Linking NASA Environmental Data with a National Public Health Cohort Study and a CDC on-line System to Enhance Public Health Decision Making

Mohammad Al-Hamdan[1], William Crosson[1], Sigrid Economou[2], Maurice Estes Jr[1], Sue Estes[1], Sarah Hemmings[1], Shia Kent[3], Mark Puckett[2], Dale Quattrochi[4], Gina Wade[1], Leslie McClure[3]

[1]Universities Space Research Association at NASA Marshall Space Flight Center, Huntsville, Alabama, USA.
[2]National Center for Public Health Informatics, Centers for Disease Control and Prevention, Atlanta, Georgia, USA.
[3]School of Public Health, University of Alabama at Birmingham, Birmingham, Alabama, USA.
[4]Earth Science Office at NASA Marshall Space Flight Center, Huntsville, Alabama, USA.

The overall goal of this study is to address issues of environmental health and enhance public health decision making by utilizing NASA remotely-sensed data and products. This study is a collaboration between NASA Marshall Space Flight Center, Universities Space Research Association (USRA), the University of Alabama at Birmingham (UAB) School of Public Health and the Centers for Disease Control and Prevention (CDC) National Center for Public Health Informatics. The objectives of this study are to develop high-quality spatial data sets of environmental variables, link these with public health data from a national cohort study, and deliver the linked data sets and associated analyses to local, state and federal end-user groups. Three daily environmental data sets were developed for the conterminous U.S. on different spatial resolutions for the period 2003-2008: (1) spatial surfaces of estimated fine particulate matter ($PM_{2.5}$) exposures on a 10-km grid utilizing the US Environmental Protection Agency (EPA) ground observations and NASA's MODerate-resolution Imaging Spectroradiometer (MODIS) data; (2) a 1-km grid of Land Surface Temperature (LST) using MODIS data; and (3) a 12-km grid of daily Solar Insolation (SI) and maximum and minimum air temperature using the North American Land Data Assimilation System (NLDAS) forcing data. These environmental datasets were linked with public health data from the UAB REasons for Geographic and Racial Differences in Stroke (REGARDS) national cohort study to determine whether exposures to these environmental risk factors are related to cognitive decline and other health outcomes. These environmental national datasets will also be made available to public health professionals, researchers and the general public via the CDC Wide-ranging Online Data for Epidemiologic Research (WONDER) system, where they can be aggregated to the county, state or regional level as per users' need and downloaded in tabular, graphical, and map formats. The linkage of these data provides a useful addition to CDC WONDER, allowing public health researchers and policy makers to better include environmental exposure data in the context of other health data available in this online system. It also substantially expands public access to NASA data, making their use by a wide range of decision makers feasible.

Linking NASA Environmental Data with a National Public Health Cohort Study and a CDC on-line System to Enhance Public Health Decision Making

Mohammad Al-Hamdan, William Crosson, Maury Estes, Sue Estes, Sarah Hemmings, Dale Quattrochi
Universities Space Research Association and Earth Science Office at NASA Marshall Space Flight Center, Huntsville, AL

Leslie McClure, Shia Kent
School of Public Health, University of Alabama at Birmingham, Alabama

Sigrid Economou, Mark Puckett
National Center for Public Health Informatics, Centers for Disease Control and Prevention, Atlanta, Georgia

Gina Wade
Von Braun Center for Science and Innovation, National Space Science and Technology Center, Huntsville, AL (Previously with USRA/MSFC)

AMS Annual Meeting
24 January 2012

Goals and Objectives

➢ **This project has dual goals in decision-making activities**
  - ➢ Providing information to decision makers about associations between environmental exposures and health conditions in a large national cohort study
  - ➢ Enriching the CDC Wide-ranging Online Data for Epidemiologic Research (WONDER) system by integrating environmental exposure data

➢ **Develop daily high-quality spatial data sets of environmental variables for the conterminous U.S. for the years 2003-2008 utilizing NASA data (Objective 1)**
  - ➢ Fine Particulates ($PM_{2.5}$) (NASA MODIS and EPA AQS)
  - ➢ Land Surface Temperature (NASA MODIS)
  - ➢ Solar Insolation and Heat-related Products (Reanalysis Data)

➢ **Link these environmental variables with public health data from a national cohort study and examine environmental health relationships (Objective 2)**
  - ➢ Cognitive Function
  - ➢ Hypertension

➢ **Make the environmental datasets available to public health professionals, researchers and the general public via the CDC WONDER system (Objective 3)**

Environmental Health Implications

➢ **Fine Particulates (PM2.5)**

  ➢ Human observation studies show that exposure to general pollution containing PM2.5 could cause inflammation, degradation, and oxidation in the brain when inhaled and could lead to altered regulation of biomarkers involved in cognitive function

  ➢ Possible risk factor for cardiovascular and respiratory diseases

➢ **Solar Insolation**

  ➢ Some research suggests that a relationship between sunlight exposure and cognition exists by affecting brain blood flow

➢ **Heat Exposure**

  ➢ Some research suggests that a relationship between heat exposure and hypertension exists by affecting stress level

National Environmental Datasets

(Objective 1)

Fine Particulate Matter (PM$_{2.5}$)

➤ Estimated ground-level PM$_{2.5}$ from MODIS AOD using published regression equations per EPA region per season (Zhang et al., JAWMA 2009)

➤ Combined with EPA PM$_{2.5}$ data from the AQS for 2003-2008

➤ Modified and ran MSFC Surfacing Algorithm (Al-Hamdan et al., JAWMA 2009) to produce continuous spatial surfaces of daily PM$_{2.5}$ for the contiguous US for 2003-2008
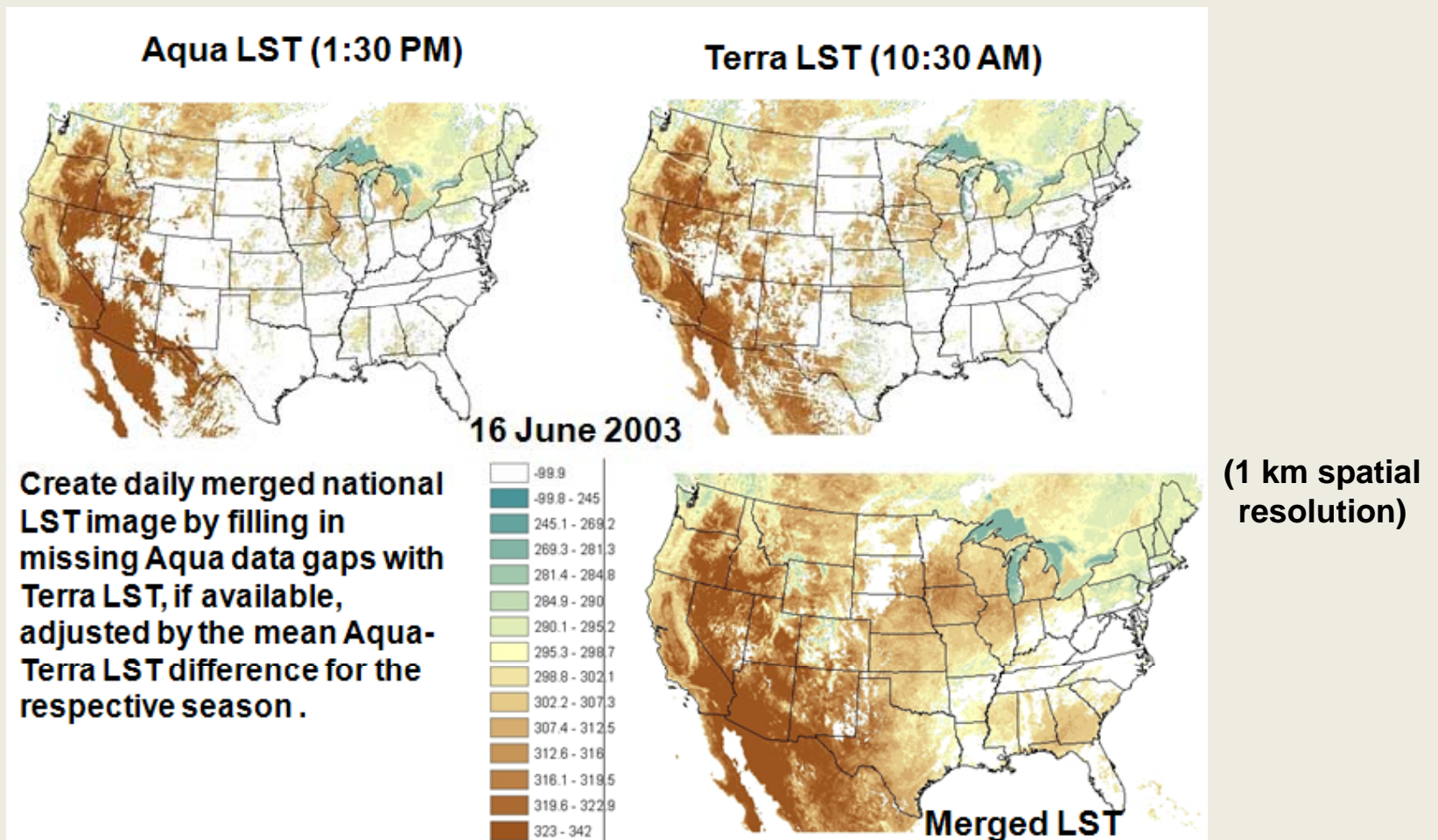

EPA AQS

NASA MODIS

PM2.5 Estimates

PM$_{2.5}$ (ug/m3)

0.00 - 2.95
2.96 - 5.90
5.91 - 8.84
8.85 - 11.79
11.80 - 14.74
14.75 - 17.69
17.70 - 20.64
20.65 - 23.59
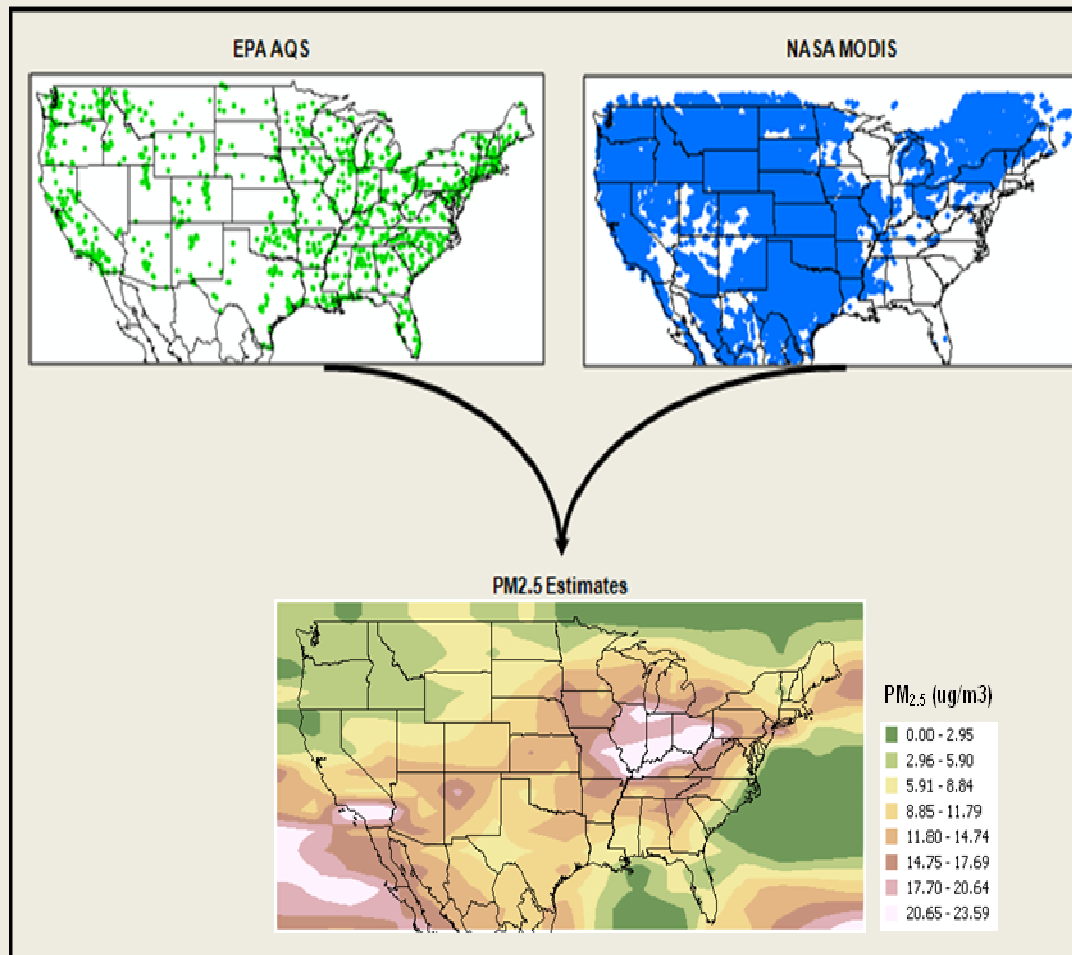
PM$_{2.5}$ on July 14, 2003

(10 km spatial resolution)

Land Surface Temperature (LST)

➢ Aqua and Terra daytime & nighttime data for 2003-2008 were processed

➢ Aqua-Terra differences were computed by season for 2003-2008

➢ Aqua data gaps were filled with Terra-adjusted LST (if available) by mean seasonal difference

➢ National merged Aqua-Terra daily LST dataset were generated for 2003-2008 for day & night (Crosson et al., RSE, in press)

Aqua LST (1:30 PM)

Terra LST (10:30 AM)

16 June 2003

Create daily merged national LST image by filling in missing Aqua data gaps with Terra LST, if available, adjusted by the mean Aqua-Terra LST difference for the respective season.

(1 km spatial resolution)

	-99.9
	-99.8 - 245
	245.1 - 269.2
	269.3 - 281.3
	281.4 - 284.8
	284.9 - 290
	290.1 - 295.2
	295.3 - 298.7
	298.8 - 302.1
	302.2 - 307.3
	307.4 - 312.5
	312.6 - 316
	316.1 - 319.5
	319.6 - 322.9
	323 - 342

Merged LST

Heat and Solar Insolation

➤ NLDAS hourly forcing data (air temperature, solar radiation, specific humidity, atmospheric pressure) for the 2003-2008 period were processed

➤ Daily statistics of Maximum Air Temperature, Minimum Air Temperature, Maximum Heat Index, and Total Solar Insolation were computed for 2003-2008

NLDAS Max Air Temperature on July 15, 2008

Temp (°F)
- 51.2 - 57.7
- 57.8 - 64.2
- 64.3 - 70.7
- 70.8 - 77.2
- 77.3 - 83.7
- 83.8 - 90.3
- 90.4 - 96.8
- 96.9 - 103.3
- 103.4 - 109.8
- 109.9 - 116.3
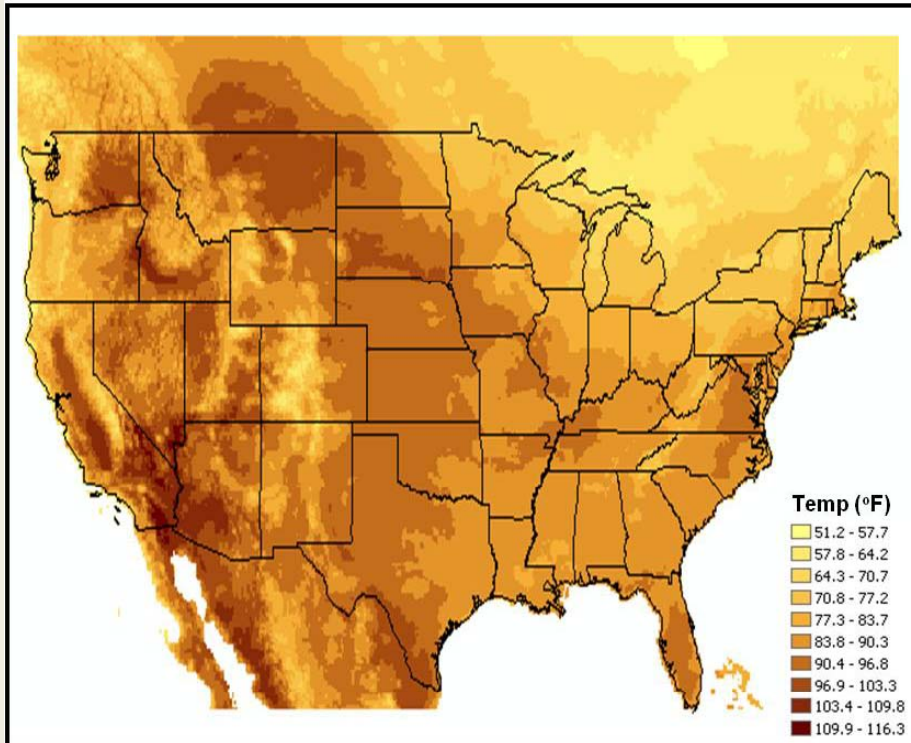
NLDAS Solar Insolation on January 1, 2008

KJ/m²/Day
- 1,512.75 - 4,276.7
- 4,276.71 - 5,700.55
- 5,700.56 - 7,040.65
- 7,040.66 - 8,548.27
- 8,548.28 - 9,888.37
- 9,888.38 - 11,144.7
- 11,144.8 - 12,401.1
- 12,401.2 - 13,657.4
- 13,657.5 - 14,913.7
- 14,913.8 - 17,761.5
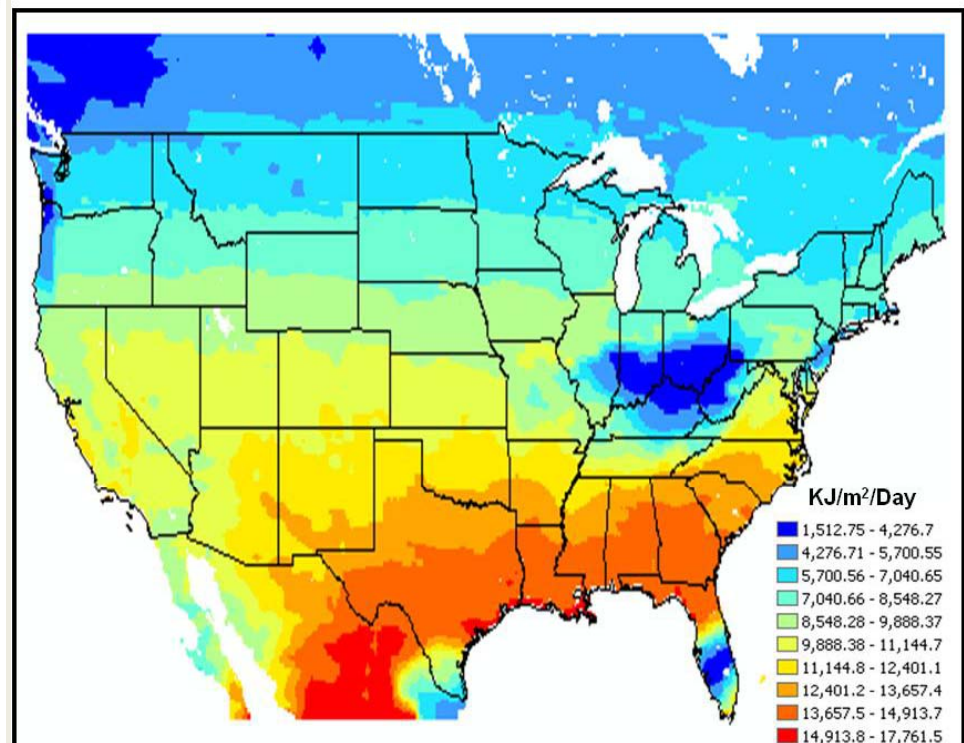
(~14 km spatial resolution)

Environmental Health Data Linkage

(Objective 2)

REasons for Geographic And Racial Differences in Stroke (REGARDS) Study Population

➢ Longitudinal population-based cohort of over 30,000 volunteers age 45 and older

➢ Racial representation
- 50% African American
- 50% white

➢ Sex representation
- 50% male
- 50% female

White
African American

➢ Geographic representation
- 20% from the buckle of the stroke belt
- 30% from the stroke belt
- 50% from the rest of the contiguous US

➢ Successfully transferred from UAB to NASA/MSFC
- BAA as per HIPPA Regulations
- Data Encryption
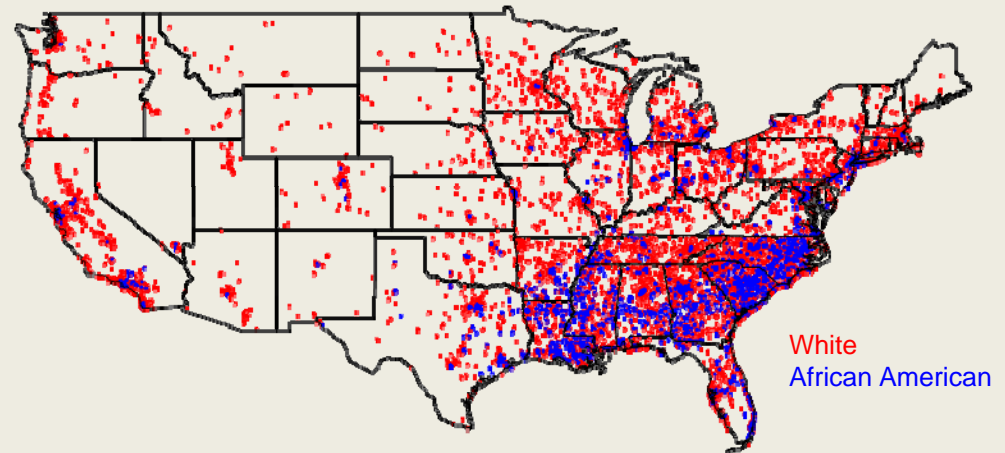
Data Linkage for Biostatistical Analyses

➢ Link in a GIS the estimates of the $PM_{2.5}$, Solar Insolation, and Air Temperature with health data from all participants in the REGARDS study on the individual level at the geographic coordinates of their residences

➢ Sort the environmental data by participant ID, and merge in with the corresponding health data from the REGARDS database

➢ Determine whether exposures to these environmental risk factors are related to cognitive decline and other health outcomes such as hypertension, inflammation, and stroke

Participant ID	Lat	Lon	Day1 Solar Insolation ($KJ/m^2/Day$)	Day2 Solar Insolation ($KJ/m^2/Day$)	Day3 Solar Insolation ($KJ/m^2/Day$)	....	Day365 Solar Insolation ($KJ/m^2/Day$)
1	99.045	-87.105	7950	8941	8945		7850
2	99.055	-89.036	7401	8501	8412		7501
3	99.065	-86.212	8001	7015	8251		8401
.							
.							
.							
.							
.							
30200	99.075	-87.855	15650	11402	15650		10750

Simulated example of the linked data set consisting of participant ID and the associated NLDAS solar insolation

Data Dissemination via CDC WONDER

(Objective 3)

Data Dissemination via CDC WONDER

Tabular Grid-level Daily Data

Grid Cell ID	County, State	FIPS	Day1 Tmax (°F)	Day2 Tmax (°F)	Day3 Tmax (°F)	. . . . . .	Day365 Tmax (°F)
1	Kern, CA	06029	71	74	66		70
2	Kern, CA	06029	70	72	67		69
3	Kern, CA	06029	72	73	66		72
.							
.							
.							
.							
.							
103936	Aroostook, ME	23003	35	31	32		34

Examples of County-level Spatial and Temporal Statistics (Map and Chart) as provided by CDC-WONDER real-time data queries

Ex. July 15, 2007

Spatial Mean (°F)
- 61 - 71
- 72 - 80
- 81 - 90
- 91 - 99
- 100 - 109
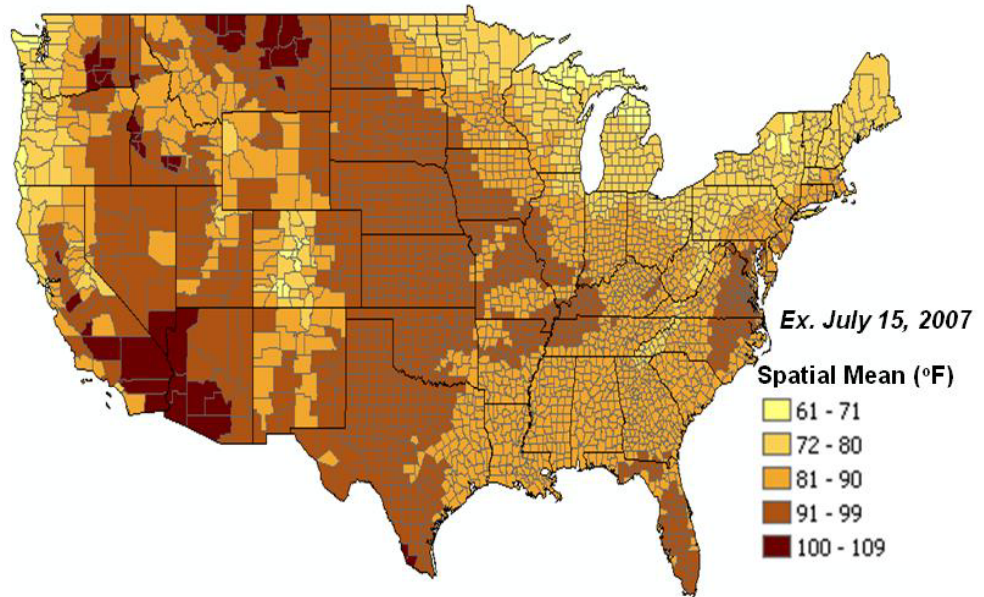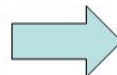
➤ Environmental exposure datasets will be made available to public health professionals, researchers and the general public via WONDER, where they can be aggregated to the county-level or higher as per users' need

➤ Users are able to spatially and temporally query datasets and create county- and higher-level maps and downloadable statistical tables and charts of data across the *contiguous* U.S.

➤ Enabling easy linkage of the environmental exposure data with other health data available via CDC WONDER
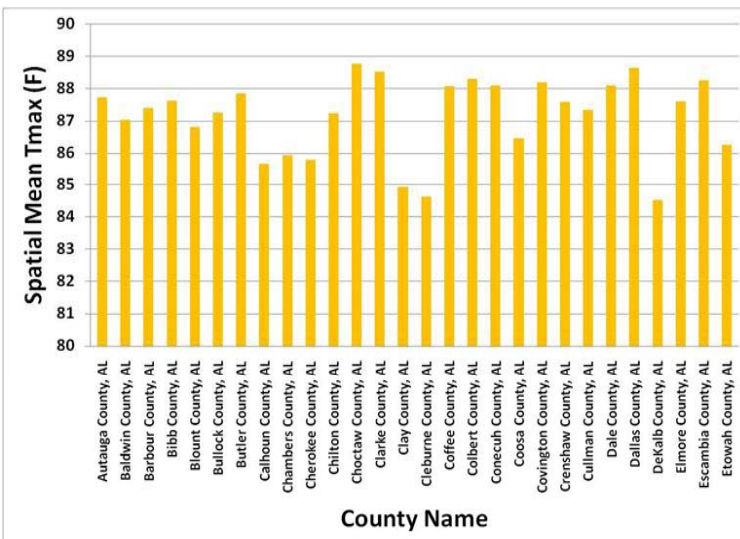
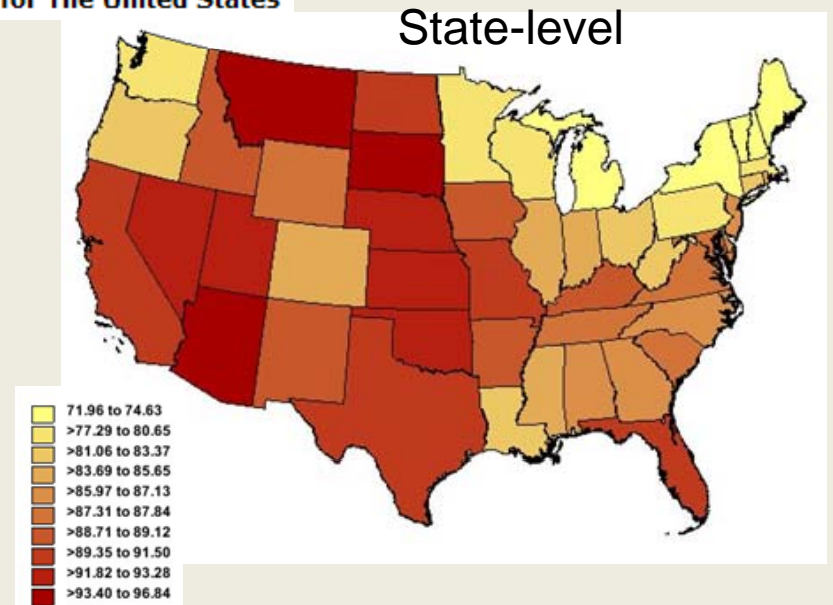| County	Avg Daily Max Air Temperature(F) # of Observations Range Standard Deviation
Autauga County, AL (01001)	87.85 11 (87.20 to 88.40) 0.43
Baldwin County, AL (01003)	85.82 26 (84.30 to 87.20) 0.61
Barbour County, AL (01005)	86.04 14 (85.50 to 86.60) 0.37
Bibb County, AL (01007)	86.92 9 (86.40 to 87.50) 0.31

CDC WONDER Main Web Page

http://wonder.cdc.gov/

NLDAS-derived Heat-related Products on CDC WONDER

Now Available at http://wonder.cdc.gov/nasa-nldas.html

CDC WONDER Spatial Aggregation

Avg Daily Max Air Temperature(F) for The United States

County-level

56.80 to 76.58
>76.60 to 81.69
>81.70 to 84.04
>84.05 to 86.03
>86.04 to 87.64
>87.65 to 89.05
>89.06 to 90.28
>90.29 to 91.58
>91.59 to 93.76
>93.77 to 109.09

State-level

71.96 to 74.63
>77.29 to 80.65
>81.06 to 83.37
>83.69 to 85.65
>85.97 to 87.13
>87.31 to 87.84
>88.71 to 89.12
>89.35 to 91.50
>91.82 to 93.28
>93.40 to 96.84

July 15, 2007

Division-level

75.14
76.87
79.37
85.26
86.53
87.07
89.60
90.36
90.88

Region-level

76.17
86.30
88.55
89.33
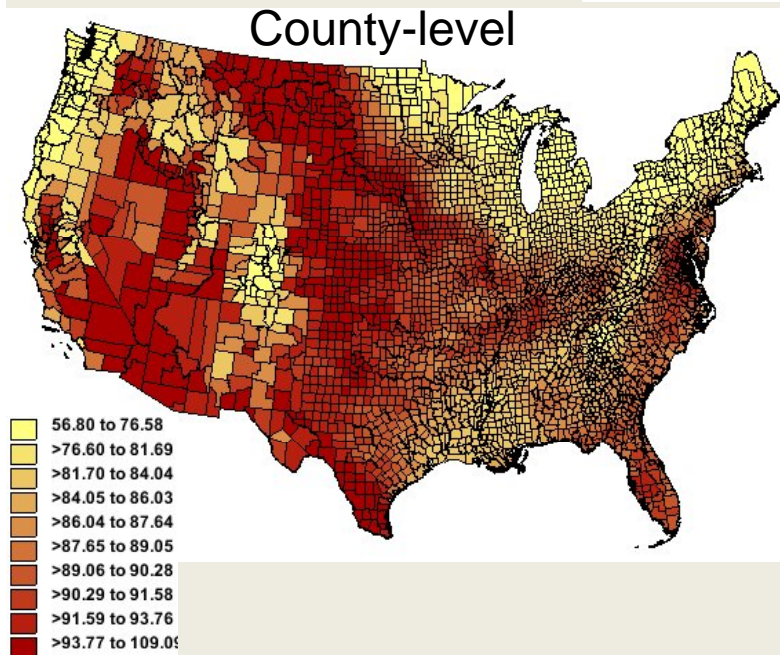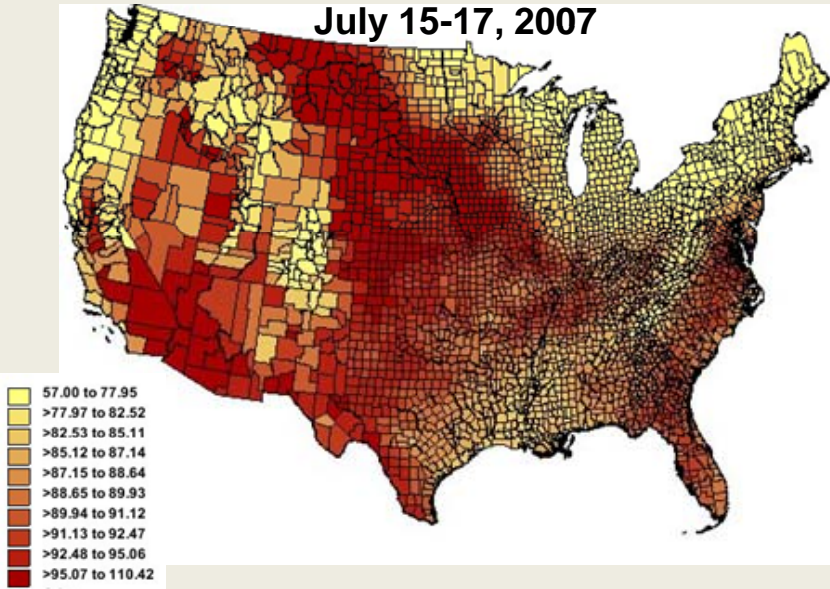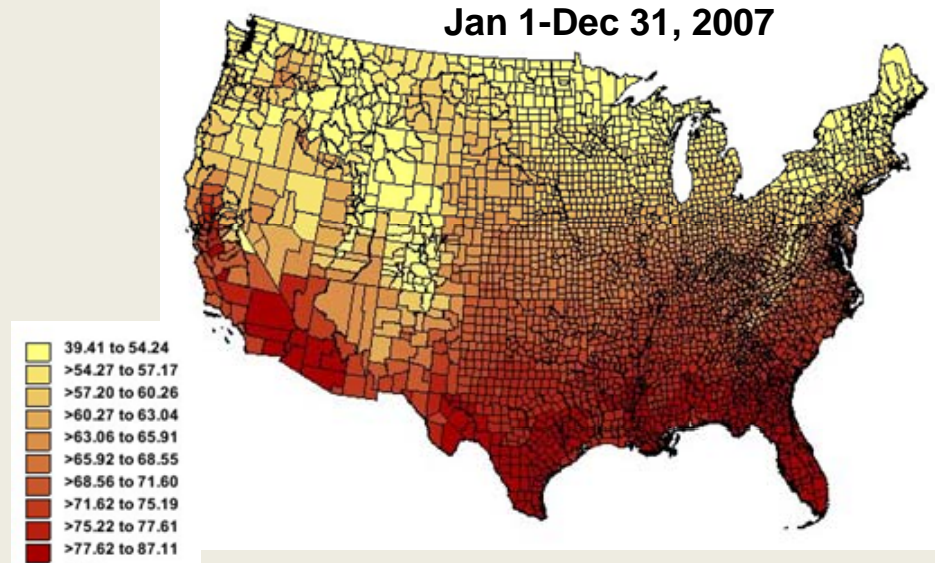
CDC WONDER Spatial/Temporal Aggregation

Avg Daily Max Air Temperature(F) for The United States

July 15-17, 2007

Jan 1-Dec 31, 2007

57.00 to 77.95
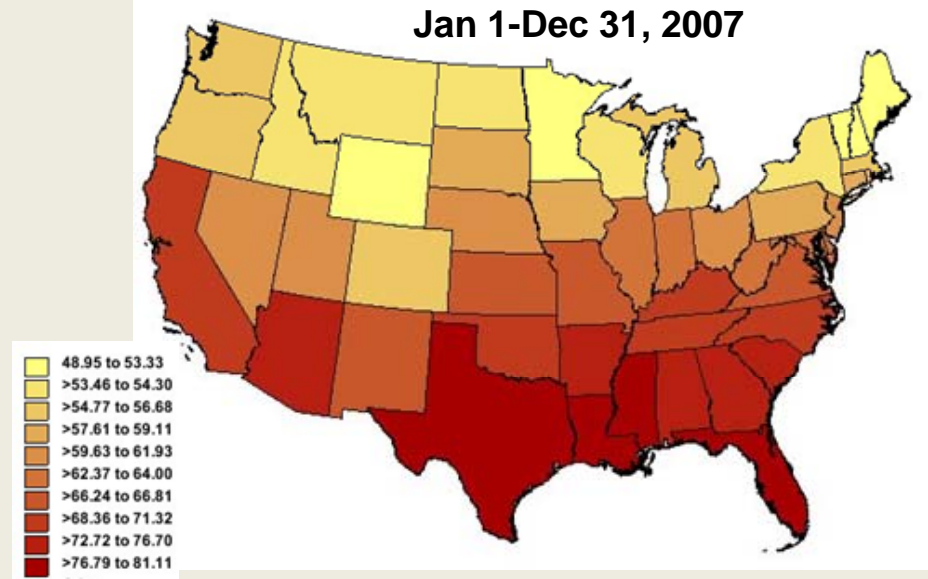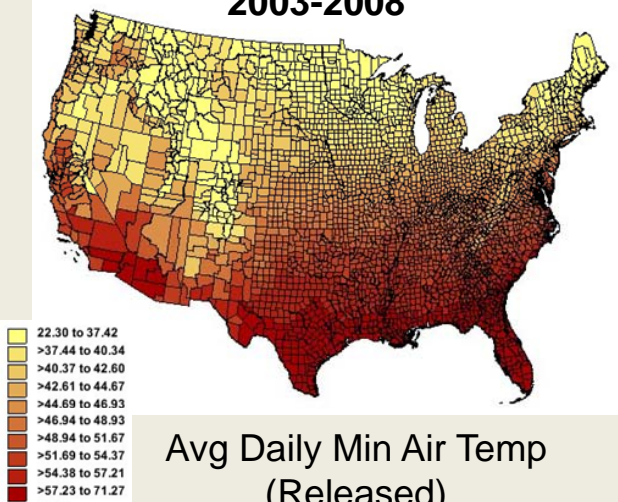>77.97 to 82.52
>82.53 to 85.11
>85.12 to 87.14
>87.15 to 88.64
>88.65 to 89.93
>89.94 to 91.12
>91.13 to 92.47
>92.48 to 95.06
>95.07 to 110.42

39.41 to 54.24
>54.27 to 57.17
>57.20 to 60.26
>60.27 to 63.04
>63.06 to 65.91
>65.92 to 68.55
>68.56 to 71.60
>71.62 to 75.19
>75.22 to 77.61
>77.62 to 87.11

July 15-17, 2007

Jan 1-Dec 31, 2007

72.94 to 75.58
>78.84 to 81.37
>81.48 to 84.16
>84.35 to 86.08
>86.58 to 87.56
>87.60 to 88.73
>88.93 to 89.45
>89.68 to 91.00
>91.28 to 94.01
>94.19 to 98.42

48.95 to 53.33
>53.46 to 54.30
>54.77 to 56.68
>57.61 to 59.11
>59.63 to 61.93
>62.37 to 64.00
>66.24 to 66.81
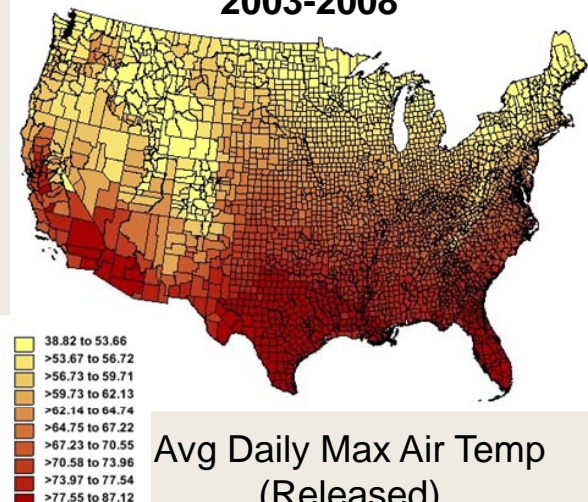>68.36 to 71.32
>72.72 to 76.70
>76.79 to 81.11

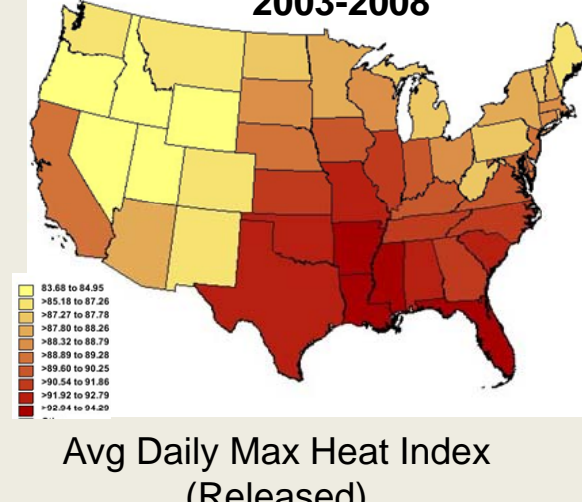# CDC WONDER Environmental Datasets Status

2003-2008

Avg Daily Min Air Temp
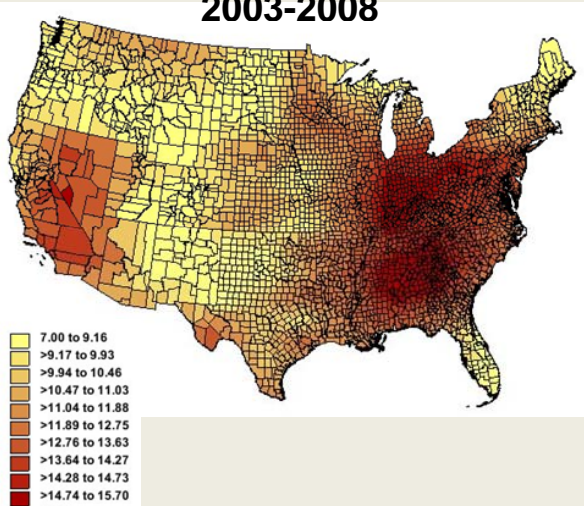(Released)

2003-2008

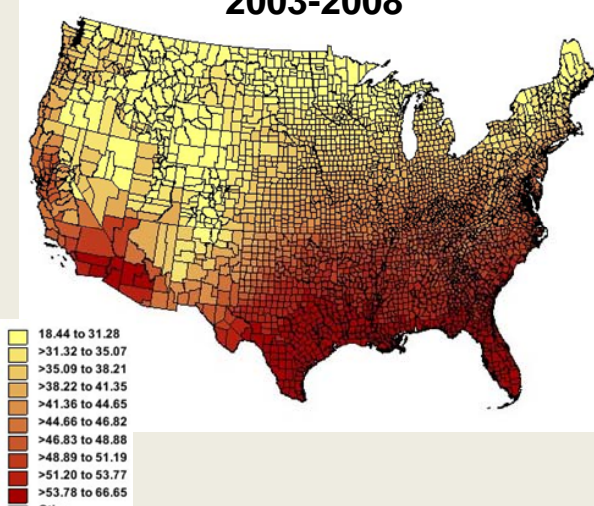Avg Daily Max Air Temp
(Released)

2003-2008

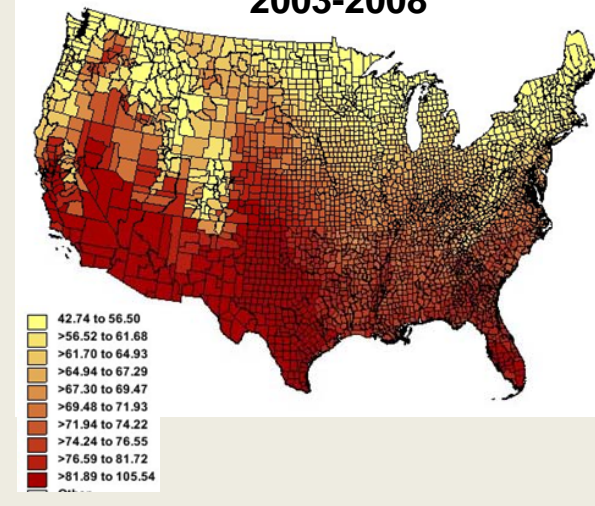Avg Daily Max Heat Index
(Released)

2003-2008

Avg PM2.5
(Will be released by March, 2012)

2003-2008

Avg Night MODIS LST
(Will be released by March, 2012)

2003-2008

Avg Day MODIS LST
(Will be released by March, 2012)

Summary

- Development of national daily products of $PM_{2.5}$, LST, maximum and minimum air temperature, maximum heat index, and solar insolation for 2003-2008

- Linkages of these data with public health data from the REGARDS national cohort study for environmental health correlation studies

- Dissemination of these environmental datasets to public health professionals, researchers and the general public via the CDC WONDER online system
  - Maximum and minimum air temperature and maximum heat index datasets *have been released at* **http://wonder.cdc.gov/nasa-nldas.html**
  - $PM_{2.5}$ and LST datasets will be released by March, 2012

- Providing a useful addition to CDC WONDER, allowing public health researchers and policy makers to better include environmental exposure data in the context of other health data available in CDC WONDER online system

- Substantially expanding public access to these NASA environmental datasets, making their use by a wide range of decision makers more feasible

Thanks!

Presenter's Contact Information:

Mohammad Z. Al-Hamdan, PhD

mohammad.alhamdan@nasa.gov