Method

This section outlines the different methods used in the project and includes descriptive statistics about the data. The first section describes the forecasting methods used, namely SARIMA, LSTM, and TCN. Following the forecasting models evaluation metrics are presented. An overview of the data used in the project is also available in this section.

Evaluation metrics

To evaluate the performance of the forecasting models, the models will be evaluated by forecasting the price given a series of data inputs and then comparing the output with the true value at that time. All models will be evaluated with the same metrics.

Correlation analysis

Pearson's correlation coefficient is used to determine the relative correlation of two dependent variables. The covariance of two variables is divided by the product of their standard deviations to get a percentage correlation between the two variables.

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \tag{8}$$

$\rho_{X,Y}$ is always a value between -1 and 1 per definition. 1 equals perfect correlation and -1 perfect negative correlation.

Pearson correlation coefficients measures the correlation between two variables. To measure the correlation for the same time series between different points in time auto correlation functions can be used. However, the main focus for the correlation analysis in this thesis is to access the correlation between the electricity price and different fundamental variables.

Mean square error (MSE)

A very common measure in statistics and regression analysis of predictive quality is the Mean Squared Error (MSE). MSE calculates the average squared predictive error from n samples. As the error is squared the relative difference at all times is considered, i.e, over and under predicting still gives a poor MSE value.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(A_t - F_t)^2 \tag{9}$$

MAPE and WAPE

One problem with MSE and RMSE is the scale dependence, the errors are relative to the scale that they are compared against. To mitigate this issue

Mean Average Percentage Error (MAPE) measures the percentage error. Let A_t and F_t be the actual value and forecasted value, respectively. Then the MAPE metric is calculated as the absolute value of the percentage difference between the forecast and the actual value:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}|\frac{A_t - F_t}{A_t}| \tag{10}$$

MAPE gives the same weight to all forecasts, regardless of the value of the forecasts. This can give a disproportional high weight to forecasts with low or close to zero forecasts. Alternatively, the percentage average of all errors normalized by the sum of all forecasts gives the WAPE estimate.

$$WAPE = \frac{\sum_{t=1}^{n}|A_t - F_t|}{\sum_{t=1}^{n}|A_t|} \tag{11}$$

WAPE is the main evaluation metric that will be used in this project due to its scale-invariant properties and easy interpretation.

Loss function

The goal of a loss function in regression problems is to fit the output outputs as close as possible to input correct values. MSE, as defined in equation (9), is the loss function used to train all models. MSE as a loss function is commonly used for financial time series as it works best for normally distributed data sets. However, for loss function risks giving to high weight to outliers due to the squaring property of the function.

Models

In total three models are created, two machine learning models and one mathematical model. The hyper parameters are shown in Table 2.

The hyper parameters are selected by a combination of sensitivity analysis performed and guidance from other experiments namely, [9], [7] and [5]. Hyperparameter selection for large machine learning models is highly complex because of the number of possible combinations. Sensitivity analysis is therefore conducted on selected parameters and considering the best outcome given training time, predicting accuracy and loss functions. The stating point for sensitivity analysis is the parameters presented in [9]. Experiments on hyper parameters are computationally expensive witch limits the possible combinations tested. Predictive steps and past history, used for the prediction, are the same for all models for both easier comparison and forecasting needs for research questions. The batch sizes are the same both for the LSTM and TCN models, using a relatively low size given the data set. The number of epochs is set at 75 although this number is never reached in practice as the early stopping prevents the models from reaching new epochs if no significant improvements are made.

Table 2: Hyper parameters used for the models

Model	Parameter	Variable used	Explanation
All	Past History	50	The number of inputs used for predictions
	Predictive steps	24	The number of forecast time steps
TCN	Blocks	2	Number of blocks in the in the TCN models. High number of blocks are useful for very large and complex input data.
	Filters	4	Number of convolutions in the network.
	Kernels	16	The number of historical data points used in the convolutions
	Dilations	[1, 2, ..., 64]	Depth of one TCN layer, always a multiple of 2
	Batch size	8	Number of training iterations to next iterations of the network
	Epochs	75	Number of allowed iterations until the best iterations of the network is selected.
LSTM	Hidden layers	2	Number of hidden layers
	Size hidden layers	64	Number of nodes per hidden layer
	Batch size	8	
	Epochs	75	
ARIMA	p	5	Number of lags used for future prediction
	q	1	Order of the moving average model
	d	1	Degree of differentiating

Data

Financial contracts are only traded on non-holidays and only from 09:00-16:00. Historical data for future price is thus only available on days when the market was open. Whereas spot prices are available with greater resolution, all hours of the day, every day. To have a simple comparison between the financial contracts and the spot prices, the average spot price for days when the market was open is used. This results in time series of length 260-262 for each year.

Table 3: Data sources for historical electricity prices

Period of interest	**Source**	**Type**	**Period**	**Price signal**
Yearly contracts	Bloomberg	Futures contracts	Year ahead	Closing price
Quarterly contracts	Bloomberg	Futures contracts	Quarter ahead	Closing price
Monthly contracts	Bloomberg	Futures contracts	Intra quarter	Closing price
Spot prices	Nord pool	Historical data	Daily	Average

The final input data to the models is a combination of financial times series data and fundamental data. The input data used are future and spot electricity prices, gas futures, temperature and wind speed and hydro reservoir levels.

Historical weather data

Weather data is collected from [18]. The data consist of daily measurements from two weather stations in SE1 and SE3. In SE3, Bromma flygplats is used as a reference point for weather data measurements and Umeå flygplats is used as a reference for the weather in SE1. Data from January 2015 to March 2022 is used. Similar to the spot prices, only daily average data points are used.

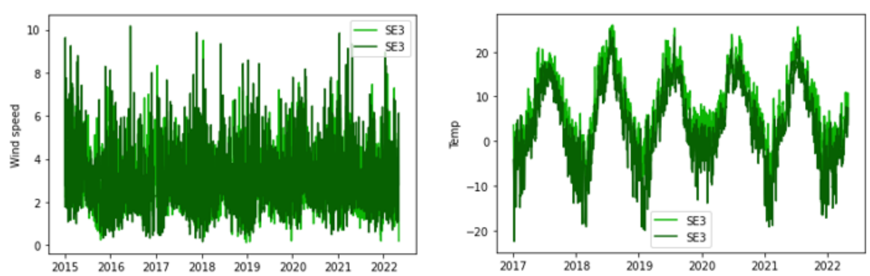

Figure 10: Measured temperature and weather in SE1 and SE3. Wind speed in m/s and temperature

Figure 10 illustrate the observed wind speed and temperature for SE1 and SE3 respectively. The observed measurements follow a clear seasonal pattern as expected with lower temperatures in the winter and higher in the summer. Similarly, the wind speed also has a seasonal component but with higher volatility.

Hydro reservoir levels

Hydro reservoir data is collected form [19]. The stored energy in water reservoirs over time for Norway and Sweden is show in Figure 11. Norway and Sweden's energy systems are dominated by hydropower and have the largest storage and production capabilities in the Nordics. A strong seasonal component can be

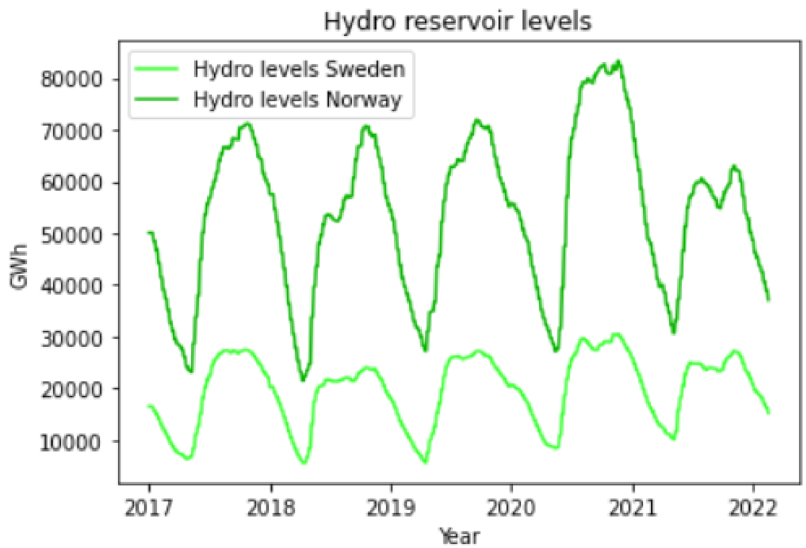

Figure 11: Historical hydro reservoirs data for Sweden and Norway in GWh

observed for both time series where the lowest values for every year are observed in the spring when the snow starts melting. The stored energy peaked in 2020 from the decreased demand during COVID-19. The lowest peaks are found in late 2021 when the energy stored did not reach the observed historical normal peaks.

Financial data

In Table 4 the average traded price for system contracts can be seen for seven short-term tenors. The lowest values for all tenors are observed in 2020 at 12.78 EUR/MWh for the 1-month contract and the highest values are observed in 2022 for the one-month contract.

Table 4: Average traded prices for SYS contracts 2017 to 2021 in EUR/MWh

SYS	**1M**	**2M**	**3M**	**1Q**	**2Q**	**3Q**	**4Q**
2017	29.25	28.26	27.89	27.78	27.06	25.84	25.61
2018	43.73	43.64	42.33	41.54	38.80	36.06	31.77
2019	39.47	40.08	40.50	38.08	38.25	37.74	36.46
2020	12.78	14.83	16.88	17.22	19.92	22.20	23.39
2021	61.84	58.61	54.76	51.87	41.86	33.64	31.31
2022	104.41	95.15	75.79	61.41	41.86	58.88	67.10

Similar to Table 4, Table 5 shows average traded price for gas futures from 2017 to 2022. The lowest observed average is 9.78 EUR MWh in 2020 and 96.83

Table 5: Average traded prices for Gas contracts 2017 to 2022 in EUR/MWh

Gas futures	**1M**	**2M**	**3M**	**1Q**	**2Q**	**3Q**	**4Q**
2017	17.48	17.50	17.48	17.47	17.28	17.26	17.35
2018	22.35	22.32	22.28	22.28	21.80	21.24	20.78
2019	15.07	15.96	16.72	16.64	18.31	18.86	19.11
2020	9.78	10.31	10.85	10.90	12.23	13.24	13.79
2021	47.54	47.54	46.70	46.74	36.61	31.71	30.21
2022	96.83	95.19	91.07	90.85	88.86	87.65	82.81

EUR/MWh in 2022.

Portfolio approach

Two portfolio approaches are evaluated in this project. One standard approach without any computational intelligence for decision making and one model based on machine learning decision making. Both models are based on a layered hedging approach with a portfolio of futures contracts. The goal of the portfolio approach is to secure 100% of forecast electricity consumption before every month. This is achieved by securing at least 25% of forecast exposure 2 years in the future, 50% the coming year, and 75% one calendar quarter before. The standard approach model assumes electricity purchase of equal weight until the desired hedging levels are reached. To achieve this, a small proportion of the electricity for the desired electricity for a delivery period is secured every trading day. The machine learning approach has the same minimum hedging bands but with a decision layer, if the price of the contract is forecasted to be lower tomorrow then do not trade the contract. The machine learning approach has the same weight as the standard portfolio. Thus the standard approach will reach the desired hedging strategy quicker than the machine learning approach.

To simulate the outcome of the hedging strategies a historical simulation is performed. The machine learning models are trained with historical data up until the year the portfolio is simulated on. The models are retrained every simulated year with historical data. This simulation scheme avoids evaluating on already seen data. One model is created for every simulated year and future contract, thus every simulated year 4 models are created for 1M, 1Q, 1Y, and 2Y contracts.

Test and validation data

The data is split into different buckets for training, validation, and testing. The split between training and testing is set to 70%. Furthermore, this project implements a roiling window approach to first train and validate the model and then test the performance given a specific unseen time period. This training approach is similar to the one used in [20]. After the testing period, the last of the training/ validation data set is incremented, extending past the testing period. The benefit of this approach is that it allows for the evaluation of the model for different time periods.

The input data is available from 2017 to March 2022, during this time three significant events have affected the electricity market. First the COVID 19 impact, then the gas shortages during the fall of 2021, and lastly the Russian invasion of Ukraine and the volatile period before the invasion. Four periods are used to evaluate the performance of the forecasting models. These periods are chosen to evaluate the performance during periods with normal conditions and during times of high volatility and uncertainty. Furthermore, the periods are chosen at regular intervals to reduce the effect of seasonal components but still give the model at least 2 years of training and validation data.

- **2019** 01/01/2019 - 01/06/2019 Normal economic activity in the area. This can be used as the baseline period.

- **2020** 01/01/2020 - 01/06/2020 COVID-19 impact with lower economic activity and subsequently lower spot and future prices. High uncertainty for all commodities markets resulting in low electricity and gas futures.
- **2021** 01/01/2021 - 01/06/2021 Relatively normal period with much of the economic activity is back from 2020 lows.
- **2022** 01/01/2022 - 08/03/2022 Impact from European gas shortages during late 2021 and early 2022. Invasion of Ukraine from 24/2. The period is ending earlier as this project is done during spring 2022.

Preprocessing

The future price in a specific pricing area is composed of two separate contracts, the SYS contract, and the EPAD contract. The SYS and EPAD contracts are added to get the final price for the bidding areas.

Financial contracts are only traded on non-holidays and only from 09:00-16:00. In contracts, the spot price is available at all times with hourly resolution. Daily average spot prices on the same days that the market is open is therefore used when comparing the data. All data is normalized between 0 and 1 before being fed to the neural networks.

Implementation and hardware

Through this project, Python is used for all data analysis and model implementation. The TensorFlow package is used for the machine learning models, combined with NumPy and pandas for prepossessing and data management. Keras [21] is used for the implementation of the models. The TCN model is implemented using the keras TCN model [17] and the LSTM model using LSTM layers. The data is trained on a 16GB system with a 3060TI GPU.