



EarthServer: Visualisation and use of uncertainty as a data exploration tool

Peter Walker, Oliver Clements, and Mike Grant
Plymouth Marine Laboratory, Plymouth, United Kingdom (petwa@pml.ac.uk)

The Ocean Science/Earth Observation community generates huge datasets from satellite observation. Until recently it has been difficult to obtain matching uncertainty information for these datasets and to apply this to their processing. In order to make use of uncertainty information when analysing "Big Data" we need both the uncertainty itself (attached to the underlying data) and a means of working with the combined product without requiring the entire dataset to be downloaded.

The European Commission FP7 project EarthServer (<http://earthserver.eu>) is addressing the problem of accessing and ad-hoc analysis of extreme-size Earth Science data using cutting-edge Array Database technology. The core software (Rasdaman) and web services wrapper (Petoscope) allow huge datasets to be accessed using Open Geospatial Consortium (OGC) standard interfaces including the well established standards, Web Coverage Service (WCS) and Web Map Service (WMS) as well as the emerging standard, Web Coverage Processing Service (WCPS). The WCPS standard allows the running of ad-hoc queries on any of the data stored within Rasdaman, creating an infrastructure where users are not restricted by bandwidth when manipulating or querying huge datasets.

The ESA Ocean Colour - Climate Change Initiative (OC-CCI) project (<http://www.esa-oceancolour-cci.org/>), is producing high-resolution, global ocean colour datasets over the full time period (1998-2012) where high quality observations were available. This climate data record includes per-pixel uncertainty data for each variable, based on an analytic method that classifies how much and which types of water are present in a pixel, and assigns uncertainty based on robust comparisons to global in-situ validation datasets. These uncertainty values take two forms, Root Mean Square (RMS) and Bias uncertainty, respectively representing the expected variability and expected offset error.

By combining the data produced through the OC-CCI project with the software from the EarthServer project we can produce a novel data offering that allows the use of traditional exploration and access mechanisms such as WMS and WCS. However the real benefits can be seen when utilising WCPS to explore the data .

We will show two major benefits to this infrastructure. Firstly we will show that the visualisation of the combined chlorophyll and uncertainty datasets through a web based GIS portal gives users the ability to instantaneously assess the quality of the data they are exploring using traditional web based plotting techniques as well as through novel web based 3 dimensional visualisation.

Secondly we will showcase the benefits available when combining these data with the WCPS standard. The uncertainty data can be utilised in queries using the standard WCPS query language. This allows selection of data either for download or use within the query, based on the respective uncertainty values as well as the possibility of incorporating both the chlorophyll data and uncertainty data into complex queries to produce additional novel data products. By filtering with uncertainty at the data source rather than the client we can minimise traffic over the network allowing huge datasets to be worked on with a minimal time penalty.