



EME Forecast:

A heuristic approach to short term load forecasting

Göran Koreneff

Email: goran.koreneff@vtt.fi

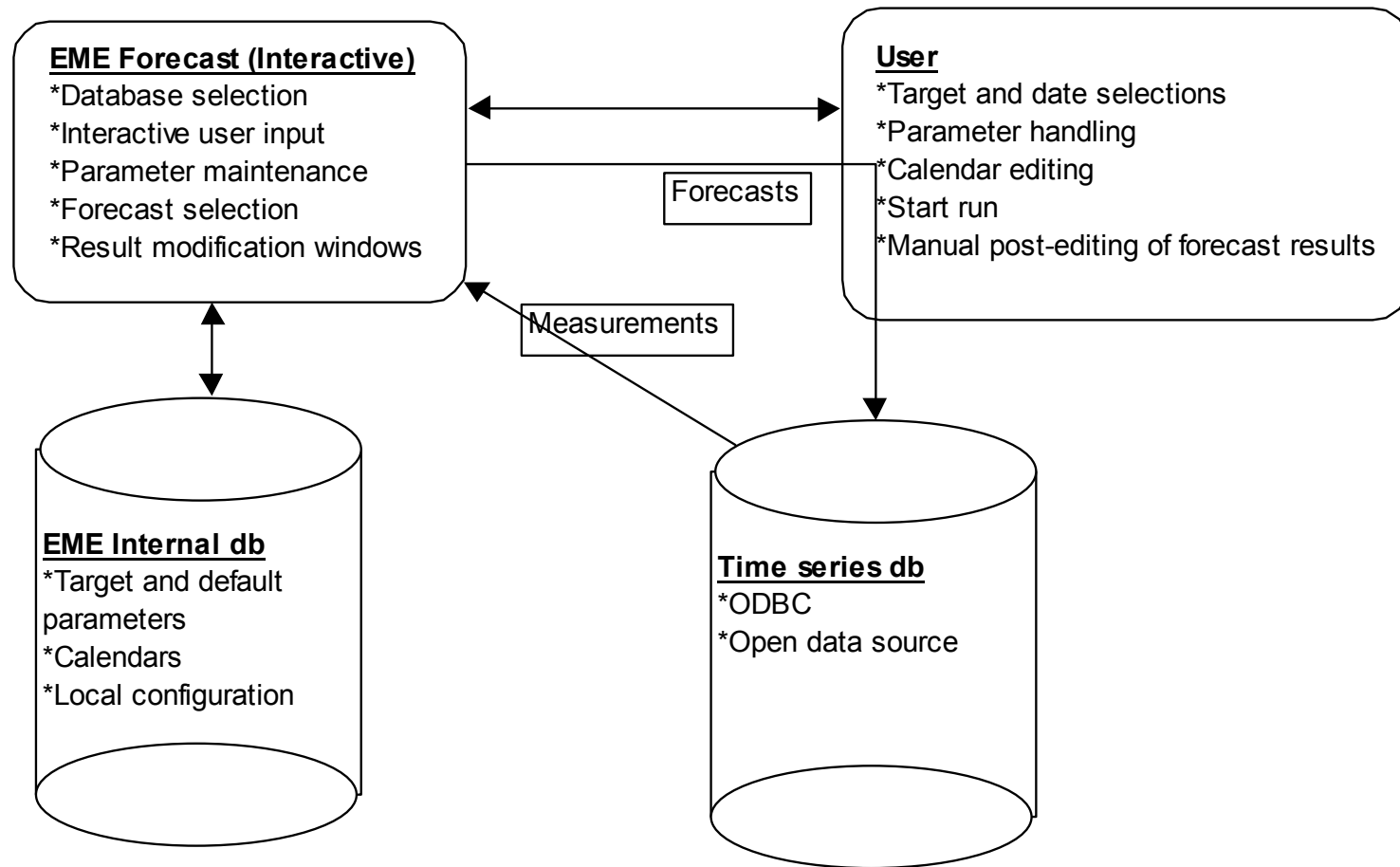
Seppo Kärkkäinen

seppo.karkkainen@vtt.fi

EME Forecast

- ◆ EME Forecast is a computer software for hourly load short ...medium term forecasting. It is a heuristic time series algorithm, trying to use the best sides of more formal approaches like ARIMA, regression analysis and load type libraries while at the same time avoiding their weak sides.
- ◆ EME Forecast was especially designed for electricity load forecasting in the de-regulated market. Several features take into account the time lap, one or more days, before measurements finally reach the user.
- ◆ Nonetheless, EME Forecast is a generic tool for all kind of energy sector forecasting, including district heat load and NordPool spot price, assuming the target has an appropriate day, week and season cycle.
- ◆ EME Forecast is used interactively or automated. It is easy to take into use, because no target specific parametrization by the user is required
- ◆ EME Forecast features a flexible and dynamic calendar system, as well as easy to use after-forecast tools .

EME Forecast: program operational environment



EME Forecast environment info

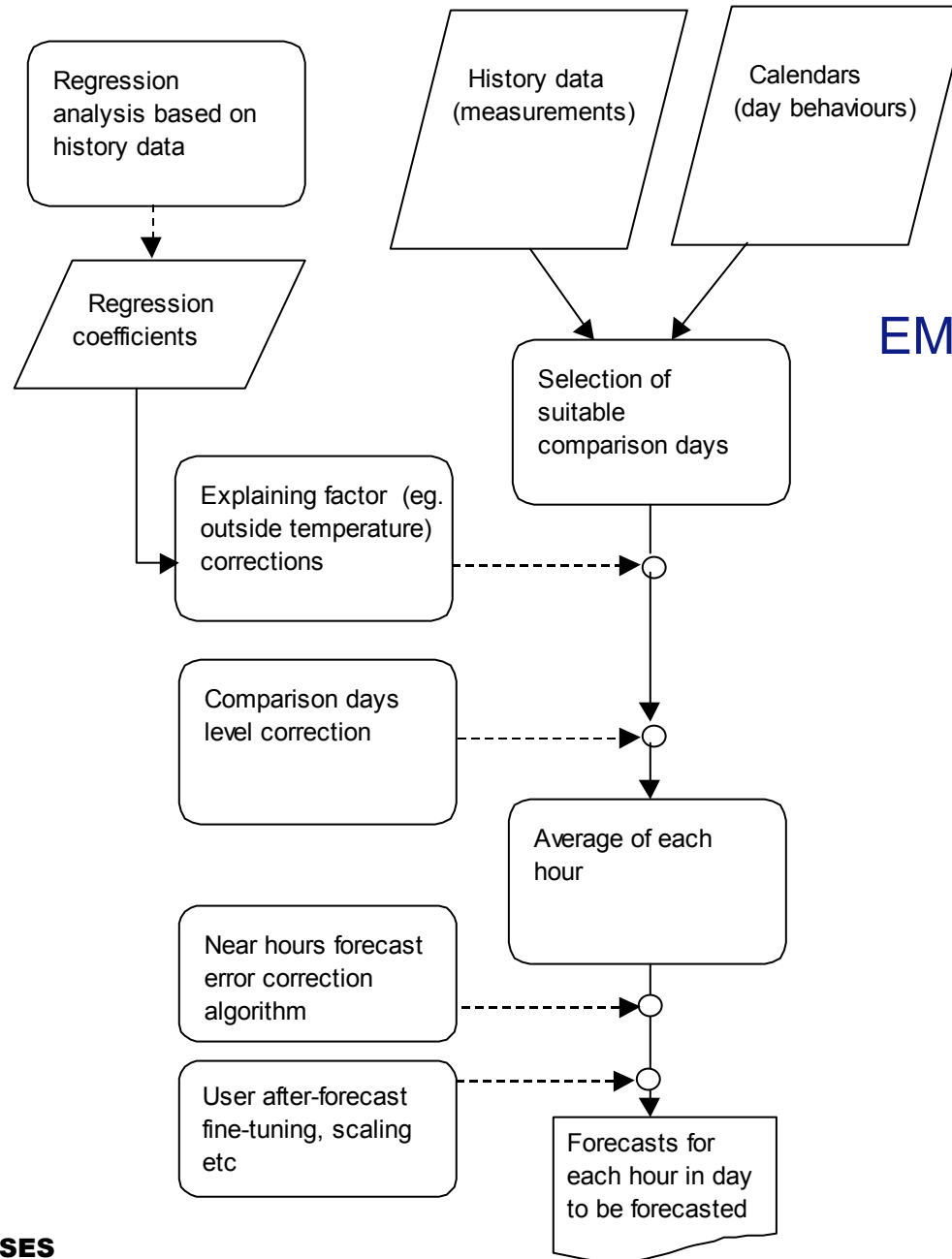
- ◆ Environment: Windows
- ◆ Programming language: APL (Dyalog APL 9.0)
- ◆ Data management: ODBC, sql-statements in text file (security, what's that?)
- ◆ Database: relational database, excel spreadsheet
- ◆ Pilot user: Turku Energia (Finnish municipal electricity utility)
 - in addition, previous version (DEM Forecast) has some other active users

(APL, what's that? See for example

<http://www.dyalog.com/>

<http://www.izap.com/~sirlin/apl/apl.faq.html>

[http://dmoz.org/Computers/Programming/Languages/APL/ \)](http://dmoz.org/Computers/Programming/Languages/APL/)



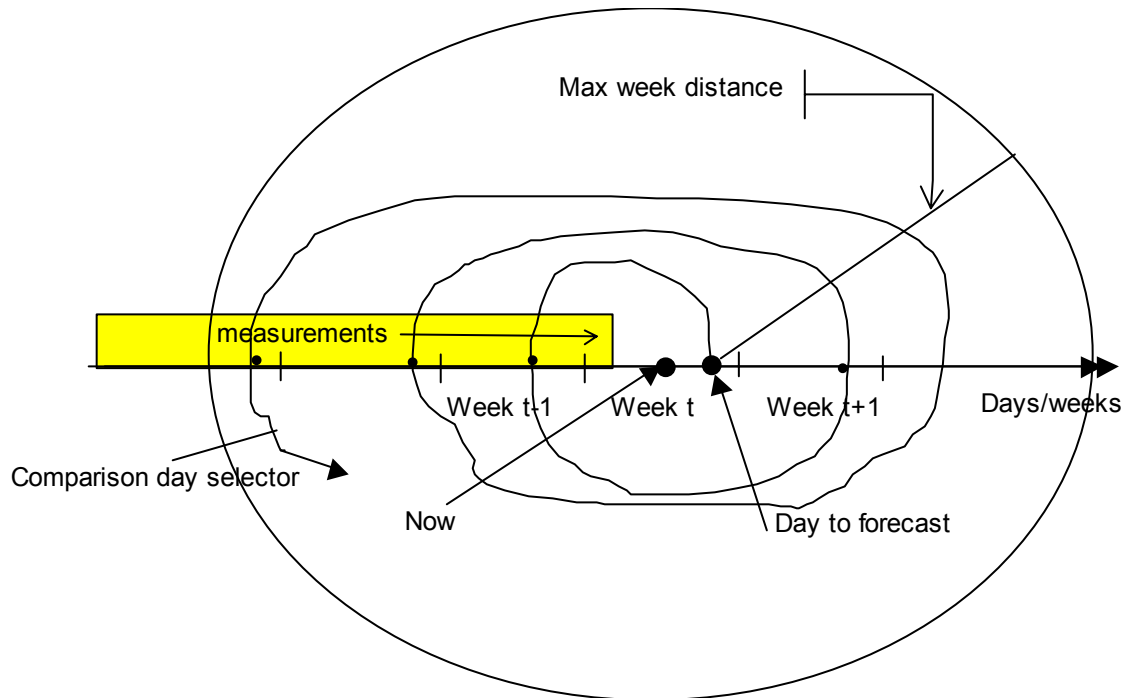
EME Forecast program schema

- Main core consists of selecting comparison days and calculating their hourly averages
- Averages can also be taken hour values cleaned from high and low values (median area average)
- Selectable features are marked with a dashed line

Main algorithm

- ◆ The forecasting method of EME Forecast is based on the usage of comparison days from history. For each day to forecast a dynamic set of comparison day selection criteria is created. The criteria takes into consideration day-of-week, day type including special holidays, week and several user definable parameters. Comparison days are sought in a iterative fashion with a widening search criteria until one or more days are found.
- ◆ Outside temperature is a strong explaining factor. EME Forecast allows for the inclusion of one user chosen explaining factor, which influence is automatically determined through a detached regression analysis process.
- ◆ The explaining factor must be a hour time series like the target. It is up to the user to decide if the data input will be pure measurements or, for example, a 12 hours moving average.

Comparison days selection method 1/4



Main rules in primary comparison days selection:

$comp\ day_i = f(\text{weekday}_{forecastday}, \text{week})$, where
 $|\text{week}(comp\ day_i) - \text{week}(forecastday)| \leq \text{Max week distance}$
 $\text{week} \in 1, \dots, 53$ and
 $i \in 1, \dots, \text{Max comparison days}$

Additional selection parameters, among others:

- quarantine of near history
- behaviour breakpoint date

Comparison days selection method 2/4

◆ Breakpoint of behaviour:

- comparison days are primarily searched from days with similar behaviour
- in an open market several customers can change supplier at the same time (e.g. at first of November), having a severe and noticeable impact on the load

◆ Quarantine:

- customer measurements are received from other network areas with a delay of one or more days
- received measurements from different senders can, at first, be offbeat resulting in inexact agglomerated loads. At the beginning of the free electricity market the measurements had a proneness to error for several days after the event, and exact history data being set several weeks afterwards. It may be best even nowadays to put the nearest history day in quarantine for forecast purposes.

Comparison days selection method 3/4

Calendar system (1)

- ◆ EME Forecast has a flexible calendar system, enabling multiple calendars. A calendar has information about which days are special days and how they are to be forecasted.
- ◆ Each forecast target is connected to either the default calendar or a named calendar.
- ◆ Each calendar consists of a collection of special days. All special days' definitions are in a common and editable table, which is available to all calendars. The same holiday can have several definitions in the table, but they must be differently named (e.g. newyear1, newyearSWE, newySpot), and only one, or none, is selected to each calendar.
- ◆ Most important, special days are not used as comparison days. The only exception is when the user chooses to have the previous same special day as a comparison day. For example, when forecasting Christmas day it might be a good idea to set the previous year's Christmas day as a comparison day.

Detached regression analysis 1/4

- ◆ Regression analysis is done on history data and with only one explaining factor
- ◆ Regression analysis is optional, and is only performed when synchronous load and explaining factor measurements are available
- ◆ The detached analysis module performs linear regression analyses by using the least squares method to fit a line through a set of observations

$$P = \beta_0 + \beta_1 \cdot T + \varepsilon$$

where

P is the forecast target, e.g. load,

T is the regressor, the explaining factor, e.g. outside temperature

β_0 is the regression constant,

β_1 is the regression coefficient and

ε is the error.

Detached regression analysis 2/4

- ◆ As the best regression analysis results are achieved when only those factors involved in the load variation are left variable and all other factors are held constant, the day, week and season variations are eliminated by analysing each hour, day type and month separately.
- ◆ $12 * 24 * 3 = \text{months} * \text{hours} * (\text{weekdays/Saturdays/Sundays}) = 864$ separate analyses.
- ◆ The regression coefficient is calculated according to:

$$\beta_1 = \frac{\sum_{i=1}^n (T_i - \bar{T})(P_i - \bar{P})}{\sum_{i=1}^n (T_i - \bar{T})^2}$$

Detached regression analysis 3/4

- ◆ The minimum correlation, R_{\min} , must be exceeded for a coefficient to stand:

$$R = \frac{\text{Cov}(P, T)}{\sigma_P \sigma_T} \leq R_{\min}$$

where

$\text{Cov}(P, T)$ is the covariance between P and T,

σ_P is standard deviation of P,

σ_T is standard deviation of T and

R_{\min} is user set minimum correlation.

- ◆ Results are further refined by comparing day hour respectively night hour results which each other for more stable and homogeneous regression coefficients:

$$\hat{\beta}_{month, daytype, hvrk} = \frac{1}{\sum_{hvrk} (R_{month, daytype, h} > R_{\min})} \sum_{hvrk} \beta_{month, daytype, h} (R_{month, daytype, h} > R_{\min}) \quad \forall month \in 1, \dots, 12$$

daytype \in weekdays, Saturdays or Sundays

hvrk \in night hours, day hours

Detached regression analysis 4/4

- ◆ Only the regression coefficients, not the regression constants, are used in converting comparison days to match forecasted temperatures or whatever chosen as explaining factor:

$$P^* \text{comp}_{c,i} = P\text{comp}_{c,i} + \beta_1 \cdot \Delta T_{c,i} \quad \forall c \in \text{comparison days}, i \in \text{hours} [1, \dots, 24]$$

where

$P^* \text{comp}_{c,i}$ is the regression corrected comparison day c hour i value,

$P\text{comp}_{c,i}$ is the uncorrected comparison day c's value at hour i,

ΔT is $T_{\text{fore}_i} - T_{\text{comp}_{c,i}}$,

T_{fore_i} is the forecasted explaining factor for hour i and

$T_{\text{comp}_{c,i}}$ is the comparison day c's explaining factor value at hour i.

- ◆ Regression correction is thus possible only when the we have forecasts for the explaining factor. The program gives an option to manually write the forecasted values, in case they are not found in the database.

Forecast level correction

- ◆ Comparison days can be, and are, weeks old. But the level can change fast, especially in spring and autumn.
- ◆ Level correction (optional) is based on scaling the day energies of the comparison days to the day energy of the most recent measured suitable day. The level correction is calculated for each comparison day separately, taking into account that the nearest measured day can be of a different day of week type:

$$\text{level correction ratio}_v = \frac{W_{\text{comp}}_{w_0, \text{day}w_M}}{W_{\text{comp}}_{w_V, \text{day}w_M}} = \frac{W_{\text{fore}}_{w_f, \text{day}w_f}}{W_{\text{comp}}_{w_V, \text{day}w_f}} \quad \forall w_V \in W$$

where

W_{comp} is the day energy of the comparison day,

$W_{\text{fore}_{w_f}}$ is the day energy of the forecasted day,

$\text{day}w_M$ is the day of week of the nearest measured and suitable day,

w_0 is the week of the nearest measured and suitable day,

$\text{day}w_F$ is the day of week of the forecast day as well as the comparison days,

w_V is the week of the comparison day v and

W is the group of all the comparison days' weeks.

EME Forecast: additional parameters

- ◆ A separate short term forecast error correction module:
 - based on forecast errors of the few, nearest hours
 - corrects the forecasts of the near future hours
 - correction basically a declining correction, but user can set correction weights for each near future hour separately

- ◆ Annual growth (%):
 - especially for long term forecasts
 - growth (or decline) linear to the distance between comparison and forecast day

- ◆ The forecast of each hour is the average of corresponding comparison days' values. The average can also be calculated as a median area average. The median area is achieved by chopping off the highest and lowest comparison days' values for the corresponding hour.

Some test results 1/5

Utility load

- ◆ The test:
 - At 00:00, the 22nd of each month, a 5 day ahead forecast is made using outside temperature as explaining factor. Measured temperatures are used as forecast temperatures.
 - Forecast target is a real utility load where measurements are available only for the year in question. History data used is always prior to forecast start hour. Default parameters and calendars are used, both regression and level corrections are on, as well as the short-term forecast error correction.
- ◆ The mean absolute percentage error (MAPE) for each forecast day (4th and 5th together) is shown in following table for the utility load:

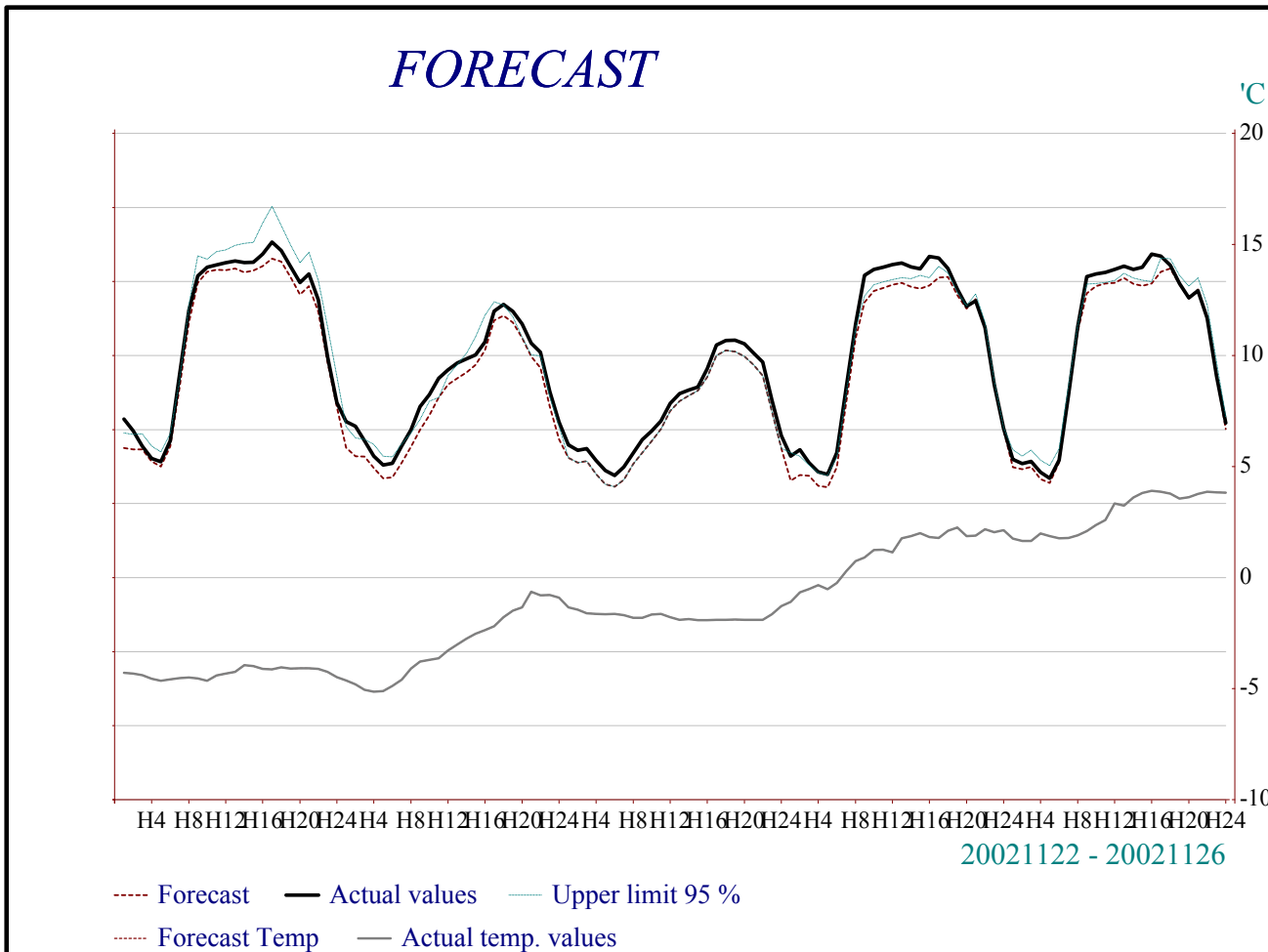
MAPE	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1st day	1,2	1,5	2,2	6,4	1,1	10,2	3,8	2,3	2,0	3,0	1,1	1,9
2nd day	2,1	2,6	2,3	2,3	1,3	10,7	1,6	1,6	3,6	3,0	2,3	6,7
3rd day	1,3	1,9	1,1	1,1	2,1	6,0	1,3	5,0	3,6	8,5	0,9	6,7
4&5th days	3,9	1,5	2,1	2,1	2,8	5,4	2,0	4,2	3,4	6,9	2,5	4,5

*Midsummer

* Christmas

Some test results 2/5

Utility load, forecasted at 24:00, the 21st of November



Upper limit 95%:
upper end of the 95%
confidence interval

Y-Axis scale
unmarked to protect
the innocent

Forecasted at
21.11.2003 24:00

X-axis scale
unintentionally
unreadable

Some test results 3/5

District heat load

- ◆ At 00:00, the 22nd of each month, a 5 day forecast ahead is simulated using outside temperature as explaining factor, and measured temperatures as forecast temperatures.
- ◆ Target is a real district heat load, measurements are available only for the year in question. History data used is always only prior to start hour. Default parameters and calendars are used, both regression and level correction are on, as well as the short-term forecast error correction.
- ◆ The mean absolute percentage error (MAPE) for each forecast day is shown in following table for the district heat load:

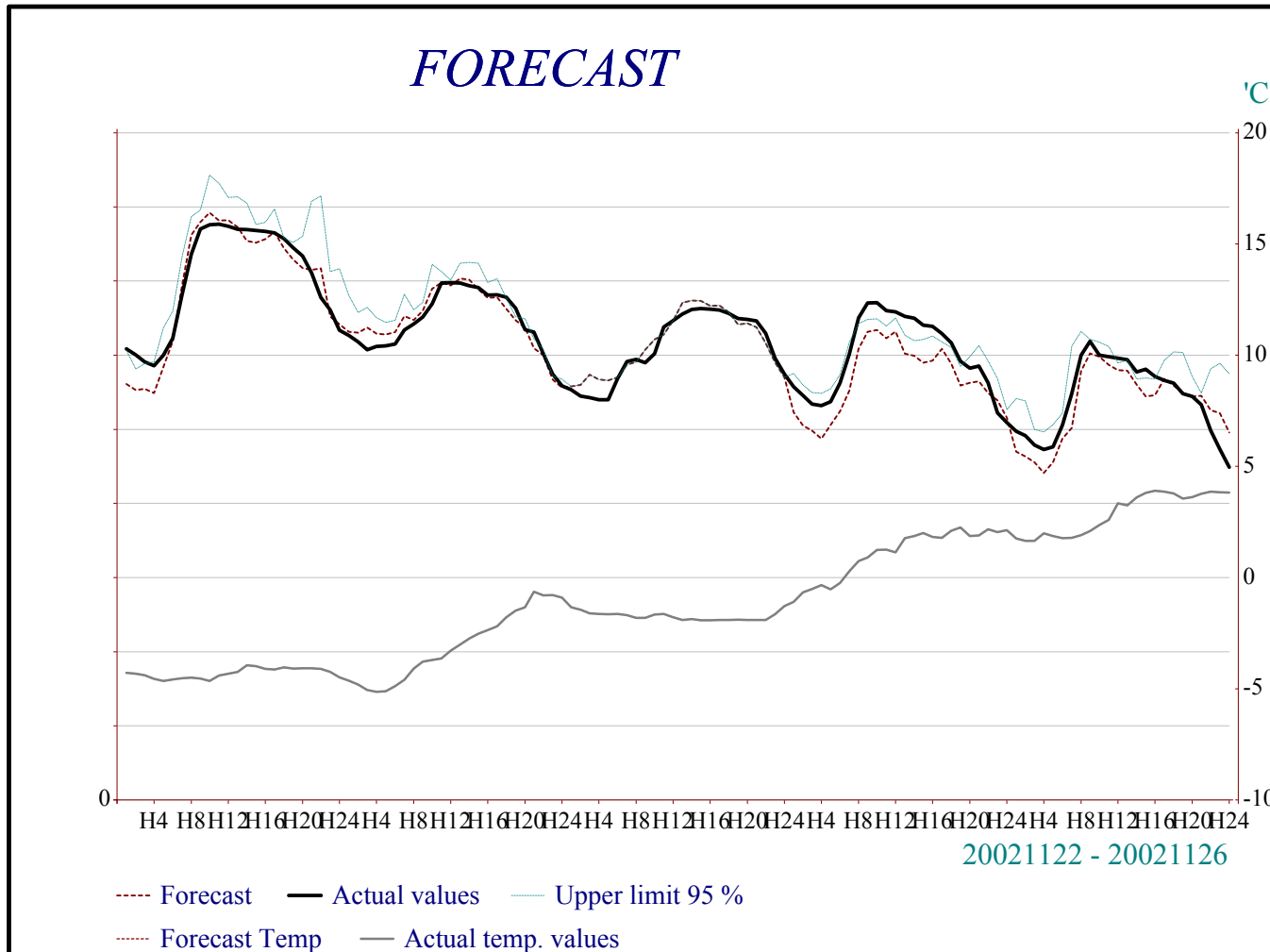
District heat load												
MAPE	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1st day	3,9	2,5	2,6	14,9	9,7	12,5	7,4	6,2	17,0	4,8	2,6	1,2
2nd day	5,2	2,0	3,3	9,0	11,2	25,9	3,5	7,6	12,3	10,7	1,6	1,4
3rd day	5,9	1,9	3,1	14,5	8,9	7,5	4,8	3,8	11,7	5,1	1,8	3,7
4&5th days	15,7	3,1	6,1	9,5	7,2	8,0	7,1	6,8	10,5	5,9	4,8	3,4

*Midsummer

* Christmas

Some test results 4/5

District heat load, forecasted at 24:00, the 21st of November



Upper limit 95%:
upper end of the 95%
confidence interval

Y-Axis scale
unmarked to protect
the innocent

Forecasted at
21.11.2003 24:00

X-axis scale
unintentionally
unreadable

Some test results 5/5

Comments and insights

- ◆ Forecasts for the first day are not always better than for later days. Because of the forecast approach of using comparison days, forecasting exactly at midnight so-to-speak is the only time where the short-term forecast error correction often misbehaves. Especially when the first hour is of a different day type as the last measurement, as in both April and July, the short-term error corrections are a hazard.
- ◆ January 26th is a Saturday. Regression coefficients for Saturdays in January were automatically dropped, because they didn't meet the preset goodness criteria, a consequence of too similar temperatures for previous Saturdays in January. Thus temperature corrections are left out, which can be noticed in the higher MAPE for days 4&5.
- ◆ June 22nd is Midsummer day, a very special holiday at least in Finland. And Christmas is Christmas. Longer history data and matching special day definitions could make a big difference for these days' forecasts.
- ◆ District heat is measured as produced, not consumed. Different physical location mix of the production, specially in spring, results in longer and different lag times and thus "different" and more stochastic load measurements.

Future improvements

- ◆ The level correction algorithm should be split into at least two parts -a day time and a night time- for better overall performance and specially to better manage behaviourally transitional time spans, e.g. spring and autumn.
- ◆ Short term forecast error correction should be deactivated when forecasting is done “at midnight”.
- ◆ Regression analysis results for Saturdays and Sundays need more refinement rules.
- ◆ Default calendar could be redone using an other philosophy than ad-hoc
- ◆ The creation of a generic, adaptive parameter management module would be a taxing and thereby also an highly interesting future task.