



Classification of Spatiotemporal Data for Epidemic Alert Systems: Monitoring Influenza-Like Illness in France

Pavel Polyakov, Cécile Souty, Pierre-Yves Böelle, Romulus Breban

► To cite this version:

Pavel Polyakov, Cécile Souty, Pierre-Yves Böelle, Romulus Breban. Classification of Spatiotemporal Data for Epidemic Alert Systems: Monitoring Influenza-Like Illness in France. *American Journal of Epidemiology*, 2019, 188 (4), pp.724-733. 10.1093/aje/kwy254 . hal-03287491

HAL Id: hal-03287491

<https://hal.science/hal-03287491>

Submitted on 30 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Classification of spatiotemporal data for epidemic alert systems: Monitoring influenza-like illness in France

Journal:	<i>American Journal of Epidemiology</i>
Manuscript ID	AJE-01021-2017.R5
Manuscript Type:	Practice of Epidemiology
Key Words:	Influenza-like-illness, Modularity, Spatial heterogeneity, Syndromic surveillance

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Classification of spatiotemporal data for epidemic alert systems:
Monitoring influenza-like illness in France

Pavel Polyakov, Cécile Souty, Pierre-Yves Böelle, and Romulus Breban

Correspondence to Dr. Romulus Breban, Unité d'Epidémiologie des Maladies
Emergentes, Institut Pasteur, 25 rue du Dr. Roux, 75724 Paris Cedex 15, France
(e-mail: romulus.breban@pasteur.fr)

Author affiliations: Institut Pasteur, UEME, Paris, France (Pavel Polyakov and
Romulus Breban); and Sorbonne Université, INSERM, Institut Pierre Louis
d'Epidémiologie et de Santé Publique, APHP, Hôpital Saint-Antoine, Paris, France
(Cécile Souty and Pierre-Yves Böelle)

This work was funded by Agence Nationale de la Recherche (Labex Integrative
Biology of Emerging Infectious Diseases) in the form of a postdoctoral
scholarship for Pavel Polyakov.

Conflict of interest: none declared.

Running head: Spatial heterogeneity of influenza epidemics

ABSTRACT

Surveillance data serving for epidemic alert systems are typically fully aggregated in space at the national level. However, epidemics may be spatially heterogeneous, undergoing distinct dynamics in distinct regions of the surveillance area. We unveil this in retrospective analyses by classifying incidence time series. We use Pearson correlation to quantify the similarity between local time series and then classify them using modularity maximization. The surveillance area is thus divided into regions with different incidence patterns. We analyzed 31 years (1985-2016) of data on influenza-like-illness from the French system Sentinelles and found spatial heterogeneity in 19/31 influenza seasons. However, distinct epidemic regions could be identified only 4-5 weeks after the nationwide alert. The impact of spatial heterogeneity on influenza epidemiology was complex. First, when the nationwide alert was triggered, 32-41% of the administrative regions were experiencing an epidemic, while the others were not. Second, the nationwide alert was timely for the whole surveillance area, but, subsequently, regions experienced distinct epidemic dynamics. Third, the epidemic dynamics were homogeneous in space. Spatial heterogeneity analyses can provide the timing of the epidemic peak and finish, in various regions, to tailor disease monitoring and control.

Key words: Influenza-like-illness, Syndromic surveillance, Spatial heterogeneity, Modularity

Abbreviations: ILI, influenza-like illness; CI, confidence interval; NUTS, nomenclature of units for territorial statistics

Syndromic surveillance systems (1-6) are operational in many countries (7-18). Such systems routinely gather data on the number of clinical cases of infectious diseases over large surveillance areas. After curation and consolidation, data can be summarized as time series of disease incidence. Nationwide epidemic alert systems (19-23) are typically based on aggregating data over the whole surveillance area (i.e., highest level of spatial aggregation), to inform top-down strategies of public health. However, epidemics may be highly heterogeneous in space. Therefore, alerts from global systems may not concur with local epidemic dynamics; the national public health message may arrive either too late or too early for the local health authorities.

In this work, we develop new analyses of surveillance data to assess spatial heterogeneity of epidemics, using elements of metapopulation theory (24). In this context, the problem of disentangling the dynamics occurring in a large surveillance area is known as the *mega-patch problem* (25). Specifically, our problem is to classify large-scale, spatiotemporal surveillance data and figure out weakly interacting epidemics of the same infectious disease in subpopulations inhabiting different demographic areas. Hence, we aim to assess epidemic heterogeneity in the surveillance area and aggregate the data by subpopulation, to retrospectively analyze the performance of global epidemic alert systems. We also show that our methodology can be used in real-time, as an epidemic unfolds, to reveal spatial heterogeneity shortly after the nationwide alert.

We apply our methods to influenza-like illness (ILI) data collected in the metropolitan France by the Sentinelles surveillance system (26, 27), a network of voluntary, unpaid, general practitioners who report weekly numbers of ILI diagnostics together with age, sex, vaccination status and clinical characteristics of patients. Influenza-like illness was defined as a sudden onset of fever over 39 degrees Celsius with myalgia and respiratory symptoms (cough, sore throat) (28).

Time series for the number of ILI cases per 100 000 individuals per week (i.e., ILI incidence), from 1984 to present day, are available at four different levels of spatial resolution, according to the nomenclature of units for territorial statistics (NUTS) in France (29): 96 departments (NUTS3), 22 administrative regions (NUTS2), 13 administrative regions (NUTS1) and the national level. The variance of incidence is estimated assuming that the number of reported cases obeys the Poisson distribution and local incidence rates are well-described by a normal distribution. Hence, 95% confidence intervals (CI) are calculated for all weekly incidence rate estimates, using these approximations (30). Furthermore, a weekly threshold for the national alert of influenza epidemic is estimated from the ILI incidence time series (18). The nationwide alert for influenza epidemic is triggered when ILI incidence exceeds this threshold for two consecutive weeks (18). Demographic data matching the time period of the Sentinelles data are available from the Institut National de la Statistique et des Etudes Economiques (31).

METHODS

Each seasonal epidemic may have distinct dynamics and spatial patterns; we propose analyses for yearly datasets. We assume that the surveillance area is divided into N smaller units for simultaneous data collection, called *catchment areas*. In each catchment area, data are collected independently and may be further used to reconstruct a time series of incidence, to be assigned to the catchment area.

Epidemics occur in spatially heterogeneous demographic environments. Hence, the epidemic dynamics at the national scale may originate from a sum of several, say C , nearly independent epidemics, each of them established into a distinct community, included in the surveillance area. A community epidemic may be localized in a region spanning several catchment areas, called *epidemic region*. See Fig. 1, where catchment areas i and j are included in epidemic region a , while catchment area k is included in epidemic region b . Studying interdependence between incidence time series, constructed for each catchment area, may suggest how to group/classify these space units and reconstruct the epidemic regions (Fig. 1).

Over a short time interval, the total number of clinical cases reported in a catchment area, comprised into an epidemic region, has two contributions. It contains (1) a fraction of the infectious disease cases reported in the epidemic region and (2) other cases, which may not belong to the epidemic, and represent a time-independent background, specific to the catchment area. Hence, the incidence of catchment area i , I_i , may be written as a fraction p_i of the infectious

disease incidence in epidemic region a , J_a , supplemented by a local background incidence q_i ; i.e., $I_i = p_i J_a + q_i$. Thus, time series from two catchment areas included in the same epidemic region fit well one versus the other through a linear model, while time series from two catchment areas, not included in the same epidemic region, do not. The purpose of our methodology is to estimate the number of epidemic regions and identify the catchment areas included in each epidemic region (Fig. 1).

The analyses are organized according to the following steps:

(i) *Assessment of time-series interdependence.* We consider each pair of catchment areas and fit their corresponding time series, one versus the other, using a linear model. In case of missing data, we select only the data at common time points. The adjusted squared coefficient of determination, R^2 , used as goodness of fit, characterizes the correlation strength between each pair of time series. R^2 is also proportional to the squared coefficient of Pearson correlation and is thus symmetric in the two time series, defining a reflexive relationship of similarity. Hence, the R^2 values from all pairs of time series can be organized as a symmetric matrix.

(ii) *Determination of epidemic regions.* We consider the matrix of R^2 values as an adjacency matrix, defining an all-to-all undirected network with weighted edges, having all catchment areas as nodes. We classify the time series corresponding to the catchment areas by determining the community structure of this all-to-all network using modularity maximization (32-35). Thus, each community contains nodes (catchment areas) whose time series correlate well to one

another, and forms an *epidemic region*. In contrast, the correlation between time series of catchment areas included in different communities (epidemic regions) may be significantly weaker. Modularity maximization is employed here as an algorithm for unsupervised classification or hierarchical clustering (35-37), using a measure of similarity (i.e., R^2). Many other classification schemes are available. However, modularity maximization is simple to implement numerically. Its methodological shortcomings are well documented (38) and minor for the application which we discuss below.

We tested two numerical algorithms (32, 34) to maximize modularity and obtained similar results for our ILI data. The output of a single run of modularity maximization is a partition of the nodes into a number of disjoint communities, determined by the algorithm. In fact, each node is assigned a numerical label, from 1 to C , for the community where it belongs. The output is organized as a vector of community labels, with length equal to the number of nodes. This data structure, fundamental for our analyses, is called *modularity vector* (32, 34).

(iii) *Impact of data uncertainty.*

Epidemic regions result from classifying a relatively small numbers of nodes (catchment areas) whose features are defined by noisy data. Bootstrap analyses can establish the robustness of the node classification with respect to noise. Here, we organized according to the following three steps.

(iiia) *Surrogate datasets.*

We built surrogates for each data point using a normal distribution with mean at the incidence estimate and variance tuned by the corresponding CI (i.e., parametric bootstrap). We thus generated 10 000 surrogate datasets, which we analyzed according to steps (i) and (ii) described above.

(iiib) *Bootstrap analyses.*

The labeling of the communities serves no particular purpose beyond single run analysis. This leads to difficulties in the case where many modularity vectors are to be processed as an ensemble; see Web Appendix 1 for illustration. Even in the case where nearly identical datasets are processed, all yielding exactly the same partition, the community labeling may differ, leading to different modularity vectors. However, in this case, we can align each modularity vector to the first, by searching among all permutations of symbols from 1 to C for a relabeling of the communities that makes each modularity vector coincide with the first.

The same principle works for the case where datasets are not too similar (still, noise in the data should be sufficiently small), so the resulting modularity vectors correspond to different, but similar partitions. In general, analysis of bootstrap data yields an ensemble of modularity vectors where the number of communities, C , ranges from 1 to some maximum value, C_{max} . For each value of C , we chose a reference vector, which we align as follows. Given two modularity vectors with C_1 and, respectively C_2 communities ($C_1 \leq C_2$), we search among all permutations of symbols from 1 to C_1 for a relabeling that maximizes the number of identical symbols between the two vectors. The remaining (i.e., non-reference) modularity vectors may be then quickly aligned to the reference vector with the

corresponding value of C , using a similar procedure. Alignment of modularity vectors ensures the same community labeling for the analyses of all bootstrap data.

(iiic) *Summary statistics for the ensemble of aligned modularity vectors*

The ensemble of aligned modularity vectors is further processed to reveal distinctions in the node classification based on details attributable to noise in the data. In particular, we compute f_a^i , the fraction of times catchment area i appeared in community a (Web Figure 1 in Web Appendix 1). To evaluate the global impact that uncertainty has on community structure, we calculate the following bootstrap score inspired by the Shannon entropy

$$B = - \sum_{a=1}^{C_{\max}} \sum_{i=1}^N \frac{f_a^i \log(f_a^i)}{N \log(N)}.$$

If all values of f_a^i are either 0 or 1, then data uncertainty has no impact on community decomposition and $B=0$; the higher the value of B , the deeper the impact of data uncertainty on community decomposition; see Web Appendix 2 for further discussion.

(iv) *Data aggregation for each epidemic region.* Incidence time series belonging to community a may be aggregated using the catchment area population multiplied by f_a^i , as weight. This yields a time series of expected incidence and corresponding CI for each community. We do not aggregate a time series for community a if $\sum_{i=1}^N f_a^i < 2$; that is, the weights for region a do not amount to represent at least two catchment areas.

(v) *Validation*. We reject the output of the above analyses if the resulting bootstrap score B is above a pre-established threshold. Furthermore, we use Hellinger distance (39) to validate statistically the distinction between two aggregated time series. Namely, for each moment of time, we calculate the Hellinger distance between the corresponding data, assuming that incidence is normally distributed. Then, summing the pairwise Hellinger distances over the duration of the time series, we obtain a Hellinger score, denoted by H . The distinction between two epidemic regions is rejected if the Hellinger score between the corresponding aggregated time series is less than a certain threshold. In this case, one may choose to merge the two epidemic regions, as the distinction revealed by the algorithm may be considered too small to have operational value.

The interdependence between aggregated, community time series may be further assessed using Pearson correlation; see step (i). Community decompositions with high bootstrap score B , where many values of f_a^i differ significantly from 0 or 1, yield high R^2 values between aggregated time series, because the time series for catchment area i may contribute significantly to several aggregated community time series. This is where our analyses do poorly. In contrast, for community decompositions with low bootstrap score B , we may expect that R^2 between aggregated time series is small, indicating weak interdependence and a meaningful disaggregation of the national epidemic curve. We also validate our analyses if the bootstrap score B is low and only one community is most likely to be found. In this case, we say that the entire surveillance area was subject to an epidemic homogeneous in space.

In the case of automated surveillance, where the lengths of the time series increase steadily with time, the above analyses may be repeated for each additional time point, to constantly monitor B , H and R^2 .

RESULTS

We split yearly data into two: a 26-week period (the *influenza season*, mid-October to mid-April) when influenza epidemics can occur in France, and its yearly complement. Analyses were carried out independently. During an influenza season, the weekly number of ILI cases per 100 000 individuals typically peaks at high values (i.e., 350-2000), while off-season, it passes through much lower values (i.e., <100). Our analyses showed that findings are robust regarding the definition of the influenza season, provided that the time interval defined by the start week (Table 1) and the 26-week duration includes the high incidence values.

We contrasted analyses of data collected during the influenza season with analyses of data collected off-season (Web Appendix 3, Web Figure 2). The adjusted R^2 values were relatively high for the data collected during the influenza season, motivating the search for epidemic regions (Web Appendix 4, Web Figure 3). In contrast, R^2 values were consistently low for the data collected off-season, indicating that ILI incidence time series are poorly correlated. Thus, the search for epidemic regions was not motivated, in this case.

We performed analyses at 3 levels of spatial resolution of the French administrative structure (Web Appendix 5, Web Figure 4). First, we considered the department level (NUTS3; 96 departments). The typical French department had ~500 000 population and 3-5 practitioners participating to the Sentinelles surveillance system. Under the circumstances, fluctuations in reported ILI cases proved important for time series dynamics and trends were not readily apparent. The resulting R^2 values were relatively low and, searching for community structure at this level, unjustified. Fluctuations in reported ILI cases were much reduced when data were aggregated at the level of the 13 NUTS1 regions; a NUTS1 region contains in median 6 departments. Spatial heterogeneity analyses were sound; however, this spatial resolution was found to be coarse and suboptimal for defining epidemic regions. A better approach was the middle ground of spatial resolution, provided by the 22 NUTS2 administrative regions; a NUTS2 region contains in median 4 departments.

Bootstrap analyses unveiled the role of data uncertainty for our results; see Table 1. Based on retrospective analyses, we divided the 31 influenza seasons into four disjoint groups: seasons where (1) $B > 0.1$ and data uncertainty had a substantial impact on ascertaining spatial homogeneity/heterogeneity of epidemics (5/31), (2) $B < 0.1$, $H < 10$ and thus, aggregated time series were very similar, considering data uncertainty (5/31), (3) $B < 0.1$, $H < 10$, $R^2 > 0.6$, and the level of spatial heterogeneity may be subject to further discussion (11/31), and (4) $B < 0.1$, $H < 10$, $R^2 < 0.6$, and spatial heterogeneity was particularly strong, playing a clear role for local surveillance (10/31).

Retrospective analyses revealed three scenarios where epidemic heterogeneity played a role relative to the current protocols for epidemic alert and epidemic threshold in France. Accordingly, we present detailed results about six illustrative influenza seasons. Figure 2 shows results for the 2015-2016 (panels A) and C)) and 1985-1986 (panels B) and D)) influenza seasons, where the surveillance area of Sentinelles was divided into two epidemic regions of comparable size. We colored a catchment area depending on the fraction of times it occurred in the first epidemic region and the fraction of times it occurred in the second epidemic region, in bootstrap analyses of 10 000 datasets. A catchment area was colored blue or red if it occurred always in the first or second epidemic region; i.e., the fractions were, respectively, either (1.0, 0.0) or (0.0, 1.0). Otherwise, we used intermediate colors, interpolated between the extremes of red and blue. However, in the case of low B , all map colors are nearly blue or nearly red.

The two epidemic regions that we found had distinct influenza dynamics; see the panels C) and D) in Fig. 2. The epidemic in one region (blue) started early, while the epidemic in the other region (red) started 3-4 weeks later. The national epidemic alert (vertical line) was triggered by the early epidemic. Hence, for 3-4 weeks, one region was under alert, yet not experiencing an epidemic, because of the epidemic in the other region. In these cases, simply mapping the difference between the local and national-level ILI incidence at the time of the national epidemic alert, suggested clear patterns of spatial heterogeneity (Web Appendix 6, Web Figure 5).

Figure 3 shows results for the 1997-1998 (panels A) and C)) and 1991-1992 (panels B) and D)) influenza seasons. Again, in each case, the surveillance area was divided into two epidemic regions of comparable size. The epidemics arrived nearly at the same time; the delay was only 1 week, which is the time resolution of Sentinelles. However, the subsequent epidemic dynamics were quite different in the two regions that might have benefited from tailored surveillance. For example, in the 1997-1998 influenza season, ILI incidence peaked 5 weeks apart in the two epidemic regions. This was not the case for the 1991-1992 influenza season. However, during the 1991-1992 influenza season, ILI incidence in one epidemic region (blue) reached below the epidemic threshold 5 weeks before the ILI incidence in the other epidemic region (red).

Finally, we report on the 2014-2015 and 2012-2013 influenza seasons (Table 1) where the whole surveillance area appeared as a single epidemic region in 97% and 85% of the bootstrap analyses, respectively. These are remarkable cases of epidemic homogeneity in space. Consequently, the national alerts were timely for the entire surveillance area and the current public health protocols worked as expected.

To assess their performance for real-time applications, we repeated our analyses with varying amount of data. Namely, we considered datasets from the beginning of the influenza season up to a certain week after the epidemic alert and investigated how B and R^2 change with the amount of data. Figure 4 shows results for the 2015-2016 (panel A)) and 1997-1998 (panel B)) influenza seasons, where the most likely divide was into two epidemic regions. The

amount of data gathered from the beginning of the influenza season up to the time of the national alert, and immediately after, was insufficient for clear results. B values were high, indicating that data uncertainty had a strong impact on our results. Furthermore, aggregated time series were highly correlated, resembling the national epidemic curve. However, data gathered 4-5 weeks after the nationwide alert was already sufficient to ascertain spatial heterogeneity of the epidemic with $B < 0.1$ and $R^2 < 0.6$. Analyses of datasets for the following weeks provided confirmation, with a slightly increasing trend for R^2 as the epidemic went extinct.

It is important to note that, even if detection of epidemic regions may occur later than the nationwide alert, it can still offer critical information on the qualitative and quantitative dynamics of regional epidemics. For example, in the case of the 2015-2016 influenza season (Fig. 4, panel A)), robust detection of epidemic regions is obtained in week 5 after the global epidemic alert. Inspection of Fig. 2 reveals that this corresponds to week 10 of the influenza season. At that time, the epidemic was past its peak in the blue epidemic region, but not in the red region. Incidentally, the epidemic was also not past its peak at the national scale. Similar results hold for the 1997-1998 influenza season (Fig. 4, panel B)). Hence, our methods can successfully deliver the spatial structure of influenza epidemics, in time for improving monitoring and control during the influenza season.

DISCUSSION

1
2
3 Timeliness is a key performance measure of public health surveillance systems
4
5 (40). However, timeliness can depend on the scale at which information is
6
7 aggregated to inform public health practitioners. For example, in France,
8
9 epidemic alerts for influenza are given once for the whole surveillance area,
10
11 based on data aggregated at the national level. There are obvious statistical
12
13 advantages in aggregating time series at the national level: noise is much
14
15 reduced and incidence trends are more prominent as they cross epidemic
16
17 thresholds. However, in case of spatial heterogeneity, this approach masks
18
19 distinct epidemic dynamics taking place in the surveillance area, and undermines
20
21 the usefulness of the surveillance system.
22
23
24
25
26
27
28

29 The case of influenza illustrates this problem perfectly. The burden of seasonal
30
31 influenza epidemics is large in western countries (41). Once the epidemic alert is
32
33 triggered, national media campaigns on prevention are launched in the press, TV
34
35 and radio, with messages on hygiene and vaccination. Timing can be critical,
36
37 since vaccine-induced protection becomes effective about 2 weeks following
38
39 vaccination (42). However, more concerning is the impact of influenza epidemics
40
41 on the routine functioning of hospital wards (43). In France, influenza epidemic
42
43 alerts put hospitals under stress, as emergency protocols may be activated (e.g.,
44
45 postponing non-urgent interventions, conscripting staff, freeing hospital beds).
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
Untimely decisions in these situations may lead to inefficiency and incumbent costs to society.

It is therefore important to identify situations where a public health message, designed for national broadcast, can be later supplemented with customized

information for the local practitioners. Mapping incidence data already gives important clues about epidemic heterogeneity (44). Furthermore, gravity/radiation models (45-49) could be employed for extensive spatiotemporal analyses, including data fitting. However, these models make strong assumptions on how two catchment areas must interact; i.e., gravity models assume that interaction declines with (squared) distance, while radiation models assume that individuals spread like radiation fluxes. With the goal of only classifying spatiotemporal data, we made minimal assumptions about the geographical structure, using concepts of metapopulation theory to discover the spatial structure of epidemics. Still, the epidemic regions that we found cluster, to a large extent, neighboring catchment areas.

Regions of distinct epidemic dynamics may occur for a variety of reasons. Factors known to influence the transmission of influenza are: susceptibility profile of the population, circulating strains, vaccine parameters and vaccination patterns, travel and daily commuting, school holidays, and the weather. In addition, the spatiotemporal coordinates of the individuals who happen to initiate the influenza epidemics (so-called *patients zero*) may also be very important. Therefore, an account for the spatial heterogeneity of influenza, starting from first principles, may be particularly challenging. Our strategy was to determine influenza epidemic regions directly, using surveillance data, by analyzing local epidemic dynamics provided by incidence time series.

Acknowledging spatial heterogeneity in a surveillance area and identifying epidemic regions may have important consequences for improving influenza

1
2
3 monitoring and control. The operational value of epidemic regions depends on
4
5 several key items. First, the spatial resolution must be appropriate (50). Here,
6
7 the signal to noise ratio was too small for a meaningful analysis at the NUTS3
8
9 level, but sufficient for defining epidemic regions at NUTS2 and NUTS1 level.
10
11

12 This is relevant for the structure of public health administration in France.

13
14 Second, the time to identification of distinct dynamics must be compatible with
15
16 staggered delivery of public health messages. Here, 4-5 weeks of data past the
17
18 time of the epidemic alert were necessary to identify epidemic regions. Even
19
20 with this delay, valuable regional updates can be passed to the local health
21
22 authorities during the epidemic, to maintain vigilance in the affected territories.
23
24
25

26
27
28 The distinction of epidemic regions may be also relevant for collecting pathogen
29
30 samples. Every influenza season, clinical samples from ILI cases are collected and
31
32 analyzed, to establish the spectrum of circulating virus strains. In turn, this
33
34 determines the composition of future influenza vaccines. Clinical samples should
35
36 be collected from each epidemic region for a better characterization of the
37
38 circulating influenza strains.
39
40
41
42
43
44

45 In conclusion, we proposed new methodology to detect spatial heterogeneity in
46
47 disease surveillance data and discussed monitoring ILI in France. However, our
48
49 methods are not tailored to influenza epidemiology and may be used for other
50
51 case diseases.
52
53
54

55 56 57 **ACKNOWLEDGEMENTS** 58 59 60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Author affiliations: Institut Pasteur, UEME, Paris, France (Pavel Polyakov and Romulus Breban); and Sorbonne Université, INSERM, Institut Pierre Louis d'Epidémiologie et de Santé Publique, APHP, Hôpital Saint-Antoine, Paris, France (Cécile Souty and Pierre-Yves Böelle).

This work was funded by Agence Nationale de la Recherche (Labex Integrative Biology of Emerging Infectious Diseases) in the form of a postdoctoral scholarship for Pavel Polyakov.

Conflict of interest: none declared.

REFERENCES

1. Henning KJ. What is Syndromic Surveillance? *Morbidity and Mortality Weekly Report* 2004;53:7-11.
2. Lemay R, Mawudeku A, Shi Y. Syndromic surveillance for influenzalike illness. *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science* 2008;6(2):161-70.
3. Lombardo JS, Burkom H, Pavlin J. ESSENCE II and the Framework for Evaluating Syndromic Surveillance Systems. *Morbidity and Mortality Weekly Report* 2004;53:159-65.
4. Mandl KD. Implementing Syndromic Surveillance: A Practical Guide Informed by the Early Experience. *Journal of the American Medical Informatics Association* 2003;11(2):141-50.
5. Sosin DM. Draft framework for evaluating syndromic surveillance systems. *Journal of Urban Health* 2003;80(2):I8-I13.
6. World_Health_Organization. Communicable disease surveillance and response systems: guide to monitoring and evaluating. Lyon [France]: World Health Organization, 2006.
7. Del Mar C, Pincus D. Incidence patterns of respiratory illness in Queensland estimated from sentinel general practice. *Australian Family Physician* 1995;24(4):625-9- 32.
8. Canada_Ministry_of_Health. Influenza in Canada--1996-1997 season. *Canada communicable disease report* 1997;23(24):185-92.
9. Matter HC, Cloetta J, Zimmermann H. Measles, mumps, and rubella: monitoring in Switzerland through a sentinel network, 1986-94. Sentinella Arbeitsgemeinschaft. *Journal of Epidemiology and Community Health* 1995;49(Suppl 1):4-8.
10. Ohshiro H, Kawamoto K, Nose T. Surveillance system of infectious diseases in Japan. *Journal of Epidemiology* 1996;6(3 Suppl):S81-5.
11. Park S, Cho E. National Infectious Diseases Surveillance data of South Korea. *Epidemiology and Health* 2014;36:e2014030.
12. Snacken R, Lion J, Van Casteren V, et al. Five years of sentinel surveillance of acute respiratory infections (1985-1990): The benefits of an influenza early warning system. *European journal of Epidemiology* 1992;8(4):485-90.
13. Sprenger MJW, Mulder PGH, Beyer WEP, et al. Influenza: Relation of Mortality to Morbidity Parameters-Netherlands, 1970-1989. *International journal of epidemiology* 1991;20(4):1118-24.
14. Thacker SB, Stroup DF. Future Directions for Comprehensive Public Health Surveillance and Health Information Systems in the United States. *American Journal of Epidemiology* 1994;140(5):383-97.
15. Tillett HE, Spencer IL. Influenza surveillance in England and Wales using routine statistics. *Journal of Hygiene* 1982;88:83-94.
16. Wu T-SJ, Shih F-YF, Yen M-Y, et al. Establishing a nationwide emergency department-based syndromic surveillance system for better public health responses in Taiwan. *BMC Public Health* 2008;8(1):18.

17. Yan W, Palm L, Lu X, et al. ISS-an electronic syndromic surveillance system for infectious disease in rural China. *PLoS One* 2013;8(4):e62749.
18. Costagliola D, Flahault A, Galinec D. A routine tool for detection and assessment of epidemics of influenza-like syndromes in France. *American journal of public health* 1991;81(1):97-9.
19. Amorós R, Conesa D, Martinez-Beneito MA. Statistical methods for detecting the onset of influenza outbreaks: a review. *REVSTAT-Statistical Journal* 2015;13(1):41-62.
20. Cowling BJ, Wong IOL, Ho LM, et al. Methods for monitoring influenza surveillance data. *International journal of epidemiology* 2006;35(5):1314-21.
21. Unkel S, Farrington CP, Garthwaite PH, et al. Statistical methods for the prospective detection of infectious disease outbreaks: a review. *Journal of the Royal Statistical Society Series A-Statistics in Society* 2012;175(1):49-82.
22. OBrien SJ, Christie P. Do CuSums have a role in routine communicable disease surveillance? *Public Health* 1997;111(4):255-8.
23. Hashimoto S, Murakami Y, Taniguchi K, et al. Detection of epidemics in their early stage through infectious disease surveillance. *International journal of epidemiology* 2000;29(5):905-10.
24. Hanski I. *Metapopulation Ecology*. Oxford: Oxford University Press; 1999.
25. Cavanaugh KC, Siegel DA, Raimondi PT, et al. Patch definition in metapopulation analysis: a graph theory approach to solve the mega-patch problem. *Ecology* 2014;95(2):316-28.
26. INSERM, UPMC, Santé_publique_France. Réseau Sentinelles. (<https://websenti.u707.jussieu.fr/sentiweb/>). (Accessed October 25, 2018).
27. Flahault A, Blanchon T, Dorléans Y, et al. Virtual surveillance of communicable diseases: a 20-year experience in France. *Statistical methods in medical research* 2006;15(5):413-21.
28. Carrat F, Tachet A, Rouzioux C, et al. Evaluation of clinical case definitions of influenza: detailed investigation of patients during the 1995-1996 epidemic in France. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 1999;28(2):283-90.
29. European_Commission. 2016. (<https://ec.europa.eu/eurostat/web/nuts/background>). (Accessed October 24, 2018).
30. Turbelin C, Souty C, Pelat C, et al. Age Distribution of Influenza Like Illness Cases during Post-Pandemic A(H3N2): Comparison with the Twelve Previous Seasons, in France. *PLoS One* 2013;8(6):e65919-9.
31. INSEE. L'Institut national de la statistique et des études économiques. (<https://http://www.insee.fr/fr/accueil>). (Accessed October 25, 2018).

32. Blondel VD, Guillaume J-L, Lambiotte R, et al. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008;2008(10):P10008.
33. Clauset A, Newman MEJ, Moore C. Finding community structure in very large networks. *Physical Review E* 2004;70(6):066111.
34. Newman MEJ. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America* 2006;103(23):8577-82.
35. Newman MEJ. Communities, modules and large-scale structure in networks. *Nature Physics* 2011;8(1):25-31.
36. Jain AK. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters* 2010;31(8):651-66.
37. Xu R, Wunsch D. Survey of clustering algorithms. *IEEE Trans Neural Netw* 2005;16(3):645-78.
38. Fortunato S, Barthelemy M. Resolution limit in community detection. *Proceedings of the National Academy of Sciences of the United States of America* 2006;104(1):36-41.
39. Hellinger distance.
(http://www.encyclopediaofmath.org/index.php?title=Hellinger_distance&oldid=16453). (Accessed October 24, 2018).
40. Jajosky RA, Groseclose SL. Evaluation of reporting timeliness of public health surveillance systems for infectious diseases. *BMC Public Health* 2004;4:29.
41. Molinari NA, Ortega-Sanchez IR, Messonnier ML, et al. The annual impact of seasonal influenza in the US: measuring disease burden and costs. *Vaccine* 2007;25(27):5086-96.
42. CDC. Key facts about Flu vaccine.
(<https://http://www.cdc.gov/flu/protect/keyfacts.htm>). (Accessed May 19, 2017).
43. Hoot NR, Aronsky D. Systematic review of emergency department crowding: causes, effects, and solutions. *Annals of emergency medicine* 2008;52(2):126-36.
44. Carrat F, Valleron AJ. Epidemiologic mapping using the "kriging" method: application to an influenza-like illness epidemic in France. *Am J Epidemiol* 1992;135(11):1293-300.
45. Gog JR, Ballesteros S, Viboud C, et al. Spatial Transmission of 2009 Pandemic Influenza in the US. *PLoS computational biology* 2014;10(6):e1003635.
46. Viboud C, Bjornstad ON, Smith DL, et al. Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science* 2006;312(5772):447-51.
47. Simini F, González MC, Maritan A, et al. A universal model for mobility and migration patterns. *Nature* 2012;484(7392):96-100.
48. Meyer S, Held L. Incorporating social contact data in spatio-temporal models for infectious disease spread. *Biostatistics* 2017;18(2):338-51.
49. Meyer S, Held L, Hohle M. Spatio-Temporal Analysis of Epidemic Phenomena Using the R Package surveillance. *J Stat Softw* 2017;77(11).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

50. Gregorio DI, Samociuk H, DeChello L, et al. Effects of study area size on geographic characterizations of health events: prostate cancer incidence in Southern New England, USA, 1994-1998. *International journal of health geographics* 2006;5:8.

For Peer Review

TABLE

Table 1. Summary of results for 31 influenza seasons (1985-2016) in France, divided into 4 disjoint categories.

Season	Start week	Bootstrap score, B	Hellinger score H between aggregated time series	Adjusted R^2 between aggregated time series	Most likely number of epidemic regions	Probability of most likely number of epidemic regions
seasons for which $B > 0.1$						
2007-2008	46	0.148	6.6	0.91	2	0.97
2005-2006	48	0.208			3	0.87
1			8.2	0.75		
2			10.0	0.65		
3			7.6	0.75		
2002-2003 ^a	47	0.253			3	0.60
1			13.3	0.39		
2			5.7	0.92		
3			10.5	0.60		
2000-2001 ^a	44	0.290			3	0.66
1			3.7	0.92		
2			9.4	0.73		
3			11.9	0.45		
1990-1991 ^a	46	0.304			2	0.65

1	1			3.1	0.97		
2	2			5.8	0.81		
3	3			3.0	0.92		
4	seasons						
5	for which						
6	$B \leq 0.1$,						
7	but $H < 10$						
8	2013-2014	47	0.096	8.1	0.83	2	0.99
9	2009-2010	32	0.082	9.9	0.76	2	1.00
10	2006-2007	46	0.042	8.7	0.82	2	1.00
11	2001-2002 ^a	45	0.100	6.6	0.89	2	0.98
12	1994-1995	52	0.070	8.8	0.80	2	0.99
13	seasons						
14	for which						
15	$B \leq 0.1$, H						
16	> 10 , but						
17	$R^2 > 0.6$						
18	2008-2009	45	0.054	10.4	0.80	2	1.00
19	2004-2005 ^a	49	0.032	12.5	0.67	2	1.00
20	2003-2004 ^a	36	0.047	12.5	0.70	2	0.98
21	1999-2000	42	0.086	11.1	0.79	2	0.99
22	1998-1999 ^a	46	0.069	10.6	0.68	2	1.00
23	1995-1996 ^a	38	0.012	13.8	0.73	2	1.00
24	1993-1994	38	0.038	12.8	0.86	2	1.00
25	1992-1993	47	0.050	11.7	0.65	2	0.99
26	1988-1989	38	0.007	14.2	0.62	2	1.00
27	1987-1988 ^a	53	0.028	12.5	0.81	2	1.00
28	1986-1987 ^a	46	0.077	12.2	0.67	2	0.96
29	seasons						

for which for which $B \leq 0.1$, H > 10 and $R^2 \leq 0.6$						
2015-2016	48	0.006	13.3	0.49	2	1.00
2014-2015	46	0.035			1	0.97
2012-2013	46	0.074			1	0.85
2011-2012	48	0.100	10.2	0.60	2	0.87
2010-2011	44	0.028	13.4	0.46	2	1.00
1997-1998 ^a	50	0.014	12.9	0.34	2	1.00
1996-1997 ^a	42	0.014	14.4	0.56	2	1.00
1991-1992	42	0.062	12.3	0.53	2	0.89
1989-1990	41	0.000	14.8	0.43	2	1.00
1985-1986 ^a	44	0.026	13.4	0.41	2	0.94

^a Corsica was excluded from the analyses due to missing data.

FIGURE LEGENDS

Figure 1. Conceptual representation of a surveillance system observing a spatially heterogeneous epidemic. The surveillance area (thick contour) is divided into N catchment areas represented as squares (e.g., i, j, k). Several (say C) nearly independent community epidemics may develop simultaneously in various regions of the surveillance area; hence we may distinguish several epidemic regions (e.g., a, b). An epidemic region may include several catchment areas; e.g., catchment areas i and j are included in epidemic region a , while catchment area k is included in epidemic region b . The incidence time series of area i fits well the incidence time series of area j through a linear model, as catchment areas i and j belong to the same epidemic region and sample the same epidemic dynamics. In contrast, the incidence time series of area i fits badly the incidence time series of area k , as catchment areas i and k belong to different epidemic regions. The purpose of our methodology is to sort out the incidence time series in all catchment areas and attribute each area to an epidemic region.

Figure 2. Analyses for the 2015-2016 and 1985-1986 influenza seasons. The panels A) and C) refer to 2015-2016, while the panels B) and D) refer to 1985-1986 (Corsica was excluded due to missing data). The panels A) and B) represent the epidemic regions that we found for France. A NUTS2 region (i.e., catchment area) was colored depending on the fraction of times it occurred in the first epidemic region and the fraction of times it occurred in the second epidemic region, in the analyses of 10 000 bootstrap datasets. Blue corresponds to fractions (1.0, 0,0) to occur in the first and second epidemic regions, respectively,

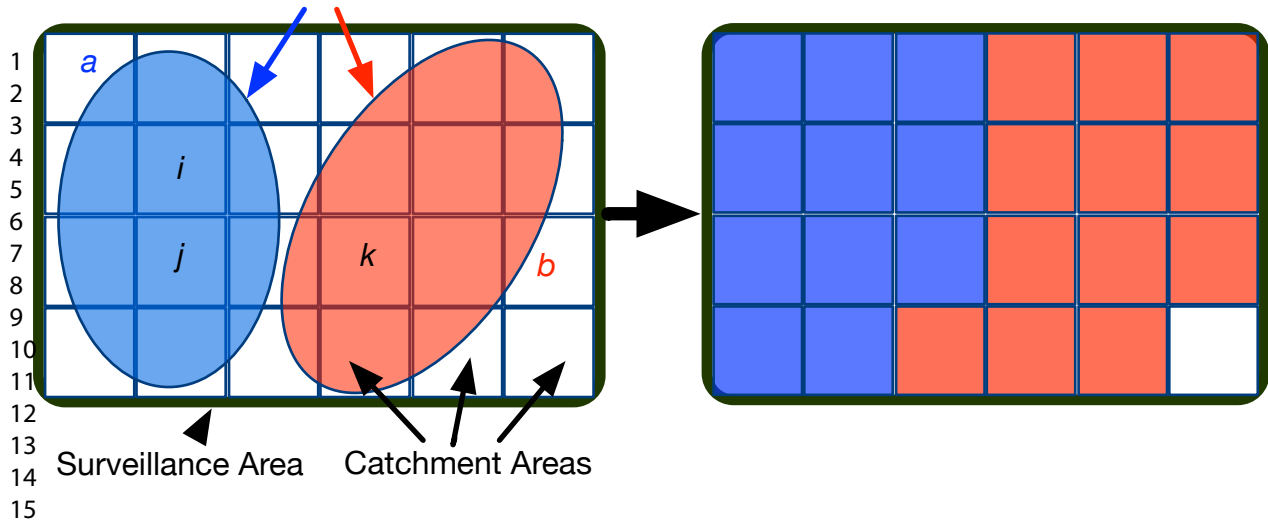
1
2
3 while red corresponds to the fractions (0.0, 1.0). For the other cases, we used
4
5 intermediate colors. The panels C) and D) represent the ILI incidence time series
6
7 (blue line with solid circle marker and red line with X marker) aggregated over
8
9 the epidemic regions represented in panels A) and B) in respective colors. We
10
11 also illustrate the national ILI incidence time series (black line) and the epidemic
12
13 threshold (dotted line). Note that the nationwide alert (vertical line) was not
14
15 timely for both epidemic regions.
16
17
18
19

20
21
22 **Figure 3.** Analyses for the 1997-1998 and 1991-1992 influenza seasons. The
23
24 panel arrangement and color code are similar to those of Fig. 2. The panels A)
25
26 and C) refer to 1997-1998 (Corsica was excluded due to missing data), while the
27
28 panels B) and D) refer to 1991-1992. The panels A) and B) represent the
29
30 epidemic regions that we found for France. A NUTS2 region (i.e., catchment area)
31
32 was colored blue (red) if it occurred all the time in the first (second) epidemic
33
34 region. Intermediate colors represent NUTS2 regions that occurred in both
35
36 epidemic regions. The panels C) and D) represent, in blue and red, the ILI
37
38 incidence time series aggregated over the epidemic regions shown in respective
39
40 colors in the panels A) and B). We also illustrate the national ILI incidence time
41
42 series (black line) and the epidemic threshold (dotted line). While the
43
44 nationwide alert (vertical line) was appropriate, the epidemic regions
45
46 experienced distinct dynamics and could have benefited from customized
47
48 monitoring.
49
50
51
52
53
54
55
56

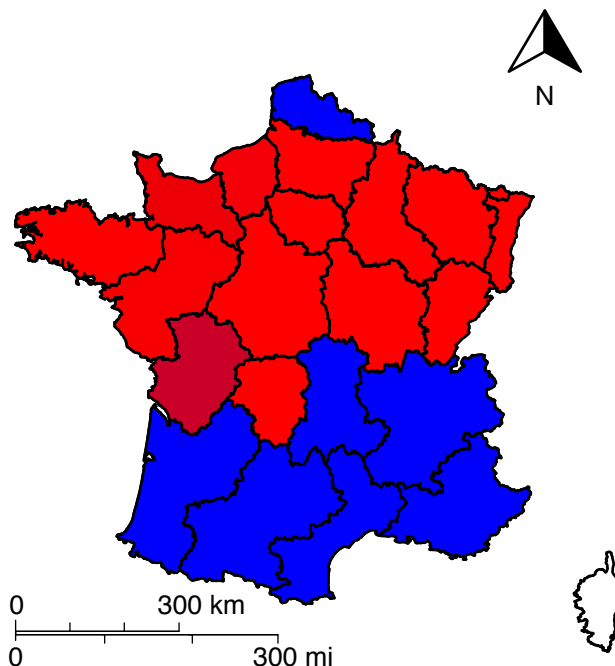
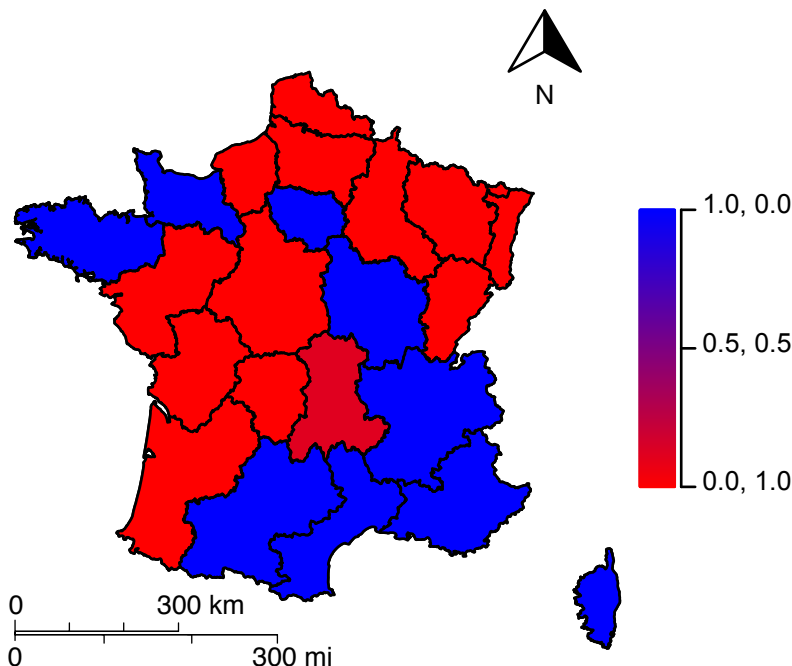
57 **Figure 4.** The bootstrap parameter B and the adjusted R^2 between the
58
59 aggregated time series of the first and second epidemic regions as a function of
60

the amount of data: results for the 2015-2016 and 1997-1998 influenza seasons. Data to estimate B and R^2 was considered from the beginning of the influenza season up to a certain week after the epidemic alert, indicated on the horizontal axis. The values for B and R^2 are plotted, respectively, in green with filled square markers and purple with + markers. Threshold values for B and R^2 at 0.1 and 0.6, respectively, are shown as horizontal dashed lines in corresponding colors. We note a marked drop in B and R^2 4-5 weeks after the epidemic alert.

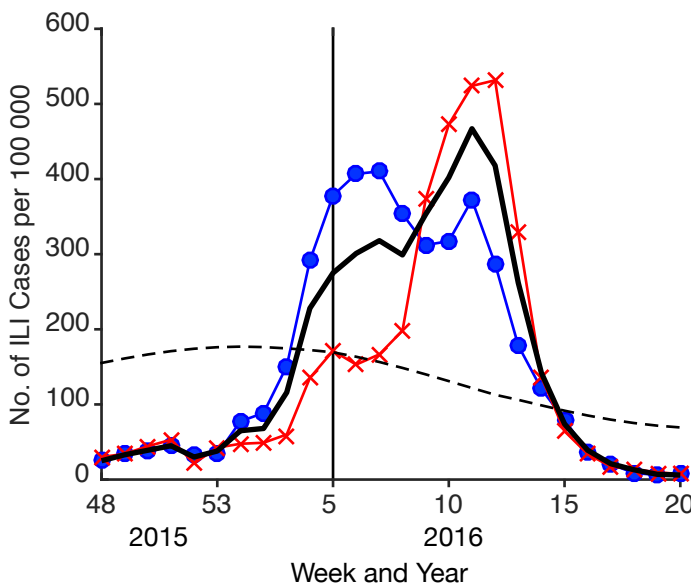
For Peer Review



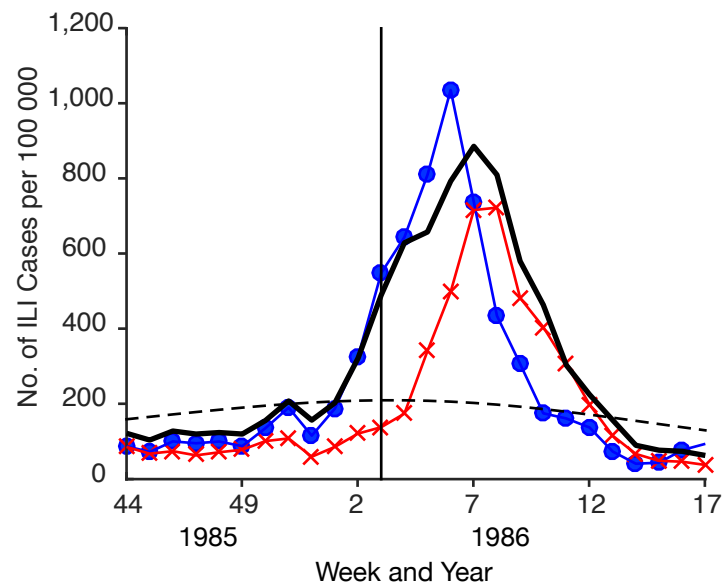
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45



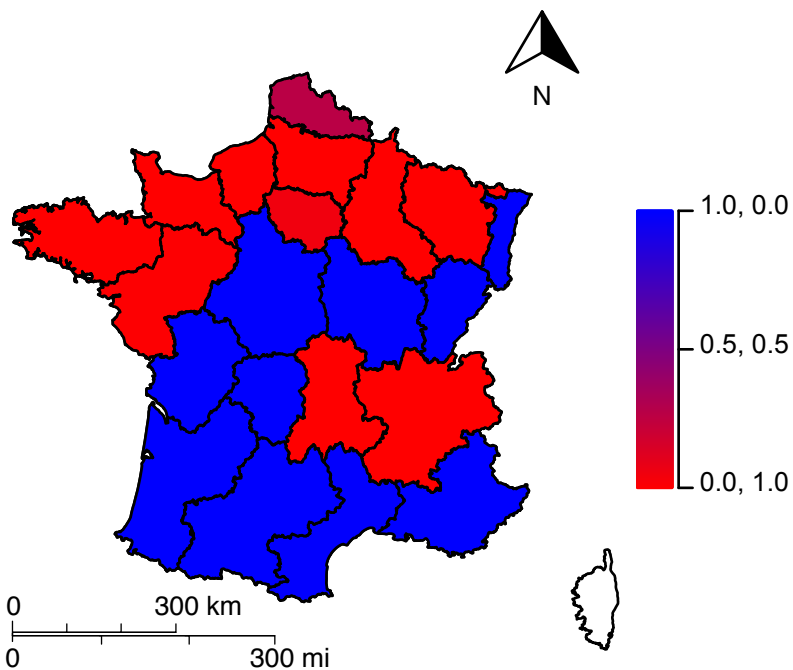
C)



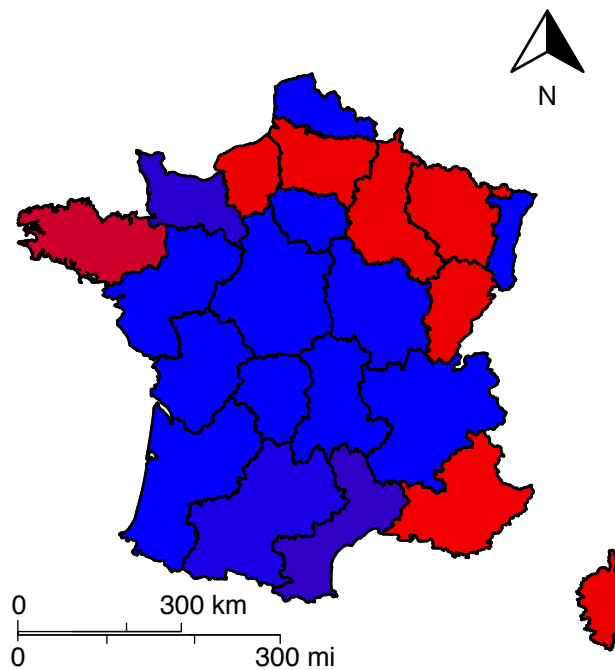
D)



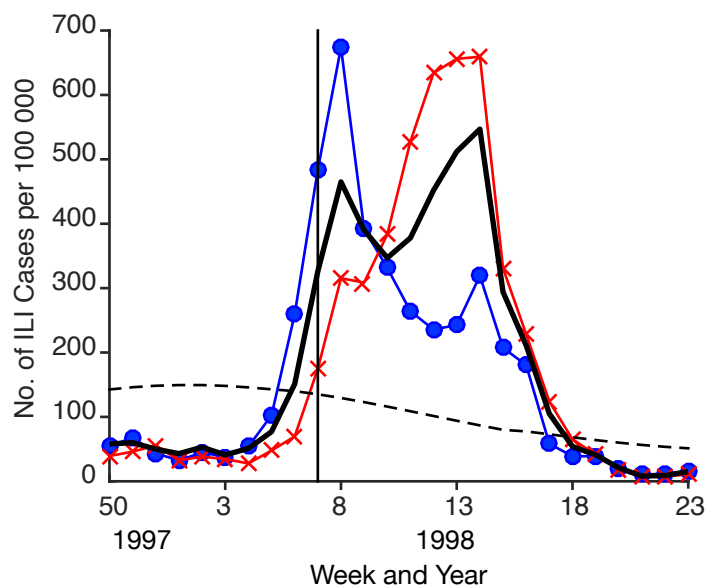
A)



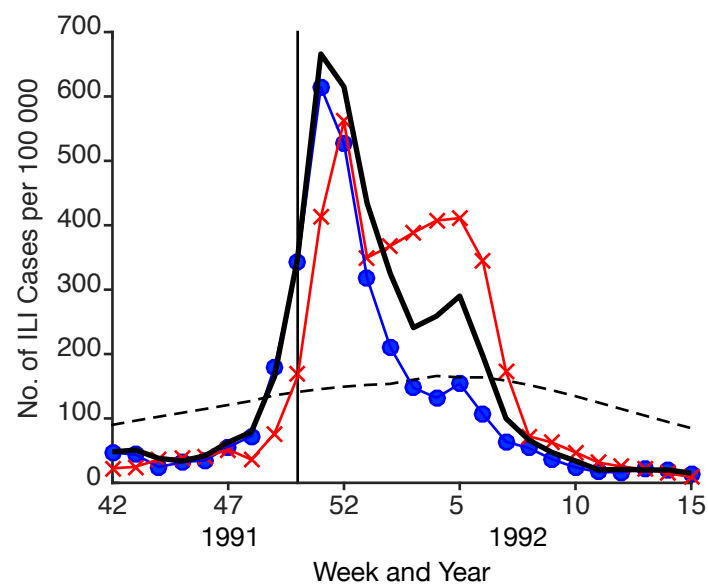
B)

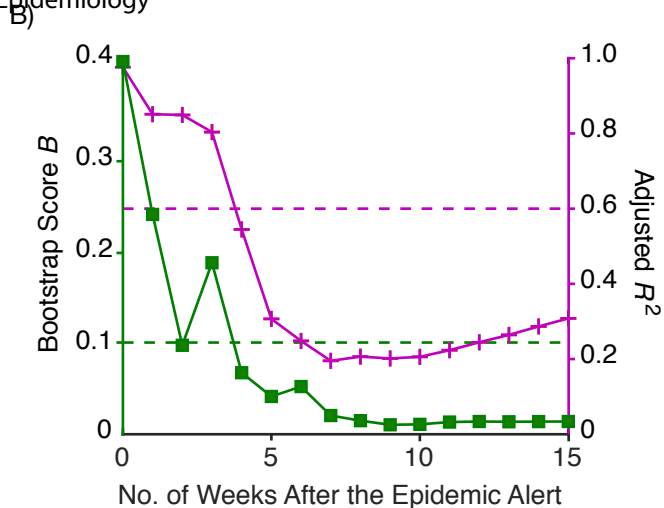
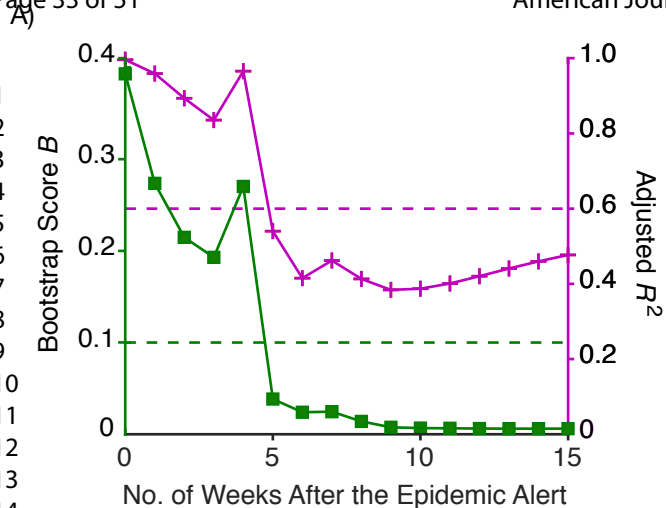


C)



D)





Classification of spatiotemporal data for epidemic alert systems: Monitoring influenza-like illness in France—WEB MATERIAL

Pavel Polyakov, Cécile Souty, Pierre-Yves Böelle, and Romulus Breban*

*Correspondence to Dr. Romulus Breban, Unit d'Epidémiologie des Maladies Emergentes, Institut Pasteur, 25 rue du Dr. Roux, 75724 Paris Cedex 15, France (e-mail: romulus.breban@pasteur.fr)

November 1, 2018

Author affiliations: Institut Pasteur, UEME, Paris, France (Pavel Polyakov and Romulus Breban); and Sorbonne Université, INSERM, Institut Pierre Louis d'Epidémiologie et de Santé Publique, APHP, Hôpital Saint-Antoine, Paris, France (Cécile Souty and Pierre-Yves Böelle)

Contents

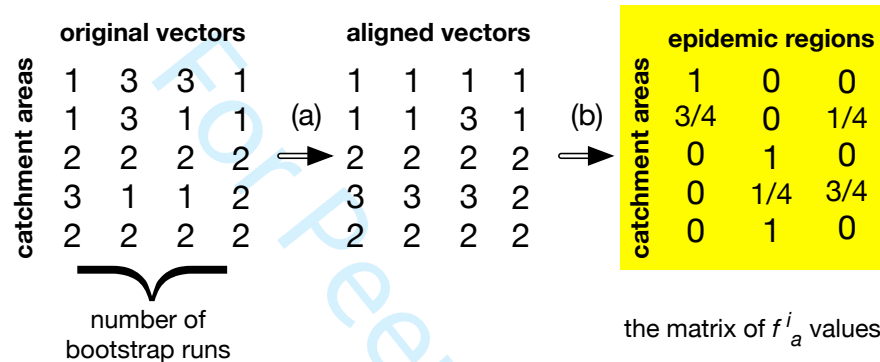
1 Web Appendix 1	3
2 Web Appendix 2	4
3 Web Appendix 3	5
4 Web Appendix 4	6
5 Web Appendix 5	7
6 Web Appendix 6	8

List of Web Figures

- 1 Alignment of modularity vectors. Two steps of analysis are illustrated: (a) relabeling of the communities for each vector such that the number of symbols in common with the reference (i.e., first) vector is maximized; (b) computation of f_a^i , the fraction of times catchment area i occurred in community (epidemic region) a 3
- 2 Comparison of results. Panels **A₁** and **B₁** show histograms of the adjusted R^2 values calculated for each pair of time series, for the 2015-2016 winter season (weeks 48-20) and the 2015 summer season (weeks 22-47) respectively. We note a substantial difference in the median R^2 values (0.587 and 0.042, respectively). For the influenza season, many time series of catchment areas correlate well, motivating a search for community structure. In contrast, off-season, the correlation is much poorer and the search for community structure not motivated. The catchment areas are represented by the 22 administrative regions, colored according to their inclusion into different epidemic regions (panels **A₂** and **B₂**, respectively). The community structure in panel **A₂** (2 epidemic regions) is meaningful, the community structure in panel **B₂** (4 epidemic regions) is spurious. No bootstrap was performed for this sample analysis. 5
- 3 Sample analyses for the 2015-2016 influenza season. Panel **A** shows the histogram of all adjusted R^2 values. Panel **B** shows the map of epidemic regions. The histograms of the adjusted R^2 values (corresponding medians: 0.723 and 0.750) for all pairs of time series in each epidemic region, blue and red, are shown in the panels **C** and **D**, respectively. The panels **E** and **F** represent the corresponding time series. Panel **G** shows the time series aggregated over the blue and red regions, respectively. The adjusted R^2 between the aggregated time series is 0.492, indicating poor correlation. The national average is plotted in black and the epidemic threshold is represented as the dashed line. Panel **H** shows the distribution of the adjusted R^2 (median: 0.354) where one time series belongs to the blue region and the other to the red region. No bootstrap was performed for this sample analysis. 6
- 4 Analyses for the 2015-2016 influenza season at three different levels of spatial resolution. Each resolution level is defined by a choice of catchment areas: 13 administrative regions (panels **A₁** and **A₂**), 22 administrative regions (panels **B₁** and **B₂**) and 96 departments (panels **C₁** and **C₂**). The top panels show graphical representations of the detected epidemic regions. The bottom panels (**A₂**, **B₂**, **C₂**) represent the histograms of the adjusted R^2 values, calculated between time series. We used blue (red) bars to show the histogram of R^2 for time series within the first (second) epidemic region. Yellow bars show the histogram of R^2 in the case one time series belongs to the blue region and the other to the red region. The median values of the adjusted R^2 for the blue and red histograms exceed that of the corresponding yellow histogram, for each resolution level (i.e., 0.811 and 0.720 > 0.444 in panel **A₂**; 0.723 and 0.750 > 0.354 in panel **B₂**; 0.477 and 0.324 > 0.117 in panel **C₂**, respectively). Analyses at the level of the 13 administrative region implies the highest signal-to-noise ratio, but the spatial resolution is rather coarse. For the highest resolution level (panels **C₁** and **C₂**), noise is important for the time series; the resulting R^2 values are relatively low and, searching for community structure at this level, unjustified. Compromise is achieved for the middle ground of spatial resolution, provided by the 22 administrative regions. No bootstrap was performed for this sample analysis. 7
- 5 Spatial heterogeneity of weekly ILI incidence per 100 000 for the 2015-2016 (panel **A**) and the 1985-1986 (panel **B**) influenza seasons, at the time of the epidemic alert. The values show the difference between local and national ILI incidence for each of the 22 regions. Data is missing for Basse-Normandie in panel **B**. For comparison, the spatial structures of time-correlated epidemic regions (cf., Figure 2 of the main text) are shown in panels **C** and **D**, respectively. 8

1 Web Appendix 1

Illustration of the alignment of modularity vectors. Starting from the original dataset, we construct a large collection of surrogate datasets (i.e., 10 000, in the main-text analyses), using parametric bootstrap. Classifying each dataset using modularity maximization yields a partition of the catchment areas (nodes) into a number of communities, C , that is determined by the algorithm in every single run. The output corresponding to each dataset is stored as a *modularity vector*, a vector assigning a community label to each catchment area. However, the number of communities and the community labeling is not preserved from run to run. To use the set of modularity vectors as a statistical ensemble, we *align* the vectors, searching for a consistent labeling of communities across the entire set. We assume that noise in the data is small, so that the partitioning of the surrogate datasets is nearly identical and modularity vectors can be made nearly identical by relabeling the communities found in each case.



Web Figure 1: Alignment of modularity vectors. Two steps of analysis are illustrated: (a) relabeling of the communities for each vector such that the number of symbols in common with the reference (i.e., first) vector is maximized; (b) computation of f_a^i , the fraction of times catchment area i occurred in community (epidemic region) a .

The left side of Fig. 1 illustrates 4 column vectors, as a hypothetical set of 4 modularity vectors, emerging from classifying 4 surrogate datasets, describing 5 catchment areas, coded 1 through 5. Entry i (i runs from 1 to 5) of a particular modularity vector provides the label of the community where catchment area i was found in classifying that particular dataset. The numbers of communities, C , found across the datasets are: 3, 3, 3 and 2; $C_{\max} = 3$. However, the community labeling is not preserved from run to run. In fact, it is easily seen that interchanging 1 and 3 (i.e., using the permutation 3 2 1 of 1 2 3) in the second vector makes the second vector identical to the first. Such a transformation is not possible for the third vector, because the partition of catchment areas is slightly different. Still, we can choose a permutation of 1 2 3 which maximizes the number of symbols that the third and first vectors have in common. A consistent relabeling of communities for each modularity vector (see main text for the general recipe) yields a set of *aligned* modularity vectors; see center of Fig. 1. This is used to compute f_a^i , the fraction of times catchment area i was classified in community a , over all bootstrap analyses; see the right side of Fig. 1. Note that we necessarily have $\sum_{a=1}^{C_{\max}} f_a^i = 1$. In Fig. 1, we found a total of $C_{\max} = 3$ communities (epidemic regions) for 5 catchment areas. Hence, the triplet $(f_1^i f_2^i f_3^i)$ helps define a color in the RGB palette for catchment area i to be represented on the map.* If colors depart significantly from pure red, green or blue (i.e., f_a^i departs significantly from 0 and 1), then we say that noise in the time series data made a significant impact on the data classification. The f_a^i values are further used to quantify the impact of noise through a *bootstrap score*.

*In fact, since $f_1^i + f_2^i + f_3^i = 1$, the color palette is the additive RGB color triangle.

2 Web Appendix 2

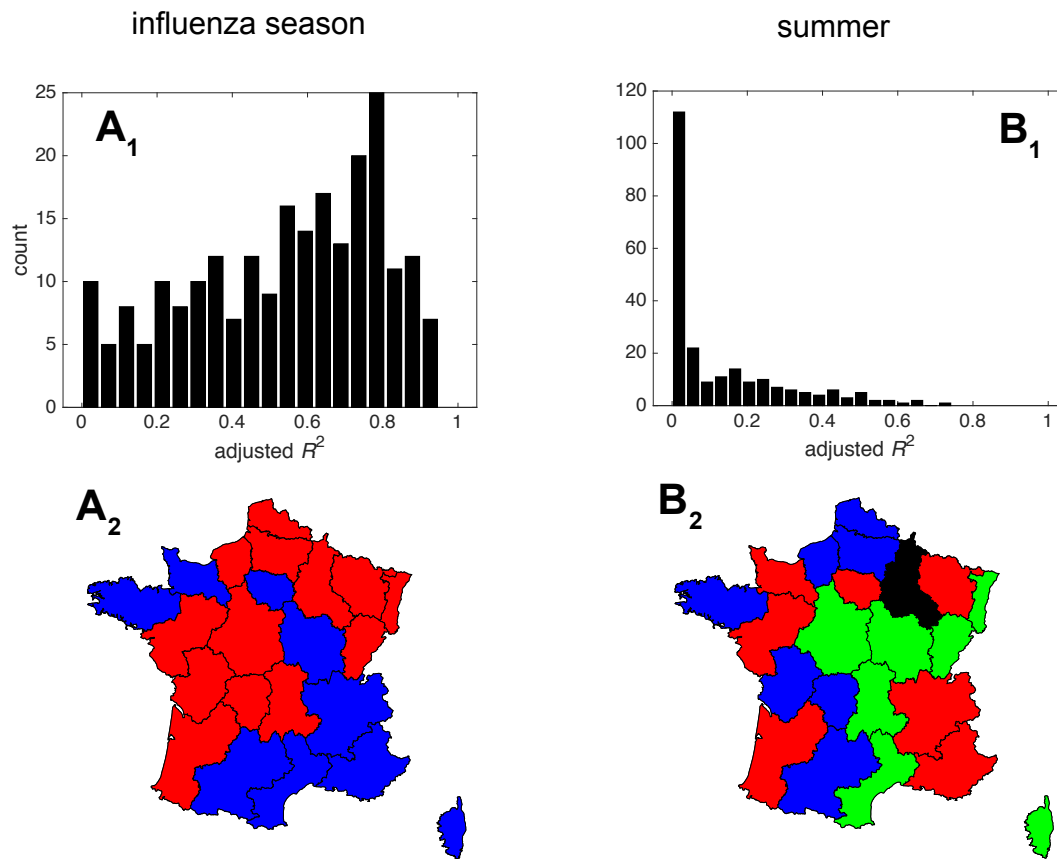
On the definition of the bootstrap score. A single run of data analyses (steps (i-ii) in the main text) uncovers the number of epidemic regions, denoted by C , and assigns each of the N catchment areas to one epidemic region. However, all this may change in a different run of analyses with bootstrapped data. The number of epidemic regions could vary from 1, when all catchment areas are grouped together, to N , when each individual catchment area forms an epidemic region; the maximum number of epidemic regions found over all bootstrap analyses C_{\max} is less or equal to N . Furthermore, the assignment of catchment areas may change even though the number of epidemic regions does not. Each catchment area i is thus characterized by the fraction of times it belonged to community a in the bootstrap analyses, f_a^i . For each fixed i , $\{f_a^i; a = 1, \dots, C_{\max} \leq N\}$ represents a set of normalized, independent probabilities, $\sum_{a=1}^{C_{\max}} f_a^i = 1$. If there exists b , $1 \leq b \leq C_{\max}$, such that $f_b^i = 1$, we say that catchment area i is robustly assigned to epidemic region b . The worse case scenario is $f_a^i = 1/N, \forall a = 1, \dots, C_{\max} = N$. We use the normalized Shannon entropy $S^i = -\sum_{a=1}^{C_{\max}} f_a^i \log f_a^i / \log N$ to quantify the information on the assignment of catchment area i . For robust assignments, $S^i = 0$, while for the worst case scenario S^i takes its maximum value of 1. We take the average Shannon entropy over the catchment areas to quantify the information for the whole surveillance area $B = \sum_{i=1}^N S^i / N$. Therefore, B may be written as

$$B = -\sum_{a=1}^{C_{\max}} \sum_{i=1}^N \frac{f_a^i \log f_a^i}{N \log N}. \quad (1)$$

It is important to note that, in general, B is not an entropy because the complete set of f_a^i values, $\{f_a^i; a = 1, \dots, C_{\max} \leq N, i = 1, \dots, N\}$ cannot always be viewed as a set of independent, normalized probabilities.

3 Web Appendix 3

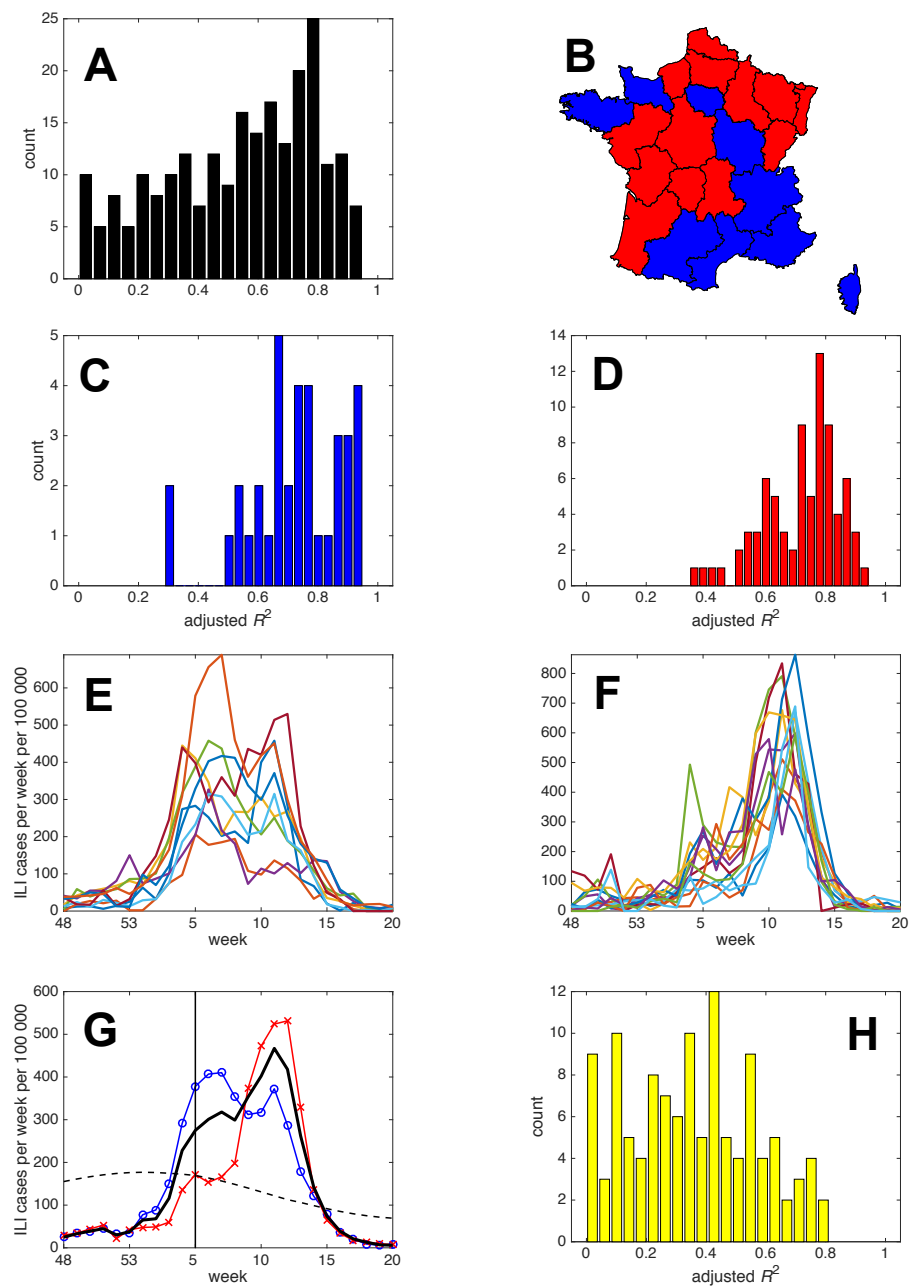
Comparison of results for the 2015-2016 influenza season and the 2015 summer season: single run analysis.



Web Figure 2: Comparison of results. Panels **A₁** and **B₁** show histograms of the adjusted R^2 values calculated for each pair of time series, for the 2015-2016 winter season (weeks 48-20) and the 2015 summer season (weeks 22-47) respectively. We note a substantial difference in the median R^2 values (0.587 and 0.042, respectively). For the influenza season, many time series of catchment areas correlate well, motivating a search for community structure. In contrast, off-season, the correlation is much poorer and the search for community structure not motivated. The catchment areas are represented by the 22 administrative regions, colored according to their inclusion into different epidemic regions (panels **A₂** and **B₂**, respectively). The community structure in panel **A₂** (2 epidemic regions) is meaningful, the community structure in panel **B₂** (4 epidemic regions) is spurious. No bootstrap was performed for this sample analysis.

4 Web Appendix 4

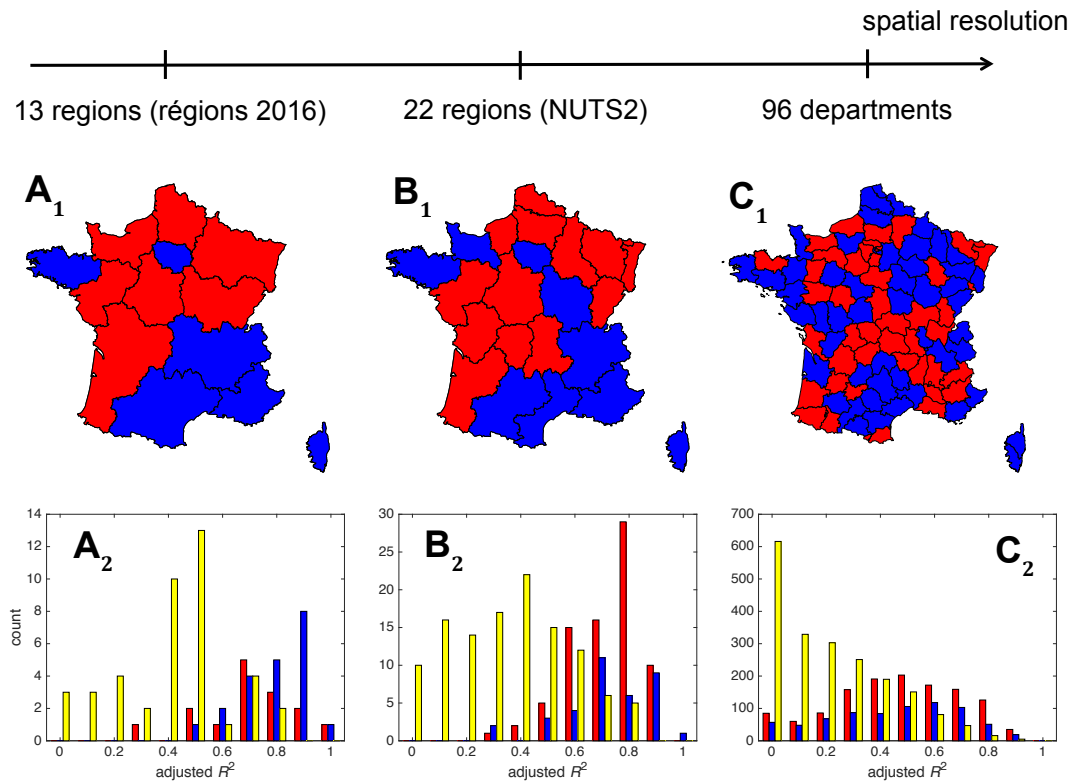
Sample, single-run analyses for the 2015-2016 influenza season.



Web Figure 3: Sample analyses for the 2015-2016 influenza season. Panel **A** shows the histogram of all adjusted R^2 values. Panel **B** shows the map of epidemic regions. The histograms of the adjusted R^2 values (corresponding medians: 0.723 and 0.750) for all pairs of time series in each epidemic region, blue and red, are shown in the panels **C** and **D**, respectively. The panels **E** and **F** represent the corresponding time series. Panel **G** shows the time series aggregated over the blue and red regions, respectively. The adjusted R^2 between the aggregated time series is 0.492, indicating poor correlation. The national average is plotted in black and the epidemic threshold is represented as the dashed line. Panel **H** shows the distribution of the adjusted R^2 (median: 0.354) where one time series belongs to the blue region and the other to the red region. No bootstrap was performed for this sample analysis.

5 Web Appendix 5

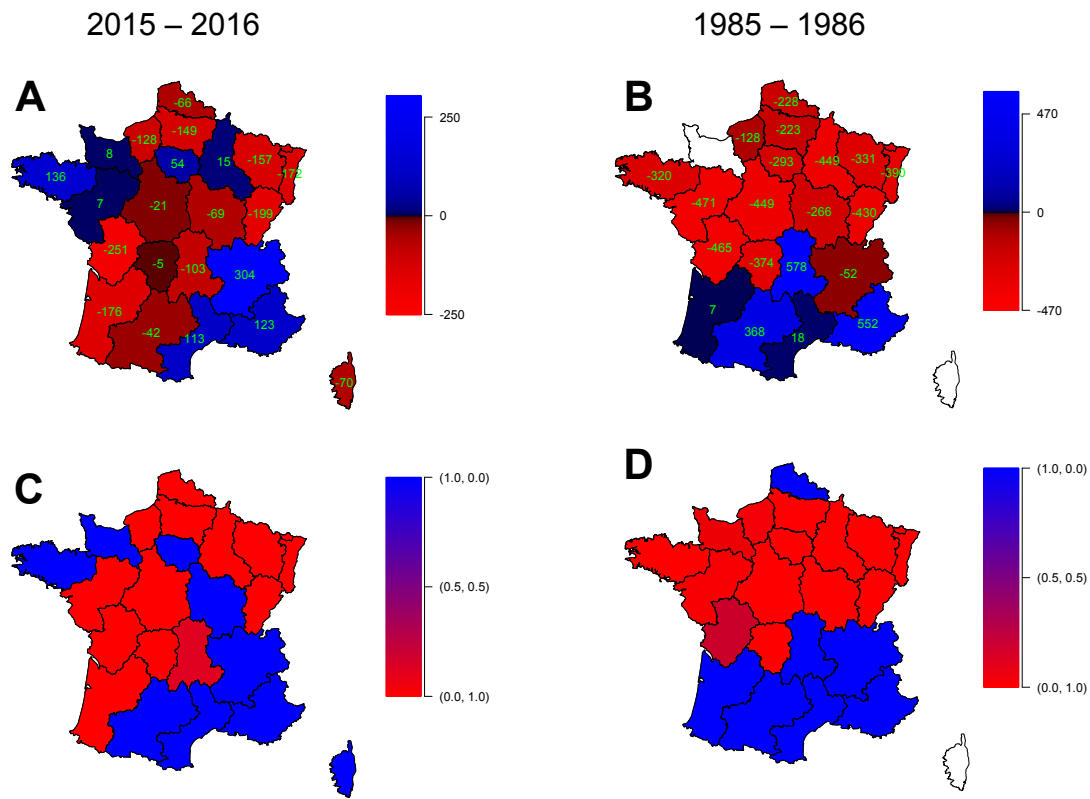
Comparison of results for three levels of spatial resolution.



Web Figure 4: Analyses for the 2015-2016 influenza season at three different levels of spatial resolution. Each resolution level is defined by a choice of catchment areas: 13 administrative regions (panels A₁ and A₂), 22 administrative regions (panels B₁ and B₂) and 96 departments (panels C₁ and C₂). The top panels show graphical representations of the detected epidemic regions. The bottom panels (A₂, B₂, C₂) represent the histograms of the adjusted R^2 values, calculated between time series. We used blue (red) bars to show the histogram of R^2 for time series within the first (second) epidemic region. Yellow bars show the histogram of R^2 in the case one time series belongs to the blue region and the other to the red region. The median values of the adjusted R^2 for the blue and red histograms exceed that of the corresponding yellow histogram, for each resolution level (i.e., 0.811 and 0.720 > 0.444 in panel A₂; 0.723 and 0.750 > 0.354 in panel B₂; 0.477 and 0.324 > 0.117 in panel C₂, respectively). Analyses at the level of the 13 administrative region implies the highest signal-to-noise ratio, but the spatial resolution is rather coarse. For the highest resolution level (panels C₁ and C₂), noise is important for the time series; the resulting R^2 values are relatively low and, searching for community structure at this level, unjustified. Compromise is achieved for the middle ground of spatial resolution, provided by the 22 administrative regions. No bootstrap was performed for this sample analysis.

6 Web Appendix 6

Incidence maps in comparison with maps of epidemic regions.



Web Figure 5: Spatial heterogeneity of weekly ILI incidence per 100 000 for the 2015-2016 (panel **A**) and the 1985-1986 (panel **B**) influenza seasons, at the time of the epidemic alert. The values show the difference between local and national ILI incidence for each of the 22 regions. Data is missing for Basse-Normandie in panel **B**. For comparison, the spatial structures of time-correlated epidemic regions (cf., Figure 2 of the main text) are shown in panels **C** and **D**, respectively.