






Artigo

MANNGA: A Robust Method for Gap Filling Meteorological Data

Thiago Meirelles Ventura¹ , Claudia Aparecida Martins¹ , Josiel Maimone de Figueiredo¹ ,
Allan Gonçalves de Oliveira¹ , Johnata Rodrigo Pinheiro Montanher¹ 

¹*Instituto de Computação, Universidade Federal de Mato Grosso, Cuiabá, MT, Brazil.*

Received: 9 December 2016 - Accepted: 9 September 2018

Abstract

This paper presents Mannga (Multiple variables with Artificial Neural Network and Genetic Algorithm), a method designed for gap filling meteorological data. The main approach is to estimate the missing data based on values of other meteorological variables measured at the same time in the same local, since the meteorological variables are strongly related. Experimental tests showed the performance of Mannga compared with other two methods typically used by researchers in this area. Good results were achieved, with high accuracy even for sequential failures, which is a big challenge for researchers. The core advantages of Mannga are the flexibility of handling different types of meteorological data, the ability of select the best variables to assist the gap filling and the capacity to deal with sequential failures. Moreover, the method is available to public use with the Java programming language.

Keywords: multivariate data, artificial neural network, genetic algorithm, open source software.

MANNGA: Um Método Robusto para Preenchimento de Falhas em Dados Meteorológicos

Resumo

Este trabalho apresenta o método Mannga (Multiple variables with Artificial Neural Network and Genetic Algorithm), desenvolvido para preencher falhas em dados meteorológicos. A ideia principal é preencher as falhas baseando-se nos valores de outras variáveis meteorológicas medidas no mesmo momento, uma vez que as variáveis meteorológicas possuem forte relação entre si. Testes foram executados para mostrar a performance do Mannga comparado com outros dois métodos comumente utilizados na área. Os resultados alcançados atingiram uma boa precisão, principalmente relacionado ao desafio de preencher valores em dados que ocorrem em sequência. As principais vantagens do Mannga são a sua flexibilidade em manipular diferentes tipos de dados meteorológicos, a habilidade de selecionar as melhores variáveis para auxiliar no preenchimento das falhas e a capacidade de lidar com falhas sequenciais. Além disso, o método está disponível publicamente na linguagem de programação Java.

Palavras-chave: dados multivariados, redes neurais artificiais, algoritmos genéticos, software livre.

1. Introduction

Meteorological data has an important position in scientific research. Based on meteorological data, explanations about climatic phenomena are made, allowing us to understand several characteristics of our planet. To aid the process of data acquisition, many types of equipment are installed in meteorological stations. Commonly, the equipment works 24 hours per day, for years. Therefore, a huge

quantity of data is generated. Unfortunately, not all data is integrally perfect, because failures appear in data series.

Missing or rejected data in these measurements is an ubiquitous problem due to equipment failures (system/sensor breakdown), maintenance and calibration, spikes in the raw data, and physical and biological constraints (e.g. storms, hurricanes, and non-optimal wind directions) (Hui *et al.*, 2004). In any case, the gap created in the data series will cause a bad interpretation in the data study. Thus, it is important to apply a gap filling method to fix the dataset.

One of the methods used for gap filling is Multiple Imputation (MI), used by [Sullivan *et al.* \(2015\)](#), a Monte Carlo technique in which the missing values are replaced by $m > 1$ simulated versions, where m is typically small, for example, between 3 and 10 ([Schafer, 1999](#)). [Horton and Ipsitz \(2001\)](#) comment on several systems to facilitate the use of the method, like Solas, Sas, S-Plus, Mice, and others. [Hui *et al.* \(2004\)](#) used the MI method for gap filling eddy covariance data, which collect data about the exchange of carbon dioxide, water vapor and heat from a vegetated surface and the atmosphere.

Other methods of gap filling are the Mean Diurnal Variation (MDV) and the Look-up Tables ([Falge *et al.*, 2001](#)). MDV replaces the gap using an average calculated from values of adjacent days ([Kato *et al.*, 2006](#)). This method was also used in [Hu *et al.* \(2009\)](#), [Alavi *et al.* \(2006\)](#) and [Mohan and Rao \(2016\)](#). The look-up table approach consists of creating a table with the flux values binned, based on the corresponding values of the external parameters. The determination of the relevant parameters and their critical values is a crucial step if this technique is to be successful ([Mishurov and Kiely, 2011](#)). This method was used in [Zhou *et al.* \(2015\)](#), [Rodrigues *et al.* \(2005\)](#), [Wilson and Baldocchi \(2001\)](#) and [Shao *et al.* \(2011\)](#).

Regression analysis is performed in order to determine the correlations between two or more variables having cause-effect relations, and to make predictions for the topic by using the relation ([Uyanik and Guler, 2013](#)). Multiple Linear Regression (MLR) can be used to simulate meteorological data, as shown in [Malik and Kumar \(2015\)](#).

Some variations of gap filling techniques were compared with the same dataset of net carbon fluxes in [Moffat *et al.* \(2007\)](#), like interpolation, probabilistic filling, look-up tables, non-linear regression, artificial neural networks, and process-based models in a data-assimilation mode. Besides, the performance of three methods for gap filling data of net ecosystem CO_2 exchange was evaluated in [Ooba *et al.* \(2006\)](#). It was concluded that a method using an Artificial Neural Network offers better performance for gap filling.

In all of them, methods for gap filling are limited to a specific climatic variable. In some cases, it is very complicated to apply the method, since you have to make different settings for each data type. These disadvantages are common in gap filling methods. Therefore, the purpose of this work is to show the development of Mannga (Multiple variables with Artificial Neural Network and Genetic Algorithm), which is an optimized method, combining two Artificial Intelligence techniques, Genetic Algorithm and Artificial Neural Network. Mannga method works with several climatic variables at the same time and avoid the user to execute a specific configuration for each variable. This method is called Mannga and it was implemented with the Java programming language.

2. Material and Methods

2.1. Proposed method

The proposed method, Mannga, takes advantage of two techniques to perform gap filling on meteorological data: Artificial Neural Network (ANN) and Genetic Algorithm (GA). Artificial Neural Network is a computational technique based on the concept of the human brain neurons. An ANN is a massively parallel distributed processor made up of simple processing units, which has a natural propensity for storing experiential knowledge and making it available for use ([Haykin, 1999](#)).

The structure of an ANN has several parameters and can be configured in many different ways. For each dataset there is a better configuration of the ANN to solve the problem. Finding the optimal structure of ANN consists of investigating an entire space of possible states. This task requires a great amount of processing, so it is necessary to use a search algorithm to find a satisfactory solution.

GA is a computational analogy of adaptive systems that is used to generate useful solutions to optimization and search problems. In this context, a Genetic Algorithm was used to assist the structure definition of the ANN, as a search method that finding optimal or good solutions by examining only a small fraction of the possible candidates ([Mitchell, 1998](#)).

The main idea of the proposed method considers that climatic variables are related to each other. Thus Mannga estimates the missing data based on the values of other available climatic variables. For example, if at 10:30 AM the value of temperature data is missing, the method calculates the temperature at this moment considering the values measured at 10:30 AM of incoming shortwave radiation, wind speed and relative humidity data. Even if there are several sequential gaps, it is possible that this method is able to fill them.

Thus, the ANN will be responsible for calculating the missing data. However, as mentioned, there are countless configurations of an ANN, each one worse or better depending on the data series. In this case, the GA was utilized to determine the best ANN for the current data series. In this approach, we have more probability to work with different types of meteorological data, because the ANN will be optimized in each test.

Based on [Ventura *et al.* \(2015\)](#), the ANN parameters determined by the GA were: training algorithm, activation functions, learning rate, momentum rate and number of neurons. Sometimes there are many climatic variables in the data series. Thus, in addition to the parameters of ANN, the GA determines which variable should or should not be used in the estimation. In this case, only the more correlated variables are used, improving the performance of the method and decreasing the error in the final estimate.

The method is shown in Figure 1. Initially, one dataset (without failures) is given to the GA. The GA will use these data to learn the patterns of the climatic variables and search for the best settings for the ANN for that specific data. This is achieved creating several neural networks with different parameters. The networks created are evaluated and those with greater precision have more chances of being selected. After several iterations, the

chosen ANN is used to gap filling on other datasets that have failures. Finally, the dataset with failures is fixed.

2.2. Experimental setup

Simulations were performed in order to evaluate the Mannga performance. The dataset used were obtained in AmeriFlux¹, which provides continuous measurements from forests, grasslands, wetlands, and croplands in North,

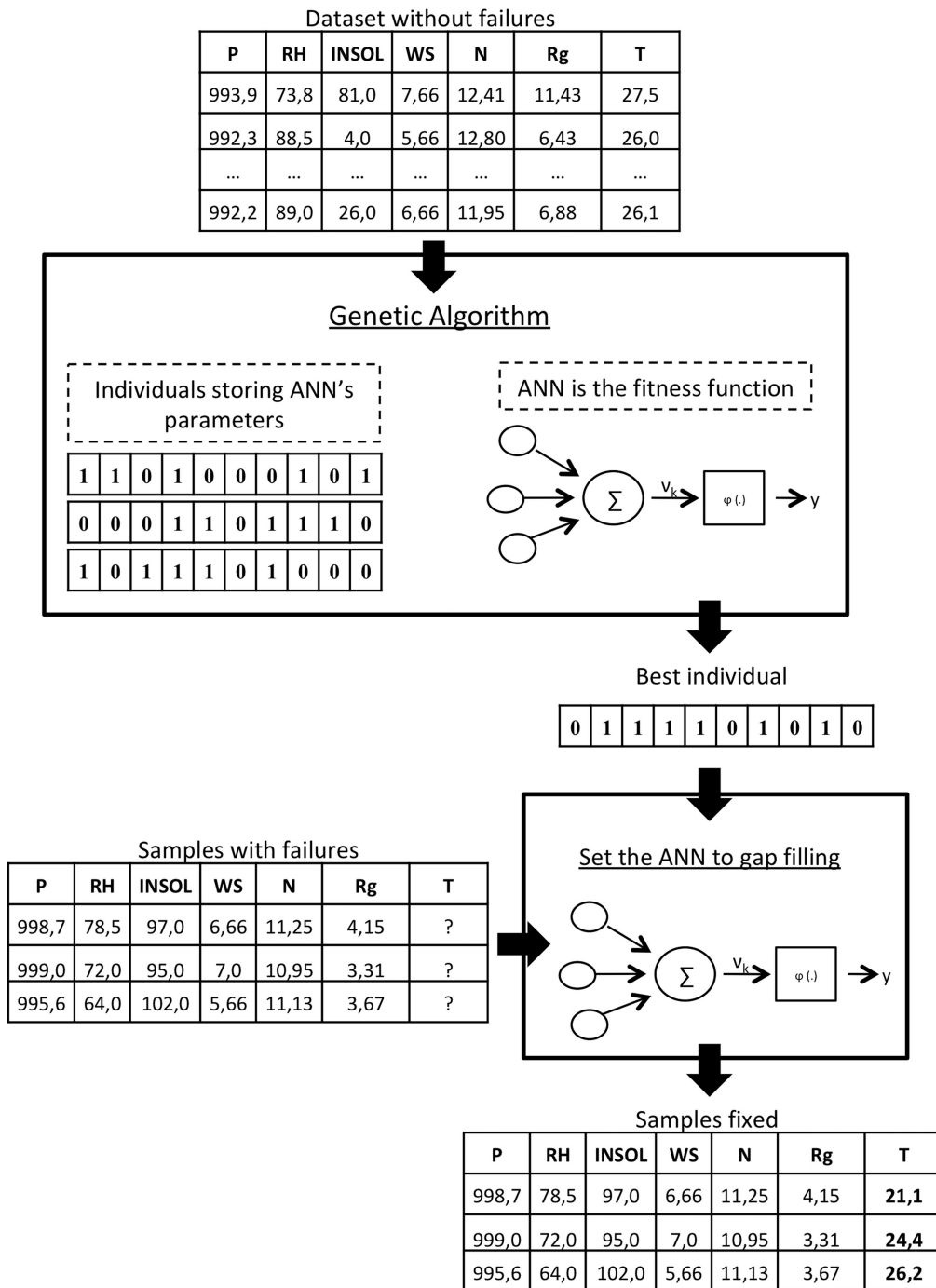


Figure 1 - Integration of GA and ANN to enable the operation of Mannga.

Central and South America (Boden *et al.*, 2013). We also evaluate a dataset from INMET². Three sites were chosen from AmeriFlux and one from INMET. The quality of several variables was not good, containing invalid and missing data. Therefore it was selected the variables and months with a minimum of quality to test Mannga performance. More information about the dataset is shown in Table 1.

Often meteorological data has a high variation during the annual cycle. Therefore to estimate values of this type of data it is necessary specific period of data to create good models. In Leauthaud *et al.* (2017) was considered only 30 days close to the gap to perform gap filling. In this case we have similar data to process, increasing the probability of a good estimation. Staub *et al.* (2017) present other advantage using only a specific amount of data, which is a decrease in computations effort to build models. For this reasons it is a good approach to select only a small sample of meteorological measurements (1 to 3 months) to perform gap filling. We do the same in this work for each dataset.

The processing time varies depending on the amount of data and computer used. In these tests, a dual-core computer with only 1GB of RAM was used, taking approximately 19 minutes for processing each month of data with Mannga gap filling method.

For each site, several variables were selected to perform gap filling. Mannga accuracy was checked by simulating gaps in data series. Three simulations were tested for each dataset:

- 5% of failures randomly inserted, to test regular scenarios on the dataset.
- 10% of failures inserted on sequence, to test the method accuracy when several gaps occur for a long period of time.
- 30% of failures randomly inserted, to test the method behavior when a lot of gaps are presented on the dataset.

To compare Mannga accuracy the same tests were performed with another two others methods: Average

(commonly used due to his facility) and Multiple Linear Regression.

2.3. Mannga implementation

To facilitate the use of the gap filling method, Mannga was developed in the Java programming language and all the complex procedures involving Artificial Intelligence were abstracted internally. It is possible to perform a complex process with a few functions, such as gap filling. Code 1 is one example of the method procedure to perform gap filling.

Code 1. Example of gap filling using Mannga implementation

```
01 ManngaParameters parameters = new ManngaParameters();
02 parameters.setErrorMaximumValid(0.05);
03
04 ManngaGapFilling gf = new ManngaGapFilling();
05 gf.setParameters(parameters);
06 gf.train("data.csv", 6, 2, true);
07
08 ManngaResult result = gf.fillGapFoundDuringTraining();
09 for (double output : result.getOutput())
10 System.out.println(output);
```

It can be set some parameter to a better method's performance. The accepted error is one of these parameters, and is set up on lines 1 and 2. Others parameters involve especially to control the ANN and GA. On lines 4 and 5 the method is created and configured. After the initial configuration it is necessary to train the structure. Line 6 shows, with only one command, the data was loaded (informing the file name, number of sensors, the column where the fails are and whether the file has a header in the first line) and the method was trained to recognize all the patterns in the data. Finally, line 8 collect the results

Table 1 - Description of the dataset.

Site	Coordinates	Year	Month(s)	Records	Variables
Lacey Township, New Jersey	39.8379; -74.3791	2009	Mar	2928	Temperature, humidity, net radiation, and incoming shortwave radiation
Florida City, Florida	25.3629; -81.0776	2012	May-Jun	4289	Temperature, wind, humidity, net radiation, incoming shortwave radiation, soil temperature, carbon concentration, and carbon flux
Lawrence, Kansas	39.0561; -95.1907	2012	Mar-Aug	11307	Temperature, humidity, incoming shortwave radiation, soil temperature, carbon concentration, and carbon flux
Campo Bom, Rio Grande do Sul	-29.6743; -51.0640	2014	Aug-Oct	2112	Temperature and humidity

of the gap filling and lines 9 to 10 shows each estimated value.

It can be observed that to run Mannga is not a difficult task. And also that can be easy to incorporate Mannga implementation in other software, even if the developer does not have any knowledge in the method used.

3. Results and Discussions

3.1. Tests results

On the first site, one month's data from New Jersey station were used, which contains 2928 records (without failures) with 15 minutes as frequency of the measurement. In these records failures were inserted in the variables of incoming shortwave radiation, net radiation, humidity and temperature. Being these randomly or in sequence: 146 (5% random), 293 (10% in sequence), and 878 (30% random). The results of the processing are shown in [Table 2](#) with their respective mean absolute error (MAE).

On the second site, data for each 20 minutes of two months from the station at Florida were used with 4289 records, and were inserted: 214 (5% random), 429 (10% in sequence), and 1287 (30% random) failures. The variables chosen were incoming shortwave radiation, net radiation, temperature, humidity and carbon concentration. The results of the processing are shown in [Table 3](#) with their respective mean absolute error (MAE).

On the third site, data collected each 30 minutes during six months from the station at Kansas were used and 11307 records were processed, with 565 (5%) random failures, 1131 (10%) in sequence failures, and 3392 (30%) random failures inserted to test how the proposed method handles multiple failures. The variables used to perform the gap filling were incoming shortwave radiation, temperature, soil temperature, carbon concentration and carbon flux. The results of the processing are shown in [Table 4](#) with their respective mean absolute error (MAE).

On the last site, data collected hourly of three months from the station at Rio Grande do Sul, Brazil, were used and 2112 records were processed, with 105 (5%)

random failures, 211 (10%) in sequence failures, and 633 (30%) random failures inserted to test the proposed method. Temperature and humidity were used to perform gap filling. The results of the processing are shown in [Table 5](#).

The results obtained in gap filling were estimated based on the values of other sensors, obtained in the same place and at the same time as the detected failures. The GA, in addition to determining the configuration parameters of the ANN, also evaluates which sensors are available to be used as input to the neural network training. This is relevant because it can happen that a sensor, which represents a particular climatic variable, has a totally different behavior from the climatic variable estimated, affecting the accuracy of the simulation.

The results showed that Mannga had a good performance with different climatic variables. Sensors such as atmosphere temperature and soil temperature obtained error like 1.42. The carbon flux also obtained good results in experiments (minor error 2.04). However, sensors such as incoming shortwave radiation and net radiation had bad results (109.88 for Kansas dataset), with MAE values far from the average. In all simulations, Mannga robustness is observed, i.e., it was seen uniformity in performance and behavior for different scenarios.

It was also observed, in the experiment using data from Kansas and Florida site, carbon concentration variable needed only one sensor, respectively carbon flux and temperature, to estimate the missing value. Unfortunately, the data related to carbon concentration did not have good accuracy (9.88 on average). It may be possible to improve its precision by using other climatic variables in data series. In order to achieve this, new tests should be performed in the future.

About the processing time to training the method, in the biggest dataset with 6 month of data, the average for training was 67 minutes and 11 seconds. It is a big difference in processing time compared with statistical methods, as can be seen in [Table 6](#), [Table 7](#), and [Table 8](#). Even so, it is an acceptable time to processing this amount of data.

Table 2 - Results from New Jersey dataset. 1st: simulation with 5% random failures, 2nd: simulation with 10% in sequence failures. 3rd: 30% random failures.

Variables	Sensor used	Time (min)			MAE		
		1 st	2 nd	3 rd	1 st	2 nd	3 rd
Incoming shortwave radiation	Wind, carbon flux, net radiation	18:02	13:28	10:38	15.59	21.66	17.84
Net radiation	Temperature	19:36	16:00	18:23	18.07	17.98	18.99
Humidity	Temperature	18:12	43:53	12:04	10.46	9.19	11.12
Temperature	Wind, carbon flux, humidity, incoming shortwave radiation	16:31	28:47	15:23	5.05	2.77	4.61

Table 3 - Results from Florida dataset. 1st: simulation with 5% random failures, 2nd: simulation with 10% in sequence failures. 3rd: 30% random failures.

Variables	Sensor used	Time (min)			MAE		
		1 st	2 nd	3 rd	1 st	2 nd	3 rd
Incoming shortwave radiation	Soil temperature	71:29	35:22	42:21	10.61	15.20	11.25
Net radiation	Temperature, carbon, concentration, carbon flux	67:22	85:14	40:00	23.42	23.07	22.06
Temperature	Wind, net radiation, incoming shortwave radiation, Humidity	37:42	69:59	53:01	1.75	1.42	1.47
Humidity	Wind, temperature	89:32	61:01	65:10	6.78	7.02	6.84
Carbon concentration	Carbon flux	34:08	36:34	25:51	11.53	6.23	9.56

Table 4 - Results from Kansas dataset. 1st: simulation with 5% random failures, 2nd: simulation with 10% in sequence failures. 3rd: 30% random failures.

Variables	Sensor used	Time (min)			MAE		
		1st	2nd	3rd	1st	2nd	3rd
Soil temperature	Temperature, carbon flux, humidity	61:23	56:06	39:05	1.8	2.01	1.88
Temperature	Carbon flux, soil temperature	74:09	53:13	58:03	2.34	2.47	2.35
Incoming shortwave radiation	Carbon concentration	98:01	89:57	41:13	105.67	109.54	109.88
Carbon flux	Humidity, incoming shortwave radiation	57:45	56:58	90:17	2.04	2.23	2.21
Carbon concentration	Temperature	65:14	111:06	55:18	10.54	10.94	10.46

Table 5 - Results from Rio Grande do Sul dataset. 1st: simulation with 5% random failures, 2nd: simulation with 10% in sequence failures. 3rd: 30% random failures.

Variables	Time (min)			MAE		
	1st	2nd	3rd	1st	2nd	3rd
Temperature	13:03	09:05	08:36	0.93	1.20	0.95
Humidity	19:28	11:30	14:22	2.52	1.94	2.16

3.2. Comparison with others methods

In order to evaluate Mannga performance, others gap filling methods were tested with the same datasets. The results can be seen in [Table 6](#), [Table 7](#) and [Table 8](#) showing the MAE obtained in each test with Mannga, Average method and Multiple Linear Regression (MLR) method.

With the simulation of 5% of random failures, Mannga was better compared to Average and MLR method in only two cases (incoming shortwave radiation and net radiation in Florida site). In all cases, Mannga was better than MLR method, except when there were just a few failures in the data series. Average method proved to be very successful in this scenario.

On the simulation of 10% of failures in sequence, in ten cases Mannga was better than the others methods. There are good precisions with several variables, like

incoming shortwave radiation, net radiation, humidity and carbon concentration. In all these tests, Mannga was always better than Average.

In the last simulation, with 30% of random failures, Mannga showed regular results. It was the best in three cases, being the second best method in all the others tests. Therefore, Mannga can be used in scenarios where exist a lot of failures in the dataset. In general, Mannga shows to be a good option to gap filling meteorological data.

3.3. Mannga public availability

As mentioned, Mannga was implemented with the Java programming language. It was included in the framework FICSED and can be downloaded on CEDA website as free software. The website has the necessary documentation to use the method.

Table 6 - Results (MAE) with 5% random failures from others methods compared with Mannga.

Dataset	Variable	Mannga	Average	MLR
New Jersey	Incoming shortwave radiation	15.59	13.14	31.64
New Jersey	Net radiation	18.07	11.29	26.18
New Jersey	Humidity	10.46	0.76	31.47
New Jersey	Temperature	5.05	0.18	5.82
Florida	Incoming shortwave radiation	10.61	27.41	23.77
Florida	Net radiation	23.42	32.52	27.02
Florida	Temperature	1.75	0.19	1.77
Florida	Humidity	6.78	1.46	9.73
Florida	Carbon concentration	11.53	3.99	21.35
Kansas	Soil temperature	1.8	0.05	1.96
Kansas	Temperature	2.34	0.2	2.58
Kansas	Incoming shortwave radiation	105.67	20.58	171.79
Kansas	Carbon flux	2.04	1.25	2.27
Kansas	Carbon concentration	10.54	2.96	94.96
Rio Grande do Sul	Temperature	0.93	1.64	0.52
Rio Grande do Sul	Humidity	2.52	5.91	2.44

Table 7 - Results (MAE) with 10% in sequence failures from others methods compared with Mannga.

Dataset	Variable	Mannga	Average	MLR
New Jersey	Incoming shortwave radiation	21.66	165.30	44.89
New Jersey	Net radiation	17.98	131.31	28.67
New Jersey	Humidity	9.19	37.43	41.62
New Jersey	Temperature	2.77	2.89	2.49
Florida	Incoming shortwave radiation	15.2	307.18	39.35
Florida	Net radiation	23.07	320.7	24.29
Florida	Temperature	1.42	3.33	1.18
Florida	Humidity	7.02	20.11	9.58
Florida	Carbon concentration	6.23	10.99	18.78
Kansas	Soil temperature	2.01	2.35	1.98
Kansas	Temperature	2.47	5.05	2.38
Kansas	Incoming shortwave radiation	109.54	232.89	157.63
Kansas	Carbon flux	2.23	2.5	1.94
Kansas	Carbon concentration	10.94	12.77	52.64
Rio Grande do Sul	Temperature	1.20	1.69	0.59
Rio Grande do Sul	Humidity	1.94	6.89	2.38

4. Conclusions

In this paper we propose a novel method for gap filling meteorological data called Mannga. The great advantage of this method is the flexibility of handle different types of meteorological data, adjusting their structure for each dataset. Another advantage is the possibility of selects the best sensors to estimate the missing value, increasing the accuracy and saving processing time. Besides, if failures occur in sequence, for example, gaps occurring in the data series for hours, days or even months, it is possible to estimate the values, considering

that other sensor variables contain valid data from the same period of failure.

We can list the method's disadvantage as the time to process the data. While Mannga takes minutes to perform the gap filling, others statistical methods takes just seconds. Furthermore, a higher accuracy was found mainly when failures occur in sequence in the dataset compared with other methods.

In general, tests were performed evaluating the proposed method and good results were achieved. Therefore, combined with its public availability, it is expected that the

Table 8 - Results (MAE) with 30% random failures from others methods compared with Mannga.

Dataset	Variable	Mannga	Average	MLR
New Jersey	Incoming shortwave radiation	17.84	21.7	34.47
New Jersey	Net radiation	18.99	18.94	27.93
New Jersey	Humidity	11.12	1.16	29.09
New Jersey	Temperature	4.61	0.27	5.24
Florida	Incoming shortwave radiation	11.25	48.18	25.43
Florida	Net radiation	22.06	45.46	25.42
Florida	Temperature	1.47	0.29	1.75
Florida	Humidity	6.84	1.74	10.34
Florida	Carbon concentration	9.56	4.69	20.95
Kansas	Soil temperature	1.83	0.11	1.9
Kansas	Temperature	2.35	0.31	2.59
Kansas	Incoming shortwave radiation	109.88	29.12	175.12
Kansas	Carbon flux	2.21	1.51	2.51
Kansas	Carbon concentration	10.46	3.24	97.31
Rio Grande do Sul	Temperature	0.95	2.19	0.55
Rio Grande do Sul	Humidity	2.16	9.22	2.36

product of this work assist several research projects in the meteorological area, making meteorological data series more consistent.

Acknowledgments

The authors acknowledge the financial support of the Fundação de Amparo a Pesquisa do Estado de Mato Grosso (FAPEMAT) process 223633/2015. In addition, we would like to thank Gregory Starr, Steven Oberbauer, Kenneth Clark and Nathaniel Brunzell for allows the use of data from Florida Everglades, Cedar Bridge and Kansas Field Station. We also acknowledge INMET for make so easy to obtain data of Brazilian meteorological stations.

References

- ALAVI, N.; WARLAND, J.S.; BERG, A.A. Filling gaps in evapotranspiration measurements for water budget studies: evaluation of a Kalman filtering approach. **Agricultural and Forest Meteorology**, v. 141, n. 1, p. 57-66, 2006.
- BODEN, T.A.; KRASSOVSKI, M.; YANG, B. The AmeriFlux data activity and data system: an evolving collection of data management techniques, tools, products and services. **Geoscientific Instrumentation, Methods and Data Systems**, v. 2, n. 1, p. 165-176, 2013.
- FALGE, E.; BALDOCCHI, D.; OLSON, R.; ANTHONI, P.; AUBINET, M.; BERNHOFER, C.; BURBA, G.; CEULEMANS, R.; CLEMENT, R.; DOLMAN, H.; GRAINER, A.; GRUNWALD, T.; HOLLINGER, D.; JENSEN, N.-O.; KATUL, G.; KERONEN, P.; KOWALSKI, A.; TA LAI, C.; LAW, B.E.; MEYERS, T.; MONCRIEFF, J.; MOORS, E.; MUNGER, J.W.; PILEGAARD, K.; RANNIK, U.; REBMANN, C.; SUYKER, A.E.; TENHUNEN, J.; TU, K.; VERMA, S.; VESALA, T.; WILSON, K.; WOFSY, S. Gap filling strategies for defensible annual sums of net ecosystem exchange. **Agricultural and forest meteorology**, v. 107, n. 1, p. 43-69, 2001.
- HAYKIN, S. **Neural networks: a comprehensive foundation**, Prentice-Hall Upper Saddle River. NJ MATH, 1999.
- HORTON, N.J.; LIPSITZ, S.R. Multiple imputation in practice: comparison of software packages for regression models with missing variables. **The American Statistician**, v. 55, n. 3, p. 244-254, 2001.
- HU, Z.; YU, G.; ZHOU, Y.; SUN, X.; LI, Y.; SHI, P.; WANG, Y.; SONG, X.; ZHENG, Z.; ZHANG, L.; LI, S. Partitioning of evapotranspiration and its controls in four grassland ecosystems: Application of a two-source model. **Agricultural and Forest Meteorology**, v. 149, n. 9, p. 1410-1420, 2009.
- HUI, D.; WAN, S.; SU, B.; KATUL, G.; MONSON, R.; LUO, Y. Gap-filling missing data in eddy covariance measurements using multiple imputation (MI) for annual estimations. **Agricultural and Forest Meteorology**, v. 121, n. 1, p. 93-111, 2004.
- KATO, T.; TANG, Y.; GU, S.; HIROTA, M.; DU, M.; LI, Y.; ZHAO, X. Temperature and biomass influences on inter-annual changes in CO₂ exchange in an alpine meadow on the Qinghai Tibetan Plateau. **Global Change Biology**, v. 12, n. 7, p. 1285-1298, 2006.
- LEAUTHAUD, C.; CAPPELAERE, B.; DEMARTY, J.; GUICHARD, F.; VELLUET, C.; KERGOAT, L.; VISCHER, T.; GRIPPA, M.; MOUHAIMOUNI, M.; BOUZOU MOUSSA, I.; MAINASSARA, I.; SULTAN, B. **A 60-year reconstructed high-resolution local meteorological data set in Central Sahel (1950–2009): evaluation, analysis and application to land surface modelling**. *International Journal of Climatology*, 37: 2699-2718, 2017.
- MALIK, A.; KUMAR, A. Pan evaporation simulation based on daily meteorological data using soft computing techniques

- and multiple linear regression. **Water Resources Management**, v. 29, n. 6, p. 1859-1872, 2015.
- MISHUROV, M.; KIELY, G. Gap-filling techniques for the annual sums of nitrous oxide fluxes. **Agricultural and forest meteorology**, v. 151, n. 12, p. 1763-1767, 2011.
- MITCHELL, M. **An introduction to genetic algorithms**. MIT press, 1998.
- MOFFAT, A. M.; PAPAIE, D.; REICHSTEIN, M.; HOLLINGER, D.Y.; RICHARDSON, A.D.; BARR, A.G.; BECKSTEIN, C.; BRASWELL, B.H.; CHURKINA, G.; DESAI, A.R.; FALGE, E.; GOVE, J.H.; HEIMANN, M.; HUI, D.; JARVIS, A.J.; KATTGE, J.; NOORMETS, A.; STAUCH, V.J. Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. **Agricultural and Forest Meteorology**, v. 147, n. 3, p. 209-232, 2007.
- MOHAN, T.S.; RAO, T.N. Differences in the mean wind and its diurnal variation between wet and dry spells of the monsoon over Southeast India. **Journal of Geophysical Research: Atmospheres**, v. 121, p. 6993-7006, 2016.
- OOBA, M.; HIRANO, T.; MOGAMI, J.-I.; HIRATA, R.; FUJINUMA, Y. Comparisons of gap-filling methods for carbon flux dataset: a combination of a genetic algorithm and an artificial neural network. **Ecological Modelling**, v. 198, n. 3, p. 473-486, 2006.
- RODRIGUES, A.; PITA, G.; MATEUS, J. Turbulent fluxes of carbon dioxide and water vapour over an eucalyptus forest in Portugal. **Silva Lusitana**, v. 13, n. 2, p. 169-180, 2005.
- SCHAFFER, J.L. Multiple imputation: a primer. **Statistical methods in medical research**, v. 8, n. 1, p. 3-15, 1999.
- SHAO, C.; CHEN, J.; LI, L.; TENNEY, G.; XU, W.; XU, J. Role of net radiation on energy balance closure in heterogeneous grasslands. **Biogeosciences Discussions**, v. 8, n. 2, p. 2001-2033, 2011.
- STAUB, B.; HASLER, A.; NOETZLI, J.; DELALOYE, R. Gap-Filling algorithm for ground surface temperature data measured in permafrost and periglacial environments. **Permafrost and Periglacial Processes**, v. 28, p. 275-285, 2017.
- SULLIVAN, T.R.; SALTER, A.B.; RYAN, P.; LEE, K.J. Bias and precision of the “multiple imputation, then deletion” method for dealing with missing outcome data. **American journal of epidemiology**, v. 182, n. 6, p. 528-534, 2015.
- UYANIK, G.K.; GÜLER, N. A study on multiple linear regression analysis. **Procedia - Social and Behavioral Sciences**, v. 106, p. 234-240, 2013.
- VENTURA, T.M.; OLIVEIRA, A.G.; MARTINS, C.A.; FIGUEIREDO, J.M.; GOMES, R.S.R. **Study of how the integration of artificial neural network and genetic algorithm should be made for modeling meteorological data**. In: 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), p. 719-722, 2015.
- WILSON, K.; BALDOCCHI, D. Comparing independent estimates of carbon dioxide exchange over 5 years at a deciduous forest in the southeastern United States. **Journal of Geophysical Research. D. Atmospheres**, v. 106, p. 34, 2001.
- ZHOU, J.; DAI, F.; ZHANG, X.; ZHAO, S.; LI, M. Developing a temporally land cover-based look-up table (TL-LUT) method for estimating land surface temperature based on AMSR-E data over the Chinese landmass. **International Journal of Applied Earth Observation and Geoinformation**, v. 34, p. 35-50, 2015.

Internet Resources

Ameriflux: <http://ameriflux.lbl.gov>

INMET: <http://inmet.gov.br>

CEDA: <http://ceda.ic.ufmt.br>

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License which permits unrestricted non-commercial use, distribution, and reproduction in any medium provided the original work is properly cited.