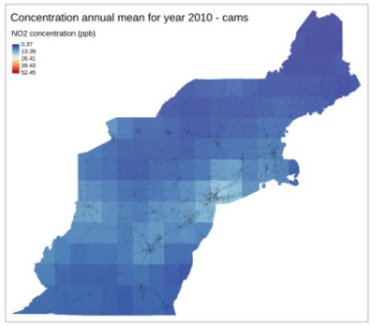
bneR - A collaborative workflow for air quality modeling and uncertainty characterization using the Bayesian Nonparametric Ensemble

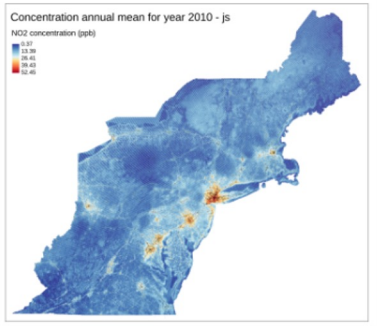
[1]Jaime Benavides, [2]Carlos Carrillo-Gallegos, [1]Vijay Kumar, [2]Arlene Fiore, [1]Marianthi-Anna Kioumourtzoglou

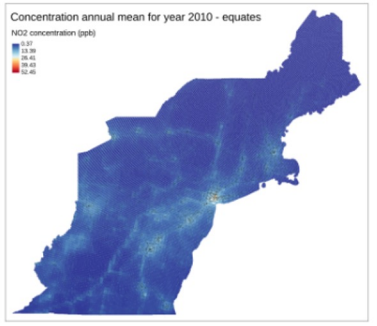
[1]Dept. of Environmental Health Sciences, Columbia University Mailman School of Public Health, NY
[2]Earth and Planetary Sciences, MIT, Cambridge, USA

06/04/2024
HAQAST meeting

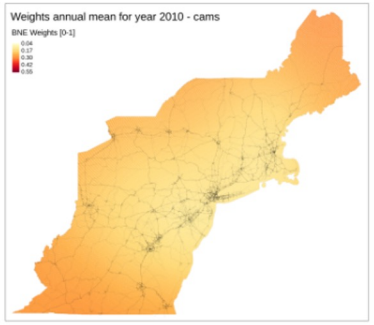
Bayesian Nonparametric Ensemble (BNE)

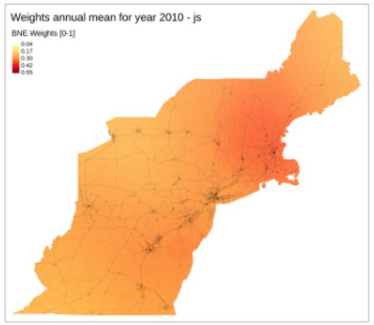


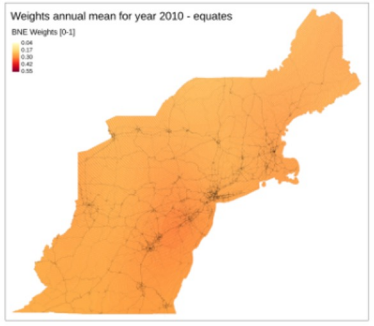




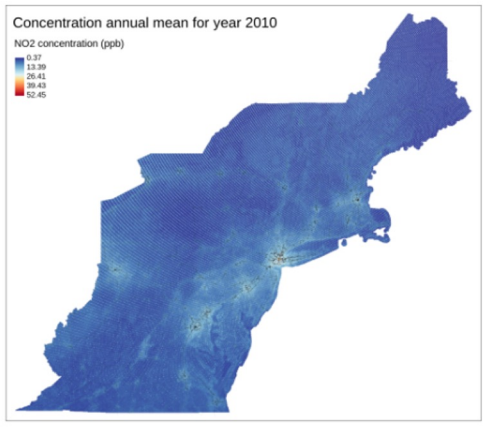
Input base models



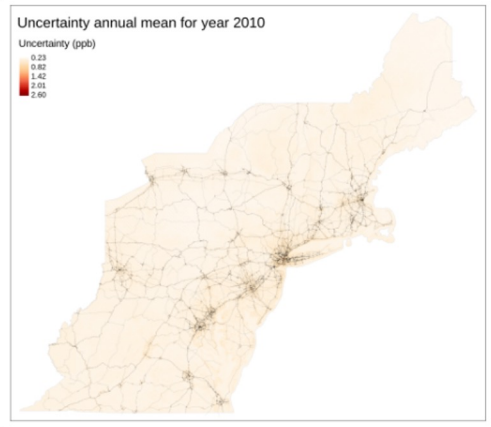




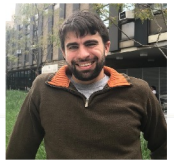
Model weights [0-1]



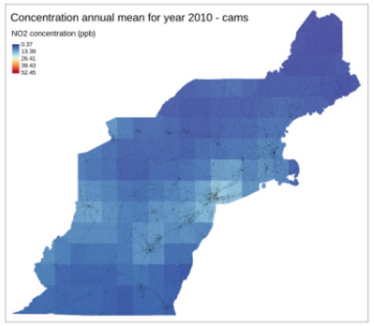
Ensemble concentrations

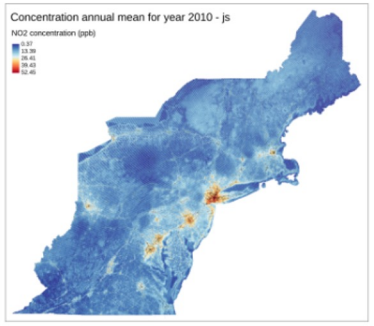


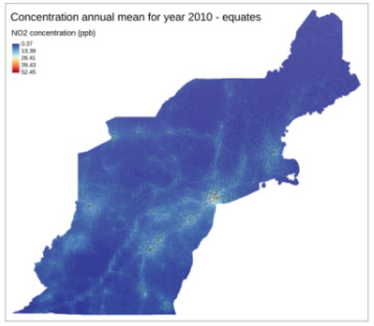
Uncertainty



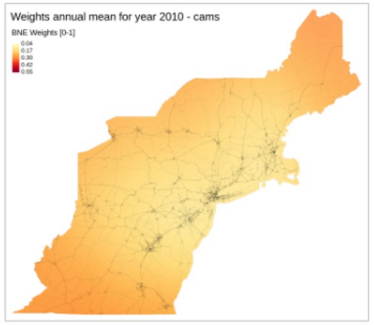
Bayesian Nonparametric Ensemble (BNE)

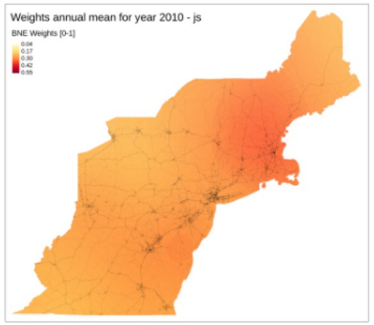


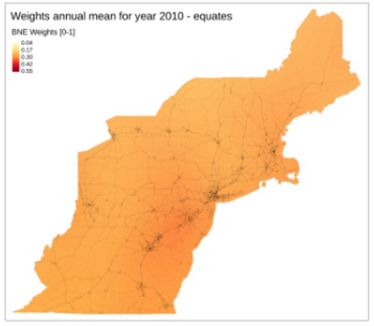




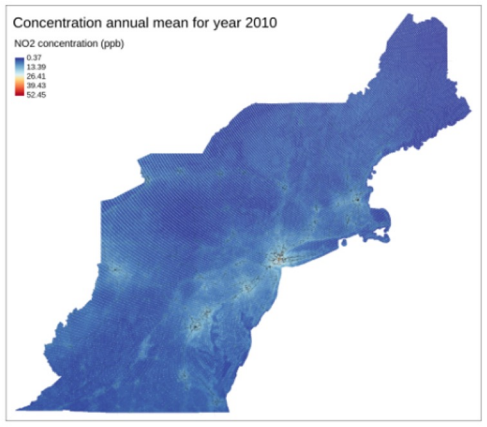
Input base models



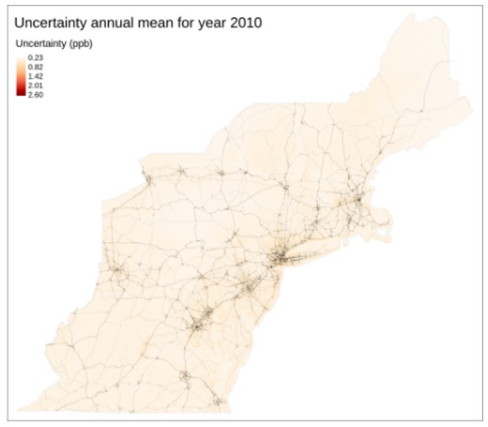




Model weights [0-1]



Ensemble concentrations



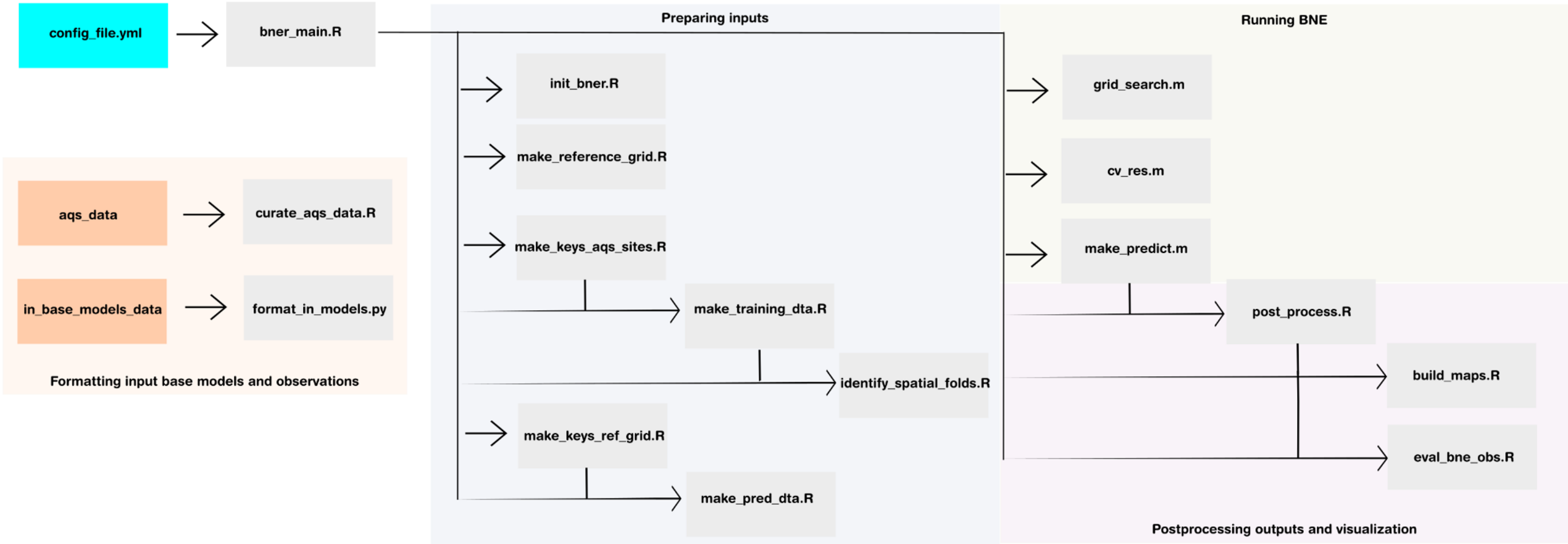
Uncertainty

For some more details, please visit Vijay's Kumar poster on BNE!

Input base models

Dataset	Pollutants	Spatial Resolution	Spatial Coverage	Temporal Coverage	Methodology
CAMS (ECMWF Atmospheric Composition Reanalysis 4)	NO_2, $PM_{2.5}$, O_3	25x 25 km^2	Global	2002-2022	Data assimilation using Integrated Forecast System (IFS) model for aerosols, chemistry, and greenhouse gasses, and several emissions datasets
EQUATES (EPA Air QUAlity Time SEries)	NO_2, $PM_{2.5}$, O_3	12 x 12 km^2	CONUS	2002-2019	WRF 4.1.1, CMAQ 5.3.2, National Emissions Inventory 2017
FAQSD (Fused Air Quality Surface Using Downscaling)	$PM_{2.5}$, O_3	12 x 12 km^2	Global	2002-2019	Fusion of CMAQ chemical transport model with ground level AQS data
JS (Di et. al)	NO_2, $PM_{2.5}$,	1 x 1 km^2	CONUS	2000-2016	Generalized additive model of three ML methods incorporating AQS data, AOD from MODIS, GEOS-CHEM, CMAQ-AQS fusion model, and land use variables
Liu (Bi et. al)	$PM_{2.5}$	1 x 1 km^2	NYS	2002-2016	AOD from MODIS, gap filled with random forest approach, incorporating AQS data, land use regression variables, and meteorology from NARR
OMI	NO_2	0.1° x 0.1°	Global	2005-2020	OMI Version 4 NO_2 Surface Estimate
WRF-CMAQ with AOD, CO	NO_2, $PM_{2.5}$, O_3	12 x 12 km^2	Global	2005-2020	Reanalysis assimilating AOD retrievals from MODIS and CO retrievals from MOPITT into CMAQ, with WRF providing meteorological input

bneR - workflow and collaborative modelling environment



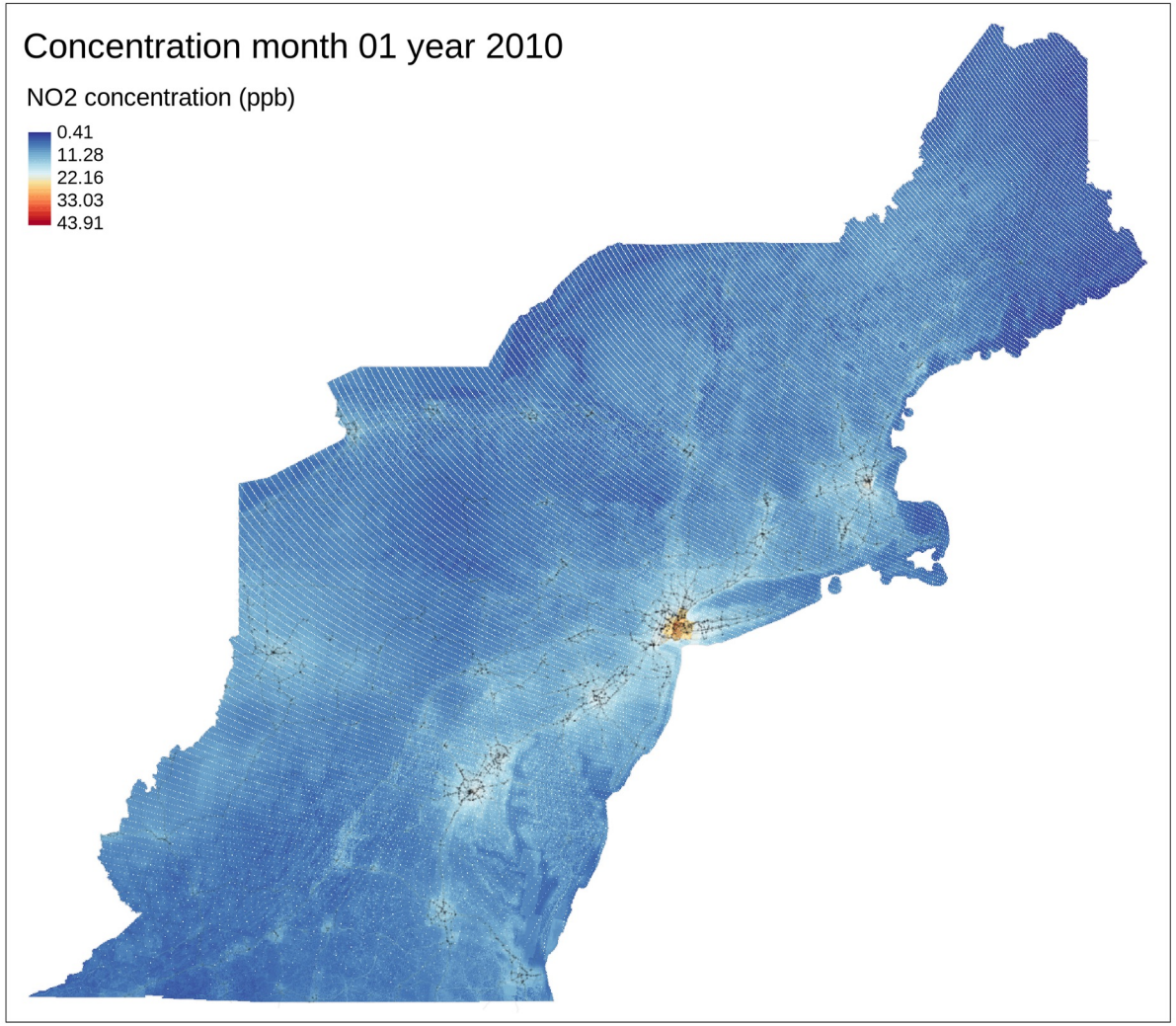
BNE workflow - configuration file

```
#------------------------------------------------------------------------#
#-------------------------BNER CONFIGURATION FILE-----------------------#
#------------------------------------------------------------------------#

## preproc processes options
run_sp_context: TRUE # if TRUE spatial context will be intersected with inp
run_create_ref_grid: TRUE
run_keys_aqs : TRUE
run_keys_preds : TRUE
run_make_training_dta : TRUE
run_make_preds_dta : TRUE
## bne processes options
run_bne_grid_search: FALSE
run_bne_cv: FALSE
run_bne_predict: FALSE
run_postprocess: FALSE


### set input parameters
exp_folder_path : "/data0/shr/bne/exp/"
data_path : "/data0/shr/bne/data/"
## general
expid : "exp_0068"
temp_resolution : "daily"
pollutant : "nox"
sp_context : "nys"
temp_context : 2015 # entire year that will be used for t
input_base_model_names : ["cams","equates","js", "omi"]
obs_dta_name : "aqs"
# defining grids
ref_grid_from_input_base_model : 3 # write number referr
this type of reference grid
ref_grid_resolution : NULL # in degrees (numeric); write
```

```
### BNE run configuration
## set parameters
time_var : 'percentOfYear'
# grid search (required to set the same number of
scale_space_w_list : [2, 1, 0.5] # [a, b, c]from a
scale_time_w_list : [2, 1, 0.5] # [a, b, c]from a
scale_space_rp_list : [2, 1, 0.5] # [a, b, c]from
scale_time_rp_list : [2,1,0.5] # [a, b, c]from a t
scale_space_wvar_list : [2,1,0.5] # [a, b, c]from
lambda_list : [0.3679, 0.1353, 0.0498, 0.0183] # s
lambda_rp_list : [0.3679, 0.1353, 0.0498, 0.0183]
seed_list : [1234] # set one seed
n_sample_list : [1000,2000]  # set two n samples
# cross-validation
```

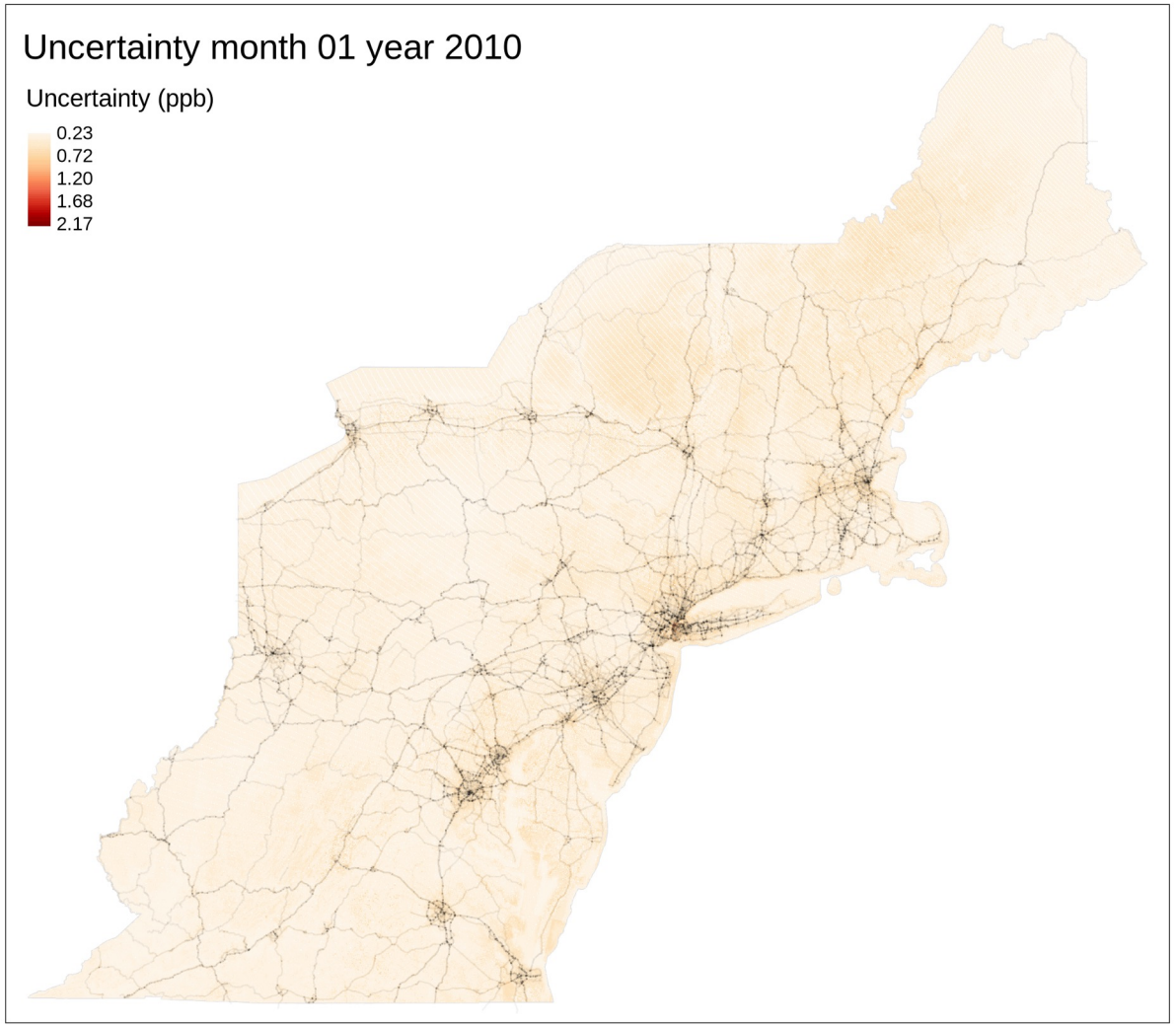
Average concentrations

BNE: [**"cams","equates","js"**]



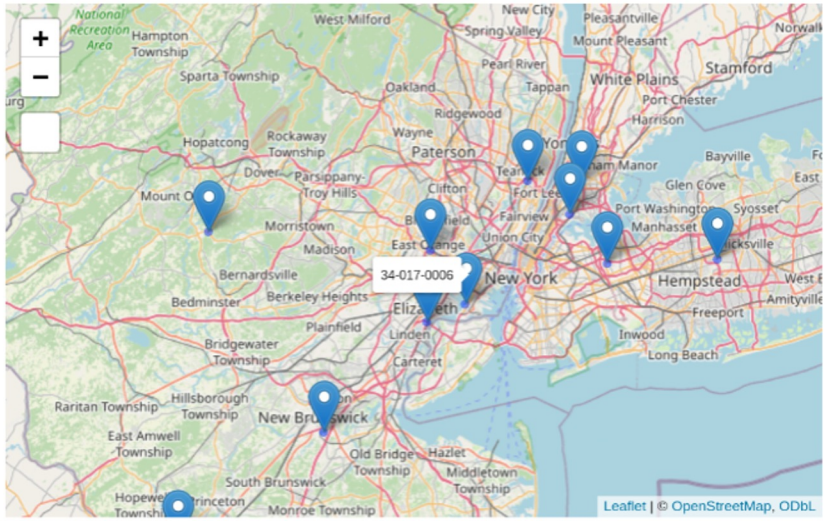
Uncertainty

BNE: [**"cams","equates","js"**]



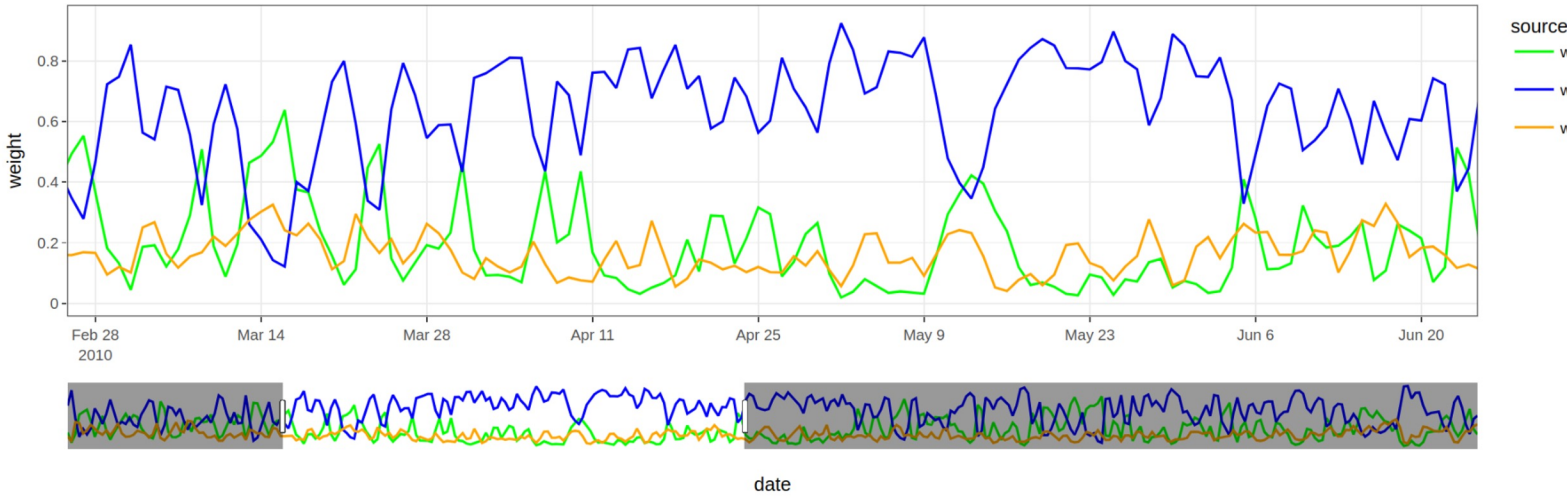
For preliminary results, please visit Vijay's Kumar poster on BNE!

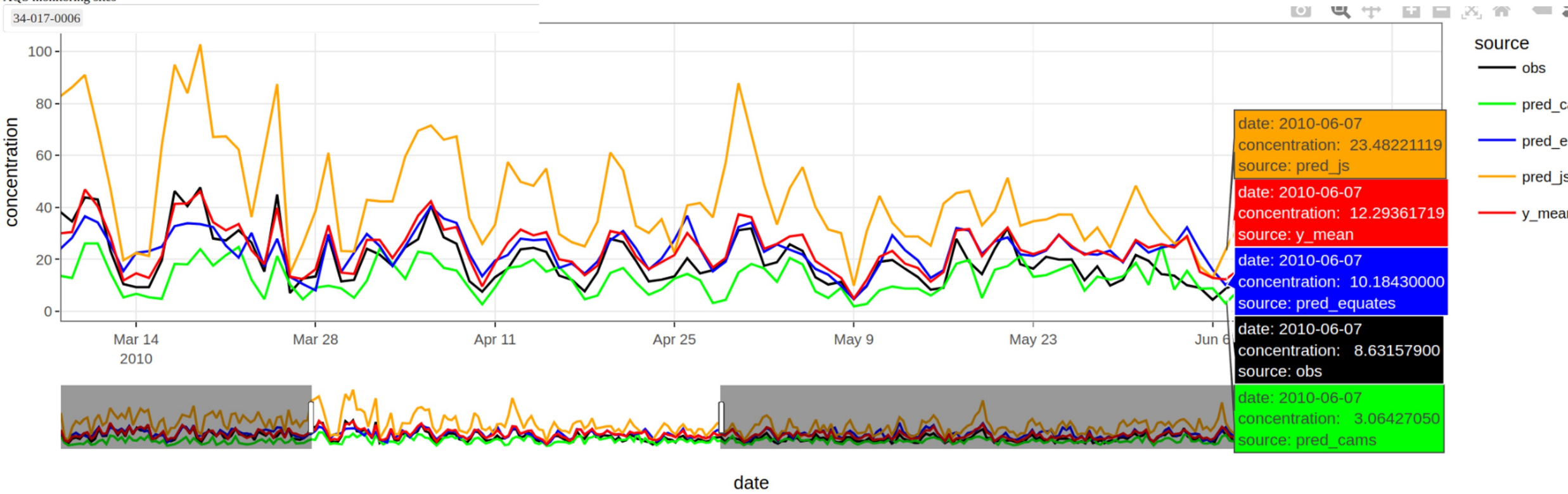
Interactive exploration of results



AQS monitoring sites

34-017-0006





Next: Co-producing data visualization for air pollutant concentration and uncertainty estimates

- **Objective:** Generate and disseminate data visualizations accounting for the needs of end-users and the integration of diverse expertises
- **Method:** Co-design of interactive visualizations to disseminate air pollutant concentrations and uncertainty estimates produced by BNE
- **Where:** Two online meetings to co-design and refine the co-developed interactive visualizations
- **When:** End of June ;)
- **Who:** Small group of potential users (e.g., health studies, policy making, etc..). **If interested, please let us know**

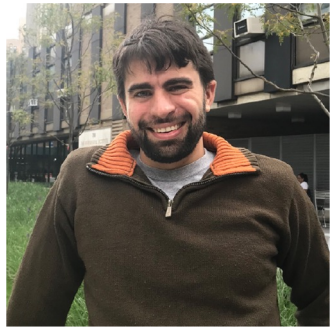
BNE team












Thank you!!

Jaime Benavides
jap2312@cumc.columbia.edu