Also, a type replacement facility was tested for updating primitive types such as such as integers that were not qualified with a size, and thus may change as the underlying architecture is upgraded. Finally, the designers leveraged a tool, called eclim, to bridge the gap between text editors and Integrated Development Environments (IDE), by running both of them simultaneously, and set up to communicate with each other. Through this mechanism, modern IDE features were made available through text editors, minimizing the learning curve for scientists already experienced with their conventions.

This work was done by Marc Abrams, Pallabi Saboo, and Mike Sonsini of Harmonia Holdings Group, LLC for Goddard Space Flight Center. Further information is contained in a TSP (see page 1). GSC-16475-1

7 Automatic Data Filter Customization Using a Genetic Algorithm

This work predicts whether a retrieval algorithm will usefully determine CO₂ concentration from an input spectrum of GOSAT (Greenhouse Gases Observing Satellite). This was done to eliminate needless runtime on atmospheric soundings that would never yield useful results. A space of 50 dimensions was examined for predictive power on the final CO₂ results.

Retrieval algorithms are frequently expensive to run, and wasted effort defeats requirements and expends needless resources. This algorithm could be used to help predict and filter unneeded runs in any computationally expensive regime.

Traditional methods such as the Fischer discriminant analysis and decision trees can attempt to predict whether a sounding will be properly processed. However, this work sought to detect a subsection of the dimensional space that can be simply filtered out to eliminate unwanted runs. LDAs (linear discriminant analyses) and other systems examine the entire data and judge a “best fit,” giving equal weight to complex and problematic regions as well as simple, clear-cut regions. In this implementation, a genetic space of “left” and “right” thresholds outside of which all data are rejected was defined. These left/right pairs are created for each of the 50 input dimensions. A genetic algorithm then runs through countless potential filter settings using a JPL computer cluster, optimizing the tossed-out data’s yield (proper vs. improper run removal) and number of points tossed.

This solution is robust to an arbitrary decision boundary within the data and avoids the global optimization problem of whole-dataset fitting using LDA or decision trees. It filters out runs that would not have produced useful CO₂ values to save needless computation. This would be an algorithmic preprocessing improvement to any computationally expensive system.

This work was done by Lukas Mandrake of Caltech for NASA’s Jet Propulsion Laboratory. For more information, contact iaoffice@jpl.nasa.gov.

This software is available for commercial licensing. Please contact Dan Broderick at Daniel.F.Broderick@jpl.nasa.gov. Refer to NPO-48253.

7 Towards Efficient Scientific Data Management Using Cloud Storage

A software prototype allows users to backup and restore data to/from both public and private cloud storage such as Amazon’s S3 and NASA’s Nebula. Unlike other off-the-shelf tools, this software ensures user data security in the cloud (through encryption), and minimizes users’ operating costs by using space- and bandwidth-efficient compression and incremental backup. Parallel data processing utilities have also been developed by using massively scalable cloud computing in conjunction with cloud storage.

One of the innovations in this software is using modified open source components to work with a private cloud like NASA Nebula. Another innovation is porting the complex backup-to-cloud software to embedded Linux, running on the home networking devices, in order to benefit more users.

This work was done by Qiming He of Open Research for Goddard Space Flight Center. For further information, contact the Goddard Innovative Partnerships Office at (301) 286-5810. GSC-16475-1