Feature Selection on the Nord Pool power market

Lukas Andersson

Master's thesis
2023:E72

Lund University

Faculty of Engineering
Centre for Mathematical Sciences
Mathematical Statistics

Centrum Scientiarum Mathematicarum

Abstract

In this thesis the four electricity bidding areas of Sweden are investigated. The aim of the work is to model the regions separately and perform feature selection to find unique characteristics of their energy markets as well as shared commonalities. To do this, a Sparse Jump Model was used. In the thesis four different data sets were used, each containing within-region energy-related features of a specific bidding area as well as foreign energy features from Denmark and Norway. For each data set the model was used to select between one and six features to most accurately cluster the data and estimate a three-state sequence, resulting in a total of 24 different model iterations.

The results from the Sparse Jump Model shows both similarities and differences between the four regions. Three features were selected in all four bidding areas; the template based energy delivery measure Schablon, the within-region temperature and the electricity load of bidding area DK2. In all bidding areas one within-region energy production feature was selected, but they differed both in terms of selection order and feature characteristics. The most interesting part of the result is the preference of the electricity load of bidding area DK2 over the within-region consumption, which was only selected as a feature in regions SE2 and SE3 and in those regions selected after the electricity load.

Contents

1 Introduction

1.1 Background

Electricity is an important commodity with unique properties. Compared to physical commodities there is no efficient way to store electricity, and you need an established transmission network to transport it. The electricity spot price also possesses an uniquely high volatility quickly moving between extreme highs and lows price-wise. Furthermore, unlike most other commodities a clear seasonal trend can be found in the electricity price on a daily, weekly and yearly level. Other characteristics of the electricity spot price is that it is mean reverting meaning it fluctuates around some normal price which it will return to even if it deviates from it for a period of time due to disturbances, and the inverse leverage effect can be observed, meaning that an increase in spot price leads to an increase in volatility (Escribano et al., 2011, Knittel and Roberts, 2005). As electricity is an integral part of day-to-day life in modern society it is of utmost importance to accurately plan the energy production to meet consumption demands.

Following complaints regarding alleged energy export abuse (de Hautecloque and Hancher, 2011), the Swedish energy market was split into four different bidding areas in 2011 with separate electricity prices; Luleå SE1, Sundsvall SE2, Stockholm SE3 and Malmö SE4, shown in fig. 1. In bidding areas SE1 and SE2 energy production exceeds consumption leading to a surplus while there in areas SE3 and SE4 generally is an energy deficit, leading to price discrepancies between northern and southern Sweden. Limitations of the transmission network can further exacerbate this problem, forcing southern Sweden to import energy from Denmark and continental Europe at times. In Sweden, more than half of the produced energy comes from renewable sources (Regeringen, 2019). While this is great from an environmental perspective, becoming more reliant on renewable energy has its' drawbacks. One of these drawbacks is the output inconsistency of due to the reliance on weather conditions to produce electricity. This inconsistency complicates forecasting and production planning and can lead to increased volatility in electricity prices (Green and Vasilakos, 2010, Lindström et al., 2015).

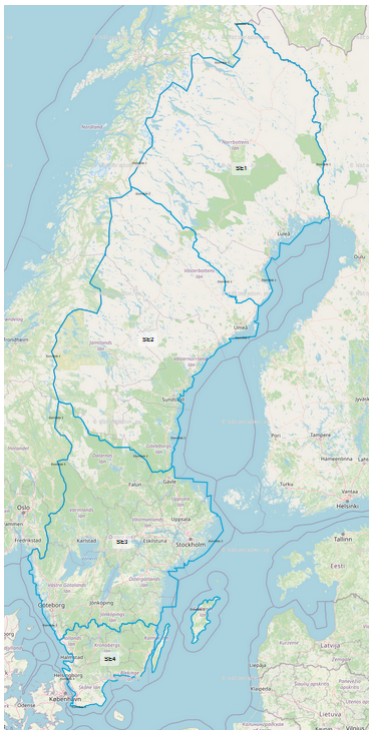


Figure 1: The four electricity price bidding areas of Sweden. Source Svenska Kraftnät.

When modeling the electricity spot price one of the more commonly used model types are regime-switching models. These type of models consists of two or more submodels which describes different states or regimes of the time series and some form of logic deciding which regime the process is at a specific point in time. In Markov regime-switching models (MRS) this logic is a Markov Chain meaning that future values are solely dependent on the current value. Previous studies have shown that MRS works well for capturing the typical characteristics of the spot price such as seasonal spike intensity and the inverse leverage effect (Janczura and Weron, 2010).

Model state estimation can oftentimes be improved with feature selection, the process of determining which variables has the most impact on the model results. It is commonly used in both machine learning and statistical methods to reduce the number of input variables by selecting those with the highest model impacts. An efficient feature selection framework by clustering is proposed by Witten and Tibshirani, 2010. It works by alternating between clustering a weighted version of the data and selecting feature weights.

In this thesis, a Sparse Jump Model (SJM) is used to investigate the four bidding areas of Sweden. Just like with MRS the SJM split the time series into different regimes or states, but in the SJM they are represented by clusters. The model translates the framework in Witten and Tibshirani, 2010 to jump models, thus incorporating the temporal ordering of the observations into a model with efficient feature selection. The model has been shown to be robust to noise, able to distinguish relevant features from irrelevant ones and it outperforms other models such as the standard Jump model, Sparse K-means clustering and PCAHMM in many scenarios (Nystrup et al., 2021).

1.2 Purpose

The purpose of this thesis is to investigate and compare the four bidding areas of Sweden using the Sparse Jump Model. Focus will be on the feature selection process and similarities and differences between the regions in terms of what features that were selected, selection order and resulting feature weights. The results are also compared to previous studies on the energy market and lastly the estimated state sequence will be briefly covered.

1.3 Outline

The structure of this thesis is as follows:

Chapter 2 - Statistical methods and theory related to clustering and Jump models are presented.

Chapter 3 - The dataset used in this thesis is presented along with data sources and transformation methods.

Chapter 4 - In this chapter, model hyperparameter selection and model for the different regions are presented.

Chapter 5 - This chapter contains the analysis, comparison and discussion of the results from chapter 4.

Chapter 6 - The final chapter contains a summary of the report and conclusions drawn.

2 Theory

This chapter presents the theory behind the Sparse Jump Model. First, some statistical methods used in the model are explained. This is followed by brief overviews of the models which the Sparse jump model is built upon and finally a presentation of the model itself and a summarization of Independent Spike Models.

2.1 Statistical methods and concepts

2.1.1 Information criterion

The use of an information criterion is a way to measure the quality of a statistical model. It measures the complexity of the model as well as how well the model fits the data and is defined as

$$IC = F + a_T M \tag{2.1}$$

where F is the model fit, M is the model complexity and a_T is a positive sequence dependant on the sample size T and possibly other factors such as number of parameters or features. Generally speaking, a model with a lower information criterion value is considered superior to one with a higher value, thus one can optimize model hyperparameter values through minimizing the information criterion.

Two frequently used information criteria are the Akaike Information Criterion (AIC) and the Bayesian information Criterion (BIC), defined as

$$\begin{aligned} AIC &= -2\ell(\hat{\theta}) + 2q \\ BIC &= -2\ell(\hat{\theta}) + q\log(T) \end{aligned} \tag{2.2}$$

where $\ell(\hat{\theta})$ is the maximized log-likelihood function value measuring model fit F, q is number of model parameters measuring complexity M with $a_T = 2$ for AIC and $a_T = \log(T)$ for BIC, differentiating the the relative weighting of model fit and complexity of the two criteria.

2.1.2 Correlation

Correlation or dependence is a relationship between two random variables. In statistics it most often refers to the Pearson product-moment correlation coefficient ρ, which establishes a linear relationship between these two variables. This coefficient can take values in the interval $[-1, 1]$ where a larger absolute coefficient value indicates a stronger linear relationship between the two variables. For random variables X, Y with means μ_X, μ_Y and standard deviations σ_X, σ_Y the Pearson product-moment correlation coefficient is defined as

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E\left[(X - \mu_x)\left(Y - \mu_Y\right)\right]}{\sigma_X \sigma_Y} \tag{2.3}$$

With a series of observations this correlation coefficient can be estimated by the sample correlation coefficient r. With n observations of variables X, Y with sample means $\bar{\mu}_x$, $\bar{\mu}_y$ and corrected sample standard deviations $\bar{\sigma}_x$, $\bar{\sigma}_y$ the sample correlation coefficient becomes

$$r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{\mu}_x)(y_i - \bar{\mu}_y)}{(n-1)\,\bar{\sigma}_x \bar{\sigma}_y} = \frac{\sum_{i=1}^{n} (x_i - \bar{\mu}_x)(y_i - \bar{\mu}_y)}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{\mu}_x)^2 \sum_{i=1}^{n} (y_i - \bar{\mu}_y)^2}}$$

(2.4)

As with ρ, the sample correlation coefficient can also take values in the interval $[-1, 1]$, where larger absolute values means stronger correlations of the variables. Heuristically, $|r| \leq 0.3$ means there's either a very weak relationship between the variables or none at all, $0.3 < |r| \leq 0.5$ means there's a weak relationship between the two, $0.5 < |r| \leq 0.7$ indicates a moderate linear relationship and for $|r| \geq 0.7$ there's a strong relationship.

A common way to visualise the correlation is through the use of a correlation matrix. For p variables the correlation matrix C has shape $p \times p$ where $c_{i,j} = r_{ij}$. Since $r_{xx} = 1$ the diagonal entries of the matrix are all equal to 1 and since $r_{ij} = r_{ji}$ it is also symmetrical. Due to these properties both the diagonal values and the values of either the upper or lower triangle of the matrix can be omitted without any loss of information. This can be seen in fig. 2 where the diagonal values have been replaced by variable names and the lower triangle shows a heatmap of the correlations between these variables.

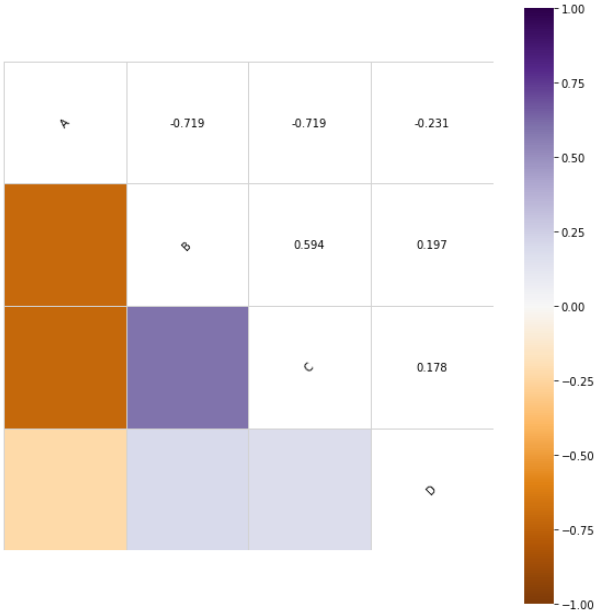


Figure 2: Correlation matrix of variables A, B, C and D. Upper triangle indices shows sample correlation r between the different variables. Diagonal values have been replaced with variable names and lower triangle values have been replaced with a heatmap of sample correlation values.

2.1.3 Coordinate descent

Coordinate descent is an optimization algorithm commonly used in machine learning and data analysis which can be used to find the minimum of a multivariate function (Wright, 2015). It does so by iterating through all function variables considering one at a time holding all other variables fixed, transforming the multivariate optimization problem of minimizing function $f(\mathbf{x})$ into several univariate optimization problems of minimizing $f(x_i)$ for $i = 1, ..., n$. These univariate problems are generally much easier to solve. While there are many different variations of coordinate descent (see Wright, 2015) the algorithm in its' most basic form is shown below.

Algorithm 1 : Coordinate descent

1. **Input:** Initial parameter vector $\mathbf{x} = (x_1, ..., x_n)$, maximum number of iterations K, tolerance tol
   (a) for $k = 1, ..., K$ or until $f(\mathbf{x}^k) - f(\mathbf{x}^{k-1}) \leq$ tol
      i. for $i = 1, ..., n$
         A. choose step size α
         B. set $x_i^k = x_i^{k-1} - \alpha \frac{\partial f}{\partial x_i}$
2. **Output:** Updated parameter vector $\mathbf{x} = (x_1, ..., x_n)$ minimizing function $f(\mathbf{x})$

There are various ways of choosing the step size α. Two common ways are using line search criteria or solving for the minimizing value of $f(x_i)$ with all other variables fixed.

2.1.4 Soft thresholding

The soft thresholding operator is the proximal mapping of the L1-norm and can oftentimes be part of the solution to different convex optimization problems. It is commonly used for wavelet denoising (Donoho, 1995) and in feature selection algorithms (Witten and Tibshirani, 2010). To derive it, some definitions and a theorem are presented below.

Definition (Subgradient) A vector $v \in \mathbb{R}$ is a subgradient of function $f : \mathbb{R}^n \to \mathbb{R}$ at $x \in \text{dom}(f)$ if

$$f(y) \geq f(x) + v^T(y - x) \quad \forall\ y \in \text{dom}(f)$$

Definition (Subdifferential) The subdifferential $\partial f(x)$ of f at point x is the set of all subgradients v

$$\partial f(x) = \left\{ v | f(y) \geq f(x) + v^T(y - x) \quad \forall\ y \in \text{dom}(f) \right\}$$

Definition (Proximal operator) The proximal operator is an operator for proper lower semi-continuous convex function f from a Hilbert space χ to $[-\infty, \infty]$ defined as

$$\text{prox}_f(x) = \text{argmin}_{y \in \chi} \left\{ \frac{1}{2} \|y - x\|_\chi^2 + f(x) \right\}$$

Theorem (Subgradient Optimality) A point x^* is a minimizer of function f iff f is subdifferentiable at point x^* and $v = 0$ is a subgradient of f at point x^*.

$$f(x^*) = \min_x f(x) \quad \Leftrightarrow \quad 0 \in \partial f(x^*)$$

Consider the function $f(x) = \lambda||x||_1$. The proximal operator of $f(x) = \lambda\|x\|_1$ is given by

$$\text{prox}_f(x) = \text{argmin}_y \left\{ \frac{1}{2} \|y - x\|_2^2 + \lambda \|y\|_1 \right\}$$

The problem is separable with respect to both y and x and can be rewritten as

$$\text{argmin}_y \left\{ \frac{1}{2}(y - x)^2 + \lambda|y| \right\}$$

Applying the subgradient optimality condition to the rewritten proximal operator we get

$$\begin{aligned} 0 &\in \partial f(x) \\ &= \partial(\frac{1}{2} \|y - x\|_2^2) + \partial(\lambda \|y\|_1) \\ &= y - x + \partial(\lambda \|y\|_1) \\ &= y - x + \lambda \,\text{sgn}(y) \end{aligned}$$

where the subdifferential of $|y|$ is the sign function sgn() defined as

$$\begin{aligned} &= -1, && x < 0 \\ \text{sgn}(x) &= [-1, 1], && x = 0 \\ &= 1, && x > 0 \end{aligned}$$

The optimum y^* is then obtained at

$$0 = y^* - x + \lambda \,\text{sgn}(y^*) \Leftrightarrow y^* = x - \lambda \,\text{sgn}(y^*)$$

With this we get three possible situations. For $y^* > 0$ we get $y^* = x - \lambda$ which holds for $x > \lambda$. If $y^* < 0$ we get $y^* = x^* + \lambda$ which holds for $x < -\lambda$. From these two situations it follows that $\|x\| > \lambda$ and $\text{sgn}(x) = \text{sgn}(y^*)$. For the final case of $y^* = 0$ we get

$$\begin{aligned} 0 \in \;& x - \lambda[-1, 1] \Leftrightarrow \\ x \in \;& [-\lambda, \lambda] \Leftrightarrow \\ |x| \leq \;& \lambda \end{aligned}$$

Summarizing this, for $f(x) = \lambda\|x\|_1$ we have

$$\text{prox}_f(x) = \text{argmin}_y \left\{ \frac{1}{2} \|y - x\|^2 + \lambda \|y\|_1 \right\} = y^* = \begin{cases} 0, & |x| \leq \lambda \\ x - \lambda \,\text{sgn}(x), & |x| > \lambda \end{cases}$$

The bracket equation on the rightmost side of the equation is the Soft Thresholding Operator on x with threshold λ. More concisely, the operator can be written as shown in eq. 2.5 with threshold Δ and the effect of applying the operator to a function is shown in fig. 3.

$$S(x, \Delta) = \text{sgn}(x) \cdot (|x| - \Delta)_+ \tag{2.5}$$

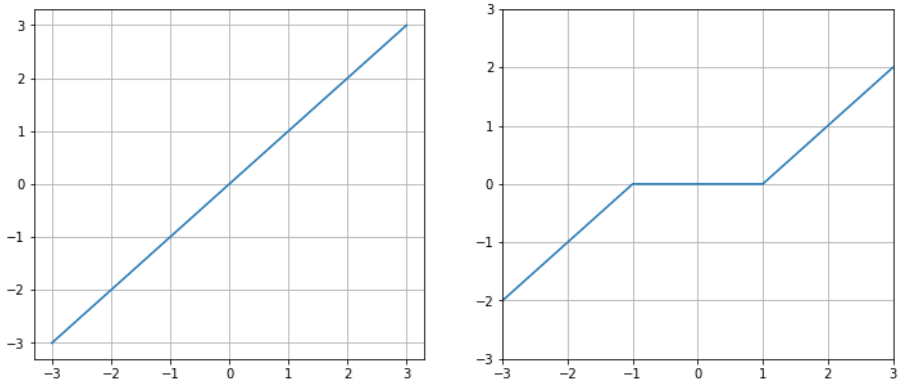


Figure 3: Illustration of the soft threshold operator. On the left side we have function f. On the right side we have the soft thresholding operator on function f with $\Delta = 1$.

2.1.5 Gap statistics

The gap statistic, proposed in *Estimating the number of clusters in a data set via the gap statistic* is a statistical method commonly used in cluster analysis to find the number of clusters in a given data set (Tibshirani et al., 2001). The method works by the steps shown in algorithm 2.

Algorithm 2 : Gap statistics

---

1. For each candidate number of clusters k = 1,...,K :
   (a) Cluster the given data set
   (b) Compute the within-dispersion measure $\boldsymbol{W_k}$, the within-cluster sum of squares
   (c) Generate $\boldsymbol{B}$ reference data sets, most easily by permuting the data set within each feature
   (d) Compute $\boldsymbol{W_{kb}}$ for b = 1,...,B
   (e) Compute the gap statistic $Gap(k) = \log(W_k) - \frac{1}{B}\sum_b \log(W_{kb})$
   (f) compute the standard deviation $sd_k = [\frac{1}{B}\sum_b \left(\log(W_{kb}) - \frac{1}{B}\sum_b \log(W_{kb})\right)^2]^{\frac{1}{2}}$
   (g) compute 1-standard error $s_k = sd_k\sqrt{1+\frac{1}{B}}$
2. choose smallest k such that $Gap(k) \geqslant Gap(k+1) - s_{k+1}$

2.2 Models

This section briefly summarizes models relevant to this thesis. The first three subsections explains the models upon which the Sparse Jump Model was built upon, and section 2.2.4 presents the model used in this thesis. The final section summarizes Independent Spike Models since the results from these models are used for comparison later on.

2.2.1 K-means

K-means clustering is a method that seeks to divide a large set of observations into different clusters (Lloyd, 1982). Initialized with $\boldsymbol{K}$ clusters $\boldsymbol{C_k}$ with respective cluster centroids $\boldsymbol{\mu_{C_k}}$, the algorithm alternates between assigning all observations to the cluster with the closest mean and updating the cluster centroids $\boldsymbol{\mu_{C_k}}$ to the mean of the sum of all observations $\boldsymbol{n_k}$ within the cluster $\boldsymbol{C_k}$. The method reaches convergence through minimizing the *within-cluster sum of squares* (WCSS), meaning given observations $\boldsymbol{Y}$ of $\boldsymbol{P}$ parameters it seeks to minimize the objective function

$$\sum_{k=1}^{K} \frac{1}{n_k} \sum_{y_{i,j,p} \in C_k} (y_{i,j,p} - \mu_{C_k})^2 \tag{2.6}$$

Similarly, instead of minimizing the WCSS one could choose to maximize the *between-cluster sum of squares* (BCSS) defined as the difference between the total sum of squares (TSS) and the WCSS

$$\sum_{p=1}^{P} \left(\frac{1}{n} \sum_{i,j} (y_{i,j,p} - \mu)^2 - \sum_{k=1}^{K} \frac{1}{n_k} \sum_{y_{i,j \in C_k}} (y_{i,j,p} - \mu_{C_k})^2 \right) \tag{2.7}$$

2.2.2 Sparse K-means

The sparse K-means method, first proposed by Witten and Tibshirani, 2010 is an extension of the regular K-means method. By introducing parameter-specific weights with constraints to the objective function it yields a form of feature selection. The feature weights shows which features are most important to the clustering of the data. Introducing the feature weights, the objective function to maximize becomes

$$\sum_{p=1}^{P} w_p \left(\frac{1}{n} \sum_{i,j} (y_{i,j,p} - \mu)^2 - \sum_{k=1}^{K} \frac{1}{n_k} \sum_{y_{i,j \in C_k}} (y_{i,j,p} - \mu_{C_k})^2 \right) \tag{2.8}$$

$$\|\mathbf{w}\|^2 \leq 1, \quad \|\mathbf{w}\|_1 \leq \kappa, \quad w_p \geq 0 \ \forall p$$

The tuning parameter $\boldsymbol{\kappa}$ governs the sparsity of the feature weights. If too large, no feature selection will occur. Witten and Tibshirani suggests modifying the Gap statistic, a method used to estimate the number of clusters in a dataset, to instead estimate the tuning parameter

(Tibshirani et al., 2001). The modified algorithm is shown below.

Algorithm 3 : Modified gap statistic for the tuning parameter

Input: Dataset $\mathbf{Y}$ containing $\mathbf{T}$ observations of $\mathbf{P}$ features with corresponding feature weights $\mathbf{w}$:

1. Obtain permuted datasets $Y_1, ..., Y_B$, permuting within each feature specifically
2. For each candidate parameter value $1 \leq \kappa \leq \sqrt{P}$
   (a) Compute $O(\kappa) = \sum_{p=1}^{P} w_p \sum_{k=1}^{K} n_k(\mu_k - \bar{\mu})^2$
   (b) Compute $O_b(\kappa)$ on Y_b for $b = 1, ..., B$
   (c) Compute $Gap(\kappa) = log(O(\kappa)) - \frac{1}{B}\sum_{b=1}^{B} log(O_b(\kappa))$
3. Choose κ that yields largest $Gap(\kappa)$

Output: κ

2.2.3 Jump models

A jump model is a model which switches between a finite amount of submodels or states to fit a dataset while taking the temporal ordering of observations into account. With a dataset $\boldsymbol{Y}$ containing $\boldsymbol{T}$ observations of $\boldsymbol{P}$ features it is suggested in *Fitting jump models* to fit a K-state jump model by minimizing the objective function shown in equation 2.9 below (Bemporad et al., 2018).

$$\sum_{t=1}^{T-1} \left[l(y_t, \mu_{s_t}) + \lambda \mathbb{I}_{s_t \neq s_{t+1}} \right] + l(y_T, \mu_{s_T}) \tag{2.9}$$

Here, $\boldsymbol{l(y_t, \mu_{s_t})}$ is the loss function and $\mathbb{I}$ is the indicator function. λ is the regularization parameter, also called the jump penalty, which governs the models' propensity to change between different states or submodels in the state sequence. A larger regularization parameter value adds a larger fixed cost to the objective function resulting in longer state persistence. If the jump penalty is too small the state sequence tends to oscillate between two states and if chosen too large the jump model will result in a single-state model. Figure 4 illustrates this problem. The choice of λ comes down to balancing fitting the data and assumptions made about process state change frequency.

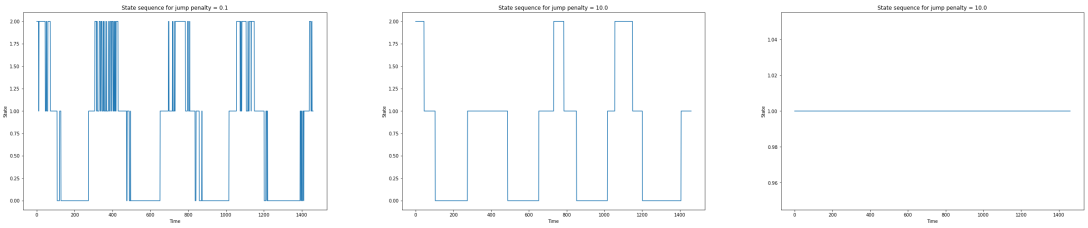


Figure 4: Estimated 3-state (state 0, 1, 2) state sequences for a data set with differing jump penalties, holding all other parameters equal. The leftmost figure with $\lambda = 0.1$ hints at a fairly obvious state sequence, but shows the oscillation around state changes caused by a too small jump penalty. The middle figure with $\lambda = 10$ shows a state sequence with similar shape to the leftmost one and through a better choice of regularization parameter the oscillation between states is prevented. The rightmost figure with $\lambda = 1000$ shows that too large a jump penalty fundamentally changes the state sequence shape, in this case reducing it to a single-state model.

Jump models can be fit by use of a coordinate descent algorithm alternating between finding the process state sequence that minimizes the objective function in eq. 2.9 and finding model parameters what minimize the loss function $\boldsymbol{l(y_t, \mu_{s_t})}$ (Bemporad et al., 2018).

2.2.4 Sparse jump models

The sparse jump model was first proposed by Nystrup et al., 2020. It extends the efficient feature selection framework from spare K-means clustering (2.2.2) to jump models (2.2.3). In this method, the obtained clusters represents different underlying process states of the observations.

Fitting the sparse jump model using BCSS and the squared Euclidean distance as a loss function the objective function to maximize becomes

$$\mathbf{w'} \sum_{k=1}^{K} n_k \left(\mu_k - \bar{\mu}\right)^2 - \lambda \sum_{t=1}^{T} \mathbb{I}_{s_t \neq s_{t+1}} \tag{2.10}$$

$$\|\mathbf{w}\|^2 \leq 1, \quad \|\mathbf{w}\|_1 \leq \kappa, \quad w_p \geq 0 \ \forall p$$

where $\boldsymbol{\mu_1, ...\mu_K}$, the model parameters, are the conditional means of the features assigned to state $\boldsymbol{k}$, $\boldsymbol{\bar{\mu}}$ is the unconditional mean of the observations, $\boldsymbol{n_k}$ are the number of observations in state $\boldsymbol{k}$, $\boldsymbol{w_1, ...w_P}$ are the feature weights, $\boldsymbol{s_1, ...s_T}$ is the state sequence, $\boldsymbol{\lambda}$ is the regularization parameter and $\boldsymbol{\kappa}$ is the tuning parameter. This objective function is solved using the coordinate descent algorithm shown in Algorithm 4, first holding weights fixed to fit model parameters and state sequence, then updating weights holding model parameters and state sequence fixed.

Algorithm 4: Sparse jump model fitting

Input: $\boldsymbol{Y}$ shape $(\boldsymbol{T}, \boldsymbol{P})$ data set of $\boldsymbol{T}$ observations of $\boldsymbol{P}$ standardized features, number of process states K, jump penalty $\boldsymbol{\lambda}$, feature limit $\boldsymbol{\kappa}$, maximum number of iterations and tolerance

1 Initialize feature weights $\mathbf{w}$ as $w_p^0 = 1/\sqrt{P}, p = 1, ..., P$

2 Iterate for $i = 1, ..., i_{\max}$ or $\left\|w^i - w^{i-1}\right\|_1 / \left\|w^{i-1}\right\| < \text{tol}$:

(a) Compute sequence of weighted features $z_t = y_t \cdot diag\left(\sqrt{w^{i-1}}\right), t = 1, ..T$

(b) Initialize state sequence $\mathbf{s} = s_1^0, ..., s_T^0$

(c) iterate for $j = 1, ..., j_{\max}$ or until $s^j = s^{j-1}$:

i. Fit model parameters $\mu^j = \text{argmin}_\mu \sum_{t=1}^T \left\|z_t - \mu_{s_t}^{j-1}\right\|^2$

ii. Fit state sequence $s_j = \text{argmin}_s \left\{ \left[\sum_{t=1}^{T-1} \left\|z_t - \mu_{s_t}^{j-1}\right\|^2 + \lambda \mathbb{I}_{s_t \neq s_{t+1}}\right] + \left\|z_T - \mu_{s_T}^{j}\right\|^2 \right\}$

(d) update weights w^i while holding parameters μ^j and state sequence s^j fixed by solving

$$\begin{aligned} &\text{maximize } w' \textstyle\sum_{k=1}^K n_k \left(\mu_k - \bar{\mu}\right)^2 - \lambda \sum_{t=1}^T \mathbb{I}_{s_t \neq s_{t+1}} \\ &\text{subject to } \|w\|^2 \leq 1, \quad \|w\|_1 \leq \kappa, \quad w_p \geq 0 \ \ \forall p \end{aligned}$$

Output: State sequence $s_1, ...s_T$ and feature weights $w_1, ..., w_P$

The feature selection in the model is done in step 2d of the algorithm through soft thresholding. The solution to the problem is

$$w = \frac{S(x_+, \Delta)}{\|S(x_+, \Delta)\|} \tag{2.11}$$

where $\boldsymbol{x}_+$ denotes the positive parts of a per-feature vector of BCSS. The threshold Δ is determined by binary search such that $\Delta = 0$ if that results in $\|w\|_1 \leq \kappa$; otherwise $\Delta \geq 0$ to yield $\|w\|_1 = \kappa$.

2.2.4.1 Selection of regularization and tuning parameter

The gap statistic used to determine tuning parameter value used in sparse K-means does not work for the sparse jump model as the feature-wise permutation of the data destroys the temporal information. In a similar fashion, the log-likelihood function cannot be used since it increases with model complexity. Instead, a General Information Criterion (GIC) is used.

In (Fan and Tang, 2013) a GIC for when number of features grows faster than number of observations is proposed and defined as

$$GIC = \frac{1}{T}\left\{2\left(\ell_T(Y, Y) - \ell_T(\hat{\theta}_{\mathcal{A}}, Y)\right) + a_T M\right\} \tag{2.12}$$

where $\ell_T(Y, Y)$ is the log-likelihood for the saturated model, $\ell_T(\hat{\theta}_{\mathcal{A}}, Y)$ is the log-likelihood for the model with the estimated active set of features $\mathcal{A}$. It is suggested setting $a_T = \log(\log(T))\log(P)$ and $M = |\mathcal{A}|$. The main differences between this criterion and the AIC and BIC is the inclusion of a saturated model for comparison and the measurement of number of active features.

This GIC is extended to Sparse Jump Models by Cortese et al., 2023. They propose replacing the log-likelihood function in eq. 2.12 with the WCSS and define the model complexity as

$$M = K_0|\mathcal{A}_0| + |\mathcal{A}_0|(K - K_0) + K_0(|\mathcal{A}| - |\mathcal{A}_0|) + \sum_{t=1}^{T} \mathbb{I}_{s_t \neq s_{t+1}} \tag{2.13}$$

where K is the number of latent model states, $|\mathcal{A}|$ is the active feature cardinality and $\mathbb{I}_{s_t \neq s_{t+1}}$ is the number of state changes in the state sequence. Using this model complexity and denoting the WCSS as $L_T(\lambda, \kappa, K; \boldsymbol{Y})$ we get the GIC proposed for the SJM as

$$GIC = \frac{1}{T}\left\{2\left(L_T\left(\bar{\lambda}, \bar{\kappa}, \bar{K}; \boldsymbol{Y}\right) - L_T\left(\lambda, \kappa, K; \boldsymbol{Y}\right)\right) + a_T M\right\} \tag{2.14}$$

where $\bar{\kappa}$, $\bar{\lambda}$ and $\bar{K}$ are the saturated model hyperparameters and a_T can be chosen as 2, $\log(T)$ or $\log(\log(T))\log(P)$ to get sparse jump model versions of AIC, BIC or the criterion of eq. 2.10. For the saturated model, setting $\bar{\kappa} = \sqrt{P}$ and $\bar{\lambda} = 0$ yields a jump model considering all features with no jump penalty. There is no obvious choice for the number of latent states of the saturated model $\bar{K}$, but Cortese et al., 2023, found that if the estimated model is to have recurrent states one should set $\bar{K} \leq 6$.

2.2.5 Independent spike models

A hidden Markov Regime Switching model (MRS) can be used to model a specific time series by two or more phases or regimes describing the time series in different process states. The idea is to have different models for the different time series regimes and which model to be used is governed by a hidden Markov Chain. A Markov Chain is a process where the future state is solely dependent on the current state and the probability of a value or change for a Markov Chain of order N is given by

$$P\left(R_t = j | R_{t-1} = i\right) = \pi_{ij}, \quad \begin{cases} \pi_{ij} \geq 0 \\ \sum_{j=1}^{N} \pi_{ij} = 1 \quad \forall\ i \end{cases} \tag{2.15}$$

The Independent Spike Model is a class of hidden Markov Regime Switching models commonly used to model the electricity spot price. When used to model the spot price it is most

common nowadays to used three regimes (Regland and Lindström, 2012). These three regimes are the Base regime B, the Spike regime S describing times of extremely high prices and the Drop regime D describing extremely low prices. In the Independent Spike Model transitions between the spike and drop regimes are not allowed, leading to the transition probability matrix P of the hidden Markov Chain R for the process taking the form

$$P = \begin{bmatrix} \pi_{BB} & \pi_{BS} & \pi_{BD} \\ \pi_{SB} & \pi_{SS} & 0 \\ \pi_{DB} & 0 & \pi_{DD} \end{bmatrix} \tag{2.16}$$

where it follows from eq. 2.15 that transition probabilities $\pi_{BB} = 1 - \pi_{BS} - \pi_{BD}$, $\pi_{ss} = 1 - \pi_{SB}$, $\pi_{DD} = 1 - \pi_{DB}$. With time varying transition probabilities Independent Spike Models have been shown to perform well when used for electricity price forecasting (Janczura and Weron, 2010).

3 Data

This section covers the data gathering and processing performed in this thesis. The final data set contains 70 features of daily observations spanning between January 1st 2019 and the 31st of December 2022, totalling 1461 observations. Initially the data set contained more features over a larger period of time, but due to large discontinuities in the data some features had to be removed and the time span had to be reduced.

3.1 Data sources

The data was gathered from five different sources. Temperature data for Sweden was collected from SMHI, the Swedish Meteorological and Hydrological Institute. Initially more weather features were gathered, but they were removed from the data set due to large and frequent temporal gaps.

Swedish energy production and consumption data was gathered from Svenska Kraftnät, the state-owned transmission system operator in Sweden. All included energy features barring the nuclear energy production was available and gathered split up by bidding area.

Nord Pool was the source for the electricity spot price data. It is an European power exchange owned by the transmission systems operators of Norway, Sweden, Denmark and Finland and it's member countries includes Scandinavia, Finland, the Baltics and parts of continental Europe. The gathered price data was available partitioned by bidding area.

All Danish data was collected from Energinet, the state-owned transmission system operator in Denmark. Denmark has two bidding areas; the western bidding area DK1 and the eastern bidding area DK2. Like with the Swedish energy data it was available partitioned by bidding area and gathered as such.

Norwegian energy data was collected from Statnett, the state-owned transmission system operator in Norway. While Norway has five bidding areas the available data was not partitioned by region.

3.2 Preprocessing

Firstly, the obtained data was examined for observations of either extreme or missing values. This was found in 11 features and as all time spans of extreme or missing values are rather small imputation of the data should have no large impact on the feature characteristics.

The missing observations were filled using *Multivariate Imputation by Chained Equations* (MICE), a method commonly used when there are multiple features with missing data points in a large data set. Algorithm 5 below outlines the imputation process.

Algorithm 5: Multivariate Imputation by Chained Equations

Input: Data set $\boldsymbol{Y}$ shape (T, P) with $\hat{P} = \hat{p}_1, ..., \hat{p}_m$ features missing observations

1 Replace the missing data points with placeholder values, e.g. the feature-specific means

2 create data sets $\boldsymbol{\hat{Y}_1, ..., \hat{Y}_m}$ all having placeholder values of one specific feature $\hat{p}_1, ..., \hat{p}_m$ removed

3 For each data set $\boldsymbol{\hat{Y}_i}$, $i = 1, ..., m$:

   (a) removing observations of all features corresponding to the gaps in feature $\hat{p}_i$, fit a model to predict $\hat{p}_i$ using all other features with e.g. predictive mean matching, linear regression, classification and regression trees or random forest imputations

   (b) use the model yielded in 3(a) to predict the missing values of $\hat{p}_i$

4 Repeat steps 2-3 for a given amount of iterations to obtain a data set of imputed values for $\hat{P}$

5 Select one of the imputations based on its' fit to the data set and add it to the original data set $\boldsymbol{Y}$

Output: Data set $\boldsymbol{Y}$ with imputed observations

The imputation was done using the *MICE* package in **R**. The default method of the program is predictive mean matching, but due to the large dissimilarities between different features in regards to mean and variance it does not work. Instead, the classification and regression tree method was used.

After imputing the missing observations the stationarity of the data set was investigated. In 22 features relating to electricity price, solar energy and wind energy we identified either an upwards trend, a non-constant variance, or both. These features along with the hydroelectric energy are shown in figures 19 - 25 in Appendix A, partitioned by type and region. To reduce the trends and stabilize the variance of the electricity price, solar energy and wind energy, they were log-transformed using equation 3.1. The constant is added so that zero-valued observations can be transformed and remain zero-valued after the transformation. The modified log-transformation in eq. 3.2 was used for the returns to keep the distribution of negative and positive observations.

$$y_{t,p} = \log(y_{t,p} + 1) \tag{3.1}$$

$$y_{t,p} = \text{sgn}(y_{t,p}) \log(|y_{t,p}| + 1) \tag{3.2}$$

Hydroelectric energy differs from the other Swedish energy sources in its' highly regulatory properties. When there's a general energy deficit, hydroelectric energy production can be increased to meet demands. To better show larger ongoing trends in the energy market, a five-day rolling mean transformation was performed on the hydroelectric energy data, shown in eq. 3.3.

$$y_{t,p} = \frac{1}{5} \sum_{k=0}^{4} y_{t-k,p} \tag{3.3}$$

Transformed features are shown in figures 26-32 found in appendix A. The figures show that the upwards trend has been significantly reduced in all log-transformed features. After transformations had been performed all features were normalized to improve the clustering of the data using eq. 3.4, where μ_p and σ_p refers to the feature-specific mean and standard deviation.

$$y_{t,p} = \frac{y_{t,p} - \mu_p}{\sigma_p} \tag{3.4}$$

3.3 Finished data set

To allow comparisons between the different bidding areas in Sweden, the transformed and normalized data set was partitioned into four separate data sets, splitting the Swedish energy and weather data by region. These four data sets each contains 1461 observations of 37 features presented in table 1 below, and their means over time are shown in fig. 5. While most are easily understandable from the descriptions found in the table, one feature in particular requires some additional information. Schablon is a template-based energy delivery where the consumption is calculated instead of measured hourly. The delivery D_i is calculated by the balance responsible entity using the equation below

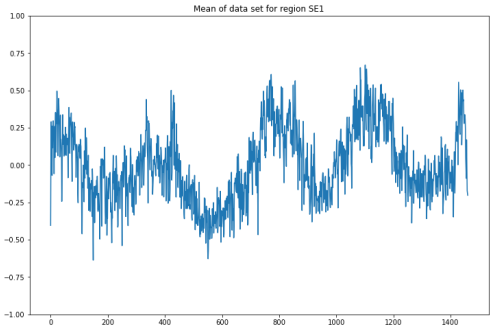
$$D_i = C\frac{s_i}{\sum_j s_j}$$

where C is the area consumption profile, s_i is the area share of the balance responsible entity and $\sum_j s$ is the shares of all different entities for the area. The balance responsible has entered an agreement with the transmission system operator and is economically responsible for matching the energy deliveries to the consumption within the specific area.

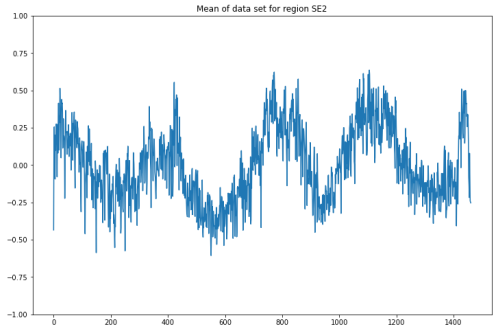
Clearly observable in all four partitioned data sets is that they have a strong seasonal trend in common. This seasonality was deliberately left in the data and not filtered for two reasons. The first reason is that as previously mentioned, the seasonality is a unique characteristic of electricity that separates it from other commodities. Secondly, the property of a seasonal trend is one of the main factors differentiating time series data from order independent data. Since the model used in this thesis was specifically developed to incorporate and make use of the temporal information in an efficient feature selection framework, this seasonal trend was therefore left in the data.

Feature name	Description
Price_SE#	Electricity spot price in the bidding area.
Rets_SE#	Electricity spot price returns in the bidding area.
Temp_SE#	Mean daily temperatures in the bidding area.
Cons_SE#	Daily energy consumption in the bidding area.
Hydro_SE#	Five-day rolling mean hydroelectric energy production in the bidding area.
Wind_SE#	Log-transformed daily wind evergy production in the bidding area.
Thermal_SE#	Daily thermal energy production in the bidding area.
Schablon_SE#	Template based energy deliveries in the bidding area.
Solar_SE#	Log-transformed daily solar energy production in the bidding area.
TotalLoad_DK#	Total daily electricity load in the bidding area.
Biomass_DK#	Daily biomass energy production in the bidding area.
Gas_DK#	Daily gas energy production in the bidding area.
HardCoal_DK#	Daily coal enery production in the bidding area.
Oil_DK#	Daily oil energy production in the bidding area.
Hydro_DK	Daily hydroelectric energy production in the bidding area.
OtherRenewable_DK#	Daily energy production in bidding arer of renewable energy from sources other than solar and wind energy.
Solar_DK#	Daily solar energy production in yjr bidding area.
Waste_DK#	Daily energy production from waste power plants in the bidding area.
OnshoreWind_DK#	Daily onshore wind energy production in the bidding area.
OffshoreWind_DK#	Daily offshore wind energy production in the bidding area.
ExchangeContinent_DK#	Daily energy import from continental Europe to the Danish bidding area.
ExchangeGreatBelt_DK1	Daily energy import from bidding area DK2 to DK1.
ExchangeNordicCountries_DK#	Daily energy import from Sweden and Norway to the Danish bidding area.
Import_NO	Daily energy imports for Norway.
Export_NO	Daily energy exports for Norway.
Flow_from_NO	Daily amount of energy export exceeding imports in Norway.

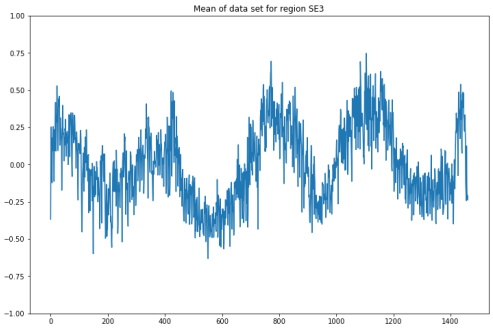
Table 1: Features included in the final data set with descriptions.



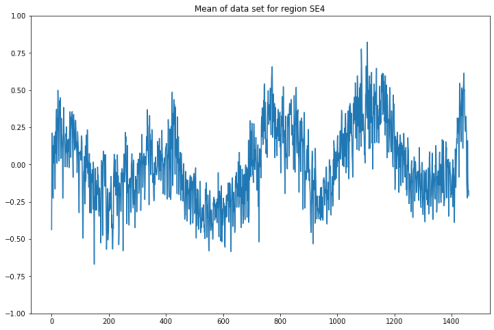
(a) Mean of data set used for region SE1.



(b) Mean of data set used for region SE2.



(c) Mean of data set used for region SE3.



(d) Mean of data set used for region SE4.

Figure 5: Means over time of the four different data sets used for modelling the electricity spot price bidding areas of Sweden. The means are calculated after transformation and normalization of the data sets. Discernible directly from the plots are that the data sets as a whole are stationary and contain a strong yearly seasonal trend.

4 Results

This section contains the results from the model hyperparameter selection followed by the model outputs. The model results are partitioned by region and shows the estimated state sequences and corresponding feature weights for 1-6 non-zero features.

4.1 Hyperparameter selection

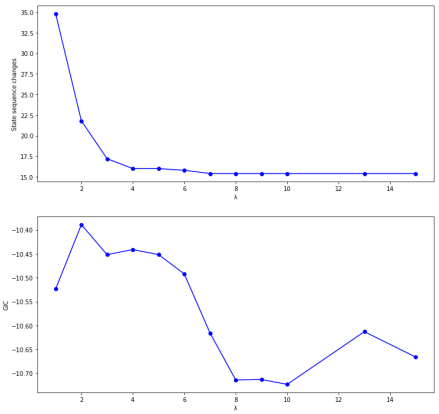
The model contains three hyperparameters; the number of latent states K, the regularization parameter λ and the tuning parameter κ. Based on previous studies (Janczura and Weron, 2010) and to improve comparability to other models, the number of latent states was set to $K = 3$.

The regularization parameter was determined using the GIC from eq. 2.14 in combination with examination of the number of state sequence changes. For the saturated model the parameters were set to $\bar{K} = 6$, $\bar{\lambda} = 0$, $\bar{\kappa} = \sqrt{P}$. The results are shown in figure 6. From the figure we see that for all regions the number of state sequence changes converges towards approximately 15 jumps when $\lambda \geq 8$. For region SE1 we have the smallest GIC value for $8 \leq \lambda \leq 10$, and in region SE2 the GIC values for $\lambda \geq 7$ are all similar. In the two other regions $\lambda = 15$ yields the lowest GIC, but the state sequence change graph indicated that it is a too large jump penalty (see fig. 4, section 2). For these reasons, the optimal choice of regularization parameter for all regions lies in the interval $8 \leq \lambda \leq 10$, and for all regions it was set to $\lambda = 10$.

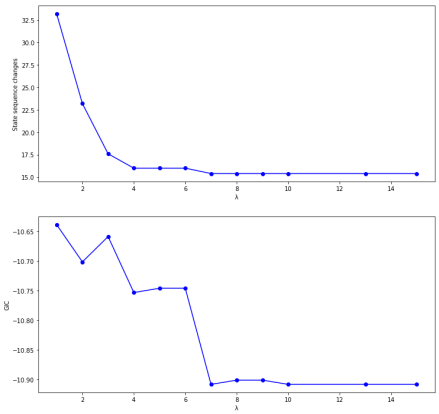
Like with the regularization parameter, the tuning parameter was selected using the GIC from eq. 2.14. As it governs the feature selection in the model, a different κ was calculated for each region and number of desired output features. For each tuning parameter calculation the model was run for 30 equidistant values $1 \geq \kappa \geq \sqrt{P}$ to determine a smaller interval which yields the desired number of selected features. After the smaller interval had been selected, $\mathcal{A}_0$ in eq. 2.13 was set to the number of desired features, K_0 was set to 3 and the GIC was calculated for 21 equidistant values of κ spanning the selected interval. The resulting GIC values are shown in figures 37-40 found in appendix B and selected tuning parameter values are shown in table 2.

	SE1	SE2	SE3	SE4
2 features	1.30	1.24	1.16	1.16
3 features	1.54	1.52	1.59	1.41
4 features	1.69	1.74	1.64	1.68
5 features	1.82	1.84	1.97	1.91
6 features	1.88	1.99	1.98	2.19

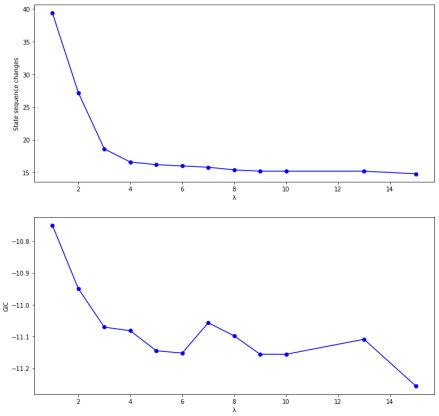
Table 2: Selected values of tuning parameter κ per region and desired output features.



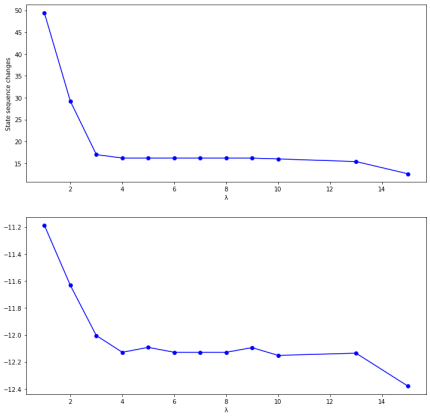
(a) Number of state sequence changes and GIC for region SE1.



(b) Number of state sequence changes and GIC for region SE2.



(c) Number of state sequence changes and GIC for region SE3.



(d) Number of state sequence changes and GIC for region SE4.

Figure 6: Number of state sequence changes and GIC value for different values of regularization parameter λ. For each value of λ, simulations were run with five equidistant values of κ spanning minimum value $\kappa = 1$ to maximum value $\kappa = \sqrt{P}$. For these five iterations, the mean value of state sequence changes and GIC was calculated and plotted against λ. Note the difference in resolution regarding the GIC values in the four figures.

4.2 Region 1 - Luleå

This section presents the results from the SJM for bidding area SE1. The selected features and their feature weights for all considered tuning parameter values are presented in a table, a correlation matrix of the selected features can be found in fig. 7 and the estimated state sequences for all scenarios is shown in fig. 8. Plots of all selected features can be found in fig. 33 in Appendix B.

	1 feature $\boldsymbol{\kappa = 1}$	2 features $\boldsymbol{\kappa = 1.30}$	3 features $\boldsymbol{\kappa = 1.54}$	4 features $\boldsymbol{\kappa = 1.69}$	5 features $\boldsymbol{\kappa = 1.82}$	6 features $\boldsymbol{\kappa = 1.88}$
Schablon_SE1	**1**	**0.7151**	**0.5173**	**0.4583**	**0.4018**	**0.3838**
Temp_SE1		0.2849	0.3781	0.3428	0.3168	0.3057
Load_DK2			0.1045	0.1347	0.1636	0.1645
Thermal_SE1				0.06417	0.1138	0.1187
Solar_DK2					0.003954	0.01753
Solar_DK1						0.009777

Table 3: Table summarizing feature weights results for region 1. First row shows specified number of desired output features and corresponding tuning parameter $\boldsymbol{\kappa}$. Leftmost column shows selected feature names, while other columns contains the yielded output feature weights dependent on feature count. The feature with highest weight is highlighted with bold font.

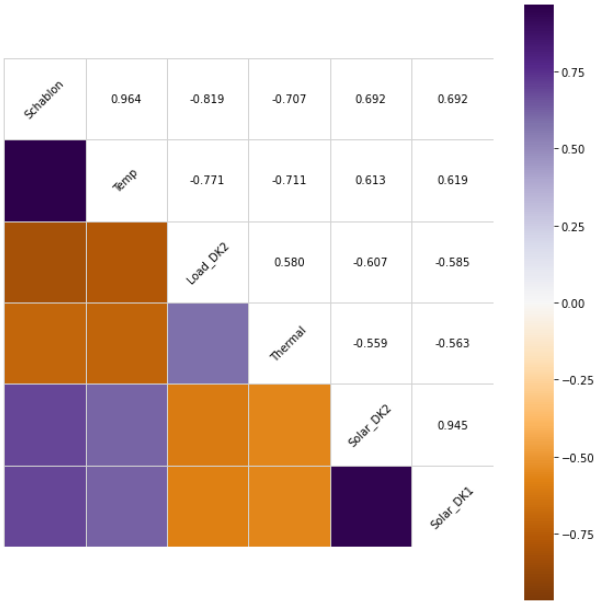


Figure 7: Correlation matrix of the features selected in bidding area SE1.

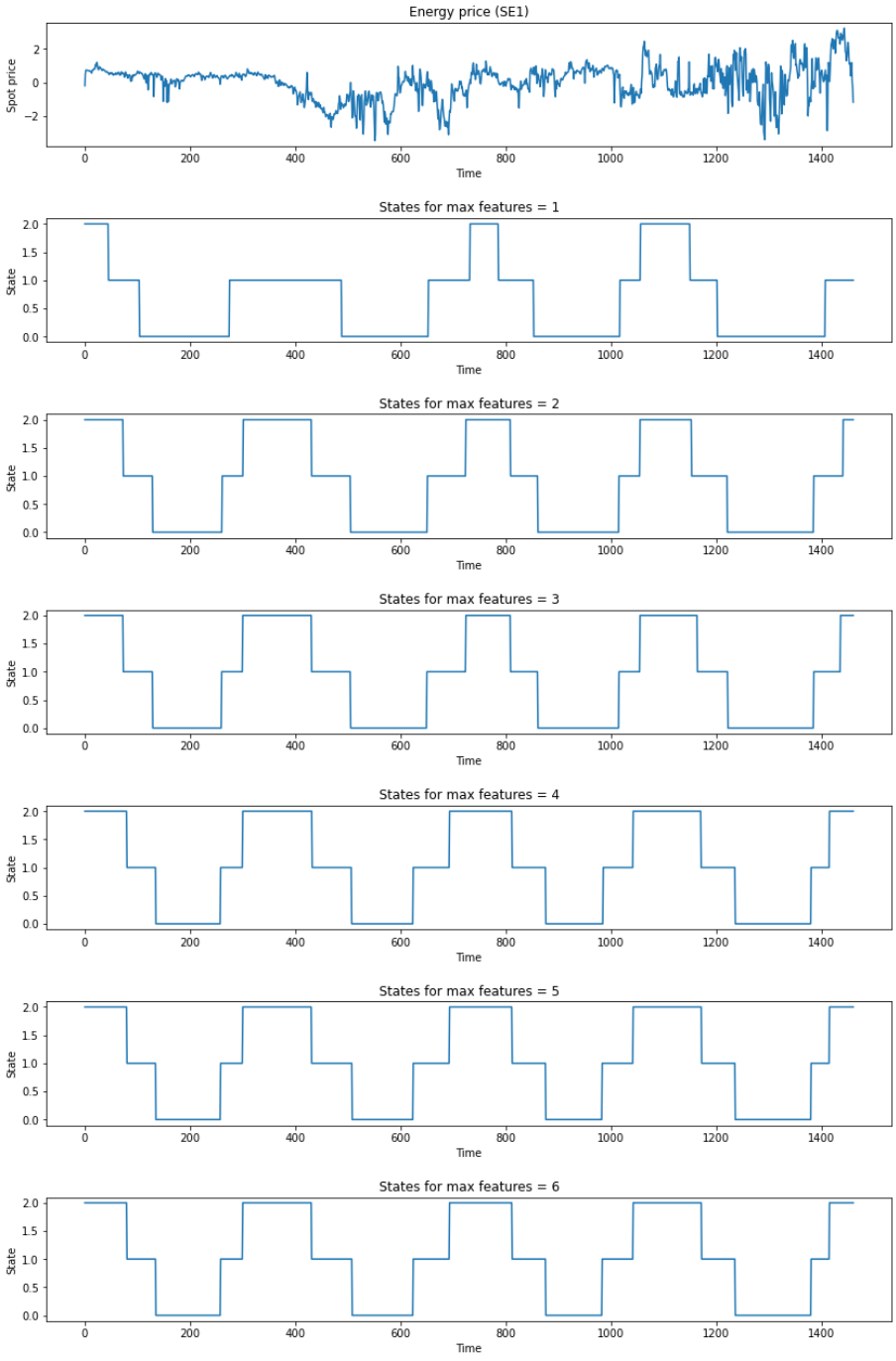


Figure 8: Estimated state sequences for the sparse jump model in region SE1. Topmost graph shows the normalized log-transformed energy price for the time period. Following graphs show the estimated states with different amount of selected features, having one feature in the top-most state sequence and six features in the bottom sequence.

4.3 Region 2 - Sundsvall

This section presents the results from the SJM for bidding area SE2. The selected features and their feature weights for all considered tuning parameter values are presented in a table, a correlation matrix of the selected features can be found in fig. 9 and the estimated state sequences for all scenarios is shown in fig. 10. Plots of all selected features can be found in fig. 34 in Appendix B.

	1 feature $\boldsymbol{\kappa = 1}$	2 features $\boldsymbol{\kappa = 1.24}$	3 features $\boldsymbol{\kappa = 1.52}$	4 features $\boldsymbol{\kappa = 1.74}$	5 features $\boldsymbol{\kappa = 1.84}$	6 features $\boldsymbol{\kappa = 1.99}$
Schablon_SE2	**1**	**0.7737**	**0.5841**	**0.4507**	**0.4130**	**0.3556**
Temp_SE2		0.2263	0.2603	0.2715	0.2604	0.2342
Load_DK2			0.1556	0.2241	0.2224	0.02049
Cons_SE2				0.05376	0.08114	
Solar_SE2					0.02303	0.2448
Solar_DK2						0.07246
Solar_DK1						0.07246

Table 4: Table summarizing feature weights results for region 2. First row shows specified number of desired output features and corresponding tuning parameter $\boldsymbol{\kappa}$. Leftmost column shows selected feature names, while other columns contains the yielded output feature weights dependent on feature count. The feature with highest weight is highlighted with bold font.

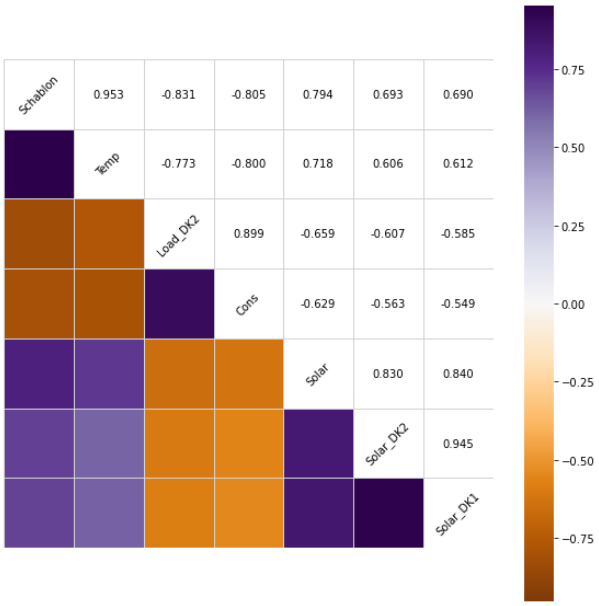


Figure 9: Correlation matrix of the features selected in bidding area SE2.

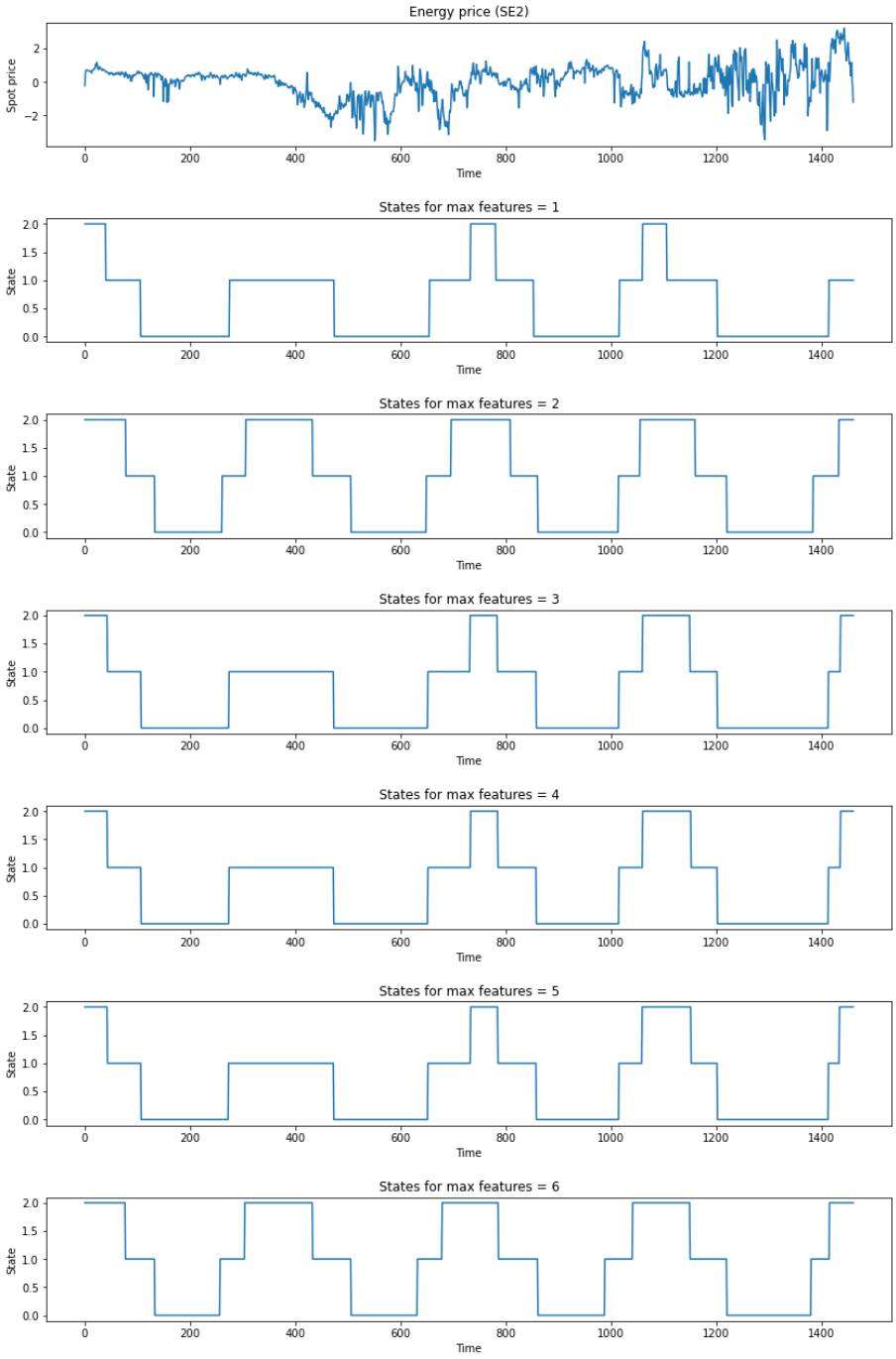


Figure 10: Estimated state sequences for the sparse jump model in region SE2. Topmost graph shows the normalized log-transformed energy price for the time period. Following graphs show the estimated states with different amount of selected features, having one feature in the topmost state sequence and six features in the bottom sequence.

4.4 Region 3 - Stockholm

This section presents the results from the SJM for bidding area SE3. The selected features and their feature weights for all considered tuning parameter values are presented in a table, a correlation matrix of the selected features can be found in fig. 11 and the estimated state sequences for all scenarios is shown in fig. 12. Plots of all selected features can be found in fig. 35 in Appendix B.

	1 feature $\boldsymbol{\kappa = 1}$	2 features $\boldsymbol{\kappa = 1.16}$	3 features $\boldsymbol{\kappa = 1.59}$	4 features $\boldsymbol{\kappa = 1.64}$	5 features $\boldsymbol{\kappa = 1.97}$	6 features $\boldsymbol{\kappa = 1.98}$
Schablon_SE3	**1**	**0.8484**	**0.4693**	**0.4796**	0.2920	0.2898
Thermal_SE3		0.1531	0.3974	0.3499	**0.3015**	**0.2991**
Temp_SE3			0.1333	0.1343	0.2068	0.2063
Load_DK2				0.03604	0.1972	0.1969
Cons_SE3					0.002438	0.006109
Biomass_DK2						0.001716

Table 5: Table summarizing feature weights results for region 3. First row shows specified number of desired output features and corresponding tuning parameter $\boldsymbol{\kappa}$. Leftmost column shows selected feature names, while other columns contains the yielded output feature weights dependent on feature count. The feature with highest weight is highlighted with bold font.

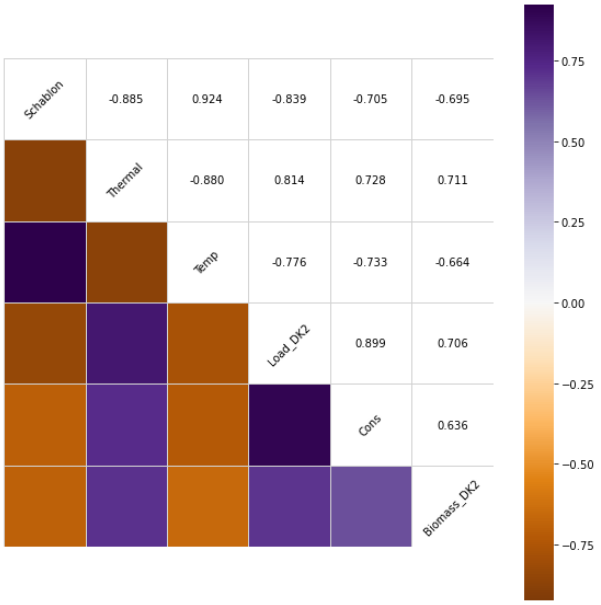


Figure 11: Correlation matrix of the features selected in bidding area SE3.

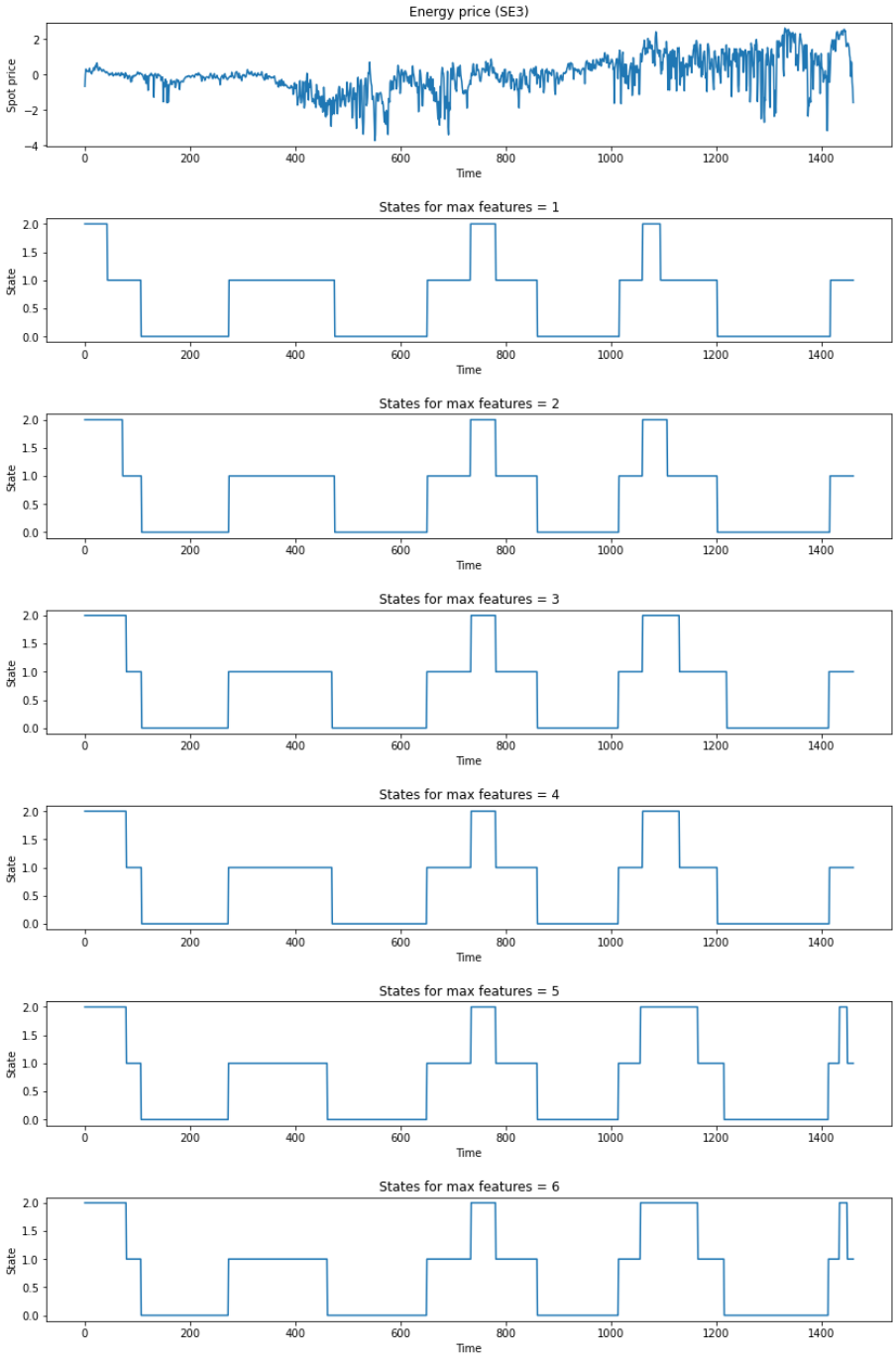


Figure 12: Estimated state sequences for the sparse jump model in region SE3. Topmost graph shows the normalized log-transformed energy price for the time period. Following graphs show the estimated states with different amount of selected features, having one feature in the topmost state sequence and six features in the bottom sequence.

4.5 Region 4 - Malmö

This section presents the results from the SJM for bidding area SE4. The selected features and their feature weights for all considered tuning parameter values are presented in a table, a correlation matrix of the selected features can be found in fig. 13 and the estimated state sequences for all scenarios is shown in fig. 14. Plots of all selected features can be found in fig. 36 in Appendix B.

	1 feature $\boldsymbol{\kappa = 1}$	2 features $\boldsymbol{\kappa = 1.16}$	3 features $\boldsymbol{\kappa = 1.41}$	4 features $\boldsymbol{\kappa = 1.68}$	5 features $\boldsymbol{\kappa = 1.91}$	6 features $\boldsymbol{\kappa = 2.19}$
Schablon_SE4	**1**	**0.8490**	**0.6470**	**0.4602**	**0.3700**	**0.2744**
Temp_SE4		0.1498	0.2814	0.2540	0.2376	0.2066
Hydro_SE4			0.07184	0.2792	0.2663	0.2339
Hydro_DK1				0.005773	0.09356	0.1574
Load_DK2					0.03246	0.1014
HardCoal_DK2						0.0264

Table 6: Table summarizing feature weights results for region 4. First row shows specified number of desired output features and corresponding tuning parameter $\boldsymbol{\kappa}$. Leftmost column shows selected feature names, while other columns contains the yielded output feature weights dependent on feature count. The feature with highest weight is highlighted with bold font.

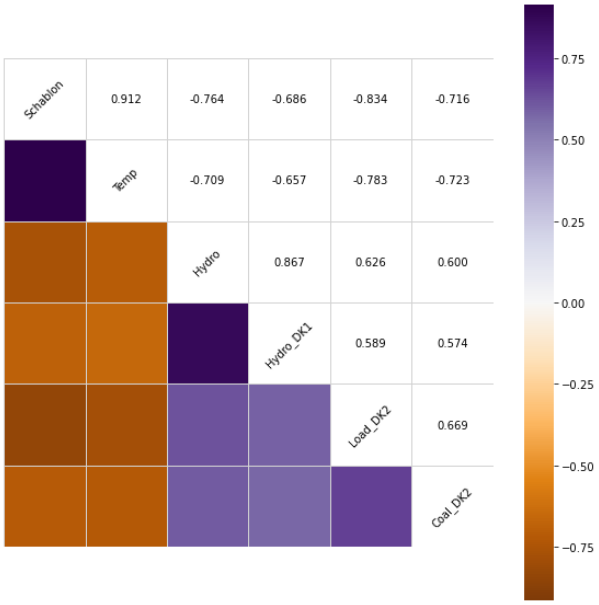


Figure 13: Correlation matrix of the features selected in bidding area SE4.

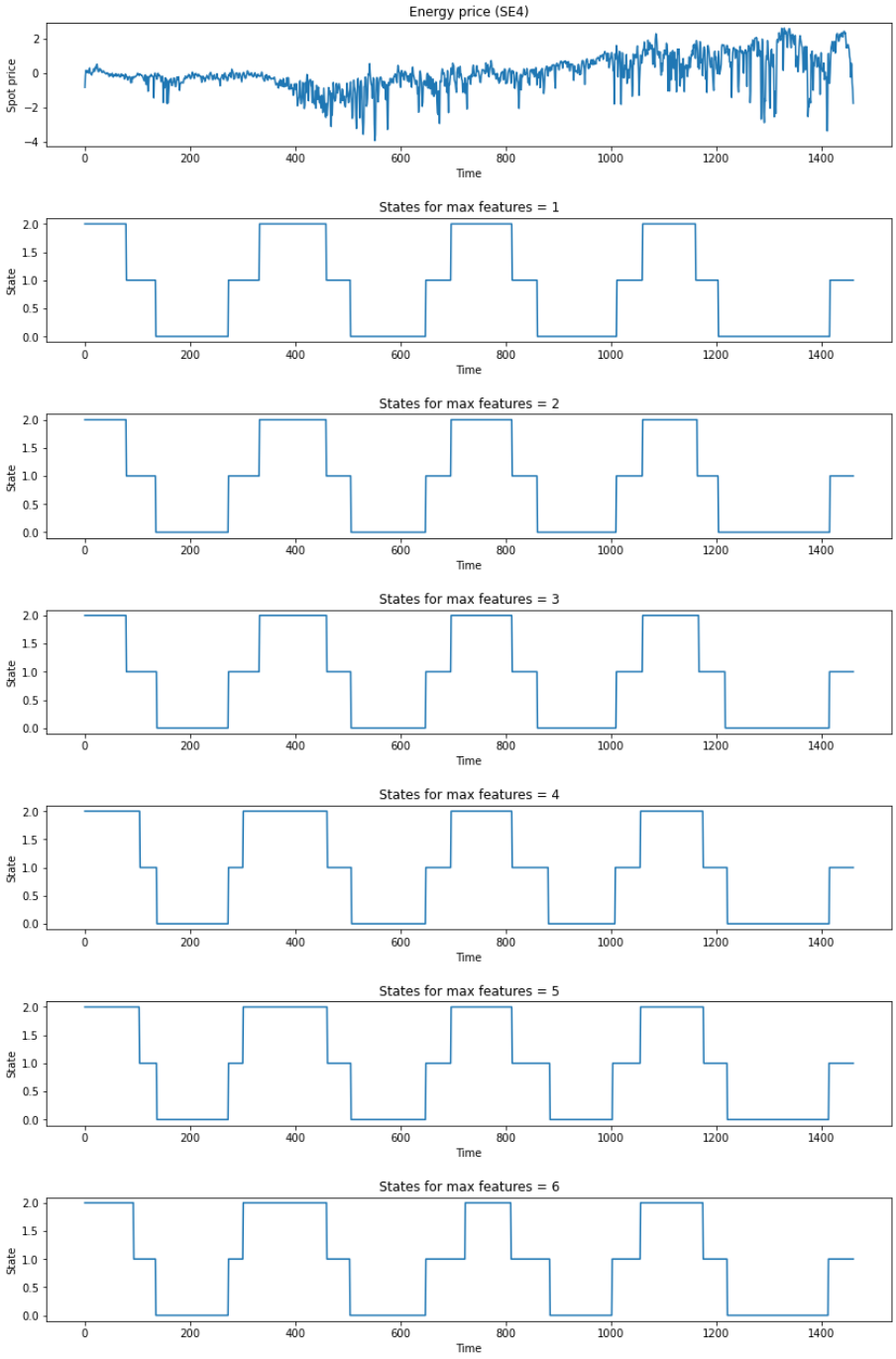


Figure 14: Estimated state sequences for the sparse jump model in region SE4. Topmost graph shows the normalized log-transformed energy price for the time period. Following graphs show the estimated states with different amount of selected features, having one feature in the topmost state sequence and six features in the bottom sequence.

5 Discussion

In this section the the results from the Sparse Jump Model are discussed. The main focus of this section is on the selected features for all bidding areas, characteristics of the features as well as commonalities and differences between the selection processes in the different areas. Furthermore, the estimated state sequences will also be analysed as well as the tuning parameter and its' impact on the results.

5.1 Selected features

This section covers the feature selection, starting with the selection process of the bidding areas in terms of feature order, then moving on to examining what the regions have in common and how they differ.

5.1.1 First feature

In all four regions, the first selected feature was Schablon, the template based energy deliveries. This comes as no surprise since it is based on forecasted energy needs within the region and should therefore be highly correlated to the other energy features. This is shown in figures 41-44 found in Appendix C, showing a correlation matrix for Schablon and energy features in the four bidding area. From the figures we can see that Schablon is strongly correlated with most energy features, only having weak correlation with the wind energy. We also see that there's a stronger correlation between Schablon and the hydroelectric energy production in the southern regions, which is interesting considering that a vast majority of the production takes places in the northern regions.

5.1.2 Second feature

The second selected feature was the temperature in bidding areas SE1, SE2 and SE4 while for area SE3 it instead was the thermal energy production. Comparing the correlation plots we see that while the temperature is highly correlated with other features in all regions, only in bidding area SE3 the thermal energy production is highly correlated to the other energy features. Examining the thermal energy production in all bidding areas, we see that region SE3 produces vastly more energy than the other three regions, shown in fig. 15. More than 95% of the time bidding area SE3 produces more thermal energy than the other three areas combined. While this might indicate why the thermal energy production was chosen in one region and not the other ones, it is worth noting that total energy production is not the sole contributing variable to getting selected as a feature. In fact, in bidding area SE3, the thermal energy production is only the largest energy source approximately 3% of the time and it contributes to a majority of the areas' total energy production less than 0.5% of the time. Despite this it

was chosen as a second feature above wind and hydroelectric energy, the two sources that often produces more energy in the region.

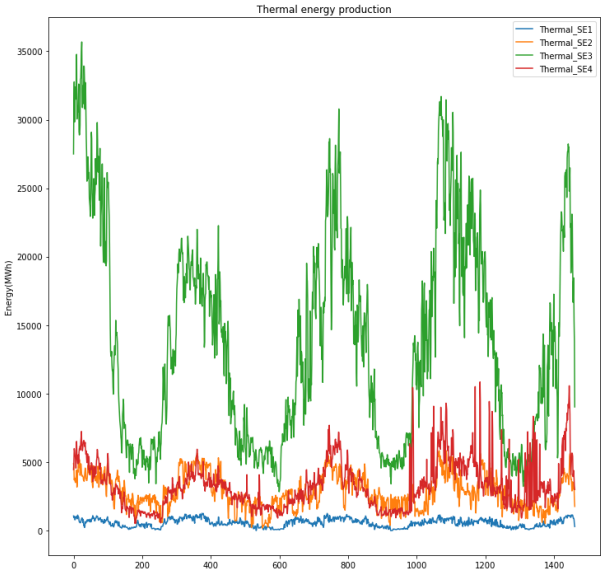


Figure 15: Thermal energy production in all Swedish bidding areas during the examined time period. Data is shown before transformation.

5.1.3 Third feature

The third selected feature was for region SE1 and SE2 the electricity load for bidding area DK2, while it for region SE3 was the temperature and for region SE4 was the hydroelectric energy production. While it may seem surprising at first to see a Danish energy feature, some explanation can be found by examining the data set. Out of the 37 features considered in the feature selection process for each bidding area 25 are Danish, from both region DK1 and DK2. Considering a large part of the data set consists of Danish features it is reasonable that one of them would be selected. Regarding the choice of Danish energy feature, electricity load is closely related to the energy consumption and should therefore be correlated to most energy features and bidding area DK2 has a larger energy exchange with Sweden than DK1. The more interesting thirdly selected feature of the four regions is the hydroelectric energy in bidding area SE4. Compared to the other three regions it is by far the smallest hydroelectric energy producer and within the bidding area it is most often outshone by the wind energy, being the largest energy source only 6% of the time. Despite this it has a stronger correlation to the other within-region features than hydroelectric energy has in other areas. It could be that due to its' role as an energy reservoir and relative scarcity, the feature coincides with times of extreme energy deficiency and is therefore chosen. Also worth noting is that as the third selected feature the hydroelectric energy of bidding area SE4 is the weakest, having a smaller feature weight.

5.1.4 Fourth feature

The fourth selected feature is the first time all regions select differently. In bidding area SE1 the thermal energy production is chosen while it in region SE2 is the consumption, in region SE3 the load of bidding area DK2 and in region SE4 the hydroelectric energy production of bidding area DK1. While this is the first time the consumption is selected in any region it could be argued that the electricity load, previously selected as a feature in both region SE1 and region SE2, is a very similar feature that also captures the energy needs at the time. Interesting is that the Danish load was selected before the within-region consumption. This is true for all regions barring bidding area SE4. Other than the fact previously mentioned that a large portion of the features considered are Danish, one possible explanation for at least the two northern bidding areas is the general energy surplus that defines them leading to the within-region consumption being less important. This does however not hold true for bidding area SE3, where the Danish energy load was selected as the fourth feature before the consumption even though the area is defined by a general energy deficiency. This indicates that the number of Danish features in the data set is the more probable reason for the load being selected over the consumption. In bidding area SE4, the first feature from area DK1 is selected. Worth noting is that there is no data for hydroelectric energy in bidding area DK2 in the data set, so the feature from area DK1 was not preferred over some similar feature from area DK2. Similarly to southernmost Sweden, the hydroelectric energy production in Denmark is minimal. Comparing the two, they show a very similar seasonal trend, shown in fig. 16. This is to be expected considering the geographical proximity of the two areas means that they experience similar climate conditions at a given time, which should lead to similar energy production and consumption. Other than to the Swedish hydroelectric energy production the feature is not strongly correlated to other selected features in the region. As a fourth feature it is by far the weakest one added, having a feature weight of less than 1%.

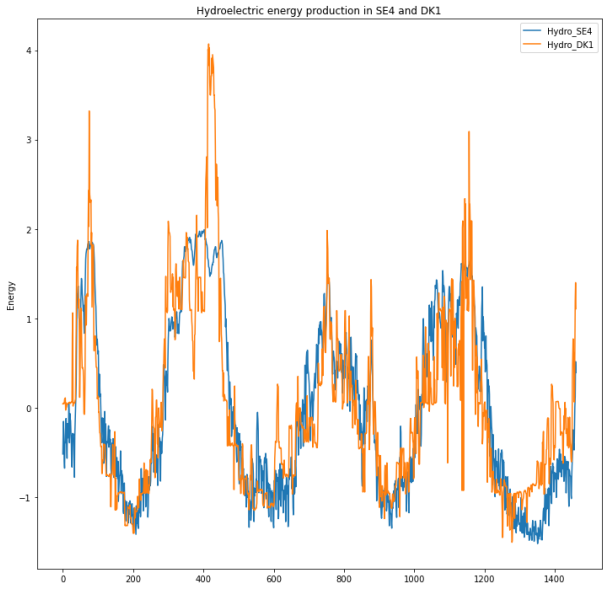


Figure 16: Hydroelectric energy production in bidding area SE4 and bidding area DK1. Data is normalized.

5.1.5 Fifth feature

Like with the fourth selected feature in bidding area SE4, the fifth selected features in areas SE1 and SE3 are very weak, both having a feature weight of less than 1%. For region SE1 it was the solar energy production in region DK2 and for area SE3 the with-region consumption. Other than a matching seasonal trend, and, once again, the large number of Danish feature in the data set, it is hard to say how or why Danish solar energy production is selected as a feature for bidding area SE1 in northern Sweden. It is also surprising that the consumption in bidding area 3, the region with the largest population, has such a low feature weight and is not selected earlier. In regions SE2 and SE4 the selected features were the within-region solar energy production and the DK2 electricity load respectively. While their weights are fairly small they're still significantly larger than the selected features in bidding areas 1 and 3. One possible reason that will be discussed later for why there is this large disparity is the tuning parameter and its' selection process.

5.1.6 Sixth feature

The sixth and final selected feature in bidding area SE1 is the solar energy production of area DK1. This feature is very strongly correlated with the fifth selected feature, the solar energy of bidding area DK2, and is probably selected for the same reasons. In region SE2 the fourth selected feature, the within-region consumption, is dropped in favour of the solar energy pro-

duction in both Danish bidding areas as the fifth and sixth features. This is the only region in which a previously selected feature is removed and worth noting is that the two added features are equally weighted and fairly largely weighted compared to the lastly selected features of the other bidding areas. With this change the two northern regions have five features in common out of six, only differing in that the thermal energy production was selected in region SE1 while in region SE2 it was the within-region solar energy production was selected instead. In bidding area SE3 the final feature selected was the biomass energy production of region DK2 and in area SE4 it was the coal energy production of region DK2. Both these features shows some seasonality but less than other selected features and they are also more volatile.

5.1.7 Regional commonalities and differences

All four regions had three selected features in common; Schablon, the temperature and the DK2 electricity load. The first selected feature is also the strongest feature in three out of the four regions regardless of number of non-zero features considered. Only in bidding area SE3 it is surpassed by the second selected feature, the thermal energy production, when five features has been selected. As previously discussed it is in an uniquely large energy source compared to the other regions and has a strong seasonality relative to the other large energy sources in the region. It was also selected over the temperature which was the second selected feature in all other regions and third in bidding area SE3. That the temperature is selected early in all regions is not that surprising considering that it has a very strong seasonal trend and is tied to the consumption in terms of heating and cooling.

While the third common feature, the electricity load of bidding area DK2, was selected in all regions, interestingly enough it was selected earlier in the two northern regions. In both these regions it was selected as the third feature while in bidding areas SE3 and SE4 it was the fourth and fifth feature respectively even though they are geographically closer to Denmark. A possible reason for this is that both northern bidding areas mainly produce hydroelectric and wind energy, both features with weak seasonality compared to other energy features.. Having weaker highly seasonal features could lead to a general preference of the Danish features when clustering the data in the northern bidding areas.

One of the more interesting observation made in the feature selection process is the preference of the Danish electricity load over the within-region consumption. As previously mentioned they are very similar features essentially measuring how much energy is used within an area. This similarity is illustrated in fig. 45 found in Appendix C showing the electricity load of the Danish bidding areas and the consumption of all Swedish bidding areas. From the figure we see that all regions share a very strong seasonal trend, and extreme events such as the dip at the turn of the year 2019-2020 and the spike in early 2021 looks to be a common denominator

in all regions. The main difference is that in bidding areas SE3, SE4 and DK1 large intraday differences seems to be more common than in the other three regions while still following the same seasonality. Despite this similarity, the electricity load of bidding area DK2 was preferred over all within-region consumption features.

A possible explanation previously mentioned was the fact that most of the feature in the data set was Danish which leads to a preference of these Danish features. This was tested by removing the features of bidding area DK1 from the data set and examining how it affects the feature selection process in terms of both weights and selected features in all regions. The results of this test can be found in table 7. Only in bidding area SE4 was there any change in the selected features and this due to the removal of the feature which was selected in the original data set, the hydroelectric energy of bidding area DK1. When included it was selected over the electricity load of DK2 even with all other features of region DK1 removed. While the test could be improved by calculating new tuning parameters for the modified data sets, it still shows that the removal of the Danish features had minimal impact on the feature selection and weighting. Only bidding area SE3 where the weight of the Danish feature was approximately halved shows any remotely significant change, and it is equally possible that this could be because of an unoptimized tuning parameter rather than the removal of the Danish features. The results from the test indicates that it seems to be some other reason than the amount of Danish features that leads to the preference of the electricity load over the within-region consumption. One possible reason is that a high electricity load might lead to increased import needs to Denmark from Sweden. This could affect not only the closest bidding area but all four considering the general energy deficit in southern Sweden would mean that to increase exports more energy would need to be transmitted from the northern regions. If that is the reason for feature selection one would however expect that the import feature specifically for the Nordic countries to Denmark included in the data set would have been selected instead, but it could be that the electricity load feature contains that information as well as other information important to the clustering process.

SE1	Schablon	Temp	Load DK2	Thermal	-
Full dataset	0.4583	0.3428	0.1347	0.06417	-
DK1 removed	0.4580	0.3423	0.1366	0.06302	-
SE2	Schablon	Temp	Load DK2	Cons	-
Full dataset	0.4507	0.2715	0.2241	0.05376	-
DK1 removed	0.4507	0.2715	0.2241	0.05376	-
SE3	Schablon	Thermal	Temp	Load DK2	-
Full dataset	0.4796	0.3499	0.1343	0.03604	-
DK1 removed	0.4480	0.3845	0.1512	0.01634	-
SE4	Schablon	Temp	Hydro	Hydro DK1	Load DK2
Full dataset	0.4602	0.2540	0.2792	0.05773	-
DK1 removed	0.4334	0.2723	0.2903	-	0.004060

Table 7: Feature weights for four selected features in all bidding areas with two different data sets. The upper row shows the yielded feature weights for the data set used earlier, while the lower row shows the feature weights for the modified data set, where all features from region DK1 was removed. In bidding area 4 the tuning parameter was set to $\kappa = 1.68$ for the original data set and changed to to $\kappa = 1.70$ for the modified to still select four features. In all other regions it was kept the same same for both data sets.

In all four regions one and only one within-area energy production feature was selected; the thermal energy production in region SE1 and SE3, the solar energy production in SE2 and the hydroelectric energy production in bidding area SE4. They are shown in fig. 46 together with their data set means and in fig. 47 as a percentage of total within-region energy production, both found in Appendix C. All four features has a one-year seasonal trend, most clearly observable in the solar energy feature. While it has the strongest seasonality it is also the feature whose seasonality differs the most from the data set mean. Firstly, the peaks of the solar data coincides with the troughs of the data set mean while for the other three features the peaks coincides with the mean peak. Compared to the other features there's also a larger temporal deviation; solar energy production does not reach its' seasonal maximum at the same time as the data set minimum. Regarding their within-region production share the two features of the northern regions contributes far less than the features of bidding areas SE3 and SE4, never exceeding a 5% share within their respective regions. It is interesting that the solar energy was selected for bidding area SE2 considering that it is a very small energy contributor to the area and out of the four features it has the weakest match to the data set in terms of seasonality. Worth noting is that the dominating energy features of bidding area SE2, the wind and hydroelectric energy, has no seasonal trend comparatively speaking, as shown in fig. 48 found in Appendix C. Furthermore it was selected as the fifth feature for region SE2 while for the other areas the within-region energy features were selected earlier, indicating that it is a weaker

feature overall. The wind and hydroelectric energy production dominates bidding area SE1 in a similar fashion, and a feature with stronger seasonality matching the data set mean was chosen instead. It is interesting that both northern areas chose features with a small production share and strong seasonality but did not choose the same feature. Regarding bidding areas SE3 and SE4, both features largely contributes to the within-area energy production and their seasonality coincides with the data set means. While the hydroelectric energy is also a large contributor in the two northern regions it serves a different purpose in those bidding areas due to their energy surplus characteristics. These features are more connected to the energy market as a whole while the hydroelectric energy produced in SE4 is more tied to the bidding area; in bidding areas SE1 and SE2 there is a much smaller seasonal trend compared to the seasonality of SE4 where the peaks coincide with the within-area consumption. This is also shown in the correlation matrices, where the hydroelectric energy is stronger correlated to the other features in the southern regions.

Regarding the features selected fifth and sixth in the four bidding areas there seems to be a fairly strong preference for danish features in all four regions. There are however some discernible characteristics between the northern and the southern regions. In both regions SE1 and SE2 the two finally selected features were the solar energy production of the two danish bidding areas, DK1 and DK2. These features are not strongly correlated to the other selected features of the regions but they do have a very strong seasonal trend and was probably selected because of that. The selected features of area SE3 was the within-region consumption and the biomass energy production of region DK2, while it for area SE4 was the electricity load and coal energy production for region DK2. While these features also have a seasonal trend they can be more strongly linked to market extremes, as coal and biomass energy production are used to combat energy deficits.

5.2 The state sequence

In all four bidding areas the state sequence is estimated quickly, showing its' approximate form as early as when only considering the first feature. This is especially true for region SE4, where there is minimal change in the state sequence when moving from one feature all the way up to six selected features. For the other regions, some changes occur when increasing the number of features. Firstly, the number of state sequence changes increases. For region SE3 this happens when moving to five selected features, while the increase happens already when the second feature is selected in bidding areas SE1 and SE2. The number of state sequence changes does however decrease when selecting for three features in bidding area SE2, and only increases again when the sixth and final feature is selected. This is the only instance where an increase of non-zero weighted features decreases the amount of state sequence changes, and it is unclear why it happens especially considering bidding area SE1 has the same first three selected features.

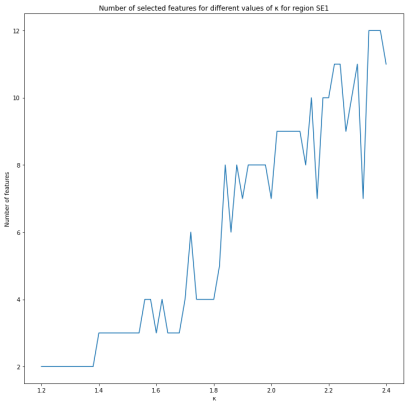
It is clear that with an increase of number of selected features the state sequence becomes more similar in shape to the data set means, exhibiting fairly strong seasonality in all bidding areas with six selected features. This is to be expected and shows that the selected features fairly accurately represents the whole data set. State 2 temporally coincides with the peaks and state 0 with the troughs, with state 1 aligning with the transitions between the two. These states can also be said to represent the different seasons of the year having state 2 represent the winter, state 0 the summer, and state 1 fall and spring. Another characteristic that all four estimated state sequences have in common is that and increase of number of selected features decreases the number of observations in state 1, moving them to states 0 and 2.

Comparing the shape of the estimated state sequences to Independent Spike Models, there are some large discrepancies. Assuming state 1 to be the base regime and state 2 and 0 the spike and drop regime respectively, he most obvious difference is the number of states in the two non-base regimes and the persistence of these states. In ISM a majority of the estimated states are generally in the base regime, state changes are much more common and the persistence of the spike and drop regimes are much lower (Janczura and Weron, 2010). The results in this thesis instead have most observations in the non-base regimes, few state changes and a strong state persistence. These large differences are explained by the fact that the regimes do not represent spikes and drops for the SJM in this thesis, but rather larger ongoing persistent trends in the energy market. Worth noting is this shape of the estimated state sequence is not an inherent characteristic of the SJM but a deliberate choice; one could easily increase the number of estimated observations in the base regime by increasing the jump penalty for transitioning from the base regime and reducing it for transitioning from non-base regimes. This would yield less persistent non-base regimes similar to what is found in ISM.

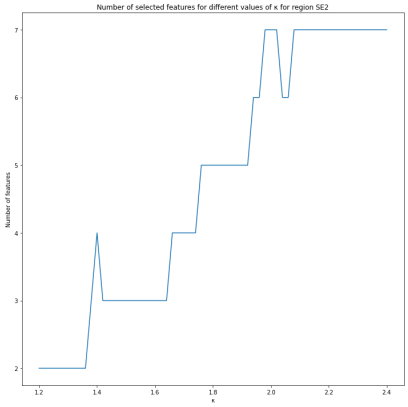
5.3 The tuning parameter

In the feature selection process, increasing the number of features sometimes has minimal impact on feature weights and the estimated state sequence. This is observed in bidding area SE1 when moving from five to six features and in bidding area SE3 both when moving from three to four features and when moving from five to six features. What these instances have in common is that they are accompanied by a small change in tuning parameter value, the worst case being that the tuning parameter for selecting five parameters in region SE3 is $\kappa = 1.97$ and to select six features it is $\kappa = 1.98$. These uneven tuning parameter changes are not by choice but rather by necessity, and fig. 17 illustrates the issue. Depending on region and amount of non-zero features, the number of candidate tuning parameter values that yields the desired outcome varies greatly, and only in bidding area SE4 did the amount of selected futures strictly increase with an increase of tuning parameter value. In region SE3 the only tuning parameter value of two decimal resolution yielding five selected features was $\kappa = 1.97$, and as seen in the figure the number of selected features for the bidding area increases from four to eight with

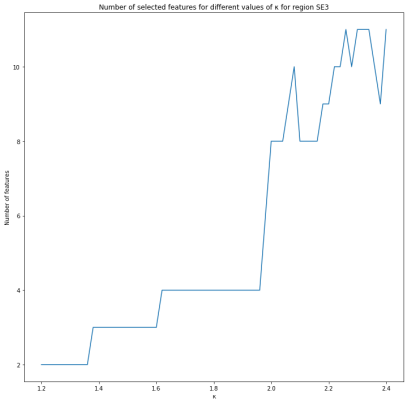
minimal change of tuning parameter value. Furthermore, as shown in fig. 18, certain tuning parameter values can select a varying amount of features. It is possible that these issues could be circumvented by changing the method of hyperparameter selection. In this thesis the jump penalty was calculated first, then using that jump penalty the tuning parameter was calculated. Since both these parameters affect the feature selection process, an improved hyperparameter selection method would probably alleviate this problem.



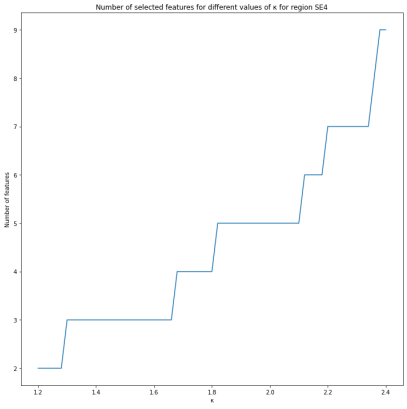
(a) Number of selected features in bidding area SE1 for varying values of κ, holding all else equal.



(b) Number of selected features in bidding area SE2 for varying values of κ, holding all else equal.



(c) Number of selected features in bidding area SE3 for varying values of κ, holding all else equal.



(d) Number of selected features in bidding area SE4 for varying values of κ, holding all else equal.

Figure 17: Number of selected features in all regions dependent on value of tuning parameter κ. For all bidding areas number of latent states was set to $K = 3$ and jump penalty to $\lambda = 10$.

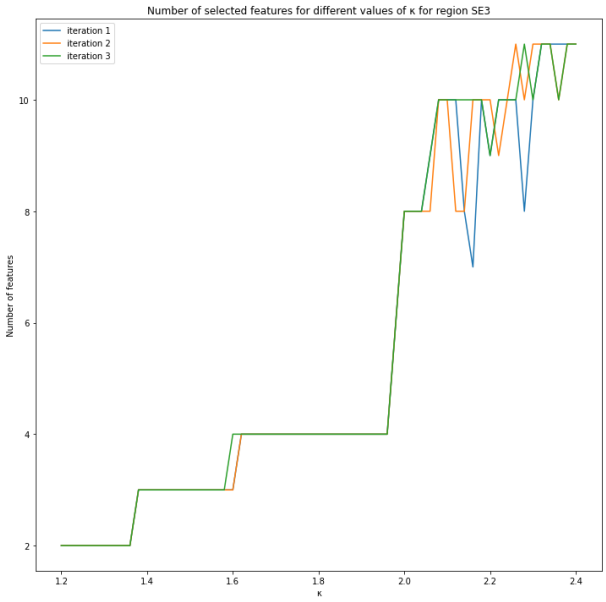


Figure 18: Three iterations of number of selected features in bidding area SE3 for varying values of κ, holding all else equal.

5.4 Improvements and further research

A lot of changes could be made in regards to the data set. The results show that the selected features and the data set as a whole contains a strong seasonality. It would be interesting to look at deseasonalized data instead and see if and how it affects the feature selection process. One could also move to an hourly time frame capturing entirely different characteristics of the energy market.

One of the more interesting results from this thesis was the prevalence of selected Danish features. A topic for further research would be to put a larger focus on the foreign markets and how they affect the different bidding areas of Sweden.

In this thesis the regularization parameter was kept uniform leading to few and persistent state sequence changes. By changing it one could get more frequent and shorter changes as seen in Independent Spike Models, and see which features correlates to more extreme spikes and drops rather than large trends.

Improvements could and probably should be made to the hyperparameter selection method. As the method stands, it leads to a very limited amount of candidate tuning parameter values in certain situations. Considering the dynamic behaviour of the tuning and regularization parameters there should be some better suited methods for the problem.

6 Conclusion

In this thesis the purpose was to investigate the four electricity bidding areas of Sweden using the Sparse Jump Model to find discernible similarities and differences between the regions when it comes to the feature selection process. The results show that all four regions had three selected features in common, all with a strong seasonal trend. Except for the first selected feature, Schablon, the order in which the two other common features were selected and their weighting differed between the regions. Furthermore, while one within-region energy production feature was selected in all bidding areas, the characteristics of the selected features varies drastically. Regarding the fifth and sixth selected features for all regions, the two northern regions seemed to prioritize features with strong seasonality while the selected features in the southern regions are more strongly correlated to market extremes.

The most interesting observation in the feature selection process was the preference of the danish electricity load over the within-area consumption in all four regions. Initially thought to be due to the large portion of danish features in the data set, it was shown that this was not the reason for the preference.

Regarding the state sequence, as expected it shows a strong seasonal trend matching the data set in all regions. Worth noting is how few features are needed to fairly accurately estimate the state sequence, especially in bidding area SE4 where the difference between the sequences for one and six features are very small.

For further use of the Sparse Jump Model in the specific way is is used in this thesis, improvements to the hyperparameter selection process should be made.

References

Bemporad, A., Breschi, V., Piga, D., & Boyd, S. P. (2018). Fitting jump models. *Automatica*, *96*, 11–21.

Cortese, F., Kolm, P., Linstrom, E., et al. (2023). Generalized information criteria for sparse statistical jump models. *Symposium i anvendt statistik-Copenhagen Business School*, 68–78.

de Hautecloque, A., & Hancher, L. (2011). The svenska kraftnät case: Introduction of bidding zones in sweden. *Network Industries Quarterly*, *13*, 20–22.

Donoho, D. L. (1995). De-noising by soft-thresholding. *IEEE transactions on information theory*, *41*(3), 613–627.

Escribano, A., Ignacio Peña, J., & Villaplana, P. (2011). Modelling electricity prices: International evidence. *Oxford bulletin of economics and statistics*, *73*(5), 622–650.

Fan, Y., & Tang, C. Y. (2013). Tuning parameter selection in high dimensional penalized likelihood. *Journal of the Royal Statistical Society: SERIES B: Statistical Methodology*, 531–552.

Green, R., & Vasilakos, N. (2010). Market behaviour with large amounts of intermittent generation. *Energy Policy*, *38*(7), 3211–3220.

Janczura, J., & Weron, R. (2010). An empirical comparison of alternate regime-switching models for electricity spot prices. *Energy economics*, *32*(5), 1059–1073.

Knittel, C. R., & Roberts, M. R. (2005). An empirical examination of restructured electricity prices. *Energy Economics*, *27*(5), 791–817.

Lindström, E., Norén, V., & Madsen, H. (2015). Consumption management in the nord pool region: A stability analysis. *Applied Energy*, *146*, 239–246.

Lloyd, S. (1982). Least squares quantization in pcm. *IEEE transactions on information theory*, *28*(2), 129–137.

Nystrup, P., Kolm, P. N., & Lindström, E. (2021). Feature selection in jump models. *Expert Systems with Applications*, *184*, 115558.

Nystrup, P., Lindström, E., & Madsen, H. (2020). Hyperparameter optimization for portfolio selection. *The Journal of Financial Data Science*.

Regeringen. (2019, December). *Sveriges femte rapport om utvecklingen av förnybar energi enligt artikel 22 i direktiv 2009/28/eg*. https://www.regeringen.se/rapporter/2019/12/sveriges-femte-rapport-om-utvecklingen-av-fornybar-energi-enligt-artikel-22-i-direktiv-200928eg/

Regland, F., & Lindström, E. (2012). Independent spike models: Estimation and validation. *Finance a Úver*, *62*(2), 180.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(2), 411–423.

Witten, D. M., & Tibshirani, R. (2010). A framework for feature selection in clustering. *Journal of the American Statistical Association*, *105*(490), 713–726.

Wright, S. J. (2015). Coordinate descent algorithms. *Mathematical programming*, *151*(1), 3–34.

7 Appendix A

This appendix contains complementary images to the Data section.

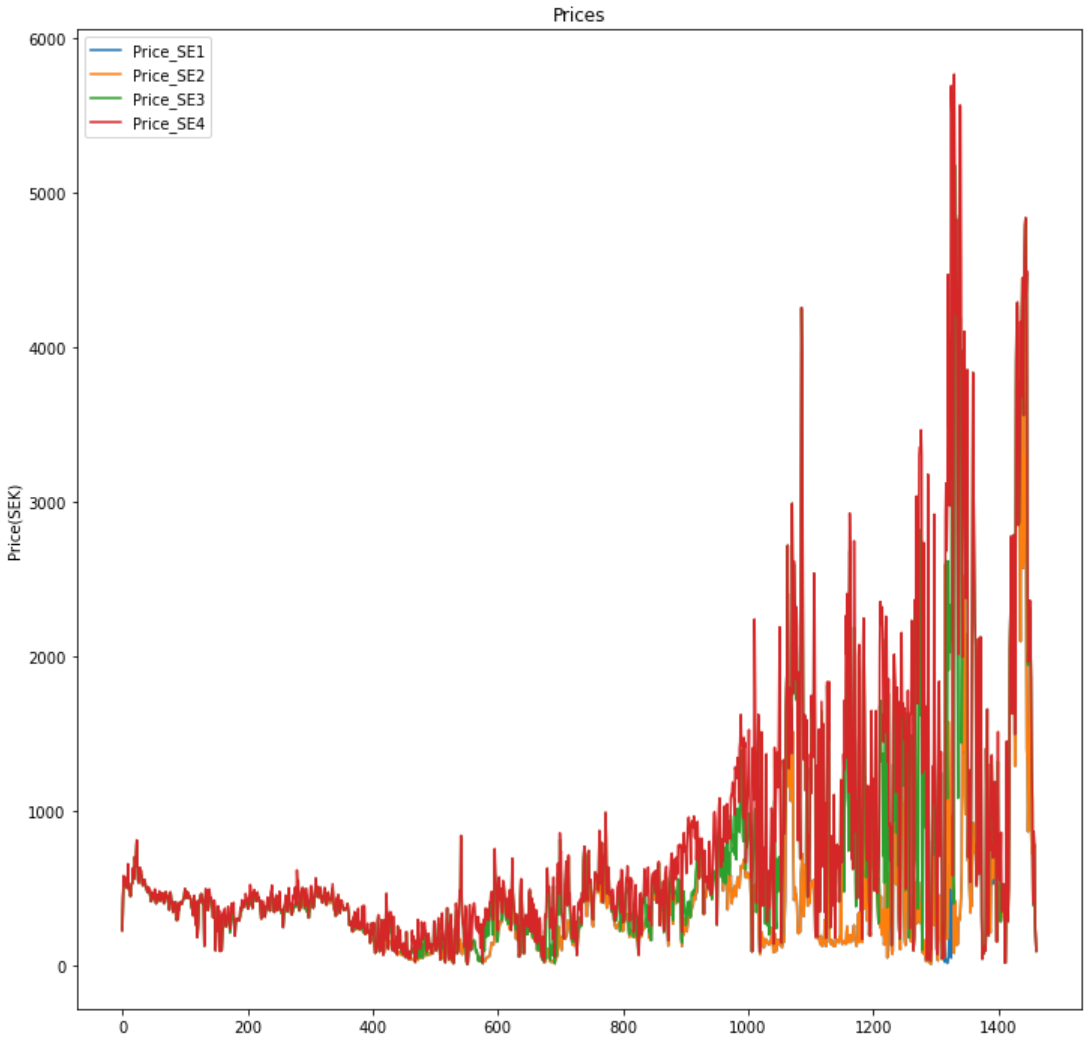


Figure 19: Mean daily electricity spot price in Sweden partitioned by region.

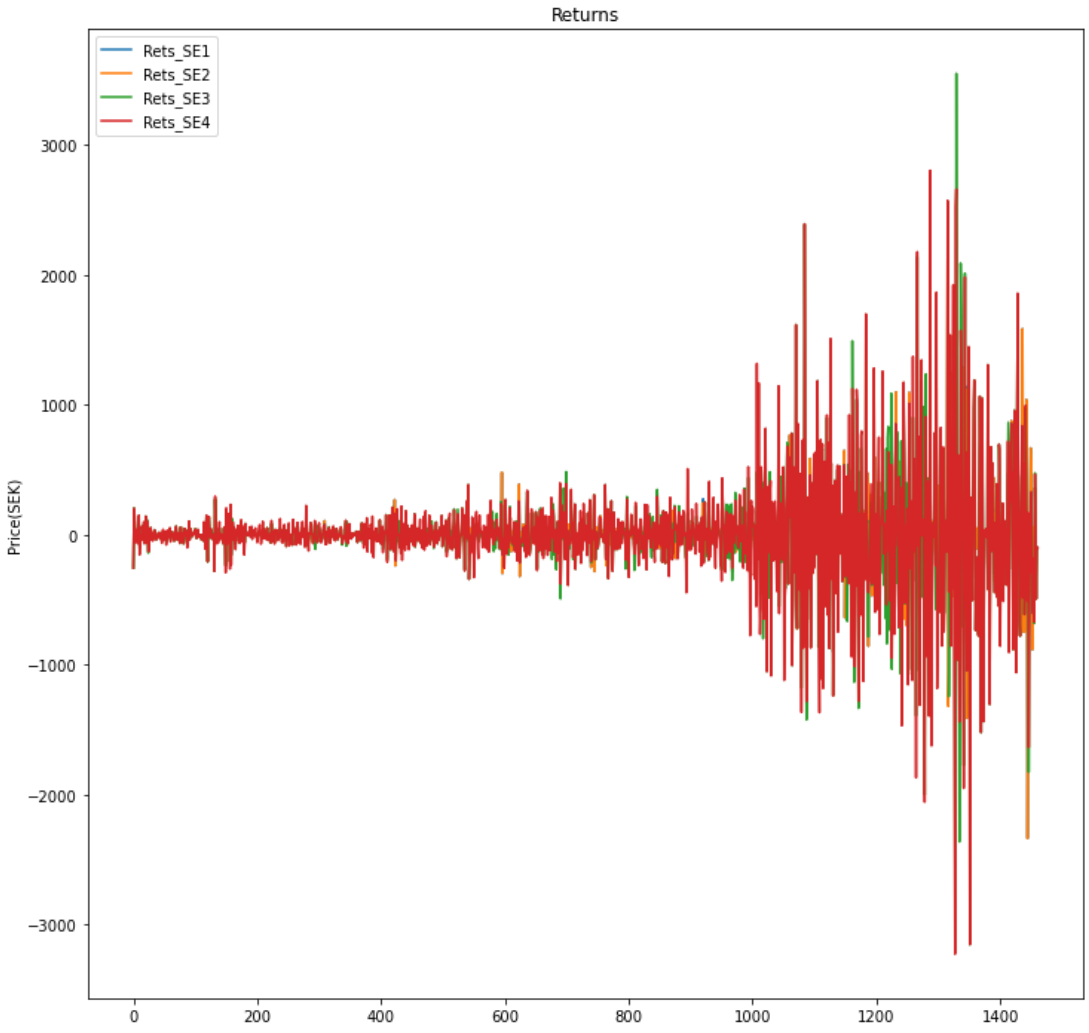


Figure 20: Mean daily electricity spot price returns in Sweden partitioned by region.

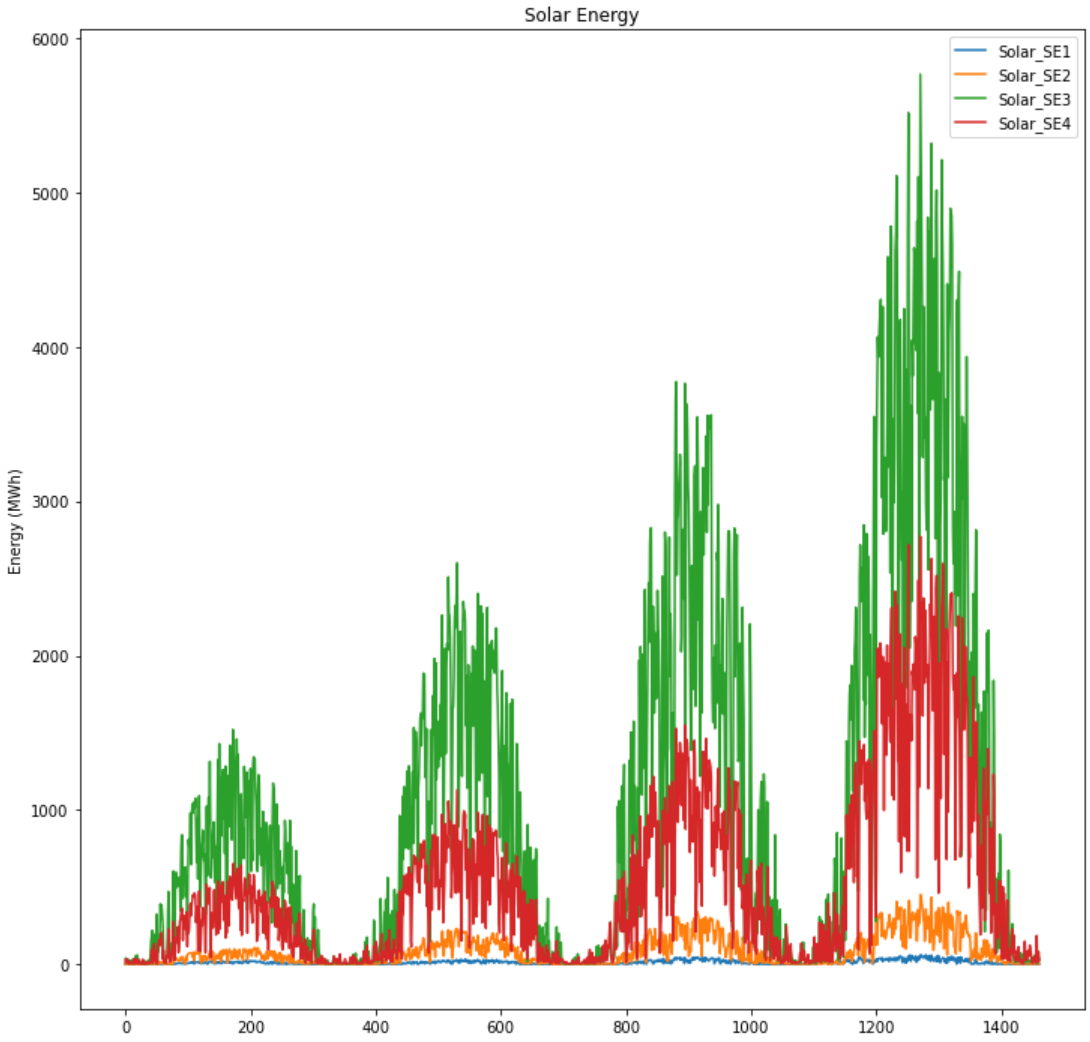


Figure 21: Total daily solar energy production in Sweden partitioned by region.

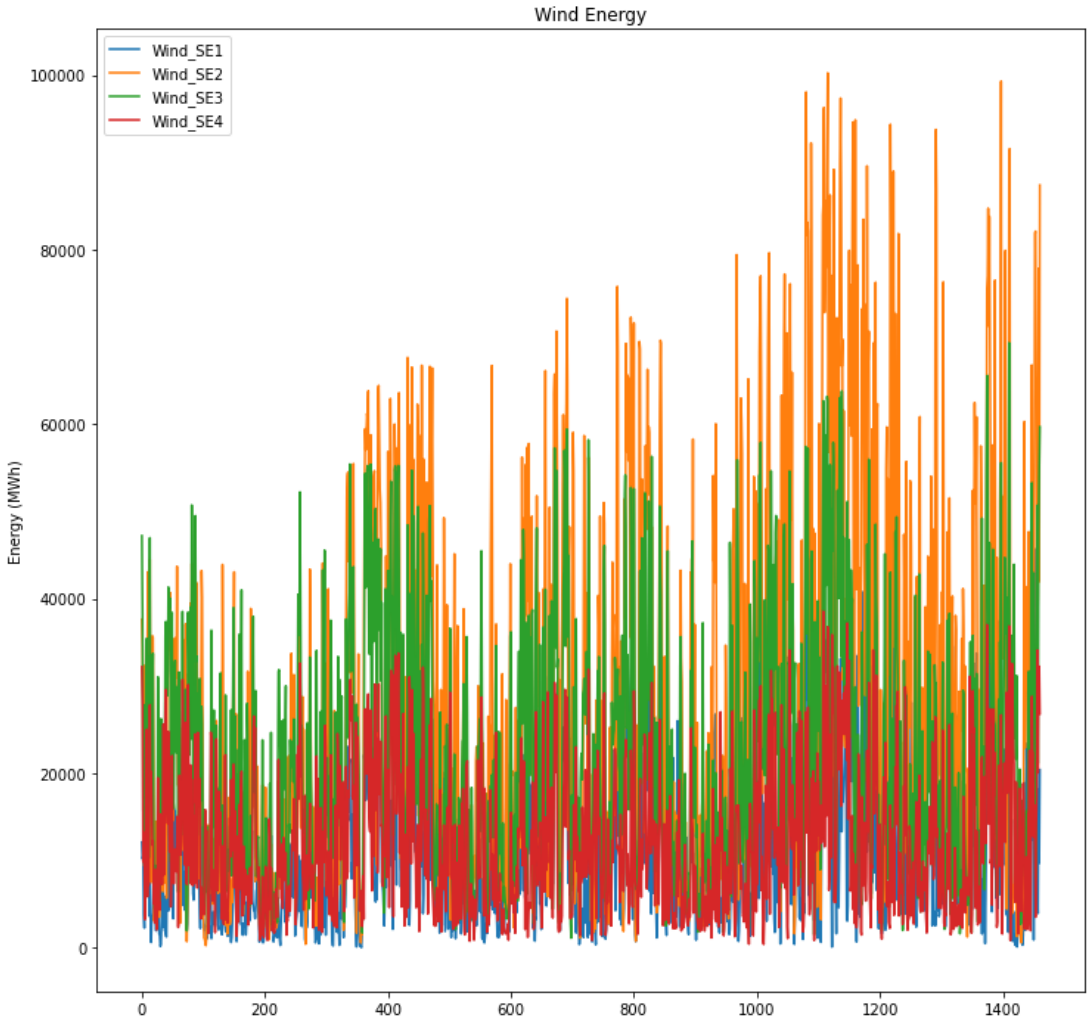


Figure 22: Total daily wind energy production in Sweden partitioned by region.

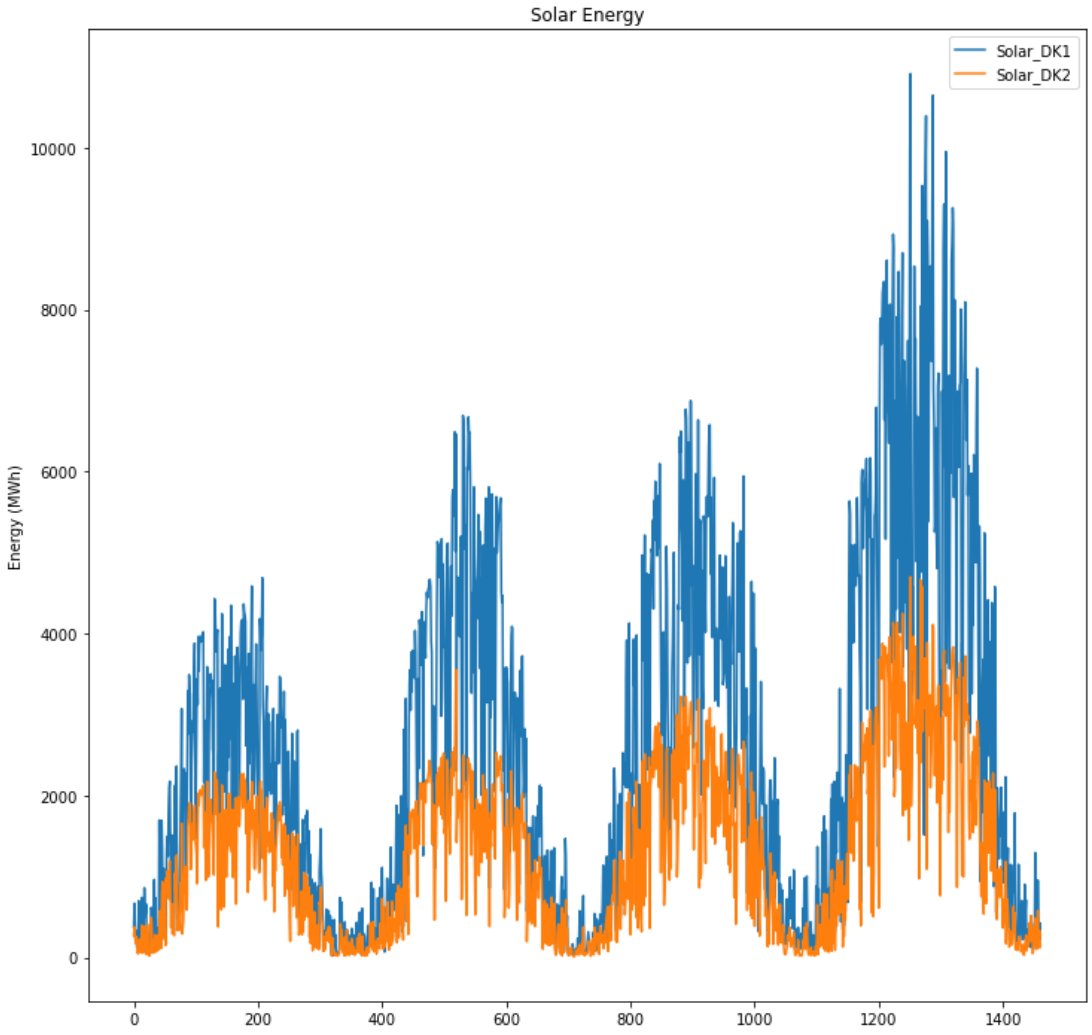


Figure 23: Total daily solar energy production in Denmark partitioned by region.

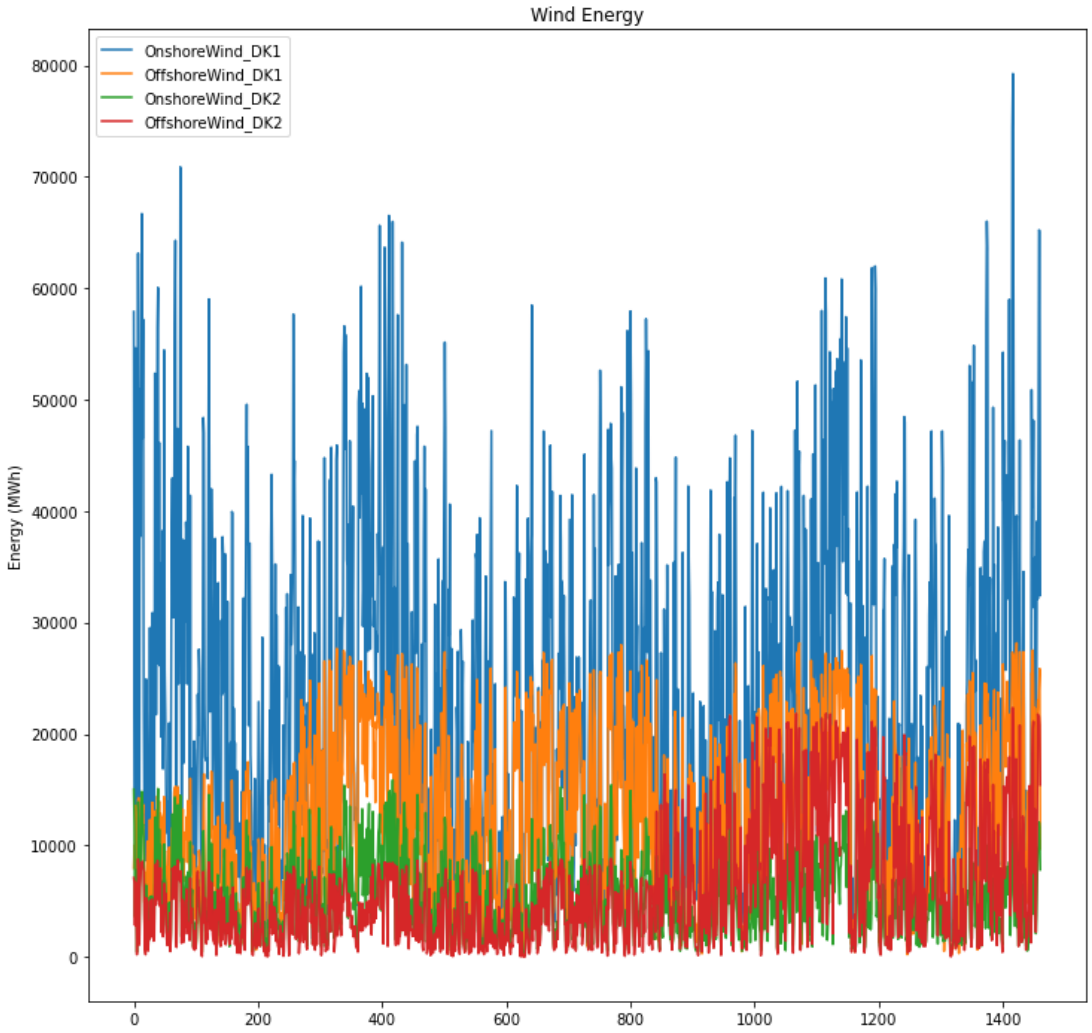


Figure 24: Total daily wind energy production in Denmark partitioned both by region and whether it's onshore or offshore wind turbines.

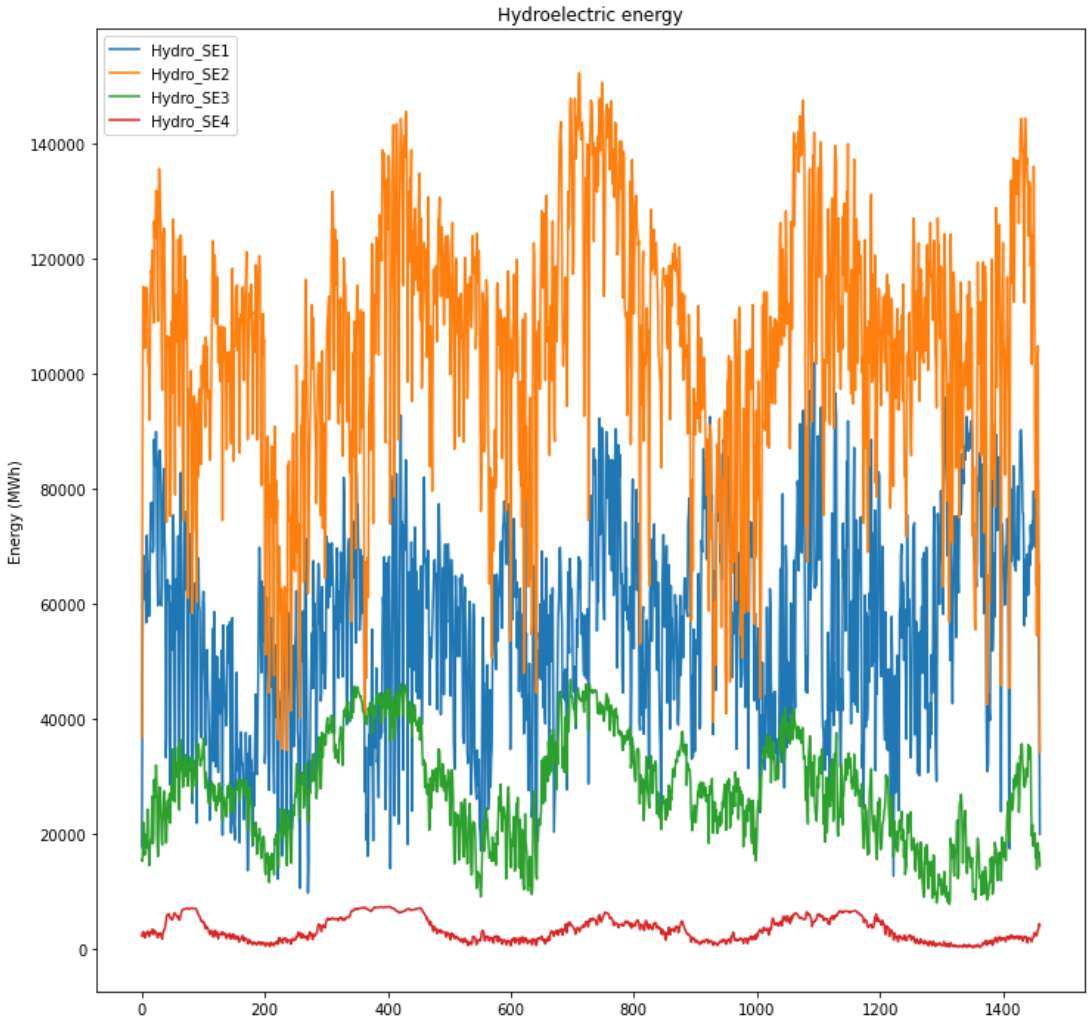


Figure 25: Total daily hydroelectric energy production in Sweden partitioned by region.

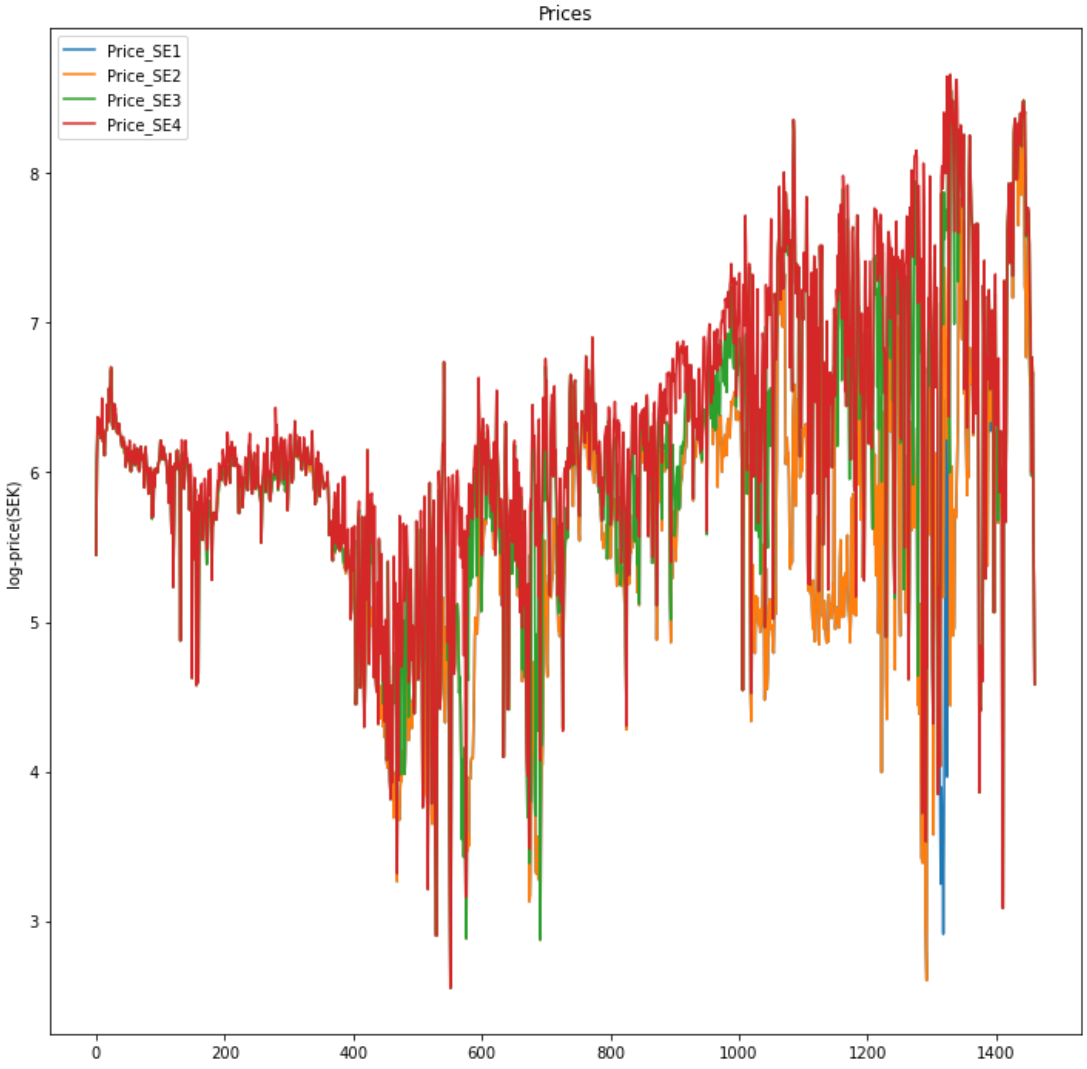


Figure 26: Log-transformed mean daily electricity spot prices.

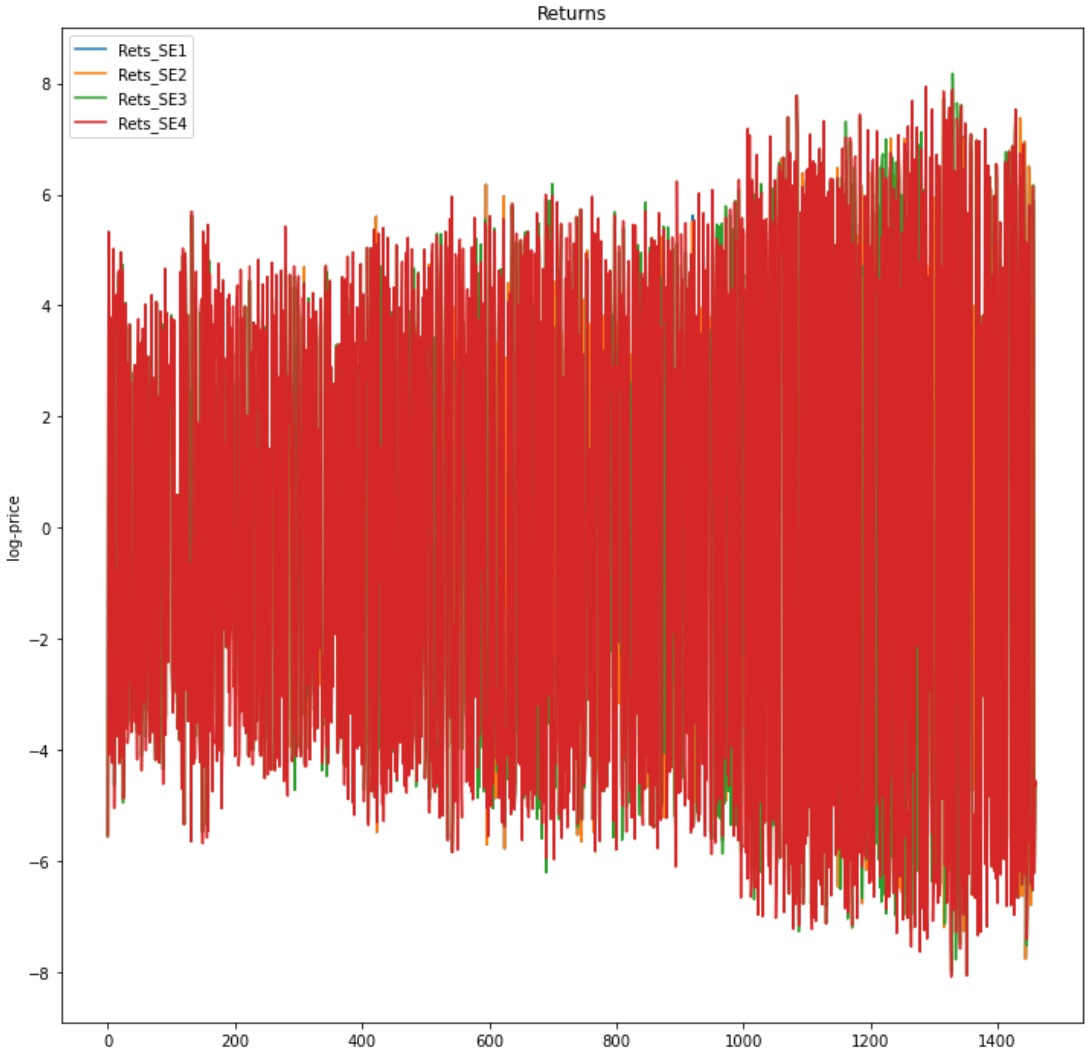


Figure 27: Log-transformed mean daily electricity spot price returns.

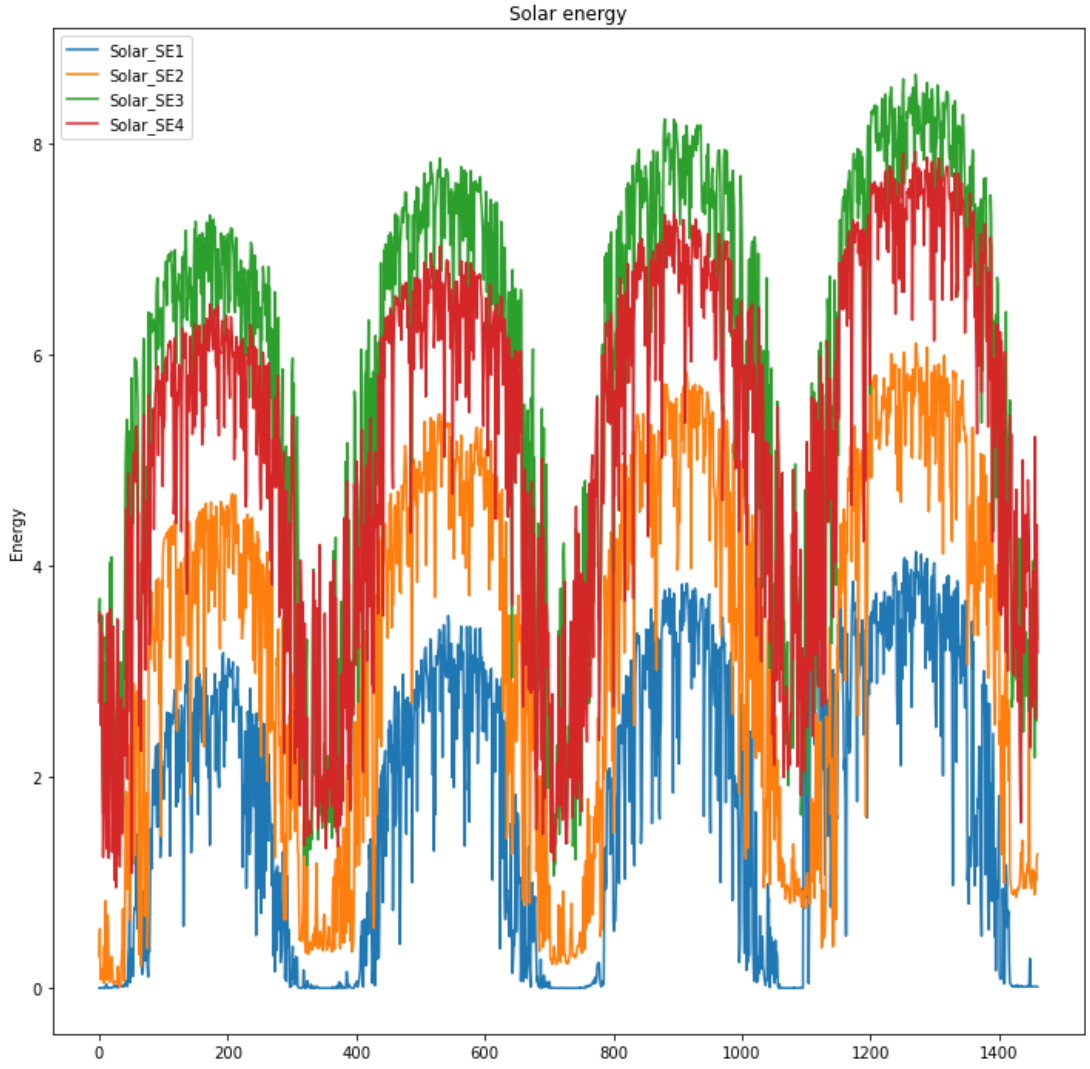


Figure 28: Log-transformed total daily solar energy production in Sweden partitioned by region.

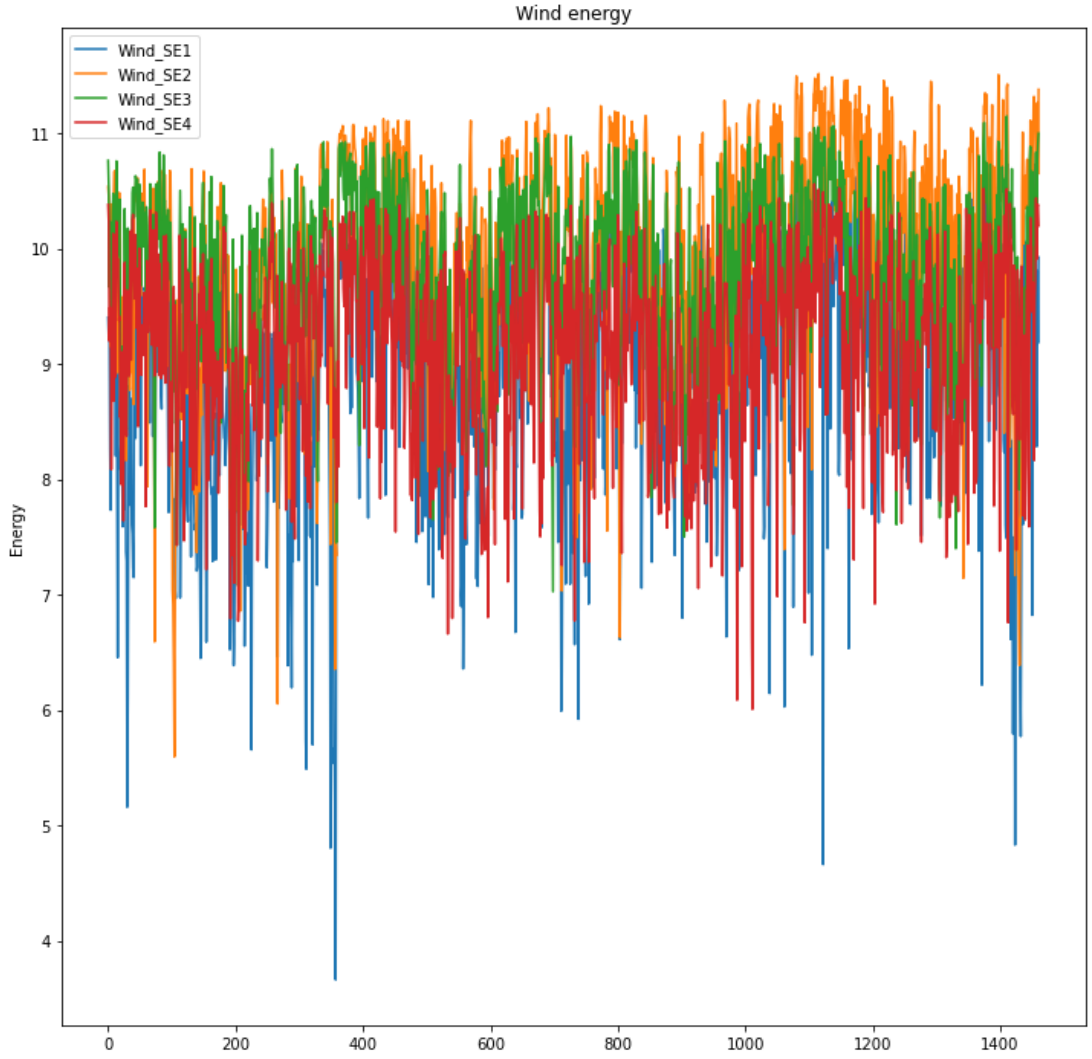


Figure 29: Log-transformed total daily wind energy production in Sweden partitioned by region.

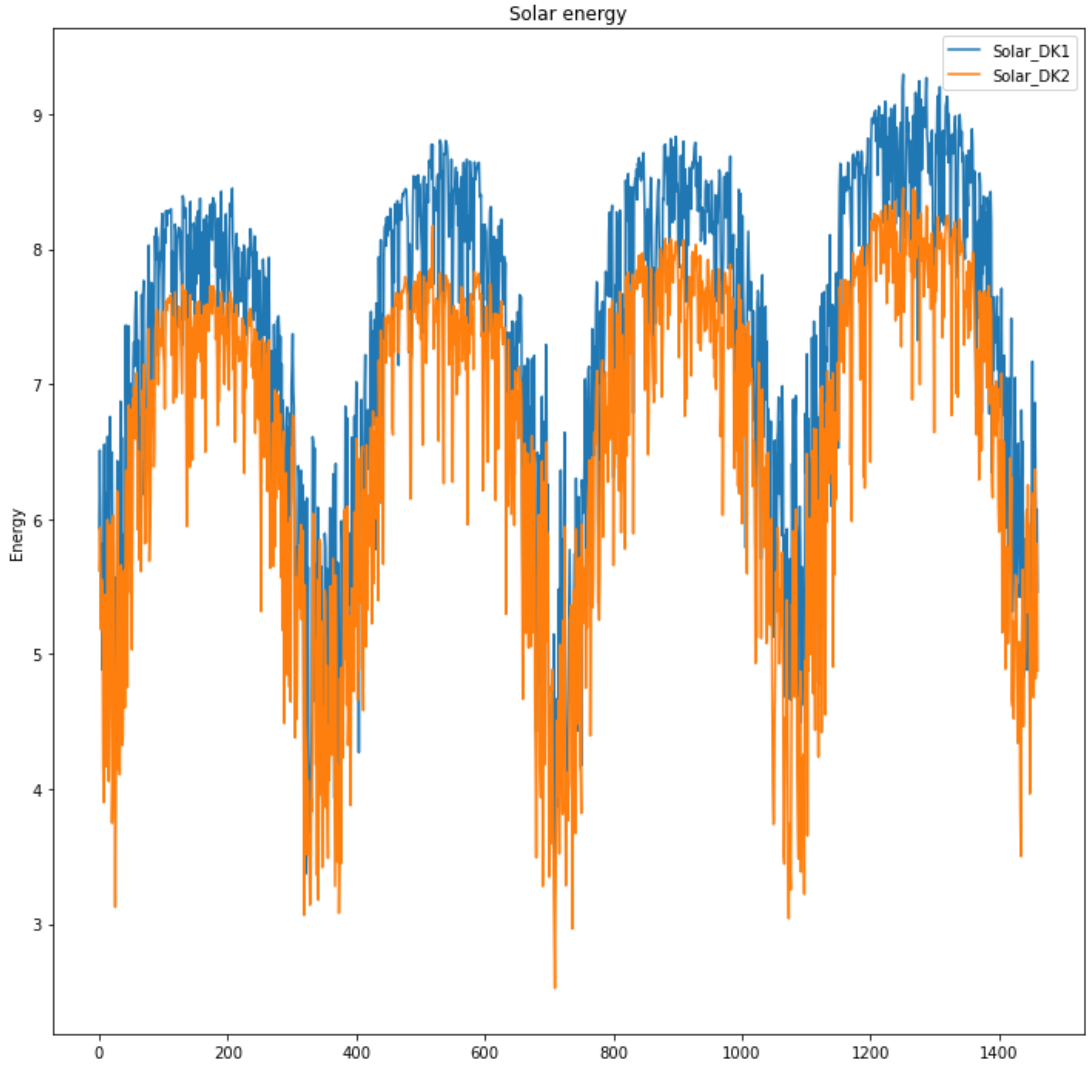


Figure 30: Log-transformed total daily solar energy production in Denmark partitioned by region.

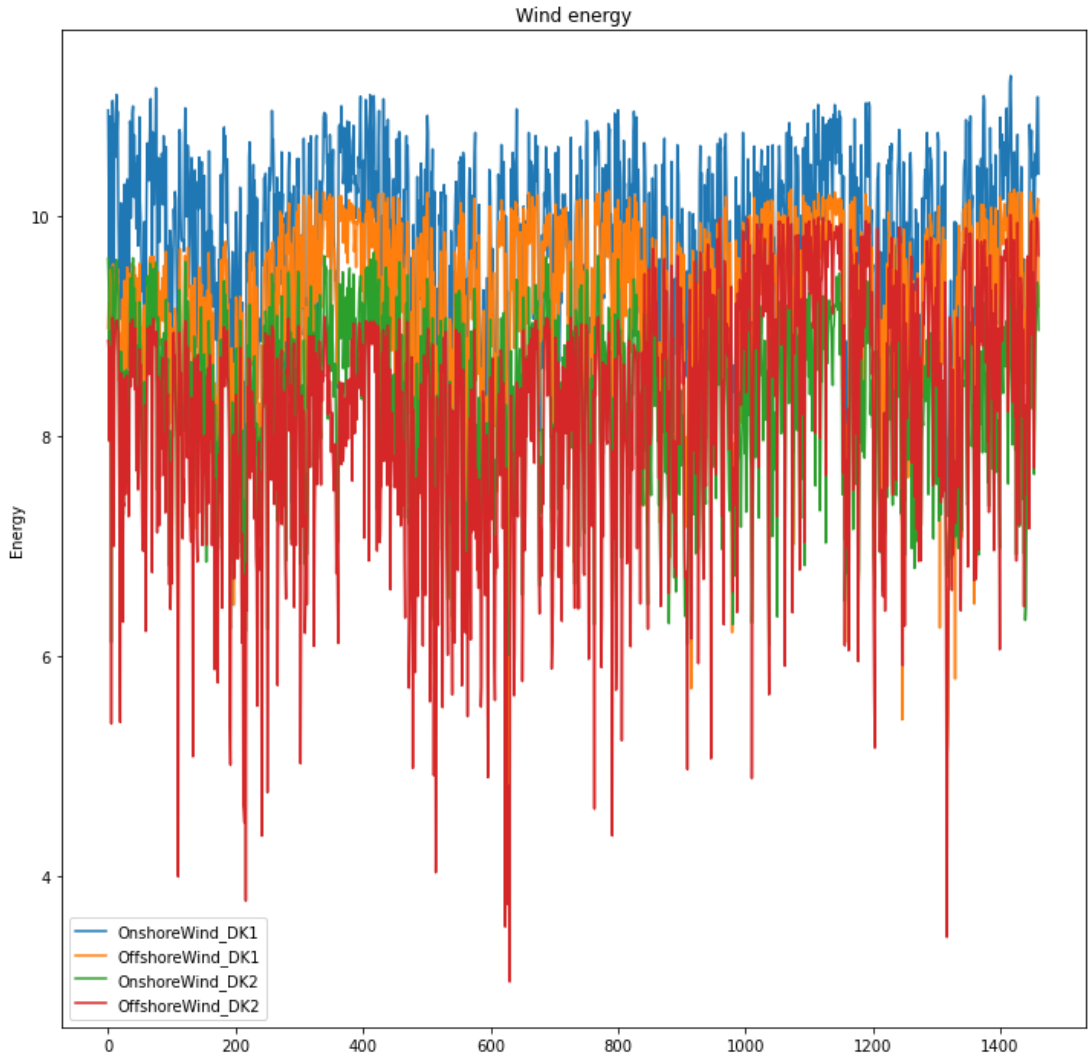


Figure 31: Log-transformed total daily wind energy production in Denmark partitioned by region and whether it's onshore or offshore wind turbines.

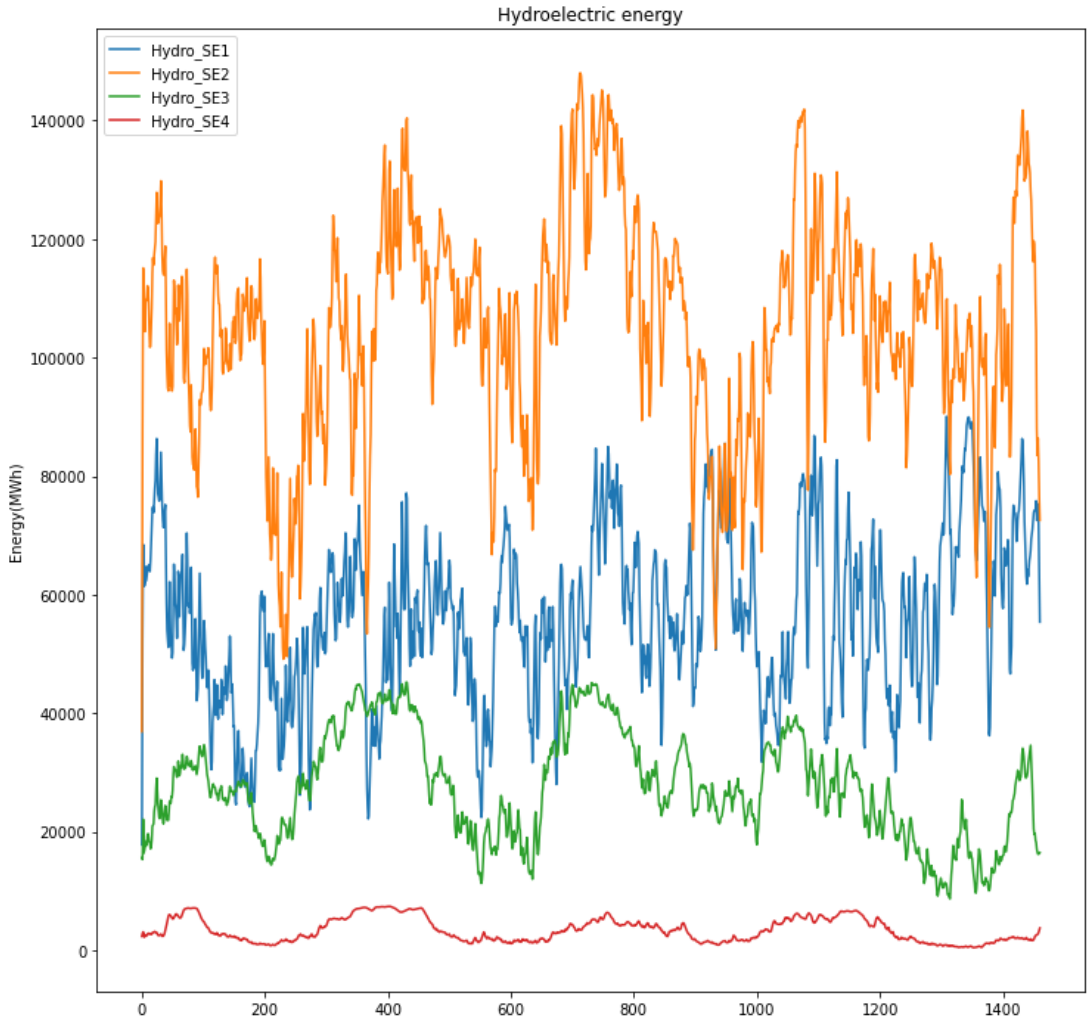
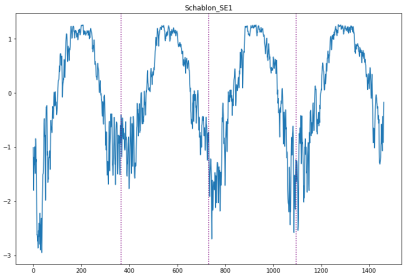


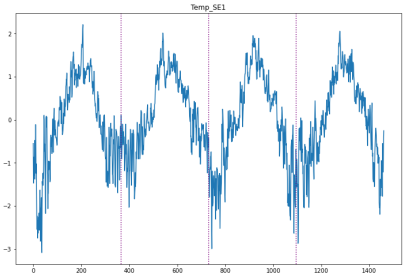
Figure 32: Five-day rolling mean-transformed hydroelectric energy production in Sweden partitioned by region.

8 Appendix B

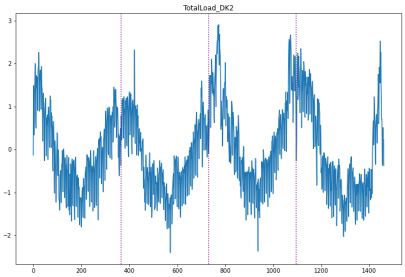
This appendix contains complementary images to the Results section.



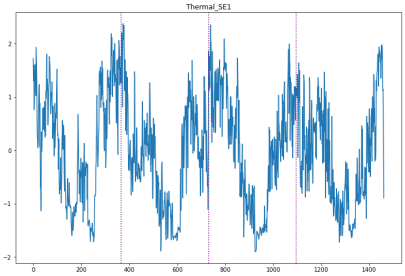
(a) First selected feature of bidding area SE1.



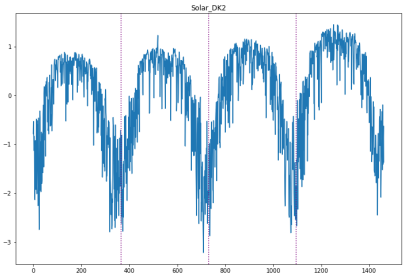
(b) Second selected feature of bidding area SE1.



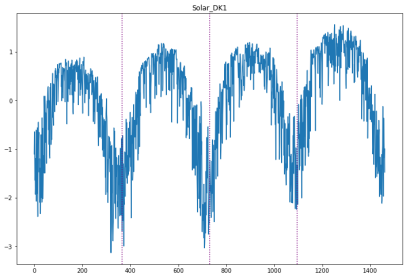
(c) Third selected feature of bidding area SE1.



(d) Fourth selected feature of bidding area SE1.

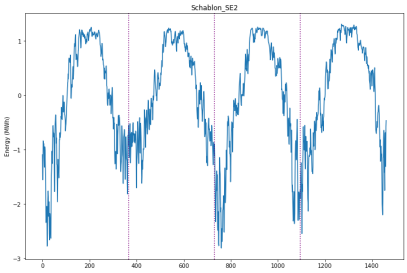


(e) Fifth selected feature of bidding area SE1.

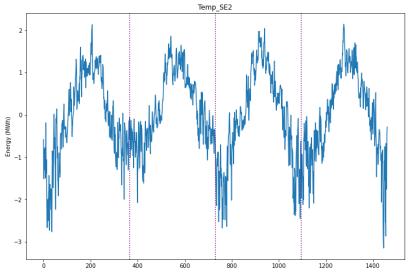


(f) Sixth selected feature of bidding area SE1.

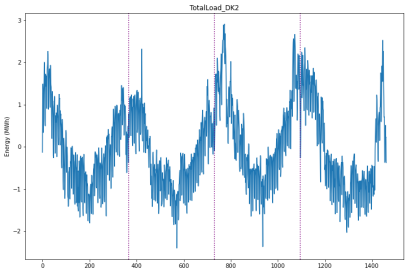
Figure 33: All selected features of bidding area SE1.



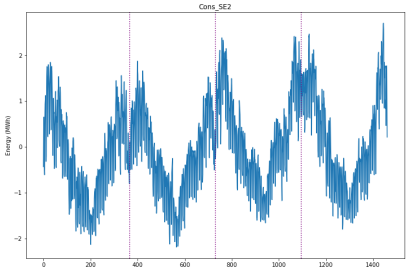
(a) First selected feature of bidding area SE2.



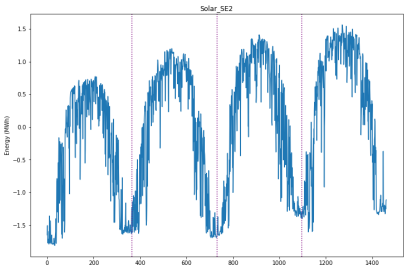
(b) Second selected feature of bidding area SE2.



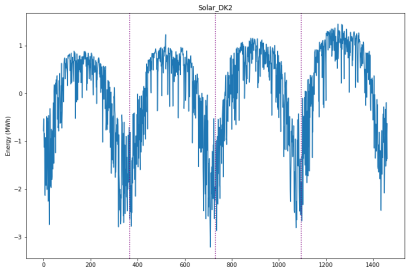
(c) Third selected feature of bidding area SE2.



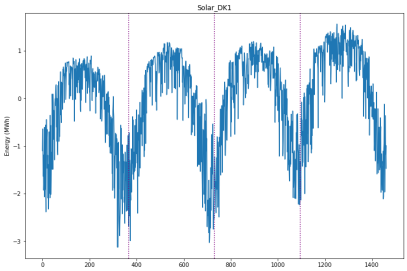
(d) Fourth selected feature of bidding area SE2.



(e) Fifth selected feature of bidding area SE2.

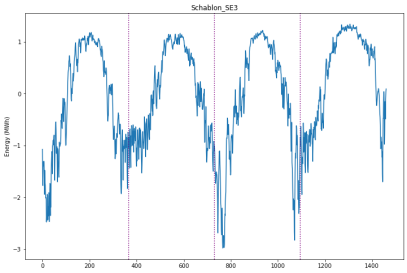


(f) Sixth selected feature of bidding area SE2.

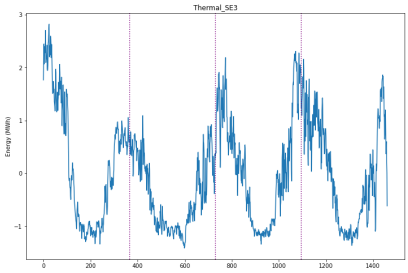


(g) Seventh selected feature of bidding area SE2.

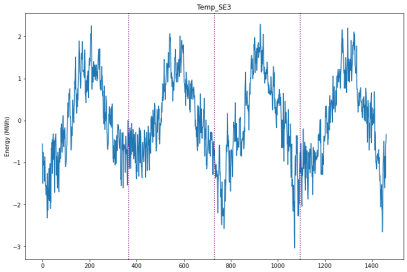
Figure 34: All selected features of bidding area SE2.



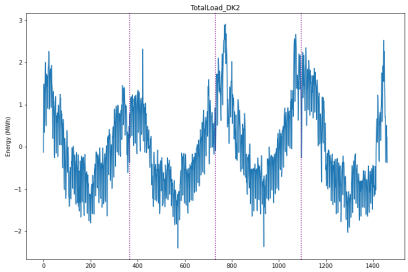
(a) First selected feature of bidding area SE3.



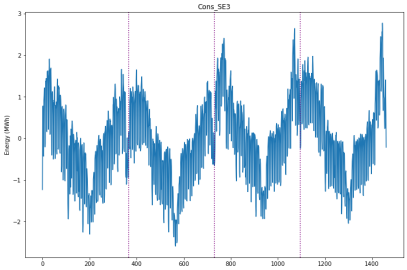
(b) Second selected feature of bidding area SE3.



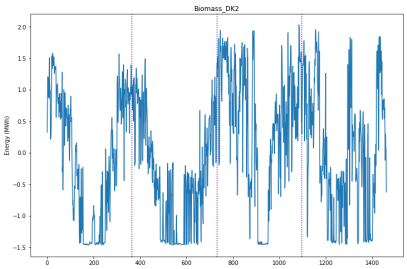
(c) Third selected feature of bidding area SE3.



(d) Fourth selected feature of bidding area SE3.

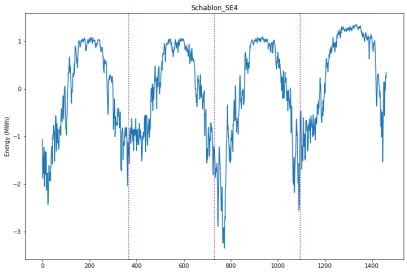


(e) Fifth selected feature of bidding area SE3.

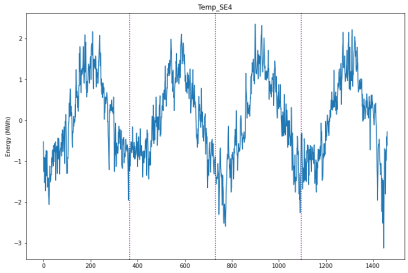


(f) Sixth selected feature of bidding area SE3.

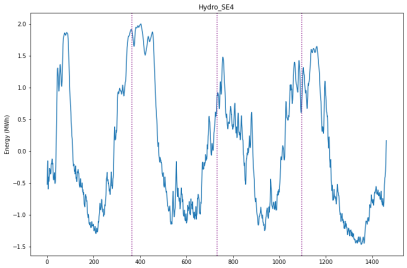
Figure 35: All selected features of bidding area SE3.



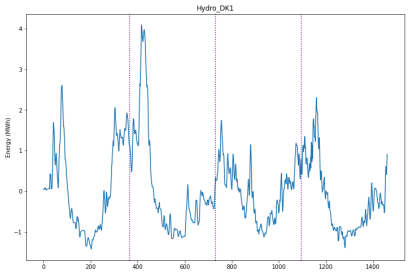
(a) First selected feature of bidding area SE4.



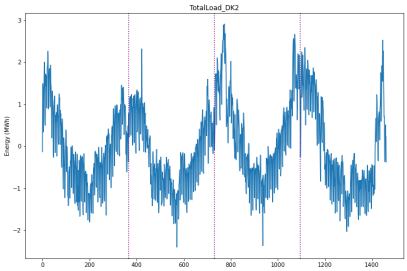
(b) Second selected feature of bidding area SE4.



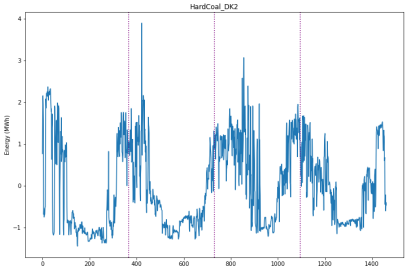
(c) Third selected feature of bidding area SE4.



(d) Fourth selected feature of bidding area SE4.

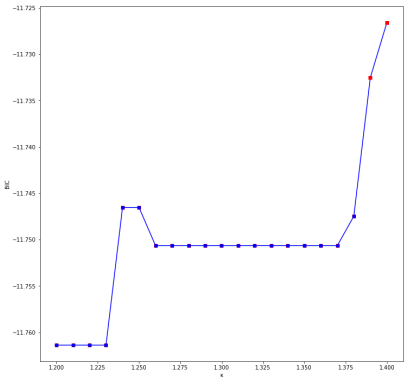


(e) Fifth selected feature of bidding area SE4.

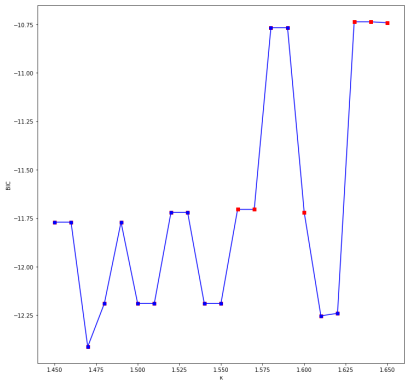


(f) Sixth selected feature of bidding area SE4.

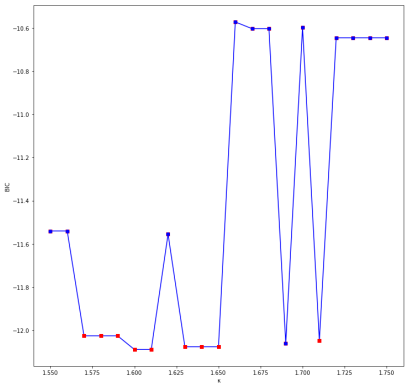
Figure 36: All selected features of bidding area SE4.



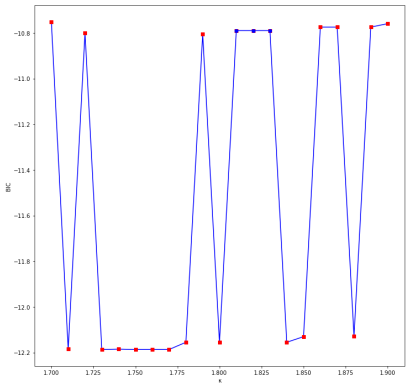
(a) GIC values and tuning parameter values for two output features.



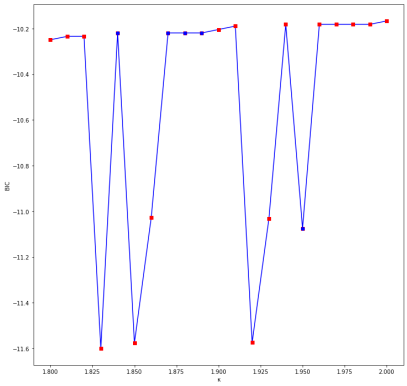
(b) GIC values and tuning parameter values for three output features.



(c) GIC values and tuning parameter values for four output features.

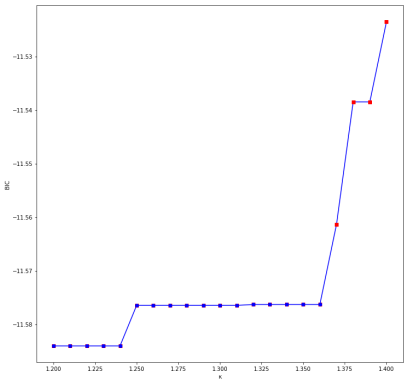


(d) GIC values and tuning parameter values for five output features.

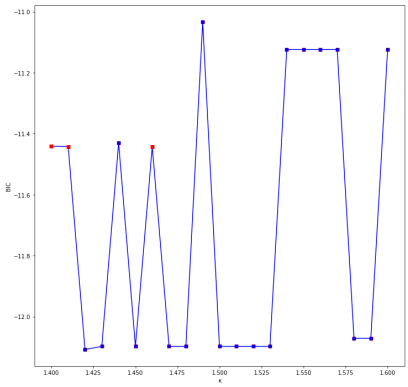


(e) GIC values and tuning parameter values for six output features.

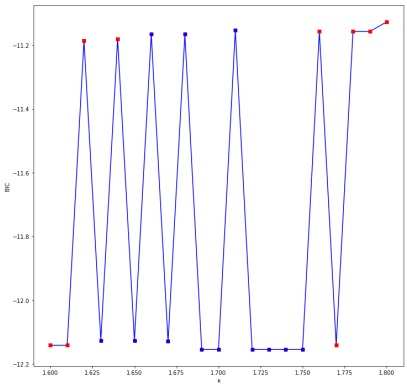
Figure 37: GIC values for different values of tuning parameter κ in region SE1 partitioned by number of desired output features. Red squares indicates tuning parameter values that resulted in incorrect number of output features and were therefore not considered.



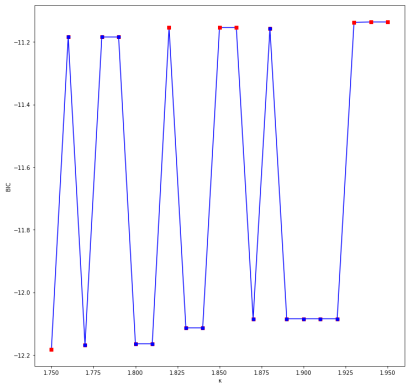
(a) GIC values and tuning parameter values for two output features.



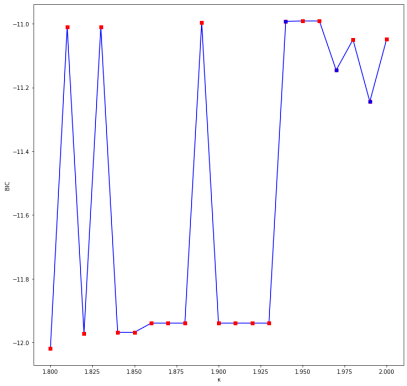
(b) GIC values and tuning parameter values for three output features.



(c) GIC values and tuning parameter values for four output features.

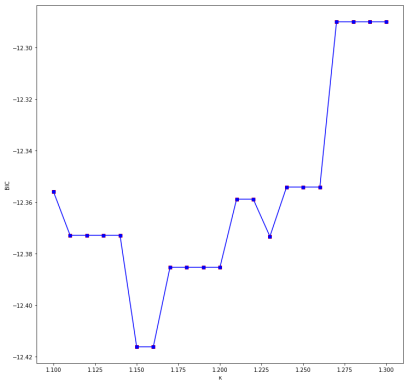


(d) GIC values and tuning parameter values for five output features.

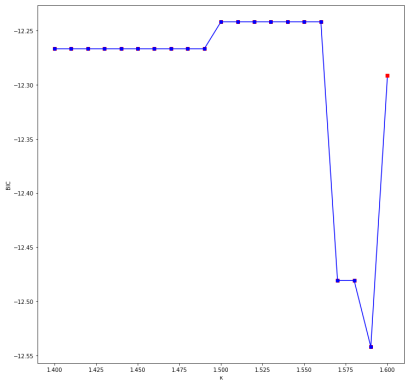


(e) GIC values and tuning parameter values for six output features.

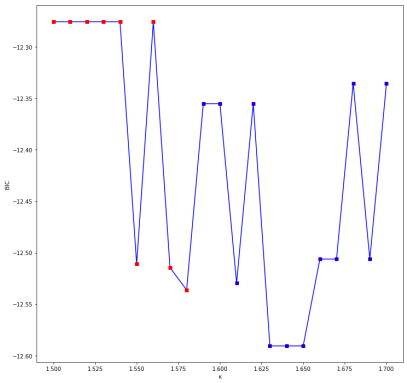
Figure 38: GIC values for different values of tuning parameter κ in region SE2 partitioned by number of desired output features. Red squares indicates tuning parameter values that resulted in incorrect number of output features and were therefore not considered.



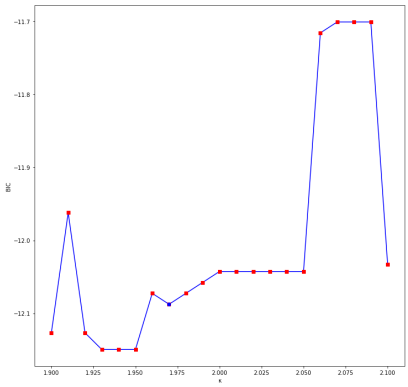
(a) GIC values and tuning parameter values for two output features.



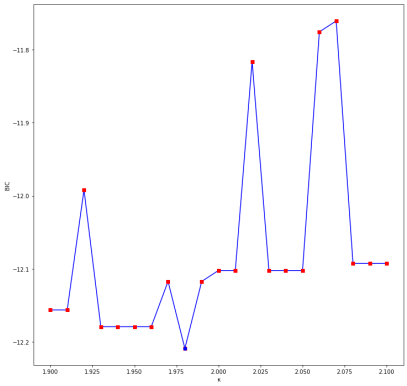
(b) GIC values and tuning parameter values for three output features.



(c) GIC values and tuning parameter values for four output features.

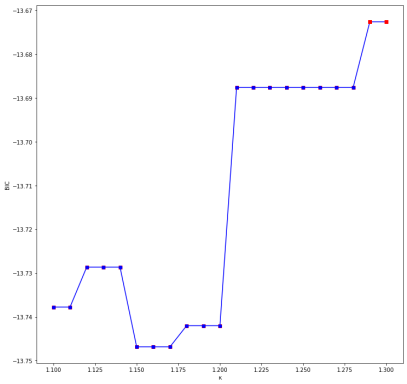


(d) GIC values and tuning parameter values for five output features.

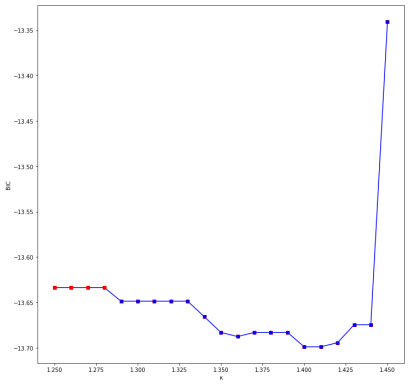


(e) GIC values and tuning parameter values for six output features.

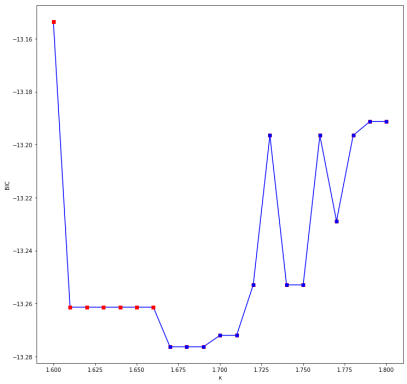
Figure 39: GIC values for different values of tuning parameter κ in region SE3 partitioned by number of desired output features. Red squares indicates tuning parameter values that resulted in incorrect number of output features and were therefore not considered.



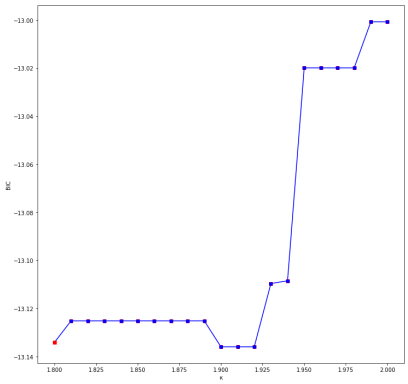
(a) GIC values and tuning parameter values for two output features.



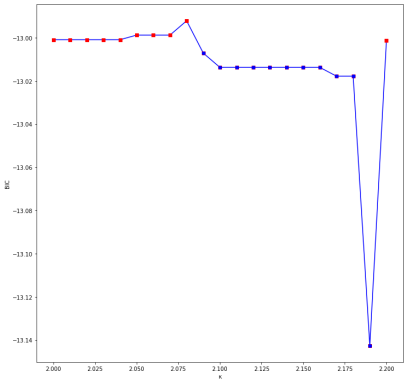
(b) GIC values and tuning parameter values for three output features.



(c) GIC values and tuning parameter values for four output features.



(d) GIC values and tuning parameter values for five output features.



(e) GIC values and tuning parameter values for six output features.

Figure 40: GIC values for different values of tuning parameter κ in region SE4 partitioned by number of desired output features. Red squares indicates tuning parameter values that resulted in incorrect number of output features and were therefore not considered.

9 Appendix C

This appendix contains complementary images to the Discussion section.

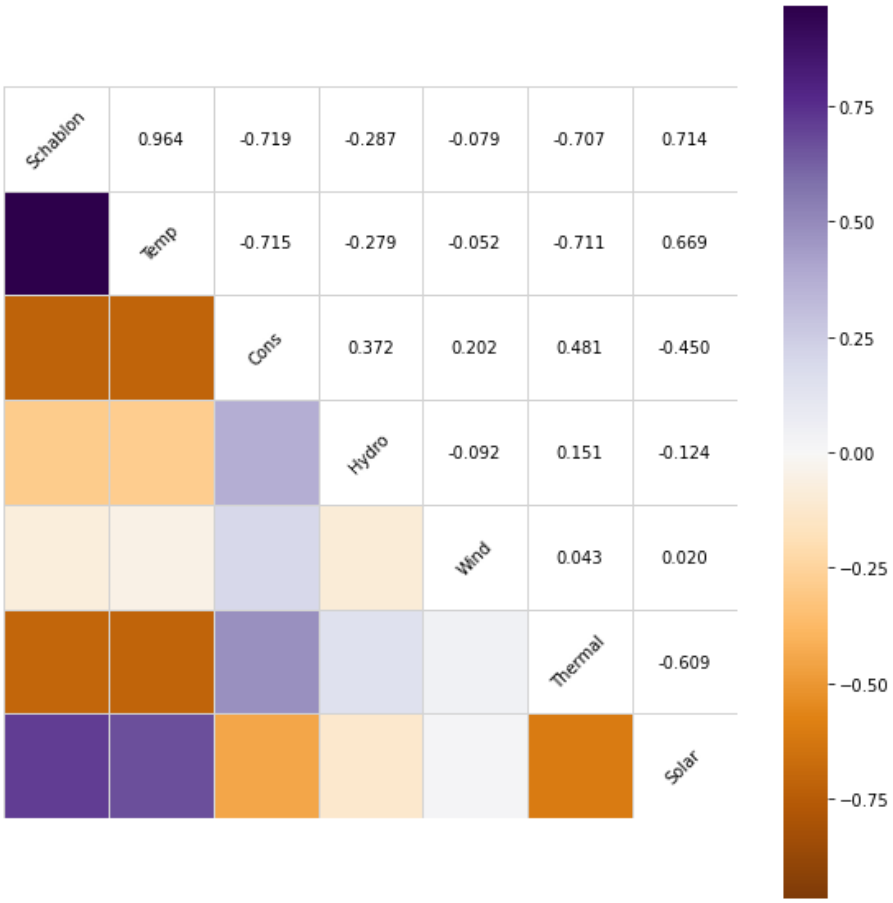


Figure 41: Correlation plot of Schablon and swedish energy features in bidding area 1. On the diagonal we have the features examined. Above the diagonal we have the r-values, and below we have a heatmap of the correlations. From the figure we see that there's either very weak or no correlation between Schablon and wind energy ($|r| \leq 0.3$), a moderate to strong correlation between Schablon and hydroelectric energy ($0.5 \leq |r| \leq 0.7$) and a strong correlation between Schablon and the other considered features ($|r| \geq 0.7$).

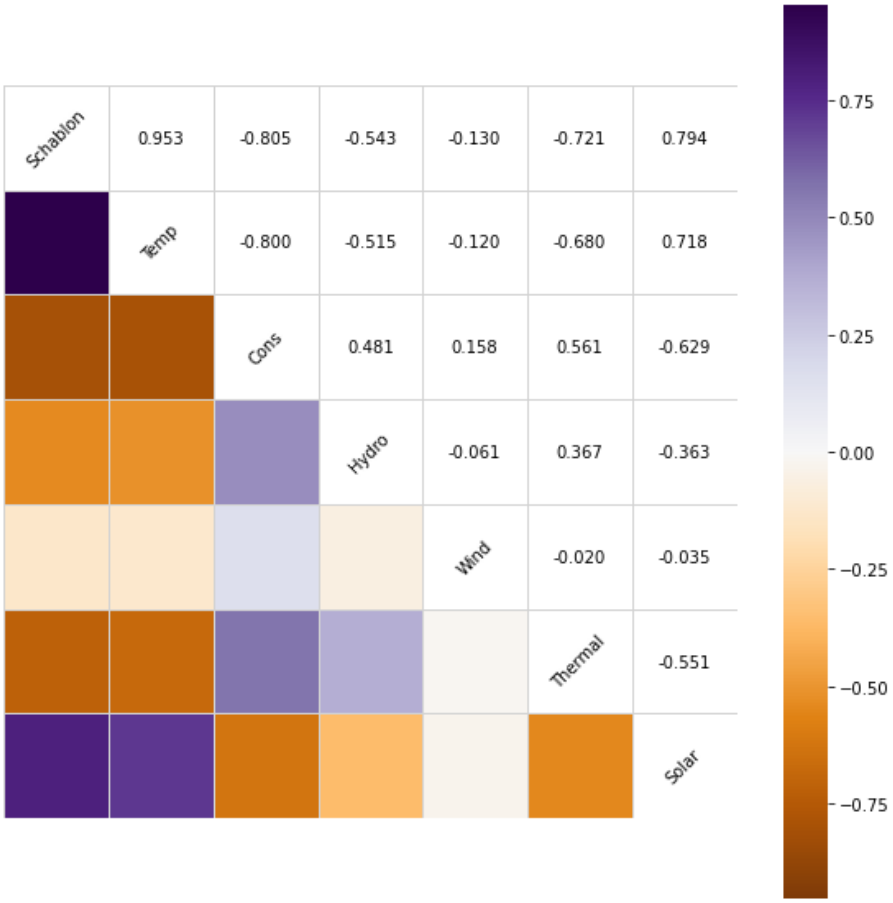


Figure 42: Correlation plot of Schablon and swedish energy features in bidding area 2. On the diagonal we have the features examined. Above the diagonal we have the r-values, and below we have a heatmap of the correlations. From the figure we see that there's either very weak or no correlation between Schablon and wind energy ($|r| \leq 0.3$), a moderate to strong correlation between Schablon and hydroelectric energy ($0.5 \leq |r| \leq 0.7$) and a strong correlation between Schablon and the other considered features ($|r| \geq 0.7$).

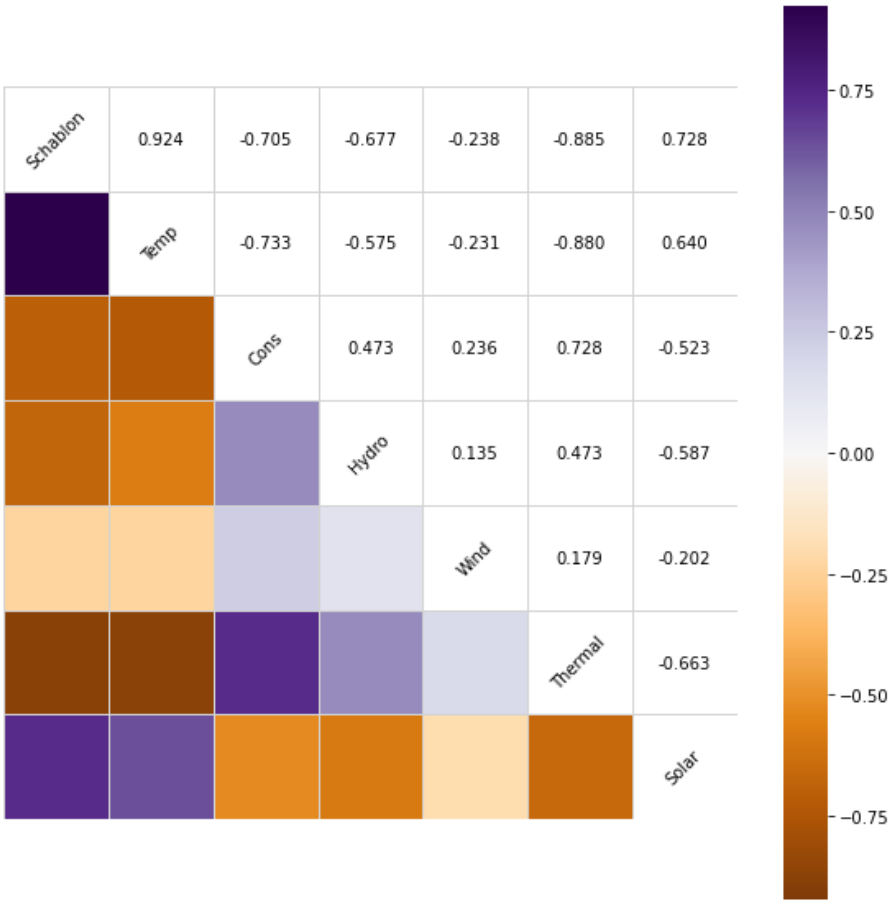


Figure 43: Correlation plot of Schablon and swedish energy features in bidding area 3. On the diagonal we have the features examined. Above the diagonal we have the r-values, and below we have a heatmap of the correlations. From the figure we see that there's either very weak or no correlation between Schablon and wind energy ($|r| \leq 0.3$), a moderate to strong correlation between Schablon and hydroelectric energy ($0.5 \leq |r| \leq 0.7$) and a strong correlation between Schablon and the other considered features ($|r| \geq 0.7$).

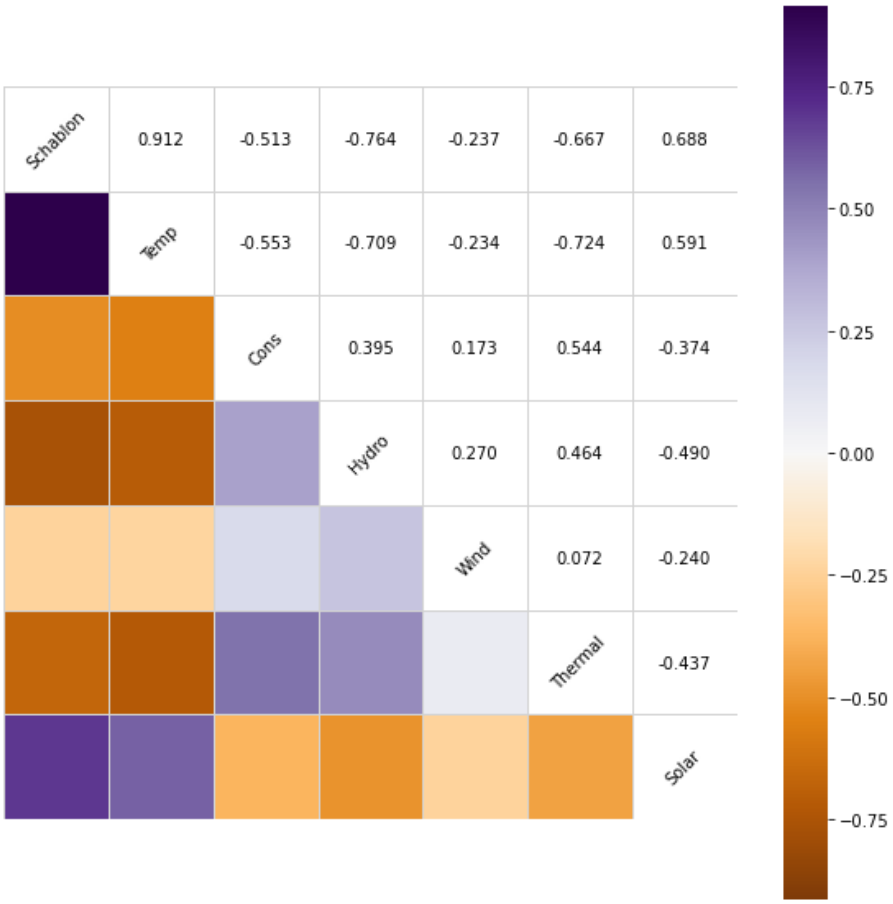
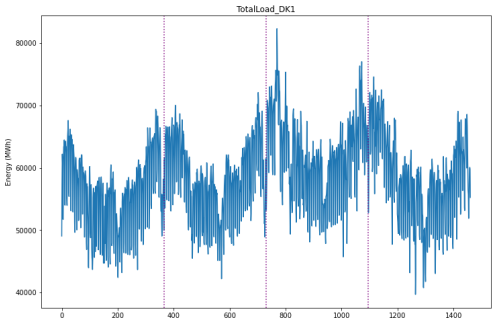
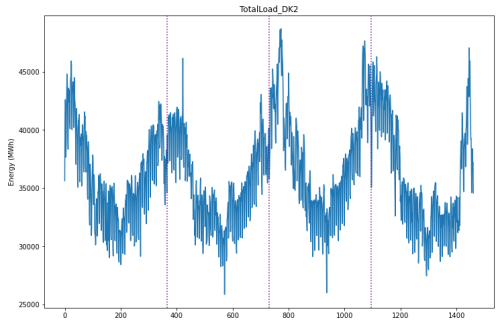


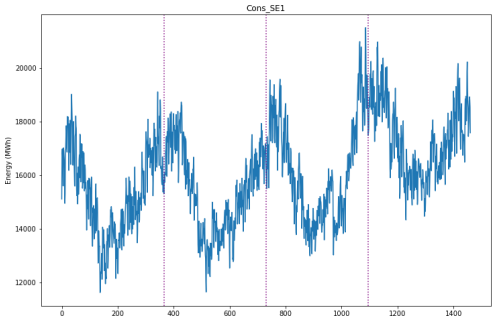
Figure 44: Correlation plot of Schablon and swedish energy features in bidding area 4. On the diagonal we have the features examined. Above the diagonal we have the r-values, and below we have a heatmap of the correlations. From the figure we see that there's either very weak or no correlation between Schablon and wind energy ($|r| \leq 0.3$), a moderate to strong correlation between Schablon and hydroelectric energy ($0.5 \leq |r| \leq 0.7$) and a strong correlation between Schablon and the other considered features ($|r| \geq 0.7$).



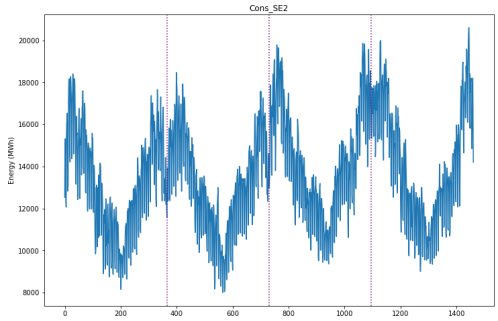
(a) Electricity load in bidding area DK1.



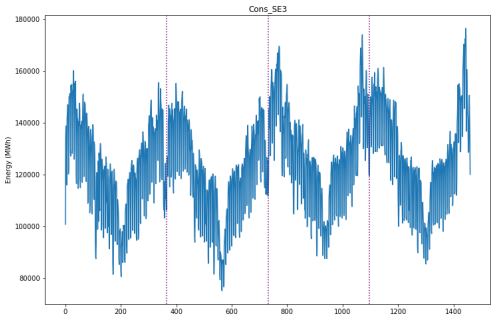
(b) Electricity load in bidding area DK2.



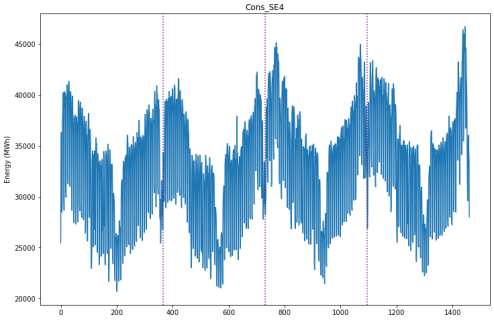
(c) Energy consumption in bidding area SE1.



(d) Energy consumption in bidding area SE2.

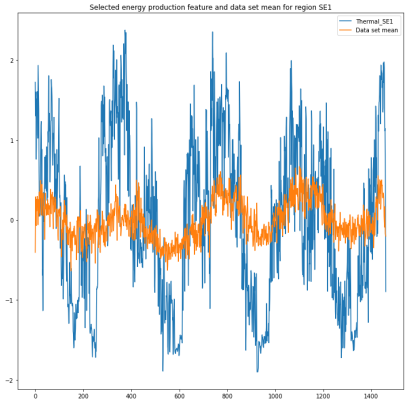


(e) Energy consumption in bidding area SE3.

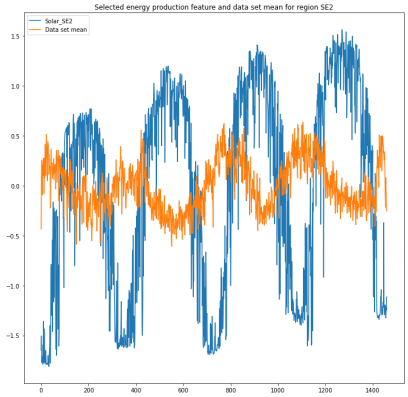


(f) Energy consumption in bidding area SE4.

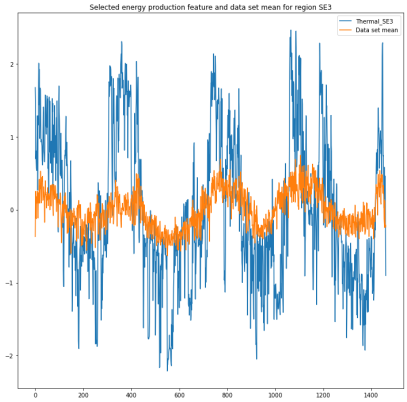
Figure 45: Electricity load of bidding area DK2 and energy consumption of bidding area SE3. Dotted purple lines marks the turn of the year. Shown data is neither transformed nor normalized.



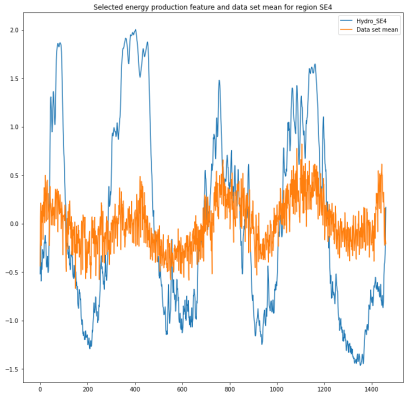
(a) The within-region thermal energy production and the data set mean of bidding area SE1.



(b) The within-region solar energy production and the data set mean of bidding area SE2.



(c) The within-region thermal energy production and the data set mean of bidding area SE3.



(d) The within-region hydroelectric energy production and the data set mean of bidding area SE4.

Figure 46: Selected energy production features of the four regions plotted with their respective data set means.

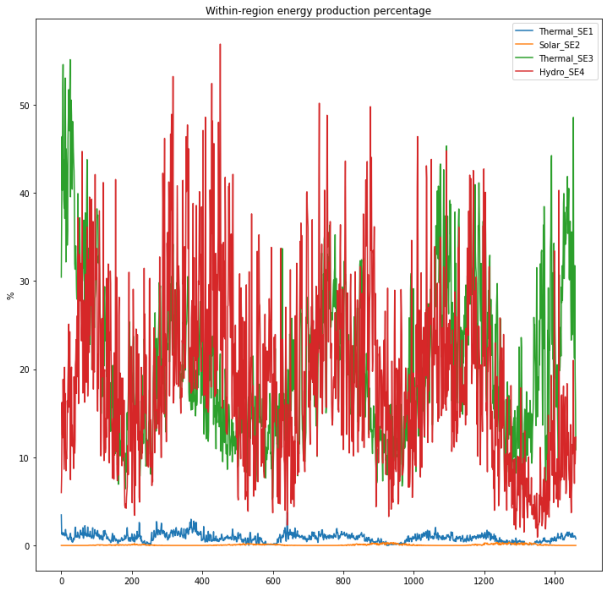
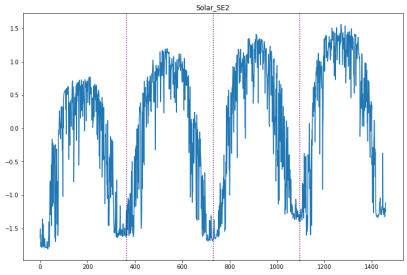
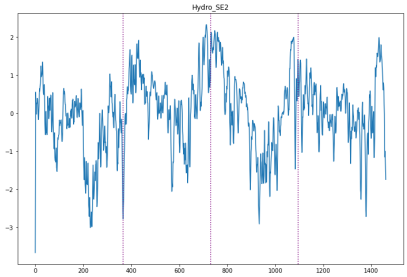


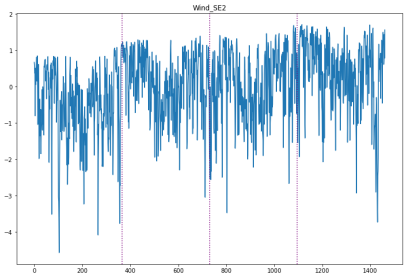
Figure 47: Plots of the selected within-region energy production features plotted as percentage of total within-region energy production



(a) The within-region solar energy production of bidding area SE2.



(b) The within-region hydroelectric energy production of bidding area SE2.



(c) The within-region wind energy production of bidding area SE2.

Figure 48: Transformed and normalized solar, wind and hydroelectric energy production in bidding area SE1.

Master's Theses in Mathematical Sciences 2023:E72
ISSN 1404-6342

LUTFMS-3491-2023

Mathematical Statistics
Centre for Mathematical Sciences
Lund University
Box 118, SE-221 00 Lund, Sweden

http://www.maths.lu.se/