A revolutionary approach to multi-dimensional data access of gridded datasets using current and proposed OGC Web Service Standards (The MetOcean Application Profile)

Paul Hershberg and Steve Olson
National Oceanic and Atmospheric Administration (NOAA)
National Weather Service (NWS)
Meteorological Development Laboratory (MDL)
Office of Science, Technology and Integration (STI)

Peter Trevelyan
UK Met Office (UKMO)
FitzRoy Road, Exeter

1. Introduction

The Open Geospatial Consortium (OGC) is the authoritative standards body providing governance on geospatial Web Services Standards. Today, there are Web Service standards most notably for Web Coverage Service (WCS), Web Feature Service (WFS), Web Map Service (WMS), Web Map Tile Service (WMTS), Web Processing Service (WPS) and Web Coverage Processing Service (WCPS). Web Services are designed for machine to machine communication of user metadata and data requests. The main focus of this paper is on the WCS, with proposed extensions to the WCS standard.

In order to understand how Web Coverage Services work, it's important to first understand what a "coverage" is. A ***coverage*** is a digital representation of some "space-time varying phenomenon", such as 2D remote sensing imagery, 3D x/y/t satellite image time series and x/y/z geophysical data, as well as 4D x/y/z/t meteorological and oceanographic data. The WCS Core used today establishes basic spatial and temporal extraction of gridded datasets. The WCS Core also has three user request types: *GetCapabilities*, *DescribeCoverage* and *GetCoverage*. *GetCapabilities* returns service properties and data offered by the service. *DescribeCoverage* returns a description of a specific coverage. *GetCoverage* returns the data coverage (or part thereof). With each request type, there are both request and response documents.

* Corresponding author address: Paul Hershberg, Room 10388, National Oceanic and Atmospheric Administration, 1325 East West Highway, Silver Spring, MD 20910; e-mail: Paul.Hershberg@noaa.gov

2. WCS Core issues and limitations

While access to meteorological and oceanographic data by data providers using OGC Web Services continues to grow, so does the volume and size of these offerings. The practice of transferring large amounts of gridded data across networks via OGC Web Coverage Services is becoming increasingly difficult. The main reasons for this include the ever increasing amounts of storage for meteorological and oceanographic data, insufficient computer resources to process user requests, and insufficient bandwidth to disseminate consumers requests for data.

One technique to help alleviate issues with resource allocation is the use of the WCS core functionality i.e. subsetting.

Subsetting is beneficial not only to consumers but also data providers. There are two types of subsetting, which can be combined: ***Trimming*** extracts a sub-area of a coverage indicated by a bounding box; the result has the same dimension (i.e., number of axes) as the original coverage. ***Slicing*** performs a cut at the position indicated, thereby reducing the dimension of the result coverage. However, the key to making subsetting work is optimizing it around its usage, e.g. point time-series, vertical profile time-series, grid series, etc, to meet the user's needs.

Even with subsetting, there are still a number of limitations and shortcomings with today's WCS core functionality. The first stems from the fact that meteorological and oceanographic data is inherently 4D, yet WCS coverages are often expressed as 2D coverages. Moreover, the size, structure, and number of web service responses with 2D coverages can get unwieldy. Secondly, interoperability of disparate web services is also an important goal, resulting in a growing need to describe 4D meteorological and oceanographic data in a community-based controlled vocabulary with additional metadata. Furthermore, today's WCS core functionality lacks specific feature and coverage types applicable for the meteorological and oceanographic communities, particularly those with use cases for data requests of trajectories, multiple points and polygons.

This paper will describe how the MetOcean Application Profile (OGC document #15-045, subsequently referred to as the MetOcean Profile) addresses all of these issues and shortcomings. The use of the MetOcean Profile represents a paradigm shift from traditional ways people think of WCS Coverages. For example, the whole atmosphere can now be depicted as a multidimensional simulation with meteorological properties (e.g. temperature, wind, and humidity). Coverages are no longer defined as these properties, but instead are defined by dimensions that can contain one or more of these properties. This allows sampling of the atmosphere as a true 4D object (hyper cube), along any of the 4 axes. In short, the MetOcean Profile will allow consumers the ability to not only extract 4D data over geography, time, and altitude, but also perform the extraction in complex patterns. All these data can now be easily represented and referenced using the MetOcean Profile. Moreover, use of the MetOcean Profile will allow users access to even higher dimensional weather characteristics, such as emergent forecast probability based on an ensemble forecast, which then can be used to understand the likelihood of certain weather conditions happening.

3. The need for a multi-dimensional approach

For meteorological and oceanographic gridded datasets, the typical dimensionality of a coverage is 4D. However, data providers offering up WCS gridded data typically store meteorological and oceanographic information in 2D coverages that are consistent in x/y/z axes. Thus, in order for a single WCS Core *GetCoverage* request to return multi-dimensional data, the coverage itself has to be multi-dimensional. This makes a *GetCoverage* operation much easier and more efficient if the geospatial object is 4D requiring only one request. The task is much more complicated when 2D coverages are used as requests need to be made for ***each*** 2D coverage comprising the 4D coverage. Prior versions of the WCS Core did not contain functionality that made identifying a 4D coverage from the numerous 2D coverages easy. However, the latest version of the WCS Core (2.1) ***does*** make identifying a 4D coverage from the 2D coverages possible (See section 4.2.2 for more detail). Furthermore, as long as all the 2D coverages share the same horizontal, vertical and temporal domains, they can be combined into a single 4D coverage. This drastically reduces the number of coverages in hand. Figure 1 denotes the visualization of a stack of 2D coverages as ***one*** true 4D coverage object.

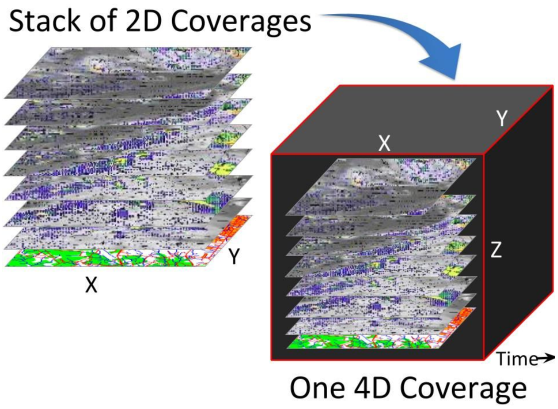

Fig 1. Coverages Paradigm Shift

Additionally, multi-dimensional coverages allow for the development of complex extraction patterns aligning with the community's needs using a simple web based API. The NWS and UK Met Office are collaborating on the development of new operations under the MetOcean Profile specifically for the extraction of trajectories and polygons (Sections 4.4 and 4.5). Extracting these new shapes represents a subsetting that is tailored to the MetOcean community.

4. Driving forces behind the need for a *profile* and a *standard*

It is important that consumers of gridded data be agnostic regarding the source of the data a web service is returning. As we move to an era of increased global cooperation and cross-domain utilization of information, harmonization of web services and the extraction of meteorological and oceanographic gridded data will be needed. In fact, harmonization is proving to be essential between organizations such as the Single European Sky Air Traffic Management Research (SESAR) project, the United States Next Generation Air Transportation System (NEXTGEN), and many other areas across the marine, defense, and general transport domains.

When an application *targets* a certain group, it's' called a "*profile*" of that group. In this case, that group is the meteorological and oceanographic user community. Hence, the MetOcean Profile is the first attempt at tailoring the OGC's web coverage services for the purpose of providing MetOcean specific data. This methodology includes both the description of a true 4D gridded data cube *and* the use of metadata describing the cube. The result is a *standardization* of the cross-domain utilization of meteorological and oceanographic information through the use of digital data exchange. Consequently, international organizations involved with meteorological and oceanographic data exchange can leverage the MetOcean Profile as the best *standard* to improve interoperability.

5. The MetOcean Application Profile: How it broadens the WCS Core Functionality and its Benefit to End Users

The MetOcean Profile can be used as an international standard, because it ***extends*** the OGC WCS Core, which is itself an international standard. Figure 2 details how the MetOcean Profile extends WCS core functionality, with the WCS Core remaining unchanged. It also shows that the MetOcean Profile incorporates all prior service extensions to the WCS Core (eg *Interpolation* and *Scaling*; these *"extensions"* to the WCS core are more generic in that they are not tailored to any one user group).

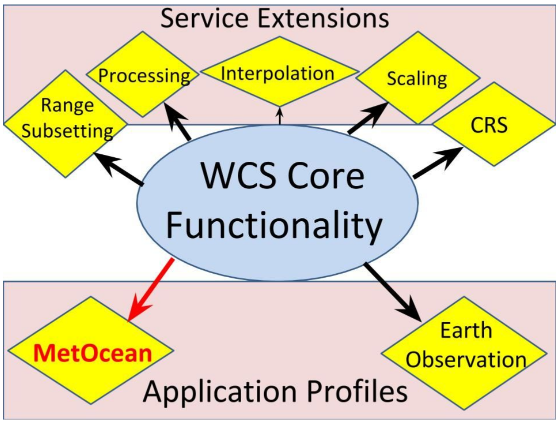

Fig 2. Adding to the WCS Core Functionality via Extensions and Profiles

A *profile* is the method a specific user community adds new functionality through the use of "hooks". The MetOcean Application is a "profile" because not only does it extend the WCS Core, but it is also designed to meet a certain group's needs. In theory, this can be applied to other disciplines outside of the MetOcean community as they add functionality.

The following subsections describe the additions to the WCS core functionality the MetOcean Profile provides.

5.1 Addressing Multi-dimensionality through *CoverageCollections* and the *DescribeCoverageCollection* operation

For the MetOcean community, being able to describe a computer simulation as a collection of coverages is a very powerful concept. Many features of the individual coverages may be shared with the collection. Therefore, a collection can be treated as a single geospatial object. The MetOcean Profile introduces the CoverageCollection as a single, uniquely identified resource specifying the member coverages. Each coverage within a CoverageCollection share their horizontal and temporal domain. Each coverage will also share characteristics such as provenance and a coordinate reference system (CRS). Figure 3 is a graphical depiction of how CoverageCollections are structured for the NWS GFS and HRRR numerical weather prediction (NWP) models.

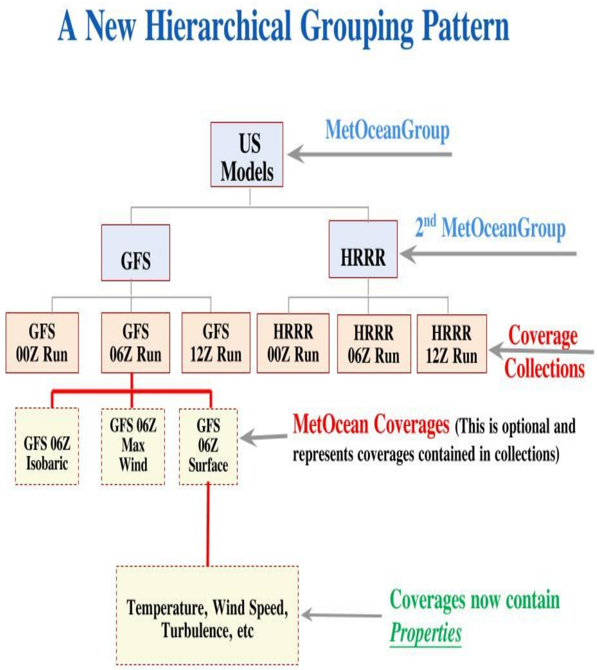

Fig 3. MetOcean and a New Structural Data Pattern

CoverageCollections (see Figure 3 center row) represent the different model run times such as the GFS 06Z run. Subsequently, the individual 4D coverages comprising each model run time collection represent different vertical levels. For example, in Figure 3, the GFS 06Z model run is a coverage collection comprised of individual coverages depicting the model's different vertical CRS's (i.e. GFS 06Z at isobaric, surface, and maximum wind levels). Finally, it is important to note in Figure 3 (bottom row) that particular meteorological and oceanic elements such as humidity or salinity are now *properties* of specific coverages, and not coverages themselves. With this new organizational capability that CoverageCollections provide, the MetOcean Profile supports a very powerful way of grouping many MetOcean datasets including Radar Mosaics, NWS Model Simulations, Climate Simulations, and Climate Observations.

To accommodate CoverageCollections, the MetOcean Profile also introduces the new operation ***DescribeCoverageCollection.*** It returns a list of all the contained coverages for a single CoverageCollection identifier. Examples of a *DescribeCoverageCollection* request and response are included in Section 7.1.

5.2 New MetOcean *Metadata*

The MetOcean Profile adds and structures metadata for the *GetCapabilities* and *DescribeCoverage* responses.

5.2.1 Additional Metadata in the *GetCapabilities* response

Currently, when making a *GetCapabilities* request, the WCS Core allows for only the return of individual coverages contained on the server. The MetOcean Profile adds to the core by supporting an extended GetCapabilities operation allowing for two new distinct metadata patterns to be returned: Groups and Collections.

5.2.1.1 Groups

Meteorological and oceanographic data is by nature hierarchical and the ability to group entities together is important. Thus, a set of simulations may be clustered together to form a logical group.

The MetOcean Profile introduces the term ***Groups*** as a way of structuring the *GetCapabilities* response to provide a hierarchical way of nesting "services". *Groups* simply allow for the organization of data, and, if required, have service endpoints. There is nothing geospatial about *Groups*. The main benefit of *Groups* is in creating a structure that reflects the organization of the intended use of data, rather than its underlying structure. Figure 3 depicts a top level *Group* as "US Models" with two subsequent *Groups* as the "GFS" and "HRRR" NWP models. The MetOcean Profile supports an organized response from a GetCapabilities request by utilizing *Groups*.

5.2.1.2 Collections

As noted in Section 4.1, the MetOcean Profile uses the CoverageCollections method to organize data. In addition, the *GetCapabilities* response has the option of only exposing CoverageCollections (i.e suppressing the individual coverage listing and further cutting the size of the response document). As previously denoted in Figure 3, there are only six CoverageCollections represented, one for each model and model run time (i.e. GFS 06Z). Also shown are three member 4D coverages per each CoverageCollection. Thus, a total of eighteen 4D coverages are indicated across the whole model run domain.

Returning CoverageCollections allows for a much smaller *GetCapabilities* response document, lessening the impact to data providers using OGC Web Services with the MetOcean Profile. In fact, a detailed analysis of the GFS model denotes for just one model run time that there are thousands of 2D coverages. So the use of both CoverageCollections and 4D coverages drastically reduces the size of the response. Clearly this is extremely important with respect to usability.

Another added benefit is that common metadata can be attributed to the entire CoverageCollection, instead of each member coverage, thus removing duplicate metadata information.

5.2.2 Additional Metadata in the *DescribeCoverage* response

Current observations are used as input to mathematical models to create simulations of weather, climate and ocean forecasts based on principles that are used to generate predictions. As such, the simulations have many properties that describe the spatial and temporal domain of the model. Thus, it makes sense to describe a MetOcean coverage that is a forecast as a simulated observation. Currently in a WCS Core *DescribeCoverage* response, the user does not receive information about the individual coverage in this manner. But the MetOcean Profile does this by defining metadata about each coverage in the *DescribeCoverage* response as a specialized Observations & Measurements (O&M) type (ISO 19156). Furthermore, not only does the MetOcean Profile leverage the O&M metadata, but it also *modifies* the metadata returned in a DescribeCoverage response. Below are a list of those modifications specifying how the response is altered in order to benefit the user.

First, for weather, climate, and ocean forecast simulations, the models used to derive them span the domain for which the simulation was made. The feature of interest is the entity about which the observation (forecast) is made. But for a forecast, the feature of interest may also be used to describe the computer simulation used to create the forecast. For this reason, the MetOcean Profile modifies the O&M feature of interest as a simulated process description in order to describe characteristics of forecast models used in the MetOcean community.

Second, in the case of the forecast simulation, many individual physical properties may be measured that are associated with a particular environmental domain (e.g. Meteorological Products, Hydrological Products, Space Products, Oceanographic Products, etc). These domains are defined in the World Meteorological Organization's (WMO) GRIB2 code tables. Use of WMO tables that are specifically referenced using links via the WMO registers allows for an authoritative vocabulary and source to be universally used. The MetOcean Profile takes advantage of this by allowing for both the O&M observed property and procedure to be referenced via WMO registers. All vertical CRS's common to the MetOcean community may also be referenced in this manner. The result is extensibility because the MetOcean Profile supports the use of controlled vocabularies through references, which are not hard coded. (*Note, at the time of writing only the WMO registers are properly supported. It is important to note that any file format (i.e. NetCDF) may be described in this way.*)

Third, the MetOcean Profile also supports Reference Time axes, allowing for analytics to be performed on analyses.

Furthermore, the MetOcean Profile takes advantage of the O&M result quality by adding a mechanism for quality control using a data *ResultMask*. In a *DescribeCoverage* response, this metadata is especially helpful in that it can denote, across the whole horizontal extent, where and/or when missing data occurs. The mask can also be used to indicate quality of data.

Finally, it is important to note that the prior version of the WCS Core, version 2.0, did not contain functionality that made identifying a 4D coverage from the many 2D coverages easy. However, the new WCS Core, version 2.1, ***does*** make identifying a 4D coverage from the 2D coverages possible through the use of OGC's Coverage Implementation Schema Version 1.1 (CIS1.1) (OGC 09-146r5). It is made possible because the CIS1.1 uses the *cis:generalGrid* type. This type allows for the characteristics of ***specific*** dimension axes to be described, including ***non-regular*** axes that define both the time and vertical domains that are so prevalent in the MetOcean community. There is a challenge when dealing with the non-regular vertical and temporal dimensions because they need to be enumerated. This is because these two dimensions often have varying time steps in the temporal domain, or varying distances from a specified height in the vertical domain. For this reason, it is ***imperative*** that the MetOcean Profile extend the WCS2.1 core functionality with CIS1.1 and not the WCS2.0 core functionality.

5.3 Effect of 2D coverages versus 4D coverages on web services: A Case study using the GFS model

In order to demonstrate the value of our approach, we captured data from the GFS model in a WCS and compared access with and without the MetOcean Profile. Specifically, we counted the number of coverages returned with the conventional 2D model versus the 4D model provided by MetOcean. The results showed that using 2D coverages required approximately 10,000 individual coverages comprising the entire GFS model for just one run time! Conversely, when using 4D coverages, the number of individual coverages comprising the same GFS model run time was reduced to just 6! The following sections describe how this reduction in number of coverages affects both the *size* of a GetCapabilities response and the *number* of GetCoverage requests.

5.3.1 Reducing the size of the WCS *GetCapabilities* response

In simple terms, a WCS *GetCapabilities* core operation returns service properties and data offered by the web service. The MetOcean Profile has made modifications to the core WCS *GetCapabilities* response. A WCS *getCapabilities* running the MetOcean profile will now return a list of coverages. From our GFS use case, when those coverages are 2D, a user querying a web service would find a *GetCapabilities* response document that included approximately 10,000 coverages and over 30,000 lines of XML code for just one GFS model run time! Conversely, when those coverages are 4D, the *GetCapabilities* response would include just 6 4D coverages and approximately 200 lines of XML code!. The reduction in the amount of lines of data returned from the server showcases the enormous increase in efficiency for OGC Web Services with the MetOcean Profile.

5.3.2 Reducing the number of WCS *GetCoverage* requests

The MetOcean Profile provides further benefits by reducing the number of GetCoverage requests needed to access a dataset. Because the MetOcean Profile supports a *GetCoverage* operation that now supports multi-dimensionality and 4D coverages, a *GetCoverage* request for data can be referenced with one *single* coverage identifier for a 4D object. In the GFS example above, using 4D coverages, a user set to request data for the entire GFS model would make only 6 *GetCoverage* requests to a WCS web service. Using 2D coverages, approximately 10,000 *GetCoverage* requests would be made to a WCS web service. No longer does it take a user thousands of individual *GetCoverage* requests to extract an entire GFS model run! This massive reduction in the number of requests can improve efficiency for both data servers and their customers.

5.4 New extraction patterns

The MetOcean Profile introduces two new operationsl named ***GetCorridor*** and ***GetPolygon***. Along with the new *DescribeCoverageCollection* operation, these three will complement the WCS Core *GetCapabilities*, *DescribeCoverage*, and *GetCoverage* operations. These two new operations, set for future release, will improve efficiency and support well-established requirements from the MetOcean user community.

5.4.1 *GetCorridor* Operation

The *GetCorridor* operation will extract data from a multi-dimensional coverage along a trajectory or corridor. Figure 4 depicts this extraction.

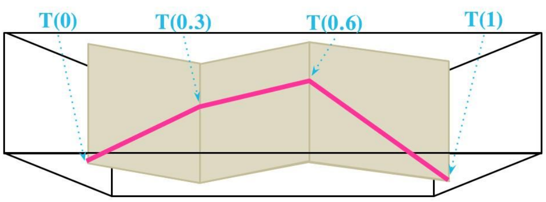

Fig. 4 *GetCorridor:* A New 4D MetOcean Data Extraction Pattern

The GetCorridor operation can yield efficiency because it enables a user to extract only the data that are deemed necessary along a fixed path. We are developing a demonstration that depicts the extraction of data along an aircraft's flight path and uses the *GetCorridor* operation.

5.4.2 *GetPolygon* Operation

The *GetPolygon* operation will extract data from a multi-dimensional coverage for a polygon. Figure 5 depicts this extraction.

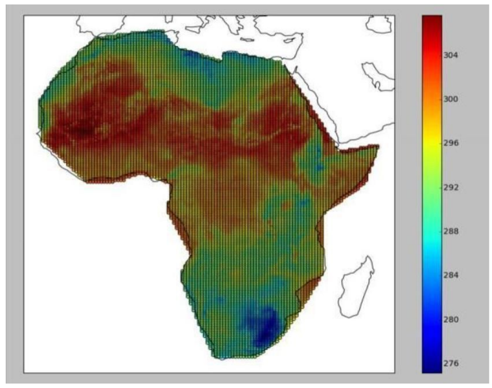

Fig 5. *GetPolygon:* A New 4D MetOcean Data Extraction Pattern

The GetPolygon operation can yield efficiency because it enables a user to extract only the data that are deemed necessary for the specific shape. For example, a user within the meteorological community may use the *GetPolygon* function to extract a data simulation of a volcanic ash plume from a 4D coverage.

6. Current Status And Results

The MetOcean Profile's development has been carried out under the auspices of the MetOcean Domain Working Group within the OGC. The profile has evolved iteratively to ensure it remains practical. The UK Met Office has been working with the NWS, and IBL Software Engineering has provided key support to create a working prototype. The prototype not only supports the common use cases (e.g.NWP output, but also supports post processed data such as the NWS's National Digital Forecast Database (NDFD). Over the past year, the MetOcean Profile's prototype has been refined, and there is a working URL that access services that provide current GFS data via the MetOcean Profile.

The prototype has proved to be very useful and the UK Met Office now supports a semi-operational service that can serve 4D coverages. This prototype implements the concepts described in this abstract. These techniques enable further progress in the evolution from a so-called data push" to a so-called data "pull" model.In this manner, bulk data are transformed into something that is much more useful to specific client applications. IBL Software is hosting a demonstration service that showcases these techniques using World Area Forecast System (WAFS) forecasts. The demonstration service is open to registered users. The demonstration service is driven by two servers, each holding a hemisphere of data. The service combines data from the servers before rendering the results.

During the presentation at the conference, we will also provide a demonstration of requests and responses for each of the WCS2.1 functions using NDFD data and the MetOcean Profile.

7. Future Plans

As of this writing, the MetOcean Profile is moving through its comment period. Once the current version is ratified, a number of improvements will be added and proposed to the OGC. It will be extended to allow for multiple slicing and trimming actions within a single "GetCoverage" operation. This will allow the extraction of multiple points or cross sections from one GetCoverage operation. The extraction patterns for corridors and polygons will also be formally added to the profile.

Other improvements to the MetOcean Profile include support for specific Coordinate Reference Systems (CRS), based on WMO standards, support for ensemble members via an additional dimension or axis, and support for JavaScript Object Notation (JSON) encoding. This work is currently being developed at the University of Reading (see https://covjson.org/spec/).

8. Annex

The following sections consist of outside links referenced in the abstract.

8.1 Links to *DescribeCoverageCollection*

The following is an example of a live *DescribeCoverageCollection* Request:

https://ws-sandbox.iblsoft.com/ndfdpp?SERVICE=WCS&VERSION=2.1.0&REQUEST=DescribeCoverageCollection&COVERAGEID=Menu-Time-Series

Next is an example link to a *DescribeCoverageCollection* Response. It depicts a CoverageCollection for the NDFD Menu Time Series product (CoverageCollection), with the coverages that comprise the collection (i.e. NDFD-Menu-Time-Series_SensibleWeather and NDFD-Menu-Time-Series_Wind_Speed_Gust):

http://www.mdl.nws.noaa.gov/~WGDS/examples/MetOcean/describeCoverageCollection/Bi-Lat_MetOcean_Examples/Responses/covcoll_responseDescribeCoverageCollection-ndfd_Menu-Time-Series.xml

8.2 Link to GFS Analyses of 2D vs 4D coverages

Below is a link showing the analysis of a GFS model for one run time and the difference in number of coverages with 2D versus 4D coverages:

http://www.mdl.nws.noaa.gov/~WGDS/examples/MetOcean/GFS-Coverage_2D_Vs_4D_stats.txt

9. Acknowledgements

The authors wish to thank Joseph Lang (NOAA/NWS/ OSTI/MDL), Daniel Gilmore (NOAA/NWS/ OSTI/MDL), Jozef Matula (IBLSoft) for their software development assistance.

10. References

Peter Trevelyan, Paul Hershberg and Steve Olson 2016 : MetOcean Application Profile Generated by WCS2.1 Services OGC Florida