Knowledge-Based Systems 293 (2024) 111639

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Echo state network and classical statistical techniques for time series forecasting: A review[☆]

Fabian Corrêa Cardoso [a,b,*], Rafael Alceste Berri [c], Eduardo Nunes Borges [c], Bruno Lopes Dalmazo [c], Giancarlo Lucca [d], Viviane Leite Dias de Mattos [e]

[a] Software Engineering Faculty, Rio Verde University, Rio Verde, GO, Brazil
[b] Graduate Program in Computational Modeling, Federal University of Rio Grande, Rio Grande, RS, Brazil
[c] Computer Science Center, Federal University of Rio Grande, Rio Grande, RS, Brazil
[d] Master Program in Electronic and Computing Engineering, Catholic University of Pelotas, Pelotas, RS, Brazil
[e] Mathematics, Statistics, and Physics Institute, Federal University of Rio Grande, Rio Grande, RS, Brazil

ARTICLE INFO

Keywords:
Systematic literature review
Reservoir computing
ESN
ARIMA–GARCH
Hybridization

ABSTRACT

Forecasting is an extensive field of study, which tries to avoid injuries, diseases, and damages but also can help in energy production, finance investments, etc. Two mathematics modeling techniques have obtained promising results: the ones based on Machine Learning (Echo State Network) and based on Statistical techniques (ARIMA/GARCH). To take advantage of both techniques, we aimed to perform a systematic literature review of Echo State Network and classical Statistical techniques for forecasting Time Series. We conducted the searches on the databases ACM, IEEE Xplore, Scopus, and Web of Science and, after, did a bibliometric and a content qualitative analysis of the selected articles. We present the techniques and sources of the data set used, the most used keywords in the articles, analyze the reservoir computing/echo state network and statistical techniques, and comment on each article selected. From the analysis of this review, it is possible to infer that it is still an area to be studied more deeply and that the academy, even if timidly, never stopped using the echo state network for time series regression in general and financial series.

1. Introduction

Systems that evolve in time or space according to probabilistic laws are called stochastic processes [2], including Time Series (TS), whose comprehension is useful for decision-making that depends on estimates of future values. According to [3], a TS is a set of observations ordered in time. By analyzing them, it is possible to find the mechanism that generates the data and make it possible to create a model and use it to make short-, medium- or long-term forecasts.

There are some ways of modeling to make short-term forecasts, either by physical equations, such as wind speed forecasting [4], the Lorenz system for weather forecasting [5], by Machine Learning (ML) techniques: Recurrent Neural Network (RNN) [6], Echo State Networks (ESN) [7], by Statistical techniques (Auto-regressive Integrated Moving Average (ARIMA) [8], and Generalized Conditional Heteroscedastic Auto-regressive (GARCH) [9]. TS modeling for forecasting is an extensive field of study, which tries to avoid injuries, diseases, and damages but also can help in energy production, weather forecasting, etc.

In [10], formally is presented ESN, proposing the implementation of a new RNN using Reservoir Computing (RC). Like the RNN, the ESN has a memory that can be short/long-term. In [11] a neural network equivalent to the ESN,[1] the Liquid State Machine (LSM), was proposed. Both are types of RNNs and use RC, according to [12]. An ESN is a particular class of RNN that uses RC and can produce forecast [13]. It uses random weights in the input and in the Reservoir, which are not changed or trained, and behaves as a complex nonlinear dynamic filter that transforms the input signals [14].

ARIMA is one of the most widely used linear models for TS forecasting, denoted as ARIMA(p, d, q), where p is the number of auto-regressive items, d is the number of differences used in TS to make it stationary, and q is the number of moving average terms. It can be determined by analyzing the Auto-correlation Function (ACF) and Partial Auto-correlation Function (PACF) and also by calculating the Akaike Information Criterion (AIC) or Bayesian Information Criterion

☆ Partially funded by Fesurv-University of Rio Verde, Federal University of Rio Grande, Capes-Brazil.
* Corresponding author at: Software Engineering Faculty, Rio Verde University, Rio Verde, GO, Brazil.
E-mail address: fabian@unirv.edu.br (F.C. Cardoso).
[1] For more practical details on ESN see [1].

https://doi.org/10.1016/j.knosys.2024.111639
Received 28 August 2023; Received in revised form 21 February 2024; Accepted 12 March 2024
Available online 26 March 2024

(BIC) [15]. The ARIMA statistical modeling for forecast uses Box–Jenkins to do forecast and is a linear TS model that assumes it is free of periodic fluctuations [16].

GARCH is a stochastic volatility model used generally to check the presence of volatility clustering in TS [17]. According to [18], it "describes the error variance as a function of the actual sizes of the error terms of the previous periods". The GARCH statistical modeling technique for the forecast is a nonlinear model that uses present and past observations and conditional variance to forecast TS [17]. We call attention to the fact that probabilistic laws control the Statistical models used in TS modeling.

Now, we need to do a differentiation, the dynamical systems are divided into deterministic and stochastic [19,20]. Weather forecasting, e.g., is a chaotic system, that is a kind of deterministic dynamical system [21] and in the beginning it was modeled by the aforementioned physical equations. TS, as mentioned in the beginning, are used to model stochastic dynamic systems. But, later, weather forecasting was also done data driven using TS [22]. ESN has different performance if the data are produced by a deterministic dynamical system or by a stochastic (or random) process. If ESN adapts to the solution of deterministic dynamical systems [23], does it also adapt to model stochastic dynamical systems with TS?

A sector of the economy that can highly benefit from the methods and techniques mentioned above for decision making is the market for financial assets or stocks. Stocks are representative securities of company ownership by stockholders [24], and can be negotiated on a Stock Exchange (SE). Stocks are also a dynamical system, but stochastic [25–27]. In this case, is it usual to apply ESN to financial TS for the modeling and the forecasting? If not, is not it being underutilized?

Another possibility for making TS forecasts is combining ML with Statistical techniques, e.g., a nonlinear deterministic model with a linear stochastic one. It is a way to potentialize one of the techniques and achieve better performance. The so-called hybridization forecasting technique consists of one modeling the linear pattern and the other modeling the nonlinear to produce a better forecast [28], although the hybridization is uncommon for forecasting financial TS, specifically stock closing prices. These techniques use data driven forecast. And we can suppose that the amount of data necessary is different for each one and for the hybridized. So, is it interesting to combine ESN, which models very well chaotic systems, with a Statistical technique, which models very well stochastic systems? Are there studies which explore the hybridization of ESN with Statistical techniques for doing more accurate financial (not only) TS forecasts?

Our interest in researching ESN and Statistical techniques is to contribute in this sense and, in addition, to know if they can be used in conjunction (hybridizing it), in order to obtain better performance. Although, the aim of this study is to perform a systematic literature review of ESN and classical Statistical techniques for forecasting TS.

The study's remainder organization is as follows: Section 2 presents a brief literature review on ESN and Statistical models for forecasting, Section 3 presents the systematic review methodology, Section 4 points out and discusses the results obtained, and then, Section 5 describes the conclusions of the research conducted.

2. Techniques for time series forecasting

The technique of modeling TS for forecasting in Statistics is known as regression, a well-stated field of research with different applications [29]. Moreover, it is possible to highlight recent techniques dealing with this kind of problem, such as Computational techniques. We highlight ML, specifically the ones that replicate regression by keeping a memory, as the RNNs [6], and as a consequence ESN.

An autoregressive process of moving averages of orders p and q, denoted by ARMA (p, q), is defined by [3,30–32]:

$$Y_t - \mu = \sum_{i=1}^{p} \phi_i (Y_{t-i} - \mu) + \sum_{j=0}^{q} \theta_j \epsilon_{t-j} + \epsilon_t, \tag{1}$$

where Y_t is the t-th observed value ($t = 1, 2, \ldots, T$), T is the number of observed values, μ is the process average, i is the lag of the autoregressive process ($i = 1, 2, \ldots, p$), j is the lag of the moving average process ($j = 1, 2, \ldots, q$), ϕ_i is the coefficient of the auto-regressive portion at the ith lag, θ_j is the moving average coefficient at the jth lag, p is the order of the autoregressive process, q is the order of the moving average process and $\epsilon_t RB(0; \sigma^2)$ is the error at the t-th time and possible changes that may be occurring in the behavior of the data.

Attention is drawn to the fact that the modeling of financial TS is normally carried out using the logarithm of the rate of return on the price of the financial stock considered (Y_t in Eq. (1)).

The methodology in question must then be used in stationary and ergodic linear processes. A stochastic process is called weakly stationary or second-order stationary if its mean and variance do not vary over time ($E(Y_t) = \mu \forall t \in Z$ and $E|Y_t^2| < \infty$), and the value of the covariance between two periods depends only on the degree of lag between observations and not on the effective period of time. where covariance is calculated ($cov(Y_t, Y_{t-j}) = \gamma_j \forall j$) [32].

A stochastic process is said to be ergodic if: (a) $\overline{Y} \underrightarrow{p} E(Y_t)$ and (b) $\hat{\gamma}_j \underrightarrow{p} \gamma_j$. This means that the sample moments must converge in probability to the population moments. According to [32], this property allows a single realization of the process to be used to estimate unknown population parameters. According to [30], this property is verified in a stochastic process for the first moment if the sum of the covariance is finite, i.e., $\sum_{j=0}^{\infty} |\gamma_j| < \infty$. [33] draws attention to the fact that, in practice, stationarity and ergodicity have the same requirements.

Although this type of modeling must be used in stationary linear processes, it is also possible to use it in linear processes that can be transformed into stationary, as is the case with homogeneous non-stationary linear processes. These can easily be transformed into stationary by determining successive differences. In this case, considering the delay operator B ($B^k Y_t = Y_{t-k}$), to represent an ARIMA model (p,d,q), Eq. (1) could be rewritten as:

$$\phi_p(B) \Delta^d Y_t = \theta_q(B) \epsilon_t, \tag{2}$$

where $\phi_p(.)$ and $\theta_q(.)$ are polynomials of degrees p and q, respectively, with B as a variable, Δ representing the difference between consecutive values of the analyzed series and d, the number of differences between consecutive values made.

According to [31], a striking feature of financial series is that, in general, they are not serially correlated but dependent. This property makes nonlinear models able to describe their behavior better than linear ones. Therefore, when there is some autocorrelation, it is usual to use hybrid models. In this case, a linear model removes the autocorrelation, like ARIMA, after a nonlinear model application is made to forecast, like GARCH.

In statistical modeling of financial TS, the Generalized Autoregressive model with Conditional Heteroskedasticity (GARCH), proposed by [34,35], has been successfully used to model this nonlinear part, more specifically volatility. This model is nothing more than a generalization of the autoregressive model with conditional Heteroscedasticity (ARCH), proposed by [36], to estimate the inflation variance. According to [31], the GARCH (m,n) model is defined as:

$$a_t = \sqrt{h_t}.\epsilon_t, \tag{3}$$

$$h_t = \alpha_0 + \sum_{i=1}^{m} \alpha_i . a_{t-i}^2 + \sum_{j=1}^{n} \beta_j . h_{t-j}, \tag{4}$$

where α_t is the innovation (ϵ_t of Eq. (1), h_t is the conditional variance of returns and ϵ_t is the error at the t-th time that quite frequently follows the Student t distribution.

ARMA-GARCH model quadratic values, which causes large residuals to have a large impact on the model. Furthermore, they also treat positive and negative returns in the same way, although there is a consensus that negative returns impact volatility more. Aiming to eliminate some of its limitations, some extensions have already been proposed, such as: exponential GARCH (EGARCH), GARCH with threshold (TGARCH) and power GARCH (PGARCH), among others.

The main advantage of statistical methods and techniques used in the TS analysis, is that TS is considered a process that evolves over time according to probabilistic laws. It means that each observation is a random variable where its behavior described by a probability distribution. Therefore, the data sequence of the analyzed TS is one of several possible sequences generated by the mechanism that created them, called a stochastic process, i.e., it is a sample. The parameters of this stochastic process (population) are estimated from the sampling distributions of their estimators, with point estimates complemented with confidence intervals defined using a probability function or probability density function. However, for the confidence intervals to be valid, some theoretical assumptions must be satisfied, which often does not happen, invalidating the results found. Furthermore, the data must also satisfy the conditions of stationarity and ergodicity.

According to [37], $u(t) \in \mathbb{R}^{Dx1}$ denotes the input value at time t of the ST, $W_{in} \in \mathbb{R}^{NxD}$ (D is the input dimension) represents the connection weights between the input layer and the hidden layer, $W \in \mathbb{R}^{NxN}$ (N is the number of neurons) denotes weights of input connections to hidden layer, and $W_{out} \in \mathbb{R}^{LxN}$(L is the output dimension) are the output weights from the hidden layer to the output layer. The state transition equation can be expressed by Eqs. (5) and (6) :

$$x(t) = (1-\alpha)x(t-1) + \alpha.tanh(W_{in}u(t) + Wx(t-1)), \tag{5}$$

$$y(t) = W_{out}x(t), \tag{6}$$

where $u(t), x(t) \in \mathbb{R}^{Nx1} and\, y(t)$ refer to the input, reservoir state and output, respectively. α denotes the leaking rate, which integrates the previous state with the current one, and $tanh()$ denotes the activation function in the reservoir.

Still according to [37], the main hyperparameters of the ESN are the input scale (IS), the spectral radius (sr) and the sparsity (s):

1. IS is used to initialize the random weights of the W_{in} matrix, and the weights are in the range between -IS and IS;
2. sr is the spectral radius of W, given by:

$$W = sr.\frac{W}{\lambda_{max}(W)}, \tag{7}$$

   where $\lambda_{max}(W)$ is the largest eigenvalue of the matrix W and elements of W are randomly generated between −0.5 and 0.5, for example;
3. s denotes the portion of non-zero weights in W.

Hybridization can be a combination of ML and Statistical techniques. In this case, both are mixed/merged with different approaches to help one to perform better. It is possible in a hybrid system to decompose the TS into low and high-volatility components, where each technique can handle separate pieces [38] to take advantage of the strength of each one. As [28] explained, “combining different models can increase the chance of capturing different patterns in the data and improve forecasting performance”. In other hand, [39] say that “hybrid models should contain a limited number of individual models to maintain model simplicity as well as accuracy”.

Any technique alone can be a universal forecasting model, be ML or Statistical, so when we hybridize techniques, we hope the hybrid techniques could at least generate much more forecasting models than each technique alone.

Many studies already developed show the superiority of the accuracy of the predictions obtained with hybrid models, regardless of the techniques or methods used in the hybridization, such as in [40–42], and others.

For better understanding, we suggest reading papers that have already studied the hybridization of Arima with various ML methods: [43], [44], [45], and [38].

3. Material and methods

To start a systematic review, it is necessary to define the research questions, according to the objectives of the work to be carried out [46]. Therefore, the following research questions were defined:

1. **Research Questions 1 (RQ1).** What are the existing RC models? What are the common terminologies to describe them? What are their components? How are these components related? Does any of the models appropriately define everyone else for any application? If it does not properly define, can we combine the knowledge of RC models and systems to create a unified model that can define all?
2. **Research Questions 2 (RQ2).** Can RC perform predictions? If so, can it be applied to financial TS? How to improve these predictions? How accurate are these predictions/how to minimize errors?

The searches were carried out in the databases Scopus,[2] ACM Digital Library,[3] Web of Science,[4] and IEEE Xplore database,[5] with the initial objective of searching for related terms such as Echo State and Reservoir Computing, covering the object of study of this specific neural network, and, as an additional line of research, we searched for terms related to the scope of the work. We also used the author’s keywords verification to do a more precise search, looking for the author’s keywords in the articles to verify if they correspond to the words used in the query.

Since RC and ESN are relatively new technologies, initially it was chosen to focus the query on these expressions and merge them with Statistical and financial expressions. In addition, the term “Machine Learning” was included, because ESN can be considered a type of Machine Learning, and also the term “Data Set”, because we were looking for papers that could be replicated, for a better understanding of ESN.

Table 1 shows the results obtained in each database after the queries with the search field “All Fields”. The first query was very broad and returned so many articles. Then a second query was created for “All Fields”, separating the Statistical and the financial expressions, keeping “Machine Learning” and “Data Set” with the latter. As the term ESN was returning articles related to Psychology, we decided to remove it, and also to eliminate the expression “Machine Learning” for not being exactly in the focus of the work. So, a third query was adopted, adding the terms “Survey”, “Review”, and “State-of-the-art”, to verify if the terms had covered all the content. In this case, the query was limited to between 2018 and 2022 (both included), i.e., the state-of-the-art. It returned 1055 articles, where 19 were in duplicity and 8 were books.

After we looked up for author’s keywords in the articles and noticed that terms like “Exchange Market”, “State of the art”, and Volatility had not appeared, but Finance and Market appeared instead, so we changed them. The last query returned 1331 articles, as 1 was in duplicity and 19 were books, leaving 1311 in the end.

As we can see, the database which returned more articles was always Scopus, followed by ACM, IEEE and Web of Science. The last one did not return any study in the final query.

After this, the steps used to select articles were read: titles, abstracts, introductions, and conclusions. In each step, inclusion criteria were used, as described below.

- **Title selection:** the first criteria were looking for titles that dealt with hybrid techniques and, when impossible, that they contained at least words/terms related to the words used in the query, the others were excluded.

[2] Access the Scopus website at http://www.scopus.com/.
[3] Access the ACM Digital Library website at https://dl.acm.org/.
[4] Access the Web of Science website at https://clarivate.com/.
[5] Access the IEEE Xplore website at https://ieeexplore.ieee.org/search/advanced.

Table 1
The relation of the considered queries and the number of obtained results per used database.

Keywords used in search	ACM Digital Library	IEEE Xplore	Scopus	Web of Science	Total
("Reservoir Computing" OR "Echo State" OR ESN) AND ("Machine Learning" OR Forecast OR Prediction OR Regression OR GARCH OR ARIMA OR "Time Series" OR Stock OR "Exchange Market" OR Volatility OR "Data Set")	1448	1456	10,186	1537	14,627
("Echo State" OR "Reservoir Computing" OR ESN) AND (Prediction OR Forecast OR "Time Series" OR GARCH OR ARIMA OR Regression) AND (Stock OR "Exchange Market" OR Volatility OR "Machine Learning" OR "Data Set")	176	151	3960	294	4581
(Survey OR Review OR "State of the art") AND ("Reservoir Computing" OR "Echo State") AND (Prediction OR Forecast OR Regression OR GARCH OR ARIMA OR "Time Series") AND ("Stock Market" OR "Exchange Market" OR Volatility OR "Data Set")	174	11	866	4	1055
(Survey OR Review) AND ("Reservoir Computing" OR "Echo State") AND (Prediction OR Forecast OR Regression OR GARCH OR ARIMA OR "Time Series") AND (Stock OR Finance OR Market OR "Data Set")	18	1	1312	0	1331

Table 2
The main characteristics of selected articles.

Articles	Sources use financial TS?	Proposes new technology related to RC/ESN?	Compares RC/ESN with Statistical techniques?	Compares other NNs with Statistical techniques?	Presents RC/ESN results analyzed statistically?
[47]	No	Yes	No	No	Yes
[48]	No	Yes	No	No	No
[49]	No	No	No	No	No
[50]	No	No	Yes	No	No
[51]	No	No	No	No	No
[52]	No	Yes	No	No	No
[53]	No	Yes	No	No	No
[54]	Yes	Yes	No	No	No
[55]	No	No	No	No	No
[56]	No	Yes	No	No	No
[57]	No	No	No	No	No
[58]	Yes	No	No	No	No
[59]	Yes	No	No	No	No
[60]	No	No	Yes	No	No
[61]	Yes	Yes	No	No	No
[62]	No	No	No	No	No
[38]	Yes	Yes	No	Yes	No
[7]	No	Yes	No	No	No
[63]	No	No	No	No	No
[64]	Yes	No	Yes	No	No
[65]	No	No	No	No	No
[66]	No	Yes	No	No	Yes
[37]	No	No	No	No	No

- **Abstract selection:** this criteria was "explains the theoretical foundations of Reservoir or Echo State Network Computation"; "describes the operation of Reservoir or Echo State Network Computation"; "focuses on software architecture/frameworks of Reservoir or Echo State Network Computation"; "describes applications of Reservoir or Echo State Network Computation in Financial Time Series", and must be related to at least one aspect of the research questions. It should be noticed that, to be included in the review, the paper must confirm at least one of the questions.
- **Introduction and Conclusion selection:** the criteria during the reading of the introductions and conclusions were "the objectives are defined clearly"; "the architecture of the system/framework/ project/ experiment is defined clearly"; "the paper presents the source of the dataset used"; "the study proposes a new technology related to Reservoir Computing or Echo State Network"; "the study presents the state-of-the-art or ideas for future work on Reservoir Computing or Echo State Network"; "the study compares Reservoir Computing or Echo State Network to Statistical techniques"; "the study presents results of Reservoir Computing or Echo State Network analyzed statistically"; "the study presents results of other ML techniques compared to Statistical techniques"; "the study uses Finance Time Series";

We read the 1311 articles' titles and, based on the exclusion criteria, reduced the number of articles to 170. Then, the abstracts of each of these were read, and, based on the inclusion criteria, we reduced the number to 86. However, of these 86 articles, 11 were not possible to be obtained in full because they were unavailable on the Internet or required payment for access. So, we read 75 introductions and conclusions and adopted the inclusion criteria described, resulting in 23 articles to be analyzed.

The JabRef[6] [67] was used to organize the bibliographic references. The biblioshiny command of the Bibliometrix[7] library [68] in RStudio[8] [69] was used for analysis and presentation of the bibliometric results.

4. Discussions

In this section, for each selected study, we present a deeper analysis. To do so, we start with a bibliometric analysis, followed by a meta-synthesis.

[6] Available at https://www.jabref.org.
[7] Available at https://www.bibliometrix.org.
[8] Available at https://www.rstudio.com/products/rstudio/download/.

Table 3
Data source, preprocessing and errors used.

Articles	Sources of data sets used	Data Preprocessing	Errors
[47]	Mackey-Glass, Laser intensity, and Spoken Arabic Digits	Laser intensity TS: feature scaling of normalization with mean zero and standard deviation of one filtered with a Gaussian of length three and standard deviation of one; Spoken Arabic Digit Recognition: feature scaling of normalization with mean zero and standard deviation of one, computed the missing values by interpolation	NRMSE (Normalized Root Mean Squared Error)
[48]	Sequence of digital bits, and Lorenz63	N.A. (not applies or not uses)	NRMSE
[49]	Positions of a patient in a bedroom by readings of an accelerometer, Types of movement in LIBRAS (Linguagem Brasileira de Sinais), RSSI data set of sensors scattered in the first floor of Waldo Library at Western Michigan University, Robot data set of 24 ultrasound sensors, and Ozone data set of atmosphere	N.A.	Friedman test
[50]	Mackey-Glass differential delay equations, Lorenz63	N.A.	N.A.
[51]	”Non-linear chaotic finance model which involves interest rate, investment demand and price index as the underlying moving parameters. (...) described by a dimensionless system of first order coupled linear differential equation”	75% training and 25% validation	MSE (Mean Squared Error)
[52]	Data generated by integrating a simplified model of a weather system developed by Lorenz in 1963, and double-scroll electronic circuit	N.A.	NRMSE
[53]	Lorenz63	N.A.	Correlation dimension calculated from simulated trajectories
[54]	S&P 500, Parking Birmingham, Chaotic data set from the Annulus experiment, Philadelphia temperature, and Bike sharing	TS additive decomposition in trend, seasonality, and residual; S&P500 daily: 80% training and 20% validation; Parking Birmingham: 80% training and 20% validation; Simulated chaotic data set: 80% training and 20% validation; Philadelphia temperature: 80% training and 20% validation; Periodic data: 80% training and 20% validation; Bike sharing: 80% training and 20% validation	MSE
[55]	138 articles	Data cleaning, scaling features, feature generation, feature selection and feature extraction	N.A.
[56]	Lorenz63, Lorenz96, and Ornstein–Uhlenbeck-like process	Feature extraction	RMSE
[57]	Synthetic data sets: Mackey-Glass and NARMA, Real-world data sets: Electromiogram of a healthy subject, and MACHO: an astronomical light curve	Electromyogram: 80% for training and validation, and 20% for test; MACHO: determine the period of RRL using MHAoV method, subsequently bin them into regular intervals, missing values are interpolated quadratically, outliers are filtered out by applying Savitzky–Golay, 80% for training and validation, and 20% for test	NMSE (Normalized MSE)
[58]	M3 Competition	Transforming the data to achieve stationarity in variance; Deseasonalizing data; Detrending data	sMAPE (symmetric Mean Absolute Percentage Error) MASE (Mean Absolute Scaled Error); Computational complexity
[59]	Data of Microsoft Stock from Yahoo Finance	Feature scaling: normalization into 0–1 of closing price, checked for null values, 80% for training and 20% for validation	MSE, RMSE

(*continued on next page*)

Table 3 (*continued*).

Articles	Sources of data sets used	Data Preprocessing	Errors
[60]	52 articles	Methodology: Planning (motivation, objective, and research questions); Development of the study (search strategy, inclusion/exclusion, and mapping); Mapping report (filtering studies and classification process)	N.A.
[61]	Intraday data of Caterpilar, eBay, and Microsoft stocks	Conversion from time series to supervised learning and equations for scaling the data	R^2, MSE, RMSE, MAPE (Mean Absolute Percentual Error)
[62]	Generalized Hénon-map and NARMA	NARMA: N.A.; Hérnon-Map: 40% training and 60% validation, removing 200 first steps from each	NMSE
[38]	Canadian Lynx, Wolf's Sunspot, British Pound/US Dollar Exchange Rate, Colorado River flow, Airline passengers, and Star Brightness	Canadian Lynx: 88% training and 12% validation; Sunspot: 77% training and 23% validation; Exchange rate: 93% training and 7% validation; Colorado river: 80% training and 20% validation; Airline: 80% training and 20% validation; Star: 80% training and 20% validation	MAE (Mean Absolute Error), MSE, MAPE, Percentage comparison
[7]	Rossler TS, annual runoff and sunspot bivariate dynamic system and monthly average temperature and rainfall series in Dalian, and Lorenz63	N.A.	MAD (Mean Absolute Deviation), RMSE, NRMSE, MAPE
[63]	Wind data sets	Data cleaning, filling the vacancy, detection and removal of outliers, feature scaling: normalization and standardization, 80% for training and 20% for validation	MAE, MSE, RMSE
[64]	BSE100 S&P Sensex Index	BSE100 S&P Sensex weekly: 79% training and 21% validation; BSE100 S&P Sensex daily: 79% training and 21% validation; Wavelet denoising	MAD, MSE, RMSE, MAPE
[65]	31 articles	Query on Scholar Google, years from 2017 to 2020, excluded articles based on other applications	N.A.
[66]	Lorenz63, and real health index	N.A.	RMSE, NRMSE, MAPE
[37]	Monthly Sunspot by Bray & Loughead (1964), Power load collected from large customers in a particular area of China, and Lorenz63	Feature scaling: normalization, 70% for training, 10% for validation, and 20% for test	MSE, NRMSE, sMAPE

4.1. Bibliometric analysis

The first analysis consists of a study of the articles and the selected articles main characteristics are shown in Table 2: 26.1% had used financial TS data sets, with all of them presenting the data set used; of those articles that propose a new technology related to RC/ESN, we can see that they were almost half: 43.5%; the comparison of RC/ESN with Statistical techniques is 13.0%; the comparison of other neural networks with Statistical techniques are 4.3%, and on presenting the results analyzed statistically, we considered that 8.7% did this because it is different just use Statistical techniques and use them to analyze.

It is important to stress that the data sets are not always TS and are of the most varied types, with Mackey-Glass differential delay equations predominating, data generated by integrating a simplified model of a weather system developed by Lorenz in 1963, and the finance TS data, generally obtained from Standard and Poor's (S&P).

About the data preprocessing used, Table 3, we see that they are, generally, scale the data, clear the data, decompose TS, make data stationary, more or less 80% of data is for training, and 20% for validation, and feature selection.

Regarding the errors for measuring accuracy, Table 2, the major of them used RMSE, after MSE and NRMSE, and MAPE. Although the RMSE tends to increase values and magnify outliers, it has the advantage of giving more weight to larger errors. MAPE and NRMSE are equivalent if the normalization of the second is between 0 and 1, they are relative errors and show percentuals.

The most relevant author's keywords are in Fig. 1: Echo State Network is the one which appeared more, followed by forecasting, prediction, recurrent neural network, and regression. Among the most relevant, financial management, machine learning, reservoir computing, stock market prediction, and time series analysis are the ones that appeared less. All are related to each other when looking for forecasting or prediction by means of regression with ML or RC using finance time series. Other keywords appeared only once.

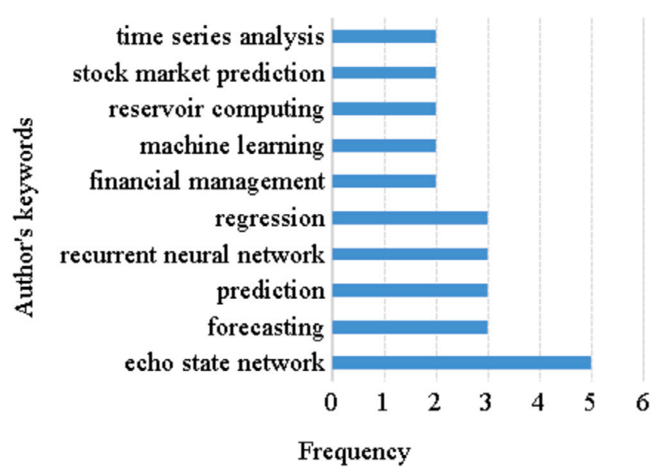

Fig. 1. Most frequent author's keywords in the articles.

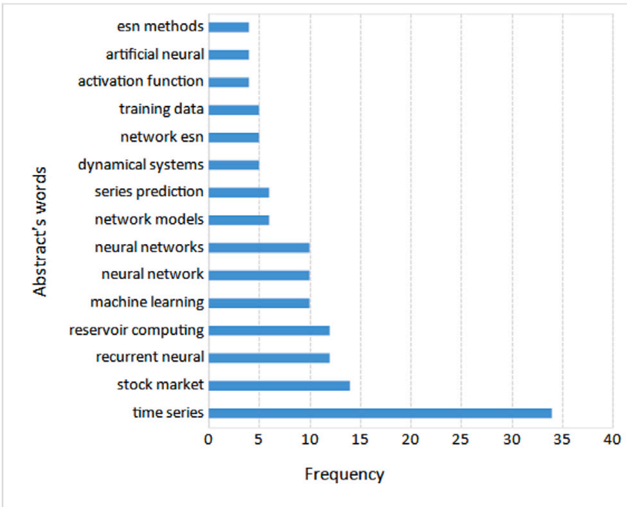

Fig. 2. Most relevant expressions (sequence of two words) in the abstracts.

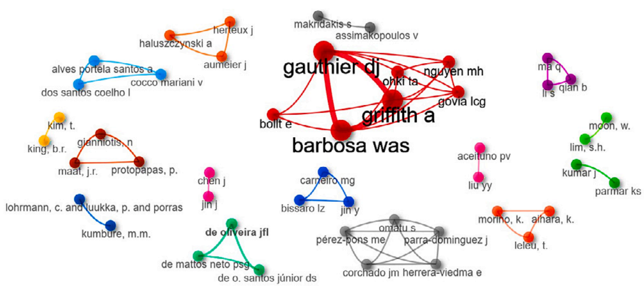

Fig. 3. Author's collaboration networks in the articles.

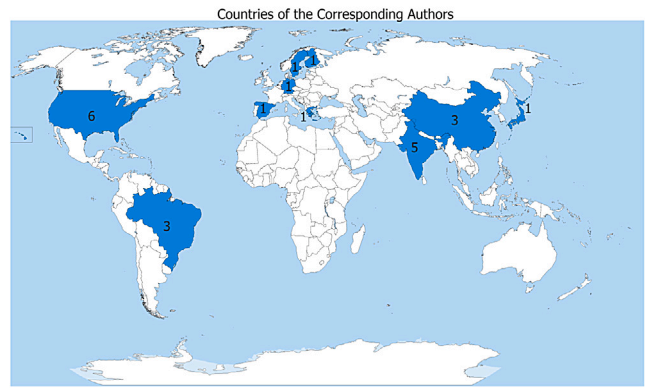

Fig. 4. Countries of the Corresponding Authors.

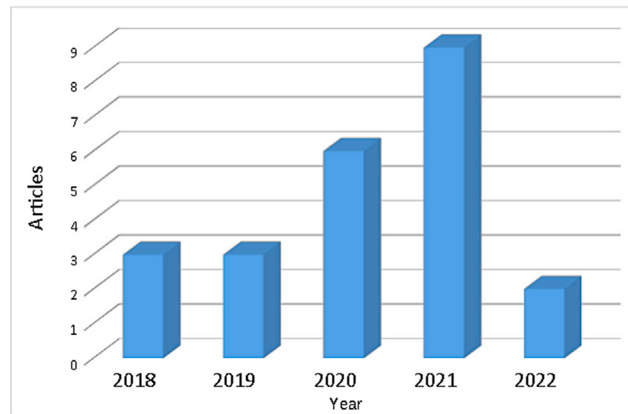

Fig. 5. Number of articles by year.

An analysis of the most relevant words done two by two (forming an expression) in the abstracts is in Fig. 2. We can see that Time Series was the expression that appeared the most, followed by the stock market, recurrent neural, reservoir computing, machine learning, and neural network(s). The subject that unites all these expressions is the application of neural networks and reservoir computing to stock market time series.

The relationship between authors and research groups is provided in Fig. 3. The stronger the connection more jointly published articles. We see that a research group is highlighted in this graphic, formed by the authors Gauthier, Griffith, Barbosa, and Bollt, which are of the same research group of Ohio State University and Clarkson University. Bollt appeared alone and as second author in 2 articles, Barbosa as first and fourth in 2 articles, Gauthier as first and eighth in 2 articles, and Griffith as second and third in 2 articles. All the others appeared just once in the other articles.

It had 2 articles just with 1 author, 2 articles with 2 authors, 9 articles with 3 authors, 5 articles with 4 authors, 3 articles with 5 authors, 1 article with 6 authors, and 1 article with 8 authors, so the average number of authors was 3.6.

Considering the author's production of the documents we have: India (5 articles), USA (4), China (3), Brazil (2), Finland (1), Germany (1), Greece (1), Japan (1), Brazil and United Kingdom (1), Sweden and USA (1), Spain, Japan and Malaysia (1), and USA, Germany and China (1). In Fig. 4 we can see the distribution map of the corresponding authors with the number of authors by country.

About the techniques used, Table 4, we see that they are divided, mainly, among Math analysis of the Reservoir, Experimental analysis of the reservoir, ESN and its variations, like DeepESN and RVESN, comparisons of techniques, forms of hyper-parameters ESN optimization. So, the techniques used are in their major part related to RC and ESN. Of the selected articles 6 do experimental analysis, being that of these 5 do an experimental analysis of the reservoir; 3 do math analysis of the reservoir; 7 use ESN, of these 3 are variations of ESN; 3 do hybridization; 1 do literature review; 1 do survey; 1 do systematic mapping; 1 compare ARIMA and RNN.

Yet about techniques used, we want to clarify the ones related to ESN:

- RVESN uses a more robust heavy-tailed Gaussian mixing distribution as the likelihood function of the model output [7];
- LSESN is a multi-reservoir structure model used to capture the multi-scale temporal characteristics of time series [37];
- DeepESN is basically a stack of reservoirs (called layers), where the output of one layer is the input of other [70].

The objectives are, mainly, to optimize ESN hyper-parameters, test activation functions, improve prediction/ forecasting, propose reductions in the reservoir size, and do reviews.

Regarding the number of articles published by year, Fig. 5 is provided a relation to the number of documents per year. It is noticeable that growing up to 2021, the smaller amount of articles in 2022 can be partially attributed to the search that was not performed in 2023.

We highlight the articles with the most citations[9]: [58], which compares Machine Learning techniques and Statistical techniques results,

[9] It was used Google Scholar to verify the number of citations in 20th July 2023.

Table 4
Techniques used and the objectives.

Articles	Techniques	Objectives
[47]	ESN; Math analysis of the Reservoir for the ESN memory; Trying a range of scalings	To optimize ESN learning; To optimize ESN hyper-parameters
[48]	Math analysis of the Reservoir; Matching the symmetry properties of a Reservoir Computing; Bayesian optimizer	To increase the power of the data processing; To optimize hyper-parameters
[49]	Experimental analysis of the reservoir structure using Regular ESN; Regular ESN and Small-World ESN, kNN, Linear SVM, Linear Regression, Logistic Regression, Generalized SVM, Decision Tree, Random Forest, Naïve Bayes, and MLP	To evaluate Regular, Small-World and random ESN models and to analyze eight classification techniques
[50]	Experimental analysis of the reservoir; Vector Auto-Regressive (VAR); Dynamic Mode Decomposition (DMD)	To show that Linear RC Machine is equivalent to VAR; To show that Linear RC Machine has a connection to DMD
[51]	ESN; Reservoir Computing; Running multiple iterations of the experiment while trying out different values; tanh	To predict/forecast long-term behavior and varying trends of a nonlinear financial system; To optimize hyper-parameters; To test the activation function
[52]	Experimental analysis of the reservoir and nonlinear vector regression; Reservoir Computing and Nonlinear Vector Auto-Regression	To demonstrate that Nonlinear Vector Auto-Regression excels at Reservoir Computing and requires even shorter training data sets and training time
[53]	Experimental analysis of the reservoir; Nodes removal; Random search procedure with uniform sampling from the parameter space	To reduce network size and improve prediction stability; To optimize the hyper-parameters
[54]	DeepESN; AD-DeepESN; ACF and PACF	To predict TS; To understand numerous characterizations of the data, including assessing potential cyclic behavior and periodicity
[55]	Literature review	To investigate machine learning techniques applied to stock market prediction
[56]	Experimental analysis of the reservoir using ESN and DeepESN; Machine Learning (ESN or DeepESN) and numerical integration; Classical static validation scheme and perform a Grid search over a prechosen hyper-parameters	To predict rare critical transition events for a class of slow–fast nonlinear dynamical systems; To optimize the hyper-parameters
[57]	ESN; Bayesian optimization and Grid search	To optimize ESN hyper-parameters efficiently; To optimize ESN hyper-parameters
[58]	Experimental analysis comparing Neural Networks and Statistical techniques; Neural Networks (MLP, BNN, RBF, GRNN, SVR, RNN, and LSTM) and Statistical techniques (kNN, CART Regression Trees, and Gaussian Processes); sigmoid	To compare the techniques; To test the activation function
[59]	ESN, regression, RNN, and LSTM	To compare these approaches of stock prediction to ESN
[60]	Systematic Mapping Study	To identify the latest applications and do a comparative study of econometric and ML models
[61]	Hybridization of ESN, Particle Swarm Optimization (PSO), and Heterogeneous Auto-regressive (HAR); Hybrid model: HAR-ESN; Select de ARIMA models in the set {0,1,2}; PSO meta-heuristic	To predict stock prices return volatility; To select the ARIMA model; To optimize the hyper-parameters
[62]	Math and experimental analysis of the reservoir; Reservoir Computing	To propose a reservoir size reduction by inputting the past or drifting states of the reservoir to the output layer
[38]	Hybridization of ARIMA, Multi-Layer Perceptron (MLP), and Support Vector Regression (SVR); ARIMA, MLP, and SVR; Automatic step-wise approach; Grid search	To propose a new strategy for combination of forecasts from linear and nonlinear models; To select ARIMA models; To optimizer MLP and SVR hyper-parameters
[7]	Robust Variational ESN (RVESN); Cross validation	To be more robust to outliers and represent the dynamic characteristics of the multivariate TS more comprehensively; To optimize the hyper-parameters
[63]	Comparison of ARIMA and RNN; ACF and PACF, and AIC and BIC; Grid search method	To predict short term wind speed; To select the ARIMA model; To optimize RNN hyper-parameters
[64]	Hybridization between ARIMA and Wavelet methods; ACF and PACF	To estimate future of stock market; To select ARIMA model
[65]	Comprehensive survey	To review the significance and need of Deep Neural Networks (DNNs) in the field of stock price and trend prediction
[66]	ESN; New particle swarm optimization-gravitational search; Hermite polynomial-based activation function	To predict TS; To converge hyper-parameters optimization fast
[37]	Long/Short-term ESN (LSESN); Genetic Algorithm	To predict TS; To optimize ESN hyper-parameters

and [52], which proposes a new technique for implementing Reservoir Computing.

In addition, it is important noticing that most of the articles analyzed deal with ML techniques (Reservoir Computing and ESN, more specifically), 3 deal with hybrid techniques, 3 did reviews, and 4 did comparisons between Statistical techniques with ESN/other ML techniques.

Finally, we found in the bibliometric analysis that the journals that appeared most as publishers were Expert Systems with Applications and Chaos, both with 3 articles, next came Neural Computing and Applications and Plos One, both with 2. The other 13 articles were from different publishers.

4.2. Content qualitative analysis and meta-synthesis

About reservoir computing, we found these articles:

1. [48]: demonstrate that matching the symmetry properties of an RC to the data increases its processing power, and state that it is possible to elevate the RC performance combining the symmetries of the reservoir and of the learning system by making direct changes to the reservoir input and output but not in the reservoir;
2. [50]: analyzes why and how an RC works at all, even using randomly selected weights,for this, connects RC to TS on VAR (Vector Auto-Regression), including theorems on representability by means of WOLD theorem, and associates with Dynamic Mode Decomposition (DMD) theory. Concludes that the concept behind RC works;
3. [52]: "demonstrate that Nonlinear Vector Auto-Regression (NVAR) overcomes Reservoir Computing benchmark tasks and requires even shorter training data sets and training time, fewer meta-parameters, and no random matrices, presenting the next generation of Reservoir Computing". Applies NVAR to three defyer problems as a benchmark: "(1) forecasting the short-term dynamics; (2) reproducing the long-term climate of a chaotic system (...); and (3) inferring the behavior of unseen data of a dynamical system";
4. [53]: "investigate the connection between properties of the reservoir and prediction quality". Using a nonlinear scaling factor in the hyperbolic tangent of the activation function they conclude that "a controlled node removal of the 'right' nodes not only leads to less variability, and thus better predictions, but also allows to reduce network size noticeably;
5. [56]: "study the problem of predicting rare critical transition events for a class of slow–fast nonlinear dynamical systems". Present "a data-driven method to predict the future evolution of the state". They tried to answer these questions: (1) Can they predict if and when a rare event will occur in a given future time window? Can they infer the characteristics of the event? (2) How far in advance can they predict the rare event? (3) With what accuracy and certainty can they achieve these goals? (4) Is it possible to answer all of these questions with a computationally inexpensive method and/or using a relatively short time series data for x? Concluding that "their method is capable of predicting a critical transition event at least several numerical time steps in advance";
6. [62]: propose to "reduce the size of the reservoir by inputting the past or drifting states of the reservoir to the output layer at the current time step", and analyzed it based on the Information Processing Capacity (IPC), reducing the reservoir size in the generalized Hénon-map and NARMA tasks. They conclude that "were able to reduce the size of the reservoir to one-tenth without a substantial increase in regression error";
7. [66]: propose "a Hermite polynomial-based activation function design with fast convergence (...) and the relation between long-term time dependence and edge-of-chaos criticality is given". The parameters to attain on the edge-of-chaos were improved by a new particle swarm optimization-gravitational search algorithm. Concluding that this "evolution makes the reservoir great potential to run on the edge of chaos with rich expression".

About the echo state network, and some ESN variation, we found these articles:

1. [47]: "show that the ensemble of eigenvalues of the network contributes to the ESN memory capacity", validating with forecasting applied to synthetic and real benchmark TS. State that reservoir characteristics, in particular, cannot be applied to general tasks. And conclude that "provide a simple way to design task-specific ESN and offer deep insights for other recurrent neural networks";
2. [49]: present ESN like a model to real-world applications. They investigate its inherent concepts and components related to the reservoir structure and readout layer. Evaluate regular and small-world network models and analyze eight classification techniques. Analyze a wide range of parameters in reservoir and readout layers through of several experiments with five real-world data sets. They conclude that ESN techniques are more efficient in embedded systems and large-scale problems;
3. [51]: employ an ESN to model a complex financial system, regenerating "the financial system unknowns with a high degree of accuracy when only limited future data is available, thereby, becoming a reliable feeder for any long-term decision making or policy formulations". They conclude that the model effectively forecasts long-term system behavior;
4. [54]: propose a novel approach based on an additive decomposition to a TS as a preprocessor to a deep ESN, evaluating it with 6 data sets: S&P 500, Parking Birmingham, Chaos data, Philadelphia temperature, Periodic data, and Bike sharing. They conclude that it has a performance superior, even when applied to non-stationary and chaotic TS;
5. [57]: study the ESN hyper-parameters determination tuning to model a group of TS of similar temporal behavior, not only individuals TS. Demonstrate their approach in synthetic data sets and astronomical light curve data set. They conclude that their "approach results in a significant reduction in the number of ESNs models required to model a whole data set while retaining predictive performance for the series in each cluster";
6. [59]: compares the efficiency of standard approaches of stock prediction to model chaotic nonlinear systems such as daily closing stock prices in NY stock market. States that "usual neural network model may fail to predict the values accurately as the initialized weights of random selection problem is prone to falling to the local optima before descending to the maximum optima, resulting in incorrect predictions". He concludes by saying that the "echo state network outperforms the conventional regression and (other) neural network approach (LSTM)";
7. [7]: propose "a robust and adaptive multivariate nonlinear time series prediction model (...) based on echo state network and variational inference, (...) utilized to handle the marginal likelihood function of the model output which is analytically intractable for the mixture distribution", called "robust variational echo state network (RVESN)". They conclude that "RVESN has better prediction performance";
8. [37]: "propose a model consisting of a multi-reservoir structure named long-short term echo state networks (LS-ESNs) to capture the multi-scale temporal characteristics of time series". LS-ESN has three independent reservoirs, each reservoir with recurrent connections of a specific time-scale, which are collected and concatenated together. They conclude that was able to "demonstrate the effectiveness of the proposed LS-ESNs".

About literature reviews, we found these articles:

1. [55]: present a literature review on machine learning techniques applied for stock market prediction examining 138 journal articles published between 2000 and 2019. The objective is "to broaden the current knowledge in stock market predictions through a systematic literature study". (...) they "established the research questions: (1) What kind of data (variables and type of variables as well as time horizons) are used? and (2) what machine learning and AI techniques were applied in stock market prediction studies?" In addition to these questions, they included "several other contributions: a discussion on state-of-the-art machine learning-based forecasting models over the last two decades, a bibliometric analysis of the selected studies with the most significant research contexts (...), and validation methods, with respect to the selected literature". And they conclude that in the articles there were 2173 unique variables used for stock market predictions and present the machine learning techniques and their variants deployed for the predictions through a bibliometric analysis;
2. [60]: they "objective identify the latest applications and do a comparative study of the performance of econometric and ML models", aiming "to find empirical evidence for the performance of ML algorithms being superior to traditional econometric models". Conclude that "there is no certainty as to the performance of ML being always superior to that of econometric models";
3. [65]: they "review the significance and need of DNNS in the field of stock price and trend prediction (...), discuss the applicability of DNN variations to the temporal stock market data and also extend (...) survey to include hybrid, as well as meta-heuristic, approaches with DNN". They "observe the potential limitations for stock market prediction using various DNNs". (...) they "also conduct a series of experiments for stock market prediction using nine deep learning-based models;" they "analyse the impact of these models on forecasting the stock market data". They "also evaluate the performance of individual models with different number of features". They conclude their work by showing the "perspectives of DNN-based stock market prediction, primarily covering research spanning from 2017 to 2020".

About hybridization techniques, we found these articles:

1. [61]: they propose a hybrid model that combines HAR (heterogeneous auto-regressive) with ESN and the PSO (particle swarm optimization) meta-heuristic for daily forecasting realized volatilities of three Nasdaq stocks, "considering 1-day, 5-days, and 21-days ahead forecasting horizons". They benchmarked "against existing specifications, such as autoregressive integrated moving average (ARIMA), HAR, multilayer perceptron (MLP), and ESN". They "evaluated in terms of r-squared and mean squared error performance metrics, and the statistical comparison is made through a Friedman test followed by a post-hoc Nemenyi test". They conclude that the proposed model produces more accurate predictions in most cases;
2. [38]: propose a hybrid system combining an ARIMA model with Multi-Layer Perceptron and Support Vector Regression applied to forecast six real-world complex TS, "that searches for a suitable function to combine the forecasts of linear and nonlinear models". (...) the "proposed system performs: (i) linear modeling of the time series; (ii) nonlinear modeling of the error series; and (iii) a data-driven combination that searches for: (iii.a) the most suitable function, between linear and nonlinear formalisms, and (iii.b) the number of forecasts of models (i) and (ii) that maximizes the performance of the combination". They conclude that the "proposed hybrid system attains superior performance compared with single and hybrid models";
3. [64]: couple soft computing models of discreet wavelet transformation and wavelet denoising with autoregressive models to forecast closing prices of the BSE100 S&P Sensex index. State that "wavelet methods, being capable of handling nonlinear data, combined with autoregressive models generate more accurate forecasts". They conclude that this complement generates considerably accurate forecasts.

In comparison between ML and statistical techniques, we found this articles:

1. [58]: use data sets of M3 Competition to compare ML methods with eight traditional Statistical techniques and conclude that the Statistical ones have high accuracy and computational requirements considerably smaller than those of ML methods;
2. [63]: they present "a comparative study of a time series Statistical model (ARIMA, . . .) and a deep learning model (RNN, . . .)" to forecast wind energy and help minimizing and also optimizing the use of conventional power plants, using two wind data sets: A- "consists of mean wind speed containing 192 hourly basis data points from 20 Jan 2019 (from time 00:00) to 27 Jan 2019 (till time 23:00). This data set is of Coimbatore taken from meteoblue weather website and the wind speed is measured from 10 m above the ground"; and B- "containing 865 data points with 10-minutes interval between each point from 1 Jan 2008 (from time 00:00) to 7 Jan 2008 (till time 00:00) which is taken from Blandford MTA and the data is collected by RERL @ Univ. of Massachusetts". Concluding "that for short-term wind speed forecasting, RNN model works well compared to ARIMA model for our datasets".

5. Final considerations

Two mathematics modeling techniques for forecasting have obtained promising results in the last decades they are Machine Learning and Statistical techniques. So, the objective of this study was to conduct a systematic literature review of ESN (Machine Learning technique) and ARIMA-GARCH (classical Statistical techniques for modeling TS and forecasting) to forecast Time Series.

From the analysis of the articles read, it was possible to answer the research questions:

- What are the existing RC models and what are the common terminologies to describe these models? The existing RC models are ESN and LSM, and these are the terminologies to describe them. What are their components and how are these components related? The main component of ESN and LSM is the reservoir. Does one of the models define all the others adequately for any application? Both define each other adequately for any application.
- Can RC make predictions? Yes, it can, as we can be seen in works [37,54], and [66]. Can it be applied to Financial Time Series? Yes, it can, as we can be seen in works [59,61]. How to improve these predictions? A way to improve predictions is hybridizing Statistical techniques with ESN. How accurate are these predictions/how to minimize error? The accuracy depends on the problems, but generally, it is reasonable.

A good part of the selected articles are proposing new technologies related to ESN or to Reservoir Computing, but few are comparing with Statistical techniques or other Neural Networks, which reveals a gap in hybridizing ESN with Statistical techniques for forecasting.

Generally, studies on Echo State Network and Reservoir Computing have two approaches: to improve hyperparameters tuning and to make adjusts in the reservoir, aiming to achieve better performance and accuracy. As we saw, most of the selected articles choose the second way.

ESN does not use TS only, consequently, its exploration with financial Time Series is small, and just one work did hybridization of Statistical techniques and ESN for stock price forecasting.

The articles that compare ESN with Statistical techniques do not analyze results statistically. This reveals a failure when hybridizing these techniques.

No studies were selected addressing the issue of the amount of data needed for each or hybridized technique. Perhaps the need for data is lower in Statistical techniques than in computational ML methods. The former cannot always forecasts, due to several restrictions necessary for their use. ML techniques, which ignore statistical restrictions, may require more data, but can always make forecasts if they can overcome the overfitting problem.

Although, ML and Statistical techniques can use the same software, like R Studio, and they can be similar underneath, they solve the problems of different manners. The ESN and Statistical techniques differ the way they adjust the hyperparameters and in the way they treat the mathematical restrictions, but the article of [50] states that one class of ESNs is similar to Vector Autoregressive (VAR), which can be considered a Statistical technique. However, VAR uses several variables, and in this work we are looking for ESN to forecast just two: stock closing prices and time.

ML and Statistics techniques also differ in how they evaluate results. While in Statistics, inference points out a probable conclusion, ML seeks generalizing results, avoiding overfitting, and applying the model to other data. Would the hybridization of these models be a way to improve this generalization?

We concluded that ESN can adapt to solve stochastic dynamical systems with TS. But, it is not usual applying it to the modeling and forecasting of financial TS, being underutilized in this research topic.

After examining the results of this review, it can be concluded that the use of a hybrid approach, combining Statistical techniques with ESN, needs to be more studied to lead to significant improvements in the accuracy of forecasting. These findings also suggest that alternative approaches may need to be explored in order to improve the accuracy of stock price forecasting in the future. Besides, is important to highlight that papers that make hybridization do not use Garch, but ARIMA. We can consider it occurs because both GARCH and ESN better model the nonlinear component of the TS and ARIMA the linear component.

Observations

The eventual opinions, hypotheses, conclusions, or express recommendations in this article are the sole responsibility of the authors and do not necessarily reflect Fesurv's [10] point of view.

CRediT authorship contribution statement

Fabian Corrêa Cardoso: Writing – original draft, Visualization, Resources, Investigation, Funding acquisition, Data curation, Conceptualization. **Rafael Alceste Berri:** Writing – review & editing, Validation, Supervision, Methodology, Formal analysis, Conceptualization. **Eduardo Nunes Borges:** Writing – review & editing, Validation, Methodology. **Bruno Lopes Dalmazo:** Writing – review & editing. **Giancarlo Lucca:** Writing – review & editing, Validation, Methodology. **Viviane Leite Dias de Mattos:** Writing – review & editing, Visualization, Validation, Supervision, Project administration, Methodology, Conceptualization.

[10] Fesurv, in Portuguese, stands for Fundação do Ensino Superior de Rio Verde

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Fabian Corrêa Cardoso was supported with scholarship by Coordination of Superior Level Staff Improvement (CAPES in Brazil).

Data availability

The authors are unable or have chosen not to specify which data has been used.

Acknowledgments

Thanks to Fesurv-UniRV[11] and the Graduate Program in Computational Modeling at the Federal University of Rio Grande, which made this work possible.

References

[1] H. Lukoševičius, A practical guide to applying echo state networks, in: Neural Networks: Tricks of the Trade: Second Edition, Springer, 2012, pp. 659–686.
[2] P. Hänggi, H. Thomas, Stochastic processes: Time evolution, symmetries and linear response, Phys. Rep. 88 (4) (1982) 207–319.
[3] P.A. Morettin, C.M.C. Toloi, Análise de Séries Temporais: Modelos Lineares Univariados, third ed., Blucher, 2018.
[4] S.S. Soman, H. Zareipour, O. Malik, P. Mandal, A review of wind power and wind speed forecasting methods with different time horizons, in: North American Power Symposium 2010, IEEE, 2010, pp. 1–8.
[5] E.N. Lorenz, Deterministic nonperiodic flow, J. Atmos. Sci. 20 (2) (1963) 130–141.
[6] C.L. Dunis, X. Huang, Forecasting and trading currency volatility: An application of recurrent neural regression and model combination, J. Forecast. 21 (5) (2002) 317–354.
[7] L. Shen, J. Chen, Z. Zeng, J. Yang, J. Jin, A novel echo state network for multivariate and nonlinear time series prediction, Appl. Soft Comput. 62 (2018) 524–535.
[8] P. Mondal, L. Shit, S. Goswami, Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices, Int. J. Comput. Sci. Eng. Appl. 4 (2) (2014) 13.
[9] A. Hossain, F. Zaman, M. Nasser, M.M. Islam, Comparison of GARCH, neural network and support vector machine in financial time series prediction, in: Pattern Recognition and Machine Intelligence: Third International Conference, PReMI 2009 New Delhi, India, December 16-20, 2009 Proceedings 3, Springer, 2009, pp. 597–602.
[10] H. Jaeger, The "echo state" approach to analysing and training recurrent neural networks-with an erratum note, Bonn, Germany: German National Research Center for Information Technology GMD Technical Report, 148, (34) 2001, p. 13.
[11] W. Maass, T. Natschläger, H. Markram, Real-time computing without stable states: A new framework for neural computation based on perturbations, Neural Comput. 14 (11) (2002) 2531–2560.
[12] V.K. Ojha, A. Abraham, V. Snášel, Metaheuristic design of feedforward neural networks: A review of two decades of research, Eng. Appl. Artif. Intell. 60 (2017) 97–116.
[13] Y. Chen, Z. He, Z. Shang, C. Li, L. Li, M. Xu, A novel combined model based on echo state network for multi-step ahead wind speed forecasting: A case study of NREL, Energy Convers. Manag. 179 (2019) 13–29.
[14] B. Schrauwen, D. Verstraeten, J. Van Campenhout, An overview of reservoir computing: theory, applications and implementations, in: Proceedings of the 15th European Symposium on Artificial Neural Networks, 2007, pp. 471–482.
[15] Z. Zhang, J. Lu, G. Zhou, X. Liao, Research on tool wear prediction based on LSTM and ARIMA, in: Proceedings of the 2018 International Conference on Big Data Engineering and Technology, 2018, pp. 73–77.
[16] M. Elsaraiti, A. Merabet, A comparative analysis of the arima and lstm predictive models and their effectiveness for predicting wind speed, Energies 14 (20) (2021) 6782.
[17] P. Murali, R. Revathy, S. Balamurali, A.S. Tayade, Integration of RNN with GARCH refined by whale optimization algorithm for yield forecasting: a hybrid machine learning approach, J. Ambient Intell. Humaniz. Comput. (2020) 1–13.
[18] W. Jiang, Applications of deep learning in stock market prediction: recent progress, Expert Syst. Appl. 184 (2021) 115537.

[11] UniRV, in Portuguese, stands for Universidade de Rio Verde

[19] M. Cattani, I.L. Caldas, S.L. de Souza, K.C. Iarosz, Deterministic chaos theory: Basic concepts, Revista Brasileira de Ensino de Física 39 (1) (2016).
[20] A.G. Hart, K.R. Olding, A.M.G. Cox, O. Isupova, J.H.P. Dawes, Using echo state networks to approximate value functions for control, 2021, arXiv preprint arXiv:2102.06258.
[21] G. Contopoulos, Order and chaos in dynamical systems, Milan J. Math. 77 (2009) 101–126.
[22] X. Li, R. Ramachandran, S. Movva, S. Graves, B. Plale, N. Vijayakumar, Storm clustering for data-driven weather forecasting, in: 24th Conference on International Institute of Professional Studies (IIPS). University of Alabama in Huntsville, 2008.
[23] T. Weng, H. Yang, J. Zhang, M. Small, Modeling chaotic systems: Dynamical equations vs machine learning approach, Commun. Nonlinear Sci. Numer. Simul. 114 (2022) 106452.
[24] J. Wang, The analysis of the financial market in China, Acad. J. Bus. Manag. 3 (2) (2021) 58–60.
[25] M.A. de Oliveira, M.A. de Ávila, W.M. da Silva, D.R. Bergmann, Estabilidade de preços de ações no mercado de capitais brasileiro: um estudo aplicando redes neurais e expoentes de Lyapunov, Revista de Administração 46 (2) (2011) 161–177.
[26] I.K. Nti, A.F. Adekoya, B.A. Weyori, A systematic review of fundamental and technical analysis of stock market predictions, Artif. Intell. Rev. 53 (4) (2020) 3007–3057.
[27] D.P. Gandhmal, K. Kumar, Systematic analysis and review of stock market prediction techniques, Comp. Sci. Rev. 34 (2019) 100190.
[28] G.P. Zhang, Time series forecasting using a hybrid ARIMA and neural network model, Neurocomputing 50 (2003) 159–175.
[29] H. Guo, X. Wang, Z. Gao, Uncertain linear regression model and its application, J. Intell. Manuf. 28 (2017) 559–564.
[30] J.D. Hamilton, Time Series Analysis, Princeton University Press, 2020.
[31] P.A. Morettin, Econometria Financeira, second ed., Blucher, 2011.
[32] T.C. Mills, R.N. Markellos, The Econometric Modelling of Financial Time Series, Cambridge University Press, 2008.
[33] R.d.L.d.S. Bueno, J. Inhasz, Econometria de séries temporais: manual de soluções, Cengage Learning, 2011.
[34] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, J. Econometrics 31 (3) (1986) 307–327.
[35] T. Bollerslev, On the correlation structure for the generalized autoregressive conditional heteroskedastic process, J. Time Series Anal. 9 (2) (1988) 121–131.
[36] R.F. Engle, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, Econometrica (1982) 987–1007.
[37] K. Zheng, B. Qian, S. Li, Y. Xiao, W. Zhuang, Q. Ma, Long-short term echo state network for time series prediction, IEEE Access 8 (2020) 91961–91974.
[38] D.S.d.O. Santos Jr., J.F.L. de Oliveira, P.S.G. de Mattos Neto, An intelligent hybridization of ARIMA with machine learning models for time series forecasting, Knowl.-Based Syst. 175 (2019) 72–86.
[39] C.N. Babu, B.E. Reddy, A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data, Appl. Soft Comput. 23 (2014) 27–38.
[40] J.D.V. Henao, V.R. Mejia, C.J.F. Cardona, Electricity demand forecasting using a SARIMA-multiplicative single neuron hybrid model, Dyna 80 (180) (2013) 4–8.
[41] C. Sigauke, Forecasting medium-term electricity demand in a South African electric power supply system, J. Energy Southern Africa 28 (4) (2017) 54–67.
[42] Y. Yang, Y. Chen, Y. Wang, C. Li, L. Li, Modelling a combined method based on ANFIS and neural network improved by DE algorithm: A case study for short-term electricity demand forecasting, Appl. Soft Comput. 49 (2016) 663–675.
[43] D.O. Faruk, A hybrid neural network and ARIMA model for water quality time series prediction, Eng. Appl. Artif. Intell. 23 (4) (2010) 586–594.
[44] M. Khashei, M. Bijari, A novel hybridization of artificial neural networks and ARIMA models for time series forecasting, Appl. Soft Comput. 11 (2) (2011) 2664–2675.
[45] I. Khandelwal, R. Adhikari, G. Verma, Time series forecasting using hybrid ARIMA and ANN models based on DWT decomposition, Procedia Comput. Sci. 48 (2015) 173–179.
[46] E.Y. Nakagawa, K.R.F. Scannavino, S.C.P.F. Fabbri, F.C. Ferrari, Revisão Sistemática da Literatura em Engenharia de Software: teoria e prática, first ed., Livros Técnicos e Científicos-LTC, 2017.
[47] P.V. Aceituno, G. Yan, Y. Liu, Tailoring echo state networks for optimal learning, iScience 23 (9) (2020).
[48] W.A.S. Barbosa, A. Griffith, G.E. Rowlands, L.C.G. Govia, G.J. Ribeill, M. Nguyen, T.A. Ohki, D.J. Gauthier, Symmetry-aware reservoir computing, Phys. Rev. E 104 (4) (2021) 045307.
[49] L.Z. Bissaro, Y. Jin, M.G. Carneiro, Regular echo state networks: simple and accurate reservoir models to real-world applications, in: Proceedings of the 36th Annual ACM Symposium on Applied Computing, 2021, pp. 1063–1069.
[50] E. Bollt, On explaining the surprising success of reservoir computing forecaster of chaos? The universal machine learning dynamical system with contrast to VAR and DMD, Chaos 31 (1) (2021) 013108.
[51] R. Budhiraja, M. Kumar, M.K. Das, A.S. Bafila, S. Singh, A reservoir computing approach for forecasting and regenerating both dynamical and time-delay controlled financial system behavior, Plos One 16 (2) (2021) e0246737.
[52] D.J. Gauthier, E. Bollt, A. Griffith, W.A.S. Barbosa, Next generation reservoir computing, Nat. Commun. 12 (1) (2021) 5564.
[53] A. Haluszczynski, J. Aumeier, J. Herteux, C. Räth, Reducing network size and improving prediction stability of reservoir computing, Chaos 30 (6) (2020) 063136.
[54] T. Kim, B.R. King, Time series prediction using deep echo state networks, Neural Comput. Appl. 32 (2020) 17769–17787.
[55] M.M. Kumbure, C. Lohrmann, P. Luukka, J. Porras, Machine learning techniques and data for stock market forecasting: A literature review, Expert Syst. Appl. (2022) 116659.
[56] S.H. Lim, L.T. Giorgini, W. Moon, J.S. Wettlaufer, Predicting critical transitions in multiscale dynamical systems using reservoir computing, Chaos 30 (12) (2020) 123126.
[57] J.R. Maat, N. Gianniotis, P. Protopapas, Efficient optimization of echo state networks for time series datasets, in: 2018 International Joint Conference on Neural Networks, IJCNN, IEEE, 2018, pp. 1–7.
[58] S. Makridakis, E. Spiliotis, V. Assimakopoulos, Statistical and machine learning forecasting methods: Concerns and ways forward, PLoS One 13 (3) (2018) e0194889.
[59] A. Mallya, Echo state networks and existing paradigms for stock market prediction, in: 2021 International Conference on Emerging Smart Computing and Informatics, ESCI, IEEE, 2021, pp. 13–18.
[60] M.E. Pérez-Pons, J. Parra-Dominguez, S. Omatu, E. Herrera-Viedma, J.M. Corchado, Machine learning and traditional econometric models: a systematic mapping study, J. Artif. Intell. Soft Comput. Res. 12 (2) (2022) 79–100.
[61] G.T. Ribeiro, A.A.P. Santos, V.C. Mariani, L.d.S. Coelho, Novel hybrid model based on echo state neural network applied to the prediction of stock price return volatility, Expert Syst. Appl. 184 (2021) 115490.
[62] Y. Sakemi, K. Morino, T. Leleu, K. Aihara, Model-size reduction for reservoir computing by concatenating internal states through time, Sci. Rep. 10 (1) (2020) 21794.
[63] K.S. Shivani, A.R. Nair, A comparative study of ARIMA and RNN for short term wind speed forecasting, in: 2019 10th International Conference on Computing, Communication and Networking Technologies, ICCCNT, IEEE, 2019, pp. 1–7.
[64] S. Singh, K.P. Singh, J. Kumar, Soft computing model coupled with statistical models to estimate future of stock market, Neural Comput. Appl. 33 (2021) 7629–7647.
[65] A. Thakkar, K. Chaudhari, A comprehensive survey on deep neural networks for stock market: The need, challenges, and future directions, Expert Syst. Appl. 177 (2021) 114800.
[66] G. Zhang, C. Zhang, W. Zhang, Evolutionary echo state network for long-term time series prediction: on the edge of chaos, Appl. Intell. 50 (2019) 893–904.
[67] JabRef Development Team, JabRef — An open-source, cross-platform citation and reference management software, 2021, URL https://www.jabref.org/.
[68] M. Aria, C. Cuccurullo, Bibliometrix: An R-tool for comprehensive science mapping analysis, J. Inform. 11 (4) (2017) 959–975.
[69] R. Core Team, R: A language and environment for statistical computing, 2013.
[70] C. Gallicchio, A. Micheli, Deep echo state network (deepesn): A brief survey, 2017, arXiv preprint arXiv:1712.04323.