A Simple Research Paradigm in the Context of the Sequoia 2000 Project and its Application to an Ocean-Atmosphere Interaction Study

D. E.Waliser[1], C. R. Mechoso[1], C. Gautier[2] , J. D. Neelin[1]

[1]Department of Atmospheric Sciences
University of California, Los Angeles
Los Angeles, CA 90024

[2]Earth Space Research Group
University of California,Santa Barbara
Santa Barbara, CA 93106

Sequoia 2000 Technical Report 93/34
University of California
Berkeley, CA 94720

June 1993

Abstract

This paper presents an application of a common research paradigm that can help enhance and facilitate the conceptual interaction between the research goals of climate and global change researchers and the design and implementation goals of the computer scientists and engineers. As presented, the paradigm fits neatly into the Sequoia 2000 architecture, and can be applied across scientific disciplines and to all levels of scientific research, from the program level to the detailed analysis level. Each of these aspects are discussed, and example applications of the paradigm to the area of ocean-atmosphere interactions are given, including a detailed application to the analysis of evaporative heat flux parameterizations for ocean general circulation models.

1. Introduction

In almost all cases of scientific research, scientist(s) start with a question and search for the answer through examination of physical observations of the phenomena and by attempting to model the system using analytical or numerical techniques. In the Earth sciences, methods of analyses, and even the expertise of the scientists themselves, can be broadly divided into three categories: observational, modeling and theoretical. The boundaries between these three methodologies are, however, somewhat blurred in present days. It is quite common for individual scientists to cross over these boundaries, and for example, perform modeling experiments to explain their observations, use observations to initialize models, or examine observations to validate their models or theories, etc. This intermingling of methodologies requires an interdisciplinary and collaborative approach to investigations on Earth science problems, and consequently a stronger emphasis on the availability and exchange of resources, such as data, models, computer systems, analysis tools, etc. To make matters more challenging, present day methods of observational analysis, modeling, and theoretical work can, in many cases, readily saturate the available computational, data storage, networking and visualization resources. The Sequoia 2000 Project [STON92, DOZ92] addresses these challenges by improving the computational and collaborative aspects of the research environment.

The principle collaborators in the Sequoia 2000 Project are climate and global change (CGC) researchers and computer scientists and engineers. Within this inherently interdisciplinary project, the goals of the climate and global change (CGC) researchers are to increase our basic understanding of the global climate system and address important questions regarding climate change and predictability on time scales of seasons to decades. Often these investigations require vast amounts of satellite and in-situ data, and/or application of complex climate models. These models generally include aspects of the atmospheric and ocean circulations, along with many complex chemical and biological subprocesses. The goals of the computer scientists and engineers are to improve and integrate the components that make up the scientific computer environment used to support CGC research. These components consist of distributed computer systems, high speed data networks, state-of-the-art visualization systems, and data base systems capable of managing multi-terabyte storage systems.

This paper illustrates a simple research paradigm that can help facilitate the integration of CGC research methodologies into the architecture of the Sequoia Project. The paradigm is described in Section 2, and discussed in the context of the Sequoia 2000 system in Section 3. An example application of the paradigm to an analysis of evaporative heat flux parameterizations for ocean general circulation models (GCMs) is presented in Section 4. A brief summary is presented in Section 5.

2. Research Paradigm

A scientific investigation usually proceeds from a basic question to the application of an analysis technique, which in turn leads to a result. Progress in science is generally made because this result generates additional inquiries, which are then investigated through further analyses, and consequently expands knowledge as it generates new, or more comprehensive, results. This scenario leads to the following research paradigm in which a scientific *question* leads to an *analysis*, which in turn leads to a *result,* which then spawns a new *question.* This paradigm, which will be referred to as QAR, is represented in Fig. 1 as a circular flow-chart with three "nodes" labelled Question, Analysis and Result, in that order. Rather than suggesting a closed "flow", the paradigm should be considered "open", in that it can be initiated from any "node" in the system. For example, by learning about the results published in the most recent literature, a scientist may generate a new question requiring a more extended analysis, or, presented with a new analysis tool, a scientist may be able to find the answer to a question already under investigation.

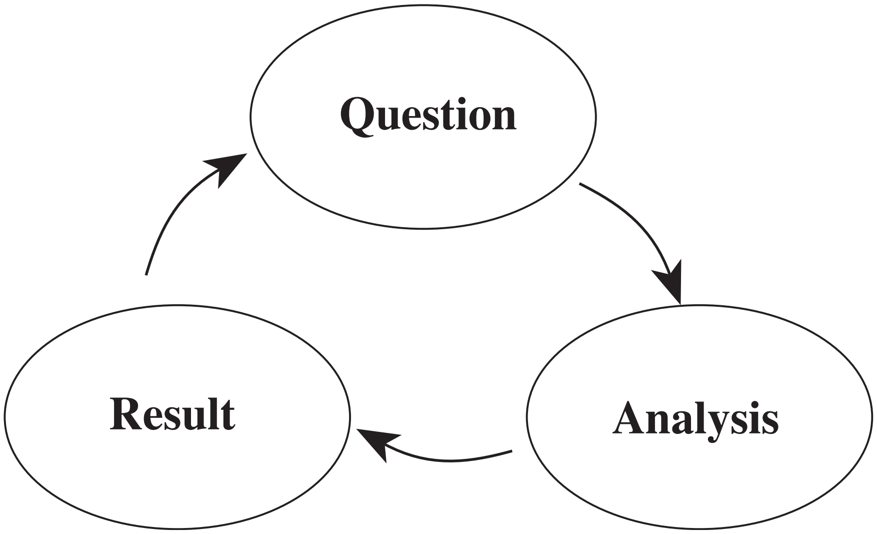

Figure 1. Schematic depiction of the QAR research paradigm.

QAR is a very simple construct, and because of this simplicity, its applicability extends not only across scientific disciplines but over all levels of organization that may encompass a scientific endeavor. This latter aspect is indicated by the diagram in Fig. 2, which is a simplified representation of the hierarchical organization of scientific research. This diagram is intended to demonstrate how a large scientific program is generally built upon the collaborative efforts of a number of investigators, and/or the succession and interrelation of a number of smaller core projects. The efforts of individual investigators, in each of these cases, usually amounts to studying one particular aspect of the problem related to their selected area of interest. The results of this research are written up in a scientific paper, or incrementally in several papers. Each of these studies, or papers, almost always consists of a significant amount of "detailed" analysis.

To illustrate how QAR applies to the hierarchy of scientific research shown schematically in Fig. 2, we next present an example from the area of climate research. The example begins at a high level of scientific organization, with the Tropical Ocean Global Atmosphere (TOGA) program. This program was designed to understand and characterize the coupling between the tropical ocean and the global atmosphere, particularly with respect to the variability associated with the El Nino - Southern Oscillation (ENSO). In short, it addresses the question: "What is ENSO?" TOGA's analysis phase constitutes 10 years of research, from 1985 to 1994, and has led to several major results such as: 1) the mechanisms which lead to large-scale tropical ENSO variability are deterministic enough that operational ENSO forecasting is possible with lead times of at least six to twelve months; 2) interannual variability not well understood may be due to interactions with other climate components such as the extratropics, land surface conditions or the cryosphere. These results, among others, have already motivated new questions that are spawning a follow up program called Global Ocean Atmosphere and Land System (GOALS). The charter of this program asks, "What is the role of the global surface boundary conditions (sea surface temperature, sea ice, snow cover and land) in determining the seasonal, interannual and interdecadal variability of the Earth climate system?"

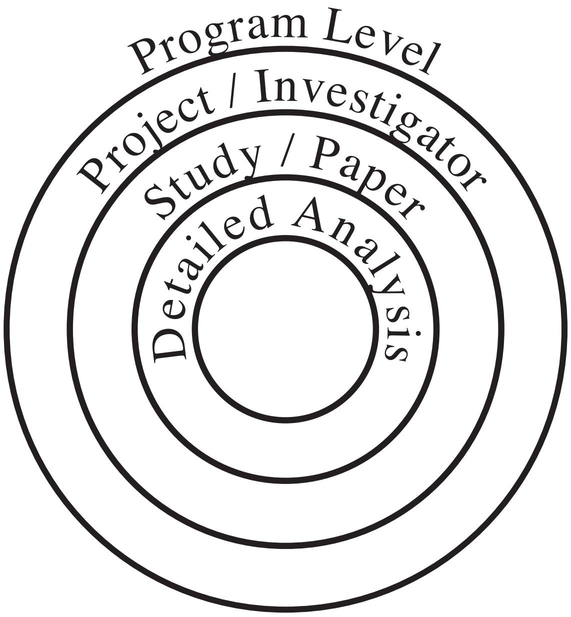

Figure 2. Successive levels of applicability of the QAR research paradigm.

Underneath the programmatic level of TOGA, many researchers are collaborating, each contributing his/her area of expertise. This expertise is usually directed towards a more specialized question that has applicability to the overall program but still may take years to find a "solution" to. For example, a numerical modeler's long-term TOGA-related research objectives may be encapsulated by the following question: "What can we learn about ENSO from coupling an atmospheric model to an ocean model?" Likewise, the research of a scientist specializing in observational analysis may be directed by the question: "What can satellite remote sensing tell us about the structure and variability of the observed heat flux between the ocean and atmosphere?" Each of these questions can support years and years of analysis, that is either refined or generalized with new results and answers. The sum total of the investigators' results adds up to the result, and thus the measure of success, of the total program.

"Investigator" level questions, such as those given above, usually break down into a lower-level set of "study" questions. For example, in the case of the numerical modeler, the first "study" level question might be to ask: "Is it possible to couple the simplified atmospheric model A to the complex ocean model O?" If the analysis, *i.e.* construction of the coupled model, does not produce reasonable results, then of course the next question is likely: "Why?". If the coupling appears to be "successful", then the next logical question might be: "What are the intrinsic modes of 'oscillation' within this coupled system?" Once this modal decomposition is performed, and the results reviewed, the next question would probably be: "How does the modeled system compare with the observed system?" The results of this comparison analysis may indicate that several aspects of the model simulation are inconsistent with observations. This would indicate that some of the assumptions or approximations in the models, or their coupling, are not valid or warrant improvement. This could lead to numerous questions concerning the detailed treatment of the model system, each of which would likely require a self-contained "study" in itself. One such study concerning the specification of the evaporative heat flux between this hypothetical coupled ocean and atmosphere model is examined more fully in Section 4. This "actual" investigation is presented to provide a more concrete example of the role the QAR paradigm can play in conceptualizing the interaction between the scientist and the computational resources at his/her disposal.

3. Contextual Place in Sequoia 2000

The previous section described how the QAR paradigm conforms to the methodologies of scientific research, particularly with regards to CGC studies. This section briefly describes how QAR fits into the Sequoia 2000 architecture, and how it may help to clarify the interactions between the hardware/software components of the Sequoia system and CGC research methods.

Figure 3 schematically illustrates the principal links between the QAR paradigm and the Sequoia architecture. The illustration indicates that the question(s) posed by the CGC researchers are primarily generated from results presented in published literature, discussed at meetings, or displayed via computer visualization. When acted on, these questions generate new analyses involving the computational, database and networking aspects of the proposed Sequoia system. These analyses can build on previous models and results stored in the database, or generate new models and results. The results from an analysis phase can take one of several forms, and often the maturity of the study dictates the most appropriate form. For example, a preliminary result is likely be output only to a computer monitor. This result would likely motivate new questions such as: "how dependent is this result to my simplifying assumptions", or "how does the solution change if I set β equal to 0.1 rather than 1.0?" Such questions would initiate new analyses and consequently produce new results. At an intermediate level of maturity, the results may be put into hardcopy or other transportable form that can be discussed with colleagues, taken to scientific meetings, or put into drafts of scientific papers. The resulting discourse may precipitate questions about the generality or sensitivity of the results, and thus spawn additional analyses and results. Finally, when the scientific work has reached a "final" stage, the results are incorporated into a scientific document for publication.

There are a number of important subtleties that are worth noting about the suggested interactions in Fig. 3. First, as indicated above, it is important that the "visualization" component of the Sequoia system be able to provide output of scientific results at many "maturity" levels. These usually begin with rather plain computer graphics or hardcopy output and extend to publishing quality layouts. It should not be overlooked that having high quality, visually intuitive, black and white output of CGC results is often more important than having it in color. However, since CGC results are often of a multi-dimensional nature, such black and white output can be more difficult to produce. Second, nearly all types of results, regardless of their format, should have the potential for being stored into the database system. This should be true of images, plotting data and formats, postscript files containing text and figures for seminar discussions, conference proceedings and published literature. Third, it is crucial that the software used to do the analysis be archived along with its history of modifications. In a noteworthy example, almost all climate models of a sufficient complexity are constantly undergoing changes, from minor "tuning" to major design changes. These changes need to be recorded and be trackable, so that previous results can be reproduced. Thus it is important to consider the processing history of models as "data" as well. Finally, it should be possible to archive the entire set of constituents from an analysis or project. Since a given study may involve a number of models, versions of models, their results, and a multitude of other data sets, it must be possible to "freeze" this configuration for later retrieval. Thus, the DBMS must also function as a "scientific notebook" for these very large, long-term, and complex investigations that typically characterize CGC research.

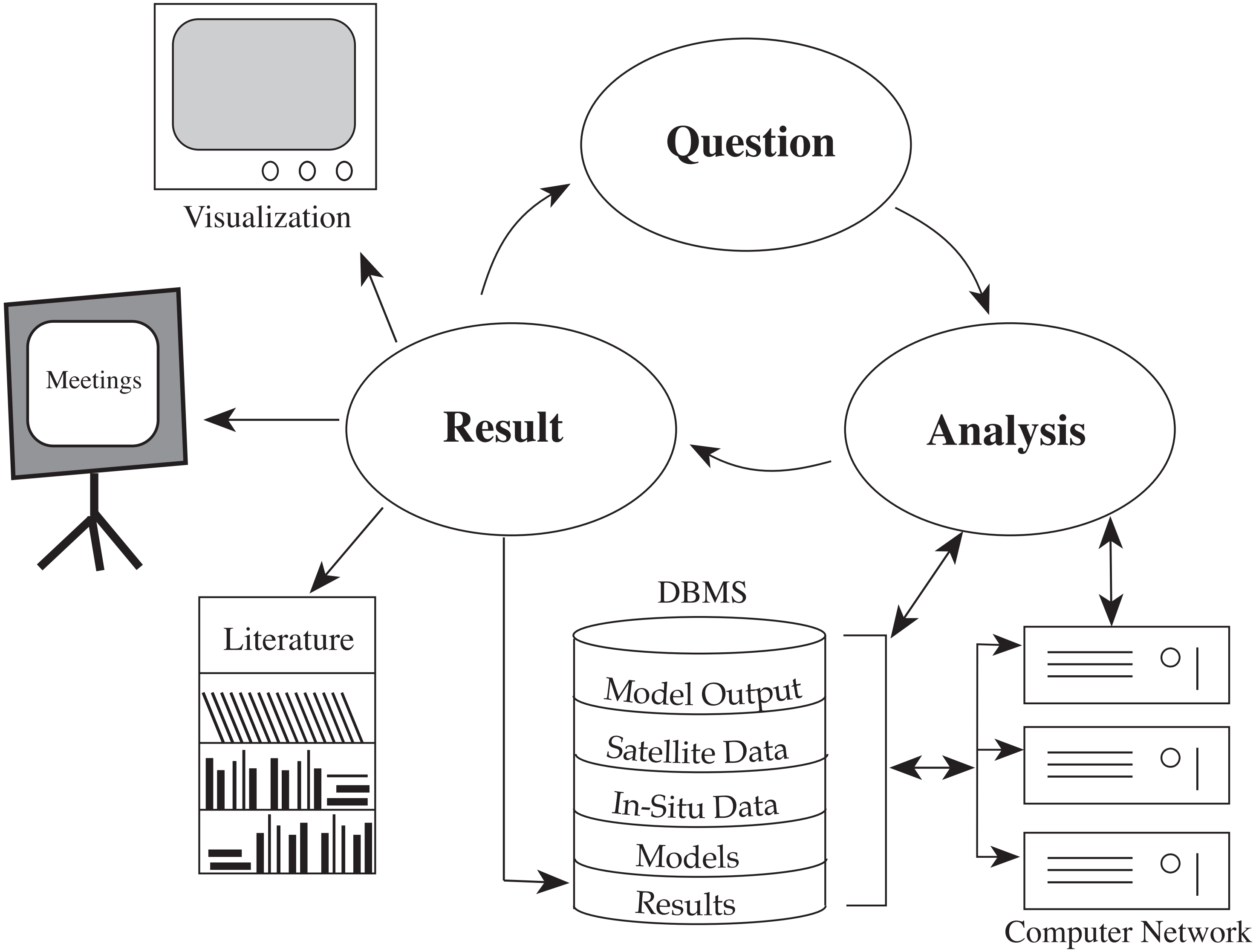
Visualization
Question
Meetings
Result
Analysis
DBMS
Literature
Model Output
Satellite Data
In-Situ Data
Models
Results
Computer Network

Figure 2. The QAR paradigm within the framework of the Sequoia 2000 Project.

4. Application of QAR to an Ocean-Atmosphere Interaction Problem

The example applications of QAR presented in Section 2 are primarily directed at the higher levels of research organization, such as the program level or the project level, and less attention was given to the detailed analysis level. The previous discussion concerning the links between the paradigm and the Sequoia system, however, makes it apparent that the benefits of the conceptual melding of these two systems may be greatest at the detailed analysis level. It is this level where scientific research can rapidly proceed through many cycles of the question, analysis and result model. In fact, limitations on scientific progress are primarily due to the speed at which these "cycles" can be performed. In the present environment, this speed is often limited by computational difficulties related to the analysis or impediments to producing intuitive and meaningful output. The example presented in this section is intended to demonstrate the applicability of the QAR paradigm to the specific steps within a detailed analysis. The analysis presented is an attempt to improve the parameterization of surface evaporative heat flux (EHF) used in forced ocean, and hybrid-coupled atmosphere-ocean, general circulation models that are designed to study and simulate El Nino.

GCMs are important tools for the analysis and prediction of seasonal, interannual and interdecadal variability of the Earth's climate system. Forced simulations are most often experiments in which observed atmospheric conditions are used as boundary conditions at the ocean-atmosphere interface to try and simulate the observed ocean conditions. These simulations are often used to study the evolution of an ocean phenomena of interest (*e.g.,* El Nino), analyzing and evaluating the behavior of the model itself, or for initiating ocean predictions. In the latter case, the ocean model is forced with observed atmospheric conditions (*e.g.,* surface wind stress, insolation, etc.) up to the present day, at which point an atmospheric model is coupled to the ocean to determine how the system might evolve in the future. Many current ENSO prediction models use this approach today [*e.g.,* CANE86]. In either the forced or coupled case, the heat flux between the atmosphere and the ocean (*i.e.,* surface heat flux) must be parameterized using some combination of model variables (*e.g.,* sea surface temperature), forcing conditions (*e.g.,* wind speed), and in some cases aspects of the observed climatology. The EHF is a particularly important component of the surface heat budget since it is the primary way the tropical ocean balances the incoming solar energy. Therefore, proper specification of the evaporative heat flux is crucial for obtaining useful ocean simulations.

In ocean GCMs, the evaporative heat flux is often modeled in bulk aerodynamic form as follows:

$$\text{Evaporative Heat Flux} = \text{EHF} = L\,\rho\,C_d\,|V|\,[Q_s(\text{SST}) - Q_a(\text{Ta, RH})], \qquad (1)$$

where L is the latent heat of condensation, ρ is the density of air, C_d is an exchange coefficient (which is itself dependent on many factors), $|V|$ is the surface wind speed, Q_s is the saturation specific humidity of the sea surface, Q_a is the specific humidity of the air, which is a function of the temperature (T_a) and relative humidity (RH). In the class of ocean simulations discussed above, all variables are provided in the above formulation by the models and/or the forcing conditions with the exception of the ρ, C_d, and Q_a. The air density and drag coefficient are often taken to be constants. This leaves the task of estimating Q_a, or equally Ta and RH, from the other available model variables, the forcing fields, and possibly climatology. It is important to note, that using observed Q_a is not a reasonable approach since this would be nearly equivalent to treating the atmosphere as having an infinite heat capacity, thus omitting the dependency of Qa on the SST itself, and therefore provide an unrealistic system with which to assess the ocean model performance. Thus the *Question* addressed in this example is "how can we best parameterize Q_a from the available information, such as SST, $|V|$, spatial or temporal aspects of the climatology of relevant variables, etc.?" This question would be best described as a "study" level question, in the terminology of Section 2. Our *Analysis* will involve exploring and exploiting observed relationships between the relevant variables to devise a scheme suitable for application in a ocean model.

Below is an outline of the manner in which a detailed analysis might progress for such a "study", given a realistic set of data resources and analysis tools within the Sequoia system.

Q: What data are available containing observations of evaporation over the ocean and/or the parameters required to compute it using Eqn. 1 (*e.g.*, $|V|$, SST, and Q_a)?

A: Query the data base management system (DBMS) with the proper specifications of variables, resolution, spatial domain, etc.

R: Visualization system (VS) displays the available data resources in a manner as shown in Plate 1 (color). The display indicates that four types of data resources are available for such an analysis. These include: 1) a 12 month, in-situ, annual climatology [OBER88]; 2) a 60 month atmospheric GCM simulation; 3) a 45 month satellite-derived data set; and 4) a 240 month, in-situ data set [WEAR80]. Since we desire observations containing monthly variability for our analysis, the 12-month annual climatology, model simulation, and satellite-derived fields are not appropriate for our purposes. This leaves only the monthly, tropical Pacific in-situ analysis, and since it covers the domain most strongly related to El Nino, this data set is chosen.

Synopsis: "We've found an appropriate data set, now we will use this data set to determine how well one of the already established parameterization schemes performs."

Q: How well does the scheme based on Seager *et al.* [SEAG87] simulate the observed evaporation? [The Seager scheme assumes $Q_a/Qs = \gamma$ = constant.]

A: Compute at each spatial point, the mean bias and root mean square (rms) difference over the 240 months of data of the following two fields:

a) $EHF_{obs} = L\, \rho\, C_d\, |V|_{obs}\, [Q_s(SST_{obs}) - Q_{a\ obs}]$ (observed)

b) $EHF_{param} = L\, \rho\, C_d\, |V|_{obs}\, Q_s(SST_{obs})\, [1 - \gamma]$ (parameterized)

where: $\rho = 1.2$ kg m^{-3},

$C_d = 1.25 \times 10^{-3}$,

$\gamma = 0.75$.

R: VS system displays the results, shown in Plate 2 (color). The results show that the Seager *et al.* scheme produces large long-term mean biases (on the order of 40 Wm^{-2}) when Q_a is parameterized as $Q_s * 0.75$, relative to using observed values of Q_a. These biases will produce an Eastern Pacific equatorial cold tongue that is too cold (*i.e.,* too much heat is removed from the ocean) and a trade wind zone that is too warm (*i.e.,* not enough heat is removed from the ocean). The rms errors are on the order of 20-50 Wm^{-2} with maxima in the central and southern areas of the ocean.

Synopsis: "These errors are greater than we would like, maybe a different approach would be more promising."

Q: How well will a scheme do that is based on a first-order Taylor expansion of observed EHF, taken about the annual cycle, and that approximates the derivative term according to the Seager *et al.* formulation, *i.e.,* Q_a/Qs = constant?

A: Compute at each spatial point, the mean bias and root mean square (rms) difference over the 240 months of data of the following two fields:

a) $EHF_{obs} = L\, \rho\, C_d\, |V|_{obs}\, [Q_s(SST_{obs}) - Q_{a\ obs}]$

b) $EHF_{param} = EHF_{anncycle} + L\, \rho\, C_d\, |V|_{obs}\, (d\Delta Q_s/dSST)\, [SST_{obs} - SST_{anncyc}]$

where: ρ, C_d and γ as above,

$EHF_{anncycle}$ and SST_{anncyc} are the observed seasonally varying climatologies of evaporative heat flux and SST,

$d\Delta Q_s/dSST$ is the derivative of $Q_s(SST)\, [1 - \gamma]$ with respect to SST.

R: VS system displays the results, shown in Plate 3 (color). The results show that because of the inclusion of the annual cycle climatology, there is no net bias. However, since significant interannual variability exists in the central Pacific (*e.g.,* El Nino), this schemes' "linear" expansion around the annual cycle does not adequately model the associated variability in EHF. Therefore, this scheme will do well for long-term mean and annual cycle simulations but will likely lead to poor results for simulations of interannual variability. Thus, since one of our primary goals is the simulation of El Nino, an improvement on the previous scheme which did not use the observed climatology, would be a better approach.

Synopsis: "This direction does not look promising, maybe we can improve on the previous scheme by examining its underlying assumptions."

Q: How good is the approximation that $\gamma = 0.75$; is there any SST or wind speed dependence?

A: From the observations, plot the bin-averaged value of γ for 0.5°C SST and 1 m/s wind speed bins, and fit a order three, two-dimensional polynomial to the result.

R: VS system displays the results, shown in Plate 4 (black & white). The plots indicate γ has a rather strong non-linear dependence on SST and wind speed. Lower values of γ (~0.68) are associated with high wind speeds and low SSTs, and with low wind speeds and high SSTs. High values of γ (~0.76) are associated with intermediate values of SST, particularly at lower wind speeds.

Synopsis: "Maybe if we take these dependencies into account, we can improve the Seager et al. scheme."

Q: Are the biases in the trade wind zones and cold tongue shown in Plate 2 due to the dependencies of γ on wind speed and SST shown in Plate 4? If so, would an incorporation of these dependencies in the EHF scheme provide an improved parameterization?

A: Use the functional relationship shown in Plate 4 to compute non-constant values of γ in the Seager *et al.* parameterization, and compute the mean bias and root mean square (rms) difference for the following two fields:

a) $EHF_{observed} = L \rho C_d |V|_{obs} [Q_s(SST_{obs}) - Q_{a\,obs}]$

b) $EHF_{seager} = L \rho C_d |V|_{obs} Q_s(SST_{obs}) [1 - \gamma]$

where: ρ, C_d as above,

γ depends on SST_{obs} and $|V|_{obs}$ according to empirical relationship shown in Plate 4.

R:VS system displays the results, shown in Plate 5 (color). Most of the biases have been reduced except for the equatorial cold tongue region. Thus, some of the biases in the parameterization may be due to dependencies not considered in the γ ratio. These dependencies most likely associated with the annual cycle since they occur away from the equator (where the annual cycle is strongest). The RMS shows little improvement, with values still on the order of 25-50 Wm^{-2}.

Synopsis: "The improvement is (incrementally) great enough that it is worthwhile to see if the modified scheme will improve our ocean simulations."

Q: How well would the current parameterization perform in an actual model simulation?

A: Use the above scheme to calculate evaporative heat flux in a new simulation with the ocean GCM. Call this simulation GCM001.EXP. Set all other parameterizations and "switches" to those used for the base case experiment GCM000.EXP (*e.g.,* case with "old" scheme). Compare the long-term mean SST from GCM000.EXP and GCM001.EXP, and compute the bias and rms difference between the monthly averaged SST fields.

R: This hypothetical simulation has not been performed, however, the above investigation is still under way.

While the ocean simulation described above has not been performed, the subsequent analysis of the numerical results would proceed in a similar pattern. Most often, analysis of El Nino simulations consists of producing a common set of result "maps", such as the long-term mean SST, a time-longitude diagram of SST anomalies at the equator, the leading modes in an empirical orthogonal function analysis, etc. These would then be compared to the same maps from observations or the "base" case, if the simulation was forced, or compared to "base" case maps, if the simulation was from a coupled experiment. It is likely that the results from this new GCM simulation would suggest further questions, requiring additional analysis of the data, or a new simulation. For a more detailed discussion of the specialized tasks and requirements unique to the processing and analysis of GCM modeling simulations, see WEIB93.

5. Summary

This paper illustrates how the use of a common research paradigm can help facilitate the integration of CGC research methodologies into the architecture of the Sequoia 2000 Project. Due to the paradigm's simplicity, it can be applied across scientific disciplines and at many organizational levels of scientific research. Further, as presented, the paradigm exhibits a complimentary melding with the high-level hardware and software components of the Sequoia system. Example applications of the paradigm to several levels of research is given, including a detailed example of the role this paradigm can play in a typical, low-level climate research study related to the numerical simulation of El Nino. We hope this presentation will help facilitate the interaction between the research goals of the climate and global change researchers and the design and implementation goals of the computer scientists and engineers.

References

[CANE86] Cane, M.A., S.E. Zebiak and S.C. Dolan, Experimental Forecasts of El Nino, *Nature,* **321**, 827-832 (1986).

[DOZ92] Dozier, J. How Sequoia 2000 Addresses Issues in Data and Information Systems for Global Change. *Sequoia 2000 Technical Report 92/14,* University of California, Berkeley, CA (1992).

[OBER88] Oberhuber, J.M. An Atlas Based on the COADS Data Set. Max Planck-Institut fur Meteorologie, Report No. 15, Hamburg, FRG (1988).

[STON92] Stonebraker, M. An Overview of the Sequoia 2000 Project. *Proceedings 1992 COMPCON Conferences,* San Francisco (1992).

[SEAG87] Seager, R., S.E. Zebiak and M.A. Cane, A model of the Tropical Pacific Sea Surface Temperature Climatology, *J. Geoph. Res.,* **93**, 1265-1280 (1987).

[WEIB93] Weibel, W, J.D. Neelin and H.H. Syu. An End-to-End Processing Scenario for Hybrid Coupled Ocean-Atmosphere Model Data Under the Sequoia 2000 Project.*Sequoia 2000 Technical Report,* University of California, Berkeley, CA (to be submitted).

[WEAR80] Weare, B.C., P.T. Strub and M.D. Samuel. Marine Climate Atlas of the Tropical Pacific Ocean. University of California, Davis, Contribution No. 20. (1980).