Plausible Model Improvement Utilizing the Information Obtained from Data Assimilation

Masayuki YOKOYAMA[1,2,3], Yuya MORISHITA[4] and Sadayoshi MURAKAMI[4]

[1]*National Institute for Fusion Science/Rokkasho Research Center, National Institutes of Natural Sciences, Rokkasho, Aomori 039-3212, Japan*

[2]*The Graduate University for Advanced Studies, SOKENDAI, Kanagawa 240-0115, Japan*

[3]*The Institute of Statistical Mathematics, Research Organization of Information and Systems, Tachikawa 190-8562, Japan (visiting)*

[4]*Department of Nuclear Engineering, Kyoto University, Kyoto 615-8540, Japan*

(Received 31 October 2023 / Accepted 13 December 2023)

Data assimilation technique implemented in fusion research has enhanced the modeling capability. The quantitative "gap" between the original model (typically based on physics considerations and/or empirical approach) and the optimized model (obtained through data assimilation) can be utilized to improve the original model to align with the measured data. Such a procedure is proposed here by taking the model of the heat diffusivity of plasmas as an example. It successfully elucidates relevant parameters recognized in the experiment but were missing in the original model, demonstrating the efficiency of the proposed procedure.

Keywords: model improvement, data assimilation, multivariate regression, information criterion

DOI: 10.1585/pfr.19.1203006

The data assimilation technique has been successfully implemented into fusion research [1] and progressed further to control the actual plasma [2].

During the data assimilation process, an original model is optimized (or modified) so that the observed data could be well reproduced with taken model uncertainties into account. Here, model uncertainties refer to elements not well considered in an original model. This generates a quantitative "gap" between original and optimized models.

Then, it comes to an idea that a model can be improved by describing this "gap" using relevant parameters and then superposing it on an initial model. In this paper, the heat transport problem in fusion plasmas is exemplified to concretely describe the above idea. All the data in this paper are obtained from the data assimilation process of Large Helical Device (LHD) [3] plasmas.

This paper is organized as follows. Firstly, it is described how the "gap" is obtained through the data assimilation. Then, an example approach for describing the "gap" using plausible physics parameters is explained, followed by summary and outlook.

In a previous study on the heat transport of LHD plasmas [4], the model for the ion heat diffusivity, χ_i, was deduced from a limited number of cases (~O(10)) as (so-called "Gyro-Bohm grad-Ti" model)

$$\chi_\mathrm{i} = C_\mathrm{i}\frac{T_\mathrm{i}}{eB}\frac{\rho_\mathrm{i}}{a}\left(a\frac{\nabla T_\mathrm{i}}{T_\mathrm{i}}\right),$$

where T_i, ρ_i, a, e, and B are the ion temperature, ion Larmor radius, plasma minor radius, electric charge, and magnetic field strength, respectively, and ∇ is a radial derivative. The radially constant factor, C_i, was estimated by best fitting the measured T_i profiles of the relevant cases. However, it was found that this model cannot reproduce the actual temporal change of T_i [5].

On the other hand, C_i is a state variable in the data assimilation and can be modified radially and temporally by fully utilizing the time series data of the measured T_i. This was done in Ref. [5] for twelve discharges, which accumulate radial profiles of C_i, as shown in Fig. 9 of Ref. [5]. It is an overplot of 275 C_i profiles based on the data assimilation cycle (40 ms). Implementing this information of C_i in an original model and setting them as χ_i in the integrated transport simulation, TASK3D [4], reasonably reproduced the temporal behavior of T_i for the twelve discharges employed [5].

It is found that "C_i"s have a common feature. Thus, it is worthwhile describing such a common feature with plausible plasma parameters. This could facilitate recognition of new important parameters and/or physics elements that were missing in the initial model.

One idea is to perform a multivariate regression analysis. Preparing dimensionless variables as explanatory variables is natural because C_i is dimensionless. The variables prepared *a priori* are $\nu_\mathrm{i}{}^*$, $\rho_\mathrm{i}{}^*$, $T_\mathrm{e}/T_\mathrm{i}$, $-(dn_\mathrm{e}/\mathrm{d}\rho)/n_\mathrm{e}$ and $-(dT_\mathrm{i}/\mathrm{d}\rho)/T_\mathrm{i}$, being the ion collision frequency, ion Larmor radius, temperature ratio, density gradient, and ion temperature gradient, respectively, all normalized by usual

author's e-mail: yokoyama@nifs.ac.jp

convention.

The statistical importance of each parameter in all the possible models is determined by evaluating the Akaike's Information Criterion (AIC) [6] (more precisely, AICc with a correction for small sample sizes) [7]. The model providing the minimum (or least) AIC value is considered a "statistically good (or reasonable)" model. Notably, in practical cases, AIC may continue decreasing with the increasing the number of explanatory variables. In such a case, the decreasing rate of AIC for the increasing number of variables (NV) can be a measure.

It is found that the exponents for $\nu_i{}^*$ and $\rho_i{}^*$, deduced from a log-linear multivariate regression, are unreasonable values (say, O(−10 or −100)), possibly due to the rather limited variations of these two variables for the data-assimilated twelve discharges. Thus, these two variables were omitted from the regression analysis, and the remaining three variables were employed below.

Figure 1 shows the results of the exhaustive search for all possible models. The bars for each NV indicate that AICc varies for different combinations of given NV variables (marked by circles). The minimum value of AICc's decreases up to NV = 2 and remains almost unchanged at NV = 3. Table 1 summarizes the appearance of variables and corresponding exponents up to NV = 3. First, $-(dn_e/d\rho)/n_e$ appears, then T_e/T_i appears at NV = 2. Lastly, $-(dT_i/d\rho)/T_i$ appears at NV = 3, with an exponent close to zero. The statistically reasonable expression for C_i, based on Fig. 1 and Table 1, could be as follows:

$$C_i \propto \left(\frac{T_e}{T_i}\right)^{0.38} \left(-\frac{dn_e}{d\rho}/n_e\right)^{0.13}.$$

Interestingly, the positive power dependence of χ_i on T_e/T_i has been recognized experimentally [8]. Thus, the obtained expression for C_i improves the initial model.

Data assimilation can provide a quantitative "gap" between the original and the optimized models. Such a "gap" was expressed with plausible plasma parameters for model improvement, enhancing the physics insights/findings beyond the data assimilation. This trial successfully revealed the parameter dependence of χ_i, which was previously experimentally recognized but not included in the original model. The statistical induction of χ_i was also reported [9], in which a regression model was directly obtained without data assimilation. The advantage of using the "gap" derived from data assimilation is the existence of a base model (original model). Nevertheless, as an outlook, it would be interesting to compare these approaches to consider the effectiveness of statistical thinking in modeling fusion plasmas and their relevance in reproducing measured data.

The authors acknowledge the LHD Experiment Group for sharing LHD experiment data. This work was supported by the JSPS KAKENHI Grant Numbers JP19K03797, JP21J14260, and JP22K03582, NIFS Collaborative Research Program (NIFS22KIPR009 and NIFS22KIPT008), QST Research Collaboration for Fusion DEMO, and "Strategic Research Projects" grant from the Research Organization of Information and Systems.

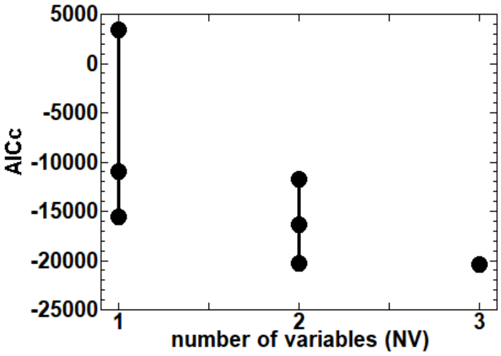

Fig. 1 Evolution of AICc values for multivariate log-linear regression for Ci as a function of number of variables (NV). The circles correspond to the AICc values.

Table 1 Appearance of variables and those exponents as a function of NV (at the lowest AICc for each NV).

NV	T_e/T_i	$-(dT_i/d\rho)/T_i$	$-(dn_e/d\rho)/n_e$
1			0.147
2	0.383		0.133
3	0.406	-0.0126	0.143

[1] Y. Morishita *et al.*, Nucl. Fusion **60**, 056001 (2020).
[2] Y. Morishita *et al.*, J. Comp. Science **72**, 102079 (2023).
[3] M. Osakabe *et al.*, Nucl. Fusion **62**, 042019 (2022).
[4] S. Murakami *et al.*, Plasma Phys. Control. Fusion **57**, 054009 (2015).
[5] Y. Morishita *et al.*, Comp. Phys. Comm. **274**, 108287 (2022).
[6] H. Akaike, "Information theory and an extension of the maximum likelihood principle", 1973 Proceedings of the 2nd International Symposium on Information Theory, Petrov, B.N., and Caski, F. (eds.), Akadimiai Kiado, Budapest 267.
[7] N. Sugiura, Communications in Statistics - Theory and Methods **7**, 13 (1978).
[8] H. Takahashi *et al.*, Nucl. Fusion **57**, 086029 (2017).
[9] M. Yokoyama and H. Yamaguchi, Nucl. Fusion **60**, 106024 (2020).