(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

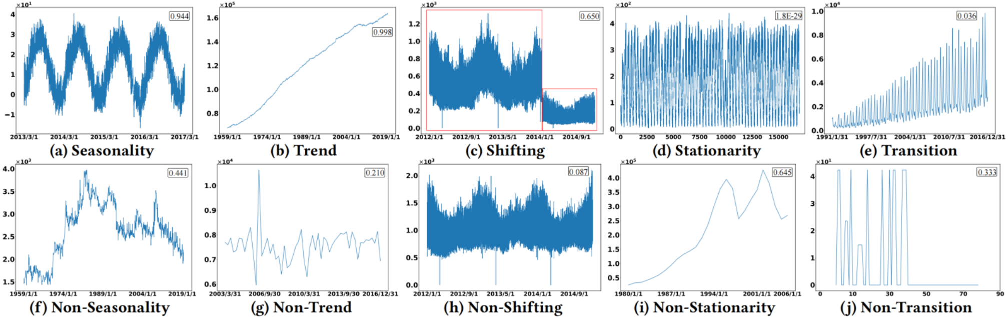


Figure 1: Visualization of data with different characteristics.

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

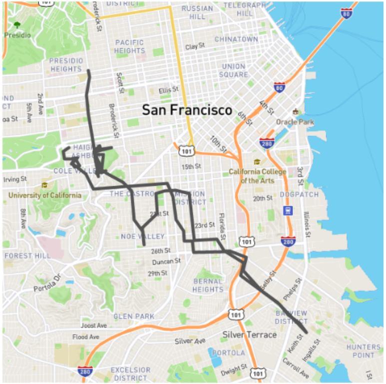
Questions

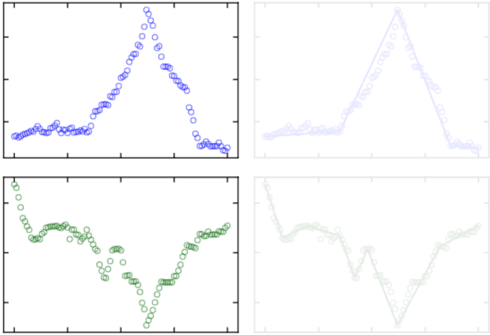
- Devices are producers of data
- Training data better stay local

What are the limits preventing *in situ* computing?

1. Data processing
2. Data storage

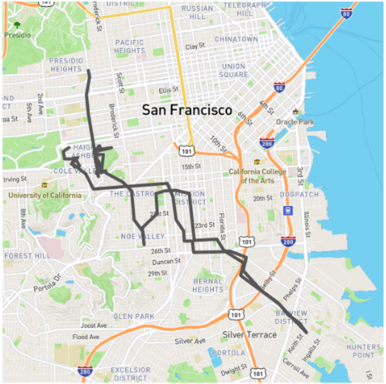
Fast linear interpolation

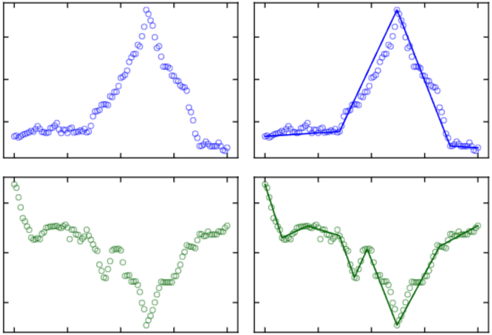




Rémy Raes, Olivier Ruas, Adrien Luxey-Bitri, Romain Rouvoy. Compact Storage of Data Streams in Mobile Devices. *DAIS'24 - 24th International Conference on Distributed Applications and Interoperable Systems, Jun 2024, Groningen, Netherlands.* **(hal-04535716v3)**

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?

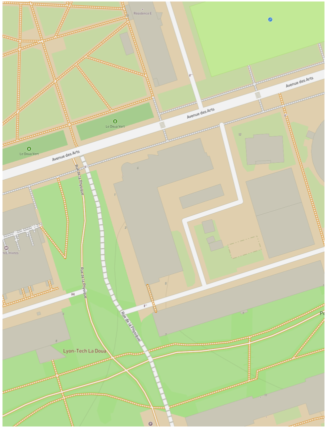
- Can we do better than straight data removal?
- "Right to be forgotten" hint from law community

Additional questions

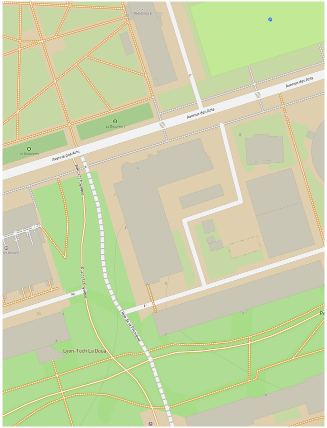
Can we go further on with time series compression?

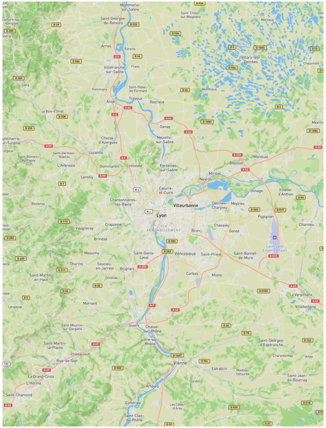
- Can we do better than straight data removal?
- ”Right to be forgotten” hint from law community

"Right to be forgotten" hint

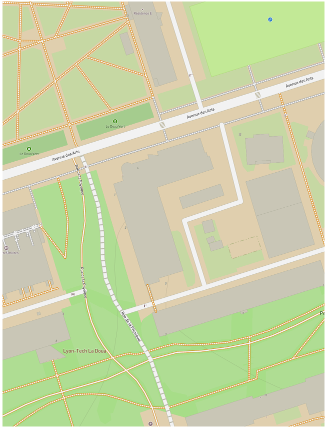


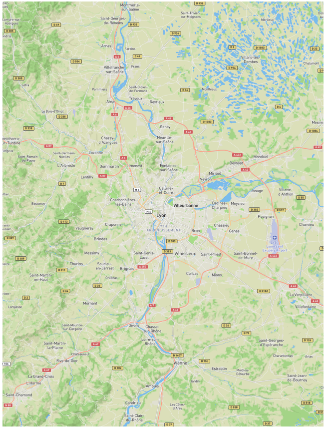
"Right to be forgotten" hint



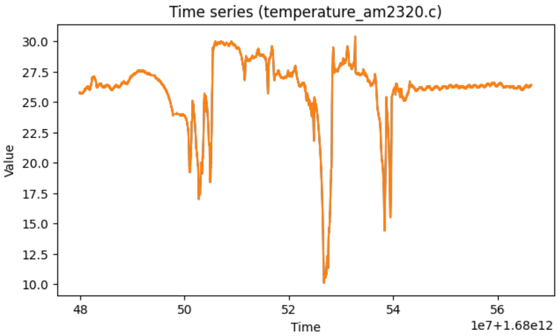


"Right to be forgotten" hint



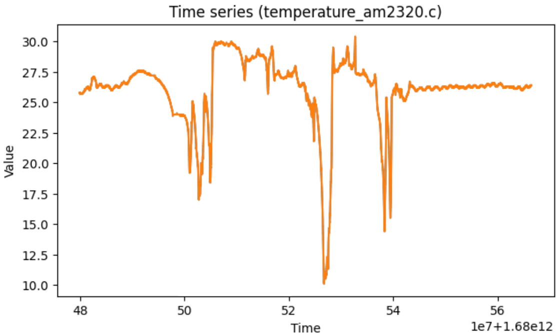


Translation into real world



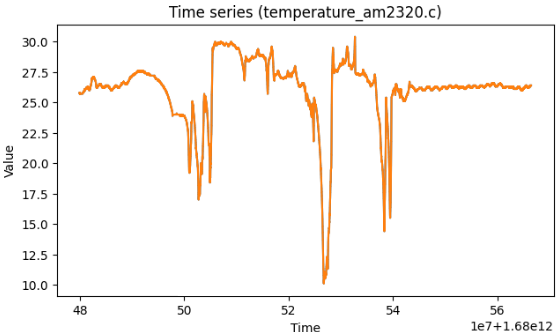
- Time series to be compressed
- Prioritize old data for compression
- Need for a time-dependent compression method

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

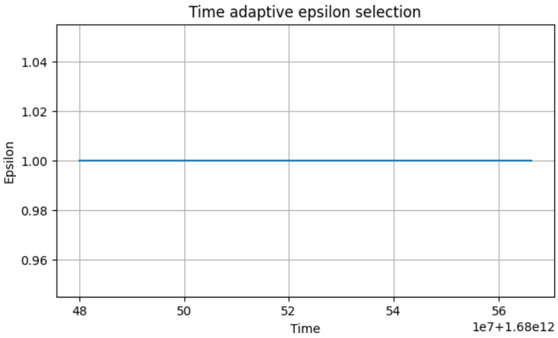
Error selection

- 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

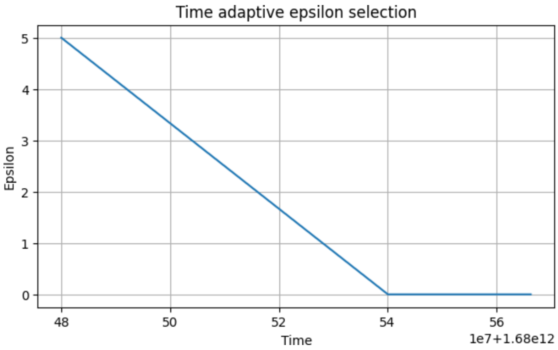


- Use a function to pick tolerated error
- Function is time-indexed
- Different behaviours:
  - Constant value (FLI)
  - Decreasing value:
    - Linear
    - by step
    - with power function

Error selection

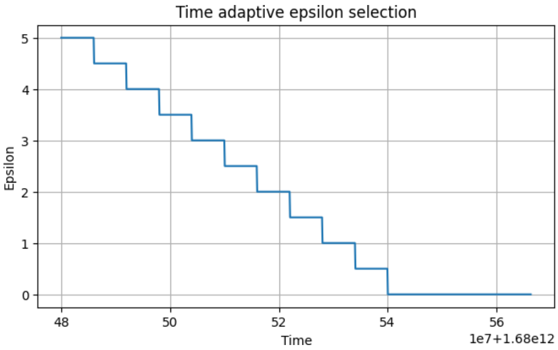
- Use a function to pick tolerated error
- Function is time-indexed
- Different behaviours:
  - Constant value (FLI)
  - Decreasing value:
    - Linear
    - by step
    - with power function

Error selection



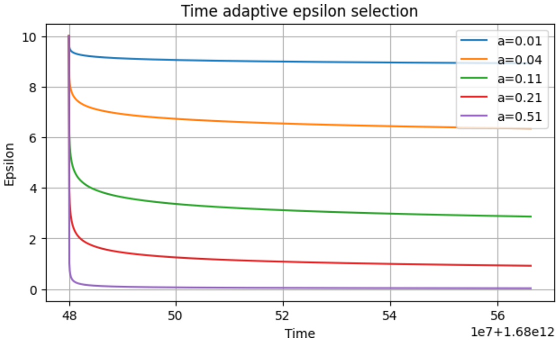
- Use a function to pick tolerated error
- Function is time-indexed
- Different behaviours:
  - Constant value (FLI)
  - Decreasing value:
    - Linear
    - by step
    - with power function

Error selection



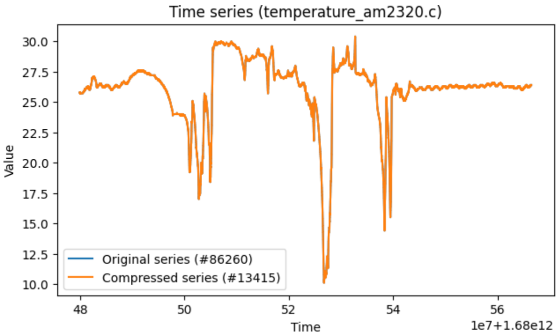
- Use a function to pick tolerated error
- Function is time-indexed
- Different behaviours:
  - Constant value (FLI)
  - Decreasing value:
    - Linear
    - by step
    - with power function

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

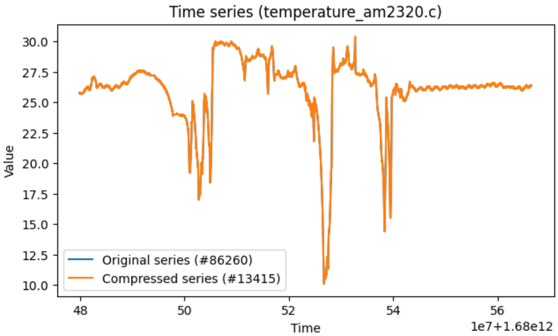
Outlayers relative conservation



Knobs to control them all

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

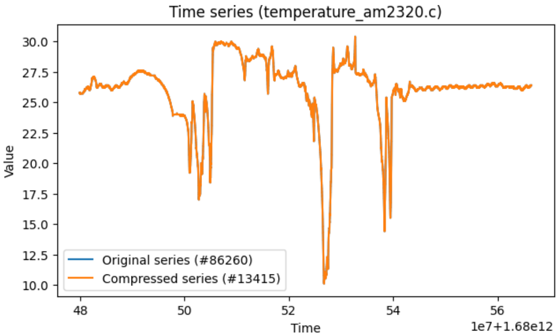
Outlayers relative conservation



Knobs to control them all

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

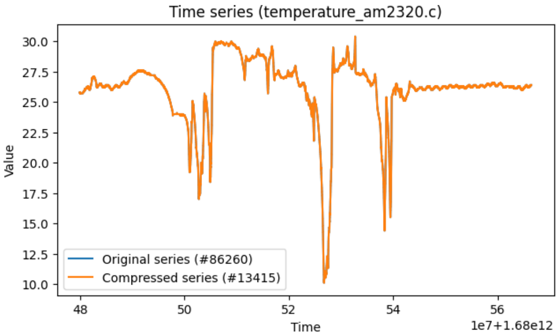
Outlayers relative conservation



Knobs to control them all

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

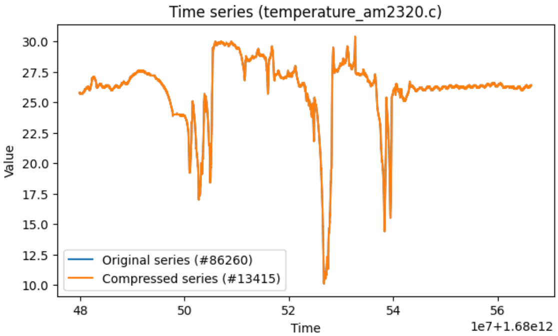
Outlayers relative conservation



Knobs to control them all

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

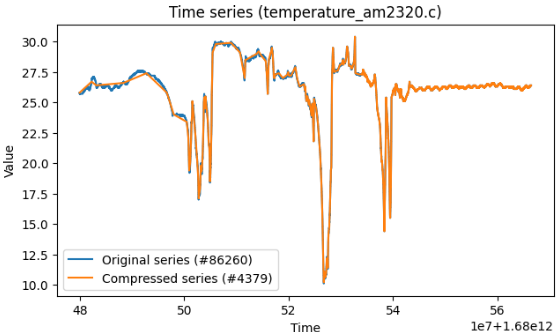
Outlayers relative conservation



Knobs to control them all

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

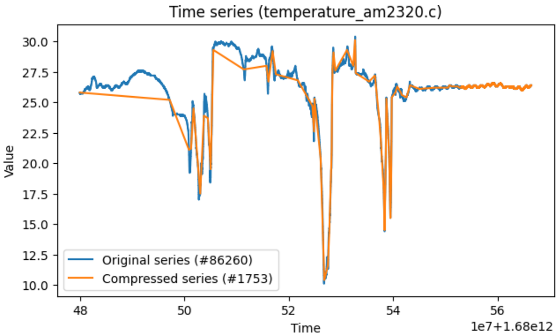
Outlayers relative conservation



Knobs to control them all

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

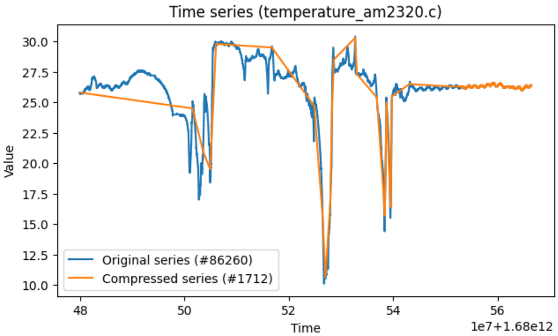
Outlayers relative conservation



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?

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

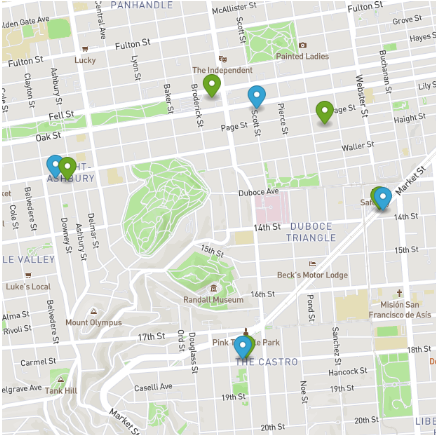


Figure – Points of Interest computed using raw data and FLI-modeled data.

Data utility with location data

Size results

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

Data utility

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

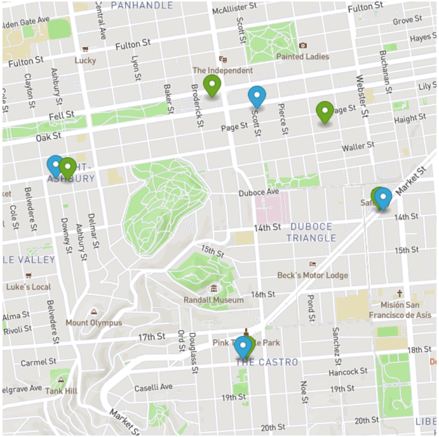


Figure – Points of Interest computed using raw data and FLI-modeled data.

Data utility with location data

Size results

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

Data utility

- Latitude
  - Tolerated error: 10^{-3} *deg* ≈ 111 *m*
  - Median error: 5.33×10^{-5}
  - RMSE: 3.72×10^{-4}
- Longitude
  - Tolerated error: 10^{-3} *deg* ≈ 88 *m*
  - Median error: 2.81×10^{-5}
  - RMSE: 3.44×10^{-4}

- Privacy utility

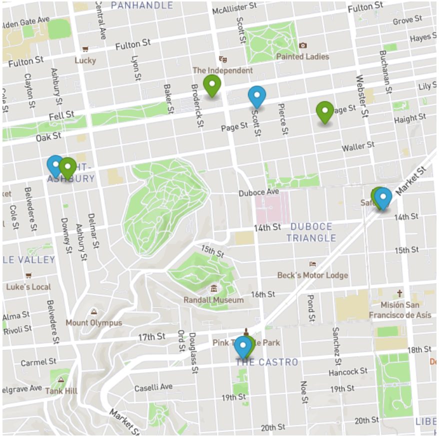


Figure – Points of Interest computed using raw data and FLI-modeled data.

Data utility with location data

Size results

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

Data utility

- Latitude
  - Tolerated error: 10^{-3} *deg* ≈ 111 *m*
  - Median error: 5.33×10^{-5}
  - RMSE: 3.72×10^{-4}
- Longitude
  - Tolerated error: 10^{-3} *deg* ≈ 88 *m*
  - Median error: 2.81×10^{-5}
  - RMSE: 3.44×10^{-4}
- Privacy utility

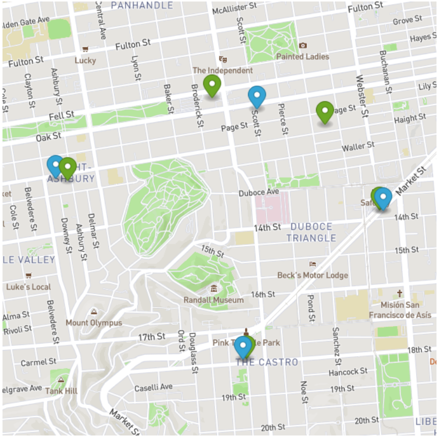


Figure – Points of Interest computed using raw data and FLI-modeled data.

(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

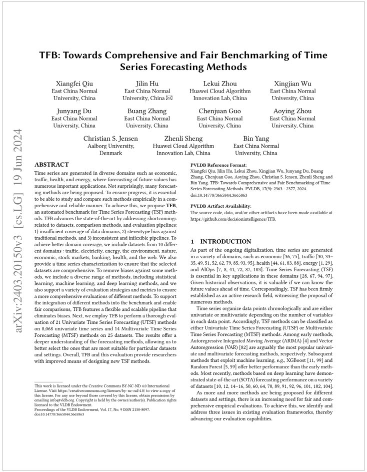
(Future) Evaluation

- 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

Benchmark



arXiv:2403.20150v3 [cs.LG] 19 Jun 2024

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 ✉

Lekui Zhou
Huawei Cloud Algorithm Innovation Lab, China

Xingjian Wu
East China Normal University, China

Junyang Du
East China Normal University, China

Buang Zhang
East China Normal University, China

Chenjuan Guo
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 in 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 in a comprehensive and reliable 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) stereotype bias against traditional methods, and 3) inconsistent and inflexible pipelines. To achieve better domain coverage, we include datasets from 10 different domains : traffic, electricity, energy, the environment, nature, economic, stock markets, banking, health, and the web. We also provide a time series characterization 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 evaluations 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 8,068 univariate time series and 14 Multivariate Time Series Forecasting (MTSF) methods on 25 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.

Proceedings of the VLDB Endowment, Vol. 17, No. 9 ISSN 2150-8097.
doi:10.14778/3665844.3665863

PVLDB Reference Format:
Xiangfei Qiu, Jilin Hu, Lekui Zhou, Xingjian Wu, Junyang Du, Buang Zhang, Chenjuan Guo, Aoying Zhou, Christian S. Jensen, Zhenli Sheng and Bin Yang. TFB: Towards Comprehensive and Fair Benchmarking of Time Series Forecasting Methods. PVLDB, 17(9): 2363 - 2377, 2024.
doi:10.14778/3665844.3665863

PVLDB Artifact Availability:
The source code, data, and/or other artifacts have been made available at https://github.com/decisionintelligence/TFB.

1 INTRODUCTION

As part of the ongoing digitalization, time series are generated in a variety of domains, such as economic [36, 75], traffic [30, 33–35, 49, 51, 52, 62, 79, 85, 93, 95], health [44, 61, 83, 88], energy [1, 29], and AIOps [7, 8, 41, 72, 87, 103]. Time Series Forecasting (TSF) is essential in key applications in these domains [28, 67, 94, 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.

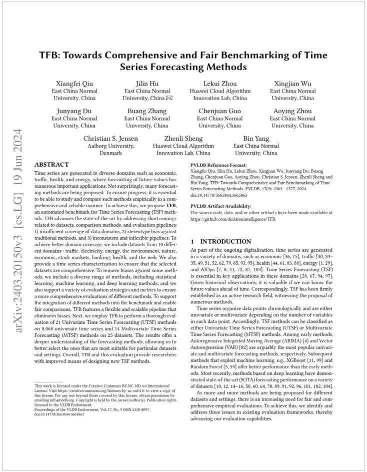
Time series organize data points chronologically and are either univariate or multivariate depending on the number of variables in each data point. Accordingly, TSF methods can be classified as either Univariate Time Series Forecasting (UTSF) or Multivariate Time Series Forecasting (MTSF) methods. Among 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, 99] and Random Forest [5, 59] 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, 14–16, 50, 60, 64, 70, 89, 91, 92, 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 comprehensive empirical evaluations. To achieve this, we identify and address three issues in existing evaluation frameworks, thereby advancing our evaluation capabilities.

TFB (PVLDB 2024)

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

Benchmark



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 ✉
Lekui Zhou, Huawei Cloud Algorithm Innovation Lab, China
Xingjian Wu, East China Normal University, China
Junyang Du, East China Normal University, China
Buang Zhang, East China Normal University, China
Chenjuan Guo, 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

arXiv:2403.20150v3 [cs.LG] 19 Jun 2024

ABSTRACT

Time series are generated in 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 in a comprehensive and reliable 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) stereotype bias against traditional methods, and 3) inconsistent and inflexible pipelines. To achieve better domain coverage, we include datasets from 10 different domains : traffic, electricity, energy, the environment, nature, economic, stock markets, banking, health, and the web. We also provide a time series characterization 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 evaluations 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 8,068 univariate time series and 14 Multivariate Time Series Forecasting (MTSF) methods on 25 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.

Proceedings of the VLDB Endowment, Vol. 17, No. 9 ISSN 2150-8097.
doi:10.14778/3665844.3665863

PVLDB Reference Format:
Xiangfei Qiu, Jilin Hu, Lekui Zhou, Xingjian Wu, Junyang Du, Buang Zhang, Chenjuan Guo, Aoying Zhou, Christian S. Jensen, Zhenli Sheng and Bin Yang. TFB: Towards Comprehensive and Fair Benchmarking of Time Series Forecasting Methods. PVLDB, 17(9): 2363 - 2377, 2024.
doi:10.14778/3665844.3665863

PVLDB Artifact Availability:
The source code, data, and/or other artifacts have been made available at https://github.com/decisionintelligence/TFB.

1 INTRODUCTION

As part of the ongoing digitalization, time series are generated in a variety of domains, such as economic [36, 75], traffic [30, 33–35, 49, 51, 52, 62, 79, 85, 93, 95], health [44, 61, 83, 88], energy [1, 29], and AIOps [7, 8, 41, 72, 87, 103]. Time Series Forecasting (TSF) is essential in key applications in these domains [28, 67, 94, 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.

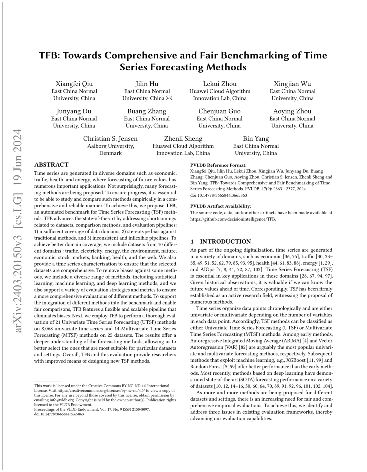
Time series organize data points chronologically and are either univariate or multivariate depending on the number of variables in each data point. Accordingly, TSF methods can be classified as either Univariate Time Series Forecasting (UTSF) or Multivariate Time Series Forecasting (MTSF) methods. Among 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, 99] and Random Forest [5, 59] 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, 14–16, 50, 60, 64, 70, 89, 91, 92, 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 comprehensive empirical evaluations. To achieve this, we identify and address three issues in existing evaluation frameworks, thereby advancing our evaluation capabilities.

TFB (PVLDB 2024)

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

Benchmark



arXiv:2403.20150v3 [cs.LG] 19 Jun 2024

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 ✉
Lekui Zhou, Huawei Cloud Algorithm Innovation Lab, China
Xingjian Wu, East China Normal University, China
Junyang Du, East China Normal University, China
Buang Zhang, East China Normal University, China
Chenjuan Guo, 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 in 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 in a comprehensive and reliable 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) stereotype bias against traditional methods, and 3) inconsistent and inflexible pipelines. To achieve better domain coverage, we include datasets from 10 different domains : traffic, electricity, energy, the environment, nature, economic, stock markets, banking, health, and the web. We also provide a time series characterization 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 evaluations 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 8,068 univariate time series and 14 Multivariate Time Series Forecasting (MTSF) methods on 25 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.

Proceedings of the VLDB Endowment, Vol. 17, No. 9 ISSN 2150-8097.
doi:10.14778/3665844.3665863

PVLDB Reference Format:
Xiangfei Qiu, Jilin Hu, Lekui Zhou, Xingjian Wu, Junyang Du, Buang Zhang, Chenjuan Guo, Aoying Zhou, Christian S. Jensen, Zhenli Sheng and Bin Yang. TFB: Towards Comprehensive and Fair Benchmarking of Time Series Forecasting Methods. PVLDB, 17(9): 2363 - 2377, 2024.
doi:10.14778/3665844.3665863

PVLDB Artifact Availability:
The source code, data, and/or other artifacts have been made available at https://github.com/decisionintelligence/TFB.

1 INTRODUCTION

As part of the ongoing digitalization, time series are generated in a variety of domains, such as economic [36, 75], traffic [30, 33–35, 49, 51, 52, 62, 79, 85, 93, 95], health [44, 61, 83, 88], energy [1, 29], and AIOps [7, 8, 41, 72, 87, 103]. Time Series Forecasting (TSF) is essential in key applications in these domains [28, 67, 94, 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.

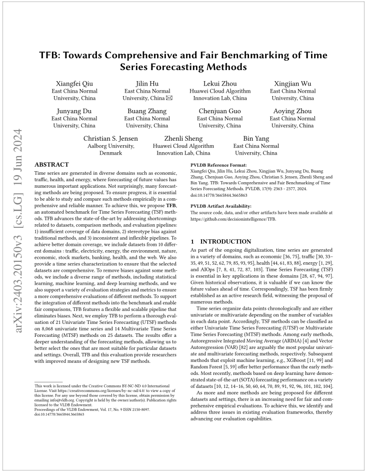
Time series organize data points chronologically and are either univariate or multivariate depending on the number of variables in each data point. Accordingly, TSF methods can be classified as either Univariate Time Series Forecasting (UTSF) or Multivariate Time Series Forecasting (MTSF) methods. Among 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, 99] and Random Forest [5, 59] 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, 14–16, 50, 60, 64, 70, 89, 91, 92, 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 comprehensive empirical evaluations. To achieve this, we identify and address three issues in existing evaluation frameworks, thereby advancing our evaluation capabilities.

TFB (PVLDB 2024)

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

Benchmark



arXiv:2403.20150v3 [cs.LG] 19 Jun 2024

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 ✉

Lekui Zhou
Huawei Cloud Algorithm Innovation Lab, China

Xingjian Wu
East China Normal University, China

Junyang Du
East China Normal University, China

Buang Zhang
East China Normal University, China

Chenjuan Guo
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 in 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 in a comprehensive and reliable 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) stereotype bias against traditional methods, and 3) inconsistent and inflexible pipelines. To achieve better domain coverage, we include datasets from 10 different domains : traffic, electricity, energy, the environment, nature, economic, stock markets, banking, health, and the web. We also provide a time series characterization 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 evaluations 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 8,068 univariate time series and 14 Multivariate Time Series Forecasting (MTSF) methods on 25 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.

Proceedings of the VLDB Endowment, Vol. 17, No. 9 ISSN 2150-8097.
doi:10.14778/3665844.3665863

PVLDB Reference Format:
Xiangfei Qiu, Jilin Hu, Lekui Zhou, Xingjian Wu, Junyang Du, Buang Zhang, Chenjuan Guo, Aoying Zhou, Christian S. Jensen, Zhenli Sheng and Bin Yang. TFB: Towards Comprehensive and Fair Benchmarking of Time Series Forecasting Methods. PVLDB, 17(9): 2363 - 2377, 2024.
doi:10.14778/3665844.3665863

PVLDB Artifact Availability:
The source code, data, and/or other artifacts have been made available at https://github.com/decisionintelligence/TFB.

1 INTRODUCTION

As part of the ongoing digitalization, time series are generated in a variety of domains, such as economic [36, 75], traffic [30, 33–35, 49, 51, 52, 62, 79, 85, 93, 95], health [44, 61, 83, 88], energy [1, 29], and AIOps [7, 8, 41, 72, 87, 103]. Time Series Forecasting (TSF) is essential in key applications in these domains [28, 67, 94, 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.

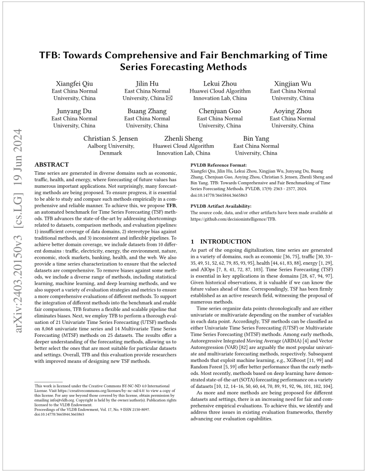
Time series organize data points chronologically and are either univariate or multivariate depending on the number of variables in each data point. Accordingly, TSF methods can be classified as either Univariate Time Series Forecasting (UTSF) or Multivariate Time Series Forecasting (MTSF) methods. Among 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, 99] and Random Forest [5, 59] 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, 14–16, 50, 60, 64, 70, 89, 91, 92, 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 comprehensive empirical evaluations. To achieve this, we identify and address three issues in existing evaluation frameworks, thereby advancing our evaluation capabilities.

TFB (PVLDB 2024)

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

Benchmark



arXiv:2403.20150v3 [cs.LG] 19 Jun 2024

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 ✉
Lekui Zhou, Huawei Cloud Algorithm Innovation Lab, China
Xingjian Wu, East China Normal University, China
Junyang Du, East China Normal University, China
Buang Zhang, East China Normal University, China
Chenjuan Guo, 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 in 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 in a comprehensive and reliable 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) stereotype bias against traditional methods, and 3) inconsistent and inflexible pipelines. To achieve better domain coverage, we include datasets from 10 different domains : traffic, electricity, energy, the environment, nature, economic, stock markets, banking, health, and the web. We also provide a time series characterization 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 evaluations 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 8,068 univariate time series and 14 Multivariate Time Series Forecasting (MTSF) methods on 25 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.

Proceedings of the VLDB Endowment, Vol. 17, No. 9 ISSN 2150-8097.
doi:10.14778/3665844.3665863

PVLDB Reference Format:
Xiangfei Qiu, Jilin Hu, Lekui Zhou, Xingjian Wu, Junyang Du, Buang Zhang, Chenjuan Guo, Aoying Zhou, Christian S. Jensen, Zhenli Sheng and Bin Yang. TFB: Towards Comprehensive and Fair Benchmarking of Time Series Forecasting Methods. PVLDB, 17(9): 2363 - 2377, 2024.
doi:10.14778/3665844.3665863

PVLDB Artifact Availability:
The source code, data, and/or other artifacts have been made available at https://github.com/decisionintelligence/TFB.

1 INTRODUCTION

As part of the ongoing digitalization, time series are generated in a variety of domains, such as economic [36, 75], traffic [30, 33–35, 49, 51, 52, 62, 79, 85, 93, 95], health [44, 61, 83, 88], energy [1, 29], and AIOps [7, 8, 41, 72, 87, 103]. Time Series Forecasting (TSF) is essential in key applications in these domains [28, 67, 94, 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 organize 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, 99] and Random Forest [5, 59] 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, 14–16, 50, 60, 64, 70, 89, 91, 92, 96, 101, 102, 104].

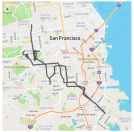
As more and more methods are being proposed for different datasets and settings, there is an increasing need for fair and comprehensive empirical evaluations. To achieve this, we identify and address three issues in existing evaluation frameworks, thereby advancing our evaluation capabilities.

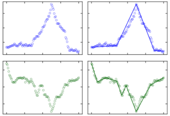
TFB (PVLDB 2024)

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

Take away

Fast linear interpolation

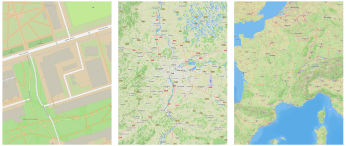




Rémy Raes, Olivier Ruas, Adrien Luxey-Bitri, Romain Rouvoy. Compact Storage of Data Streams in Mobile Devices. *DAIS'24 - 24th International Conference on Distributed Applications and Interoperable Systems, Jun 2024, Groningen, Netherlands.* **(hal-04535716v3)**

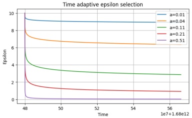
27/11/2024 7/19

"Right to be forgotten" hint



27/11/2024 10/19

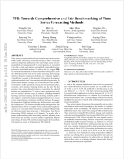
Error selection



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

27/11/2024 12/19

Benchmark



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

TFB (PVLDB 2024)

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

27/11/2024 17/19

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): $|y_2 - y_1|/|x_2 - x_1|$.

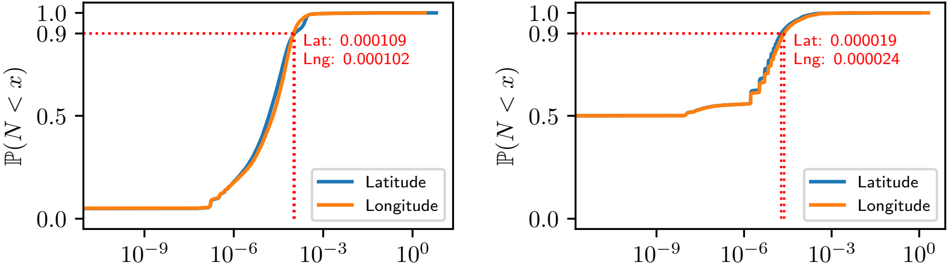


Figure – CDF of latitude and longitude variations of successive locations in Cabspotting and PrivaMov.

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