University of Tehran Press

Online ISSN: 2345-475X

DESERT

Home page: https://jdesert.ut.ac.ir/

Proposing an Intelligent Hybrid Algorithm: Group Method of Data Handling – Harmony Search for Dust Storm Modeling in Western Iran

Ramtin Tavoosi Rad[1], Mohammad Ansari Ghojghar[1*], Hassan Khosravi[1]

[1] Department of Reclamation of Arid and Mountainous regions Engineering, Faculty of Natural Resources, University of Tehran, Karaj, Alborz, Iran. Email: ansari.ghojghar@ut.ac.ir

Article Info.

Article type:
Research Article

Article history:
Received: 22 Apr. 2026
Received in revised from: 13 May 2026
Accepted: 18 May 2026
Published online: 27 May 2026

Keywords:
Forecasting,
Horizontal visibility,
Harmony algorithm,
Data group processing.

ABSTRACT

This study addresses this imperative by proposing an intelligent hybrid algorithm, integrating Group Method of Data Handling (GMDH) with Harmony Search (HS), to offer a novel, scalable, and cost-effective framework. The objective is to significantly enhance dust storm prediction accuracy by effectively capturing intricate environmental interactions, thereby critically advancing the state-of-the-art in environmental hazard forecasting. This study aims to model the frequency index of dust storm days (FDSD) using integrated Group Method of Data Handling - Harmony Search (GMDH-HS) intelligent algorithm at eight synoptic stations in Kurdistan Province, Iran over a 50-year statistical period (1971–2020). The performance of two intelligent models, Group Method of Data Handling (GMDH) and Group Method of Data Handling - Harmony Search (GMDH-HS), was evaluated and compared. The results, based on goodness-of-fit criteria during the training and testing phases, indicated a marginal but statistically insignificant difference in accuracy between the two models. The Hybrid model of Data Handling - Harmony Search (GMDH-HS) model exhibited superior performance, achieving lower error values of 0.113, 1.231 percent respectively for NRMSE and MAPE metrics. The selection of a suitable model should be guided by the specific characteristics of the project, the dataset, and existing constraints. By leveraging the strengths of both components—GMDH's ability to automatically select the optimal network structure and HS's efficiency in fine-tuning the model parameters—this hybrid framework offers a state-of-the-art, scalable, and computationally efficient predictive tool, thereby critically advancing the accuracy and efficiency of environmental hazard forecasting models.

Cite this article: Tavoosi Rad, R., Ansari Ghojghar, M., Khosravi, H. (2026). Proposing an Intelligent Hybrid Algorithm: Group Method of Data Handling – Harmony Search for Dust Storm Modeling in Western Iran. DESERT, 31 (1), DOI: 10.22059/jdesert.2026.107638

DOI: 10.22059/jdesert.2026.107638

Publisher: University of Tehran Press

1. Introduction

Urbanization, industrialization, and climate change have made dust storms one of the main sources of air pollution globally (Manisalidis *et al*., 2020). This phenomenon usually occurs in arid and semi-arid regions with loose soils or surfaces without vegetation cover (Du *et al*., 2025). According to the World Meteorological Organization (WMO), a dust storm is a phenomenon that transports a large volume of dry soil particles or other fine material from the Earth's surface into the atmosphere and reduces horizontal visibility to less than 1,000 meters (Midelton & Kang, 2017). One of the well - known types of this phenomenon is the cyclone storm which occurs in arid and semi - arid regions and is often accompanied by thunderstorms. This storm is characterized by a wall of dust and sand that moves rapidly, reducing horizontal visibility to a few meters or less (Golkar *et al*., 2019). In addition to strong winds, vertical air movement is also necessary to transport dust particles over long distances. Dust storms can occur in a variety of sizes and intensities, from local, short-lived events to large, long-lasting storms that affect large areas (Liang *et al*., 2020; Wang *et al*., 2022). This phenomenon is associated with hot and dry weather conditions, low humidity, and atmospheric instability, significantly contributing to air pollution and increasing health risks (Han *et al*., 2022). In addition to severely reducing visibility, dust storms can cause damage to roads and airport runways, significantly disrupting transportation activities (Miri & Middleton, 2022). These storms also increase the concentration of bioaerosols in the air, which poses a serious threat to human health, especially those suffering from respiratory problems (Miri *et al*., 2023). On the other hand, the deposition of dust particles on soil and vegetation can disrupt the balance of ecosystems and reduce agricultural productivity (Midelton & Kang, 2017; Zhang *et al*., 2021; Tang *et al*., 2023).

Nowadays, much research has been conducted to predict dust storms using various modeling approaches. For example, Yarmohammadi *et al*. (2025) focused on predicting the path of dust storms using deep learning models by employed deep learning architectures, CNN-LSTM model. The results show that the performance of both models is improved by considering textual information. Furthermore, the increase of 0.2 in the kappa coefficient measure across all forecast hours indicates that the convolutional neural network-long short-term memory network (CNN-LSTM) model performs better when considering textual information compared to the convolutional long short-term memory network model. Similarly, Al-Shammari *et al*. (2024) predicted the frequency of dust storms in Saudi Arabia using machine learning. Their findings demonstrated the superiority of temporal convolutional networks and LSTM models over traditional statistical methods in capturing complex temporal dynamics. Ansari ghojghar & Pourmohammad (2024) modeled dust storms using a triple hybrid model of a general regression neural network-support vector machine-long short-term memory (GRNN-SVM-LSTM). The results of their findings showed that the aforementioned triple hybrid metamodel has the best performance compared to the individual and dual hybrid models studied in predicting the frequency index of days with dust storms. Pourgholam Amiji *et al*. (2020) compared the performance of integrated seasonal autoregressive moving average time series models and Holt-winters in predicting dust storms in Sistan and Baluchestan province. The results showed that the Adaptive Neural Fuzzy Inference System (ANFIS) had the best performance among the other models studied. Sobhani *et al*. (2020) studied dust modeling and forecasting in western Iran. Based on the output results of the artificial neural network model based on radial basis functions in predicting dust for the future years of the studied stations, in both the average and maximum dust frequency scales, the western and southwestern stations of the studied region were more exposed to dust in the future years. Despite these studies, several gaps remain in the existing literature. First, many of the aforementioned deep learning and hybrid models, require intensive computational resources and large datasets. This issue making these models difficult to implement for real-time early warning systems. Second, there is a lack of research on models that can automatically optimize their own functional structures while maintaining low computational complexity. Most existing studies focus on fixed-architecture models that may struggle with the highly non-linear and random nature of dust storm data in specific climatic regions like Kurdistan Province. This research aims to fill these gaps by proposing a GMDH-HS hybrid intelligent algorithm. The GMDH-HS approach combines the automated structure-selection capability of GMDH with the robust optimization of the Harmony Search algorithm, providing a highly accurate yet computationally efficient solution for modeling the FDSD index.

Accurate prediction of the Frequency Index of Dust Storm Days (FDSD) remains a significant and complex environmental challenge. A review of the current literature reveals two critical limitations in existing modeling techniques: first, standalone conventional models often lack the fidelity and robust non-linearity handling capability required to accurately capture the intricate atmospheric dynamics governing dust events, resulting in inadequate predictive accuracy. Second, while some complex hybrid models have been explored, they are typically characterized by excessive computational resource demands and high operational costs, rendering them neither economically viable nor scalable for practical, regional environmental forecasting. To overcome this critical deficiency in the state-of-the-art, this study introduces, for the first time in the field of atmospheric hazard prediction, a novel and computationally efficient intelligent hybrid algorithm: the Group Method of Data Handling–Harmony Search (GMDH-HS) model. The core innovation lies in the synergistic integration: the GMDH framework is utilized for its inherent self-organizing polynomial network structure, which is exceptional at automatically determining the optimal model complexity and capturing the non-linear relationship between atmospheric predictors and the FDSD index.

Despite of significant developments in hydrological forecasting, several challenges still persist in long-term discharge series modelling. Traditional deterministic models often struggle to find the complexities, non-linear dynamics of long-term flow data, especially in regions with high climatic variability like Kurdsitan Provinse in Iran. Also while various meta-heuristic algorithms have been employed to optimize machine learning models, achieving a balance between high predictive accuracy and computational efficiency remains a significant problem. There is a lack of robust, scalable, and economically applicable digital-based solutions specifically designed for long-term streamflow forecasting in semi-arid regions like Kurdistan Province, Iran. Responding to these gaps, the present study develops a hybrid GMDH-HS model to forecast long-term FDSD data (1971–2020) in Kurdistan Province, Iran. The primary objective of this research is to establish a reliable predictive framework that optimizes model performance through meta-heuristic integration. Specifically, the study aims to implement a hybrid architecture combining Group Function Data Driven (GMDH) with Harmony Search (HS) for parameter optimization, evaluate the model's capability in capturing long-term hydrological patterns and assess the scalability and computational efficiency of the proposed approach for real-time hydrological management. The findings of this study are expected to provide a robust tool for managing dust storms and a more efficient and accurate alternative to conventional modeling techniques in the region.

2. Materials and methods

2. 1. Study area and methodology

In the present study, the application of the intelligent hybrid model of Group Management of

Data-Harmony (GMDH-HS) to predict the frequency of days with dust storms in eight meteorological stations of Kurdistan Province with a statistical period of 50 years (1971-2020) was investigated (Figure 1). According to the World Meteorological Organization definition, a day with dust storms is a day on which, in at least one of the eight SINOPS, one of the dust-related codes (06, 07, 08, 09, 30, 31, 32, 33, 34, 35 and 98) is declared in the current weather report section (Table 1). Provided that the horizontal visibility data corresponding to the declared code is recorded at less than 1000 meters (Goudi *et al*., 2014). Therefore, in this study, hourly data on horizontal visibility and codes standardized by the World Meteorological Organization for the detection of dust storms were used for all codes. A stringent, multi-step Quality Control (QC) protocol was implemented to ensure the reliability and representativeness of the time-series data, which involved subjecting all records to standard range and step tests, employing the Interquartile Range (IQR) method for objective outlier detection, and conducting internal consistency checks to verify the data's physical plausibility. Crucially, missing values were addressed systematically: short, intermittent gaps (defined as less than three consecutive hours) were filled using linear interpolation to preserve the underlying trend, while longer or critical gaps were explicitly treated as Not Available (NA) and excluded from segment-specific analyses to prevent potential bias. Prior to model development, the processed time series underwent formal statistical validation: the Run Test was utilized to formally assess the homogeneity and randomness of the data, which was accepted at a 95% confidence level, verifying the data's suitability for time-series analysis. Furthermore, an initial descriptive analysis was performed using the Decompose function in R software to effectively separate and visualize the core time-series components, while the stationarity of the training data—a prerequisite for the predictive modeling—was rigorously evaluated using the Augmented Dickey-Fuller (ADF) test.

Table 1. World Meteorological Organization codes related to wind erosion and dust phenomena (O'Loingsigh *et al*., 2014).

Code	Description
06	Massive dust storm outside weather station
07	Soil or sand lifted from the ground within the area of a weather station
08	Tornado sightings within or outside the meteorological station area during the observation hours or the past hour
09	Sand or dust storm during observation outside the station or within the last hour within the range of the weather station
30	Mild or moderate sand or dust storm with a decreasing trend over the past hour - with wind speeds of 15 m/s or more - reduced visibility to less than 1000 but more than 200 m
31	Light or moderate sand or dust storm with no change in intensity over the past hour - with wind speeds of 15 m/s or more - visibility reduced to less than 1000 but more than 200 m
32	Mild or moderate sand or dust storm with increasing intensity over the past hour - with wind speeds of 15 m/s or more - reduced visibility to less than 1000 but more than 200 m
33	Severe sand or dust storm with a decreasing trend over the past hour - wind speed of 15 m/s or more - reduced visibility to less than 200 m
34	Severe sand or dust storm with no change in intensity over the past hour - wind speed 15 m/s or more - reduced visibility to less than 200 m
35	Severe sand or dust storm with increasing intensity over the past hour - wind speed 15 m/s or more - reduced visibility to less than 200 m
98	Thunderstorm - no precipitation - accompanied by sand or dust storm

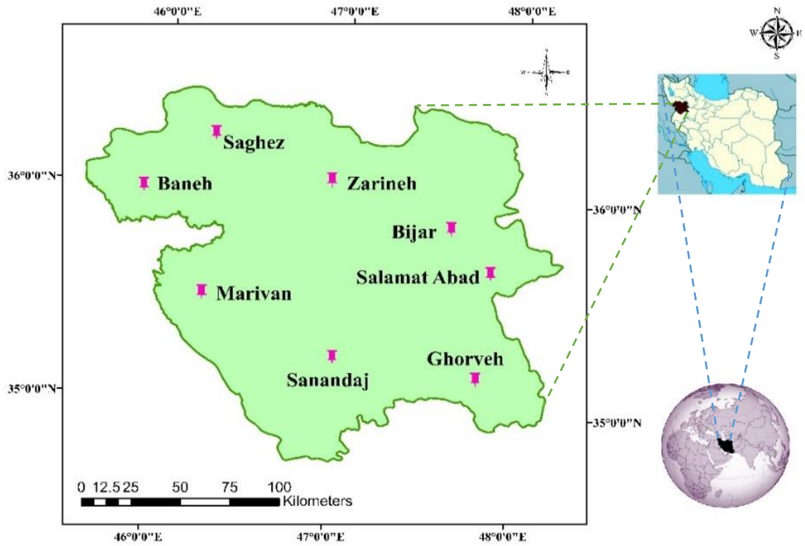

Figure 1. Introduction of the stations studied in the present study (Source: Research findings)

2.2. Harmony Search Algorithm (HSA)

This algorithm is one of the simple and innovative meta-heuristic algorithms that is inspired by putting musical notes together to create a melody with aesthetic standards, in order to search for the optimal solution. Compared to other meta-heuristic methods, it has fewer mathematical requirements and can be adapted and used in various engineering problems by changing the parameters and operators. The algorithm's working process consists of three main steps. First, a set of initial solutions called harmony memory is randomly generated (Yang, 2009). In the second stage, this memory is updated by generating new solutions. The generation of new solutions is based on three main operations: selecting from the harmony memory, fine-tuning the values, and generating new random values when needed. The new solution is then evaluated against the objective function and, if it performs better, it replaces the weakest solution in memory. This process continues until a stopping criterion is reached, such as a certain number of iterations or a minimum optimum is reached (Abualigah *et al.,* 2020).

2.3. Group Method of Data Handling (GMDH)

The data clustering method was first proposed by Ivakhnenko (1968) to create high-order polynomial equations that are capable of analyzing and evaluating complex systems. This algorithm is actually a type of self-organizing learning method that has the ability to model nonlinear systems and overcome the complexity of regression equations. Due to its ability to model complex regression systems with high degrees, it is called a type of meta-heuristic system (heuristic learning) (Olfatmiri *et al*., 2022). The progressive structure of this algorithm includes neurons with at least two inputs, and their transfer function is defined as polynomials that are fitted using the least squares method (Amiri *et al*., 2024).

2.4. Intelligent Hybrid Model of Group Method of Data Handling - Harmony Search Algorithm (GMDH-HS)

This model has been developed with the aim of combining the self-organizing modeling capabilities of the data group management algorithm and the optimization capabilities of the harmony search (HS) algorithm in order to solve complex and nonlinear problems. In this method, the Group Management Data algorithm, as an initial step, creates the model structure and, through the analysis of input and output data, extracts nonlinear relationships in the form of successive layers based on polynomial functions. Then, by evaluating and eliminating low-quality models at each stage, a simpler and more efficient model is presented for more accurate prediction (Mahdavi-Meymand *et al*., 2021). The self-organizing feature in the GMDH algorithm allows for automatic adjustment of the model structure and reduction of its complexity, while the harmony search algorithm uses harmony memory, fine-tuning of values, and generation of random values to search for the most optimal combination of parameters to increase the accuracy of the model. This process is completed by repeatedly evaluating the solutions based on prediction error criteria, and finally, the model with the highest accuracy and lowest error is selected as the output (Figure 2). Thus, the Intelligent Hybrid Model of Group Method of Data Handling - Harmony Search Algorithm (GMDH-HS) has remarkable flexibility and efficiency in solving complex problems, in addition to its effective learning capability (Khajeh *et al*., 2024).

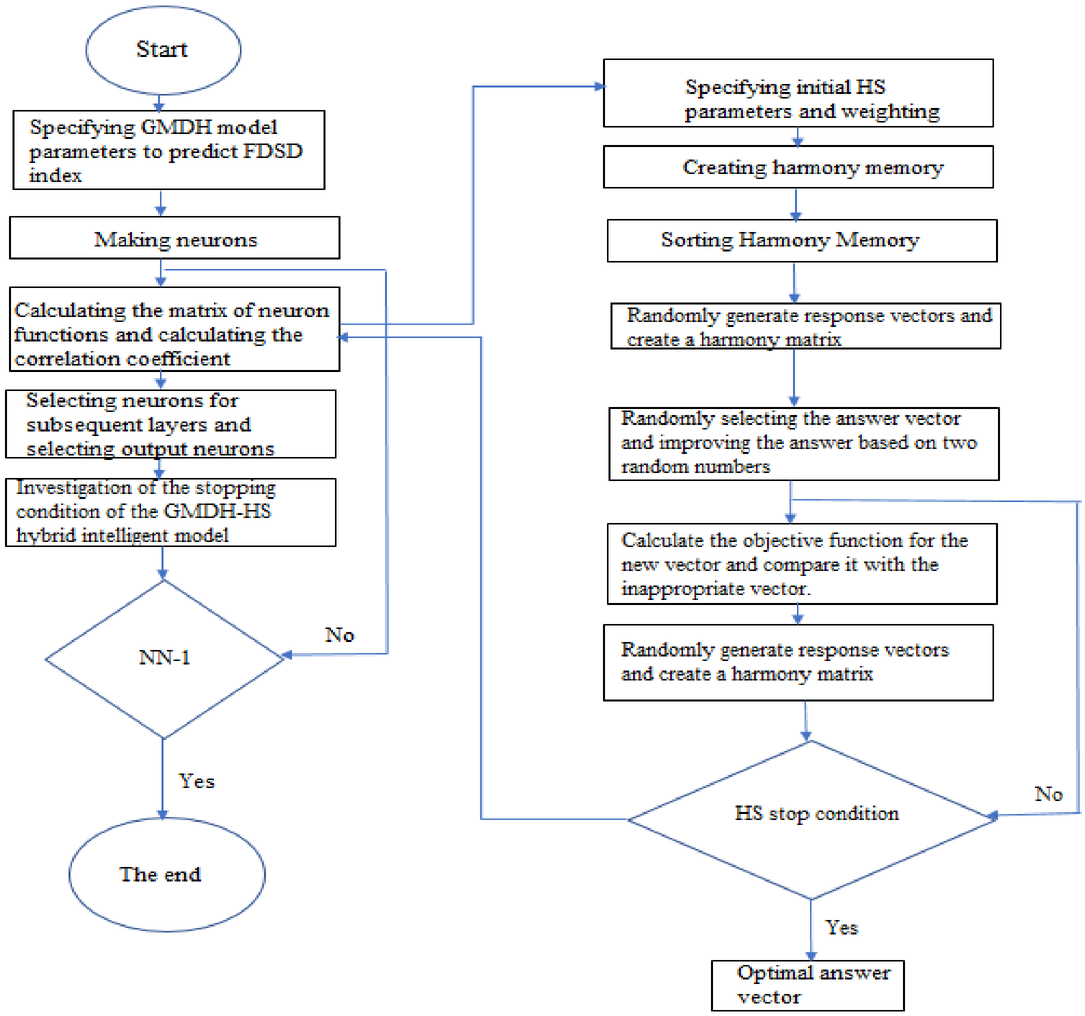

Figure 2. Flowchart of the intelligent algorithm for the integrated group data management - Harmony GMDH-HS (Source: Research findings)

2.5. Forecasting models

In this study, the variable of interest is the number of past seasons or seasons in order to predict the future season or seasons. The number of predictable seasons can include one, two, three, and four seasons prior to the index of the frequency of days with dust storms. In model number one, in order to predict dust storms for the next season, a time series of the frequency of days with dust storms with one lag step is used. In other words, in order to predict at time t+1, its value at time t is used. In model number two, the prediction for the next season is made using the values of the frequency of days with dust storms for the previous two seasons. In models 3 and 4, the predictions are made based on the FDSD values in the previous three and four seasons. The following relationships represent the predictions made in models 1 to 4. Also, the models used in selecting the previous season or seasons for forecasting in future seasons are shown in Figure 5.

Model 1 $FDSD_{(t+1)} = f\left(FDSD_{(t)}\right)$

Model 2 $FDSD_{(t+1)} = f\left(FDSD_{(t)}.FDSD_{(t-1)}\right)$

Model 3 $FDSD_{(t+1)} = f\left(FDSD_{(t)}.FDSD_{(t-1)}.FDSD_{(t-2)}\right)$

Model 4 $FDSD_{(t+1)} = f\left(FDSD_{(t)}.FDSD_{(t-1)}.FDSD_{(t-2)}.FDSD_{(t-3)}\right)$

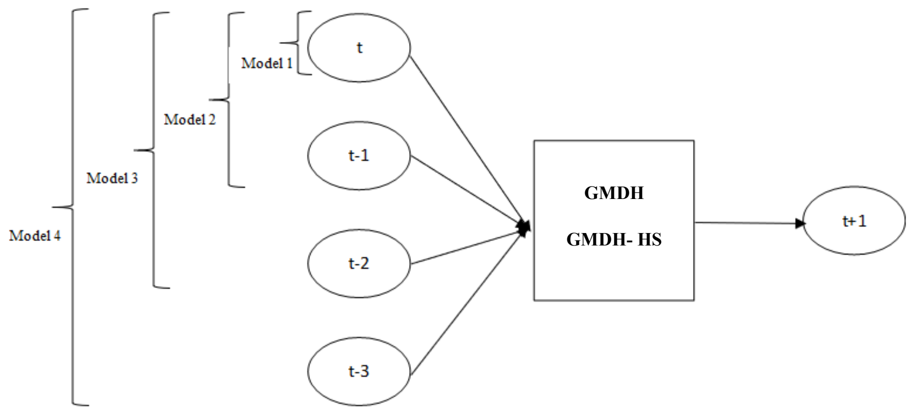

Figure 3. Structures used in dust forecasting (Ansari ghojghar *et al*., 2022)

2.6. Evaluation criteria

In this study, to evaluate the accuracy and efficiency of the models, the correlation coefficient (R), normalized root mean square error (NRMSE), mean absolute percentage error (MAPE), and Nash-Sutcliffe coefficient (NS) were used based on equations 1 to 4.

$$NRMSE = \frac{\sqrt{\frac{\sum_{i=1}^{n}(f_i - O_i)^2}{N}}}{O_i} \quad (1)$$

$$MAPE = \frac{100}{n} \times \sum_{1}^{n} \left|\frac{f_i - O_i}{f_i}\right| \quad (2)$$

$$NS = \frac{1}{n}\sum_{i=1}^{n}|O_i - f_i| \tag{3}$$

$$R = \frac{\frac{1}{n}\sum_{i=1}^{n}(O_i - \bar{O})(f_i - \bar{f})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - \bar{O})^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(f_i - \bar{f})^2}} \tag{4}$$

In equations 1 to 4, O_i is the observed values at time i, f_i is the predicted values at time i, $\bar{f}$ is the mean of the predicted values, $\bar{O}$ is the mean of the observed values, and n is the number of data.

3. Results and discussion

The results of examining the performance of the GMDH and GMDH-HS models in predicting dust storms at eight synoptic stations in Kurdistan Province over a 5-year period (1971-2020) indicate the high accuracy of both models in predicting the FDSD index and a slight and insignificant difference in their results. For example, at Baneh station, the GMDH model has NRMSE, MAPE, and correlation (R) values of 0.161, 1.517 percent, and 0.957, respectively, while the GMDH-HS model performs better with NRMSE, MAPE, and R values of 0.113, 1.231 percent, and 0.986, respectively. The findings indicate high accuracy of both models and a slight improvement in the performance of the GMDH-HS model compared to the individual GMDH model. At Saqqez station, the GMDH model shows NRMSE, MAPE, and R values of 0.169, 1.712 percent, and 0.953, respectively, while the GMDH-HS model shows more accurate performance with NRMSE values of 0.117, MAPE of 1.721 percent, and R of 0.985. These comparisons indicate the high ability of both models in predicting the desired index, while the GMDH-HS model also shows a slight advantage at this station. Overall, the results at all stations indicate the high accuracy of both models in predicting the FDSD index, but the GMDH-HS model provides more accurate and reliable performance at the stations under study.

The evaluation indices R, NRMSE, MAPE, and NS were used to examine and compare the performance of the GMDH and GMDH-HS models to predict the frequency index of days with dust storms in Kurdistan Province over a 50-year statistical period (1971-2020) (Figure 5). Tables 2 and 3, respectively, show the values of the evaluation criteria related to the training and testing stages of the GMDH and GMDH-HS models for predicting the frequency index of days with dust storms at eight synoptic stations: Baneh, Saqqez, Marivan, Sanandaj, Zarrineh, Qorveh, Salamatabad, and Bijar. In these tables, the evaluation criteria in 4 combinations for each of the aforementioned models have been examined. According to Table 2 and the evaluation indicators used, it can be concluded that all the stations under study show better performance in the first and second combinations than in the other combinations. Therefore, using one and finally two previous combinations to model dust storms at all stations improves modeling performance. Among them, combination number one has been selected as the optimal combination for modeling the FDSD index with the GMDH model. According to the results of modeling dust storms at Baneh station (with the highest frequency of days with dust storms on a seasonal scale) and Bijar station (with the lowest seasonal frequency), it can be concluded that as the number of dusty days at the studied stations decreases, moving from Baneh and Saqqez stations towards Selmatabad and Bijar stations, the accuracy of modeling the FDSD index decreases And the third and fourth combinations in Table 2 are not so acceptable. For example, at Baneh station, the correlation coefficient and Nash-Sutcliffe error increased from

0.954 and 0.904 days in the fourth step to 0.957 and 0.907 days in the first step, respectively. At Saqqez, Sanandaj, and Qorveh stations, the error criteria decreased from 1.712, 3.001, and 4.127 days to 2.001, 3.427, and 4.927 days, respectively, after predicting dust storms using the first season instead of the fourth season. The results of this section are consistent with the results of Amiri *et al*. (2024), Golkar *et al*. (2019), Ni *et al*. (2025), and Mo *et al*. (2018).

Table 2. Statistical results of input data to the GMDH model for predicting the FDSD index (Source: Research findings)

Station	Combination number	Train dataset				Test dataset			
		NRMSE	MAPE (%)	NS	R	NRMSE	MAPE (%)	NS	R
Baneh	1	0.147	1.327	0.918	0.971	0.161	1.517	0.907	0.957
	2	0.149	1.341	0.917	0.971	0.162	1.519	0.906	0.956
	3	0.146	1.425	0.916	0.969	0.164	1.612	0.905	0.955
	4	0.146	1.512	0.916	0.969	0.167	1.693	0.904	0.954
Saqqez	1	0.149	1.641	0.916	0.968	0.169	1.712	0.903	0.953
	2	0.151	1.644	0.919	0.968	0.166	1.723	0.903	0.953
	3	0.152	1.703	0.921	0.967	0.168	1.745	0.902	0.952
	4	0.153	1.941	0.915	0.966	0.171	2.001	0.901	0.951
Zarrineh	1	0.154	2.232	0.914	0.965	0.171	2.325	0.901	0.949
	2	0.156	2.325	0.913	0.964	0.172	2.426	0.902	0.949
	3	0.156	2.517	0.913	0.963	0.173	2.612	0.904	0.948
	4	0.156	2.693	0.913	0.962	0.174	2.812	0.899	0.947
Marivan	1	0.157	2.921	0.912	0.961	0.175	3.001	0.898	0.947
	2	0.159	3.003	0.912	0.961	0.176	3.015	0.897	0.946
	3	0.161	3.127	0.911	0.959	0.177	3.225	0.897	0.945
	4	0.162	3.246	0.914	0.958	0.178	3.427	0.896	0.944
Sanandaj	1	0.158	3.706	0.913	0.957	0.179	3.728	0.895	0.943
	2	0.157	3.827	0.912	0.956	0.181	3.926	0.894	0.943
	3	0.163	3.976	0.911	0.955	0.182	4.004	0.894	0.942
	4	0.163	4.004	0.909	0.954	0.183	4.125	0.893	0.941
Bijar	1	0.164	4.106	0.908	0.953	0.184	4.127	0.892	0.941
	2	0.165	4.221	0.907	0.952	0.185	4.321	0.891	0.941
	3	0.166	4.431	0.907	0.951	0.186	4.527	0.891	0.939
	4	0.167	4.851	0.907	0.951	0.187	4.927	0.891	0.938
Qorveh	1	0.168	4.995	0.906	0.949	0.188	5.001	0.889	0.937
	2	0.169	5.007	0.905	0.948	0.189	5.231	0.888	0.937
	3	0.171	5.224	0.904	0.948	0.192	5.321	0.887	0.936
	4	0.172	5.123	0.903	0.947	0.193	5.426	0.886	0.935
Salamatabad	1	0.173	5.321	0.902	0.946	0.201	5.721	0.885	0.934
	2	0.174	5.421	0.902	0.945	0.203	5.823	0.884	0.933
	3	0.175	5.823	0.901	0.944	0.204	5.927	0.883	0.932
	4	0.181	5.903	0.901	0.943	0.205	5.941	0.882	0.931

Table 3. Statistical results of input data to the GMDH-HS model for predicting the FDSD index (Source: Research findings)

Station	Combination number	Train dataset				Test dataset			
		NRMSE	MAPE (%)	NS	R	NRMSE	MAPE (%)	NS	R
Baneh	1	0.093	1.007	0.943	0.997	0.113	1.231	0.923	0.986
	2	0.096	1.113	0.943	0.997	0.114	1.325	0.923	0.986
	3	0.097	1.114	0.942	0.996	0.115	1.421	0.922	0.985
	4	0.099	1.201	0.941	0.998	0.115	1.521	0.921	0.985
Saqqez	1	0.101	1.207	0.939	0.995	0.117	1.721	0.921	0.985
	2	0.101	1.301	0.938	0.995	0.112	1.728	0.921	0.984
	3	0.103	1.315	0.937	0.995	0.119	1.741	0.919	0.979
	4	0.104	1.346	0.936	0.994	0.121	1.912	0.918	0.976
Zarrineh	1	0.099	1.378	0.935	0.993	0.121	1.918	0.918	0.983
	2	0.105	1.512	0.934	0.994	0.122	2.001	0.922	0.983
	3	0.106	1.713	0.933	0.996	0.123	2.003	0.917	0.982
	4	0.107	1.719	0.932	0.993	0.124	2.125	0.924	0.981
Marivan	1	0.108	1.803	0.931	0.992	0.125	2.136	0.916	0.979
	2	0.108	1.825	0.929	0.991	0.126	2.205	0.916	0.978
	3	0.108	1.827	0.929	0.989	0.126	2.301	0.916	0.977
	4	0.109	1.903	0.929	0.987	0.127	2.305	0.915	0.975
Sanandaj	1	0.112	1.924	0.928	0.988	0.127	2.307	0.914	0.975
	2	0.113	1.978	0.927	0.989	0.128	2.802	0.914	0.975
	3	0.113	2.006	0.926	0.987	0.129	2.904	0.913	0.974
	4	0.114	2.125	0.926	0.986	0.129	2.907	0.912	0.973
Bijar	1	0.115	2.224	0.925	0.985	0.131	2.923	0.912	0.972
	2	0.116	2.321	0.924	0.985	0.132	2.945	0.912	0.971
	3	0.117	2.325	0.923	0.984	0.133	3.007	0.911	0.971
	4	0.118	2.426	0.923	0.983	0.134	3.012	0.909	0.971
Qorveh	1	0.119	2.512	0.927	0.983	0.135	3.221	0.908	0.969
	2	0.121	2.614	0.925	0.983	0.136	3.334	0.907	0.969
	3	0.122	2.712	0.925	0.982	0.137	3.412	0.906	0.968
	4	0.123	2.803	0.922	0.979	0.138	3.419	0.905	0.967
Salamatabad	1	0.124	2.912	0.921	0.981	0.139	3.511	0.904	0.966
	2	0.125	2.956	0.919	0.978	0.141	3.526	0.904	0.965
	3	0.126	2.978	0.918	0.977	0.142	3.596	0.903	0.964
	4	0.127	3.003	0.917	0.976	0.143	4.001	0.902	0.963

Table 3 presents the results of dust storm forecasting over a 50-year period (1970-2020) in Kurdistan Province. By examining the results of modelling the FDSD index with the combined GMDH-HS model, it can be concluded that in this method, as in the previous method, in all stations studied, the modeling results of these storms improved by using one or two previous seasons, and by using more seasons, the NRMSE and MAPE error values had an increasing trend. For example, at Marivan and Selmatabad stations, the NRMSE and MAPE error values have decreased relatively from 0.174, 2.812, 0.193, and 5.426 in the fourth seasonal combination to 0.171, 2.325, 0.188, and 5.001 days as a result of applying a previous season to model dust storms. Therefore, it can be concluded that using a previous season to predict the FDSD index in future seasons provides the best results. Considering that in Table 3, the order of the stations according to the FDSD index is descending from Baneh to Bijar, it can be concluded that the performance and accuracy of the GMDH-HS model for modeling dust storms is directly related to the increase in the number of days with dust storms. On the other hand, as can be seen in Table 3, the error value at all studied stations for all combinations has changed slightly compared to the individual GMDH model. Also, the correlation and Nash-Sutcliffe coefficients have slightly increased compared to the modeling results with GMDH. For example, at Saqez station, the NRMSE and MAPE values in the test phase in combination number four have decreased from 0.171 and 2.001 in the GMDH model to 0.121 and 1.912 in the GMDH-HS model. Thus, it can be concluded that the GMDH-HS intelligent model method is superior to the individual GMDH model in modeling dust storms. The comparative analysis presented in Tables 2 and 3 reveals a slight difference in predictive performance between the two predicting models of this study. It shows that the GMDH model has a high ability to predict the frequency of dust storms in Kurdistan Province. This high level of performance can be attributed to the GMDH's inherent ability to automatically select optimal functional structures, which effectively handles the non-linear and stochastic nature of duststorm-related atmospheric variables. The observed results align with previous studies, such as Qaderi *et al*. (2020), Olfat Miri *et al*. (2022), and Ghorbani *et al*. (2013), suggesting that the complex meteorological drivers in the semi-arid regions of Iran necessitate advanced non-linear modeling approaches to achieve reliable forecasting. These findings indicate that the GMDH model provides a robust and reliable alternative for long-term dust storm monitoring under similar climatic conditions.

The cost and complexity of modeling dust storms is directly dependent on the choice of model and analysis method. The GMDH model is known as an efficient and low-cost tool for modeling nonlinear phenomena due to its self-organizing structure and ability to automatically select effective variables. This model, utilizing limited computational resources and a simple process, offers a suitable option for analyzing complex data with acceptable accuracy. In contrast, the GMDH-HS hybrid model, which is formed by combining the GMDH method and the Harmony algorithm, has greater ability to model complex phenomena by utilizing more advanced optimization processes. This model, although it increases the prediction accuracy, has a higher computational cost due to the need for more complex settings and longer execution time. Choosing between these two models requires a careful assessment of the project needs, including the expected level of accuracy, time constraints, and available computational resources. In cases where the volume of calculations and the cost of project construction are more important, using the GMDH model seems more logical. While for situations where forecast accuracy and uncertainty reduction are of higher priority, the combined GMDH-HS model would be an optimal option, despite higher implementation costs. In addition, the difference in the results of the models is not significant at the 95 and 99 percent confidence levels. Therefore, despite the high performance and considerable accuracy of the GMDH-HS hybrid model, the results of the GMDH model also showed acceptable performance with a slight difference compared to the hybrid method. The GMDH intelligent model requires fewer computational resources due to its simple structure and training and testing processes. This model is a suitable option in situations where the data is simpler and the relationships between variables are more linear, and it requires less time for processing and analysis. While the GMDH-HS model requires more computational resources due to the use of more complex algorithms and consideration of nonlinear interactions and more complex changes in the data. Therefore, this model may not be a suitable option in projects that have limitations in terms of computational resources or require fast predictions. In terms of implementation costs and processing time, the GMDH model, with its lower complexity and lower computational costs, is more cost-effective for projects that require fast and relatively accurate predictions. While the GMDH-HS model, despite its higher costs, can be beneficial in sensitive projects requiring more accurate predictions due to its higher accuracy and ability to simulate more complex relationships. Comparing the mean of observed and predicted values using the t-test indicates acceptance of the null hypothesis based on the equality of the mean of the observed and predicted time series of the frequency index of days with dust storms in Kurdistan Province (Table 4). Thus, it can be concluded that both models have retained the average of the observational time series to predict the FDSD index. The relationship between the observed and predicted values at all eight stations studied is a line with an approximate slope of 45 degrees. Among the models studied, the GMDH-HS hybrid intelligent algorithm has the lowest dispersion (highest accuracy and therefore highest agreement) with the first quadrant bisector (y=x). The results of this section are consistent with the studies of Goudie & Middleton, (2006) and Pourgholam Amiji *et al*. (2020) (Figure 5).

Table 4. Comparison test of the observed and predicted time series averages of the FDSD index (Source: Research findings)

	average			T statistic	
	observations	GMDH	GMDH- HS	GMDH	GMDH- HS
Baneh	0.993	0.981	0.987	2.127	2.236
Saqqez	0.984	0.962	0.983	2.124	2.212
Marivan	0.976	0.957	0.973	2.078	2.178
Sanandaj	0.965	0.943	0.961	2.023	2.143
Zarrineh	0.961	0.931	0.958	2.004	2.131
Qorveh	0.951	0.921	0.949	1.993	2.117
Salamatabad	0.943	0.912	0.938	1.984	2.104
Bijar	0.939	0.907	0.931	1.878	2.005

4. Conclusion

This study developed and evaluated a hybrid GMDH-HS intelligent algorithm to model the frequency index of dust storm days (FDSD) in Kurdistan Province, Iran, using long-term data (1971–2020). The integration of the Harmony Search (HS) meta-heuristic with the Group Function Data-Driven (GMDH) approach aimed to overcome the limitations of traditional modeling in capturing the complex, non-linear dynamics of dust storm phenomena.

The results demonstrated that both the GMDH and the hybrid GMDH-HS models possess high predictive potential for FDSD indices. The performance difference between the two models was not statistically significant, but the GMDH-HS model recorded superior performance in minimizing prediction errors and optimizing internal parameters through its automatic evolutionary process. This confirms that the hybrid approach provides a more robust framework for simulating the highly stochastic and non-linear nature of dust storm patterns in semi-arid regions. The findings of this research offer practical implications for environmental management and public health authorities in Kurdistan Province. The developed GMDH-HS model can serve as a reliable tool for establishing early warning systems and aiding in the development of proactive mitigation strategies to reduce the socio-economic and health-related impacts of dust storms.

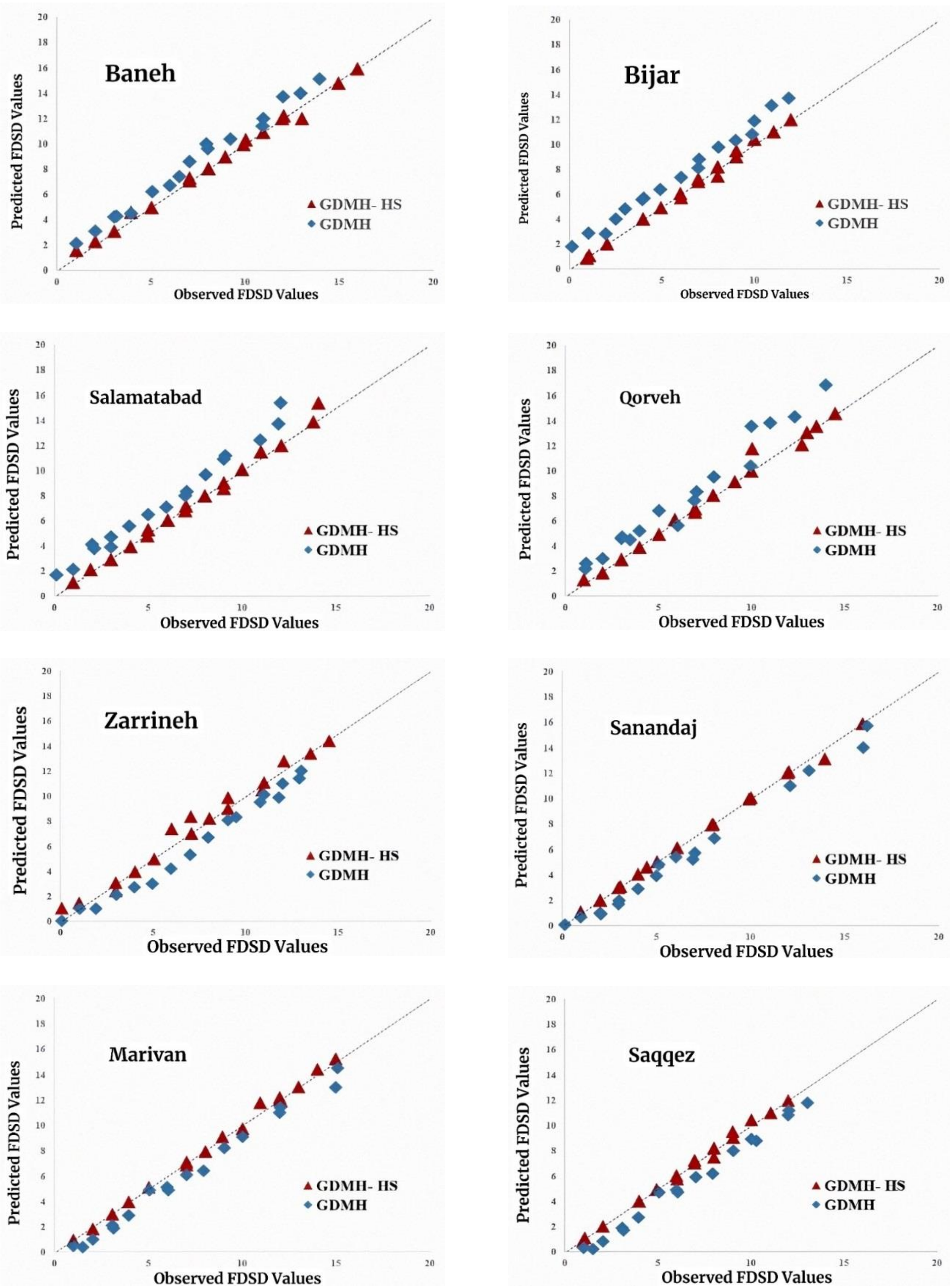

Figure 4. Comparison of observed and predicted values of the FDSD index in Kurdistan Province (Source: Research findings)

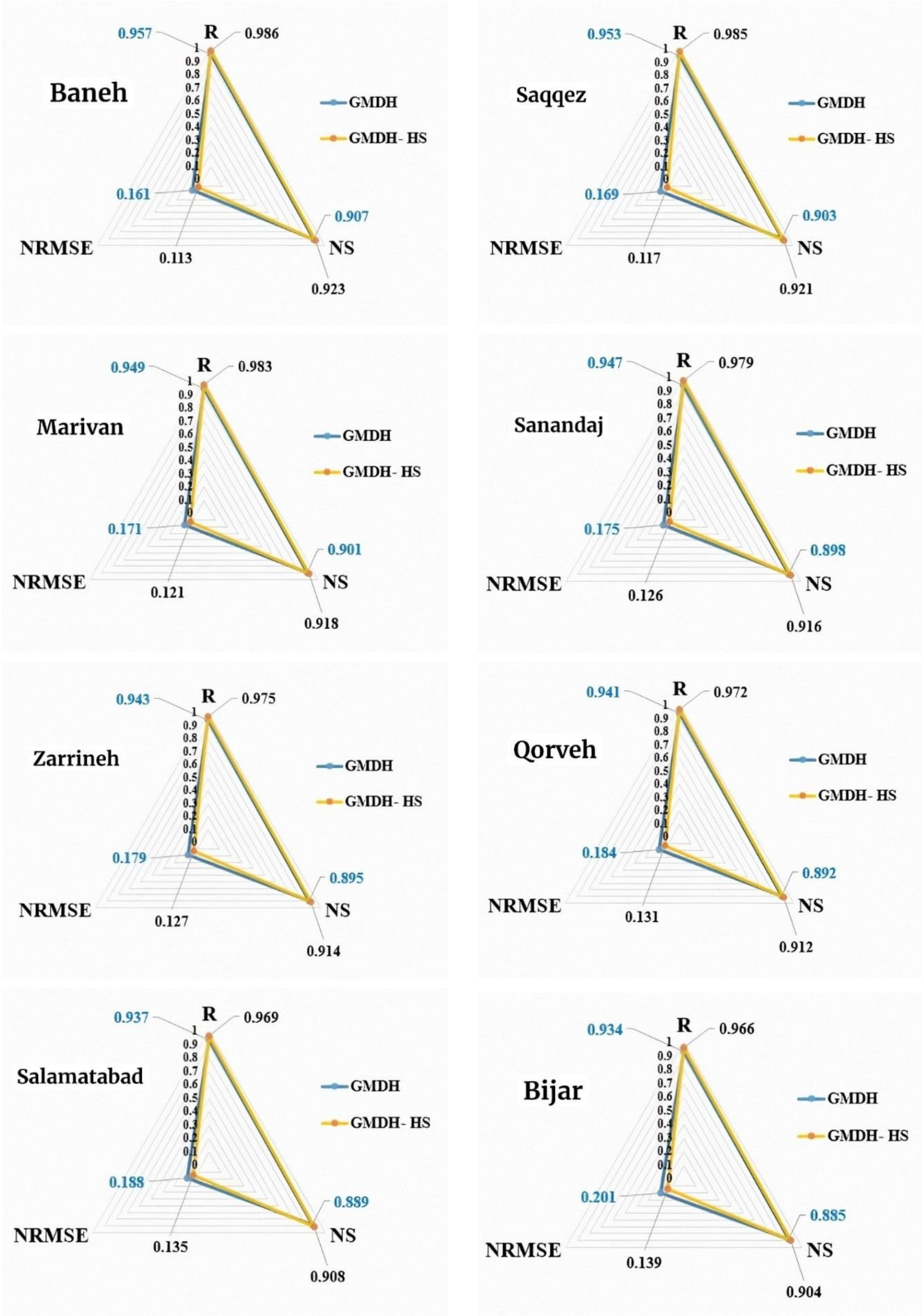

Fig. 5. Comparison of Behbaneh model evaluation indices for modeling dust storms (Source: Research findings)

Despite the high accuracy achieved, this study is limited by the reliance on historical FDSD indices. Future research should focus on integrating real-time meteorological variables—such as wind speed, humidity, and aerosol optical depth—into the GMDH-HS framework to enhance the temporal resolution and reliability of short-term dust storm forecasting.

Acknowledgment
The authors would like to thank all the experts who contributed to the completion of this study.

Author Contributions
Conceptualization, Ansai ghojghar, M. and Khosravi, H.; methodology, Ansari ghojghar, M.; software, Ansai ghojghar, M.; validation, Tavoosi Rad, R., Ansai ghojghar, M. and Khosravi, H.; formal analysis, Tavoosi Rad, R.; investigation, Tavoosi Rad, R.; resources, Tavoosi Rad, R.; data curation, Khoshravi, H.; writing—original draft preparation, Tavoosi Rad, R.; writing—review and editing, Tavoosi Rad, R.; visualization, Tavoosi Rad, R.; supervision, Ansai ghojghar, M, Khosravi, H., project administration, Ansai ghojghar, M. All authors have read and agreed to the published version of the manuscript

Funding
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Competing interests
The authors declare that they have no competing interests.
Data availability

References

Abualigah, L., Diabat, A., & Geem, Z. W. (2020). A Comprehensive Survey of the Harmony Search Algorithm in Clustering Applications. *Applied Sciences*, *10*(11), 3827. https://doi.org/10.3390/app10113827

Alshammari, R. K., Alrwais, O., & Aksoy, M. S. (2024). Machine Learning Forecast of Dust Storm Frequency in Saudi Arabia Using Multiple Features. *Atmosphere*, *15*(5), 520. https://doi.org/10.3390/atmos15050520

Amiri, M., Amiri, F., Pourasad, M. H., & Soleimani, S. (2024). Optimization of the Results of the ML-Based GMDH Algorithm in order to Increase the Accuracy of Walnut Pollen Detection and Horizontal Optical Depth through the TLBO Algorithm. *Iranian Journal of Remote Sensing & GIS*, *16*(1), 49-66. https://doi.org/10.48308/gisj.2022.102900

Ansari ghojghar, M., & Pourmohammad, P. (2024). Dust Storm Modeling Using the Triple Hybrid Model GRNN-SVM-LSTM (Case Study: Sistan and Baluchestan Province). *Iranian Journal of Soil and Water Research*, *55*(10), 1961-1979. https://doi.org/10.22059/ijswr.2025.385882.669844

Ansari ghojghar, M., Bazrafshan, J., & Araqi Nejad, S. (2022). Evaluating the efficiency of hybrid meta-models combining machine learning and Box-Jenkins for dust storm modeling (Case study: Khuzestan Province). *Iranian Journal of Water and Soil Research, 53*(8), 1695–1714. https://doi.org/10.22059/ijswr.2022.346694.669335

Du, S.; Chen, S.; Cheng, S.; He, J.; Zhao, D.; Zhu, X.; Lian, L.; Tu, X.; Zhao, Q.; & Zhang, Y. (2025). Dust Storm Detection for Ground-Based Stations with Imbalanced Machine Learning. *Environ. Model. Softw*, *188*, 106420. https://doi.org/10.1016/j.envsoft.2025.106420

Gholkar, Ahmadi, M. M., Qaderi, K., & Rahimpour, M. (2019). Prediction of Peak Velocity of Pollutant Transport in Rivers Using Group Method Data Handling (GMDH) and Intelligent Hybrid Method GMDH-HS. *Journal of Water and Wastewater; Ab va Fazilab (in persian)*, *30*(1), 64-76. https://doi.org/10.22093/wwj.2018.83092.2389

Ghorbani, A., Eslami, A., & Ebrahimi, H. (2013). ESTIMATION OF THE TIP BEARING CAPACITY OF SINGLE PILE FROM CPTU DATA USING GMDH TYPE NEURAL NETWORKS. *Sharif Journal of Civil Engineering*, *Volume 29-2*(3), 71-80.

Goudie, A. S. (2014). Desert dust and human health disorders. Environment International, 63, 101-113. https://doi.org/10.1016/j.envint.2013.10.011

Goudie, A. S., & Middleton, N. J. (2006). *Desert dust in the global system*. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-32355-4

Han, X., Liu, L., Shi, P., Zhang, G., Liu, Y., Zhang, X., Yang, Y., Qu, Z., Wang, G., & Hu, Z., *et al*. (2022). Dust storm and haze convergence process and its effect on air pollution in Northern China. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4155201

Ivakhnenko, A. G. (1971). Polynomial theory of complex systems. *IEEE Transactions on Systems, Man, and Cybernetics*, *1*(4), 364–378. https://doi.org/10.1109/TSMC.1971.4308320

Khajeh, M., Ghaffari-Moghaddam, M., Piri, J., Barkhordar, A., & Ozturk, T. (2024). Machine learning-aided enhancement of white tea extraction efficiency using hybridized GMDH models in microwave-assisted extraction. *Scientific Reports*, *14*(1), 25900. https://doi.org/10.1038/s41598-024-25900-6

Liang, L., Han, Z., Li, J., Xia, X., Sun, Y., Liao, H., Liu, R., Liang, M., Gao, Y., & Zhang, R. (2022). Emission, transport, deposition, chemical, and radiative impacts of mineral dust during severe dust storm periods in March 2021 over East Asia. *Science of the Total Environment*, 852, 158459. https://doi.org/10.1016/j.scitotenv.2022.158459

Mahdavi-Meymand, A., Zounemat-Kermani, M., & Qaderi, K. (2021). Prediction of hydro-suction dredging depth using data-driven methods. *Frontiers of Structural and Civil Engineering*, *15*(3), 652–664. https://doi.org/10.1007/s11709-021-0732-5

Manisalidis, I., Stavropoulou, E., Stavropoulos, A., & Bezirtzoglou, E. (2020). Environmental and health impacts of air pollution: A review. *Frontiers in Public Health*, *8*, 14. https://doi.org/10.3389/fpubh.2020.00014

Middleton, N., & Kang, U. (2017). Sand and Dust Storms: Impact Mitigation. *Sustainability*, *9*(6), 1053. https://doi.org/10.3390/su9061053

Miri, A., & Middleton, N. (2022). Long-term impacts of dust storms on transport systems in south-eastern Iran. *Natural Hazards, 114*, 291–312. https://doi.org/10.1007/s11069-022-05379-3

Miri, A., Shirmohammadi, E., & Sorooshian, A. (2023). Impacts of dust storms on indoor and outdoor bioaerosol concentration in the Sistan region of Iran. *Journal of Building Engineering, 76*, 107302. https://doi.org/10.1016/j.jobe.2023.107302

Mo, L., Xie, L., Jiang, X., Teng, G., Xu, L., & Xiao, J. (2018). GMDH-based hybrid model for container throughput forecasting: Selective combination forecasting in nonlinear subseries. *Applied Soft Computing*, *62*, 478-490. https://doi.org/10.1016/j.asoc.2017.10.033

Ni, C., Huang, H., Cui, P., Ke, Q., Tan, S., Ooi, K. T., & Liu, Z. (2025). Integrated multi-objective optimization of a horizontal evaporator structure in domestic refrigerators: Comparison between the semi-empirical model and GMDH neural networks for enhanced Pareto frontiers. *Applied Thermal Engineering*, *260*, 124949. https://doi.org/10.1016/j.applthermaleng.2024.124949

O'Loingsigh, T., McTainsh, G. H., Tews, E. K., Strong, C. L., Leys, J. F., Shinkfield, P., & Tapper, N. J. (2014). The Dust Storm Index (DSI): a method for monitoring broadscale wind erosion using meteorological records. *Aeolian Research*, *12*, 29-40. https://doi.org/10.1016/j.aeolia.2013.10.004

Olfatmiri, Y., Jabbari, E., Bahreinimotlagh, M., Alizadeh, H., & Hassanabadi, A. (2022). Application of Acoustic Tomographic Data in Short-Term Forecasting of Streamflow Using Combinatorial GMDH Algorithm (CGA). *Amirkabir Journal of Civil Engineering*, *54*(5), 1755-1772. https://doi.org/10.22060/ceej.2021.19850.7270

Pourgholam Amiji, M., Ansari Ghojghar, M., Bazrafshan, J., Liaghat, A., & Araghinejad, S. (2020). Comparing the Performance of SARIMA and Holt-Winters Time Series Models With Artificial Intelligence Methods in Dust Storms Forecasting (Case Study: Sistan and Baluchestan Province). *Physical Geography Research*, *52*(4), 567-587. https://doi.org/10.22059/jphgr.2021.303847.1007524

Qaderi, K., Bakhtiari, B., Madadi, M. R., & Afzali-Gorouh, Z. (2020). Evaluating GMDH-based models to predict daily dew point temperature (case study of Kerman province). *Meteorology and Atmospheric Physics, 132*, 667–682. https://doi.org/10.1007/s00703-020-00723-2.

Sobhani, B., Safarian Zengir, V., & Faizollahzadeh, S. (2020). Modeling and Prediction of Dust in Western Iran. *Physical Geography Research*, *52*(1), 17-35. https://doi.org/10.22059/jphgr.2020.284389.1007408

Tang, H., Haugvaldstad, O. W., Stordal, F., Bi, J., Groot Zwaaftink, C. D., Grythe, H., ... & Kaakinen, A. (2023). Modelling the 2021 East Asia super dust storm using FLEXPART and FLEXDUST and its comparison with reanalyses and observations. *Frontiers in Environmental Science*, *10*, 1013875. https://doi.org/10.3389/fenvs.2022.1013875

Wang, W., Samat, A., Abuduwaili, J., Ge, Y., De Maeyer, P., & Van de Voorde, T. (2022). Temporal characterization of sand and dust storm activity and its climatic and terrestrial drivers in the Aral Sea region. *Atmospheric Research*, 275, 106242. https://doi.org/10.1016/j.atmosres.2022.106242

Yang, X. S. (2009). Harmony search as a metaheuristic algorithm. *Music-inspired harmony search algorithm: theory and applications*, 1-14. https://doi.org/10.48550/arXiv.1003.1599

Yarmohamadi, M., Alesheikh, A. A., & Sharif, M. (2025). Using Hybrid Deep Learning Models to Predict Dust Storm Pathways with Enhanced Accuracy. *Climate*, *13*(1), 16. https://doi.org/10.3390/cli13010016

Zhang, J., Zhou, C. H., Gui, H. L., & Zhang, X. Y. (2021). Biological crust in sand and dust storm source areas of Asia and its impact on dust emission. *Advances in Climate Change Research*, *12*(3), 395-408. https://doi.org/10.1016/j.accre.2021.05.007