The Effect of Length of Forecast Horizon on Rational Aggregation in Long-Term Forecasting of Energy Systems Development

Yu. D. Kononov*, D. Yu. Kononov, and S. V. Steklova

Melentiev Energy Systems Institute Siberian Branch of the Russian Academy of Sciences, Irkutsk, Russia

***Abstract* — The paper examines the rational aggregation of models that are employed to address energy sector forecasting challenges specific to various forecasting time frames. Possible approaches are proposed. The paper concludes with estimates of the potential impact of the magnitude and nature of input data uncertainty on forecast and aggregation errors.**

***Index Terms* — aggregation, energy sector, forecast error, forecasting time frame, Monte Carlo simulation, uncertainty.**

I. Introduction

The ongoing shift to the new social and technological order, the accelerating rate of scientific and technological advance, drastic changes that are expected to take place within the structure of energy production and consumption, changing requirements to environmental and energy security all contribute to the growth of uncertainty with regard to future conditions of the energy sector development. Obviously, the longer the forecasting time frame, the more uncertain are future conditions and the less reliable are forecasts.

Published forecasts of energy sector development in the USA and Europe that are made for 15 - 20 years ahead prove the non-linear nature of the escalation of the uncertainty range as the forecasting time frame extends (see Fig. 1). The minimum to maximum range of values of primary energy consumption volume in the USA for all cases and scenarios covered by the forecasts grows from the low 5-10% for the 5-year time frame to the high 13-23% and 22-38% for the forecasts made for 15 and 25 years ahead, respectively. The "Energy Strategy of Russia to 2030" (approved in 2009) claims that the difference between total energy consumption volumes under the worst and the best case scenarios amounts to 7% for the first 5 years and subsequently grows to 22% and 31% for the forecasts made for 15 and 20 years ahead, respectively. One of the lines of research undergoing active development and aiming at making long-term forecasts more evidence-based is the growing sophistication of research tools.

The state of the art of computer and information technologies makes it possible to build arbitrarily complex systems of models. However, under enormous and ever growing uncertainty of input data the following considerations are likely to challenge the practicality of making research tools ever more complex: 1) more granular treatment of data, increase in the number of entities subject to forecasting, and disaggregation of employed models all require additional information inputs,

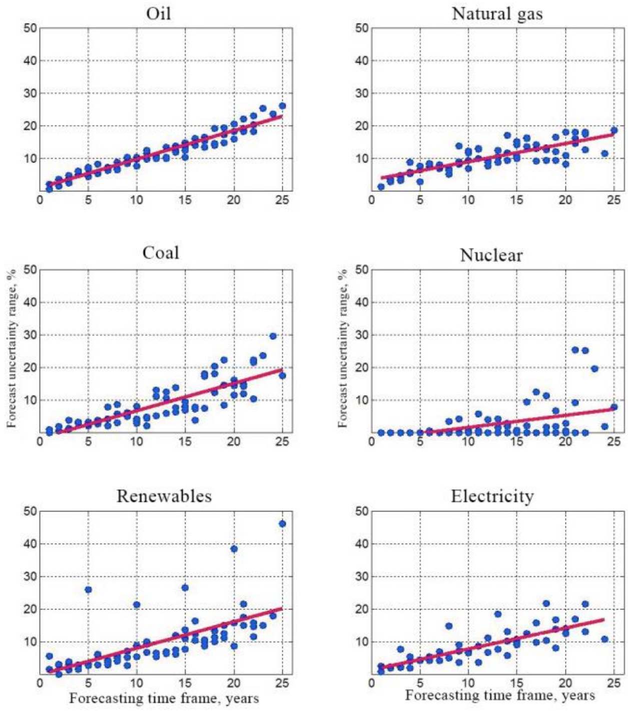

Figure 1. The uncertainty ranges of forecasts of the US energy consumption volumes as a function of forecasting time frames.

* Corresponding author.
E-mail: kononov@isem.irk.ru

http://dx.doi.org/10.25729/esr.2018.01.0006

Received February 07, 2018. Revised March 15, 2018. Accepted: April 11, 2018. Available online: April 25, 2018.

which increases the likelihood of a higher forecast error; 2) merging industry-level and regional systems and models within a unified system of models with fully automated calculations entails the problem of the optimality criterion that these models are to share; 3) the more complex such systems of models grow the more difficult it is to track and interpret unforeseen results; and, finally, 4) a high opinion of the complexity of the tools oftentimes results in excessive and unjustified confidence in long-term forecasts.

The above and other deficiencies are all the more poignantly manifest themselves when the very same models that share the identical level of aggregation are applied to forecasting of the energy sector development for both medium-term (up to 10-15 years) and long-term forecasts.

A key principle that guides systems studies and the process of improvement of research tools is the trade-off between the required precision of calculation results and the precision of the information used to generate them [2]. The principle is analogous to the well-known Occam's razor principle and assumes building models that are as simple as possible yet capable of accounting for defining properties of the studied system that are required to appropriately tackle the task under given conditions. This echoes the following quote attributed to Albert Einstein as well: "Everything should be made as simple as possible, but not simpler" [3].

Striving for utterly comprehensive while mathematically tractable treatment of development dynamics and non-linear relationships within the studied system as well as the detailed representation of its structure can go against the grain of the inherent uncertainty of input economic data and the mutable nature of properties of complex systems that are being modeled, which can even entail negative outcomes.

The principle of correspondence between research tools and actual uncertainty of the input data fed into them as well as the required degree of forecast accuracy has so far been implemented based on intuitions held by model developers and model users and remains more of an art than a science. A more evidence-based approach to the implementation of the principle can be developed by means of the quantitative analysis and juxtaposition of the actual uncertainty of input data and the import of calculation results obtained thereby in order to identify possible issues and make decisions more substantiated. It is obvious that the value of forecasts and requirements for their validity depend on the forecasting time frame and the actual problem being solved.

The time-honored tradition is to treat the aggregation problem as a problem of reducing the dimensionality of a model so that the losses of information generated by the model are kept to a bare minimum. It appears that given large uncertainty of input data and large dimensions and complexity of forecasting models it is reasonable to raise the problem of the rational aggregation of such models. To this end, it is necessary to account not only for the magnitude and nature of the input data uncertainty, but also for the possible and maximum acceptable error of key variables to be forecast. There are no such versatile methods that can be applied to solving the problem for an arbitrary system and an arbitrary time frame.

This paper covers possible approaches to rational aggregation of optimization models employed for long-term forecasting of the national energy sector development and regional energy supply systems. These approaches include the following: an evaluation of the input data accuracy (its uncertainty range), a study of the effect that various aggregation levels of these data and models have on the results generated by multi-variant calculations, an identification of an acceptable accuracy level for the variables to be forecast.

II. Aggregation In Long-Term Forecasts Of The National Energy Sector Development

To illustrate the above, it will suffice to refer to the beginning of the widespread use of optimization models back in the middle of the 20th century with the emergence of more elaborate computerized systems of models built on top of them in the decades that followed (see e.g. [3,4]). In a number of cases, their composition and the level of aggregation remain independent of the length of the forecast horizon.

The principle of correspondence between research tools employed and the uncertainty of input data is fulfilled by a multi-stage approach to narrowing down the uncertainty range of conditions and results of forecasting studies [5,6]. The approach implies the multi-stage narrowing down of the length of the forecast horizon, iterative calculations generated by models of various hierarchical levels used to handle specific forecasting time frames, and the reconciliation of totals in time. In doing so, the initial stage covers the time range of over 15-20 years and the minimum number of levels and models (see Fig. 2). It is worth noting that most of the models of energy-related industries development and those of regional energy supply systems are optimization models.

Iterative calculations (carried out in top-down and bottom-up fashions) make it possible to account for features specific to the development of systems of various hierarchical levels that make up the integral national energy system. Within each of the time frames, it is the problems that are deemed most significant that are to be solved. To this end, there are various possible aggregation levels for the energy facilities, energy links, and geographical areas that are subject to being modeled.

When using multi-level systems of models, one can employ well-known methods of iterative aggregation [7,8,9].

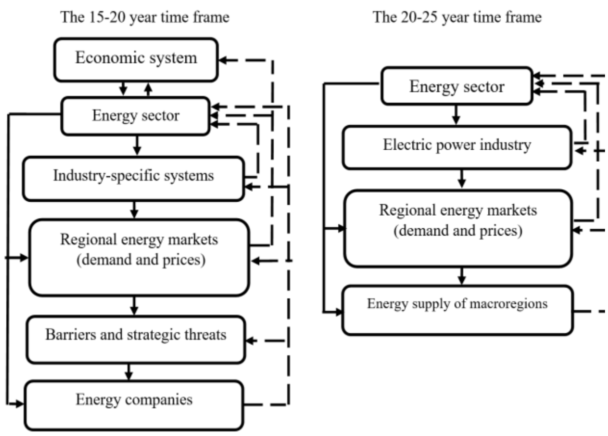

Figure 2. Interactions between hierarchical levels, problems, and models specific to various time frames of long-term energy sector forecasting studies.

In Russia, methods of iterative information aggregation in hierarchically built systems of models underwent active development in the 1970s and 1980s. Back then they were applied to the coordination of decisions generated by industry-level and regional model hierarchies of energy systems that account for both producer and consumer behavior patterns [12,13]. Such methods assume aggregation and disaggregation of all interrelated models at each iteration step. In so doing, the end of calculations is marked by achieving an acceptable level of aggregation. The latter is defined as the optimality criterion for the upper level model taking the same value for two successive iterations.

The limitation inherent in such an approach is that multi-level model systems that are designed to forecast the national energy sector development, have different top-level models for medium- and long-term forecasts (see Fig. 2) with each of the hierarchical levels applying their own optimization criteria. Furthermore, under major uncertainty, there is no need to strive for the perfect match of results of iterative calculations.

It may be feasible to favor separate aggregation of models for each hierarchical level over their combined aggregation. To this end, it is possible to implement the following calculation steps:

1. Building a basic (reference) model that is as comprehensive as possible in its most detailed consideration of energy system facilities, energy links, and system properties.
2. Using multi-variant calculations to identify the input data that have the most decisive impact on key variables to be forecast.
3. Estimating the input data uncertainty range. The estimates are possible to attain if assisted by an analysis of available forecasts of their assumed behavior. Such analysis when backed by expert judgment can also provide insights into the nature of the data uncertainty (a probability distribution of values of a given variable within a certain uncertainty range).
4. Finding an approximate value of the minimum possible calculation error for the reference model calculations under given uncertainty of input data. This value can serve as a reference point for the minimum forecast error, which is always larger than zero.
5. Benchmarking calculation results under a varying level of model aggregation against reference model calculation results, and identifying corresponding aggregation errors.
6. Based on the results of a comparative analysis of such errors against the allowed (acceptable) error one arrives at a rational level of model aggregation.

III. Aggregation Of Geographical Areas In Long-Term Forecasts Of The Electric Power Industry Development

Models and methods used to substantiate development strategies of the Russian electric power industry are covered in sufficient detail in [14]. When optimizing the power generation mix 10 to 15 years in advance, one accounts for power plant operating conditions and cross-system flows of generation capacity and power. This calls for highly detailed optimization models. A case in point is the well-known SOYUZ model [15] that treats the national territory of Russia as divided into regional energy systems. A system of models used for long-term forecasting of the national energy sector, the rational electric power industry development can be represented in a less aggregated way.

This class of models includes MISS, a stochastic statistics-based simulation model and its software implementation [16]. The model is developed to tentatively assess the competitiveness of available types of power plants and options to fulfill the energy demand of macroregions under ambiguous information on expected conditions.

The optimality criterion used in this model is the minimum cost of power generation (production) in a given region under the following constraints: the demand for electric power in a given area, its export or import potential, the capacity of already operating plants and the potential for generation capacity additions for various types of power plants, and constraints on gas production and supply in the area. All of the above constraints are specified as ranges of possible values. The upper and lower boundaries of possible values are also provided for fuel prices, capital intensity, and technical and economic indicators that influence the cost of electricity.

Variables of interest of the model are as follows: the capacity of new power plants, the amount of electricity they generate, consumption volumes for various types of fuels, producer prices at each of the plants as well as the weighted average and marginal electricity generation price in a given region.

To account for information uncertainty, one has to generate and study multiple optimal solutions (hundreds thereof) under various combinations of input data. This implies the use of well-established repeated random sampling Monte Carlo methods (experiments) by the MISS model.

When possible combinations of input data values (all treated as interval estimates) are generated, the numeric parameters that define the type of probability distribution of values within the ranges are varied. This allows generating random variables of the most diverse types of probability distributions ranging from uniform to normal, to lognormal, to exponential, etc.

The MISS model was used to estimate the effect of aggregation on the mix of generation capacity additions and the generation cost in European Russia and Siberia [17].

For basic models, each macroregion was made up of six regional energy systems so as to account for energy and fuel supply conditions and cross-system energy links specific to them. During the aggregation process, the six regions were consolidated into a single macroregion. In doing so, instead of specific uncertain ranges of fuel prices we used generalized ones as well as overall constraints on the maximum possible additions of gas-fired, coal-fired, nuclear power, and hydroelectric plants.

It follows from the calculations (see Table 1) that model aggregation leads to the increase in the error of the electricity cost calculations by a mere 1-2%, while the uncertainty of the electricity cost forecast itself is approximately twice as low as the fuel cost forecast error for balancing power plants in a given region. The effect of aggregation on the mix of electricity generation capacity additions is much more pronounced. The corresponding share of combined cycle gas turbine (CCGT) plants grows by 1.2-1.5 times.

The aggregation error increases when a fixed range of probable values is fed into the MISS model which are assumed to be best modeled not as a normal distribution but as interval (uniform) uncertainty. The error also increases as the input data uncertainty range extends, which is inevitable when the forecasting time frame grows larger (see Table 2).

In the long run, it is power plants of emerging types that will play an increasingly important role. That is why as the forecasting time frame increases, the impact of model aggregation on the mix of electric power generation sources will remain more significant than that on the projections of electricity prices.

The calculations carried out using the MISS model demonstrate that under assumed conditions the level of investment risk associated with options of electric power supply in Siberia is higher than that in European Russia. The risk level of investing in a given plant was defined as an inverse value of the frequency (probability) of its

Table 1. Deviation of aggregated model calculation results relative to reference values.

Input data specification	Electricity price		Share of CCGT plants	
	European Russia	Siberia	European Russia	Siberia
Average values	1	0.5	6	9
Normal distribution	2.1	0.1	12	10
Interval uncertainty	2.4	0.2	18	12

Note: Calculation results are for the assumed 2020-2025 conditions. The deviation is presented as percentage, while structural changes (the share of CCGT plants in the total added generation capacity of power plants) are in percentage points (pp).

inclusion in optimal solutions generated by multi-variant calculations.

It is reasonable then to assume that energy supply forecasting studies for the regions of higher investment risks should be carried out with reliance on more granulated (less aggregated) models.

IV. Approaches To Judging Acceptable Accuracy Of Forecasts

Multi-variant calculations by way of optimization and stochastic models enable plotting the curve of changes in the model's objective function values and key variables to be forecast as a function of the aggregation level. It is more challenging to identify an acceptable level of the forecast error. As of now, there are no versatile methods that would provide such an assessment. Therefore, in practice one has to trust one's intuition backed up by the knowledge of task-specific factors and one's accumulated experience.

One of the major objectives of long-term forecasts of

Table 2. The effect of the increase in the uncertainty range and the average gas price on calculation results generated by reference and disaggregated models.

	Units	Gas price increase, %		
		5	10	25
Average electricity price				
before aggregation	%	2	4	12
after aggregation	%	3	5.5	14
Decrease in the CCGT plants share				
before aggregation	pp	14	24	28
after aggregation	pp	13	15	17

Note: Calculated for European Russia.

energy systems development is to provide government agencies on a par with companies with takeaways to be used for making timely investment decisions. To this end, it is crucial to estimate risk and returns of large-scale projects of electricity generation capacity additions. Such estimates have to be based on forecasts of plausible energy price and demand behavior.

The investor values more remote rewards less than more immediate ones. By varying the values of key input data variables for any of the time periods and by estimating the effect of such variations on the project value, it is possible to arrive at conclusions bearing on the acceptable decrease in the forecast precision for more remote time periods within the forecasting time frame.

Such an approach was applied, in particular, to assess the sensitivity of investment projects returns calculated as the Net Present Value (NPV) a) for investment projects of nuclear power plants construction to the changes in demand (production) volume that occur over time and b) for projects of Combined Cycle Gas Turbines (CCGT) construction to the changes in gas prices.

The calculation results (see Fig. 2) show a notable non-linear decrease in sensitivity of the project value to changes in gas prices in more remote time periods. Under assumed input data the surge in demand for electricity by as high as 20% exerts significant effect on the NPV only within the time frame limited by the first 15 years. Accordingly, forecast performance requirements can safely be relaxed for electricity demand forecasts at the end of the nuclear power plant life cycle.

As the length of the forecast horizon extends, requirements for fuel price forecast performance notably relax as well in the case of the CCGT construction project valuation. Even under the scenario of a 1.5-2-time increase in the gas price at the end of the forecasting time frame, the decrease in the project's net present value does not exceed 2-3% (see Fig. 3).

In the case that the results of forecasting studies are used to inform investment decision making, the risk value

Table 3. Correspondence between the uncertainty of input data and forecast performance

Variable	Units	Generation cost: CCGT plants	Generation cost: System's average	Generation cost: Marginal
Uncertainty range	cents per kilowatt hour	6.7-7.4	7.0-7.6	7.8-8.1
	%	10.4	8.6	3.8
Correspondence between inaccuracy of data for electricity and gas prices	% / %	0.32	0.26	0.12

Note. Calculations for the energy systems of Siberia with gas prices assumed to fall within the \$100-\$133 / sm^3 range.

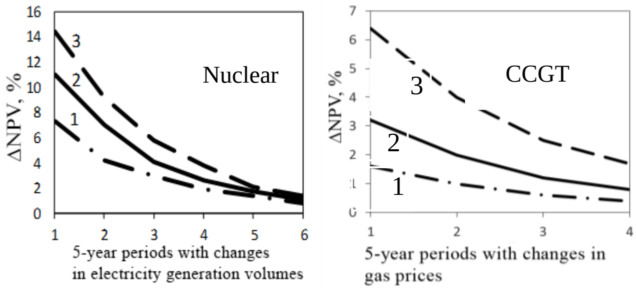

Figure 3 Changes in the project's NPV as a function of changes in the electricity generation volume (an NPP construction project) or gas prices (a CCGT plant construction project) in one of the five-year periods within the forecasting time frame. Increase by 25% - Curve 1, increase by 50% - Curve 2, increase by 100% - Curve 3.

assumed for the project valuation can serve as a plausible reference value, alongside the error inherent in projections of the key variable (for example, electricity prices).

The error for a given time frame can be identified by a sensitivity analysis of forecast variables as they respond to changes in input data within a predefined range of their possible values.

To illustrate this point, Table 3 lists the results of the analysis of the effect that the gas price has on the power plant electricity generation cost. Calculations were carried out using the MISS model as applied to one of the scenarios of Siberian electric power industry development within the 2020 to 2025 time period.

The calculations indicate that under assumed conditions each 1% decrease in the accuracy of a gas price forecast leads to an increase by approximately 0.26% in the minimum forecast error with respect to the average electricity cost.

V. Conclusion

Rational aggregation of models employed in practical forecasting work entails assessing and accounting for the effect of uncertainty of the input data on the probable error in key variables to be forecast. It is also essential to understand what magnitude of the forecast error can be safely deemed acceptable when making timely decisions (be they investment, managerial, or strategic ones).

Obviously, the priority and complexity of efforts to accommodate these factors are determined by the forecasting time frame and the particulars of the problem. The wider the range of the input data uncertainty (that is known to grow with longer time frames), the greater the unavoidable forecast error, which hence makes the use of more aggregated models all the more justified.

The approaches proposed herein to identify rational aggregation of energy facilities and geographical areas in various stages of forecasting studies include: an evaluation of input data accuracy (its uncertainty range) as changing over time, a study of the effect that various aggregation levels of these data and models have on the results yielded by multi-variant calculations, an identification of acceptable accuracy of the variables to be forecast.

As of now, there are no universally applicable methods that would facilitate such an assessment. In the case that the results of forecasting studies are used to inform investment decision making, the risk value assumed for large-scale project valuations can serve as a plausible reference value, alongside the error inherent in projections of key variables. The latter is dependent on the input data uncertainty range and increases as the length of the forecast horizon extends into the future.

Acknowledgement

This research was supported by a RFBR grant (Project No. 16-06-00091).

References

[1] U.S. Energy Information Administration. Annual Energy Outlook (1995-2014). [Online]. Available: http://www.eia.gov/forecasts/aeo/

[2] L.A. Melentiev. *Energy Systems Analysis*. Moscow: Nauka, 1979.

[3] U.S. Energy Information Administration. The National Energy Modeling System: An Overview [Online]. Available: http://www.eia.gov/oiaf/aeo/overview/

[4] The SCANER modeling and information system. Ed. by A.A. Makarov. Moscow: ERI RAS, 2011.

[5] Yu.D. Kononov, "A multi-stage approach to improve the validity of long-term forecasts of the energy sector development and strategic threat assessment", *Proceedings of the Russian Academy of Sciences. Power Engineering*, No. 14, pp. 61-70, 2014.

[6] Yu.D. Kononov. *Approaches to improve the validity of long-term forecasts of the energy sector development*. Novosibirsk: Nauka, 2015.

[7] D.F. Rogers, R.D. Plante, R.T. Wong, J.R. Evans, "Aggregation and Disaggregation Techniques and Methodology in Optimization", *Operations Research*, Vol. 39, No. 4, pp. 553-582, 1991.

[8] W.D. Fisher, "Criteria for Aggregation in Input-Output Analysis", *The Review of Economics and Statistics*, Vol. 40, No. 3, pp. 250-260, Aug. 1958.

[9] P.H. Zipkin, "Bounds for Row-Aggregation in Linear Programming", *Operations Research*, Vol. 28, No. 4, pp. 903-916, August 1980.

[10] *Iterative aggregation and its application to planning*. Ed. by L.M. Dudkin. Moscow: Ekonomika, 1979.

[11] V.I. Tsurkov, "Aggregation in problems of industry-level planning", *Economics and Mathematical Methods*, Vol. XVI, No. 3, pp. 533-544, 1983. (in Russian)

[12] *Methods and models for reconciliation of hierarchical decisions*. Ed. by A.A. Makarov. Novosibirsk: Nauka, 1979.

[13] A.A. Makarov, A.S. Makarova, B.G. Saneev, "Investigation of the mechanism of interaction of hierarchically organized economic systems", *Economics and Mathematical Models*, Vol. XI, No. 5, 1975. (in Russian)

[14] *Justification of electric power energy systems development*. Ed. By N.I. Voropai. Novosibirsk: Nauka, 2015.

[15] N.I. Voropai, V.V. Trufanov, "Mathematical modeling of energy systems development in the modern world", *Electricity*, No. 10, pp. 6-12, 2000.

[16] Yu.D. Kononov, V.N. Tyrtyshny, "Asssessment of the effect of the nature of uncertainty of future conditions on competitiveness of fuel supply options", *Energeticheskaya Politika*, No. 4, pp. 50-56, 2011. (in Russian)

[17] D.Yu. Kononov, Yu.D. Kononov, "Rational aggregation of geographical areas for long-term forecasting of electricity prices", *Studies on Russian Economic Development*, Vol. 27, No. 6, pp. 649-655, 2016.

Yuri D. Kononov. - Doctor of Economics, Professor, Honored Worker of science of the Russian Federation, chief researcher of Melentiev Energy Systems Institute of Siberian Branch of the Russian Academy of Sciences, Russia, Irkutsk. His research interests are modeling and research of interaction between energy and economy; methods to enhance the validity of long-term forecasts of energy development.

Dmitry Yu. Kononov. - Ph. D., Senior researcher of Melentiev Energy Systems Institute of Siberian Branch of the Russian Academy of Sciences, Russia, Irkutsk. His research interests are consideration of the impact of the uncertainty factors on the methods for the evaluation of strategic threats to power industry development.

Steklova Svetlana V. - Senior engineer of Melentiev Energy Systems Institute of Siberian Branch of the Russian Academy of Sciences, Russia, Irkutsk. Her research interests are an analysis of reliability of national and international energy development forecasts.