

A software architecture for seasonal climate forecasts in the tropical Pacific

Andrew Charles¹, David McClymont², Roald de Wit¹, David Jones¹

1. Australian Bureau of Meteorology,
2. DHM Environmental Software Engineering Pty. Ltd.
Email: a.charles@bom.gov.au

Abstract: The high predictability of seasonal climate in the tropical Pacific provides opportunities for using seasonal forecasts to improve the resilience of climate sensitive sectors throughout the region. Since 2004 the Pacific Island-Climate Prediction Project (PI-CPP) managed by the Australian Bureau of Meteorology (BoM) has built seasonal prediction capabilities within National Meteorological Services (NMS) of Pacific Island countries through the development and provision of decision support software and training. The software, SCOPIC (Seasonal Climate Outlooks for Pacific Island Countries) uses a statistical approach to generate seasonal outlooks based on discriminant analysis using relationships between local predictand variables and ENSO-based (El Niño Southern Oscillation) climate drivers (the predictors). Now six years on, the BoM is looking to better capture sources of predictability and adjust for the negative impacts of climate change on statistical models.

BoM are preparing the transition from statistically based seasonal climate forecasts to forecasts derived from dynamical physics-based coupled ocean-atmosphere general circulation models (CGCM). There are several barriers hindering the integration of dynamically based seasonal forecasting practices into NMS operations. Two key issues are that the forecasts require super-computing capabilities to generate numerical predictions and methodologies for the interpretation of ensemble outputs as probabilistic forecasts are evolving and this is an active and specialised area of research. Therefore, a multi-pronged approach is being pursued to integrate these forecasts into tools that are readily accessible to the NMS. Currently, BOM generates the CGCM output from which regional and point-based dynamical seasonal outlooks are generated and calibrated against observational datasets before uploading to a data server. The outlooks are then published to the Internet using an interactive web-based map interface. The SCOPIC software itself has been extended by (1) using web-services to directly integrate CGCM based probabilistic forecasts into its reporting tools, and (2) by including CGCM climate predictors as input to the statistical discriminant analysis model to increase useful forecast lead-times. The underlying raw data is stored as netCDF and is served through an open standards based web service (OPEN DAP). Exposing the raw outlook data through services using open standards has the advantage of giving users the freedom to extract all of the relevant data and verification scores, facilitating the ingestion into other decision-support systems, analysis and mapping tools. The web portal displaying the seasonal outlooks utilises a client side Javascript map interface (OpenLayers), dynamically displaying images generated by a custom web mapping service (WMS). Interactive mapping provides a sense of locality and user control, opens up new opportunities for data 'mash-up', and for the rapid display of outlook data and skill information in a way that is controlled by the end user. The site-based outlooks are made available as XML-based web services for input into SCOPIC's reporting tools. Publishing outlooks in this way provides for future flexibility in the display of outlooks.

SCOPIC displays the same location-specific content as the portal in a format more familiar to its users and that can be directly compared against the statistical outlooks. SCOPIC has also been extended by integrating new predictors based on CGCM predictions of climate drivers. This statistical-dynamical approach to forecasting potentially allows prediction at longer lead-times. These modifications to SCOPIC have been designed to be extensible to allow the integration of outputs from different dynamical models and future versions of BOM's CGCM. Managing the transition from statistical to dynamical seasonal forecasting is a complex task. The use of open standards, interoperable systems and simple, clean designs simplify the challenge of integrating data from multiple streams into usable seasonal forecasts. This paper provides an overview of the architectural, design and technology decisions made in the development of systems to support this and future climate prediction projects for the Pacific.

Keywords: *Open Source, Statistical-Dynamical, Seasonal Forecasting, Services Architecture*

1. INTRODUCTION

Seasonal climate outlooks are used by Pacific Island Countries (PICs) to inform decision across a range of sectors including water management, agriculture and hydroelectricity. Statistically based seasonal outlooks use the El Niño-Southern Oscillation (ENSO) based predictors to compare the current observed ENSO state and the historical relationship between ENSO and local climate variables. Statistical forecasting depends on the assumption of stationary relationships between predictors and predictands (the variable we wish to predict). In a changing climate, these relationships may 'trend' with time, reducing the effectiveness of statistical schemes.

CGCMs solve a set of dynamical equations ('primitive equations') to project the current analysed state of the ocean-atmosphere system into the future. The primitive equations are not sensitive to a particular climate regime, and so dynamical models are less compromised by climate change than statistical models. CGCMs allow longer lead-time prediction of major climate drivers like ENSO, and therefore can support longer lead seasonal forecasts. CGCMs explicitly take into account climate processes that are important for seasonal climate prediction, and are not constrained by what has occurred in the past. CGCMs can potentially provide more detailed seasonal climate information (For further background see Anderson: 'Overview of Seasonal Forecasting' p45 in Troccoli et al 2005).

The Pacific Adaptation Strategy Assistance Program (PASAP) has been developed with the aim of strengthening the capacity of PICs and Timor Leste to adapt to climate change through the provision of information, tools and training. One facet of this is the improvement of seasonal forecasts in partner countries in a project led by BoM. Having identified dynamical model-based seasonal forecasts as the key means by which to strengthen climate prediction capabilities in partner countries, the next step was to develop the systems and tools to achieve this goal.

SCOPIC is a PC Windows-based decision support software with features to analyse and make statistical predictions for arbitrary time series, currently used by the NMS of PICPP partner countries. Seasonal outlooks generated by SCOPIC are used by PIC meteorological services for publicly issued forecasts, and in turn by industries and agencies sensitive to climate variability. The group of users is diverse, with a range of levels of atmospheric science background, probability and decision theory. The introduction of SCOPIC has broken new ground in the development of an understanding of probability forecasts and an appreciation of the decisions that can be influenced by climate forecasts. SCOPIC is used in 13 countries, some for over 6 years. The popularity of SCOPIC means integration of new seasonal forecasts into SCOPIC has been a focus of this activity.

This paper explains the systems we developed to integrated a new, and fundamentally different model into an existing forecasting and decision-support system. Section 2 outlines the overall system architecture and guiding design principles. Section 3 outlines our approach to storing and publishing the model-based outlook data. Section 4 details the means by which CGCM outlooks were incorporated into SCOPIC. Section 5 gives an overview of the design of the PASAP Seasonal Outlook Portal, a rich web interface providing access to the outlooks.

2. AN ARCHITECTURE FOR LINKING MODEL-BASED OUTLOOKS TO RICH USER-FRIENDLY DECISION-SUPPORT SYSTEMS

The system is composed of size main elements, numbered in Figure 1:

1. The CGCM is run by the BoM out for 9 months every day.
2. Forecast products are generated from dynamical model output using data analysis software. The resulting derived forecast products are persisted in self-describing files with additional metadata to support the clients that deliver the outlooks.
3. Forecast data is exposed via a data server.
4. Scheduled processes access and reformat the data for SCOPIC access.
5. Custom web services use the data server's interface to the forecast data to provide maps, data, line plots.
6. The PASAP Seasonal Outlook Portal consumes the outputs of the custom web services, and displays model-based outlooks as overlays on dynamical maps and standard plots.

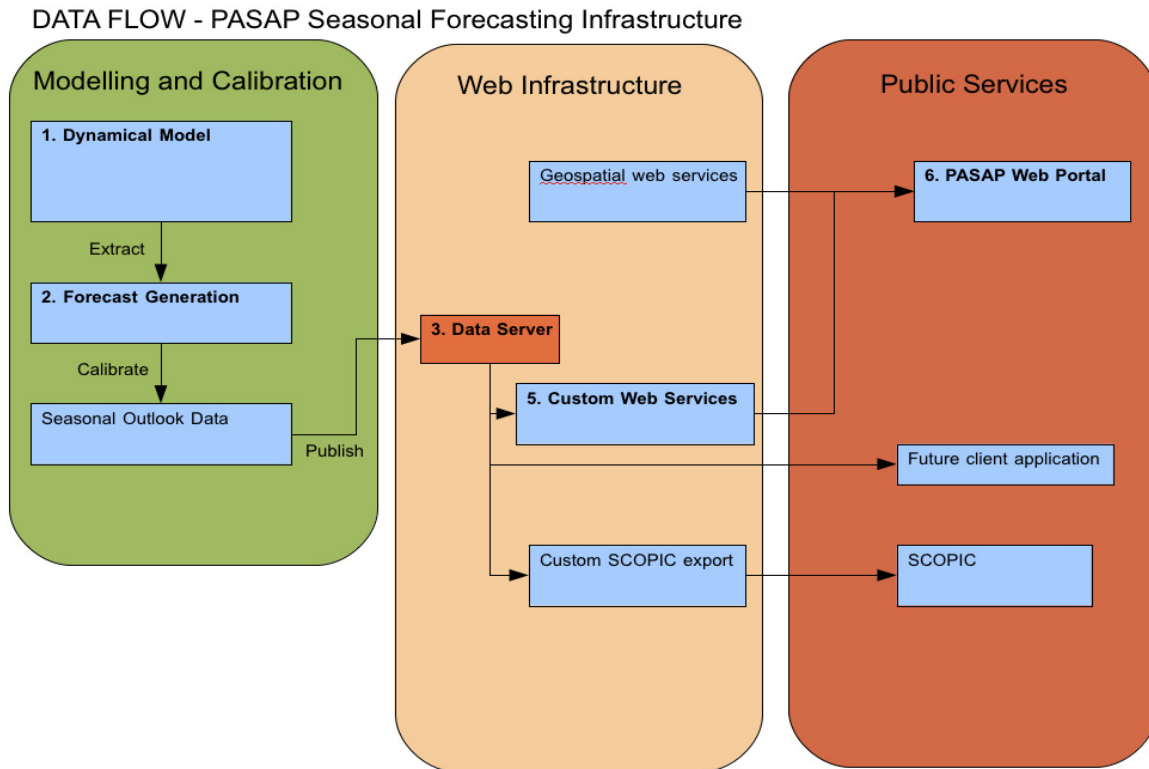


Figure 1. High level architecture of the seasonal forecast generation and publication system.

Our system for distributing model-based outlook data to client applications and models is based on the use of open standards, clean interfaces, sufficient metadata and modular components. Figure 1 shows the main components and their connections, which we will outline in this section. Systems need to be interoperable, and the use of open standards supports this interoperability. Open standards means that where a standard exists for a certain type of data, figure or service that we should prefer to use it rather than create a new *ad hoc* format or interface. Sufficient metadata reduces the amount of intelligence that must be programmed into the client, and supports the goals of modularity and clean interfaces by ensuring all data provided is described as completely as possible. Modular components means that system components have a single function or clearly defined set of functions such that they may be replaced without major impact on other system components. Clean interfaces means that coupling between system modules is kept to a minimum, and that system modules communicate with each other as far as possible using the standards described above.

3. THE DESIGN OF MODEL DATA AND DATA ACCESS SYSTEMS TO SUPPORT INTEGRATION INTO DECISION-SUPPORT SYSTEMS AND DOWNSTREAM MODELS

It is a requirement that BOM to support the integration of model outputs with currently existing clients (SCOPIC) and in development clients (PASAP Seasonal Outlook Portal) as well as supporting integration with clients that do not yet exist. The integration of model outputs into arbitrary decision-support systems and downstream models is supported by providing the model output, and post-processed model-based forecasts in standard formats, and exposing web services that provide access to data and meta-data via clearly defined protocols.

The system has been developed to ingest output data from the Predictive Atmospheric-Ocean Model for Australia (POAMA) GCM (Wang *et al.*, 2008). Model output data is in a self-describing format called netCDF (network Common Data Format, Rew *et al.* 1990), a widely used standard format for gridded meteorological and oceanographic data which has recently been formally adopted the Open Geospatial Consortium (OGC).

The generation of seasonal outlooks from CGCM outputs requires a post-processing, or 'calibration' step to be directly used as forecasts of real fields. The forecasts are corrected for mean state bias and over-confidence in the ensemble distribution. Conceptually this calibration step can be considered the application of a statistical model to the direct model output. There is a range of outstanding scientific questions concerning optimal strategies for bias-correcting model outputs in a way that preserves the predictable signal, and we expect this component to change in the future. Forecast generation is accomplished using readily available tools such as netCDF Operators (Zender, 2008) for simple procedures, and numerical python (van der Walt *et al.*, 2011) modules for more complex transformations. By storing the

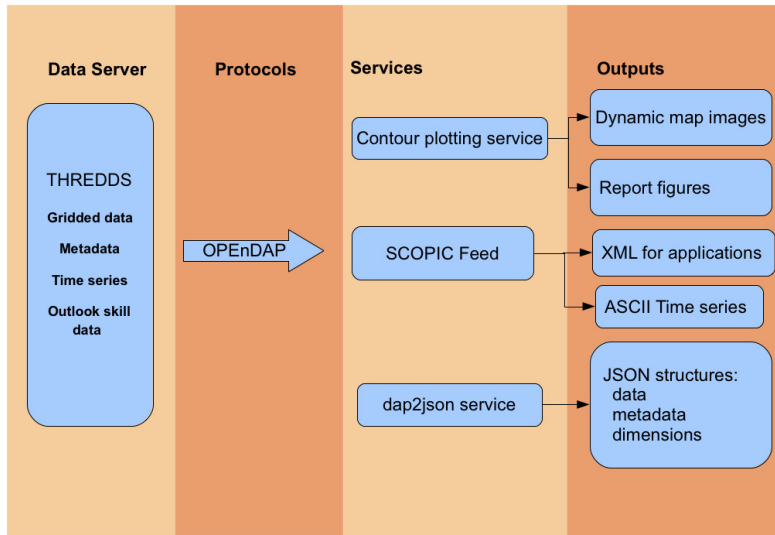


Figure 2. Multiple output formats are supported by layers over an OPeNDAP enabled data server.

generated forecasts in the same accessible, metadata rich format as the raw model output they are easily ingestible into downstream clients, whether these are simple viewers or more complex models that use the calibrated model data as one input among many.

Dynamical model outputs may be used to drive downstream models (for example hydrological models) for which statistical adjustment of the model output may not be useful. Our design facilitates this by allowing for programmatic access to raw model outputs.

Seasonal forecasts of broad-scale ocean and atmospheric mean states, major climate drivers and point-based outlooks calibrated to observational sites are generated. The final outlooks are data products, not graphical products. The format of the output is not dependent on the particular dynamical model, or even that the model is dynamical: the forecast is simply an ensemble time series of gridded data with descriptive metadata. Metadata is preserved as data is processed and new metadata added to describe transformations. Metadata is curated by an additional process to ensure the metadata associated with the generated forecasts is complete and describes the forecast generation process itself.

Integration of CGCM outputs into models and decision support systems necessarily involves remote access to model data because the computational footprint of a CGCM prohibits their practical execution on desktop computers. Remote access to the model data is provided by the THREDDS data server. This provides access to compatible datasets via a number of standard protocols over HTTP, in particular the DAP standard defined in ESE-RFC-004.1.1 (Gallagher et al 2007). Multiple layers of access to data and data-derived products are then built on this foundation (Figure 2).

THE INTEGRATION OF MODEL OUTPUTS INTO EXISTING MODEL-BASED DECISION-SUPPORT SYSTEMS

SCOPIC has been a part of the climate forecasting landscape in NMSs in the Pacific since 2004. It is deployed in 13 countries and has been used for over 6 years. The wide adoption and popularity of SCOPIC makes integration with SCOPIC a requirement of any attempt to engage with and enhance seasonal forecasting in Pacific country NMSs.

SCOPIC serves multiple functions. Firstly it provides tools for time series visualisation and analysis of predictor and predictand data. Secondly it contains an implementation of a robust statistical model for using relationships between variables to generate seasonal timescale predictions, and tools to verify the historical performance of these predictions. It also contains features to support the publication and communication of the generated outlooks through the generation of high quality interactive chart and map outputs, and automated reporting in English and local languages such as Samoan and Tongan. Finally, it contains tools for analysing and monitoring current and historical droughts.

We have integrated the CGCM outputs into the SCOPIC tool in two ways (Figure 3). Firstly we use SCOPIC as a publishing platform to provide pre-generated seasonal outlooks using statistical corrections computed at the BoM. This allows for the straightforward comparison of outlooks generated in-country by SCOPIC with the GCM-based outlooks. Data delivery is accomplished using the web services described in the previous section, whereby a process accesses the data through OPeNDAP and generates outlook data in XML format that is later read and presented by SCOPIC (Figure 2).

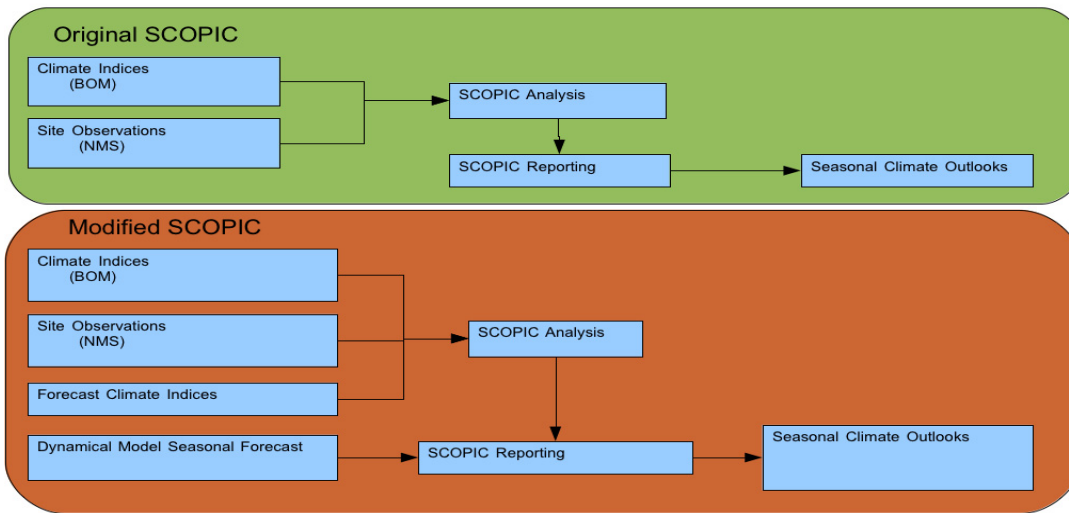


Figure 3. Modifications to SCOPIC to integrate CGCM based outlooks.

Secondly we enabled the ingestion of the dynamical model output by the statistical model at the heart of SCOPIC. The model outputs are imported as predictor data in SCOPIC, and a statistical relationship between the dynamically-derived predictor and the predictand is computed and presented as probabilistic forecasts. This process is equivalent in character (but not in the specific statistical method) to the calibration performed on model output in the BOM before it is considered a 'forecast'. Building this capability into the SCOPIC software enables scientists in partner countries to conduct their own investigations of the efficacy of combining dynamical model and observational predictors ('statistical-dynamical' forecasting).

RICH WEB INTERFACES FOR PUBLISHING MODEL OUTPUTS AND INTEGRATING MODEL OUTPUTS WITH DOWNSTREAM SYSTEMS

SCOPIC provides strong support for the integration of point-based outlooks and the generation of statistical outlooks using dynamically predicted climate drivers, but little support for the display of broad-scale gridded outlooks. These broad-scale patterns are a valuable means of evaluating and understanding the outputs of CGCMS, side-by-side with SCOPIC, to allow for access to new outlooks and facilitate training.

We have developed a rich web-based interface to provide a one-stop shop for access to dynamical model-based outlooks. Based on experience with experimental CGCM based outlooks and with operational seasonal outlooks issued by the BoM's National Climate Centre we carefully crafted an intuitive user interface for the most common use cases. An iterative approach to the development of the web portal UI included testing of early development versions of the portal with users at the PASAP/PICPP 2010 workshop in Auckland and a series of in-country visits in 2011. These sessions validated the overall UI design and provided valuable feedback for improvements.

The central component of the web portal is a dynamic map view with options to see model-based outlooks overlaid on a base map. Selected forecast stations can be visualised and queried, as well as climate indices based on ENSO.

The interface is a thin client built on top of BOM.Map, an in-house developed and maintained JavaScript library that integrates a number of popular open source JavaScript libraries. The mapping component of this package is OpenLayers, a popular open-source mapping library. The user interface is driven by ExtJS, a powerful library that makes it easy to create cross-browser UIs. UI Components such as tabs, panels and dialogs are created using the ExtJS JavaScript framework, using jQuery and GeoExt to integrate OpenLayers and ExtJS (Figure 4.).

On the server side, we expose some WSGI web services. The most important ones are a figure plotting script and a custom built WMS service that dynamically generates gridded and contour plots used as map overlays by the mapping framework. Both fetch the data on the fly from the OPENDAP service and create images using the python-matplotlib (Hunter, 2007) library. Another service converts metadata from OPENDAP to a JSON format.

Graphical displays of broad-scale, point and climate driver forecasts are generated. The web application displays the contextual information provided as meta-data by the data service layer.

Much of the information required by the client pertains to the forecast variables that are available, where the data is located and how they are to be displayed in the client. Maintaining this logic in the client requires a coupling between the data that is produced and the data that is asked for by the client. It is possible to remove this coupling by being

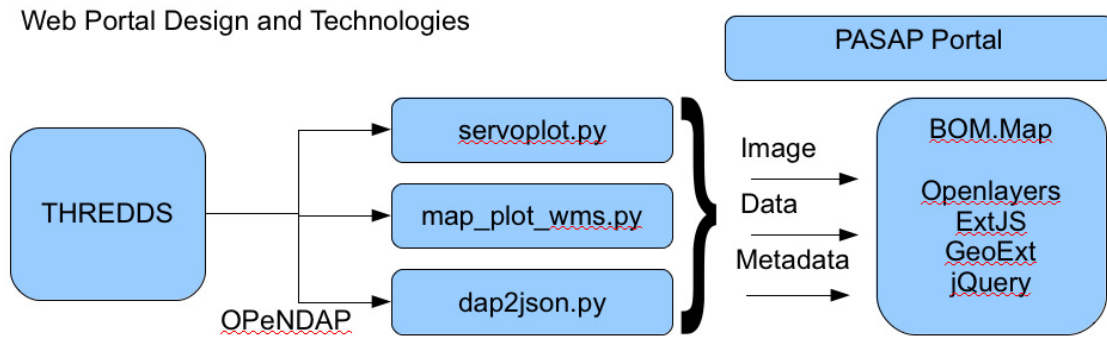


Figure 4. PASAP seasonal outlook web portal's main technologies.

prescriptive about what is to be displayed on the client in the data service layer. However, if it is desired that multiple clients could be developed, each providing a specialised view into the forecast data and potentially performing computational value-adding then it is clear the client must contain this information. The location of this information on the client side has an upside, as there is a clear potential to build flexibility for the user to add new data sources and customise graphic displays.

The access to the data provided by the data service layer allows for the future design of web clients that perform computational value-adding using processing services, for example the ingestion and subsequent combination and calibration of multiple selected models. Access is given not just to application graphics but also to outlook data. User-friendly options for data extraction from the web portal are provided to support users of the range of tools from Excel to R.

The web portal consumes the outputs of the custom web services, with spatial visualisation provided by Javascript libraries GeoExt, ExtJs and OpenLayers. Mapserver provides geographic overlays and the spatial grid.

The integration of data into a dynamical mapping tool provides opportunities for data mash-up in which data from different sources is displayed in composite. The provision of geospatial information in such a way that data from multiple sources can be integrated has been identified as an area for the development of new 'killer apps' (Abbott, 2011). For example, one potential future application for seasonal forecasting might be the display of agricultural or fishery yield data overlaid with outlook reliability data.

CONCLUSION

The technical challenges involved in building climate prediction capacity in Pacific partner country NMSs revolve around the challenge of integrating heterogeneous models into existing decision support systems while building the tools to support future systems. We have described a set of technical approaches to this problem in the context of the development of seasonal prediction services in the Pacific region. We have used these approaches to develop software systems which support current requirements and have built in interoperability to support future requirements

Data availability is key to supporting future extensions and new applications, as is good quality metadata that allows other layers such as the SCOPIC xml outlook generator to be developed rapidly with minimal knowledge of the source dataset. Throughout the development process we have obtained feedback from end users and adjusted the product in response. Success of the project will ultimately be determined by the level of utility Pacific Island country meteorological services and the communities that use the outlooks are able to derive from seasonal outlooks made available through the system.

ACKNOWLEDGEMENTS

This work has been undertaken through the Pacific Adaptation Strategy Assistance Program, which forms part of the Australian Government's International Climate Change Adaptation Initiative, funded by the Department of Climate Change and Energy Efficiency. SCOPIC was developed under the Pacific Island Climate Prediction Project, an AusAid funded initiative. We also gratefully acknowledge other project team members Yuriy Kuleshov, Harry Hendon, Eun-pa Lim, Andrew Cottril, Sally Langford and Janita Pahalad. We are grateful to Dr Brett Anderson for reviewing this manuscript and in advance to the two anonymous reviewers whose comments have/will greatly improve the manuscript.

REFERENCES

- Abbott, S., (2010), Enabling and Encouraging Transparency in Earth Science Data for Decision Making, Abstract U12A-02 presented at 2010 *Fall Meeting, AGU, San Francisco, Calif., 13-17 Dec.*
- Domenico, B., Caron, J., Davis, E., Kambic, R., Nativi, S., (2002). Thematic Real-time Environmental Distributed Data Services (THREDDS): Incorporating Interactive Analysis Tools into NSDL. *Journal of Digital Information*, Vol 2, No 4.
- Eaton et al, (2010), NetCDF Climate and Forecast (CF) Metadata Conventions Version 1.5. <http://cfpmdi.llnl.gov/documents/cf-conventions/1.5/cf-conventions.pdf>, 25 October.
- Gallagher, J., Potter, N., Sgouros, T., Hankin, S., Flierl, G. (2007), The Data Access Protocol – DAP 2.0. ESE-RFC-004.1.1. *NASA Earth Science Data Systems Standards*. <http://www.esdswg.org/spg/rfc/ese-rfc-004>
- Wang, G., Alves, O., Hudson, D., Hendon, H., Liu, G., and Tseitkin, F., (2008). SST skill assessment from the new POAMA-1.5 System. [BMRC Research Letter No.8.](#)
- Hunter, J. D., (2007), Matplotlib: A 2D Graphics Environment, *Computing in Science & Engineering*, Vol. 9, No. 3. pp. 90-95.
- Rew, R. K. and Davis, G. P., (1990). NetCDF: An Interface for Scientific Data Access, *IEEE Computer Graphics and Applications*, Vol. 10, No. 4, pp. 76-82.
- Rew, R., (2010). CF Metadata Conventions: ES-DS-RFC-021v0.02 *NASA Earth Science Data Systems Standards*. <http://www.esdswg.org/spg/rfc/esds-rfc-021>
- Troccoli, A., Harrison, M., Anderson, D.L.T., and Mason, S. J., (2008). *Seasonal Climate: Forecasting and Managing Risk*. Springer.
- Van der Walt, S., Colbert, S.C., Varoquax, G., (2011) The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science and Engineering*.
- Zender, C. S. (2008), Analysis of Self-describing Gridded Geoscience Data with netCDF Operators (NCO), *Environmental Modeling and Software*, 23(10), 1338–1342.

SOFTWARE GLOSSARY/REFERENCES

- SCOPIIC: Seasonal Climate Outlooks for Pacific Island Countries (SCOPIIC). Version 2.7 [Computer Software] (2011). Australian Government, Bureau of Meteorology. Available from <http://www.bom.gov.au/climate/pi-cpp>
- THREDDS: Developed and maintained by Unidata, THREDDS implements the DAP protocol providing a simple catalogue and access service for netCDF and other self-describing files. <http://www.unidata.ucar.edu/software/tds/>
- WMS: OpenGIS Web Map Service (WMS) Implementation Specification. http://portal.opengeospatial.org/files/?artifact_id=14416
- OPeNDAP: Open-source Project for a Network Data Access Protocol. <http://www.opendap.org/>
- NetCDF: A set of conventions for binary files designed for representing gridded earth science datasets. In addition to references above see Open Geospatial Consortium (OGC), OGC Network Common Data Form (NetCDF) Core Encoding Standard version 1.0 Standards document 10-090r3. 2011. <http://www.opengeospatial.org/standards/netcdf>
- Openlayers: An opensource javascript library to load, display and render maps from multiple sources on web pages. openlayers.org
- SCOPIIC: Windows application developed for the generation of seasonal outlooks in Pacific Island Countries using linear discriminant analysis, developed under the PI-CPP project.
- ExtJS: a cross-browser JavaScript library for building rich internet applications. <http://www.sencha.com/products/extjs3/>
- R: A popular programming environment and language for statistical analysis.
- SciPy: Scientific modules for the Python programming language. Jones, E., Oliphant, T., Peterson, P., and others, 2001. [SciPy](http://www.scipy.org): Open Source Scientific Tools for Python. <http://www.scipy.org>