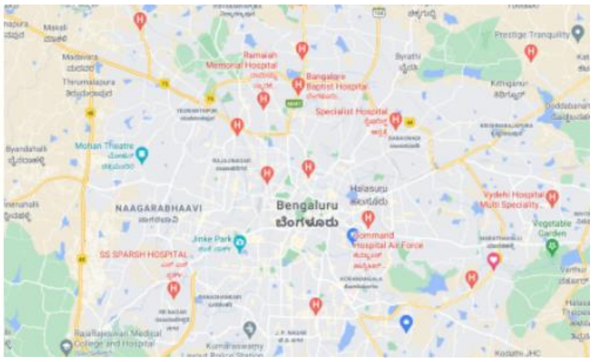
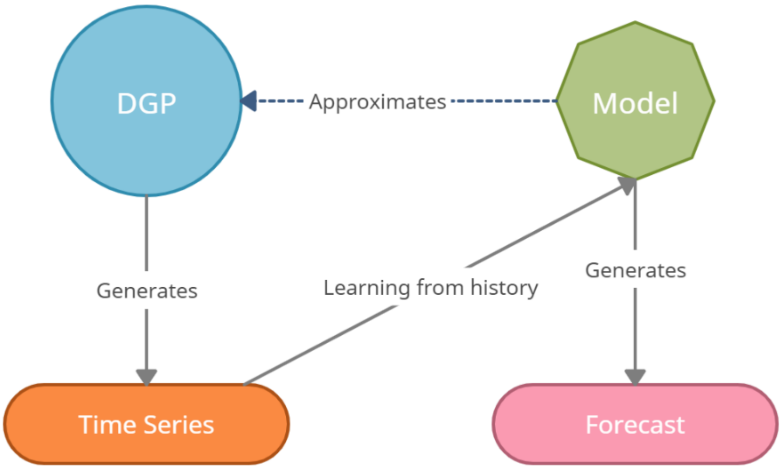
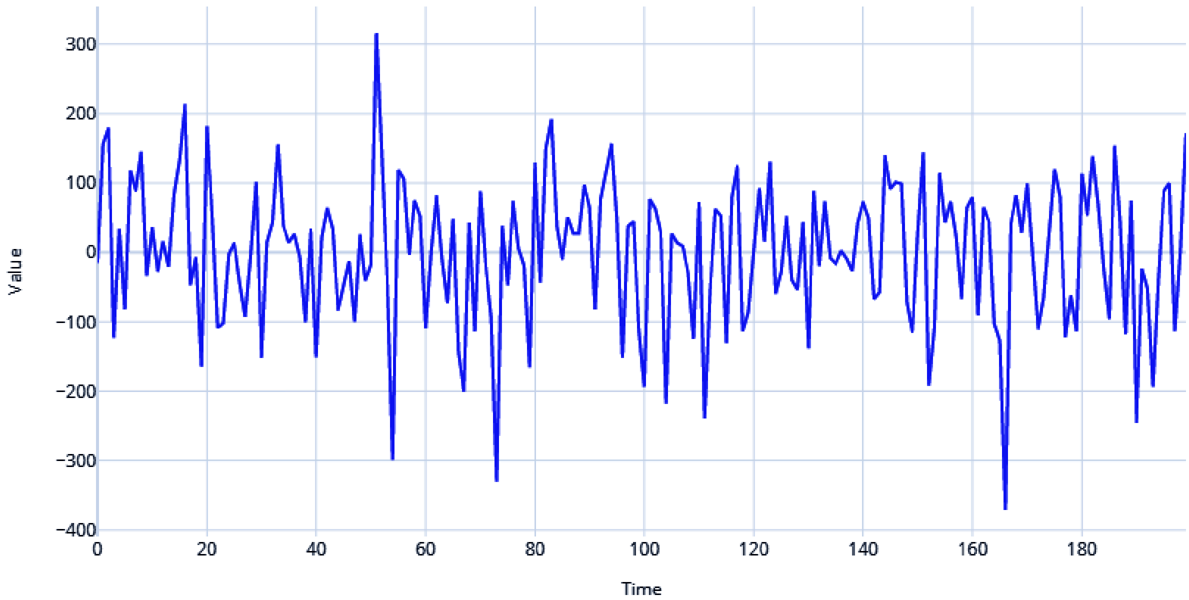
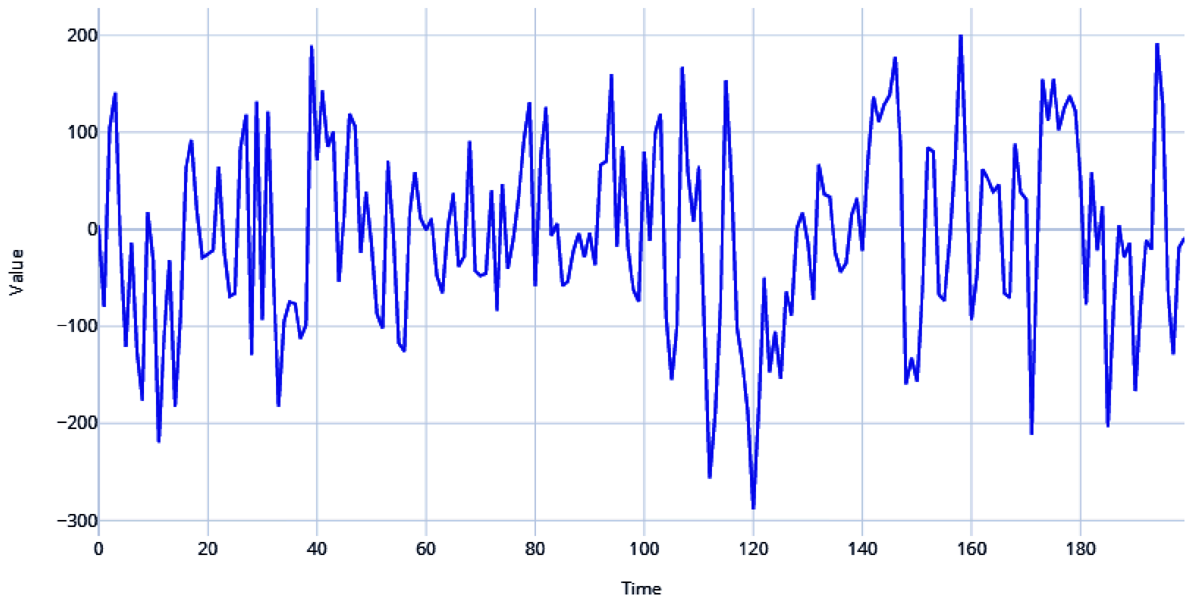
Chapter 1: Introducing Time Series

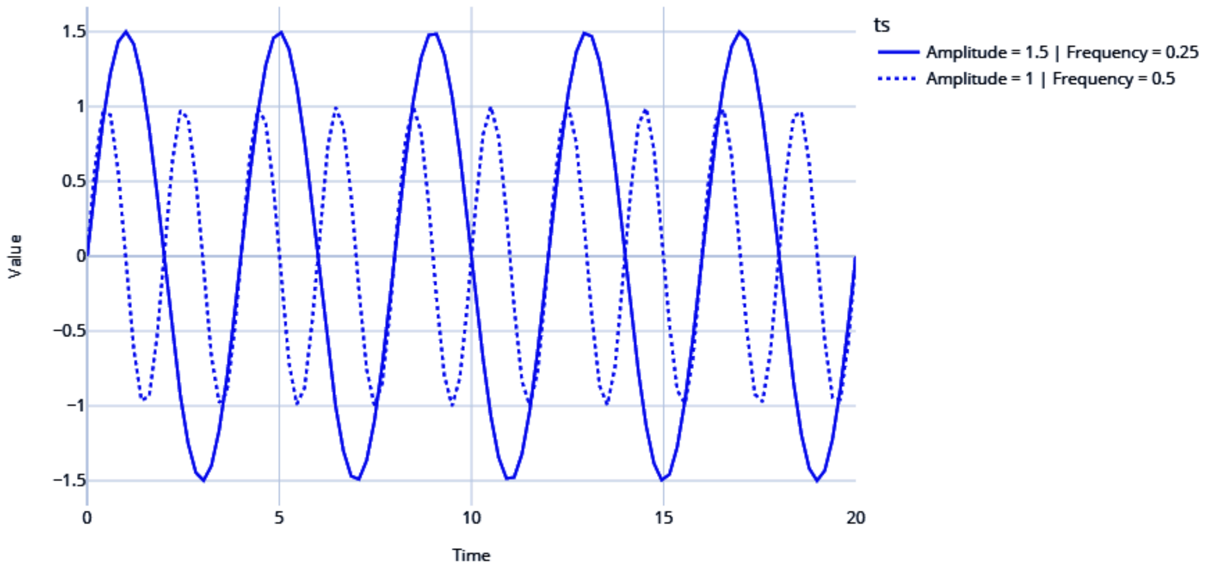




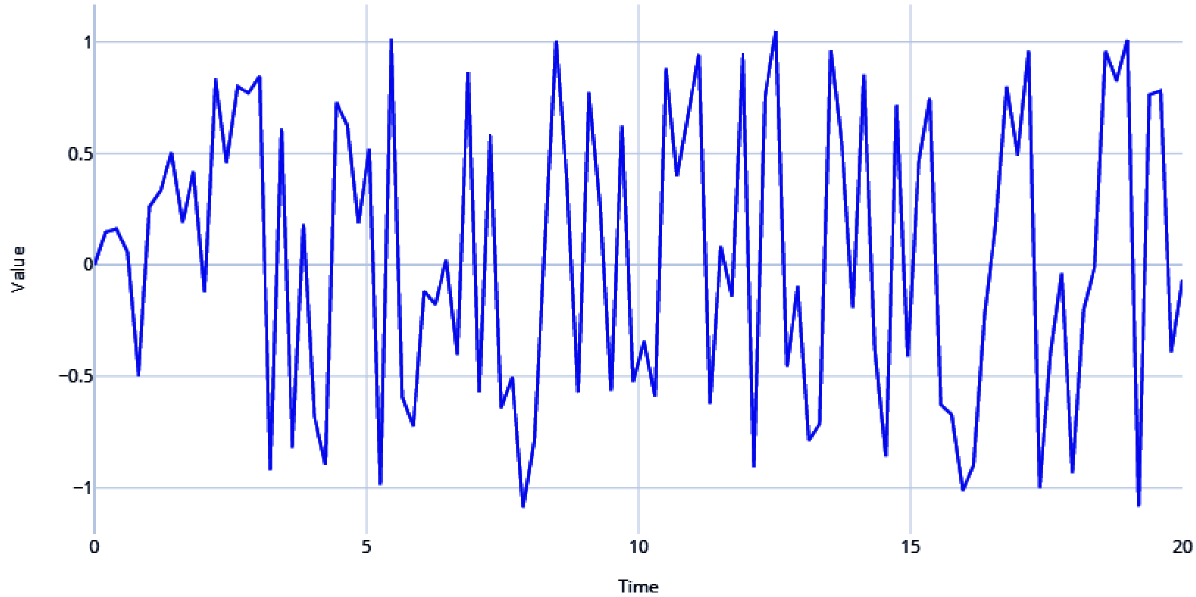




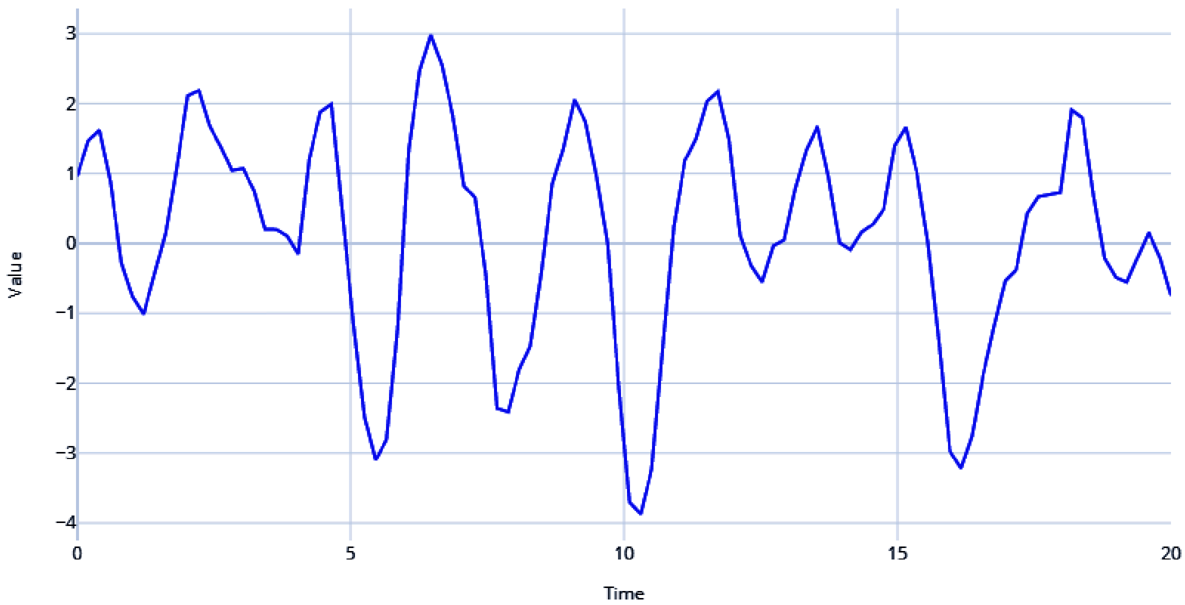
200
100
0
−100
−200
−300
Value
0
20
40
60
80
100
120
140
160
180
Time

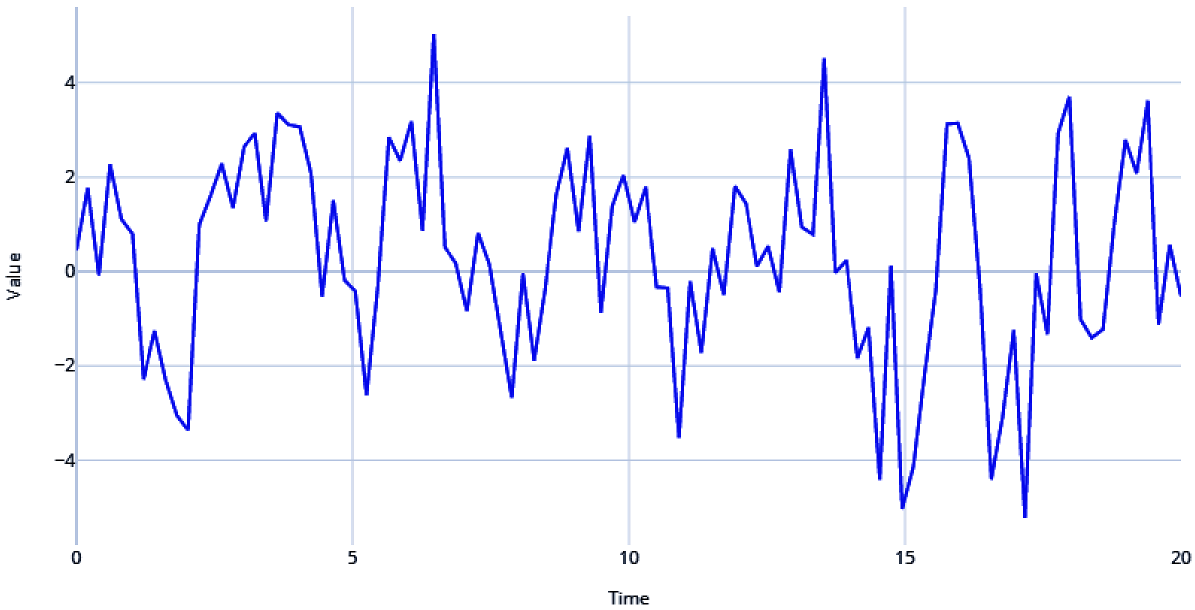


ts
Amplitude = 1.5 | Frequency = 0.25
Amplitude = 1 | Frequency = 0.5
1.5
1
0.5
0
−0.5
−1
−1.5
Value
0
5
10
15
20
Time



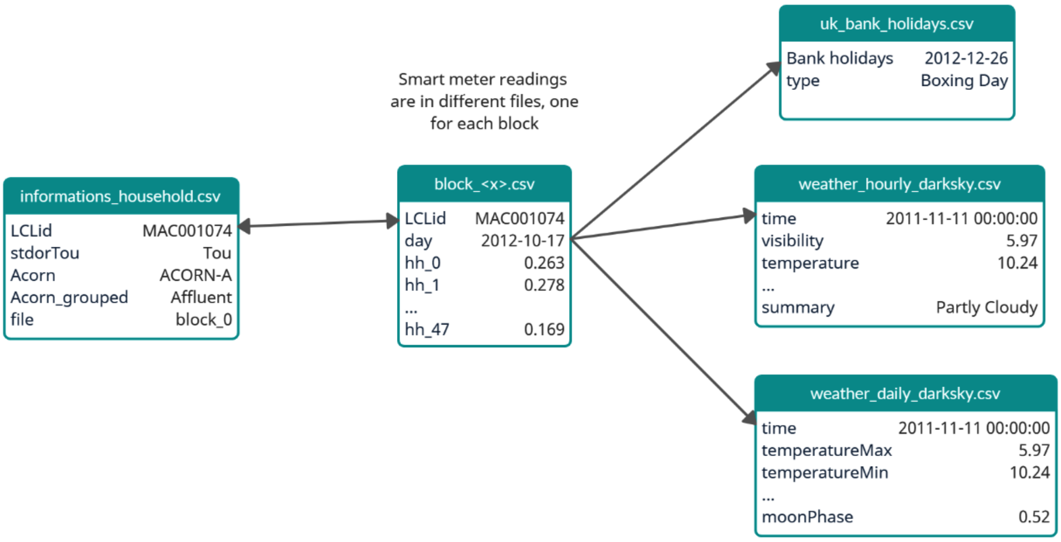
1
0.5
0
−0.5
−1
Value
0
5
10
15
20
Time


3
2
1
0
−1
−2
−3
−4
Value
0
5
10
15
20
Time

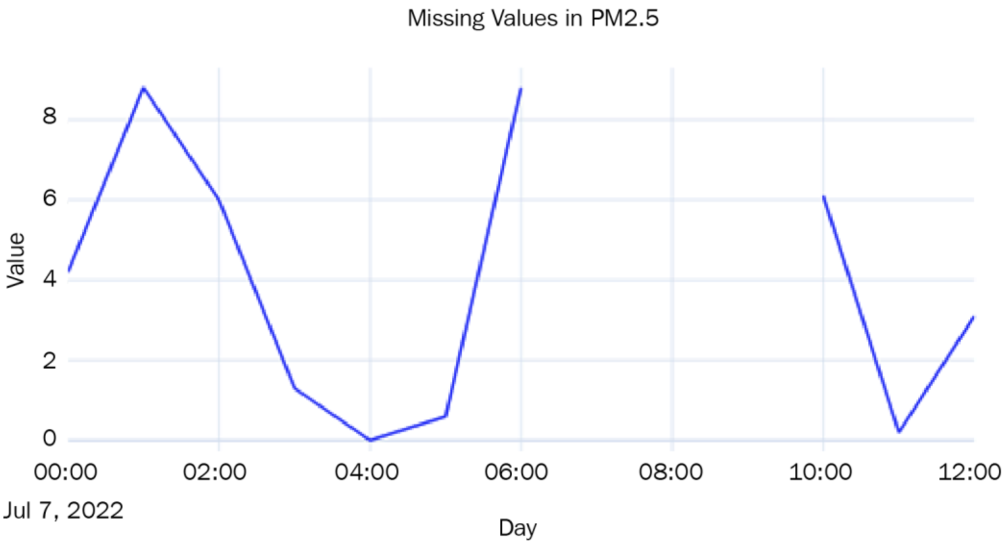

4
2
0
−2
−4
Value
0
5
10
15
20
Time

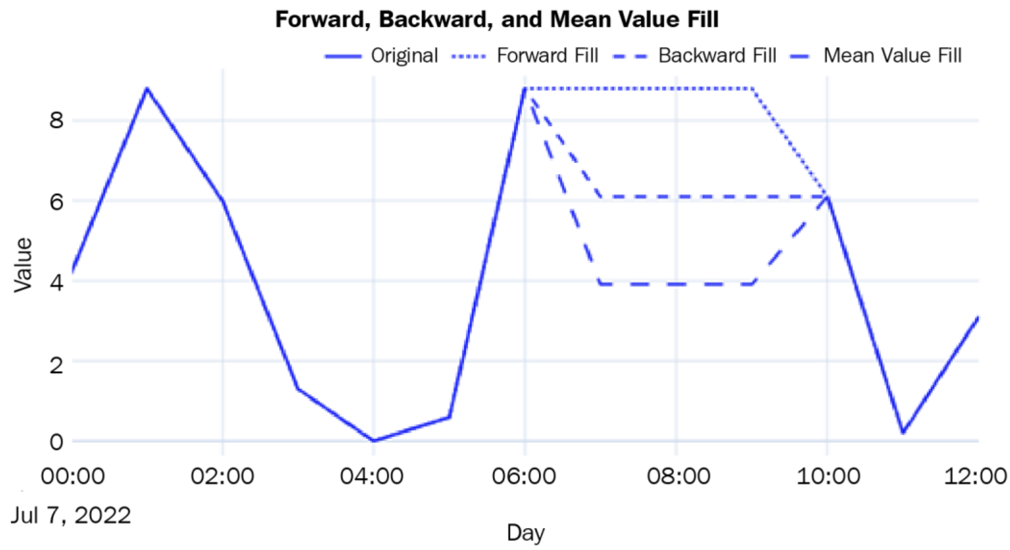
$t = 1, 2, 3, \ldots, L$	Index denoting time step of interest. L is the total length of the time series
$n = 1, 2, 3, \ldots, N$	In cases where we are talking about more than one time series: n is the index denoting the time series and N is the total number of timeseries in the set of timeseries we are considering.
y_t	Value of the timeseries at timestep t
$y_{n,t}$	Value of the n^{th} time series at timestep t
Y_L or $Y = \{y_1, y_2, y_3, \ldots y_L\}$	A complete time series of L timesteps (if mentioned)
$Y_{N,L} = \begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,L} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,L} \\ \cdots & \cdots & \cdots & \cdots \\ y_{N,1} & y_{N,2} & \cdots & y_{N,L} \end{bmatrix}$	A set of N timeseries with L timesteps.
f_t	Forecast at timestep t
$f_{[t,t+H]}$	Multi-step Forecast from timesteps t to $t + H$, where H is the forecast horizon
$F_{N,H} = \begin{bmatrix} f_{1,t} & f_{1,t+1} & \cdots & f_{1,t+H} \\ f_{2,t} & f_{2,t+1} & \cdots & f_{2,t+H} \\ \cdots & \cdots & \cdots & \cdots \\ f_{N,t} & f_{N,t+1} & \cdots & f_{N,t+H} \end{bmatrix}$	Multi-step Forecast for a set of time series (N) and timesteps t to $t + H$, where H is the forecast horizon

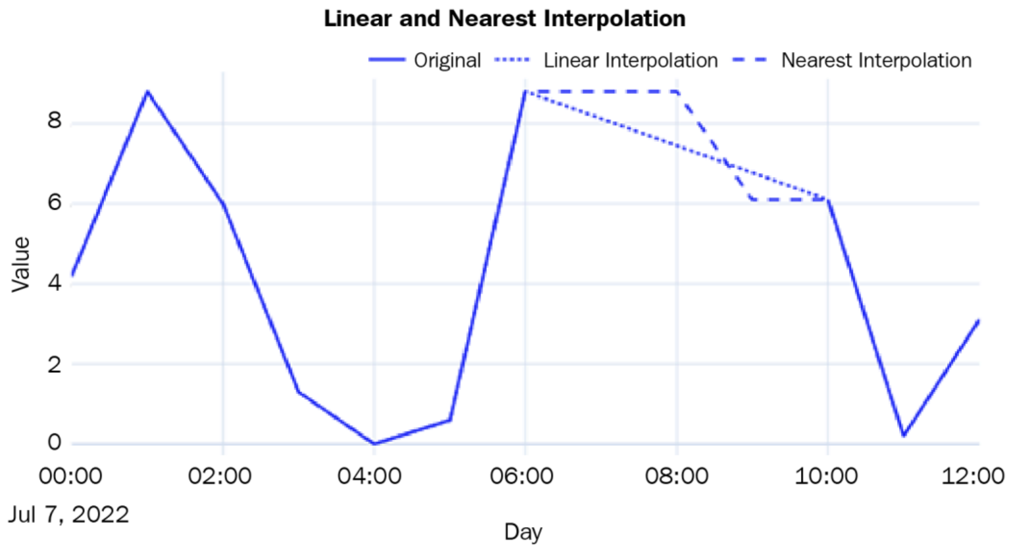
Chapter 2: Acquiring and Processing Time Series Data

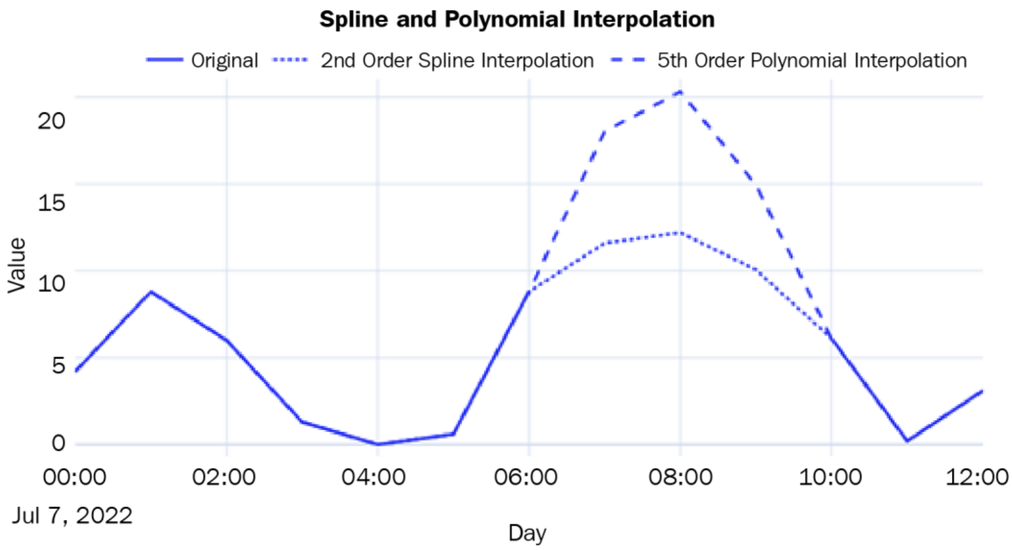


	date	ISE	ISE.1	SP	DAX	FTSE	NIKKEI	BOVESPA	EU	EM
0	2009-01-05	0.035754	0.038376	-0.004679	0.002193	0.003894	0.000000	0.031190	0.012698	0.028524
1	2009-01-06	0.025426	0.031813	0.007787	0.008455	0.012866	0.004162	0.018920	0.011341	0.008773
2	2009-01-07	-0.028862	-0.026353	-0.030469	-0.017833	-0.028735	0.017293	-0.035899	-0.017073	-0.020015
3	2009-01-08	-0.062208	-0.084716	0.003391	-0.011726	-0.000466	-0.040061	0.028283	-0.005561	-0.019424
4	2009-01-09	0.009860	0.009658	-0.021533	-0.019873	-0.012710	-0.004474	-0.009764	-0.010989	-0.007802



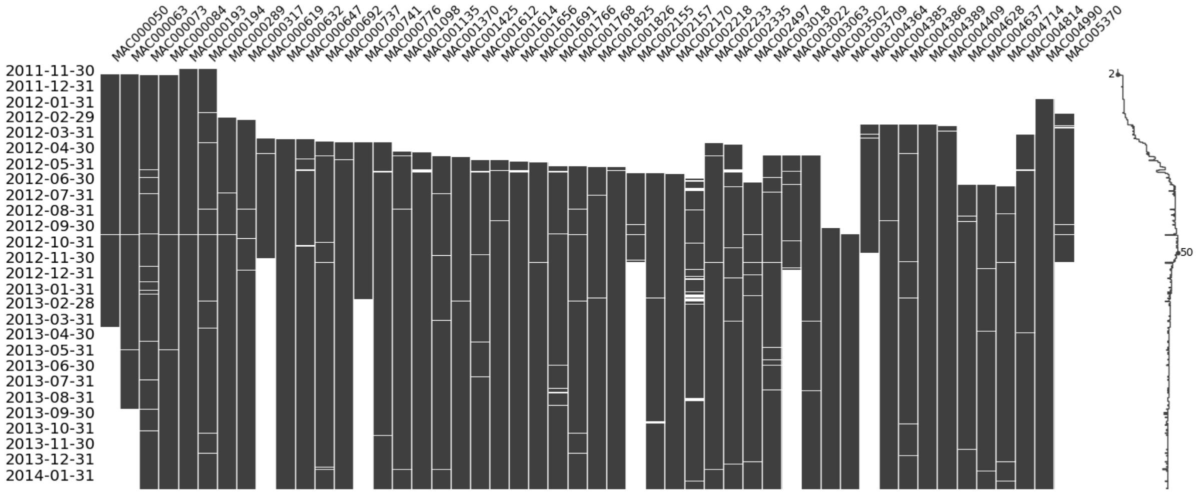

Forward, Backward, and Mean Value Fill
Original
Forward Fill
Backward Fill
Mean Value Fill
Value
0
2
4
6
8
00:00
Jul 7, 2022
02:00
04:00
06:00
08:00
10:00
12:00
Day

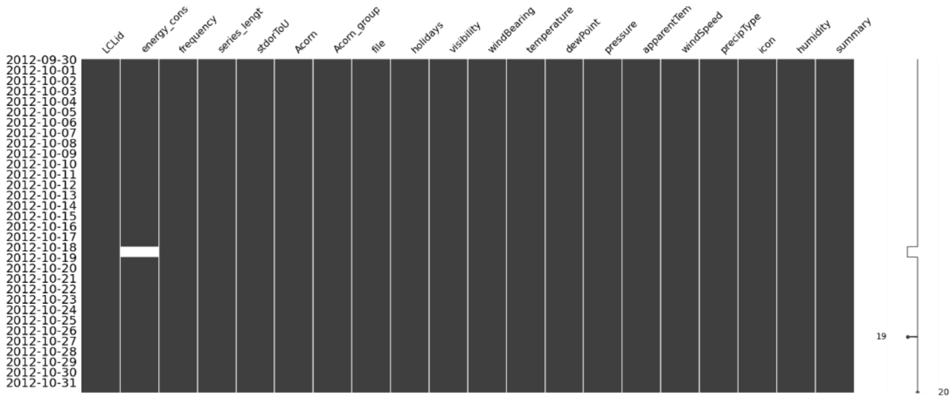

Linear and Nearest Interpolation
Original
Linear Interpolation
Nearest Interpolation
Value
0
2
4
6
8
00:00
Jul 7, 2022
02:00
04:00
06:00
08:00
10:00
12:00
Day

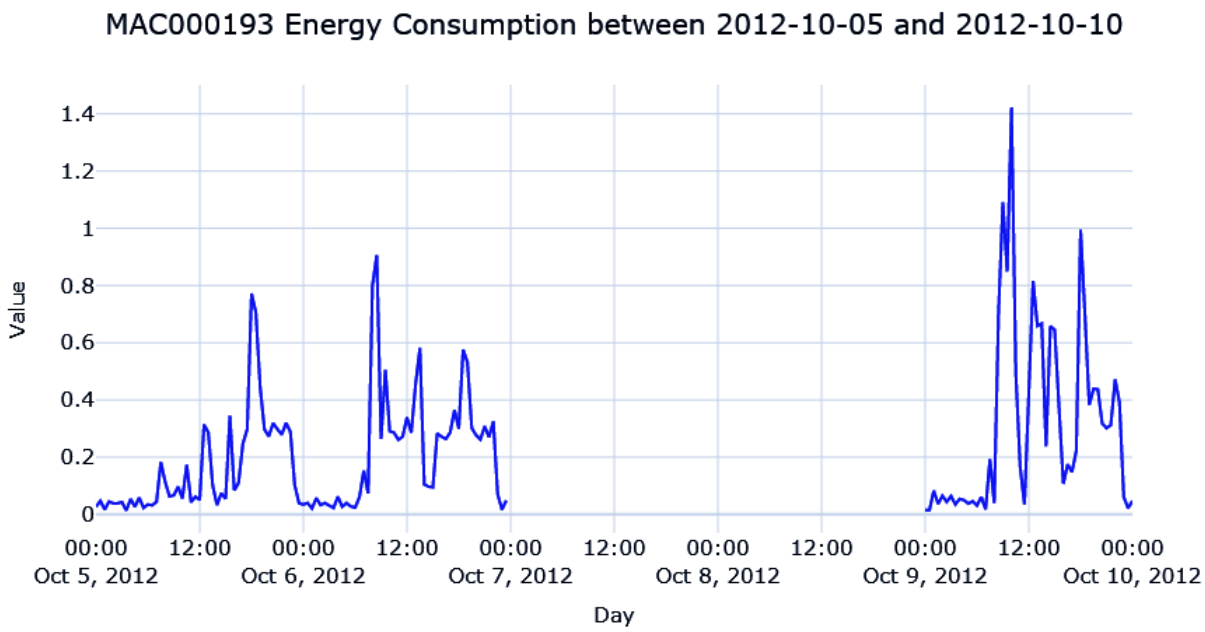

Spline and Polynomial Interpolation
Original
2nd Order Spline Interpolation
5th Order Polynomial Interpolation
Value
0
5
10
15
20
00:00
Jul 7, 2022
02:00
04:00
06:00
08:00
10:00
12:00
Day

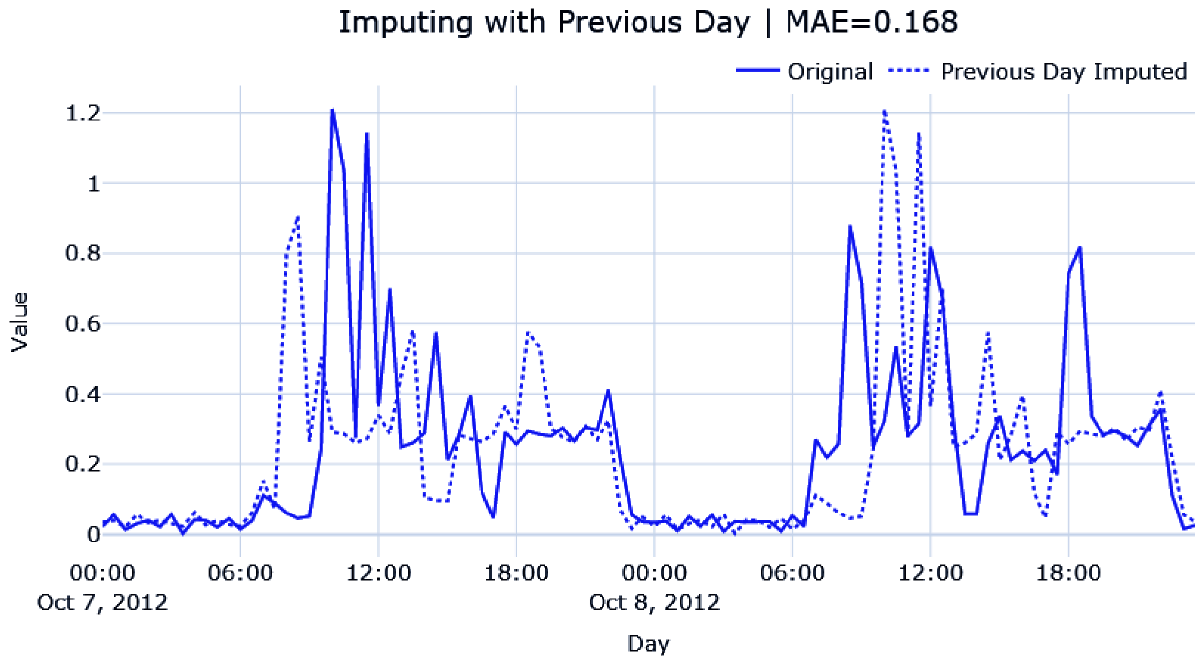
LCLid	start_timestamp	frequency	energy_consumption	series_length
MAC000002	2012-10-13	30min	[0.263, 0.2689999999999999, 0.275, 0.256, 0.21...	24144
MAC000246	2011-12-04	30min	[0.175, 0.098, 0.144, 0.065, 0.071, 0.037, 0.0...	39216
MAC000450	2012-03-23	30min	[0.337, 1.426, 0.996, 0.971, 0.994, 0.952, 0.8...	33936
MAC001074	2012-05-09	30min	[0.18, 0.086, 0.106, 0.173, 0.146, 0.223, 0.21...	31680
MAC003223	2012-09-18	30min	[0.076, 0.079, 0.123, 0.109, 0.051, 0.069, 0.0...	25344

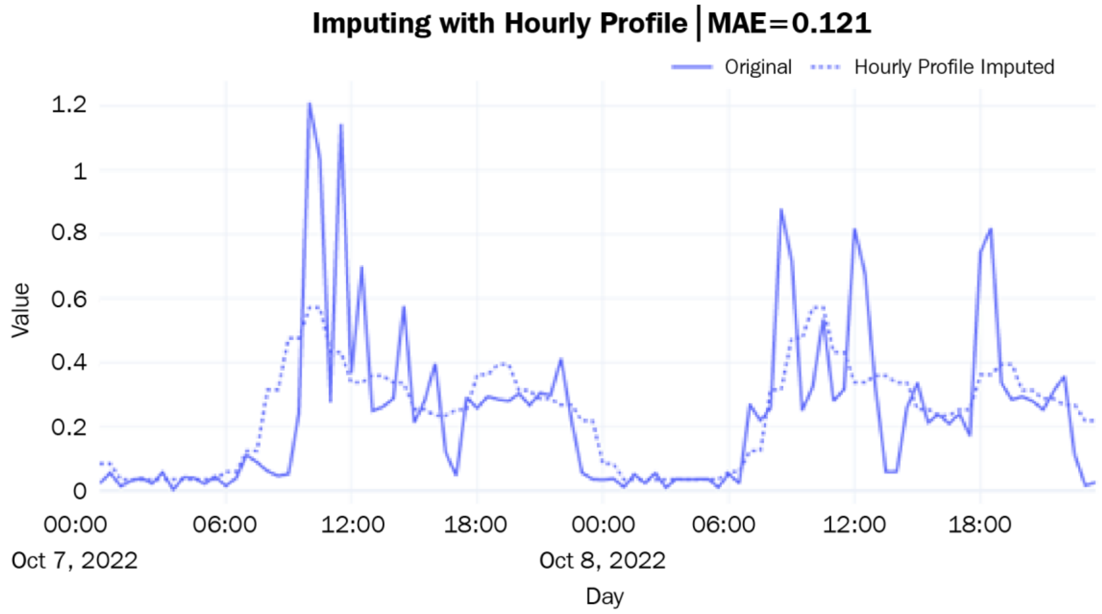
LCLid	energy_consumption	series_length	timestamp	frequency
MAC000002	0.263	24144	2012-10-13 00:00:00	30min
MAC000002	0.269	24144	2012-10-13 00:30:00	30min
MAC000002	0.275	24144	2012-10-13 01:00:00	30min
MAC000002	0.256	24144	2012-10-13 01:30:00	30min
MAC000002	0.211	24144	2012-10-13 02:00:00	30min

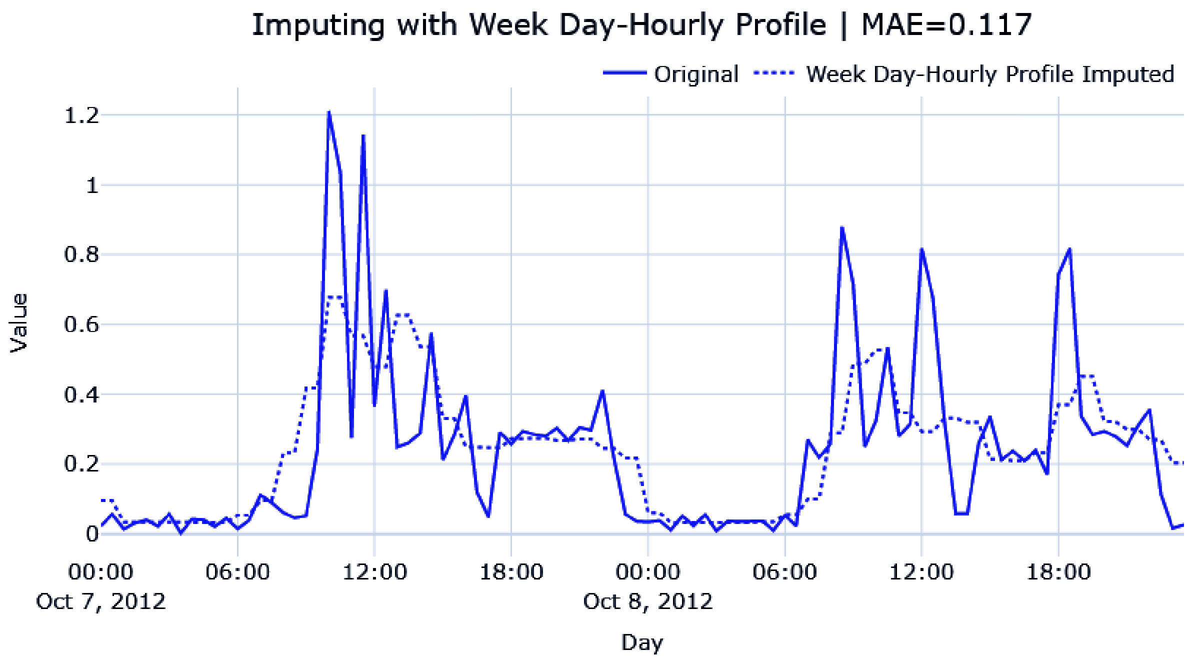


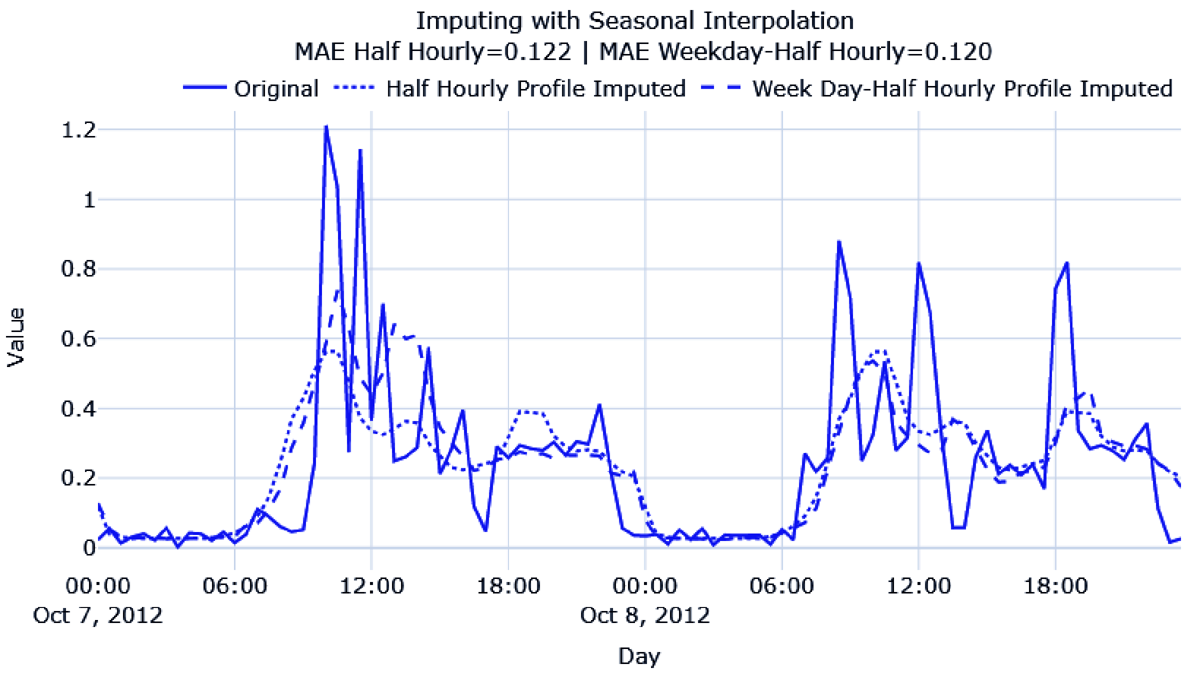



MAC000193 Energy Consumption between 2012-10-05 and 2012-10-10
Value
1.4
1.2
1
0.8
0.6
0.4
0.2
0
00:00
Oct 5, 2012
12:00
00:00
Oct 6, 2012
12:00
00:00
Oct 7, 2012
12:00
00:00
Oct 8, 2012
12:00
00:00
Oct 9, 2012
12:00
00:00
Oct 10, 2012
Day

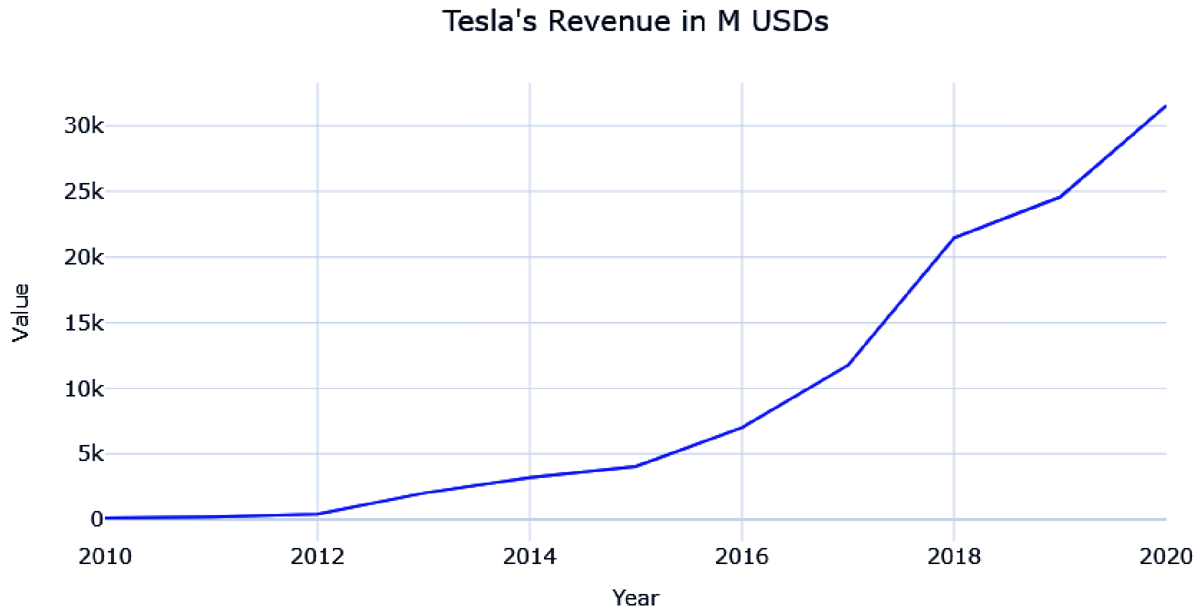

Imputing with Previous Day | MAE=0.168
Original
Previous Day Imputed
Value
1.2
1
0.8
0.6
0.4
0.2
0
00:00
Oct 7, 2012
06:00
12:00
18:00
00:00
Oct 8, 2012
06:00
12:00
18:00
Day

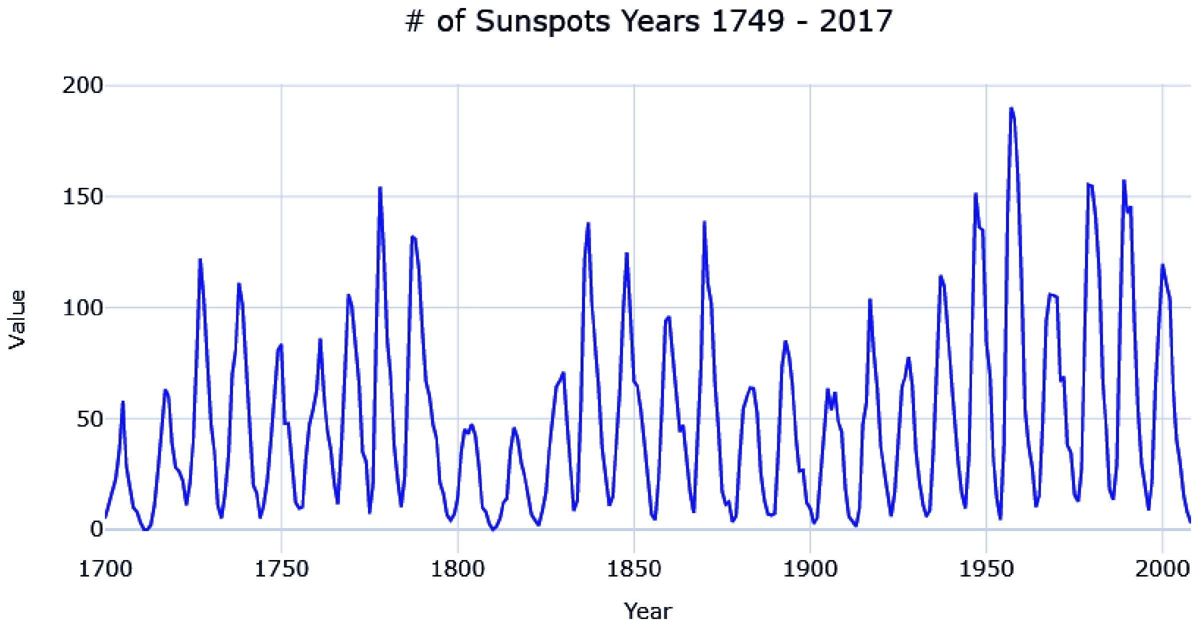

Imputing with Hourly Profile | MAE=0.121
Original
Hourly Profile Imputed
Value
1.2
1
0.8
0.6
0.4
0.2
0
00:00
Oct 7, 2022
06:00
12:00
18:00
00:00
Oct 8, 2022
06:00
12:00
18:00
Day

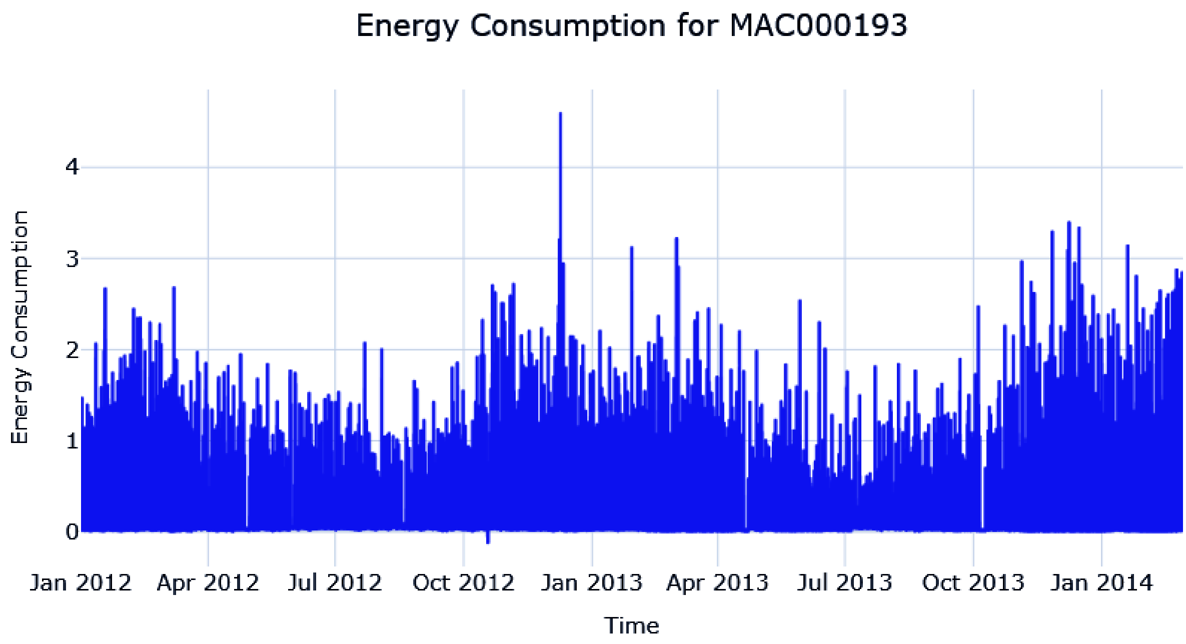

Imputing with Week Day-Hourly Profile | MAE=0.117
Original
Week Day-Hourly Profile Imputed
Value
1.2
1
0.8
0.6
0.4
0.2
0
00:00
Oct 7, 2012
06:00
12:00
18:00
00:00
Oct 8, 2012
06:00
12:00
18:00
Day

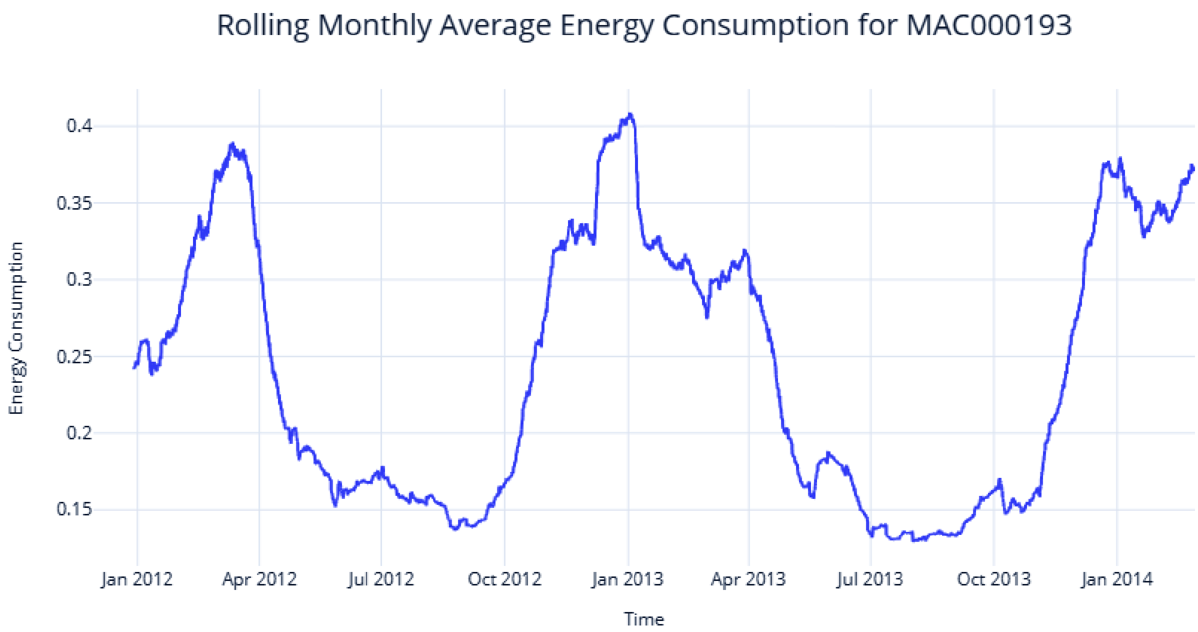

Imputing with Seasonal Interpolation
MAE Half Hourly=0.122 | MAE Weekday-Half Hourly=0.120
Original
Half Hourly Profile Imputed
Week Day-Half Hourly Profile Imputed
Value
1.2
1
0.8
0.6
0.4
0.2
0
00:00
Oct 7, 2012
06:00
12:00
18:00
00:00
Oct 8, 2012
06:00
12:00
18:00
Day

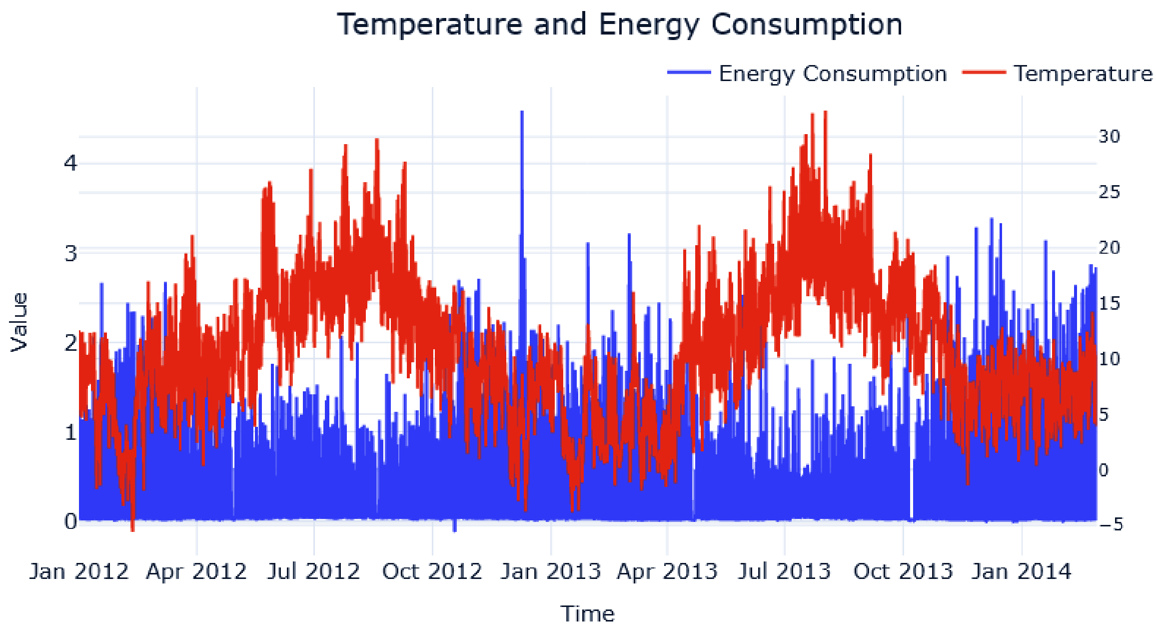
Chapter 3: Analyzing and Visualizing Time Series Data

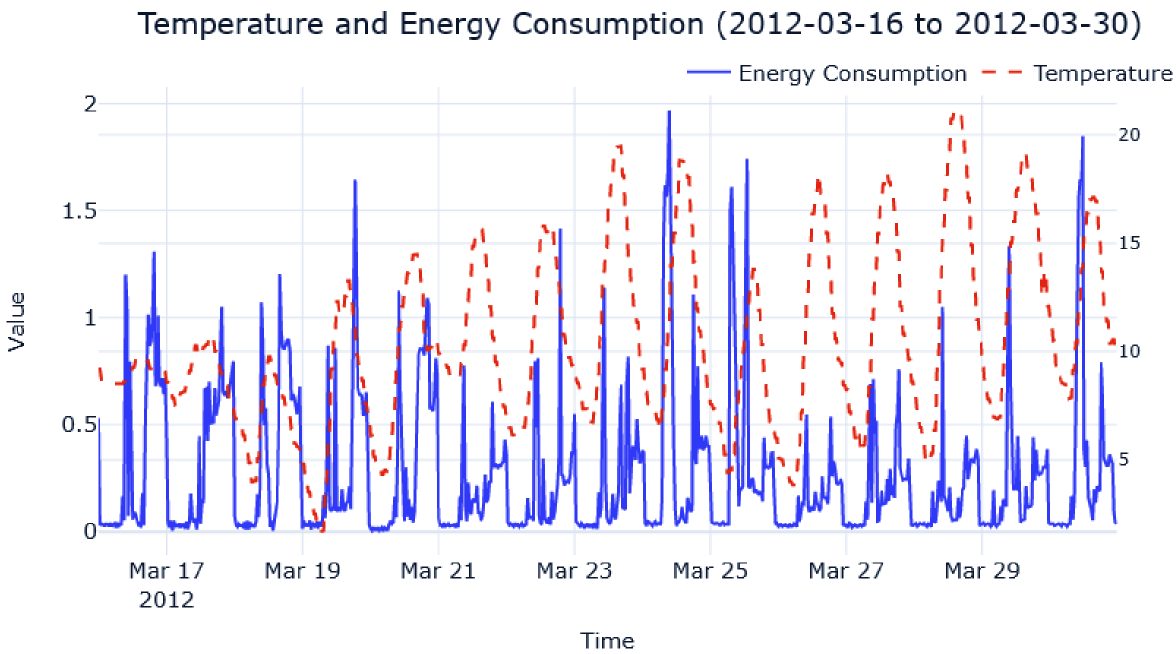


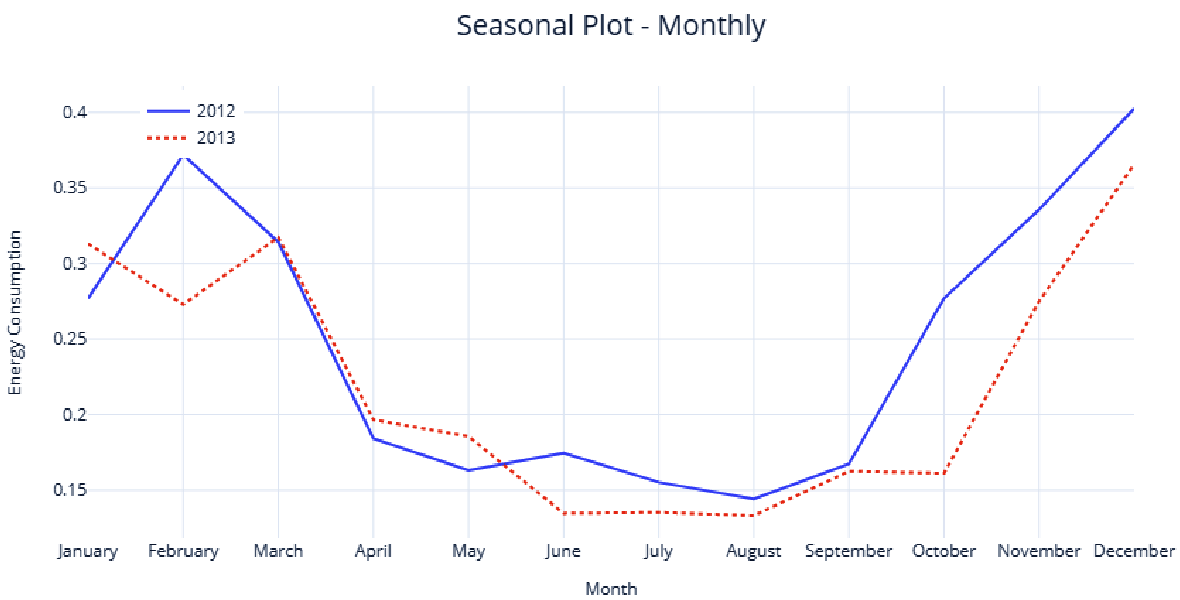


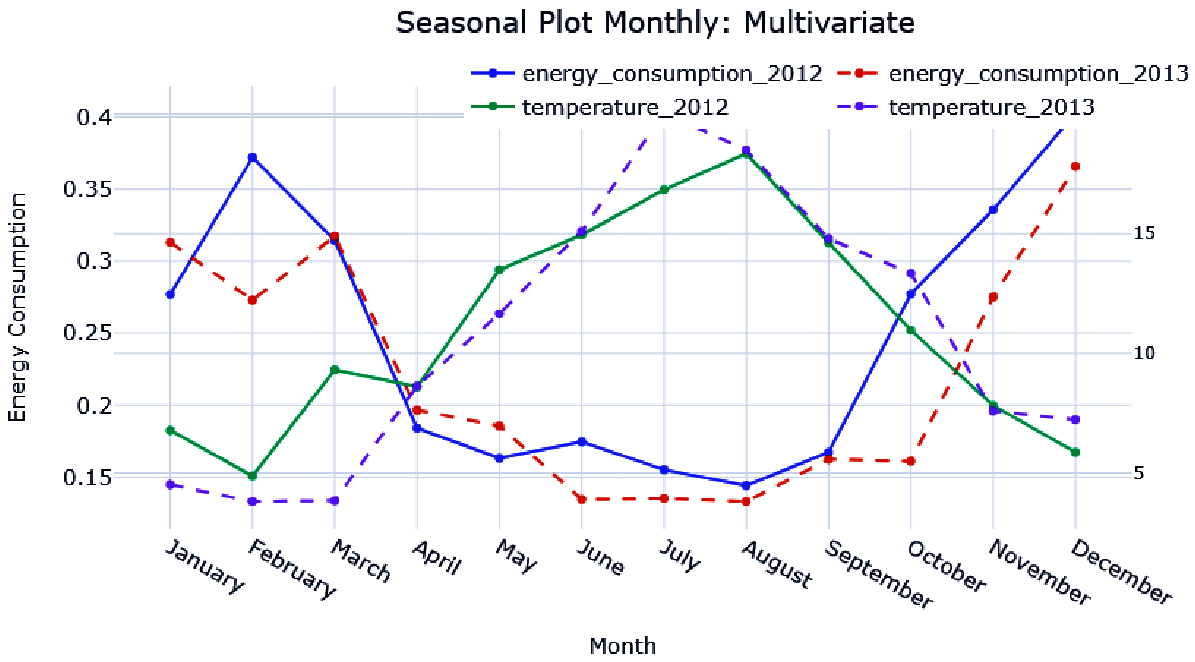

Energy Consumption for MAC000193
Energy Consumption
0
1
2
3
4
Jan 2012
Apr 2012
Jul 2012
Oct 2012
Jan 2013
Apr 2013
Jul 2013
Oct 2013
Jan 2014
Time

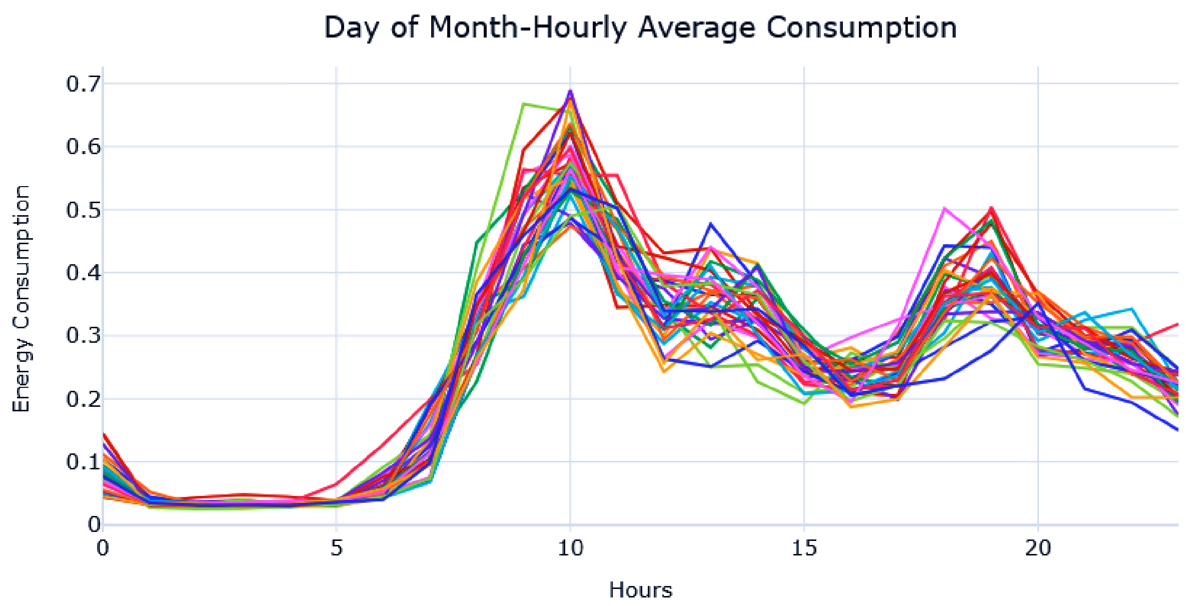

Rolling Monthly Average Energy Consumption for MAC000193
Energy Consumption
0.15
0.2
0.25
0.3
0.35
0.4
Jan 2012
Apr 2012
Jul 2012
Oct 2012
Jan 2013
Apr 2013
Jul 2013
Oct 2013
Jan 2014
Time

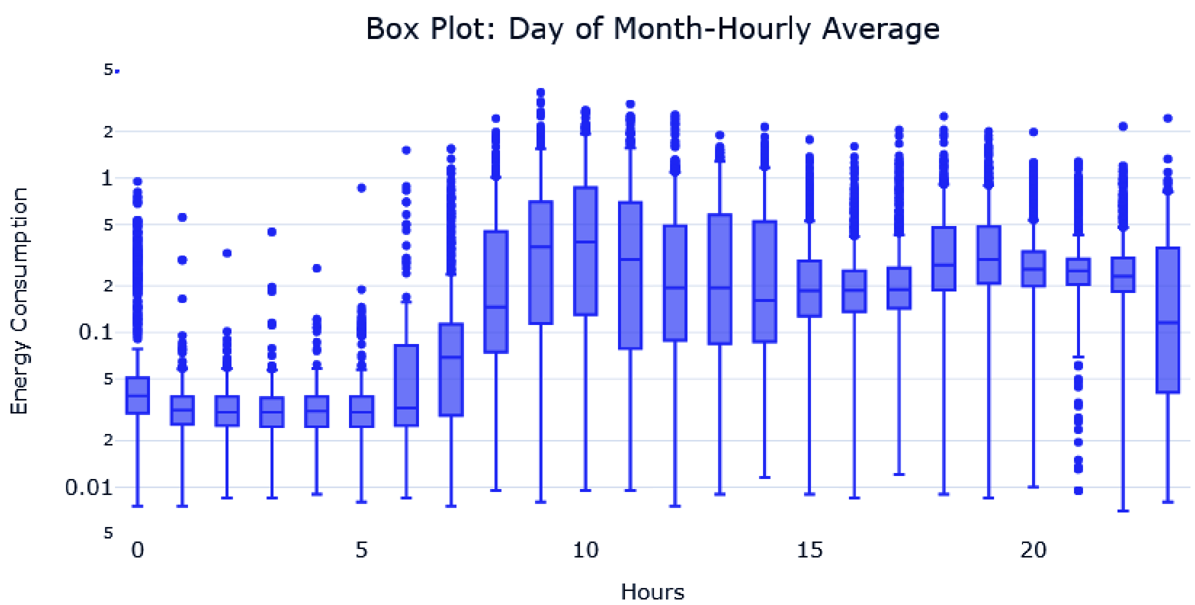

Temperature and Energy Consumption
Energy Consumption
Temperature
Value
Time
Jan 2012
Apr 2012
Jul 2012
Oct 2012
Jan 2013
Apr 2013
Jul 2013
Oct 2013
Jan 2014
0
1
2
3
4
−5
0
5
10
15
20
25
30

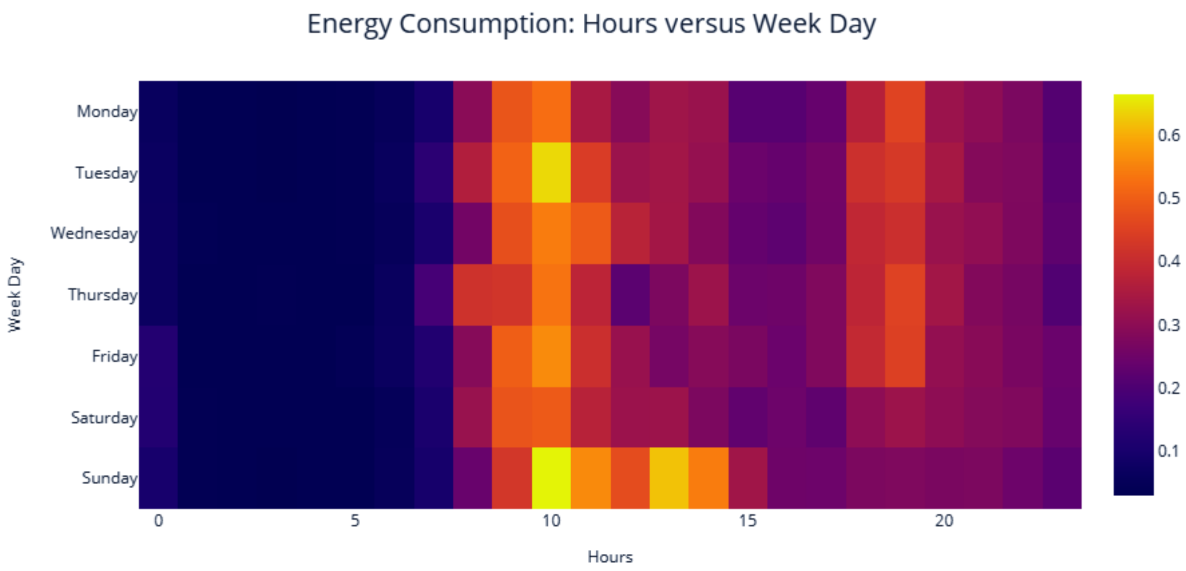

Temperature and Energy Consumption (2012-03-16 to 2012-03-30)
Energy Consumption
Temperature
Value
Time
Mar 17
2012
Mar 19
Mar 21
Mar 23
Mar 25
Mar 27
Mar 29
0
0.5
1
1.5
2
5
10
15
20

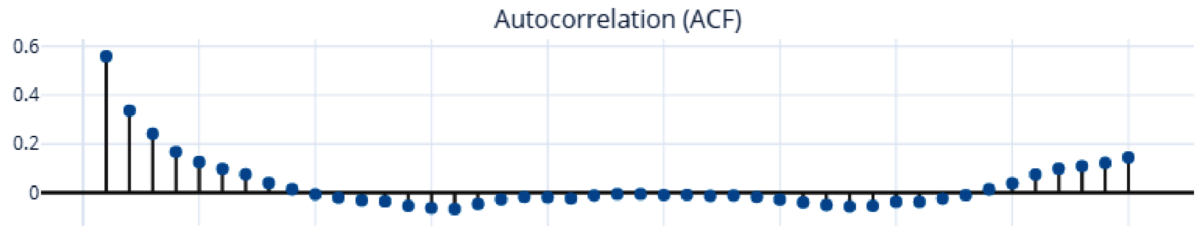

Seasonal Plot - Monthly
2012
2013
Energy Consumption
0.4
0.35
0.3
0.25
0.2
0.15
January
February
March
April
May
June
July
August
September
October
November
December
Month

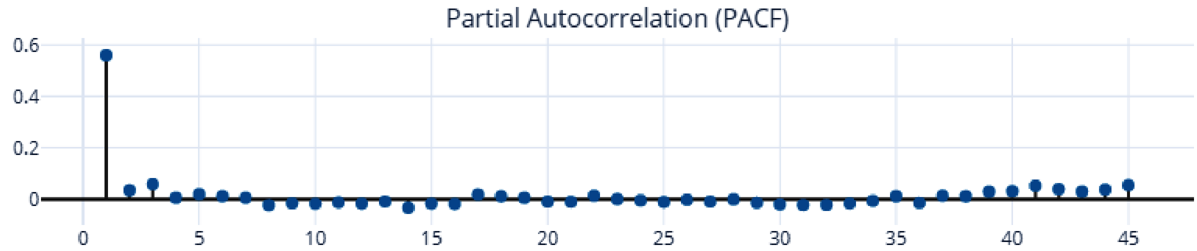

Seasonal Plot Monthly: Multivariate
energy_consumption_2012
energy_consumption_2013
temperature_2012
temperature_2013
Energy Consumption
0.4
0.35
0.3
0.25
0.2
0.15
15
10
5
January
February
March
April
May
June
July
August
September
October
November
December
Month

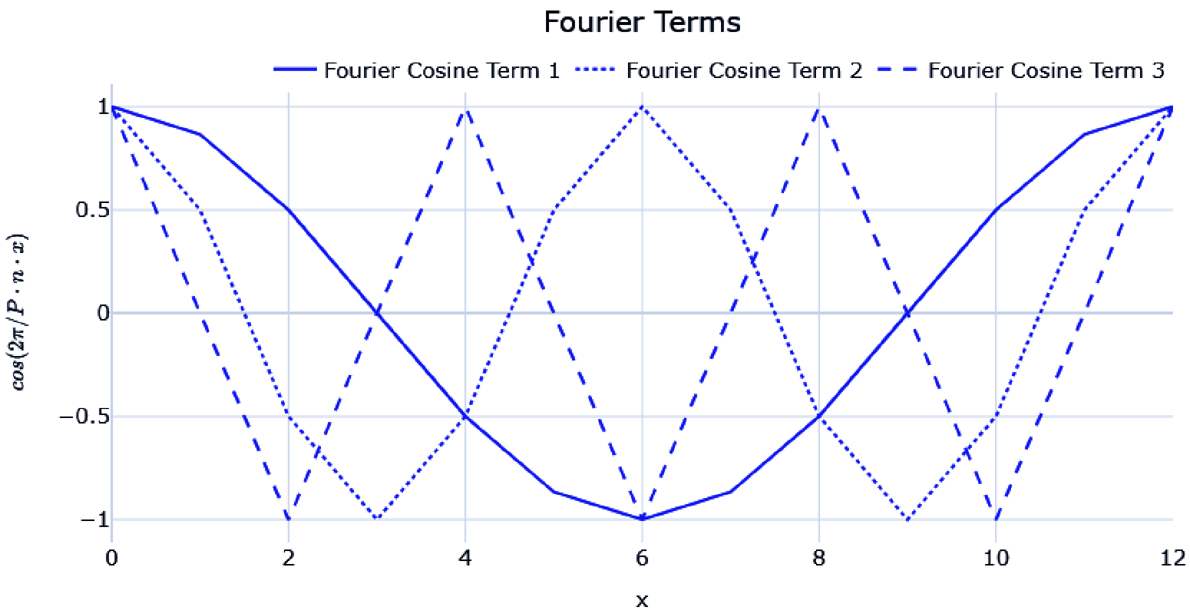

Day of Month-Hourly Average Consumption
Energy Consumption
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
5
10
15
20
Hours

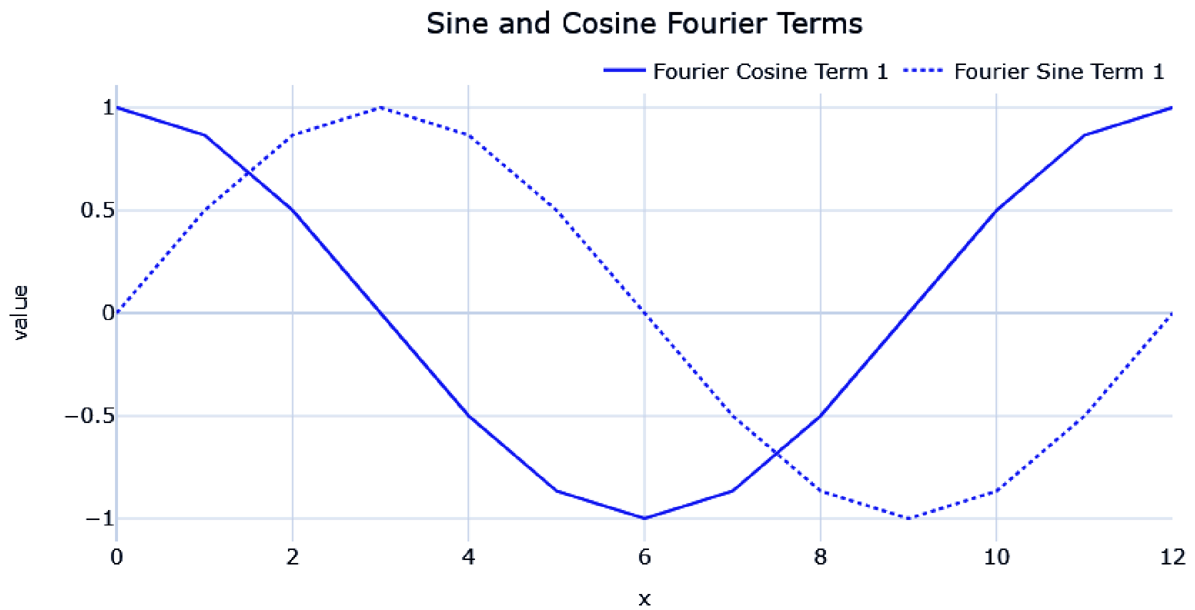

Box Plot: Day of Month-Hourly Average
Energy Consumption
5
2
1
5
2
0.1
5
2
0.01
5
0
5
10
15
20
Hours


Energy Consumption: Hours versus Week Day
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Week Day
0
5
10
15
20
Hours
0.6
0.5
0.4
0.3
0.2
0.1

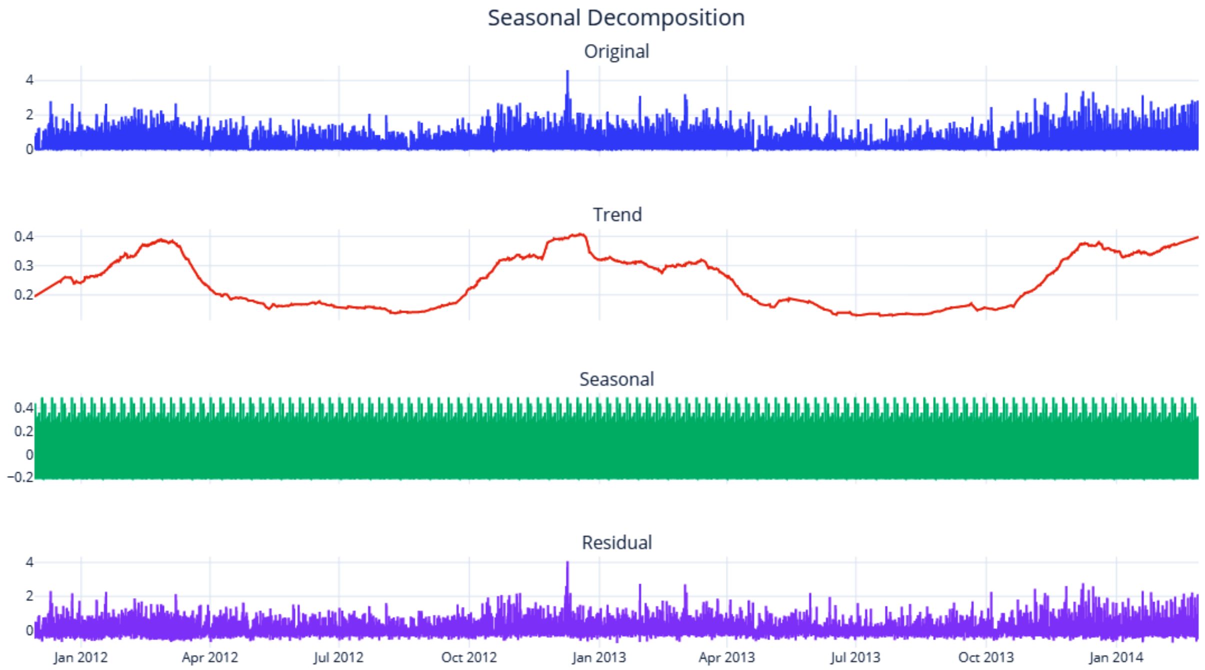

Autocorrelation (ACF)
0.6
0.4
0.2
0


Partial Autocorrelation (PACF)
0.6
0.4
0.2
0
0
5
10
15
20
25
30
35
40
45

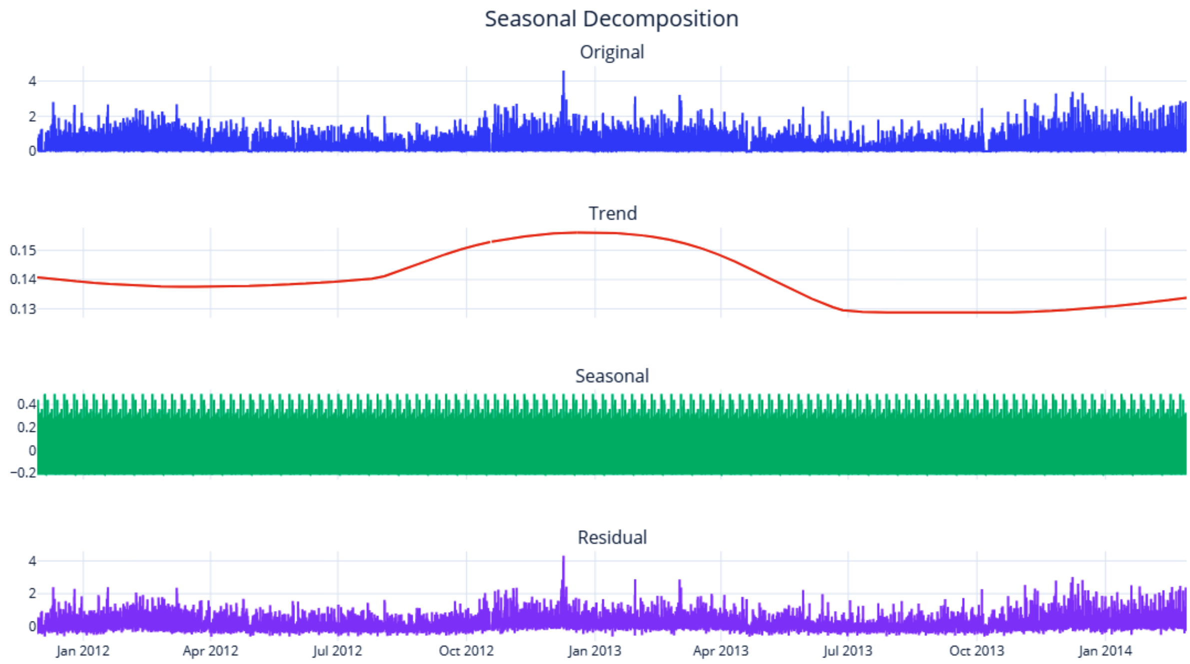

Fourier Terms
Fourier Cosine Term 1
Fourier Cosine Term 2
Fourier Cosine Term 3
1
0.5
0
−0.5
−1
$cos(2\pi/P \cdot n \cdot x)$
0
2
4
6
8
10
12
x

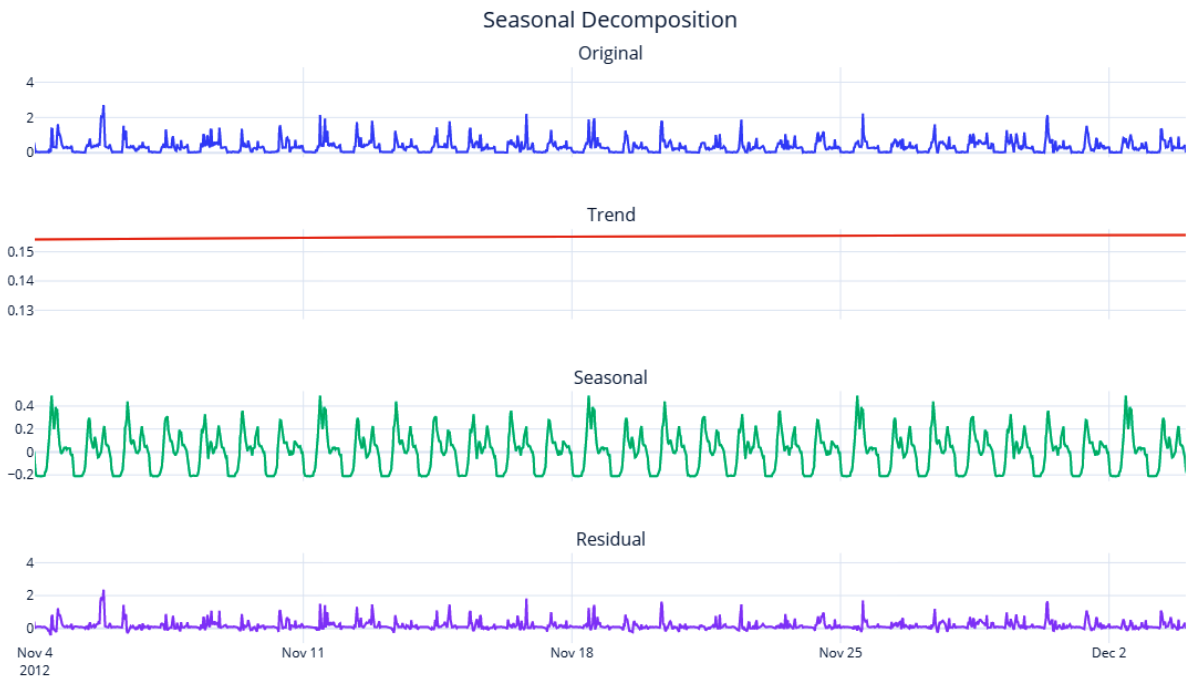


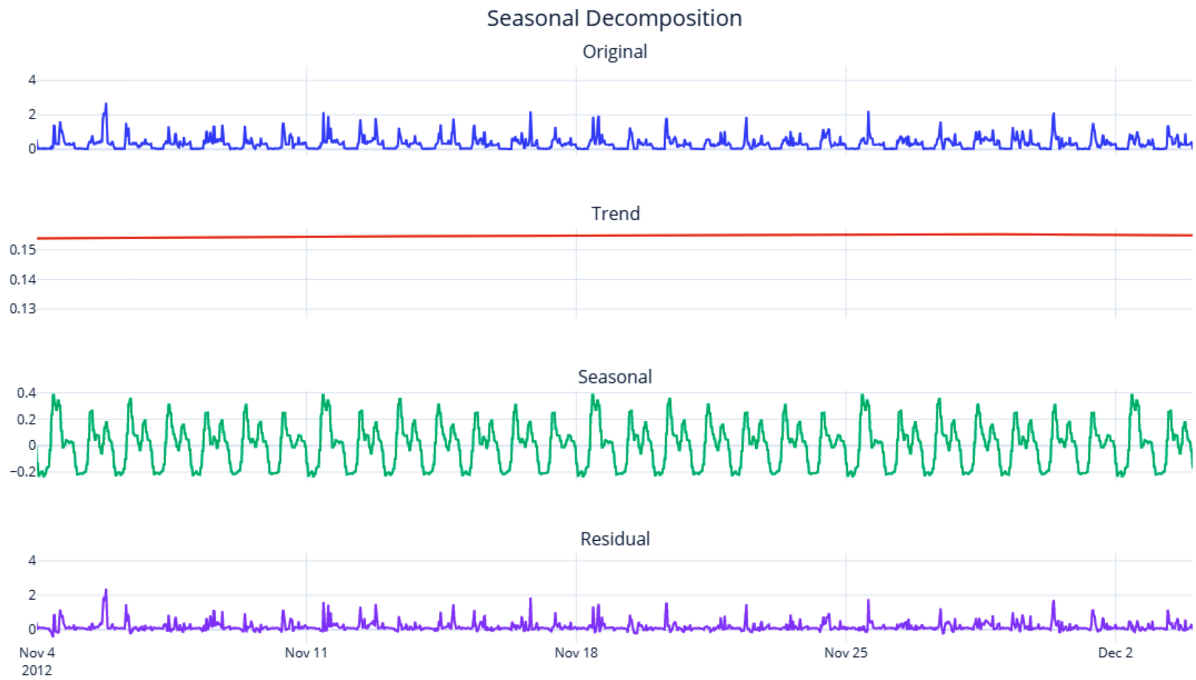
Sine and Cosine Fourier Terms
Fourier Cosine Term 1
Fourier Sine Term 1
value
x



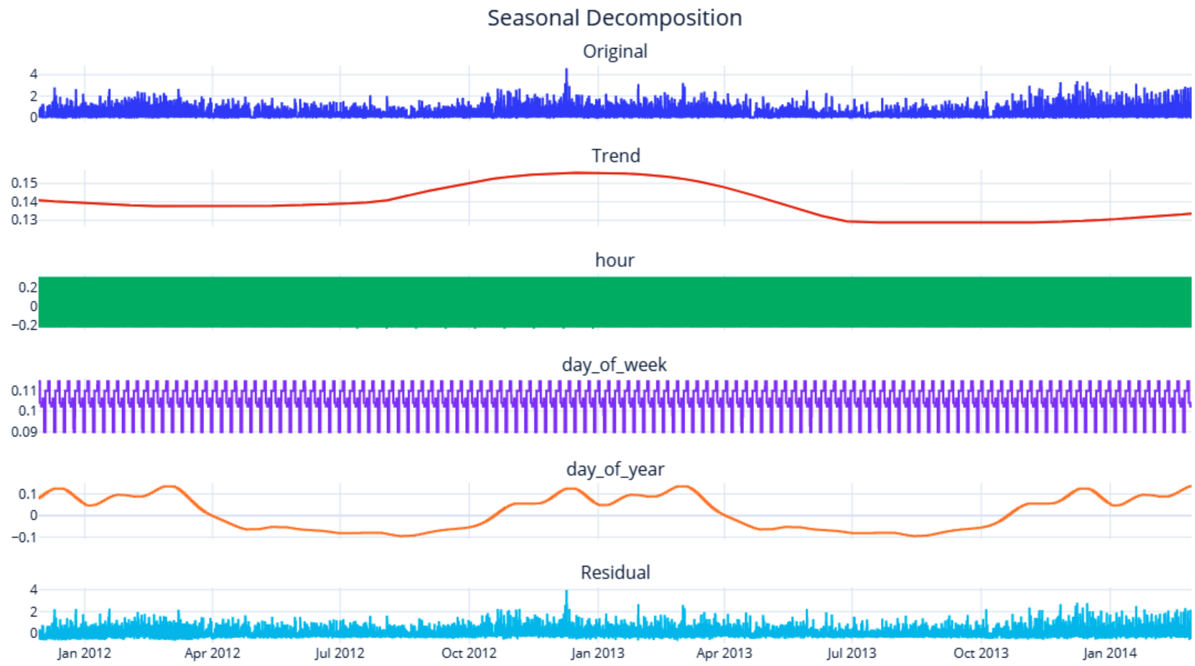
Seasonal Decomposition
Original
Trend
Seasonal
Residual
Jan 2012
Apr 2012
Jul 2012
Oct 2012
Jan 2013
Apr 2013
Jul 2013
Oct 2013
Jan 2014


Seasonal Decomposition
Original
Trend
Seasonal
Residual
Jan 2012
Apr 2012
Jul 2012
Oct 2012
Jan 2013
Apr 2013
Jul 2013
Oct 2013
Jan 2014

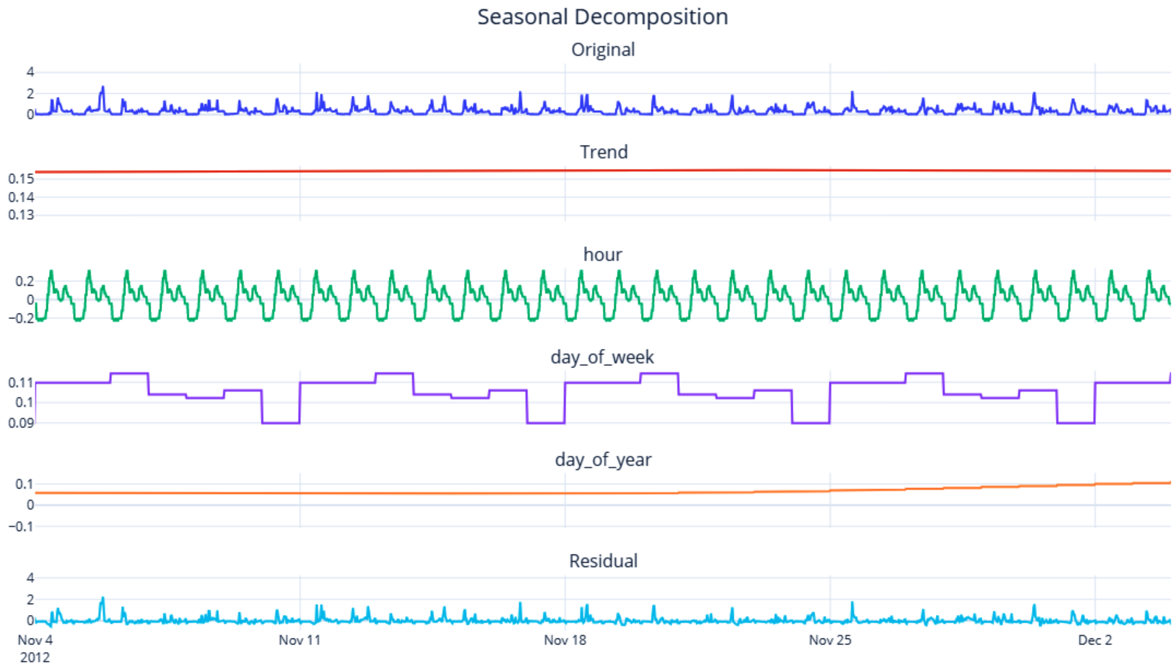

Seasonal Decomposition
Original
Trend
Seasonal
Residual
Nov 4
2012
Nov 11
Nov 18
Nov 25
Dec 2



Seasonal Decomposition
Original
4
2
0
Trend
0.15
0.14
0.13
Seasonal
0.4
0.2
0
−0.2
Residual
4
2
0
Nov 4
2012
Nov 11
Nov 18
Nov 25
Dec 2

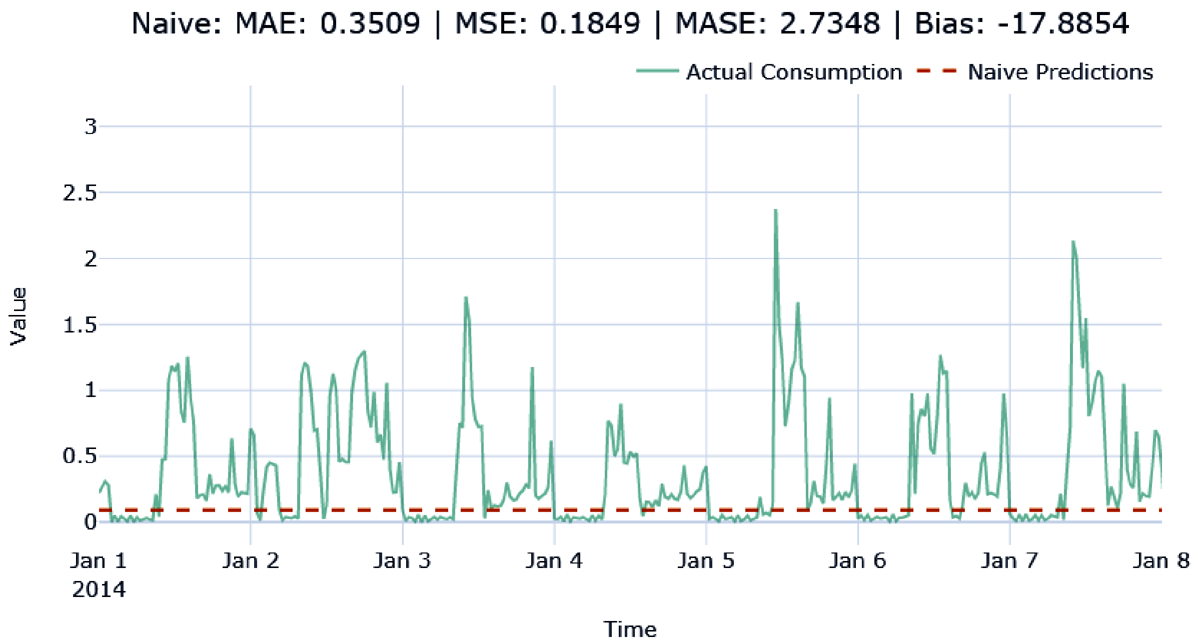


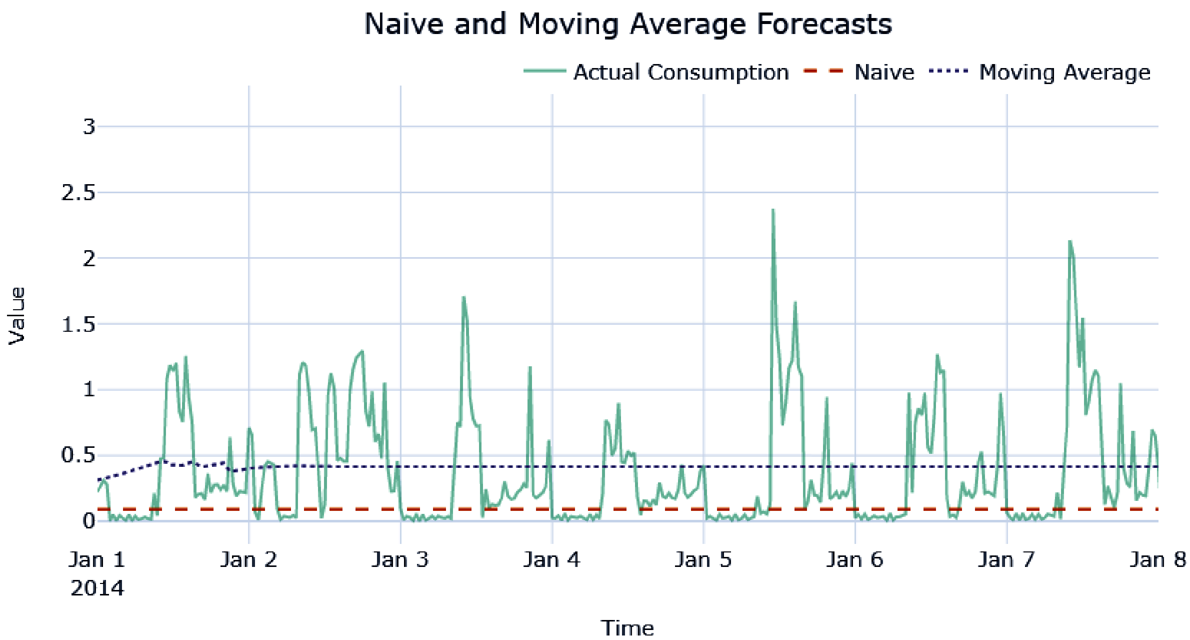
Seasonal Decomposition
Original
4
2
0
Trend
0.15
0.14
0.13
hour
0.2
0
−0.2
day_of_week
0.11
0.1
0.09
day_of_year
0.1
0
−0.1
Residual
4
2
0
Jan 2012
Apr 2012
Jul 2012
Oct 2012
Jan 2013
Apr 2013
Jul 2013
Oct 2013
Jan 2014

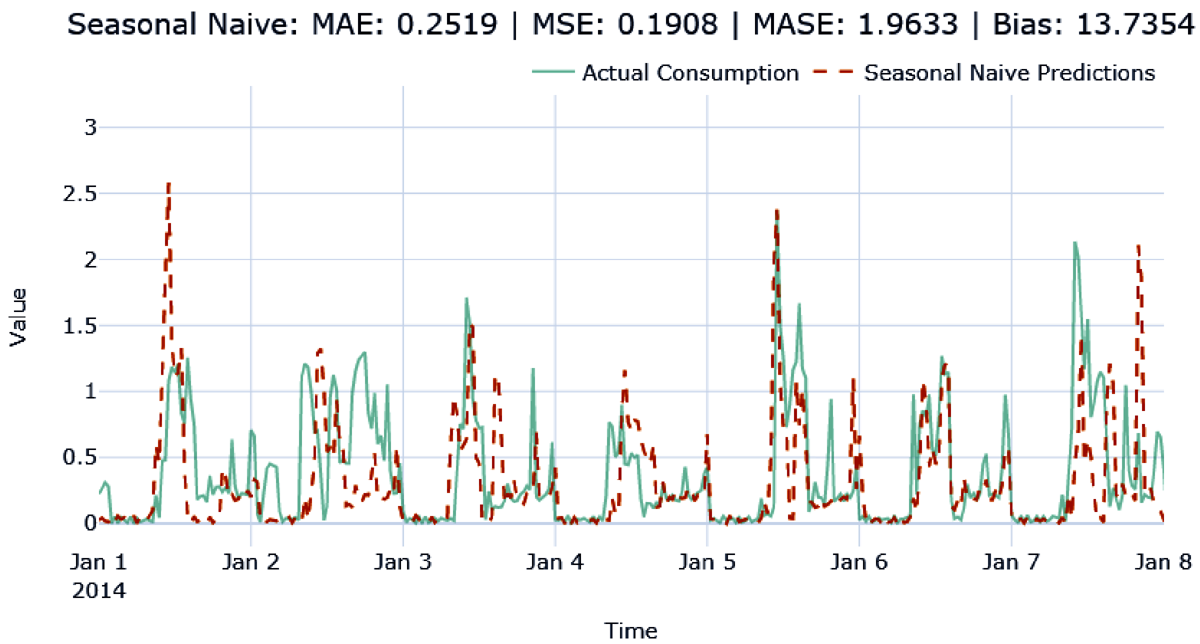


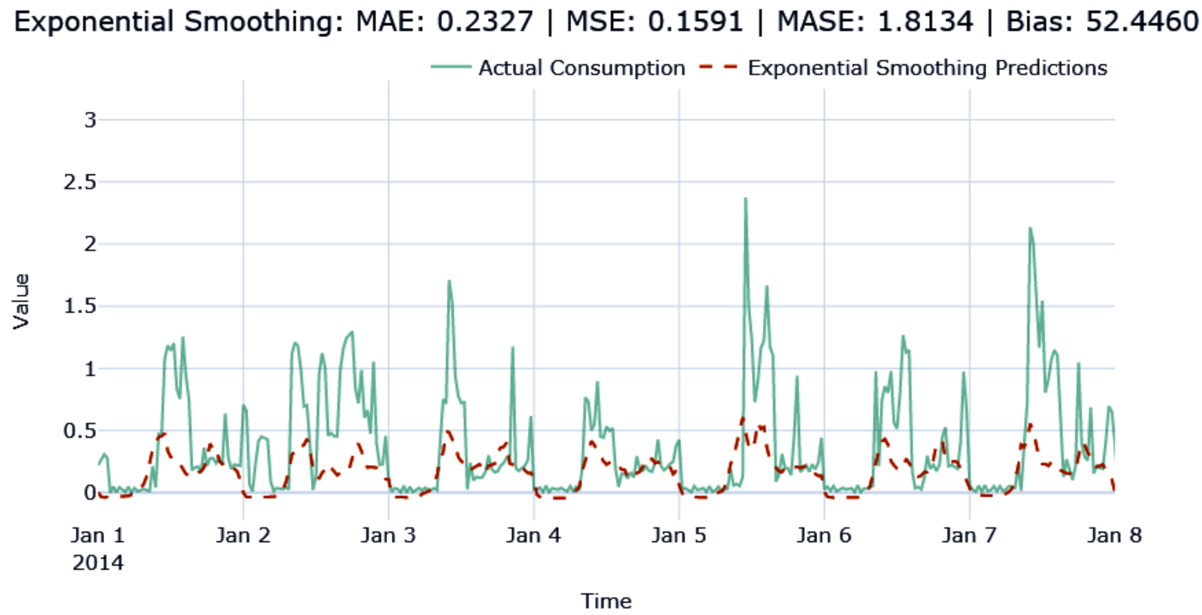
	# of Outliers	% of Outliers
3SD	802	2.12%
2SD on Residuals	728	1.92%
4IQR	747	1.97%
4SD on Residuals	468	1.24%
Isolation Forest	364	0.96%
Isolation Forest on Residuals	359	0.95%
ESD	420	1.11%
S-ESD	424	1.12%

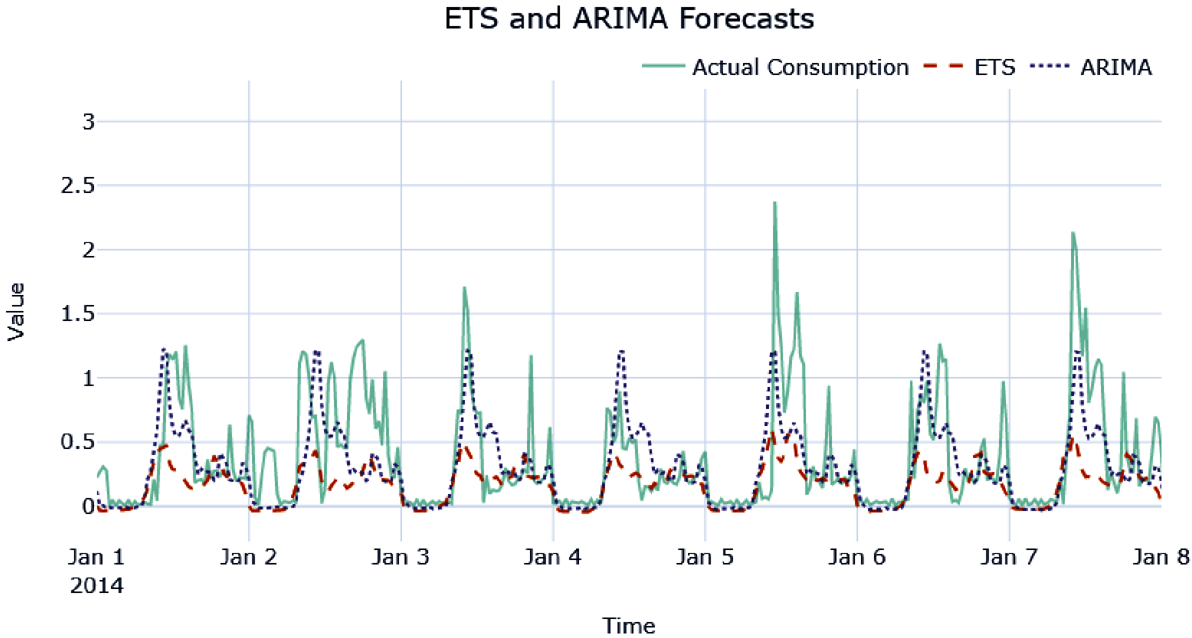
Chapter 4: Setting a Strong Baseline Forecast

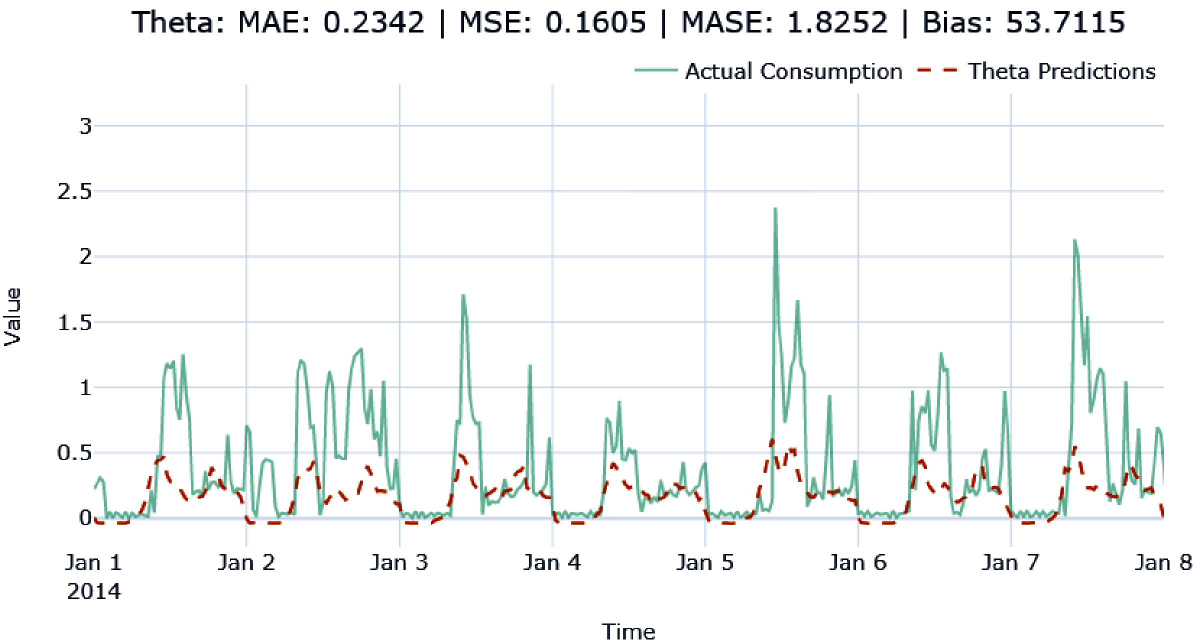


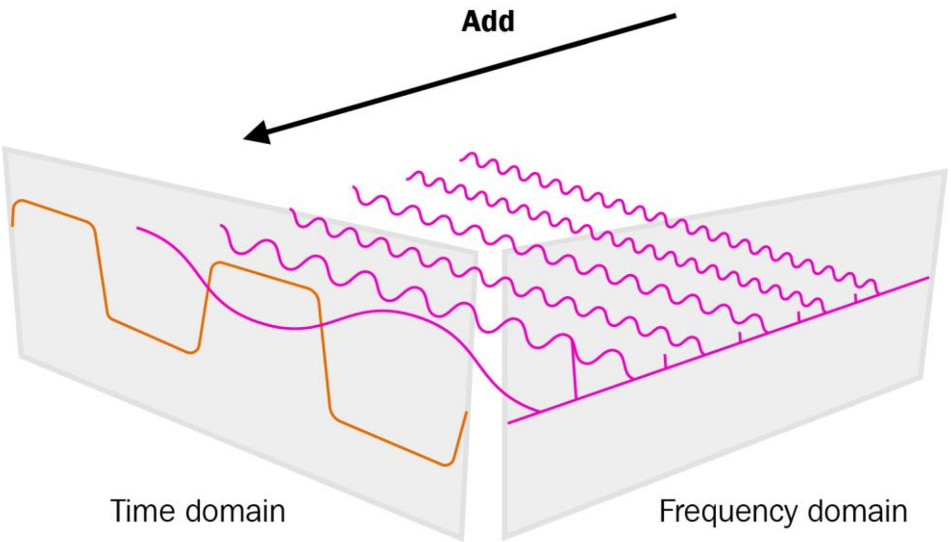


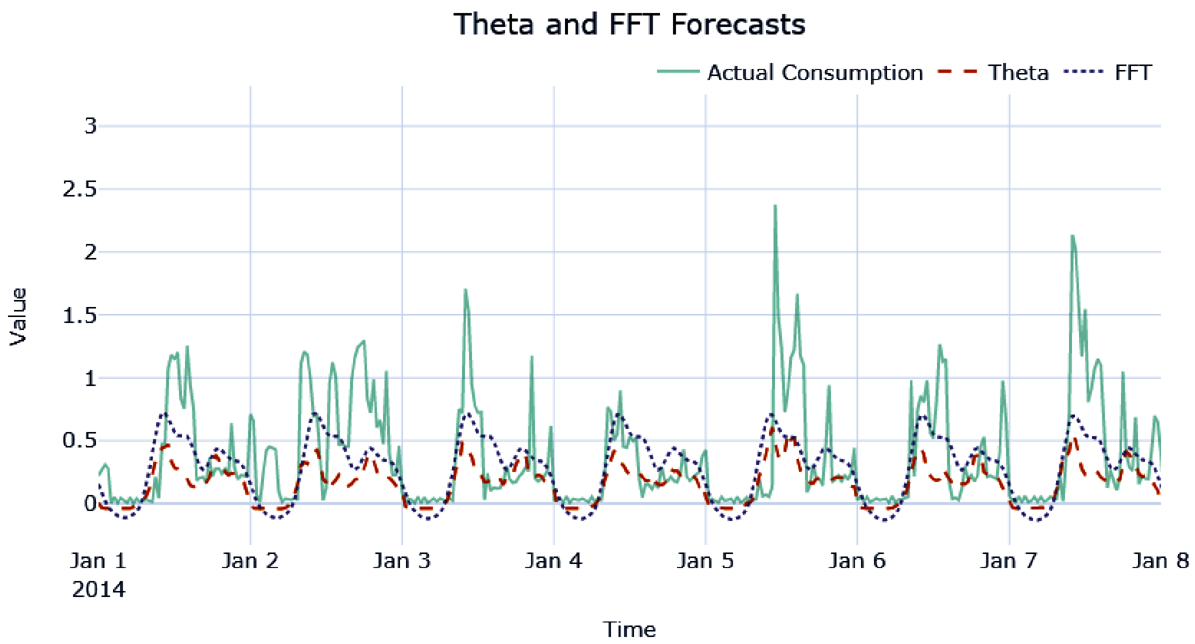

Seasonal Naive: MAE: 0.2519 | MSE: 0.1908 | MASE: 1.9633 | Bias: 13.7354
Actual Consumption
Seasonal Naive Predictions
Value
Time
Jan 1
2014
Jan 2
Jan 3
Jan 4
Jan 5
Jan 6
Jan 7
Jan 8


Exponential Smoothing: MAE: 0.2327 | MSE: 0.1591 | MASE: 1.8134 | Bias: 52.4460
Actual Consumption
Exponential Smoothing Predictions
Value
Time
Jan 1
2014
Jan 2
Jan 3
Jan 4
Jan 5
Jan 6
Jan 7
Jan 8


ETS and ARIMA Forecasts
Actual Consumption
ETS
ARIMA
Value
3
2.5
2
1.5
1
0.5
0
Jan 1
2014
Jan 2
Jan 3
Jan 4
Jan 5
Jan 6
Jan 7
Jan 8
Time


Theta: MAE: 0.2342 | MSE: 0.1605 | MASE: 1.8252 | Bias: 53.7115
Actual Consumption
Theta Predictions
Value
3
2.5
2
1.5
1
0.5
0
Jan 1
2014
Jan 2
Jan 3
Jan 4
Jan 5
Jan 6
Jan 7
Jan 8
Time


Add
Time domain
Frequency domain



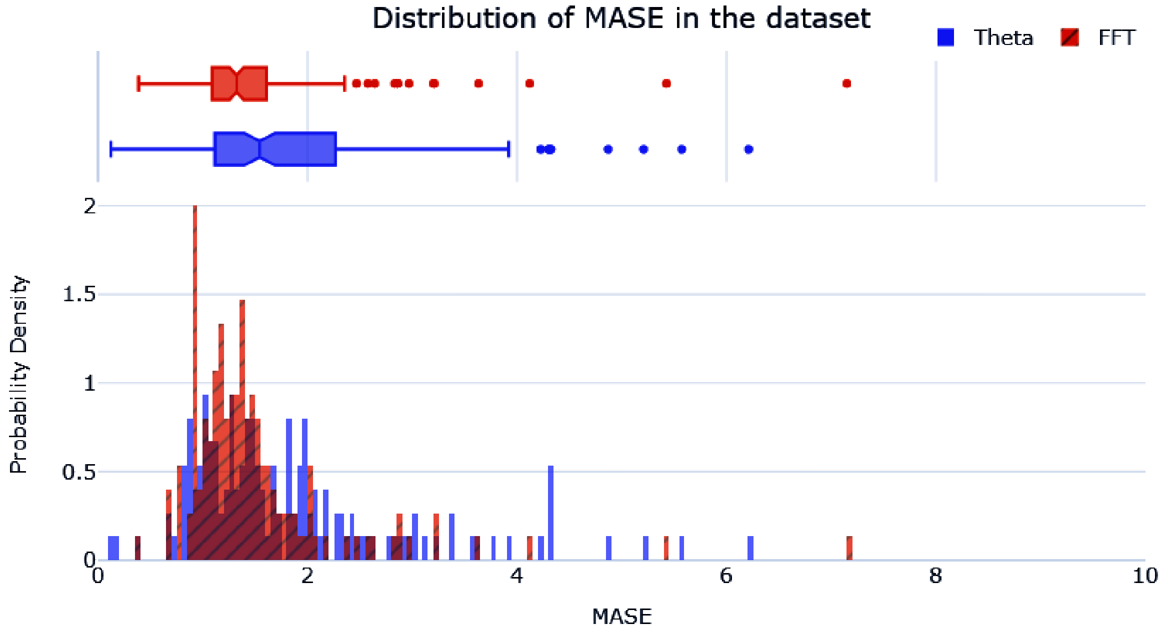
Algorithm	MAE	MSE	MASE	Forecast Bias	Time Elapsed
Naive	0.305	0.249	2.380	74.34%	0.045541
Moving Average Forecast	0.351	0.185	2.735	-17.89%	0.096654
Seasonal Naive Forecast	0.252	0.191	1.963	13.74%	0.055316
Exponential Smoothing	0.233	0.159	1.813	52.45%	29.352878
ARIMA	0.203	0.107	1.639	24.00%	319.322491
Theta	0.234	0.160	1.825	53.71%	0.268956
FFT	0.239	0.120	1.860	23.15%	0.592196

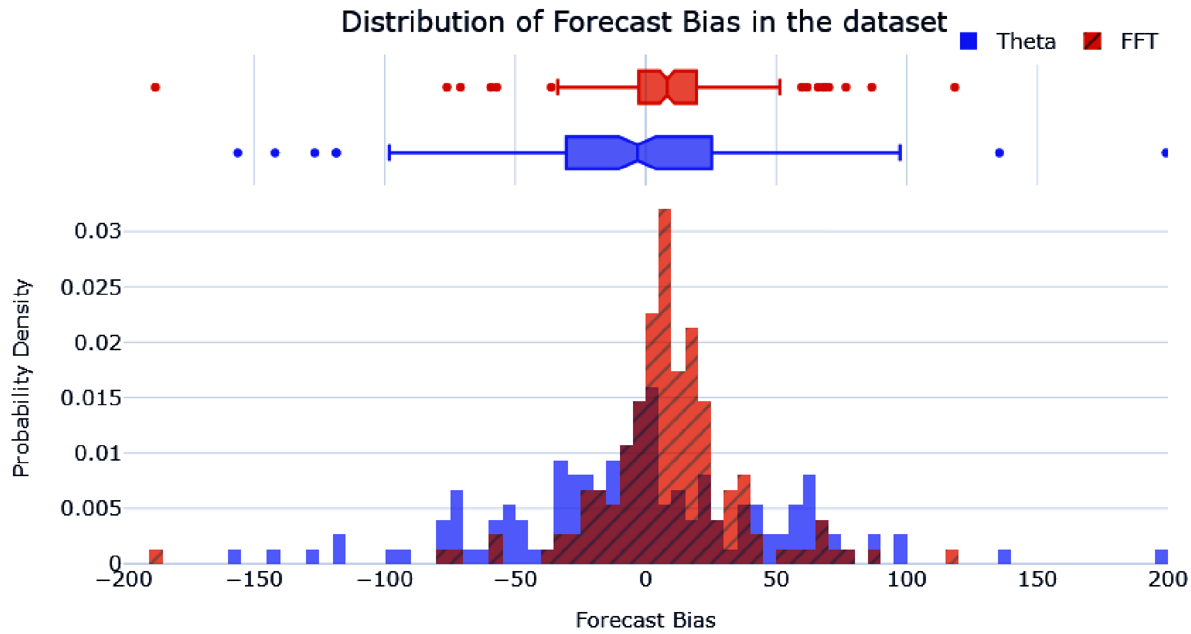
Validation

Algorithm	MAE	MSE	meanMASE	Forecast Bias
FFT	0.206	0.128	2.179	16.73%
Theta	0.282	0.245	2.274	11.80%

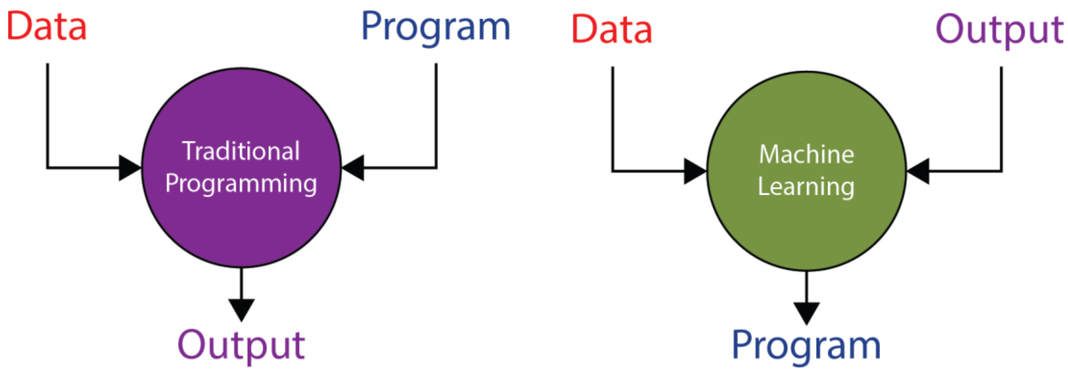
Test

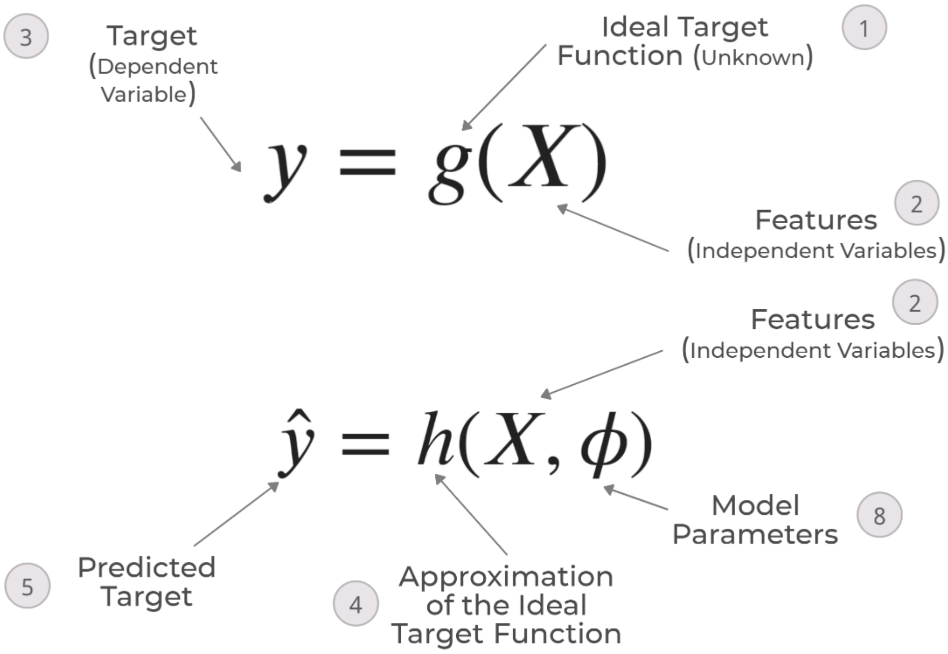
Algorithm	MAE	MSE	meanMASE	Forecast Bias
FFT	0.198	0.113	2.014	8.54%
Theta	0.226	0.139	1.913	7.64%

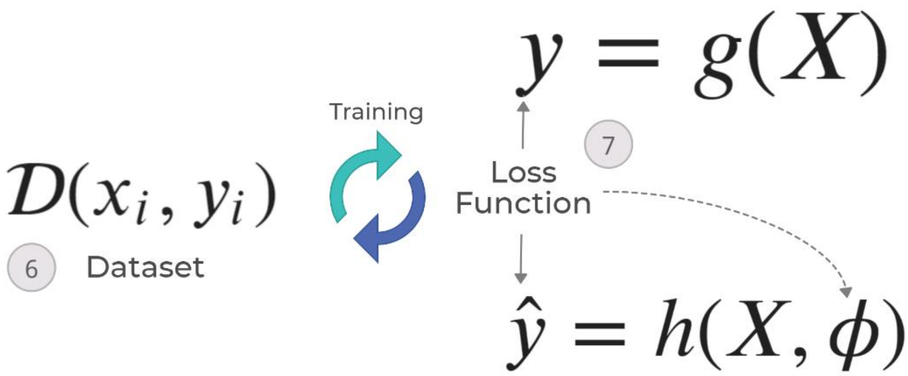


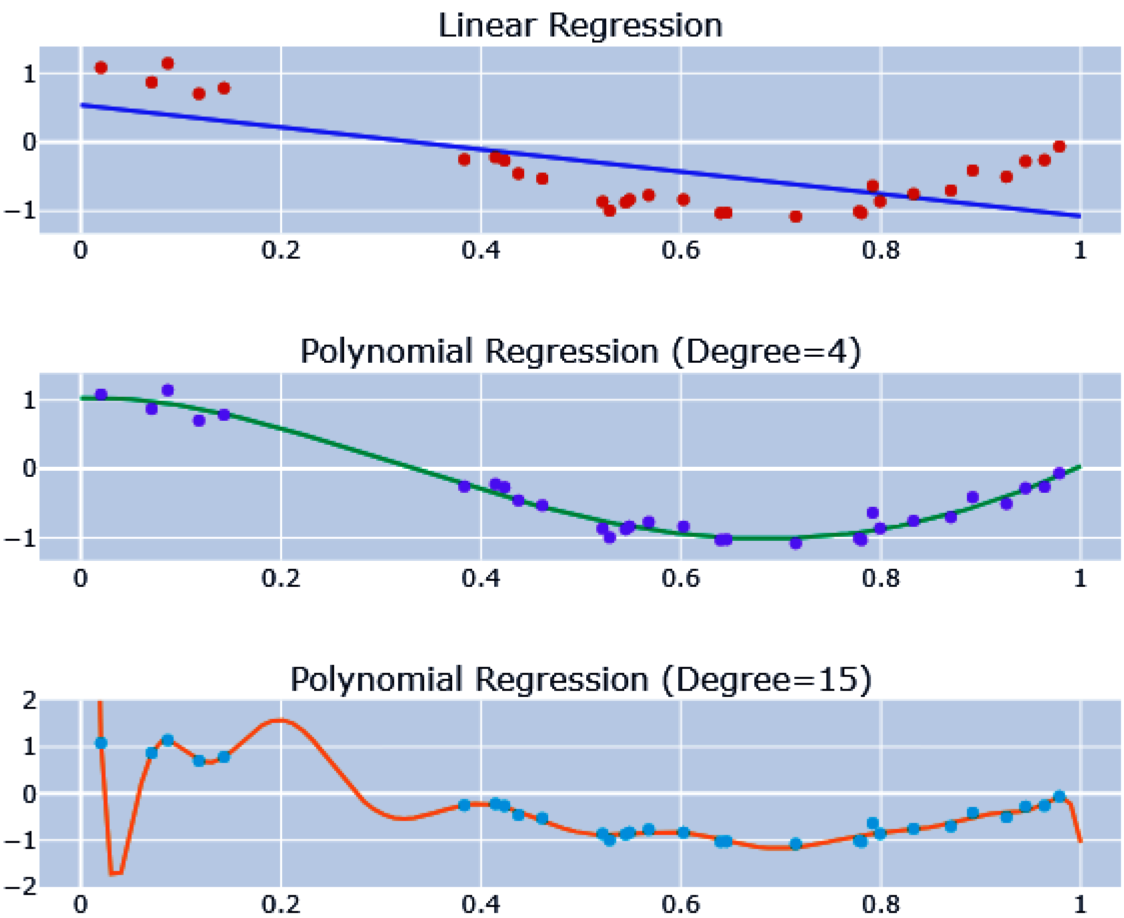


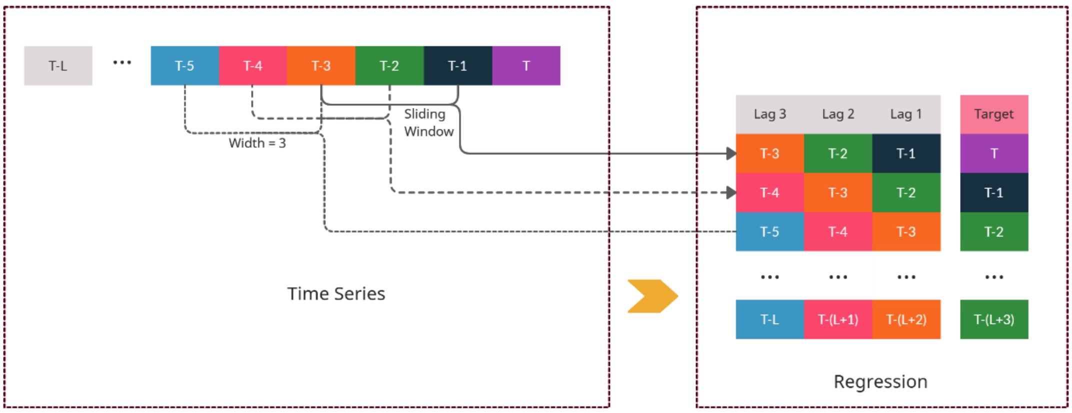
Chapter 5: Time Series Forecasting as Regression



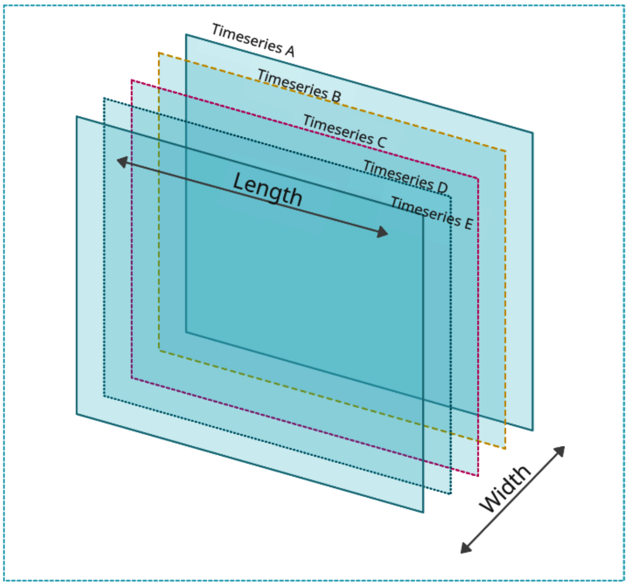




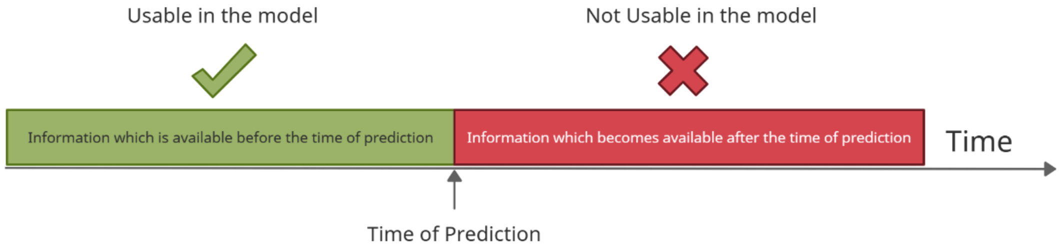

Linear Regression
Polynomial Regression (Degree=4)
Polynomial Regression (Degree=15)
0
0.2
0.4
0.6
0.8
1
2
1
0
−1
−2

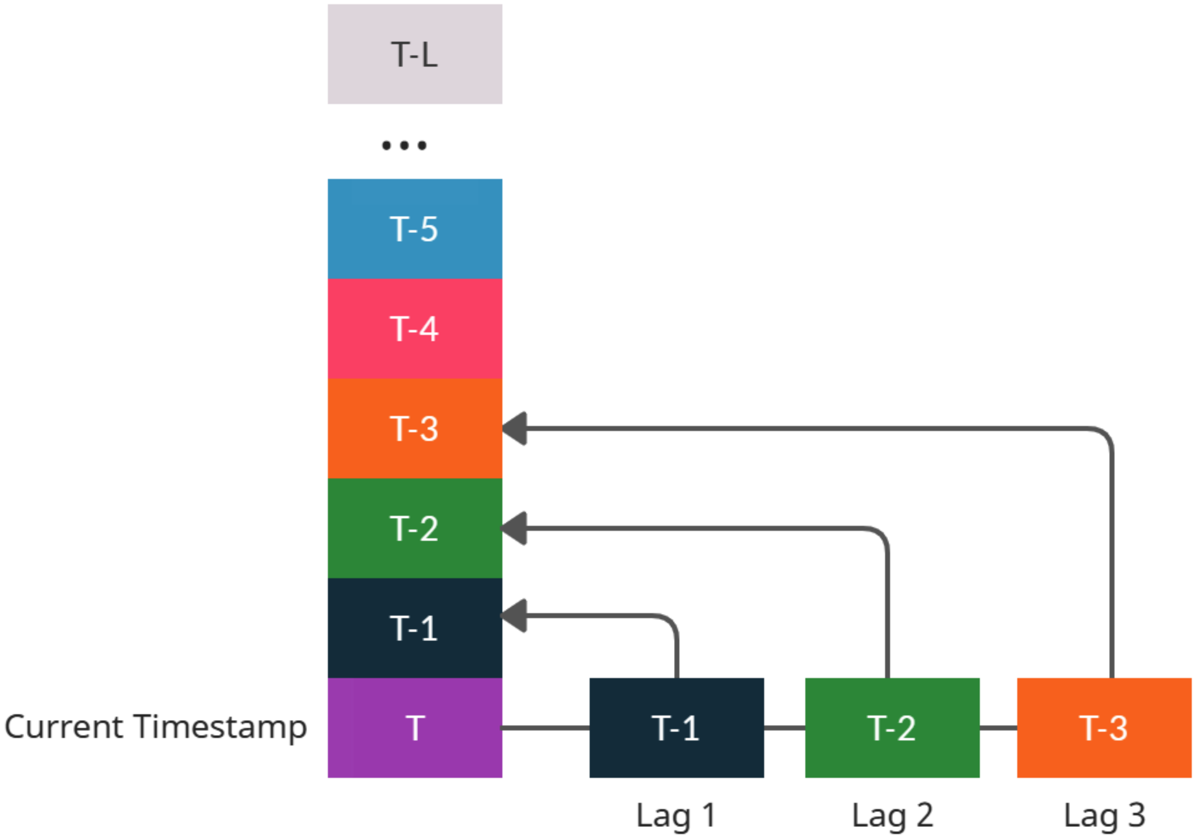

T-L
T-5
T-4
T-3
T-2
T-1
T
Sliding Window
Width = 3
Time Series
Lag 3
Lag 2
Lag 1
Target
T-(L+1)
T-(L+2)
T-(L+3)
Regression

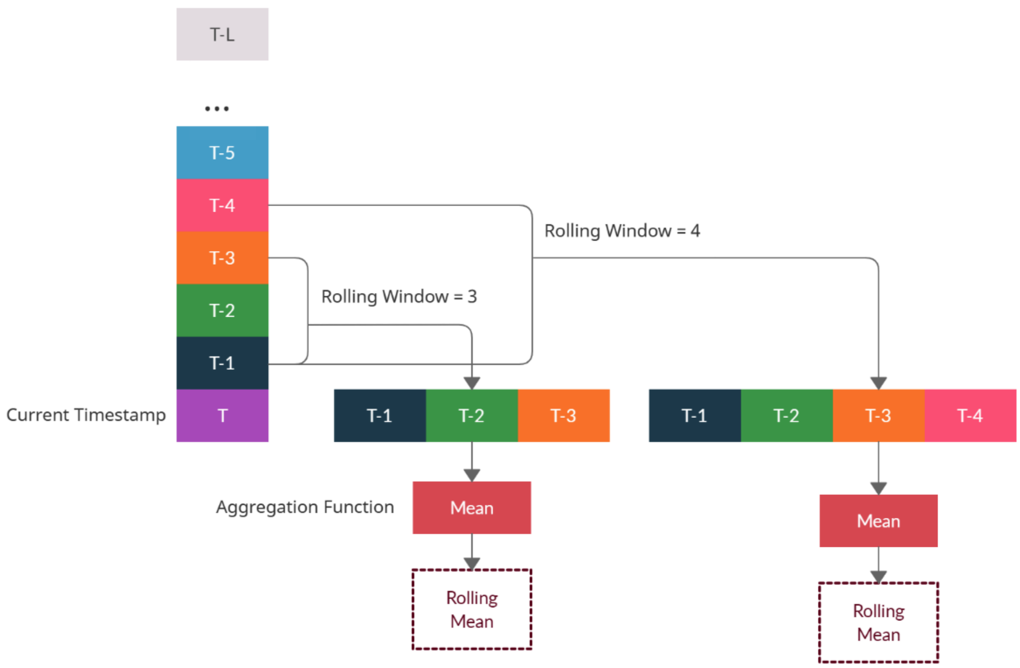
Time Series Dataset

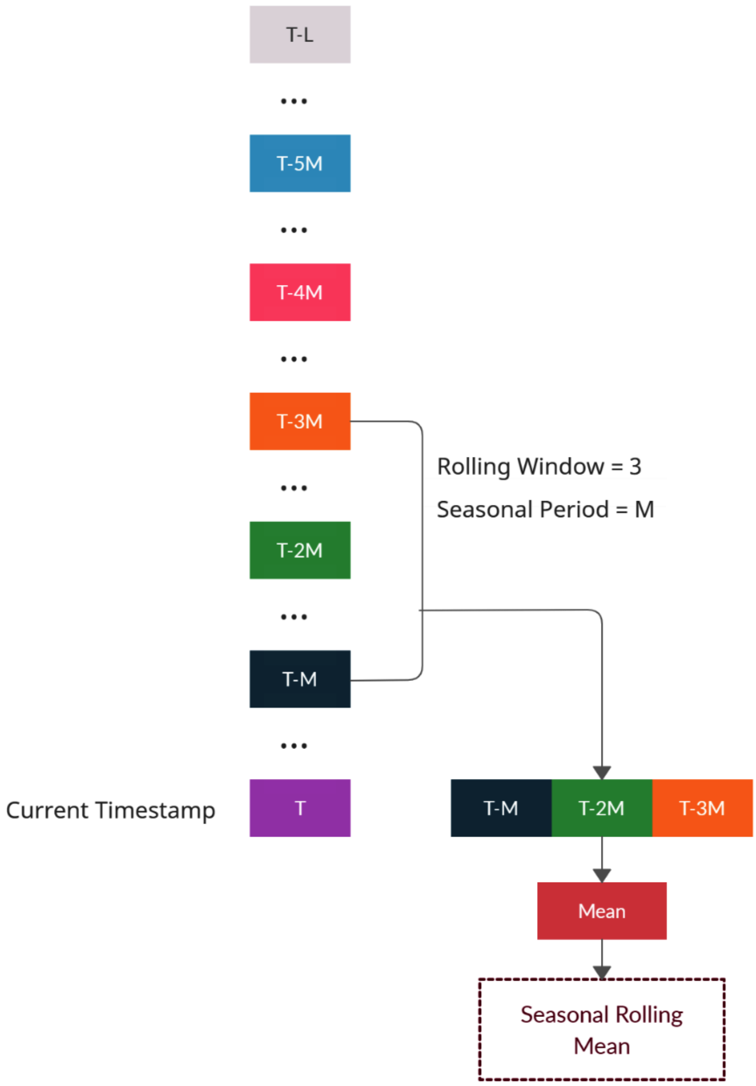


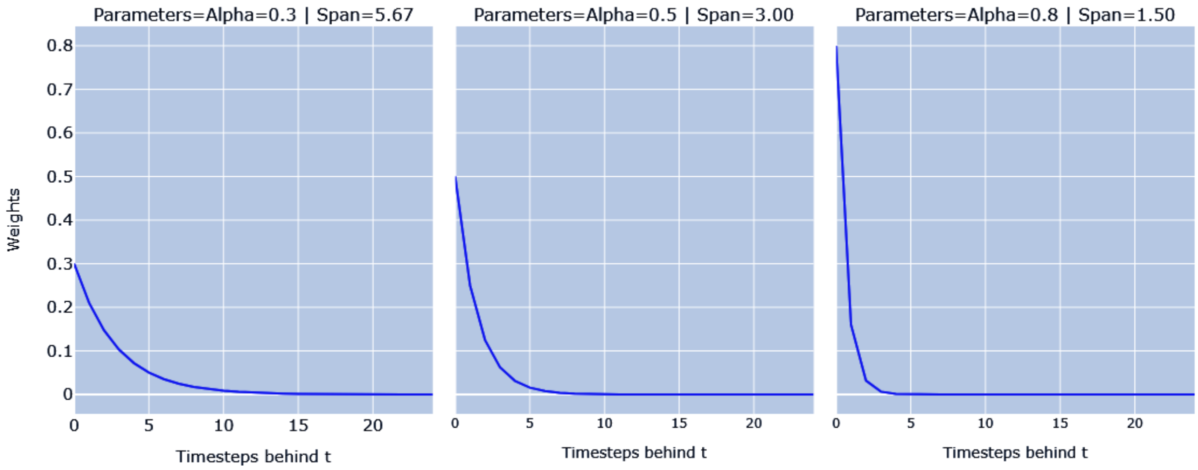
Chapter 6: Feature Engineering for Time Series Forecasting

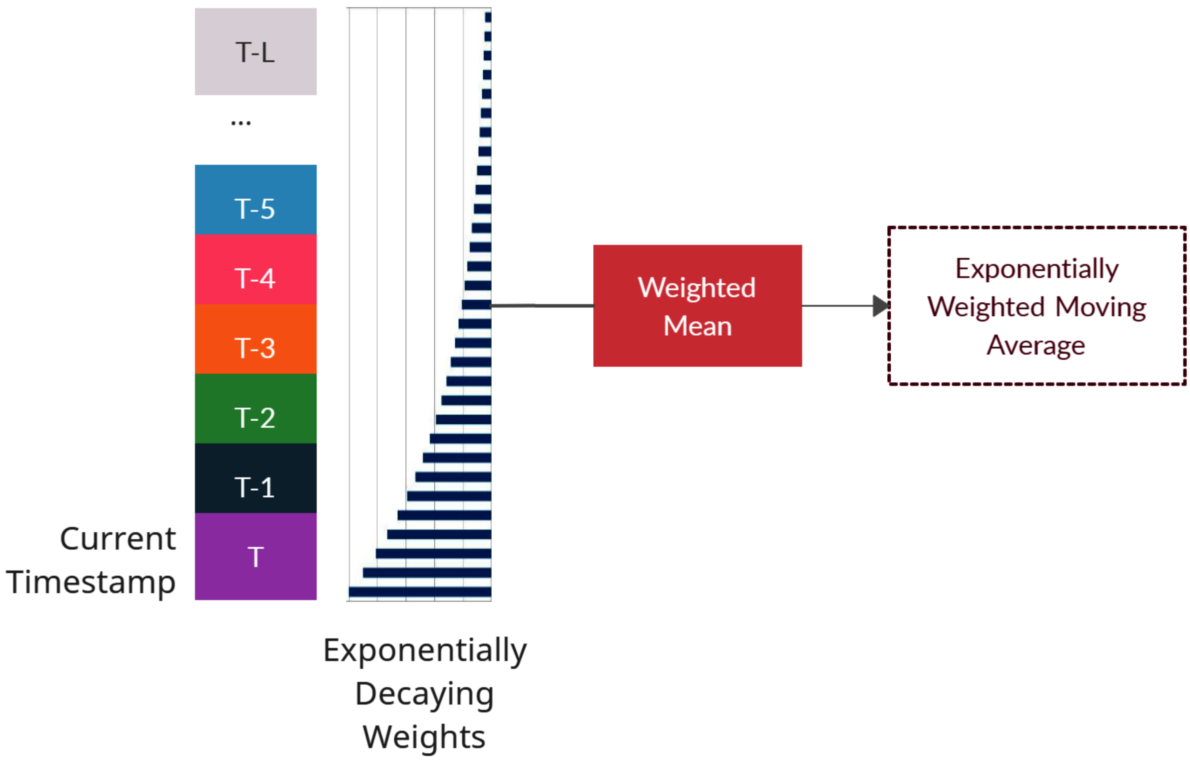


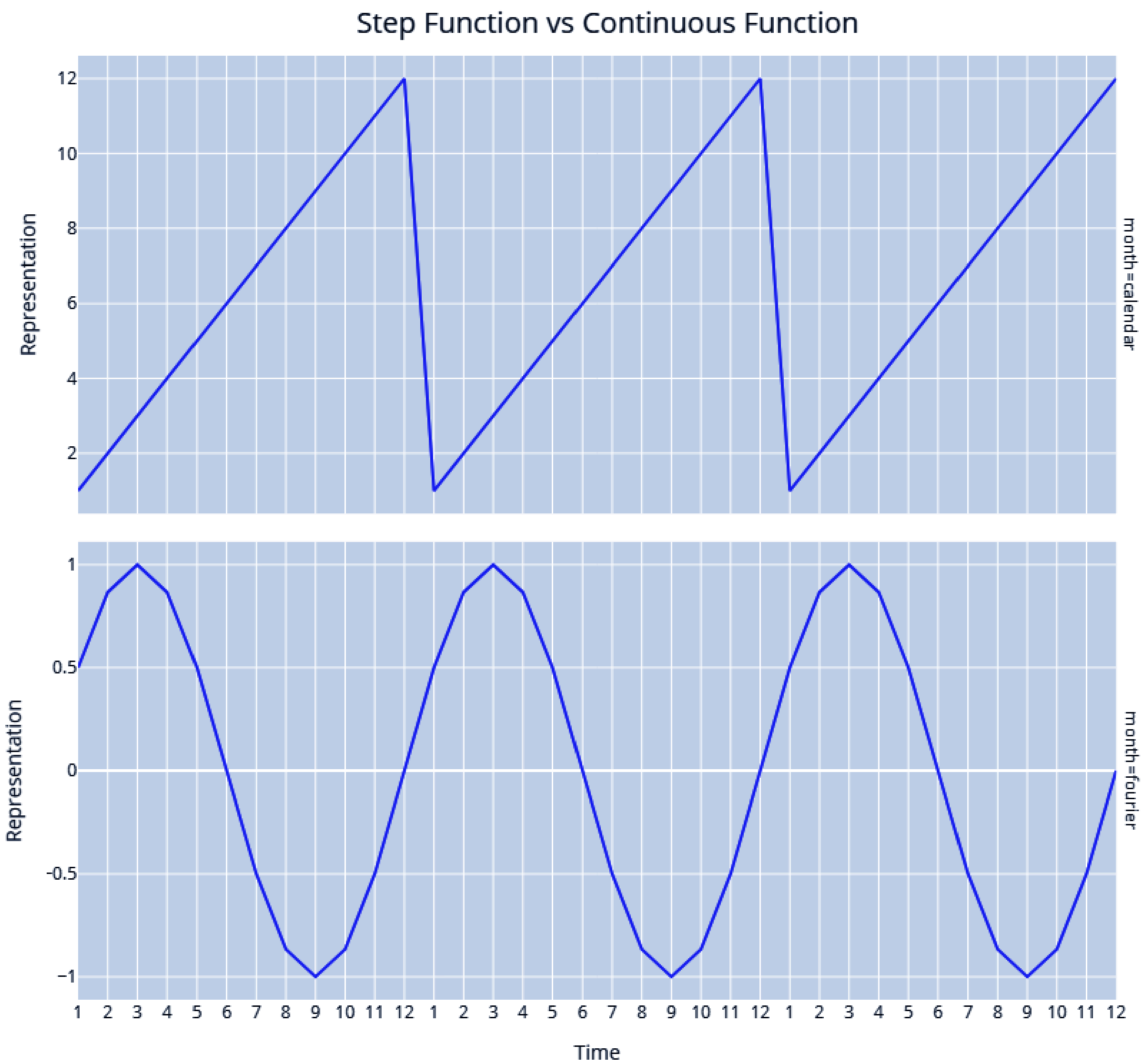





T-L
T-5M
T-4M
T-3M
T-2M
T-M
T
Current Timestamp
Rolling Window = 3
Seasonal Period = M
T-M
T-2M
T-3M
Mean
Seasonal Rolling Mean

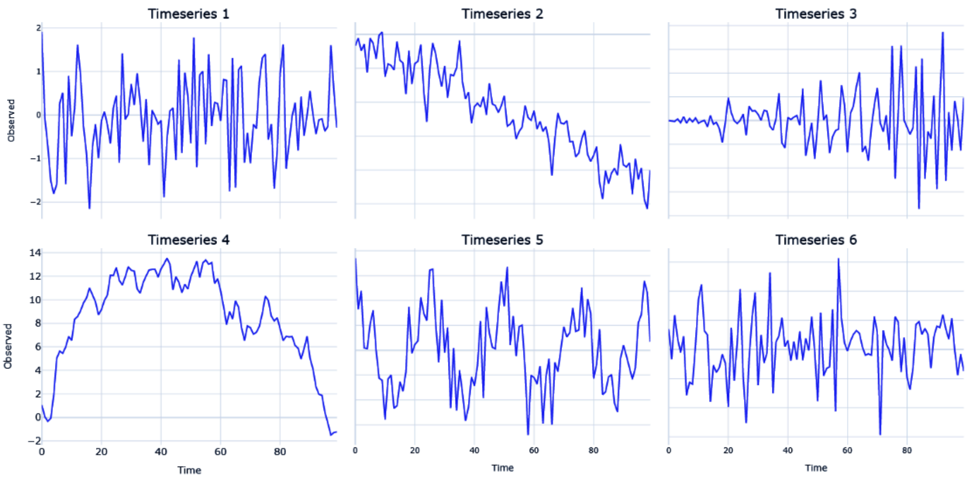

Parameters=Alpha=0.3 | Span=5.67
Parameters=Alpha=0.5 | Span=3.00
Parameters=Alpha=0.8 | Span=1.50
Weights
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
5
10
15
20
Timesteps behind t
Timesteps behind t
Timesteps behind t

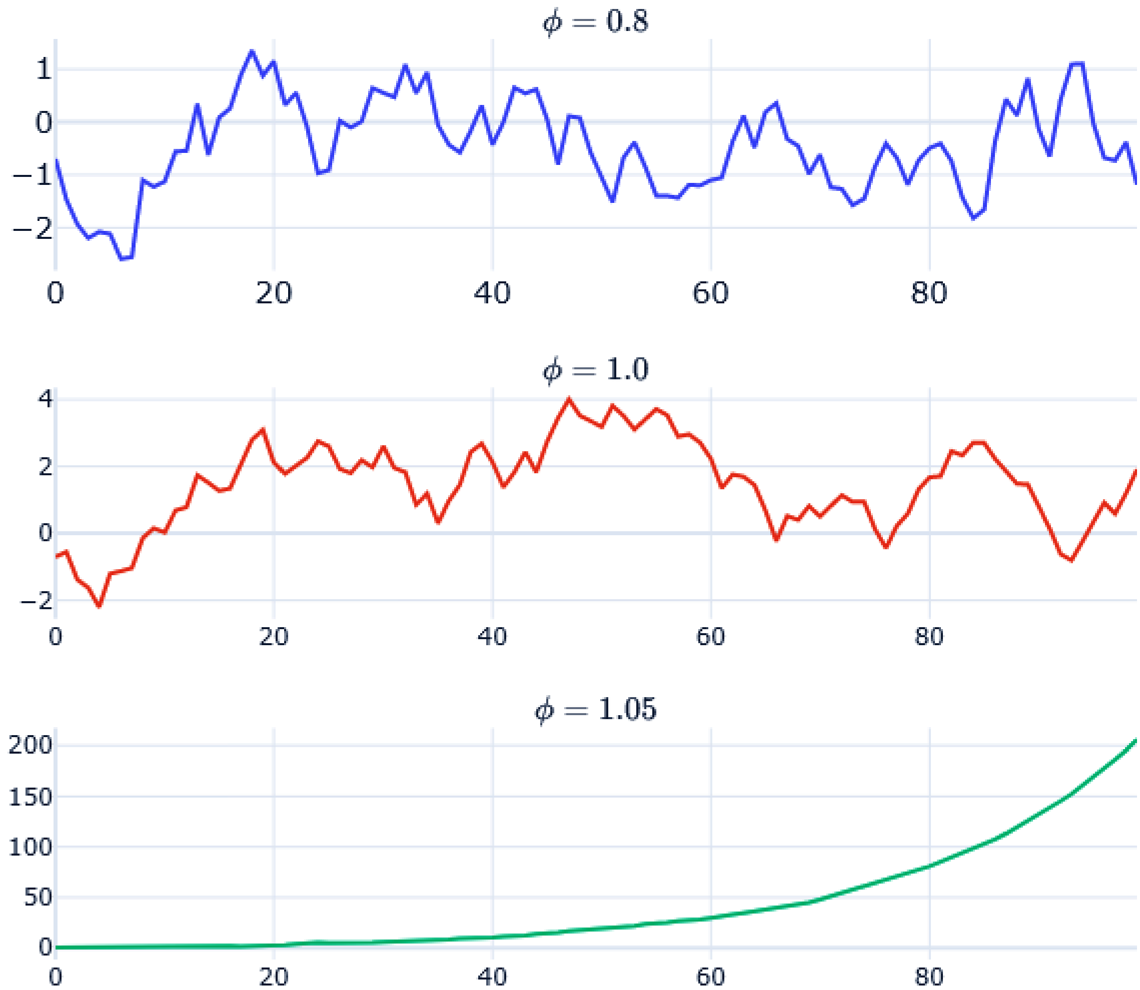

T-L
...
T-5
T-4
T-3
T-2
T-1
T
Current Timestamp
Exponentially Decaying Weights
Weighted Mean
Exponentially Weighted Moving Average

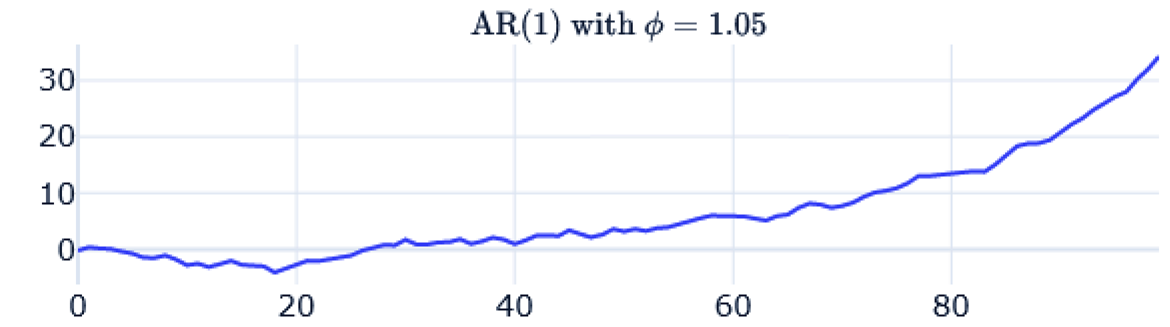


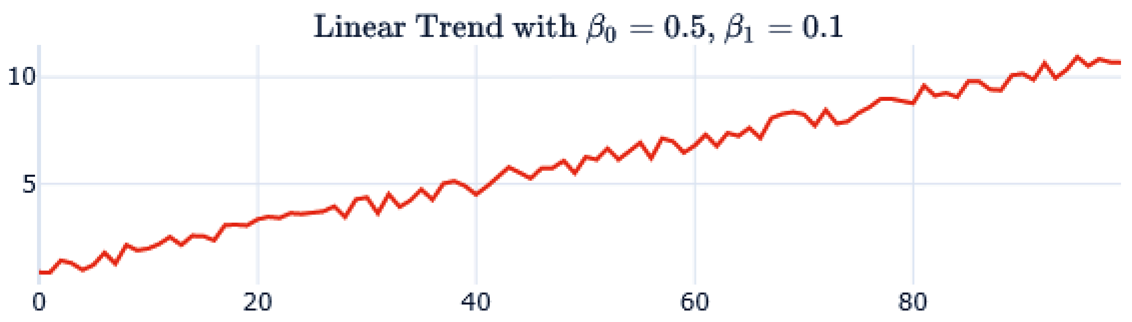
Step Function vs Continuous Function
Representation
month=calendar
month=fourier
Time

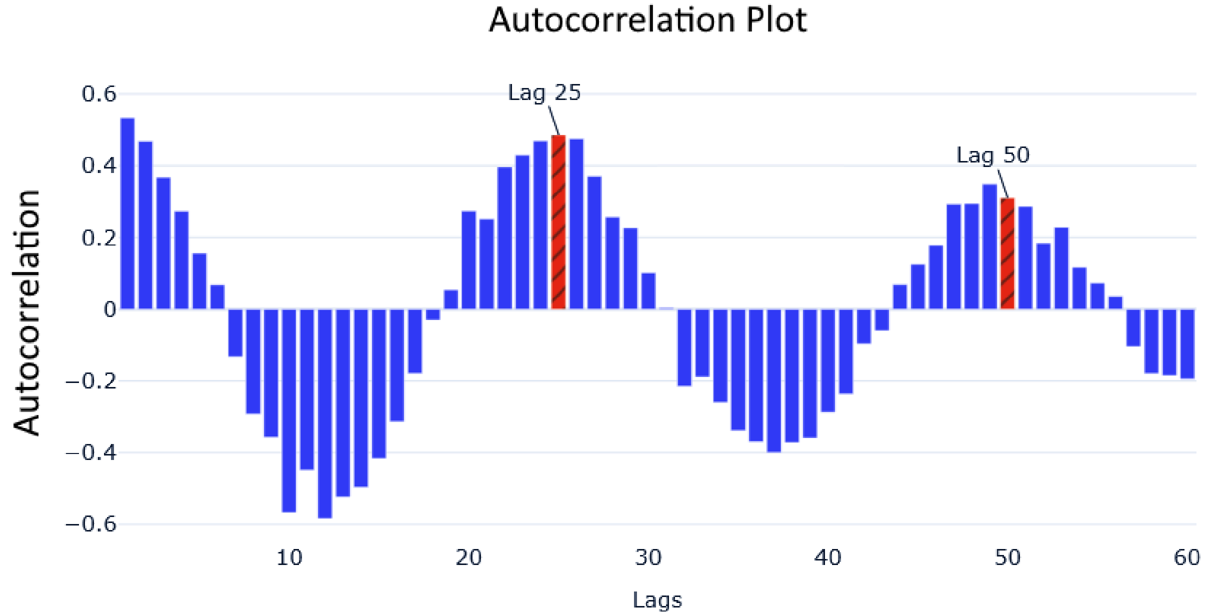
Chapter 7: Target Transformations for Time Series Forecasting

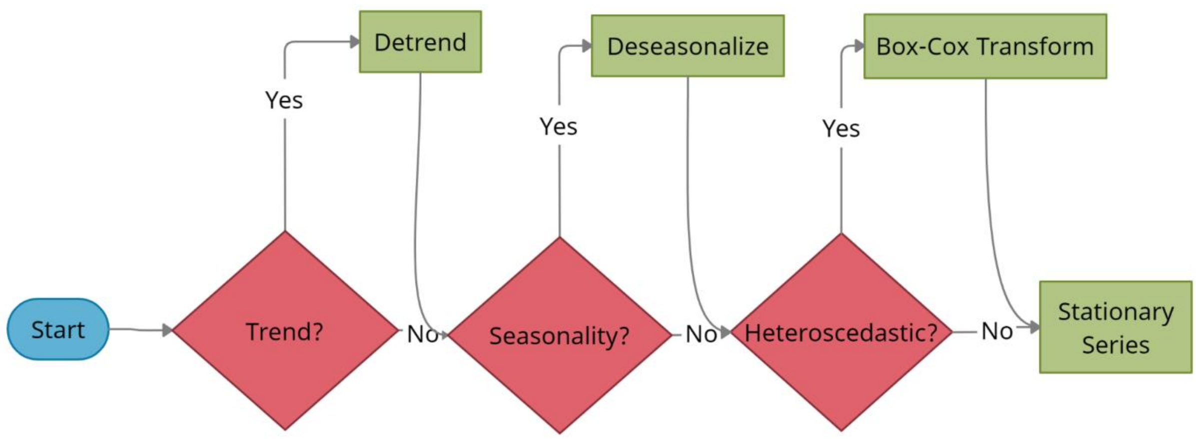





AR(1) with $\phi = 1.05$
30
20
10
0
0
20
40
60
80


Linear Trend with $\beta_0 = 0.5$, $\beta_1 = 0.1$
10
5
0
20
40
60
80

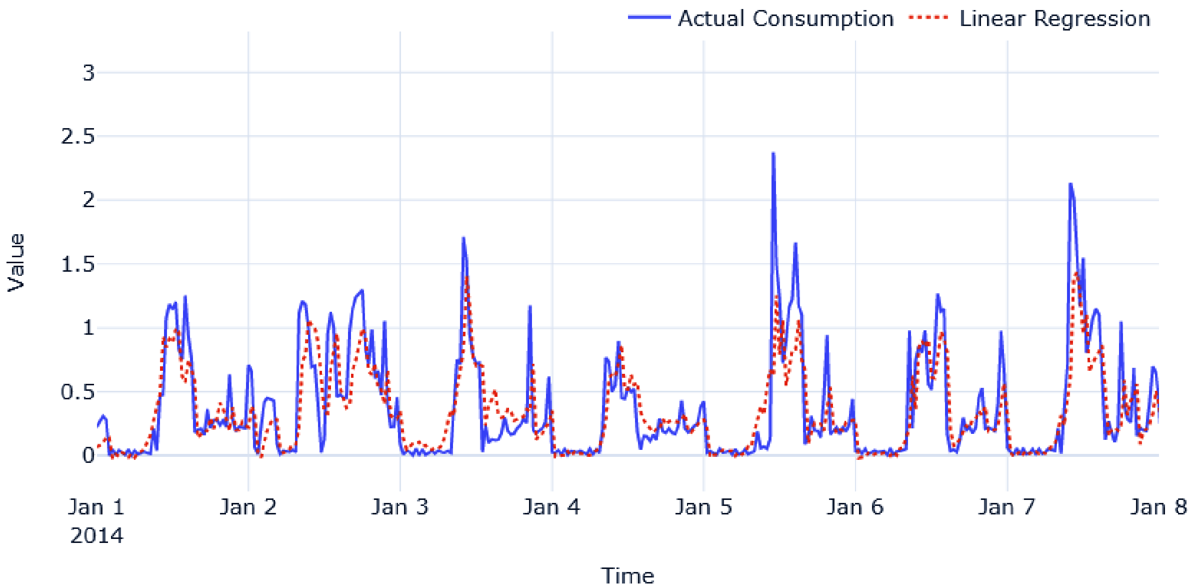

Autocorrelation Plot
Autocorrelation
Lag 25
Lag 50
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
10
20
30
40
50
60
Lags


Start
Trend?
Yes
Detrend
No
Seasonality?
Yes
Deseasonalize
No
Heteroscedastic?
Yes
Box-Cox Transform
No
Stationary Series

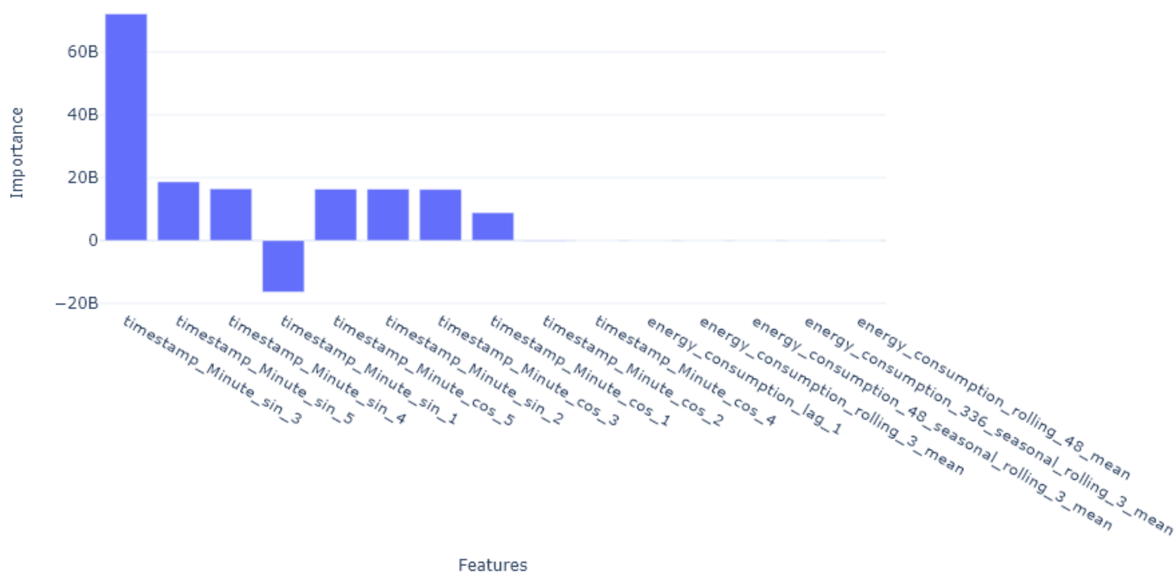
Chapter 8: Forecasting Time Series with Machine Learning Models

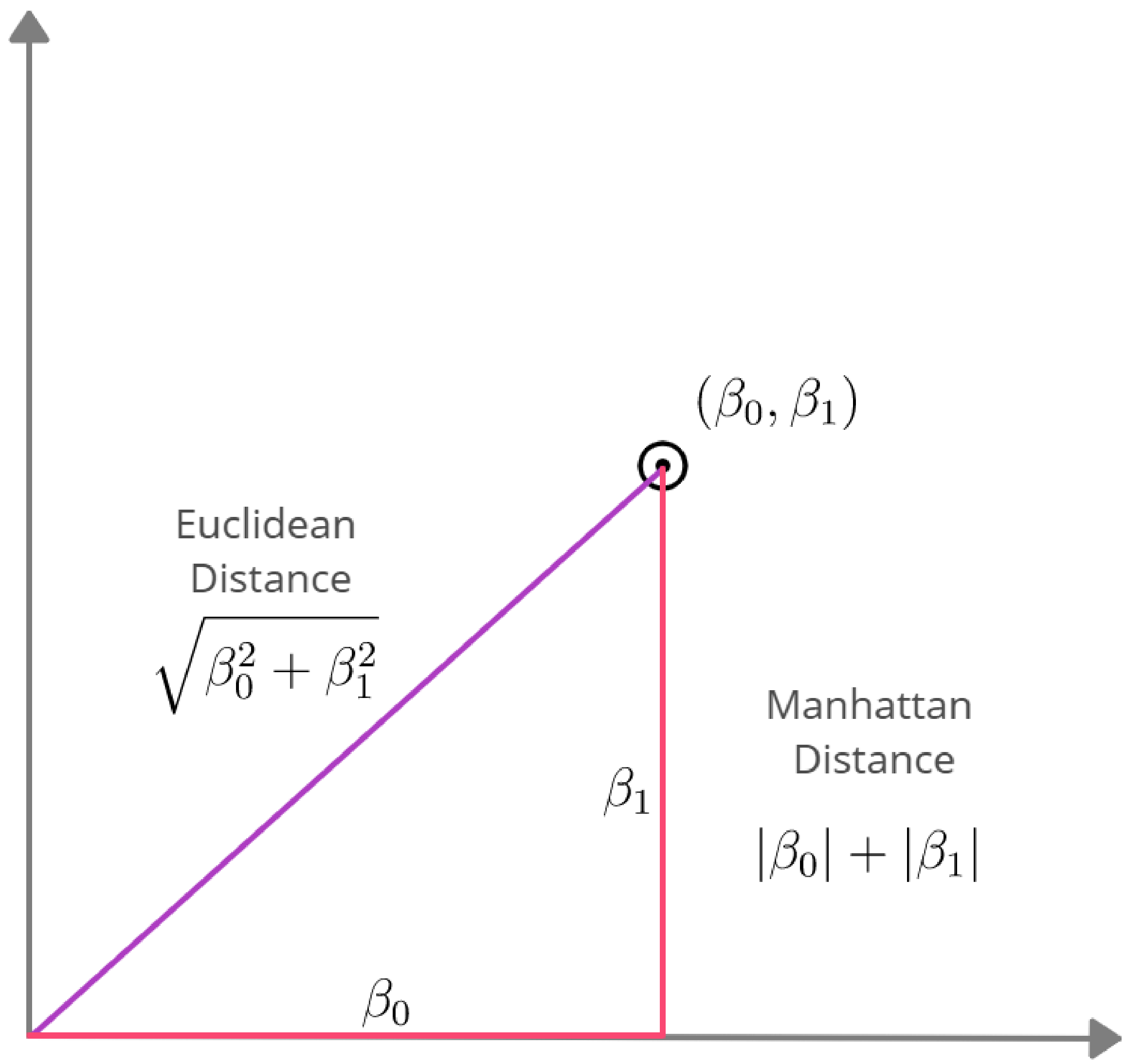
	MAE	MSE	meanMASE	Forecast Bias
Naive	0.086	0.045	1.050	0.02%
Seasonal Naive	0.122	0.072	1.487	4.07%

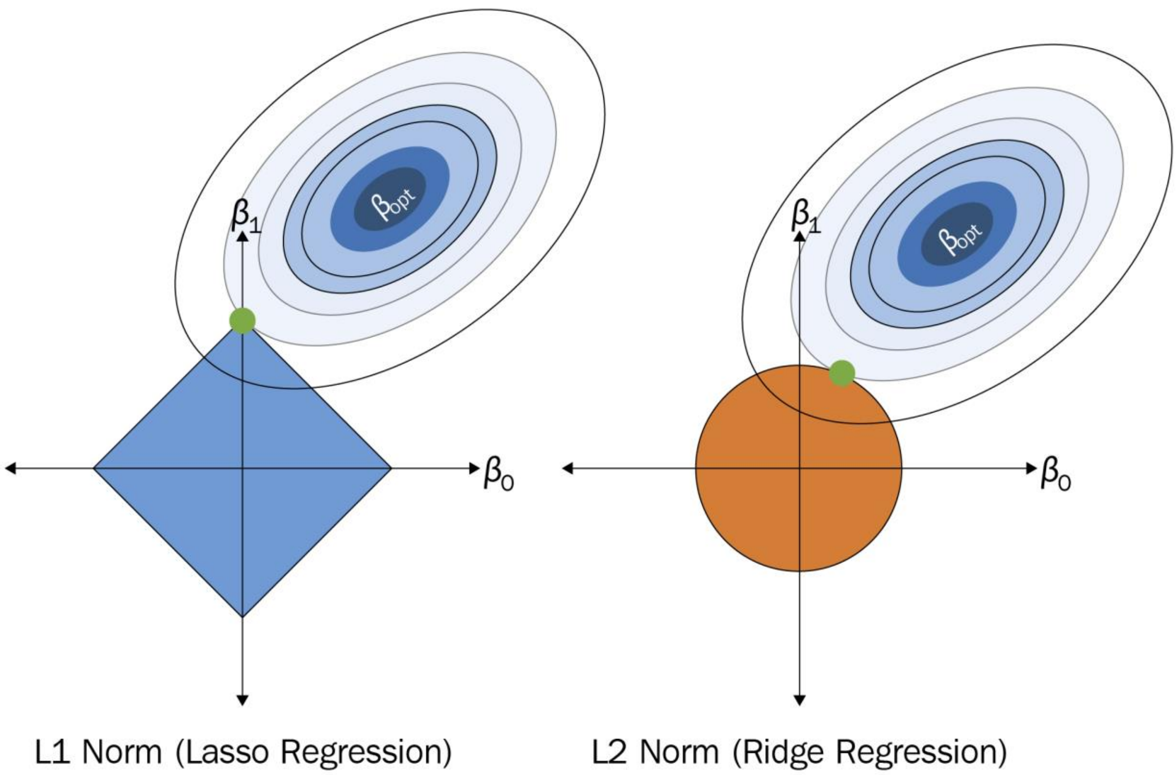
Linear Regression: MAE: 0.1595 | MSE: 0.0748 | MASE: 1.2431 | Bias: 6.1844

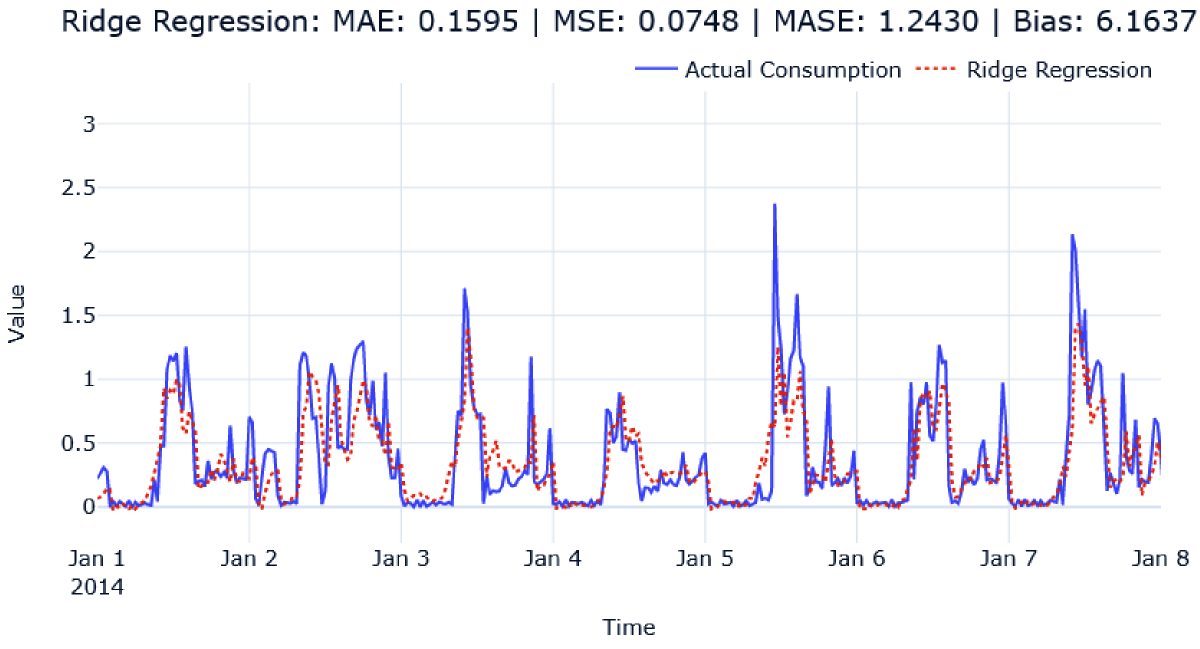


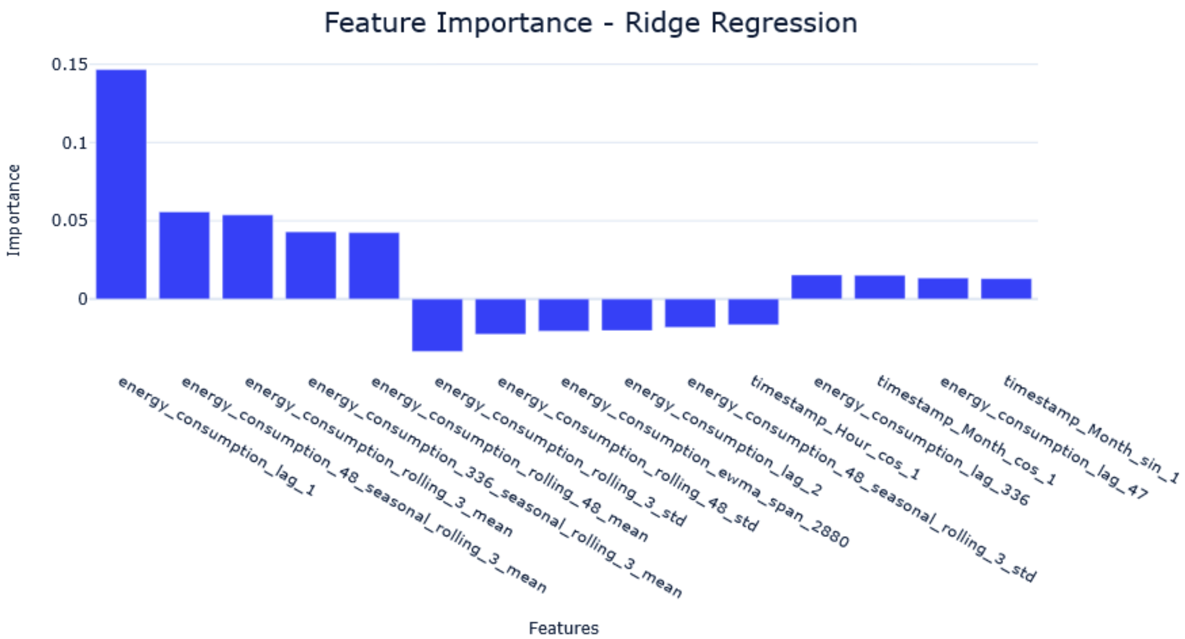
Feature Importance - Linear Regression



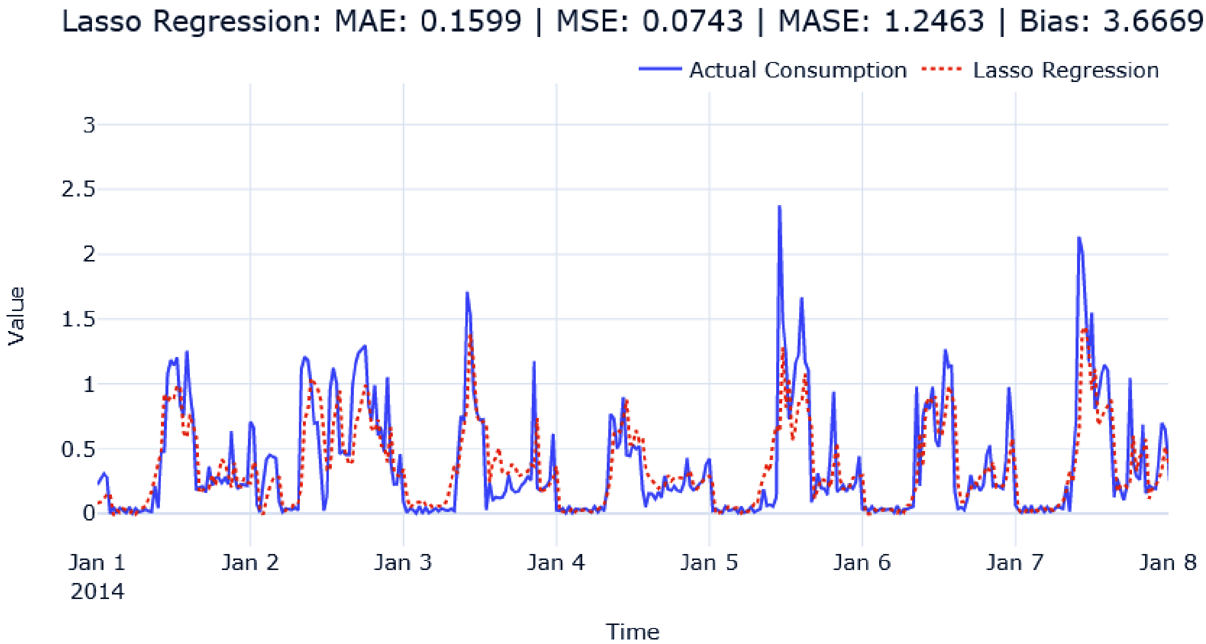

(β_0, β_1)
Euclidean Distance
$\sqrt{\beta_0^2 + \beta_1^2}$
Manhattan Distance
$|\beta_0| + |\beta_1|$
β_1
β_0


β_opt
β_1
β_0
β_opt
β_1
β_0
L1 Norm (Lasso Regression)
L2 Norm (Ridge Regression)

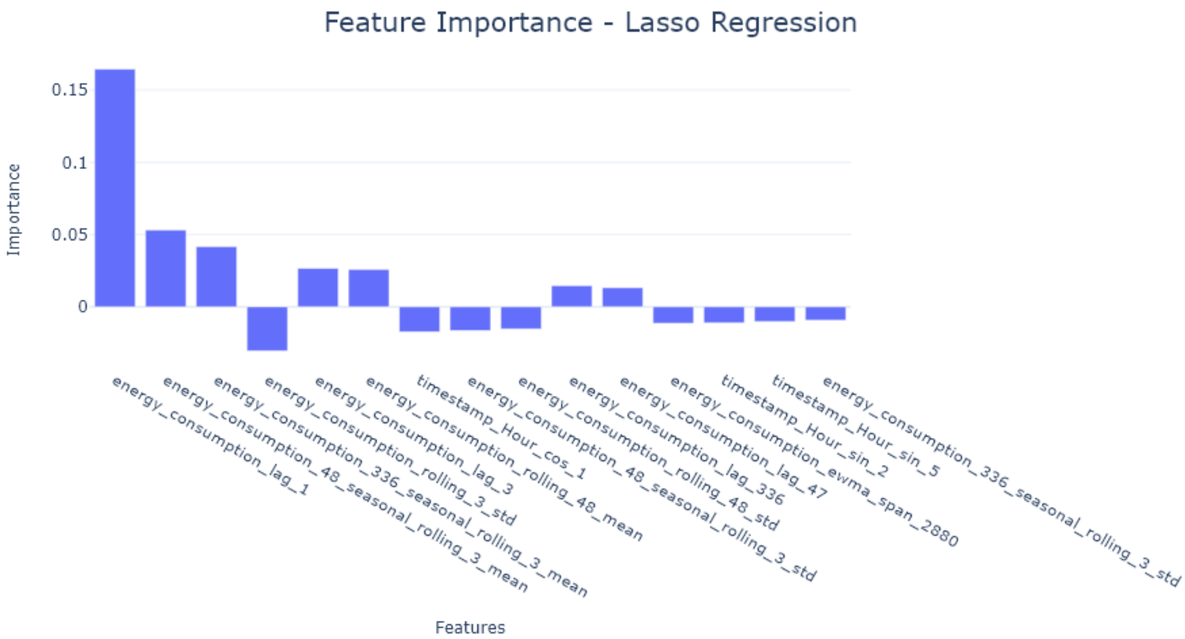

Ridge Regression: MAE: 0.1595 | MSE: 0.0748 | MASE: 1.2430 | Bias: 6.1637
Actual Consumption
Ridge Regression
Value
3
2.5
2
1.5
1
0.5
0
Jan 1
2014
Jan 2
Jan 3
Jan 4
Jan 5
Jan 6
Jan 7
Jan 8
Time

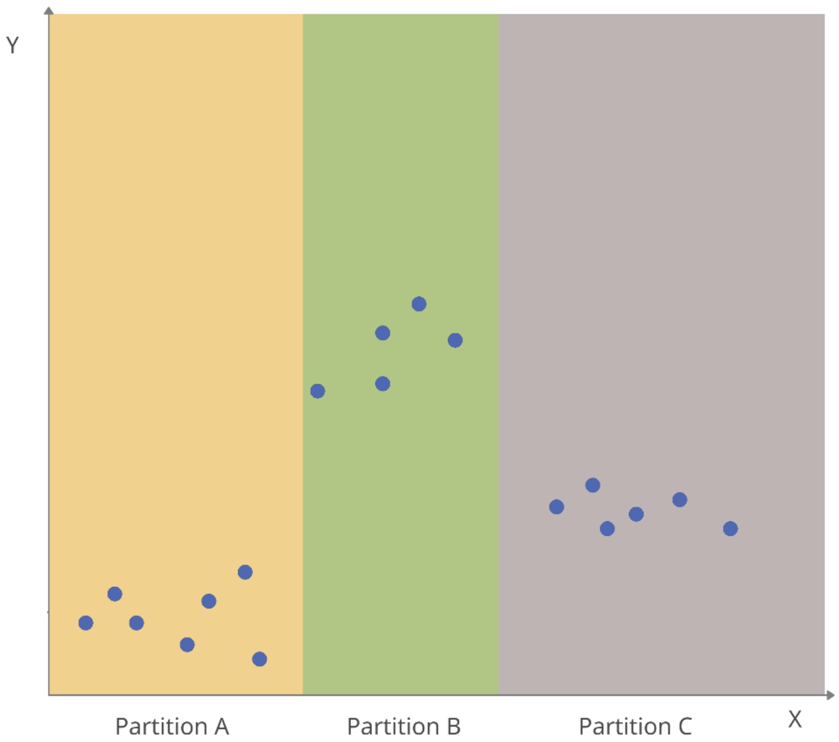


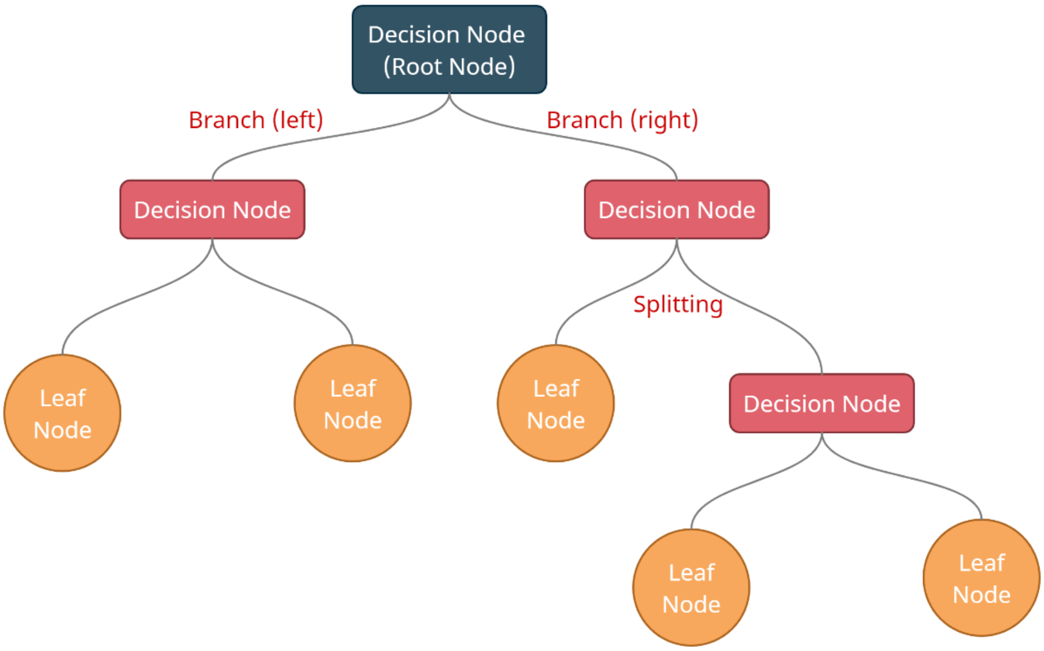
Feature Importance - Ridge Regression
Importance
0.15
0.1
0.05
0
energy_consumption_lag_1
energy_consumption_48_seasonal_rolling_3_mean
energy_consumption_rolling_3_mean
energy_consumption_336_seasonal_rolling_3_mean
energy_consumption_rolling_48_mean
energy_consumption_rolling_3_std
energy_consumption_rolling_48_std
energy_consumption_ewma_span_2880
energy_consumption_lag_2
energy_consumption_48_seasonal_rolling_3_std
timestamp_Hour_cos_1
energy_consumption_lag_336
timestamp_Month_cos_1
energy_consumption_lag_47
timestamp_Month_sin_1
Features

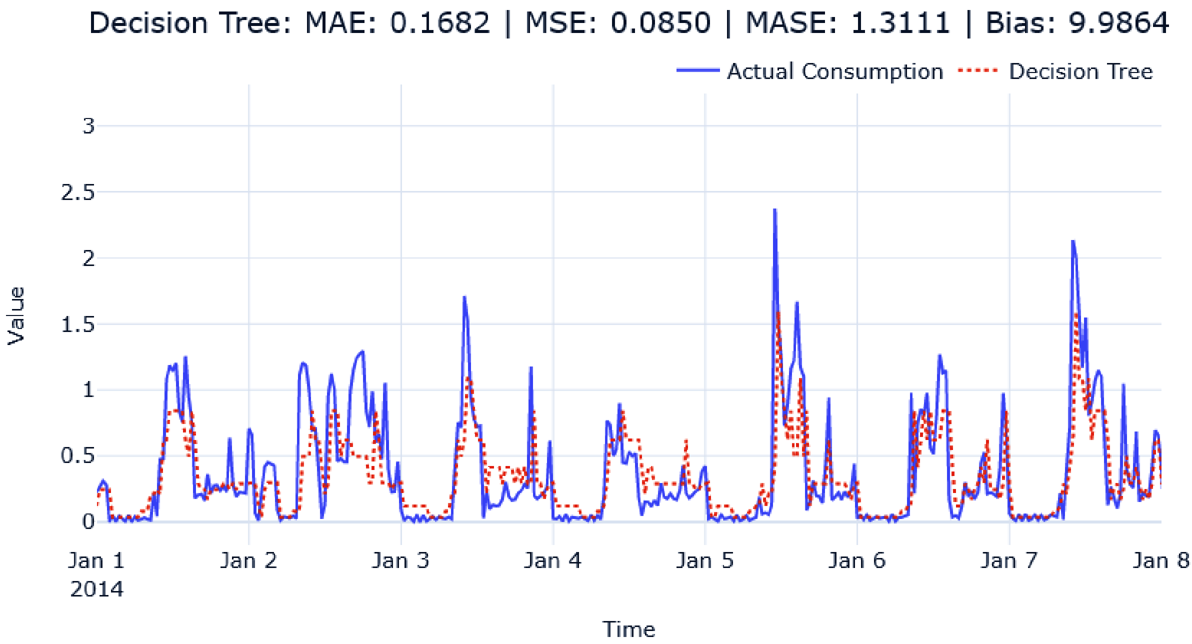


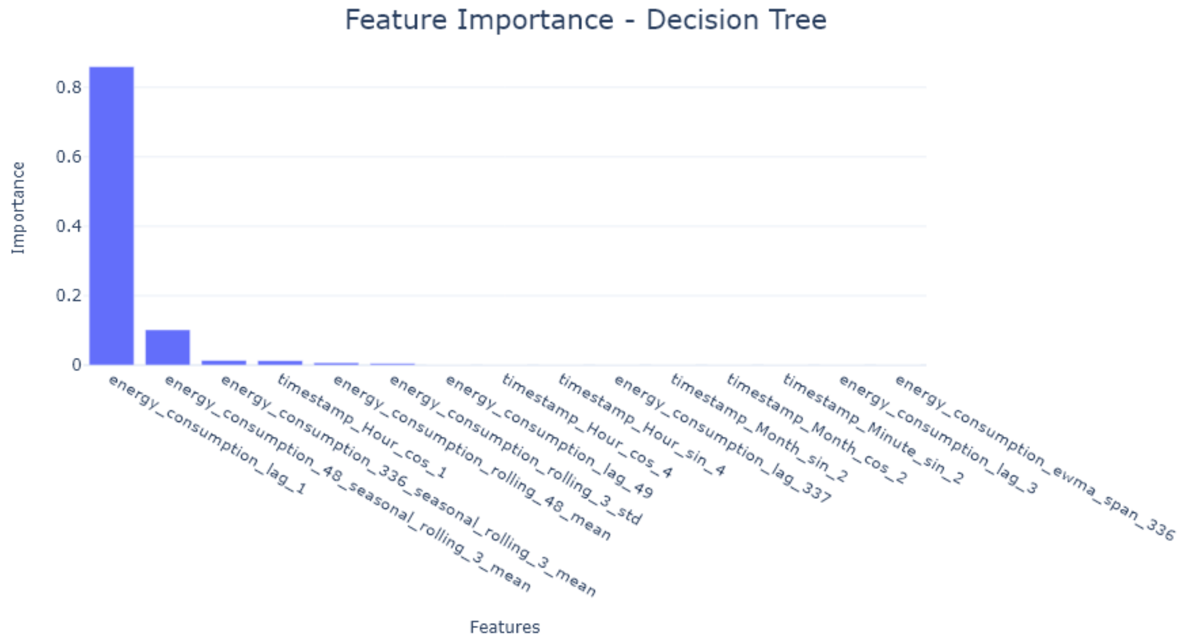
Lasso Regression: MAE: 0.1599 | MSE: 0.0743 | MASE: 1.2463 | Bias: 3.6669
Actual Consumption
Lasso Regression
Value
3
2.5
2
1.5
1
0.5
0
Jan 1
2014
Jan 2
Jan 3
Jan 4
Jan 5
Jan 6
Jan 7
Jan 8
Time


Feature Importance - Lasso Regression
Importance
0.15
0.1
0.05
0
energy_consumption_lag_1
energy_consumption_48_seasonal_rolling_3_mean
energy_consumption_336_seasonal_rolling_3_mean
energy_consumption_rolling_3_std
energy_consumption_lag_3
energy_consumption_rolling_48_mean
timestamp_Hour_cos_1
energy_consumption_48_seasonal_rolling_3_std
energy_consumption_rolling_48_std
energy_consumption_lag_336
energy_consumption_lag_47
energy_consumption_ewma_span_2880
timestamp_Hour_sin_2
timestamp_Hour_sin_5
energy_consumption_336_seasonal_rolling_3_std
Features

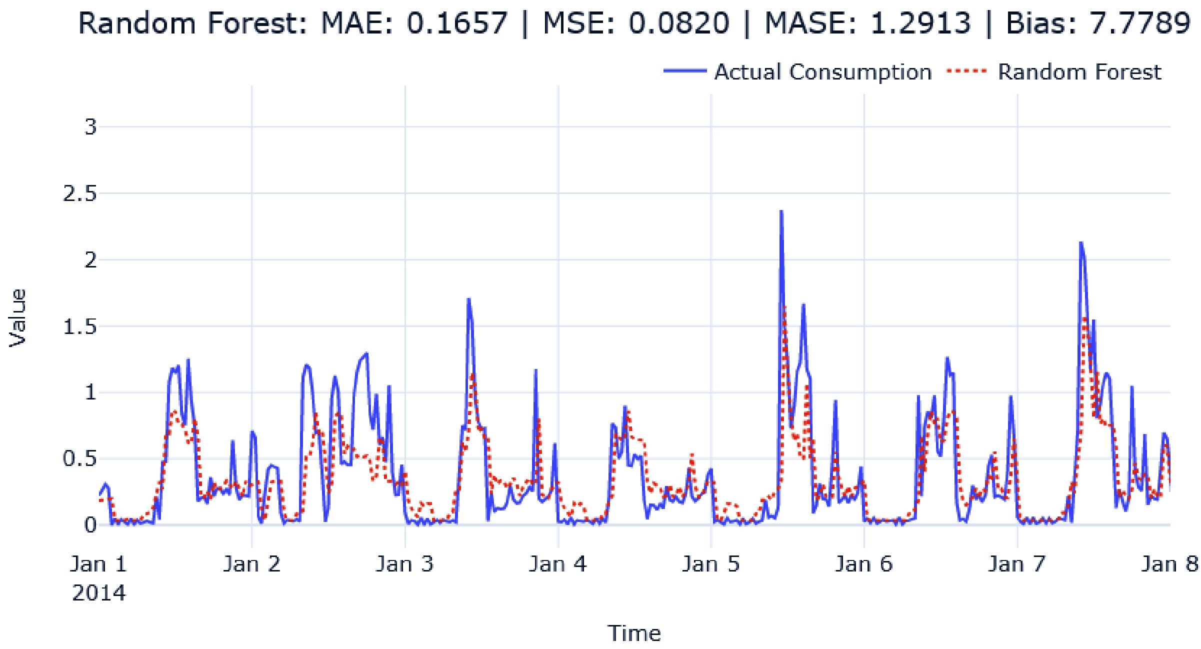

Y
Partition A
Partition B
Partition C
X


Decision Node
(Root Node)
Branch (left)
Branch (right)
Decision Node
Decision Node
Splitting
Leaf
Node
Leaf
Node
Leaf
Node
Decision Node
Leaf
Node
Leaf
Node


Decision Tree: MAE: 0.1682 | MSE: 0.0850 | MASE: 1.3111 | Bias: 9.9864
Actual Consumption
Decision Tree
Value
3
2.5
2
1.5
1
0.5
0
Jan 1
2014
Jan 2
Jan 3
Jan 4
Jan 5
Jan 6
Jan 7
Jan 8
Time

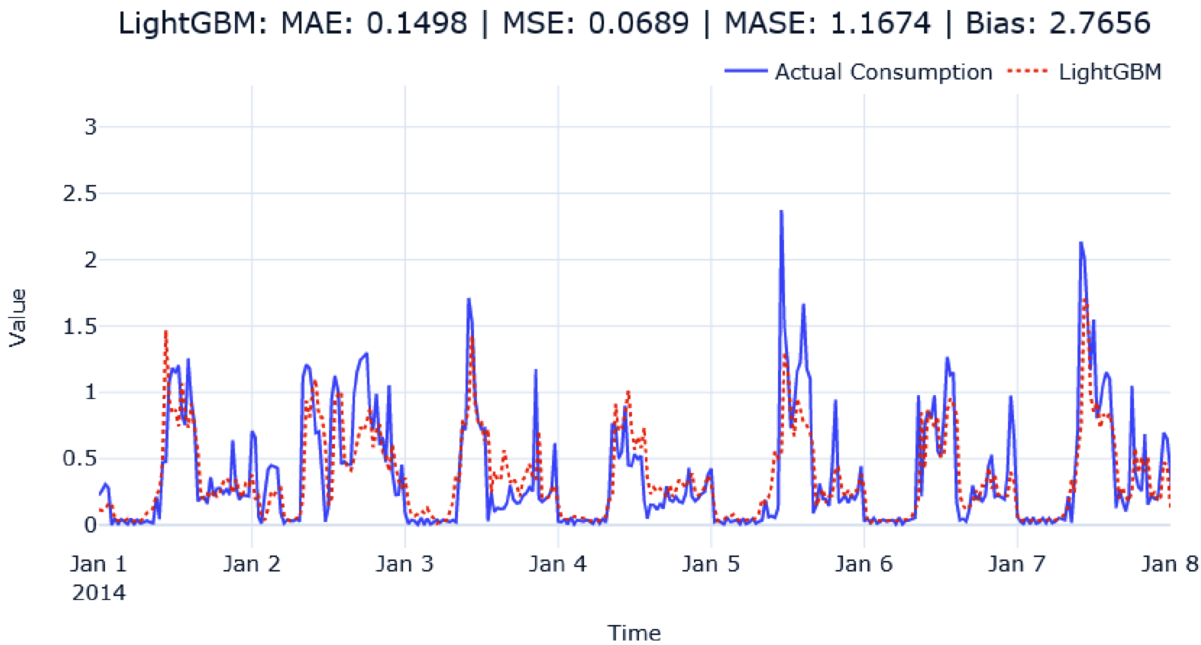


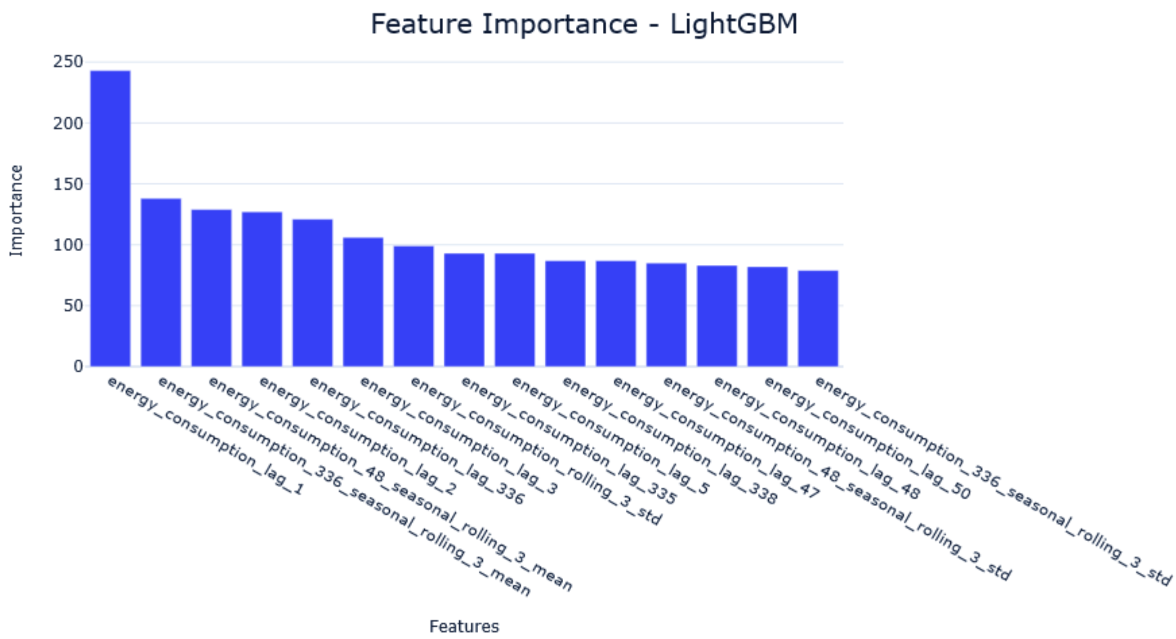
Feature Importance - Decision Tree
Importance
0.8
0.6
0.4
0.2
0
energy_consumption_lag_1
energy_consumption_48_seasonal_rolling_3_mean
energy_consumption_336_seasonal_rolling_3_mean
timestamp_Hour_cos_1
energy_consumption_rolling_48_mean
energy_consumption_rolling_3_std
energy_consumption_lag_49
timestamp_Hour_cos_4
timestamp_Hour_sin_4
energy_consumption_lag_337
timestamp_Month_sin_2
timestamp_Month_cos_2
timestamp_Minute_sin_2
energy_consumption_lag_3
energy_consumption_ewma_span_336
Features



Random Forest: MAE: 0.1657 | MSE: 0.0820 | MASE: 1.2913 | Bias: 7.7789
Actual Consumption
Random Forest
Value
3
2.5
2
1.5
1
0.5
0
Jan 1
2014
Jan 2
Jan 3
Jan 4
Jan 5
Jan 6
Jan 7
Jan 8
Time


Feature Importance - Random Forest
Importance
0.8
0.6
0.4
0.2
0
energy_consumption_lag_1
energy_consumption_48_seasonal_rolling_3_mean
energy_consumption_336_seasonal_rolling_3_mean
energy_consumption_lag_47
timestamp_Hour_cos_1
energy_consumption_rolling_6_std
energy_consumption_lag_336
timestamp_Hour_sin_1
energy_consumption_lag_46
energy_consumption_rolling_48_mean
energy_consumption_rolling_48_std
energy_consumption_rolling_3_std
energy_consumption_lag_335
energy_consumption_lag_48
energy_consumption_lag_334
Features


LightGBM: MAE: 0.1498 | MSE: 0.0689 | MASE: 1.1674 | Bias: 2.7656
Actual Consumption
LightGBM
Value
3
2.5
2
1.5
1
0.5
0
Jan 1 2014
Jan 2
Jan 3
Jan 4
Jan 5
Jan 6
Jan 7
Jan 8
Time



Algorithm	MAE	MSE	MASE	Forecast Bias	Time Elapsed
Naive	0.1753	0.1050	1.3664	0.03%	nan
Seasonal Naive	0.2377	0.1709	1.8521	4.80%	nan
Linear Regression	0.1595	0.0748	1.2431	6.18%	0.490227
Ridge Regression	0.1595	0.0748	1.2430	6.16%	0.425553
Lasso Regression	0.1599	0.0743	1.2463	3.67%	0.949244
Decision Tree	0.1682	0.0850	1.3111	9.99%	0.474296
Random Forest	0.1657	0.0820	1.2913	7.78%	26.353781
XGB Random Forest	0.1644	0.0818	1.2808	9.35%	1.786139
LightGBM	0.1498	0.0689	1.1674	2.77%	0.435123

Algorithm	MAE	MSE	meanMASE	Forecast Bias
Naive	0.0882	0.0450	1.1014	-0.00%
Seasonal Naive	0.1292	0.0777	1.6004	-1.00%
Lasso Regression	0.0802	0.0271	1.0052	-0.29%
XGB Random Forest	0.0808	0.0306	1.0177	-2.43%
LightGBM	0.0772	0.0275	0.9781	0.05%

Algorithm	MAE	MSE	meanMASE	Forecast Bias
Naive	0.086	0.045	1.050	0.02%
Seasonal Naive	0.122	0.072	1.487	4.07%
Lasso Regression	0.077	0.026	0.946	0.99%
XGB Random Forest	0.078	0.030	0.966	-0.18%
LightGBM	0.075	0.027	0.914	2.57%

Algorithm	MAE	MSE	meanMASE	Forecast Bias
Naive	0.088	0.045	1.101	-0.00%
Seasonal Naive	0.129	0.078	1.600	-1.00%
Lasso Regression	0.080	0.027	1.005	-0.29%
XGB Random Forest	0.081	0.031	1.018	-2.43%
LightGBM	0.077	0.028	0.978	0.05%
Lasso Regression_auto_stat	0.083	0.030	1.055	-3.50%
XGB Random Forest_auto_stat	0.086	0.033	1.098	-8.33%
LightGBM_auto_stat	0.079	0.029	1.002	-4.41%

Chapter 9: Ensembling and Stacking

LCLid	timestamp	FFT	Lasso Regression	Lasso Regression_auto_stat	LightGBM	LightGBM_auto_stat	Theta	XGB Random Forest	XGB Random Forest_auto_stat	energy_consumption
MAC000002	2014-01-01 00:00:00	0.255057	0.455158	0.418127	0.419273	0.414226	0.398836	0.39043	0.347244	0.496
	2014-01-01 00:30:00	0.234834	0.471182	0.422963	0.425602	0.403881	0.380464	0.39043	0.328894	0.427
	2014-01-01 01:00:00	0.207254	0.443879	0.418469	0.373510	0.378720	0.369484	0.39043	0.317926	0.469
	2014-01-01 01:30:00	0.175136	0.438553	0.382969	0.345316	0.333053	0.328537	0.39043	0.276990	0.362
	2014-01-01 02:00:00	0.144155	0.309471	0.295905	0.304884	0.277279	0.306195	0.36219	0.242762	0.452

Algorithm	MAE	MSE	meanMASE	Forecast Bias
LightGBM	0.0751	0.0271	0.9142	2.57%
best_fit	0.0740	0.0269	0.8971	0.14%

Algorithm	MAE	MSE	meanMASE	Forecast Bias
LightGBM	0.0751	0.0271	0.9142	2.57%
best_fit	0.0740	0.0269	0.8971	0.14%
median_ensemble	0.0767	0.0279	0.9304	-0.86%
average_ensemble	0.0828	0.0285	1.0159	1.58%

Algorithm	MAE	MSE	meanMASE	Forecast Bias
LightGBM	0.0751	0.0271	0.9142	2.57%
best_fit	0.0740	0.0269	0.8971	0.14%
median_ensemble	0.0767	0.0279	0.9304	-0.86%
average_ensemble	0.0828	0.0285	1.0159	1.58%
greedy_ensemble	0.0733	0.0251	0.8951	0.81%

Algorithm	MAE	MSE	meanMASE	Forecast Bias
LightGBM	0.0751	0.0271	0.9142	2.57%
best_fit	0.0740	0.0269	0.8971	0.14%
median_ensemble	0.0767	0.0279	0.9304	-0.86%
average_ensemble	0.0828	0.0285	1.0159	1.58%
greedy_ensemble	0.0733	0.0251	0.8951	0.81%
stochastic_hillclimb__ensemble	0.0751	0.0257	0.9206	1.18%

Algorithm	MAE	MSE	meanMASE	Forecast Bias
LightGBM	0.0751	0.0271	0.9142	2.57%
best_fit	0.0740	0.0269	0.8971	0.14%
median_ensemble	0.0767	0.0279	0.9304	-0.86%
average_ensemble	0.0828	0.0285	1.0159	1.58%
greedy_ensemble	0.0733	0.0251	0.8951	0.81%
stochastic_hillclimb__ensemble	0.0751	0.0257	0.9206	1.18%
simulated_annealing_ensemble	0.0735	0.0248	0.9041	0.26%

Forecast	Weights
LightGBM	0.4221
LightGBM_auto_stat	0.2991
Lasso Regression_auto_stat	0.1266
Lasso Regression	0.1012
XGB Random Forest	0.0510
FFT	0.0000
Theta	0.0000
XGB Random Forest_auto_stat	0.0000

Algorithm	MAE	MSE	meanMASE	Forecast Bias
LightGBM	0.0751	0.0271	0.9142	2.57%
best_fit	0.0740	0.0269	0.8971	0.14%
median_ensemble	0.0767	0.0279	0.9304	-0.86%
average_ensemble	0.0828	0.0285	1.0159	1.58%
greedy_ensemble	0.0733	0.0251	0.8951	0.81%
stochastic_hillclimb__ensemble	0.0751	0.0257	0.9206	1.18%
simulated_annealing_ensemble	0.0735	0.0248	0.9041	0.26%
optimal_combination_ensemble	0.0732	0.0248	0.8956	0.81%

Algorithm	MAE	MSE	meanMASE	Forecast Bias
LightGBM	0.0751	0.0271	0.9142	2.57%
best_fit	0.0740	0.0269	0.8971	0.14%
median_ensemble	0.0767	0.0279	0.9304	-0.86%
average_ensemble	0.0828	0.0285	1.0159	1.58%
greedy_ensemble	0.0733	0.0251	0.8951	0.81%
stochastic_hillclimb__ensemble	0.0751	0.0257	0.9206	1.18%
simulated_annealing_ensemble	0.0735	0.0248	0.9041	0.26%
optimal_combination_ensemble	0.0732	0.0248	0.8956	0.81%
linear_reg_blending	0.0755	0.0245	0.9260	4.35%
ridge_reg_blending	0.0737	0.0243	0.9082	1.84%
lasso_reg_blending	0.0736	0.0243	0.9068	1.94%
huber_reg_blending	0.0704	0.0246	0.8989	-6.42%

Chapter 10: Global Forecasting Models

Algorithm	MAE	MSE	meanMASE	Forecast Bias	Time Elapsed
LightGBM	0.077183	0.027510	0.978056	0.050231	NaN
GFM Baseline	0.079581	0.027326	1.013393	0.218127	28.718087

Acorn_grouped
Comfortable
Comfortable
Adversity
Affluent
Affluent

One-Hot Encoding

Comfortable	Adversity	Affluent
1	0	0
1	0	0
0	1	0
0	0	1
0	0	1

Algorithm	MAE	MSE	meanMASE	Forecast Bias	Time Elapsed
LightGBM	0.077183	0.027510	0.978056	0.050231	NaN
GFM Baseline	0.079581	0.027326	1.013393	0.218127	28.718087
GFM+Meta (CountEncoder)	0.079411	0.027233	1.011801	0.037475	68.020298

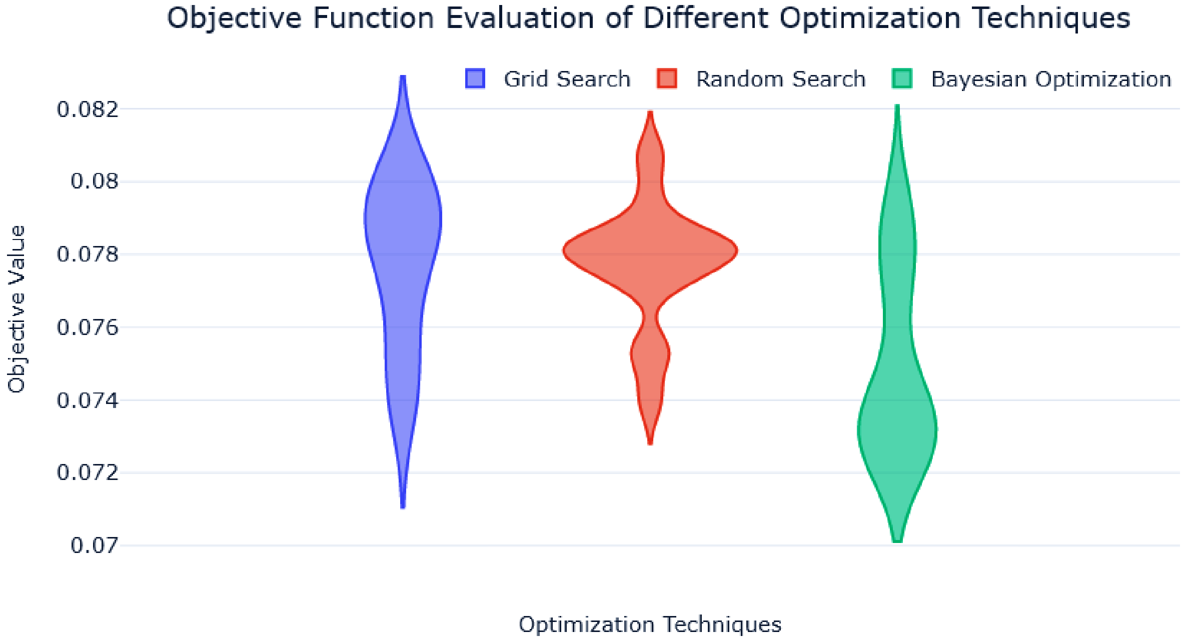
Acorn_grouped	energy_consumption
Comfortable	10
Comfortable	15
Adversity	5
Affluent	15
Affluent	20
Adversity	8
Adversity	7

Mean Encoding

Acorn_grouped	Encoded Value
Comfortable	(10 + 15)/2
Adversity	(5 + 8 + 7)/3
Affluent	(15 + 20)/2

Algorithm	MAE	MSE	meanMASE	Forecast Bias	Time Elapsed
LightGBM	0.077183	0.027510	0.978056	0.050231	NaN
GFM Baseline	0.079581	0.027326	1.013393	0.218127	28.718087
GFM+Meta (CountEncoder)	0.079411	0.027233	1.011801	0.037475	68.020298
GFM+Meta (TargetEncoder)	0.079537	0.027218	1.012400	0.335610	43.607325

Algorithm	MAE	MSE	meanMASE	Forecast Bias	Time Elapsed
LightGBM	0.077183	0.027510	0.978056	0.050231	NaN
GFM Baseline	0.079581	0.027326	1.013393	0.218127	28.718087
GFM+Meta (CountEncoder)	0.079411	0.027233	1.011801	0.037475	68.020298
GFM+Meta (TargetEncoder)	0.079537	0.027218	1.012400	0.335610	43.607325
GFM+Meta (NativeLGBM)	0.079209	0.027329	1.002630	-0.083755	30.316029

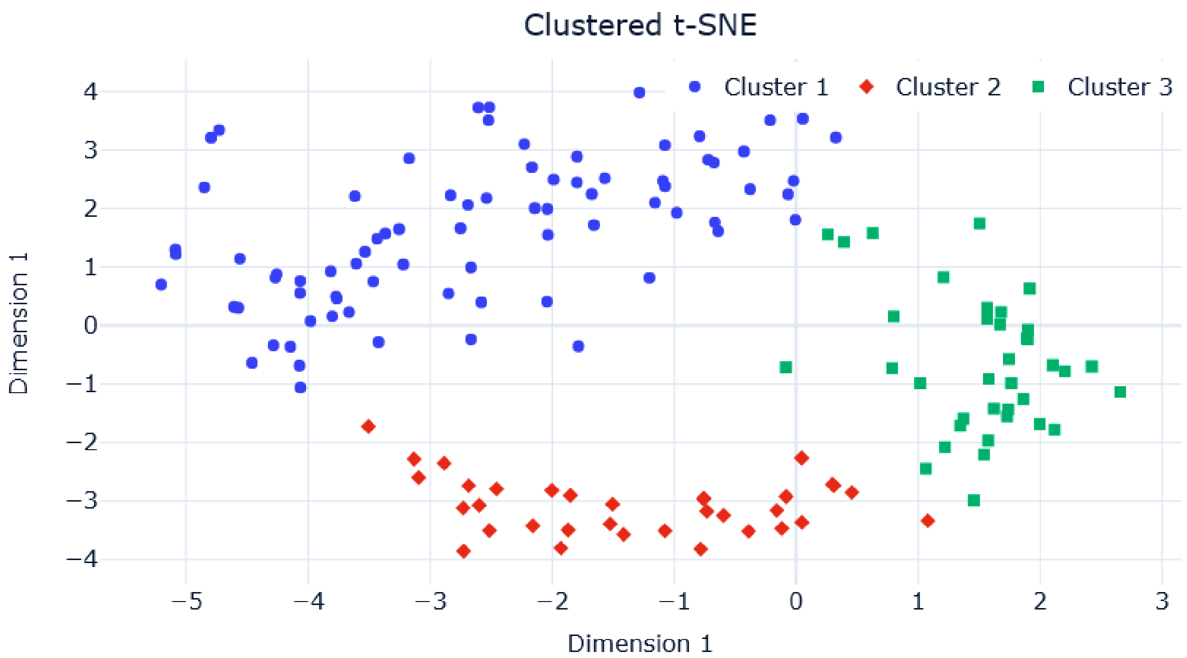


Algorithm	MAE	MSE	meanMASE	Forecast Bias	Time Elapsed
LightGBM	0.077183	0.027510	0.978056	0.050231	NaN
GFM Baseline	0.079581	0.027326	1.013393	0.218127	28.718087
GFM+Meta (CountEncoder)	0.079411	0.027233	1.011801	0.037475	68.020298
GFM+Meta (TargetEncoder)	0.079537	0.027218	1.012400	0.335610	43.607325
GFM+Meta (NativeLGBM)	0.079209	0.027329	1.002630	-0.083755	30.316029
Tuned GFM+Meta	0.072918	0.030641	0.900749	-12.412786	57.936451

Algorithm	MAE	MSE	meanMASE	Forecast Bias	Time Elapsed
LightGBM	0.077183	0.027510	0.978056	0.050231	NaN
GFM Baseline	0.079581	0.027326	1.013393	0.218127	28.718087
GFM+Meta (CountEncoder)	0.079411	0.027233	1.011801	0.037475	68.020298
GFM+Meta (TargetEncoder)	0.079537	0.027218	1.012400	0.335610	43.607325
GFM+Meta (NativeLGBM)	0.079209	0.027329	1.002630	-0.083755	30.316029
Tuned GFM+Meta	0.072918	0.030641	0.900749	-12.412786	57.936451
Tuned GFM+Meta+Random Part	0.072598	0.030681	0.898618	-12.361642	49.178089

Algorithm	MAE	MSE	meanMASE	Forecast Bias	Time Elapsed
LightGBM	0.077183	0.027510	0.978056	0.050231	NaN
GFM Baseline	0.079581	0.027326	1.013393	0.218127	28.718087
GFM+Meta (CountEncoder)	0.079411	0.027233	1.011801	0.037475	68.020298
GFM+Meta (TargetEncoder)	0.079537	0.027218	1.012400	0.335610	43.607325
GFM+Meta (NativeLGBM)	0.079209	0.027329	1.002630	-0.083755	30.316029
Tuned GFM+Meta	0.072918	0.030641	0.900749	-12.412786	57.936451
Tuned GFM+Meta+Random Part	0.072598	0.030681	0.898618	-12.361642	49.178089
Tuned GFM+Meta+ACORN Part	0.072567	0.030786	0.898071	-12.316822	52.118687

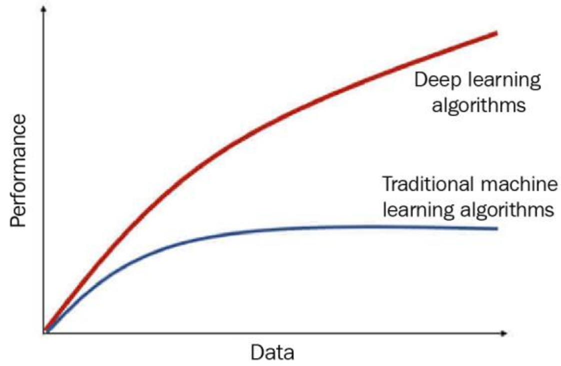
	0_ECDF_0	0_ECDF_1	0_ECDF_2	0_ECDF_3	0_ECDF_4	0_ECDF_5	0_ECDF_6	0_ECDF_7	0_ECDF_8	0_ECDF_9	...	0_Median diff
LCLid												
MAC000061	0.000028	0.000057	0.000085	0.000114	0.000142	0.000171	0.000199	0.000228	0.000256	0.000285	...	0.000
MAC000062	0.000028	0.000057	0.000085	0.000114	0.000142	0.000171	0.000199	0.000228	0.000256	0.000285	...	-0.003
MAC000066	0.000028	0.000057	0.000085	0.000114	0.000142	0.000171	0.000199	0.000228	0.000256	0.000285	...	0.000
MAC000086	0.000028	0.000057	0.000085	0.000114	0.000142	0.000171	0.000199	0.000228	0.000256	0.000285	...	-0.002
MAC000126	0.000028	0.000057	0.000085	0.000114	0.000142	0.000171	0.000199	0.000228	0.000256	0.000285	...	-0.002

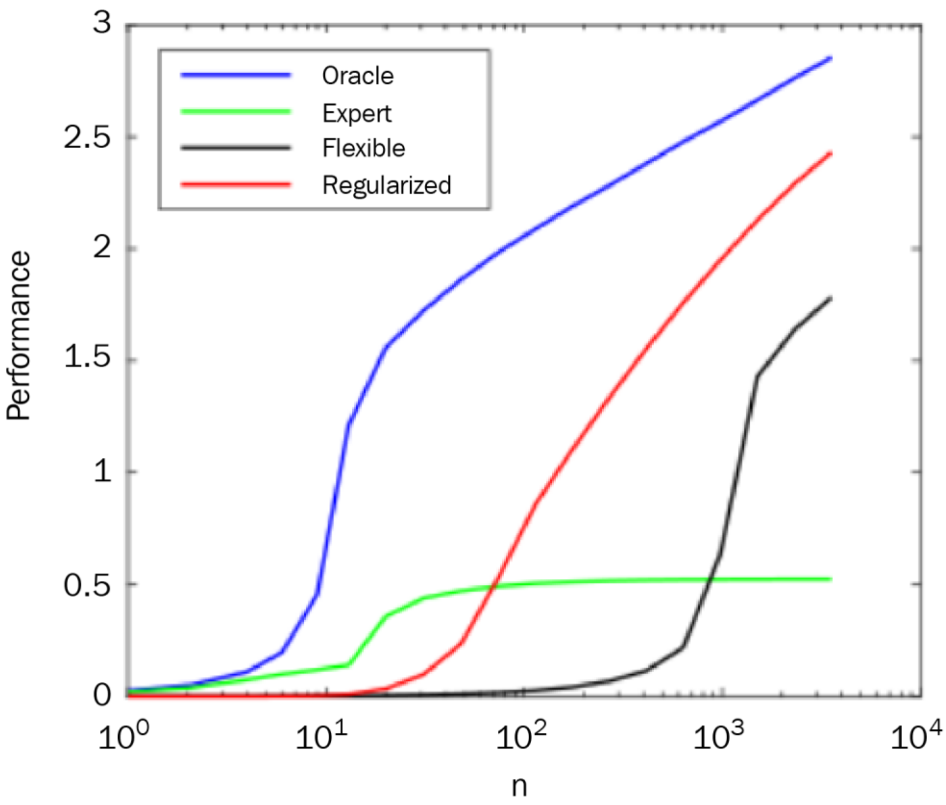


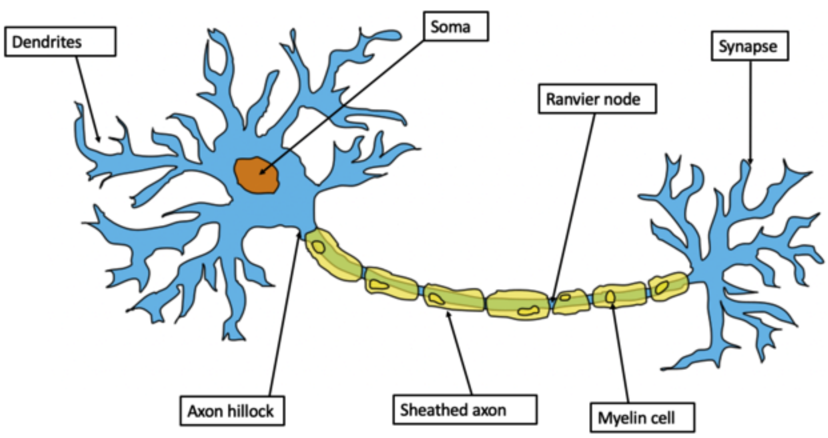
	Algorithm	MAE	MSE	meanMASE	Forecast Bias	Time Elapsed
0	LightGBM	0.077183	0.027510	0.978056	0.050231	NaN
1	GFM Baseline	0.079581	0.027326	1.013393	0.218127	28.718087
2	GFM+Meta (CountEncoder)	0.079411	0.027233	1.011801	0.037475	68.020298
3	GFM+Meta (TargetEncoder)	0.079537	0.027218	1.012400	0.335610	43.607325
4	GFM+Meta (NativeLGBM)	0.079209	0.027329	1.002630	-0.083755	30.316029
5	Tuned GFM+Meta	0.072918	0.030641	0.900749	-12.412786	57.936451
6	Tuned GFM+Meta+Random Part	0.072598	0.030681	0.898618	-12.361642	49.178089
7	Tuned GFM+Meta+ACORN Part	0.072567	0.030786	0.898071	-12.316822	52.118687
8	Tuned GFM+Meta+Clustered Part	0.072347	0.029976	0.905182	-12.521149	66.373510

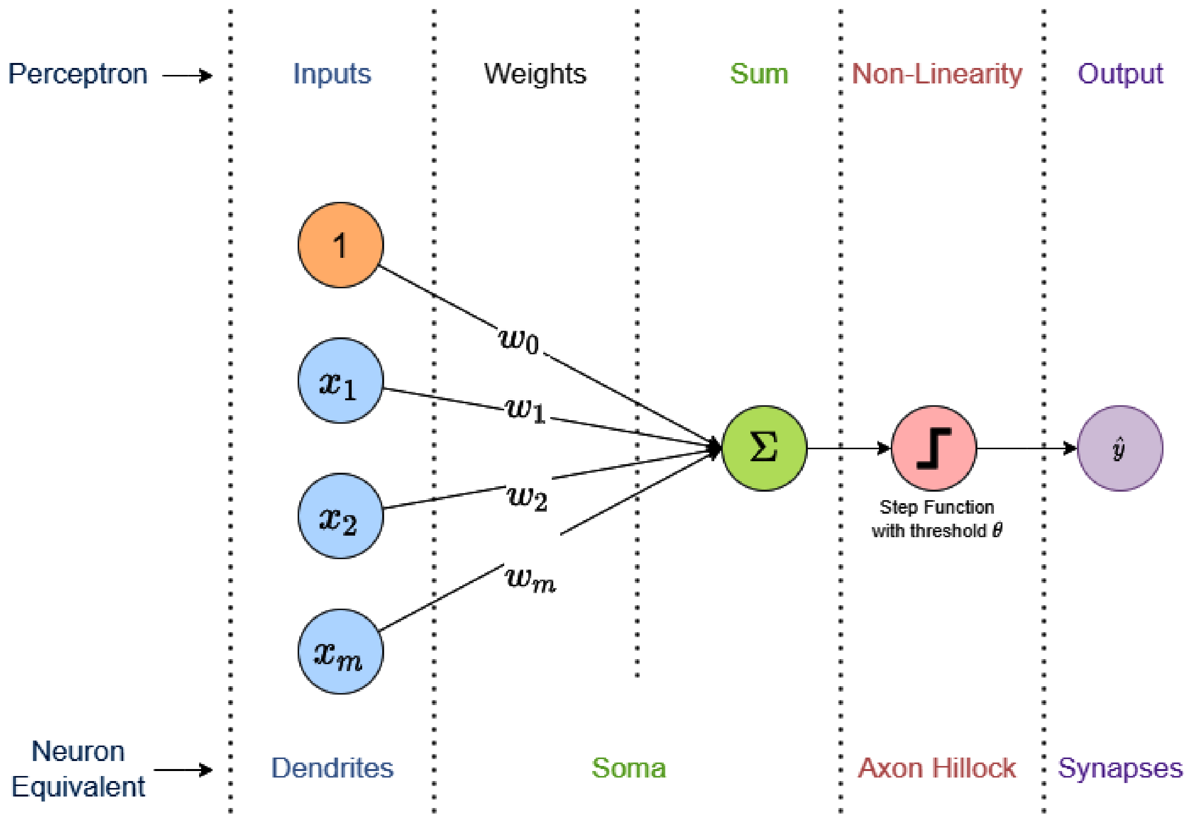
Algorithm	MAE	MSE	meanMASE	Forecast Bias	Time Elapsed
LightGBM	0.0751	0.0271	0.9142	2.57%	nan
GFM Baseline	0.0773	0.0280	0.9586	0.71%	38.147337
GFM+Meta (CountEncoder)	0.0772	0.0276	0.9600	0.69%	71.797225
GFM+Meta (TargetEncoder)	0.0773	0.0276	0.9612	0.99%	48.736455
GFM+Meta (NativeLGBM)	0.0770	0.0279	0.9483	0.84%	32.467879
Tuned GFM+Meta	0.0700	0.0310	0.8384	-12.38%	62.091281
Tuned GFM+Meta+Random Part	0.0706	0.0339	0.8405	-12.96%	55.361825
Tuned GFM+Meta+ACORN Part	0.0696	0.0305	0.8342	-12.43%	52.066995
Tuned GFM+Meta+Clustered Part	0.0685	0.0285	0.8282	-11.94%	57.231065

Chapter 11: Introduction to Deep Learning









Linear Combination of Inputs

Output

Bias

$$\hat{y} = g\left(w_0 + \sum_{i=1}^{m} x_i w_i\right)$$

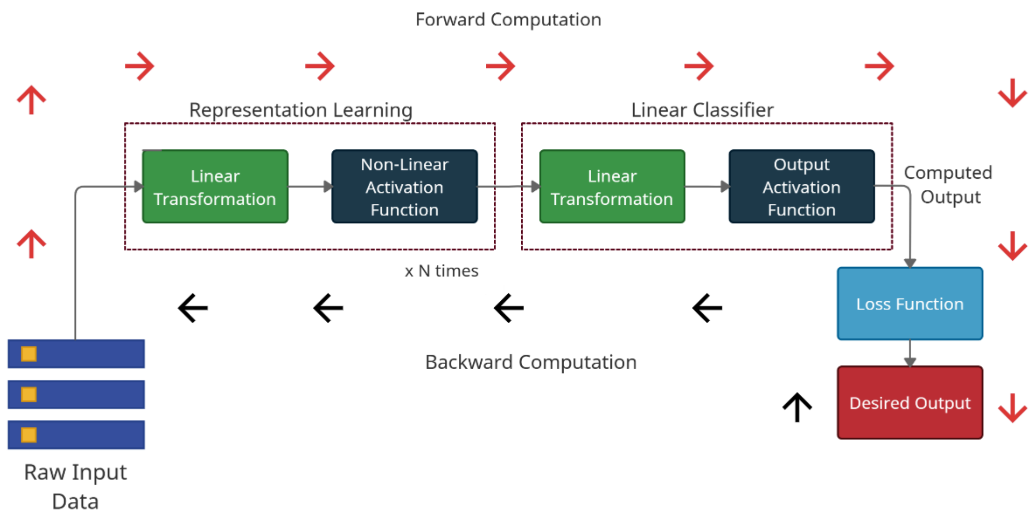
Non-Linearity

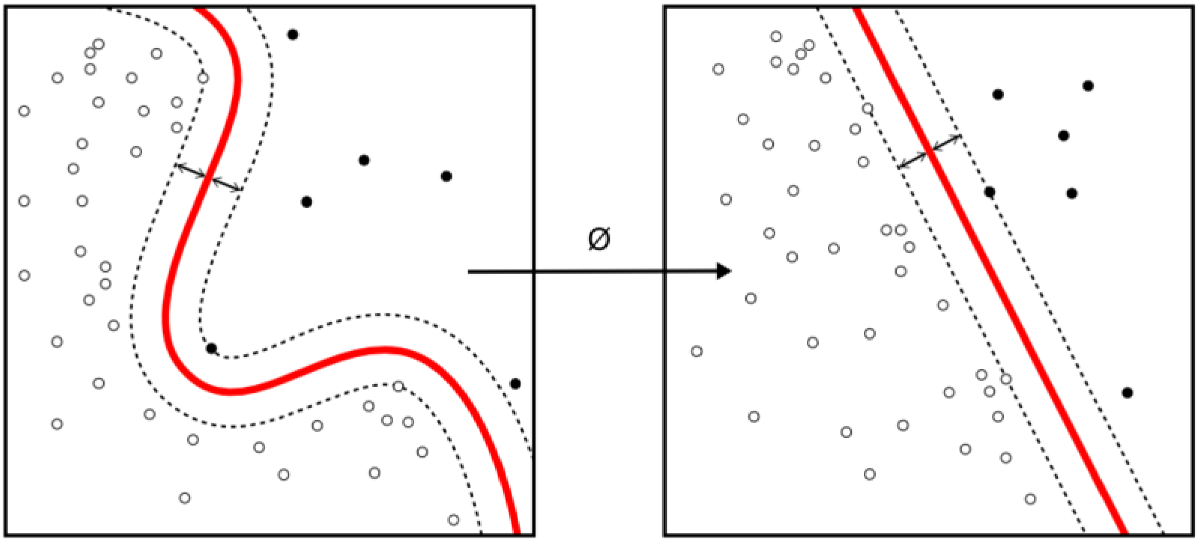
Vector Form →

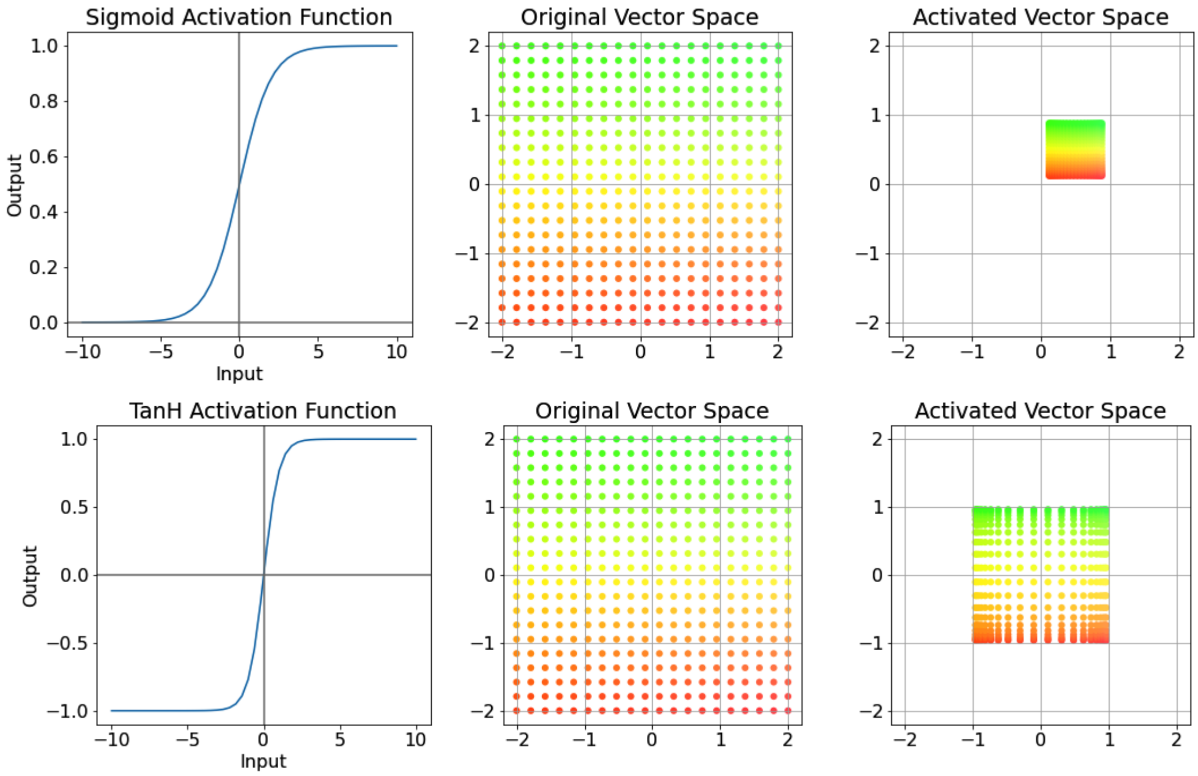
$$\hat{y} = g\left(\mathbb{X}_T \mathbb{W}\right)$$

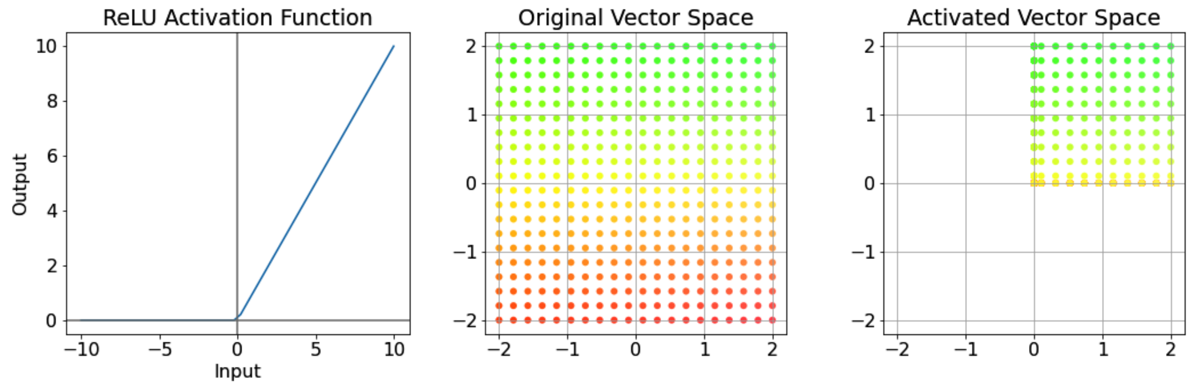
where

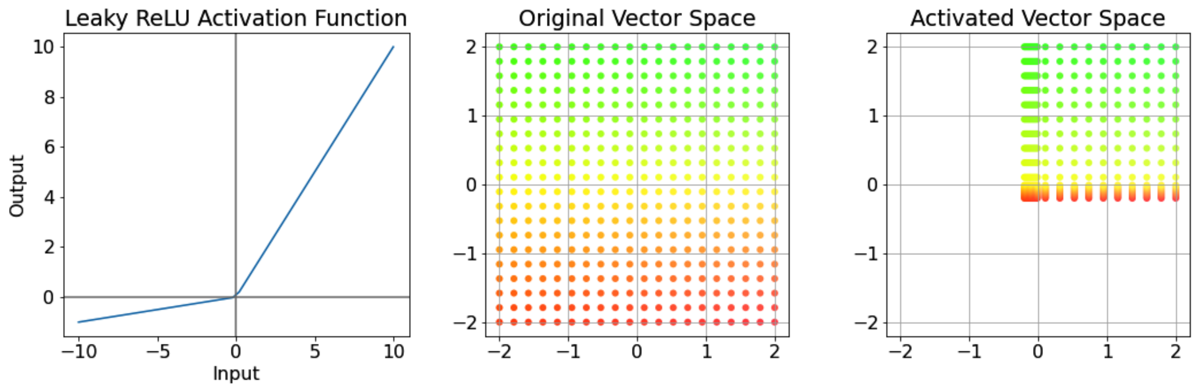
$$\mathbb{X} = \begin{bmatrix} 1 \\ x_1 \\ \dots \\ x_m \end{bmatrix} \text{ and } \mathbb{W} = \begin{bmatrix} w_0 \\ w_1 \\ \dots \\ w_m \end{bmatrix}$$

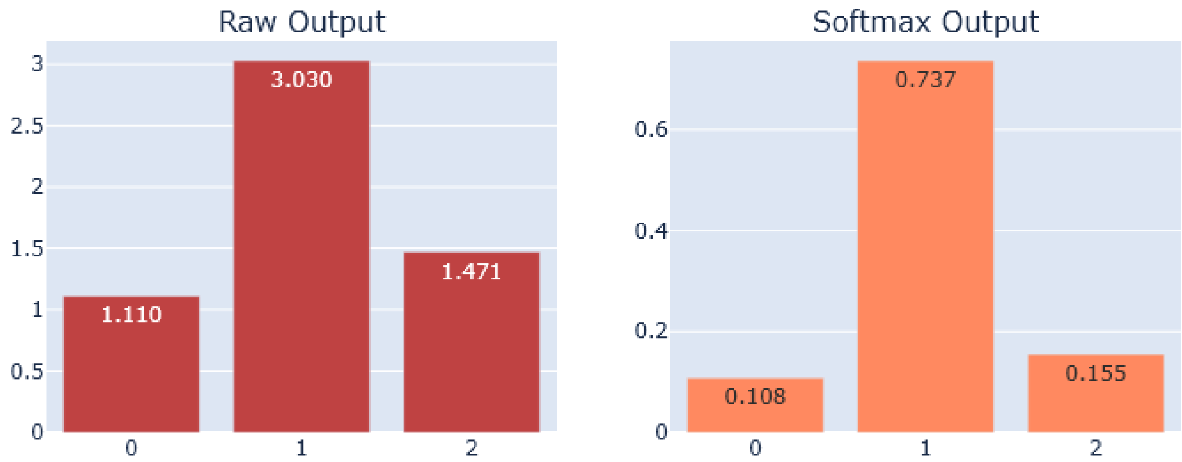

Forward Computation
Representation Learning
Linear Classifier
Linear Transformation
Non-Linear Activation Function
Linear Transformation
Output Activation Function
Computed Output
x N times
Loss Function
Backward Computation
Desired Output
Raw Input Data


Ø

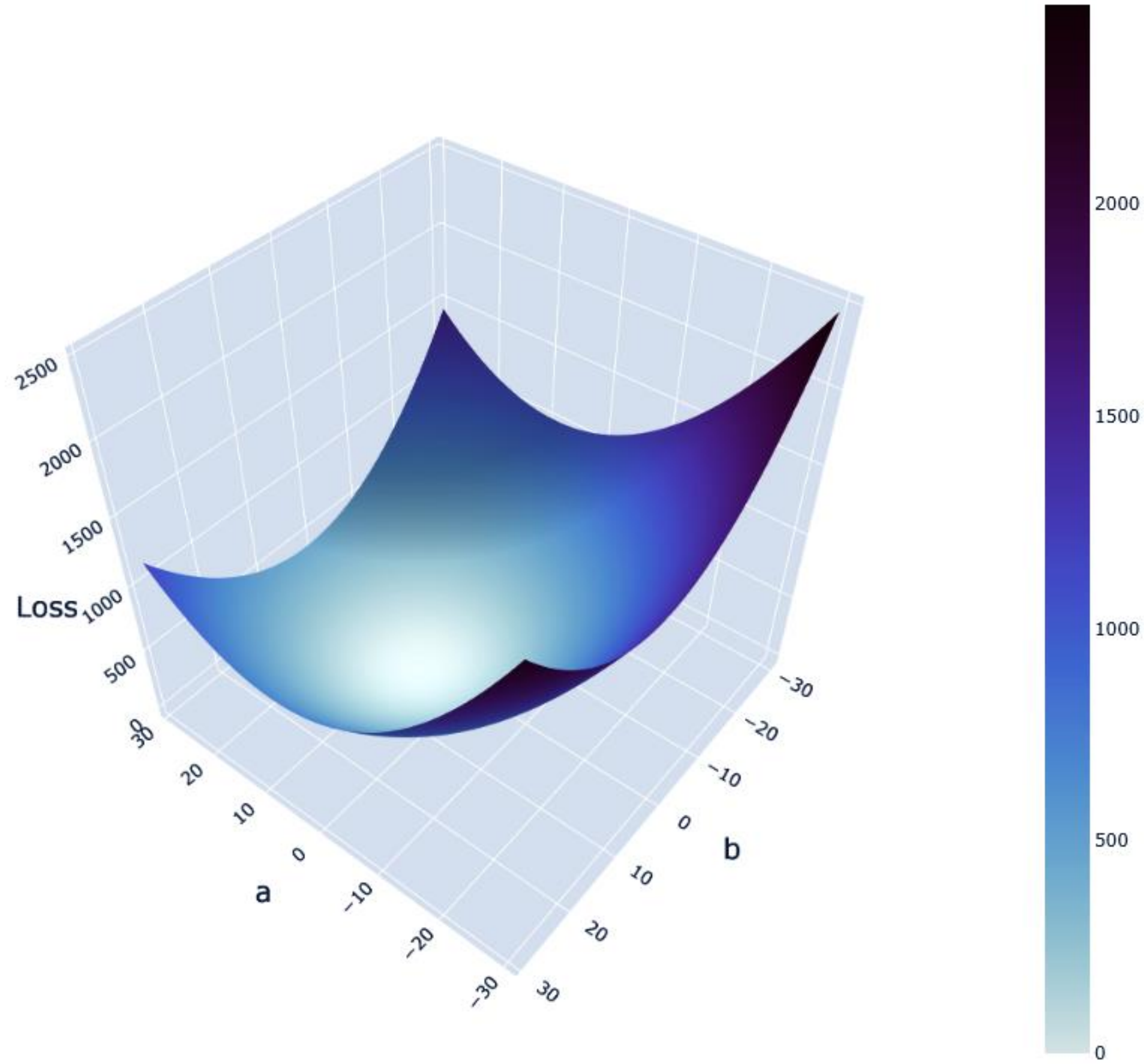

Sigmoid Activation Function
Output
Input
Original Vector Space
Activated Vector Space
TanH Activation Function
Output
Input
Original Vector Space
Activated Vector Space

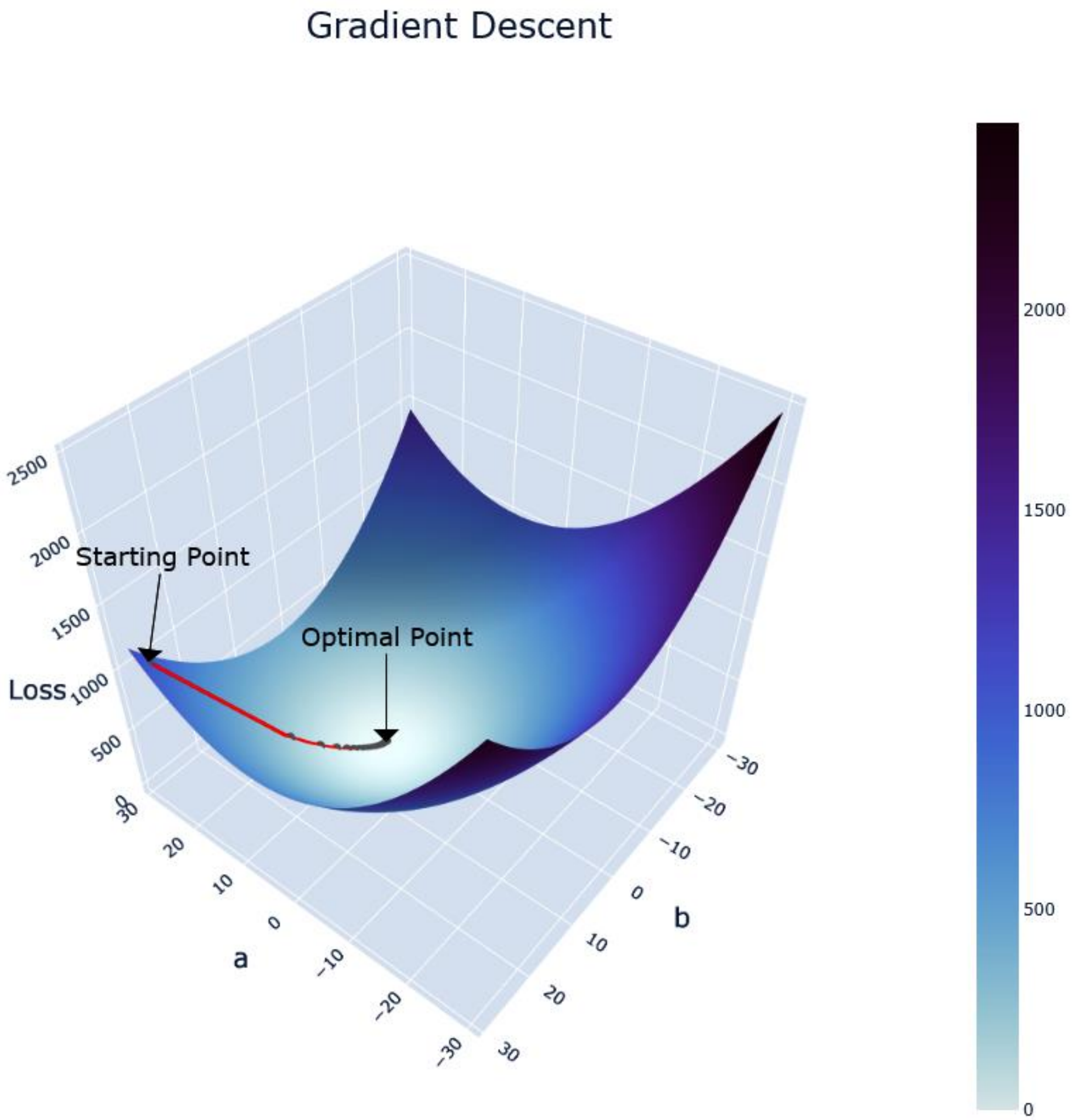

ReLU Activation Function
Output
Input
Original Vector Space
Activated Vector Space


Leaky ReLU Activation Function
Output
Input
Original Vector Space
Activated Vector Space


Raw Output
1.110
3.030
1.471
Softmax Output
0.108
0.737
0.155

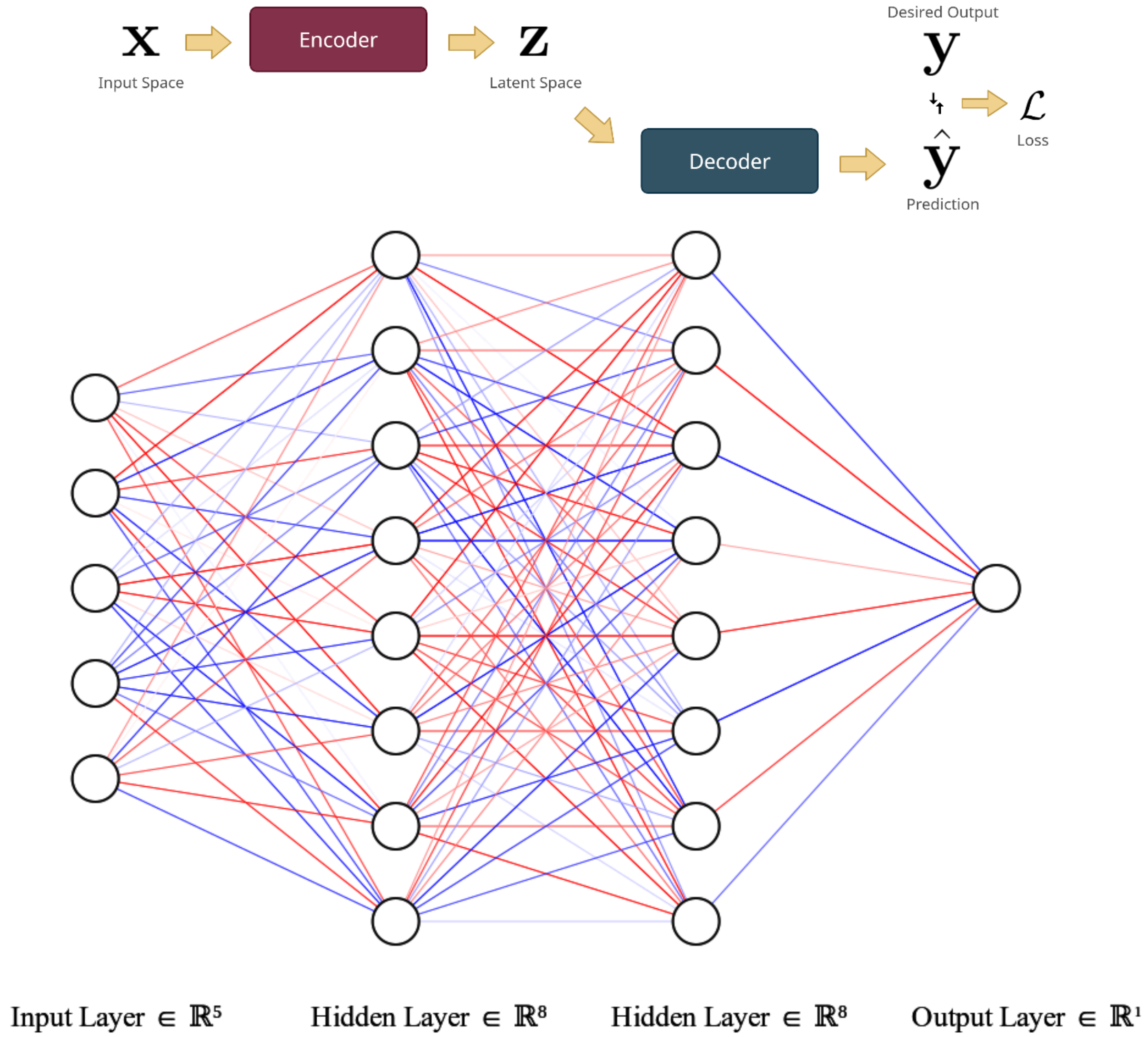
Loss Surface for $\mathcal{L}(a, b) = (a - 8)^2 + (b - 2)^2$

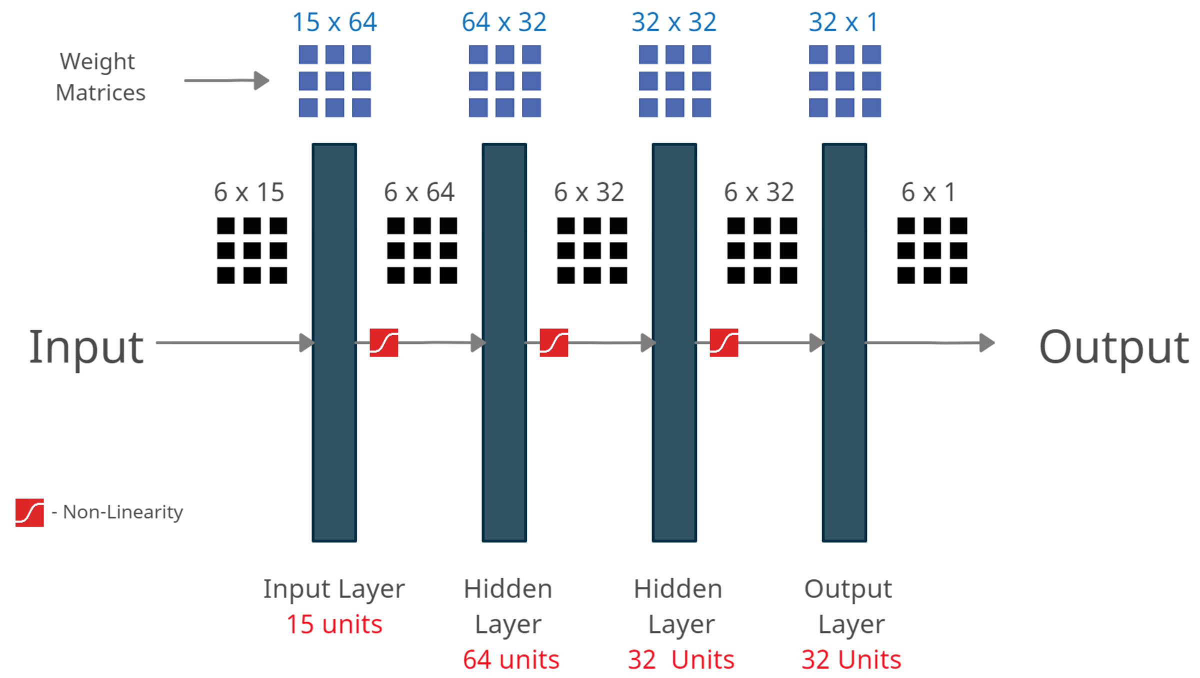


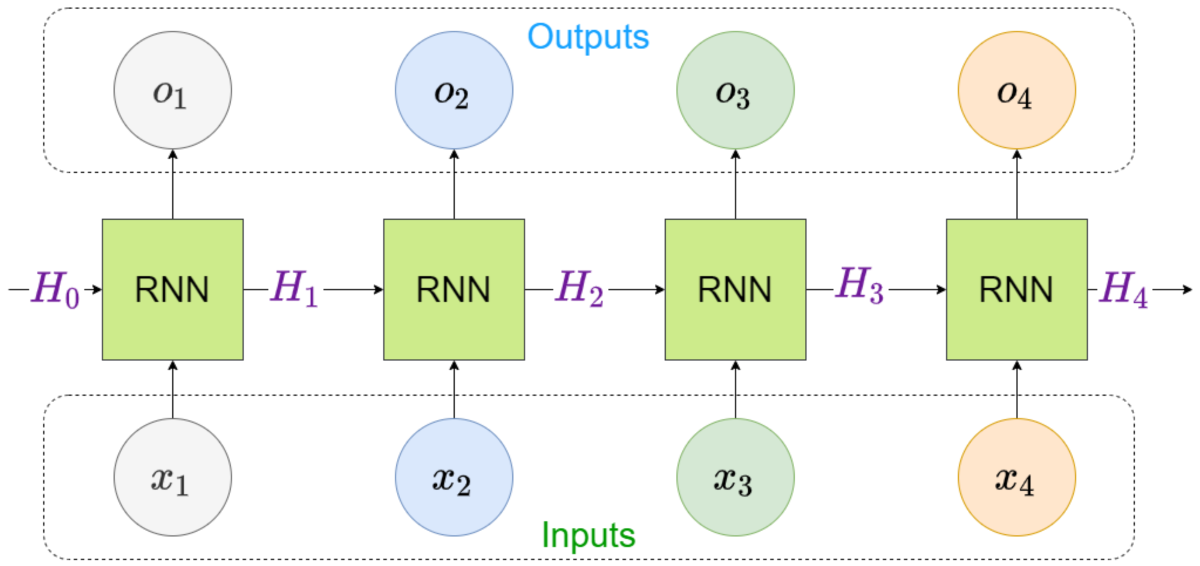


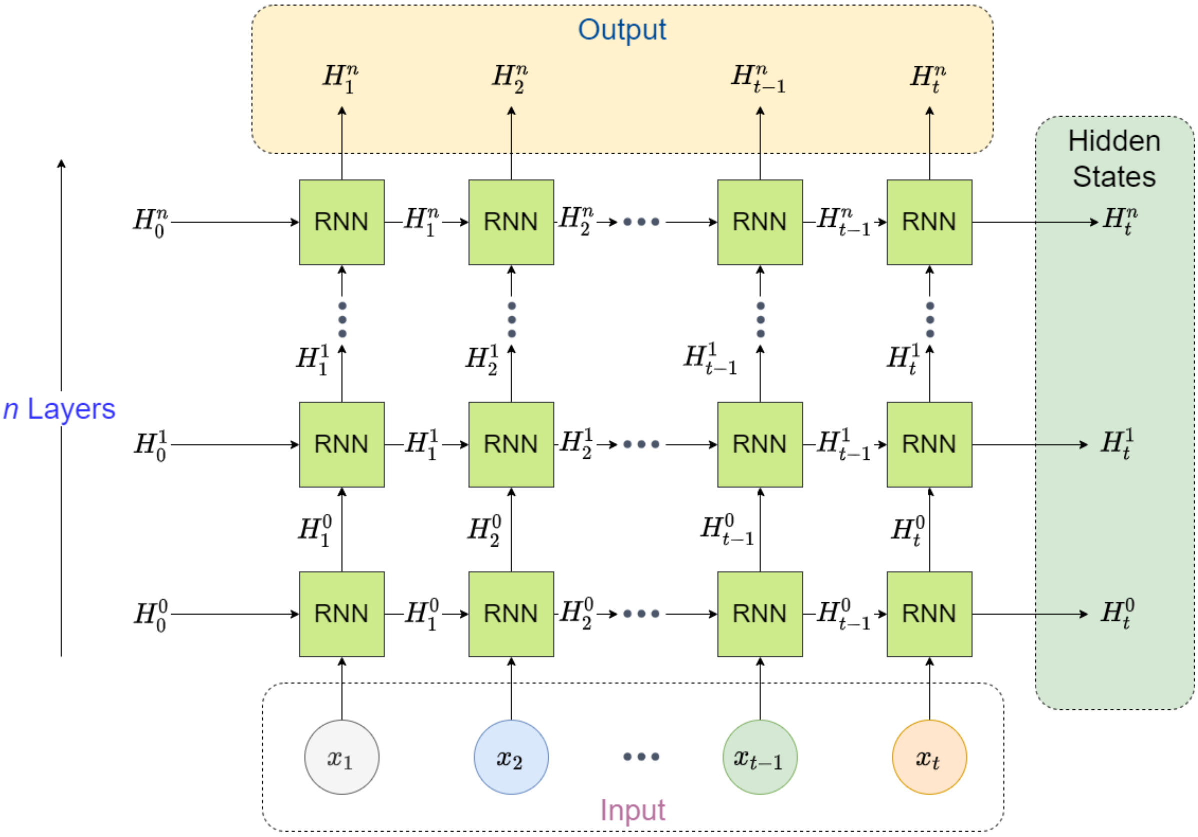
Gradient Descent
Starting Point
Optimal Point
Loss
a
b
2500
2000
1500
1000
500
0
30
20
10
0
−10
−20
−30

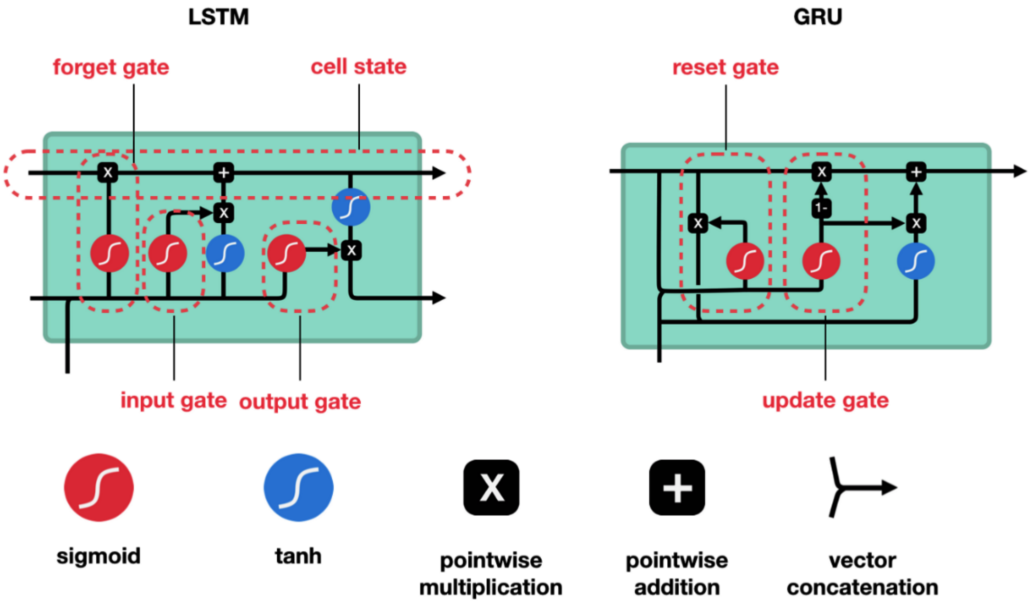
Chapter 12: Building Blocks of Deep Learning for Time Series

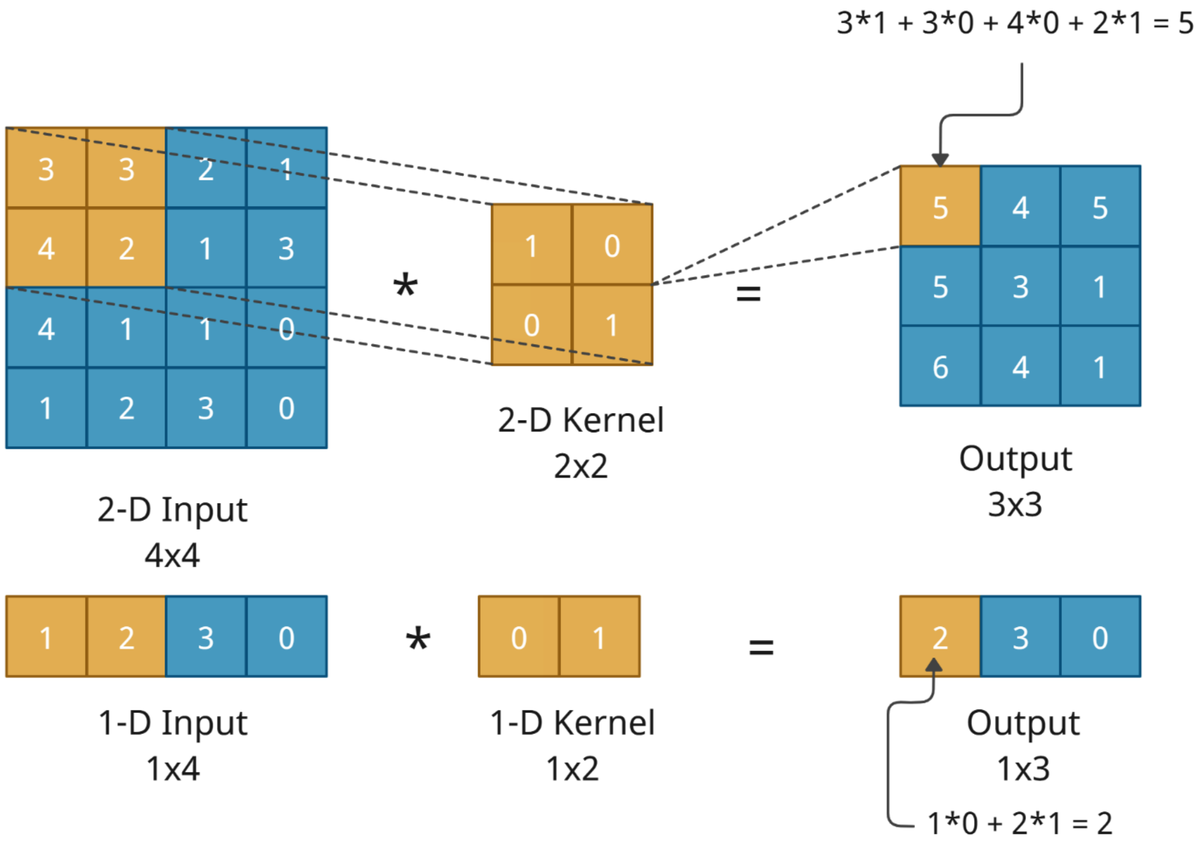


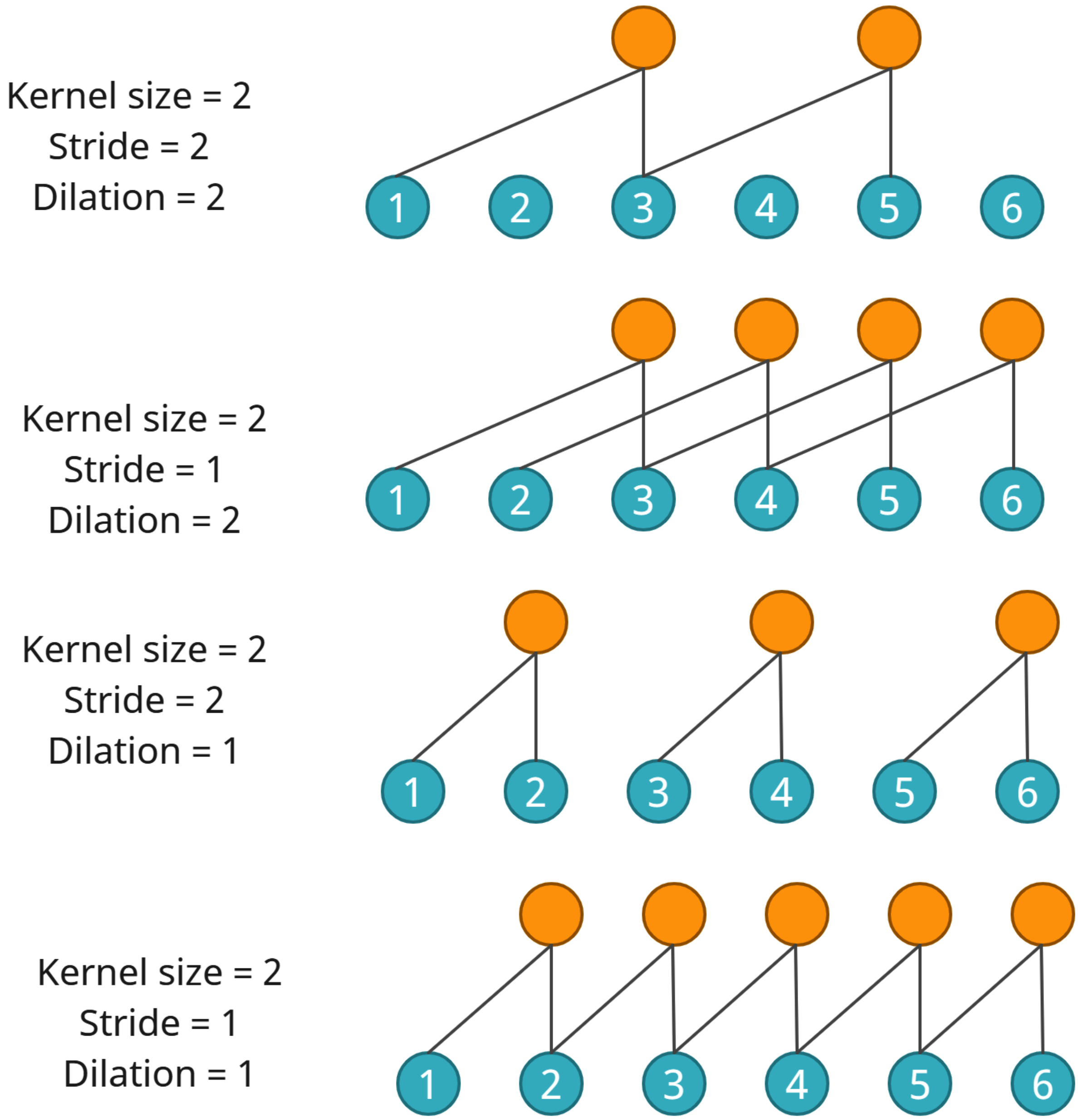

15 x 64
64 x 32
32 x 32
32 x 1
Weight Matrices
6 x 15
6 x 64
6 x 32
6 x 32
6 x 1
Input
Output
- Non-Linearity
Input Layer 15 units
Hidden Layer 64 units
Hidden Layer 32 Units
Output Layer 32 Units


Outputs
o_1 o_2 o_3 o_4
H_0 RNN H_1 RNN H_2 RNN H_3 RNN H_4
x_1 x_2 x_3 x_4
Inputs


Output
H_1^n
H_2^n
H_{t-1}^n
H_t^n
Hidden States
n Layers
RNN
H_0^n
H_0^1
H_0^0
H_t^1
H_t^0
x_1
x_2
x_{t-1}
x_t
Input

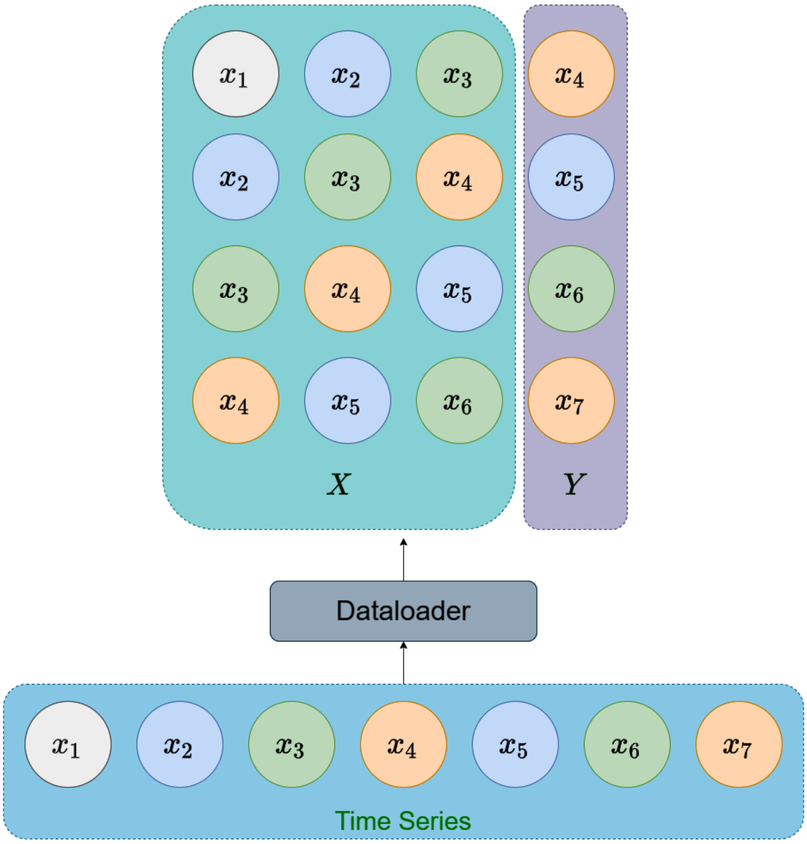

LSTM
GRU
forget gate
cell state
input gate
output gate
reset gate
update gate
sigmoid
tanh
pointwise multiplication
pointwise addition
vector concatenation

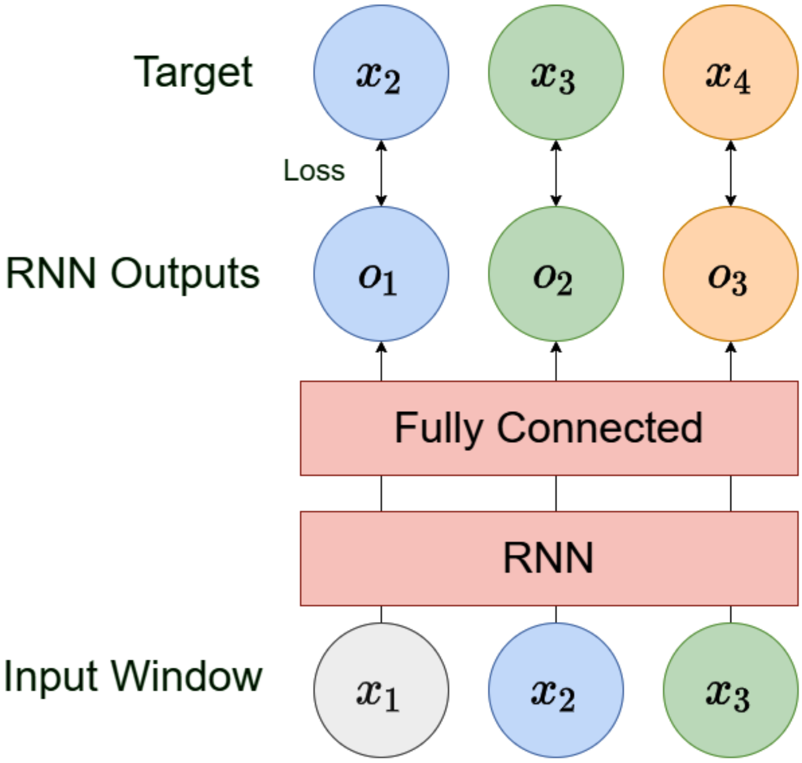

3*1 + 3*0 + 4*0 + 2*1 = 5
3 3 2 1
4 2 1 3
4 1 1 0
1 2 3 0
*
1 0
0 1
=
5 4 5
5 3 1
6 4 1
2-D Kernel
2x2
Output
3x3
2-D Input
4x4
1 2 3 0
*
0 1
=
2 3 0
1-D Input
1x4
1-D Kernel
1x2
Output
1x3
1*0 + 2*1 = 2


Kernel size = 2
Stride = 2
Dilation = 2
Kernel size = 2
Stride = 1
Dilation = 2
Kernel size = 2
Stride = 2
Dilation = 1
Kernel size = 2
Stride = 1
Dilation = 1

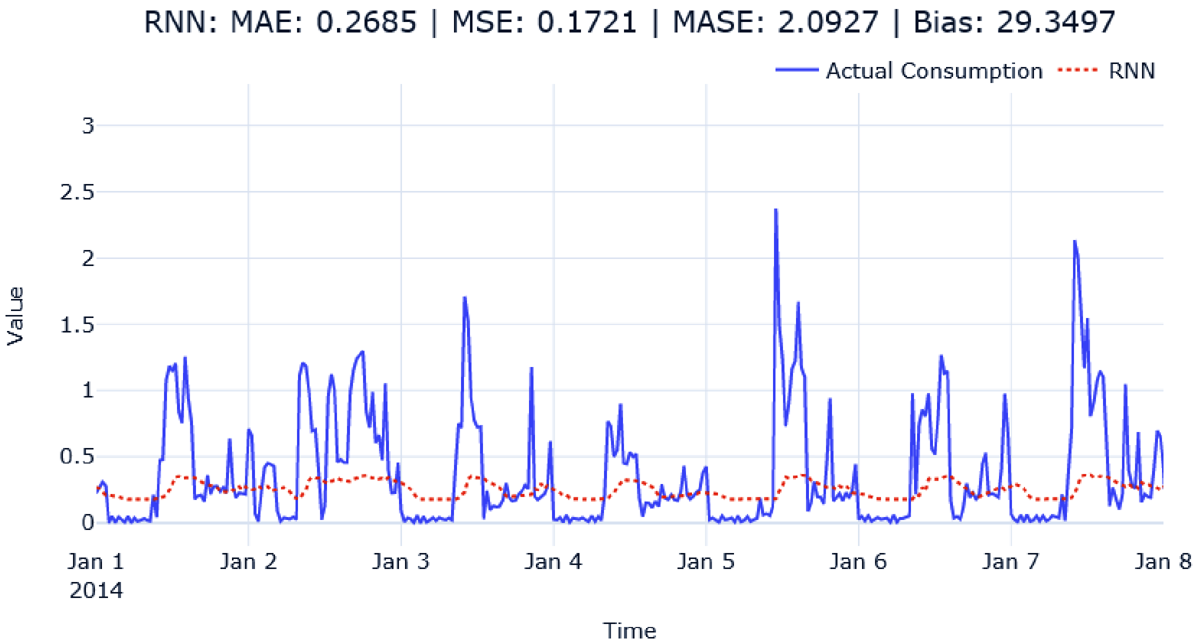
Chapter 13: Common Modeling Patterns for Time Series

Algorithm	MAE	MSE	meanMASE	Forecast Bias
GFM+Meta (NativeLGBM)	0.0873	0.0340	1.0627	-0.68%
FTTransformerModel	0.0913	0.0332	1.1598	5.90%

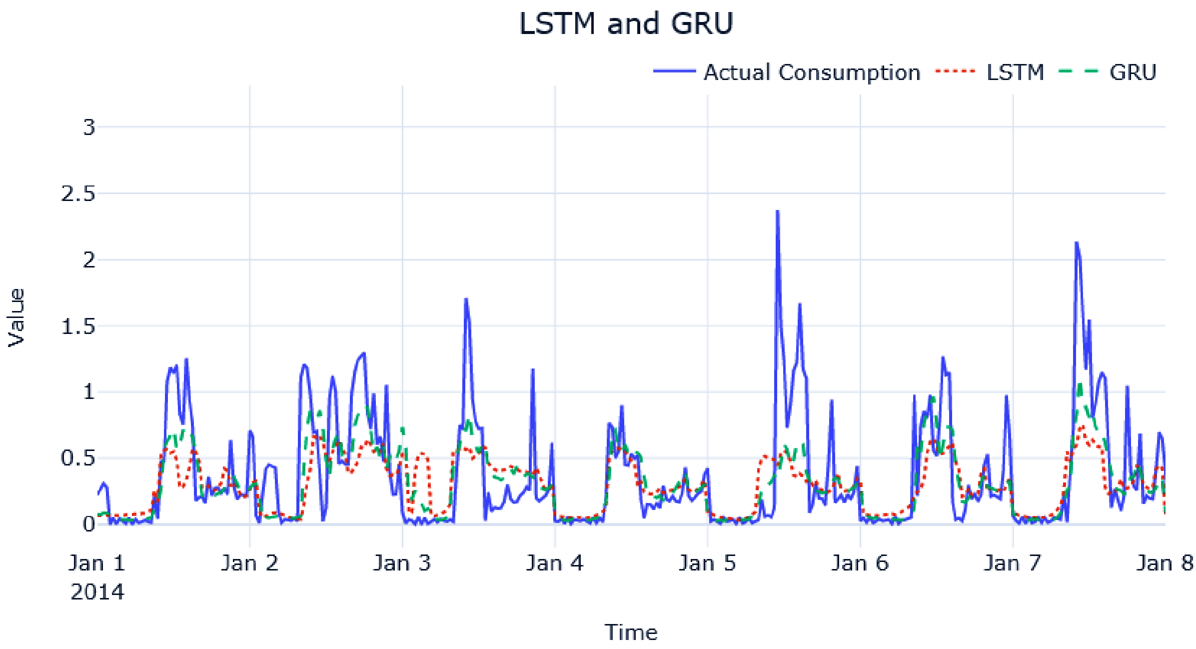


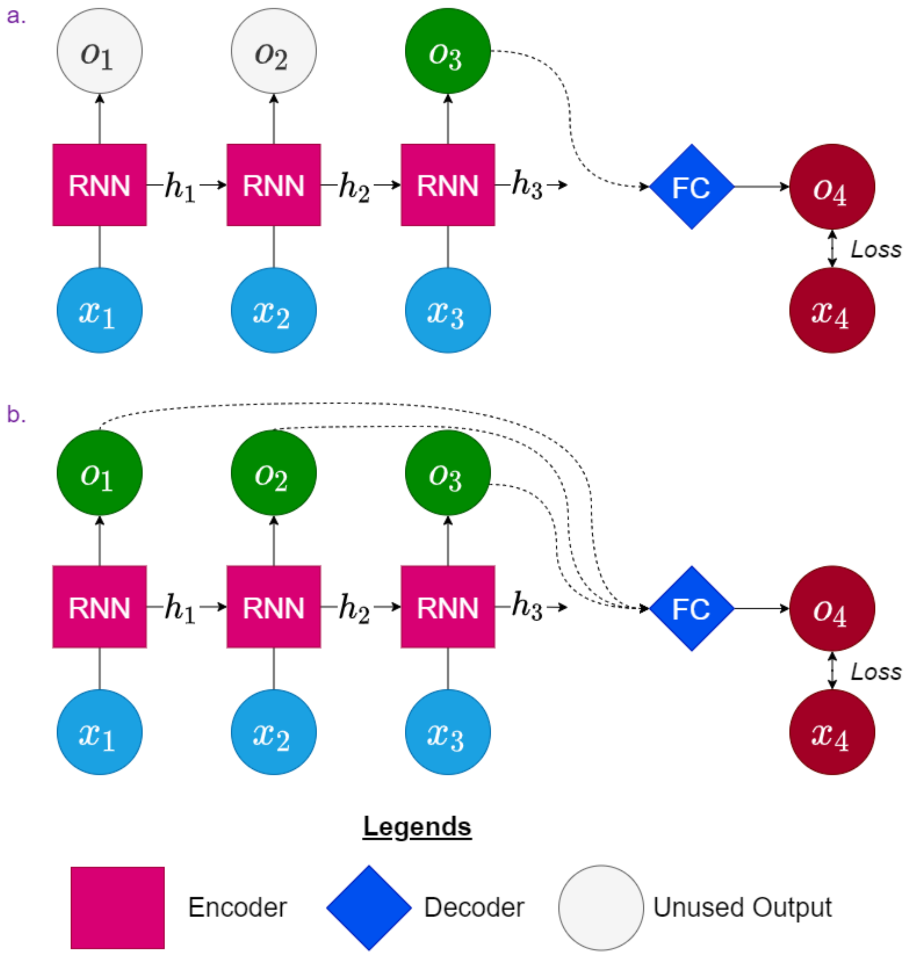


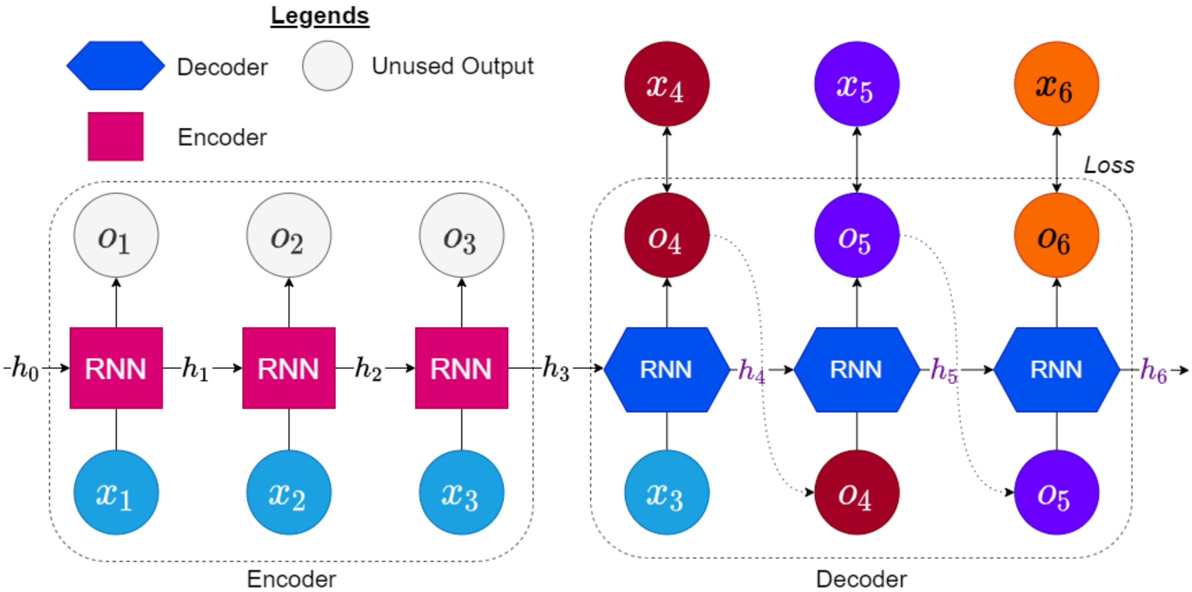
Algorithm	MAE	MSE	MASE	Forecast Bias
Lasso Regression	0.1598	0.0743	1.2452	3.78%
XGB Random Forest	0.1641	0.0819	1.2792	9.30%
LightGBM	0.1470	0.0666	1.1457	3.36%
RNN	0.2685	0.1721	2.0927	29.35%



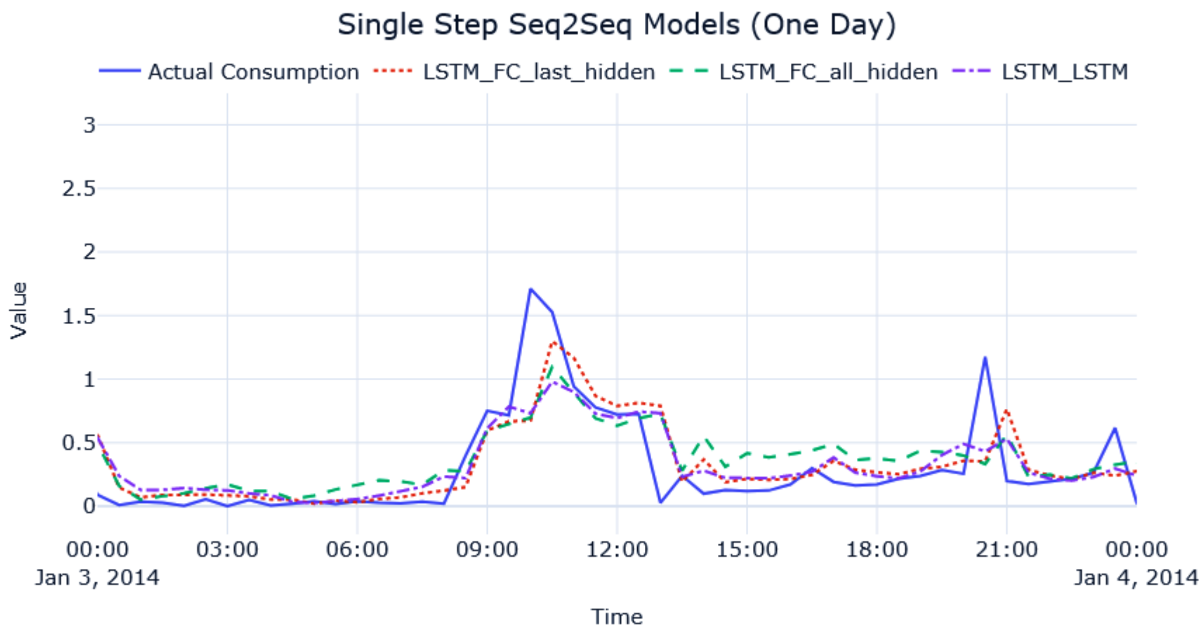
Algorithm	MAE	MSE	MASE	Forecast Bias
Lasso Regression	0.1598	0.0743	1.2452	3.78%
XGB Random Forest	0.1641	0.0819	1.2792	9.30%
LightGBM	0.1470	0.0666	1.1457	3.36%
RNN	0.2685	0.1721	2.0927	29.35%
LSTM	0.1982	0.1125	1.5442	17.94%
GRU	0.1714	0.0899	1.3358	14.48%


LSTM and GRU
Actual Consumption
LSTM
GRU
Value
3
2.5
2
1.5
1
0.5
0
Jan 1
2014
Jan 2
Jan 3
Jan 4
Jan 5
Jan 6
Jan 7
Jan 8
Time

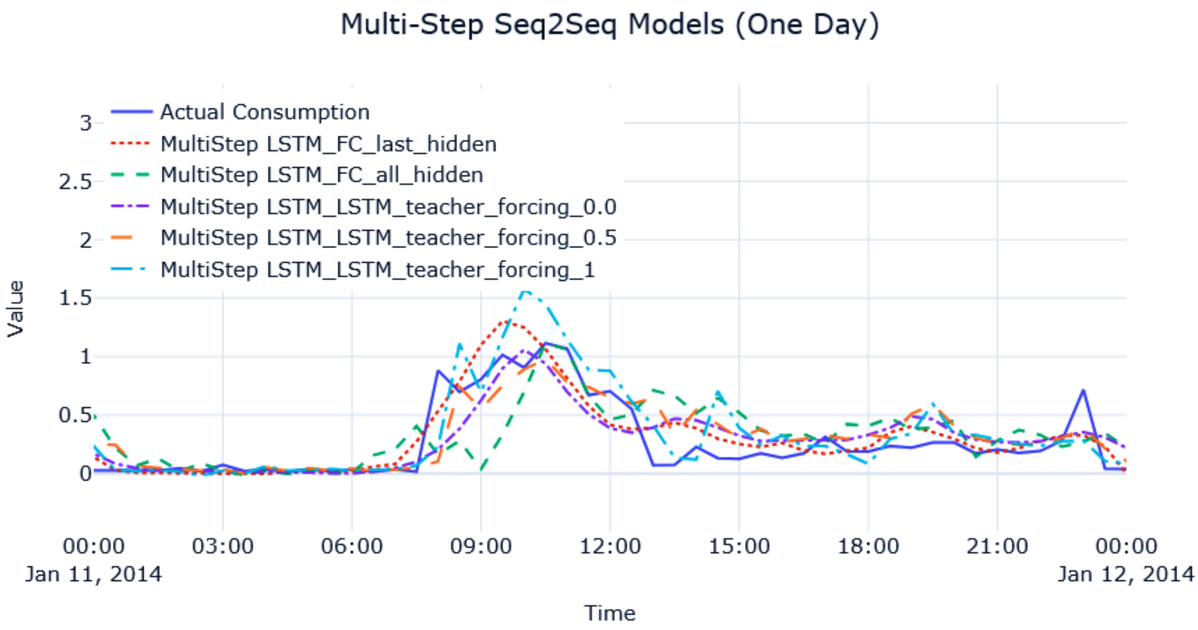

a.
o_1
o_2
o_3
RNN
h_1
h_2
h_3
FC
o_4
Loss
x_1
x_2
x_3
x_4
b.
Legends
Encoder
Decoder
Unused Output



Algorithm	MAE	MSE	MASE	Forecast Bias
Lasso Regression	0.1598	0.0743	1.2452	3.78%
XGB Random Forest	0.1641	0.0819	1.2792	9.30%
LightGBM	0.1470	0.0666	1.1457	3.36%
RNN	0.2685	0.1721	2.0927	29.35%
LSTM	0.1982	0.1125	1.5442	17.94%
GRU	0.1714	0.0899	1.3358	14.48%
LSTM_FC_last_hidden	0.1642	0.0815	1.2797	5.87%
LSTM_FC_all_hidden	0.1667	0.0799	1.2993	10.00%
LSTM_LSTM	0.1600	0.0795	1.2472	13.31%

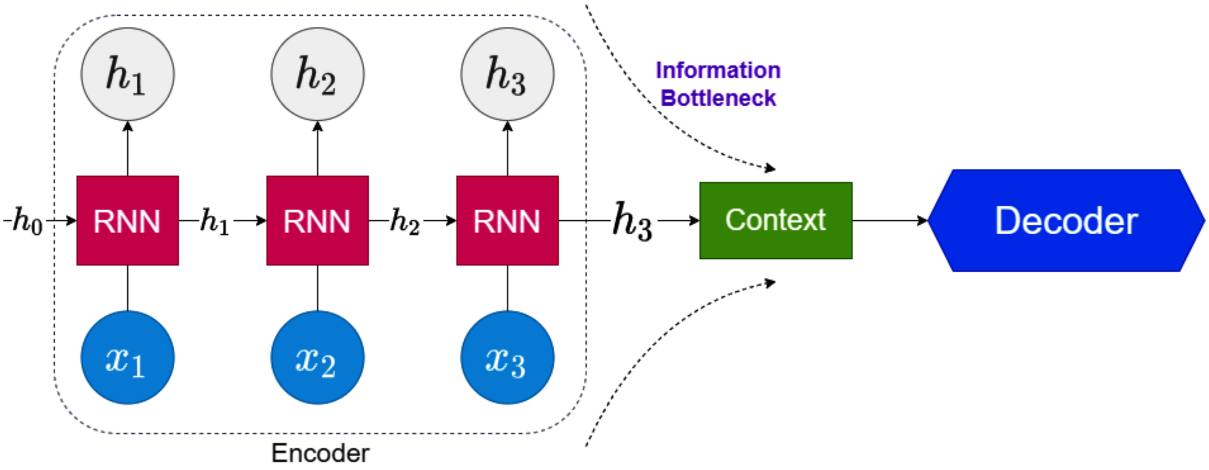


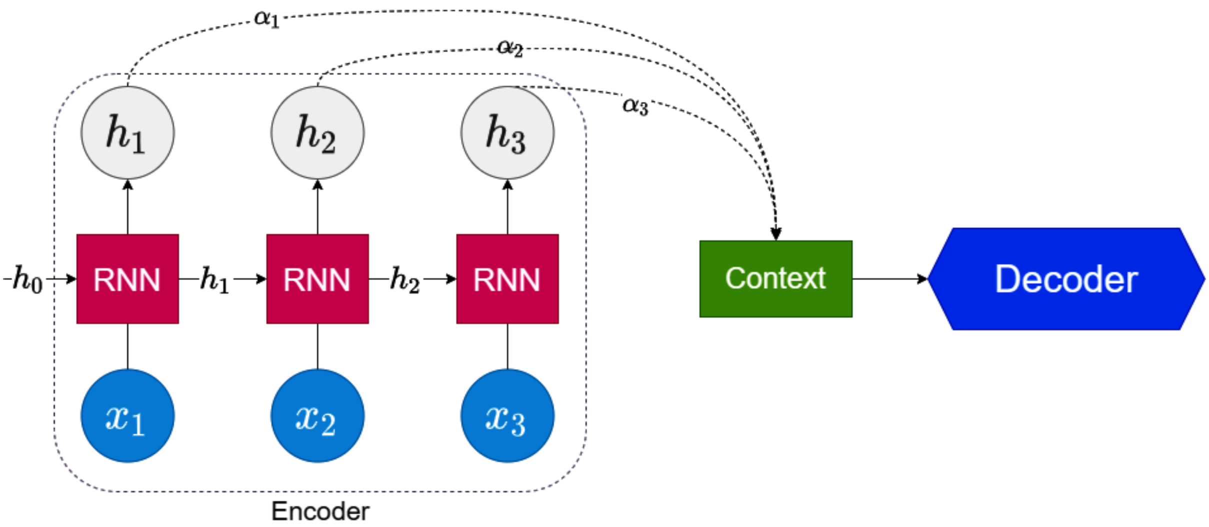
Algorithm	MAE	MSE	MASE	Forecast Bias
Lasso Regression	0.1598	0.0743	1.2452	3.78%
XGB Random Forest	0.1641	0.0819	1.2792	9.30%
LightGBM	0.1470	0.0666	1.1457	3.36%
RNN	0.2685	0.1721	2.0927	29.35%
LSTM	0.1982	0.1125	1.5442	17.94%
GRU	0.1714	0.0899	1.3358	14.48%
LSTM_FC_last_hidden	0.1642	0.0815	1.2797	5.87%
LSTM_FC_all_hidden	0.1667	0.0799	1.2993	10.00%
LSTM_LSTM	0.1600	0.0795	1.2472	13.31%
MultiStep LSTM_FC_last_hidden	0.2177	0.1305	1.6967	10.59%
MultiStep LSTM_FC_all_hidden	0.2344	0.1317	1.8265	9.33%
MultiStep LSTM_LSTM_teacher_forcing_0.0	0.2058	0.1241	1.6039	15.32%
MultiStep LSTM_LSTM_teacher_forcing_0.5	0.1866	0.0997	1.4544	11.90%
MultiStep LSTM_LSTM_teacher_forcing_1	0.1754	0.0912	1.3671	12.93%

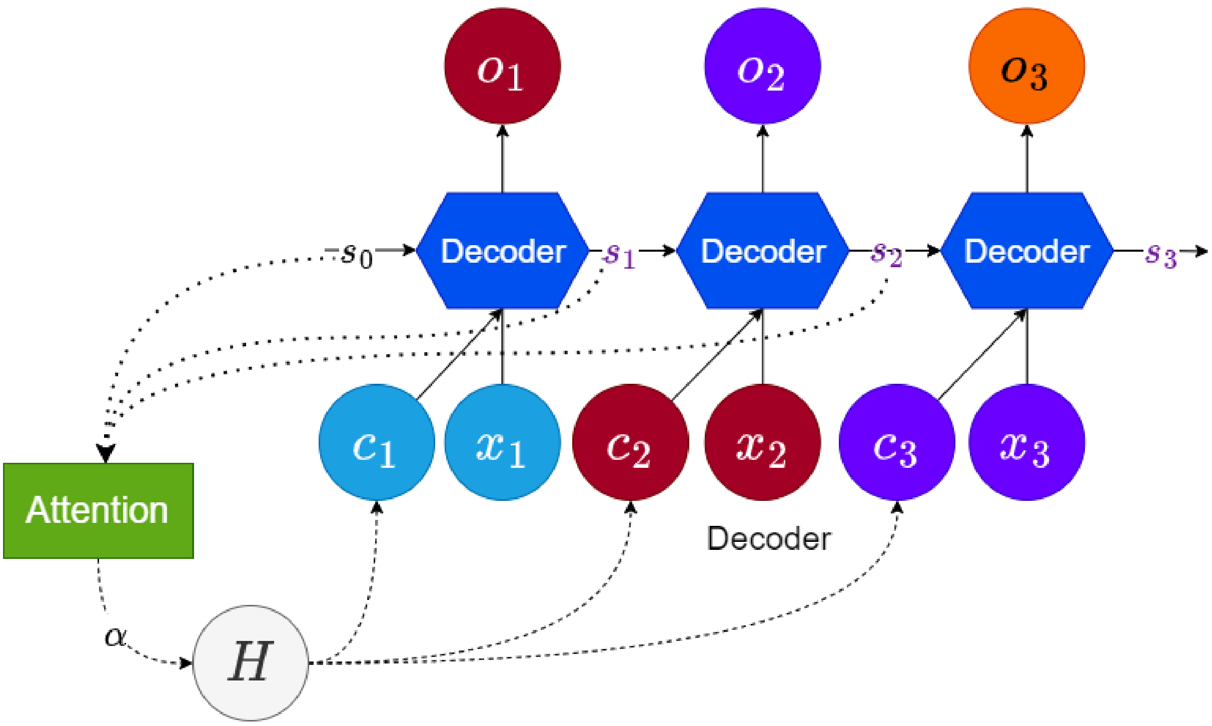


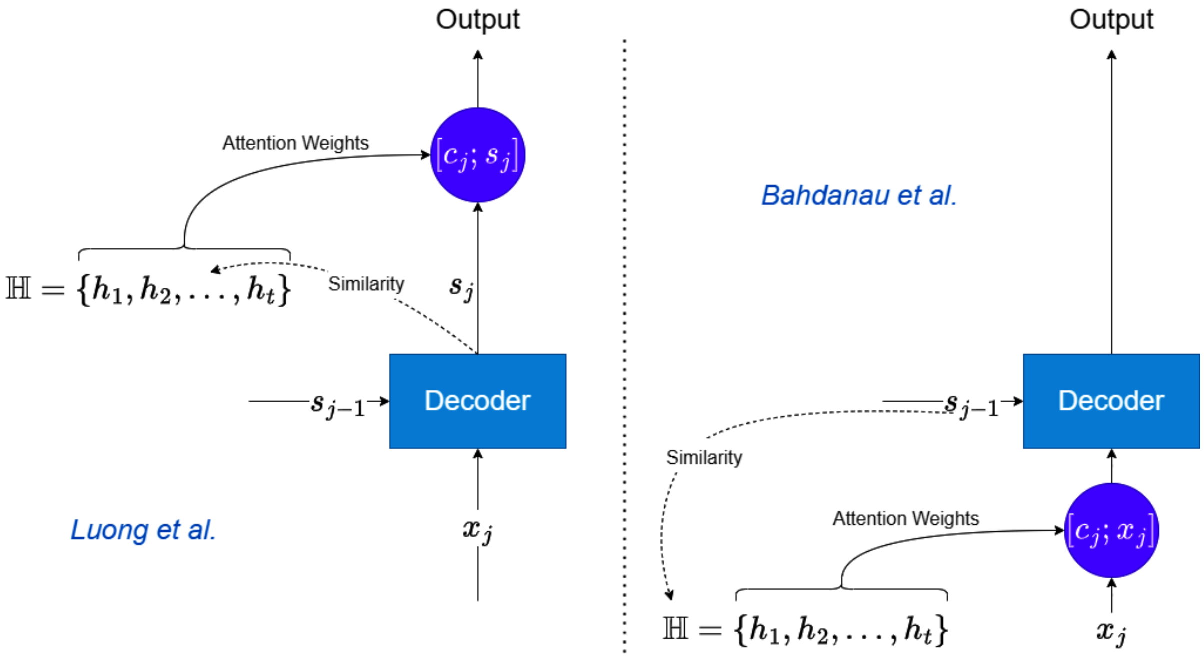
Multi-Step Seq2Seq Models (One Day)
Actual Consumption
MultiStep LSTM_FC_last_hidden
MultiStep LSTM_FC_all_hidden
MultiStep LSTM_LSTM_teacher_forcing_0.0
MultiStep LSTM_LSTM_teacher_forcing_0.5
MultiStep LSTM_LSTM_teacher_forcing_1
Value
3
2.5
2
1.5
1
0.5
0
00:00
Jan 11, 2014
03:00
06:00
09:00
12:00
15:00
18:00
21:00
00:00
Jan 12, 2014
Time

Chapter 14: Attention and Transformers for Time Series

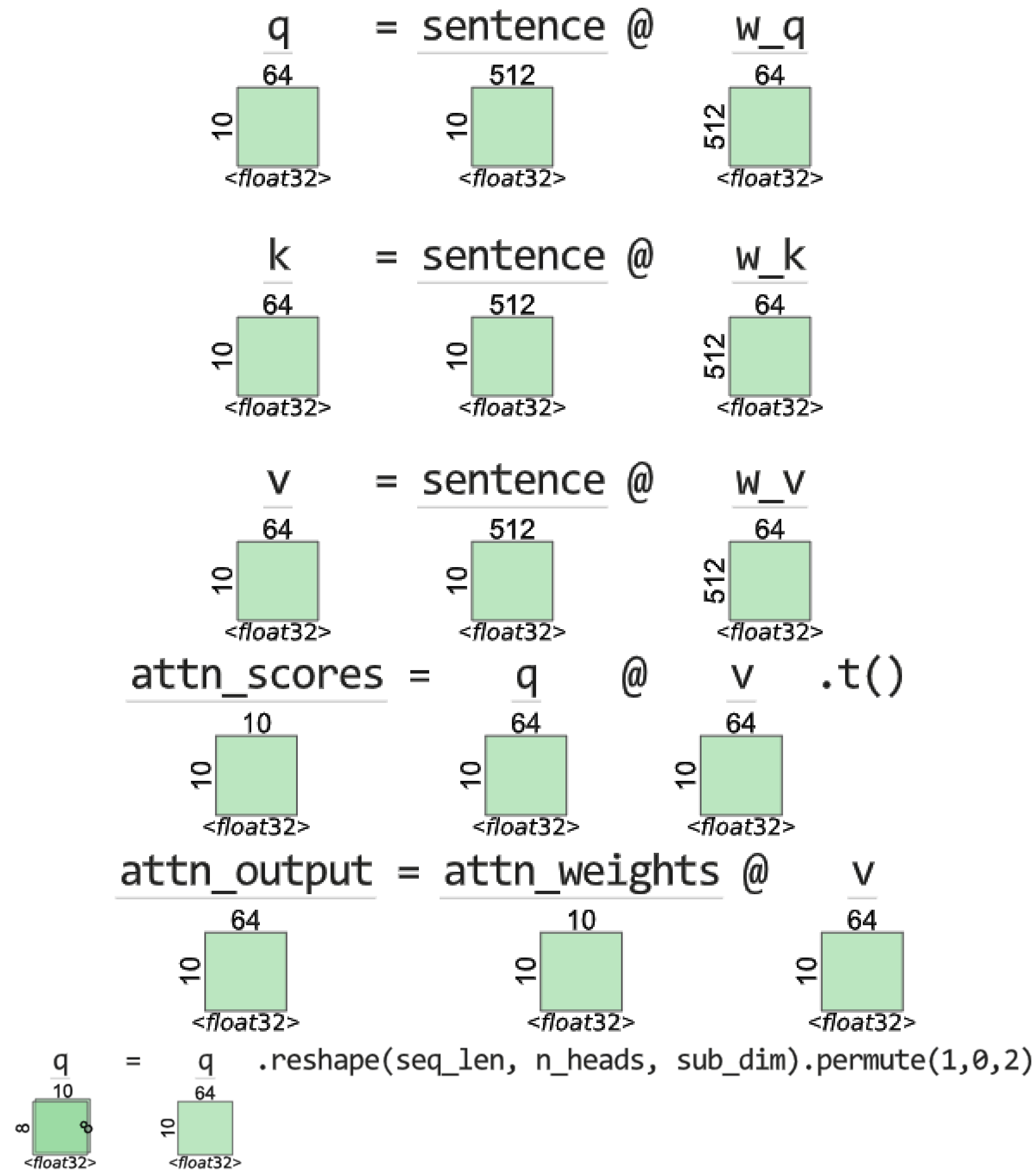






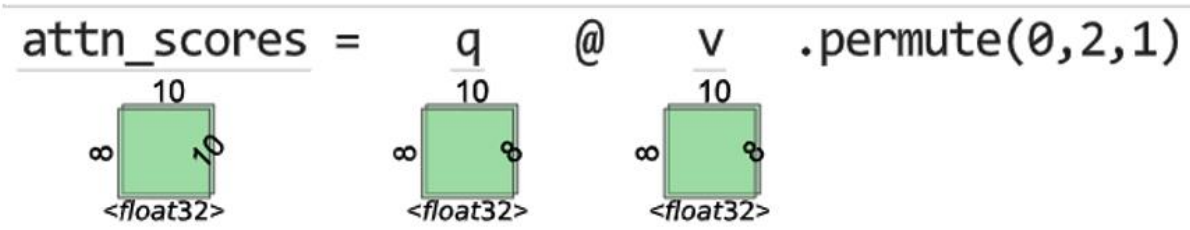


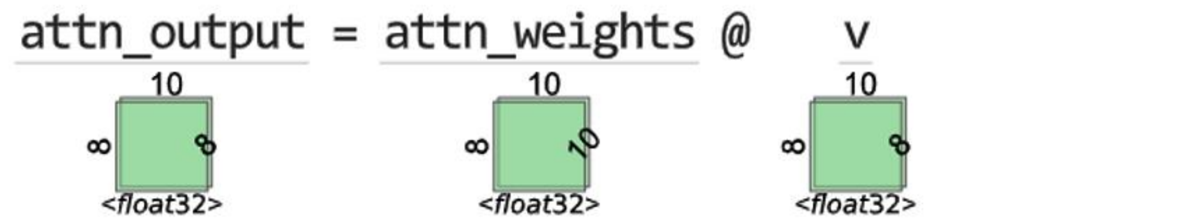
Algorithm	MAE	MSE	MASE	Forecast Bias
MultiStep LSTM_LSTM_teacher_forcing_0.0	0.2058	0.1241	1.6039	15.32%
MultiStep LSTM_LSTM_teacher_forcing_1	0.1754	0.0912	1.3671	12.93%
MultiStep_Seq2Seq_dot_Attn_teacher_forcing_1	0.1536	0.0717	1.1967	8.25%
MultiStep_Seq2Seq_scaled_dot_Attn_teacher_forcing_1	0.1772	0.0878	1.3812	6.00%
MultiStep_Seq2Seq_general_Attn_teacher_forcing_1	0.1665	0.0818	1.2977	8.97%
MultiStep_Seq2Seq_concat_Attn_teacher_forcing_1	0.1632	0.0778	1.2717	6.65%
MultiStep_Seq2Seq_additive_Attn_teacher_forcing_1	0.1561	0.0734	1.2168	10.40%



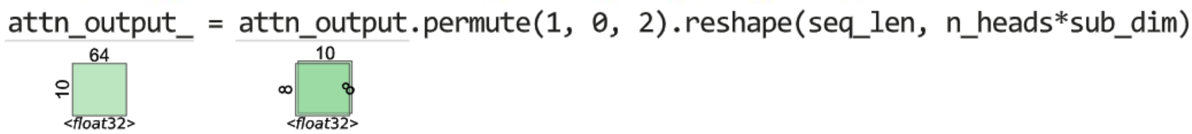
---

q, k, v dimensions: n_heads, seq_len, sub_dim : 8 x 10 x 8

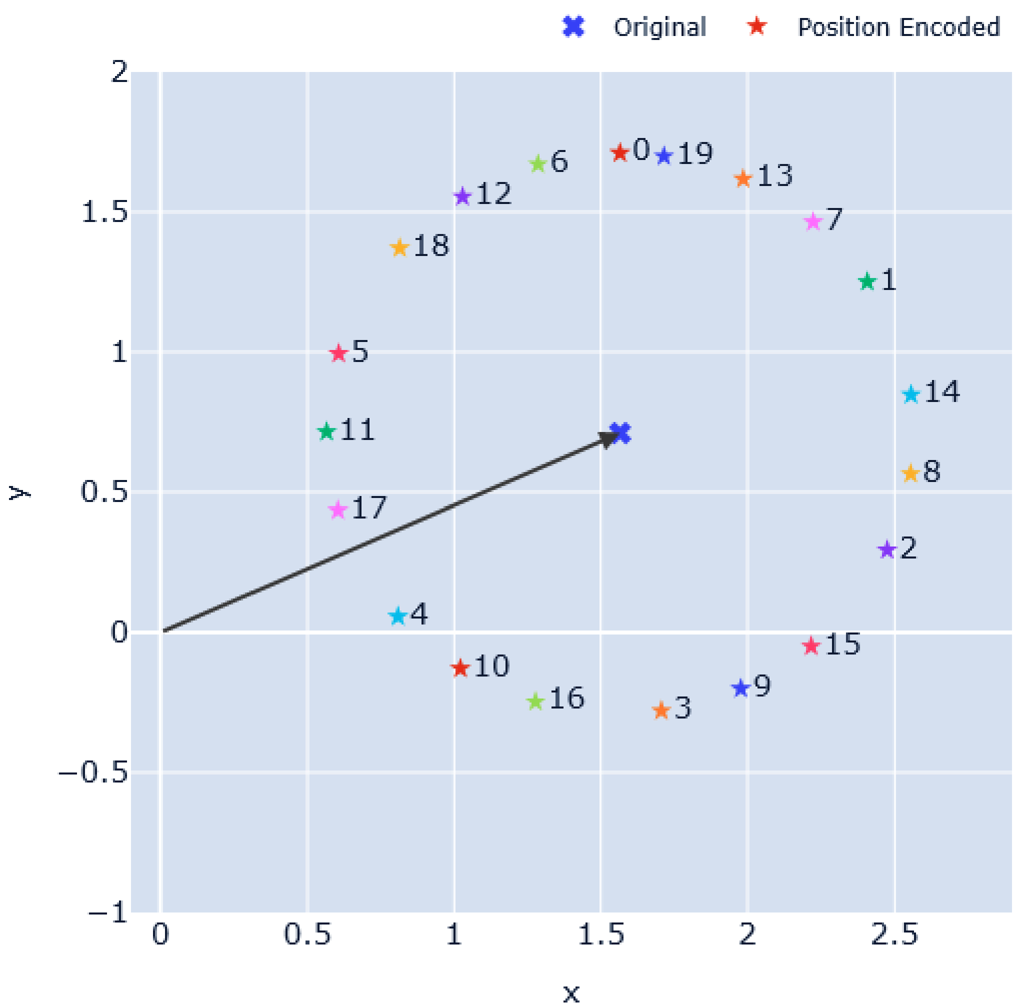


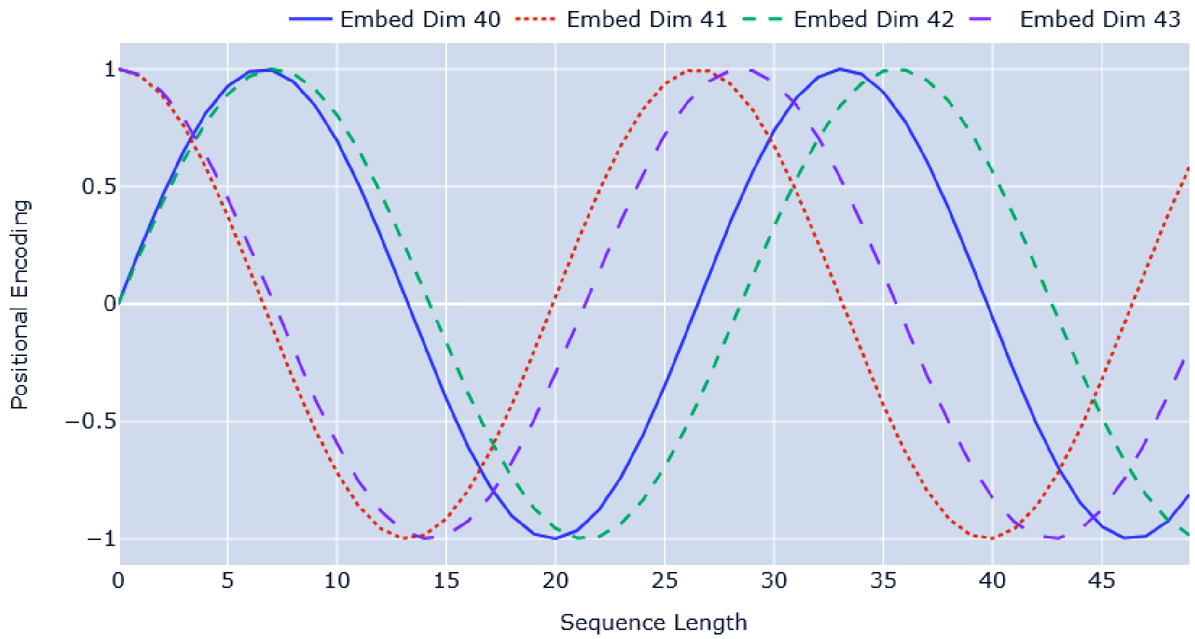


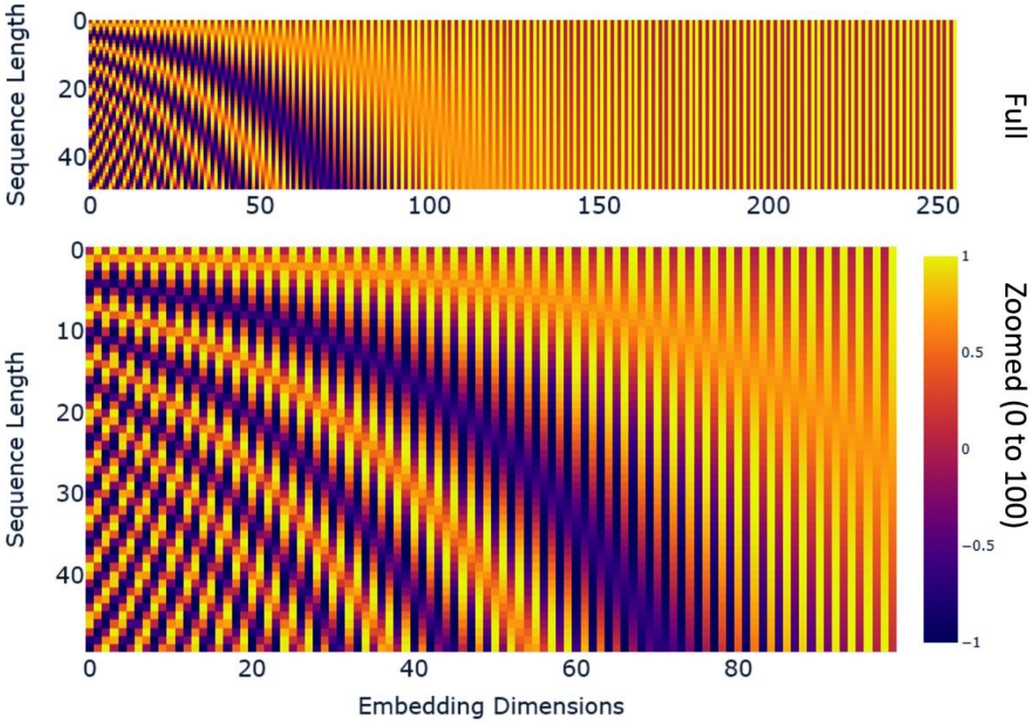
attn_output dimensions: n_heads, seq_len, sub_dim : 8 x 10 x 8



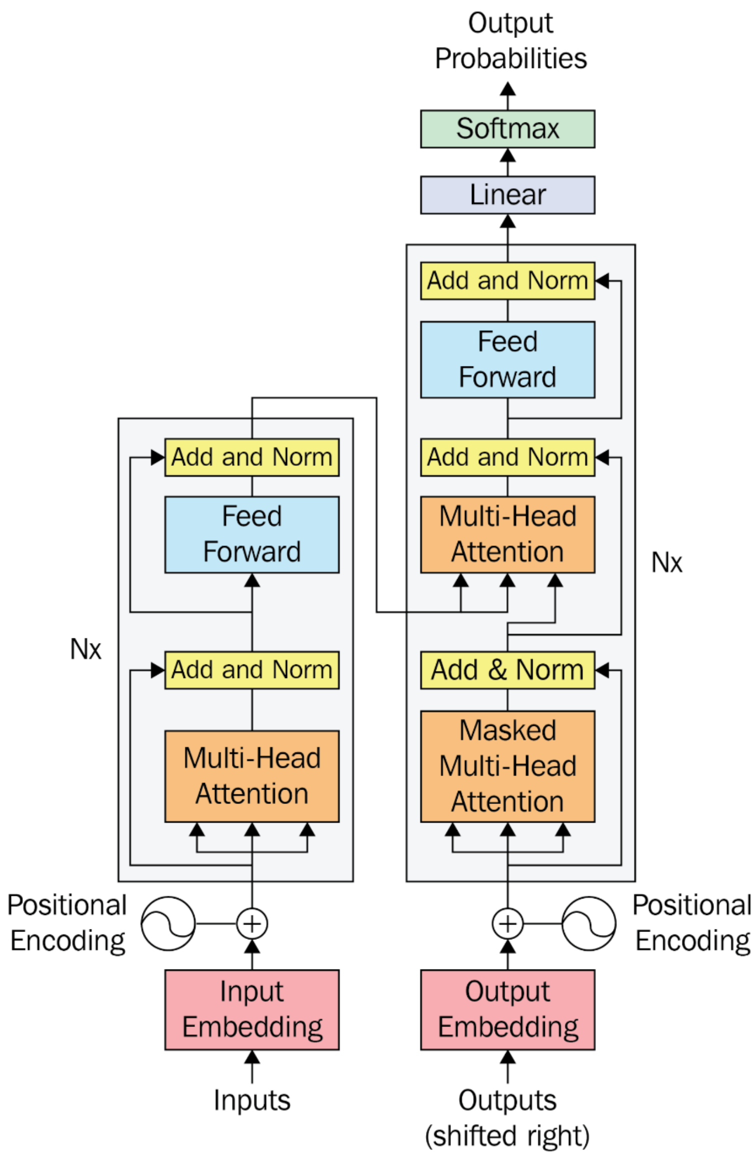
attn_output dimensions: seq_len, n_heads*sub_dim (attn_dim) : 8 x 10 x 8

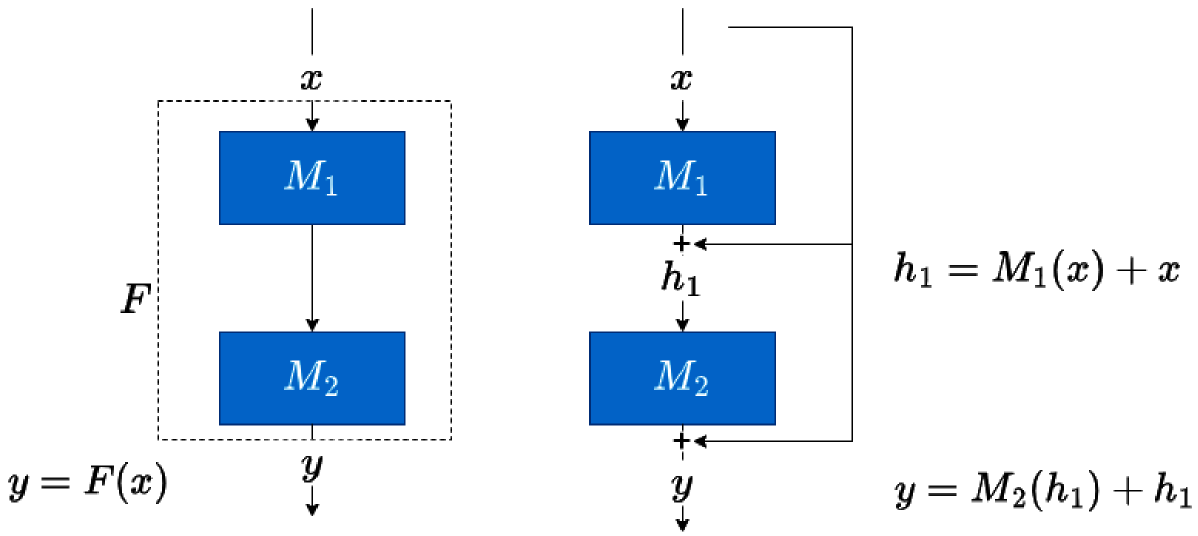

Original
Position Encoded
y
x
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19


Embed Dim 40
Embed Dim 41
Embed Dim 42
Embed Dim 43
Positional Encoding
Sequence Length

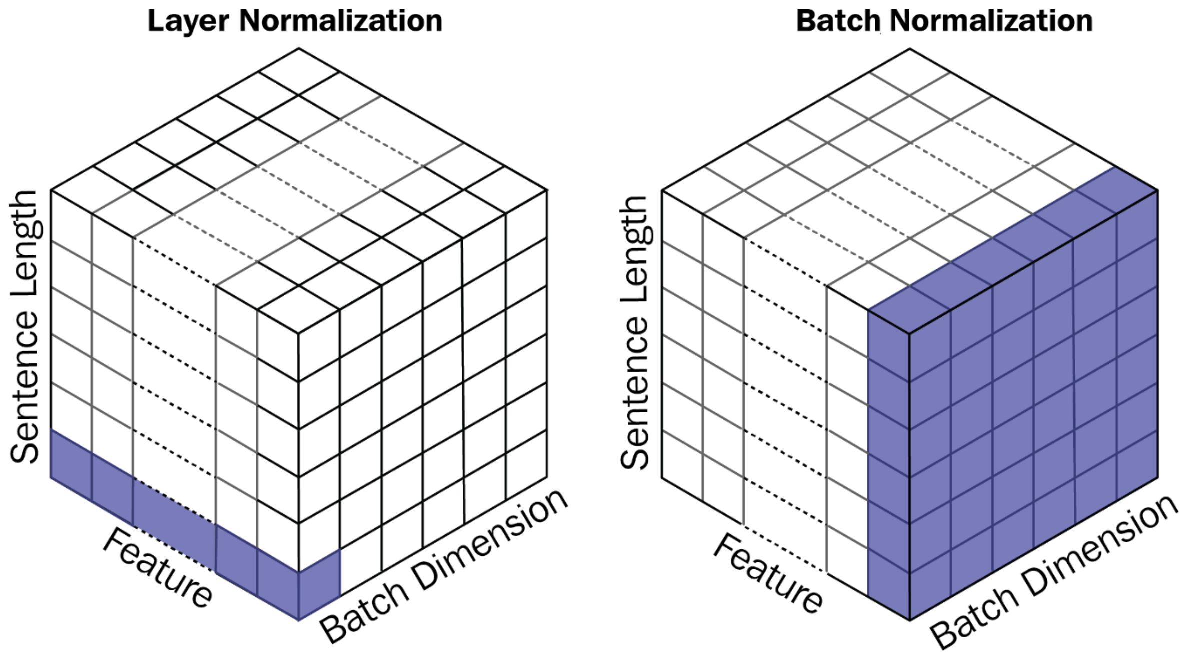


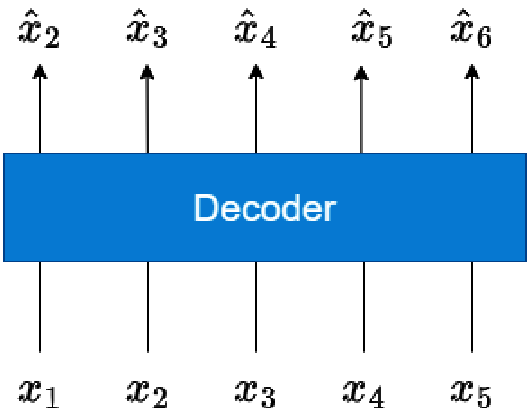
Sequence Length
0
20
40
0
50
100
150
200
250
Full
Sequence Length
0
10
20
30
40
0
20
40
60
80
Embedding Dimensions
1
0.5
0
−0.5
−1
Zoomed (0 to 100)

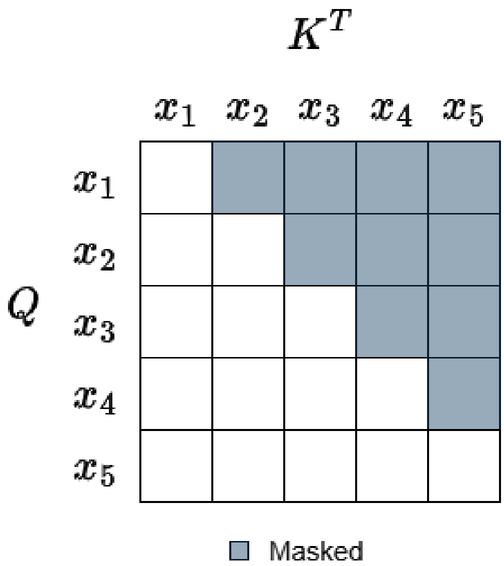




a. Regular Neural Network b. Residual Neural Network

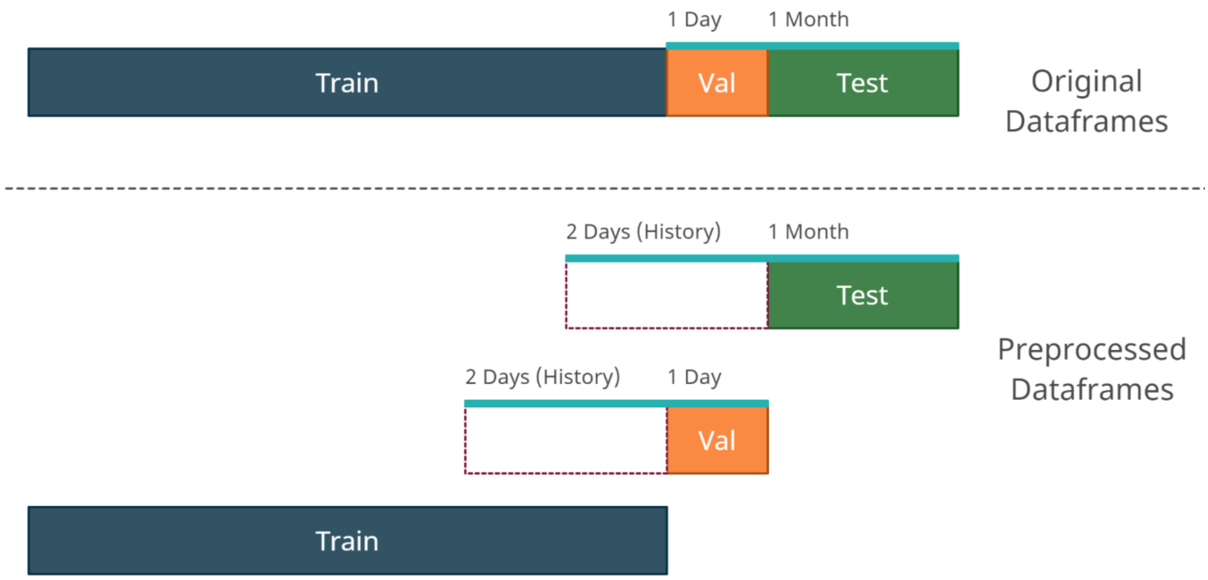






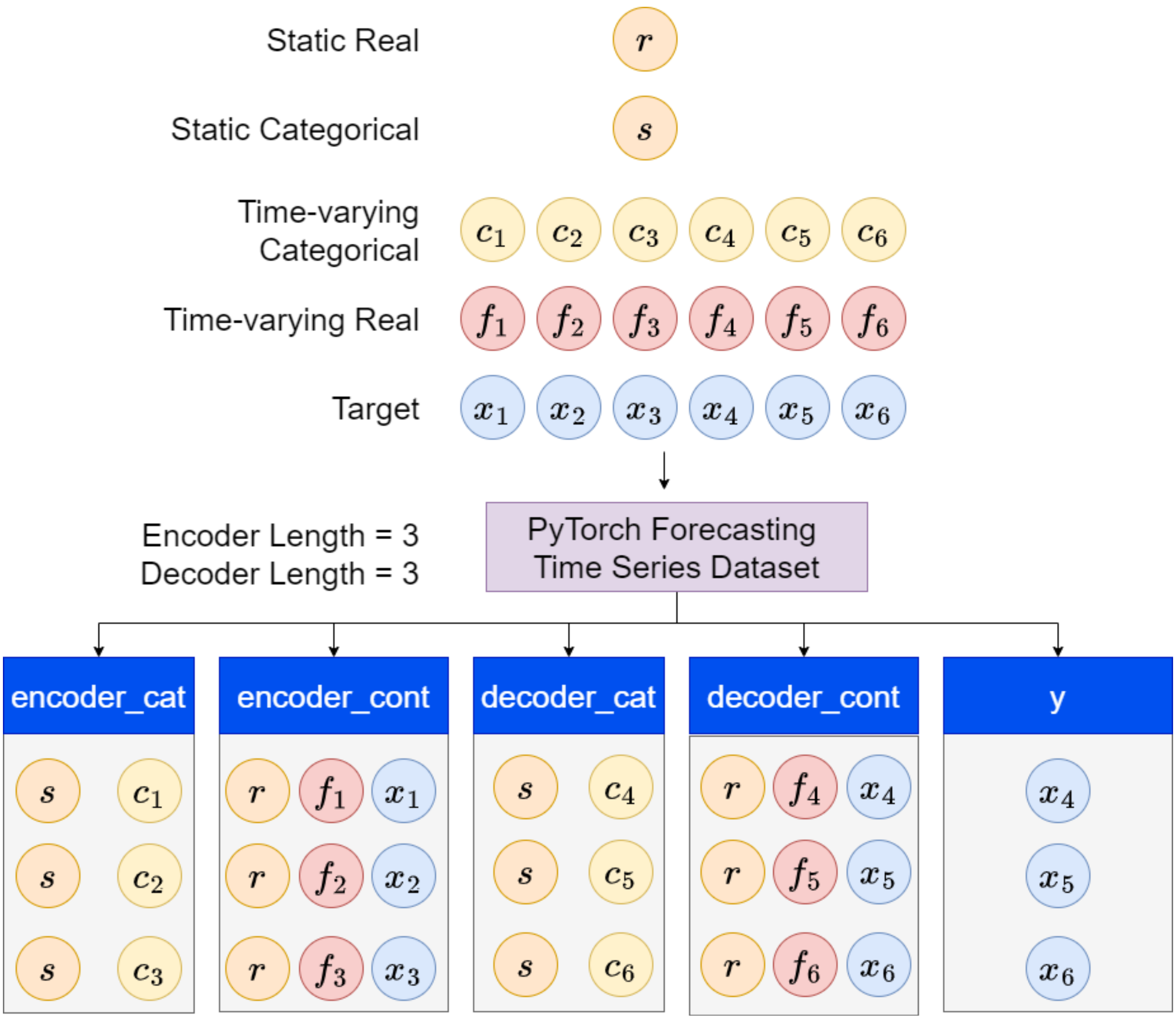
Algorithm	MAE	MSE	MASE	Forecast Bias
MultiStep LSTM_LSTM_teacher_forcing_1	0.1754	0.0912	1.3671	12.93%
MultiStep_Seq2Seq_dot_Attn_teacher_forcing_1	0.1536	0.0717	1.1967	8.25%
MultiStep_Transformer_Multi_Step_FF_decoder	0.1949	0.1104	1.5188	16.24%

Chapter 15: Strategies for Global Deep Learning Forecasting Models

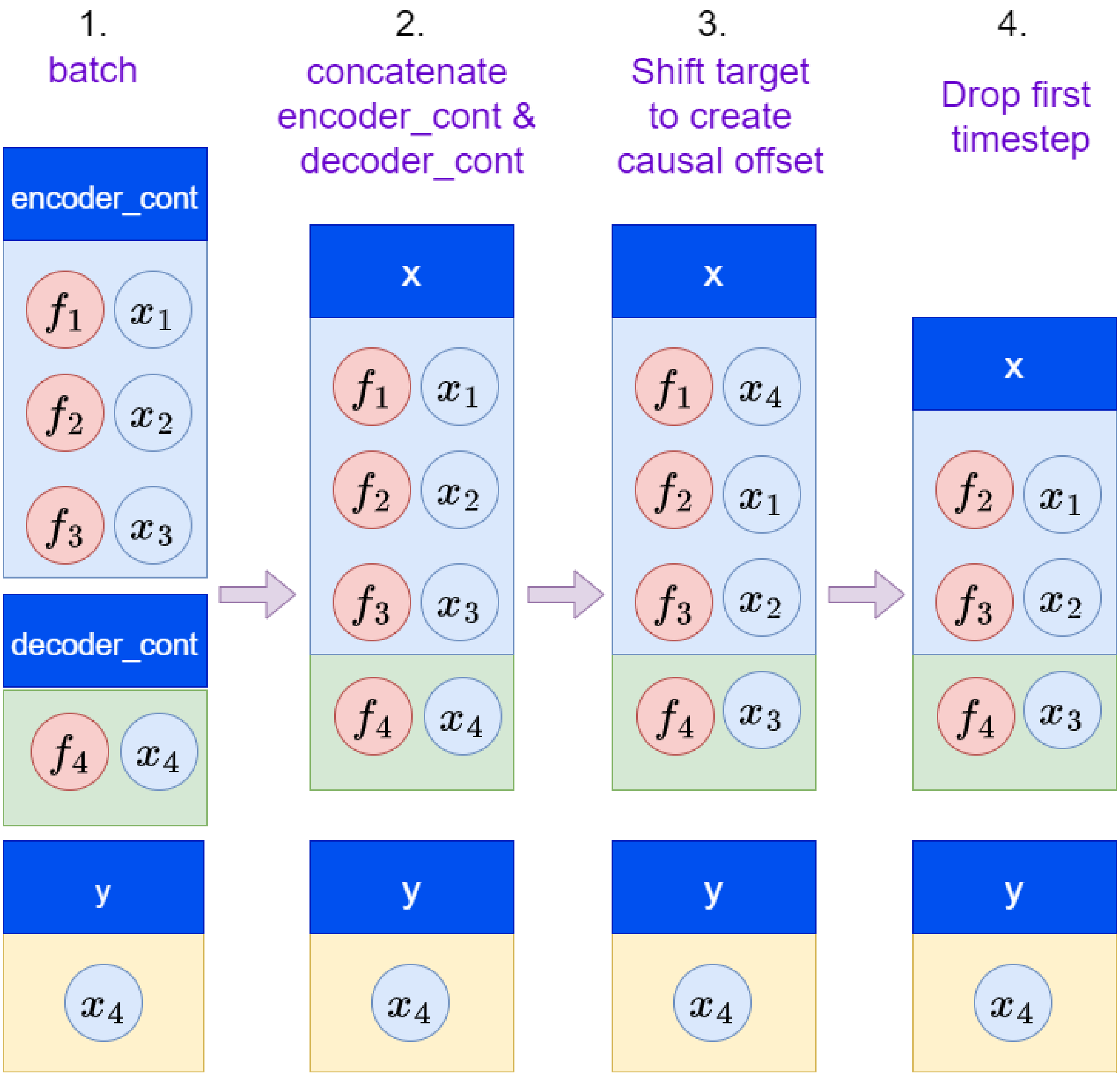



```
sizes of x =
        encoder_cat = torch.Size([512, 96, 0])
        encoder_cont = torch.Size([512, 96, 1])
        encoder_target = torch.Size([512, 96])
        encoder_lengths = torch.Size([512])
        decoder_cat = torch.Size([512, 1, 0])
        decoder_cont = torch.Size([512, 1, 1])
        decoder_target = torch.Size([512, 1])
        decoder_lengths = torch.Size([512])
        decoder_time_idx = torch.Size([512, 1])
        groups = torch.Size([512, 1])
        target_scale = torch.Size([512, 2])

size of y =
        y = torch.Size([512, 1])
```



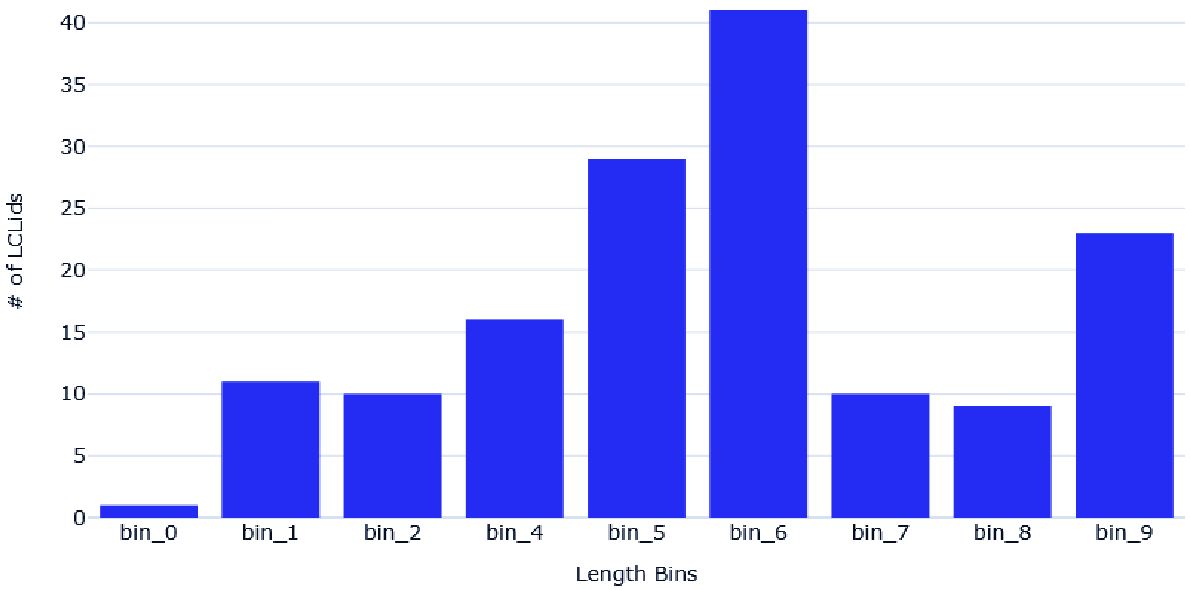
Algorithm	MAE	MSE	meanMASE	Forecast Bias
simple	0.085441	0.030798	1.078673	1.410458

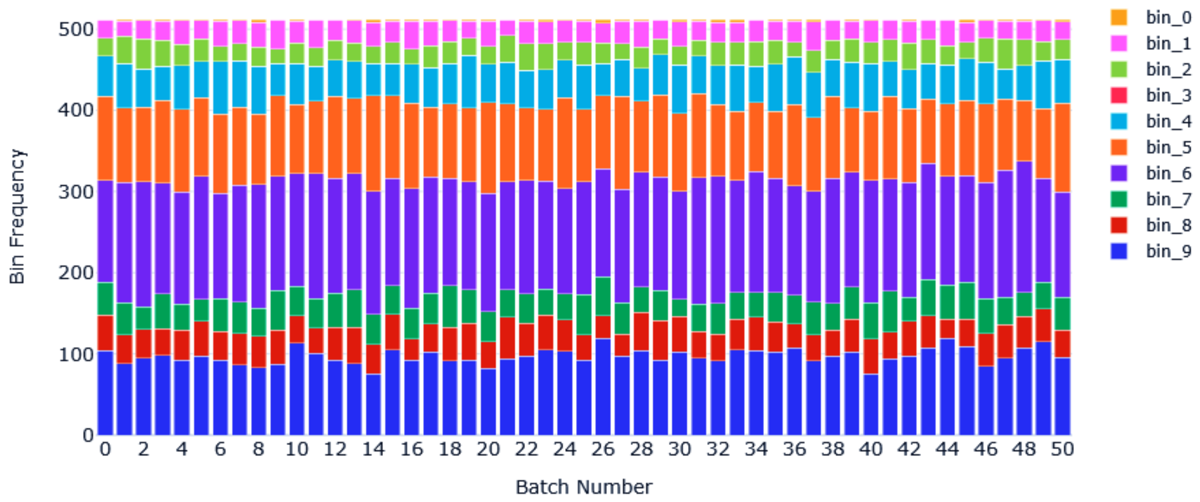


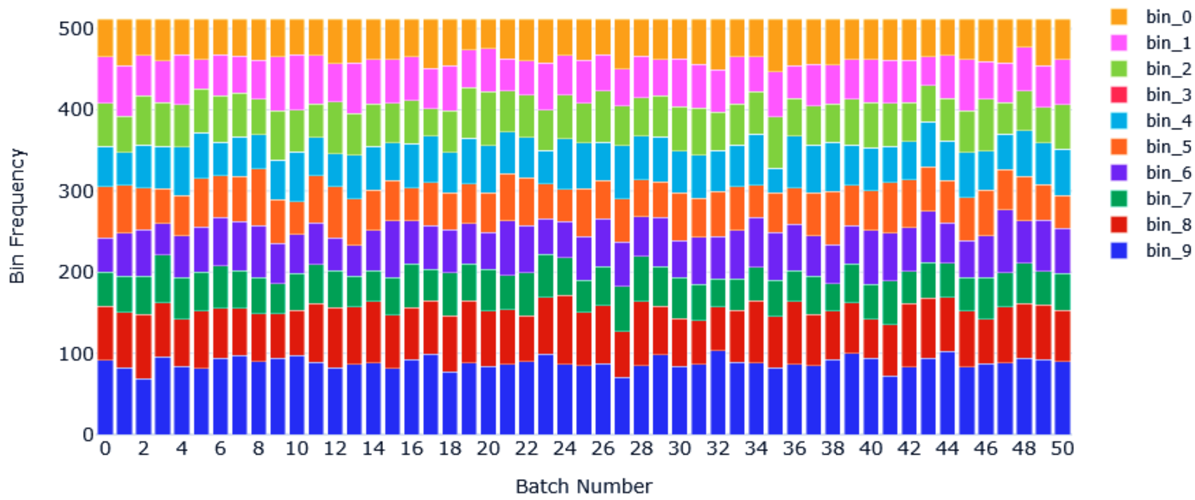
Algorithm	MAE	MSE	meanMASE	Forecast Bias
simple	0.0854	0.0308	1.0787	1.41%
simple+time_varying	0.0851	0.0310	1.0764	0.73%

Algorithm	MAE	MSE	meanMASE	Forecast Bias
simple	0.0854	0.0308	1.0787	1.41%
simple+time_varying	0.0851	0.0310	1.0764	0.73%
simple+static+time_varying	0.0843	0.0297	1.0685	0.94%

Algorithm	MAE	MSE	meanMASE	Forecast Bias
simple	0.0854	0.0308	1.0787	1.41%
simple+time_varying	0.0851	0.0310	1.0764	0.73%
simple+static+time_varying	0.0843	0.0297	1.0685	0.94%
simple+static+time_varying+scale	0.0822	0.0298	1.0395	-3.20%

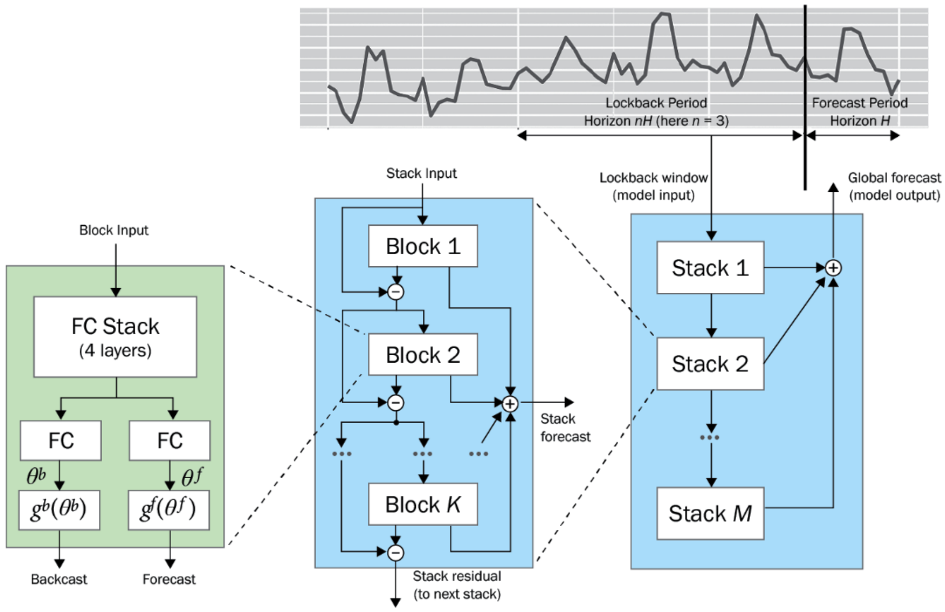


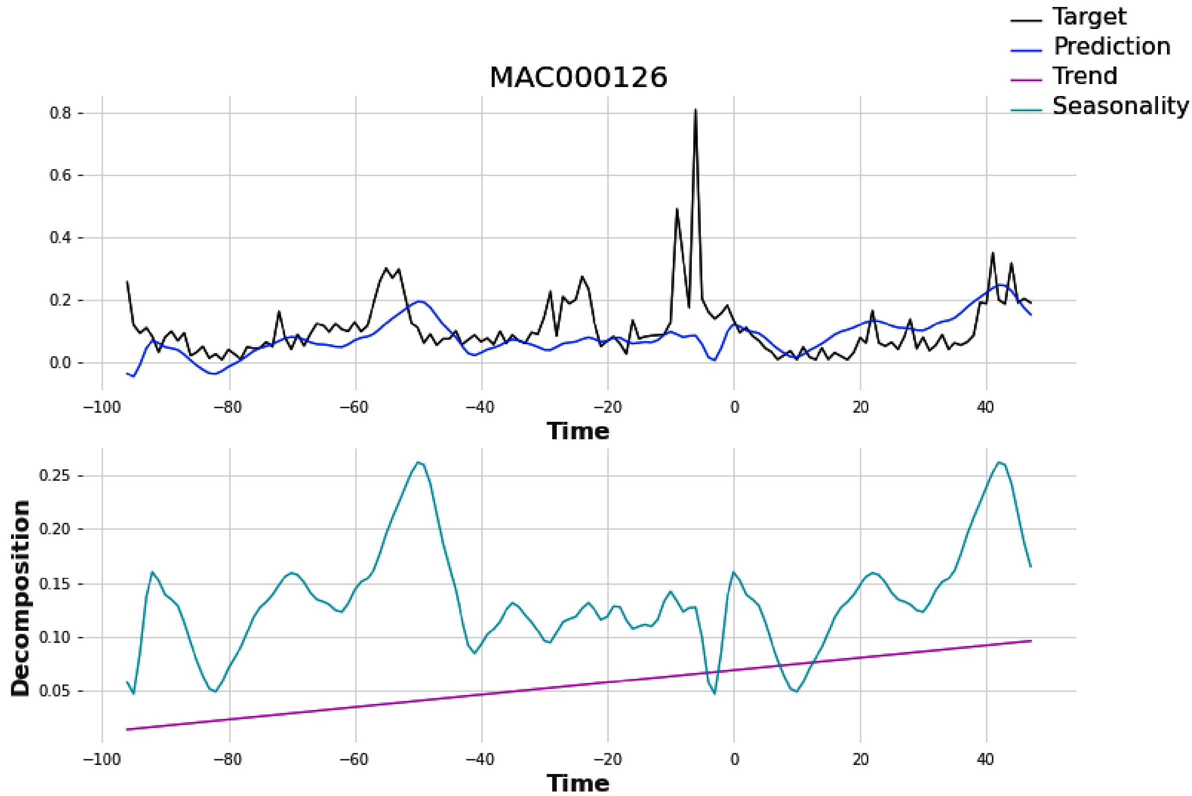


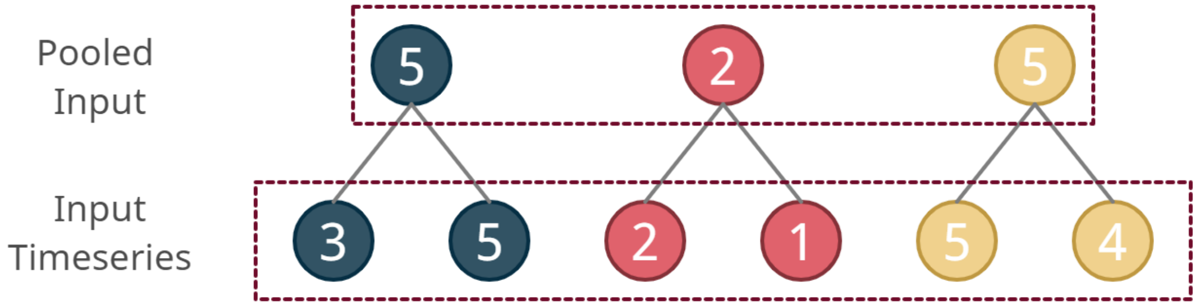


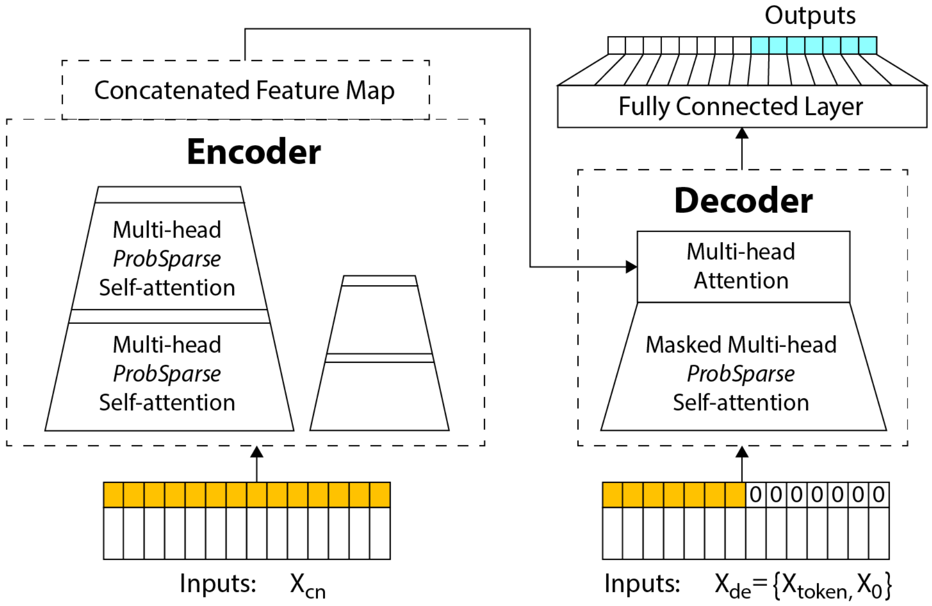
Algorithm	MAE	MSE	meanMASE	Forecast Bias
simple	0.0854	0.0308	1.0787	1.41%
simple+time_varying	0.0851	0.0310	1.0764	0.73%
simple+static+time_varying	0.0843	0.0297	1.0685	0.94%
simple+static+time_varying+scale	0.0822	0.0298	1.0395	-3.20%
simple+static+time_varying+num_sampler	0.0815	0.0297	1.0372	-4.06%

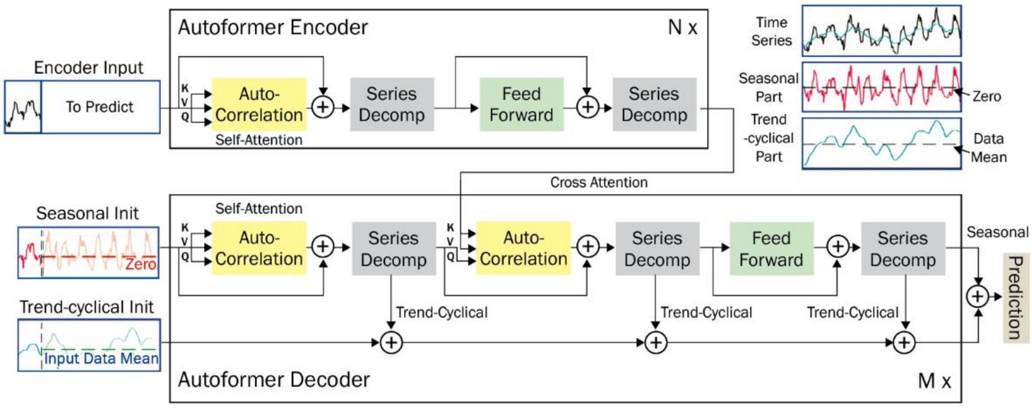
Chapter 16: Specialized Deep Learning Architectures for Forecasting

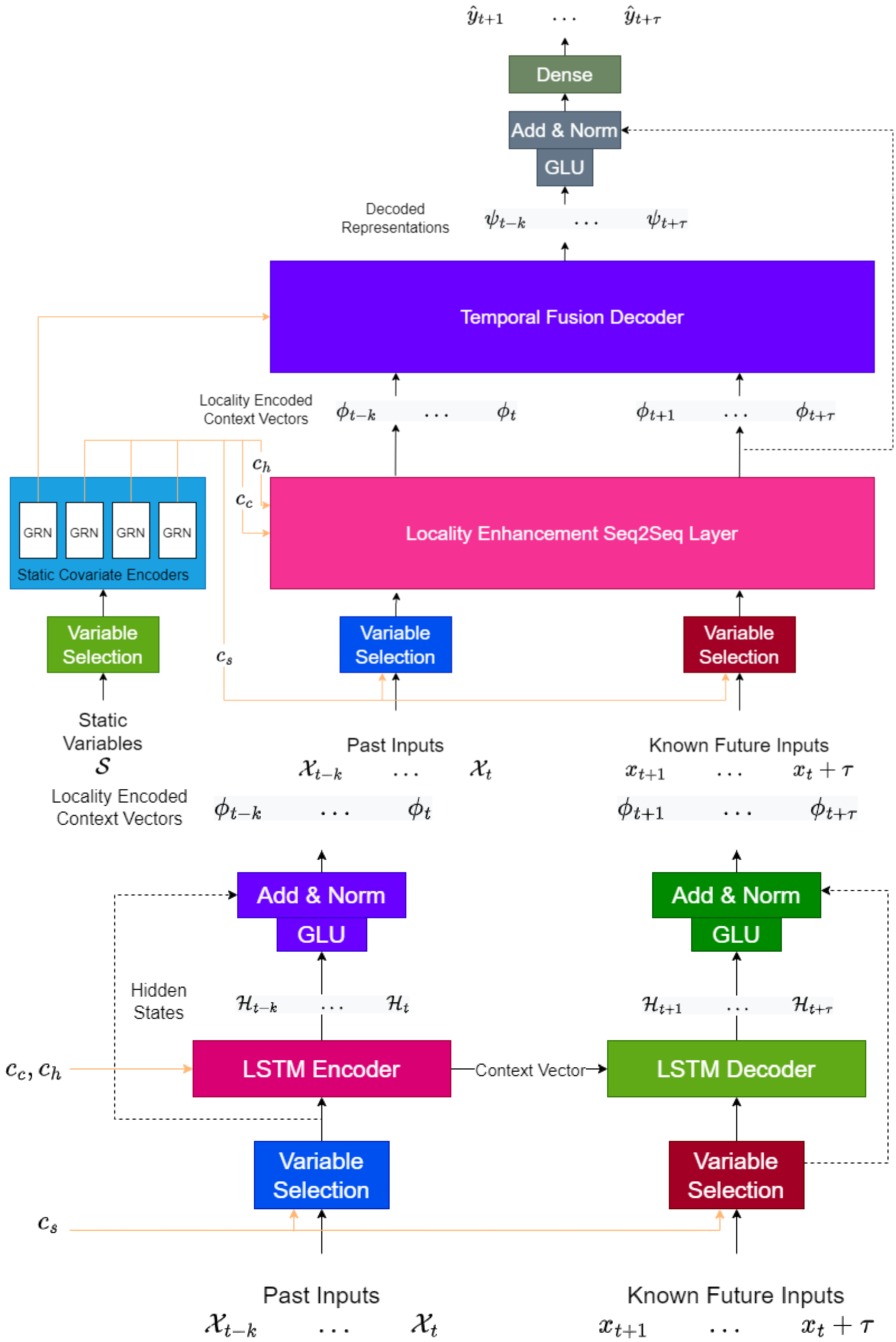




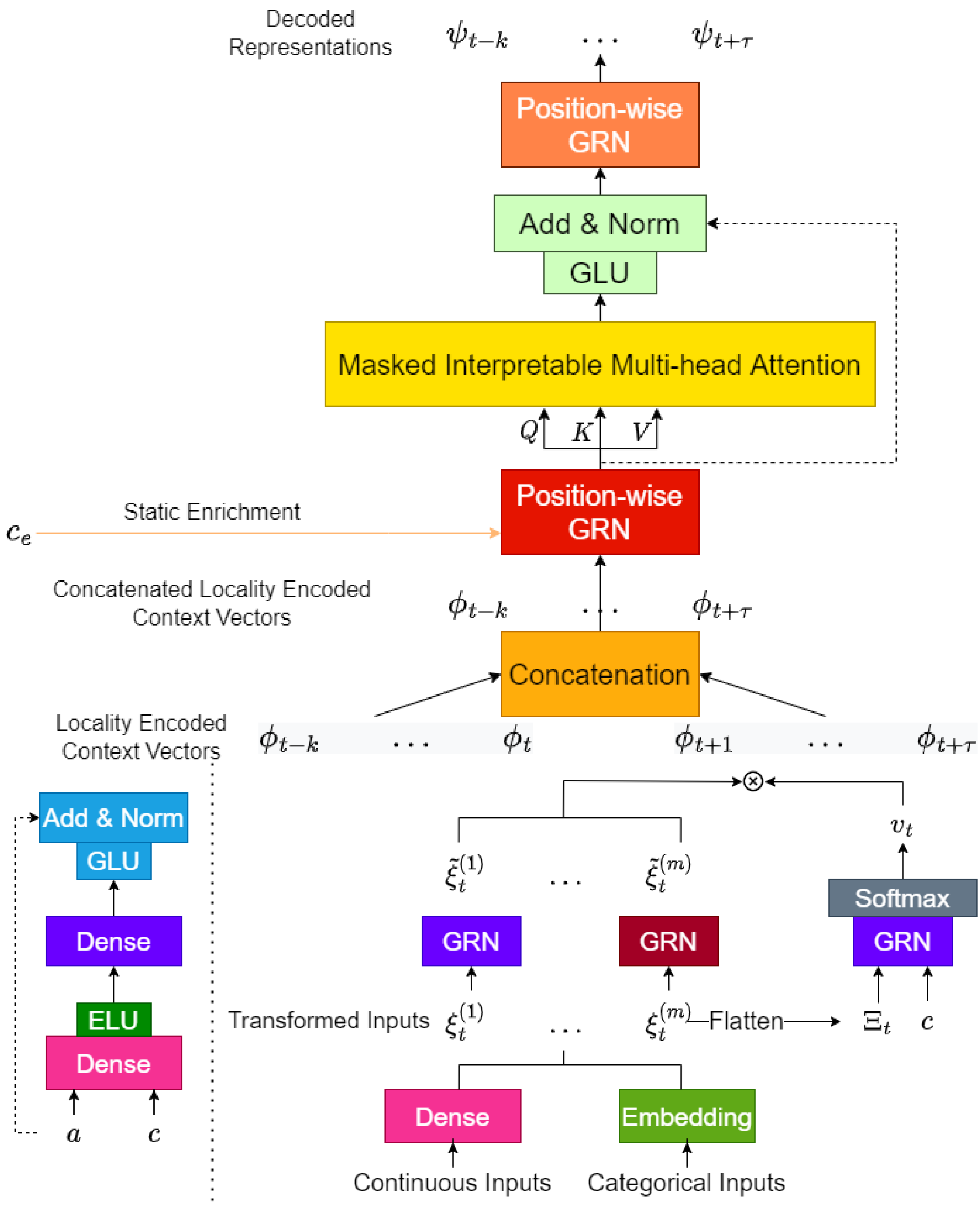

Pooled
Input
Input
Timeseries
5
2
5
3
5
2
1
5
4

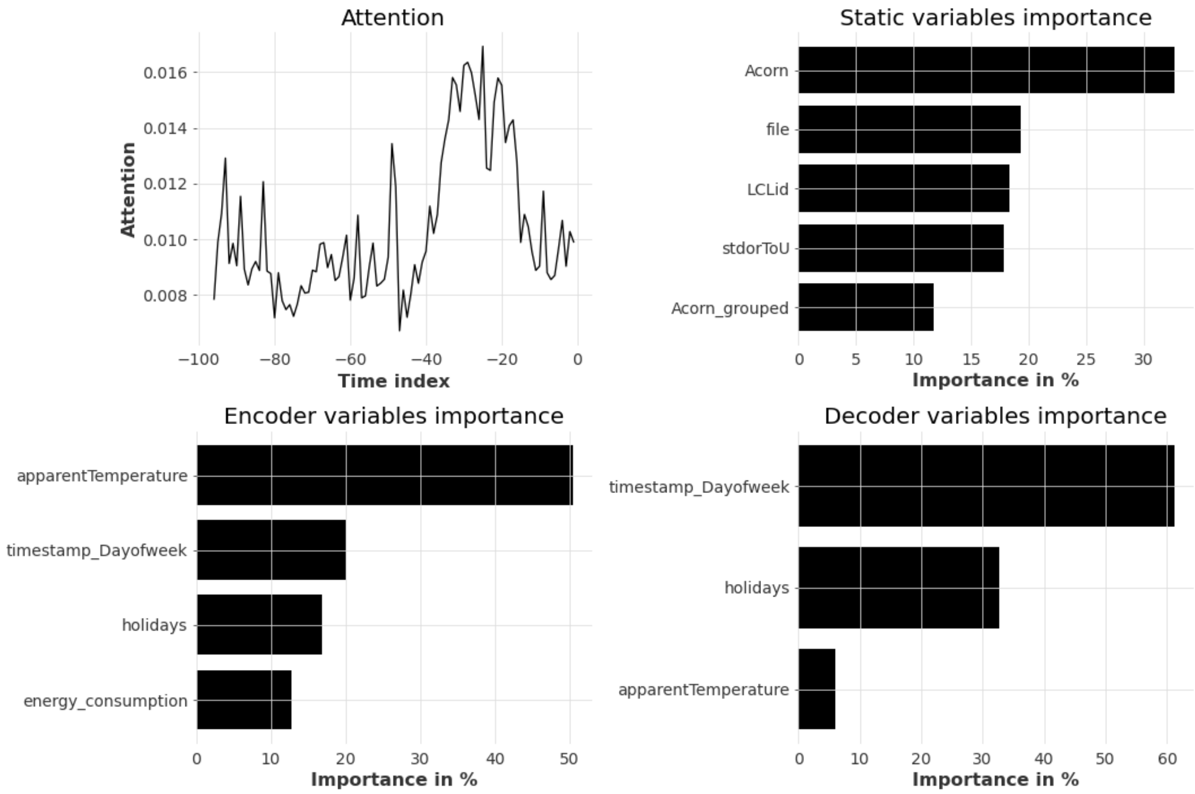

Outputs
Concatenated Feature Map
Fully Connected Layer
Encoder
Multi-head
ProbSparse
Self-attention
Multi-head
ProbSparse
Self-attention
Decoder
Multi-head
Attention
Masked Multi-head
ProbSparse
Self-attention
0 0 0 0 0 0 0
Inputs: X_{en}
Inputs: $X_{de}=\{X_{token}, X_0\}$


Autoformer Encoder
N x
Encoder Input
To Predict
K
V
Q
Auto-
Correlation
Self-Attention
Series
Decomp
Feed
Forward
Series
Decomp
Time
Series
Seasonal
Part
Zero
Trend
-cyclical
Part
Data
Mean
Cross Attention
Seasonal Init
Zero
Trend-cyclical Init
Input Data Mean
Self-Attention
Auto-
Correlation
Series
Decomp
Auto-
Correlation
Series
Decomp
Feed
Forward
Series
Decomp
Seasonal
Prediction
Trend-Cyclical
Trend-Cyclical
Trend-Cyclical
Autoformer Decoder
M x

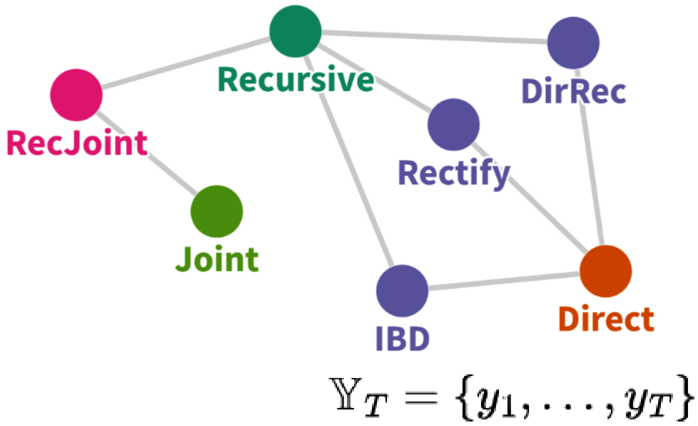


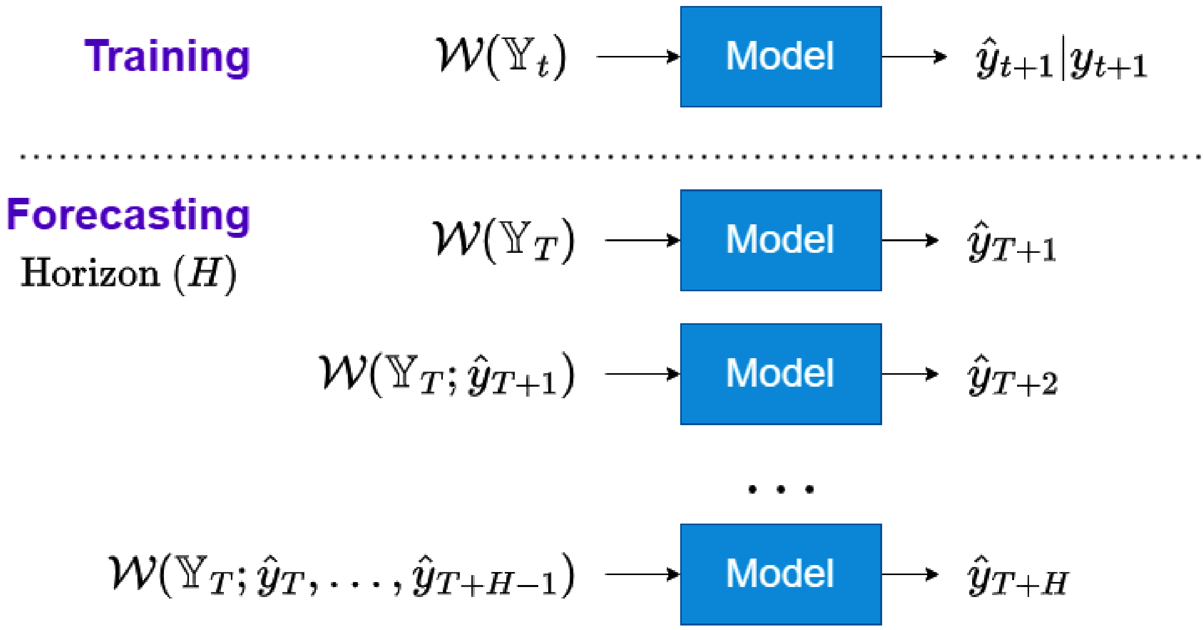
$\hat{y}_{t+1}$ ... $\hat{y}_{t+\tau}$
Dense
Add & Norm
GLU
Decoded Representations
ψ_{t-k} ... $\psi_{t+\tau}$
Temporal Fusion Decoder
Locality Encoded Context Vectors
ϕ_{t-k} ... ϕ_t
ϕ_{t+1} ... $\phi_{t+\tau}$
c_h
c_c
GRN
GRN
GRN
GRN
Static Covariate Encoders
Locality Enhancement Seq2Seq Layer
Variable Selection
c_s
Variable Selection
Variable Selection
Static Variables
$\mathcal{S}$
Past Inputs
$\mathcal{X}_{t-k}$ ... $\mathcal{X}_t$
Known Future Inputs
x_{t+1} ... $x_t + \tau$
Locality Encoded Context Vectors
ϕ_{t-k} ... ϕ_t
ϕ_{t+1} ... $\phi_{t+\tau}$
Add & Norm
GLU
Add & Norm
GLU
Hidden States
$\mathcal{H}_{t-k}$ ... $\mathcal{H}_t$
$\mathcal{H}_{t+1}$ ... $\mathcal{H}_{t+\tau}$
c_c, c_h
LSTM Encoder
Context Vector
LSTM Decoder
Variable Selection
Variable Selection
c_s
Past Inputs
$\mathcal{X}_{t-k}$ ... $\mathcal{X}_t$
Known Future Inputs
x_{t+1} ... $x_t + \tau$

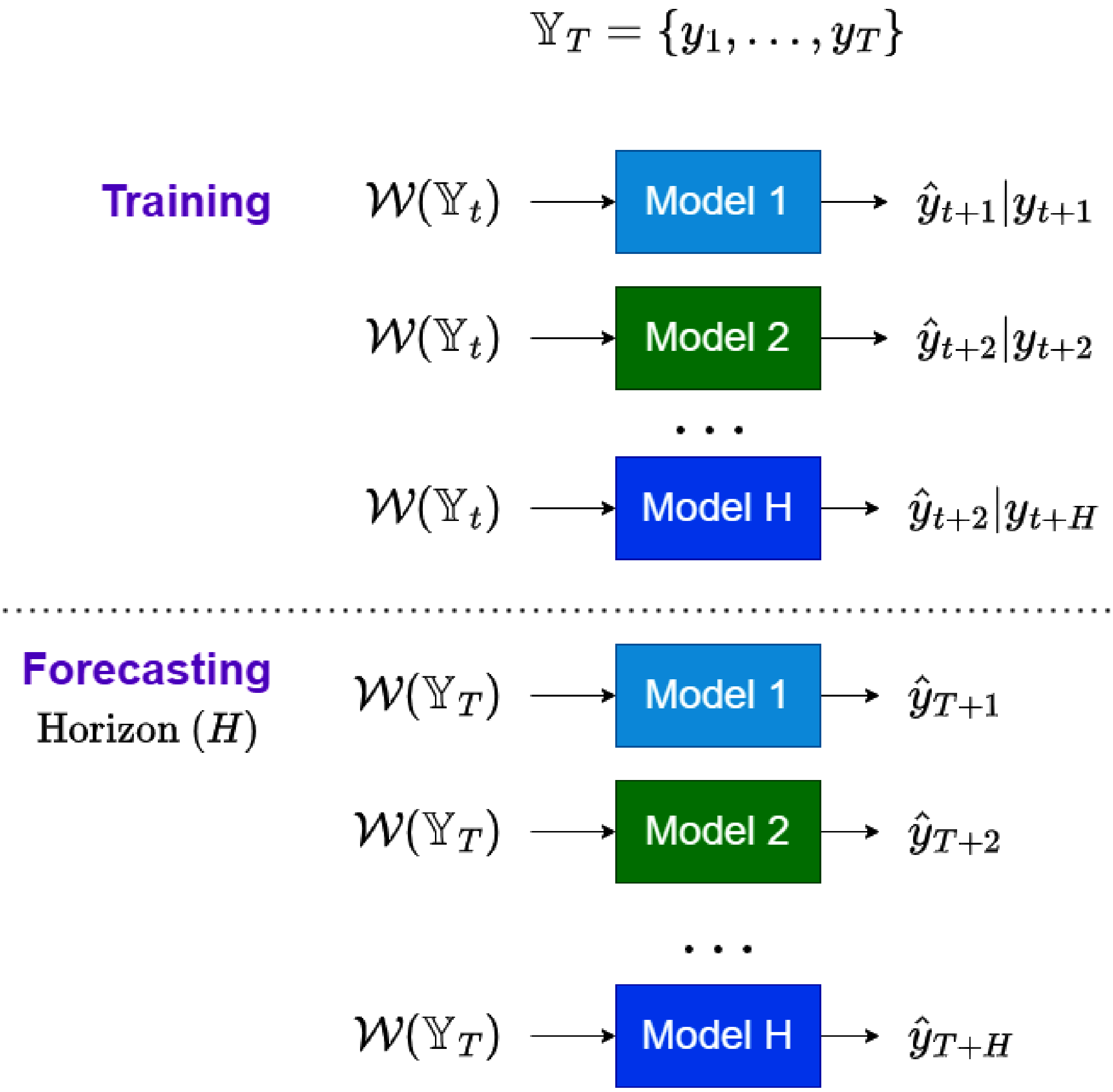

Decoded
Representations
ψ_{t-k} ... $\psi_{t+\tau}$
Position-wise
GRN
Add & Norm
GLU
Masked Interpretable Multi-head Attention
Q K V
Static Enrichment
c_e
Position-wise
GRN
Concatenated Locality Encoded
Context Vectors
ϕ_{t-k} ... $\phi_{t+\tau}$
Concatenation
Locality Encoded
Context Vectors
ϕ_{t-k} ... ϕ_t ϕ_{t+1} ... $\phi_{t+\tau}$
v_t
$\tilde{\xi}_t^{(1)}$ ... $\tilde{\xi}_t^{(m)}$
Softmax
GRN
GRN
GRN
Add & Norm
GLU
Dense
ELU
Dense
a c
Transformed Inputs
$\xi_t^{(1)}$ ... $\xi_t^{(m)}$
Flatten
Ξ_t c
Dense
Embedding
Continuous Inputs
Categorical Inputs


Attention
Attention
Time index
0.016
0.014
0.012
0.010
0.008
−100
−80
−60
−40
−20
0
Static variables importance
Acorn
file
LCLid
stdorToU
Acorn_grouped
Importance in %
0
5
10
15
20
25
30
Encoder variables importance
apparentTemperature
timestamp_Dayofweek
holidays
energy_consumption
Importance in %
0
10
20
30
40
50
Decoder variables importance
timestamp_Dayofweek
holidays
apparentTemperature
Importance in %
0
10
20
30
40
50
60

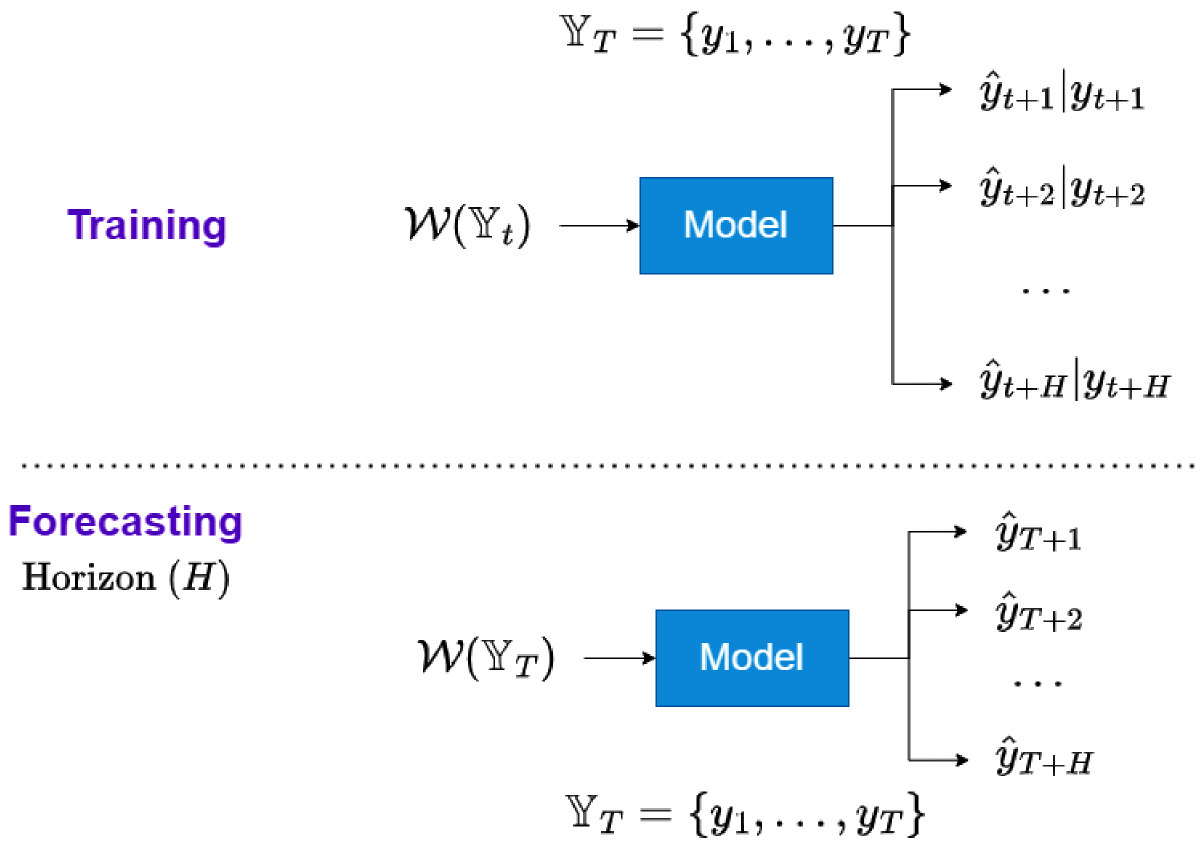
Chapter 17: Multi-Step Forecasting



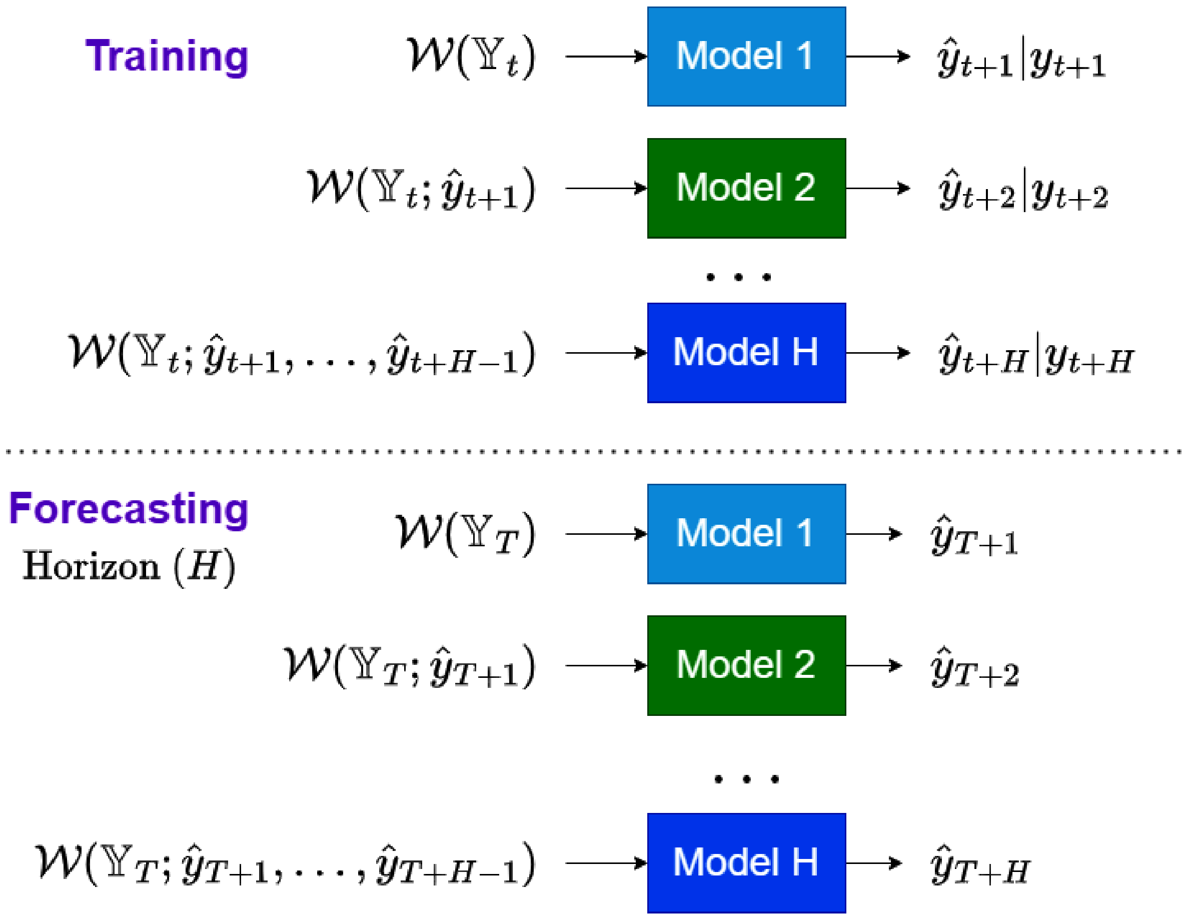




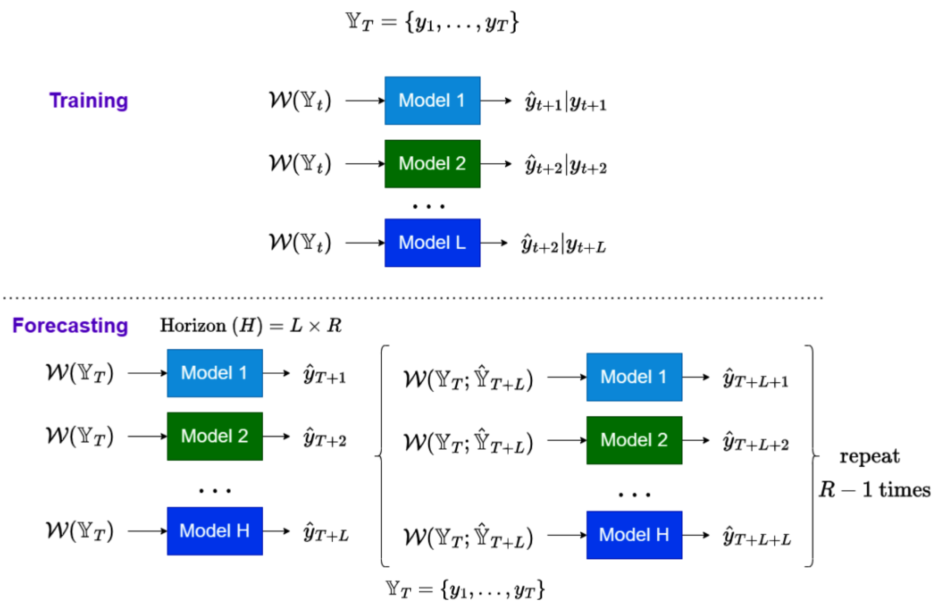
$\mathbb{Y}_T = \{y_1, \ldots, y_T\}$
Training
$\mathcal{W}(\mathbb{Y}_t)$
Model 1
$\hat{y}_{t+1}|y_{t+1}$
$\mathcal{W}(\mathbb{Y}_t)$
Model 2
$\hat{y}_{t+2}|y_{t+2}$
$\mathcal{W}(\mathbb{Y}_t)$
Model H
$\hat{y}_{t+2}|y_{t+H}$
Forecasting
Horizon (H)
$\mathcal{W}(\mathbb{Y}_T)$
Model 1
$\hat{y}_{T+1}$
$\mathcal{W}(\mathbb{Y}_T)$
Model 2
$\hat{y}_{T+2}$
$\mathcal{W}(\mathbb{Y}_T)$
Model H
$\hat{y}_{T+H}$

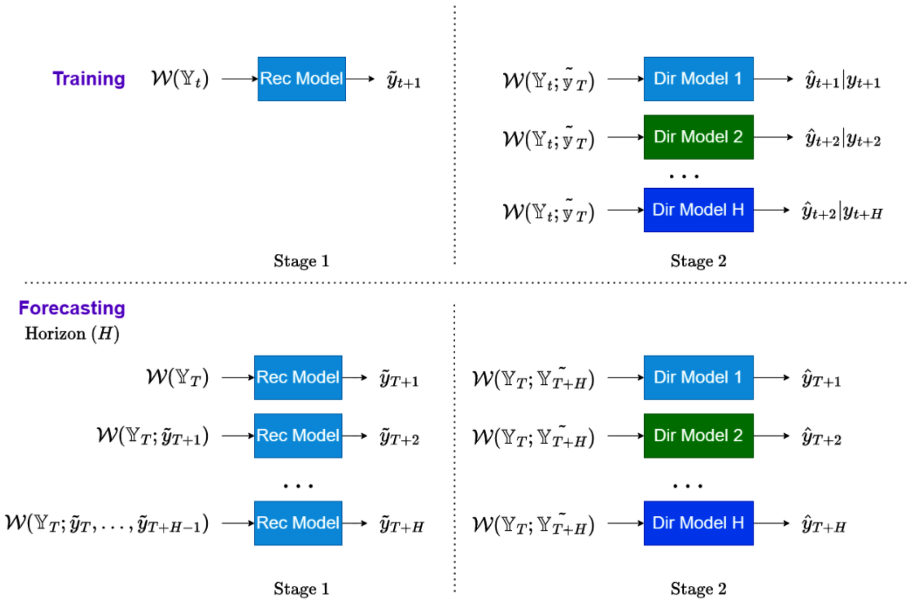


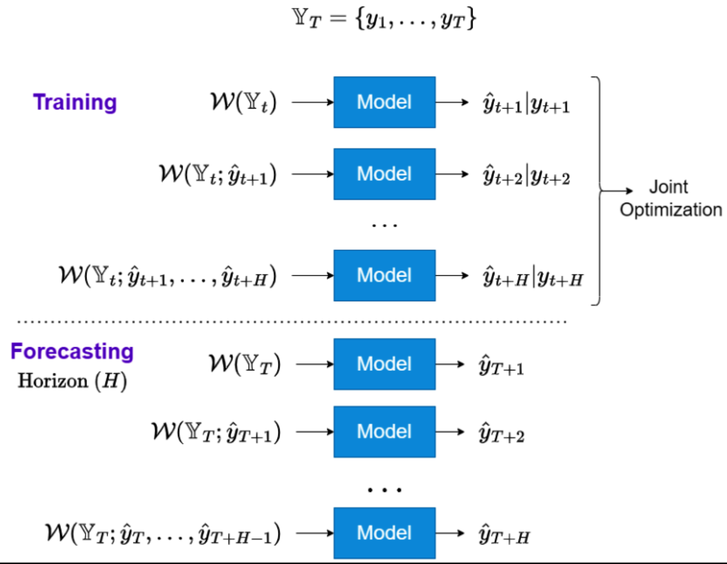
$\mathbb{Y}_T = \{y_1, \dots, y_T\}$
Training
$\mathcal{W}(\mathbb{Y}_t)$
Model
$\hat{y}_{t+1}|y_{t+1}$
$\hat{y}_{t+2}|y_{t+2}$
$\dots$
$\hat{y}_{t+H}|y_{t+H}$
Forecasting
Horizon (H)
$\mathcal{W}(\mathbb{Y}_T)$
Model
$\hat{y}_{T+1}$
$\hat{y}_{T+2}$
$\dots$
$\hat{y}_{T+H}$
$\mathbb{Y}_T = \{y_1, \dots, y_T\}$



Training
$\mathcal{W}(\mathbb{Y}_t)$
Model 1
$\hat{y}_{t+1}|y_{t+1}$
$\mathcal{W}(\mathbb{Y}_t; \hat{y}_{t+1})$
Model 2
$\hat{y}_{t+2}|y_{t+2}$
$\dots$
$\mathcal{W}(\mathbb{Y}_t; \hat{y}_{t+1}, \dots, \hat{y}_{t+H-1})$
Model H
$\hat{y}_{t+H}|y_{t+H}$
Forecasting
Horizon (H)
$\mathcal{W}(\mathbb{Y}_T)$
Model 1
$\hat{y}_{T+1}$
$\mathcal{W}(\mathbb{Y}_T; \hat{y}_{T+1})$
Model 2
$\hat{y}_{T+2}$
$\dots$
$\mathcal{W}(\mathbb{Y}_T; \hat{y}_{T+1}, \dots, \hat{y}_{T+H-1})$
Model H
$\hat{y}_{T+H}$

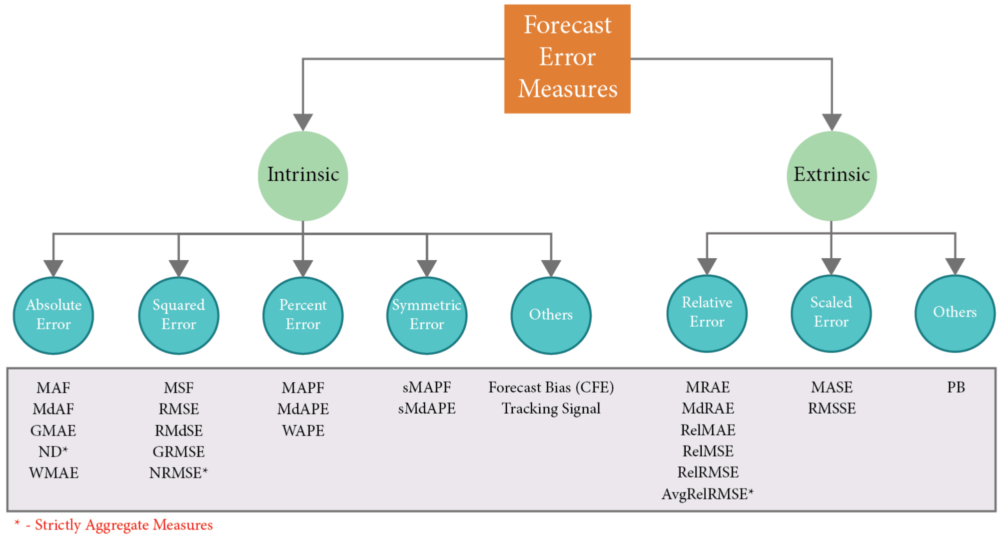

$\mathbb{Y}_T = \{y_1, \dots, y_T\}$
Training
$\mathcal{W}(\mathbb{Y}_t)$
Model 1
$\hat{y}_{t+1}|y_{t+1}$
Model 2
$\hat{y}_{t+2}|y_{t+2}$
Model L
$\hat{y}_{t+2}|y_{t+L}$
Forecasting
Horizon $(H) = L \times R$
$\mathcal{W}(\mathbb{Y}_T)$
Model 1
$\hat{y}_{T+1}$
Model 2
$\hat{y}_{T+2}$
Model H
$\hat{y}_{T+L}$
$\mathcal{W}(\mathbb{Y}_T; \hat{\mathbb{Y}}_{T+L})$
$\hat{y}_{T+L+1}$
$\hat{y}_{T+L+2}$
$\hat{y}_{T+L+L}$
repeat
$R-1$ times
$\mathbb{Y}_T = \{y_1, \dots, y_T\}$

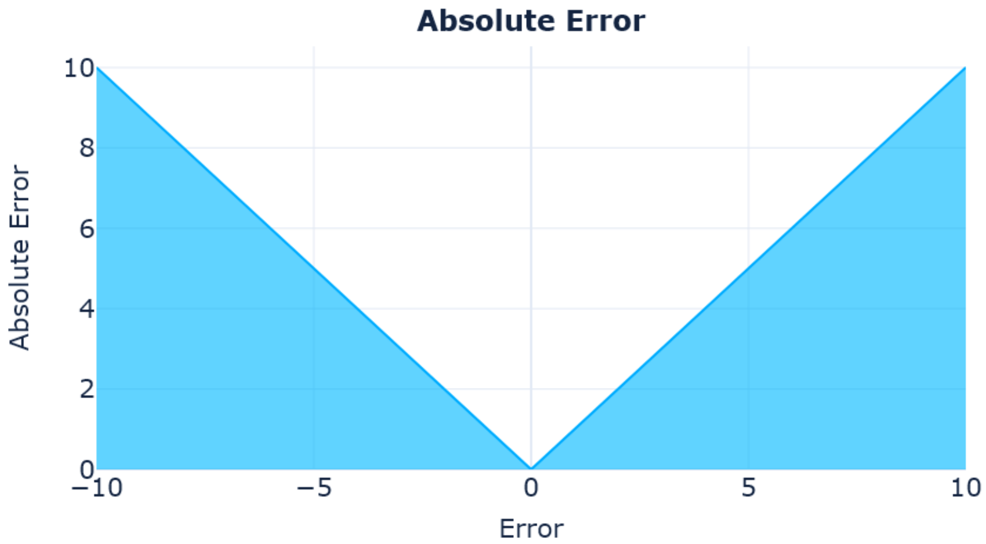

Training
$\mathcal{W}(\mathbb{Y}_t)$
Rec Model
$\tilde{y}_{t+1}$
$\mathcal{W}(\mathbb{Y}_t; \tilde{\mathbb{Y}}_T)$
Dir Model 1
$\hat{y}_{t+1}|y_{t+1}$
Dir Model 2
$\hat{y}_{t+2}|y_{t+2}$
Dir Model H
$\hat{y}_{t+2}|y_{t+H}$
Stage 1
Stage 2
Forecasting
Horizon (H)
$\mathcal{W}(\mathbb{Y}_T)$
$\tilde{y}_{T+1}$
$\mathcal{W}(\mathbb{Y}_T; \tilde{y}_{T+1})$
$\tilde{y}_{T+2}$
$\mathcal{W}(\mathbb{Y}_T; \tilde{y}_T, \dots, \tilde{y}_{T+H-1})$
$\tilde{y}_{T+H}$
$\mathcal{W}(\mathbb{Y}_T; \tilde{\mathbb{Y}}_{T+H})$
$\hat{y}_{T+1}$
$\hat{y}_{T+2}$
$\hat{y}_{T+H}$
Stage 1
Stage 2

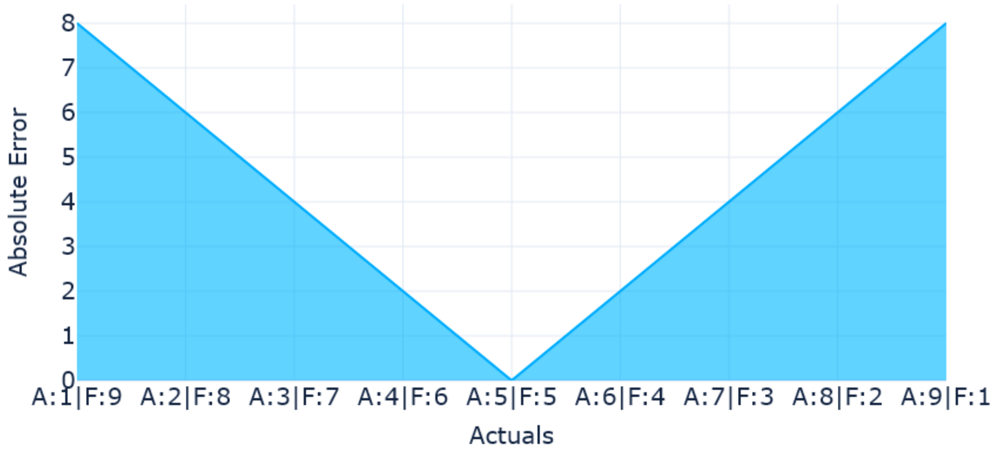


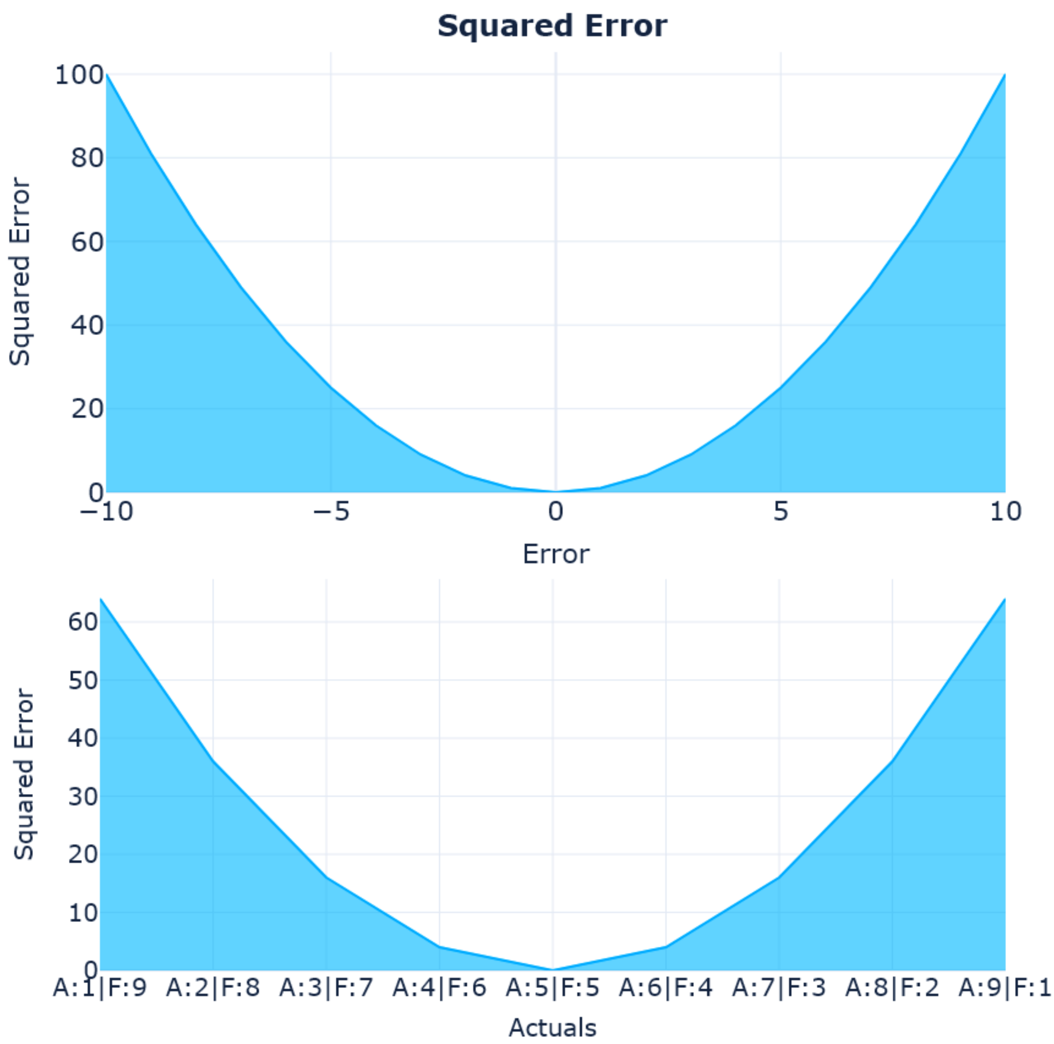
Strategy	# of Models	Type	Output Size	Training Time	Prediction Time
Recursive	1	S.O	1	$1 \times T_{so}$	$H \times I_{so}$
Direct	H	S.O	1	$H \times T_{so}$	$H \times I_{so}$
DirRec	H	S.O	1	$H \times T_{so}$	$H \times I_{so}$
IBD	L	S.O	1	$L \times T_{so}$	$H \times I_{so}$
Rectify	H+1	S.O	1	$(H+1) \times T_{so}$	$2 \times H \times I_{so}$
Joint	1	M.O	H	$1 \times T_{mo}$	$1 \times I_{mo}$
RecJoint	1	M.O	1	$1 \times (T_{mo} + \delta)$	$H \times I_{mo}$

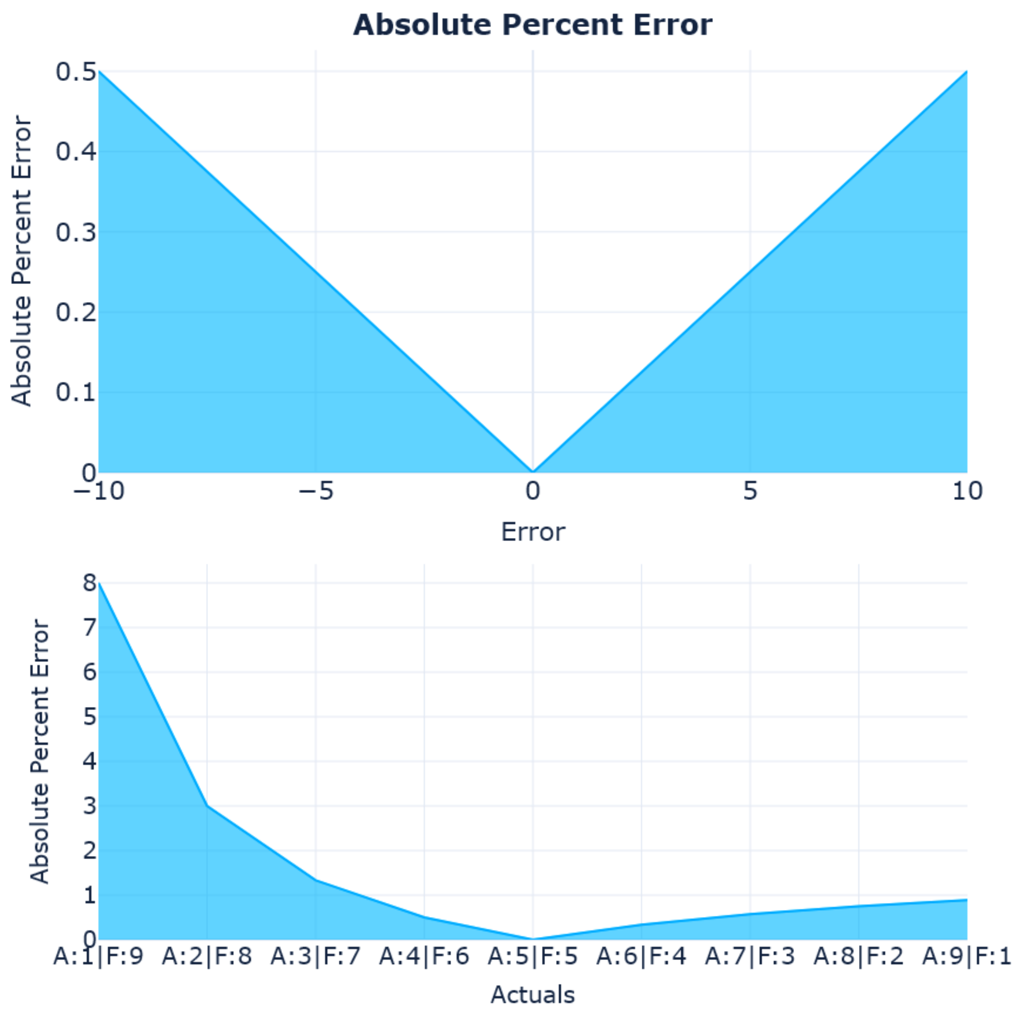
Chapter 18: Evaluating Forecasts – Forecast Metrics



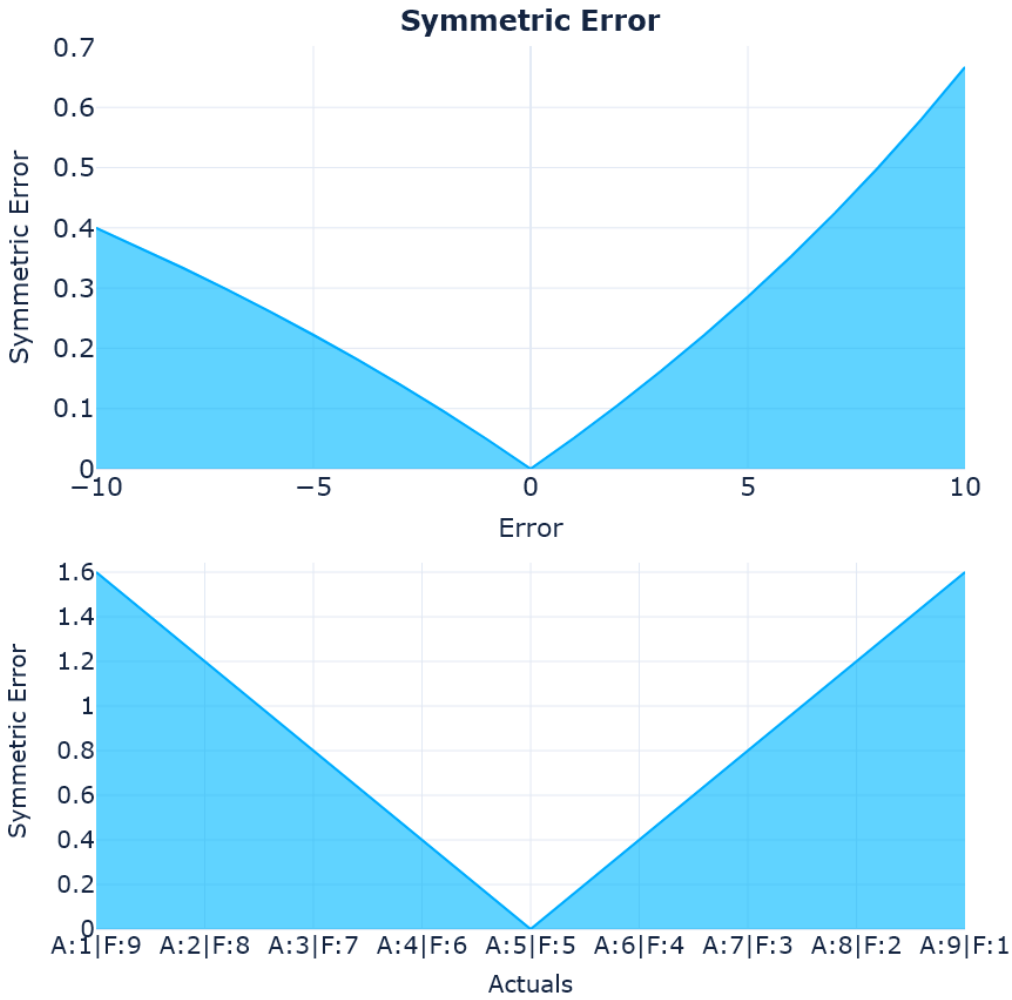


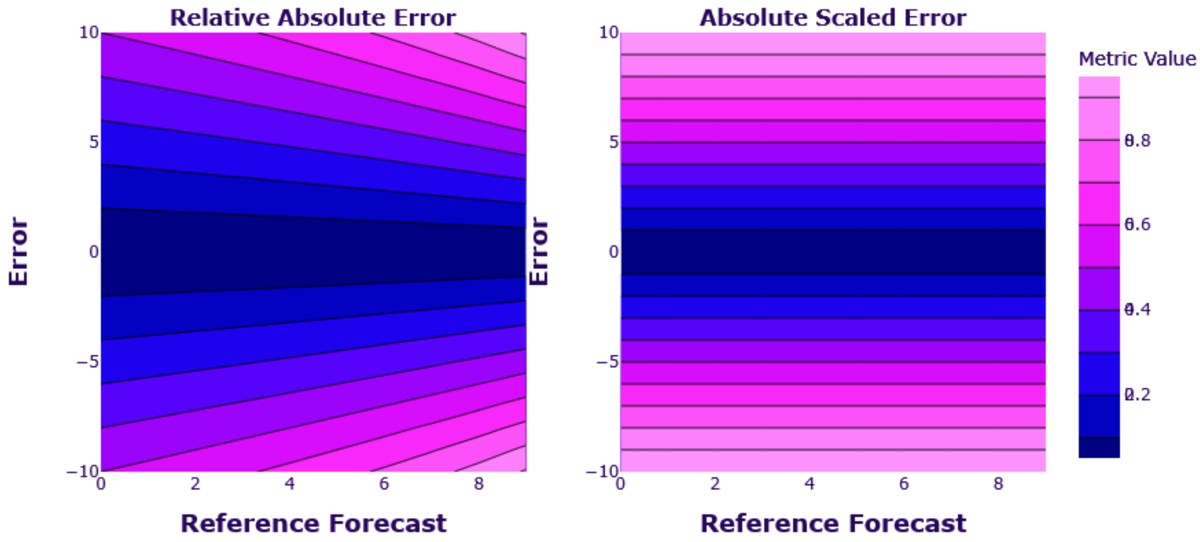


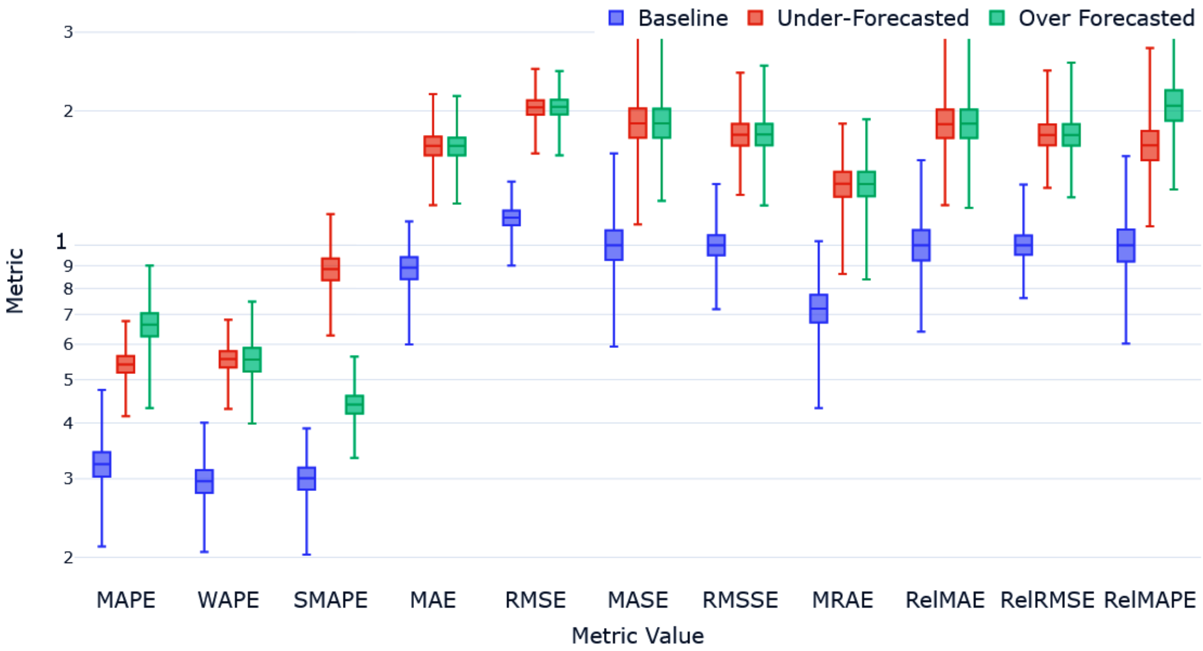

Squared Error
Squared Error
100
80
60
40
20
0
−10
−5
0
5
10
Error
Squared Error
60
50
40
30
20
10
0
A:1|F:9
A:2|F:8
A:3|F:7
A:4|F:6
A:5|F:5
A:6|F:4
A:7|F:3
A:8|F:2
A:9|F:1
Actuals

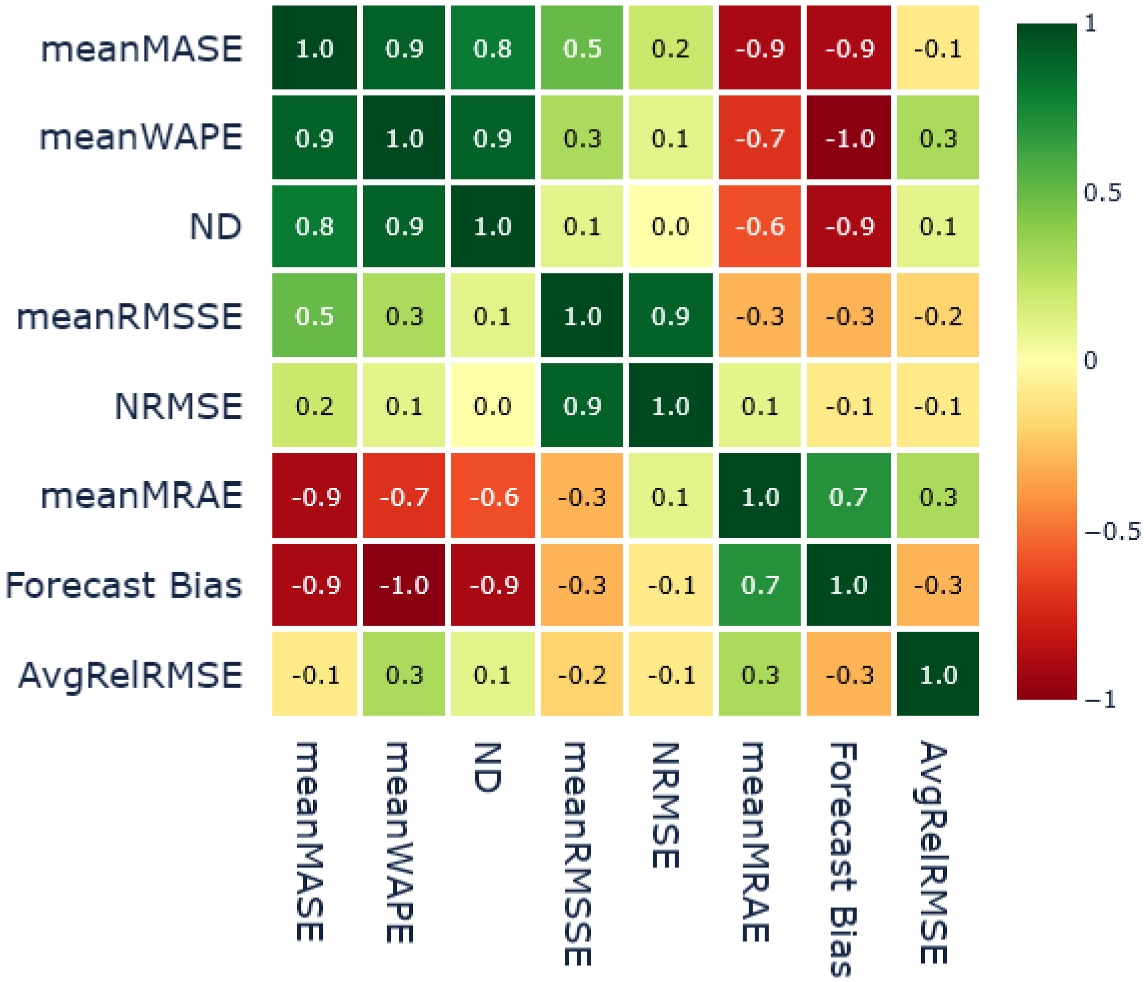


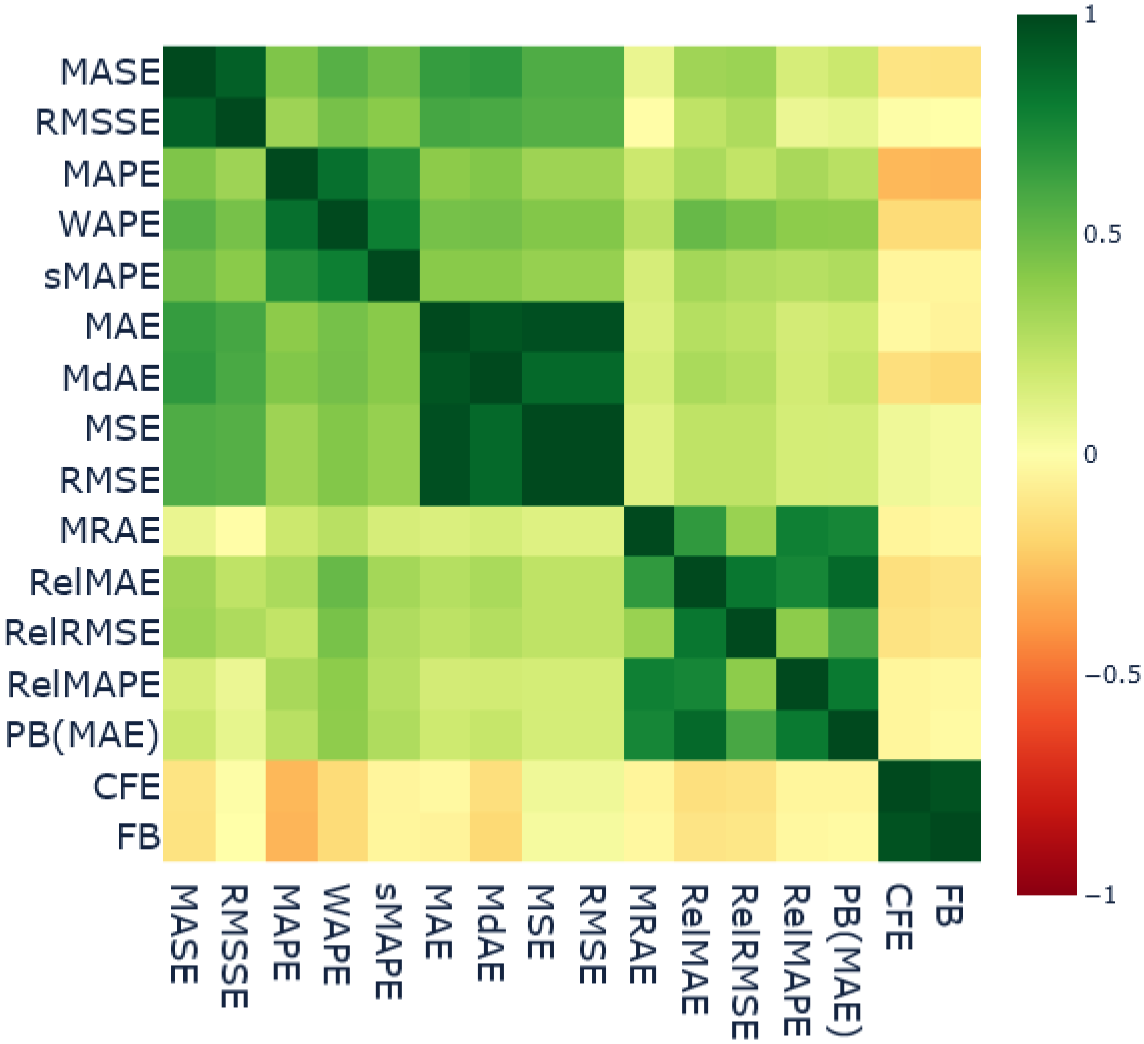
Absolute Percent Error
Absolute Percent Error
0.5
0.4
0.3
0.2
0.1
0
−10
−5
0
5
10
Error
Absolute Percent Error
8
7
6
5
4
3
2
1
0
A:1|F:9
A:2|F:8
A:3|F:7
A:4|F:6
A:5|F:5
A:6|F:4
A:7|F:3
A:8|F:2
A:9|F:1
Actuals


Symmetric Error
Symmetric Error
Error
Symmetric Error
A:1|F:9
A:2|F:8
A:3|F:7
A:4|F:6
A:5|F:5
A:6|F:4
A:7|F:3
A:8|F:2
A:9|F:1
Actuals


Relative Absolute Error
Absolute Scaled Error
Error
Error
Reference Forecast
Reference Forecast
Metric Value

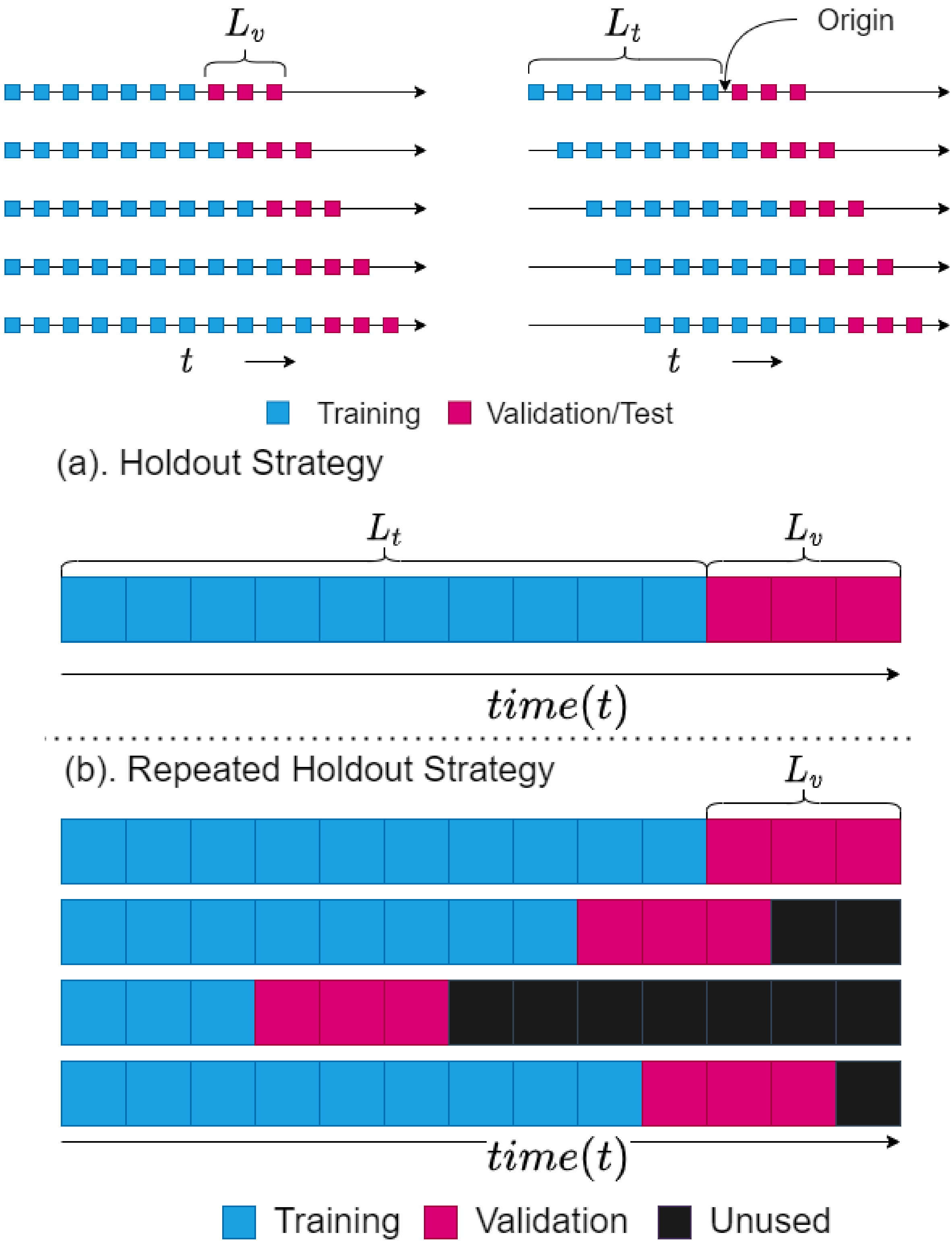

Baseline
Under-Forecasted
Over Forecasted
Metric
3
2
1
9
8
7
6
5
4
3
2
MAPE
WAPE
SMAPE
MAE
RMSE
MASE
RMSSE
MRAE
RelMAE
RelRMSE
RelMAPE
Metric Value

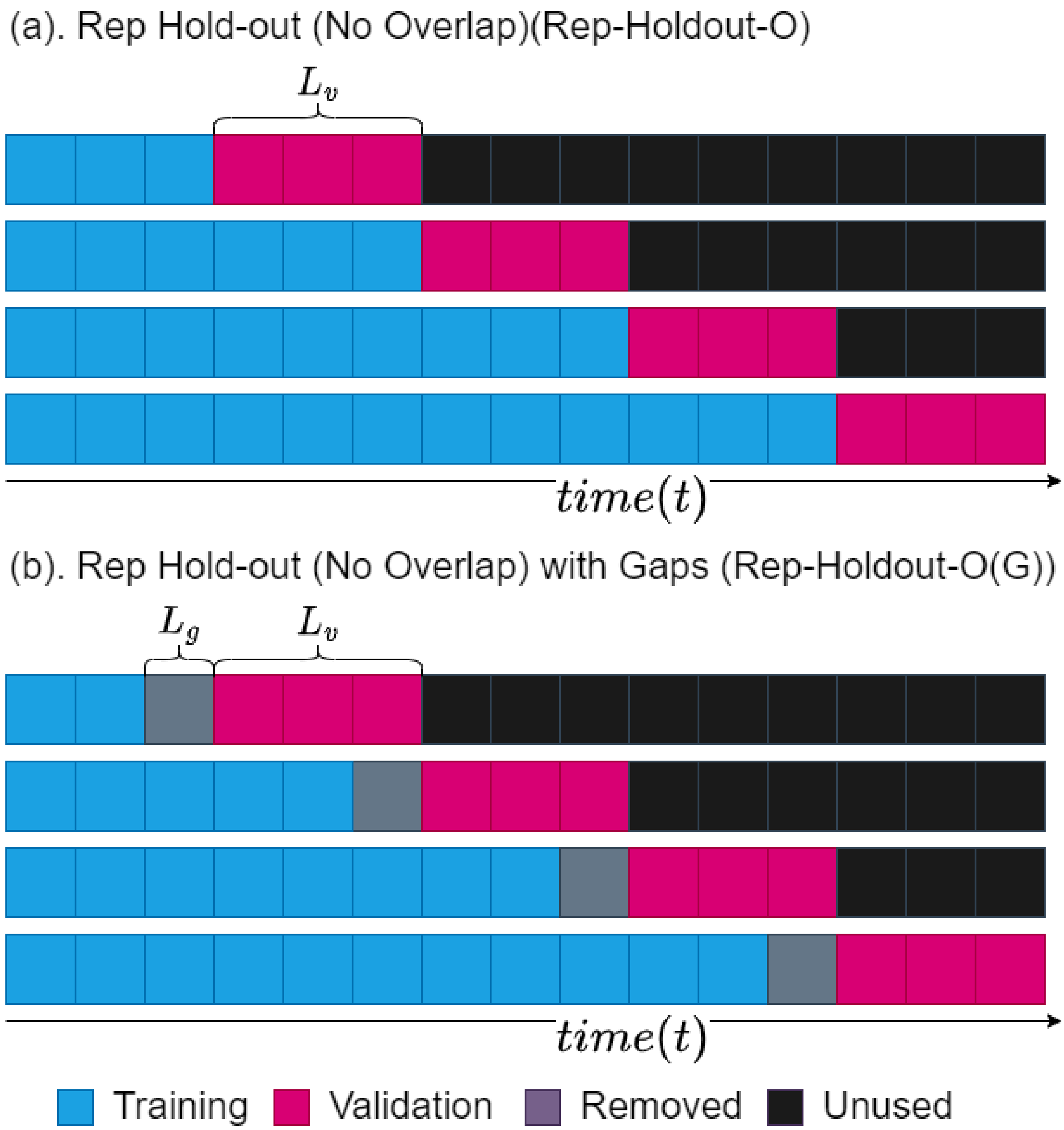

meanMASE 1.0 0.9 0.8 0.5 0.2 -0.9 -0.9 -0.1
meanWAPE 0.9 1.0 0.9 0.3 0.1 -0.7 -1.0 0.3
ND 0.8 0.9 1.0 0.1 0.0 -0.6 -0.9 0.1
meanRMSSE 0.5 0.3 0.1 1.0 0.9 -0.3 -0.3 -0.2
NRMSE 0.2 0.1 0.0 0.9 1.0 0.1 -0.1 -0.1
meanMRAE -0.9 -0.7 -0.6 -0.3 0.1 1.0 0.7 0.3
Forecast Bias -0.9 -1.0 -0.9 -0.3 -0.1 0.7 1.0 -0.3
AvgRelRMSE -0.1 0.3 0.1 -0.2 -0.1 0.3 -0.3 1.0
meanMASE
meanWAPE
ND
meanRMSSE
NRMSE
meanMRAE
Forecast Bias
AvgRelRMSE
1
0.5
0
−0.5
−1

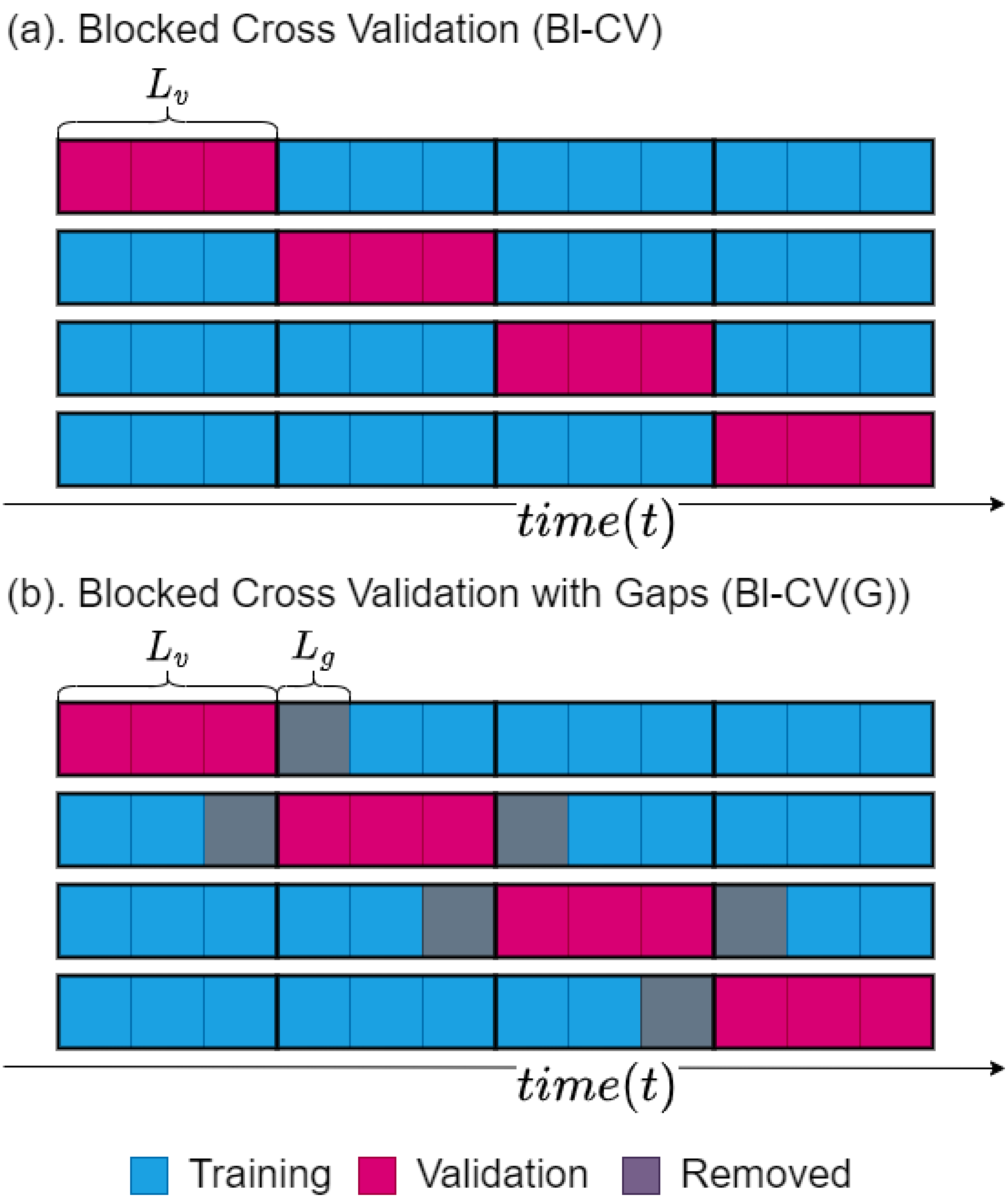


MASE
RMSSE
MAPE
WAPE
sMAPE
MAE
MdAE
MSE
RMSE
MRAE
RelMAE
RelRMSE
RelMAPE
PB(MAE)
CFE
FB
1
0.5
0
−0.5
−1

Chapter 19: Evaluating Forecasts – Validation Strategies




(a). Rep Hold-out (No Overlap)(Rep-Holdout-O)
L_v
$time(t)$
(b). Rep Hold-out (No Overlap) with Gaps (Rep-Holdout-O(G))
L_g
L_v
$time(t)$
Training
Validation
Removed
Unused


(a). Blocked Cross Validation (Bl-CV)
L_v
$time(t)$
(b). Blocked Cross Validation with Gaps (Bl-CV(G))
L_v
L_g
$time(t)$
Training
Validation
Removed