Journal of Nonlinear and Convex Analysis
Volume 23, Number 10, 2022, 2383–2394

THE SYSTEMATIC TRACEABILITY AND ANALYSIS OF COVID-19

ZEKAI LU*, LINGXI PENG*, XUAN LIU, YINGTING ZHONG, RONG HU†, AND HAOHUAI LIU†

Abstract. The traceability of COVID-19 is of great importance for Epidemiological implications. Two robust and non-parametric traceability methods were employed in the study to evaluate COVID-19's traceability for the prevention of epidemics. We first corrected the initial dates for eight countries using Solow and smith's method, and then the OLE method was applied to the corrected dataset for traceability. Our model suggests that the first global case of COVID-19 originated on 25 September 2019 (95%CI 23 September) and that the first case emerged in Europe and spread rapidly to neighbouring countries, expanding globally in early January 2021. Our study suggests that the spread of the epidemic may be more rapid and earlier than we thought. The two methods we use can be used in a robust traceability approach for small sample estimates, providing additional explanations for epidemic traceability studies.

1. Introduction

COVID-19 is a highly contagious form of pneumonia caused by a coronavirus known as SARS-CoV-2 [14]. Tracing the origin of COVID-19, it can identify the source of transmission of COVID-19 , the initial transmission route, the mutation pattern and the potential risk of COVID-19. This will enable more targeted and effective preventive and control measures. In addition, only by identifying the source of COVID-19 can the mode of mutation of the virus can be studied and the resurgence of the epidemic can be stopped. The traceability and prediction of COVID-19 are among the hot spots of academic research.

Several studies have been carried out on the origin of COVID-19. These studies can be broadly divided into two categories, the first of which is traceability through the biological route [3]. The review by M.Morens et al. points out that the prevailing view in biology currently uses a genetic backtracking approach. They suggest that the SARS-CoV-2 virus evolved from a coronavirus in mammals like bats and pangolins; however, there is still much debate about where it arose [3, 14, 21]. Some argue that the virus could have been synthe-sized; for example, Law's research suggests that SARS-CoV-2 is highly unlikely to exist in nature as a highly complex virus

2020 *Mathematics Subject Classification.* 62M10, 62P10.

Key words and phrases. COVID-19, virus traceability, ecological approach, mathematical model.

*These authors contributed equally to this work and should be considered co-first authors.

†These authors contributed equally to this work and should be considered co-corresponding authors.

This work was supported by The National Social Science Fund of China with No.21BSH023, the Science and Technology projects from Guangzhou City with NO.202102010481, the University Scientific Research Project of Guangzhou Education Bureau with grant No. 202235165 and the National Natural Science Foundation of China No. 12171114.

and could be the product of human genetic engineering [18]. Notably, the study by Segreto et al. also supports this opinion [27]. Inconclusively, due to the biological complexity of the viruses present. It often takes years to make progress, as shown in the traceability research of SARS-CoV and HIV viruses [24, ?Hon2020Overview:].

Another type of traceability is through mathematical models, Cooper et al. used the SIR model to investigate the origin of COVID-19, and their study revealed that the outbreak developed much more rapidly than we thought [8]. Yang et al. investigated this issue using curve fitting methods, and their model showed that the United States was the likely origin of the pandemic [34]. There is also a study by Nsoesie et al. using a big data approach to internet activity in Wuhan, China, in late 2019, and their study side-by-side suggests that Wuhan may be the origin of the epidemic [22]. These studies generally suffer from the following problems. Firstly, the models have a wide range of errors, with some models estimating errors in years, which is not applicable to accurate traceability studies. Secondly, the models are too idealistic to simulate realistic and complex situations of all kinds, a problem that is prevalent in the current widely popular models of infectious diseases [32].In addition, the data sets used were not rigorous enough for science. For example, the study by Nsoesie et al. has been widely debated due to the one-sided and subjective nature of the data used [20] . Also, with the development of machine learning and blockchain technology, researchers have been able to use several new techniques to study the COVID-19 outbreak [23]. For example, Srivastava et al. have developed a series of medical support platforms [11, 19, 31, 35], which have given researchers new ideas for virus traceability.

The following improvements have been made to address these problems, which are common in mathematical models. Firstly, we use data from a more reliable and authoritative source (the WHO data) [33]. In addition, we imputed the data by combining the virus detection results reported by countries in 2019. These data were then tested for isolated cases using Solow and Smith's ecological approach, and the corrected dates were used as the dataset. Finally, another ecological model called OLE was used for traceability studies.

2. Dataset

We use global epidemic data published by the World Health Organization (WHO), which has a high degree of completeness and credibility as reported to the WHO by national/regional health statistics departments. The dataset has eight columns, reporting date, country, country code, region, new cases, new deaths and cumulative cases and deaths [33]. The period is from 22 January 2020 to 14 September 2021. The dataset contains data from 233 countries or territories, covering almost all countries and territories in the world. The dataset was first compiled on 22 January 2020, and the data was supplemented with data published by national health authorities and relevant studies for previous cases [2, 5, 9, 17].

The basis for the addition to the data set is as follows, according to the National Institute of Oncology in Milan, Italy, the first case in Italy was reported in the Veneto region precisely in September 2019 [2]. Based on this fact, we have added an example of data for Italy on 30 September 2019. On 10 November 2019, molecular traces of the SARS-CoV-2 virus were found in biopsy tissue collected from a

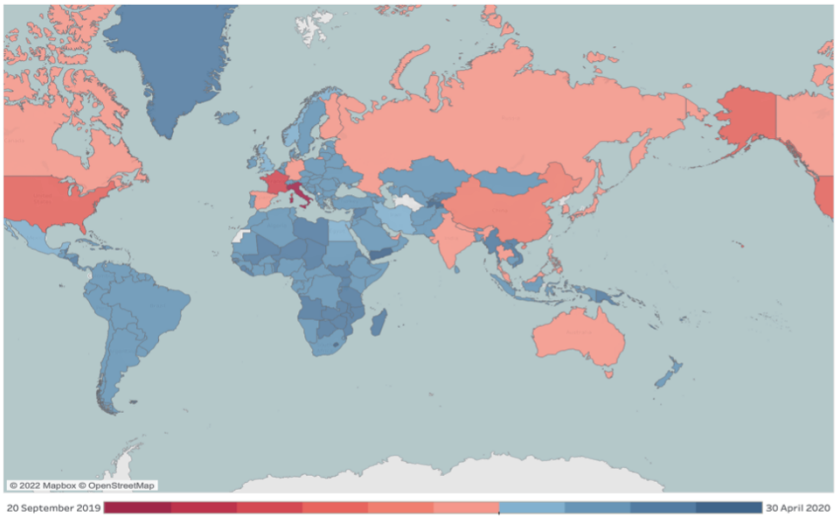

FIGURE 1. Date of the first case reported. The order of dates varies from blue to red, and this is the bluer the colour, the earlier the report, and vice versa.

reddened skin area of the arm of an Italian woman who tested positive for COVID-19 antibodies in blood tests after June 2020 [17]. According to this, another case data was added on 10 November 2019. A retrospective study of approximately 2,500 chest scans performed between 1 November 2019 and 30 April 2020 at Albert Schweitzer Hospital in France determined that the earliest case of novel coronavirus infection at the hospital dates back to 16 November 2019. Accordingly, a case was added in France on 16 November 2019. 106 of the archived samples from regular blood donations from 13 December 2019 to 17 January 2020 tested positive for antibodies to novel coronavirus in nine states in the United States, but the date of onset of the first officially confirmed case in the US state was 19 January 2020. Accordingly, novel coronaviruses were already present in the United States in December 2020 [4].

Furthermore, we therefore, added a case from the United States on 13 December 2019. On 27 December 2019, a pharyngeal swab sample from a patient with haemoptysis in France tested positive for novel coronavirus nucleic acid by RT-PCR [10]. Finally, we added a case to the French dataset on 27 December 2019.

In order to better represent the time of the first positive COVID-19 test in each country that is currently publicly available, we have plotted Figure 1 as a distribution of the time of the first case reported in each country.

3. Method and result

3.1. **Test for isolated cases.** Due to the existence of isolated cases, that is, cases that are isolated upon entry into customs, these cases do not cause mass transmission. So using isolated cases as the date of the first test report may bias our

estimates. It was, therefore, necessary to first test the first detection dates for each country to screen out countries where isolated cases existed. Moreover, to correct the dates.

The test proposed by Solow and Smith was therefore used to test whether cases were isolated [29,30]. Solow and Smith's test is a non-parametric method for testing whether there is specificity in the extremes of a sequence and was first applied in ecology to detect whether the reappearance of a species is a reintroduction [29, 30]. Because of its simplicity and robustness, the method has also been used in fields such as sports science to test whether a break in the sports record is worth being questioned . Briefly, the null hypothesis of the method (H_0) is that there is no specificity in the extreme values of the series, and the alternative hypothesis (H_1) is that there is some specificity in the extreme values of the series, that is that there is an abnormal deviation.

Using Solow and Smith's method, it is possible to test whether the date of the first report is an isolated case, giving a p-value for the confidence interval and thus allowing us to have further confidence that a case is isolated. In addition, Solow and Smith's method can also correct for the date of the first test report. We can correct for a reporting date of $p < 0.05$ so that the reporting date rejects the alternative hypothesis of the Solow and Smith method. This makes the OLE estimates more accurate than the uncorrected ones. It is also a non-parametric estimation method, which does not rely on a specific distribution. The estimates are highly robust.

To test whether the earliest reported date of detection in each country was an isolated case, we made the null hypothesis (H_0) that the virus was generated from within the local population, which indicates that the virus had begun to spread on a large scale in the local community. The alternative hypothesis (H_1) is that the virus was introduced in isolation, which indicates that the case was isolated at the time of introduction (e.g. it was isolated upon repatriation and did not contact the community).

Let $x_1 > x_2 > \cdots > x_k$ Be the k earliest cases of COVID-19, ordered from the most recent to the earliest. If the earlier cases of COVID-19 are recorded at time y, we are interested in whether the earliest reporting date is on the isolated case report. Solow and Smith's approach suggests that under the null hypothesis, the same process generates the new record as the earlier record, the quantity t_k Can be determined by the following equation 3.1 ,

$$t_k = \frac{y - x_1}{(y - t_1) + \sum_{j=1}^{k-1} (j+1)\,(x_j - x_{j+1})} \tag{3.1}$$

The random variable T_k corresponding to it has a β distribution with parameters one and $k-1$, so the formula for testing the p-value is 3.2,

$$p = \operatorname{prob}(T_k > t_k) = (1 - t_k)^{k-1} \tag{3.2}$$

We performed this test using the dates of the first five earliest COVID-19 cases generated in each country. The abnormal dates were subsequently corrected so

Country (incl. region)	First case test date	p	Need to be corrected?	Corrected date
Belgium	2020/2/4	0.02	Yes	2020/3/2
Cambodia	2020/1/27	0.03	Yes	2020/3/8
Sri Lanka	2020/1/27	0.01	Yes	2020/3/12
Sweden	2020/1/31	0.02	Yes	2020/2/27
Vanuatu	2020/11/11	0.05	Yes	2020/12/9
Viet Nam	2020/1/24	0.01	Yes	2020/3/24
Wallis and Futuna	2020/10/19	0.02	Yes	2020/11/23

TABLE 1. Test results for the earliest cases in each country

that the corrected dates accepted the null hypothesis ($p \geq 0.05$). The results are presented in Table 1 [1] .

A total of eight countries rejected the null hypothesis ($p < 0.05$), which indicates that the first cases detected in these countries may be at risk of isolated isolation. They were Belgium, Cambodia, Sri Lanka and Sweden, among others. We subsequently corrected the dates for these countries.

3.2. **Traceability using OLE methods.** The OLE method (Optimal Linear Estimation) is a non-parametric estimation method of ecological dates proposed by Roberts and Solow [26]. It can estimate the historical date of species extinction given a series of sightings. Clements et al. have experimentally demonstrated that the OLE method is robust to extinction extrapolation under various scenarios and can effectively estimate biological extinction dates.

As the underlying assumptions of the OLE algorithm are not just specific to organisms and species extinction processes, the method does not include any biologically specific parameters. It considers the time interval at which sightings occur and their distribution, regardless of the type of phenomenon under study. This makes it easily applicable to various phenomena, such as archaeology and environmental science [15, 25].

In addition, OLE does not require us to provide a complete and continuous record of sightings, which makes it suitable for exceptional cases where data are enormously missing, such as the missing cases that often occur in the tracing of new crown outbreaks. Finally, OLE is a non-parametric method that does not make assumptions about sighting rates or data distribution, making it more flexible than other methods [28]. Finally, OLE is based on extreme value theory, which suggests that the distribution of maximum values can be approximated by a generalized extreme value distribution regardless of the actual distribution of records [6].

For these reasons, in this study, we used the OLE algorithm to trace the COVID-19 epidemic in each country to estimate when the COVID-19 epidemic appeared in each country. Since COVID-19 tracing can be seen as tracking the date of origin of a species, which means the opposite direction of species extinction. We, therefore, need to replace the n_{end} dates of the last observation used in the original algorithm with the n_{start} dates of the original COVID-19 report.

[1]All tables in this paper can be found in https://github.com/williamlorder/The-systematic-traceability-and-analysis-of-COVID-19-Supplementary-materials

We make $T_1 > T_2 > \cdots > T_k$ be the first K earliest reporting dates for each country COVID-19. In this case, the OLE algorithm is consistent with the fact that the joint distribution of the first K earliest reporting dates has the exact and approximate extreme Weibull distribution, regardless of the original distribution of the complete reporting dates.In short, the OLE form of θ is a weighted sum of the reporting dates, calculated as equation 3.3,

$$\hat{\theta} = \sum_{i-1}^{k} a_i T_i \tag{3.3}$$

The following equation 3.4 give the weight vector α

$$a = \left(e^t \Lambda^{-1} e\right)^{-1} \Lambda^{-1} e \tag{3.4}$$

where e is a vector of k ones, and Λ is a $k \times k$ matrix with typical elements of λ_{ij}, which the following equation 3.5 can determine as,

$$\lambda_{ij} = \left(\Gamma\left(2\hat{\nu} + i\right)\Gamma\left(\hat{\nu} + j\right)\right) / \left(\Gamma\left(\hat{\nu} + i\right)\Gamma\left(j\right)\right), j \leq i \tag{3.5}$$

In addition, Γ is the standard gamma function and $\hat{\nu}$ is an estimate of the shape parameter of the joint Weibull distribution for the k most recently observed times, which is given by the following equation 3.6,

$$\hat{\nu} = \frac{1}{k-1} \sum_{i-1}^{k-2} log \frac{T_1 - T_k}{T_1 - T_{i+1}} \tag{3.6}$$

The approximate one-sided upper bound on the $1 - \alpha$ confidence interval for the final θ is T_E^u can be determined by the following equation 3.7,

$$T_E^u = \frac{t_n - c\left(\alpha\right) t_{n-k+1}}{1 - c\left(\alpha\right)} \tag{3.7}$$

where the $c\left(\alpha\right)$ the function is equal to equation 3.8,

$$c\left(\alpha\right) = \left(\frac{k}{-log\alpha}\right)^{-\hat{v}} \tag{3.8}$$

As OLE assumes an extreme Weibull distribution for sightings, it should only be used for the most recent sightings [28] . However, there is still debate about the optimal number of sightings, according to Collen, Purvis, and Mace et al. [7], increasing the number of sightings used (tested to a maximum of 18) would improve prediction accuracy. We therefore compromised by selecting up to 12 initial new crown reporting dates in each country as sightings for estimation. This means that the starting point is the date of the first day of reporting with new cases in each country and the endpoint is the twelfth date of reporting with new cases. We used this period for our estimates. The Solow and Smith's method was used first to correct the reporting date for the first example earlier. In the following estimates,

Country (incl. region)	Corrected date	N	Time interval	θ	T_E^u	Estimated date (θ)	Estimated date (T_E^u)
Italy	2019/9/30	12	173	4.52	6.54	2019/9/25	2019/9/23
France	2019/11/16	12	38	7.32	23.1	2019/11/8	2019/10/23
United States of America	2019/12/13	12	68	2.81	11.49	2019/12/10	2019/12/1
Spain	2020/1/20	12	41	20.64	84.54	2019/12/30	2019/10/27
China	2020/1/4	12	22	3.8	9.87	2019/12/31	2019/12/25
Republic of Korea	2020/1/19	10	18	15.38	20.71	2020/1/3	2019/12/29
Japan	2020/1/14	12	23	9.97	43.1	2020/1/4	2019/12/1
Thailand	2020/1/13	12	18	7.16	21.2	2020/1/5	2019/12/22
The United Kingdom	2020/2/1	12	29	20.88	4.1	2020/1/11	2020/1/27

TABLE 2. Estimation of the first 10 first origin dates using the OLE method

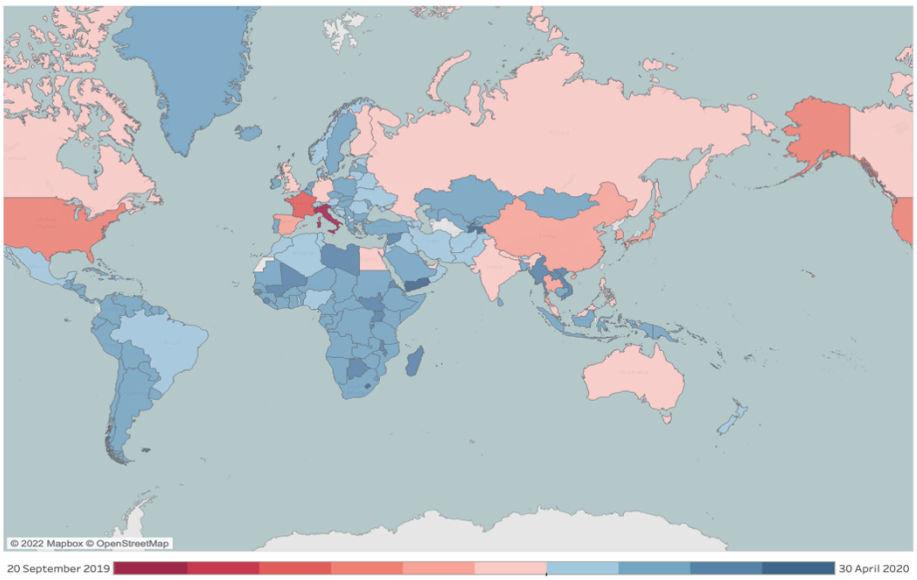

FIGURE 2. Estimated origin date by OLE(θ).The order of dates varies from blue to red, and this is the bluer the colour, the earlier the report, and vice versa.

the corrected dates will be used for estimation. The estimation results using the OLE method are given here in Table 2.

4. Discussion

Our results show Italy as the first country where the new crown epidemic originated. This is followed by the USA, France, Spain and China.

Figure 2 shows the estimated start dates for each country using the OLE point estimate (θ), and Figure 3 shows the estimated start dates for each country after taking the upper limit of the OLE estimate.

According to our modelling results, the first global cases of COVID-19 infection appeared in Italy, followed by France, the USA, Spain and China. After the initial cases appeared in Italy, they seem to have spread rapidly to the countries around

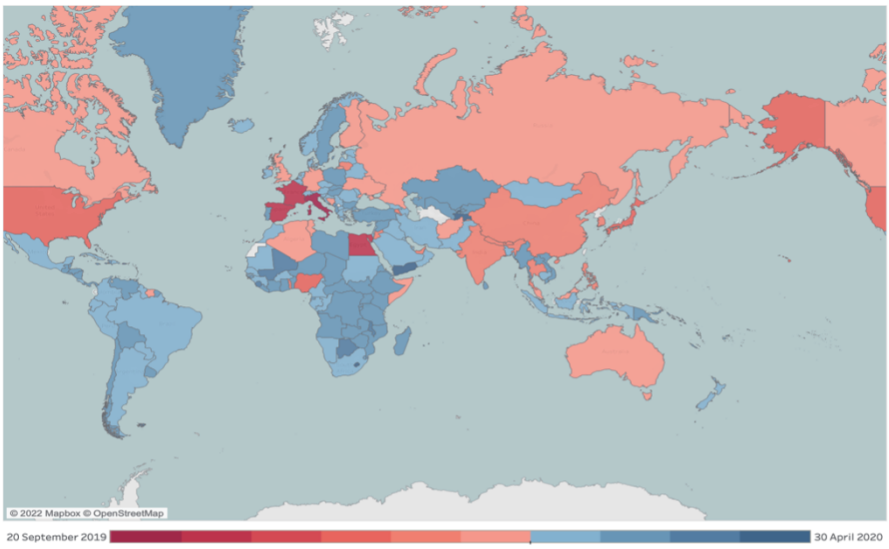

FIGURE 3. Estimated origin date by OLE(Upper).The order of dates varies from blue to red, and this is the bluer the colour, the earlier the report, and vice versa.

it (France and Spain), making Italy the center of infection in Europe, which is consistent with the published epidemic data for each country and the WHO study.

According to the results of our model, the epidemic broke out after China in South Korea, Japan and Thailand in Asia, and we assume that the virus spread from Europe and the United States to China through the global transport network, turning China into the center of transmission in Asia. This led to the further spread of the virus to various countries and regions in East Asia. Many studies and reports have pointed out that the early symptoms of COVID-19 are fever, cough and malaise, which can easily be mistaken for common influenza and thus overlooked [1, 12, 36]. Before being recognised by humans, the virus spreads silently through a tight global network of countries. This partly explains how epidemics in various countries can break out almost simultaneously in a concise period [16]. Our study raises the possibility that the origin of COVID-19 was not China, that the virus may have spread in Europe before spreading to China, and that humans only gradually recognised it in a large outbreak in Wuhan, China.

This paper has some innovations over previous studies. We have used more credible and authoritative sources (WHO data). We have also supplemented the data with a synthesis of country-reported virus testing results for 2019. The data were also tested for isolated cases using the method of Solow and smith [2, 5, 9, 17] and estimated using corrected dates [29, 30]. Finally, we chose up to 12 reporting dates as the estimation period according to Collen et al. According to their simulation experiments, the confidence interval of the model reached convergence when the number of reporting dates was 12 [7].

5. Conclusion

In this study, we used the authoritative WHO dataset for our research modelling, and the model reveals that the time and place of origin of the epidemic was far less straightforward than we thought, with rapid spread through the population as early as late 2019. The study offers a new way of thinking about epidemiological investigations, using a combination of the Solow and Smith method and the OLE method to make relatively robust inferences despite small samples. This approach allows for more accurate traceability in the absence of early cases, which often occurs in epidemiological surveys.

Future researchers could extend our research by several points.Firstly,Using more estimation methods to compare results more comprehensively, some of the current time series methods in the field of deep learning cope better with extreme values, and this could be transferred to retrospective studies; Secondly,more tuning and study of the parameters used in the model, we used the parameter estimates suggested by Collen et al. suggested parameter estimates. However, they used a simulated dataset, which is at variance with reality.

References

[1] Y. Alimohamadi, M. Sepandi, M. Taghdir and H. Hosamirudsari, *Determine the most common clinical symptoms in COVID-19 patients: a systematic review and meta-analysis*, Journal of Preventive Medicine and Hygiene **61** (2020), E304–E312.

[2] A. Amendola, S. Bianchi, M. Gori, D. Colzani, M. Canuti, E. Borghi, M. C. Raviglione, G. V Zuccotti and E. Tanzi, *Evidence of SARS-CoV-2 RNA in an oropharyngeal swab specimen, Milan, Italy, early December 2019*, Emerging Infectious Diseases **27** (2021): 648.

[3] K. G. Andersen, A. Rambaut, W. I. Lipkin, E. C. Holmes and R. F. Garry, *The proximal origin of SARS-CoV-2*, Nature Medicine **26** (2020), 450–452.

[4] S. V. Basavaraju, M. E. Patton, K. Grimm, U. M. A. Rasheed, S. Lester, L. Mills, M. Stumpf, B. Freeman, Brandi, A.Tamin and J. Harcourt, *Serologic testing of US blood donations to identify severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)–reactive antibodies: December 2019–January 2020*, Clinical Infectious Diseases **72** (2021), e1004–e1009.

[5] G. Chavarria-Miro,E. Anfruns-Estrada, A. Martínez-Velázquez, M. Vázquez-Portero, S. Guix, M. Paraira, B. Galofré and G. Sánchez, R. M. Pintó, A. Bosch, *Time evolution of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in wastewater during the first pandemic wave of COVID-19 in the metropolitan area of Barcelona, Spain*, Applied and Environmental Microbiology **87** (2021), e02750–20.

[6] S. Coles, J. Bawa, L. Trenner and P. Dorazio, *An Introduction to Statistical Modeling of Extreme Values*, vol. 208, Springer, 2001.

[7] B. Collen, A. Purvis and G. M. Mace, *BIODIVERSITY RESEARCH: When is a species really extinct? Testing extinction inference from a sighting record to inform conservation assessment*, Diversity and Distributions **16** (2010), 755–764.

[8] I. Cooper, A. Mondal and C. G. Antonopoulos, *A SIR model assumption for the spread of COVID-19 in different communities*, Chaos, Solitons & Fractals **139** (2020): 110057.

[9] A. Deslandes, V. Berti, Y. Tandjaoui-Lambotte, C. Alloui, E. Carbonnelle, J. R. Zahar, S. Brichler and Y. Cohen, *SARS-CoV-2 was already spreading in France in late December 2019*, International Journal of Antimicrobial Agents **55** (2020): 106006.

[10] France's early COVID-19 case may hold clues to pandemic's start, Reuters 2020, `https://www.reuters.com/article/us-health-coronavirus-france-retests-idUSKBN22H15R`, [Online; accessed 2022-02-14].

[11] S. Gopikrishnan, P. Priakanth, G. Srivastava and G. Fortino, Giancarlo, *EWPS: emergency data communication in the internet of medical things*, IEEE Internet of Things Journal **8** (2021), 11345–11356.

[12] R. Han, L. Huang, H. Jiang, J. Dong, H. Peng and D. Zhang, *Early clinical and CT manifestations of coronavirus disease 2019 (COVID-19) pneumonia*, AJR Am J Roentgenol **215** (2020), 338–343.

[13] K. L. Hon, K. K. Y. Leung, A. K. C. Leung, S. Sridhar, S. Qian, S. L. Lee and A. A. Colin, *Overview: the history and pediatric perspectives of severe acute respiratory syndromes: novel or just like SARS*, Pediatric Pulmonology **55** (2020), 1584–1591.

[14] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang and B. Cao, *Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China*, The Lancet **395** (2020), 497–506.

[15] A. J. M. Key, I. Jarić and D. L. Roberts, *Modelling the end of the acheulean at global and continental levels suggests widespread persistence into the middle palaeolithic*, Humanities and Social Sciences Communications **8** (2021), 1–12.

[16] N. C. Khanh, P. Q. Thai, H.-L. Quach, N.-A. H. Thi, P. C. Dinh, T. N. Duong, L T. Q. Mai, N. D. Nghia, T. A. Tu and L. N. Quang, T. D. Quang, T.-T. Nguyen, F. Vogt and D. D. Anh, *Transmission of SARS-CoV 2 during long-haul flight*, Emerging Infectious Diseases **26** (2020), 2617–2624.

[17] G. La Rosa, P. Mancini, G. B. Ferraro, C. Veneri, M. Iaconelli, L. Bonadonna, L. Lucentini and E. Suffredini, *SARS-CoV-2 has been circulating in northern Italy since December 2019: evidence from environmental monitoring*, Science of the Total Environment **750** (2021): 141711.

[18] P. K. Law, *COVID-19 pandemic: its origin, implications and treatments*, Open Journal of Regenerative Medicine **9** (2020), 43–64.

[19] H. Li, K. Yu, B. Liu, C. Feng, Z. Qin and G. Srivastava, *An efficient ciphertext-policy weighted attribute-based encryption for the internet of health things*, IEEE Journal of Biomedical and Health Informatics **PP** (2021): PMID: 33905340.

[20] Misleading to the Public: Students, Experts Criticize Harvard Study Suggesting Early Emergence of Covid-19 in Wuhan | News | The Harvard Crimson, `https://www.thecrimson.com/article/2021/4/22/lack-transparency-covid-wuhan-study/`, [Online; accessed 2022-02-24].

[21] D. M. Morens, J. G. Breman, C. H. Calisher, P. C. Doherty, B. H. Hahn,G. T. Keusch, L D. Kramer, J. W. LeDuc, T. P. Monath and J. K. Taubenberger, *The origin of COVID-19 and why it matters*, The American Journal of Tropical Medicine and Hygiene, **103** (2020), 955–959.

[22] E. O. Nsoesie, B. Rader, Y. Barnoon, L. Goodwin and J. Brownstein, *Analysis of hospital traffic and search engine data in Wuhan China indicates early disease activity in the fall of 2019*, 2020.

[23] D. Polap, G. Srivastava and K. Yu, *Agent architecture of an intelligent medical system based on federated learning and blockchain technology*, **58** (2021): 102748.

[24] A. Rambaut, D. L. Robertson, O. G. Pybus, M. Peeters and E. C. Holmes, *hylogeny and the origin of HIV-1*, Nature **410** (2001), 1047–1048.

[25] M. M. Rivadeneira, G. Hunt and K. Roy, *The use of sighting records to infer species extinctions: an evaluation of different methods*, Ecology **90** (2009), 1291–1300.

[26] D. Roberts and A. R. Solow, *When did the dodo become extinct?*, Nature **426** (2003), 245–245.

[27] R. Segreto, Y. Deigin, K. McCairn, A. Sousa, D. Sirotkin, K. Sirotkin, J. J. Couey, A. Jones and D, Zhang, *Should we discount the laboratory origin of COVID-19?*, Environmental Chemistry Letters **19** (2021), 2743–2757.

[28] A. R. Solow, *Inferring extinction from a sighting record*, Mathematical Biosciences **195** (2005), 47–55.

[29] A. R. Solow, A. C. Kitchener, D. L. Roberts and J. D. S. Birks, *Rediscovery of the Scottish polecat, Mustela putorius: Survival or reintroduction?*, Biological Conservation, **128** (2006), 574–575.

[30] A. R. Solow and W. Smith, *How surprising is a new record?*, The American Statistician **59** (2005), 153–155.
[31] G. Srivastava and R. Venkataraman, *A review of the state of the art in business intelligence software*, Enterprise Information Systems **16** (2022), 1–28.
[32] H. H. Weiss, *The SIR model and the foundations of public health*, Materials Matematics (2013), 1–17.
[33] WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data, `https://covid19.who.int/info?openIndex=2`, [Online; accessed 2022-01-26].
[34] Z. Yang, Y. Hu, Z. Ding and T. Guo, *Dating the First Case of COVID-19 Epidemic from a Probabilistic Perspective*, pages 14.
[35] K. Yu, L. Tan, X. Shang, J. Huang, G. Srivastava and P. Chatterjee, *Efficient and privacy-preserving medical research support platform against COVID-19: a blockchain-based approach*, IEEE Consumer Electronics Magazine **10** (2020), 111–120.
[36] Z. Y. Zu, M. D. Jiang, P. P. Xu, W. Chen, Q. Q. Ni, G. M. Lu and L. J. Zhang, *Coronavirus disease 2019 (COVID-19): a perspective from China*, Radiology **296** (2020), E15–E25.

Manuscript received March 4, 2022
revised March 14, 2022

Z. Lu
School of Public Administration,Guangzhou University,China
E-mail address: `1903500028@e.gzhu.edu.cn`

L. Peng
School of Mechanical and Electrical Engineering,Guangzhou University,China
E-mail address: `flyingday@139.com`

X. Liu
School of Education,Guangzhou University,China
E-mail address: `1300048991@qq.com`

Y. Zhong
College of Life Sciences,Guangzhou University,China
E-mail address: `2690485414@qq.com`

R. Hu
School of Public Administration,Guangzhou University,China
E-mail address: `hurong@gzhu.edu.cn`

H. Liu
School of Chemistry,Guangzhou University,China
E-mail address: `huaihuai99@163.com`