

## Prediction of wildfire and air pollution with machine learning over the five domains of Eastern Eurasia

\*Teppei J Yasunari<sup>1,2</sup>, Ichigaku Takigawa<sup>3</sup>, Kyu-Myong Kim<sup>4</sup>, Arlindo M da Silva<sup>4</sup>

1. Faculty of Engineering, Hokkaido University, 2. Arctic Research Center, Hokkaido University, 3. Graduate School of Information Science and Technology, Hokkaido University, 4. NASA Goddard Space Flight Center

Under the ongoing global warming, necessity of better forecasting on wildfires and its air pollutions is increasing. In the last year, we carried out a preliminary wildfire prediction using the areal-averaged monthly mean data over a large Siberian domain (Yasunari et al., 2018, presented at AGU Fall Meeting 2018: <http://bit.ly/2PG1bmK>). Here we extend our previous analysis of machine learning and prediction into the five separated domains over the Arctic and sub-Arctic regions over Eastern Eurasia including Siberia, East Asia, Far East, etc. We use the same monthly mean dataset during 2003-2017 as we did in the last year above (i.e., NASA's FEER fire data, MODIS Snow Cover Fraction, SCF, Modern-Era Retrospective analysis for Research and Applications, Version 2, MERRA-2). The time periods of 2003–2014 and 2015–2017 focus on learning and prediction. Areal averaged data for objective (i.e., wildfire and air pollution variables) and explanatory variables (i.e., snow, temperature, wind, surface soil wetness, precipitation, geopotential height, etc.) over the defined five domains over Eastern Eurasia (Domain 1: 57°–70°N, 60°–90°E; Domain 2: 57°–70°N, 90°–140°E; Domain 3: 60°–70°N, 140°–180°E; Domain 4: 42.5°–57°N, 60°–90°E; Domain 5: 42.5°–57°N, 90°–140°E) are calculated, together with producing monthly time-lagged data for explanatory variables up to six months (i.e., total 174-month data available during 2003 and 2017). In addition, we made different areal-averaged fire pixel count data in which some smaller fires were removed with defined thresholds to remove expected human-made fires such as bonfires (i.e., assumed natural fire cases). If the areal-averaged fire data under the thresholds on a specific month showed no fire, we replaced that data to zero (i.e., used for our analyses but treated as zero fire data on that month).

Using Random Forest Regression, for example, in the machine learning, except for Domain 5, all the fire-included case showed the worse prediction score than those of the assumed natural fire cases. This implies that smaller fire cases, which were likely ignited by human activities or others, would be one of the causes for making wildfire predictions worse. However, over Domain 5, all the fire-included case showed the best probably because the assumed natural fire case was probably affected by the unusual large-scale wildfire events (i.e., outliers). These significantly large-scale wildfire cases also likely to make predictions worse because of not enough such cases existing in the training data (i.e., not frequent). In addition, these outliers might happen under unusual climatic and environmental conditions. How to include more large-scale fire cases in training data would be one of the important future targets on wildfire prediction with machine learning. On the day of the presentation, we'll also show the results for air pollution prediction.

Keywords: Arctic, wildfire, air pollution, PM2.5, machine learning, Siberia