Timeseries prediction on energy consumption employing weather data

Anna Kirchner, Arnau Morancho Tardà, Christopher Pfeiffer, Neus Rodeja Ferrer, Chrysoula Stathaki

Agenda

- Data
- Objective
- Approach 1: Weather prediction + regression
- Approach 2: RNN
- Results/Evaluation

Data

- Kaggle dataset from Spain
- Hourly observations during 4 years (2015-2018)
- 14 weather features measured in 5 cities

  (Valencia, Madrid, Bilbao, Sevilla and Barcelona):

```
['dt_iso', 'city_name', 'temp', 'temp_min', 'temp_max', 'pressure',
'humidity', 'wind_speed', 'wind_deg', 'rain_1h', 'rain_3h', 'snow_3h',
'clouds_all', 'weather_id', 'weather_main', 'weather_description']
```

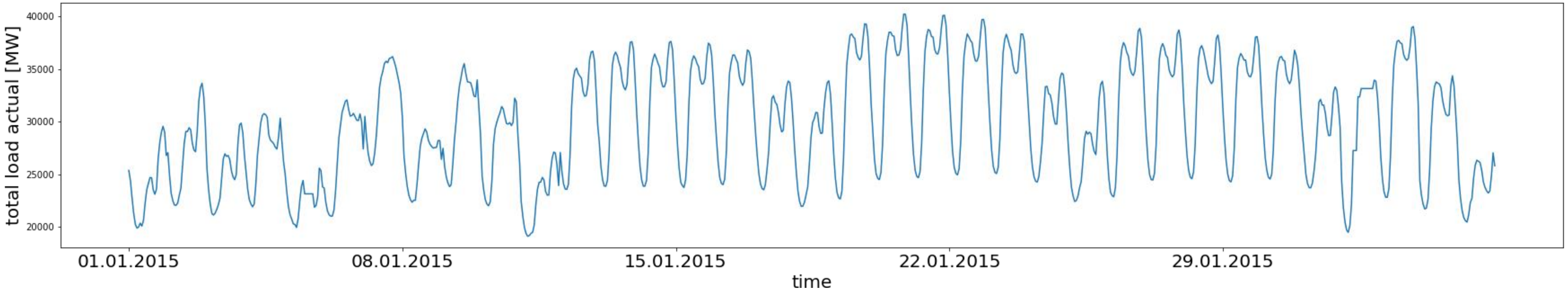
- 28 energy features for the whole country:

```
['time', 'generation biomass', 'generation fossil brown coal/lignite','generation fossil coal-derived gas',
'generation fossil gas','generation fossil hard coal', 'generation fossil oil','generation fossil oil shale',
'generation fossil peat','generation geothermal', 'generation hydro pumped storage aggregated','generation hydro
pumped storage consumption','generation hydro run-of-river and poundage','generation hydro water reservoir',
'generation marine','generation nuclear', 'generation other', 'generation other renewable','generation solar',
'generation waste', 'generation wind offshore','generation wind onshore', 'forecast solar day ahead','forecast wind
offshore eday ahead', 'forecast wind onshore day ahead','total load forecast', 'total load actual', 'price day ahead',
'price actual']
```

Objective

- Predict an energy variable in Spain from weather information
- Selecting target variable
  - Wind onshore generation, solar generation, energy consumption (total load actual)

➡ **Goal: Predict energy consumption in Spain from weather information 1 day ahead**



- Using different approaches and models + comparing their performance

Data Preprocessing

- Join weather variables for all cities and energy dataset by time
- Remove duplicates
- Categorical variables:
  - Remove weather_id and weather_description
  - Encode weather_main

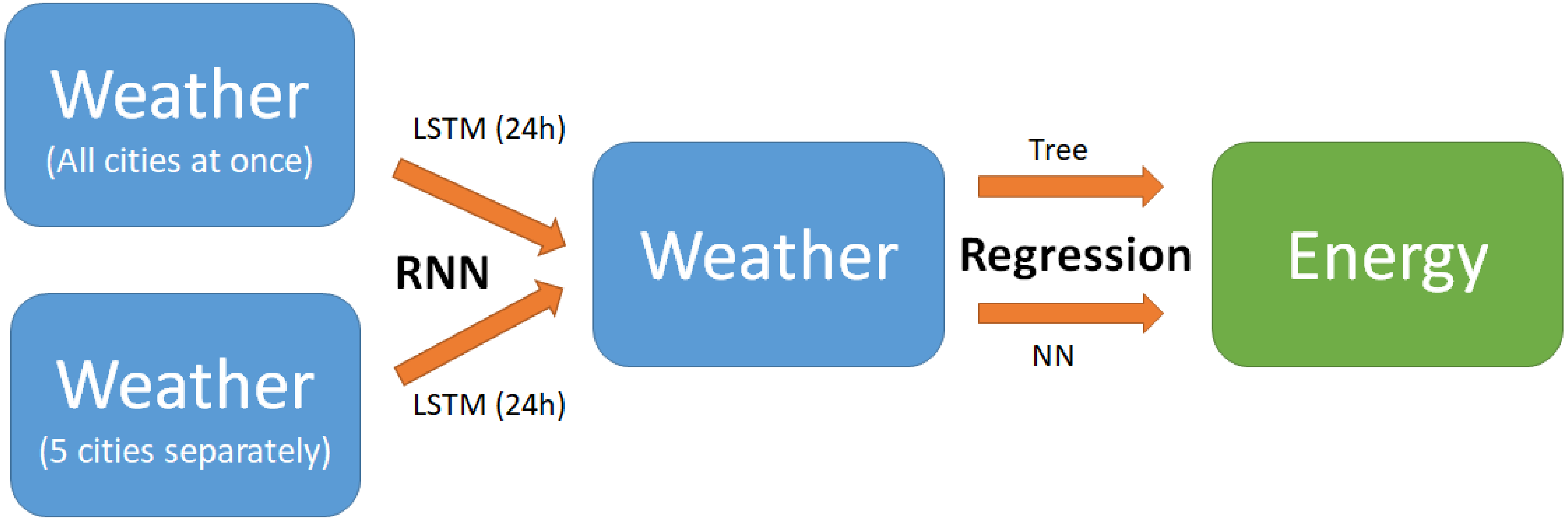
clear	clouds	haze	dust	mist	smoke	fog	drizzle	squall	rain	thunder-storm	snow
0	1	2	3	4	5	6	7	8	9	10	11

Input data processing

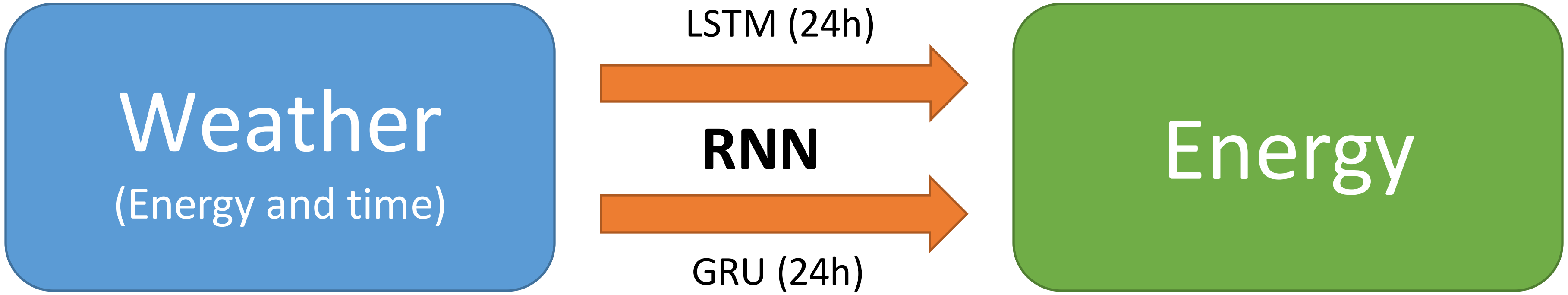
- Split into train and validation set (validation = last 68 weeks(=33%))
- Scale both input features and target
- Create sequence for training
  - (sequence length = 72 hours, forecast horizon = 24 hours)

Approaches

- Approach 1:

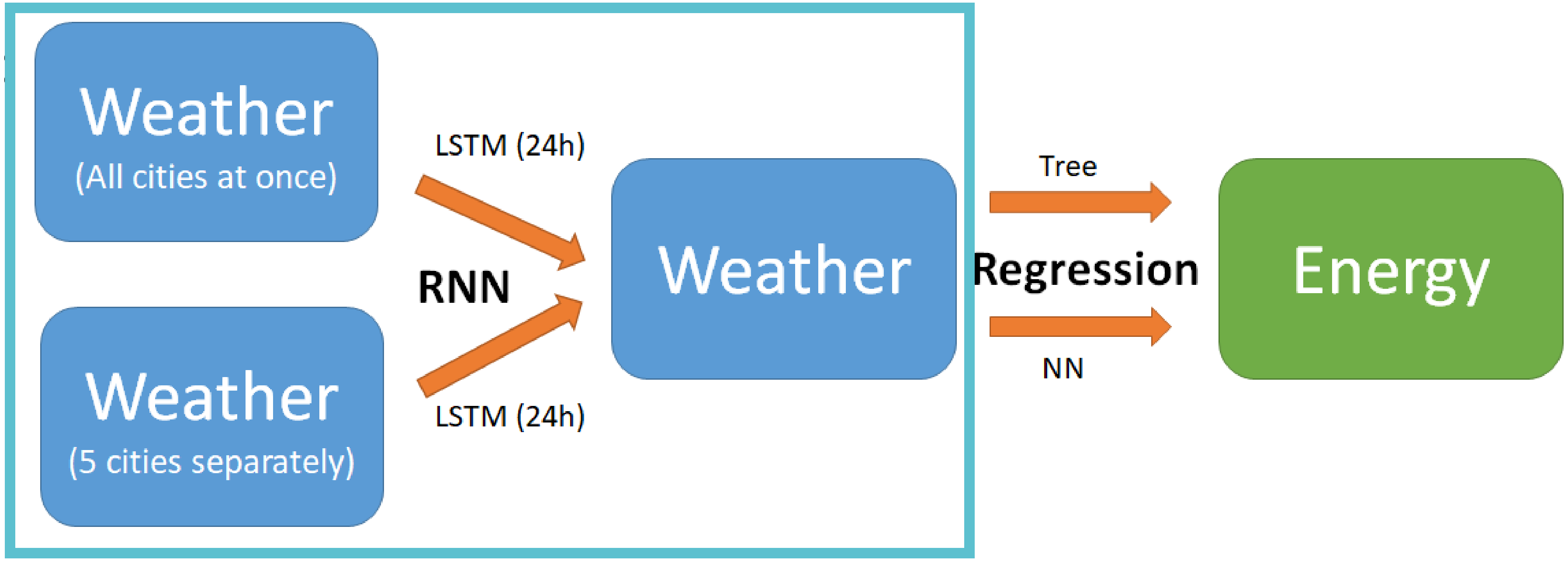


- Approach 2:

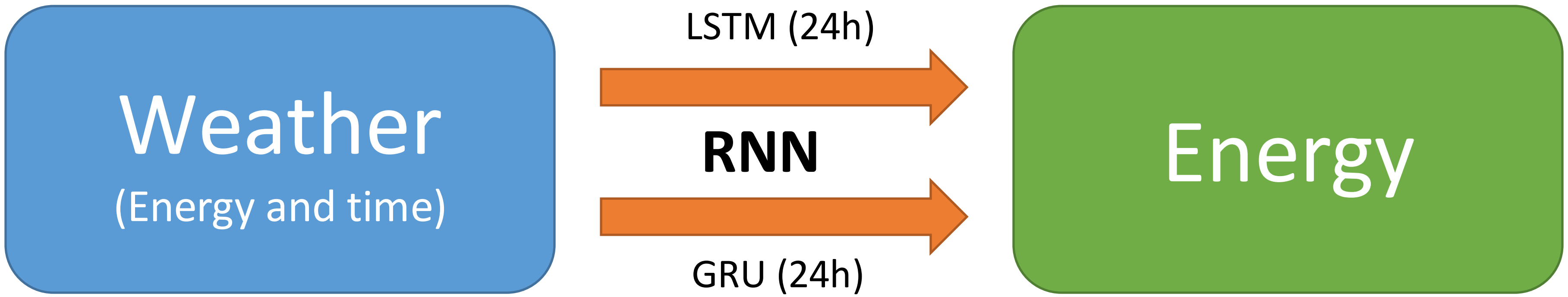


Weather predictions

- Approach 1:

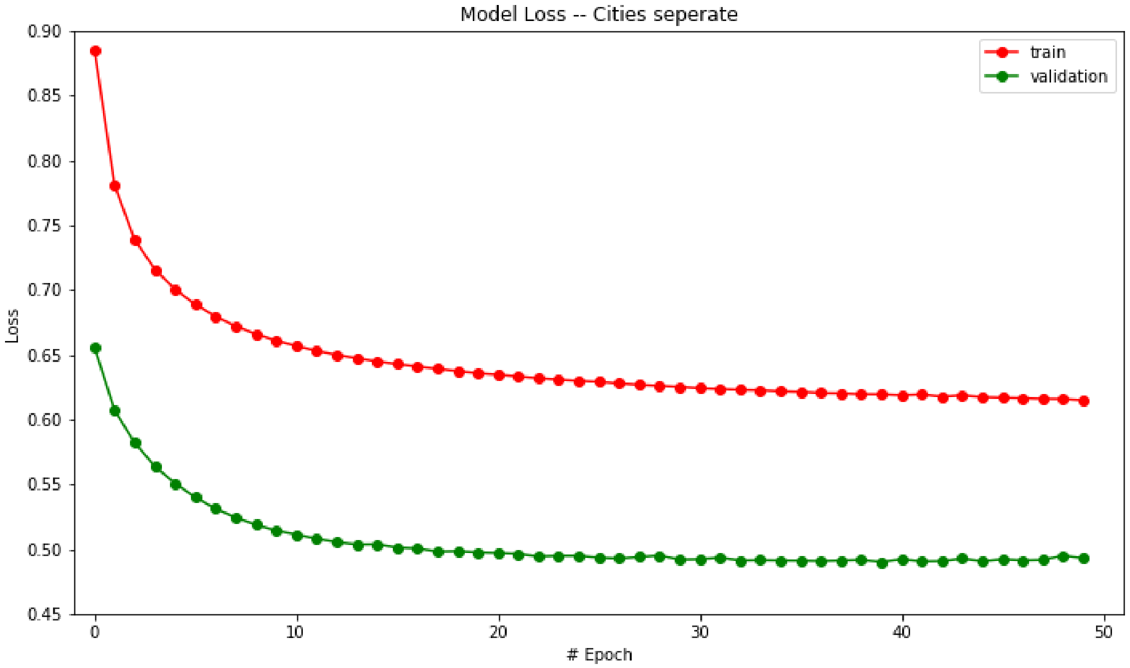


- Approach 2:

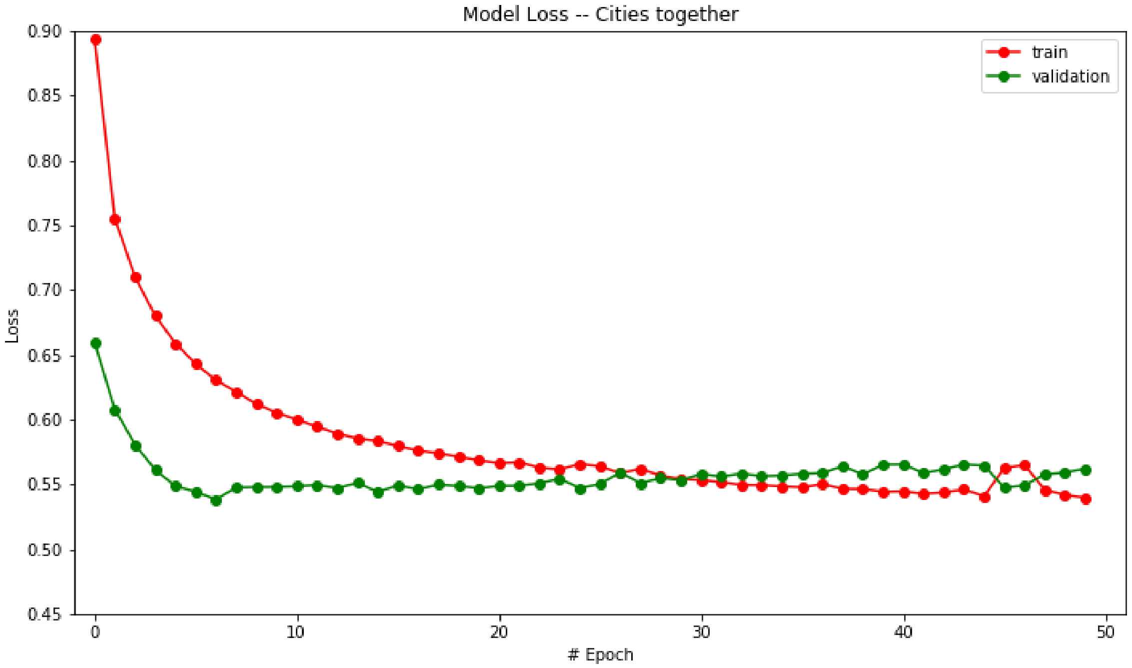


Weather predictions – Training the Model

Cities separate	Cities together
5 parallel layers with 12 features/nodes each	**1 single LSTM layer** with 60 features/nodes
ca. **5.5 min** for 50 epochs	ca. **1 min** for 50 epochs



(used EarlyStopping with patience 10)



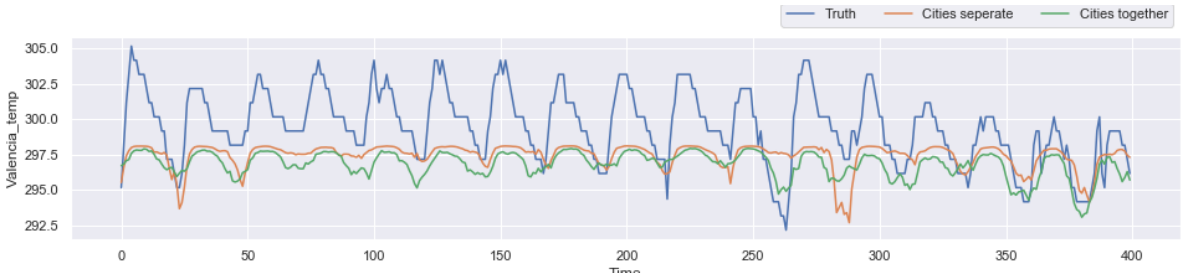
→ decided to only use 6 epochs

Weather predictions – Losses

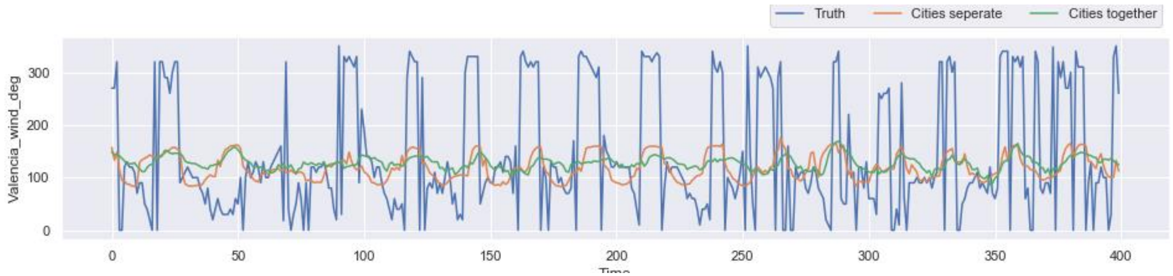
Feature	Mean Error cities sperate		Mean Error cities together	
	Root MSE	MAPE	Root MSE	MAPE
Valencia_temp	2.37	0.81%	2.66	0.92%
Valencia_temp_min	2.60	0.89%	2.92	1.00%
Valencia_temp_max	2.32	0.80%	2.48	0.85%
Valencia_pressure	3.86	0.38%	4.59	0.45%
Valencia_humidity	11.63	18.16%	11.42	17.56%
Valencia_wind_speed	1.35	59.72%	1.24	49.87%
Valencia_wind_deg	94.36	64.01%	92.90	63.35%
Valencia_rain_1h	0.05	249.80%	0.05	582.90%
Valencia_rain_3h	0.00	-	0.00	-
Valencia_snow_3h	0.00	-	0.00	-
Valencia_clouds_all	18.02	91.55%	17.88	94.97%
Valencia_weather_main	0.95	102.99%	1.04	90.18%

Weather predictions – Time Series

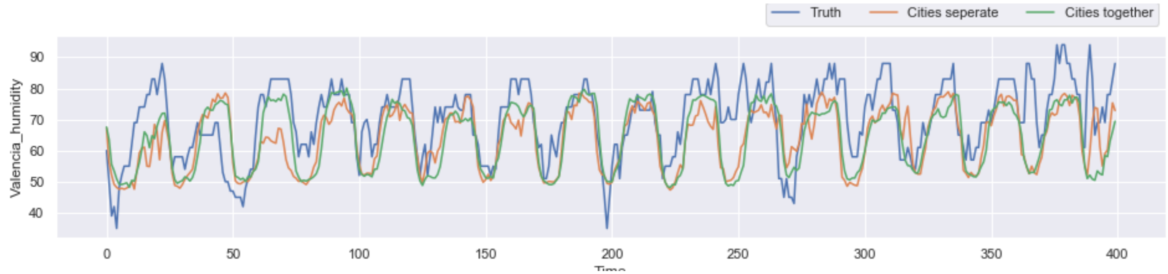
Temperature



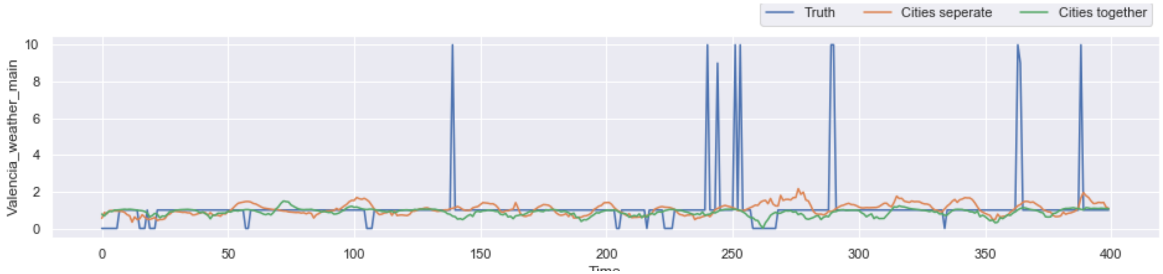
Wind direction



Humidity

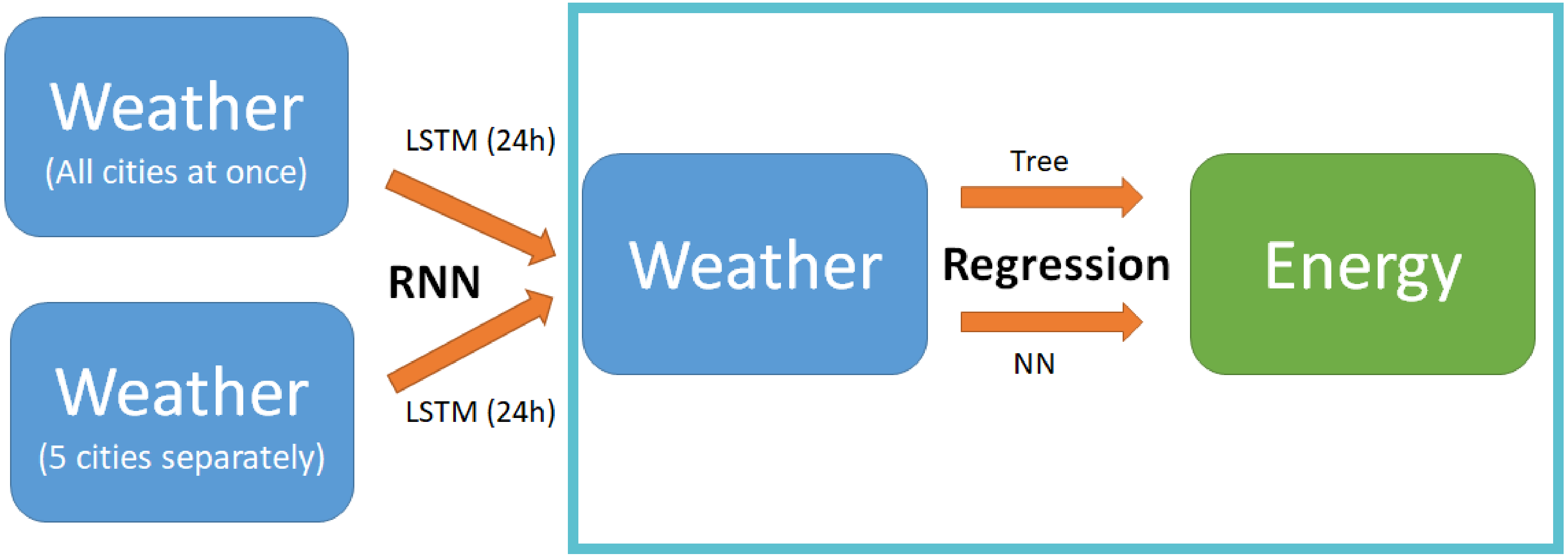


weather_main

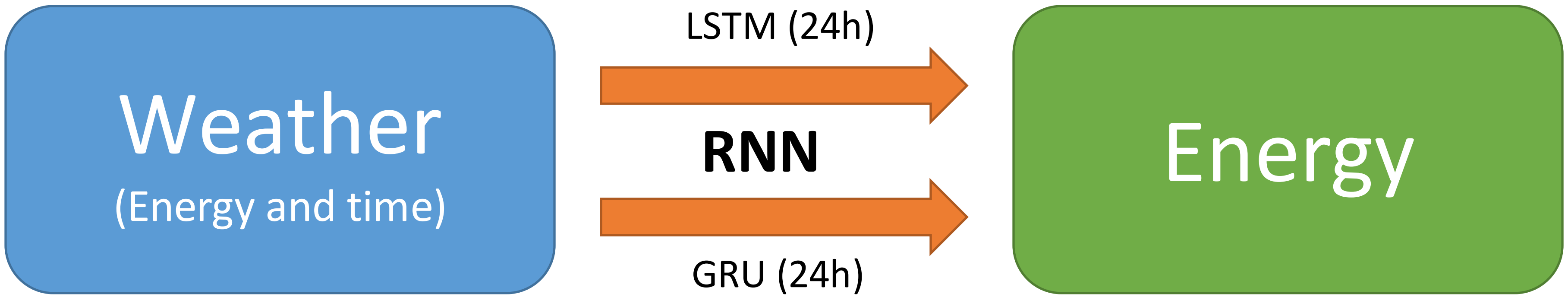


Approach

- Approach 1:

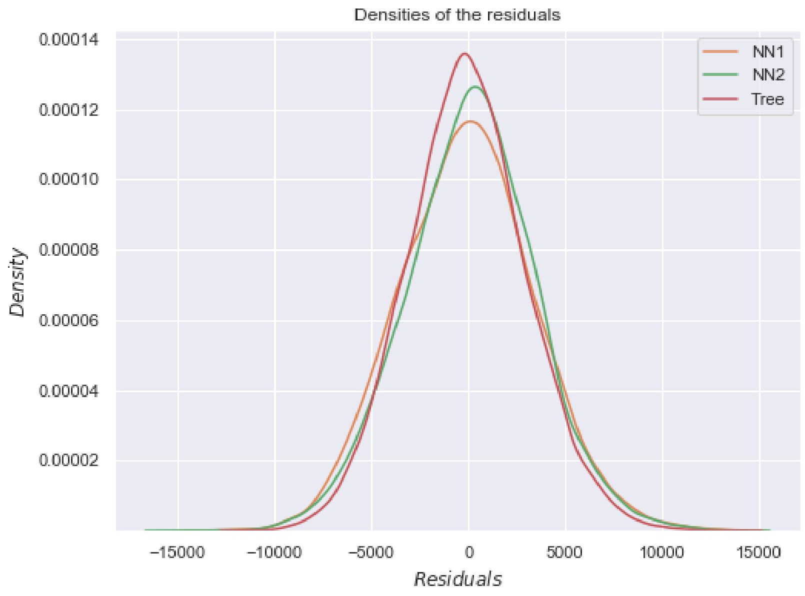


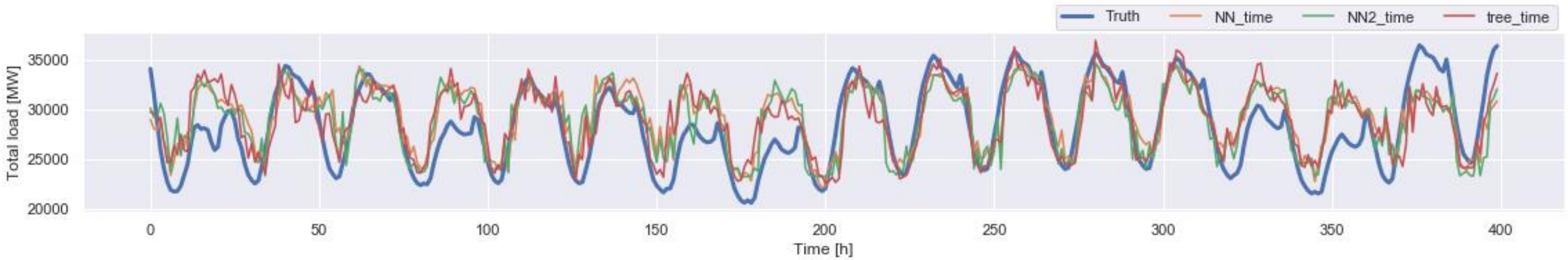
- Approach 2:



Energy regression

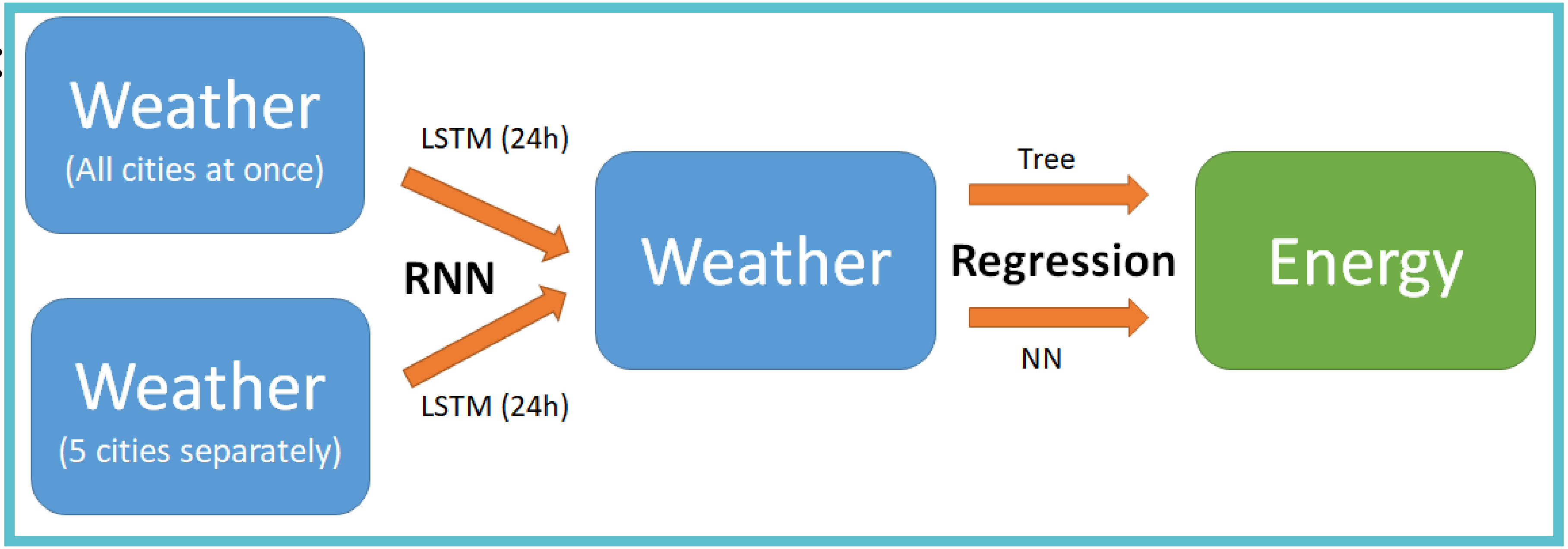
Algorithm	Root MSE	MAPE
NN1 (3Layer)	3482.97	9.69%
NN2 (6Layer)	3341.23	9.69%
Tree	3058.15	8.47%



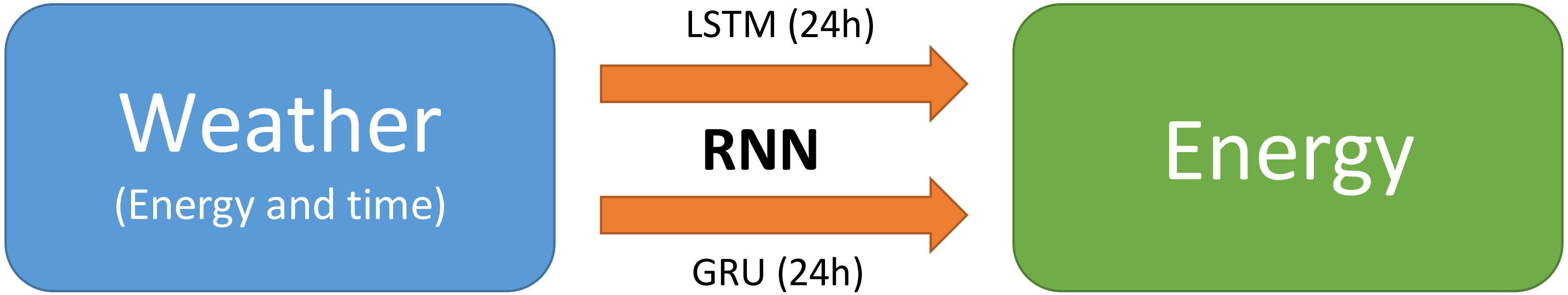


Approach

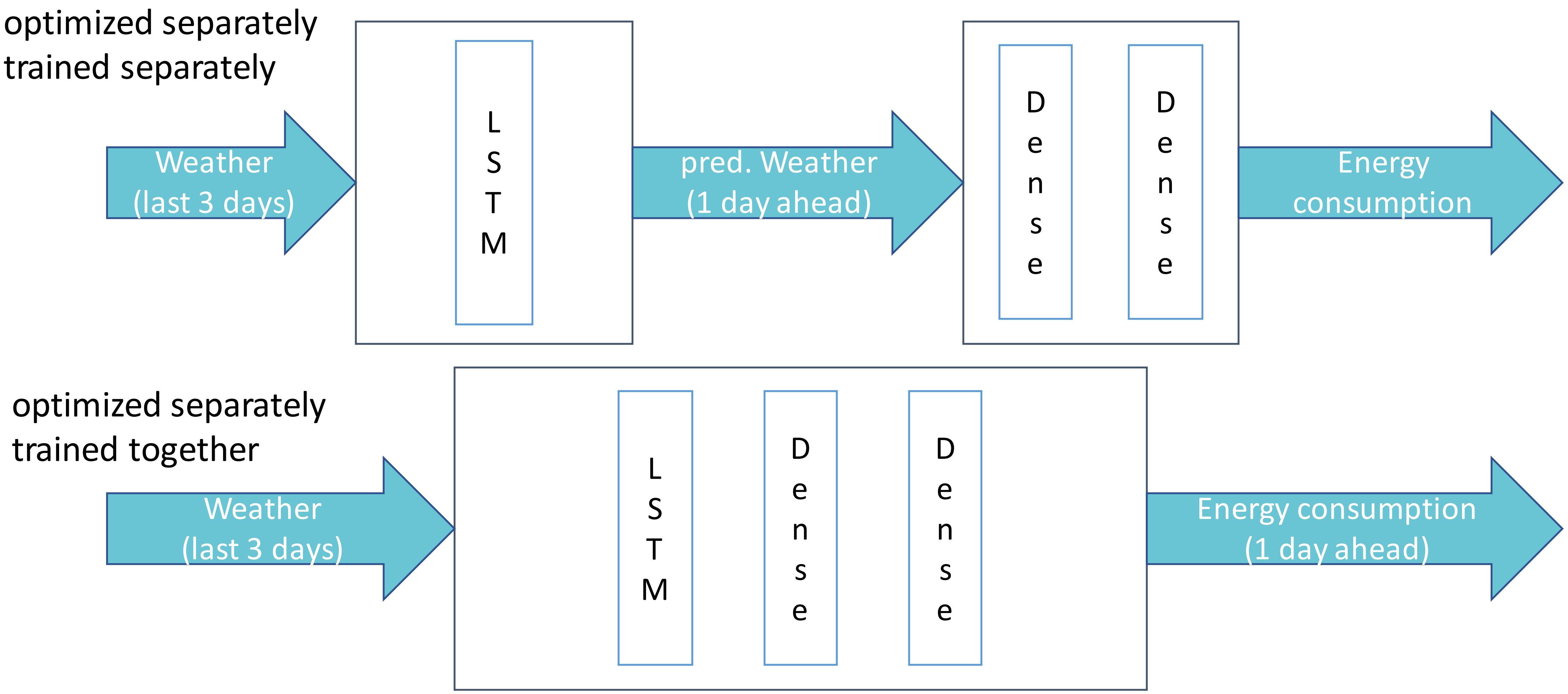
- Approach 1:



- Approach 2:

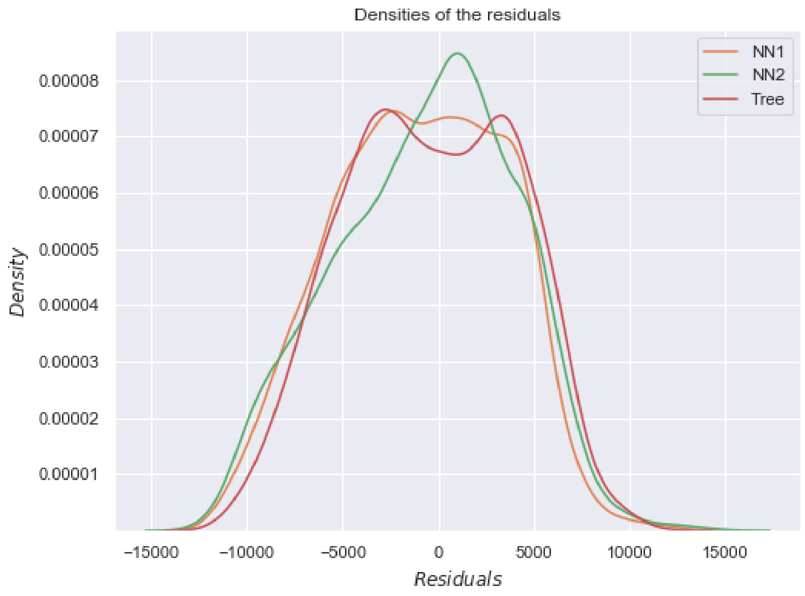


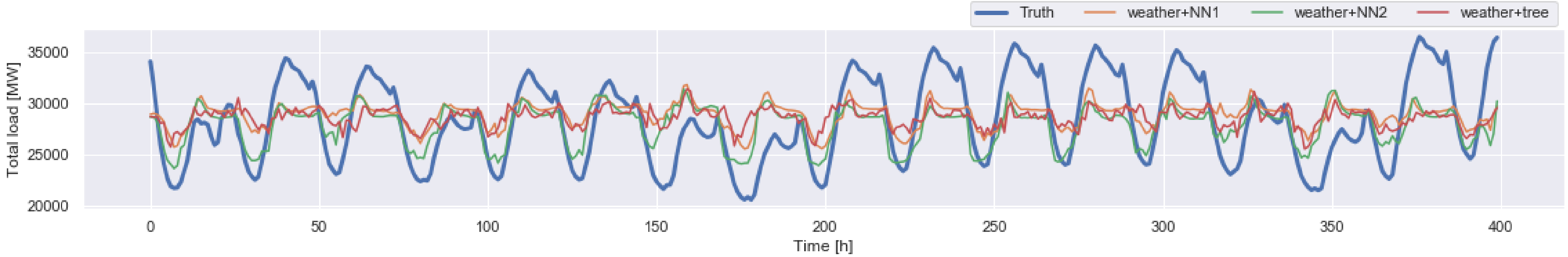
Approach 1


optimized separately
trained separately
Weather
(last 3 days)
LSTM
pred. Weather
(1 day ahead)
Dense
Dense
Energy
consumption
optimized separately
trained together
Weather
(last 3 days)
LSTM
Dense
Dense
Energy consumption
(1 day ahead)

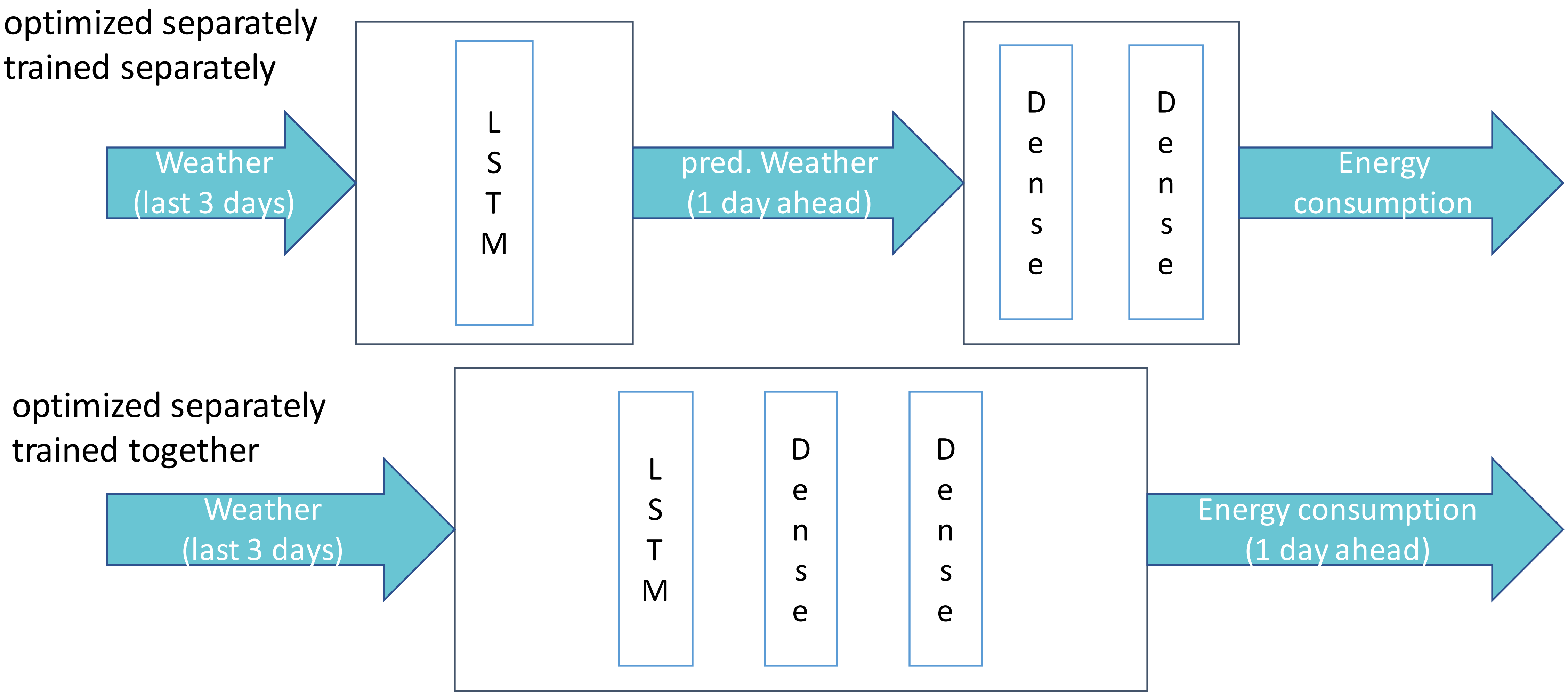
Energy regression with pred. weather

Algorithm	Root MSE	MAPE
NN1 (3Layer)	4652.38	12.63%
NN2 (6Layer)	4688.32	12.71%
Tree	4456.26	12.75%



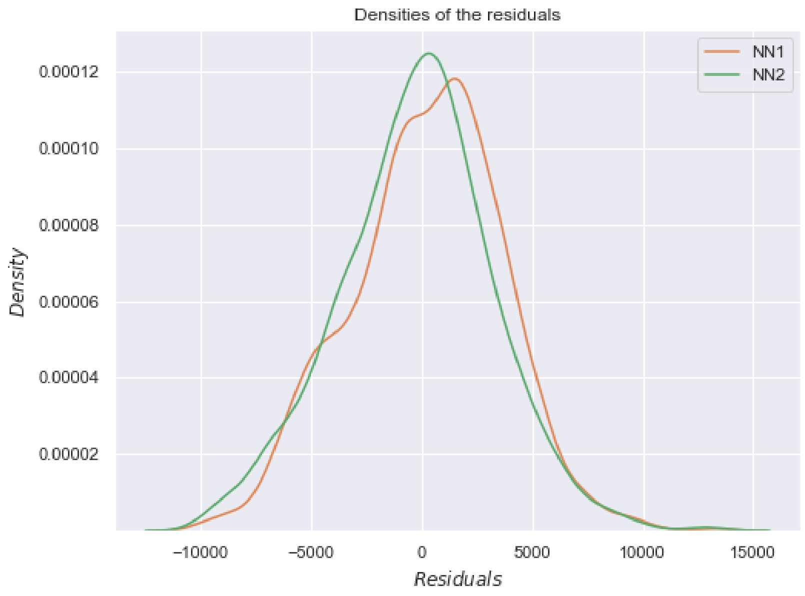


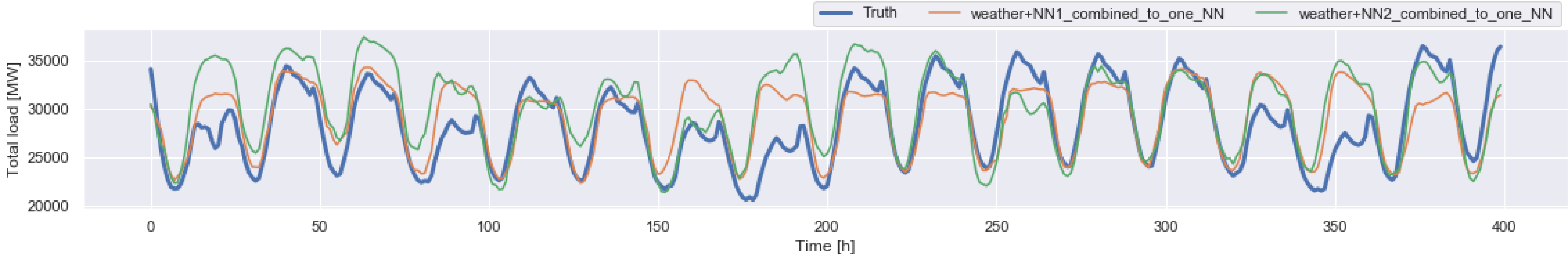
Approach 1


optimized separately
trained separately
Weather
(last 3 days)
LSTM
pred. Weather
(1 day ahead)
Dense
Dense
Energy
consumption
optimized separately
trained together
Weather
(last 3 days)
LSTM
Dense
Dense
Energy consumption
(1 day ahead)

Energy regression with combined network

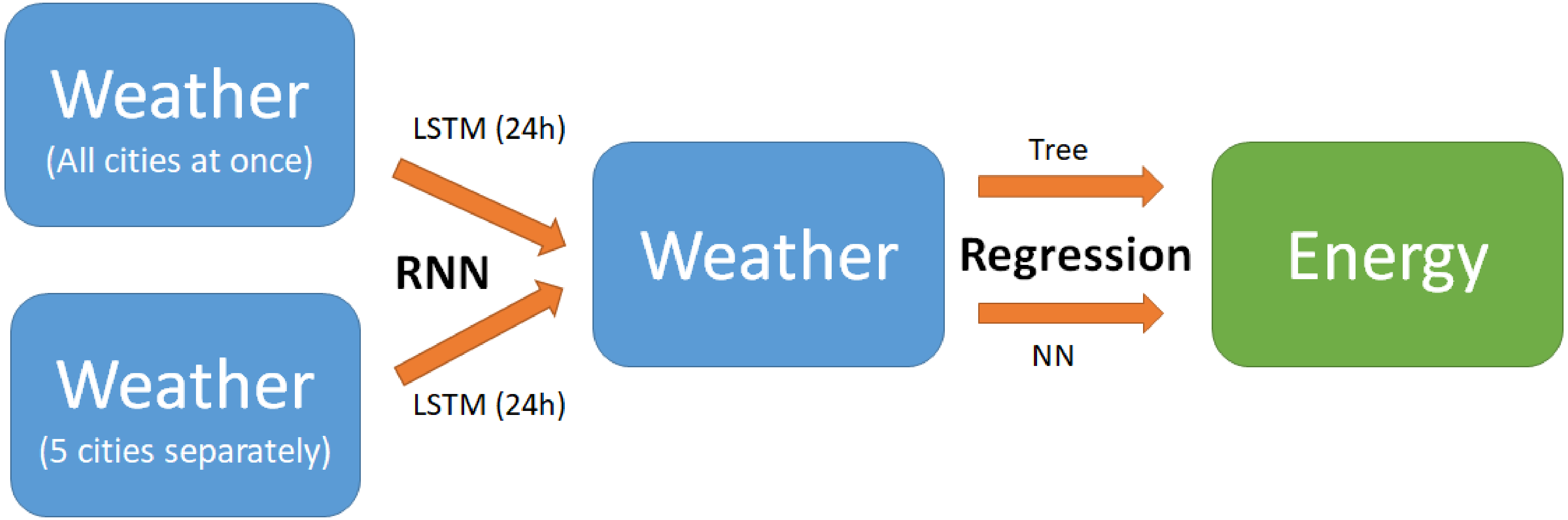
Algorithm	Root MSE	MAPE
LSTM + NN1	3512.39	9.91%
LSTM + NN2	3844.89	10.23%



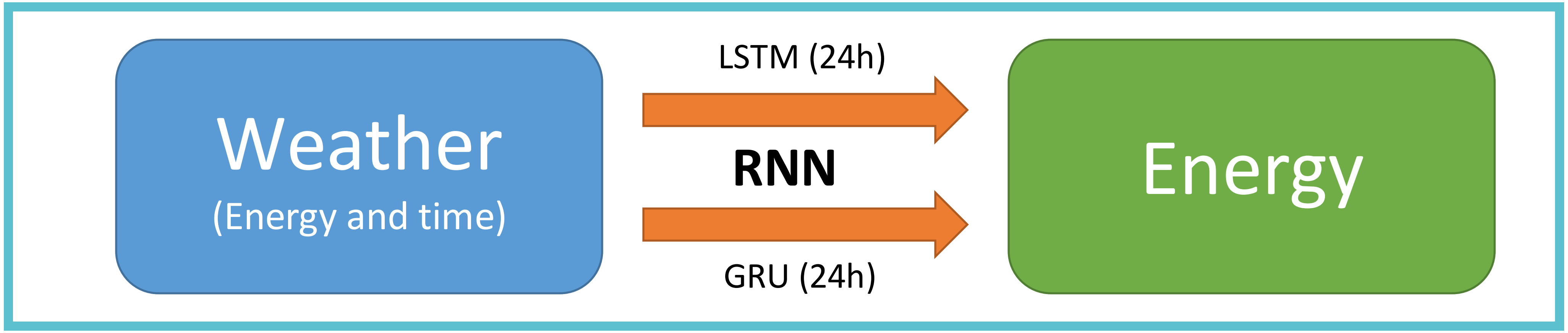


Approach 2

- Approach 1:

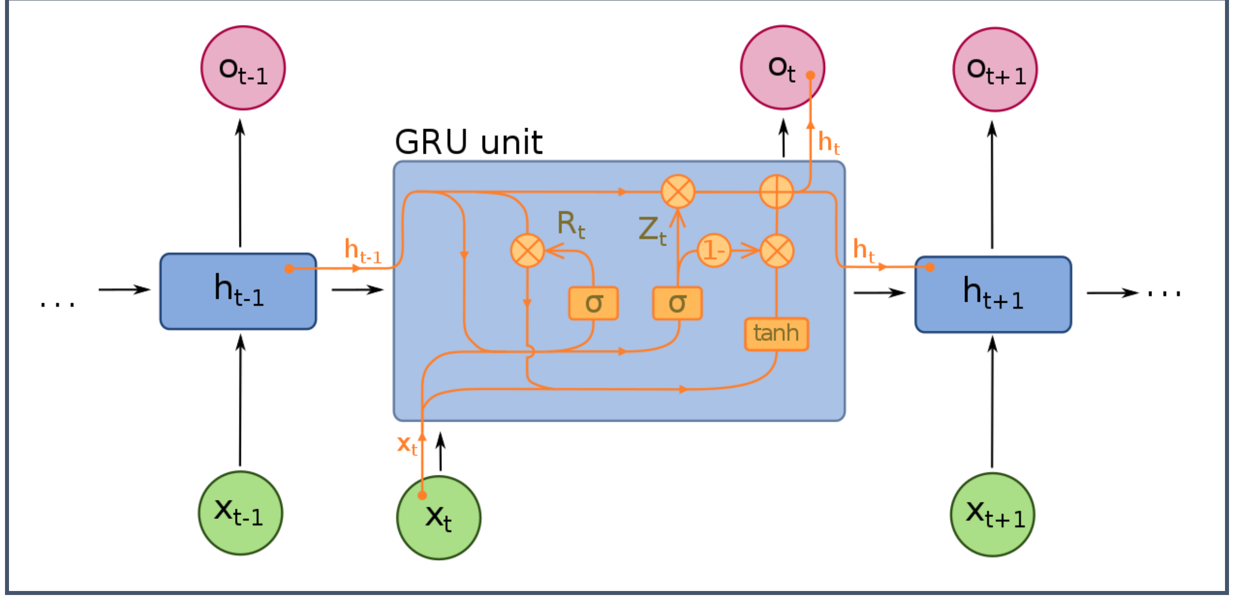


- Approach 2:

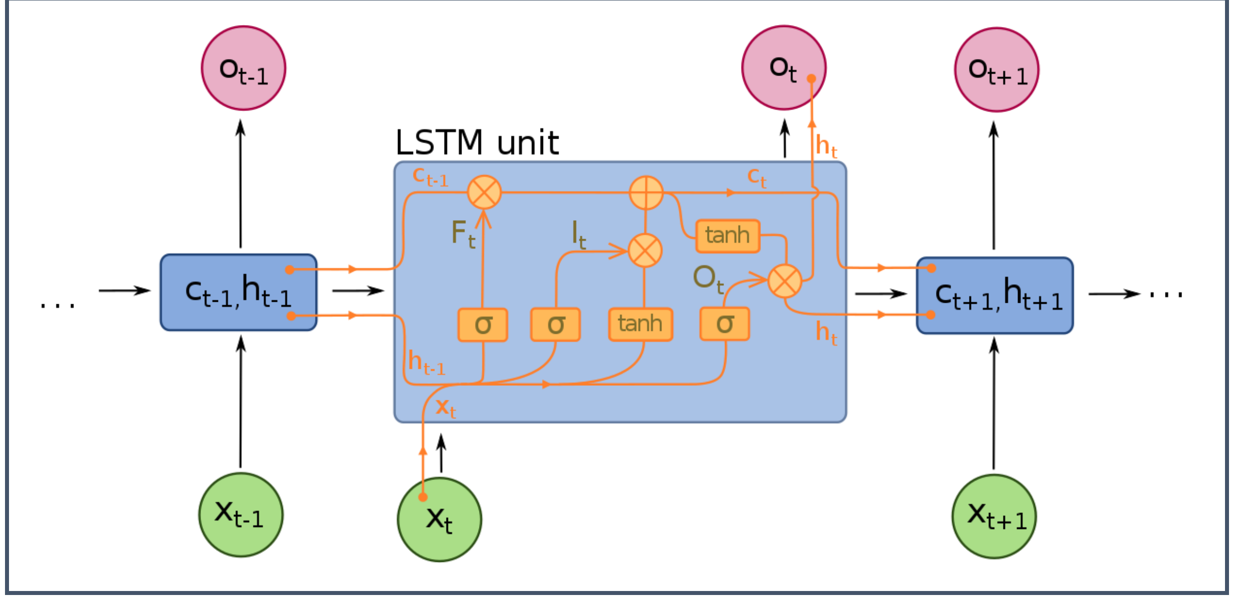


RNNs

Gated recurrent unit



Long short-term memory

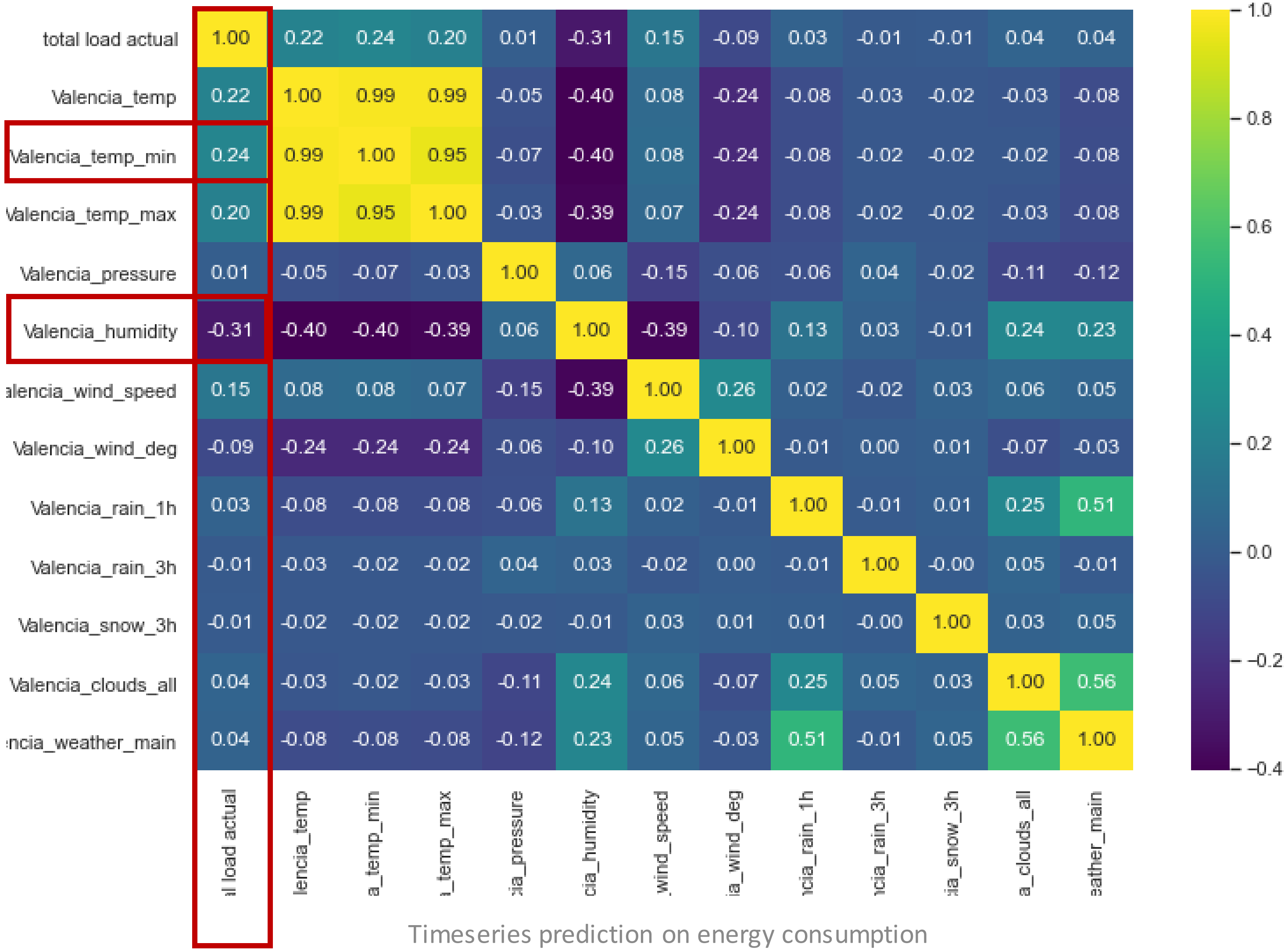


picture source: https://en.wikipedia.org/wiki/Recurrent_neural_network

Model architecture:

- Models (hyperparameters optimized with kerastuner-bayesian):
  - LSTM: 1 layer, 2 layers + 1 Dense layer (output)
  - GRU: 1 layer, 2 layers + 1 Dense layer (output)
- Input feature combinations:
  - All 62 features (weather from 5 cities, time, energy)
  - 3 features: Valencia minimum temp, Valencia humidity + time of the day
  - 4 features: 2 weather features + time + energy consumption
- Weather features chosen based on correlation-values

Correlations of features



Model architecture:

- Models (hyperparameters optimized with kerastuner-bayesian):
  - LSTM: 1 layer, 2 layers + 1 Dense layer (output)
  - GRU: 1 layer, 2 layers + 1 Dense layer (output)

- Input feature combinations:
  - All 62 features (weather from 5 cities, time, energy)
  - 3 features: Valencia minimum temp, Valencia humidity + time of the day
  - 4 features: 2 weather features + time + energy consumption

- Weather features chosen based on correlation-values

Model evaluation - Statistical analysis

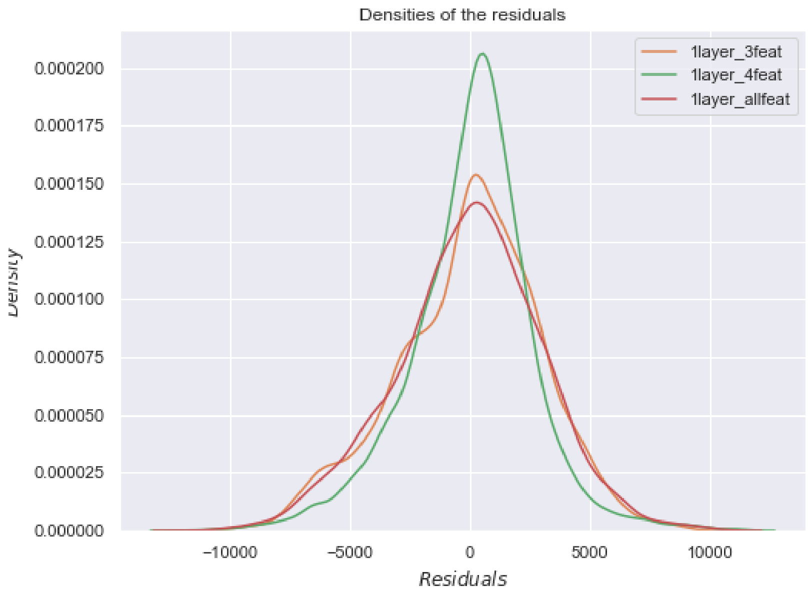
- Training + optimization with MSE
- 2 different evaluation metrics: RMSE, MAPE
- Training each optimized model 10 times, save metrics
- Compute mean loss to find the best performing one
- t-test to see if difference is statistically significant

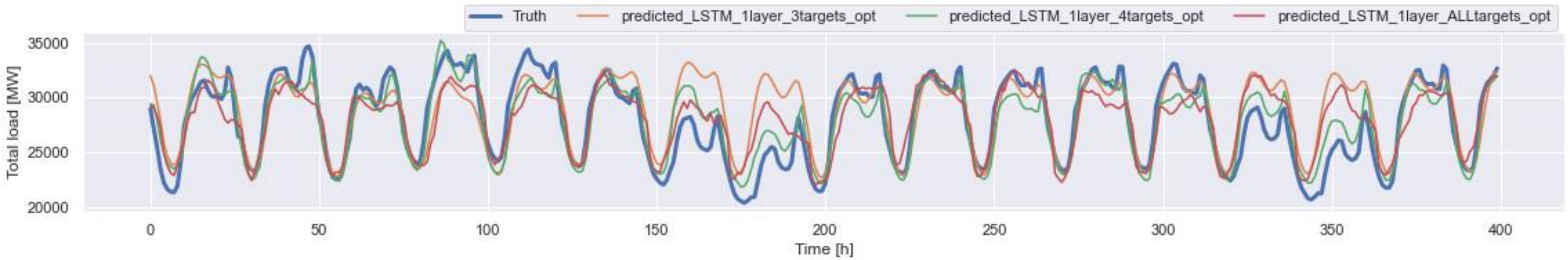
Input features - results

(1 layer LSTM)

Algorithm	RMSE	MAPE
LSTM,1layers,3feat	3476.29	9.87%
LSTM,1layers,4feat	2970.27	7.21%
LSTM,1layers,allfeat	3715.20	9.05%

t-test (with RMSE of 10 model trainings):
4 features always significantly different ($p < 0.05$) from 3 and all features.
3 and all in one model not significantly different from each other.

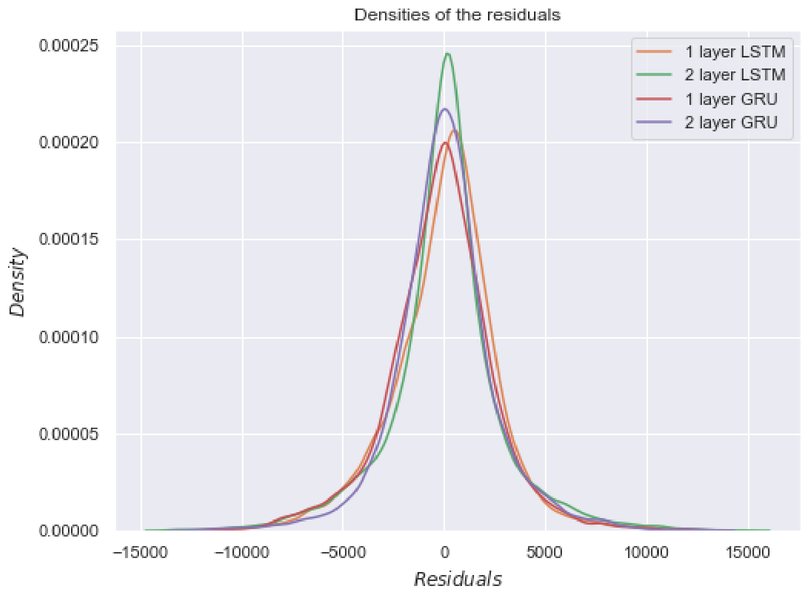




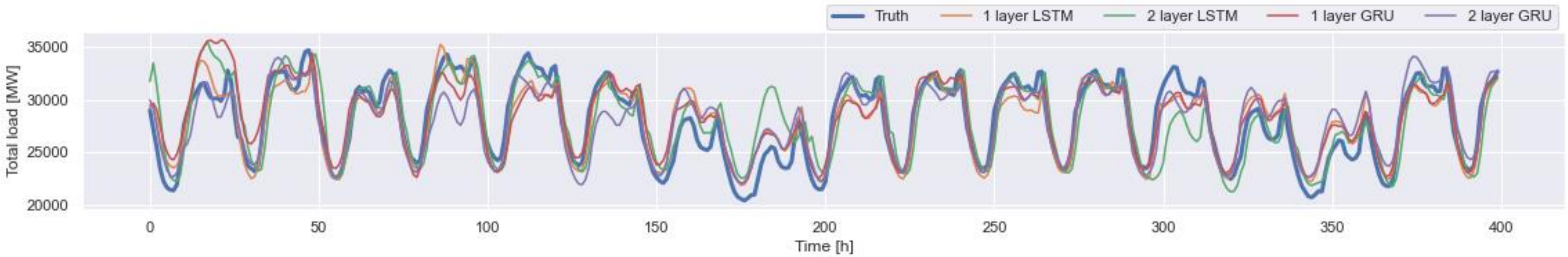
Model architecture - results

(all using 4 input features)

Algorithm	Root MSE	MAPE
LSTM, 1layer	2970.27	7.21%
LSTM, 2layer	2818.29	6.72%
GRU, 1layer	2855.61	6.89%
GRU, 2layer	2602.95	6.15%

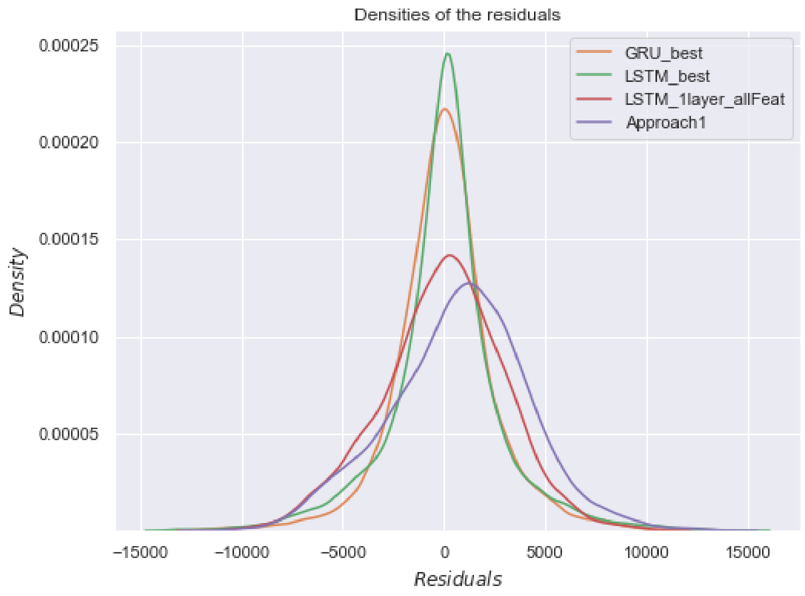


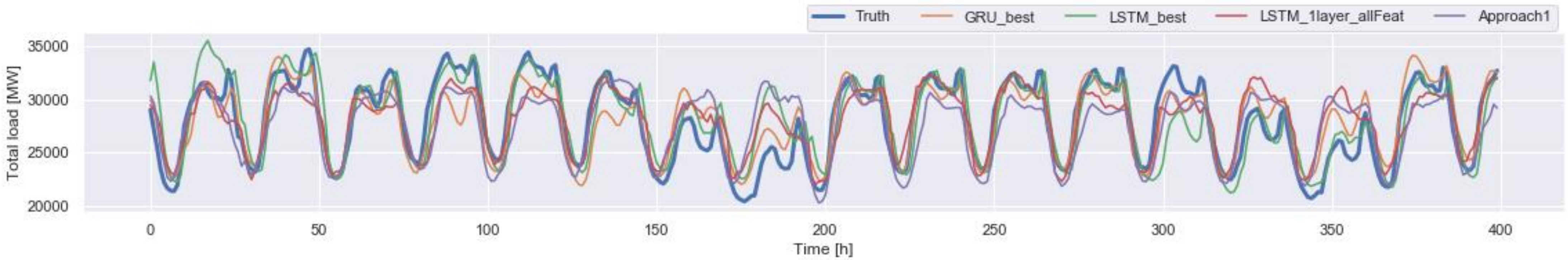
t-test (with RMSE of 10 model trainings):
LSTM 2layer and GRU 1 layer not statistically significant different ($p < 0.05$)
For all others the difference is statistically significant



Results / Evaluation

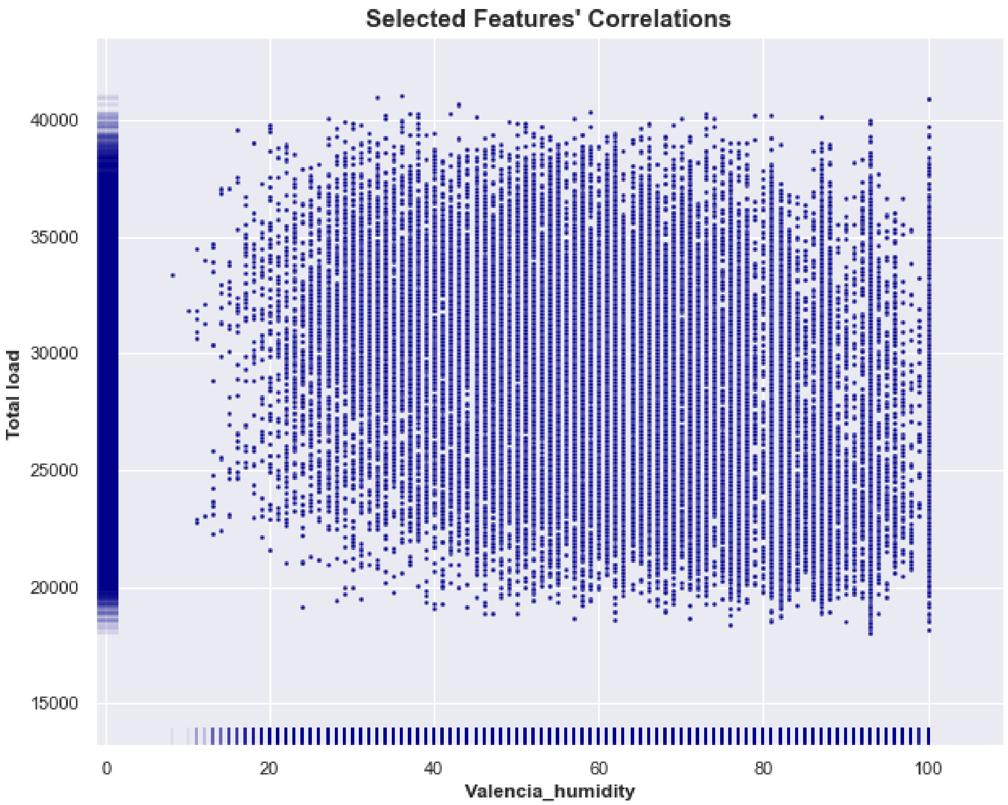
Algorithm	Root MSE	MAPE
GRU (2 layers)	2602.95	6.15%
LSTM (2 layers)	2818.29	6.72%
LSTM (1layers,allfeat)	3715.20	9.05%
Appr. 1: LSTM + 3 Dense (all weather feat)	3512.39	9.91%

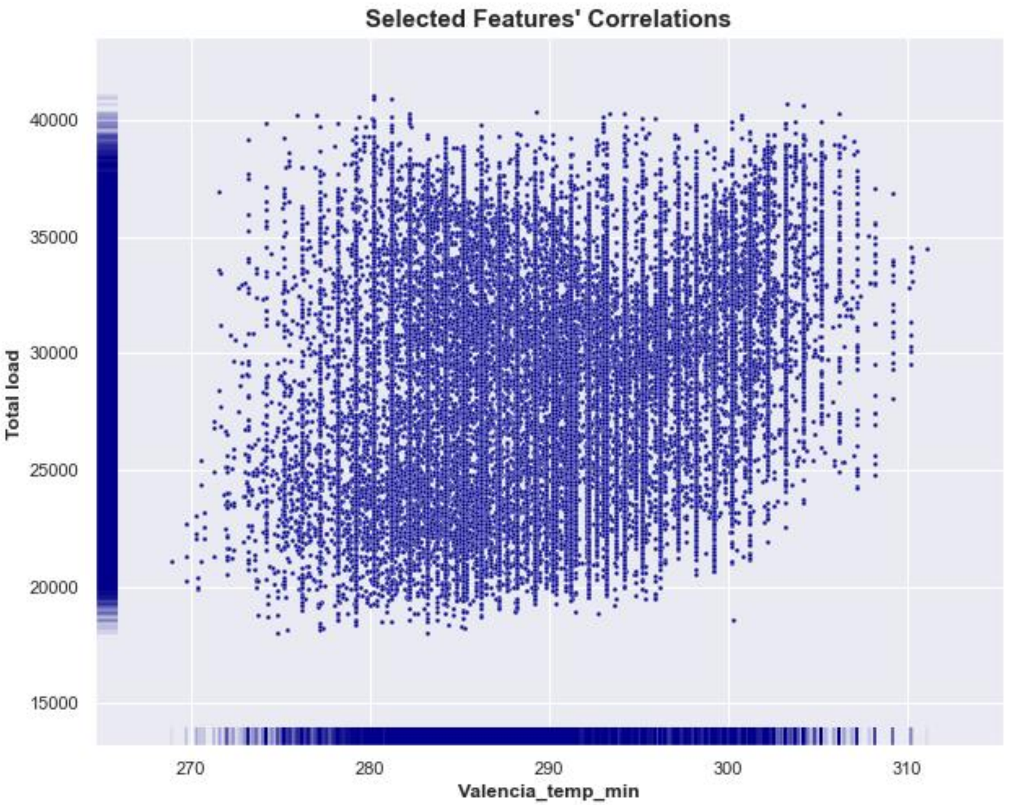




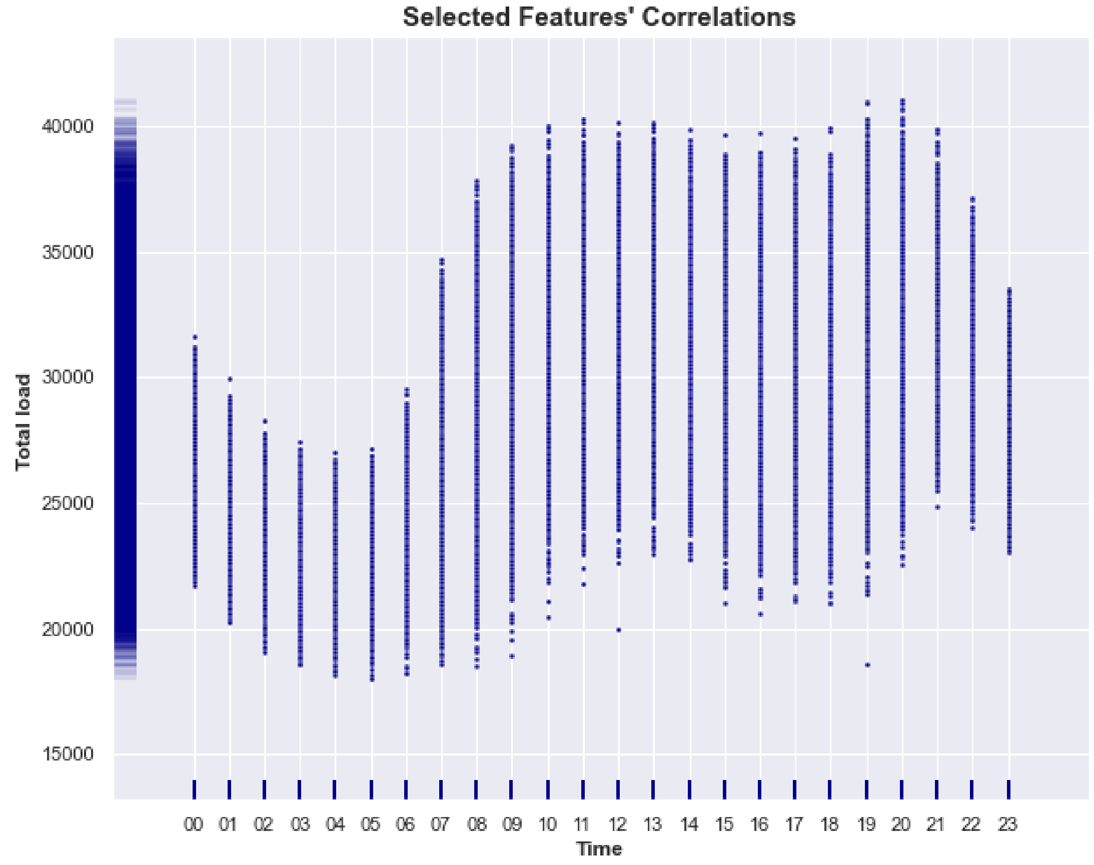
Appendix

Correlations of features

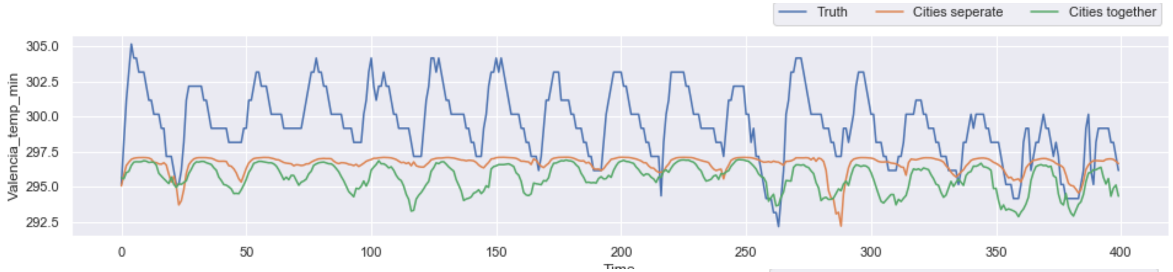


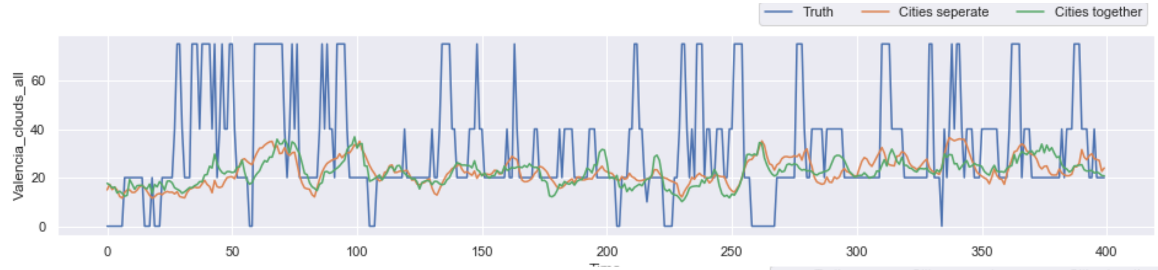


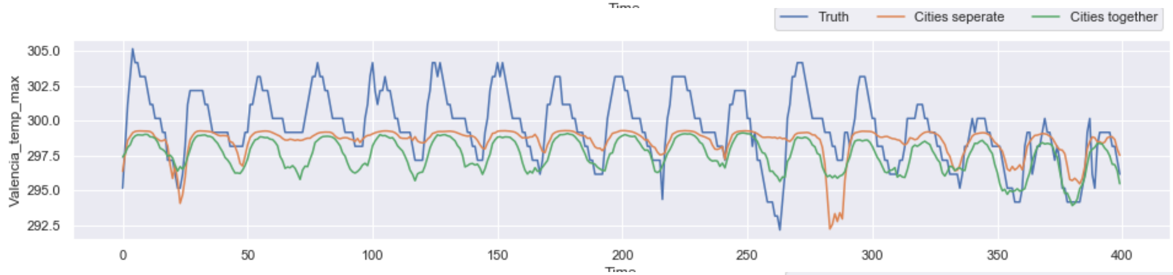
Correlations of features

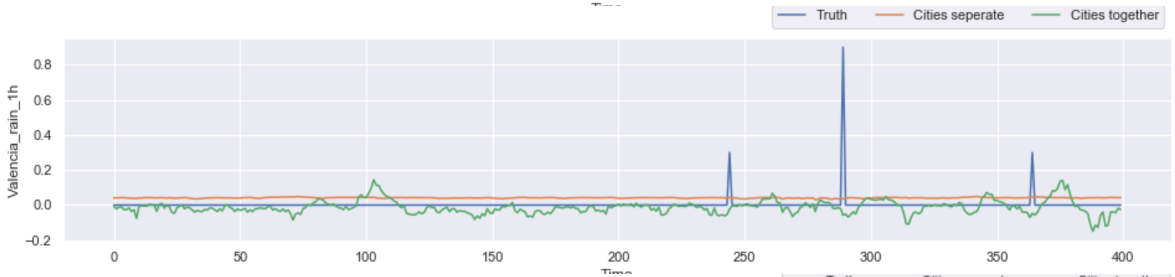


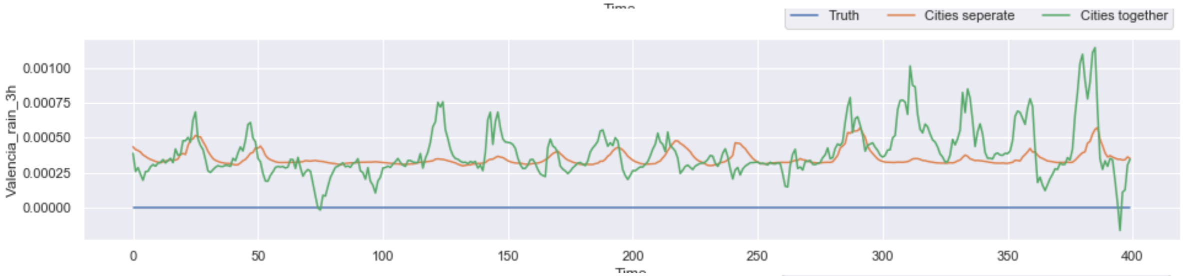
Weather predictions – Further Time Series

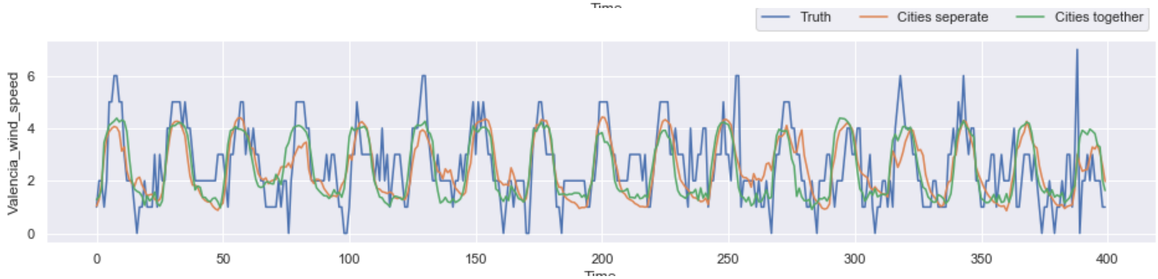


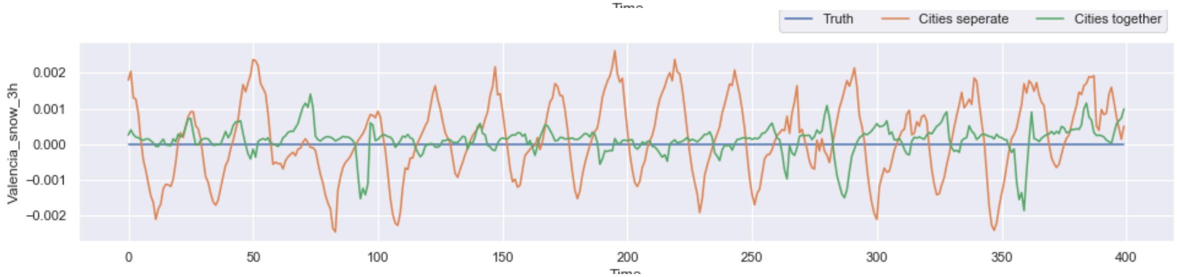












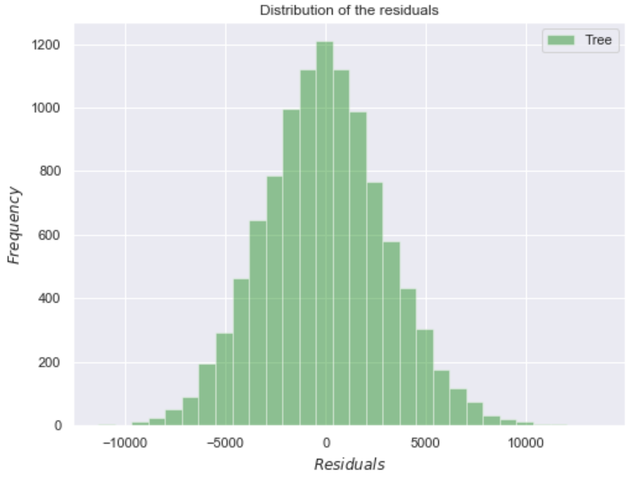
Weather predictions

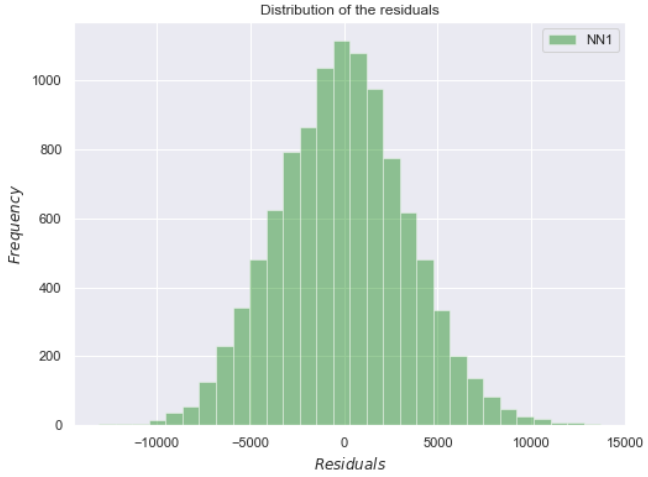
Algorithm	Cities seperate	Cities together
Input features	All columns	All columns
Optimization	optuna with TPESampler and Hyperband	optuna with TPESampler and Hyperband
Hyperparameters	N-epochs: 50 (using EarlyStopping with patience 10) Batch size: 148 Learning rate: 0.00082 Num layers: 5 in parallel, so technically 1 Num neurons (/units): 5*12 Dropout rate: 0 Activation function: tanh	N-epochs: 6 Batch size: 576 Learning rate: 0.0013 Num layers: 1 Num neurons (/units): 60 Dropout rate: 0 Activation function: tanh
RMSE (avg over all predicted features)	123.310240006	273.606269337
MAPE (avg over all predicted features)	96.08117119074457	100.68384238279901
Comments:		

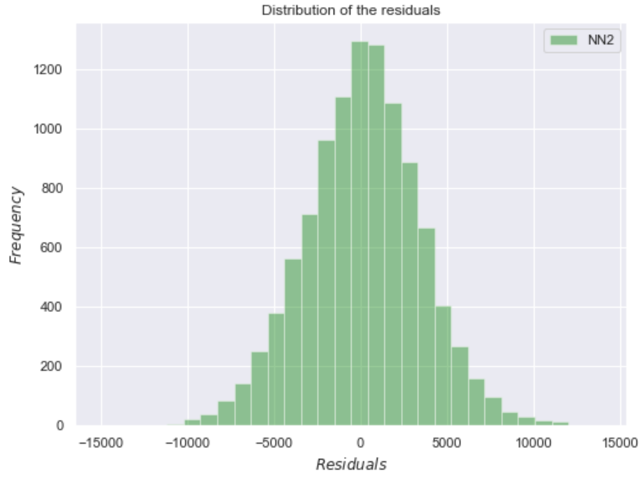
Energy regression

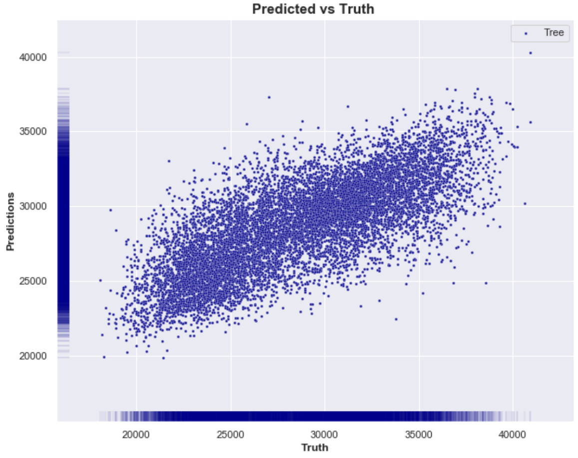
Algorithm	LightGBM	NN1	NN2
Input features	All columns	All columns	All columns
Optimization	Bayesian	KerasTuner, Baysian	KerasTuner, Hyperband
Hyperparameters	Max-depth: 58 N-leaves: 183 learning rate: 0.1 reg_alpha: 0.01 Min split gain: 0.1 Min child weight: 10	N-epochs: 86 Batch size: 100 Learning rate: 0.00069 Num layers: 3 Num neurons (/units): 416,416,1 Dropout rate: 0 Activation function: relu,relu,linear	N-epochs: 78 (Early Stopping, 10) Batch size: 100 Learning rate: 0.00065 Num layers: 6 Num neurons (/units): 64,224,32,32,32,1 Dropout rate: 0 Activation function: relu,relu,relu,relu,relu,linear
RMSE	3058.1516312419	3481.9675932047	3341.233452609
MAPE	0.08466964811682642	0.09694302899721959	0.09694302899721959
Comments:			

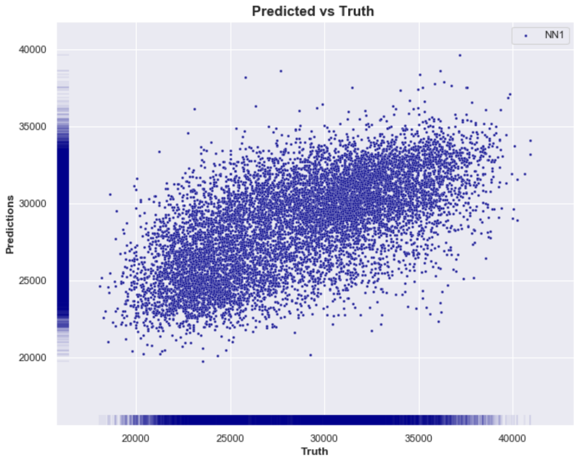
Energy regression

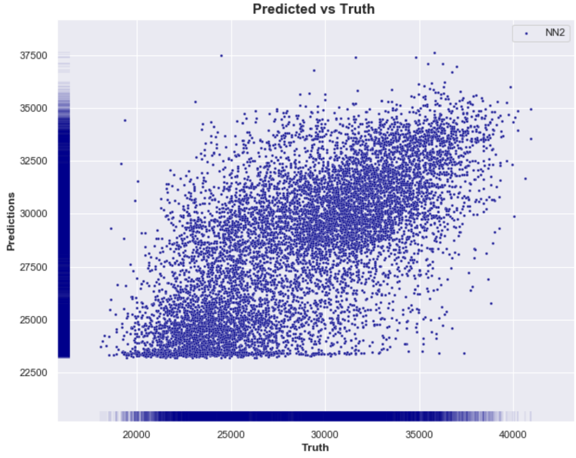








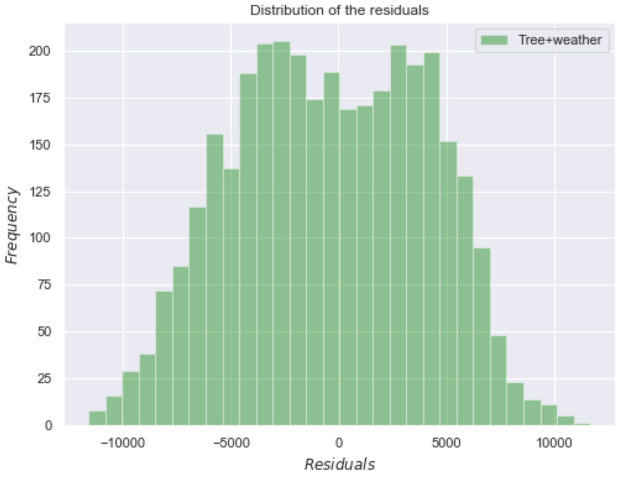


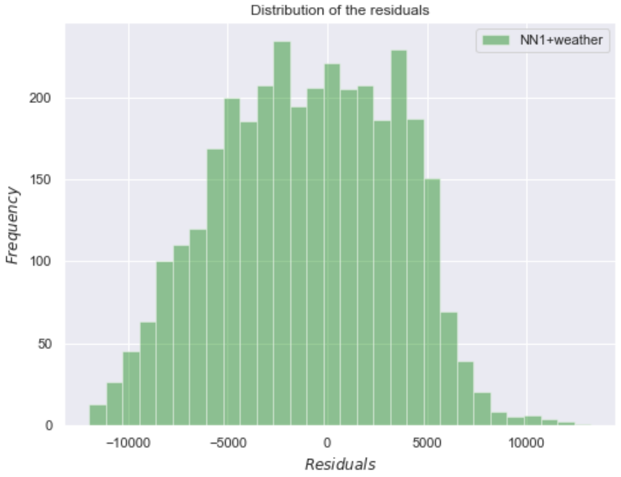


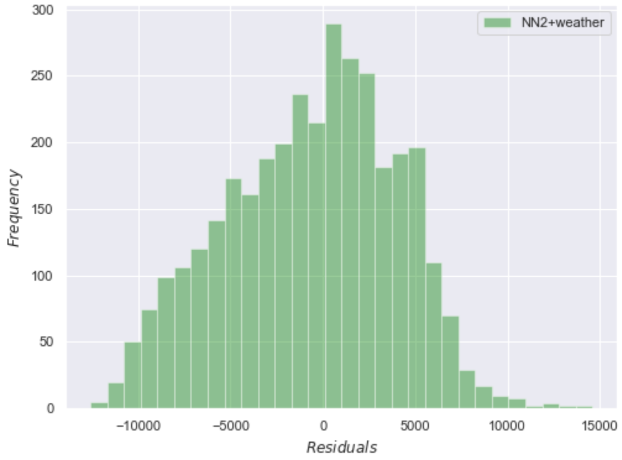
Energy prediction from predicted weather

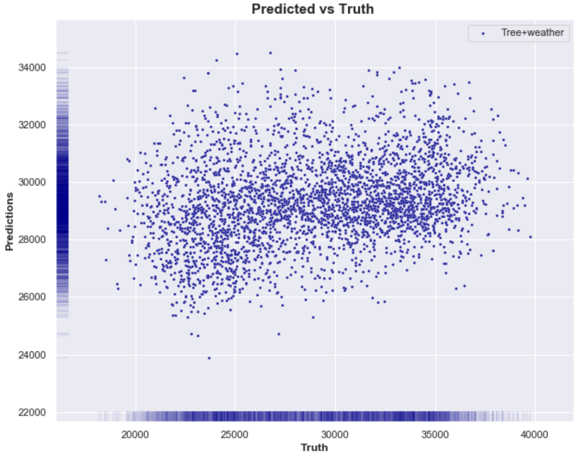
Algorithm	LightGBM	NN1	NN2
Input features	All columns	All columns	All columns
Optimization	see previous models	see previous models	see previous models
Hyperparameters	see previous models	see previous models	see previous models
RMSE	4456.259144140454	4652.381930039405	4688.31977344518
MAPE	0.12749983719854924	0.12625481749535658	0.12708823742347305
Comments:			

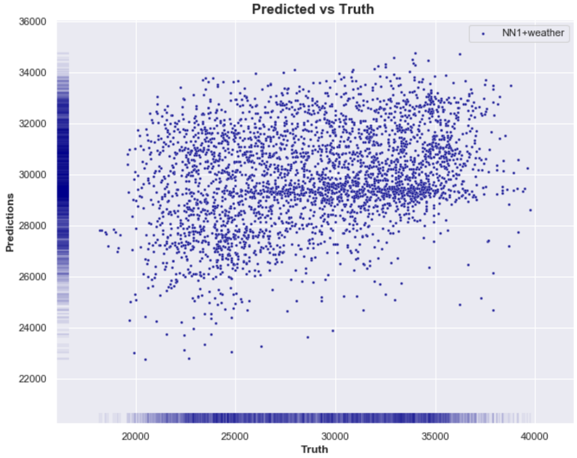
Energy prediction from predicted weather

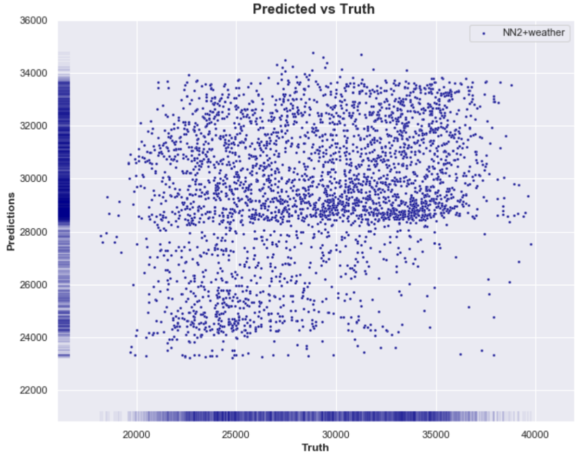






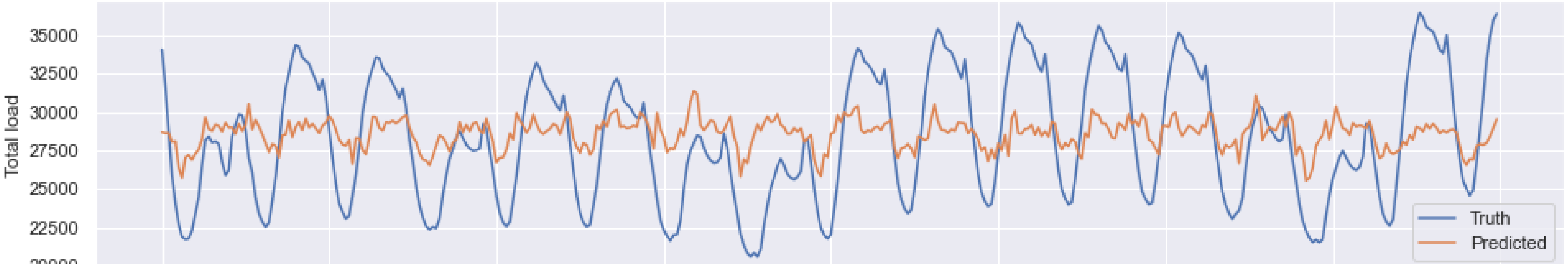




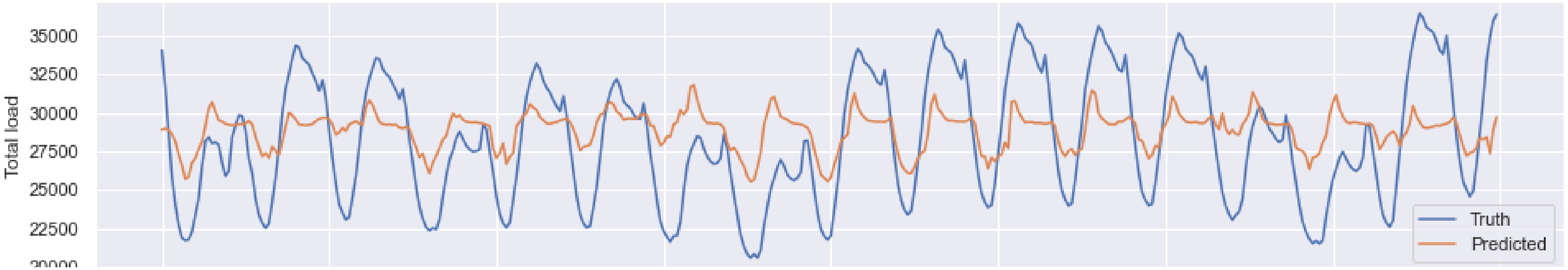


Energy prediction from predicted weather

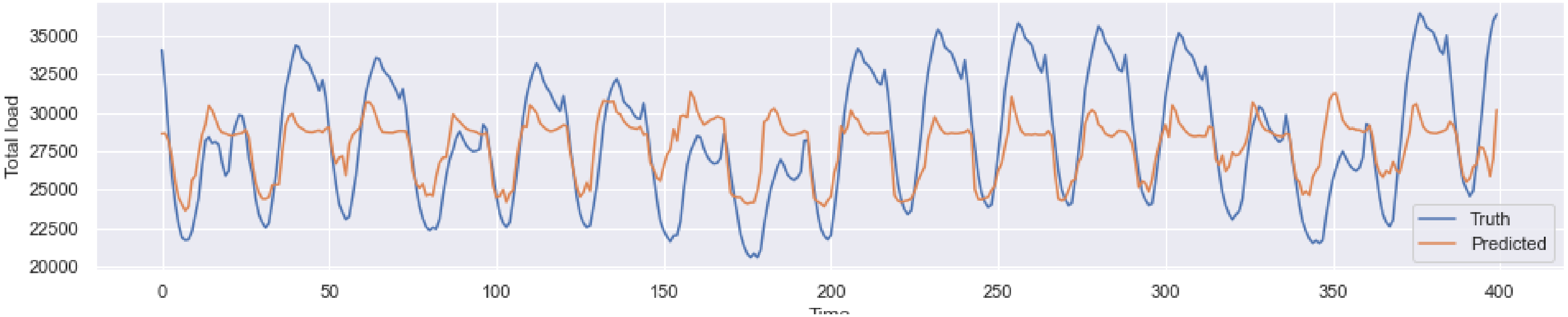
Tree



NN1



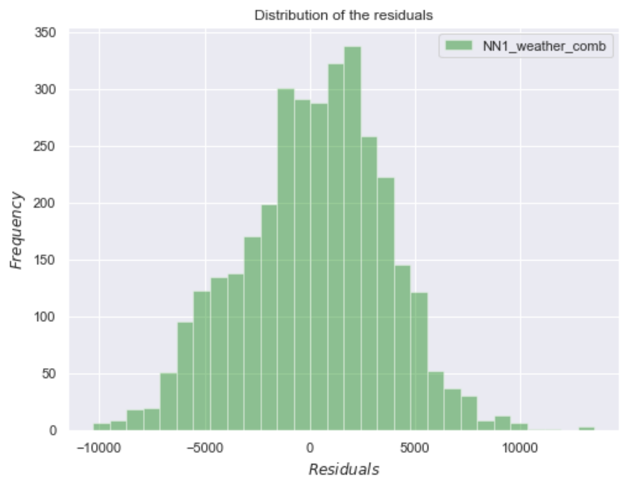
NN2

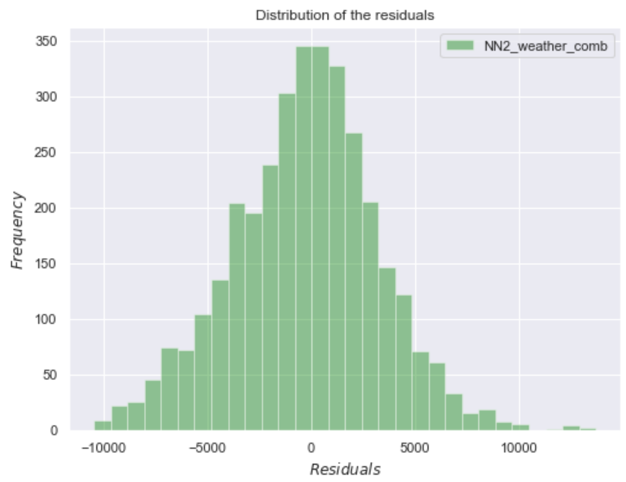


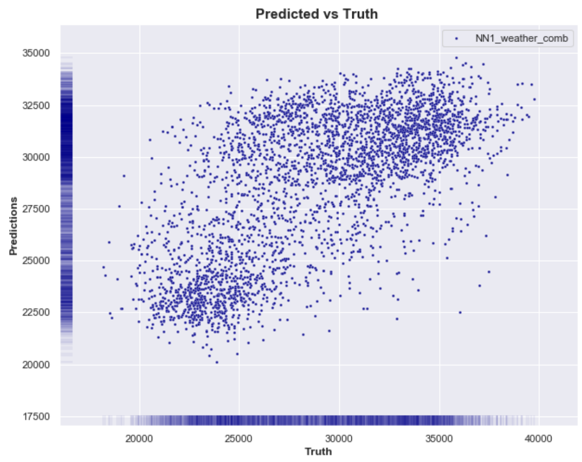
Combined layers for weather and regression

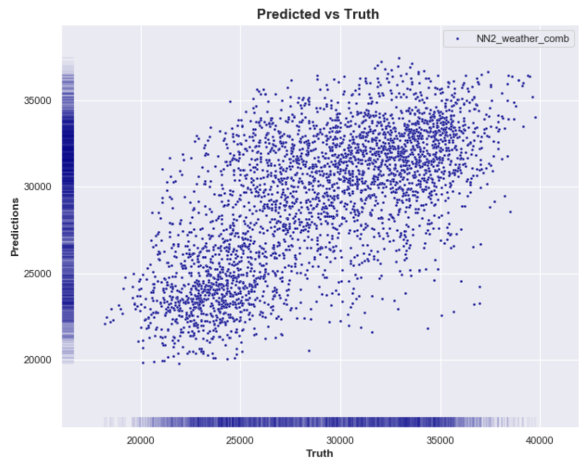
Algorithm	**weather + NN1**	**weather + NN2**
Input features	All columns	All columns
Optimization	see previous models	see previous models
Hyperparameters	see previous models	see previous models
RMSE	3512.390026707932	3844.8934557636785
MAPE	0.0990748177263861	0.10233166635126485
Comments:		

Combined layers for weather and regression



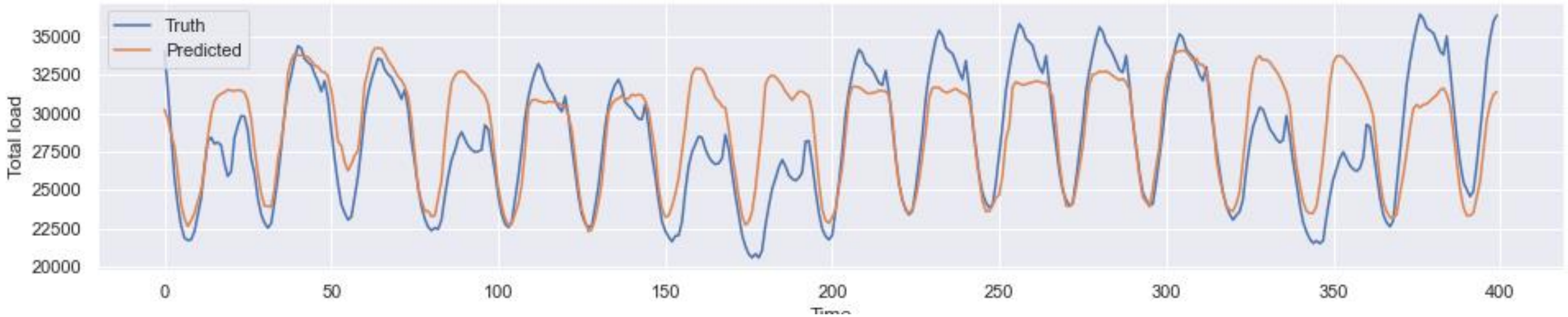




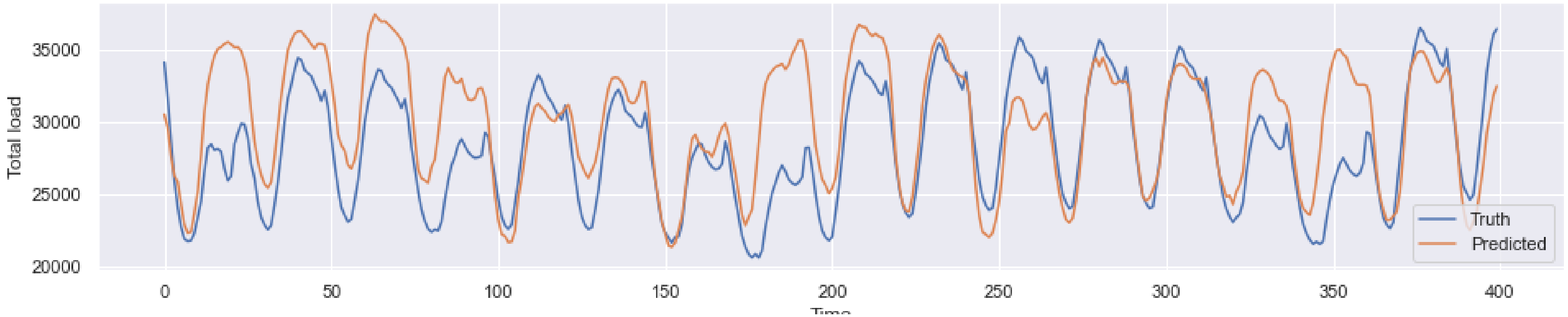


Combined layers for weather and regression

NN1



NN2



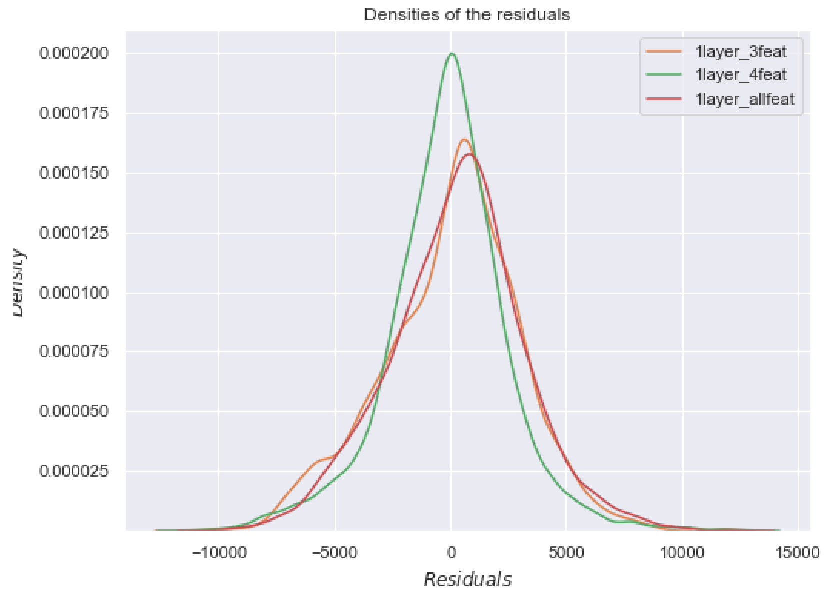
GRU – 1 layer

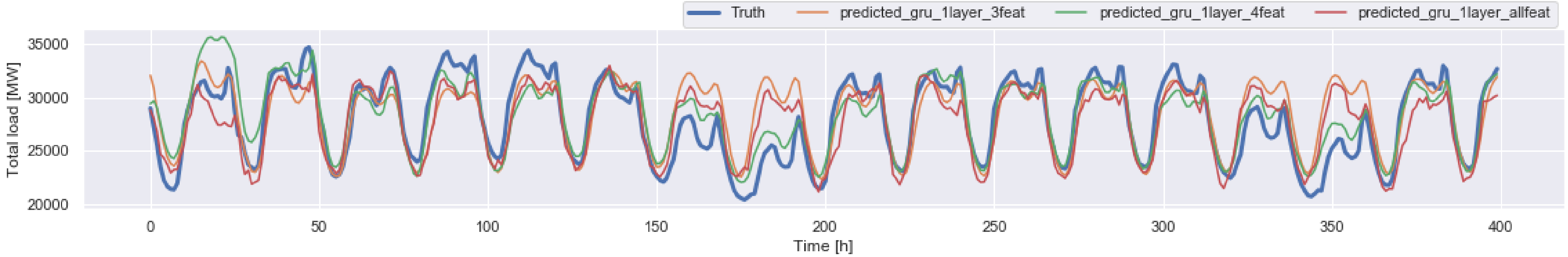
Algorithm	GRU – 1 LAYER	GRU – 1 LAYER	GRU – 1 LAYER
Input features	Valencia_temp_min, Valencia_humidity, time	Valencia_temp_min, Valencia_humidity, time, total actual load	All columns
Optimization	Bayesian	Bayesian	Bayesian
Hyperparameters	N-epochs: earlyStopping Batch size: 210 Learning rate: 0.001105 Num layers: 1 Num neurons (/units): 243 Dropout rate: 0.2 Activation function: tanh	N-epochs: earlyStopping Batch size: 90 Learning rate: 0.000538 Num layers: 1 Num neurons (/units): 297 Dropout rate: 0.35 Activation function: tanh	N-epochs: earlyStopping Batch size: 270 Learning rate: 0.003526 Num layers: 1 Num neurons (/units): 217 Dropout rate: 0.0 Activation function: tanh
RMSE	60.026586	53.434370	56.734219
MAPE	0.097150	0.068926	0.087607
Comments:			

GRU – 1 layer evaluation

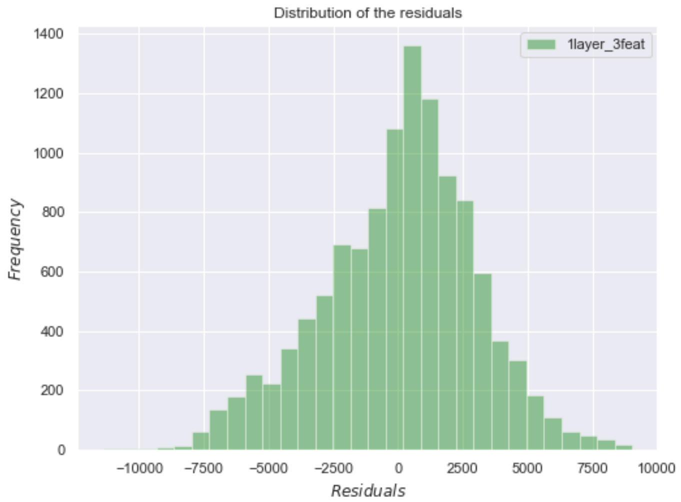
Algorithm	Root MSE	MAPE
GRU,1layer,3feat	3603.61	9.72%
GRU,1layer,4feat	2855.61	6.89%
GRU,1layer,allfeat	3224.60	8.76%

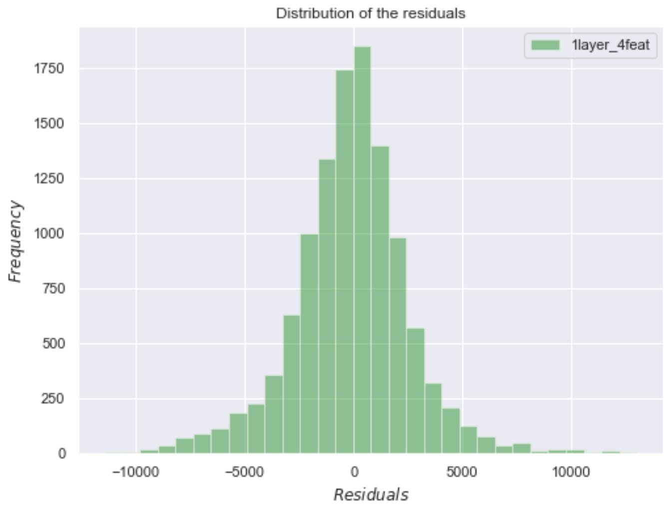
t-test (with RMSE of 10 model trainings): all models are significantly different ($p < 0.05$) from each other

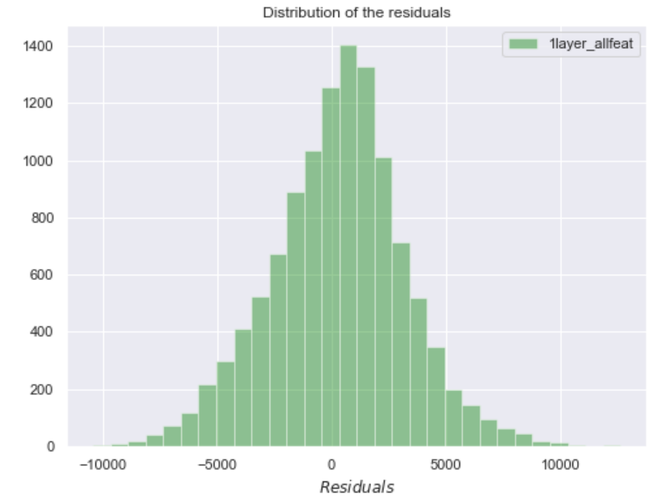


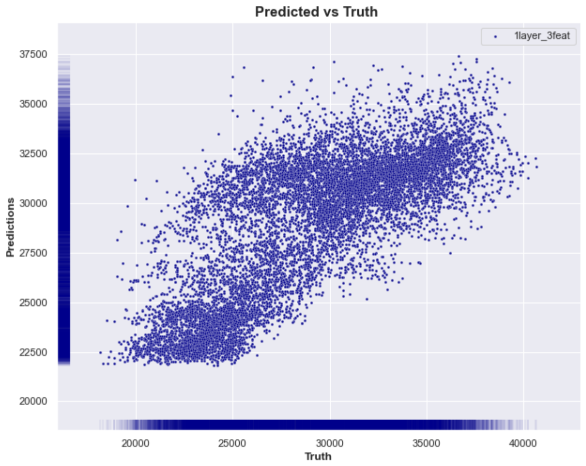


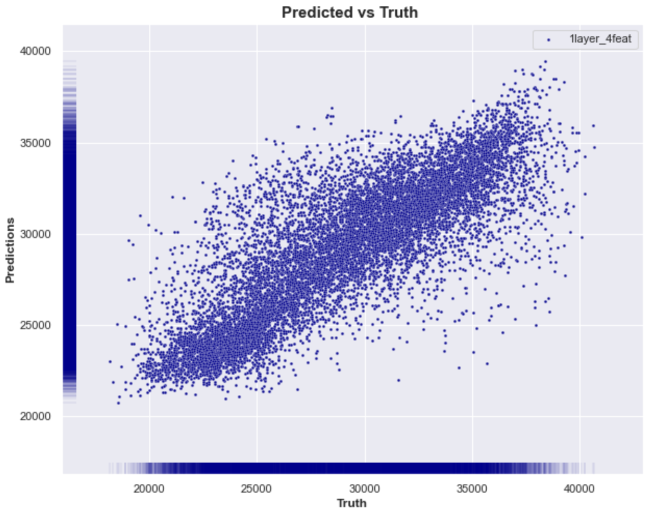
GRU – 1 layer

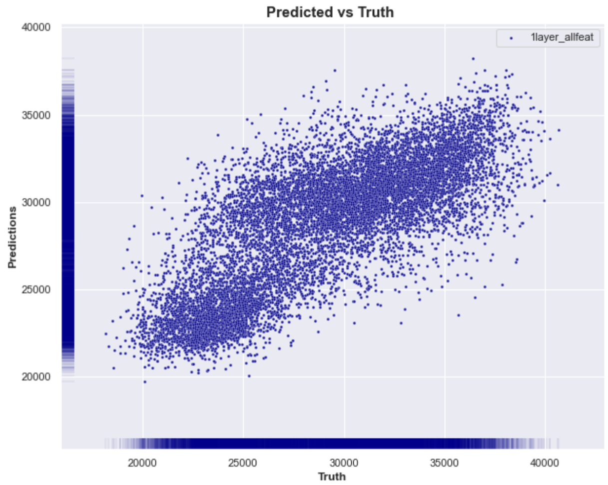

Distribution of the residuals
1layer_3feat
Frequency
Residuals


Distribution of the residuals
1layer_4feat
Frequency
Residuals


Distribution of the residuals
1layer_allfeat
Residuals


Predicted vs Truth
1layer_3feat
Predictions
Truth


Predicted vs Truth
1layer_4feat
Predictions
Truth


Predicted vs Truth
1layer_allfeat
Predictions
Truth

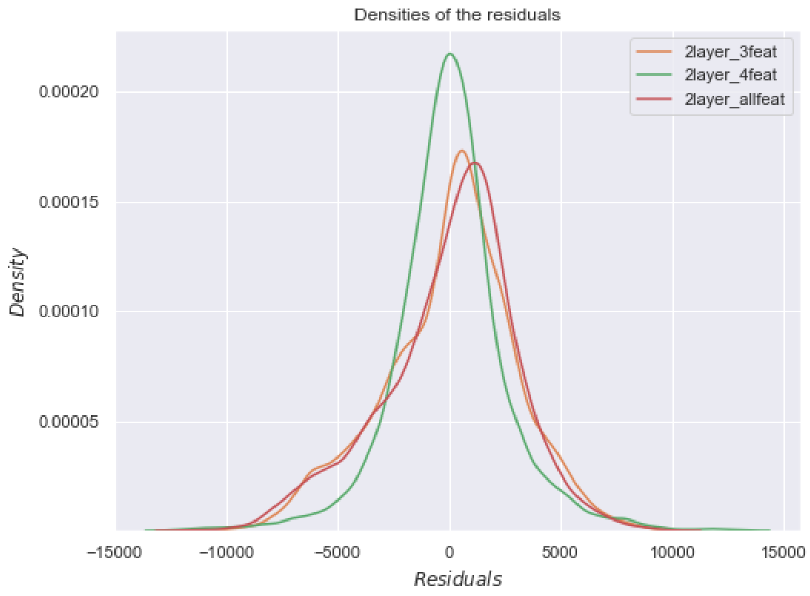
GRU – 2 layer

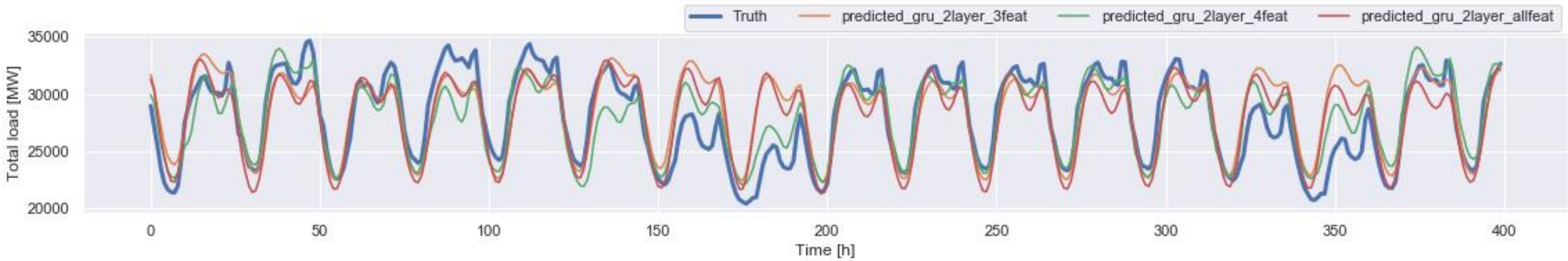
Algorithm	GRU – 2 LAYERS	GRU – 2 LAYERS	GRU – 2 LAYERS
Input features	Valencia_temp_min, Valencia_humidity, time	Valencia_temp_min, Valencia_humidity, time, total actual load	All columns
Optimization	Bayesian	Bayesian	Bayesian
Hyperparameters	N-epochs: earlyStopping Batch size: 190 Learning rate: 0.002375 Num layers: 2 Num neurons (/units): 105/195 Dropout rate: 0.35/0.1 Activation function: tanh	N-epochs: earlyStopping Batch size: 250 Learning rate: 0.004231 Num layers: 2 Num neurons (/units): 10/130 Dropout rate: 0.5/0.35 Activation function: tanh	N-epochs: earlyStopping Batch size: 50 Learning rate: 0.000100 Num layers: 2 Num neurons (/units): 10/300 Dropout rate: 0.5/0.4 Activation function: tanh
RMSE	61.099674	51.017209	58.840459
MAPE	0.098331	0.061522	0.092708
Comments:			

GRU – 2 layers, evaluation

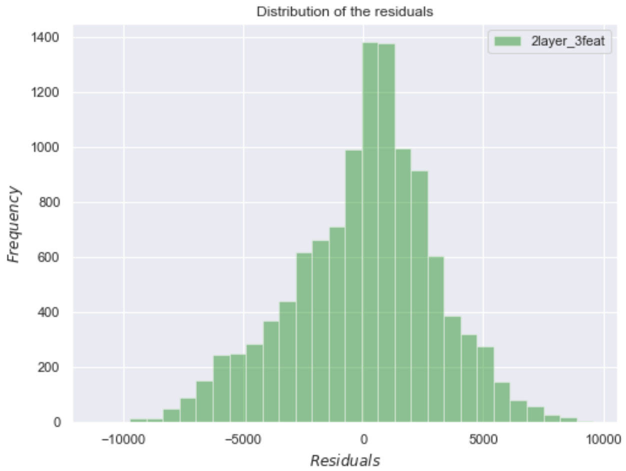
Algorithm	Root MSE	MAPE
GRU,2layer,3feat	3733.50	9.83%
GRU,2layer,4feat	2602.95	6.15%
GRU,2layer,allfeat	3465.12	9.27%

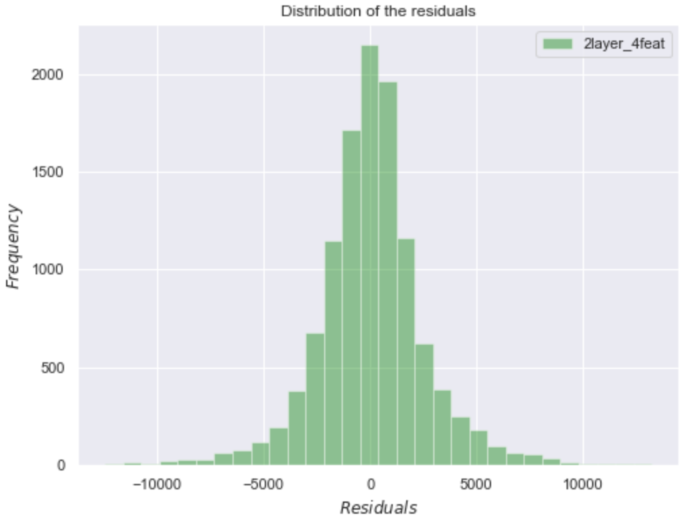
t-test (with RMSE of 10 model trainings): all models are significantly different ($p < 0.05$) from each other

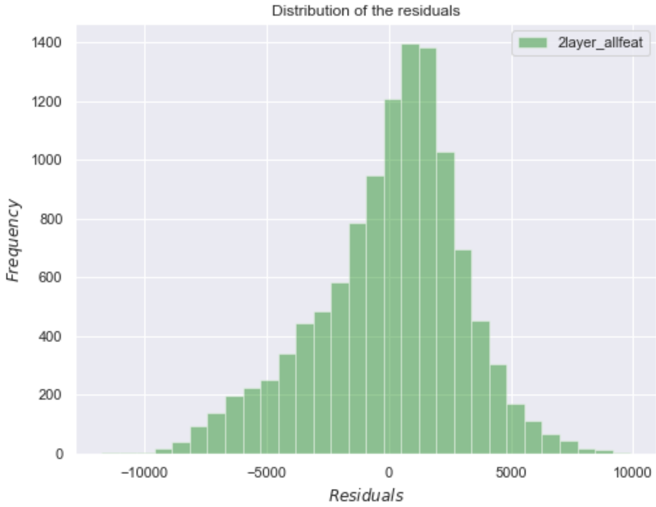


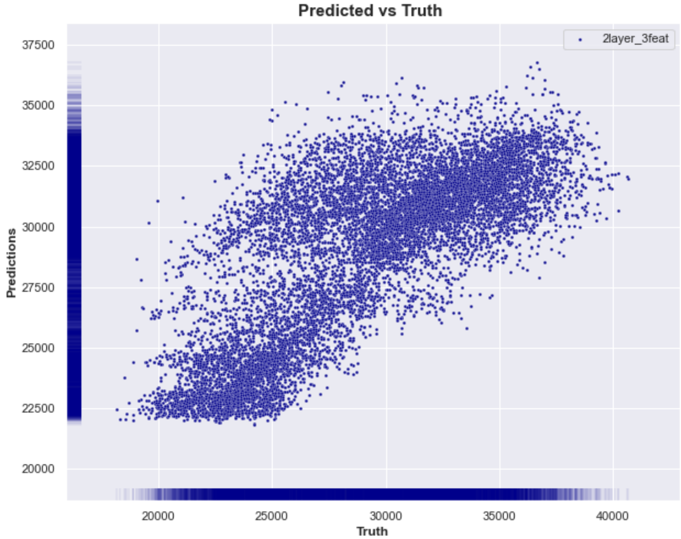


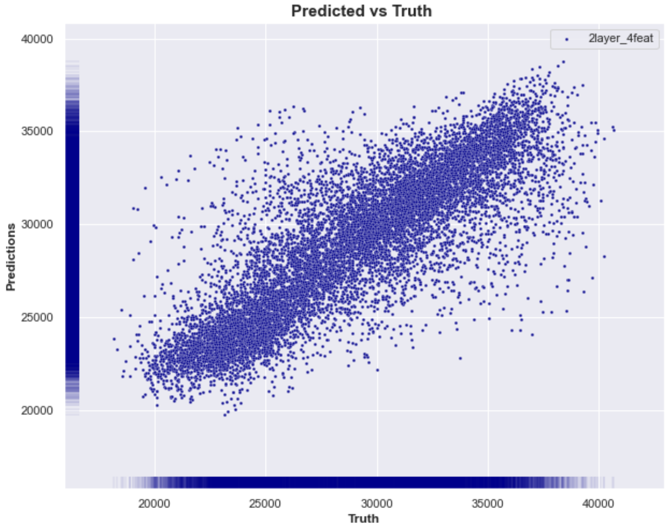
GRU – 2 layers

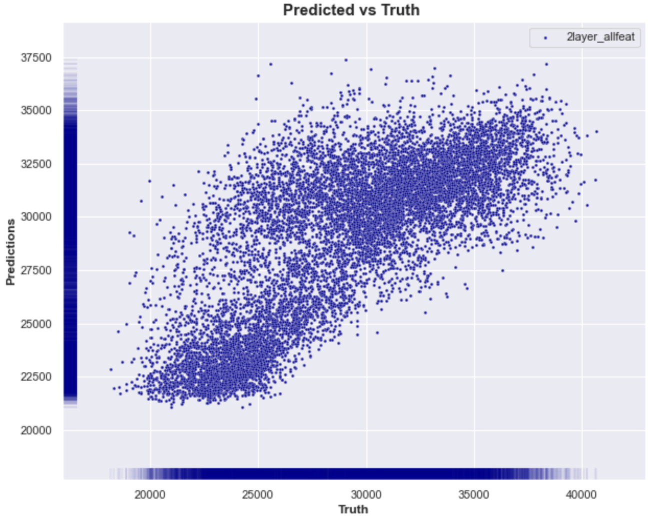

Distribution of the residuals
2layer_3feat
Frequency
Residuals


Distribution of the residuals
2layer_4feat
Frequency
Residuals


Distribution of the residuals
2layer_allfeat
Frequency
Residuals


Predicted vs Truth
2layer_3feat
Predictions
Truth


Predicted vs Truth
2layer_4feat
Predictions
Truth


Predicted vs Truth
2layer_allfeat
Predictions
Truth

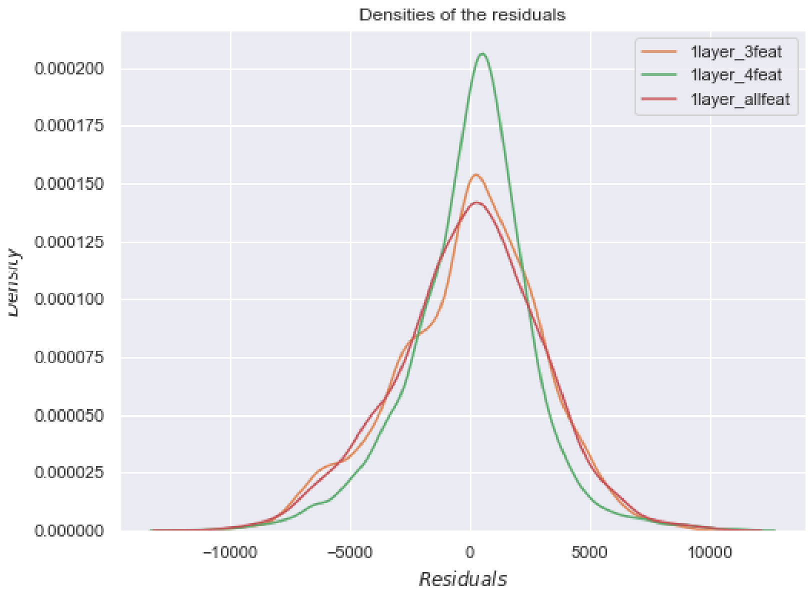
LSTM – 1 layer

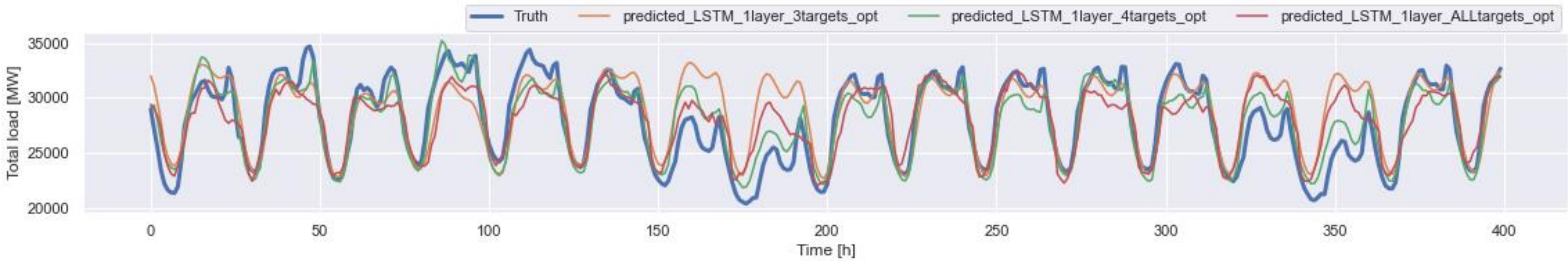
Algorithm	LSTM - 1 LAYER	LSTM - 1 LAYER	LSTM - 1 LAYER
Input features	All columns	Valencia_temp_min, Valencia_humidity, time, total actual load	Valencia_temp_min, Valencia_humidity, time
Optimization	Bayesian	Bayesian	Bayesian
Hyperparameters	N-epochs: earlyStopping Batch size: 330 Learning rate: 0.003279 Num layers: 1 Num units: 11 Dropout rate: 0.45 Activation function: tanh	N-epochs: arlyStopping Batch size: 390 Learning rate: 0.008959 Num layers: 1 Num units: 56 Dropout rate: 0.5 Activation function: tanh	N-epochs: earlyStopping Batch size: 230 Learning rate: 0.003151 Num layers: 1 Num units: 77 Dropout rate: 0.45 Activation function: tanh
MSE (unscaled)	12084602.37	8822483.31	13802724.46
MSE (scaled)	0.581009	0.424171	0.663613
MAPE (unscaled)	0.090475	0.072125	0.098662
Comments:			

LSTM – 1 layer, evaluation

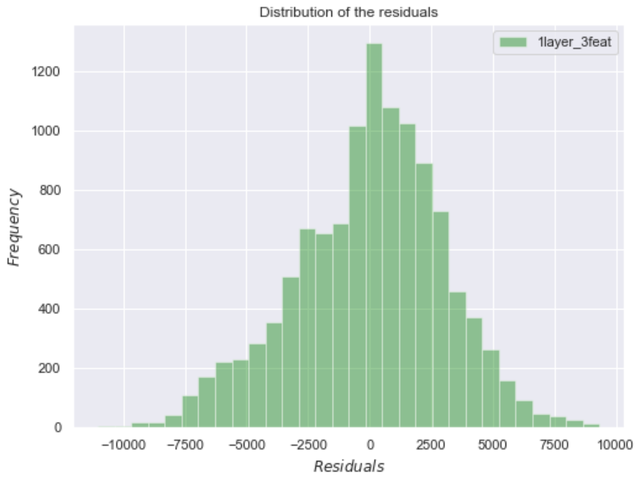
Algorithm	RMSE	MAPE
LSTM,1layer,3feat	3476.29	9.87%
LSTM,1layer,4feat	2970.27	7.21%
LSTM,1layer,allfeat	3715.20	9.05%

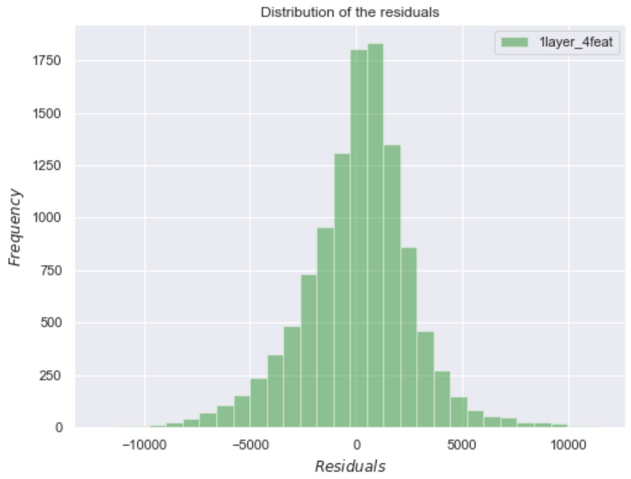
t-test (with RMSE of 10 model trainings): all models are significantly different ($p < 0.05$) from each other

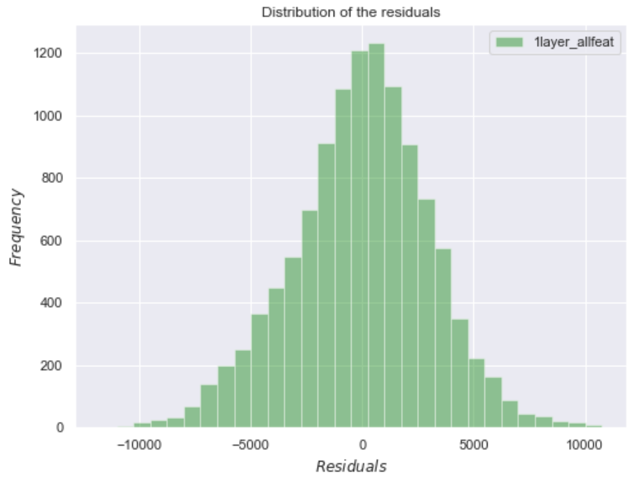


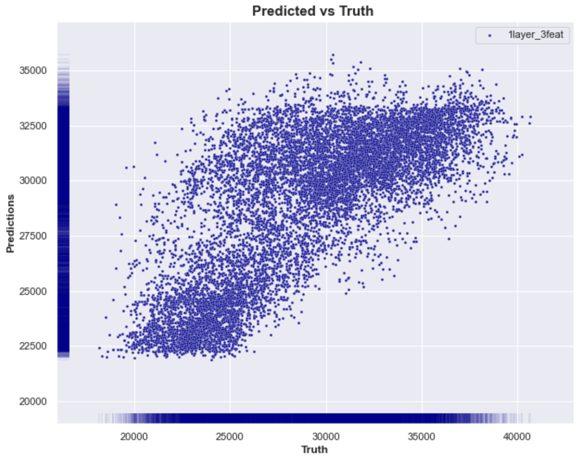


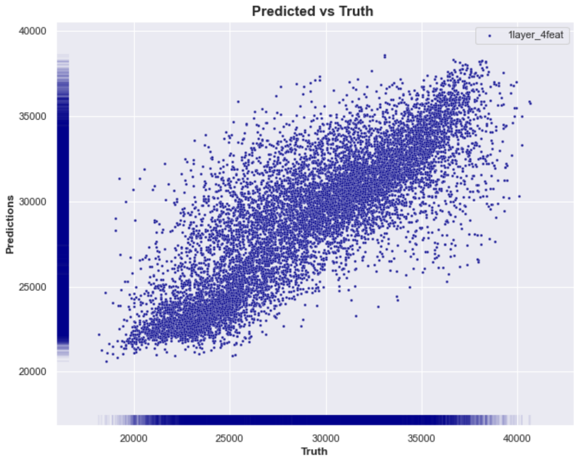
LSTM – 1 layer

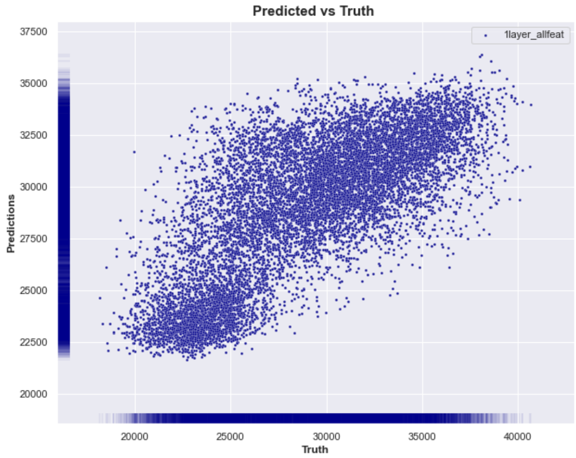

Distribution of the residuals
1layer_3feat
Frequency
Residuals


Distribution of the residuals
1layer_4feat
Frequency
Residuals


Distribution of the residuals
1layer_allfeat
Frequency
Residuals


Predicted vs Truth
1layer_3feat
Predictions
Truth


Predicted vs Truth
1layer_4feat
Predictions
Truth


Predicted vs Truth
1layer_allfeat
Predictions
Truth

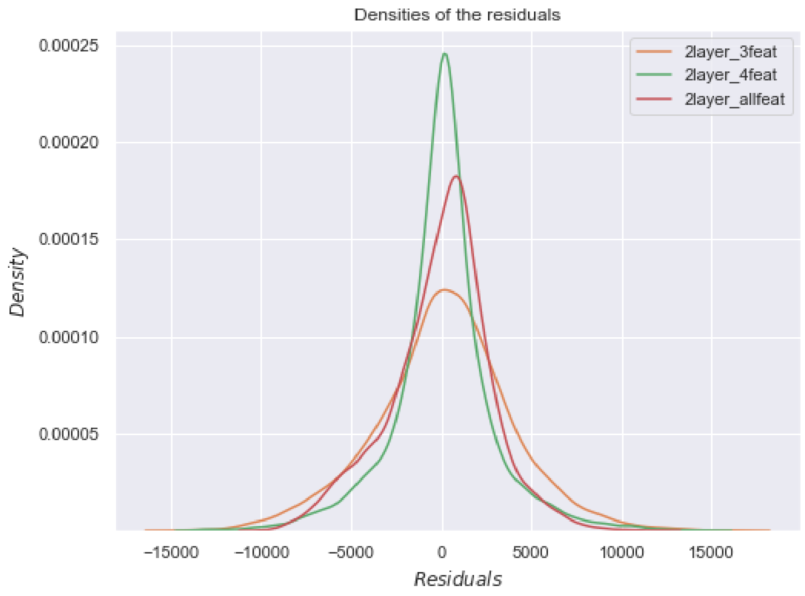
LSTM – 2 layers

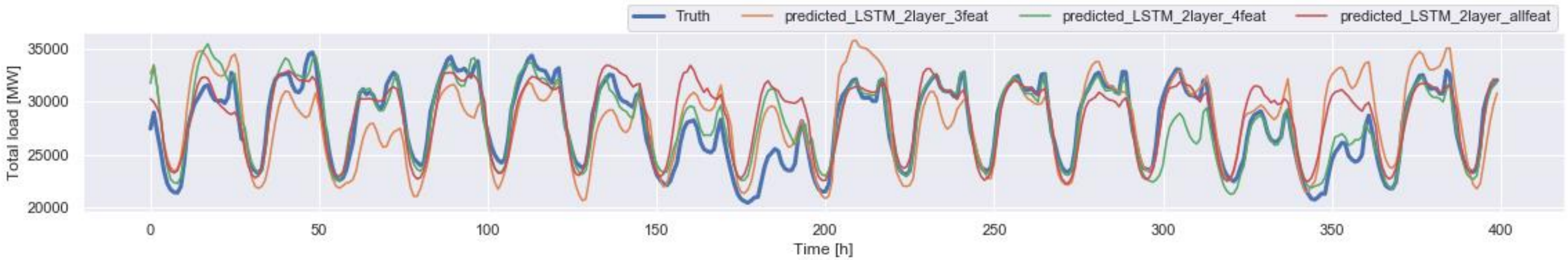
Algorithm	LSTM - 2 LAYERS	LSTM - 2 LAYERS	LSTM – 2 LAYERS
Input features	All columns	Valencia_temp_min, Valencia_humidity, time, total actual load	Valencia_temp_min, Valencia_humidity, time
Optimization	Bayesian	Bayesian	Bayesian
Hyperparameters	N-epochs: earlyStopping Batch size: 390 Learning rate: 0.002795 Num layers: 2 Num units: 10, 123 Dropout rate: 0.0, 0.0 Activation function: tanh	N-epochs: earlyStopping Batch size: 390 Learning rate: 0.006347 Num layers: 2 Num units: 10, 108 Dropout rate: 0.0, 0.4 Activation function: tanh	N-epochs: earlyStopping Batch size: 290 Learning rate: 0.000856 Num layers: 1 Num units: 10, 10 Dropout rate: 0.15, 0.5 Activation function: tanh
MSE (unscaled)	1.412757e+07	7.942777e+06	1.373014e+07
MSE (scaled)	0.679232	0.381876	0.660123
MAPE (unscaled)	0.099327	0.067231	0.098061
Comments:			

LSTM – 2 layer, evaluation

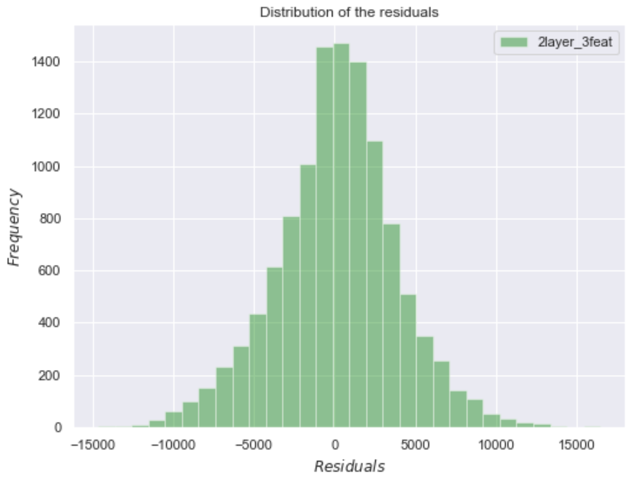
Algorithm	RMSE	MAPE
LSTM,2layers,3feat	3705.42	9.81%
LSTM,2layers,4feat	2818.29	6.72%
LSTM,2layers,allfeat	3758.67	9.93%

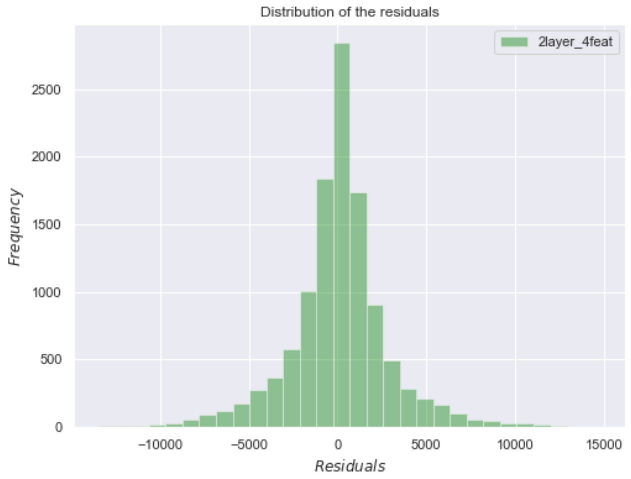
t-test (with RMSE of 10 model trainings): 4 features significantly different ($p < 0.05$) from 3 and all features, but 3 and all not significantly different from each other

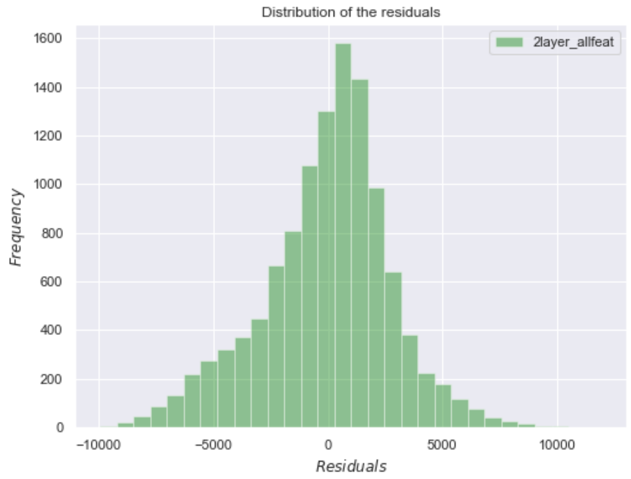


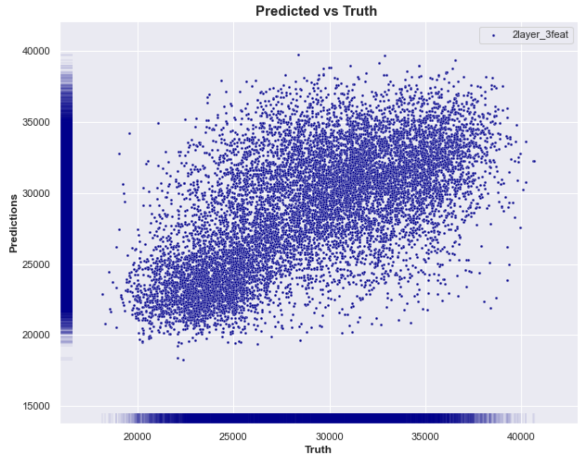


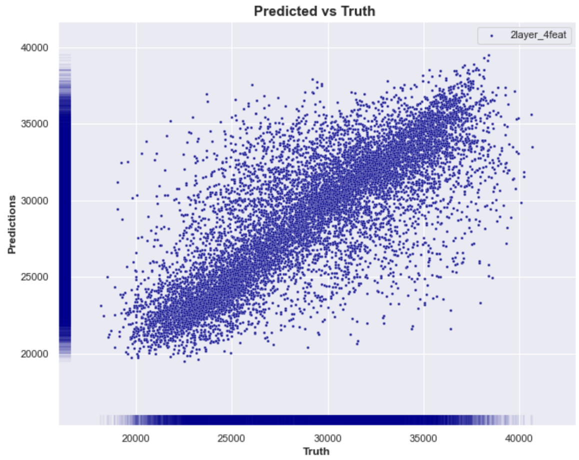
LSTM – 2 layer

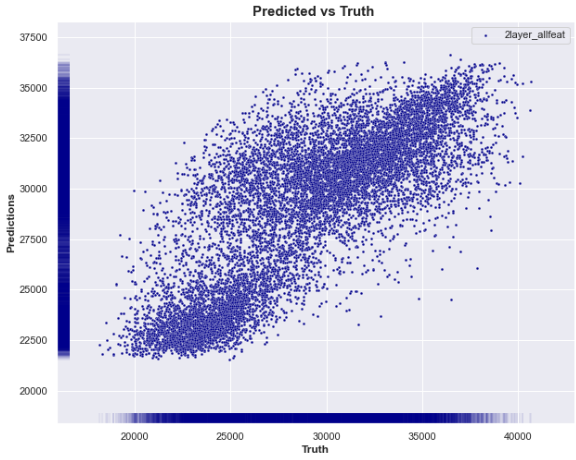

Distribution of the residuals
2layer_3feat
Frequency
Residuals


Distribution of the residuals
2layer_4feat
Frequency
Residuals


Distribution of the residuals
2layer_allfeat
Frequency
Residuals


Predicted vs Truth
2layer_3feat
Predictions
Truth


Predicted vs Truth
2layer_4feat
Predictions
Truth


Predicted vs Truth
2layer_allfeat
Predictions
Truth