1	CoViD-19, learning from the past: A wavelet and cross-correlation analysis of the epidemic					
2	dynamics looking to emergency calls and Twitter trends in Italian Lombardy region					
3	Wavelet and cross-correlation analysis of CoViD-19 spread dynamics					
4						
5	Bruno Alessandro Rivieccio ^{1*¶} , Alessandra Micheletti ^{2¶} , Manuel Maffeo ^{3,4} , Matteo Zignani ⁵ ,					
6	Alessandro Comunian ⁶ , Federica Nicolussi ⁷ , Silvia Salini ⁷ , Giancarlo Manzi ⁷ , Francesco Auxilia ^{3,8} ,					
7	Mauro Giudici ⁶ , Giovanni Naldi ² , Sabrina Gaito ⁵ , Silvana Castaldi ^{3,9&} , Elia Biganzoli ^{10&}					
8						
9	¹ Department of Laboratory Medicine, Division of Anatomic Pathology, Niguarda Hospital, Milan,					
10	Italy					
11	² Department of Environmental Science and Policy, University of Milan, Milan, Italy					
12	³ Department of Biomedical Sciences for Health, University of Milan, Milan, Italy					
13	⁴ Public Health Post Graduate School, University of Milan, Milan, Italy					
14	⁵ Department of Computer Science, University of Milan, Milan, Italy					
15	⁶ Department of Earth Sciences, University of Milan, Milan, Italy					
16	⁷ Department of Economics, Management and Quantitative Methods & Data Science Research					
17	Center, University of Milan, Milan, Italy					
18	⁸ ASST FBF-Sacco, Milan, Italy					
19	⁹ Fondazione IRCCS Ca' Granda Ospedale Maggiore, Milan, Italy					
20	¹⁰ Department of Clinical Sciences and Community Health & Data Science Research Center,					
21	University of Milan, Milan, Italy					
22						
23	* Corresponding author					
24	E-mail: brunoalessandro.rivieccio@ospedaleniguarda.it					
25						
26	[¶] These authors contributed equally to this work.					
27	^{&} These senior authors also contributed equally to this work.					

28 Abstract

The first case of Coronavirus Disease 2019 in Italy was detected on February the 20th in 29 30 Lombardy region. Since that date, Lombardy has been the most affected Italian region by the epidemic, and its healthcare system underwent a severe overload during the outbreak. From a public 31 32 health point of view, therefore, it is fundamental to provide healthcare services with tools that can reveal possible new health system stress periods with a certain time anticipation, which is the main 33 34 aim of the present study. Moreover, the sequence of law decrees to face the epidemic and the large 35 amount of news generated in the population feelings of anxiety and suspicion. Considering this 36 whole complex context, it is easily understandable how people "overcrowded" social media with 37 messages dealing with the pandemic, and emergency numbers were overwhelmed by the calls. 38 Thus, in order to find potential predictors of possible new health system overloads, we analysed data both from Twitter and emergency services comparing them to the daily infected time series at a 39 regional level. Particularly, we performed a wavelet analysis in the time-frequency plane, to finely 40 41 discriminate over time the anticipation capability of the considered potential predictors. In addition, 42 a cross-correlation analysis has been performed to find a synthetic indicator of the time delay 43 between the predictor and the infected time series. Our results show that Twitter data are more related to social and political dynamics, while the emergency calls trends can be further evaluated 44 as a powerful tool to potentially forecast new stress periods. Since we analysed aggregated regional 45 data, and taking into account also the huge geographical heterogeneity of the epidemic spread, a 46 future perspective would be to conduct the same analysis on a more local basis. 47

- 48
- 49
- 50
- 51
- 52

53 Introduction

54

On February the 20th the first Italian case of Coronavirus Disease 2019 (CoViD-19) due to secondary transmission outside China was identified in Codogno, Lombardy region [1]. In the following days the number of cases started to rise not only in Lombardy but also in other Italian regions, although Lombardy remained and is still the most affected region in Italy [2]. At the time of writing (October the 4th), 325,329 cases have been identified in Italy, out of which 108,065 in Lombardy region [3].

The progressive decrease of CoViD-19 cases should not let our guard down, indeed it is 61 62 clear that, since the beginning of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-63 2) pandemic, public health has suffered from the absence of a proper preparedness plan to face an 64 episode which was unexpected and unpredictable and has heavily impacted on the territorial and 65 hospital healthcare services. Planning has a fundamental role nowadays but it can be adequate only 66 if the next possible pandemic peak can be effectively foreseen by means of a predictive tool which 67 accounts for all the available signals. In order to do so, it is of paramount importance to learn from 68 what happened during the first peak to be prepared for the potential next one.

The SARS-CoV-2 outbreak in Italy has been characterised by a massive spread of news coming from both official and unofficial sources leading to what has been defined as infodemia, an over-abundance of information – some accurate and some not – that has made hard for people to find trustworthy sources and reliable guidance needed [4].

73 Infodemia on SARS-CoV-2 created the perfect field to build suspicion in the population,
74 which was scared and not prepared to face this outbreak. It is understandable how the rapid increase
75 of the cases number, the massive spread of news and the adoption of laws to face this outbreak led
76 to a feeling of anxiety in the population, whose everyday life changed very quickly.

A way to assess the dynamic burden of social anxiety is a context analysis of major social networks activities over the internet. To this aim Twitter represents a possible ideal tool, because of the focused role of the tweets according to the more urgent needs of information and communication, rather than general aspects of social projection and debate as in the case of Facebook, which could provide slower responses for the fast individual and social context evolution dynamics [5].

Taking into account this specific context, it is easy to understand why the 112 emergency number service in Lombardy region was suddenly overwhelmed by an enormous number of calls that rapidly overcame its capacity to cope and compromised the possibility to identify those patients who needed immediate medical assistance [6].

As pointed out by the Scientific Italian Society for Medical Emergency (SIEMS), the number of calls to 112 for the Milan province was 5,086 on February the 16th, before the outbreak, and rapidly increased to 6,798 on February the 21st and to 10,657 on February the 22nd [7].

The emergency service in Lombardy region is organised through three first-level PSAPs (public-safety answering points), called CUR-NUE (Unique answering operating room / point – European emergency number), which forward the call to the most appropriate service, i.e. Police, Fire Department or Medical emergency rescue service. So, after the first assessment, calls requiring medical assistance are sent to one of the four second-level PSAPs called SOREU (Regional Operating Rooms for Medical Emergency and Urgency), depending on the geographical area the call is coming from in order to evaluate the patient and decide the most appropriate intervention.

Lombardy region, from an administrative point of view, is made up of twelve provinces, and
the management of medical emergency calls occur in four operating rooms which receive these
calls from different sets of provinces, namely:

• SOREU delle Alpi (SRA) from the provinces of Bergamo, Brescia and Sondrio;

• SOREU dei Laghi (SRL) from the provinces of Como, Lecco and Varese;

• SOREU Metropolitana (SRM) from the provinces of Milan and Monza-Brianza;

• SOREU della Pianura (SRP) from the provinces of Cremona, Lodi, Mantova and Pavia.

Some of the non-urgent calls received by the four SOREU do not need, though, an 104 105 ambulance dispatch, so not all the calls result in a medical rescue mission: in the latter cases the 106 patients are recommended by SOREU medical technicians to consult other medical services such as 107 general practitioners. Moreover, during the epidemic, according to specific internal procedures, 108 SOREU medical technicians answering to the emergency calls, in the case of signs and/or 109 symptoms evocative of CoViD-19 but not life threatening, advised patients to wait for a recall by a 110 public health medical doctor: after this re-evaluation call, it was up to these medical doctors the 111 final decision on the management of the case (ambulance dispatch or home quarantine).

112 To reduce the burden of calls to the emergency number which occurred during the first days 113 of the outbreak, it was necessary to redirect non-urgent calls, especially those asking for 114 information, to other services. According to European Emergency Number Association guidelines 115 [8], Lombardy region created a regional toll-free number for CoViD-19, the first one in Italy. Other 116 Italian regions created their own one in the following weeks, as well as other European countries 117 like Spain, Germany, Croatia, which were facing similar issues [9].

118 The 24/24-hour toll-free number was settled on February the 23rd by AREU (Regional 119 Emergency Service Agency) and, although it helped to funnel non-urgent calls, it was not enough 120 because of the huge number of calls: for example on the second day it received more than 400,000 121 calls.

Calls to the emergency services could be an important and helpful indicator of the spread of the infection among the population, taking into account the possibility to analyze data regarding the municipality from which the calls originated and the motivations that induced people to ask for fast medical support. Statistical models could be used to assess the association of these data with new cases of CoViD-19 in order to predict new epidemic hotspots on a municipal scale, or with a smaller spatial scale for big cities.

In addition to usual public health indicators, social media data may also be used as probes of the people behaviour according to the recent trends of digital epidemiology. As mobile technology continues to evolve and proliferate, social media are expected to occupy an increasingly prominent role in the field of infectious diseases [10-12].

A recent systematic review concluded that the inclusion of online data in surveillance systems has improved the disease prediction ability over traditional syndromic surveillance systems and Twitter was the most common social network analysed for this aim [13].

Despite some limitations and concerns, a better understanding of the behavioural change induced by social media can strengthen mathematical modelling efforts and assist in the development of public policy so as to make the best use of this increasingly ubiquitous resource in controlling the spread of disease [11].

Aim of the study is to understand the correlation between the users calls to the emergency 139 140 services and the spread of the infection in the population during the first peak of the CoViD-19 141 outbreak in Lombardy region of Italy, the first world hotspot after the Chinese raise in Wuhan. 142 Furthermore, the joint analysis with Twitter trends related to emergency was performed to better 143 understand the most important population concerns according to the infection dynamics. Overall, 144 the joint active monitoring of the communication dynamics over emergency calls and social 145 networks like Twitter could provide an integrated means for the adaptive management of 146 information delivery as well as the optimization of the rescue logistic and finally it could provide 147 relevant anticipation on the outbreak. These aspects appear of critical importance for CoViD-19 148 surveillance, and for the preparedness of emergency and strategic plans [14].

149

150

151

153 Materials and methods

154

155 **Data**

- 156 In the present work, we analysed the following time series:
- PSAP-II SOREU-118 daily incoming calls from 2020.02.18 to 2020.03.30 [15];
- PSAP-I CUR NUE-112 daily incoming calls from 2020.02.18 to 2020.03.30 [16];
- toll-free number daily incoming calls from 2020.02.23 to 2020.03.30 [16];

• daily Twitter data (tweets, replies, likes, retweets) from 2020.02.18 to 2020.06.29 [17];

daily infected from 2020.02.24 (the first day since which Italian Department for Civil Defense
 has provided data) to 2020.06.29 [3].

102 has provided data) to 2020.00.25 [5].

Data about SOREU-118, NUE-112 and toll-free number daily incoming calls were only 163 164 available at a regional level. The very first days of the NUE and toll-free number time series have 165 been discarded due to the very intense population panic reaction which reflected into a very huge 166 amount of calls (whose peak was even higher than the following, new cases-related one). 167 Particularly, in the case of NUE they were inappropriate non-urgent calls (most of all for 168 information need), so they were not forwarded to the corresponding SOREU: indeed, in the SOREU 169 time series we do not observe any peak in the very first days. Moreover, this choice is justified if we 170 consider that – in case of a new epidemic burden – there would not be such a powerful reaction, so 171 to the aim of predictability we can take into account just the subsequent new increase in the calls to 172 NUE and toll-free number, which is more related to the CoViD-19 dynamics. Twitter data, instead, 173 differently from the emergency calls, were not geolocalised. Finally, daily new cases have been 174 collected at the province level and then aggregated at the regional level.

175

176

178 **Twitter data analysis**

179 The monitoring of the communication dynamics on online social media has been conducted 180 on Twitter [18]. Specifically, the Twitter Search API (Application Programming Interface) was used to collect all the original tweets in Italian language containing the keywords "112" or "118" in 181 182 the body text, discarding retweets. The data span the period from 2020.02.18 to 2020.06.29: "(112 or 118) lang:it since:2020-02-18 until:2020-06-29" was the query submitted to the Twitter Search 183 184 API. We tokenised and lemmatised the text of each tweet by the POS (Part of Speech)-tagger provided by the state-of-art natural language processing (NLP) library in Python, spaCy [19]. After 185 186 the lemmatization, for each tweet we only kept the lemma of nouns and verbs. By inspecting the set 187 of lemmas over the whole corpus of tweets, we manually identified the most common terms related 188 to emergency, bringing to the identification of 283 keywords (see the Support Information for the 189 list of keywords). As a rationale in the choice of the keywords, we kept terms related to: a) the 190 reasons why people called for medical assistance (CoViD-19-related symptoms, parts of the body, 191 words referring to family members or relatives); b) health workers or medical tools involved in the 192 emergency calls (such as "doctor", "nurse", "oximeter"...); and c) the general context of the 193 emergency situation (words dealing with public health policies such as "lockdown", complaints for 194 the saturation of the emergency number services...). Finally, all the tweets that contained only the search keywords "112" or "118" in the text were excluded from the Twitter dataset, leading to a 195 196 total amount of 16,216 tweets (or "statuses", in Twitter jargon) which were used for the purpose of 197 this paper. The identification of the 283 keywords related to the emergency and the filtering based on the co-presence of these keywords with the search keywords "112" or "118" (e.g. "112" and 198 199 "doctor", "118" and "ambulance"...) impacted on the reduction of the number of tweets: indeed tweets semantically not related to the topics, which however contained the terms "112" or "118", 200 201 have been discarded.

202 Since both the timestamp and the number of likes, retweets and replies at the moment of the 203 data collection were available for each tweet, it was possible to reconstruct four time series related 204 to the dynamics of the emergency calls and to the perception of the situation on Twitter: 1) the 205 production of new tweets per day, 2) the number of retweets of new tweets per day, 3) the number 206 of likes per day, and 4) the number of replies per day. The last three indicators have been computed 207 by aggregating the number of the interactions got by the tweets, day by day. Moreover, they capture 208 different aspects in the perception of the emergency. Indeed, the daily number of replies is a first 209 indicator of the level of discussions triggered by the tweets, and it might represent the reaction to a 210 particular policy or the support of the Twitter community to other users needing medical assistance. 211 Instead, the daily number of retweets and of likes represent an early approximation of the 212 endorsement to the content of the tweet, and, in this specific context, a further involvement with the 213 content of the tweets, in terms of emotional and situational closeness. For instance, a person may 214 retweet a content reporting a call for assistance because he has experienced the same situation.

215

216 Wavelet analysis

217 Wavelets represent a powerful tool to analyze localised variations of non-stationary power at 218 many different frequencies in time series: the decomposition of the signal information in the time-219 frequency domain, indeed, allows to bring out variability and its changes over time [20-23]. 220 Particularly, wavelets can "capture" and detect in the time-frequency (scale) plane both long-period 221 and short-period trends. In other words, since the period and the frequency of a signal are inversely 222 proportional, long-period trends correspond to low-frequency components, and are properly called 223 "trends" or "backgrounds", while short-period trends correspond to high-frequency components, 224 and are called "anomalies" or "discontinuities". Anomalies, despite their limited spatio-temporal 225 location, possess a huge amount of information content, thus it is of primary importance to reveal 226 them adequately. Among the others, the wavelet transform has relevant features such as a good

capability of time-frequency localization (useful to analyze signals changing over time), and offers
the possibility of a multi-resolution representation over different scales [20]. Wavelets transforms,
differently from other series expansions, are thus a suitable tool to identify both a trend in a
nonparametric form, keeping significant local peaks, and variations from the trend [24-25].

231 Wavelet analysis is therefore useful to detect stress moments in health systems during 232 emergencies, as in the case of CoViD-19 pandemic. By comparing peaks of two time series in a 233 given time interval, one can be informed with anticipation by a surrogate endpoint about the 234 forthcoming stress peaks of the healthcare system. Namely, in the early pandemic outbreak, peaks 235 in the emergency calls should possibly anticipate those for new infections and hospital admissions. 236 However, for the aforementioned aims, time-frequency localization of signal components in highly 237 discontinuous processes as a pandemic is a critical task requiring the modelling features of the 238 wavelets analysis. Methods as those based on conventional spectral analysis such as discrete Fourier 239 transform are not suited for the same aims since it assumes periodicity in the time series [26]. 240 Moreover, while wavelet transform allows a variable higher time-frequency resolution according to 241 the needs, other alternative methods which are able to detect frequencies changing over time display 242 some limitations. Indeed, short time Fourier transform has a fixed time-frequency resolution, and 243 Hilbert-Huang transform applied after empirical mode decomposition of the signal only computes 244 the instantaneous frequency of each empirical mode, without highlighting the non-stationarity of the 245 frequencies contributions [26-30].

In order to identify the trends, all the time series were first smoothed using a moving average linear filter of a 7-days amplitude, decomposed through the wavelet transforms and then normalised to their maximum values. Tweets-dependent data (daily number of replies, likes and retweets) were also previously normalised to the correspondent number of daily tweets.

Details on the adopted continuous wavelet transform (CWT) and related measures are reported in the Supplementary material. Briefly, first of all, the CWT of all the signal was computed, together with the magnitude of the wavelet transform (using the modulus of the complex

253 values). After that, for each pair of time series, the following quantities were calculated: the 254 magnitude-squared wavelet coherence (MSWC), the wavelet cross-spectrum (WCS), the cone of 255 influence, the phase coherence relationship (using the argument of the complex values), the time 256 delay between the two signals (using the phase lag values). Each value of scale was converted to the 257 equivalent Fourier value of frequency, and thus to the correspondent period. However, note that the 258 relationship between scale and frequency is only an approximation, since there is not a precise correspondence between the two: among others, Meyers and coll. (1993) proposed a method for the 259 260 conversion from scales to "pseudo-frequencies" [31].

Wavelet coherence and cross-spectrum analysis provides a detailed both time- and frequency-localised information on the phase lag and thus on the time delay between the compared signals, thanks to the decomposition through CWT, but lacks a global view of the trends and their relative shifts over time. Thus, in order to give a synthetic and unique indicator of the similarity over time between the signals, we also performed a time domain analysis estimating their crosscorrelation sequence.

267

Time lag estimation through cross-correlation analysis

In signal processing, another way to determine the similarity of two discrete time sequences 269 270 is the cross-correlation. Indeed, since cross-correlation measures the relation between a vector x and 271 time-shifted (lagged) copies of another vector y as a function of the lag itself, it allows also a time-272 delay analysis between the signals. Moreover, by revealing their relative displacement in time, the 273 lag at which there is the maximum correlation can be considered as the time shift necessary to align 274 the series x and y by sliding y backward (negative lag, in case y has a delay compared to x) or forward (positive lag, in case y displays an anticipation respect to x). To this aim, we computed the 275 276 cross-correlation function between the time series, at different lags. The lag with the maximum 277 cross-correlation value was thus identified as the "typical time delay" leading the lagged signal.

278	A 90% confidence interval was computed both for the cross-correlation values and for the					
279	lag corresponding to the cross-correlation function peak, i.e. for the days of delay. We used two					
280	different methods to calculate the interval estimation in order to compare parametric and non-					
281	parametric estimates: 1) a Fisher's <i>z</i> statistics through the Fisher's <i>z</i> -transformation [32]; 2) a Monte					
282	Carlo method through a surrogate time series with the same auto-correlation of the original one, for					
283	1,000 simulations [33]. Details and notes on these two confidence interval computation methods are					
284	provided in the Supplementary material.					
285	Cross-correlation could also be estimated using wavelets, specifically through a maximal					
286	overlap discrete wavelet transform (MODWT): it would have been our intention to complete the					
287	analysis with wavelets, but since emergency calls time series are made up of few samples, there are					
288	not enough non-boundary coefficients, even at the first level, to compute the wavelet cross-					
289	correlation function for a sufficient number of lags (see explanation about the edge effects in the					
290	Supplementary material).					
291						
292	All the previous reported analyses were performed using MatLab R2020a, The MathWorks					
293	Inc.					
294						
295						
296	Results					
297						
298	Wavelet analysis					
230						
299	As already pointed out before, the following daily regional aggregated data were considered:					
300	• toll-free number incoming calls,					
301	• NUE-112 incoming calls,					
302	• SOREU-118 incoming calls.					

303 In addition, we compared also Twitter data to the number of regional daily infected patients.

The WCS and the MSWC were calculated for each of these time series in relation to the data of regional daily infected. Indeed, both wavelet cross-power spectrum and coherence, through the CWT, can show areas in the time-frequency space where two signals share common harmonic components. In particular, the focus will be on the areas for which coherence is higher than 0.5 (indicated by the arrows, see Figs 1-7), since for lower values of coherence the phase lag is not reliable.

310 In the following figures (Figs 1-7), the time courses of the smoothed and normalised series 311 are displayed on the left, whereas the WCS/MSWC representation is on the right. In the time 312 courses charts, the daily infected curve is in red, while the potential predictor curve is in blue. In the 313 WCS/MSWC charts, the x axis represents time (days), the y axis (logarithmic scale) represents scale 314 (which has been converted to the equivalent Fourier frequency, cycles/day), and the color scale 315 represents the MSWC. The cone of influence, where edge effects should be considered, is shown as 316 a white dashed line. For areas where the coherence exceeds 0.5, the charts display arrows to show 317 the phase lag between the two signals. The arrows, which do not represent vectors since the length 318 is not proportional to the intensity, are spaced in time and scale. The direction of the arrows 319 designates the relative phase on the unit circle: a rightward-pointing arrow indicates in-phase coherence relationship of the two signals ($\Delta \varphi = 0$); a leftward-pointing arrow indicates anti-phase 320 321 coherence relationship ($\Delta \varphi = \pi$). The corresponding lag in time depends on the duration of the 322 cycle (period).

323

Fig 1. Regional toll-free number daily incoming calls vs. daily infected time courses and wavelet analysis. On the left, the smoothed (7-days amplitude moving average) and normalised time courses are displayed (toll-free number calls in blue, daily infected in red); on the right, WCS and MSWC chart is shown (see text for explanation).

329 Fig 2. Regional NUE daily incoming calls vs. daily infected time courses and wavelet analysis.

On the left, the smoothed (7-days amplitude moving average) and normalised time courses are displayed (NUE calls in blue, daily infected in red); on the right, WCS and MSWC chart is shown (see text for explanation).

333

Fig 3. Regional SOREU daily incoming calls vs. daily infected time courses and wavelet analysis. On the left, the smoothed (7-days amplitude moving average) and normalised time courses are displayed (SOREU calls in blue, daily infected in red); on the right, WCS and MSWC chart is shown (see text for explanation).

338

Fig 4. Daily number of tweets vs. regional daily infected time courses and wavelet analysis. On the left, the smoothed (7-days amplitude moving average) and normalised time courses are displayed (tweets in blue, daily infected in red); on the right, WCS and MSWC chart is shown (see text for explanation).

343

Fig 5. Daily number of likes vs. regional daily infected time courses and wavelet analysis. On the left, the smoothed (7-days amplitude moving average) and normalised time courses are displayed (likes in blue, daily infected in red); on the right, WCS and MSWC chart is shown (see text for explanation).

348

Fig 6. Daily number of retweets vs. regional daily infected time courses and wavelet analysis. On the left, the smoothed (7-days amplitude moving average) and normalised time courses are displayed (retweets in blue, daily infected in red); on the right, WCS and MSWC chart is shown (see text for explanation).

353

354 Fig 7. Daily number of replies vs. regional daily infected time courses and wavelet analysis. On

the left, the smoothed (7-days amplitude moving average) and normalised time courses are displayed (replies in blue, daily infected in red); on the right, WCS and MSWC chart is shown (see text for explanation).

358

359 Looking at the time courses, it is evident a large anticipation (about two weeks) of the 360 emergency calls trends with respect to the epidemic dynamics: if we consider the peaks of the 361 curves, indeed, while the infected time series reaches its maximum at day 29, the toll-free number 362 and SOREU calls peak occur at day 15 (time delay -14 days), and the NUE calls curve reaches the 363 peak at day 16 (time delay -13 days). However, wavelet analysis does not reveal any strong 364 coherence for all these pair of signals: specifically, among all the previously mentioned time series, the only one for which wavelet analysis does not display any relevant coherence compared to daily 365 366 new cases data is that of regional toll-free number incoming calls (Fig 1). Instead, NUE regional 367 data and daily infected signals (Fig 2) display coherence over days from 18 to 22 at frequencies around 0.25 cycles/day, with a phase lag from -126.4° to -134.7°, corresponding to a time delay 368 369 from 2.5 days to 2.6 days. Interestingly, days from 18 to 22 are confined between the two peaks, 370 since the NUE calls curve reaches the peak at day 16, while the infected curve reaches its maximum 371 at day 29. Not surprisingly, wavelet cross-spectrum and coherence analysis between regional daily 372 incoming calls to SOREU and infected people (Fig 3) shows an anomaly less limited over time and 373 over scale (frequency band ranging from 0.25 cycles/day to 0.37 cycles/day, time interval from day 374 17 to day 28), with a phase shift between -72.1° and -111.2°, leading to a time delay of about 2.4 375 days.

Even if not geolocalised, we finally compared regional epidemic time series with Twitter data (Figs 4–7): just considering the time courses, it is evident that the best potential predictor is the tweets time series (peaks at day 22 and day 29, respectively for daily number of tweets and infected curve). Indeed, since replies, likes and retweets are variables dependent on the original tweets, they

200	are deleved in time: likes and retweats reach the peak at day 28 just one day before new cases						
500	are delayed in time: fixes and retweets reach the peak at day 28, just one day before new cases,						
381	loosing almost all the anticipation capability, while the maximum value of replies is achieved even						
382	later than the peak of the infected curve, at day 33. Wavelet analysis, focused on the predictability,						
383	confirms this consideration:						
384	• in the case of daily tweets (Fig 4) it detects two relevant areas of high coherence in the time-						
385	frequency plane:						
386	\circ a trend at the lowest frequencies (from 0.02 cycles/day to 0.03 cycles/day), with the						
387	phase relationship shifting from in-phase coherence to a maximum lag of 31.8° (time						
388	delay 2.6 days),						
389	\circ and a time-localised anomaly before the infected peak around the frequency of 0.25						
390	cycles/day, with a phase-lag ranging from 73.0° (day 20, 0.37 cycles/day) to 116.3° (day						
391	27, 0.23 cycles/day) and a subsequent time delay between 0.5 and 1.4 days;						
392	• for the number of daily likes (Fig 5) just a trend localised in the first half of the observation						
393	period and at the lowest frequencies is captured, with a maximum phase lag of 65.3° at 0.06						
394	cycles/day, corresponding to a time shift of 3.0 days;						
395	• daily number of retweets (Fig 6) display a similar trend more localised in the frequencies						
396	domain (from 0.04 cycles/day to 0.08 cycles/day), with a phase shift always near 0° (in the						
397	range between -26.7° and 19.6°);						
398	• finally, the replies time series (Fig 7) shows a background even more localised both in time and						
399	frequency, and an anomaly around the frequency of 0.14 cycles/day from day 19 to day 33, both						
400	with a coherence value around 0.6.						
401							

402 **Twitter trends**

403 The next figures (Figs 8-9) display raw data about Twitter trends (tweets, replies, retweets404 and likes).

Fig 8. Twitter trends (1). The trends of daily number of new tweets about emergency calls (blue line) and of replies (pink line) they sparked are shown here. The vertical green dotted lines indicate the principal episodes related to the lockdown policies in Italy. The orange box indicates the first peaks of activities, while the red circle highlights the highest peak in the number of replies.

409

410 Fig 9. Twitter trends (2). The trends of daily number of retweets (yellow line) and of likes (purple
411 line) are shown here.

412

413 It is evident that, concerning the dynamics of communications and the context on social 414 media, Twitter activity (Figs 8–9) is not so strictly related to the epidemic dynamics, since it is 415 triggered most of all by social, political and chronicle news, which drive an emotional participation 416 of the users. Indeed, as highlighted by the orange box in Fig 8, the first increase in all these time 417 series (tweets, replies, likes and retweets), from 2020/02/21 to 2020/02/25, precedes just the 418 establishment of the red areas in Codogno and Vo' Euganeo, while the second peak of daily tweets 419 on 2020/03/14 (red circle in Fig 8) is related to the death of an operator of SRA due to CoViD-19. 420 Moreover, likes and retweets (Fig 9) trends look more aligned to the announcements about 421 lockdown policies. Consequently, from the viewpoint of the management of the emergency, and 422 comparing the Twitter trends to the emergency calls dynamics, Twitter data, despite their real-time 423 nature, represent a weak signal for the real-time monitoring and forecasting of the outbreaks, rather 424 they are more suitable to measure the perception of the outbreak and the lockdown policies.

425

426 **Cross-correlation and time delay analysis**

427 While wavelet decomposition allows a detailed analysis in both the time and frequency 428 domains, it lacks a unique, global indication of the shift over time between two signals: to this aim, 429 we performed also a time delay analysis estimating the cross-correlation sequence for each pair of

430 time series. This analysis has been conducted only for the three cases for which WCS/MSWC 431 revealed a strong coherence in the ascending phase (before both the peaks or between them), which 432 is the most important from a public health monitoring point of view. These time series are the following: (i) daily regional incoming calls to NUE-112 (Fig 10); (ii) daily regional incoming calls 433 434 to SOREU-118 (Fig 11); (iii) daily number of new tweets (Fig 12). In the following figures, the 435 maximum of each function is depicted in red, and the confidence limits for the peak lag, deduced by 436 the 90% *z*-Fisher confidence bounds of the cross-correlation values, are reported below the figures, 437 in the caption. Negative lags denote by how many days the time series of infected patients should be 438 shifted backward over time to be "aligned" with the predictor.

439

Fig 10. Cross-correlation sequence estimate between NUE regional incoming calls and daily
infected. The blue lines represent the 90% confidence interval (CI) limits computed through a *z*transformation. The maximum value of the cross-correlation function is depicted in red. Peak lag
[CI]: -3 days [-8,1].

444

Fig 11. Cross-correlation sequence estimate between SOREU regional incoming calls and
daily infected. The blue lines represent the 90% confidence interval (CI) limits computed through a
z-transformation. The maximum value of the cross-correlation function is depicted in red. Peak lag
[CI]: -5 days [-11,1].

449

Fig 12. Cross-correlation sequence estimate between daily number of new tweets and daily
infected. The blue lines represent the 90% confidence interval (CI) limits computed through a *z*transformation. The maximum value of the cross-correlation function is depicted in red. Peak lag
[CI]: -6 days [-8,-2].

454

455	In addition, a sensitivity analysis of these results with respect to the amplitude of the initial
456	smoothing with a moving average filter was performed. The results are reported in Table 1.

457

458 Table 1. Sensitivity tests on the uncertainty in the location of the cross-correlation function

459 peak. Different moving-average amplitudes have been used to test the robustness of the results of

the confidence intervals (C.I.) computed through the *z*-transformation.

461		NUE calls		SOREU	J calls	tweets	
462	n. days	peak lag	C.I.	peak lag	C.I.	peak lag	C.I.
463	4	-4	[-10,0]	-5	[-11,1]	-6	[-8,-3]
464							
465	5	-4	[-9,0]	-5	[-11,1]	-6	[-8,-3]
466	6	-4	[-9,1]	-5	[-11,1]	-5	[-8,-2]
467 468	7	-3	[-8,1]	-5	[-11,1]	-6	[-8,-2]
469	8	-3	[-8,1]	-4	[-11,1]	-5	[-8,-2]
470	9	-2	[-8,1]	-4	[-10,0]	-5	[-8,-2]
471	10	-2	[-8,1]	-3	[-10,1]	-5	[-8,-2]
472							

473

474 Changes in the results when the amplitude of the window is varying are relatively small475 (Table 1), consequently it can be assumed that our results are robust with respect to this parameter.

476 Similar and consistent results have been obtained with the Monte Carlo method described in

the Supplementary material (Figs 13-15).

478

Fig 13. Cross-correlation analysis through a Monte Carlo simulation method for the regional
NUE daily incoming calls vs. the daily infected time series. A random phase test (see

181	Supplementary material) has been performed (1,000 simulations) to compute the time lag to "align"
182	the signals and the corresponding confidence interval (C.I.): time lag = -4 days (C.I11,1).
183	
184	Fig 14. Cross-correlation analysis through a Monte Carlo simulation method for the regional
185	SOREU daily incoming calls vs. the daily infected time series. A random phase test (see
186	Supplementary material) has been performed (1,000 simulations) to compute the time lag to "align"
187	the signals and the corresponding confidence interval (C.I.): time $lag = -6 days$ (C.I13,1).
188	
189	Fig 15. Cross-correlation analysis through a Monte Carlo simulation method for the daily new
190	tweets vs. the daily infected time series. A random phase test (see Supplementary material) has
191	been performed (1,000 simulations) to compute the time lag to "align" the signals and the
192	corresponding confidence interval (C.I.): time lag = -3 days (C.I11,4).
193	
194	
195	Discussion

496

2

Lombardy region has been the epicentre of the CoViD-19 epidemic in the Western Countries [34]. After the detection of the first case, on February the 20th, national and regional health authorities put in place several strategies to limit the spread of the infection and deal with the consequences of the increasing number of cases [35-37]. Indicators that can reveal and anticipate a rise of cases are of paramount importance to support the planning and interventions of the health service organization. Our study shows that the number of calls to emergency services could be a good indicator that can anticipate the need for hospitalization in the early pandemic outbreak.

Tô et al. [38] have recently adopted the wavelet theory to model the number I(t) of infected individuals at time t in the context of the classical susceptible – infected – recovered (SIR)

506 epidemiological model, but to the best of our knowledge this is the first time that wavelet analysis is 507 used to detect health system stressful periods during the CoViD-19 pandemic according to the 508 coherence analysis of the peaks from different time series related to specific marker events.

509 However, some considerations about the analysis we conducted seem to be appropriate. 510 With respect to wavelet analysis, wavelet decomposition in the time-scale plane has the advantage 511 of giving a huge, precise and detailed amount of information both about time localization and 512 frequency components. Nonetheless, this kind of analysis has also some limitations. First of all, 513 time and frequency resolutions, according to uncertainty Heisenberg principle, are inversely 514 proportional. Secondly, the edge effects affect the reliability of the results in the time-scale plane 515 outside the cone of influence (see Supplementary material for an explanation), so that for the lowest 516 frequencies the time interval with reliable results is very short. Lastly, one must consider also the 517 band-pass filtering action of the CWT: indeed, the frequency spectrum bounds are a function of the 518 number of samples of the signal (see Supplementary material for a detailed explanation), so the 519 availability of data over time could represents a limitation for this kind of analysis. In our specific 520 instance, it can be noticed that the inferior limit of the frequency domain in the case of Twitter data 521 is much lower than the one of the daily regional emergency calls time series (112, 118, toll-free 522 number), just because much more samples for the Twitter data are available (Figs 1-7). 523 Consequently, in the cases of the calls to the emergency services, the ability of wavelet 524 decomposition in revealing hidden signals and trends at the lowest frequencies is limited by the 525 poor number of collected data. This limitation also affects the WCS/MSWC computation and the 526 possibility of detecting large time delays between the signals. This occurs because the lowest 527 frequencies are cut off by the wavelet band-pass filter, so the duration of a cycle corresponding to 528 the inferior limit of the frequency spectrum (i.e. the longest period of the band) is shorter than the 529 time delay between the two signals. Thus, the large anticipation in the ascending phase evident from 530 the time courses (see in the results the delays between the peak days) cannot be captured by 531 WCS/MSWC. A future perspective would be to obtain and analyze a more complete set of data

532 over the time scale: as already explained, with a greater number of samples, indeed, the minimum 533 spectrum frequency gets lower, the corresponding duration of a cycle (i.e. the maximum period) 534 increases and consequently wavelet analysis through CWT could be able to detect larger time 535 delays between the signals. Moreover, even if aggregated regional data allow us to capture the 536 "sum" of the effects of different local situations with probable a greater anticipation capability, if 537 we consider the enormous geographical heterogeneity of CoViD-19 spread, regional data possess a 538 limited usefulness for public health monitoring and preparedness with respect to a possible second 539 wave of the epidemic. One more future direction would be to analyze these data at a more local 540 level, such as for each SOREU and each province (NUTS-3 level), or even municipalities, of 541 Lombardy region. To supply the lack of generality of wavelet analysis, we performed also a time 542 lag analysis through the cross-correlation function. Once again, the limited number of data affected our results: indeed, while the peaks of the emergency calls anticipated that of daily new cases by 543 544 about two weeks (Figs 2-3), the delay of the cross-correlation function peak is much lower both for 545 NUE calls (-3 days and -4 days, respectively for the original data and the modelled time series, see 546 Figs 10 and 13) and for SOREU calls (-5 days and -6 days, respectively for the original data and the 547 surrogate time series, see Figs 11 and 14). This can be explained just looking at the time courses 548 (Figs 2-3): the availability of emergency calls data is such that we can observe a great part of the 549 descending phase, whereas for the infected curve an initial decrease is visible only during the last 550 days of the corresponding observation period. With this perspective, consequently, it would be 551 useful to analyze a more complete dataset. Several other indicators are currently under investigation 552 and many of them will provide useful information, but we should not only rely on indicators 553 focused on detecting an increase in new cases, because the main impact on the health system is 554 more related to the characteristics of the infected population rather than to the number of infected 555 people.

The current results are referred to the early pandemic outbreak, and the same anticipation for the new cases is not necessarily expected for subsequent outbreaks since the improved capability of

558 the healthcare systems in surveillance and early diagnosis. However, the proposed approach should 559 be useful in forthcoming post-CoViD-19 era as a robust monitoring and convenient tool according 560 to the needs of the early detection of new viral outbreaks. Actually, this approach has the additional 561 advantage of a higher stability of the signal related to calls and tweets in a short term time interval: 562 indeed, the patterns of new positive cases are extremely variable across the days of the week 563 because of the notification delay in the reporting. On the contrary, the signal coming from the 564 emergency calls and tweets is in real time since it is not filtered by the administrative steps and the 565 capability of the diagnostic infrastructure.

566 The severe countermeasures put in place, such as the national lockdown, had a deep impact 567 on the population from several points of view, not only on the health system. It is therefore 568 important to take into account the social reaction to the crisis and analyzing it is part of the public 569 health response. Our analysis shows that Twitter trends correlate more with social factors rather 570 than with the number of cases (Figs 8-9). This finding suggests that a thorough analysis of social 571 media would improve our understanding about what the most common worries, fears and feelings 572 of the population are, in order to address them through a public health strategy that should include a proper use of social media to inform the population. Among all the Twitter data, only the daily 573 574 number of new tweets reveals some anticipation capability with respect to the epidemic curve: 575 wavelet analysis, indeed, detects a trend at the lowest frequencies, and a phase-lagged anomaly in a 576 frequency range centred around 0.25 cycles/day that occurs just between the two peaks (Fig 4). This 577 finding is confirmed and consistent both with the 7-days distance of the two curve peaks and with 578 the cross-correlation analysis (maximum at -6 days and -3 days lags, respectively for the original 579 data and the modelled time series, with small confidence intervals, see Figs 12 and 15).

580

581

583 **References**

- 585 1. Cereda, D, Tirani M, Rovida F, Demicheli V, Ajelli M, Poletti P, et al. The early phase of the
- 586 COVID-19 outbreak in Lombardy, Italy. arXiv:2003.09320v1 [Preprint]. 2020 [submitted 2020
- 587 March 20; cited 2020 October 5]. Available from: https://arxiv.org/abs/2003.09320v1
- 2. Rivieccio BA, Luconi E, Boracchi P, Pariani E, Romanò L, Salini S, et al. Heterogeneity of
 COVID-19 outbreak in Italy. Acta Biomed. 2020;91:31-4.
- 590 3. Italian Department for Civil Defense [Internet]. CoViD-19 Italia. Monitoraggio situazione.
- 591 [cited 2020 October 5]. Available from: <u>https://github.com/pcm-dpc/COVID-19</u>. Italian.
- 4. WHO. Novel Coronavirus (2019-nCoV) situation report 13. 2020 Feb 2. Available from:
- 593 <u>https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-</u>
- 594 <u>ncov-v3.pdf</u>
- 595 5. Odlum M, Yoon S. What can we learn about the Ebola outbreak from tweets? Am J Infect
 596 Control. 2015;43(6):563 71.
- 597 6. Castaldi S, Maffeo M, Rivieccio BA, Zignani M, Manzi G, Nicolussi F, et al. Monitoring
- emergency calls and social networks for COVID-19 surveillance. To learn for the future: the
- 599 outbreak experience of the Lombardia region in Italy. Acta Biomed. 2020;91:29-33.
- 600 7. vita.it [Internet]. Coronavirus, numeri di emergenza presi d'assalto. [posted 2020 February 24;
- 601 cited 2020 October 5]. Available from: <u>http://www.vita.it/it/article/2020/02/24/coronavirus-</u>
- 602 <u>numeri-di-emergenza-presi-dassalto/154125/</u>. Italian.
- 8. EENA. EENA recommendations for Emergency Services Organisations during the COVID-19
- 604 outbreak. Available from: <u>https://eena.org/knowledge-hub/documents/eena-recommendations-</u>
- 605 for-emergency-services-organisations-during-the-covid-19-outbreak/

- 606 9. EENA. Appendix. Data and strategies per country on emergency calls & public warning during
- 607 COVID-19 outbreak. Available from: <u>https://eena.org/knowledge-hub/documents/data-and-</u>
- 608 <u>strategies-per-country-on-emergency-calls-public-warning-during-covid-19</u>
- 10. Chen E, Lerman K, Ferrara E. Tracking social media discourse about the COVID-19 pandemic:
- development of a public coronavirus Twitter data set. JMIR Public Health Surveill [Internet].
- 611 2020 Apr-Jun [cited 2020 October 5];6(2):e19723. Available from:
- 612 <u>https://publichealth.jmir.org/2020/2/e19273/</u>
- 613 11. Sooknanan J, Comissiong DMG. Trending on social media: integrating social media into
 614 infectious disease dynamics. Bull Math Biol. 2020;82:86.
- 12. twitter.com [Internet]. The COVID tracking project. [cited 2020 October 5]. Available from:
- 616 <u>https://twitter.com/covid19tracking</u>
- 617 13. Gupta A, Katarya R. Social media based surveillance systems for healthcare using machine
 618 learning: a systematic review. J Biomed Inform. 2020;108:103500.
- 619 14. Castaldi S, Romanò L, Pariani E, Garbelli C, Biganzoli E. COVID-19: the end of lockdown
- 620 what next? Acta Biomed. 2020;91:236-8.
- 15. ilsole24ore.com [Internet]. Chiamate al 118, in Lombardia oltre il 30% per motivi respiratori e
- 622 infettivi. Picco il 16 marzo. [posted 2020 April 7; cited 2020 October 5]. Available from:
- 623 <u>https://www.ilsole24ore.com/art/chiamate-118-lombardy-oltre-30percento-motivi-respiratori-e-</u>
- 624 <u>infettivi-picco-16-marzo-ADEYHnI?refresh_ce=1</u>. Italian.
- 16. ilsole24ore.com [Internet]. In Lombardia calano le telefonate al 112, il picco il 12 marzo. Resta
- elevato il rapporto con i ricoveri. [posted 2020 April 4; cited 2020 October 5]. Available from:
- 627 <u>https://www.ilsole24ore.com/art/in-lombardy-calano-telefonate-112-picco-13-marzo-resta-</u>
- 628 <u>elevato-rapporto-i-ricoveri-ADp1r3H</u>. Italian.
- 629 17. twitter.com [Internet]. [cited 2020 October 5]. Available from:
- 630 <u>https://twitter.com/search?q=(118%20OR%20112)%20lang%3Ait%20until%3A2020-06-</u>
- 631 <u>29%20since%3A2020-02-18&src=typed query</u>

- 632 18. Bali R, Sarkar D, Lantz B, Lesmeister C. R: unleash machine learning techniques. Birmingham
- 633 (UK): Packt Publishing Ltd; 2016.
- 634 19. ExplosionAI GmbH. Industrial-strength natural language processing in Python: spaCy. Version
- 635 2.3.0 [software]. 2020 Jun 16 [cited 2021 Jan 30]. Available from: <u>https://spacy.io</u>
- 636 20. Daubechies I. The wavelet transform time-frequency localization and signal analysis. IEEE
- 637 Trans Inform Theory. 1990;36:961-1004.
- 638 21. Stoica P, Moses R. Spectral analysis of signals. Upper Saddle River (NJ): Prentice Hall; 2005.
- 639 22. Tian F, Tarumi T, Liu H, Zhang R, Chalak L. Wavelet coherence analysis of dynamic cerebral
- autoregulation in neonatal hypoxic-ischemic encephalopathy. Neuroim Clin. 2016;11:124-32.
- 641 23. Torrence C, Compo GP. A practical guide to wavelet analysis. Bull Am Metereol Soc.
 642 1998;79:61-78.
- 643 24. Ogden T. Essential wavelets for statistical applications and data analysis. Basel (CH):
 644 Birkha user; 1997.
- 645 25. Percival DB, Walden AT. Wavelet methods for time series analysis. Cambridge (UK):
 646 Cambridge University Press; 2000.
- 647 26. Boashash B. Time-frequency signal analysis and processing. A comprehensive reference. 2nd
 648 ed. Cambridge (MA): Academic Press; 2015.
- 649 27. Lilly JM. Element analysis: a wavelet-based method for analysing time-localised events in noisy
- time series. Proc Math Phys Eng Sci [Internet]. 2017 Apr [cited 2021 Jan
- 651 30];473(2200):20160776. Available from:
- 652 <u>https://royalsocietypublishing.org/doi/10.1098/rspa.2016.0776?url_ver=Z39.88-</u>
- 653 2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%200pubmed
- 28. Kiymik MK, Güler I, Dizibüyük A, Akin M. Comparison of STFT and wavelet transform
- 655 methods in determining epileptic seizure activity in EEG signals for real-time application.
- 656 Comput Biol Med. 2005;35(7):603-16.

- 657 29. Ryan ED, Cramer JT, Egan AD, Hartman MJ, Herda TJ. Time and frequency domain responses
 658 of the mechanomyogram and electromyogram during isometric ramp contractions: a comparison
 659 of the short-time Fourier and continuous wavelet transforms. J Electromyogr Kinesiol.
 660 2008;18(1):54-67.
- 30. Puliafito V, Vergura S, Carpentieri M. Fourier, wavelet and Hilbert-Huang transforms for
 studying electrical users in the time and frequency domain. Energies. 2017;10(2):188-201.
- 31. Meyers SD, Kelly BG, O'Brien JJ. An introduction to wavelet analysis in oceanography and
 meteorology: with application to the dispersion of Yanai waves. Monthly Weather Rev.
 1993;121:2858-66.
- 32. Fisher RA. On the 'probable error' of a coefficient of correlation deduced from a smallsample. Metron. 1921;1:3-32.
- 33. Ebisuzaki W. A method to estimate the statistical significance of a correlation when the data are
 serially correlated. J Climate. 1997;10(9):2147-53.
- 670 34. euronews.com [Internet]. Coronavirus: more than a third of people in Italy's COVID-19
- epicentre estimated to have had disease. [posted 2020 April 28; cited 2020 October 5].
- 672 Available from: <u>https://www.euronews.com/2020/04/28/coronavirus-more-than-a-third-of-</u>
- 673 people-in-italy-s-covid-19-epicentre-estimated-to-have-ha
- 35. Disposizioni attuative del decreto-legge 23 febbraio 2020, n. 6, recante misure urgenti in
- 675 materia di contenimento e gestione dell'emergenza epidemiologica da COVID-19, Italian Prime
- 676 Minister Decree, Gazzetta Ufficiale Serie Generale n. 45 (Feb 23, 2020). Italian.
- 677 36. Misure straordinarie ed urgenti per contrastare l'emergenza epidemiologica da COVID-19 e
- 678 contenere gli effetti negativi sullo svolgimento dell'attività giudiziaria. Italian Law Decree n.
- 679 11, Gazzetta Ufficiale Serie Generale n. 60 (Mar 8, 2020). Italian.
- 680 37. Costituzione dell'Unità di Crisi dell'emergenza sanitaria in ordine all'emergenza
- epidemiologica da COVID-19 e relativa Task Force. Lombardy Region Secretariat Decree n.

682	3287	(Mar	12,	2020).	Available	from:	https://www.openpolis.it/wp-
683	content	/uploads/2	020/05/E	DECRETO-3	3287-DEL-12N	MARZO-2	2020.pdf. Italian.

- 684 38. Tô TD, Protin F, Nguyen TTH, Martel J, Nguyen DT, Charles P, et al. Epidemic dynamics via
- wavelet theory and machine learning with applications to Covid-19. Biology. 2020;9(12): 477.
- 686 39. Dziedziech K, Nowak A, Hasse A, Uhl T, Staszewski WJ. Wavelet-based analysis of time-
- variant adaptive structures. Philos Trans A Math Phys Eng Sci. 2018;13:376(2126):20170245.
- 40. Grinsted A, Moore JC, Jevrejeva S. Application of the cross-wavelet transform and wavelet
- coherence to geophysical time series. Nonlin Proc Geophys. 2004;11:561-6.
- 690 41. Maraun, Kurths DJ, Holschneider M. Nonstationary Gaussian processes in wavelet domain:
- 691 synthesis, estimation and significance testing. Phys Rev E. 2007;75(1 Pt 2):016707.
- 42. Torrence C, Webster P. Interdecadal changes in the ESNO-Monsoon System. J Clim.
 1999;12:2679-90.

































NUE daily regional incoming calls vs. daily infected Cross-Correlation Sequence Estimate Fisher's Z-transformation 90% Confidence Bounds





Daily new tweets vs. daily infected Cross-Correlation Sequence Estimate Fisher's Z-transformation 90% Confidence Bounds



NUE daily regional incoming calls vs. daily infected Monte Carlo Cross-Correlation Sequence Estimate , n = 1000 simulations - 90% Confidence Bounds





-20 -10 0 10 20 Time Lag (days) Daily new tweets vs. daily infected Monte Carlo Cross-Correlation Sequence Estimate n = 1000 simulations - 90% Confidence Bounds

