Classical and Contemporary Approaches to Big Time Series Forecasting

Christos Faloutsos
CMU and Amazon Research
christos@cs.cmu.edu

Jan Gasthaus
AWS AI Labs
gasthaus@amazon.com

Tim Januschowski
AWS AI Labs
tjnsch@amazon.com

Yuyang Wang
AWS AI Labs
yuyawang@amazon.com

ABSTRACT

Time series forecasting is a key ingredient in the automation and optimization of business processes: in retail, deciding which products to order and where to store them depends on the forecasts of future demand in different regions; in cloud computing, the estimated future usage of services and infrastructure components guides capacity planning; and workforce scheduling in warehouses and factories requires forecasts of the future workload. Recent years have witnessed a paradigm shift in forecasting techniques and applications, from computer-assisted model- and assumption-based to data-driven and fully-automated. This shift can be attributed to the availability of large, rich, and diverse time series corpora and result in a set of challenges that need to be addressed such as the following. How can we build statistical models to efficiently and effectively learn to forecast from large and diverse data sources? How can we leverage the statistical power of "similar" time series to improve forecasts in the case of limited observations? What are the implications for building forecasting systems that can handle large data volumes?

The objective of this tutorial is to provide a concise and intuitive overview of the most important methods and tools available for solving large-scale forecasting problems. We review the state of the art in three related fields: (1) classical modeling of time series, (2) scalable tensor methods, and (3) deep learning for forecasting. Further, we share lessons learned from building scalable forecasting systems. While our focus is on providing an intuitive overview of the methods and practical issues which we will illustrate via case studies, we also present some technical details underlying these powerful tools.

SIGMOD '19, June 30-July 5, 2019, Amsterdam, Netherlands

ACM ISBN 978-1-4503-5643-5/19/06…$15.00
https://doi.org/10.1145/3299869.3314033

KEYWORDS

Forecasting, Tensor Analysis, Neural Network, Time Series

ACM Reference Format:
Christos Faloutsos, Jan Gasthaus, Tim Januschowski, and Yuyang Wang. 2019. Classical and Contemporary Approaches to Big Time Series Forecasting. In *2019 International Conference on Management of Data (SIGMOD '19), June 30-July 5, 2019, Amsterdam, Netherlands.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3299869.3314033

1 INTRODUCTION

Time series data occur naturally in countless domains including medical analysis [50], real estate [60], financial analysis [78], sensor network monitoring [56] and social activity mining [48, 49]. Of all the time series related data mining tasks, forecasting is one of the most sought-after applications of data-driven methods (and arguably the most difficult one) due to its importance in industrial, social and scientific applications. For example, forecasting plays a key role in automating and optimizing operational processes in most businesses and enables data driven decision making. Forecasts of product supply and demand can be used for optimal inventory management, staff scheduling and topology planning, and are more generally a crucial technology for most aspects of supply chain optimization. Outside of the retail use-case, the increasing volume of online, time-stamped activities represents a vital new opportunity for data scientists and analysts to measure the collective behavior of social, economic, and other important evolutions [39]. Facing rapidly growing data sets, the most fundamental requirements are the efficient and effective forecasting of "big" time series sequences.

Time series forecasting is a well known topic that has attracted interest in various research communities (e.g., statistics, machine learning, econometrics, operational research, databases, data mining, networking) for several decades. In the statistics and econometrics communities, the prevalent forecasting methods in use today have been developed in the setting of forecasting individual or small groups of time series with complex models designed and tuned by domain experts. Simultaneously, data mining and database researchers have been focusing on finding patterns in thousands or millions of related time series. Examples include forecasting the energy consumption of individual households, forecasting the load for servers in a data center, or forecasting the demand for all products that a large retailer offers. In these scenarios, a substantial amount of data on past behavior of similar, related time series can be leveraged for making a forecast for an individual time series. Using data from related time series not only allows fitting more complex (and hence potentially more accurate) models without overfitting, it can also alleviate the time and labor intensive human selection and preparation of co-variates and model selection steps required by classical techniques. Recent studies has revealed some new directions for research on large scale time series forecasting, including:

- **Scalable classical models:** Classical state-space models serve as a reliable workhorse for forecasting, with appealing properties such as interpretability, robustness, and theoretical guarantees. Modern extensions include scalable implementations, as well as the support for missing data and multiple data types.
- **Large-scale tensor analysis:** Time series data can be modeled as tensors, and tensor analysis is an important data mining tool that has various applications including sensor streams, hyperlinks, medical records and social networks over time.
- **Neural forecasting models:** With its dominance in machine learning applications such as image recognition and machine translation, deep learning has recently also received revived interests in the field of time series forecasting. Modern deep learning techniques not only improve the state-of-art forecasting performance but also, from a systems perspective, greatly reduce the complexity of the forecasting pipeline, and therefore increase maintainability.

This tutorial aims to bring together classical forecasting techniques, time series data mining techniques, and neural forecasting methods through a concise and intuitive overview of the most important tools and techniques that we can use to help us understand and forecast time series. We will provide a comprehensive overview of proven and current directions for time series forecasting, and deal specifically with the following key topics: (1) classical linear and non-linear modeling of time series, (2) scalable tensor methods, (3) deep learning for forecasting, and (4) lessons learned developing large scale forecasting systems. We shall supply IPython notebooks to illustrate commonly used forecasting techniques that are covered in this tutorial.

Who should attend. The target audience consists of database and data mining researchers who wish to familiarize themselves with the main techniques and recent developments in time series forecasting. Additionally, this tutorial is suitable for practitioners who want a concise, intuitive overview of the state of the art.

Prerequisites. The tutorial assumes familiarity with basic linear algebra, calculus, and discrete math, as well as with fundamentals of machine learning.

Related tutorials and how the current proposal differs. Related tutorials have been presented, e.g., (a) *Forecasting Big Time Series: Old and New*, by Christos Faloutsos, Jan Gasthaus, Tim Januschowski, and Yuyang Wang, VLDB 2018, (b) *Mining and Forecasting of Big Time Series Data*, by Yasuhi Sakurai, Yasuko Matsubara, and Christos Faloutsos, SIGMOD 2015, WWW 2016, KDD 2017, (c) *Indexing and Mining Streams*, by Christos Faloutsos, SIGMOD 2004, (d) *Indexing and Mining Time Sequences*, by Christos Faloutsos and Lei Li, SIGKDD 2010, and (e) *Mining Shape and Time Series Databases with Symbolic Representations*, by Eamonn Keogh, SIGKDD 2007.

Comparing to [18] (90 mins tutorial)[1], the proposed tutorial aims to expand the coverage of both the classical linear and nonlinear methods, as well as discussing more neural network methods and in greater depth. In addition, we plan to add large scale tensor analysis, the latest development of the neural forecasting, including recent work on combining classical and neural forecasting techniques and empirical comparisons of different neural structures. In general, our proposal differs from (b) – (e) such that: (1) we zoom in and purely focus on the *forecasting* part of time series analysis; (2) we bring together the *statistical* and *econometric* aspects with *data mining and management* of large scale time series forecasting, and (3), our major addition is the inclusion of recent *deep learning models for forecasting*, which have become increasingly popular in various domains and frequently shown superior performance compared to classical methods.

2 OUTLINE

The three-hour tutorial is tentatively structured as follows:

(1) **Introduction to Forecasting (15 mins)**

[1] https://lovvge.github.io/Forecasting-Tutorial-VLDB-2018/

- Basic (explanatory) analysis and decomposition of time series, i.e., trend, level, seasonality, etc.
- Point forecast vs. probabilistic forecast
- Forecast accuracy metrics

(2) **Classical Methods: Linear and Non-linear Models (45 mins)**
- Linear regression
  - Parameter estimation, least squares (LS), recursive LS
  - Time series transformations, i.e., power transform, Box-Cox, etc.
- Linear dynamical systems and exponential smoothing
  - Exponential smoothing (ES), Holt-Winters, and general Innovation state space models (ISSM)
  - ISSM with features, missing data, and different likelihood functions
- Non-linear dynamical systems
  - lag-plots, fractal dimension, power-law, and non-linear equations
  - non-linear dynamical system for online activities
  - information diffusion in social networks

(3) **Modern Methods: Tensor Analysis and Deep Learning (115 mins)**
- Scalable Tensor Analysis
  - Basics of matrix and tensor factorization
  - Decomposition of higher-order tensors
  - Big sparse tensors and forecasting of complex time-stamped events
- Deep learning for forecasting
  - Multi-layer perceptron (feedforward neural networks (NNs))
  - Recurrent neural networks (RNN)s: classic, Sequence-to-Sequence and other architectures
  - Others: Convolution, Wavenet, Generative Adversarial Networks (GANs), and all that

(4) **Conclusions and Lessons learned (15 mins)**
- Building large scale forecasting systems
- Developing Deep Autoregressive Network (DeepAR) and other models in AWS Sagemaker and Amazon Forecast

1. Introduction to Forecasting. In the opening chapter of the tutorial, we introduce the basic forecasting concepts and terminology. The classical time series analysis tools such as time series decomposition, lag plots, autocorrelations, etc. are also introduced [10, 11]. We discuss how to evaluate the accuracy of a forecast with metrics such as mean absolute percentage error (MAPE) and quantile losses [30]. In particular, we discuss the problem of assessing the quality of a *probabilistic* forecast and introduce the notion of a *proper scoring rule* and show why it is important for producing a calibrated forecast [22, 23].

2. Classical Methods: Linear and Non-linear models. We next provide a comprehensive review of the classical methods for forecasting, which mainly focus on individual time series (local).

2.1 Linear models. First, we cover the classic linear methods for forecasting, including the linear regression, autoregressive and moving average models (ARIMA), as well as useful tools in data management system such as MUSCLES [73] and AWSOM [55]. We also introduce linear dynamical systems (LDS), Kalman filter (KF) and their variants [31, 40, 41, 68]. In particular, we focus on the exponential smoothing models (Innovation state space models (ISSM)) [29] and structural time series models [16, 24, 62] along with their Bayesian counterparts [63, 64]. We show how to incorporate trend, seasonality factors, external signals such as promotional and other types of events and missing (or partially missing) observations [63]. We shall also cover how to model different types of time series observations, e.g. real, positive, integer, or binary data. We close this part by discussing the topic of scalable implementations of (Bayesian) state-space models [63, 64].

2.2. Non-linear dynamic models. Next, we introduce non-linear dynamic models. We start by explaining non-linear forecasting methods and introduce some fundamental concepts such as lag-plots [12], which is based on nearest-neighbor search, along with fractal dimension and power law [4, 52]. We then review the most common non-linear equations, including the logistic function (LF), the susceptible-infected (SI) model [2], the independent cascade (IC) model [17], the so-called "bass" model [5], the Lotka-Volterra (LV) model [51] and other non-linear equations [54]. We explain the importance of non-linear equations and the concept of gray-box non-linear mining. In this part, we also review recent work on understanding the non-linear time evolution of online user activities. Analyses of epidemics, blogs, social media, propagation and the cascades they create have attracted much interest.

3. Modern methods: Tensor Analysis and Deep Learning. In this part, we move to the territory of modern forecasting techniques, in particular with tensor methods and deep learning models. In contrast to the classical methods, the modern approaches learn across multiple related time series (global).

3.1 Scalable Tensor Analysis. We present large-scale studies of complex time-stamped events and big sparse tensors. We first introduce classic matrix factorizations (MF), and forecasting methods that are based on MF [74]. Next, the basic approaches in tensor analysis are reviewed, including Tucker, PARAFAC (CP), and higher-order SVD (HOSVD) [35, 36, 66]. Complex time-stamped events can be represented as a tensor with several dimensions. For example, given a set of time-stamped event entries of the form {object, actor, timestamp} (e.g., web-clicks: URL, userID, timestamp), we can treat them as a 3rd order tensor. Here, one subtle, but important issue is that the complex time-stamped tensor is very sparse, which derails all typical time series mining and forecasting tools. We introduce a scalable algorithm, TriMine [48] to deal with this issue. TriMine has the ability to find meaningful patterns in complex time-stamped tensors, and forecast future events. Furthermore, TensorCast [3, 13] is introduced to deal with additional contextual information and forecasting the disappearance of existing relations. Other forecasting algorithms that based on tensor factorization include [28, 46, 67]. We show new directions for both tensor analysis, including automatic non-linear analysis for big time series tensors [50] and other applications [27, 47, 65],

3.2 Deep learning for forecasting. In the 90s, Feedforward NNs were popular among forecasters [76] with applications in electrical load [43, 57], financial time series [21], and others [25]. Recent ground-breaking successes of deep neural network in other areas of machine learning have brought revived interests in applying deep learning techniques, especially recurrent neural networks and their variants [26], to time series forecasting [7, 33, 34, 45]. In this part, we first introduce the multi-layer perceptron (feedforward NN) as an extension of the linear regression models introduced in part 1, and review recent examples [1].

Then we give an overview of different types of RNNs, which capture the sequential nature of time series data. Different RNN forecasters are introduced [6, 19, 38, 42, 53, 70, 75] and we explain the intuitions behind different structures (canonical and seq2seq) and demonstrate their performances on a variety types of time series. We shall discuss convolutional NNs [8, 72], WaveNet [69] and illustrate how they can be used for forecasting. Finally, we discuss new directions for deep generative models for forecasting, in particular, with models that combines the strengths of both RNNs and classical probabilistic graphical models [20, 37, 44, 59]. We also discuss new areas of using deep learning for a variety of forecasting problems, such as spatio-temporal forecasting [14, 61, 77] and rare event forecasting, temporal point processes [15, 71].

4. Conclusions and Lessons learned. We conclude the tutorial with a summary of the previous parts and share the lessons learned developing the scalable forecasting system for retail within Amazon [9] and deep learning based forecasting algorithms [19] in AWS SageMaker [32] and Amazon Forecast [58].

3 PRESENTERS' SHORT BIO

Christos Faloutsos is a Professor at Carnegie Mellon University. He has received the Presidential Young Investigator Award by the National Science Foundation (1989), the Research Contributions Award in ICDM 2006, the SIGKDD Innovations Award (2010), twenty "best paper" awards (including two test of time awards), and four teaching awards. Five of his advisees have attracted KDD or SCS dissertation awards. He is an ACM Fellow, he has served as a member of the executive committee of SIGKDD; he has published over 300 refereed articles, 17 book chapters, and two monographs. He holds eight patents and has given over 40 tutorials and over 20 invited distinguished lectures. His research interests include data mining for graphs and streams, fractals, database performance, and indexing for multimedia and bioinformatics data.

Jan Gasthaus is a Senior Machine Learning Scientist in the Amazon AI Labs, working mainly on time series forecasting and large-scale probabilistic machine learning. He is passionate about developing novel machine learning solutions for addressing challenging business problems with scalable machine learning systems, all the way from scientific ideation to productization. Prior to joining Amazon, Jan obtained a BS in Cognitive Science from the University of Osnabrueck, an MS in Intelligent Systems from UCL, and pursued a PhD at the Gatsby Unit, UCL, focusing on Nonparametric Bayesian methods for sequence data.

Tim Januschowski is a Machine Learning Science Manager in Amazon AI Labs. He has worked on forecasting since starting his professional career. At Amazon, he has produced end-to-end solutions for a wide variety of forecasting problems, from demand forecasting to server capacity forecasting. Tim's personal interests in forecasting span applications, system, algorithm and modeling aspects and the downstream mathematical programming problems. He studied Mathematics at TU Berlin, IMPA, Rio de Janeiro, and Zuse-Institute Berlin and holds a PhD from University College Cork.

Yuyang Wang is a Senior Machine Learning Scientist in Amazon AI Labs, working mainly on large-scale probabilistic machine learning with its application in Forecasting. He received his PhD in Computer Science from Tufts University, MA, US and he holds an MS from the Department of Computer Science at Tsinghua University, Beijing, China. His research interests span statistical machine learning, numerical linear algebra, and random matrix theory. In forecasting, Yuyang has worked on all aspects ranging from practical applications to theoretical foundations.

REFERENCES

[1] John Alberg and Zachary C Lipton. Improving factor-based quantitative investing by forecasting company fundamentals. *arXiv preprint arXiv:1711.04837*, 2017.

[2] Roy M. Anderson and Robert M. May. *Infectious diseases of humans: Dynamics and control.* Oxford Press, 2002.

[3] Miguel Araújo, Pedro Ribeiro, and Christos Faloutsos. Tensorcast: forecasting time-evolving networks with contextual information. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pages 5199–5203. AAAI Press, 2018.

[4] Albert L. Barabasi. The origin of bursts and heavy tails in human dynamics. *Nature*, 435, 2005.

[5] Frank M. Bass. A new product growth for model consumer durables. *Management Science*, 15(5):215–227, 1969.

[6] Konstantinos Benidis, Jan Gasthaus, Yuyang Wang, Syama Sundar Rangapuram, David Salinas, Valentin Flunkert, and Tim Januschowski. Probabilistic forecasting with spline quantile function rnns. In *Artificial Intelligence and Statistics*, 2019.

[7] Filippo Maria Bianchi, Enrico Maiorino, Michael C Kampffmeyer, Antonello Rizzi, and Robert Jenssen. An overview and comparative analysis of recurrent neural networks for short term load forecasting. *arXiv preprint arXiv:1705.04378*, 2017.

[8] Mikołaj Bińkowski, Gautier Marti, and Philippe Donnat. Autoregressive convolutional neural networks for asynchronous time series. *arXiv preprint arXiv:1703.04122*, 2017.

[9] Joos-Hendrik Böse, Valentin Flunkert, Jan Gasthaus, Tim Januschowski, Dustin Lange, David Salinas, Sebastian Schelter, Matthias Seeger, and Yuyang Wang. Probabilistic demand forecasting at scale. *PVLDB*, 10(12):1694–1705, 2017.

[10] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time series analysis: forecasting and control.* John Wiley & Sons, 2015.

[11] P. J. Brockwell and R. A. Davis. Time series: Theory and method. Springer-Verlag, 1991.

[12] Deepay Chakrabarti and Christos Faloutsos. F4: Large-scale automated forecasting using fractals. *CIKM 2002*, November 2002.

[13] Miguel Ramos de Araujo, Pedro Manuel Pinto Ribeiro, and Christos Faloutsos. Tensorcast: Forecasting with context using coupled tensors. In *Data Mining (ICDM), 2017 IEEE International Conference on*, pages 71–80. IEEE, 2017.

[14] Dingxiong Deng, Cyrus Shahabi, Ugur Demiryurek, Linhong Zhu, Rose Yu, and Yan Liu. Latent space model for road networks to predict time-varying traffic. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1525–1534. ACM, 2016.

[15] Nan Du, Hanjun Dai, Rakshit Trivedi, Utkarsh Upadhyay, Manuel Gomez-Rodriguez, and Le Song. Recurrent marked temporal point processes: Embedding event history to vector. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1555–1564. ACM, 2016.

[16] James Durbin and Siem Jan Koopman. *Time series analysis by state space methods*, volume 38. OUP Oxford, 2012.

[17] David Easley and Jon Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World.* Cambridge University Press, September 2010.

[18] Christos Faloutsos, Jan Gasthaus, Tim Januschowski, and Yuyang Wang. Forecasting big time series: old and new. *Proceedings of the VLDB Endowment*, 11(12):2102–2105, 2018.

[19] Valentin Flunkert, David Salinas, Jan Gasthaus, and Tim Januschowski. Deepar: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting, arXiv:1704.04110*, To appear.

[20] Marco Fraccaro, Simon Kamronn, Ulrich Paquet, and Ole Winther. A disentangled recognition and nonlinear dynamics model for unsupervised learning. In *Advances in Neural Information Processing Systems*, pages 3604–3613, 2017.

[21] Edward Gately. *Neural networks for financial forecasting.* John Wiley & Sons, Inc., 1995.

[22] Tilmann Gneiting, Fadoua Balabdaoui, and Adrian E Raftery. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2):243–268, 2007.

[23] Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

[24] Andrew C Harvey. *Forecasting, structural time series models and the Kalman filter.* Cambridge university press, 1990.

[25] Tim Hill, Leorey Marquez, Marcus O'Connor, and William Remus. Artificial neural network models for forecasting and decision making. *International journal of forecasting*, 10(1):5–15, 1994.

[26] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[27] Bryan Hooi, Shenghua Liu, Asim Smailagic, and Christos Faloutsos. Beatlex: Summarizing and forecasting time series with patterns. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 3–19. Springer, 2017.

[28] Bryan Hooi, Hyun Ah Song, Amritanshu Pandey, Marko Jereminov, Larry Pileggi, and Christos Faloutsos. Streamcast: Fast and online mining of power grid time sequences. In *Proceedings of the 2018 SIAM International Conference on Data Mining*, pages 531–539. SIAM, 2018.

[29] Rob Hyndman, Anne B Koehler, J Keith Ord, and Ralph D Snyder. *Forecasting with exponential smoothing: the state space approach.* Springer Science & Business Media, 2008.

[30] Rob J Hyndman and George Athanasopoulos. Forecasting: Principles and practice. *www. otexts. org/fpp.*, 987507109, 2017.

[31] Ankur Jain, Edward Y. Chang, and Yuan-Fang Wang. Adaptive stream resource management using kalman filters. In *SIGMOD Conference*, pages 11–22, 2004.

[32] Tim Januschowski, David Arpin, David Salinas, Valentin Flunkert, Jan Gasthaus, Lorenzo Stella, and Paul Vazquez. Now available in amazon sagemaker: Deepar algorithm for more accurate time series forecasting. *https://aws.amazon.com/blogs/machine-learning/now-available-in-amazon-sagemaker-deepar-algorithm-for-more-accurate-time-series-forecasting/*, 2018.

[33] Tim Januschowski, Jan Gasthaus, Yuyang Wang, Syama Sundar Rangapuram, and Laurent Callot. Deep learning for forecasting: Current trends and challenges. *Foresight: The International Journal of Applied Forecasting*, (51), 2018.

[34] Tim Januschowski, Jan Gasthaus, Yuyang Wang, Syama Sundar Rangapuram, Laurent Callot, et al. Deep learning for forecasting. *Foresight: The International Journal of Applied Forecasting*, (50):35–41, 2018.

[35] Tamara G. Kolda and Brett W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, 2009.

[36] Tamara G. Kolda, Brett W. Bader, and Joseph P. Kenny. Higher-order web link analysis using multilinear algebra. In *ICDM 2005: Proceedings of the 5th IEEE International Conference on Data Mining*, pages 242–249, November 2005.

[37] Rahul G Krishnan, Uri Shalit, and David Sontag. Structured inference networks for nonlinear state space models. In *AAAI*, pages 2101–2109, 2017.

[38] Vitaly Kuznetsov and Zelda Mariet. Foundations of sequence-to-sequence modeling for time series. 2019.

[39] Jure Leskovec, Lars Backstrom, and Jon Kleinberg. Meme-tracking and the dynamics of the news cycle. *ACM SIGKDD*, 2009.

[40] L. Li, J. McCann, N.S. Pollard, and C. Faloutsos. Dynammo: mining and summarization of coevolving sequences with missing values. In *ACM SIGKDD*, pages 507–516, 2009.

[41] Lei (CMU) Li, B. Aditya (CMU) Prakash, and Christos (CMU) Faloutsos. Parsimonious Linear Fingerprinting for Time Series. In *VLDB*, Singapore, 2010.

[42] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. Graph convolutional recurrent neural network: Data-driven traffic forecasting. *ICLR*, 2018.

[43] C-N Lu, H-T Wu, and S Vemuri. Neural network based short term load forecasting. *IEEE Transactions on Power Systems*, 8(1):336–342, 1993.

[44] Danielle C. Maddix, Yuyang Wang, and Alex Smola. Deep factors with gaussian processes for forecasting. In *Workshop on Bayesian Deep Learning, NIPS*, 2018.

[45] Spyros Makridakis, Evangelos Spiliotis, and Vassilios Assimakopoulos. Statistical and machine learning forecasting methods: Concerns and ways forward. *PLOS ONE*, 13(3):1–26, 03 2018.

[46] Yasuko Matsubara, Yasushi Sakurai, and Christos Faloutsos. Non-linear mining of competing local activities. In *WWW*, pages 737–747. ACM, 2016.

[47] Yasuko Matsubara, Yasushi Sakurai, and Christos Faloutsos. Ecosystem on the web: non-linear mining and forecasting of co-evolving online activities. *World Wide Web*, 20(3):439–465, 2017.

[48] Yasuko Matsubara, Yasushi Sakurai, Christos Faloutsos, Tomoharu Iwata, and Masatoshi Yoshikawa. Fast mining and forecasting of complex time-stamped events. In *KDD*, pages 271–279. ACM, 2012.

[49] Yasuko Matsubara, Yasushi Sakurai, B. Aditya Prakash, Lei Li, and Christos Faloutsos. Rise and fall patterns of information diffusion: model and implications. In *KDD*, pages 6–14. ACM, 2012.

[50] Yasuko Matsubara, Yasushi Sakurai, Willem G van Panhuis, and Christos Faloutsos. Funnel: automatic mining of spatially coevolving epidemics. In *KDD*, pages 105–114. ACM, 2014.

[51] Robert May and Angela McLean. *Theoretical Ecology: Principles and Applications*. Oxford University Press, 3rd edition, 2007.

[52] M. McGlohon, J. Leskovec, C. Faloutsos, N. Glance, and M. Hurst. Finding patterns in blog shapes and blog evolution. In *International Conference on Weblogs and Social Media.*, Boulder, Colo., Helen Martin 2007.

[53] Srayanta Mukherjee, Devashish Shankar, Atin Ghosh, Nilam Tathawadekar, Pramod Kompalli, Sunita Sarawagi, and Krishnendu Chaudhury. Armdn: Associative and recurrent mixture density networks for eretail demand forecasting. *arXiv preprint arXiv:1803.03800*, 2018.

[54] Martin A Nowak. *Evolutionary dynamics*. Harvard University Press, 2006.

[55] Spiros Papadimitriou, Anthony Brockwell, and Christos Faloutsos. Adaptive, hands-off stream mining. In *VLDB*, pages 560–571. Morgan Kaufmann, 2003.

[56] Spiros Papadimitriou and Philip S. Yu. Optimal multi-scale patterns in time series streams. In *SIGMOD Conference*, pages 647–658, 2006.

[57] Dong C Park, MA El-Sharkawi, RJ Marks, LE Atlas, and MJ Damborg. Electric load forecasting using an artificial neural network. *IEEE transactions on Power Systems*, 6(2):442–449, 1991.

[58] Danilo Poccia. Amazon forecast — time series forecasting made easy. *https://aws.amazon.com/blogs/aws/amazon-forecast-time-series-forecasting-made-easy/*, 2018.

[59] Syama Sundar Rangapuram, Matthias Seeger, Jan Gasthaus, Lorenzo Stella, Yuyang Wang, and Tim Januschowski. Deep state space models for time series forecasting. In *Advances in Neural Information Processing Systems*, 2018.

[60] You Ren, Emily B Fox, and Andrew Bruce. Achieving a hyperlocal housing price index: Overcoming data sparsity by bayesian dynamical modeling of multiple data streams. *arXiv preprint arXiv:1505.01164*, 2015.

[61] Lucas Roberts, Leo Razoumov, Lin Su, and Yuyang Wang. Gini regularized optimal transport with an application to spatio-temporal forecasting. *NIPS Workshop on Optimal Transport*, 2017.

[62] Steven L Scott and Hal R Varian. Predicting the present with bayesian structural time series. *International Journal of Mathematical Modelling and Numerical Optimisation*, 5(1-2):4–23, 2014.

[63] Matthias Seeger, Syama Rangapuram, Yuyang Wang, David Salinas, Jan Gasthaus, Tim Januschowski, and Valentin Flunkert. Approximate bayesian inference in linear state space models for intermittent demand forecasting at scale. *arXiv preprint arXiv:1709.07638*, 2017.

[64] Matthias W Seeger, David Salinas, and Valentin Flunkert. Bayesian intermittent demand forecasting for large inventories. In *NIPS*, pages 4646–4654, 2016.

[65] Hyun Ah Song, Bryan Hooi, Marko Jereminov, Amritanshu Pandey, Larry Pileggi, and Christos Faloutsos. Powercast: Mining and forecasting power grid sequences. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 606–621. Springer, 2017.

[66] Jimeng Sun, Dacheng Tao, and Christos Faloutsos. Beyond streams and graphs: dynamic tensor analysis. In *KDD*, pages 374–383. ACM, 2006.

[67] Koh Takeuchi, Hisashi Kashima, and Naonori Ueda. Autoregressive tensor factorization for spatio-temporal predictions. In *ICDM*, pages 1105–1110. IEEE, 2017.

[68] Yufei Tao, Christos Faloutsos, Dimitris Papadias, and Bin Liu. Prediction and indexing of moving objects with unknown motion patterns. In *SIGMOD Conference*, pages 611–622. ACM, 2004.

[69] Aäron Van Den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew W Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio. In *SSW*, page 125, 2016.

[70] Ruofeng Wen, Kari Torkkola, and Balakrishnan Narayanaswamy. A multi-horizon quantile recurrent forecaster. *NIPS Workshop on Time Series, arXiv:1711.11053*, 2017.

[71] Shuai Xiao, Mehrdad Farajtabar, Xiaojing Ye, Junchi Yan, Le Song, and Hongyuan Zha. Wasserstein learning of deep generative point process models. In *NIPS*, pages 3247–3257, 2017.

[72] SHI Xingjian, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *NIPS*, pages 802–810, 2015.

[73] B.K. Yi, N.D. Sidiropoulos, T. Johnson, HV Jagadish, C. Faloutsos, and A. Biliris. Online data mining for co-evolving time sequences. In *Data Engineering, 2000. Proceedings. 16th International Conference on*, pages 13–22. IEEE, 2000.

[74] Hsiang-Fu Yu, Nikhil Rao, and Inderjit S Dhillon. Temporal regularized matrix factorization for high-dimensional time series prediction. In *NIPS*, pages 847–855, 2016.

[75] Rose Yu, Stephan Zheng, Anima Anandkumar, and Yisong Yue. Long-term forecasting using tensor-train rnns. *arXiv preprint arXiv:1711.00073*, 2017.

[76] Guoqiang Zhang, B Eddy Patuwo, and Michael Y Hu. Forecasting with artificial neural networks:: The state of the art. *International journal of forecasting*, 14(1):35–62, 1998.

[77] Junbo Zhang, Yu Zheng, and Dekang Qi. Deep spatio-temporal residual networks for citywide crowd flows prediction. In *AAAI*, pages 1655–1661, 2017.

[78] Yunyue Zhu and Dennis Shasha. Statstream: Statistical monitoring of thousands of data streams in real time. *PVLDB*, pages 358–369, 2002.