# Swat Team Trip Report

## Scripps Institute of Oceanography
## San Diego, CA
## Monday, 29 September 1992 – Wednesday, 1 October 1992

### 1. Introduction

Beginning on Monday, 29 September 1992, the Sequoia 2000 Swat Team visited researchers at Scripps Institute of Oceanography (SIO) in San Diego. The meeting included swat team members (Jim Frew, Allison Woodruff, Duane Waliser, Peter Kochevar, and Mike Olson) as well as researchers from SIO and from the San Diego Supercomputer Center (SDSC).

The meeting spanned three days, with presentations by SIO researchers and computer scientists from other sitest on the state of Sequoia research. As was the case in Santa Barbara, visualization of scientific data received the most attention. However, emphasis was also placed on tools for analyzing data.

### 2. Background

Scripps is a department of the University of California at San Diego. Researchers there are largely independent of one another, but collaborate on some projects. Sequoia principal investigators at SIO are Richard Somerville, John Roads, Tim Barnett, and Warren White. Other researchers working on Sequoia-related projects include Dan Cayan, Shyh-Chin Chen, Nicholas Graham, Kyozo Ueyoshi, and Alejandro Pares-Sierra. There are a number of computer scientists at Scripps working on Sequoia-related projects. These include Mary Meyers, Tony Tubbs, Larry Riddle, Emelia Villaros-Bainto, Norm Hall, and Bill Borbridge.

Projects in progress are looking at land, oceanic, and atmospheric systems using observations and model data. Because of the size of the Scripps group and the diversity of the projects they are working on, this report will not describe individual projects in any detail. Instead, it will concentrate on areas in which the SIO researchers need common services.

### 3. Data and Storage in Bigfoot

The Bigfoot data store in Berkeley provides several terabytes of storage on several different devices, including magnetic disk and optical disk and tape jukeboxes. This storage is available to both database users and to users who want to store flat files outside of the database system. At present, the various devices have different interfaces; different programs are used to move data on and off of them. By January, Berkeley will provide FTP, NFS, and Postgres access to all supported devices.

Climate researchers should load the data they're interested in using. Frew will handle the allocation of space and will compile a list of all data available on the system. When data is loaded, it must include a description of its format. Other researchers will be free to use the data.

The computer scientists in Berkeley are very interested in moving large amounts of data into Bigfoot, so climate researchers are encouraged to move ahead with loading as quickly as they can. Since climatologists may not have the time or expertise to load data themselves, SIO will hire a staff member with DEC money to handle data loading.

Frew will compile the list of all data currently on Bigfoot and will give it to Warren White and other interested researchers immediately.

## 4. Datatypes and Schema Issues

To make it easier to share data, researchers should convert their datasets into a standard format that all Sequoia applications can understand. In order to do this, the project needs to define a set of datatypes (such as *temperature*, *depth*, and *salinity*) and their representations (including number of bits stored, precision, and encoding of units). A group of Sequoia members, chaired by Frew, is working on designing these types. The group is called *sequoia-schema*, and interested parties are encouraged to join.

The schema group has so far designed common table formats for data stored in Postgres tables. Work on datatypes and representations is proceeding. Until this work is finished, researchers should load data using whatever type is currently available, and makes sense for the values stored. For example, latitude and longitude might be stored as four-byte integer values.

If you have questions when you choose types for data stored in Postgres, contact Mike Olson or Frew for help.

John Roads pointed out that common standard formats won't be useful in all cases. For example, to checkpoint a model run, it may be necessary to store temperatures as double-precision floating point values, since the model's internal state must be restored from the checkpoint exactly. In cases like this, researchers should choose the datatype required to do the job. The goal of standardization is to make it easy to share data, not to make it impossible to do work.

## 5. Large Objects

Since many of the datasets Sequoia researchers work with are large — for example, grid output from Global Climate Model runs — the Sequoia standard schema should include standards for large object data. Large objects may be stored inside Postgres or as flat Unix files.

There are several issues in representing large objects:

- We would like all large objects of a single variety to have the same representation inside the database system.
- We would like all large objects of a single variety to have the same representation outside of the database system. That is, we should support a common import/export format for large object data, so that it can be shared with researchers who are not on Project Sequoia.
- We would like to avoid designing entirely new reresentations for large object data.

Sequoia should permit scientists to work with data in familiar formats. At the same time, to encourage sharing of data among different projects, it is important that a standard interchange format be developed. In principle, any dataset in the Sequoia database should be available in a standard format, and may also be available in other formats.

The sequoia-schema working group will come up with a standard representation for large object data. Interested researchers are encouraged to help define this standard.

Two important issues in the representation of large objects are *metadata management*, or how information describing the large object data is stored, and *representation*, or appropriate formats for the actual large object data. Much work has been done elsewhere on managing self-describing data objects, which include a description of the data format with the data itself. Net-CDF and HDF are two standards for representing self-describing objects. Zahid Ahmed and Peter Kochevar of SDSC have some experience in this area, and will work with the Sequoia schema group to incorporate large object representations into the standard schema.

Large object types of interest to Sequoia researchers include raster images and volumetric data. Volumetric data is a *k*-dimensional grid of *n*-tuples — for example, output from a GCM run, which is a four-dimensional volume with wind velocity, temperature, and other measurements at every point in the grid. Volumetric data is of especial interest. Mike Olson will look

at providing a general-purpose gridded large object data type by the end of October. If none can be written at Berkeley by then, the computer scientists at SDSC may write one.

## 6. Scientific Data Management

Sequoia researchers at SIO all do preprocessing and conversion of datasets they receive, visualize them, and would like to run statistical analyses on them after they identify interesting features in the visualization. This section describes the processing steps common to all SIO researchers and lists the services that the Sequoia computer science community should provide.

### 6.1. Data Conversion

Data from external sources must typically be processed in order to be useful to climate researchers. The first processing step requires translating the data from the format in which it arrives to a format that can be managed by the hardware and software used for analysis. Important issues include the size and precision of integer and floating point values, how to store dates and times, and minimizing the size of the processed dataset.

In general, this processing will happen before the data is loaded into Bigfoot, and the storage system will not have to provide conversion routines. Researchers who want the storage system to handle conversion automatically must provide routines to do the conversion; these can be invoked automatically when new data arrives.

Data conversion is not a very hard problem. It is related to the Sequoia standard schema, since the schema will define the format of the output dataset. However, no serious work is required here.

### 6.2. Data Transformation and Interpolation Techniques

Once the data is in a format suitable for analysis by the software tools that a researcher uses, it must generally be transformed in some way. The transformations of interest to Sequoia researchers at SIO are the same as those used by researchers at the Earth Space Research Group in Santa Barbara.

Data that arrives on an irregular grid must be mapped to a regular grid in some way. Since the two grids have different spatial resolutions, actual data values must be interpolated to produce values on the new grid.

The interpolation technique used by Warren White and other researchers at Scripps is called *optimal interpolation*. First, a variogram is produced from the data that represents how similar values are at different distances from one another. By finding the sill of the variogram[1], a researcher can estimate the degree to which values are affected by their neighbors, and the distance at which such interaction stops.

Second, the variogram is used to produce a weighting curve that assigns more weight to nearby points than to distant ones[2].

Third, the weighting curve is used on points on the input grid to produce points on the output grid. Since the weighting curve is bell-shaped and typically falls off quickly, the number of source points required to compute destination points is small.

This interpolation technique yields much better results than less sophisticated techniques like bilinear interpolation. SIO researchers have code that does optimal interpolation, and

[1] In computer science terms, the knee of the curve

[2] Of course, the variogram need not depend on distance, but it generally does.

will contribute it to Sequoia for inclusion in a data processing tool.

Optimal interpolation produces a good weighting curve to use in computing target values. Other interpolation techniques use other curves, but typically apply them to source points in the same way. Sequoia should develop a general-purpose interpolation tool for use on scientific data. Users should be able to select the variety of curve they want (for example, optimal or bilinear) and produce gridded output data easily. Frew will work with Warren White at SIO to specify the interface for this tool and to begin constructing it.

### 6.3. Visualization

This section describes current visualization efforts at SIO. Section 7 lists nine visualization benchmarks proposed for Sequoia by Warren White.

Data visualization is a tool to help scientists identify interesting properties that they can then analyze more rigorously. SIO scientists feel strongly that any visualization tool must be closely integrated with the analysis tools described in the next section. Peter Kochevar pointed out that visualizing data and visualizing the results of data analysis are equally feasible, and should both be supported.

SIO researchers are using many different visualization tools, including AVS, IDL, NCAR graphics, and other packages developed by researchers elsewhere or at Scripps. There is a good deal of IDL expertise. Users of NCAR graphics like it because of its strong bias toward the Earth sciences. Of the packages in common use, AVS and IDL are by far the easiest to extend, but there is not much experience with AVS so far.

Peter Kochevar of SDSC is working with John Roads of SIO to do visualization of some of John's regional climatological data. Once this project is completed, Peter will give a seminar at Scripps on how AVS can be used on data from other researchers.

In order to make it easier for Scripps to use to AVS, they will hire someone to become the local visualization expert. That person will provide support for the scientists, and will work with computer scientists at other sites. Jim Frew will work with Richard Somerville to generate a job description for this position.

In the short term, Scripps researchers are interested in hosting an exchange student from some other site to help with scientific data management and visualization. Tim Barnett has agreed to find funding for this person. Mike Olson and Jim Frew will to find an interested student[3], and Barnett will work with them to be sure that the student is right for the position.

### 6.4. Statistical Analysis of Data

Looking at data with a visualization tool suggests features that can be more rigorously analyzed. This section describes some of the types of analysis that SIO researchers do.

First, a common analysis technique is to concentrate on anomalies, rather than actual measurements. Anomalies are differences from mean (or otherwise expected) values. Anomalies tend to cluster, and anomalies in one domain (sea surface temperature, for example) tend to be clustered with anomalies in other domains (such as precipitation).

Second, virtually all SIO researchers use Complex Empirical Orthogonal Functions (CEOF) to simplify and characterize the data they work with. This technique is similar to Fourier series analysis in computer science, except that CEOF may choose to use non-sinusoidal curves in constructing the target curve. For a given dataset, values are plotted on

[3] So if you're interested, send us mail. San Diego winters are mild.

some grid, and the curve that passes through the points is constructed. Next, the curve's principal components are extracted. Generally, the first few components provide more than ninety percent of the curve's shape. This small CEOF set can then be used to characterize the dataset in subsequent analysis and in models.

Finally, several different statistical and mathematical techniques are used frequently by SIO scientists. Correlations among datasets, analysis of variance (ANOVA), and function integration and differentiation are common analytical techniques.

All of these analytical techniques should be part of the data management toolkit.

### 6.5. Summary

Using data in science involves a number of different processing steps. Any scientific data management tool adopted by Sequoia should provide services for each step. At SIO, researchers commonly must convert data to manageable types, transform it to a reference system with which they can work, visualize the data, and then analyze it. Current systems make preprocessing, visualization, and analysis entirely distinct. Sequoia should try to integrate these steps, so that they could all be done using a single toolkit.

## 7. Nine Visualization Benchmarks

Warren White produced a list of nine visualization benchmarks for Sequoia to run in the short term. These benchmarks are of strong interest to researchers at SIO, and are likely to be interesting to researchers at other sites, as well. Researchers at other sites are strongly encouraged to elaborate on Warren's list, or to draw up a similar list of their own.

This section briefly describes the nine benchmarks for the computer science community. Climatologists already familiar with the discipline's terminology may find Warren's original document, provided as Appendix A, more interesting.

### 7.1. Maintain a Mean Annual Cycle Dataset

For a given dataset that represents some mean annual cycle (*e.g.*, equatorial rain measurements), researchers want to be able to acquire new measurements, visualize and analyze them, and incorporate them into the mean annual cycle dataset.

This requires that the new measurements be projected on some map. Differences from the mean annual cycle are produced, and these anomalies are then re-gridded, using optimal interpolation, so that they are at the same resolution as the mean annual cycle data. The anomalies can now be examined with a visualization tool.

If visualization turns up questionable values, the original measurements that produced them should be annotated by the researcher, and the questionable values discarded. Optimal interpolation and visualization repeat until the dataset is acceptable.

At this point, the processed new measurements can be added into the mean annual cycle dataset, producing a more accurate version of the cycle.

Jim Frew reports that Berkeley is working on handling the map projection problem. Warren volunteered to donate his optimal interpolation code to the project.

### 7.2. 4D Display

This benchmark requires that three-dimensional data (such as ocean bathymetry) be plotted over the globe according to some map projection. After the ocean floor is rendered, three-dimensional shapes must be drawn in the water. For example, temperature variation can be displayed geometrically. All the points at which water temperature is 6°C in a given region form an *isosurface*, and this isosurface can be drawn.

It should also be possible to animate these pictures, to watch the isosurface move over time. At any timestep, the user should be able to interrupt the animation, and slice through the displayed image to see a planar plot of values.

From this visualization, researchers will want to compute statistical properties of the data and the CEOF. The visualization tool must support sophisticated analysis.

The project should acquire high-resolution maps of political boundaries, geographical boundaries, and land topography and ocean bathymetry for the whole planet.

### 7.3. Surface-Volume Animations

This benchmark requires that volume and surface data (for example, the ocean surface and the atmosphere above it) be visualized simultaneously. Data from the two sources must be mapped to a common grid and displayed. Changes in time must then be shown in step.

An example is the relationship between heat flux at the sea surface and wind velocity and humidity in the atmosphere. In slow-moving, wet air, less sea water evaporates, minimizing heat loss at the ocean's surface.

Again, statistical analysis follows naturally from the visualization.

### 7.4. Volume-Volume Animations

This benchmark extends the last one. Just as the atmosphere and ocean surface are related systems, the atmosphere and ocean volume are. The same sorts of animations described in the previous section should be possible on two three-dimensional datasets.

### 7.5. Large Numbers of Datasets

This benchmark requires that data from $n < 12$ different sources be converted to the same grid, and that statistical analyses on them be done (for example, are different pairs strongly correlated? At what space or time scale?).

Visualizing this many datasets simultaneously is very hard; so far, we have no good way of dealing with clutter. According to Warren, it would be sufficient (and an improvement over what's currently possible) simply to tile the screen into $n$ regions, and display one dataset in each region. Animations on the datasets should happen in lockstep.

CEOF analysis on the $n$ datasets should be supported.

### 7.6. Model/Data Combination

This benchmark is similar to one proposed at the swat team's trip to Santa Barbara. Model data and observations must be used simultaneously in a single analysis and visualization. This requires gridding data from different sources in the same way, and displaying several different datasets simultaneously.

### 7.7. Model/Data Assimilation

This is an interactive process by which a researcher can refine a climatology model. First, the model is run to produce some data. Then real observations are collected. The model output and actual measurements are mapped to the same grid, and the differences between model output and observed values are displayed. From these differences, the researcher changes formulae in the model to produce answers more like the observed values.

This process can be repeated many times.

### 7.8. Climate Model Forecasting

This benchmark uses the previous one to refine a model, and then runs the model into the future. The values output from the model are a forecast. The correlation between the forecast and observed data can be computed. This technique can be used to do ‘‘hindcasting,’’ in which a model produces a forecast for some year in the past, and the model’s output is compared to measurements collected that year.

### 7.9. Designing Observation Networks

An important issue in the sciences is determining where, and how often, observations should be made. Temperatures vary little over small distances or time periods, so measuring temperature every inch or every second is not productive. Using analysis and visualization, the best scale for collecting measurements can be determined.

Using statistical analysis techniques, the space/time decorrelation scale for a given dataset can be computed. This scale can be displayed by the visualization tool. The researcher can identify the point on the curve at which changes become important, and can use that information to design a network of observing stations.

## 8. Miscellaneous Issues

### 8.1. Text

Richard Somerville expressed interest in how Sequoia will support textual data. Allison Woodruff described her research plans for him, which include an intelligent indexing technique for textual data. Richard is interested in storing text in Bigfoot, and using a smart bibliography system to construct customized textbooks for classes.

### 8.2. Documentation

SIO researchers encouraged Berkeley to produce a document describing the general philosophy and lore of the project. Such a document would include a guide to the available software, why it was chosen, and what to do about it if you don’t like it. Mike Olson agreed to contribute a high-level guide to Postgres to such a document.

Jim Frew agreed to supply a list of all data currently available on Bigfoot, and to produce regular updates as new data is loaded. The anonymous FTP directories currently contain some of this information, and should eventually become a project document.

### 8.3. Computer Systems

Jim Frew controls hardware and storage space for use by project investigators. He will come up with a formal guide to both, including some information on where programs should be run and how data storage is allocated.

## 9. Open Issues

The day after the swat team trip to San Diego ended (October 1), a meeting of the visualization group was held via PictureTel. At the meeting, the swat team was charged with investigating support by the visualization system for data lineage and computational steering. Under close examination, these two terms turned out not to have definitions.

At the upcoming swat team meetings in Los Angeles, Santa Barbara, and Davis, definitions for these terms should be agreed upon, and a proposal for their support by the visualization package should be made.

## Appendix A

## Visualization Benchmarks for Global and Regional Climate Change

This section contains the nine visualization benchmarks as they were proposed by Warren White. The benchmarks list specific applications and the names of researchers who would benefit from the benchmark. The researcher who can give more details on the benchmark is named in **bold face**.

### A.1. Map Non-Uniformly Distributed Data

Display data locations, interpolate data, form anomalies, contour data, evaluate data, re-interpolate and re-contour, combine anomalies and mean.

Specific applications: map ocean physical, chemical, and biological station data onto a uniform grid at a uniform time; place time series data at ocean buoy stations or airports or radiosond stations onto a uniform grid at a uniform time; place satellite data onto a uniform grid at a uniform time.

Interested researchers: **Norm Hall**, Warren White, John Roads, Dan Cayan.

**A.1.a.** Take any 4D data set distributed randomly over some specific domain and display the data locations on some projection (be able to post the values if desired).

**A.1.b.** Calculate the mean, variance, and covariance statistics (*i.e.*, spatial and temporal decorrelation scales and the signal-to-noise ratios).

**A.1.c.** Form anomalies fromthe mean annual cycle.

**A.1.d.** Interpolate anomaly data onto a specific uniform 4D grid using optimal interpolation, and compute the interpolation errors.

**A.1.e.** Contour or colorize the gridded anomaly values and their errors (be able to overlay with original data locations).

**A.1.f.** Open a data window on individual data locations to check questionable anomaly values, and be able to insert a quality control flag (metadata) on the original data.

**A.1.g.** Re-interpolate and re-contour anomaly data of known quality (*i.e.*, repeat steps c to f), producing quality-controlled 4D gridded fields of anomaly data.

**A.1.h.** Add the anomalies back onto the mean annual cycle.

### A.2. 4D Display

Topographic relief, render 3D volume, animate 3D volume, slice 4D volume, correlation matrix.

Specific applications: render salinity tongues in the ocean; render current jets in the ocean and atmosphere; render temperature anomalies in the ocean and atmosphere.

Interested researchers: **John Roads**, Shyh-Chin Chen, Tim Barnett, Richard Somerville, Norm Hall, Warren White, Kyozo Ueyoshi, Alejandro Pares-Sierra.

**A.2.a.** Plot the surface of the ocean or land bathymetry as flat boundaries or in relief over any portion (or all) of the globe on *n* map projections (*e.g.*, spherical, Mercator, Lambert equal-area).

**A.2.b.** Render the surface of a scalar contour in 3 spatial dimensions encompassing a volume, allowing resulting ‘‘worms’’ and ‘‘slugs’’ to be traced through the 3D space of the ocean or atmosphere system.

**A.2.c.** Be able to enter into the 3D space to examine rendered scalar volumes from a variety of different angles and zoomed distances.

**A.2.d.** If data have a time dimension (*i.e.*, the 4th dimension), allow the ‘‘worms’’ and ‘‘slugs’’ to ‘‘burrow’’ (extending the metaphor) through the 4D space of the ocean or atmosphere system.

**A.2.e.** Display a contour of a 2D slice through the 4D space-time volume.

**A.2.f.** Compute the auto-covariance matrix, and display the complex EOF.

### A.3. Surface Effect to Volume Response

Surface contour, render 3D volume, join fields, animate join fields, cross-correlation of surface with volume.

Specific applications: Map SST anomalies over the global ocean, with air temperature anomalies rendered over the entire globe; map wind stress curl over the North Pacific Ocean and render the subsurface temperature anomalies below it; map snow cover anomalies over the Himalayas and render the air temperature anomalies above the entire globe.

Interested researchers: **Richard Somerville**, Nicholas Graham, John Roads, Tim Barnet, Shyh-Chin Chen, Kyozo Ueyoshi, Alejandro Pares-Sierra.

**A.3.a.** Plot the contour or relief of surface effects (*e.g.*, SST, radiation flux, primary productivity, snow cover, ice cover, biomass, popualtion density, etc.).

**A.3.b.** Render the 3D volume response, either below the surface in the ocean or above the surface in the atmosphere.

**A.3.c.** Insert the contoured surface and the rendered volume on the same window.

**A.3.d.** Animate the surface and the 3D volume above or below it.

**A.3.e.** Compute the cross-covariance matrix between the surface and the volume either above or below it, and display the multi-variate CEOF.

### A.4. Ocean Volume and Atmosphere Volume Interaction

Render ocean 3D volume, render atmosphere 3D volume, join the two fields, animate ocean-atmosphere interaction, cross-correlation between ocean and atmosphere.

Specific applications: Animate subsurface termperature anomalies in the global ocean and air temperature (or clouds, moisture content, etc.) anomalies in the global atmosphere above it to see how anomalies in one fluid medium influence those in the other.

Interested researchers: **Nicholas Graham**, Tim Barnett, Warren White, John Roads, Kyozo Ueysohi, Alejandro Pares-Sierra.

**A.4.a.** Render 3D volume of some oceanic anomalies over some portion (or all) of the global ocean on one of *n* projections.

**A.4.b.** Render 3D volume of some atmospheric anomalies over some portion (or all) of the global ocean on the same projection.

**A.4.c.** Merge the two rendered volume anomalies over the same or different portions of the globe.

**A.4.d.** Animate the rendered volumes in both fluid media.

**A.4.e.** Compute the cross-covariance matrix between oceanic and atmospheric anomalies, and display the multivariate CEOF.

### A.5. Multiple Interactions

Define *n* 2D datasets, normalize *n* 2D datasets, cross-correlation statistics, multivariate principal components.

Specific applications: Examine the interaction between ocean temperature anomalies off the coast of California, wind-air temperature/cloud-precipitation anomalies over California, snow cover and biomass anomalies on the land of California, and river runoff and reservoir capacity anomalies.

Interested researchers: **Dan Cayan**, Tim Barnett, Nicholas Graham, Warren White, John Roads.

**A.5.a.** Define $n < 12$ datasets of gridded 2D data over some portion (or all) of the globe.

**A.5.b.** Normalize the datasets and display ANOVA statistics (including decorrelation space and time scales) in a table with *n* entries.

**A.5.c.** Compute the cross-covariance matrix between the *n* datasets, and display the possible interactions (*i.e.*, maximum cross-correlation at *x* space lag and *y* time lag) in an $n \times n$ matrix.

**A.5.d.** Compute the multivariate CEOF between the *n* datasets, and display the principal components in a table with 2*n* spatial patterns and one time sequence.

### A.6. Model-Data Intercomparison

Define model and observed datasets, difference model and observed datasets, cross-correlation statistics of differences, multivariate principal components of differences.

Specific applications: examine the ability of an ocean model to simulate ocean temperature anomalies off the coast of California measured from AVHRR; examine the ability of an ocean-atmosphere coupled model to simulate the observed upper ocean temperature and lower atmosphere winds over the global equatorial wave.

Interested researchers: **Tim Barnett**, Nicholas Graham, Alejandro Pares-Sierra, John Roads, Kyozo Ueyoshi.

**A.6.a.** Define model and observed datasets in 4D over some portion (or all) of the globe.

**A.6.b.** Difference the model and observed datasets and display ANOVA statistics (including decorrelation space and time scales).

**A.6.c.** Compute the cross-covariance matrix between the model and observed datasets, and display possible interactions (*i.e.*, maximum cross-correlation at $x$ space lag and $y$ time lag).

**A.6.d.** Compute the multivariate CEOF between the model and observed datasets, and display the principal components in a table with $2n$ spatial patterns and one time sequence.

### A.7. Model-Data Assimilation

Operate prognostic climate model, define model and observed datasets, interpolate model data to location of observed data, difference model and observed data, decorrelation statistics of differences, interpolate differences to model grid, update model first-guess field.

Specific applications: Conduct model-data assimilation on a global ocean model using observed XBT temperature data collected over the ocean by the Global Ocean Observing System (GOOS).

Interested researchers: **Warren White**, Tim Barnett, Nicholas Graham, John Roads.

**A.7.a.** Load into memory a community prognostic climate (coupled) model, integrate the model, and display $n$ fields.

**A.7.b.** Define model datasets that will be updated by observed datasets.

**A.7.c.** Interpolate the model data to the location of non-uniformly distributed observed data.

**A.7.d.** Difference the interpolated model data with the observed data, and display decorrelation statistics (including the signal-to-noise ratio).

**A.7.e.** Interpolate the model-data differences onto the model grid using optimal interpolation.

**A.7.f.** Update the first-guess field of the model with model-data differences under the nudging criterion.

**A.7.g.** Integrate the model over the time period coincident with the observed data.

### A.8. Climate Model Forecasting

Operate prognostic climate model, assimilate observational datasets, integrate the model into the future, determine skill of forecast.

Specific applications: Conduct hindcasting over the global ocean-atmosphere system for the period 1972 to 1992, using a coupled model assimilated with subsurface temperature and surface wind stress.

Interested researchers: **Tim Barnett**, Nicholas Graham, Warren White.

**A.8.a.** Load into memory a community prognostic climate (coupled) model, integrate the model, and display $n$ fields.

**A.8.b.** Call the model-data assimilation recipe, yielding A.7.e above.

**A.8.c.** Integrate the coupled model from model-data assimilated state into the future in the absence of observed data, and display *n* forecast fields at *m* lag times.

**A.8.d.** Compute the cross-correlation of observed data with the forecasted fields.

### A.9. Observing System Network Design

Decorrelation scales, dynamical decorrelation scales, assimilate observational datasets, integrate the model into the future, determine skill of forecast.

Specific applications: Design an upper ocean temperature observing system over the global ocean in support of the Global Ocean Observing System; design a wind vector measuring buoy array along the coast of California in support of the Coastal Ocean Program.

Interested researchers: **Warren White**, John Roads, Alejandro Pares-Sierra.

**A.9.a.** Examine fields of decorrelation time-space scales and signal-to-noise ratios of differences from A.7.d above (or from the anomalies in A.1.f above).

**A.9.b.** Calculate the interpolation error as a function of data density, given the decorrelation scales and signal-to-noise ratios, for every location where the decorrelation scales are given.