Energy Price Predictions using SSMs

GROUP 15

CYRIL SMETANKA

RAPHAEL SUTTER

LE MINH TRUONG

Project Objectives

- Infer energy prices in the Netherlands based on weather, energy consumption/production and temporal features.
- Develop and compare deep learning models for price forecasting
- Models evaluated
  - Deep State-Space Model (Gu et al. 2021)
  - Structured State Spaces for Sequence Modeling (Gu et al. 2022)
  - Mamba (Gu & Dao. 2023)
  - Linear Regression

Project Outline

- Data extraction and transformation
- Training each model
- Evaluating each model
- Comparison between models
- Discussion

The Dataset

Input Features:

- Spanning back 3 years: n = 26280
- Temperature, Solar Radiation, Wind Speed, Energy Consumption, Solar Production, Wind Production in the Netherlands
- Temporal features
  - Is_daytime, sin_hour, cos_hour, price_lag_1h

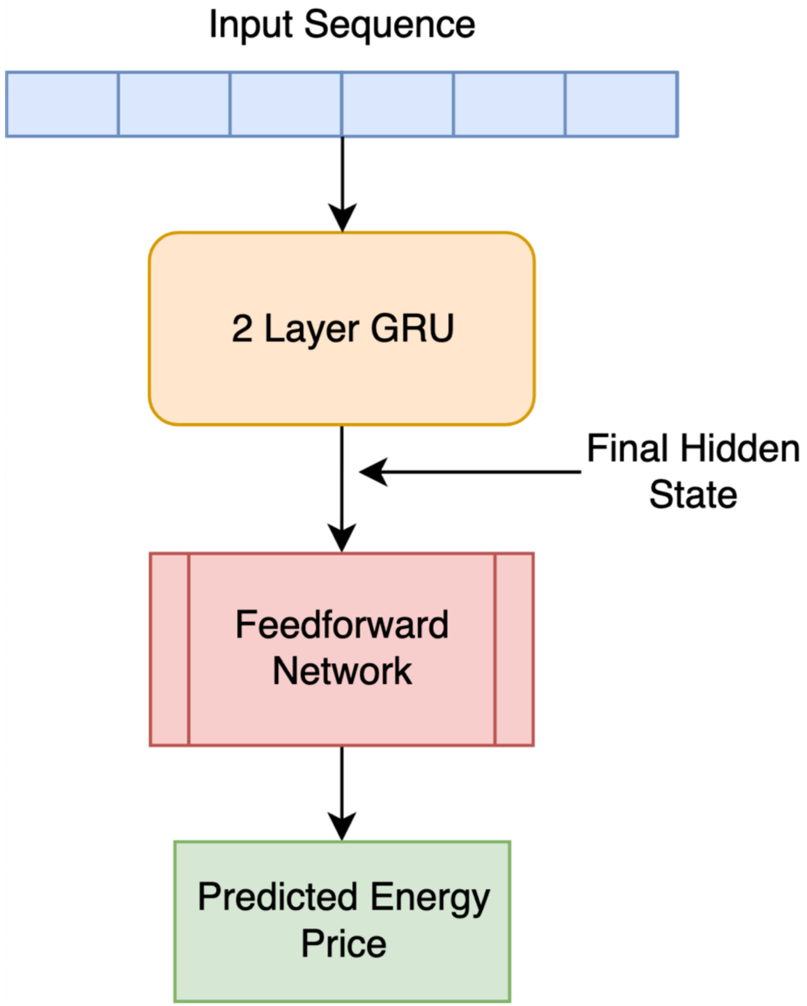
Target:

- Price (EUR/MWh)

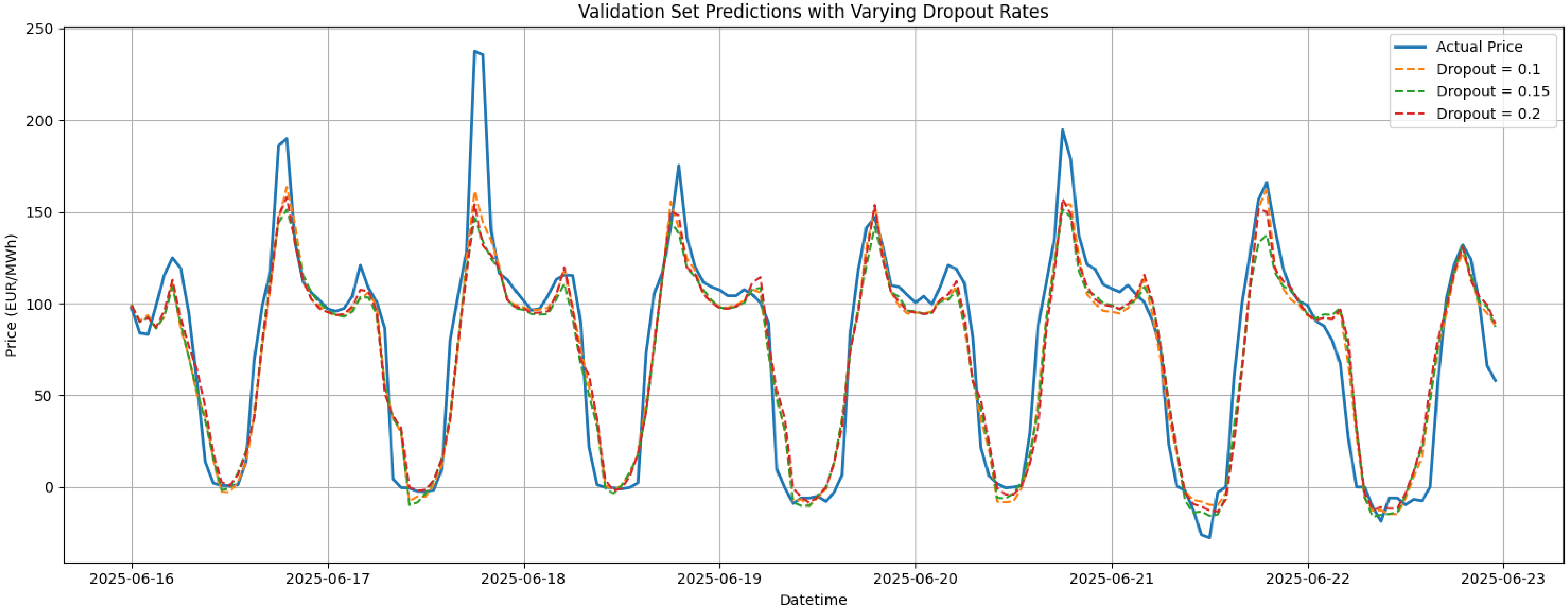
Deep State-Space Model (DSSM)

Architecture:

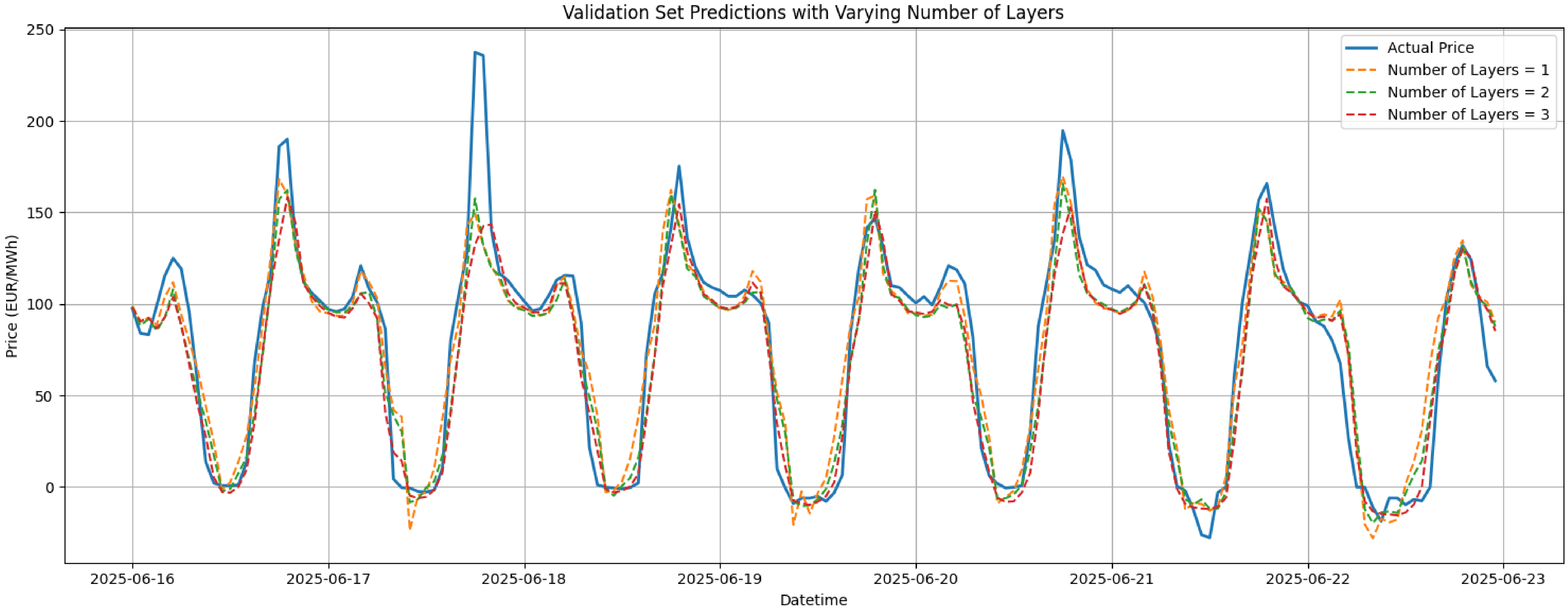
- 2 layer GRU for state representation
  - Dropout (default = 0.1) between layers
- Feedforward Network (Observational model)
  - Two hidden layers with RELU activation
  - Final linear output layer with output neuron



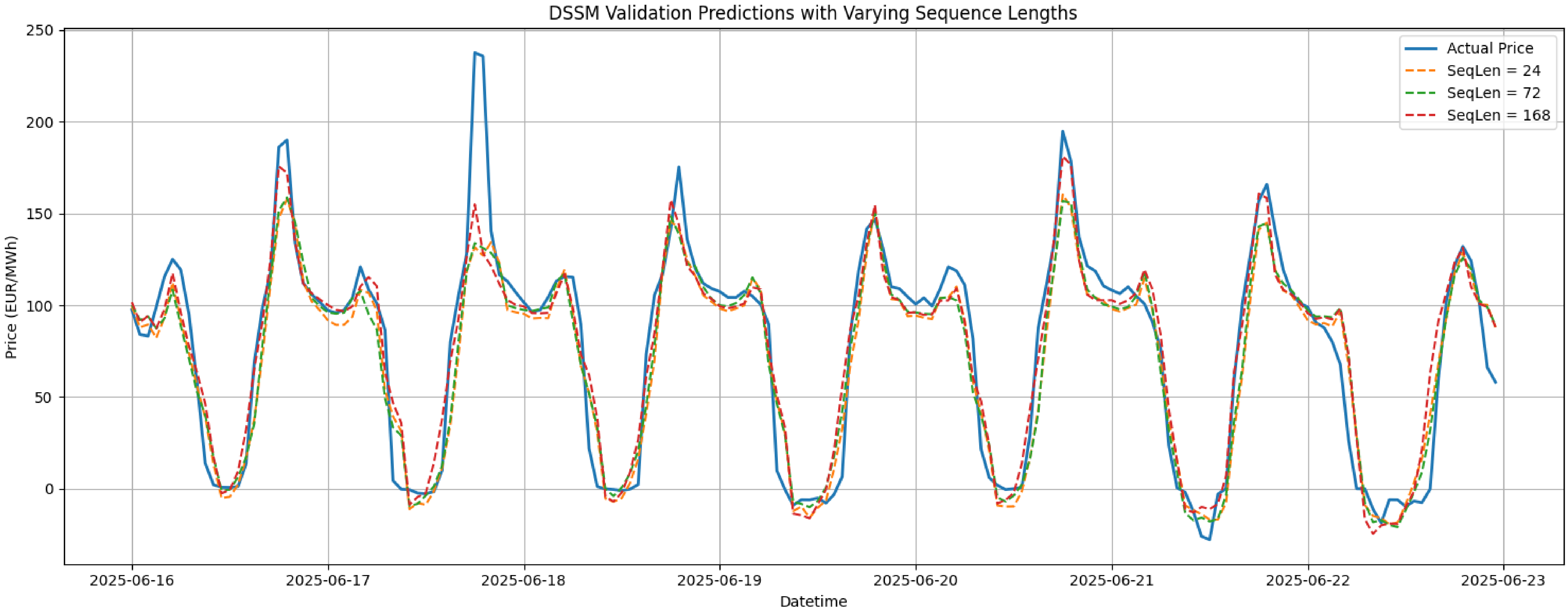
DSSM – Dropout Rate



DSSM – Number of Layers



DSSM – Sequence Length



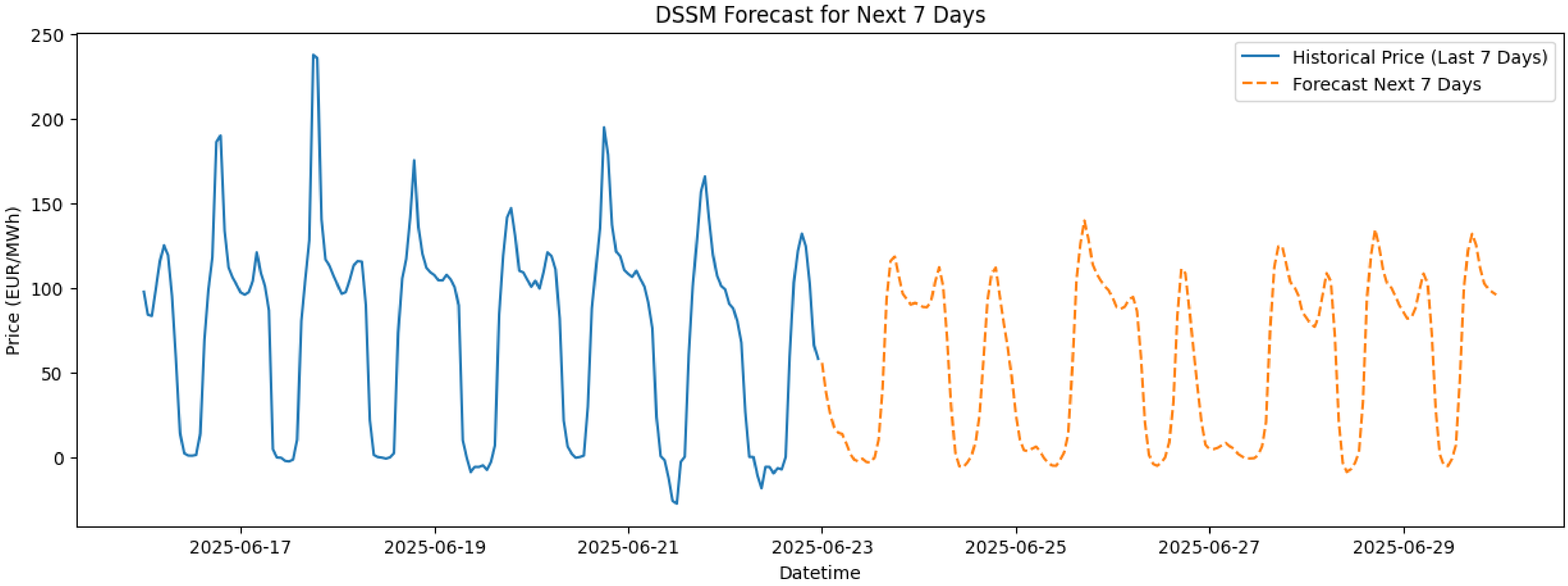
DSSM - Hyperparameter Results

	Drop out	Sequence length	Number of layers	Evaluation Metrics			
				MAE	MSE	SMAPE	Accuracy
Drop out	0.1	24	2	13.38	366.87	50.48%	96.43%
	0.15			13.38	368.47	49.29%	98.81%
	0.2			13.85	418.23	48.70%	98.21%
Sequence length	0.15	24	2	14.52	432.39	50.30%	96.66%
		72		13.83	412.63	50.04%	98.21%
		168		12.77	386.94	44.99%	98.21%
Num layers	0.15	24	1	14.53	423.21	50.38%	97.62%
			2	14.46	408.66	49.99%	98.81%
			3	13.26	397.73	47.25%	98.21%

DSSM Results

	Short-term (0-24h)	Medium-term (24-72h)	Long-term (72-168h)
MAE	10.67	14.05	14.83
MSE	211.01	381.52	372.25
SMAPE	25.14%	71.59%	41.76%
Accuracy	100.00%	95.83%	98.96%

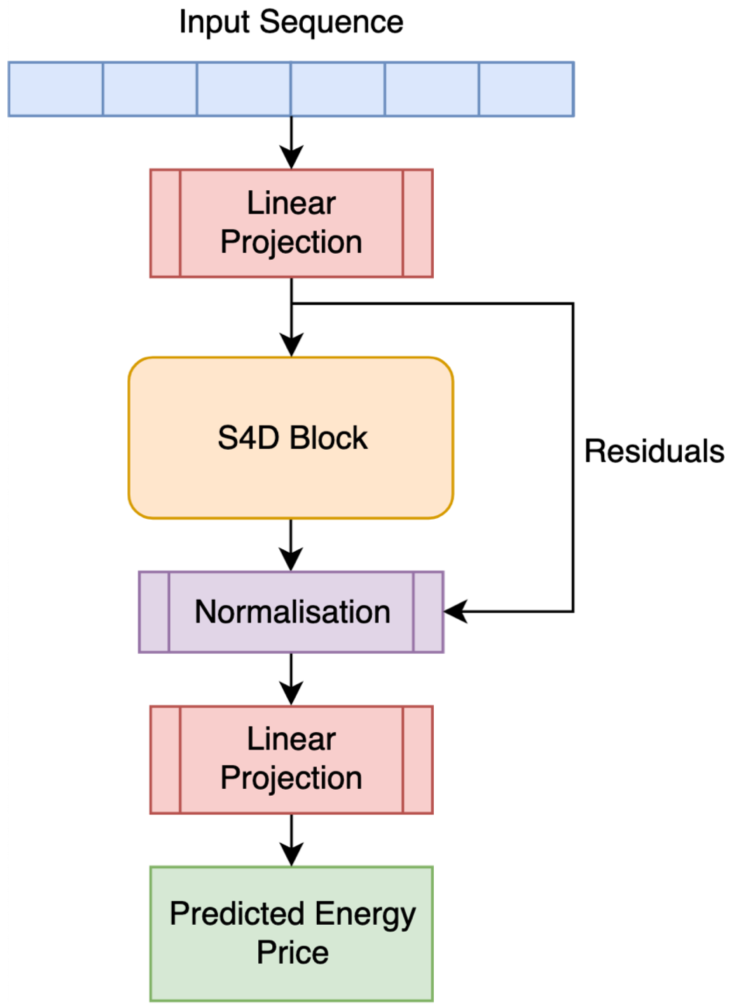
DSSM – Week Forecast



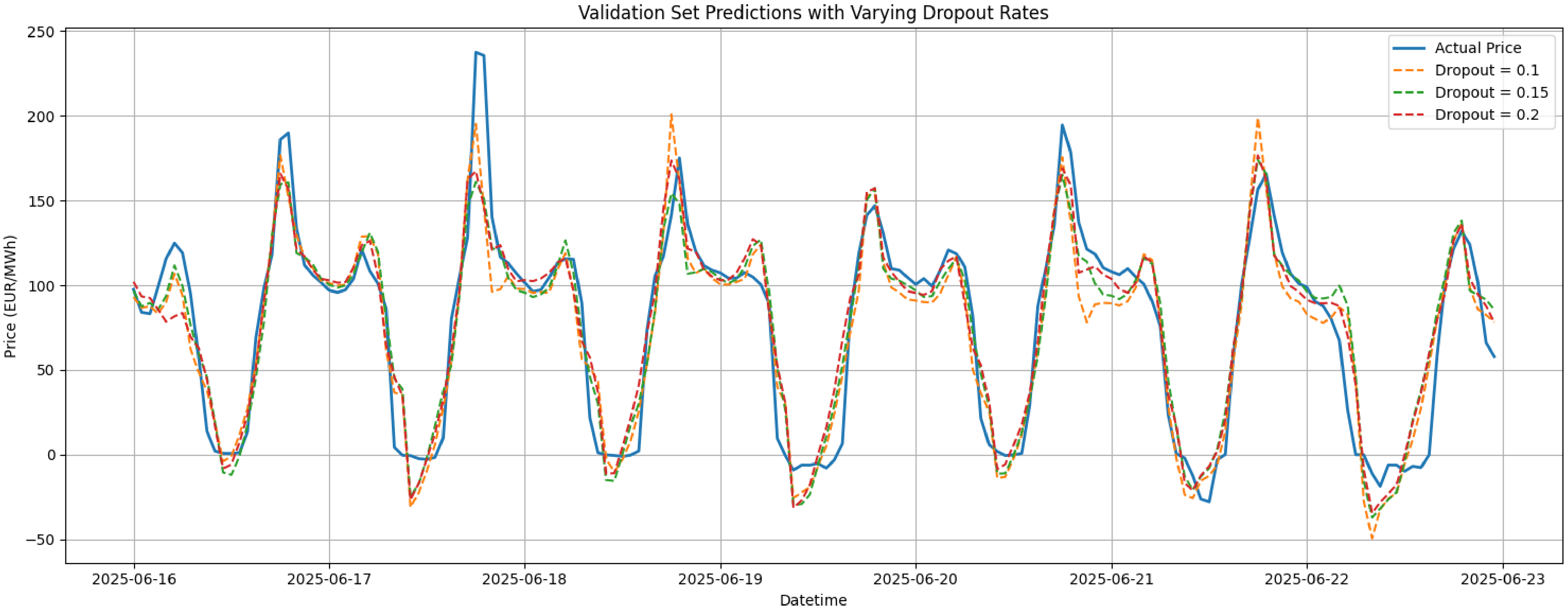
Structured State Spaces for Sequence Modeling (S4)

Architecture:

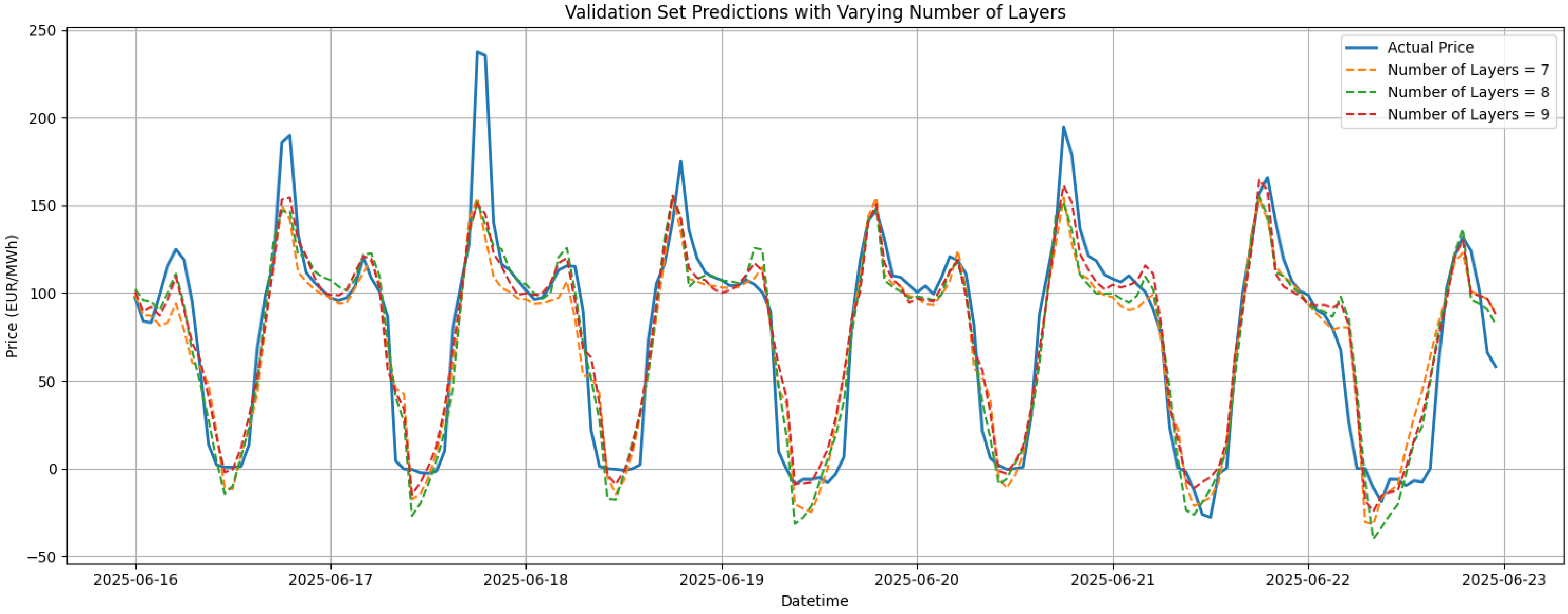
- Linear layer to project input features to hidden_size
- S4D Block (Structured State Space Layers) containing
  - Diagonal state-space dynamics (A, B, C, D matrices)
  - Learned time step (Δt) per hidden unit
  - Option for recurrent or convolutional mode
  - Residual connections and LayerNorm
- GELU activation and dropout (default=0.15) for regularization
- Final linear output layer projecting from hidden_size → 1



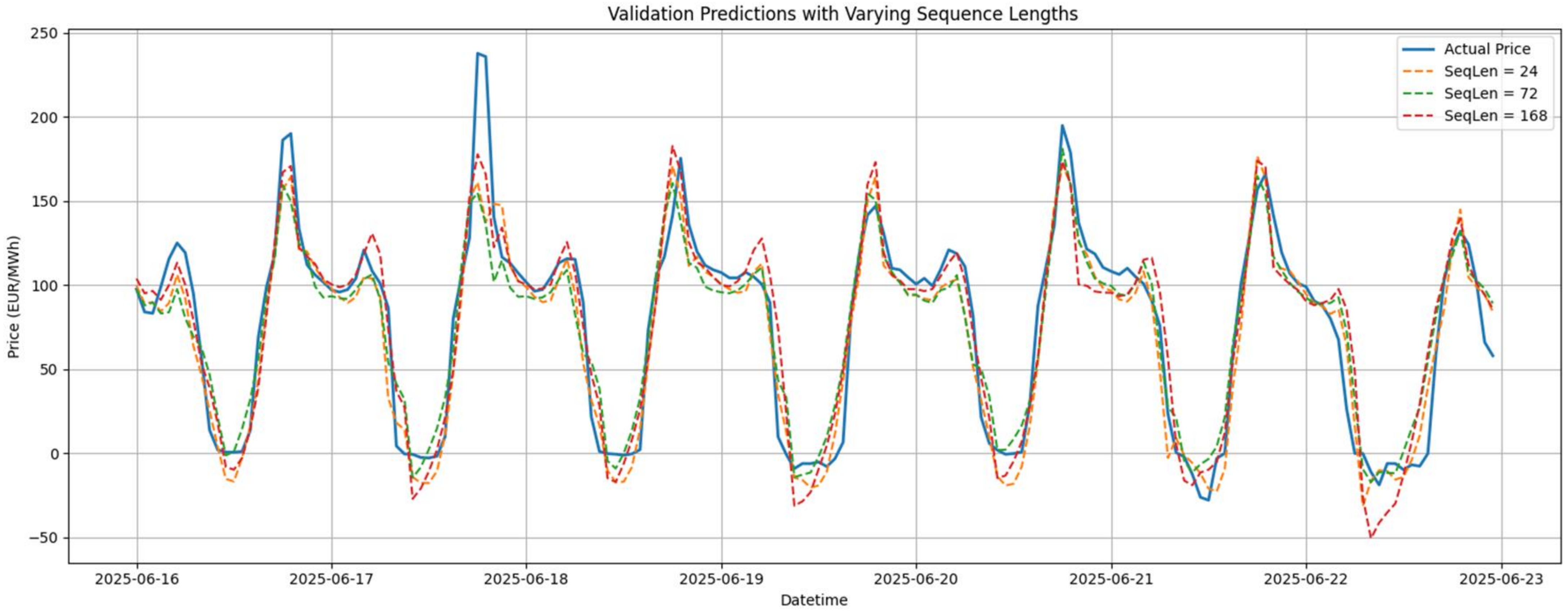
S4 – Dropout Rate



S4 – Number of Layers



S4 – Sequence Length



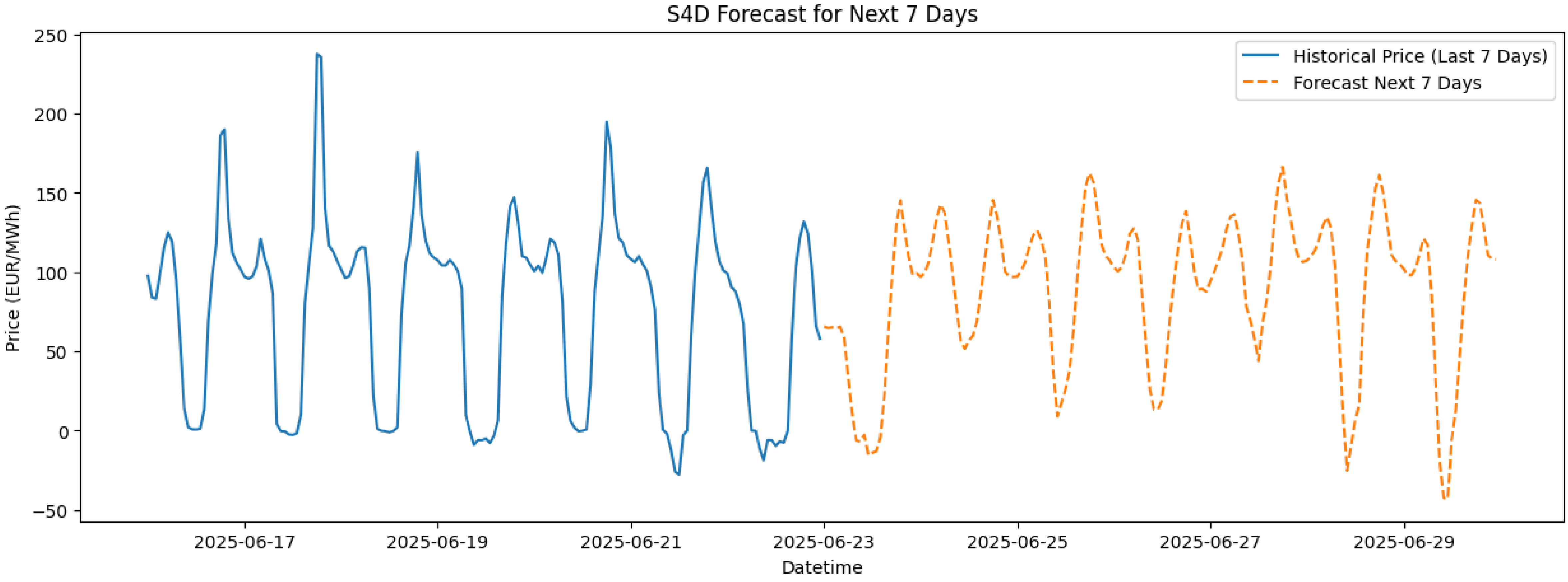
S4 - Hyperparameter Results

	Drop out	Sequence length	Number of layers	Evaluation Metrics			
				MAE	MSE	SMAPE	Accuracy
Drop out	0.1 0.15 0.2	72	9	16.72 15.65 15.42	459.46 429.22 435.02	55.78% 55.59% 56.62%	98.81% 98.21% 97.62%
Sequence length	0.2	24 72 168	9	14.05 15.16 15.09	356.33 431.47 400.22	51.22% 54.20% 54.63%	98.81% 98.81% 97.02%
Num layers	0.2	72	7 8 9	16.18 14.71 13.76	512.82 412.35 397.96	54.55% 53.08% 53.62%	98.21% 98.21% 98.81%

S4 Results

	Short-term (0-24h)	Medium-term (24-72h)	Long-term (72-168h)
MAE	18.42	14.41	18.40
MSE	446.33	374.43	537.71
SMAPE	32.20%	59.14%	43.10%
Accuracy	96.81%	96.81%	96.81%

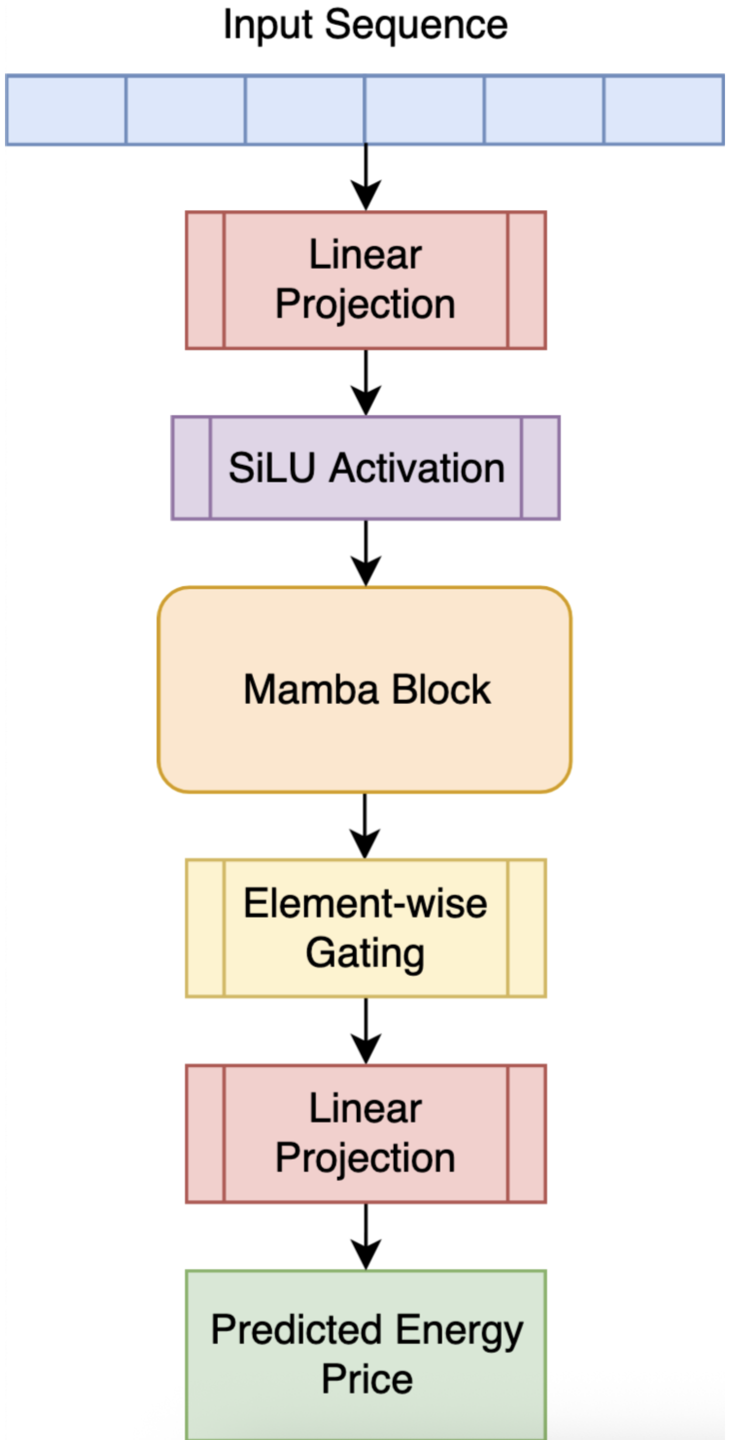
S4 – Week Forecast



Mamba

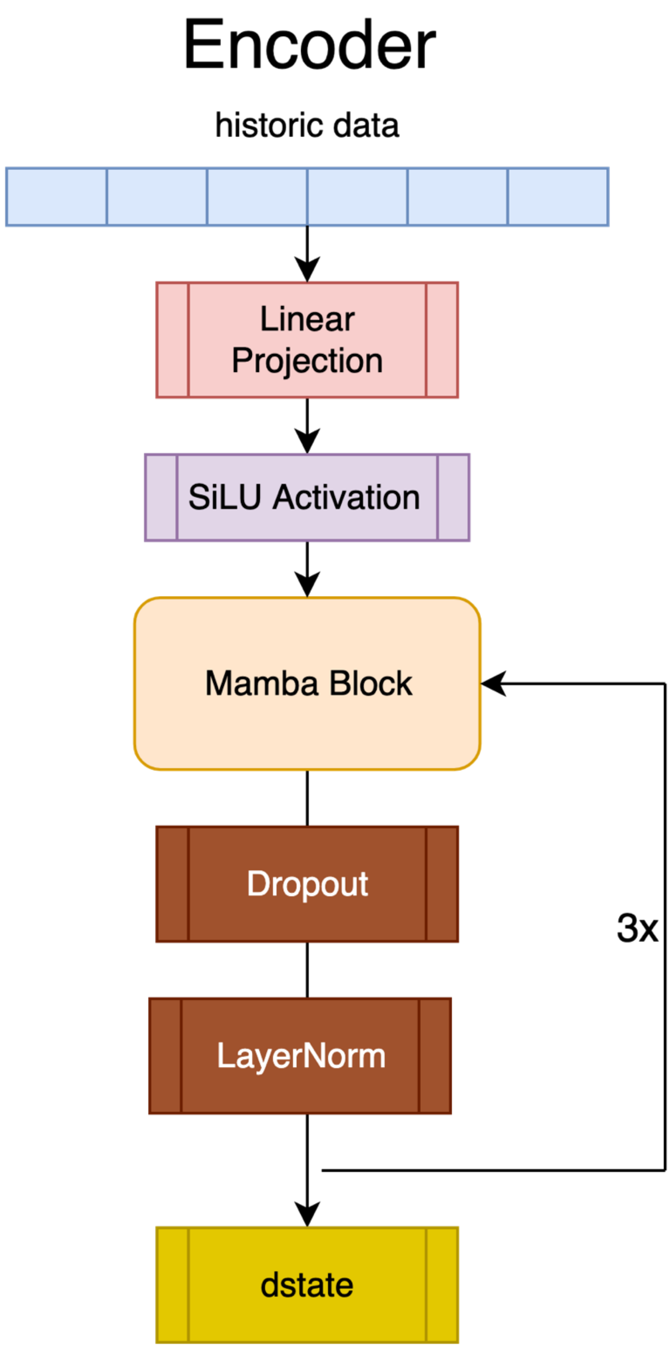
Architecture:

- Linear projection maps raw input features to hidden_size
- Mamba block
  - Depth-wise 1D convolution for token mixing
  - Structured State Space Model (SSM) with:
    - Diagonal state matrix A (log-parameterized)
    - Input-dependent B, C, and Δt (learned per time step)
    - Dynamic recurrent state update per hour
  - Query-key interaction (q_proj, k_proj) fused with SSM output
  - Element-wise gated output: SSM_out $\odot$ SiLU(query)
- Final linear layer maps to a single energy price per hour



Mamba

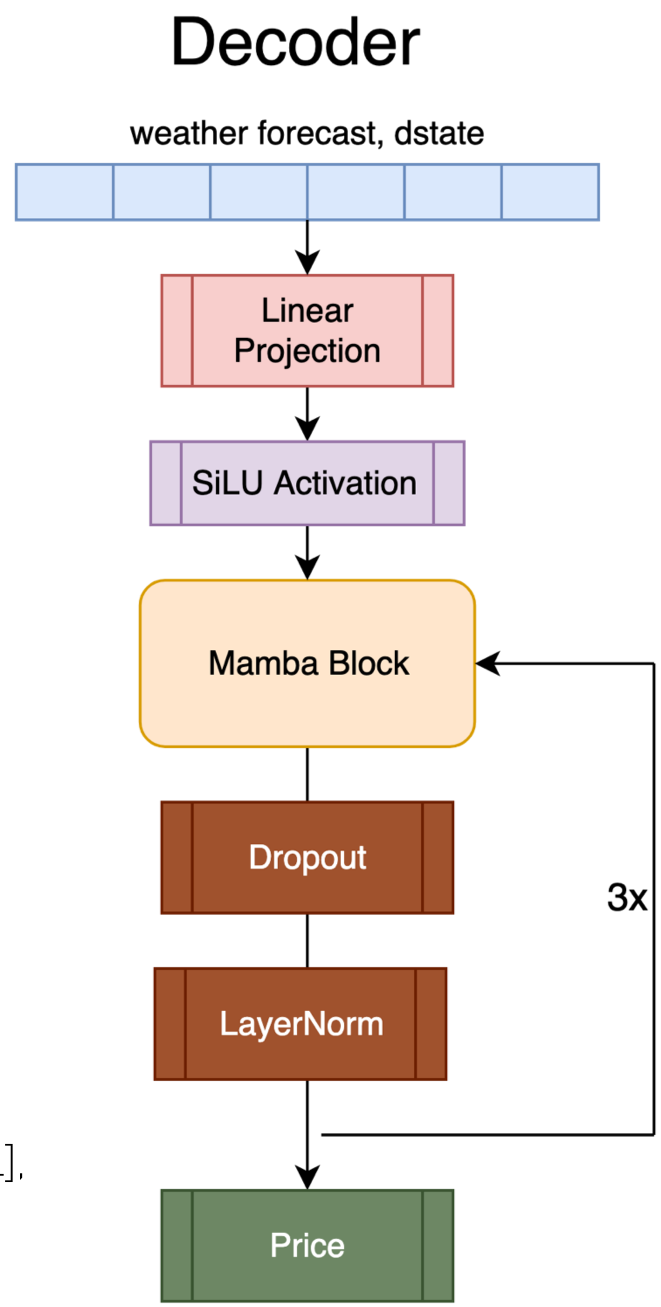
- Linear Projection
- Projects raw historical features [B,Thist,nhist][B,Thist ,nhist] into model dimension [B,Thist,dmodel][B,Thist ,dmodel].
- Stacked Mamba Blocks (repeated layers 3x)
- Mamba SSM (diagonal AA, time- and input-dependent B,C,ΔtB,C,Δt, per-step recurrent update)
- Dropout (p=0.1)
- LayerNorm over the dmodeldmodel axis
- Residual connection around each block
- Summary Extraction
- Take the final time-step output as the encoder state ("dstatedstate"), shape [B,dmodel][B,dmodel].



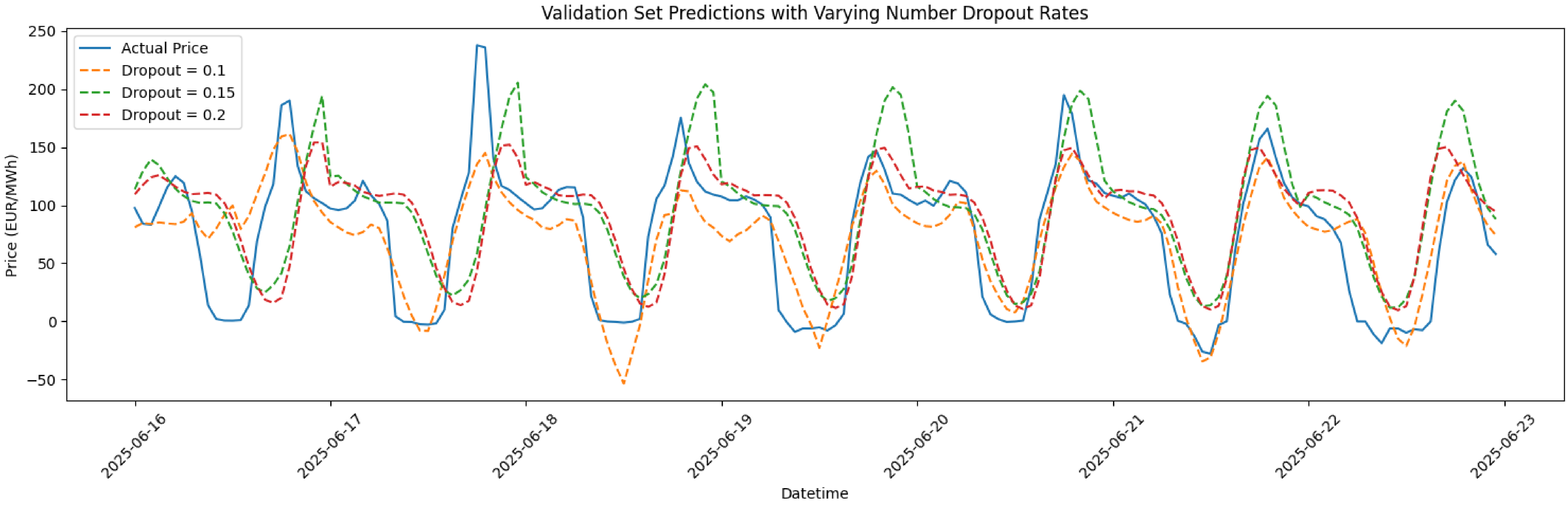
Mamba

Decoder

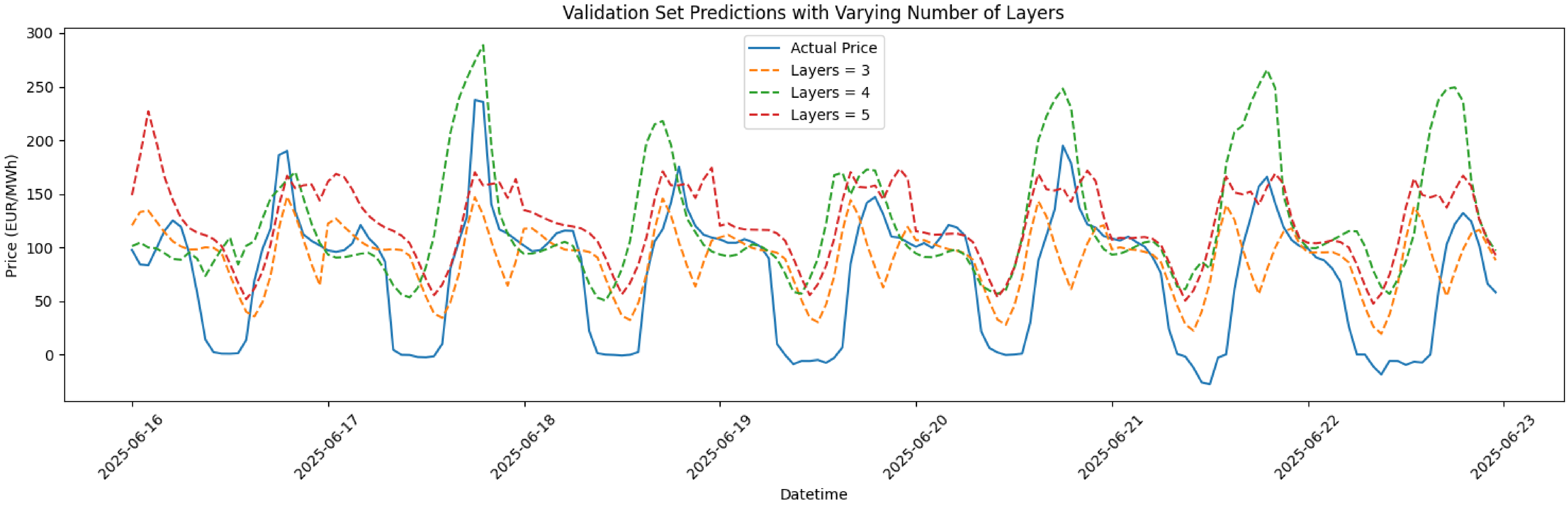
- Summary Injection
  - Expand the encoder state to [B,Tfore,dmodel][B,Tfore ,dmodel] and **concatenate** with forecast features [B,Tfore,nfore][B,Tfore ,nfore], yielding [B,Tfore,nfore+dmodel][B,Tfore ,nfore +dmodel].
- Linear Projection
  - Map to [B,Tfore,dmodel][B,Tfore ,dmodel].
- Stacked Mamba Blocks (× nlayersnlayers)
  - Same Mamba + Dropout + LayerNorm + residual pattern as the encoder.
- Output Head
  - Final linear layer from [B,Tfore,dmodel][B,Tfore ,dmodel] to [B,Tfore,1][B,Tfore ,1], producing the hourly energy price forecast.



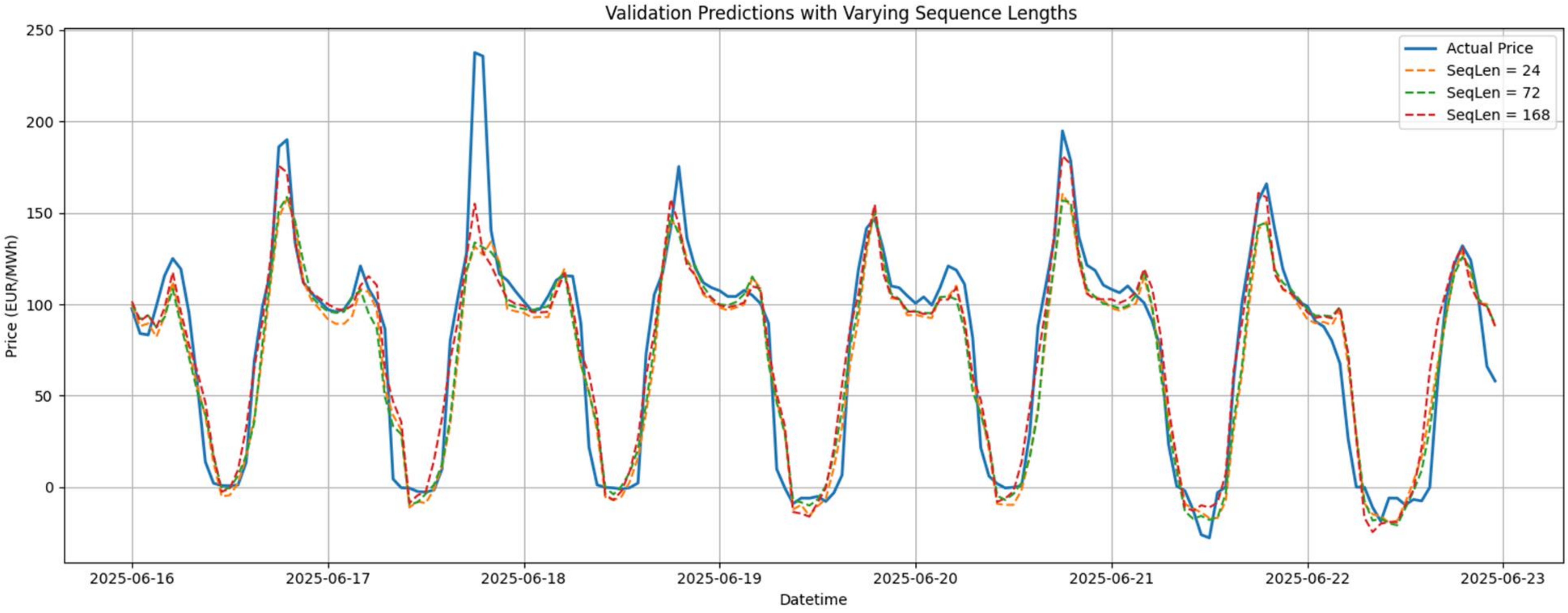
Mamba – Dropout Rate



Mamba – Number of Layers



Mamba – Sequence Length



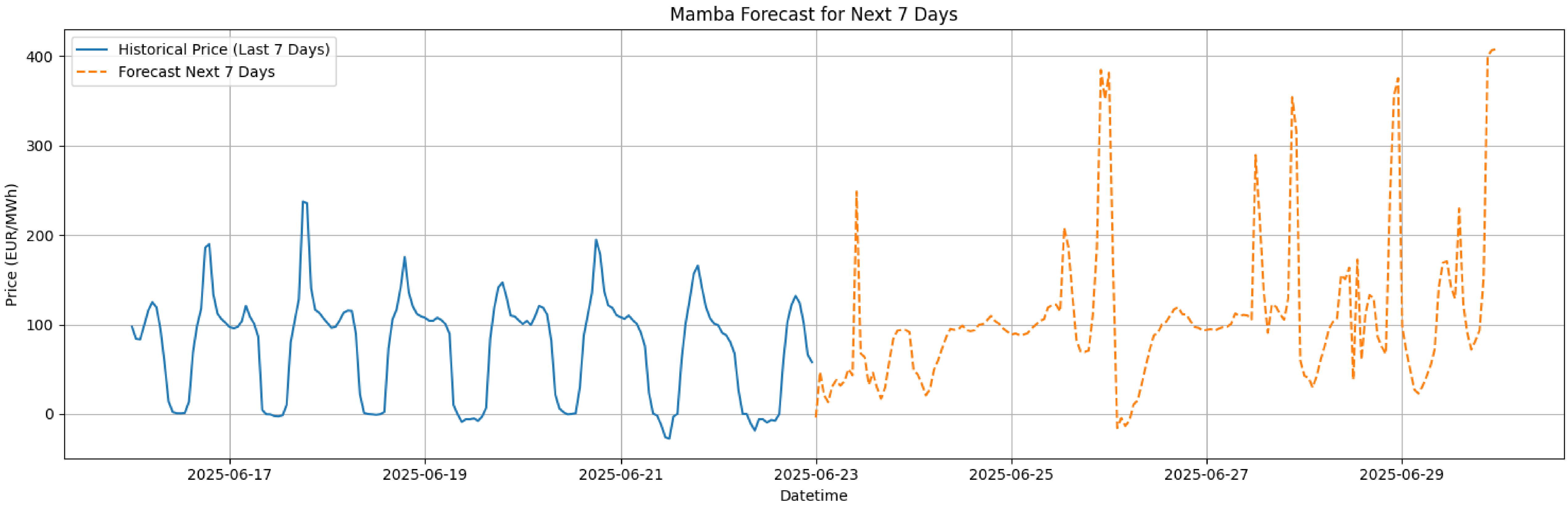
Mamba – Hyperparameter Results

	Drop out	Sequence length	Number of layers	Evaluation Metrics			
				MAE	MSE	SMAPE	Accuracy
Drop out	0.1	24	3	23.92	989.03	62.94%	91.67%
	0.15	24	3	41.78	2870.68	81.31%	72.02%
	0.2	24	3	39.14	2855.42	80.74%	76.19%
Sequence length	0.1	24	3	30.49	1881.77	69.34%	83.93%
		72	3	21.85	830.15	61.40%	96.43%
		168	3	32.56	1714.57	72.32%	83.33%
Num layers	0.1	24	3	39.74	2696.02	76.73%	75.60%
			4	54.24	5260.57	77.98%	60.71%
			5	51.51	4165.23	78.38%	61.90%

Mamba Results

	Short-term (0-24h)	Medium-term (24-72h)	Long-term (72-168h)
MAE	23.59	20.77	21.95
MSE	784.97	965.13	773.96
SMAPE	57.29%	54.45%	45.94%
Accuracy	100.00%	95.83%	95.83%

Mamba – Week Forecast



Model comparison

Model	Hyperparameters			Evaluation Metrics			
	Drop out	Sequence length	Number of layers	MAE	MSE	SMAPE	Accuracy
DSSM	0.15	168	2	13.32	373.24	51.28%	97.50%
S4D	0.2	72	9	16.39	466.67	56.19%	97.50%
Mamba	0.15	72	3	21.85	830.15	61.40%	96.43%
LR	n/a	n/a	n/a	48.74	4395.58	60.03%	69.52%

Discussion

- All models performed better than baseline (linear regression)
- Models are not great at predicting price volatility; high SMAPE values (51.28%-60.03%)
  - Model was trained on MAE loss; focused on averages and not extremes
  - Use probabilistic forecasting (Gaussian likelihood)
  - Add more volatility features (momentum indicators)
- Diminishing returns from layers, more layers didn't necessarily improve performance but increased training times
- Models get worse the further they are from the current day
- DSSM outperformed all models
- Mamba model did not include temporal features
  - Could not learn patterns within time intervals as effectively when inferring data
- Possible improvements
  - More specific temporal features (seasonal)
  - Dataset larger than 3 years for deeper training

Q&A

References

- https://github.com/gnovack/easy-ssm
- **Gu, A., Johnson, I., Goel, K., Saab, K., Dao, T., Rudra, A., & Ré, C. (2021).** Linear State-Space Layers: Combining Recurrent, Convolutional, and Continuous-time Models with Linear State Space Layers. *arXiv preprint arXiv:2110.13985*. https://arxiv.org/abs/2110.13985
- **Gu, A., Goel, K., Gupta, A., & Ré, C. (2022).** On the parameterization and initialization of diagonal state space models. *arXiv preprint arXiv:2206.11893*. https://arxiv.org/abs/2206.11893
- **Gu, A., & Dao, T. (2023).** Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*. https://arxiv.org/abs/2312.00752