Evaluation of a Business Intelligence Tool with Grey Forecasting Model in a Cloud Environment

M. M. Silva[1], A. A. A. Silva[2], N. Ferraz Jr.[3], A. E. Guelfi[4], M. T. Azevedo[5], C. L. C. Larieira[6], S. T. Kofuji[7]
[1, 2] IPT – Instituto Pesquisas Tecnológicas, São Paulo, Brazil
marcello.maier@outlook.com

[2] UNIP – Universidade Paulista, São Paulo, Brazil
SENAC – Centro Universitário SENAC, São Paulo, Brazil
anderson@uol.com.br

[3, 5, 7] USP – Universidade de São Paulo, São Paulo, Brazil
{norisjunior@usp.br} {marcelo.azevedo@pad.lsi.usp.br} {kofuji@usp.br}

[4]UNOESTE - Universidade do Oeste Paulista, São Paulo, Brazil
guelfi@unoeste.br

[6]FGV – Fundação Getúlio Vargas, São Paulo, Brazil
larieira@hotmail.com

ABSTRACT: *The adoption of cloud computing by small and medium businesses generates several different challenges, such as cost management, lack of human and technological resources, and difficulty in managing the cloud environment in an aligned manner, which makes the decision-making process difficult. Business Intelligence (BI) tools are considered decision support systems and can assist in managing cloud resources, but they lack validation. An approach with other predictive tools, which validates the results of BI tools, can solve this limitation. This work uses the predictive capabilities of a BI, validated by predictive methods, to create scenarios foreseen for the cloud environment that correlate to information relevant to the business area. The results show that the Grey Forecasting Model (GM) can be used to validate forecasts generated by BI. In addition, it was found that the accuracy of forecasts also improves in certain scenarios. These scenarios address different moments, showing correlations between the cloud environment and the business area, further improving the decision-making process in small and medium-sized companies.*

Keywords: Cloud Computing, Business Intelligence (BI), Decision Support Systems, Grey Theory GM(1,1), Small and Medium Enterprises (SME)

Received: 2 May 2021, Revised 10 August 2021, Accepted 29 August 2021

DOI: 10.6025/jism/2021/11/4/117-133

1. Introduction

Cloud computing provides several benefits to Small and Medium Enterprises (SMEs), such as the possibility of cost reduction, increased competitiveness in Information Technology (IT), flexibility, and scalability [1]. However, a cloud environment can also create obstacles to SMEs, such as a lack of control over the technology and management of the environment. [2]. This is usually the result of a limited IT environment, that is, with few human and technological resources [3].

Figure 1 shows that although SMEs face challenges after adopting cloud computing, there is scarce literature about this post-adoption scenario. A Decision Support System (DSS) could be the outlet for SMEs to better manage their cloud environment. DSS are systems that help decision-making through analysis and estimates [4].

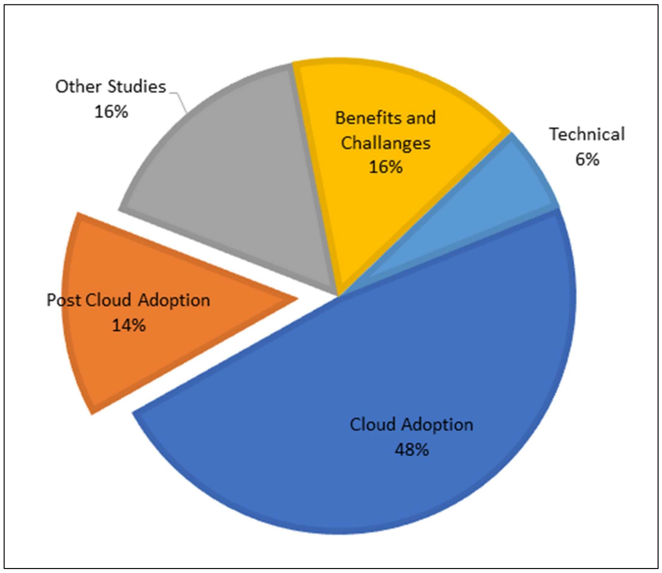

Figure 1. Distribution of studies on cloud computing and SMEs, adapted from [5]

Business Intelligence (BI) tools are examples of data-driven DSS. Some BI tools have predictive capabilities, which can generate estimates. Once validated, these estimates can help decision-makers in managing the environment. IT management and timely decision-making positively influence SMEs, generating competitive advantages for businesses, such as scalability, accessibility, and access to more modern IT resources [5]. For this reason, the current paper demonstrates the use of a BI tool in a predictive manner with results validated by the Grey Forecasting Model (GM(1,1)) in an SME cloud environment.

BI is not seen as a model or a predictive system but as a decision support tool that offers predictive characteristics, being able to estimate results to aid decision-making. The aim of the current paper is to validate the estimates generated by BI with GM(1,1), which is a predictive method established in the academic environment. The experiment demonstrated in this work starts before the COVID-19 pandemic and runs until February 2021, when the pandemic is still a reality in Brazil. As a result, the experiment identifies SME behavior patterns found during the pandemic, which is an additional and unexpected contribution.

In addition to the Introduction section, this paper brings some theoretical concepts in section II Background, a bibliographical survey on state of the art in section 3 Related Works, the proposal in section 4 Predictive Decision Support System, details about the experiments in section 5 Experiment and Proposal Application, a discussion of the results in section 6 Results and General Discussion, and conclusions in section 7 Conclusion.

2. Background

This section introduces the fundamental concepts of the technologies cited in this paper.

2.1. Cloud Computing

Cloud computing is defined by the National Institute of Standards and Technology (NIST) as a model for accessing computing resources on demand, with fast provisioning and minimal interaction with the service provider [6]. It is divided into three models: (1) Software as a Service (SaaS) – Applications hosted in the cloud, with no control over the infrastructure of these applications; (2) Platform as a Service (PaaS) – Applications in progress or application support structure; and (3) Infrastructure as a Service (IaaS) - Allocation of complete computing resources.

In addition, there are other service models such as Storage as a Service (StaaS) [7], BI as a Service (BIaaS) [8], API as a Service (APIaaS), and others defined as Everything as a Service (XaaS) [9].

2.2. Decision Support Systems

DSS are systems used to support decision-making and problem solving, derived from research on decision-making in the 1950s [10]. These systems usually help decision-making through the production of reports or mathematical simulations [11], with BI being an umbrella term for these DSS [12], [13].

BI tools are considered data-oriented DSS, with the main function of accessing and manipulating databases [4]. Microsoft Power BI is an example of a DSS, capable of addressing the flow of analytics in a complete way, with a local predictive capability.

2.3. Predictive Methods

Also, in the context of predictions, predictive methods are a set of actions, usually based on mathematical functions, with the ability to identify patterns and thus allow the presentation of predictions based on previous training in a data set. One of the basic characteristics of the decision-making process is precisely to anticipate future situations or events in order to create a basis for planning [14].

There are several ways to generate predictions, such as applying linear regression models, time series analysis, Artificial Neural Networks (ANN) [15], and Fuzzy Theory [16], ANN being a popular method with universal approach characteristics and learning capability [17]. Both ANN and linear regression tend to require a large amount of data, making it difficult to use them in small or incomplete databases. [16]. Difficulty in dealing with a neural network structure and unstable training can also impede this method [17]. A similar criticism is made of Fuzzy Theory, which is used for cognitive uncertainties and makes learning difficult through extensive parameters and complex structures [15].

Unlike the methods mentioned, the Grey Theory stands out, which is used in small data samples and requires minimal information to process uncertainties [18]. As part of the Grey Theory concepts, GM(1,1) is a prediction model used in: the energy industry, systems control, integrated circuits, among others [16].

2.4. Grey Forecasting Model

The *GM*(*n, m*) model is composed of three basic operations, Accumulated Generation (AGO), which has the function of solving the GM(1,1) differential equation, smoothing randomness, and determining the value n-steps ahead, and Inverse Accumulated Generation (IAGO), an operation that provides the data from the determination of the AGO. An example of the operation can be seen in the calculation of [19].

$$X^{(0)} = \left(X^{(0)}(1), X^{(0)}(2), \ldots, X^{(0)}(n)\right), \quad n \geq 4 \tag{1}$$

Where $X^{(0)}$ represents a non-negative time sequence, and n is the sample size. AGO ($X^{(0)}$) = $X^{(1)}$ is the result of the AGO operation in $X^{(0).}$

$$X^{(0)} = \left(X^{(0)}(1), X^{(0)}(2), \ldots, X^{(0)}(n)\right), \quad n \geq 4 \tag{2}$$

$$X^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i), k = 1,2, \ldots, n \tag{3}$$

A new mean sequence is generated from $X^{(1)}$ called $Z^{(1)}$ which is the mean value of the adjacent data.

$$Z^{(1)} = \left(Z^{(1)}(1), Z^{(1)}(2), \ldots, Z^{(1)}(n)\right), \quad n \geq 4 \tag{4}$$

$$Z^{(1)} = 0.5x^{(1)}(k) + 0.5x^{(1)}(k-1), \ldots k = 2, \ldots, n \tag{5}$$

The GM differential equation (1.1) is estimated by least squares.

$$X^{(0)}(k) + az^{(1)}(k) = b \tag{6}$$

having as an alternative:

$$\frac{dx^1(t)}{dt} + ax^{(1)}(t) = b \tag{7}$$

The solution of $X^{(1)}$ (t) in time k:

$$x_p^{(1)}(k+1) = \left[x^{(0)}(1) - \frac{b}{a}\right] e^{-ak} + \frac{b}{a} \tag{8}$$

Where $[a,\ b]^T$ is a sequence of parameters, defined by:

$$[a, b]^T = (B^T B)^{-1}\, B^T Y \tag{9}$$

Where $Y = [X^{(0)}(2), X^{(0)}(3), \ldots, X^{(0)}(n)]^T$ (10)

$$B = \begin{bmatrix} -Z^{(1)}(2) & 1 \\ -Z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -Z^{(1)}(n) & 1 \end{bmatrix} \tag{11}$$

Using the IAGO operation, it is possible to stabilize the Grey model and obtain the predicted value from the original data in time $(k + 1)$ and $(k + H)$:

$$x_p^{(0)}(k+1) = \left[x^{(0)}(1) - \frac{b}{a}\right] e^{-ak}(1 - e^{a}) \tag{12}$$

$$x_p^{(0)}(k+H) = \left[x^{(0)}(1) - \frac{b}{a}\right] e^{-a(k+H-1)}(1 - e^{a}) \tag{13}$$

3. Related Works

Regarding the proposal addressed in this article, [2] present a model for adopting cloud computing for SMEs, demonstrating the benefits and challenges. The work highlights trends and performs an analysis of cloud deployment in different types of SMEs, as well as citing the ease of scaling the business, cost reduction, and access to technologies previously limited to large companies as reasons for SMEs to migrate to the cloud.

On the other side of the spectrum, challenges are also cited, such as lack of human resources, decision-making without trained IT professionals, lack of time to implement, and runaway cost growth. The authors suggest a conceptual three-phase implementation model that helps SMEs plan for cloud migration. However, the model is only conceptual and does not have validation substantiated by practical experiments. In addition, the model does not address the challenges encountered after the adoption of cloud computing, being reserved only to aid migration.

[20] demonstrate the Analytical Hierarchy Process (AHP) decision-making method in cloud computing adoption by SMEs.

The study includes two SMEs with different interests, which allows the authors to define criteria that correspond to the best alternatives for each SME.

The AHP generates a priority index, indicating the best alternatives for SMEs within cloud computing models such as SaaS, PaaS, and IaaS. Despite demonstrating the use of the AHP in decision-making in a practical way, the model is fed with characteristics defined by the authors, and not real data from SMEs, and focuses only on the cloud migration phase, not a post-adoption moment.

In [21], a literature review is presented that highlights the lack of research on post-adoption to cloud computing by SMEs. The important research gap observed in the cited works is also reinforced in [5]. The study uses Resources-Based View Theory (RBV) to assess an SME's performance in its market environment, adapting to understand the relationship between cloud computing resources and an SME's performance. The work demonstrates hypotheses based on the RBV, however, the use of the proposed method is not presented. This indicates that the proposed method could help future research on the use of cloud computing by SMEs. Hand in hand with the adoption of cloud computing by SMEs, yet it does not make a simulation using real data nor presents results of its methods.

The work of [4] presents concepts about DSS. In general, the authors emphasize the Model-driven DSS, in which computer systems use financial models to assist decision-making. However, the work also cites definitions for several other DSS, including Data-driven, which use management reports, data analysis, and BI tools to access and process data and assist decision-making. The work presents a broad theoretical basis and describes the use of DSS with mathematical models and simulations but does not demonstrate the use of any DSS in a practical way.

[22] present the importance of using BI in SMEs, showing that BI was a technology previously reserved for large companies. These tools are fundamental in the decision-making process, bringing advantages to the SME, such as the reduction in dispersed information, ease of access to information, and flexibility in adapting to the company. A four-phase framework is proposed as an example of the applicability of BI as a DSS. The phases go through planning, software definition, intelligence, and dissemination. The authors conclude that BI tools are essential for decision-making in companies of any size. Despite this, the work also does not present practical uses of a BI tool.

Power BI is ranked as a market leading tool, above others like [23]. In addition, for an SME with limited resources, free or low-cost tools are the first option. In this regard, Power BI again stands out as the cheapest tool available [24], being another highlight its ability to generate forecasts locally, which is one of the focuses of this paper.

In [16], the use of the GM(1,1) model part of the Grey Theory is presented. The work demonstrates a forecast study using China's electricity consumption data. The work shows that the GM(1,1) performs forecasts in various fields such as transport, business, and electricity. The GM(1,1) is suitable for predictions with a small sample, however, the authors suggest improving the model's accuracy using Genetic Programming (GP). GP can generate more assertive solutions through the crossing, mutating, and reproduction rules while also showing good performance in small databases. The effectiveness of the proposal is compared using two evaluation indices, which are Percentage Error (PE) and Mean Absolute Percentage Error (MAPE), that measure prediction accuracy statistically. The authors achieve good results when using GM(1,1) for data prediction and can also improve prediction accuracy using GP. The results presented corroborate the proposal of this paper, which aims to use the GM(1,1) prediction validation with a BI tool. The use of PE and MAPE statistical methods is also important for comparing results with more directly related works. However, it is still fair to note that the authors' work is applied to small samples, leaving room for using GM(1,1) with larger samples.

The work of [25] demonstrates the measurement of the financial performance of a venture capital company between the years 2001 and 2003. Using variables from financial ratios and company attributes, the paper aims to extract the most significant indicators, such as the ability to pay a short-term debt, profitability, cash flow. This is performed using Globalization Grey Relational Analysis (GRA), and the Grey Decision-Making (GDM) method is applied to evaluate the performance of the samples. GRA is a quantitative analysis used to explore similarity and dissimilarity and measure the correlation between factors. The GDM is a model for choosing a strategy to better handle a situation, used in the paper as a means of obtaining the weighted average value for each year. The work concludes by indicating the attribute that most affects the company's financial performance, and the authors are successful in using Grey models in the analysis of historical data.

Other examples of the use of Grey Theory can be found in the work of [26], who compared predictions of different GM(1,1) variants. Comparing the parity data between USD and Euro, the authors analyze the result of different forecasting methods based on Grey Theory, such as the original GM(1,1), GM(1,1) modified with error modeling and Fourie, and the Grey Verhulst Model (GVM). Still in predictive methods, [18] also use GM(1,1) as the main forecasting method, comparing other methods derived from the Grey Theory using renewable energy consumption data to predict usage trends. The Grey Theory and its methods derived from the Grey Forecasting Model are adequate to generate forecasts when the data sample is small, this being the main choice among the cited authors when compared to other means of forecasting such as ANN, Fuzzy, or regression. Figure 2 aims to demonstrate some similarities of subjects between this paper and the related works.

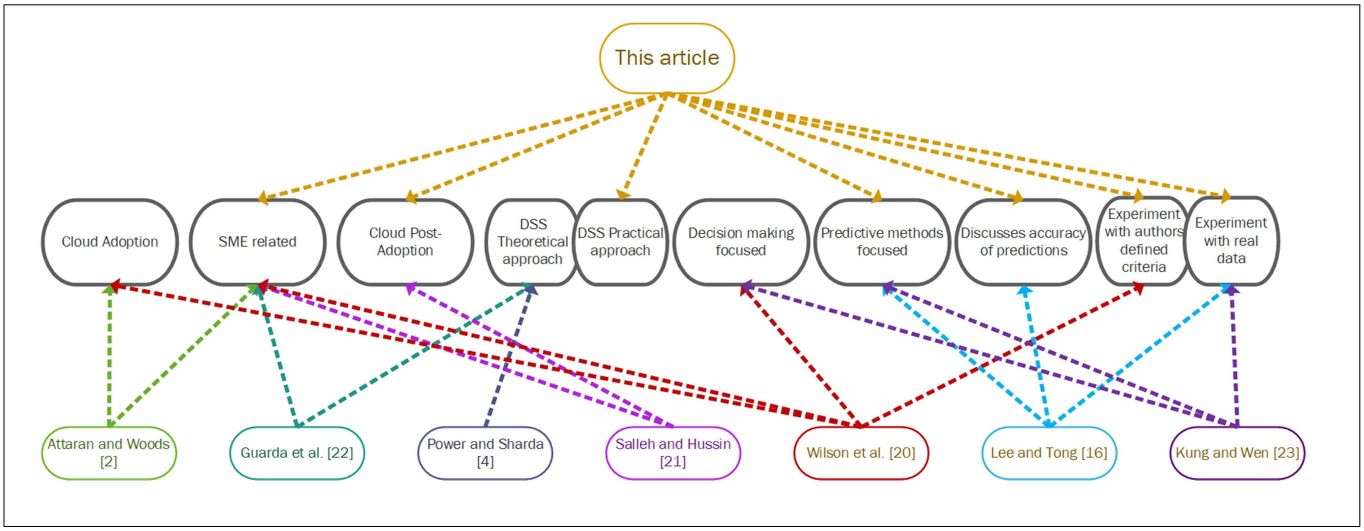

Figure 2. Connections between related works and this paper

These subjects, shown in Figure 2 in grey, are some of the bases used in this paper. In multicolor patterns are the authors of the related work, the arrows indicating the subjects each work relates to, the dark yellow color indicates this paper emphasizing the subjects in common.

As is described in the introduction section, Cloud Computing brings benefits to SMEs. A major part of these studies focuses on making possible or facilitating SMEs' migration to the cloud [2][20], but that is not the only scenario found in SMEs, they also face challenges after migration, like management, cost, security, and others. The lack of research in post-adoption pointed by [21] encourages this paper. It is also mentioned the benefits that BI can bring to the SME, making easy access to information and analysis. A step further can be taken when using predictive methods to compare and validate results given by a BI tool, some methods are well described like GM(1,1), but none of the researched works have yet used cloud computing data to generate forecasts, being a new contribution of this paper, the same can be said about the practical application of these papers using real-world data, which seems not to be common as these works often carry theoretical approaches or data defined by the authors.

4. Utilization Proposal of a Predictive DSS

This section presents the proposal of the current work, which consists of using a BI tool as a DSS validated by GM(1,1). The scenario used in this work comprises the public cloud and belongs to an SME in the field of Business to Business (B2B) in retail, with cloud computing consumption data being the main objective of analysis. The proposal is divided into three phases; phase one, data collection and definition of variables; phase two, predictive tools; and phase three, comparison and analysis. The phases can be seen in Figure 3.

Phase one of the proposal includes the elaboration of a logbook (external logs) that follows the evolution of the consumption data of resources in the SME's cloud. The purpose of the logbook is to improve understanding of the results predicted by BI, based on the analysis of external variables (discussed later). In this phase, the consumption data of the SME cloud

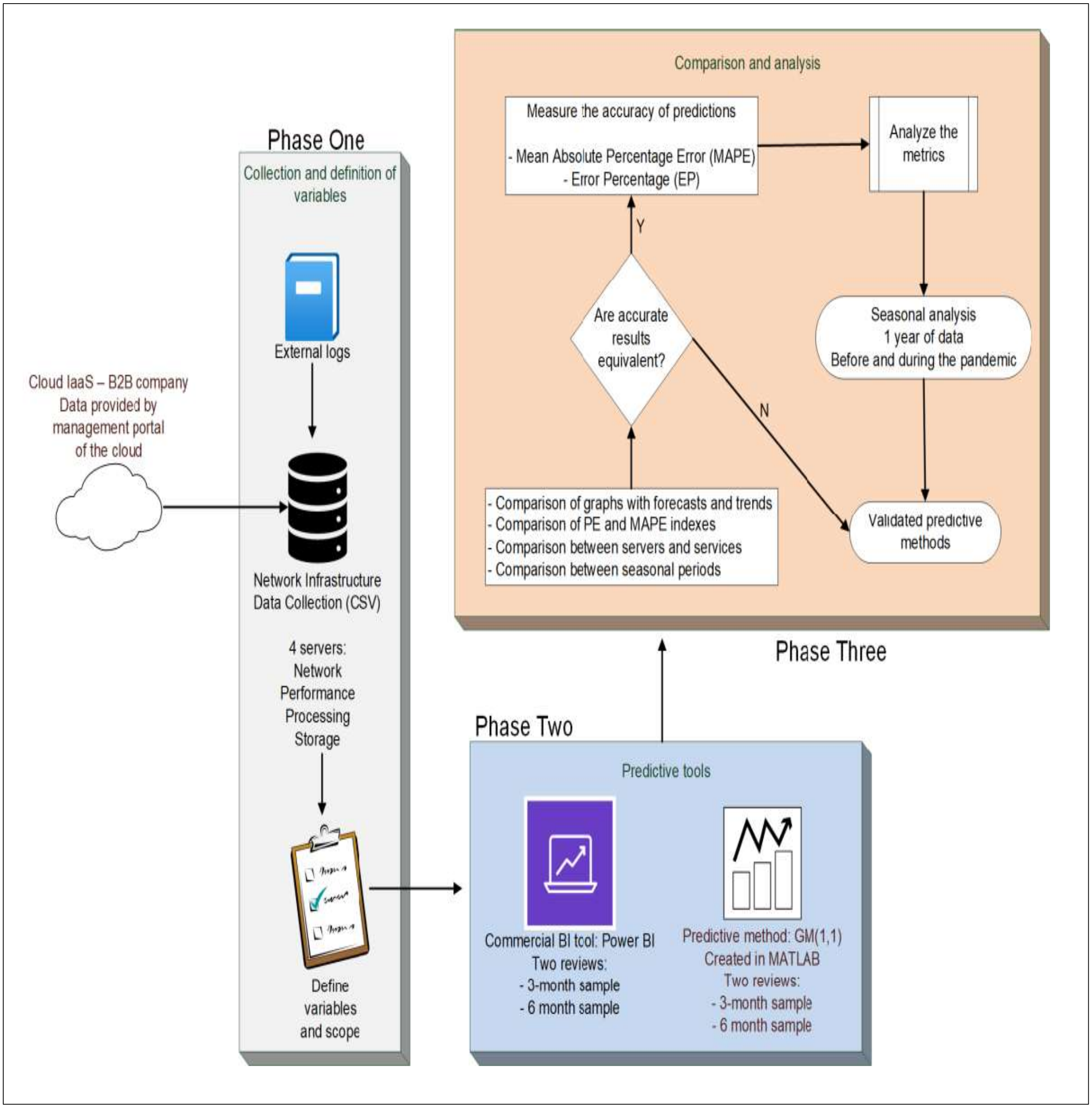

Figure 3. Predictive DSS phases

environment are also collected, and the data periods that will be worked on are defined.

Phase two of the proposal consists of the application of two distinct predictive means, the first being a commercial BI tool (Power BI) and the second the GM(1, 1) method applied via MATLAB. Predictive media receive two equal data samples to generate a forecast of equal size.

Phase three is the final phase of the proposal and addresses the complete comparison and analysis of the expected results, using statistical methods (PE and MAPE) and also considering the different periods of the data collected.

5. Experiment and Proposal Application

For the application of BI in SME data, it is first necessary to define the data that will form the knowledge base of the proposed system [26]. The data used are collected from the SME cloud environment. These data quantitatively indicate the use of resources in a given period. The use of this cloud resource consumption data follows [27] guidelines, which indicates that the evaluation of predictive methods should be performed using data from situations for which predictions will be used. A feature of this proposal is the long data collection, carried out over 15 months, which enables the visualization of different scenarios in different periods.

The environment has four servers hosted on the IaaS model and offers eight monitoring measures: Central Processing Unit (CPU) percentage (1) minimum; (2) maximum; (3) mean; Data written to disk (4) minimal; (5) maximum; (6) mean; (7) Data In (network) mean; (8) Data Out (network) mean.

The proposal presented contemplates the prediction using two approaches: first, the use of a BI tool as a DSS is addressed, as defined by [4]. This BI tool contains a local predictive function and allows the generation of forecasts for a certain period on the collected data. As reported by [28], the predictive capability is a key function of this type of tool.

In parallel to the use of the BI tool, predictions are performed using the predictive model GM(1,1) [16], [18],[26]. The comparison of the results of the GM(1,1) application with the results presented by the BI tool allows quantitative and comparative data analysis.

By comparing the result, it is possible to validate the use of the BI tool as a predictive DSS. This enables the tool to be used to assist decision-making. [27] reports that comparing predictions is a means of validation that brings more confidence about the results. The use of PE and MAPE are also metrics that allow the comparison and validation of results, as noted in [16] and [18]. The PE calculation is performed following equation 14.

$$PE = \frac{Y_{(k)} - X_{(k)}}{X_{(k)}} \times 100\% \tag{14}$$

Calculation of MAPE is performed using equation 15.

$$MAPE = \frac{\sum_{k=1}^{n} \frac{X_{(k)} - Y_{(k)}}{X_{(k)}}}{n} \times 100 \tag{15}$$

MAPE also classifies the predictions according to the result of the equation, as shown in Table 1.

MAPE	**Forecasting level**
< 10	High forecasting
10 - 20	Good forecasting
20 - 50	Reasonable forecasting
> 50	Weak Forecasting

Table 1. MAPE criteria, adapted from [21]

In addition, in both forecast models, there are four forecast windows, three of them monthly (Figure 4) and one quarterly (Figure 5).

The intention in the windows of Figures 4 and 5 is to discover the behavior of forecasts in the same forecasting length (three forecasted months), however, with different data inputs, as mentioned, three monthly forecasts (three 30-day forecasts) resulting in three forecast months, and a quarterly forecast, also resulting in three forecast months (a 90-day forecast).

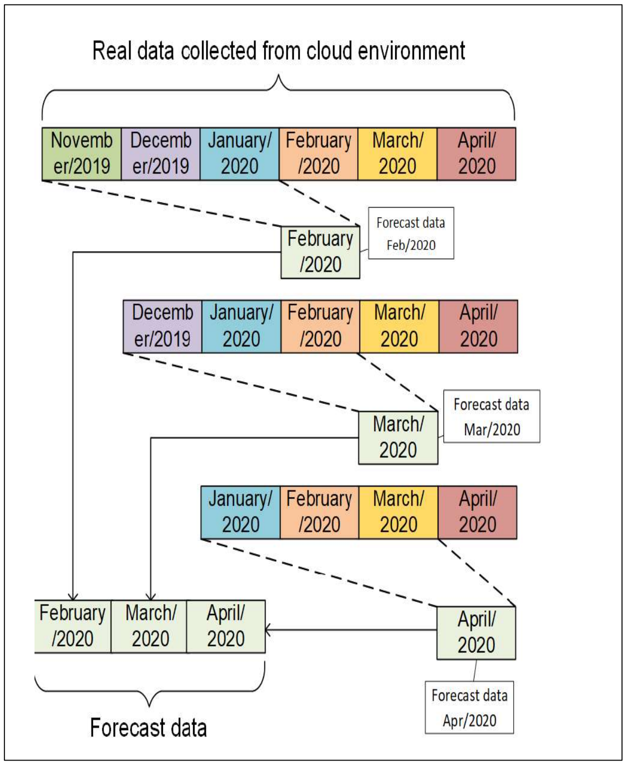

Figure 4. Detail of the use of three months of data to generate one monthly forecast

Figure 4 shows the use of three months of data entry to generate one monthly forecast, this process is performed three times, creating three monthly forecasts.

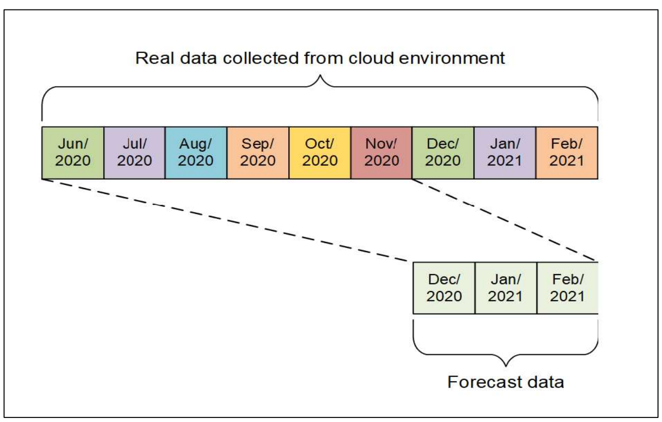

Figure 5. Detail of the use of six months of data to generate three monthly forecasts

Like the monthly forecast, Figure 5 demonstrates the quarterly forecast, which uses six months of input data to generate three forecast months. However, this time only a quarterly forecast is created.

These forecast windows also help the experiment to divide the results into different moments of the COVID-19 pandemic, which, despite not being in the initial plan of this work, ended up becoming part of the experiment due to the long duration of data collection (15 months), and a divider of the results.

After comparing and validating the results, a seasonal analysis was carried out in conjunction with the logbook. This makes possible to identify possible influences caused by external variables such as the home office policy adopted by the company, audience behavior patterns, webinars, and events held by the SME. The proposed analysis addresses different time periods to identify similarities. The periods are divided into three moments of the pandemic, the first moment being before (pre-pandemic), the second moment being considered as the beginning of the pandemic, and a final stage, called in this work the consolidated pandemic.

6. Results and General Discussion

This section presents the results obtained with the experiment proposed in the previous section. Cloud computing data from an SME are used to generate forecasts through two different means, a BI tool (Power BI) and the GM(1,1) method. The results of the forecasts generated will be compared, and their accuracy measured and validated by applying MAPE.

The data collection resulted in a database containing information regarding the usage of cloud resources (CPU, disk, network in and out) of the four servers. During the 15 months of data collection, it was possible to record a great amount of cloud usage, including the behavioral changes caused by the COVID-19 pandemic, which significantly influences the environment by reducing the usage of some resources.

In the period considered as pre-pandemic, between November 2019 and January 2020, the servers present usage of resources that is considered normal. At this time, there were still no signs of a possible pandemic, and the SME is normally operated, as shown in Figure 6.

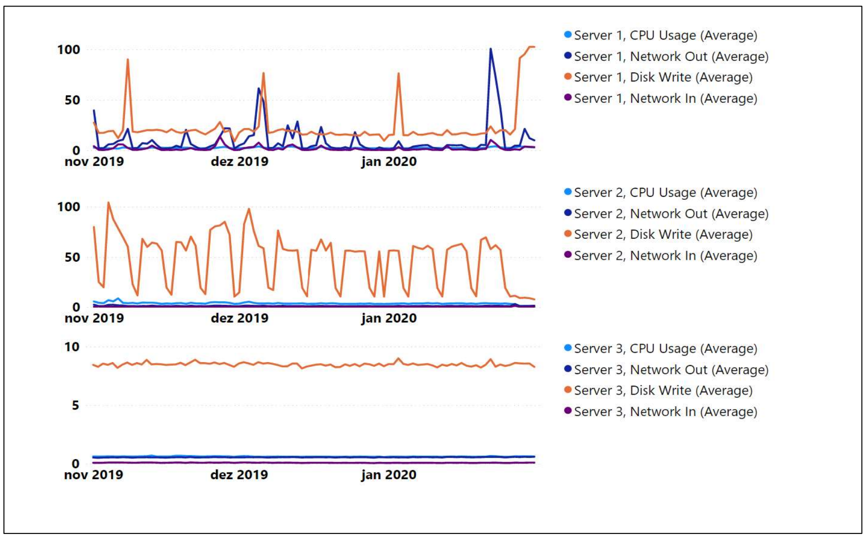

Figure 6. Behavior of the servers in the pre-pandemic period

Between February and April 2020, the first news about COVID-19 appeared. The first case of contamination was reported in February, and the state of the pandemic was officially decreed by the Brazilian government in April. There are enough facts to consider this period as the beginning of the pandemic, and it is precisely the window that demonstrates the beginning of the change in the SME's resource usage, as shown in Figure 7.

Between June 2020 and February 2021, the use of resources follows a new pattern, considered by this work as a period of the consolidated pandemic. The servers show a certain regularity during this period, and there is a little more activity in the disk resources of servers 1 and 3. However, the clearest change is in server 2's disk usage, which remains almost null, very different from the usage presented in the pre and initial pandemic periods. Figure 8 demonstrates this new pattern of behavior.

With the usage of resources and recorded periods, the forecasts are executed and result in 256 different scenarios, with the

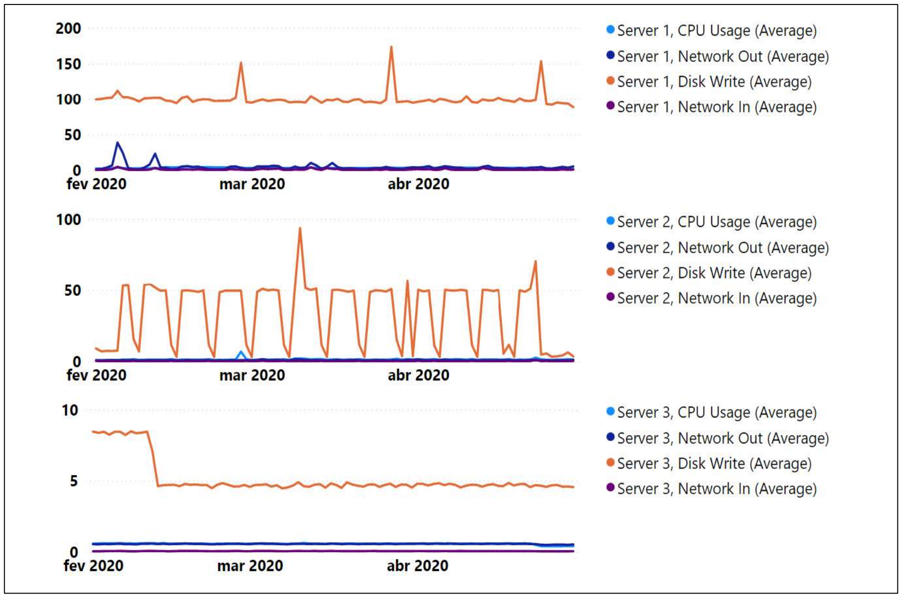

Figure 7. Behavior of the servers at the beginning of the pandemic period

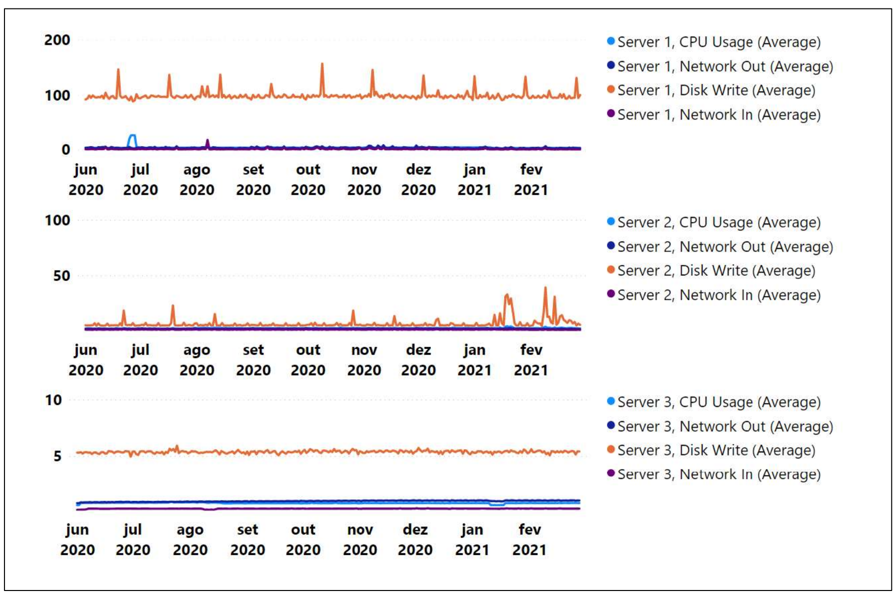

Figure 8 behavior of the servers in the consolidated pandemic period

combination of the four servers, eight measures (min, max and mean for CPU; min, max and mean for Disk; mean for Network In and Out), two predictive tools (BI and GM), and four forecast windows (three-monthly, and one quarterly forecast).

Allied to the size of the data collection, the forecasts were divided into two periods called pandemic onset and consolidated pandemic. The division of periods reveals behavior patterns in servers such as the change in resource usage (CPU, Disk, Network) on three of the servers, servers that host internal services (applications, network services, fileserver) of the SME. Resource consumption reduced drastically to the point of suggesting some server downtime.

On the other hand, one of the servers, which, unlike the previous servers, hosts external services, such as images on the SME website, did not show major changes, demonstrating a regular consumption of resources during the 15 months observed.

The predictions made in the initial period of the pandemic try to replicate a non-regular behavior of the servers that service the SME internally. This causes difficulty in reproducing the real scenario in both predictive means. The behavior of forecasts improves in the period of the consolidated pandemic, which demonstrates regular usage. The regularity is replicated very similarly in the predicted consolidated pandemic results.

As for the server that serves the SME audience, having a regular usage of resources already in the initial period of the pandemic, it makes the forecasts of both periods more assertive than the previous ones. Despite the regularity of usage in both periods, it is also possible to identify greater similarities in the quarterly forecasts.

In addition to usage behavior, the experiment reveals that predictions made by BI tend to follow changes in consumption, with more variation between highs and lows. Unlike BI, predictions made with GM(1,1) tend to result in more linear estimates, as shown in Figure 9.

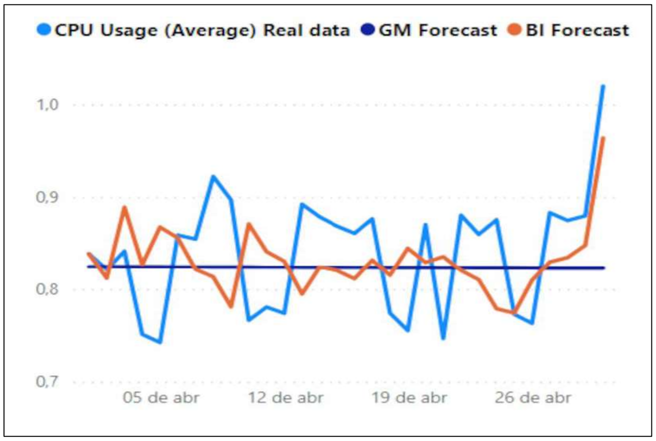

Figure 9. Graph representing data behavior of one of the servers

Despite the visual difference in the results, the comparison and validation of predictions via PE and MAPE reveal that both estimates have similar behavior in practically all measures. The PE calculations demonstrate a very similar percentage error between the predictions of both BI and GM. Even in cases where the numerical results were not similar, the behavior curves show high similarity, as shown in Figure 10.

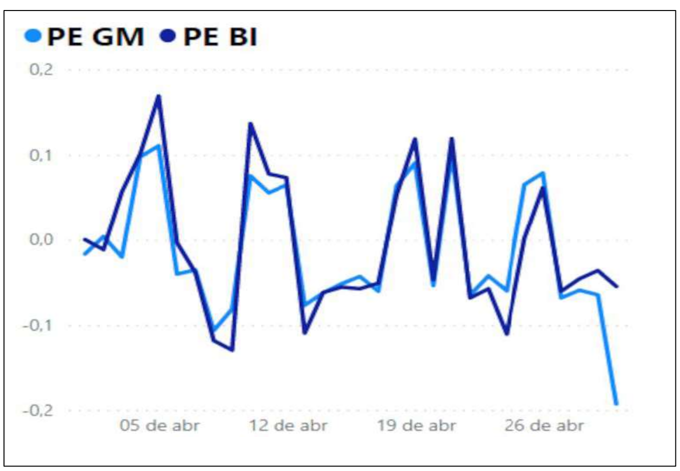

Figure 10. Graph representing PE behavior of one of the servers

The MAPE calculations classify the accuracy of predictions from poor to excellent, as per Table 1, with the results being mostly classified as fair to excellent for measurements of CPU, network input, and output, and some predictions classified as weak in write to disk.

By comparing the MAPE classifications, it is possible to validate the predictions through similarity between the results, as shown in the example in Table 2.

MAPE	Forecasting level	Server	MAPE NetIn Average GM(1,1)	MAPE NetIn Average BI	Results
< 10	High	4	5.36	5.25	equivalent
10 - 20	Good	3	14.70	16.60	equivalent
20 - 50	Reasonable	2	45.38		divergent
> 50	Weak	2		65.52	
	Weak	1	159.78	316.64	equivalent

Table 2. Sample of MAPE equivalence comparison

It is observed that BI and GM forecasts are equivalent in more than 66% of all measures for forecasts executed month-to-month, as shown in Figure 11.

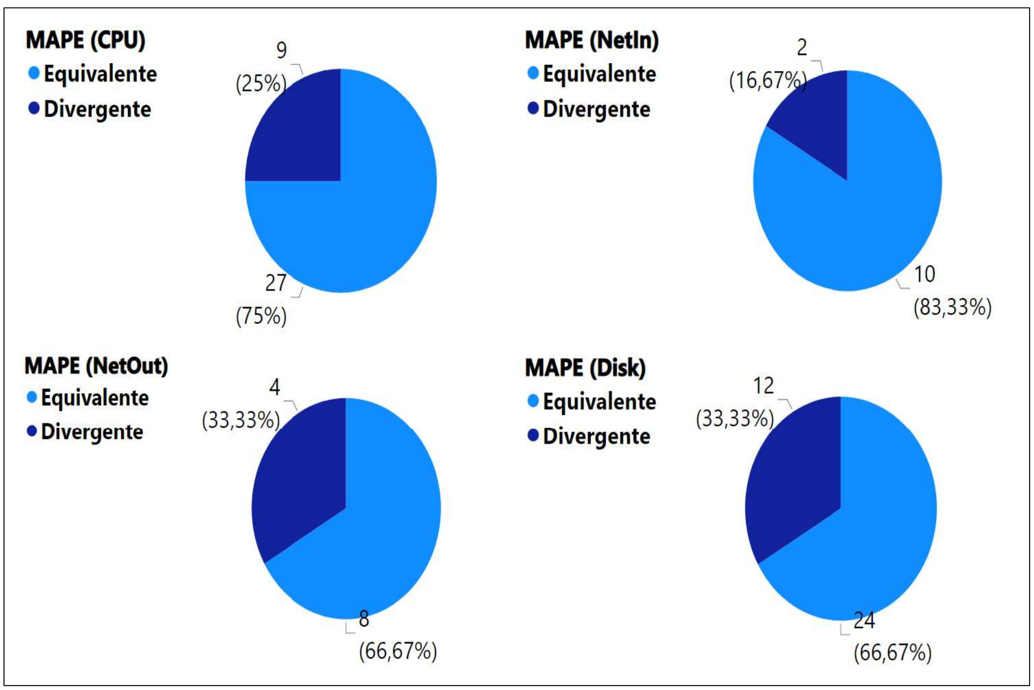

Figure 11. Equivalence between MAPE classifications of GM x Power BI forecasts – Month by month

When it comes to quarterly forecasts, the result is even greater, reaching up to 100% in some of the forecasts, as shown in Figure 12.

This result confirms the validation of the forecasts when using the MAPE method scale.

Despite the MAPE validation confirming the statistical equivalence of the forecasts, it is worth elaborating more about the forecasts. At first, none of the methods could predict the drastic changes that the pandemic caused in the cloud consumption behavior, only after a few months of new behavior (consolidated period) that the prediction starts to show more regular

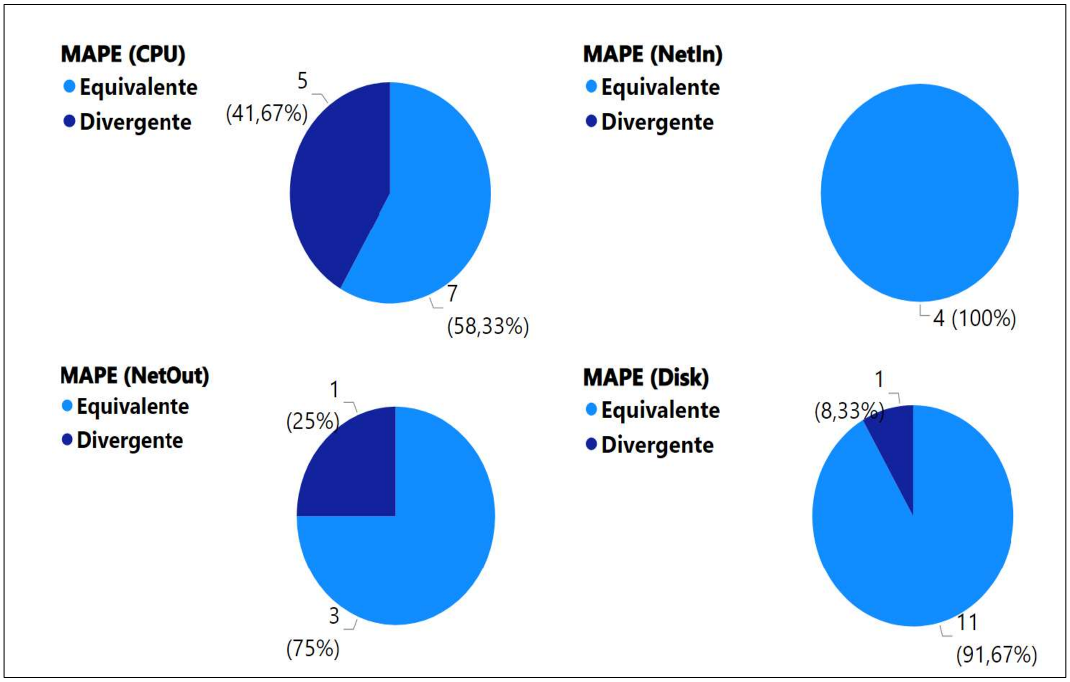

Figure 12. Equivalence between MAPE classifications of GM x Power BI forecasts - quarterly

results. Some factors may indicate why the forecasts of the consolidated period were more similar to the real data, one of them is the bigger size of the data entry, being six months long, basically being double the size of the first forecasts entry. Another factor that should be mentioned is the fact that after the pandemic, the consumption of resources was scaled down, this may have helped the forecasts to be similar not only in scale but also in the behavior of variation spikes, being these very close to the real data.

Still on the forecasts, considering all servers, the GM(1,1) forecasts results for CPU usage were classified as 33% of High Forecast, 17% Good, 33% Reasonable, and 17% weak. For the Power BI forecasts, the classification came to be 28% High, 14% Good, 36% Reasonable, and 22% weak. In the forecasts regarding disk writing, both GM(1,1) and Power BI had 50% classified as weak. The NetIn forecasts were 33% and 42%, classified as High for GM(1,1) and Power BI, respectively. On the other side, the NetOut GM(1,1) forecast presented 42% High, opposed to the 41% weak of Power BI. More in-depth analysis and full results comparison can also be found in the appendix.

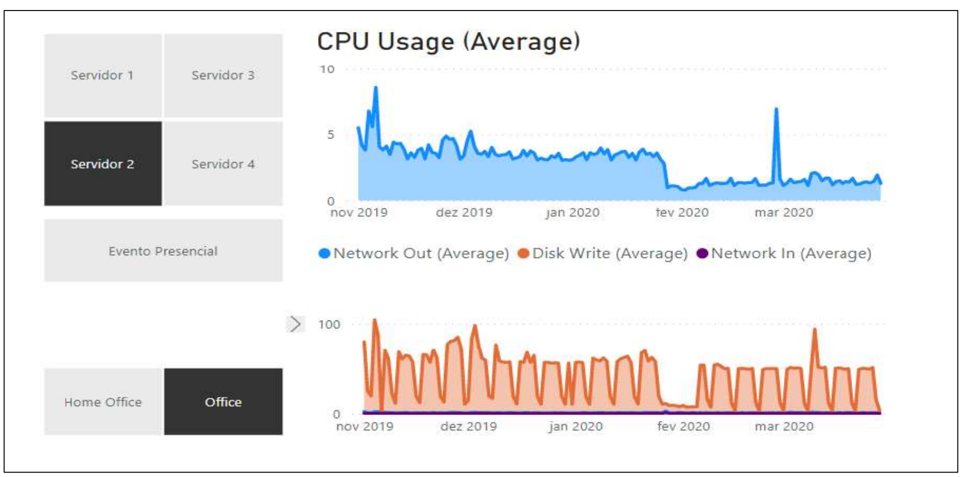

Figure 13. Resource consumption relating server 2 to the working period before the home office policy

Throughout the experiment, the logbook records information related to the SME, such as corporate policies, calendar of events and webinars, data referring to the SME website, among others. By including these data in BI, it is possible to observe which external variables modify or not the behavior of resource usage (CPU, Disk, Network) of the cloud. A comparative analysis is able to demonstrate a link between the period in which the home office policy was introduced in the SME to the detriment of the COVID-19 pandemic, resulting in a drastic drop in resource usage (Disk and Network), bringing the measurement practically to zero, as shown in Figures 13 and 14.

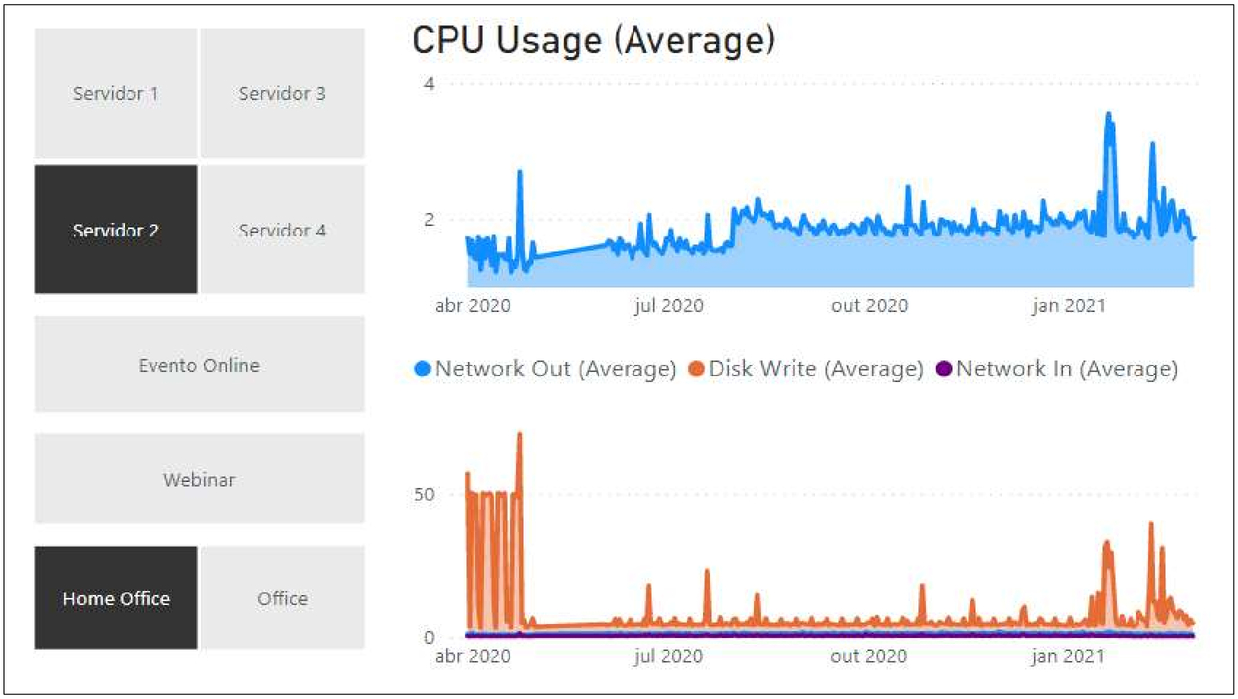

Figure 14. resource consumption relating server 2 to the working period after the home office policy was implemented

Another fact observed is the reason for the regular usage of server resources that serve the external audience of the SME. The data from this server exactly follow the behavior obtained on the website, with page views following some measures of this server, such as writing to disk. Figure 15 demonstrates this similar behavior between measurements.

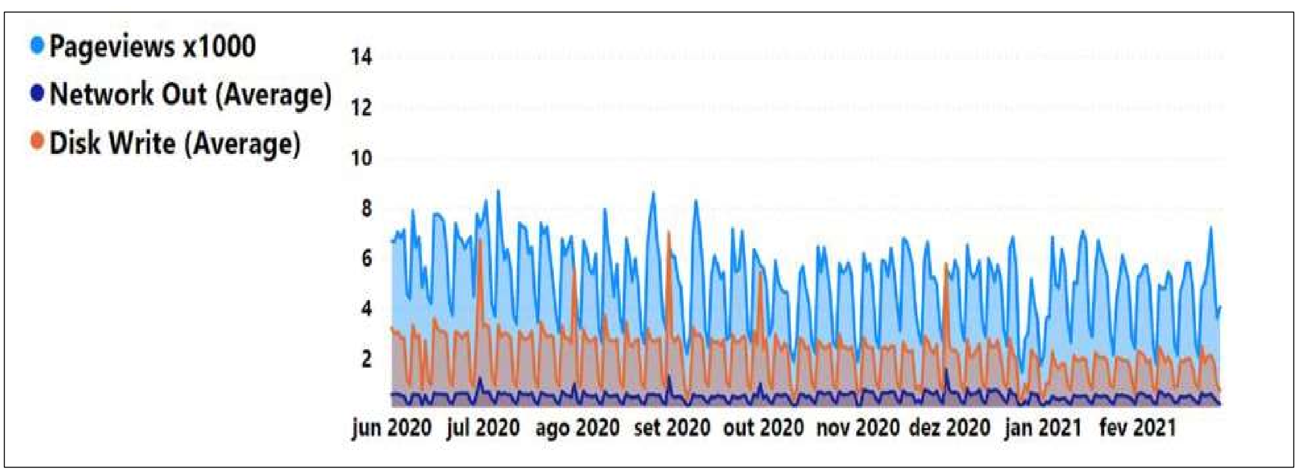

Figure 15. Correlation of the behavior of server 4 with external data between June and February 2021

The results presented demonstrate examples of what can be obtained when relating data from external sources to BI, and it is possible to find relationships between these and the behavior of the cloud environment, and even on specific servers. These relationships provide even more depth to the information that a predictive BI can bring, being a way to obtain relevant information and help decision-making with a predictive DSS.

7. Conclusion

This paper demonstrates an experiment that addresses the use of a BI tool as a predictive DSS to aid decision-making. Cloud resource usage data from an SME were used as the basis for the BI tool.

The estimates provided by BI are validated using the GM(1,1) as a second predictive method, as this method is already established in the academic world. Even though BI already has the local capacity to generate estimates for certain periods, the GM(1,1) belonging to Grey Theory brings a second level of confidence to the predicted scenarios.

The validation is performed by comparing the results of the predictions by the MAPE method, revealing similarities in the classification of the predictions. Over the 15 months, the SME's cloud resource consumption behavior data are collected and used to generate a timeline that enables the relation of some behaviors to external variables. This was possible by using a logbook to record these SME variables.

Despite meeting expectations, the experiment may benefit from some modifications in future works. For example, the use of real-time cloud consumption monitoring, eliminating the need for prior collection. Another point that can be expanded is the use of more predictive methods such as ANN or Fuzzy Theory, even GM(1,1) can be improved to decrease the linearity of predictions. These methods could even be used to further improve BI accuracy.

In the field of relationship with external data, there is still room for integration of other systems to BI, enriching the amount of data and allowing for greater and more relevant discoveries. Finally, this paper provides a sample of the use of a BI tool such as a DSS validated by the GM(1,1) predictive method. An additional contribution includes the application of a predictive method using real data from an SME.

Appendix: The database, results, and scripts are available at https://github.com/marcello-maier/IPT_Masters/.

Acknowledgment: technical support of the Huawei-USP partnership.

References

[1] Widyastuti, D., Irwansyah, M. (2018) Benefits And Challenges Of Cloud Computing Technology Adoption In Small And Medium Enterprises (SMEs) *Proc. 4th BCM 2017*, 41, 241–246 https://doi.org/10.2991/bcm-17.2018.46

[2] Attaran, M., Woods, J. (2018) Cloud computing technology: improving small business performance using the Internet *Journal of Small Business and Entrepreneurship*, 6331 https://doi.org/10.1080/08276331.2018.1466850

[3] Carcary, M., Doherty, E., Conway, G. (2014) The Adoption of Cloud Computing by Irish SMEs – an Exploratory Study *Eletronic Journal Information Systems Evaluation*, 17 (1), 3–15

[4] Power, D. J., Sharda, R. (2007) Model-driven decision support systems: Concepts and research directions. *Decision Support Systems*, 43 (3), 1044–1061 https://doi.org/10.1016/j.dss.2005.05.030

[5] Salleh, N. A., Hussin, H., Suhaimi, M. A., Md Ali, A. (2018). A systematic literature review of cloud computing adoption and impacts among small medium enterprises (SMEs) *Proc. ICT4M 2018*, 278–284 https://doi.org/10.1109/ICT4M.2018.00058

[6] Mell, P., Grance, T. (2011). Definition of Cloud Computing Recommendations of the National Institute of Standards and Technology NIST*, Information Technology Laboratory*, 145, 7 https://doi.org/10.1136/emj.2010.096966

[7] Liu, M., Pan, L., Liu, S. (2019) To Transfer or Not: An Online Cost Optimization Algorithm for Using Two-Tier Storage-as-a-Service Clouds *IEEE Access*, *7*, 94263–94275 https://doi.org/10.1109/ACCESS.2019.2928844.

[8] Chang, V. (2014) The Business Intelligence as a Service in the Cloud *Future Generation Computer Systems*, *37*, 512–534 https://doi.org/https://doi.org/10.1016/j.future.2013.12.028

[9] Dahunsi, F., Idogun, J., Olawumi, A. (2021) Commercial Cloud Services for a Robust Mobile Application Backend Data Storage *Indonesian Journal of Computing, Engineering and Design*, 3 (1), 31–45 https://doi.org/10.35806/ijoced.v3i1.139

[10] Shim, J. P., Warkentin, M., Courtney, J. F., Power, D. J., Sharda, R., & Carlsson, C. (2002) Past, present, and future of decision support technology *Decision Support Systems*, 33 (2), 111–126 https://doi.org/10.1016/S0167-9236(01)00139-7

[11] Asemi, A., Safari, A., Asemi Zavareh, A. (2011) The Role of Management Information System (MIS) and Decision Support System (DSS) for Manager's Decision Making Process *International Journal of Business and Management*, 6 (7), 164–173 https://doi.org/10.5539/ijbm.v6n7p164

[12] Arnott, D., Lizama, F., Song, Y. (2017) Patterns of business intelligence systems use in organizations *Decision Support Systems*, 97, 58–68 https://doi.org/10.1016/j.dss.2017.03.005

[13] Gudfinnsson, K., Strand, M. (2018) Challenges with BI adoption in SMEs. *8th IISA 2017*, *2018-Janua*, 1–6 https://doi.org/10.1109/IISA.2017.8316407

[14] Jasemi, M., Kimiagari, A. M. (2011) Developing a modular portfolio selection model for short-term and longterm market trends and mass psychology *South African Journal of Industrial Engineering*, 22 (1), 67–81

[15] Herrera, M., et al. (2010) Predictive models for forecasting hourly urban water demand *Journal of Hydrology*, *387*(1–2), 141–150 https://doi.org/10.1016/j.jhydrol.2010.04.005

[16] Lee, Y. S., Tong, L. I. (2011) Forecasting energy consumption using a grey model improved by incorporating genetic programming *Energy Conversion and Management*, 52 (1), 147–152 https://doi.org/10.1016/j.enconman.2010.06.053

[17] Dudek, G. (2016) Pattern-Based Local Linear Regression Models for Short-Term Load Forecasting *Electric Power System Research*, 130 (August), 139–147 https://doi.org/http://dx.doi.org/10.1016/j.epsr.2015.09.001

[18] Tsai, S. B. et al. (2017) Models for forecasting growth trends in renewable energy *Renewable and Sustainable Energy Reviews*, 77, 1169–1178 https://doi.org/10.1016/j.rser.2016.06.001

[19] Silva, A. et al. (2017) Energy-efficient node position identification through payoff matrix and variability analysis *Telecommunication Systems*, 65 (3), 459–477 https://doi.org/10.1007/s11235-016-0245-4

[20] Wilson, B. M. R., Khazaei, B., & Hirsch, L. (2017) Cloud adoption decision support for SMEs using Analytical Hierarchy Process (AHP) Proc. 4th *AIEEE 2016*, 10–12 https://doi.org/10.1109/AIEEE.2016.7821809

[21] Salleh, N. A., Hussin, H. (2017) A conceptual model on cloud computing and SME performance: Resource-based view *Proceedings - 6th International Conference on Information and Communication Technology for the Muslim World, ICT4M* 2016, 105–111 https://doi.org/10.1109/ICT4M.2016.032

[22] Guarda, T., Santos, M., Pinto, F., Augusto, M., & Silva, C. (2013) Business Intelligence as a Competitive Advantage for SMEs *International Journal of Trade, Economics and Finance*, *January*, 187–190 https://doi.org/10.7763/ijtef.2013.v4.283

[23] Kung, C. Y., Wen, K. L. (2007) Applying Grey Relational Analysis and Grey Decision-Making to evaluate the relationship between company attributes and its financial performance-A case study of venture capital enterprises in Taiwan *Decision Support Systems*, 43(3), 842–852 https://doi.org/10.1016/j.dss.2006.12.012

[24] Kayacan, E., Ulutas, B., Kaynak, O. (2010) Grey system theory-based models in time series prediction *Expert Systems with Applications*, 37(2), 1784–1789 https://doi.org/10.1016/j.eswa.2009.07.064

[25] G. M. NEGREIROS, J., L. MARQUES, J. A. (2019) Microsoft Power BI Desktop for Data Analytics. *Proc. eedings of the Multidisciplinary Academic Conference*, 132 http://search.ebscohost.com/login.aspx?direct=true&db=edb&AN=137297569&site=eds-live&scope=site

[26] Gowthami, K., Kumar, M. R. P. (2017) Study on Business Intelligence Tools for Enterprise Dashboard Development *International Research Journal of Engineering and Technology*(*IRJET)*, 4 (4) 2987–2992

[27] Armstrong, J. S. (2001) Evaluating Forecasting Methods In *Principles of Forecasting: A Handbook for Researchers and Practitioners* (p 443–472) https://doi.org/10.1007/978-0-306-47630-3_20

[28] Chugh, R., Grandhi, S. (2013) Why Business Intelligence Significance of Business Intelligence Tools and Integrating BI Governance with Corporate Governance *International Journal of E-Entrepreneurship and Innovation*, 4 (2) 1–14. https://doi.org/10.4018/ijeei.2013040101