

IMM  
DEPARTMENT OF MATHEMATICAL MODELLING

Technical University of Denmark  
DK-2800 Lyngby – Denmark

J. No. x599-8  
1998-11-23  
HAN

# **EMPIRISK BESTEMT MODEL FOR ELFORBRUGET I ØSTDANMARK**

**Henrik Aalborg Nielsen  
Klaus Kaae Andersen  
Henrik Madsen**

**TECHNICAL REPORT**

**IMM-REP-1998-18**

**IMM**

# **EMPIRISK BESTEMT MODEL FOR ELFORBRUGET I ØSTDANMARK**

**Henrik Aalborg Nielsen  
Klaus Kaae Andersen  
Henrik Madsen**

**TECHNICAL REPORT**

**IMM-REP-1998-18**

**IMM**

# Empirisk bestemt model for elforbruget i Østdanmark

Henrik Aalborg Nielsen, Klaus Kaae Andersen & Henrik Madsen

23. november 1998



Institut for Matematisk Modellering  
Danmarks Tekniske Universitet



<i>Model for elforbruget i Østdanmark</i>	1
---	---

## **Indhold**

<b>1</b>	<b>Introduktion</b>	<b>2</b>
<b>2</b>	<b>Data</b>	<b>2</b>
<b>3</b>	<b>Model</b>	<b>3</b>
<b>4</b>	<b>Evaluering af model</b>	<b>4</b>
<b>5</b>	<b>Konklusion</b>	<b>8</b>
<b>A</b>	<b>Data</b>	<b>9</b>
<b>B</b>	<b>Matematisk formulering af model</b>	<b>10</b>
<b>C</b>	<b>Generering af den anvendte B-spline basis</b>	<b>11</b>

## 1 Introduktion

Nielsen & Madsen (1997) beskriver empirisk baserede modeller for elforbruget øst for Storebælt. Det har været ønsket, at udvikle en samling SAS-makroer (SAS Institute Inc. 1990) til simpel og forholdsvis hurtig håndtering af en statisk version af en sådan model (Nielsen & Madsen 1998). Af denne grund er årsvariationen modelleret på en anden måde end i (Nielsen & Madsen 1997). Denne rapport beskriver kort de opnåede resultater, når data for 1991-96 anvendes.

## 2 Data

Data er timeværdier af udetemperatur, vindhastighed, globalstråling og elforbrug. For klimadata gælder, at disse er gennemsnit for den forløbne time. Klimadata er målt på Højbakkegård i Taastrup af Institut for Jordbrugsvidenskab, Den Kgl. Veterinær- og Landbohøjskole (Jensen 1997). Enhederne er  $^{\circ}C$  for udetemperaturen,  $m/s$  for vindhastigheden og  $W/m^2$  for globalstrålingen.

Elforbruget, registreret ved et givet klokkeslet, er det totale elforbrug i den forløbne time i  $MWh$ . Af denne grund kan det også betragtes som en gennemsnitlig effekt og anføres da i enheden  $MWh/h$ . Elforbrugsdata er leveret af planlægningsafdelingen hos Elkraft, Ballerup. Direkte målinger af det samlede forbrug eksisterer ikke, i stedet er data fremkommet som en sum af

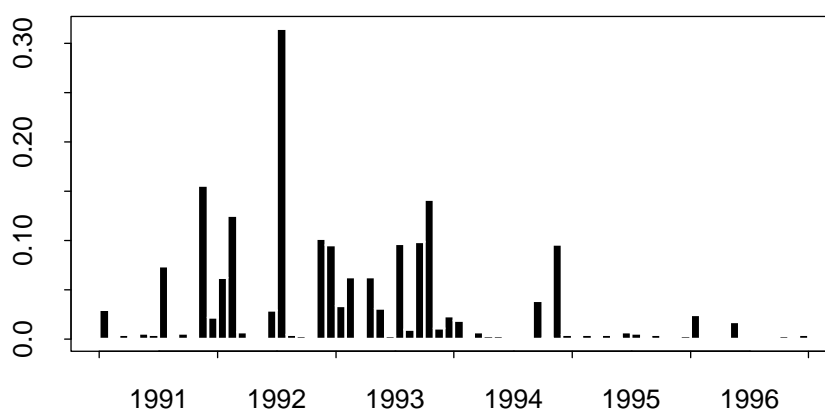
- produktion på centrale anlæg, korrigeret for udveksling med udlandet,
- decentral produktion på anlæg ejet af Sjællandske Kraftværker (SK), skaleret til SK-ejet, privat og distributionsejet decentral produktion, og
- produktion på elværksejede vindmøller, skaleret til elværksejet og privat vindkraftproduktion.

Hver af de fire variable er blevet undersøgt for manglende og atypiske værdier. Manglende værdier er blevet erstattet med en talkode, således at andelen af manglende observationer pr. måned kan beregnes. I det følgende redegøres kort for den anvendte fremgangsmåde ved rensning af data.

Indledningsvis er det sikret at alle indeks (dato, klokkeslet) mellem den 1/1-1991 kl. 01:00 og 31/12-1996 kl. 24:00 er tilstede i data. Herefter er data undersøgt for atypiske værdier. Med en atypisk værdi forstås en observation, hvor den registrerede værdi ikke skønnes at være en mulig værdi, eller hvor det dynamiske forløb indikerer, at observationen er fejlagtig. I det førstnævnte tilfælde er de fundne værdier korrigeret eller sat til manglende. Eksempelvis er negative værdier af globalstrålingen korrigeret til værdien nul, mens ekstreme udetemperaturer er blevet sat til manglende. I det andet tilfælde, er det dynamiske forløb af variablene undersøgt. Dette er gjort, da det ikke umiddelbart kan afsløres ud fra

observationens numeriske værdi, hvorvidt den er atypisk. Fremgangsmåden for at identificere sådanne atypiske observationer er, at estimerer en  $ARMA(2, 1)$  model (Madsen 1995) for hver tidsrække (variabel). En sådan model dækker både den hurtige og den langsomme dynamik i tidsrækken, og er dermed i stand til at afsløre visse atypiske observationer. Residualerne fra de estimerede  $ARMA(2, 1)$  modeller er blevet sorteret i dalende rækkefølge, da atypiske observationer typisk vil generere store residualer. Herefter er serierne undersøgt visuelt i omegne, af de hertil svarende, 200 numerisk største residualer. Værdier, der ved den visuelle undersøgelse vurderes at være fejlagtige, er blevet sat til manglende, og udelades således blot af analysen, der beskrives i de følgende afsnit.

På baggrund af rensning af data kan det konstateres, at de manglende værdier i datasættet dels skyldes fejlagtige observationer dels at variable ikke er blevet registreret. Figur 1 viser andelen af timer per måned hvor en eller flere af de fire variable mangler i det endelige datasæt. Det skal understreges, at figuren domineres af værdier, der oprindeligt var manglende i klimadata.



Figur 1. Andelen af timer per måned hvor en eller flere af elforbrug, temperatur, vindhastighed eller globalstråling mangler.

I appendiks A (side 9) er de anvendte data gengivet grafisk.

### 3 Model

I (Nielsen & Madsen 1997) er årsvariationen modelleret v.h.a. en Fourierudvikling, der er aktiv hele året, overlejret med en Fourierudvikling, der kun er aktiv i sommerperioden og en ændring i middelværdi, der kun er aktiv mellem jul og nytår. Specielt mht. sommerperioden er det ikke let at garantere anvendeligheden på ukendte data, og estimation af periodens længde og placering er overordentlig vanskelig når Fourierudviklinger anvendes på den beskrevne måde. Af disse grunde modelleres årssvingningen ved en kubisk B-spline basis med ortogonale søjler (de Boor 1978, Chambers & Hastie 1991). Den anvendte basis har en overvægt af knuder placeret i sommerperioden, samt i starten og slutningen af kalenderåret, se appendiks C, og har i alt 27 interne knuder, svarende til 30 frihedsgrader.

Det bemærkes, at den herved opnåede årsvariation ikke har samme værdi ved starten og slutningen af året. Den vil således også beskrive en tendens i middelværdien.

En præcis redegørelse for den valgte placering af knuder vil ikke blive givet i denne rapport. Det skal dog bemærkes, at knudeplaceringen er valgt således, at kun langsomme årssvingninger modelleres, på nær omkring jul/nytår og i sommerperioden. Dette er begrundet ved analysen af årvariationen i (Nielsen & Madsen 1997).

Af hensyn til beregningstiden arbejdes der i den implementerede model med tre døgn-grupper, hvorimod der i (Nielsen & Madsen 1997) anvendes fem grupper. De her anvendte grupper er

1. Hverdage
2. Lørdage og halve helligdage
3. Søndage og andre helligdage

Ud fra bl.a. Figur 4.5 i (Nielsen & Madsen 1997) vurderes denne modifikation at være af mindre betydning. I appendiks B er modellen beskrevet i flere detaljer. Hverdage mellem jul og nytår behandles som dage i gruppe 2.

Ovenstående model er i denne rapport sammenlignet med en tilsvarende model, hvor der anvendes 10 knuder per måned til modellering af årssvingningen.

## 4 Evaluering af model

Som nævnt i afsnit 3 sammenlignes den valgte model, hvor en B-spline basis med 27 interne knuder anvendes til modellering af årsvariationen, med en model med 10 knuder per måned, svarende til 118 interne knuder. Denne sammenligning er valgt for at illustrere, at den valgte model er relativ ufølsom overfor om et eller flere år anvendes til estimation, fordi den tvinger en større del af forklaringen over på klimaet.

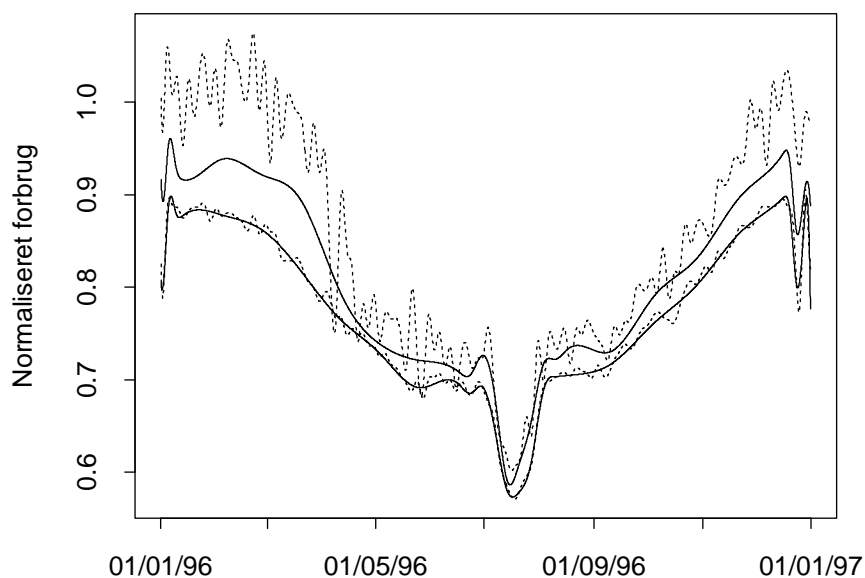
I analyserne er det normaliserede elforbrug per time anvendt. Dette er defineret som det aktuelle elforbrug i en given time delt med det observerede timegennemsnit for det pågældende år.

Figur 2 viser årsvariationen<sup>1</sup> som den estimeres ved anvendelse af de to modeller, når h.h.v. data for 1996 og for 1991-96 anvendes. Når data for 1991-96 anvendes, er de to estimater stærkt sammenfaldende, dog giver estimatet svarende til 118 interne knuder anledning til nogle højfrekvente, men små fluktuationer omkring estimatet svarende til 27 interne knuder. Når kun data for 1996 anvendes afviger de to estimater stærkt, og estimatet svarende til 27 interne knuder, er det der mest ligner de for 1991-96 opnåede estimater. Som det ses af Figur 3 skyldes dette at jo mindre amplituden i årsvariationen er jo større andel af elforbruget bliver forklaret af klimavariablene, især udetemperaturen.

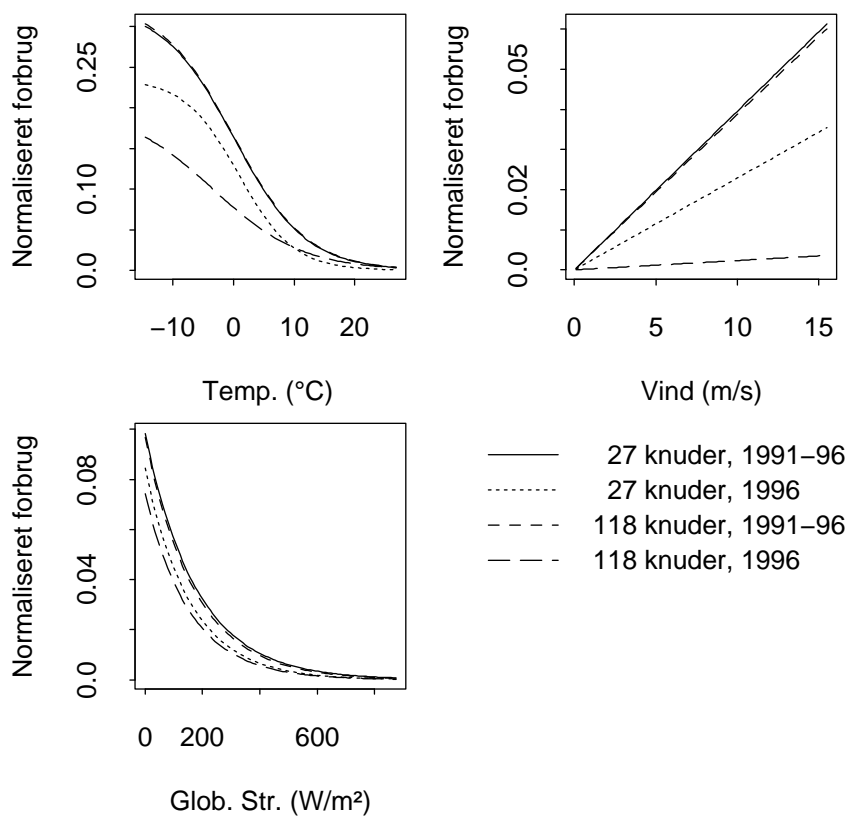
---

<sup>1</sup>1. aksen viser datoer i 1996, omend årstallet i realiteten er uden betydning.





Figur 2. Årsvariation ved 27 (fuld) og 118 interne knuder (stiplet). Data for 1996 (øverst) og 1991-96 (nederst).



Figur 3. Estimeret klimaafhængigheder, beregnet over intervaller svarende til de i 1996 observerede minimum og maksimum. Bemærk at skala på 2. akserne *ikke* er ens.

For at undersøge modellernes anvendelighed på fremtidige data, er modellerne blevet krydsvalideret på to måder. I begge tilfælde gennemløbes data for 1991-96 år for år og for hvert år der betragtes deles det samlede datasæt i to; (i) estimationsdata og (ii) valideringsdata, hvor (ii) altid svarer til alle observationer der ikke er medtaget i (i). For de to forskellige valideringsmetoder defineres estimationsdata, når et givet år betragtes, ved:

1. Alle observationer i det betragtede år.
2. Alle observationer *på nær* observationer i det betragtede år.

Disse to metoder er analoge til den grafiske sammenligning ovenfor. For hvert sæt estimer (seks per model) beregnes prædiktionsfejlene for valideringsdata, herefter kaldet  $v_i$ , hvor  $i$  indikerer observationen. Størrelsen af disse måles ved forklaringsgraden

$$R^2 = 1 - \frac{\sum v_i^2}{\sum (P_i - \bar{P})^2} \quad (1)$$

hvor  $P_i$  er det normaliserede elforbrug,  $\bar{P}$  er gennemsnittet heraf og begge summationer forløber over tidspunkter svarende til valideringsdata. Forklaringsgraden er således et udtryk for hvor stor en del af den samlede variation i data, der forklares af modellen. Desuden er  $RMS$  (Root Mean Square) anvendt. Idet  $N$  angiver antallet af prædiktionsfejl, er denne størrelse er defineret ved

$$RMS = \sqrt{\sum v_i^2 / N}. \quad (2)$$

Såfremt middelværdien af prædiktionsfejlene er nul, er  $RMS$  tilnærmelsesvis lig standardafvigelsen, hvor  $\pm 2$  standardafvigelser tilnærmelsesvis angiver et 95% prædiktionsinterval<sup>2</sup>.

Tabel 1 og 2 viser forklaringsgraderne for de to modeller når h.h.v. et og fem år anvendes ved estimationen. Forklaringsgraderne er både beregnet for estimations- og valideringsdata. Valideringsresultaterne er gengivet grafisk i Figur 4. Ikke overraskende giver anvendelsen af flere års data i forbindelse med estimationen det bedste resultat og her er 27 interne knuder kun marginalt bedre end 118. Hvis derimod kun et år anvendes til estimationen er 27 interne knuder langt bedre en 118. Når 27 interne knuder anvendes er følsomheden overfor hvorvidt 1 eller 5 år anvendes til estimation begrænset, hvorfor denne spline basis er at foretrække.

I Tabel 3 er  $RMS$  vist. Med 27 interne knuder og fem års estimationsdata synes 0.045 at være en typisk værdi. Med et gennemsnitligt timeforbrug på  $1600 MWh$  (svarende til 1996) giver dette et approksimativt 95% prædiktionsinterval på  $\pm 144 MWh$ . Når et år anvendes er det tilsvarende interval  $\pm 160 MWh$ .

---

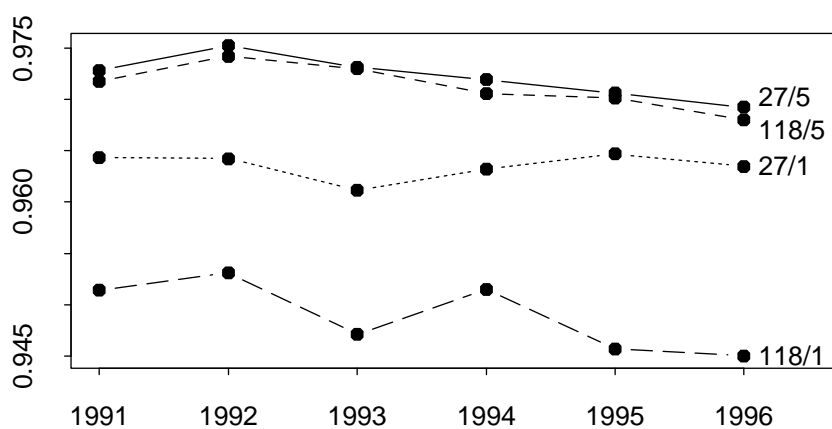
<sup>2</sup>Et interval der med 95% sandsynlighed indeholder den sande værdi.

	Estimation		Validering	
	27	118	27	118
1991	0.984	0.989	0.964	0.951
1992	0.986	0.990	0.964	0.953
1993	0.986	0.990	0.961	0.947
1994	0.983	0.987	0.963	0.951
1995	0.979	0.985	0.965	0.946
1996	0.982	0.987	0.963	0.945

Tabel 1. Forklaringsgrad når et år anvendes ved estimationen (metode 1 på side 6).

	Estimation		Validering	
	27	118	27	118
1991	0.976	0.977	0.973	0.972
1992	0.976	0.976	0.975	0.974
1993	0.976	0.977	0.973	0.973
1994	0.977	0.977	0.972	0.971
1995	0.977	0.978	0.971	0.970
1996	0.977	0.978	0.969	0.968

Tabel 2. Forklaringsgrad når fem år anvendes ved estimationen (metode 2 på side 6).



Figur 4. Valideringsresultater svarende til Tabel 1 og 2. Til højre for hver kurve er antallet af interne knuder og antallet af år anvendt til estimation anført som <knuder>/<år>.

	Metode 1 (Et år til est.)		Metode 2 (Fem år til est.)	
	27	118	27	118
1991	0.0490	0.0572	0.0419	0.0426
1992	0.0492	0.0563	0.0397	0.0405
1993	0.0508	0.0592	0.0433	0.0434
1994	0.0495	0.0568	0.0440	0.0450
1995	0.0485	0.0601	0.0448	0.0452
1996	0.0495	0.0607	0.0453	0.0462

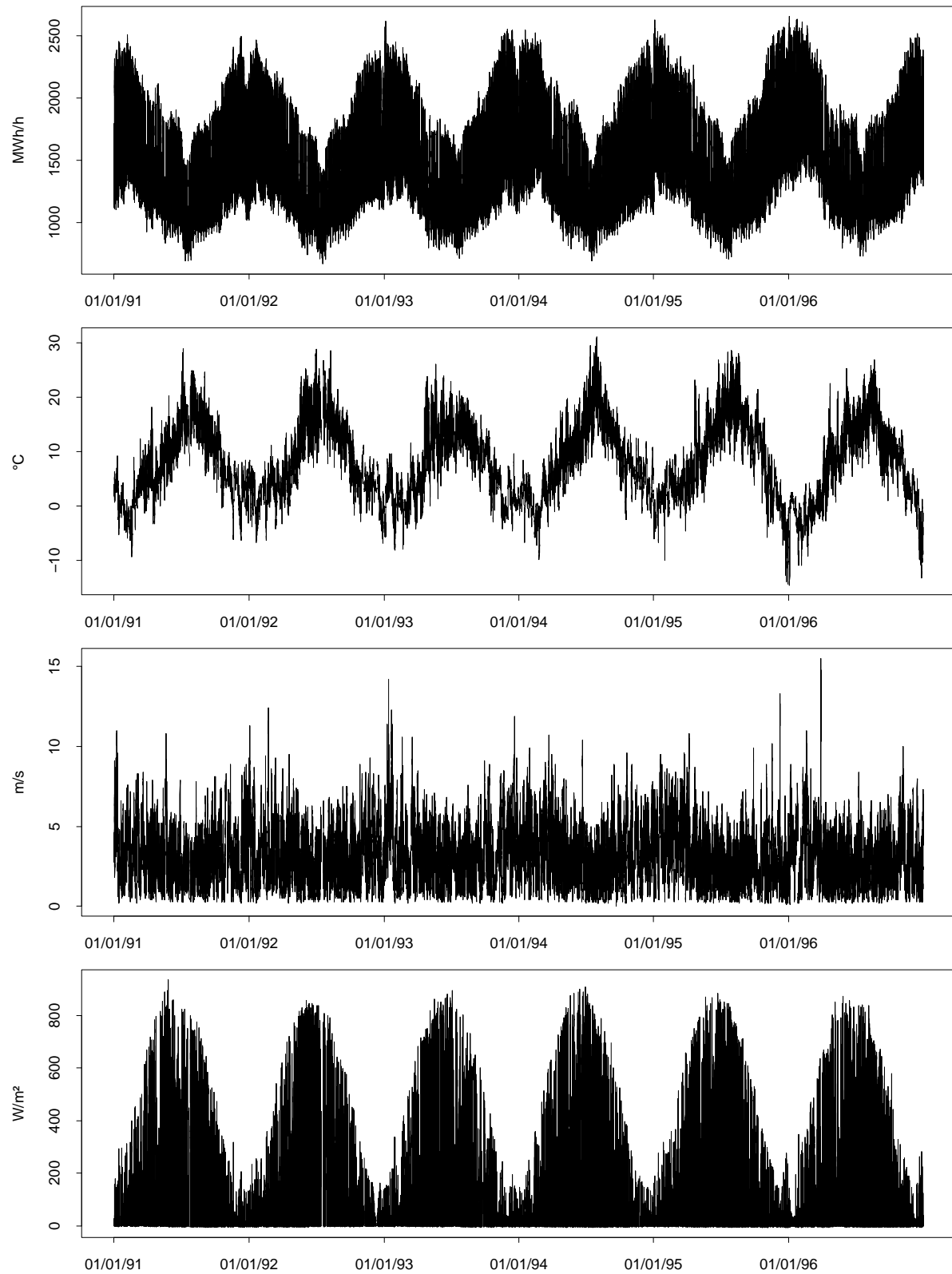
Tabel 3. *RMS* for valideringsfejlen for hver af de to metoder nævnt på side 6.

## 5 Konklusion

En statistisk model til prædiktions af timeforbruget af elektricitet i Østdanmark er præsenteret og sammenlignet med en alternativ model, der fitter data bedre. De to modeller afviger udelukkende ved, at der i den ene anvendes en B-spline med 27 interne knuder til modellering af årssvingningen (den valgte model), medens der i den anden anvendes 118 interne knuder (den alternative model). Det er demonstreret, at den valgte model er bedre til prædiktions end den alternative og at den er rimelig robust overfor om et eller flere år anvendes til estimation. Modellernes prædiktionssevne er evalueret ved anvendelse af krydsvalidering.

Det er karakteristisk for den valgte model, at den tvinger forklaring fra en tidlig årsvariation til en forklaring via klimavariabel – primært udetemperaturen. I (Nielsen & Madsen 1997) er det demonstreret at dette også kan opnås ved at inkludere en beskrivelse af den dynamiske respons på udetemperatur og globalstråling. I (Nielsen & Madsen 1998) er der givet forslag til hvorledes dette kan inkluderes uden at ændre på den anvendte model.

# A Data



## B Matematisk formulering af model

I matematiske termer kan den implementerede model repræsenteres som

$$P_t = \mu + Y_t + D_t + f_1(T_t) + f_2(W_t) + f_3(R_t) + e_t, \quad (3)$$

hvor  $P_t$  er elforbruget i tidsrummet  $t - 1h$  til  $t$ ,  $\mu$  er en konstant,  $Y_t$  er årssvingningen,  $D_t$  er døgnsvingningen,  $f_1$ ,  $f_2$  og  $f_3$  er funktioner,  $T_t$ ,  $W_t$ , og  $R_t$  er gennemsnit for tidsrummet  $t - 1h$  til  $t$  af h.h.v. udetemperatur, vindhastighed og globalstråling. Endelig er  $e_t$  afvigelsen mellem den observerede værdi af elforbruget og den værdi, der må forventes, under antagelse af at modellen giver en rimelig beskrivelse.

Temperaturafhængigheden modelleres som en logistisk kurve

$$f_1(T_t) = \frac{a_{T2}}{1 + \exp(a_{T1}(T_t - a_{T0}))}, \quad (4)$$

der udviser mætning for både små og store temperaturer. Vindafhængigheden modelleres som en lineær afhængighed

$$f_2(W_t) = a_W W_t. \quad (5)$$

Endelig modelleres afhængigheden af globalstrålingen som en (aftagende) eksponentialfunktion

$$f_3(R_t) = a_{G2} \exp(-a_{G1} R_t). \quad (6)$$

Parametrene  $a..$  er ukendte konstanter, der søges fastlagt ud fra målinger.

Årssvingningningen modelleres ved en kubisk B-spline basis med ortogonale søjler (de Boor 1978, Chambers & Hastie 1991). Den anvendte basis har en overvægt af knuder placeret i sommerperioden, samt i starten og slutningen af året, se appendiks C, i alt 27 interne knuder, svarende til 30 frihedsgrader. Det bemærkes, at den herved opnåede årsvariation ikke har samme værdi ved starten og slutningen af året. Den vil således også beskrive en tendens i middelværdien.

Døgnvariationen modelleres som i (Nielsen & Madsen 1997). Med henblik på at nedsætte beregningstiden, skelnes der ikke mellem forskellige typer af hverdage, dvs.

$$D_t = \sum_{i=1}^3 I_i(t) \left[ \mu_i(t) + \sum_{j=1}^5 \left( c_{1ij}(t) \cos \frac{2\pi j h_d(t)}{24} + c_{2ij}(t) \sin \frac{2\pi j h_d(t)}{24} \right) \right], \quad (7)$$

hvor  $I_i(t)$  er en 0-1 variable der er 0, pånær når dagen svarende til  $t$  høre til dagsgruppe  $i$  (1 = hverdage, 2 = lørdage og halve helligdage, 3 = søndage og andre helligdage),  $h_d(t)$  er tiden på døgnnet, og  $\mu_i(t)$  er en dagsgruppe bestemt afvigelse fra den overordnede middelværdi bestemt af  $\mu$  i (3),  $\mu_i(t)$  kan variere med tiden på året. Af hensyn til entydigheden af estimerne kræves det, at

$$\mu_3(t) = -\mu_1(t) - \mu_2(t).$$

Både  $\mu_1(t)$  og  $\mu_2(t)$ , samt koefficienterne i døgnsvingningen  $c_{1ij}(t)$  og  $c_{2ij}(t)$  ( $i = 1, \dots, 3$ ,  $j = 1, \dots, 5$ ) modelleres som 5. ordens Fourierudviklinger af årssvingningen i disse koefficienter. Idet  $\alpha(t)$  betegner en af disse koefficienter gælder således

$$\alpha(t) = \gamma_0 + \sum_{i=1}^5 \left( \gamma_{1i} \cos \frac{2\pi i h_y(t)}{24 \times N_d} + \gamma_{2i} \sin \frac{2\pi i h_y(t)}{24 \times N_d} \right),$$

hvor parametrene  $\gamma_{..}$  er forskellige for alle koefficienter i døgnsvingningen.  $N_d$  er antallet af dage i det pågældende år, og  $h_y(t)$  er tiden på året målt i timer.

## C Generering af den anvendte B-spline basis

Den anvendte basis er genereret v.h.a. følgende S-PLUS 3.4<sup>3</sup> kommandoer (> angiver den primære prompt og + (helt til venstre) angiver den sekundære prompt):

Først vælges interne knuder<sup>4</sup>:

```
> iknots <- unique(sort(c(seq(0, 0.4, length = 4)[-1],
+   seq(0.4, 5.5, length = 8)[ - c(1, 8)],
+   seq(5.5, 7.5, length = 10),
+   seq(7.5, 11.5, length = 6)[ - c(1, 6)],
+   seq(11.5, 12, length = 5)[-5])))
```

Hvilket giver følgende knudeplacering i enheden "forløbne måneder":

```
> iknots
 [1] 0.1333333 0.2666667 0.4000000 1.1285714 1.8571429 2.5857143
 [7] 3.3142857 4.0428571 4.7714286 5.5000000 5.7222222 5.9444444
[13] 6.1666667 6.3888889 6.6111111 6.8333333 7.0555556 7.2777778
[19] 7.5000000 8.3000000 9.1000000 9.9000000 10.7000000 11.5000000
[25] 11.6250000 11.7500000 11.8750000
```

For normale år dannes nu en kubisk B-spline basis med ortogonale søjler som

```
> bs(seq(from=0, to=12, length=24*365), knots=iknots))
```

mens følgende anvendes for skudår

```
> bs(seq(from=0, to=12, length=24*366), knots=iknots))
```

Se evt. (Chambers & Hastie 1991).

<sup>3</sup>Se evt. <http://www.mathsoft.com/splus/>

<sup>4</sup>Kommandoen er anført for at kunne genskabe knudernes placering i dobbelt præcision.

## Litteratur

- Chambers, J. M. & Hastie, T. J., eds (1991), *Statistical Models in S*, Wadsworth, Belmont, CA.
- de Boor, C. (1978), *A Practical Guide to Splines*, Springer Verlag, Berlin.
- Jensen, S. E. (1997), 'Agroclimate at taastrup 1966....1996. agrohydrology and bioclimatology', Department of Agricultural Sciences. The Royal Vet. and Agric Univ., Copenhagen.
- Madsen, H. (1995), 'Tidsrækkeanalyse, 2. udgave', Institut for Matematisk Modellering, Danmarks Tekniske Universitet.
- Nielsen, H. A. & Madsen, H. (1997), *Development of methods for evaluation of electricity savings and load levelling measures, Part 1: Aggregated Power Consumption Models for the Eastern Part of Denmark*, Department of Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark in collaboration with NES A/S, DK-2900 Hellerup, Denmark. EFP95/1753/95-0001.
- Nielsen, H. A. & Madsen, H. (1998), *SAS-makroer til estimation og prediktion i elforbrugsmode*, Teknisk rapport nr. 19, Institut for Matematisk Modellering, Danmarks Tekniske Universitet, Lyngby.
- SAS Institute Inc. (1990), *SAS Guide to Macro Processing, Version 6*, second edn, SAS Institute Inc., Cary, NC.