MATHEMATICAL MODELING AND COMPUTING, Vol. 12, No. 1, pp. 101–107 (2025)

Impact of information on solar flares and earthquakes on the prediction of the annual dynamics of the infrasound wave envelope

Soroka O., Melekh B., Karnaushenko V., Chornodolskyy Ya.

Ivan Franko National University of Lviv,
1 Universytetska Str., 79000, Lviv, Ukraine

(Received 27 October 2024; Accepted 27 January 2025)

The research results on the effectiveness of using data on solar flares and earthquakes to predict the infrasound wave envelope are presented. The resulting SARIMAX model, enhanced with the aforementioned external factors, exhibits a 30% reduction in mean squared error and a 29% increase in the coefficient of determination compared to the previously presented ARIMA model. Additionally, a significant achievement of the new approach, compared to previous ones, is the successful reproduction of the sharp intensity drop in the envelope during the August–September–October period. The proposed approach significantly improves the detection process of aperiodic planetary-scale events, thereby enhancing its practical value.

Keywords: *Earth's atmosphere; space physics; infrasound; solar activity; earthquakes; modeling; time series.*

2010 MSC: 85A20, 62M10, 62M20, 62P35 **DOI:** 10.23939/mmc2025.01.101

1. Introduction

The study of physical processes in the Earth's atmosphere offers significant opportunities for the identification, localization and prediction of large-scale events in the Earth's crust and space. Atmospheric infrasound is one such process that is widely used in practice and has gained popularity due to its high sensitivity and ability to propagate responses to disturbances over long distances [1]. For example, in 1996, the Conference on Disarmament in Geneva initiated the International Monitoring System (IMS), which monitors atmospheric infrasound, along with seismic, hydroacoustic, and radionuclide components [2].

Infrasound analysis is conducted using specialized measuring equipment deployed around the globe. Despite the ongoing evolution of modern measurement tools, the cost of their production and the imperfections of sensors remain significant obstacles to progress. Therefore, it is important to explore new approaches for analyzing already collected data to enhance the efficiency of existing systems. In particular, machine learning techniques can significantly increase the amount of useful information that can be extracted from atmospheric infrasound measurements, thereby expanding its practical applications [3].

In a previous study [4], the predictability of the annual dynamics of infrasound was confirmed, making it possible to further identify individual random events that cause corresponding acoustic disturbances and cannot be explained by the proposed statistical model. Improving the accuracy of the model remains relevant, as the predictions obtained in the study [4] do not reproduce the actual measurements, particularly in months with sharp value changes. This can be critical for identifying individual random planetary-scale events, as they can provoke significant spikes in atmospheric infrasound. To address this issue, it was decided to expand the model's dimensionality by incorporating data on earthquakes and solar flares. The quantitative comparison of the model's explanatory power regarding infrasound dynamics variations was conducted using the coefficient of determination, also

This work was supported by grant ID: 0123U101836 provided by the Ministry of Education and Science of Ukraine.

known as the R-squared statistic. The development of such a model enables the assessment of the impact of including earthquake and solar flare data on infrasound prediction and allows for more accurate identification of aperiodic events caused by processes of a different nature [5].

2. Analysis of experimental measurements

2.1. Solar flares

A solar flare is a short-term, sudden increase in radiation intensity emitted in the area of sunspots. For many years, solar flares have been best tracked in the H-alpha wavelength in the chromosphere, although white-light solar flares are sometimes observed in the photosphere. Currently, solar flares are detected via satellite monitoring of the Sun in the X-ray spectrum. Flares are characterized by a rise time (on the order of minutes) and a decay time (on the order of tens of minutes). The total energy released in a typical flare is about 10^{30} ergs; the magnetic field is extremely high, ranging from 100 to 10 000 gauss. Optical H-alpha flares are usually accompanied by bursts of radio and X-ray emissions. Changes in solar radiation characteristics during solar flares lead to the generation of infrasound waves in the atmosphere by causing atmospheric heating and reorganizing chains of atmospheric chemical reactions [1].

Solar flares are characterized by several parameters, including brightness and size. The National Geophysical Data Center (NGDC) [6] maintains archives for approximately 80 stations, covering the period from 1938 to the present. Currently, five stations regularly send their data monthly to the NGDC in Boulder. Flare reports have been processed and published in the Quarterly Bulletin on Solar Activity [7].

The study utilized a dataset consisting of the "Comprehensive Flare Index" (1955–2008) and the "Kandilli Flare Index" (1976–2010). The Comprehensive Flare Index was developed by Dodson and Hedeman from the McMath-Hulbert Solar Observatory [8–10]. The term "flare index" was introduced by Kleczek [11], who derived a measure for quantitatively determining daily (24 hours) flare activity:

$$Q = i \cdot t. \tag{1}$$

Here i represents the intensity scale or optical importance of the flare, and t represents its duration in minutes. The researcher assumed that this dependence approximately reflects the total energy emitted by the flares. Daily flare indices for the 21st, 22nd, and 23rd solar cycles were determined using compiled groups of solar flares prepared by the National Oceanic and Atmospheric Administration's National Geophysical Data Center [2].

Before use, the solar activity index data were normalized relative to the amplitude of the infrasound wave envelope. The result is presented in Figure 1 (red line). It is also worth noting the intense peak of solar activity in July–August 2000, corresponding to one of the most well-known solar storms known as Bastille Day [12].

2.2. Earthquakes

For many years, researchers have been seeking mechanisms and characteristics that reflect the preparatory processes of earthquakes in various geophysical fields [13, 14]. It is known that before an earthquake occurs, infrasound with energies up to $10^6 - 10^{10}$ Joules can be generated by the movement of lithospheric plates. This leads to atmospheric heating by infrasound and changes in variations of geomagnetic and geoelectric fields, as well as anomalous manifestations of nighttime radiation in the middle and upper atmosphere [15]. Therefore, atmospheric infrasound can serve as a precursor to earthquakes.

In particular, in the study [15], it was found that several days before a major earthquake, there are significant changes in the daily spectrum and daily rhythm of infrasound, and the use of infrasound data for forecasting anomalous seismic events was proposed.

For the analysis of earthquake impacts in this study, a dataset from the Significant Earthquake Database [16] was used, containing information on more than 5700 earthquakes from 2150 BC to the present. To be included in this dataset, an earthquake must meet one of three criteria: it resulted in

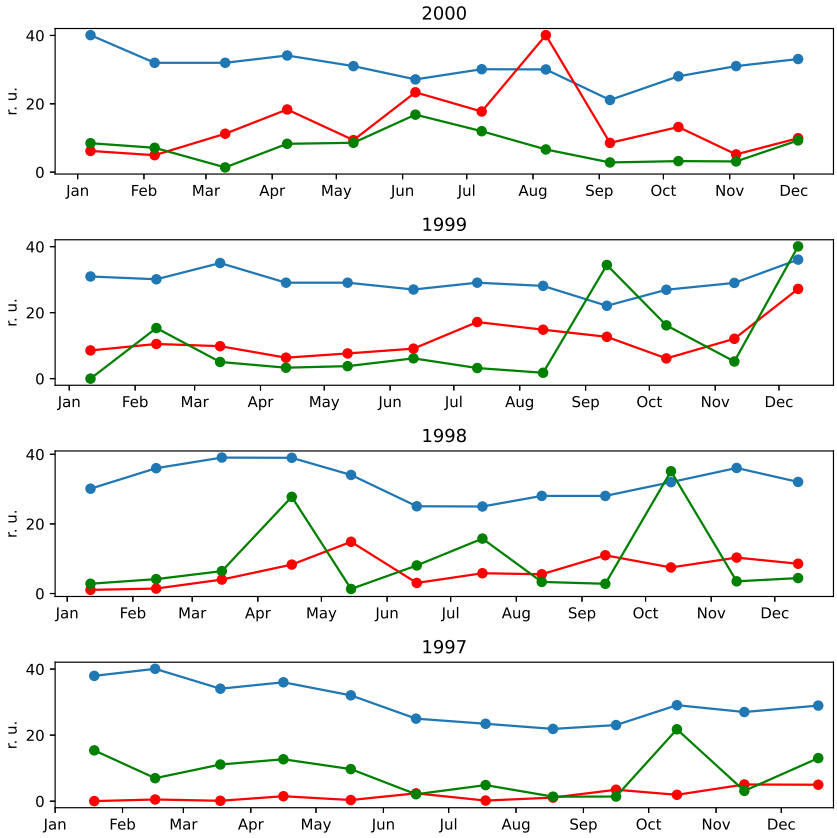

Fig. 1. The blue dashed line represents the monthly measurements of infrasound envelope from 1997 to 2000 [4]; the red line represents the solar flare index [8–10]; the green line represents the estimated earthquake amplitude [16] at the location of the infrasound observations. All values are normalized relative to the infrasound envelope.

human fatalities, caused damages estimated at more than 1 million, had a magnitude of 7.5 or higher, or generated a tsunami. The dataset includes information on the date and time of the event, geographic latitude, longitude, and focal depth of the earthquake source, as well as its magnitude. Additionally, socio-economic data on the number of deaths, the number of destroyed or damaged buildings, and the extent of damages are available but were not utilized in this study.

To assess the intensity of an earthquake, the concept of magnitude is used – a relative measure that characterizes the amount of energy released at the earthquake's epicentre and is proportional to the logarithm of the maximum amplitude of ground particle displacement recorded by seismic stations [17]. Since the data on infrasound intensity were obtained at the Western Regional Center of Special Monitoring of the National Space Agency of Ukraine, while information about earthquakes pertains to the entire world, a model is needed to assess the earthquake's impact on a point with specified geographic coordinates.

Since the calculation concerns global scales, it is reasonable to consider the earthquake source as a point disturbance source and the corresponding generated wave as spherical. The intensity of such a spherical wave at a given point will be inversely proportional to the square of the distance from its source:

$$I = \frac{M_W}{r^2}. \tag{2}$$

In formula (2), I represents the estimated intensity of the magnitude at the specified point on the planet, M_W is the earthquake magnitude, and r is the geodetic distance between the specified point and the earthquake source. After estimating the intensity of the magnitude at the specified point, it was scaled relative to the magnitude of the infrasound. The results are presented in Figure 1 (green line).

2.3. Infrasound envelope

As previously noted, experimental measurements of infrasound [1] in the frequency range of 0.003 – 0.2 Hz were conducted at the Western Regional Center of Special Monitoring of the National Space Agency of Ukraine during the years 1997–2000. Since the typical interval of infrasound oscillation periods ranges from a few seconds to several days [1], it is appropriate to characterize changes in the amplitude of infrasound oscillations as a modulated random process [18, 19]. The corresponding data are presented in Figure 1 (blue line). In the previous study [4], by analyzing the Fourier spectrum and autocorrelation function of the infrasound envelope, the harmonic nature of the annual dynamics of the infrasound envelope with a clearly defined 1-year period was demonstrated. It is also known that during the summer, the infrasound envelope is formed by oscillations of the first ($T = 24$ hours) and second ($T = 12$ hours) harmonics, whereas during the winter period, only the first harmonic is dominant. This can be explained by the fact that in the winter period, the atmospheric conditions during the day and night are similar, unlike in the summer [1].

Table 1. Correlation coefficients of the index of solar flares and earthquakes with the infrasound envelope by year.

Year	Earthquake	Solar flare index
1997	0.55	−0.41
1998	0.12	0.15
1999	−0.01	0.41
2000	0.13	−0.17

To assess the impact of including information about earthquakes and solar flares on the quality of predicting the annual dynamics of the infrasound envelope, correlation coefficients for each of the exogenous variables were calculated over the years. The obtained results are presented in Table 1. The instability of the correlation coefficient may indicate a nonlinear relationship between these natural processes.

3. Modeling methods

3.1. SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors)

The main difference between the SARIMAX model [20] and the popular ARIMA model [4] lies in the inclusion of seasonal parameters and the ability to incorporate additional exogenous dependencies. The SARIMA approach addresses seasonality by applying the ARIMA model to historical lags that are multiples of the specified seasonal parameter [21]. Exogenous parameters are considered additive components.

The SARIMAX method was chosen because it was previously established [4] that there is a clear periodicity in the infrasound with aperiodic influences from external (exogenous) factors, including earthquakes, solar flares, galactic cosmic radiation, anthropogenic, and other atmospheric disturbances [22]. In this study, a variation of the SARIMAX approach from the statsmodels library [23] of the Python programming language was used.

3.2. Coefficient of determination R-squared

In addition to the correlation coefficient and root mean square deviation, in this study, the coefficient of determination R-squared [24] was also estimated. This metric allows us to estimate the proportion of variance in the dependent variable that can be explained by the proposed statistical model. The coefficient of determination can be calculated using the following formulas:

$$R^2 = \frac{SS_{res}}{SS_{tot}}, \tag{3}$$

$$SS_{res} = \sum_i (y_i - f_i)^2, \tag{4}$$

$$SS_{tot} = \sum_i (y_i - \overline{y})^2. \tag{5}$$

Here y_i is i-th component of the input vector, f_i is i-th component of the output (predicted) vector, $\overline{y}$ is vector's mean value.

According to formula (1), the coefficient of determination can be interpreted as the explained percentage of data variance.

4. Training and forecast

To forecast annual infrasound envelope dynamics, leveraging solar flares and earthquake data, a SARIMAX model was designed with parameters (1, 1, 4) for the autoregressive (AR), differencing, and moving average (MA) orders, and (5, 0, 12) for their respective seasonal components. The model parameters have been determined by minimizing the root mean square deviation (RMSE) between the predicted and experimental measurements for three years (1998, 1999, and 2000).

Table 2 contains a comparison of the model metrics investigated in this study and the previous one [4]. The provided data of root mean square deviations show that the proposed SARIMAX model exhibits a 41% higher accuracy than the most accurate ARIMA model from the previous study [4].

Table 2. RMSE and R-squared metrics for different models.

Model	RMSE	R-squared
Autoregression	2.83 [4]	0.57
Moving average	3.24 [4]	0.43
ARIMA	2.63 [4]	0.63
SARIMAX	1.86	0.81

Comparing the coefficient of determination (R-squared) of the models allows us to assess the impact of incorporating data about earthquakes and solar flares. The obtained value of 0.81 for the SARIMAX model indicates that 81% of the variance in the dependent variable can be fully explained via training on historical data and incorporating information about earthquakes and solar flares. In contrast, the previous ARIMA model [4], which only used historical information for training, can only explain 63% of the variance of the infrasound envelope function. Two conclusions can be drawn from this. Firstly, during the year 2000, there were aperiodic or low-frequency events (frequency of less than once every three years) that were not caused by earthquakes or solar flares, which are responsible for 19% of the variance in infrasound. Secondly, the difference in the coefficient of determination between the ARIMA and SARIMAX models is 0.18, indicating that incorporating information about earthquakes and solar flares as additional exogenous predictors in the model allows for an additional explanation of 18% of the variance in infrasound.

Figure 2 illustrates the SARIMAX model prediction (red line) of the annual dynamics of the infrasound envelope for the year 2000, alongside its experimental values (green line). From the provided graph, it is evident that the model adequately predicts the dynamics of the infrasound envelope for most months. However, there are some discrepancies observed in March and December, which will be discussed further. It is worth noting the particularly successful prediction of a double sharp change in the infrasound envelope during the August–September–October period, which was not achieved with previous ARIMA models. This improvement is attributed to the inclusion of earthquakes and solar flare data as new additional dimensions of the model. Importantly, the obtained model surpasses the accuracy (root mean square error) of the ARIMA model and the correlation level of the autoregressive model, which had the best performance in the mentioned characteristic [4].

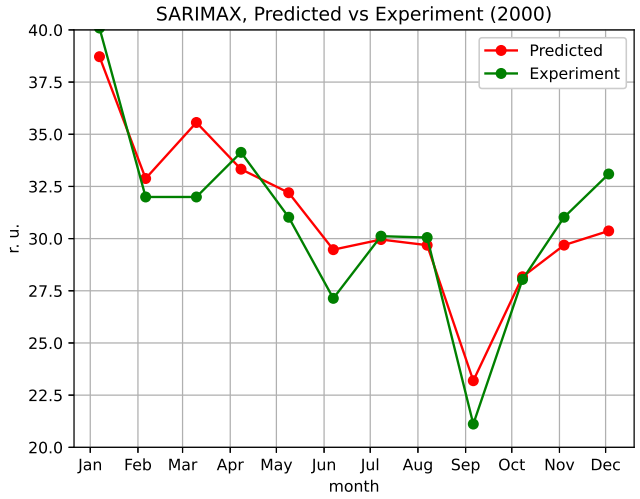

Fig. 2. Prediction of the annual dynamics of the infrasound envelope using information on solar flares and earthquakes, year 2000. The red line represents predicted values, the green line represents experimental data.

The identified discrepancies in the prediction of the dynamics of the infrasound envelope in March–April and December open up possibilities for further improvement of the model by including other factors which are capable of producing atmospheric infrasound besides the considered solar flares and earthquakes. One such additional source of atmospheric infrasound could be meteor showers, particularly the Lyrids in March–April and the Geminids in December [25]. Since the model presented

in the study does not incorporate data on meteor showers, all aperiodic events associated with this source will be unpredictable.

Another factor is the greater influence of relatively recent historical data compared to older data. From this perspective, the most significant data for training and prediction are concentrated within the year 1999. In Figure 1, it can be observed that the experimental measurements of infrasound this year had a local maximum in March, which was successfully reproduced by the model even in the absence of correlation with earthquake or solar flare data. Utilizing a broader range of data will address this issue in the future.

5. Conclusions

Using the SARIMAX model in conjunction with data on solar flares and earthquakes, the accuracy of predicting the annual dynamics of infrasound has been improved (reducing the root mean square error) by 30%. Additionally, the obtained model is characterized by a higher r-squared value, surpassing the previous ARIMA model [4] by 29%. The predicted curve of the annual dynamics of the infrasound envelope for the year 2000 accurately reproduces the sharp intensity change during the period of August–September–October, mitigating a significant deficiency of the ARIMA model [4]. These findings indirectly demonstrate the significant impact of earthquakes and solar flares on the generation of atmospheric infrasound, as previously noted in studies [1, 11].

However, it has been observed that the predicted dynamics of infrasound diverged in March and December, which were caused by an insufficient number of external influences considered by the model. Further expansion of the model with such external regressors can mitigate this issue and improve prediction accuracy. Additionally, interesting avenues for further research include using causal inference methods to investigate the influence of individual irregular factors on the generation of atmospheric infrasound.

[1] Soroka S., Mezencev V., Karatayeva L., Soroka O. Infrasound of space origin and its influence on terrestrial processes. Kosmichna nauka i texnolohiya. **14** (6), 73–88 (2008), (in Ukrainian).

[2] https://www.ctbto.org/our-work/international-monitoring-system

[3] Fields M. P., Bennett H., Scoggins R. Machine learning for source classification utilizing infrasound data. Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications III. **11746**, 117462N (2021).

[4] Soroka O. S., Melekh B. Ya., Karnaushenko V. O., Chornodolskyy Ya. M. Journal of Physical Studies. **28** (3), 3403 (2024).

[5] Albert S., Linville L. Benchmarking current and emerging approaches to infrasound signal classification. Seismological Research Letters. **91** (2A), 921–929 (2020).

[6] https://www.ngdc.noaa.gov/

[7] https://www.ngdc.noaa.gov/stp/space-weather/solar-data/solar-features/solar-radio/iau-qbsa/documentation/qbsa-history.pdf

[8] Dodson H. W., Hedeman E. R. An Experimental Comprehensive Flare Index and Its Derivation for "Major" Flares, 1955–1969, UAG-41, World Data Center A (NOAA), 25 pp. Available from the NOAA National Geophysical Data Center (1971).

[9] Dodson H. W., Hedeman E. R. Experimental Comprehensive Flare Indices for Certain Flares, 1970–1974, UAG-52, World Data Center A (NOAA), 27 pp. Available from the NOAA National Geophysical Data Center (1975).

[10] Dodson H. W., Hedeman E. R. Experimental Comprehensive Flare Indices for "Major" and Certain Lesser Flares, 1975–1979, UAG-80, World Data Center A (NOAA), 38 pp. Available from the NOAA National Geophysical Data Center (1981).

[11] Kleczek J. Publ. Inst. Centr. Astron., No. 22, Prague (1952).

[12] Andrews M. D. Lasco and eit Observations of the Bastille day 2000 Solar Storm. Solar Physics. **204**, 179–196 (2001).

[13] Rikitake T. Earthquake prediction. Earth-Science Reviews. **4**, 245–282 (1968).

[14] Kanamori H. Earthquake prediction: An overview. International Geophysics. **81** (B), 1205–1216 (2003).

[15] Soroka S. O., Mezentsev V. P., Karataieva L. M. Zviazok seismichnoi aktyvnosti z atmosfernym infrazvukom. Vidbir i obrob. informatsii: Mizhvid. zb. nauk. pr. **32** (108), 61–66 (2010), (in Ukrainian).

[16] National Geophysical Data Center / World Data Service (NGDC/WDS): NCEI/WDS Global Significant Earthquake Database. NOAA National Centers for Environmental Information.

[17] Mala hirnycha entsyklopediia u 3 t. Za red. V. S. Biletskoho. **2**, 670 s. (2007), (in Ukrainian).

[18] Arce G. R. Nonlinear Signal Processing: A Statistical Approach. John Wiley & Sons (2004).

[19] Wu R.-S., Luo J., Wu B. Seismic envelope inversion and modulation signal model. Geophysics. **79** (3), WA13–WA24 (2014).

[20] Vagropoulos S. I., Chouliaras G. I., Kardakos E. G., Simoglou C. K., Bakirtzis A. G. Comparison of SARIMAX, SARIMA, modified SARIMA and ANN-based models for short-term PV generation forecasting. 2016 IEEE international energy conference (ENERGYCON). 1–6 (2016).

[21] Chatfield C. The Analysis of Time Series: An Introduction. Chapman & Hall/CRC, Boca Raton, Fla (2004).

[22] Negoda A. A., Soroka S. A. Prospects in the investigations of the atmosphere and ionosphere with the use of artificial acoustic influence. Space Science and Technology. **5** (2/3), 3–12 (1999).

[23] Seabold S., Perktold J. Statsmodels: Econometric and Statistical Modeling with Python. Proceedings of the 9th Python in Science Conference. 92–96 (2010).

[24] Miles J. R-squared, adjusted R-squared. Encyclopedia of Statistics in Behavioral Science. John Wiley & Sons, Ltd (2005).

[25] Mezentsev V. P., Soroka O. S., Soroka S. O. Infrazvukova reaktsii atmosfery na meteorni potoky. Vidbir i obrobka informatsii. **28** (104), 61–66 (2008), (in Ukrainian).

Вплив відомостей про сонячні спалахи та землетруси для передбачення річної динаміки огинаючої інфразвукової хвилі

Сорока О., Мелех Б., Карнаушенко В., Чорнодольський Я.

Львівський національний університет імені Івана Франка,
вул. Кирила і Мефодія, 8, 79005 Львів, Україна

Представлено результати дослідження ефективності використання даних про сонячні спалахи та землетруси для передбачення огинаючої інфразвуку. Отримана модель SARIMAX, розширена з-за допомогою вищезгаданих сторонніх чинників, характеризується на 30% меншим середньо-квадратичним відхиленням та на 29% більшим коефіцієнтом детермінації, у порівнянні з раніше представленою моделлю ARIMA. Також, вагомим досягненням нового підходу, у порівнянні з попередніми, є успішне відтворення різкого перепаду інтенсивності огинаючої у період серпень–вересень–жовтень. Запропонований в роботі підхід дозволяє значно покращити процес виявлення аперіодичних подій планетарного масштабу, тим самим підвищуючи його практичну цінність.

Ключові слова: *атмосфера Землі; фізика космосу; інфразвук; сонячна активність; землетруси; моделювання; часові ряди.*