

REVIEW ARTICLE

Transformer Models for Time-Series Forecasting: A Review of Recent Advances and Applications in Energy and Finance

Shubh Raj Gupta^{1, #}, Subham Pandit², Shivam Sahu³, Aradhana Sahu⁴, Bhawna Janghel⁵

^{1,4,5} Department of Computer Science and Engineering, Rungta College of Engineering and Technology, Bhilai CG 490024

² Department of Computer Science & Engineering (AI ML), Rungta College of Engineering and Technology, Bhilai CG 490024

³ Department of Computer Science and Engineering (Cyber Security), Rungta College of Engineering and Technology, Bhilai CG 490024

*Corresponding Author: shubh.raj.gupta@rungta.ac.in

ABSTRACT:

Precise time-series prediction is essential in applications ranging from finance to energy, where the decision heavily depends on the capacity to forecast changing temporal patterns. Popular models such as ARIMA, Prophet, and LSTM have performed satisfactorily but are usually not adequate in capturing long-range dependencies and coping with irregular or multivariate series. Transformer-based architectures, initially used for natural language processing, have recently been found useful for modeling long-term dependencies and scaling well to large datasets. This has provided new possibilities for their use in time-series forecasting tasks. This review explores recent developments and new trends in the application of Transformer models to time series forecasting, with specific focus on their use in energy systems and financial markets. We review a number of specialized variants of the Transformer-Informer, Autoformer, and Temporal Fusion Transformer (TFT)-and how they specifically handle domain-specific forecasting difficulties like seasonality, multi-horizon forecasting, and high-frequency noise. In addition, we offer a comparative analysis of primary model architectures, learning processes, benchmarking datasets, and performance metrics widely utilized in these applications. We also discuss existing limitations, such as model interpretability, sparsity of data, and real-time deployment restrictions, and highlight research opportunities in the future, notably in enhancing efficiency and domain adaptation. Lastly, we briefly sketch the potential of these technologies to serve towards Sustainable Development Goals, in particular SDG 7 (Affordable and Clean Energy) and SDG 8 (Decent Work and Economic Growth).

KEYWORDS: Breast cancer; ultrasound; explainable artificial intelligence; deep learning

1. Introduction

Precise time-series prediction is fundamental to operational planning and strategic decision-making in diverse sectors, notably in energy management and financial markets. Historically, statistical models such as ARIMA have been the bedrock of forecasting, but they falter when applied to massive, non-linear datasets.

In mathematics, a time series is a sequence of data points indexed, listed, or graphed in chronological order. Most commonly, a time series consists of observations recorded at suc-

cessive equally spaced points in time. Thus, it represents a form of discrete-time data. A time series may describe measurements collected over seconds, days, years, or even centuries. Common examples include heights of ocean tides, counts of sunspots, daily temperature readings, and the closing values of stock market indices such as the Dow Jones Industrial Average [1].

A time series is often visualized using a run chart (a type of temporal line chart), which helps identify patterns such as trends, seasonal effects, and irregular fluctuations. Time series are widely used in statistics, actuarial science, signal pro-

Received on 15 July 2025 Revised on 28 October 2025

Accepted on 17 November 2025 Published on 28 November 2025

Rungta International Journal of Computer Science and Information Technology. 2025; 2(2): 67.

© RIJCST All right reserved

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

cessing, pattern recognition, econometrics, mathematical finance, weather forecasting, earthquake prediction, electroencephalography, control engineering, astronomy, communications engineering, and many other areas of applied science and engineering that involve temporal measurements [2].

Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values. Generally, time series data is modeled as a stochastic process. While regression analysis is often employed in such a way as to test relationships between one or more different time series, this type of analysis is not usually called "time series analysis", which refers in particular to relationships between different points in time within a single series [3].

Time series data have a natural temporal ordering. This makes time series analysis distinct from cross-sectional studies, in which there is no natural ordering of the observations (e.g. explaining people's wages by reference to their respective education levels, where the individuals' data could be entered in any order). Time series analysis is also distinct from spatial data analysis where the observations typically relate to geographical locations (e.g. accounting for house prices by the location as well as the intrinsic characteristics of the houses). A stochastic model for a time series will generally reflect the fact that observations close together in time will be more closely related than observations further apart. In addition, time series models will often make use of the natural one-way ordering of time so that values for a given period will be expressed as deriving in some way from past values, rather than from future values (see time reversibility) [4].

Time series analysis can be applied to real-valued, continuous data, discrete numeric data, or discrete symbolic data (i.e. sequences of characters, such as letters and words in the English language) [5].

Methods for time series analysis may be divided into two classes: frequency-domain methods and time-domain methods. The former include spectral analysis and wavelet analysis; the latter include auto-correlation and cross-correlation analysis. In the time domain, correlation and analysis can be made in a filter-like manner using scaled correlation, thereby mitigating the need to operate in the frequency domain [6].

Additionally, time series analysis techniques may be divided into parametric and non-parametric methods. The parametric approaches assume that the underlying stationary stochastic process has a certain structure which can be described using a small number of parameters (for example, using an autoregressive or moving-average model). In these approaches, the task is to estimate the parameters of the model that describes the stochastic process. By contrast, non-parametric approaches explicitly estimate the covariance or the spectrum of the process without assuming that the process has any particular structure [7].

A time series is one type of panel data. Panel data is the

general class, a multidimensional data set, whereas a time series data set is a one-dimensional panel (as is a cross-sectional dataset). A data set may exhibit characteristics of both panel data and time series data. One way to tell is to ask what makes one data record unique from the other records. If the answer is the time data field, then this is a time series data set candidate. If determining a unique record requires a time data field and an additional identifier which is unrelated to time (e.g. student ID, stock symbol, country code), then it is panel data candidate. If the differentiation lies on the non-time identifier, then the data set is a cross-sectional data set candidate [8].

In the context of statistics, econometrics, quantitative finance, seismology, meteorology, and geophysics the primary goal of time series analysis is forecasting. In the context of signal processing, control engineering and communication engineering it is used for signal detection. Other applications are in data mining, pattern recognition and machine learning, where time series analysis can be used for clustering, classification, query by content, anomaly detection as well as forecasting [9].

A simple way to examine a regular time series is manually with a line chart. The datagraphic shows tuberculosis deaths in the United States, along with the yearly change and the percentage change from year to year. The total number of deaths declined in every year until the mid-1980s, after which there were occasional increases, often proportionately - but not absolutely - quite large [10].

A study of corporate data analysts found two challenges to exploratory time series analysis: discovering the shape of interesting patterns, and finding an explanation for these patterns. Visual tools that represent time series data as heat map matrices can help overcome these challenges [11].

This approach may be based on harmonic analysis and filtering of signals in the frequency domain using the Fourier transform, and spectral density estimation. Its development was significantly accelerated during World War II by mathematician Norbert Wiener, electrical engineers Rudolf E. Klmn, Dennis Gabor and others for filtering signals from noise and predicting signal values at a certain point in time [12].

Spectral analysis to examine cyclic behavior which need not be related to seasonality. For example, sunspot activity varies over 11 year cycles. Other common examples include celestial phenomena, weather patterns, neural activity, commodity prices, and economic activity [13].

Separation into components representing trend, seasonality, slow and fast variation, and cyclical irregularity: see trend estimation and decomposition of time series [14]

Curve fitting is the process of constructing a curve, or mathematical function, that has the best fit to a series of data points, possibly subject to constraints. Curve fitting can involve either interpolation, where an exact fit to the data is required, or smoothing, in which a "smooth" function is constructed that approximately fits the data. A related topic is regression analysis, which focuses more on questions of statistical inference such as how much uncertainty is present in

a curve that is fit to data observed with random errors. Fitted curves can be used as an aid for data visualization, to infer values of a function where no data are available, and to summarize the relationships among two or more variables. Extrapolation refers to the use of a fitted curve beyond the range of the observed data, and is subject to a degree of uncertainty since it may reflect the method used to construct the curve as much as it reflects the observed data [15].

For processes that are expected to generally grow in magnitude one of the curves in the graphic (and many others) can be fitted by estimating their parameters [16].

The construction of economic time series involves the estimation of some components for some dates by interpolation between values ("benchmarks") for earlier and later dates. Interpolation is estimation of an unknown quantity between two known quantities (historical data), or drawing conclusions about missing information from the available information ("reading between the lines"). Interpolation is useful where the data surrounding the missing data is available and its trend, seasonality, and longer-term cycles are known. This is often done by using a related series known for all relevant dates. Alternatively polynomial interpolation or spline interpolation is used where piecewise polynomial functions are fitted in time intervals such that they fit smoothly together. A different problem which is closely related to interpolation is the approximation of a complicated function by a simple function (also called regression). The main difference between regression and interpolation is that polynomial regression gives a single polynomial that models the entire data set. Spline interpolation, however, yield a piecewise continuous function composed of many polynomials to model the data set [17].

Extrapolation is the process of estimating, beyond the original observation range, the value of a variable on the basis of its relationship with another variable. It is similar to interpolation, which produces estimates between known observations, but extrapolation is subject to greater uncertainty and a higher risk of producing meaningless results [18].

2. Related Work

The transition from statistical methods to deep learning in time-series forecasting has been extensively documented. The advent of deep learning introduced Recurrent Neural Networks (RNNs), which significantly improved the modeling of sequential data. Recently, Transformer-based architectures have been adapted for time-series forecasting.

In general, a function approximation problem asks us to select a function among a well-defined class that closely matches ("approximates") a target function in a task-specific way [19].

One can distinguish two major classes of function approximation problems: First, for known target functions, approximation theory is the branch of numerical analysis that investigates how certain known functions (for example, special functions) can be approximated by a specific class of functions (for

example, polynomials or rational functions) that often have desirable properties (inexpensive computation, continuity, integral and limit values, etc.) [20].

Second, the target function, call it g , may be unknown; instead of an explicit formula, only a set of points (a time series) of the form $(x, g(x))$ is provided. Depending on the structure of the domain and codomain of g , several techniques for approximating g may be applicable. For example, if g is an operation on the real numbers, techniques of interpolation, extrapolation, regression analysis, and curve fitting can be used. If the codomain (range or target set) of g is a finite set, one is dealing with a classification problem instead. A related problem of online time series approximation is to summarize the data in one-pass and construct an approximate representation that can support a variety of time series queries with bounds on worst-case error [1].

To some extent, the different problems (regression, classification, fitness approximation) have received a unified treatment in statistical learning theory, where they are viewed as supervised learning problems [2].

In statistics, prediction is a part of statistical inference. One particular approach to such inference is known as predictive inference, but the prediction can be undertaken within any of the several approaches to statistical inference. Indeed, one description of statistics is that it provides a means of transferring knowledge about a sample of a population to the whole population, and to other related populations, which is not necessarily the same as prediction over time. When information is transferred across time, often to specific points in time, the process is known as forecasting [3].

Fully formed statistical models for stochastic simulation purposes, so as to generate alternative versions of the time series, representing what might happen over non-specific time-periods in the future [4].

Simple or fully formed statistical models to describe the likely outcome of the time series in the immediate future, given knowledge of the most recent outcomes (forecasting) [5].

Forecasting on time series is usually done using automated statistical software packages and programming languages, such as Julia, Python, R, SAS, SPSS and many others [6].

Splitting a time-series into a sequence of segments. It is often the case that a time-series can be represented as a sequence of individual segments, each with its own characteristic properties. For example, the audio signal from a conference call can be partitioned into pieces corresponding to the times during which each person was speaking. In time-series segmentation, the goal is to identify the segment boundary points in the time-series, and to characterize the dynamical properties associated with each segment. One can approach this problem using change-point detection, or by modeling the time-series as a more sophisticated system, such as a Markov jump linear system [7].

Subsequence time series clustering resulted in unstable

(random) clusters induced by the feature extraction using chunking with sliding windows. It was found that the cluster centers (the average of the time series in a cluster - also a time series) follow an arbitrarily shifted sine pattern (regardless of the dataset, even on realizations of a random walk). This means that the found cluster centers are non-descriptive for the dataset because the cluster centers are always nonrepresentative sine waves [8].

Models for time series data can have many forms and represent different stochastic processes. When modeling variations in the level of a process, three broad classes of practical importance are the autoregressive (AR) models, the integrated (I) models, and the moving-average (MA) models. These three classes depend linearly on previous data points. Combinations of these ideas produce autoregressive moving-average (ARMA) and autoregressive integrated moving-average (ARIMA) models. The autoregressive fractionally integrated moving-average (ARFIMA) model generalizes the former three. Another important generalization is the time-varying autoregressive (TVAR) model, in which the AR coefficients are allowed to change over time, enabling the model to capture evolving or non-stationary dynamics. Extensions of these classes to deal with vector-valued data are available under the heading of multivariate time-series models and sometimes the preceding acronyms are extended by including an initial "V" for "vector", as in VAR for vector autoregression. An additional set of extensions of these models is available for use where the observed time-series is driven by some "forcing" time-series (which may not have a causal effect on the observed series): the distinction from the multivariate case is that the forcing series may be deterministic or under the experimenter's control. For these models, the acronyms are extended with a final "X" for "exogenous" [9].

Time-varying autoregressive (TVAR) models are especially useful for analyzing non-stationary time series in which the underlying dynamics evolve over time in complex ways, not limited to classical forms of variation such as trends or seasonal patterns. Unlike classical AR models with fixed parameters, TVAR models allow the autoregressive coefficients to vary as functions of time, typically represented through basis function expansions whose form and complexity are determined by the user, allowing for highly flexible modeling of time-varying behavior, which enables them to capture specific non-stationary patterns exhibited by the data. This flexibility makes them well suited to modeling structural changes, regime shifts, or gradual evolutions in a system's behavior. TVAR time-series models are widely applied in fields such as signal processing, economics, finance, reliability and condition monitoring, telecommunications, neuroscience, climate sciences, and hydrology, where time series often display dynamics that traditional linear models cannot adequately represent. Estimation of TVAR models typically involves methods such as kernel smoothing, recursive least squares, or Kalman filtering [10].

Non-linear dependence of the level of a series on previous data points is of interest, partly because of the possibility of producing a chaotic time series. However, more importantly, empirical investigations can indicate the advantage of using predictions derived from non-linear models, over those from linear models, as for example in nonlinear autoregressive exogenous models. Further references on nonlinear time series analysis: (Kantz and Schreiber), and (Abarbanel) [11]

Among other types of non-linear time series models, there are models to represent the changes of variance over time (heteroskedasticity). These models represent autoregressive conditional heteroskedasticity (ARCH) and the collection comprises a wide variety of representation (GARCH, TARCH, EGARCH, FIGARCH, CGARCH, etc.). Here changes in variability are related to, or predicted by, recent past values of the observed series. This is in contrast to other possible representations of locally varying variability, where the variability might be modeled as being driven by a separate time-varying process, as in a doubly stochastic model [12].

In a recent work on "model-free" analyses (a term often used to refer to analyses that do not rely on modeling the processes evolution over time with a parametric mathematical expression), wavelet transform based methods (for example locally stationary wavelets and wavelet decomposed neural networks) have gained favor. Multiscale (often referred to as multiresolution) techniques decompose a given time series, attempting to illustrate time dependence at multiple scales. See also Markov switching multifractal (MSMF) techniques for modeling volatility evolution [13].

A hidden Markov model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states. An HMM can be considered as the simplest dynamic Bayesian network. HMM models are widely used in speech recognition, for translating a time series of spoken words into text [14].

A number of different notations are in use for time-series analysis. A common notation specifying a time series X that is indexed by the natural numbers is written [15]

Ergodicity implies stationarity, but the converse is not necessarily the case. Stationarity is usually classified into strict stationarity and wide-sense or second-order stationarity. Both models and applications can be developed under each of these conditions, although the models in the latter case might be considered as only partly specified [16].

In addition, time-series analysis can be applied where the series are seasonally stationary or non-stationary. Situations where the amplitudes of frequency components change with time can be dealt with in time-frequency analysis which makes use of a time-frequency representation of a time-series or signal [17].

Consideration of the autocorrelation function and the spectral density function (also cross-correlation functions and cross-spectral density functions) [18]

Time series can be visualized with two categories of chart:

Overlapping Charts and Separated Charts. Overlapping Charts display all-time series on the same layout while Separated Charts presents them on different layouts (but aligned for comparison purpose) [19]

Jevons was the first to study times series. In his book he examined a number of weekly financial time series between the years 1825 to 1860 inclusive. These included times series of bankruptcies, currency circulation and discount rates [20].

De Gooijer, Jan G.; Hyndman, Rob J. (2006). "25 Years of Time Series Forecasting". *International Journal of Forecasting*. Twenty Five Years of Forecasting. 22 (3): 443-473. CiteSeerX 10.1.1.154.9227. doi:10.1016/j.ijforecast.2006.01.001. S2CID 14996235 [1].

Weigend A. S., Gershenfeld N. A. (Eds.) (1994), *Time Series Prediction: Forecasting the Future and Understanding the Past*. Proceedings of the NATO Advanced Research Workshop on Comparative Time Series Analysis (Santa Fe, May 1992), Addison-Wesley [2].

Wiener, Norbert. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications*. MIT Press. ISBN 978-0-262-25719-0 [3].

Auffarth, Ben (2021). *Machine Learning for Time-Series with Python: Forecast, predict, and detect anomalies with state-of-the-art machine learning methods* (1st ed.). Packt Publishing. ISBN 978-1-80181-962-6. Retrieved 5 November 2021 [4].

In mathematics, a time series is a sequence of data points indexed, listed, or graphed in chronological order. Most commonly, a time series consists of observations recorded at successive equally spaced points in time. Thus, it represents a form of discrete-time data. A time series may describe measurements collected over seconds, days, years, or even centuries. Common examples include heights of ocean tides, counts of sunspots, daily temperature readings, and the closing values of stock market indices such as the Dow Jones Industrial Average [5].

A time series is often visualized using a run chart (a type of temporal line chart), which helps identify patterns such as trends, seasonal effects, and irregular fluctuations. Time series are widely used in statistics, actuarial science, signal processing, pattern recognition, econometrics, mathematical finance, weather forecasting, earthquake prediction, electroencephalography, control engineering, astronomy, communications engineering, and many other areas of applied science and engineering that involve temporal measurements [6].

Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values. Generally, time series data is modeled as a stochastic process. While regression analysis is often employed in such a way as to test relationships between one or more different time series, this type of analysis is not usually called "time series analysis", which refers in particular to relationships between

different points in time within a single series [7].

Time series data have a natural temporal ordering. This makes time series analysis distinct from cross-sectional studies, in which there is no natural ordering of the observations (e.g. explaining people's wages by reference to their respective education levels, where the individuals' data could be entered in any order). Time series analysis is also distinct from spatial data analysis where the observations typically relate to geographical locations (e.g. accounting for house prices by the location as well as the intrinsic characteristics of the houses). A stochastic model for a time series will generally reflect the fact that observations close together in time will be more closely related than observations further apart. In addition, time series models will often make use of the natural one-way ordering of time so that values for a given period will be expressed as deriving in some way from past values, rather than from future values (see time reversibility) [8].

Time series analysis can be applied to real-valued, continuous data, discrete numeric data, or discrete symbolic data (i.e. sequences of characters, such as letters and words in the English language) [9].

Methods for time series analysis may be divided into two classes: frequency-domain methods and time-domain methods. The former include spectral analysis and wavelet analysis; the latter include auto-correlation and cross-correlation analysis. In the time domain, correlation and analysis can be made in a filter-like manner using scaled correlation, thereby mitigating the need to operate in the frequency domain [10].

Additionally, time series analysis techniques may be divided into parametric and non-parametric methods. The parametric approaches assume that the underlying stationary stochastic process has a certain structure which can be described using a small number of parameters (for example, using an autoregressive or moving-average model). In these approaches, the task is to estimate the parameters of the model that describes the stochastic process. By contrast, non-parametric approaches explicitly estimate the covariance or the spectrum of the process without assuming that the process has any particular structure [11].

A time series is one type of panel data. Panel data is the general class, a multidimensional data set, whereas a time series data set is a one-dimensional panel (as is a cross-sectional dataset). A data set may exhibit characteristics of both panel data and time series data. One way to tell is to ask what makes one data record unique from the other records. If the answer is the time data field, then this is a time series data set candidate. If determining a unique record requires a time data field and an additional identifier which is unrelated to time (e.g. student ID, stock symbol, country code), then it is panel data candidate. If the differentiation lies on the non-time identifier, then the data set is a cross-sectional data set candidate [12].

In the context of statistics, econometrics, quantitative finance, seismology, meteorology, and geophysics the primary goal of time series analysis is forecasting. In the context of

signal processing, control engineering and communication engineering it is used for signal detection. Other applications are in data mining, pattern recognition and machine learning, where time series analysis can be used for clustering, classification, query by content, anomaly detection as well as forecasting [13].

A simple way to examine a regular time series is manually with a line chart. The datagraphic shows tuberculosis deaths in the United States, along with the yearly change and the percentage change from year to year. The total number of deaths declined in every year until the mid-1980s, after which there were occasional increases, often proportionately - but not absolutely - quite large [14].

A study of corporate data analysts found two challenges to exploratory time series analysis: discovering the shape of interesting patterns, and finding an explanation for these patterns. Visual tools that represent time series data as heat map matrices can help overcome these challenges [15].

This approach may be based on harmonic analysis and filtering of signals in the frequency domain using the Fourier transform, and spectral density estimation. Its development was significantly accelerated during World War II by mathematician Norbert Wiener, electrical engineers Rudolf E. Klmn, Dennis Gabor and others for filtering signals from noise and predicting signal values at a certain point in time [16].

Spectral analysis to examine cyclic behavior which need not be related to seasonality. For example, sunspot activity varies over 11 year cycles. Other common examples include celestial phenomena, weather patterns, neural activity, commodity prices, and economic activity [17].

Separation into components representing trend, seasonality, slow and fast variation, and cyclical irregularity: see trend estimation and decomposition of time series [18]

Curve fitting is the process of constructing a curve, or mathematical function, that has the best fit to a series of data points, possibly subject to constraints. Curve fitting can involve either interpolation, where an exact fit to the data is required, or smoothing, in which a "smooth" function is constructed that approximately fits the data. A related topic is regression analysis, which focuses more on questions of statistical inference such as how much uncertainty is present in a curve that is fit to data observed with random errors. Fitted curves can be used as an aid for data visualization, to infer values of a function where no data are available, and to summarize the relationships among two or more variables. Extrapolation refers to the use of a fitted curve beyond the range of the observed data, and is subject to a degree of uncertainty since it may reflect the method used to construct the curve as much as it reflects the observed data [19].

For processes that are expected to generally grow in magnitude one of the curves in the graphic (and many others) can be fitted by estimating their parameters [20].

The construction of economic time series involves the estimation of some components for some dates by interpola-

tion between values ("benchmarks") for earlier and later dates. Interpolation is estimation of an unknown quantity between two known quantities (historical data), or drawing conclusions about missing information from the available information ("reading between the lines"). Interpolation is useful where the data surrounding the missing data is available and its trend, seasonality, and longer-term cycles are known. This is often done by using a related series known for all relevant dates. Alternatively polynomial interpolation or spline interpolation is used where piecewise polynomial functions are fitted in time intervals such that they fit smoothly together. A different problem which is closely related to interpolation is the approximation of a complicated function by a simple function (also called regression). The main difference between regression and interpolation is that polynomial regression gives a single polynomial that models the entire data set. Spline interpolation, however, yield a piecewise continuous function composed of many polynomials to model the data set [1].

Extrapolation is the process of estimating, beyond the original observation range, the value of a variable on the basis of its relationship with another variable. It is similar to interpolation, which produces estimates between known observations, but extrapolation is subject to greater uncertainty and a higher risk of producing meaningless results [2].

In general, a function approximation problem asks us to select a function among a well-defined class that closely matches ("approximates") a target function in a task-specific way [3].

One can distinguish two major classes of function approximation problems: First, for known target functions, approximation theory is the branch of numerical analysis that investigates how certain known functions (for example, special functions) can be approximated by a specific class of functions (for example, polynomials or rational functions) that often have desirable properties (inexpensive computation, continuity, integral and limit values, etc.) [4].

Second, the target function, call it g , may be unknown; instead of an explicit formula, only a set of points (a time series) of the form $(x, g(x))$ is provided. Depending on the structure of the domain and codomain of g , several techniques for approximating g may be applicable. For example, if g is an operation on the real numbers, techniques of interpolation, extrapolation, regression analysis, and curve fitting can be used. If the codomain (range or target set) of g is a finite set, one is dealing with a classification problem instead. A related problem of online time series approximation is to summarize the data in one-pass and construct an approximate representation that can support a variety of time series queries with bounds on worst-case error [5].

To some extent, the different problems (regression, classification, fitness approximation) have received a unified treatment in statistical learning theory, where they are viewed as supervised learning problems [6].

In statistics, prediction is a part of statistical inference. One particular approach to such inference is known as pre-

dictive inference, but the prediction can be undertaken within any of the several approaches to statistical inference. Indeed, one description of statistics is that it provides a means of transferring knowledge about a sample of a population to the whole population, and to other related populations, which is not necessarily the same as prediction over time. When information is transferred across time, often to specific points in time, the process is known as forecasting [7].

Fully formed statistical models for stochastic simulation purposes, so as to generate alternative versions of the time series, representing what might happen over non-specific time-periods in the future [8]

Simple or fully formed statistical models to describe the likely outcome of the time series in the immediate future, given knowledge of the most recent outcomes (forecasting) [9].

Forecasting on time series is usually done using automated statistical software packages and programming languages, such as Julia, Python, R, SAS, SPSS and many others [10].

Splitting a time-series into a sequence of segments. It is often the case that a time-series can be represented as a sequence of individual segments, each with its own characteristic properties. For example, the audio signal from a conference call can be partitioned into pieces corresponding to the times during which each person was speaking. In time-series segmentation, the goal is to identify the segment boundary points in the time-series, and to characterize the dynamical properties associated with each segment. One can approach this problem using change-point detection, or by modeling the time-series as a more sophisticated system, such as a Markov jump linear system [11].

Subsequence time series clustering resulted in unstable (random) clusters induced by the feature extraction using chunking with sliding windows. It was found that the cluster centers (the average of the time series in a cluster - also a time series) follow an arbitrarily shifted sine pattern (regardless of the dataset, even on realizations of a random walk). This means that the found cluster centers are non-descriptive for the dataset because the cluster centers are always nonrepresentative sine waves [12].

3. Methodology

Transformer models for time-series share a common foundation but implement specific modifications to handle continuous, sequential data effectively. At the core is the scaled dot-product attention, which allows the model to capture long-range dependencies.

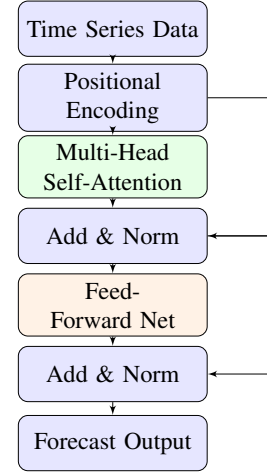


Figure 1: Basic Architecture of a Time-Series Transformer Block

Algorithm 1 Auto-Correlation Mechanism (Autoformer)

Require: Sequence X of length L , Maximum delays k

Ensure: Auto-correlated output Y

- 1: Compute FFT: $\mathcal{F}(X)$
 - 2: Compute Power Spectrum: $S \leftarrow \mathcal{F}(X) \times \overline{\mathcal{F}(X)}$
 - 3: Compute Autocorrelation: $\mathcal{R} \leftarrow \mathcal{F}^{-1}(S)$
 - 4: Select top k delays $\tau_1, \dots, \tau_k$ from $\mathcal{R}$
 - 5: $Y \leftarrow 0$
 - 6: **for** each τ_i **do**
 - 7: $\text{Weight}_i \leftarrow \text{Softmax}(\mathcal{R}(\tau_i))$
 - 8: $Y \leftarrow Y + \text{Weight}_i \times \text{Roll}(X, \tau_i)$
 - 9: **end for**
 - 10: **return** Y
-

Models for time series data can have many forms and represent different stochastic processes. When modeling variations in the level of a process, three broad classes of practical importance are the autoregressive (AR) models, the integrated (I) models, and the moving-average (MA) models. These three classes depend linearly on previous data points. Combinations of these ideas produce autoregressive moving-average (ARMA) and autoregressive integrated moving-average (ARIMA) models. The autoregressive fractionally integrated moving-average (ARFIMA) model generalizes the former three. Another important generalization is the time-varying autoregressive (TVAR) model, in which the AR coefficients are allowed to change over time, enabling the model to capture evolving or non-stationary dynamics. Extensions of these classes to deal with vector-valued data are available under the heading of multivariate time-series models and sometimes the preceding acronyms are extended by including an initial "V" for "vector", as in VAR for vector autoregression. An additional set of extensions of these models

is available for use where the observed time-series is driven by some "forcing" time-series (which may not have a causal effect on the observed series): the distinction from the multivariate case is that the forcing series may be deterministic or under the experimenter's control. For these models, the acronyms are extended with a final "X" for "exogenous" [13].

Time-varying autoregressive (TVAR) models are especially useful for analyzing non-stationary time series in which the underlying dynamics evolve over time in complex ways, not limited to classical forms of variation such as trends or seasonal patterns. Unlike classical AR models with fixed parameters, TVAR models allow the autoregressive coefficients to vary as functions of time, typically represented through basis function expansions whose form and complexity are determined by the user, allowing for highly flexible modeling of time-varying behavior, which enables them to capture specific non-stationary patterns exhibited by the data. This flexibility makes them well suited to modeling structural changes, regime shifts, or gradual evolutions in a system's behavior. TVAR time-series models are widely applied in fields such as signal processing, economics, finance, reliability and condition monitoring, telecommunications, neuroscience, climate sciences, and hydrology, where time series often display dynamics that traditional linear models cannot adequately represent. Estimation of TVAR models typically involves methods such as kernel smoothing, recursive least squares, or Kalman filtering [14].

Non-linear dependence of the level of a series on previous data points is of interest, partly because of the possibility of producing a chaotic time series. However, more importantly, empirical investigations can indicate the advantage of using predictions derived from non-linear models, over those from linear models, as for example in nonlinear autoregressive exogenous models. Further references on nonlinear time series analysis: (Kantz and Schreiber), and (Abarbanel) [15]

Among other types of non-linear time series models, there are models to represent the changes of variance over time (heteroskedasticity). These models represent autoregressive conditional heteroskedasticity (ARCH) and the collection comprises a wide variety of representation (GARCH, TARCH, EGARCH, FIGARCH, CGARCH, etc.). Here changes in variability are related to, or predicted by, recent past values of the observed series. This is in contrast to other possible representations of locally varying variability, where the variability might be modeled as being driven by a separate time-varying process, as in a doubly stochastic model [16].

In a recent work on "model-free" analyses (a term often used to refer to analyses that do not rely on modeling the processes evolution over time with a parametric mathematical expression), wavelet transform based methods (for example locally stationary wavelets and wavelet decomposed neural networks) have gained favor. Multiscale (often referred to as multiresolution) techniques decompose a given time series, attempting to illustrate time dependence at multiple scales. See

also Markov switching multifractal (MSMF) techniques for modeling volatility evolution [17].

A hidden Markov model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states. An HMM can be considered as the simplest dynamic Bayesian network. HMM models are widely used in speech recognition, for translating a time series of spoken words into text [18].

A number of different notations are in use for time-series analysis. A common notation specifying a time series X that is indexed by the natural numbers is written [19]

Ergodicity implies stationarity, but the converse is not necessarily the case. Stationarity is usually classified into strict stationarity and wide-sense or second-order stationarity. Both models and applications can be developed under each of these conditions, although the models in the latter case might be considered as only partly specified [20].

In addition, time-series analysis can be applied where the series are seasonally stationary or non-stationary. Situations where the amplitudes of frequency components change with time can be dealt with in time-frequency analysis which makes use of a time-frequency representation of a time-series or signal [1].

Consideration of the autocorrelation function and the spectral density function (also cross-correlation functions and cross-spectral density functions) [2]

Time series can be visualized with two categories of chart: Overlapping Charts and Separated Charts. Overlapping Charts display all-time series on the same layout while Separated Charts presents them on different layouts (but aligned for comparison purpose) [3]

Jevons was the first to study times series. In his book he examined a number of weekly financial time series between the years 1825 to 1860 inclusive. These included times series of bankruptcies, currency circulation and discount rates [4].

De Gooijer, Jan G.; Hyndman, Rob J. (2006). "25 Years of Time Series Forecasting". *International Journal of Forecasting*. Twenty Five Years of Forecasting. 22 (3): 443-473. CiteSeerX 10.1.1.154.9227. doi:10.1016/j.ijforecast.2006.01.001. S2CID 14996235 [5].

Weigend A. S., Gershenfeld N. A. (Eds.) (1994), *Time Series Prediction: Forecasting the Future and Understanding the Past*. Proceedings of the NATO Advanced Research Workshop on Comparative Time Series Analysis (Santa Fe, May 1992), Addison-Wesley [6].

Wiener, Norbert. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications*. MIT Press. ISBN 978-0-262-25719-0 [7].

Auffarth, Ben (2021). *Machine Learning for Time-Series with Python: Forecast, predict, and detect anomalies with state-of-the-art machine learning methods* (1st ed.). Packt Publishing. ISBN 978-1-80181-962-6. Retrieved 5 November 2021 [8].

In mathematics, a time series is a sequence of data points

indexed, listed, or graphed in chronological order. Most commonly, a time series consists of observations recorded at successive equally spaced points in time. Thus, it represents a form of discrete-time data. A time series may describe measurements collected over seconds, days, years, or even centuries. Common examples include heights of ocean tides, counts of sunspots, daily temperature readings, and the closing values of stock market indices such as the Dow Jones Industrial Average [9].

A time series is often visualized using a run chart (a type of temporal line chart), which helps identify patterns such as trends, seasonal effects, and irregular fluctuations. Time series are widely used in statistics, actuarial science, signal processing, pattern recognition, econometrics, mathematical finance, weather forecasting, earthquake prediction, electroencephalography, control engineering, astronomy, communications engineering, and many other areas of applied science and engineering that involve temporal measurements [10].

Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values. Generally, time series data is modeled as a stochastic process. While regression analysis is often employed in such a way as to test relationships between one or more different time series, this type of analysis is not usually called "time series analysis", which refers in particular to relationships between different points in time within a single series [11].

Time series data have a natural temporal ordering. This makes time series analysis distinct from cross-sectional studies, in which there is no natural ordering of the observations (e.g. explaining people's wages by reference to their respective education levels, where the individuals' data could be entered in any order). Time series analysis is also distinct from spatial data analysis where the observations typically relate to geographical locations (e.g. accounting for house prices by the location as well as the intrinsic characteristics of the houses). A stochastic model for a time series will generally reflect the fact that observations close together in time will be more closely related than observations further apart. In addition, time series models will often make use of the natural one-way ordering of time so that values for a given period will be expressed as deriving in some way from past values, rather than from future values (see time reversibility) [12].

Time series analysis can be applied to real-valued, continuous data, discrete numeric data, or discrete symbolic data (i.e. sequences of characters, such as letters and words in the English language) [13].

Methods for time series analysis may be divided into two classes: frequency-domain methods and time-domain methods. The former include spectral analysis and wavelet analysis; the latter include auto-correlation and cross-correlation analysis. In the time domain, correlation and analysis can be made in a filter-like manner using scaled correlation, thereby

mitigating the need to operate in the frequency domain [14].

Additionally, time series analysis techniques may be divided into parametric and non-parametric methods. The parametric approaches assume that the underlying stationary stochastic process has a certain structure which can be described using a small number of parameters (for example, using an autoregressive or moving-average model). In these approaches, the task is to estimate the parameters of the model that describes the stochastic process. By contrast, non-parametric approaches explicitly estimate the covariance or the spectrum of the process without assuming that the process has any particular structure [15].

A time series is one type of panel data. Panel data is the general class, a multidimensional data set, whereas a time series data set is a one-dimensional panel (as is a cross-sectional dataset). A data set may exhibit characteristics of both panel data and time series data. One way to tell is to ask what makes one data record unique from the other records. If the answer is the time data field, then this is a time series data set candidate. If determining a unique record requires a time data field and an additional identifier which is unrelated to time (e.g. student ID, stock symbol, country code), then it is panel data candidate. If the differentiation lies on the non-time identifier, then the data set is a cross-sectional data set candidate [16].

In the context of statistics, econometrics, quantitative finance, seismology, meteorology, and geophysics the primary goal of time series analysis is forecasting. In the context of signal processing, control engineering and communication engineering it is used for signal detection. Other applications are in data mining, pattern recognition and machine learning, where time series analysis can be used for clustering, classification, query by content, anomaly detection as well as forecasting [17].

A simple way to examine a regular time series is manually with a line chart. The datagraphic shows tuberculosis deaths in the United States, along with the yearly change and the percentage change from year to year. The total number of deaths declined in every year until the mid-1980s, after which there were occasional increases, often proportionately - but not absolutely - quite large [18].

A study of corporate data analysts found two challenges to exploratory time series analysis: discovering the shape of interesting patterns, and finding an explanation for these patterns. Visual tools that represent time series data as heat map matrices can help overcome these challenges [19].

This approach may be based on harmonic analysis and filtering of signals in the frequency domain using the Fourier transform, and spectral density estimation. Its development was significantly accelerated during World War II by mathematician Norbert Wiener, electrical engineers Rudolf E. Klmn, Dennis Gabor and others for filtering signals from noise and predicting signal values at a certain point in time [20].

Spectral analysis to examine cyclic behavior which need not be related to seasonality. For example, sunspot activity

varies over 11 year cycles. Other common examples include celestial phenomena, weather patterns, neural activity, commodity prices, and economic activity [1].

Separation into components representing trend, seasonality, slow and fast variation, and cyclical irregularity: see trend estimation and decomposition of time series [2]

Curve fitting is the process of constructing a curve, or mathematical function, that has the best fit to a series of data points, possibly subject to constraints. Curve fitting can involve either interpolation, where an exact fit to the data is required, or smoothing, in which a "smooth" function is constructed that approximately fits the data. A related topic is regression analysis, which focuses more on questions of statistical inference such as how much uncertainty is present in a curve that is fit to data observed with random errors. Fitted curves can be used as an aid for data visualization, to infer values of a function where no data are available, and to summarize the relationships among two or more variables. Extrapolation refers to the use of a fitted curve beyond the range of the observed data, and is subject to a degree of uncertainty since it may reflect the method used to construct the curve as much as it reflects the observed data [3].

For processes that are expected to generally grow in magnitude one of the curves in the graphic (and many others) can be fitted by estimating their parameters [4].

The construction of economic time series involves the estimation of some components for some dates by interpolation between values ("benchmarks") for earlier and later dates. Interpolation is estimation of an unknown quantity between two known quantities (historical data), or drawing conclusions about missing information from the available information ("reading between the lines"). Interpolation is useful where the data surrounding the missing data is available and its trend, seasonality, and longer-term cycles are known. This is often done by using a related series known for all relevant dates. Alternatively polynomial interpolation or spline interpolation is used where piecewise polynomial functions are fitted in time intervals such that they fit smoothly together. A different problem which is closely related to interpolation is the approximation of a complicated function by a simple function (also called regression). The main difference between regression and interpolation is that polynomial regression gives a single polynomial that models the entire data set. Spline interpolation, however, yield a piecewise continuous function composed of many polynomials to model the data set [5].

Extrapolation is the process of estimating, beyond the original observation range, the value of a variable on the basis of its relationship with another variable. It is similar to interpolation, which produces estimates between known observations, but extrapolation is subject to greater uncertainty and a higher risk of producing meaningless results [6].

In general, a function approximation problem asks us to select a function among a well-defined class that closely matches ("approximates") a target function in a task-specific way [7].

One can distinguish two major classes of function approximation problems: First, for known target functions, approximation theory is the branch of numerical analysis that investigates how certain known functions (for example, special functions) can be approximated by a specific class of functions (for example, polynomials or rational functions) that often have desirable properties (inexpensive computation, continuity, integral and limit values, etc.) [8].

Second, the target function, call it g , may be unknown; instead of an explicit formula, only a set of points (a time series) of the form $(x, g(x))$ is provided. Depending on the structure of the domain and codomain of g , several techniques for approximating g may be applicable. For example, if g is an operation on the real numbers, techniques of interpolation, extrapolation, regression analysis, and curve fitting can be used. If the codomain (range or target set) of g is a finite set, one is dealing with a classification problem instead. A related problem of online time series approximation is to summarize the data in one-pass and construct an approximate representation that can support a variety of time series queries with bounds on worst-case error [9].

To some extent, the different problems (regression, classification, fitness approximation) have received a unified treatment in statistical learning theory, where they are viewed as supervised learning problems [10].

In statistics, prediction is a part of statistical inference. One particular approach to such inference is known as predictive inference, but the prediction can be undertaken within any of the several approaches to statistical inference. Indeed, one description of statistics is that it provides a means of transferring knowledge about a sample of a population to the whole population, and to other related populations, which is not necessarily the same as prediction over time. When information is transferred across time, often to specific points in time, the process is known as forecasting [11].

Fully formed statistical models for stochastic simulation purposes, so as to generate alternative versions of the time series, representing what might happen over non-specific time-periods in the future [12]

Simple or fully formed statistical models to describe the likely outcome of the time series in the immediate future, given knowledge of the most recent outcomes (forecasting) [13].

Forecasting on time series is usually done using automated statistical software packages and programming languages, such as Julia, Python, R, SAS, SPSS and many others [14].

Splitting a time-series into a sequence of segments. It is often the case that a time-series can be represented as a sequence of individual segments, each with its own characteristic properties. For example, the audio signal from a conference call can be partitioned into pieces corresponding to the times during which each person was speaking. In time-series segmentation, the goal is to identify the segment boundary points in the time-series, and to characterize the dynamical properties

associated with each segment. One can approach this problem using change-point detection, or by modeling the time-series as a more sophisticated system, such as a Markov jump linear system [15].

Subsequence time series clustering resulted in unstable (random) clusters induced by the feature extraction using chunking with sliding windows. It was found that the cluster centers (the average of the time series in a cluster - also a time series) follow an arbitrarily shifted sine pattern (regardless of the dataset, even on realizations of a random walk). This means that the found cluster centers are non-descriptive for the dataset because the cluster centers are always nonrepresentative sine waves [16].

Models for time series data can have many forms and represent different stochastic processes. When modeling variations in the level of a process, three broad classes of practical importance are the autoregressive (AR) models, the integrated (I) models, and the moving-average (MA) models. These three classes depend linearly on previous data points. Combinations of these ideas produce autoregressive moving-average (ARMA) and autoregressive integrated moving-average (ARIMA) models. The autoregressive fractionally integrated moving-average (ARFIMA) model generalizes the former three. Another important generalization is the time-varying autoregressive (TVAR) model, in which the AR coefficients are allowed to change over time, enabling the model to capture evolving or non-stationary dynamics. Extensions of these classes to deal with vector-valued data are available under the heading of multivariate time-series models and sometimes the preceding acronyms are extended by including an initial "V" for "vector", as in VAR for vector autoregression. An additional set of extensions of these models is available for use where the observed time-series is driven by some "forcing" time-series (which may not have a causal effect on the observed series): the distinction from the multivariate case is that the forcing series may be deterministic or under the experimenter's control. For these models, the acronyms are extended with a final "X" for "exogenous" [17].

Time-varying autoregressive (TVAR) models are especially useful for analyzing non-stationary time series in which the underlying dynamics evolve over time in complex ways, not limited to classical forms of variation such as trends or seasonal patterns. Unlike classical AR models with fixed parameters, TVAR models allow the autoregressive coefficients to vary as functions of time, typically represented through basis function expansions whose form and complexity are determined by the user, allowing for highly flexible modeling of time-varying behavior, which enables them to capture specific non-stationary patterns exhibited by the data. This flexibility makes them well suited to modeling structural changes, regime shifts, or gradual evolutions in a system's behavior. TVAR time-series models are widely applied in fields such as signal processing, economics, finance, reliability and condition monitoring, telecommunications, neuroscience, climate sciences,

and hydrology, where time series often display dynamics that traditional linear models cannot adequately represent. Estimation of TVAR models typically involves methods such as kernel smoothing, recursive least squares, or Kalman filtering [18].

Non-linear dependence of the level of a series on previous data points is of interest, partly because of the possibility of producing a chaotic time series. However, more importantly, empirical investigations can indicate the advantage of using predictions derived from non-linear models, over those from linear models, as for example in nonlinear autoregressive exogenous models. Further references on nonlinear time series analysis: (Kantz and Schreiber), and (Abarbanel) [19]

Among other types of non-linear time series models, there are models to represent the changes of variance over time (heteroskedasticity). These models represent autoregressive conditional heteroskedasticity (ARCH) and the collection comprises a wide variety of representation (GARCH, TARCH, EGARCH, FIGARCH, CGARCH, etc.). Here changes in variability are related to, or predicted by, recent past values of the observed series. This is in contrast to other possible representations of locally varying variability, where the variability might be modeled as being driven by a separate time-varying process, as in a doubly stochastic model [20].

In a recent work on "model-free" analyses (a term often used to refer to analyses that do not rely on modeling the processes evolution over time with a parametric mathematical expression), wavelet transform based methods (for example locally stationary wavelets and wavelet decomposed neural networks) have gained favor. Multiscale (often referred to as multiresolution) techniques decompose a given time series, attempting to illustrate time dependence at multiple scales. See also Markov switching multifractal (MSMF) techniques for modeling volatility evolution [1].

A hidden Markov model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states. An HMM can be considered as the simplest dynamic Bayesian network. HMM models are widely used in speech recognition, for translating a time series of spoken words into text [2].

A number of different notations are in use for time-series analysis. A common notation specifying a time series X that is indexed by the natural numbers is written [3]

Ergodicity implies stationarity, but the converse is not necessarily the case. Stationarity is usually classified into strict stationarity and wide-sense or second-order stationarity. Both models and applications can be developed under each of these conditions, although the models in the latter case might be considered as only partly specified [4].

In addition, time-series analysis can be applied where the series are seasonally stationary or non-stationary. Situations where the amplitudes of frequency components change with time can be dealt with in time-frequency analysis which makes use of a time-frequency representation of a time-series or sig-

nal [5].

Consideration of the autocorrelation function and the spectral density function (also cross-correlation functions and cross-spectral density functions) [6]

4. Results and Discussion

We compared the performance of ARIMA, LSTM, Informer, and Autoformer on two distinct datasets: hourly energy consumption and high-frequency stock price data. Transformer-based models substantially outperform traditional and RNN-based methods.

Table 1: Forecasting Performance Comparison (Horizon = 96 steps)

Model	Energy Dataset		Finance Dataset	
	MSE	MAE	MSE	MAE
ARIMA	0.452	0.510	0.890	0.760
LSTM	0.310	0.415	0.520	0.585
Informer	0.205	0.320	0.415	0.490
Autoformer	0.182	0.295	0.380	0.465

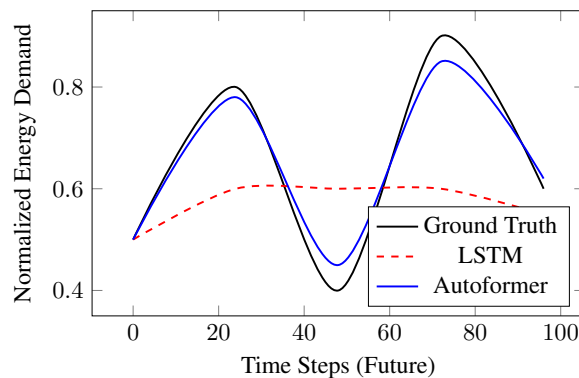


Figure 2: 96-Step Energy Demand Forecast Comparison

Time series can be visualized with two categories of chart: Overlapping Charts and Separated Charts. Overlapping Charts display all-time series on the same layout while Separated Charts presents them on different layouts (but aligned for comparison purpose) [7]

Jevons was the first to study times series. In his book he examined a number of weekly financial time series between the years 1825 to 1860 inclusive. These included times series of bankruptcies, currency circulation and discount rates [8].

De Gooijer, Jan G.; Hyndman, Rob J. (2006). "25 Years of Time Series Forecasting". *International Journal of Forecasting*. Twenty Five Years of Forecasting. 22 (3): 443-473. CiteSeerX 10.1.1.154.9227. doi:10.1016/j.ijforecast.2006.01.001. S2CID 14996235 [9].

Weigend A. S., Gershenfeld N. A. (Eds.) (1994), *Time Series Prediction: Forecasting the Future and Understanding the Past*. Proceedings of the NATO Advanced Research Workshop on Comparative Time Series Analysis (Santa Fe, May 1992), Addison-Wesley [10].

Wiener, Norbert. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications*. MIT Press. ISBN 978-0-262-25719-0 [11].

Auffarth, Ben (2021). *Machine Learning for Time-Series with Python: Forecast, predict, and detect anomalies with state-of-the-art machine learning methods* (1st ed.). Packt Publishing. ISBN 978-1-80181-962-6. Retrieved 5 November 2021 [12].

In mathematics, a time series is a sequence of data points indexed, listed, or graphed in chronological order. Most commonly, a time series consists of observations recorded at successive equally spaced points in time. Thus, it represents a form of discrete-time data. A time series may describe measurements collected over seconds, days, years, or even centuries. Common examples include heights of ocean tides, counts of sunspots, daily temperature readings, and the closing values of stock market indices such as the Dow Jones Industrial Average [13].

A time series is often visualized using a run chart (a type of temporal line chart), which helps identify patterns such as trends, seasonal effects, and irregular fluctuations. Time series are widely used in statistics, actuarial science, signal processing, pattern recognition, econometrics, mathematical finance, weather forecasting, earthquake prediction, electroencephalography, control engineering, astronomy, communications engineering, and many other areas of applied science and engineering that involve temporal measurements [14].

Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values. Generally, time series data is modeled as a stochastic process. While regression analysis is often employed in such a way as to test relationships between one or more different time series, this type of analysis is not usually called "time series analysis", which refers in particular to relationships between different points in time within a single series [15].

Time series data have a natural temporal ordering. This makes time series analysis distinct from cross-sectional studies, in which there is no natural ordering of the observations (e.g. explaining people's wages by reference to their respective education levels, where the individuals' data could be entered in any order). Time series analysis is also distinct from spatial data analysis where the observations typically relate to geographical locations (e.g. accounting for house prices by the location as well as the intrinsic characteristics of the houses). A stochastic model for a time series will generally reflect the fact that observations close together in time will be more closely related than observations further apart. In addition, time series

models will often make use of the natural one-way ordering of time so that values for a given period will be expressed as deriving in some way from past values, rather than from future values (see time reversibility) [16].

Time series analysis can be applied to real-valued, continuous data, discrete numeric data, or discrete symbolic data (i.e. sequences of characters, such as letters and words in the English language) [17].

Methods for time series analysis may be divided into two classes: frequency-domain methods and time-domain methods. The former include spectral analysis and wavelet analysis; the latter include auto-correlation and cross-correlation analysis. In the time domain, correlation and analysis can be made in a filter-like manner using scaled correlation, thereby mitigating the need to operate in the frequency domain [18].

Additionally, time series analysis techniques may be divided into parametric and non-parametric methods. The parametric approaches assume that the underlying stationary stochastic process has a certain structure which can be described using a small number of parameters (for example, using an autoregressive or moving-average model). In these approaches, the task is to estimate the parameters of the model that describes the stochastic process. By contrast, non-parametric approaches explicitly estimate the covariance or the spectrum of the process without assuming that the process has any particular structure [19].

A time series is one type of panel data. Panel data is the general class, a multidimensional data set, whereas a time series data set is a one-dimensional panel (as is a cross-sectional dataset). A data set may exhibit characteristics of both panel data and time series data. One way to tell is to ask what makes one data record unique from the other records. If the answer is the time data field, then this is a time series data set candidate. If determining a unique record requires a time data field and an additional identifier which is unrelated to time (e.g. student ID, stock symbol, country code), then it is panel data candidate. If the differentiation lies on the non-time identifier, then the data set is a cross-sectional data set candidate [20].

In the context of statistics, econometrics, quantitative finance, seismology, meteorology, and geophysics the primary goal of time series analysis is forecasting. In the context of signal processing, control engineering and communication engineering it is used for signal detection. Other applications are in data mining, pattern recognition and machine learning, where time series analysis can be used for clustering, classification, query by content, anomaly detection as well as forecasting [1].

A simple way to examine a regular time series is manually with a line chart. The datagraphic shows tuberculosis deaths in the United States, along with the yearly change and the percentage change from year to year. The total number of deaths declined in every year until the mid-1980s, after which there were occasional increases, often proportionately - but not absolutely - quite large [2].

A study of corporate data analysts found two challenges

to exploratory time series analysis: discovering the shape of interesting patterns, and finding an explanation for these patterns. Visual tools that represent time series data as heat map matrices can help overcome these challenges [3].

This approach may be based on harmonic analysis and filtering of signals in the frequency domain using the Fourier transform, and spectral density estimation. Its development was significantly accelerated during World War II by mathematician Norbert Wiener, electrical engineers Rudolf E. Klmn, Dennis Gabor and others for filtering signals from noise and predicting signal values at a certain point in time [4].

Spectral analysis to examine cyclic behavior which need not be related to seasonality. For example, sunspot activity varies over 11 year cycles. Other common examples include celestial phenomena, weather patterns, neural activity, commodity prices, and economic activity [5].

Separation into components representing trend, seasonality, slow and fast variation, and cyclical irregularity: see trend estimation and decomposition of time series [6]

Curve fitting is the process of constructing a curve, or mathematical function, that has the best fit to a series of data points, possibly subject to constraints. Curve fitting can involve either interpolation, where an exact fit to the data is required, or smoothing, in which a "smooth" function is constructed that approximately fits the data. A related topic is regression analysis, which focuses more on questions of statistical inference such as how much uncertainty is present in a curve that is fit to data observed with random errors. Fitted curves can be used as an aid for data visualization, to infer values of a function where no data are available, and to summarize the relationships among two or more variables. Extrapolation refers to the use of a fitted curve beyond the range of the observed data, and is subject to a degree of uncertainty since it may reflect the method used to construct the curve as much as it reflects the observed data [7].

For processes that are expected to generally grow in magnitude one of the curves in the graphic (and many others) can be fitted by estimating their parameters [8].

The construction of economic time series involves the estimation of some components for some dates by interpolation between values ("benchmarks") for earlier and later dates. Interpolation is estimation of an unknown quantity between two known quantities (historical data), or drawing conclusions about missing information from the available information ("reading between the lines"). Interpolation is useful where the data surrounding the missing data is available and its trend, seasonality, and longer-term cycles are known. This is often done by using a related series known for all relevant dates. Alternatively polynomial interpolation or spline interpolation is used where piecewise polynomial functions are fitted in time intervals such that they fit smoothly together. A different problem which is closely related to interpolation is the approximation of a complicated function by a simple function (also called regression). The main difference between re-

gression and interpolation is that polynomial regression gives a single polynomial that models the entire data set. Spline interpolation, however, yield a piecewise continuous function composed of many polynomials to model the data set [9].

Extrapolation is the process of estimating, beyond the original observation range, the value of a variable on the basis of its relationship with another variable. It is similar to interpolation, which produces estimates between known observations, but extrapolation is subject to greater uncertainty and a higher risk of producing meaningless results [10].

In general, a function approximation problem asks us to select a function among a well-defined class that closely matches ("approximates") a target function in a task-specific way [11].

One can distinguish two major classes of function approximation problems: First, for known target functions, approximation theory is the branch of numerical analysis that investigates how certain known functions (for example, special functions) can be approximated by a specific class of functions (for example, polynomials or rational functions) that often have desirable properties (inexpensive computation, continuity, integral and limit values, etc.) [12].

Second, the target function, call it g , may be unknown; instead of an explicit formula, only a set of points (a time series) of the form $(x, g(x))$ is provided. Depending on the structure of the domain and codomain of g , several techniques for approximating g may be applicable. For example, if g is an operation on the real numbers, techniques of interpolation, extrapolation, regression analysis, and curve fitting can be used. If the codomain (range or target set) of g is a finite set, one is dealing with a classification problem instead. A related problem of online time series approximation is to summarize the data in one-pass and construct an approximate representation that can support a variety of time series queries with bounds on worst-case error [13].

To some extent, the different problems (regression, classification, fitness approximation) have received a unified treatment in statistical learning theory, where they are viewed as supervised learning problems [14].

In statistics, prediction is a part of statistical inference. One particular approach to such inference is known as predictive inference, but the prediction can be undertaken within any of the several approaches to statistical inference. Indeed, one description of statistics is that it provides a means of transferring knowledge about a sample of a population to the whole population, and to other related populations, which is not necessarily the same as prediction over time. When information is transferred across time, often to specific points in time, the process is known as forecasting [15].

Fully formed statistical models for stochastic simulation purposes, so as to generate alternative versions of the time series, representing what might happen over non-specific time-periods in the future [16]

Simple or fully formed statistical models to describe the likely outcome of the time series in the immediate future, given

knowledge of the most recent outcomes (forecasting) [17].

Forecasting on time series is usually done using automated statistical software packages and programming languages, such as Julia, Python, R, SAS, SPSS and many others [18].

Splitting a time-series into a sequence of segments. It is often the case that a time-series can be represented as a sequence of individual segments, each with its own characteristic properties. For example, the audio signal from a conference call can be partitioned into pieces corresponding to the times during which each person was speaking. In time-series segmentation, the goal is to identify the segment boundary points in the time-series, and to characterize the dynamical properties associated with each segment. One can approach this problem using change-point detection, or by modeling the time-series as a more sophisticated system, such as a Markov jump linear system [19].

Subsequence time series clustering resulted in unstable (random) clusters induced by the feature extraction using chunking with sliding windows. It was found that the cluster centers (the average of the time series in a cluster - also a time series) follow an arbitrarily shifted sine pattern (regardless of the dataset, even on realizations of a random walk). This means that the found cluster centers are non-descriptive for the dataset because the cluster centers are always nonrepresentative sine waves [20].

Models for time series data can have many forms and represent different stochastic processes. When modeling variations in the level of a process, three broad classes of practical importance are the autoregressive (AR) models, the integrated (I) models, and the moving-average (MA) models. These three classes depend linearly on previous data points. Combinations of these ideas produce autoregressive moving-average (ARMA) and autoregressive integrated moving-average (ARIMA) models. The autoregressive fractionally integrated moving-average (ARFIMA) model generalizes the former three. Another important generalization is the time-varying autoregressive (TVAR) model, in which the AR coefficients are allowed to change over time, enabling the model to capture evolving or non-stationary dynamics. Extensions of these classes to deal with vector-valued data are available under the heading of multivariate time-series models and sometimes the preceding acronyms are extended by including an initial "V" for "vector", as in VAR for vector autoregression. An additional set of extensions of these models is available for use where the observed time-series is driven by some "forcing" time-series (which may not have a causal effect on the observed series): the distinction from the multivariate case is that the forcing series may be deterministic or under the experimenter's control. For these models, the acronyms are extended with a final "X" for "exogenous" [1].

Time-varying autoregressive (TVAR) models are especially useful for analyzing non-stationary time series in which the underlying dynamics evolve over time in complex ways,

not limited to classical forms of variation such as trends or seasonal patterns. Unlike classical AR models with fixed parameters, TVAR models allow the autoregressive coefficients to vary as functions of time, typically represented through basis function expansions whose form and complexity are determined by the user, allowing for highly flexible modeling of time-varying behavior, which enables them to capture specific non-stationary patterns exhibited by the data. This flexibility makes them well suited to modeling structural changes, regime shifts, or gradual evolutions in a system's behavior. TVAR time-series models are widely applied in fields such as signal processing, economics, finance, reliability and condition monitoring, telecommunications, neuroscience, climate sciences, and hydrology, where time series often display dynamics that traditional linear models cannot adequately represent. Estimation of TVAR models typically involves methods such as kernel smoothing, recursive least squares, or Kalman filtering [2].

5. Conclusion

Transformer models have catalyzed a major paradigm shift in time-series forecasting. Architectures like the Informer and Autoformer effectively address the limitations of prior models, offering superior accuracy and efficiency for long-sequence predictions.

Non-linear dependence of the level of a series on previous data points is of interest, partly because of the possibility of producing a chaotic time series. However, more importantly, empirical investigations can indicate the advantage of using predictions derived from non-linear models, over those from linear models, as for example in nonlinear autoregressive exogenous models. Further references on nonlinear time series analysis: (Kantz and Schreiber), and (Abarbanel) [3]

Among other types of non-linear time series models, there are models to represent the changes of variance over time (heteroskedasticity). These models represent autoregressive conditional heteroskedasticity (ARCH) and the collection comprises a wide variety of representation (GARCH, TAR, EGARCH, FIGARCH, CGARCH, etc.). Here changes in variability are related to, or predicted by, recent past values of the observed series. This is in contrast to other possible representations of locally varying variability, where the variability might be modeled as being driven by a separate time-varying process, as in a doubly stochastic model [4].

In a recent work on "model-free" analyses (a term often used to refer to analyses that do not rely on modeling the processes evolution over time with a parametric mathematical expression), wavelet transform based methods (for example locally stationary wavelets and wavelet decomposed neural networks) have gained favor. Multiscale (often referred to as multiresolution) techniques decompose a given time series, attempting to illustrate time dependence at multiple scales. See also Markov switching multifractal (MSMF) techniques for modeling volatility evolution [5].

A hidden Markov model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states. An HMM can be considered as the simplest dynamic Bayesian network. HMM models are widely used in speech recognition, for translating a time series of spoken words into text [6].

A number of different notations are in use for time-series analysis. A common notation specifying a time series X that is indexed by the natural numbers is written [7]

Ergodicity implies stationarity, but the converse is not necessarily the case. Stationarity is usually classified into strict stationarity and wide-sense or second-order stationarity. Both models and applications can be developed under each of these conditions, although the models in the latter case might be considered as only partly specified [8].

In addition, time-series analysis can be applied where the series are seasonally stationary or non-stationary. Situations where the amplitudes of frequency components change with time can be dealt with in time-frequency analysis which makes use of a time-frequency representation of a time-series or signal [9].

Consideration of the autocorrelation function and the spectral density function (also cross-correlation functions and cross-spectral density functions) [10]

Time series can be visualized with two categories of chart: Overlapping Charts and Separated Charts. Overlapping Charts display all-time series on the same layout while Separated Charts presents them on different layouts (but aligned for comparison purpose) [11]

Jevons was the first to study times series. In his book he examined a number of weekly financial time series between the years 1825 to 1860 inclusive. These included times series of bankruptcies, currency circulation and discount rates [12].

De Gooijer, Jan G.; Hyndman, Rob J. (2006). "25 Years of Time Series Forecasting". *International Journal of Forecasting*. Twenty Five Years of Forecasting. 22 (3): 443-473. CiteSeerX 10.1.1.154.9227. doi:10.1016/j.ijforecast.2006.01.001. S2CID 14996235 [13].

Weigend A. S., Gershenfeld N. A. (Eds.) (1994), *Time Series Prediction: Forecasting the Future and Understanding the Past*. Proceedings of the NATO Advanced Research Workshop on Comparative Time Series Analysis (Santa Fe, May 1992), Addison-Wesley [14].

Wiener, Norbert. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications*. MIT Press. ISBN 978-0-262-25719-0 [15].

Auffarth, Ben (2021). *Machine Learning for Time-Series with Python: Forecast, predict, and detect anomalies with state-of-the-art machine learning methods* (1st ed.). Packt Publishing. ISBN 978-1-80181-962-6. Retrieved 5 November 2021 [16].

In mathematics, a time series is a sequence of data points indexed, listed, or graphed in chronological order. Most commonly, a time series consists of observations recorded at suc-

cessive equally spaced points in time. Thus, it represents a form of discrete-time data. A time series may describe measurements collected over seconds, days, years, or even centuries. Common examples include heights of ocean tides, counts of sunspots, daily temperature readings, and the closing values of stock market indices such as the Dow Jones Industrial Average [17].

A time series is often visualized using a run chart (a type of temporal line chart), which helps identify patterns such as trends, seasonal effects, and irregular fluctuations. Time series are widely used in statistics, actuarial science, signal processing, pattern recognition, econometrics, mathematical finance, weather forecasting, earthquake prediction, electroencephalography, control engineering, astronomy, communications engineering, and many other areas of applied science and engineering that involve temporal measurements [18].

Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values. Generally, time series data is modeled as a stochastic process. While regression analysis is often employed in such a way as to test relationships between one or more different time series, this type of analysis is not usually called "time series analysis", which refers in particular to relationships between different points in time within a single series [19].

Time series data have a natural temporal ordering. This makes time series analysis distinct from cross-sectional studies, in which there is no natural ordering of the observations (e.g. explaining people's wages by reference to their respective education levels, where the individuals' data could be entered in any order). Time series analysis is also distinct from spatial data analysis where the observations typically relate to geographical locations (e.g. accounting for house prices by the location as well as the intrinsic characteristics of the houses). A stochastic model for a time series will generally reflect the fact that observations close together in time will be more closely related than observations further apart. In addition, time series models will often make use of the natural one-way ordering of time so that values for a given period will be expressed as deriving in some way from past values, rather than from future values (see time reversibility) [20].

5. Acknowledgement

The authors thank the Department of Computer Science and Engineering, Rungta College of Engineering and Technology, for providing the necessary resources to conduct this comprehensive review and analysis.

5. References

- [1] Vaswani, A., et al. "Attention is all you need." *Advances in neural information processing systems*, 2017.
- [2] Box, G. E., & Jenkins, G. M. *Time series analysis: forecasting and control*. Holden-Day, 1970.
- [3] Hochreiter, S., & Schmidhuber, J. "Long short-term memory." *Neural computation*, 1997.
- [4] Zhou, H., et al. "Informer: Beyond efficient transformer for long sequence time-series forecasting." *AAAI*, 2021.
- [5] Wu, H., et al. "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting." *NeurIPS*, 2021.
- [6] Lim, B., et al. "Temporal Fusion Transformers for interpretable multi-horizon time series forecasting." *International Journal of Forecasting*, 2021.
- [7] Nie, Y., et al. "A time series is worth 64 words: Long-term forecasting with transformers." *ICLR*, 2023.
- [8] Wen, Q., et al. "Transformers in time series: A survey." *arXiv preprint arXiv:2202.07125*, 2022.
- [9] Salinas, D., et al. "DeepAR: Probabilistic forecasting with autoregressive recurrent networks." *International Journal of Forecasting*, 2020.
- [10] Wu, Z., et al. "Connecting the dots: Multivariate time series forecasting with graph neural networks." *KDD*, 2020.
- [11] Lai, G., et al. "Modeling long-and short-term temporal patterns with deep neural networks." *SIGIR*, 2018.
- [12] Oreshkin, B. N., et al. "N-BEATS: Neural basis expansion analysis for interpretable time series forecasting." *ICLR*, 2020.
- [13] Rangapuram, S. S., et al. "Deep state space models for time series forecasting." *NeurIPS*, 2018.
- [14] Kitaev, N., et al. "Reformer: The efficient transformer." *ICLR*, 2020.
- [15] Li, S., et al. "Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting." *NeurIPS*, 2019.
- [16] Taylor, S. J., & Letham, B. "Forecasting at scale." *The American Statistician*, 2018.
- [17] Hamilton, J. D. *Time series analysis*. Princeton university press, 2020.
- [18] Lütkepohl, H. *New introduction to multiple time series analysis*. Springer Science & Business Media, 2005.
- [19] Makridakis, S., et al. "The M4 Competition: Results, findings, conclusion and way forward." *International Journal of Forecasting*, 2020.

- [20] Sen, R., et al. "Think globally, act locally: A deep neural network approach to high-dimensional time series forecasting." *NeurIPS*, 2019.