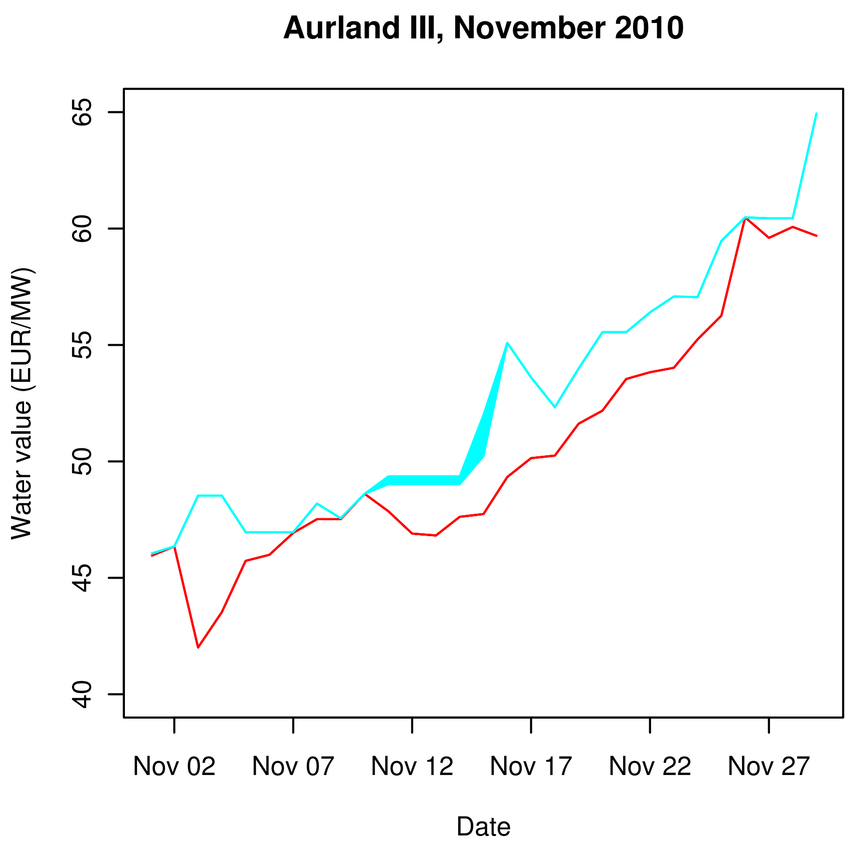
Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Note

Estimation of water values from live power production data



Note no SAMBA/05/11

Authors Egil Ferkingstad
Anders Løland

Date 31st January 2011

Norwegian Computing Center

Norsk Regnesentral (Norwegian Computing Center, NR) is a private, independent, non-profit foundation established in 1952. NR carries out contract research and development projects in the areas of information and communication technology and applied statistical modelling. The clients are a broad range of industrial, commercial and public service organizations in the national as well as the international market. Our scientific and technical capabilities are further developed in co-operation with The Research Council of Norway and key customers. The results of our projects may take the form of reports, software, prototypes, and short courses. A proof of the confidence and appreciation our clients have for us is given by the fact that most of our new contracts are signed with previous customers.

Norsk Regnesentral
Norwegian Computing Center
Postboks 114, Blindern
NO-0314 Oslo, Norway

Besøksadresse
Office address
Gaustadalléen 23
NO-0373 Oslo, Norway

Telefon · telephone
(+47) 22 85 25 00
Telefaks · telefax
(+47) 22 69 76 60

Internett · internet
www.nr.no
E-post · e-mail
nr@nr.no

Title	**Estimation of water values from live power production data**

Authors	**Egil Ferkingstad** <egil.ferkingstad@nr.no> **Anders Løland** <anders.loland@nr.no>

Date	31st January 2011
Publication number	SAMBA/05/11

Abstract

The Ably Power Live system (http://www.ablylive.com) provides real-time data on power production for several Norwegian hydro power plants. Of particular interest is estimation of the so-called *water value*, defined as the opportunity cost of using one unit of water for power production instead of leaving it in the magazine. If power plant managers were rational, one would expect them to produce power if the electricity price is above the water value, and shut down production otherwise. Thus, knowledge of water values would give crucial insight into the power markets as well as the strategic planning and operations of hydro power plants. We have developed a method for estimation of water values based on power production and electricity price data. Our method is based on a segmentation algorithm fitting a piecewise constant function to the production data.

Keywords	Power markets, breakpoint estimation, process segmentation
Target group	Ably AS
Availability	Confidential
Project	Vannverdi
Project number	220487
Research field	Finance, insurance and commodity markets
Number of pages	13

Contents

Summary

We present a method for daily water value estimation based on production and price data. The following is an outline of the algorithm, with a running example of how it works:

1. Define *production intervals*, for example "zero, positive" or "negative, zero, low, moderate, high" by specifying limits of each interval.

   Example:

   Zero: (0,100)
   Positive: (100,500)

2. Estimate a piecewise constant function based on the production data.

   Example:

   Hour 0-6: 10 MW
   Hour 7-20: 200 MW
   Hour 21-23: 5 MW

3. For each time point, determine the current production interval based on the estimated piecewise constant function and the given production interval limits.

   Example:

   Hour 0-6: Zero
   Hour 7-20: Positive
   Hour 21-23: Zero

4. For each production interval i, estimate an interval $[w_i^{min}, w_i^{max}]$ for the water value w_i by the *minimum-value method*:
   Consider all prices at each production interval. For each production interval, except the lowest interval ($i = 0$), the interval estimate $[w_i^{min}, w_i^{max}]$ is given by

   $$\begin{aligned} w_i^{min} &= \text{maximal price for all intervals below interval } i, \\ w_i^{max} &= \text{minimum price at interval } i, \end{aligned}$$

   where we require $w_i^{min} \leq w_i^{max}$.

   Example:

   Zero: Water value not defined
   Positive: $[w_1^{min}, w_1^{max}] = [43, 48]$

1 Introduction

The Ably Power Live system (`http://www.ablylive.com`) delivers power plant production data in real-time (currently at a four-minute resolution). Hourly price data is available form Nord Pool Spot (`http://www.nordpoolspot.com/`). We here present a methodology for estimating daily power-plant specific water values based on the production and price data.

2 Methodology

Visual inspection of the power production data suggests that they can be modelled well by a piecewise constant function, with breakpoints corresponding to changes in production interval. To estimate this function, we have used the segmentation algorithm described by Picard et al. (2005), which we will now explain.

2.1 Estimating a piecewise constant function with K segments

Let $Y_t = y_t, t = 1, \ldots, n$ be the production at time t, and let $(t_1, t_2, \ldots, t_{K-1})$ be the $K-1$ unknown breakpoints, where K is known. Assume that

$$\forall t \in (t_{k-1}, t_k],\ Y_t = \mu_k + \varepsilon_t,\ \varepsilon_t \sim N(0, \sigma^2).$$

If the breakpoints were known, the variance σ and the segment means μ_k could easily be estimated in the usual way. Thus, the real problem lies in estimating the breakpoints $(t_1, t_2, \ldots, t_{K-1})$. We will now describe an exact maximum likelihood solution. The log-likelihood can be written as a sum $L_K = \sum_{k_1}^{K} l_k$, where

$$l_K = -\frac{1}{2} \sum_{t=t_{k-1}+1}^{t_k} \left\{ \log(2\pi\sigma^2) + \left[\frac{y_t - \mu_k}{\sigma} \right]^2 \right\}.$$

Unless K is very small, brute-force maximisation is impossible, since the number of different segmentations is $\binom{n-1}{K-1}$ and the computational complexity is $O(n^K)$. Fortunately, the exact solution can be found using a dynamic programming algorithm with complexity $O(n^2)$: Let $\hat{L}_{k+1}(i, j)$ be the maximum likelihood for partitioning $Y_{i+1}, \ldots, Y_j$ into $k+1$ segments (with k breakpoints). First, for all $0 \leq i < j \leq n$, we calculate the maxima $\hat{L}_1(i, j)$ for all one-segment partitions. Then, for $k = 1, 2, \ldots, K-1$, and $j = 2, \ldots, n$

$$\hat{L}_{k+1}(1, j) = \max_h \{\hat{L}_k(1, h) + \hat{L}_1(h+1, j)\},$$

the maximised full likelihood is $\hat{L}_K = \hat{L}_K(1, n)$, and the breakpoints are the sequence of maximising h in the recursion for $\hat{L}_K(1, n)$. For example, for $K = 3$,

$$\begin{aligned} \hat{L}_3(1, n) &= \max_{h_2}\{\hat{L}_2(1, h_2) + \hat{L}_1(h_2 + 1, n)\} \\ &= \max_{h_2}\{\max_{h_1}\{\hat{L}_1(1, h_1) + \hat{L}_1(h_1 + 1, h_2)\} + \hat{L}_1(h_2 + 1, n)\} \end{aligned}$$

and the breakpoints are the $\hat{h}_1, \hat{h}_2$ which maximise $\hat{L}_3 = \hat{L}_3(1, n)$.

2.2 Choosing the number K of segments

In practice, the optimal number $\hat{K}$ of segments is unknown. Assume that K_{max} is the maximal number of segments that we are willing to consider, so $K = 1, 2, \ldots, K_{max}$. The maximised log-likelihood $\hat{L}_K$ measures the fit of the model with K segments. However, $\hat{L}_K$ is increasing with K, so it cannot be used by itself to find $\hat{K}$, since this will always give $\hat{K} = K_{max}$. The problem of choosing $\hat{K}$ can be seen as a model selection problem, and classical methods such as the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) could be used. These methods penalise for model complexity, so that we avoid always choosing the most complex (and best-fitting) model. For example, using BIC, we would maximise

$$BIC(K) = \max(L_K) - (\log n)p,$$

where p is the total number of parameters. In our case $p = 2K$, since we need to estimate $K - 1$ breakpoint locations, the K segment means $\mu_1, \ldots, \mu_K$ and the variance σ.

Using AIC or BIC often leads to an overestimation of the number of segments. As an alternative, Picard et al. (2005) suggest looking at $\hat{L}_K$ as a function of K, and choosing $\hat{K}$ where $\hat{L}_K$ ceases to increase significantly. The rate of increase can be measured be studying the *second-order differences* (i.e., differences of differences) of $\hat{L}_K$ as K increases. The second-order differences are given by

$$D_K = (\hat{L}_{K+1} - \hat{L}_K) - (\hat{L}_K - \hat{L}_{K-1}) = \hat{L}_{K+1} - 2\hat{L}_K + \hat{L}_{K-1}.$$

Note that D_K (seen as a function of K) is the discrete equivalent of the second derivative of $\hat{L}_k$. Even though the first-order differences $\hat{L}_{K+1} - \hat{L}_K$ are always positive, the second order differences may be negative, meaning that the rate of increase is decreasing. To apply this to choosing $\hat{K}$, it seems natural to choose the largest K such that D_K is below some (negative) threshold:

$$\hat{K} = \max\{K = 1, 2, \ldots, K_{max} : D_K < s \cdot n\},$$

where n is the number of data points and s is some user-specified constant. We follow Picard et al. (2005) and choose $s = -0.5$.

2.3 Inferring water values

We need to estimate water values for each day. Our estimates are based on price data $p_1, p_2 \dots, p_n$ and the piecewise constant function $\hat{y}_1, \dots, \hat{y}_n$ based on the production data from Sections 2.1-2.2. In addition, we need to specify *production intervals* for which different water values are defined: For given constants $\gamma_0 < \cdots < \gamma_m$ (where we allow $\gamma_0 = -\infty$ and $\gamma_m = \infty$) we are at production interval i at time t if $\gamma_i \leq \hat{y}_t < \gamma_{i+1}$. Denote the production interval at time t by i_t. For each interval $i > 0$ a corresponding (daily) water value w_i is defined. The number of intervals as well as the constants γ_i should be based on knowledge of the power plant (such as the installed capacity and the number of turbines). Often, it is most natural to only consider two intervals: interval 0 representing zero production, and interval 1 representing positive production, where γ_1 represents the lowest smoothed production value that we consider to be positive. In this case we only estimate a single daily water value w_1. However, the added flexibility in allowing for more than two production intervals can be useful in some cases. A description of our two proposed water value estimation methods follows.

2.3.1 Breakpoint-change method

Let the estimated breakpoints (see Sections 2.1-2.2) be $\hat{t}_1, \hat{t}_2, \dots, \hat{t}_{K-1}$. We have electricity prices $p_1, \dots, p_n$ and estimated production intervals $\hat{i}_1, \dots, \hat{i}_n$. Define a neighbourhood of plus/minus c minutes around each breakpoint, such that neighbourhood k is given by $[\hat{t}_k - c, \hat{t}_k + c]$. We propose $c = 60$ as default value (c should be at least 30). It seems irrational to decrease production if prices increase, or increase production if prices decrease. Therefore, we want to exclude breakpoints where the change in production interval and change in price have opposite sign. To achieve this, we define a breakpoint k as *valid* if and only if

$$(p_{t_k+c} - p_{t_k-c})(\hat{i}_{t_k+c} - \hat{i}_{t_k-c}) > 0,$$

and *invalid* otherwise (note that a breakpoint is invalid if $\hat{i}_{t_k+c} = \hat{i}_{t_k-c}$). Each valid breakpoint gives an interval estimate $[\min_{t_k-c<t<t_k+c} p_t, \max_{t_k-c<t<t_k+c} p_t]$ for water value $w_{\max(\hat{i}_{t_k-c}, \hat{i}_{t_k+c})}$ (i.e., the water value at the neighbourhood border having the highest production interval). If there are several possible interval estimates for w_i, we choose the estimate which gives the narrowest interval on the current day. Thus, for each day, we have interval estimates $[\hat{w}_1^{min}, \hat{w}_1^{max}], \dots, [\hat{w}_m^{min}, \hat{w}_m^{max}]$ for the water value at each interval for which there is a valid breakpoint. Further, we demand that $\hat{w}_1^{min} \leq \hat{w}_1^{max} \leq \hat{w}_2^{min} \leq \hat{w}_2^{max} \cdots \leq \hat{w}_m^{min} \leq \hat{w}_m^{max}$, and replace $(\hat{w}_1^{min}, \hat{w}_1^{max}, \hat{w}_2^{min}, \hat{w}_2^{max}, \dots, \hat{w}_m^{min}, \hat{w}_m^{max})$ by cumulative maxima if this is not the case. If point estimates $\hat{w}_i$ are needed, the most natural choice is $\hat{w}_i = \hat{w}_i^{max}$. Figure 1 illustrates the method.

2.3.2 Minimum-value method

Recall that the estimated production intervals are denoted $\hat{i}_1, \ldots, \hat{i}_n$. As in Section 2.3.1, we estimate intervals $[\hat{w}_1^{min}, \hat{w}_1^{max}], \ldots, [\hat{w}_m^{min}, \hat{w}_m^{max}]$ and demand that the sequence $(\hat{w}_1^{min}, \hat{w}_1^{max}, \ldots, \hat{w}_m^{min}, \hat{w}_m^{max})$ is non-decreasing. For each interval i, a natural upper limit estimate $\hat{w}_i^{max}$ of the water value w_i is given by the minimum price for which we have production at interval i:

$$\hat{w}_i^{max} = \min_{\hat{i}_t = i} p_t.$$

Conversely, a natural lower limit estimate $\hat{w}_i^{min}$ is given by

$$\hat{w}_i^{min} = \max_{\hat{i}_t < i} p_t.$$

We modify the estimates above by excluding prices at valid breakpoints from the maximum/minimum (to avoid always getting $\hat{w}_i^{min} = \hat{w}_i^{max}$). Also, if $\hat{w}_i^{min} > \hat{w}_i^{max}$, we simply set $\hat{w}_i^{min} = \hat{w}_i^{max}$. The method is illustrated in Figure 2.

2.3.3 Online estimation

For both of the estimation methods presented above, the current estimates of the daily water values may be updated as new data is gathered. For example, for the data shown in Figure 1, at 10:00 our best interval estimate would be given by the blue dashed lines, since we have only seen the first breakpoints. However, at 23:00, we have seen both breakpoints, and the current best estimate (which is also the final daily estimate) is given by the red dashed lines.

2.3.4 Water value time series

It is interesting to study the development of water values over time. As an example, Figure 3 shows daily water value estimates for the plant Aurland III in November 2010. A rising trend is apparent.

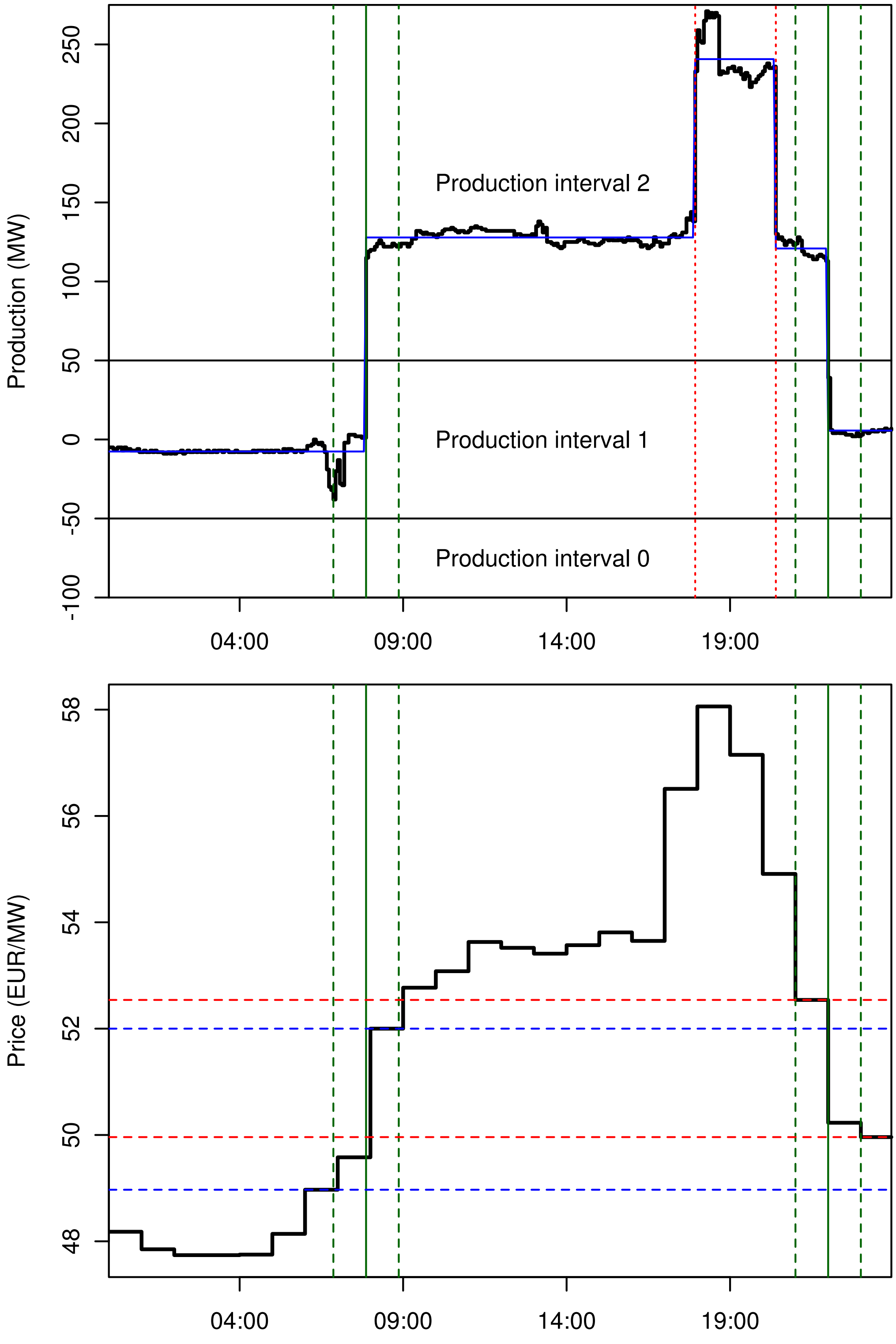


Figure 1. Illustration of the breakpoint-change method for water value estimation, with production and price data from Aurland III on 15.11.2010. The estimated piecewise constant function is shown in blue. The solid green vertical lines show the locations of the two valid breakpoints, with neighbourhoods shown by dashed green lines, and the two dotted red lines indicate invalid breakpoints. At each valid breakpoint, a possible water value interval is determined, shown by the blue and red dashed lines, respectively. The estimated daily water value at production interval 2, $[\hat{w}_2^{min}, \hat{w}_2^{max}]$ = [49.96, 52.54], is at the red dashed lines, since this interval is the narrowest of the two.

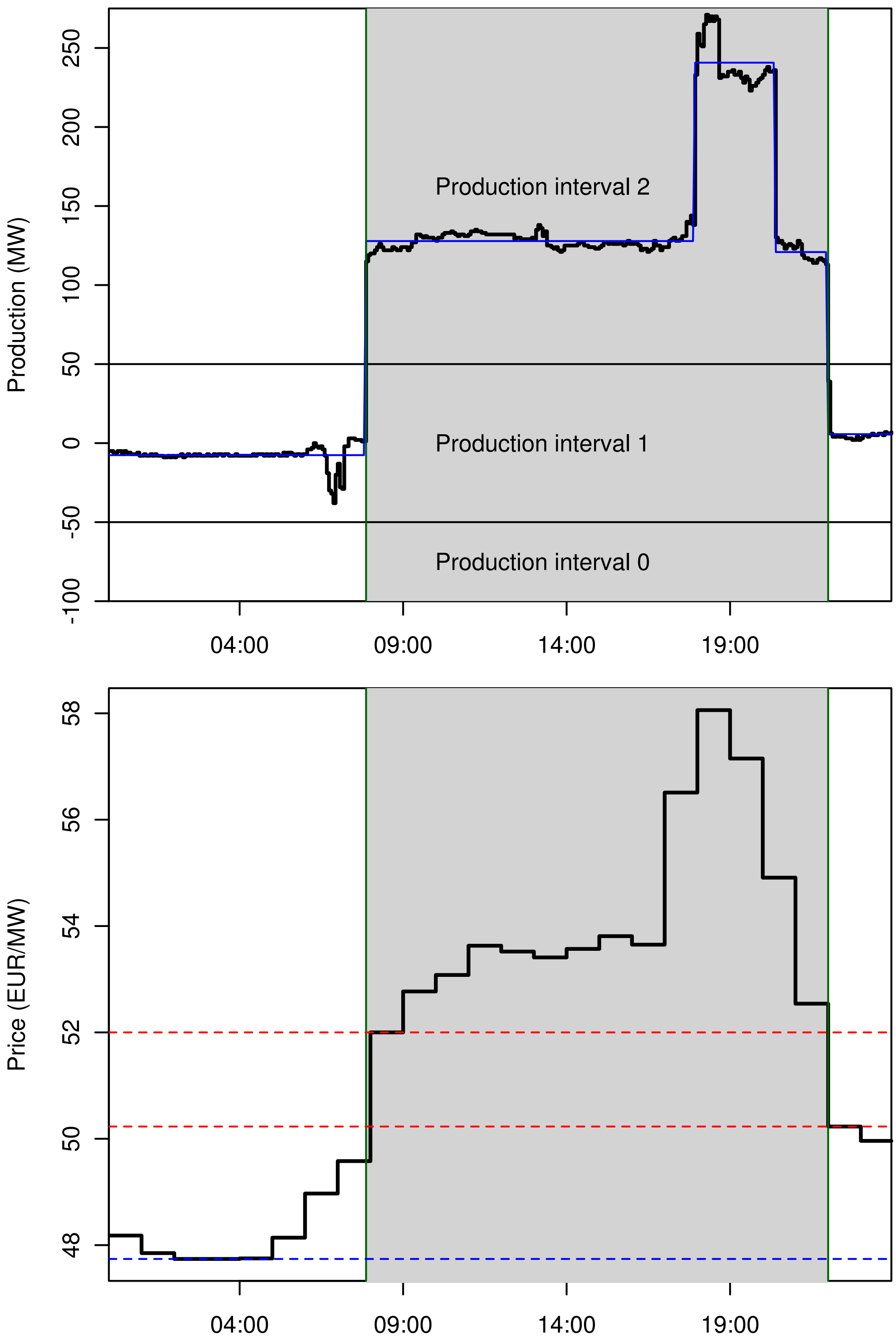


Figure 2. Illustration of the minimum-value method for water value estimation, with production and price data from Aurland III on 15.11.2010. The grey area indicates the time period when the production is in interval 2; in the white area production is in interval 1. The lower limit of the interval is given by the maximum price in the white area, while the upper limit is given by the minimum price in the grey area. The estimate (shown by the dashed red lines) at production interval 2 is $[\hat{w}_2^{min}, \hat{w}_2^{max}]$ = [50.23, 52.00]. For production interval 1, we find an upper limit for the water value at the minimum price in the white area (shown by the dashed blue line), so $\hat{w}_1^{max}$ = 47.74.

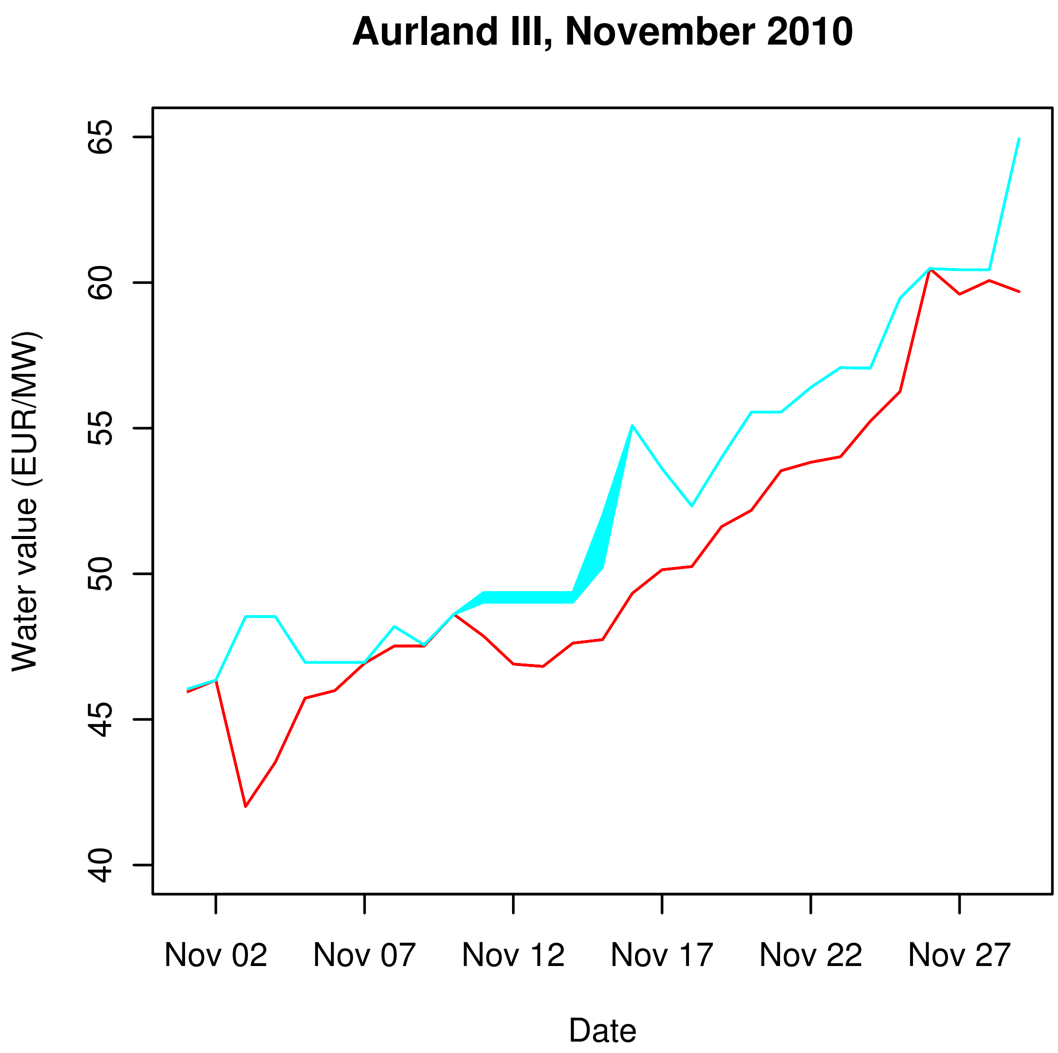


Figure 3. Estimated water values (using the minimum-value method) for Aurland III in November 2010. The red curve shows water values (interval estimates) for production interval 1, while the blue curve shows water values for production interval 2. Notice that the interval estimates often collapse to a single point.

References

Picard, F., Robin, S., Lavielle, M., Vaisse, C., and Daudin, J. (2005). A statistical approach for array CGH data analysis. *BMC Bioinformatics*, 6(1):27.