Letter of Submission

THIS NEEDS TO BE UPDATED – SEE THE DOCUMENT ON GITHUB

Title of the Project: Statistical Methods for Daily Mortality and Multiple Environmental Risk Factors

Name and affiliations of lead investigators

- Patrick E. Brown
  - Centre for Global Health Research, St. Michael's Hospital
  - Department of Statistical Sciences, University of Toronto
- Fateh Chebana
  - Centre Eau Terre Environnement, Institut national de la recherche scientifique, Québec.
- Cindy Feng
  - School of Epidemiology and Public Health, University of Ottawa
- Meredith Franklin
  - Keck School of Medicine, University of Southern California (until July 2021)
  - School for the Environment and Department of Statistical Sciences, University of Toronto (from July 2021)

List of proposed collaborators, titles, and affiliations

Name: Kamal Rai
Title: PhD student
Affiliations: Centre for Global Health Research, St. Michael's Hospital
Department of Statistical Sciences, University of Toronto

Name: Daniel Rainham
Title: Professor
Affiliations: School of Health and Human Performance, Dalhousie University
Healthy Populations Institute, Dalhousie University

Name: Hwashin Shin
Title: Scientist
Affiliation: Environmental Health Science and Research Bureau, Health Canada

Name: Céline Campagna
Title: Responsable scientifique
Affiliation: Équipe Changements climatiques et santé, Institut National de Santé Publique du Québec

Name: Pierre Masselot
Title: Research Fellow
Affiliation: London School of Hygiene & Tropical Medicine, United Kingdom

List of potential partner organizations

- The **Centre for Global Health Research, St. Michael's Hospital** will lead the health sciences research component of the project, providing data from the US and India and time of research staff to work on manuscripts. The Toronto-based component of the team will be located at the Centre for Global Health Research (CGHR) and will

be integrated into the Geospatial Mortality research group Dr. Brown leads. A portion of the Toronto PhD student's salary will be funded though a CGHR research grant.

- **L'Institut National de Santé Publique du Québec** has the mandate to analyse, monitor and evaluate determinants of health, including environmental pollutants, and will give access to the Quebec provincial health databases. In-kind support will also include health expertise for the statistical design and interpretation of research, including co-supervision of students.
- Hwashin Shin at **Health Canada** is central to this project. Her needs for an improved air quality indicator instigated the discussions which lead to this application. Dr. Shin has been funding Dr. Brown's team through Health Canada research contracts, and intends to provide at least $15,000 per year to the project.

Research Aims

THIS NEEDS TO BE UPDATED – SEE THE DOCUMENT ON GITHUB

Anticipated roles of trainees (students and post-doctoral fellows)

THIS NEEEDS TO BE UPDATED TO MATCH THE RESPONSE TO THE SAC's QUESTIONS

Kamal Rai will complete his PhD in 2021 and will work on this project as a postdoc. He will develop the Bayesian implementation of the GAIM models in Stan. This includes exploring determining appropriate prior distributions for the weights α, developing visualizations that communicate modeling results, and assisting other project members in developing shape constraints. He will be responsible for producing paper(s) summarizing the results of this model when run on Canadian air pollution and mortality data. To facilitate team communication and cohesion, he will also split time between Toronto (at the Centre for Global Health Research) and Ottawa (at the University of Ottawa), and use the proximity of the University of Ottawa to Quebec to occasionally visit project collaborators located there.

The University of Toronto PhD student will develop the INLA-like Bayesian computations to conduct inference on the GAIM, and compare its results from those obtained from the Stan implementation. A University of Laval or University of Ottawa PhD student will develop methods to conduct shape-constrained (Bayesian) inference, and examine the relationship between COVID-19 deaths and air pollution levels.

Equity, Diversity, Inclusion

THIS NEEDS TO BE COMPLETED

Anticipated organization of collaboration

THIS NEEEDS TO BE UPDATED TO MATCH THE RESPONSE TO THE SAC's QUESTIONS

The research team is situated in three cities - Toronto, Quebec, and Halifax. Two investigators, Patrick Brown and Meredith Franklin, will be in Toronto at the University of Toronto. One investigator and one collaborator, Cindy Feng and Daniel Rainham, will be in Halifax at Dalhousie University. Finally, one investigator, Fateh Chebana, will be in Quebec at INRS.

Plans for dissemination and communication

THIS NEEDS TO BE UPDATED

The lead investigators of this proposal have a track record of publishing research results in leading statistical and epidemiological journals, and aim to publish the results of this project in high-impact journals. The results and findings of this multiple pollutant inquiry will also be shared with Health Canada and the Institut National de Santé Publique du Québec. Drs Shin and Campagna will use the methodologies developed in their ongoing research and programatic work, and facilitate the adoption of the methods more widely in their organizations.

Schedule of events

THIS NEEDS TO BE UPDATED

Project Milestones

Task: Develop the bcGAIM

- 6-8 months: Implement prior(s) for shape-constrained inference for 1^{st}-order and 2^{nd}-order random walks.
- 8-12 months: Implement the bcGAIM in Stan and apply to the multi-pollutant problem.
- 12-16 months: Iterate development of priors until modeling results are satisfactory for the multi=pollutant model.
- 16-20 months: Write a paper summarizing these modeling results and submit for publication.
- 16-20 months: Release an R package so that these models are readily available.
- 20-24 months: Implement additional prior(s) for shape-constrained Bayesian inference.
- 24-30 months: Extend to a hierarchical model.
- 30-36 months: Write a paper summarizing the results of the hierarhical extensions for the exact and approximate inference models, and submit for publication.

Task: Approximate Inference Algorithm

- 4-6 months: Implement Laplace approximations for $\pi(\eta|\theta, \alpha, Y)$ and $\pi(\theta|Y)$ in Stan.
- 4-8 months: Implement an approximate inference algorithm for $\pi(\alpha|\theta, Y)$ in Stan.
- 8-12 months: Implement both approximations outside of Stan. Compare estimation results to those achieved in Stan.
- 12-16 months: Iterate development of approximation schemes outside of Stan.
- 16-20 months: Write a paper summarizing these results and submit for publication.
- 20-24 months: Apply both approximations to the multi-pollutant model.
- 20-24 months: Add the approximate Bayesian inference models to the R package.
- 24-30 months: Extend the approximation algorithm to hierarchical models. Compare to results obtained by Stan.
- 30-36 months: Add hierarchical approximate inference model to the R package.

Task: Multi-Pollutant Application

Note: The Stan implementation of the bcGAIM and the approximate inference algorithms will be developed against the multi-pollutant model, so the tasks for this application are mostly listed above. We include it on its own to give a specific breakdown of the development of the mulit-pollutant model.

- 8-12 months: Explore the performance of bcGAIM across regions and mortality outcomes.
- 12-16 months: Iteratively refine the bcGAIM (including the shape-constraining priors).

- 24-30 months: Extend the multi-pollutant model to a hierarchical model (exact and approximate versions).
- 30-36 months: Write a paper summarizing these results and submit for publication.
- 30-36 months: Add hierarhical bcGAIM model to the R package.

Task: COVID-19 Application

- 12-14 months: Identify COVID-19 confounders and data sets that may be used to fit a COVID-19 bcGAIM model.
- 14-16 months: Fit the bcGAIM model to COVID-19 mortality data.
- 16-20 months: Write a paper summarizing these results and submit for publication.

Task: Collaborative Applications

- 12-36 months: Once the bcGAIM is implemented, work with collaborations on appropriate epidemological studies.

Dissemination and Publication Activities

THIS NEEDS TO BE UPDATED

Year 1

- Milestone: Implement the bcGAIM with shape-constrained priors that are applicable for the multi-pollutant problem.
- Milestone: Implement the INLA-like approximation to the target density of this model in Stan.

Year 2

- Milestone:
- Milestone:
- Submit paper: A multi-pollutant air quality index.
- Submit Paper: Approximate Bayesian inference for the bcGAIM model
- Dissemination: Discuss shape-constrained Bayesian inference at 1-2 conferences.
- Dissemination: Discuss approximate Bayesian inference at 1-2 conferences.

Year 3

- Milestone:
- Submit Paper: Shape-constrained Bayesian inference with interpretable priors.
- Submit paper: The effects of multiple pollutant mixtures on COVID-19 mortality.
- Submit paper: A hierarchical extension to Approximate Bayesian inference
- Dissemintation: Discuss the multi-pollutant air quality index at 1-2 conferences.

- Dissemintation: Discuss hierarchical extensions to the multi-pollutant air quality index at 1-2 conferences.
- Dissemintation: Discuss the COVID-19 inquiry into air pollution exposure at 1-2 conferences.

Major Collaborative Activities

THIS NEEDS TO BE UPDATED

The different components of the bcGAIM project are naturally related. The bcGAIM is being developed in Stan in the first year, as is the first version of the approximate inference algorithm. Therefore, the bcGAIM model should be written to faciliate incorporating these approximations, and they should be developed knowing they will be implemented in Stan. In the second year, the two Stan models and the approximate inference algorithm will be extended to a hierarchical formulation. Although the hierarchical structure is at the city-level, nearby cities differ in their distance from each other. Ideally, a hierarchical model should account for how the composition of a mixture of pollutants varies by distance. The numerical difficulties and more complicated hierarchical structure should encourage strong collaboration at this stage of the project. The third year is devoted to applications – applying the fully developed bcGAIM model to the multi-pollutant problem, COVID-19 data, and other epiodemiological applications that arise during the course of the project – as well as writing papers and producing a useful R package. There is again natural collaboration between those writing and maintaining the R package.

Three-year budget

THIS NEEDS MILD UPDATES - I CAN DO THESE

The table below lists income and expenses for each of the three years of the project. This budget assumes that CANSSI funding over the 3 year period is $200,000.

Expenses	Contributions (in 1,000's)								
	Year 1			Year 2			Year 3		
	CANSSI	Non-CANSII		CANSSI	Non-CANSSI		CANSSI	Non-CANSSI	
Postdotoral Fellow	\$35	HC[1]	\$20	\$35	HC	\$20	\$35	HC	\$20
Graduate Students	\$24	INRS[2]	\$20	\$24	INRS	\$10	\$24	INRS	\$10
		NSERC[3]	\$7		NSERC	\$7		NSERC	\$7
Undergrad Students		USRA[4]	\$9		USRA	\$9		USRA	\$9
		DG[5]	\$3		DG	\$3		DG	\$3
Research Assistant		CGHR[6]	\$10		CGHR	\$10		CGHR	\$10
INSPQ Staff Time		INSPQ[7]	\$10		INSPQ	\$10		INSPQ	\$10
Travel	\$7			\$7			\$9		
Total	**\$66**		**\$69**	**\$66**		**\$69**	**\$68**		**\$69**

[1] HC: Health Canada

[2] INRS: Institut national de la recherche scientifique

[3] NSERC: The Natural Sciences and Engineering Research Council of Canada

[4] USRA: NSERC Undergraduate Student Research Award

[5] DG: NSERC Discovery Grant

[6] CHGR: Centre for Global Health Research

[7] INSPQ: Institut national de santé publique du Québec (Quebec Public Health Institute)

Annual Expenses

1. Postdoctoral Fellow: The postdoctoral fellow will be funded from the CANSSI CRT grant and Health Canada. He will help organize team meetings, split time between Toronto and Halifax, and help onboard other students as they join the project.

2. Graduate Students: 1 PhD student and 1 Master's student will be involved in this project. One of the graduate students will be based at INRS, the other will be at the University of Toronto or Dalhousie University.

3. URSA Students: This project will have a number of self-contained projects suitable for undergraduate research assistants. We intend to involve 2 URSAs at \$6,000/year each, whose work will directly contribute to the project's research aims.

4. Research Assistant: Hana Fu at CGHR will contribute roughly 3 days/month to the project. She will help maintain project data files and perform preliminary analysis.

5. INSPQ Staff Time: Céline Campagna at INSPQ will devote 0.5 days/week to the project.

6. Travel/Equipment: The travel expenses will cover attending conferences and travel between the three institutions by the project trainees. The equipment spending is intended to cover new computing equipment or cloud computing costs.

Contributions

1. CANSSI: The CANSSI funding is $200,000 over three years, or $66,666 per year.
2. Health Canada: Health Canada will contribute $20,000 per year via research contracts.
3. INRS: Fateh Chebana will contribute $10,000 in graduate student funding.
4. NSERC: Cindy Feng will contribute $7,000/year in graduate student funding.
5. INSPQ: The INSPQ will contribute staff time to the project, estimated at $10,000/year for 3 years.
6. CGHR: CGHR's support will be in-kind, in the form of funding the research assistant and providing research facilities.
7. Undergraduate Summer Students: Two USRAs will be applied for each year, which will pay for undergraduate summer students. NSERC also requires a contribution from the Discovery Grant of the supervisor.