

# Application: Bayesian Disease Mapping and Model-based Analysis for Spatial Epidemiology

YOON BAE JUN - yjun@iastate.edu

EOSS5: Essential Open Source Software for Science (Cycle 5)

## Summary

**ID:** EOSS5-0000000107

**Last submitted:** Apr 19 2022 01:25 PM (CDT)

## 1. Applicant Details

**Completed** - Apr 19 2022

## 1. Applicant Details

**Complete the following information for the Applicant (required)**

The information entered should be for the individual submitting the application who will act as the main person responsible for the application and as its point of contact. **To edit your name or email**, navigate to Account Information by clicking your name in the upper right corner.

**Name:** YOON BAE JUN

**Email:** yjun@iastate.edu

Add your home institution, company, or organization. This does not need to be the organization to which a grant would ultimately be awarded, if selected for funding.

Institution/Affiliation	Iowa State University
-------------------------	-----------------------

## 2. Proposal Details

**Completed** - Apr 19 2022

## 2. Proposal Details

**a. Proposal Title: Bayesian Disease Mapping and Model-based Analysis for Spatial Epidemiology**

*To edit your proposal title, navigate to the main page; click on the three dots to the right of the application title; and select Rename from the dropdown menu. Proposal title is limited to 60 characters including spaces.*

**b. Amount Requested**

*Enter requested budget in USD, including indirect costs. This number should be between \$100k and \$400k over a two year period. Enter whole numbers only (no dollar signs, commas, or cents).*

200000

**c. Proposal Summary/Scope of Work**

*Provide a short summary of the work being proposed (maximum of 500 words)*

In the past two years, due to the worldwide pandemic breakout with COVID-19, there has been a considerable demand in epidemiology and public health studies to come up with spatio-temporal models and maps of disease mortality/morbidity in human populations. It has become imperative that all developed sophisticated models are designed so that any clinical practitioner and enthusiastic user can quickly implement them to monitor the disease progression in the population. As an infectious disease spreads through proximity or direct contact of the virus over a period, the data thus associated is inherent of spatio-temporal nature (data collected across both space and time that describe a phenomenon). In this context, modern Bayesian/probabilistic methods enable us to utilize spatial and/or temporal correlation between observations and develop an effective modeling and data analysis scheme. A Bayesian approach provides us the necessary tool for uncertainty estimation regarding any of the "health effects of interest." This leads to an overall improve the quality and accessibility of health effect estimation. An innovative visualization is of utmost importance to understand the core issue related to any pandemic breakout and translate the modern model-based analysis results to clinical practitioners and public health experts. However, such desired statistical software that can implement current statistical model-based analysis results with innovative visualization targeted towards clinicians and

public health officials is unavailable or require strong programming skills. We have recently studied and published a comprehensive research article on the Bayesian spatio-temporal model to estimate the association between COVID-19 death counts (<https://pubmed.ncbi.nlm.nih.gov/35106052/>) and several social and environmental risk factors, including ambient air pollution exposure. Based on the knowledge and experience we gathered from the mentioned research work, we would like to develop and provide a simple, unified, publicly available cloud-based software platform that can be implemented in various disease mapping studies under the contemporary Bayesian/probabilistic framework. Medical practitioners and public health officials can use our developed software without computer coding knowledge. Our software will also provide an intuitive, bright graphical result output that can be well communicated among the general public with a clear understanding of the analysis results. Our primary target is to make these sophisticated Bayesian spatio-temporal models that are currently only limited to a niche academic group accessible, usable, and interpretable to the general public and medical scientists.

Below are some widely popular pandemic models and corresponding visualization we plan to create under this umbrella:

Bayesian Spatial Negative Binomial model

Bayesian Spatio-Temporal Negative Binomial model

Bayesian Spatio-Temporal Zero-Inflated Negative Binomial model

Data Summary Table Generator

Data Summary Figure Generator

Model Fit Summary, Model Quality Check, and Model Comparison Table Generator

Spatial Map visualization

Temporal Bar visualization

Temporal Trend visualization

Visualization of associated uncertainties with all relevant "health effects of interest" and model outputs.

#### **d. Value to Biomedical Users**

*Described the expected value the proposed work to the biomedical research community (maximum of 250 words)*

Our focus is the study of the geographical or spatial distribution of infectious disease and spatial visualization. Epidemiologists and public health researchers often face multifaceted challenges, (i) how to model and draw a meaningful conclusion from the data during the early stage of a pandemic when there are very few non-zero counts of incidents, (ii) how the spatial pattern and correlation structure of the data be integrated into the model; and (iii) how to account for the progression of the disease over time and quantify related uncertainties.

There might be statistical software that can accommodate the challenges mentioned above. However, epidemiologists and medical practitioners find them difficult to utilize without substantial coding knowledge or to rely on Statistics or computer sciences experts. Our software is designed for epidemiological and biomedical experts to get customized and easily interpretable health effect estimation outputs and map visualizations. Health effect modeling intended in the proposed software will include a zero-inflation model component to recognize excess zero counts (no-event) over a region or time. Such excess zero counts (no-event) are often the case in the early stages of an infectious disease. Bayesian spatio-temporal random effect model structure will be implemented in the proposed software to capture the disease's spatial correlation and/or temporal dynamics. The proposed software would operate on a real-time web-based platform and is intuitively easy to use and interpret without programming familiarity. The analysis results will be summarized and presented in a customized dashboard format that an epidemiologist or healthcare professional already knows.

---

#### **e. Open Source Software Projects**

Number of software projects are involved in your proposal (maximum of five):

2

Complete the table with the following information for each software project. If there is no homepage URL, re-enter the main code repository URL.

	Software project name	Main code repository URL	Homepage URL
1	R-package	<a href="https://github.com/junpeea">https://github.com/junpeea</a>	<a href="https://github.com/junpeea">https://github.com/junpeea</a>
2	Python-package	<a href="https://github.com/junpeea">https://github.com/junpeea</a>	<a href="https://github.com/junpeea">https://github.com/junpeea</a>

## f. Landscape Analysis

*Briefly describe the other software tools (either proprietary or open source) that the audience for this proposal primarily uses. How do the software project(s) in this proposal compare to these other tools in terms of user base size, usage, and maturity? How do existing tools and the project(s) in this proposal interact? (maximum of 250 words)*

SpatialEpi ([github.com/rudeboybert/SpatialEpi](https://github.com/rudeboybert/SpatialEpi)) is an R-package that implements statistical models used in Spatial Epidemiology. The Python spatial analysis library (PySAL, [pysal.org](https://pysal.org)) is another open-source library for geospatial data science.

SpatialEpi does not cover zero-inflation (early-stage pandemic) models nor spatio-temporal (disease progression over time and space) models. PySAL does not cover statistical models popularly used in Spatial Epidemiology. This software is not customized for general users and medical practitioners. Their model-based solutions are not intuitive to understand and interpret. None of the above two software can streamline the "data to knowledge" of infectious disease dynamics through spatio-temporal map visualizations.

Our analysis is illustrated in our recently published paper (<https://pubmed.ncbi.nlm.nih.gov/35106052/>) and the online repository (<https://github.com/junpeea/COVID-PM-STZINB>). We have three primary goals in our software design and development. Firstly, expand the user base from purely quantitative academics to a broader range of users like biomedical practitioners and public health researchers. Secondly, we will design a customized solution to conduct model-based disease risk estimation and map visualization varying over space and time. Thirdly, our software will be developed based on R-Shiny with cloud deployment. It will be easy to use with interactive interfaces. Our designed application can also run on mobile platforms (iOS and Android) and is seamlessly integrated into file-sharing platforms like Dropbox and OneDrive for data access and digital report storage.

Existing software (SpatialEpi and PySAL) will be used as our base libraries. Beyond that, we will expand to more advanced Bayesian/probabilistic models explicitly designed for infectious disease mapping and associated uncertainty quantification.

### g. Category

Choose the two categories that best describe the software project(s) audience.

	Category
Category 1	Infectious disease
Category 2	Visualization

### h. Previous CZI Funding

Did you previously apply for funding for this or a related proposal under the CZI EOSS program?	No
Have you previously received funding for this proposal under the CZI EOSS program?	No

## 3. Equal Opportunity & Diversity

Completed - Apr 19 2022

### Equal Opportunity & Diversity

CZI Science supports the science and technology that will make it possible to cure, prevent, or manage all diseases by the end of this century. Everyone is affected by disease, yet different communities are affected by or experience disease in different ways. Moreover, due to systemic barriers, the scientific enterprise itself is not a place where all voices and talents thrive. We believe the strongest scientific teams — encompassing ourselves, our grantees, and our partners — incorporate a wide range of backgrounds, lived experiences, and perspectives that guide them to the most important unsolved problems. To enable our work, we incorporate diverse perspectives into our strategy and processes, and we also seek to empower community partners to engage in science.

We request demographic information associated with applications submitted to CZI in response to our open calls. This information helps us learn from the RFA process, as well as improve our strategies to help ensure members of underrepresented or marginalized groups in science are aware of and able to apply to CZI opportunities. **Please note that answering all questions below is voluntary, and demographic information will not be used to make final grant funding decisions.** All responses will be shared

only with limited personnel, who will use that information only for the purposes described in this paragraph.

If you have any additional questions about why we ask this, what we do with the data, or to share suggestions for improvement, please reach out to [sciencegrants@chanzuckerberg.com](mailto:sciencegrants@chanzuckerberg.com).

**Please complete for the Applicant who is the person submitting the application. Please note that completing the below is voluntary, and demographic information will not be used to make final grant funding decisions.**

**What is your race/ethnicity? (optional)**

Asian (a person having origins in the Far East, Southeast Asian, the Indian Subcontinent; this does not include Native Hawaiian or Other Pacific Islander)

**What is the year of your last academic degree? (optional)**

2021

**What is your gender? (optional)**

Man

**Are you transgender? (optional)**

No

**Are you a member of the LGBTQ community? (optional)**

No



**Do you have one or more disabilities? (optional)**

No