

# A New Approach to Streaming Data from the Cloud

NIALL H. ROBINSON, RACHEL PRUDDEN, AND ALBERTO ARRIBAS

**TOO-BIG DATA.** With the rise of high-resolution environmental measurements and simulation, extremely large scientific datasets are becoming increasingly commonplace. Many of these large datasets are highly multidimensional. Earth science is generating such extremely large datasets on a daily basis, from weather forecasts to climate simulations: the United Kingdom's Met Office will soon generate ~400 TB daily, with our archive expected to reach the exabyte scale.

The scientific community is in the process of learning how to efficiently make use of these unwieldy datasets. Increasingly, data analysis and storage are being moved to “cloud computing,” where remote servers can let analysts process and disseminate data in a way that is scalable and economic. The web browser is emerging as the de facto interface to perform analysis of this cloud-hosted data, with the resultant information visualized locally in the web browser on the user's device. This approach, with the real computing power sitting in some central location, accessed by relatively low-powered user devices, is known as a “thin-client.”

At the same time, web browsers are becoming increasingly powerful. While remote servers are certainly required for heavy lifting analysis, many moderately intensive analysis and visualization tasks can now be handled by the web browser on the user's computer. For tasks such as interactive visualization, this can make the experience more responsive while reducing the demands on central servers. Clients need not be as “thin” as they used to be.

Currently, data can only be moved to users' web browsers when it has been downsized far enough. Efficient on-demand transfer of multidimensional data would allow us to flexibly move the boundary between specialized remote data servers for processing big datasets and local user machines for interrogation, visualization, and understanding. This flexibility is essential to allow users (be they analysts or members of the public) to fluently interact with cloud-hosted data.

To explore these principals, we rendered a large cloud-hosted atmospheric dataset in a web page by repurposing the technology used to stream high-definition videos. There are two salient qualities of this approach. First, it is compatible with web browsers, which will be the key user interface as data moves to the compute cloud. Second, it gives more data compression than standard atmospheric science algorithms.

**THROWING AWAY WHAT WE DO NOT NEED.** We explored these ideas by setting ourselves the challenge of visualizing a full weather forecast field as an animated 3D web page visualization. These data are richly spatiotemporal; however, they are routinely communicated to the public as a 2D map, and scientists are largely limited to visualizing data via static 2D maps or 1D scatterplots. We wanted to present Met Office weather forecasts in a way that represents all the generated data.

Web browsers have recently attained the ability to utilize the graphics card of the user's computer, allowing for immersive 3D visualizations in web pages. However, web pages are limited by the timely transfer of data to the user's browser via the internet. They also have a limited amount of memory available to them compared to traditional computer programs—typically less than 1.5 GB. As such, we needed to reduce the amount of data we transmitted as much as possible.

Many popular compression algorithms used in atmospheric science are reversible, often called “lossless”: they encode all the original information using less storage by taking advantage of redundancy in the

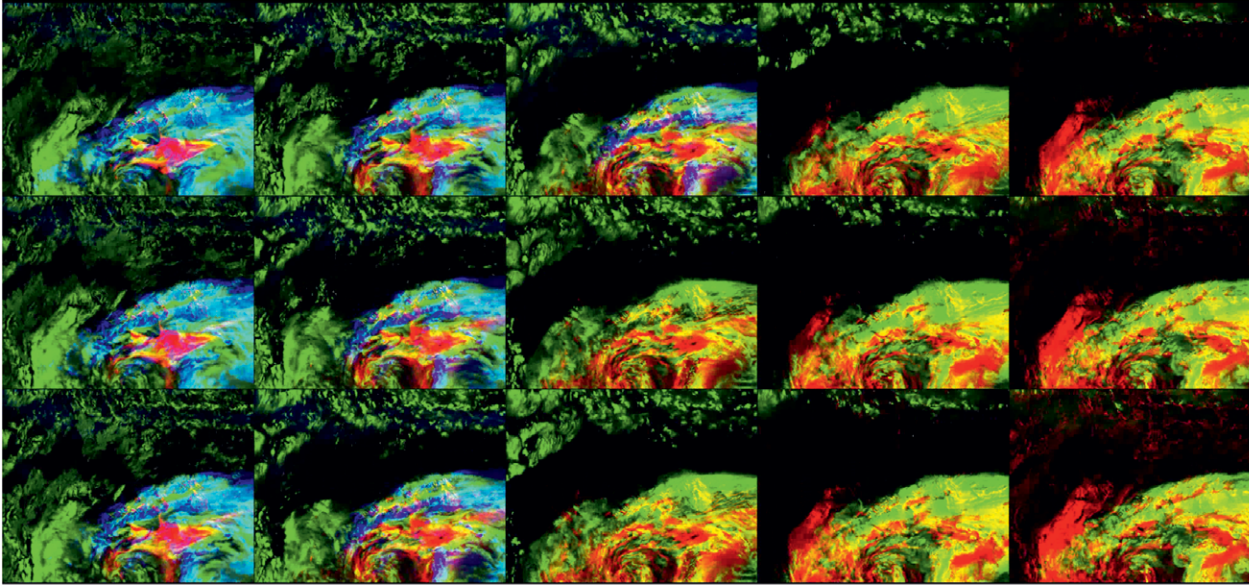
**AFFILIATIONS:** ROBINSON, PRUDDEN, AND ARRIBAS—  
Informatics Lab, Met Office, Exeter, United Kingdom

**CORRESPONDING AUTHOR:** Niall H. Robinson,  
niall.robinson@informaticslab.co.uk

*The abstract for this article can be found in this issue, following the table of contents.*

DOI:10.1175/BAMS-D-16-0120.1

For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#).



**FIG. 1.** 3D atmospheric data of cloud fraction encoded as pixels in an image.

data. By contrast, “lossy” compression can discard nonessential information to reduce storage space. This involves a judgement (often encapsulated in the compression algorithm) as to which kinds of information can be safely discarded.

Encoding high-resolution video is one application that has had a great deal of prior development, fueling technologies we use every day, such as YouTube and Netflix. We realized that this application had important parallels with our application. Specifically, it is optimized to vastly compress data with spatiotemporal coherence. It is lossy, but optimized to retain information that is discernible to the human eye.

In order to make use of these desirable properties, we repurposed video files so that instead of storing moving images, they stored gridded weather forecast data. These videos are never viewed directly by the end user; instead, they act as a vehicle to efficiently transmit data to the users' computer, where it is converted into a realistic, interactive 3D visualization.

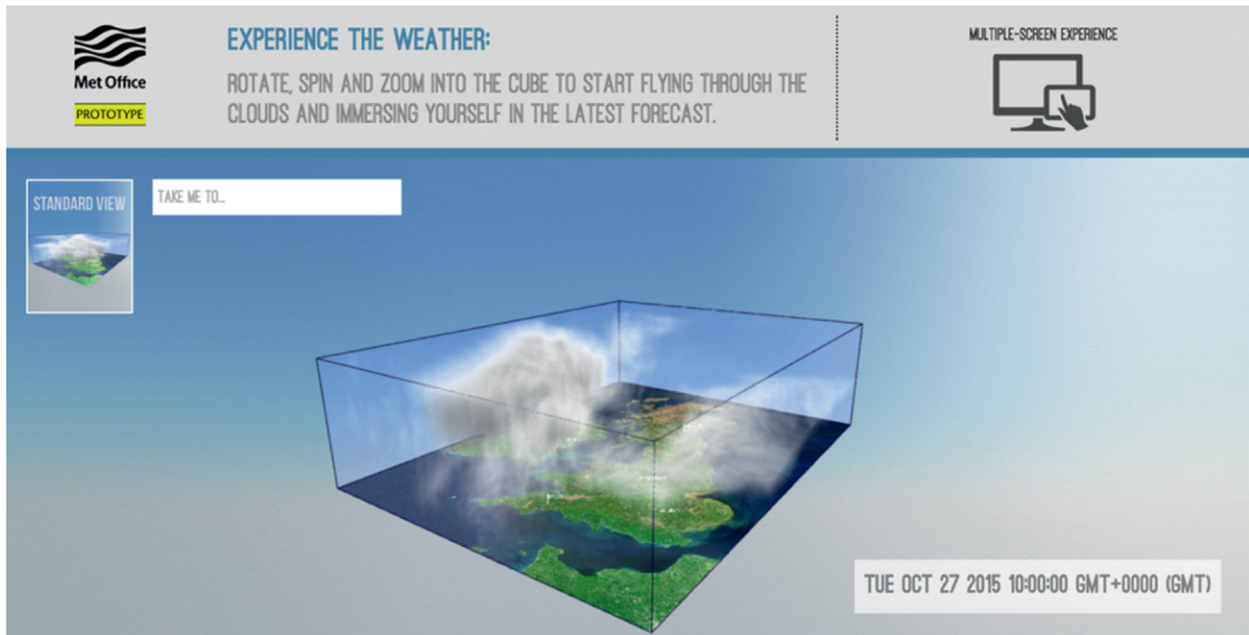
First, data from each forecast time were encoded as a separate image file, such as you might use to store a photograph. Each pixel in an image comprises several numbers representing the amount of red, green, and blue. However, instead of storing color, we used these channels to store our data values—that is, one color of one pixel in the image contains the value of one cell of the gridded data of the forecast. To do this, the 3D data array at a given forecast time, covering latitude, longitude, and altitude, was divided up into

slices comprising 2D latitude-by-longitude data. The slices from the different altitudes were then arranged adjacently in an image file. Data for different altitudes were also stored in the three different color channels of the image. An example of this image-encoded data for one time in the forecast is shown in Fig. 1.

These images, representing different times in the forecast, were then joined together as individual frames of a video using a high-resolution video compression algorithm (in our case, Ogg Theora), allowing it to be efficiently sent to the user. You can view an example of the data, encoded as a video, at [www.youtube.com/watch?v=FKv\\_lzTcjB8](http://www.youtube.com/watch?v=FKv_lzTcjB8).

A single, time-varying 3D field from the Met Office weather forecast is approximately 5 GB when it is stored as NetCDF, a common atmospheric data format. Various lossless (e.g., NetCDF+gzip) and lossy (e.g., Grib+JPEG 2000) compression formats are available to compress atmospheric science data; however, in our experience they are not routinely used.

Comparison of lossy compression is difficult, as the quality is inherently subjective. However, using video compression certainly results in smaller data files than routinely available geospatial compression algorithms. We chose to use Ogg Theora video compression as the algorithm is open source and supported by all the major web browsers. This approach compressed our ~5 GB data fields to between 10–20 MB—that is, ~400 times smaller than lossless NetCDF, and ~4–8 times smaller than lossy Grib compression.



**FIG. 2.** The web application rendering the video-encoded data as a 3D field.

### USING THESE DATA IN THE BROWSER.

This video of atmospheric data is then streamed into our web application for rendering (Fig. 2). We created an interactive 3D animation of the data using custom-made graphics card routines, which simulated the passage of light rays through the data (a technique known as “volume rendering” or “ray tracing”). This application allows the users to interact with the animated 3D data field over a standard internet connection without installing specialized software or hardware.

Employing video encoding is particularly useful in this context of delivery to web browsers. Browsers natively support the decoding of video data, as opposed to atmospheric data formats, which have to be decoded before being transmitted. Video can be easily streamed into the browser, meaning client memory is used efficiently. As the data are stored in a standard graphical data format, they can easily be transferred to the graphics card, where it can be rendered on the fly as the user interacts with the visualization. Other video functionality is also useful, such as playback controls and on-the-fly scaling.

**WHAT DID WE LEARN?** The use of video compression allowed us to significantly reduce the data load. The 400:1 reduction in data volume is due to

the loss of data. Crucially though, the information relevant to this application is retained as the major features of the data field are apparent in the final data rendering. Visual codecs are optimized to lose data that cannot be seen, a feature that is not just optimal for traditional 2D visualization, but also largely suitable for such 3D rendering.

Lossless compression has long been the default choice for scientific data. This means that, since all the original information is conserved, it can be safely applied in any situation. However, this universality comes at the price of unwieldy data volumes. Incorporating lossy compression at appropriate points in the data supply chain can reduce data volumes without impacting the overall quality.

**WHAT NEXT?** The approach presented here compresses by using coherence in the latitude  $\times$  longitude  $\times$  time dimensions, but there is coherence in other dimensions (for instance, altitude) that could be leveraged to give even greater compression ratios. Also, video compression is optimized to preserve visual information, but work could be done to preserve more esoteric properties (for instance, atmospheric turbulence information). While the appropriate data precision is clearly application-specific, we assert that it is routinely

spuriously high in atmospheric science, both in terms of the generated data and the subsequent use case. Given the data volume challenge we face, we should consider the wider use of more aggressive data compression techniques, such as lossy compression, as a matter of course.

Simulations and measurements of the environment we live in seem set to increase in application as well as volume. As virtual and augmented reality technologies gain traction (both in user entertainment and data analysis), it is imperative that we can broadcast dynamic content for users; “3D video” should be allowed to become a fundamental and common type of data.

There is beginning to be a change in how we deal with modern data volumes, with some organizations starting to move analysis and dissemination to the compute cloud. As such, we think that approaches similar to the one presented here, which allow effective streaming of big datasets over the internet into web browsers, will become a fundamental portal to data in coming years

**ACKNOWLEDGMENTS.** The authors wish to thank Ed Campbell of the Met Office for helpful discussions regarding video compression, and the members of the Met Office Informatics Lab.

## FOR FURTHER READING

Becker, P., L. Plesea, and T. Maurer, 2015: Cloud optimized image format and compression. *Int. Arch. Photogramm., Remote Sens. Spat. Inf. Sci.*, **XL-7/W3**, 613–615. [Available online at [www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XL-7-W3/613/2015/isprsarchives-XL-7-W3-613-2015.pdf](http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XL-7-W3/613/2015/isprsarchives-XL-7-W3-613-2015.pdf).]

Hübbe, N., A. Wegener, J. M. Kunkel, Y. Ling, and T. Ludwig, 2013: Evaluating lossy compression on climate data. *Supercomputing: 28th International Supercomputing Conference, ISC 2013*, Leipzig, Germany, June 16-20, 2013. Proceedings, J. M. Kunkel, T. Ludwig, and H. W. Meuer, Eds., Springer, 343–356.

Kim, B., K. H. Lee, K. J. Kim, T. Richter, H.-S. Kang, S. Y. Kim, Y. H. Kim, and J. Seo, 2009: JPEG2000 3D compression vs 2D compression: An assessment of artifact amount and computing time in compressing thin-section abdomen CT images. *Med. Phys.*, **36**, 835, doi:10.1118/1.3075824.

Lucero A., S. D. Cabrera, A. Aguirre, and E. Vidal, 2003: Compressing three-dimensional GRIB meteorological data using KLT and JPEG 2000. Proc., *Int. Geoscience and Remote Sensing Symp. (IGARSS '03)*, Toulouse, France, 1836–1838. [Available online at <http://ieeexplore.ieee.org/iel5/9010/28603/01294266.pdf>.]

## NEW FROM AMS BOOKS!

### Weather in the Courtroom Memoirs from a Career in Forensic Meteorology

**William H. Haggard**

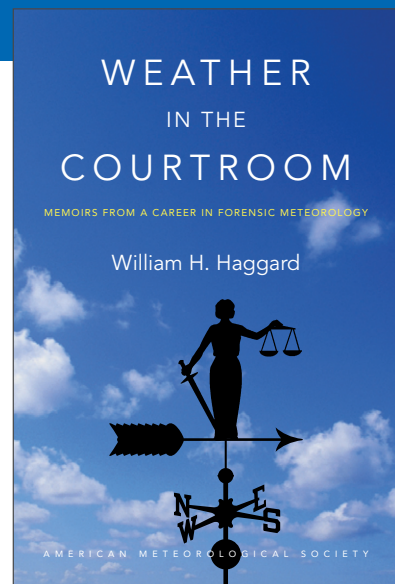
From a pioneering forensic meteorologist...

The inside scoop on legendary litigations where the weather may—or may not—have played a crucial role, including:

- The disappearance of an Alaskan congressman's airplane in 1972
- The collapse of Tampa Bay's Skyway Bridge in 1980
- The crash of Delta Flight 191 in Dallas/Fort Worth in 1985

**William H. Haggard** is former director of the NCDC and AMS certified consulting meteorologist (CCM), fellow, and honorary member.

© 2016, paperback, 240 pages  
ISBN 13: 978-1-940033-95-2  
List price: \$30    AMS Member price: \$20



**AMS BOOKS**  
RESEARCH APPLICATIONS HISTORY  
[bookstore.ametsoc.org](http://bookstore.ametsoc.org)