Sakarya University Journal of Science
SAUJS

ISSN 1301-4048 e-ISSN 2147-835X Period Bimonthly Founded 1997 Publisher Sakarya University
http://www.saujs.sakarya.edu.tr/

Title: The Support Vector Regression with L1 Norm: Application to Weather Radar Data in Adjusting Rainfall Errors

Authors: Arzu OZKAYA, Asım Egemen YILMAZ

Recieved: 2022-03-19 00:00:00

Accepted: 2023-03-22 00:00:00

Article Type: Research Article

Volume: 27
Issue: 3
Month: June
Year: 2023
Pages: 634-642

How to cite
Arzu OZKAYA, Asım Egemen YILMAZ; (2023), The Support Vector Regression with L1 Norm: Application to Weather Radar Data in Adjusting Rainfall Errors. Sakarya University Journal of Science, 27(3), 634-642, DOI: 10.16984/saufenbilder.1090178
Access link
https://dergipark.org.tr/en/pub/saufenbilder/issue/78131/1090178

Sakarya University Journal of Science 27(3), 634-642, 2023

The Support Vector Regression with L_1 Norm: Application to Weather Radar Data in Adjusting Rainfall Errors

Arzu ÖZKAYA *1, Asım Egemen YILMAZ2

Abstract

This paper presents the results of the research on radar rainfall estimate errors with the support vector regression (SVR) method using the observed rain gauge data. The paper depicts the methodological base of the algorithm that covers additive and multiplicative corrections and the practical implementations considering the locations of gauge measurements. The preliminary results show that the SVR has a location-oriented performance. The multiplicative and additive correction factors show decreasing and unstable trends respectively, as the distance from the radar location increase. Another particular outcome is that the SVR shows better results for almost all stations in decreasing the error in maximum rainfall amounts measured with weather radar.

Keywords: Support vector regression, weather radar data, flood, error minimization

1. INTRODUCTION

With the increasing world population, climate change, and rapid urbanization, extreme weather events are expected to occur with growing frequencies. According to World Meteorological Organization (WMO), 44% of disasters have been associated with floods all over the world [1]. This outcome makes studies in forecasting floods extremely important. Moreover, success in flood estimation primarily depends on accurate rainfall data [2]. Rain gauge stations, treated as ground-truth measurements, are the basic source in flood forecasting studies. However, rainfall has a dynamic spatio-temporal pattern and rain gauge stations, pointwise measurements, are generally too sparse to capture this variability [3]. Representation of rainfall distribution in ungauged or poorly-gauged areas, remote sensing products can be used because of their wide coverage and fine resolutions. In the field of hydrology, weather radar products have been used for decades with some inaccuracy [4]. The quality of radar products can be improved with the implementation of methods using ground reference data [5]. Although there is a wide range of studies to assess and increase the performance of radar rainfall data, the generalization of the methods for a broader area causes a limitation in error reduction [6]. And this points out that future analysis should focus on working with more radar products.

* Corresponding author: a.ozkaya@aybu.edu.tr (A. OZKAYA)
1 Ankara Yıldırım Beyazıt University
2 Ankara University
E-mail: aeyilmaz@eng.ankara.edu.tr
ORCID: https://orcid.org/0000-0003-3983-8831, https://orcid.org/0000-0002-4156-4238

Since rainfall depth is indirectly measured by the weather radar systems, uncertainties in the products are inevitable. And, errors that come from these uncertainties can be alleviated by various kinds of methods in the pre-or post-processing stages of data production. Although uncertainties coming from the error sources like signal attenuation, ground clutter, anomalous propagation, beam blockage, miscalibration, vertical air motion, and precipitation drift [7] can be reduced by pre-processing techniques, our knowledge about these errors is limited. As for post-processing techniques, geostatistical methodologies, statistical objective analysis, and deep learning models are generally used. In statistical analysis, vector norms with correction factors like multiplicative and additive can also be useful in error minimization. And, the recent application with Taxicab Norm showed that error reduction performed well in the multiplicative model but, correction values were dominated by location [8]. Since there is a nonlinear systematic error in the radar data, error generalization with norm methods may not fit completely, and obviously, it is not the unique way for the solution.

In learning algorithms, the support vector regression (SVR) is a famous one based on Vapnik's concept [9]. And, SVR indicates powerful results in time series analysis in most of the studies [10–13]. Moreover, since SVR is based on computational learning theory, optimum weights and thresholds for the trained network can be found conveniently. In searching parameters, although SVR has a lack of knowledge memory, problems related to inefficiency or time consumption do not appear especially for the processes that are nonlinear and nonstationary [14]. Since atmospheric processes have complex interactions, especially over complex terrain, hydrologic elements exhibit high nonlinearity. Rainfall, the driving force in flood studies, is one of the important nonlinear variables among hydrologic elements. However, an error that is the difference between the amount of rainfall recorded from the gauge and that estimated from the weather radar, can be reduced with SVR.The study in this paper is motivated by a desire to apply the SVR method in error minimization of radar rainfall estimates recorded in flood events in Mugla, Turkey for 16 flood events. To the best of our knowledge, the SVR application using weather radar data is the first in Turkey.

The paper is organized as follows. In Section II, the study area and datasets are described. Section III gives the methodology. In Section IV, results are given and the paper ends with conclusions in Section V.

2. STUDY AREA AND DATASETS

The study area covers the city of Muğla in Turkey (Figure 1). According to the Köppen-Geiger climate classification, cold-rainy winters and Mediterranean hot summer climates prevail over the region. Due to its climate characteristics, heavy rainfalls and floods are observed in late autumn and early winter, and this causes loss of lives and damage to infrastructures [15]. In the time between 2015 and 2019, 16 flood events were observed within Muğla province (Table 1). For the southwest part of Turkey, the spatiotemporal distributed rainfall data can be acquired by Muğla weather radar. It is a C-band Doppler weather radar and has a 120-km range with 333.33-m spatial resolution.

Table 1 Information about the flood events

No	Date (dd/mm /yyyy)	Location	Max. Cum. Rainfall (Gauge/ Radar)	Max. Rainfall (Gauge/ Radar)	Stand. Dev. (Gauge/ Radar)	data (hrs)
1	11/5/2015	Muğla	60.5/14.9	27.8/3.9	3.9/0.6	71
2	21/09/2015	Bodrum	211.4/24.1	36.8/5.8	7.9/1.0	71
3	21/10/2015	Bodrum	109.4/31.5	29.9/6.2	4.6/1.1	71
4	16/01/2017	Muğla	85.6/82.6	10.2/13.4	2.1/2.9	71
5	7/2/2017	Bodrum	58.4/26.3	17.6/9.5	3.0/1.5	71
6	7/3/2017	Marmaris	173.8/69.2	23.2/8.2	3.9/1.5	71
7	10/3/2017	Ortaca	41.3/43.6	5.3/5.0	1.1/1.1	71
8	23/10/2017	Muğla	63.0/24.1	14.8/3.7	2.5/0.8	71
9	13/11/2017	Datça	42.1/28.5	9.8/9.1	1.9/1.6	71
10	28/12/2017	Fethiye	127.8/69.1	36.2/10.3	4.9/1.8	71
11	16/11/2018	Bodrum	149.5/71.0	41.0/15.9	7.2/3.0	71
12	16/07/2019	Muğla	44.4/22.9	16.8/8.8	2.8/1.5	71
13	23/09/2019	Dalaman	45.7/22.0	21.5/10.2	3.4/1.6	71
14	4/10/2019	Köyceğiz	101.6/23.3	58.5/12.7	12.3/2.8	31
15	3/11/2019	Marmaris	185.8/49.0	47.8/11.8	8.7/2.0	71
16	24/11/2019	Ortaca	126.5/60.4	46.2/13.6	6.0/1.9	71
Av.	-	-	101.7/41.4	27.7/9.3	4.8/1.4	

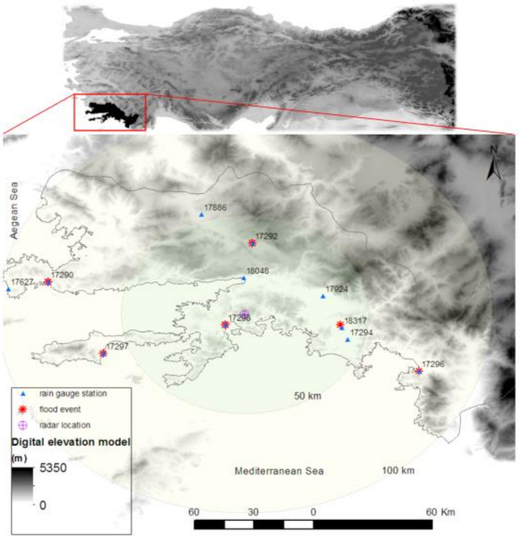

Figure 1 Study area, locations of rain gauge stations, flood events, and weather radar station with radar ranges (50 km and 100 km) over digital elevation model

For application, the flood data from the observational stations and Muğla weather radar were provided by the Turkish State Meteorological Service (TSMS). The locations of flood events and stations over the digital elevation model can be seen in Figure 1. The information about the flood events that cover maximum accumulated rainfall amounts both gauge and radar datasets for each event and corresponding maximum hourly rainfall values and standard deviations are given in Table 2. Given datasets, the length of the data is 71 hrs except for event number 14, lasting 31 hours.

Table 2 Station information

Station No	longitude	latitude	Elevation (m)
17290	27.440	37.033	26
17292	28.367	37.210	646
17294	28.799	36.772	12
17296	29.124	36.627	3
17297	27.692	36.708	28
17298	28.245	36.840	16
17886	28.137	37.340	365
17924	28.687	36.970	24
18317	28.772	36.826	13
18048	28.327	37.051	1
17627	27.260	37.000	6

3. METHOD

The SVR algorithm is a nonlinear Generalized Portrait algorithm proposed by Vapnik, used for solving classification and regression problems. It is based on the principle of structural risk minimization (SRM). A visualization of the problem is depicted in Figure 2. The primary goal of this study is to correct radar data with gauge measurements by giving tolerated errors, ε.

As the increase of ε, the number of adjusted radar data in the ε-intensive tube increases, therefore evaluation is performed with regression parameters variation instead of ε.

A simple linear regression problem trained using the dataset with k vector size can be given as

$$[(x_{n-k+1}, y_{n-k+1}), (x_{n-k+2}, y_{n-k+2}), \dots, (x_{n-1}, y_{n-1}), (x_n, y_n)], \quad (1)$$

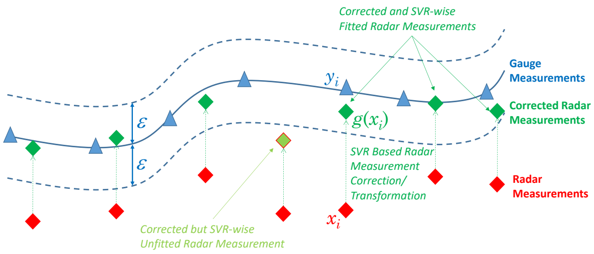

Figure 2 Illustration of an SVR regression function

where x_i is the input data, rainfall value obtained from radar measurement, and y_i is the corresponding output data, rainfall value gathered from gauge observation. By a nonlinear function $\emptyset(x)$, the regression function can be defined as

$$f = g(x) = w^T.\emptyset(x) + b, \quad (2)$$

where w and b are the regression function parameters. The coefficients (w and b) are estimated by minimizing the following general function;

$$r(f) = C\frac{1}{n}\sum_{i=1}^{n} L_\varepsilon(y_i) + \frac{1}{2}||w||^2, \quad (3)$$

where C and ε are predefined parameters. $L_\varepsilon(y_i)$ is the ε-insensitive loss function. When the adjusted value is within the ε-tube, the loss becomes zero. The tolerated errors can be stated with penalized loss as

$$L_\varepsilon(y_i) = \begin{cases} 0, & for\ |y_i - [g(x)]| < \varepsilon, \\ |y_i - [g(x)]| - \varepsilon, & for\ |y_i - [g(x)]| \geq \varepsilon. \end{cases} \quad (4)$$

The regression problem can be expressed with the following optimization problem,

$$\min_{w^T,b} \sum_{i=1}^{k} ||g(x_i) - y_i||, \quad (5)$$

where ‖…‖ is the L_1 norm. By these definitions x_i (the rainfall value gathered from radar measurement) is corrected as $g(x_i)$ by means of the support vector regression.

For the selection of ε value, the input noise value should be known [16]. However, such noise is not generally known nor is this study. In the determination of the ε value, the maximum rainfall amount which is significant in flood studies due to being of the main driving force for the maximum streamflow amount is analyzed for all flood events (Table 1). The average value of maximum observed rainfall is calculated as 27.7 mm/hr. Since epsilon has a direct effect on the results of the SVR, the value selection is constrained as 2 mm. C is the other parameter that controls the penalty on observations. In this study, C is taken as unity, 1.

4. RESULTS AND DISCUSSIONS

Using the SVR algorithm with L_1 norm, regression function parameters are determined for each station. And, variation of averaged regression function parameters for each station is evaluated considering the radar distance (Figure 3). From the results, it is noticed that the multiplicative factors have a decreasing trend as radar distance increases (Figure 3 (a)), whereas the additive factors show an unstable trend in which stations located in mid-distance (40 km - 60 km) have the highest value (Figure 3 (b)).

In order to determine SVR power, four criteria are used in the evaluation. These are the difference between gauge and radar data that is given as error (e), standard deviation (std), the relative error in maximum rainfall (re), and root mean square percentage error (RMSPE). The equations that represent the four criteria are given below,

$$e = \frac{1}{k}\sum_{i=1}^{k}|g(x_i)-y_i|, \quad (6)$$

$$std = \sqrt{\frac{1}{k-1}\sum_{i=1}^{k}(e_i-\overline{e})^2}, \quad (7)$$

$$re = (\frac{1}{k}\sum_{i=1}^{k}|g(x_{i,max})-y_{i,max}|/y_{i,\max})100, \quad (8)$$

$$RMSPE = \sqrt{\frac{\sum_{i=1}^{k}((g(x_i)-y_i)/y_i)^2}{k}}, \quad (9)$$

where $\overline{g(x)}$, $\overline{y}$ and $\overline{e}$ are averages of corrected radar values, gauge rainfall values and mean absolute error values, respectively. $g(x_{i,max})$ and $y_{i,max}$ are the maximum corrected radar rainfall data and the maximum gauge rainfall data for the ith event. k is the number of data pairs used for each flood event.

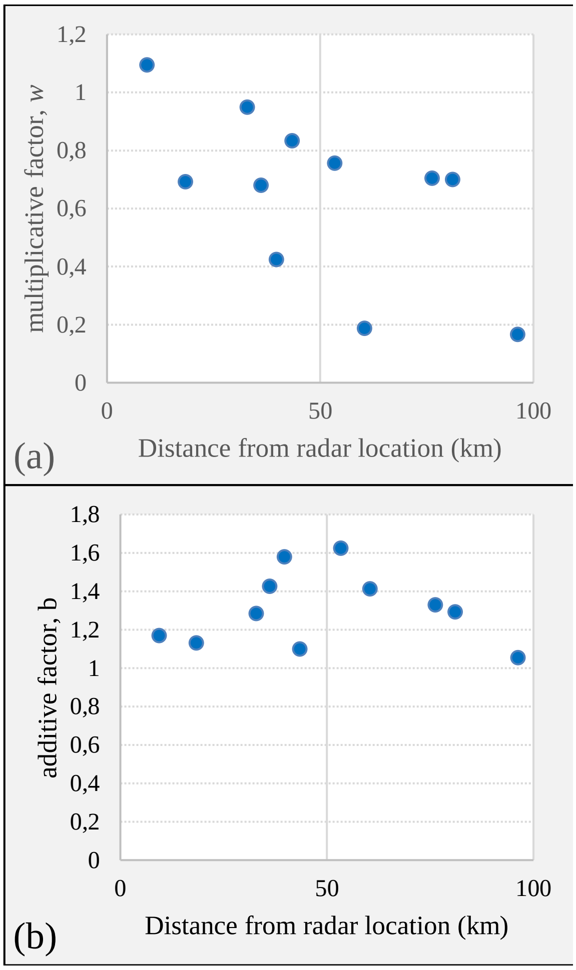

Figure 3 Variation of regression parameters to distance between stations and radar location

The averages of the difference between gauge observations and radar measurements considering the order of stations from closely located stations to further ones are given in Figure 4 (a) and (b). Although there is an increase in the rainfall differences at all stations regardless of proximity to radar location, a decrease in the interquartile range of the box plots belongs to the stations located in mid and long-distance to the radar point (17292, 17886, 17297, 17296, 17290, and 17627) is observed.This outcome is related to error increase and can be explained by the fact that constraining the radar data in the ε-intensive tube makes an increase in the differences.

When the results of standard deviations are analyzed it is seen that all median values show an improvement with the SVR method (Figure 4 (c) and (d)). Moreover, the lengths of the whiskers and quartiles decrease (Figure 4 (d)).

The results for the maximum rainfall amounts show that the error in these data decreases with the SVR application except for the station located in the outermost (Figure 5(a)). With RMSPE calculations, it is seen that the original dataset (depicted with blue color) shows an erorr increase as distance increases in general. With SVR, the stations that show high RMSPE values achieve a decrease but, stations that show low RMSPE values get small improvements or increases in the error. Furthermore, the number of data pairs that fall into the ε-intensive tube does not give apparent relation between the station closeness. The highest three values are noticed in the mid-located stations (in the range of 40-60 km) (Figure 5 (c)).

5. SUMMARY AND CONCLUSION

To save human lives and properties, accurate rainfall forecasting is significant for the areas that are prone to flash floods. The rainfall data of stations that are located under the umbrella of Muğla weather radar show that extreme events are likely to be seen in cold seasons. This study introduces an SVR technique for adjusting weather radar rainfall data belonging to 16 flood events. The results reveal that the SVR model with the L_1 norm is a promising alternative in correcting amounts of rainfall data. In the error-adjusting process, additive and multiplicative correction factors are used simultaneously.

The main conclusions of this study are the following:

- Since the data from weather radar is based on the scaling law formulation that is obtained from raindrop size distribution and each event has different atmospheric characteristics, the error improvement is not the same for all stations and shows a location-oriented performance.

- The values for the correction factor show that as the distance between the radar and the station increases, the values of the multiplicative factor show a decreasing trend and the majority of the values are less than one. Meanwhile, the values of additive factors indicate a variable trend in which stations that are located in the mid-range have the greatest ones, almost 1.6. Furthermore, all stations have additive factors greater than the value of 1. These results point out that although weather radars generally underestimate the rainfall amounts, the SVR algorithm with defined tolerated error makes a variation in additive factor instead of a multiplicative one.

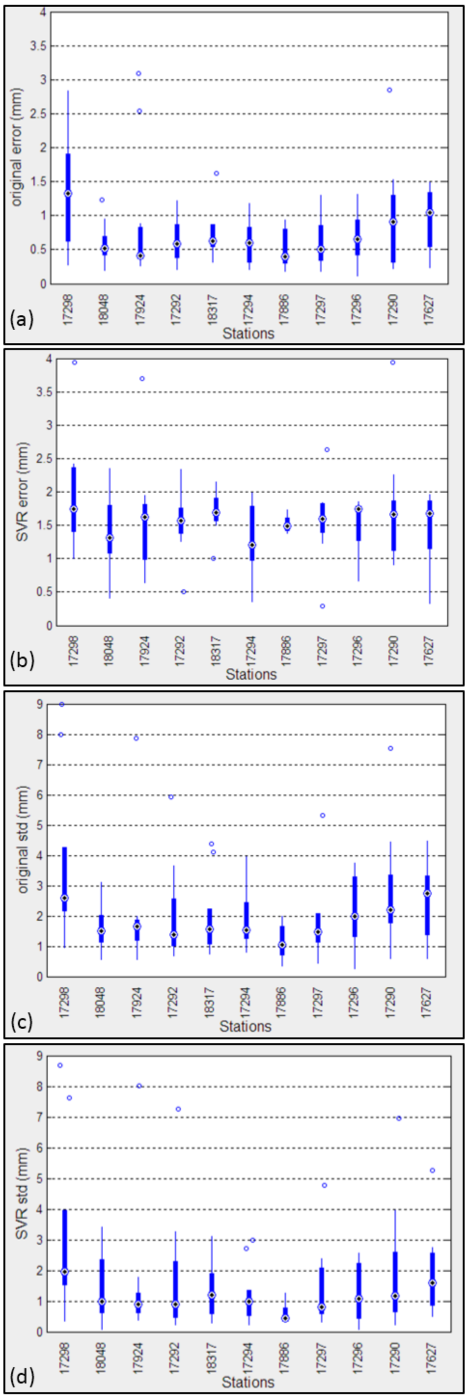

Figure 4 Error and standard deviation results for original and SVR applied datasets ((a) and (c) show the statistics related to original data, (b) and (d) show the SVR applied data, the order of stations in the x-axis is the same as the order of radar closeness given in the map)

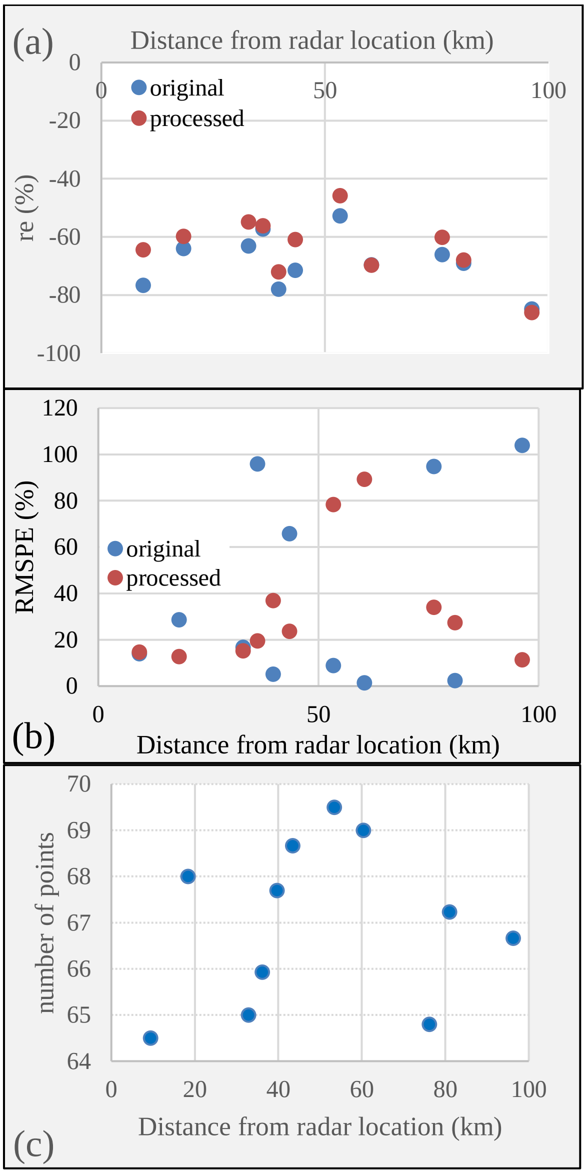

Figure 4 Statistical resultts for original and SVR applied (processed) datasets

- In the point of error reduction, it is discerned that the SVR has poor results independent of location. Contrary to this, the standard deviation of the error decreases for all stations.

- Since other norms, L_2 and L_∞ norms are not working well most probably owing to the non-Gaussian distribution of datasets, detailed investigations considering the physical properties of the topology and time will probably present a clearer conclusion. Moreover, some other optimization techniques, like ant search algorithms, particle swarm optimization algorithms, and genetic algorithms can be used for tuning the SVR parameters. For future studies, revealing the performance in determining SVR parameters by using different optimization techniques will be a challenging issue.

Acknowledgments

The radar and gauge rainfall data were provided by the Turkish State Meteorological Service (TSMS).

Funding

The authors have not received any financial support for the research, authorship, or publication of this study.

The Declaration of Conflict of Interest/ Common Interest

No conflict of interest or common interest has been declared by the authors

Authors' Contribution

The authors contributed equally to the study.

The Declaration of Ethics Committee Approval

This study does not require ethics committee permission or any special permission.

The Declaration of Research and Publication Ethics

The authors of the paper declare that they comply with the scientific, ethical, and quotation rules of SAUJS in all processes of the paper and that they do not make any falsification on the data collected. In addition, they declare that Sakarya University Journal of Science and its editorial board have no responsibility for any ethical violations that may be encountered and that this study has not been evaluated in any academic publication environment other than Sakarya University Journal of Science.

REFERENCES

[1] WMO, "WMO atlas of Mortality and Economic Losses from Weather, Climate and Water Extremes (1970–

2019)", 2021. [Online]. Available: https://library.wmo.int/doc_num.php?explnum_id=10989

[2] X. Sun, R. G. Mein, T. D. Keenan, J. F. Elliott, “Flood estimation using radar and raingauge data,” Journal of Hydrology, vol. 239, no. 1, pp. 4–18, 2000.

[3] C. Chwala, H. Kunstmann, “Commercial microwave link networks for rainfall observation: Assessment of the current status and future challenges,” WIREs Water, vol. 6, no. 2, pp. 1337, 2019.

[4] L. J. Battan, “Vertical Air Motions and the Z-R Relation,” Journal of Applied Meteorology, vol. 15, no. 10, pp. 1120–1121, 1976.

[5] I. Emmanuel, H. Andrieu, P. Tabary, “Evaluation of the new French operational weather radar product for the field of urban hydrology,” Atmospheric Research, vol. 103, pp. 20–32, 2012.

[6] I. V. Sideris, M. Gabella, R. Erdin, U. Germann, “Real-time radar–rain-gauge merging using spatio-temporal co-kriging with external drift in the alpine terrain of Switzerland,” Quarterly Journal of the Royal Meteorological Society, vol. 140, no. 680, pp. 1097–1111, 2014.

[7] G. Villarini, W. F. Krajewski, “Review of the Different Sources of Uncertainty in Single Polarization Radar-Based Estimates of Rainfall,” Surveys in Geophysics, vol. 31, no. 1, pp. 107–129. 2010.

[8] A. Ozkaya, A. E. Yılmaz, “Muğla Radarı Yağış Tahmin Hatalarının L2 ve L∞ Normlarıyla Analizi,” in International Symposium on Remote Sensing in Meteorology, İstanbul, Dec. 2021. [Online]. Available: https://meteouzal.itu.edu.tr/#/en/home

[9] V. Vapnik, The Nature of Statistical Learning Theory. Springer Science and Business Media, 1999.

[10] P.-F. Pai, W.-C. Hong, “A recurrent support vector regression model in rainfall forecasting,” Hydrological Processes, vol. 21, no. 6, pp. 819–827, 2007.

[11] W.-C. Hong, P.-F. Pai, “Potential assessment of the support vector regression technique in rainfall forecasting,” Water Resources Management, vol. 21, no. 2, pp. 495–513, 2007.

[12] Z. Zeng, W. W. Hsieh, A. Shabbar, W. R. Burrows, “Seasonal prediction of winter extreme precipitation over Canada by support vector regression,” Hydrology and Earth System Sciences, vol. 15, no. 1, pp. 65–74, 2011.

[13] A. Danandeh Mehr, V. Nourani, V. Karimi Khosrowshahi, M. A. Ghorbani, “A hybrid support vector regression–firefly model for monthly rainfall forecasting”, International journal of Environmental Science and Technology, vol. 16, no. 1, pp. 335–346, 2019.

[14] Y. Xiang, L. Gou, L. He, S. Xia, W. Wang, “A SVR–ANN combined model based on ensemble EMD for rainfall prediction,” Applied Soft Computing, vol. 73, pp. 874–883, 2018.

[15] A. G. Yilmaz, “The effects of climate change on historical and future extreme rainfall in Antalya, Turkey,” Hydrological Sciences Journal, vol. 60, no. 12, pp. 2148–2162, 2015.

[16] C.-C. Chuang, C.-C. Hsu, and C. W. Tao, “Embedded support vector

regression on Cerebellar Model Articulation Controller with Gaussian noise," Applied Soft Computing, vol. 11, no. 1, pp. 1126–1134, 2011.