Linear Fuzzy Space Based Framework for Air Quality Assessment

Endre Pap, Đorđe Obradović, Zora Konjović, Ivan Radosavljević
Singidunum University
Belgrade, Serbia

Emails: {epap, djobradovic, zkonjovic, iradosavljevic}@singidunum.ac.rs

***Abstract*— Air quality is one of the most critical issues humankind is facing today. There are diverse types of indices measuring the air pollution which are mainly based on aggregation functions. This paper proposes a model aimed at forecasting aggregated air pollution indices based on our theory of the linear fuzzy space. The proposed original model consists essentially of two sub models. The first one models concentrations of pollutants, while the second one models Air Quality Index (AQI). We model concentrations of pollutants 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 modelled as an aggregation of Pollutant Standard Index (PSI) obtained via fuzzy linear transformation defined by fuzzy breakpoints. Some preliminary results are presented indicating model performance in terms of prediction mean absolute errors.**

Keywords: fuzzy set; linear fuzzy space; AQI index; aggregation operator.

I. INTRODUCTION

In the last decade, mankind has been facing air pollution as one of the most important issues with adverse effect on human health, but also on the economy of societies. According to WHO's annual World Health Statistics report from 2016, outdoor air pollution causes approximately 4.2 million deaths per year [1]. As reported by the European Environmental Agency (EEA) in 2018, the number of deaths in Europe related to concentrations of the particles PM2.5 was about 379,000 [2]. Therefore, there is a great need for the air pollution forecasting models which will express 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. Therefore, it is not easy to give a precise air pollution forecast, which would be of great importance for public health.

To cope with uncertainty and imprecision, we use a fuzzy approach. More precisely, the one based on our previous results presented in [3] - [8], where we use mathematical models for basic concepts: fuzzy point, fuzzy spatial relation, fuzzy ordering, and fuzzy distance. For modelling the 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). This model enables a more adequate air pollution forecast.

Multi-contaminant Air Quality Index (AQI) manages multiple effects due to the exposure to more pollutants, gives more complete information on the possible impacts of air pollutants and 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.

The rest of the paper is organized in four sections. Section 2 presents theoretical foundations together with related work, while Section 3 presents the model of the proposed framework. Section 4 presents the simulation results for the real data set (82457 samples/16 variables/24h measurements) while Section 5 summarizes the research results, identifies model deficiencies, and outlines future research.

II. PRELIMINARIES

The theoretical foundations of our model are based on 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, fuzzy space metrics, and fuzzy linear combination, as defined in [3] – [7].

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

i) $(\forall \mu \in \mathcal{F}^2)(\exists_1 P \in R^2)\ \mu(P) = 1,$
ii) $(\forall X_1, X_2 \in R^2)(\lambda \in [0,1])\ \mu(\lambda X_1 + (1-\lambda)X_2) \geq min\big(\mu(X_1), \mu(X_2)\big),$
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 $\mathcal{H}^2 \subset \mathcal{F}^2$ 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.

 ISBN: 978-1-61208-977-5

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.

The 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 $\mathcal{H}^2$ be a linear fuzzy space and $\tilde{d}: \mathcal{H}^2 \times \mathcal{H}^2 \to \mathcal{H}^+$, $\mathsf{L}, \mathsf{R}: [0,1] \times [0,1] \to [0,1]$ be symmetric, associative, and non-decreasing for both arguments, and $\mathsf{L}(0,0) = 0$, $\mathsf{R}(1,1) = 1$. The ordered quadruple $(\mathcal{H}^2, \tilde{d}, \mathsf{L}, \mathsf{R})$ 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) $\tilde{d}(\tilde{X}, \tilde{Y}) = \tilde{d}(\tilde{Y}, \tilde{X})$, $\widetilde{\forall X}, \tilde{Y} \in \mathcal{H}^2$

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

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

$\tilde{d}(\tilde{X}, \tilde{Y})(s+t) \leq \mathsf{R}\big(d(x,z)(s), d(z,y)(t)\big)$
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
$\left[\tilde{d}(\tilde{X}, \tilde{Y})\right]^\alpha = [\lambda_\alpha(x,y), \rho_\alpha(x,y)]$ $(x, y \in R^+, 0 < \alpha \leq 1)$.

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

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

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

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

B. *Fuzzy Air pollution indices*

As shown in [7], a multi-contaminant model of AQI, in which aggregation functions (aggregation operators) are applied to combine several numerical values into a single representative, is predominant by far. An aggregation operator has natural properties such as monotonicity and boundary conditions. In practice, the data is usually 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 $A^{(n)}(1, \ldots, 1) = 1$.

Aggregation applies to various fields and takes diverse forms, from the simple to quite sophisticated ones, modelling the interaction between criteria which are managed by monotone set functions and corresponding integrals [8], [10-12] that will be incorporated in our future research.

The simplest AQI model calculates a sub-index (AQI_i) for each pollutant i by the following linear interpolation formula:

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

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 is ongoing research for new aggregation functions, which involve the influence of multiple pollutants (see [13] - [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 representing a complex multi-contaminant situation:

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

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

An interesting modification of the United States Environmental Protection Agency (EPA) AQI is proposed in [13], giving a new index $RAQI$ which is the product of three terms:

$$RAQI = F_1 * F_2 * F_3$$

were

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

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

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

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

This model strives to avoid ambiguity (indicating a less polluted air as highly polluted) and ellipticity (indicating highly polluted air as less polluted) by introducing entropy.

In addition, there are interesting approaches like in [14] that model the air pollution index via a mixture of distributions based on its structure and descriptive status. Each of these models can be easily fuzzified by simply mapping crisp space to a linear fuzzy space by the suitable fuzzy function over a crisp domain.

There are also results that utilize fuzzy logic for modelling air quality indices, like those presented in [15], [16], [19], and [20].

In [15], the input variables are air pollutant criteria (PM10, SO2, CO, NO2, O3), and the 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, and triangular for output variables). The rule base representing the relationship between input variables and output variables contains 243 rules. The max-min inference strategy and centroid method are chosen for the inference and defuzzification process.

In [16], ten parameters (selected concentrations of pollutants) are divided into two groups. Firstly, the parameters in each group are processed by the inference systems, and then grouped and normalized between 0 and 100, resulting in two new groups. These new groups are processed in the second step by new inference systems and result in the Fuzzy Air Quality Index (FAQI). All rules (72 in total) have only one antecedent. A classical fuzzy inference system (Mamdani) is used for this purpose. 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.

In [19], the authors present a comparative study of the results obtained from several models for air pollution index forecast which shows that the fuzzy time series models outperformed the other models in terms of forecasting accuracy and computation time. Finally, [20] utilizes a fuzzy time series-Markov chain model for predicting the daily air pollution index.

C. Fuzzy time series

Most of the real-world tasks that utilize time series rely on multivariate time series models [21]-[24]. The common multivariate time series model is [21]:

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 $E(Z_{i,t}) = \mu_i$ 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 of the time difference $(s - t)$.

On the other hand, the original definition of the univariate first fuzzy time series model is [24]:

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 apply the same common multivariate model which is modified to support imprecise values. In our model, we simply replace a crisp point with a linear fuzzy space point [2]:

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) whilst spatial relations defined on the linear fuzzy space hold.

In our framework, diverse 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 to model air quality forecast.

A. Data model

The data model used in this paper consists of temporal georeferenced samples. Each sample is a time series covering the previous 24h in 1h sample rate (total 385 real values). Each time series corresponds to one variable. Variables are divided in two groups: meteorological parameters, the Global Data Assimilation System (GDAS), and six common air pollutants known as “criteria air pollutants.”

GDAS parameters [25] are described in Table I.

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	Irradiance/solar power	W/m2

Criteria air pollutants [26] are given in Table II.

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	Sulphur 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, the first step is to transform them into a common domain (0-500). This transformation is usually defined by breakpoint tables and the resulting values are called Pollutant Standard Index (PSI). Instead of using discrete functions, we propose a fuzzy linear transformation defined by fuzzy breakpoints (Figure 1).

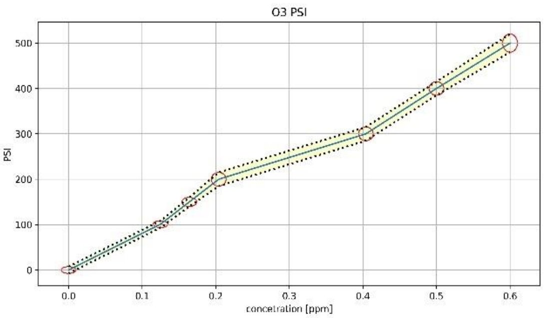

Figure 1. Fuzzy linear transformation

A 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:

$$\widetilde{FPSI_i} = \widetilde{linterp}\left(C, \left[\tilde{P_0}, \ldots \tilde{P_n}\right]\right) = (\, FPSI_i, r_{FPSI}) \,,$$

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

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

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

$FPSI$ can be further represented by a linguistic variable, or it can be used directly in the 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 the decision-making process and/or facilitate general understanding, a fuzzy linguistic variable defined by corresponding fuzzy sets can be easily introduced in such a model.

C. Prediction model

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

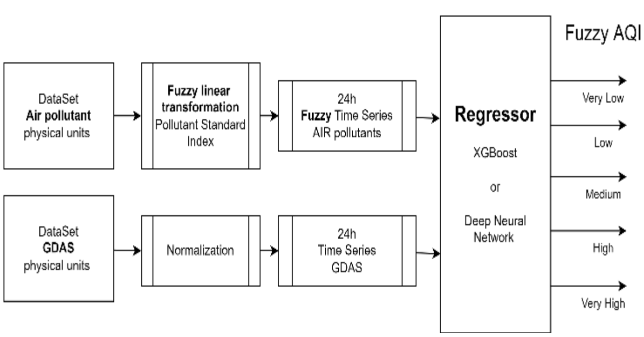

Figure 2. Prediction model

IV. MODEL APPLICATION

In order to present the proposed model/methodology, we ran one experiment on a large and diverse data set. The used data set contains more than 82000 samples each with 385 real values. GDAS values are interpolated to fit five geo locations and merged with measurements of the concentration of the air pollutants.

A. Data set

In this experiment, we used five data sets from five distinct locations in USA, each in the same format. The sources of data are [25] and [26]. Samples are indexed by temporal attribute, datetime, ranging from January 1, 2015 to December 31, 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 presents the data in more details (sample sizes per locations).

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

The same source provides data about land use (COMMERCIAL, RESIDENTIAL) and type of location (URBAN, SUBURBAN, RURAL), as shown in Table IV.

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 using PSI functions (Table V), which transform the physical value domain into a real value interval [0, 500].

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:

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

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

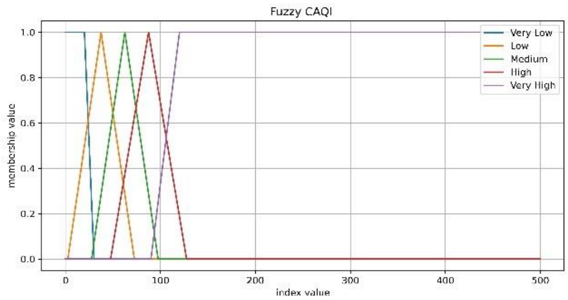

Figure 3. Fuzzy CAQI

C. Prediction

In our experiments, we applied two multivariate regressors. The first multivariate predictor regressor in this experiment is XGBoostRegressor with 24*10 GDAS and 24*6 air pollutant variables as input, and 5 real valued outputs, each corresponding to a single fuzzy set (FAQI_*very low* to *very high*), as depicted in Figure 3. Data set is split up into train (80%) and test (20%) subsets and trained with 1000 estimators with *max_depth* 4 and enabled early stopping method to avoid overfitting.

The second multivariate predictor regressor in this experiment is a deep neural network with 24*10 GDAS and 24*6 air pollutant variables as input, and 5 real valued outputs, each corresponding to a single fuzzy set (FAQI_*very low* to *very high*), with one hidden layer consisting of 20 Rectified Linear Units (ReLU) nodes. The activation functions in output layer are Sigmoid. The data set is split 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

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

TABLE VI. XGBOOST

ID	very low	low	medium	high	very high
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	very low	Low*	medium	high	very high
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

Tables VI and VII show that both regressors behave similarly. Moreover, they are good in prediction for categories *medium*, *high* and *very high* and poor in prediction for categories *very low* and *low*. Having that in mind and the main purpose of the FAQI to alarm of dangerous air pollution (*high* and *very high*, possibly *medium*), the results indicate that further research is needed and justified.

V. CONCLUSION

This paper proposes a model aimed at forecasting the aggregated air pollution index that is based on our theory of the linear fuzzy space. The proposed model consists of two sub models. The first one models the concentration of pollutants, while the second one models multi-contaminant air quality index. We model the concentration of pollutants by regression, utilizing fuzzy time series of two groups of data: measured concentrations of pollutants and meteorological parameters. The 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 an 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.

Air pollution is a result of an extremely complex and interdependent interaction among multiple factors (air pollutants, environment, time, climate conditions, etc.) additionally burdened with uncertainty and imprecision in data. This makes a single index a rough approximation of the considered pollution situation.

Indeed, there is a potential for improvements in the research topics tackled in this paper which shapes further research directions. The possible improvements could be further divided in two rough partitions. The first, which is of fundamental kind, is about rethinking the air pollution index concept (for example, making it contextually dependent, or making it multidimensional). The second one is about improvement of the model proposed in this paper: use of new parameters (like those in Table IV), training data balancing, learning shapes of membership functions from historical data, and alike. Improvements should specifically address creation of precision metrics in linear fuzzy space, enabling estimations of sensitivity of interval partitions selection in time series, aggregation models, and fuzzy sets parameters. Recent theory development (see [27], [28]) gives a method for identification of the optimal solution for convex and non-convex optimization in fuzzy approach that could help to do this.

The two partitions intersect at utilization of Artificial Intelligence (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] World health statistics 2016: monitoring health for the SDGs, sustainable development goals. 1.Health Status Indicators. 2.Global Health. 3.Health Priorities. 4.Mortality. 5.Universal Coverage. 6.Life Expectancy. 7.Statistics. I.World Health, E-ISBN 978 92 4 069569 6, https://www.who.int/docs/default-source/gho-documents/world-health-statistic-reports/world-heatlth-statistics-2016.pdf [Accessed: May 15th 2022]

[2] Air quality in Europe – 2018 report, ISSN 1977-8449, https://www.eea.europa.eu/publications/air-quality-in-europe-2018/at_download/file [Accessed May 15th, 2022]

[3] Đ. Obradović, "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.

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

[5] Đ. Obradović, Z. Konjović, E. Pap, and I. Rudas. "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, 2013, pp 137-153.

[6] Đ. Obradović, Z. Konjović, E. Pap, and N. Ralević "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

[7] Đ. Obradović, Z. Konjović, E. Pap, and I. J. Rudas “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

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

[9] X. Tan et al., "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.

[10] E. P. Klement, R. Mesiar, and E. Pap, “Triangular norms,” Kluwer, 2000.

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

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

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

[14] N. A. AL-Dhurafi, N. Masseran, Z. H. Zamzuri, and M. A. Mohd Safari “Modeling the Air Pollution Index based on its structure and descriptive status,” Air Qual Atmos Health 11, pp 171–179, 2018. https://doi.org/10.1007/s11869-017-0528-21

[15] I. Atacak, N. Arici, and D. Guner, “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.

[16] A. Fino , F. Vichi, Ch. Leonardi, and K. Mukhopadhyay, "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].

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

[18] M. H. Sowlat, H. Gharibi, M. Yunesian, M. T. Mahmoudi, and S. Lotfi “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.

[19] J.W. Koo, S.W. Wong, G. Selvachandran, H. V. Long, and L. H. Son, “Prediction of Air Pollution Index in Kuala Lumpur using fuzzy time series and statistical models,” Air Qual Atmos Health 13, pp 77–88, 2020. https://doi.org/10.1007/s11869-019-00772-y

[20] Y. Alyousifi, M. Othman, R. Sokkalingam, I. Faye, and P. C. L. Silva, “Predicting Daily Air Pollution Index Based on Fuzzy Time Series Markov Chain Model,” Symmetry 2020, vol. 12, no. 293, 2020; doi:10.3390/sym12020293

[21] W. W. S. Wei, “Multivariate Time Series Analysis and Applications”, John Wiley & Sons Ltd., Book Series: Wiley Series in Probability and Statistics, 2019, doi:10.1002/9781119502951

[22] Q. Song and B.S. Chissom, “Fuzzy time series and its model”, Fuzzy Sets and Systems vol.3 no. 54, pp. 269–277, March 1993, doi:10.1016/0165-0114(93)90372-O

[23] S. Panigrahi and H.S. Behera, "Fuzzy Time Series Forecasting: A Survey, " in H. Behera, J. Nayak, B. Naik, and D. Pelusi 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

[24] A. Sahin, M. F. Dodurka, T. Kumbasar, E. Yesil, and S. Siradag, "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.

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

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

[27] S. Das and D. Chakraborty, An Introduction to Analytical Fuzzy Plane Geometry, STUDFUZZ, vol. 381, Springer, 2019.

[28] S. Das and D. Chakraborty, Conceptualizing fuzzy line as a collection of fuzzy points, Inf. Sci. 598, pp. 86-100, 2022.