Predicting the Future is Like Completing a Painting: Towards a Novel Method for Time-Series Forecasting

NADIR MAAROUFI[1], **MEHDI NAJIB**[1], **AND MOHAMED BAKHOUYA**[2]

[1]TICLab, College of Engineering and Architecture, International University of Rabat, Sala Al Jadida 11000, Morocco
[2]LERMA Lab, College of Engineering and Architecture, International University of Rabat, Sala Al Jadida 11000, Morocco

Corresponding author: Mohamed Bakhouya (mohamed.bakhouya@uir.ac.ma)

ABSTRACT This article is an introductory work towards a larger research framework relative to *Scientific Prediction*. It is a mixed between science and philosophy of science, therefore we can talk about *Experimental Philosophy of Science*. As a first result, we introduce a new forecasting method based on image completion, named *Forecasting Method by Image Inpainting (FM2I)*. In fact, time series forecasting is transformed into fully images- and signal-based processing procedures. After transforming a time series data into its corresponding image, the problem of *data forecasting* becomes essentially a problem of *image inpainting*, i.e., completing missing data in the image. Extensive experimental evaluation was conducted using the shortest series of the dataset proposed by the well-known M3-competition. Results show that FM2I, despite still being in its infancy, represents an efficient and robust tool for short-term time series forecasting. It has achieved prominent results in terms of accuracy and outperforms the best M3 methods. We have also investigated the effectiveness of the FM2I against the Smyl method, the winner of the M4 competition. Using the same category of shortest series, results show a close accuracy compared to the Smyl method. The FM2I is also able to generate ensemble data forecasts, which contain the best and more accurate forecast compared to existing and considered methods.

INDEX TERMS Scientific prediction and experimental philosophy of science, extensive structural realism, bridging philosophy, time series forecasting, fully integrated modeling and processing framework, ensemble data autocorrelation forecasting, augmented dimension prediction, image and signal processing.

I. INTRODUCTION

It is well known that good forecasts are of utmost importance in all scientific disciplines [1]. However, the relevance of accurate forecasts has become increasingly palpable since the emergence of Big-Data as a new paradigm of scientific activities for many researchers [2]. One of the suggested tracks was due to the distinguished statistician Leo Breiman when he wrote with reference to predictions: "*If our goal as a field is to use data to solve problems, then we need to move away from exclusive dependence on data models and adopt a more diverse set of tools.*" [3]. Therefore, we believe that understanding the *nature of prediction* can help creating novel forecasting methods. First of all, let us precise that forecasting is considered as a process of predicting the future according to past and present data. We emphasize that this terminology may vary according to each scientific discipline (see [4]).

This paper is not mainly concerned by *philosophy of science*, but we want to share with the research community the methodology we have undertaken, starting by deeper philosophical considerations, with the aim is to elaborate a new forecasting method. More precisely, we first start by raising philosophical questions, which are related to the nature of prediction itself, and then translate these questions following a scientific methodology for being applied for real-sitting forecasting scenarios. We name this way of thinking the *Experimental Philosophy of Science*. Thus, we will switch, in this article, between science and philosophy of science as much as required in order to better describe our approach and highlight obtained results.

As a first result, we introduce a new short-term time series forecasting method based on *image completion*, we name it *Forecasting Method by Image Inpainting (FM2I)*.

The associate editor coordinating the review of this manuscript and approving it for publication was Geng-Ming Jiang.

119918

VOLUME 9, 2021

It represents a performing and robust tool for short-term Time Series (TS) forecasting. Basically, the main idea behind the FM2I method is as follows. We first transform a TS into an image and then complete it using adapted image inpainting techniques. The completed image is converted back to the TS in order to obtain forecast values. Our aim is to perform these tasks without usual pre-processing (e.g., seasonality, trends, stationarity) of the original TS. The process of transforming TS into images and from the inpainted images into the original TS, including forecast values, must be possible.

In summary, the main general contributions of this work are three folds: *i)* interlinking philosophy of science and science, named *experimental philosophy of science*, *ii)* an extensivity testing framework and *iii)* a novel forecasting method by image inpainting. The remainder of this paper is structured as follows. Section 2 presents the formulation of the scientific prediction and philosophical problem. The scientific transformation of the problem is introduced in Section 3. In Section 4, materials and methods are described. The proposed FM2I algorithm is detailed in Section 5. Experimental results and comparison are described in Section 6. Scientific and philosophical discussions together with potential directions are given at the end of this document. Section 7 provides an overview of existing approaches, from literature, for time series classification and regression. Concluding remarks and perspectives are given in Section 8. An appendix is included at the end in order to describe some notions and fundamentals related to signal theory.

II. SCIENTIFIC PREDICTION AND PHILOSOPHICAL PROBLEM FORMULATION

It is well known in the philosophy of science field that *Scientific Explanation* and *Scientific Prediction* are one of the main goals of sciences [5], [6] and [7]. Nevertheless, philosophers of science have focused much more on developing a theory of explanation rather than introducing a prediction theory [8], [9] and [10]. Indeed, several models of explanation were arised during the past half century. The pioneering and most famous models of explanation, so-called covering-law due to Carl Gustav Hempel and Paul Oppenheim, are DN (Deductive Nomological), DS (Deductive Statistical) and IS (Inductive Statistical) [11], [12]. According to their models, we get a scientific explanation of an event/phenomenon E, which occurred (called *Explanandum*) when E can be deduced, induced or inferenced from the *Explanans* (I, L), i.e., a set of initial conditions I and a set of laws L. In other words, the explanandum E must be a logical consequence of the explanans (I, L), which meet certain conditions [11]. There are other subsequent alternative models of explaination; we can mention, for instance, the Causal [13] or Unificationist [14] model.

However, despite the extensive work in this area, still a logical prediction model is required to formalize the structure of predictive reasoning, see [9] and [10]. One of the consequences is the weakening of the explanation theory itself [5]. In order to overcome the difficulty of constructing a theory

of prediction, early renowned works of Hempel and Oppenheim [11] and Hempel [12] introduced the controversial ([15] and [16]) *Structural Identity Thesis* (also called *Symmetry Thesis*) between explanation and prediction. The structural identity thesis assumes that explanation and prediction have the same logical form and are, somehow, equivalent. In other words, every adequate explanation is a potential prediction and every adequate prediction is a potential explanation. In particular, as it was recalled in [12], in the Hempel-Oppenheim model, predictions concern exclusively phenomena that occur in the future and, therefore, it seems that the difference between explanation and prediction is purely chronological or temporal. This structural identity thesis has been heavily criticized by many philosophers [15]–[17].

Nowadays, there is no clear consensus concerning the definition of scientific prediction, see [9] and [10]. Most particularly, the notion of scientific prediction itself is still ambiguous. Finding, however, an adequate and unified definition of prediction is challenging. It must be quite restrictive in order to exclude *Divination*, since it is based on occultic and metaphysics process of predicting, for instance as in astrology. Moreover, it must be fairly too broad by including all forms arising from different scientific disciplines. Indeed, according to the definition of Hempel-Oppenheim's model, the evolutionary biology, palaeontology or the geology are automatically excluded, since these disciplines deal with the past, which means that they are exclusively explicative sciences. Furthermore, their model also excludes any prediction not obtained through scientific laws (see a counter-example in [18]).

Actually, in current literature, a better understanding of the notion of scientific prediction appears in two major areas of the philosophy of science, *Confirmation Theory* and *Scientific Realism* debate (see [10] and [19]). Briefly, the two questions asked in these two fields are respectively: should a hypothesis or a theory be better confirmed when it generates accurate predictions [20]? Are accurate predictions a reasonable criterion to assess a theory's maturity in order to affirm the reality of its assumed structures [21]? These two philosophical domains reach the same problem related to scientific prediction and the need to redefine this fundamental notion is required. The article [22] was the pioneering work, which explicitly makes the connection between these two philosophical areas and provides the current form of the issue [19]. We wish to emphasize that [22] marks a certain breakthrough from the existing literature since it addresses a realistic predictive process from applied mathematics. Indeed, authors in [22] analyze in a concrete way how existing regression analysis, from statistics theory, proceeds to predict, in order to make a powerful contribution to the philosophical debate.

In this work, we want to take the research a step further by philosophically formulating one of our questions concerning the nature of predictions. We will turn it into a scientific issue, and then formalize our own methodology, test it and then contribute to engaging both philosophical and scientific debate related to the prediction. That is what

we mean by *experimental philosophy of science*. We must stress that unlike the above-mentioned confirmation theory and scientific realism, our investigation is motivated by the prediction itself and not serving primarily another theory. Since, we are not yet be able to formalize a definition of prediction, we need, however, a framework to better clarify our approach. Thus, we consider the definition given in [23] or its extended version in [10]. More precisely, the prediction is a result of a *predictive process*. This later is considered as a series of inferences allowing setting, without measures or observations, the value of one or several variables (or the relationships between variables), variables are used here in a broad sense. However, this definition remains barely ambiguous since it is difficult to characterize a *scientific predictive process* [2].

Thus, we start by wondering the simple following question, which we call the *nature of predictions question*. Using the definition above, predictions are a result of a predictive process, then is this last possess a signature related to the own nature of the phenomenon, the event or the type of data we want to predict? In other words, if we dispose or establish a predictive process in a particular framework, can this process be potentially a predictive process in a completely different setting? More precisely, assume that we have an accurate predictive process arising from a theory or a model, concerning a specific scientific field, and able to predict a phenomenon *A*, is this process capable of predicting, in another scientific field, a phenomenon *B*? This question may seem to be trivial for theories, which are specifically concerned by prediction. For example, the statistical theory is able to predict a variety of different phenomena, since we can model them according to known tools or classes of statistics-based models. In our research program, since the aim is to exhibit new prediction methods, we want to go further by investigating "unthinkable" links between phenomena.

After careful considerations, it appeared suitable to reformulate this question in relationship with the notion of *Predictive Capacity*. First of all, we want to precise the difference between the latter and *Predictive Power*, see [9] for more details. The predictive power refers to the ability to derive, from a theory, at least one testable prediction. It can be obtained *a priori* by ensuring that the theory might have empirical consequences. While the predictive capacity can be used to evaluate the variety and the accuracy of the predictions generated by a theory, there are no ways, however, to assess it *a priori* since the validity of these predictions have to be checked after their future realizations (or not) and that in relationship with the confirmation theory, as stated by [9], [19]. The same author gives two dimensions or two epistemic virtues of predictive capacity: the *Intensivity*, which refers to the accuracy of predictions with regard to a particular phenomenon, while the *Extensivity* refers to the ability to predict in a large variety of phenomena. Accordingly, we can formulate the following question: how can we test the extensivity of a theory's or a model's predictive capacity?
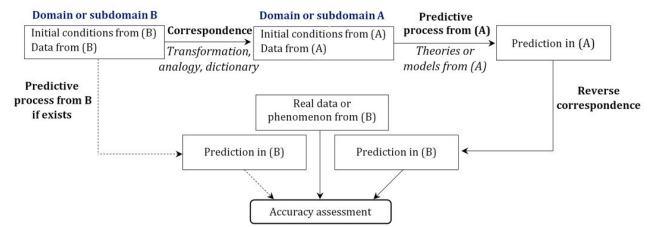

FIGURE 1. ETF: extensivity testing framework.

III. SCIENTIFIC TRANSFORMATION OF THE PROBLEM

As described in the previous section, the problem is formulated in a very general way, but the power of this type of formulation is to generate potential research directions in the field of scientific prediction. More precisely, throughout an adequate *correspondence* procedure, the purpose is to test the extensivity of a theory's or a model's predictive capacity from a specific scientific setting or domain *A*. Thus, we have to choose a predictive process, which is designed, established and used in *A*, then test it in the frame of another specific domain *B*. This *correspondence* must enable switching, going back and forth, between the two domains or sub-domains. The results of this procedure are assessed in terms of accuracy according to real and/or existing predictions, which are, eventually, obtained by a predictive process from *B*. This procedure, named *extensivity testing framework (ETF)*, can be formalized as depicted in Fig. 1.

While this extensivity testing framework could certainly bring a first answer to the previous question, it will not obviously tackle the whole above-mentioned philosophical problem. Nevertheless, getting a conclusive test, means that we have found a new, unknown and "unthinkable" prediction method and thus reaching our objective announced in the introduction. Therefore, it will be crucial to narrow the scope by only focusing, for instance, on the prediction by *Extrapolation/Interpolation*. So, let us recall that in practice, extrapolation or interpolation are operations, which are mainly used for predicting the values of one or several variables, denoted *Y*, according to known values of one or several variables, denoted *X*. The unique difference between the extrapolation and interpolation is as follows: the latter is concerned by the variables, which are situated inside the experimental or the studied sampling area, while the former is related to those that are outside this area [9]. However, since the coming trend is not known, the extrapolation seems a difficult task than the interpolation. But, both the extrapolation and interpolation can be clearly considered as predictive processes according to the definition given in the previous section.

After careful considerations, we have focused on the extrapolation by testing the predictive capacity of *Image Completion* for performing *Short-term Time Series Forecasting*. Basically, an image is a distributed amplitude of colors [24], it is a two-dimensional spatial pixels organization. Furthermore, an image possesses a specific structure [25], which is constituted, for example, of edges and texels; while a temporal series is a one-dimensional temporal sequence

of successive values. As a matter of fact, there is a significant difference between these two data structures. Moreover, the image-based predictive process arises from image restoration techniques, which are designed for spatial patterns analysis and governed by *Visual Logic* assessment as well, such as image quality [26] and *Psychovisual* quality metrics [27]. Thus, there is no reason to expect that this image-based predictive process could work for temporal series forecasting, since it should be based on a *Chronological Logic* by analysing past/present patterns. In other words, finding a way to draw past/present events and completing the painting, obtained by a purely pictorial reasoning, allows to 'see' the future!

Following the above-mentioned considerations, we have reached the required ingredients that allow developing our scientific and technical framework, which is considered as an instantiation of the proposed extensivity testing framework. It is basically a fully integrated modeling and processing framework for short term time-series forecasting, named FM2I, as depicted in Fig. 2.
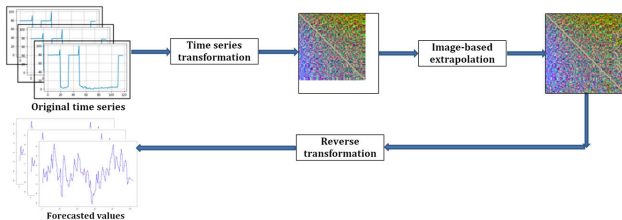

FIGURE 2. A schematic view of the proposed fully integrated framework FM2I.

IV. MATERIALS AND METHODS

The methodology presented in this section covers the entire process, from time series transformations into 2D image structure to image-based forecasting (inpainting), back to the 1D sequence representing the original times series including the forecast data. It's worth noting that time series can be either univariate or multivariate. In this work, we focus only on univariate time series. As shown in Fig. 2, the FM2I approach is composed of three main processing phases: transformation of 1D sequence representing time series into two 2D image-based representation, image extrapolation by adapting inpainting methods, and reverse data techniques to transform a 2D image-based into 1D sequence to further represent the original time series together with forecast data.

Regarding the first phase, the main question is how to transform 1D TS to 2D image-based structures by enhancing the initial information? We have investigated three main increasingly enhanced approaches. The first basic one, named *naive approach*, considers the plot of the TS as an input image to the adapted inpainting method. In the second one, named *enhanced approach*, we transform directly the TS into an augmented TS based-matrix using signal processing principles and tools. Based on this enhanced matrix, more adequate transformations are proposed in the deeper and final approach.

The second phase is dedicated to image-based forecasting using adapted image inpainting methods. It is worth noting that these methods have been mainly proposed for filling in part of an image (restoration of damaged regions). Existing inpainting methods can be classified into two main categories [28]: *diffusion-based* approaches and *examplar-based* approaches. Diffusion-based approaches have been mainly inspired by techniques from computational fluid dynamics. The pioneer work in this area was performed by Bertalmio *et al.* in [29] and [30] by using Navier-Strokes equation. Authors built an approach, at pixel level, by performing an analogy between the stream function in a 2D incompressible fluid and image intensity. As stated in [28], these approaches have provided excellent results for small inpainted regions, but they tend to introduce smooth effects in large regions. Exemplar-based approaches operate, however, at patch level by propagating information from known regions into missing ones. Unlike diffusion-based approaches, exemplar-based approaches are more adapted when inpainting large missing regions [28].

In this work, we have adapted the exemplar-based approach, in particular, the patch-based method, which is basically proposed for completing missing (or deteriorated) parts of images. In fact, for TS forecasting, after transforming the TS into an image, the missing parts (i.e., unknown region) are considered as our forecast area. More precisely, the patch-based method is diverted from its basic function and then used for completing the area, which is related to the forecast TS values. It is worth noting that we are developing a modified Navier-Stokes based method. However, in this work, we choose to focus only on the adapted patch-based method.

The rest of this section presents the three abovementioned TS transformation techniques (i.e., Naive, Enhanced, and Deeper) together with issues we have tackled in order to maximize the transformation precision while considering a back and forth *TS-image correspondence*.

A. NAIVE APPROACH

In this approach, we have considered the plotted TS as a 2D image (original image), as depicted in Fig. 3. The plotted TS represents CO_2 values from the dataset described in [31]. As shown in this figure, the forecast area (extrapolated zone) is added to the original image as a deteriorated part for being inpainted by the patch-based method. The latter considers the extrapolated area as a regular hole to be reconstructed or inpainted, i.e., filling the missing pixels. The adapted patch-based method is able to fill on it; however, it generates several forecasts with high errors compared to the original TS (i.e., real plot). In fact, it is difficult, even impossible, for the patch-based method to infer the forecast area. It is due to the fact that the plotted TS 2D image holds less rich information. Indeed, there are much more white pixels than those representing TS values. Moreover, we have noticed, from the obtained experimental results, that the patch-based method does not deal well with such representations. This constraint compelled us to
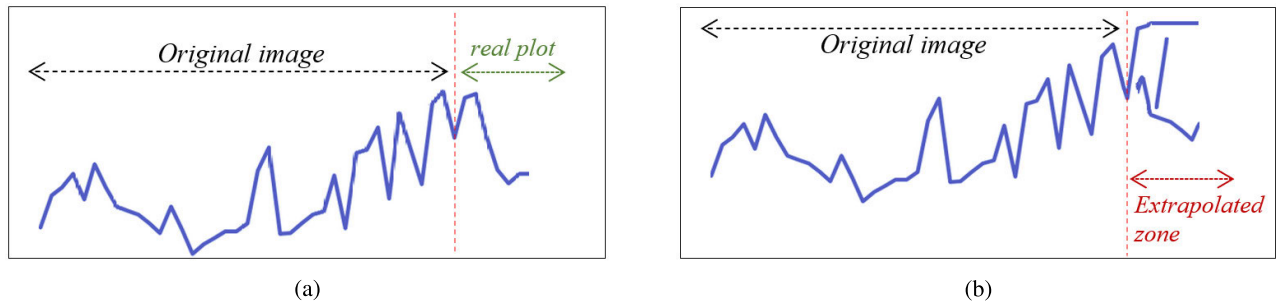
FIGURE 3. Actual TS with a) real data, b) predicted data.

find suitable and richer 2D representations of TS, as described in the next section. Our aim is to extend the original TS with richer information, which could lead to more image patterns' structures and features.

B. ENHANCED APPROACH

The aim of this enhanced approach is to generate the TS's corresponding matrix that could increase the information gain against such 1D TS representation. Basically, a time series TS, having a set of values indexed in a time order, can be represented by $TS = \{s_0, s_1, s_2, s_3, \ldots s_n\} = (s_i)_{i=0,1,\ldots n}$, where s_0 is the first value obtained at $t = 0$ while s_n represents the value obtained at $t = n$. We have investigated existing techniques, from the signal processing theory, in order to transform a 1D signal into a 2D matrix. More precisely, we have used the *Time Autocorrelation Function* as a tool for 1D-to-2D transformations. Let us consider a TS as a *random signal with a finite average power*, the aim is to represent the majority of real-life time series. We can consider a TS as a *digital signal* resulting from the sampling of a random and permanent analog bounded signal $x(t)$ having a finite average power. In fact, let us consider $x(t)$ a signal having a finite average power, we can then define the temporal autocorrelation of this signal by the following formula, but more details are presented in the Appendix A:

$$\Gamma(\tau) = \lim_{T \to +\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t)x(t-\tau)dt. \quad (1)$$

It is worth noting that the function $\Gamma(\tau)$ is well defined. Indeed, by applying the Cauchy-Schwarz inequality taking into consideration that $x(t)$ is bounded, we can obtain $\Gamma(\tau) \leq P(x(t)) < +\infty$. Actually, the maximum of the function Γ is found at the origin, in fact, we have $\Gamma(0) = P(x(t))$.

This autocorrelation function measures the degree of similarity between the signal and its offset version, so it is natural that this function is maximum for $\tau = 0$. $\Gamma(\tau)$ compares the signal to itself, it is, therefore, a temporal average of the product of the signal by itself shifted by a time τ. Moreover, $\Gamma(\tau)$ measures the similarity degree or internal signal's dependency. Mathematically, this function can be seen in several ways; namely, as a scalar product between the signal and its offset. Thus, if $\Gamma(\tau) = 0$, it means that the signal and its shift of one step τ are completely uncorrelated, they

are two orthogonal signals. We can also see Γ as a sort of convolution of the signal with itself. This autocorrelation function does not only contain important information about the signal, but also its Fourier transform and then its spectrum. Indeed, this relationship manifests itself through the power spectral density (PSD). Thus, this autocorrelation function seems to contain valuable information regarding the properties of the signal $x(t)$. More details regarding the PSD relationship with the auto-correlation function are presented in the Appendix B.

For a digital signal $(x_n)_{n \in \mathbb{Z}}$, the discrete version of $\Gamma(\tau)$ can be written as follows:

$$\gamma(n) = \lim_{M \to +\infty} \frac{1}{M} \sum_{k=-\frac{M-1}{2}}^{\frac{M-1}{2}} x_k x_{k-n}. \quad (2)$$

So, the autocorrelation formula for a TS, for $i \in [0, 1, 2 \ldots n]$, is $\gamma(i) = \frac{1}{n+1} \sum_{k=i}^{n} s_k s_{k-i}$. This allows defining a symmetric temporal autocorrelation matrix in the following form:

$$
\gamma_{(i \in [0,1,\ldots n])}
$$
$$
= \begin{pmatrix}
\gamma(0) & \gamma(1) & \gamma(2) & \cdots & \cdots & \gamma(n) \\
\gamma(1) & \gamma(0) & \gamma(1) & \gamma(2) & \cdots & \gamma(n-1) \\
\gamma(2) & \gamma(1) & \ddots & \ddots & \vdots & \cdots \\
\vdots & \gamma(2) & \gamma(1) & \ddots & \ddots & \vdots \\
\vdots & \vdots & \gamma(2) & \ddots & \ddots & \gamma(1) \\
\gamma(n) & \gamma(n-1) & \cdots & \cdots & \gamma(1) & \gamma(0)
\end{pmatrix}
$$

It is worth noting that, it is not only the symmetric temporal autocorrelation matrix (STAM) that can be generated, but also the STAM version by excluding its average. Indeed, any signal $x(t)$ can be written according to the form $x(t) = \bar{x} + y(t)$, where $\bar{x}$ is a constant, it is the mean value of the signal $x(t)$, and $y(t)$ is a signal of zero mean value. So, we can get the relation $\frac{dx(t)}{dt} = \frac{dy(t)}{dt}$. This will, therefore, allow using the STAM of the signal $y(t)$ with zero mean. More specifically, for a TS, differentiating it by doing $s_{i+1} - s_i$, provides a new TS, where the mean is eliminated before generating the STAM. Therefore, it will be easy to return back to the original TS afterwards, including the forecast area. More details about particular signals, which are stationary and ergodic are given in Appendix C.
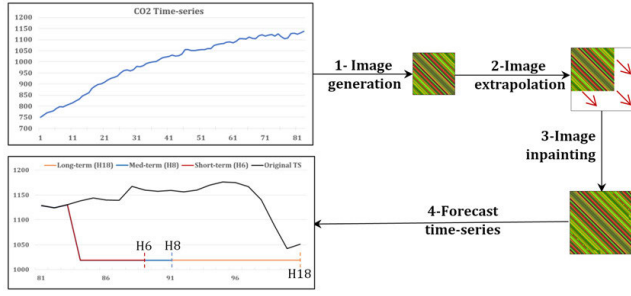
FIGURE 4. Forecasting using the auto-correlation matrix.

The auto-correlation function, which is used to generate the STAM, is used, in turn, for TS forecasting. In fact, this latter is applied to the original TS in order to generate its corresponding matrix, as depicted in Fig. 4 (image generation). Then, the image, including the area to be predicted, is generated (delimited forecast area generation). The size of this area corresponds to the size of the forecast horizon. The adapted patch-based approach is then applied on this latter to extrapolate the forecast area (image inpainting). The final phase is the backtracking from the original image, including the forecast area (i.e., short, medium and long horizon), to its corresponding TS (forecast time-series). All these transformation techniques must take into consideration two major objective functions, maximizing the transformation precision (i.e., information loss induced by the transformation), minimizing the execution time, while considering a back and forth *TS-image correspondence*. These techniques are described in Section 5.

In order to assess the effectiveness of this enhanced method, we have used the most common metrics, the MSE (Mean Squared Error), RMSE (Root Mean Square Error), MAE (Mean Absolute Error), the MAPE (Mean Absolute Percentage Error) and sMAPE (symmetric Mean Absolute Percentage Error). The MSE is used to assess the quality of the estimation by computing the average squared difference between actual and estimated values. It is at all times positive, but when closer to zero the estimation is considered of high quality. Alike MSE, RMSE is used to measure the deviation of predicted values from actual values. It is computed as square root of average square deviation. The MAE is used to measure the average magnitude of the errors between forecast values and their corresponding actual ones. It is worth noting that the lower values of RMSE and MAE are better, i.e., the smaller an RMSE/MAE value is observed more the predicted values are closer. The MAPE measures the average of mean absolute percent errors, while sMAPE measures the average of symmetric mean absolute percent errors (%). The metrics are formally described as follows: $MSE = \frac{1}{n}\sum_{i=1}^{n}(F_i - A_i)^2$, $RMSE = \sqrt{\sum_{i=1}^{n}(F_i - A_i)^2}$, $MAE = \frac{\sum_{i=1}^{n}|F_i - A_i|}{n}$, $MAPE = \frac{1}{n}\sum_{i=1}^{n}|\frac{F_i - A_i}{A_i}|$, and $SMAPE = \frac{2}{n}\sum_{i=1}^{n}\frac{|F_i - A_i|}{|F_i| + |A_i|}$, where n is the forecast horizon (n-step-ahead), F_i are forecast values, which are produced at time i by the established forecasting algorithm, whereas A_i is the value, which is actually observed at time i.

TABLE 1. Performance metrics for short, medium and long horizon forecasting.

Metrics	Short-Horizon	Med-Horizon	Long-Horizon
MSE	16864.50	17546.19	16375.84
RMSE	129.86	132.46	127.96
MAE	129.37	132.02	122.05
MAPE	11.26%	11.47%	10.59%
sMAPE	11.94%	12.17%	11.24%

Table 1 shows the MSE, RMSE, MAE, MAPE and the sMAPE for short (6-step-ahead), medium (8-step-ahead) and long (18-step-ahead) forecasting. The obtained results show similar patterns, like the naive approach, but with high errors for all horizons, short, medium and long. More precisely, this enhanced method is able to generate only constant forecast values, which is an obvious behaviour since they show homogeneous patterns (i.e., diagonal lines form). In fact, the adapted patch-based method performs well by forecasting data, but the auto-correlation matrix is not suitable in this case. Moreover, despite the fact that we have used one of the best patch-based methods, its behaviour is mainly related to the image's patterns, which are related to the relevance of the selected TS-image transformations. This boils down to the fact that the information is too concentrated and needs to be broken down. The next section will introduce deeper transformation techniques including new representations of original TS.

C. DEEPER APPROACH

We have first investigated a new matrix representation, with the aim to enrich data through self-signal autocorrelation, called modified auto-correlation (MAC), by combining all possible offsets. This matrix can be performed by transforming the auto-correlation matrix into another one (MAC), for instance, by separating the sums, which are seem represented on the diagonal and the sub diagonals. We then obtain the matrix, depicted in Fig. 5a, in which the sum of the diagonal and each diagonal makes the $\gamma(i) \times (N + 1)$. This structure leads to a gramian matrix (GM).

It is worth noting that in [32] authors have reached similar results (Gramian Angular Field, GAF), by encoding TS (1D cartesian coordinates) as images (2D polar coordinates). In fact, images are represented by GAF in which each element is computed by the trigonometric sum between different time intervals. Furthermore, they have presented other techniques to allow transforming 1D sequence time series data into image-based representation for TS classification using traditional machine learning approaches. In fact, after transforming time series data into images, conventional neuronal approaches are used for classification purposes, e.g., learning features and identifying structure in TS. More precisely, instead of using TS data as input, these approaches use its corresponding encoded images (e.g., traditional image recognition). For instance, we can cite the gramian angular summation field (GASF), gramian angular differed field (GADF), markov transition field (MTF). Similarly, authors in [33] present another technique, called relative position

$$\begin{pmatrix} x_0.x_0 & \cdots & \cdots & x_0.x_n \\ x_1.x_0 & \ddots & \cdots & x_1.x_n \\ \vdots & \cdots & \ddots & \vdots \\ x_n.x_0 & \cdots & \cdots & x_n.x_n \end{pmatrix}$$

(a) MAC

$$\begin{pmatrix} \cos(\theta_0 + \theta_0) & \cdots & \cdots & \cos(\theta_0 + \theta_n) \\ \cos(\theta_1 + \theta_0) & \ddots & \cdots & \cos(\theta_1 + \theta_n) \\ \vdots & \cdots & \ddots & \vdots \\ \cos(\theta_n + \theta_0) & \cdots & \cdots & \cos(\theta_n + \theta_n) \end{pmatrix}$$

(b) GASF

$$\begin{pmatrix} x_0 - x_0 & \cdots & \cdots & x_0 - x_n \\ x_1 - x_0 & \ddots & \cdots & x_1 - x_n \\ \vdots & \cdots & \ddots & \vdots \\ x_n - x_0 & \cdots & \cdots & x_n - x_n \end{pmatrix}$$

(c) RPM

$$\begin{pmatrix} cosg(\theta_0, \theta_0) & \cdots & \cdots & cosg(\theta_0, \theta_n) \\ cosg(\theta_1, \theta_0) & \ddots & \cdots & cosg(\theta_1, \theta_n) \\ \vdots & \cdots & \ddots & \vdots \\ cosg(\theta_n, \theta_0) & \cdots & \cdots & cosg(\theta_n, \theta_n) \end{pmatrix}$$

(d) GC

FIGURE 5. Matrix representations of TS.

matrix (RPM), for transforming 1D TS into image-based representations (see Fig. 5c). They mainly combined the RPM and CNN (Convolutional Neural Network) in an unified framework in order to enhance the accuracy of the time series classification. Reported results show the effectiveness of the proposed approaches against the state-of-art classification techniques.

To the best of our knowledge, all research works, which use image-based representations of TS, are dedicated to classification. However, only the work presented in [34], [35] dealt with time series forecasting. However, in [34], [35] authors used the same principle adapted for classification purposes, i.e., from 1D vector into 2D representation, in the form of image, to data classification using conventional neuronal approaches. Furthermore, they have used an image-based representation, which is similar to the above-mentioned naive representation. For instance, as described in [34], the TS (i.e., the daily load data) is transformed into a graphical representation instead of a matrix-based representation. A dual-branch deep convolutional network is then fed by the generated graphical (i.e. plotted) representation in order to extract features for clustering purposes.

After a detailed study of the above-mentioned representations, we have selected GASF and RPM, which allow easily a return back to the 1D (i.e., 1D-2D and 2D-1D) representation. For the GASF transformation matrix, authors, in [32], presented the transformation as a transition to polar coordinates, assuming that $\theta_0 = \arccos(x_0), \theta_1 = \arccos(x_1), \theta_2 = \arccos(x_2) \ldots \ldots \theta_n = \arccos(x_n)$. More specifically, if $\theta_0 = \arccos(x_0), \theta_1 = \arccos(x_1), \theta_2 = \arccos(x) \ldots \ldots \theta_n = \arccos(x_n)$, the matrix can be written as depicted in Fig. 5b. According to the trigonometry formulas, the following expression, $\cos(\theta_i + \theta_j) = x_i.x_j - \sqrt{(1 - x_i^2)}\sqrt{1 - x_j^2}$, can be easily obtained. Unlike MAC representation, the GASF uses a modified scalar product representation.

Alike the GASF representation, we have introduced another matrix, named Generalized cosine (GC). After rescaling the TS values to be within unified interval (e.g., $[-1, 1]$), if we denote by $cosg$ the generalized cosine function, as introduced in [36], the two-variables function is defined by $cosg(\theta_i, \theta_j) = \frac{\sin(\theta_i)}{\sin(\theta_i + \theta_j)}$, and therefore, the corresponding GC matrix can be generated, as depicted in Fig. 5d. Using the formulas of trigonometry, we can get: $cosg(\theta_i, \theta_j) = \frac{\sqrt{(1 - x_i^2)}}{x_i\sqrt{1 - x_j^2} + x_j\sqrt{1 - x_i^2}}$, and $cosg(\theta_i, \theta_i) = \frac{1}{2x_i}$. As shown in Fig. 5d, the matrix is not symmetric. Therefore, in order to reinforce TS representation patterns and features, we have proposed two symmetric forms of the basic GC, GCS1 and GCS2 as follows: $GSS1 = \frac{1}{2}(GS +^t GS)$, where ^tA is the transpose of matrix A and $GSS2(i, j) = GS(j, i)$ for $i < j$.

In order to evaluate the efficiency of these matrix representations for TS forecasting, we have computed their accuracy using a set of TS (3003 time series) from M3 competition [37], [38]. Fig.6 represents an example of a given TS using the above-mentioned image-based representations. Extensive experiments have been conducted in order to compute the MSE, RMSE, MAE, MAPE and the sMAPE of selected TS (i.e., TS1, TS3, TS4, TS11, TS20, TS24). Obtained results are reported in Table 2 to show whether these matrices affect the forecasting accuracy. As shown in this table, none of them outperforms for all considered TS.

V. FM2I ALGORITHM

The FM2I algorithm is structured into four main phases, as depicted in Algorithm 1. The *first phase* includes all steps that are required for preparing TS. The first step splits TS into two parts, training (i.e., parameters tuning) and testing. The testing part represents the forecasting values according to the horizon's sizes. It is mainly used once the best forecasting model is specified. It is worth noting that since the above-mentioned matrix transformations can be considered as a
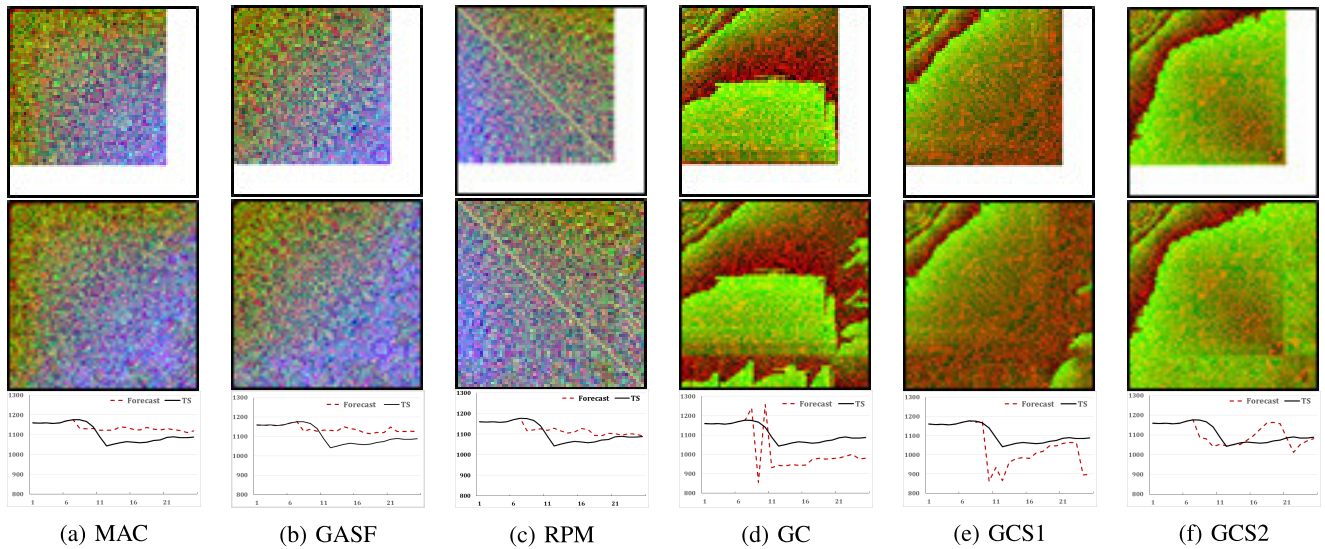
(a) MAC   (b) GASF   (c) RPM   (d) GC   (e) GCS1   (f) GCS2

FIGURE 6. Image-based representation: without and with forecast area/values.

TABLE 2. Matrices ranking for various short time-series: short-term forecasting.

	Matrices ranking						sMAPE(%)					
	TS1	**TS3**	**TS4**	**TS11**	**TS20**	**TS24**	**TS1**	**TS3**	**TS4**	**TS11**	**TS20**	**TS24**
GASF	5	2	4	2	**1**	2	9,984	6,563	9,285	4,36	**11,271**	4,335
GC	3	5	5	3	5	**1**	5,255	13,413	9,566	4,471	25,871	**4,242**
GCS1	4	4	**1**	5	4	4	8,225	12,258	**4,415**	6,063	25,069	4,898
GCS2	6	6	2	**1**	6	3	14,328	13,78	8,465	**3,707**	27,705	4,353
MAC	**1**	3	6	4	2	5	**1,18**	10,459	10,865	5,321	12,648	5,946
RPM	2	**1**	3	6	3	6	1,426	**5,082**	8,995	6,482	12,959	6,44

correlation between every two values of the TS, both TS parts need to be rescaled in order to have the same data encoding (i.e., within the same data range or same data magnitude). However, rescaling the TS values, for instance to $[-1, 1]$, represents an issue for some matrices (e.g., GASF). This could happen mainly when bringing matrices' values back to the original TS. As a solution, we paid more attention regarding each transformation in order to select the most suitable rescale. For instance, for the GASF, we have rescaled the original TS to $[0, 1]$, instead of $[-1, 1]$, in order to allow the transformation back and forth, i.e., between the original TS to its corresponding image.

The *second phase* consists of an exhaustive testing of various forecasting models through a grid-search approach. In fact, we test all possible combinations enriched by a progressive exploration of the studied TS. Consequently, the matrices (2D representation) representing TS are built on an incremental way, so as to be used to select the best models for each sub-part of the TS. The 2D representation of each TS has to be generated as an image. Thus, techniques for image encoding (RGB), from the generated matrix, have to be applied. We first start by a technique proposed in [32], which uses a (RGB) scale-based solution in order to transform TS values into image encoding values. This solution, despite its encoding capabilities, shows its limits when returning back to the original TS. In order to tackle this issue, we have first introduced a *dictionary-based technique*. It is a RGB dictionary, which is used to establish a static mapping of

each numerical value, from the scaled matrix into a unique three values (RGB), required for colored pixel creation. This solution allows image encoding without precision loss; albeit, it is a time consuming task as it requires numerous research requests in a high dimension dictionary of 256^3 features. The second technique generates a RGB dictionary with minimum features representing only the numerical values existing in the scaled matrix. Though, this solution does improve the execution time, it restricts the domain of the forecast values to previous values existing in the scaled matrix and hence hinders the generation of new values. This issue represents a huge information loss. The third technique generates a gray shaded image based on a static dictionary with only 256 values, representing the possible gray levels. This is quite similar to the second technique; it reduces drastically the image generation time, however, it induces an important information loss. The fourth technique, named, *dynamic image encoding*, uses a bijective function allowing encoding the scaled matrix values into three values RGB (256^3). Moreover, it supports the unique generation of RGB code for the new forecast values while reducing the image generation time.

The generated images are then extrapolated and inpainted using the adapted patch-based algorithm. Once images are inpainted, the reverse process is applied in order to both extract the forecast values from the image and reverse all the scaling steps, in adequacy with the applied matrices. The generated models are sorted on based on their accuracy (i.e., sMAPE) and are saved into a log file.

Algorithm 1 FM2I Algorithm

```
/* Phase 1: params initializing                                           */
[TS_tr, TS_test] ← structuringTS(TS_raw, horizon);
TS_test ← minMaxScalingTS(TS_test);
TS_scaled ← minMaxScalingTS(TS_tr);
stepsCount = (size(TS_scaled) / horizon) − 1;
matrixCount = 12; // 6 matrices applied/differenced TS or not
/* Phase 2: models generation                                            */
for i ← 1 to matrixCount by 1 do
    // test all matrices
    for j ← 1 to stepsCount by 2 do
        // varying TS size for cross validation
        TS = TS_scaled[1 : horizon * (j + 1)];
        m ← genMatrix(TS, i);
        img ← dynamicRGB(m);
        for k ← 3 to k ≥ horizon by 2 do
            extImg ← inPainting(pic, k, horizon);
            TS_f ← tsFromImg(extImg, horizon); 2n + 1 forcasts
            TS_fbest ← bestSelect({TS_f});
            TS_real ← TS[(j*horizon) + 1: (j + 1) * horizon];
            [TS_real, TS_fbest] ← reverseScaling(TS_real, TS_fbest)
            [mse, rmse, mape, smape] ← accuracyEval(TS_fbest, TS_real);
            modelsLog.saveModel(i, k, [mse, rmse, mape, smape]);
        end for
    end for
end for
/* Phase 3: Best model selection                                         */
strategy ← [freqItems, shortMemory];
if mode == 0  then
    [matrix, patch] = bestModel(modelsLog, strategy[0]);
else
    [matrix, patch] = bestModel(modelsLog, strategy[1]);
end if
/* Phase 4: Testing best forecasting model                               */
TS = TS_scaled[1 : setpsCount * horizon];
m ← genMatrix(TS, matrix);
img ← dynamicRGB(m);
extImg ← inPainting(pic, k, horizon);
TS_f ← tsFromImg(extImg, horizon); 2n + 1 forcasts
TS_fbest ← bestSelect({TS_f});
TS_real ← TS_test;
[TS_real, TS_fbest] ← reverseScaling(TS_real, TS_fbest);
[mse, rmse, mape, smape] ← accuracyEval(TS_fbest, TS_real);
```

The third phase exploits the log file as the basic Dataset to be mined in order to extract the best model for being used for TS forecasting. For this purpose, two strategies are proposed: the first one sorts all the models according to their performance and then applies frequent items search in order to extract the most frequent configuration based on the progressive models generation. The second strategy, named the short memory, aims to reduce the TS size used for image generation and seeks to extract the frequent items based on the last exploration of the TS. Both strategies are tested and could be alternated on a case by case basis.

The fourth phase generates forecasts based on the best model configuration specified in the previous phase. The forecast is executed through an image generation according to the best matrix. This image is extrapolated and inpainted according to the best patch size. The forecasts are extracted from the image and rescaled to rescind the scaling and differencing changes. Once the forecasts are generated they are compared to the TS_{test} part, representing the real values to forecast, extracted at the parameters initializing step.

In summary, the original TS is first scaled to $[-1, 1]$ for equalizing its initial values taking into consideration

TABLE 3. M3-competition TS.

	Types of Time Series Data						
Interval	Micro	Industry	Macro	Finance	Demog	Other	Total
Yearly	146	102	83	58	245	11	645
Quarterly	204	83	336	76	57	0	756
Monthly	474	334	312	145	111	52	1428
Other	4	0	0	29	0	141	174
Total	828	519	731	308	413	204	3003

the possible TS upper and lower bounds. The scaled TS is then used to generate its corresponding matrix by, eventually rescaling again its values to be within the interval [0, 1]. This step is related to the matrix used with the aim to make easy the back-and-forth (*B&F*) transformation process. The rescaled matrix is used to generate its corresponding image using the above-mentioned techniques (e.g., dynamic image encoding). Regarding image extrapolation, the patch-based algorithm, which was proposed mainly for image inpainting in order to fill damaged images, is adapted and applied for image extrapolation. In fact, the patch-based algorithm is enhanced for image extrapolation from the borders with a partial neighborhood, which does not surround the area of the image to be completed as conventionally applied in a given image inpainting field. The reverse techniques are then applied to get forecast values and compute the accuracy metrics.

VI. EXPERIMENTAL RESULTS

This section presents the experimental evaluation of FM2I against several top methods from both M3 and M4 competitions.

A. DATASETS AND METHODS

In order to assess the effectiveness of the FM2I, we have used the dataset from M3 competition, which is considered as the main driver for developing and comparing forecasting methods. The M3 dataset includes 3003 TS [37], which are categorized according to their types (e.g., industry, finance) and time intervals (e.g., Monthly, quarterly, yearly). In other words, the dataset includes 645 yearly TS, 756 quarterly TS, and 1428 monthly TS, which could be used respectively for short-, medium-, and long-term forecasting. The TS, which are classified as ''Other'', are an additional set, which is dedicated for medium-term forecasting. Table 3 summarizes the number of TS from each category and according to the period (e.g., Monthly, quarterly, yearly) [37].

It's worth noting that FM2I, despite still being in its infancy, it is more accurate for short term series forecasting. Indeed, the FM2I-based models are able to generate ensemble data forecasts. However, it remains very difficult to automatically select the best forecast accordingly. In fact, the huge number of the generated forecasts is mainly related to the size of the TS initial given values (see Section IX). In the meanwhile, and in order to investigate new techniques, we have focused, in this work, only on the 192 first shortest time series from M3 (i.e., all TS having 14 (152 TS) and 15 (40 TS) given values) for yearly forecasting (i.e., 6 forecast values).

In our experiments, we have used the forecasts that are made available by participants of the M3 competition. We also selected the best methods, which are featured from the competition, and achieved great performance [37]. In fact, we have considered the following top methods, from M3 competition [38], in order to assess the performance of FM2I for TS forecasting: ARARMA, Theta, ForecastPro. As stated in [38], the Theta method performs extremely well in almost all considered short TS and short-term forecasting horizon. ForecastPro was also considered as a new method, which performs well in a similar way to the Theta. ARARMA is a variant of the ARIMA method, which is considered the most sophisticated statistical-based forecasting method. We have also included the Naive method as a baseline, as also described in the M3 competition, in order to assess the performance of FM2I methods against this benchmark.

In fact, we have used these representative methods, because, on one hand, we have both the dataset and their forecasting results for comparing the accuracy of FM2I, on the other hand, these methods were tuned by specialists/experts for TS forecasting and, despite they are classical, are considered among the best performing methods in the field. This is way they have also been used for comparing new forecasting methods accordingly. Other methods will be considered in our forthcoming work, mainly those used in the new M5 competition.

We have also investigated the FM2I using similar shortest series from the M4-competition dataset (i.e., all TS having 14 (31 TS) and 15 (2657 TS) given values) with the aim to study its effectiveness for yearly forecasting (i.e., 6 forecast values) [40]. The FM2I was compared against the Smyl method [41], the winner of the competition.

B. COMPARISON AND DISCUSSIONS

Experiments have been conducted on a computer having the following characteristics, Intel Core i7-8565, 8 GHzx8, 16.00 GB RAM, x64 based processor, Ubuntu 18 LTS. Table 4 presents the comparison, in terms of forecasting accuracy, of the FM2I against the top methods, from M3 competition. The accuracy is assessed in terms of MAE, MSE, RMSE, MASE and sMAPE. As shown in this Table, the FM2I outperforms all methods in terms of MAPE and sMAPE. Similarly, reported results, as depicted in Table 5, show a close accuracy compared to the Smyl method when considering the selected TS from M4 dataset.

Table 6 depicts the average computational time (CT) for both training and forecasting phases of FM2I. In order to compute the average CT, several experiments have been conducted for a set of TS with different sizes. As shown in this Table, the FM2I method, in its current version, allows forecasting short TS in a reasonable time. However, for medium and large TS, as mentioned above, a sophisticated computing infrastructure together with innovative techniques for models' parameters tuning, are required for conducting extensive experiments with the whole dataset.

TABLE 4. Accuracy and ranks of error for short-term horizon: 192 shortest TS from M3 dataset.

	Average errors					Rank across all methods				
	MSE.103	**MAE**	**RMSE**	**MAPE(%)**	**sMAPE(%)**	**MSE**	**MAE**	**RMSE**	**MAPE(%)**	**sMAPE(%)**
FM2I	10,067.56	1,757.22	1,503.50	**35.62**	**23.77**	3	3	3	**1**	**1**
Theta	12,603.95	1,865.18	1,602.92	37.28	23.91	4	4	4	4	2
Naive	4,416.60	1,619.83	1,396.04	32.24	24.83	1	1	1	2	3
ANN	4,786.18	1,664.05	1,424.27	34.75	25.04	2	2	2	3	4
ForecastPro	18,707.97	2,102.02	1,796.87	38.92	25.61	5	5	5	5	5
ARARMA	173,228.23	3,211.93	2,632.34	48.96	27.83	6	6	6	6	6

TABLE 5. Accuracy (sMAPE) for short-term horizon: 2688 shortest TS from M4 dataset.

			sMAPE(%)	
Series	**TS_Size**	**Size+Horizon**	**Smyl**	**FM2I**
31	14	20	10.27	11.18
2657	15	21	17.19	19.60

It's worth noting that, in summary, the FM2I outperforms all top methods in terms of accuracy for shortest TS (i.e., 14 and 15 given values). However, regarding selected shortest TS from M4, the FM2I provides a good accuracy compared to the Smyl method, the winner of the competition. Nevertheless, FM2I-based models are able, compared to the considered forecasting methods (i.e., in short-, med-, and large-term forecasting), to generate the best and very accurate forecast (named the best fit forecast) with a significant gap. More details and related results are clearly presented and discussed in Section IX, but according to the obtained results and the mentioned new directions, more tuned FM2I-based models will be able to select accurate forecasts for short-, med-, and large-term forecasting.

VII. RELATED WORK

The rapid growth and deployment of IoT (Internet of Things) and wireless sensors technologies have shown a great potential for collecting high-volume, high-velocity, and high-variety of time series data. In fact, a myriad of sensors can be deployed for gathering contextual data that could be integrated with other data, such as location, weather data, and social media data [42]. The processing and analysis of these times series data allows the development of context-aware applications and services in many applications domains, such as in e-health [43], transportation [44], and energy management [39]. For instance, short-term forecasting of solar power production and utility demand could allow dynamic and predictive control of micro-grid energy systems [45].

There are two main goals in analyzing and processing time series data, classification and regression. Time series data classification allows revealing patterns and features (e.g., anomalous values), while regression allows forecasting/predicting future values according to past previous ones. Time series forecasting and analysis are key elements in many applications. Their aim is to analyze time-series trends by building a forecasting model for being used for either classification or for predicting/forecasting n-steps-ahead values.

Several approaches have been proposed in the past decades for time series data analysis and forecasting. They can be classified into three main categories, as depicted in Fig. 7: *model-driven*, *data-driven*, and *data structures-driven* approaches. In the model-based approach, time-series are transformed into state-space representations. The forecasting process is then performed by simply assuming that the forecast period observations are missing and hence apply existing methods to complete them [46]. Kalman Filter and Particle Filter [47] are among the model-based methods that can be applied to reveal repeated patterns and for forecasting time series data [48], [49]. While these approaches provide very good accuracy, they remain insufficient in catching the complexity and the dynamics of complex systems.

Data-driven approaches have been proposed to circumvent this issue by exploiting the high data volumes of measured data. The aim is to establish a more complete data model, which can be used for predicting and forecasting [3]. These approaches can be classified, in turn, into two main categories: *statistical-based* and *Neural Network-based* approaches. Statistical-based approaches have been proposed to predict future values as the product of several past observations. Examples of these approaches are ARIMA [50] and exponential smoothing (ES) [46]. Artificial Neural Network-based approaches have been proposed as information processing models for time series data analysis and prediction [51]. Examples are multi-layer feed forward neural network and LSTM (Long Short Term Memory). However, while these models did provide accurate forecasting results, they typically require larger training datasets in order to learn accurate models [49]. As stated in [52], realistic TS are somehow characterized by both nonstationarity and non-linearity, and therefore, it is difficult for data-driven methods to achieve high accuracy within a reasonable computational time.

Data structures-based approaches have been proposed for time series data analysis and forecasting. These approaches can be classified into three main categories: *Tree-based*, *Network-based* and *Image-based* approaches. Classification and regression trees, random forests and gradient boosting trees are the main Tree-based methods that have been proposed for both classification and regression problems [51]. Network-based concepts have been proposed recently for studying and analyzing time series. The aim is to extract and reveal dynamically relevant statistical properties of time series [53]. Studying TS from network (weighted and unweighted) viewpoints has created a new perspective for time series analysis [54]. In this latter, authors proposed a bridge between a time series and a complex network with well-conserved structural features and properties. For instance, inspired by this network-based approach, a novel method for CCI (Construction Cost Index) forecasting is proposed in [55]. It is mainly based on the visibility graph and
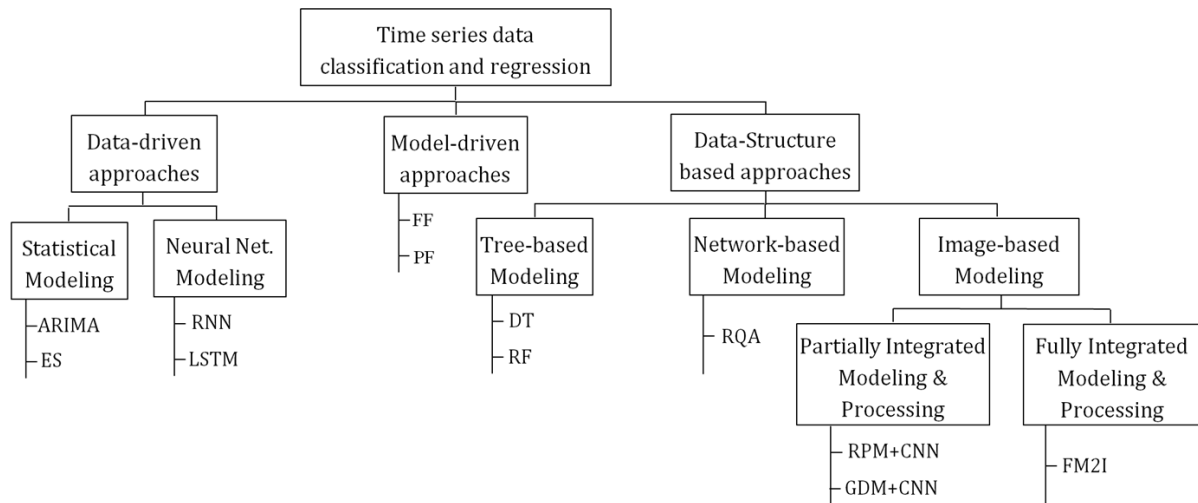
FIGURE 7. Approaches for time series classification and regression.

TABLE 6. Computational time (CT) for a set of shortest TS for short-term horizon: training, forecasting, min, max, and total.

Series	TS_Size	Size+Horizon	CT-Training(s)	Min-CT-Forecast(s)	Max-CT-Forecast(s)	Forecast average CT(s)	Total CT(s)
S1	15	21	2.84	0.17	0.58	0.27	3.11
S2	14	20	2.61	0.17	0.52	0.26	2.88

links' prediction methods in order to compute the network nodes' similarities. Reported results showed that this prediction method provides good prediction accuracy. However, this method, despite its bridging originality, it was used only for predicting the next TS value (i.e., One-step-ahead forecast). In fact, this method could be extended and tested, not only for CCI predictions, as stated by the Authors, but also for other TS predictions (i.e., other areas). It could be also extended to handle short-, med-, and long-term TS forecasting. Similarly, in [52], a multiple time-frequency spaces fuzzy interval forecasting approach, based on visibility network, is proposed for energy and finance forecasting. One-step-ahead forecasting is mainly performed by using link prediction methods. However, despite the accuracy of the proposed approach against state-of-the-art ones, the process of transforming TS into complex networks is strictly irreversible, and requires depth investigation. Still, its application for multivariate TS could not be handled, as also state by the authors.

Recently, many research studies propose to transform and encode time series into different types of images, such as GASF/GADF (Gramian Angular Summation/Difference Fields) and MTF (Markov Transition Fields) [32]. These approaches fall into partially integrated modeling and processing categories, as described in Fig. 7. The main aim of most of these studies is to enable and use techniques from image recognition and signal processing communities for time-series classification. All approaches use Artificial Neural Networks in order to reveal topological properties from time series-based images. For instance, in [32], Wang and Oates propose to use deep Tiled Convolutional Neural Networks in order to classify time series for multi-level features extraction after encoding time series as GASF/GADF/MTF images. Similarly, Yang et al. in [56] have introduced a

framework, which allows encoding multivariate time series data into 2D GASF/GADF/MTF-based images, for features classification through a Convolutional Neural Network. The experimental results, using two multivariate datasets, show that these image-based structures do not have a significant impact on the accuracy. Furthermore, Chen and Shi in [33] propose to encode univariate time series data into a Relative Position Matrix and feed it to a Convolutional Neural Network for classification purposes. Unlike the time series classification, similar research studies have been conducted for time series forecasting. For instance, Li et al., in [34], propose an approach that transforms time series data into graphical representations (in the form of 2D plotted images) and then apply deep learning methods for data forecasting. Similarly, authors in [35] introduce an approach for extracting time series features using computer vision techniques (e.g., spatial bag-of-features [57], deep CNN). Alike the approach proposed in [34], authors propose to first transform time series into graphical representations (recurrence plots) and then use computer vision techniques to extract time series features and reveal the characteristics for being used by the CNN framework for time series forecasting.

However, despite the combination of these image-based structures with state-of-the-art deep learning methods provide the best accuracy, the relationships between extracted topological features and time series data remains unclear, since no inverse operations have been proposed to transform the produced and enhanced 2D images into the new 1D sequence of time series data [32]. Furthermore, most of the approaches proposed so far use partially integrated image-based modeling and processing techniques combined with conventional neuronal approaches as depicted in Fig. 7.

In the proposed FM2I, a fully integrated modeling and processing framework, as depicted in Fig. 7, we transform the time series forecasting into fully images- and signal-based processing procedures. In fact, forecasting time series data is transformed into a problem of completing a picture (i.e., without falling back on machine learning techniques). After transforming a time series data into its corresponding image, the problem of *data forecasting* becomes essentially a problem of *image inpainting* problem, i.e., completing the missing pixel in the image. In summary, the main contributions of this work against the above-mentioned approaches are five folds:

- New matrices are proposed for 2D image-based representations of time series data,
- New transformation techniques, from time series data to image-based representation,
- Image-based forecasting techniques based on inpainting methods,
- Re-inverse transformation techniques, from 2D representation to 1D vector representing the original time series data.
- Performance evaluation using datasets from M3 and M4 competitions together with extensive comparison of the FM2I against the top M3 and M4 approaches.

It is worth noting that the work presented in this paper initiates a new branch of TS forecasting (see Fig. 7). In fact, instead of learning TS patterns and features directly from temporal dimension, the spatial representation is used along with image-based processing techniques and tools. More precisely, a 2D image representation could reveal interesting and richer patterns and features, which could not be explicitly represented in a 1D TS.

VIII. PHILOSOPHICAL DISCUSSION AND OTHER POTENTIAL DIRECTIONS

In this work, we have started by pointing out philosophical considerations and reached unexpected and astonishing results: *completing a painting enables us to see the future in color*! A 'visual' 2D predictive procedure works for a 1D chronological data! Even after completing this work, we are not yet able to answer completely the central question concerning the nature of predictions developed in section 2. Nevertheless, we would like to draw some philosophical conclusions from this fruitful endeavor. It is obvious that we need more than a discussion section and certainly more than an article to present in details all our reflections, which we will surely do in forthcoming works. But, we want at least to summarize some of our queries about the implications of this work. More precisely, we have chosen to focus on the two following aspects.

A. THE VALUE OF PREDICTION VS EXPLANATION: THE NEW EMERGING ROLE OF PREDICTION AND BRIDGING PHILOSOPHY

As mentioned in [8], in contrast with scientific explanation, the value or the role of scientific prediction seems to have always been clear. Indeed, common sense has it that predictions are used for their own interest: *predict* what will be going on in order to make better decisions. At the same time, it helps testing the strength of the efficient way of understanding the surrounding world. In addition, as indicated in Section 2, according to explicit or tacit scientific practices, philosophers of science wish to formalize other interests of predictions as validating or invalidating theories and models and then assessing the maturity of theories ([10] and [19]). When the value of explanation is unclear as indicated in [8] with reference to the work [7] entitled ''Why Ask Why?'', see also [58], [59]. More precisely, the argument is: suppose that we possess a perfect power in predicting everything in this world (like Laplace's demon), then the need of explanation becomes unnecessary or serves at most to assuage our thirst for understanding and therefore should be seen as a simple 'psychological satisfaction' which makes it epistemically suspect [60]. The conclusion from this thought-experiment can be corroborated by the actual success of *Data Sciences* based on *Statistics* and *Machine Learning*, which are *black boxes* with explanation free knowledge.

Now suppose we hold the reverse of the argument: we have a perfect explanation power making us capable to explain everything, do we always need predictions? Certainly, yes. Indeed, we need them in order to make, accordingly, the most adequate decisions. At this stage, we point out a certain *Dissymmetry* between predictions and explanations as the value of the first one appears better compared to the second one. There is a bias in this reasoning, though can we consider an explanation without prediction or the reverse? A development concerning the first part of this question can be found in [8] and [60]. More precisely, authors highlight that we can talk about scientific explanations *if and only if* this last is able to generate new testable predictions in order to explore theories implications, which would probably justify why scientists appreciate explanations. This ties in with, in some way, the famous Karl Popper's idea of *falsifiability* or *refutability*. When predictions can arise from explanations but don't necessarily need them (see the definition of pediction considered in Section 2), then the dissymmetry persists. In fact, in the absence of perfect predictive processes, explanations are still needed to properly assess the accuracy of predictions and to accept them by scientists [61]. All in all, we think that as our predictive power grows, so does the need of explanation decrease.

In light of the conclusion of this work, we can point out new perspectives on the role of scientific prediction. Indeed, as depicted in the general framework (Fig. 1), the bottom line of our idea is to test the extensivity of the theories' or models' predictive capacity. We could actually conclude that: on the one hand, achieving our difficult challenge by succeeding our first test proves that it is doable, and that all without necessarily being experts in the fields studied. On the other hand, it is worth a try it, since this sort of *modus operandi* offers scientists the opportunity to create *Bridges* between different scientific fields/subfields or different scientific theories

by following predictions success. At that moment of our experiments, we wonder if there are other examples of these bridges in science. After careful and extensive research, we found some examples of the tacit applications of the Fig. 1. We chose to present a significant example from theoretical physics, which also focuses the attention of some philosophers of science [62] and [63], but never from the angle developed here. Thus, let us mention the famous so called *AdS/CFT Correspondence*, first introduced by Juan Maldacena in the mostly cited article of high energy physics (20000 citations) [64] and most significant achievement in string theory in the last twenty years. From our perspective, this example illustrates perfectly our ideal conclusion.

Indeed, in theoretical physics, the anti-de Sitter/conformal field claims the equivalence between the strongly coupled four dimensional gauge theory and gravitational theory in five dimensions. In particular, it can be seen as a bridge between two completely different fields of physical theory, since it is an unexpected link between Gravity theory in five dimensions and Quantum field theory in four dimensions. This correspondence deserved to be explained in details, for the sake of conciseness, we refer the interested readers to [65] and [64]. Nonetheless, we must point out that there is no formal mathematical proof of this correspondence, even if, there exists an *AdS/CFT dictionary* permitting to switch between the two theories. An example of applying the previous process can be clearly identified in nuclear physics. Briefly, in order to study quark-gluon plasma, instead of using its insoluble mathematical formalism, in [66], according to the AdS/CFT correspondence, authors predict the value of the ratio of shear viscosity to the entropy density associated with the quark-gluon plasma, which is close to the experimental results. This leads to a new understanding and then new explanations concerning quark-gluon plasma through this correspondence. Here, we reach a new benefit of prediction since using this precess we may find, if even possible, new explanations by reversing the conventional closed loop, depicted in Fig. 8a.

Starting by testing the extensivity of theories or models' predictive capacity, we select the theory or model, which provides the most accurate predictions (i.e., there is a bridge) and only after we go back to the previous process by trying to obtain explanations and so on (see Fig. 8b).

In fact, by systematizing the approach, described in this work, and by identifying the possible bridges according to their predictions accuracy, we will surely find more correspondences, and should not be dependent on the scientists' genius like Maldacena. In the next subsection, we will try to go further by investigating the deeper reason for the existence of these bridges. Nevertheless, this interesting gain relies on overcoming the major difficulty of how we can switch between scientific disciplines/sub-disciplines. More precisely, it is worth experimenting how we can transform initial conditions or data's field/subfield to be an input of a theory or a model of another one? In other words, we wish to develop in forthcoming works a more philosophical/mathematical framework, called *Correspondence*, when

we hope to formalize in details this idea (we have very promising leads). In our example (FM2I), we have moved from 1D temporal series data structure to a suitable 2D image data structure. First of all, let us begin by designating the benefits of these bridges by giving three essential aspects, which can also be considered as novel roles of scientific predictions:

- *New Models and Methods:* Prediction has always helped scientists to select between models and methods inside a field/subfield. But, in the light of this work, it appears that this selection could be generalized between models and methods from other fields/subfields, since we found a correspondence permitting us to switch between them. In our example, we have found new forecasting models and methods arising from image inpainting methods.
- *New Correspondences and Analogies:* We will develop our perspective in details concerning these notions in a forthcoming work. So far, we have presented above an example of correspondence based on a dictionary permitting to switch between two theories. We can also obtain other types of correspondence through other transformations like those developed in this paper. Of course, the different types of correspondences should be chosen according to their predictions accuracy. Here, we meet another new role of prediction. Analogies are also a type of correspondence, which are very fruitful processes used by many greatest scientist like J.C. Maxwell and H. Poincaré (see [67] and [68]). Then, we hope we can find new analogies following this procedure.
- *Towards New Theories:* The correspondences selected by predictions could lead to new theories, a discussion concerning an example related to analogies between *Heat and Electricity* can be found in [69].

B. WHY DO THESE BRIDGES EXIST? EXTENSIVE STRUCTURAL REALISM

Wondering the reason of these bridges' existence, we have naturally converged towards a modified version of the Poincaré's *Structural Realism* thesis. Let us begin by briefly recalling some philosophy of scientific modes of thinking in order to better clarify our position. Concretely, *Scientific Realism* posits that mature theories with an undisputed predictive success are true or approximatively true in the sense that the structure assumed in these theories reflect the actual reality. The main argument in support of scientific realism is called *Miracle Argument*, which is introduced by Hilary Putnam when he wrote "*The positive argument for realism is that it is the only philosophy that doesn't make the success of science a miracle*" [70]. This means that it is virtually impossible that our theories succeed to describe observable phenomena and at the same time be completely false. The principal argument against scientific realism is called *Pessimistic Induction*. Based on history of science, this theory claims that since most of our past theories, even strongly
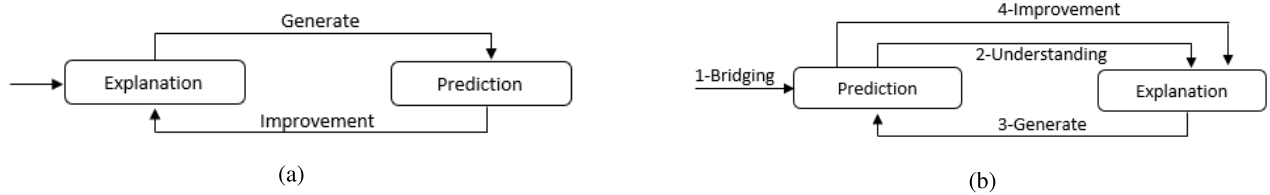
FIGURE 8. Closed loop: a) conventional, b) new.

confirmed, have turned out to be incorrect, thus, it is very improbable that our current theories be true. According to this argument, the present theories are ephemeral and are likely to change; it is just a matter of time.

In an article entitled *"Structural Realism: The Best of Both Worlds?"* [71], John Worrall elaborated a position inspired by Poincaré's works. More precisely, structural realism takes into the account the pessimistic induction concerning the theoretical entities involved in current science, but at the same time it maintains the validity of the miracle argument at the mathematical structures level. Indeed, the famous example proposed in [71] and also analyzed by Poincaré in [72] is very relevant. Specifically, Fresnel's wave theory is based on the wrong postulate of an entity existence, called *Luminiferous Aether*, as the medium for the propagation of light. Augustin-Jean Fresnel developed a mathematical formula, called *Fresnel diffraction integral*, which was validated since it succeeded in observing and predicting the famous Frenel's spot. This example proves that there is some truth in Frenel's theory, which is clearly its mathematical structure and not the existence of an entity permitting the propagation of light. In fact, Frenel's equations can be seen as a particular case of Maxwell's equations, which does not assume any ether existence. Therefore, Frenel's mathematical structure can be viewed as approximatively true since it is conserved as a particular case in the latter theories dealing with light propagation like electromagnetism, including quantum theory.

In this respect, we think that we should not be confined to conserve mathematical structures in latter theories treating the same phenomenon. We have to distinguish between *Intensive* and *Extensive* structural realism. Indeed, we can posit that this propriety of conserved mathematical structure is an *Intensive Structural Realism*. While, in the light of this article development, we are able to introduce what an *Extensive Structural Realism* can be. Specifically, in most cases, testing the extensivity of theories' predictive capacity means testing the extensivity of the predictive capacity of its mathematical structures. Indeed, the above example of the prediction concerning the quark-glon plasma according to the AdS/CFT correspondence is specifically due to the mathematical language of the string theory. More accurately, we can conclude that the capacity of certain mathematical structures of working in several domains means that there exists a kind of universality of these structures. This allows us to think about the miracle argument at the level of the mathematical structures extensivity, in the sense that it should be a miracle

that a mathematical equation or formula appears in many completely different domains without reaching any reality. We can find many examples (see [67]) in physical analogies when an equation or formula works in different fields. In this sense, extensive structural realism is the capacity for some mathematical structures to arise in domains other than those for which it was initially elaborated.

Let us provide another example in relationship with this work. We are currently developing a forecasting promising method based on some partial differential equations like Heat and Navier-Stokes equations through adequate bridges (image). Knowing that heat equation, for example, was developed by Joseph Fourier to model heat diffusion; nevertheless, this equation (or its modified version) appears in many other domains like electricity, image restoration, Brownian motion, finance, quantum mechanics, forecasting, etc, which means that this equation possesses something universal, at least at a first order. Thus, it should be a miracle that this equation be completely divorced from a certain reality. That is what we called the *extensive structural realism*. Hence, as one of the objectives of this work is to highlight the interest of creating new bridges between theories and models we naturally highlighted the question of their existence. After careful consideration, we strongly think that this existence is due to the existence of mathematical structures, like heat equation; in other words, the miracle argument at the level of the mathematical structures recurrence or persistence. The basic idea is then, is that we are sure that these types of mathematical structures exist, since we possess many examples, then we have to strengthen the existing and find new ones by trying to test bridges according to their predictions' accuracy.

We must point out that the discussion above suggests a significant link with some modes of thinking from *Philosophy of Mathematics*. For example, it should be very interesting to further analyze these ideas in light of the *Mathematical Realism* concerning the question of mathematics and mathematical objects' objectivity. We will focus on such links in a forthcoming work. Let us conclude this subsection by the following reflection. Elaborating a theory of everything, capable of describing and explaining all phenomena in the universe, is a precious physicists' hope. One of the main impediments is the difficulty of finding a theory unifying gravitation and standard model of particle physics. Whether this unification is really possible, it is reasonable to think that it should be realized through a mathematical structure. Hence, this last must be able to explain and predict everything. We think

that it will be more suitable to proceed by trying to find this structure through its capacity of predicting everything and only after that check (or not) its capacity of explaining and describing everything, and not the reverse, which seems much more difficult. Indeed, if we are able to evacuate the major difficulty of bridging theories and models through adequate correspondences, we will be able to test this procedure. Who knows? Perhaps the equation of everything already exists in the literature; it just needs to be tested!

IX. SCIENTIFIC DISCUSSION AND OTHER POTENTIAL DIRECTIONS

The ideas developed in this article are clearly the beginning of a large series of works since there is an important variety of other potential directions. We briefly focus on two aspects or questions: is the FM2I improvable? Why it works? Should it be adapted for real-time data streaming?

A. PARAMETERS TUNING AND OPTIMIZATION

Parameters tuning of the FM2I forecasting algorithm is based on a grid search through testing numerous combinations of matrices (conventional and differenced ones), min-max TS scaling and patch size. It generates a set of candidate forecasting models during the progressive TS exploration. The aim is to extract the most frequent and best suitable configuration (i.e., scenarios). Generally, this model provides accurate forecasts, however, our exhaustive forecasting approaches show that this model is not necessarily the best fitting one.

Fig. 9 and Table 7 show, as an initial proof of this statement, the improvement of FM2I when using the best fitting model (FM2I, BF model). The accuracy differences show that adequate improvement could be performed in order to select the best model (i.e., Forecast-S1 in the current case, near the real TS). This could be achieved through adapted and innovative approaches.

It is worth noting that we have also conducted extensive experiments of the FM2I using the best fitting model, using M3 competition dataset, for short-term horizon (645 TS), med-term horizon (756 TS and other 174 TS) as shown in Tables 8, 9 and 10. As shown in Table 10, the FM2I BF model outperforms the considered methods, in the three forecasting horizons. Furthermore, a representative subset of the M4 competition dataset [73] is used, in particular, for short-term horizon forecasting, as depicted in Table 12. This latter shows similar performance in terms of accuracy and rank.

These preliminary results are promising, but still finding the optimization methods, which could select the best fitting model is required and represent our challenging ongoing work.

B. TOWARDS ENSEMBLE DATA AUTOCORELATION FORECASTING

Several methods, as depicted in Fig. 7, have been proposed for TS forecasting. Many recent studies have shown that none of these forecasting methods is alone able to model

TABLE 7. Metrics for the best scenario via the best fitting model: Forecast-S1.

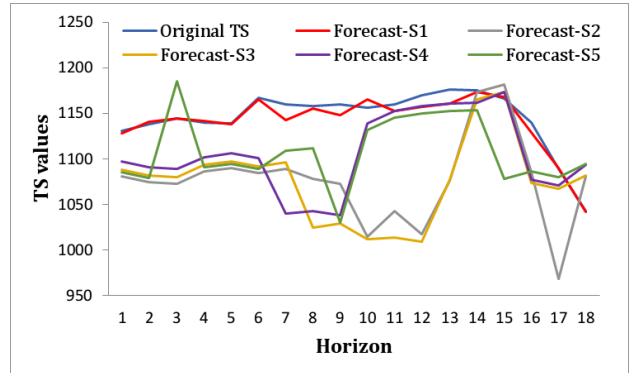| Scenarios	MSE	RMSE	MAE	MAPE(%)	sMAPE(%)
Forecast-S1	61,44	7,83	5,61	0,48	0,49
Forecast-S2	7161,03	84,62	74,93	6,54	6,81
Forecast-S3	7886,58	88,80	74,86	6,50	6,80
Forecast-S4	3537,93	59,48	46,42	4,06	4,18
Forecast-S5	3043,92	55,17	47,21	4,12	4,22

FIGURE 9. Forecasting values: Forecast-S1 represents the best fitting model.

real-world TS. Hybrid ensemble models, named also ensemble learning model, have been proposed for accurate TS forecasting, combining then strengths of each forecasting method. For instance, they have been heavily used in the past decades for weather forecasting [74]. These models' aim is to combine several forecasting models in order to increase the accuracy against the usage of a single model [75], [76]. In fact, an ensemble of independent models (i.e., multiple and diverse) could be established for being used in the collective prediction process [77]. For instance, in [78] authors propose a hybrid model, combining ARIMA and Elman artificial neural network (EANN), for TS forecasting. Reported results show that the proposed hybrid model outperforms both ARIMA and EANN when considered separately. Similarly, authors in [79] propose an ensemble learning model composed of the decision tree, gradient boosted trees and random forest models and results show the effectiveness of this hybrid approach. Most of work to-date show that combining multiple forecasts, which are generated by an ensemble models, could generally provide better forecasts with higher accuracy [80]. Their efficiency was effectively highlighted in the M4 competition by the Smyl's approach, which won the competition. This latter, named ensemble of specialists, combines the LSTM (long short term memory) neural network with an ES model (exponential smoothing) (ES) model, for TS forecasting.

Unlike ensemble learning models, in our proposed FM2I, we are able to forecast $O(n^h)$ possibilities, where n is the TS size and h is the considered horizon. Thus, FM2I can be seen as *ensemble data autocorrelation forecasting* in which, instead of generating an ensemble of learning models, several ensemble data are generated (i.e., forecast scenarios). However, methods are required to select the best suitable forecast that maximizes the accuracy, which still a challenging task.

TABLE 8. sMAPE and ranks of error for short-term horizon: 645 TS.

Method	Average errors					Rank across all methods				
	MSE $.10^3$	MAE	RMSE	MAPE(%)	sMAPE(%)	MAE	MSE	RMSE	MAPE(%)	sMAPE(%)
FM2I-BF	**644.67**	**497.92**	**386.06**	**9.22**	**6.90**	1	1	1	1	1
Theta	6626.00	1091.46	1252.70	22.58	16.97	4	4	4	5	2
ForecastPro	10706.26	1176.78	1354.30	22.23	17.27	5	5	5	4	3
Naive	2732.26	1025.84	1178.58	20.88	17.88	2	2	2	2	4
ARARMA	70250.21	1598.18	1896.22	26.73	18.35	6	6	6	6	5
ANN	3201.64	1049.43	1215.62	21.83	18.56	3	3	3	3	6

TABLE 9. sMAPE and ranks of error for med-term horizon: 756 TS.

Method	Average errors					Rank across all methods				
	MSE $.10^3$	MAE	RMSE	MAPE (%)	sMAPE(%)	MAE	MSE	RMSE	MAPE(%)	sMAPE(%)
FM2I-BF	**242.06**	**297.73**	**226.61**	**7.01**	**4.41**	1	1	1	1	1
Theta	850.32	475.41	557.23	11.67	8.95	2	2	2	2	2
ForecastPro	1130.45	528.18	615.92	12.94	9.81	4	5	4	6	3
Naive	1026.69	523.73	611.44	12.38	9.95	3	4	3	3	4
ANN	995.24	534.46	642.69	12.81	10.15	5	3	6	4	5
ARARMA	1358.64	549.01	637.41	12.88	10.18	6	6	5	5	6

TABLE 10. sMAPE and ranks of error for med-term horizon: other 174 TS.

Method	Average error					Rank across all methods				
	MSE $.10^3$	RMSE	MAE	MAPE(%)	sMAPE(%)	MSE	RMSE	MAE	MAPE(%)	sMAPE(%)
FM2I, BF model	**41.96**	**95.8**	**75.89**	**1.81**	**1.78**	1	1	1	1	1
ARARMA	140.02	213.53	186.04	4.67	4.38	2	2	2	2	2
Theta	208.93	223.98	197.11	4.87	4.40	3	3	3	3	3
ForecastPro	222.49	235.21	204.94	5.10	4.60	5	4	4	4	4
ANN	214.77	279.10	226.83	5.75	5.23	4	5	5	5	5
Naive	278.35	309.88	278.43	7.02	6.30	6	6	6	6	6

TABLE 11. FM2I BF model rank against other methods.

		FM2I, BF model	Naive	ARARMA	Theta	ANN	ForecastPro
Short-Horizon	times ranked best	557	18	33	9	22	6
	% ranked best model	86.36%	2,79%	5.12%	1.4%	3.41%	0.93%
Med-Horizon	times ranked best	563	28	48	56	13	48
	% ranked best model	74.47%	3.7%	6.35%	7.41%	1.72%	6.35%
Other	times ranked best	148	1	11	7	2	5
	% ranked best model	85.06%	0.57%	6.32%	4.02%	1.15%	2.87%

TABLE 12. FM2I BF model rank against other methods from M4 competition, short-term horizon (6) for 100 TS.

Method	Average errors					Rank across all methods				
	MSE $.10^3$	RMSE	MAE	MAPE%	sMAPE %	MSE	RMSE	MAE	MAPE%	sMAPE %
FM2I, BF model	**1582.757**	**454.24**	**334.26**	**4.52**	**4.10**	1	1	1	1	1
Smyl	2545.43	960.21	847.53	13	12.28	2	2	2	4	2
Theta	3547	1033.63	912.21	12.09	12.73	3	3	3	2	3
Naive	3958.41	1118.07	984.38	12.44	13.49	4	4	4	3	4
ANN	6450.3	1379.16	1293.3	17.36	16.93	5	5	5	5	5
ARIMA	14151.88	1458.77	1300.24	18.95	17.69	6	6	6	6	6

C. AUGMENTED DIMENSION PREDICTION AND ENTROPY REDUCTION

Revisiting this works' outcome, we naturally put more emphasize on the following question: Why does this bridge works so well? The first idea that we have highlighted is: it should be a link with augmenting dimension! Indeed, we start by wishing to predict forthcoming values of 1D TS, we transform it to a richer 2D structure and then predict a region related to these values. It turns out that this forecasting method is very accurate, and then we test the effect of augmenting dimension. We will explain in details this link in a forthcoming work. The second idea we are investigating concerns the entropy. Indeed, it is well known that the entropy is related to TS's complexity and then to the forecasting performance [81], [82]. Thus, basically, according to the accuracy of FM2I, we have absolutely augmented the *forecastability* of our initial TS. This indicates that we have reduced the initial entropy (somehow) of TS by augmenting its information through our TS-image transformations. In other words, the idea is to measure the amount of information gained for each TS-image transformations.

D. REAL-TIME TS FORECASTING

Real-time TS streams processing induces two main challenges, limited resources and frequent data distribution changes. The first challenge requires using distributed computing platforms for handling the high volume and velocity of data streams. Regarding the second challenge, time series are usually received from data sources with several imperfections, such as noise and/or redundancies, missing values and inconsistencies. Consequently, predictive algorithms will perform poorly using low quality data. Data preprocessing or data preparation techniques, such as cleaning integration, normalization and transformation are highly required in order

to deal with this issue. In addition, the amount and the dimension of received data is growing considerably with the emergence of IoT devises will be connected and embedded in our environments (e.g., buildings, vehicles), and reduction techniques become mandatory for dimensions reduction and simplification (e.g., feature selection). These techniques will help in providing high quality data with reduced dimensions, which are necessary for faster training and better understandability of results. Another important characteristic of data is their continuous and high speed arrivals (i.e., data streams). Many emerging real-world applications are generating data streams, and therefore, dataset ever-grow inconsiderably. However, techniques are required to cope with the time and memory constraints of the high arrivals of data streams. These streams might have a non-stationary behavior that could lead to the concept drift in which the distribution of data streams change frequently [83]. So, learning techniques must be adaptive and learn from new arrival data (i.e., adapt to changes in the processes that generate data streams) especially for context-driven applications requiring real-time or near time decisions (e.g., road accident avoidance, patient monitoring). In this direction, FM2I is under deployment in real-sitting scenarios in order to study its effectiveness in dealing with real-time TS streams. We are also adapting it to deal with multivariate TS for both batch and stream processing.

X. CONCLUSION AND FUTURE WORK
In this paper, we have first presented a research framework for scientific prediction. This latter was used to develop a new forecasting method, FM2I. We have conducted extensive experiments using the well-known M3 dataset for short TS forecasting. Results have been reported and showed that FM2I outperforms the top M3 methods in terms of accuracy for short-term time series forecasting. Our first investigations using the M4 dataset showed that FM2I is able to generate ensemble data forecasts, which contain the very accurate one, compared to existing and considered methods. Our ongoing work focuses, according to our scientific and philosophical considerations, on the development of the enhanced version of the proposed method for medium and large TS forecasting. In particular, the focus will be on the development of new techniques, which allow selecting, from the generated ensemble data forecasts, the best suitable and accurate forecast. The method is also under investigation for being deployed in real-sitting scenarios taking into consideration the required computing infrastructures. The main aim is to study its effectiveness, in terms of accuracy and computational complexity, in dealing with real-time TS streams including multivariate TS.

APPENDIX
APPENDIX A
Generally, data are transmitted as electrical signals. Thus, by electrical analogy, signals could also be classified according to their energy characteristics, signals with finite energy or with finite power. In our case, we consider a signal $x(t)$ as

a random with a finite average power. This type of signals is also called a permanent signal. In other words, a signal $x(t)$ is permanent if $P(x(t)) = \lim_{T \to +\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} |x(t)|^2 dt < +\infty$. Actually, it is practical and reasonable to add the condition of being bounded. In other words, we will consider finite power signals, which are in addition an element of $L^\infty(\mathbb{R}) = \{x(t)/ \sup_{t \in \mathbb{R}} |x(t)| < +\infty\}$. The class of functions $L^\infty(\mathbb{R})$ is a smaller than the class of functions with finite average power because if $x \in L^\infty(\mathbb{R})$ then it is a finite power:

$$\lim_{T \to +\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} |x(t)|^2 dt \leq \sup_{t \in \mathbb{R}} |x(t)|^2 = \|x\|_{L^\infty(\mathbb{R})}^2 < +\infty \tag{3}$$

The class of finite average power signals is a fairly broad class because it also contains finite energy signals. Indeed, a signal $y(t)$ is said to be of finite energy, also called transient signal if

$$W(y(t)) = \int_{\mathbb{R}} |y(t)|^2 dt < +\infty \Leftrightarrow \|y\|_{L^2(\mathbb{R})} < +\infty \tag{4}$$

Thus, the finite energy signals are the functions of squares summable on $\mathbb{R}$. In addition, a signal with a finite energy is a signal with zero average power and, therefore, is finite. Indeed, we have:

$$P(y(t)) = \lim_{T \to +\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} |y(t)|^2 dt \leq \lim_{T \to +\infty} \frac{1}{T} W(y(t)) = 0 \tag{5}$$

In practice, signals are given over a finite duration and their average power P is, therefore, calculated over a window of finite size. This implies that signals with finite energy will have a small but not zero average power. For this reason, we have considered a large class, which includes practically all varieties of random signals. Furthermore, the average power and the energy can be calculated in the discrete case, namely, for a digital signal. So, to summarize, we can consider our time series as digital signals resulting from the sampling of permanent random and analog signals $x(t)$ with finite and bounded power.

APPENDIX B
The PSD allows specifying the frequency content of a signal. It is important to note that the spectrum of a signal $x(t)$ is the modulus of its Fourier transform $|\mathcal{F}(x(t))| = |X(\xi)|$, where $\mathcal{F}$ denotes the Fourier transform. We then define the PSD of the signal $x(t)$ as the square of the spectrum of $x(t)$, namely, $S(\xi) = |X(\xi)|^2$. In our case, the signal $x(t)$ is assumed to be at finite average power. It is, therefore, not necessarily integrable over $\mathbb{R}$ and its Fourier transform does not necessarily exist, at least in this form. We can, however, obtain the same formula asymptotically. More precisely, we can define the Fourier transform on a window of size T, denoted by $X(\xi, T)$. Thus, we can speak of spectral density on a window, which is named by $S(\xi, T) = |X(\xi, T)|^2$.

Therefore, the PSD of this type of signal can be defined by $S(\xi) = \lim_{T \to +\infty} \frac{1}{T} |X(\xi, T)|^2$.

Now, we can make explicitly a relationship between $\Gamma(\tau)$ and $S(\xi)$. Indeed, using Parseval's theorem we can show that the PSD is the Fourier transform of the temporal autocorrelation function (here the Fourier transform can be in the sense of a distribution), $S(\xi) = \mathcal{F}(\Gamma(\tau)) = \int_{\mathbb{R}} \Gamma(\tau) \exp(-2i\tau\xi) d\tau$.

APPENDIX C

So far, we have clarified, from signal processing perspectives, the relationships between time series and signals. We also explained the choice of (STAM) in relation to its properties and highlighted its features from signal processing principles. This part focuses, however, on the particular case dealing with stationary and ergodic random signals. More precisely, we have considered (TS) as a random digital signal, so it can be considered as a discrete stochastic (or random) process indexed by time $(X_i)_{i \in \mathbb{N}}$. It is a family of random variables, at each instant $i \in \mathbb{N}$, the random variable X_i has a certain probability distribution and the value s_i of (TS) is an instance of it. We suppose that all these random variables $(X_i)_{i \in \mathbb{N}}$ have probability densities that we denote by $(f_i)_{i \in \mathbb{N}}$. We also denote by $f_{i,j}$ the joint probability density of the two random variables X_i and X_j.

Let us denote by $\mathbb{E}(X)$ the expectation of X. The *Statistical Autocorrelation Function* between time i and j is written:

$$R(i,j) = \mathbb{E}(X_i X_j) = \int_{\mathbb{R}} \int_{\mathbb{R}} x_i x_j f_{i,j}(x_i, x_j) dx_i dx_j \quad (6)$$

This function is linked to an important function, it is the function of auto-covariance. Now, if we want to use the tools of frequentist statistics (or classical statistics), we start by assuming, at least, that the stochastic process $(X_i)_{i \in \mathbb{N}}$ is *stationary*.

Definition 1: A random process is said to be stationary in the strict/strong sense (or to order 1) if all its statistical properties are invariant at any time scale. More precisely, all the $(X_i)_{i \in \mathbb{N}}$ have the same probability distribution and for any indices $i \neq j$ the joint distribution $f_{i,j}$ is invariant by translation of time, i.e., $f_{i,j} = f_{l,k}$ si $i - j = l - k$.

The process can be seen as stationary in the broad/weak sense (or of order 2) if the statistical properties of order 1 (expectation and variance) and 2 (statistical autocorrelation function) are invariant over time.

As strict stationary $\Rightarrow$ stationary broad, if the stochastic process $(X_i)_{i \in \mathbb{N}}$ is stationary strict or broad, then we will have:

$$R(i,j) = R(\tau) \quad where \quad i - j = \tau, \quad (7)$$

$$\forall l \in \mathbb{N}, \quad R(i,j) = \mathbb{E}(X_l X_{l-\tau}) \quad where \quad i - j = \tau \quad (8)$$

Theorem 1 (Wiener-Khinchin-Einstein Theorem): For a stationary stochastic process in the broad sense, the PSD is the Fourier transform of its statistical autocorrelation function:

$$S(\xi) = \mathcal{F}(R(\tau)) = \int_{\mathbb{R}} R(\tau) \exp(-2i\tau\xi) d\tau \quad (9)$$

This result is very important; it has been independently demonstrated by Wiener and Khinchin. Its statement was also anticipated by Einstein. This first theorem gives us a first link between the functions Γ and R, the temporal and statistical autocorrelation. Indeed, the Fourier transform of the two functions gives the PSD.

Definition 2: A random process is said to be ergodic if the statistical means are equal to the temporal average values:

$$\mathbb{E}(h(X)) = \int_{\mathbb{R}} h(y) f(y) dy = \lim_{T \to +\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} h(x(t)) dt \quad (10)$$

Thus, if our process is stationary (wide) and ergodic, we have an equality between the statistical autocorrelation function and the temporal autocorrelation function:

$$\Gamma(\tau) = R(\tau) \quad (11)$$

Thus, our (STAM), defined for any (ST), is equal to the statistical autocorrelation matrix for an ergodic stationary stochastic process. Stationarity is testable but the ergodicity, however, could not be tested, it is only assumed.

Remark 1: If we use the Wiener-Khinchin-Einstein theorem, as the PSD is the Fourier transform of Γ and R. Thus, in the cases where the Fourier transform is bijective we will have the following, without any assumption on the ergodicity of the stochastic process:

$$\mathcal{F}^{-1}(S)(\tau) = \Gamma(\tau) = R(\tau) \quad (12)$$

This is done in particular for Γ and R belonging to the Schwartz space or $L^2(\mathbb{R})$. Thus, if we assume that the signals are fairly regular, we can assume that they are ergodic.

REFERENCES

[1] Y. Kang, E. Spiliotis, F. Petropoulos, N. Athiniotis, F. Li, and V. Assimakopoulos, "Déjà vu: A data-centric forecasting approach through time series cross-similarity," *J. Bus. Res.*, vol. 132, pp. 719–731, Aug. 2021.

[2] T. H. Stewart and T. K. Tolle, *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond, WA, USA: Microsoft Research, 2009.

[3] L. Breiman, "Statistical modeling: The two cultures," *Statist. Sci.*, vol. 16, no. 3, pp. 199–231, 2001.

[4] N. Silver, *The Signal and the Noise: Why So Many Predictions Fail-But Some Don't*. New York, NY, USA: Penguin Press, 2015, p. 560.

[5] A. Hofstadter, "Explanation and necessity," *Philosophy Phenomenol. Res.*, vol. 11, no. 3, pp. 344–347, 1951.

[6] H. Longino, *The Fate of Knowledge*. Princeton, NJ, USA: Princeton Univ. Press, 2002.

[7] W. Salmon, "Why ask, why? An inquiry concerning scientific explanation," in *Proc. Addresses Amer. Philos. Assoc.*, vol. 132. Dordrecht, The Netherlands: Springer, 1978, pp. 683–705, doi: 10.1007/978-94-009-9404-1_15.

[8] H. E. Douglas, "Reintroducing prediction to explanation," *Philosophy Sci.*, vol. 76, no. 4, pp. 444–463, Oct. 2009.

[9] G. Leconte, "L'Activité prédictives des sciences empirique: Analyse d'un succès scientifique et de sa portée," Ph.D. dissertation, Philosophie, Univ. Panthéon-Sorbonne, Paris, France, 2017. [Online]. Available: https://tel.archives-ouvertes.fr/tel-01936647/document

[10] G. Leconte. (2017). *Prédiction Scientifique. Version Académique, Dans M Kristanek (dir.), l'Encyclopédie philosophique.* [Online]. Available: http://encyclo-philo.fr/predictions-scientifiques-a/

[11] C. G. Hempel and P. Oppenheim, "Studies in the logic of explanation," *Philosophy Sci.*, vol. 15, no. 2, pp. 135–175, 1948.

[12] C. G. Hempel, *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York, NY, USA: Free Press, 1965.

[13] W. Salmon, *Scientific Explanation and the Causal Structure of the World*. Princeton, NJ, USA: Princeton Univ. Press, 1984.

[14] M. Friedman, "Explanation and scientific understanding," *J. Philosophy*, vol. 71, no. 1, pp. 5–19, 1974.

[15] N. Rescher, "On explanation and prediction," *Brit. J. Philosophy Sci.*, vol. 8, no. 32, pp. 281–290, Feb. 1958.

[16] I. Scheffler, "Explanation, prediction, and abstraction," *Brit. J. Philosophy Sci.*, vol. 7, no. 28, pp. 293–309, 1957.

[17] H. N. Russell, "On the symmetry between explanation and prediction," *Phil. Rev.*, vol. 68, no. 3, pp. 349–358, 1959.

[18] M. Forster and E. Sober, "How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions," *Brit. J. Philosophy Sci.*, vol. 45, no. 1, pp. 1–35, Mar. 1994.

[19] G. Leconte, "Le problème des prédictions dans les sciences expérimentales," *Philonsorbonne*, vol. 7, pp. 81–99, 2013, doi: 10.4000/philonsorbonne.495.

[20] A. Musgrave, "Logical versus historical theories of confirmation," *Brit. J. Philosophy Sci.*, vol. 25, no. 1, pp. 1–23, 1974.

[21] J. Worral, "Structural realism: The best of both worlds?" *Dialectica*, vol. 43, nos. 1–2, pp. 99–124, 1989.

[22] C. Hitchcock and E. Sober, "Prediction versus accommodation and the risk of overfitting," *Brit. J. Philosophy Sci.*, vol. 55, no. 1, pp. 1–34, Mar. 2004.

[23] J. Worral, "Prediction and accommodation revisited," *Stud. Hist. Philosophy Sci. A*, vol. 45, pp. 54–61, Mar. 2014.

[24] P. Chakravorty, "What is a signal? [lecture notes]," *IEEE Signal Process. Mag.*, vol. 35, no. 5, pp. 175–177, Sep. 2018.

[25] R. M. Haralick, "Statistical and structural approaches to texture," *Proc. IEEE*, vol. 67, no. 5, pp. 786–804, May 1979.

[26] Z. Wang and A. C. Bovik, "Modern image quality assessment," *Synth. Lect. Image, Video, Multimedia Process.*, vol. 2, no. 1, pp. 1–156, Jan. 2006.

[27] C. Charrier and T. Eude, "A psychovisual color image quality metric integrating both intra and inter channel masking effect," in *Proc. 12th Eur. Signal Process. Conf.*, 2004, pp. 2123–2126.

[28] Z. Xu and J. Sun, "Image inpainting by patch propagation using patch sparsity," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1153–1165, May 2010.

[29] M. Bertalmio, A. L. Bertozzi, and G. Sapiro, "Navier-strokes, fluid dynamics, and image and video inpainting," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Dec. 2001, pp. 417–424.

[30] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proc. SIGGRAPH*, 2000, pp. 417–424.

[31] L. M. Candanedo and V. Feldheim, "Accurate occupancy detection of an office room from light, temperature, humidity and CO_2 measurements using statistical learning models," *Energy Buildings*, vol. 112, pp. 28–39, Jan. 2016.

[32] Z. Wang and T. Oates, "Imaging time-series to improve classification and imputation," in *Proc. 24th Int. Joint Conf. Artif. Intell.*, 2015, pp. 1–7.

[33] W. Chen and K. Shi, "A deep learning framework for time series classification using relative position matrix and convolutional neural network," *Neurocomputing*, vol. 359, pp. 384–394, Sep. 2019.

[34] L. Li, K. Ota, and M. Dong, "Everything is image: CNN-based short-term electrical load forecasting for smart grid," in *Proc. 14th Int. Symp. Pervas. Syst., Algorithms Netw., 11th Int. Conf. Frontier Comput. Sci. Technol., 3rd Int. Symp. Creative Comput. (ISPAN-FCST-ISCC)*, Jun. 2017, pp. 344–351.

[35] L. Xixi, K. Yanfei, and L. Feng, "Forecasting with time series imaging," *Expert Syst. Appl.*, vol. 160, Dec. 2020, Art. no. 113680.

[36] M. Derraz, "The general sinus," *Int. J. Swarm Intell. Evol. Comput.*, vol. 6, no. 2, pp. 1–10, 2017.

[37] *M3 Competition*. Accessed: Jun. 28, 2021. [Online]. Available: https://forecasters.org/resources/time-series-data/m3-competition/

[38] S. Makridakis and M. Hibon, "The M3-competition: Results, conclusions and implications," *Int. J. Forecasting*, vol. 16, no. 4, pp. 451–476, Oct. 2000.

[39] A. Elmouatamid, R. Ouladsine, M. Bakhouya, N. El Kamoun, K. Zine-Dine, and M. Khaidar, "MAPCAST: An adaptive control approach using predictive analytics for energy balance in micro-grid systems," *Int. J. Renew. Energy Res.*, vol. 10, no. 2, pp. 945–954, 2020.

[40] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "The M4 competition: 100,000 time series and 61 forecasting methods," *Int. J. Forecasting*, vol. 36, no. 1, pp. 54–74, Jan. 2020.

[41] S. Smyl, "A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting," *Int. J. Forecasting*, vol. 36, pp. 75–85, Jan. 2020.

[42] S. Fiore, M. Bakhouya, and W. W. Smari, "On the road to exascale: Advances in high performance computing and simulations—An overview and editorial," *Future Gener. Comput. Syst.*, vol. 82, pp. 450–458, May 2018.

[43] A. Kharbouch, Y. Naitmalek, H. Elkhoukhi, M. Bakhouya, V. De Florio, M. D. El Ouadghiri, S. Latre, and C. Blondia, "IoT and big data technologies for monitoring and processing real-time healthcare data," *Int. J. Distrib. Syst. Technol.*, vol. 10, no. 4, pp. 17–30, Oct. 2019.

[44] Y. N. Malek, M. Najib, M. Bakhouya, and M. Essaaidi, "Multivariate deep learning approach for electric vehicle speed forecasting," *Big Data Mining Analytics*, vol. 4, no. 1, pp. 56–64, Mar. 2021.

[45] M. Majidpour, H. Nazaripouya, P. Chu, H. Pota, and R. Gadh, "Fast univariate time series prediction of solar power for real-time control of energy storage system," *Forecasting*, vol. 1, no. 1, pp. 107–120, Sep. 2018, doi: 10.3390/forecast1010008.

[46] J. Durbin and S. Jan Koopman, *Time Series Analysis by State Space Methods*, vol. 38, 2nd ed. New York, NY, USA: Oxford, 2012, doi: 10.1093/acprof:oso/9780199641178.001.0001.

[47] X. Ping, Q. Chen, G. Liu, J. Su, and F. Ma, "Particle filter based time series prediction of daily sales of an online retailer," in *Proc. 11th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Oct. 2018, pp. 1–6.

[48] M. A. Pinsky, "Time-series forecast with adaptive feedback controlled predictor," *Universal J. Comput. Math.*, vol. 1, no. 2, pp. 47–55, Aug. 2013.

[49] S. S. Rangapuram, M. Seeger, Y. Wang, J. Gasthaus, L. Stella, and T. Januschowski, "Deep state space models for time series forecasting," in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst. (NIPS)*, Dec. 2018, pp. 7796–7805.

[50] R. Hyndman, A. B. Koehler, J. K. Ord, and R. D. Snyder, *Forecasting With Exponential Smoothing: The State Space Approach* (Springer Series in Statistics). Berlin, Germany: Springer-Verlag, 2008.

[51] T. Fischer, C. Krauss, and A. Treichel, "Machine learning for time series forecasting—A simulation study," Inst. Econ., Friedrich-Alexander Univ. Erlangen-Nuremberg, Erlangen, Germany, Tech. Rep. 02/2018, 2018.

[52] G. Liu, F. Xiao, C.-T. Lin, and Z. Cao, "A fuzzy interval time series energy and financial forecasting model using network-based multiple time-frequency spaces and the induced ordered weighted averaging aggregation operation," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 11, pp. 2677–2690, Nov. 2020, doi: 10.1109/TFUZZ.2020.2972823.

[53] R.-V. Donner, M. Small, J.-F. Donges, N. Marwan, Y. Zou, R.-X. Xiang, and J. Kurths, "Recurrence-based time series analysis by means of complex network methods," *Int. J. Bifurcation Chaos*, vol. 21, no. 4, pp. 1019–1046, 2011, doi: 10.1142/S0218127411029021.

[54] P. Xu, R. Zhang, and Y. Deng, "A novel visibility graph transformation of time series into weighted networks," *Chaos, Solitons Fractals*, vol. 117, pp. 201–208, Dec. 2018.

[55] S. Mao and F. Xiao, "A novel method for forecasting construction cost index based on complex network," *Phys. A, Stat. Mech. Appl.*, vol. 527, Aug. 2019, Art. no. 121306.

[56] C.-L. Yang, Z.-X. Chen, and C.-Y. Yang, "Sensor classification using convolutional neural network by encoding multivariate time series as two-dimensional colored images," *Sensors*, vol. 20, no. 1, p. 168, Dec. 2019.

[57] Y. Cao, C. Wang, Z. Li, L. Zhang, and L. Zhang, "Spatial-bag-of-features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 3352–3359, doi: 10.1109/CVPR.2010.5540021.

[58] W. C. Salmon and P. Kitcher, Eds., *Wesley Salmon Four Decades of Scientific Explanation*. Minneapolis, MN, USA: Univ. Minnesota Press, 1989.

[59] P. Kitcher and W. C. Salmon, *Scientific Explanation* (Minnesota Studies in the Philosophy of Science), vol. 13. Minneapolis, MN, USA: Univ. Minnesota Press., 1989.

[60] J. D. Trout, "Scientific explanation and the sense of understanding," *Philosophy Sci.*, vol. 69, pp. 212–233, Jun. 2002.

[61] O. Biran and C. Cotton, "Explanation and justification in machine learning: A survey," in *Proc. IJCAI Workshop Explainable Artif. Intell. (XAI)*, 2017, pp. 8–13.

[62] S. de Haro, D. Dieks, G. Hooft, and E. Verlinde, "Forty years of string theory reflecting on the foundations," *Found. Phys.*, vol. 43, no. 1, pp. 1–7, Jan. 2013.

[63] S. De Haro, "The heuristic function of duality," *Synthese*, vol. 196, no. 12, pp. 5169–5203, Dec. 2019.

[64] J. Maldacena, "The large N limit of superconformal field theories and supergravity," *Adv. Theor. Math. Phys.*, vol. 2, no. 4, pp. 231–252, 1998.

[65] M. Natsuume, *AdS/CFT Duality User Guide* (Lecture Notes in Physics), vol. 903. Japan: Springer, 2015, doi: 10.1007/978-4-431-55441-7.

[66] P. K. Kovtun, D. T. Son, and A. O. Starinets, "Viscosity in strongly interacting quantum field theories from black hole physics," *Phys. Rev. Lett.*, vol. 94, no. 11, Mar. 2005, Art. no. 111601.

[67] N. Chaumat, C. Etasse, J. Hauchemaille, B. Jech, M. Lobry, B. Maheu, and D. Thorez, *L'Analogie en Physique, Tome I: Exemple Historique*. Rouen, France: IREM de Rouen, 1989.

[68] M. Paty, "Les analogies mathématiques au sens de poincare et leur fonction en physique," Arch. Ouverte Sci. l'Homme la Société, 2007. [Online]. Available: https://halshs.archives-ouvertes.fr/halshs-00167540

[69] E. Bolmont, "Evolution des Analogies. L'Exemple de l'Analogie Chaleur-electricite," in *Philosophia Scientiae*, vol. 7, no. 2, pp. 115–146, 2003.

[70] H. Putnam, *What is Mathematical Truth?* Cambridge, U.K.: Cambridge Univ. Press, 1979, pp. 60–78.

[71] J. Worrall, "Structural realism: The best of both worlds?" *Dialectica*, vol. 43, no. 2, pp. 99–124, 1989.

[72] H. Poincaré, *Science and Hypothesis*. Paris, France: Flammarion, 1902.

[73] *M4 Competition*. Accessed: Jun. 28, 2021. [Online]. Available: https://github.com/Mcompetitions/M4-methods

[74] T. Palmer, "The ECMWF ensemble prediction system: Looking back (more than) 25 years and projecting forward 25 years," *Quart. J. Roy. Meteorol. Soc.*, vol. 145, no. S1, pp. 12–24, Sep. 2019, doi: 10.1002/qj.3383.

[75] L. K. Hansen and P. Salamon, "Neural network ensembles," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 10, pp. 993–1001, Oct. 1990.

[76] U. Naftaly, N. Intrator, and D. Horn, "Optimal ensemble averaging of neural networks," *Netw. Comput. Neural Syst.*, vol. 8, no. 3, pp. 283–296, 1997.

[77] J. D. Wichard and M. Ogorzałek, "Time series prediction with ensemble models applied to the CATS benchmark," *Neurocomputing*, vol. 70, nos. 13–15, pp. 2371–2378, Aug. 2007.

[78] Y. Xiao, J. Xiao, and S. Wang, "A hybrid model for time series forecasting," *Human Syst. Manage.*, vol. 31, no. 2, pp. 133–143, Aug. 2012.

[79] A. Galicia, R. Talavera-Llames, A. Troncoso, I. Koprinska, and F. Martínez-Álvarez, "Multi-step forecasting for big data time series based on ensemble learning," *Knowl.-Based Syst.*, vol. 163, pp. 830–841, Jan. 2019.

[80] B. Schneider, D. Jackle, F. Stoffel, A. Diehl, J. Fuchs, and D. Keim, "Integrating data and model space in ensemble learning by visual analytics," *IEEE Trans. Big Data*, vol. 7, no. 3, pp. 483–496, Jul. 2021, doi: 10.1109/TBDATA.2018.2877350.

[81] M. Ponce-Flores, J. Frausto-Solis, G. Santamaria-Bonfil, J. Perez-Ortega, and J. J. Gonzalez-Barbosa, "Time series complexities and their relationship to forecasting performance," *Entropy*, vol. 22, no. 1, pp. 1–18, 2020.

[82] X. Wang, K. Smith, and R. Hyndman, "Characteristic-based clustering for time series data," *Data Mining Knowl. Discovery*, vol. 13, no. 3, pp. 335–364, 2006, doi: 10.1007/s10618-005-0039-x.

[83] J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, and A. Bouchachia, "A survey on concept drift adaptation," *ACM Comput. Surv.*, vol. 46, no. 4, Apr. 2014, Art. no. 44, doi: 10.1145/2523813.

NADIR MAAROUFI received the B.S., M.S., and Ph.D. degrees in mathematics from University of Picardie Jules Verne, France, in 2004, 2007, and 2010, respectively. He is currently an Associate Professor of mathematics with the International University of Rabat. He is the author/coauthor of about 20 articles in refereed journals and conferences in the areas of fundamental and applied mathematics. His interests include dynamical systems in finite and infinite dimension, partial differential equations, functional analysis, complexity, ergodic theory, probability, statistics, and mainly dimension theory. He is also concerned with philosophy of mathematics and science.

MEHDI NAJIB received the M.S. degree from the Faculty of Sciences Semlalia, Marrakech, Morocco, in 2009, and the Ph.D. degree in computer science from the University of Le Havre, France, in 2014. He is currently an Assistant Professor of computer science with the International University of Rabat. His research interests include AI and prediction, smart buildings, adaptive systems for workflow management, decision support systems, simulation, and risk management. He has published more than 20 papers in international conferences, journals and books, and participated to many research projects.

MOHAMED BAKHOUYA received the Ph.D. degree from UTBM, France, in 2005, and the H.D.R. degree from UHA, France, in 2013. He is currently a Professor of computer science with the International University of Rabat. He has more than ten years experiences in participating and working in sponsored ICT projects. He has published more than 100 papers in international journals, books, and conferences. His research interests include various aspects related to the design, validation, and implementation of distributed and adaptive systems, architectures, and protocols. He was an EiC of *IJARAS* journal and serves as a Guest Editor for a number of international journals, such as *ACM Transactions on Autonomous and Adaptive Systems*, *International Journal of Product Development*, *Concurrency and Computation: Practice and Experience*, *FGCS*, and *MICRO*.

● ● ●