stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE BESLISKUNDE BW 153/81 DECEMBER
(DEPARTMENT OF OPERATIONS RESEARCH)

J.D. VAN DER BIJ & J.H. VAN SCHUPPEN

SHORT TERM ADAPTIVE PREDICTION OF RAILWAY POWER DEMAND

Preprint

kruislaan 413 1098 SJ amsterdam

Printed at the Mathematical Centre, 413 Kruislaan, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

1980 Mathematics subject classification: 93E11, 93E03, 93E12

Short term adaptive prediction of railway power demand*)

by

J.D. van der Bij**)
J.H. van Schuppen

ABSTRACT

The short term prediction problem is considered for power demand at railway energy supply stations of the dutch national railway company. A model and an adaptive prediction algorithm are presented. The algorithm is based on a new selftuning predictor for the output of a Gaussian system.

KEY WORDS & PHRASES: *railways, power demand prediction, adaptive stochastic filtering, selftuning predictor*

---

*) This report will be submitted for publication elsewhere.

**) Technical University of Eindhoven, The Netherlands

1. INTRODUCTION

The purpose of this paper is to present an algorithm for short term prediction of power demand at energy supply stations of the dutch national railway system.

The N.V. Nederlandse Spoorwegen, the dutch national railway company, uses electrical power for its rolling stock, for safety purposes, and for buildings and related objects. Electric power is obtained locally from the provincial power companies. At energy supply stations, that are located at about every 22 kilometers of the network, the power is converted from AC to DC and supplied to the railway power network. This power is mainly used for traction, but also for light and heating. The power demand of the railway power network at an energy supply station has rather peculiar characteristics. One characteristic is the peak demand in the hours between 7 and 9, and 16 and 19, which may run as high as twice the demand at 12 o'clock noon. Another characteristic is the relatively high variance of the data.

What is the motivation for this investigation? The energy cost that the railway company is faced with is based on the total power used, and on the peak load used in certain time periods. The price structure varies between the different power companies. With the objective to save on energy costs the railway company is considering to install energy storage facilities at the energy supply stations to store energy in off-peak hours with which to flatten the peak loads. The reduced peak demand of the railway company on the power network is financially advantageous for the former while increasing the efficiency for the latter. A feasibility study has been started to evaluate this plan. One aspect of the scheme is to develop a control algorithm for the energy storage and use. For this, two subproblems have been distinguished namely: 1) to develop an algorithm for short term prediction of power demand at an energy supply station; 2) to develop a control algorithm for the energy storage. The prediction algorithm is likely to play a part in the control algorithm.

The problem is then to develop an algorithm for short term prediction of the power demand at energy supply stations. The short term should be interpreted as ranging from 15 minutes to 3 hours. Moreover, at each time moment predictions are required for one to say t_1 periods ahead. This type of prediction will be called a multi-step prediction. As a basis for prediction only power data are to be used. A general formulation of the algorithms is required that can be used at the different locations of the railway company.

The above formulated problem is similar to the prediction problem of power demand in a national electric network. For the latter problem there is a considerable literature. In general the load is decomposed into a nominal and a residual part. The estimation of the nominal part may be done via Fourier series expansion including adaptation of the coefficients, exponential smoothing, Wiener filtering, Kalman filtering, or an application of recursive least squares [1,3,4,6,7,13,17,18,21,22]. Estimation of the residual part may be accomplished via Karhunen-Loève expansion, Kalman filtering, or adaptive Kalman filtering. For the problem under consideration it has been decided to use the adaptive prediction approach.

The adaptive prediction algorithm for power demand data is based on the Kalman filter and on an adaptive prediction algorithm for Gaussian systems. The problem of constructing and evaluating adaptive prediction algorithms for Gaussian systems has received quite some attention recently [1,5,10,11, 13,19,20,24]. Most of the references quoted consider the stochastic prediction problem for a fixed time increment, while here multistep predictions are required. Therefore a new algorithm for adaptive stochastic prediction of the output of a Gaussian system is derived in which multistep predictions are computed in a recursive manner. The state space view is emphasized.

The problem formulation is presented in Section 2, while the selftuning predictor for Gaussian systems is derived in Section 3. A model for power demand prediction is developed in Section 4, based on three different time scales. The model is an extension of one proposed by BOHLIN [1]. A computer program for this algorithm has been written, and has been implemented on the computer of the railway company. Some numerical results are presented in Section 5.

Acknowledgements are due to Mr. K.G.M. Jeurissen of the N.V. Nederlandse Spoorwegen for his cooperation on the problem, and prof. J.C. Willems of the University of Groningen for helpful advice.

2. PROBLEM FORMULATION

As explained in Section 1 electric energy is supplied to the railway power network at energy supply stations. To produce predictions of the power demand in the railway power network one would need, in principle, information about the total network. As a first step in model reduction it is therefore assumed that energy supply stations only serve a local area and do not interchange energy with other stations. A detailed model for the power demand at an energy supply station is developed in Section 4.

Over which time intervals must predictions be made? The railway company has let it be known that a general algorithm is desired because the data processing differs between the local power companies. Therefore some remarks on prediction intervals follow.

In a discrete time prediction model one may distinguish four different time intervals. These intervals are:

Δs - the time interval over which data are or become available;

Δt - the time interval between the moments that predictions must be made;

Δf - the time interval between the time of a prediction and the start of the prediction interval Δp;

Δp - the time interval for which a prediction must be made.

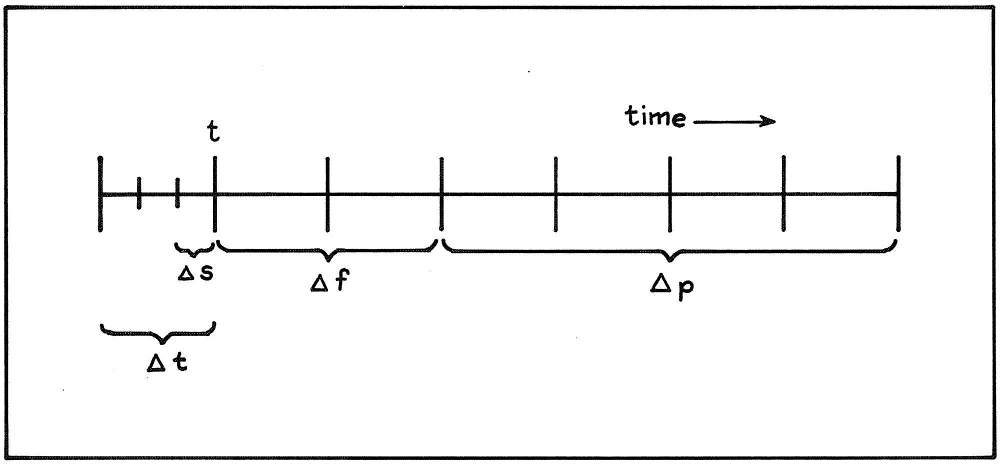

Figure 1. Time intervals of the prediction problem

As an example one may think of $(\Delta s,\Delta t,\Delta f,\Delta p) = (5,15,30,60)$ minutes see figure 1. To simplify the algorithm somewhat the following assumptions are made: there exist $k_t,k_f,k_p \in \{0,1,2,\ldots\}$ such that $\Delta t = k_t\Delta s$, $\Delta f = k_f\Delta t$, $\Delta p = k_p\Delta t$. Under these assumptions the prediction algorithm will be designed such that it aggregates k_t data points of power demand in intervals of Δs minutes to Δt minute totals, and at every Δt minutes it predicts from $k_f + 1$ to $k_f + k_p$ time periods of Δt minutes ahead. The general problem will then be taken to predict at every time step from one to say t_1 time periods ahead.

A s-step prediction, with $s \in \{1,2,\ldots\}$ fixed is defined to be a prediction s steps ahead. A *multistep* prediction, say with t_1 steps, is defined to be a prediction for $1,2,\ldots,t_1$ steps ahead.

2.1. PROBLEM. To produce multistep predictions of the power demand at an energy supply station on the basis of load data only.

3. ADAPTIVE PREDICTION FOR GUASSIAN SYSTEMS

The power demand prediction algorithm to be presented in Section 4 is based on a selftuning adaptive prediction algorithm for Gaussian systems. The latter algorithm will be derived below.

Let (Ω,F,P) be a complete probability space and $T = Z = \{\ldots,-1,0,1,\ldots\}$ the time index set. A Gaussian random variable $x: \Omega \to R^n$ with parameters $\mu \in R^n$ and $Q \in R^{n\times n}$, satisfying $Q = Q^T \geq 0$, is denoted by $x \in G(\mu,Q)$. A Gaussian white noise process with intensity $V: T \to R^{n\times n}$, satisfying for all $t \in T$, $V(t) = V(t)^T \geq 0$, is a stochastic process $v: \Omega \times T \to R^n$ such that $\{v(t), t \in T\}$ is an independent sequence and for all $t \in T$, $v(t) \in G(0,V(t))$.

3.1. ASSUMPTION. Let be given a stationary real valued Gaussian process on T with zero mean function and covariance function $c: T \to R$. Assume that the process has a rational spectral density and that $\lim_{t\to\infty} c(t) = 0$.

3.2. PROBLEM. The *multistep stochastic prediction problem* for the Gaussian process defined in 3.1, denoted by y, is to determine for all $t \in T$ and $s \in Z_+$ an expression for

$$E[\exp(iuy(t+s)) \mid F_t^y],$$

where $F_t^y = \sigma(\{y(\tau), \forall\tau \leq t\})$ is the σ-algebra generated by y up to $t \in T$. □

To solve this problem a representation of the Gaussian process is needed. In stochastic system theory the concept of a stochastic dynamical system has been defined. Loosely speaking such a system consists of a state and output process such that for all time moments the past and future of these processes are conditional independent given the state at that time moment. In the following only the definition of a Gaussian system is needed.

3.3. DEFINITION. A *Gaussian stochastic dynamical system,* for short a Gaussian system, is a collection

$$\{T, R^n, B_n, R^k, B_k, R^m, B_m, A, B, C, D, V\}$$

where $T = Z$, $n,k,m \in Z_+$, R^n, B_n is the n-dimensional Euclidean space with its Borel σ-algebra, $A: T \to R^{n\times n}$, $B: T \to R^{n\times m}$, $C: T \to R^{k\times n}$, $D: T \to R^{k\times m}$, $V: T \to R^{m\times m}$ such that for all $t \in T$, $V(t) = V(t)^T \geq 0$. If (Ω, F, P) is a complete probability space and $v: \Omega \times T \to R^m$ a Gaussian white noise process with intensity V, then define $x: \Omega \times T \to R^n$, $y: \Omega \times T \to R^k$

$$x(t+1) = A(t)x(t) + B(t)v(t),$$

$$y(t) = C(t)x(t) + D(t)v(t).$$

The process x will be called the *state process* and y the *output process*. A *time-invariant Gaussian system* is a Gaussian system for which A, B, C, D and V are time-invariant. □

It may be shown that a Gaussian system is a stochastic dynamical system as defined above 3.3. The question now is whether the Gaussian process defined in 3.1 can be represented as the output of a Gaussian system?

3.4. PROPOSITION. *Given the Gaussian process specified in* 3.1. *Then there exists a time-invariant Gaussian system*

$$\{T, R^n, B_n, R^k, B_k, R^m, B_m, A, B, C, D, V\}$$

(1) $$x(t+1) = Ax(t) + Bv(t),$$

(2) $$y(t) = Cx(t) + Dv(t),$$

such that the process y *defined by* (1,2) *and the Gaussian process defined in* 3.1 *have the same family of finite dimensional distributions. Then one calls this Gaussian system a weak Gaussian stochastic realization of the process specified by* 3.1. *Furthermore the dimension* n *of the state space* (R^n, B_n) *may be chosen to be minimal. For such a minimal stochastic realization* (A,C) *is an observable pair, and* A *is exponentially stable.*

PROOF. [8, thm. 3.4; 9, thm. 8.9]. □

The stochastic realization (1,2) of the given process is not unique. Consider the asymptotic Kalman filter for the Gaussian system (1,2)

$$\hat{x}(t+1|t) = A\hat{x}(t|t-1) + K(y(t) - C\hat{x}(t|t-1))$$

where $\hat{x}(t+1|t) = E[x(t+1) \mid F_t^y]$. This may also be written as

(3) $$\hat{x}(t+1|t) = A\hat{x}(t|t-1) + K\bar{v}(t), \hat{x}(0|-1),$$

(4) $$y(t) = C\hat{x}(t|t-1) + \bar{v}(t),$$

where $\bar{v}: \Omega \times T \to R$ is the innovation process which is known to be a Gaussian white noise process. Clearly (3,4) is another weak Gaussian stochastic realization of the given process. Since in this realization the state process is such that for all $t \in T$, $\hat{x}(t+1|t)$ is F_t^y measurable it will be called the past-output induced stochastic realization of the process specified in 3.1. On the basis of the output process y only the realizations (1,2) and (3,4) are indistinguishable. Because our goal is output prediction, attention is in the following restricted to the past-output induced stochastic realization (3,4).

What is the solution of the multistep stochastic prediction problem for the Gaussian process y having the stochastic realization (3,4)?

3.5. PROPOSITION. *Consider the past-ouput induced stochastic realization of the Gaussian process defined in* 3.1, *say*

$$\{T,R^n,B_n,R,B,R,B,A,K,C,1,\bar{q}\}$$

$$\hat{x}(t+1|t) = A\hat{x}(t|t-1) + K\bar{v}(t),$$

$$y(t) = C\hat{x}(t|t-1) + \bar{v}(t).$$

The solution to the multistep stochastic prediction problem for the given process is given by, for $t \in T$, $s \in Z_+$,

$$E[\exp(iuy(t+s)) \mid F_t^y] = \exp(iu\hat{y}(t+s|t) - \frac{1}{2} u^2 q(t+s|t))$$

where $\hat{y}$ and q are determined by

(5) $$\hat{x}(t+1|t) = A\hat{x}(t|t-1) + K(y(t) - C\hat{x}(t|t-1))$$

(6) $$Q(t+1|t) = 0,,$$

(7) $$\hat{x}(t+s+1|t) = A\hat{x}(t+s|t), \hat{x}(t+1|t),$$

(8) $$Q(t+s+1|t) = AQ(t+s|t)A^T + \bar{q}KK^T, Q(t+1|t) = 0,$$

(9) $$y(t+s|t) = C\hat{x}(t+s|t),$$

(10) $$q(t+s|t) = CQ(t+s|t)C^T + \bar{q}.$$

PROOF. This follows from applying to (3.4) the Kalman filter [14]. □

From now on the time index set is taken to be $T = \{0,1,2,...\}$, and the representation (3,4) is initialized at $\hat{x}(0|-1) = 0$. Because A is stable the effect of the initial condition is negligible in the long run.

In the following an alternative representation of the multistep predictor is needed that is linear in the parameters.

3.6. PROPOSITION.

a) *The one-step output predictor or Kalman filter*

$$\hat{x}(t+1|t) = A\hat{x}(t|t-1) + K(y(t) - C\hat{x}(t|t-1)), \quad \hat{x}(0|-1) = 0,$$
$$\hat{y}(t+1|t) = C\hat{x}(t+1|t),$$
$$q(t+1|t) = \bar{q},$$

has the equivalent representation

(11) $$h(t+1|t) = L_n h(t|t-1) + M_n \hat{y}(t|t-1) + N_n y(t), \quad h(0|-1) = 0,$$

(12) $$\hat{y}(t+1|t) = h(t+1|t)^T p,$$

(13) $$q(t+1|t) = \bar{q},$$

where $h: \Omega \times T \to R^{2n}$

(14) $$h(t+1|t)^T = (-\hat{y}(t|t-1), \dots, -\hat{y}(t-n+1|t-n), y(t), \dots, y(t-n+1)),$$

$p \in R^{2n}$ *is determined by* A, K, *and* C, *the elements of which are denoted by*

(15) $$p^T = (\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n),$$

$$L_n = \left(\begin{array}{ccc|ccc} 0 & \dots & 0 & & & \\ & & \vdots & & 0 & \\ & I_{n-1} & 0 & & & \\ \hline & & & 0 & \dots & 0 \\ & 0 & & & & \vdots \\ & & & & I_{n-1} & 0 \end{array}\right) \in R^{2n \times 2n},$$

(16) $$M_n = -e_1 \in R^{2n}, \quad N_n = e_{n+1} \in R^{2n}$$

$e_j \in R^{2n}$ *being the* j-*th unit vector.*

b) *The multistep predictor* (7) - (10) *has the equivalent representation*

(17) $$hh(t+s+1|t) = L_n hh(t+s|t) + (M_n+N_n)\hat{y}(t+s|t),\quad hh(t+1|t) = h(t+1|t),$$

(18) $$\hat{y}(t+s+1|t) = hh(t+s+1|t)^T p,$$

(19) $$Q^r(t+s+1|t) = F(p)Q^r(t+s|t)F(p)^T + \bar{q}G(p)G(p)^T,\quad Q^r(t+1|t) = 0,$$

(20) $$q(t+s+1|t) = H(p)Q^r(t+s+1|t)H(p)^T + \bar{q},$$

where, if p *is given by* (15), *then*

(21) $$F(p) = \begin{pmatrix} -\alpha_1+\beta_1 & 1 & & 0 \\ \vdots & 0 & \ddots & \\ & & & 1 \\ -\alpha_n+\beta_n & 0 & \dots & 0 \end{pmatrix},\quad G(p) = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix},$$

(22) $$H(p) = (1 \quad 0 \quad \dots \quad 0).$$

c) *The output process* y *has the representation*

(23) $$y(t) = h(t|t-1)^T p + \bar{v}(t).$$

PROOF.

a) One has

$$\hat{x}(t+1|t) = (A-KC)\hat{x}(t|t-1) + Ky(t),\quad \hat{x}(0|-1) = 0,$$

$$\hat{y}(t+1|t) = C\hat{x}(t+1|t).$$

Because by 3.4 (A,C) is an observable pair, so is (A-KC,C). Then there exists a state space transformation such that the above form is input-output equivalent to

$$\hat{r}(t+1|t) = \begin{pmatrix} -\alpha_1 & 1 & & 0 \\ \vdots & 0 & \ddots & \\ & & & 1 \\ -\alpha_n & 0 & \dots & 0 \end{pmatrix} \hat{r}(t|t-1) + \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} y(t),$$

$$\hat{y}(t+1|t) = (1 \quad 0 \quad \ldots \quad 0)\hat{r}(t+1\ t),$$

and by a calculation to

$$\hat{y}(t+1|t) = -\sum_{i=1}^{n} \alpha_i \hat{y}(t+1-i|t-i) + \sum_{i=1}^{n} \beta_i y(t+1-i) = h(t+1|t)^T p.$$

The recursion for h then follows from its definition.

b) $$\hat{x}(t+s+1|t) = A\hat{x}(t+s|t)$$

$$= (A-KC)\hat{x}(t+s|t) + K\hat{y}(t+s|t)$$

$$\hat{y}(t+s+1|t) = C\hat{x}(t+s|t),$$

and the result follows along the lines of the proof of a).
The expressions for the variances follows similarly.

c) $$y(t) - h(t|t-1)^T p = y(t) - \hat{y}(t|t-1) = \bar{v}(t).$$ □

The parameters of the past-output induced Gaussian stochastic realization (3.4) are n, A, K, C, q, or in the equivalent form (11,12) n, q, p.

3.7. PROBLEM. Given the Gaussian process specified in 3.1 having the past-output induced stochastic realization (3,4). Assume that the values of the parameters n, q are known but that the value of p is unknown. The *adaptive multi-step stochastic prediction problem* for the above defined process y is to determine for all $t \in T$ and $s \in Z_+$ an estimate of the conditional distribution of y(t+s) given F_t^y. □

There are several synthesis procedures to solve the above problem. The *selftuning synthesis procedure* will be used here. This procedure prescribes that at each time step an estimate is to be made of the values of the parameters of the past-output induced stochastic realization, which estimate is then substituted for the parameter in the formula's for the predictor, formula's that are obtained with knowledge of the parameters. This approach has been introduced by WITTENMARK [24]. The procedure involves a kind of stepwise certainly equivalent because the output prediction is used at subsequent time moments in the parameter estimation algorithm, which fact is not specified a priori.

3.8. THEOREM. *Consider the representation*

(24) $$p(t+1) = p(t) + r(t),\ p(0),$$

(25) $$y(t) = h(t|t-1)^T p(t) + \bar{v}(t),$$

where (25) *is the representation of the past-output induced stochastic realization of* y *as derived in* 3.6.c, *and* (24) *a representation for the time-varying parameters. Assume that* $p(0) \in G(0, v_0 I_{2n})$, $r\colon \Omega \times T \to R^{2n}$ *is a Gaussian white noise process with intensity* $v_1 I_{2n}$, $v_1 \in R_+$, *and* $p(0)$, r, $\bar{v}$ *are independent objects. The solution to the adaptive stochastic prediction problem via the selftuning synthesis procedure is then that*

$$E[\exp(iuy(t+s)) \mid F_t^y]$$

may be approximated by

(26) $$\exp(iu\hat{y}(t+s|t) - \frac{1}{2}u^2 q(t+s|t))$$

where $\hat{y}$ *and* q *are here determined by respectively the parameter estimation algorithm*

(27) $$k^p(t) = Q^p(t|t-1)h(t|t-1)[h(t|t-1)^T Q^p(t|t-1)*h(t|t-1) + \bar{q}]^{-1},$$

(28) $$\hat{p}(t+1|t) = \hat{p}(t|t-1) + k^p(t)[y(t) - h(t|t-1)^T \hat{p}(t|t-1)],\ \hat{p}(0|-1) = 0,$$

(29) $$Q^p(t+1|t) = Q^p(t|t-1) + v_1 I_{2n} - Q^p(t|t-1)h(t|t-1)*$$

$$*[h(t|t-1)^T Q^p(t|t-1)h(t|t-1) + \bar{q}]^{-1}*$$

$$*h(t|t-1)^T Q^p(t|t-1),\ Q^p(0|-1) = v_0 I_{2n},$$

the adaptive Kalman filter

(30) $h(t+1|t) = L_n h(t|t-1) + M_n \hat{y}(t|t-1) + N_n y(t), \; h(0|-1) = 0,$

(31) $\hat{y}(t+1|t) = h(t+1|t)^T \hat{p}(t+1|t),$

(32) $q(t+1|t) = \bar{q};$

and the adaptive multistep predictor:

(33) $hh(t+s+1|t) = L_n hh(t+s|t) + (M_n+N_n)\hat{y}(t+s|t), \; hh(t+s|t) = h(t+1|t),$

(34) $\hat{y}(t+s+1|t) = hh(t+s+1|t)^T \hat{p}(t+1|t),$

(35)
$$Q^r(t+s+1|t) = F(\hat{p}(t+1|t))Q^r(t+s|t)F(\hat{p}(t+1|t))^T + \bar{q}G(\hat{p}(t+1|t))G(\hat{p}(t+1|t))^T, \quad Q^r(t+s|t) = 0,$$

(36) $q(t+s+1|t) = H(\hat{p}(t+1|t))Q^r(t+s+1|t)H(\hat{p}(t+1|t))^T + \bar{q}.$

PROOF. The proof is by induction. Suppose one is at $t \in T$ having obtained $\hat{p}(t|t-1)$, $Q^p(t|t-1)$ and $h(t|t-1)$ as functions of $(y(0),\ldots,y(t-1))$. The selftuning synthesis procedure now prescribes to estimate anew the value of the parameter p. The assumptions imply that the representation (24,25) is like a Gaussian system, the difference with such a representation being that $h(t|t-1)$ is dependent on $(y(0),\ldots,y(t-1))$. Using the conditional Kalman filter, see [15, II thm. 13.4], for which the formula's are similar to the standard Kalman filter, one obtains the parameter estimation algorithm (27, 28,29). The selftuning synthesis procedure prescribes next to substitute the estimate $\hat{p}(t+1|t)$ for p in the formula's for the prediction algorithm as given in 3.6. The result then follows. □

REMARKS.

1. What is new in the adaptive stochastic prediction algorithm is the recursive computation of the estimates and the variances. In most of the references quoted in section 1 the prediction of $y(t+s)$ given F_t^y for a

fixed $s \in Z_+$ is considered. HOLST [13], considers also multistep predictions but his implementation consists of a predictor for every increment of $s \in Z_+$. This in general leads to a rather high demand on computer memory and in addition has a very serious theoretical disadvantage, see remark 4. BOHLIN [1] does not consider multistep predictions explicitly.

2. It is possible to use other parameter estimation algorithms in the adaptive stochastic prediction algorithm. The formula's easily follow from 3.6 using the selftuning synthesis procedure.
3. It has been suggested [16] to include in the parameter estimation algorithm a stability test on the eigenvalues of (A-KC), or on the zeroes of the polynomial $\Sigma_{i=0}^{n} \alpha_i z^{n-i}$, with $\alpha_0 = 1$. Stability is a sufficient condition for convergence of the parameter estimates. Under the assumption 3.1 and a minimal stochastic realization, stability is ensured. In the algorithm for power demand prediction no stability test has been incorporated. No instabilities have been noticed.
4. HOLST [13] suggests another multistep adaptive stochastic prediction algorithm based on the following derivation. For each time increment of $s \in Z_+$ another predictor is constructed. Then, along the lines of the proof of 3.6,

$$\hat{y}(t+s|t) = h(t+s|t)^T p,$$

$$y(t+s|t) = \sum_{i=1}^{n} \alpha_i \hat{y}(t+s-i|t-i) + \sum_{i=1}^{n} \beta_i y(t+1-i) + \bar{\bar{v}}(t),$$

$$= h(t+s|t)^T p + \bar{\bar{v}}(t),$$

$$\bar{\bar{v}}(t) = \bar{v}(t) + \sum_{r=t-s+1}^{t-1} CA^{t-r-1} K\bar{v}(r).$$

   If $s > 1$ then $\bar{\bar{v}}$ is not Gaussian white noise. Yet HOLST [13] suggests to ignore this fact and to apply the least-squares system identification algorithm to estimate p. It is well known that estimates obtained this way may not converge. Therefore this approach to the adaptive stochastic prediction problem is doubtful.
5. In 3.8 the contribution of the variance of the parameter prediction error on the variance of the output prediction error has been neglected. The

selftuning synthesis procedure does not prescribe to account for this.

6. The algorithm 3.8 has the parameters v_0, v_1, q. The value of v_0 is not very critical since its influence decreases when time proceeds. The values of v_1, q are initially unknown. Here the stochastic approximation-like algorithm suggested by BOHLIN [1, p.454] to estimate a multiplicative factor in v_1, q is useful.
7. The convergence analysis of the adaptive stochastic predictor 3.8 has not yet been completed. A first question is what should converge? Two views have been taken on this question. L. LJUNG [16], considers the convergence of the estimates of the values of the parameters. Another view is to consider the difference between the predictions with known parameters and the adaptive predictions, see [11]. The references quoted contain sufficient conditions for convergence.
8. What is an optimal adaptive stochastic predictor? Since by definition an adaptive stochastic predictor is an approximation, the optimality is not clear. Consider the error between the prediction with known parameters and the adaptive predictor. It seems that an optimal adaptive predictor should at least satisfy two conditions: 1) the above defined error process must converge in some sense; 2) the adaptive predictor should have minimal asymptotic variance for the above error process.

4. THE LOAD MODEL AND PREDICTION ALGORITHM

As explained in section 1 energy is supplied to the railway power network at energy supply stations. In this section a model for the power demand at an energy supply station is proposed and a prediction algorithm derived.

Several time intervals that are relevant for prediction have been mentioned in section 2. In the following it is assumed that time is discretized in intervals of length Δt, thus each time moment a prediction must be made. It will be supposed that if data are available over intervals of length Δs with $\Delta s < \Delta t$ and $\Delta t = k_t \Delta s$, $k_t \in Z_+$, then these are aggregated to totals of length Δt.

The power demand data are clearly nonstationary. Several periods in the data may be distinguished such as an hour due to the railway schedule, a day with a difference between weekdays and weekenddays, a week and a year.

There may also be a trend throughout the years. It has been decided to separate out the week and day periods for special treatment. In addition attention is given to short term fluctuations. The variations between the seasons will be taken care of through the adaptive nature of the model. The full model has thus three time scales, a week, a day, and a short term period.

The week model

Assume that the power demand data y may be modelled as

(37) $$w_{s+1} = w_s + r_s,\ w_0,$$

(38) $$y(t) = w_s(t \bmod \tau_w) + u(t),$$

where (Ω,F,P) is a probability space, $s \in T = \{0,1,2,\ldots\}$, τ_w is the number of Δt periods in one week, $w_0\colon \Omega \to R^{\tau_w}$, $w_0 \in G(\mu^{w0},Q^{w0})$, $r\colon \Omega \times T \to R^{\tau_w}$ is a Gaussian white noise process with intensity $v_1 I_{\tau_w}$, $v_1 \in R_+$, $t = s\tau_w + t_1$, $1 \le t_1 \le \tau_w$ so $t_1 = t \bmod \tau_w$, $u\colon \Omega \times T \to R$ is a Gaussian white noise process with intensity $v_2 \in (0,\infty)$, $w\colon \Omega \times T \to R^{\tau_w}$ defined by formula (37), and w_0, v, u are independent objects. For $s \in T$, w_s will be called the week pattern of the load demand in week s, and y(t) the power demand at time $t = s\tau_w + t_1$.

The week model (37,38) is such that for each period of the week there is a first order model driven by a Gaussian white noise process. At each period of the week the value of the week pattern of that period is read out with a certain noise component u. There is no interaction between the periods of the week; this point is taken care of by the day and the short term model.

The asymptotic Kalman filter for the representations (37,38) is, for $t \in \{(s-1)\tau_w+1,\ldots,(s-1)\tau_w+\tau_w\}$

(39) $$k_s^w = q^w_{s|s-1}(q^w_{s|s-1} + v_2)^{-1},$$

(40) $$\hat{w}_{s+1|s}(t \bmod \tau_w) = \hat{w}_{s|s-1}(t \bmod \tau_w) + k_s^w[y(t) - \hat{w}_{s|s-1}(t \bmod \tau_w)],\quad \hat{w}_{0|-1} = \mu^{w0},$$

(41) $$q^w_{s+1|s} = q^w_{s|s-1} + v_1 - (q^w_{s|s-1})^2[q^w_{s|s-1} + v_2]^{-1}, \quad q_{0|-1} = v_0,$$

(42) $$q^y_{s+1|s} = q^w_{s+1|s} + v_2.$$

Note that the gain k^w and the variance q^w are real valued and need to be computed only once a week. By choosing v_1 and v_2 appropriately the above algorithm will approximately follow the seasonal fluctuations. In applications the algorithm is initialized with the first week of data.

The day model

Next a model is proposed for the difference between the power demand data and the estimate of the week pattern. This model has a multiplicative form with a period of a day and is supposed to model the effect of meteorological factors on the data. Multiplicative models have been considered earlier [12].

Assume that for all time periods in day $r \in T$

$$y(t) = g(r)\hat{w}_{s|s-1}(t \bmod \tau_w).$$

Let τ_d be the number of Δt periods in one day. A least-squares approach to estimate $g(r)$ yields:

(43) $$g(r) = [\sum_{t=1}^{\tau_d} y(t)\hat{w}_{s|s-1}(t \bmod \tau_w)]/[\sum_{t=1}^{\tau_d} \hat{w}_{s|s-1}(t \bmod \tau_w)^2]$$

where the sum is over all time periods in day r. Assume further that $\{g(r)-1, r \in T\}$ can be modelled as a Gaussian process, specifically as the output of a Gaussian system of order n_d. One may then associate with this process the representations (3,4). Let v_3, v_4 be respectively the variances (v_1,q) in (24,25). The one-step adaptive stochastic predictor of $\{g(r)-1, r \in T\}$ is then

(44) $$k^d(r) = Q^d(r|r-1)h^d(r|r-1)[h^d(r|r-1)^T Q^d(r|r-1)h^d(r|r-1) + v_4]^{-1},$$

(45) $$\hat{p}^d(r+1|r) = \hat{p}^d(r|r-1) + k^d(r)[g(r) - 1 - h^d(r|r-1)^T * \hat{p}^d(r|r-1)],$$
$$\hat{p}^d(0|-1) = 0,$$

(46) $$Q^d(r+1|r) = Q^d(r|r-1) + v_3 I_{2n_d} - Q^d(r|r-1)h^d(r|r-1)*$$
$$*[h^d(r|r-1)^T Q^d(r|r-1)h^d(r|r-1) + v_4]^{-1}*$$
$$*h^d(r|r-1)^T Q^d(r|r-1), \quad Q^d(0|-1) = I_{2n_d},$$

(47) $$h^d(r+1|r) = L_{n_d} h^d(r|r-1) + M_{n_d}(\hat{g}(r)-1) + N_{n_d}(g(r)-1), \quad h(0|-1) = 0,$$

(48) $$\hat{g}(r+1|r) = h^d(r+1|r)^T \hat{p}(r+1|r) + 1.$$

Here L_{n_d}, M_{n_d}, N_{n_d} are as defined in 3.6.

The short term model

Short term fluctuations, of the order of several time periods, are modelled next. Let the prediction error process after the application of the week and day prediction algorithm be, for the s-th week and the r-th day, $e: \Omega \times T \to R$

(49) $$e(t) = y(t) - \hat{g}(r|r-1)\hat{w}_{s|s-1}(t \bmod \tau_w).$$

Assume that e is a Gaussian process satisfying the assumptions of 3.1, and that the order of the associated Gaussian system (3,4) is $n_e \in Z_+$. Denote the variances (v_1, q) in the representation (24,25) in this case by v_5, q. Then 3.8 yields the following multistep adaptive prediction algorithm

(50) $$k^e(t) = Q^e(t|t-1)h^e(t|t-1)[h^e(t|t-1)^T Q^e(t|t-1)*h^e(t|t-1) + q]^{-1},$$

(51) $$\hat{p}^e(t+1|t) = \hat{p}^e(t|t-1) + k^e(t)[e(t) - h^e(t|t-1)^T \hat{p}^e(t|t-1)],$$
$$\hat{p}^e(0|-1) = 0,$$

(52) $$Q^e(t+1|t) = Q^e(t|t-1) + v_5 I_{2n_e} - Q^e(t|t+1)h^e(t|t-1)*$$
$$*[h^e(t|t-1)^T Q^e(t|t-1)h^e(t|t-1) + q]^{-1}*$$
$$*h^e(t|t-1)^T Q^e(t|t-1), \quad Q^e(0|-1) = I_{2n_e},$$

(53) $$h^e(t+1|t) = L_{n_e} h(t|t-1) + M_{n_e} \hat{e}(t|t-1) + N_{n_e} e(t),\ h^e(0|-1) = 0,$$

(54) $$\hat{e}(t+1|t) = h^e(t+1|t)^T \hat{p}^e(t+1|t),$$

(55) $$q(t+1|t) = q,$$

(56) $$hh^e(t+s+1|t) = L_{n_e} hh^e(t+s|t) + (M_{n_e} + N_{n_e})\hat{e}(t+s|t),$$

$$hh^e(t+1|t) = h^e(t+1|t),$$

(57) $$\hat{e}(t+s+1|t) = hh^e(t+s+1|t)^T \hat{p}^e(t+1|t).$$

The formula's for the variance of the prediction error of $\hat{e}(t+s+1|t)$ follow as in 3.8.

The final prediction of the power demand in week s, in day r, at time t, is then

(58) $$\hat{y}(t+\tau|t) = \hat{g}(r|r-1)\hat{w}_{s|s-1}((t+\tau) \bmod \tau_w) + \hat{e}(t+\tau|t)$$

for $\tau \in Z_+$ and with the associated variance as the variance of the prediction error of $\hat{e}(t+\tau|t)$. In this algorithm the effect that occurs at a day change has been neglected. The complete adaptive stochastic prediction algorithm for power demand is formed by (39-58).

The above model and algorithm differ from those suggested by HOLST [13] and BOHLIN [1]. The difference with [13] is in the estimation of the week pattern via the Kalman filter rather than via exponential smoothing, in the multiplicative model with the time scale of a day, and in the recursive computation of the multistep predictions. The difference with [1] is in the adaptive stochastic prediction algorithm for Gaussian systems, in the full week model, the day model, and in the somewhat more general short term model. Yet the model and algorithm are very much inspired by BOHLIN [1] and [2]. As mentioned in section 3 one may add an algorithm to recursively estimate a multiplicative factor in the variance of the innovation process.

5. NUMERICAL RESULTS

Data from several energy supply stations in a region of the Netherlands have been made available for the investigation by the railway company. The data are of the total power demand in 15 minute intervals. Initially data have been used over 4 months of 1979, about 11000 data points. In the latter part of the investigation data of the full year 1979 have been used, about 35000 data points.

A computer program has been written that generates predictions according to the algorithm presented in section 4. The software has been implemented on the computer of the railway company. The computer program produces predictions in the form: for $t \in T$, and $s = 1,2,\ldots,t_1$, $\{t+s, \hat{y}(t+s|t), q(t+s|t)^{\frac{1}{2}}\}$, and for evaluation purposes $y(t)$, $\hat{y}(t|t-1)$, $y(t) - \hat{y}(t|t-1)$, $q(t|t-1)^{\frac{1}{2}}$.

An evaluation of the performance of the prediction algorithm has been made. The following properties of the prediction errors have been determined: 1) the sample mean; 2) the sample variance; 3) the relative sample variance

$$s_{\%} = [\sum_{t=1}^{t_2} ((y(t) - \hat{y}(t|t-1))/y(t))^2/t_2]^{\frac{1}{2}}*100;$$

4) the correleogram; 5) an estimate of the distribution. In addition an hypothesis test on the correleogram and the sample distribution of the prediction errors has been made.

For a certain energy supply station some numerical results are mentioned in table 1. The statistics are of the prediction errors in: 1) the 4 month period minus the first three weeks; 2) the same but only for the hours of 6 to 22 of a day; 3)monday mornings from 6 to 10 hours; 4) friday afternoons from 15 to 20 hours. In figure 2 is displayed the power demand on an arbitrary day, and in figure 3 a sample prediction. Monday morning and friday afternoon have been selected because in these time periods the highest power demand occurs.

Hours	Sample means	Sq.rt. sample variance	Relative variance %	Range of data
0 -24h.	34.9 kw	334.7 kw	41.44%	(0 ,5500) kw
6 -22h.	4.6 kw	353.8 kw	16.18%	(1500,5500) kw
6 -10h. mondays	-18.4 kw	355.4 kw	10.23%	(2000,5000) kw
15-20h. fridays	17.3 kw	419.0 kw	12.47%	(2000,5500) kw

Table 1. Evaluation of the one-step prediction for an energy supply station.

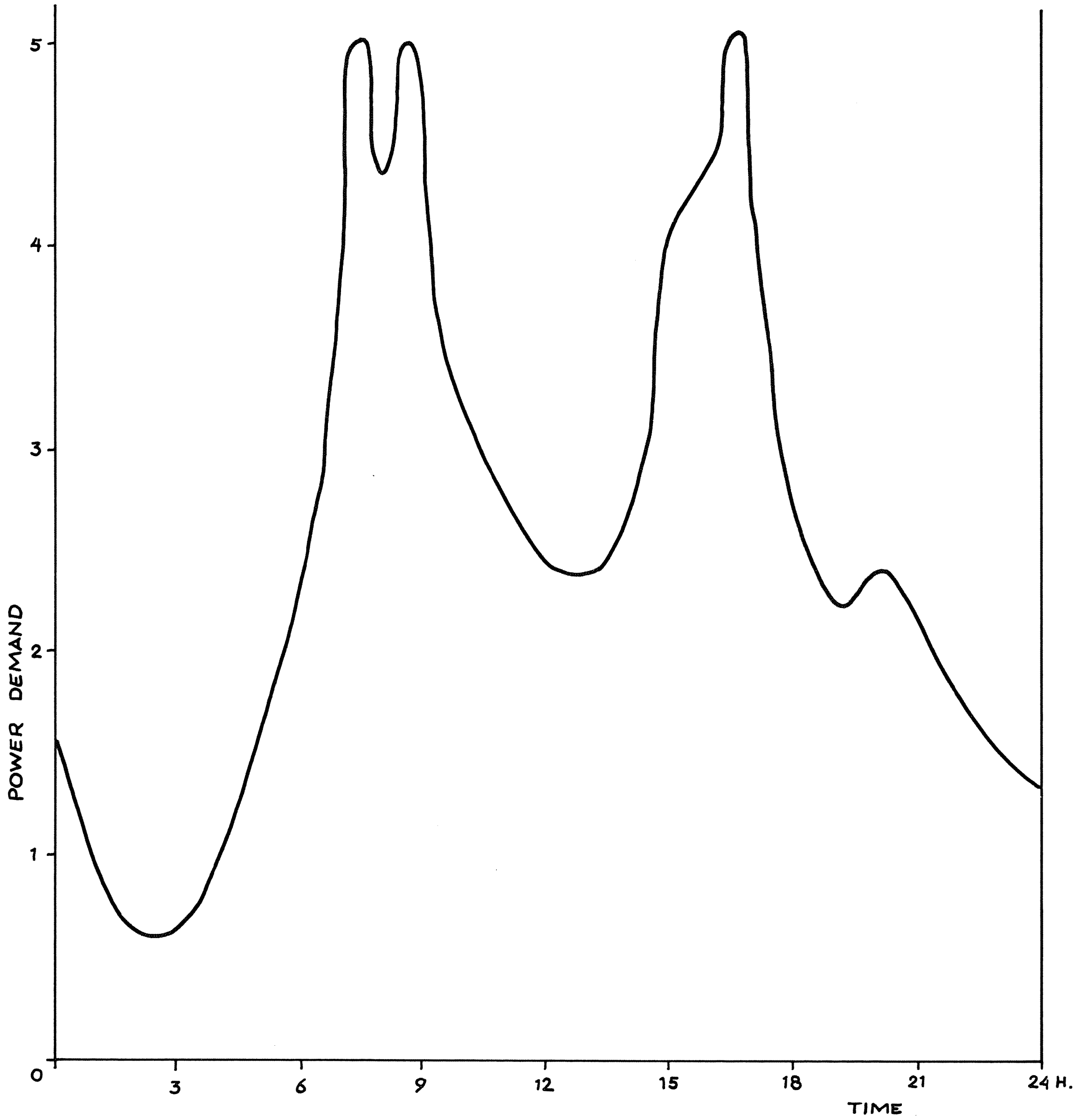

Figure 2. The power demand at an energy supply station on an arbitrary day.

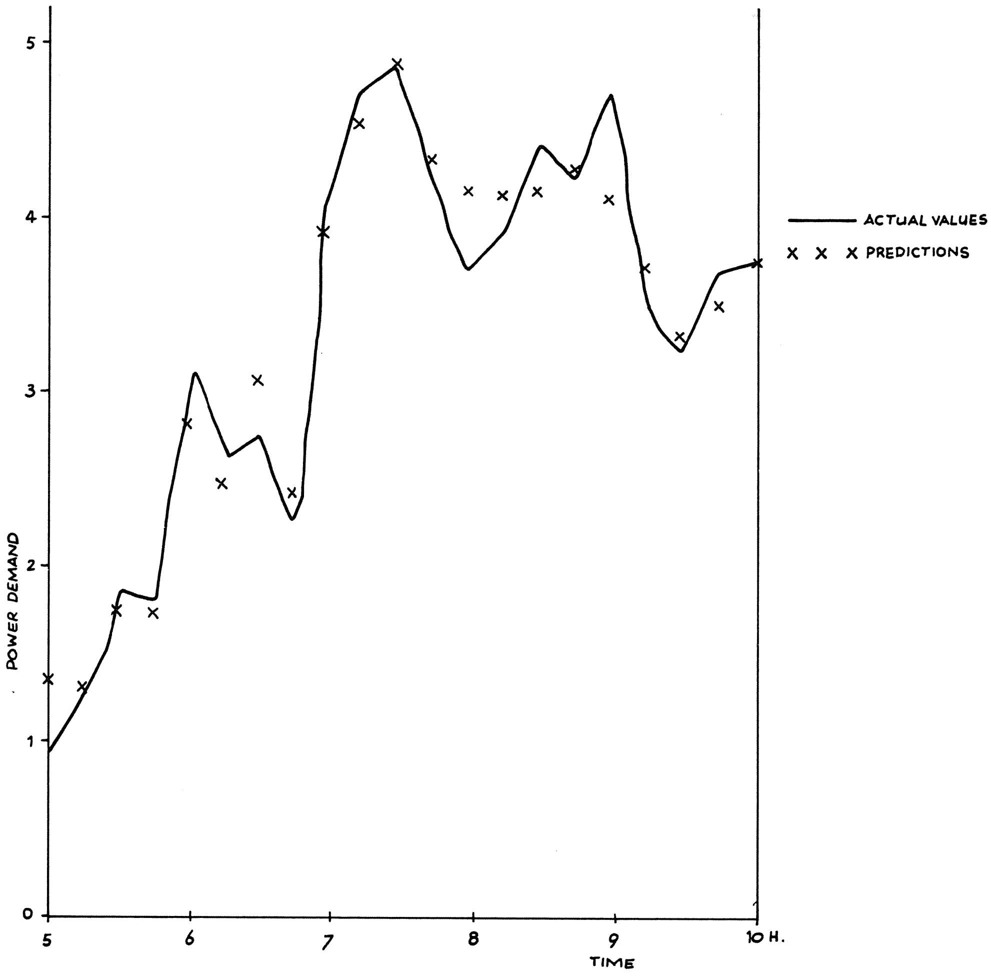

Figure 3. One step adaptive prediction of power demand at an energy supply station.

The relative sample variance is not very useful as an evaluation measure of the algorithm when considering data over 24 hours a day. At night, when due to irregular freight traffic the power demand is rather erratic and of rather low value, the relative prediction errors are high and consequently contribute disproportionately to the relative sample variance. During day hours the relative sample variance is useful.

Multistep predictions have been generated based on the selftuning adaptive prediction algorithm mentioned in remark 4 of section 3. For the above mentioned energy supply station the square root of the sample variance in statistic 1) increases over 1 to 12 periods ahead from about 360 to 420 kw.

The parameters of the algorithm have been varied. Based on the variance of the prediction error the choices $n_d = 2$ and $n_e = 2$ have been made. No specific order test has been used. The performance of the algorithm as measured by the sample variance of the prediction error proved to be rather sensitive for the ratio v_1/v_2. Finally the choice $v_1 = 15$ and $v_2 = 500$ has been made. With this choice the algorithm seems to follow the seasonal fluctuation rather well. The variances v_3, v_4 in the day model have been chosen to be 0 resp. 1. Finally in the short term model v_5 has been chosen to be zero. In both the day and the short term model a stochastic approximation-like algorithm for the variance has been tried, see [13, p. 127]. This corresponds to taking $v_3, v_5 > 0$ in the algorithm presented here. These measures resulted in a slight improvement of the variance of the prediction error.

A summary of the evaluation follows. The variance of the prediction error is relatively high. During peak hours the square root of the relative sample variance is of the order of 12%. The correlogram showed that the prediction errors are approximately uncorrelated. The empirical distribution of the prediction errors has a normal shape, but an hypothesis test at 95% indicated a deviation from normality. Special holidays are an additional problem.

Some causes for the high relative sample variance of the prediction errors have been suggested by persons at the railway company. These causes are: 1) the driving style of the machinist; 2) meteorological factors, primarily wind; 3) deviations from the railway schedule; 4) and composition of the trains. Apparently the high relative sample variance is just a characteristic of the data.

Predictions have also been generated for several other energy supply stations, for a station with data aggregated to 1 hour totals, and for 15 minute data of the sum of several stations. The results of these predictions for the aggregated data are considerably better than those mentioned above for one station.

To conclude this evaluation, the predictions produced by the algorithm are of reasonable quality considering the characteristics of the data. The selftuning prediction algorithm needed only little tuning and performed quite well.

6. CONCLUSIONS

An adaptive stochastic prediction algorithm has been presented for short term prediction of power demand at an energy supply station of the dutch railway company. The algorithm is based on a new multistep adaptive prediction algorithm for Gaussian systems.

REFERENCES

[1] BOHLIN, T. (1976), *Four cases of identification of changing systems*, In: *System Identification-Advances and Case Studies*, R.K. Mehra and D.G. Lainiotis (eds), Academic Press, New York, pp. 441-518.

[2] BOHLIN, T. & P. KAMJOU (1977), *A batch program for forecasting electric consumption*, Report TRITA-REG-7701, Dept. Automatic Control, Royal Institute of Technology, Stockholm, Sweden.

[3] BUNN, D.W. (1980), *Experimental study of a Bayesian method for daily electricity load forecasting*, Appl. Math. Modelling 4, pp. 113-116.

[4] CHRISTIAANSE, W.R. (1971), *Short-term load forecasting using general exponential smoothing*, IEEE Trans. Power Apparatus & Systems, 90, pp. 900-910.

[5] KEYSER, R.M.C. DE & A.R. VAN CAUWENBERGHE (1981), *A selftuning multistep predictor application*, Automatica 17, pp. 167-175.

[6] FARMER, E.D. (1964), *A method of prediction for non-stationary processes and its application to the problem of load estimation*, In: *Automatic and Remote Control*, V. Broida (ed.), Proc. 2nd IFAC Congress, Butterworths, London, pp. 47-54.

[7] FARMER, E.D. & M.J. POTTON (1966), *The prediction of load on a power system*, Proc. 3rd IFAC Congress, London, paper 21F.

[8] FAURRE, P. (1973), *Réalisations markoviennes de processus stationnaires*, Rapport de Recherche no. 13, INRIA, Le Chesnay, France.

[9] FAURRE, P., M. CLERGET & F. GERMAIN (1979), *Opérateurs rationnels positiefs*, Dunod, Paris.

[10] GOODWIN, G.C., KWAI SANG SIN & K.K. SALUJA (1980), *Stochastic adaptive control and prediction - The general delay-colored noise case*, IEEE Trans. Automatic Control 25, pp. 946-950.

[11] GOODWIN, G.C., P.J. RAMADGE & P.E. CAINES (1981), *A global convergent adaptive predictor*, Automatica 17, pp. 135-140.

[12] HARRISON, P.J. & C.F. STEVENS (1971), *A Bayesian approach to short-term forecasting*, Oper. Res. Quarterly 22, pp. 341-362.

[13] HOLST, J. (1977), *Adaptive prediction and recursive estimation*, Report LUTFD2/(TFRT-1013)/1-206/(1977), Dept. of Automatic Control, Lund Institute of Technology, Lund, Sweden.

[14] KALMAN, R.E. (1960), *A new approach to linear filtering and prediction problems*, Trans. ASME Ser. D, J. Basic. Eng. 82, pp. 35-45.

[15] LIPSTER, R.S. & A.N. SHIRYAYEV (1977), *Statistics of random processes I, General theory II*, Applications, Springer-Verlag, Berlin.

[16] LJUNG, L. (1978), *Convergence of an adaptive filter algorithm*, Int. J. Control 27, pp. 673-693.

[17] MAHALANABIS, A.K. & M. HANMANDLU (1980), *Further studies on the problem of load modelling of an interconnected power system for short-term prediction*, Int. J. Systems Sci. 11, pp. 937-946.

[18] SHARMA, K.L.S. & A.K. MAHALANABIS (1974), *Recursive short-term load-forecasting algorithm*, Proc. IEE, 121, pp. 59-62.

[19] SIN, K.S., G.C. GOODWIN & R.R. BITMEAD (1980a), *An adaptive* d-*step ahead predictor based on least squares*, IEEE Trans. Automatic Control 25, pp. 1161-1165.

[20] SIN, K.S., G.C. GOODWIN & R.R. BITMEAD (1980b), *An adaptive* d-*step ahead predictor based on least squares*, Proc. 19th IEEE Conference on Decision and Control, pp. 962-967.

[21] SINGH, A., K.K. BISWAS & A.K. MAHALANABIS (1977), *Load modelling for real time monitoring of power systems*, IEEE Trans. Power Apparatus & Systems 96, pp. 1908-1914.

[22] TOYODA, J., MO-SHING CHEN & Y. INOUE (1970), *An application of state estimation to short-term load forecasting*, Part I: *Forecasting modelling*, IEEE Trans. Power Apparatus & Systems 89, pp. 1678-1682.

[23] VAN DER BIJ, J.D. (1981), *Voorspelling energieafname voedingspunten (Prediction of energy demand at energy supply stations)*, Report St&O/SOG/620.002(310), Nederlandse Spoorwegen, Utrecht, The Netherlands. (In dutch).

[24] WITTENMARK, B. (1974), *A selftuning predictor*, IEEE Trans. Automatic Control 19, pp. 848-851.