Centrum voor Wiskunde en Informatica
REPORT*RAPPORT*

Interactive Exploration and Modeling of Large Data Sets: A Case Study with Venus Light Scattering Data

J.J. van Wijk, H.J.W. Spoelder, W.J.J. Knibbe, K.E. Shahroudi

Software Engineering (SEN)

SEN-R9719 August 31, 1997

Report SEN-R9719
ISSN 1386-369X

CWI
P.O. Box 94079
1090 GB Amsterdam
The Netherlands

P.O. Box 94079, 1090 GB Amsterdam (NL)
Kruislaan 413, 1098 SJ Amsterdam (NL)
Telephone +31 20 592 9333
Telefax +31 20 592 4199

Interactive Exploration and Modeling of Large Data Sets: A Case Study with Venus Light Scattering Data

Jarke J. van Wijk[1], Hans J.W. Spoelder[2], Willem-Jan Knibbe[2] Kamran Eftekhari Shahroudi
CWI
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

ABSTRACT
We present a system where visualization and the control of the simulation are integrated to facilitate interactive exploration and modeling of large data sets. The system was developed to estimate properties of the atmosphere of Venus from comparison between measured and simulated data. Reuse of results, distributed computing, and multiple views on the data were the major ingredients to create an effective environment.

1991 Computing Reviews Classification System: I.3.2 [**Computer Graphics**]: Graphics Systems - *Distributed/network graphics*; I.3.6 [**Computer Graphics**]: Methodology and Techniques - *Interaction techniques*; I.3.8 [**Computer Graphics**]: Applications; I.6.4 [**Simulation and Modeling**]: Model Validation and Analysis; J.2 [**Physical Sciences and Engineering**]: *Astronomy, Earth and atmospheric sciences*
Note: Presented at Visualization'96, San Francisco, 27-31 October, 1996.
Work carried out under CWI project SEN-1.3, Interactive Visualization Environments.

1. INTRODUCTION

With the ever growing availability of computational power and graphical capabilities of computer systems it becomes feasible to interactively explore and model increasingly large and complex data sets. This interaction is crucial if the model-parameters for the experimental observations are not or only partially known and when no clear-cut optimization criterion can be defined for the matching of experimental data and theoretical models.

Challenges in this field lie primarily in graphical support for the matching process, and the computational techniques that can be exploited to accelerate the model evaluation to enable an interactive approach.

To study the feasibility of such a combination of interactive exploration and modeling we have performed a case study on light scattering data which have been gathered for more than a decade by the Pioneer Venus Orbiter encircling Venus. In this so-called OCPP (Orbiter Cloud Photo Polarimeter) experiment [1] daily measurements of the visible part of the Venus atmosphere have been made. In these measurements intensity and degree of linear polarization at four different wavelengths were determined for the visible, sunlit part of the planet. Such a set, often referred to as a map, characterizes the time dependent state of the atmosphere. Knowledge of the theory of multiple light scattering has allowed for the development of compute intensive programs which calculate the afore mentioned experimental parameters for model atmospheres. Here scattering is mainly caused by the presence of one or two species of spherical particles with sizes characterized by their effective radius r and their effective standard deviation σ.

An interactive approach to gain insight is necessary here, because

- the proper model-parameters are not known in advance;
- physical constraints on the model-parameters and the a priori knowledge of the researcher about them are hard to formulate explicitly, yet must be satisfied to guarantee realistic results;
- no integral minimization criterion can be defined. This requires a weighting of the difference between calculated and measured results for different days, wavelengths and locations, which is hard to formulate.

[1] Netherlands Energy Research Foundation ECN, P.O. Box 1, 1755 ZG Petten, The Netherlands.
[2] Department of Physics and Astronomy, Vrije Universiteit, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands.

- each single simulation takes considerable time on super computers (30 s on a Cray C98/4256). The insight of the researcher is needed to reduce the number of steps in the search space.

Therefore, the graphical user-interface to the simulation becomes a crucial component. This interface has to show model results, allow for easy comparison with experimental data, and enable the researcher to perform simulations with new settings of the parameters.

2. APPROACH

The first key problem to be solved was the large simulation time. If new results can be calculated instantaneously, then the user can be enabled to change parameter values by dragging a mouse, while viewing the effect in real-time. However, since each simulation takes about 30 s, this option was not useful here. The best solution we found was to be economically with these precious results. The result of each simulation is saved to file, and the user is enabled to browse easily through these results. Meanwhile, he can ask for new results to be calculated.

3. INTERFACE

The second key problem is the visualization of many experimental data and simulation results, and interaction with model parameters. The best results were achieved by presentation of multiple simple displays simultaneously, where each display shows a different selection of the data, augmented with many options to simplify browsing through the data. We have experimented with several presentation methods of the data. There seemed to be no single optimum. What and how to present depends on the particular features and aspects of interest – which will change as a result of getting insight. Moreover, the presentation has to depend on the particular hypothesis to be tested. Here we have studied two different hypotheses: First, the properties of the particles are uniform over the surface of the planet; Second, these properties vary. Especially in the latter case a human in the loop is indispensable, first to judge if the results are physically relevant, second to steer the matching process.

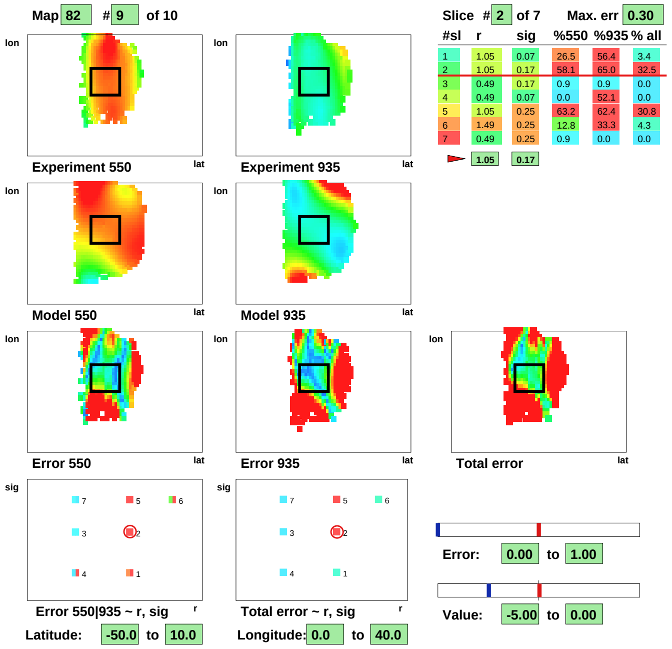
3.1 Constant radius

The underlying assumption of the user interface shown in fig. 1 is that r and σ are constant over the surface of the planet. Various versions of the spatial distribution are shown in a matrix. The two columns correspond to the two most relevant wavelengths (550 nm and 935 nm). The value presented here is the ratio of the Stokes parameters Q and I, which is a measure for the degree of polarization of the reflected light. The ratio Q/I depends on the properties of the particles scattering the light, and can hence be used to determine these [2]. From top to bottom the experimental values, the results of the simulation, and the deviation of model and experiment are shown.

The table in the upper-right corner shows that in total 7 such results (called slices) have been computed already. For each slice the r and σ, and three error-measures are shown. The value shown is the percentage of all points for which the deviation of the calculated value from the measured value is within a threshold. It is given per wavelength, as well as combined. The user can select the area for which these values must be evaluated. By dragging and resizing the rectangle he can study the spatial dependency of the goodness of the fit. The value of the error threshold is shown in the upper-right corner. The spatial distribution of the error combined over the wavelengths (root of squared sum), is shown at the right of the two displays of the error per wavelength.

The two displays below the spatial distributions show an alternative presentation of the calculated slices. Here these are shown in the parameter space: horizontal axes denote the value of r, vertical axes denote the value of σ. In the right display each slice is shown as a square, which color corresponds to the total error found for that slice. In the left display each slice is presented via two rectangles, one per wavelength, where the color shows the error per wavelength. The slice for which currently the spatial distribution is shown is encircled, and the corresponding row in the table is also marked.

The researcher can interact with the visualization in several ways. The day of measurement can be set by clicking on its number in the upper-left corner and entering a new value. The result is that the measurements and the previously computed slices for that day are retrieved. If no computed slice is available, a new simulation is started with the current values of r and σ. Hence, the results for these parameters can easily be compared for multiple measurement days. Previously computed slices can be selected by clicking at the corresponding row



Map 82 # 9 of 10

Slice # 2 of 7 Max. err 0.30

#sl	r	sig	%550	%935	% all
1	1.05	0.07	26.5	56.4	3.4
2	1.05	0.17	58.1	65.0	32.5
3	0.49	0.17	0.9	0.9	0.0
4	0.49	0.07	0.0	52.1	0.0
5	1.05	0.25	63.2	62.4	30.8
6	1.49	0.25	12.8	33.3	4.3
7	0.49	0.25	0.9	0.0	0.0
►	1.05	0.17			

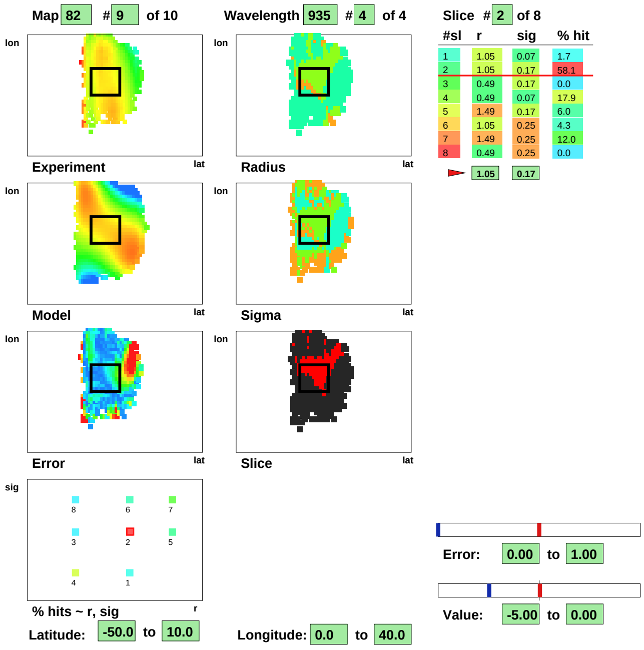
Error: 0.00 to 1.00

Value: -5.00 to 0.00

Latitude: -50.0 to 10.0

Longitude: 0.0 to 40.0

Figure 1: Interface for uniform particle distribution



Map 82 # 9 of 10

Wavelength 935 # 4 of 4

Slice # 2 of 8

#sl	r	sig	% hit
1	1.05	0.07	1.7
2	1.05	0.17	58.1
3	0.49	0.17	0.0
4	0.49	0.07	17.9
5	1.49	0.17	6.0
6	1.05	0.25	4.3
7	1.49	0.25	12.0
8	0.49	0.25	0.0
►	1.05	0.17	

Error: 0.00 to 1.00

Value: -5.00 to 0.00

Latitude: -50.0 to 10.0

Longitude: 0.0 to 40.0

Figure 2: Interface for varying particle distribution

in the table, or by clicking at the squares and rectangles in the r/σ displays. A new slice can be computed by selecting the current value of r or σ and entering a new value. Furthermore, the error-threshold can be set and the color scales for the displayed value and errors can be modified.

3.2 Varying radius

Another hypothesis is to assume that the parameters vary over the surface of the planet. This assumption is used in the interface shown in fig. 2. The results are shown for a single wavelength. The left column of the surface distribution displays show again the measured data, the computed result, and the deviation. In the two upper-right displays the optimal values for r and σ are shown. For each point the slice with the smallest deviation is looked up, and the corresponding r and σ are shown. Hence, each slice gives for some part of the surface the best result. This area is shown in the third display of the second column. In addition to the interaction options presented before, the user can click also in the distribution displays at the right, upon which the corresponding calculated slice is shown.

In all displays color was intensively used as a means to visualize data, not only in the distribution displays, but also in the table with available slices. An alternative representation of combinations of r and σ pairs is shown in fig. 3. Here r is mapped to the radius of the circle and σ is mapped to the width of a ring around the circle. This representation is helpful to investigate the inverse correlation between radius size and size of distribution which is suggested in literature. From figure 2 one sees that, according to this modeling, the radius varies only slightly over the planet, which is in agreement with earlier analysis of earth based polarimetry [3], and may be obtained for droplets consisting of a concentrated solution of sulfuric acid.

This approach can also be helpful to investigate the hypothesis that larger particles are found near the poles of Venus and exhibit a time dependent behavior over a period of years which has been suggested in literature.

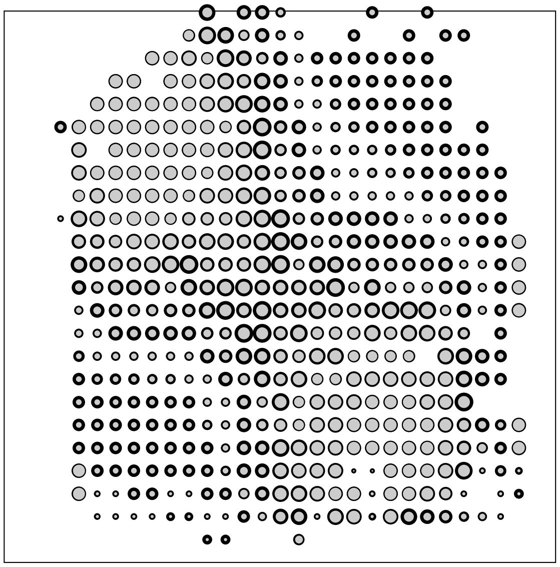


Figure 3: Simultaneous display of r and σ

4. IMPLEMENTATION

This application involved user-interface, visualization, data base and distributed computing aspects. A data flow diagram of the system is shown in fig. 4.

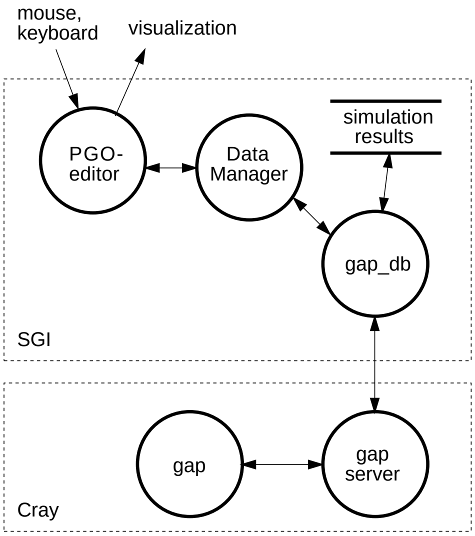


Figure 4: Data flow diagram

For the development of the user-interface we used the Computational Steering Environment (CSE), developed at CWI [4, 5]. The CSE provides a powerful graphics editor for the definition of user-interfaces and visualizations. This editor is based on Parametrized Graphics Objects. The user can draw graphics objects, which properties (geometry, color) are parametrized to data. This parametrization is bidirectional: calculated data can be shown as graphics, graphics objects can be manipulated to change data. As an example, fig. 5 shows the construction of fig. 3. Two circles are defined. Each circle is defined by two points: one for the center and one on the edge. The two circles share the center point, which is parametrized to latitude and longitude. The distance of the edge point of the inner circle is parametrized to r, the edge point of the outer circle is linked to the other edge point, and its distance from this point is parametrized to σ. When the PGO-editor is switched from edit-mode to run-mode, this visual specification is interpreted: the data are retrieved and mapped on the graphics objects. Arrays give multiple instances of the objects.

The data are exchanged with the central process in the CSE: the Data Manager. A separate process (gap_db), developed for this application, is also connected to the Data Manager. It takes care of management of a data base of results, and handles requests for new simulations. If the requested data are available, these are returned, else a request for a new simulation is issued. These requests are passed to a compute server, running at a Cray C98/4256. This compute server calls the simulation programs (gap), and returns the results to gap_db, which in turn saves these results in the data base and transfers them to the Data Manager.

5. CONCLUSIONS

The development of models that match experimental results for a wide range of conditions is a complex task. Automated approaches fall short, interactive exploration with suitable interfaces is indispensable to gain insight and to improve the model. We have presented examples of such interfaces for the simulation of the atmosphere of Venus, an application which we think to be typical for a wide class of applications. The reuse of previously computed results, their simultaneous visualization in different ways, and options for easy browsing are important keys to effective systems. The CSE enabled us to develop this application in short time.

In the future we will expand on this application further. More complex distributions of particle size and deviation will be studied, hence higher dimensional parameter spaces have to be visualized. More support for navigation will be required here. Furthermore, efforts will be made to reduce the CPU times. A multi-processor IBM SP2 will be used, and the user will be enabled to run several simulations simultaneously. Another strategy that will be explored is to calculate results in advance. This requires a good understanding and a model of the

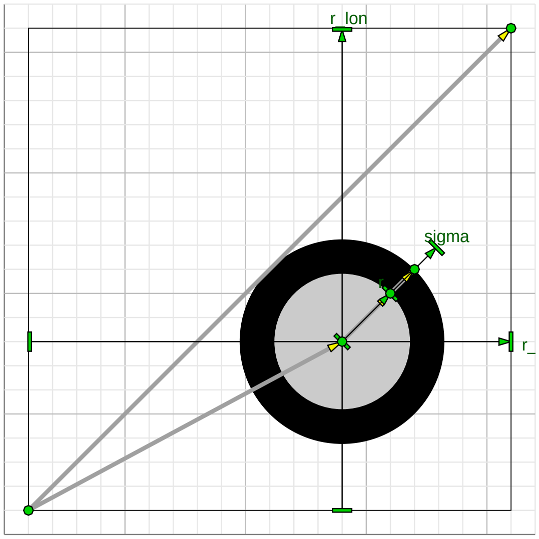


Figure 5: Construction of figure 3

user behavior in order to make reasonable predictions of desired results.

6. ACKNOWLEDGEMENTS

We thank Frans Groen and Frits Post for their stimulating contributions in the discussions. This work is supported by the Netherlands Computer Science Research Foundation (SION), with financial support of the Netherlands Organization for Scientific Research (NWO).

REFERENCES

1. E. Russel, L. Watts, S. Peelicori, and D. Coffen. Orbiter cloud photopolarimeter for the Pioneer Venus mission. *Proc. Soc. Photo. Opt. Intrumen. Eng.*, 112, pp. 28–44, 1977.

2. W.J.J. Knibbe, J.F. De Haan, J.W. Hovenier, and L.D. Travis. Spatial variations of Venus' cloud properties derived from polarimetry. In R. Santer (ed.) *Atmospheric Sensing and Modeling*, SPIE Proceedings, 2311, pp. 47–57, 1994.

3. J.E. Hansen, and J.W. Hovenier. Interpretation of the polarization of Venus. *J. Atmos. Sci.* 31, pp. 1137–1160, 1975.

4. J.J. van Wijk, and R. van Liere. An environment for computational steering. Presented at the Dagstuhl Seminar on Scientific Visualization, 23-27 May 1994, Germany, Proceedings to be published.

5. J.D. Mulder, and J.J. van Wijk. 3D Computational steering with parametrized geometric objects. In: G.M. Nielson, and D. Silver (eds.), *Proceedings IEEE Visualization'95*, CS Press, pp. 304–311, October 1995.