TEMPORAL ANALYSIS OF ATMOSPHERIC DATA USING OPEN STANDARDS

Piero Campalani[a,*], Alan Beccati[a], Simone Mantovani[b], Peter Baumann[a]

[a] Center for Advanced Systems Engineering (CASE), Jacobs University Bremen, 28759 Bremen, Germany –
{p.campalani,a.beccati,p.baumann}@jacobs-university.de
[b] MEEO S.r.l, 44122 Ferrara, Italy – mantovani@meeo.it

KEY WORDS: Geospatial Web Services, Big Data, OGC, WCS, WCPS

ABSTRACT:

The continuous growth of remotely sensed data raises the need for efficient ways of accessing data archives. The classical model of accessing remote sensing (satellite) archives via distribution of large files is increasingly making way for a more dynamic and interactive data service. A challenge, though, is interoperability of such services, in particular when multi-dimensional data and advanced processing are involved. Individually crafted service interfaces typically do not allow substitution and combination of services. Open standards can provide a way forward if they are powerful enough to address both data and processing model.

The OGC Web Coverage Service (WCS) is a modular service suite which provides high-level interface definitions for data access, subsetting, filtering, and processing of spatio-temporal raster data. WCS based service interfaces to data archives deliver data in their original semantics useful for further client-side processing, as opposed to the Web Map Service (WMS) (de la Beaujardière, 2006) which performs a pre-rendering into images only useful for display to humans.

In this paper we present a case study where the OGC coverage data and service model defines the client/server interface for a climate data service. In particular, we show how flexible temporal analysis can be performed efficiently on massive spatio-temporal coverage objects. This service, which is operational on a several Terabyte data holding, has been established as part of the *EarthServer* initiative focusing on Big Data in the Earth and Planetary sciences.

1 INTRODUCTION

Sensor, image, simulation, and statistics data comprise, to a large extent, what makes up Big Data in the Earth sciences today. The sheer volume of data stored, as well as the velocity with which further data keep pouring in, calls for more efficient ways of accessing data archives than the classical model of accessing remote sensing archives via distribution of large files. A first step beyond such an *ftp* based download of files, whose granularity is fixed and imposed by the service, is user-driven spatial subsetting where the resulting files are generated on the fly. Additionally, sometimes reprojection and recoding into some client-chosen delivery format is offered.

Clearly, standardization of such Web interfaces is instrumental for achieving interoperability. One and the same client can be directed to different archives, and the functionality faced can be anticipated uniformly. While for humans this can be considered a convenience, for machine-to-machine communication it is essential to have a clear, exact data and service semantics.

The Open Geospatial Consortium (OGC) offers interface standards for various data categories and purposes. At the heart is the *feature* data type which resembles a geospatial object of any kind. Vectorial data are one particular sub-category, coverages another. A coverage loosely can be characterized as the digital representation of some space-time varying phenomenon; most commonly, they are known as raster data – such as 1-D sensor timeseries, 2-D satellite imagery, 3-D x/y/t image timeseries, or 4-D xy/z/t climate datasets –, but encompass more general structures as well, like point clouds and general meshes. As such, coverages qualify as Big Data when looking merely at their typical volumes as well as the variety encountered.

An abstract definition of coverages is given by OGC Abstract Topic 6 (OGC, 2006) and ISO 19123 (ISO, 2005), a concrete, interoperable, and compliance testable specification based on ISO

19123 is the GML Application Schema - Coverages (Baumann, 2010a), nicknamed GMLCOV.

From both vector-type features and raster-type coverages, a WMS service can render a 2-D image for display. Wile this is good for human consumption, often results need to be fed into further processing, or specific visualization tools. To this end, features (and, therefore, coverages) can be served out through a Web Feature Service (WFS) (Vretanos, 2010). However, a WFS will always only deliver the complete feature or coverage; while sometimes this is sufficient, normally subsetting is indispensable on the high-volume coverages. Therefore, a specific WCS interface (Baumann, 2012) is provided by OGC. Actually, WCS is a whole suite consisting of a Core which every implementation must support, plus a set of extensions adding further functionality facets. Hence, WCS offers a functionality spectrum ranging from simple spatio-temporal subsetting and encoding of the result in any suitable format up to complex ad-hoc analytics through a coverage query language, the Web Coverage Processing Service (WCPS) (Baumann, 2009).

In the remainder of this article we describe relevant related standards and technologies, with a section dedicated to the *Earth-Server* project; Section 4 presents a comprehensive walkthrough of WCPS queries that can enable different kinds of online analysis, for either existing features within the project and as well for a range of future value-adding capabilities that EO services can implement. Conclusions are exposed in Section 5. For a preliminary introduction to the WCS processing extension, namely WCPS, the reader is redirected to (Baumann, 2010b) where the standard has been formally presented, along with some first use case scenarios inside and outside the EO domain.

2 STATE OF THE ART

There is a legion of raster data formats, all with their particular sets of metadata, degree of informational completeness, the

---

*Corresponding author.

number of dimensions and pixel types they can support, etc. In terms of operations, often libraries are available with APIs for complete or partial access. However, each format is crafted individually and independently, and with its individual fingerprint it is not possible to combine them into interoperable services without heavy programming. Hence, a suitable abstraction is needed which is powerful enough to support all raster types appearing, with a unified metadata concept, encapsulated in an Abstract Data Type style Web interface. From the variety of approaches existing, in the context of this paper only adopted standards are of relevance, due to the focus on interoperability. Standardization of geo service interfaces is mainly driven by OGC which has an outreach into ISO TC211 and liaisons into other standardization bodies like W3C and OASIS-Open.

OGC Abstract Topic 6 (AT 6) — which is identical to ISO 19123 — defines an abstract model of coverages, which is concretized by the Geography Markup Language (GML) 3.2, "an XML grammar written in XML Schema for the description of application schemas as well as the transport and storage of geographic information" (Baumann, 2010a).

GMLCOV is a concrete application profile, which provides the unified coverage representation. There a function-like mapping in the domain space is held by one of the three fundamental constituting elements of a GMLCOV, the `domainSet`.

This work considers mainly data having 2D (geographical coordinates: x, y), 3D (x, y and temporal or height coordinates: x, y, t or x, y, h) and 4D (x, y, p, t, where pressure level p proxies height). A complete domain definition needs a Coordinate Reference System (CRS) for each of its dimensions to be consistently and effectively understood. GMLCOV utilizes an OGC compliant, URI based labelling of CRS, paired with a semantic resolver to ensure interoperability. A resolver also allows user defined CRS to be readily understood by mapping them to their GML definition (Baumann et al., 2012).

The actual data payload is contained in the `rangeSet` part. To allow values in the `rangeSet` to be properly decoded, structural information about them is required, along with metadata about their semantic meaning; both of these are defined in the third key element of any GMLCOV, the `rangeType` element. Figure 1 shows these aforementioned coverage constituting elements and their logical relation.

The GMLCOV coverage model ensures a semantically rich set of metadata by means of a direct inclusion the Sensor Web Enablement (SWE) data model (Robin, 2011) as means to describe the feature space, i.e. the `rangeType`. It lets the service specify many aspects of a sensor's measurement like units of measure, allowed values, quality flags, reference frames, and others. In the case of spaceborne imagers and radiometers, this can reach the resolution of the single spectral band.

The concrete definition of a coverage, together with a format encoding specification, allow instances to be created and transmitted among different software elements, thus ensuring interoperability: a key aspect to be achieved in modern archives to foster data dissemination and exploitation. Coverages can be understood by machines, processed automatically in complex workflows and easily shared.

GMLCOV is the core data model over which web services are built to allow interoperable access to data, such as the WCS service and the Processing extension to WCS. The former provides basic access to the underlying coverage data, such as subsetting over the domain, and an extension mechanism by which further
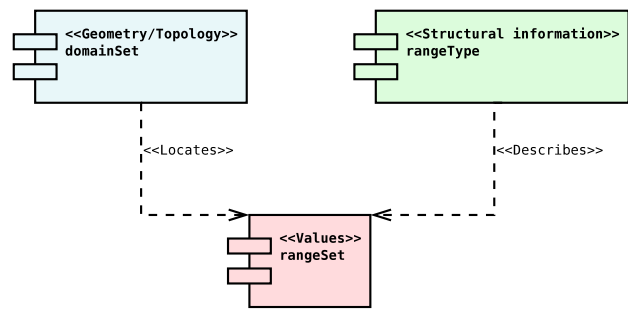

Figure 1: Principal elements of a GMLCOV instance and their relationship.

functions can be added (Baumann, 2012), the latter, providing advanced server side processing capabilities.

Naturally extending the GMLCOV coverage model, the WCS Earth Observation (EO) application profile — now open OGC standard (Baumann et al., 2014) — further enriches the descriptive metadata that can be associated with single EO granules, like phenomenon time, result time, dataset status. It also defines new, more complex, coverage types with homogeneous and heterogeneous grouping options over time (Meissl et al., 2013).

Support for time-series of coverages, a crucial aspect for EO archive services, is also promoted by the OGC Temporal Domain Working Group (OGC, 2013). It enables a single coverage to represent a full multi-dimensional time-series by intrinsically adding one or more temporal dimensions to the CRS of a coverage, building a spatio-temporal compound CRS. This way, coordinate tuples in the `domainSet` can include both geospatial and temporal numbers, whose meaning is publicly encoded via the actionable URIs which identify each single CRS. See (Campalani et al., 2013) for further insights on the matter.

3 THE EARTHSERVER INITIATIVE

The EU funded project *EarthServer* (`http://earthserver.eu/`), coordinated by Jacobs University Bremen (JUB), aims at the development of data access and data processing access services in several geoscience and planetary application domains. These services are aimed at delivering flexible yet user friendly interfaces to data archives, enhanced with processing capabilities. On the server side, the *rasdaman* array database technology (Raster Data Manager, 2009) is used to ensure efficient and scalable archive storage while including the processing features offered by its WCPS engine component, also known as *PetaScope* (Aiordăchioaie and Baumann, 2010). On the client side, a wide array of open-source libraries and technologies is leveraged for making data analysis available on a wide range of platforms, such as mobile devices, standard web browsers and up to immersive virtual reality devices. The project targets data volumes in the range of hundreds of Terabytes, thus tackling several of the challenges of Big Data. This work collects and showcases current results of the ongoing project from the perspective of data archive enhancement through WCPS and interoperability. Application domains considered in the project are Cryosphere, Atmosphere (Climate), Marine (Ocean), Geology (Solid Earth) and Planetary (Mars).

In the Atmosphere domain, the Multi-sensor Evolution Analysis (MEA) platform developed at MEEO Srl (`http://www.meeo.it/`) provides easy-to-use gateway portals to multi-dimensional 2D up to 5D datasets retrieved from spaceborne observations (e.g. MODIS), numerical modelling (e.g. MACC) or ground observations, totaling terabytes of data. It offers Data Curation and
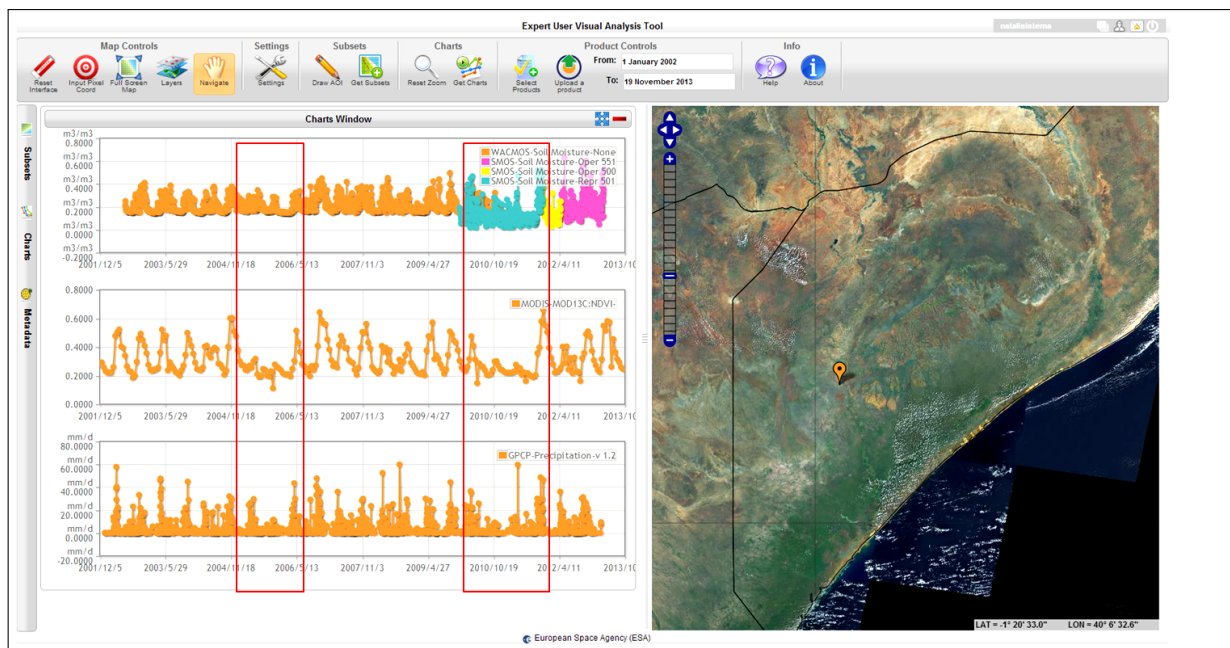
Figure 2: Identification of drought/famine events in Eastern Africa (Somalia) and evolution trends using multi-temporal data analysis of Soil Moisture (top), NDVI (middle) and precipitation (bottom) time series.

temporal analysis of atmospheric profiles, aerosol content and optical depth, cloud properties, among other Earth Observation (EO) products (Mantovani et al., 2013, VELISAR 2.0, 2006).

The data itself was provided by European Space Agency (ESA), National Aeronautics and Space Administration (NASA) and entities such as the European Centre for Medium-range Weather Forecast (ECMWF), the Austrian Meteorological Service (ZAMG), the Italian National Agency for new technologies, energies and sustainable development (ENEA), and the Sweden's Meteorological and Hydrological Institute (SMHI). Software and data requirements have been collected from User Interest Groups (UIGs) belonging Earth and atmosphere communities, and include temporal evolution on a specific point with the creation of time series of a single field, comparison of different fields on the same point and visualization of specific fields superimposed with high resolution background maps (Natali et al., 2011).

Figure 2 presents the current potential of the platform which, in addition to the mere visualization of the selected datasets, provides useful plots that show the temporal evolution back in time over selected points of interest, with the chance to evaluate the same evolution for a set of related products. Specifically, the figure shows how the visual temporal cross-comparison of data from different EO-derived products like soil moisture, vegetation index and precipitation could help identifying known famine events in Eastern Africa of the recent years. In the next section, we will propose a cookbook of WCPS queries that can suggest new advanced features to web services like MEA.

4 QUERYING THE ATMOSPHERE

In today's data-intensive science, users and community experts are overwhelmed with data sets from many different sources (Hey et al., 2009). They need tools that can help them moving the processing to the data, and hence ease their research studies to at least provide preliminary analysis capabilities on the bulk data archives, being able to select the subset of fields they need to download for the more complex processes.

```
for   variableName in ( coverageList )
 *(, variableName in ( coverageList ) )
  [ where booleanScalarExpr ]
  return encode ( coverageExpr, formatName )
```

Listing 1: Template syntax for a WCPS encoded `processCoveragesExpr`.

Querying the atmosphere can involve a multiplicity of different products: raw atmospheric dataset comes as multispectral images whose spatial resolution and temporal frequencies can widely differ between sensors; derived spaceborne products might represent aggregated fields with missing data and averaged over a time interval; statistical model simulations and forecasts usually provide the prediction error as at full spatial extent. In this section it will be shown how WCPS can handle and support typical access and processing of atmosphere products.

Based on the GMLCOV coverage model, WCPS offers a language to retrieve information from one or more coverages stored on a server. As formalized in Listing 1, the basic request structure consists of:

  - a loop over a list of coverages offered by the server;

  - an optional filter predicate (`where` clause);

  - an expression indicating the desired processing operation.

The `return` clause executes the specified operations for each coverage in the input coverage list, given that the filter condition is met. One or more variables are used to reference the groups of coverages. The WCPS standard describes a long list of basic processing operation (*expressions*) available for use. They mainly divide into three big types: probing functions, scalar and coverage expressions. Probing functions extract descriptive metadata of a coverage, like its bounding box, its native coordinate system or its number of dimensions. Scalar expressions are formally operations which *condense* the values provided by a set of

```
#MIN
min(c[t("#DATE")])
#MIN with Nil exclusion
min((c[t("#DATE")] = #NIL) * (-(#NIL)) +
    (c[t("#DATE")]!= #NIL) * c[t("#DATE")])
```

Listing 2: Minimum of an image in a time-series.

coverage points to a single scalar value, representative of a specified data aggregation, like the maximum value. Finally, coverage expressions take coverage objects as input and operate on a point-to-point basis: the returning value is a coverage of the same kind. For a thorough introduction on WCPS see (Baumann, 2010b).

Now focusing again on WCPS capabilities for EO visualization services, in the first place WCPS can be used to provide basic sample univariate and bivariate statistics from a coverage with missing data. Online exploratory analysis for variable selection in (geostatistical) regression models is then faced in Section 4.2, followed by examples of how to achieve fruitful time-series analysis on both single pixel or area of interest. Section 4.4 describes how to couple products from different sensors onto a single WCPS query, so to compare their fields over time. Finally, Section 4.5 presents some use case for practical scenarios of air quality risk assessment.

Time-series of 2D coverages will be assumed for simplicity. To help query readability in the listings, previously shown coverage expressions can be referenced through the red header which entitles each expression. Additionally, symbol n stands for a null value and the user's spatial region of interest is referred to as r. This latter defines a spatial selection over some spatial axes, like Latitude and Longitude, and is formally equivalent to as 'Lat(lo:hi), Long(lo:hi)', in case of a geographic projection. Simply including pressure or height parameters in the r can turn the same WCPS to 4D. Irrelevant parts of a WCPS query like the output format or the coverage(s) name will be left out, while only showing the core of the query, that is the `coverageExpr` (see Listing 1). Every further listing presented here will show a WCPS `coverageExpr`, and this detail will be left out from the captions. Additionally, coverage names are out of scope here, while related variable names will be `c` for single-product queries, `c1` and `c2` in case of multiple products.

4.1 Sample Statistics

The WCPS query language currently offers basic bricks of so-called condensers, which apply some operation on a set of grid points to retrieve a (scalar) score. Such operations enable extremes detection (`min`,`max`), average (`avg`) and sum (`add`). A general condenser is also available to build up customized operations.

Basic univariate statistics, like the minimum of a coverage, need special care in case missing values are present, like shown in Listing 2. By exploiting the boolean logic of the query language, one can *shift* Nil values outside the scope of the condensing operation. In the listing, the notation 'c[t("1950-01-04")]' means extracting the 2D slice image over some date in the 3D time-series, based on the syntax defined in ISO 8601 (ISO, 2004).

Collapsing Nil values is also shown in Listing 3. In the first expression, the built-in `avg` operator is replaced by a manual computation of what an average is, that is the sum of the values, divided by their cardinality ($\frac{1}{N}\sum_{i}^{N} c_i$); secondly, the corrected sample standard deviation $s_c = \left(\frac{1}{N-1}\sum_{i}^{N}(c_i - \bar{c})\right)^{0.5}$ is computed. In real queries the inclusion of WCPS macros, like the

```
#AVG
add((c[#ROI, t("#DATE")]  = #NIL) * (0)
  + (c[#ROI, t("#DATE")] != #NIL) *
    (c[#ROI, t("#DATE")]))
  / count(c[#ROI, t("#DATE")] != #NIL)
#STDEV
sqrt(add(
  pow(
    (c[#ROI, t("#DATE")] = #NIL) * (0) +
    (c[#ROI, t("#DATE")]!= #NIL) * (
     c[#ROI, t("#DATE")] - #AVG(c)), 2))
  / (count(c[#ROI, t("#DATE")] != #NIL) - 1))
```

Listing 3: Univariate statistics: sample mean and standard deviation.

```
#COVARIANCE
add(( #NIL(c1) or   #NIL(c2)) * (0) +
    (!#NIL(c1) and !#NIL(c2)) * (
    (c1[#ROI, t("#DATE")] - (#AVG(c1))) *
    (c2[#ROI, t("#DATE")] - (#AVG(c2))))
/ count(
    (c1[#ROI, t("#DATE")] != #NIL and
     c2[#ROI, t("#DATE")] != #NIL)) - 1)
#PCC
add(( #NIL(c1) or   #NIL(c2)) * (0) +
    (!#NIL(c1) and !#NIL(c2)) * (
    (c1[#ROI, t("#DATE")] - (#AVG(c1)))
    (c2[#ROI, t("#DATE")] - (#AVG(c2)))))
/ (sqrt(add(
    ( #NIL(c1) or   #NIL(c2)) * (0) +
    (!#NIL(c1) and !#NIL(c2)) * (
 pow(c1[#ROI, t("#DATE")] - (#AVG(c1)), 2 ))))
 * sqrt( add(
    ( #NIL(c1) or   #NIL(c2)) * (0) +
    (!#NIL(c1) and !#NIL(c2)) * (
 pow(c2[#ROI, t("#DATE")] - (#AVG(c2)), 2 )))))
```

Listing 4: Bivariate statistics: sample covariance and Pearsons'r linear Correlation Coefficient (see Eq. 1).

average of a coverage while computing s_c can be either obtained by separate AJAX queries or by literal replacement of the expression.

While the computation of further coefficients of univariate statistics is clearly possible via WCPS, we now proceed to show some more interesting bivariate statistic. Listing 4 shows how to compute the covariance function $q_{cd} = \frac{1}{N-1}\sum_{i}^{N}(c_i - \bar{c})(d_i - \bar{d})$ and the coefficient of linear correlation between two coverages c1 and c2, which might have independently distributed missing data, see Equation 1.

$$r_{cd} = \sum_{i}^{N}(c_i - \bar{c})(d_i - \bar{d}) / \sqrt{\sum_{i}^{N}(c_i - \bar{c})^2 \sum_{i}^{N}(d_i - \bar{d})^2} \tag{1}$$

The general mechanism to handle missing data is hence to to split the operation in the sum of two terms whose data is masked by complementary boolean expression. Such mechanism will not be shown own in future sections to achieve a better query readability.

4.2 Exploratory Data Analysis

The ever increasing amount of data that is daily provided by geostationary and polar-orbiting satellites, makes it hard to atmospheric modellers to use whole archives of datasets without data curation services.

```
#HISTOGRAM
coverage histogram
over     $i x(0:99)
values   count(
  c[t("#DATE")] >  (1+#MIN) +
    $i  * (($MAX-$MIN) / 100) and
  c[t("#DATE")] <= (1+#MIN) +
 (1+$i) * (($MAX-$MIN) / 100))
#HISTOGRAM_LOG
coverage histogram
over     $i x(0:99)
values   count(
  ln(1 + #NIL + c[t("#DATE")])
  >  ln(1+#NIL+#MIN) +
      $i * (ln(
      (1+#NIL+#MAX)/(1+#NIL+#MIN)) / 100) and
  ln(1 + #NIL + c[t("#DATE")])
  <= ln(1+#NIL+#MIN) +
      (1+$i) * (ln(
      (1+#NIL+#MAX)/(1+#NIL+#MIN)) / 100) and
  c[t("#DATE")] != #NIL))
```

Listing 5: 100-bin histogram retrieval of a coverage and its logarithmic transform.

```
#LM_RESIDUALS
c1[#ROI, t("#DATE")] - c2[#ROI, t("#DATE")] * (
  add(c1[#ROI, t("#DATE")] *
    c2[#ROI, t("#DATE")] ) /
  add(pow(c2[#ROI, t("#DATE")], 2)))
```

Listing 6: Map of the residual for the model $c1 \sim c2$.

Although advanced geostatistical analysis and modelling is surely better executed locally with ad-hoc tools like R (R Development Core Team, 2011), it is hereby shown how to provide some first exploratory data analysis directly on-the-fly, to let users *evaluate* models before downloading data. This can be a crucial factor, especially in the phase of covariables selection (Gupta and Christopher, 2009).

A first visual comparison between fields is usually achieved with scatterplots, which can be easily plotted on a browser by fetching abscissa and ordinate series with separate WCPS queries. A more complex use case, frequently used in any sort of preliminary data analysis, is the visualization of the histogram, the discretized density function.

Listing 5 show how to retrieve the necessary data to build up and visualize online the histogram of a selected field. In case of highly skewed distributions, like is the usual case for e.g. aerosols and particulate matter in general, it can be useful to check the histogram of the log-transformed variable, like shown in the second example again in Listing ref lst:histograms.

Often in regression analysis there is need to inspect a 2D residual map derived from the application of a linear model, to check for homogeneity in the spatial distribution of the residuals over the ROI. Listing 6 shows how to reproduce such map on a linear model without intercept $y \sim x$, where the normalized coefficient $\beta_0 = \sum xy / \sum x^2$. Alternatively the residual plot could be retrieved to evaluate the degree of heteroskedasticity.

4.3 Time-series Analysis

Data analysis *across* time is certainly a major application in the Atmosphere domain (Zha et al., 2010), as in many other. Evaluating information over an historical archive of data serves to different actors, from environmental policy makers for decision making, to environmental modellers to feed their models, up to the simple citizen to evaluate the air quality of his geographic area across the years.

```
#SINGLE_PXH
coverage single_pixel_history
over     $t t ( imagecrsdomain(c, t) )
values   c[#ROI, t:"CRS:1"($t)
#MERGED_PXH
coverage averaged_pixel_history
over     $t t ( imagecrsdomain(c, t) )
values   avg(c[#ROI, t:"CRS:1"($t)])
#AGGREGATED_PXH
coverage aggregated_pixel_history
over $t t ( 0,
    (count(c[#SLICE(Lat,Long)])/\#AGG_DAYS)-1 )
values   avg(c[#SLICE(Lat,Long),
        t:"CRS:1"($t:$t+(\#AGG_DAYS-1))])
```

Listing 7: Pixel histories, with spatial and temporal aggregation.

In this section we present the template queries to extrapolate so called *pixel histories*, that is 1D series of a selected variable of interest, within a specified time interval. As shown in Listing 7, there are different types of pixel histories which could be retrieved, depending on the application scenario, and they are hereby explained:

Single-pixel history 1D series of values over a specified 0D geographic point (e.g. which AOT values over $30\,°\mathrm{E}-45\,°\mathrm{N}$?);

Merged-pixel history 1D series of values representing a specified condensing operations over a region of interest (e.g. which average $PM_{2.5}$ values over the urban area of Rome?);

Aggregate-pixel history 1D series of values which have been aggregated over a specified time interval (e.g. which weekly average of daily $PM_{2.5}$ over this location?).

The WCPS function `imagecrsdomain` gives the number of cells of a coverage along a specified dimension: to address then this internal indexed CRS, the notation `CRS:1` has been implemented. Variable `#AGG_DAYS` represents the number of days of time aggregation, e.g. 7 for weekly averages. Finally, `#SLICE(Lat,Long)` determines a slicing operation over the geographic axes, i.e. a point.

4.4 Products Cross-Comparison

An eagerly awaited functionality of atmospheric data services is the possibilty to compare and combine different satellite products and model data. This can be required for data validation, quality enhancements or data fusion for the reduction of the holes in the satellite data (Leptoukh et al., 2007).

As an example of the potential of WCPS processing capabilities, Listing 8 shows some use case query that can be exploited to achieve a visual intercomparison of two different satellite products' time-series: `c1` having daily frequency, and `c2` having a higher spatial resolution and M times the temporal frequency of `c1`.

`#Δ_MERGED_PXH` retrieves a 1D merged pixel history of the differences (residuals) between the two coverages: this is independent of the spatial and temporal resolution thanks to the `avg` condenser which can span both spatial and temporal coordinates in the compound CRS of the coverage. `#AGGREGATED_Δ_MAP` shows a 2D map of the "distance" between the products, aggregating in time the values of `c2` to the more to the coarser temporal resolution of `c1`. This is achieved thanks to the scaling operator `scale`. Finally, `#1:1_Δ_MAP` on the other hand, make a one-to-one comparison of products, by upscaling the coarser spatial resolution of `c1`. Time subsettings are explicitly shown this time to better appreciate the meaning of each query.

```
#Δ_MERGED_PXH
coverage delta_merged_pixel_history
over     $t t ( imagecrsdomain(c1, t) )
values avg(c1[#ROI, t:"CRS:1"($t)]) -
       avg(c2[#ROI, t:"CRS:1"($t:$t+(#M-1))])
#AGGREGATED_Δ_MAP
scale(c1[#ROI, t("2014-01-01")],
  imagecrsdomain(c2[#ROI, t("2014-01-01")]) -
  coverage delta_pixel_history
  over $x x ( imagecrsdomain(
       c2[#ROI, t("2014-01-01")], Long) ),
       $y y ( imagecrsdomain(
       c2[#ROI, t("2014-01-01")], Lat) )
  values avg( c2[Lat($y), Long($x),
   t("2014-01-01T00":"2014-01-01T23:59")])
#1:1_Δ_MAP
scale(c1[#ROI, t("2014-01-01")],
  imagecrsdomain(
    c1[#ROI, t("2014-01-01")]) -
    c2[#ROI, t("2014-01-01T12:00")] )
```

Listing 8: Expressions for the intercomparison of two satellite products with independent spatial and temporal resolutions.

```
#TRAFFIC_LIGHT
{ red: (char)( c >= LO_THRESHOLD ) * 255;
green: (char)( c <  HI_THRESHOLD ) * 255;
 blue: (char)( c * 0 )
{ [#ROI, t("#DATE")]
```

Listing 9: Traffic light maps for the visualization of the exceedance of a lower and a higher concentration threshold.

```
#95%_EXCEEDANCE
((c[#ROI, t("#DATE")]).prediction +
 (c[#ROI, t("#DATE")]).variance * 1.644854)
  > #HI_THRESHOLD
```

Listing 10: Threshold exceedance map from a model simulation, with 95% of confidence interval.

4.5 Hazard Assessment

Whereas atmospheric scientists and modellers require exact data retrieval to apply statistics and simulations, a more user (and policy makers) oriented approach is to yield visual imagery of a certain product, e.g. related to the air quality, signalling hazard situations, like the case of mass concentrations of pollutants that go far beyond the recommended threshold.

A green-yellow-red "traffic light" map can be fetched, as shown in Listing 9: being able to define the values of the output map separately on the different bands, it is possible to use logical expression to set the red, green and blue bands respectively.

Like shown in Listing 10, a different situation occurs when running a-posteriori analysis of model forecasts (Emili et al., 2011): a user could upload his geostatistical maps to the WCPS server, and gain information by actively using the estimation error associated with every predicted pixel to retrieve exceedance maps of the selected pollutant over a specified threshold, and with a custom confidence interval.

5 CONCLUSIONS

In this article we have presented a comprehensive set of WCPS queries that can be adopted by EO service providers to enable high-level processing frontend functionalities to their spatio-temporal datasets, with a focus on the typical datasets that are provided in the atmospheric imagery.

Several demonstrations were shown, from sample univariate and bivariate statistic coefficients to exploratory variables analysis, from time-series pixel and ROI histories to the intercomparison of different satellite products. Risk assessment capabilities for typical air quality use cases are finally shown.

The proposed approach supports the paradigm of moving the processing to the data, opening new possibilities to scientific modellers and EO analysts. The use of the OGC open standard and FOSS database technologies like *rasdaman* promotes interconnectedness and interoperability among Big Data services.

REFERENCES

Aiordăchioaie, A. and Baumann, P., 2010. PetaScope: An open-source implementation of the OGC WCS geo service standards suite. In: Scientific and Statistical Database Management, Springer, pp. 160–168.

Baumann, P., 2009. OGC Web Coverage Processing Service (WCPS) Language Interface Standard. OGC 08–068r2.

Baumann, P., 2010a. GML Application Schema for Coverages. OGC 09–146.

Baumann, P., 2010b. The OGC Web Coverage Processing Service (WCPS) standard. Geoinformatica 14(4), pp. 447–479.

Baumann, P., 2012. OGC Web Coverage Service (WCS) – Core. OGC 09–110r4.

Baumann, P., Campalani, P., Misev, D. and Yu, J., 2012. Finding my CRS: A systematic way of identifying CRSs. In: ACM SIGSPATIAL GIS, pp. 71–78.

Baumann, P., Meissl, S. and Yu, J. Y., 2014. OGC Web Coverage Service 2.0 Interface Standard — Earth Observation Application Profile. OGC 10–140r1.

Campalani, P., Misev, D., Beccati, A. and Baumann, P., 2013. Making time just another axis in geospatial services. In: 20th International Symposium on Temporal Representation and Reasoning (TIME).

de la Beaujardière, J., 2006. Web map service. OpenGIS Web Map Server Implementation Specification pp. 06–042.

Emili, E., Popp, C., Wunderle, S., Zebisch, M. and Petitta, M., 2011. Mapping particulate matter in alpine regions with satellite and ground-based measurements: An exploratory study for data assimilation. Atmospheric Environment.

Gupta, P. and Christopher, S., 2009. Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: Mutliple regression approach. Journal of geophysical research 114(D14), pp. D14205.

Hey, A. J., Tansley, S., Tolle, K. M. et al., 2009. The fourth paradigm: data-intensive scientific discovery.

ISO, 2004. ISO 8601:2004: Data elements and interchange formats – Information interchange – Representation of dates and times.

ISO, 2005. ISO 19123:2005: Geographic information – Schema for coverage geometry and functions.

Leptoukh, G., Zubko, V. and Gopalan, A., 2007. Spatial aspects of multi-sensor data fusion: Aerosol optical thickness. In: Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International, IEEE, pp. 3119–3122.

Mantovani, S., Natali, S., Barboni, D. and Grazia Veratelli, M., 2013. Climate data service in the FP7 EarthServer project. In: EGU General Assembly Conference Abstracts, Vol. 15, p. 12849.

Meissl, S., Aime, A. and Giannecchini, S., 2013. WCS and EO-WCS status in open source. In: FOSS4G.

Natali, S., Beccati, A., D'Elia, S., Veratelli, M., Campalani, P., Folegani, M. and Mantovani, S., 2011. Multitemporal data management and exploitation infrastructure. In: The 6[th]International Workshop on the Analysis of Multi-temporal Remote Sensing Images (MultiTemp).

OGC, 2006. The OpenGIS Abstract Specification — Topic 6: Schema for coverage geometry and functions. OGC 07–011.

OGC, 2013. Temporal DWG Web. `http://external.opengis.org/twiki_public/TemporalDWG/WebHome`. Accessed online on 2013-aug-22.

R Development Core Team, 2011. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Raster Data Manager, 2009. `www.rasdaman.org`.

Robin, A., 2011. OGC SWE Common Data Model Encoding Standard. OGC 08–094r1.

VELISAR 2.0, 2006. Public access to crustal deformation data. `http://mea.eo.esa.int/InSARItaly`.

Vretanos, P., 2010. OpenGIS Web Feature Service 2.0 Interface Standard. OGC 09–025r1.

Zha, Y., Gao, J., Jiang, J., Lu, H. and Huang, J., 2010. Monitoring of urban air pollution from MODIS aerosol data: effect of meteorological parameters. Tellus B 62(2), pp. 109–116.