Epidemics 28 (2019) 100341

Contents lists available at ScienceDirect

Epidemics

journal homepage: www.elsevier.com/locate/epidemics

Real-time prediction of influenza outbreaks in Belgium

Gisele H.B. Miranda[a,b,*], Jan M. Baetens[b], Nathalie Bossuyt[c], Odemir M. Bruno[d], Bernard De Baets[b]

[a] Institute of Mathematics and Computer Science, University of São Paulo, São Carlos - SP, Brazil
[b] KERMIT, Department of Data Analysis and Mathematical Modelling, Faculty of Bioscience Engineering, Ghent University, Gent, Belgium
[c] Epidemiology of Infectious Diseases, Epidemiology and Public Health, Sciensano, Brussels, Belgium
[d] Scientific Computing Group, São Carlos Institute of Physics, University of São Paulo, São Carlos - SP, Brazil

ARTICLE INFO

Keywords:
Influenza
Influenza-like illness
Influenza prediction
Confidence region

ABSTRACT

Seasonal influenza is a worldwide public health concern. Forecasting its dynamics can improve the management of public health regulations, resources and infrastructure, and eventually reduce mortality and the costs induced by influenza-related absenteism. In Belgium, a network of Sentinel General Practitioners (SGPs) is in place for the early detection of the seasonal influenza epidemic. This surveillance network reports the weekly incidence of influenza-like illness (ILI) cases, which makes it possible to detect the epidemic onset, as well as other characteristics of the epidemic season. In this paper, we present an approach for predicting the weekly ILI incidence in real-time by resorting to a dynamically calibrated compartmental model, which furthermore takes into account the dynamics of other influenza seasons. In order to validate the proposed approach, we used data collected by the Belgian SGPs for the influenza seasons 2010–2016. In spite of the great variability among different epidemic seasons, providing weekly predictions makes it possible to capture variations in the ILI incidence. The confidence region becomes more representative of the epidemic behavior as ILI data from more seasons become available. Since the SIR model is then calibrated dynamically every week, the predicted ILI curve gets rapidly tuned to the dynamics of the ongoing season. The results show that the proposed method can be used to characterize the overall behavior of an epidemic.

1. Introduction

According to the World Health Organization (WHO) (WHO, 2016), seasonal influenza is a worldwide public health concern and one of the leading causes of death in the human population. The annual influenza epidemics result in 3 to 5 million severe cases, and about 250,000 to 500,000 deaths (WHO, 2016; WIV-ISP, 2017). In countries with a temperate climate, the seasonal influenza is more frequent during the winter, but in tropical regions the outbreaks are irregular and can occur year round. Monitoring campaigns of influenza epidemics include an assessment of their intensity, duration, peaks and the timing of epidemic outbreaks. The definition of the onset of an epidemic is a crucial point for the management of public health policies and infrastructure, involving for instance alerts for hospitals and the implementation of other health protocols (Nsoesie et al., 2014).

In Belgium, the Unit of Epidemiology and Public Health and the National Influenza Centre (NIC) of Sciensano are responsible for the influenza surveillance system (WIV-ISP, 2017). Since 2007 there exists a network of so-called Sentinel General Practitioners (SGPs) whose main goal is the early detection of an influenza epidemic, and an assessment of its intensity and duration. The SGPs network includes about 1.5% of all Belgian GPs and is representative of them in terms of age, sex and geographical spread. On a weekly basis the GPs provide standardized reports on the number of patients with an influenza-like illness (ILI). High fever, respiratory and systemic symptoms are among the general criteria for ILI (WIV-ISP, 2017).

In spite of the seasonality of influenza epidemics in temperate regions, predicting the onset is still challenging due to differences in meteorological conditions, virus types and human population dynamics. Indeed, across consecutive seasons there is much variability for what concerns the onset, peak and duration. The epidemic of 2014–2015 in Belgium, for instance, was an intense one and had a long duration. Influenza A(H3N2) was the predominant virus type during this season in Europe and North America. There was a higher number of hospitalizations with laboratory-confirmed influenza as compared to the past seasons. In addition, as reported by the European system for monitoring excess mortality for public health action, there was an excess all-cause mortality among elder people (more than 65 years old)

* Corresponding author at: Institute of Mathematics and Computer Science, University of São Paulo, São Carlos - SP, Brazil.
E-mail address: gimiranda@usp.br (G.H.B. Miranda).

https://doi.org/10.1016/j.epidem.2019.04.001
Received 3 September 2018; Received in revised form 14 April 2019; Accepted 16 April 2019
Available online 18 April 2019

during the season 2014–2015 in most of the reporting countries (ECDC Summary, 2015; WHO Review, 2015). This illustrates the importance of public health tools that are capable of real-time monitoring and forecasting the incidence of ILI cases.

Different modelling approaches for forecasting influenza can be found in literature, such as time series, agent-based, metapopulation and compartmental models (Brauer, 2008; Kane et al., 2014; Soebiyanto et al., 2010). In recent reviews (Nsoesie et al., 2014; Chretien et al., 2014), authors highlight the importance of a clear definition of what is to be predicted, the scope and the limitations of the proposed approach, as well as a clear description of the global evaluation measures that ease the comparison of different approaches. Quantifying the performance of real-time approaches remains a challenge (Nsoesie et al., 2014), and is further complicated by the fact that they use different data sources. In the last decade, digital data sources have been used, such as Google Flu Trends (GFT) (Dugas et al., 2013; GFT, 2017; Nsoesie et al., 2013; Shaman and Karspeck, 2012), Wikipedia logs (Generous et al., 2014; McIver and Brownstein, 2014), Twitter (Paul et al., 2014), and others (Ginsberg et al., 2009). For instance, a recent study using GFT data concludes that such data can be used to forecast the week with the highest ILI incidence when these data allow to infer trends, but not to forecast the highest ILI incidence (Nsoesie et al., 2013). Although Google stopped publishing real-time estimates on-line in 2015, the data is still available for researchers.

In contrast to retrospective approaches, their real-time counterparts usually involve a dynamic adjustment of the underlying model parameters as additional data become available. Shaman and Karspeck (2012) propose a method to identify the onset of an epidemic through the use of a data assimilation method (the ensemble adjustment Kalman filter), which combines data observations with an ensemble of simulations obtained from a model in order to provide a posterior estimate of the model parameters. Using a retrospective approach, the authors show that it is possible to predict the weekly incidence of ILI cases. Nsoesie et al. (2013) also present an approach for influenza forecasting that combines an individual-based model and an optimization procedure for finding the optimal parameter set based on GFT data.

In a recent study, Won et al. (2017) present a framework for the early detection of influenza using data from three different sources: the European Influenza Surveillance Network (EISN), GFT and Saúde 24 (Serviço Nacional de Saúde, 2017). The authors show that a combination of data from different sources improves the model accuracy for what concerns the identification of the onset of seasonal influenza. Besides, another study shows the predictive potential of GFT data using generalized linear models and generalized linear autoregressive moving average models, claiming that GFT can be used to underpin a tool for an early identification of the epidemics (Dugas et al., 2013).

In this paper, we propose an approach for the real-time prediction of the weekly incidence of ILI cases of an ongoing epidemic up to its end. It involves two steps, being the construction of a confidence region based on the ILI data of previous seasons and a dynamic calibration of a compartment model. The confidence region accounts for the variability of the weekly ILI incidence among the seasons and the time of occurrence of a specific number of ILI cases. The compartment model is the so-called SIR epidemic model (Kermack and McKendrick, 1927), and is used to predict weekly incidence of ILI cases throughout the remainder of an ongoing season from a given week on. For the dynamic calibration, an optimization algorithm is used to find the parameter set that minimizes the difference between the simulated and the observed weekly ILI incidence up to a given week. We evaluated the proposed approach using Belgian SGP data for the influenza seasons 2010–2011 to 2015–2016.

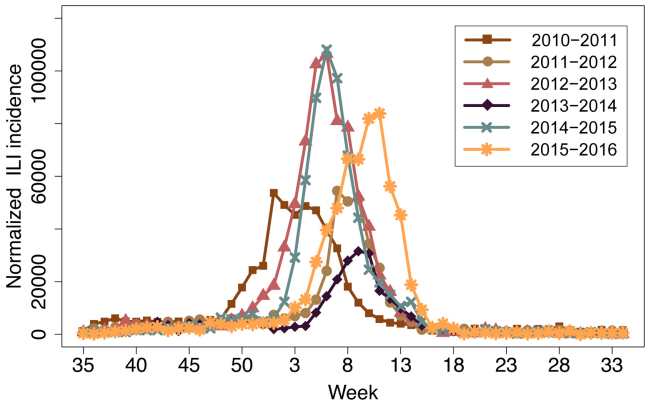

Fig. 1. Normalized weekly incidence of influenza-like illness (ILI) cases in Belgium for the seasons 2010–2011 to 2015–2016. The normalized ILI data are shown from week 35 of year i to week 34 of year $i + 1$.

2. Methods

2.1. Data

Fig. 1 shows the weekly ILI incidence in Belgium throughout the seasons 2010–2011 to 2015–2016. These data were obtained from Sciensano which supports health-policy-making in Belgium through research (Sciensano, 2017). The reported values are normalized with respect to the catchment population of the 43 Belgian districts, which are administrative regions of that country. The catchment population represents the average number of patients that visits the SGP network per district. This normalization is carried out for each district as follows: the weekly ILI number is divided by the catchment population of the corresponding district, and, then multiplied by its total population. It was performed in order to extrapolate the ILI cases in the catchment population of each district to the district's total population. We will refer to the resulting data as the *normalized ILI data*.

In this paper, week 35 is considered as the first week of a new season. Therefore, a season runs from week 35 of year i to week 34 of year $i + 1$. We can observe the seasonality of the weekly ILI incidence during the last six seasons in Fig. 1. Yet, there is considerable variability among the epidemic seasons for what concerns the peak of the ILI curve, the duration of the season and the week in which the highest weekly ILI incidence is reported.

The highest ILI incidence was observed in seasons 2012–2013 and 2014–2015. Accordingly the epidemics during these seasons were classified by Sciensano as seasons of high intensity with an ILI consultation rate of about 1000 per 100,000 inhabitants (WIV-ISP, 2017). The duration of these epidemics was also longer as compared to other seasons. Season 2015–2016 may also be considered as a high intensity epidemic, but with a lower ILI peak than in seasons 2012–2013 and 2014–2015. Seasons 2010–2011 and 2011–2012 have a similar behavior, except for what concerns their onset, which occurred much later for the latter. Still, both were characterized by moderate epidemics. Season 2013–2014 is unique among the analyzed seasons because the corresponding influenza epidemic was mild and of short duration. Table 1 summarizes the main characteristics of the epidemics during the studied seasons in terms of their onset, duration, the week of the highest ILI incidence, the consultation rate (the number of consultations per 100,000 inhabitants), the epidemic threshold calculated by the European Centre for Disease Prevention and Control (ECDC) using the MEM (Moving Epidemic Method) (Vega et al., 2013, 2015) method and their intensity. The epidemic threshold accounts for the weekly ILI incidence per 100,000 inhabitants.

Table 1
Characteristics of the Belgian influenza seasons 2010–2011 to 2015–2016. The onset is the first week at which the epidemic threshold was exceeded. This threshold is provided by ECDC and calculated using the MEM method.

Season	Onset (week/year)	ILI peak (week/year)	Duration (weeks)	Consult. rate (consult/10^5 inhabitants)	Epidemic threshold (ILI/10^5 inhabitants)	Epidemic intensity
2010–2011	50/2010	01–04/2011	12	500	148	Moderate
2011–2012	05/2012	07/2012	6	531	138	Moderate
2012–2013	52/2012	06/2013	12	1000	139	High
2013–2014	06/2014	09/2014	6	311	141	Mild
2014–2015	02/2015	06/2015	10	979	140	High
2015–2016	04/2016	09/2016	10	734	143.9	Moderate

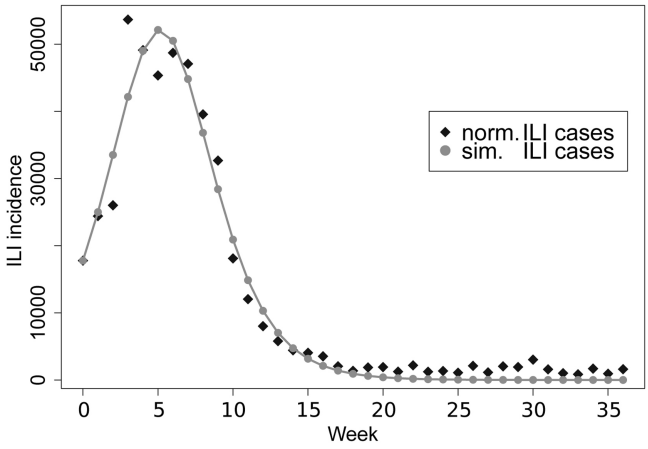

Fig. 2. Normalized weekly incidence of ILI cases (norm. ILI cases) and the SIR model solution representing the simulated counterpart (sim. ILI cases) for the season 2010–2011. Week 0 of the plot corresponds to week 50 of 2010, which is the onset of the corresponding season.

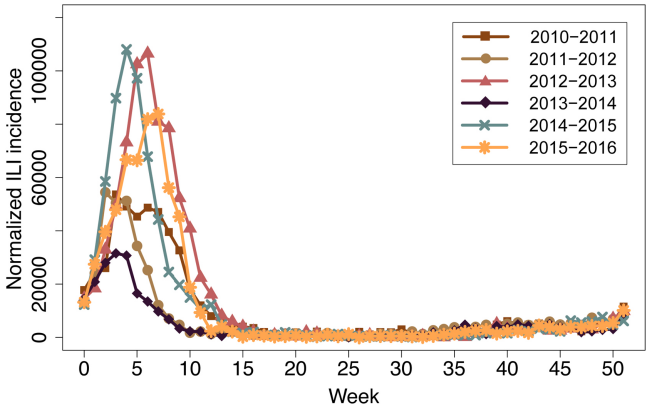

Fig. 3. Normalized weekly incidence of ILI cases for every season starting from the onsets (week 0) given in Table 1.

2.2. The SIR model

The SIR model is an established compartmental model that can be used to describe infectious disease dynamics (Kermack and McKendrick, 1927). It assumes that the individuals of a population can be placed in 3 compartments, namely Susceptible (S) individuals, those able to contract the infection; Infected (I) individuals, those able to propagate the infection, and Recovered or Removed (R) individuals, who recovered from the infection or died. There are many variations of this model (Allen, 1994; Anderson and Robert, 1991; Hethcote, 2000). In such models, the total population is typically fixed, and hence they neglect processes such as migration and birth. This is, however, not an issue because an epidemic usually evolves much faster than birth-death processes (Hethcote, 2000). The SIR model can be described by the following system of ordinary differential equations (Kermack and McKendrick, 1927):

$$\frac{dS}{dt} = -\frac{\beta IS}{N}, \tag{1}$$

$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I, \tag{2}$$

$$\frac{dR}{dt} = \gamma I, \tag{3}$$

where β is the transmission rate, γ the recovery rate and N the size of the total population. Since the latter is constant, we have $S(t) + I(t) + R(t) = N$. If the basic reproduction number, denoted R_0, and equalling $\frac{\beta}{\gamma}$, is larger than 1, then the number of infected individuals grows exponentially. R_0 is also defined as the expected number of new infections derived from a single infection.

The SIR and related models have been used by others to describe the dynamics of influenza, and other infectious diseases such as measles (Grenfell, 1992; Grenfell et al., 2002) and whooping cough (Lavine et al., 2011). In order to show that the SIR model is able to capture the dynamics of the Belgian influenza epidemics, we calibrated parameters β and γ in Eqs. (1)–(3) for each season separately using all available ILI data for that particular season by means of the Nelder-Mead method (Nelder and Mead, 1965) for minimizing the sum of squared differences between the normalized and the simulated weekly incidence of ILI cases. Fig. 2 presents the normalized and simulated weekly ILI incidence for season 2010–2011. For this season, the simulated period started in week 50 of 2010 in accordance with the onset reported in Table 1. Therefore, week 0 corresponds to week 50 of this season. We note that the normalized and simulated ILI dynamics are in very close agreement, while also the peak of the weekly ILI incidence is accurately described.

2.3. Real-time prediction of ILI incidence

2.3.1. The confidence region

Despite the variability regarding the epidemic onset, the seasonality of influenza epidemics is one of their most distinctive features. For that reason, time series are often used in the analysis of ILI incidence due to its periodic behavior (Kane et al., 2014; Soebiyanto et al., 2010; Held and Paul, 2012; Viboud et al., 2003). In this paper, we have at our disposal a time series of ILI incidence for every epidemic season, which can be shifted in such a way that week 0 corresponds to the onset, as such allowing for the mutual comparison of different seasons on a common basis, as shown in Fig. 3. This on its turn allows to characterize the similarities among the seasons, model previous epidemics, and underpin predictions for upcoming seasons. We will refer to the onset of an epidemic season as week 0. The onsets of the investigated seasons are shown in Table 1.

In order to obtain a single construct that captures the variability among the studied seasons, we first account for the variability among the weekly ILI incidence, denoted by n, at a given week across the seasons. For this purpose, we calculate the mean weekly ILI incidence,

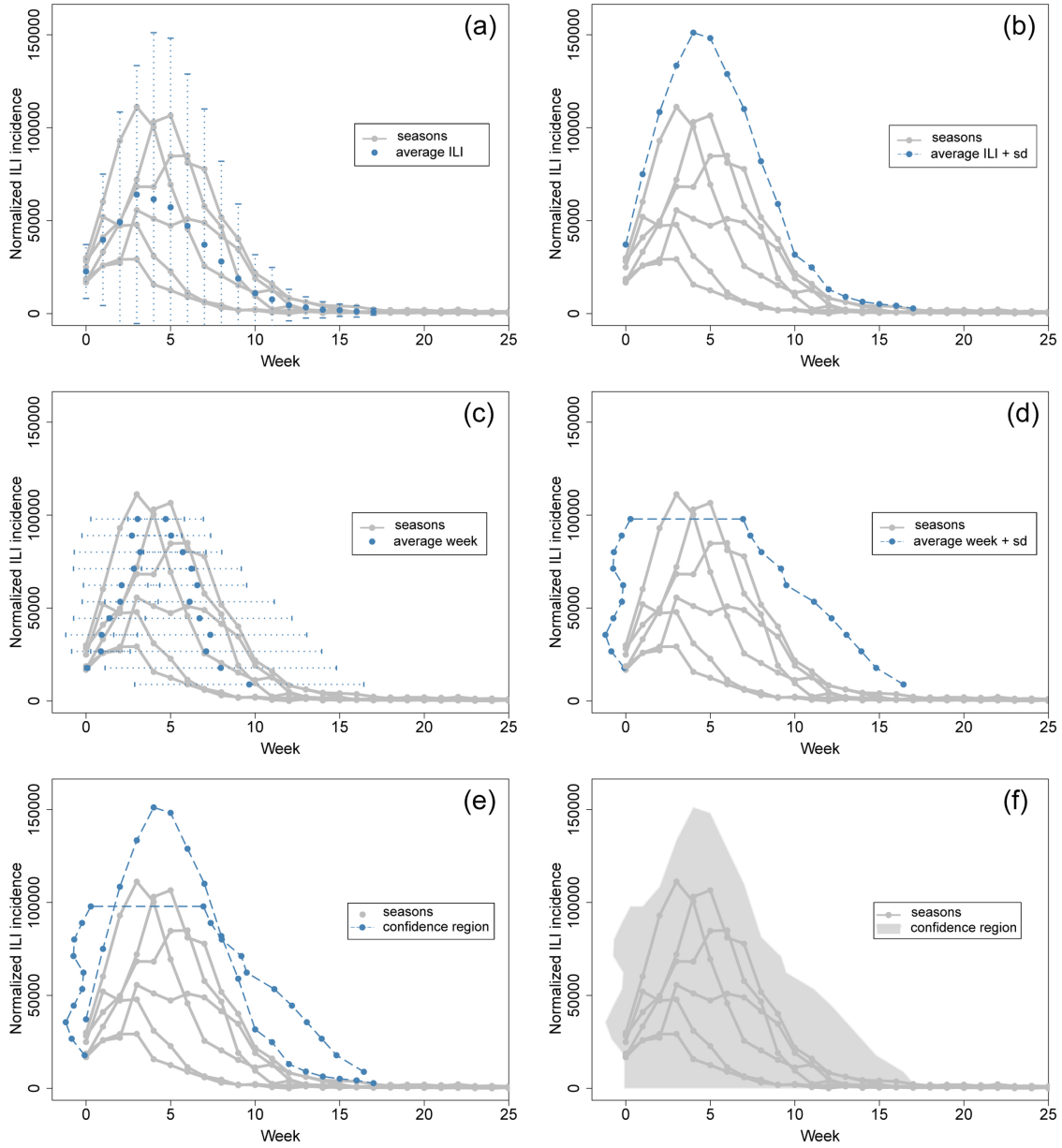

Fig. 4. Construction of the confidence region. (a) Average weekly incidence of ILI cases, μ_n, and the corresponding standard deviation (sd), σ_n. (b) Boundary of the confidence region along the vertical axis defined by $\mu_n + \sigma_n$. (c) Average time of occurrence per ILI incidence value, μ_t, and the corresponding standard deviation (sd), σ_t. (d) Boundary of the confidence region along the horizontal axis defined by $\mu_t + \sigma_t$. (e) Union of the envelopes defined in (b) and (d). (f) Final confidence region.

denoted by μ_n and the corresponding standard deviation σ_n. Then, $\mu_n + \sigma_n$ is used to define the boundaries of the so-called confidence region along the vertical axis (Fig. 4(a) and (b)). Similarly, we analyze the variability in time regarding the week in which a specific n was observed. For that purpose, the interval between the minimum and the maximum weekly ILI incidence (n) among the seasons was sampled to obtain equally spaced ILI incidence values along the vertical axis. Then, for every season separately, linear interpolation was used to retrieve the time of occurrence, denoted by t, of every such ILI incidence value. Subsequently, the resulting weeks were averaged per ILI incidence value across the seasons, as such leading to μ_t and σ_t, after which $\mu_t + \sigma_t$ was used to define the boundaries of the confidence region along the horizontal axis (Fig. 4(c) and (d)).

In a final step (Fig. 4(e)), the two regions are merged to a single region that accounts for the variability among both the weekly ILI incidence in a certain week and the time of occurrence of a certain weekly ILI incidence. This was achieved by taking the union of the envelopes shown in Fig. 4(b) and (d). The final confidence region on the basis of the ILI data collected throughout all studied seasons is shown in Fig. 4(f). Note that both σ_n and σ_t can be scaled in order to account for less or more variability in the magnitude of the weekly incidence of ILI cases and the time at which these occur.

2.4. Prediction of ILI incidence

Once the confidence region is constructed, it can be used together with the calibrated SIR model to predict the weekly ILI incidence from week i on, given data up to week $i-1$ for the current season.

Let us denote the epidemic onset of season j as w_0^j, then in week w_0^j ILI data of season j up to that week are available, so the reported weekly ILI incidence from some week w_{-m}^j up to week w_0^j can be used for predicting the weekly ILI incidence from week w_1^j on. This is achieved by using data from weeks w_{-m}^j up to w_0^j to calibrate the parameters of Eqs. (1)–(3) using the Nelder-Mead optimization algorithm (Nelder and

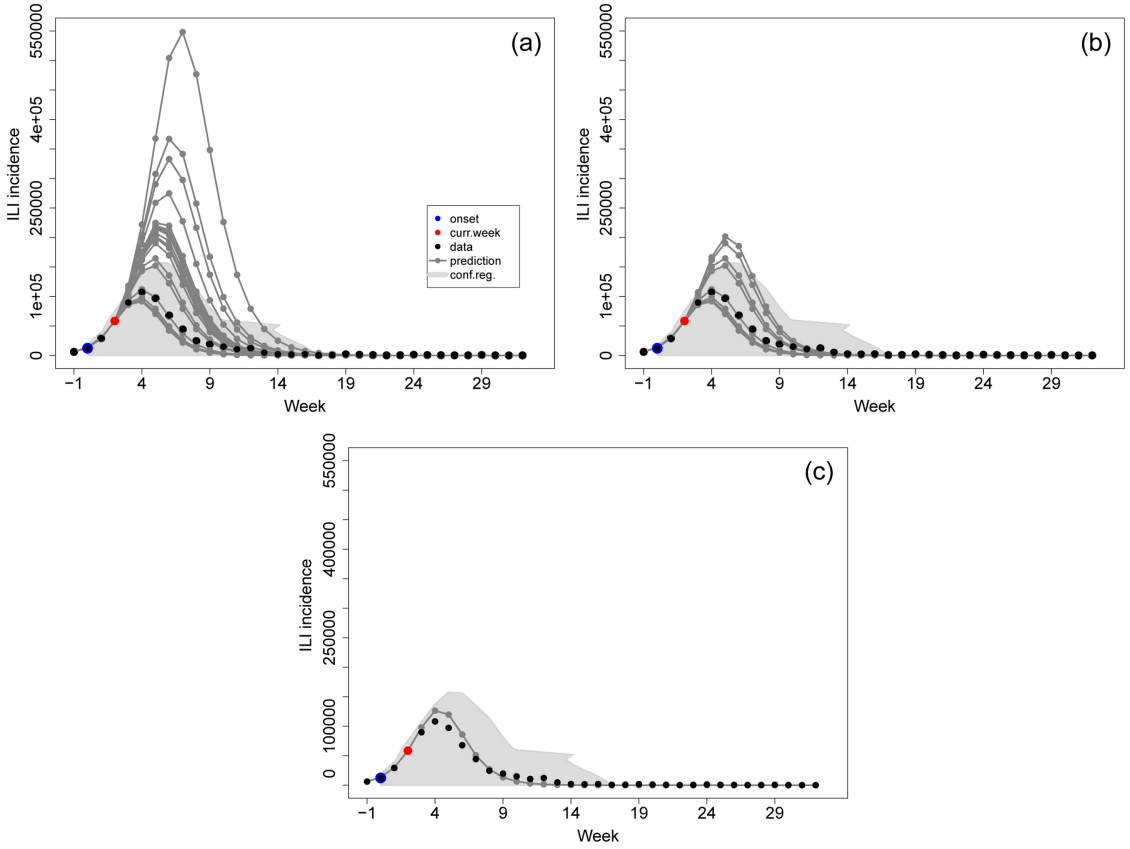

Fig. 5. Selection of acceptable model parameters. (a) ILI incidence curves obtained for 30 runs of the optimization algorithm starting from 30 initial guesses of β and γ. (b) ILI incidence curves whose prediction for week $i + 1$, after current week (curr. week), red dot, falls within the confidence region (conf. reg.). (c) ILI incidence curve generated using the average of the parameter values ($\hat{\beta}$ and $\hat{\gamma}$) that give rise to ILI incidence curves falling within the confidence region. The simulated periods start one week before the onset (week 0). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

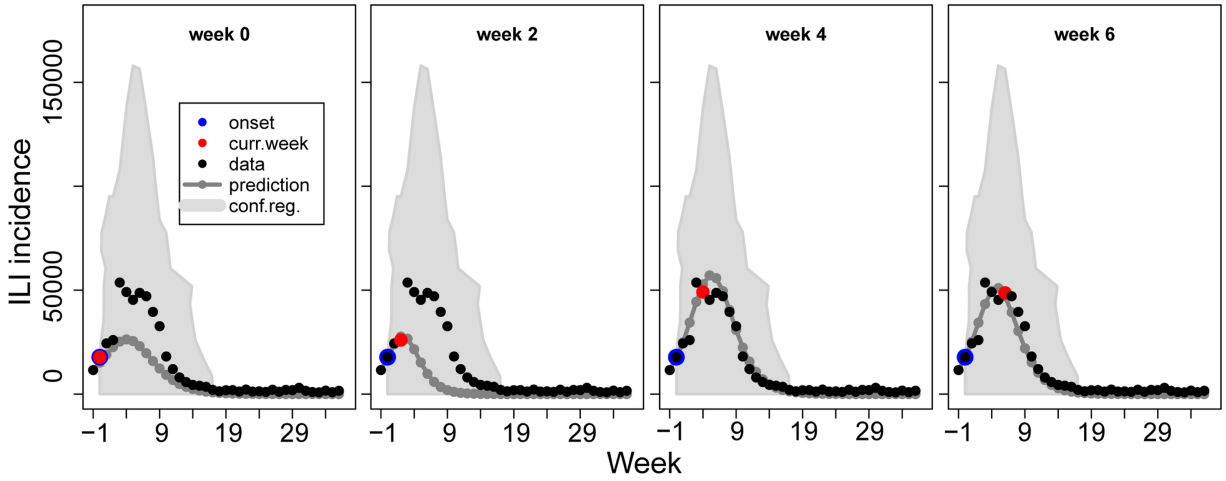

Fig. 6. Predicted weekly ILI incidence (gray) throughout season 2010–2011 for different starting weeks of the simulated period, together with the normalized ILI data (black). The epidemic onset and the start of the simulated period, the current week (curr. week), are shown in blue and red, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Mead, 1965). Given the heuristic nature of this optimization algorithm, initial guesses are needed for the parameters β and γ to initialize the search. Yet, these affect the values of the calibrated parameters, with consequent effects on the predicted weekly ILI incidence. In order to account for this issue, the confidence region is used to select meaningful parameter sets. More specifically, for each initial guesses of β and γ and given ILI data up to week i, it is verified whether the predicted ILI incidence for week $i + 1$ that is obtained using the resulting calibrated model falls within the confidence region. If so, the corresponding parameter set is added to the set of acceptable parameters sets.

Fig. 5(a) shows all ILI curves for season 2014–2015 corresponding to 30 random initial guesses of β between 0 and 12.5, and γ between 0 and β, therefore, each curve corresponds to a single run of the optimization algorithm. The simulated periods start one week before the onset (week 0). In Fig. 5(b), we see only those curves whose prediction for the subsequent week in relation to the current week (red dot) falls

Table 2
Estimated model parameters β and γ, mean absolute error (MAE), root mean squared error (RMSE) and Pearson correlation coefficient between the normalized and simulated weekly ILI incidence for seasons 2010–2011 to 2015–2016.

Season	$\hat{\beta}$	$\hat{\gamma}$	MAE	RMSE	Corr.
2010–2011	4.635539	4.280132	2091.08	3035.25	0.985
2011–2012	8.376870	7.657043	1825.54	2722.60	0.989
2012–2013	4.231897	3.722846	2072.34	3448.29	0.994
2013–2014	8.607093	8.152276	860.92	1252.99	0.993
2014–2015	6.068778	5.288085	2609.47	3982.24	0.994
2015–2016	4.714837	4.207067	2676.26	4479.31	0.986

within the confidence region constructed on the basis of the ILI data of the other seasons. The parameter values corresponding to the latter curves are averaged ($\hat{\beta}$ and $\hat{\gamma}$) and the resulting ILI incidence curve for $\hat{\beta}$ and $\hat{\gamma}$ is plotted (Fig. 5(c)).

Fig. 6 depicts the predicted weekly ILI incidence for season 2010–2011 when increasingly more ILI data become available for that season and after 300 runs of the optimization algorithm for each plot. The shaded area corresponds to the confidence region that was constructed on the basis of the other seasons. In the first plot of Fig. 6, for instance, the simulated period starts at the onset of the 2010–2011 epidemic (Table 1) and only ILI data from weeks 0 and -1 were used for calibrating the model parameters. As ILI data of more weeks become available, these can then be used for model recalibration. The remaining plots of Fig. 6 present similar results for different starting weeks of the simulated period.

3. Results

3.1. Analysis of the SIR model for the Belgium epidemic seasons

Table 2 lists for every season the calibrated SIR model parameters, together with the mean absolute error (MAE) and root mean squared error (RMSE) quantifying the discrepancies between the normalized and simulated weekly ILI incidence, and the Pearson correlation between those. From this table it is clear that the SIR model accurately describes the dynamics of the seasonal influenza epidemics in Belgium, but also that the values of the calibrated parameters β and γ vary considerably across the seasons. Still, the MAE and RMSE are relatively small compared to the total Belgian population size during the studied seasons. Besides, the Pearson correlation coefficients are always very high, indicating a distinct linear relationship between the normalized and simulated weekly ILI incidence.

As explained before, the epidemic onsets of the consecutive seasons are identified with the first week in which the epidemic threshold, as defined by Sciensano, was exceeded. This is an important point because a more accurate description of an epidemic can be obtained as the start of the simulated period approaches the true onset. Therefore, knowing the epidemic onset is a prerequisite for reliably simulating influenza dynamics. Moreover, it gives a common ground to mutually compare seasonal influenza epidemics by shifting the outbreak curves in such a way that week zero corresponds to the onset given in Table 1. From Fig. 3 it is once more clear that there exists a considerable variability among the studied seasons in terms of both the weekly ILI incidence and the week in which, for instance, the highest ILI incidence was observed.

3.2. Accuracy of the proposed approach

Our dataset consists of time series of ILI incidence during the seasons 2010–2011 to 2015–2016. We evaluated the accuracy of the proposed approach in predicting the weekly ILI incidence using cross-validation where each season was considered once as the "current" season, i.e., the validation season, while the remaining seasons were used for constructing the confidence region and hence made up the calibration set. The final model parameters are the averages of those obtained at the end of 300 runs of the optimization algorithm, excluding the parameter sets that led to ILI incidence curves not enclosed by the confidence region. We adopted the average population size in Belgium from 2010 up to 2016[1] as the value of N in Eqs. (1)–(3). Fig. 7 presents the predicted weekly ILI incidence during season 2010–2011 using data from season 2011–2012 up to 2015–2016 for different starting weeks of the simulated period. In this case, the confidence region was generated using ILI data from the latter seasons. From now on, we will denote as simulated period the time lapse between the first week of which ILI data are used and the last week for which the weekly ILI incidence is predicted.

Each plot in Fig. 7 demonstrates the agreement between the normalized and predicted weekly ILI incidence. As an example, consider Fig. 7(a), where only ILI data up to the epidemic onset are available. The start of the simulated period is initially set to -1, i.e., one week before the onset and the predictions are obtained from week 1 on. As ILI data for a next week arrive, the model can be recalibrated, which is visualized by the plot sequence in Fig. 7(a). Similar results are presented in Fig. 7(b) and (c), but for a simulated period starting in weeks -2 and -3, respectively. From Fig. 7 we infer that the start of the simulation period does not significantly impact the obtained results. Yet, using data up to week 3, the best agreement between the simulated and the normalized weekly ILI incidence was obtained if the simulated period started one week before the epidemic onset (Fig. 7(a)).

Fig. 8 presents similar outputs, but now for season 2014–2015, which was of high intensity (Fig. 1). Together with 2012–2013, these two seasons have a great influence on the shape of the upper boundary of the confidence region. For season 2014–2015, the predictions obtained using data up to the peak (weeks 4 and 5) are not influenced by the choice of the start of the simulated period (Fig. 8(a)–(c)), but the number of weeks used before the onset has a greater influence.

One issue that was observed for season 2014–2015 was the low number of sets of model parameters leading to curves falling within the confidence region. When starting three weeks before the onset there was no parameter set leading to predictions up to week 4 (Fig. 8(c)), and, the success rate is less than 3% when using data up to weeks 2 and 3. Moreover, when using data up to five or more weeks, more than 96% of sets of model parameters passed the check. This is due to the fact that season 2014–2015 has one of the highest peaks of ILI cases among the six seasons (Fig. 3), and, therefore, since it is left out of the calibration set for the construction of the confidence region (generated with the remaining seasons), the latter may not cover any predictions for some weeks of this season. However, since there is another season (season 2012–2013) with a similar peak, although with a different onset, and which was used to construct the confidence region, it was possible to obtain predictions for the majority of the weeks in season 2014–2015. However, when considering season 2012–2013 as the validation set, we could obtain predictions for all the weeks. This can be explained by the fact that the peak of this season was observed a few weeks later when compared to the peak of season 2014–2015, and, there is much more variability of the ILI incidence among the seasons after the peak (Fig. 3). Therefore, the descending part of the confidence region generated for season 2012–2013 covers predictions for all the weeks.

An overview of the predicted weekly ILI incidence for the other seasons can be seen in Fig. 9. Fig. 9(a) presents the normalized ILI cases for season 2011–2012 starting from week -1, while Fig. 9(b)–(d) present the same number for seasons 2012–2013 up to 2015–2016, respectively, starting from week -2. In general, moderate and mild seasons lead to ILI incidence curves that are enclosed by the confidence

[1] Source: Directorate-general Statistics - Statistics Belgium (https://statbel.fgov.be/en).

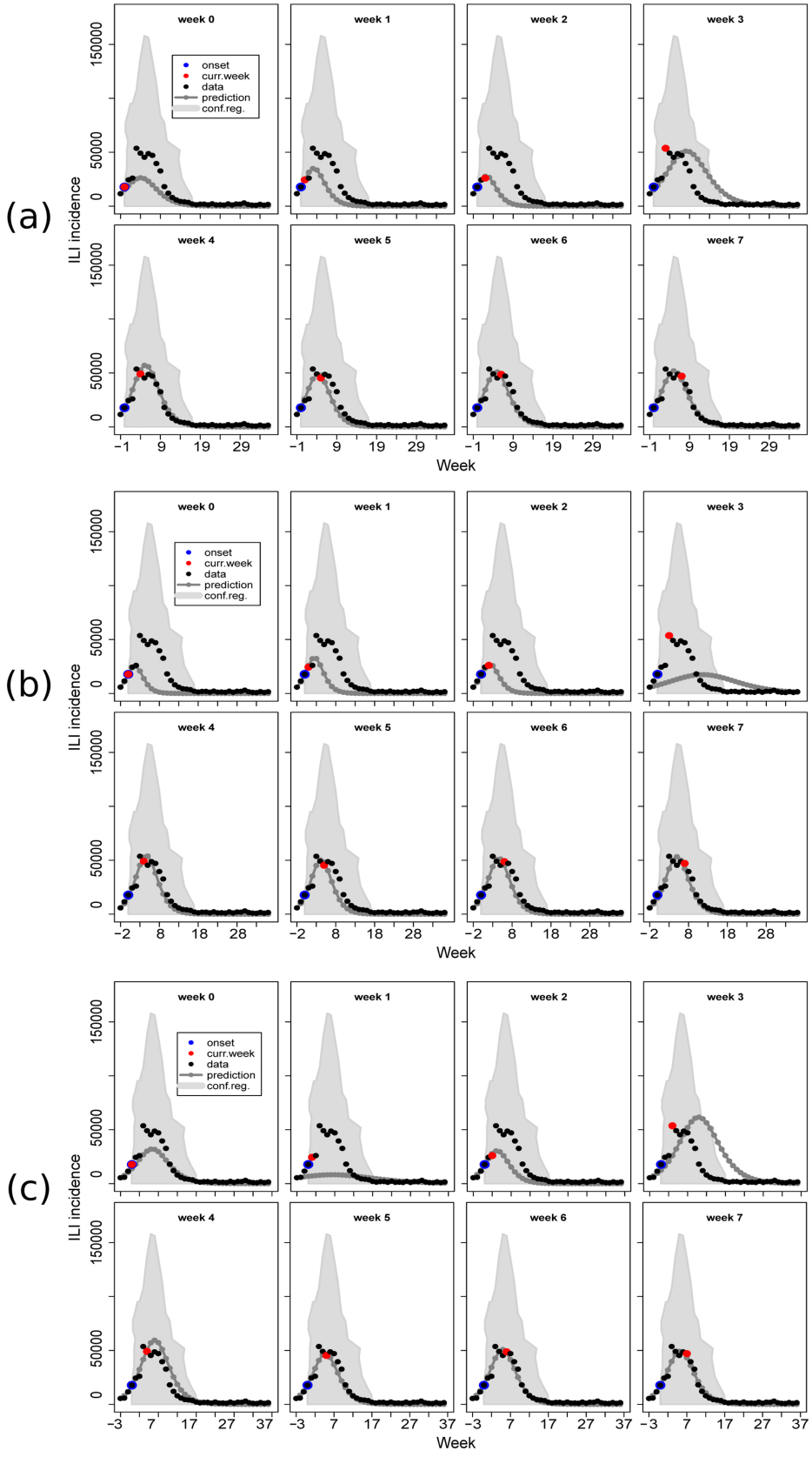

Fig. 7. Simulated (gray) and normalized (black) weekly ILI incidence in Belgium for season 2010–2011. The blue dot corresponds to the epidemic onset and the red dot to the current week (curr. week) up to which the normalized ILI data were used for model calibration. Simulated period starting (a) the week before the onset, (b) 2 weeks before the onset and (c) 3 weeks before the onset. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

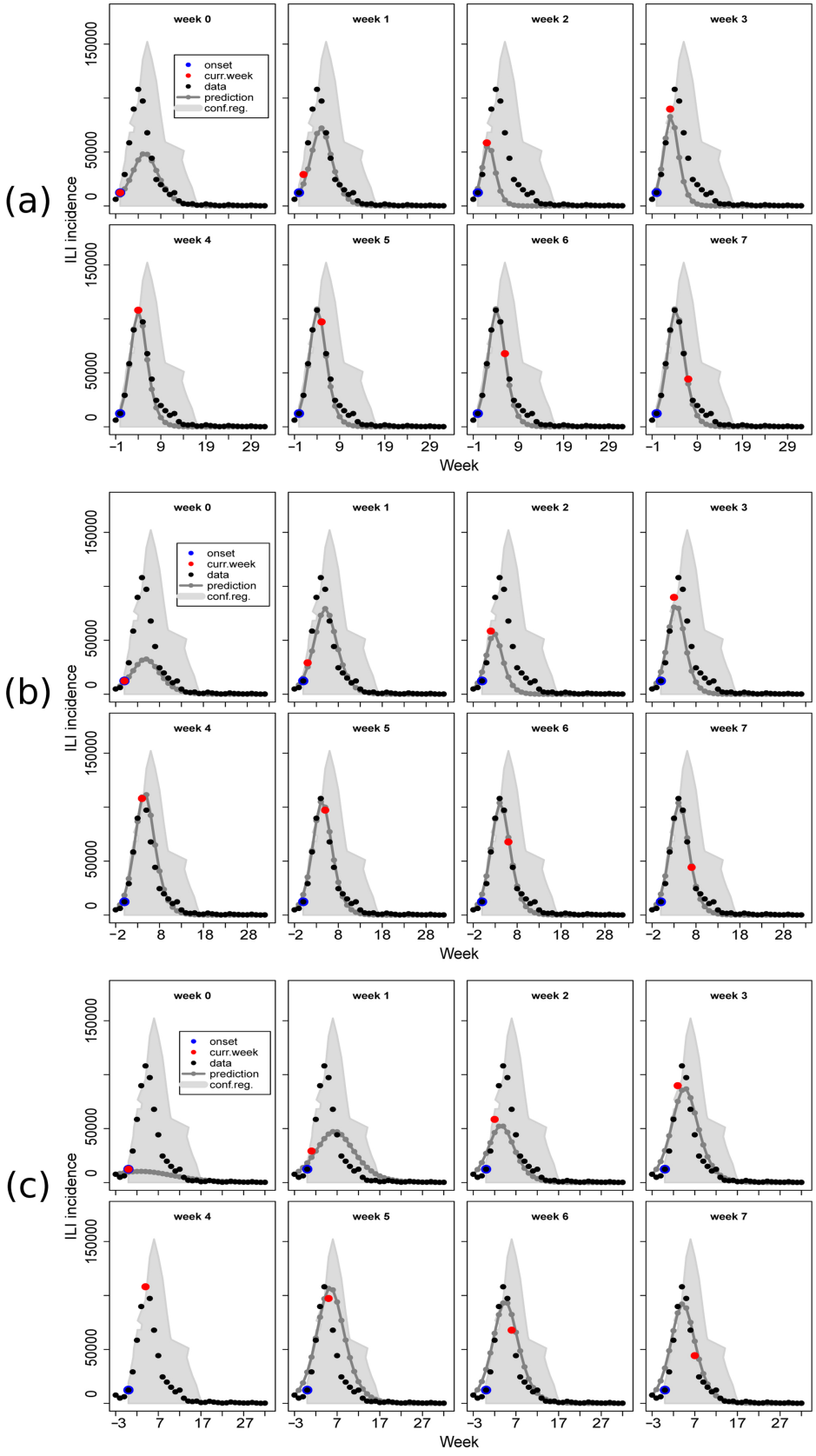

Fig. 8. Simulated (gray) and normalized (black) weekly ILI incidence in Belgium for season 2014–2015. The blue dot corresponds to the epidemic onset and the red dot to the current week (curr. week) up to which the normalized ILI data were used for model calibration. Simulated period starting (a) the week before the onset, (b) 2 weeks before the onset and (c) 3 weeks before the onset. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

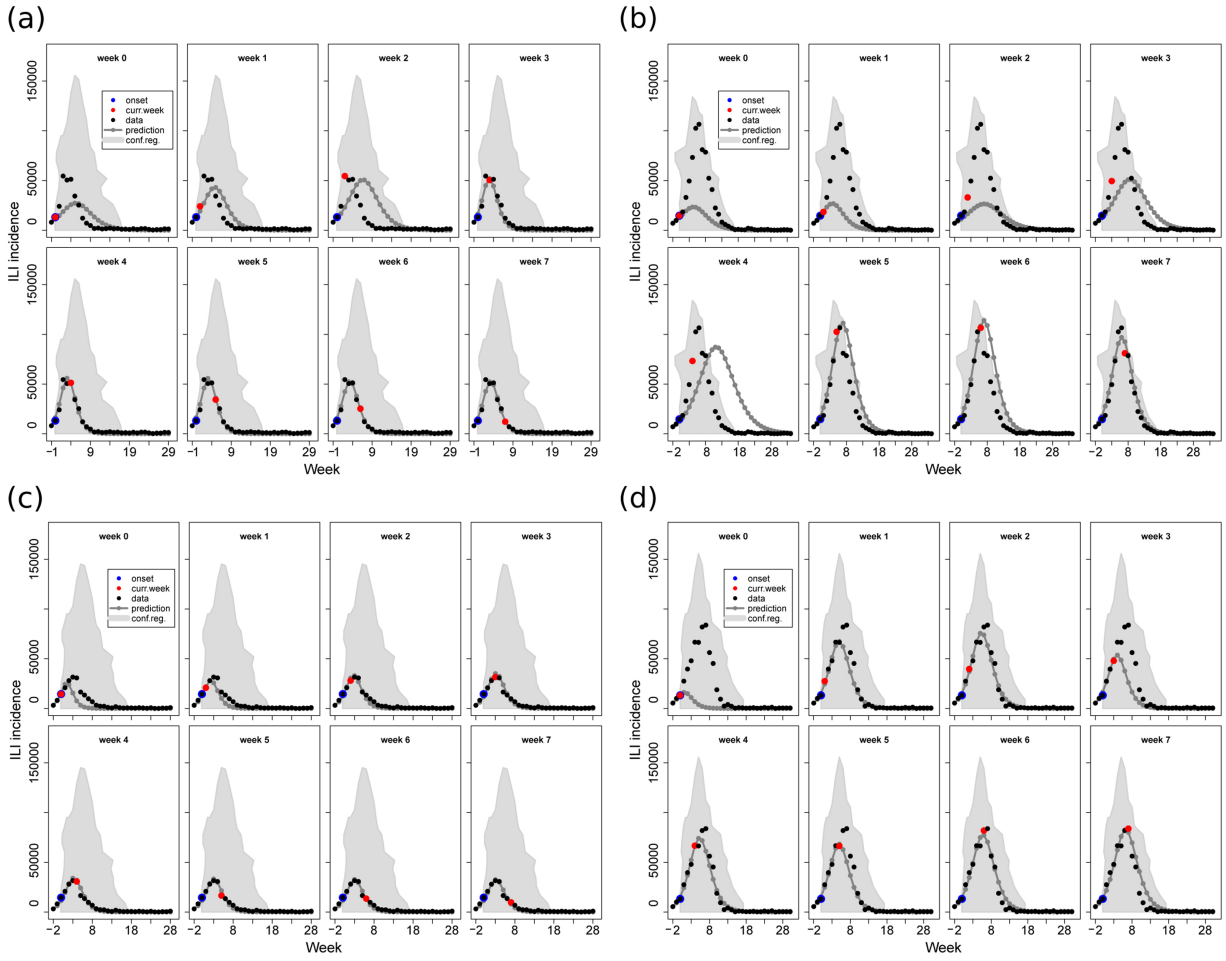

Fig. 9. Simulated (gray) and normalized (black) weekly ILI incidence in Belgium for different epidemic seasons. The blue dot corresponds to the epidemic onset and the red dot to the current week (curr. week) up to which the normalized ILI data were used for model calibration. (a) Prediction for season 2011–2012 starting the simulated period from the week before the onset. (b–d) Prediction for seasons 2012–2013, 2013–2014 and 2015–2016, respectively, starting the simulated period from 2 weeks before the onset. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

region, since the high intensity seasons are the ones that determine the boundaries of the confidence region.

3.3. Error analysis

Fig. 10 shows the mean absolute error (MAE) and the root mean square error (RMSE) between the normalized and the simulated weekly ILI incidence, as well as the cross-correlation considering different time lags. Each row in this figure corresponds to a different season and each column to a different starting week. Each dot in the plots corresponds to the average error over 10 predicted ILI incidence values (weeks 0–9), together with the corresponding standard deviation.

As expected, the smallest errors were obtained for a zero time lag for the majority of the seasons and starting weeks, which indicates that there is no shift between the time series of the normalized and the predicted values. In addition, the highest correlation was also observed for time lag zero. Yet, when comparing the errors across the consecutive seasons we see that the largest errors were obtained for season 2014–2015. This corroborates the results shown in Fig. 8, since this season defines the upper boundary of the confidence region. The error analysis also shows that the best agreement between the normalized and predicted weekly ILI incidence was obtained for season 2013–2014. Moreover, in general, the start of the simulated period did not affect the error indices and correlation coefficients.

4. Discussion

The epidemic threshold is provided by ECDC and used by several European countries in order to increase international comparability. This way, in a weekly-reporting system of ILI cases, as in Belgium, the epidemic onset can be rapidly tracked. However, there is a great variability among different epidemic seasons for what concerns the behavior of the epidemic after its onset. This variability concerns for instance the time to reach the peak and the duration of the epidemic. Providing weekly predictions makes it possible to capture variations in the ILI incidence, on the basis of which public health strategies can be updated and or revised. Many related studies discuss how purely mechanistic models, such as the SIR model, can be improved in order to arrive at more accurate and more reliable predictions. Among the proposed methods, some rely on filtering or data assimilation methods, which make use of recursive approaches so that the evolving characteristics of the epidemic can also be modeled and the model parameters can be dynamically updated (Yang et al., 2014; Ray and Reich, 2018). This is the case, for instance, of methods relying on the Ensemble Kalman Filter (EnKF) or the Ensemble Adjustment Kalman Filter (EAFK) (Shaman and Karspeck, 2012) and particle filtering (Huang and Provan, 2016). However, the optimal method also depends on the final application (Reich et al., 2018).

In comparison, the method presented in this paper makes use of a confidence region generated from data of other epidemic seasons that is used as a filter for the weekly SIR predictions. Therefore, it accounts for

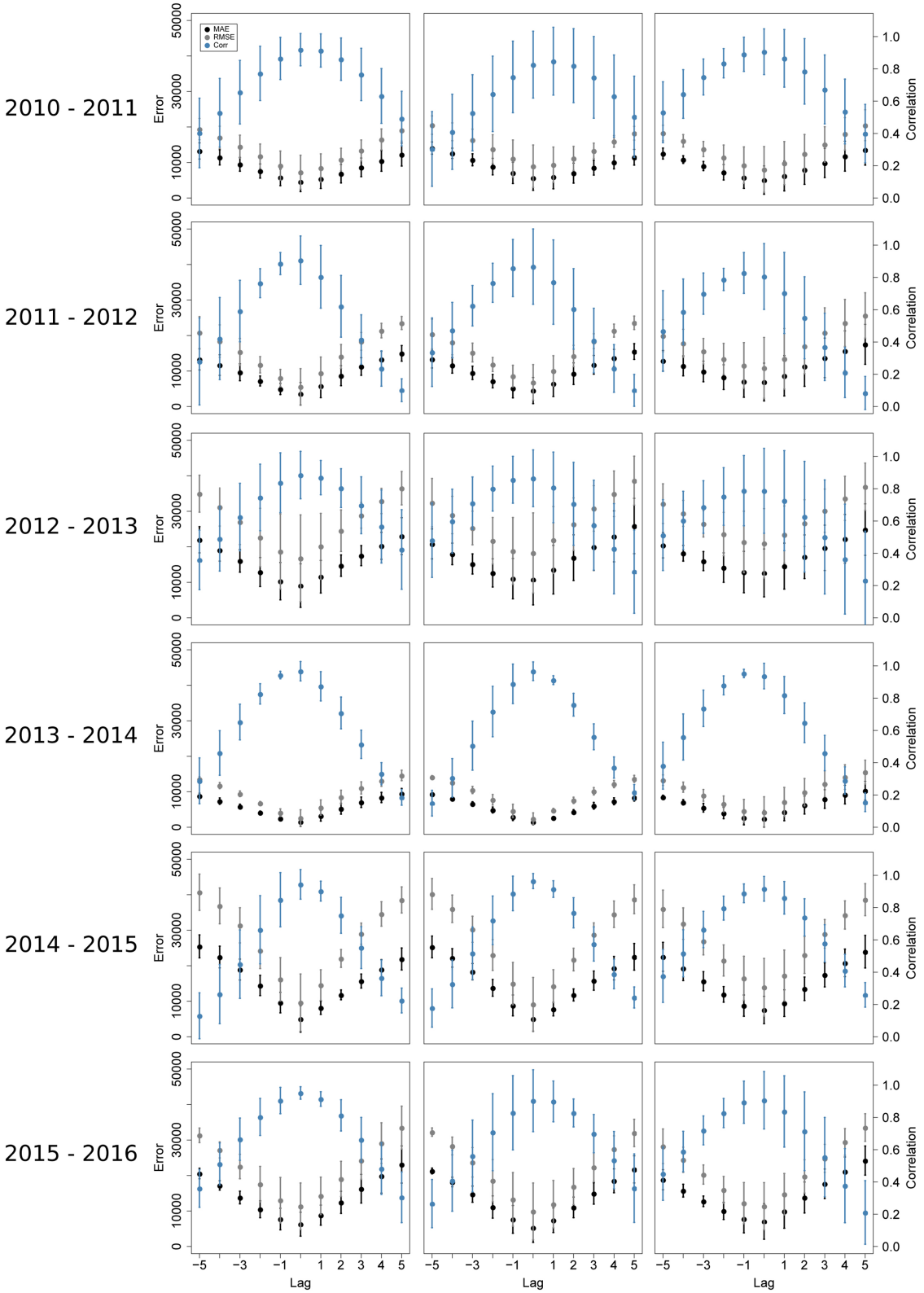

Fig. 10. MAE (mean absolute error), RMSE (root mean square error) and correlation between the normalized and the predicted weekly ILI incidence per season (rows) and per starting week of the simulated period (columns) for different time lags.

the statistical properties of the epidemic data. The proposed approach can be used for the real-time prediction of the weekly ILI incidence, where the model parameters could be updated as new counts of the number of ILI cases become available. We demonstrated the applicability of the proposed method for what concerns the seasonal influenza epidemics in Belgium during seasons 2010–2011 to 2015–2016. It takes into account the records of seasonal influenza epidemics in Belgium using a confidence region, and in this way allows to incorporate the variability existing among the consecutive influenza seasons. Clearly, as ILI data from more seasons become available, the more representative

this confidence region becomes. Still, if the ILI incidence curve of a new season lies close to and above the one(s) that define(s) the boundaries of the confidence region, then the predicted weekly ILI incidence for that particular season might be underestimated. Likewise, seasons characterized by a steep ILI incidence curve can also be hard to predict. This problem can, however, be partially overcome by scaling the standard deviations σ_n and σ_t by accounting for the amount of variability in the magnitude of the weekly ILI incidence and the time at which these occur. Moreover, the SIR model is calibrated dynamically every week as new ILI data from the SGP becomes available, such that the predicted ILI curve gets rapidly tuned to the dynamics of the ongoing season.

The proposed approach can be used as a decision-supporting tool for medical centers and health-care services, as it yields scientifically underpinned predictions of the weekly ILI incidence throughout the ongoing epidemic season. Its implementation is straightforward, few computational resources are needed and it is based on an established modeling framework. This implies that it could as well be used for predicting the weekly ILI incidence in other countries that suffer from seasonal influenza outbreaks, provided that the confidence region is constructed on the basis of ILI data of those countries. For what concerns the shift of the seasons to their onset, we have demonstrated that this procedure allows to characterize the behavior of the epidemics despite the temporal displacement of the beginning of the epidemic observed for the different seasons. Therefore, the confidence region accounts only for the epidemic period. As a future analysis, when estimating the parameters, it would be useful to take into account the circulating strains and their match (or mismatch) with the vaccine of a specific influenza season, since this may have a contribution to the dynamics of the epidemic.

Acknowledgments

The authors thank all participating General Practitioners of the sentinel network and the support from Sciensano, in Belgium, regarding the data acquisition process. This work has been supported by the São Paulo Research Foundation (FAPESP) [#2015/05899-7 and #2018/00147-5]; the Coordination for the Improvement of Higher Education Personnel (CAPES); the Special Research Fund (BOF) of Ghent University, Belgium [#01SF0516]; and the National Council for Scientific and Technological Development (CNPq), Brazil [#307797/2014-7].

References

Allen, L.J., 1994. Some discrete-time SI, SIR, and SIS epidemic models. Math. Biosci. 124, 83–105.
Anderson, R.M., Robert, M., 1991. Infectious Diseases of Humans: Dynamics and Control. Oxford University Press.
Brauer, F., 2008. Compartmental Models in Epidemiology. Springer, Berlin, Heidelberg, pp. 19–79.
Chretien, J.P., George, D., Shaman, J., Chitale, R.A., McKenzie, F.E., 2014. Influenza forecasting in human populations: a scoping review. PLoS One 9, e94130.
Dugas, A.F., Jalalpour, M., Gel, Y., Levin, S., Torcaso, F., Igusa, T., Rothman, R.E., 2013. Influenza forecasting with Google flu trends. PLoS One 8, e56176.
ECDC Summary, 2015. Summarising the 2014–2015 influenza season in Europe. (accessed 14 March 2018). https://ecdc.europa.eu/en/news-events/summarising-2014-2015-influenza-season-europe.
Generous, N., Fairchild, G., Deshpande, A., Del Valle, S.Y., Priedhorsky, R., 2014. Global disease monitoring and forecasting with Wikipedia. PLoS Comput. Biol. 10, e1003892.
GFT, 2017. Google Flu Trends. (accessed 27 July 2017). https://www.google.org/flutrends/about/.
Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S., Brilliant, L., 2009. Detecting influenza epidemics using search engine query data. Nature 457, 1012–1014.
Grenfell, B., 1992. Chance and chaos in measles dynamics. J. R. Stat. Soc. Ser. B (Methodological) 54, 383–398.
Grenfell, B.T., Bjørnstad, O.N., Finkenstädt, B.F., 2002. Dynamics of measles epidemics: scaling noise, determinism, and predictability with the TSIR model. Ecol. Monogr. 72, 185–202.
Held, L., Paul, M., 2012. Modeling seasonality in space-time infectious disease surveillance data. Biom. J. 54, 824–843.
Hethcote, H.W., 2000. The mathematics of infectious diseases. SIAM Rev. 42, 599–653.
Huang, W., Provan, G., 2016. An improved state filter algorithm for SIR epidemic forecasting. ECAI 2016: 22nd European Conference on Artificial Intelligence. IOS Press.
Kane, M.J., Price, N., Scotch, M., Rabinowitz, P., 2014. Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks. BMC Bioinform. 15, 276.
Kermack, W.O., McKendrick, A.G., 1927. A contribution to the mathematical theory of epidemics. Proc. R. Soc. Lond. A: Math. Phys. Eng. Sci. 115, 700–721.
Lavine, J.S., King, A.A., Bjørnstad, O.N., 2011. Natural immune boosting in pertussis dynamics and the potential for long-term vaccine failure. Proc. Natl. Acad. Sci. 108, 7259–7264.
McIver, D.J., Brownstein, J.S., 2014. Wikipedia usage estimates prevalence of influenza-like illness in the united states in near real-time. PLoS Comput. Biol. 10, e1003581.
Nelder, J.A., Mead, R., 1965. A simplex method for function minimization. Comput. J. 7, 308–313.
Nsoesie, E., Mararthe, M., Brownstein, J., 2013. Forecasting peaks of seasonal influenza epidemics. PLoS Curr. 5.
Nsoesie, E.O., Brownstein, J.S., Ramakrishnan, N., Marathe, M.V., 2014. A systematic review of studies on forecasting the dynamics of influenza outbreaks. Influenza Other Respir. Viruses 8, 309–316.
Paul, M.J., Dredze, M., Broniatowski, D., 2014. Twitter improves influenza forecasting. PLoS Curr. 6.
Ray, E.L., Reich, N.G., 2018. Prediction of infectious disease epidemics via weighted density ensembles. PLoS Comput. Biol. 14, e1005910.
Reich, N.G., Brooks, L., Fox, S., Kandula, S., McGowan, C., Moore, E., Osthus, D., Ray, E.L., Tushar, A., Yamana, T., et al., 2018. Forecasting seasonal influenza in the US: a collaborative multi-year, multi-model assessment of forecast performance. bioRxiv 397190.
Sciensano, 2017. Scientific Institute of Public Health. (accessed 17 October 2017). https://www.sciensano.be/en.
Servićo Nacional de Saúde, Saúde 24. http://www.saude24.pt (accessed 17 October 2017).
Shaman, J., Karspeck, A., 2012. Forecasting seasonal outbreaks of influenza. Proc. Natl. Acad. Sci. 109, 20425–20430.
Soebiyanto, R.P., Adimi, F., Kiang, R.K., 2010. Modeling and predicting seasonal influenza transmission in warm regions using climatological parameters. PLoS One 5, e9450.
Vega, T., Lozano, J.E., Meerhoff, T., Snacken, R., Beauté, J., Jorgensen, P., Ortiz de Lejarazu, R., Domegan, L., Mossong, J., Nielsen, J., et al., 2015. Influenza surveillance in Europe: comparing intensity levels calculated using the moving epidemic method. Influenza Other Respir. Viruses 9, 234–246.
Vega, T., Lozano, J.E., Meerhoff, T., Snacken, R., Mott, J., Ortiz de Lejarazu, R., Nunes, B., 2013. Influenza surveillance in Europe: establishing epidemic thresholds by the moving epidemic method. Influenza Other Respir. Viruses 7, 546–558.
Viboud, C., Boëlle, P.Y., Carrat, F., Valleron, A.J., Flahault, A., 2003. Prediction of the spread of influenza epidemics by the method of analogues. Am. J. Epidemiol. 158, 996–1006.
WHO, 2016. Influenza (Seasonal). (accessed on 19 July 2017). http://www.who.int/mediacentre/factsheets/fs211/en/.
WHO Review, 2015. WHO. Review of the 2014–2015 influenza season in the northern hemisphere. (accessed 14 March 2018). http://www.who.int/wer/2015/wer9023.pdf?ua=1.
WIV-ISP, 2017. Flu surveillance system in Belgium. (Online; accessed 30 September 2017).
Won, M., Marques-Pita, M., Louro, C., Gonçalves-Sá, J., 2017. Early and real-time detection of seasonal influenza onset. PLoS Comput. Biol. 13, e1005330.
Yang, W., Karspeck, A., Shaman, J., 2014. Comparison of filtering methods for the modeling and retrospective forecasting of influenza epidemics. PLoS Comput. Biol. 10, e1003583.