21, rue d'Artois, F-75008 PARIS
http : //www.cigre.org

C2-202

CIGRE 2008

Various methods for forecasting cross-border power flows in the Hungarian transmission system

BÜRGER László, GÖLÖNCSÉR Péter, DECSI Tamás, SEBESTYÉN Géza, OROSZKI Lajos*

MAVIR ZRt., PROVENTUS Kft.

Hungary

SUMMARY

The deregulation of electricity market is a challenge for transmission system operators all over the world. The era of the well predictable electrical environment is slowly going away. Some time ago those rapid and large exchange program fluctuations that characterize today's European interconnected power system were not conceivable. The loop flows that they generate have a severe impact not only on the grid of the countries directly involved in the trading but their neighbours as well. These impacts are undoubtedly the most observable in the variations of cross-border interface flows. Foreseeing the cross-border interface flows may indicate the possibility of congestions and can help prevent dangerous operating situations. The Hungarian Transmission System Operator has been producing day-ahead forecasts with hourly resolution since 2004. From that time up to now three different methods were introduced, each one complementing the other. Historically the first one relies on a PTDF (Power Transfer Distribution Factor) based model, the second one utilizes a purely statistical algorithm, while the third one is an ancillary method to improve the quality of the previous two forecasts. This paper is the detailed description of the aforementioned methods and the experiences obtained during the past three years.

KEYWORDS

Power - Flow - Interface - Forecast - PTDF - Statistical - Gradient - Tie-line

oroszki@mavir.hu

POSING THE PROBLEM

At the first glance forecasting cross-border power flows seems to be evident, nothing noticeable, it happens each time somebody solves a load-flow case. But to do that there is a need for a vast amount of data. Producing a day ahead, hourly resolution forecast requires 24 network models from all TSO-s running in parallel. Merging the models of the individual countries is a further challenge and on top of that the calculations must be fulfilled within a narrow timeframe. In spite of that there are serious attempts to forecast network flows in this manner. To pick up just one real example there is probably no one in the industry who has not heard about DACF files [1] whose name is an acronym for Day Ahead Congestion Forecast. UCTE countries have been exchanging their DACF models for about a decade. At the very beginning this cooperation was limited to one network model per day that referred to the time 10:30 CET. By now most participants provide network models for 03:30, 10:30, 12:30 and 19:30 CET but it is still not sufficient to produce hourly resolution forecasts for cross-border power flows and not only because of the missing models, there are other problems as well. Officially the DACF models must be published on the Electronic Highway by 18:00 CET. That is too late on the one hand and data delivery is often delayed on the other hand. This time the working hours are normally over which practically means that the network merging and solving process is deferred for the next day that is the forecasted day. Mainly that explains that the available time for creating the forecasts is highly pressing and not sufficient in many cases. Oppositely, the European exchange programs are still open when the network models must have already been finalized. When it comes to render the forecast results it leads to uncertainties and undoubtedly this is the most severe objection to the methods that try to restrict the prediction of cross-border flows to load flow calculations. To worsen the situation the network models are in a so called UCTE format that is not directly handled by many network simulation programs of the TSO-s. The back and forth conversions from the native data format of the applied network simulation programs to UCTE format slow the whole process further down and lead to many model errors that must be corrected at the very last stage of calculations. Sometimes the network models are so faulty or so many TSO models are missing that building of the whole network model becomes impossible. In these cases strictly insisting on the load flow approach would be disastrous, as the forecast users would not have any idea how the cross-border flows will evolve. The problems enumerated so far formed the basis for developing more reliable, computationally less demanding methods that can circumvent the lack of appropriate load flow data and for which the adaptation to the frequent variation of power exchanges is not so difficult.

SURVEYING THE INPUT DATA

To reach the goals worded in Section 1 there was a need for a completely new set of input data. First and foremost it was straightforward that the new methods should utilize the historical power flow data. In this respect there were two possibilities: either using the high accuracy hourly consumption amounts that were registered by the energy accounting system or resorting to the EMS/ SCADA system. While the accounting system provided the necessary average values of the power flows, the SCADA system recorded only the 6 second instant values. Despite of that the SCADA system proved to be the most favourable. It was not as accurate as the accounting system but when it came the time it never missed to provide the necessary data. Beyond the historical power flow values there was a UCTE homepage that turned out to be indispensable. It was VULCANUS. At this homepage published the day ahead forecasted exchange programs of the European countries, moreover it suited well the next presented forecast methods as all of its data was in hourly resolution.

THE PTDF BASED APPROACH

The first attempt to get rid of the burden of the load flow calculations was inspired by the flow based capacity allocation methods [2]. The calculations still relied on forecasted load flow data, but the applied PTDF coefficients eliminated the rigid coupling between forecasting the cross-border power flows and building the forecasted network models.

To get an understanding how the PTDF based method works let us have a look at Figure 1 that shows a system consisting of N control areas. The control areas are connected via tie-lines which are grouped into interfaces. For instance the interface between control areas 2 and 3 is made up of 4 tie-lines and the power flow along these lines is denoted by P^{23}. Equation (1) approximates the power flowing through the interfaces with a linear formula. The PTDF coefficients implied are normally derived from the admittance matrix of the network load flow model.

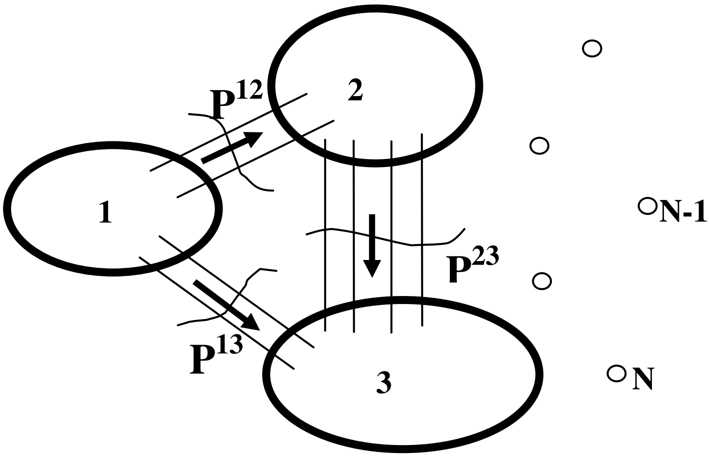


FIGURE 1

$$\mathrm{P}^{ij}(\mathrm{t}) = \mathrm{P}_0^{ij}(\mathrm{t}) + \sum_{1 \le k, l \le N}^{k<l} \mathrm{PTDF}_{kl}^{ij}(\mathrm{t}) \bullet \mathrm{E}_{kl}(\mathrm{t}) \tag{1}$$

As the geographical distribution of power generation and consumption inside a control area never coincide, in parallel running systems there are always interface flows even if all of the power exchanges are zero. The first constant term on the right hand side of (1) takes into account these natural flows. The remaining part that describes the effect of power exchanges may be interpreted as follows: the power exchange E_{kl} between the systems K and L changes the interface flow of the systems I and J by the product of E_{kl} and $PTDF_{kl}^{ij}$, where $PTDF_{kl}^{ij}$ is a power transfer distribution factor that expresses the sensitivity of power flows between the systems I and J to the power exchanges from system K to L.

Assuming that within a narrow timeframe the natural flows and the distribution factors do not change significantly one can derive a relationship that is suitable for predicting the cross-border power flows. A linear equation can be gained to predict interface flow from the following equations.

To prove that, the above assumptions are formed from (2) to (6) in five mathematical expressions then the desired relationship (7) is obtained by subtracting (2) from (3) and rearranging the result.

$$P^{ij}(t_0) = P_0^{ij}(t) + \sum_{1 \le k, l \le N}^{k<l} PTDF_{kl}^{ij}(t) \bullet E_{kl}(t_0) \tag{2}$$

$$P^{ij}(t_1) = P_0^{ij}(t) + \sum_{1 \le k, l \le N}^{k<l} PTDF_{kl}^{ij}(t) \bullet E_{kl}(t_1) \tag{3}$$

$$P_0^{ij}(t) = P_0^{ij}(t_0) \tag{4}$$

$$PTDF_{kl}^{ij}(t) = PTDF_{kl}^{ij}(t_0) \tag{5}$$

$$\mathrm{t}_0 \le \mathrm{t} \le \mathrm{t}_1 \tag{6}$$

$$P^{ij}(t_1) = P^{ij}(t_0) + \sum_{1 \le k, l \le N}^{k<l} PTDF_{kl}^{ij}(t) \bullet (E_{kl}(t_1) - E_{kl}(t_0)) \tag{7}$$

$$t_0 \le t \le t_1 \tag{8}$$

Equation (7) says that the cross-border power flows can be forecasted by taking their values for a reference time in the past and correcting that value with the weighted sum of power exchange program differences, where the weights are equal to the power transfer distribution factors in Equation (1) .

Based on this formula the Hungarian TSO has been producing day ahead, hourly resolution forecasts since 2004 February. As for the natural flows the applied assumption is less restrictive than what is formed in equation (4). Putting down a daily periodicity to $P_0^{ij}(t)$ the difference between t_1 and t_0 is chosen exactly to 24 hours. The PTDF based forecasts are available in three different versions. Each version adapts a single set of power transfer distribution factors. Version 3 that is the most accurate uses the distribution factors that are calculated from the 10:30 CET DACF model of the forecasted day. Version 2 works with the 10:30 distribution factors of the reference day. Version 3 is a bit special as it applies an average PTDF data set. Applying average distribution factors turned out to be very useful as they give reasonable good results and in the early hours on Mondays there is no any other PTDF data set. Version 3 is also indispensable when the merging of DACF models is not feasible.

The PTDF based method has solved many of the problems raised in the first section although it had some inherent weaknesses. The amount of load flow data demanded by the method has decreased significantly. Instead of 24 complete load flow models only single one was needed. Due to the simplicity of equation (7) the adaptation to the variations of power exchanges became very easy, the necessary calculations may have been done even in real time circumstances. In case of the most accurate forecasts the remarks concerning to the narrow time frame remained still relevant but for versions 2 and 3 there were no such problems. The assumptions regarding to the natural flows proved to be acceptable on weekdays but not on red letter days. For this reason the forecasts prepared for Mondays and Saturdays were less reliable. Grid malfunctions and topology changes made the method's limitations more apparent. Once one of these events happened the forecasts lost their accuracy for the whole day. The reasons of that were twofold: first the assumption about the 24 hour periodicity of the natural flows did not apply; then equation (5), prescribing that the PTDF data set must be updated daily, lost its validity. Finally, the PTDF based method did not provide any information about the expected range of forecast errors.

STATISTICAL APPROACH

The need of forecasting not only the expected values but the confidence intervals of the cross-border power flows led to the application of statistical methods. The forecasts were calculated based on the day ahead hourly resolution power exchange programs and the hourly resolution cross border power flows computed from the SCADA system measured values. This means that this approach does not use the PTDF data to produce the forecasts.

Since long term forecasts are always less reliable than the shorter term ones, three different statistical models were set up to produce the forecasts for the next 24 hours. The first is a short term model that uses the actual power flow data for the *t-1* and earlier hours to calculate the forecast for the *t* hour. Certainly, this model is also the most reliable from the three. The second model has a little longer time-horizon, it uses the actual power flow data for the *t-2* and earlier hours to calculate the forecast for the *t* hour. Finally, the third model has a longer time-horizon, since it only uses the actual power flow data for the *t-26* and earlier hours to calculate the forecast for the *t* hour. This last model is the less reliable one.

The construction of the statistical models was done in four main steps. In the first step the reliability of the power exchange programs were improved. This step was necessary because of the situations – like missing data for example – when the raw data was not reliable or directly applicable. The improvement was made by simple rules based on the statistical analysis of the earlier data and the observable trends. In the second step the variables used by the forecast models were specified. These are the power exchange programs for the given hour, some lagged power flow data and some lagged differentials of the power flow data for the given border. In the third step the forecast coefficients for the specified variables were calculated. It was done by linear regressions. And finally, in the fourth step the confidence intervals were defined. Using these data one can determine not only the expected cross-border power flows, but also an interval for any desired significance level. This means, that one can not only say, that the power flow on one border will be around 98 MW, but also that it will be between 45 and 151 MW with a 95% probability.

Table I shows the main statistics of the one hour forecast models for the six Hungarian borders. Each border is represented by the appropriate column.

TABLE I: One hour model statistics.

Model	A>H	RO>H	SH>H	SK>H	UA>H	YU>H
Interval	34.25	26.59	38.39	47.81	28.85	21.95
Probability	84.92%	82.94%	84.41%	82.09%	89.28%	70.76%
Correlation	0.9883	0.9866	0.9874	0.9831	0.9911	0.9755
R^2	0.9768	0.9733	0.9750	0.9665	0.9823	0.9516
Std. of the error term	26.24	21.82	29.99	43.93	19.25	24.29

The rows are showing the main statistics for each border. The first row, labelled as "Interval" shows us the standard confidence range, which is calculated from the standard deviation of the power flow data. The second row, labelled as "Probability" gives us the ratio of the cases, where the difference between the hourly actual power flow data and the forecasted value is less, than the above defined interval. This means for example, that for the Austrian border the difference between the actual and the forecasted values is less than 34.25 MW for the 84,92% of the cases. The third row, labelled as "Correlation" shows us the correlation between the forecasted and the actual values. The fourth row, labelled as "R^2" shows us the goodness of the models. A perfect model has a correlation and an R^2 of 1. And finally, the fifth row, labelled as "Std. of the error term" gives us the standard deviation of the error term, which is the difference between the forecasted and the actual values.

The one hour forecasts are very accurate. A relatively narrow interval is sufficient to contain the actual power flow data with a probability of 71-89%. The correlation between the actual and the forecasted values is more than 0.975, while the R^2 of the models is more than 0.951. The error terms have also a relatively low standard deviation. Although the 2 and 24 hour models are less accurate than the 1 hour model, they are still quite good, the correlation between the forecasted and the actual power flow values are always larger than 0.945.

Figure 2 shows the forecasted interval and the actual power flow for the one hour model of the Rumanian border. The blue and the red graphs represent the bottom and the top of the forecasted interval, while the green graph shows the actual data. For the presented 24 hour timeframe the actual data is always in the forecasted interval.

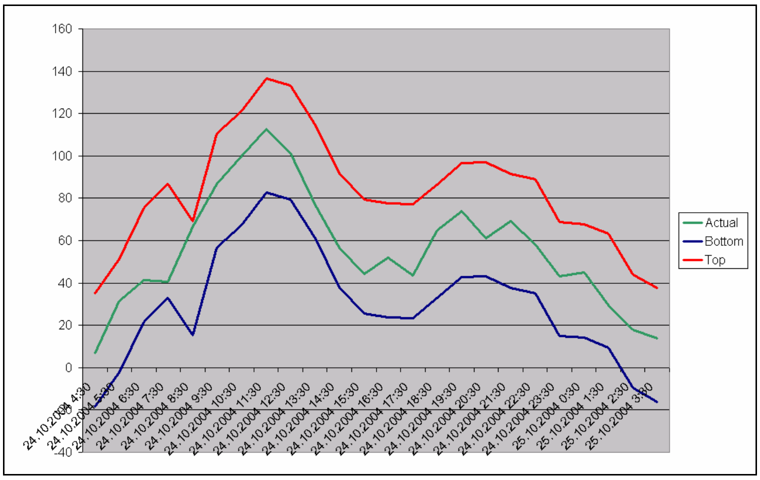


FIGURE 2: Forecasted range and actual values for the Romanian border.

The statistical approach has some serious advantages over the PTDF based method. First, it uses more input data, so it is more reliable. And in fact, the error terms of the statistical forecasts are on the average 50% smaller, than those of the PTDF based ones. Second, this approach does not need PTDF coefficients, thus it is less labour-intensive and exceedingly easy to automate.

FORECAST IMPROVEMENTS

To improve the short-term accuracy of the previously presented methods, a supplementary algorithm was developed. Hereafter it is referred to GRADINT.

The GRADINT algorithm currently does real-time adjustments on the forecasts that were produced by the PTDF method It disaggregates the interface-flows into tie-line Its mainstays are the followings: it is able to forecast the flows in non-average system states, it requires no additional data and it is able to follow the fast changes with smaller time-constant then the other algorithms.

The idea of the method is based on the gradient-monitoring algorithms that are used in the new digital protection systems. Because the topological changes such as tapping of a branch or a generator cause offset type, fast increasing errors the gradient-monitoring algorithms are applicable here too. Regarding the 6 sec. sampling time of the SCADA system switching event caused errors behave very similar to the step function that produces an infinite gradient at the time of switching. That makes the recognition of switching events possible. If the gradient of the forecast error exceeds an appropriate limit, it can be ascertained that the topology of the network has changed significantly and the formerly produced forecasts will introduce permanent, offset type errors. To get rid of these errors the forecasts should be biased with an offset that coincides with the forecast error at the instant of topology changes.

Gradient-monitoring is a useful tool to correct fast evolving forecast errors, but what about with the slowly increasing errors, caused by intraday schedule modifications.

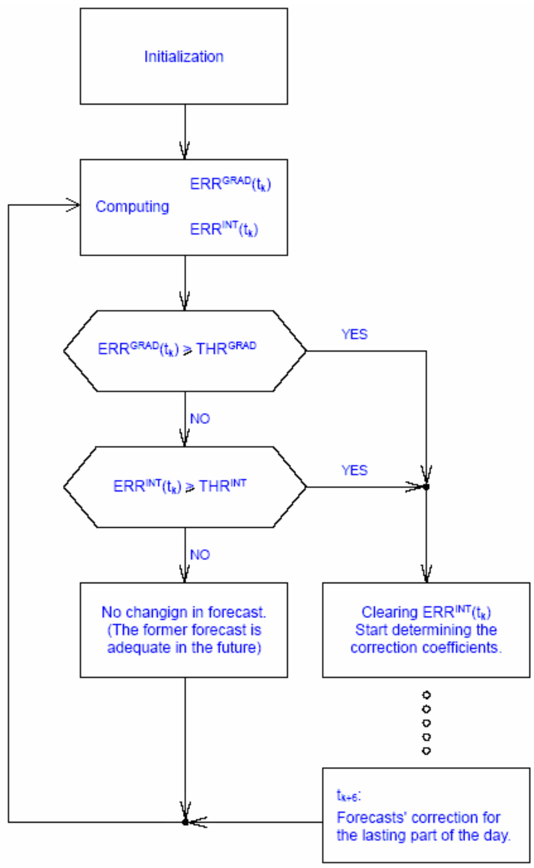


FIGURE 3

The gradient-monitoring is futile for this scenario. This drawback can be eliminated with the integration of each tie-line forecast's error and doing corrections if any of the integrated error is higher then a threshold value.

Let us see now, how the algorithm in practice works. On Figure 3 the flow-chart of the algorithm can be seen. Separation of the different error type detection is noticeable on the figure. The definition of the variables, that can bee seen on the figure are the followings: $ERR^{GRAD}(t_k)$ – the gradient error value at t_k, $ERR^{INT}(t_k)$ – the integral error value at t_k, THR^{GRAD} – threshold value for the tie-line gradient error, THR^{INT} – threshold value for the integration error.

Let us see the equations that are determining the operation of the algorithm. At first the initialization is done in two steps. In the first step the interface-flow forecasts are distributed to tie-line flow forecasts. In the second step the tie-line flow forecasts that are now given in hourly resolution are framed into 6 sec resolution data. The two steps are described by the Equation 9 and 10.

$$P_{l.f.h.(tie-line.j)}(i) = P_{intersection.forecast}(t_i)\cdot\frac{P_{line.flow(tie-line.j)}(t_0)}{\sum\limits_{intersection} P_{line.flow}(t_0)}, \quad \text{i=1..24} \tag{9}$$

where: t_i – the given hour in de day, t_0 – the starting time of the GRADINT algorithm, $P_{intersection.forecast}$ – forecasted flow value for the intersection by PTDF, $P_{line.flow}$ – thel power-flow value of the tie-line, $P_{l.f.h.(tie-linej)}$ – the hourly forecasted power-flow value of the tie-line j

$$P_{line.forecast}(t_i) = \begin{cases} if\ \_(i-600k)\in[-50;+50], then:\ \_P_{l.f.h}(k) + (P_{l.f.h}(k+1) - P_{l.f.h}(k))\cdot\dfrac{i-600k+50}{100}, \\ else:\ \_P_{l.f.h}(k+1) \end{cases} \tag{10}$$

$$k = \left\lfloor \frac{i}{600} \right\rfloor, \quad \text{i=1..14400}$$

where: $P_{line.forecast}(t_i)$ – the power-flow forecast of a tie-line at time of t_i,, $P_{l.f.h.(tie-linej)}$ – the hourly forecasted power-flow value of a tie-line.

The algorithm runs in infinite cycle after the initialization. In the first step of the loop the gradient (ERR^{GRAD}) and integration error (ERR^{INT}) values are calculated for each tie-line.

$$ERR^{GRAD} = \Delta P_{line.flow}(t_k) - \Delta P_{line.forecast}(t_k) \geq THR^{GRAD} \tag{11}$$

where: t_0 – the time of the last correction, t_k – the investigated timeframe, THR^{INT} – integral threshold value for the tie-line, $\Delta P_{line.flow}$ – the actual gradient of power-flow value of a tie-line, $\Delta P_{line.forecast}$ – gradient of the forecasted power-flow value of the tie-line.

$$ERR^{INT}(t_k) = ABS\left(\sum_{t=j}^{k}\left[P_{line.flow}(t_i) - P_{line.forecast}(t_i)\right]\right) \geq THR^{INT} \qquad (12)$$

where: t_k – the investigated time (present), t_j – the time of the last correction, $\Delta P_{line.flow}$ – the gradient of the power-flow value of the tie-line, $\Delta P_{line.forecast}$ – the forecasted power-flow value of the tie-line.

Secondly the threshold values are compared to the beforehand computed error values. Due to the outcome of the comparison the loop starts again or the forecast correction procedure starts.

Data of the minute that are following the comparison are used to determine the offset value for each tie-line (Equation 13). If an error or errors occur, offset values are calculated for each tie-line independently. The forecasts are corrected with the computed offset values, as it is described by Equation 13 and 14.

$$CORR_{line.forecast} = \frac{1}{6}\sum_{i=0}^{5}\left[P_{line.flow}(t_{k+5-i}) - P_{line.forecast}(t_{k+5-i})\right] \qquad (13)$$

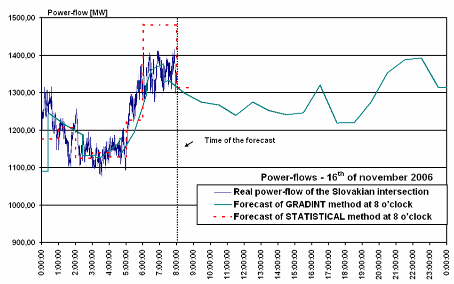
$$P_{line.forecast}^{new}(t_{k+5+j}) = P_{line.forecast}^{old}(t_{k+5+j}) + CORR_{line.forecast} \quad , \text{ j=1..(14400-k-5)} \qquad (14)$$

where: $CORR_{line.forecast}$ – correction offset value, t_k – the time when the correction method started, $P_{line.flow}$ – the actual power-flow value of the tie-line, $P_{line.forecast}$ – the forecasted power-flow value of the tie-line, $P_{line.forecast}$ – the former forecasted power-flow, $P_{line.forecast}$ – the corrected power-flow forecast.

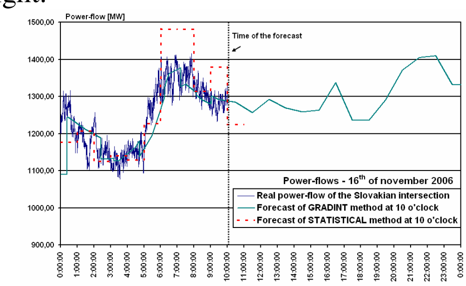
ALGORITHM AT WORK

In this chapter an example for the operation of the new algorithms is presented. The forecast for the Slovakian interface flow, which include two 400kV tie-lines: Győr – Gabcikovo and Göd – Levice, on 16th of November 2006 can be seen in Figure 4. The forecasts, which are made at 8 o'clock, are shown in Figure 4.a. The blue continuous line is the realized power-flow, the red broken line is the one-hour forecast that is made by the statistical method and the green continuous line is the summation of the GRADINT forecast that are made for the tie-lines of the interface.

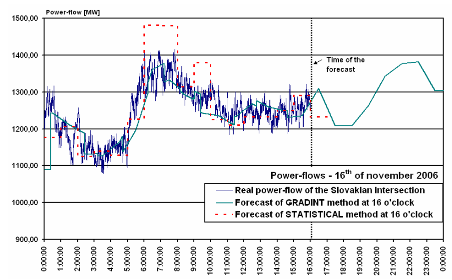
In Figure 4.b, 4.c the forecasts are made at 10 o'clock and 16 o'clock. In Figure 4.d the forecasted values can be seen for the whole day at midnight.



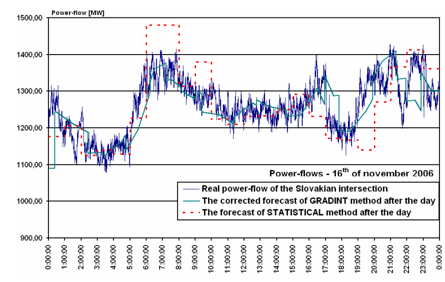
A



b



C



d

FIGURE 4

THE COMPARISON OF THE NEW ALGORITHMS

Study was executed in 2006, to get information about the reliability of the algorithms.

The study was made on the data of the 4th quarter of 2006. To compare the algorithms the one-hour predictions from 1 am to 12 pm were considered. The error is considered in every hour as the absolute value of the difference between the predicted and actual values of the intersections' power-flows. In this way there were 24 observations every day and that meant 24*92=2208 observations for each interface. (Just the short-term forecasting part of the statistical algorithm was available at the time of comparison.) The results of the study can be seen in Table II.

TABLE II

OVERALL ERROR STATISTICS FOR THE 4TH QUARTER OF 2006

Algorithm	Interface	Mean [MW]	Std. Deviation [MW]	Maximum [MW]
STATISTICAL	HU-SK	67,7	70,7	1081,8
GRADINT	HU-SK	55,5	62,5	984,3
STATISTICAL	HU-HR	55,9	55,8	598,3
GRADINT	HU-HR	40,9	44,2	984,8
STATISTICAL	HU-AUT	41,1	43,3	401,0
GRADINT	HU-AUT	40,5	36,4	304,3
STATISTICAL	HU-UA	44,7	38,7	642,0
GRADINT	HU-UA	33,2	37,0	464,2
STATISTICAL	HU-SRB	35,1	34,4	553,3
GRADINT	HU-SRB	36,1	37,4	564,6
STATISTICAL	HU-RO	32,2	33,5	243,3
GRADINT	HU-RO	24,1	29,6	225,7

There is no significant difference between the precision of the two new algorithms and they are tolerable precise if the system is close to its normal operating point.

Algorithms supplement each other well due to their different solve aspects of the problem. For long-term (24 hour) forecast the statistical algorithm should be used, because it has higher data insufficiency tolerance and smaller computational demand. The GRADINT should be used for short term (<4 hour) forecast, due to its fast response for switching events, which makes the algorithm able to handle emergency cases. The presently available statistical and PTDF based forecasts are going to be improved by this method. The organizational level adaptation and software implementation are under investigation.

BIBLIOGRAPHY

[1] UCTE Operational Handbook Policy 4 Co-ordinated planning, (March, 2006, pages 1-6) (http://www.ucte.org/_library/ohb/Policy%204_v2.0_03.05.2006.pdf)

[2] P. Gölöncsér, Z. Sulyok, Crossing Power Flow of the Hungarian Electric Power, (Elektortechnika, volume 97 number 12, December 2004, pages 336-339.) (In Hungarian).