

Research Article

Alternative Definitions of Complexity for Practical Applications of Model Selection Criteria

Murari Andrea ¹, Riccardo Rossi ² and Teddy Craciunescu ³

¹Consorzio RFX, CNR, ENEA, INFN, Università di Padova, Acciaierie Venete SpA, Corso Stati Uniti 4, Padova 35127, Italy

²Department of Industrial Engineering, University of Rome Tor Vergata, Via del Politecnico 1, Rome 00133, Italy

³National Institute for Laser, Plasma and Radiation Physics, Măgurele 077126, Romania

Correspondence should be addressed to Teddy Craciunescu; teddy.craciunescu@gmail.com

Received 7 September 2020; Revised 8 June 2021; Accepted 22 October 2021; Published 2 November 2021

Academic Editor: Andrea Roli

Copyright © 2021 Murari Andrea et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Defining and quantifying complexity is one of the major challenges of modern science and contemporary societies. This task is particularly critical for model selection, which is aimed at properly identifying the most adequate equations to interpret the available data. The traditional solution of equating the complexity of the models to the number of their parameters is clearly unsatisfactory. Three alternative approaches are proposed in this work. The first one estimates the flexibility of the proposed models to quantify their potential to overfit. The second interprets complexity as lack of stability and is implemented by computing the variations in the predictions due to uncertainties in their parameters. The third alternative is focused on assessing the consistency of extrapolation of the candidate models. All the upgrades are easy to implement and typically outperform the traditional versions of model selection criteria and constitute a good set of alternatives to be deployed, depending on the priorities of the investigators and the characteristics of the application.

1. The Evaluation of Complexity for the Purpose of Data Analysis

The science of complexity continues to make significant progress in clarifying many phenomena in science and engineering [1]. As a field of study, complexity theory is aimed at understanding the behaviour of systems difficult to analyse and predict. The body of knowledge produced is impressive and general interpretative paradigms exist, such as the heuristic that complexity is a middle ground between randomness and rigid order [2]. On the other hand, detailed analysis of individual systems and phenomena typically requires specific tools and techniques. The objective of developing reliable models for these systems is therefore particularly important.

Given the nature of complex systems, very often multiple models are good candidates for the interpretation of the phenomena under study and it is not necessarily simple to discriminate between them and identify the most appropriate. In this context, Model Selection Criteria (MSC)

would be expected to play an important role and be systematically used [3]. They consist of a series of indicators aimed at determining to what extent a mathematical model is supported by the available data. Unfortunately, even the more advanced model selection criteria are difficult to deploy in practice, because typically several of their underlying assumptions are violated.

Among the weaknesses of the traditional formulation of MSC, the quantification of complexity is certainly a major issue. Sound mathematical theories for estimating the complexity of mathematical functions do exist. Among the most developed are certainly the Vapnik-Chervonenkis [4] and the Rademacher [5] criteria. On the other hand, these quantifiers are very difficult to calculate for the vast majority of mathematical equations encountered in practice. They are also based on an interpretation of complexity, which is not necessarily the most useful in the context of model selection (see the next section).

The typical solution of assuming that the complexity of an equation is determined by the number of its parameters is

well known to be unsatisfactory. Particularly when it comes to overfitting, as will be shown in the next sections, models with the same number of parameters can have a completely different behaviour. Moreover, the definition of complexity itself is not unique in this context. Depending on the application, different types of model, clearly of different level of complexity, can prove to be more or less adequate. Therefore, in this work, three different quantifiers of model complexity are proposed. The first one is explicitly conceived to penalise the candidate models for their potential to overfit. The second favours equations, which are less sensitive to uncertainties in their parameters. The third is more devoted to guaranteeing a smooth extrapolation of the models out of sample.

With regard to the structure of the paper, the next section is an overview of the way traditional MSC deal with the issue of quantifying model complexity. Section 3 introduces the main rationale behind the proposed alternative views of complexity and how they can be inserted in the MSC criteria, with the help of some didactic examples. The list of function families and noise statistics investigated is provided in Section 4. The results of systematically applying the new versions of the MSC to these functions are also covered in Section 4. The conclusions are drawn in the last section of the paper, with a discussion about possible further developments.

2. How Information Theoretic and Bayesian Model Selection Criteria Handle Complexity

In the science of complex systems, measurements are the basic inputs required to provide quantitative knowledge about systems. However, all the measurements provide limited information about the phenomena to be investigated, since they are affected by uncertainties. Such uncertainties, also referred to as measurement errors, present quite a challenge for model selection. Indeed, the main aim of identifying equations to describe phenomena resides in the possibility to use them to predict situations not already encountered in experiments. One qualifying aspect of mathematical models is therefore their generalisation capability. Errors in the data are a problem in this perspective because models reproducing too well the available data can be too influenced by the noise and therefore can generalise poorly. In the literature, this issue is typically indicated with the term “overfitting.”

Information theoretic and Bayesian model selection criteria address this aspect by penalising the complexity of the models. The main argument behind this approach is that more complex mathematical models typically are more flexible and therefore are inherently more prone to overfitting the input examples. This conceptual framework is problematic because it is not obvious why simpler equations should be more adequate to model the behaviour of complex systems. More importantly, the implementation of this approach is flawed in practice, as will be discussed in detail later in this section.

The most widely used MSC belong to the family of information theoretic criteria and Bayesian information

criteria. The main representative of the first family is the Akaike Information Criterion (AIC), which is meant to minimise the extrapolation errors [6]. The Bayesian Information Criterion (BIC) exemplifies the second class and is conceived to select the most likely model given the data [7]. Their derivations and properties are described in detail in [3].

Under the traditional assumption that the data are identically distributed and independently sampled from a normal distribution, it can be demonstrated that the AIC can be written (up to an additive constant, which depends only on the number of entries in the database and not on the model) as

$$\text{AIC} = n \cdot \ln(\text{MSE}) + 2k, \quad (1)$$

where MSE is the mean-squared error of the residuals, n is the number of entries in the database, and k is the number of parameters in the model. Similar assumptions allow expressing the BIC criterion as

$$\text{BIC} = n \cdot \ln(\sigma_{(\epsilon)}^2) + k \cdot \ln(n), \quad (2)$$

where $\sigma_{(\epsilon)}^2$ is the variance of the residuals and again n is the number of entries in the database and k is the number of parameters in the model.

Both criteria are cost functions to be minimised, in the sense that the better the model the lower their value. This can be intuitively understood simply by inspection of their structure. The first term favours models that are closer to the data. The second addend is the penalty term for complexity. On the other hand, assuming that more complex functions have a higher number of parameters is a very poor approximation. Two classical examples, showing the limitations of this approach, are discussed in the following. Let us suppose that the model generating the data is a 5th-degree polynomial depending on 7 parameters:

$$y = 1 + 2x + 0.01x^2 + 3e^{-5}x^5. \quad (3)$$

A sinusoidal function at high frequency, a function depending on only three parameters (amplitude, frequency, and phase) can fit the data equally well (see Figure 1). Moreover, in case of added noise, the frequency of the sinusoidal function can be increased to the point that it can fit the data even better than the original function. Therefore, the traditional versions of the AIC and BIC would select the sinusoid as the best model, since both terms in equations (1) and (2) would be smaller than for the actual model generating the data.

Another case, emphasising the difficulty of quantifying the complexity of a mathematical function, is the comparison of an exponential and a polynomial function. In principle, when comparing the complexity of two classes of mathematical functions, one should use the same representation of the functions. In this example, the exponential would have infinite complexity, once expressed as a series. On the other hand, it is much less prone to overfitting than a high-order polynomial, even if formally depending on a much higher number of parameters. These intuitive

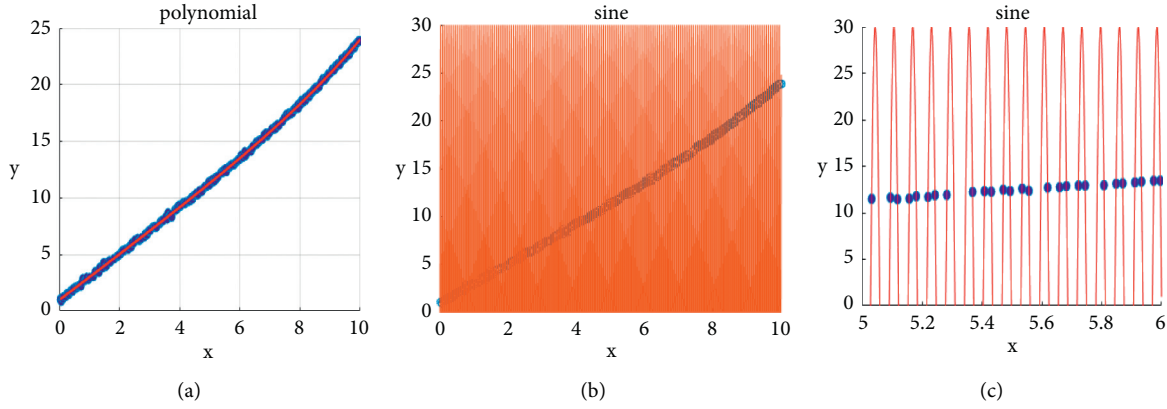


FIGURE 1: (a) Data generated by a fifth-order polynomial. (b) Fit with a high frequency sinusoidal function. (c) Zoom of the plot in the centre to show the quality of the fit of the sinusoid.

considerations have been confirmed with a series of numerical tests.

3. Alternative Definitions of Complexity

The examples, presented in the last section, show clearly that quantifying the complexity of a function with the number of parameters is clearly not adequate for model selection. An appropriate indicator for this application would have to satisfy various requirements. It should be easy to compute, if possible, independently from the number of database entries. It should also properly quantify the tendency of the models to overfit the data. A good MSC should also be robust against small errors in the determination of the parameters and should also generalise and extrapolate well. Of course, it is probably impossible to devise a single indicator capable of fulfilling all these desiderata. Three possible alternatives are introduced in the next two subsections. The first one interprets complexity as flexibility of the equations, resulting in the potential to overfit the data. The second one is more orientated toward assessing stability, the capability of the models to provide consistent predictions in the presence of unavoidable uncertainties in the parameters. The third is more orientated toward guaranteeing a smooth extrapolation out of sample. To illustrate in a simple way the main rationale behind the proposed new versions of the criteria, functions of only one independent variable are discussed in this section. It should be noted however that, as shown in Section 5, they can be naturally extended to higher numbers of regressors.

3.1. Quantification of Model Flexibility. A possible practical approach to address the potential of a model to overfit is to quantify its flexibility, in the region covered by the database. In this respect, a good indicator is the moving average standard deviation of the model. Such an indicator, called Model Flexibility (MF) in the following, can be easily calculated by computing the moving average of $y = f(x)$ over a reasonable interval and then summing the squares as

$$\text{MovSTD}_{y,x}(x_i) = \sqrt{\frac{\sum_{j=i-\Delta}^{i+\Delta} (df/dx - \bar{df}/dx)^2}{2\Delta}}, \quad (4)$$

$$\text{MF} = \frac{\sum_{i=1}^N \text{MovSTD}_{y,x}^2(x_i)}{N}. \quad (5)$$

According to the line of thought, quantified by equations (4) and (5), model A is more complex than model B if its derivative varies more in the considered interval of the independent variable x . A very effective version of the AIC and BIC, to implement this approach, is

$$\begin{aligned} \text{AIC}_{MF} &= N \ln(\text{MSE}) + 2k + \sqrt{N} \ln(1 + \text{MF}), \\ \text{BIC}_{MF} &= N \ln(\sigma_{(\epsilon)}^2) + k \ln(N) + \sqrt{N} \ln(1 + \text{MF}). \end{aligned} \quad (6)$$

To interpret these equations, they can be profitably rewritten as

$$\begin{aligned} \text{AIC}_{MF} &= N \ln(\text{MSE}) + 2k + \sqrt{N} \ln(1 + \text{MF}) = N \ln\left(\text{MSE}(1 + \text{MF})^{1/\sqrt{N}}\right) + 2k \\ \text{BIC}_{MF} &= N \ln(\sigma_{(\epsilon)}^2) + k \ln(N) + \sqrt{N} \ln(1 + \text{MF}) = N \ln\left(\sigma_{(\epsilon)}^2(1 + \text{MF})^{1/\sqrt{N}}\right) + k \ln(N). \end{aligned} \quad (7)$$

Consequently, the higher the MF factor, and therefore the more flexible the model, the more penalising are considered the prediction errors.

Of course, the indicator MF, in order to have a real impact, cannot be calculated only for the entries of the database; otherwise, it would tend to reproduce the

classification of the first term in the previous equations. This is not a problem since the MF indicator is meant to quantify the complexity of the model. Consequently, it can be computed using much more synthetic points, albeit in the interval of the independent variables quantified by the DB. The evaluation of the new version AIC_{MF} and BIC_{MF} of the indicators can therefore be implemented with the following procedure:

- (i) Given the number of entries in the database, generate a suitable number N_{model} of independent variable points in the domain $x_{\text{min}}-x_{\text{max}}$
- (ii) Calculate the corresponding predictions of the models
- (iii) Compute the MF indicator to be inserted in AIC_{MF} and BIC_{MF}

The two main free parameters of the procedure are N_{model} and Δ . The best way to choose N_{model} consists of progressively increasing it until the values of the indicators stabilise; convergence typically requires a multiple of the entries in the database (between 3 and 10 but this problem is problem-dependent). For Δ , a safe choice is $\Delta = f \sqrt[3]{N}$ where f is equal to N_{model}/N .

For the problematic example of the sinusoid reported in Section 2, the new proposed version of the indicators performs significantly better as can be seen by inspection of Table 1.

3.2. Quantification of Sensitivity to Parameter Errors: Parameters Stability. The second alternative definition of complexity, for the purpose of model selection, is more focused on stability. The main rationale behind this approach is that the estimation of the parameters in the model is always affected by some uncertainties. Other things being equal, it is assumed that a model is to be preferred if its predictions are less sensitive to modifications of its parameters. In this perspective, models, providing more consistent predictions when their parameters are varied within their confidence intervals, are deemed more reliable and therefore to be preferentially chosen.

The implementation of the idea just described is based on the knowledge (or educated guess) of the uncertainties in the parameters of the candidate models. The procedure consists of the following:

- (i) Generating a large number of parameter combinations, sampling the probability distribution function of their uncertainties
- (ii) Calculating the predictions of the models for these combinations of parameters
- (iii) Computing a suitable estimator of the range of variability of the predictions with the variations of the parameters

For the estimator of the stability, various alternatives are viable: mean, standard deviation, max value, etc. The user can choose the most suited to the application. For example, if

the worst-case scenario is particularly relevant, an appropriate choice could be the maximum variation in the predictions. With regard to the pdf of the parameter uncertainties, the choice must of course be driven by the knowledge of the application and the type of errors affecting the model estimates. In the numerical cases presented in the following, the indicator chosen is the MSE and Gaussian noise is assumed to affect the parameters. In any case, indicating with PS the estimator of stability, the proposed version of the criteria reads

$$\begin{aligned} BIC_{PS} &= N \ln(\text{MSE}) + k \ln(N) + \sqrt{N} \ln(\text{PS}), \\ AIC_{PS} &= N \ln(\text{MSE}) + 2k + \sqrt{N} \ln(\text{PS}). \end{aligned} \quad (8)$$

For the problematic example of the sinusoid reported in Section 2, again the upgraded version of the indicators performs significantly better as can be seen by inspection of Table 1.

3.3. Quantification of Extrapolation Behaviour: Boundary Stability. In many applications, the extrapolation properties of the models are of great importance. Some models can behave very well in sample but vary wildly out of sample. When designing new devices or experiments, this fact can cause serious difficulties. To remedy or at least to alleviate this problem, the criteria can be fine-tuned to reduce the likelihood of selecting models with a wild behaviour out of sample. A good alternative consists in fitting the data excluding the boundary of the database and then quantifying how the models perform in this region. An effective procedure, to implement this approach, calculates a different boundary stability (BS) coefficient for the upper and one for the lower boundary. For the upper boundary, a suitable window dx is chosen and the points in the interval between $x_{\text{max}} - dx$ and x_{max} are not considered:

- (i) The model is fitted to the data in the reduced domain
- (ii) The points in the discarded interval are predicted and a suitable indicator of the residuals is calculated (the MSE in the following examples is called $\text{MSE}_{\text{sup,red}}$)
- (iii) The model is fitted to the whole set of data and its prediction in the reduced domain are used to calculate again the indicator of their quality, MSE_{sup}

The stability parameter BS_{sup} is calculated as follows:

$$BS_{\text{sup}} = \max\left(\frac{\text{MSE}_{\text{sup,red}}}{\text{MSE}_{\text{sup}}} - 1, 0\right). \quad (9)$$

The same algorithm is implemented also for the boundary in the lowest part of the regressor BS_{inf} . The total boundary stability is then defined as follows:

$$BS = BS_{\text{sup}} + BS_{\text{inf}}. \quad (10)$$

Inserting this indicator in the AIC and BIC criteria leads to the following formulations:

TABLE 1: Estimates of the various versions of the indicator for the problematic examples of the comparison between a polynomial (model 1) and a sinusoid (model 2) introduced in Section 2.

	BIC	AIC	BIC _{MF}	AIC _{MF}	BIC _{BS}	AIC _{BS}	BIC _{PS}	AIC _{PS}
Polynomial (correct)	-1151	-1172	-1151	-1172	-1151	-1172	-1321	-1342
Sine (incorrect)	-1167	-1178	-922	-933	-1167	-1178	-1150	-1161
Model selected	2	2	1	1	2	2	1	1
Result	Wrong	Wrong	Correct	Correct	Wrong	Wrong	Correct	Correct

$$\begin{aligned} \text{BIC}_{BS} &= N \ln(\text{MSE}(1 + \text{BS})) + k \ln(N), \\ \text{AIC}_{BS} &= N \ln(\text{MSE}(1 + \text{BS})) + 2k. \end{aligned} \quad (11)$$

For the problematic example of the sinusoid reported in Section 2, again the new proposed version of the indicators is quite competitive with the traditional AIC and BIC, as can be seen again by inspection of Table 1.

4. Systematic Tests with Synthetic Data

A series of systematic tests have been performed to investigate the properties of the developed upgrades of the criteria. Since the objective is the qualification of additional criteria for practical applications, the most commonly used classes of functions have been considered: polynomial, trigonometric, power laws, and exponential. The criteria can be applied also to density estimation (i.e., to models aimed at fitting probability distribution functions). Even for this type of task, the most popular pdfs have been tested: Gaussian, Poisson, and uniform.

In terms of results, typically the situation of Table 1 represents quite well the performance of the various criteria. The traditional versions are by far the weakest. The proposed upgrades practically never perform worse than the original versions. When the AIC and BIC identify the right model, they typically discriminate better and therefore are less vulnerable to noise. In various cases, as the one shown in Table 1, they can even converge on the right equating, when the traditional forms of AIC and BIC fail to do so.

In general, the Model Flexibility and the Parameter Stability criteria are the most coherent and reliable. The boundary stability criterion comes into its own when models are particularly problematic out of sample but in many cases does not outperform the traditional AIC and BIC.

The previous results and considerations apply also for the case of multiple regressors. An example, representative of many tests performed, is the case of the data being generated by the polynomial:

$$y = 1 + 2x_1^2 - 0.01y_2^3. \quad (12)$$

The candidate models tested are

$$y_{m,1} = a + b_1x_1 + c_1x_1^2 + d_1x_1^3 + b_2x_2 + c_2x_2^2 + d_2x_2^3, \quad (13)$$

$$y_{m,2} = A_1 \sin(b_1x_1 + c_1) + A_2 \sin(b_2x_2 + c_2). \quad (14)$$

The candidate equations (13) and (14) are so close to the original function generating the data that the traditional AIC and BIC cannot identify the right model. On the other hand,

AIC_{MF}, BIC_{MF}, AIC_{PS}, and BIC_{PS} correctly converge on the right solution. The boundary stability version of the criteria is competitive with AIC and BIC but not enough to improve the selection, as can be deduced from Table 2.

$$y = 3x^{3.3}. \quad (15)$$

The candidate models are assumed to be

$$y_{m,1} = \sum_{i=1}^8 A_i \sin(b_i x + c_i), \quad (16)$$

$$y_{m,2} = ax^b + c. \quad (17)$$

The classical BIC and AIC have problems discriminating the right model (17) from the competitive but wrong one (16). Indeed, the average values of the two indicators are very similar (Table 3) and therefore, in 25% of the cases, even the small level (10%) of noise can mislead the indicators. Using the new definitions, the difference between the two models is increased, especially in the case of the BS method. In fact, as shown in Figure 2, the sum of sines model (16) is very unstable on the boundary region of the available data, and the BS method is very sensitive to this aspect by design, preferring the right model.

The last example presented is a two-dimensional case, where the data are generated from the equation

$$y = 1 + 2x_1^2 - 0.01y_2^3. \quad (18)$$

And the candidate models are

$$y_{m,1} = a + b_1x_1 + c_1x_1^2 + d_1x_1^3 + b_2x_2 + c_2x_2^2 + d_2x_2^3, \quad (19)$$

$$y_{m,2} = A_1 \sin(b_1x_1 + c_1) + A_2 \sin(b_2x_2 + c_2). \quad (20)$$

Again, the sine functions tend to overfit the data particularly at high frequencies. In the new proposed versions of the indicators, the sensitivity to parameter changes (detected by PS) and the penalty for high model flexibility (implemented by MF) ensures the correct detection of the right model (increasing the average correct detection probability from 50% to 100%). The average results are shown in Table 4.

5. Discussion and Conclusions

Complexity is a property of systems, which is difficult to define in absolute generality. To a certain extent, the details vary with the application and the priorities of the observer. Such a context dependence is particularly evident in the case of model selection, because quantifying the complexity of mathematical functions is a notoriously arduous task. The

TABLE 2: Estimates of the various versions of the indicators for the examples of equations (16)–(18).

	BIC	AIC	BIC _K	AIC _K	BIC _{BS}	AIC _{BS}	BIC _{PS}	AIC _{PS}
Model 1	-1692	-1710	-1692	-1710	-1692	-1710	-1940	-1982
Model 2	-1693	-1735	-1325	-1342	-1693	-1735	-1703	-1720
Model selected	2	2	1	1	2	2	1	1
Result	Wrong	Wrong	Correct	Correct	Wrong	Wrong	Correct	Correct
Relative difference	-0.03%	-1.46%	21.69%	21.50%	-0.03%	-1.46%	12.22%	13.25%

A representative example to show the potential of the BS method is reported in the following. The data have been generated with the equation

TABLE 3: Estimates of the various versions of the indicators for the regressor example of equations (15)–(17).

	BIC	AIC	BIC _{MF}	AIC _{MF}	BIC _{BS}	AIC _{BS}	BIC _{PS}	AIC _{PS}	BIC _{All}	AIC _{All}
Model 1	6409	6291	6539	6422	95968	95850	6372	6254	96062	95944
Model 2	6010	6000	6045	6035	6010	6000	5735	5725	5770	5760
Model selected	2	2	2	2	2	2	2	2	2	2
Correct model detection probability	~75%	~75%	100%	100%	100%	100%	100%	100%	100%	100%

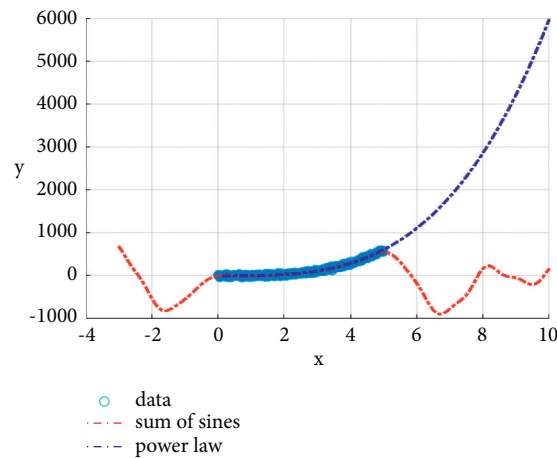


FIGURE 2: Generated data (from equation (15)) and fitted models (equations (16) and (17)). The sum of sines model clearly shows out-of-sample instability.

TABLE 4: Estimates of the various versions of the indicators for the regressor example of equations (18)–(20).

	BIC	AIC	BIC _{MF}	AIC _{MF}	BIC _{BS}	AIC _{BS}	BIC _{PS}	AIC _{PS}	BIC _{All}	AIC _{All}
Model 1	-1692	-1710	-1692	-1710	-1692	-1710	-1940	-1982	-1939	-1982
Model 2	-1693	-1735	-1325	-1342	-1725	-1737	-1703	-1720	-1335	-1352
Model selected	2	2	1	1	2	2	1	1	1	1
Correct model detection probability	~50%	~50%	100%	100%	~95%	~95%	100%	100%	100%	100%

popular solution of equating the number of parameters of a model to its complexity presents serious drawbacks. The alternatives proposed in the present work are all quite easy to implement in practice and they do not pose unrealistic requirements in terms of data availability and computational resources. They are based on three different interpretations of complexity in the context of model selection: flexibility, robustness against parameter errors, and extrapolation consistency. The first one indeed penalises flexibility as a potential for overfitting. The second favours stable models, whose predictions do not change much with small variations in the parameters. The third privileges solutions, which extrapolate smoothly out of sample.

Of course, also the criteria proposed in this paper are not a panacea and cannot claim absolute generality. On the contrary, as probably any other definition, they are better suited for certain applications. On the other hand, the systematic tests performed indicate that they have a great potential to at least complement the indicators available. They have proven to perform very well for a very large number of classes of functions of practical interest. In most cases, they show a significantly higher discriminatory power and are less prone to be completely misleading than the traditional AIC and BIC. The weakest of the upgrades proposed is certainly the Boundary Stability version; on the other hand, to test the behaviour of the models out of

sample, it can turn out to be very useful and can profitably be used as a complement to the other indicators.

In terms of developments, a significant activity has already started to include these new versions of the criteria in the genetic programmes for the automatic analysis of large databases, implementing also more advanced metrics [8] and better treatments of the error bars [9]. With regard to future applications, high-temperature plasmas and environmental sciences are obvious targets [10, 11]. In nuclear fusion, some crucial quantities for the design of new experiments, such as the energy confinement time or the power threshold to access the H mode, are derived from empirical databases; the new versions of the indicators could lead to more robust empirical scaling laws [12, 13]. Other relevant topics could be impurity studies [14] and the control of the current profile [15], particularly in metallic devices such as JET with the ITER Like Wall [16]. With regard to the Earth sciences, better model selection could help not only in the investigation of complex interactions between atmospheric phenomena [17], but also in optimising remote sensing techniques [18].

The physical problems just mentioned are natural fields of application for the indicators developed in this work, since they typically require modelling systems with a relatively limited number of variables but with strong nonlinear interactions. A different, extremely interesting, and challenging task would be deployment of the new complexity definitions to phenomena that require many parameters to be fitted. This is another frontier of complexity typical, for example, of “sloppy models” in biochemical networks [19, 20]. These systems usually need a huge number of parameters to describe the reaction kinetics, whose details are not known and must be derived by experimental data. Given the limited observability of the kinetics and the noisy character of the measurements [21], the collected data present large uncertainties, which render parameter estimation quite problematic. Moreover, it is usually found that this data can be described by completely different models but with the same complexity and MSE. In this context, the new definitions of complexity, introduced in this paper, could provide some useful guidance about the most appropriate models to select. Indeed, the MF complexity can help in understanding the “internal stability” (avoiding solutions like the sine overfitting shown in this paper); the PS complexity would ensure the choice of more stable parameters, while the BS could guarantee more stable extrapolation.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; and in the decision to publish the results.

Authors' Contributions

Andrea Murari and Riccardo Rossi have contributed equally to the paper.

Acknowledgments

One of the authors (T.C.) acknowledges the financial support received from the contract 1EU- 4/2 funded by Romanian Research and Innovation Ministry.

References

- [1] S. C. Rohilla, *Methods and Techniques of Complex Systems Science: An Overview in Complex Systems Science in Biomedicine*, Springer, Berlin, Germany, 2006.
- [2] Y. Bar-Yam, *General Features of Complex Systems*, New England Complex Systems Institute, Cambridge, MA, USA, 2002.
- [3] K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, Springer-Verlag, Berlin, Germany, 2nd edition, 2002.
- [4] V. N. Vapnik, *The Nature of Statistical Learning Theory. Information Science and Statistics*, Springer-Verlag, Berlin, Germany, 2000.
- [5] G. Giorgio and S. Marcello, “Approximation error bounds via rademacher’s complexity,” *Applied Mathematical Sciences*, vol. 2, no. 4, pp. 153–176, 2008.
- [6] H. Akaike, “Information theory and an extension of the maximum likelihood principle,” in *Proceedings of the 2nd International Symposium on Information Theory*, Tsahkadsor, Armenia, September 1971.
- [7] G. E. Schwarz, “Estimating the dimension of a model,” *Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.
- [8] A. Murari, P. Boutot, J. Vega et al., “Clustering based on the geodesic distance on Gaussian manifolds for the automatic classification of disruptions” *Nucl. Fusion*, vol. 53, p. 33006, 2013.
- [9] A. Murari, E. Peluso, M. Gelfusa, M. Lungaroni, and P. Gaudio, “How to handle error bars in symbolic regression for data mining in scientific applications,” *Statistical Learning and Data Sciences*, vol. 9047, pp. 347–355, 2015c.
- [10] F. Romanelli and R. Kamendje, “Overview of JET results,” *Nuclear Fusion*, vol. 49, no. 10, p. 104006, 2009.
- [11] J. Ongena, “Towards the realization on JET of an integrated H-mode scenario for ITER,” *Nuclear Fusion*, vol. 44, no. 1, pp. 124–133, 2004.
- [12] A. Murari, E. Peluso, M. Lungaroni, M. Gelfusa, and P. Gaudio, “Application of symbolic regression to the derivation of scaling laws for tokamak energy confinement time in terms of dimensionless quantities,” *Nuclear Fusion*, vol. 56, no. 2, p. 26005, 2015.
- [13] A. Murari, E. Peluso, M. Gelfusa, I. Lupelli, and P. Gaudio, “A new approach to the formulation and validation of scaling expressions for plasma confinement in tokamaks,” *Nuclear Fusion*, vol. 55, no. 7, p. 73009, 2015.
- [14] M. E. Puiatti, M. Mattioli, G. Telesca et al., “Radiation pattern and impurity transport in argon seeded ELMy H-mode discharges in JET,” *Plasma Physics and Controlled Fusion*, vol. 44, no. 9, pp. 1863–1878, 2002.
- [15] D. Mazon, “Tokamak edge turbulence: background theory and computation,” *Plasma Physics and Controlled Fusion*, vol. 45, p. 7, 2003.

- [16] J. Pamela, "The JET programme in support of ITER," *Fusion Engineering and Design*, vol. 82, no. 5–14, pp. 590–602, 2007.
- [17] T. Craciunescu and A. Murari, "Geodesic distance on Gaussian manifolds for the robust identification of chaotic systems," *Nonlinear Dynamics*, vol. 86, no. 1, pp. 677–693, 2016.
- [18] A. F. Villaverde and A. Barreiro, "Identifiability of large nonlinear biochemical networks," *Communications in Mathematical and in Computer Chemistry*, vol. 76, pp. 259–269, 2016.
- [19] P. Gaudio, P. Gaudio, M. Gelfusa et al., "In-cell measurements of smoke backscattering coefficients using a CO₂ laser system for application to lidar-dial forest fire detection," *Optical Engineering*, vol. 49, no. 12, p. 124302, 2010.
- [20] M. K. Transtrum, "Perspective: sloppiness and emergent theories in physics, biology, and beyond," *The Journal of Chemical Physics*, vol. 143, no. 1, 2015.
- [21] J. R. Banga and E. Balsa-Canto, "Parameter estimation and optimal experimental design," *Essays in Biochemistry*, vol. 45, pp. 195–210, 2008.