Highlights of Spanish Astrophysics XI, Proceedings of the XV Scientific Meeting of the Spanish Astronomical Society held on September 4–9, 2022, in La Laguna, Spain. M. Manteiga, L. Bellot, P. Benavidez, A. de Lorenzo-Cáceres, M. A. Fuente, M. J. Martínez, M. Vázquez- Acosta, C. Dafonte (eds.), 2023

Fully Adaptive Bayesian Algorithm for Data Analysis. FABADA.

Sánchez-Alarcón, Pablo M.[1,2]**, and Ascasibar, Yago**[3]

[1] Instituto de Astrofísica de Canarias, c/ Vía Láctea s/n, E-38205, La Laguna, Tenerife, Spain
[2]Departamento de Astrofísica, Universidad de La Laguna, E-38206, La Laguna, Tenerife, Spain
[3]Departamento de Física Teórica, Universidad Autónoma de Madrid, E-28049 Madrid, Spain

Abstract

The discovery potential from astronomical and other data is limited by their noise. We introduce a novel non-parametric noise reduction technique based on Bayesian inference, `FABADA`, that automatically improves the signal-to-noise ratio of one- and two-dimensional data, such as astronomical images and spectra.
The algorithm iteratively evaluates possible smoothed versions of the data, the smooth models, estimating the underlying signal that is statistically compatible with the noisy measurements. Iterations stop based on the evidence and the χ^2 statistic of the last smooth model. We then compute the expected value of the signal as a weighted average of the whole set of smooth models. We evaluate its performance in terms of the peak signal to noise ratio using a battery of real astronomical observations.
Our Fully Adaptive Bayesian Algorithm for Data Analysis (`FABADA`) yields results that, without any parameter tuning, are comparable to standard image processing algorithms whose parameters have been optimized based on the true signal to be recovered, something that is impossible in a real application. On the other hand, state-of-the-art non-parametric methods, such as BM3D, offer a slightly better performance at high signal-to-noise ratio, while our algorithm is significantly more accurate for extremely noisy data, a situation usually encountered in astronomy. The source code of the implementation of the method, is publicly available at https://github.com/PabloMSanAla/fabada.

1 Introduction and Methodology

The acquisition of any experimental data is affected by several sources of statistical error, which ultimately translate into a random noise component in the measurements to be recorded. In astronomy, the noise introduced can sometimes be comparable to or even larger than the signal, and different image processing algorithms may be used to recover the information that is buried deep in the data.

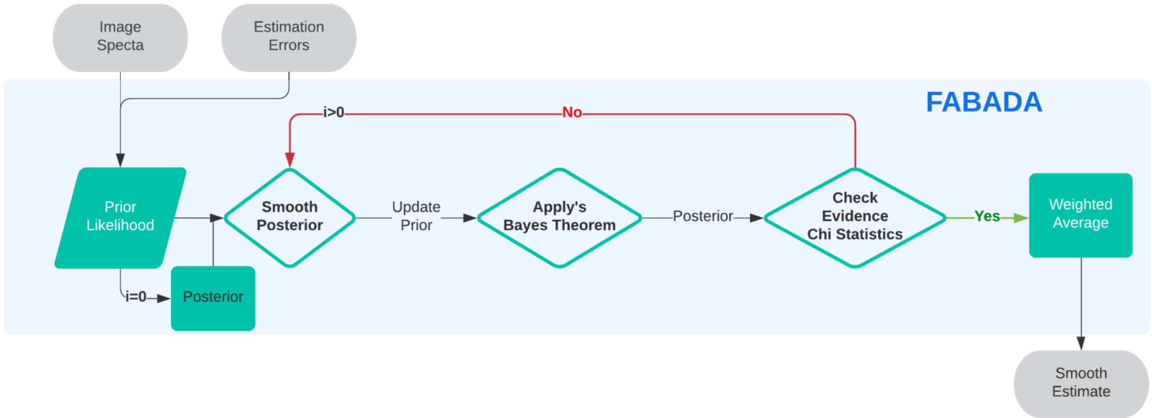

Figure 1: Visual layout of the Fully Adaptive Algorithm for Data Analysis.

The purpose of our Fully Adaptive Algorithm for Data Analysis is the estimation of the underlying signal in a way that is compatible with the measurements and the variance associated with it. Then, the inputs of our algorithm are the one- and two-dimensional data (such as astronomical spectra and images) and the variance associated with it (which can be heterogeneous). In Fig. 1 we show the layout of the method. `FABADA` is a novel non-parametric technique that reduces the noise in the image by applying Bayes' theorem in an iterative way. Therefore, we must define a suitable likelihood function to evaluate the models to be tested, and specify a prior probability distribution for the signal to initiate the process. Our likelihood is based on the statistics of a Gaussian process, and we start from an improper, constant prior. Then, we evaluate different smooth versions (by using the average of the surrounding points) of the posterior probabilities until a certain condition is reached. This new Bayesian technique incorporates an automatic selection criterion based on the statistical properties of the residuals, and therefore it yields a fully non-parametric method.

2 Astronomical Examples

In Fig. 2 we show two examples of the performance of `FABADA` in comparison with optimized standard algorithms (LOWESS [8], WIENER [7], and Gaussian Filter [3] and the state-of-the-art no parametric method, the Block-matching and 3D filtering, BM3D [5] of one-dimensional data (a-left) and two-dimensional images (b-right).

In figure 2-a we present the recoveries obtained for two random realizations with high, and low signal-to-noise ratios (SNR) for the spectrum of a pair of interacting galaxies, *Arp 256* [2]. At high SNR (left column), all algorithms display not only a similar performance, but actually converge to very similar solutions. All algorithms are able to correctly trace the presence of the most prominent emission lines, as well as the strong absorption line near the peak at the left end of the spectrum. Nevertheless, it is important to note that, while the

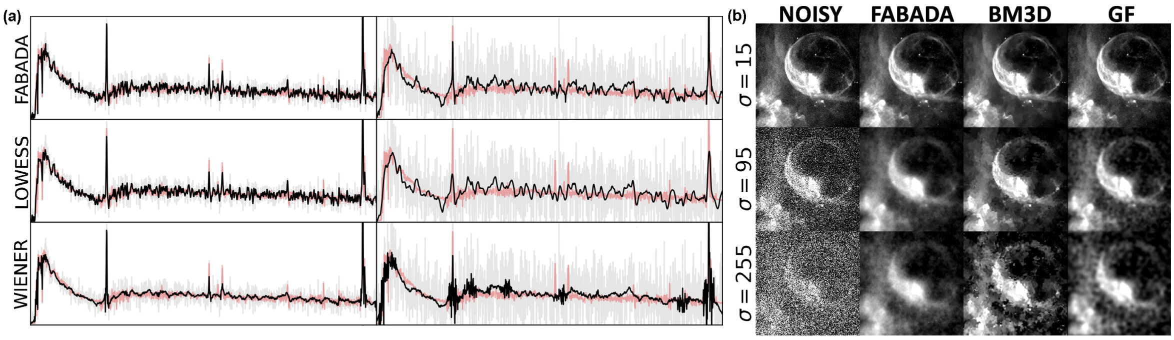

Figure 2: Examples of the recoveries obtained by FABADA on one-dimensional astronomical data (a-left) spectra, and two-dimensional images (b-right).

Wiener filter provides the best recovery in terms of to the overall noise reduction, `FABADA` tends to preserve the true intensity of these features slightly better than any of the other algorithms. This trend becomes more significant as the noise increases, and it is more difficult to discriminate significant spectral features from Gaussian random fluctuations. In the right panel, at low SNR, all models are able to reproduce the overall shape of the continuum. However, they fail to recover even the strongest absorption and emission lines, although hints of the brightest emission lines are still present in the standard methods. Only our prescription is able to provide a good description of these prominent features with this level of noise in the input data, albeit weaker absorption and emission lines are completely lost.

Similar trends are observed in the results obtained for the 2D data. Figure 2-b shows the recovery of the *Bubble* nebula image, taken from the Hubble Space Telescope gallery produced by NASA and the Space Telescope Science Institute (STScI). We compare the recovery with the state-of-the-art method (BM3D) and an optimized Gaussian Filter (GF) represented in each column, with three different noise levels ($\sigma = 15\,95,\ 255$ out of a dynamical range of 255) along the rows. All models yield fairly similar reconstructions for the highest signal-to-noise case ($\sigma = 15$). In particular, several edges in `FABADA` and GF seem to be a little bit more blurry compared to BM3D, together with a clearly visible *salt and pepper* noise component. Once again, the advantages of our algorithm become more evident as the noise increases. It is remarkable how the state-of-the-art BM3D method starts to add some artificial edges to the recovered image while `FABADA` is able to trace smoothly the global structure of the nebula. The GF can trace the structure but shows a preference for circular regions that can be related to the radius of the filters. These trends become more prominent if we push the noise even further, as is often the case in practical astrophysical applications, the signal itself is comparable to or even lower than the statistical uncertainties ($\sigma = 255$). While BM3D and the GF introduce some kind of artifacts, rectangular sharp gradients and circle regions, respectively, `FABADA` recovers the global structure in a smooth way.

3 Comparison Results

In Fig. 3 we show the results of the comparison between `FABADA`, the optimized standard methods, and the state-of-the-art BM3D method (only in 2D). We use the Peak Signal to Noise Ratio (PSNR) to evaluate the quality of the reconstruction by each algorithm. For simplicity we merge all the optimized standard methods (LOWESS, Savitzky–Golay filter [6], Gaussian filter, Wiener filter, and median filter) in one line (orange line), showing the region from the lowest to highest values of the metric obtained with this methods. Fig. 3-a (top panel) show the results for one-dimension for the three different spectra used in the comparison. The first, second and third column shows the results for a Kurucz stellar model [4], a emission line spectra from a supernova remnant SN132D [1], and for an interacting pair of galaxies, the pair group Arp256. Taking into account these results, one can see how `FABADA` performs as well as the best possible solutions of the standard methods typically used in astronomy.

Fig. 3-a (bottom panel) show the PSNR obtained for all the methods for different images. We consider eight different targets that are intended to sample the wide range of features that may be encountered in the field, including planets (Saturn), stars, diffuse nebulae (bubble, crab, ghost and eagle), and galaxies, either alone or in potentially blended groups (cluster). We can see similar behavior as the one-dimensional results. `FABADA` seems to offer the best solution possible obtained by the standard methods without the fine tuning of any parameter. However, in general, BM3D stands over the other methods, including `FABADA`, at high SNR ($\sigma \lesssim 95$ dB), in particular for the *Saturn* image. Its collaborative filter is particularly well suited for periodic data, or images with repetitive patterns, which are virtually absent in other test cases. The *stars* image would be a paradigmatic example, and the difference in this test is insignificant. On the other hand, `FABADA` perform better than BM3D at low signal-to-noise ratios.

4 Conclusions

In this work, we present the theory and implementation of a novel automatic algorithm for noise reduction: the Fully Adaptive Bayesian Algorithm for Data Analysis (`FABADA`). Our method iteratively evaluates progressively smoother models of the underlying signal and then combines them according to their Bayesian evidence and χ^2 statistic. The source code is publicly available at `https://github.com/PabloMSanAla/fabada`.

We compare `FABADA` with other methods that are representative of the current state of the art in image analysis and digital signal processing. For this comparison, we used the most typical metrics, the Peak-Signal-to-Noise-Ratio (PSNR), which is a measure of the Mean Square Error (MSE). One important advantage of our method, shared by BM3D over classical algorithms is the absence of free parameters to be tuned by the user. Our results suggest that `FABADA` and BM3D achieve values of PSNR comparable to or better than the best possible solution attainable by the classical methods.

Beyond the precise values of the global quantitative metrics, both `FABADA` and BM3D are

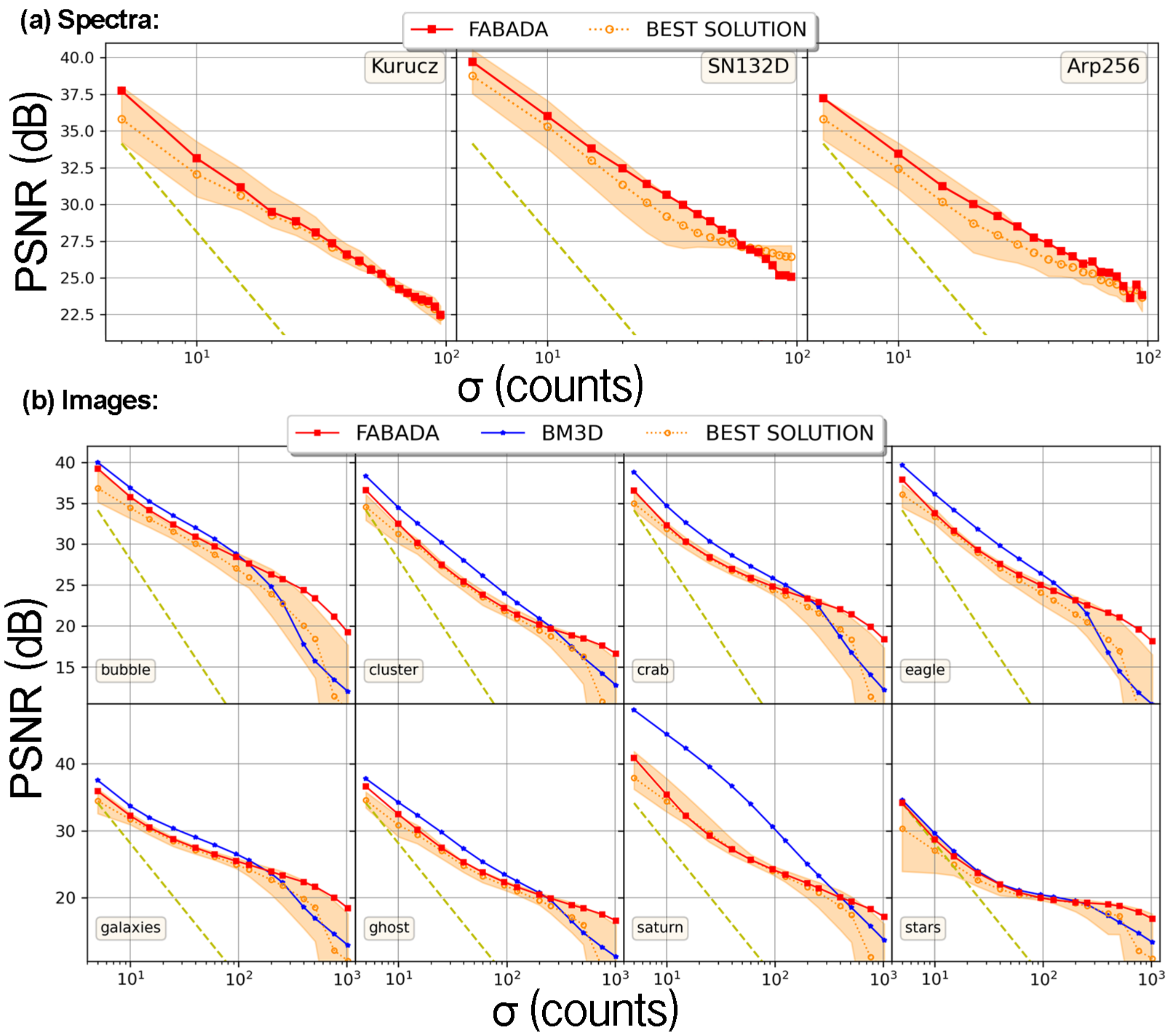

Figure 3: Results of the PSNR comparison between standard and state-of-the-art algorithms with respect to FABADA on one-dimensional astronomical data (a-top) spectra, and two-dimensional images (b-bottom).

quite successful in adapting to the structures present in the input data. Perhaps the most significant difference between them is that `FABADA`'s priors assume that the signal is smooth, whereas BM3D uses block-matching to look for repetitive patterns. This might be relevant when one must recover the height and shape of the spectral features in 1D or the gradients and boundaries in 2D. We argue that `FABADA` appears to offer a trade-off between noise reduction, increasing the metric values significantly, in a way that is statistically compatible with the data, keeping significant features without introducing considerable artifacts.

Acknowledgments

PMSA acknowledge financial support from the State Research Agency (AEI-MCINN) of the Spanish Ministry of Science and Innovation under the grant "The structure and evolution of galaxies and their central regions" with reference PID2019-105602GB-I00/10.13039/501100011033, from the ACIISI, Consejería de Economía, Conocimiento y Empleo del Gobierno de Canarias and the European Regional Development Fund (ERDF) under grant with reference PROID2021010044, and from IAC project P/300724, financed by the Ministry of Science and Innovation, through the State Budget and by the Canary Islands Department of Economy, Knowledge and Employment, through the Regional Budget of the Autonomous Community. Yago Ascasibar acknowledges financial support from grant "Starbursts throughout the evolution of the Universe" (PID2019-107408GB-C42/AEI/10.13039/501100011033) from the AEI-MICINN, Spain.

References

[1] Blair, W., Morse, J., Raymond, J., Kirshner, R., Hughes, J., Dopita, M., Sutherland, R., Long, K. & Winkler, P. Hubble Space Telescope Observations of Oxygen-rich Supernova Remnants in the Magellanic Clouds. II. Elemental Abundances in N132D and 1E 0102.2-7219. *Astrophysical Journal.* **537**, 667-689 (2000,7)

[2] Brown, M., Moustakas, J., Smith, J., Da Cunha, E., Jarrett, T., Imanishi, M., Armus, L., Brandl, B. & Peek, J. An Atlas of Galaxy Spectral Energy Distributions from the Ultraviolet to the Mid-infrared. *Astrophysical Journal, Supplement.* **212**, e18 (2014,6)

[3] Bromiley, P. Products and Convolutions of Gaussian Distributions. *Division Of Informatics, Imaging And Data Sciences, School Of Health Sciences, University Of Manchester.* pp. 1-2 (2003,1)

[4] Castelli, F. & Kurucz, R. New Grids of ATLAS9 Model Atmospheres. *Modelling Of Stellar Atmospheres.* **210** pp. A20 (2003,1)

[5] Dabov, K., Foi, A., Katkovnik, V. & Egiazarian, K. Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering. *IEEE Transactions On Image Processing.* **16**, 2080-2095 (2007)

[6] Savitzky, A. & Golay, M. Smoothing and Differentiation of Data by Simplified Least Squares Procedures.. *Analytical Chemistry.* **36**, 1627-1639 (1964),

[7] Lim, J. Two-dimensional signal and image processing. (Englewood Cliffs, N.J. ; Prentice Hall,1990)

[8] William S. Cleveland Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal Of The American Statistical Association.* **74**, 829-836 (1979),