Astrostatistics in Canada and Beyond

Pauline Barmby (Western University),
Gwen Eadie (University of Toronto),
Gregory Sivakoff (University of Alberta),
David Stenning (Simon Fraser University)

2023 Oct 30–Nov 3

1 Overview of the Field

Astrostatistics is the development and application of statistical methods to measurements of the Universe beyond the Earth's atmosphere. As one of the oldest quantitative sciences, astronomy was once linked closely with statistics—for example, problems in Newtonian celestial mechanics were the driver for the development of the least-squares method [7]. In the last century, this link weakened as the field shifted from solely collecting astronomical observations to considering their astrophysical intepretation. In the past few decades, astrophysics and statistics have begun to re-approach one another [3, 4, 6].

Astrophysics has gone from being a data-poor field to one with an embarrassment of data riches, with complex, high-dimensional datasets that require sophisticated methods for statistical inference. Astronomy runs the gamut in terms of data type, including data products such as images, spatial data, time series data, and categorical data. Most of these data are collected through observational studies without the ability for repeated sampling or controlled experiments. Thus, characteristic challenges of astrophysics data include but are not limited to heteroscedastic measurement errors, truncated samples, incomplete data, and non-repeatability [8]. At the same time, issues in astrophysics models include parameter degeneracy, uncertainty quantification, over-simplification, and over-fitting [10]. The statistical community is becoming more aware of the interesting problems and opportunities provided by astrophysics, and the astronomy community is recognising the need for increased connections and collaborations with many different fields in statistics.

2 Recent Developments and Open Problems

The Canadian statistical and astrophysical communities have not worked together extensively in the past, but times are changing, with several joint faculty positions recently being filled and with the potential for interdisciplinary collaboration through grant programs of the Canadian Statistical Sciences Institute (CANSSI; e.g. the CANSSI Collaborative Research Team Projects). The explosive interest in big data is driving a need for the development of new statistical tools and approaches, and astrophysics datasets provide an ideal avenue for exploration and testing. Canadians are involved in many new projects that generate large quantities of data, including the Canadian Hydrogen Intensity Mapping Experiment (CHIME; [2]), the Legacy Survey of Space and Time (LSST; [9]) with the Vera C. Rubin Observatory, and the SKA Observatory [11] and its precursors. The massive datasets produced by these world-class facilities require statistical expertise to be most efficiently exploited for physical understanding.

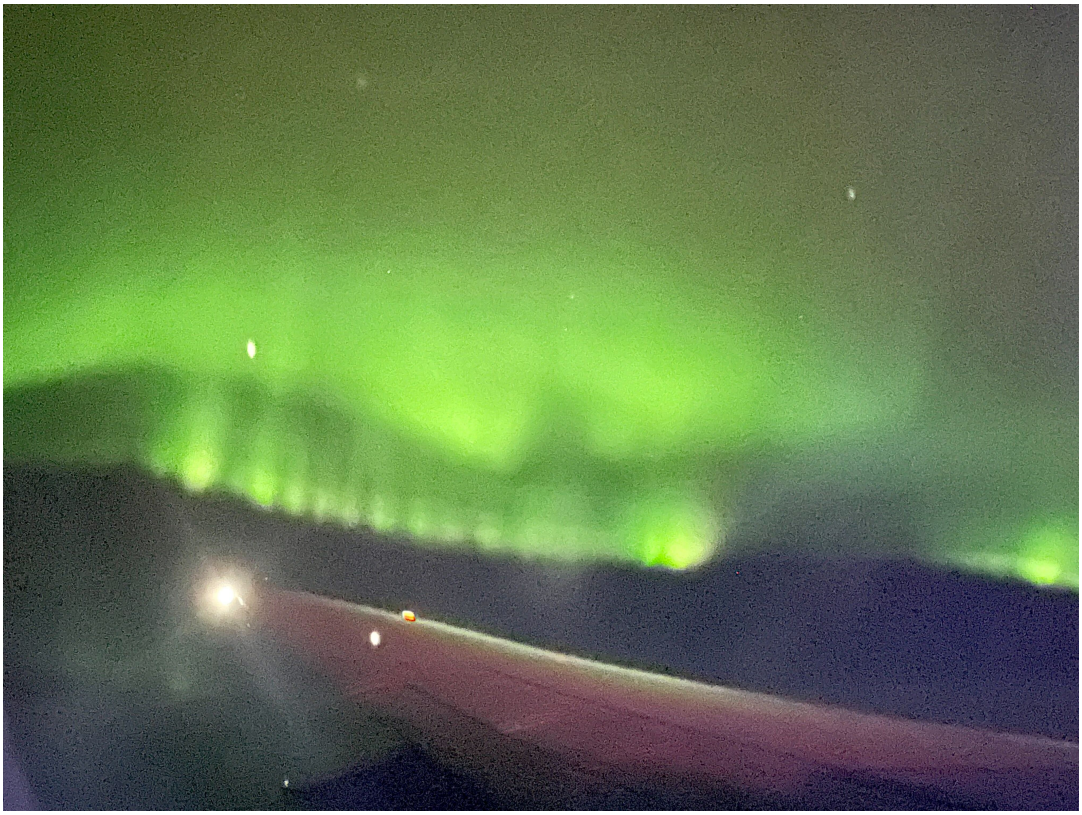

Figure 1: Airborne aurora borealis from "Astrostatistics in Canada and Beyond" return trip. Credit: D. Li (PhD student, University of Toronto)

The novelty of this workshop was both in the imminent arrival of massive datasets from new facilities and on the emphasis on what astrophysicists and statisticians can offer to one another as research partners. Unlike other recent efforts, the workshop focused not on a particular astrophysical application but on providing a broad introduction to the possibilities for collaboration, particularly between the Canadian communities. We provided structured space and time for groups of participants who do not normally meet to identify links between astronomy and statistics and start new collaborations and projects. Remote participation worked well for participants who were unable to travel for various reasons; one travel highlight was an amazing view of the aurora borealis from a participant's return flight to Toronto (Figure 1)!

3 Presentation Highlights

The first day of the meeting included four review talks, on the history of astrostatistics, big datasets in astronomy, statistical learning methods for big data, and astrostatistics with 'small data.' All meeting participants introduced themselves with 'lightning talks': 60-second talks using a single presentation slide. This format was new to the statistician participants, and enjoyed by all. The day finished with a panel discussion about funding opportunities for interdisciplinary projects, such as in astrostatistics.

All participants who wanted to present their research were able to do so: early-career researchers gave 20-minute talks on the mornings of day 2 and 4, and more senior researchers gave 10-minute talks on the morning of day 3. Most talks were in person, but several participants appreciated the opportunity to present and attend remotely; the Zoom hybrid meeting technology worked very well for presentations. The ECR talks in particular were all very well done and well-received by both astronomers and statisticians. There was a wide range of presentation topics, from fast radio bursts [5] to stratified learning [1], solar flares [12] and copulas [13], with slides available at `bit.ly/birs-astro-talks`. Two presentations later led

to an impromptu self-directed discussion on copulas in astronomy. Notably, two ECR presentations sparked productive afternoon discussions between astronomers and statisticians about statistical methods for fast radio bursts and galaxies. The morning of day 5 featured summary talks from two of the organizers ("Statistics" Summary Talk from an "Astronomy" Point-of-View and vice versa) and a workshop-wide discussion of next steps (see Section 5).

4 Scientific Progress Made

The afternoons of days 2 and 4 were spent in self-organized discussions: from a list of topics, participants voted on the ones they were most interested in, and broke into smaller groups for discussions. Some sessions were repeated so that participants could discuss multiple topics. Discussion topics included:

- Simulation-based and likelihood free inference
- MCMC / Bayesian
- Noisy / biased / incomplete data
- Electromagnetic transients
- Data handling
- Astronomical surveys
- Probabilistic catalogues
- Astrostatistics resources

For most of the discussions summary notes were taken and shared with all participants via Google Docs. These notes included links to relevant projects, publications, and tutorials as well as records of the discussion. The shared meals and excursions also led to several small group impromptu discussions about individual students' work, which were very productive and led to initiation of new collaborations. Most remote participants indicated in advance that they would likely not participate in these discussions; we were able to arrange Zoom participation for those who were interested.

5 Outcome of the Meeting

Participants were polled for their feedback during the closing session using Mentimeter. They reported becoming more familiar with the terminology of either statistics or astronomy (Figure 2) and appreciated the chance to meet researchers from other fields and learn from them (Figure 3).

The 'next steps' discussion on the workshop's last day led to a host of ideas for further developing astrostatistics in Canada and beyond. Joining existing professional groups, such as the International Astrostatistics Association (IAA), the Astrostatistics Interest Group (AIG) of the American Statistical Association, or the Working Group on Astroinformatics and Astrostatistics (WGAA) of the American Astronomical Society, is one way to keep in touch with the community. The idea of creating a Canadian group such as a chapter of the Statistical Society of Canada, or a mailing list of interested researchers, was also discussed. The workshop Slack channel workspace is another possible forum, although like many conference Slacks it has seen little use after the meeting. Maintaining a more permanent web presence for the Canadian astrostatistics community, including a list of funding opportunities, would be desirable: one possibility is to partner with one of the existing groups listed above.

Potential follow-on activities include a future BIRS workshop, perhaps focused on explainability and expanded to include researchers in machine learning and computer science, or taking advantage of the American Astronomical Society's "Meeting in a Meeting" structure through WGAA. Existing and future astrostatistics training opportunities were also discussed. Workshops already exist in Italy, Spain, Chile, USA (Penn State), Greece and Germany; adding a workshop to a future Canadian Astronomical Society meeting, or pursuing funding for a wider interdisciplinary training program through NSERC's CREATE program, are other

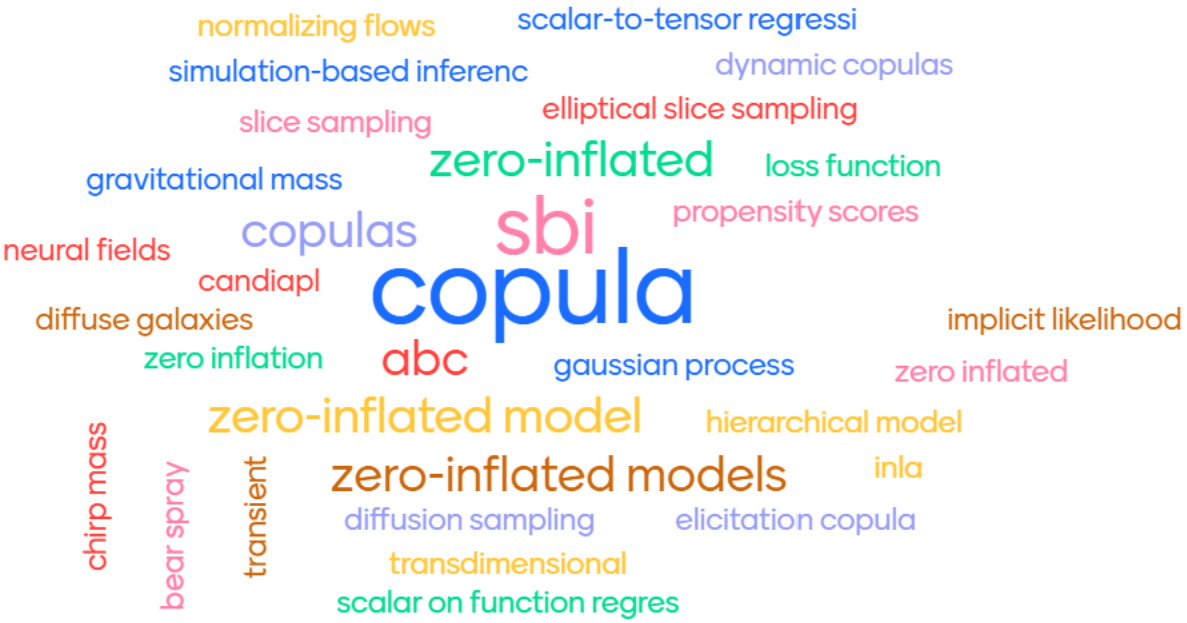

Figure 2: Terminology and lingo participants learned at the workshop

possibilities. Co-supervision of graduate students is another, smaller-scale way to pursue interdisciplinary training. Writing community papers (also called white papers) that promote astrostatistics for the development of human capital, guides to documentation, or updates to the dictionaries maintained by the International CHASC Astro-Statistics Collaboration (`https://hea-www.harvard.edu/astrostat/` are further examples of possible future activities for interested participants.

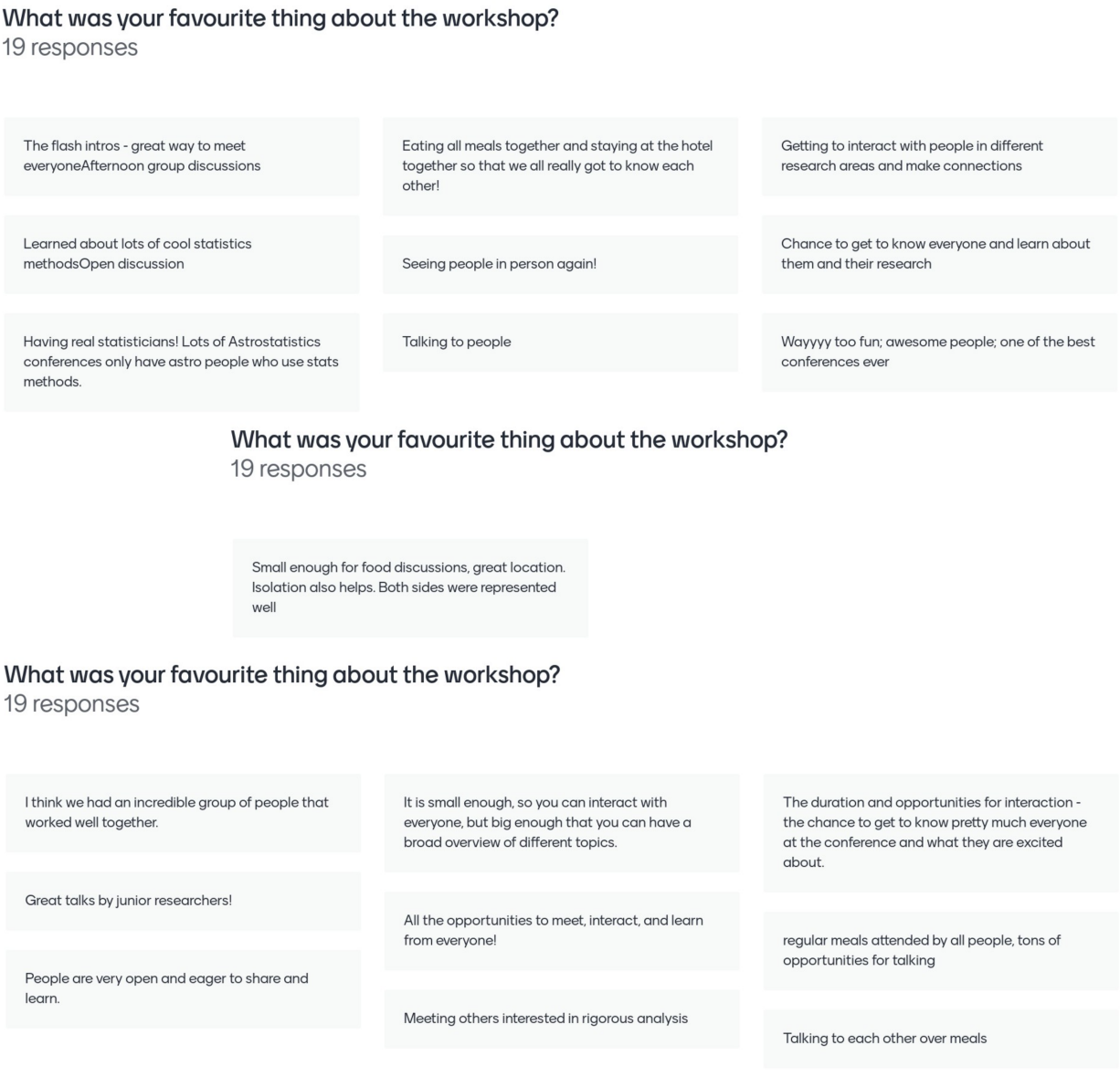

Figure 3: Participant feedback gathered with the Mentimeter platform during summary talk

References

[1] M. Autenrieth, D.A. van Dyk, R. Trotta & D. Stenning, Stratified Learning: A General-Purpose Statistical Method for Improved Learning under Covariate Shift, *Statistical Analysis and Data Mining*, 116, 2023.

[2] K. Bandura et al., Canadian Hydrogen Intensity Mapping Experiment (CHIME) pathfinder, in SPIE *Ground-based and Airborne Telescopes*, vol. 9145, 2014.

[3] G.J. Babu & E.D. Feigelson, Astrostatistics, Chapman & Hall, 1996

[4] G.J. Babu & E.D. Feigelson, Statistical Challenges in Modern Astronomy IV, Astro. Soc. Pacific, 2007

[5] A.M. Cook et al., An FRB Sent Me a DM: Constraining the Electron Column of the Milky Way Halo with Fast Radio Burst Dispersion Measures from CHIME/FRB, *The Astrophysical Journal*, 946, 2, 2023

[6] G. Eadie et al., Astrostatistics in Canada. In Canadian Long Range Plan for Astronomy and Astrophysics White Papers, 2019.

[7] E. Feigelson, Astrostatistics. In *Encyclopedia of Mathematics*, EMS Press, 2016.

[8] E.D. Feigelson, R.S. de Souza, E.E.O. Ishida, & G.J. Babu, 21st Century Statistical and Computational Challenges in Astrophysics, *Annual Review of Statistics and Its Application*, vol. 8, pp. 493517, 2021

[9] R. Hložek et al., Science with the Large Synoptic Survey Telescope. Canadian Long Range Plan for Astronomy and Astrophysics White Papers, 2019.

[10] C.M. Schafer, A Framework for Statistical Inference in Astrophysics *Annual Review of Statistics and Its Application* 2, 141162, 2015

[11] K. Spekkens et al., Canada and the SKA from 2020–2030. Canadian Long Range Plan for Astronomy and Astrophysics White Papers, 2019.

[12] H. Sun, W. Manchester, M. Jin, Y. Liu, and Y. Chen, Tensor Gaussian process with contraction for multi-channel imaging analysis. In *Proceedings of the 40th International Conference on Machine Learning*, 2023.

[13] R. Zimmerman, R.V. Craiu & V. Leos-Barajas, Copula Modeling of Serially Correlated Multivariate Data with Hidden Structures, *Journal of the American Statistical Association*, 2023.