



ISSN: 2348-5906
CODEN: IJMRK2
IJMR 2022; 9(2): 111-116
© 2022 IJMR
www.dipterajournal.com
Received: 10-01-2022
Accepted: 12-02-2022

Stavelin Abhinandithe K
Assistant Professor, Division of
Medical Statistics, School of Life
Sciences, JSS AHER, Mysore,
Karnataka, India

Madhu B
Professor, Community Medicine,
JSS Medical College, JSS AHER,
Mysore, Karnataka, India

Somanathan Balasubramanian
Former, Director Research,
JSS AHER, Mysore, Karnataka,
India

Sridhar Ramachandran
HOD, Ramakrishna Mission
Vidyalaya, MCA, Mysore,
Karnataka, India

Corresponding Author:
Stavelin Abhinandithe K
Assistant Professor, Division of
Medical Statistics, School of Life
Sciences, JSS AHER, Mysore,
Karnataka, India

Time series forecasting using GMDH neural networks for Chikungunya in Mysore district, India

Stavelin Abhinandithe K, Madhu B, Somanathan Balasubramanian and Sridhar Ramachandran

DOI: <https://doi.org/10.22271/23487941.2022.v9.i2b.607>

Abstract

Infectious diseases are the diseases which are caused by microorganisms. It plays a vital role in day to day life. Infectious diseases prediction is the growing field in the present scenario. Predicting and forecasting the infectious diseases helps as early warning to prepare for medical emergencies. Among Infectious diseases, chikungunya is an infection spread by mosquitos that carry the dengue and Zika viruses. Infectious illness transmission spread not only depends on vector and host but also depends on environmental/meteorological factors which are a complex process that requires advanced computer approaches such as soft computing and artificial intelligence to estimate and anticipate the complex phenomenon to predict and forecast infectious diseases. In order to predict and forecast the infectious diseases, GMDH techniques are used. The data is collected from Mysore district. We have considered confirmed cases of chikungunya from 2006-2019 along with 19 meteorological variables. A lag period of 0-7 was considered. GMDH models showed that for chikungunya cases, the RMSE value was low for 5/7 with less parameter. Here we have considered the Minimum temperature, mean relative humidity and average sunshine as the significant predictors in predicting and forecasting the chikungunya incidences.

Keywords: Chikungunya, meteorological parameters, GMDH, time series, RMSE

1. Introduction

Infectious diseases prediction is an important issue for public health authorities and industry stakeholders all over the world^[1]. Employee absence and a rapid increase in hospital workload also affect economic output. This has prompted public health officials to forecast the Infectious disease in several countries. For the purposes of assessment and evaluating disease control, multiple methodologies are used to predict future events based on past experiences. Many linear and nonlinear models are used to forecast epidemics^[2] and these models adopt epidemiological time series data to make short- and long-term predictions.

Chikungunya is a disease spread by mosquitos that carry the dengue and Zika viruses. It may also possibly spread through infected blood^[3]. The Chikungunya disease has caused a number of epidemics in Africa and Asia. A large epidemic occurred in the Indian Ocean in 2005-2006. Cases were discovered in Asia, Australia, the United States, Canada, and continental Europe. For the first time in Europe (Italy), an outbreak of autochthonous chikungunya virus infection occurred in 2007. Huge incidences were reported in France between 2010 and 2014. Chikungunya first appeared in the Caribbean in December 2013, and it soon moved across the United States. At present, the virus had spread throughout America's (sub) tropical areas, as well as Africa and Asia^[3].

It was found that nearly 1.2 million suspected and confirmed cases have chikungunya diseases were recorded in Caribbean and other parts of United States till the end of February 2015 followed by chikungunya outbreaks in the Indian Ocean in 2005-2006 and in 2007 it spread to Italy^[4, 5]. A network for arthropod vector surveillance for human public health to improve the surveillance of competent vectors for infectious diseases was established through ECDC (European Centre for disease prevention and control).

The majority of those who are afflicted experience severe symptoms. They typically begin three to seven days after being bitten by an infected mosquito. Fever and joint pain are the most typical symptoms. Headache, muscle soreness, joint swelling, and rashes are some of the other symptoms^[6, 7].

Infectious illness transmission spread not only depends on vector and host but also depends on environmental/meteorological factors which are a complex process that requires advanced computer approaches such as soft computing and artificial intelligence to estimate and anticipate the complex phenomenon to predict and forecast infectious diseases such as chikungunya^[6, 7, 8].

In the present study, we have examined the impact of meteorological factors with respect to the incidence of chikungunya for the district of Mysuru using GMDH (Group method of data handling) technique. This article intends to provide experiences for working on the chikungunya control arrangement and, as a result, to assist in making progress prior to the time for the public area.

2. Materials and Methods

2.1. Study Area and Climatic Conditions

Mysore shapes the southern piece of Karnataka, southeast part has Tamil Nadu, the area of Kodagu to its west, the region of Mandya lies in the north, Hassan in the northwest, and the Bangalore to its upper east.

2.2. Geographical Area

The district covers an area of 6, 76,382 hectares, with 62,851 hectares of forestland. Mysore City is one of the Karnataka's third most populous city, with an estimated population of 920,550 people in 2011, with 461,042 men and 459,508 women for the same year.

2.3. Climatic Condition

According to the Köppen climatic classification, Mysore has a tropical savanna climate (Aw) that borders on a hot semi-arid climate (BSh)^[9]. Summer, which runs from March to June, the monsoon season, which runs from July to November, and winter, which runs from December to February, are the three main seasons. As per the Indian Meteorological Department (IMD) records, the average maximum temperature per annual is 30.2 °C and the average minimum temperature is 19.1 °C^[10].

2.4. Data and Climatic Factors

We have collected the data of chikungunya incidences from District Health Office, Mysore, Ministry of Health and Family Welfare, Government of Karnataka from 2006 to 2019. Month wise incidences were collected and these are considered as 'n' observations. Here n=156 months of observation.

Meteorological data for Mysore district was collected from Indian Meteorological Division, Govt of India for the same period i.e. from the year 2006 to the year 2019. From the meteorological division month-wise data of min temperature (°C), max temperature (°C), mean temperature (°C), minimum rainfall (mm), maximum rainfall (mm), mean rainfall (mm), total rain during month (mm), mean of daily velocity (Km/hr.), minimum relative humidity (%), maximum relative humidity (%), mean relative humidity (%), minimum vapor pressure (mill bars), maximum vapor pressure (mill bars), mean of daily vapor pressure (mill bars), minimum

evaporation (mm), maximum evaporation (mm), total evaporation during month (mm), mean of daily evaporation (mm) and average sunshine (hours) were considered. Total of 19 meteorological variables were considered for the present study.

We have used Spearman's correlation analysis to find the relationship between infectious disease and meteorological variables by using IBM SPSS 22. The meteorological variables along with infectious diseases which showed significant at 5% were used for model build. The results are shown for 5/7 and 7/7 with Less parameter (LP) and Full parameter (FP) for the chikungunya disease. All the significant parameters of meteorological variables are taken as full parameter and the variables which were significant at 1% and 5% are considered as less parameter. Further trial and error was used to select the variables which satisfied the selection criteria. GMDH shell is further used to get the results.

2.5. Objectives

This paper intends to test the applicability of a polynomial GMDH-type neural network method in time series prediction in the context of infectious diseases along with environmental variables.

2.6. GMDH Procedure

The GMDH approach was created to handle problems involving higher order regression polynomials, including modelling and classification. A polynomial series in the form of the Volterra series, known as the Kolmogorov-Gabor polynomial, can be used to explain the general link between input and output variables^[11, 12].

$$y = a + \sum_{i=1}^m b_i \cdot x_i + \sum_{i=1}^m \sum_{j=1}^m c_{ij} \cdot x_i \cdot x_j + \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m d_{ijk} \cdot x_i \cdot x_j \cdot x_k + \dots \quad (1)$$

GMDH is a collection of generalization techniques for computer-based mathematical modeling of multi parametric data groups, including completely automatic structural and parametric model optimization. In the context of uncertainty, GMDH is a revolutionary strategy for resolving structural and parametric identification challenges. It is a sub-model of the ANN model that deals with time series data (Artificial Neural Network)^[11, 12].

The basic idea behind the GMDH is to use a quadratic node transfer function with regression coefficients to produce an analytical function in a feed-forward network. The GMDH is a self-organizing, unidirectional structure with many layers, each made up of neurons with similar structures. Using the singular value decomposition approach and solving normal equations, weight is introduced as definite and constant values inside each neuron.

All pairwise combinations of p lagged time series are considered by the GMDH method. As a result, each neuron receives each combination. A model is built using these two inputs to approximate the intended outcome. In other terms, a neuron receives two input variables and outputs one result. GMDH's fundamental function is based on the forward propagation of signals through the network's nodes, which is similar to the approach utilized in traditional neural networks. Every layer is made up of simple nodes that each conduct their own polynomial transfer function and pass their output to the next layer's nodes^[13, 14, 15].

Because of their capability to forecast any time series and continuous nonlinear function with arbitrary precision, GMDH models are used.

- GMDH is a strong and appealing method that is the cornerstone of many types of polynomial neural networks with a significant improvement in prediction errors when compared to the traditional MLP structure with regression analysis.
- The algorithm employed in this method finds interrelationships in data automatically, resulting in an optimal structure/network and increasing the accuracy of existing algorithms.

2.7. Time Series Prediction Using GMDH

As compared to the traditional time series forecasting problem, the number of input nodes of nonlinear model, the GMDH is equal to the number of lagged variables ($Y_{t-1}, Y_{t-2}, Y_{t-3}, \dots, Y_{t-p}$), where p is the number of chosen lagged. The outputs, y_t , the predicted value of a time series defined as

$$y_t = f(y_{t-1}, y_{t-2}, y_{t-3}, \dots, y_{t-p}) \tag{2}$$

However, no systematic method for order to determine the number of lagged p has been proposed. The number of lags p

is computed or using traditional Box Jenkins approaches. The most significant variables to employ as input nodes in the GMDH model's input layer are the lagged variables acquired from the Box-Jenkins analysis. In our study, a time series model is considered as nonlinear function of several past observations and random errors as follows:

$$y_t = f[(y_{t-1}, y_{t-2}, y_{t-3}, \dots, y_{t-p}), (\dots)] \tag{3}$$

Where f is a nonlinear function determined by the GMDH [15, 16, 17, 18, 19].

3. Results

For Chikungunya disease, we have considered the period from (January 2006- December 2015) as training set and from (January 2006-December 2019) as testing data. The following Figure 1 shows the time series plot of the chikungunya disease for a period from 2006-2019. In order to become the series into stationary, we adopted logarithm to the actual data sets. Figure 2 shows the Autocorrelation plot for different Lag period for Chikungunya data and Figure 3 shows the residual plot for the same data.

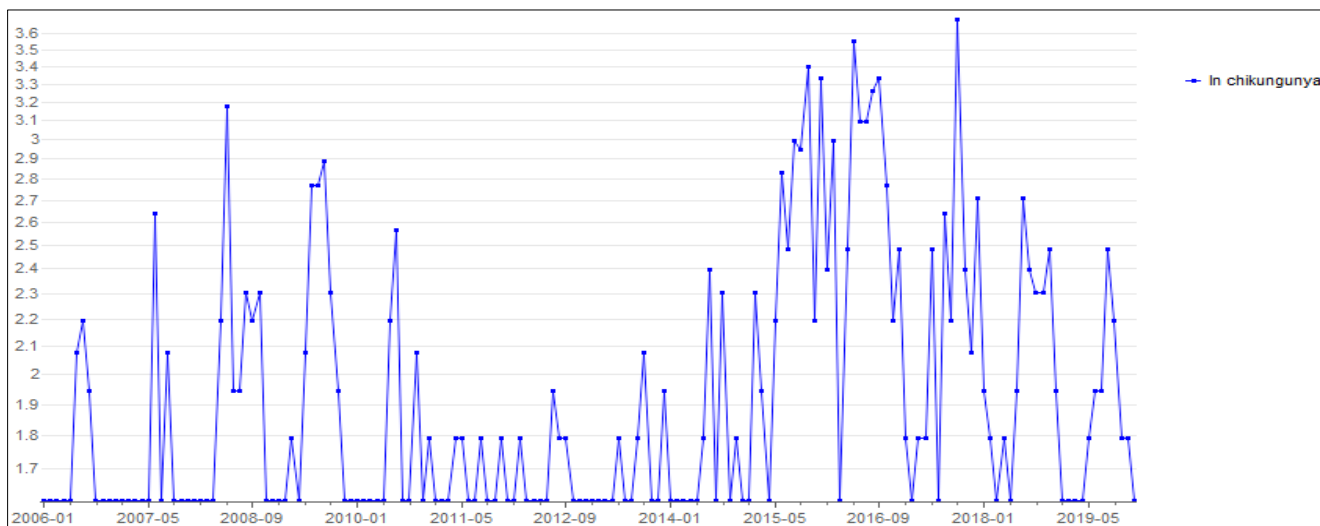


Fig 1: Plot of the Chikungunya disease for a period from 2006-2019

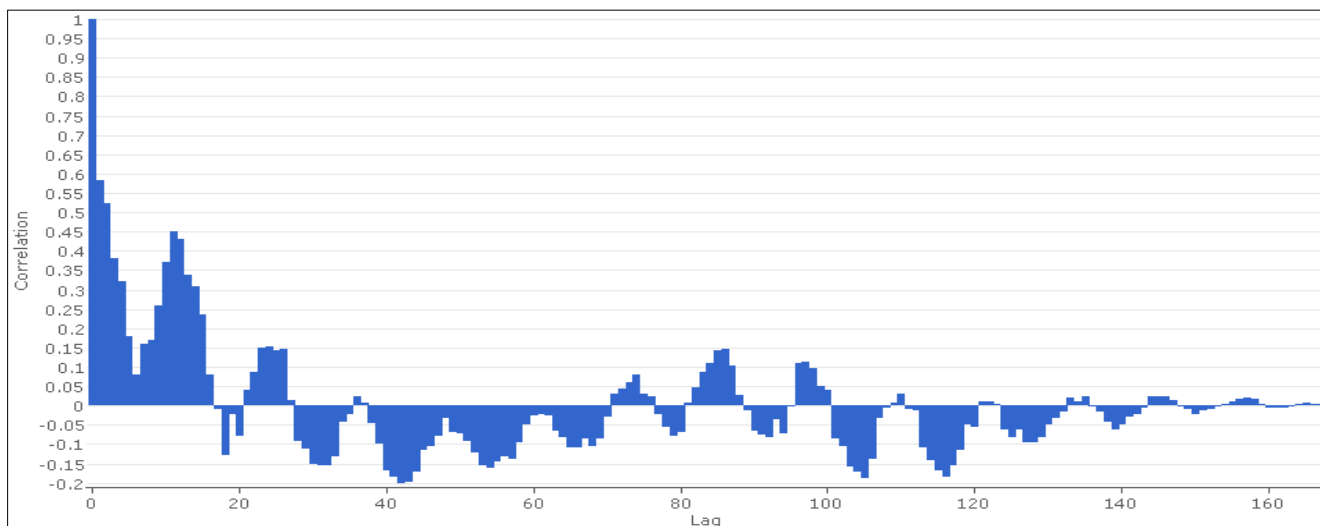


Fig 2: Autocorrelation plot for different Lag period for Chikungunya incidences.

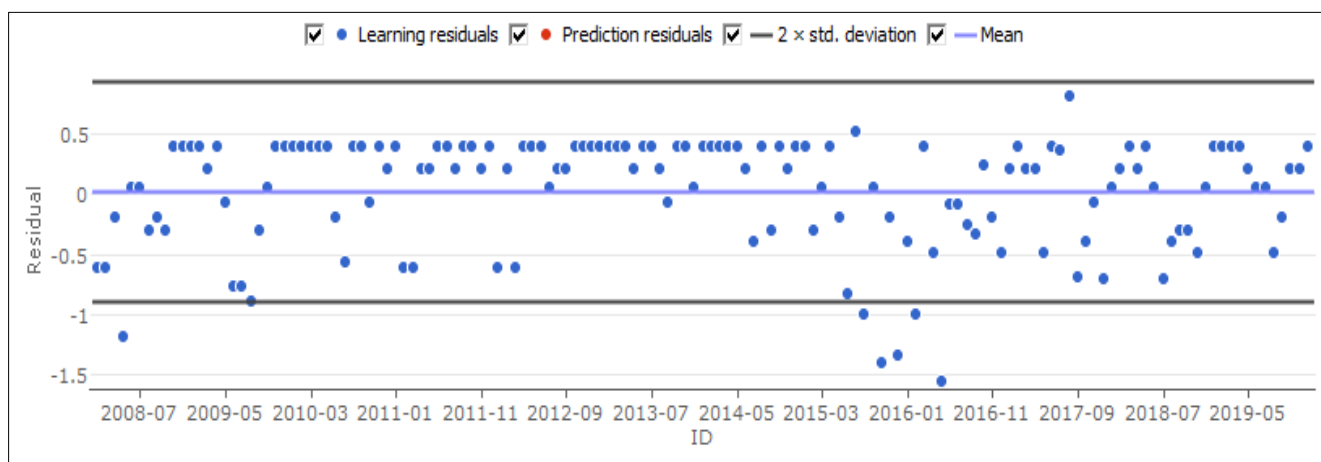


Fig 3: Residual Plot for the Chikungunya incidences

Table 1: Comparison of GMDH models based stepwise forward and mixed selection and GMDH neural networks, Average sMAPE and normalized residual sum.

	Chikungunya			
	5/7 LP(120)	5/7 FP(120)	7/7 LP(168)	7/7 FP(168)
Combinatorial (quick)	0.683205	0.683205	0.371888	0.765305
Stepwise forward selection	0.683205	0.683205	0.430957	0.765305
Stepwise mixed selection	0.683205	0.683205	0.344098	0.765305
GMDH neural network	0.680864	0.683205	0.344098	0.765305
Average sMAPE	27.15%	23.95%	15.24%	26.24%
Normalized residual sum	-18.09%	-14.94%	2.08%	-21.74%

From table 1, the results from different models are tabulated. Four models were considered, 5/7 with less parameter, 5/7 with full parameters (all significant parameters at 5% and 1% level of significance, 7/7 with less parameter and 7/7 with full parameters. From the table we notice that GMDH neural

network values for four models along with normalized residual sum. Here we notice that normalized residual sum is less i.e., -21.74% for the 7/7 with all the significant parameters at 5% and 1% level of significance.

Table 2: RMSE, Average sMAPE, MAE values of fitted GMDH Neural network models for Chikungunya disease.

S.No	Variable	Model	RMSE	Average sMAPE	MAE	R ²
1	5/7(LP)	GMDH	0.333598	27.15%	0.311191	0.404833
2	5/7(FP)	GMDH	0.44935	23.95%	0.384959	-0.0798
3	7/7(LP)	GMDH	0.46091	15.24%	0.385689	0.190291
4	7/7(FP)	GMDH	0.45498	26.24%	0.378107	0.20589

From the table no 2, for Chikungunya disease, 5/7(LP) model has smallest RMSE, highest R² value when compared with others models. Here we have considered Minimum temperature, Maximum rain, Average sunshine and RH mean

and RH minimum as the predictors for the 5/7LP model. Following figure 4 shows the forecasting results for chikungunya diseases for the year 2021.

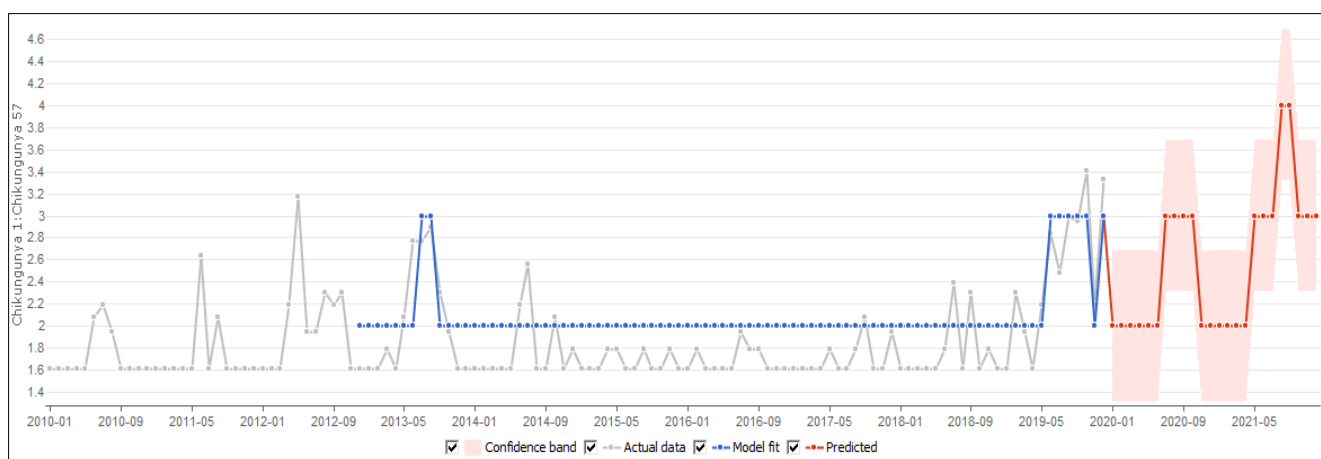


Fig 4: Forecasting Chikungunya cases using GMDH for the year 2021.

In table 3, correlation analysis was performed between Observed and forecasted Chikungunya cases.

Table 3: Correlation between Observed and forecasted Chikungunya Cases

Correlations			
		Observed	Forecast
Observed	Pearson Correlation	1	.457**
	Sig. (2-tailed)		.000
	N	135	135
Forecast	Pearson Correlation	.457**	1
	Sig. (2-tailed)	.000	
	N	135	135

From the above table, Correlation is significant ($p < 0.0001$) and Q-Q plot of predicted values with residual values were plotted. Here R^2 value is 0.208

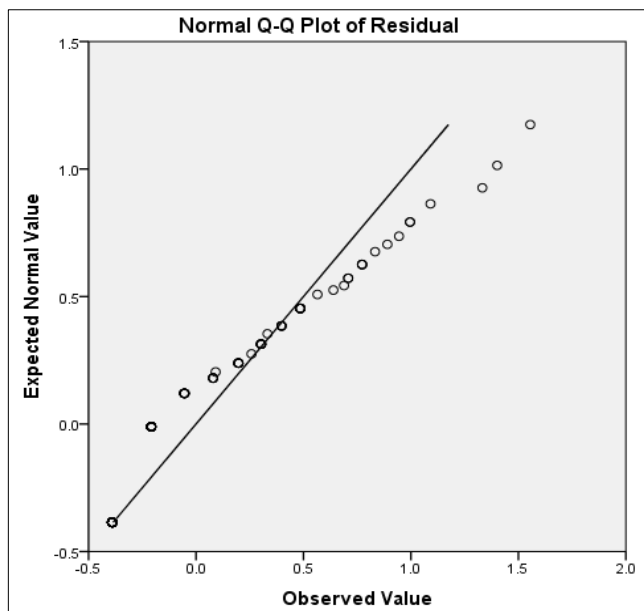


Fig 5: Normal Q-Q plot of predicted values Vs Residual Values

4. Discussion and Conclusion

GMDH shell is used as predictive analytics. It is based on the classical GMDH algorithm and used to solve clustering, classification and time series prediction problems. GMDH shell uses the data mining algorithm which is self-organized neural networks. For our study, we adopted GMDH models in order to check the interrelationship in the data which will provide an optimal structure and increases the accuracy of the existing algorithms by framing a network [18, 19]. GMDH models showed that for chikungunya cases, the RMSE value was low for 5/7 with less parameter. Here we have considered the parameters such as the Minimum temperature, mean relative humidity and average sunshine as the significant predictors in predicting and forecasting the chikungunya incidences.

Thus based on our study, we conclude that GMDH can successfully replace neural networks in identification and forecasting tasks [20]. GMDH has a viable competitor in the modeling of real-world systems attributed to her ability to overcome the bulk of neural network defects.

5. References

- Bloom DE, Cadarette D. Infectious disease threats in the twenty-first century: Strengthening the global response. *Frontiers in immunology*. 2019 Mar 28;10:549.
- Ellner SP, Bailey BA, Bobashev GV, Gallant AR, Grenfell BT, Nychka DW. Noise and nonlinearity in measles epidemics: combining mechanistic and statistical approaches to population modeling. *The American Naturalist*. 1998 May;151(5):425-40.
- Van Bortel W, Dorleans F, Rosine J, Blateau A, Rousset D, Matheus S, *et al.* Chikungunya outbreak in the Caribbean region, December 2013 to March 2014, and the significance for Europe. *Euro surveillance*. 2014 Apr 3;19(13):20759.
- Morrison TE. Reemergence of chikungunya virus. *Journal of virology*. 2014 Oct 15;88(20):11644-7.
- Trampuz A, Jereb M, Muzlovic I, Prabhu RM. Clinical review: Severe malaria. *Critical care*. 2003 Aug;7(4):1-9.
- Tsatsarkin KA, Chen R, Sherman MB, Weaver SC. Chikungunya virus: Evolution and genetic determinants of emergence. *Current opinion in virology*. 2011 Oct 1;1(4):310-7.
- Vega-Rúa A, Zouache K, Girod R, Failloux AB, Lourenço-de-Oliveira R. High level of vector competence of *Aedes aegypti* and *Aedes albopictus* from ten American countries as a crucial factor in the spread of Chikungunya virus. *Journal of virology*. 2014 Jun 1;88(11):6294-306.
- Cherry CC, Beer KD, Fulton C, Wong D, Buttke D, Staples JE, Ellis EM. Knowledge and use of prevention measures for chikungunya virus among visitors-Virgin Islands National Park, 2015. *Travel medicine and infectious disease*. 2016 Sep 1;14(5):475-80.
- Medlock JM, Hansford KM, Schaffner F, Versteirt V, Hendrickx G, Zeller H, *et al.* A review of the invasive mosquitoes in Europe: ecology, public health risks, and control options. *Vector-borne and zoonotic diseases*. 2012 Jun 1;12(6):435-47.
- Alvares CA, Stape JL, Sentelhas PC, Gonçalves JD, Sparovek G. Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift*. 2013 Dec 1;22(6):711-28.
- Horsthemke W. Noise induced transitions. In *Non-equilibrium dynamics in chemical systems*. Springer, Berlin, Heidelberg, 1984, 150-160.
- Ivakhnenko AG. The group method of data of handling; a rival of the method of stochastic approximation. *Soviet Automatic Control*. 1968;13:43-55.
- Muller JA, Ivakhnenko AG. Self-organizing modelling in analysis and prediction of stock market. In *Proceedings of the Second International Conference on Application of Fuzzy Systems and Soft Computing-ICAFS*. 1996;96:491-500.
- Ivakhnenko AG. The group method of data of handling; a rival of the method of stochastic approximation. *Soviet*

- Automatic Control. 1968;13:43-55.
15. Ivakhnenko AG. Sorting methods for modeling and clusterization (Survey of GMDH papers for the years 1983-1988) the present stage of GMDH development. Sov. J Autom. Inf. Sci. 1988;21(4):1-3.
 16. Ivakhnenko AG. Polynomial theory of complex systems. IEEE transactions on Systems, Man, and Cybernetics. 1971 Oct;4:364-78.
 17. Madala HR, Ivakhnenko AG. Inductive learning algorithms for complex systems modeling. CRC press, 2019, Aug 8.
 18. Stepashko VS, Yurachkovskiy YP. The present state of the theory of the group method of data handling. Sov. J Autom. Inf. Sci. 1986(4):36-43.
 19. Yurachkovsky YP. Restoration of polynomial dependencies using self-organization. Soviet Automatic Control. 1981;14:17-22.
 20. Ivakhnenko A, Krotov G. Modernized gmdh algorithms for identification and 2-level long-term prediction of the baikal ecosystem. Avtomatika. 1983 Jan;1(2):11-7.