Annual Progress Report:

Parallelization and Visual Analysis of Multidimensional Fields: Application to Ozone Production, Destruction, and Transport in Three Dimensions

Grant No. NAGW-3886

Karsten Schwan

January, 1996

College of Computing
Georgia Institute of Technology
Atlanta, Georgia 30332
email: schwan@cc.gatech.edu
fax: 404-853-9378

List of Investigators:

Karsten Schwan, College of Computing
Fred Alyea, Earth and Atmospheric Sciences
M. William Ribarsky, College of Computing
Mary Trauner, Information Technology
Greg Eisenhauer, student, College of Computing
Yves Jean, student, College of Computing
Weiming Gu, student, College of Computing (PhD, 12/95)
Ray Wang, postdoc, Earth and Atmospheric Sciences
Jeffrey Waldrop, student, College of Computing
Beth Schroeder, student, College of Computing
Jeremy Heiner, student, College of Computing

List of associated investigators (not funded by NASA):

Thomas Kindler, student, Earth and Atmospheric Sciences(PhD, 12/95)
Dilma Silva, student, College of Computing
Jeffrey Vetter, student, College of Computing (funded by separate NASA GSRP grant.)
Vernard Martin, student, College of Computing

I. Scientific Application and Analysis

Scientific Progress

The three-dimensional, spectral transport model used in the current project was first successfully integrated over climatological time scales by Dr. Guang Ping Lou for the simulation of atmospheric N_2O using the United Kingdom Meteorological Office (UKMO) 4-dimensional, assimilated wind and temperature data set. A non-parallel, FORTRAN version of this integration using a fairly simple N_2O chemistry package containing only photo-chemical reactions was used to verify our initial parallel model results. The integrations reproduced the gross features of the observed stratospheric climatological N_2O distributions but also simulated the structure of the stratospheric Antarctic vortex and its evolution. A paper describing this work was presented at the Spring, 1994 AGU meeting (Lou, et al, 1995) and an enlarged version suitable for publication is currently in preparation (Lou, et al, 1996).

Subsequently, Dr. Thomas Kindler, who produced much of the parallel version of our model, enlarged the N_2O model chemistry package to include N_2O reactions involving $O(^1D)$ and also introduced assimilated wind data from NASA as well as UKMO. Initially, transport calculations without chemistry were run using Carbon-14 as a non-reactive tracer gas with the result that large differences in the transport properties of the two assimilated wind data sets were apparent from the resultant Carbon-14 distributions. Subsequent calculations for N_2O, including its chemistry, with the two input winds data sets with verification from UARS satellite observations have refined the transport differences between the two such that the model's steering capabilities could be used to infer the correct climatological vertical velocity fields required to support the N_2O observations. During this process, it was also discovered that both the NASA and the UKMO data contained spurious values in some of the higher frequency wave components, leading to incorrect local transport calculations and ultimately affecting the large scale properties of the model's N_2O distributions, particularly at tropical latitudes. Subsequent model runs with wind data that had been filtered to remove some of the high frequency components produced much more realistic N_2O distributions. A paper presenting these results and data limitations was given at the Fall, 1995 AGU meeting (Kindler, et. al, 1995) and, in more detail, in Dr. Kindler's Ph.D. Thesis at Georgia Tech (Kindler, 1995).

During the past few months, the UKMO wind data base for a complete two-year period was processed into spectral form for model use. This new version of the input transport data base now includes complete temperature fields as well as the necessary wind data. This was done to facilitate advanced chemical calculations in the parallel model which often depend upon temperature. Additional UKMO data is being added as it becomes available.

How Interactive Steering Contributes To Scientific Applications

When combining on-line visualization and steering tools with our group's parallel version of a global spectral atmospheric transport model, users have the unique ability to compare model results with observational data during the model run. Should discrepancies between model results and observations occur, model execution can be stopped, rolled back in time, model parameters may be changed, whereupon the the user can then rerun the model with new parameter settings. Our experiences with this new approach in model validation are quite positive. Specifically, when applying these interactive validation methods to the scientifically relevant problem of simulating the global distribution and transport of Nitrous Oxide (N_2O), interesting scientific outcomes result from a comparison of results using simulated windfields for transport versus using assimilated (measured) windfields for driving the transport inside the model. When comparing the results from using two different sets of assimilated windfields (NASA, UKMO), the model shows an underprediction of vertical mass transport in the equatorial area with the UKMO windfields and an overprediction of the vertical transport with the NASA winds. Online model interactions permitted us to adjust and "play with" vertical windfields to investigate in detail the sensitivity of the biased model results to changes in the vertical advection term. As a result and compared with with observation data, model results were improved significantly for both windfield sets.

II. High Performance Computing: Implementation Summaries

Parallelization of the Atmospheric Modeling Code

The original atmospheric model was written in FORTRAN. To facilitate on-line monitoring, it was necessary to rewrite the code in C since the current FALCON system (which supports on-line monitoring) only supports the C programming language. An effort is underway to develop a version of FALCON that will work with other languages. (The FALCON system is described fully in the Infrastructure section.)

The parallelized model was targeted for two different computer architectures. The first approach was to target a shared memory machine model. A Kendall Square KSR 2 supercomputer was first chosen for this implementation. This shared memory model parallelizes the computations by both atmospheric level and by term in the Navier Stokes equation. When Kendall Square Research closed its doors, the model was ported to the Power Challenge Series of supercomputers manufactured by Silicon Graphics. Both machines run a variant of the UNIX operating system and support the shared memory paradigm so porting the software was of minimal effort.

The second approach was to target a distributed memory machine model called a message passing model. The machine used to implement the model was a group of three high end IBM workstations and an IBM SP-2. We used the MPI message passing library to communicate between the machines. Each processor was assigned work to do by a master process. Each of the slave processors calculated its portion of work, occasionally communicating with other processor to get needed information. Results were sent back to the master processor at the end of each timestep. In addition to the responsibility of gathering information for the entire application, the master processor also was given a portion of work to do. This distributed memory model parallelizes the application by level only in order to keep communication costs down.

On-line Program Monitoring and Steering

The monitoring is accomplished by inserting what we call "sensors" into the actual code at compile time. Sensors are declared earlier by the programmer and are used to gather interesting information about a program's state at a particular moment. At run time, when the code encounters a "sensor", the sensor will gather up whatever information it needs and will send that information to a centralized routing station called a "hub". The hub is a program, usually running on another machine, that gathers information and stores or forwards it to other applications as necessary. For our system, the visualization system connects to the "hub" to request the monitoring information about the application. Although it is theoretically possible to have the application communicate directly to the visualization, the hub provides more functionality in that it can offload work to another processor and it provides facilities for allowing an arbitrary number of applications to connect and access the same monitoring information without affecting the application in any way. Also, the hub allows us to easily provide a communications interface for the steering function that will accommodate multiple steerable components without unduly affecting the running time of the application.

To steer the application, special sensors are inserted into the application that check for steering commands from the hub. If a command is waiting to be received from a hub, then the sensor interprets the steering command and modifies the application state accordingly. In most cases, steering is accomplished by first stopping the application from proceeding further in its simulation, changing the program state, and then re-starting the application so that it may continue with the new state information. This start/modify/stop sequence is done to insure that all parts of the application are synchronized and have the necessary state information so that the model is in a consistent state and thus the calculations are consistent as well. (Monitoring and steering are described fully in the Infrastructure section.)

Our applications include monitoring sensors for the wind fields and for the concentrations of various (single) chemical constituents. A selectable 2D or 3D interactive visual interface allows the scientist to move through the data at each timestep using various projections as desired. Steering sensors allow the scientist to evaluate and test new values for the wind fields in conjunction with simple checkpoint and restart facilities which are required to assure stable and accurate simulation behavior. As mentioned above, the spurious wind values were discovered through the use of this interface. (The visualization system is described fully in the Visualization section.)

Parallel Code Performance Evaluation

This work concerns the parallel implementation of a grand challenge problem: global atmospheric modeling. The novel contributions of our work include: (1) a detailed investigation of opportunities for parallelism in atmospheric transport based on spectral solution methods, (2) the experimental evaluation of overheads arising from load imbalances and data movement for alternative parallelization methods, and (3) the development of a parallel code that can be monitored and steered interactively based on output data visualizations and animations of program functionality or performance. Code parallelization takes advantage of the relative independence of computations at different levels in the earth's atmosphere, resulting in parallelism of up to 40 processors, each independently performing computations for different atmospheric levels and requiring few communications between different levels across model time steps. Next, additional parallelism is attained within each level by taking advantage of the natural parallelism offered by the spectral computations being performed (eg., taking advantage of independently computable terms in equations).

Performance measurements are performed on a 64-node KSR2 supercomputer. However, the parallel code has been ported to several shared memory parallel machines, including SGI multiprocessors, and it has now been ported to several modern platforms, including the IBM SP-2 machine, the SGI Powerchallenge , and workstations. Additional performance evaluation is ongoing on those machines.

III. Infrastructure

In order to enable our integrated approach to simulation and visualization/analysis, we have developed Falcon, a toolkit that collectively supports on-line monitoring, steering, visualization, and analysis of parallel and distributed simulations. The general usefulness of the toolkit is demonstrated by its diverse application to areas such as interactive molecular dynamics simulation and interactive simulation of fault containment strategies in telecommunication systems. It is anticipated that the Falcon toolkit will be available for distribution on the WWW in the near future. The Falcon toolkit includes sensors and probes, an on-line steering system, DataHub for routing information from a server to multiple clients, and Portable Binary I/O (PBIO) for transmitting binary data between machines.

Sensors, probes, and steering objects inserted in the simulation code are generated from monitoring and steering specifications. Their partially analyzed monitoring information is sent to graphical and visualization displays. Once steering decisions are made by the user, changes to the application's parameters and states are made by Falcon's steering mechanism which invokes the steering objects embedded in the application code.

Falcon's on-line steering component consists of a steering server on the target machine that performs steering, and a steering client that provides the user interface and control facilities remotely. The steering server is typically created as a separate execution thread of the application to which local monitors forward only those monitoring event that are of interest to steering activities. The steering client receives application run-time information from the application, displays the information to the user, accepts steering commands from the user, and enacts changes that affect the application's execution. Communication between application and steering client and steering client and server is handled by the transmission tool, DataHub.

DataHub is a transmission tool for routing messages between multiple clients where clients can be broadly classified as applications, visualization/analysis/steering tools, or other DataHubs. Messages are identified by their format names and registered with DataHub by both senders and receivers. When a message is received, it is routed to those clients who have registered their interest in receiving that message type. Communication is done either through sockets or file I/O. The hub server can provide additional functionality such as event reordering before data is routed to clients. DataHub and PBIO taken together provide a flexible display system for attaching different types of graphical and visualization displays to an application's execution. Graphics intensive clients, which run on high performance front-end workstations to take advantage of better graphics and visualization support, can be dynamically attached to and detached from the display system.

The program steering environment demands speed and compactness of binary data transmission in a heterogeneous environment. These needs are met by Portable Binary I/O (PBIO), a set of services for transmitting binary data between machines in heterogeneous environments. PBIO provides a low overhead service by not requiring data to be translated into a "standard" or "network" representation and portability by transferring data between machines despite differences in byte ordering, sizes of datatypes, and compiler structure layout differences.

Though PBIO uses a metaformat in which the actual formats of binary records could be described, the representation of the metadata is hidden. Writers of data provide a description of names, types, sizes, and positions of fields in records through calls to the PBIO library. Readers provide similar information. No translation is done on the writer's end; meta information describing the senders format is sent in the PBIO data stream. On the reader's end, the format of the incoming data is compared with the format the reading program expects. Where discrepancies exist, PBIO performs the appropriate translations.

IV. Visualization

We have integrated the Glyphmaker visualization system, including modules developed within the Iris Explorer environment, with the Falcon steering system and the atmospheric model. As the model generates timesteps, the visualization is updated in an on-line fashion. Additions to the visualization capability include modules to immediately display the data or to pass it along to PV-Wave for alternative visualizations and analysis.

By direct manipulation steering we mean that we can interact directly with visualizations of atmospheric simulations to alter the future course of the simulations. We do this, for example, by scaling, rotating, translating, or inputting data for graphical objects bound to the data. Thus we could use the conditional box (a tool from Glyphmaker) to define spatial regions in the data where one could change chemical concentrations or other parameters. We could also employ data probes from Glyphmaker to locate localized behavior of interest and to adjust parameter values where desired. We have extended the rendering module in Glyphmaker to support these direct manipulation capabilities. (The standard rendering module in Explorer is not nearly flexible enough for our purposes, so we have made extensive modifications to it using SGI Inventor.) Our direct manipulation techniques involve interactions with both 3D and 2D representations of the data. This hybrid approach is attractive because it recognizes that while new and innovative methods are necessary to explore spatially complex and multidimensional data or to control simulations that produce these data, familiar tools such as 2D plots are succinct ways of expressing user intent.

In our current version of the visualization/analysis tools, we have added a graphical steering mechanism to our Glyphmaker visualization system. The system allows the user to select from a set of geometric forms. The user can then deform the geometry to encompass a desired spatial region, within which one can change parameters in the atmospheric model. In addition to interactive control of position and deformation of the geometric steering objects, the user receives visual feedback from both the steering object's geographic position and from the model datastructure indices. The visual feedback is enhanced by allowing the user to choose from a variety of projections (spherical or flat) with the graphical attributes of the geometric form adjusted to the type of projection.

The direct manipulation steering approach is a new and powerful way to control spatially complex and dynamic simulations, such as those from atmospheric models. It allows the user to do side-by-side probing and analyzing of the correlations in the data while being able to redirect the simulation in a spatially intuitive way to better understand how the physical processes evolve. It requires the capability for direct, quantitative probing of data that we have built into Glyphmaker through the formulation of elaborate data structures that always connect the visual representations to the original data, allowing investigation down to the individual datum. The data structures change and expand dynamically as new bindings between visual representations and data are made. The steering also requires a close coupling with the steering control and data transfer mechanisms provided by Falcon. The first stages of this integration has been completed and future development will require updating of the Falcon system to respond to new needs placed on it by enhancement of the visual interface as well as modification of the modes of visual interaction necessitated by improvements in the Falcon system.

The flexibility of the Glyphmaker system allows the use of the steering objects for analysis as well. The data elements within the region could be reclassified with their own glyphs (e.g., with different shapes or colors than the surroundings) so that their behavior could be highlighted and followed in detail. We have added the capability to take these selected data and list any values or show their distribution in 2D plots. This is the process of mixing 3D visualizations with 2D quantitative analyses that we mentioned above.

During the reporting period, we have also worked closely with atmospheric scientists. We instrumented the atmospheric model with a mechanism for deferred steering. Our design allows model changes to be scheduled rather than applied immediately. This is necessary because the parallelized execution is kept efficient by minimal synchronization. Additionally we focused on steering the vertical windfields. The windfields are an important transport mechanism and are derived from observed data. Our steering system permits both human interactive control and automated input from weather data sources (satellites, etc.).

V. Funding and Indirect Effects

Our work should have broad relevance not only for atmospheric scientists but also for other members of the NASA community who must look at or control the production of large amounts of dynamic, multidimensional data. The direct manipulation steering mechanisms especially form a new set of tools for controlling simulations where the output is 3D, complicated, and dynamic. These tools are even more important for understanding atmospheric processes since they evidently depend on the detailed behavior of highly correlated variables. Thus the steering tools are of particular interest to the atmospheric scientists with which we work and should engender broader interest as well.

Specific impacts include the awarding of a grant from IBM to work on real-time visualization and on the development of hierarchical visualization schemes for navigating and exploring very large datasets. This research will use a heterogeneous environment consisting of an SP-2 multiprocessor machine and graphics workstation. In addition we have written proposals for the DOD Multidisciplinary University Research Initiative and for other programs based, in part, on this work. Finally the visualization/analysis and steering tools were presented during the site visit on a proposal from the College of Computing for an NSF Infrastructure Grant. The grant was subsequently awarded based, in part, on a demonstration of broad-based, innovative, and collaborative efforts using computing.

By leveraging the NASA grant, the investigators were able to acquire two major equipment grants from the National Science Foundation, entitled:

1) Karsten Schwan, Principal Investigator, jointly with M. Ahamad, M. Ammar, R. Fujimoto, and S. Hudson, "Interactive Computing on Cluster Computers", NSF CISE equipment grant, $105,000 from NSF, $50,000 cost sharing, July 1995 - June 1997.

2) Karsten Schwan, Principal Investigator, jointly with R. Fujimoto, S. Hudson, J. Limb, and M. Ahamad, "Distributed Laboratories", NSF CISE Research Infrastructure Grant, approx. $1,200,000, Aug. 1995 - July 1998.

These grants are helping us direct our efforts into the following new directions.

The continuing merger of computer and communication technologies is leading to a new computing/communications infrastructure of unprecedented magnitude, enabling new applications with broad economic and social impact. Yet, such applications pose major challenges to researchers in Computer Science and in application domains. We are constructing an infrastructure consisting of computer and communication equipment to support collaborative experimental research among five research projects that attack some of these challenges. These projects are closely integrated and build on each other. As concrete demonstrations of our joint work, we will realize systems from two different "driver" applications that integrate research results, system and networking software, and tools from each of the five projects.

The topic of our joint research is the realization of distributed laboratories, where individuals can interact with each other, and more importantly, with powerful, distributed computational tools as readily as if all were located in a single site. Our intent is to permit scientists, engineers, and managers at geographically distinct locations (including individuals "tele-commuting" from home) to combine their expertise in solving shared problems, by allowing them to simultaneously view, interact with, and steer sophisticated computations executing on high performance distributed computing platforms. The research results and tools resulting from these efforts will have broad application in many domains. However, we are using two specific applications to focus our efforts, and to help ensure that our results and software tools are properly integrated:

- A distributed laboratory for experimentation with high performance numeric computations for applications in molecular physics, atmospheric sciences, working with high performance atmospheric and pollution modelling, and manufacturing systems.
- A distributed laboratory for studying the behavior of future-generation, large-scale telecommunication networks through high performance parallel and distributed simulation models of wired and wireless networks, called the virtual telecommunication networks application.

The NASA application and some of the infrastructure developed with this grant will also be employed in a major new ARPA-funded effort:

> Karsten Schwan, Principal Investigator, jointly with M. Ahamad, C. Codella, and B. Mukherjee (IBM TJ Watson), "Object Technology for High Performance Systems", ARPA/CSTO, Mark Gersh, Program Manager, Oct. 1995 - Sept. 1998, approx. $1,200,000, awarded July 1995.

VI. Papers and Presentations

This section provides a summary of papers and presentations given at a number of conferences and meetings around the country. Abstracts and descriptions are provided for detail and clarification.

1. AUG Spring, 1995 meeting.

Reference:

Guang Ping Lou[1],, Fred Alyea, and Derek Cunnold, "3-D Simulations of N_2O Transport and Antarctic Vortex Evolution", presented at AGU 1995 Spring Meeting, Baltimore, MD, May 30-June 2, 1995, paper no. A51B-5.

Abstract:

[1] (now at General Services Corporation, Data Assimilation Office, GSFC, NASA, 7501 Forbes Blvd., Suite 200, Seabrook, MD 20706; ph: 301-805-6996; e-mail: glou@dao.gsfc.nasa.gov)

3-D Simulations of N_2O Transport and Antarctic Vortex Evolution

This study focuses on three areas: (a) the structure of the stratospheric Antarctic vortex and its evolution; (b) the transport of N_2O and dynamical forces that dominate these processes; (c) the climatology of the N_2O mixing ratio distribution and its driving factors. A 3-dimensional spectral chemical transport model was employed to simulate N_2O transport and study the driving forces that affect the processes. The dynamical driving fields are from the UKMO 4-dimensional assimilated data set. UARS CLAES N_2O mixing ratio are used for the N_2O initial conditions. Model results show that the N_2O distribution and transport closely resemble the CLAES measurements, especially at high latitudes. The correlation coefficients between CLAES N_2O temperatures, and model N_2O and temperatures are remarkably similar in terms of their meridional distributions. Diagnostic study and model simulation results reveal that while large-scale Eulerian mean vertical motion fields are upward inside the vortex, the mean residual circulation vertical velocity is downward. The monthly mean maximum sinking residual velocity is -0.40 cm/s at about 1.5 mb and -0.07 cm/s in the 30-9 mb layer inside the Antarctic vortex in September. The vortex first breaks in the upper stratosphere during September. Then the breaking process propagates downward to the 3-10 mb level in the middle of October. At the lower levels, 10-20 mb, the vortex breaks up in early November. These breaking processes continue to penetrate to lower levels at about 20-30 mb by late November. In the meridional transport of N_2O, eddy transport is the chief process. Especially at higher altitudes, there seems to be persistent eddy mixing going on at the middle latitudes during the early spring. However, the residual circulation transport dominates the long term vertical mixing. The bulge of the elevated N_2O mixing ratio in the tropical stratosphere is determined by the uplifting of mass by the residual circulation. During the Southern Hemisphere summer, the uplifting of N_2O by the residual circulation reaches above 1 ppb/day. The downward transport inside the vortex can exceed 2 ppb/day in the winter hemisphere. The climatological distribution of the N_2O mixing ratio follows the seasonal variations of the solar radiation. The bulge of the elevated N_2O shifts toward the summer hemisphere by up to 15 degrees in latitude. The slopes of the N_2O mixing ratios are sharper in the winter hemisphere and the surf zone is well defined in the middle latitudes on the zonal mean plots.

Presented at AGU 1995 Spring Meeting, Baltimore, MD, May 30-June 2, 1995, paper no. A51B-5.

2. AUG Fall, 1995 Meeting

Reference:

Kindler, T.P., D.M. Cunnold, F.N. Alyea, G.P. Lou, and W.L. Chameides. "A Comparison of CLAES N_2O Simulations using 3D Transport Models Driven by UKMO and GSFC Assimilated Winds", presented at AGU 1995 Fall Meeting, San Francisco, CA, December 11-15, 1995, paper no. A52D-9.

Abstract:

A Comparison of CLAES N_2O Simulations using 3D Transport Models Driven by UKMO and GSFC Assimilated Winds

A three dimensional chemical model has been developed. The model has a vertical resolution of approximately 1.25 km (on-half a UARS layer) and is spectrally truncated at T21. In this paper we will compare N_2O simulations from two calculations in which the model is driven by the windfields provided by the assimilation models of UKMO and GSFC. The calculations were initialized on September 1, 1992 with a distribution based on UARS CLAES N_2O measurements and were run for 13 months. The zonal mean gradients of N_2O are found to steepen using the GSFC wind fields whereas they flatten out using the UKMO fields (as we have previously reported). Consequently the calculated atmospheric lifetime of N_2O changes from 180 years initially to less than 100 years and longer than 200 years respectively using the GSFC and UKMO winds. The budgets of N_2O in the two calculations will be compared in terms of contributions by the residual mean circulation and mixing along isentropes. The degree of isolation of the polar vortices and the extent of iteration between the tropics and the extratropics will also be examined using area mapping analyses.

Presented at AGU 1995 Fall Meeting, San Francisco, CA, December 11-15, 1995, paper no. A52D-9.

3. Supercomputing '95, GII Testbed

Reference:

M. C. Trauner, V. C. Martin. "A Parallel Spectral Model for Atmospheric Transport Processes", GII Testbed and HPC Challenge Applications on the I-Way, Virtual Environments and Distributed Computing at SC '95, Supercomputing '95 Conference, San Diego, CA, December 3-8, 1995, project no. 13.

Abstract:

A Parallel Spectral Model for Atmospheric Transport Processes

Earth and atmospheric scientists at Georgia Tech have developed a global chemical transport model that uses assimilated windfields for the transport calculations. These models are important tools to answer scientific questions about the stratospheric-tropospheric exchange mechanism or the distribution of species such as chlorofluorocarbons, hydrochlorofluorocarbons, and ozone. This model uses a spectral approach common to global models to solve the transport equation for each species.

Ideally, in large-scale atmospheric simulations, the observational database should be closely coupled to the visualization/analysis process. In fact, there should be feedback in the form of steering between the latter and the simulation in order to yield more accurate representations of atmospheric processes and a significantly more focused investigation. Because the data have complicated 3D

structures and are highly time-dependent, the visualization approach must handle this dynamic data in a highly interactive fashion.

In this project, the researchers have combined all these aspects into a single, integrated approach. This has required a collaborative, interdisciplinary process involving atmospheric scientists and experts in high-performance parallel computing, visualization, and user interfaces. The process used here could serve as a template for building highly effective and powerful applications (and tools supporting them), a process where the developer comes away with a deeper understanding of user needs.

Discussion:

This application was accepted for execution over the GII testbed and visualization on the I-Way Wall.

A working prototype of the distributed memory model with visualization and minor steering was exhibited. The atmospheric transport model was running on 32 nodes of the IBM SP-2 supercomputer at the Cornell Theory Center. On-line monitoring data was shipped over a dedicated ATM network to San Diego to an SGI Challenge server which acted as a centralized resource manager and router (the Datahub.) The custom visualization was running on an SGI Onyx connected to the Wall.

The prototype allowed for the user to interactively (while the model is running) view both the wind fields and the N_2O concentrations at any part of the globe in a variety of interesting formats including spherical levels extending from the earth's surface into the stratosphere, a flat map Cartesian view with strict longitudinal and latitudinal planes, and simply x-y plots. The data viewed could be chosen via explicit selection or relative position using sliding bars and dials.

4. Penn State University

Reference:

Karsten Schwan, "Interactive High Performance Programs: From On-line Scientific Applications to Operating Systems", Penn State University, College Park, Dec. 1994.

5. Workshop on Debugging and Performance Tuning for Parallel Computing Systems

Reference:

Karsten Schwan, Weiming Gu, Greg Eisenhauer, Jeffrey Vetter, "Interactive Parallel Programs: The On-line Steering of Large-Scale Parallel Codes", invited lecture at the Workshop on Debugging and Performance Tuning for Parallel Computing Systems, Cape Cod, Oct. 1994.

VII. Looking Ahead

The Science and High Performance Computing

A more direct method of transforming the UKMO and NASA data to spectral form is being developed that will not require linear interpolation processes to "move" data from one grid system to a different one for spectral transformation. Although the interpolation process that has been used to date is not thought to contribute in any important way to the introduction of any spurious high frequency waves to the data, in view of the now-known existence of such waves in the wind data base, it is thought that the elimination of any potential high frequency noise that may be introduced numerically in preparation for transformation be undertaken. The distribution of energy as a function of spatial resolution for the transformed assimilated data base will then be compared with observational data in order to delineate the frequencies that contain spurious values.

A major upgrade of the parallel model that is currently under way involves the simultaneous integration with a number of atmospheric species and the inclusion of the necessarily complex chemical packages that will be required. For this purpose, we propose to make use of a substantially modified version of a large atmospheric chemical model obtained from the "Laboratoire de Physique et Chimie de l'Environment", CNRS, Orleans, France. This model is to be included as a separate module linked and interacting with the current parallel transport model and should thus permit state-of-the-art simulations of stratospheric mixes of important atmospheric constituents.

Minor changes to the parallel model that are planned for the next few months include the installation of new fourth-order numerical scheme for the spectral vertical diffusion calculations and the introduction of wind data at the lowest model levels to better simulate the effects of the Earth's surface boundary layer.

The infrastructure grant concerns tool development and distributions, especially focussing on steering and its use for scientific processors, with extensions of these tools to address entire distributed laboratories.

The Visualization

We are extending Glyphmaker in ways to increase its power in the analysis of atmospheric simulations that will grow significantly in size and complexity as the parallel approaches are scaled up. It will be necessary to manage levels of detail in the visualizations so that we can retain highly interactive exploratory analysis as the data grows. This will require both automatic and user-directed methods, since the user will not know at the outset what the data contains but will want to direct and refine the visualization process. We are working on general methods for detail management that are based on an understanding of the nature of physical data and that include both approaches for 3 and 4D (including time) pattern recognition and for feature recognition and extraction. These approaches will allow a natural organization of the data for further study including higher level visualization (e.g., surface and volumes) of general unstructured or scattered data. These approaches will also permit us to represent the data with visual abstractions at multiple levels of complexity. We will work closely with application scientists so that the visual abstraction process matches the physical abstraction process that they use to simplify and then understand their data.

We plan to extend the visual representations and interactions for steering. One extension will allow the user to specify distribution functions with a few parameters so that, for example, more physically accurate concentration profiles can be inserted into the simulation. Thus the user can easily specify how model changes are distributed within the extent of the steering object. Also, we are incorporating the ability to acquire steering specifications from the visualization output. For example, if an isosurface specification produces a surface in the visualization, we will be able to use the surface as a spatial parameter for steering. We plan to write a paper shortly on our present and some of our new steering capabilities.

In order to achieve our ultimate goal of real-time exploratory visualization, steering and control of simulations, regardless of the size of data output, we must investigate alternatives to our present visualization approach. Among other things, this means looking at tools other than SGI Iris Explorer and Inventor. The reason for this is that we must have fast rendering of thousands of potentially independent objects; neither Explorer or Inventor are optimized for this case. We are considering, for example, the use of the CAVE libraries from NCSA. These are built for scientific visualization in immersive virtual environments. They thus are built for real-time use, have been employed on big data, and have some tools for exploratory navigation built in. By integrating the CAVE libraries with Open Inventor, we can retain several of our interaction and direct manipulation tools. As an alternative, we are also considering building our own renderer using OpenGL. This will give us optimal efficiency and control over visualization capabilities. However, we will have to rebuild most of our interaction capabilities and some of our visualization techniques.

Whichever path we take for our rendering tools, we will move them from GL to OpenGL. This coupled with use of libraries like Open Inventor will make available a large number of platforms for use by our system.

Collaborative Steering

We plan to incorporate support for collaborative work in the monitoring/steering infrastructure beyond the simple example of replicating the pixels of a visualization on several workstations' screens. Support will be needed to allow the collaborators to have different views of a single visualization (or possibly different visualizations of the same data) and to coordinate the steering interactions and feedback among the views. For rendering the visualizations we use an object-oriented graphics library which allows one to arrange objects into a tree structure to describe a scene. This library includes several objects which respond to user input (mouse, keyboard, etc.) which we use for steering. To support collaboration we add a mechanism to this library which maintains consistent copies of the scene tree structure on two or more machines.

VIII. Bibliography

Lou, Guang Ping, Fred Alyea, and Derek Cunnold, "N_2O transport dynamics and its climatology", in preparation, 1996.

Kindler, Thomas Paul. "The Development of Supercomputing Tools In a Global Chemistry Transport Model (CTM) and Its Application to Selected Problems in Global Atmospheric Chemistry", Ph.D. thesis, Georgia Institute of Technology, December 1995.

Yves Jean, Thomas Kindler, William Ribarsky, Weiming Gu, Gregory Eisenhauer, Karsten Schwan, and Fred Alyea, "An Integrated Approach for Steering, Visualization, and Analysis of Atmospheric Simulations", Report GIT-GVU-95-15, *Proceedings Visualization '95*, pp. 383-387.

In this research we have constructed at tightly coupled set of methods for monitoring, steering, and applying visual analysis to large scale simulations. This work shows how a collaborative, interdisciplinary process that teams application and computer scientists can result in a powerful integrated approach. The integrated design allows great flexibility in the development and use of analysis tools. This work also shows that visual analysis is a necessary component for full understanding of spatially complex, time-dependent atmospheric processes.

Martin deBoer, Yves Jean, William Ribarsky, Gregory Newton, Frits Post, and Robert Sumner, "A New Data Organization for Analysis in Visualization Systems", Report GIT-GVU-96-10, to be submitted to *Transactions in Visualization and Computer Graphics.*
In this paper we present a general data organization for exploratory visualization/analysis. It supports mappings between (user-defined) visual representations and the original data. We present tools that take advantage of this organization and allow selection of data using spatial or other constraints. These selected data can be reclassified for binding to new visual representations (or hidden from view). The selection and binding is performed in an intuitive manner by direct manipulation so that even users who are not graphics experts can do it. The users then have easy-to-use tools to calculate new quantities from selected variables and thus to formulate new data structures. It is important to put these intuitive and easy-to-use tools directly in the hands of users who, though not graphics experts, are experts in their own fields and thus are the best ones to control the visualization/analysis process.

Yves Jean, Thomas Kindler, William Ribarksy, Weiming Gu, Gregory Eisenhauer, Karsten Schwan, and Fred Alyea, "Case Study: An Integrated Approach for Steering, Visualization, and Analysis of Atmospheric Simulations", *Visualization '95*, Oct. 1995.

Thomas Kindler, Karsten Schwan, Dilma Silva, Mary Trauner, and Fred Alyea, "Parallelization of Spectral Models for Atmospheric Transport Processes", *Concurrency: Practice and Experience,* to appear 1996.

Ph.D., Weiming Gu, "On-line Monitoring and Steering of Parallel Programs", August 1995.

Weiming Gu, Greg Eisenhauer, Karsten Schwan, and Jeffrey Vetter, "Falcon: On-line Monitoring and Steering of Large-Scale Parallel Programs", GIT-CC-94-21, journal submission, Dec. 1994, revised Aug. 1995.

Karsten Schwan, Fred Alyea, William Ribarsky, and Mary Trauner, "Integrating Programming Steering, Visualization, and Analysis in Parallel Spectral Models of Atmospheric Transport", *NASA Science Information Systems Newsletter,* Issue 36, July 1995.

Weiming Gu, Jeffrey Vetter, and Karsten Schwan, "An Annotated Bibliography of Interactive Program Steering", ACM SIGPLAN Notices, Sept. 1994

Greg Eisenhauer, Weiming Gu, Karsten Schwan,and Niru Mallavarupu, "Falcon -- Toward Interactive Parallel Programs: The On-line Steering of a Molecular Dynamics Application", High Performance *Distributed Computing (HPDC-3),* San Francisco, CA, Aug. 1994.

Weiming Gu, Greg Eisenhauer, Eileen Kraemer, Karsten Schwan, John Stasko, and Jeffrey Vetter, "Falcon: On-line Monitoring and Steering of Large-Scale Parallel Programs", *Frontiers 95*, Feb. 1995.

Greg Eisenhauer, Weiming Gu, Dilma Silva, Karsten Schwan, Jeffrey Vetter, Eileen Kraemer, "Opportunities and Tools for Highly Interactive Parallel Computing", *Proceedings of the Workshop on Debugging and Performance Tuning for Parallel Computing Systems: Toward a Unified Environment*, Los Alamos National Laboratories, IEEE PRESS, May 1996.

Ph.D., Jeffrey Vetter, "Interactive Steering of Large-Scale Parallel Applications" (in progress).

Jeffrey Vetter and Karsten Schwan, "Progress: A Toolkit for Interactive Program Steering", *24th International Conference on Parallel Programming*, IEEE, Aug. 1995.

Ph.D., Vernard Martin, "Performance of Heterogeneous Parallel Applications" (in progress).