Development of IDL-based geospatial data processing framework for meteorology and air quality modeling

Hyun Cheol Kim [1,2,3], Fong Ngan [1,2,4], Pius Lee [2,3], and Daniel Tong [1,2,5]

[1] Cooperative Institute for Climate and Satellites, University of Maryland, College Park, Maryland
[2] Air Resources Laboratory, NOAA, College Park, Maryland
[3] Co-Principal Investigator
[4] Investigator
[5] Principal Investigator

Outline

- Project goal
- Development of IGDP
  - Polygon clipping algorithms
  - Raster data processing
- Applications
- Advanced applications
  - Conservative downscaling: Satellite NO2 column comparison
  - Spatial allocating: Emission surrogates

Project goal

- To investigate basic computational algorithms to handle Geographic Information System (GIS) data and satellite data, which are essential in regional meteorological and chemical modeling.
- To develop a set of generalized libraries within a geospatial data processing framework aiming to process geospatial data more <u>efficiently</u> and <u>accurately</u>.

Geospatial Inputs for model simulation

Meteorology model

- Land use land cover (USGS LULC, IGBP, …)
- Vegetation fraction (e.g. NDVI)
- Leaf Area Index (LAI)
- Soil type

Geospatial data

Chemistry model

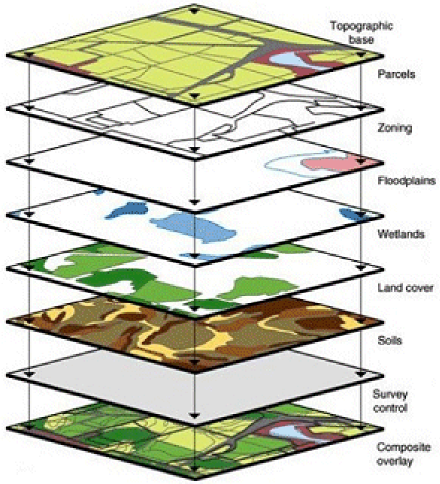
- Land use land cover (USGS LULC, IGBP, …)
- Land file (land type, tree speciation)
- Ocean file (Open ocean fraction, surf zone)

Emission model

- Surrogates
- LULC, Land/Ocean file
- Vegetation data for biogenic emission

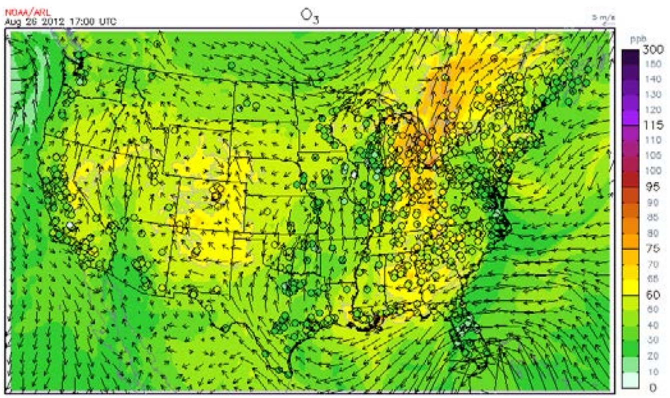
IDL-based Geospatial Data Processor (IGDP)

- Developed by NOAA/ARL funded by TCEQ/AQRP.
- Fast and accurate polygon, polyline, and pixel data processing capability
- Conservative conversion (e.g. regridding) between different map projections and domain settings.
- GIS, model comparison
- Emission processing (e.g. surrogate files)



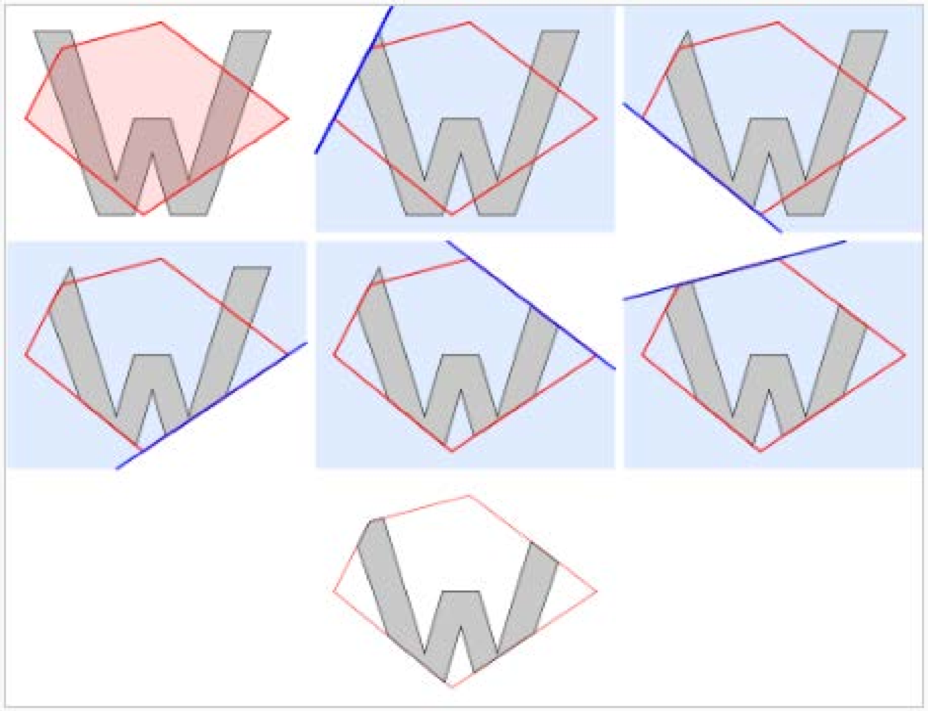
Polygon, polyline, and pixels (GIS, Satellites, Emission)

IGDP



Gridded world (Model)

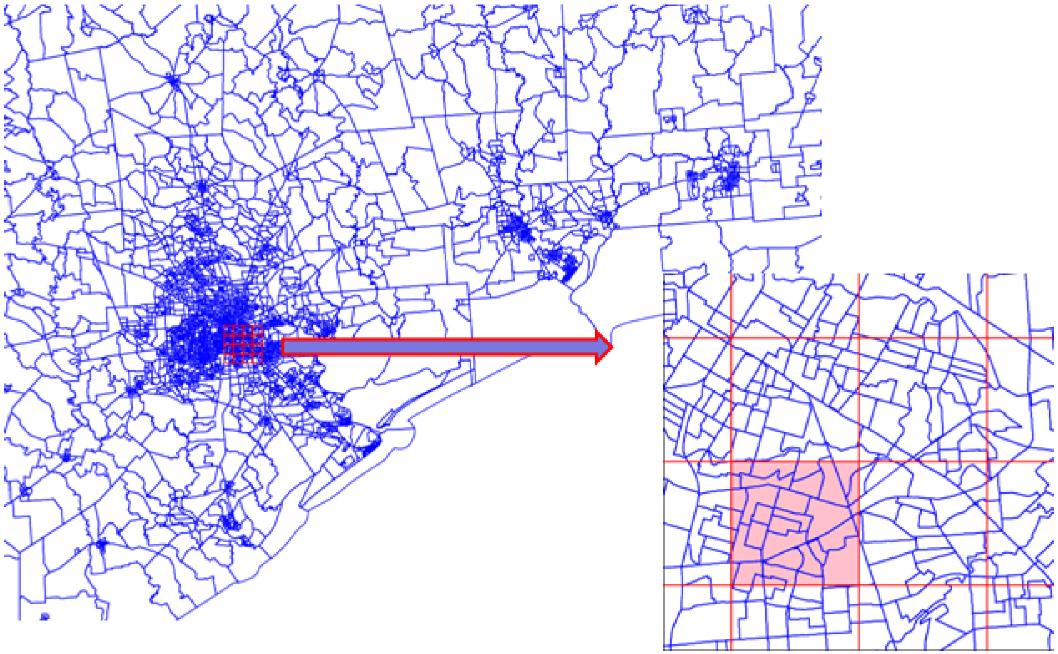
Polygon clipping algorithms



The Sutherland-Hodgman algorithm is used for clipping polygons. It works by extending each line of the convex clip polygon in turn and selecting only vertices from the subject polygon that are on the visible side.

Spatial allocator

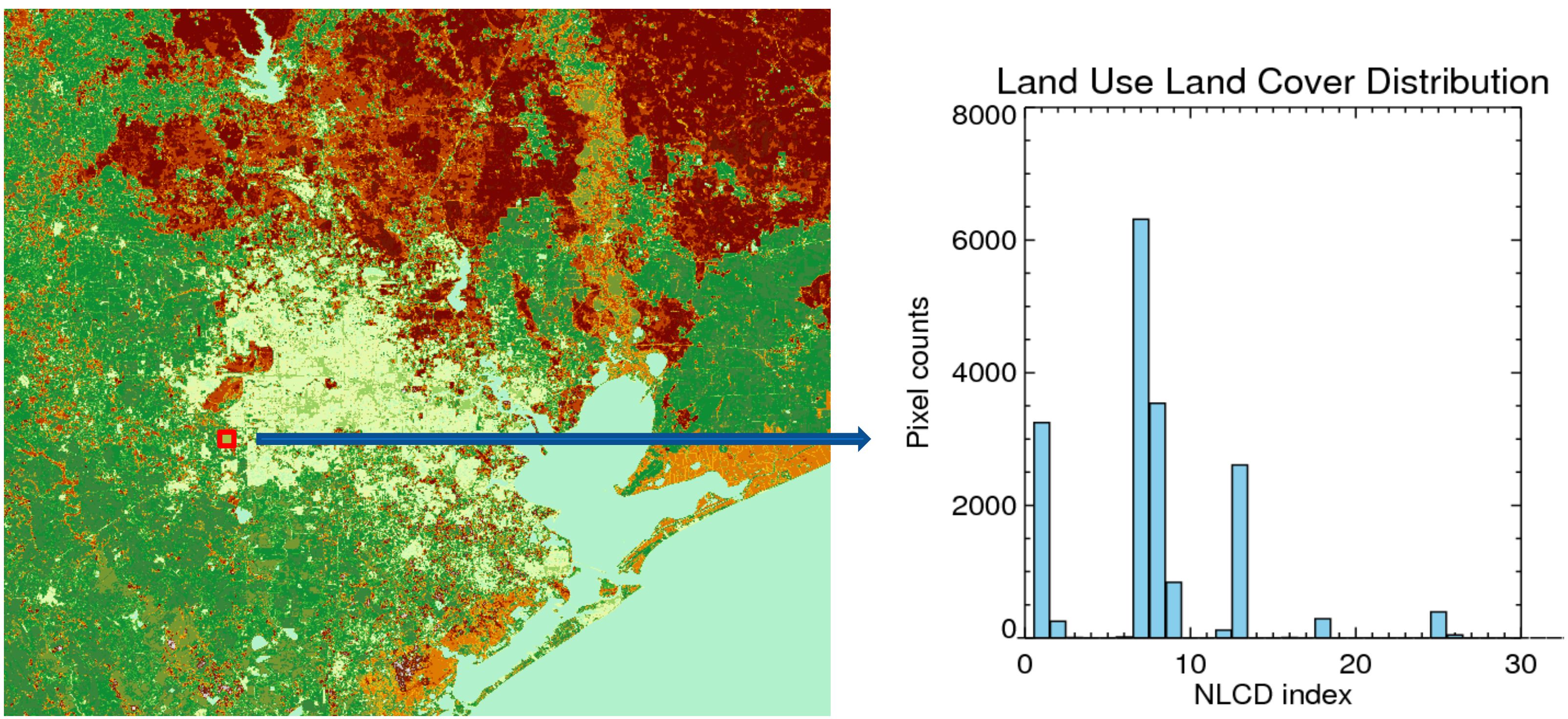
- Spatial allocator requires huge computational power to calculate fractional weightings between GIS polygons and/or polylines and gridded cells, so an efficient polygon/polyline cutting algorithm is important. Some GIS data have more than millions of entities.
- A key for faster spatial allocator is to optimize computational iterations in both polygon clipping and map projection calculations.



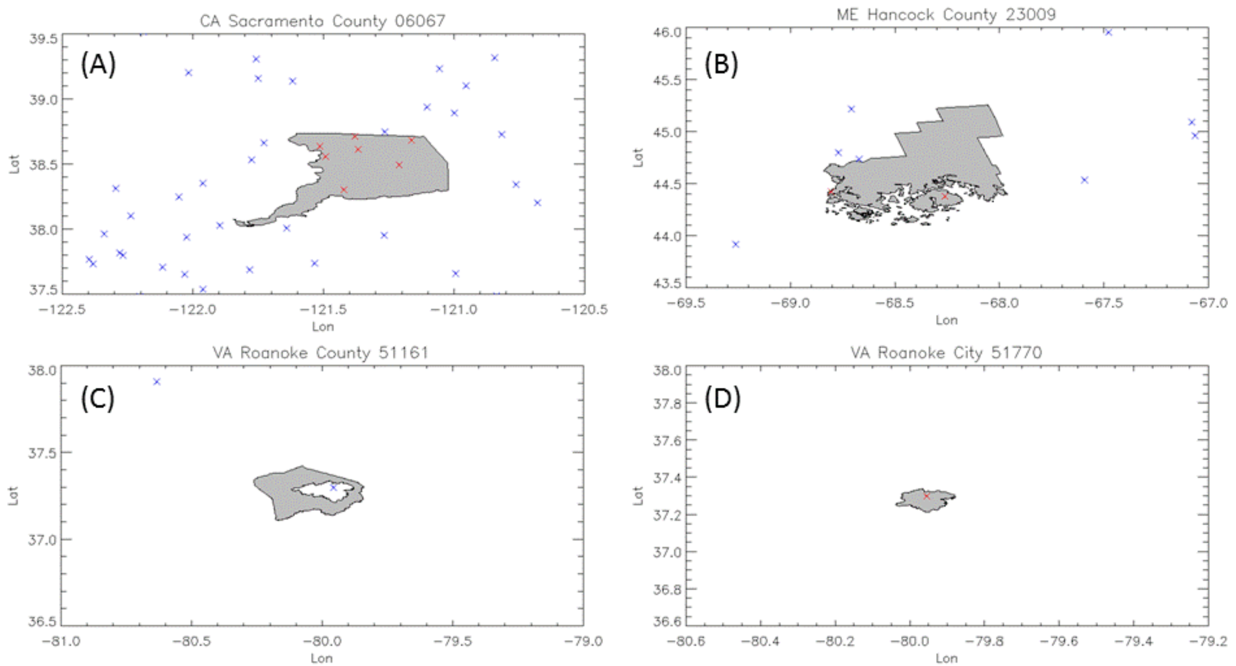
GIS data for population (census track) in Houston, with 4km grid cells (red line)

Raster data processor

The raster tool uses a histogram reverse-indexing method in IDL histogram function, and is capable of faster access of grouped pixels. For each grid cells, the raster tool provides histogram and statistics of pixels inside. Figure shows an example of 30-m NLCD LULC data near Houston region.



Raster data in irregular-shaped boundary

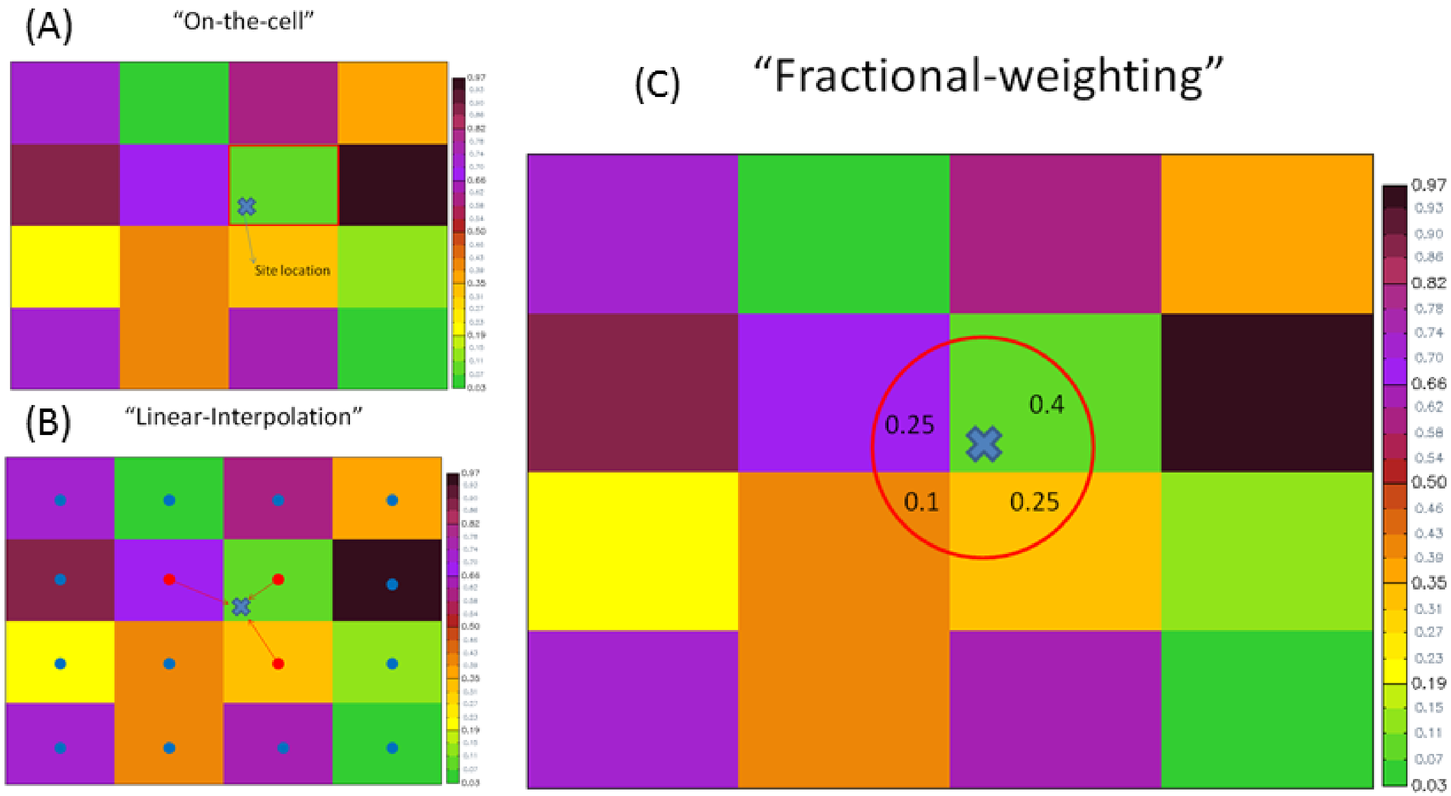


Examples of “inside polygon” determination. Shaded areas are county/FIPS boundaries for (a) Sacramento County, CA (06067), (b) Hancock county, ME (23009), (c) Roanoke County, VA (51161), and (d) Roanoke City, VA (21770). Blues crosses indicate AIRNow site locations outside given county/FIPS boundary and red crosses indicate inside county/FIPS boundary.

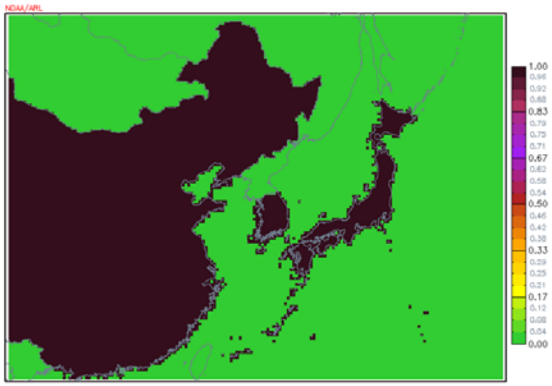
Algorithm optimization

- Procedures are straightforward, but need numerous repeated calculations for polygon clipping, map projection conversion, and data pixel indexing.
- Optimization of calculation algorithm and efficient memory usage are two key points in building efficient geospatial data processing tool

Application (1) fractional weighting

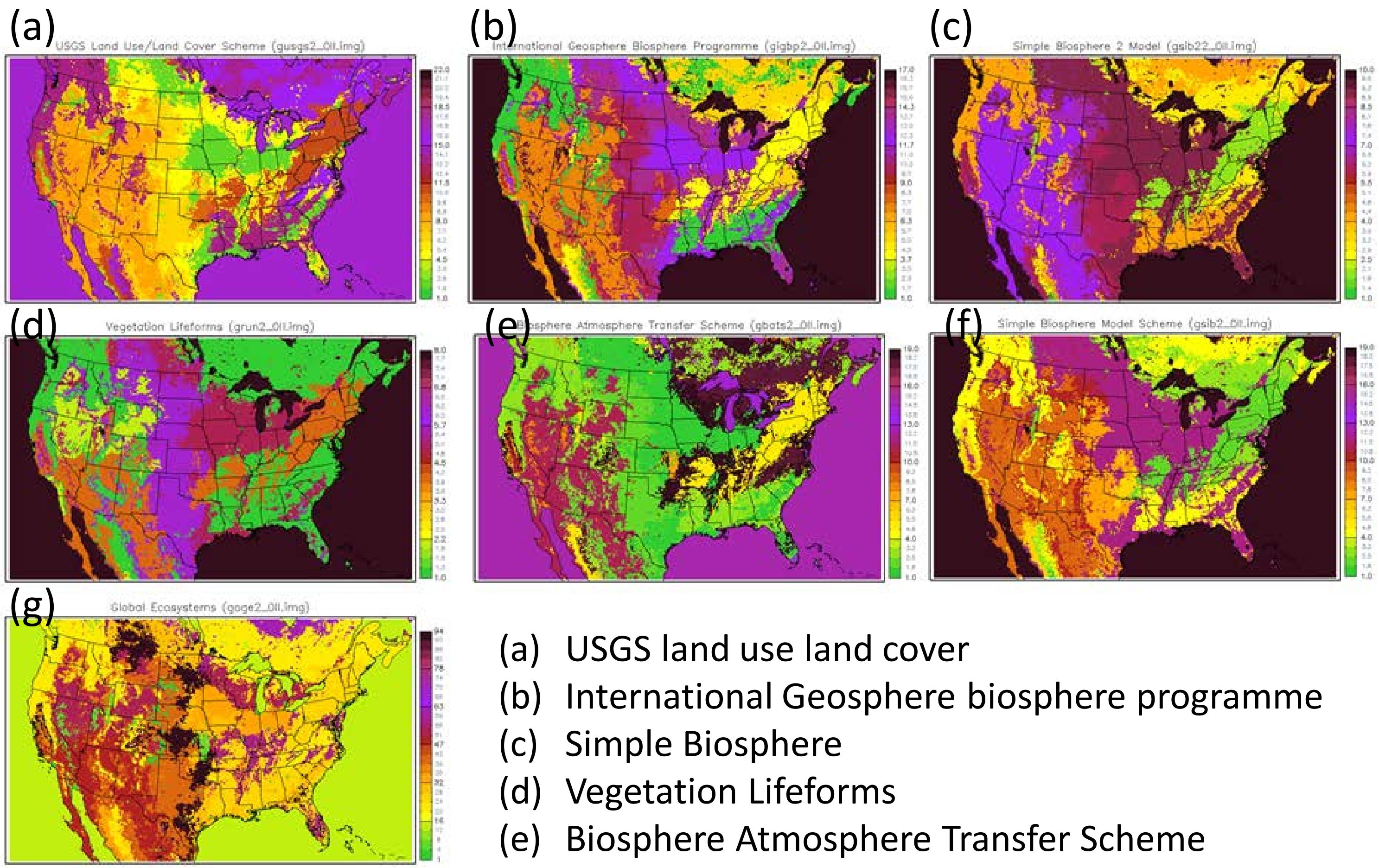


Application (2) model cell masking



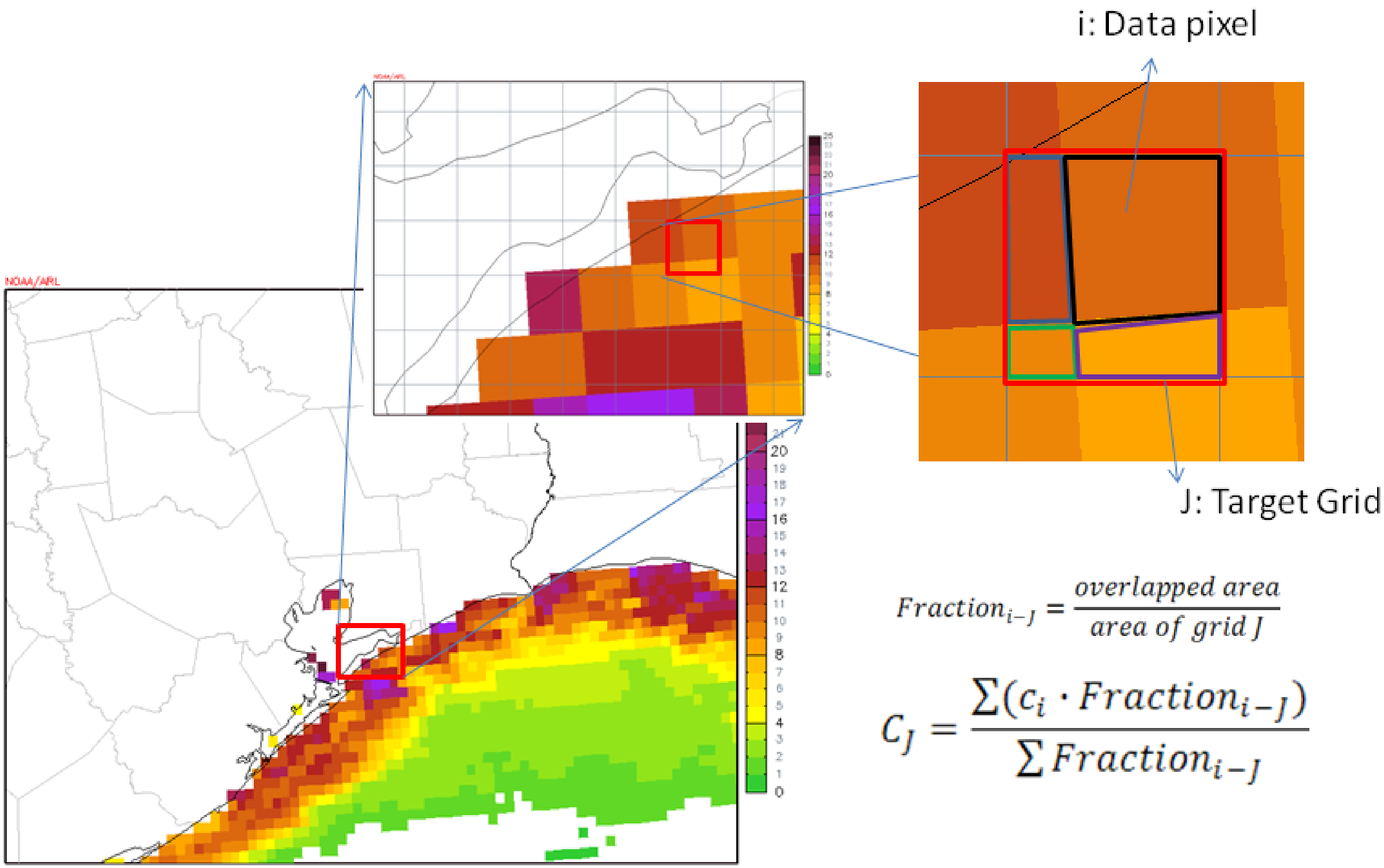
Examples of grid masking routine. Level 1 masking (state level) for Texas and Maryland (left), and level 0 masking (country level) for China, S. Korea, and Japan

Application (3) LULC data processing



(a) USGS land use land cover
(b) International Geosphere biosphere programme
(c) Simple Biosphere
(d) Vegetation Lifeforms
(e) Biosphere Atmosphere Transfer Scheme
(f) Simple Biosphere Model Scheme
(g) Global Ecosystems

Application (4) Conservative remapping



$$Fraction_{i-J} = \frac{overlapped\ area}{area\ of\ grid\ J}$$

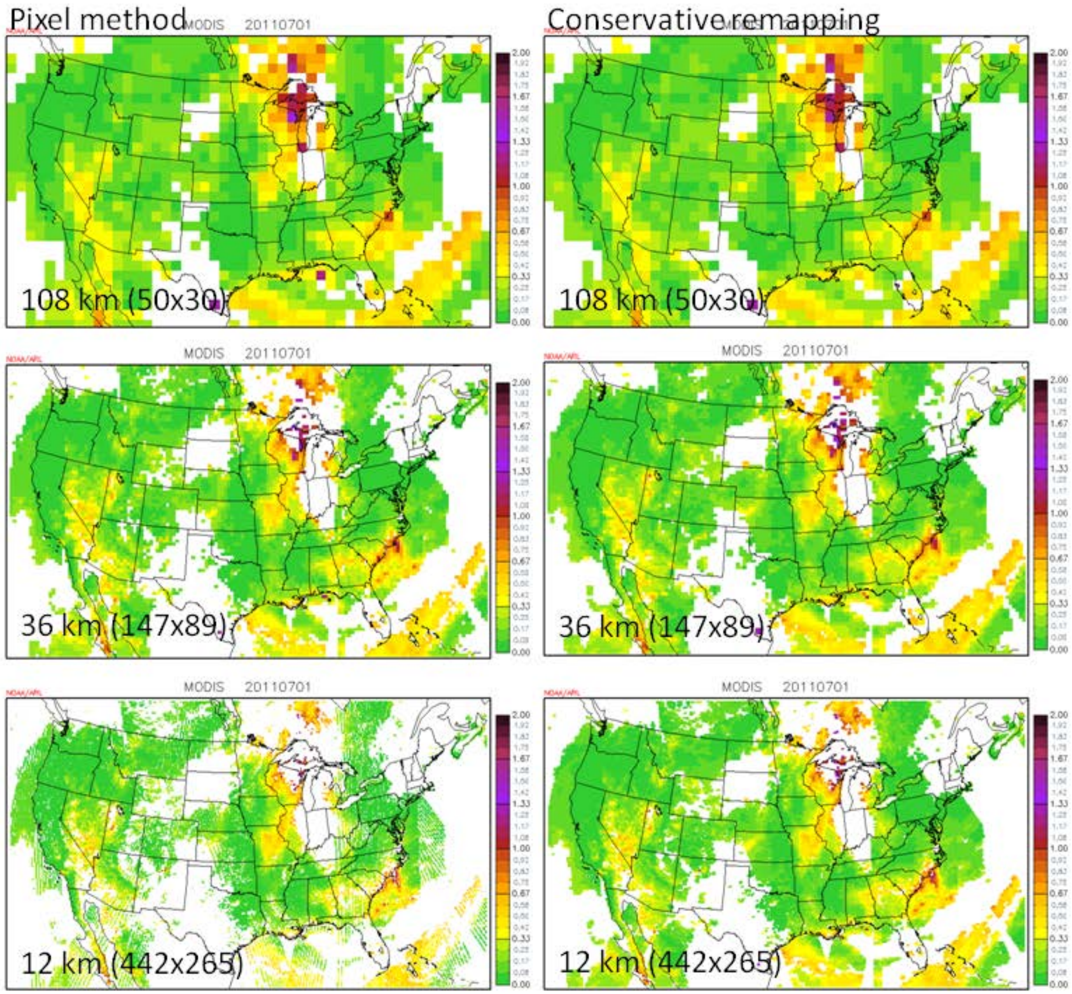
$$C_J = \frac{\sum(c_i \cdot Fraction_{i-J})}{\sum Fraction_{i-J}}$$

Example: Chlorophyll a concentration regridding in Houston

Application (5) MODIS AOD regridding

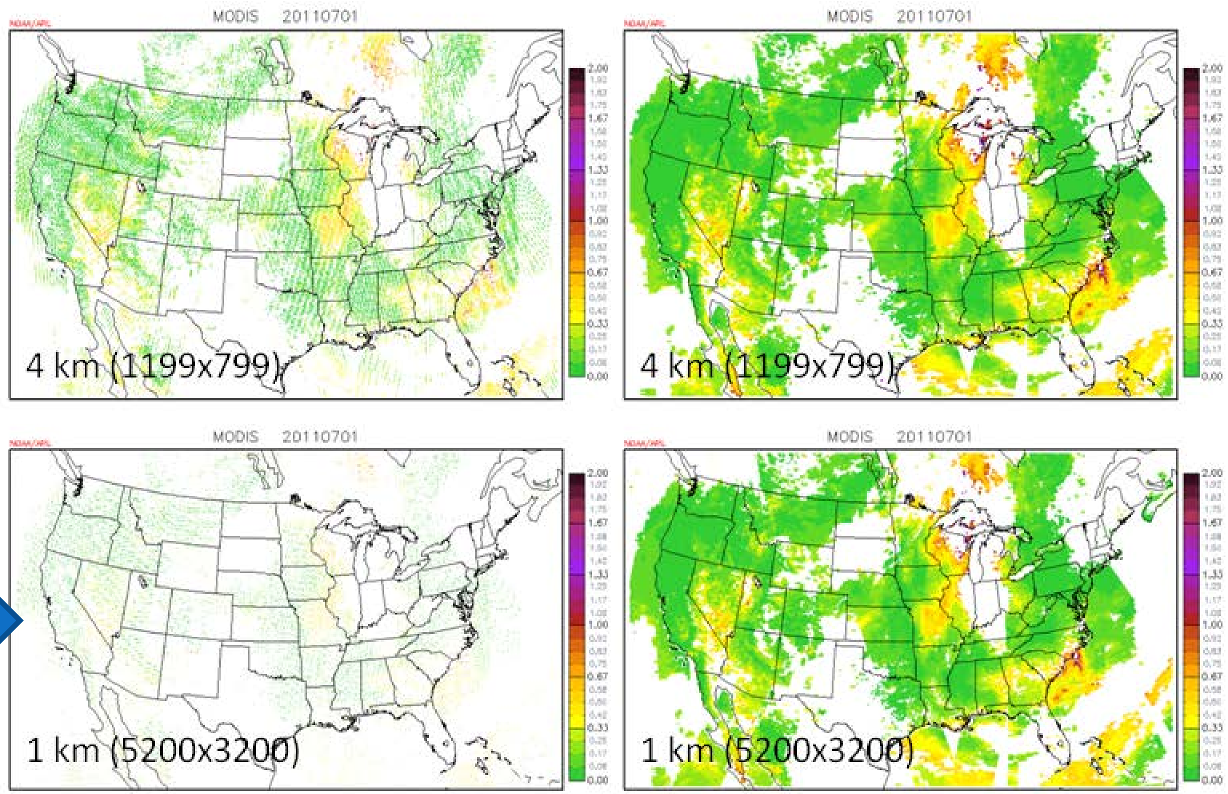
Pixel aggregation method

Conservative remapping method



AOD regridding - continued

Pixel aggregation method



Conservative remapping method

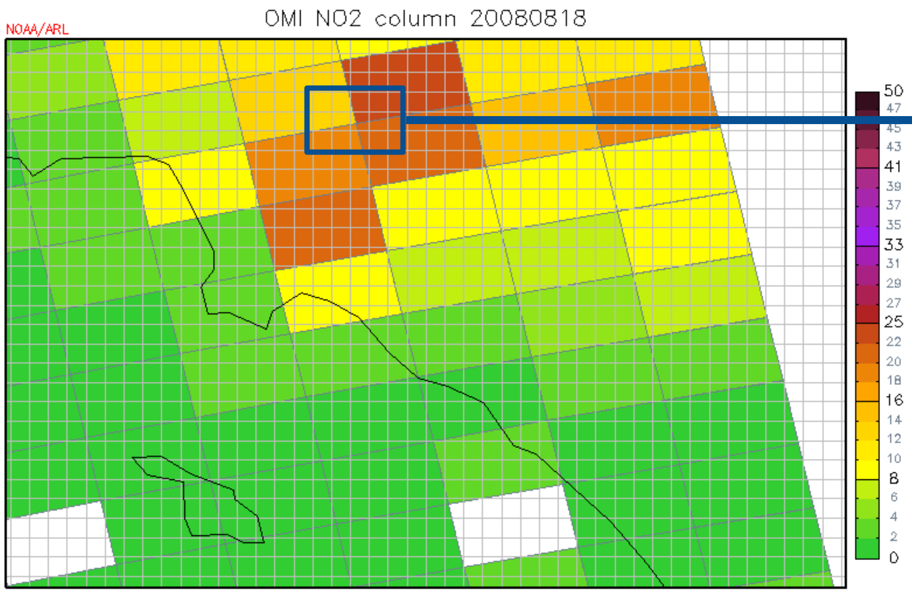
- Pixel aggregation method does not work for fine resolution grid cells.
- Conservative remapping method works for any resolution

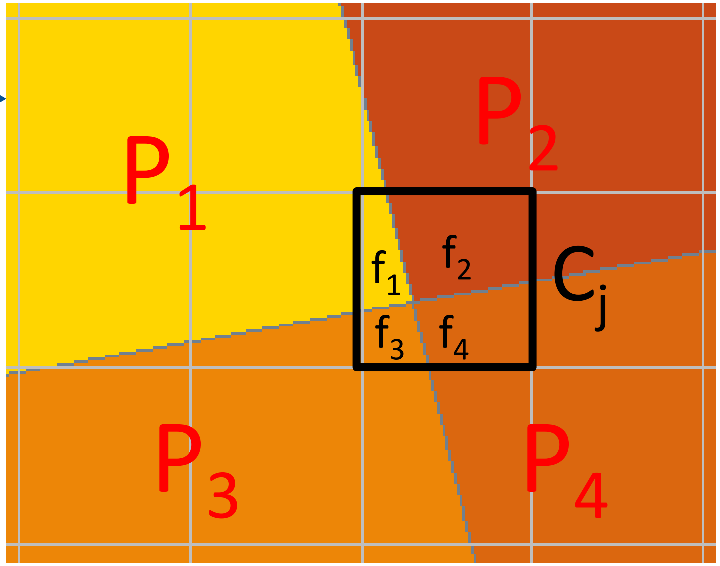
Advanced application (1)
NO_2 column observations from space

Instrument	Satellite	Local pass	Resolution	Data
GOME	ERS	10:30 LT	320x40 km	1996-2003
SCIAMACHY	ENVISAT	10:00 LT	30x60 km	2002-2012
OMI	Aura	13:30 LT	13x24 km	2004-
GOME-2	MetOp-A MetOp-B	09:30 LT	40x80 km	2007- 2013-

Question: How to compare different satellite measurements with various data resolution, and fine-scale regional air quality modeling

Step 1: Conservative remapping





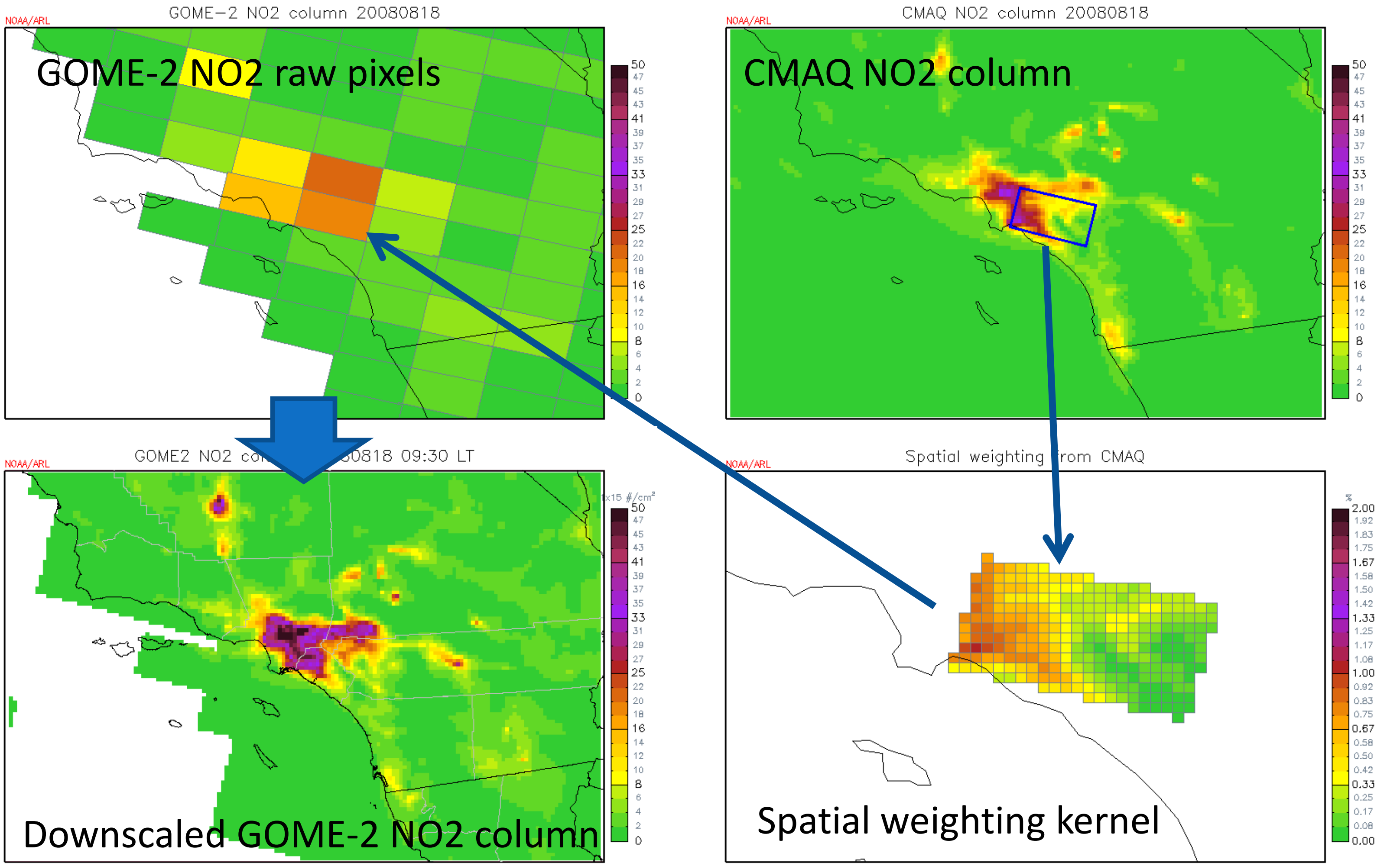
$$f_{i,j} = \frac{Area(P_i \cap C_j)}{Area(C_j)}$$

Cj = (P1·f1+P2·f2+P3·3+P4·f4)/(f1+f2+f3+f4)

$$C_j = \frac{\sum P_i \cdot f_{i,j}}{\sum f_{i,j}}$$

Overlapping fractions ($f_{i,j}$) are calculated using IDL-based Geospatial Data Processor (IGDP)

Step 2: Downscaling

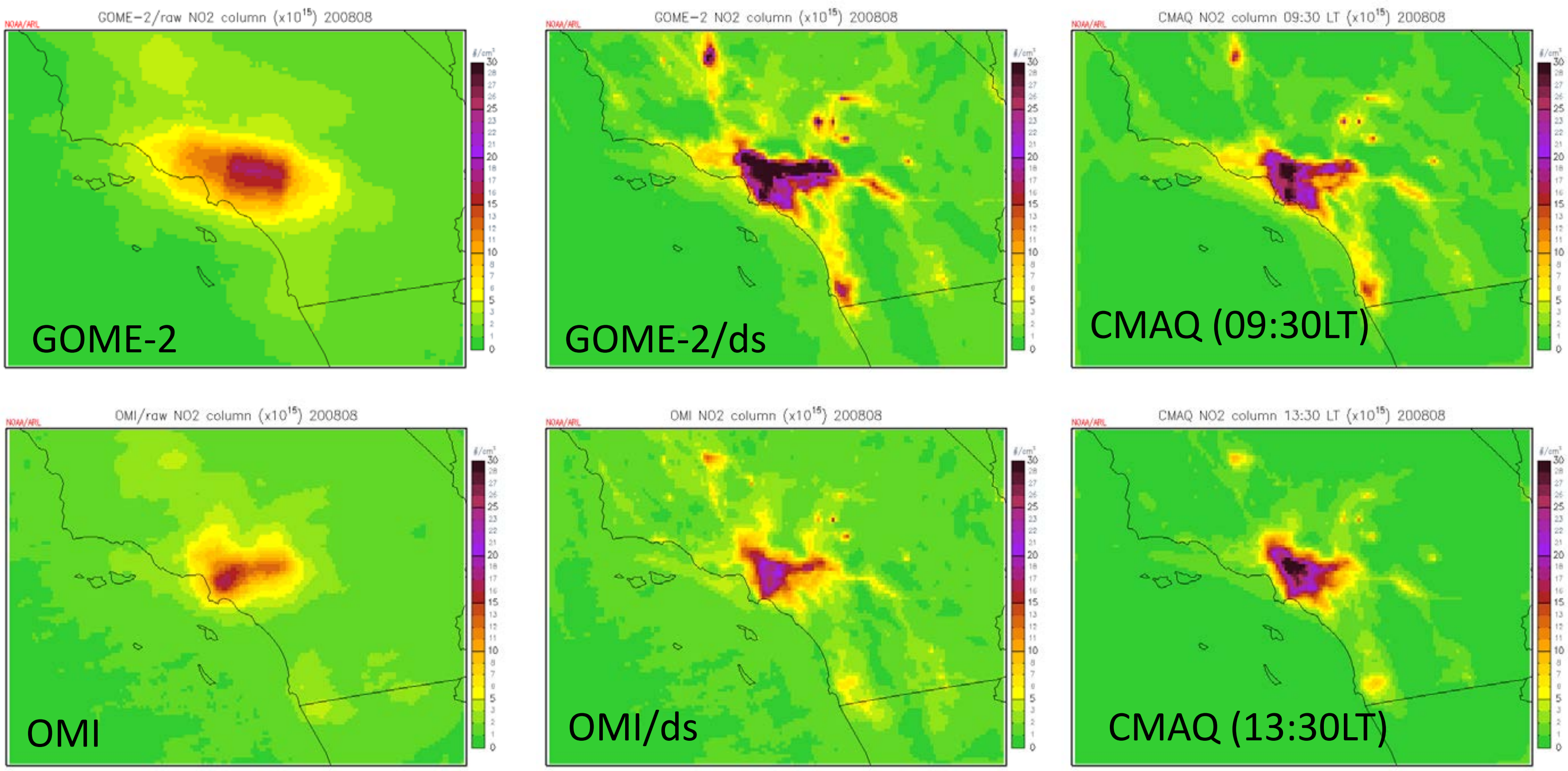


$$C_j = \frac{\sum P_i \cdot k_{i,j} \cdot f_{i,j}}{\sum f_{i,j}}$$

where $k_{i,j}$ is the spatial weighting kernel of P_i in the position of C_j

Each spatial weighting kernel is satellite pixel-specific, so the quantity in the original pixel never propagates out of original data pixel's coverage

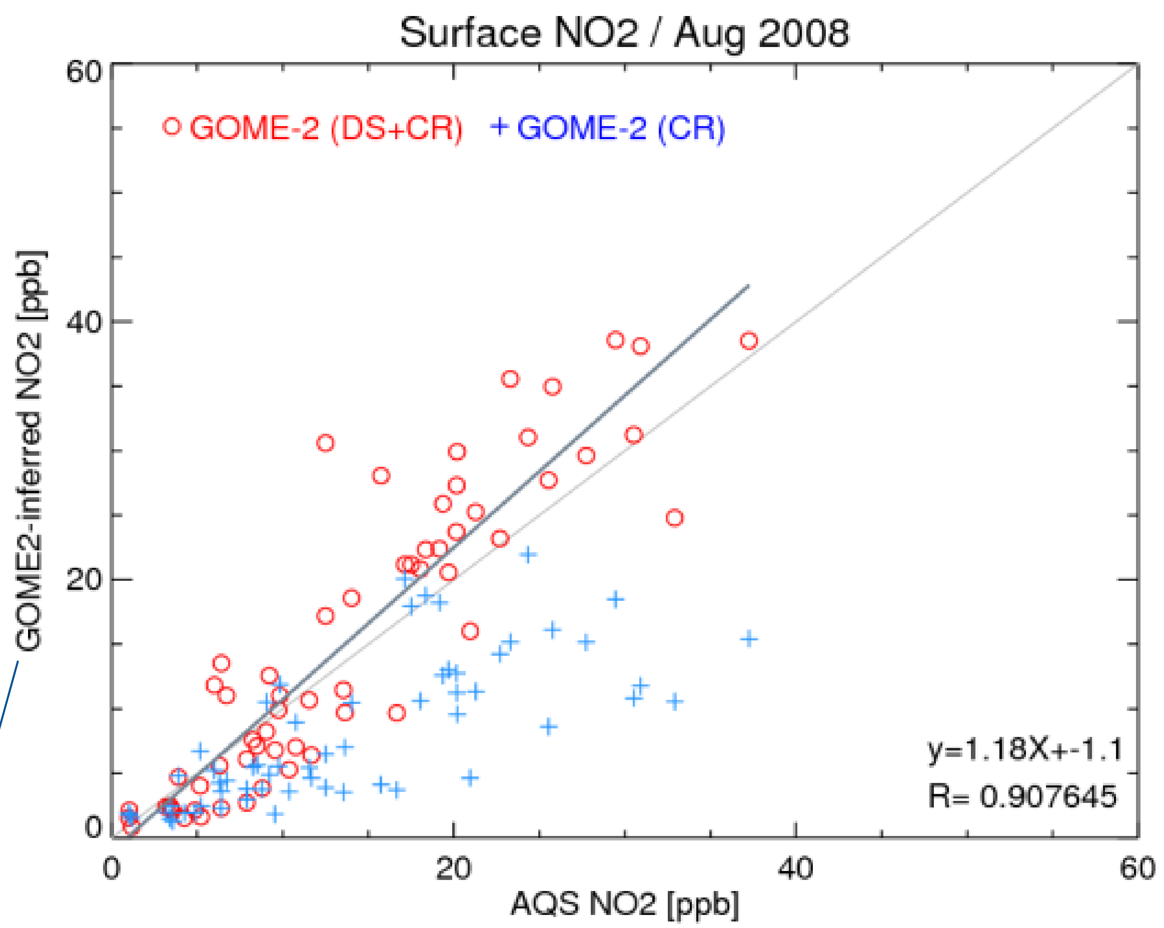
Monthly averaged NO_2 column
S. California Aug. 2008



How downscaling method works

- Satellite measurement is strictly preserved within each footprint. Strength of model does not affect satellite measurement's strength at all.
- The only information that passed from model to satellite is relative spatial information within each satellite footprint.
- Eventually, this method converts systematic negative bias due to coarse resolution to random errors, which can be cancelled out by temporal averaging.

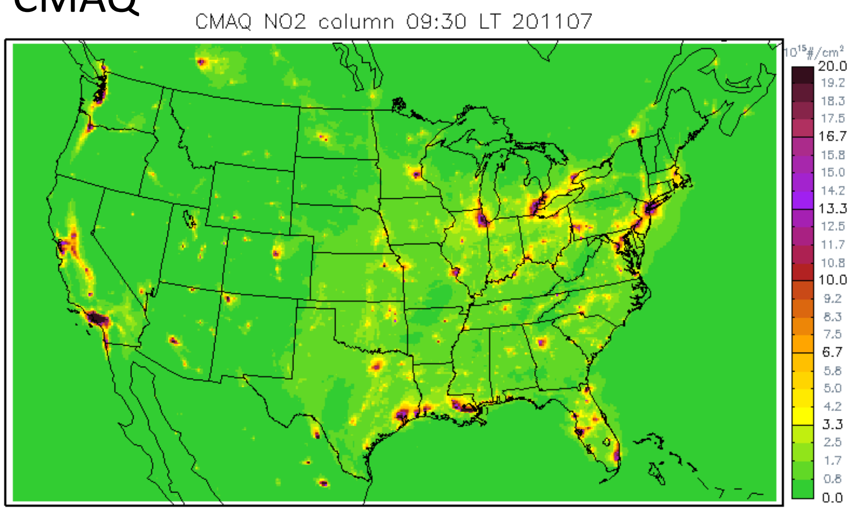
Surface NO2 concentration comparison



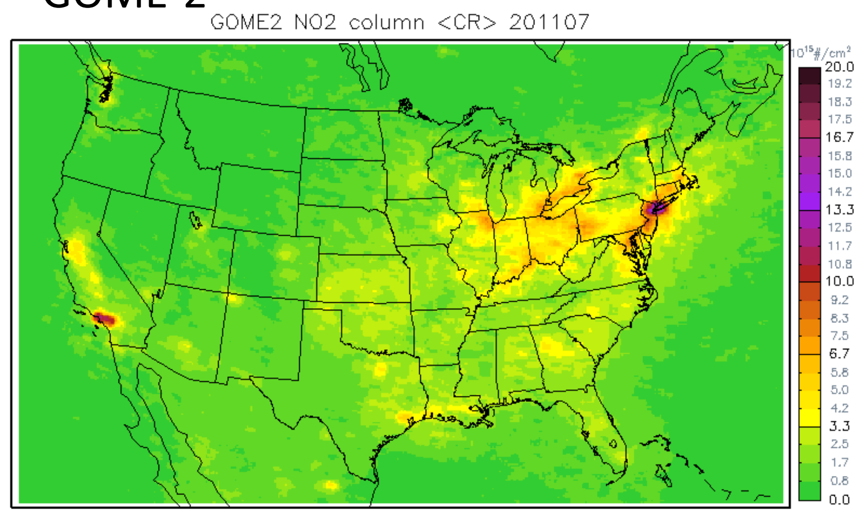
$$\frac{GOME2\ column}{CMAQ\ column} \times CMAQ\ surface\ conc.$$

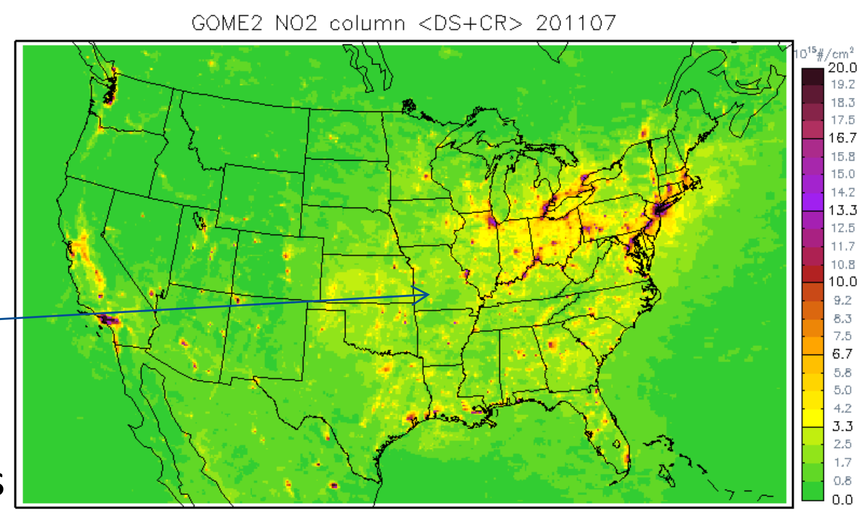
Monthly mean NO2 columns (Jul 2011) CMAQ, GOME-2, Adjusted GOME-2

CMAQ



GOME-2



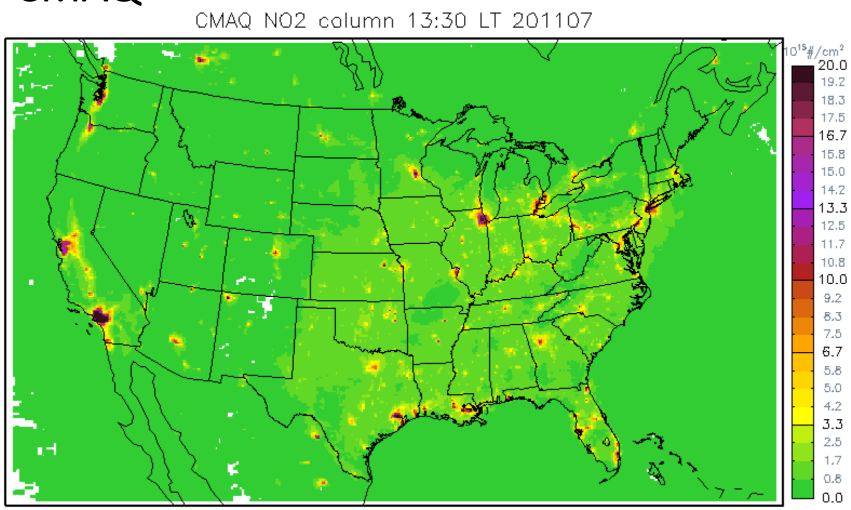


Small-scale features are well reconstructed by downscaling method

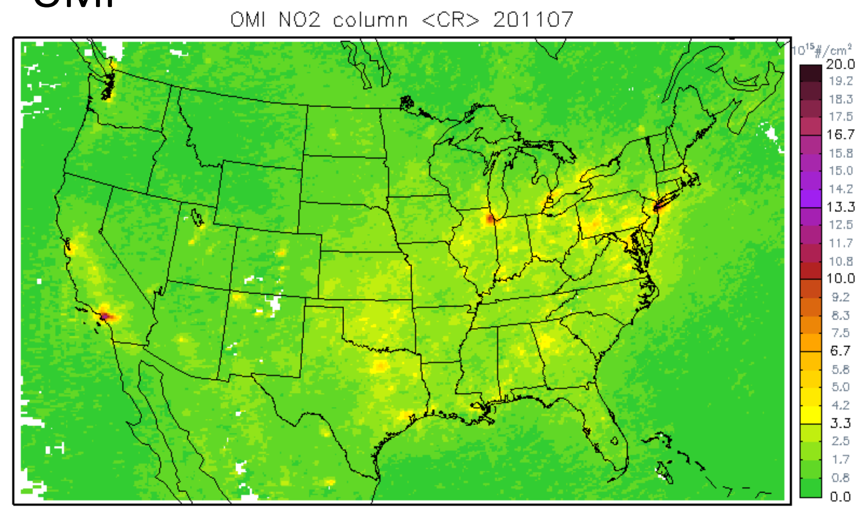
GOME-2/ds

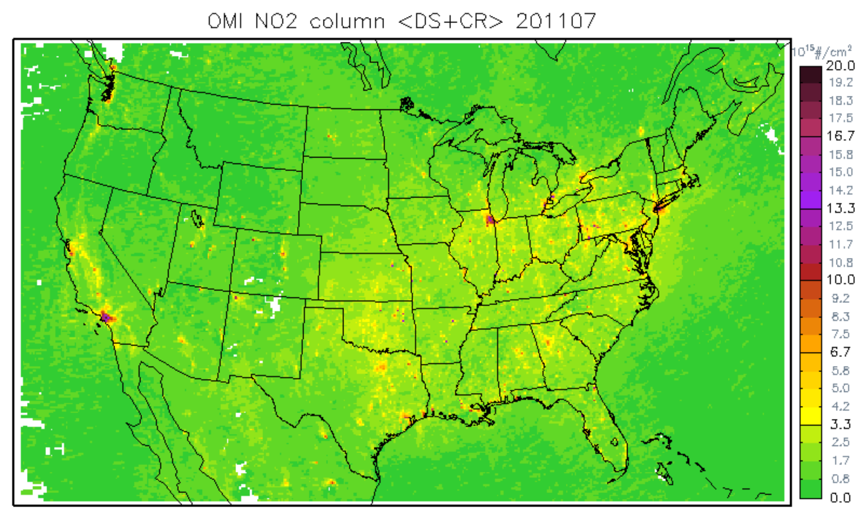
Monthly mean NO2 columns (Jul 2011)
CMAQ, OMI, Adjusted OMI

CMAQ



OMI

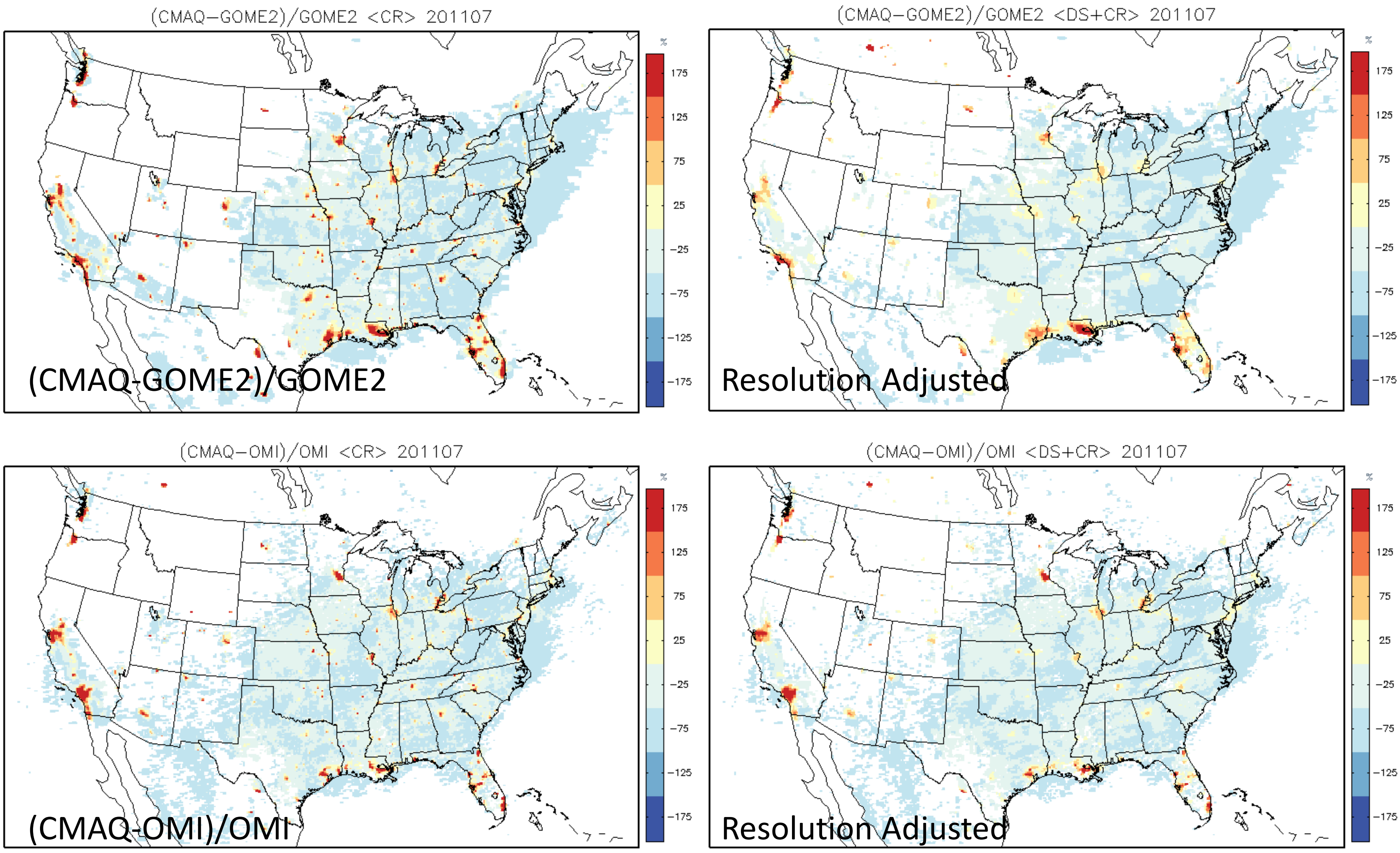




Since OMI has better resolution than GOME-2, the adjustment effect in small that GOME-2 case, but it still shows clear improvement in urban cities.

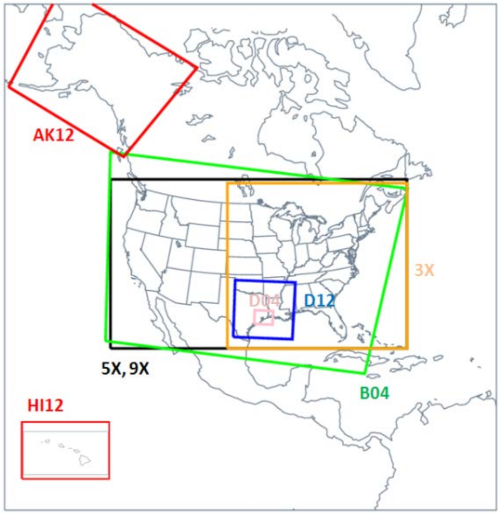
OMI/ds

Difference (CMAQ-Sat)/Sat [%]



Advanced application (2) emission surrogates

- We have tested new spatial allocator for various domains settings
- The IGDP tools support both Lambert Conformal Conic (LCC) and Rotated Lat/Lon (RLL) map projection



Name	5X	3X	D12	D04	HI12	AK12	B04	7X
Resolution	442X265 (12 km)	268X259 (12 km)	89X89 (12 km)	83X65 (4km)	80X52 (12 km)	199X163 (12 km)	1114X880 (4 km)	1199X799 (4 km)
Projection	LCC	LCC	LCC	LCC	LCC	LCC	RLL	LCC
# of surrogates	130	130	130	57	62	104	130	130
Processing Time	**13.4 hr**	**9.8 hr**	**5.5 hr**	**3.5 hr**	**1.2 hr**	**3.2 hr**	**25.8 hr**	**19 hr**

Result with a single CPU. 10~100 times faster than current Spatial Allocator

IGDP user guide

- Project final report will include a user guide with IDL library descriptions, and a sample package for NO2 column processing.
- The sample package will show fully automatized procedures to generate model NO2 column and satellite NO2 column (traditional and downscaled) for given CMAQ output file names.

Summary

- IGDP is designed to perform accurate and fast geospatial data processing to support meteorology and air quality modeling.
- Conservative remapping method provides seamless regridding of any data regardless of target domain resolution.
- Conservative downscaling method can implement additional information (e.g. fine-scale spatial distribution) while conducting conservative regridding.
- Accurate and efficient processing of GIS data is crucial for fine-scale modeling.

Weakness in current modeling system

- Geospatial data inputs for modeling have various sources (usually from GIS data) and time of production. There is no unified framework to assure consistency in data processing.
- As more fine resolution data are available, current data processing tools are too slow, or cause memory problem.
- Map projection errors are not fully considered, which could be critical in fine resolution modeling.