Web Services for Comparative Data Analysis of Emissions Inventories

Stefan Falke
Washington University in St. Louis, Campus Box 1180, One Brookings Drive, St. Louis, MO 63130
stefan@wustl.edu

Gregory Stella
Alpine Geophysics, LLC, 387 Pollard Mine Road, Burnsville, NC 28714
gms@alpinegeophysics.com

Terry Keating
U.S. EPA, Office of Air & Radiation, 1200 Pennsylvania Ave. NW, Mail Code 6103A,
Washington, DC 20460
keating.terry@epa.gov

ABSTRACT

This paper describes the development and application of web services for comparing US and global emission inventories. Web services and service oriented architectures are providing new ways to support emissions modelers and analysts as they provide a web browser based environment for accessing and analyzing emissions data and inventories. Traditional emissions data analyses that can be conducted using web applications reduce the effort associated with accessing data, formatting data, and running analyses. The web application presented here supports an “offline” analysis conducted by the authors for reconciling and comparing regional and global emissions inventories. The focus of the analysis was the reconciliation of global inventories developed using “top down” methods with regional inventories developed using localized “bottom up” methods for a domain covering the continental United States. Part of that analysis is reconstructed using web services, including the spatial allocation of point source US EPA NEI data to a 1x1 degree grid and calculation of differences and ratios between the gridded NEI and EDGAR inventories. The online application is made possible through dynamic internet access to emissions inventory data and web standards for connecting data access services with data analysis and map visualization services. The same analysis services can be quickly re-applied when emissions data are updated or applied to new comparisons with other inventories. The analysis web services are openly available online (www.neisgei.org) for other emissions or non-emissions applications.

INTRODUCTION

A variety of approaches are presently employed in deriving emissions inventories. Being able to understand the differences and similarities between inventories can help in integrating them for purposes of creating more comprehensive inventories or for highlighting their respective strengths and weaknesses. Currently, comparisons of emissions inventories require substantial manual effort in acquiring the data, reformatting them, importing them into desktop applications, and running analyses. The complexities of the end-to-end analysis process limits who can analyze the data and how readily those analyses can be conducted. The next generation of web services provides an opportunity to help make this process more efficient and available to the wider community. The multiple data sources, data types, and spatial and temporal resolutions of the data used in air emissions analyses offers an attractive test environment for web service applications and exemplifies a broader class of applications that could be addressed using web technologies.

Advances in information technology are promising to help achieve the next generation of emission inventory systems. A NARSTO Emissions report highlights new database management approaches and information systems in envisioning a future inventory that “includes all significant emissions from all sources, time periods and areas, with quantified uncertainties, and timely accessibility. From this vision, the overall goal is to make inventories complete, accurate, timely, transparent, and affordable."[1] The hope is that new information technologies can make multi-spatial, temporal and composition scale air emissions data and tools easier to find, use and integrate.

The work presented here is part of the Networked Environmental Information System for Global Emissions Inventories project (NEISGEI, pronounced "nice-guy,") is an EPA-supported initiative to develop information technology components needed for a global air emissions inventory network. Part of this effort includes the development and sharing of web services that allow user-driven data processing and analysis of emissions data. To this end, we created a web application within a service oriented framework to support the type of emission reconciliation analysis described in a previous study by Gregory Stella at Alpine Geophysics, Inc.[2] where the purpose of the analysis was to compare emission estimates contained among various regional and international emission inventories.

The Stella analysis compared the U.S. National Emissions Inventory (NEI) to the Emissions Database for Global Atmospheric Research (EDGAR), which has been used in many global atmospheric modeling studies examining transboundary air pollution. EDGAR was developed by a consortium of institutes in the Netherlands using a uniform methodology for the entire globe and is reported on a 1 x 1 degree grid. As a result, EDGAR data is not consistent necessarily with more spatially resolved and official emissions estimates generated at a national level, such as the U.S. National Emissions Inventory (NEI). To better understand the intercontinental flow of air pollutants, it is important to compare, evaluate, and where possible resolve the spatial and temporal differences between these emissions inventories developed at different scales and the Stella analysis did so for NEI and EDGAR. The project presented here applied web service technologies to create web applications that support aspects of the Stella analysis in spatially comparing different emissions inventories. Specifically, we create data access services for serving point source or county level emissions (e.g. NEI and similar inventories for Mexico and Canada) and gridded emissions inventories (e.g., EDGAR and another globally gridded inventory, RETRO) through a standard web interface. We developed analysis services to spatially allocate point source or county level emissions to a grid for quantitative comparison with gridded emissions inventories.

The development of the web services and the emissions analysis application relied on existing components to the greatest extent possible and only developed new components to fill gaps. This approach to development is a key principle of cyberinfrastructure and web 2.0, two trends in the next generation of web applications. In our development we use Open Geospatial Consortium standards, the DataFed air quality web service infrastructure and openly available mapping tools. These technologies and the method we used in supplementing them and integrating them into an emissions comparison application are described in this paper. The section titled, "Web Service Technologies" outlines the underlying methods and technologies used in developing the emissions comparison web application. The section, "Service Oriented Emissions Analysis," describes the datasets, data access web services, and analysis web services used in building the web application. The "Comparative Emissions Analysis Application" section explains how the various services and other technologies were combined to create the application and how the user interacts with the application to conduct emissions data analyses.

WEB SERVICE TECHNOLOGIES

Many of today's cutting edge web applications are generated by linking together multiple existing components rather than building every piece from scratch. The novel development is created the pieces needed to get disparate pieces to communicate and work together in an application framework. This section describes some of the technologies and development principles used in creating the emissions comparison application, including cyberinfrastructure components, Open Geospatial Consortium standards, and DataFed web services.

Web 2.0 and Cyberinfrastructure

The development of new information technologies and network technologies has changed the way people interact through the web. Web 2.0 is a term commonly used to describe the new web where information is not simply served for consumption through websites but is interactively created as user-driven content. Within the science and engineering domains, this next phase of the web is referred to as cyberinfrastructure[3,4], e-science[5], and service oriented science[6]. These terms are used to represent new computing environments of hardware, software and web services that create new capabilities for sharing information, conducting research in a distributed environment, and achieving new insights that would have taken longer, or not occurred at all, in the old system of stand-alone research labs and centers. Cyberinfrastructure projects have demonstrated inroads in the new approach for accessing and visualizing environmental data.[7,8,9]

The web services approach uses standards-based interfaces for connecting data providers with data users. The network strives to go beyond only searching and visualizing data to include data processing and analysis services to allow users to create new content. These network users can function on an independent level, each addressing local issues of importance. These individual components can be integrated or modified to handle differing data types dynamically, on demand. Web service technology is still evolving and does not currently provide a complete out-of-the-box software solution. However, many required components are considered standards in web programming applications and therefore make it possible to create an operational data web service network and to develop web applications upon that network.

One of the more popular aspects of web 2.0 is "mashups." Mashups refer to applications constructed by combining services from disparate sources. Over half of all mashups use some form of a map interface, and of those, by far the most prevalent is Google Maps.[10] We explored the use of the Google Maps API in building our web applications. Google Maps is an attractive interface for working with point data. Image, or gridded data, worked fairly well, as long as that was not dynamically created, such as in the grid spatial allocation of emissions data. Google Maps uses its own projection rather than a common projection which leads to difficulties in overlaying imagery and grids. We ultimately restricted our use of Google Maps to handle point-only applications and used other map tools for our gridded analysis application as described later in this paper. In creating our comparative emissions analysis application (or mashup), standards and services for accessing, processing, and visualizing data were combined. Javascript was used as the "glue" to hold the components together to form the web application.

OGC Specifications

Standards for finding, accessing, displaying, and processing geospatial information, from map images to numeric datasets, are defined by the Open Geospatial Consortium (OGC)[11]. The Web Map Service (WMS) [12] provides an interface for exchanging map images while the Web Feature Service (WFS) [13], Web Coverage Service (WCS) [14], and Sensor Observation Service (SOS)[15] define interfaces for exchanging numeric data. The WFS is designed for traditional Geographic Information System (GIS) features, such as buildings and state borders. A WCS service allows access to multi-dimensional data that represent continuous phenomena coverages, such as satellite imagery grid coverages or weather monitoring network point coverages. The SOS defines interfaces for describing and access data from a broad range of sensors.

We focused on implementations of the WCS for both gridded emissions data and point/area emissions. In the future, we will explore the use of the WFS because point facilities and counties can be considered to be feature types with attributes of emissions. We use the WMS for creating the map images for display in the emissions comparison tool.

Three types of requests are generally made to a WCS server, 1) *GetCapabilities* for retrieving information on what the server offers, including a list of available coverages 2) *DescribeCoverage* for getting information about a particular coverage and 3) *GetCoverage* for retrieving data contained within the coverage. In the emissions comparison application developed here, we only use the GetCoverage request as we have *a priori* linked the specific coverages to the application. However, the other requests were implemented in the NEISGEI WCS and are available for other applications to use.

An example of a WCS data request is shown in Figure 1. The URL for the request is a single text string but is broken into multiple rows in the figure to make its elements easier to read. The URL request points to the web server providing the data service and specifies that it wants to make a GetCoverage request for the EDGAR NO_x data for the year 2000 over a geographic area that corresponds to the contiguous US. The GetCoverage request will return the EDGAR NOx emissions grid for the specified space and time constraints in a netCDF format, a common data format for multi-dimensional atmospheric science data.[16]

Figure 1. Example URL for an HTTP Get request for the OGC Web Coverage Service.

```
http://niceguy2.wustl.edu/wcs/NEISGEI.wcs?
SERVICE=WCS
&REQUEST=GetCoverage
&VERSION=1.1.0
&CRS=EPSG:4326
&COVERAGE=Edgar.NOX
&BBOX=-126,24,-66,50
&TIME=2000-01-01
&FORMAT=NetCDF
```

DataFed

The federated data system, DataFed (http://datafed.net), is a web infrastructure that provides the foundation for accessing distributed air quality data and for processing and visualizing these data through web services.[17] The key role of DataFed is to mediate the flow of data between data providers and users. DataFed provides mediator software for creating "views" of data, including maps, time series, and tables, that are distributed among multiple web servers. The views are created using web services thereby allowing them to be used and reused in custom applications with standard web programming languages.

DataFed provides a web service chaining environment in which "views files," XML-based files that specifies the data layers that comprise a view along with the services associated with creating that data layer (e.g., data access services, analysis services, and display services). [18] The order in which the layers are stacked in the view file determines the order in which the services are executed in a service flow chain. The ability to describe a chain of services allows for the creation of flexible, multi-data layer interactions with analysis and display services. Outputs from one layer can be used as input for another layer's services, the output from that layer can be used "downstream" with other services, and so forth. Each service's setting can be controlled through a set of parameters, thereby creating a dynamic service flow for web applications.

We use DataFed's analysis services framework and service flow engine to incorporate the grid operator service, create a new point-to-grid conversion service, and connect our data access services with the analysis services. The result is a DataFed view that can be controlled externally through controls in our emissions analysis application. The services used are described in the following section and the web application in which they are integrated is described in the final section of this paper.

SERVICE ORIENTED EMISSIONS ANALYSIS

The availability of service oriented architectures, the OGC specifications, and the DataFed framework allow the development of an emissions comparison web application. The application uses a combination of data access and data analysis services and a web browser user interface to interact with emissions data and calculate comparisons among emissions inventories.

Data Services

Most emission databases are already accessible through Internet-based methods either through direct data-file download or web query tools. The query systems allow users to filter and access data at multiple levels of detail. These systems meet the needs of individual end users who log in to the online system, complete forms for defining their query, and then view the results in tables/graphics or download the data for use in other tools. While these systems serve the individual user, they do not easily come together to form a distributed emission inventory network where automated computer-to-computer interaction among services is possible for supporting dynamic web applications. This computer level interaction is possible through web standards and services, such as the Open Geospatial Consortium (OGC) specifications. In order to support data access to the emissions comparison application, we downloaded a variety of point, county and gridded emissions datasets, stored them on NEISGIE servers and created OGC-based services for dynamically accessing the data through URL requests as described in the previous section.

<u>USNEI</u>

The EPA's National Emissions Inventory (NEI) includes estimates of annual emissions of air pollutants, including volatile organic compounds (VOCs), carbon monoxide (CO), nitrogen oxides (NOx), sulfur dioxide (SO2), particulate matter (PM10 and PM2.5), and ammonia (NH3).[19] Emission estimates for individual point sources and county level estimates for area and mobile sources, are available for years 1990, 1996, 1999 and 2002.

The NEI data files are available for download from the EPA Technology Transfer Network Clearinghouse for Inventories & Emissions Factors as ASCII text files in NEI Input Format (NIF). We imported the tables into a Microsoft SQL database and modified the tables in order to facilitate their access through the OGC interfaces. A facility location code was created by combining the StateCountyFIPs, TribalCode, StateFacilityIdentifier, and EmissionReleasePointID fields. Most, but not all, emissions in the NEI files are in units of short tons. The emission records defined by units other than tons, were converted to short tons. A location table with the location code and latitude, longitude was created for joining with, and geolocating, the emissions values. For area and mobile source data, we used the county centroid as the location latitude and longitude coordinates. Emissions returned from the data access query are summed based on the facility location code. In other words, emissions from multiple processes with the same EmissionReleasePointID are summed together to represent a single emissions source.

<u>Mexico NEI</u>

The Mexico National Emissions Inventory (MNEI) was created by a partnership between Mexico's Secretariat of the Environment and Natural Resources *(Secretaría de Medio Ambiente y Recursos Naturales—SEMARNAT)* and National Institute of Ecology *(Instituto Nacional de Ecología—INE)*, the U.S. EPA, Western Governors' Association (WGA), and the North American Commission for Environmental Cooperation (CEC).[20] MNEI provides emissions in NIF format for the year 1999. Pollutants include the nitrogen oxides (NOx), sulfur oxides (SOx), volatile organic compounds (VOC), carbon monoxide (CO), and PM10, PM2.5, and ammonia (NH_3). The Mexico NEI includes emissions for five source types: point, area, motor vehicle, nonroad mobile, and natural.

MNEI data are provided as downloaded Microsoft Access files for interior and border states. We merged the interior and border state Microsoft Access files and imported the resultant database to a Microsoft SQL Server. We then processed the data using the same approach as outlined for the US NEI data (creation of location code, unit conversion to tons where needed, use of municipality centroids, and creation of location tables).

NPRI

The Canadian National Pollutant Release Inventory (NPRI) contains annual emissions for a wide range of pollutants, including total particulate matter (TPM), particulate matter less than or equal to 10 microns (PM10), particulate matter less than or equal to 2.5 microns (PM2.5), sulfur oxides (SOx), nitrogen oxides (NOx), volatile organic compounds (VOC), carbon monoxide (CO) and ammonia (NH3). [21] NPRI is reported annually. Emissions since 2001 are used in the emissions comparison application because that was the first year criteria pollutants were included in NPRI.

The NPRI data are made available through Microsoft Access database files. We downloaded the individual files for each year from the NPRI website, merged the years 2001-2004 into a single Microsoft Access database, stored the file on the NEISGEI server, and provided an OGC Web Coverage Service interface to it.

EDGAR

The Emission Database for Global Atmospheric Research (EDGAR) project has been carried out jointly by the National Institute for Public Health (RIVM) and the Netherlands Organization for Applied Scientific Research (TNO).[22] It combines information on many different anthropogenic emission sources. EDGAR provides global annual emissions at a 1 degree X 1 degree resolution for CO2, CH4, N2O, CO, NOx, NMVOC, and SO2 for the years 1990, 1995, and 2000.

EDGAR emissions data are available for download through the Global Emissions Inventory Analysis (GEIA) website as ASCII text files. A single file contains gridded emissions for one pollutant for one year. The ASCII format included header information describing the content of the grid followed by rows or emissions values. We created a script written in IronPython to read the ASCII files and create a hypercube, a multidimensional binary data structure (we used the DataFed hypercube structure). The hypercube allows for more efficient web interface data access compared with the ASCII files because each data request queries a single, integrated data source rather than multiple ASCII files. The hypercube was stored on the NEISGEI server and made web accessible as an OGC Web Coverage Service (WCS).

RETRO

The REanalysis of the TROpospheric chemical composition over the past 40 years project (RETRO) created emissions gridded data sets.[23] Within the RETRO project, global gridded data sets for anthropogenic and vegetation fire emissions of several trace gases were generated, covering the period from 1960 to 2000 with a monthly time resolution. These data sets are made available through GEIA. The data are provided with spatial resolution of 0.5x0.5 degrees and in netCDF format. The RETRO emissions are based on the Netherlands Organization for Applied Scientific Research (TNO) emissions estimates. Source types include power generation, residential and, commercial combustion, industrial combustion, industrial processes, extraction distribution of fossil fuels, solvent use, road transport, other mobile sources, waste treatment and disposal, agriculture and landuse change.

Each RETRO netCDF file contains emissions for one pollutant for one year, at a monthly temporal resolution. We created an IronPython script to read the netCDF files and created hypercubes for the NO_x and CO. Each RETRO netCDF provides emissions flux information (kg/m2/s) at a 0.5X0.5 degree resolution grid and grid area information. We calculated emissions (in kg) by multiplying the emissions flux by the grid area by the number of seconds in the month. The hypercube was stored on the NEISGEI server and made web accessible as an OGC Web Coverage Service (WCS). A temporal aggregator web service was used to calculate annual emissions from the monthly values.

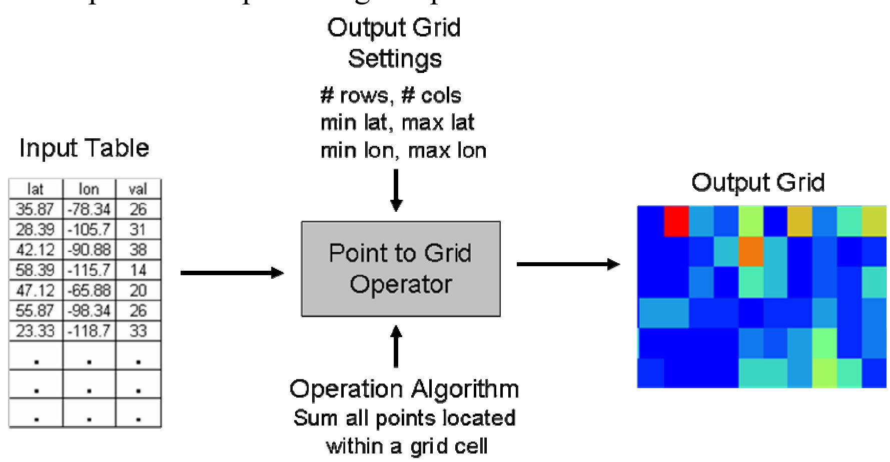
Analysis Services

Most cyberinfrastructure research to date has focused on the exchange and interoperability of data. But making data readily accessible and more easily shared is just one step. Services for using that data are a next step in adding value and increasing the capabilities available to data consumers. With the emissions data accessible through OGC service interfaces as described in the previous section, analysis services were created for processing the results of OGC data access requests and creating new datasets that could also be served through OGC interfaces. From our service oriented perspective, the interoperability requirements for analysis services is that their input and output interfaces adhere to OGC specifications, but the actual analysis algorithm can be implemented any way the service provider wishes. Two analysis services are used in this project: a point-to-grid creation service and a grid analysis service.

<u>Point-to-Grid</u>

The point-to-grid service developed here takes a set of latitude, longitude points and sums the associated values for all points that fall within each grid cell of an output grid (Figure 2). The resolution of the grid (rows and columns) is determined by the user. The service incrementally steps over the rows and cols of the output grid and conducts a mathematical operation to arrive at an output value for each grid cell. In our case, the emissions at points (facility points or county centroids) located within a grid cell are summed so that each grid cell has a value of tons of emissions.

Figure 2. Inputs and outputs for the point-to-grid operator service.

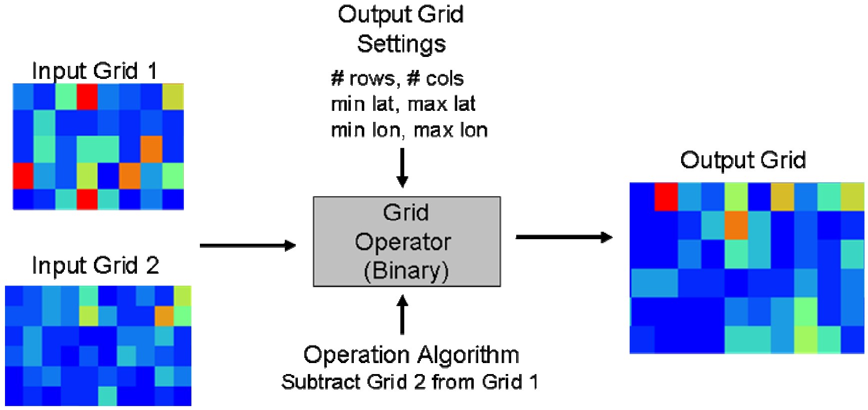


The point-to-grid web service was created within the DataFed Microsoft .NET environment and exposed for access through web service standards. It provides a general structure for any service that has points as input and a grid as the output. The service framework traverses each grid cell in the grid and executes a specified operation for each grid cell. The actual operation is defined by a script. In our case, we created a spatial aggregation script that identifies all points within the grid cell and sums their emissions. Other operations can be run using the point-to-grid operator. Since the structure is already defined, all that is needed for a different operation is a new script. A script for spatially interpolating point values via an inverse distance weighted algorithm has been added. Other scripts will likely be added by others in the future.

<u>Grid Operator</u>

The Grid Operator is a binary service, meaning it takes two grids as input, conducts a mathematical operation using their values, and creates a single grid output (Figure 3). This is a common operator type in Geographic Information Systems (GIS) and is commonly called “map algebra.” Examples include simple operations, such as ‘Grid A + Grid B’ or ‘Grid B / Grid A’ and expression evaluations, such as ‘(2*Grid A)/(Grid B / 1000).’ As with the point-to-grid service, the grid operator service provides a script interface for the calculation executed on each grid cell.

Figure 3. Inputs and outputs for the grid operator service.

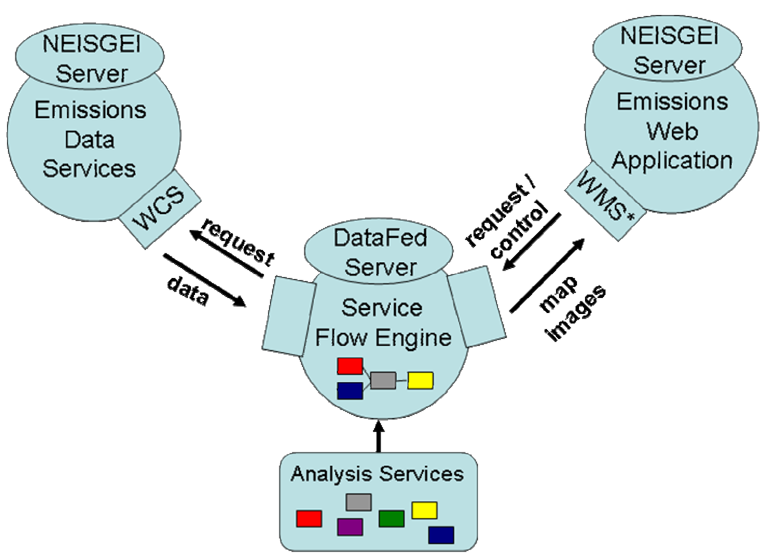


The grid operator includes settings to accommodate differences in the two input grid cells. If the two grids do not cover the same geographic area, a user can specify whether the output grid should be restricted to the area covered by one of the grids, the intersection of the two grids, the union of the two grids, or a user defined geographic area. In case the two grids having different resolutions (number of rows and columns), an option is available to make the output adhere to the resolution of one of the grids or a user defined resolution.

Service Framework

Web services are modular components that gain value when connected to form a chain of services, thereby creating a web application. The services can be geographically distributed among servers. The services come together by way of a workflow, which constructs and manages a set of services chained together. The data services used for emissions inventories and the web application that provides a user interface to the services reside on the NEISGEI server. The analysis services and the service flow management framework reside on the DataFed server. The services interact because they adhere to standard web standards that the service flow engine can then use to pass data between services. As shown in Figure 4, the web application interface provides service settings (e.g., which datasets to use and what type of analysis to conduct) to the service flow engine for executing an analysis service chain. The service engine requests data from the emissions Web Coverage Service (WCS) for input in one or more analysis services. The output of the service chain is sent to web application where it is displayed for the user.

Figure 4. The interaction and interoperability among the data access services, analysis services, service flow engine, and emissions comparison application. *The WMS request from the web application is extended to include non-standard WMS elements that control the service flow.



COMPARATIVE EMISSIONS ANALYSIS APPLICATION

Reconciliation of Regional and Global Emissions Inventories

The application presented here is based on an "offline" analysis conducted for reconciling and comparing regional and global emissions inventories. The focus of that analysis was the reconciliation of global inventories developed using "top down" methods with regional inventories developed using localized "bottom up" methods for a domain covering the continental United States. The regional and global emissions reconciliation study examined similarities and differences among a variety of emissions inventories but focused its analysis on the US NEI and the EDGAR inventories. Substantial effort was involved in getting the multiple emissions datasets into formats that could be meaningfully compared through GIS and summary statistic analyses. Part of the analysis included comparison of emissions spatial distribution patterns. The study concluded that the spatial patterns among regional and global inventories agreed fairly well but there were notable differences among the annual emission estimates at local scales. The analysis was supported by summary tables of source emissions as well as maps depicting emissions represented by colored grid cells.

The web application developed here is designed to support the spatial analysis and make it more readily available through web tools. The goals for the web application are to make the analysis easier to execute (e.g., reduce amount of effort in accessing and reformatting data), more flexible, by being able to rapidly create spatial comparison maps with new data, and available to a broader emissions analysis community.

Web Application

A web application was developed to provide a user interface for conducting spatial comparisons between emissions inventories. In order calculate spatial comparisons, a common spatial data structure is required. A service is provided in the application for gridding point and area emissions inventories. Once the point and area emissions inventories are spatially aggregated to a grid, they can be compared with gridded emissions models.

The application's user interface is shown in Figure 5. It consists of three maps, each with its own set of user controls. The larger map at the top of the application window is the emissions comparison output. The controls for specifying the type of comparison and its associated settings are described in table 2. The output grid is dependent on the two input grids which are represented by the two smaller maps. Each of the two input grids can be either an existing gridded emissions dataset, such as EDGAR or RETRO, or a newly calculated gridded emissions dataset derived from point or area emissions, such as US NEI, Mexican NEI or the Canadian NPRI. Therefore, the emissions comparison can be run on two gridded emissions models, two point or area emissions, or one gridded emissions model with one point/area inventory. The settings for the two input grids, grid a and grid b, are described in tables 3 and 4. Each input grid can be previewed in a map before the grid comparison is run.

The settings in the web application controls dictate the web service flow, as depicted in Figure 6. The Grid Operator has two input flows stemming from either the output of a point to grid operator or a data access request to a gridded emissions data source. Changing the settings of the web application controls and submitting a new calculation request, reruns the process to generate an updated map.

Figure 5. Screenshot of the emissions comparison web application. Numbers correspond to control descriptions in Tables 2-4.

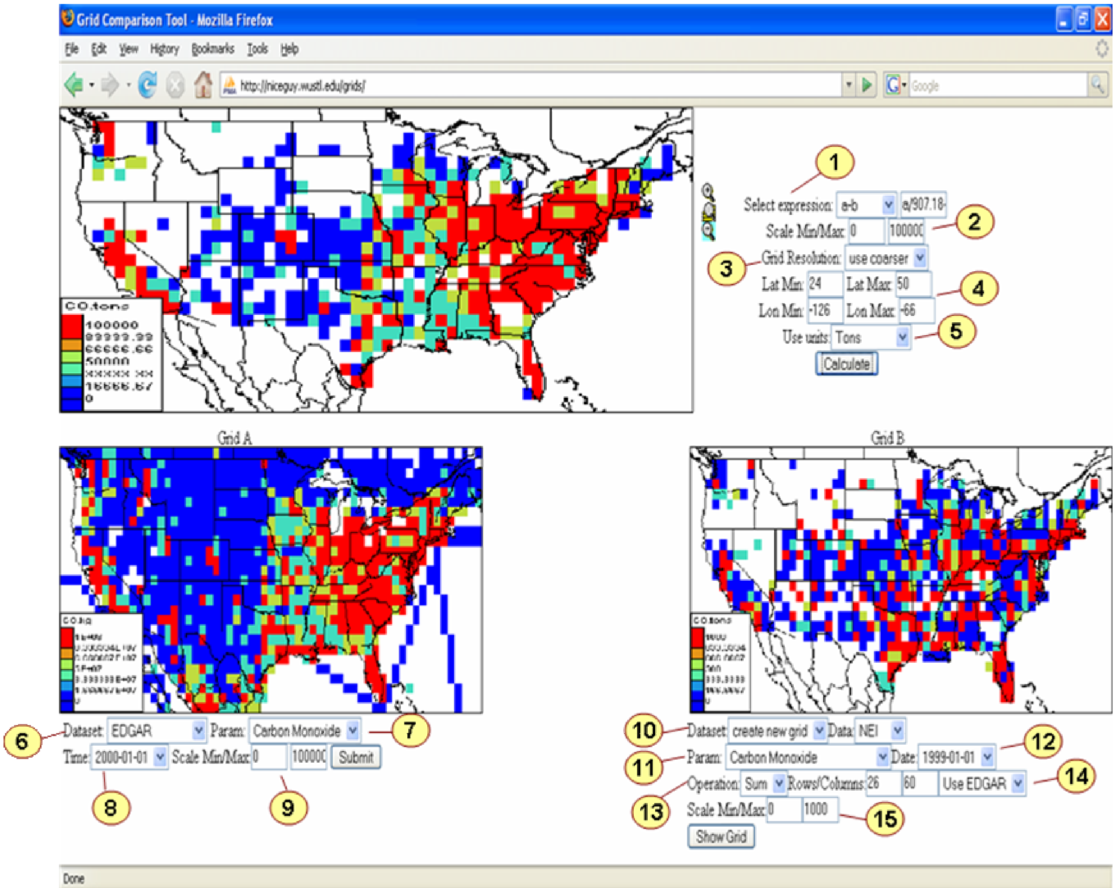


Figure 6. Web service flow for the emissions comparison application. Datasets are depicted as blue ovals. Web services are depicted as orange rectangles.

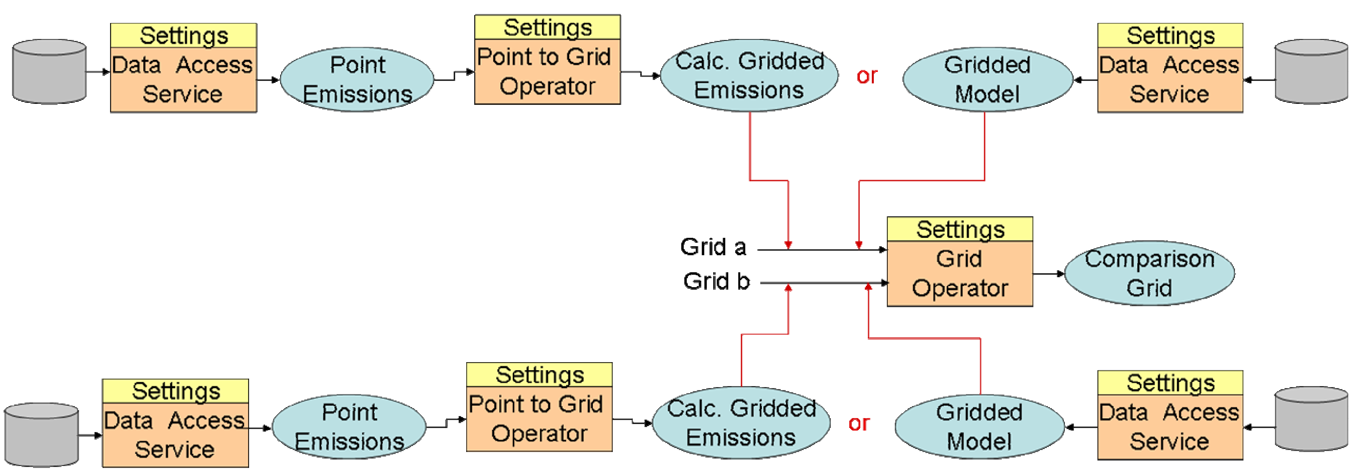


Table 2. Controls for output grid from emissions comparison calculation.

Control	Description	Corresponding # in Figure 4
Comparison Specification	Selection or setting of the calculation to be used in comparing grid a with grid b	1
Scale	The range used for setting the color range in displaying the output values from the calculation. Values correspond to the legend in the map.	2
Resolution	Selection of the output grid resolution. Options are the coarser grid or finer grid of the two grid inputs.	3
Bounding Box	Setting the minimum and maximum latitudes and longitudes of the area over which the comparison should be made	4
Units	The units of the comparison output. During the calculation the input data from grid a and grid b will be converted, if necessary, to the selected units.	5

Table 3. Controls for Input Grid a.

Control	Description	Corresponding # in Figure 4
Dataset	The gridded emissions dataset to be used as input to the grid comparison calculation. Current options are EDGAR and RETRO.	6
Parameter	The pollutant used for the input grid.	7
Date/Time	The emissions year for the input grid.	8
Scale	The range used for setting the color range in displaying the grid emissions. Values correspond to the legend in the map.	9

Table 4. Controls for Input Grid b.

Control	Description	Corresponding # in Figure 4
Dataset	The point or area emissions dataset to be used for creating a new gridded emissions database. Current options are NEI, MNEI, and NPRI	10
Parameter	The pollutant used in creating the grid.	11
Date/Time	The emissions year to use in creating the grid.	12
Operation	The mathematical operation to apply in aggregating the emissions that fall within a grid cell. Currently the only option is ‘sum.’	13
Rows/Columns	The number of rows and columns for the new gridded emissions. Options include specifying custom numbers of rows/columns or matching the resolution from EDGAR or RETRO.	14
Scale	The range used for setting the color range in displaying the grid emissions. Values correspond to the legend in the map.	15

The settings in the web application controls are used by Javascript code to assemble the URL text strings that pass the web service settings to the service control engine. Figure 7 captures this flow of information from user interface to URL to service flow engine in an example that calculates the absolute difference in EDGAR and US NEI point emissions for carbon monoxide.

The online application is made possible through dynamic internet access to emissions inventory data and web standards for connecting data access services with data analysis and map visualization services. An advantage of web service approach compared with traditional analysis tools, is that the analysis can be run dynamically through a web browser. The same analysis services can be quickly re-applied when emissions data are updated or applied to new comparisons with other inventories, such as changing the gridded model input from EDGAR to RETRO in running a comparative analysis between US NEI and the 0.5x0.5 degree RETRO inventory.

We are in the process of testing the application and conducting various emissions comparisons to determine its performance, accuracy, and time/cost savings in conducting analysis. Others are encouraged to use the application and web services in order to assess their usefulness in emissions analysis and to suggest improvements. The emissions comparison application and its associated web services are openly available online (www.neisgei.org) for other user-defined applications.

Future Enhancements

Some of the enhancements being pursued for the emissions comparison tool include:

Elevated Source Types – The Stella emissions comparison analysis explored an isolated comparison of elevated source types since those are most influential on long range transport. Including filters on the data access service to only retrieve the elevated source types could accommodate such analysis in the web application.

Open Layers – OpenLayers is an open source javascript library for creating map applications similar to GoogleMaps or Yahoo!Maps.[24] It provides map navigational tools and a framework in which to import distributed data in maps and build navigational and manipulating interfaces.

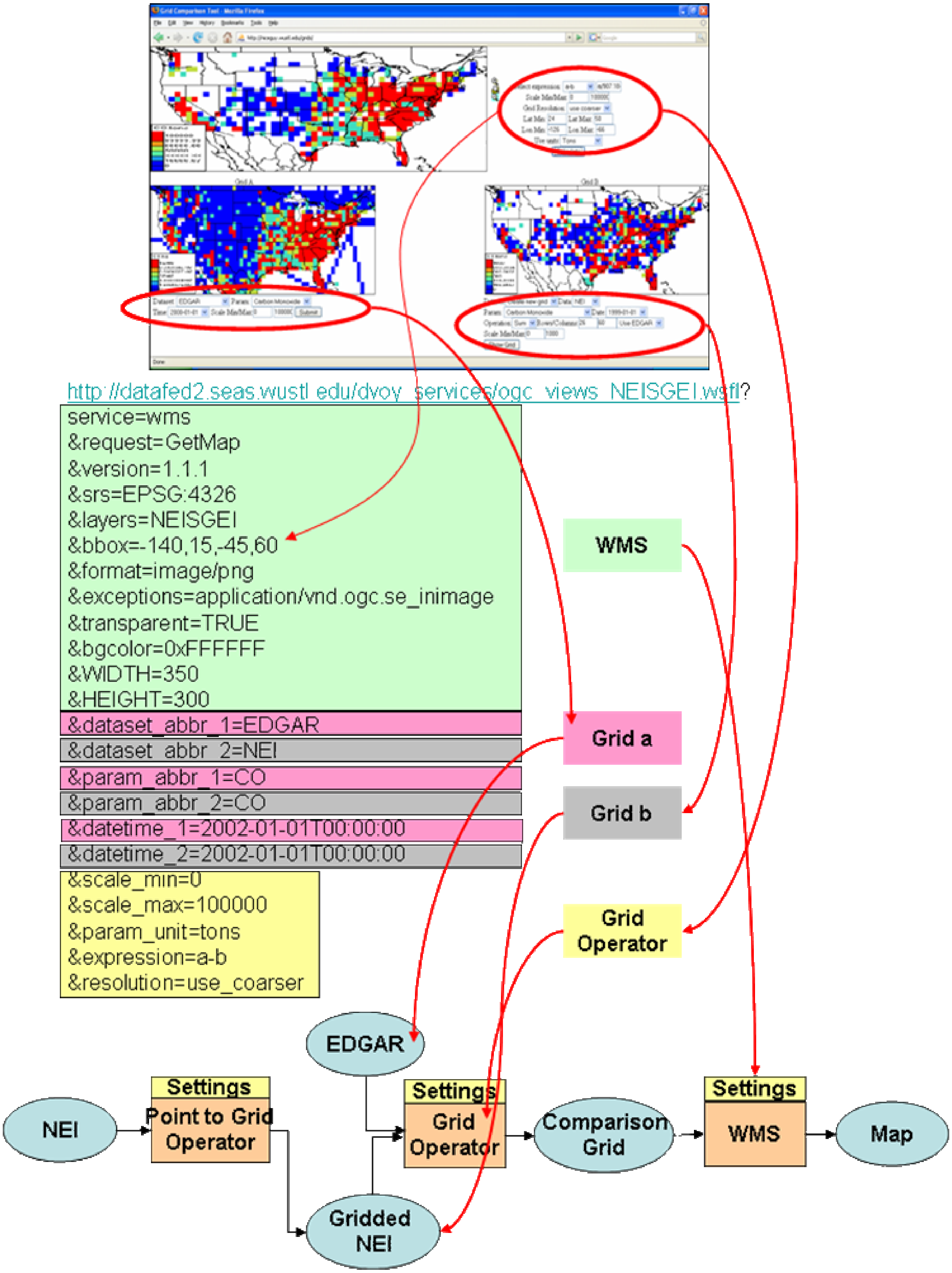
Distributed services – The data access and analysis services currently reside on independent servers but the ultimate goal is to demonstrate interoperability with distributed analysis services. We aim to assemble an application that uses data services from multiple servers, analysis services from multiple servers, and a service flow engine that resides on yet another independent server.

Satellite data – Remote sensing imagery is now being used to assess emission inventories. We are working on integrating satellite data as part of the emissions comparison tool and defining methods for meaningfully comparing satellite data with emissions inventories.

Other spatial aggregation functions – Currently, the point emissions in a grid cell are simply summed. We foresee a need for other types of spatial aggregation functions, such as weighted algorithms, that could be included as scripts in the point to grid operator.

Export data – The output grids are only available as images in the current version of the user interface. There is a need to be able to export the gridded data so that they can be imported into other online or offline analysis tools.

Figure 7. The relationship and data flow between the emissions analysis application's user interface, URL requests for passing control information, and the service flow engine that interprets the URL and executes the services.



CONCLUSIONS

Web services are being successfully used in creating web applications (ala mashups) to work with data beyond simple visualization or download. We created a web application that supports emissions data analysis in spatially allocating emissions to a grid for comparison with other gridded emissions inventories. The application is based on web services and constructed so that the individual data access and analysis services could be reconfigured (e.g. adding a different spatial allocation algorithm) or rewired (e.g. new combination of services) to support a different type of application. The web application simplifies data access and provides an environment for quickly running gridded comparisons based on various inventory-inventory combinations. The process for chaining web services needs to be simplified before it achieves its full potential of providing easy to use analysis tools for a broad audience but prototypes, such as the application built in this project, are incrementally moving forward toward this goal.

REFERENCES

[1] *The NARSTO Emission Inventory Assessment Team, Improving Emission Inventories for Effective Air Quality Management Across North America: A NARSTO Assessment*, NARSTO 05-001, 2005.

[2] Stella, G. Reconciliation and Comparison of Regional and Global Emission Inventories, Contract No: 68D-02-066, Assignment No: 4-07, prepared for US EPA Office of Air and Radiation, Washington DC, September 2006.

[3] Atkins, D.E.; Droegemeier, K.K; Feldman, S.I.; Garcia-Molina, H.; Klein, M.L.; Messerschmitt, D.G.; Messina, P.; Ostriker, J.P.; Wright, M.H. *Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*. Technical report, National Science Foundation, 2003.

[4] National Science Foundation, *Cyberinfrastructure Vision for 21st Century Discovery*, National Science Foundation Cyberinfrastructure Council, March 2007, online: http://www.nsf.gov/od/oci/CI_Vision_March07.pdf

[5] Taylor, J. *Defining e-Science*, National e-Science Centre, online: http://www.nesc.ac.uk/nesc/define.html (accessed April 2007).

[6] Foster, I. "Service Oriented Science," Science, 2005, Vol 308, Issue 5723, 814-817.

[7] Falke, S., Stella, G., Keating, T., and B. Hemming, "Cyber infrastructure for Emissions Data and Tools," Presented at the 15th Emission Inventory Conference of the U.S. Environmental Protection Agency, New Orleans, LA, May 2006.

[8] Brodaric, B., and M. Gahegan, *Representing Geoscientific Knowledge in Cyberinfrastructure: challenges, approaches and implementations*, In: Sinha, A.K. (Ed.), GeoInformatics, Data to Knowledge, Geological Society of America Special Paper 397, 2006, 1-20.

[9] Droegemeier, K., Chandrasekar, V., Clark, R., Gannon, d., Graves, S., Joseph, E., Ramamurthy, M., Wilhelmson, R., Brewster, K., Domenico, B., Leyton, T., Morris, V., Murray, D., Plale, B., Ramachandran, R., Reed, D., Rushing, J., Weber, D., Wilson, A., Xue, M., and S. Yalda. 2004: *Linked environments for atmospheric discovery (LEAD): A cyberinfrastructure for mesoscale meteorology research and education.* Preprints, 20th Conf. on Interactive Info. Processing Systems for Meteorology, Oceanography, and Hydrology, Seattle, WA, Amer. Meteor. Soc.

[10] ProgrammableWeb.com, *API Dashboard, Top APIs for Mashups*, online: http://www.programmableweb.com/apis (accessed April 2007)

[11] The Open Geospatial Consortium, *The OpenGIS® Abstract Specification, Topic 6: The Coverage Type and its Subtypes, Version 4*, Wayland, Massachusetts, 2000.

[12] de la Beaujardiere, J. (editor), 2002, *Web Map Service Implementation Specification(WMS),* Version 1.1.1, OGC Document OGC 01-068r3.

[13] Vretanos, P., (editor), 2004, *Web Feature Service Implementation Specification V1.1*, OGC Document OGC-04-094.

[14] Evans J., Bernhard B., Di L., (eds.) 2003. *Web Coverage Service (WCS)*, Version 1.0.0, OGC 03-065r6. OpenGIS© Implementation Specification. Open GIS Consortium.

[15] Na A., Priest M. (editors), 2006. *OpenGIS Sensor Observation Service*. OGC 05-088r1. Open Geospatial Consortium. GML

[16] Unidata, netCDF website, online: http://www.unidata.ucar.edu/software/netcdf/ (accessed April 2007)

[17] Husar, R.; Poirot, R., "DataFed and Fastnet: Tools for Agile Air Quality Analysis"; *Environmental Manager* 2005, September, 39-41

[18] Husar, R. and E. Robinson, "Architectures and Technologies Enabling the Diffusion of Atmospheric Science Information" presented at European Geosciences Union General Assembly, Vienna, Austria, April 18, 2007.

[19] US EPA, *What is the National Emissions Inventory (NEI)*, online: http://www.epa.gov/ttn/chief/net/neiwhatis.html (accessed April 2007)

[20] Eastern Research Group, Inc. and TransEngineering, *Mexico National Emissions Inventory, 1999*, 3393-00-011-002 ,prepared for Secretariat of the Environment and Natural Resources and the National Institute of Ecology of Mexico with support by United States Environmental Protection Agency, Western Governors' Association and North American Commission for Environmental Cooperation, October 2006.

[21] Environment Canada, National Pollutant Release Inventory (NPRI) Data, online: http://www.ec.gc.ca/pdb/npri/npri_dat_rep_e.cfm (accessed April 2007)

[22] Olivier, J., Van Aardenne, J., Dentener, F., Ganzeveld, L. and J. Peters, 2005. *Recent trends in global greenhouse gas emissions: regional trends and spatial distribution of key sources*. In: "Non-CO2 Greenhouse Gases (NCGG-4)", A. van Amstel (coord.), 325-330. Millpress, Rotterdam, ISBN 90 5966 043 9.

[23] Schultz, M., Pulles, T., Brand, R., van het Bolscher, M., and S. Dalsøren, 2005, *A global data set of anthropogenic CO, NOx, and NMVOC emissions for 1960-2000* (paper in preparation)

[24] OpenLayers, online: http://www.openlayers.org/ (accessed April 2007)

ACKNOWLEDGEMENTS

Ed Fialkowski developed the NEISGEI Web Coverage Service and the javascript, php and python code for the comparison web application. Kari Hoijarvi provided key technical guidance on the data access and analysis services, the service flow engine and in working with netCDF. Rudy Husar's insights on service oriented architectures were invaluable in designing and implementing the web service based application. Sunil Shrestha and Mark Bober assisted with the creation of the emissions data access services. The work presented was funded in part by the EPA Office of Air and Radiation cooperative agreement, "Cyberinfrastructure for Air Quality Management."

KEY WORDS
Emissions inventories, web services, emissions analysis, regional emissions, global emissions, service oriented architecture, cyberinfrastructure