FluHMM: A simple and flexible Bayesian algorithm for sentinel influenza surveillance and outbreak detection

HELLENIC CENTER FOR DISEASE CONTROL & PREVENTION (H.C.D.C.P.)
MINISTRY OF HEALTH

Theodore Lytras[1],Kassiani Gkolfinopoulou[2]

[1]Office of Scientific Support, Hellenic Centre for Disease Control and Prevention, Athens, Greece. [2]Department of Epidemiological Surveillance and Intervention, Hellenic Centre for Disease Control and Prevention, Athens, Greece

Introduction

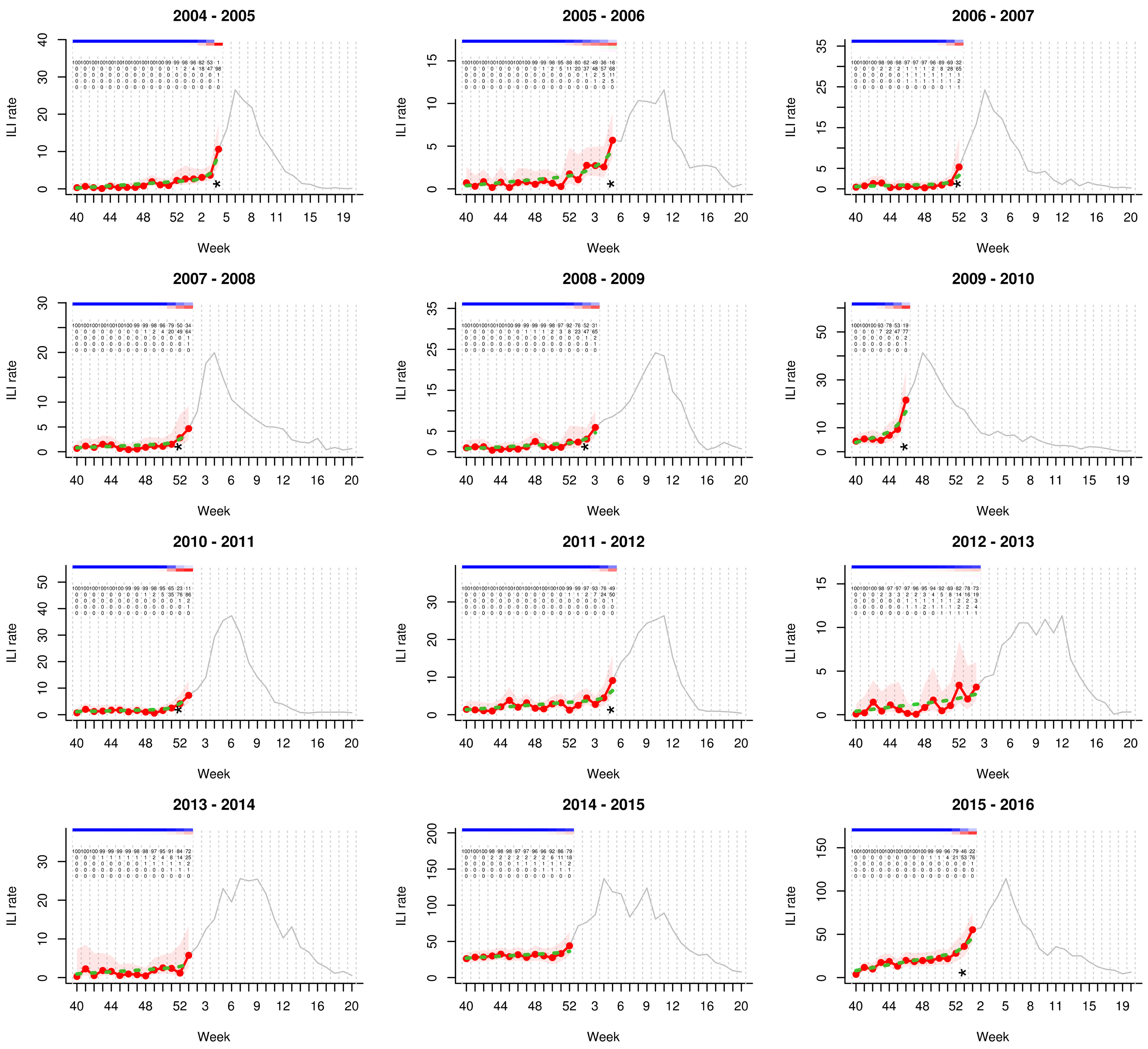
A variety of influenza surveillance systems have been established in many countries, particularly sentinel surveillance for Influenza-like Illness (ILI) / Acute Respiratory Infection (ARI). One of their primary objectives is **timely detection of increased influenza activity**, both seasonal and pandemic. This requires flexible automated algorithms that maximize the utility of all available data, while being sufficiently user-friendly to facilitate widespread application. Our objective was to develop such an algorithm for sentinel influenza surveillance.

Methods

We used a Hidden Markov Model, estimated under a Bayesian framework, to separate a given time period in five phases: (1) *pre-epidemic*, (2) *epidemic growth*, (3) *epidemic plateau*, (4) *epidemic decline*, and (5) *post-epidemic*. The model provides the **posterior probability of being in each phase**, for every week in the surveillance period, using the available ILI/ARI data up to the target week. Each phase is modelled as a Gaussian process, with the mean as a linear function of time for the first four phases and as exponentially decaying for the post-epidemic phase. Only information from the current season is utilized; data from comparable past seasons can optionally be incorporated into the model as informative priors, instead of the default noninformative ones.

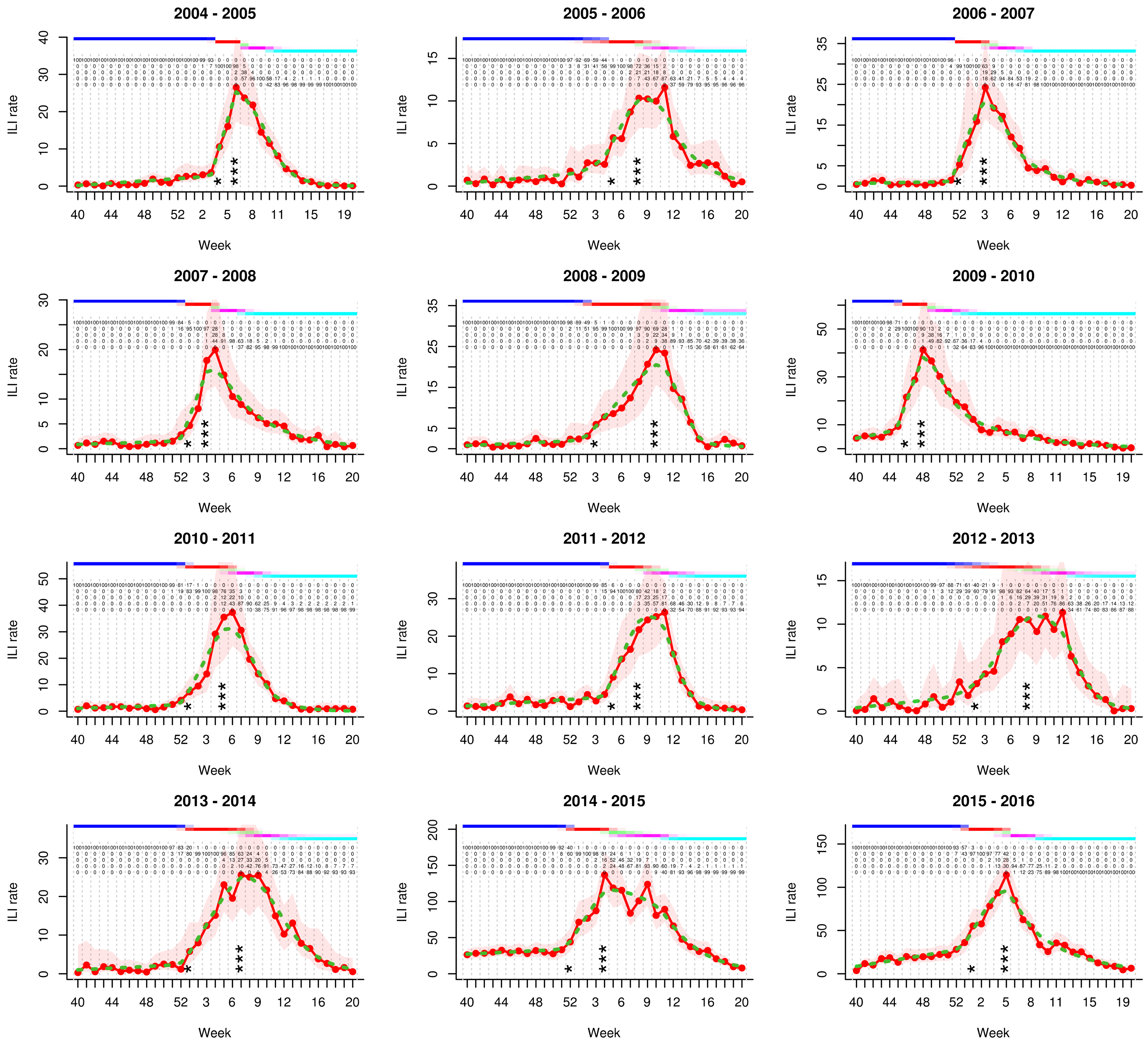
The model is fitted with the JAGS software, and we have created an opensource R package named "FluHMM" to facilitate running the model, summarizing the output and plotting the results. The package is available at: https://github.com/thlytras/FluHMM .

Figure 2. FluHMM model output at the most likely first epidemic week, seasons 2004-05 to 2015-16



Legend for Figures 1 & 2: Red line – observed ILI rates. Grey line – unobserved (future) ILI rates. Green dotted line – fitted FluHMM mean ILI rate. Top colored bars & vertically stacked values – posterior probabilities per epidemic phase. Bottom single star – most likely first epidemic week. Bottom triple star – most likely peak intensity week.

Figure 1. End-of-season FluHMM model output, 2004-05 to 2015-16



Results

We applied the model retrospectively to 12 seasons of sentinel surveillance ILI data from Greece, using only current season data. In all but one cases, the increasing influenza activity was reliably detected at or within a week of its onset, with no false alarms raised early. Setting a posterior probability threshold of 65% for the start of the epidemic wave, the two-week VUTROCS (Volume Under the Time-ROC Curve Surface) was 79.2%, with 100% specificity.

Conclusions

The FluHMM model provides **a flexible, reliable and intuitive method** to detect the onset of the seasonal influenza epidemic, **without relying on comparable historical data**. Performance is unsurprisingly dependent on the quality and precision of the data, but the Bayesian approach still quantifies the certainty associated with the given data in an optimal way, allowing public health decisions to be made.

Box 1. FluHMM example usage in R. (*r* is a vector of ILI rates, and s is an optional vector of log standard errors for the rates)

```
> m <- FluHMM(r, logSE=s)
> update(m, 10000)
> autoUpdate(m)
> plot(m)
> describe(m)
Current week: i=15 (week 201601)
Weeks fixed on phase 1 (pre-epidemic): K=3

Probability that the epidemic has started = 77.9%
Most likely first epidemic weeks:
   i firstEpiWeek Probability
1 14       201553        32.6
2 15       201601        24.2
3 13       201552        17.5

Probability that peak intensity has been reached = 1.4%
```