

(Don't you) Forget about me : On the importance of time in data



Too much data to handle?
Let's see what we can do!

Rémy Raes

01 Context



The dream team



Me, myself and I

- ▶ Rémy Raes
- ▶ Previously research engineer
- ▶ 2nd year Ph.D student

Distributed Machine Learning in
Ubiquitous Environments using
Location-dependent Models

- ▶ Systems for ML
- ▶ Shared between Spirals (Lille) and
WIDE (Rennes)



The dream team



Me, myself and I

- ▶ Rémy Raes
- ▶ Previously research engineer
- ▶ 2nd year Ph.D student

Distributed Machine Learning in Ubiquitous Environments using Location-dependent Models

- ▶ Systems for ML
- ▶ Shared between Spirals (Lille) and WIDE (Rennes)



The dream team



Me, myself and I

- ▶ Rémy Raes
- ▶ Previously research engineer
- ▶ 2nd year Ph.D student

Distributed Machine Learning in Ubiquitous Environments using Location-dependent Models

- ▶ Systems for ML
- ▶ Shared between Spirals (Lille) and WIDE (Rennes)

The dream team



Me, myself and I

- ▶ Rémy Raes
- ▶ Previously research engineer
- ▶ 2nd year Ph.D student

Distributed Machine Learning in Ubiquitous Environments using Location-dependent Models

- ▶ Systems for ML
- ▶ Shared between Spirals (Lille) and WIDE (Rennes)

Time series

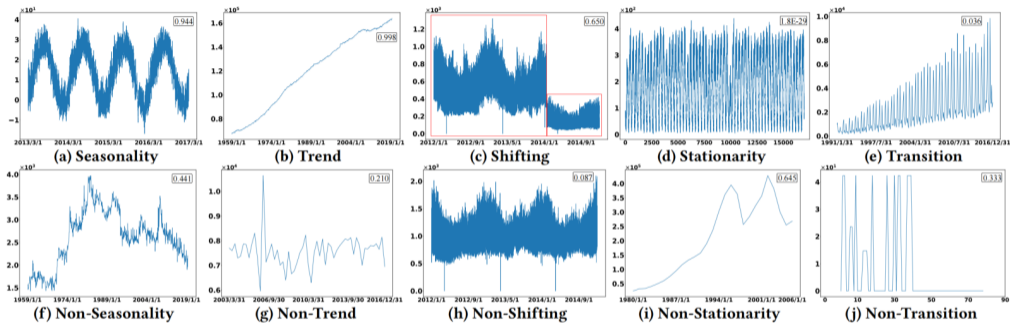


Figure 1: Visualization of data with different characteristics.

Qiu et al. TFB: Towards Comprehensive and Fair Benchmarking of Time Series Forecasting Methods. *Proceedings of the VLDB Endowment*, Vol. 17, No. 9 ISSN 2150-809 (10.14778/3665844.3665863)

Questions

- ▶ Devices are producers of data
- ▶ Training data better stay local

What are the limits preventing *in situ* computing?

1. Data processing
2. Data storage

Questions

- ▶ Devices are producers of data
- ▶ Training data better stay local

What are the limits preventing *in situ* computing?

1. Data processing
2. Data storage

Questions

- ▶ Devices are producers of data
- ▶ Training data better stay local

What are the limits preventing *in situ* computing?

1. Data processing
2. Data storage

Questions

- ▶ Devices are producers of data
- ▶ Training data better stay local

What are the limits preventing *in situ* computing?

1. Data processing
2. Data storage

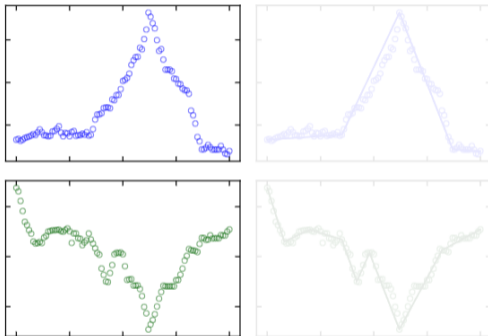
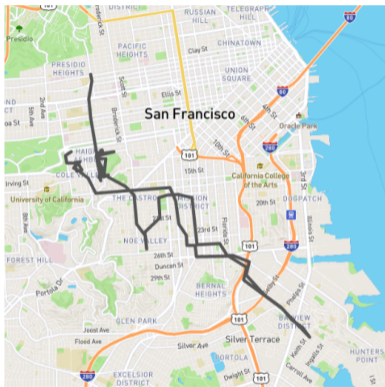
Questions

- ▶ Devices are producers of data
- ▶ Training data better stay local

What are the limits preventing *in situ* computing?

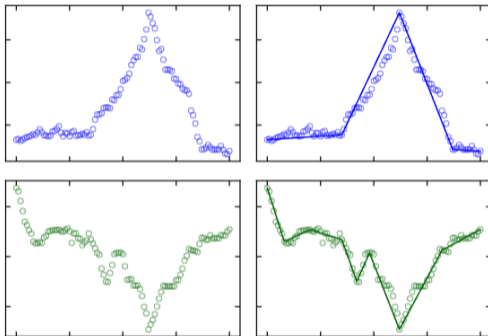
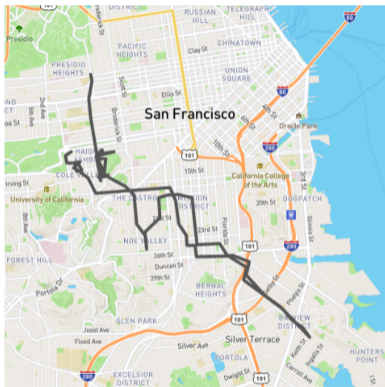
1. Data processing
2. Data storage

Fast linear interpolation



Rémy Raes, Olivier Ruas, Adrien Luxey-Bitri, Romain Rouvoy. Compact Storage of Data Streams in Mobile Devices. DAIS'24 - 24th International Conference on Distributed Applications and Interoperable Systems, Jun 2024, Groningen, Netherlands. (hal-04535716v3)

Fast linear interpolation



Rémy Raes, Olivier Ruas, Adrien Luxey-Bitri, Romain Rouvoy. Compact Storage of Data Streams in Mobile Devices. *DAIS'24 - 24th International Conference on Distributed Applications and Interoperable Systems*, Jun 2024, Groningen, Netherlands. (hal-04535716v3)

02 Works





Additional questions

Can we go further on with time series compression?

- ▶ Can we do better than straight data removal?
- ▶ "Right to be forgotten" hint from law community



Additional questions

Can we go further on with time series compression?

- ▶ Can we do better than straight data removal?
- ▶ "Right to be forgotten" hint from law community



Additional questions

Can we go further on with time series compression?

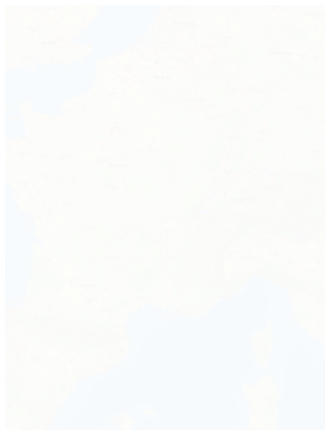
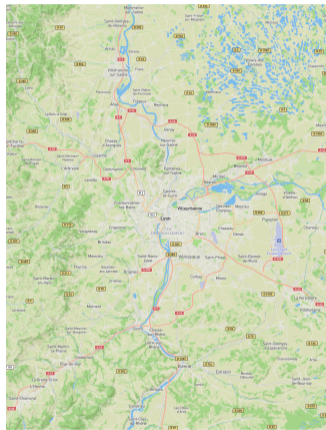
- ▶ Can we do better than straight data removal?
- ▶ "Right to be forgotten" hint from law community



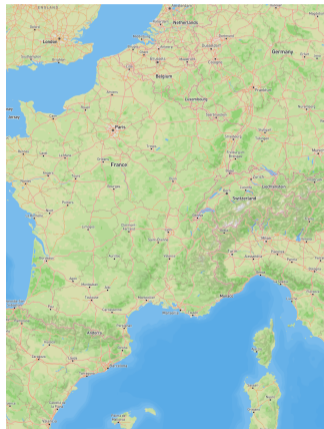
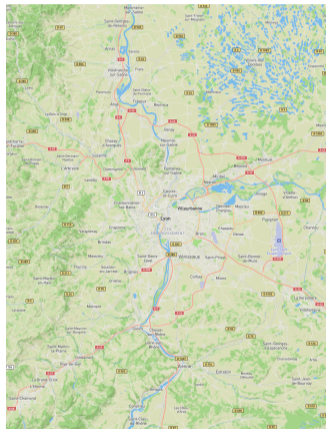
"Right to be forgotten" hint



"Right to be forgotten" hint

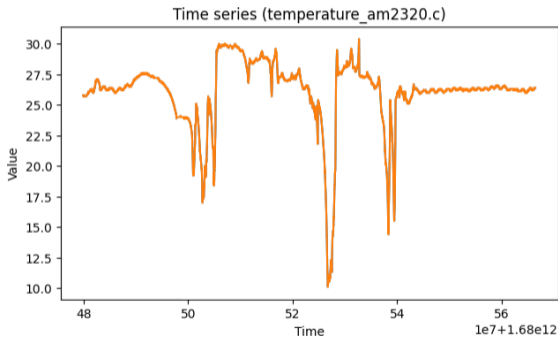


"Right to be forgotten" hint





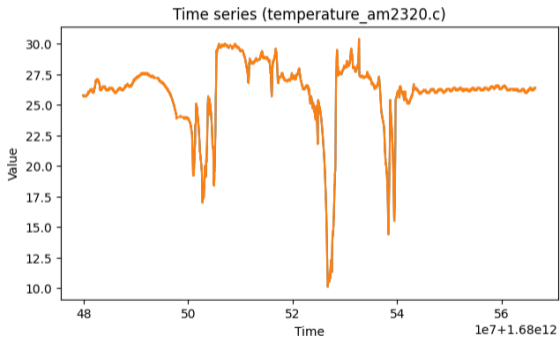
Translation into real world



- ▶ Time series to be compressed
- ▶ Prioritize old data for compression
- ▶ Need for a time-dependent compression method



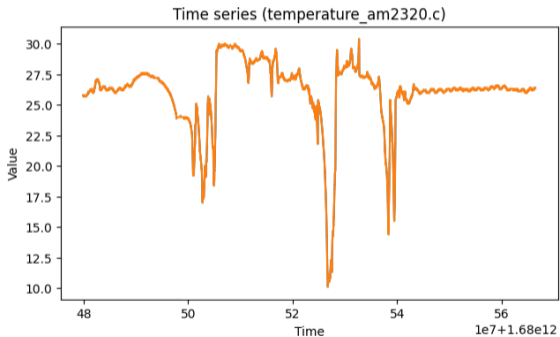
Translation into real world



- ▶ Time series to be compressed
- ▶ Prioritize old data for compression
- ▶ Need for a time-dependent compression method



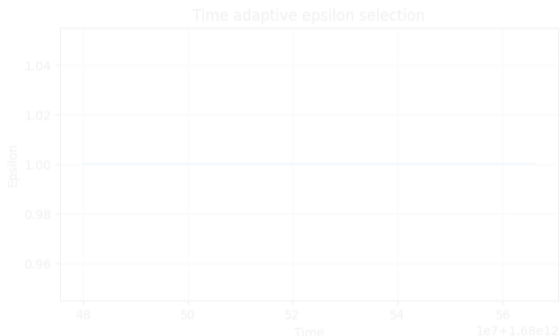
Translation into real world



- ▶ Time series to be compressed
- ▶ Prioritize old data for compression
- ▶ Need for a time-dependent compression method



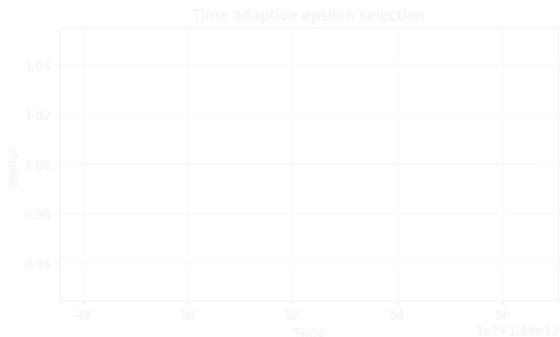
Error selection



- ▶ Use a function to pick tolerated error
- ▶ Function is time-indexed
- ▶ Different behaviours:
 - Constant value (FLI)
 - Decreasing value:
 - Linear
 - by step
 - with power function



Error selection



► Use a function to pick tolerated error

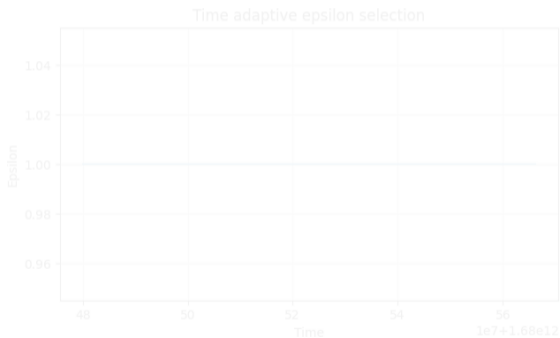
► Function is time-indexed

► Different behaviours:

- Constant value (FLI)
- Decreasing value:
 - Linear
 - by step
 - with power function



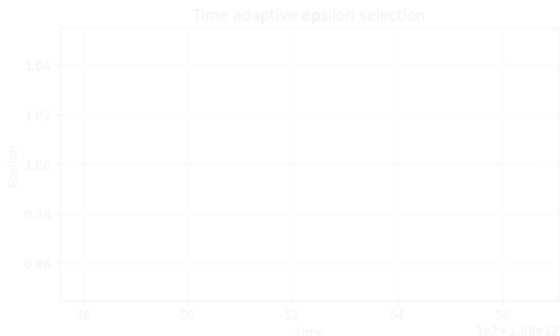
Error selection



- ▶ Use a function to pick tolerated error
- ▶ Function is time-indexed
- ▶ Different behaviours:
 - Constant value (FLI)
 - Decreasing value:
 - Linear
 - by step
 - with power function



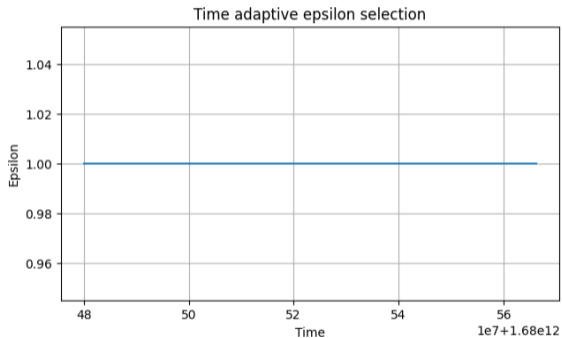
Error selection



- ▶ Use a function to pick tolerated error
- ▶ Function is time-indexed
- ▶ Different behaviours:
 - Constant value (FLI)
 - Decreasing value:
 - Linear
 - by step
 - with power function



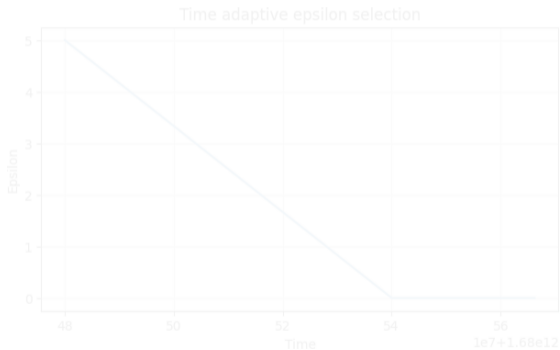
Error selection



- ▶ Use a function to pick tolerated error
- ▶ Function is time-indexed
- ▶ Different behaviours:
 - Constant value (FLI)
 - Decreasing value:
 - Linear
 - by step
 - with power function



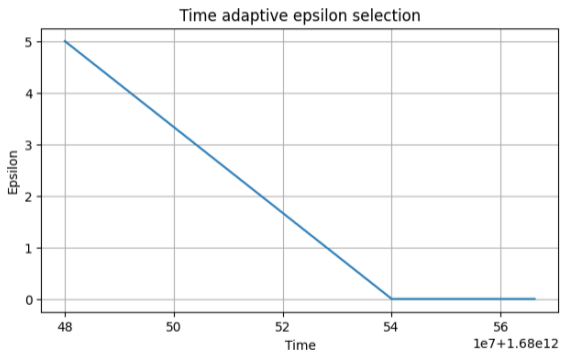
Error selection



- ▶ Use a function to pick tolerated error
- ▶ Function is time-indexed
- ▶ Different behaviours:
 - Constant value (FLI)
 - Decreasing value:
 - Linear
 - by step
 - with power function



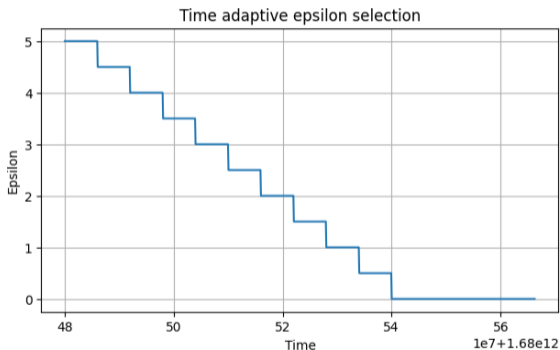
Error selection



- ▶ Use a function to pick tolerated error
- ▶ Function is time-indexed
- ▶ Different behaviours:
 - Constant value (FLI)
 - Decreasing value:
 - Linear
 - by step
 - with power function



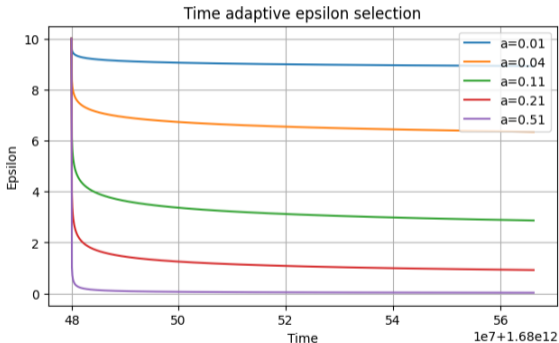
Error selection



- ▶ Use a function to pick tolerated error
- ▶ Function is time-indexed
- ▶ Different behaviours:
 - Constant value (FLI)
 - Decreasing value:
 - Linear
 - by step
 - with power function



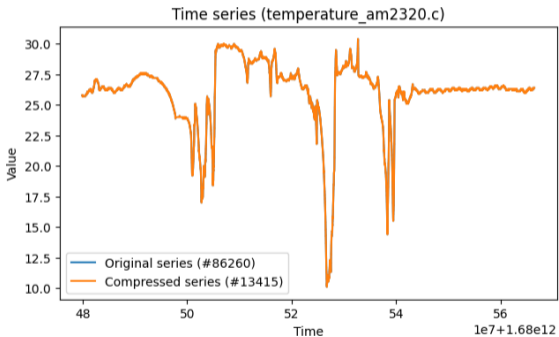
Error selection



- ▶ Use a function to pick tolerated error
- ▶ Function is time-indexed
- ▶ Different behaviours:
 - Constant value (FLI)
 - Decreasing value:
 - Linear
 - by step
 - with power function



Outlayers relative conservation

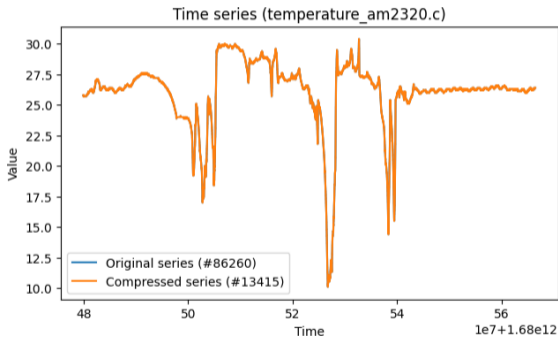


Knobs to control them all

- ▶ Double target
 - Size
 - Data quality
- ▶ Control theory?



Outlayers relative conservation

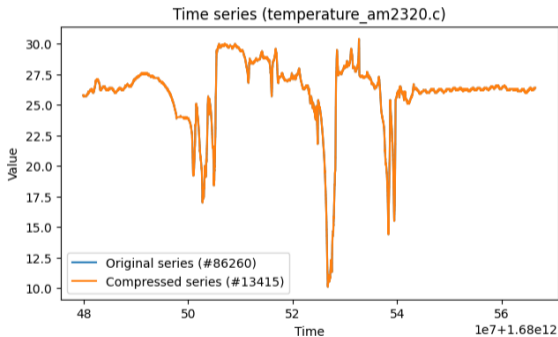


Knobs to control them all

- ▶ Double target
 - Size
 - Data quality
- ▶ Control theory?



Outlayers relative conservation

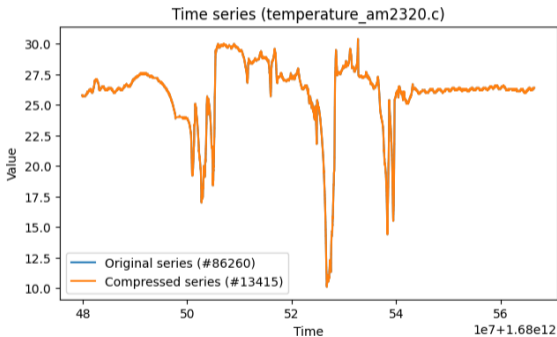


Knobs to control them all

- ▶ Double target
 - Size
 - Data quality
- ▶ Control theory?



Outlayers relative conservation

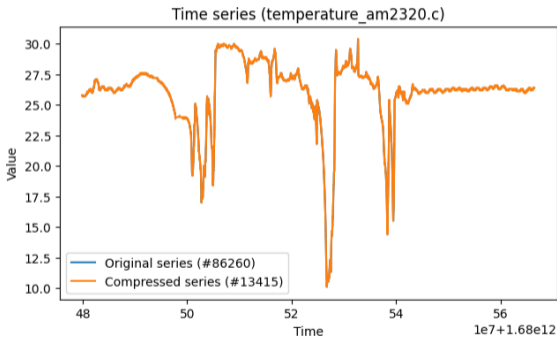


Knobs to control them all

- ▶ Double target
 - Size
 - Data quality
- ▶ Control theory?



Outlayers relative conservation

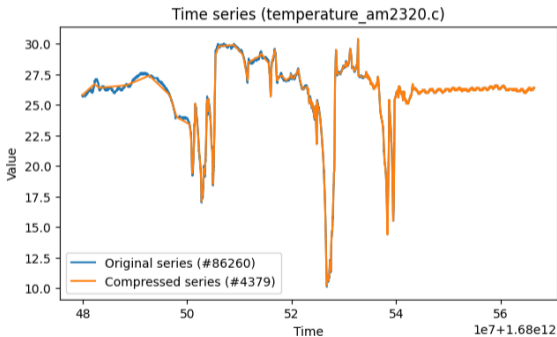


Knobs to control them all

- ▶ Double target
 - Size
 - Data quality
- ▶ Control theory?



Outlayers relative conservation

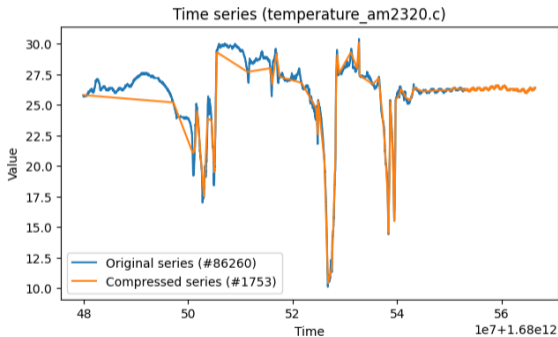


Knobs to control them all

- ▶ Double target
 - Size
 - Data quality
- ▶ Control theory?



Outliers relative conservation

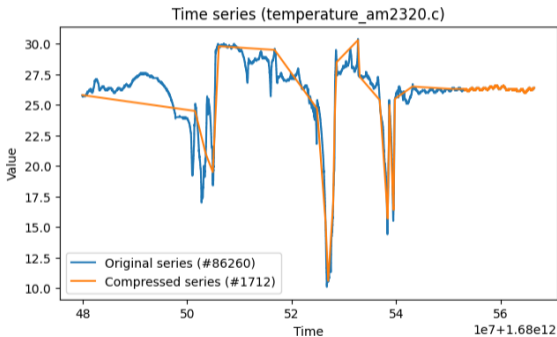


Knobs to control them all

- ▶ Double target
 - Size
 - Data quality
- ▶ Control theory?



Outlayers relative conservation




Knobs to control them all

- ▶ Double target
 - Size
 - Data quality
- ▶ Control theory?

03

Future works



Data utility with location data

Size results

▶ $\epsilon = 10^{-3}$

▶ From 7.2 GB to 25 MB

Data utility

▶ Latitude

- Tolerated error: $10^{-3} \text{ deg} \approx 111 \text{ m}$
- Median error: 5.33×10^{-5}
- RMSE: 3.72×10^{-4}

▶ Longitude

- Tolerated error: $10^{-3} \text{ deg} \approx 88 \text{ m}$
- Median error: 2.81×10^{-5}
- RMSE: 3.44×10^{-4}

▶ Privacy utility

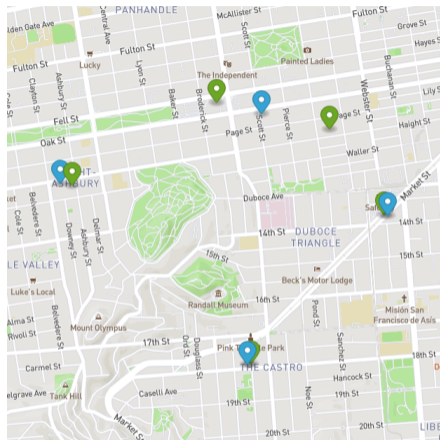


Figure – Points of Interest computed using raw data and FLI-modeled data.

Data utility with location data

Size results

- ▶ $\epsilon = 10^{-3}$
- ▶ From 7.2 GB to **25 MB**

Data utility

- ▶ Latitude
 - Tolerated error: $10^{-3} \text{ deg} \approx 111 \text{ m}$
 - Median error: 5.33×10^{-5}
 - RMSE: 3.72×10^{-4}
- ▶ Longitude
 - Tolerated error: $10^{-3} \text{ deg} \approx 88 \text{ m}$
 - Median error: 2.81×10^{-5}
 - RMSE: 3.44×10^{-4}
- ▶ Privacy utility

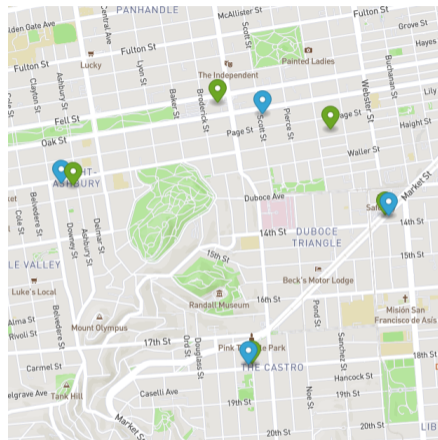


Figure – Points of Interest computed using raw data and FLI-modeled data.

Data utility with location data

Size results

- ▶ $\epsilon = 10^{-3}$
- ▶ From 7.2 GB to **25 MB**

Data utility

- ▶ Latitude
 - Tolerated error: $10^{-3} \text{ deg} \approx 111 \text{ m}$
 - Median error: 5.33×10^{-5}
 - RMSE: 3.72×10^{-4}
- ▶ Longitude
 - Tolerated error: $10^{-3} \text{ deg} \approx 88 \text{ m}$
 - Median error: 2.81×10^{-5}
 - RMSE: 3.44×10^{-4}

▶ Privacy utility

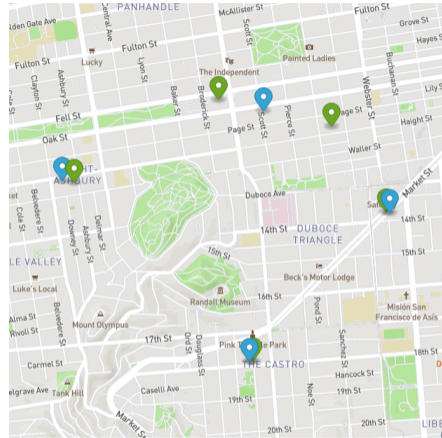


Figure – Points of Interest computed using raw data and FLI-modeled data.

Data utility with location data

Size results

- ▶ $\epsilon = 10^{-3}$
- ▶ From 7.2 GB to **25 MB**

Data utility

- ▶ Latitude
 - Tolerated error: $10^{-3} \text{ deg} \approx 111 \text{ m}$
 - Median error: 5.33×10^{-5}
 - RMSE: 3.72×10^{-4}
- ▶ Longitude
 - Tolerated error: $10^{-3} \text{ deg} \approx 88 \text{ m}$
 - Median error: 2.81×10^{-5}
 - RMSE: 3.44×10^{-4}
- ▶ Privacy utility

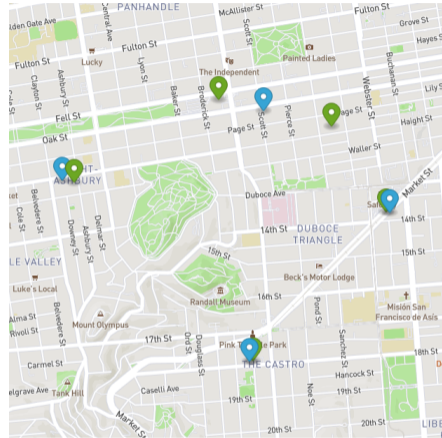


Figure – Points of Interest computed using **raw data** and **FLI-modeled data**.



(Future) Evaluation

- ▶ Time series tasks: forecasting, anomaly/pattern detection
- ▶ Compressing training dataset should have an impact on the model accuracy
- ▶ Study prediction error with transformed training dataset



(Future) Evaluation

- ▶ Time series tasks: forecasting, anomaly/pattern detection
- ▶ Compressing training dataset should have an impact on the model accuracy
- ▶ Study prediction error with transformed training dataset



(Future) Evaluation

- ▶ Time series tasks: forecasting, anomaly/pattern detection
- ▶ Compressing training dataset should have an impact on the model accuracy
- ▶ Study prediction error with transformed training dataset

TFB: Towards Comprehensive and Fair Benchmarking of Time Series Forecasting Methods

Xiangfei Qiu East China Normal University, China	Jilin Hu East China Normal University, China ^{2S}	Lelei Zhou Huawei Cloud Algorithm Innovation Lab, China	Xingjian Wu East China Normal University, China
Junyang Du East China Normal University, China	Huang Zhang East China Normal University, China	Chenjuan Gao East China Normal University, China	Aoying Zhou East China Normal University, China
Christian S. Jensen Aalborg University, Denmark	Zhenli Sheng Huawei Cloud Algorithm Innovation Lab, China	Biao Yang East China Normal University, China	

ABSTRACT

Time series are generated by diverse domains such as economic, traffic, health, and energy, where forecasting of future values has numerous important applications. Not surprisingly, many forecasting methods are being proposed. To ensure progress, it is essential to be able to study and compare such methods especially to cover preferable and suitable manner. To achieve this, we propose TFB, an automated benchmark for Time Series Forecasting (TSF) methods. TFB advances the state-of-the-art by addressing shortcomings related to datasets, comparison methods, and evaluation pipelines: 1) insufficient coverage of data domains, 2) coverage bias against traditional methods, and 3) inconsistent and unfair pipelines. To achieve better domain coverage, we include datasets from 19 different domains: traffic, electricity, energy, the environment, nature, economic, stock markets, banking, health, and the web. We also provide a time-series characteristic to ensure that the selected datasets are comprehensive. To remove biases against some methods, we include a diverse range of methods, including statistical learning, machine learning, and deep learning methods, and we also support a variety of evaluation strategies and metrics to ensure a more-comprehensive evaluation of different methods. To support the integration of different methods into the benchmark and enable fair comparisons, TFB features a flexible and scalable pipeline that eliminates biases. Next, we employ TFB to perform a thorough evaluation of 21 Univariate Time Series Forecasting (UTSF) methods on 1681 univariate time series and 14 Multivariate Time Series Forecasting (MTSF) methods on 28 datasets. The results offer a deeper understanding of the forecasting methods, allowing us to better select the ones that are most suitable for particular datasets and settings. Overall, TFB and this evaluation provide researchers with improved means of designing new TSF methods.

This work is licensed under the Creative Commons Attribution 4.0 International License. You can find a copy of the license at <https://creativecommons.org/licenses/by/4.0/>. For any reuse or distribution, you must make clear to others that you copied the work from the original author. For any reuse or distribution, you must make clear to others that you copied the work from the original author. For any reuse or distribution, you must make clear to others that you copied the work from the original author. For any reuse or distribution, you must make clear to others that you copied the work from the original author.

TFB Reference Format:
Shangqi Qiu, Jilin Hu, Lelei Zhou, Xingjian Wu, Junyang Du, Huang Zhang, Chenjuan Gao, Aoying Zhou, Christian S. Jensen, Zhenli Sheng and Biao Yang. TFB: Towards Comprehensive and Fair Benchmarking of Time Series Forecasting Methods. *TFB*, 2024. 2021, 2024.
doi:10.47792/2690441.2690441

TFB Artifact Availability:
The source code, data, and/or other artifacts have been made available at <https://github.com/donotoverfit/tfb>.

1 INTRODUCTION

As part of the ongoing digitalization, time series are generated in a variety of domains, such as economic [36, 75], traffic [34, 35, 39, 85, 92, 93, 95, 96, 99, 95, 96], health [84, 85, 90, 96], energy [1, 29], and AIops [1, 8, 41, 72, 87, 301]. Time Series Forecasting (TSF) is essential for applications in these domains [28, 62, 78, 97]. Given historical observations, it is valuable if we can know the future values ahead of time. Correspondingly, TSF has been firmly established as an active research field, witnessing the proposal of numerous methods.

Time series originate data points chronologically and are either univariate or multivariate depending on the number of variables in each data point. Accordingly, TSF methods can be classified as either Univariate Time Series Forecasting (UTSF) or Multivariate Time Series Forecasting (MTSF) methods. Among early methods, Autoregressive Integrated Moving Average (ARIMA) [4] and Vector Autoregression (VAR) [82] are arguably the most popular univariate and multivariate forecasting methods, respectively. Subsequent methods that exploit machine learning, e.g., XGBoost [11, 95] and Random Forest [5, 98] offer better performance than the early methods. Most recently, methods based on deep learning have demonstrated state-of-the-art (SOTA) forecasting performance on a variety of datasets [10, 12, 18, 30, 39, 68, 84, 70, 89, 91, 92, 90, 191, 302, 186].

As more and more methods are being proposed for different datasets and settings, there is an increasing need for fair and open preferable empirical evaluations. To achieve this, we identify and address these issues in existing evaluation frameworks, thereby advancing our evaluation capabilities.

TFB (PVLDB 2024)

- ▶ 25 multivariate datasets
- ▶ 8068 univariate datasets
- ▶ 22 models
- ▶ (open-source!)

TFB: Towards Comprehensive and Fair Benchmarking of Time Series Forecasting Methods

Xiangfei Qiu
East China Normal
University, China

Jilin Hu
East China Normal
University, China^{2S}

Lelei Zhou
Huawei Cloud Algorithm
Innovation Lab, China

Xingjian Wu
East China Normal
University, China

Junyang Du
East China Normal
University, China

Huang Zhang
East China Normal
University, China

Chenjian Gao
East China Normal
University, China

Aoying Zhou
East China Normal
University, China

Christian S. Jensen
Aalborg University,
Denmark

Zhenli Sheng
Huawei Cloud Algorithm
Innovation Lab, China

Biao Yang
East China Normal
University, China

ABSTRACT

Time series are generated by diverse domains such as economic, traffic, health, and energy, where forecasting of future values has numerous important applications. Not surprisingly, many forecasting methods are being proposed. To ensure progress, it is essential to be able to study and compare such methods especially to cover preferable and suitable manner. To achieve this, we propose TFB, an automated benchmark for Time Series Forecasting (TSF) methods. TFB advances the state-of-the-art by addressing shortcomings related to datasets, comparison methods, and evaluation pipelines: 1) insufficient coverage of data domains, 2) coverage bias against traditional methods, and 3) inconsistent and unfair pipelines. To achieve better domain coverage, we include datasets from 19 different domains: traffic, electricity, energy, the environment, nature, economic, stock markets, banking, health, and the web. We also provide a time-series characteristic to ensure that the selected datasets are comprehensive. To remove biases against some methods, we include a diverse range of methods, including statistical learning, machine learning, and deep learning methods, and we also support a variety of evaluation strategies and metrics to ensure a more-comprehensive evaluation of different methods. To support the integration of different methods into the benchmark and enable fair comparisons, TFB features a flexible and scalable pipeline that eliminates biases. Next, we employ TFB to perform a thorough evaluation of 21 Univariate Time Series Forecasting (UTSF) methods on 161 univariate time series and 14 Multivariate Time Series Forecasting (MTSF) methods on 28 datasets. The results offer a deeper understanding of the forecasting methods, allowing us to better select the ones that are most suitable for particular datasets and settings. Overall, TFB and this evaluation provide researchers with improved means of designing new TSF methods.

This work is licensed under the Creative Commons Attribution 4.0 International License. You may obtain a copy of the license at <http://creativecommons.org/licenses/by/4.0/>. For any reuse or distribution, you must give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

PVLDB Reference Format:

Shangyi Qiu, Jilin Hu, Lelei Zhou, Xingjian Wu, Junyang Du, Huang Zhang, Chenjian Gao, Aoying Zhou, Christian S. Jensen, Zhenli Sheng and Biao Yang. TFB: Towards Comprehensive and Fair Benchmarking of Time Series Forecasting Methods. PVLDB, 17(6):2104–2121, 2024. doi:10.1177/21584604241268303

PVLDB Artifact Availability:

The source code, data, and/or other artifacts have been made available at <https://github.com/donotmeforgetme/TFB>.

1 INTRODUCTION

As part of the ongoing digitalization, time series are generated in a variety of domains, such as economic [36, 75], traffic [34, 35–38, 49, 51, 52, 65, 79, 89, 95, 96], health [34, 61, 66], energy [1, 26], and AIops [1, 8, 41, 72, 87, 90]. Time Series Forecasting (TSF) is essential for applications in these domains [20, 62, 78, 92]. Given historical observations, it is valuable if we can know the future values ahead of time. Correspondingly, TSF has been firmly established as an active research field, witnessing the proposal of numerous methods.

Time series originate data points chronologically and are either univariate or multivariate depending on the number of variables in each data point. Accordingly, TSF methods can be classified as either Univariate Time Series Forecasting (UTSF) or Multivariate Time Series Forecasting (MTSF) methods. Among early methods, Autoregressive Integrated Moving Average (ARIMA) [4] and Vector Autoregression (VAR) [82] are arguably the most popular univariate and multivariate forecasting methods, respectively. Subsequent methods that exploit machine learning, e.g., XGBoost [11, 95] and Random Forest [5, 98] offer better performance than the early methods. More recently, methods based on deep learning have demonstrated state-of-the-art (SOTA) forecasting performance on a variety of datasets [10, 12, 18–30, 38, 64, 70, 89, 91, 93, 96, 101, 102, 104].

As more and more methods are being proposed for different datasets and settings, there is an increasing need for fair and open preferable empirical evaluations. To achieve this, we identify and address these needs in existing evaluation frameworks, thereby advancing our evaluation capabilities.

TFB (PVLDB 2024)

- ▶ 25 multivariate datasets
- ▶ 8068 univariate datasets
- ▶ 22 models
- ▶ (open-source!)

TFB: Towards Comprehensive and Fair Benchmarking of Time Series Forecasting Methods

Xiangfei Qiu East China Normal University, China	Jilin Hu East China Normal University, China ²⁵	Lelei Zhou Huawei Cloud Algorithm Innovation Lab, China	Xingjian Wu East China Normal University, China
Jiuyang Du East China Normal University, China	Huang Zhang East China Normal University, China	Chenjuan Gao East China Normal University, China	Aoying Zhou East China Normal University, China
Christian S. Jensen Aalborg University, Denmark	Zhenli Sheng Huawei Cloud Algorithm Innovation Lab, China	Bao Yang East China Normal University, China	

ABSTRACT

Time series are generated by diverse domains such as economic, traffic, health, and energy, where forecasting of future values has numerous important applications. Not surprisingly, many forecasting methods are being proposed. To ensure progress, it is essential to be able to study and compare such methods especially to cover preferable and suitable manner. To achieve this, we propose TFB, an automated benchmark for Time Series Forecasting (TSF) methods. TFB advances the state-of-the-art by addressing shortcomings related to datasets, comparison methods, and evaluation pipelines: 1) insufficient coverage of data domains, 2) coverage bias against traditional methods, and 3) inconsistent and inflexible pipelines. To achieve better domain coverage, we include datasets from 19 different domains: traffic, electricity, energy, the environment, nature, economic, stock markets, banking, health, and the web. We also provide a time-series characteristic to ensure that the selected datasets are comprehensive. To remove biases against some methods, we include a diverse range of methods, including statistical learning, machine learning, and deep learning methods, and we also support a variety of evaluation strategies and metrics to ensure a more-comprehensive evaluation of different methods. To support the integration of different methods into the benchmark and enable fair comparisons, TFB features a flexible and scalable pipeline that eliminates biases. Next, we employ TFB to perform a thorough evaluation of 23 Univariate Time Series Forecasting (UTSF) methods on 1681 univariate time series and 14 Multivariate Time Series Forecasting (MTSF) methods on 28 datasets. The results offer a deeper understanding of the forecasting methods, allowing us to better select the ones that are most suitable for particular datasets and settings. Overall, TFB and this evaluation provide researchers with improved means of designing new TSF methods.

This work is licensed under the Creative Commons Attribution 4.0 International License. You may obtain a copy of the license at <http://creativecommons.org/licenses/by/4.0/>. For any reuse or distribution, you must give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

PVLDB Reference Format:

Shangyu Qiu, Jilin Hu, Lelei Zhou, Xingjian Wu, Jiuyang Du, Huang Zhang, Chenjuan Gao, Aoying Zhou, Christian S. Jensen, Zhenli Sheng and Bao Yang. TFB: Towards Comprehensive and Fair Benchmarking of Time Series Forecasting Methods. PVLDB, 17(6):2104–2117, 2024. doi:10.1177/2158458417683003

PVLDB Artifact Availability:

The source code, data, and/or other artifacts have been made available at <https://github.com/donotforgetme/TFB>.

1 INTRODUCTION

As part of the ongoing digitalization, time series are generated in a variety of domains, such as economic [36, 75], traffic [34, 35–38, 49, 51, 52, 65, 79, 89, 95, 96], health [34, 61, 69, 80], energy [1, 26], and AI Ops [1, 8, 41, 72, 87, 90]. Time Series Forecasting (TSF) is essential for applications in these domains [20, 62, 78, 92]. Given historical observations, it is valuable if we can know the future values ahead of time. Correspondingly, TSF has been firmly established as an active research field, witnessing the proposal of numerous methods.

Time series originate data points chronologically and are either univariate or multivariate depending on the number of variables in each data point. Accordingly, TSF methods can be classified as either Univariate Time Series Forecasting (UTSF) or Multivariate Time Series Forecasting (MTSF) methods. Among early methods, Autoregressive Integrated Moving Average (ARIMA) [4] and Vector Autoregression (VAR) [82] are arguably the most popular univariate and multivariate forecasting methods, respectively. Subsequent methods that exploit machine learning, e.g., XGBoost [11, 95] and Random Forest [5, 98] offer better performance than the early methods. Most recently, methods based on deep learning have demonstrated state-of-the-art (SOTA) forecasting performance on a variety of datasets [10, 12, 18–30, 38, 64, 70, 89, 91, 93, 96, 101, 102, 106].

As more and more methods are being proposed for different datasets and settings, there is an increasing need for fair and open preferable empirical evaluations. To achieve this, we identify and address these issues in existing evaluation frameworks, thereby advancing our evaluation capabilities.

TFB (PVLDB 2024)

- ▶ 25 multivariate datasets
- ▶ 8068 univariate datasets
- ▶ 22 models
- ▶ (open-source!)

TFB: Towards Comprehensive and Fair Benchmarking of Time Series Forecasting Methods

Xiangfei Qiu East China Normal University, China	Jilin Hu East China Normal University, China ^{2S}	Lelei Zhou Huawei Cloud Algorithm Innovation Lab, China	Xingjian Wu East China Normal University, China
Jiuyang Du East China Normal University, China	Huang Zhang East China Normal University, China	Chenjuan Gao East China Normal University, China	Aoying Zhou East China Normal University, China
Christian S. Jensen Aalborg University, Denmark	Zhenli Sheng Huawei Cloud Algorithm Innovation Lab, China	Bin Yang East China Normal University, China	

ABSTRACT

Time series are generated by diverse domains such as economic, traffic, health, and energy, where forecasting of future values has numerous important applications. Not surprisingly, many forecasting methods are being proposed. To ensure progress, it is essential to be able to study and compare such methods especially to cover preferable and suitable manner. To achieve this, we propose TFB, an automated benchmark for Time Series Forecasting (TSF) methods. TFB advances the state-of-the-art by addressing shortcomings related to datasets, comparison methods, and evaluation pipelines: 1) insufficient coverage of data domains, 2) overtype bias against traditional methods, and 3) inconsistent and inflexible pipelines. To achieve better domain coverage, we include datasets from 19 different domains: traffic, electricity, energy, the environment, nature, economic, stock markets, banking, health, and the web. We also provide a time-series characteristic to ensure that the selected datasets are comprehensive. To remove biases against some methods, we include a diverse range of methods, including statistical learning, machine learning, and deep learning methods, and we also support a variety of evaluation strategies and metrics to ensure a more-comprehensive evaluation of different methods. To support the integration of different methods into the benchmark and enable fair comparisons, TFB features a flexible and scalable pipeline that eliminates biases. Next, we employ TFB to perform a thorough evaluation of 23 Univariate Time Series Forecasting (UTSF) methods on 1681 univariate time series and 14 Multivariate Time Series Forecasting (MTSF) methods on 28 datasets. The results offer a deeper understanding of the forecasting methods, allowing us to better select the ones that are most suitable for particular datasets and settings. Overall, TFB and this evaluation provide researchers with improved means of designing new TSF methods.

This work is licensed under the Creative Commons Attribution 4.0 International License. You can find a copy of the license at <https://creativecommons.org/licenses/by/4.0/>. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or by any information storage or retrieval system, without the prior written permission of the IEEE Publishing Group.

Proceedings of the IEEE Information, Vol. 00, No. 0000, 00-00, 2024.

doi:10.1109/0000.0000000

PVLDB Reference Format:

Shangyi Qiu, Jilin Hu, Lelei Zhou, Xingjian Wu, Jiuyang Du, Huang Zhang, Chenjuan Gao, Aoying Zhou, Christian S. Jensen, Zhenli Sheng and Bin Yang. TFB: Towards Comprehensive and Fair Benchmarking of Time Series Forecasting Methods. PVLDB, 17(6): 2300–2311, 2024. doi:10.1177/215838582412683003

PVLDB Artifact Availability:

The source code, data, and/or other artifacts have been made available at <https://github.com/donotforgetme/TFB>.

1 INTRODUCTION

As part of the ongoing digitalization, time series are generated in a variety of domains, such as economic [36, 75], traffic [34, 35–38, 49, 51, 52, 65, 79, 89, 95, 96], health [34, 61, 69, 80], energy [1, 26], and AIops [1, 8, 41, 72, 87, 90]. Time Series Forecasting (TSF) is essential for applications in these domains [20, 62, 78, 93]. Given historical observations, it is valuable if we can know the future values ahead of time. Correspondingly, TSF has been firmly established as an active research field, witnessing the proposal of numerous methods.

Time series originate data points chronologically and are either univariate or multivariate depending on the number of variables in each data point. Accordingly, TSF methods can be classified as either Univariate Time Series Forecasting (UTSF) or Multivariate Time Series Forecasting (MTSF) methods. Among early methods, Autoregressive Integrated Moving Average (ARIMA) [4] and Vector Autoregression (VAR) [82] are arguably the most popular univariate and multivariate forecasting methods, respectively. Subsequent methods that exploit machine learning, e.g., XGBoost [11, 95] and Random Forest [5, 98] offer better performance than the early methods. Most recently, methods based on deep learning have demonstrated state-of-the-art (SOTA) forecasting performance on a variety of datasets [10, 12, 18–30, 39, 68, 84, 70, 89, 91, 92, 96, 101, 102, 106].

As more and more methods are being proposed for different datasets and settings, there is an increasing need for fair and open preferable empirical evaluations. To achieve this, we identify and address these needs in existing evaluation frameworks, thereby advancing our evaluation capabilities.

TFB (PVLDB 2024)

- ▶ 25 multivariate datasets
- ▶ 8068 univariate datasets
- ▶ 22 models
- ▶ (open-source!)

TFB: Towards Comprehensive and Fair Benchmarking of Time Series Forecasting Methods

Xiangfei Qiu East China Normal University, China	Jilin Hu East China Normal University, China ²⁵	Lelei Zhou Huawei Cloud Algorithm Innovation Lab, China	Xingjian Wu East China Normal University, China
Junyang Du East China Normal University, China	Huang Zhang East China Normal University, China	Chenjuan Gao East China Normal University, China	Aoying Zhou East China Normal University, China
Christian S. Jensen Aalborg University, Denmark	Zhenli Sheng Huawei Cloud Algorithm Innovation Lab, China	Biao Yang East China Normal University, China	

ABSTRACT

Time series are generated by diverse domains such as economic, traffic, health, and energy, where forecasting of future values has numerous important applications. Not surprisingly, many forecasting methods are being proposed. To ensure progress, it is essential to be able to study and compare such methods empirically to cover preferable and suitable manner. To achieve this, we propose TFB, an automated benchmark for Time Series Forecasting (TSF) methods. TFB advances the state-of-the-art by addressing shortcomings related to datasets, comparison methods, and evaluation pipelines: 1) insufficient coverage of data domains, 2) overuse bias against traditional methods, and 3) inconsistent and inflexible pipelines. To achieve better domain coverage, we include datasets from 19 different domains: traffic, electricity, energy, the environment, nature, economic, stock markets, banking, health, and the web. We also provide a time-series characteristic to ensure that the selected datasets are comprehensive. To remove biases against some methods, we include a diverse range of methods, including statistical learning, machine learning, and deep learning methods, and we also support a variety of evaluation strategies and metrics to ensure a more-comprehensive evaluation of different methods. To support the integration of different methods into the benchmark and enable fair comparisons, TFB features a flexible and scalable pipeline that eliminates biases. Next, we employ TFB to perform a thorough evaluation of 23 Univariate Time Series Forecasting (UTSF) methods on 1681 univariate time series and 14 Multivariate Time Series Forecasting (MTSF) methods on 28 datasets. The results offer a deeper understanding of the forecasting methods, allowing us to better select the ones that are most suitable for particular datasets and settings. Overall, TFB and this evaluation provide researchers with improved means of designing new TSF methods.

This work is licensed under the Creative Commons Attribution 4.0 International License. You may obtain a copy of the license at <http://creativecommons.org/licenses/by/4.0/>. For more information, contact the publisher or the author.

PVLDB Reference Format:

Shangyi Qiu, Jilin Hu, Lelei Zhou, Xingjian Wu, Junyang Du, Huang Zhang, Chenjuan Gao, Aoying Zhou, Christian S. Jensen, Zhenli Sheng and Biao Yang. TFB: Towards Comprehensive and Fair Benchmarking of Time Series Forecasting Methods. PVLDB, 17(6): 2104–2111, 2024. doi:10.1177/21572455241268303

PVLDB Artifact Availability:

The source code, data, and/or other artifacts have been made available at <https://github.com/donotoverfit/TFB>.

1 INTRODUCTION

As part of the ongoing digitalization, time series are generated in a variety of domains, such as economic [36, 75], traffic [34, 35–38, 49, 51, 52, 65, 79, 89, 95, 96], health [84, 85, 86], energy [5, 26], and AIops [1, 8, 41, 72, 87, 301]. Time Series Forecasting (TSF) is essential for applications in these domains [20, 62, 78, 97]. Given historical observations, it is valuable if we can know the future values ahead of time. Correspondingly, TSF has been firmly established as an active research field, witnessing the proposal of numerous methods.

Time series originate data points chronologically and are either univariate or multivariate depending on the number of variables in each data point. Accordingly, TSF methods can be classified as either Univariate Time Series Forecasting (UTSF) or Multivariate Time Series Forecasting (MTSF) methods. Among early methods, Autoregressive Integrated Moving Average (ARIMA) [4] and Vector Autoregression (VAR) [82] are arguably the most popular univariate and multivariate forecasting methods, respectively. Subsequent methods that exploit machine learning, e.g., XGBoost [11, 95] and Random Forest [5, 98] offer better performance than the early methods. Most recently, methods based on deep learning have demonstrated state-of-the-art (SOTA) forecasting performance on a variety of datasets [10, 12, 18–30, 38, 68, 70, 89, 91, 92, 93, 101, 102, 104].

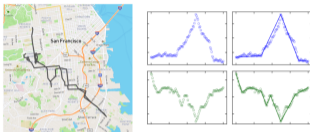
As more and more methods are being proposed for different datasets and settings, there is an increasing need for fair and open preferable empirical evaluations. To achieve this, we identify and address these needs in existing evaluation frameworks, thereby advancing our evaluation capabilities.

TFB (PVLDB 2024)

- ▶ 25 multivariate datasets
- ▶ 8068 univariate datasets
- ▶ 22 models
- ▶ (open-source!)

Take away

Fast linear interpolation

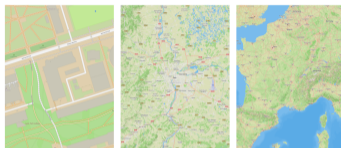


Féry Rous, Olivier Rous, Adrien Lusty-Bérj, Romain Rouvoy: Compact Storage of Data Streams in Mobile Devices. DAIS'24 - 24th International Conference on Distributed Applications and Interoperable Systems, Jun 2024, Groninger, Netherlands. (hal-04535716v2)

2019.0204

loria 310

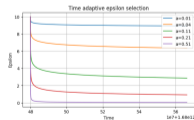
"Right to be forgotten" hint



2019.0204

loria 1019

Error selection



- ▶ Use a function to pick tolerated error
- ▶ Function is time-indexed
- ▶ Different behaviours:
 - Constant value (FLI)
 - Decreasing value:
 - Linear
 - by step
 - with power function

2019.0204

loria 1019

Benchmark



TFB (PVLDB 2024)

- ▶ 25 multivariate datasets
- ▶ 8068 univariate datasets
- ▶ 22 models
- ▶ (open-source!)

2019.0204

loria 1019

Merci.





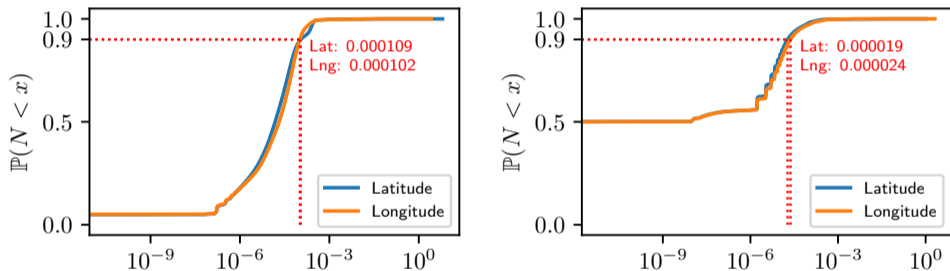
Research questions

Distributed Machine Learning in Ubiquitous Environments using Location-dependent Models

- ▶ How to store unbounded data streams on constrained mobile devices?
- ▶ How to exchange relevant model samples among nearby devices?
- ▶ How to program DML algorithms for the masses?

About the *epsilon* value

- ▶ Selecting a good ϵ value requires **data domain knowledge**
- ▶ Drift between consecutive values (x_1, y_1) and (x_2, y_2) : $|y_2 - y_1|/|x_2 - x_1|$.



- ▶ We used $\epsilon = 10^{-3}$ as a baseline value in the FLI paper