On Collaboration

Nancy Reid
University of Toronto

Concordia University
November 21, 2018

www.incass.ca

www.canssi.ca

CANSSI
INCASS

Looking Back – CANSSI 2014

Vision

A growing and strongly engaged research force in the statistical sciences in Canada, working on problems at the leading edge of the discipline, collaborating across disciplines and internationally, and tackling issues of importance to science and society

Mission

To promote statistical sciences research and training in Canada and internationally, through its role as a member as a member of the Canadian network of mathematical and statistical sciences institutes

NPCDS 2002 – 2008

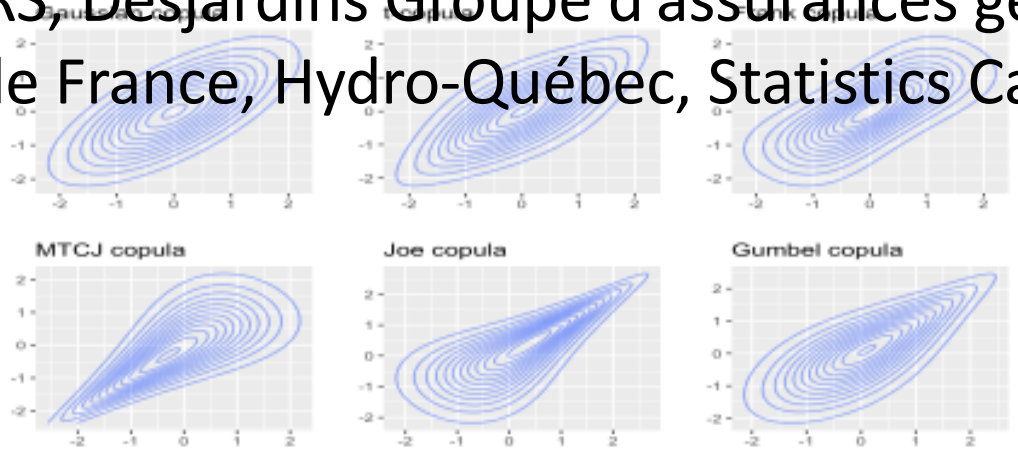
- Proactive partnerships with scientists in other disciplines
- Building interdisciplinary research capacity
- A new generation of statistical scientists willing to lead national research collaborations
- Creation of interdisciplinary teams with statistical leadership

Why CANSSI?

- Ongoing demand for statistical expertise from scientists in many disciplines
- Funding for fundamental research largely confined to single disciplines
- Very small grants for statistical scientists

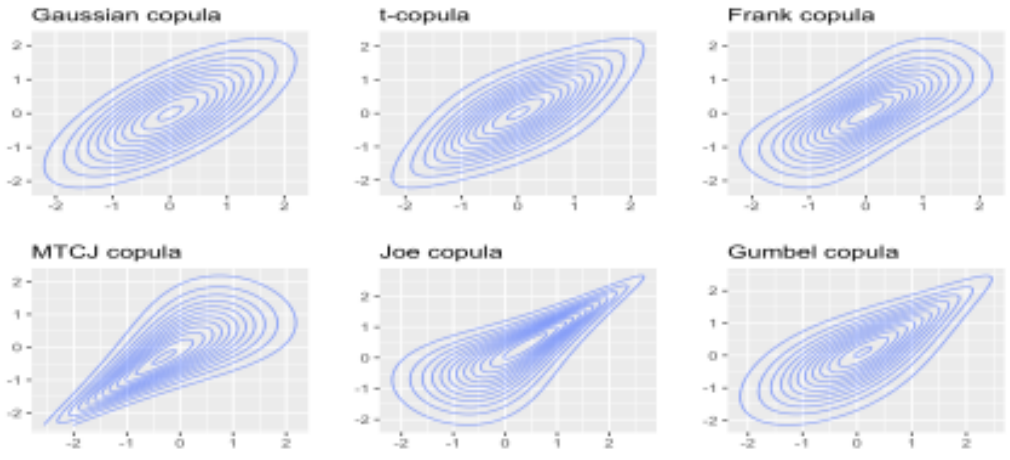
Early successes

- Copula dependence modelling – theory and applications
- Statisticians – McGill, Laval, UQTR, HEC, Toronto, UBC, Manitoba
- Collaborators – Banque de Montréal, Banque Nationale du Canada, INRS, Desjardins Groupe d'assurances générales, Électricité de France, Hydro-Québec, Statistics Canada



Team 1 – Copula dependence modelling

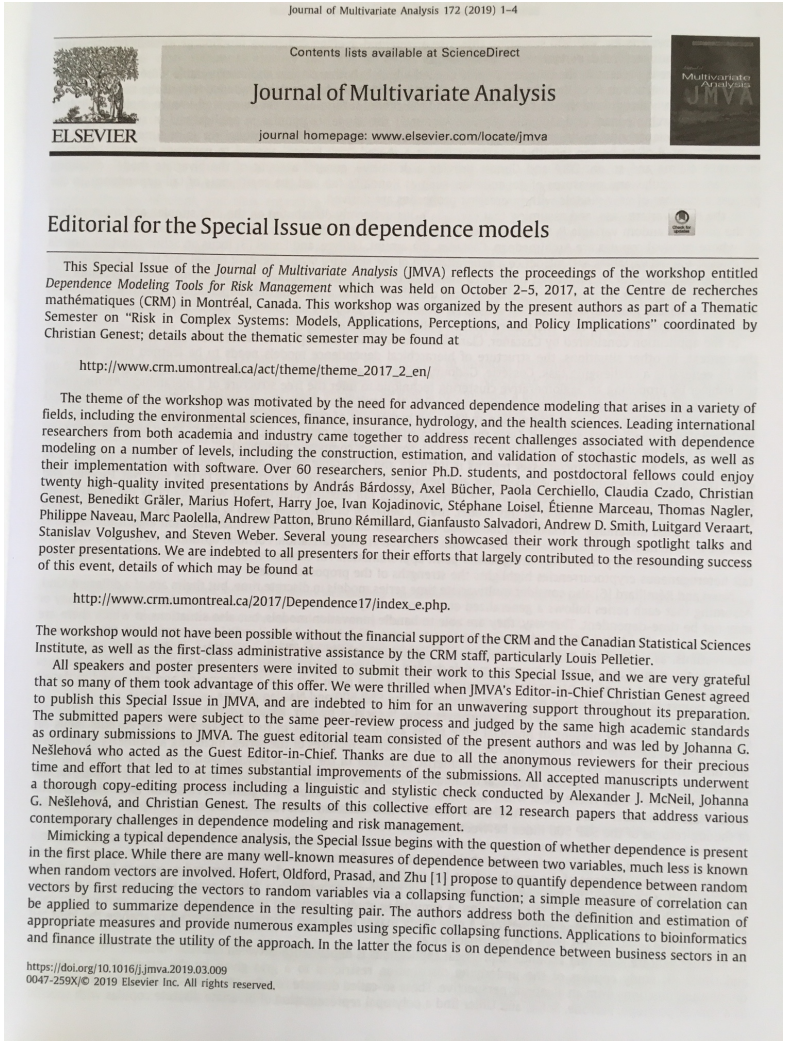
- 18 MSc students; 18 PhD students
- 40 national and international presentations
- 28 publications



2017

- Environmental Risk and Extreme Events • Regulatory Issues in Business • Control of Systemic Risk • Tools for Quantitative Risk Management • Risk in Health Sciences

2019



Journal of Multivariate Analysis 172 (2019) 1–4

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Editorial for the Special Issue on dependence models

This Special Issue of the *Journal of Multivariate Analysis* (JMVA) reflects the proceedings of the workshop entitled *Dependence Modeling Tools for Risk Management* which was held on October 2–5, 2017, at the Centre de recherches mathématiques (CRM) in Montréal, Canada. This workshop was organized by the present authors as part of a Thematic Semester on "Risk in Complex Systems: Models, Applications, Perceptions, and Policy Implications" coordinated by Christian Genest; details about the thematic semester may be found at

http://www.crm.umontreal.ca/act/theme/theme_2017_2_en/

The theme of the workshop was motivated by the need for advanced dependence modeling that arises in a variety of fields, including the environmental sciences, finance, insurance, hydrology, and the health sciences. Leading international researchers from both academia and industry came together to address recent challenges associated with dependence modeling on a number of levels, including the construction, estimation, and validation of stochastic models, as well as their implementation with software. Over 60 researchers, senior Ph.D. students, and postdoctoral fellows could enjoy twenty high-quality invited presentations by András Bárdossy, Axel Bücher, Paola Cerchiello, Claudia Czado, Christian Genest, Benedikt Gräler, Marius Hofert, Harry Joe, Ivan Kojadinovic, Stéphane Loisel, Étienne Marceau, Thomas Nagler, Philippe Naveau, Marc Paolella, Andrew Patton, Bruno Rémillard, Gianfausto Salvadori, Andrew D. Smith, Luitgard Veraart, Stanislav Volgushev, and Steven Weber. Several young researchers showcased their work through spotlight talks and poster presentations. We are indebted to all presenters for their efforts that largely contributed to the resounding success of this event, details of which may be found at

http://www.crm.umontreal.ca/2017/Dependence17/index_e.php.

The workshop would not have been possible without the financial support of the CRM and the Canadian Statistical Sciences Institute, as well as the first-class administrative assistance by the CRM staff, particularly Louis Pelletier.

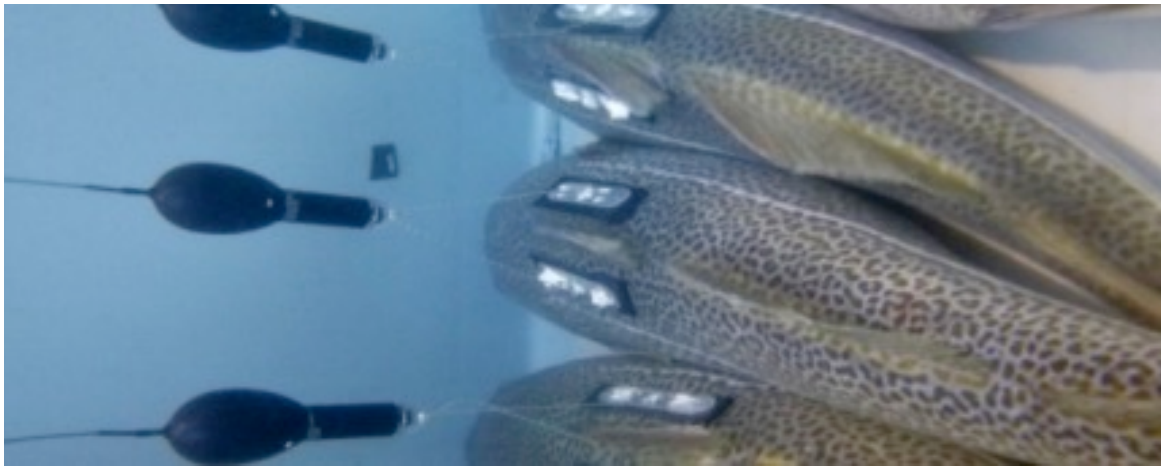
All speakers and poster presenters were invited to submit their work to this Special Issue, and we are very grateful that so many of them took advantage of this offer. We were thrilled when JMVA's Editor-in-Chief Christian Genest agreed to publish this Special Issue in JMVA, and are indebted to him for an unwavering support throughout its preparation. The submitted papers were subject to the same peer-review process and judged by the same high academic standards as ordinary submissions to JMVA. The guest editorial team consisted of the present authors and was led by Johanna G. Nešlehová who acted as the Guest Editor-in-Chief. Thanks are due to all the anonymous reviewers for their precious time and effort that led to at times substantial improvements of the submissions. All accepted manuscripts underwent a thorough copy-editing process including a linguistic and stylistic check conducted by Alexander J. McNeil, Johanna G. Nešlehová, and Christian Genest. The results of this collective effort are 12 research papers that address various contemporary challenges in dependence modeling and risk management.

Mimicking a typical dependence analysis, the Special Issue begins with the question of whether dependence is present in the first place. While there are many well-known measures of dependence between two variables, much less is known when random vectors are involved. Hofert, Oldford, Prasad, and Zhu [1] propose to quantify dependence between random vectors by first reducing the vectors to random variables via a collapsing function; a simple measure of correlation can be applied to summarize dependence in the resulting pair. The authors address both the definition and estimation of appropriate measures and provide numerous examples using specific collapsing functions. Applications to bioinformatics and finance illustrate the utility of the approach. In the latter the focus is on dependence between business sectors in an

https://doi.org/10.1016/j.jmva.2019.03.009
0047-259X/

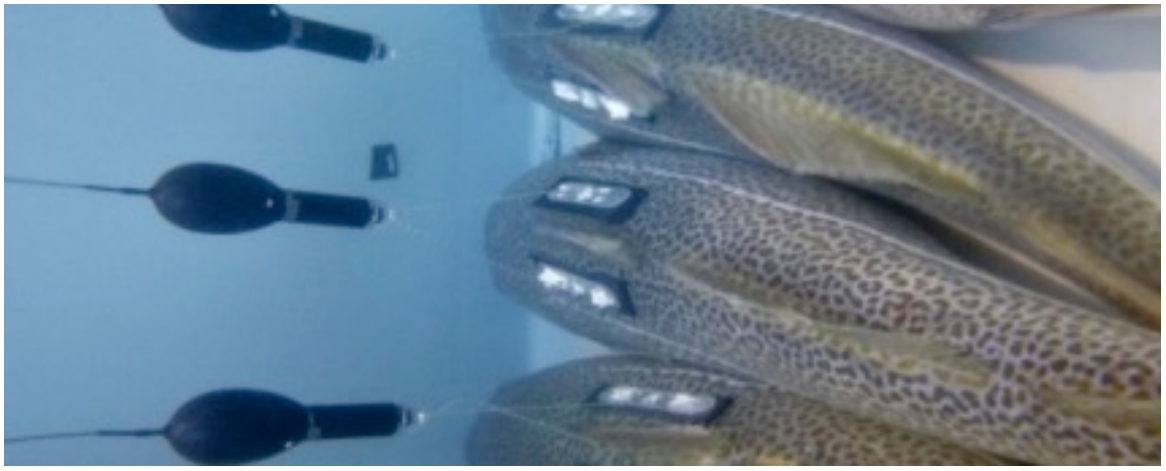
Early successes

- Advancements to State-Space Models for Fisheries Science
- Statisticians – Dalhousie, Memorial, SFU, CH, DK, NO, China
- Collaborators – Carleton, DFO, Nw Fisheries Science, Dalhousie, AU
- Aaron MacNeil recruited as CRC Chair in Fisheries Science



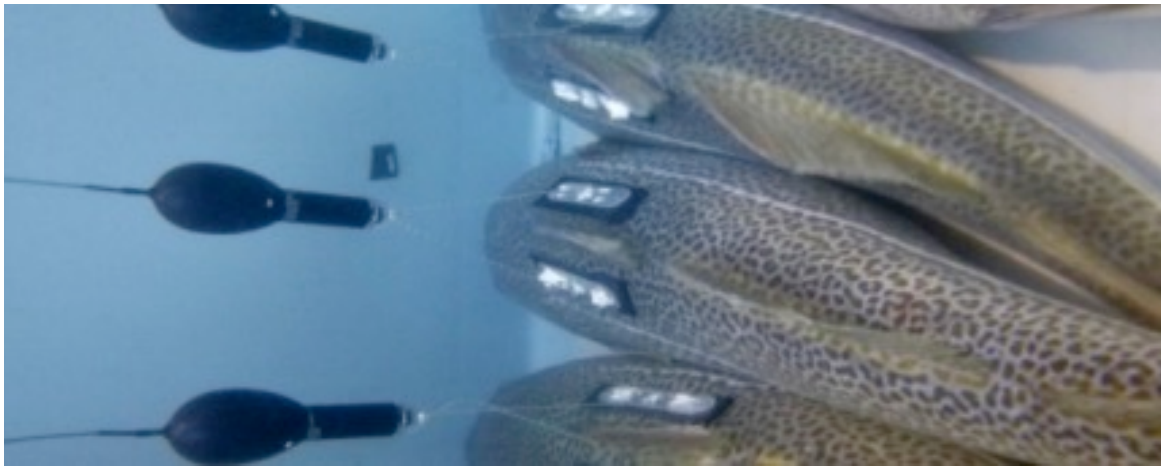
Team 3 – Fisheries Science

- Two-way dialogue with Fisheries and Oceans Canada (DFO)
- Proposal for short course on stock assessment
- Joint supervisions of graduate students
- Team members participating in DFO meetings



Team 3 – Fisheries Science

- Two-way dialogue with Fisheries and Oceans Canada (DFO)
- Proposal for short course on stock assessment
- Joint supervisions of graduate students
- Team members participating in DFO meetings



2018

- CRT 10 – Towards Sustainable Fisheries
- Partnership with Ocean Tracking Network, Ocean Frontiers Institute

2019

OCEAN FRONTIER INSTITUTE

IN THE NEWS

New study by OFI researcher provides a potential solution to malnu

A new paper by OFI researcher Aaron MacNeil published in the journal Nature demons

In a recent interview Dr. MacNeil discussed the global challenge of health and micron
of some local communities.

You can find the full interview here, or check out the research paper here.

nature

Letter | Published: 25 September 2019

Harnessing global fisheries to tackle micronutrient deficiencies

Christina C. Hicks, Philippa J. Cohen, Nicholas A. J. Graham, Kirsty L. Nash, Edward H. Allison, Coralie D'Lima, David J. Mills, Matthew Roscher, Shakuntala H. Thilsted, Andrew L. Thorne-Lyman & M. Aaron MacNeil

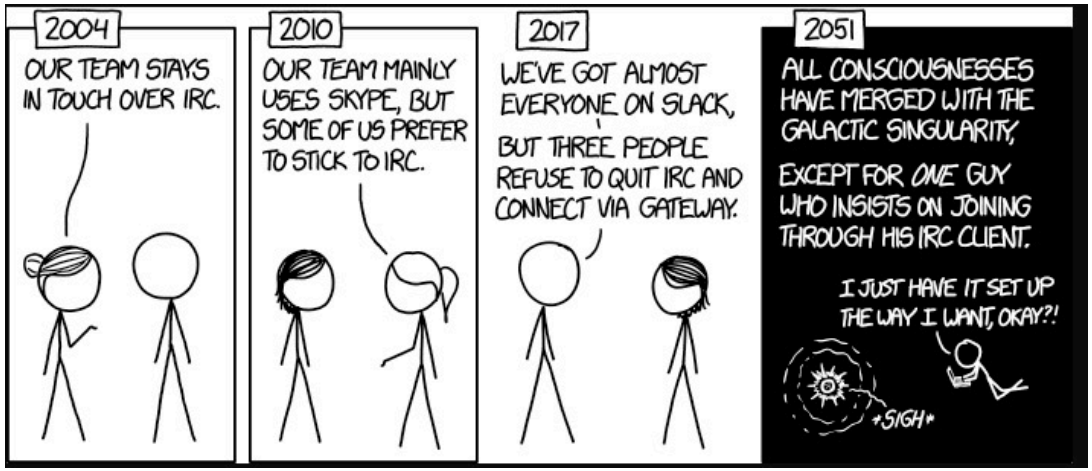
Nature **574**, 95–98(2019) | Cite this article

7112 Accesses | **4** Citations | **614** Altmetric | Metrics

CANSSI INCASS

Why CANSSI?

- Collaboration is Difficult
  - Unpredictable
  - Takes time
  - Needs a lot of luck
  - High risk research
- Collaboration is Essential
  - Science is specialized
  - Statistical science is specialized



Why CANSSI?

- Aren't most scientists collaborating already?
  - Of course
  - Serendipity, mutual interests, scientific curiosity, love to travel
- Interdisciplinary collaborations are difficult to fund
  - Higher risk, especially if they are new
  - Funding is limited
  - Needs some incentives

14/24
09/19

CANSSI
INCASS

National Science Foundation
WHERE DISCOVERIES BEGIN

Other Special Research Programs Available to DMS Communities

- Harnessing the Data Revolution (HDR): Institutes for Data-Intensive Research in Science and Engineering - Ideas Labs (I-DIRSE-IL) N
- Harnessing the Data Revolution (HDR): Institutes for Data-Intensive Research in Science and Engineering - Frameworks (I-DIRSE-FW) N
- Harnessing the Data Revolution (HDR): Data Science Corps (DSC) N
- Secure and Trustworthy Cyberspace (SaTC)
- Cyberinfrastructure for Sustained Scientific Innovation (CSSI): C
- Computational and Data-Enabled Science and Engineering (CDS&E)
- Enabling Quantum Leap: Quantum Idea Incubator for Transformational Advances in Quantum Systems (QII - TAQS) N
- Quantum Leap Challenge Institutes (QLCI) N
- Connections in Quantum Information Science (CQIS) C
- Designing Materials to Revolutionize and Engineer our Future (DMREF)
- Enabling Quantum Leap: Convergent Accelerated Discovery Foundries for Quantum Materials Science, Engineering and Information (Q-AMASE-i) C
- Understanding the Rules of Life: Epigenetics N
- Understanding the Rules of Life: Building a Synthetic Cell N
- Collaborative Research in Computational Neuroscience (CRCNS)
- Cracking the Olfactory Code (Olfactory) C
- Innovations at the Nexus of Food, Energy and Water Systems (INFEWS)
- EMERGING FRONTIERS IN RESEARCH AND INNOVATION 2020
- GROWING CONVERGENCE RESEARCH (GCR) N
- NSF-NIST Interaction in Basic and Applied Scientific Research in BIO, ENG & MPS (NSF-NIST)
- Secure and Trustworthy Cyberspace Frontiers (SaTC Frontiers)

CANSSI INCASS

Interdisciplinary Collaboration can be rocket science

rsistent link using digital object identifier

ELSEVIER

Physics Letters B

Volume 800, 10 January 2020, 135069

Searches for lepton-flavour-violating decays of the Higgs boson in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector

The ATLAS Collaboration *

⊞ Show more

https://doi.org/10.1016/j.physletb.2019.135069

Interdisciplinary Collaboration can be rocket science

7. Statistical analysis

The searches for $H \to e\tau$ and $H \to \mu\tau$ are treated independently. For each search, the analysis exploits the four signal regions and the two control regions specified in Table 2. The BDT score distributions of all signal regions are analysed to test the presence of a signal, simultaneously with the event yields in control regions, which are included to constrain the normalizations of the major backgrounds estimated from simulation. The statistical analysis uses a binned likelihood function $\mathcal{L}(\mu, \theta)$, constructed as a product of Poisson probability terms over all bins considered in the search. This function depends on the parameter μ, defined as the branching ratio $\mathcal{B}(H \to \ell\tau)$, and a set of nuisance parameters θ that encode the effect of systematic uncertainties in the signal and background expectations. All nuisance parameters are implemented in the likelihood function as Gaussian or log-normal constraints. The normalization factors of the single-top-quark and $t\bar{t}$ backgrounds in the $\ell\tau_{\ell'}$ channel and of the $Z \to \tau\tau$ background component are unconstrained parameters of the fit. Estimates of the parameters of interest are calculated with the profile-likelihood-ratio test statistic $\tilde{q}_\mu$ [89], and the upper limits on the branching ratios are derived by using $\tilde{q}_\mu$ and the CL_S method [90].

Interdisciplinary Collaboration can be rocket science

7. Statistical analysis

The searches for $H \to e\tau$ and $H \to \mu\tau$ are treated indepen-
alysis exploits the four signal regions
ecified in Table 2. The BDT score dis-
are analysed to test the presence of
the event yields in control regions,
rain the normalizations of the ma-
m simulation. The statistical analysis
ion $\mathcal{L}(\mu, \theta)$, constructed as a product
ver all bins considered in the search.
parameter μ, defined as the branch-
et of nuisance parameters θ that en-
uncertainties in the signal and back-
ground expectations. All nuisance parameters are implemented in the likelihood function as Gaussian or log-normal constraints. The normalization factors of the single-top-quark and $t\bar{t}$ backgrounds in the $\ell\tau_{\ell'}$ channel and of the $Z \to \tau\tau$ background component are unconstrained parameters of the fit. Estimates of the parameters of interest are calculated with the profile-likelihood-ratio test statistic $\tilde{q}_\mu$ [89], and the upper limits on the branching ratios are derived by using $\tilde{q}_\mu$ and the CL_S method [90].

8. Results

The best-fit branching ratios and upper limits are computed while assuming $\mathcal{B}(H \to \mu\tau) = 0$ for the $H \to e\tau$ search and $\mathcal{B}(H \to e\tau) = 0$ for the $H \to \mu\tau$ search. The best-fit values of the LFV Higgs boson branching ratios are equal to $(0.15^{+0.18}_{-0.17})\%$ and $(-0.22 \pm 0.19)\%$ for the $H \to e\tau$ and $H \to \mu\tau$ search, respectively.

Interdisciplinary Collaboration can be rocket science

The Annals of Applied Statistics
2019, Vol. 13, No. 3, 1884–1926
https://doi.org/10.1214/19-AOAS1258

APPROXIMATE INFERENCE FOR CONSTRUCTING ASTRONOMICAL CATALOGS FROM IMAGES[1]

BY JEFFREY REGIER*, ANDREW C. MILLER†, DAVID SCHLEGEL‡, RYAN P. ADAMS§, JON D. MCAULIFFE*,¶ AND PRABHAT‡

*University of California, Berkeley**, *Columbia University*†, *Lawrence Berkeley National Laboratory*‡, *Princeton University*§ *and The Voleon Group*¶

We present a new, fully generative model for constructing astronomical catalogs from optical telescope image sets. Each pixel intensity is treated as a random variable with parameters that depend on the latent properties of stars and galaxies. These latent properties are themselves modeled as random. We compare two procedures for posterior inference. One procedure is based on Markov chain Monte Carlo (MCMC) while the other is based on variational inference (VI). The MCMC procedure excels at quantifying uncertainty, while the VI procedure is 1000 times faster. On a supercomputer, the VI procedure efficiently uses 665,000 CPU cores to construct an astronomical catalog from 50 terabytes of images in 14.6 minutes, demonstrating the scaling characteristics necessary to construct catalogs for upcoming astronomical surveys.

Interdisciplinary Collaboration can be rocket science

- Our first contribution is a statistical model that can simultaneously find centroids, determine photometry, deblend overlapping light sources, perform star/galaxy separation and adjust estimates of all quantities based on prior information
- Our second contribution is to develop and compare two approximate posterior inference procedures for our model: one based on MCMC and the other based on VI

Interdisciplinary Collaboration isn't always rocket science

Interdisciplinary Collaboration isn't always rocket science

"In some of the individual studies, people who owned dogs were on average about eight years younger than people who didn't"

Oops

Interdisciplinary Collaboration isn't always rocket science

A limitation of our meta-analysis is that it was not possible to perform an analysis adjusted for confounders of the association between dog ownership and the outcomes.

Google owning a dog reduces risk of death

All News Images Videos Shopping More Settings Tools

Want to live longer? Try getting a **dog**.

Washington Post - Oct. 9, 2019

"**Dog ownership**," the authors conclude, "is associated with **lower** the long term, which is possibly driven by a reduction in ...

Dog ownership linked to longer life in two new American Heart ...

CBS News - Oct. 8, 2019

Dog People Live Longer. But Why?
NPR - Oct. 26, 2019
We've known forever that **owning a dog** increases one's physical activity ... studies suggested that **dog ownership reduces** the **risk of death** from ...

American Heart Association news: **Dogs** boost cardiovascular ...
dvm360 - Nov. 13, 2019
The **lower risk of death** associated with **dog ownership** could be explained by increased physical activity and decreased depression and ...

Do Purebreds (But Not Mutts) **Reduce Dog** Owner **Death** Rates?
Psychology Today (blog) - Oct. 31, 2019
Indeed, **owning a dog** was associated with an impressive 21% **lower risk of death** among the heart attack victims and an 18% **lower risk of** ...

CANSSI INCASS

On Collaboration

- Statistical colleagues
- Biostatistical colleagues
- Health and physical scientists
- Social scientists
- Humanists

CANSSI INCASS

On Collaboration

- National research labs
- Government research departments
- Industrial partners

- Policy analysts
- Policy makers
- Politicians

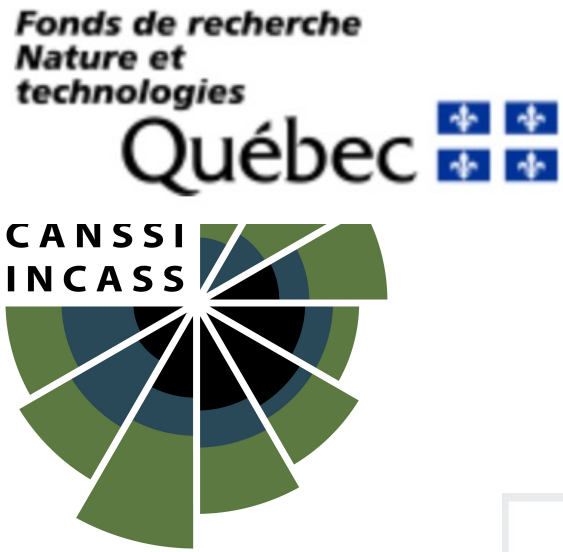
- Journalists
- Not for profit organizations
- Interested public

MITACS

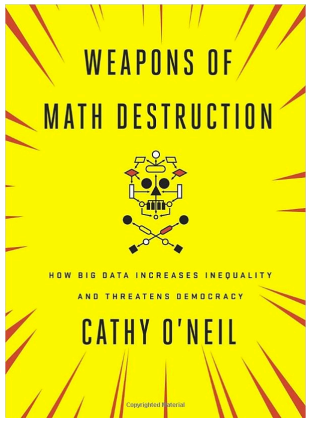
Scotiabank

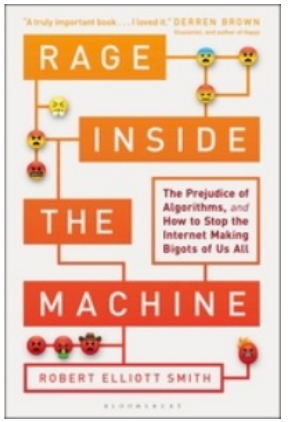
Public Health Agency of Canada

Looking Ahead



Looking Ahead





Looking Abroad

HDSR

HARVARD DATA SCIENCE REVIEW

A Microscopic, Telescopic, and Kaleidoscopic View of Data Science

CANSSI
INCASS