Visualization Tools for Data Assimilation

Suzana Djurcilov and Alex Pang

Computer Science Department
University of California
Santa Cruz, CA 95064

ABSTRACT

Weather forecasts are typically produced once or twice each day. Each run usually covers several forecast periods. Over the course of the day, as measured sensor data becomes available, discrepancies between observations and model forecasts are resolved and integrated so as to update and improve the next forecast run. The process of resolving the differences between model output and sensor measurements is known as data assimilation. Traditional methods include kriging and optimal interpolation. They involve statistical and historical information on reliability of sensor measurements (including desirability of sensor location, calibration, etc.), variability of the field, model resolution, initial and boundary conditions, etc.

Some of the parameters of a data assimilation model are integration techniques, choice and frequency of incremental update methods, interpolation algorithms, resolution of the model grid, estimation filtering and smoothing algorithms and finite differencing schemes. All of these parameters can have a profound effect on the tendencies displayed by a forecasting model. Hence, having visualization tools to display these possibly conflicting information is very useful for the scientists in quickly identifying regions of high conflict and/or regions of low confidence levels. Allowing the scientists to control the data assimilation variables can assist in constructing a protocol that is appropriate for a specific geographical region.

This paper presents a suite of visualization tools to aid scientists in performing their data assimilation analyses. These tools will provide an integrated display of 3D model outputs with 0D point measurements from meteorological stations, 1D measurements from wind profilers, sonde, and floating buoys, 2D measurements from CODAR current measurements and GOES satellite feed, and, when available, 3D volume measurements from NEXRAD data. The tools provided here will help extend the 2D domain in which data assimilation is currently being performed to include analysis of the overall method as well as visualizations of the separate weather measurements.

Keywords: Comparative Visualization, Differences, Kriging, Spatial and Temporal Interpolation, Multiquadrics, Forecasting

1 Introduction

The primary goal of visualization in meteorology is to provide better understanding of weather patterns in a specific geographic region. Furthermore, interactivity with a visual output allows a scientist to isolate and explore particular aspects of the environmental data. For these reasons, an array of visualization tools have been developed and are widely used for display of meteorological data. Environmental data such as those found in meteorology can be grouped into two main types:

- *Observations* - includes direct measurements gathered from meteorological sensors (e.g. pressure, wind, temperature) located on meteorological stations, floating and drifting buoys, acoustic doppler current profilers and ocean surface current measurements, radiosondes (balloons) and vertical wind profilers, instrumented aircrafts, and others. These sensors provide 0-dimensional point measurements, 1-dimensional measurements (e.g. vertical wind profilers), 2-dimensional measurements (e.g. surface currents), and 3-dimensional measurements (e.g. integrated NEXRAD radar returns).
- *Derived data* - includes numerical model output and other interpolated data. Measurements are frequently interpolated both spatially and temporally to allow 2D and 3D representations. These interpolated fields are often used to as initial conditions for forecast models.

One major difference between these two data types is the amount of data that needs to be handled. It is usually possible to display most of if not all geo-referenced observation data either as color-coded glyphs (e.g. arrows and wind-barbs),[1] or even simply as text. However, it is gridded data that is more valuable in terms of extracting features and studying the distribution of the data. Because of their density, gridded data are commonly displayed with the use of iso-contours, streamlines, volume rendering, etc. Data are also often converted into gridded form both as input and output of weather forecast models.

There has been extensive work in creating powerful visualization packages for multiple meteorological variables of model datasets: Vis5D,[2] GEMPAK,[3] WXP,[4] and Spray[5] to name a few. While these tools adequately address the need of displaying large amounts of environmental data, one important question is inevitably left out in the visualization: "What is the quality/reliability/uncertainty of the data being shown in the visualization?" In meteorology, and in general, there are several factors that contribute to the quality of the forecast. The two main factors that we consider are errors that might be introduced in the data acquisition stage, and additional errors that are introduced during the processing stage.

We see two driving motivations for developing visualization techniques for presenting meteorological data in conjunction with associated errors. Firstly, ignorance of the possible errors associated with the displayed data can lead to unreliable and misguided interpretations, Secondly, analysis of the errors found in the forecast models can help evaluate the model effectiveness, their sensitivity to parameter changes, and initial and boundary conditions. 0 are constantly struggling to produce more reliable models. With visualization tools to help analyze and improve the reliability of weather models, the general population can be better forewarned of catastrophic weather conditions and losses may be reduced to a minimum.

A measure of reliability of weather models is the correlation between the model output and the observations for the forecasted times. Visualizing the correlation (or alternatively, the difference) between the two facilitates understanding of the nature of the errors and provides additional feedback in terms of the model effectiveness. In this paper, we use the term "uncertainty" (introduced in[6]) interchangeably with the term "error".

This paper gives an overview of the problem domain and examples of practical methods that map uncertainty of model output to a variety of graphical elements. We isolate and concentrate on the error produced by the forecasting model itself, leaving aside other elements contributing to data uncertainty.

2 Background

Output from forecast models comes as 2D or 3D grids containing important physical information such as surface pressure, relative humidity, air temperature, and others. These gridded values are also used as input to weather models and forward integrated in time to produce forecasts. In general, the farther out the predictions, the less reliable they are. Typically, predictions are constantly validated and updated with actual measurements, and the models recalculate a new prediction. This repetitive process is enhanced by the use of data assimilation - a method in which observational data is statistically integrated into the forecasting process to resolve differences between predictions and observations. There are two main modes to data assimilation: batch mode and continuous mode. In the early days, data assimilation was performed at predetermined intervals where observational data was interpolated with the previous prediction to create initial conditions for the next run. In contrast, *continuous* (also known as four-dimensional) data assimilation, merges observations into the model as they become available. Both of these approaches have their strengths and weaknesses. For example, continuous data assimilation amplifies inertia-gravity wave shocks in models which can cause non-convergence of the model and data rejection,[7] i.e. data not getting assimilated because of a large discrepancy with the current state of the model. This problem is somewhat alleviated with the introduction of dynamic relaxation techniques (Newtonian relaxation or nudging[8] which can damp some of the model errors. Consider one of the standard equations used in atmospheric modeling - the continuity equation:

$$\frac{\partial h}{\partial t} + \frac{\partial h}{\partial x} + \frac{\partial h}{\partial y} + h\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = 0$$

where x and y are the eastward and northward directions respectively, u and v the eastward and northward wind components, and h the height of the surface. With the use of Newtonian relaxation this equation is transformed to

$$\frac{\partial h}{\partial t} + \frac{\partial h}{\partial x} + \frac{\partial h}{\partial y} + h * \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = -\gamma_N(h - h_0) + \gamma_D \bigtriangledown^2 (h - h_0)$$

where h_0 is the observed geopotential height, and γ_N and γ_D the Newtonian and diffusive relaxation coefficients. Both additional terms in the second equation nudge the geopotential height towards the observation.

These techniques for resolving observations with predictions are very general and can be applied to a variety of forecast models. Forecast models can range from a regional weather model that produces an output every few hours to a climate models that encompasses the planet and tries to predict the effects of the ozone layer to global warming in the next decades.[9] The techniques can also be used to compare differences arising from different models, resolution, model parameters, and assimilation cycle time.

Forecasting inherits several sources of error and uncertainty from the process of measuring and transporting weather data: e.g. improper instrument calibration, communication errors, and incorrect observation registration. As with many other mathematical models of natural phenomena, interpolation and numerical integration is used, and errors are added at each step of the way. It is also important to note that the atmosphere is a chaotic system, that is to say that a small and local change in atmospheric conditions can cause a cumulative and unexpected effect on the larger scale. It should come as no surprise then, that forecasts cannot predict accurately or with much confidence too far into the future; nor at a high resolution over a large area. It is a continuous challenge to discover, measure, and implement new and improved methods of modeling the atmosphere that will include and appropriately account for all aspects of change.

Measuring error in Model data

The importance of quantitative ways of measuring the error in measured and derived data has long been recognized in the scientific community. Historically, the analysis of the model forecast error, like in many other disciplines, has been implemented by statistical error analysis,[10],[11],[12],[13].[14] A classic case is one in which the normalized variance of the differences between the model output and the actual atmospheric condition is taken as the measure of the model effectiveness. Standard deviation and mean error provide good summary measurements of the size and spread of error, but they tell little about the nature and physical distribution of the errors. Graphical techniques often complement and present these measurements in the form of histograms and plots of error findings to further determine the distribution of errors.[15] These approaches provide excellent quantitative measures. What we are looking for are qualitative ways of describing model performance.

An important step in this direction was NCGIA's Research Initiative 7: "Visualizing the Quality of Spatial Information".[16] This initiative united the activities within the geographic information systems (GIS) community and challenged them to the issues of defining, measuring, and visualizing geospatial quality information (e.g. accuracy, precision, consistency, currency, and completeness) together with the base data. Some of the resulting work from this community focus on classification and accounting of different types of uncertainty e.g.,[17],[18].[19] More recently, the International Cartographic Association's Commission on Visualization has fostered exciting research in surface representations, e.g. map comparisons of uncertainty in cartographic data.[20] Furthermore, research has been initiated to analyze the value and "usability" of graphic depictions of data reliability.[21] However, most of these efforts have not been applied to the study of data processing in meteorology. The distinguishing feature of visualizing uncertainty of environmental data is the dynamic nature (shorter time scales) of the events compared to those commonly found in GIS applications. Also, the "error" of interest in meteorology is the difference between the forecasted and observed data, regardless of the uncertainties accumulated in the process. For this reason we decided to focus on analyzing the point-for-point error, defined as the absolute difference of point value of a model and the corresponding point on an observational grid. In order to gain true insight into the specifics of the error in a weather model area one needs a method that will visually correlate the error with its location, and also enable the the person to interactively explore the findings in a 3-dimensional fashion. Eventually, this will lead to better fine-tuning of the model in general. The goal is obviously to reduce the error by finding the choice of variables best suited for the given task of reliably predicting weather.

3 Implementation

3.1 The Model

We use the forecast model developed by Pennsylvania State University and the National Center for Atmospheric Research(NCAR), called MM5 (Mesoscale Model Version 5[8]), centered over the greater area of Monterey Bay in California. MM5 uses Newtonian relaxation in order to nudge the model towards observational data and allows the user to set several variables used in the numerical analysis. Instead of MM5's Cressman-type (distance-weighed) optimal interpolation, an alternative scheme was used to create the initial conditions. Observational data from the area was interpolated using Hardy's multiquadric interpolation (see[22] for details) from gridded data received from the National Weather Service (NWS).

Two different grid sizes where used to create our visualizations:

1. *Surface*- a 60x48 grid containing water vapor, cloud water, rain water, rain rate, near surface temperature and pressure values.
2. *Upper levels* - a 60x48x20 grid containing wind, temperature, cloud water and rain water values.

3.2 The Data

We used initial data from June 8th, 1996 at 00:00 hours to produce a 12-hour forecast. Two versions of the model were compared: one in which only the interpolated NWS data was used to initialize the grid, and another version where additional temperature and vapor soundings were added to create initial contidions. The additonal measurements were blended into the grid by using inverse distance weighing of the soundings for each grid point. The model outputs obtained from the two cases were then compared to observational data for June 8th, 1996 at 12:00. The grid for observational data was created using the same interpolation method as for the initial analysis. The purpose behind this step was to avoid any artifacts due to the choice of interpolation schemes.

It was found that the two versions performed very similar for most parameters but varied highly in the forecast of water-vapor values, as can be seen in figures 1 and 2. Version 1 gave surprisingly good predictions along the model boundary with the largest errors along hillsides, while version 2 had large error values at the model edges, but correctly predicted for the coastal and inland areas.

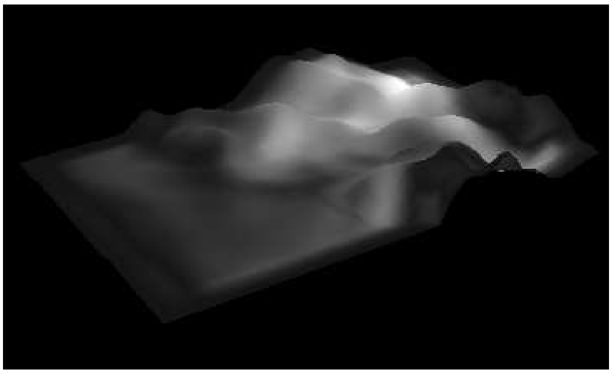


Figure 1: Difference between gridded observations and predictions of water-vapor from method 1.

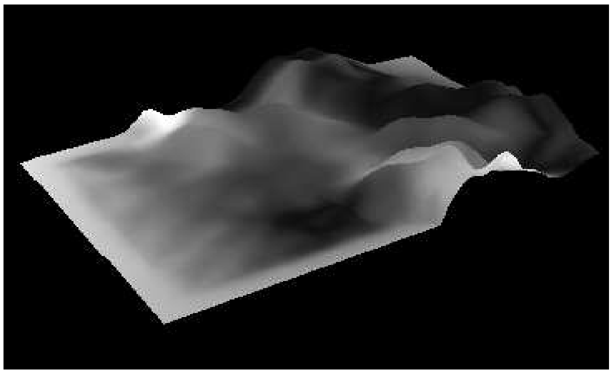


Figure 2: Same for method 2. The flat areas are values collected over the ocean surface.

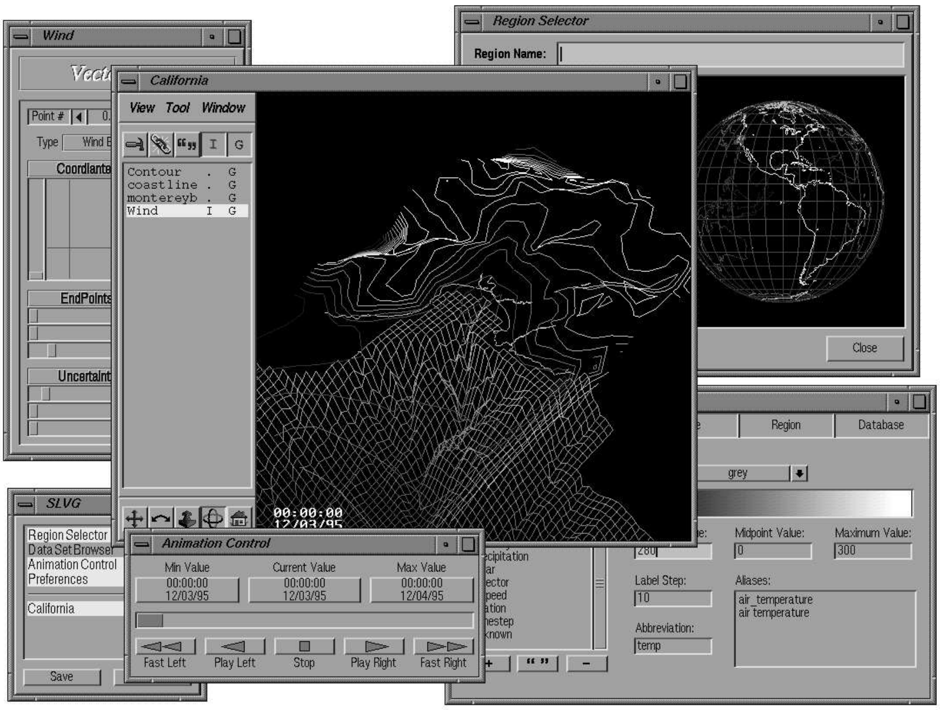
Alternatively, the observational data was compared against the values from the grid cell within which it was positioned, thereby providing a means of comparing specific station data against the forecast. The observational and model data was converted to HDF format, which was then read in by the visualization package. This way if it is found that a model consistently produces data of low certainty in a specific region where only sparse measurements are available, the meteorologist may attribute the error to the interpolation scheme and change it or possibly add a new station in the center of the region.

4 Visualization Tools

The visualizations described in this paper were created using SLVG, a visualization toolkit created at University of California, Santa Cruz, specializing in interactive displays of real-time weather data. SLVG utilizes a set of libraries providing a direct connection to a database containing last 3 years of observational data from the greater region of Monterey Bay. SLVG is written in C++ and OpenGL, using Xforms as the interface package.

Figure shows a sample SLVG session.

Below we list several options in displaying the model-to-observation differences. In all cases only a few graphical attributes (usually size or color) were employed for the display of uncertainty, leaving room for the use of the remaining components in the display of base data values.



1. *Pseudo-colored surface*. The user is allowed to select a 2D slice of the model grid, which is then color-mapped according to the difference values found at the grid points. This method is very useful for examining horizontal layers of data.

2. *Station point values*. Depending on the environmental parameter being displayed, there is either a color-mapped sphere or an arrow (in case of vector data) representing the magnitude and orientation difference between model versus observation values. The user has the option of displaying text values for all environmental parameters. Useful for evaluating interpolation methods.

3. *Iso-Contours* for displaying ranges of similar values. We expect this feature to allow 0 to place the display in the familiar context of isobaric pressure levels. The other purpose of iso-contours is to isolate the regions above or below a critical error threshold value.

4. *Volume probes*. Due to the density of data, it is frequently not feasible to display glyphs representing data values at each grid cell. Therefore, we provide the user with variable resolution volume probes. The available probes are :

   - *Spheres* - where the radius of the sphere is mapped to error value and color to the observational scalar value. The side-effect of this method is that areas of high error values appear more dense, thereby providing a natural visual cue.
   - *Spheres with Arrows* - intended for display of wind uncertainty. Again radius and color are mapped to wind magnitude error and observed magnitude values respectively. The length of the arrow represents the angular error, while the direction of the arrow follows the observed values.
   - *Mini-cubes / point clouds* - similar to the spheres, these computationally inexpensive methods are provided as alternatives to volume rendering. Semi-transparent cubes, not larger than half the size of a grid cell provide a quick overview of volumetric distribution of the data. For very high density data, visualization using colored points to represent the error at each grid cell is needed to preserve interactivity. However, static views using the point cloud can be confusing and works better when the user has the opportunity to interactively rotate and view the results from various angles.

All of the above options can be combined to produce a more conclusive view of the data. Drawing the terrain also provides visual cues for registering the dataset to a geographic location.

One difficulty we encountered was the problem of scaling the data, since there is no expected range of values for the error. To keep the size of the glyphs within a viewable range and to provide a consistent color coding scheme, we had to scale the error values. We first scanned the dataset for the absolute value of the maximum difference $maxDiff$ for the given parameter. We then scaled the rest of the error values to lie within 0 and 1 using the following formula:
$adjustedDifferenceValue = abs(modelValue - observationValue)/maxDiff$
In case of pseudo-coloring to the gray scale, we mapped low values to black and high error values to white. The drawback of this approach that errors are always scaled to lie between 0 and 1 from one comparison time to another. Hence, the visual output does not provide a method for evaluating the relative size of the error since the range of glyph sizes and colors are the same for all cases. Weighing this problem against the option of a cluttered or busy display, we decided to simply provide the information on the range of error values to the user as an auxiliary presentation (e.g. with a separate legend).

5 Conclusion and Future Work

We have presented several methods of scientific data visualization used for improved understanding of uncertainty associated with forecasting models. Although this paper concentrates on the use of visualization techniques in meteorology, this approach can easily be applied to other application domains where alternative ways of analyzing error in gridded data are needed, such as experimental and simulated wind measurements, mining and various GIS modeling schemes. Specifically in meteorology an analyst can gain a better understanding of the effects of their parameter choice by visualizing the difference between the given model output and observational data and comparing it to a difference produced had a different set of model parameters been used.

As our project evolves, we would like to expand our suite of tools to include new instruments and data types (e.g. volumetric data from NEXRAD) and other methods of visualizations such as iso-surfaces and volume visualization. Also, more continuity is required in the process of assessing a model before a conclusive decision can be made about the quality of the forecasts it produces. At the moment, it is premature to evaluate the effectiveness of our visualization tools. A long-term study would be a more adequate framework for both model and tool evaluation.

ACKNOWLEGMENTS

We would like to thank Dr. Douglas Miller and Prof. Wendell Nuss for providing model output data from the Naval Postgraduate School. We would also like to thank members of the SLVG research group including: Michael O'Neil, Chris Oates, Elijah Saxon, Jeremy Story, and Zoe Wood. This work is funded in part by ONR grant N00014-92-J-1807 and NSF grant IRI-9423881.

6 REFERENCES

[1] Craig M. Wittenbrink, Alex T. Pang, and Suresh K. Lodha. Glyphs for visualizing uncertainty in vector fields. *IEEE Transactions on Visualization and Computer Graphics*, 2(3):266–279, September 1996.

[2] William Hibbard and David Santek. The Vis-5D system for easy interactive visualization. In *Visualization*

'90, pages 28–35, 1990.

[3] M. L. desJardins, K. F. Brill, and S. S. Schotz. Use of gempak on unix workstations. In *Seventh Conference on Interactive Information Processing Systems for Meteorology, Hydrology, and Oceanography,New Orleans, LA*, pages 449–451, 1991.

[4] Dan Vietor. Wxp-the weather processor. URL: http://wxp.atms.purdue.edu.

[5] Alex Pang and Craig Wittenbrink. Spray rendering as a modular visualization environment. *Computer Graphics*, 29(2):33 – 36, 1995.

[6] A. Pang, C.M. Wittenbrink, and S. K. Lodha. Approaches to uncertainty visualization. Technical Report UCSC-CRL-96-21, University of California, Santa Cruz, 1996.

[7] Roger Daley. *Atmospheric Data Analysis*. Pergammon, Oxford, U.K., first edition, 1994.

[8] Georg A. Grell, Jimy Dudhia, and David R. Stauffer. Description of the fifth-generation penn state/ncar mesoscale model (mm5). URL: http://www.mmm.ucar.edu/mm5/mm5-desc-doc.html. NCAR/TN-398+STR.

[9] G. W. Paltridge. The value of climate forecasting. *Geophysical Surveys*, 7(3):273–90, June 1985.

[10] G. Wills et al. Statistical exploration of spatial data. In *Proceedings of the 4th International Symposium on Spatial Data Handling, Vol. 1*, pages 491–500. IGU, July 1990.

[11] R. A. Anthes, Kuo Ying-Hwa, Hsie Eirh-Yu, S. Low-Nam, et al. Estimation of skill and uncertainty in regional numerical models. *Quarterly Journal of the Royal Meteorological Society*, 115(488):763–806, July 1989.

[12] Roger Daley. Estimating observation error statistics for atmospheric data assimilation. *Annales Geophysicae*, 11(7):634–47, July 1993.

[13] T.T. Warner, L.E. Key, and A.M. Lario. Sensitivity of mesoscale-model forecast skill to some initial-data characteristics, data density, data position, analysis procedure and measurement error. *Monthly Weather Review*, 117(6):1281–310, June 1989.

[14] G. J. Boer. Systematic and random error in an extended-range forecasting experiment. *Monthly Weather Review*, 121(1):173–88, January 1993.

[15] Y.-T. Hou, K.A. Campana, K. E. Mitchell, S.-K. Yang, et al. Comparison of an experimental noaa avhrr cloud dataset with other observed and forecast cloud datasets. *Journal of Atmospheric and Oceanic Technology*, 10(6):833–49, December 1993.

[16] M. Kate Beard, Barbara P. Buttenfield, and Sarah B. Clapham. NCGIA research initiative 7: Visualization of spatial data quality. Technical Paper 91-26, National Center for Geographic Information and Analysis, October 1991.

[17] Yee Leung et al. Visualization of fuzzy scenes and probability fields. In *Proceedings 5th International Symposium on Spatial Data Handling*, pages 480 – 490. University of South Carolina, August 1992.

[18] M. Goodchild, B. Buttenfield, and J. Wood. on introduction to visualizing data validity. In H.M. Hearnshaw and D.J. Unwin, editors, *Visualization in Geographical Information Systems*, pages 141–149. Wiley, 1994.

[19] Sarah B. Clapham. A formal approach to the visualization of spatial data quality. In *GIS/LIS Proceedings, Vol. 1*, pages 138–149. American Society for Photogrammetry and Remote Sensing, November 1992.

[20] Trevor J. Davis and C. Peter Keller. Modelling multiple spatial uncertainties. URL: http://geography.geog.uvic.ca/dept/announce/fuzzy/cgpaper, 1995. University of British Columbia and University of Victoria.

[21] Beverley J. Evans. Cartographic display of data-reliability:does it benefit the map user? URL: http://www.gis.psu.edu/ica/icavis/evans/evans.html, 1995. Pennsylvania State University.

[22] Wendell Nuss. Use of multiquadric interpolation for meteorological objective analysis. *Monthly Weather Review*, 122(7):1611–31, July 1994.

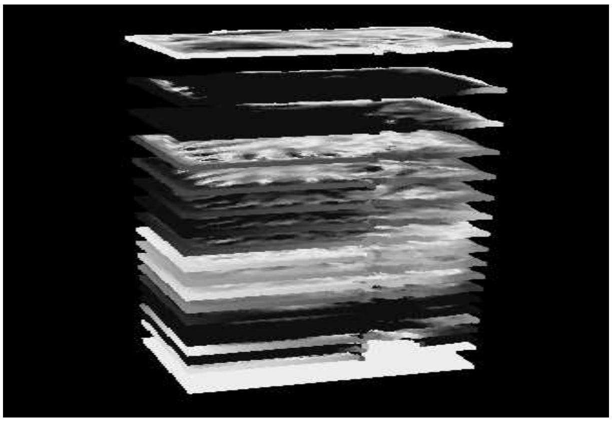


Figure 3: Layers of upper atmosphere temperature differences, produced using the minicubes method.

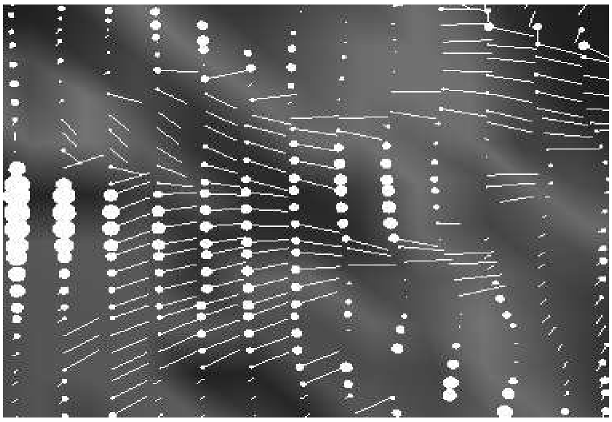


Figure 4: Wind magnitude (spheres) and angle (lines) differences, with background topology.

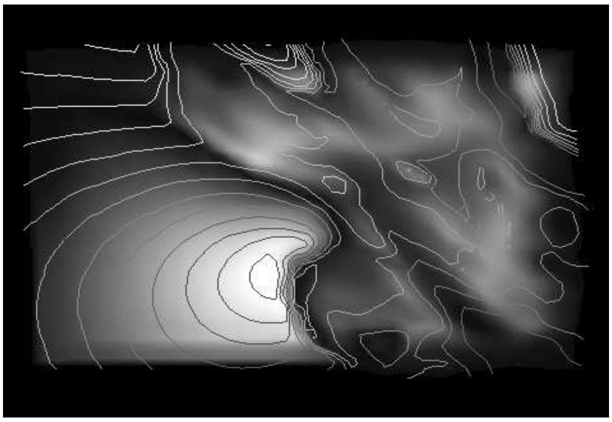


Figure 5: Pseudo-colored difference of surface temperature values between observation and method 1.

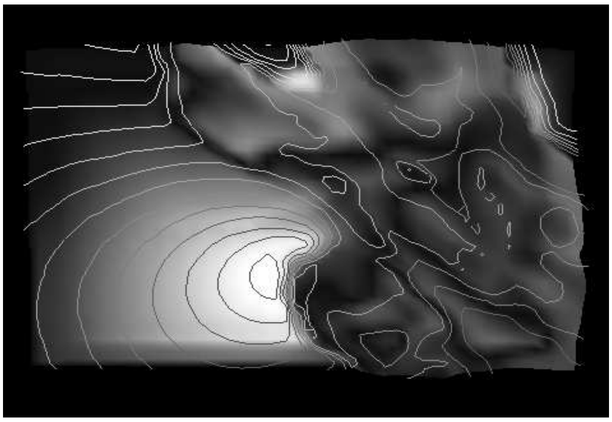


Figure 6: Same for observation and method 2. Superimposed are contours of surface temperature values.

Figure 7: Spheres of variable size are used to represent difference between observation and model values for method 1.

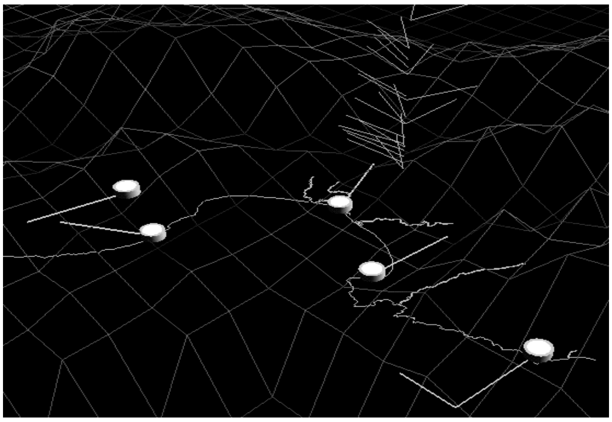


Figure 8: For this image we combined the station plot, coastline and terrain tools for a view of the model area.