CANSSI - Letter of Submission

SAC Review

1. *Provide details about the mentoring of the HQP, linking HQP directly to specific mentors. Include details on the mentoring of the student at Laval and how the student will be a full participant in the team. Keep in mind that it is unusual for a postdoctoral fellow to play a major role in mentoring, without the involvement of a senior researcher.*

   The HQP involved in the project are 1 postdoc fellow (PDF), 1 PhD student, 1 MSc student and 2 undergraduate trainees per year.

   The main advisors of the PDF will be Patrick Brown and Meredith Franklin and co-advisors Fateh Chebana and Cindy Feng. For financial and family reasons, the PDF will be located mainly in Toronto and Halifax and occasionally in Quebec City (1 week per semester when traveling between Toronto and Halifax).

   The main advisor of the PhD student will be Fateh Chebana and co-advisors Patrick Brown, Meredith Franklin and Cindy Feng. The PhD student will be located in Quebec City with short visits to Toronto and Halifax, once a semester. The main advisor of the MSc student will be Cindy Feng and will be located in Halifax. Patrick Brown, Meredith Franklin and Fateh Chebana will contribute as co-advisors.

   The undergraduate students usually as summer trainee where one (per year) will be located in Toronto to assist the PDF under the supervision of Patrick Brown whereas the second one (one per year) will be located in Quebec City to assist the PhD student under the supervision of Fateh Chebana. PDF and PhD student are encouraged to participate in the supervision of the undergraduate trainees.

   The collaborators at each location, under the request of the main advisors, will be helpful to the HQP by their practical expertise and by allowing HQP to extend their network.

2. *Clarify how the team will function and how team members will interact and collaborate. Describe the roles of the lead investigators in the project and research.*

   The PIs play the crucial role of the HQP supervision. Each PI acts as a supervisor of one main HQP (PDF, PhD and MSc) and one undergraduate trainee, and also as a co-supervisor for the other main HQP. Even though their contribution is essential for the project, the collaborators will play a complementary role in the supervision of the HQP.

   Regular meeting (weekly) between each HQP and the associated main advisor, and monthly (or as needed) with the co-advisors. Each month, all team members (PIs, HQP, collaborators) involved in a given sub-project will meet (in person or virtual) to discuss that specific sub-project. Bimonthly all team members will meet for updates, discussing issues, presenting results and methods, depending on the stage of the whole project (e.g. at the beginning discussions more about hiring, getting data, but at the end more about results).

   The HQP can visit their co-supervisors and the collaborators. This will be very beneficial to the HQP to get more abroad point of views. Each PI is responsible for one sub-project and all the PIs together are responsible for the whole project in all aspects (scientific, managing, administrative, etc).

   All members can meet at the SSC annual conference or any other opportunity.

3. *Provide more details on how the aims associated with the application are related to the statistical methodology.*

The Aims have been revised to focus on methodological contributions, and how the applications motivate the statistical work is more clearly explained. The practical problem to be addressed is assessing how multiple air quality and environmental risk factors (at different time lags) affect short term mortality when the effects are non-linear and the variables strongly correlated. Environmental epidemiology traditionally focuses on the relationship between single pollutant (i.e., particulate matter, nitrogen dioxide, sulfur dioxide, carbon monoxide, or ozone) and health outcomes (i.e., respiratory or circulatory diseases). In recent years, there has been great research interest in characterizing the relationship between multipollutant mixtures and health outcomes to better inform the air quality management decision-making and health risk assessment. The most commonly used method is to include multiple air pollutants in a regression model to investigate their independent association with the health outcome. Nevertheless, the models can be very unstable due to the multicollinearity or concurvity (the nonparametric analogy of multicollinearity) among the air pollutants, leading to significance tests with inflated type 1 error rates. As a result, new and innovative statistical methods for assessing how multiple highly correlation air pollutants affect short term morbidity/mortality are in great demand.

The proposed bcGAIM is anticipated to effectively address the above mentioned challenges (multicollinearity and concurvity) with the weights in the multi-pollutant air quality index estimated based on the data, as well as allowing for nonlinear relationships between pollutants and health outcomes. More broadly, the proposed research provides new ideas for modeling multiple correlated time-series predictors with potential lagged effects on an outcome. Nevertheless, developing this method is non-trivial and can not be easily implemented in any exiting software, since the likelihood surface could well be flat and have many ridges. We need to be able to identify which parameters or combinations thereof can and cannot be well identified by the data, which motivates the proposed methodology in this grant application.

4. *Briefly describe the available data sets and the confounders present therein.*

Daily air pollution data will be acquired from the National Air Pollution Surveillance (NAPS) Program, a network of 250 stations across Canada that is managed by Environment Canada and Environment and Climate Change Canada. Data on important environmental confounders of air pollution, such as temperature, wind speed and humidity will be acquired from weather stations operated by Environment and Climate Change Canada.

Daily health outcome data, which are not publicly availble, will be provided courtesy of Health Canada and INSPQ (the Quebec Public Health Institute). These data provide daily counts of cause-specific hospital admissions and mortality at the census tract or city level. Confounders in the health dataset include age, gender, race, and various markers of socio-economic status at the tract/city level.

COVID-19 data, which are publicly available, will be acquired from provincial authorities. Depending on the province in question, the COVID-19 data provide individual records of cases or mortality and include confounders such as age and socio-economic status. We will focus on building models for COVID-19 mortality as they are more reliable than the case data, which may suffer from bias since not everyone was tested due to limited testing capacity at the beginning of the pandemic.

The multi-pollutant cause-specific morbidity and mortality models will be developed separately from the multi-pollutant COVID-19 models. Both the nature of the data (count vs individual) and the available confounders are different between the two health datasets requiring different applications of the bcGAIM approach.

5. *Discuss why you model mortality from asthma, rather than occurrence.*

Our proposed model, bcGAIM, can be fit to any health outcome of interest. The outcomes will be chosen in consultation with epidemiologists with whom we are collaborating.

However, we do have a few outcomes in mind and they will be chosen based on the application of bcGAIM. For example, with the Health Canada data we will look at cause-specific morbidity and mortality including cardiopulmonary outcomes that have been examined in our previous studies (Franklin et al 2007,2008; Zanobetti et al 2009). To clarify the SAC's question, asthma will only be examined as a morbidity (i.e. occurrence), not as a cause of mortality.

As stated in the previous comment, COVID-19 mortality will be of interest due to issues with under-reporting of cases when tests were not available.

In summary, the outcome will be chosen based on which dataset we apply in bcGAIM, with expertise from our epidemiological collaborators, and with our previous work in mind.

6. *Detail the technical challenges to be addressed, provide references to what is already known and clearly state what is to be developed. In your discussion, include the following.*

   (a) *Provide explicit information on the kinds of constraints you will place on the prior distributions. Provide more extensive references to the literature on shape constrained models.*

   We now elaborate on the shape constrained non-parameteric effects. We will consider three approaches: basis function methods along the lines of @ramsay1998, reparameterizing Gaussian processes to accomdate shape-constrained priors, and a novel methodology of a random walk with skew-normal increments.

   (b) *Describe the challenges of introducing Bayesian approaches.*

   There are three main challenges to be addressed: finding a parametrization of bcGAIM which clearly separates combinations of parameters which are well identified and those which are weakly identified by the data; creating an MCMC algorithm which mixes well when the dimensionality is large and the likelihood surface flat; developing an INLA-like algorithm which is more easily automated than MCMC. We now elaborate on these challenges in the proposal.

   (c) *You propose to initiate the models in STAN. Does this mean the research is very straightforward? If not, what are the challenges?*

   It is unlikely that the most straightforward parametrization of bcGAIM will work well in Stan when the dimensionality is high. Finding an efficient implementation in Stan and developing alternatives to Stan are two of the key challenges. Developing a non-parametric model which is monotone (or encourages monotonicity) is another challenge.

Reviewer 1

1. *From a methodological perspective, the proposed extensions are not particularly novel. However, the application of Bayesian nonparametric regression in the context of air pollution epidemiology is novel.*

   The *Research Aims* contains additional details on the proposed extensions in the bcGAIM. Of these, the "nested model" construction of shape-constrained priors for Gaussian processes and the approximations proposed to render the hierarhical model less computationally demanding are perhaps more novel than they may have appeared in the LOI.

Reviewer 2

1. *What sort of research would be needed to construct the new bcAQHI once the new bcGAIM is built and computationally implemented?*

   With a Poisson response distribution and the response variable being the appropriate health outcome, the bcGAIM will output a relative risk for every combination of (measured or forecasted) pollutant values input into the model. The bcAQHI will take the relative risks as inputs and output warnings, based on cutpoints for the relative risk. The key task is determining the cutpoints, which we may do in a model-based way or using expert knowledge. Finally, forecasted pollutant values for next-day predictions will be provided by Environment Canada.

2. *The cGAIM itself is by no means a new idea. The original model for a single group (K=1) goes back to Hardle (1993). In fact, one seems to get Hardle's model if one drops the fk's from the model. Wang et. al. (2015) presents a multigroup version (K>1) to get around the curse of dimensionality, the whole point of this approach. But special cases were published between 1993 and 2015.*

   There are new features in the cGAIM – it considers constraints and groupwise additive index terms, while much of the existing literature only considers one or the other. While Hardle, Hall, and Ichimura (1993) examine a single index model and Wang et. al. (2015) consider a multiple index model, neither consider constraints. Two papers that consider constrained estimation are Xia and Tong (2006), where the authors constrain s to be monotonic and the components of α to be non-decreasing, and Fawzi et al. (2016), where the authors constrain the components of α to be non-negative and sum to one but do not constrain s. In comparison, the cGAIM allows any linear constraint to be placed on α and different shape constraints on s including monotonicity, convexity, and concavity (Masselot et. al., 2020). Additional comparisons between the cGAIM and bcGAIM are given in the *Research Aims* section.

3. *The third main topic seems the most novel in as much as it will show how the new bcAQHI might be used to assess COVID-19 mortality. Of course, it would seem more reasonable to me to build a new bcAQHI designed specifically for that purpose. And that led me to wonder about the health outcome to be used to fit the bcGAIM-the all-causes mortality-to get the alpha and in turn the index. For example, ozone would seem more relevant as a risk index for asthma and PM 2.5 for COPD. Why not just publish the pollutant concentrations themselves?*

   The bcGAIM is a modeling framework - it describes fixed effects, smooth functions of confounders, and a smooth function of a linear combination of covariates. For the multi-pollutant model, the bcGAIM estimates the relative risk of an observed air pollutant mixture, which is transformed into the bcAQHI by identifying cutpoints for the relative risk. If we would like to estimate relative risks, we would use the bcGAIM. If we would like to determine levels to provide warnings, we would transform the bcGAIM into an application-specific bcAQHI. For the COVID-19 model, we would like to report relative risks. For this application, we would fit the bcGAIM using COVID-19 mortality to estimate the relative risk of the pollutant mixture. We would also use COVID-19 specific fixed effects and confounders for this model. If desired, we could apply cutpoints to the estimated relative risks to obtain warning levels COVID-19 specific air quality index.

   Regarding the pollutant concentrations, in Canada they are publicly available via the National Air Pollution Surveillance (NAPS) Program. However, it is very difficult to understand their health effects without using a model-based approach. For the multi-pollutant model, the key benefit of the bcGAIM (and the bcAQHI) is that it provides an ease of interpretation for these health effects that just publishing data does not. Compared to single-pollutant models, it provides a measure of the relative risk of the mixture of air pollutants in the ambient air. Moreover, this is a much better representation of the health risks an individual will face than what is estimated by a single pollutant model.

4. *The conversion of the cGIAM to a bcGIAM should be feasible. A major challenge will be the big data problem since daily health counts for all-cause mortality will be modeled, and this using a semi-parametric model. But the task would be simplified by the intended approach of designing a different model for each Canadian city, how many we don't know. But they do not intend to incorporate random city effects*

in the Bayesian framework to enable strength to be borrowed deficiency in the intended approach, but perhaps a compromise needed for feasibility.

There are significant challenges relating to the size of the data. We will be examining 25+ regions in Canada over a 20+ year period (6,000+ days) of daily data. Fitting the bcGAIM to each city independently is computationally feasible, while fitting a hierarchical model across 25 regions is less feasible but perhaps more desirable. We have given further thought to this goal since submitting the LOI, and will be implementing various Laplace approximations in the target density to lessen the computational burden. This would allow us to produce estimates of the relative risks of the pollutant mixture using a hierarchical model. The details of the proposed approximation are discussed in the *Research Aims* section.

5. *A complicated issue and I had to do a lot of digging to figure out how the team of Investigators was assembled. In part, this involves those listed as Collaborators. The LOI could have done a better job of clarifying the links of the Investigators and Collaborators in the proposed project.*

   First in terms of locations, in Toronto there are as PIs Patrick Brown and Meredith Franklin are at the University of Toronto along with the postdoctotal fellow and one undergraduate trainee (1 per year) and XXX as collaborator at Centre for Global Health Research, St. Michael's Hospital. In Quebec city, Fateh Chebana is the PI and Céline Campagna is collaborator from INSPQ along with the PhD student and one undergraduate trainee (1 per year). In Halifax, Cindy Feng is the PI and Daniel Rainham as collaborator along with the MSc student. In Ottawa, we have our collaborators from Health Canada in particular Hwashin Shin and his team. Finally from the UK, we have Pierre Masselot who is a former PhD and postdoc with Fateh Chebana.

   Each PI is responsible for one sub-project and all the PIs together are responsible for the whole project in all aspects (scientific, managing, administrative, etc). Even though their contribution is essential for the project, the collaborators will play a complementary role in the supervision of the HQP. They also play a role in providing data, giving their expertise and knowledge of the data and the variables, their understanding of the practical problematic and the interpretation of the results. Pierre Masselot has an important role in terms of the codes of the modeling and providing the basics about the cGAIM model (the frequentist version).

   In terms of meetings, there will be regular meeting (weekly) between each HQP and the associated main advisor, and monthly (or as needed) with the co-advisors. Each month, all team members (PIs, HQP, collaborators) involved in a given sub-project will meet (in person or virtual) to discuss that specific sub-project. Bimonthly all team members will meet for updates, discussing issues, presenting results and methods, depending on the stage of the whole project (e.g. at the beginning discussions more about hiring, getting data, but at the end more about results). The HQP can visit their co-supervisors and the collaborators. This will be very beneficial to the HQP to get more abroad point of views. All members can meet at the SSC annual conference or any other opportunity.

6. *It is challenging to coordinate and run such a program successfully and we don't get a clear impression from the LOI that the applicants have thought about this issue very much. What is clear is that it is designed to provide the pipeline from data through to the AQHI. That is excellent. What is not clear is how the collaboration is to be managed and what I for one would like to see if a proposal is invited is an active group interaction plan.*

   Indeed, such of multidisciplinary and multi-location projects are challenging but interesting. This is particularly true in our project where members are at least in three provinces (Ontario, Quebec and Nova-Scotia) and from different disciplines (including statistics, epidemiology, environment, public health). Some PIs have already been involve in similar projects (e.g. Cindy Feng in a previous CRT by CANSSI) and as leader (e.g. Fateh Chebana with a major project with INSPQ). Hence, PIs have experience in successfully managing such kind of projects. In addition, in this specific project, the clarity of the role distributions, the different levels of contributions and responsibilities, the planned meetings (at different locations, for different purposes, etc), the complementarity between all team members, all are ingredient to project successful.

7. *The discussion of the bcGIAM is well done including the parts about the implementation of the computation strategies. But the LOI should have explained it better. For a start the big picture with its three main objectives should have been better described in general terms and how they relate to one another before diving into details re the cGIAM.*

   The bcGAIM project has three research objects: to develop a multi-pollutant air quality health index (AQHI), to develop multi-pollutant exposure models for various health effects, and to investigate how mixtures of pollutants affects daily COVID-19 mortality. These three tasks are all applications of the bcGAIM. The bcAQHI is derived from the relative risks estimated by the bcGAIM, the exposure models are bcGAIM models (with different fixed effects, smooth functions, and mixtures), and the COVID-19 model is the bcGAIM with daily COVID-19 mortality as the outcome. Thus, the methodological innovations are found in the bcGAIM, namely in developing interpretable priors and implementing approximate inference algorithms. The research aims - the bcAQHI, epidemiological studies, and COVID-19 studies - are potential high-impact applications of the bcGAIM that we will be pursuing.

Reviewer 3

1. *The naive case fatality rate that the investigators plan to explore with regard to Objective 3 is subject to errors caused by an undercount of both the numerator and the denominator and known to be a poor measure of the mortality risk of the disease.*

   While reported COVID-19 mortality rates are subject to reporting error and there may be inaccuracies in the reported data, it is the best data available. If there are anomalies such as persistent under-reporting, under-reporting on weekends or holidays, different inclusion criteria (long-term care cases and deaths may not be reported in some regions), or any others we can address them during data processing or model-fitting. For instance, we can add a day-of-the-week effect to account for delayed reporting on weekends, have the model impute missing values, or relate cases and deaths to help detect systematic under-reporting.

2. *Although four methodological advancements of the bcGAIM are listed in the Methods section, those seem to be rather incremental changes of the cGAIM.*

   The *Research Aims* section has additional material explaining the differences between the cGAIM and bcGAIM. The major benefit is in quantifying the uncertainty of α. Secondary benefits include being able to specify the strength of the monotonicity constraint, and the ability to extend the bcGAIM to additional smooth functions s or additional covariates in each s. As detailed in *Research Aims*, quantifying the uncertainty in the estimate of α is crucial for conducting inference with the multi-pollutant model. The extensibility of the bcGAIM and being able to control the strength of the shape constraint are also important features that will play an important role when in developing the multi-pollutant model.

3. *How the bcGAIM helps with developing a simple, intuitive air quality index that simultaneously accounts for the health effects of multiple air pollutants is not explained.*

   This was clarified in the response to the first question by Reviewer 2.

4. *The potential for impact in statistics and inferential data science seems to be marginal considering that methodological advancements of the bcGAIM listed in the Methods section are deemed to be rather incremental changes of the cGAIM.*

   The bcGAIM has novel innovations that will enable new statistical reasoning to applied to mixtures of interest, such as in the multi-pollutant problem. There are a number of statistical challenges in developing the bcGAIM. For one, fitting a model to daily observations across 25+ regions in Canada is a significant computational task. The bcGAIM attempts to address this by applying a (non-linear) one-dimensional function to a linear combination of related covariates, which both eases the computational burden and improves the interpretability of the model. To accomplish this, we must set shape-constrained priors on s and estimate α. These are both challenging problems and are discussed in more detail in the *Research Aims* section. It also has additional material to clarify the differences between the cGAIM and bcGAIM.

Reviewer 4

1. *The team has the potential to provide an excellent environment for interdisciplinary training of students. It would be helpful to add the names of the faculty supervisors/collaborators to the mentoring plan. The part for the roles of personnel could be further clarified.*

   Clarification can be found in the answer to the second question posed by Reviewer 2, as well as the *Anticipated Roles of Trainees* and *Anticipated Organization of Collaboration* sections. These two sections contain more details on how collaboration is organized, and the role of the investigators.

Reviewer 5

1. *A potential weakness would be that while well motivated by multi-pollutant modeling, the investigators could strengthen the proposal by identifying other applications where this form of model would be applicable.*

   The *Research Aims* section contains a brief discussion of other applications, with some further discussion in *Anticipated Roles of Trainees.*

2. *What is the relationship between the linear combination that goes into the smooth function and the air quality index? How will the index provide measures that indicate it is safe or not based on the smooth function and or linear combination of the exposures? In terms of data, is this based on hospital admissions (asthma or other conditions or only mortality). More details on data sources related to aims would be helpful.*

   The bcGAIM estimates the relative risk of the linear combination of the pollutants, and is translated into an air quality index based on cutpoints on the levels of the relative risk. For the air quality index, we are collaborating with epidemiologists and they will advise on the appropriate health outcome to use for the air quality index.

3. *While the researchers describe this as a constrained or shape constrained model, the proposal lacks details about what shape constraints are desired. There is a significant literature on Bayesian shape constrained modeling (monotonicity in particular) but limited references are provided. How does the proposed research build on this and what will be novel and contribute to general statistical methodology? Is this a straightforward model to fit in STAN or code directly or are there methodological advancements to be made there?*

   The *Research Aims* section contains additional details on the bcGAIM, more extensive references to the literature on shape-constrained inference, and details on the challenges to fitting the bcGAIM in Stan.

4. *The researchers cite Stringer et. al. (2020) as developing a Bayesian single pollutant version of a case-crossover model using non-MCMC methods such as Integrated Nested Laplace Approximations (INLA). Given the space limitations of the proposal it is not clear that such results will immediately carry over to the Bayesian shape constrained model where both the (constrained) smooth function of the linear combination of the exposures, the weights in the linear combination of exposures and the smooth functions of confounders has to be estimated, in addition to the other smooth functions of confounders. INLA like methods have been used to solve an array of complex problems, so this may be feasible. While this might provide computational efficiencies in point estimation, how does this address the importance of uncertainty quantification of the Bayesian model over the frequentist model of Masselot et. al. (2020)?*

   The *Research Aims* section gives more detail on the INLA-like approximation. To summarize those additions, the bcGAIM has link function $g(\lambda_t) = X^T\beta + s(\alpha^T Z_t) + f_1(W_{1,t}) + \ldots + f_K(W_{K,t})$. Conditional on α, we can simplify the estimation problem by considering parameters ϕ, θ, and α and estimating $\pi(\eta|Y,\theta,\alpha)$, $\pi(\alpha|Y,\theta)$, $\pi(\theta|Y)$, and $\pi(\eta|Y) = \int \pi(\eta|Y,\theta,\alpha)\pi(\theta|Y,\alpha)\pi(\alpha|Y)d\theta d\alpha$ (the last one numerically). The Laplace approximation can be applied to $\pi(\eta|Y,\theta,\alpha)$ and $\pi(\theta|Y)$ (within Stan), and $\pi(\theta|Y,\alpha)$ can be estimated using HMC. We will also develop an approximation of $\pi(\theta|Y,\alpha)$ outside of Stan. We will release both Stan models (exact and approximate), as well as the standalone approximate inference algorithm, in an R package.

5. *In discussing priors to induce shape constraints, the proposal rejects the idea of placing a prior on the expansion of the smooth function, but rather to place it directly on s. What types of priors on functions spaces are going to be used; Gaussian Process priors or others? How are the constraints incorporated?*

   We will be considering shape constraints on Gaussian processes, a widely used and flexible class of functions. The shape constraints would be incorporated by re-parameterizing the Gaussian process, or exploiting its structure to enforce the desired shape constraint. Ideally, a re-paramterization would create a parameter whose value directly relates to the shape constraint. For example, Kamal et. al. (2020) proposes a re-parameterization of the anisotropic Matern function where one parameter controls the anisotropic ratio of the Matern random field, and another the anisotropic angle. Alternatively, the

Guassian process contains mathematical structure that allows us to introduce an object that plays a similar role. The *Research Aims* section contains a brief discussion of shape-constrained Bayesian inference, and additional discussion on our approach to developing shape-constrained priors.

6. *While the impact of pollutant exposure and COVID-19 exposure is suggested by Wu et. al. (2020), how does this model relate to the models for total mortality? Will this be a joint model for COVID versus non COVID deaths with common smooth s function or different smoothed functions? Or will there be different linear combinations of exposures or the same? Or will this utilize the proposed Air-Quality index? What data are available and do they provide the necessary information about potential confounders or other covariates? (socio-economic status, access to health insurance, housing status (group living such as nursing homes, dorms, single family, number of family members etc) co-morbidities, etc. Are individual level data available or is this aggregated data at say a county level? Missing data is clearly an issue with COVID-19 deaths but is not addressed.*

   To investigate the relationship between air pollution and COVID-19 mortality, we will fit the bcGAIM to to the data we acquire from the provices. The granularity (individual or aggregate) of COVID-19 case and mortality reporting varies widely depending on the reporting region. For example the Government of Ontario's Treasury Board Secretariat provides many COVID-19 data sets at the individual level, including "Long-Term Care Home COVID-19 Data", "Confirmed positive cases of COVID-19 in Ontario", and "Status of COVID-19 cases in Ontario". The "Confirmed positive cases of COVID-19 in Ontario" data set contains age, gender, location (by public health unit), and the patient outcome. The "Status of COVID-19 cases in Ontario" data set contains daily tests completed, test outcomes, case outcomes, current hospitalizations, and current patients in ICUs. Other provinces only provide data at the census tract level, with breakdowns for covariates such as age and gender.

   It is important to note that none of these data sets were available at the time of our LOI submission. The Government of Ontario has gradually made additional information available to the public, a trend that will likely continue in Ontario and other reporting regions. It's very reasonable to expect more data in Ontario (and elsewhere) to be made available over time. Already, we have access to rich COVID-19 data sets for Ontario. For individual data we will develop bcGAIM models in a case-crossover framework addressing multiple pollutants and their lags. In these models individuals are condidtioned on themselves and the exposures vary between the case period (e.g. pollutant concentrations 2 days before death) and the control period (e.g pollutant concentrations on bi-directional referent days around the case period). Since individuals are conditioned on themselves, we do not need to adjust for individual-level confounders. For aggregate data available at the census tract or city level (COVID-19 from provinces other than Ontario, mortality and morbidity data from Health Canada) we will develop bcGAIM models in a Poisson "time series" framework. Here we will again address multiple pollutants but will also adjust for individual demographic and socio-economic confounders at the aggregate level. This approach is similar to the county-level aggregation seen in Wu et. al. (2020). In the Poisson bcGAIM we could also, for example, compare COVID-19 mortality in low-income vs. high-income regions, in majority white vs. majority minority regions, or in densely populated vs. less densely populated regions.

7. *As this is a very application motivated proposal it would be useful to know what data are available for each of the aims and how they will be integrated.*

   Daily air pollution data will be acquired from the National Air Pollution Surveillance (NAPS) Program, a network of 250 stations across Canada that is managed by Environment Canada and Environment and Climate Change Canada. Data on important environmental confounders of air pollution, such as temperature, wind speed and humidity will be acquired from weather stations operated by Environment and Climate Change Canada.

   Daily health outcome data, which are not publicly availble, will be provided courtesy of Health Canada and INSPQ (the Quebec Public Health Institute). These data provide daily counts of cause-specific hospital admissions and mortality at the census tract or city level. Confounders in the health dataset include age, gender, race, and various markers of socio-economic status at the tract/city level.

   COVID-19 data, which are publicly available, will be acquired from provincial authorities. Depending on the province in question, the COVID-19 data provide individual records of cases or mortality and include confounders such as age and socio-economic status. We will focus on building models for COVID-19 mortality as they are more reliable than the case data, which may suffer from bias since not everyone was tested due to limited testing capacity at the beginning of the pandemic.

The multi-pollutant cause-specific morbidity and mortality models will be developed separately from the multi-pollutant COVID-19 models. Both the nature of the data (count vs individual) and the available confounders are different between the two health datasets requiring different applications of the bcGAIM approach. It is also important to distinguish that the multi-pollutant estimates from the mobidity/mortality models (not the COVID-19 models) will be used as the generated bcAQHI. This allows us to provide air quality indices for particular health outcomes.

The bcGAIM is a general model - it contains fixed effects β, smooth functions of potential confounders $f = (f_1, \ldots, f_k)$, and a smooth function of a linear combination s whose coefficients α are also of interest. In our applications, the response is a mortality or morbidity outcome of interest. We will be fitting two versions of the bcGAIM to these two problems - one for the multi-pollutant model that will generate the bcAQHI, and one for the multi-pollutant COVID-19 investigation. The response for the COVID-19 model will be COVID-19 mortality, and it will also have different fixed effects and confounders.

8. *The proposal mentions developing random effects in the smooth function. What do these capture and what is the motivation? i.e spatial random effects, treating the weights as random effects (to allow spatial variation). Additional clarity would be helpful.*

   By random effects, we are referring to Gaussian processes such as random walks. For the hierarchical multi-pollutant model, we will likely also include a city-level spatial random effect.