Technical Information about Fused Air Quality Surface Using Downscaling Tool: Metadata Description

I. IDENTIFICATION TAB

A. CITATION PAGE

FIELD	FIELD CONTENT
* CATEGORY	Air quality
* PUBLICATION DATE	May 17, 2016
*TITLE	Fused Air Quality Predictions Using Downscaling
URL	https://www.epa.gov/hesc/rsig-related-downloadable-data-files
* NATIVE DATASET ENVIRONMENT	Predictive output files (.csv) include the following variables: Date, 11-digit Federal Information Processing Standard (FIPS) code (http://quickfacts.census.gov/qfd/meta/long_fips.htm) that denotes state FIPS code (digits 1-2), county FIPS code (digits 3-5), census tract base (digits 6-9), and census tract suffix (digits 10-11), Latitude (degrees) of US Census tract centroids, Longitude (degrees) of US Census Tract centroids, Pollutant Prediction, Predictive standard error. Except for Date, all variables are floating-point real numbers. Projection NAD83 was used to determine the census tract centroid locations. For 2001-06, 2000 US Census tract locations are used. Then in 2007-2010 US Census tract locations are provided.

B. DESCRIPTION PAGE

FIELD	FIELD CONTENT
* ABSTRACT	A Bayesian space-time downscaling fusion model approach described in a series of three published journal papers: 1) (Berrocal, V., Gelfand, A. E. and Holland, D. M. (2012). Space-time fusion under error in computer model output: an application to modeling air quality. *Biometrics* **68**, 837-848; 2) Berrocal, V., Gelfand, A. E. and Holland, D. M. (2010). A bivariate space-time downscaler under space and time misalignment. *The Annals of Applied Statistics* **4**, 1942-1975; and 3) Berrocal, V., Gelfand, A. E., and Holland, D. M. (2010). A spatio-temporal downscaler for output from numerical models. J. of Agricultural, Biological,and Environmental Statistics **15**, 176-197) is used to provide daily, predictive $PM_{2.5}$ (daily average) and O_3 (daily 8-hr maximum) surfaces for 2001-2006. The downscaling fusion model uses both air quality monitoring data from the National Air Monitoring Stations/State and Local Air Monitoring Stations (NAMS/SLAMS) and numerical output from the Models-3/Community Multiscale Air Quality (CMAQ). Currently, predictions at the US census tract centroid locations within the 12 *km* CMAQ domain are archived. Please refer to Section D below for actual spatial domain coordinates. Predictions at the CMAQ grid cell centroids, or any desired set of locations could be provided if needed. Predictions are not provided on December 31st of any year because complete

	daily CMAQ output are not available on those days after conversion of GMT to local eastern time
* PURPOSE	These archived predictive surfaces are intended for use by statisticians and environmental scientists interested in the spatial distribution of pollution for daily time periods. The predictive surfaces result from a research program designed to develop novel statistical fusion models at the US EPA/ORD/NERL in RTP, NC. As new statistical models are developed, we will evaluate their use for providing improved predictive space-time surfaces. **Model Limitations and Uses:** 1. These daily surfaces were developed primarily for health studies that relate daily pollution levels to daily measures of public health outcomes. 2. For air regulatory purposes, these daily predictions could easily be averaged to seasonal or annual temporal scales, with subsequent determination of areas exceeding air quality standards based on averages. The prediction standard errors could also be used to determine the probability of exceeding air standards. For air standards that are based on maximum values (e.g. ozone), using these daily predictions is still appropriate. However, extreme value ozone models (now under development) that focus solely on predicting maximum concentrations will be better suited to this specific application. 3. The use of these daily predictions to determine the number of days above or below air standard thresholds without consideration of the prediction uncertainty is discouraged. For counting daily exceedances, using these predictive surfaces is not the same as using air monitoring data (which are only influenced by measurement error). 4. There are areas of the conterminous U.S. with few or no air monitoring sites. In these areas, it may be difficult to adjust gridded CMAQ output to provide accurate predictions of air quality. For such areas, the bias-adjustments have been made based on the available monitoring data which may be located in areas with different emissions and geography.
SUPPLEMENTAL INFORMATION	
* PROGRESS	Complete
* UPDATE FREQ.	As needed

C. TIME & DATE PAGE

FIELD	FIELD CONTENT
* CURRENTNESS	May 17, 2016

* DATE TYPE	Range
* SINGLE DATE	
* MULTIPLE DATES	
Date 1	
Date 2	
Date 3	
* RANGE OF DATES	FROM: January 1, 2001 TO: December 30, 2011

D. GEOGRAPHIC AREA PAGE

2001 12 km FIELD	**FIELD CONTENT (longitude, latitude)**
* SOUTHWEST Corner	-99.5, 28.4
* NORTHEAST Corner	-67.0, 45.3
* NORTHWEST Corner	-100.4, 48.7
* SOUTHEAST Corner	-74.2, 25.9

2002-06 12 km FIELD	**FIELD CONTENT (longitude, latitude)**
* SOUTHWEST Corner	-106.8, 25.0
* NORTHEAST Corner	-65.3, 47.6
* NORTHWEST Corner	-111.1, 50.6
* SOUTHEAST Corner	-74.6, 22.9

2007- 12 km FIELD	**FIELD CONTENT (longitude, latitude)**
* SOUTHWEST Corner	-119.01, 24.04
* NORTHEAST Corner	-63.03, 47.92
* NORTHWEST Corner	-128.26, 48.59
* SOUTHEAST Corner	-72.99, 23.57

E. KEYWORDS PAGE

FIELD	**FIELD CONTENT**
* THEME	ISO
* THEME KEYWORDS	Environment
THEME 2	EPA
THEME 2 KEYWORDS	Air Pollution
THEME 3	
THEME 3 KEYWORDS	
* PLACES	Conterminous United States
* PLACES KEYWORDS	
PLACES 2	
PLACES 2 KEYWORDS	
PLACES 3	
PLACES 3 KEYWORDS	

F. SECURITY PAGE

FIELD	**FIELD CONTENT**
* SECURITY CLASSIFICATION SYSTEM	EPA classification system
* CLASSIFICATION	Medium Confidentiality
* SECURITY HANDLING DESCRIPTION	May be shared within the research community
* ACCESS CONSTRAINTS	Access for specific applications within use constraints
* USE CONSTRAINTS	The data are intended for use by statisticians and environmental scientists interested in the spatial distribution of daily air pollution over multiple years. Collaboration with EPA in these studies is expected.

II. DATA QUALITY TAB

FIELD	**FIELD CONTENT**
* PROCESS DATE	May 17, 2016
* PROCESS DESCRIPTION	The best available statistical fusion model and air quality data/numerical model output are used to develop these surfaces. It should be noted that data inputs and modeling assumptions can change over time, as we improve our our ability to statistically combine multiple sources of data.
PROCESS DATE	
PROCESS DESCRIPTION	
PROCESS DATE	
PROCESS DESCRIPTION	
PROCESS DATE	
PROCESS DESCRIPTION	
PROCESS DATE	
PROCESS DESCRIPTION	
PROCESS DATE	
PROCESS DESCRIPTION	
PROCESS DATE	
PROCESS DESCRIPTION	
PROCESS DATE	
PROCESS DESCRIPTION	
PROCESS DATE	
PROCESS DESCRIPTION	
PROCESS DATE	
PROCESS DESCRIPTION	

* LOGICAL CONSISTENCY REPORT	The predictive surfaces are based on using two sources of spatial information: air monitoring data and CMAQ numerical output (http://www.epa.gov/asmdnerl/CMAQ).
* COMPLETENESS REPORT	Predictive surfaces are provided for all days of the year.

III. ENTITY AND ATTRIBUTES TAB

FIELD	FIELD CONTENT
* OVERVIEW	The predictive surfaces are intended for use by statisticians and environmental scientists in modeling efforts that require high resolution air quality inputs for the period.
* DETAILED CITATION	**Input data**: The NAMS/SLAMS air quality monitoring data were downloaded from the Air Quality System (AQS) database. The CMAQ numerical output were created from version 4.6 of the model using CBIV mechanism. The output are 24-hour integrated $PM_{2.5}$ and daily 8-hr maximum O_3 areal average concentration over 12 km CMAQ cells. These CMAQ results are based on: (1) emissions data from the EPA's National Emissions Inventory developed using mobile emissions model Mobile 6; and (2) daily continuous emissions monitoring data for the major NOx point sources. Further, the meteorological data to produce these model results are from the Mesoscale Model 5 (MM5) version 3.6.3 simulations. **Downscaler Model Description**: The space-time Bayesian fusion model combines the monitoring data and CMAQ output to predict daily air pollution at the point level. The use of the term, downscaler, occurs by scaling the areal grid-cell CMAQ output to the point-level air monitoring data. The downscaler differs from other fusion efforts by not assuming the existence of a true air pollution process driving both the monitoring data and CMAQ output. Instead, downscaling relates air data and model output using a linear regression with bias coefficients (additive and multiplicative) that can vary in space and time. This approach to modeling provides a new answer to the “change-of-support” problem where we would like to predict air pollution at a certain spatial resolution, but must reconcile the difference between point monitoring data and areal CMAQ concentrations. Model parameters of the downscaler model are fit just to paired CMAQ and air monitoring data, thus CMAQ output that do not contain monitoring sites are not used in model fitting. The entire CMAQ surface is used when predicted to a desired set of point locations that could be the CMAQ grid centroids, census tract centroids, or any set of point locations.

IV. DISTRIBUTION TAB

FIELD	FIELD CONTENT
RESOURCE DESCRIPTION	Downloadable zipped data files (.csv) containing daily predictive surfaces for each pollutant by year.
DISCLAIMER/LIABILITY	Although these data have been processed successfully on a computer system at the Environmental Protection Agency, no warranty expressed or implied is made regarding the accuracy or utility of the data on any other system or for general or scientific purposes, nor shall the act of distribution constitute any such warranty. It is also strongly recommended that careful attention be paid to the contents of the metadata file associated with these data to evaluate data set limitations, restrictions or intended use. The U.S. Environmental Protection Agency shall not be held liable for improper or incorrect use of the data described and/or contained herein.
CUSTOM ORDER PROCESS	

V. METADATA TAB

FIELD	FIELD CONTENT
* DATE CREATED	May 17, 2017
* STANDARD NAME	
* ACCESS CONSTRAINTS	Access for specific applications within use constraints
* USE CONSTRAINTS	The data are intended for use by statisticians and environmental scientists in modeling efforts that require high-resolution predictive spatial fields of air pollution.

VI. CONTACTS TAB

A. ORIGINATORS PAGE

FIELD	FIELD CONTENT

* PERSON	David M. Holland
* ORGANIZATION	US Environmental Protection Agency
* TITLE	Mathematical Statistician
USERID	
HOURS	
INSTRUCTIONS	
* PHONE NO. 1	919-541-3126
PHONE NO. 2	
* FAX	919-541-1138
* E-MAIL	holland.david@epa.gov
TDD/TTY	
* STREET ADDRESS	109 T. W. Alexander Drive, E243-05 (NERL)
* CITY	Research Triangle Park
STATE	NC
COUNTRY	USA
* ZIP	27711

B. DISTRIBUTORS PAGE

FIELD	FIELD CONTENT
* PERSON	Adam Reff
* ORGANIZATION	US Environmental Protection Agency
* TITLE	Atmospheric Scientist
USERID	
HOURS	
INSTRUCTIONS	
* PHONE NO. 1	919-541-5683
PHONE NO. 2	
* FAX	
* E-MAIL	Reff.adam@epa.gov
TDD/TTY	
* STREET ADDRESS	109 T. W. Alexander Drive
* CITY	RTP
STATE	NC
COUNTRY	USA
* ZIP	27711