Could Emergency Calls and Twitter Activity Help to Prevent Health System Overloads Due to CoViD-19 Epidemic? Wavelets and Cross-Correlation as Useful Tools for Time-Frequency Signal Analysis: Lessons from the Italian Lombardy Region

Bruno Alessandro Rivieccio[1], Alessandra Micheletti[2], Manuel Maffeo[3], Matteo Zignani[4], Alessandro Comunian[5], Federica Nicolussi[6], Silvia Salini[6], Giancarlo Manzi[6], Francesco Auxilia[3], Mauro Giudici[5], Giovanni Naldi[2], Sabrina Gaito[4], Silvana Castaldi[3] and Elia Biganzoli[7]

[1]Department of Laboratory Medicine, Division of Anatomic Pathology, Niguarda Hospital, Milan, Italy – brunoalessandro.rivieccio@ospedaleniguarda.it
[2]Department of Environmental Science and Policy, University of Milan, Milan, Italy – alessandra.micheletti@unimi.it, giovanni.naldi@unimi.it
[3]Department of Biomedical Sciences for Health, University of Milan, Milan, Italy – manuel.maffeo@unimi.it, francesco.auxilia@unimi.it, silvana.castaldi@unimi.it
[4]Department of Computer Science, University of Milan, Milan, Italy – matteo.zignani@unimi.it, sabrina.gaito@unimi.it
[5]Department of Earth Sciences, University of Milan, Milan, Italy – alessandro.comunian@unimi.it, mauro.giudici@unimi.it
[6]Department of Economics, Management and Quantitative Methods, University of Milan, Milan, Italy – federica.nicolussi@unimi.it, silvia.salini@unimi.it, giancarlo.manzi@unimi.it
[7]Department of Clinical Sciences and Community Health – elia.biganzoli@unimi.it

The history of the Italian CoViD-19 epidemic began on 2020, February the 20th, in Lombardy region, which quickly became the most stricken geographical area of the world. This first outbreak caught national health system unprepared, and hospitals experienced patients overload, facing an unknown infectious disease. Thus, it is of primary importance to provide public health services with tools which can help to potentially prevent health system stress periods. To this aim, we performed a time-frequency analysis of regional emergency calls, CoViD-19-related Twitter data and daily new cases through wavelets, and a comparison of the signals in the time domain using cross-correlation. Our findings show that emergency calls could be a good predictor of health service burdens, while Twitter activity is more related to personal and emotional involvement in the emergency and to socio-political dynamics. Social media should therefore be used to improve institutional communication in order to prevent "infodemia".

DOI: 10.54103/milanoup.73.65

1 Introduction

After the first Italian case of CoViD-19 was identified in Lombardy region, on 2020, February the 20th,[1] several countermeasures were put in place until the national lockdown.[2,3] The outbreak displayed a huge geographical heterogeneity,[4] and Lombardy was the worst hit region of the world.[5–7]

Many factors can explain this disaster and hospitals overloading. Among these, unpreparedness[8,9] and the consequences of the fragmentation of the National Health Service into different regional systems[10] had a prominent role. Indeed, both national and regional pandemic plans were obsolete,[11–13] and in Lombardy a local reformation dismantled public territorial medicine services in favor of a hospital-centered system and of private healthcare facilities.[9,14]

Moreover, the debate on CoViD-19 was not limited within the scientific community, having being conducted even on mass media and social networks. Thus, an “infodemia"[15] led to different psycho-emotional reactions in common people.

Therefore: (i) regional emergency numbers were overwhelmed by calls from people searching for information and reassurance and/or needing medical assistance;[16,17] (ii) the Emergency Departments became the first line against the virus, despite the opposite regional pandemic plan recommendation;[12] (iii) social media were “overcrowded" by messages from people sharing “thoughts" about the pandemic.[16]

Given these features, emergency calls could be a potential predictor of new health system overloads, and social networks might represent a tool to monitor feelings about CoViD-19 spread and management. Social media analysis has already been used in digital epidemiology,[18,19] and in the case of CoViD-19 it has been applied even for contact tracing and outbreak control.[20] Particularly, Twitter high-speed communication capability allows super-fast spread of information, and its “follow model" can be described as an interest graph, differently from Facebook.[21] Thus, as a very dynamic platform, Twitter is the ideal probe to study CoViD-19 social context.

The aim of this work is to investigate the anticipation capability of daily emergency calls and Twitter trends with respect to CoViD-19 spread dynamics. It is of crucial importance to provide health services with useful tools for surveillance, monitoring, control and prevention of new burdens.

2 Materials and Methods

Data analysis was conducted using MatLab R2020a.

2.1 Data

The following daily time series were considered:

- calls to unique emergency number 112 from February 18, 2020 to March 30, 2020;[22]
- calls to medical emergency number 118 from February 18, 2020 to March 30, 2020;[23]
- calls to regional CoViD-19 toll-free number from February 23, 2020 to March 30, 2020;[23]
- Twitter data (tweets, replies, likes, retweets) from February 18, 2020 to June 29, 2020;[24]
- CoViD-19 new cases from February 24, 2020 to June 29, 2020.[25]

All data were regionally aggregated, except Twitter ones which were not geolocalised. The first days of 112 and toll-free number series were discarded as outliers.

2.2 Twitter Data Analysis

We submitted a query to the Twitter Search Application Programming Interface in order to select all the tweets in the period of interest containing the term "112" or "118" or both ("search terms"), discarding retweets.[24] Then, the text of each tweet was lemmatised and tokenised through spaCy:[26] only nouns and verbs were considered. After that, we manually identified 283 common words related to the emergency ("keywords"). The co-presence of one (or both) of the search terms with one (or more) of the keywords led us to the selection of 16,216 tweets. In this way, we discarded the tweets in which the search terms were used with a meaning not related to the pandemic.

2.3 Wavelet Analysis

For non-stationary signals, wavelets represent the most suitable tool to perform a time-frequency analysis. In fact wavelet transform displays a variable time-frequency resolution and is able to highlight the changes over time (i.e. the non-stationarity) of the frequencies contributions.[27–35]

Wavelets can detect in the time-frequency (scale) plane both long-period backgrounds (trends) and short-period discontinuities (anomalies). So, anomalies represent high-frequency hidden signals and - despite their limited temporal location - possess a huge amount of information content.[36,37]

Wavelet analysis is therefore useful to "capture" stress moments in health systems during emergencies, as in the case of CoViD-19 pandemic. Namely, in the early epidemic outbreak, peaks in the emergency calls should possibly anticipate those for new infections and hospitalizations.

All the time series were first pre-processed to identify trends. Since all the signals displayed a typical weekly "seasonality", a moving average linear filter of a 7-days amplitude was applied to smooth them. The resulting data sequences were decomposed through the continuous wavelet transform (CWT) and then normalised to their maximum values. Tweets-dependent data (replies, likes, retweets) were also normalised to the corresponding number of tweets.

To measure the similarity between signals, we computed the wavelet cross-spectrum (WCS) and the magnitude-squared wavelet coherence (MSWC) in the time-frequency plane, resulting in a value of coherence for each point, i.e. for each pair of time-frequency coordinates. To discard edge-effects, coherence values outside the boundaries of the cone of influence were not considered. Moreover, since we were interested in quantifying the time delay between signals, we used a complex analytic Morlet (Gabor) wavelet through which the phase lag was computed. Finally, we converted phase data into time information.

2.4 Cross-Correlation Analysis

While coherence measures signals similarity in the frequency domain, cross-correlation is its counterpart in the time domain. The lag corresponding to the maximum value of the cross-correlation sequence represents a synthetic estimate of the time delay between two discrete sequences. Consequently, to quantify the time anticipation of the leading signal with respect to the lagged one, we also computed the cross-correlation function.

A 90% confidence interval was calculated for the lag corresponding to the cross-correlation function peak, through: (i) a z-Fisher statistics on the original data;[38] (ii) a random phase test for 1,000 simulations.[39]

3 Results

3.1 Twitter Trends

Figure 1 shows Twitter raw data. The first increase, shared by all the time series, occurs from 2020/02/21 to 2020/02/25 (days -3 to 1), and can be ascribed to the detection of the CoViD-19 first case and the subsequent establishment of the 'red area' in the involved municipalities. A second sharp peak in common is clearly visible on 2020/03/14 (day 19) and is due to the death of a healthcare worker in Bergamo, the most stricken province. Therefore, Twitter activity seems mostly triggered by political decisions and chronicle news, and only partly by the epidemic dynamics.

3.2 Time Course Analysis

Table 1 shows the signals time-to-peak differences. All the emergency calls series share a relevant time anticipation (almost two weeks) and thus can be further considered as potential predictors. Regarding to Twitter data, instead, only tweets display a 7-days anticipation, while the remaining tweets-dependent series lose this capability, and replies are even lagged with respect to new cases.

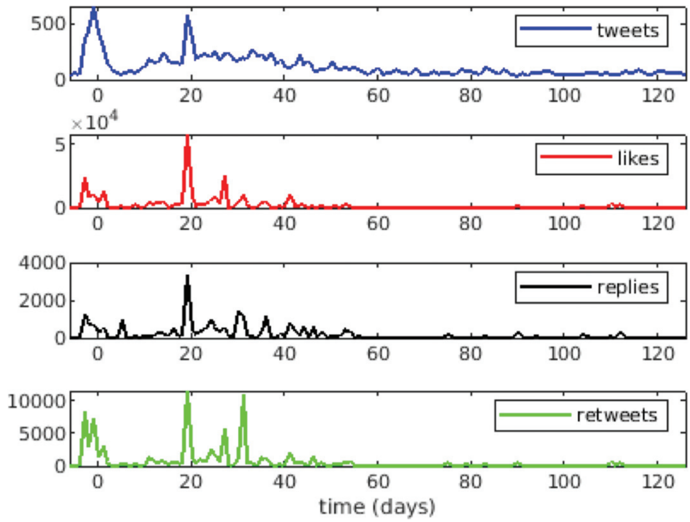

Figure 1: Twitter trends. Conventionally, "day 0" corresponds to the date of February 24, 2020, when the Italian Department for Civil Defense released the first official epidemic report.

3.3 Wavelet Analysis

The WCS and the MSWC were calculated for each time series in relation to new cases data. The focus will be for coherence values higher than 0.5. Besides, we will consider only high frequency discontinuities (the most informative variations from the trend). Finally, we will take into account only the ascending phase due to its importance from a public health viewpoint: consequently, we will exclude the replies whose peak is delayed compared to that one of the new cases.

Calls to toll-free number, likes and retweets do not reveal any strong coherence with respect to the infected signal. Thus, we will consider only calls to 112, calls to 118 and tweets data, and the remaining time series will not be further discussed.

Table 1: Time-to-peak analysis. Differences are smoothed and normalised values.

x	TTP_x (days)	TTP_i (days)	ΔTTP (days)
Calls to 112	16		-13
Calls to 118	15		-14
Calls to 800.89.45.45	15		-14
Tweets	22	29	-7
Likes	28		-1
Replies	33		+4
Retweets	28		-1

Notes: 800.89.45.45 is the regional CoViD-19 toll-free number. x = potential predictor; TTP_x = time-to-peak of the potential predictor; TTP_i = time-to-peak of the infected time series; $\Delta TTP = TTP_x - TTP_i$ = time-to-peak difference.

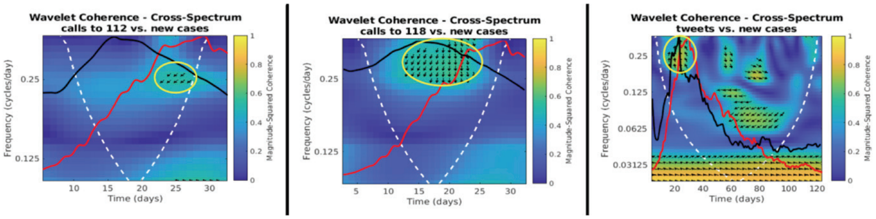

Figure 2: Wavelet analysis. WCS/MSWC representations are superimposed to the corresponding time plots (smoothed and normalised data): on the left, calls to 112 vs. new cases; in the middle, calls to 118 vs. new cases; on the right, tweets vs. new cases. The infected time course is depicted in red, the potential predictor in black. The yellow circles highlight the discontinuities between the peaks.

In Figure 2 the time plots of the smoothed and normalised series are superimposed to the WCS/MSWC charts. The cone of influence is shown as a white dashed line. For areas where the coherence exceeds 0.5, the charts display arrows representing the phase lag between signals. The direction of the arrows designates the relative phase on the unit circle (rightward meaning in-phase, leftward indicating anti-phase). The corresponding lag in time depends on the duration of the cycle (period).

In all the cases the anomalies are time-confined between the peaks of the curves, and a frequency component of about 0.25 cycles/day (i.e. a period duration of 4 days) is shared. These high-coherence discontinuities display different phase lags, and thus time delays, for the three pairs of signals:

- 2.5-2.6 days for calls to 112 (Fig 2, left panel);
- 2.4 days for calls to 118 (Fig 2, middle panel);
- 0.5-1.4 days for tweets (Fig 2, right panel).

3.4 Cross-Correlation Analysis

In Table 2 for each pair of signals the time delay with the corresponding 90% confidence interval is reported. The computation was done in two ways, to compare parametric and non-parametric estimates. Negative values denote by how many days the time series of infected patients should be shifted backward along time axis to be ’aligned’ with the potential predictor.

4 Discussion

Indicators that can anticipate a rise of cases are of paramount importance to plan interventions of the health service organization. Our study shows that the number of calls to emergency services could be a good indicator that can precede

Table 2: Cross-correlation analysis. For each pair of signals the time delay with the corresponding 90% confidence bounds is indicated. The confidence interval has been constructed via both the Fisher's z-transformation (left column) and a random phase test (right column).

	Time lag of the cross-correlation function peak (days) 90% confidence intervals	
Time series	**Fisher's z transformation**	**Monte Carlo simulation**
Calls to 112	-3 [-8, 1]	-4 [-11, 1]
Calls to 118	-5 [-11, 1]	-6 [-13, 1]
Tweets	-6 [-8, -2]	-3 [-11, 4]

the need for hospitalization in the early pandemic outbreak. The current results are not necessarily expected to be found for subsequent burdens.

Our step-by-step analysis made us discard four time series as potential predictors. In the end, only calls to 112, calls to 118 and tweets were analysed in the time domain.

To the best of our knowledge, this is the first time that wavelet analysis and cross-correlation are used to detect health system stressful periods during the CoViD-19 pandemic. However, a first limitation of our study is the short period of availability of emergency calls data. In fact, due to mathematical reasons, this deficiency "shaped": (i) the time-frequency resolution of wavelet analysis; (ii) the location of the cross-correlation function peak. A future perspective would be to obtain a more complete dataset. Moreover, even if aggregated regional data probably possess a greater anticipation capability, taking into account the geographical heterogeneity of CoViD-19 spread, there is a limited usefulness for public health monitoring. One more direction would be to analyse these data at a more local level. Other indicators are under investigation, keeping in mind also that the impact on the health system is more related to the patients' characteristics rather than to the number of new cases.

Finally, our analysis shows that Twitter trends correlate more with social factors rather than with the number of cases. This finding suggests that social media analysis could improve and address public health strategies and institutional communication to the population.

References

[1] Cereda, D., Tirani, M., Rovida, F., Demicheli, V., Ajelli, M., Poletti, P., Trentini, F., Guzzetta, G., Marziano, V., Barone, A. et al. (2020). The early phase of the COVID-19 outbreak in Lombardy, Italy. arXiv:2003.09320v1 [Preprint].

[2] Italian Prime Minister Decree, Gazzetta Ufficiale Serie Generale n. 45 (2020). Disposizioni attuative del decreto-legge 23 febbraio 2020, n. 6, recante misure urgenti in materia di contenimento e gestione dell'emergenza epidemiologica da COVID-19. [mplementing provisions of the decree-law 23 February 2020, n. 6, containing urgent measures regarding

the containment and management of the epidemiological emergency from COVID-19] (In Italian).

[3] Italian Law Decree n. 11, Gazzetta Ufficiale Serie Generale n. 60 (2020). Misure straordinarie ed urgenti per contrastare l'emergenza epidemiologica da COVID-19 e contenere gli effetti negativi sullo svolgimento dell'attivitá giudiziaria. [Extraordinary and urgent measures to counter the epidemiological emergency from COVID-19 and contain the negative effects on the conduct of judicial activity] (In Italian).

[4] Rivieccio, B.A., Luconi, E., Boracchi, P., Pariani, E., Romanó, L., Salini, S., Castaldi, S., Biganzoli, E., Galli, M. (2020). Heterogeneity of COVID-19 outbreak in Italy. *Acta Biomedica*, 91: 31–34.

[5] Agenzia Giornalistica Italiana (2020). Quali sono le regioni piú colpite dall'epidemia nel mondo?. [What are the regions most affected by the epidemic in the world?] (In Italian). https://www.agi.it/cronaca/news/2020-04-10/coronavirus-regioni-piu-colpite-lombardia-new-york-madrid-8301518/. Accessed: February 6, 2021.

[6] Los Angeles Times (2020). Perfect storm: virus disaster in Italy's Lombardy region is a lesson for the world. https://www.latimes.com/world-nation/story/2020-04-26/perfect-storm-virus-disaster-in-italys-lombardy-region-is-a-lesson-for-the-world. Accessed: February 6, 2021.

[7] Euronews (2020). Coronavirus: more than a third of people in Italy's COVID-19 epicentre estimated to have had disease. https://www.euronews.com/2020/04/28/coronavirus-more-than-a-third-of-people-in-italy-s-covid-19-epicentre-estimated-to-have-ha. Accessed: February 6, 2021.

[8] Nuclear Threat Initiative, Johns Hopkins University Center for Health Security, The Economist Intelligence Unit (2019). Global Health Security index - 2019 report.

[9] Castaldi, S., Luconi, E., Rivieccio, B.A., Boracchi, P., Marano, G., Pariani, E., Romano, L., Auxilia, F., Nicolussi, F., Micheletti, A. et al (2021). Are health indicators able to describe the ability to cope with COVID-19 epidemic? *Risk Management and Healthcare Policy*, 14: 2221–2229.

[10] Italian Constitutional Law n. 3, Gazzetta Ufficiale Serie Generale n. 248 (2001). Modifiche al titolo V della parte seconda della Costituzione. [Amendments to Title V of Part Two of the Constitution] (In Italian).

[11] Open Online (2020). Coronavirus. Il servizio di Report e il pdf del piano pandemico mai aggiornato dal 2006. [Coronavirus. The Report service and the pdf of the pandemic plan, never updated since 2006]. (In Italian). https://www.open.online/2020/12/01/coronavirus-servizio-report-pdf-piano-pandemico-mai-aggiornato-2006/. Accessed: February 6, 2021.

[12] Italian Ministry of Health (2006). National plan for preparedness and response to an influenza pandemic. http://www.salute.gov.it/imgs/C_17_pubblicazioni_501_ulterioriallegati_ulterioreallegato_0_alleg.pdf.

[13] Lombardy Region (2009). Piano pandemico regionale – Aggiornamento a seguito del manifestarsi della pandemia influenzale A/H1N1v. [Regional pandemic plan - Update following the outbreak of the A / H1N1v influenza pandemic]. Lombardy Regional Circular n. 19/SAN/2009 of the General Directorate for Health, Bollettino Ufficiale Regione Lombardia Serie Ordinaria n. 39. (In Italian).

[14] Lombardy Regional Law n. 33 (2009). Testo unico delle leggi regionali in materia di sanitá. [Consolidated text of regional health laws] Bollettino Ufficiale Regione Lombardia Serie Ordinaria n. 52, 3° Supplemento Ordinario. (In Italian).

[15] WHO (2020). Novel Coronavirus (2019-nCoV) situation report - 13. 2020 Feb 2. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf.

[16] Castaldi, S., Maffeo, M., Rivieccio, B.A., Zignani, M., Manzi, G., Nicolussi, F., Salini, S., Micheletti, A., Gaito, S., Biganzoli, E. (2020). Monitoring emergency calls and social networks for COVID-19 surveillance. To learn for the future: the outbreak experience of the Lombardia region in Italy. *Acta Biomedica*, 91: 29–33.

[17] vita.it. Coronavirus, numeri di emergenza presi d'assalto. [Coronavirus, emergency numbers overwhelmed] http://www.vita.it/it/article/2020/02/24/coronavirus-numeri-di-emergenza-presi-dassalto/154125. (In Italian). Accessed February 7, 2021.

[18] Marquetoux, N., Stevenson, M.A., Wilson, P., Ridler, A., Heuer, C. (2016). Using social network analysis to inform disease control interventions. *Preventive Veterinary Medicine*, 126: 94–104.

[19] Silk, M.J., Croft, D.P., Delahay, R.J., Hodgson, D.J., Boots, M., Weber, N., et al. (2017). Using social network measures in wildlife disease ecology, epidemiology, and management. *BioScience*, 67(3): 245–57.

[20] Saraswathi, S., Mukhopadhyay, A., Shah, H., Ranganath, T.S. (2020). Social network analysis of COVID-19 transmission in Karnataka, India. *Epidemiology & Infections*, 148: e230.

[21] Bali, R., Sarkar, D., Lantz, B., Lesmeister, C.R. (2016). *Unleash machine learning techniques*. Birmingham (UK): Packt Publishing Ltd; 2016.

[22] Il Sole 24Ore (2020). Chiamate al 118, in Lombardia oltre il 30% per motivi respiratori e infettivi. Picco il 16 marzo. [Calls to 118, in Lombardy more than 30% for infective and respiratory reasons. Peak on March 16]. https://www.ilsole24ore.com/art/chiamate-118-lombardy-oltre-30percento-motivi-respiratori-e-infettivi-picco-16-marzo-ADEYHnI?refresh_ce=1. (In Italian). Accessed: February 7, 2021.

[23] Il Sole 24Ore (2020). In Lombardia calano le telefonate al 112, il picco il 12 marzo. Resta elevato il rapporto con i ricoveri. [In Lombardy 112 calls decrease, peak on March 12. Calls- admissions ratio still high]. https://www.ilsole24ore.com/art/in-lombardy-calano-telefonate-112-picco-13-marzo-resta-elevato-rapporto-i-ricoveri-ADp1r3H. (In Italian). Accessed: February 7, 2021.

[24] twitter.com [(2020). https://twitter.com/search?q=(118%20OR%20112)%20lang%3Ait%20until%3A2020-06-29%20since%3A2020-02-18&src=typed_query. (In Italian). Accessed: February 7, 2021.

[25] Italian Department for Civil Defense (2020). CoViD-19 Italia. Monitoraggio situazione. https://github.com/pcm-dpc/COVID-19. (In Italian). Accessed: February 7, 2021.

[26] ExplosionAI GmbH (2020). Industrial-strength natural language processing in Python: spaCy. Version 2.3.0 [software]. https://spacy.io. Accessed: February 7, 2021.

[27] Boashash, B. (2015).*Time-frequency signal analysis and processing. A comprehensive reference. 2nd ed.* Cambridge (MA): Academic Press.

[28] Kiymik, M.K., Güler, I., Dizibüyük, A., Akin, M. (2005). Comparison of STFT and wavelet transform methods in determining epileptic seizure activity in EEG signals for real-time application. *Computers in Biology and Medicine*, 35(7): 603–16.

[29] Ryan, E.D., Cramer, J.T., Egan, A.D., Hartman, M.J., Herda, T.J. (2008). Time and frequency domain responses of the mechanomyogram and electromyogram during isometric

ramp contractions: a comparison of the short-time Fourier and continuous wavelet transforms. *Journal of Electromyography and Kinesiology*, 18(1): 54–67.

[30] Puliafito, V., Vergura, S., Carpentieri, M. (2017). Fourier, wavelet and Hilbert-Huang transforms for studying electrical users in the time and frequency domain. *Energies*, 10(2): 188–201.

[31] Daubechies, I. (1990). The wavelet transform time-frequency localization and signal analysis. *IEEE Transactions on Information Theory*, 36: 961–1004.

[32] Stoica, P., Moses, R. (2005). *Spectral analysis of signals*. Upper Saddle River (NJ): Prentice Hall.

[33] Tian, F., Tarumi, T., Liu, H., Zhang, R., Chalak, L. (2016). Wavelet coherence analysis of dynamic cerebral autoregulation in neonatal hypoxic-ischemic encephalopathy. *NeuroImage: Clinical*, 11: 124–32.

[34] Torrence, C., Compo, G.P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, 79: 61–78.

[35] Lilly, J.M. (2017). Element analysis: a wavelet-based method for analysing time-localised events in noisy time series. *Proceedings of the Royal Society A - Mathematics, Physics Engineering Sciences*, 473(2200): 20160776.

[36] Ogden, T. (1997). *Essential wavelets for statistical applications and data analysis*. Basel (CH): Birkhäuser; 1997.

[37] Percival, D.B., Walden, A.T. (2000). *Wavelet methods for time series analysis*. Cambridge (UK): Cambridge University Press.

[38] Fisher, R.A. (1921). On the 'probable error' of a coefficient of correlation deduced from a small sample. *Metron*. 1: 3–32.

[39] Ebisuzaki, W. (1997). A method to estimate the statistical significance of a correlation when the data are serially correlated. *Journal of Climate*, 10(9): 2147–2153.