Improving Efficiency of Grid Representation in GML

Piero Campalani, Alan Beccati, Peter Baumann

Abstract

Earth observation and remote sensing today are a primary resource for a miscellanea of geoservices providing massive timeseries of raster-based datasets. This leads to more complex grid structures than before – for example, in a 3-D x/y/t timeseries data cube, regular lat/long axes of some orthoimagery may be combined with irregular time axis. The OGC unified coverage model allows to describe such generalized grid structures through its concept of a Referenceable Grid Coverage. This standard, utilizing GML, is extended by GML 3.3 with new types of geometries that can handle irregular and warped gridded topologies. A shortcoming is, however, that they are efficient only in the special cases described initially. Hybrid situations like the one above can become inconveniently verbose. In this paper, we document first considerations to improve flexibility and modularity of the GML grid models as part of our work to fully unleash all grid types for OGC-type Web services. Indeed, in case of hybrid rectilinear and curvilinear grid models there is room for optimization in the existing GML types, presented by means of formal structure definitions and practically motivated examples.

1. Introduction

The importance of spaceborne and airborne imagery for the environmental sciences is constantly increasing, along with the growing capabilities of web services to offer scalable access to location-based information. Interoperability between the plethora of scientific realms involved is a determinant aspect of the game, and should appear in the list of design goals of any geoservice today.

The Open Geospatial Consortium (OGC) Geography Markup Language (GML) is an open standard for modeling geographical features, both vector and raster based. It was actually designed for the Internet and was meant to create a whole Web of geographic interoperable systems, going beyond the mere transport of data between separate services (Raper 1999). GML 3.3 (Portele 2012) integrated new schema definitions, comprising a set of types to address the definition of irregularly gridded datasets, therein called referenceable grids.

In August 2010 the fully refactored Web Coverage Service (WCS) 2.0 has been adopted by OGC: based on GML 3.2.1 (Portele 2007), the new service permits a highly interoperable mean for accessing and exchanging digital geolocated data in the spacetime cube, which are represented in GML as coverages (Baumann 2010b). When it comes to the description of a coverage geometry (or topology), the explicit location of each single point could be expressed in a straightforward way. In the era of big data this method can however rapidly bring to excessively verbose descriptions: more conciseness must be capitalized whenever possible, while still pertaining a clear understandable model.

Gridded coverages are widely adopted by the scientific community to model their raster-based datasets, and offer interesting use cases in relation to their geometric representation. As depicted in Fig. 1, starting from a perfectly regular and aligned grid, then several affine (scaling, rotations,
translational, shearing) or non-affine transformation can change its shape in the Coordinate Reference System (CRS), up to warp it to an irregularly spaced and/or curvilinear set of points.

A grid could then be categorized into different tiers of geometric regularity, from the simplest (and more concisely described) regular grids, to the irregular and still rectilinear ones, up to warped (and highly verbose) curvilinear topologies.

However, there are practical cases — time series of grids above all — where the topology results in a sort of hybrid pattern, and subparts of a multidimensional grid belong to different geometric categories: the currently available GML grid types are not flexible enough to accommodate an optimized description of such situations, and they instead compel the adoption of verbose descriptions where unneeded.

In this article we are going to describe our first efforts in the elaboration of a more flexible grid model for such hybrid situations, by applying little changes to the GML 3.3 specification. In Sec. 2 an overview of related works is given; Sec. 3 will formally describe the proposed changes to the current GML standard, and practical examples will be then exposed in Sec. 4; final conclusions are drawn in Sec. 5.

2. Related work

As previously mentioned, the whole discussion carried out in the present work has originated from the need to smoothly accommodate complex gridded geometries inside the OGC coverage model.

GMLCOV (Baumann 2010a) is the shorthand name for this coverage model and it distinctly generalizes coverages beyond the traditional 2-D imagery and beyond rasters to more complex n-dimensional data structures inside spatiotemporal CRSs (Campalani et al. 2012): the increasing functionalities of the coverage WCS and WCPS (Baumann 2009) services have indeed spotted the need for new more comprehensive model for GML grids.

The hereby reasoned optimizations are certainly stressed when dealing with very large gridded datasets, and a renewed flexible grid concept would spread its benefits towards several scientific communities which adopted the coverage model: remote sensing for the environment (Natali et al.
2011), planetary multispectral analysis (Oosthoek et al. 2012), geological sciences (Lake 2005), public sector geodata (Baumann 2011), to name a few. In parallel to the issue presented here, serving large multidimensional datasets calls for new flexibility in the representation of the CRS as well. Closely intermingled in the elaboration of a comprehensive coverage model is hence the seamless integration of the temporal dimension inside and outside the GMLCOV/WCS suite of standards (Campalani et al. 2013). OGC has recently established an open and interactive domain working group for this activity, whose proceedings are publicly available at http://external.opengeospatial.org/twiki_public/TemporalDWG/.

3. Change proposals
In these following sections we will discuss some idea for optimization of grid models defined within the GML 3.3 gmlrgrid namespace, which is assumed to be known by the reader. A rectilinear grid with irregular spacings (ReferenceableGridByVectors – RGBV) extends the behaviour of a regular grid by letting the spacing between points be non-uniform. As with regular grids — known as rectified in the GML ontology — a n-D RGBV is defined by an origin point and a set of n vectors which act as directional resolutions, but additionally each vector comes with a list of coefficients which describe the irregular spacing of every point along an axis.

There can be cases where one or more rectilinear axes which regular spacing, and others that do not. The explicit listing of the coefficients of the former axes is unnecessary in this case, since knowing the resolution vectors is clearly enough for the position of the points on the regular axes to be determined. Therefore the coefficients should be optional within the description of a grid axis, GeneralGridAxis, and Listings 1 formalize this: this way, the redundant series of coefficients associated to a regular axis, could be omitted from resulting in a significantly reduced description.

Extending this concept to mixed curvilinear/rectilinear grids, it is worth defining a new hybrid type which we might call ReferenceableGridByArrayAndVectors (RGBAV). This new type merges pre-existing GML notations in order to reflect the inherent mixture of such grids so that no redundant information is kept.

Generally, these structures could be seen as the repetition of a warped geometry along one or more rectilinear axes, at either regular or irregular intervals, which can then act as the multidimensional origin of the full grid. Sec. 4.2 will give more insight by means of a numerical example.

4. Practical examples
This section will provide some use cases to provide further details on the functioning behind the proposed gmlrgrid environment.

4.1. Mixed regular/irregular rectilinear grids
This kind of geometry could easily take place when either irregular time series of regular 2D/3D spatial datasets, or with irregularly spaced layouts over fixed intervals of time are considered. More specifically, this could be the case of an archive of satellite rectified images, a single (or a series of) satellite image(s) on different pressure levels, or maybe a row of soil samples with irregular distance in space, taken at regular intervals in time (see Fig. 2).

\[ c_n = \frac{n}{\alpha} \quad \text{for} \quad n \in [0, |\alpha| - 1], \] being $|\alpha|$ the cardinality of the set of points along the a regular axis.
Changes proposal: `gmlrgrid:GeneralGridAxisType` to let the coefficients be optional (upper-left side) and `gmlrgrid:ReferenceableGridByArrayAndVectors` new type definition (lower-left side, with an instance document on the right side).

The advantages of having optional axis coefficients are particularly evident when the irregularity of one single axis coerces the use of a referenceable grid representation: having a dataset of $10^4$ 2D $10^5 \times 10^5$ spaceborne rectified images with irregular overpass hours (e.g. polar-orbiting satellites), the chance to silently assume the regular spacing in the spatial axes would reduce the coefficients from $21e5$ to $10e4$ in the description of the 3D grid.

4.2. Mixed curvilinear/rectilinear grids

Temporal series of unrectified satellite products represent the most self-evident use case for this situation. For ease of representation, we can think of the irregular repetition along time of a fixed $3 \times 3$ warped topology. This geometric structure can be conveniently modeled as an RGBA, as shown in Listings 1, which represents an irregular time series starting from 21st December 2012 (ANSI day 150470) of a lat/long warped grid. Looking at this GML description there are several
things to observe and comment:

- the `@dimension` of the grid is 3, equal to the dimensionality of the spatio-temporal CRS (see `@srsName` and `@srsDimension`);

- the `axisLabels` are custom names associated to the axes of the grid and they are not to be confused with the labels of the CRS;

- the array of coordinates is a triplet of lat/long/time values and, being the xy pattern constant through time, only the location of the base layer requires to be listed;

- the values in the `posList` array are correctly parsed thanks to the information in the `grid-AxesSpanned` and `sequenceRule` elements, which explains that the tuples of coordinates linearly list the $3 \times 3$ points of the base grid section by first spanning the x axis;

- after defining the curvilinear pattern of the grid, an additional `generalGridAxis` is needed to define the remaining rectilinear geometry represented by the `axis t` along time: taking the map defined in the `posList` as a 2D origin plane, a vector $\{0,0,1\}$ can be used to obtain the location of points in the i-th level in time; e.g. the coordinate of the lower-left corner of the 2D spatial grid on the last layer in time could be simply evaluated as $\{2,8,150470\} + c_t[5] \cdot \{0,0,1\} = \{2,8,150479\}$, being $c_t$ the array of coefficients for the t axis $[0,1,3,5,9]$.

Whereas current standard GML would have force the explicit listing of all the grid points, by using the hybrid grid model we could reduce the overall verbosity and at the same time more clearly express the nature of the full grid as repetition of a warped pattern along a rectilinear axis.

5. Conclusions

This article documents the results of a preliminary optimization analysis of the currently available GML models for irregular grids, with the intention to extend their flexibility for use cases of complex multidimensional coverages.

There are common practical situations, temporal series of coverages in particular, that result in hybrid topologies in one and the same coverage, something not addressed by the currently available

---

2The `@srsName` presents a concatenation of the geographic EPSG:4326 and the temporal reference system, the ANSI dates in this case, i.e. days from the 1st of January 1601.
GML grid types as they do not permit assignment of regularity types to individual coordinate space axes. Rather, the type of grid is chosen for the whole coverage and all of its axes.

In order to achieve a clearer and less verbose grid description, we propose two changes and present them by means of XML definitions and simple examples. These consist in a more flexible concept of the rectilinear grid axis and a first idea for a new grid type (hereby called ReferenceableGrid-ByArrayOfVectors), to accommodate mixed topologies.

Future work will focus on the deeper elaboration of a coherent, modular conceptual model to encompass all complex grid models in a GML retro-compatible manner.

Acknowledgment

The research leading to the results presented here has received funding from the European Community’s Seventh Framework Programme (EU FP7) under grant agreement n. 283610 “European Scalable Earth Science Service Environment (EarthServer)”.

References