ARCHITECTURE OF THE MERCURY MESOSCALE METEOROLOGICAL DATA FUSION

C. Fields, C. Cavendish, M. Coombs, T. Eskridge, R. Hartley, H. Pfeiffer, and C. Soderlund

Knowledge Systems Group, Computing Research Laboratory
New Mexico State University, Las Cruces, NM 88003-OWI USA

S. Kirby and G. McWilliams

U.S. Army Atmospheric Sciences Laboratory
White Sands Missile Range, NM 88002-5501 USA

1. INTRODUCTION

The MERCURY mesoscale meteorological data fusion system is being developed as an intelligent interface between the U.S. Army's Integrated Meteorological System IM[ETS) mesoscale database and tactical decision aids (TDAs) that require meteorological data as input (McWilliams or al., this volume). MERCURY addresses, at the mesoscale level of resolution, the operational problem that meteorological data are often needed for locations from which data have not been obtained, or from which data cannot be obtained. Ile task of MERCURY is to generate an optimal set of values of relevant meteorological variables to describe a given location within a mesoscale region of interest, given a database of measurements made at other locations within the region of interest, synoptic objective analyses, numerical predictions, and regional climatology. Ile long-term MERCURY specification calls for up to 6 hour nowcasting capability; the near-term goal is accurate fusion of current data.

The task environment in which MERCURY will operate imposes a number of fairly severe requirements. First, MERCURY must operate autonomously. Decisions made by MERCURY will be neither guided nor chocked by an expert meteorologist, the results obtained by MERCURY will, in the operational environment, only be available to users via the user interfaces of the client TDAs, which will in general be used by nonmeteorologists. MERCURY cannot, therefore, rely on expert input to resolve ambiguities or otherwise deal with unexpected features of the data environment. Second, MERCURY must be usable in any geographical area. 'no design of MERCURY cannot, therefore, impose any a priori restrictions other than size on the choice of the mesoscale domain in which it is applied. Third, the types, quality, and geographical extent of data available to MERCURY will be highly variable. MERCURY must be capable of deriving required meteorological variables from very sparse or otherwise nonoptimal data, and must be capable of dealing with relatively large data voids within the region of interest. Finally, MERCURY will be required to serve multiple client TDAs with different data requirements in near real time. Ile design philosophy being employed in the MERCURY projects reflects the tradeoffs imposed by these requirements between resolution and accuracy on the one hand, and autonomy, portability, and response time on the other.

An open, heterogeneous data analysis architecture incorporating approximate numerical models, qualitative models, heuristic rules, and mesoscale objective analysis has been developed to meet the above requirements. no architecture is designed to use the faster, more reliable quantitative models when input data to drive them arc available, and to fall back on a combination of qualitative modeling and climatology when data are unavailable. Heuristic metarules am used to direct the flow of control to either quantitative or qualitative models. A geotopographic database provides high-resolution terrain and land use data for the region of interest to the models, and supplies the underlying coordinate system to which data are anchored.

MERCURY has been under development since August, 1987. Due to uncertainty about the feasibility of meeting all of the stated requirements with available computing technology, a prototyping strategy, in which both specification and design are allowed to evolve as successive prototypes are developed, was chosen for the project. A 300 km x 300 km region surrounding the Los Angeles basin for which a dense set of surface and upper-air data is available has been selected for testing MERCURY prototypes. This region includes coastline, mountains, and desert, and exhibits a number of mesoscale phenomena of interest. The current development effort is focussed on constructing a prototype capable of data fusion in this domain, with the requirement of nowcasting capability set

This work was supported by U.S. Army Atmospheric Sciences Laboratory contract #DAAD07-86-C-0034 to New Mexico State University. The views, opinions, and findings contained in this report are those of the authors, and should not be construed as an official Department of the Army position, policy, or decision unless so designated by other documentation.

aside until the data fusion capability is achieved. Initial prototype versions of most of the system components have been implemented; some components are now in the second round of prototyping. no system is currently distributed between a Symbolics 3650 running Genera 7.1 and a Sun 3 running Unix via a TCP message passing service (Soderlund, 1988). The evolution of the MERCURY design and prototype over the last year is described in Fields (1988).

This paper describes the MERCURY architecture as currently designed. While MERCURY is being developed to meet specific Army requirements, the architectural principles being employed may have wider application in geographic information systems, mesoscale data fusion and nowcasting systems, and general data interpretation in the environmental sciences.

2. OVERVIEW OF THE ARCHITECTURE

A data flow diagram of die MERCURY architecture is shown in Fig. 1. MERCURY accepts input from meteorological data and derived products sources, and from a number of databases. It accepts requests from, and returns answers to client TDAs.

The development version of MERCURY includes a developer interface for set-up and testing. In the operational version, all user interaction with MERCURY will be via the user interface from which the client TDAs are accessed. This user interface will be external to MERCURY. In the operational setting, as currently planned, user interaction will be limited to informing MERCURY of the coordinates of the region of interest; all other MERCURY activities will be autonomous.

Current meteorological data from measurement sites within the region of interest flow from a data source to a meteorological data representation that is based on the MERCURY geotopographic representation. no data source is assumed to perform error checking, and to format the data in a specified way. Only mesonet ground station and rawinsonde data are ingested and used by the current MERCURY prototype; capabilities for using data from profilers, Doppler radars, remotely piloted vehicles, and various satellite instrumentation are planned in future versions of the system.

Synoptic objective analyses am provided to MERCURY by the derived products source. Analyses am provided as gridded fields of real numbers representing heights at both mandatory and significant levels, temperatures at 850, 700, and 500 mb, and vortices at 500 mb. Gridded fields representing numerical predictions will be added as an additional derived input when nowcasting is attempted.

Data and derived products for input to MERCURY are currently obtained from the domestic data plus (DD+) and NMC numerical products broadcasts of Zephyr Weather Information Service, Inc. The Local Data Manager (LDM) component of the Unidata System for Scientific Data Management (USSDM) software developed by the UCAR Unidata Program Center (Campbell and Row, 1988) is used to ingest and manage both data and derived products. This software runs on the Sun 3.

MERCURY is demand driven by its client TDAs. Client TDAs query MERCURY by passing a request for data for a particular location and time to the client TDA interface (Fig. 1). The location must be in the current region of interest, and the time must be within a specified interval from the current time. Data requirements of client TDAs are provided in a database; the TDA does not, therefore, have to explicitly request values for the variables that it needs. Ile client TDA interface requests values of the variables requireA by the TDA from the data analysis system, which produces the required values as output. The client TDA interface reformats this output, using formats specified in the TDA requirements database, before passing the result to the requesting TDA as an answer.

3. DATA REPRESENTATIONS

The organization of the geotopographic and meteorological data representations is shown in Fig. 2. The two representations effectively form a set of overlays on a latitude-longitude coordinate system. All dam or derived products associated with a latitude-longitude point may be accessed by the data analysis system via the coordinates of the point. Data may also be accessed by specifying a distance, or a function of distance, from a point. The data analysis system thus treats the geotopographic and meteorological data representations as a single database of point data tied to coordinates.

Point elevation and land use data at 100 in nominal grid resolution (for midlatitudes) are provided by Defense Mapping Agency (DMA) digital terrain databases. DMA land use codes are converted to roughness estimates using average roughness values for desert, forest, agricultural, urban, and marine areas (Hansen, 1984). The elevation data are contoured using the NCAR Graphics contouring routine; both elevation contours and land use region boundaries may be displayed via the development user interface.

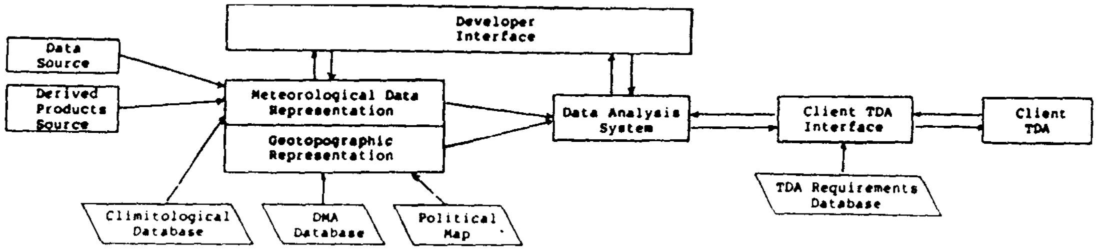

Fig. 1: Block data flow diagram of the MERCURY architecture. Diamonds represent databases. The data representations and data analysis system are shown in more detail in Figs. 2 and 4, respectively.

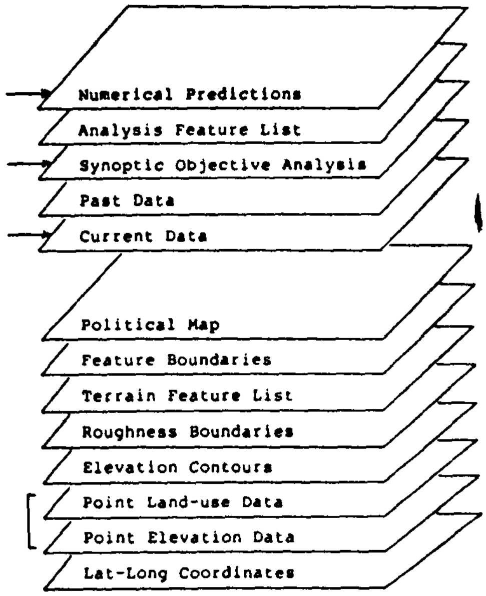

Fig. 2: Organization of the meteorological data and geotopographic representations. Each layer is tied to the coordinate system. Incoming arrows indicate input from the data and derived products sources; the DMA database is indicated by a square bracket. Other layers are calculated from those below them as described in the text.

The local slope of the terrain at each grid point is calculated as the cross product of vectors representing the changes in elevation from the grid point to its nearest neighbors in positive latitude and longitude, Regions bounded by slopes above a specified cutoff am identified as significant terrain features, and are placed in a terrain feature list. Such regions are represented by their boundaries at full resolution, and by the slopes associated with the points inside the boundaries, which are averaged on a 500 m nominal grid. This averaging procedure results in a mesoscale, as opposed to microscale, representation of orography, and in a factor of 25 decrease in the number of points representing slope data that must be maintained in the terrain feature list. Also placed in the terrain feature list are the boundaries, again at full resolution, of marine or desert regions above a specified size. Ile terrain feature list thus provides a compact representation of all features of the terrain that am likely to have significant effects on the mesoscale weather pattern. Boundaries of objects in the terrain features list may be displayed on the development user interface.

A map showing political boundaries is included in the representation as an aid to the developer. Figure 3 shows the map of the Los Angeles basin area chosen for system testing, with land use regions and terrain features indicated. This map was produced using the GeoFlavors software developed by Ball Systems Engineering Division, which is written in Symbolics Common Lisp and runs on Symbolics 3600 series machines. This software was used for the initial Symbolics-based prototype of the data representations, Both data representations arc currently being moved to the Sun 3, using the map and contouring facilities provided by the NCAR graphics package. The slope analysis, boundary identification, and distance measurement software all written in C, and run on the Sun 3.

Meteorological data am tied to the coordinate system as time-stamped point data associated with the location at which they were measured. The elevation and land use (roughness) of the location of a measurement station are obtained directly from the relevant overlay. Distances from measurement stations to terrain features are calculated as straight-line distances to their boundaries. Gridded objective analyses and numerical predictions are represented as additional time-stamped point data overlays on the coordinate system. Only most-recent data and objective analyses are used in the current prototype; selected past data will be stored in future enhancements that use numerical predictions. The pressure heights from the synoptic objective analysis are further analyzed using a slope analysis procedure similar to that employed to identify orographic features. Regions in the height fields having slopes above a specified value are stored in an analysis feature list similar to the terrain feature list. This list provides a rudimentary representation of synoptic features that may be expected to dominate or interact with mesoscale effects in the region of interest. This procedure, if successful, will also be employed for predicted heights. Climatological data will be represented as point data for given locations. Details of the form of the climatological data to be used have yet to be determined.

Figure 3 also shows the menu layout and interaction window of the development user interface (Symbolics version). The menu facilities allow display manipulation and calls to the data analysis system; the interaction window allows the display of data files too large to display on the map. such as the mesonet station data file shown in the figure. These facilities are also being moved to the Sun, where they will be merged with those provided by the menu-driven USSDM user interface.

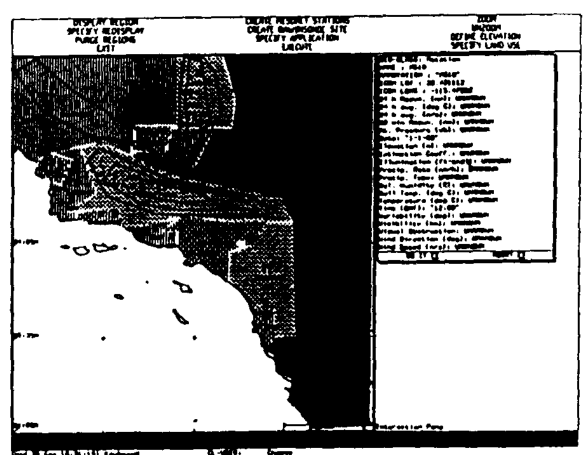

Fig. 3: Symbolics implementation of developer interface, showing map of Los Angeles basin area selected for testing, function menu, and mesonet station data.

. 4. DATA ANALYSIS SYSTEM

The most common method for estimating the values of meteorological variables at locations for which data are unavailable is to use an interpolation scheme, typically as embedded in an objective analysis procedure. Interpolation is often unreliable, however, in regions containing significant terrain features such as coastlines, mountain ranges, or deserts. The distance-weighted averaging procedures used in typical interpolation schemes arc, moreover, quite unreliable when used for extrapolation to regions for which no representative data are available.

The MERCURY task environment requires the estimation of values for variables at a mesoscale level of accuracy in regions containing complex terrain, and in which a significant fraction of the region may be void of data. Often regions of interest will be divided roughly in two, with measurements possible on one side of the dividing line and impossible on the other. The two sides of the region may be separated by complex terrain, e.g. mountain ranges or coastlines. Extrapolation using an mesoscale objective analysis procedure is very likely to be unreliable under these circumstances.

The MERCURY data analysis system employs mesoscale objective analysis, approximate numerical models, heuristic rules, and qualitative models to estimate the values of required variables under different conditions of terrain complexity and data availability. Because MERCURY must be applicable in any locale, these procedures have been designed to be as general as possible. Thus far, no region-specific approximations or heuristics are being considered in the design. AN procedures must, moreover, be autonomous. All inputs are obtained from the meteorological and geotopographic data representations directly; -there are no user queries. Response time is minimized by using the computationally- simplest procedure Judged to be reliable under the circumstances to generate a response.

The organization of the data analysis system is shown in Fig. 4. The four analysis modules each have access to the data representations, and operate independently from each other. The analysis module or modules to employ in a particular situation are determined by the metarule base, which receives the request for data from the client TDA interface. Ile use of metarules for control considerably simplifies the structure of the system, and increases the autonomy of each of the analysis modules (cf. Clancey and Bock, 1988). The metarules evaluate the data request with respect to the type of data required. the availablity of dam, and the terrain in the vicinity of the point for which data are required. If wind velocity data are required for a location near a coastline, for example, objective analysis is employed if the required values can be obtained by interpolation from coastal stations. An approximate numerical model for sea breeze penetration is employed if the available data are too few for interpolation, but if at least one measurement from a coastal station is available. Qualitative modeling is employed if no coastal data arc available.

Approximate diagnostic models for seabreeze penetrate tion, for correction of temperatures and pressures for differences in elevation, and for correction of wind velocities for differences in roughness are in the prototype stage. Heuristic rules are used to estimate the effects of synoptic features from either local upper-air data or synoptic objective analyses, to select optimal measurement sites to provide data for locations in complex terrain, and to select sites to include in objective analyses. Objective analysis is typically limited to measurement sites in similar land-use regions that have similar surrounding terrain to the location for which data are required.

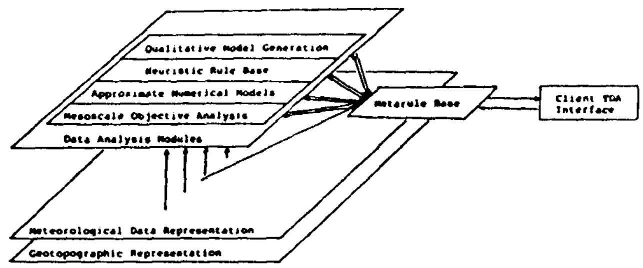

Fig. 4: Organization of the data analysis system. Metarules are used to regulate the flow of control to the various data analysis modules. All modules have independent access to the data and geotopographic representations.

Qualitative modeling will be used primarily to estimate values in subregions for which data are unavailable, or for which terrain complexity or synoptic effects disallow the use of quantitative methods. Thus far, qualitative models of meteorological processes have been restricted to expert systems for forecasting particular phenomena, usually in a limited geographic region (reviewed by Dyer, 1987). Such systems can rely on region- or phenomenonspecific heuristics. This approach is not feasible in the MERCURY task environment, in which any phenomenon may, in principle, be encountered in any region; under such circumstances, the number of individual rules that would be required grows without bound (Coombs et al, 1988; Fields and Dietrich, 1988).

An alternative to the expert-system approach is the use of nonspecific qualitative knowledge of physics together with a constructive reasoning procedure to generate qualitative models of phenomena on demand (cf. Bobrow, 1985). A flexible architecture for constructive reasoning, the model *generative reasoning* (MGR) architecture, has been developed and implemented on the Symbolics, in part motivated by the demands of MERCURY (Coombs and Hartley, 1987; Fields et W., 1988). The MGR architecture allows the generation of multiple models of a phenomenon from both qualitative and quantitative data. Model evaluation is decoupled from model generation; this allows the generation of models that may conflict with some of the data. MGR can, therefore, function in the presence of uncertain or incoherent data. MGR includes facilities for incorporating numerical functions and constraints into both model generation and model evaluation. Unsuccessful models may be decomposed into components, and these components may then be recombined to form new models. MGR thus provides a flexible automated problem solving architecture to support the MERCURY qualitative model generation module. Development of qualitative physical models of sea breeze-mountain breeze and sea breezeconvective system interactions in the Los Angeles basin region using MGR are currently in progress.

5. SUMMARY

Ile objective of the MERCURY project is to integrate a number of analytic techniques and computing technologies in a single software system for autonomous mesoscale data fusion. Ile MERCURY architecture is open, allowing the incorporation of additional modules as required to either the data representation or the data analysis system, and heterogeneous, with straightforward transfer of data and control between modules having different underlying software architectures. The implcmentation of MERCURY is in the early prototype stage. The next two years of MERCURY development and testing should reveal whether the approach that has been taken is adequate to the task.

6. REFERENCES

Bobrow, D. (ed.), 1985: *Qualitative Reasoning About Physical Systems.* Cambridge, MA: MIT/Bradford.

Campbell, D. and R. Rew, 1988: Design issues in the UNIDATA local data management system. *Preprints of the Fourth* International Conference *on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology.* AMS. pp. 208-212

Clancey, W. and C. Bock, 1988: Representing control knowledge as abstract tasks and metarules. In: L. Bolc and M. Coombs (eds.) *Expert System Applications.* Berlin: Springer. pp. 1-77.

Coombs, M. and R. Hartley, 1987: 'Me MGR algorithm and its application to the generation of explanations for novel events. *Int. J. Man-Machine Studies* 27: 679-708.

Coombs, M., C. Fields, and G. McWilliams, 1988: *Artificial Intelligence Methods for Optimizing the Use of Meteorological Databases: Recommendations for Implementing the MERCURY System.* Final Report, Contract #DAAD07-86-C-0034, WAO #87.2-10.5, U.S. Army Atmospheric Sciences Laboratory.

Dyer, R., 1987: Expert systems in weather forecasting and other meteorological applications. *AI Applications in Natural Resource Management 1:* 19-24.

Fields, C., 1988: *Artificial Intelligence Methods for Optimizing the Use of Meteorological Databases: Architecture of the MERCURY System.* Final Report, Contract #DAAD07-86-C-0034, WAO #88.2-10.2. U.S. Army Atmospheric Sciences Laboratory.

Fields, C., M. Coombs, and R. Hartley, 1988: MGR: An architecture for problem solving in unstructured task environments. *Proc. Third Int. Symp. on Methodologies for Intelligent Systems.* Amsterdam: Elsevier (in press).

Fields, C. and E. Dietrich, 1988: Engineering artificial intelligence applications in unstructured task environments. In: D. Partridge (ed.) *Artificial Intelligence and Software Engineering.* Norwood, NJ: Ablex (in press).

Hansen, F. V., 1984: Tactical smoke and chemical models Technical Report, U.S. Army Atmospheric Sciences Laboratory.

Soderlund, C. 1988: A TCP remote message passing service. *Memoranda in Computer and Cognitive Science* MCCS-88-130, Computing Research Laboratory, New' Mexico State University.