



Regional Method for Israel COVID-19 Forecast

Gaidai O¹, Liu Z¹, Xing Y^{2*} and Wang K¹

¹Shanghai Ocean University, China

²University of Stavanger, Norway

*Corresponding author: Yihan Xing, University of Stavanger, Norway, Email: yihan.xing@uis.no

Research Article

Volume 6 Issue 4

Received Date: December 07, 2022

Published Date: December 30, 2022

DOI: 10.23880/eij-16000253

Abstract

Novel coronavirus disease 2019 (COVID-19) is a contagious disease having high transmissibility to spread worldwide along with low morbidity and mortality, presenting certain burden on worldwide public health. Due to non-stationarity and complicated nature of epidemic waves, it is challenging to accurately model such phenomenon. Traditional statistical methods dealing with spatio-temporal observations of multi-observational processes do not have an advantage of dealing with extensive observational dimensionality and cross-correlation between different local observations. This study has chosen COVID-19 daily numbers of recorded new and dead patients in Israel.

Keywords: COVID-19; Epidemic Outbreak; Probability Forecast; Public Health; Mathematical Biology

Abbreviations: CI: Confidence Interval; GEV: Generalized Extreme Value; SODP: Second-Order Difference Plot; AI: Artificial Intelligence; EVT: Extreme Value Theory.

Introduction

In the contemporary research community, the statistical aspects of COVID-19 and other recent epidemics of a similar nature were receiving considerable attention [1]. In general, using traditional theoretical statistical methods to determine outbreak probabilities and realistic health system reliability factors under actual epidemic conditions is quite difficult [1-3]. The only observation figures for COVID-19, however, are constrained to the first of 2020. Israel was obviously selected because to its COVID-19 origin, substantial health monitoring, and available internet research [4,5]. This study relied on the quasi-stationarity assumption; the underlying epidemiological process was assumed, although seasonally varying but still statistically representative during two 2020-2022 years of continuous observation. The underlying trend must be addressed in the longer time horizon. Note

that although analysed data set is quite limited both in time (only two years) as well as in space (only state of Israel), the proposed methodology has already been well validated in other studies by authors Gaidai O, et al. [6-19].

Methods

A suitably enough period of time has been used to assess the MDOF health system response vector process $(t) = (X(t), Y(t), Z(t), \dots) (o, T)$. Unidimensional

global maxima over the full period (o, T) are

$$X_T^{\max} = \max_{0 \leq t \leq T} X(t), \quad Y_T^{\max} = \max_{0 \leq t \leq T} Y(t),$$

$$Z_T^{\max} = \max_{0 \leq t \leq T} Z(t), \dots \text{ see Fig. 2. One generally implies a big}$$

value of with regard to the dynamic system auto-correlation time when they refer to a suitably lengthy duration. Let X_1, \dots, X_{N_x} represent subsequent local maxima of the

process $X(t)$ at discrete temporally increasing time instants $t_1^X < \dots < t_{N_X}^X$ in time $(0, T)$. For other MDOF response components, such as $Y(t), Z(t)$, etc, the analogous definition is as follows: $Y_1, \dots, Y_{N_Y}; Z_1, \dots, Z_{N_Z}$ and so on. For the sake of simplicity, it is assumed that all of $\mathbf{R}(t)$

components and their maxima are positive. The goal is to calculate the likelihood of a system failing, specifically the likelihood of exceeding, accurately

$$1 - P = \text{Prob}(X_T^{\max} > \eta_X \cup Y_T^{\max} > \eta_Y \cup Z_T^{\max} > \eta_Z \cup \dots) \quad (1)$$

$$P = \int_{(0, 0, 0, \dots)}^{(\eta_X, \eta_Y, \eta_Z, \dots)} p_{X_T^{\max}, Y_T^{\max}, Z_T^{\max}, \dots}(X_T^{\max}, Y_T^{\max}, Z_T^{\max}, \dots) dX_T^{\max} dY_T^{\max} dZ_T^{\max} \dots$$

is the likelihood that critical values of the response components $\eta_X, \eta_Y, \eta_Z, \dots$ will not be exceeded; stands for the logical unity operation «or»; and $p_{X_T^{\max}, Y_T^{\max}, Z_T^{\max}, \dots}$ is the joint

probability density of the global maxima throughout the whole time $(0, T)$. Due to its large dimensionality and the

limits of the provided data set, it is not practical to directly estimate the latter joint probability distribution.

More specifically, the health system is considered to have failed instantly when either $X(t)$ exceeds η_X , $Y(t)$ exceeds η_Y , $Z(t)$ exceeds η_Z , and so on. For each unidimensional response component of $\mathbf{R}(t)$, fixed failure levels $\eta_X, \eta_Y, \eta_Z, \dots$

are, of course, unique $\mathbf{R}(t)$. $X_{N_X}^{\max} = \max\{X_j; j = 1, \dots, N_X\} = X_T^{\max}$,

$$Y_{N_Y}^{\max} = \max\{Y_j; j = 1, \dots, N_Y\} = Y_T^{\max},$$

$$Z_{N_Z}^{\max} = \max\{Z_j; j = 1, \dots, N_Z\} = Z_T^{\max}, \text{ and so forth. Time}$$

instants at local maxima $[t_1^X < \dots < t_{N_X}^X; t_1^Y < \dots < t_{N_Y}^Y; t_1^Z < \dots < t_{N_Z}^Z]$ are

combined into a single time vector, $t_1 \leq \dots \leq t_N$, in

monotonously non-decreasing order. Keep in mind that $t_N = \max\{t_{N_X}^X, t_{N_Y}^Y, t_{N_Z}^Z, \dots\}$, $N = N_X + N_Y + N_Z + \dots$

In this instance, the local maxima of one of the MDOF structural response components, such as $X(t)$ or $Y(t)$, or

$Z(t)$, and so on, are represented by \cdot . This means that, given

a $\mathbf{R}(t)$ time record, all that is required to record a component's MDOF limit vector $(\eta_X, \eta_Y, \eta_Z, \dots)$ exceedance in any of

its components is to continuously and simultaneously screen for local maxima in the unidimensional response component. Following the merged time vector $t_1 \leq \dots \leq t_N$, local

unidimensional response component maxima are combined into a single temporal non-decreasing vector $\bar{\mathbf{R}} = (R_1, R_2, \dots, R_N)$. In other words, each local maxima R_j

corresponds to an actual encountered local maxima for either $X(t)$ or $Y(t)$, or $Z(t)$, and so on. The unified limit

vector (η_1, \dots, η_N) is then introduced, and each component

η_j is either an η_X, η_Y or η_Z component, depending on which

of $X(t)$ or $Y(t)$, or $Z(t)$, etc, corresponds to the running

index 's current local maxima.

The new MDOF limit vector $(\eta_X^\lambda, \eta_Y^\lambda, \eta_Z^\lambda, \dots)$ with

$$\eta_X^\lambda \equiv \lambda \cdot \eta_X, \eta_Y^\lambda \equiv \lambda \cdot \eta_Y, \eta_Z^\lambda \equiv \lambda \cdot \eta_Z, \dots$$

is introduced, along with the caling parameter $0 < \lambda \leq 1$, to artificially simultaneously decrease limit values for all response components. Each component of the unified limit vector $(\eta_1^\lambda, \dots, \eta_N^\lambda)$ is named η_j^λ and is either $\eta_X^\lambda, \eta_Y^\lambda$ or η_Z^λ , and

so on. The latter defines probability $P(\lambda)$ automatically as a

function of λ ; take note that $P \equiv P(1)$ from Eq. (1). Non-

exceedance probability $P(\lambda)$ will be now estimated as follows

$$\begin{aligned} P(\lambda) &= \text{Prob}\{R_N \leq \eta_N^\lambda, \dots, R_1 \leq \eta_1^\lambda\} \\ &= \text{Prob}\{R_N \leq \eta_N^\lambda | R_{N-1} \leq \eta_{N-1}^\lambda, \dots, R_1 \leq \eta_1^\lambda\} \cdot \text{Prob}\{R_{N-1} \leq \eta_{N-1}^\lambda, \dots, R_1 \leq \eta_1^\lambda\} \\ &= \prod_{j=2}^N \text{Prob}\{R_j \leq \eta_j^\lambda | R_{j-1} \leq \eta_{j-1}^\lambda, \dots, R_1 \leq \eta_1^\lambda\} \cdot \text{Prob}\{R_1 \leq \eta_1^\lambda\} \end{aligned} \quad (2)$$

In practice, dependence between neighbouring R_j is not always negligible; thus, the following one-step (will be called here conditioning level $k = 1$) memory approximation being

introduced

$$\text{Prob}\{R_j \leq \eta_j^\lambda \mid R_{j-1} \leq \eta_{j-1}^\lambda, \dots, R_1 \leq \eta_1^\lambda\} \approx \text{Prob}\{R_j \leq \eta_j^\lambda \mid R_{j-1} \leq \eta_{j-1}^\lambda\} \quad (3)$$

for $2 \leq j \leq N$ (conditioning level). The approximation introduced by Eq. (3) will be further expressed as

$$\text{Prob}\{R_j \leq \eta_j^\lambda \mid R_{j-1} \leq \eta_{j-1}^\lambda, \dots, R_1 \leq \eta_1^\lambda\} \approx \text{Prob}\{R_j \leq \eta_j^\lambda \mid R_{j-1} \leq \eta_{j-1}^\lambda, R_{j-2} \leq \eta_{j-2}^\lambda\} \quad (4)$$

where $3 \leq j \leq N$ (will be called conditioning level $k = 3$), and so on. Subsequent improvements to the statistical independence assumption are presented in Eq. (4). This study makes the assumption that the entire health system, including all of its dimensions, is stationary. As a result, the constructed process $\mathbf{R}(t)$ is also assumed to be stationary. Since original MDOF process $\mathbf{R}(t)$ is stationary because it is assumed to be ergodic, the probability will be independent of j but solely dependent on conditioning level k for $j \geq k$. So, it is possible to approximate non-exceedance probability using the average conditional exceedance rate method.

$$P_k(\lambda) \approx \exp(-N \cdot p_k(\lambda)), \quad k \geq 1. \quad (5)$$

It should be noted that Eq. (5) deduces from Eq. (1) by ignoring $\text{Prob}(R_1 \leq \eta_1^\lambda) \approx 1$ and assuming $N \gg k$ because the probability of a design failing must be extremely low. The well-known mean up-crossing rate equation for the probability of exceeding is comparable to Eq. (5) [20]. There is evident convergence with respect to the conditioning parameter k

$$P = \lim_{k \rightarrow \infty} P_k(1); \quad p(\lambda) = \lim_{k \rightarrow \infty} p_k(\lambda) \quad (6)$$

Note that Eq. (5) for $k = 1$ turns into a known non-exceedance probability expression for the mean up-crossing rate function

$$P(\lambda) \approx \exp(-\nu^+(\lambda)T); \quad \nu^+(\lambda) = \int_0^\infty \zeta P_{RR}(\lambda, \zeta) d\zeta \quad (7)$$

Where $\nu^+(\lambda)$ stands for the mean up-crossing rate of the response level for the non-dimensional vector $R(t)$ above that was constructed from scaled MDOF system response $\left(\frac{X}{\eta_X}, \frac{Y}{\eta_Y}, \frac{Z}{\eta_Z}, \dots\right)$. Rice's formula, which is given in Eq. (7),

yields the mean up-crossing rate, where is the joint probability density for (R, \dot{R}) , where is the time derivative $R'(t)$. The Poisson assumption, which is the basis for Eq.

(7), states that up-crossing events with high λ levels (in the current study, it is $\lambda \geq 1$) can be assumed to be independent. The stationarity assumption was applied in the examples above. The nonstationary case, however, can also be handled by the suggested methodology. Each short-term seasonal state has a probability of q_m in the nonstationary case's

scattered diagram of $m = 1, \dots, M$ seasonal epidemic conditions, meaning that $\sum_{m=1}^M q_m = 1$. Next, let one introduce

the long-term equation

$$p_k(\lambda) \equiv \sum_{m=1}^M p_k(\lambda, m) q_m \quad (8)$$

With $p_k(\lambda, m)$ being same function as in Eq. (6), corresponding to a specific short-term seasonal epidemic state with the number m . The $p_k(\lambda)$ functions mentioned above are frequently regular in the tail, especially for values of approaching and exceeding 1. More specifically, for $\lambda \geq \lambda_0$, the distribution tail behaves as $\exp\{-(a\lambda + b)^c + d\}$, where a, b, c, d are appropriately fitted constants for an appropriate tail cut-on λ_0 value. Therefore, one can write

$$p_k(\lambda) \approx \exp\{-(a_k \lambda + b_k)^{c_k} + d_k\}, \quad \lambda \geq \lambda_0 \quad (9)$$

Next, nearly perfectly linear tail behavior is seen by plotting $\ln\{\ln(p_k(\lambda)) - d_k\}$ versus $\ln(a_k \lambda + b_k)$.

Sequential quadratic programming (SQP), a tool included in the NAG Numerical Library, can also be used to find the parameters a_k, b_k, c_k, p_k, q_k ideal values [8-19].

Results

In both epidemiology and mathematical biology, the prediction of influenza-like epidemics has long been the main focus. The website that presents COVID-19 data from Israel is where the statistical information in this section is taken from [4]. The website provides numbers of newly diagnosed and death cases every day in Israel from 22 January 2020 to 6 April 2022. In order to illustrate a two-dimensional (2D) dynamic health system, patient numbers from two distinct Israel health recorded categories—daily recorded patients and deaths, respectively—were selected as components. In order to unify both two measured time series X, Y the following scaling was performed

$$X \rightarrow \frac{X}{\eta_X}, \quad Y \rightarrow \frac{Y}{\eta_Y} \quad (10)$$

Making both two responses non-dimensional and having the same failure limit equal to 1. In the current study, failure limits, or alternatively epidemic thresholds, were set at the observed two-year maxima, twice increased, for each category of the health system. Then, all local maxima from both measured time series were combined into a single time series while maintaining their temporal non-decreasing order: $\bar{R} = (\max\{X_1, Y_1\}, \dots, \max\{X_N, Y_N\})$ with the

entire vector being sorted in temporally non-decreasing order of these local maxima's occurrence.

Figure 1 presents 100-years return level extrapolation in accordance with Eq. (9) towards epidemic outbreak with

100-year return period. This extrapolation is shown by the horizontal dotted line, and a cut-off value of λ was used to further extrapolate it. According to Eq. (10), dotted lines represent the extrapolated 95% confidence interval (CI). The goal failure probability $1 - P$ from Eq. (5) is directly connected to $p(\lambda)$ according to that equation Eq.(1). In Figure 1, the star symbolizes the expected non-dimensional λ level, which denotes the likelihood that an epidemic would break out at any given level of the Israeli health system in the years to come. For the method validation, the asymptotic Gumbel generalized extreme value (GEV) distribution was applied, namely extrapolation was done on Gumbel probability plot, prediction was found located within 95% CI, depicted in Figure 1.

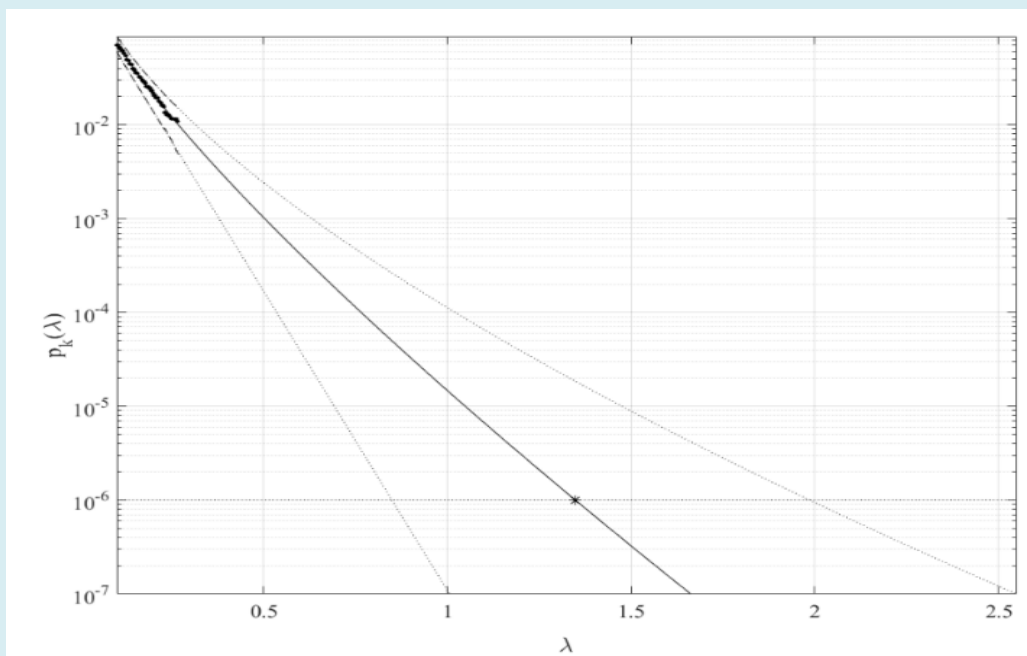
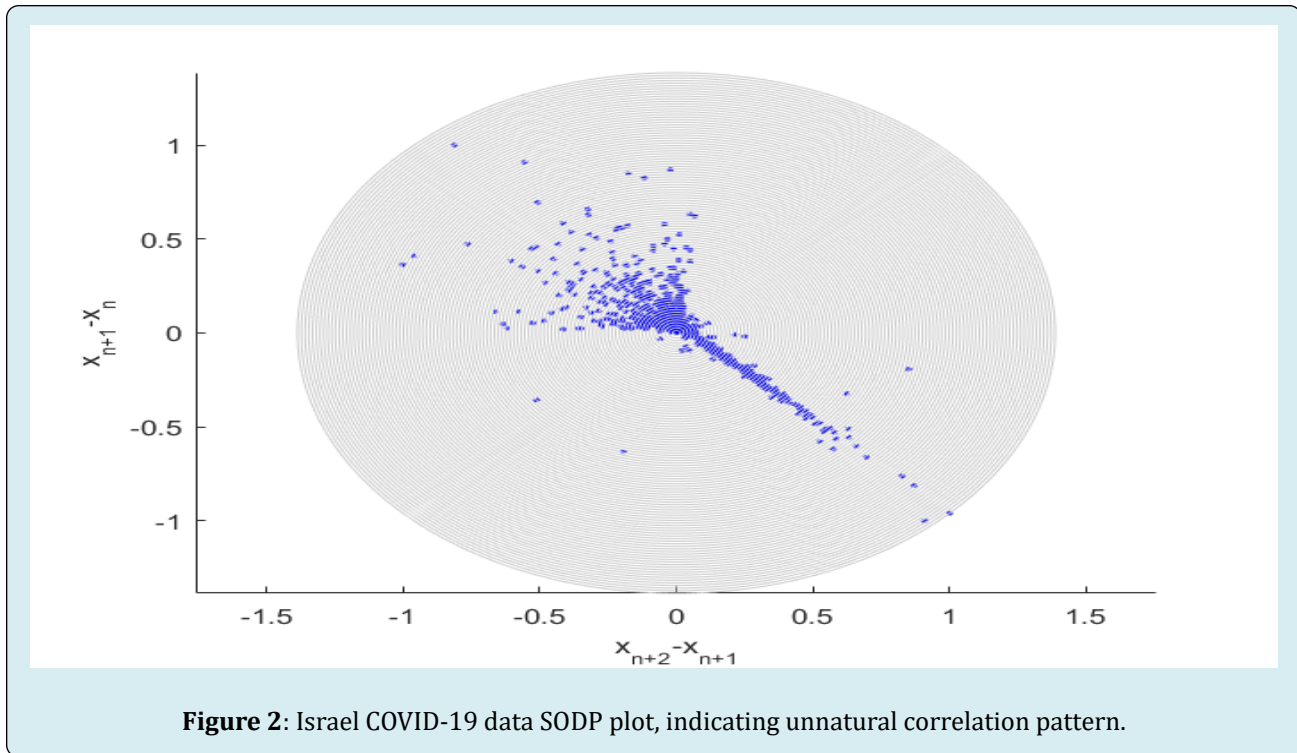


Figure 1: 100-years return level (horizontal dotted line) extrapolation of $p_k(\lambda)$ towards target level (indicated by star) and beyond. Extrapolated 95% CI indicated by dotted lines.

The Poincare plot is where the second-order difference plot (SODP) got its start. Time series data from SODP can be used to observe the statistical situation of consecutive differences.

Figure 2 presents SODP plot, this type of plot can also be used to identify data patterns and compare them to other data sets, such as when using an artificial intelligence (AI) recognition method based on entropy [21]. Figure 2 indicates unnatural correlation pattern - the reason for that is presence of the straight diagonal line pattern Figure 2, not

present on other national COVID-19 data SODP plots, for example Norway. SODP plot straight diagonal line means that large portion of the data consecutive differences $x_{n+2} - x_{n+1}$ being almost equal to $x_{n+1} - x_n$, which is un-physical and at least raises suspicions, at most may indicate the data rigging. Keep in mind that while this paper presents MDOF and a sub-asymptotic method, extreme value theory (EVT) is asymptotic and 1DOF. Despite its apparent simplicity, the current work presents a unique multidimensional modeling technique and a methodological route to apply epidemic forecasting while it is still in progress.



Discussion

The disadvantage of coping effectively with systems with high dimensionality and cross-correlation between various systems responses is not present in traditional health systems dependability approaches dealing with observed time series. The capacity to examine the dependability of high dimensional non-linear dynamic systems is the main benefit of the approach that has been developed. Despite its apparent simplicity, the current work presents a unique multidimensional modeling technique and a methodological route to apply epidemic forecasting while it is still in progress. This study looked at recently reported COVID-19 patient and death figures from Israel, which is an illustration of a two-dimensional (2D) phenomenon that was seen in 2020–2022. It should be noted that the provided strategy works effectively for more than simply two dimensional levels. As a result, the recommended technique may be useful for a range of dependability investigations on non-linear dynamic health systems.

Acknowledgements

The authors declare no conflict of interest. No funding was received. All authors contributed equally. The datasets analysed during the current study are available online.

Conflict of Interest

There are no conflicts to declare.

References

1. Chen J, Lei X, Zhang L, Peng B (2015) Using Extreme Value Theory Approaches to Forecast the Probability of Outbreak of Highly Pathogenic Influenza in Zhejiang, China. *PLoS One* 10(2): e0118521.
2. Mugglin AS, Cressie N, Gemell I (2002) Hierarchical statistical modelling of influenza epidemic dynamics in space and time. *Stat Med* 21(18): 2703-2721.
3. Naess A, Gaidai O (2009) Estimation of extreme values from sampled time series. *Structural Safety* 31(4): 325-334.
4. Israel COVID data.
5. Thomas M, Rootzen H (2019) Real-time prediction of severe influenza epidemics using Extreme Value Statistics. *arXiv*.
6. Gaidai O, Xing Y, Xu X (2022) COVID-19 epidemic forecast in USA East coast by novel reliability approach. *Research Square*.
7. Xu X, Xing Y, Gaidai O, Wang K, Patel K, et al. (2022) A novel multi-dimensional reliability approach for floating wind turbines under power production conditions. *Frontiers in Marine Science*.
8. Gaidai O, Xing Y, Balakrishna R (2022) Improving extreme response prediction of a subsea shuttle tanker hovering in ocean current using an alternative highly correlated

- response signal. Results in Engineering 15: 100593.
9. Cheng Y, Gaidai O, Yurchenko D, Xu X, Gao S (2022) Study on the Dynamics of a Payload Influence in the Polar Ship. The 32nd International Ocean and Polar Engineering Conference, Paper Number: ISOPE-I-22-342.
 10. Gaidai O, Storhaug G, Wang F, Yan P, Naess A, et al. (2022) On-Board Trend Analysis for Cargo Vessel Hull Monitoring Systems. The 32nd International Ocean and Polar Engineering Conference, Paper Number: ISOPE-I-22-541.
 11. Gaidai O, Xu X, Naess A, Cheng Y, Ye R, et al. (2020) Bivariate statistics of wind farm support vessel motions while docking. Ships and Offshore Structures 16(2): 135-143.
 12. Gaidai O, Fu S, Xing Y (2022) Novel reliability method for multidimensional nonlinear dynamic systems. Marine Structures 86: 103278.
 13. Gaidai O, Yan P, Xing Y (2022) A novel method for prediction of extreme wind speeds across parts of Southern Norway. Front Environ Sci.
 14. Gaidai O, Yan P, Xing Y (2022) Prediction of extreme cargo ship panel stresses by using deconvolution. Front Mech Eng.
 15. Balakrishna R, Gaidai O, Wang F, Xing Y, Wang S (2022) A novel design approach for estimation of extreme load responses of a 10-MW floating semi-submersible type wind turbine. Ocean Engineering 261: 112007.
 16. Gaidai O, Yan P, Xing Y, Xu J, Wu Y (2022) A novel statistical method for long-term coronavirus modelling. F1000 Research.
 17. Gaidai O, Xu J, Yan P, Xing Y, Zhang F, et al. (2022) Novel methods for wind speeds prediction across multiple locations. Scientific Reports 12: 19614.
 18. Gaidai O, Xing Y (2022) Novel reliability method validation for offshore structural dynamic response. Ocean Engineering 266(5): 113016.
 19. Gaidai O, Wu Y, Yegorov I, Alevras P, Wang J, et al. (2022) Improving performance of a nonlinear absorber applied to a variable length pendulum using surrogate optimization. Journal of Vibration and Control.
 20. Rice SO (1944) Mathematical analysis of random noise. Bell System Tech J 23(3): 282-332.
 21. Yayık A, Kutlu Y, Altan G (2019) Regularized HessELM and Inclined Entropy Measurement for Congestive Heart Failure Prediction. arXiv.

