Data Assimilation and Objectively Optimized Earth Observation

David J. Lary[ab] and Anuradha Koratkar[b]

[a]Global Modelling and Assimilation Office, NASA Goddard Space Flight Center, MD, USA;
[b]GEST at the University of Maryland Baltimore County, MD, USA

ABSTRACT

This paper describes a vision for a future objectively optimized earth observation system. The system envisioned will dynamically adapt the what, where, and when of the observations made in an online fashion to maximize information content, minimize uncertainty in characterizing the systems state vector, and minimize both the required storage and data processing time for a given observation capability. Here we describe a prototype system with two relatively mature symbiotic components that seeks to achieve this goal. One component is the science goal monitor (SGM), the other is an Automatic code generation system for chemical modeling and assimilation (AutoChem). The Science Goal Monitor (SGM) is a prototype software tool to determine the best strategies for implementing science goal driven automation in missions. The tools being developed in SGM improve the ability to monitor and react to the changing status of scientific events. The SGM system enables scientists to specify what to look for and how to react in descriptive rather than technical terms. The system monitors streams of science data to identify occurrences of key events previously specified by the scientist. When an event occurs, the system autonomously coordinates the execution of the scientist's desired reactions. The data assimilation system can feed multivariate objective measures to the SGM such as information content and system uncertainty so that SGM can schedule suitable observations given the observing system constraints. The observing system may of course be a sensor web suite of assets including orbital and suborbital platforms.

Keywords: Optimized Earth Observation, Science Goal Monitor, Data Assimilation

1. INTRODUCTION

NASA science missions have traditionally operated on the assumption that we can only manage scheduling priorities and scientific processing on the ground with significant human interaction, and that all scientific data must be downloaded and archived regardless of its scientific value. However, increases in onboard processing and storage capabilities of spacecraft, as well as increases in rates of data accumulation will soon force NASA operations staff and scientists to re-evaluate the assumption that all science must be done on the ground. In order to take advantage of these new in-flight capabilities, improve science return and contain costs, we must develop strategies that will help reduce the perceived risk associated with increased use of automation in all aspects of spacecraft operations. An important aspect of science operations is the ability to respond to science driven events in a timely manner. For such investigations, we must teach our observing platforms to intelligently achieve the scientists goals. The principles presented here are generic but a specific example will be taken from atmospheric chemistry. The assimilation system described is being used in NASA Aura data validation.

Most of the key Science Questions from the NASA Earth Science Research Strategy involve the tracking of dynamically evolving geophysical fields. It is desirable to make the best use of a given earth observation capability by using an objective dynamic data retrieval control system that dynamically adapts the observations made in an online fashion. This facilitates the dynamic tracking of time-evolving sharp gradients. One example would be those in chemical tracer fields often located at the polar vortex edge, the tropopause and the day-night division. An example is shown in Figure 1. On the left the sharp gradient in NO (nitric-oxide) at the terminator can be seen. On the right a visualization of both the tropopause and the polar vortex can be seen, both are important mixing barriers.

This approach fits in well with the Sensor Web Concept . A system composed of multiple science instrument/processor platforms that are interconnected by means of a communications fabric for the purpose of collecting measurements and processing data for Earth or Space Science objectives.

Further author information: (Send correspondence to David.Lary@umbc.edu, Telephone: 1 301 614 6405)

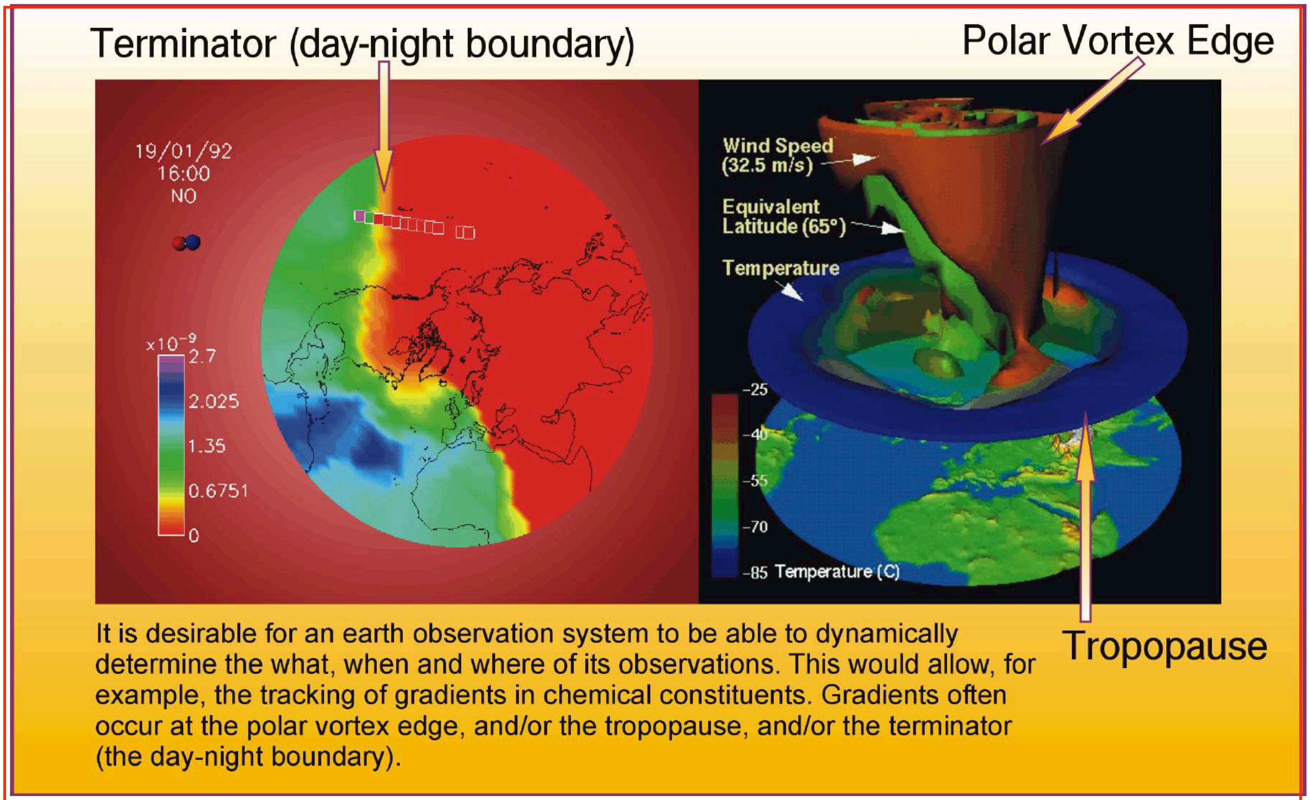

Figure 1. It would be very useful for an earth observing system to dynamically track evolving features. For example, on the left the sharp gradient in NO (nitric-oxide) at the terminator can be seen. On the right a visualization of both the tropopause and the polar vortex can be seen, both are important mixing barriers.

2. DYNAMIC DATA

The first and most important element is the concept of dynamic data . The dynamic data retrieval control system envisioned here dynamically adapts what measurements are made, where they are made, and when they are made. The dynamic adaption is performed an online to maximize the information content, minimize the uncertainty in characterizing the systems state vector, and minimize both the required storage and data processing time, and minimize the data heterogeneity minimization within analysis grid cells. (It is conceivable that these ideas could be used in future to direct additional observations from unmanned automated sub-orbital platforms.)

The ability to develop a dynamic data retrieval control system for an objectively optimized earth observation system depends in large part on products made available when data assimilation is an integrated part of the earth observation system. Making data assimilation an integral part of the earth observation system is a prudent step since assimilation seeks to bring together heterogeneous information together with its associated uncertainty from a variety of sources (both observational and theoretical) in a self-consistent mathematical framework.

3. SCIENCE GOAL MONITOR

At the heart of the dynamic data retrieval control system is the Science Goal Monitor. The Science Goal Monitor (SGM; http://aaa.gsfc.nasa.gov/SGM) is a prototype software tool to explore strategies for implementing science goal driven operations for multiple sensors/platforms.[1] A space science SGM is being prototyped for dynamic automated reactions to intrinsically varying astronomical phenomenon using one of the Small and Moderate Aperture Research Telescope System (SMARTS) telescopes. An Earth science prototype has been built for Earth Observing 1 (EO-1) to evaluate how multiple sensors can react dynamically to obtain rapid observations of evolving earth science events (see paper by Koratkar et al.

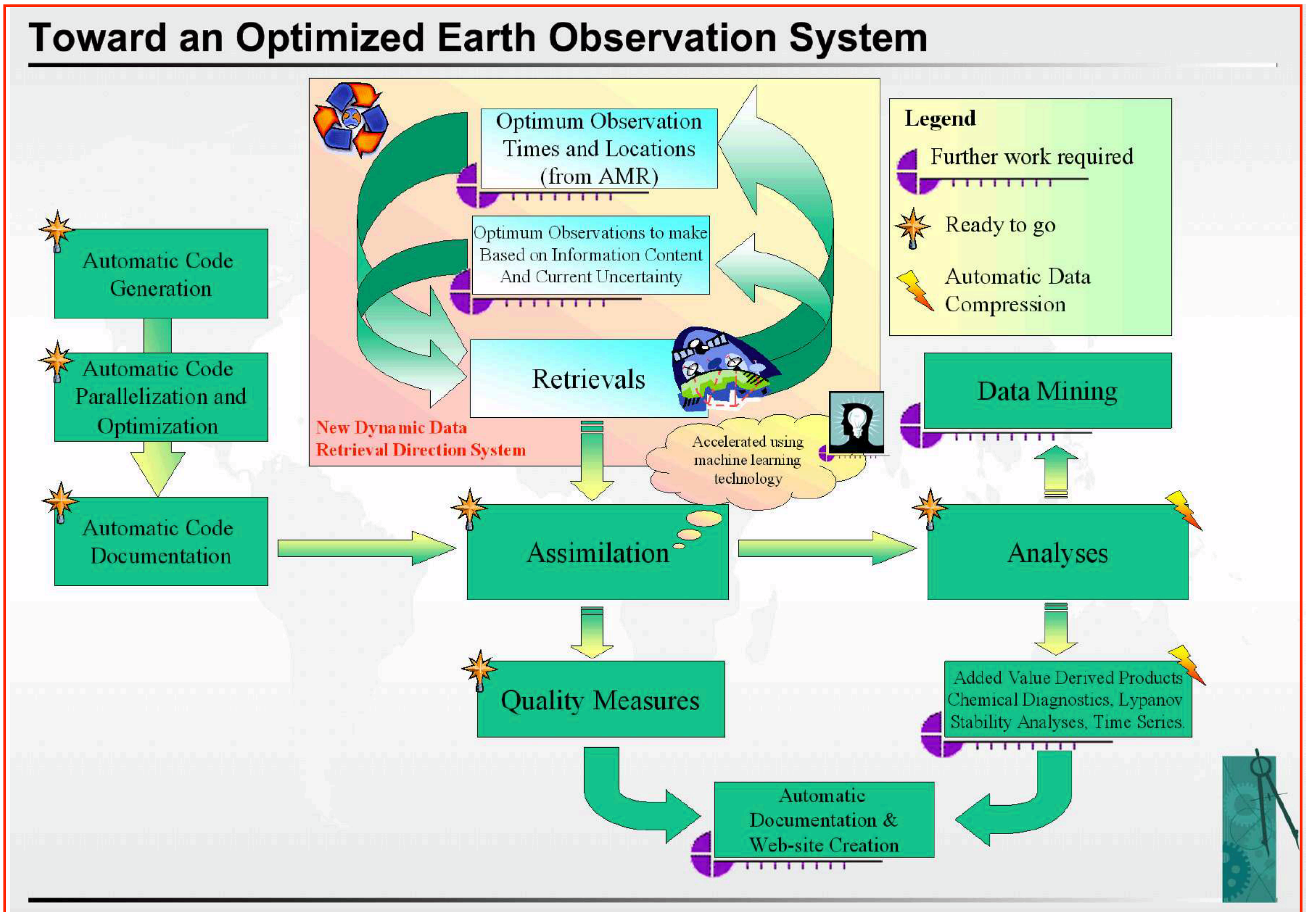

Figure 2. Schematic overview.

presented at this meeting). Here we envision extending these previous prototypes to use objective metrics such as information content and system uncertainty so that SGM can schedule suitable observations that objectively optimize the use of our assets.

4. INFORMATION CONTENT AND STATE VECTOR UNCERTAINTY

As a dynamic system evolves with time not all of the state variables within the state vector contain equal amounts of information (information content), and not all state variables are known to the same precision. It is therefore clearly desirable that the observations made both contain the maximum information content possible with a given observing platform capability and allow the systems state to be characterized with a minimum uncertainty.

Information content is a broad term that could be quantified in any number of ways depending on the system or problem being studied. Therefore, although we propose to use a specific measure of information content for the atmospheres chemical system, these measures could easily be substituted with alternative measures that may be more suitable depending on the given objectives of an investigation . Although we describe a specific example from atmospheric chemistry, the principle is clearly more general. The key new concept in this approach is that information content and system uncertainty are used in determining: what should be measured, when and where. Thus providing a cost effective strategy for using resources and minimizing the data storage required to characterize a system with a given level of precision.

One measure of information content/ranking that could be used is described by[2] coupled with the so-called goal attainment algorithm to provide the information content ranking. The chemical assimilation system will provide analyses of the state vector together with an associated uncertainty. The information content/ranking software uses the analyzed state vector to provide the information content ranking. This information is then passed to the SGM to allow it to objectively determine the following days observation schedule.

5. AUTOMATIC CODE GENERATION

The complexity of the atmospheres chemical system varies tremendously with location: From the relatively simple chemistry of the mesosphere involving primarily O, H, and N containing species. To the more complex chemistry of the stratosphere also involving Cl, Br, F and S containing species and simple hydrocarbons such as CH_4 and other greenhouse gasses. To the very complex chemistry of the troposphere, which also involves volatile organic hydrocarbons (VOCs) and their host of oxidation products. Therefore, any tool that is going to be involved in implementing a dynamic objectively optimized earth observation strategy must be capable of dealing with these very different chemical regimes. Consequently, it is most desirable to have an automatic code generator that is capable of creating and reusing code for the deterministic models required to describe the chemistry of these different regimes together with the entire data assimilation infrastructure required (i.e. time derivatives, Jacobians, Hessians, adjoints, and information content). The AutoChem code generation and modeling/assimilation system has these capabilities and has already been validated in a range of studies.

6. DATA ASSIMILATION

The information content metrics and uncertainty characterization will be supplied by the chemical assimilation system, AutoChem. AutoChem (http://pdfcentral.shriver.umbc.edu/AutoChem/index.html) is an automatic code generation system, documenter and symbolic differentiator for atmospheric chemical modeling and data assimilation[3].[4] An advantage of assimilation is that it propagates information from data-rich regions to data-poor regions. Data assimilation also offers a mathematical framework to check and quantify the chemical consistency of multispecies observations with one another and with photochemical theory through the use of objective skill scores. That is, the analysis can examine both the consistency between different instruments observing the same constituent, and the photochemical self-consistency between multiconstituent observations and photochemical theory.

7. AUTOMATIC DATA COMPRESSION

After the raw radiance data observed by a satellite is processed higher level one and two datasets are generated. These higher-level datasets are usually stored at a uniform precision, where the stored precision is usually significantly greater than the certainty with which the level one and two data are known. For example, the data may be stored with eight significant figures when we are only confident in the first three or four. If the total data volume is small then this does not have significant cost implication. However, when we are dealing with very high data volumes this does have a significant cost implication for storage and/or data transfer. For many years now a variety of data compression techniques have been used that could be adapted to reduce the amount of space required for data storage and time for data transmission. The degree of data compression can be chosen to make the compression non lossy for the accuracy characterized by the assimilation system, i.e. to three significant figures if that is how well we know the variable instead of to eight or sixteen significant figures if we do not know the variable to that precision. If it is found at a later date that reprocessing is required then this can still be done as the raw radiance data is stored to the full machine precision. Automatic data compression uses the dynamic data concept in the addition of value added products without incurring prohibitive space requirements.

8. MACHINE LEARNING

The whole approach described depends in large part on the integration of a data assimilation system. When considering data assimilation of atmospheric chemistry, one of the computationally most expensive tasks is the time integration of a large and stiff set of ordinary differential equations (ODEs). However, very similar sets of ODEs are solved at adjacent grid points and on successive days, so similar calculations are repeated many thousands of times. This is the type of application that benefits from adaptive, error monitored, machine-learning technology. Our ODE solver already employs adaptive time stepping with error monitoring, if this is extended to an adaptive use of machine learning then there are literally massive potential savings in computational expense. A prototype code has been developed that we would like to extend here for use within the ODE solver. Early work seems promising that such an approach would work.[5] A success in this area would mean a dramatic reduction in the computational cost of assimilation and hence of the entire dynamic data retrieval control system.

9. CONCLUSION

A schematic overview of the objectively optimized earth observation system envisioned is shown in Figure 2. The elements of the dynamic data retrieval control system can help both in objectively planning mission goals and in the cost effective operation of future optimized earth observing systems. During the planning stage the objective measures of information content are invaluable in determining what the instrument capabilities should be. During the operation of future earth observing systems the dynamic data retrieval control system could dynamically adapt what measurements are made, where they are made, and when they are made, in an online fashion to maximize the information content, minimize the uncertainty in characterizing the systems state vector, and minimize both the required storage and data processing time.

The same technology could be applied to the analyses and design of ground based pollution monitoring networks to provide regular pollution analyses. These could then be used for epidemiological studies in the precise quantification on the impacts of pollution on human health. For example, it was noted by Shallcross (personal communication) that high levels of benzene were associated with high hospital admissions of cardiovascular conditions.

At a more basic level the idea of symbiotic communication and dynamic data could be used in many applications to optimize monitoring and observing systems. The ideas of automatic code generation and automatic documentation to facilitate system implementation on a variety of hardware is also of quite general applicability. As is the concept of automatic data compression to minimize the required cost of both storage and dissemination.

REFERENCES

1. A. Koratkar, S. Grosvenor, J. E. Jones, A. Memarsadeghi, and K. R. Wolf, "Science goal driven observing: A step towards maximizing science returns and spacecraft autonomy," *SPIE* **4844**, p. 250, 2002.
2. B. Khattatov, J. Gille, L. Lyjak, G. Brasseur, V. Dvortsov, A. Roche, and J. Waters, "Assimilation of photochemically active species and a case analysis of UARS data," *J. Geophys. Res. (Atmos.)* **104**(D15), pp. 18715–18737, 1999.
3. M. Fisher and D. Lary, "Lagrangian 4-dimensional variational data assimilation of chemical-species," *Q. J. R. Meteorol. Soc.* **121**(527 Part A), pp. 1681–1704, 1995.
4. D. J. Lary, B. Khattatov, and H. Y. Mussa, "Chemical data assimilation: A case study of solar occultation data from the atlas 1 mission of the atmospheric trace molecule spectroscopy experiment (atmos)," *J. Geophys. Res. (Atmos.)* **108**(D15), 2003.
5. D. J. Lary, M. D. Muller, and H. Y. Mussa, "Using neural networks to describe tracer correlations," *Atmospheric Chemistry and Physics* **4**, pp. 143–146, 2004.