Linear Fuzzy Space Based Framework for Air Quality Assessment

Endre Pap, Đorđe Obradović, Zora Konjović, Ivan Radosavljević
Singidunum University
Belgrade, Serbia
epap@singidunum.ac.rs,
djobrad@singidunum.ac.rs,
zkonjovic@singidunum.ac.rs,
iradosavljevic@singidunum.ac.rs

***Abstract*— Air quality is one of the most important issues mankind is facing today. There are different types of indices measuring the air pollution which are mainly based on aggregation functions. This paper proposes the model aimed at forecasting aggregated air pollution indices based on our theory of the linear fuzzy space. The proposed model consists essentially of two sub models. The first one models pollutants concentration, while the second one models multi-contaminant air quality index (FAQI). We model pollutants concentration by regression (XGBoost and deep neural network) utilizing fuzzy time series of two groups of data (measured concentrations and meteorological parameters). Multi-contaminant air quality index is modeled as an aggregation of Pollutant Standard Index (PSI) obtained via fuzzy linear transformation defined by fuzzy breakpoints. We also provide some preliminary results indicating model performance in terms of mean absolute errors for FAQI prediction**

Keywords- fuzzy set, linear fuzzy space, AQI index, aggregation operator.

I. INTRODUCTION

In the last decade mankind is facing air pollution as one of the most important issues causing very negative consequences on human health, but also on the economy of societies. By the data from 2016 of the World Health Organization (WHO), outdoor air pollution causes approximately 4.2 million deaths per year. By the European Environmental Agency (EEA) in 2018 in Europe the number of deaths related to concentrations of the particles PM2.5 was about 379,000. Therefore, there is a great need for the air pollution forecasting models expressing the air pollution as a simple value that is understandable for a wide audience.

Air pollution is an extremely complex spatio-temporally determined dynamic system distinctly characterized by the presence of imprecision and uncertainty.

To cope with uncertainty and imprecision, we use a fuzzy approach. More precisely, based on our previous results presented in [1, 2, 3, 4, 5, 6], we use mathematical models for basic concepts (fuzzy point, fuzzy spatial relation, fuzzy ordering, and fuzzy distance). For modelling temporal dimension of air pollution, we use a combination of time series models with techniques supporting the manipulation of imprecise and uncertain data, known under the umbrella term Fuzzy Time Series (FTS).

There are different types of the air quality index (AQI) expressing the air pollution as a single value, most of them based on some aggregation functions. Multi-contaminant oriented AQIs manage multiple effects due to the exposure to more pollutants, give more complete information on the possible impacts of air pollutants and give a direction for a more accurate, consistent, and comparable AQI system. Hence, we opt for multi-contaminant AQIs as a model of air pollution estimate.

In Section 2 we present preliminaries comprising linear fuzzy space theory, multi-contaminant AQIs, and fuzzy time series. Section 3 presents the proposed Framework, while Section 4 presents simulation results for the real data set consisting of 82457 samples (16 variables, 24h measurements). Section 6 brings conclusions, identifies model deficiencies, and outlines future research.

II. PRELIMINARIES

In this section we present the fundamentals of our model: linear fuzzy space theory, multi-contaminant fuzzy AQIs, and fuzzy time series.

A. *Linear fuzzy space*

In this subsection we present the fundamental concepts of the linear fuzzy space: fuzzy point, linear fuzzy space, fuzzy space ordering, and fuzzy space metrics as defined in [1].

Definition 1 *Fuzzy point* $P \in R^2$, denoted by $\tilde{P}$ is defined by its membership function $\mu_{\tilde{P}} \in \mathscr{F}^2$, where the set $\mathscr{F}^2$ contains all membership functions $u: R^2 \to [0,1]$ satisfying following conditions:

i) $(\forall \mu \in \mathscr{F}^2)(\exists_1 P \in R^2)\ \mu(P) = 1$,
ii) $(\forall X_1, X_2 \in R^2)(\lambda \in [0,1])\ \mu(\lambda X_1 + (1-\lambda)X_2) \geq min(\ \mu(X_1), \mu(X_2))$,
iii) function μ is upper semi-continuous,
iv) $[\mu]^\alpha = \{X | X \in R^2, \mu(X) \geq \alpha\}$ α-cut of function μ is convex.

Here, a point from R^2 with a membership function $\mu_{\tilde{P}}(P) = 1$, is denoted by P (P is the core of the fuzzy point

$\tilde{P}$), the membership function of point $\tilde{P}$ is denoted by $\mu_{\tilde{P}}$, while $[P]^\alpha$ stands for the α-cut (a set from R^2) of the fuzzy point.

Definition 2 R^2 *Linear fuzzy space* is the set $\mathscr{H}^2 \subset \mathscr{F}^2$ of all functions which, in addition to the properties given in Definition 1, are:

i) Symmetrical with respect to the core $S \in R^2$ $(\mu(S) = 1)$,
$\mu(V) = \mu(M) \wedge \mu(M) \neq 0 \Rightarrow d(S,V) = d(S,M)$,
where $d(S,M)$ is the distance in R^2.

ii) Inverse-linearly decreasing regarding points' distance from the core, i.e.:

If $r \neq 0$: $\mu_{\tilde{S}}(V) = \max\left(0, 1 - \frac{d(S,V)}{|r_S|}\right)$,

If $r = 0$: $\mu_{\tilde{S}}(V) = \begin{cases} 1 & if \quad S = V \\ 0 & if \quad S \neq V \end{cases}$,

where $d(S,V)$ is the distance between point V and the core S $(V, S \in R^2)$ and $r \in R$ is a constant.

Elements of that space are represented as ordered pairs $\tilde{S} = (S, r_S)$ where $S \in R^2$ is the core of $\tilde{S}$, and $r_S \in R$ is the distance from the core for which the function value becomes 0.

Measurement in the space, especially the distance between plane geometry objects is defined as a generalization of the concept of physical distance:

Definition 3 Let $\mathscr{H}^2$ be a linear fuzzy space and $\tilde{d}: \mathscr{H}^2 \times \mathscr{H}^2 \to \mathscr{H}^+$, $\mathrm{L}, \mathrm{R}: [0,1] \times [0,1] \to [0,1]$ be symmetric, associative, and non-decreasing for both arguments, and $\mathrm{L}(0,0) = 0$, $\mathrm{R}(1,1) = 1$. The ordered quadruple $(\mathscr{H}^2, \tilde{d}, \mathrm{L}, \mathrm{R})$ is called fuzzy metric space and the function $\tilde{d}$ is a *fuzzy metric*, if and only if the following conditions hold:

i) $\tilde{d}(\tilde{X}, \tilde{Y}) = \tilde{0} \Leftrightarrow [\tilde{X}]^1 = [\tilde{Y}]^1$

ii) $\tilde{d}(\tilde{X}, \tilde{Y}) = \tilde{d}(\tilde{Y}, \tilde{X})$, $\widetilde{\forall X}, \tilde{Y} \in \mathscr{H}^2$

(i) $\forall \tilde{X}, \tilde{Y} \in \mathscr{H}^2$:

$\tilde{d}(\tilde{X}, \tilde{Y})(s+t) \geq \mathrm{L}(d(x,z)(s), d(z,y)(t))$
if $s \leq \lambda_1(x,z) \wedge t \leq \lambda_1(z,y) \wedge s + t \leq \lambda_1(x,y)$

$\tilde{d}(\tilde{X}, \tilde{Y})(s+t) \leq \mathrm{R}(d(x,z)(s), d(z,y)(t))$
if $s \geq \lambda_1(x,z) \wedge t \geq \lambda_1(z,y) \wedge s + t \geq \lambda_1(x,y)$,

The α-cut of a fuzzy number $\tilde{d}(x,y)$ is given by
$[\tilde{d}(\tilde{X}, \tilde{Y})]^\alpha = [\lambda_\alpha(x,y), \rho_\alpha(x,y)]$ $(x, y \in R^+, 0 < \alpha \leq 1)$.

The fuzzy zero, $\tilde{0}$ is a non-negative fuzzy number with $[\tilde{0}]^1 = 0$.

Definition 4 Let $\mathscr{H}^2$ be a linear fuzzy space. Then, function $f: \mathscr{H}^2 \times \mathscr{H}^2 \times [0,1] \to \mathscr{H}^2$ called a *linear combination* of the fuzzy points $\tilde{A}, \tilde{B} \in \mathscr{H}^2$ is given by:

$f(\tilde{A}, \tilde{B}, u) = \tilde{A} + u \bullet (\tilde{B} - \tilde{A})$,

where $u \in [0,1]$ and the operator $\bullet$ is the scalar multiplication of the fuzzy point.

B. Fuzzy Air pollution indices

Aggregation is a process of combining several numerical values into a single representative. The corresponding numerical function is called aggregation function (aggregation operator). It has some natural properties as monotonicity and boundary conditions. Aggregation applies to various fields: applied mathematics, computer science, economy and finance, pattern recognition and image processing, data fusion, multicriteria decision aid, automated reasoning, etc. see [6, 7, 9]. In practice, usually the data is normalized, so the definition of aggregation becomes:

Definition 5. An aggregation function (operator) is a function $A^{(n)}: [0,1]^n \to [0,1]$ which satisfies the following conditions

1. is nondecreasing (in each variable)
2. $A^{(n)}(0, \ldots, 0) = 0$ and $A^{(n)}(1, \ldots, 1) = 1$.

Some well-known examples are arithmetic mean, geometric mean, harmonic mean, triangular norms, etc. (see [6, 7, 9]). More sophisticated aggregation functions modelling the interaction between criteria are managed by monotone set functions and corresponding integrals [6, 7, 8].

One possible application of the aggregation functions is measuring the air quality by air quality index (AQI) [10]. What follows is the simplest model where a sub-index (AQI_i) is calculated for each pollutant *i* by the following linear interpolation formula

$$ACI_i = \frac{I_{high} - I_{low}}{C_{high} - C_{low}}(C - C_{low}) + I_{low} .$$

Here C is the monitored ambient average concentration of pollutant *i*; C_{low} is the breakpoint lower than or equal to C; C_{high} is the breakpoint higher than or equal to C; and I_{low} and I_{high} are the sub-index values corresponding to C_{low} and C_{high}, respectively. The overall AQI is then calculated as a simple max aggregation:

$$AQI = \max_{i=1}^{m}(AQI_i).$$

There are further investigations for new aggregation functions, which involve the influence of several pollutants [9, 11, 12, 13, 14, 20]. Among these AQIs, arithmetic pollutant aggregation integrates pollutants in a linear or nonlinear way, and weighted pollutant aggregation further assigns varied weights from different approaches. The general air quality health index ($GAQHI$) is proposed as a pollutant-aggregated, local health-based AQI paradigm suitable for the representing complex multi-contaminant situation:

$$Is = \left(\sum_{i=1}^{n}(AQI_i)^\alpha\right)^{\frac{1}{\alpha}},$$

where $\alpha \in [1, \infty]$.

Other modifications of the EPA AQI are proposed in [12], giving a new index $RAQI$ which is the product of three terms:

$$RAQI = F_1 * F_2 * F_3$$

were

$$F_1 = \max(I_i), \qquad i = 1{,}5$$

$$F_2 = \frac{\sum_{i=1}^{5} Ave_{daily}(I_i)}{Ave_{annual} \bullet \left(\sum_{i=1}^{5} Ave_{daily}(I_i)\right)}$$

and the Shannon entropy function is introduced in the third term:

$$F_3 = \frac{Ave_{annual} \bullet Entropy_{daily}\left(\max_{i=1}^{5}(I_i)\right)}{Entropy_{daily}\left(\max_{i=1}^{5}(I_i)\right)}$$

Here, the first factor F_1 represents the main effect of the air pollution on the human body current PSI value to decrease the eclipse aberration. The second factor accounts for the single contribution of the subindices to the final one to mitigate eclipsicity, and the third helps to avoid both ambiguity and eclipsicity[1].

Each of these models can be easily fuzzified by simply mapping crisp space to a linear fuzzy space by the suitable fuzzy function over the crisp domain.

There are also results that employ fuzzy logic for modelling air quality indices like [11, 15].

In the paper [11] the input variables are air pollutant criteria (5 in total), and output variable is fuzzy AQI. The fuzzification process is defined via the boundary values of the universal sets and the corresponding fuzzy sets (trapezoidal for input variables and triangular for output). The rule base representing the relationship between input variables and output variables contains 243 rules. The max-min inference strategy and centroid method are decided for the inference and defuzzification process.

In the paper [15] ten parameters (selected pollutants concentrations) are divided into two groups. At the first step, the parameters in each group were processed by the inference systems, and then grouped and normalized between 0 and 100, resulting in two new groups. These new groups were processed in the second step by new inference systems and resulted in the FAQI. All rules (72 in total) have only one antecedent. A classical fuzzy inference system (Mamdani) is used.

Common to those results is a reliance on an approach that utilizes fuzzy rule-based inference and does not include a time series-based prediction.

C. Fuzzy time series

Most of the real-world tasks that utilize time series rely on multivariate time series models [17, 18, 19, 20]. The common multivariate time series model is [17]:

Let $Z_t = \left[Z_{1,t}, Z_{2,t}, \dots, Z_{m,t}\right]'$, be an m-dimensional jointly stationary real-valued vector process such that $E\left(Z_{i,t}\right) = \mu_i$ is a constant for each $i = 1,2,\dots,m$ and the cross-covariances between $Z_{i,t}$ and $Z_{j,s}$ for all $i = 1,2,\dots,m$ and $j = 1,2,\dots,m$ are functions only on the time difference $(s - t)$.

On the other hand, the original definition of the univariate first fuzzy time series model is [20]:

Definition 6. Let $Y(t)(t = \cdots,0,1,2,\dots)$, a subset of R^1 be the universe of discourse on which fuzzy sets $f_i(t)(i = 1,2,\dots)$ are defined and $F(t)$ is the collection of $f_i(t)(i = 1,2,\dots)$. Then $F(t)$ is called a fuzzy time series on $Y(t)(t = \cdots,0,1,2,\dots)$.

Our time series model is a combination of the previous two where we basically apply the same common multivariate model which is modified to support imprecise values. In our model [2], we simply replace a crisp point with a linear fuzzy space point.

Definition 7. Let $Y(t)(t = \cdots,0,1,2,\dots)$, a subset of R^1 be the universe of discourse. Let $H^l(l = 1,2)$ be a linear fuzzy space. Furthermore, let $f_i(t)(i = 1,2,\dots)$ be fuzzy sets defined as points on a linear fuzzy space over the given universe of discourse, and $\tilde{F}_j(t)(j = 1,2,\dots m)$ be collections of these fuzzy points. Then, $\tilde{F}_t = \left[\tilde{F}_{1,t}, \tilde{F}_{2,t}, \dots, \tilde{F}_{m,t}\right]'$ is called a linear fuzzy space based fuzzy time series on $Y(t)(t = \cdots,0,1,2,\dots)$.

This definition enables all features of linear fuzzy space to be utilized. For example, a process vector can be of a mixed type (some components can be crisp, some can be fuzzy linguistic or non-linguistic) Also, spatial relations defined on the linear fuzzy space hold, providing for efficient representations and processing of measurement errors.

In our model, machine learning techniques can be used to create complex, non-linear relations.

III. Fuzzy model of Air pollution indices prediction

In this example we demonstrate how the linear fuzzy space is used for time series-based forecasting. Fuzzy time series defined by means of the fuzzy linear space as described in subsection C are used for modelling air quality forecasting.

A. Data model

Data model used in this paper consists of temporal georeferenced samples. Each sample is a timeseries covering previous 24h in 1h sample rate (total 385 real values). Each timeseries corresponds to one variable.

Variables are divided in two groups, meteorological parameters GDAS (The Global Data Assimilation System) and six common air pollutants known as “criteria air pollutants”. GDAS parameters [21] are described by Table I, while criteria air pollutants [22, 23] are given in Table II.

TABLE I. GDAS parameters

ID	Description	Unit
PRSS	Pressure at surface	hPa
TPP6	Accumulated precipitation (6 h accumulation)	m
RH2M	Relative Humidity at 2m AGL	%
TO2M	Temperature at 2m AGL	K
TCLD	Total cloud cover (3- or 6-h average)	%
U10M	U-component of wind at 10 m AGL	m/s
V10M	V-component of wind at 10 m AGL	m/s
TMPS	Temperature at surface	K
PBLH	Planetary boundary layer height	m
irradiance[2]	Irradiance/solar power	W/m2

[1] Ambiguity and eclipsicity here refer to effects caused by using max operator mainly.

[2] This is a calculated value

TABLE II. AIR POLLUTANTS PARAMETERS

ID	Description	Unit
PM10	Suspended particles smaller than 10 μm	μg/m^3
PM25	Suspended particles smaller than 2.5 μm	μg/m^3
SO2	Sulfur dioxide	ppb
CO	Carbon Monoxide	ppm
NO2	Nitrogen Dioxide	ppb
O3	Ground-level Ozone	ppm

B. Linear fuzzy space based air pollution index

Since air pollutants are measured in different physical units and scales, first step is to transform them into a common domain (0-500). This transformation is usually defined by breakpoints tables and resulting values are called Pollutant Standard Index (PSI). Instead of using discrete functions, we propose fuzzy linear transformation defined by fuzzy breakpoints (Fig. 1).

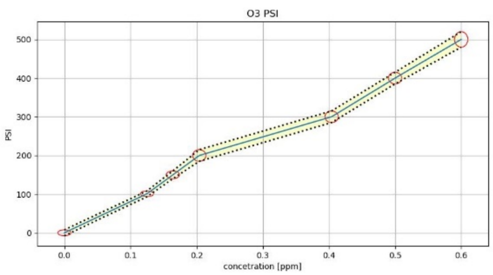

Figure 1. Fuzzy linear transformation

Fuzzy linear transformation is defined by an ordered list of 2D Fuzzy points $\tilde{P} = (\tilde{X}, \tilde{Y})$. Each 2D fuzzy point consists of two components $\tilde{X} = (X, r_x)$ and $\tilde{Y} = (Y, r_Y)$ which are 1D fuzzy points. Then, fuzzy PSI (FPSI) is defined as:

$$\widetilde{FPSI_\iota} = \widetilde{\iota nterp}\left(C, \left[\widetilde{P_0}, \dots \widetilde{P_n}\right]\right) = (\ FPSI_i, r_{FPSI}),$$

$$FPSI_i = \frac{Y_{high} - Y_{low}}{X_{high} - X_{low}}(C - X_{low}) + Y_{low}$$

$$r_{PSI} = \frac{r_{Yhigh} - r_{Ylow}}{r_{Xhigh} - r_{Xlow}}(C - X_{low}) + r_{Ylow}$$

where $\widetilde{\iota nterp}$ is a fuzzy linear transformation from concertation fuzzy space into index fuzzy space. Fuzzy points $\widetilde{P_{h\iota gh}}$ and $\widetilde{P_{low}}$ are fuzzy points whose roots of $\widetilde{X}$ components are nearest to the concertation $\boldsymbol{C}$.

$FPSI$ can be further represented by a linguistic variable, or it can be used directly in aggregation process.

A single fuzzy value $FAQI$ is obtained by applying some fuzzy aggregation operator (aggreg) to all (n) component $FPSI$ indices:

$$FAQI = aggreg(FPSI_i), i = 1, n$$

To simplify a decision-making process and/or facilitate understanding for humans, a fuzzy linguistic variable defined by corresponding fuzzy sets can be easily introduced in such model.

C. Prediction model

In our model we opt for multivariate regression to forecast $FAQI$ (Fig. 2). However, other classification method can be easily incorporated in the proposed model.

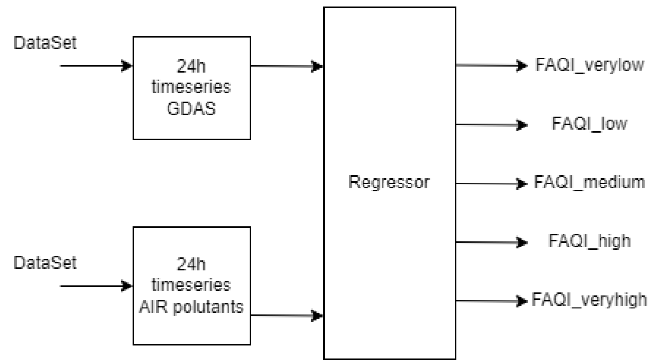

Figure 2. Prediction model

IV. MODEL APPLICATION

To present proposed model/methodology we run one experiment on large and diverse data set. Dataset contains more than 82000 samples each with 385 real values. GDAS values are interpolated to fit five geo locations, merged with air pollutant measurement.

A. Data set

In this experiment we used five datasets from five different locations in USA, each in same format. The sources of data are [21] and [23]. Samples are indexed by temporal attribute measurement datetime from January 1st , 2015 to December 31st, 2021. All ten meteorological GDAS and six air pollutants are stored in 24 hours' time slot with 1h sample rate (385 real values in total). Table III shows data in more details (sample sizes per locations). The same source provides data about land use (COMMERCIAL, RESIDENTIAL) and type of location (URBAN, SUBURBAN, RURAL) as shown in Table IV. These data are not yet used in our experiment.

TABLE III. DATA SETS

site_id	site	Samples
11-001-0043	Washington, DC	27,981
13-089-0002	Near Atlanta, GA	21,468
18-097-0078	Indianapolis, IN	16,774
22-033-0009	Baton Rouge, LA	6,569
32-003-0540	Las Vegas, NV	9,665

TABLE IV. SITE TYPES

ID	City	Land use	Location
11-001-0043	Washington, DC	COMMERCIAL	URBAN
13-089-0002	Near Atlanta, GA	RESIDENTIAL	SUBURBAN
18-097-0078	Indianapolis, IN	RESIDENTIAL	SUBURBAN
22-033-0009	Baton Rouge, LA	COMMERCIAL	URBAN
32-003-0540	Las Vegas, NV	RESIDENTIAL	URBAN

PSI calculation was done by using PSI functions, which transform physical value domain into real value interval [0, 500] (Table V).

TABLE V. PSI BREAKPOINTS

PSI	PM10 μg/m^3	SO2 ppm	CO ppm	NO2 ppm	O3 ppm
0	0	0	0	0	0
50	50	0.03	4.5	-	0.06
100	150	0.14	9	-	0.12
200	350	0.3	15	0.6	0.2
300	420	0.6	30	1.2	0.4
400	500	0.8	40	1.6	0.5
500	600	1	50	2	0.6

B. Fuzzy Air Quality index

In our framework the fuzzy air quality index is modelled via a simple max aggregation function applied to five $FPSI$ indices of each criteria air pollutants:

$$FAQI = \max\left(FPSI_{\mathrm{CO}}, FPSI_{\mathrm{PM10}}, FPSI_{NO2}, FPSI_{O3}, FPSI_{SO2}\right)$$

Finally, we introduce a fuzzy linguistic variable (*very low*, *low*, *medium*, *high*, *very high*) defined by corresponding fuzzy sets as depicted on Fig. 3.

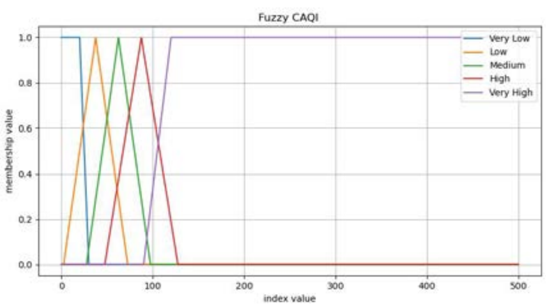

Figure 3. Fuzzy CAQI

C. Prediction

In our experiments we applied two multivariate regressors. The first multivariate predictor regressor in this experiment was XGBoostRegressor with 24*10 GDAS and 24*6 AirPollutatns variables as input and 5 real valued outputs each corresponding to a single fuzzy set (FAQI_*verylow* to FAQI_*veryhigh*) as depicted on Fig. 3. Data set is splitted up into train (80%) and test (20%) subsets and trained with 1000 estimators with *max_depth* 4 and turned on early stopping method to avoid overfitting.

The second multivariate predictor regressor in this experiment was a deep neural network with 24*10 GDAS and 24*6 AirPollutatns variables as input and 5 real valued outputs each corresponding to a single fuzzy set (FAQI_*verylow* to FAQI_*veryhigh*), one hidden layer with 20 ReLU nodes. Activation functions in output layer are Sigmoid. Data set is splitted up into train (80%) and test (20%) subsets. Two dropout layers with 10% random filters are incepted between active layers to prevent overfitting.

D. Prediction results

Mean absolute errors for FAQI prediction are shown in Table VI (XGBoost) and Table VII (deep neural network).

TABLE VI. XGBOOST

ID	verylow	low	medium	high	veryhigh
11-001-0043	0.229	0.230	0.049	0.003	0.001
13-089-0002	0.239	0.217	0.036	0.004	0.001
18-097-0078	0.213	0.213	0.065	0.006	0.001
22-033-0009	0.224	0.218	0.063	0.009	0.000
32-003-0540	0.083	0.236	0.185	0.045	0.012

TABLE VII. DEEP NEURAL NETWORK

ID	verylow	low	medium	high	veryhigh
11-001-0043	0.405	0.399	0.065	0.003	0.001
13-089-0002	0.460	0.398	0.045	0.004	0.002
18-097-0078	0.425	0.406	0.074	0.005	0.002
22-033-0009	0.426	0.392	0.056	0.008	0.001
32-003-0540	0.109	0.362	0.291	0.039	0.010

What could be concluded from the above tables is that both regressors behave similarly. Moreover, they are good in prediction for categories *medium*, *high* and *veryhigh* and poor in prediction for categories *verylow* and *low*. This means that this model should be improved to perform equally good for all categories. But, having in mind one of the main purposes of the FAQI to warn of dangerous air pollution (*high* and *veryhigh*, possibly *medium*) the results point out that further research deserves attention.

V. CONCLUSION

This paper proposes the model aimed at forecasting aggregated air pollution index founded on our theory of the linear fuzzy space. The proposed model consists essentially of two sub models. The first models pollutants concentration, while the second one models multi-contaminant air quality index. We model pollutants concentration by regression utilizing fuzzy time series of two groups of data, measured pollutants concentrations and meteorological parameters. Multi-contaminant air quality index is modeled as a fuzzy aggregation of PSI obtained via fuzzy linear transformation defined by fuzzy breakpoints.

Preliminary results show that our model is characterized by a distinct property which is a good performance for higher values of air quality index, and significantly worse (mean absolute errors higher for order of magnitude) performance for lower values. This is a notable deficiency of the model calling for improvement that will ensure equally good performance for all categories.

On the other hand, air pollution is an extremely complex dependency among many factors (air pollutants, environment, time, climate conditions, etc.) additionally burdened with uncertainty and imprecision in data. All this makes a single index a rough approximation of the considered pollution situation only.

Indeed, there is a large space for improvements in the research topics tackled in this paper which shapes further research directions. This space could be partitioned in two rough partitions. The first, which is of fundamental kind, is about rethinking the air pollution index concept (for example, make it contextually dependent, or make it multidimensional). The second one is about "local" issues concerning the improvement of the model proposed in this paper. Those could comprise introduction of new parameters (like those in Table IV), training data balancing, learning shapes of membership functions from historical data, etc. The two partitions intersect at utilization of AI methods, particularly fuzzy approach and machine learning techniques.

ACKNOWLEDGMENT

The authors acknowledge funding provided by the Science Fund of the Republic of Serbia #GRANT No. 6524105, AI - ATLAS.

REFERENCES

[1] Obradović, Đ., "Model of the computer simulation system for managing geospace under uncertainty conditions," PhD thesis, Faculty of Technical Sciences University of Novi Sad (in Serbian), 2011.

[2] Obradović Đ., Konjović, Z., Pap, E, Šoštarić. "The linear fuzzy space: Theory and applications. Pap, E. (Ed.) Artificial Intelligence: Theory and Applications, " pp 227-253, Springer, 2021.

[3] Obradović, Đ., Konjović, Z., Pap, E., Rudas, I. "Fuzzy Geometry in Linear Fuzzy Space, Intelligent Systems: Models and Applications," Revised and Selected Papers from the 9th IEEE International Symposium on Intelligent Systems and Informatics SISY 2011. Springer Berlin Heidelberg, pp 137-153, 2013.

[4] Obradović, Đ., Konjović, Z., Pap, E. , Ralević, N. "The maximal distance between imprecise point objects," Fuzzy Sets and Systems 170 (1), pp. 76-94, 2011, doi: https://doi.org/10.1016/j.fss.2010.12.005

[5] Obradović, Đ., Konjović, Z., Pap, E. , Rudas, I. J. Linear Fuzzy Space Based Road Lane Detection, Knowledge Based Systems 38 (1), pp. 37-47, 2013, doi: https://doi.org/10.1016/j.knosys.2012.01.002

[6] Grabisch, M., Marichal, J. L., Mesiar, R., Pap, E. "Aggregation functions (Encyclopedia of mathematics and its applications, series number 127)", Cambridge University Press, 2009.

[7] Klement, E. P., Mesiar, R., Pap, E., "Triangular norms", Kluwer, 2000.

[8] Pap, E. (Ed.), "Artificial Intelligence: Theory and Applications," Springer, 2021.

[9] Pap, E. (Ed.), "Intelligent Systems: Models and Applications. Topics in Intelligent Engineerinng and Informatics 3," Springer, 2013.

[10] Wikipedia, "Air Quality Index", November 2021, Available: https://en.wikipedia.org/wiki/Air_quality_index, [Accessed: December 2021]

[11] Atacak, I., Arici,N., Guner, D. "Modelling and evaluating air quality with fuzzy logic algorithm-Ankara-Cebeci sample", Int J Intell Syst Appl Eng, vol. 5, no. 4, pp. 263–268, Dec. 2017.

[12] Cheng, W., "Revised air quality index derived from an entropy function," Atmospheric Environment, 38(3), pp 383-391, 2004

[13] Fino, A., Vichi, F., Leonardi, Ch., Mukhopadhyay, K., "An overview of experiences made and tools used to inform the public on ambient air quality," Atmosphere 12(1), pp 1: 1524, 2021, Available: https://doi.org/10.3390/atmos12111524, [Accessed: January 12, 2022].

[14] Kyrkilis, G., Chaloulakou, A., Kassomenos, P.A.: "Development of an aggregate air quality index for an urban mediterranean agglomeration: Relation to potential health effects, " Environ. Int. 33, 670–676, 2007.

[15] Sowlat, M. H., Gharibi, H., Yunesian, M., Mahmoudi, M. T., Lotfi, S. "A novel, fuzzy-based air quality index (FAQI) for air quality assessment, Atmospheric Environment," 45(12), pp. 2050-2059, 2011, https://doi.org/10.1016/j.atmosenv.2011.01.060.

[16] Tan, X., Han, L., Zhang, X., Zhou, W., Li, W., Qian, Y., "A review of current quality indexes and improvements under the multi-contaminant air pollution exposure," J. Environ. Manag. 279, https://doi.org/10.1016/j.jenvaman.2020.111681 2021.

[17] Wei, W. W. S. "Multivariate Time Series Analysis and Applications", | John Wiley & Sons Ltd., Book Series: Wiley Series in Probability and Statistic, 2019, DOI:10.1002/9781119502951

[18] Song, Q., Chissom, B.S. "Fuzzy time series and its model", Fuzzy Sets and Systems 54(3):269–277, March 1993, DOI:10.1016/0165-0114(93)90372-O

[19] Panigrahi S., Behera, H.S.: "Fuzzy Time Series Forecasting: A Survey, " In: Behera H., Nayak J., Naik B., Pelusi D. (eds) Computational Intelligence in Data Mining. Advances in Intelligent Systems and Computing, vol 990. Springer, Singapore, 2020, https://doi.org/10.1007/978-981-13-8676-3_54

[20] Sahin A., Dodurka, M. F., Kumbasar, T., Yesil, E., Siradag, S., "Review Study on Fuzzy Time Series and Their Applications in the Last Fifteen Years," International Fuzzy Systems Symposium (FUZZYSS'15), Istanbul, 4, pp 166-171, 2015.

[21] NOAA-Air Resources Laboratory, Global Data Assimilation System (GDAS1) [Dataset], Archive began: December 1th 2004, https://www.ready.noaa.gov/gdas1.php, [Acessed: January 10, 2022]

[22] Huetteman, J., Johnson, T., Schreifels, J., "Using eGRID for environmental footprinting of electricity purchases," 2021, Available: https://www.epa.gov/egrid/download-data, [Accessed: January 12, 2022].

[23] EPA United States Environment Protection Agency, eGRID2020 [Dataset], January 27th 2022, https://www.epa.gov/egrid/download-data [Accessed: February 1, 2022].