



Open and Reproducible Science: Desirable or Even Mandatory, But Not So Simple!

In previous editorials, SPS President Athina Petropulu and I had the opportunity to say a few words about ethics, especially taking into account the usefulness of our research projects, for humanity and Earth, in a wide sense. In the current energy crisis and the explosion of costs, this issue becomes still more important, and I believe that it must be considered carefully in all our projects. Scientific integrity is another topic that I often discuss as it is actually a duty for all researchers for many of the reasons I developed in my November 2022 editorial [1].

In addition to ethics and scientific integrity, another important issue in research is related to open and reproducible science [2]. This implies that scientists must share datasets and codes when publishing new results. And reproducibility means that when using these data and codes and the attached information, any scientist should be able to reproduce the results, or to use the data and codes for conducting as-fair-as-possible benchmarks. Open and reproducible science is essential for ensuring confidence in scientific results, and more widely to help society feel more confident in science and scientists. But reproducibility is not evident. In a *Nature* article [3], Baker writes: “More than 70% of researchers have tried and failed to reproduce another scientist’s experiments, and more than half have failed to reproduce their own experi-

ments.” Those are some of the telling figures that emerged from *Nature*’s survey of 1,576 researchers who took a brief online questionnaire on reproducibility in research.

In this issue of *IEEE Signal Processing Magazine (SPM)*, Shenouda and Bajwa [A1] address the issue of reproducibility from a practical point of view and provide a set of recommendations for sharing data and codes efficiently for the purpose of reproducible research. Of course, this implies that you must first be able to reproduce your own results. This article explains the main pitfalls to achieving reproducible experiments and then provides common tools and techniques that can be used to overcome each of those pitfalls, bearing in mind that making experiments reproducible can entail extra effort that may divert attention away from our primary research task.

But, in addition to the issues discussed in this article, I believe that we must be aware of other replicability problems related to software and hardware architectures. The question of uncertainty in computing has been studied by computer scientists, explaining differences that can be obtained even for very simple codes when changing software versions, or when changing the hardware (e.g., running on a 32- or 64-bit processor) [4]. It has also been shown that these uncertainties are not independent. For instance, compile-time and runtime options can interplay on performance [5].

In many domains, e.g., in neuroimaging, datasets and codes are shared

by scientists from many countries. But processing the same data with different pipelines, or even with the same pipeline, by changing a few parameters (even just one), can provide very different results [6]. Other results show changes can appear with different operating systems, software packages, or workstation types [7], [8]. Because I think that these issues deserve to be published in *SPM*, I invited scientists working on computational uncertainty to write a tutorial for IEEE Signal Processing Society members.

In this issue

This issue of *SPM* is primarily comprised of the second part of the special issue on “Physics-Driven Machine Learning for Computational Imaging,” with nine articles detailed in [A2]. These articles consider a large variety of imaging, including optical imaging, tomographic imaging, hyperspectral image unmixing, magnetic resonance imaging, electromagnetic imaging, and terahertz computational imaging. Although the physics behind these various technics is very different, the main message to take home is the impact of physics-driven learning for designing simpler, faster, and more explainable methods, or the ability for providing, when few data are available, physics-driven synthetic data, which are relevant.

This issue also contains three column and forum articles in addition to [A1]. In [A3], Dolecek explores some

(continued on page 12)

than 30. The prize is US\$500 per author (up to a maximum of US\$1,500 per award) and a certificate. Eligibility is based on a four-year window preceding the year of election, and judging is based on general quality, originality, subject matter, and timeliness. Two Young Author Best Paper Awards are being presented this year:

- Ashutosh Pandey for the paper coauthored with DeLiang Wang, “A New Framework for CNN-Based Speech Enhancement in the Time Domain,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, July 2019.
- Qiuqiang Kong and Turab Iqbal for the paper coauthored with Yin Cao, Yuxuan Wang, Wenwu Wang, and Mark D. Plumbley, “PANNs: Large-Scale Pretrained Audio Neural

Networks for Audio Pattern Recognition,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, October 2020.

SPS members receive 2023 IEEE awards

The IEEE James L. Flanagan Speech and Audio Processing Technical Field Award will be presented to Alexander Waibel, “for pioneering contributions to spoken language translation and supporting technologies.”

The IEEE Fourier Award for Signal Processing will be presented to Rabab Kreidieh Ward, “for outstanding contributions to advancing signal processing techniques and their practical applications, and for technical leadership.”

IEEE has announced the recipients of the 2023 IEEE medals, which are

the highest-honor awards it presents. The medals will be given at the 2023 IEEE Honors Ceremony. Two SPS members have been awarded an IEEE medal for 2023:

- The IEEE Jack S. Kilby Signal Processing Medal for outstanding achievements in signal processing will be presented to José M.F. Moura, “for contributions to theory and practice of statistical, graph, and distributed signal processing.”
- The IEEE James H. Mulligan, Jr. Education Medal for outstanding contributions to education will be presented to James J. Truchard, “for the development of LabVIEW and establishing worldwide programs to enhance hands-on learning in laboratories and classrooms.”



FROM THE EDITOR *(continued from page 3)*

tips and tricks to decrease the number of additions per output sample in a cascaded integrator-comb multistage decimation filter. Two “Lecture Notes” focus on simple signal processing examples for understanding graph convolutional neural networks [A4] and making more explainable deep learning [A5]. Although these two articles use examples related to a simple linear filtering, for which we can wonder, what is the interest in using a nonlinear model, I think that these articles are interesting from a didactic point of view. Especially, in [A5], the same data (related to a two- or three-taps filter) are trained with four different neural architectures, all very simple. Although after training the different architectures achieve good fit of the filter, the explainability is not possible despite the network simplicity. Due to the black-box nature of the networks, even simple (six weights and three neurons for three of them), discussion clearly shows the impossibility of relating the weights of the network to the physical parameters of the filter. In the last part, the author suggests what is called a system-centric philosophy, which, in fact, suggests the use of some steps based on prior

knowledge of the system to learn. This is exactly the same philosophy as the one supported in all the articles of the special issue on “Physics-Driven Machine Learning for Computational Imaging.”

I wish everyone an enjoyable and rewarding read.

Appendix: Related Articles

- [A1] J. Shenouda and W. U. Bajwa, “A guide to computational reproducibility in signal processing and machine learning,” *IEEE Signal Process. Mag.*, vol. 40, no. 2, pp. 141–151, Mar. 2023, doi: 10.1109/MSP.2022.3217659.
- [A2] B. Wen, S. Ravishankar, Z. Zhao, R. Giryes, and J. C. Ye, “Physics-driven machine learning for computational imaging: Part 2,” *IEEE Signal Process. Mag.*, vol. 40, no. 2, pp. 13–15, Mar. 2023, doi: 10.1109/MSP.2023.3236492.
- [A3] G. J. Dolecek, “Update on the CIC multistage decimation filter with a minimum number of additions per output sample (APOS): Can we still decrease the number of APOS?” *IEEE Signal Process. Mag.*, vol. 40, no. 2, pp. 151–154, Mar. 2023, doi: 10.1109/MSP.2022.3216720.
- [A4] L. Stanković and D. Mandić, “Understanding the basis of graph convolutional neural networks via an intuitive matched filtering approach,” *IEEE Signal Process. Mag.*, vol. 40, no. 2, pp. 155–165, Mar. 2023, doi: 10.1109/MSP.2022.3207304.
- [A5] M. Narwaria, “Explainable machine learning – The importance of a system-centric perspective,” *IEEE Signal Process. Mag.*, vol. 40, no. 2, pp. 165–172, Mar. 2023, doi: 10.1109/MSP.2022.3211368.

References

- [1] C. Jutten, “Scientific integrity: A duty for researchers [From the Editor],” *IEEE Signal Process.*

Mag., vol. 39, no. 6, pp. 3–84, Nov. 2022, doi: 10.1109/MSP.2022.3198298.

[2] National Academies of Sciences, Engineering, and Medicine et al., *Reproducibility and Replicability in Science*. Washington, DC, USA: National Academy Press, 2019. [Online]. Available: <https://www.ncbi.nlm.nih.gov/books/NBK547532/>

[3] M. Baker, “1,500 scientists lift the lid on reproducibility,” *Nature*, vol. 533, no. 7604, pp. 452–454, May 2016, doi: 10.1038/533452a.

[4] S. Bernardi, M. Famelis, J.-M. Jézéquel, R. Mirandola, D. Perez Palacin, F. A. C. Polack, and C. Trubiani, “Living with uncertainty in model-based development,” in *Composing Model-Based Analysis Tools*, R. Heinrich, F. Durán, C. Talcott, and S. Zschaler, Eds. Cham, Switzerland: Springer-Verlag, 2021, pp. 159–185.

[5] L. Lesoil, M. Acher, X. TeRnava, A. Blouin, and J.-M. Jézéquel, “The interplay of compile-time and run-time options for performance prediction,” in *Proc. 25th ACM Int. Syst. Softw. Product Line Conf. (SPLC)*, Leicester, U.K., Sep. 2011, vol. A, pp. 100–111, doi: 10.1145/3461001.3471149.

[6] R. Botvinik-Nezer et al., “Variability in the analysis of a single neuroimaging dataset by many teams,” *Nature*, vol. 582, no. 7810, pp. 84–88, Jun. 2020, doi: 10.1038/s41586-020-2314-9.

[7] T. Glatard et al., “Reproducibility of neuroimaging analyses across operating systems,” *Frontiers Neuroinformatics*, vol. 9, Apr. 2015, Art. no. 12, doi: 10.3389/fninf.2015.00012.

[8] E. H. Gronenschild, P. Habets, H. I. L. Jacobs, R. Mengelers, N. Rozendaal, J. van Os, and M. Marcelis, “The effects of FreeSurfer version, workstation type, and Macintosh operating system version on anatomical volume and cortical thickness measurements,” *PLoS One*, vol. 7, no. 6, Jun. 2012, Art. no. e38234, doi: 10.1371/journal.pone.0038234.

