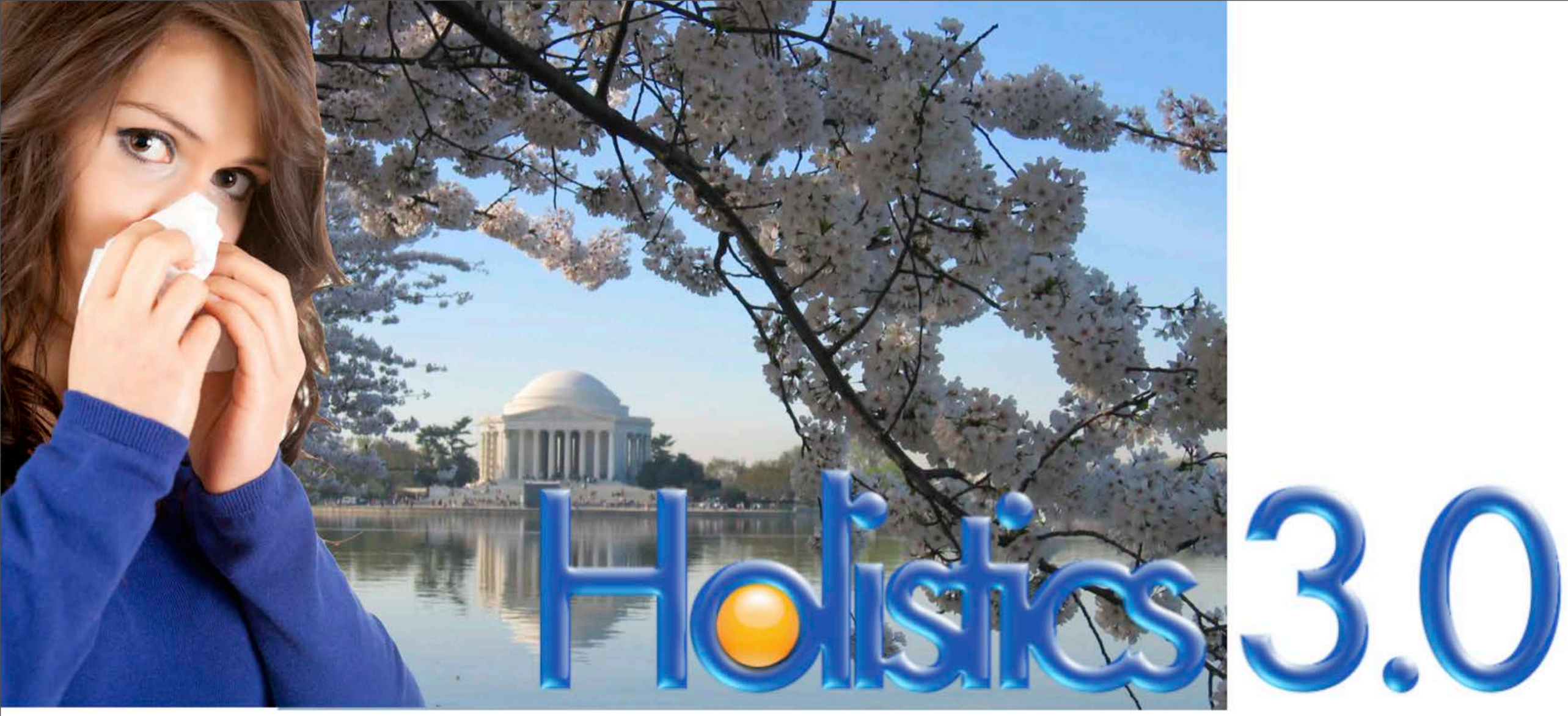


http://holistics3.com

David.Lary@utdallas.edu

Comprehensive Data

Machine Learning

Holistics 3.0

Comprehensive Big Data

+

Machine Learning

+

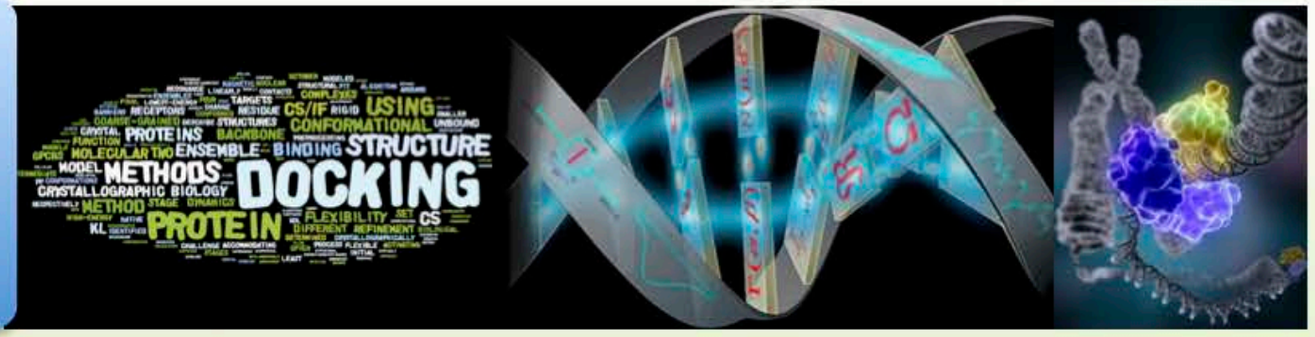
Causality

→ Holistics 3.0

California Children Example

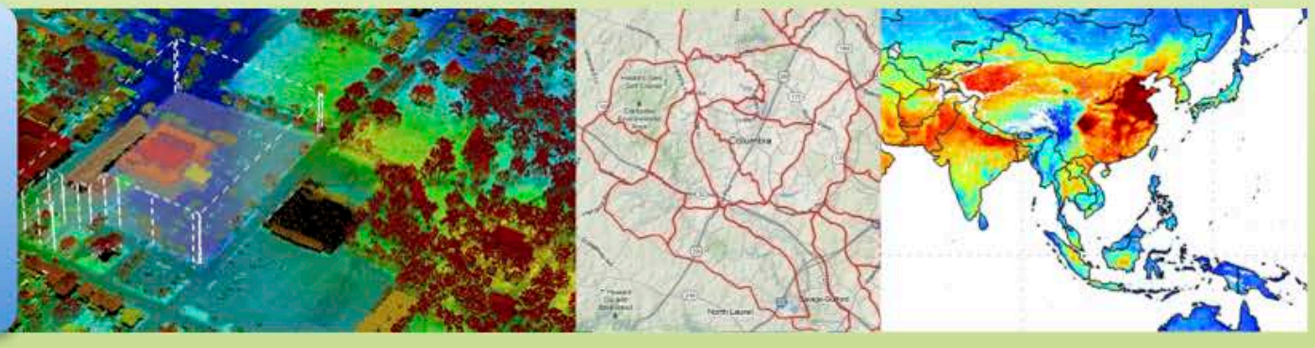
Think Big: Holistic & Comprehensive Informatics

Bio Informatics



Medical Informatics

Environmental Informatics



Holistic Informatics:

Multiple Big Data + EMR + Social Media + Machine Learning + Causality

http://holistics3.com

Prof David Lary

MODIS Aqua July 21, 2013.

North Sumatra

Malacca Strait

Malaysia

Rupat Island

Riau

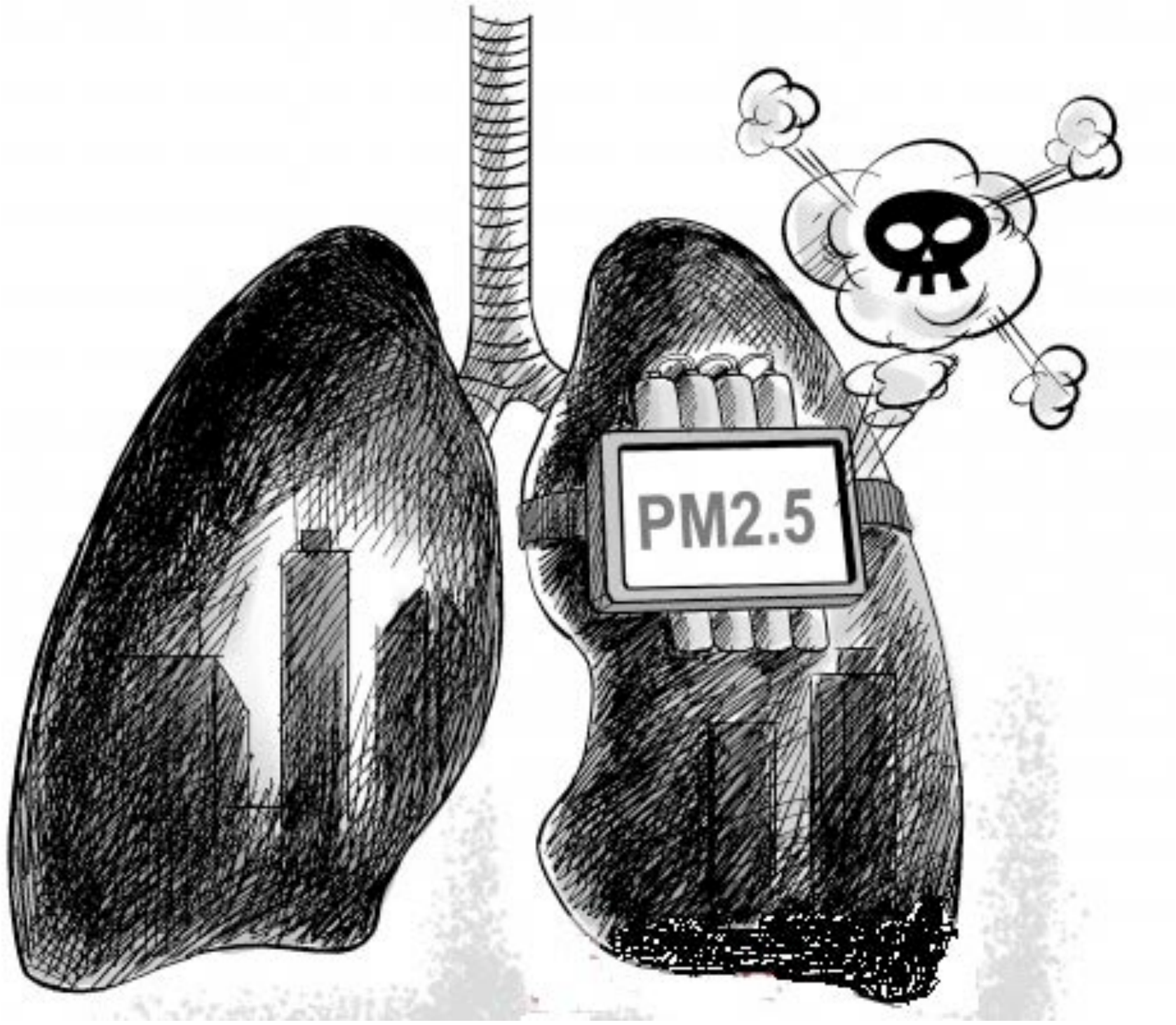
David Lary

Unprecedented levels of air pollution in Singapore and Malaysia in June led to respiratory illnesses, school closings, and grounded aircraft. This year it was so bad that in some affected areas there was a 100 percent rise in the number of asthma cases, and the government of Malaysia distributed gas masks.

Air pollution in Ulaanbaatar, Mongolia

Thursday, October 24, 13

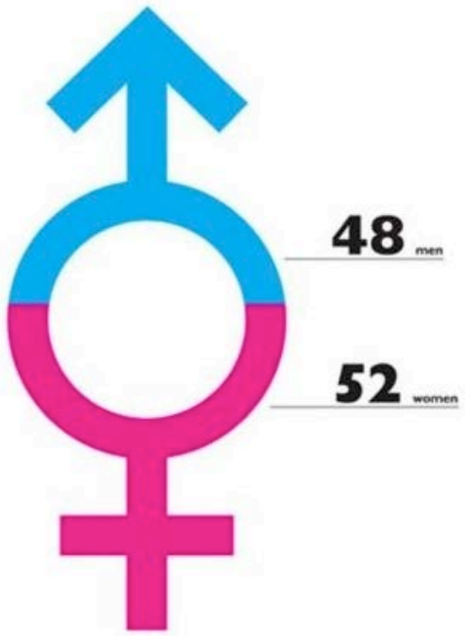
PM2.5 Invisible Killer



Outdoor Air Pollution Kills 2 Million People A Year

If the world were a village of 100 people

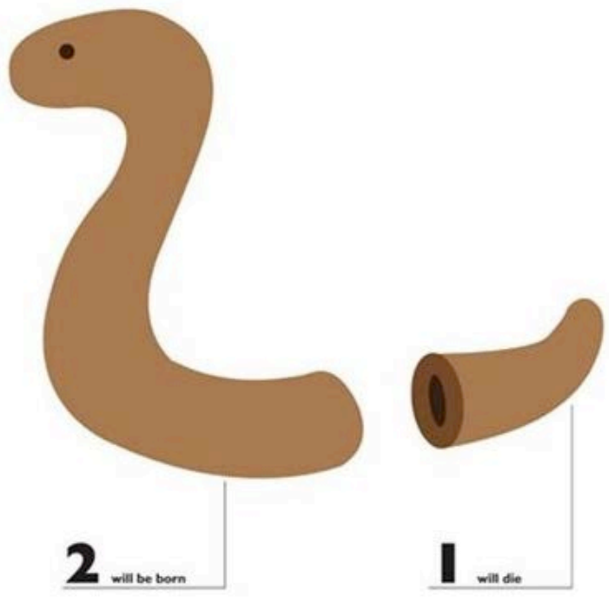
GENDER



If the world were a village of 100 people

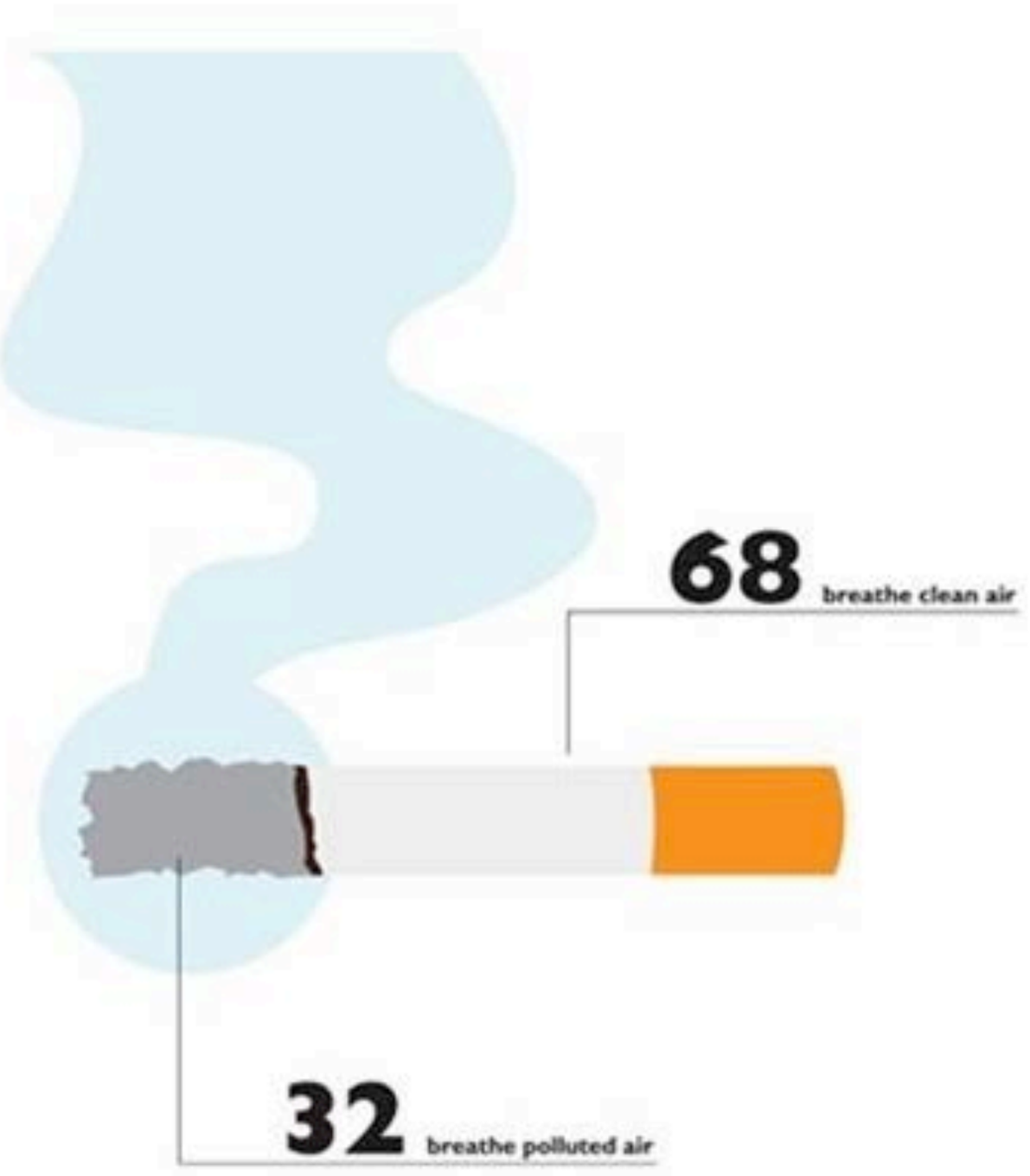
BIRTHS/DEATHS

(IN ONE YEAR)

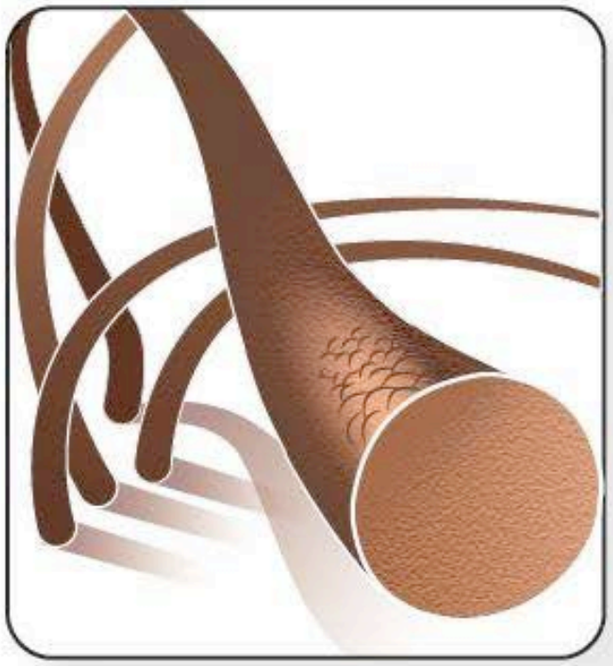


If the world were a village of 100 people

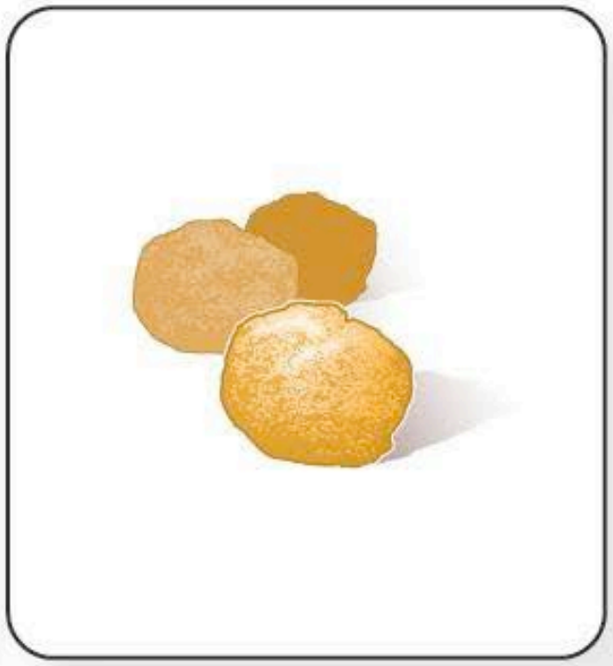
AIR



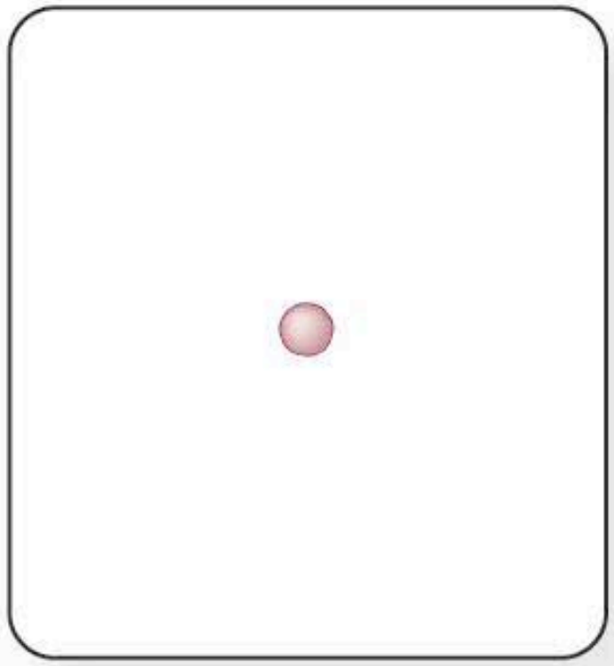
Fine Particulate Matter Size Comparison



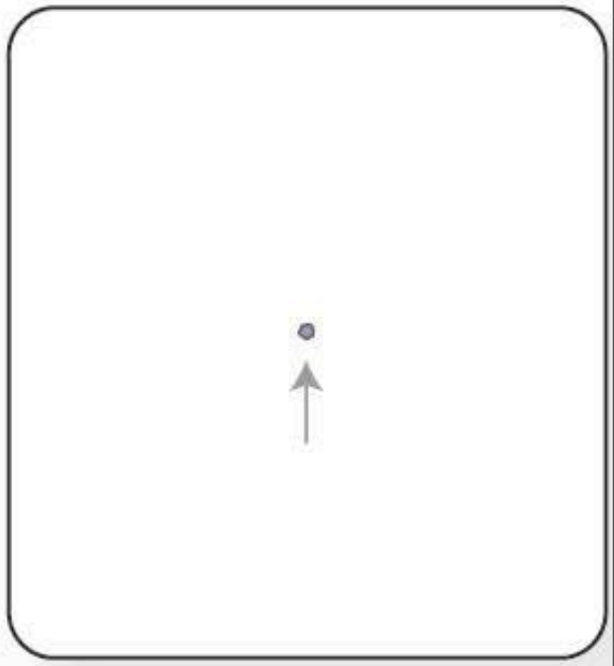
Human hair (about 70µm wide)



Grain of sand (about 50µm wide)



PM_{10} (less than 10µm wide)



$PM_{2.5}$ (less than 2.5µm wide)

µm = micrometer

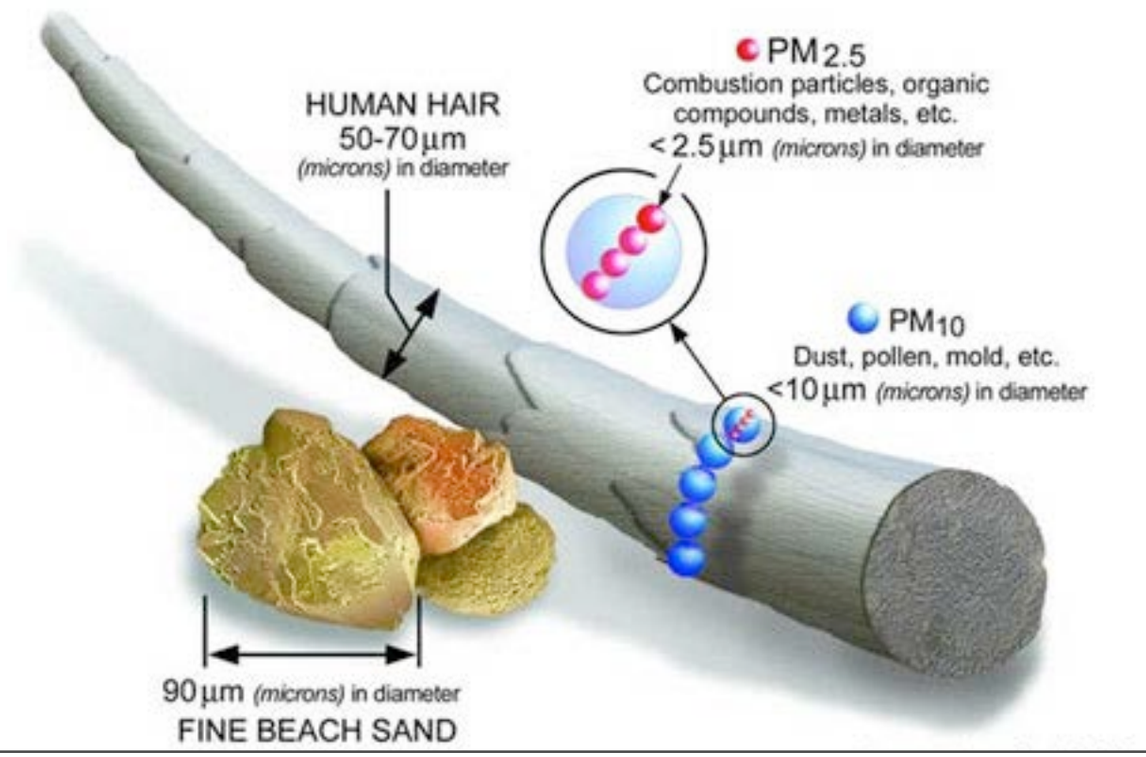
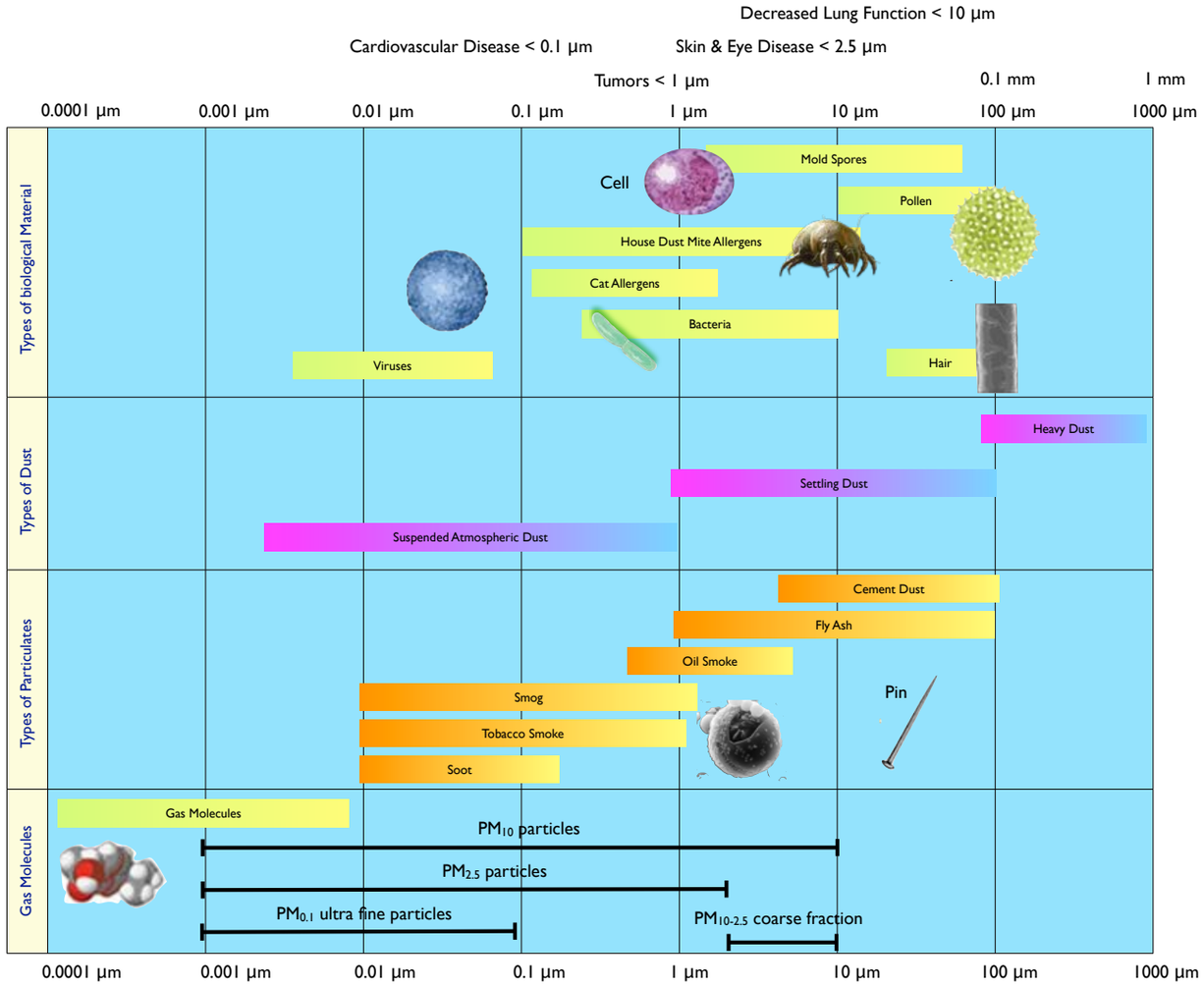


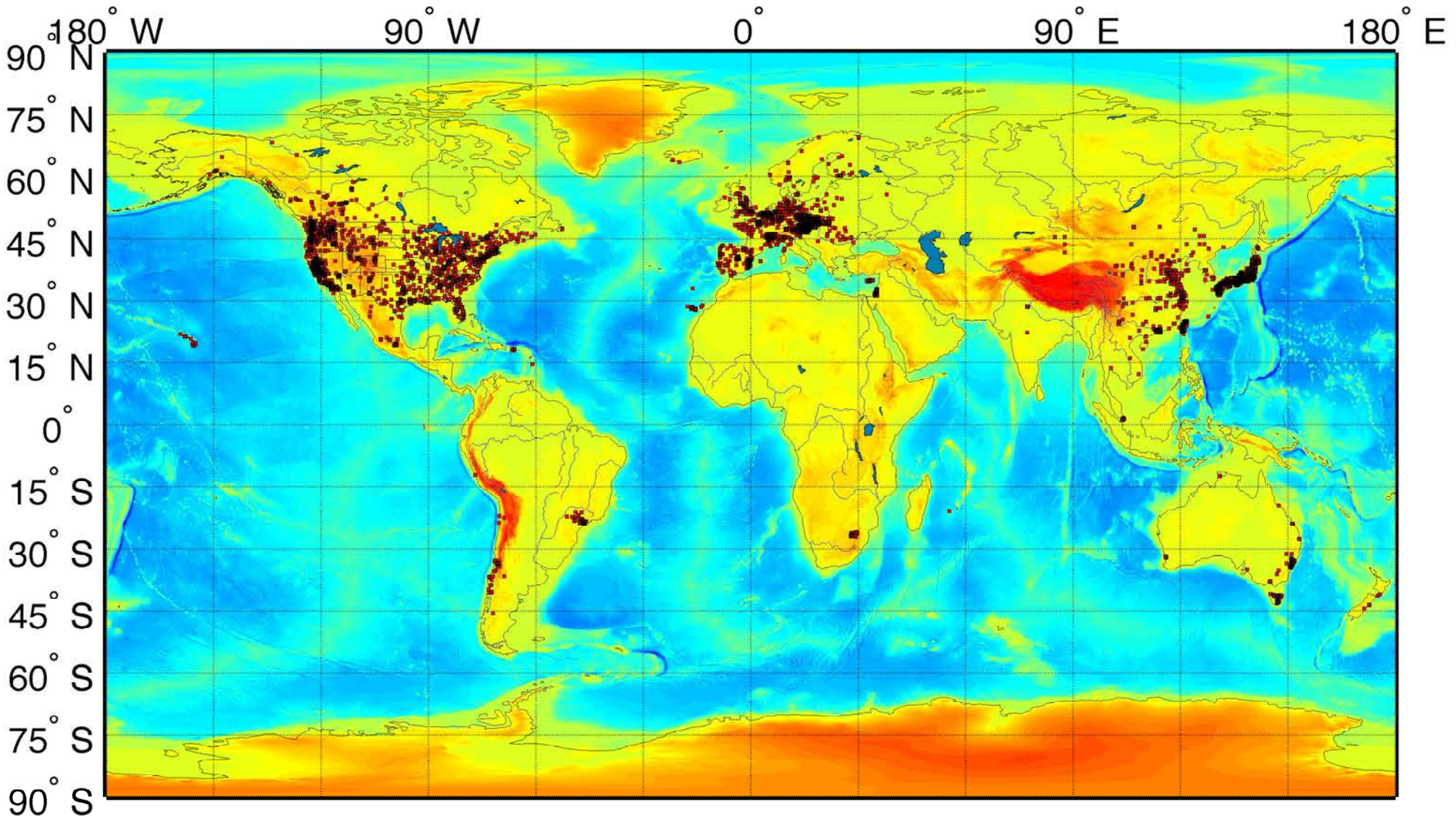
Table 1. PM and health outcomes (modified from *Ruckerl et al.* (2006)).

Health Outcomes	Short-term Studies			Long-term Studies		
	PM10	PM2.5	UFP	PM10	PM2.5	UFP
Mortality						
All causes	xxx	xxx	x	xx	xx	x
Cardiovascular	xxx	xxx	x	xx	xx	x
Pulmonary	xxx	xxx	x	xx	xx	x
Pulmonary effects						
Lung function, e.g., PEF	xxx	xxx	xx	xxx	xxx	
Lung function growth				xxx	xxx	
Asthma and COPD exacerbation						
Acute respiratory symptoms		xx	x	xxx	xxx	
Medication use			x			
Hospital admission	xx	xxx	x			
Lung cancer						
Cohort				xx	xx	x
Hospital admission				xx	xx	x
Cardiovascular effects						
Hospital admission	xxx	xxx		x	x	
ECG-related endpoints						
Autonomic nervous system	xxx	xxx	xx			
Myocardial substrate and vulnerability		xx	x			
Vascular function						
Blood pressure	xx	xxx	x			
Endothelial function	x	xx	x			
Blood markers						
Pro inflammatory mediators	xx	xx	xx			
Coagulation blood markers	xx	xx	xx			
Diabetes	x	xx	x			
Endothelial function	x	x	xx			
Reproduction						
Premature birth	x	x				
Birth weight	xx	x				
IUR/SGA	x	x				
Fetal growth						
Birth defects	x					
Infant mortality	xx	x				
Sperm quality	x	x				
Neurotoxic effects						
Central nervous system		x	xx			

x, few studies; xx, many studies; xxx, large number of studies.



Hourly Observations from 52 Countries from 1996-present at a total of 6931 $PM_{2.5}$ Measuring Sites



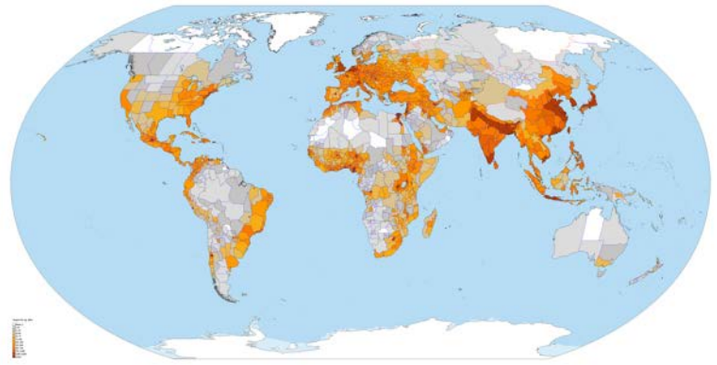
PM2.5 - AOD Relationship

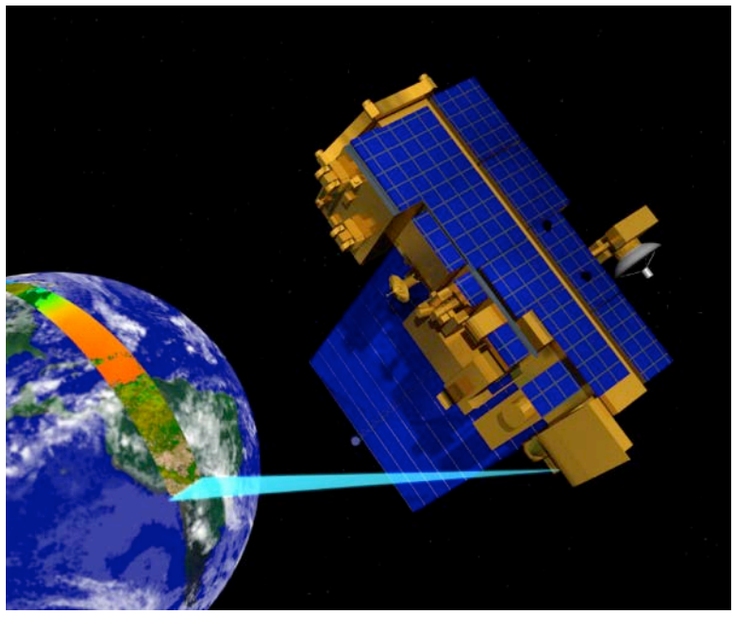
Non-linear, function of wind speed, T, P, humidity, surface type, etc and prone to large biases in MODIS AOD algorithm due to surface reflectivity.

Literature studies have major issue in Western US (typical correlation coefficient 0.2) better in Eastern US (typical correlation coefficient 0.6)

Terra DeepBlue

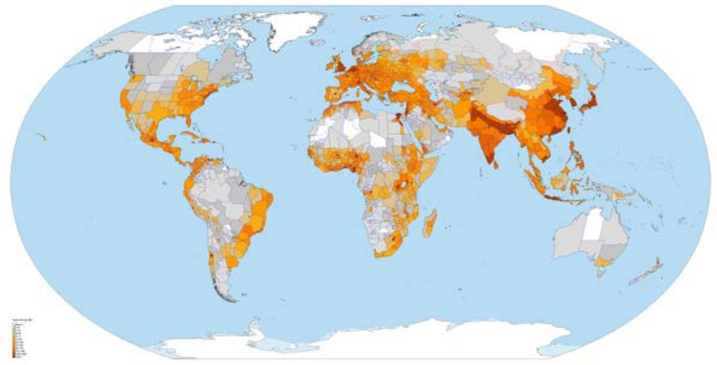
Rank	Source	Variable	Type
1		Population Density	Input
2	Satellite Product	Tropospheric NO_2 Column	Input
3	Meteorological Analyses	Surface Specific Humidity	Input
4	Satellite Product	Solar Azimuth	Input
5	Meteorological Analyses	Surface Wind Speed	Input
6	Satellite Product	White-sky Albedo at 2,130 nm	Input
7	Satellite Product	White-sky Albedo at 555 nm	Input
8	Meteorological Analyses	Surface Air Temperature	Input
9	Meteorological Analyses	Surface Layer Height	Input
10	Meteorological Analyses	Surface Ventilation Velocity	Input
11	Meteorological Analyses	Total Precipitation	Input
12	Satellite Product	Solar Zenith	Input
13	Meteorological Analyses	Air Density at Surface	Input
14	Satellite Product	Cloud Mask Qa	Input
15	Satellite Product	Deep Blue Aerosol Optical Depth 470 nm	Input
16	Satellite Product	Sensor Zenith	Input
17	Satellite Product	White-sky Albedo at 858 nm	Input
18	Meteorological Analyses	Surface Velocity Scale	Input
19	Satellite Product	White-sky Albedo at 470 nm	Input
20	Satellite Product	Deep Blue Angstrom Exponent Land	Input
21	Satellite Product	White-sky Albedo at 1,240 nm	Input
22	Satellite Product	Scattering Angle	Input
23	Satellite Product	Sensor Azimuth	Input
24	Satellite Product	Deep Blue Surface Reflectance 412 nm	Input
25	Satellite Product	White-sky Albedo at 1,640 nm	Input
26	Satellite Product	Deep Blue Aerosol Optical Depth 660 nm	Input
27	Satellite Product	White-sky Albedo at 648 nm	Input
28	Satellite Product	Deep Blue Surface Reflectance 660 nm	Input
29	Satellite Product	Cloud Fraction Land	Input
30	Satellite Product	Deep Blue Surface Reflectance 470 nm	Input
31	Satellite Product	Deep Blue Aerosol Optical Depth 550 nm	Input
32	Satellite Product	Deep Blue Aerosol Optical Depth 412 nm	Input
	In-situ Observation	**$PM_{2.5}$**	**Target**

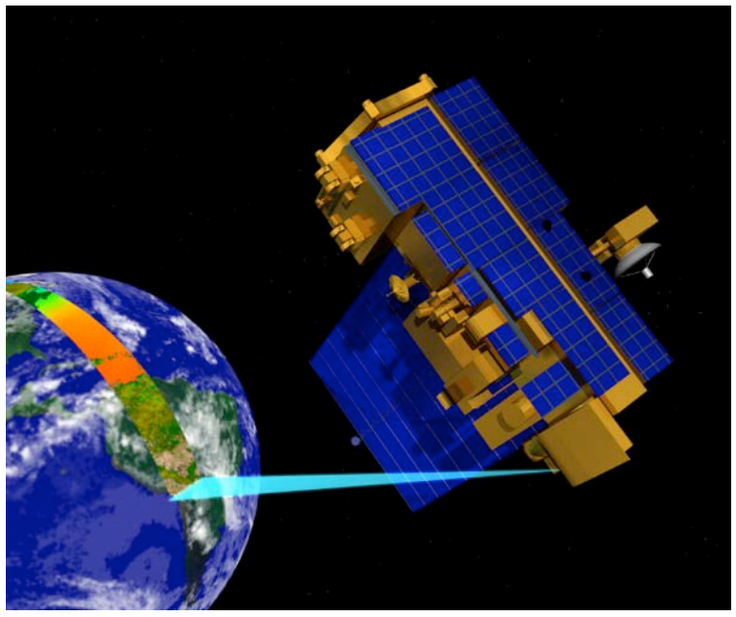




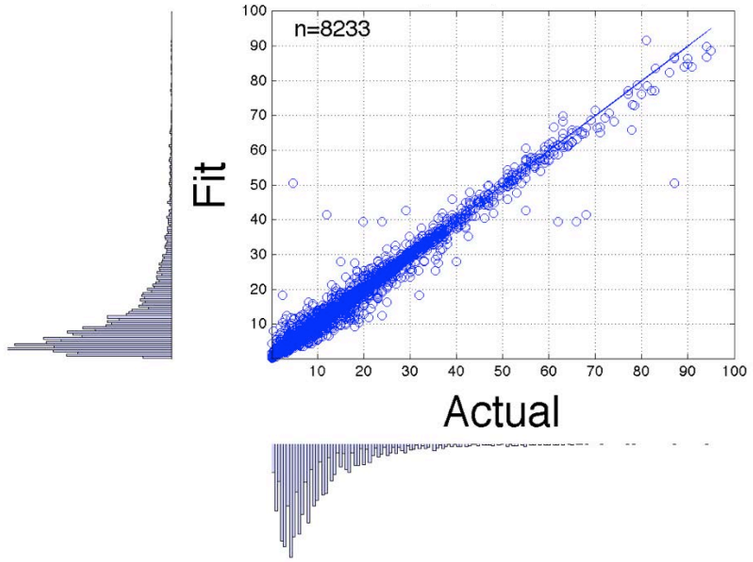
Aqua DeepBlue

Rank	Source	Variable	Type
1	Satellite Product	Tropospheric NO_2 Column	Input
2	Satellite Product	Solar Azimuth	Input
3	Meteorological Analyses	Air Density at Surface	Input
4	Satellite Product	Sensor Zenith	Input
5	Satellite Product	White-sky Albedo at 470 nm	Input
6		Population Density	Input
7	Satellite Product	Deep Blue Surface Reflectance 470 nm	Input
8	Meteorological Analyses	Surface Air Temperature	Input
9	Meteorological Analyses	Surface Ventilation Velocity	Input
10	Meteorological Analyses	Surface Wind Speed	Input
11	Satellite Product	White-sky Albedo at 858 nm	Input
12	Satellite Product	White-sky Albedo at 2,130 nm	Input
13	Satellite Product	Solar Zenith	Input
14	Meteorological Analyses	Surface Layer Height	Input
15	Satellite Product	White-sky Albedo at 1,240 nm	Input
16	Satellite Product	Deep Blue Surface Reflectance 660 nm	Input
17	Satellite Product	Deep Blue Surface Reflectance 412 nm	Input
18	Satellite Product	White-sky Albedo at 1,640 nm	Input
19	Satellite Product	Sensor Azimuth	Input
20	Satellite Product	Scattering Angle	Input
21	Meteorological Analyses	Surface Velocity Scale	Input
22	Satellite Product	Cloud Mask Qa	Input
23	Satellite Product	White-sky Albedo at 555 nm	Input
24	Satellite Product	Deep Blue Aerosol Optical Depth 550 nm	Input
25	Satellite Product	Deep Blue Aerosol Optical Depth 660 nm	Input
26	Satellite Product	Deep Blue Aerosol Optical Depth 412 nm	Input
27	Meteorological Analyses	Total Precipitation	Input
28	Satellite Product	White-sky Albedo at 648 nm	Input
29	Satellite Product	Deep Blue Aerosol Optical Depth 470 nm	Input
30	Satellite Product	Deep Blue Angstrom Exponent Land	Input
31	Meteorological Analyses	Surface Specific Humidity	Input
32	Satellite Product	Cloud Fraction Land	Input
	In-situ Observation	**$PM_{2.5}$**	**Target**

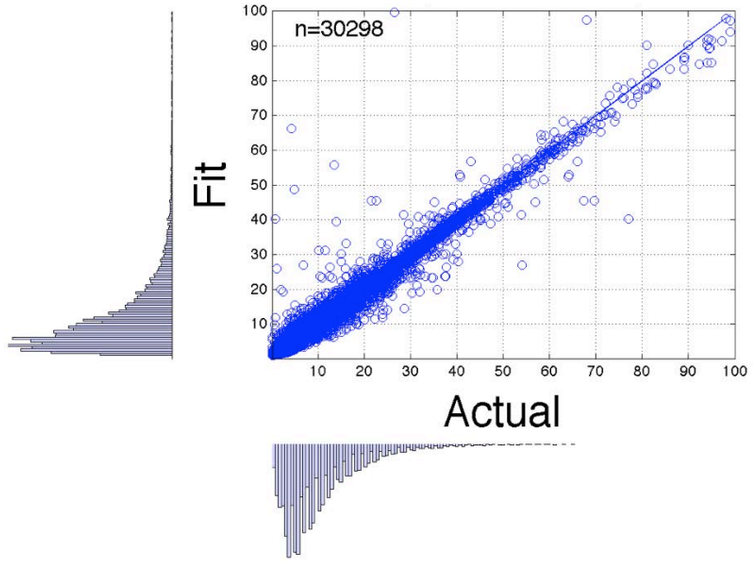




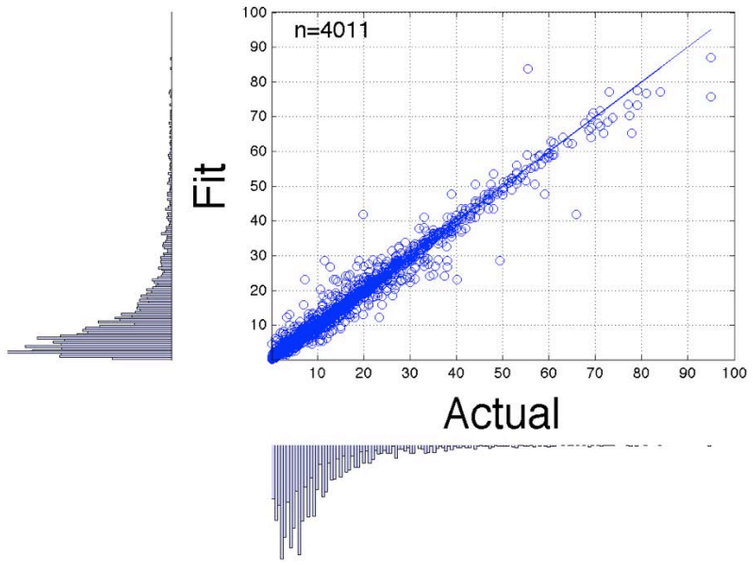
Aqua DeepBlue, R=0.99, R^2=0.98



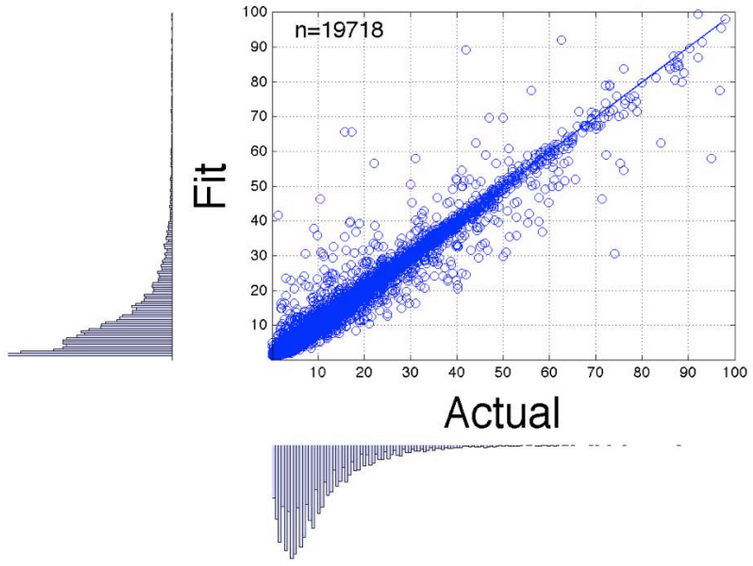
Aqua Standard, R=0.99, R^2=0.98



Terra DeepBlue, R=0.99, R^2=0.99



Terra Standard, R=0.98, R^2=0.97



Training dataset statistics and global 2000-2012 correlation coefficients.

	n	R	R^2
Aqua Deep Blue	8,233	0.99	0.98
Aqua Standard	30,298	0.99	0.98
Terra Deep Blue	4,011	0.99	0.98
Terra Standard	19,718	0.98	0.97

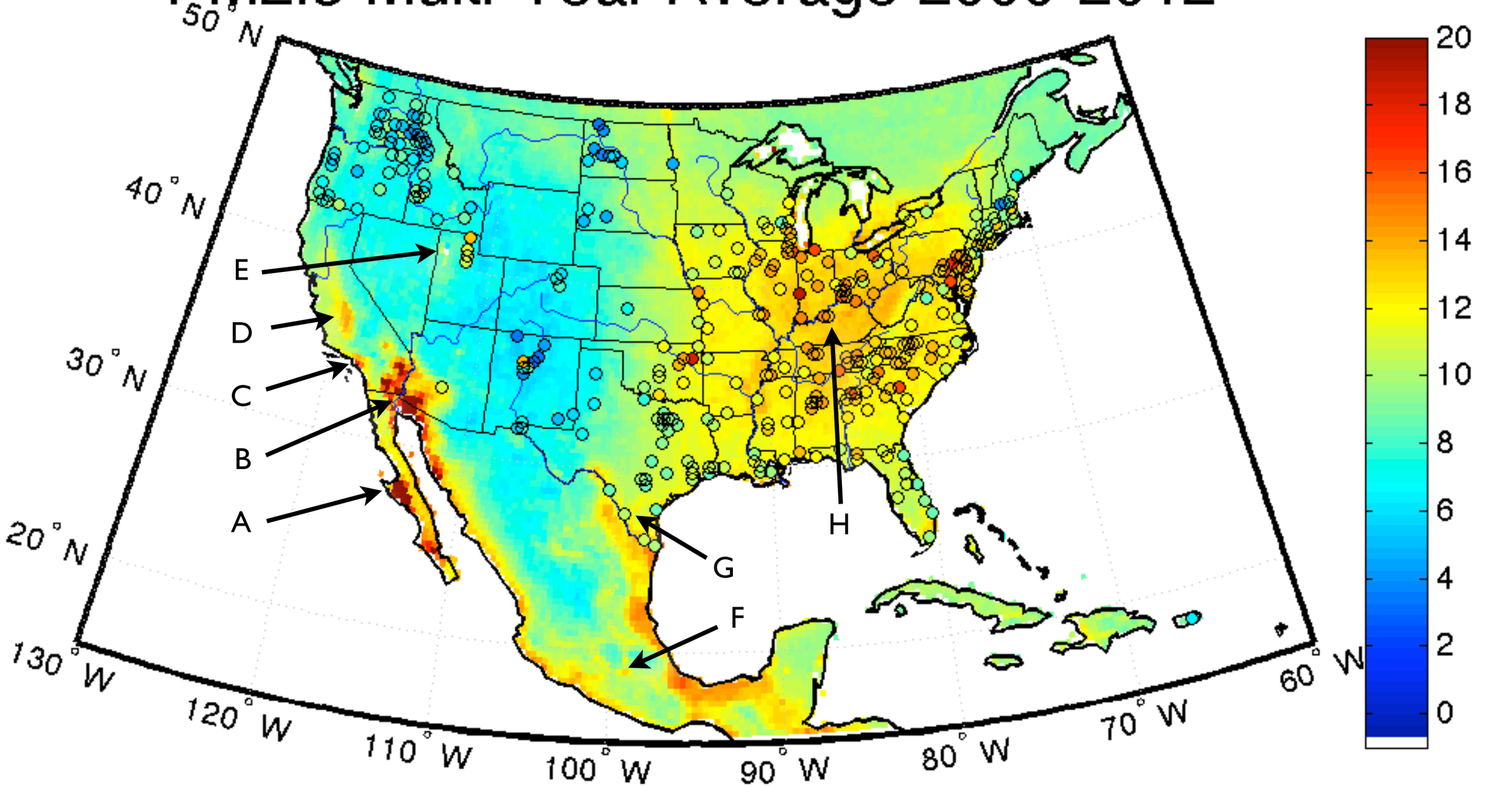
Machine Learning Estimate of $PM_{2.5}$

Combine a total of about 30 variables

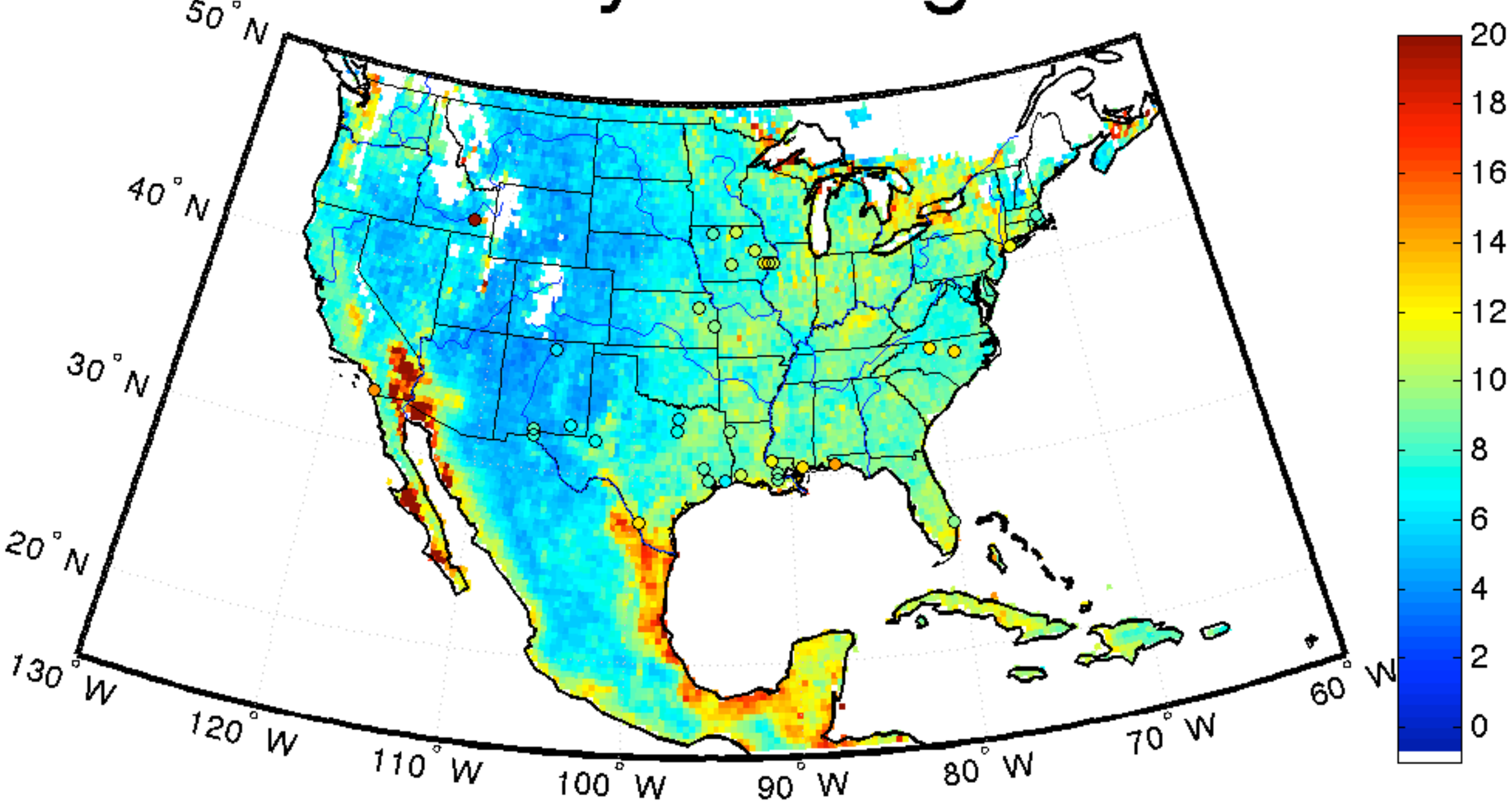
- In-situ observations
- Population Density
- Meteorological Variables
- Satellite Variables

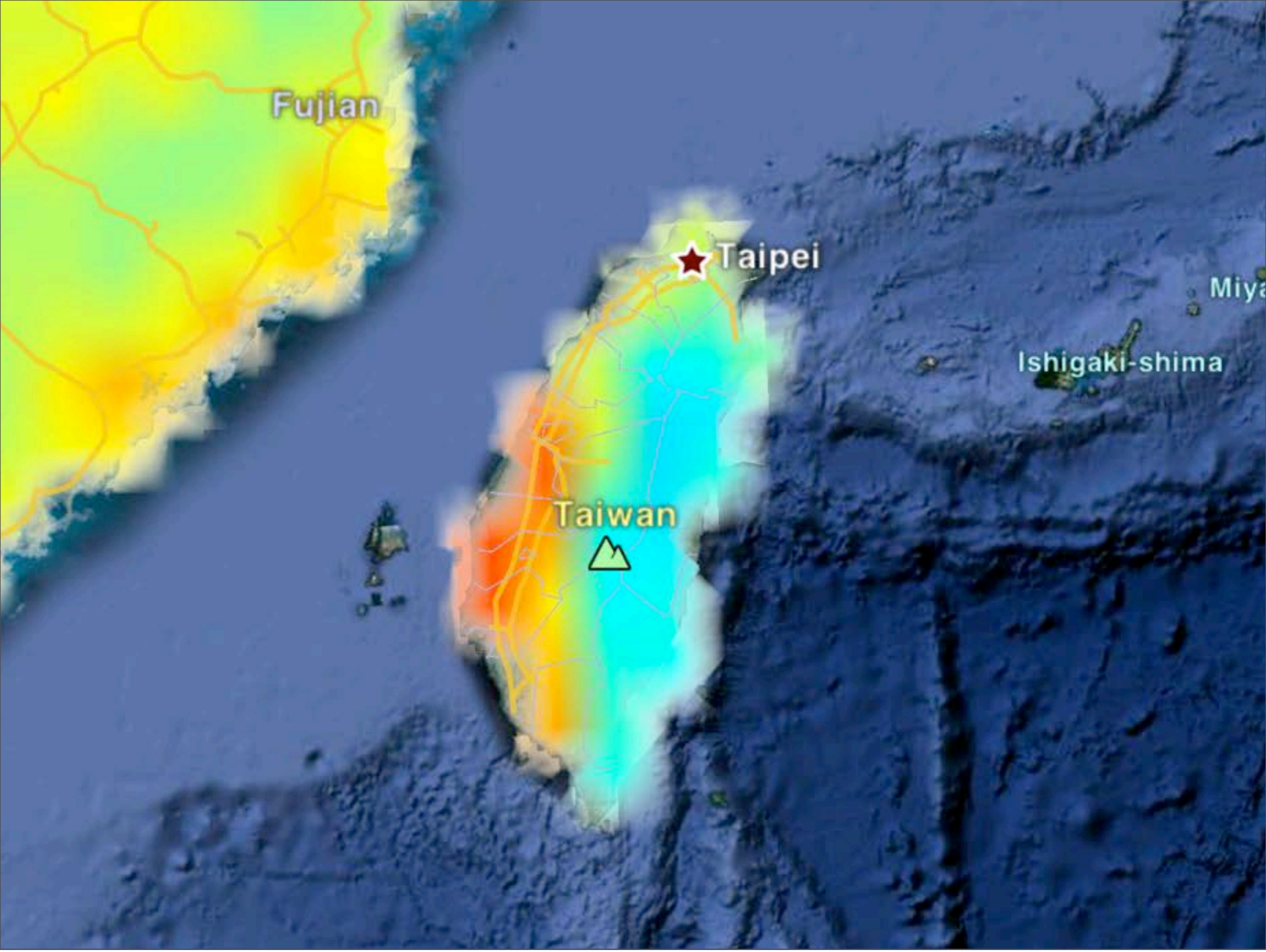
Globally, every day from March 1, 2000-present

PM2.5 Multi-Year Average 2000-2012



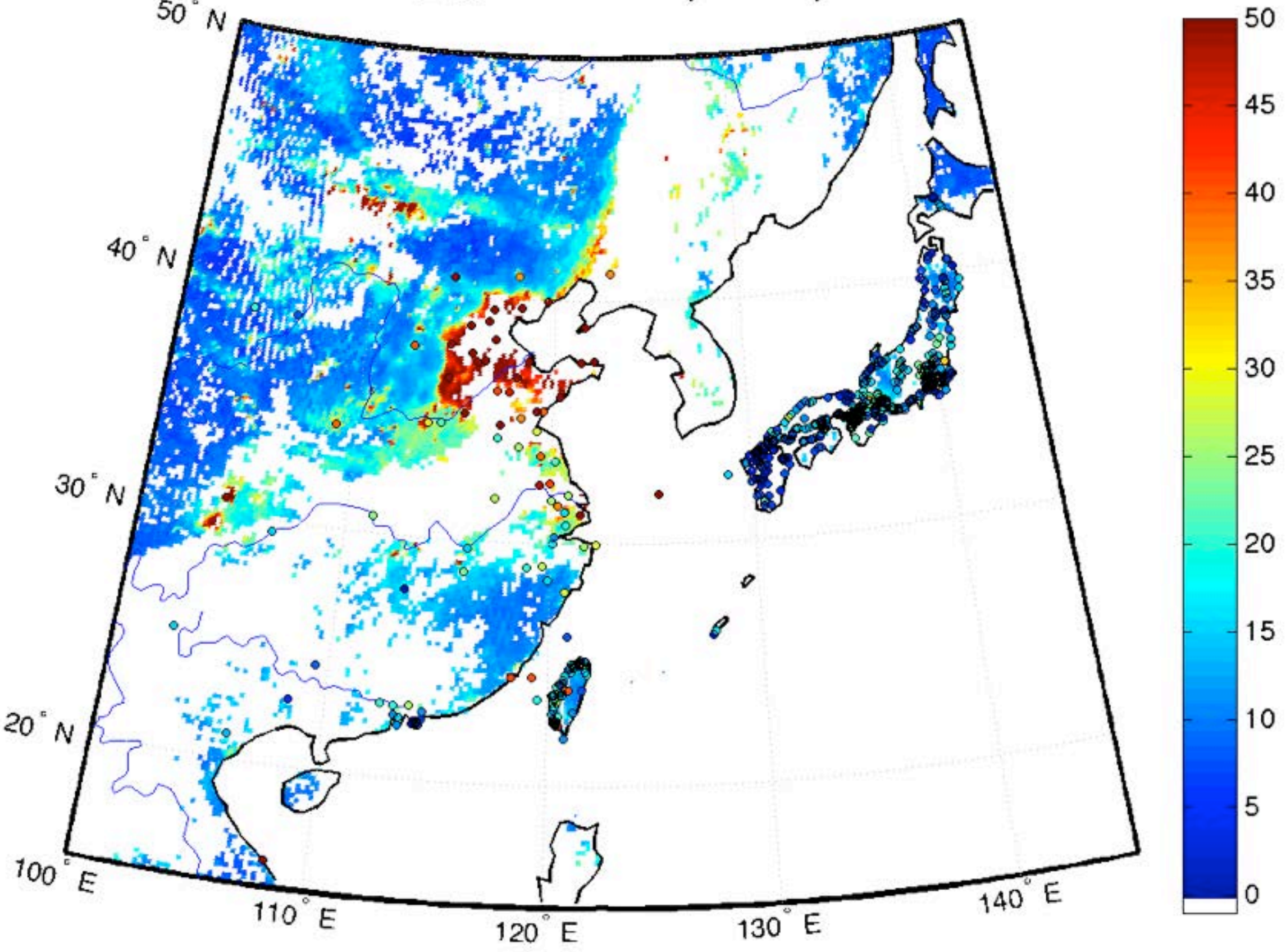
PM2.5 Monthly Average 2000/03



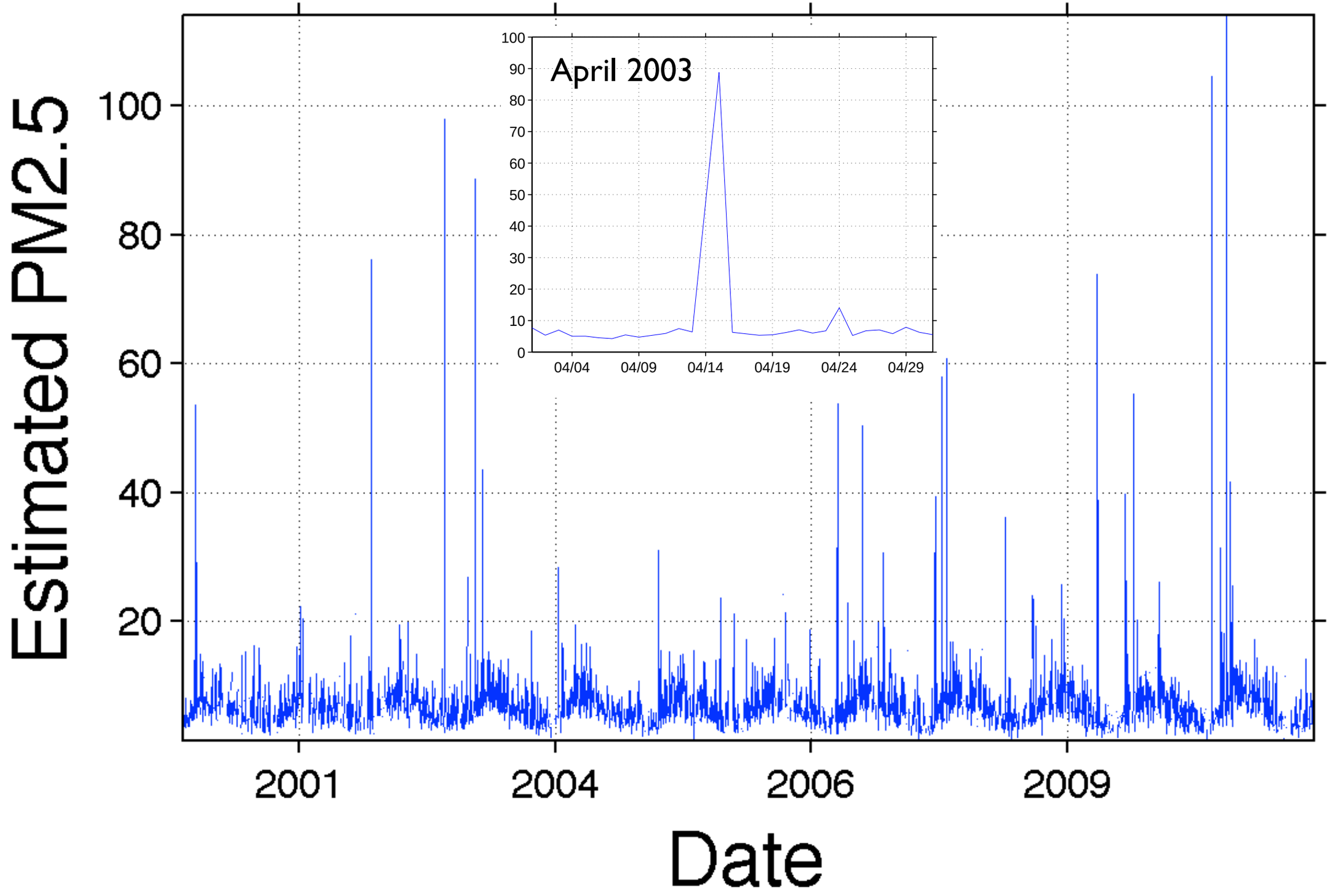


Thursday, October 24, 13

$PM_{2.5}$ 2013/08/04



El Paso, United States of America
2010 Population 0.8 Million

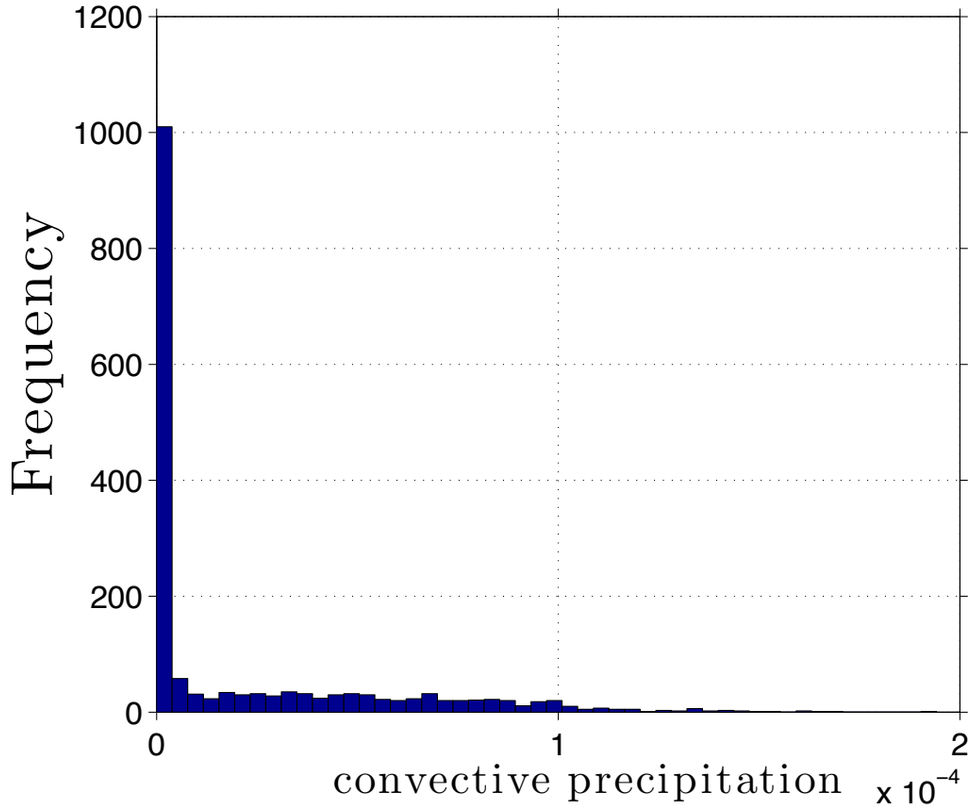


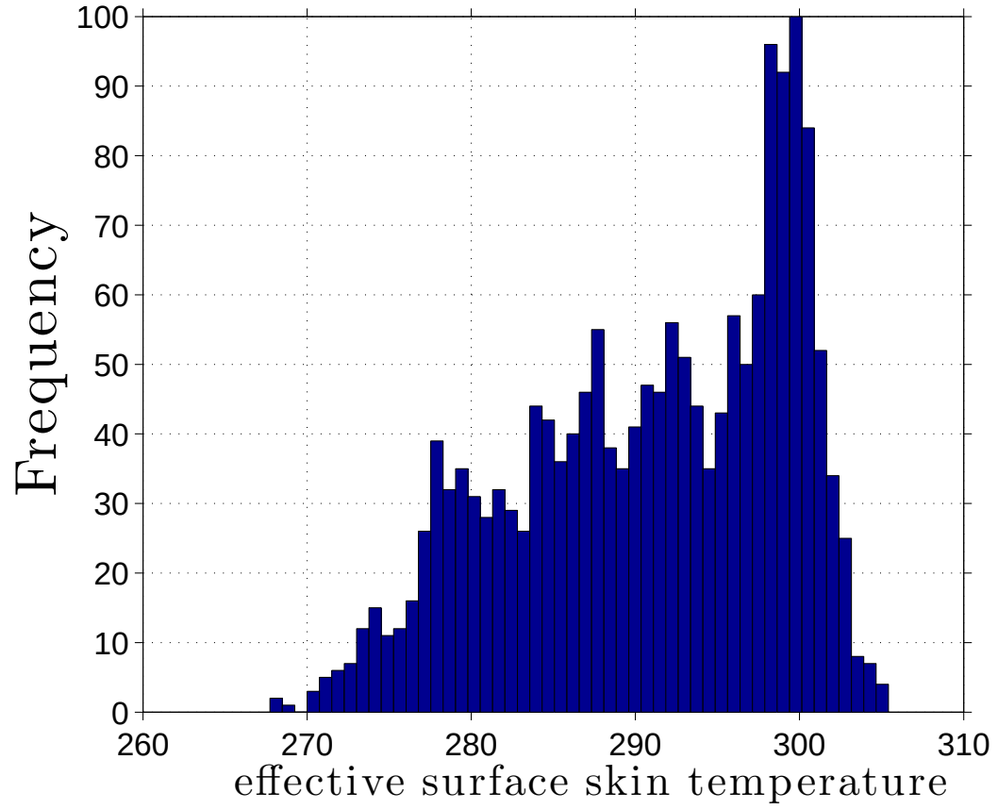
Big Picture

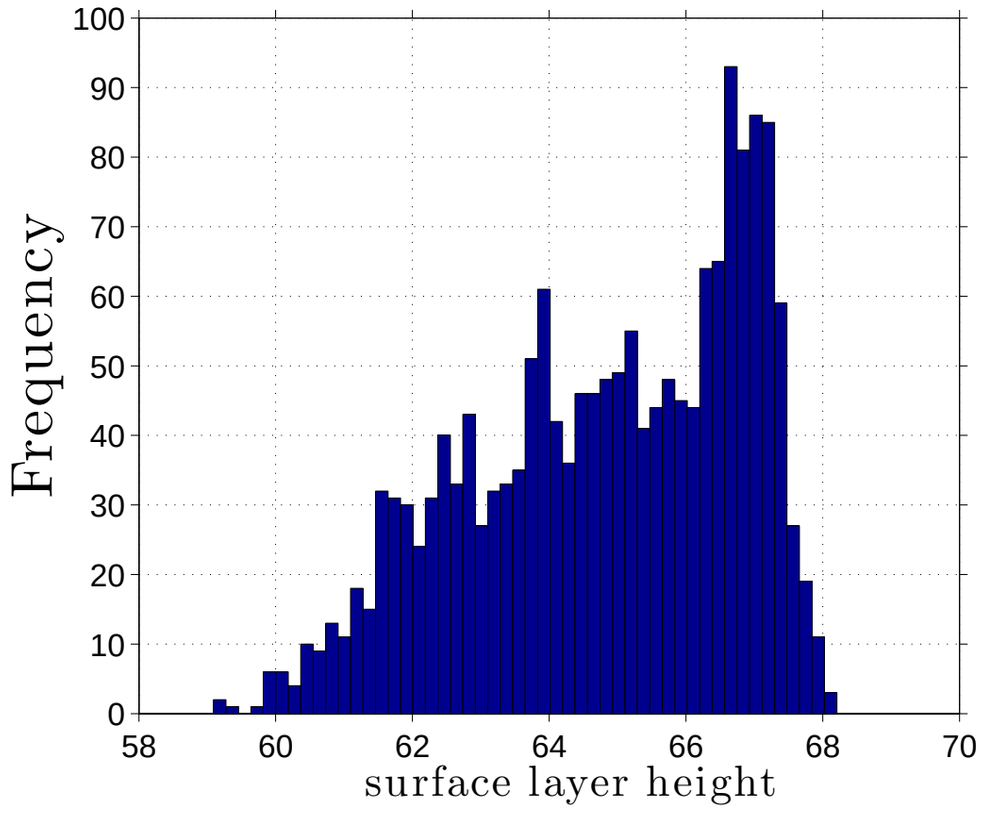
The relationships between Asthma admissions and environmental variables are clearly:

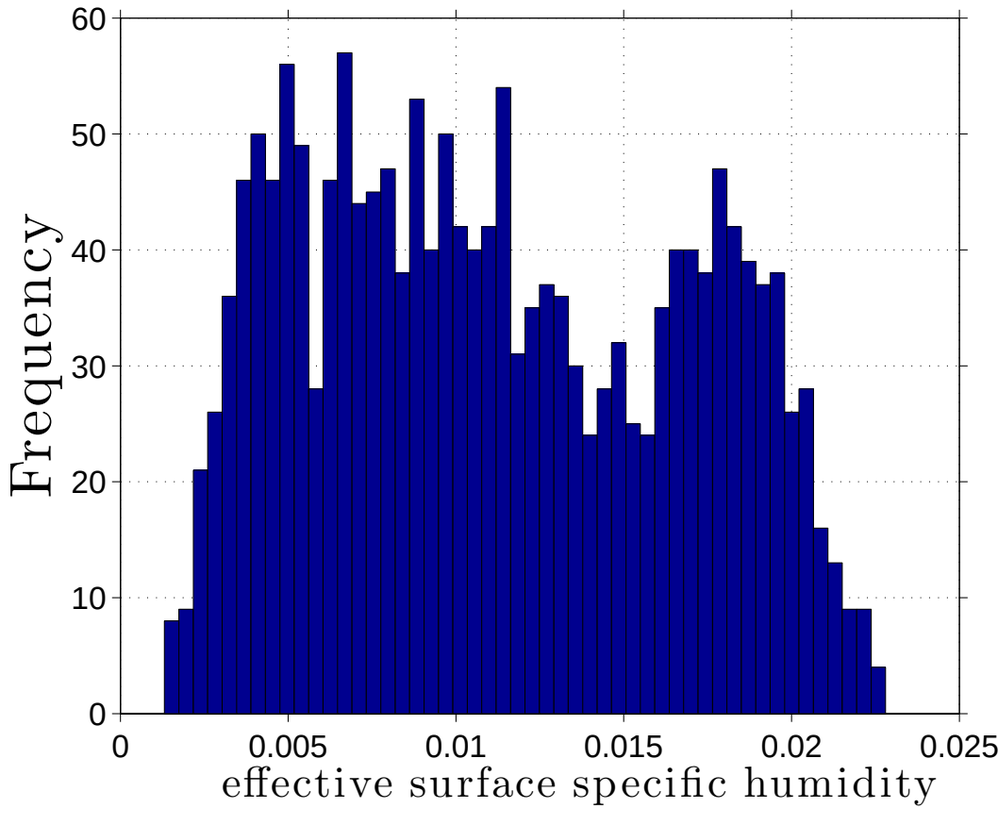
- Non-Gaussian
- Multi-variate
- Non-linear
- There is a different relationship to environmental variables for ER, In-Patient, and Out-Patient admissions

Non Gaussian Distributions









Mutual Information

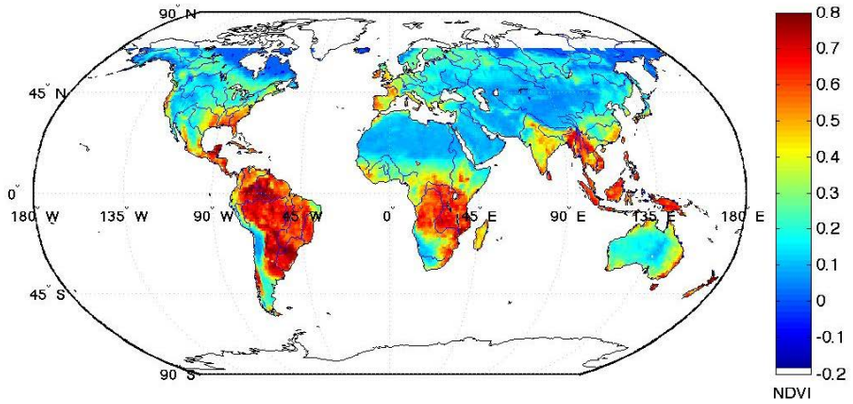
The mutual information of two variables is a quantity that measures the mutual dependence of the two variables.

$$I\left(X,Y\right)=\sum_{x\in X}\sum_{y\in Y}p\left(x,y\right)\log\frac{p\left(x,y\right)}{p\left(x\right)p\left(y\right)}$$

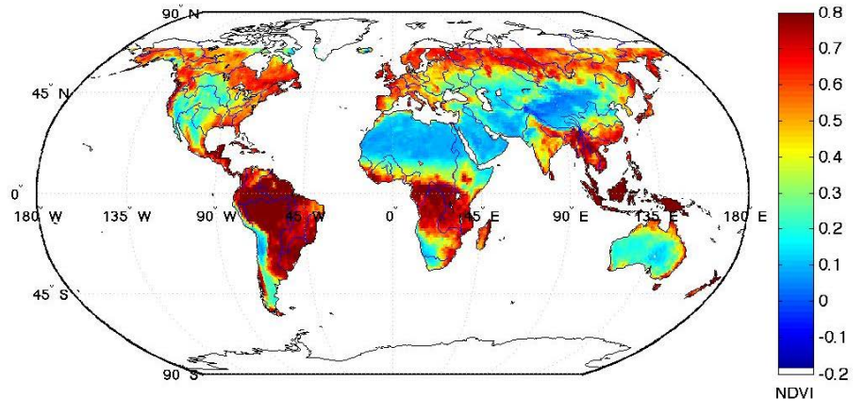
Intuitively, mutual information measures the information that X and Y share: it measures how much knowing one of these variables reduces uncertainty about the other.

Multi-Variate

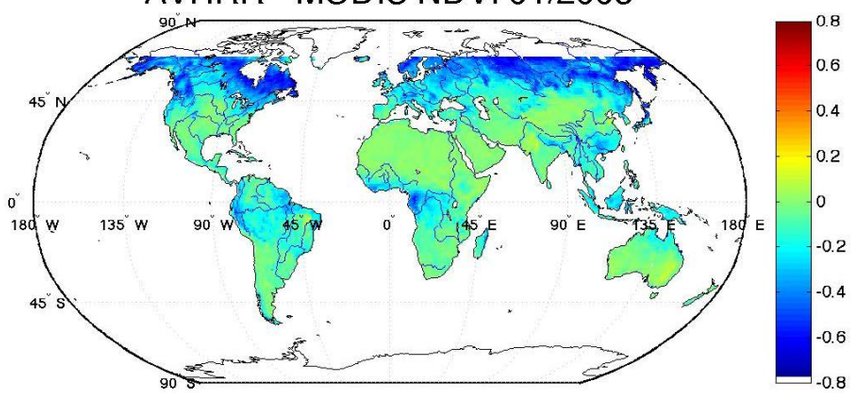
AVHRR NDVI 01/2003



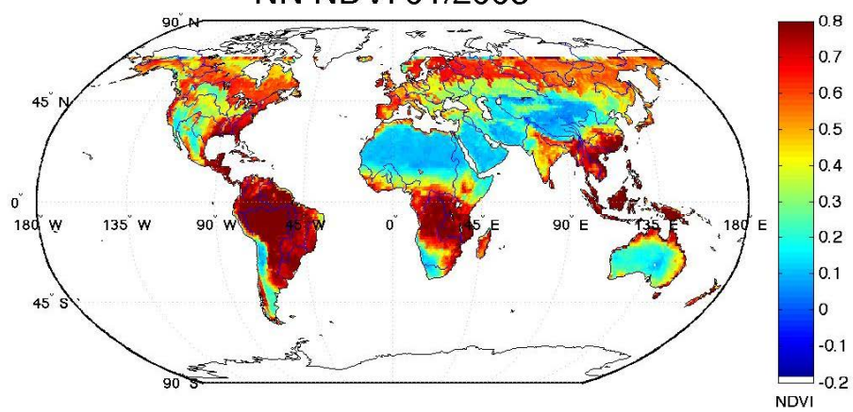
MODIS NDVI 01/2003

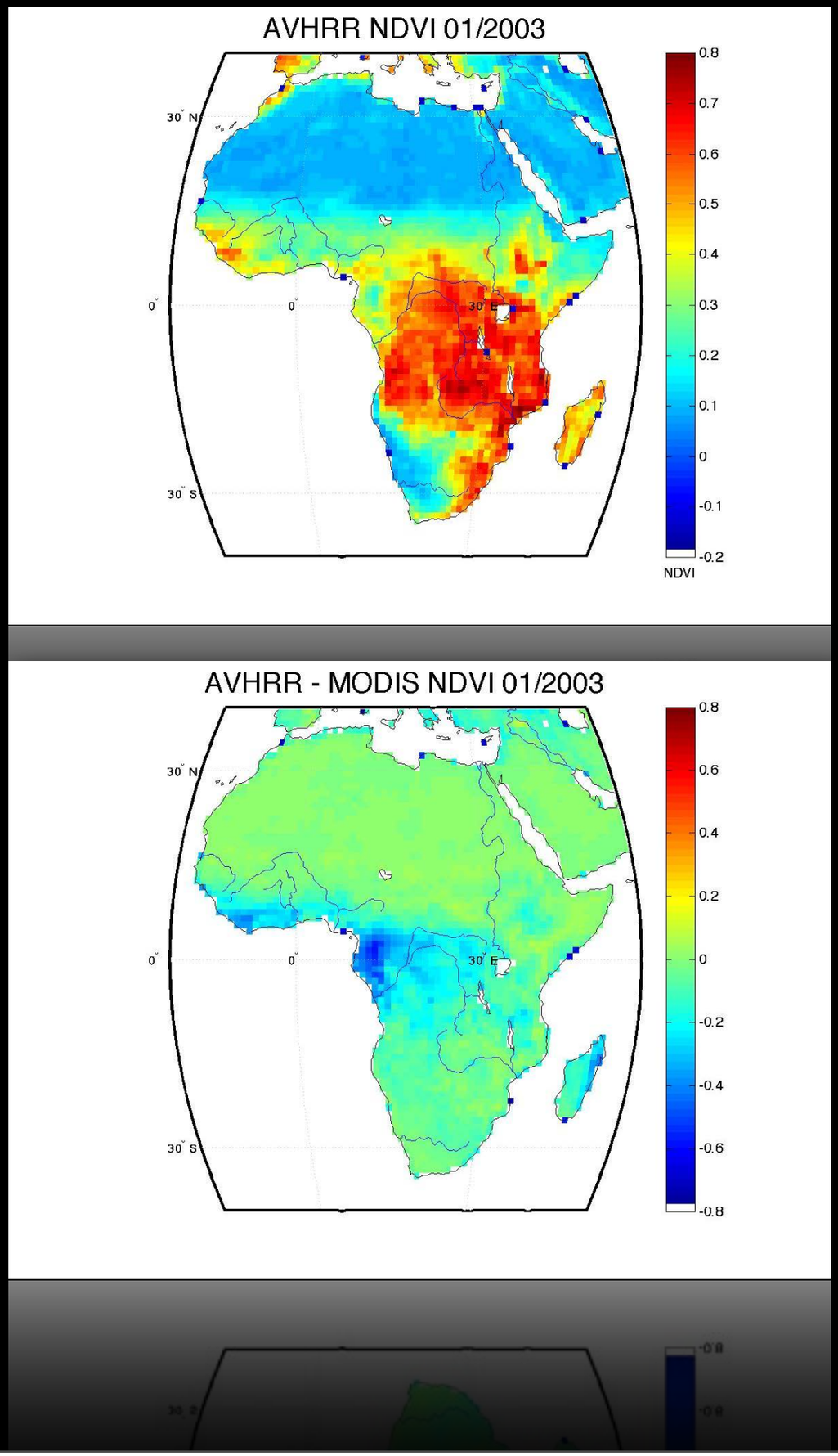


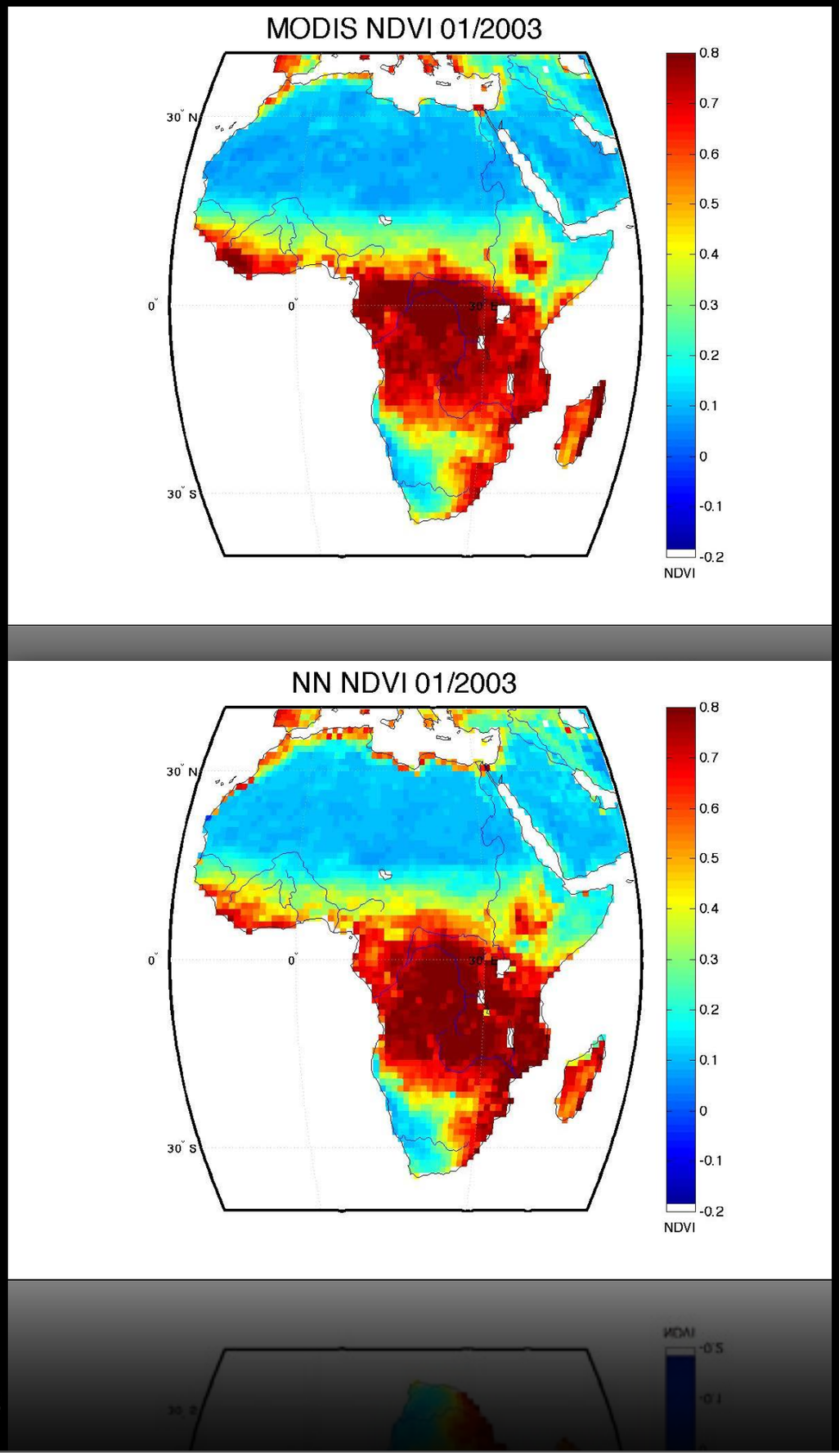
AVHRR - MODIS NDVI 01/2003



NN NDVI 01/2003



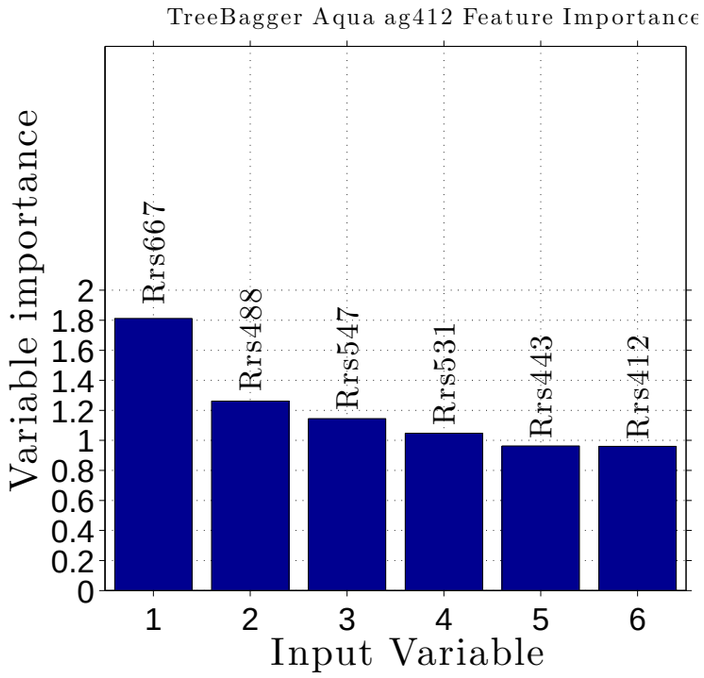

AVHRR NDVI 01/2003
30˚ N
0˚
0˚
30˚ E
30˚ S
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
NDVI

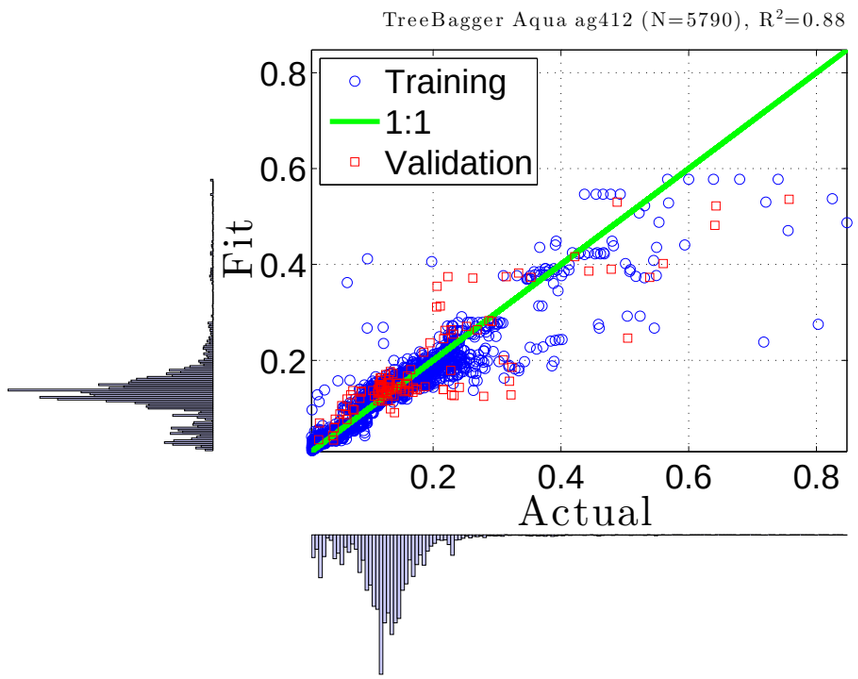

MODIS NDVI 01/2003
30˚ N
0˚
0˚
30˚ E
30˚ S
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
NDVI

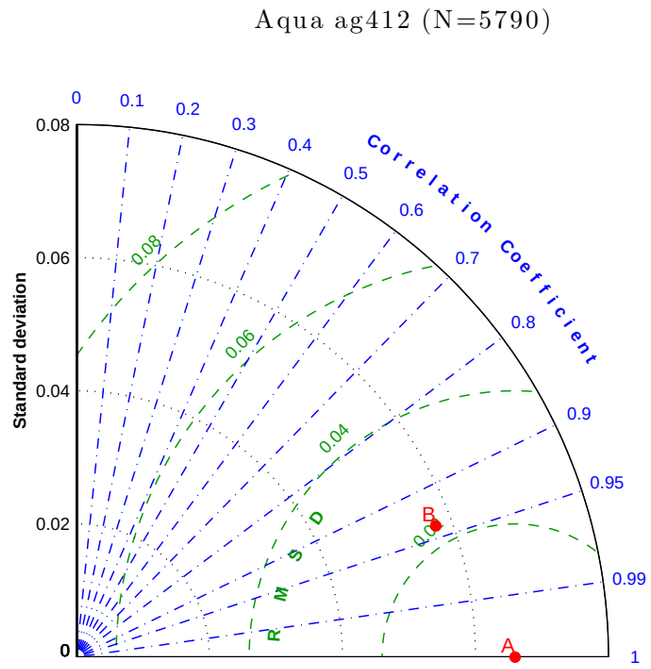
AVHRR - MODIS NDVI 01/2003
30˚ N
0˚
0˚
30˚ E
30˚ S
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8

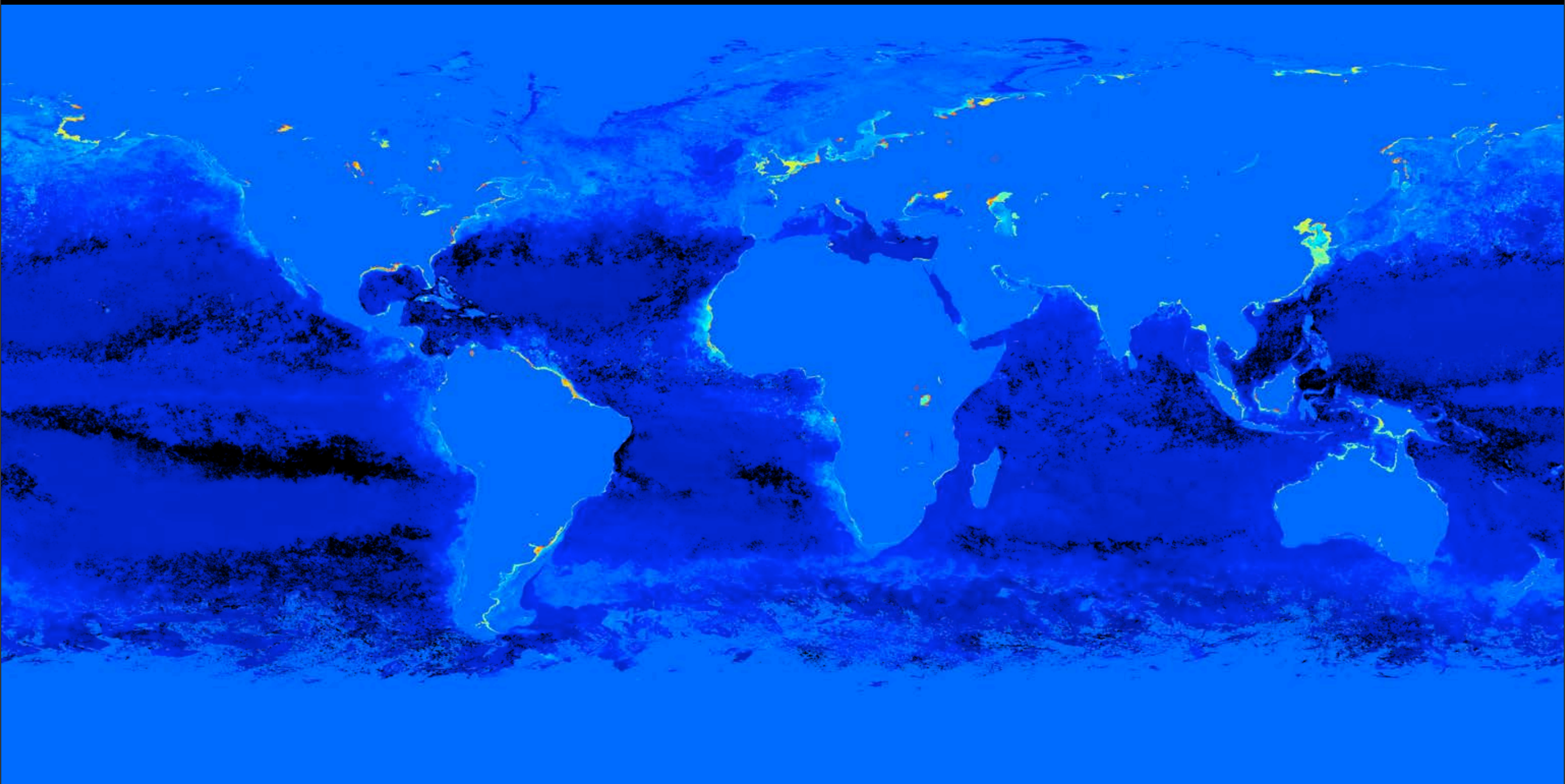
NN NDVI 01/2003
30˚ N
0˚
0˚
30˚ E
30˚ S
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
NDVI

Aqua ag412


TreeBagger Aqua ag412 Feature Importance
Variable importance
Input Variable
Rrs667
Rrs488
Rrs547
Rrs531
Rrs443
Rrs412


TreeBagger Aqua ag412 (N=5790), R²=0.88
Training
1:1
Validation
Fit
Actual

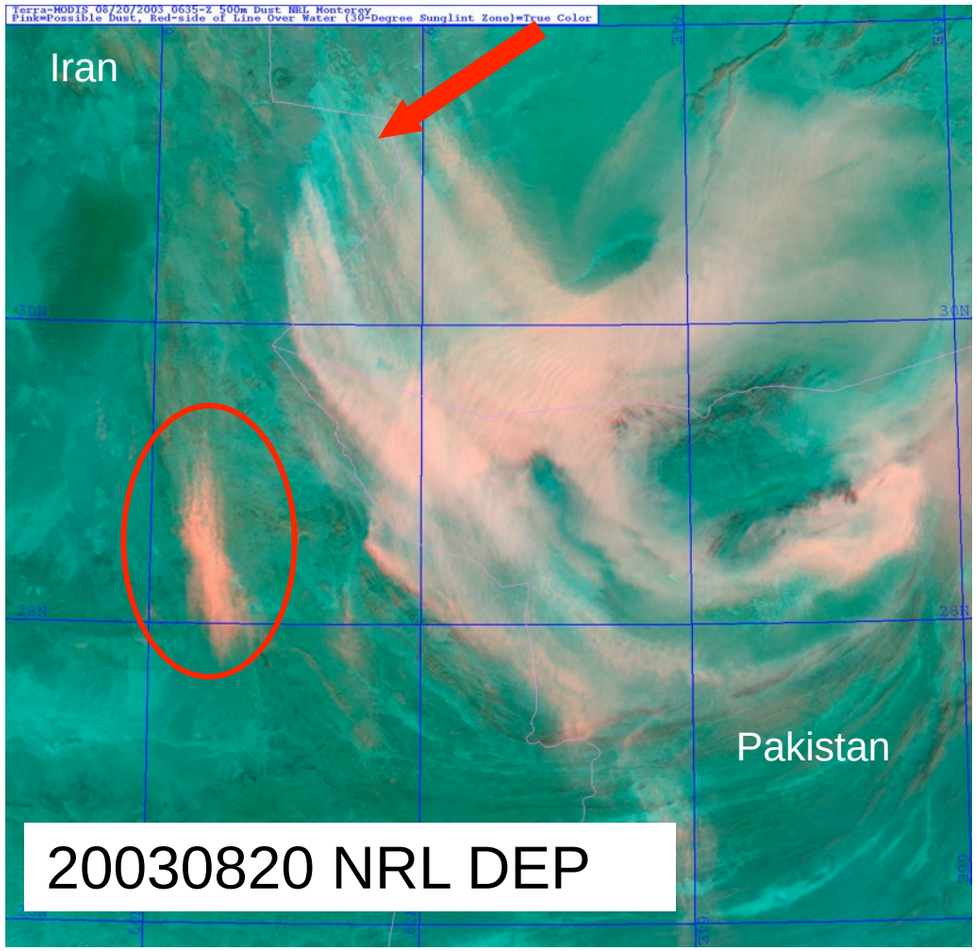

Aqua ag412 (N=5790)
Correlation Coefficient
Standard deviation
RMSD
A
B



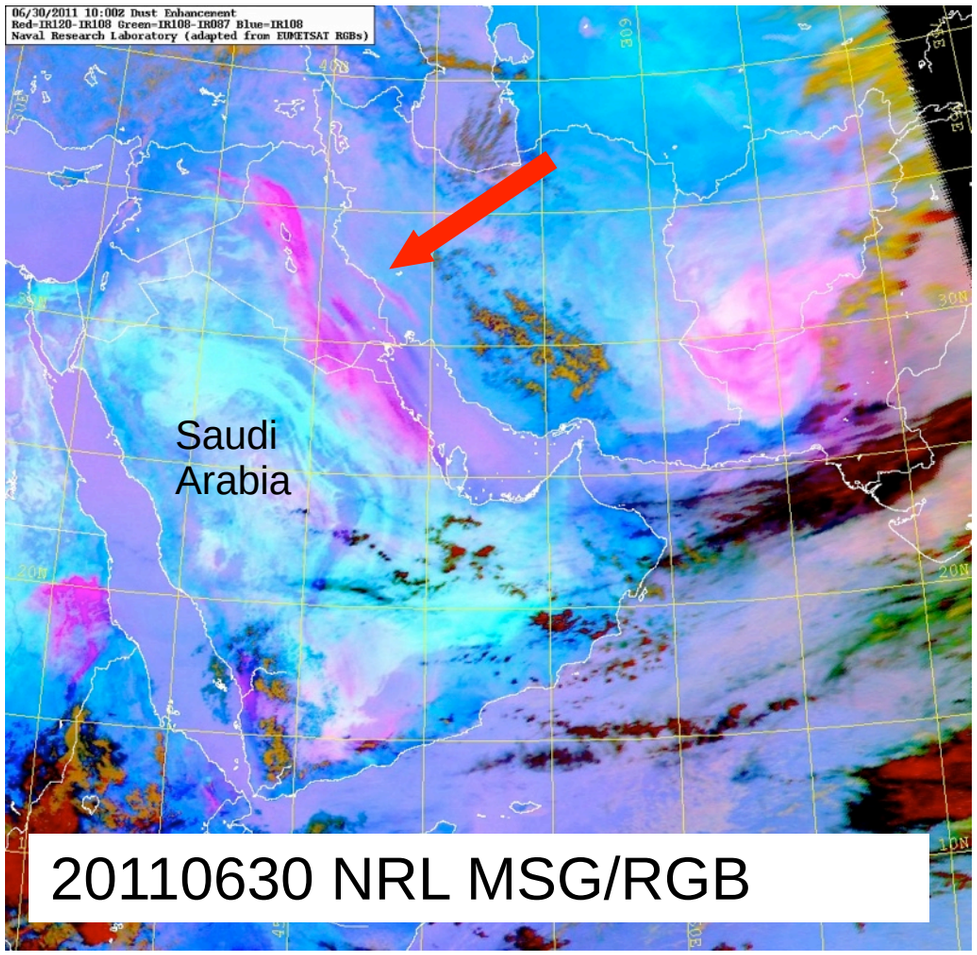
Thursday, October 24, 13

A Haboob (Arabic: هَبوب "strong wind", or "blowing furiously.")

NRL High-resolution Dust Source Database



20030820 NRL DEP



20110630 NRL MSG/RGB

Approach and Methodology

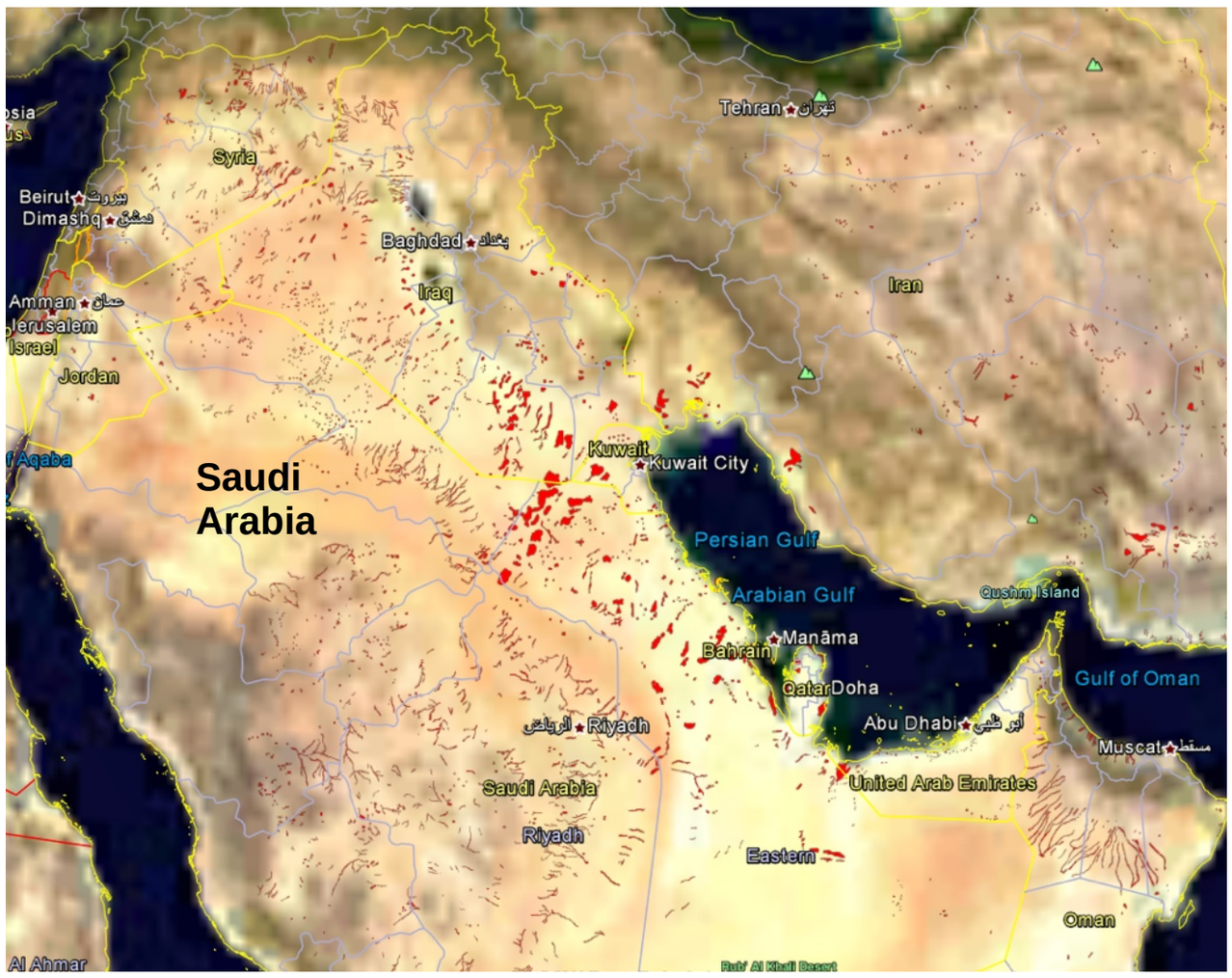
- **10 years of DEP (2 yr MSG/RGB) imagery**
- COAMPS 10 m wind overlays
- Surface weather plots
- **ENVI (Gis-like software)**
- NGDC topographical 10ºX10º tiles
- **Overlay 0.25º grid** or use **Google Earth (GE)**

- **Dust source area entered into database** (cursor location tool = 1km precision)
- Cross-correlate land and water features using maps, atlases, Landsat images (detailed topographic, geographic, and geomorphic information, **GE**)
- Technical and governmental reports

NRL High-resolution Dust Source Database

Solid red and purple shapes identify dust source areas located using DEP and MSG.

SW Asia DSD



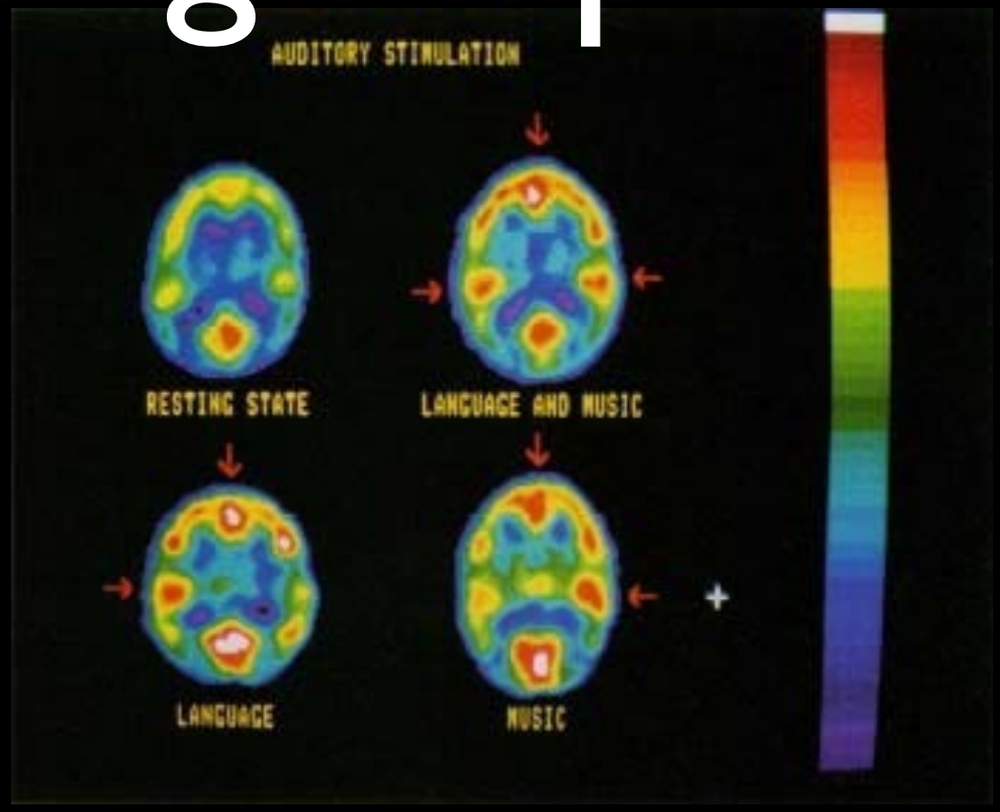
East Asia DSD

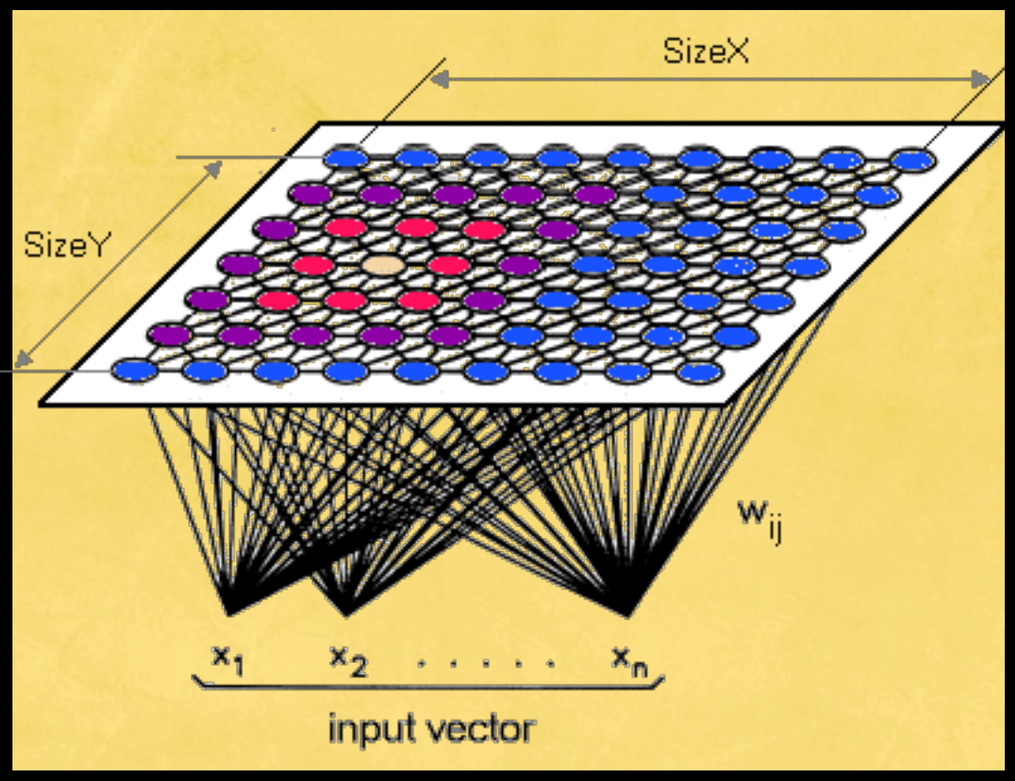
Self-Organizing Map

SOMs reduce dimensionality by producing a map that objectively plots the similarities of the data by grouping similar data items together.

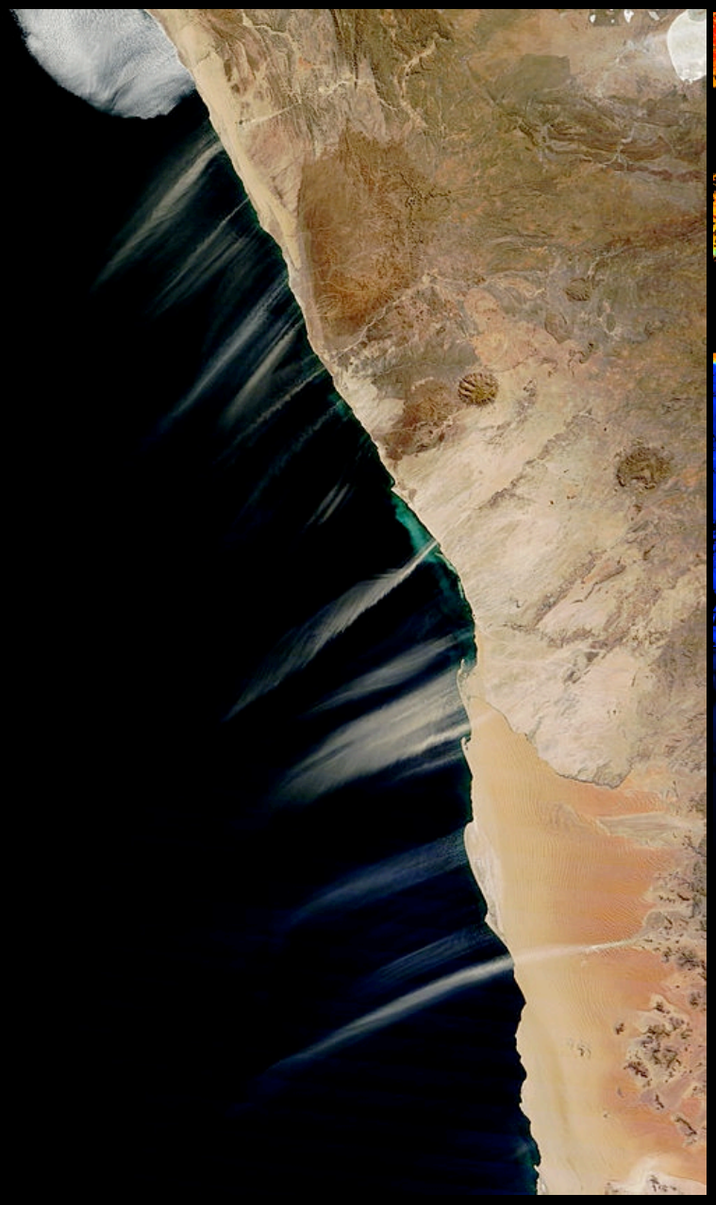
SOMs learn to classify input vectors according to how they are grouped in the input space.

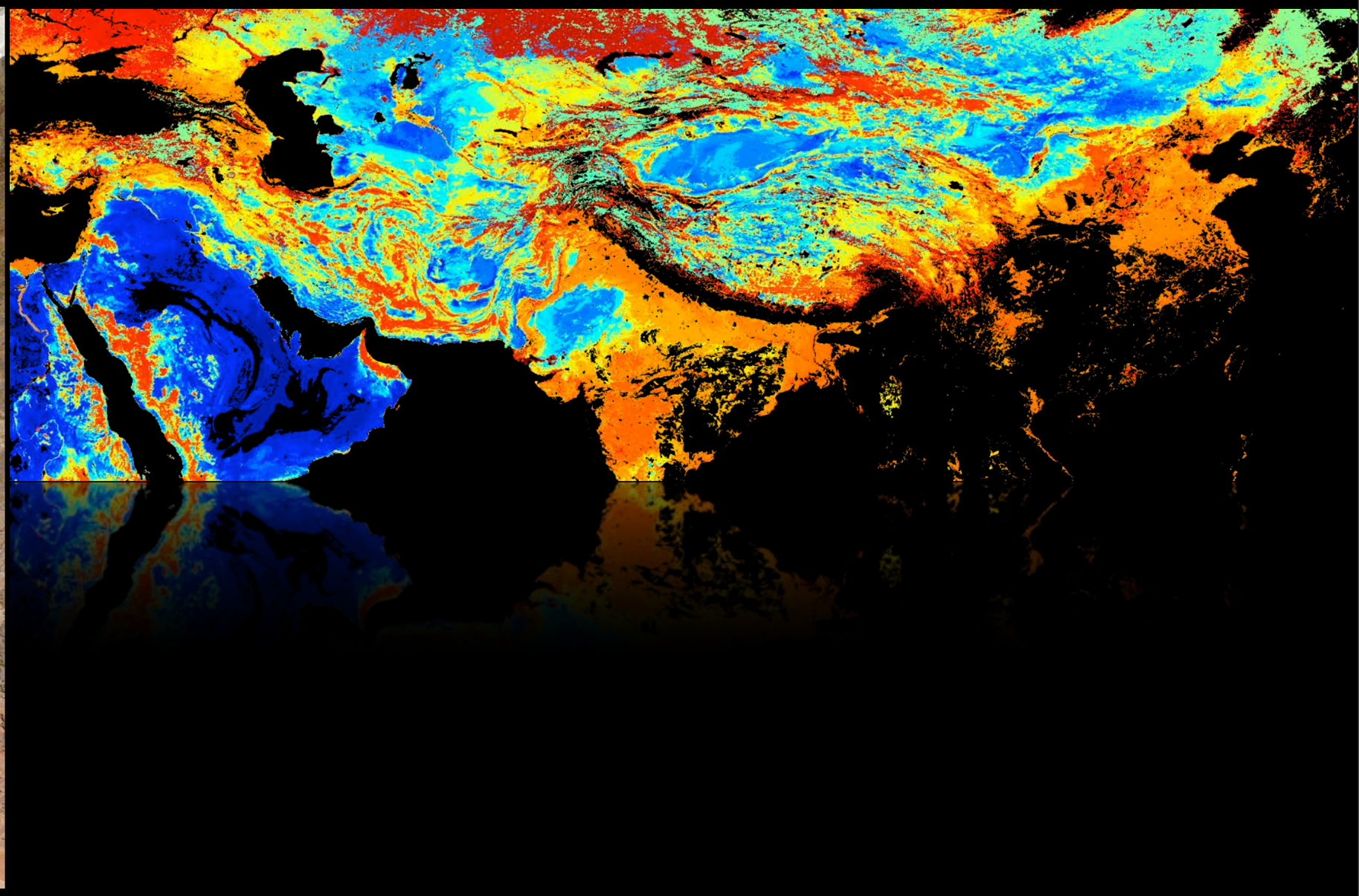
SOMs learn both the distribution and topology of the input vectors they are trained on. This approach allows SOMs to accomplish two things, reduce dimensions and display similarities.



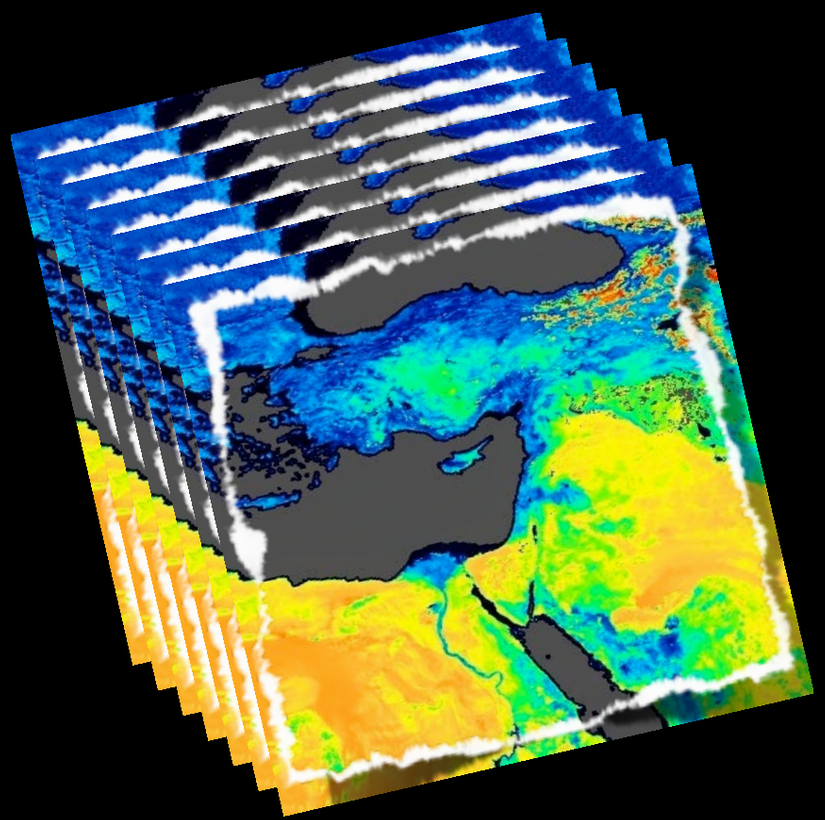


Detecting Dust Sources

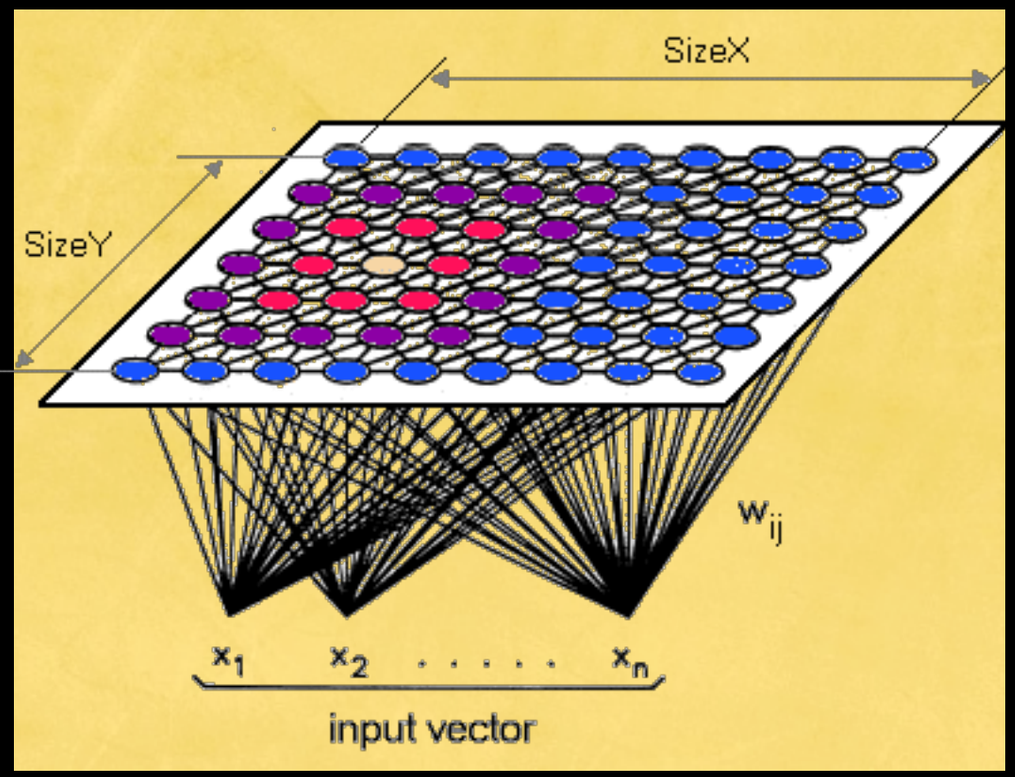


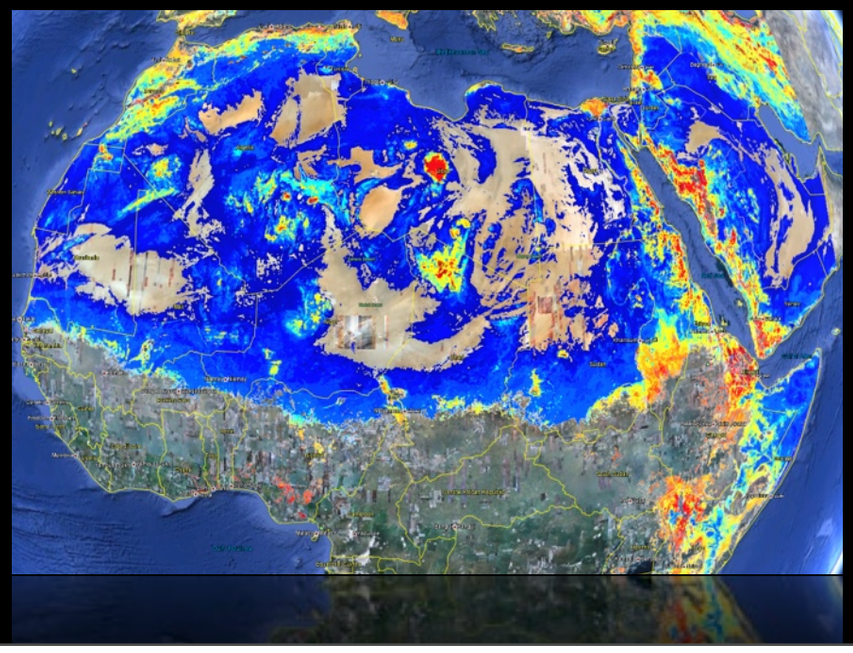


Self Organizing Map Classification

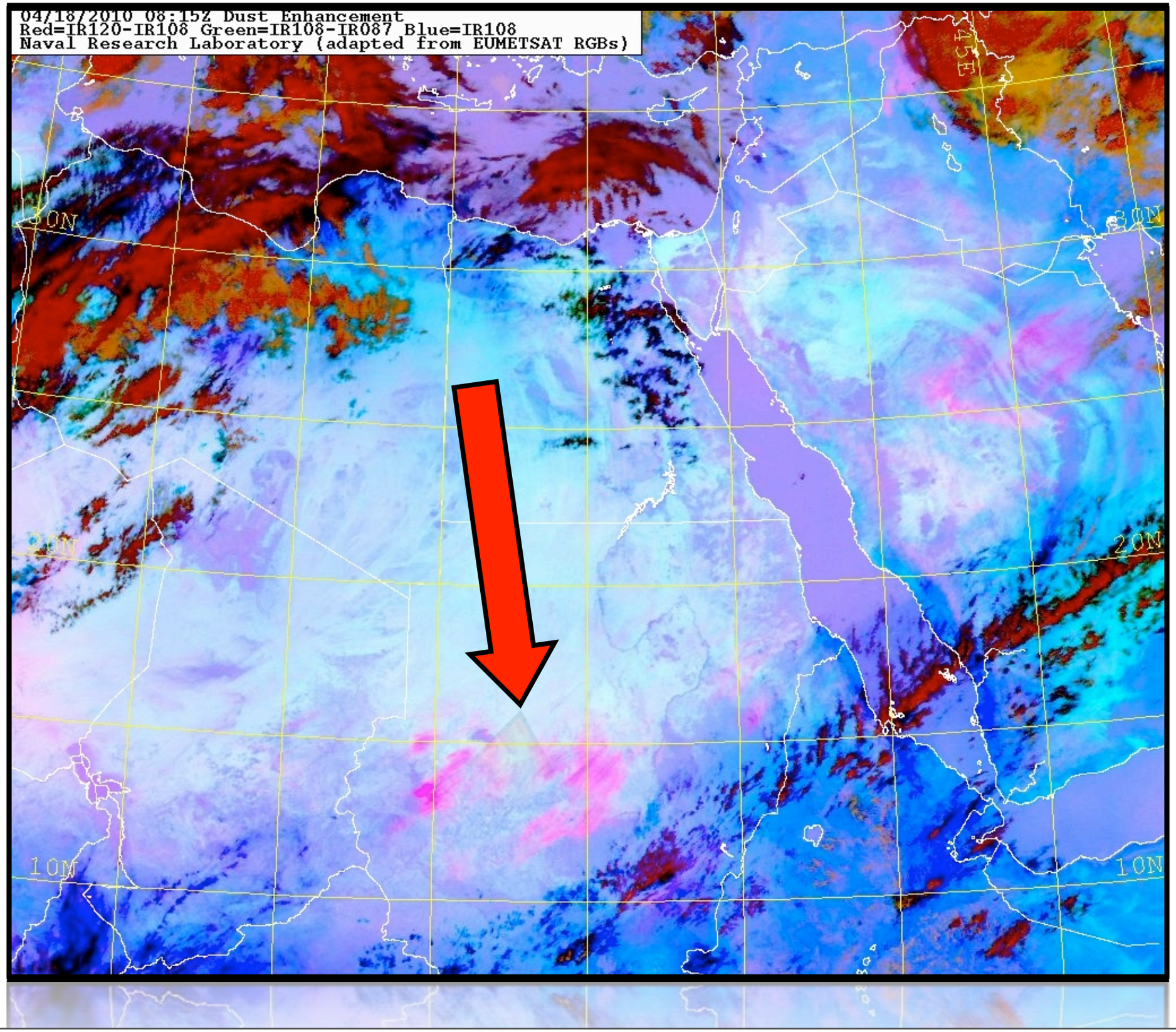


7 Bands
MODIS MCD43C3
bihemispherical reflectance





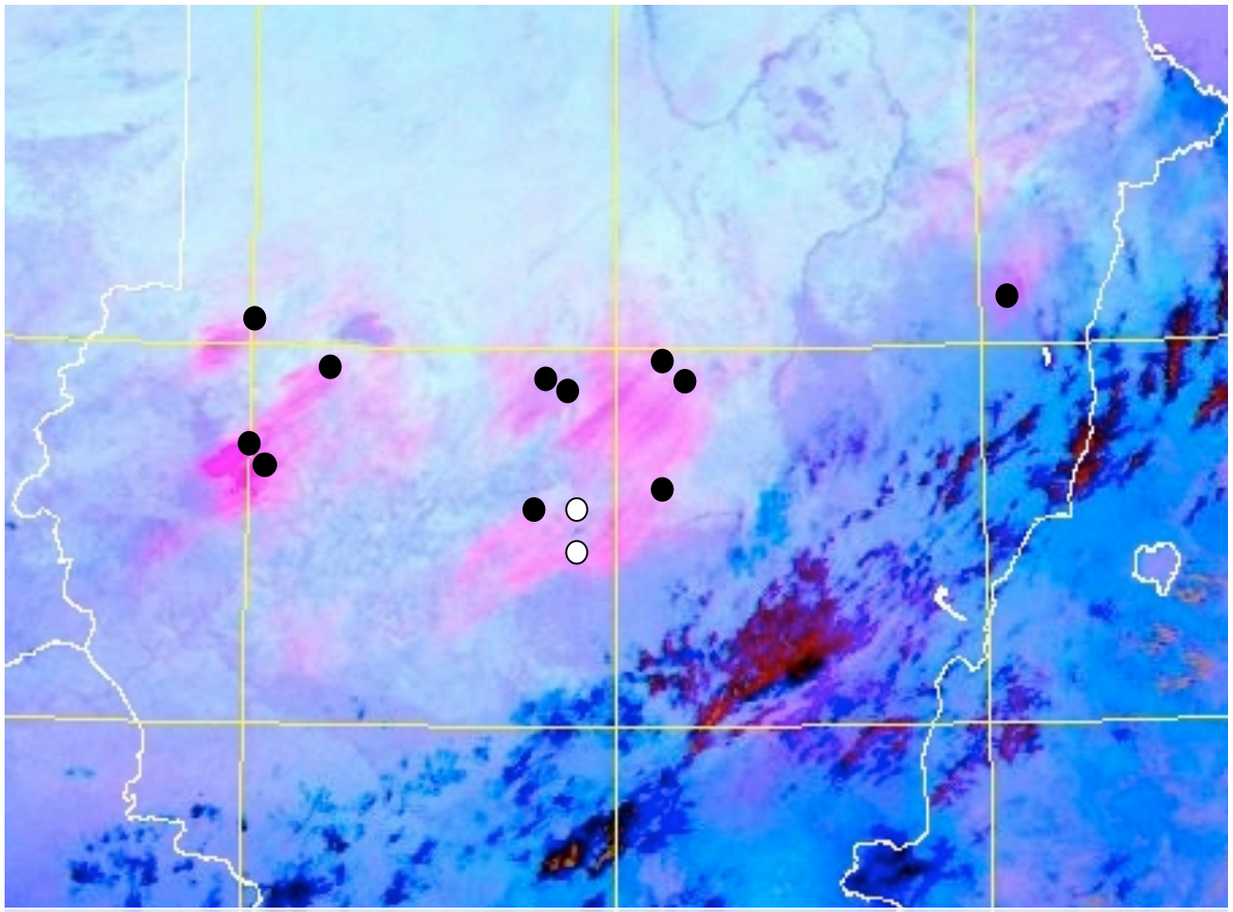
Sudan Dust Event: April 18, 2010 (6Z – 8Z)

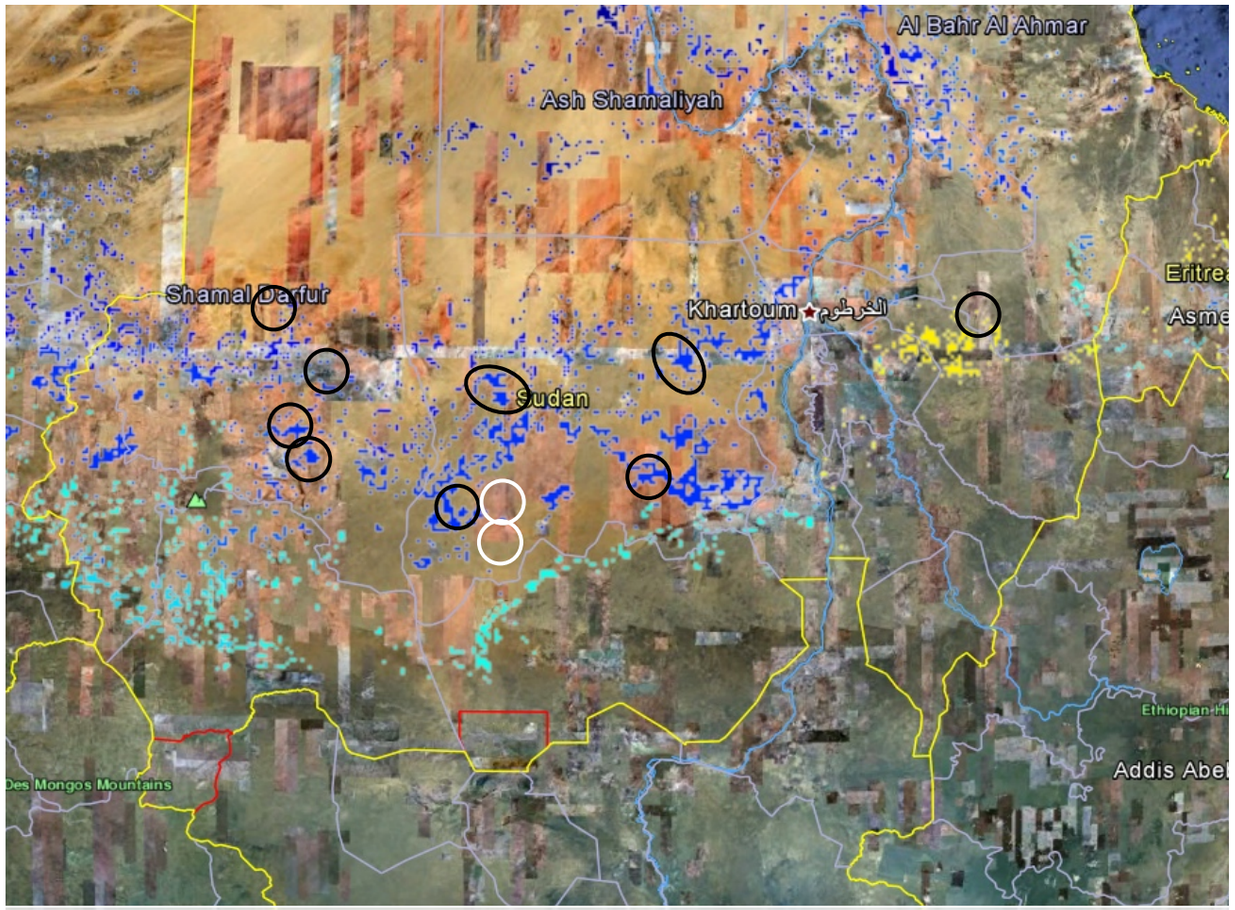


Sudan Dust Event: April 18, 2010 (6Z – 8Z)

Plumes originate from southern Sahel

Corresponding SOM-Classes: 175, 177, 648

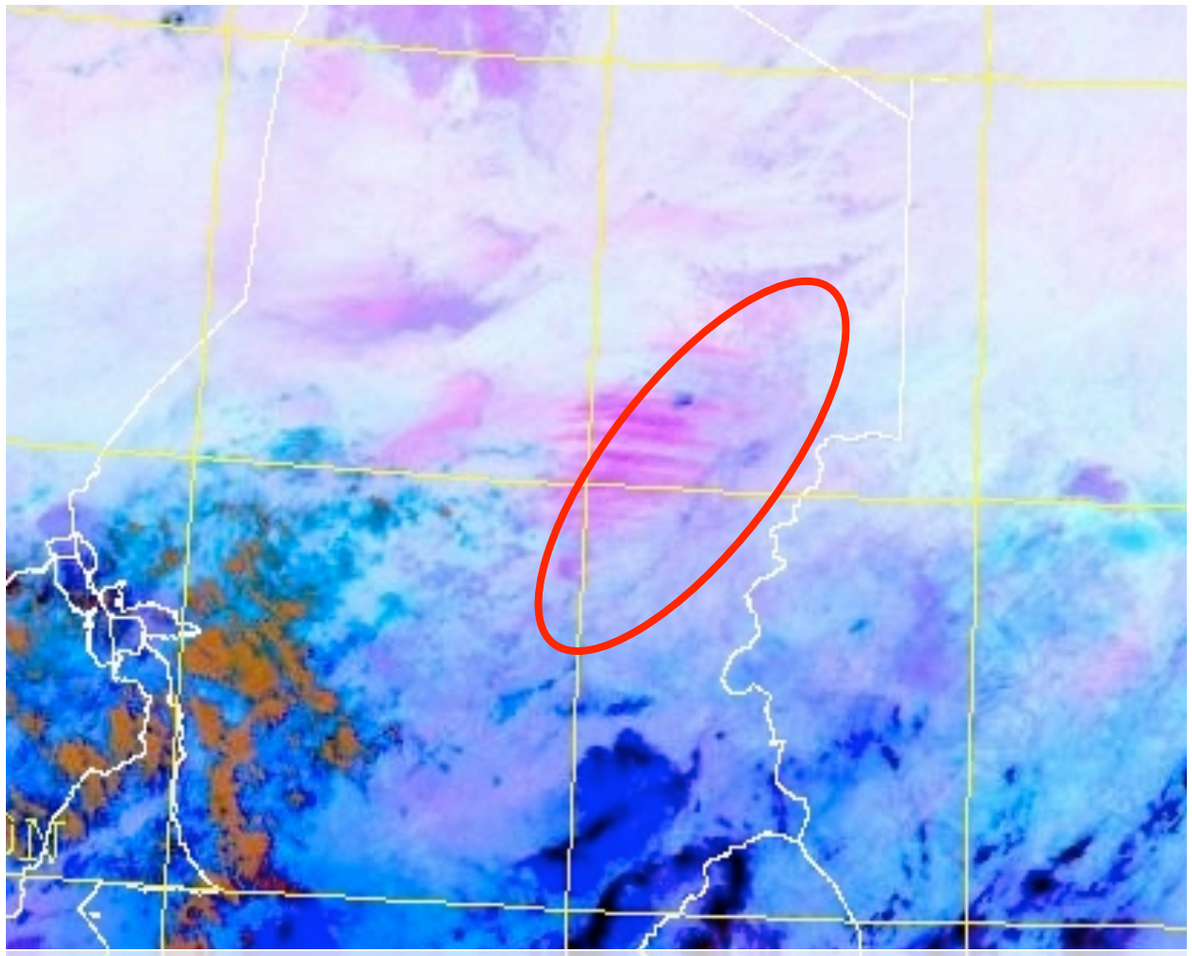


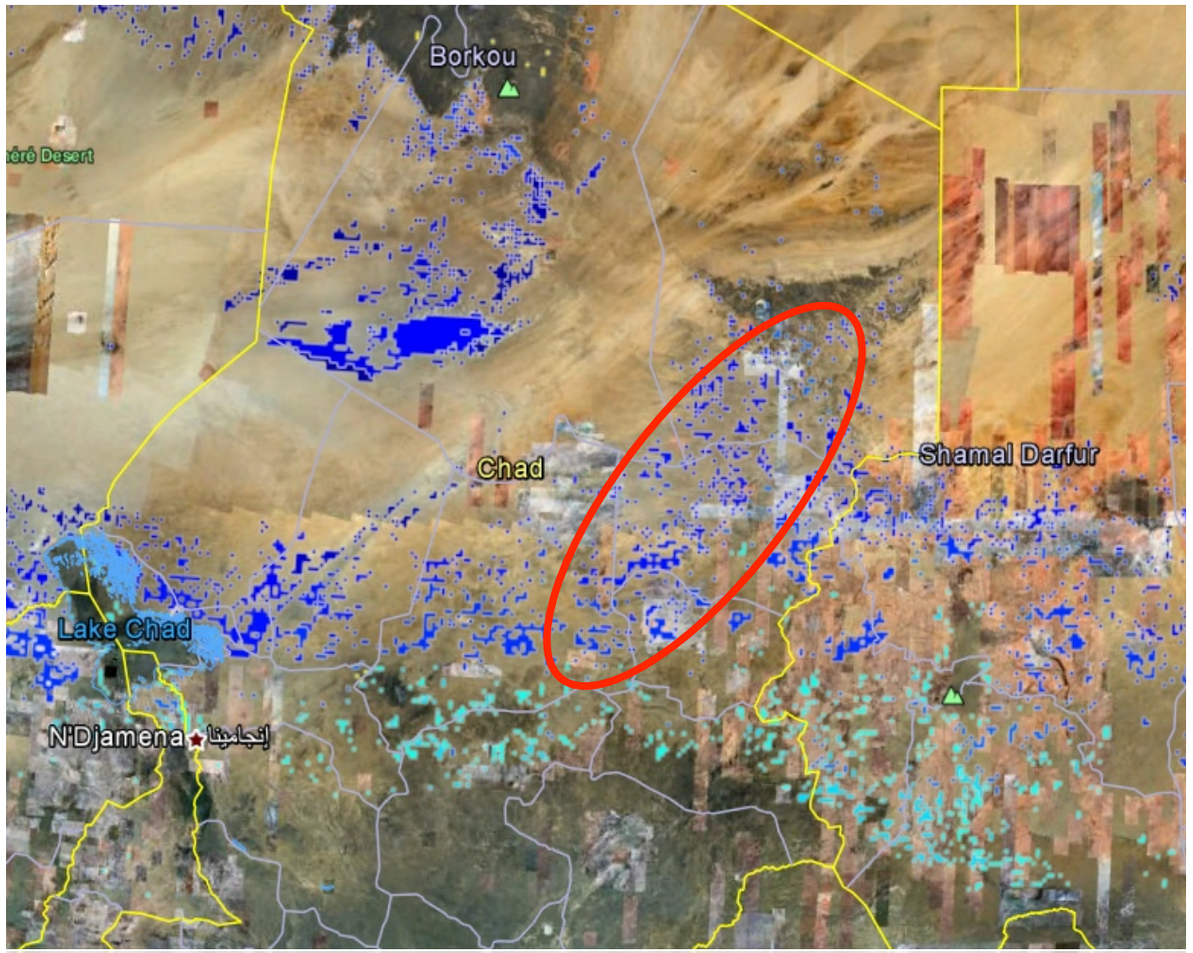


Chad: Ennedi Plateau Dust Event: April 9, 2010

Plumes originate from southern Sahel

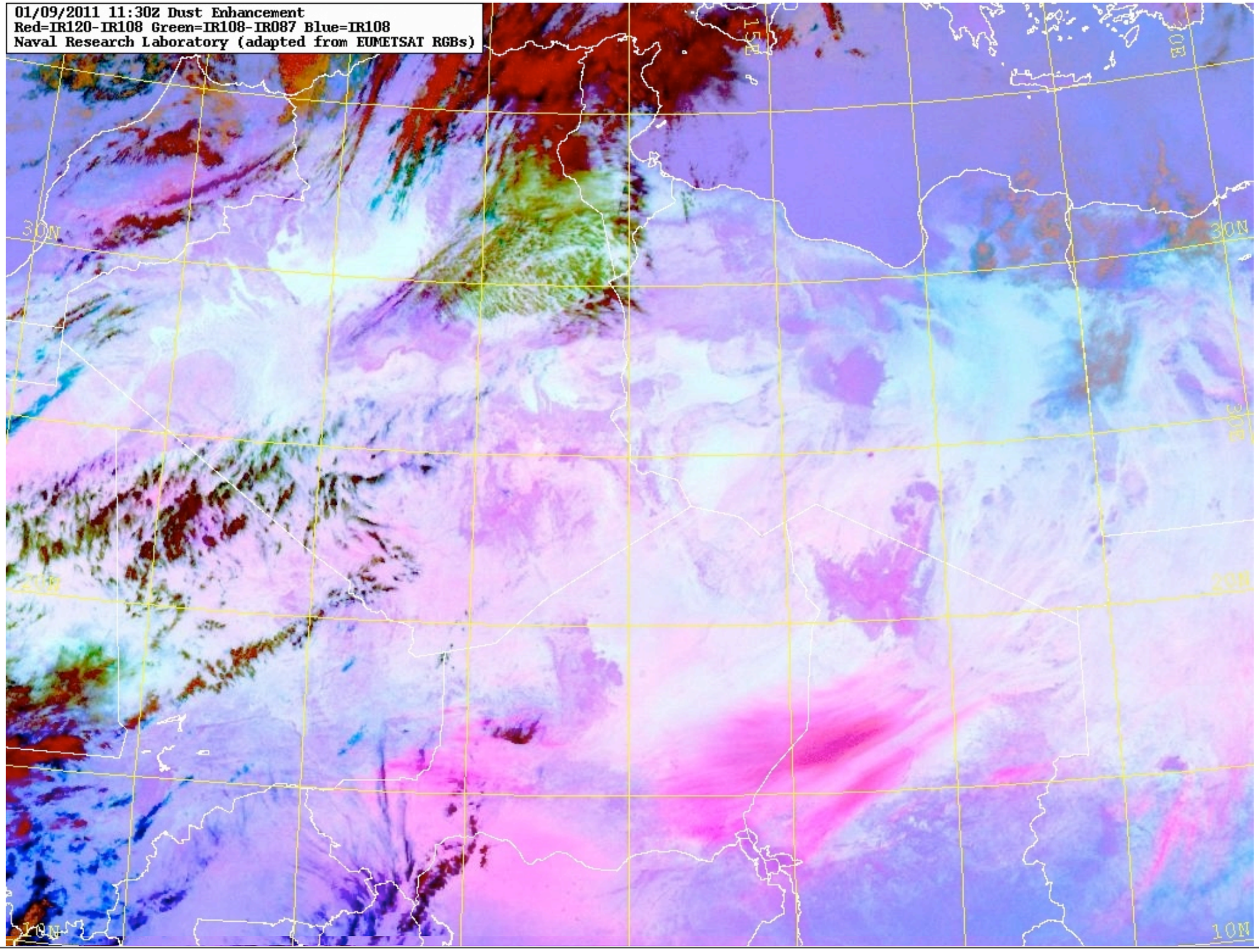
Corresponding SOM-Classes: 175, 177, 648



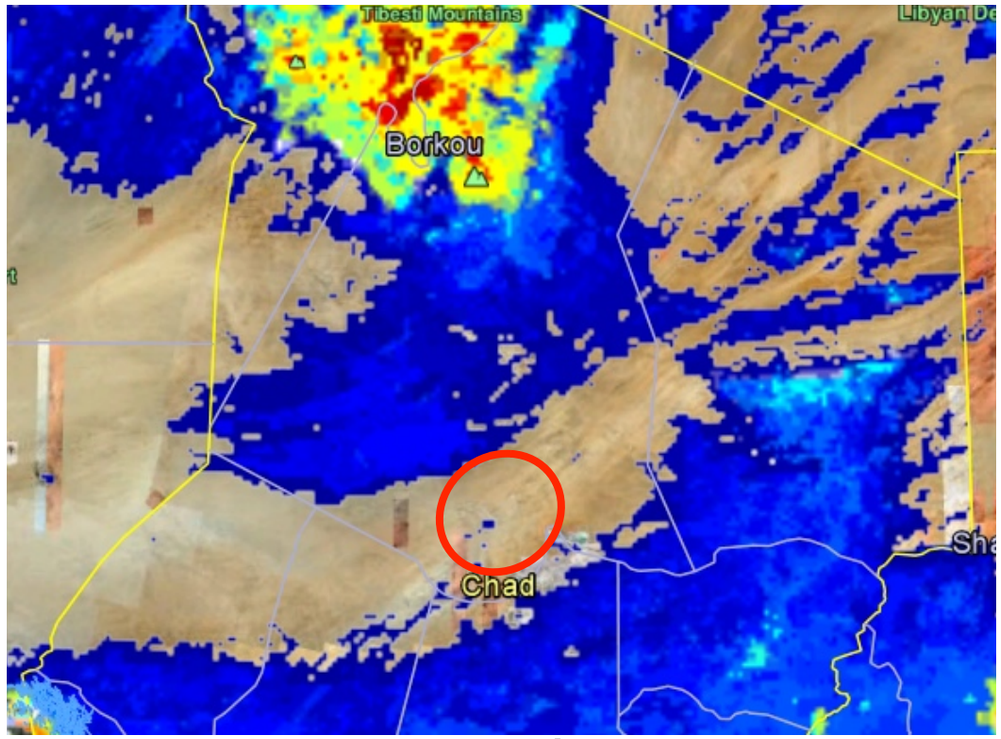


Chad: Bodélé Depression
Dust Event: March 16, 2010 (7Z -12Z)

Located at the southern edge of the Sahara Desert in north central Africa, is the lowest point in Chad. Dust storms from the Bodélé Depression occur on average about 100 days per year. The Bodélé depression is a single spot in the Sahara that provides most of the mineral dust to the Amazon forest.

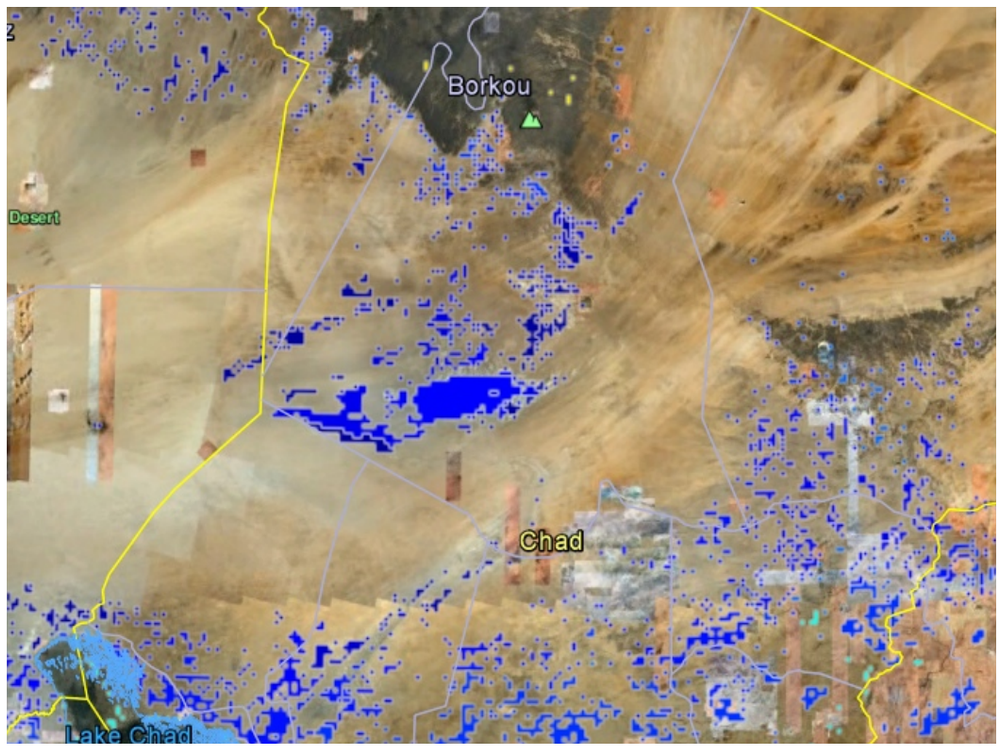


Chad: Bodélé Depression

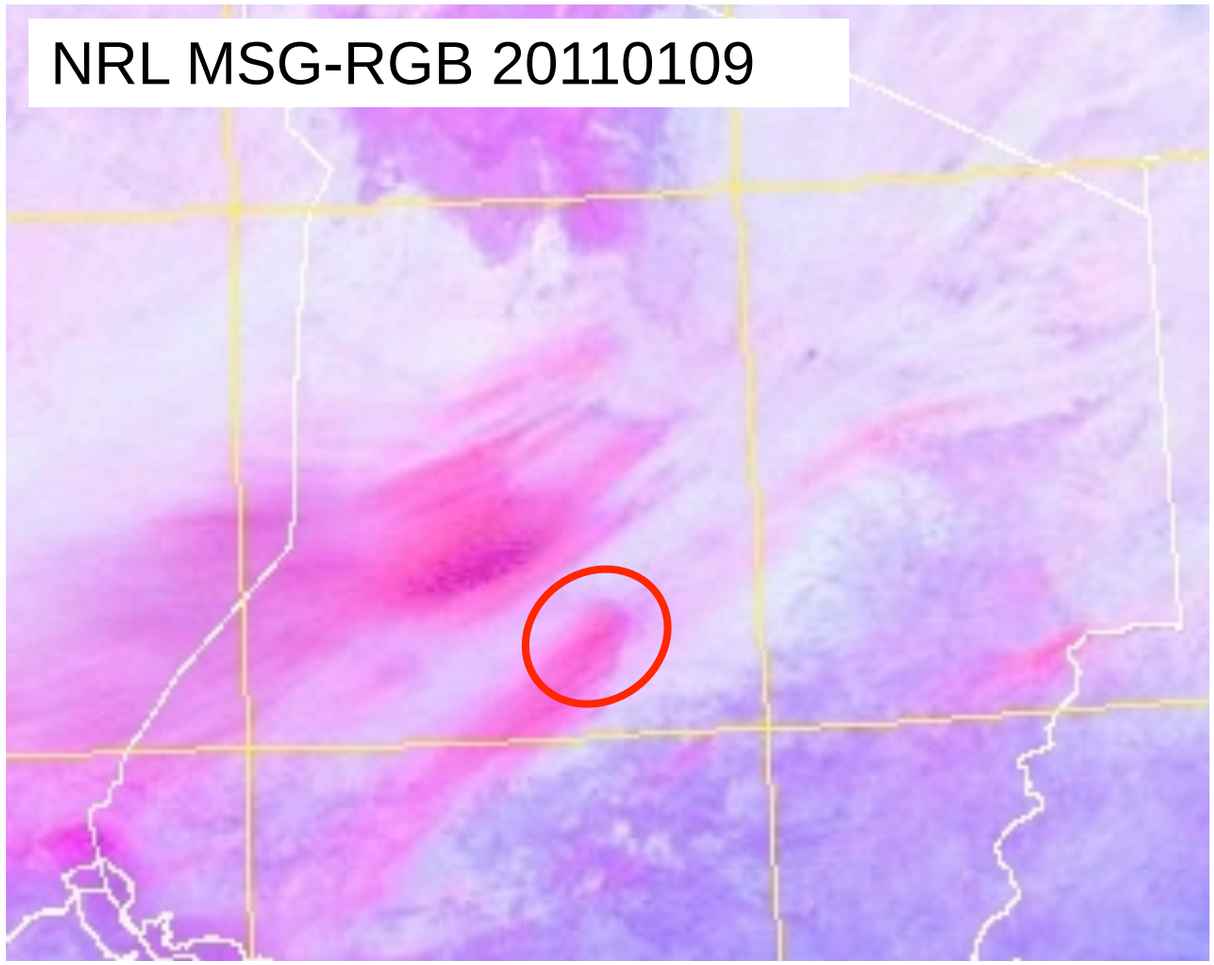


1000 SOM Classes

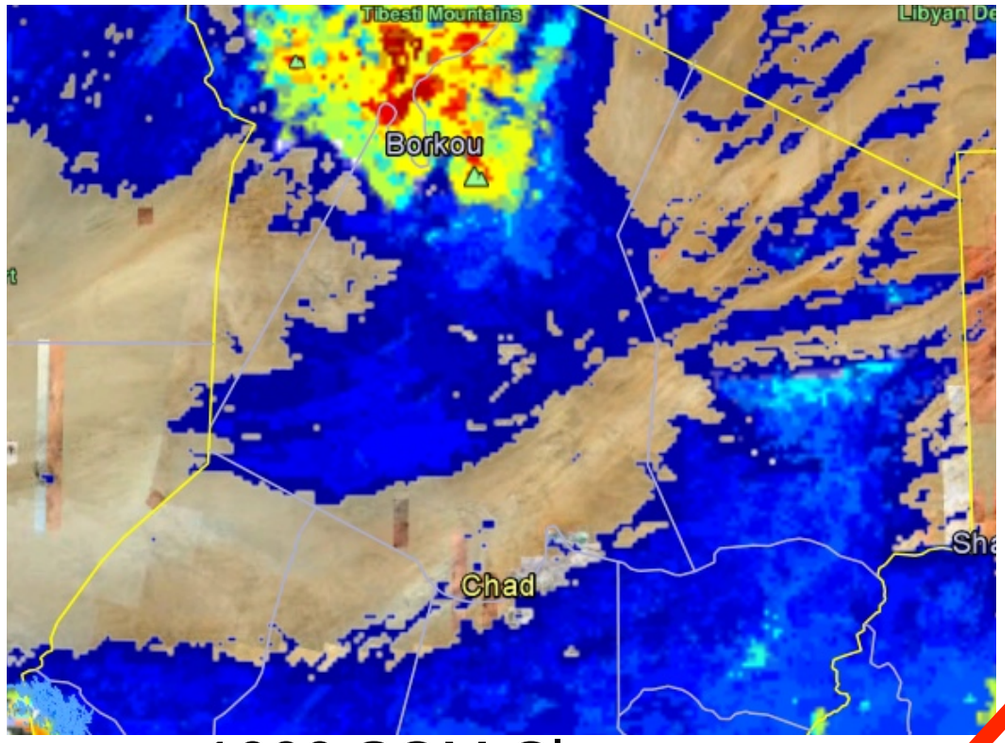
Source area is not designated in first pass of MODIS reflectance and land surface classification.



Selected SOM Classes

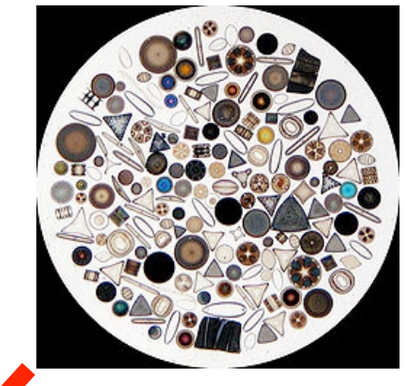


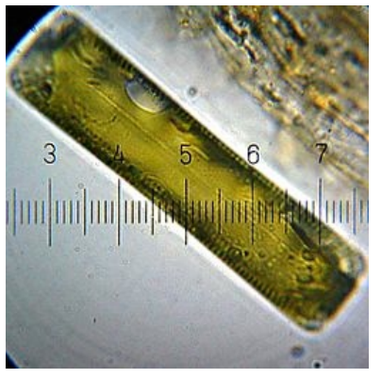
Chad: Bodélé Depression

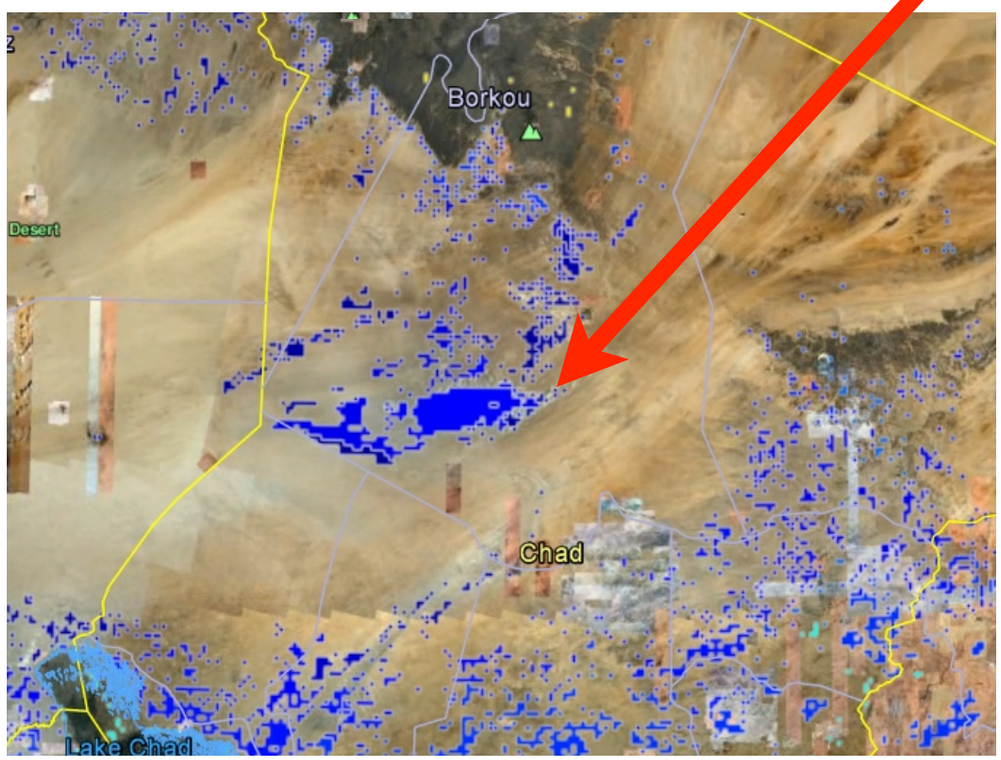


1000 SOM Classes

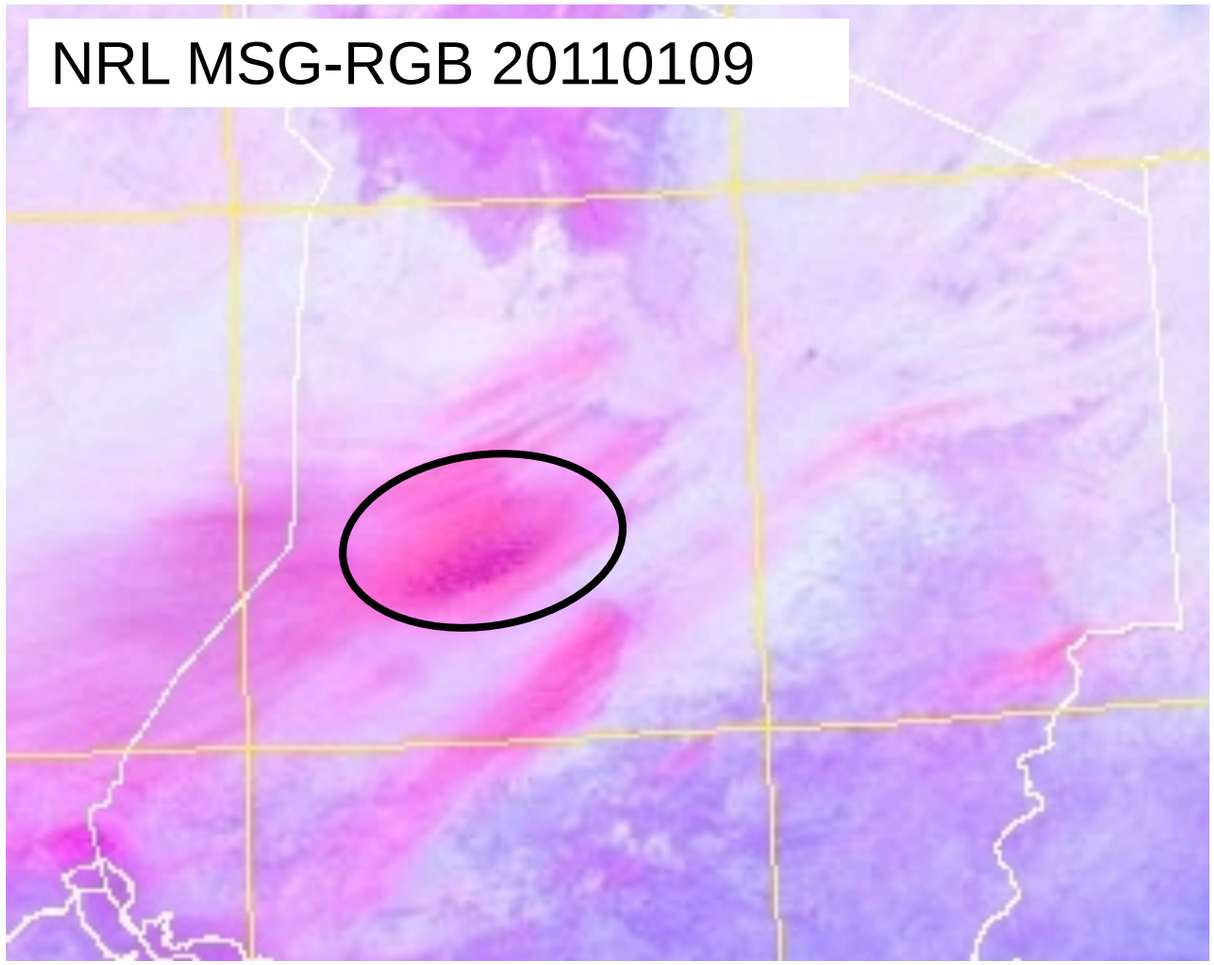
Class 137 maps diatom sediment in depression.



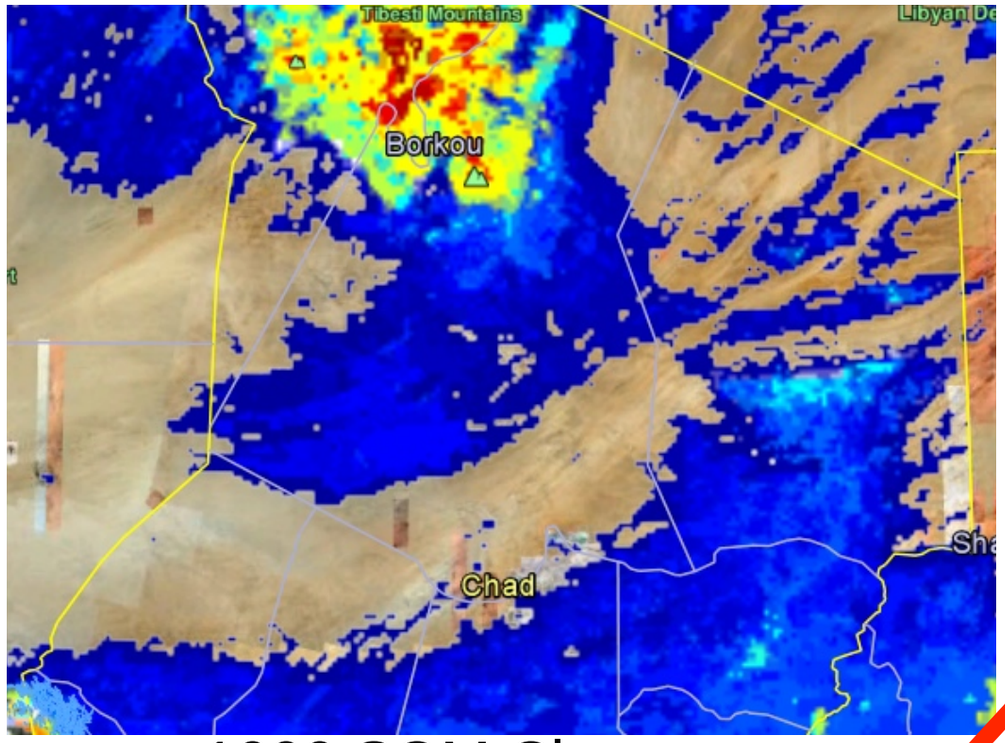




Selected Classes with Class 137

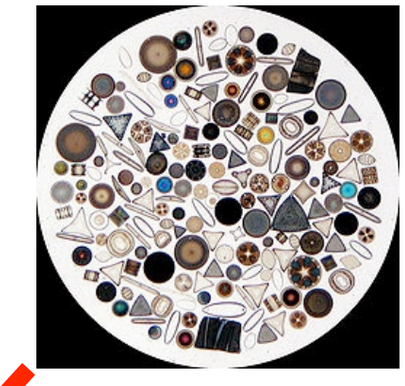


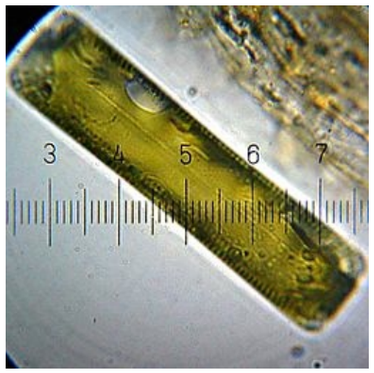
Chad: Bodélé Depression

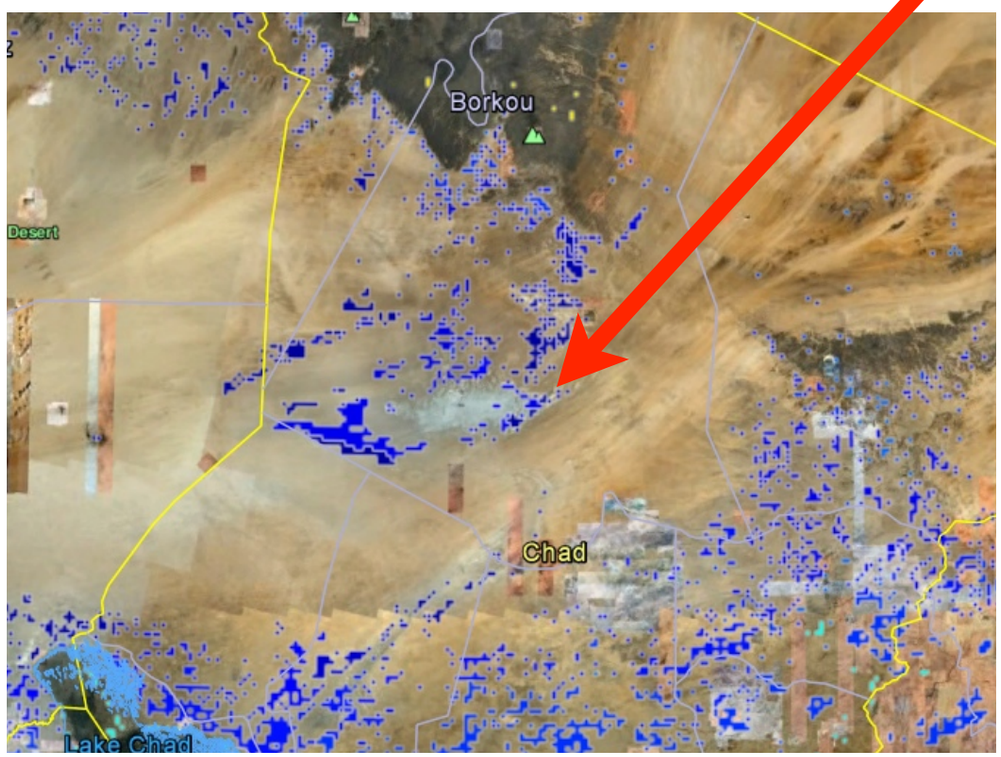


1000 SOM Classes

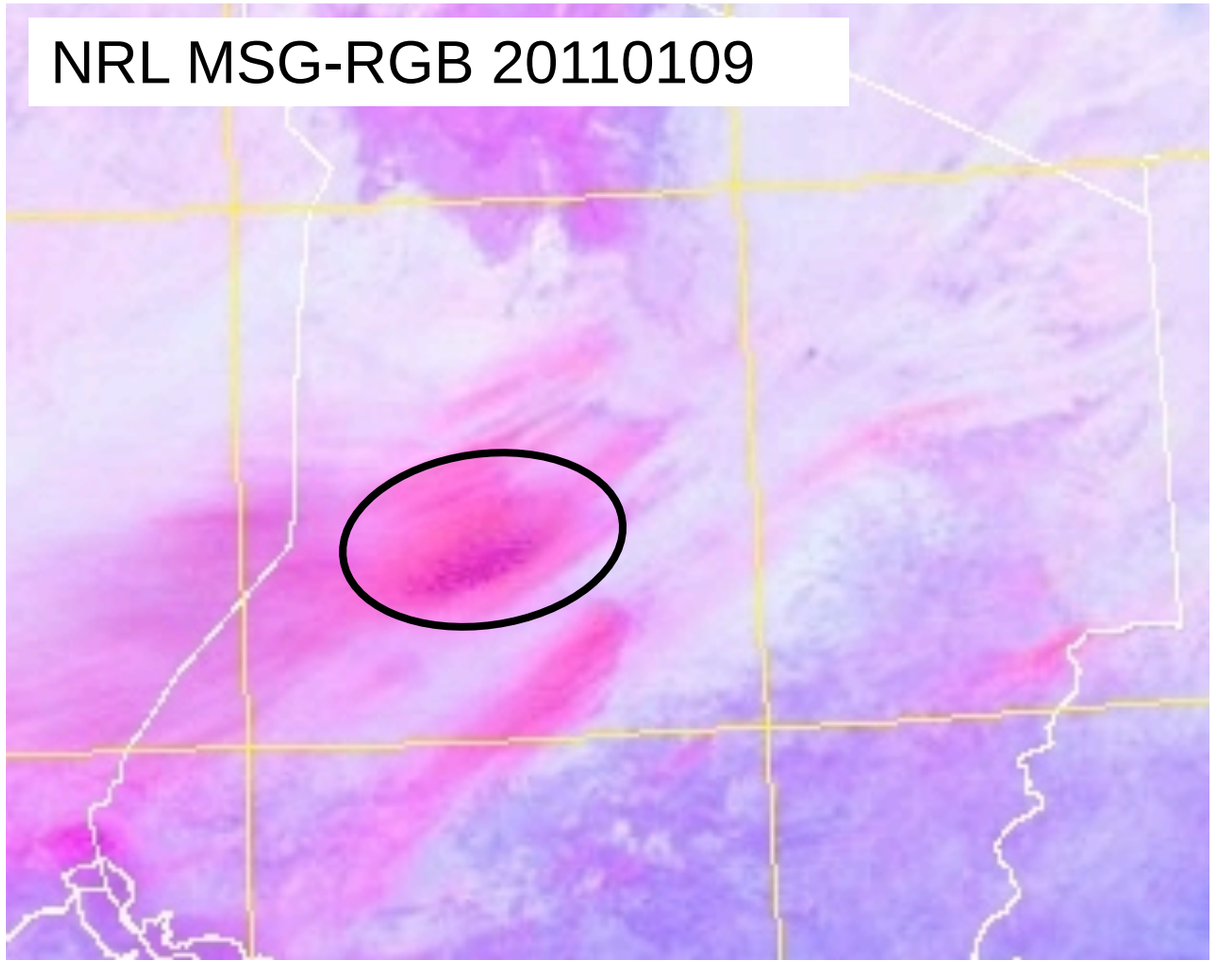
Class 137 maps diatom sediment in depression.







Selected Classes **Without Class 137**

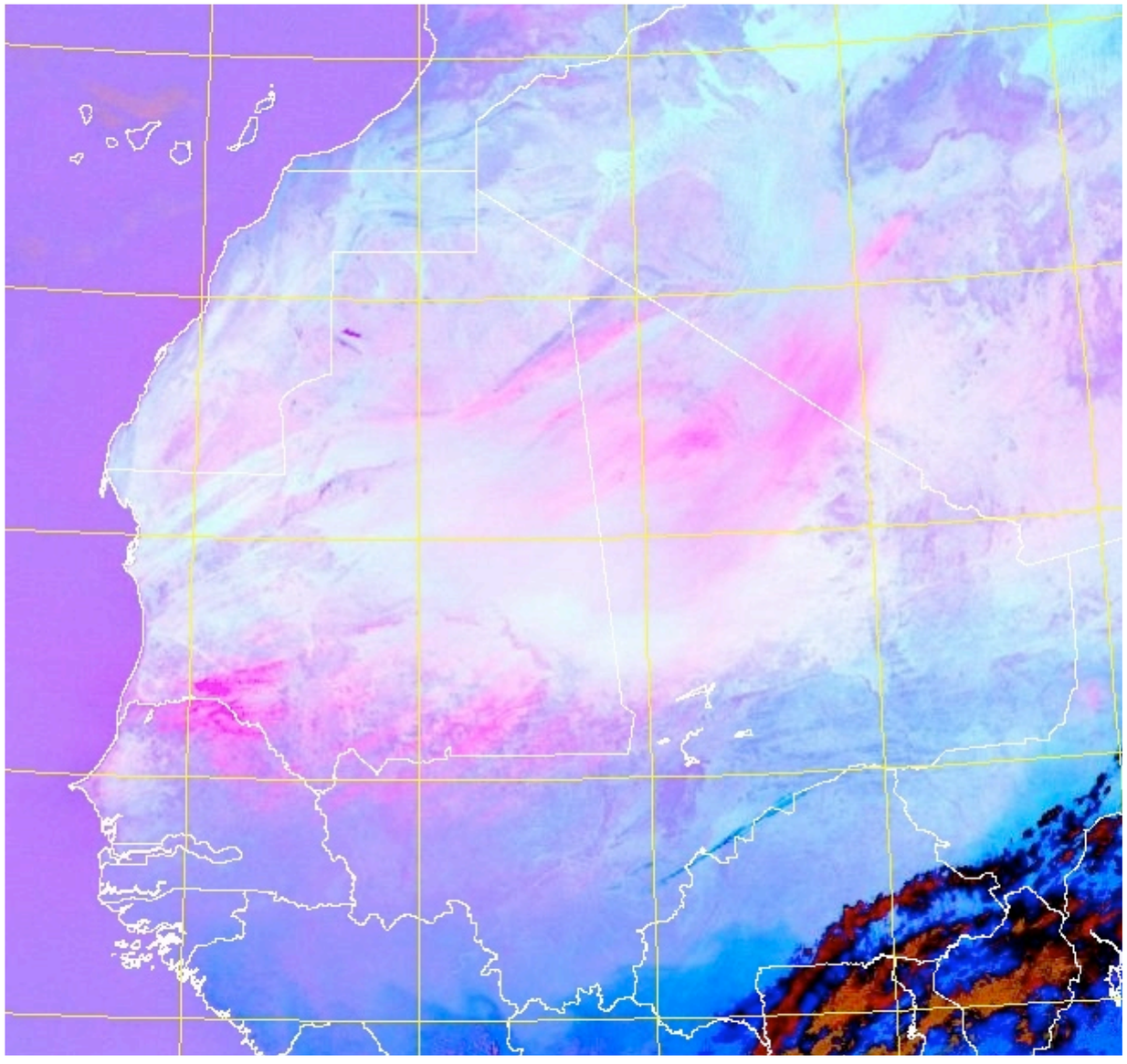


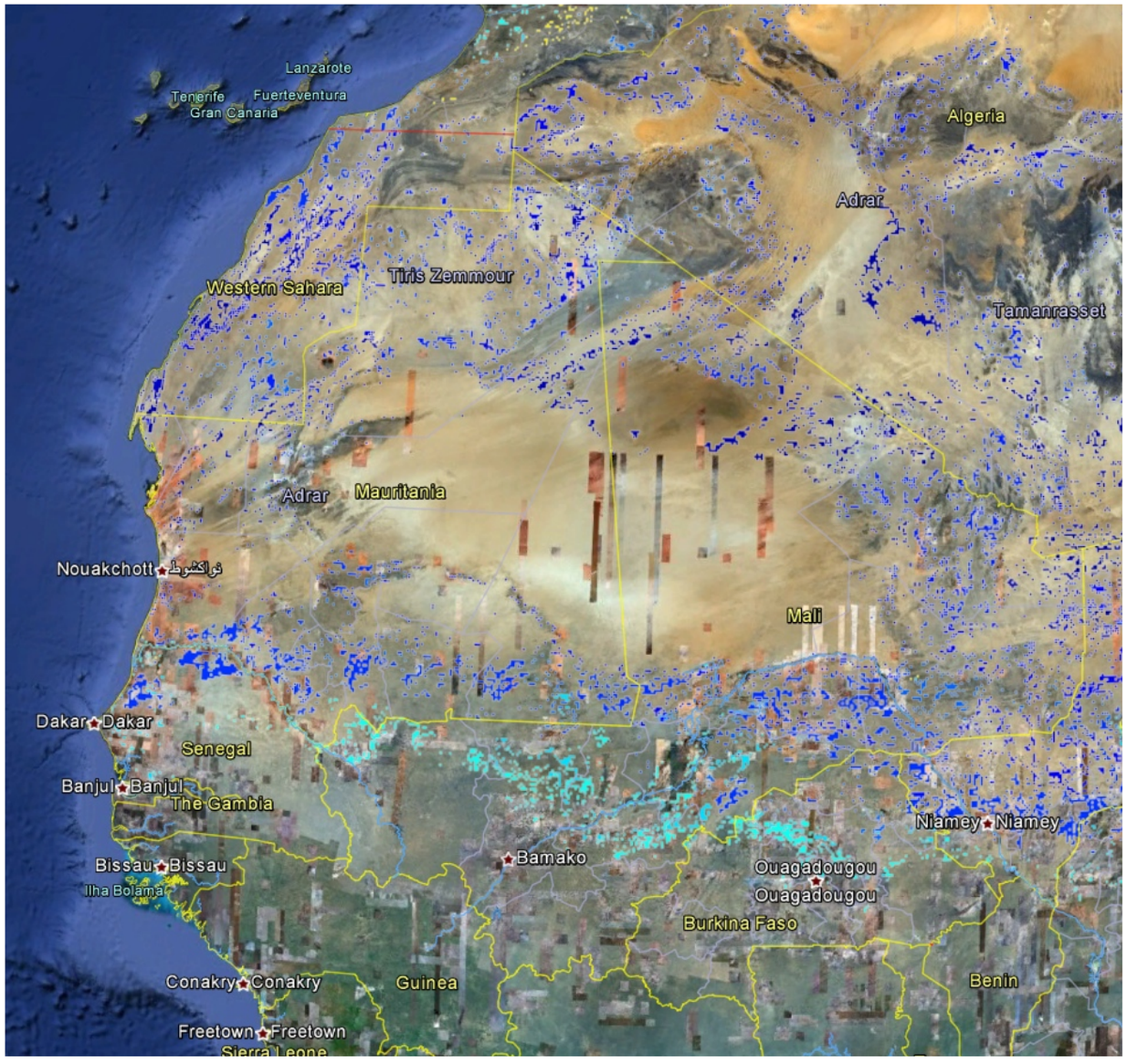
West Africa: Feb 2, 2011 13Z

Dust plumes regional location

Northern (Sahara: Algeria, Mali)
- Tidikelt
- Highlands west of Hoggar mountains
- Edges of Taoudeni basin

Southern (Sahel: Mauritania, Senegal)
- Between Tangant plateau and Bambouk mountains



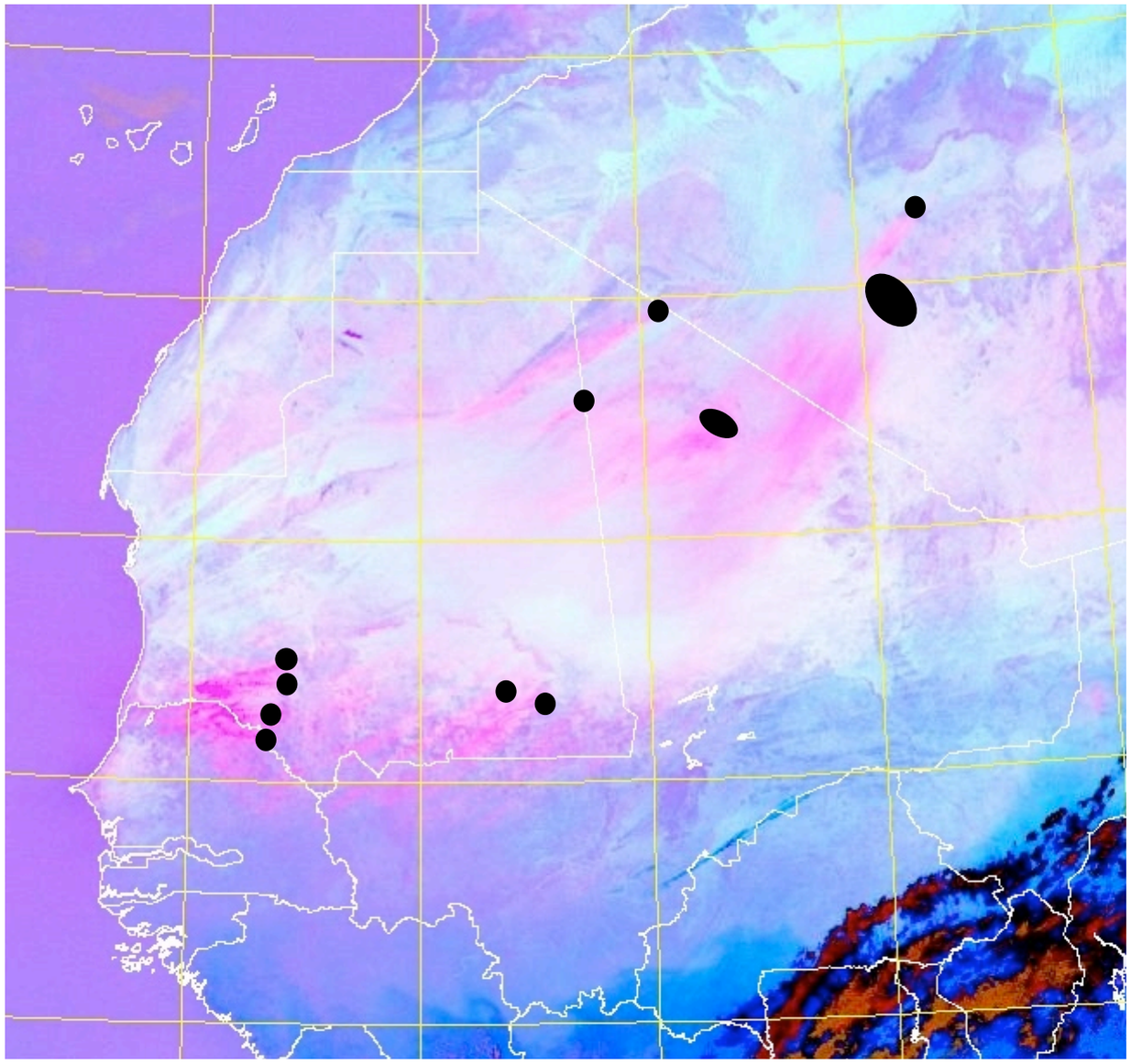


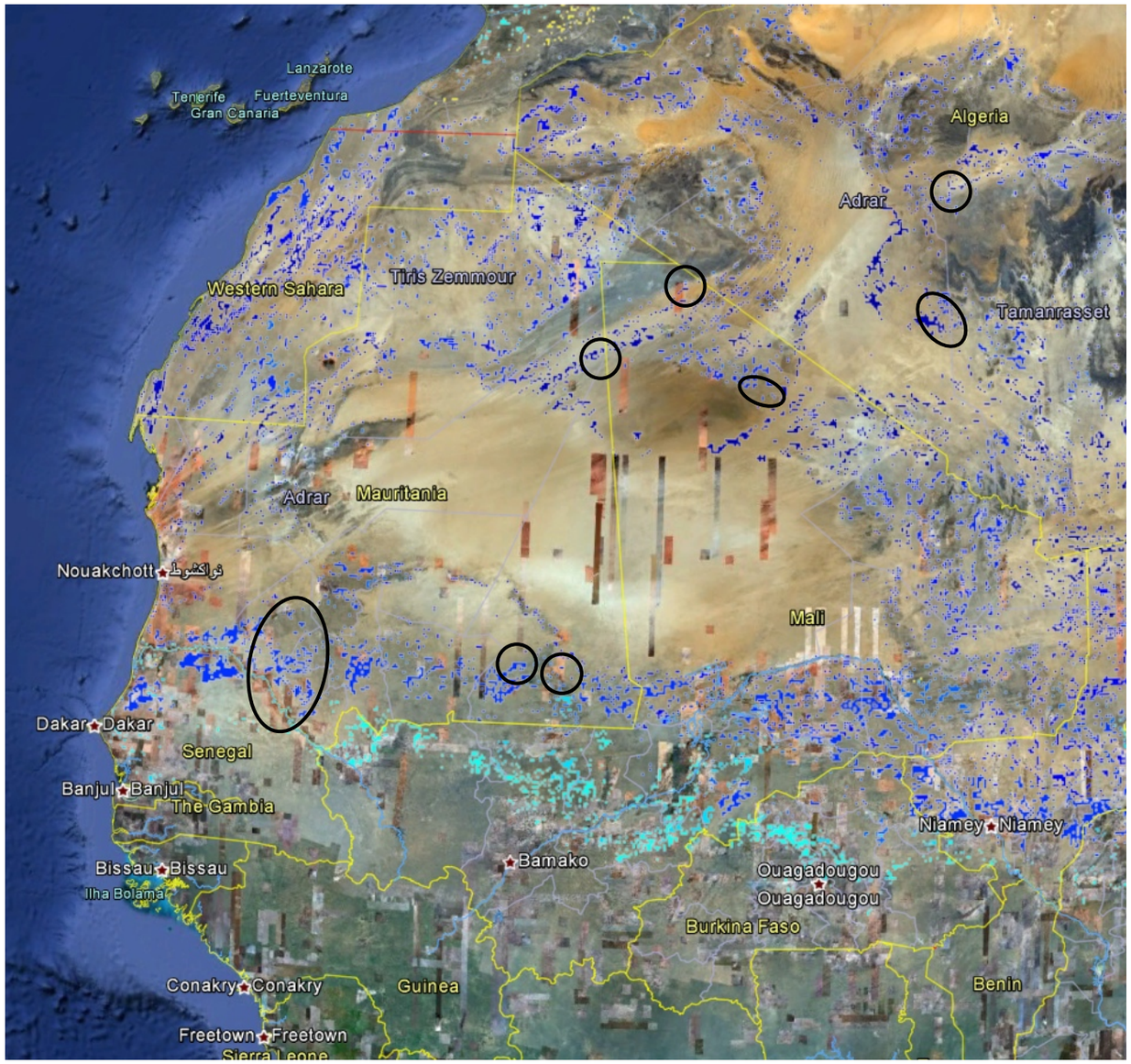
West Africa: Feb 2, 2011 13Z

Solid black circles/ovals show plume source

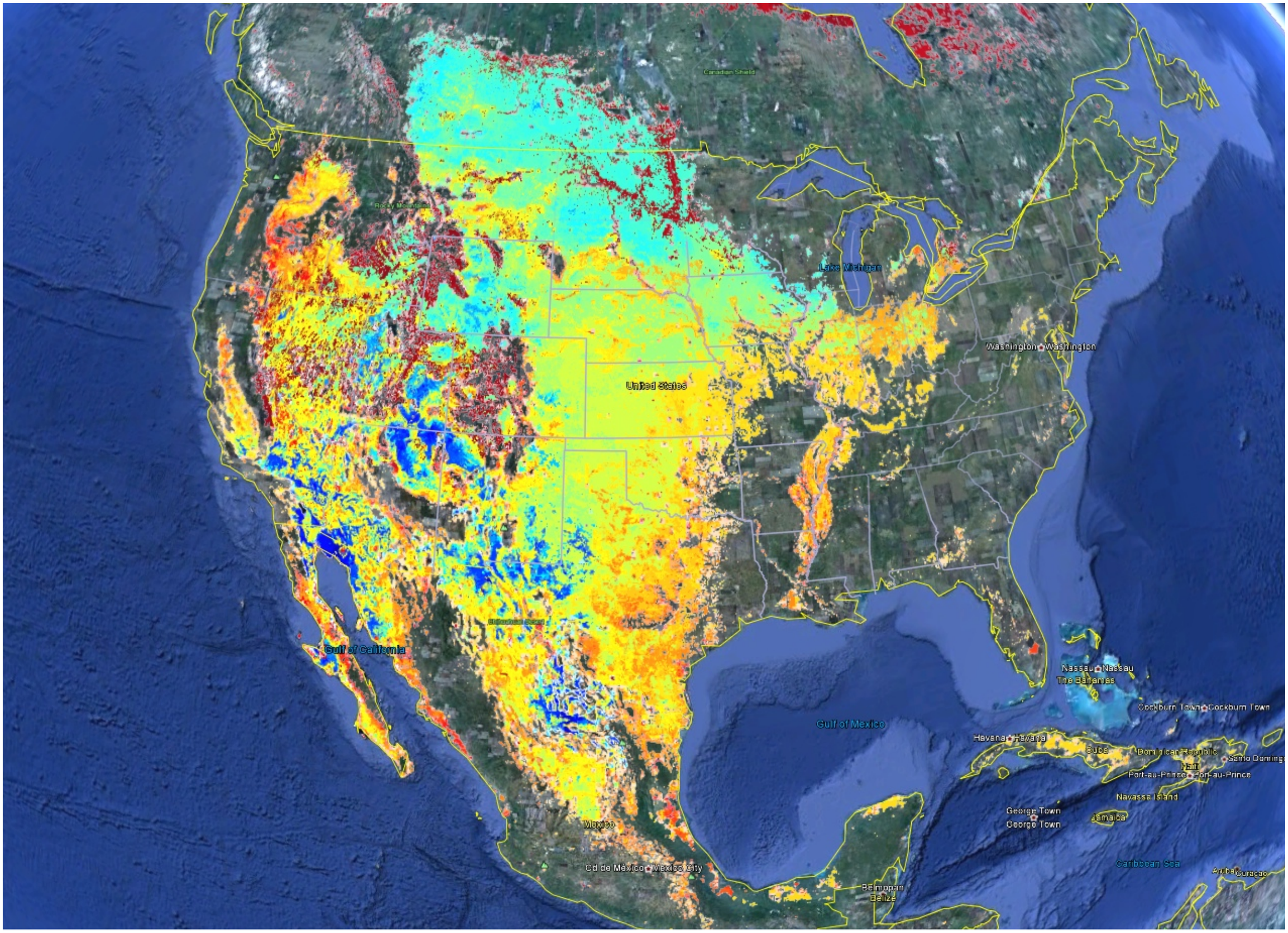
Corresponding SOM Classes within open circles/ovals

Northern Sahara: 36, 40, 63, 100
Sahel: 147, 229, 230, 405

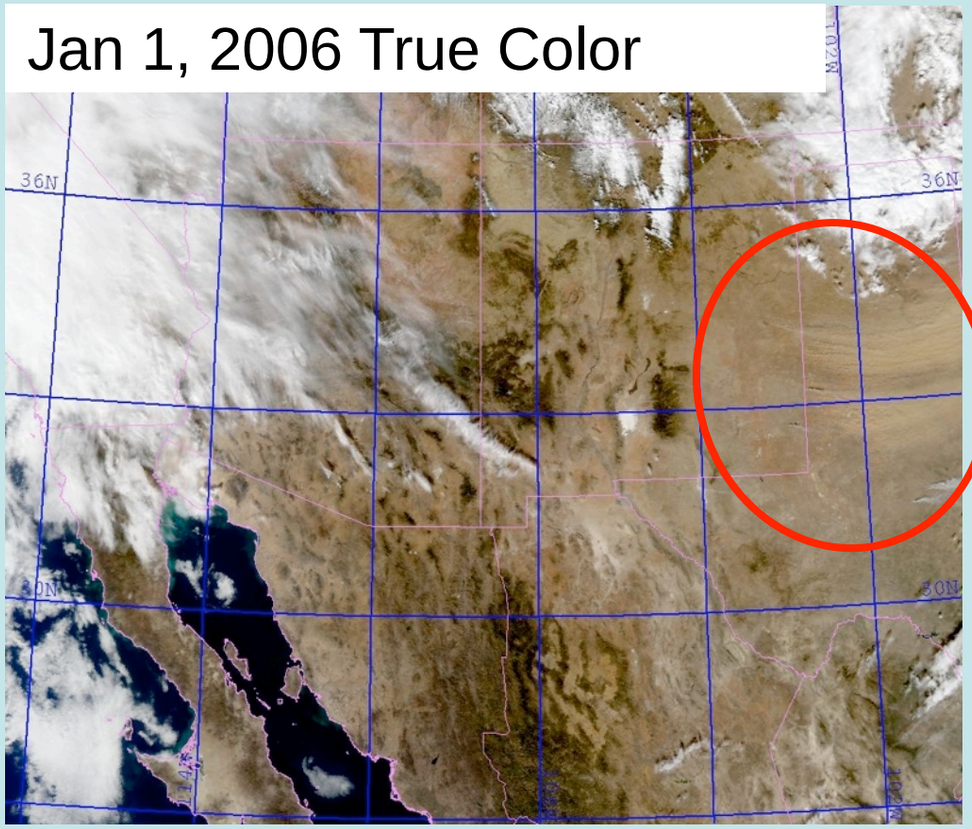


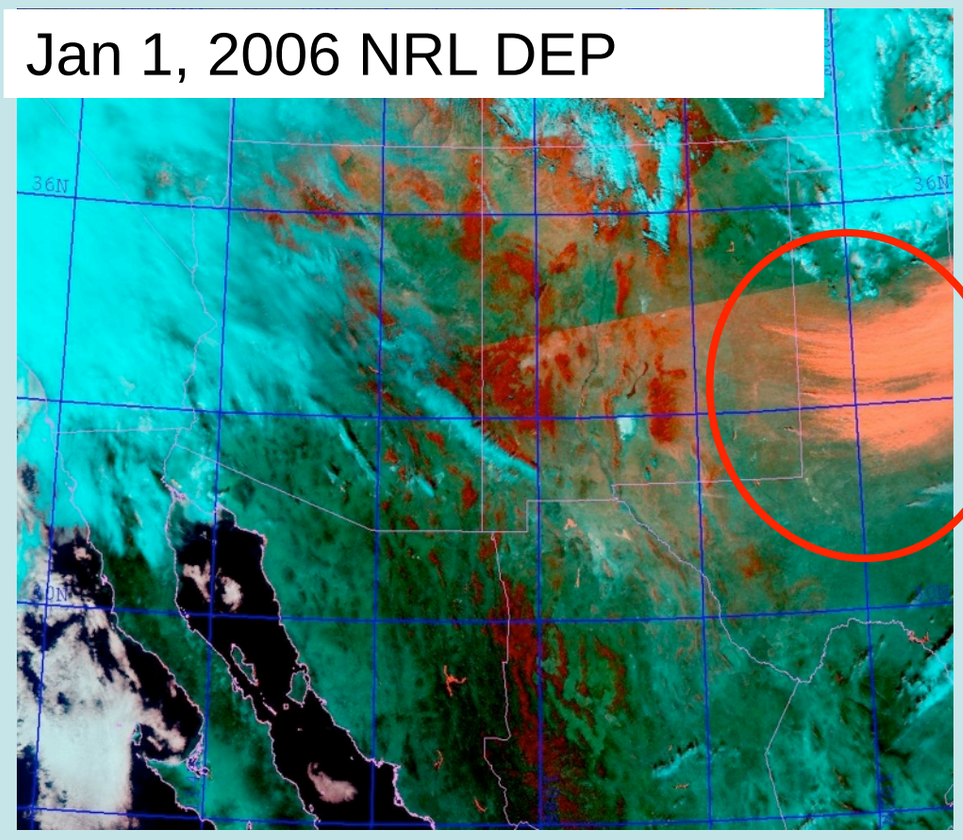


All 1000-Classes mapped for North America

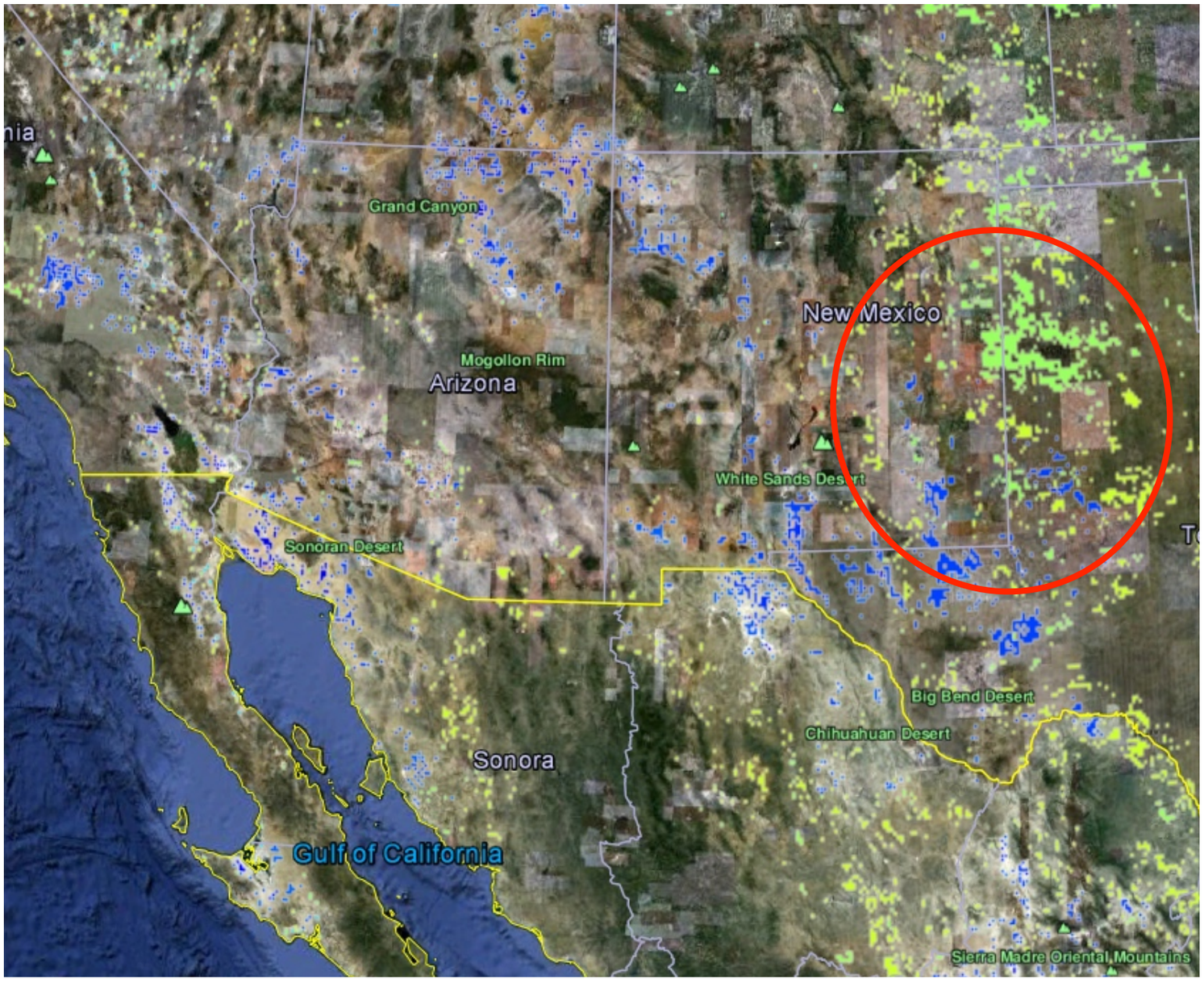


Sources along New Mexico/Texas border





Agricultural on high planes
Blue dessert areas



The North American sources have a different spectral signature than those we saw in SW Asia

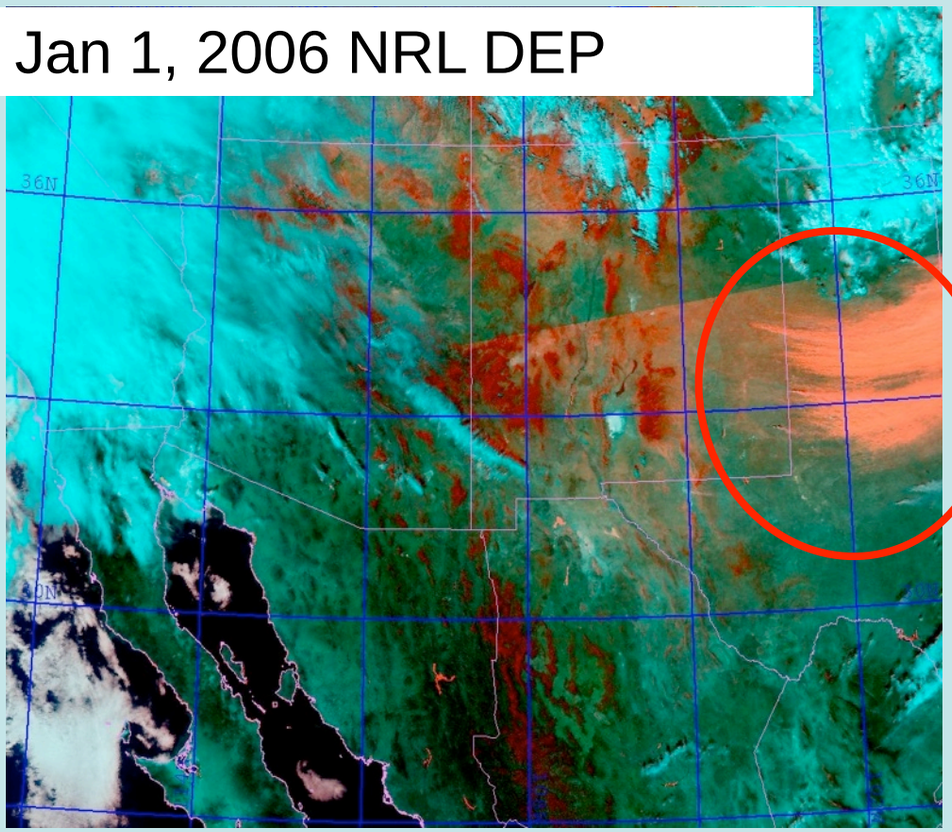
Sources along New Mexico/Texas border

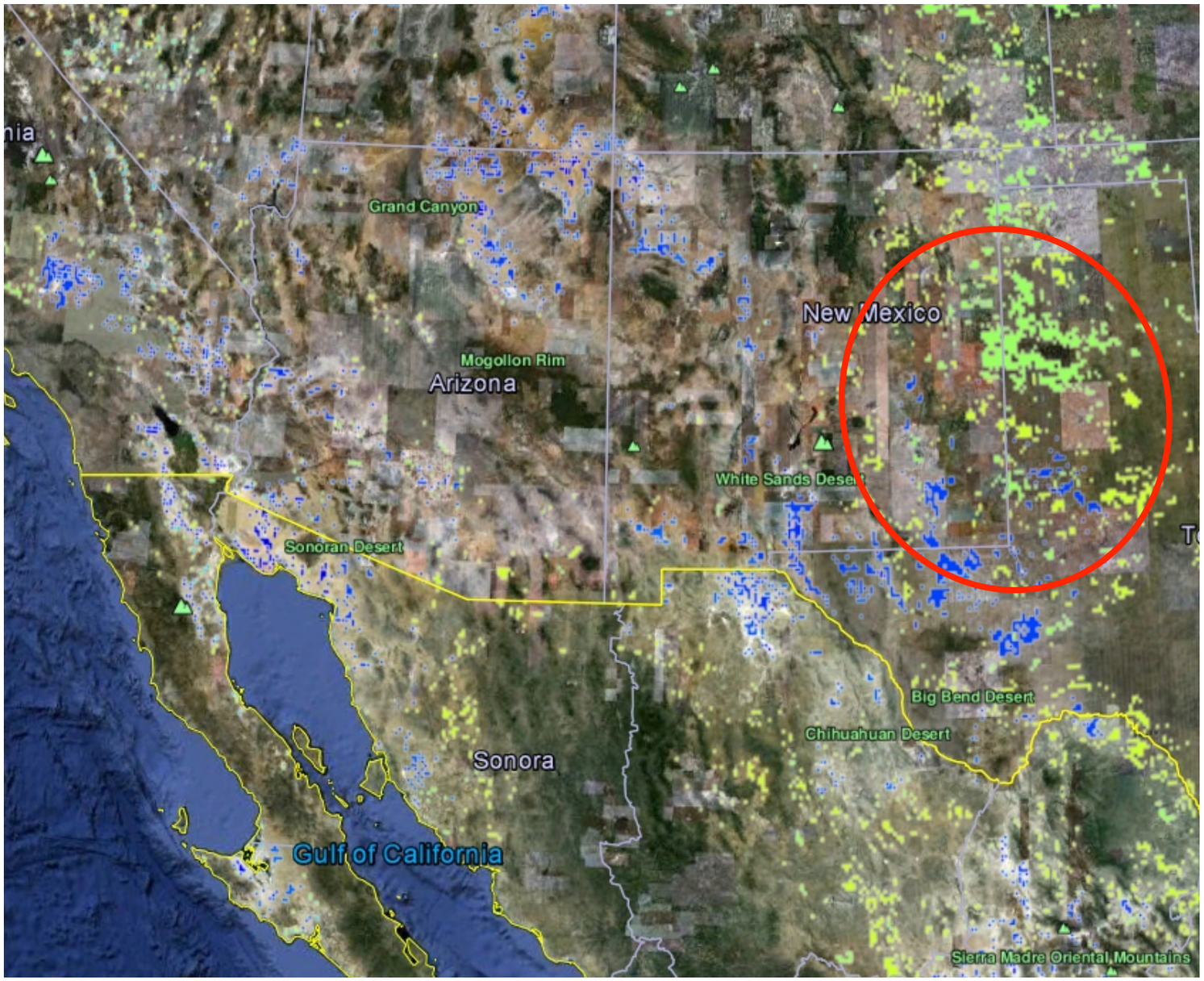
Plume Head Locations

34.45 N, 103.15 W	East of Clovis, NM
34.29 N, 103.65 W	NW of Floyd, NM
33.78 N, 103.15 W	SW of Lingo, NM
32.72 N, 103.25 W	West of Hobbs, NM

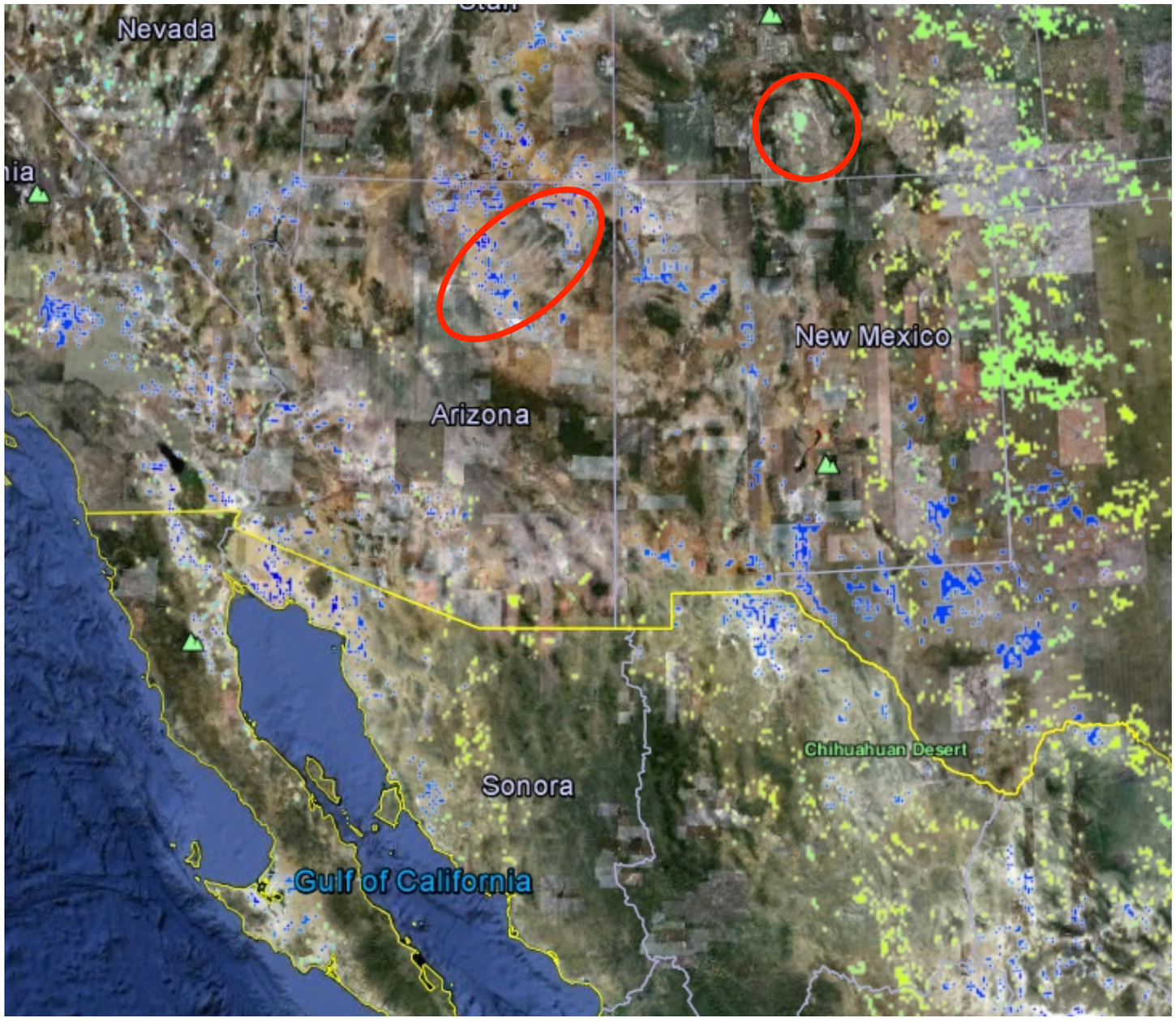
Corresponding SOM-Classes

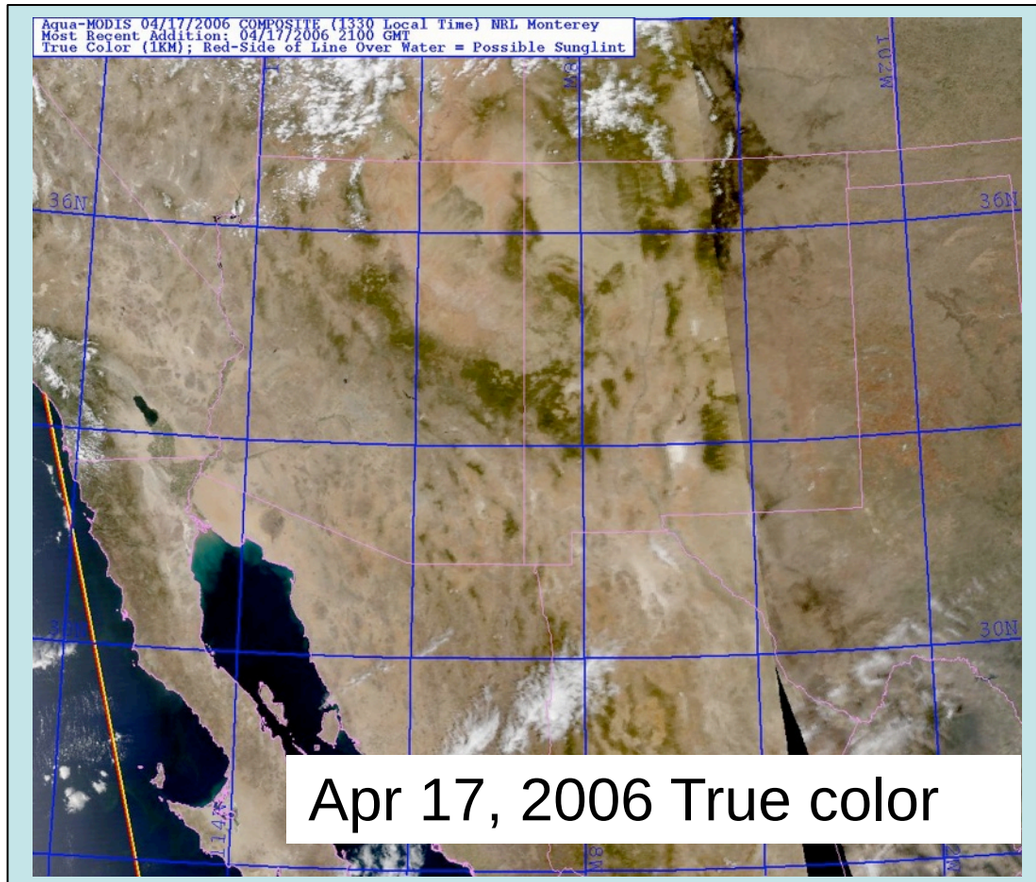
551, 552, 608 N/S High Plains

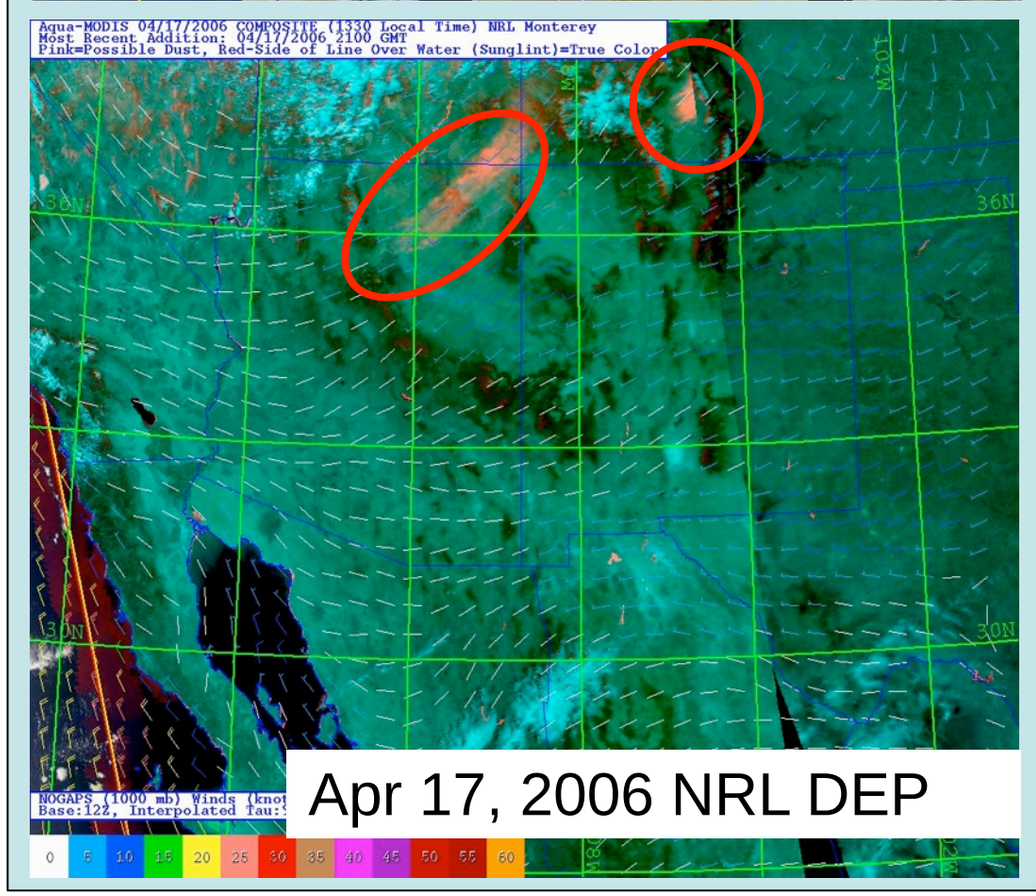




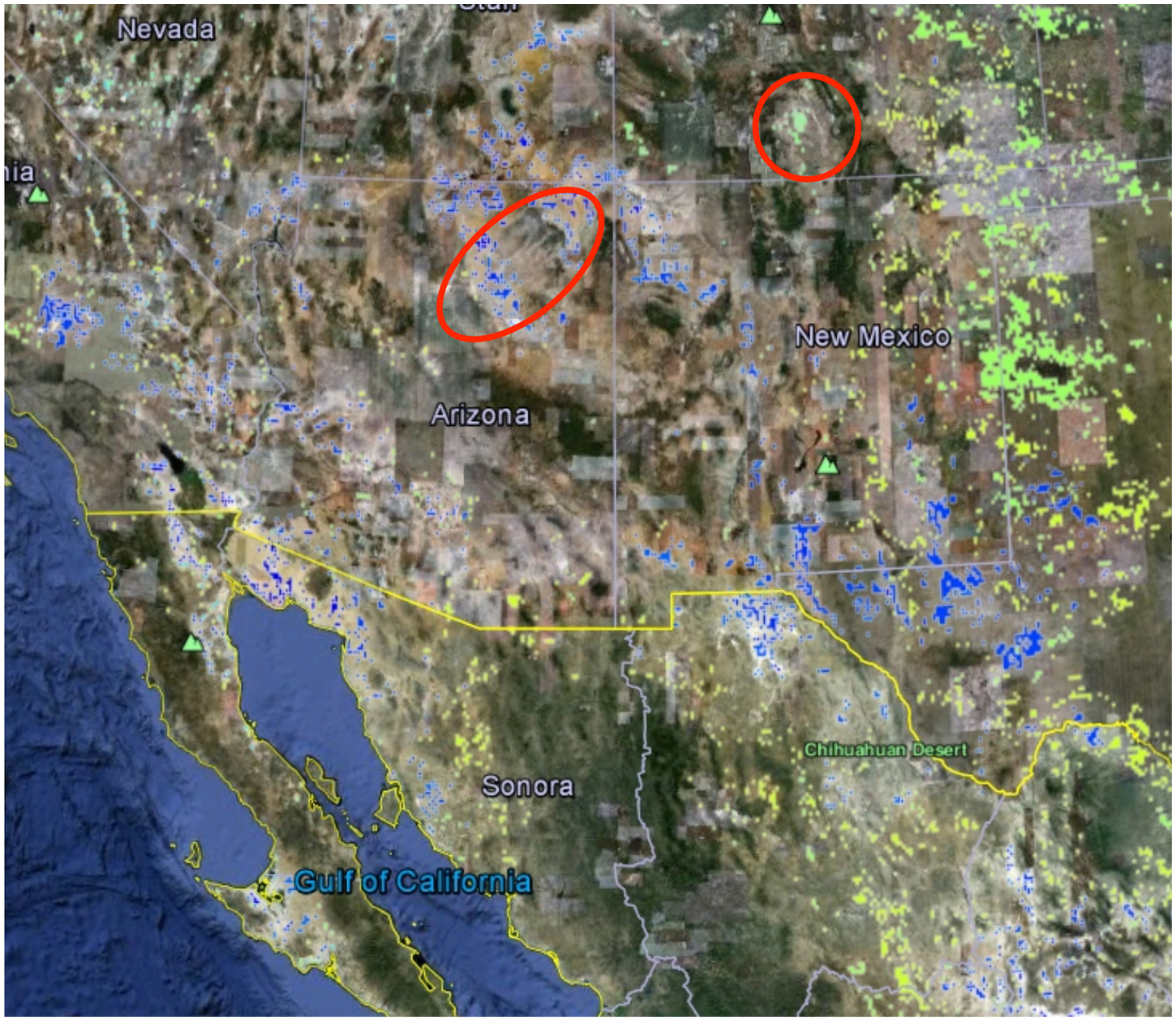
Sources in Arizona and Colorado







Sources in Arizona and Colorado

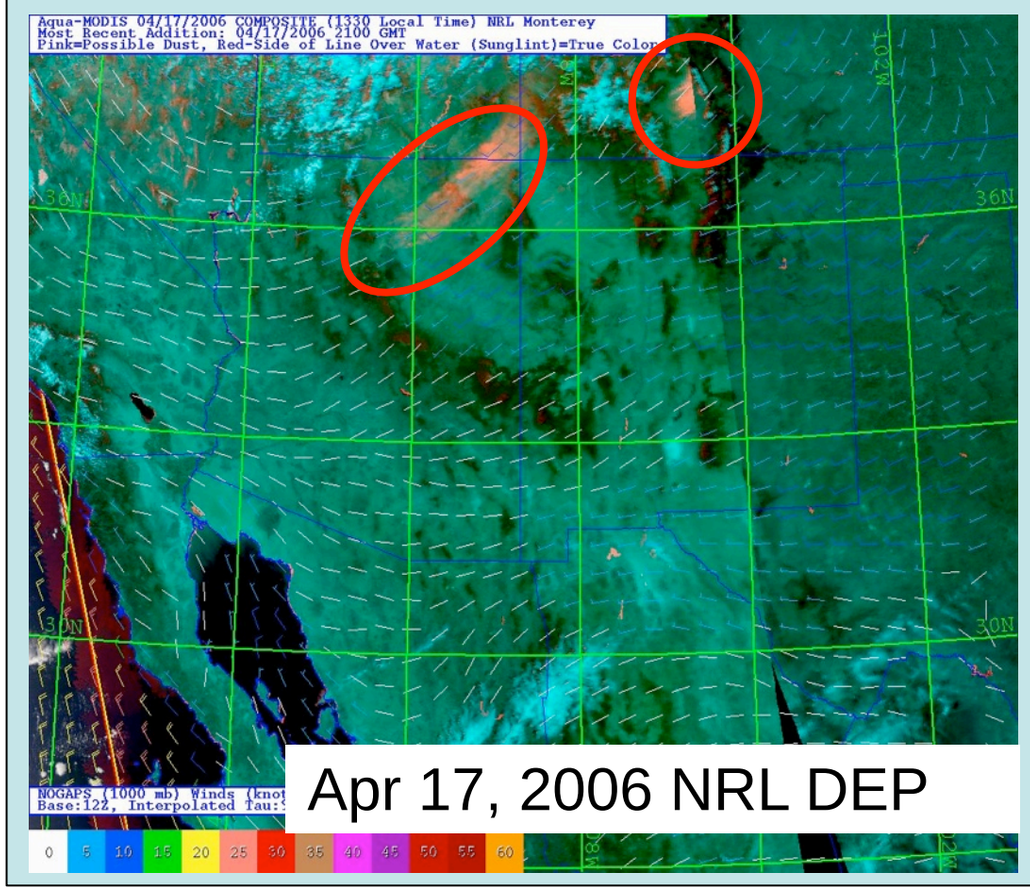


Plume Head Locations

35.85 N, 111.34 W	Painted Desert, AZ
35.75 N, 111.15 W	Painted Desert, AZ
37.85 N, 106.22 W	N of La Garita, CO
37.60 N, 106.20 W	S of Torres, CO

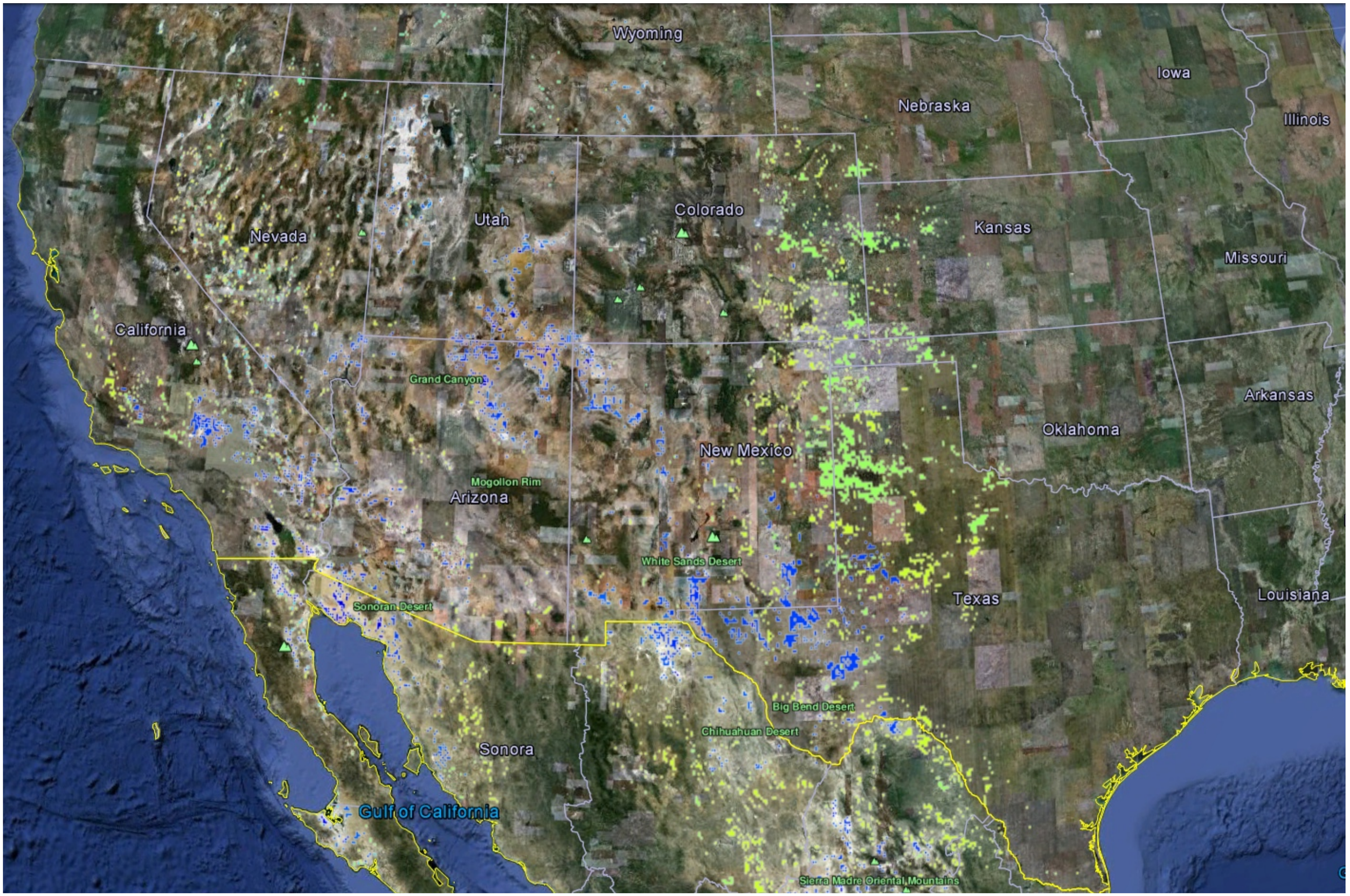
Corresponding SOM-Classes

218, 228, 229, 249 258, 260	Painted Desert, AZ
513, 521, 525, 526	San Luis Valley, CO

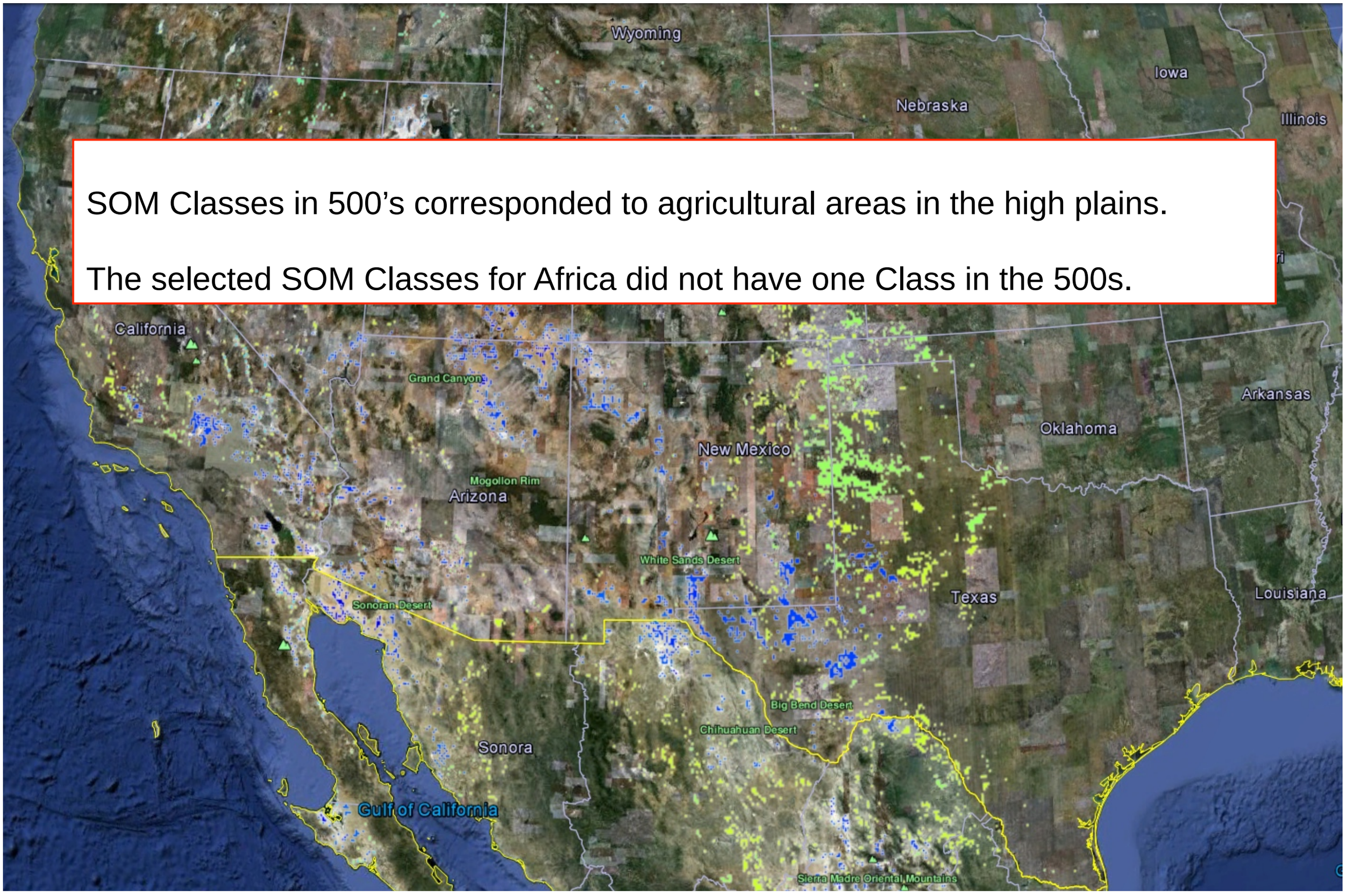


Apr 17, 2006 NRL DEP

Selected Classes for North America (n=64)



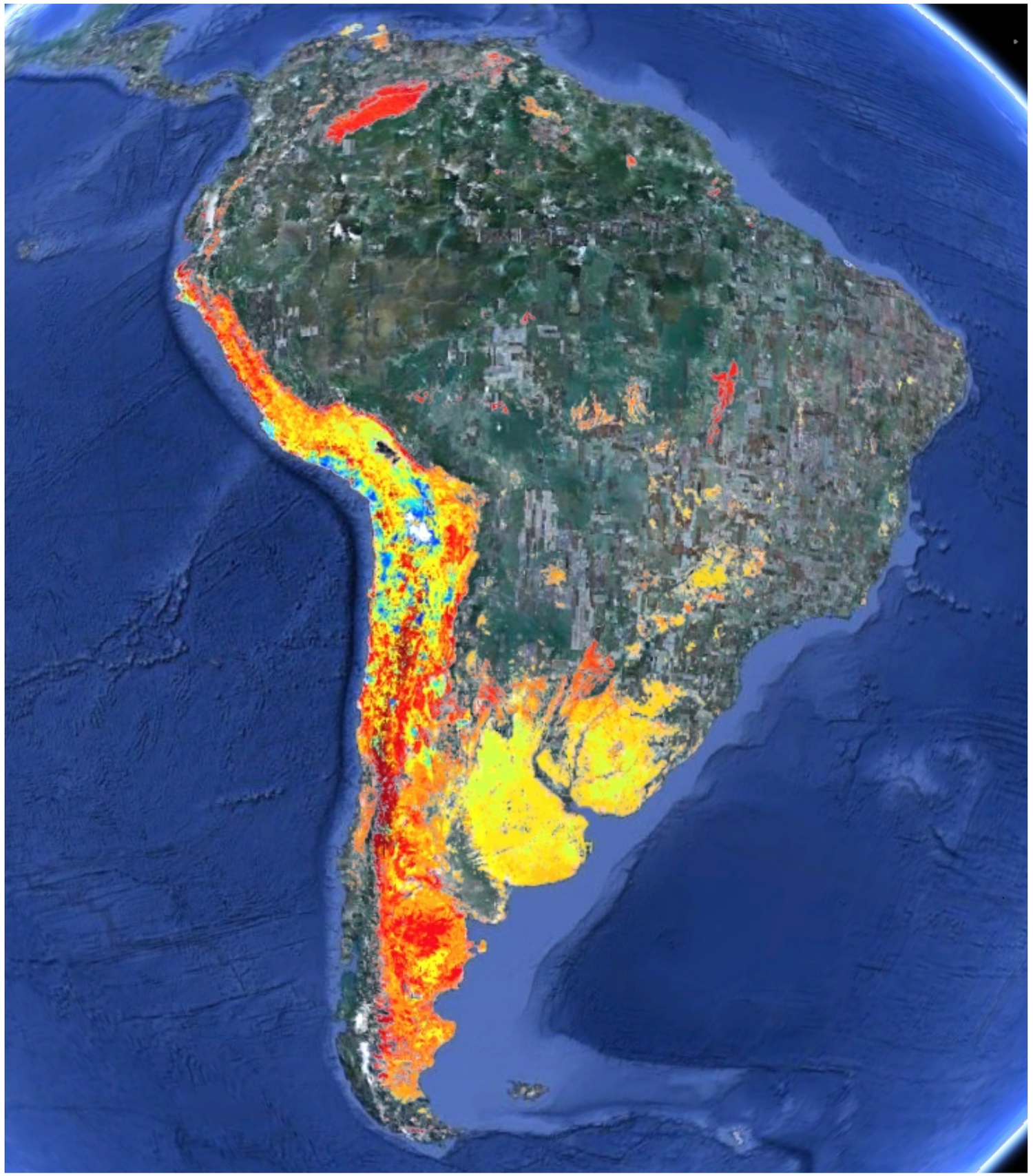
Selected Classes for North America (n=64)



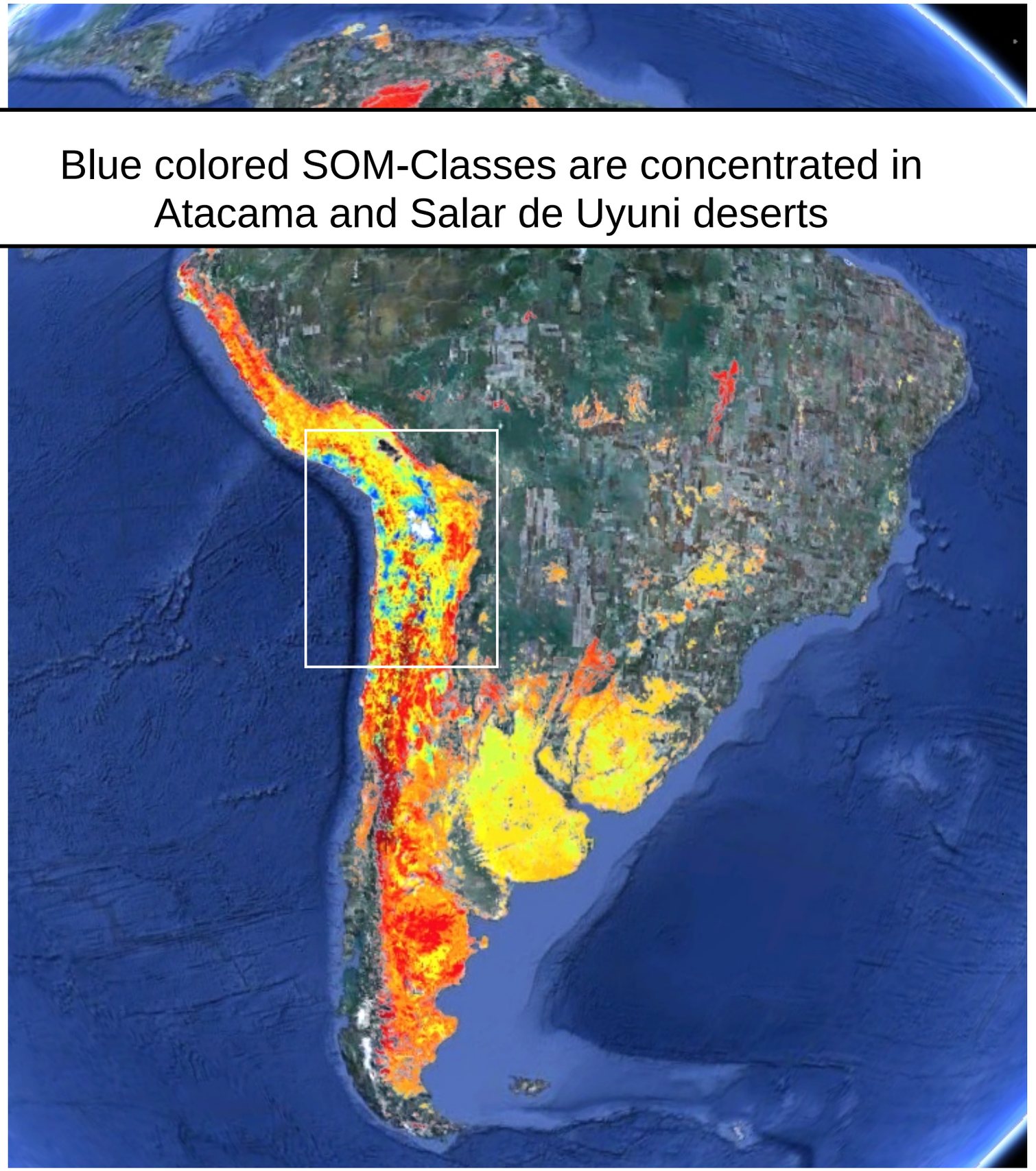
SOM Classes in 500's corresponded to agricultural areas in the high plains.

The selected SOM Classes for Africa did not have one Class in the 500s.

All 1000-Classes mapped for South America



All 1000-Classes mapped for South America



Blue colored SOM-Classes are concentrated in
Atacama and Salar de Uyuni deserts

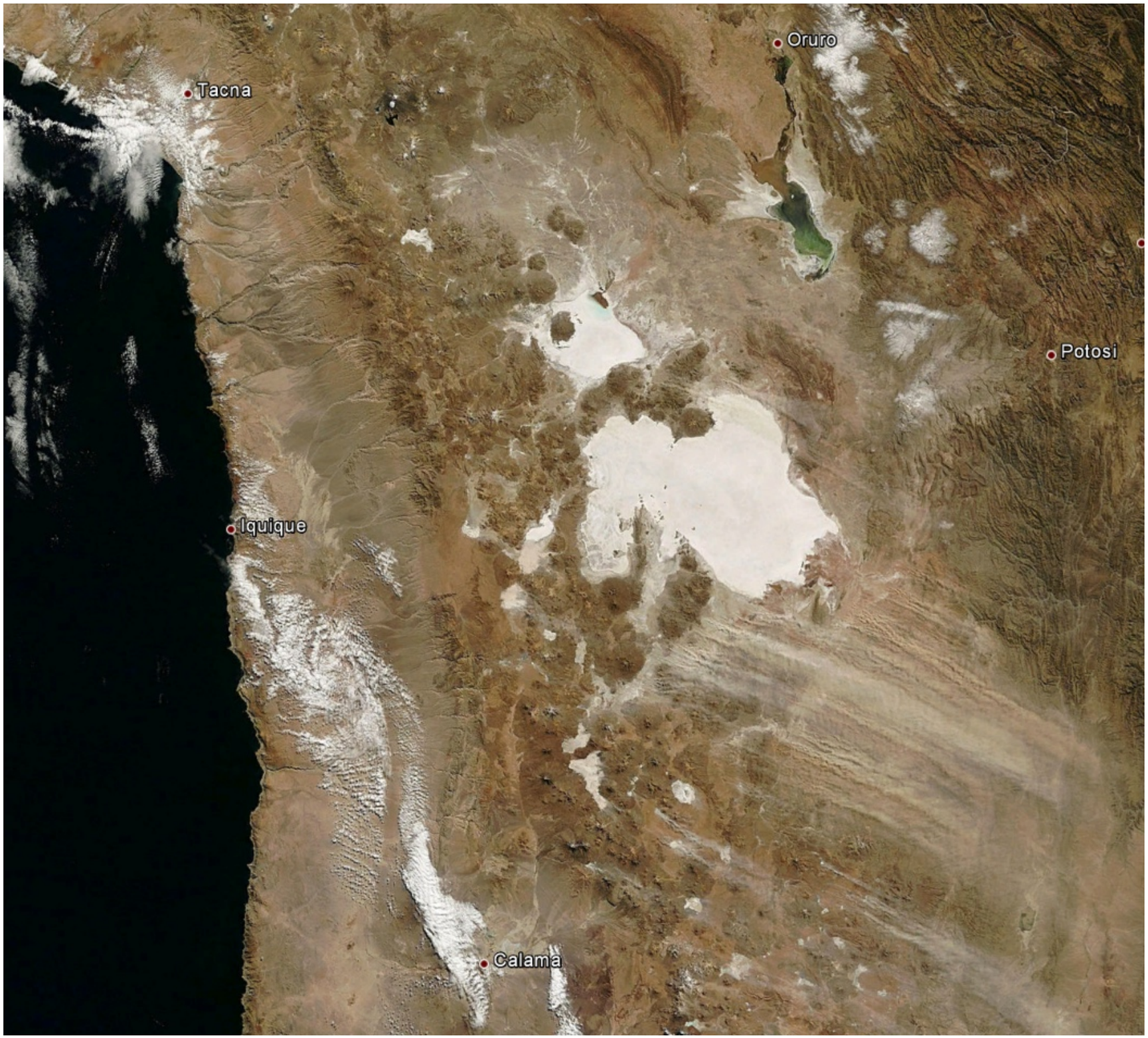
All 1000-Classes mapped for South America

Blue colored SOM-Classes are concentrated in
Atacama and Salar de Uyuni deserts

White areas are salt flats

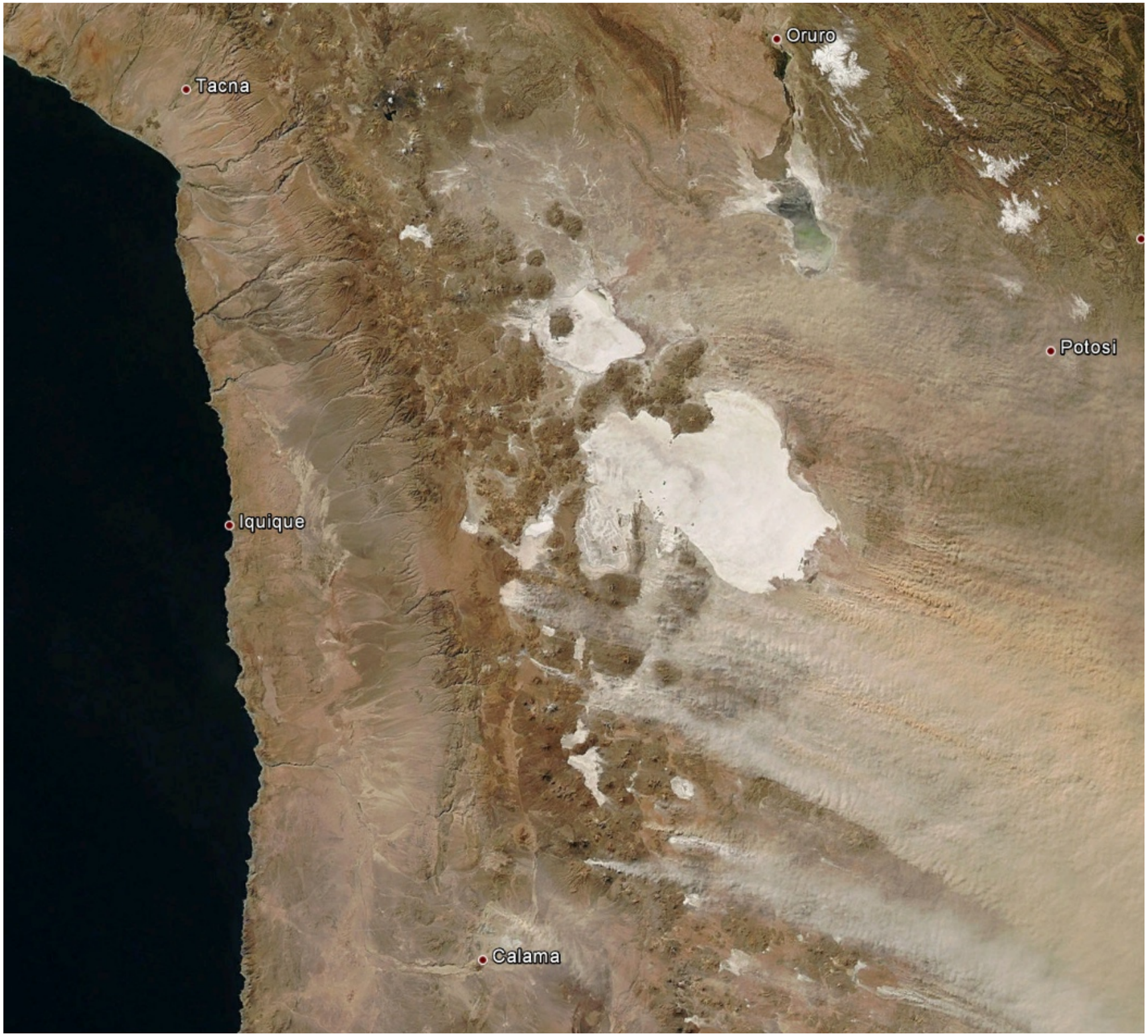
South America: Bolivia and Chile

July 21, 2009 MODIS Terra True Color



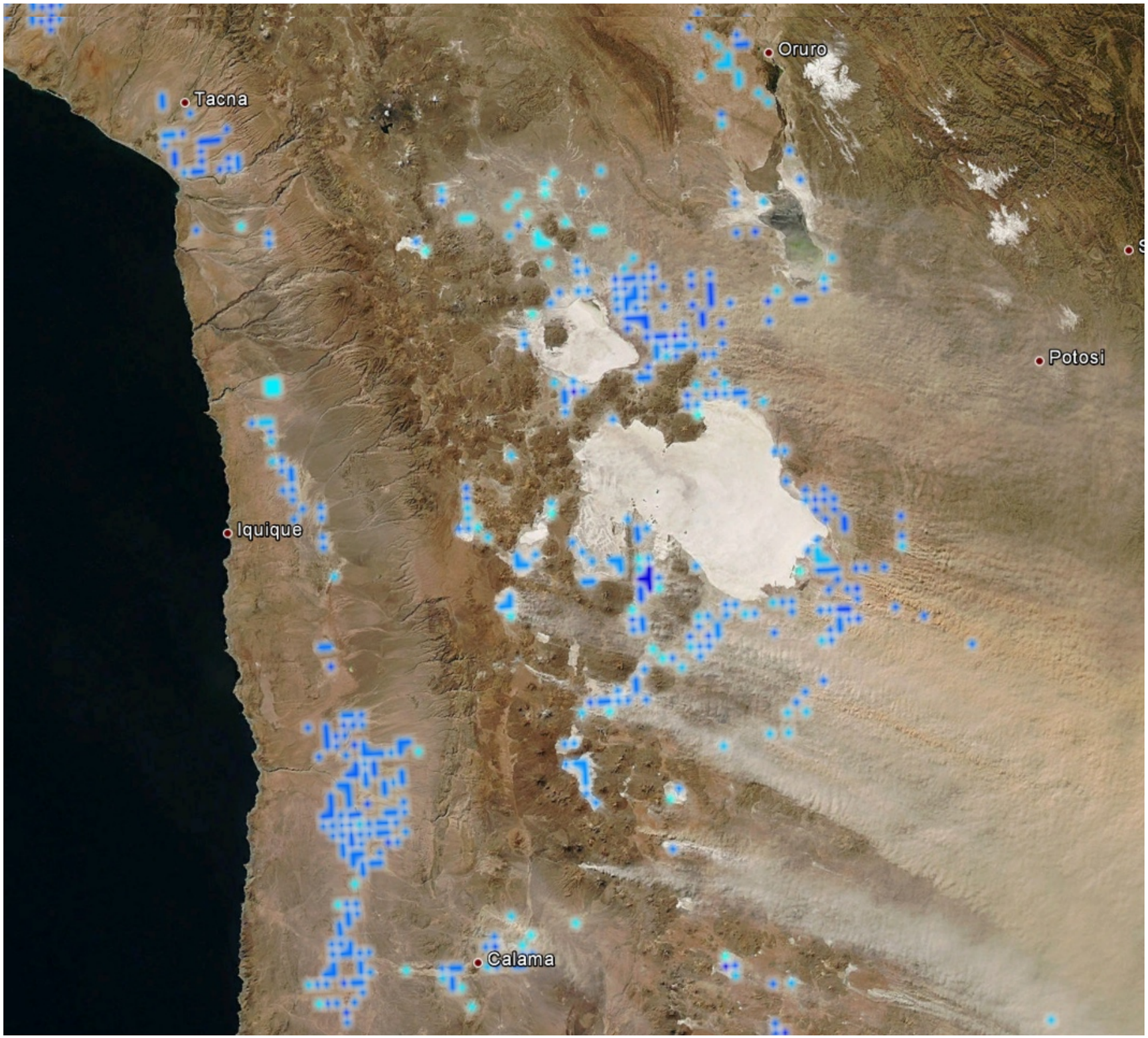
South America: Bolivia and Chile

July 18, 2010 MODIS Terra True Color



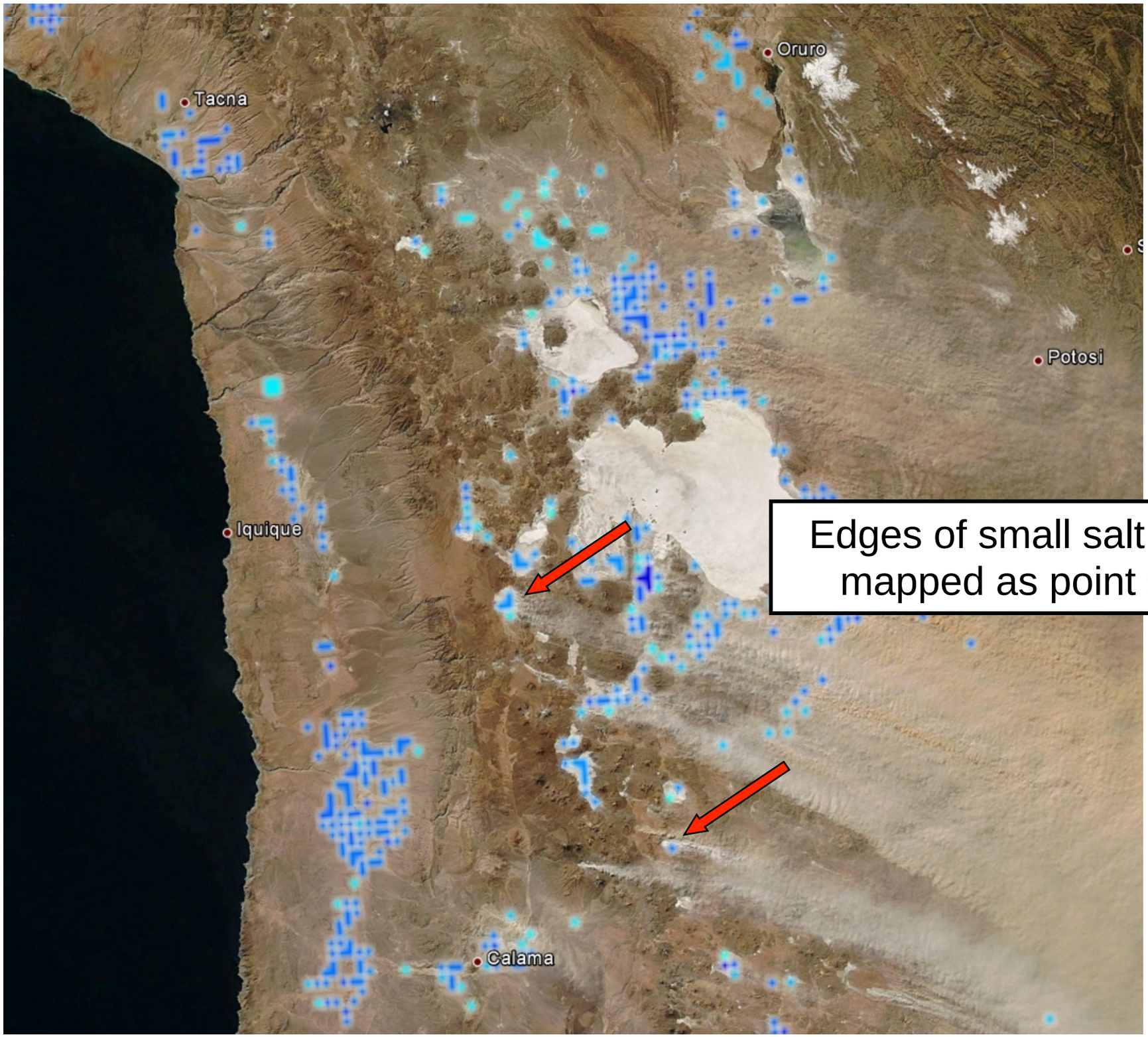
South America: Bolivia and Chile

Selected SOM-Classes in 200s, 300s, and 400s



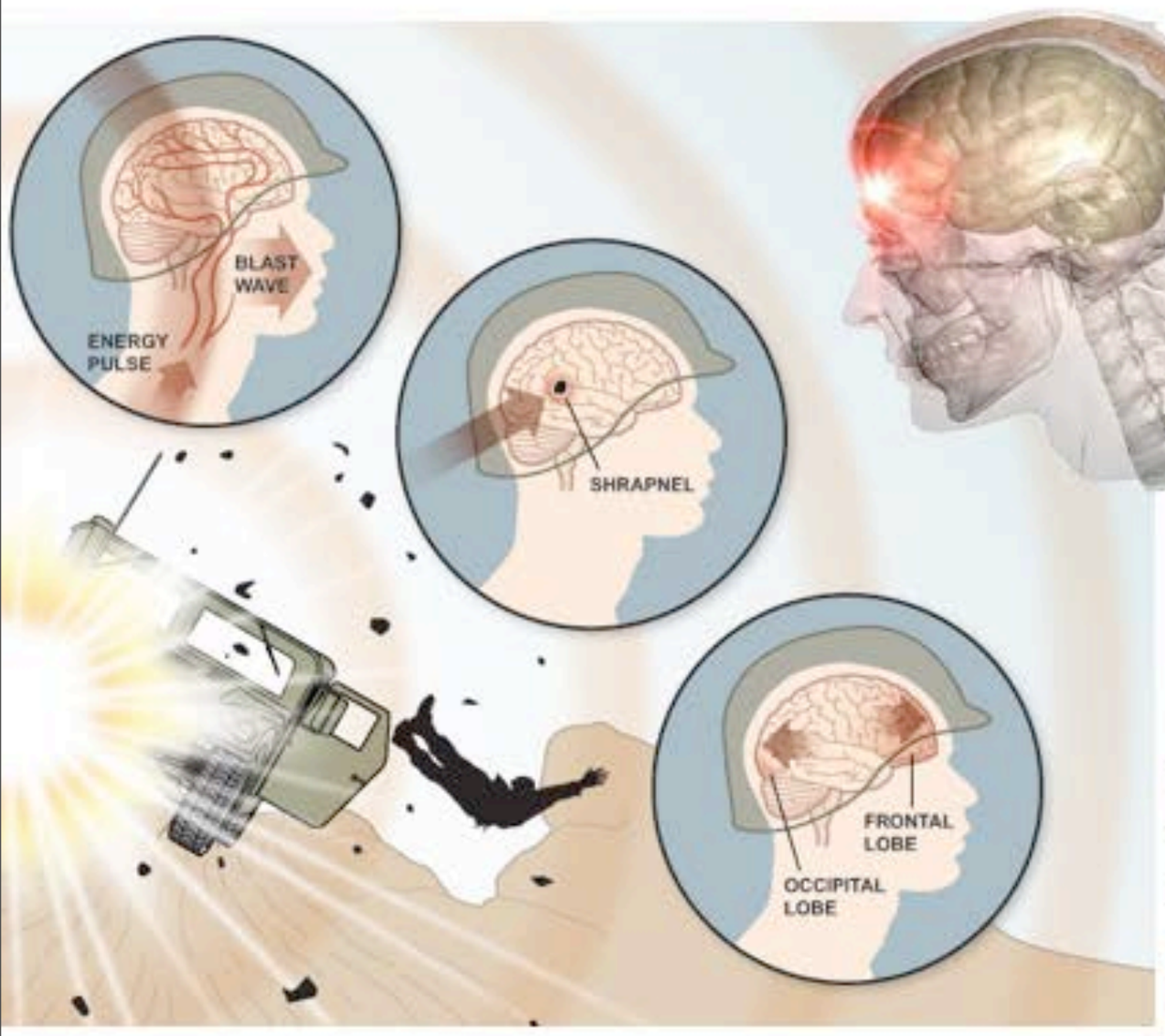
South America: Bolivia and Chile

Selected SOM-Classes in 200s, 300s, and 400s



FORECASTING AIRBORNE MOLD

Exposure to mold spore aeroallergens can trigger numerous allergic reactions ranging from mild to severe. Persons vulnerable to mold spore exposure can benefit from timely warnings of increased spore levels, and health care professionals can use this information to develop strategies to prevent some adverse outcomes. Currently, estimates of mold spore levels are based on data acquired from a limited number of ground-monitoring stations that are unavailable in many parts of the country. This method of data collection and processing is also quite expensive.



HELPING WOUNDED WARRIORS

Traumatic brain injury (TBI) is a serious public health problem in the United States. Each year, traumatic brain injuries contribute to a substantial number of deaths and cases of permanent disability. Every year, at least 1.7 million TBIs occur either as an isolated injury or along with other injuries.

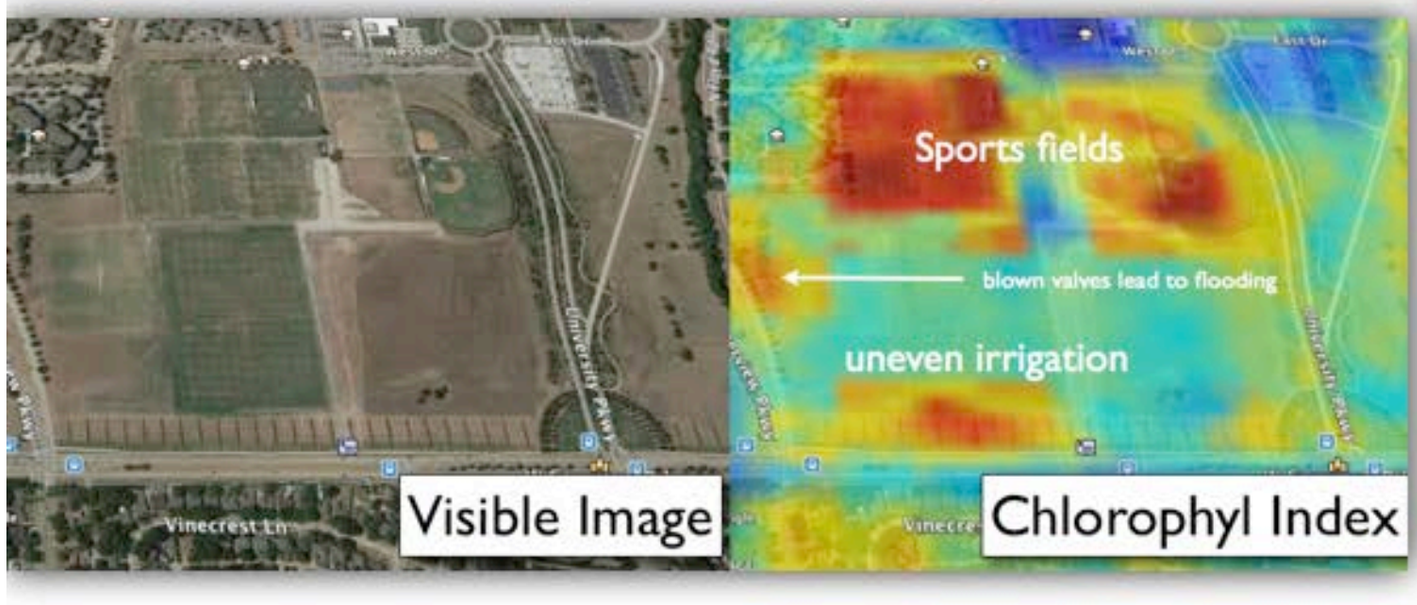
LIFE EXPECTANCY

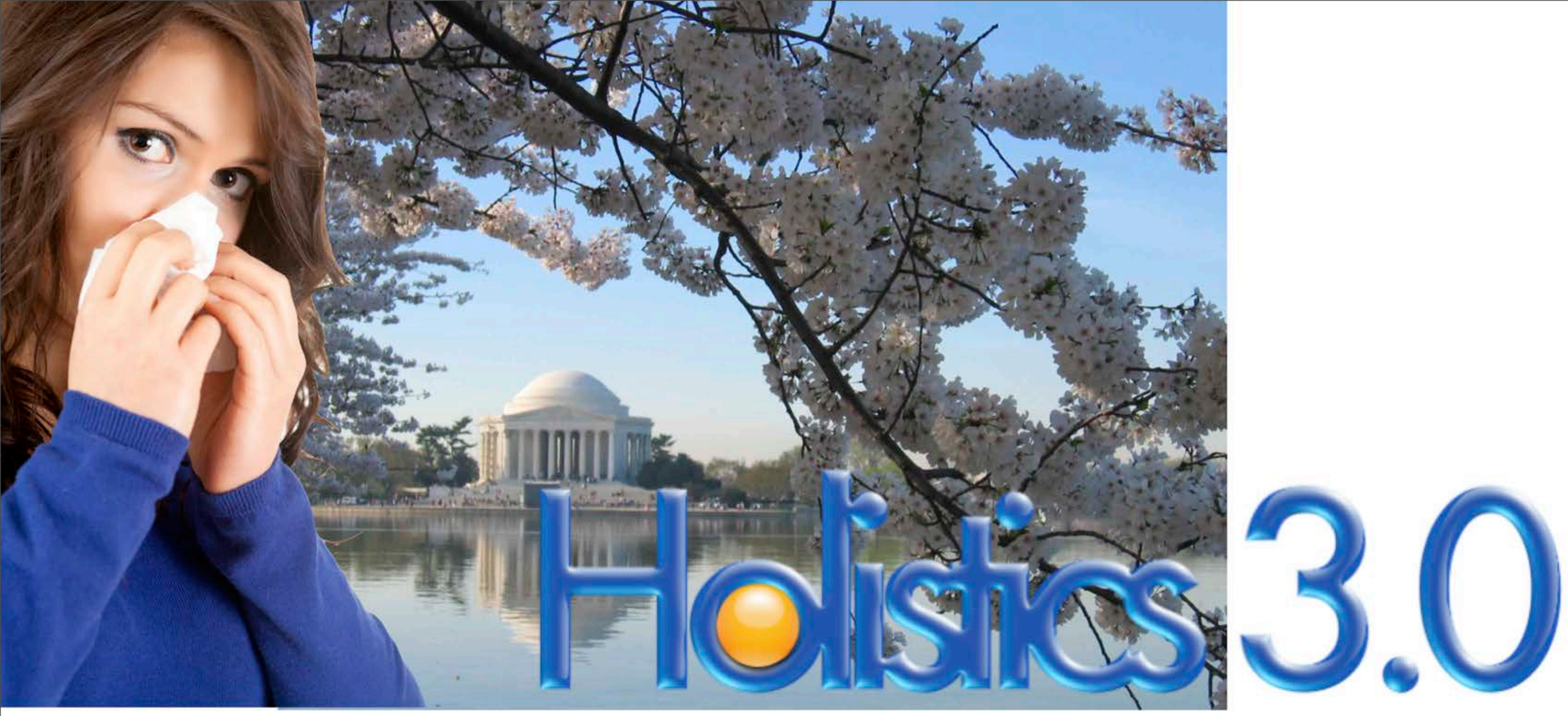
EFFECTIVE PUBLIC HEALTH POLICY

A new age is dawning with actionable insights being provided by a combination of holistic datasets comprehensively describing issues, coupled with unprecedented computing power and machine learning. For the first time in man's history we have the chance to comprehensively characterize problems and objectively focus on the key issues.

WATER CRISIS THREATENS US FOOD SECURITY

The global water crisis – caused by drought, flood, and climate change – is about recognizing water's true value, using it efficiently, and planning for a different future.





http://holistics3.com

David.Lary@utdallas.edu