Monitoring Industrial Machine Power Consumption Using Non- and Semiparametric Quantile Regression

Master Thesis Presented by:

Sven Ballentin

(524563)

to

Prof. Dr. Brenda López Cabrera

Prof. Dr. Wolfgang K. Härdle

Ladislaus von Bortkiewicz Chair of Statistics

C.A.S.E. - Center for Applied Statistics and Economics

Humboldt-Universität zu Berlin

in partial fullfillment of the requirements

for the degree of

Master of Economics and Management Science

Berlin, February 13, 2014

Abstract

This work presents an innovative approach on how to design and implement a power consumption monitoring system for industrial machines. The approach bases on the comparison of predictions of a conditional quantile model estimated in a reference period with the realized power consumption in an evaluation period. Three different non- and semiparametric quantile regression models including lagged explanatory variables were adopted. The findings indicate that these models are capable to model the conditional quantile well. Additionally, two simple monitoring warning methods were defined, the first one to detect an increased share of violations retrospectively and the other one to detect an elevated number of consecutive violations in a live-monitoring approach. In an example it could be shown that the warning methods are indeed able to detect an increase in power consumption.

Contents

List of Figures

List of Tables

List of abbreviations

LR	Likelihood ratio
LR-CC test	Likelihood ratio test on conditional coverage
LR-Ind test	Likelihood ratio test on independence
LR-Ind test	Likelihood ratio test on unconditional coverage
ADF	Augmented Dickey Fuller test
EU	European Union
NP	Nonparametric Model
SP-Add	Semiparametric Additive Model
SP-PL	Semiparametric Partial Linear Model

1 Introduction

Germany claims to be one of the leading nations when it comes to increasing energy efficiency. Many of its ambitious goals are stated in the ”Energiekonzept”, e.g. power consumption is to fall by 25% by 2050 compared to 2008 (BMWi (2010)). Of particular importance are energy efficiency gains in the industrial sector since it is the sector with the highest share of energy consumption, 28.97% in 2011 (Eurostat (2014)). Mainly motivated by increasing energy prices, many industrial companies have already realized the need for a more energy efficient production. Nevertheless, in order to incentivize companies further, the European Union (EU) adopted the Directive 2012/27/EU, establishing a common framework of measures for the promotion of energy efficiency (European Parliament and Council of the EU (2012)). On the national German level, one of the principle energy efficiency measures, introduced by the Federal Ministry of Economics and Technology, is the subsidization of energy management systems (Bundesministerium für Wirtschaft und Energie (2013)). A fundamental part of any energy management system is the monitoring of the level of power consumption. Due to wear and tear or changes in the settings, energy efficiency might deteriorate over time. This can lead to substantial economic losses for the operator of the machine and unnecessary carbon dioxide emissions. Especially, large scale cooling machines and air compressors can consume sizable amounts of electricity. Monitoring and detection of states of decreasing energy efficiency can help to reduce costs and prevent the waste of electricity.

The existing literature on the analysis of energy efficiency of machines mainly focuses on measuring the power consumption of different types of machine tools as a basis for identifying optimization potentials. Devoldere et al. (2007) focused on the potential to improve energy efficiency of manufacturing equipment for discrete part production. They investigated the power consumption of a machine tool and classified the activities into productive and non-productive periods. Other research was done on machine monitoring. Behrendt et al. (2012) developed a systematic method to assess power consumption of machines. Hu et al. (2012) described the introduction of an on-line energy efficiency monitoring system that decreases implementation cost. Only few papers try to actually model the power consumption of machines. Dietmair and Verl (2009) introduced a lean and scalable modeling formalism that allows making predictions about the energy efficiency depending on machine design and operation. Draganescu et al. (2003) modeled machine tool efficiency as functions of different machine parameters by using response surface methodology. Thereby, they were able to identify high-efficiency machine settings.

The author is not aware of any research that has been done in the statistical modeling of machine power consumption on a broad level. This thesis aims at developing an approach for the monitoring of industrial machine power consumption. The idea is to compare an estimated conditional quantile of energy consumption based on a reference period with the realized energy consumption in an evaluation period. The goal is to identify periods where the machine's energy efficiency is reduced. In order, to model as many different machine types as possible, flexible models are needed that do not require a specified functional form and that can capture nonlinear dependencies in the variables. I propose nonparametric models, since they can adapt to many different data generating processes without imposing a certain functional form. Nevertheless, they suffer from the curse of dimensionality. For this reason, I further propose semiparametric models that avoid the curse of dimensionality at the cost of loosing flexibility by the imposition of structure. This enables me to use additional explanatory variables, such as lagged ones. In order to establish a benchmark for excessive power consumption, I estimate conditional quantiles. I propose two methods to detect situations where there is an abnormal sequence of violations (values exceeding the estimated quantile). The first one can be used retrospectively to identify an elevated share of violations in the evaluation period and the second one reveals too many consecutive violations in a live monitoring approach. For the evaluation of the different models I propose to use backtesting techniques.

The results of the proposed monitoring approach are encouraging. Visual inspection indicates that the predicted values from the different models capture the movement of the realized power consumption. Model evaluation provides mixed results, since no clear best model can be identified. However, the semiparametric models seem to overall perform better than the nonparametric model, which seems to justify the inclusion of lagged variables at cost of imposing more structure. A first example shows that the proposed warning methods are indeed able to detect an increase in the mean of power consumption. Nonetheless, the approach can not clearly distinguish between an increase in the mean and an increase in the variance.

The remainder of this thesis is organized as follows. Section 2 gives an overview of the methodology. This section consists of a brief overview of the retrospective evaluation period approach and the live-monitoring approach, and a summary of the model requirements. Furthermore, I describe the models and their estimators, the evaluation method, and the monitoring warning methods. In section 3, an empirical example is given. First, the different models are evaluated on data for two different cooling machines. Second, in an example the functionality of the monitoring warning methods is investigated. Section 4 concludes and makes suggestions for further research.

2 Methodology

2.1 Approach

The proposed monitoring can be performed in two different ways, the retrospective evaluation period approach and the live-monitoring approach. A short scheme of the two approaches is given in figure 1.

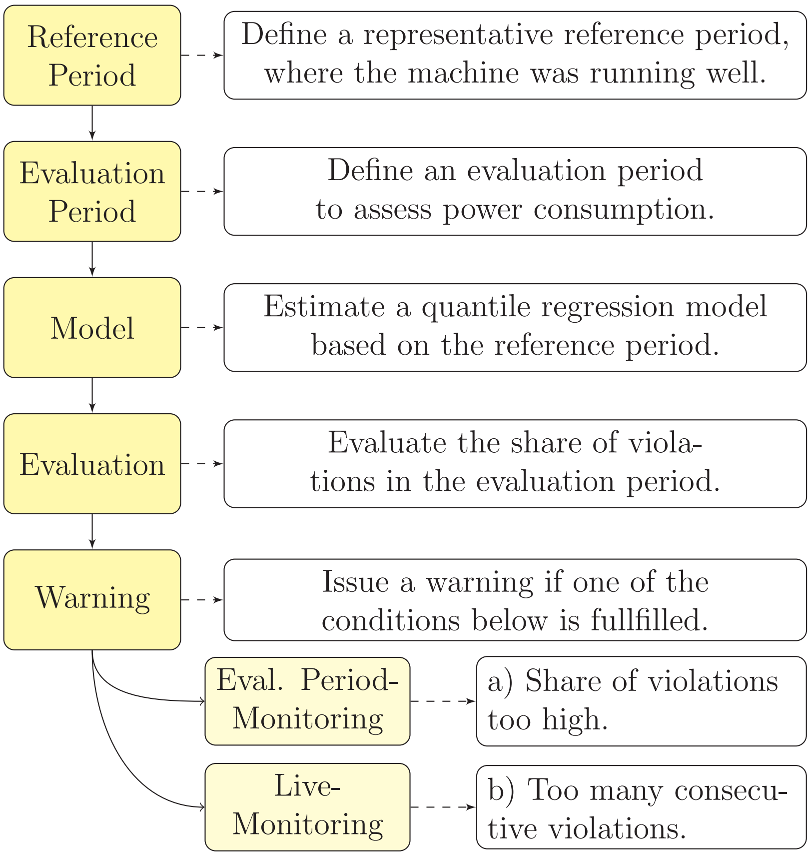


Figure 1: Scheme of the approach.

Both approaches are similar for the first steps. The idea is to model power consumption of a machine in an evaluation period using values from a reference period that is representative for a period where the machine was running well (very energy efficient). First, I choose an adequante reference and evaluation period. In a second step, I find an adequate model based on data from the reference period with T_R observations. Following, I use the model estimated based on the energy efficient reference period to predict values in an evaluation period with T_E observations. By comparing these values to the realized values it can be evaluated if the machine was running as energy efficient as in the reference period. I use quantile regression to establish a benchmark that determines which observations are violations (lying above the estimated conditional quantile) and which are not (below the estimated conditional quantile). An illustration is shown in figure 2. However, one cannot simply take all violations as states of excessive energy consumption, since, e.g. for a 95% quantile, one would expect 5% violations. Therefore, for the first *evaluation period approach* I propose a method that assesses if the share of violations in the evaluation period is too high and issues a warning in that case. This approach is backward looking since an already passed evaluation period is assessed.

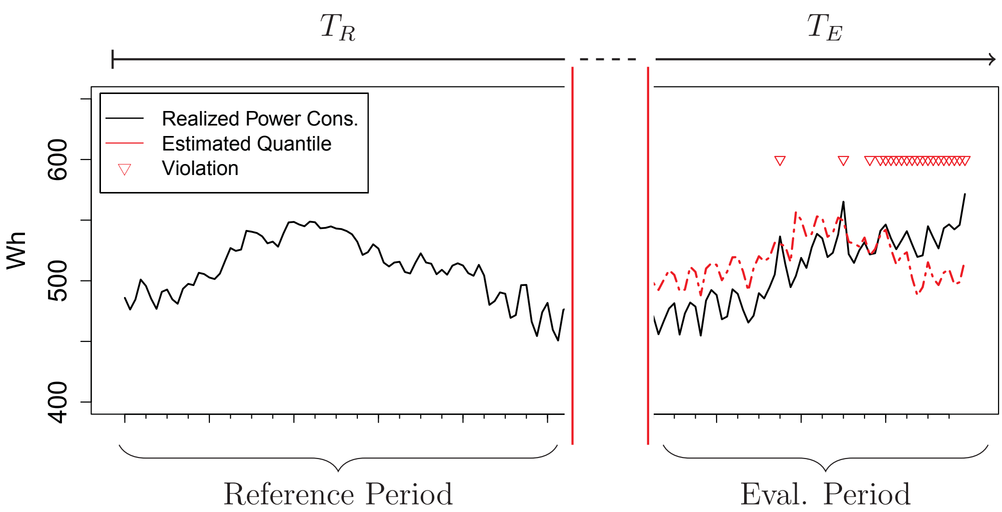


Figure 2: Illustration of the approach.

The *live monitoring approach* differs slightly in its structure. The quantile regression model is still estimated based on the reference period but no fixed evaluation period is defined. Instead, one estimates the conditional quantile for each new incoming (live) observation. To assess power consumption, I propose a method that bases on the assumption of independent violations. Whenever, a certain number of consecutive violations is observed, the monitoring system issues a warning that power consumption might be elevated.

2.2 General model requirements

As mentioned before, the goal of this thesis is to develop a monitoring system that is fast and easy to implement and applicable to many different machine types. This leads to certain model requirements. Since, it would be very time consuming to specify a functional form for each machine, the used quantile regression model has to be able to capture the dynamics of power consumption without specification of such a form. Furthermore, given the different machine types, it is very likely that there exist nonlinear dependencies between the power consumption and the explanatory variables. The model has to be able to capture this. Moreover, there might exist time dependencies in the data, since for example cooling machines exhibit inertia due to self-cooling and after-cooling effects, e.g. the machine first has to cool itself down before providing a certain cooling power. Finally, for some machines we might need to use a model with many explanatory variables (e.g. including lagged explanatory variables).

Non- and semiparametric models provide the required flexibility and are able to model nonlinear dependencies. Furthermore, under certain assumption they are able to model dependent data. Purely nonparametric models do not need any specification of the functional form. However, due to the curse of dimensionality, it is not possible to include many exogeneous variables. Semiparametric models, such as partial linear regression models or additive models, do not suffer from the curse of dimensionality but are not as flexible as purely nonparametric models. Both models are advantageous in certain situations, therefore I will consider semi- and nonparametric quantile regression models.

2.3 Nonparametric quantile regression - The local linear quantile estimator

Parametric regression models are widely used in research and provide an efficient way to estimate models. There exist various nonparametric regression models. Among the most common ones are kernel regression models such as the Nadaraya-Watson estimator introduced by Nadaraya (1964) and Watson (1964), the local linear estimator and the local polynomial estimator (see Fan (1996)). Note that the first two estimators are special cases of the local polynomial estimator. In the following analysis I use the local linear approach since, in contrast to the local constant estimator, the local linear fit can improve the function estimation in regions with sparse observations and in boundary regions. The boundary effect might be quite substantial, especially in the multivariate setting, for which the boundary can include a large amount of data points (see Fan (1996)). Furthermore, I will not use polynomials of higher order since this increases variability of the estimate and also increases computational time.

2.3.1 The local linear quantile estimator

One can define a local linear quantile estimator as $\hat{m}_\alpha(x_t) = \hat{a}$, where $\hat{a}$ and $\hat{b}$ minimize

$$\sum_{s=1}^{T_R} \rho_\alpha \left\{ Y_t - a - b^T(X_s - x_t) \right\} \mathcal{K}\left(\frac{x_t - X_s}{h} \right), \tag{1}$$

where $\mathcal{K}\left(\frac{x_t - X_s}{h}\right) = \mathrm{k}\left(\frac{x_t^1 - X_s^1}{h_1}\right) \times \cdots \times \mathrm{k}\left(\frac{x_t^d - X_s^d}{h_d}\right)$ is a product kernel, where X_t^j with $j = 1, \ldots, d$ is the j-th component of $X_t \in \mathbb{R}^d$, $h = (h_1, \ldots, h_d)$ is a vector of bandwidths and ρ_α is the check function given as

$$\rho_\alpha(z) = \alpha z \mathbb{1}_{[0,\infty)}(z) - (1 - \alpha) z \mathbb{1}_{(-\infty,0)}(z). \tag{2}$$

2.3.2 Bandwidth selection

The correct choice of the bandwidth is of great importance, since the bandwidth controls the amount of local smoothing in the local linear quantile regression and thus balances between local curvature and stochastic variability. In the literature there exist a few methods to choose the bandwidth in a local quantile regression setting. Fan (1996) developed an approach that is derived from the asymptotically optimal bandwidth for a univariate quantile function, replacing the unknown terms by some adequate estimates. Nevertheless, it is not appropriate for our purpose, since it is only applicable on univariate regression. Cai and Xu (2008) followed a different approach by using the nonparametric version of the Akaike information criterion for the estimation of dynamic smooth coefficient models. Unfortunately, they only show how to apply this approach in the univariate case and do not give a theoretical basis for the proposed methodology. Spokoiny et al. (2011) developed a more sophisticated method by introducing a data-driven locally adaptive bandwidth selection procedure. Yu and Jones (1998) proposed a

rule of thumb bandwidth selection method that is based on the transformation of the optimal bandwidth of the local mean regression (e.g. selected by cross-validation). In the derivation of their method they assume the quantile functions are parallel. However, this assumption might be too restrictive for many situations due to heteroscedasticity.
Given the fast and easy implementation of the last-mentioned method, I use the rule-of-thumb method proposed by Yu and Jones (1998). Their bandwidth selection strategy can be summarized as follows:

1. Use method to select the opt. bandwidth for the mean-regression h_m.
2. Transform h_m using, $h_\alpha = h_m \left\{\alpha(1-\alpha)/\phi(\Phi^{-1}(p))^2\right\}^{1/5}$.

There exist various methods to select the bandwidth h_m for the mean-regression

$$Y_t = m(X_t) + \varepsilon_t, \tag{3}$$

where ε_t is a r.v. with mean zero and $m(X_t)$ an unknown function. One method that is widely used in applications is the data-driven least squares cross-validation. For this approach one minimizes the following function by choosing $h_1, \dots, h_d$

$$CV_{ll}(h_1, \dots, h_q) = T_R^{-1} \sum_{t=1}^{T_R} (Y_t - \hat{m}_{-t}(x_t))^2, \tag{4}$$

where $\hat{m}_{-t}(x_t)$ is the local linear estimator with bandwidth $h = h_1, \dots, h_d$ without using the t-th observation (leave-one-out estimator). One has to use the leave-one out estimator avoid the problem of predicting Y_t with itself in $g(X_t)$.[1] One can show that bandwidth parameters $\hat{h}$ selected via

[1] One could make $CV_{ll}(h_1, \dots, h_q)$ arbitrarily small by letting $(h_1, \dots, h_d) \to \mathbf{0}_{\boldsymbol{q}}$. That would be equivalent to explaining Y_t with itself.

cross-validation are asymptotically equivalent to the deterministic optimal bandwidth parameters h (see Li and Racine (2007) for details). To find the optimal bandwidth, one can use standard numerical optimization procedures (as for instance the Nelder-Mead method explained in Dennis and Woods (1987)) to find the global minimum of function (4).

2.4 Semiparametric additive quantile regression - A two-stage estimator

As already mentioned, purely nonparametric models are often unattractive in settings with many exogenous variables because of the curse of dimensionality. Semiparametric models do not suffer from this problem because they achieve dimension reduction in one way or another. One class of semiparametric models are additive models. They allow for a component-wise analysis and combine flexible nonparametric modelling of multidimensional inputs with the statistical precision of a univariate nonparametric analysis. Nevertheless, they are not as flexible as purely nonparametric models, since an additive structure is imposed.

2.4.1 The semiparametric additive quantile regression model

A semiparametric additive quantile regression model can be written as

$$(5) \qquad Y_t = \mu_\alpha + m_{1,\alpha}(X_t^1) + \cdots + m_{d,\alpha}(X_t^d) + u_{t,\alpha}, \quad t = 1, \ldots, T_R,$$

where α is the quantile, $Y_t \in \mathbb{R}$, X_t^j with $(j = 1, \ldots, d)$ is the j-th component of $X_t \in \mathbb{R}^d$, $\{(Y_t, X_t) : t = 1, \ldots, T_R\}$ is assumed to be a stationary and α-mixing process (see 2.6), μ_α is an unknown constant, $m_{1,\alpha}(\cdot), \ldots, m_{d,\alpha}(\cdot)$ are unknown real-valued univariate functions and $u_{t,\alpha}$ is an unobserved r.v. with $Q_\alpha(u_{t,\alpha}|X_t = x) = (F^{-1}_{u_{t,\alpha}}(\alpha)|X_t = x) = 0$.

2.4.2 The two-stage estimator

In the literature there exist basically three methods to estimate the conditional quantile model in equation (5): splines, backfitting, and marginal integration. Doksum and Koo (2000) consider a spline estimators but do not provide pointwise rates of convergence or an asymptotic distribution, making inference difficult. De Gooijer and Zerom (2003) developed a marginal integration estimator of the model (5). The estimator is asymptotically normal and, thus, makes inference relatively easy. Nevertheless, the proposed marginal integration estimator begins with an unrestricted, d-dimensional, nonparametric quantile regression and, therefore, this estimator suffers also from the curse of dimensionality. Horowitz and Lee (2005) developed an estimator that is asymptotically normally distributed and avoids the curse of dimensionality. The basic idea of the estimator is to do a quantile series approximation in the first stage (e.g. based on B-splines) and then, in the second stage, use a marginal local polynomial quantile estimator.

In the following, I describe the two-stage estimator by Horowitz and Lee (2005). For any $x \in \mathbb{R}^d$, define $m(x) = m_1(x^1) + \cdots + m_d(x^d)$, where x^j is the j-th component of x. Let us assume that the support of X is $\mathcal{X} \equiv [-1, 1]$ and, without loss of generality (since the location of m_j is not identified), normalize $m_{1,\alpha}, \ldots, m_{d,\alpha}$ so that

$$\int_{-1}^{1} m_{j,\alpha}(v)dv = 0, \tag{6}$$

for $j = 1, \ldots, d$. For the **first-stage series** estimator, let $\{p_\kappa : \kappa = 1, 2, \ldots\}$ denote a basis for smooth functions, e.g. B-Splines. In order to obtain asymptotic results, κ must satisfy certain conditions as $n \to \infty$, which are given in section 4 of Horowitz and Lee (2005). For any positive integer κ fulfilling these conditions, define a vector consisting of a one and the basis function values for all explanatory variables

$$P_\kappa(x) = [1, p_1(x^1), \dots, p_\kappa(x^1), p_1(x^2), \dots, p_\kappa(x^2), \dots \\ , p_1(x^d), \dots, p_\kappa(x^d)]^T. \tag{7}$$

Furthermore, define a vector of corresponding coefficients

$$\Theta_{\kappa,\alpha} = [\mu, \underbrace{\theta^1_{1,\alpha}, \dots, \theta^1_{\kappa,\alpha}}_{\Theta^1_{\kappa,\alpha}}, \underbrace{\theta^2_{1,\alpha}, \dots, \theta^2_{\kappa,\alpha}}_{\Theta^2_{\kappa,\alpha}}, \dots, \underbrace{\theta^d_{1,\alpha}, \dots, \theta^d_{\kappa,\alpha}}_{\Theta^d_{\kappa,\alpha}}]^T, \tag{8}$$

with $\Theta_{\kappa,\alpha} = [\mu, \Theta^1_{\kappa,\alpha}, \dots, \Theta^d_{\kappa,\alpha}]$. Then $P_\kappa(x)^T\Theta_{\kappa,\alpha}$ is a series approximation of the additive quantile model $\mu_\alpha + m_{1,\alpha}(x^1), \dots, m_{d,\alpha}(x^d)$. Given a random sample $\{(Y_t, X_t) : t = 1, \dots, T_R\}$ that follows a stationary and α-mixing process (see 2.6), the estimator $\tilde{\Theta}_{\kappa,\alpha}$ is given as the minimizer of

$$S^{(1)}_{\kappa,\alpha} = T_R^{-1} \sum_t^{T_R} \rho_\alpha[Y_t - P_\kappa(X_t)^T\Theta_{\kappa,\alpha}], \tag{9}$$

where $\rho_\alpha(\cdot)$ is the check function given in equation (2). The first-stage series estimator is defined as

$$\tilde{\mu}_\alpha + \tilde{m}_{1,\alpha}(x^1) + \dots + \tilde{m}_{d,\alpha}(x^d) = P_\kappa(x)^T\tilde{\Theta}_{\kappa,\alpha}. \tag{10}$$

For any $j = 1, \dots, d$ and any $x^j \in [-1, 1]$, the series estimator $\tilde{m}_{j,\alpha}$ for a single additive function $m_{j,\alpha}$ is given as

$$\tilde{m}_{j,\alpha}(x^j) = P_\kappa(x^j)\tilde{\Theta}^j_{\kappa,\alpha}. \tag{11}$$

Horowitz and Lee (2005) proposed a second-stage local polynomial quantile estimator in their paper. In the following I will only consider a local linear quantile estimator as in section 2.3.1, since it achieves good precision without assuming the existence of higher order derivatives of the functions $m_{\alpha,j}(\cdot)$ (see Fan (1996)). To describe the **second-stage local linear quantile estimator** of for example $m_j(x^j)$ define

$$\tilde{Y}_{t-j} = Y_t - \tilde{\mu}_\alpha - \sum_{i \neq j}^{d} \tilde{m}_{i,\alpha}(X^i_t), \tag{12}$$

where $\tilde{\mu}_\alpha$ and $\tilde{m}_{i,\alpha}(X_t^i)$ is the first-stage estimate. For a random sample $\{(Y_t, X_t) : t = 1, \ldots, T_R\}$, the local second-stage estimator $\hat{m}_{j,\alpha}(x^j)$ of $m_{j,\alpha}(x^j)$ is defined as $\hat{m}_{j,\alpha}(x^j) = \hat{a}$, where $\hat{a}$ and $\hat{b}$ minimize

$$S_{j,\alpha}^{(2)} = (T_R h)^{-1} \sum_{t=1}^{T_R} \rho_\alpha \left[\tilde{Y}_{t,-j} - a - b(X_t^j - x^j) \right] k \left(\frac{x^j - X_t^j}{h_j} \right), \tag{13}$$

where $k(\cdot)$ is a univariate kernel function with support $[-1, 1]$, $\rho_\alpha(\cdot)$ is the check function described in section 2 and h_j is the bandwidth corresponding to regressor x^j.

2.4.3 Number of basis functions and bandwidth selection

For the first-stage estimator the **optimal number of basis functions** κ has to be estimated. Horowitz and Lee (2005) propose to use the following Schwartz-type information criterion (based on He and Shi (1996) and Doksum and Koo (2000)) where the estimator $\hat{\kappa}$ of κ minimizes

$$QBIC(\kappa) = n \log \left(\sum_{t=1}^{T_R} \rho_\alpha [Y_t - P_\kappa(X_t)^T \tilde{\Theta}_{\kappa,\alpha}] \right) + 2(\log n)\kappa, \tag{14}$$

where the second term represents a penalty for the number of basis functions. Furthermore, Horowitz and Lee (2005) state that in the first stage overfitting is needed to reduce the asymptotic bias, therefore they set $\kappa = \hat{\kappa} + 1$.
The **bandwidths** $h = (h_1, \ldots, h_d)$ are chosen by using a rule-of-thumb described in Fan (1996). Their rule-of-thumb bandwidth is derived from the asymptotically optimal bandwidth for a univariate quantile function (based on the minimization of the conditional weighted mean integrated squared error of a local linear estimator) and is given as

$$\hat{h}_{ROT} = C(\mathcal{K}) \left[\frac{\alpha(1-\alpha)[\hat{f}\{\hat{F}^{-1}(\alpha)\}]^{-2} \int w_0(x)dx}{\sum_{t=1}^{T_R} \{\hat{m}_\alpha''(X_t)\}^2 w_0(X_t)} \right]^{1/5}, \tag{15}$$

with

$$C(\mathcal{K}) = \left[\frac{\int \mathcal{K}^2(u)du}{\{\int u\mathcal{K}(u)du\}^2}\right], \tag{16}$$

where $w_0(x)$ is a weight function, which is set to be $w_0(x) = \mathbb{1}_{[-2,2]}(x)$ by Horowitz and Lee (2005), to eliminate extreme values.[2] The quantile function $m_\alpha(\cdot)$ in equation (15) is estimated by a global polynomial fit with degree four, $\hat{m}_\alpha = \hat{a}_0 + \hat{a}_1 x + \hat{a}_2 x^2 + \hat{a}_3 x^3 + \hat{a}_4 x^4$, with the second derivative being $\hat{m}''_\alpha = 2a_2 + 6a_3 x + 12a_4 x^2$.[3] $\hat{F}^{-1}(\alpha)$ is the α-th sample quantile of the residuals of the global polynomial fit and $\hat{f}(\cdot)$ is obtained from a kernel density estimate of these residuals. Finally, we have to calculate the value for $C(\mathcal{K})$, which only depends on the chosen Kernel. As mentioned in section 3.2, I follow Horowitz and Lee (2005) and use a Gaussian Kernel for which $C(\mathcal{K}) = 0.776$.

2.5 Semiparametric partial linear model - The average quantile regression estimator:

A second class of semiparametric models are partial linear models. These models, just as the semiparametric additive models, avoid the curse of dimensionality by imposing some structure on the functional form of the regression model and thus achieving dimension reduction. For the partial linear model, one assumes that one part of the model is linear and the other follows some unknown multivariate function.

[2] As mentioned before, X is normalized to have standard deviation of one and mean of zero, thereby values above two times the standard deviation are considered as "extreme".

[3] Fan (1996) use a polynomial up to degree $p+3$, where p stands for the degree of the polynomial kernel regression used, for the local linear model $p = 1$.

2.5.1 The semiparametric partial linear quantile regression model:

A partial linear quantile regression model can be written as

$$Y_t = m_\alpha(X_t) + Z_t^T \beta_\alpha + u_{t,\alpha}, \quad t = 1, \dots, T_R, \tag{17}$$

where α is the quantile, $Y_t \in \mathbb{R}$, $X_t \in \mathbb{R}^{d_x}$, $Z_t \in \mathbb{R}^{d_z}$, $\{(Y_t, X_t, Z_t)\} : t = 1, \dots, T_R\}$ is assumed to be a stationary and α-mixing process (see 2.6), $m_\alpha(\cdot)$ is an unknown real-valued function and $u_{t,\alpha}$ is an unobserved r.v. with $Q_\alpha(u_{t,\alpha}|x_t) = (F^{-1}_{u_{t,\alpha}}(\alpha)|x_t) = 0$.

2.5.2 The three-stage estimator

The following three-stage estimator was proposed by Cai and Xiao (2012) and was originally constructed for another class of semiparametric models, the partially varying coefficient models. These models can be written as

$$Y_t = m_\alpha(X_t)V_t + \beta_\alpha^T Z_t + u_{t,\alpha}, \quad t = 1, \dots, T_R, \tag{18}$$

with $X_t \in \mathbb{R}^{d_x}$, $Z_t \in \mathbb{R}^{d_z}$ and $V_t \in \mathbb{R}^{d_v}$, $\{(Y_t, V_t, X_t, Z_t) : t = 1, \dots, T_R\}$ is a stationary and α-mixing process (see 2.6) and $u_{t,\alpha}$ as in equation (17). If $V_t = 1$ for $t = 1, \dots, T_R$, the model reduces to a semiparametric partial linear quantile regression model as in equation (17) and we can use the estimator developed for this model.

To estimate β_α, Cai and Xiao (2012) proposed to use an adjusted local linear estimator in the **first stage**. This estimator is similar to the estimator mentioned in equation (1), but includes also the linear part of equation (17). The local estimator $\hat{\beta}_\alpha(x_t)$ of β_α is defined by $\hat{\beta}_\alpha(x_t) = \hat{c}$, where $(\hat{a}, \hat{b}, \hat{c})$ is the solution to the following minimization problem:

$$\min_{a,b,c} \sum_{s=1}^{T_R} \rho_\alpha[Y_t - a - b^T(X_s - x_t) - c^T Z_t] \mathcal{K}\left(\frac{x_t - X_s}{h^{(1)}}\right), \quad \forall\, t = 1, \dots, T_R, \tag{19}$$

where $h^{(1)}$ denotes the vector of bandwidths of length d_x for the first-stage estimator, $\mathcal{K}(\cdot)$ is a product kernel as in equation 1 and $\rho_\alpha(\cdot)$ is the check function defined in equation (2). Note that β_α is a global parameter but the first-stage estimator involves only local data points in the neighborhood of x. Cai and Xu (2008) show that the local estimator $\hat{\beta}_\alpha(x_t)$ converges in probability to β_α at a nonparametric rate, $\hat{\beta}(\cdot) - \beta = O_p((nh))^{-1/2}$. To obtain a more efficient estimator for β_α we have to use all data points and the optimal convergence rate should be $\sqrt{n}$, instead of $\sqrt{nh}$. To obtain a $\sqrt{n}$-consistent estimator for β_α, Cai and Xiao (2012) proposed the simple averaging method

$$\hat{\beta}_\alpha = \frac{1}{T_R} \sum_{t=1}^{T_R} \hat{\beta}_\alpha(x_t). \tag{20}$$

Lee (2003) proposed to use another **second-stage** averaging method given as

$$\hat{\beta}_\alpha = \frac{\sum_{t=1}^{T_R} \tau_X(x_t)\hat{\beta}_\alpha(x_t)}{\sum_{t=1}^{T_R} \tau_X(x_t)}, \tag{21}$$

where $\tau_X(\cdot)$ is a trimming function such that $\tau_X(x) = \mathbb{1}(x \in \mathcal{X})$ with a compact subset $\mathcal{X}$ of $\mathbb{R}^{d_x}$. The function is introduced to estimate β_α without being influenced by the tail behavior of the distribution of X. Lee (2003) use the following trimming function, $w_0(x) = \mathbb{1}_{[-2,2]}(x)$. The motivation behind this threshold is that before applying the estimator, the exogenous variables X, Z are normalized such that the standard deviation is one and the mean is zero. The coefficient vector β_α is always estimated based on the reference period.

After having estimated the parametric component β_α, in the **third stage** we still have to estimate the unknown function $m_\alpha(\cdot)$. Since the parametric component β_α can be estimated with an $n^{-1/2}$ rate (faster than the fastest possible rate of convergence for the nonparametric component), we can estimate $m_\alpha(\cdot)$ as asymptotically efficient as if β_α were known (see Cai and Xiao (2012)). Therefore, we can regress $Y_t - Z_t^T \hat{\beta}_\alpha$ on X_t by using the nonparametric local linear quantile estimator explained in section 2.3.1 using a different bandwidth vector $h^{(2)}$.

2.5.3 Bandwidth Selection

For first-stage and the third-stage estimator we have to choose the bandwidth denoted as $h^{(1)}$ and $h^{(2)}$. For the first stage Cai and Xiao (2012) used a rule-of-thumb idea to obtain an estimate $\hat{h}^{(1)}$ for $h^{(1)}$. First, they choose a data-driven bandwidth selector as suggested in Cai and Xu (2008) to obtain an initial bandwidth, denoted as $\hat{h}_0^{(1)}$. Furthermore, they state that for the first-stage estimation of $c(x)$, the bandwidth should be smaller than optimal (undersmoothing) to reduce the bias. Therefore, by following the idea in Cai (2002) for a two-step approach, they take the bandwidth as $\hat{h}_1^{(1)} = A_0 \times \hat{h}_0^{(1)}$, where $A_0 = n^{-\omega_0}$, with $\omega_0 = 1/10$. To choose the initial bandwidth estimate $\hat{h}_0^{(1)}$, Cai and Xu (2008) proposed a method based on the nonparametric version of the AIC for the univariate quantile regression case. Nevertheless, they do not give a theoretical basis for the bandwidth selection method (in the quantile regression case) and lack to explain how to perform this approach for the multivariate case. Additionally, the method is computationally intensive. Therefore, I will use the simple mean-bandwidth transformation method explained in section 2.3.2 and proposed by Yu and Jones (1998), to select the initial bandwidth $\hat{h}_0^{(1)}$. Note, that for this method, we have to estimate the optimal bandwidth for the mean regression $h_{mean}^{(1)}$ first. This mean regression for the first-stage estimator is simply $Y_t = m(X_t) + \varepsilon_t$ as in equation (3).

For the third-stage estimator we also need a bandwidth $h^{(2)}$. Nevertheless, this bandwidth does not have to be smaller than optimal (undersmoothing). Therefore, we can skip the second step of the just mentioned approach and only use mean-bandwidth transformation method proposed by Yu and Jones (1998). In this case the mean regression is slightly different, since we use the model $Y_t - Z^T\hat{\beta}_\alpha = m(X_t) + \varepsilon_t$ and perform least-squares cross-validation on this model to obtain $h^{(2)}_{mean}$.

2.6 Assumptions for time series models

In order to perform statistical inference on time series data, it is necessary to assume that at least some features of the underlying probability law are sustained over a time period of interest. For nonlinear time series analysis, this leads to the assumptions of *strict stationarity.* Following Fan (2003), a time series $\{X_t : t = 0 \pm 1, \pm 2, \dots\}$ follows a strictly stationary process if $(X_1, \dots, X_n)$ and $(X_{1+k}, \dots, X_{n+k})$ have the same joint distributions for $n \in \mathbb{N}_0$ and $k \in \mathbb{Z}$.

Furthermore we need the assumption of *mixing* for the implementation of the estimators presented in the previous sections. Intuitively, a mixing time series can be viewed as s sequence of random variables for which the past and distant future are asymptotically independent. There exist various mixing conditions to reflect different kinds of dependencies. For our analysis we need α-mixing. For $n \in \mathbb{N}_0$, define

$$\alpha(n) = \sup_{A \in \mathcal{F}^0_{-\infty}, B \in \mathcal{F}^\infty_n} |P(A)P(B) - P(AB)|, \tag{22}$$

where $\mathcal{F}^j_i$ denotes the σ-algebra generated by $\{X_t, i \le t \le j\}$ and $X_t \in \mathcal{L}^2(\mathcal{F}^j_i)\ \forall\ i \le t \le j$. The process is then said to be α-mixing if $\alpha(n)$ converges to zero as $n \to 0$.

Honda (2000) proved the uniform convergence and asymptotic normality of nonparametric local polynomial estimators in a time series setting for α-mixing and stationary processes. Under the condition of α-mixing and strict stationarity Huang and Shen (2004) established the consistency and rates of convergence of spline estimates in a functional coefficient time series setting.[4]

2.7 Model testing

Model testing for quantile models is most commonly applied in financial risk modeling, as for instance the Value-at-Risk model, which is defined as a conditional quantile of the return distribution. One of the key evaluation techniques used in this area is backtesting. Among the most basic and widely applied backtesting methods are the test for conditional and unconditional coverage; and the test for independence proposed by Christoffersen (1998), which I will explain in the following section.

First, define the event where the realized value Y_t exceeds the estimated conditional quantile $\hat{Q}_\alpha(Y_t|X_t = x_t)$ as a *violation*. Based on this idea, define the sequence of violations, called hit sequence, as

$$I_t(\alpha) = \begin{cases} 1 & \text{if} \quad Y_t > \hat{Q}_\alpha(Y_t|X_t = x_t) \quad t = 1, \ldots, T, \\ 0 & \text{else.} \end{cases} \tag{23}$$

The test of Christoffersen (1998) consists of two tests. First, one tests the null hypothesis that

$$I_t(\alpha) \sim i.i.d. \;\; Bernoulli(1 - \alpha), \tag{24}$$

[4]Horowitz and Lee (2005) showed consistency and rates of convergence of their estimator only for independent data. Nevertheless, it is assumed that the estimator is also consistent for dependent data, since the spline based estimator and the local linear estimator are also consistent. A prove is left to further studies.

and tests against the alternative

$$I_t(\alpha) \sim i.i.d.\ \ Bernoulli(\pi_1). \tag{25}$$

This test is called the test of correct unconditional coverage

$$H_{0,uc} : \pi_1 = 1 - \alpha, \tag{26}$$

which is a test that on average the coverage, the number of realized values lying below the estimated quantile, is correct. Christoffersen (1998) uses a likelihood ratio test (LR-test) to test this hypothesis. The corresponding likelihood functions under the null and alternative hypothesis are

$$L(I, p) = (1 - \alpha)^{T_1} \alpha^{(T - T_1)}, \tag{27}$$

where T_1 is the number of ones in the sample; and under the alternative

$$L(I, \pi_1) = \pi_1^{T_1} (1 - \pi_1)^{(T - T_1)}. \tag{28}$$

The maximum-likelihood estimate of the unknown probability parameter π_1 is

$$\hat{\pi}_1 = T_1/T \tag{29}$$

and therefore one can write the test statistic of the LR-test of unconditional coverage (LR-UC test) as

$$LR_{uc} = -2 \log \left[\ln L(I, 1 - \alpha) / \ln L(I, \hat{\pi}_1) \right] \overset{asy}{\sim} \chi^2(s - 1) = \chi^2(1), \tag{30}$$

where $s = 2$ is the number of possible outcomes of I_t. This test assumes that the violations are independent. To test this hypothesis, one defines an alternative where the hit sequence follows a first-order Markov sequence with switching probability matrix

$$\Pi_1 = \begin{bmatrix} 1 - \pi_{01} & \pi_{01} \\ 1 - \pi_{11} & \pi_{11} \end{bmatrix}, \tag{31}$$

where $\pi_{ij} = \Pr(I_t = j | I_{t-j} = i)$. One can then test the independence as

$$H_{0,ind} = \pi_{01} = \pi_{11}. \tag{32}$$

For this test, the likelihood under the alternative hypothesis is

$$L(I, \pi_{01}, \pi_{11}) = (1 - \pi_{01})^{(T_0 - T_{01})} \pi_{01}^{T_{01}} (1 - \pi_{11})^{(T_1 - T_{11})} \pi_{11}^{T_{11}}, \tag{33}$$

where T_{ij} denotes the number of observations with a j following an i. The maximum-likelihood estimates are

$$\hat{\pi}_{01} = T_{01}/T_0, \tag{34}$$

$$\hat{\pi}_{11} = T_{11}/T_1. \tag{35}$$

Under the null hypothesis the switching matrix becomes

$$\Pi_2 = \begin{bmatrix} 1 - \pi_1 & \pi_1 \\ 1 - \pi_1 & \pi_1 \end{bmatrix}. \tag{36}$$

The likelihood under the null is

$$L(I, \pi_1) = (1 - \pi_1)^{(T_{00} + T_{10})} \pi_1^{(T_{01} + T_{11})}. \tag{37}$$

Given these two likelihood functions one can define the test statistic of the LR-test for independence (LR-Ind test) as

$$LR_{ind} = -2 \log \left[L(I, \hat{\pi}_1) / L(I, \hat{\pi}_{01}, \hat{\pi}_{11}) \right] \overset{asy}{\sim} \chi^2 \left((s-1)^2 \right) = \chi^2(1). \tag{38}$$

Having defined a test for unconditional coverage and independence separately, we only have to combine them to arrive at the joint test of coverage and independence, also called conditional coverage. The null hypothesis of this test is given as

$$H_{0,ind} = \pi_{01} = \pi_{11} = 1 - \alpha. \tag{39}$$

In effect one tests the null of unconditional coverage against the alternative of the independence test. Christoffersen (1998) show that the LR-test of conditional coverage (LR-CC test) is given as

$$LR_{cc} = -2\log\left[L(I, 1-\alpha)/L(\hat{\pi}_{01}, \hat{\pi}_{11})\right] \overset{asy}{\sim} \chi^2\left(s(s-1)\right) = \chi^2(2). \tag{40}$$

This LR-test enables joint testing of independence and correct coverage.

2.8 Monitoring warning

As mentioned in section 2.1 we still have to define a method that indicates what share of violations or how many consecutive violations are maximally acceptable before the monitoring indicates an abnormally high energy consumption and issues a warning. Following, I present two ideas how to develop such a method, the first one is based on the LR-UC test and the second one on the LR-Ind test. The first method can be used in the retrospective evaluation period monitoring approach. Assume that in the evaluation period one observes the hit sequence

$$I_t^{eval}(\alpha) = \begin{cases} 1 & \text{if} \quad Y_t > \hat{Q}_\alpha(Y_t|X_t = x_t) \quad t = 1, \dots, T_E, \\ 0 & \text{else.} \end{cases} \tag{41}$$

If the quantile regression model performs well, then the share of violations $\pi_1 = 1/T_E \sum_{t=1}^{T_E} I_t^{eval}$ should be close to $1-\alpha$ or in other words the coverage rate is correct. If the machine runs less energy efficient than in the reference period, π_1 should lie above $1-\alpha$, and of course vice versa. Now, to define a rule for the for the assessment of the power consumption in the evaluation period, one could use the LR-UC test and denote power consumption as too high, if the null hypothesis of the LR-UC test is rejected and $\pi_1 > 1-\alpha$. In short, define the evaluation period monitoring warning as

$$M^{uc}(\alpha) = \begin{cases} 1 & \text{if} \quad LR_{uc}(\alpha) > 6.635 \text{ and } \hat{\pi}_1 > 1 - \alpha, \\ 0 & \text{else,} \end{cases} \tag{42}$$

where $\hat{\pi}_1 = T_1/T$, $LR_{uc}(\alpha)$ as in equation (30) and where the critical value 6.635 is taken from a chi squared distribution with $df = 1$ and a significance level of 1%. Whenever $M^{uc}(\alpha) = 1$ the power consumption is evaluated as being too high and a warning is issued. The high significance level is chosen here to reduce the probability of committing a type I error. The motivation is that an incorrect rejection of a true null hypothesis (correct coverage) would result in a wrong warning of the monitoring system and this could undermine the credibility of the monitoring system. However, if a higher sensitivity of the test is desired, the significance level can be reduced to 5% or 10%.

The second method bases on the LR-Ind test of the previous section and can be used for the live-monitoring approach. Assuming the quantile regression model performs well and one obtains independent violations in the reference period. Then, one would expect independent violations in the evaluation period as well. Under this assumption of independent violations and for a large quantile, e.g. 95%, it would be extremely unlikely to observe four or more consecutive violations. The probability of four consecutive violations would be only $P(I_t^{eval}(\alpha) = I_{t-1}^{eval}(\alpha) = I_{t-2}^{eval}(\alpha) = I_{t-3}^{eval}(\alpha) = 1) = 0.05^4 = 0.00000625$, in percentage 0.000625%. Thus, a second indicator for increased energy consumption could be the number of consecutive violations. As already mentioned, this indicator is especially useful for the live-monitoring approach, since no fixed evaluation period is needed.

The live-monitoring warning is defined as

$$M_t^{ind}(\alpha) = \begin{cases} 1 & \text{if} \quad I_t^{eval}(\alpha) = I_{t-1}^{eval}(\alpha) = I_{t-2}^{eval}(\alpha) = I_{t-3}^{eval}(\alpha) = 1, \\ 0 & \text{else.} \end{cases} \tag{43}$$

A warning is issued if at least the last four observations in the hit sequence were violations. The number of consecutive violations could be set differently. A higher number would imply a less sensitive monitoring system, meaning a warning would be issued only in the case of a substantial and lasting upward deviation from the usual power consumption.

Obviously, both tests only work if the estimated quantile model performs well. In the first case this means that the estimated conditional quantile should actually result in share of violations that is close to $1 - \alpha$ for the reference period. This has to be tested with the LR-UC test in the reference period. For the second test, evidently, we also need a well performing quantile regression model but in this case also the independence of the violations in the reference period is crucial. Independence and correct coverage can be tested using the LR-CC test.

In the case of a warning, there should always be a visual check by an expert to ascertain that there was no major error in the estimation of the quantile, due to missing or false explanatory variables for example.

3 Empricial Results

3.1 Data

To evaluate the performance of the different models applied on real data I choose two different machines, a cooling machine and a dry cooler, since cooling machines are among the machines with the highest power consumption in the industry. All data were provided by Ökotec Energiemanagement GmbH, an overview is given in table 1. The different time periods were chosen with the support of an engineering expert such that there was no evidence for major wear and tear of the machine and no changes in the operational settings of the machine were done in the reference period. No evaluation period for the dry cooler is given because in the practical example in section 3.4 only the cooling machines was analyzed. The different variables were chosen with engineering expertise to best describe the power consumption of the respective machine. For both machines the frequency of the data is 15 minutes. Stationarity of the different time series was tested using the augmented Dickey-Fuller Test (ADF test, see Dickey and Fuller (1979)).[5] The null hypothesis of an existent unit root was rejected for all time series from the respective reference period on a 1% significance level (see table 2).

3.2 Analysis

For all machines the response variable is the power consumption. As already mentioned, the explanatory variables were chosen with the support of

[5]The ADF-Test only tests on the existence of a unit root and is not sufficient to test strict stationarity. Some test on strict stationarity exist (e.g. see Lima and Neri (2013)) but are still new and an implementation and critical reflection would go beyond the scope of this thesis.

	Cooling Machine	Dry Cooler
Ref. Period	23.07.12 12:00 PM -	11.02.13 00:00 AM
	07.08.12 11:45 PM	03.02.13 11:45 PM
Number of Obs.	1392	1920
Eval. Period	08.08.12 00:00 AM -	
	15.08.12 00:00 AM	
Number of Obs.	673	
Interval Length	15 Minutes	15 Minutes
Response Variable	Y_t-Power cons., Wh	Y_t-Power cons., Wh
Explanatory Var.	X_t^1-Gen. cooling energy, Wh	X_t^1-Gen. cooling energy, Wh
	X_t^2-Outlet temp., °C	X_t^2-Outlet temp., °C
	X_t^3-Inlet temp., °C	X_t^3-Inlet temp., °C
	X_t^4-Outside temp., °C	X_t^4-Outside temp., °C

Table 1: Data description.

	Y_t	X_t^1	X_t^2	X_t^3	X_t^4
Dry Cooler	-4.27^{***}	-12.39^{***}	-4.53^{***}	-5.10^{***}	-4.89^{***}
Cooling Machine	-4.22^{***}	-4.36^{***}	-6.37^{***}	-6.78^{***}	-6.67^{***}

Table 2: Test statistics of the ADF-Test for data from the reference period.

an engineering expert. For the semiparametric models lagged explanatory variables were included to make use of the additional functionality of the semiparametric models due to the curse of dimensionality for purely non-parametric models. Table 3 gives an overview of the different estimated models. I estimated two different quantiles $\alpha_1 = 0.9$ and $\alpha_2 = 0.95$.

For the local linear quantile estimator and the semiparametric estimators the Gaussian kernel was chosen whenever kernel smoothing was used.

Local Linear Model (NP)	
Cooling Machine	$Y_t = m_\alpha(X_t^1, X_t^2, X_t^3, X_t^4)$
Dry Cooler	$Y_t = m_\alpha(X_t^1, X_t^2, X_t^3, X_t^4)$
Semiparametric Additive Model (SP-Add)	
Cooling Machine	$Y_t = \sum_{i=1}^4 m_\alpha^i(X_t^i) + m_\alpha^4(X_{t-1}^1) + m_\alpha^5(X_{t-1}^4) + m_\alpha^6(X_{t-2}^4)$
Dry Cooler	$Y_t = \sum_{i=1}^4 m_\alpha^i(X_t^i) + m_\alpha^4(X_{t-1}^1) + m_\alpha^5(X_{t-1}^4) + m_\alpha^6(X_{t-2}^4)$
Semiparametric Partial Linear Model (SP-PL)	
Cooling Machine	$Y_t = m_\alpha(X_t^1, X_t^2, X_t^3, X_t^4) + \beta_{\alpha,1}X_{t-1}^1 + \beta_{\alpha,2}X_{t-1}^4 + \beta_{\alpha,3}X_{t-2}^4$
Dry Cooler	$Y_t = m_\alpha(X_t^1, X_t^2, X_t^3, X_t^4) + \beta_{\alpha,1}X_{t-1}^1 + \beta_{\alpha,2}X_{t-1}^4 + \beta_{\alpha,3}X_{t-2}^4$

Table 3: Model specifications.

For the additive model cubic B-Splines were used in the first-stage, following Horowitz and Lee (2005). All models were estimated based on the data from the reference period. The estimated models were used to predict power consumption in the evaluation period.

Figure 3 shows the estimated 95%-quantile for all models and the two different machines in a segment of the evaluation period (I only show a segment for better illustration, the entire evaluation period of both quantiles is given in the appendix). All models seem to perform well as the estimated quantile follows the movement of the realized power consumption over the entire segment and no major deviations are evident. For all machines the estimated quantile is close to the realized power consumption for all three models. For the purpose of monitoring this result is desirable since relatively small changes in machine efficiency will result in power consumption surpassing the estimated quantile. Furthermore, visually inspecting the figure, there are no major differences between the three different models since the lines are nearly superimposable, which leads to the conclusion that lagged variables only have a minor influence on the estimated quantile. However, this is only a visual deduction and needs further examination using the tests described in section 2.7.

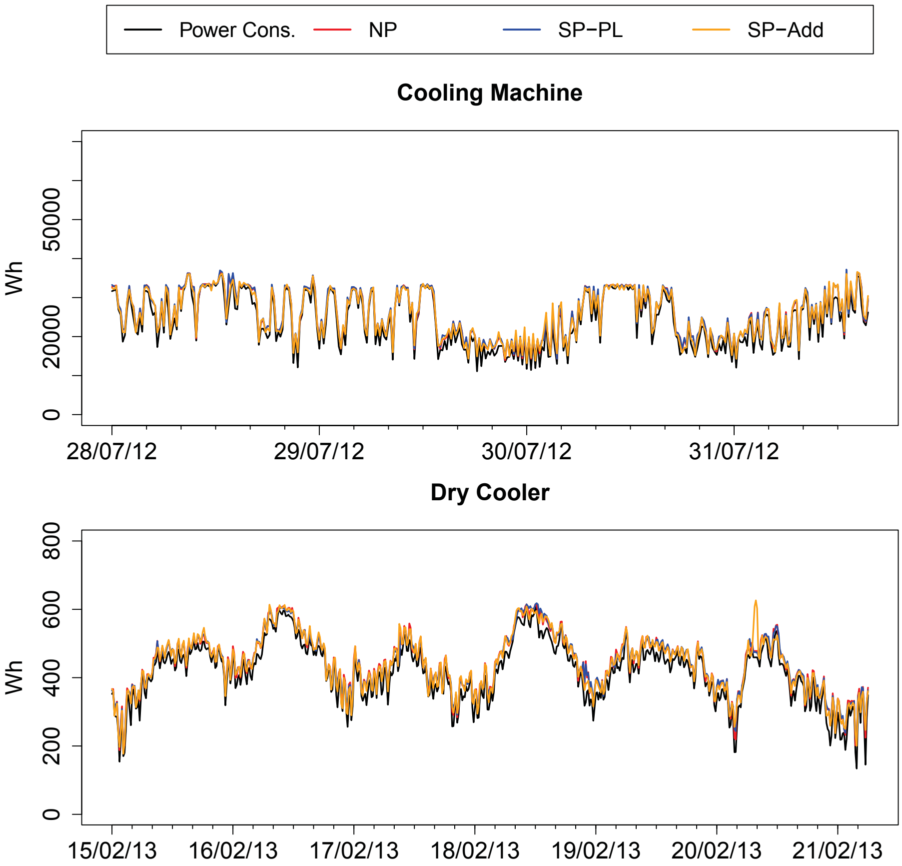


Figure 3: Realized power consumption and estimated 95%-quantile for the two machines (only an evaluation period segment for better illustration).

3.3 Evaluation

The evaluation of the performance of the different models and estimators is done using the LR-CC, LR-Ind and LR-UC test based on the reference period. The results are shown in table 4 and 5.

		Quantile	NP	SP-PL	SP-Add
Cooling Machine	LR-UC	90%	98.43***	86.53***	2.73
		95%	87.03***	79.09***	2.53
	Share of violations	90%	3.09%	3.45%	8.69%
		95%	0.65%	0.79%	4.09%
Dry Cooler	LR-UC	90%	284.53***	152.29***	19.94***
		95%	156.82***	117.65***	36.76***
	Share of violations	90%	0.89%	2.76%	7.08%
		95%	0.26%	0.68%	2.29%

Table 4: Test-statistic of the LR-UC test and share of violations for the reference period.

As we can see for the LR-UC test, in most of the cases the null hypothesis of correct unconditional coverage is rejected on a 1%- significance level. Only for the additive model based on the cooling machine the null hypothesis is not rejected on any significance level. Here the share of violations is close to the values expected based on the quantile. In all the other cases the share of violations lies well below the share we would expect. Thus, we "overestimate" the quantile, which means that our estimated quantile lies on average above the true quantile. This has implications for the monitoring process since an abnormally high energy consumption level would be reported only if realized values exceed the overestimated quantile, which in turn means that the monitoring warning methods M^{uc} and M^{ind} become less sensitive. Essentially, this is not a negative result, since the probability of issuing a false warning is reduced. On the other hand, the probability of committing a type II error is increased and consequently the power of the test is reduced. As already mentioned, for the M^{uc}-method, one can counteract this effect by choosing a critical value which corresponds to a lower significance level.

		Quantile	NP	SP-PL	SP-Add
Cooling Machine	LR-Ind	90%	0.09	4.86**	28.68***
		95%	0.12	0.18	19.72***
	LR-CC	90%	98.53***	91.39***	31.41***
		95%	87.14***	79.26***	22.26***
Dry Cooler	LR-Ind	90%	0.30	5.72**	99.71***
		95%	0.03	0.18	31.99***
	LR-CC	90%	284.84***	158.01***	119.65***
		95%	156.85***	117.82***	68.75***

Table 5: Test-statistic of the LR-Ind and LR-CC test for the reference period.

Table 5 shows the results of the LR-Ind and LR-CC test. The LR-Ind tests null hypothesis of independence is not rejected on any significance level for the nonparametric model for both machines and quantiles. For the partial linear model the null hypothesis is not rejected for the 95%-quantile and rejected on a 5%-significance level for the 90%-quantile. For the additive model the null hypothesis is rejected on a 1%-significance level for both machines and quantiles. The rejection of the null hypothesis in the case of the additive model is closely linked to the occurrence of too many consecutive violations, hence $\pi_{11} > \pi_{01}$ (the different estimated probabilities $\hat{\pi}_{01}$ and $\hat{\pi}_{11}$ for all machines are shown in the appendix). This is shown in figure 4, especially for the additive model one can see many occurrences of consecutive violations (indicated by the red triangle above). This has an adverse effect on the second monitoring method M_t^{ind}, since the occurrence of many consecutive violations could also be caused by a weak model.

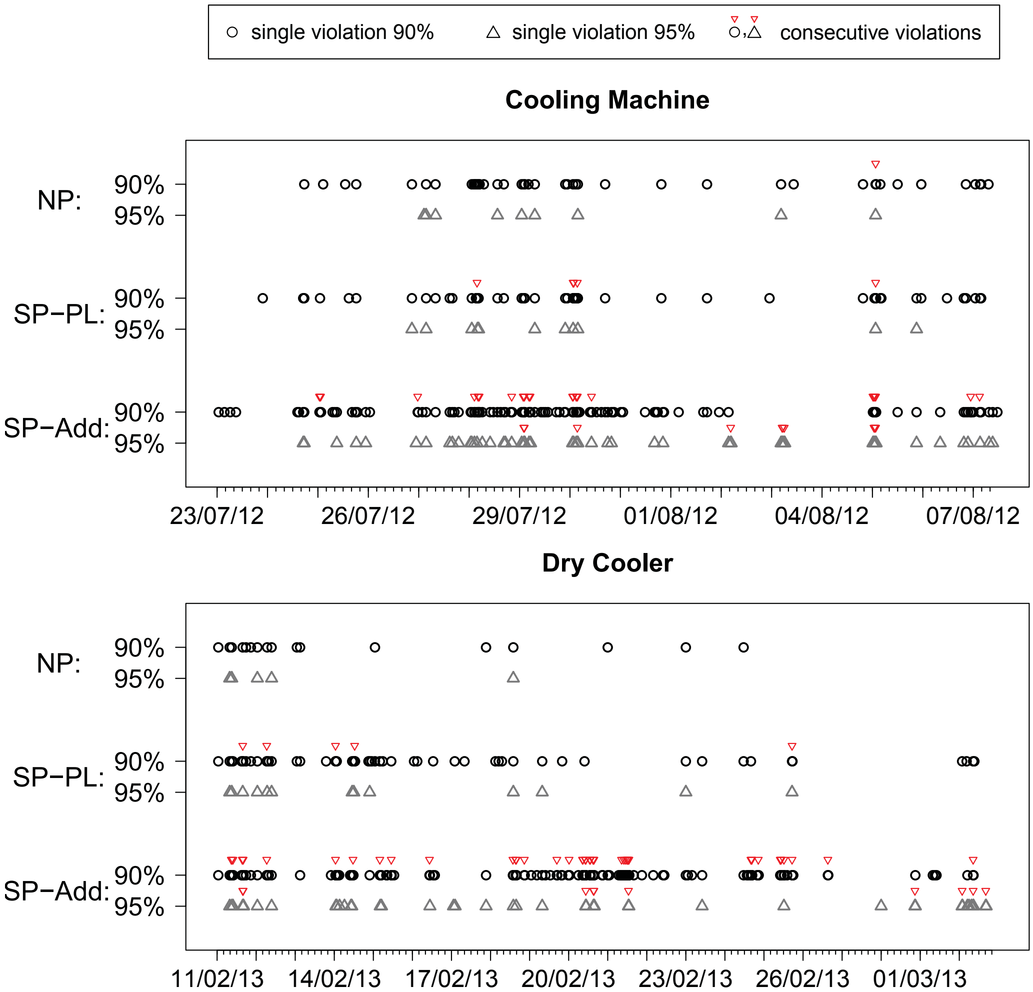


Figure 4: Single (black or grey symbol) and consecutive violations (red triangle above) for the different models and quantiles for the two machines.

As for the LR-CC test one can see in table 5 that it is rejected on the 1%-significance level for all models for both machines and quantiles. This result is straightforward since the LR-CC test is a combination of the LR-UC and LR-Ind test and the models deliver either a correct coverage or independent violations but not both. Hence, in terms of the LR-CC test none of the models performs well. However, examining the values of the test-statistic,

one can see an interesting result. For both machines and quantiles the nonparametric model has the highest value, the semiparametric model has the second highest and the additive model the lowest value of the test statistic. Thus, the two semiparametric models and especially the additive model perform best in terms of the LR-CC test, since their statistic is closest to the null hypothesis of correct conditional coverage.
Concluding, one can not identify a best model since the test results are mixed. In terms of correct unconditional coverage the additive model performs best, in terms of independence of violations the nonparametric and semiparametric partial linear model perform well, with a slight advantage for the nonparametric model. In terms of the combined test for conditional coverage, there is no well performing model but both semiparametric models have a lower test-statistic than the nonparametric model, which supports the usage of lagged variables in the models . However, none of the models is capable to deliver independent violations and a correct unconditional coverage.
For the broad implementation of the monitoring system the additive model might be advantageous, given that it is computationally less intensive than the other two models. For the proposed estimation of the additive model one only has to perform a fast spline regression, and several univariate bandwidth selections and nonparametric regressions, whereas for the estimators of the other two models we have a computationally much more intensive multivariate bandwidth selection and kernel regression.

3.4 Example

To illustrate the monitoring in a practical example, I altered the realized data from the evaluation period of the cooling machine in two different ways.[6] In the first example, I increased the power consumption abruptly by a stochastic factor after the first half of the evaluation period. Define the mean-altered power consumption in the evaluation period as

$$Y_t^{mean} = \begin{cases} Y_t & \text{if} \quad t \in T_{E1}, \\ A_t Y_t & \text{else}, \end{cases} \tag{44}$$

where T_{E1} is the first half of the observation in the evaluation period T_E (08.08.12 00:00 AM - 11.08.12 12:00 PM) and $A_t \sim N(1.05, 0.02)$. The stochastic factor A_t moderately increases the mean power consumption and the variance after the first half of the evaluation period. Secondly, I increased only the variance of the power consumption. Define the variance-altered power consumption as

$$Y_t^{var} = \begin{cases} Y_t & \text{if} \quad t \in T_{E1}, \\ B_t Y_t & \text{else}, \end{cases} \tag{45}$$

where $B_t \sim N(1, 0.05)$. This represents a small change in the variance of power consumption but no change in the mean. Thus, a monitoring system that is only designed to detect an increase in power consumption, should issue no warning. I estimated the conditional 95%-quantile using the semiparametric partial linear model, since the model delivered independent violations in the reference period and performed better than the pure nonparametric model in terms of correct unconditional coverage.

[6] Alteration was necessary since in the evaluation period no major wear and tear was evident and the operational settings of the machine were not changed.

I applied both the $M^{uc}(\alpha)$- and the $M^{ind}(\alpha)$-method on the original and the two altered evaluation periods. Figure 5 shows the violations for the three different evaluation periods and indicates where a warning was issued by the $M^{ind}(\alpha)$-method. As we can see, in the first half there are very few violations and the occurrence is, as designed, identical for Y_t^{mean}, Y_t^{var} and Y_t. In the second half, the number of violations for Y_t^{var} is higher than for Y_t but lower than for Y_t^{mean}. Additionally, the number of $M^{ind}(\alpha)$-warnings issued for Y_t^{mean} is much higher than for Y_t^{var}, where only one warning was issued. Thus, we already can see that a small change in the variance and especially a small change in the mean of the power consumption already leads to a much higher number of violations and warnings compared to the original series.

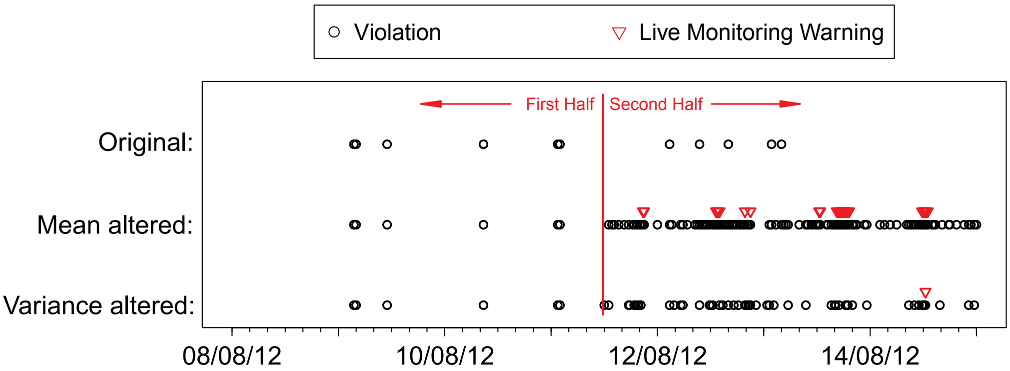


Figure 5: Violations and issued live-monitoring warnings for the different power consumption series of the example.

Table 6 shows the results of the LR-tests and the monitoring methods for the original and altered evaluation periods. One can observe that for the original series, the test results are similar to those of the reference period and no warning was issued neither by the $M^{uc}(\alpha)$- nor by the $M^{ind}(\alpha)$-method. For the altered evaluation periods, one sees that the share of violations is increased, especially for the mean-altered evaluation period. This result is reflected by the rejection of the null hypothesis of the LR-UC test. The $M^{uc}(\alpha)$-method issued a warning for both series, which is a direct consequence of the elevated share of violations. The $M^{ind}(\alpha)$-method issued several warning for both but much less for the variance altered evaluation period.

	Y_t	Y_t^{mean}	Y_t^{var}
LR-CC Test Statistic	21.79***	239.14***	26.95***
LR-UC Test Statistic	21.42***	183.52***	15.47***
LR-Ind Test Statistic	0.37	55.63***	11.48***
Share of Violations	1.63%	19.75%	8.62%
$M^{uc}(\alpha)$ warning issued	NO	YES	YES
Number of $M^{ind}(\alpha)$ warnings	0	28	1

Table 6: Results of the LR-tests and monitoring methods for the altered and original evaluation period.

Thus, both monitoring warning methods are not capable to differentiate between a stochastic increase in the mean and a pure increase in the variance. This is an important result, since the monitoring should only detect an increase in the mean of the power consumption, which results in a reduced energy efficiency of the machine. Nevertheless, both tests are capable to detect the mean increased power consumption, even though the mean was only increased moderately. The M_t^{ind}-method issues the first warning relatively quick at 8:45 PM, less than nine hours after the abrupt change on 11.08.2013 at 12:00 PM (see figure 5). Therefore, I am confident that the monitoring warning methods are quite useful in the detection of increased power consumption. Nevertheless, as we have seen, a warning does not necessarily indicate an increased power consumption, since a change in variance could also result in a warning. An idea of how to identify such situations is to estimate the conditional mean using one of the above mentioned models and compare the mean of these estimated values with the mean of the realized values in the evaluation period using a t-test.[7] This represents only a first idea, detailed considerations on this issue are left to further studies.

[7] In the case of estimating the conditional mean one has to replace the asymmetric loss function with a squared loss function.

4 Conclusion

This work presents an innovative approach on how to design and implement a power consumption monitoring system for industrial machines. The approach bases on the comparison of predictions of a conditional quantile model estimated in a reference period with the realized power consumption in an evaluation period. Three different non- and semiparametric models were adopted to estimate the conditional quantile. Additionally, semiparametric models were used to avoid the curse of dimensionality of nonparametric models and to enable the use of additional (lagged) explanantory variables. The performance of the quantile regression models was evaluated using backtesting to test the independence of violations and the correct coverage of violations. Additionally, two simple monitoring warning methods were defined, the first one to detect an increased share of violations retrospectively and the other one to detect an elevated number of consecutive violations in a live-monitoring approach.

The results of the different models are promising. Visual inspection shows that the predicted values from all models seem to follow the movement of the realized power consumption without major deviations. Nevertheless, all models and especially the partial linear and the nonparametric model seem to overestimate the conditional quantile, since the share of violations is below the expected share based on the chosen quantile. However, this is essentially not a bad result, since the probability that the monitoring warning test issues a false warning is reduced, implying a less sensitive and thus a more reliable warning. The test on independence of violations indicates independent violations for nearly all models, only for the additive model the null hypothesis is clearly rejected as there are too many consecutive violations. Given the evaluation test results no clear superior model can be identified. However, according to the LR-CC test, the semiparametric models seem to perform better which implies a beneficial effect of the inclusion of additional lagged explanatory variables. Furthermore, the additive model has the advantage of a computationally less intensive estimator, which is beneficial if the monitoring system is implemented for many machines.

In a short practical example for one of the models, it could be shown that the monitoring approach is capable to detect a modest increase in the mean of power consumption. Both monitoring warning tests issued a warning and in the case of the live-monitoring, the warning was issued relatively quick after the mean was increased. Furthermore, the example showed that in the case of a variance increase, a warning was released as well although the mean of the power consumption was not elevated. An additional test on equality of an estimated conditional mean and the realized power consumption could be employed to detect such cases. Further elaborations on this test are left to future studies.

In future works, other nonparametric regression techniques could be compared to the here proposed approach. For example, spline based methods are computationally very fast to implement and seem to provide promising results (e.g. see Huang and Shen (2004) and Kim (2007)). Furthermore, other backtesting methods could be employed to improve the model evaluation (e.g. the conditional duration test by Christoffersen and Pelletier (2004)).

References

Behrendt, Thomas, André Zein, and Sangkee Min. 2012. *Development of an energy consumption monitoring procedure for machine tools*, {CIRP} Annals - Manufacturing Technology **61**, no. 1, 43 –46.

BMWi. 2010. *Energy concept for an environmentally sound, reliable and affordable energy supply*, Federal Ministry of Economics and Technology (BMWi).

Bundesministerium für Wirtschaft und Energie. 2013. *Richtlinie für die Förderung von Energiemanagementsystemen*, Bundesanzeiger, 1–7.

Cai, Zongwu. 2002. *Two-step likelihood estimation procedure for varying-coefficient models*, Journal of Multivariate Analysis **82**, no. 1, 189 –209.

Cai, Zongwu and Zhijie Xiao. 2012. *Semiparametric quantile regression estimation in dynamic models with partially varying coefficients*, Journal of Econometrics **167**, no. 2, 413–425.

Cai, Zongwu and Xiaoping Xu. 2008. *Nonparametric quantile estimations for dynamic smooth coefficient models*, Journal of the American Statistical Association **103**, no. 484.

Christoffersen, Peter and Denis Pelletier. 2004. *Backtesting value-at-risk: A duration-based approach*, Journal of Financial Econometrics **2**, no. 1, 84–108.

Christoffersen, Peter F. 1998. *Evaluating interval forecasts*, International economic review, 841–862.

De Gooijer, Jan G and Dawit Zerom. 2003. *On additive conditional quantiles with high-dimensional covariates*, Journal of the American Statistical Association **98**, no. 461, 135–146.

Dennis, JE and Daniel J Woods. 1987. *Optimization on microcomputers: The nelder-mead simplex algorithm*, New Computing Environments: Microcomputers in Large-Scale Computing, 116–122.

Devoldere, Tom, Wim Dewulf, Wim Deprez, Barbara Willems, and Joost R Duflou. 2007. *Improvement potential for energy consumption in discrete part production machines*, Springer.

Dickey, David A and Wayne A Fuller. 1979. *Distribution of the estimators for autoregressive time series with a unit root*, Journal of the American statistical association **74**, no. 366a, 427–431.

Dietmair, A and A Verl. 2009. *Energy consumption forecasting and optimisation for tool machines*, Energy **62**, 63.

Doksum, Kjell and Ja-Yong Koo. 2000. *On spline estimators and prediction intervals in nonparametric regression*, Computational statistics & data analysis **35**, no. 1, 67–82.

Draganescu, F., M. Gheorghe, and C.V. Doicin. 2003. *Models of machine tool efficiency and specific consumed energy*, Journal of Materials Processing Technology **141**, no. 1, 9 –15.

European Parliament and Council of the EU. 2012. *Directive 2012/27/EU*, Official Journal of the European Union **L 315**, 1.

Eurostat. 2014. *Final energy consumption, by sector*, Code: tsdpc320.

Fan, Jianquing. 1996. *Local polinomial modelling and its applications*, Vol. 66, CRC Press.

Fan, Jianquing. 2003. *Nonlinear time series: nonparametric and parametric methods*, Springer.

He, Xuming and Peide Shi. 1996. *Bivariate tensor-product b-splines in a partly linear model*, Journal of Multivariate Analysis **58**, no. 2, 162 –181.

Honda, Toshio. 2000. *Nonparametric estimation of a conditional quantile for alpha-mixing processes*, Annals of the Institute of Statistical Mathematics **52**, no. 3, 459–470.

Horowitz, Joel L and Sokbae Lee. 2005. *Nonparametric estimation of an additive quantile regression model*, Journal of the American Statistical Association **100**, no. 472, 1238–1249.

Hu, Shaohua, Fei Liu, Yan He, and Tong Hu. 2012. *An on-line approach for energy efficiency monitoring of machine tools*, Journal of Cleaner Production **27**, no. 0, 133 – 140.

Huang, Jianhua Z and Haipeng Shen. 2004. *Functional coefficient regression models for non-linear time series: A polynomial spline approach*, Scandinavian Journal of Statistics **31**, no. 4, 515–534.

Kim, Mi-Ok. 2007. *Quantile regression with varying coefficients*, The Annals of Statistics, 92–108.

Lee, Sokbae. 2003. *Efficient semiparametric estimation of a partially linear quantile regression model*, Econometric theory **19**, no. 1, 1 –31.

Li, Qi and Jeffrey Scott Racine. 2007. *Nonparametric econometrics: Theory and practice*, Princeton University Press.

Lima, Luiz Renato and Breno Neri. 2013. *A test for strict stationarity*, Uncertainty analysis in econometrics with applications, pp. 17–30.

Nadaraya, E. 1964. *On estimating regression*, Theory of Prob. and Appl. **9**, 141–142.

Spokoiny, V., W. Wang, and W. Härdle. 2011. *Local quantile regression*, SFB 649 Discussion Paper 2011-005, Humboldt-Universität zu Berlin.

Watson, Geoffrey S. 1964. *Smooth regression analysis*, Sankhyā: The Indian Journal of Statistics, Series A, 359–372.

Yu, Keming and MC Jones. 1998. *Local linear quantile regression*, Journal of the American statistical Association **93**, no. 441, 228–237.

Appendix

	Quantile	Est.	NP	SP-PL	SP-Add
Cooling Machine	90%	$\hat{\pi}_1$	0.031	0.034	0.087
		$\hat{\pi}_{01}$	0.031	0.032	0.072
		$\hat{\pi}_{11}$	0.023	0.104	0.240
	95%	$\hat{\pi}_1$	0.006	0.008	0.041
		$\hat{\pi}_{01}$	0.007	0.008	0.034
		$\hat{\pi}_{11}$	0.000	0.000	0.193
Dry Cooler	90%	$\hat{\pi}_1$	0.009	0.028	0.071
		$\hat{\pi}_{01}$	0.009	0.026	0.050
		$\hat{\pi}_{11}$	0.000	0.094	0.356
	95%	$\hat{\pi}_1$	0.003	0.007	0.023
		$\hat{\pi}_{01}$	0.003	0.007	0.018
		$\hat{\pi}_{11}$	0.000	0.000	0.227

Table 7: Estimated π's for the two machines and different models in the reference period.

	Quantile	Est.	Y_t	Y_t^{mean}	Y_t^{var}
Cooling Machine	95%	$\hat{\pi}_1$	0.016	0.198	0.086
		$\hat{\pi}_{01}$	0.017	0.137	0.073
		$\hat{\pi}_{11}$	0.000	0.447	0.224

Table 8: Estimated π's for the example given in section 3.4.

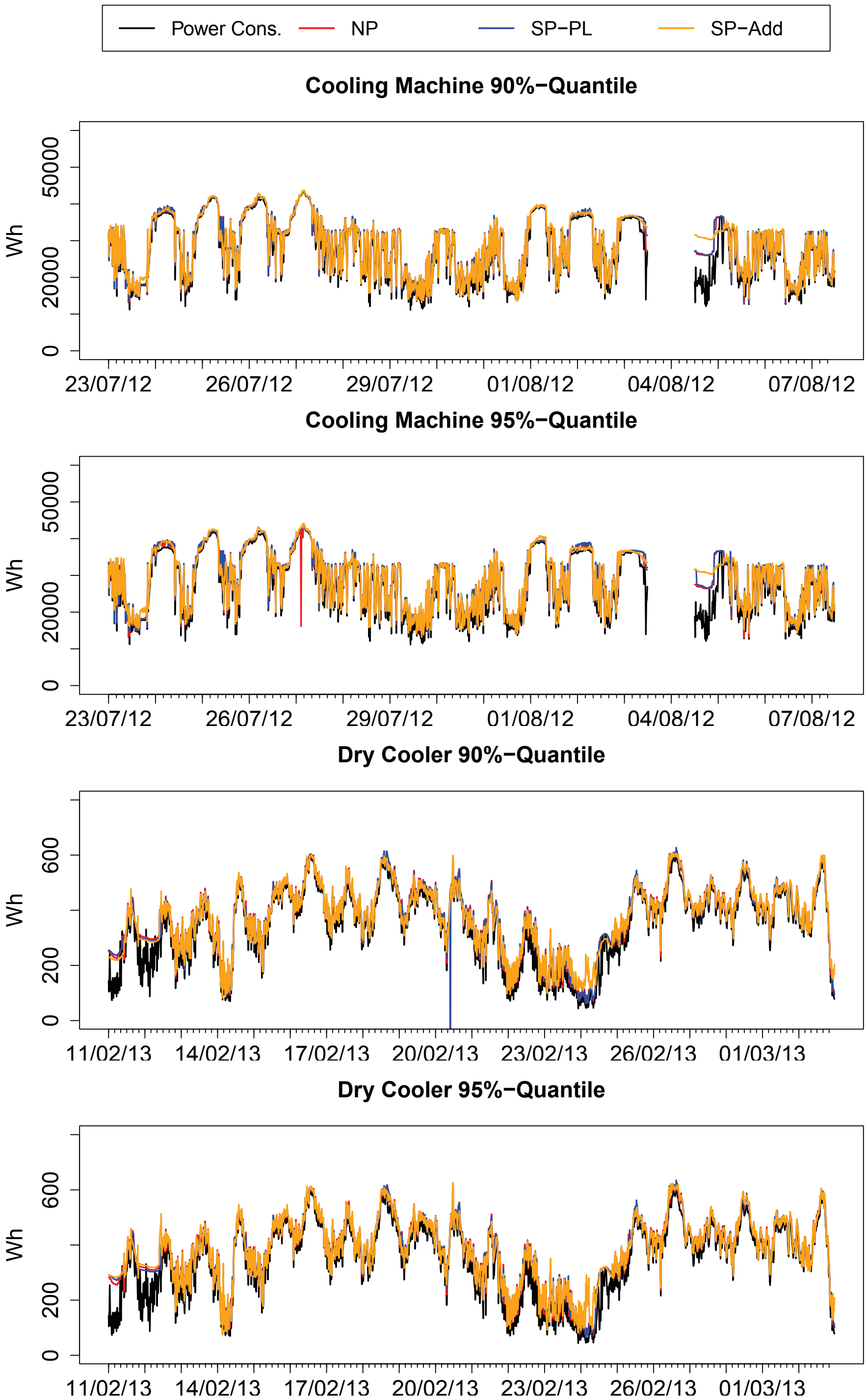
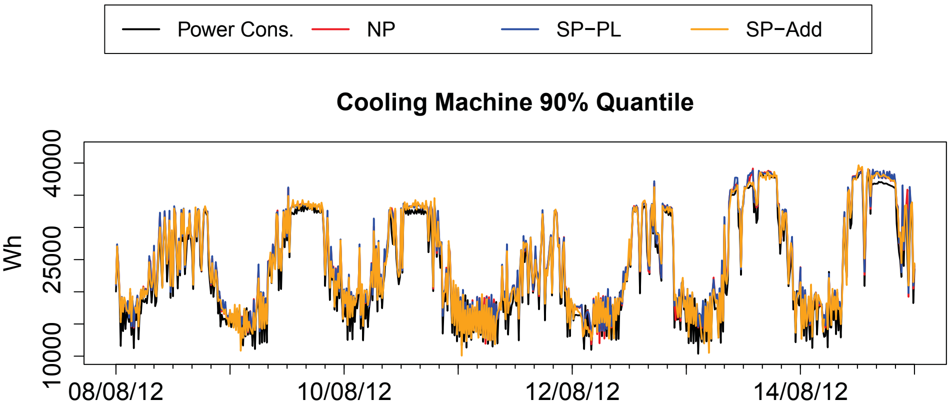
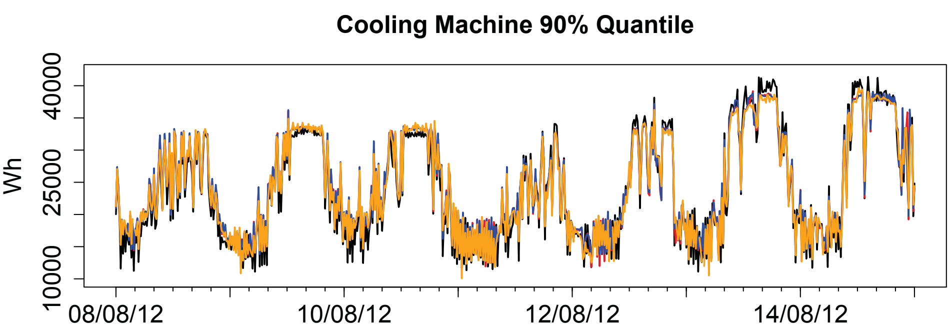


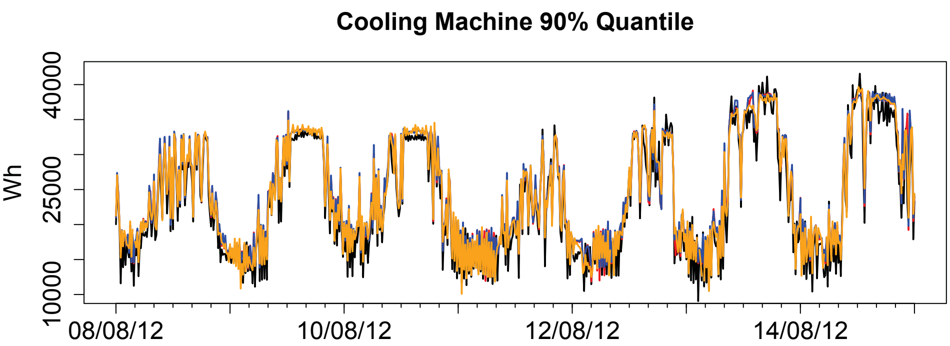
Figure 6: Realized power consumption and estimated quantiles for the two machines in the reference period (gap in cooling machine data, due to default of sensor).



(a) Original power consumption.

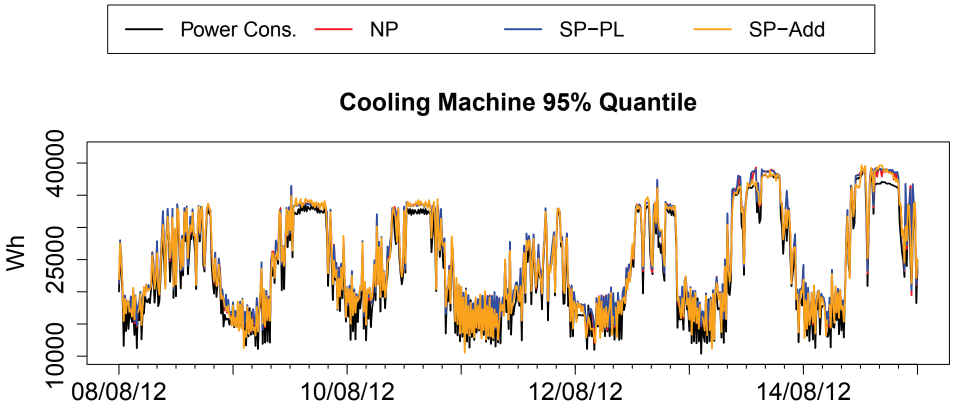


(b) Mean-altered power consumption.

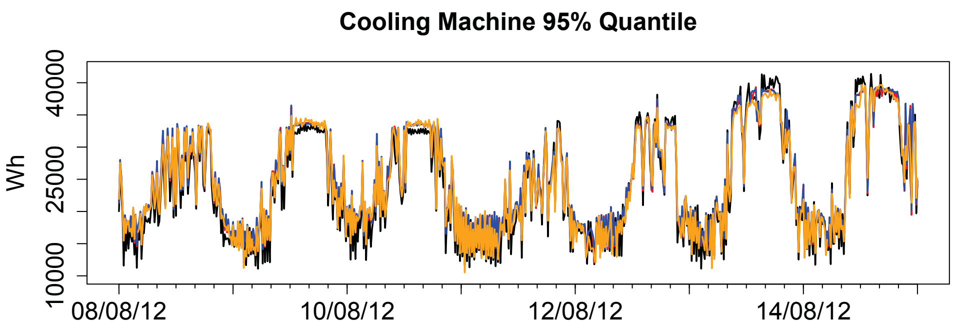


(c) Variance-altered power consumption.

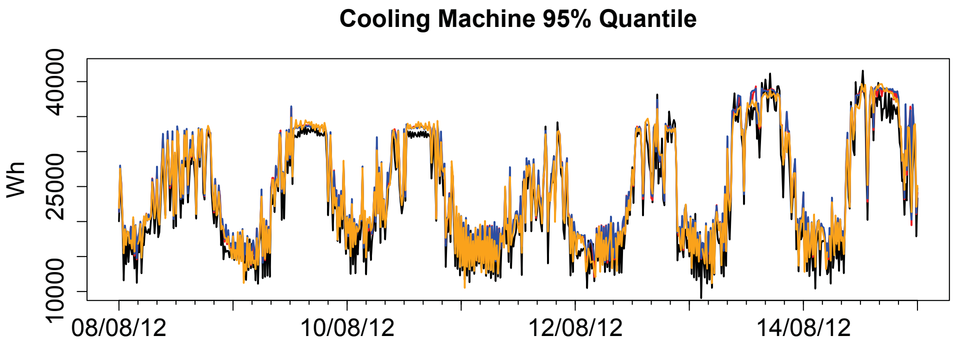
Figure 7: Different power consumption series of the example and estimated 90% quantile.



(a) Original power consumption.



(b) Mean-altered power consumption.



(c) Variance-altered power consumption.

Figure 8: Different power consumption series of the example and estimated 95% quantile.

Declaration of Authorship

I hereby certify that this master thesis has been composed by me and is based on my own work, unless stated otherwise. No other person's work has been used without due acknowledgement in this master thesis. All references have been quoted, and all sources of information, including graphs and data sets, have been specifically acknowledged.

Sven Ballentin

Berlin, February 13, 2014