

Opportunity Title: Chemical Source Modeling Using Data Mining, Statistics, and Machine Learning

Opportunity Reference Code: EPA-ORD-NRMRL-LMMD-2018-09

Organization U.S. Environmental Protection Agency (EPA)

Reference Code EPA-ORD-NRMRL-LMMD-2018-09

How to Apply A complete application consists of:

- An application
- Transcripts – [Click here for detailed information about acceptable transcripts](#)
- A current resume/CV, including academic history, employment history, relevant experiences, and publication list
- Two educational or professional references

All documents must be in English or include an official English translation.

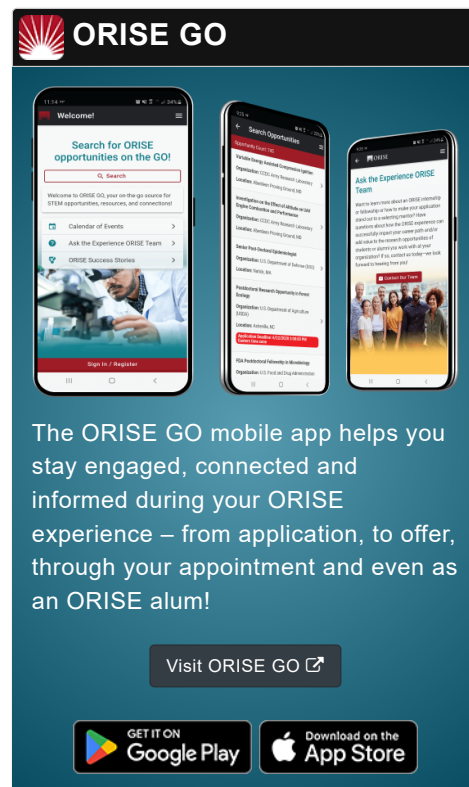
If you have questions, send an email to EPArpp@orau.org. Please include the reference code for this opportunity in your email.

Description

This research project will develop methods to model sources of chemical releases throughout the life cycle of a chemical, including manufacturing, processing, distribution, use, and end-of-life activities, for application in human exposure models as part of the Agency's high-throughput chemical risk assessment program. In collaboration with other ORD research, this research project will apply data mining, machine learning, and transport modeling principles to quickly and accurately estimate chemical releases.

The research project will involve the collection, curation, modeling, classification, regression, and prediction of chemical release data for risk assessment purposes. Collection will include searching for, extracting, documenting, and warehousing data throughout the world wide web. Curation will require evaluating and preprocessing data according to big data principles, with emphasis on data quality analysis. Modeling will involve the use of engineering knowledge to fill gaps in release data throughout the life cycles of chemicals. Classification refers to the use of machine learning to categorize collected data based on similarities in specified data descriptors, including physical properties, chemical quantities, and the nature of the activities involving the chemicals. Regression and other statistical methods will be applied as fit for purpose to model trends in the data. Prediction will be used as appropriate to extrapolate beyond the specific chemicals and circumstances studied in the previous steps. The research participant will interact with a team to develop methods and computer tools and to publish appropriate methodology and case study results.

The research participant will learn innovative ways to apply data mining and transport modeling skills within the field of chemical risk assessment to support next-generation high-throughput modeling approaches. The research participant will gain experience with the application of machine learning to big data for predictive analysis. The research participant will interact with leading exposure modelers and gain a better understanding of contemporary and emerging trends in human exposure modeling within a regulatory context. The research participant will learn about procedures for generating and managing high quality scientific data. The research participant will receive training on writing and publishing peer-reviewed research manuscripts. The research participant will have opportunities to learn about topics related to the primary research area of chemical risk assessment, such as



Opportunity Title: Chemical Source Modeling Using Data Mining, Statistics, and Machine Learning

Opportunity Reference Code: EPA-ORD-NRMRL-LMMD-2018-09

materials management and sustainability, through interactions with various parts of the Agency.

This program, administered by ORAU through its contract with the U.S. Department of Energy (DOE) to manage the Oak Ridge Institute for Science and Education (ORISE), was established through an interagency agreement between DOE and EPA. The initial appointment is for one year, but may be renewed upon recommendation of EPA and is contingent on the availability of funds. The participant will receive a monthly stipend commensurate with educational level and experience. Proof of health insurance is required for participation in this program. The appointment is full-time in the Cincinnati, Ohio area. Participants do not become employees of EPA, DOE or the program administrator, and there are no employment-related benefits.



Completion of a successful background investigation by the Office of Personnel Management (OPM) is required for an applicant to be on-boarded at EPA. OPM can complete a background investigation only for individuals, including non-US Citizens, who have resided in the US for the past three years.

Qualifications

The qualified candidate must have received a doctoral level degree in one of the related fields. Degree must have been received within five years of the appointment start date.

A background and/or expertise with all or a combination of the following is desired: computer programming (Python, R), data mining, statistics and regression analysis, machine learning, chemical process modeling, transport phenomena modeling, and life-cycle inventory modeling.

Eligibility Requirements

- **Degree:** Doctoral Degree received within the last 60 month(s).
- **Discipline(s):**
 - **Computer, Information, and Data Sciences** (4 )
 - **Engineering** (3 )

Affirmation

I certify that I have lived in the United States for the past three years.