

Article

Upgrading Model Selection Criteria with Goodness of Fit Tests for Practical Applications

Riccardo Rossi ^{1,*},[†] , Andrea Murari ^{2,†}, Pasquale Gaudio ¹  and Michela Gelfusa ¹ 

¹ Department of Industrial Engineering, University of Rome “Tor Vergata”, via del Politecnico 1, 01100 Roma, Italy; gaudio@ing.uniroma2.it (P.G.); gelfusa@ing.uniroma2.it (M.G.)

² Consorzio RFX (CNR, ENEA, INFN, Università di Padova, Acciaierie Venete SpA), Corso Stati Uniti 4, 35127 Padova, Italy; Andrea.murari@euro-fusion.org

* Correspondence: R.Rossi@ing.uniroma2.it

† These authors contributed equally to the work.

Received: 31 March 2020; Accepted: 13 April 2020; Published: 15 April 2020



Abstract: The Bayesian information criterion (BIC), the Akaike information criterion (AIC), and some other indicators derived from them are widely used for model selection. In their original form, they contain the likelihood of the data given the models. Unfortunately, in many applications, it is practically impossible to calculate the likelihood, and, therefore, the criteria have been reformulated in terms of descriptive statistics of the residual distribution: the variance and the mean-squared error of the residuals. These alternative versions are strictly valid only in the presence of additive noise of Gaussian distribution, not a completely satisfactory assumption in many applications in science and engineering. Moreover, the variance and the mean-squared error are quite crude statistics of the residual distributions. More sophisticated statistical indicators, capable of better quantifying how close the residual distribution is to the noise, can be profitably used. In particular, specific goodness of fit tests have been included in the expressions of the traditional criteria and have proved to be very effective in improving their discriminating capability. These improved performances have been demonstrated with a systematic series of simulations using synthetic data for various classes of functions and different noise statistics.

Keywords: model selection criteria; Bayesian information criterion (BIC); Akaike information criterion (AIC); Shannon entropy; goodness of fit tests; Kolmogorov–Smirnov test

1. Introduction to Model Selection Criteria Based On Bayes and Information Theory

The selection of the most appropriate model, to describe the phenomena under study, is a major concern of modern science [1,2]. Statisticians have naturally been involved in this task, and so it is not surprising that many statistical approaches have been proposed over the years for dealing with this key issue. Indeed, model selection has been investigated extensively from both frequentist and Bayesian perspectives. Many tools for identifying the “best model” among a set of candidates have been suggested in the literature [3,4]. Two of the most widely used model selection families of indicators are the Akaike information criterion (AIC) [5] and the Bayesian information criterion (BIC) [6]. The AIC is an information-theoretic indicator based on the Kullback–Leibler Divergence [7]; it essentially quantifies the information lost by a given model. Therefore, the basic principle underlying the AIC criterion is the assumption that the less information a model loses, the higher is its quality. Bayesian theory informs the BIC criterion, which is designed to maximize the posterior probability of a model given the data [7].

The theoretical derivations of these metrics result in the following unbiased forms of the criteria:

$$AIC = -2 \ln(L) + 2k \quad (1)$$

$$BIC = -2 \ln(L) + k \ln(n) \quad (2)$$

where L is the likelihood of the model given the data, k the number of estimated parameters in the model, and n the number of entries in the database. Both AIC and BIC metrics are basically cost functions, which have to be minimized; they favor models with a high likelihood but implement a penalty for complexity (the term proportional to k).

The AIC and BIC criteria are very well known and have found many interesting applications. On the other hand, their original formulation is not necessarily easy to implement in practice. A very delicate part is the likelihood of the models, which can be virtually impossible to calculate. This difficulty can be due to various causes, ranging from the type of noise affecting the data to the nature of the models to be tested and the “a priori” information about their properties [8]. To overcome the practical difficulties of calculating the likelihood, one typically makes recourse to the traditional assumption that the model and data errors are identically distributed and independently sampled from a normal distribution. If this hypothesis is valid, 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):

$$AIC = n \cdot \ln(MSE) + 2k \quad (3)$$

where $\sigma_{(\epsilon)}^2$ is the variance of the residuals.

Similar assumptions allow expressing the BIC criterion as:

$$BIC = n \cdot \ln(\sigma_{(\epsilon)}^2) + k \cdot \ln(n) \quad (4)$$

where MSE is the mean-squared error of the residuals.

In Equations (3) and (4), derived, for example, in [7], the MSE and variance are calculated based on the residuals and the differences between the data and the estimates of the models.

As can be easily appreciated by inspection of Equations (3) and (4), which constitute the most widely used forms of AIC and BIC, the statistical information, originally provided by the likelihood, is reduced to the MSE and variance of the residuals. The natural question is, therefore, whether some additional statistical information about the distribution of the residuals could be taken into account and improve the performance of the two criteria. The practical importance of this question is not to be underestimated because, in many applications, the assumptions leading to expressions (3) and (4) are clearly violated. The first attempt in this direction, reported in [9], relied on the Shannon entropy as an indicator of the residual distribution. The encouraging results, obtained with this upgrade, has motivated the insertion of more sophisticated summary statistics of the distribution of the residuals into the AIC and BIC, to see whether they could further increase their performance. The paper described this line of investigation in detail and is structured as follows. The goodness of fit tests, implemented to improve AIC and BIC, are described in Section 2. The families of models and the typologies of statistical noise investigated are the subjects of Section 3. The following Section 4 describes the results obtained for three very important families of models: exponential functions, polynomials, and power laws. Section 5 reports the results for a scan of the noise amplitude. Summary and future developments are the subjects of the last Section 6.

2. Model Selection and Goodness of Fit Tests

As mentioned in the introduction, a previous attempt to improve the practical implementation of AIC and BIC was based on the Shannon entropy of the residuals. The main idea behind that approach was the observation that, if a model was perfect, the residuals should reflect the statistics of the noise affecting the data. Assuming additive random noise and other things being equal, models, whose residuals present a more uniform probability density function (pdf), should, therefore, be preferred. Indeed, the residuals of inferior models should contain patterns present in the data and not properly identified by the models. The Shannon entropy H can be interpreted as an indicator of how uniform is

the distribution of the residuals and can, therefore, be included in the AIC and BIC to favor models with higher values of H . In this perspective, the following versions of the BIC and AIC criteria were tested:

$$BIC_H = n \cdot \ln\left(\frac{\sigma(\epsilon)^2}{H}\right) + k \cdot \ln(n) \quad (5)$$

$$AIC_H = n \cdot \ln\left(\frac{MSE}{H}\right) + 2k \quad (6)$$

where $H = -\sum_i p_i \ln p_i$ indicates the Shannon entropy of the residuals, and p_i is the probability of the i -th residual.

These new versions of the criteria clearly outperformed the traditional version of the AIC and BIC, and they had also asymptotic convergence [9]. On the other hand, entropy is a quite crude indicator of the residual distributions. Moreover, in many practical applications, the statistics of the noise is not necessarily Gaussian. It was, therefore, reasonable to investigate whether more sophisticated tests could further improve the discriminatory power of the AIC and BIC criteria. To this end, various goodness of fit tests had been implemented. The main rationale was that, very often, it was possible to determine experimentally the actual statistics of the noise affecting the data. At this point, better models should present residuals with a distribution more similar to the one of the noise.

The goodness of fit tests investigated in this work were:

1. Chi-squared;
2. Anderson Darling;
3. Kolmogorov–Smirnov.

The null distribution of all these statistics is calculated under the null hypothesis that the sample is drawn from the reference distribution, the pdf of the noise in our case.

The “chi-squared test”, or χ^2 test, implemented in this paper was the Pearson’s chi-squared test [10], which is often used to determine whether there is a statistically significant difference between two probability distribution functions (pdfs). In our application, as already discussed, the two distributions to be compared were one of the residuals and one of the noises affecting the data.

The “Anderson–Darling” test can also be used to assess whether a sample is drawn from a specified distribution function [11]. It is based on the fact that, assuming the data does arise from this distribution, its cumulative distribution function (CDF) is expected to follow the one, which can be derived by the original distribution. In our application, the residuals could be then tested for uniformity with a distance test [12]. The test statistics could then be compared with the critical values of the distribution expected on the basis of the knowledge of the noise.

The “Kolmogorov–Smirnov” test (K–S test) is another nonparametric test to determine the equality between one-dimensional probability distributions or, as in our case, between a pdf and a sample [13]. The statistic of the K–S quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution. The null distribution of this statistic is calculated under the null hypothesis that the sample is drawn from the reference distribution.

The results of the goodness of fit tests could be summarized with their Z score; the lower its value, the closer the residuals to the expected pdf of the noise. Since the AIC and BIC criteria were indicators to be minimized, the Z scores of the goodness of fit tests could be included in their mathematical expressions as follows:

$$AIC_{GF} = n \cdot \ln\left(\frac{MSE}{H}(1 + Z_{score}^2)\right) + 2k \quad (7)$$

$$BIC_{GF} = n \cdot \ln\left(\frac{\sigma_\epsilon^2}{H}(1 + Z_{score}^2)\right) + k \ln(n) \quad (8)$$

The interpretation of this new version of the AIC and BIC criteria was very intuitive. The better the model, the lower the Z score of the residuals (since they are closer to the pdf of the noise) and, therefore,

the lower the numerical value of the criteria. Equations (7) and (8) suggested also a consideration about the asymptote stability of AIC_{GF} and BIC_{GF} (where GF means Goodness of Fit); if the model was perfect, the Z score would tend to zero with increasing the number of points, and the AIC_{GF} and BIC_{GF} would converge to AIC_H and BIC_H . The asymptotic converge of the criteria using the entropy of the residuals has already been demonstrated in [9], proving the asymptotic convergence of AIC_{GF} and BIC_{GF} as well. The AIC_{GF} and BIC_{GF} of Equations (7) and (8) were, therefore, the new versions of the model selection criteria that had been tested and whose performance is described in the rest of the paper.

3. Families of Functions and Noise Statistics

To investigate and quantify the performance of the alternative formulations of the model selection criteria, AIC_{GF} and BIC_{GF} , a series of systematic tests were performed. The analysis was focused mainly on three classes of models, which were among the most useful and used in practice. They were the class of exponential functions, polynomials, and power laws. For each class, a representative example is discussed in the following. All the analyses were performed on bidimensional functions ($z = f(x, y)$).

The exponential functions investigated in this paper had the form:

$$z(x, y) = axe^{(bx+cy)} + dx e^{(ex+fy)} + g \quad (9)$$

Polynomials are mathematical expressions that contain two or more algebraic terms, which can be added, subtracted, or multiplied. In general, polynomial expressions include at least one variable and typically also constants and positive exponents. Polynomial functions had the following form:

$$z(x, y) = p_{00} + p_{10}x + p_{20}x^2 + p_{01}y + p_{02}y^2 + p_{03}y^3 + p_{11}x + p_{21}x^2y + p_{12}xy^2 \quad (10)$$

The power laws considered in this paper were monomials of the form:

$$z(x, y) = cx^a y^b \quad (11)$$

where the exponents could be either positive or negative.

With regard to the noise statistics, four of the most relevant distribution functions have been tested: Gaussian, uniform, Poisson, and gamma-distributed noise [14]. For the reader convenience, the mathematical formulation of these types of noise is reported in the following, while one representative case for each distribution is shown graphically in Figure 1.

Gaussian distribution:

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2}$$

Uniform distribution:

$$f(x) = \frac{1}{b-a} \text{ for } a \leq x \leq b$$

$$f(x) = 0 \text{ otherwise}$$

Poisson distribution:

$$f(x) = \frac{\lambda^n}{n!} e^{-\lambda} \quad \forall n \in \mathbb{N}$$

Gamma distribution:

$$f(x) = \frac{x^{k-1} e^{-x/\theta}}{\theta^k \Gamma(k)} \quad \forall x > 0 \text{ and } k, \theta \in \mathbb{N}$$

The parameters chosen for the examples reported in the paper were:

Gaussian distribution: $\mu = 0$ e $\sigma = 3$.

Uniform distribution: $a = -1$ e $b = 1$;
 Poisson distribution: $\lambda = 10$;
 Gamma distribution: $k = 3$ e $\theta = 2$.

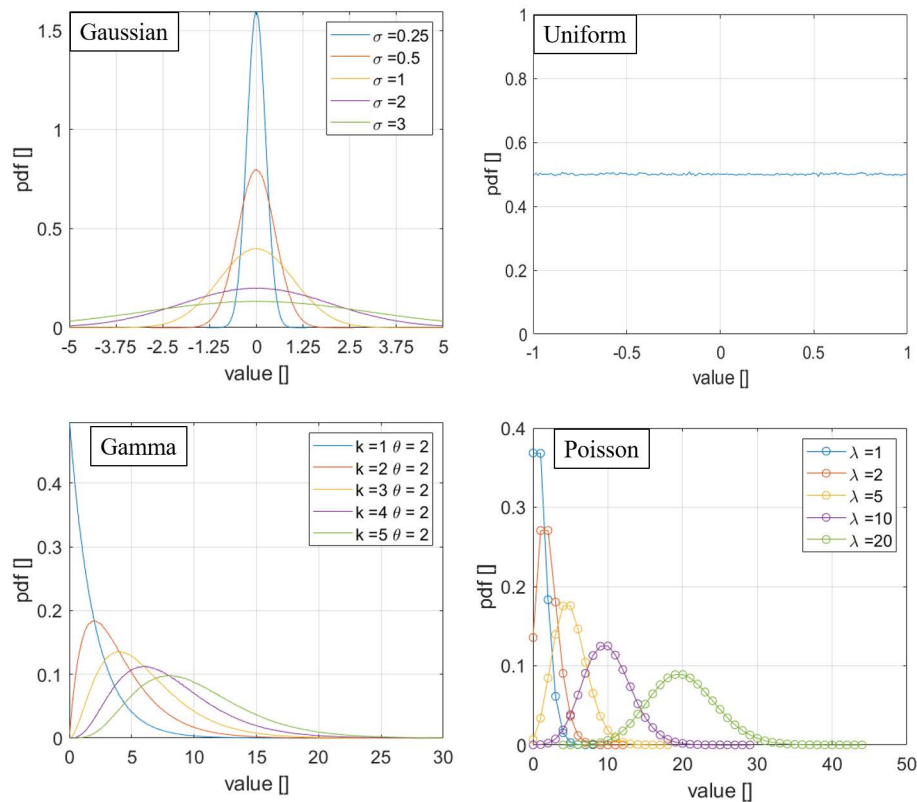


Figure 1. The probability density function of the four distributions used in the analyses reported in this paper.

4. Results for Exponential Functions, Polynomials, and Power Laws

All the analyses were performed by means of numerical tests. A systematic analysis of many cases for each class of functions had been performed. Given a domain of x and y , the function $z = f(x, y)$ was built. Then, the noise with a specific distribution was added to the function. These data represented the points to fit.

To summarize the results in a concise way, in the following, for each class of functions, a representative case was reported. The values of the traditional AIC and BIC (Equations (3) and (4)) and of AIC_H and BIC_H were compared with those of AIC_{GF} and BIC_{GF} for two cases: the actual correct model, called “Correct”, and a very competitive alternative equation, called “Alternative”. The alternative equation had been obtained by fitting a different function to the data in such a way that the traditional indicator of the quality of the fit—RMSE and variance—were very competitive with the ones of the correct model. This fit had been performed with a nonlinear least-squares minimization method, which finds the parameters of a nonlinear equation by an iterative approach [15]. Being this method based on the minimization of the least squares, the function found was the one, which minimized also the classic versions of BIC and AIC. Therefore, the quality of the alternative criteria could be analyzed to see whether they showed better performance in the convergence to the right model, the one actually used to generate the synthetic data. The domain of each example was $0 < x < 10$ and $-10 < y < 10$. A scan in the number of entries had been performed for all four types of noise distributions.

The exponential function used to generate the data was:

$$z(x, y) = 100(x + y)e^{-x} \quad (12)$$

while the function for the fit was the one given in Equation (9). A plot of the original function and with added noise of Poisson distribution is shown in Figure 2.

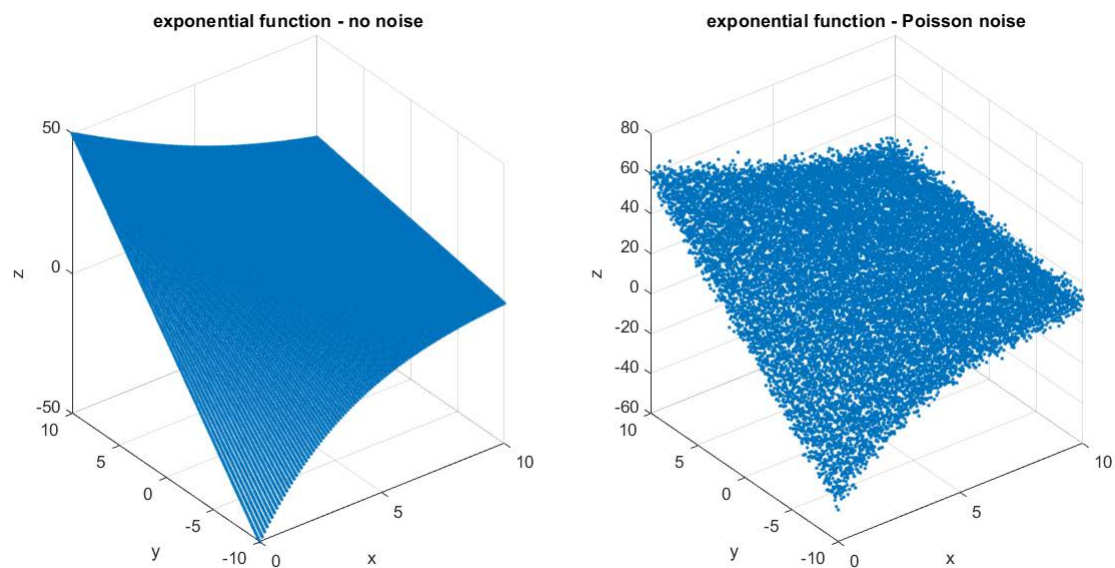


Figure 2. Exponential function scatter plot in the case of the no noise (**left**) and Poisson noise (**right**).

Figure 3 reports the results for the exponential functions. In the plots of this figure, as in those of Figures 4–6, the x-axis reported the number of points generated from the original function and used to test the various criteria. Moreover, each result was an average of over 10 repetitions of the calculations. The smoothness of the obtained curves was, of course, an implicit demonstration of their stability. Again on the nomenclature, the “Correct” function referred to the original function used to generate the data (Equation (12)), while “Alternative” represented the function calculated by a regression tool, which was based on the minimization of the classic BIC and AIC. The tool used for the regression was the Curve Fitting Toolbox of Matlab, which allows performing both linear and non-linear fitting by using the least-square minimization method. For non-linear least squares, the robustness of the model could be achieved by the LAR (Least Absolute Residuals) or Bisquare approach, while the algorithms implemented were the Trust-Region and the Lavemberg–Marquardt.

The same analysis had been performed for polynomial type functions. The function used to generate the data was:

$$z(x, y) = 3x + x^2 + y + 3y^4 \quad (13)$$

while the fit function was reported in Equation (10). The results are reported graphically in Figure 4.

The same analysis had been performed for power-law type functions. The function used to generate the data was:

$$z(x, y) = cx^3y^{-1} \quad (14)$$

While the fit function was reported in Equation (11). The results are reported graphically in Figure 5.

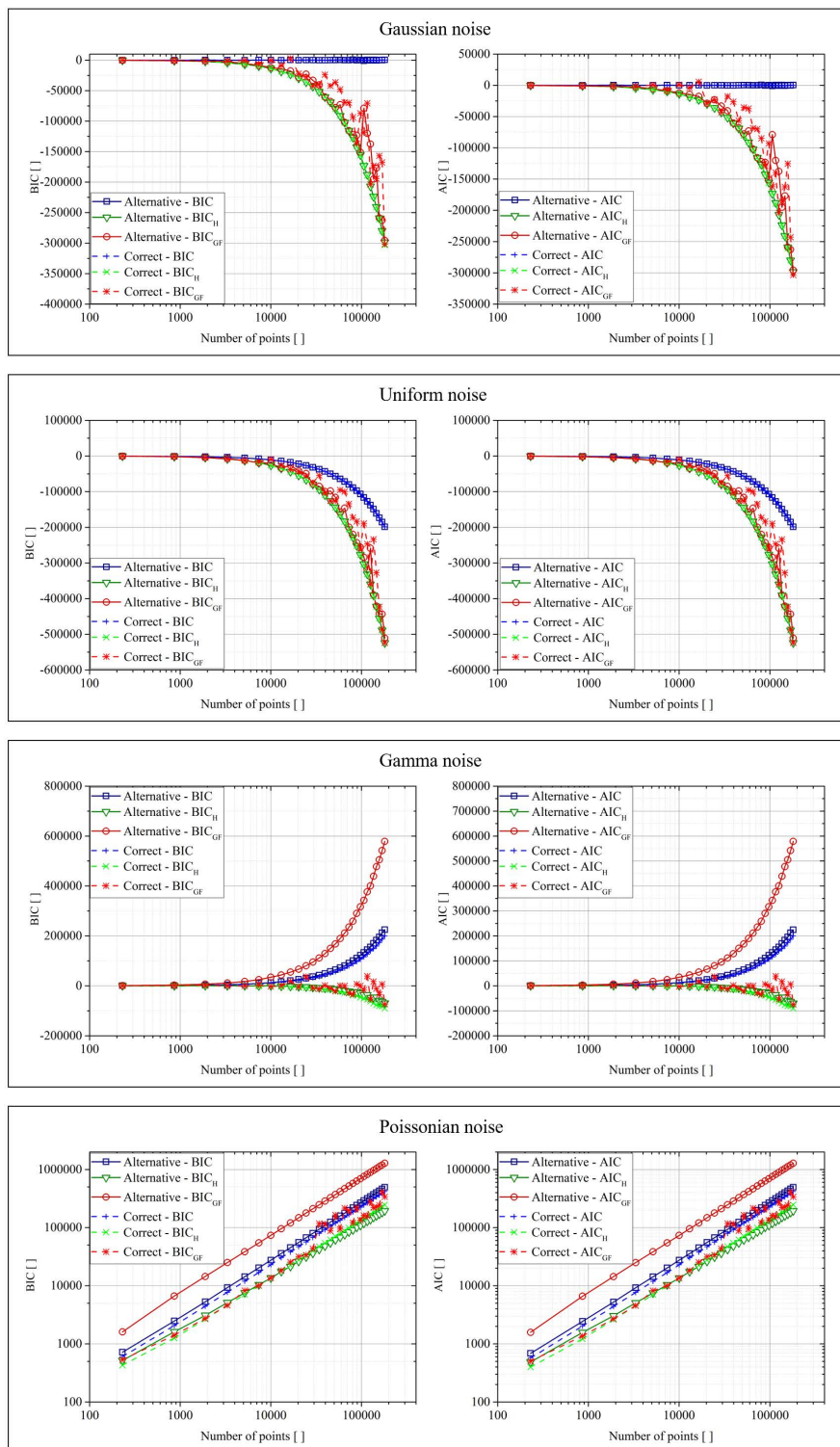


Figure 3. BIC (Bayesian information criterion) and AIC (Akaike information criterion) for the two fit functions in the case of the four different types of noise for the exponential function.

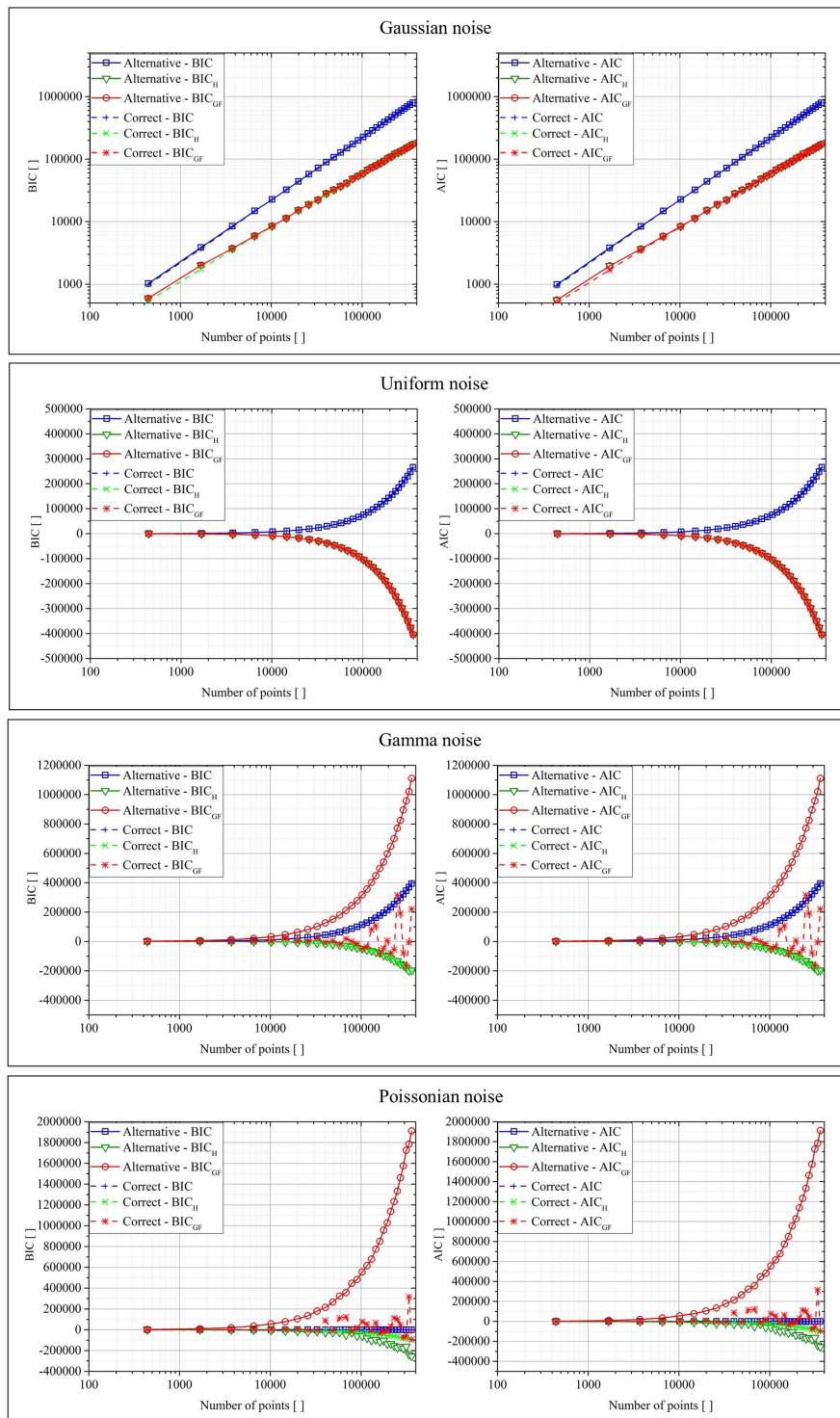


Figure 4. BIC and AIC for the two fit functions in the case of the four different types of noise for the polynomial function.

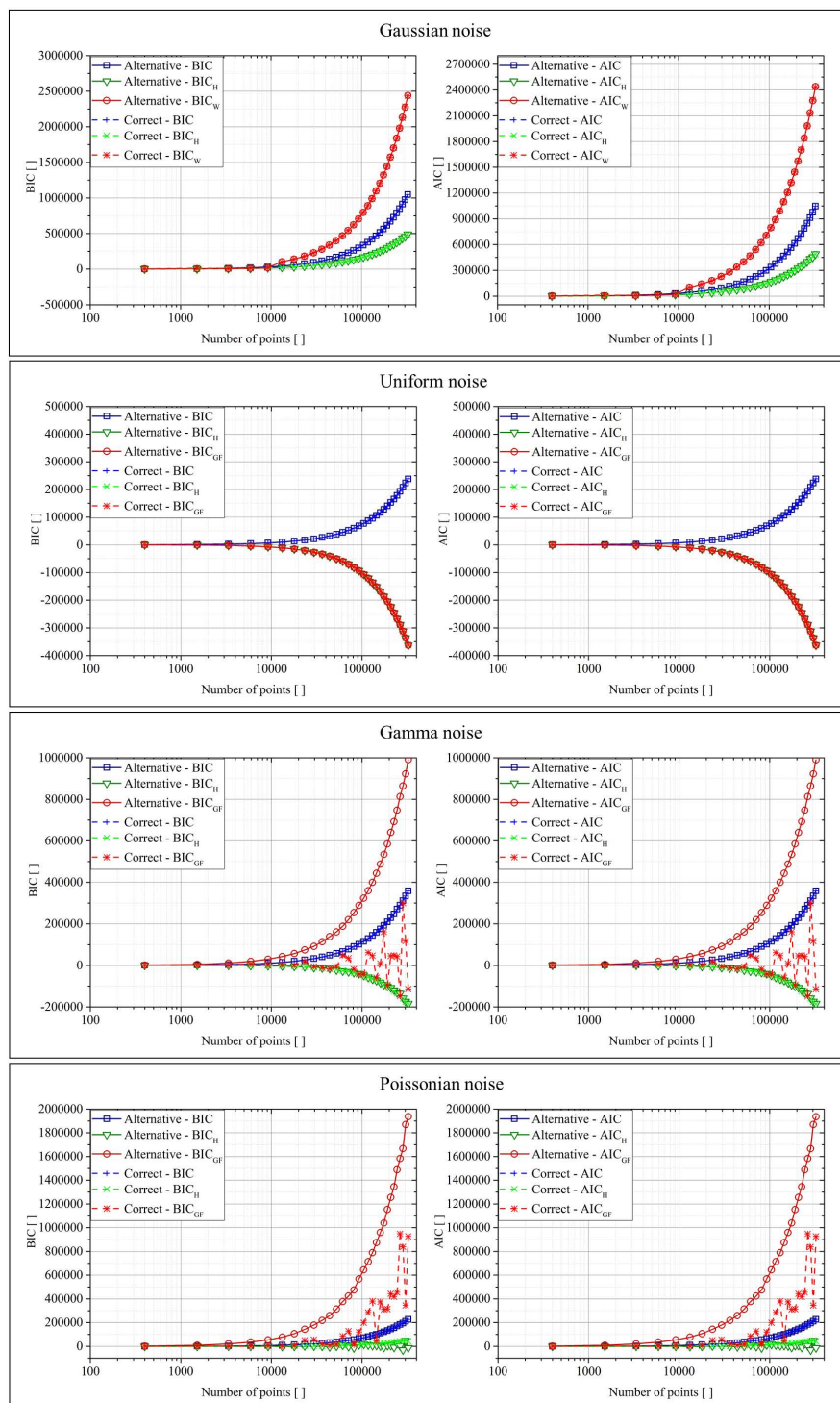


Figure 5. BIC and AIC for the two fit functions in the case of the four different types of noise for the power-law function.

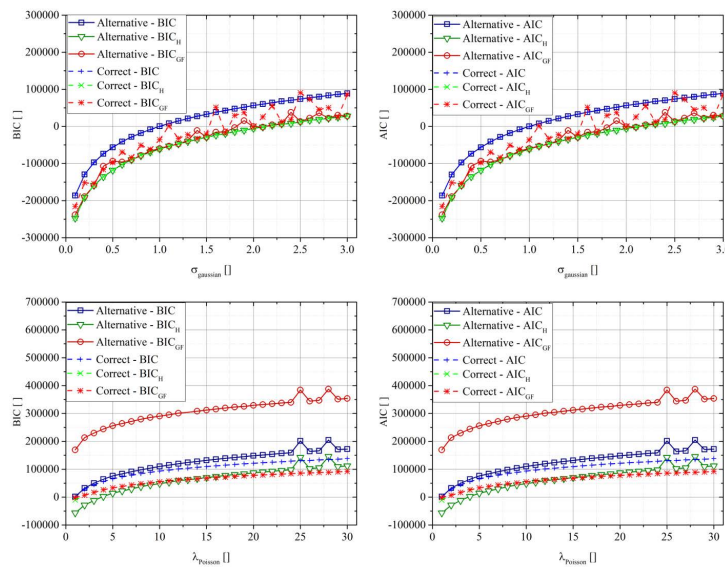


Figure 6. BIC and AIC for the two fit functions in the case of the four different types of noise for the power-law function.

The interpretation of the previously reported results was the same for all the classes of functions. For Gaussian and flat distributed noise, all versions of both AIC and the BIC (traditional formulation, Shannon entropy, and goodness of fit) returned basically equal values in all cases. As shown in [9], for these cases, the upgraded versions— AIC_H , BIC_H and AIC_{GF} , BIC_{GF} —improve only the convergence to the right model in limited and very difficult cases. For the other two types of noise, the situation was dramatically different. The use of the “goodness of fit” approach gave a much smaller BIC_{GF} and AIC_{GF} in the case of the correct function, improving systematically the convergence to the right model. Basically, this result was due to the fact that, in the traditional AIC and BIC, the minimization of the errors automatically implies that the fit acts with the purpose of minimizing the mean of the errors (by bringing it as close as possible to zero). This requirement was not appropriate in the case of data affected by the noise of Poisson and gamma distributions, which were non-symmetric with respect to zero. For these cases, the goodness of fit tests significantly improved the capability of the criteria to select the right model. In general, BIC_{GF} and AIC_{GF} are expected to provide appreciable better performance in all cases whose data are affected by no zero-sum distributed noise. Concerning the rate of convergence and the effects of data size, the most important thing to consider is that all goodness of fit tests need an adequate amount of data in order to return a reliable Z-score. On the other hand, from the plots in the previous figures, it could be seen how AIC_{GF} and BIC_{GF} provided better performance than AIC and BIC, even for a quite limited number of points, of the order of 100, which is a very real demand in most practical applications. Therefore, the rate of convergence of the proposed criteria seemed also to be quite satisfactory.

5. Effect of Noise Intensity

Figure 6 compares the various BIC and AIC definitions in the case of the exponential function, changing the intensity of two types of noise, Gaussian and Poisson. In the case of the Gaussian distribution, the parameter was the standard deviation of the distribution, while in the case of the Poisson, the parameter λ .

The parametric analysis validated the previous results for any tested noise intensity. In the case of Gaussian distributions, the use of the GF approach did not lead to real improvements. In the case of the Poisson distribution, both AIC_{GF} and BIC_{GF} significantly improved the selection criteria, leading to a large difference between the “Correct” and the “Alternative” cases, independently from the noise intensity.

6. Conclusions

The most widely used versions of the model selection criteria AIC and BIC are valid under the assumption that the data are affected by Gaussian, zero-sum additive noise. This is a consequence of the fact that, in most practical applications in science and engineering, often it is very difficult, if not impossible, to compute the likelihood of the data given the model. In this situation, the AIC and BIC have been reformulated as reported in Equations (3) and (4), in which the statistical information about the residuals is completely summarized by their MSE and variance. If the data is affected by noise non-symmetric with respect to zero, these versions of the criteria can fail badly because they favor models with zero-sum residuals. On the contrary, the new proposed versions of the selection criteria, AIC_{GF} and BIC_{GF} , present significantly better performance because they take into account the fact that the residuals should present a probability distribution function as close as possible to the one of the noise. Indeed, all the numerical tests performed indicate that, if the noise distribution is known and sufficient amount of data are available to compute reliably the Z scores, AIC_{GF} and BIC_{GF} are the versions of the criteria to rely on. Under these conditions, they have always proved to have much better convergence properties in all respects. As a consequence, the proposed new versions of the selection criteria are expected to be deployed quite systematically in various fields of complex science, ranging from high-temperature plasmas [16–23] to atmospheric physics [24,25]. Another interesting application could be found in the regularization of recent tomographic inversion methods [26–28].

From a methodological standpoint, it should be mentioned that the upgraded versions, AIC_{GF} and BIC_{GF} , include frequentist indicators in information-theoretic and Bayesian criteria. This approach seems quite promising and worth further investigations. The proposed criteria could also be further improved by implementing more advanced metrics, such as the geometric distance and the Venn definition of probability [29–31]. In this perspective, a combination of the proposed upgrades with recently developed extensions of the AIC criterion, using a class of pseudo distances instead of the Kullback-Leiber divergence, should also be seriously pursued [32].

Author Contributions: Data curation, R.R.; Formal analysis, A.M., R.R.; Funding acquisition, P.G., M.G.; Methodology, A.M., R.R.; Project administration, M.G.; Software, R.R.; Validation, R.R., A.M., M.G.; Writing—original draft, A.M.; Writing—review and editing, R.R., A.M. R.R and A.M. authors contributed equally to the work. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest. 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.

References

1. Bailly, F.; Longo, G. *Mathematics and the Natural Sciences*; Imperial College Press: London, UK, 2011.
2. D’Espagnat, B. *On Physics and Philosophy*; Princeton University Press: Oxford, UK, 2002.
3. Claeskens, G. Statistical model choice. *Annu. Rev. Stat. Appl.* **2016**, *3*, 233–256. [[CrossRef](#)]
4. Corder, G.W.; Foreman, D.I. *Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach*; Wiley: Hoboken, NJ, USA, 2009; ISBN 978-0-470-45461-9.
5. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723. [[CrossRef](#)]
6. Schwarz Gideon, E. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464. [[CrossRef](#)]
7. Kenneth, P.B.; Anderson, D.R. *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2002.
8. Dubois, G. *Modeling and Simulation*; Taylor & Francis: Abingdon, UK; CRC Press: Boca Raton, FL, USA, 2018.
9. Murari, A.; Peluso, E.; Cianfrani, F.; Gaudio, P.; Lungaroni, M. On the use of entropy to improve model selection criteria. *Entropy* **2019**, *21*, 394. [[CrossRef](#)]

10. Pearson, K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1900**, *50*, 157–175. [[CrossRef](#)]
11. Anderson, T.W.; Darling, D.A. Asymptotic theory of certain “goodness-of-fit” criteria based on stochastic processes. *Ann. Math. Stat.* **1952**, *23*, 193–212. [[CrossRef](#)]
12. Shapiro, S.S. How to test normality and other distributional assumptions. In *The ASQC Basic References in Quality Control: Statistical Techniques 3*; American Society for Quality: Milwaukee, WI, USA, 1980; pp. 1–78.
13. Marsaglia, G.; Tsang, W.W.; Wang, J. Evaluating Kolmogorov’s distribution. *J. Stat. Softw.* **2003**, *8*, 1–4. [[CrossRef](#)]
14. Bonamente, M. *Statistics and Analysis of Scientific Data*; Graduate Texts in Physics; Springer Science + Business Media: Berlin/Heidelberg, Germany, 2017.
15. Kelley, C.T. *Iterative Methods for Optimization*, *SIAM Frontiers in Applied Mathematics*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 1999; ISBN 0-89871-433-8.
16. Romanelli, F.; Kamendjeon Behalf of JET-EFDA Contributors. Overview of JET results. *Nucl. Fusion* **2009**, *49*, 104006. [[CrossRef](#)]
17. Ongena, J.; Monier-Garbet, P.; Suttrop, W.; Andrew, P.; Bécoulet, M.; Budny, R.; Corre, Y.; Cordey, G.; Dumortier, P.; Eich, T. Towards the realization on JET of an integrated H-mode scenario for ITER. *Nucl. Fusion* **2004**, *44*, 124–133. [[CrossRef](#)]
18. Murari, A.; Lupelli, I.; Gelfusa, M.; Gaudio, P. Non-power law scaling for access to the H-mode in tokamaks via symbolic regression. *Nucl. Fusion* **2013**, *53*, 043001. [[CrossRef](#)]
19. Murari, A.; Peluso, E.; Gelfusa, M.; Lupelli, I.; Lungaroni, M.; Gaudio, P. Symbolic regression via genetic programming for data driven derivation of confinement scaling laws without any assumption on their mathematical form. *Plasma Phys. Control. Fusion* **2015**, *57*, 014008. [[CrossRef](#)]
20. Murari, A.; Peluso, E.; Lungaroni, M.; Gelfusa, M.; Gaudio, P. Application of symbolic regression to the derivation of scaling laws for tokamak energy confinement time in terms of dimensionless quantities. *Nucl. Fusion* **2015**, *56*, 26005. [[CrossRef](#)]
21. Murari, A.; Lupelli, I.; Gaudio, P.; Gelfusa, M.; Vega, J. A statistical methodology to derive the scaling law for the H-mode power threshold using a large multi-machine database. *Nucl. Fusion* **2012**, *52*, 063016. [[CrossRef](#)]
22. Murari, A.; Pisano, F.; Vega, J.; Cannas, B.; Fanni, A.; Gonzalez, S.; Gelfusa, M.; Grosso, M.; JET EFDA Contributors. Extensive statistical analysis of ELMs on JET with a carbon wall. *Plasma Phys. Control. Fusion* **2014**, *56*. [[CrossRef](#)]
23. Orsitto, F.P.; Boboc, A.; Gaudio, P.; Gelfusa, M.; Giovannozzi, E.; Mazzotta, C.; Murari, A.; JET-EFDA Contributors. Mutual interaction of faraday rotation and cotton-mouton phase shift in JET polarimetric measurements. *Rev. Sci. Instrum.* **2010**, *81*, 10D533. [[CrossRef](#)] [[PubMed](#)]
24. Gaudio, P.; Gelfusa, M.; Malizia, A.; Richetta, M.; Antonucci, A.; Ventura, P.; Murari, A.; Vega, J. Design and development of a compact Lidar/Dial system for aerial surveillance of urban areas. In Proceedings of the SPIE—The International Society for Optical Engineering, Dresden, Germany, 23–26 September 2013; p. 8894.
25. Gaudio, P. New frontiers of forest fire protection: A portable laser system. *WSEAS Trans. Environ. Dev.* **2013**, *9*, 195–205.
26. Craciunescu, T.; Peluso, E.; Murari, A.; Gelfusa, M.; JET Contributors. Maximum likelihood bolometric tomography for the determination of the uncertainties in the radiation emission on JET TOKAMAK. *Rev. Sci. Instrum.* **2018**, *89*, 053504. [[CrossRef](#)]
27. Craciunescu, T.; Bonheure, G.; Kiptily, V.; Murari, A.; Tiseanu, I.; Zoita, V.; JET-EFDA Contributors. A comparison of four reconstruction methods for JET neutron and gamma tomography. *Nucl. Instrum. Methods Phys. Res. Sect. A* **2009**, *605*, 374–383. [[CrossRef](#)]
28. Craciunescu, T.; Murari, A. Geodesic distance on Gaussian manifolds for the robust identification of chaotic systems. *Nonlinear Dyn.* **2000**, *86*, 677–693. [[CrossRef](#)]
29. Amari, S.; Nagaoka, H. *Methods of Information Geometry*; Oxford University Press: Oxford, UK, 2000.
30. Murari, A.; Boutot, P.; Vega, J.; Gelfusa, M.; Moreno, R.; Verdoolaege, G.; de Vries, P.C. Clustering based on the geodesic distance on Gaussian manifolds for the automatic classification of disruptions. *Nucl. Fusion* **2013**, *53*, 033006. [[CrossRef](#)]

31. Dormido-Canto, S.; Vega, J.; Ramírez, J.M.; Murari, A.; Moreno, R.; López, J.M.; Pereira, A.; JET-EFDA Contributors. Development of an efficient real-time disruption predictor from scratch on JET and implications for ITER. *Nucl. Fusion* **2013**, *53*, 113001. [[CrossRef](#)]
32. Toma, A.; Karagrigoriou, A.; Trentou, P. Robust model selection criteria using pseudodistances. *Entropy* **2020**. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).