MRES PROJECT

Carrow Morris-Wiltshire[1]

c.morris-wiltshire@ncl.ac.uk

Supervised by: Professor Stuart Barr[1], Professor Phil James[1], Keith Hermiston[2]

[1] School of Engineering, Newcastle University, Newcastle upon Tyne, UK
[2] DSTL, Newcastle upon Tyne, UK

Thesis

Introduction

Carrow Morris-Wiltshire

Supervisee

MRes Student

Geospatial Systems CDT, Newcastle University

Prof. Stuart Barr

Primary Supervisor

Professor of Geospatial Systems Engineering

School of Engineering, Newcastle University

Prof. Phil James

Secondary Supervisor

Professor of Urban Data

School of Engineering, Newcastle University

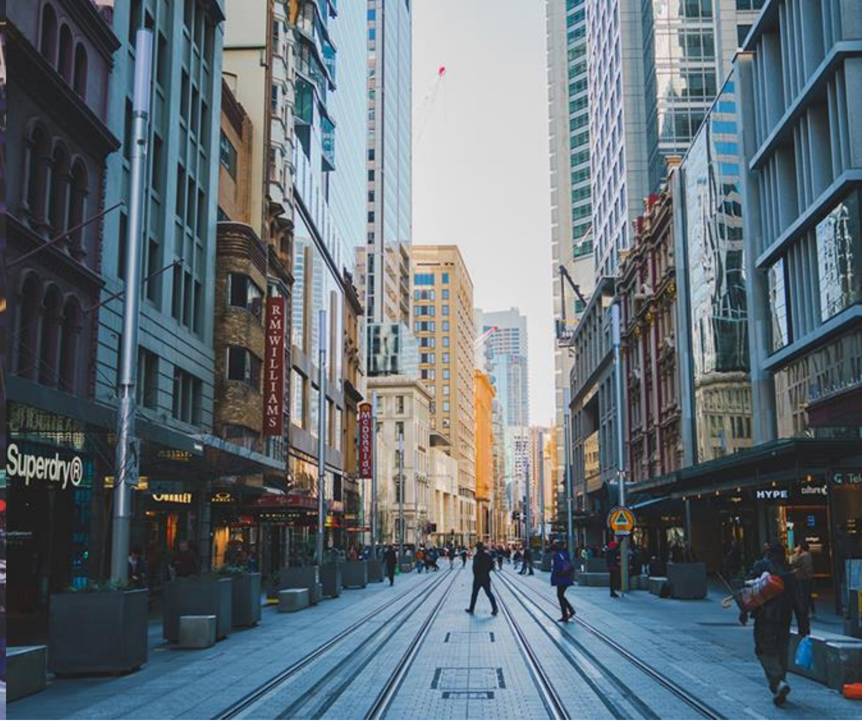
Keith Hermiston

Industry Supervisor

Principal Scientist

Cyber and Information Systems Division, DSTL

TOWARDS SITUATIONAL AWARENESS OF URBAN PEDESTRIAN FLOWS: **AN ASSESSMENT OF DATA QUALITY ON THE PERFORMANCE OF PREDICTIVE MODELS**



Project Deliverables:

- Literature Review
- Data Exploration Report
- Methodology
- Results/Discussion/Future Steps

[dstl] GEOSPATIAL SYSTEMS Newcastle University

Presentation Contents

Background

- **What is situational awareness?**
- **How might situational awareness of pedestrians aid urban decision making?**
- **What are the critical barriers to situational awareness of pedestrians in Newcastle upon Tyne?**

The DSTL Perspective

"The perception of environmental elements and events concerning time or space, the comprehension of their meaning, and the projection of their status in the future. "
(Endsley 1995)

1. **Perception**: Automated data visualisation
2. **Comprehension**: Automated data analytics and insights
3. **Projection**: Automated predictive capabilities

How might SA of pedestrians be useful?

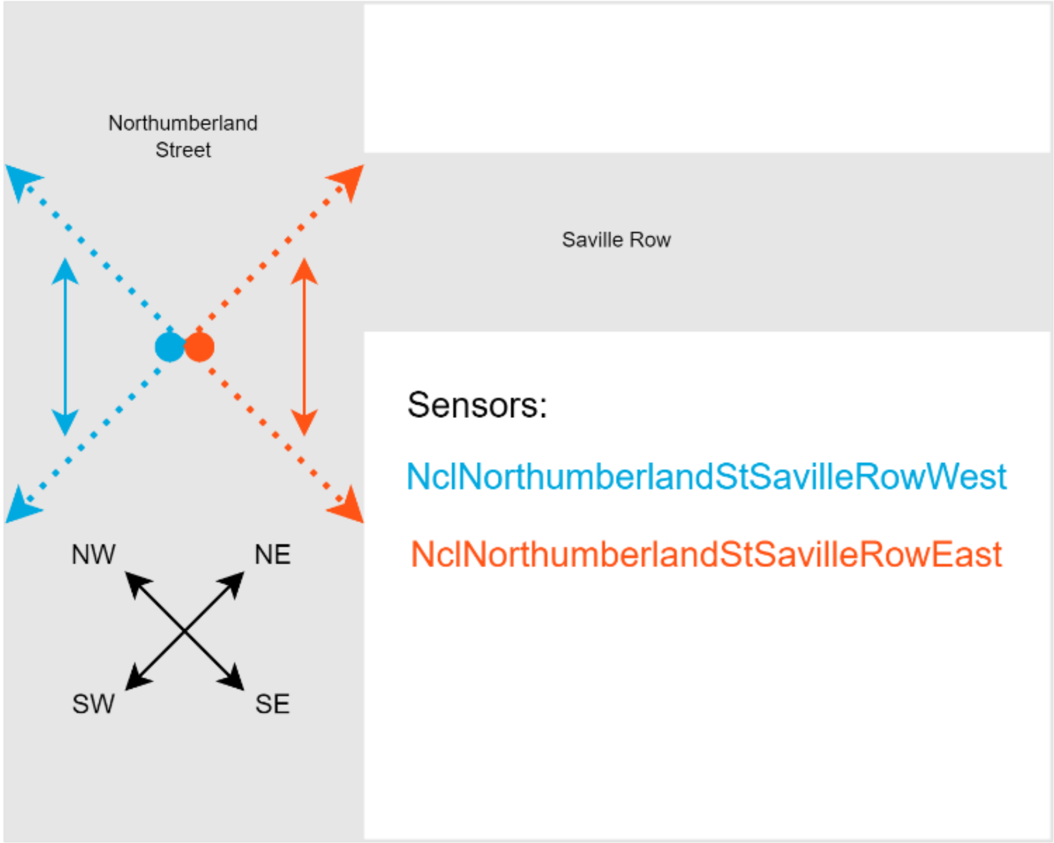
5. **Automated Reporting and Regulation**

Pedestrian flows can be used in automated systems that require real-time data to trigger alerts or actions, such as adjusting traffic light timings before/during peak pedestrian movement or initiating crowd-control measures before/during events.

8. **Agile, Data-Centric Decision-Making**

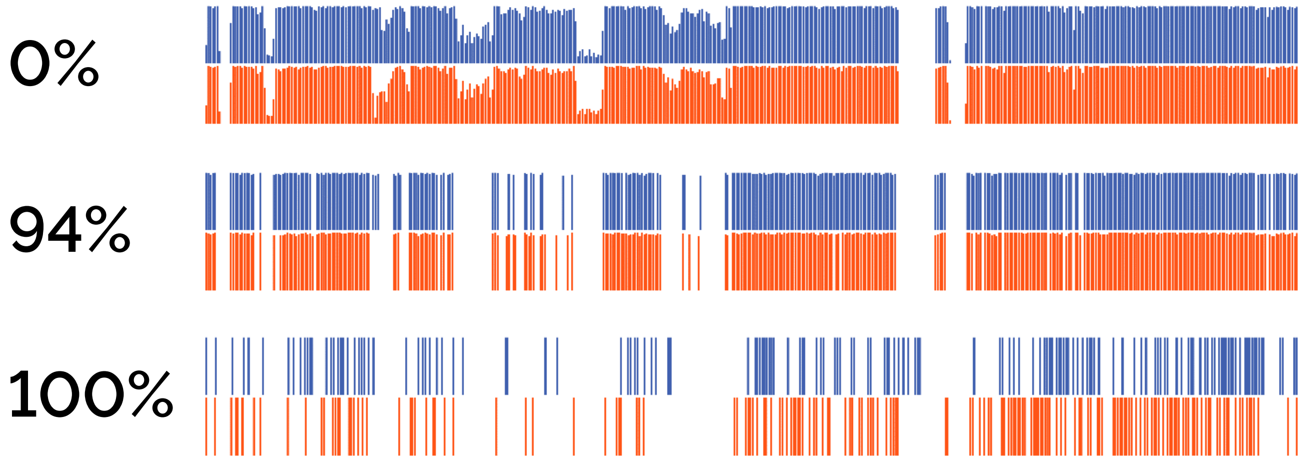
Real-time pedestrian flow data can inform immediate decisions, like directing traffic, deploying police for crowd control, or sending push notifications to guide people towards less crowded areas during events.

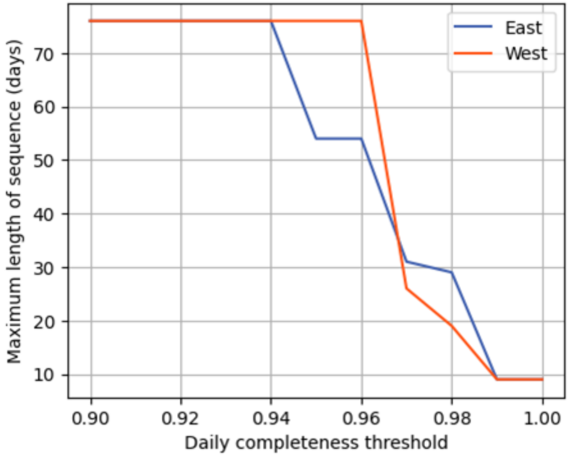
The Data


Northumberland Street
Saville Row
Sensors:
NclNorthumberlandStSavilleRowWest
NclNorthumberlandStSavilleRowEast
NW
NE
SW
SE

But….

- **Data is not immediately useable due to its quality…**





There fore...

What quality of data is needed to create a model that can predict pedestrian flow?

Test Objectives

- **Anomaly prediction test**

1. Develop a modelling workflow that identifies anomalies in the data using single-step univariate prediction validating on unseen data.

2. Generate additional input features (feature engineering) to assess how the model's performance changes for anomaly prediction.

- **Horizon test**

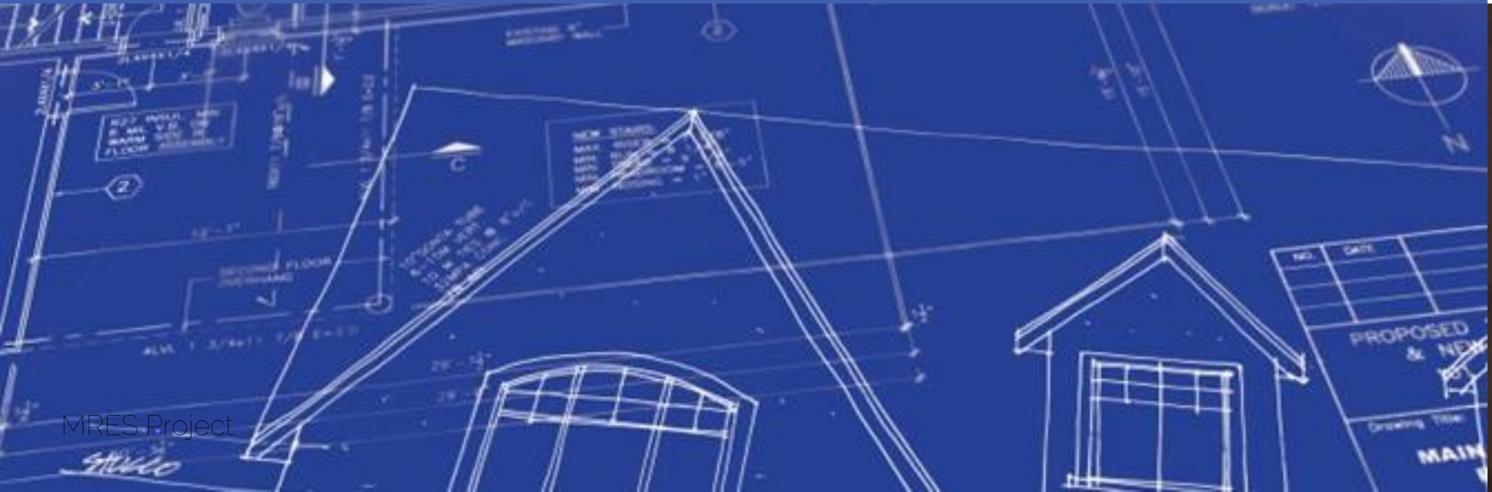
3. Measure the change in prediction accuracy as the prediction horizon increases for univariate and multivariate models.

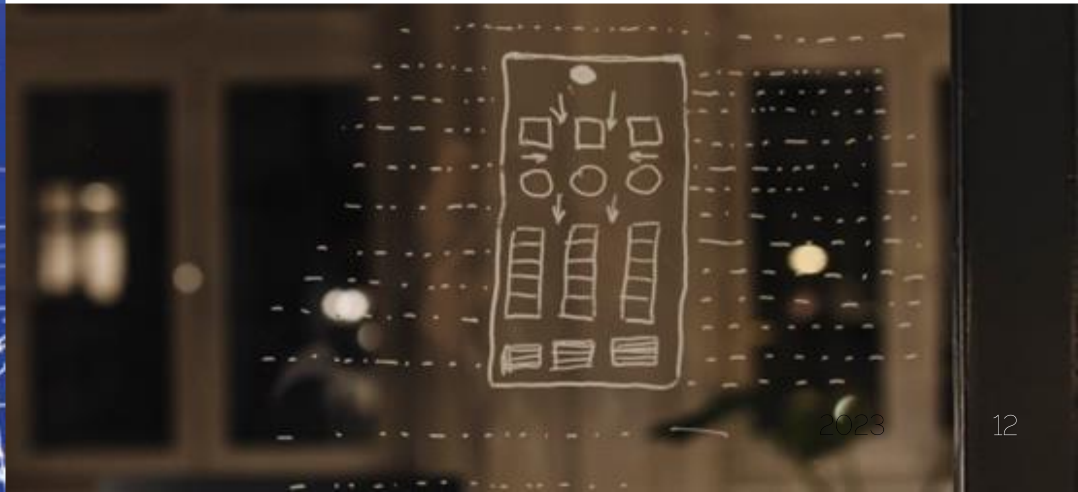
- **Data robustness test**

4. Measure the change in prediction accuracy as the data completeness of the training data is reduced for univariate and multivariate models.

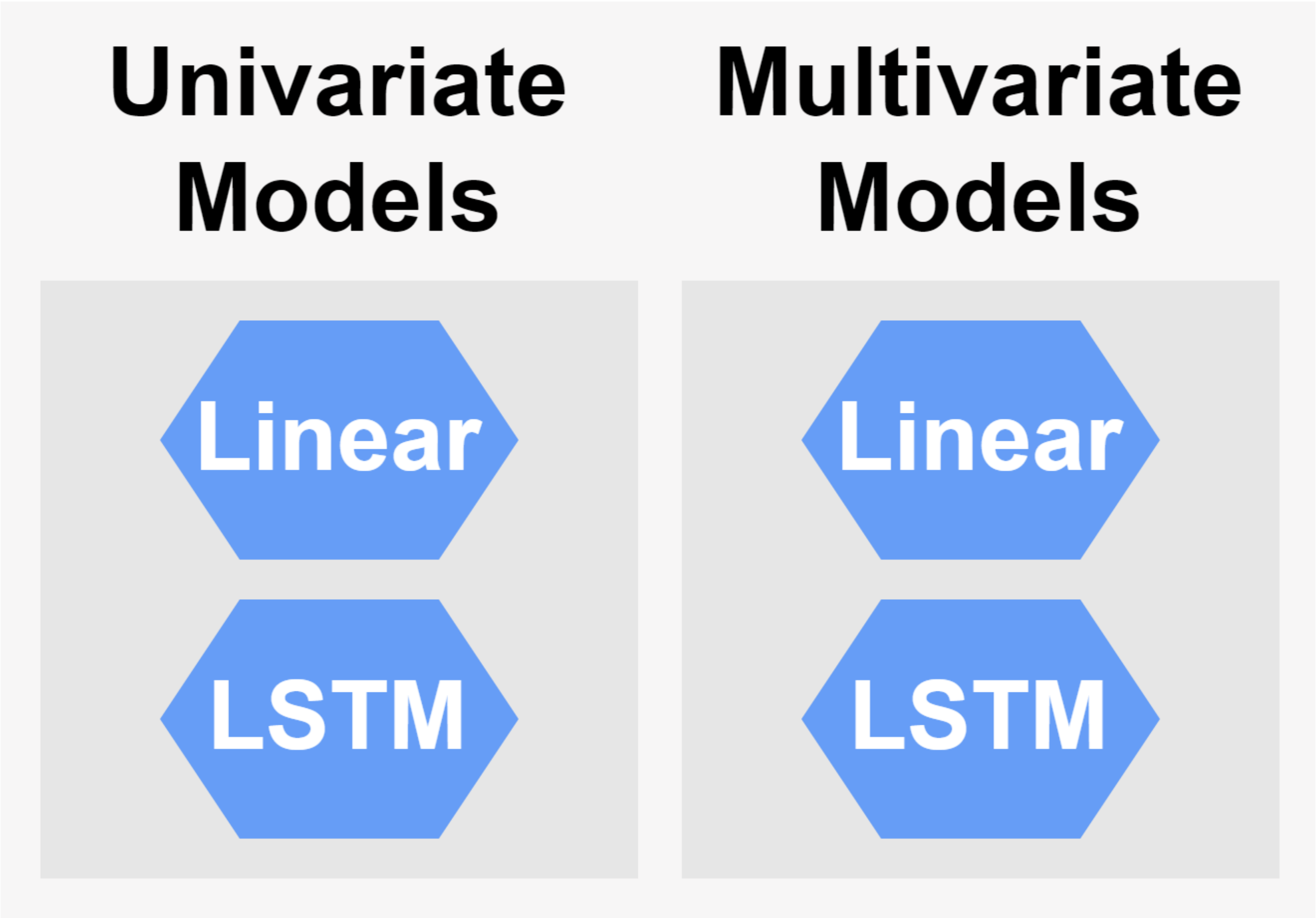
Methodology

- **What models are used?**
- **What additional features are chosen for multivariate models?**
- **Which hyperparameters have been chosen?**



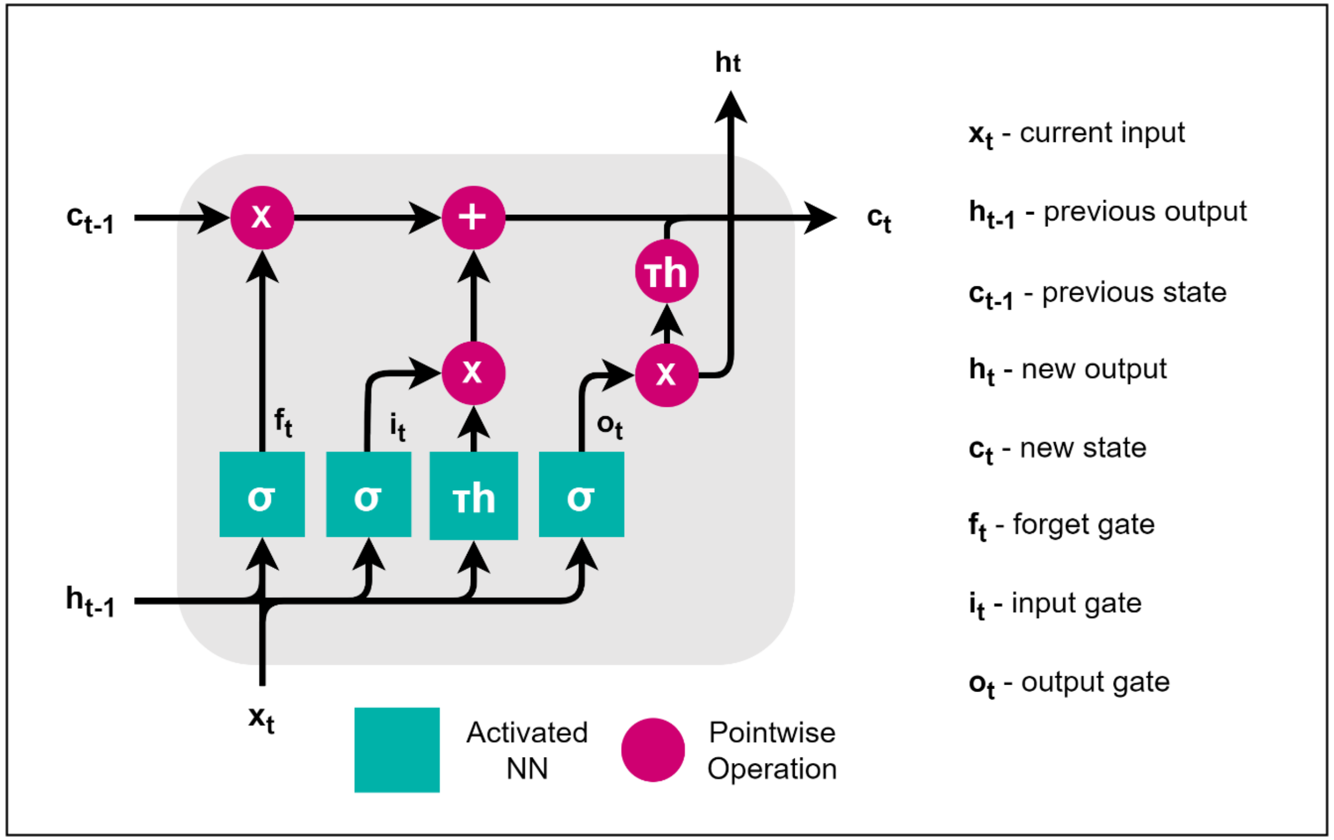


Model Selection



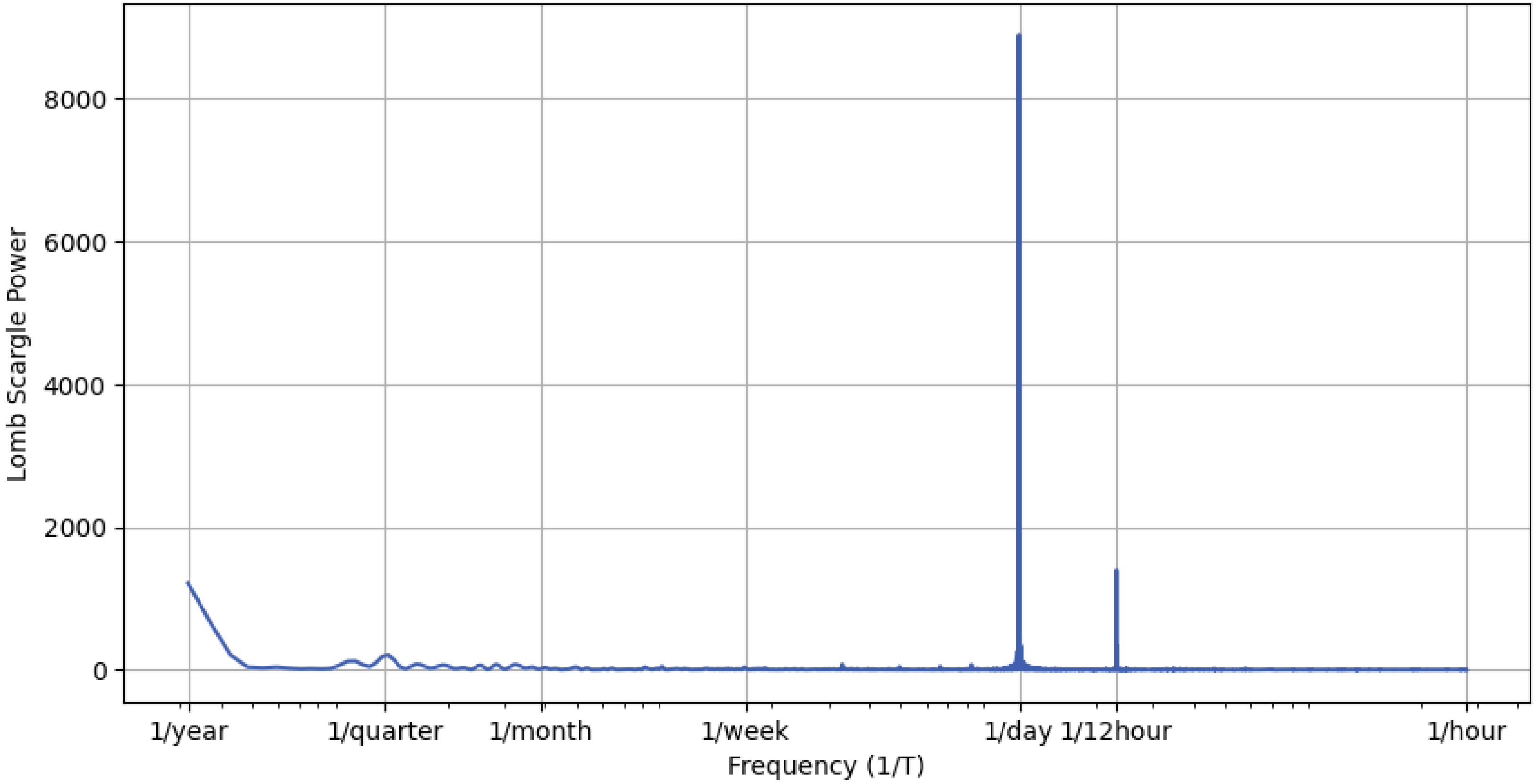
LSTM (Long Short-Term Memory)

An LSTM cell includes a forget gate which allows long term patterns to be stored

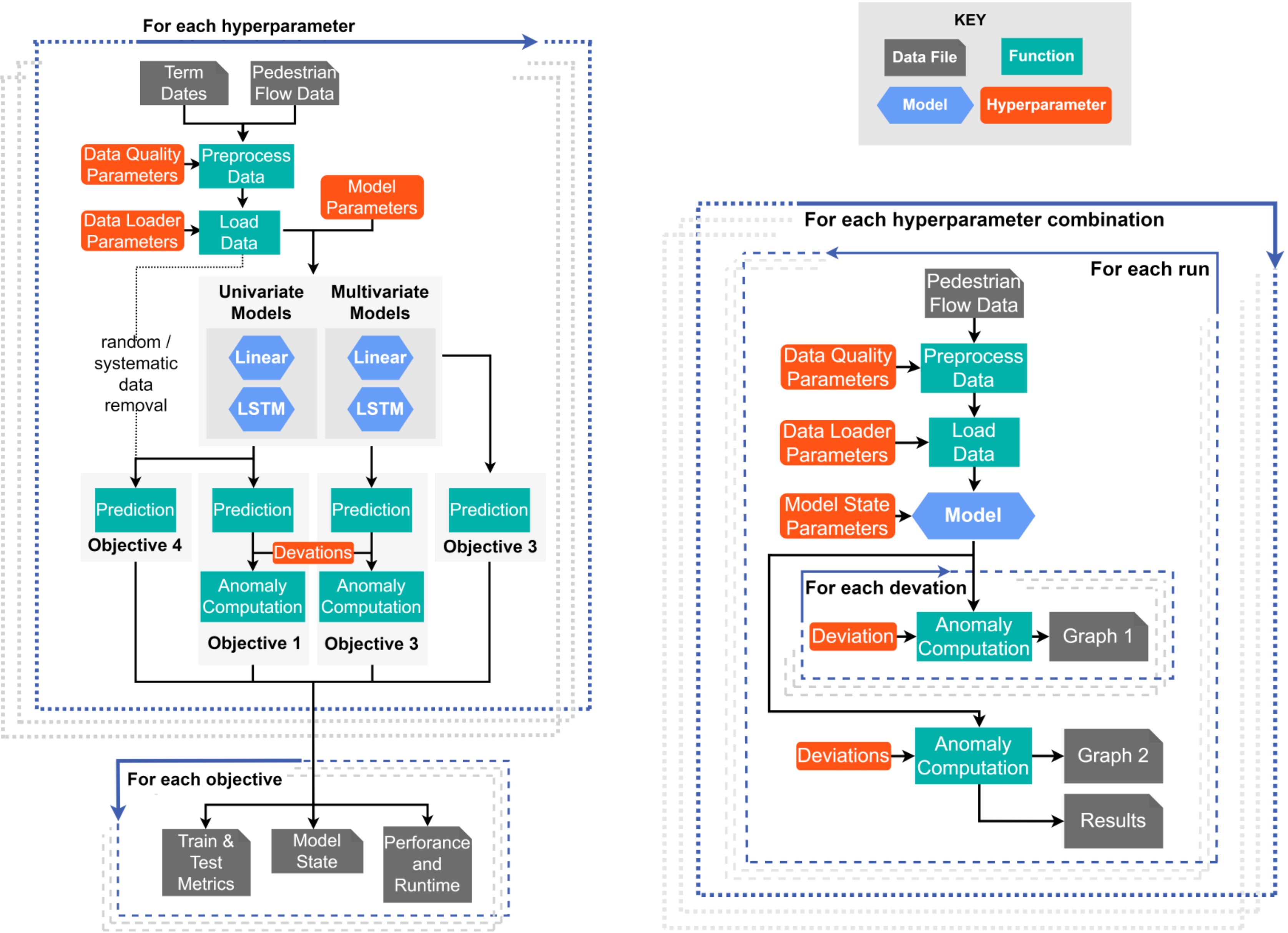


Feature Extraction

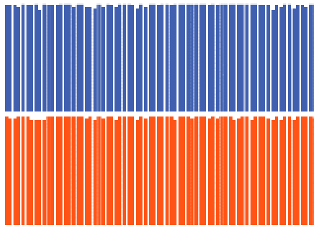
1/ Year
1/Quarter
1/Day
2/Day



Main Workflow



Bespoke Data Manipulation Tools



```
def check_daily_completeness(df, completeness_threshold)
```

- Removes all records contained within a day that does not meet the threshol

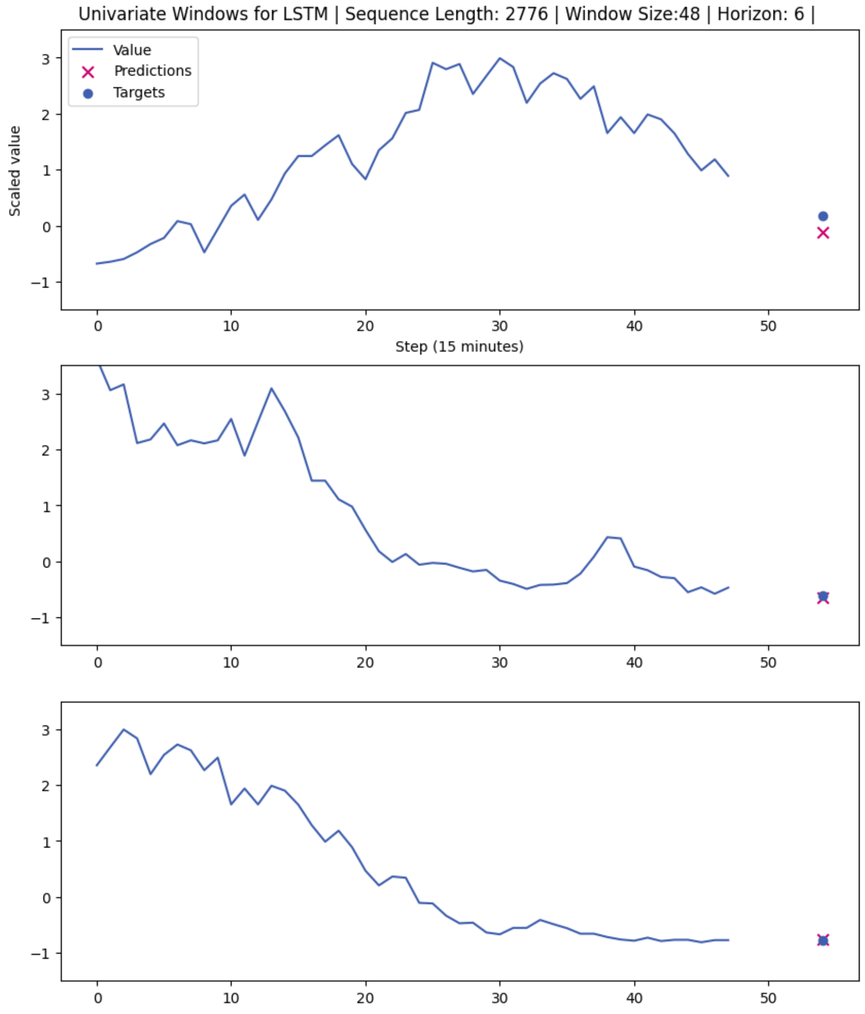
```
def find_longest_sequence(df)
```

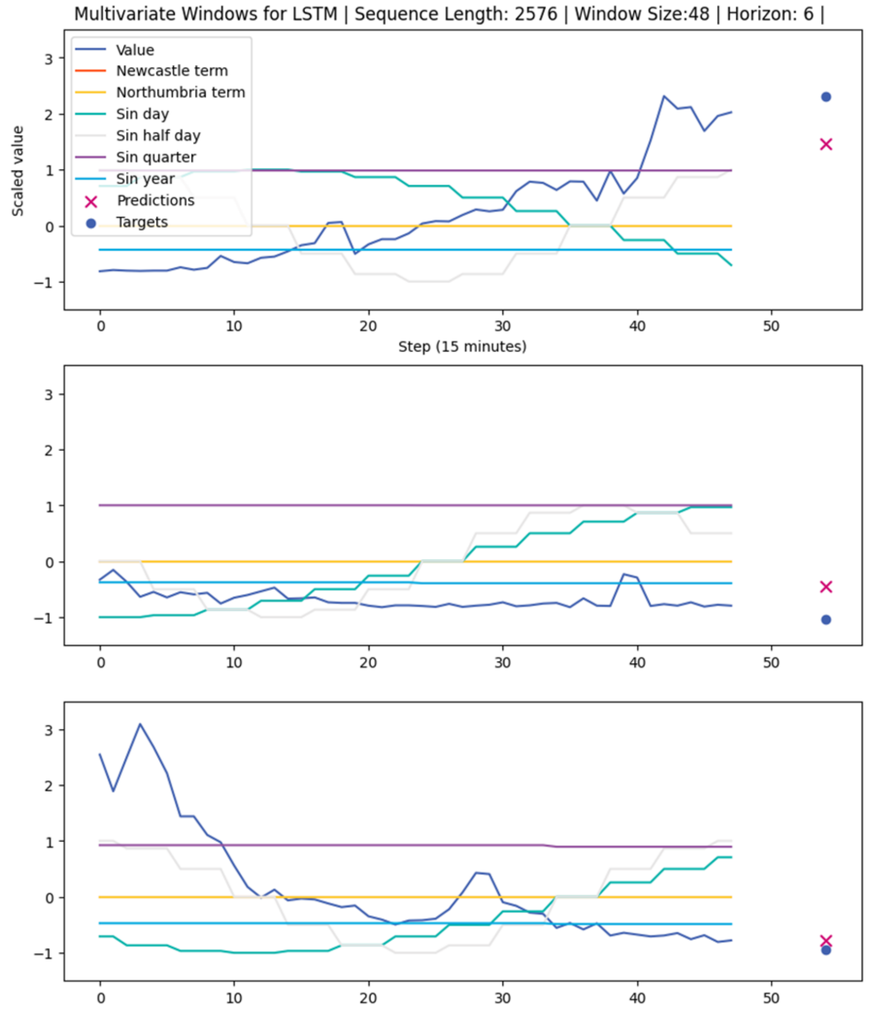
- Finds the longest continuous sequence of days that are above the threshold

```
def prepare_dataloaders(df, window_size, input_indices,
target_index, horizon)
```

- Generates sliding windows for training the models

Input Windows



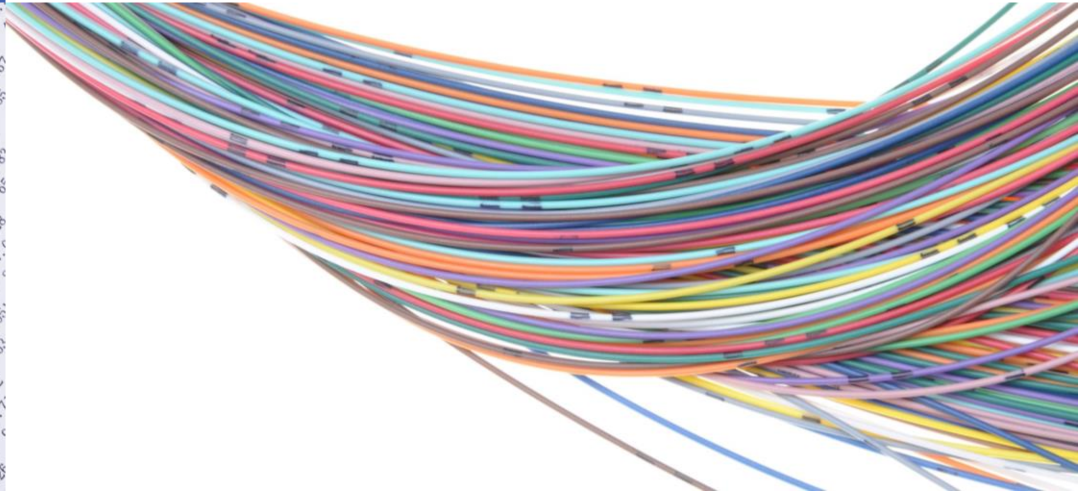


Hyperparameter Search Space

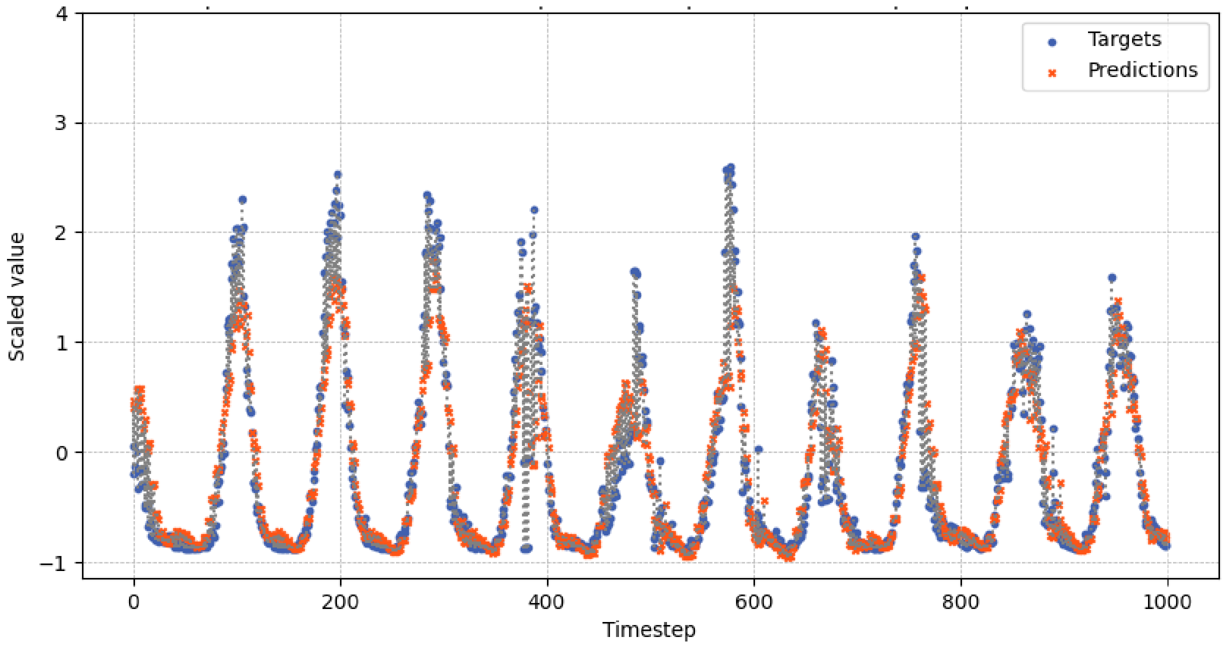
Hyperparameter	Values
COMPLETENESS_THRESHOLD	1.0, 0.98, 0.96, 0.94 or 1.0, 0.95, 0.90, 0.80
INPUT_INDICES	0 or [0,1,2,3,4,5,6,7,8,9,10]
TARGET_INDEX	0
WINDOW_SIZE	1 (linear) or 3, 6, 12, 24, 48, 96 (LSTM)
SEQUENCE_LENGTH	¼ , ½ , ¾ , 1 (proportion of the full sequence length used)
HORIZON	1, 3, 6, 12, 24
BATCH_SIZE	8 (for both linear and LSTM)
LEARNING_RATE	0.1 (with scheduler)
EPOCHS	20 (LSTM), 50 (Linear)
OPTIMISER	torch.optim.Adam
CRITERION	torch.nn.MSELoss()
ERROR_DEV	list(range(2, 10.01, step=0.1))

Results

- Can contextual anomalies be detected?
- To what extent does horizon decrease prediction accuracy?
- How does the completeness threshold affect model performance?



Anomaly Prediction



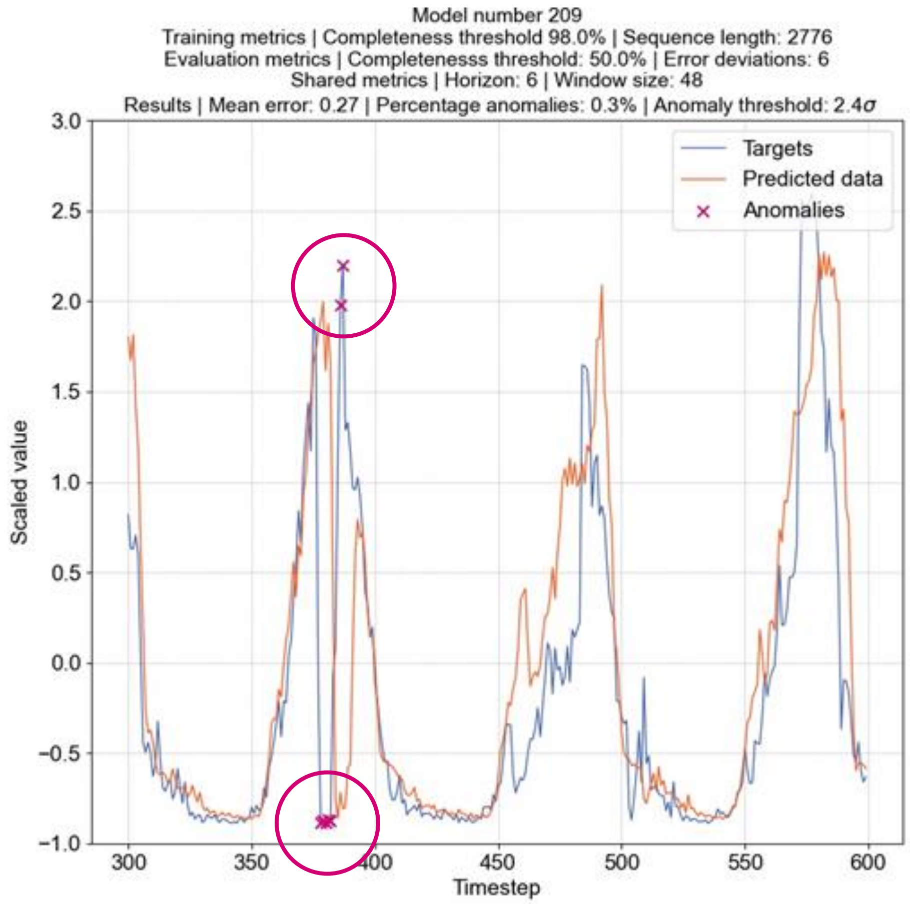
Sample Output

$$\overrightarrow{errors} = \overrightarrow{prediction} - \overrightarrow{targets}$$

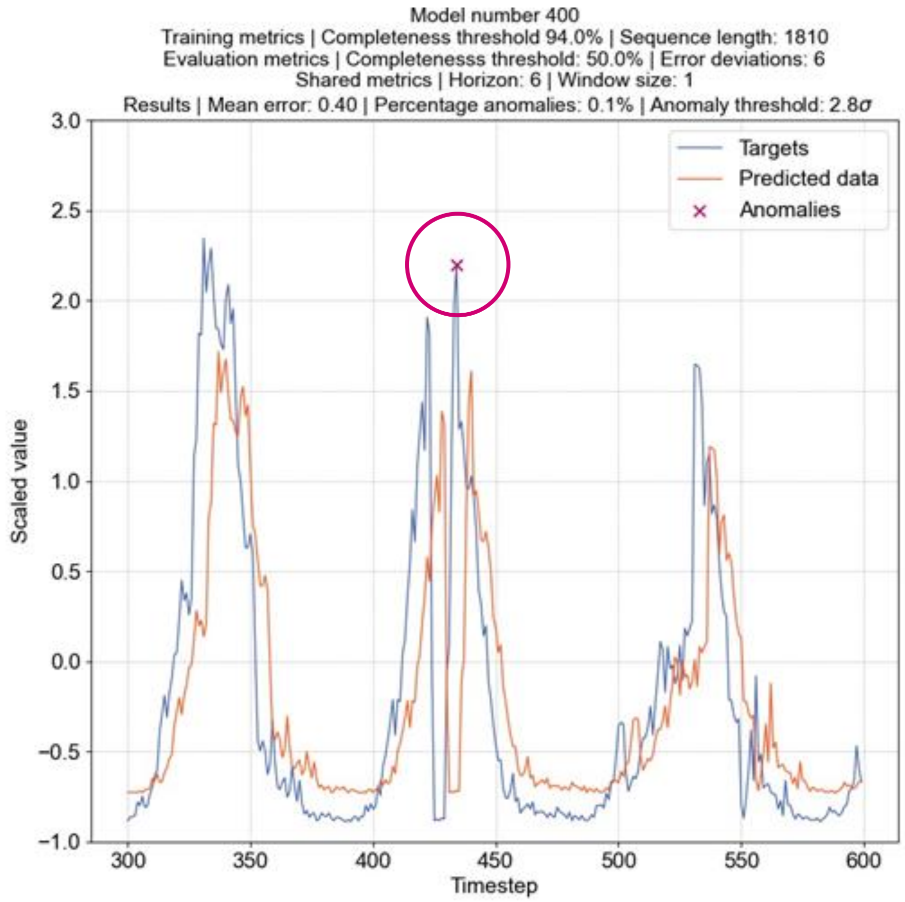
$$threshold = \overline{error} + \sigma_{error} * ERROR\_DEV$$

$$\overrightarrow{anomalies} = \overrightarrow{errors} > threshold$$

Anomaly Prediction (Univariate)

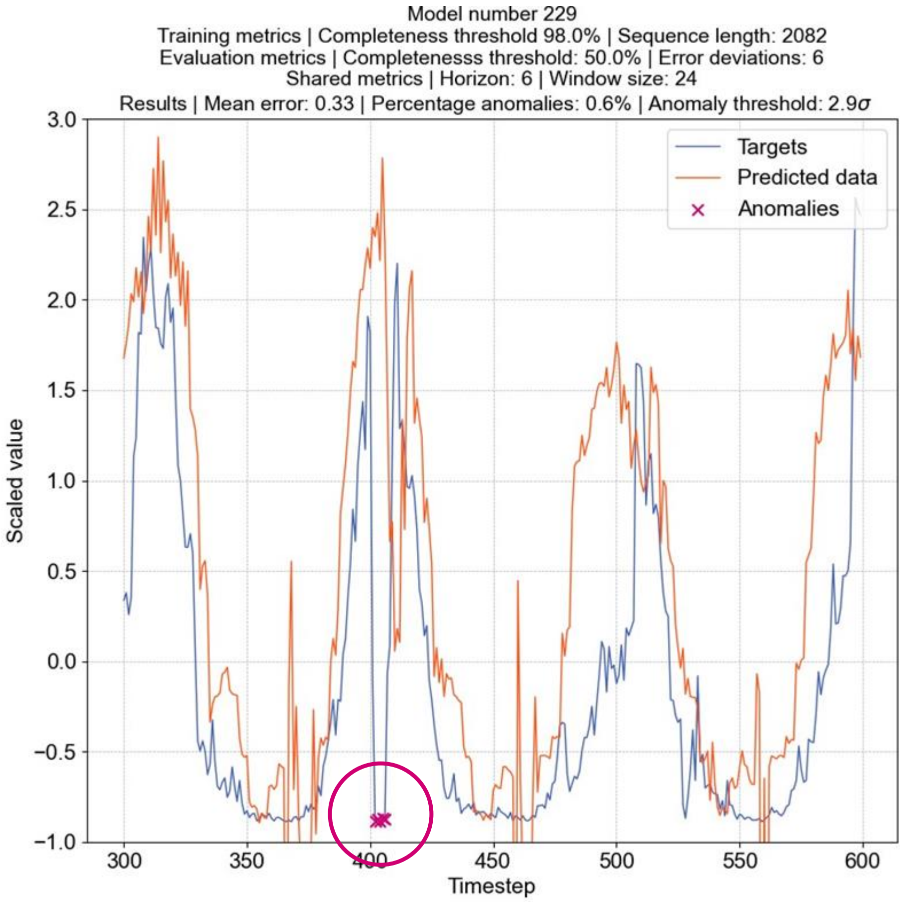


Univariate LSTM

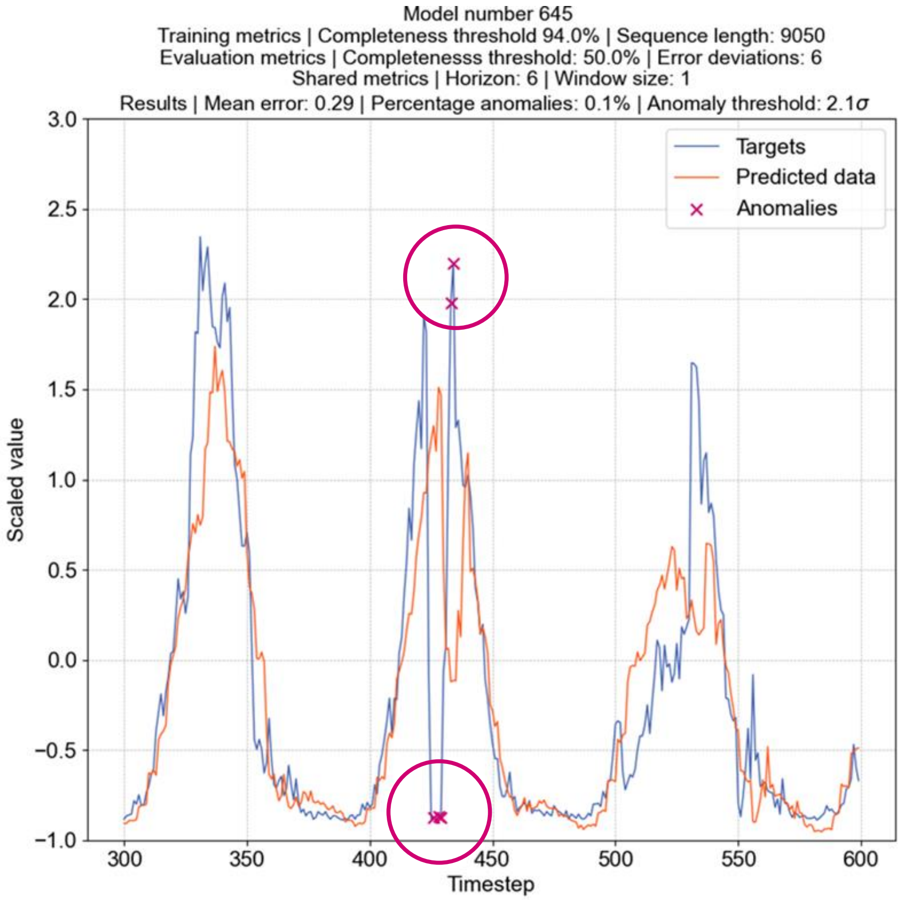


Univariate Linear

Anomaly Prediction (Multivariate)



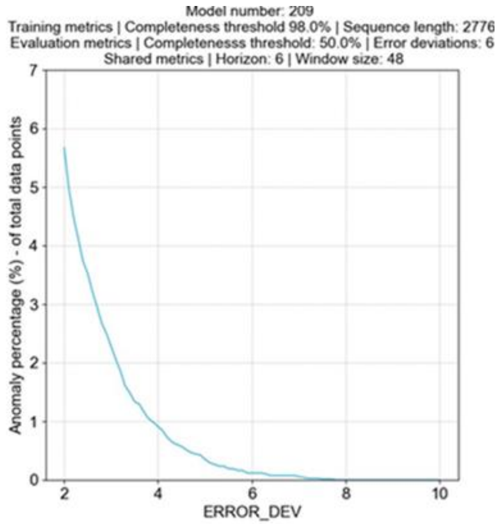
Multivariate LSTM



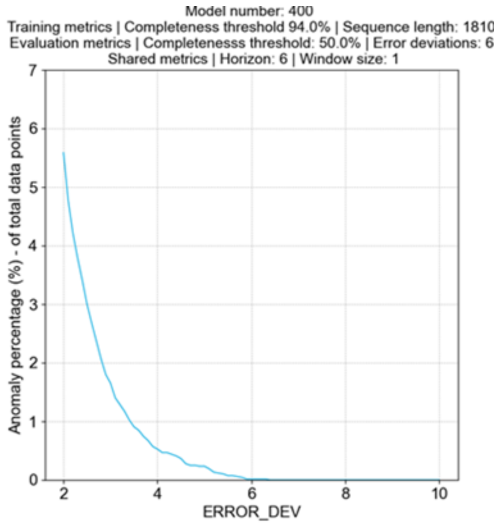
Multivariate Linear

Standard Deviations from Error

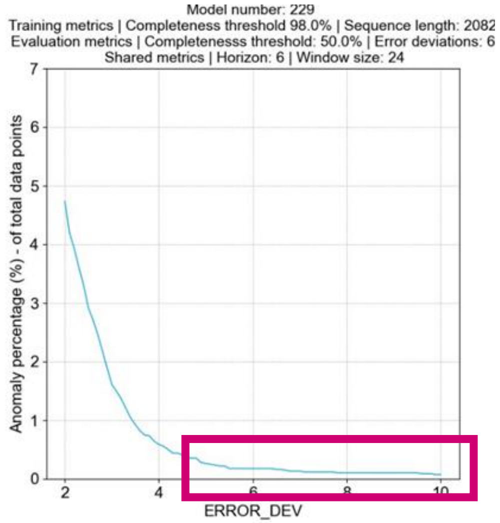
Univariate LSTM



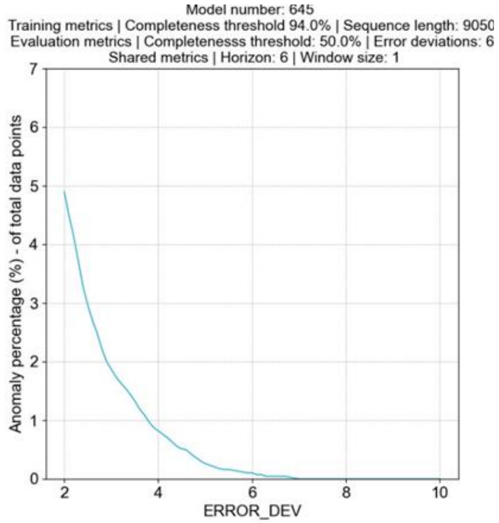
Univariate Linear



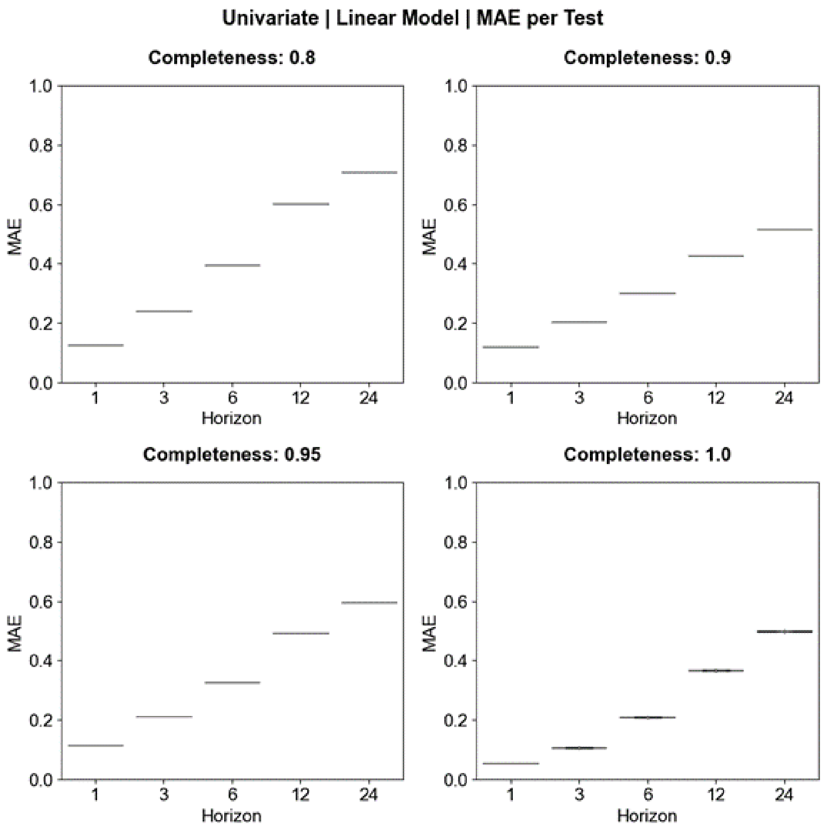
Multivariate LSTM



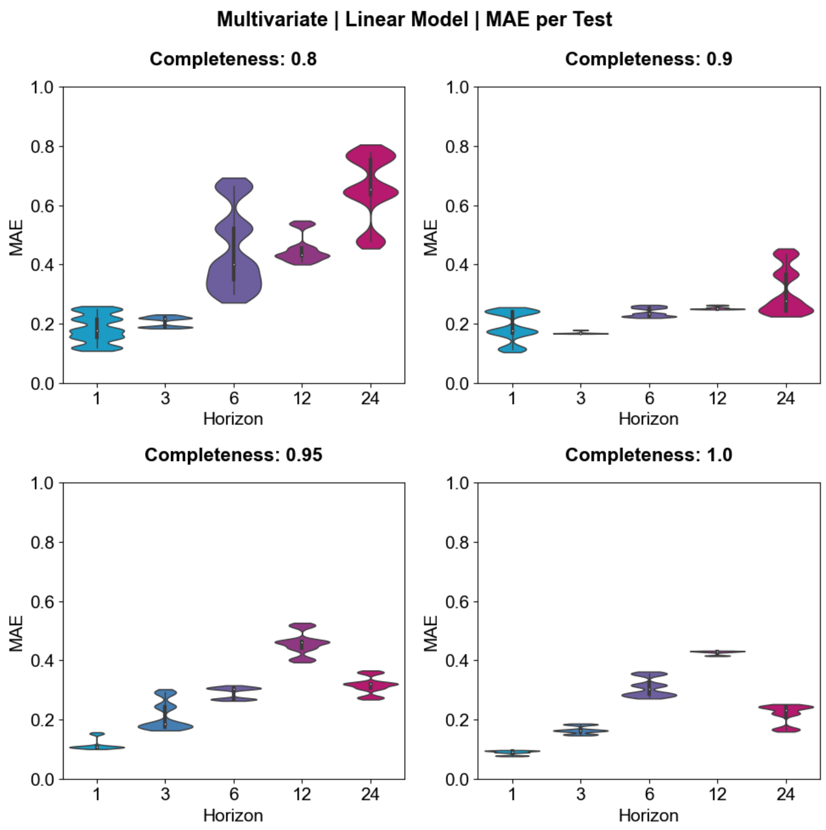
Multivariate Linear



Prediction Horizon (Linear)

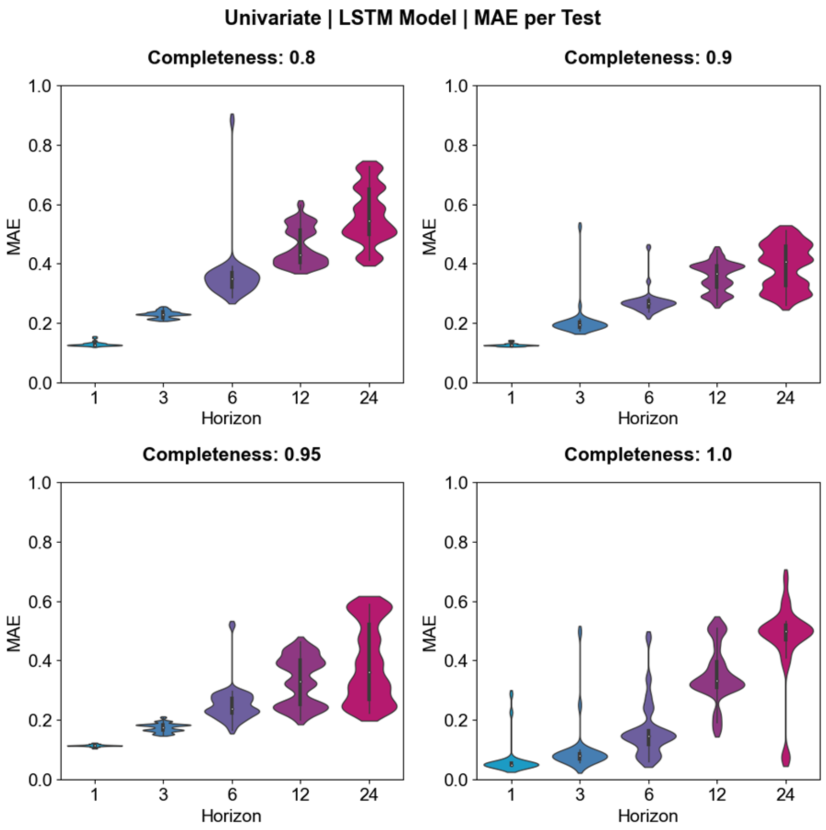


Univariate Linear

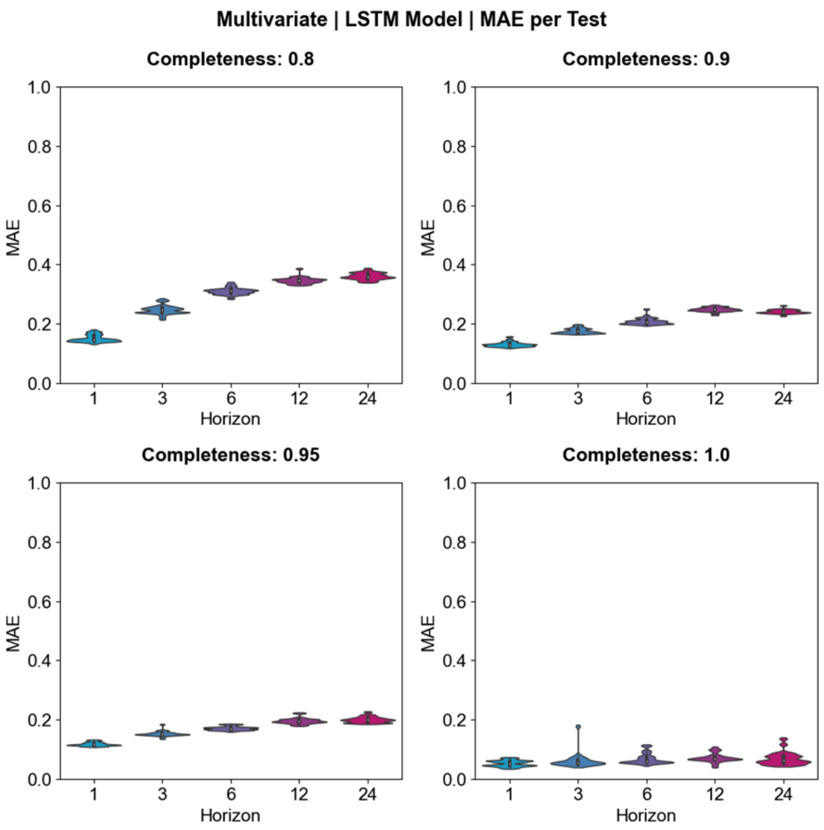


Multivariate Linear

Prediction Horizon (LSTM)

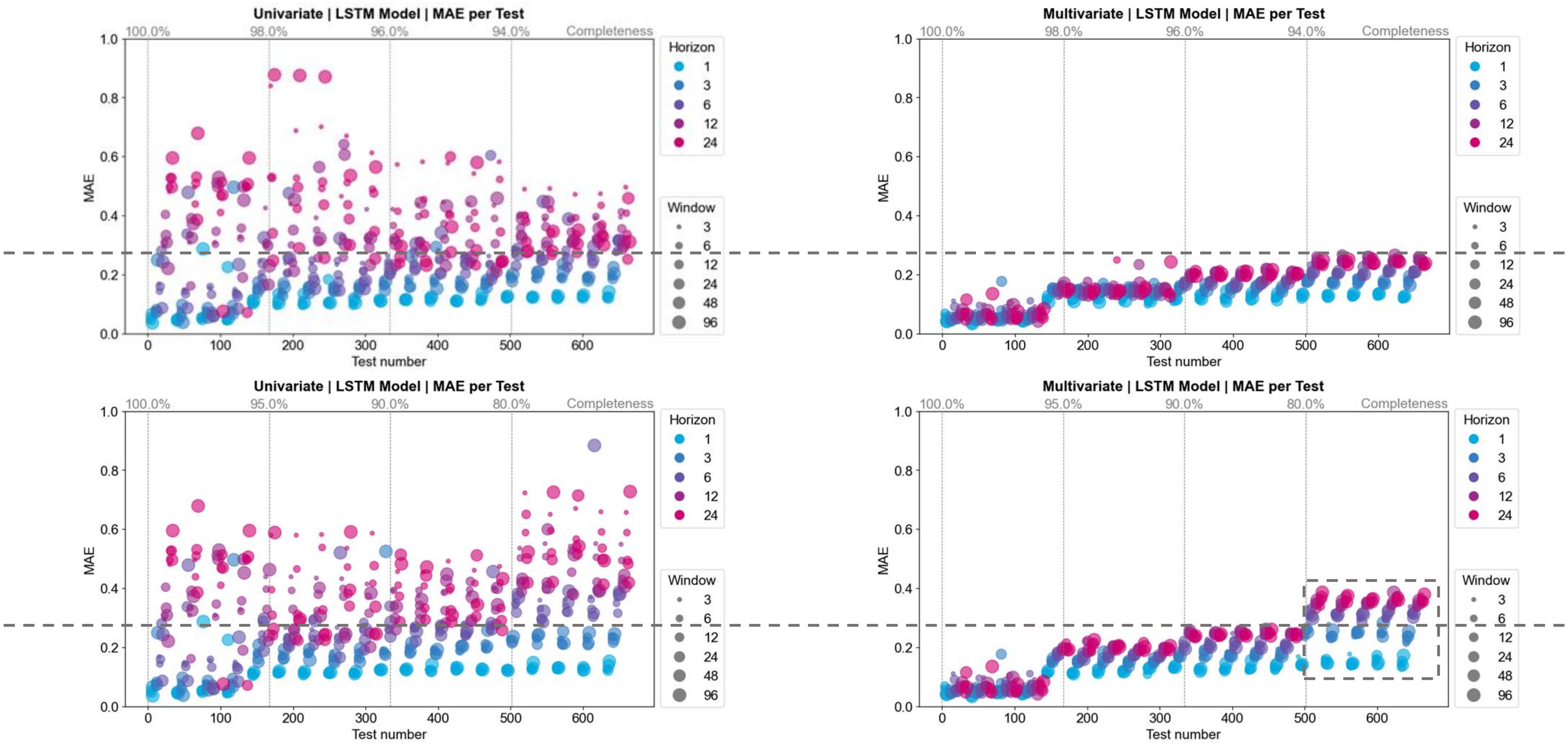


Univariate LSTM



Multivariate LSTM

Data Robustness (LSTM)



Summary of Findings

Finding 1

- Using multivariate inputs seems to increase model robustness against low quality data.

Finding 2

- The best performing models (MAE/RMSE) used 30 days of training data with 98% daily completeness.

Finding 3

Next Steps...

1

- MORE EPOCHS! (For the most promising models)

2

- Look at other anomaly detection methods
- Investigate other models – LSTM Autoencoders / Transformersw

3

- Expand to data from other sensors
- Simplify and deploy in the cloud for use in the urban observatory

Situational Awareness Infrastructure

Data Collection

Data Storage

Data Preprocessing

Data Analytics

Data Driven Decisions

Questions?

EDA Report

Thesis

GitHub Repo