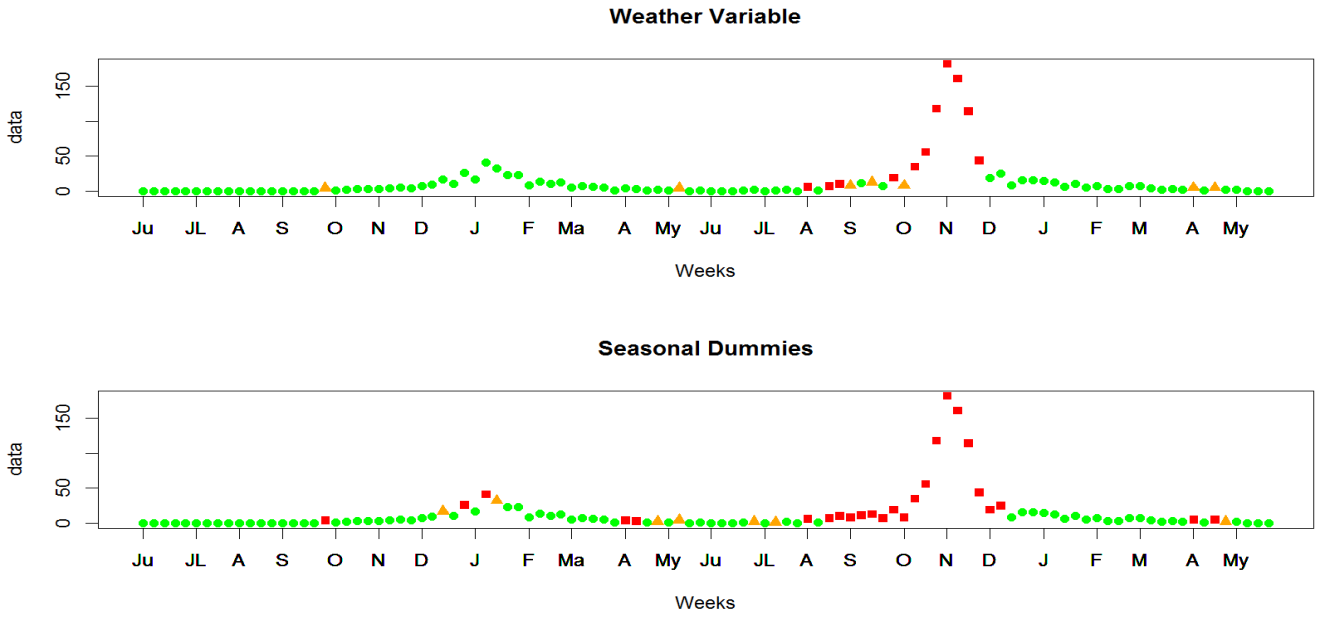


Figure 5: Weekly data processed by the EEDP for 2008-2010. Circles, triangles and squaresn indicate clear weeks, warnings weeks and alert weeks respectively.