Data Imputation in Related Time Series Using Fuzzy Set-Based Techniques

Adam Kiersztyn
Department of Computer Science
Lublin University of Technology
Lublin, Poland
adam.kiersztyn.pl@gmail.com

Paweł Karczmarek
Department of Computer Science
Lublin University of Technology
Lublin, Poland
pawel.karczmarek@gmail.com

Rafał Łopucki
Centre for Interdisciplinary Research
The John Paul II Catholic University of Lublin
Lublin, Poland
lopucki@kul.pl

Witold Pedrycz
Department of Electrical & Computer Engineering
University of Alberta
Edmonton, Canada
Department of Electrical and Computer Engineering
King Abdulaziz University
Jeddah, Saudi Arabia
Systems Research Institute
Polish Academy of Sciences
Warsaw, Poland
wpedrycz@ualberta.ca

Ebru Al
Ekol Lojistik Inc.
Ekol Caddesi No:2, TR-34935
Sultanbeyli-Istanbul, Turkey
ebru.al@ekol.com

Ignacy Kitowski
The State School of Higher Education in Chełm
Chełm, Poland
ignacyk@autograf.pl

Adam Zbyryt
The Polish Society for Bird Protection (PTOP)
Białystok, Poland
adam.zbyryt@wp.pl

***Abstract*—One of the main challenges faced by people who use data from empirical research in their work is missing data. In many scientific disciplines and industries there are references to time series. The suitability of several methods to imputation of the missing data in the study of mutual links between the analysed time series have been presented and tested in this work. In this paper, known methods of supplementing data in time series were enriched by the use of fuzzy sets and their processing was tested on unique data from experimental research and a transport company database. Fuzzy linguistic descriptors-based methods of missing data imputation in databases containing time series are discussed. The proposed method has a high efficiency, which have been proven in a series of experiments with both artificial and real datasets. The proposed methodologies have been tested on theoretical example and empirical data sets from various fields: (1) ecological data on changes in bird arrival dates in the context of climate change and (2) data describing the transport of containers between ports on the Mediterranean. Moreover, an important novelty of this work is, in particular, an application of fuzzy techniques to the correction of the datasets containing bird migration descriptions.**

***Index Terms*—missing data, data imputation, linguistic descriptors, fuzzy sets, transport data, ecological data**

Funded by the National Science Centre, Poland under CHIST-ERA programme (Grant no. 2018/28/Z/ST6/00563)

I. Introduction

Time series analysis has a number of applications and is used both in scientific research and in everyday life by numerous institutions and enterprises [1]–[5]. The fields that are often based on the analysis of time series are broadly understood biological sciences as well as sciences dealing with transport logistics. In biological sciences, for example, time series are used to study the consequences of climate change and great scientific effort is being made to predict how wildlife will respond to climate change and what management or conservation measures should be applied [6]–[10]. In transport logistics, on the other hand, the goal is to achieve the highest transport efficiency [11], [12]. In both areas (biology and transport logistics) changes or trends of selected parameters over time are most often compared, and for such comparisons, contemporary and historical data sets of the best quality and reliability are necessary. Unfortunately, the quality of historical data is often insufficient, and one of the main problems are data gaps (data are rarely collected systematically for many years) [13], [14].

In this paper, our objective is to propose and test the suitability of several fuzzy methods to fill in the missing data

978-1-7281-6932-3/20/$31.00 ©2020 IEEE

in the study of mutual links between the analysed time series. Our methodology has been tested on theoretical synthetic data and experimental data coming from various fields: (1) ecological data on changes in bird arrival dates in the context of climate change and (2) data describing the transport of containers between ports on the Mediterranean. The novelty of the work is the application of methods based on fuzzy sets to filling gaps in the data coming from field ecological studies and transportation.

It is worth mentioning that the topic of an application of fuzzy methods to the missing (incomplete) data has already been raised in the literature. However, the application areas considered in this paper have not been widely discussed so far. From a wide plethora of publications the following notions attract attention: Fuzzy K-Means [15] or Fuzzy C-Means [16]–[18], neuro-fuzzy structures [19], rough methods [20], or Granular Computing [21], [22].

In the proposed solution to make up for missing data in time series, tests are performed on full data, from which values are randomly removed, and then such artificially created gaps are filled. Missing values are generated in three ways:

(a) deleting a random value within each series;
(b) deleting 3 random values within each series;
(c) removing one time, common to all ranks;
(d) removing at most 3 common moments of time shared by all series.

For such artificially introduced data gaps, 5 methods of filling gaps were proposed in two variants: in the crisp version and in the fuzzy version. The latter is based on an introduction of the linguistic descriptors based on fuzzy numbers constructed on the basis of the appropriate crisp counterparts. This kind of approach brings high efficiency proven in a series of experiments.

The structure of the present study is as follows. In Section II and Section III discussed is the problem of missing data filling using crisp values and fuzzy linguistic descriptors, respectively. In Section IV the results of numerical experiments are presented while the last section covers summary and future work directions.

II. Crisp filling in the missing data

This paragraph discusses a crisp method of data imputation. Its fuzzy counterpart is proposed in the next section as its efficient enhancement. Consider a set of K time series describing similar phenomena (having the same set of values) in the same time window with the length N. Let $X_i(n), 1 \leq i \leq K, 1 \leq n \leq N$, denote the value of the i-th time series at the moment in time n. Without losing the generality of considerations, it can be assumed that all time series have values in a set of real numbers. One of the most commonly used techniques to make up for missing values in time series (but not only in time series) is the use of the average for the examined series in the analysed time window (where only the moments of time, for which the value of the series is determined, are used to determine the average), i.e.,

$$\bar{X}_i = \frac{\sum_{n=1}^{N} X_i(n)}{\sum_{n=1}^{N} \chi\left(X_i(n) \neq null\right)} \tag{1}$$

where *null* means the lack of a given item in the time series (database).

An alternative way to make up for missing values is to use full information about the behaviour of all the series being analysed in subsequent moments. It can be assumed that if the time series in question describe similar phenomena, they behave similarly at given moments of time in the sense that either they all increase or decrease. With this assumption, one can use the average increases for individual moments of time or the median increases in subsequent moments of time. At the same time, if knowledge of the existence of specific groups in the analysed subsets is available, it is possible to apply increments of the above measures for each of the subgroups independently.

More precisely, within each moment of time, the averages are calculated, i.e.,

$$\bar{X}(n) = \frac{\sum_{i=1}^{K} X_i(n)}{\sum_{i=1}^{K} \chi\left(X_i(n) \neq null\right)}. \tag{2}$$

Then the average increases are determined, namely

$$\Delta\bar{X}(n) = \bar{X}(n) - \bar{X}(n-1) \text{ for } n = 2, 3, \ldots, N \tag{3}$$

for the next moments of time. The final missing value in the time series is calculated according to the formula

$$X_i(n) = X_i(n-1)\left(1 + \Delta\bar{X}(n)\right). \tag{4}$$

In the above considerations it is possible to replace the average with the median, as well as use the value of statistics determined within individual of the highlighted groups. In the experimental section, it was limited to presenting only 4 methods, which were selected on the basis of previous research.

III. Fuzzy filling in the missing data

An alternative approach to the use of classic statistical measures to fill data gaps is to use fuzzy sets. This non-classical approach has a very logical justification, because in many cases there can be no absolute certainty of the values recorded.

In many cases, for example, data from field observations, the exact value of the analysed features or phenomenon cannot be determined, but only its approximation. Then, the use of fuzzy numbers is not only fully justified but even advisable.

For fuzzy methods, fuzzy numbers are used instead of the recorded time series values. More specifically, it is permissible to blur a symmetrical discrete value two values less and two values more (see Fig. 4).

In order to increase the transparency of considerations and simplify the model, linguistic descriptors based on the results of statistical analysis of available data are introduced.

Belonging to individual descriptors depends, of course, on the fuzzy form of the analysed value. It is possible to use different numbers of descriptors; however, the use of 5 or 7 seems to be the best compromise between the accuracy and complexity of the calculations (see the experimental section). In the case of 5 descriptors, their ranges and symbols representing them are presented in Table I. In order to facilitate calculations, each descriptor has been assigned a numerical value representing it, the values given in Table I are only examples of symbols describing individual descriptors.

TABLE I
DESCRIPTORS AND THEIR RANGE

Descriptor	Range	Symbol
Significantly ahead of time	$(-\infty; \bar{X} - 1.5S]$	-2
Slightly ahead of time	$(\bar{X} - 1.5S; \bar{X} - 0.5S]$	-1
On time	$[\bar{X} - 0.5S; \bar{X} + 0.5S]$	0
Slightly over time	$(\bar{X} + 0.5S; \bar{X} + 1.5S]$	1
Significantly over time	$(\bar{X} + 1.5S; \infty)$	2

For such symbolic representations of linguistic descriptors, similar considerations can be made as in the case of crisp ones, with blurring of the recorded value of the empirical series allowed. In addition, each value achieved by the time series is blurred according to the characteristics of the data being analysed. Triangular membership functions are most commonly used because of their intuitiveness, but there is nothing to prevent other forms of fuzzy membership function being used.

Within each of the proposed methods of completing data, a full analysis is carried out on linguistic descriptors and the artificially entered missing value is supplemented, which is compared with the fuzzy input value.

IV. Numerical experiments

A. Theoretical example

In order to accurately present the discussed methods of solving the problem of filling data gaps in time series, let us consider the following example. Let $N = 6$ and time series $X_1, X_2, \ldots, X_6$ be given, of which the first 3 belong to one subgroup A, and the last 3 belong to the second subgroup B. Let us assume that for each of these series there are values at moments of time $n = 1, 2, \ldots, 10$. As part of the test, each time series values are random. To increase the transparency of the values of individual time series, it is convenient to present them in the form of a Table II. These values have been drawn and serve only as a reference.

The last row of the table contains averages for each of the time series considered. The average and median increments for all series and within each group are presented in the Table III.

For this kind of sample values, various types of efficiency tests of the proposed methods can be performed. In the case of removing a one random value in each series, the percentage of correctly completed data gaps by applying the average is presented in Fig. 1.

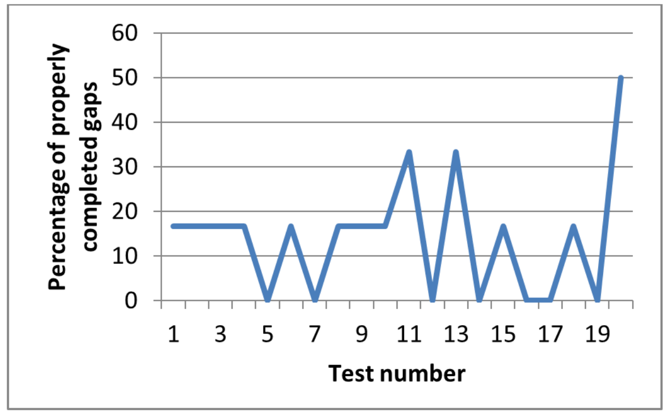

Fig. 1. Percentage of correctly completed deficiencies using a simple average for one value randomly removed in each time series

It can be observed that the use of a simple average calculated for a given time series allows to make up for missing data. However, the results obtained are not very satisfying.

Comparisons of the effectiveness of the introduced methods (expressed as a percentage of correctly reproduced deleted values) are presented in Fig. 2 and Fig. 3. It can be observed that most methods are more effective for removing the same moments of time within all time series considered. In addition, it should be noted that the correctness of completing the data depends on which values have been drawn for deletion. This is evidenced by the large difference between the results for individual methods in individual tests.

It also needs to be pointed out that for synthetic data it is difficult to clearly indicate the most effective method. The value of properly filling out the missing data depends on

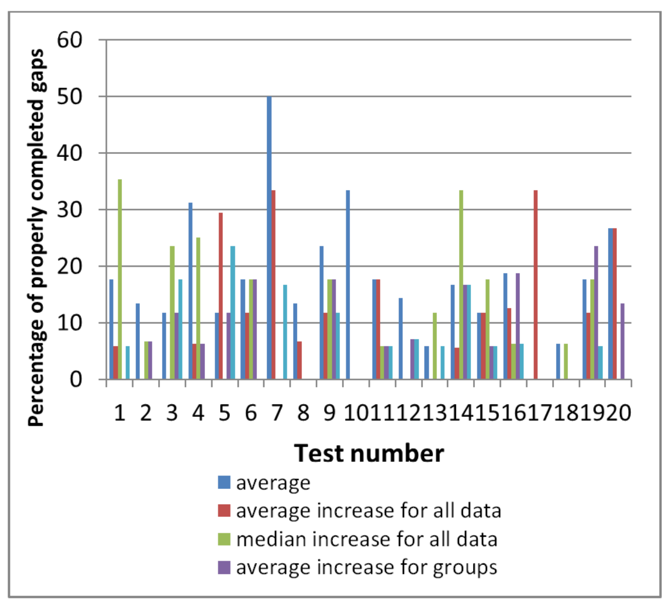

Fig. 2. Comparison of the percentage of correctly completed data gaps in the case of random removal of at most three (random value) elements for each time series

TABLE II
SAMPLE (RANDOM) VALUES OF TIME SERIES WITH DETERMINED VALUES ALLOWING TO REMOVE MISSING DATA

$\mathbf{X_1}$	$\mathbf{X_2}$	$\mathbf{X_3}$	$\mathbf{X_4}$	$\mathbf{X_5}$	$\mathbf{X_6}$	$\bar{\mathbf{X}}$	$\mathbf{Me}$	$\bar{\mathbf{X}}_\mathbf{A}$	$\mathbf{Me_A}$	$\bar{\mathbf{X}}_\mathbf{B}$	$\mathbf{Me_B}$
9	19	32	19	10	21	18.5	19.0	20.0	19.0	16.7	19.0
11	21	39	22	5	10	18.3	16.0	23.7	21.0	12.3	10.0
7	15	27	15	16	30	18.8	15.5	16.3	15.0	20.3	16.0
7	14	26	16	5	11	13.8	12.5	15.7	14.0	10.7	11.0
6	13	24	15	10	20	15.5	14.0	14.3	13.0	15.0	15.0
6	13	23	12	15	30	17.5	14.0	14.0	13.0	19.0	15.0
5	11	20	13	11	22	14.8	12.0	12.0	11.0	15.3	13.0
12	22	41	25	9	18	22.5	20.0	25.0	22.0	17.3	18.0
8	16	29	16	12	23	18.8	16.0	17.7	16.0	17.0	16.0
9	16	29	17	7	15	17.2	15.5	18.0	16.0	13.0	15.0
9	19	32	19	10	21	18.5	19.0	20.0	19.0	16.7	19.0
8	16	29	17	10	20	$\bar{\mathbf{X}}$					

TABLE III
AVERAGE AND MEDIAN

$\nabla\bar{\mathbf{X}}$	$\nabla\mathbf{Me}$	$\nabla\bar{\mathbf{X}}_\mathbf{A}$	$\nabla\mathbf{Me_A}$	$\nabla\bar{\mathbf{X}}_\mathbf{B}$	$\nabla\mathbf{Me_B}$
0,99	0,84	1,18	1,11	0,74	0,53
1,03	0,97	0,69	0,71	1,65	1,60
0,73	0,81	0,96	0,93	0,52	0,69
1,12	1,12	0,91	0,93	1,41	1,36
1,13	1,00	0,98	1,00	1,27	1,00
0,85	0,86	0,86	0,85	0,81	0,87
1,52	1,67	2,08	2,00	1,13	1,38
0,84	0,80	0,71	0,73	0,98	0,89
0,91	0,97	1,02	1,00	0,76	0,94

the nature of the time series considered and their internal consistency. In the case of synthetic data, random values random values have been used. Therefore it is not possible to clearly identify trends in time series.

A summary of the best results for 20 random data sets is presented in Table IV.

TABLE IV
A SUMMARY OF THE BEST RESULTS FOR ARTIFICIAL DATASET

	One gap	At most 3 gaps	One gap year	At most 3 gaps years
Average	50	31.3	50	44.4
Average increase for all data	33.3	29.4	33.3	50
Median increase for all data	16.7	35.3	33.3	50
Average increase for groups	33.3	23.5	100	50
Median increase for groups	16.7	23.5	50	50

Comparing the data contained in Table IV, it should be noted that the methods based on the average and median increments for individual subgroups have the highest efficiency in the case of missing data for whole years.

In the case of using fuzzy equivalents of the discussed methods and for the proposed membership function, the apropriate results are presented in Fig. 4.

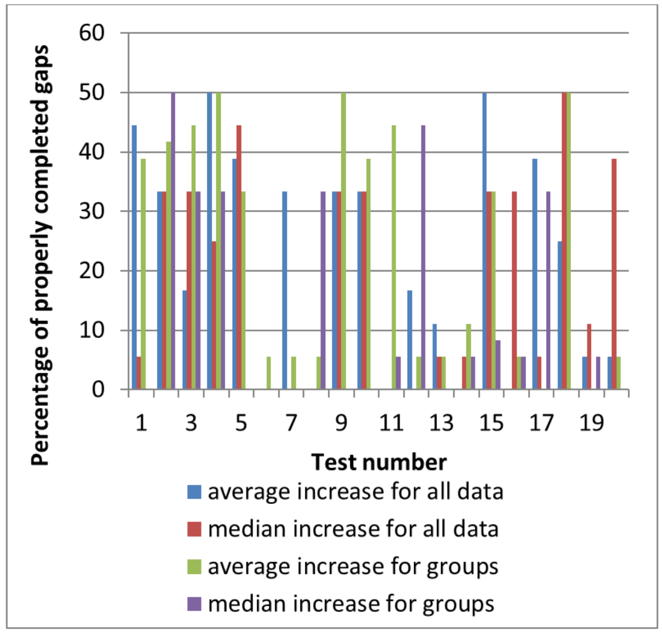

Fig. 3. Comparison of the percentage of correctly completed data gaps in the case of random removal of at most three (random value) moments of time – common for all time series

The introduction of a fuzzy set-based approach allows to significantly improve the results of correcting the lack of data. Fig. 5 shows the percentage of correction of data deficiencies by means of a simple average using the fuzzy membership function presented in Fig. 4. As can be seen, the introduction of fuzzy numbers has increased the degree of proper restoration of missing data, up to 100%.

As one can see, the introduction of fuzzy set-based methods significantly improves the effectiveness of the proposed methods as evidenced by the Table V.

Comparing the results contained in Tables IV and V, it should be stressed that the use of fuzzy sets significantly increases the efficiency of correcting data gaps. For each of the analysed methods, the version using fuzzy numbers brings significantly better results. The question remains whether the use of more complex forms of membership functions will achieve even better results.

TABLE V
A SUMMARY OF THE BEST RESULTS FOR ARTIFICIAL DATASET

	One gap	At most 3 gaps	One gap year	At most 3 gaps years
Average	100	86.7	100	100
Average increase for all data	83.3	50	66.7	77.8
Median increase for all data	66.7	50	83.7	66.7
Average increase for groups	66.7	52.9	83.3	55.6
Median increase for groups	50	43.8	50	50

B. Ecological data

The processing of the proposed solutions was tested on two completely different empirical data sets. One is the arrival date of migrating birds in Podlasie, a region in north-eastern Poland. Seasonal bird migrations are strongly associated with climate change and can be used to analyse the ecological effects of climate warming [7], [13], [14], [23], [24]. The spring arrival dates of 79 species of birds (so-called first arrival dates or FADs) collected in northern Poland (the North Podlasie Lowland: $21°51' - 23°57'$ E; $52°17' - 53°54'$ N) in the years 1996-2016 [25] have been used. For the needs of analysis, a data matrix consisting of 79 columns (species) and 21 rows (years) was built. For each species, data gaps have been artificially introduced and various methods for completing them have been tested.

For ecological data, similar tests were carried out as in the case of synthetic data, with the difference that a maximum of 5 deficiencies were allowed for each species or 5 years without data. The best results were obtained when the original data gaps were supplemented by means of the median increase in

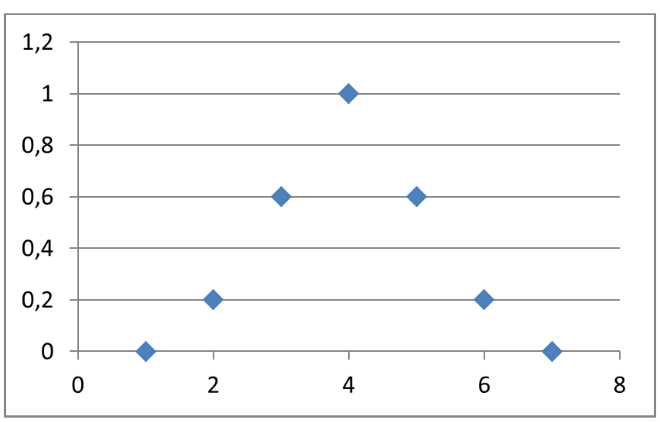

Fig. 4. Sample fuzzy function for discrete data, value 4 is the observed value

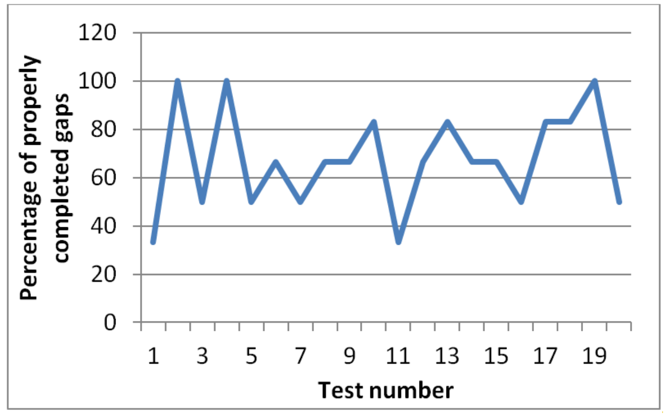

Fig. 5. Percentage of correctly completed deficiencies using a fuzzy average for one value randomly removed in each time series

bird groups. The average results of correct additions (expressed as a percentage) are presented in Table VI.

As one can see, the results obtained are far from satisfactory. However, they indicate that the crisp approach has the potential to be used for ecological data. It is advisable to refine these methods or use techniques using fuzzy sets. The maximum values achieved in the tests for correcting missing values are presented in Fig. 6.

It can be observed that filling out the missing data with a simple mean calculated for each species separately gives the worst results. After using seven fuzzy linguistic descriptors and conducting 20 random tests of introducing data gaps, the maximum percentage of correctly completed data gaps was presented in Fig. 7. As expected, the use of fuzzy sets has significantly improved the results obtained. Comparing the results presented in Fig. 6 and Fig. 7, it should be noted that the use of fuzzy sets has allowed to increase efficiency by almost four times.

TABLE VI
AVERAGE PERCENTAGES OF CORRECTING DATA GAPS BASED ON 20 RANDOM TESTS

	One gap	At most 5 gaps	One gap year	At most 5 gaps years
Average	4.2	4.5	4.3	4.7
Average increase for all data	4,.3	2.6	2.3	2.3
Median increase for all data	3.6	3,.1	3.3	3.7
Average increase for groups	4.9	3.6	3.7	3.7
Median increase for groups	5.0	3.5	3.6	6.0

C. Roro shipment data

The last test set, which checks the effectiveness of the proposed solutions are data describing the transport of containers between Mediterranean Sea ports. Anonymous data from a transport company were used for the analysis. From the

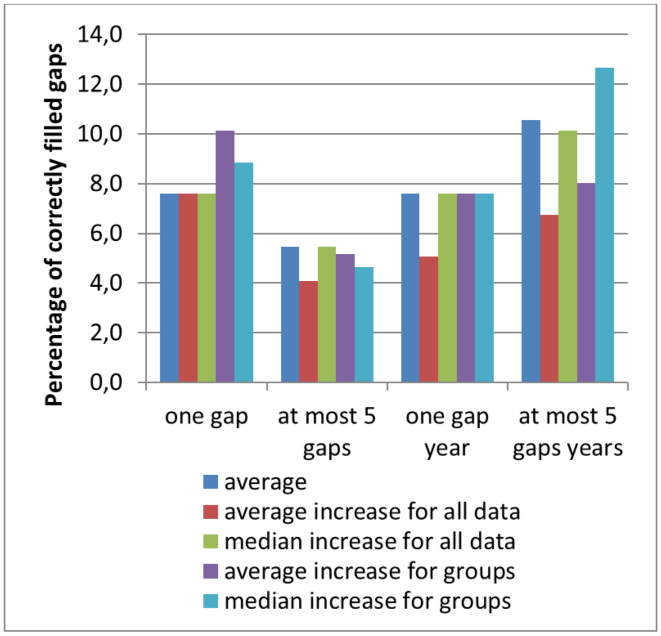

Fig. 6. Averaged results based on 20 random samples

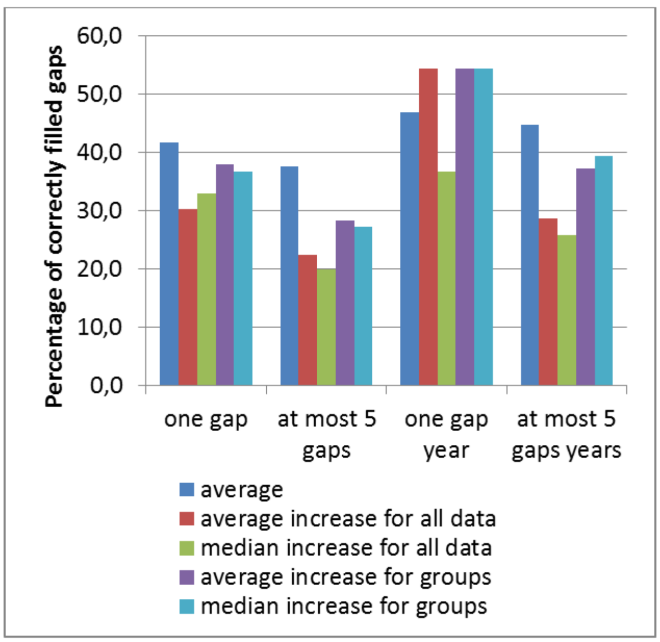

Fig. 7. Maximum values of the percentage of correctly filled gaps

TABLE VII

SELECTED TRANSPORT DATA SHOWING THE TIME OF CONTAINER SHIP VOYAGE BETWEEN PORTS IN THE FOLLOWING MONTHS (IN HOURS)

Month	B-C	C-B	C-D	D-C	D-F	F-D
1	52	54	46	53	50	44
2	53	55	47	61	50	45
3	55	55	49	62	50	45
4	56	56	55	63	51	45
5	59	57	60	63	51	46
6	59	57	61	63	51	46
7	59	58	62	63	51	46
8	60	58	65	68	52	46
9	60	58	66	68	52	46
10	60	58	66	69	53	46
11	60	58	66	75	53	46
12	60	58	66	82	53	47

available data the route between three port pairs was selected for analysis and the duration of voyages was determined. To illustrate the working of the proposed method, 3 pairs of ports were randomly selected and the cruise time in hours was determined on the basis of the time and the date of leaving the departure port and entering the destination port. The results for the following months are presented in Table VII.

Test results obtained with a crisp approach suggest that also for transport data this type of gap correction methods can be successfully used also for transport data (Fig. 8 and Fig. 9). It should be noted that, due to the specificity of the transport data (high repeatability of results), the proposed crisp version methods work correctly and achieve relatively satisfactory results. The best results are obtained by methods based on the use of increases in the value of series at subsequent moments of time. Depending on the data items drawn for deletion, in some cases, better methods are those using increments for all data work better, while in others increments within groups.

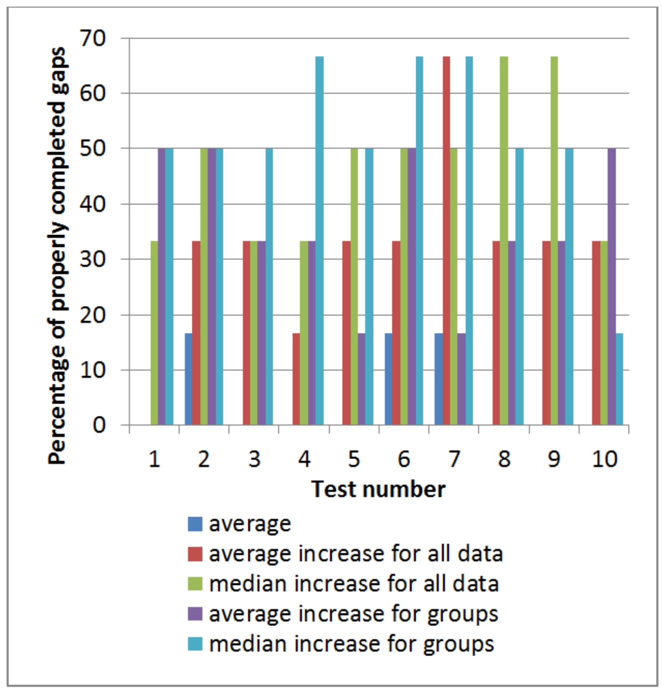

Fig. 8. Percentages of correctly completed data gaps, with one random gap within each route (first 10 tests).

The best results achieved by individual methods within 20 tests are presented in the Table VIII. Comparing the data contained in Table VIII, as in the case of synthetic data, it can be seen that the methods of growth based on average and

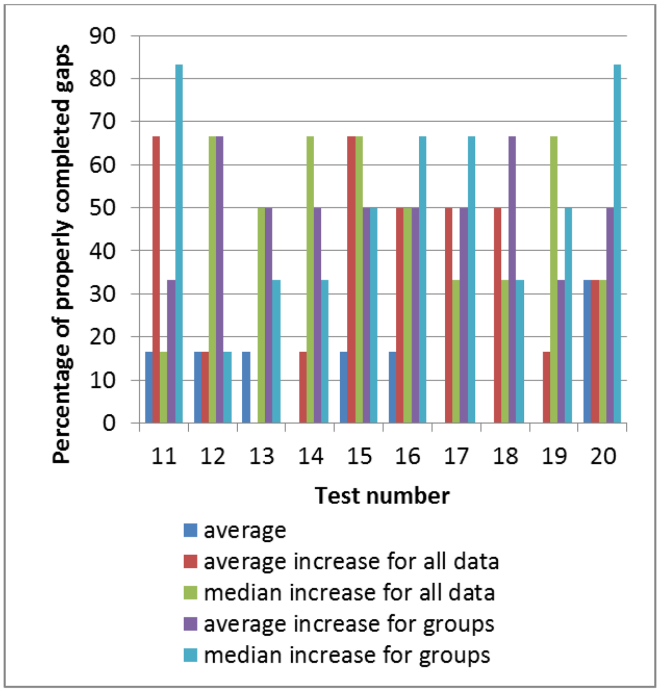

Fig. 9. Percentages of correctly completed data gaps, with one random gap within each route (last 10 tests).

TABLE VIII
SELECTED TRANSPORT DATA SHOWING THE TIME OF CONTAINER SHIP VOYAGE BETWEEN PORTS IN THE FOLLOWING MONTHS (IN HOURS)

	One gap	**At most 3 gaps**	**One gap year**	**At most 3 gaps years**
Average	33.3	84.6	50	91.7
Average increase for all data	66.7	57.1	83.3	100
Median increase for all data	66.7	62.5	83.3	72.2
Average increase for groups	66.7	82.3	83.3	83.3
Median increase for groups	83.3	64.7	50	88.9

median increments for individual subgroups have the highest efficiency.

Despite such good values of filling in the lack of data with the help of crisp methods, it turns out that the use of fuzzy numbers significantly improves the already good result. The comparison of crisp and fuzzy approaches for one random gap for each series is presented in the Fig. 10.

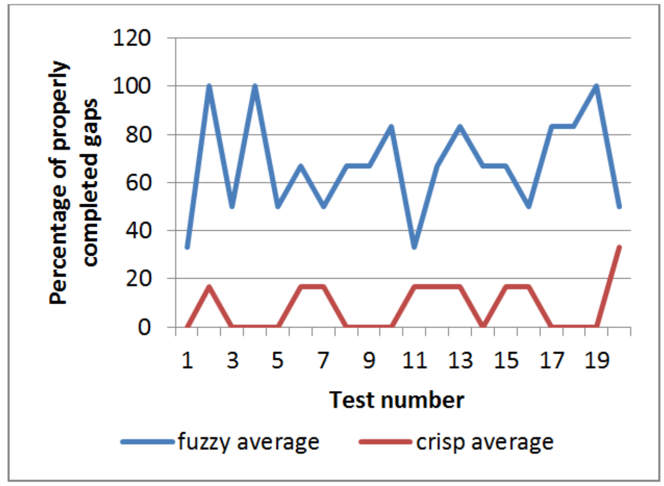

Fig. 10. Comparison of results obtained with the help of a simple average in crisp and fuzzy versions.

V. Conclusions and Future Work

The work presents an intuitive approach to filling up data gaps in a group of related time series. Our results can be widely used in biological sciences, because ecological data due to their non-deterministic nature have great potential to use fuzzy logic [26].

Further development of the proposed concept is planned, among others, through in-depth analysis of the possibilities of using other linguistic descriptors and more sophisticated techniques of Computational Intelligence and Granular Computing. Work is currently underway on the application of other forms of membership functions and the initial results achieved are optimistic. In addition, it is planned to compare the proposed solutions based on fuzzy sets with other competitive methods. Very thorough statistical examination is planned for the analyses performed, allowing for a detailed comparison of alternative methods.

Acknowledgment

Funded by the National Science Centre, Poland under CHIST-ERA programme (Grant no. 2018/28/Z/ST6/00563).

References

[1] A. M. A. Athick, K. Shankar, and H. R. Naqvi, "Data on time series analysis of land surface temperature variation in response to vegetation indices in twelve Wereda of Ethiopia using mono window, split window algorithm and spectral radiance model," *Data Brief*, vol. 27, p. 104773, 2019.

[2] S. Chen and P. Shang, "Financial time series analysis using the relation between MPE and MWPE," *Physica A*, vol. 537, p. 122716, 2020, available online.

[3] R. Mead, R. N. Curnow, and A. M. Hasted, *Statistical methods in agriculture and experimental biology*. Chapman & Hall/CRC, 2017.

[4] K. Mise *et al.*, "Time-series analysis of phosphorus-depleted microbial communities in carbon/nitrogen-amended soils," *Appl. Soil Ecol.*, vol. 145, 2020, available online.

[5] Y. Ophir and K. H. Jamieson, "The effects of Zika virus risk coverage on familiarity, knowledge and behavior in the US – A time series analysis combining content analysis and a nationally representative survey," *Health Commun*, vol. 35, no. 1, pp. 35–45, 2020.

[6] M. Ahola *et al.*, "Variation in climate warming along the migration route uncouples arrival and breeding dates," *Glob Chang Biol*, vol. 10, no. 9, pp. 1610–1617, 2004.

[7] T. Usui, S. H. Butchart, and A. B. Phillimore, "Temporal shifts and temperature sensitivity of avian spring migratory phenology: A phylogenetic meta-analysis," *J Anim Ecol*, vol. 86, no. 2, pp. 250–261, 2017.

[8] L. Sokolov, "Effect of global warming on the timing of migration and breeding of passerine birds in the 20th century," *Entomol Rev*, vol. 86, no. 1, pp. 59–81, 2006.

[9] O. V. Askeyev, T. H. Sparks, I. V. Askeyev, and P. Tryjanowski, "Is earlier spring migration of Tatarstan warblers expected under climate warming?" *Int J Biometeorol*, vol. 51, no. 6, pp. 459–463, 2007.

[10] N. Saino *et al.*, "Climate change effects on migration phenology may mismatch brood parasitic cuckoos and their hosts," *Biol. Lett.*, vol. 5, no. 4, pp. 539–541, 2009.

[11] B. L. Smith, W. T. Scherer, and J. H. Conklin, "Exploring imputation techniques for missing data in transportation management systems," *Transp Res Rec*, vol. 1836, no. 1, pp. 132–142, January 2003.

[12] G. Shen and S. G. Aydin, "Origin – destination missing data estimation for freight transportation planning: a gravity model-based regression approach," *Transp. Plan. Technol.*, vol. 37, no. 6, pp. 505–524, August 2014.

[13] W. Biaduń, I. Kitowski, and E. Filipiuk, "Trends in the arrival dates of spring migrants in Lublin (E Poland)," *Acta Ornithol*, vol. 44, no. 1, pp. 89–94, 2009.

[14] C. Kullberg *et al.*, "Change in spring arrival of migratory birds under an era of climate change, Swedish data from the last 140 years," *Ambio*, vol. 44, no. 1, pp. 69–77, 2015.

[15] D. Li, J. Deogun, W. Spaulding, and B. Shuart, "Towards missing data imputation: a study of fuzzy K-Means clustering method," in *International Conference on Rough Sets and Current Trends in Computing RSCTC*. Springer, 2004, pp. 573–579.

[16] L. Zhang, W. Lu, X. Liu, W. Pedrycz, and C. Zhong, "Fuzzy C-Means clustering of incomplete data based on probabilistic information granules of missing values," *Knowl Based Syst*, vol. 99, pp. 51–70, May 2016.

[17] J. Tang, G. Zhang, Y. Wang, H. Wang, and F. Liu, "A hybrid approach to integrate fuzzy C-Means based imputation method with genetic algorithm for missing traffic volume data estimation," *Transp Res Part C Emerg Technol*, vol. 51, pp. 29–40, 2015.

[18] A. G. Di Nuovo, "Missing data analysis with fuzzy C-Means: A study of its application in a psychological scenario," *Expert Syst. Appl.*, vol. 38, no. 6, pp. 6793–6797, 2011.

[19] R. Nowicki, "Rough neuro-fuzzy structures for classification with missing data," *IEEE Trans. Syst., Man, Cybern. B*, vol. 39, no. 6, pp. 1334–1347, 2009.

[20] M. Amiri and R. Jensen, "Missing data imputation using fuzzy-rough methods," *Neurocomputing*, vol. 205, pp. 152–164, 2016.

[21] C. Garcia, D. Leite, and I. Skrjanc, "Incremental missing-data imputation for evolving fuzzy granular prediction," *IEEE Trans. Fuzzy Syst.*, 2019, in press.

[22] C. Zhong, W. Pedrycz, D. Wang, L. Li, and Z. Li, "Granular data imputation: a framework of granular computing," *Appl. Soft Comput.*, vol. 46, pp. 307–316, September 2016.

[23] M. Zalakevicius, G. Bartkeviciene, L. Raudonikis, and J. Janulaitis, "Spring arrival response to climate change in birds: a case study from eastern Europe," *J. Ornithol.*, vol. 147, no. 2, pp. 326–343, 2006.

[24] S. E. Newson *et al.*, "Long-term changes in the migration phenology of UK breeding birds detected by large-scale citizen science recording schemes," *Ibis (Lond. 1859)*, vol. 158, no. 3, pp. 481–495, 2016.

[25] Database of the Polish Society for Birds Protection (PTOP), http://www.ptop.org.pl, 2016, online; accessed on 2019-10-01.

[26] R. Łopucki and A. Kiersztyn, "The city changes the daily activity of urban adapters: Camera-traps study of Apodemus agrarius behaviour and new approaches to data analysis," *Ecol. Indic.*, vol. 110, p. 105957, 2020, available online 2019-11-30.