"C-Squares", a New Spatial Indexing System and its Applicability to the Description of Oceanographic Datasets

Tony Rees
CSIRO Marine Research • Hobart, Tasmania Australia

Abstract

A new method is described for representing, querying, displaying and exchanging dataset spatial extents at the metadata level. This method, entitled "c-squares" for *Concise Spatial Query And Representation System*, represents dataset spatial extents (footprints) as regular or irregular shapes built up from smaller component units (grid squares) coded according to a globally applicable system, and permits more reliable spatial queries than are presently enabled by the commonly used "bounding rectangles" method in metadata systems. The method is equally applicable to marine and terrestrial datasets but is particularly useful for marine data (which frequently have irregular, voyage-specific footprints), plus other dataset types not well represented by bounding rectangles.

Introduction

Scientists have, for many years, had access to abstracting services such as "Oceanographic Abstracts", "Biological Abstracts" and others, which provide a resource discovery and description facility for articles in the scientific literature. Over the last decade, similar techniques have begun to be applied to the scientific datasets held by many research agencies, in the form of metadata records (metadata = "data about data") which are then assembled into metadata catalogs or data directories. Thus, by searching an internet-enabled metadata catalog such as NASA's "Global Change Master Directory" (gcmd.gsfc.nasa.gov/), the European Space Agency's "INFEO" directory (www.infeo.org/), or the "Australian Spatial Data Directory" (www.auslig.gov.au/asdd/), an inquirer can discover datasets of potential interest, view the summary descriptive information (metadata) about any dataset and, in an increasing number of cases, be presented with an on-line access point for the data.

The ability to specify a geographic location as part of the inquirer's search criteria is one of the key aspects of any metadata catalog dealing with so-called "spatial" (or geospatial) data—that is, data which are related to position on the earth's surface, and could thus be represented on a map. From the time of early metadata standards, such as NASA's DIF or Directory Interchange Format (e.g. NASA, 1993) and the first draft US Federal Geographic Data Committee standard (FGDC, 1994), up to three possible mechanisms for spatial indexing of datasets have been supported. The first is *location keyword* or equivalent; however for much oceanographic data, the inquirer may wish to search for data in regions for which there is no finer scale locality name than, for example, "south-east Pacific Ocean". The second is the *bounding polygon* (supported by FGDC and some other national standards), which indicates a regular or irregular border deemed to enclose the dataset. Such polygons can potentially be uploaded to GIS (Geographic Information System) software where searching for spatial overlaps can be performed, but are generally not searchable by the simpler, numeric- or text-based query operations supported by most metadata catalogs.

The third is the *bounding rectangle*—alternatively known as *bounding box*, or *minimum bounding rectangle* (MBR)—four numeric values indicating the northernmost, southernmost, westernmost and easternmost limits of the data (e.g. in decimal degrees). This functions as a simplified surrogate for the bounding polygon, and is straightforward for metadata systems to search using simple numeric comparisons with an inquirer's designated "search rectangle". A bounding rectangle can be a good fit to dataset spatial extents where:

- data "footprints" are rectangular or nearly so, with no significantly concave portions of the perimeter;
- the boundaries of the data are reasonably closely aligned with parallels of latitude and longitude;
- data coverage within its designated perimeter is essentially complete without holes or significant data gaps; and
- the data form a single, contiguous block.

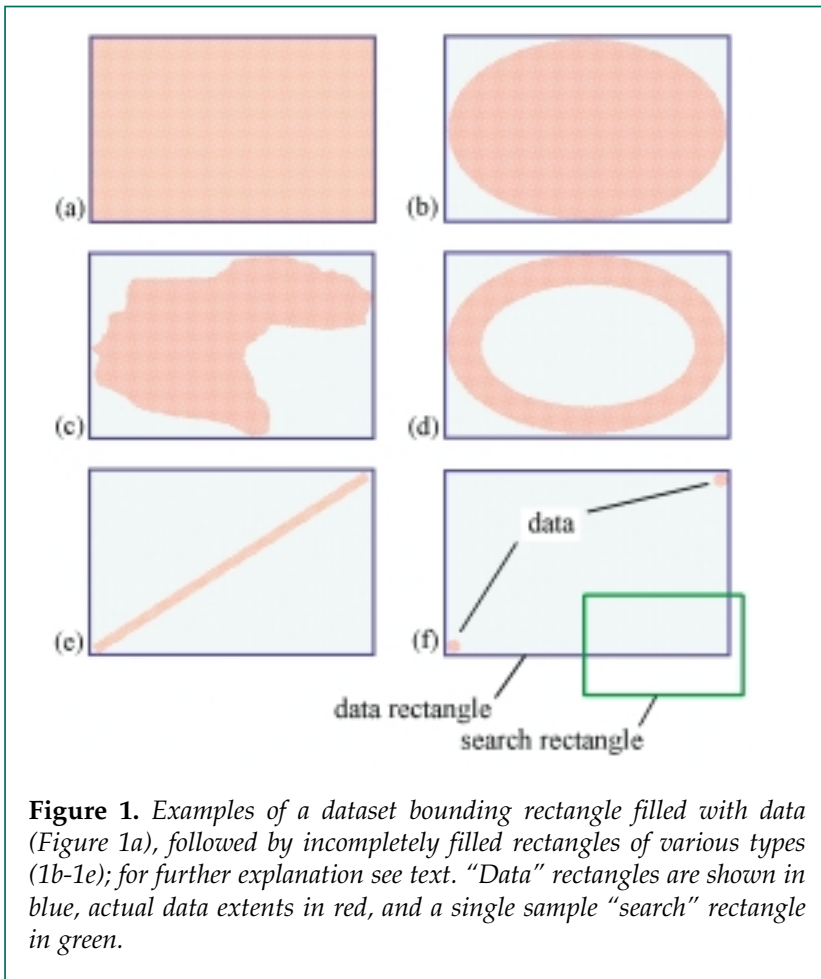Such an "ideal" data type is represented in Figure 1a.

Figure 1. *Examples of a dataset bounding rectangle filled with data (Figure 1a), followed by incompletely filled rectangles of various types (1b-1e); for further explanation see text. "Data" rectangles are shown in blue, actual data extents in red, and a single sample "search" rectangle in green.*

These conditions are, however, frequently not met by various types of dataset in practice—represented by the generalized cases shown in Figures 1b-1f. In Figure 1b, the dataset footprint fills the rectangle reasonably well, but unfilled regions of the bounding rectangle occur at the corners. Figure 1c is a more prevalent real-world case, with concave as well as convex portions of the dataset boundary, and a consequent increase in the unfilled percentage of the data rectangle. Figure 1d is typical of datasets with one or more unsampled regions or "holes". In Figure 1e, the borders of the dataset footprint are essentially linear, but are aligned obliquely with parallels of latitude and longitude; while Figure 1f illustrates a dataset comprising more than one disjunct sampling region. In any one of these situations, a sample "search" rectangle (illustrated in green) can potentially intersect an unfilled region of the data rectangle, and produce a false hit.

In practice, marine data usually display at least one, and commonly combinations of the above characteristics (e.g. Figures 2 a-d); in fact the majority of spatial datasets, both marine and terrestrial, are potentially susceptible to these problems. For example, natural features or administrative areas are rarely rectangular; marine data from around the perimeter of an island or continent will result in a dataset with a central "hole", which may occupy a significant portion of the bounding rectangle; while discontinuous sampled areas or dispersed, naturally occurring phenomena (such as "sea ice in the polar regions") can easily result in dataset footprints with multiple regions of interest and (potentially considerable) intervening gaps.
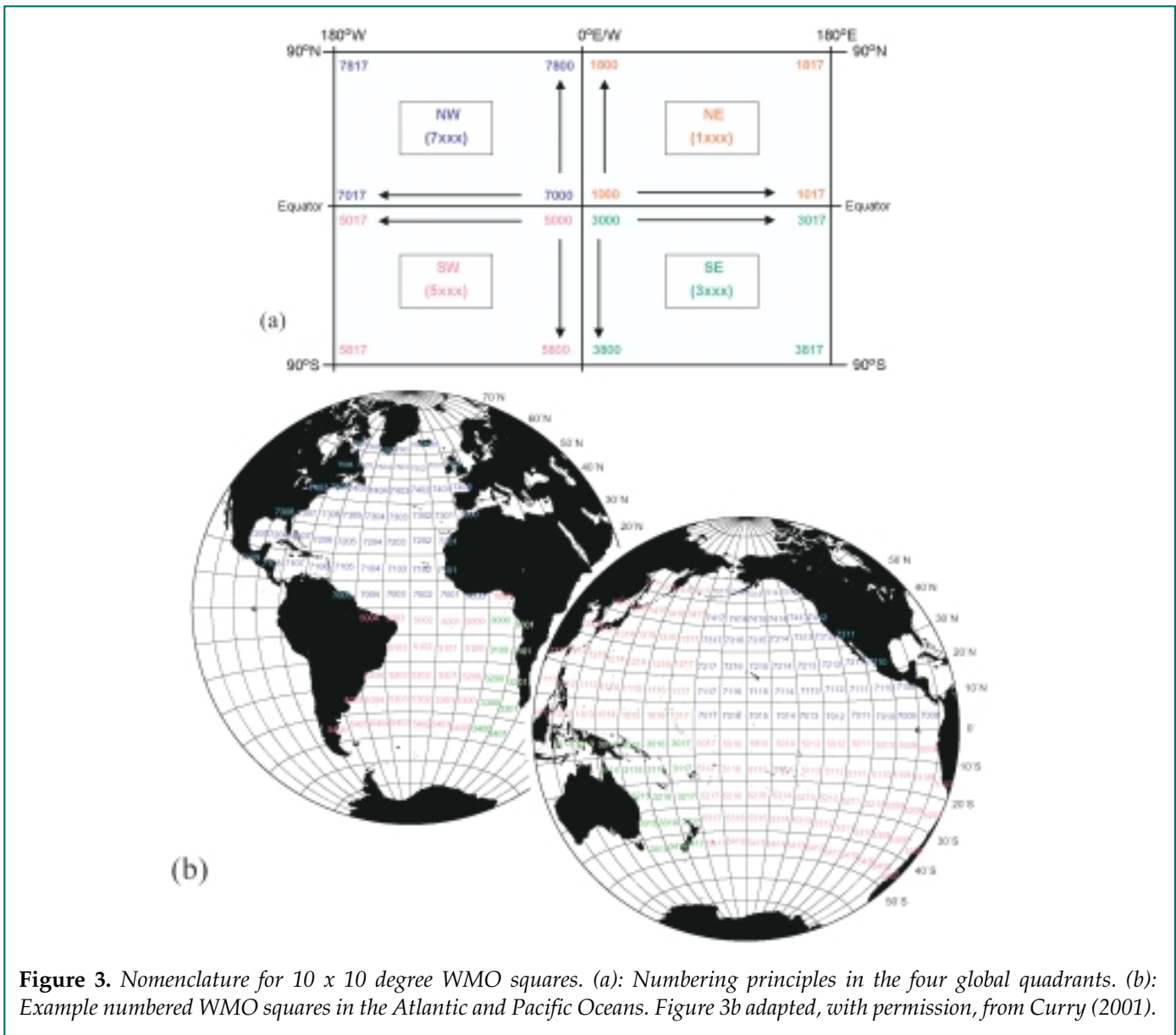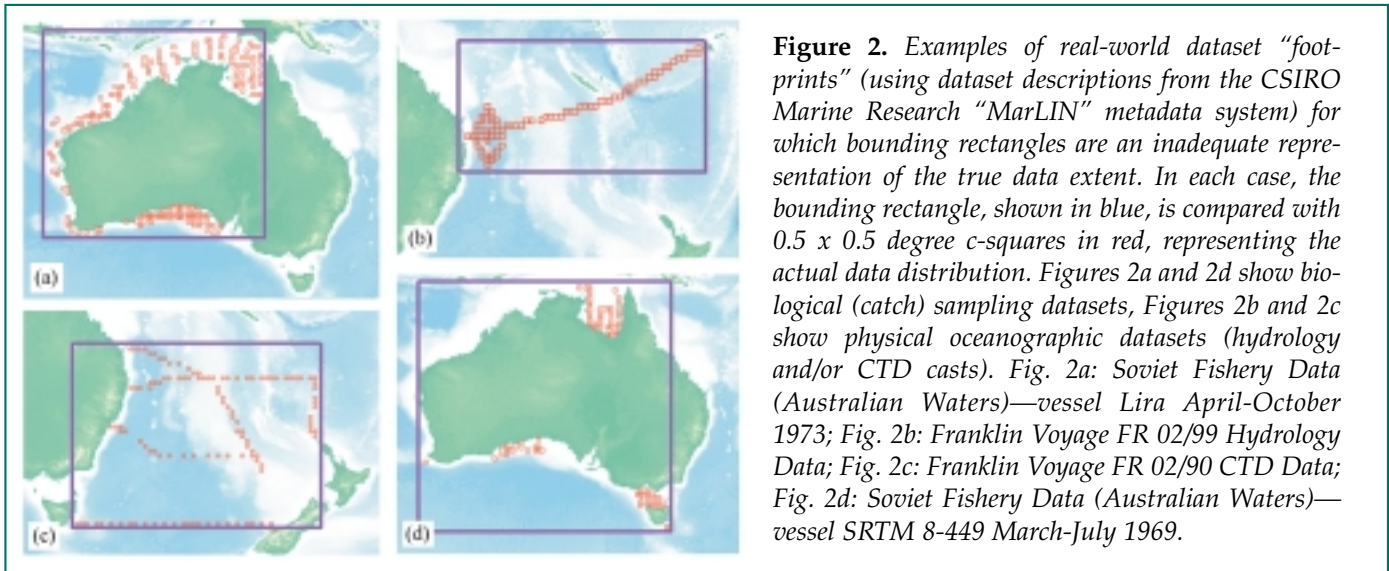
For the past five years, the author's agency (CSIRO Marine Research in Australia) has operated a metadata catalog "MarLIN" (www.marine.csiro.au/marlin/) to describe its data holdings (see Rees and Finney, 2000). Until recently, as with similar systems elsewhere, spatial searches of MarLIN could only be conducted using location keywords or bounding rectangles. Increasing awareness of the inherent problems with "bounding rectangles" searches has resulted in the development of a new, globally applicable system for spatial indexing and searching entitled "c-squares"—for *Concise Spatial Query And Representation System* (Rees, 2002a; 2002b)—which is described in detail for the first time below.

"C-Squares" Description
System Principle

C-squares is built on the principle that the surface of the earth can be divided up into a grid of labeled squares (individual "c-squares"), at one of a range of scales, and that any type of dataset spatial extent can then be represented as a coded list of those squares in which the data occur. This list, referred to hereafter as the *c-squares string*, is stored using simple (ASCII) text comprising numbers and separator characters, and if required, can be matched to an inquirer's designated search region (itself expressed as one or more c-square codes) using standard text search methods. It can also be used to generate a representation of the dataset footprint on a map; included in metadata export or exchange; and can provide the capability for remote spatial searching of stored metadata by any system which is capable of interpreting the numbering convention employed.

Using c-squares, search reliability can be substantially improved compared with bounding rectangles, since testing for "hits" or "misses" now takes place at the resolution of the individual c-square (e.g. 1 degree resolution or finer) rather than at the resolution of the overall rectangle which can frequently be one, or several orders of magnitude larger. Therefore, all "hits" returned should now be true "data" hits, within the limit of resolution implicit in the choice of grid

Figure 2. *Examples of real-world dataset "footprints" (using dataset descriptions from the CSIRO Marine Research "MarLIN" metadata system) for which bounding rectangles are an inadequate representation of the true data extent. In each case, the bounding rectangle, shown in blue, is compared with 0.5 x 0.5 degree c-squares in red, representing the actual data distribution. Figures 2a and 2d show biological (catch) sampling datasets, Figures 2b and 2c show physical oceanographic datasets (hydrology and/or CTD casts). Fig. 2a: Soviet Fishery Data (Australian Waters)—vessel Lira April-October 1973; Fig. 2b: Franklin Voyage FR 02/99 Hydrology Data; Fig. 2c: Franklin Voyage FR 02/90 CTD Data; Fig. 2d: Soviet Fishery Data (Australian Waters)— vessel SRTM 8-449 March-July 1969.*

Figure 3. *Nomenclature for 10 x 10 degree WMO squares. (a): Numbering principles in the four global quadrants. (b): Example numbered WMO squares in the Atlantic and Pacific Oceans. Figure 3b adapted, with permission, from Curry (2001).*

square size. A certain degree of economy (conciseness) at the metadata level is also implied, in that multiple data occurrences within the same square are only coded once.

A final, but important, element of the c-squares system is that the nomenclature adopted for the individual squares should be capable of being applied at a range of scales to suit different types of data. The preferred mechanism for this should be a hierarchical system, with the resulting smaller and smaller sub-units retaining a simple relationship to latitude and longitude. The latter should be expressed in decimal degrees, which is the accepted unit for dataset spatial extents in most national metadata standards as well as in the forthcoming ISO (International Organization for Standardization) standard 19115 for "Geographic Information—Metadata" (see www.anzlic.org.au/asdi/isomap.htm).

C-Squares Nomenclature

As described above, the basic principle of c-squares could in theory be applied using any of a variety of notation systems for the individual squares. In practice, however, it is necessary to standardize on a single nomenclature in order for the results to be seamlessly interoperable between computer systems on both inter-agency, and international levels. The grid system selected to form the basis of "c-squares" is the WMO (World Meteorological Organization) grid, which divides the world into 648 10 x 10 degree squares, each with a unique 4-digit identifier or code. The first digit of this code indicates the square's "global quadrant", i.e. north-east (1), south-east (3), south-west (5) or north-west (7) as indicated in Figure 3a. The remaining three digits are derived from the minimum (smallest absolute value) bounding parallels measured in tens of degrees: digit 2 (0 through 8) for 0+ to 80+ degrees of latitude (north or south, depending on the global quadrant designated), and digits 3-4 (00 through 17) for 0+ to 170+ degrees of longitude (east or west, depending on the global quadrant). As an example, the 10 x 10 degree square bounded on its "minimum" sides by 30° N and 160° E—in the Pacific Ocean, to the east of Japan—is allocated the code 1316, the square with its origin at 30° S and 160° E is allocated 3316, and so on. A global coverage of WMO square nomenclature is
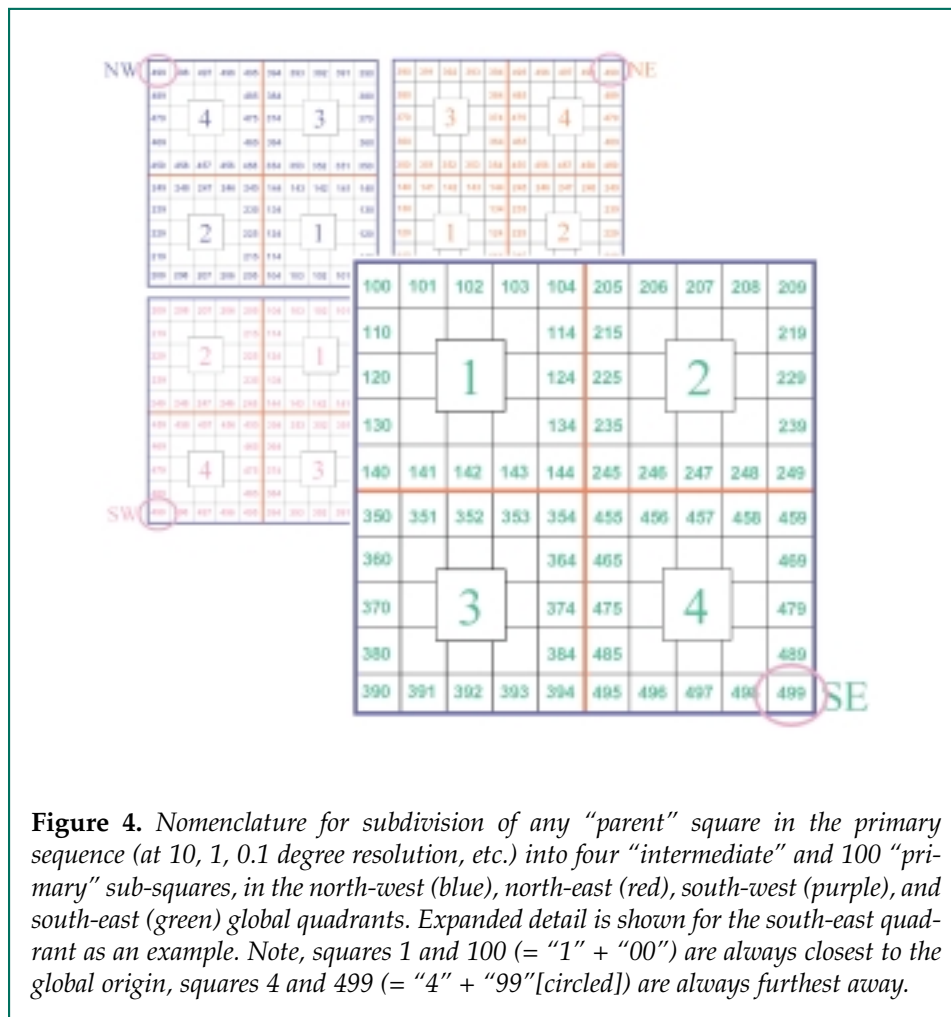

Figure 4. *Nomenclature for subdivision of any "parent" square in the primary sequence (at 10, 1, 0.1 degree resolution, etc.) into four "intermediate" and 100 "primary" sub-squares, in the north-west (blue), north-east (red), south-west (purple), and south-east (green) global quadrants. Expanded detail is shown for the south-east quadrant as an example. Note, squares 1 and 100 (= "1" + "00") are always closest to the global origin, squares 4 and 499 (= "4" + "99"[circled]) are always furthest away.*

given in Appendices 10A and 10B of NODC (1998), while coverage of principal ocean basins (including polar aspects) is illustrated in Curry (2001), from which diagrams for the Atlantic and Pacific Oceans are reproduced, with minor modification, in Figure 3b.

C-squares extends this numbering system by recursive subdivision according to two sequences: a primary sequence at 10 x 10, 1 x 1, 0.1 x 0.1 degree squares, etc., and an "intermediate" set of subdivisions (quadrants) at 0.5 of each of these values - giving 5 x 5, 0.5 x 0.5, 0.05 x 0.05 degree squares, etc. The notation adopted builds on a format previously utilized in the Australian "Blue Pages" metadata system for data searching (AODC/ERIN, 1996), as follows:

- 10 x 10 degree square:
  1316 (= WMO 10-degree square notation)
- 5 x 5 degree square:
  1316:2 (= "Blue Pages" 5-degree square notation)
- 1 x 1 degree square: 1316:225 (= "Blue Pages" 1-degree square notation)
- 0.5 x 0.5 degree square:
  1316:225:4 (c-squares extended notation)
- 0.1 x 0.1 degree square: 1316:225:469 (c-squares extended notation)
- etc.

This notation can extended indefinitely to give any desired resolution, each succeeding set of three digits (plus colon separator) representing a further tenfold reduction in square size.

Following the initial 4-digit WMO square code, each additional group of three digits includes one digit (1 through 4) indicating the relevant "intermediate" quadrant number (Figure 4), where 1 indicates additional units of latitude and longitude are both "low" (e.g. between 0 and 5 on a 0-10 scale), 2 indicates latitude low and longitude high, 3 indicates latitude high and longitude low, and 4 indicates latitude and longitude both high (e.g. between 5 and 10 on a 0-10 scale). Following this initial digit are two further digits, 00 through 99, which indicate the position of the relevant "primary" sub-square (one tenth linear division, one hundredth division by area) by units/sub-units of latitude (digit 2) and longitude (digit 3), within its parent square ten times larger. When these digits are combined, individual sub-squares have 3-digit identifiers as follows:

- *Quadrant 1:* 100 through 104, 110 through 114, 120 through 124, 130 through 134, 140 through 144
- *Quadrant 2:* 205 through 209, 215 through 219, 225 through 229, 235 through 239, 245 through 249
- *Quadrant 3:* 350 through 354, 360 through 364, 370 through 374, 380 through 384, 390 through 394
- *Quadrant 4:* 455 through 459, 465 through 469, 475 through 479, 485 through 489, 495 through 499

arranged according to the system illustrated in Figure 4.

It should also be noted from Figure 4 that, as with the 10 x 10 degree WMO squares, the arrangement of numbered sub-squares forms mirror images in the various global quadrants, both across the equator and across the Greenwich Meridian. Thus in North America, for example (north-west global quadrant), square 499 is always at the upper left of its parent square, while in Australia (south-east global quadrant) it is at the lower right.

In use, this coding system has the important characteristic that the minimum latitude and longitude boundaries of any square (expressed in decimal degrees) are represented directly as numbers within the relevant c-square code. Thus, for example, the 0.1 x 0.1 degree square extending from 32.6° to 32.7° N and 165.9° to 166.0° E is represented as *1316:225:469* which, on inspection, can be seen to encode *1316:2<u>2</u>5:4<u>6</u>9* (32.6° of latitude) and *13<u>1</u>6:22<u>5</u>:46<u>9</u>* (165.9° of longitude), in global quadrant 1 (*<u>1</u>316*), i.e. measured N and E, respectively. The remaining digits (*<u>2</u>25* and *<u>4</u>69*) are automatically assigned according to the principle of intermediate quadrants illustrated in Figure 4.

The actual length of the code indicates the spatial resolution encoded, which is also translatable (as a first approximation) into real-world resolution, since 1 degree of latitude is equal to approximately 110 km on the earth's surface, while 1 degree of longitude is never more than 110 km in distance (and tends towards zero at the poles). Thus, to achieve a resolution of the order of 100 km or better, a 1 x 1 degree square is required (8 characters), while around a 1 km resolution (or better) would be achieved with a 0.01 x 0.01 degree square (16 characters).

Storage of Dataset Footprints Using C-squares

A dataset footprint expressed in c-squares is simply a string of the relevant codes, at a resolution selected by the metadata creator, separated by the vertical bar or "pipe" character "|", e.g. (at 1 x 1 degree resolution) *3311:144|3311:134|3311:124*, etc. The resolution adopted for a particular c-squares string will vary according to the needs of the user and the spatial scale of query which the metadata system will be intended to support, balanced against the overhead of holding increasingly long c-squares strings which would be required as resolution becomes finer. Obviously this can vary greatly from one application, and user's requirement, to another. As an example, in the "MarLIN" data directory we have chosen to encode a range of data types from discrete oceanographic research voyages at 0.5 x 0.5 degree resolution, resulting in the creation of strings ranging from 1 to 239 codes (<3000 characters), with 80% of these strings being within the range 5 to 80 codes (see examples in Figures 2 a-d). By contrast, in another database (the "CAAB" master taxon list for aquatic species), data points (individual catch records for the species in question) have been encoded at 0.1 x 0.1 degree resolution so as to support more detailed spatial queries; c-squares strings created in this context can exceed 1000 codes, requiring >13000 characters since each code at this resolution is represented by 12 characters, plus a separator.

From the above examples, it follows that interoperable searching of dataset footprints expressed as c-squares strings across multiple systems will only be supportable at the minimum resolution of all contributing databases (this is equivalent to location keywords for oceanographic datasets, for example, being implemented at different scales by different users, and therefore not interoperable except at the highest level). Thus, in version 1.0 of the draft c-squares specification (available via the c-squares website at www.marine.csiro.au/csquares/) it is suggested that a resolution of at least 1 x 1 degree squares be used by metadata creators wishing to guarantee that their data will be suitable for distributed spatial searching using c-squares. Of course, this does not preclude individual databases using finer subdivisions as appropriate to individual data needs, since due to the hierarchical nature of the codes, the

C-squares is built on the principle that the surface of the earth can be divided up into a grid of labeled squares…

relevant "1 degree" portion is embedded in any higher resolution code and will always be available for searching at that level (see further discussion below).

To address situations where large dataset extents may be required to be described while maintaining a relatively fine resolution, a "wildcard" notation has been introduced into the c-squares specification, which permits an additional level of economy in creating c-squares strings. A shorthand notation such as "7311:*" indicates all four 5 x 5 degree squares within the 10 degree square 7311, "7311:***" indicates all one hundred 1 x 1 degree squares in the same parent square, and so on. By this means, where appropriate, internal contiguous blocks of squares may be encoded using the "wildcard" notation (resulting in shorter c-squares strings), while coding at the ultimate fine-scale resolution may only be required around these filled blocks—for example, closer to the footprint periphery. This is somewhat similar to the efficiencies introduced by variable-resolution arrays known as quadtrees (e.g. Samet, 1984), in which the depth of recursive subdivision (branching) is matched to the level of detail required to be encoded in different regions of a particular subject. However, for quadtrees the subdivisions are typically by a recursive factor of two (giving four quadrants with each successive subdivision) rather than the alternating "base two then base five" subdivisions of the present system.

Methods of Creation for C-Squares Strings

C-squares strings can potentially be created by several methods. The method used so far at CSIRO Marine Research involves running an algorithm over spatially referenced point data in a table or database which will generate the relevant c-square code for each point at a user-specified resolution, then creating an automated procedure to test each code to see if it is new (to prevent multiple instances of the same code in the c-squares string), and only then add it to the string. Examples of algorithms for c-squares conversion (encoders) are available on the c-squares website at www.marine.csiro.au/csquares/, in three different languages: the initial Oracle PL/SQL version developed at CSIRO Marine Research, and versions in Java, and ColdFusion, courtesy of OBIS and FishBase developers Phoebe Zhang and Eli Agbayani, respectively.

Potential also exists to generate c-square strings from vector (e.g. polygon) information stored in a GIS or similar system, however an additional stage of vector-to-raster conversion would most likely be required as a preliminary step. Tools for such encoding are currently under development. A third option (requiring no technological overhead at all) is that the strings of codes can simply be created manually, with reference to a map on which squares delimited by divisions of latitude and longitude have been annotated with the relevant c-square codes.

Spatial Searching Using C-Squares

Spatial searching of dataset footprints expressed as c-squares strings is very simple, and merely requires looking for a match between a user's search region (itself expressed as one or more c-square codes) and any part of the c-squares string—that is, a straightforward text search of the type already supported by any text-based database. Because the codes are hierarchical, a user's search, for example on the 5 x 5 degree square "1316:2" will be capable of matching codes at multiple resolutions such as 1316:2, 1316:227, 1316:227:466 and so on. A simple extension of the search procedure can also be included so that a match will also be returned on "wildcard" notation such as 1316:*, 1316:***, etc. Implicit in this matching procedure, however, is that such a search should not match a higher resolution code such as 1316 alone, unless some qualification is returned to the user to the effect that the result is only a "possible" rather than a "confirmed" hit (the limitation being the resolution encoded in the original metadata).

Spatial searching can be implemented as a process for selecting a single search square from many possibilities at a range of scales (e.g. the current MarLIN c-squares search interface at www.marine.csiro.au/marlin/csq-chooser.htm), or by constructing a user interface which would permit the designation of more complex search regions composed of multiple c-squares—either user-defined, or to match pre-defined spatial objects such as administrative areas, etc. For search regions composed of multiple c-squares, the search algorithm would need to test each component c-square within the user's search region in turn, against stored c-squares strings representing individual dataset extents. While taking more time to execute, this requires no more complex logic than a normal text search including the Boolean "OR" (alternative) clause—for example, searching for "Atlantic OR Pacific"—and could be made efficient by incorporation of standard text database indexing functions if required, e.g. by treating every c-square code within a string as a separate "word".

Display of Dataset Footprints Using C-Squares

A benefit of using c-squares to encode and store dataset footprints is that either pre-existing, or specially created tools can be used for display of the dataset extent on a range of base maps. Initial testing of the prototype c-squares system was carried out using the (then) web-accessible *Xerox PARC Map Viewer* at pubweb.parc.xerox.com/map/ (see Wood et al., 2000 for example output), which currently has been suspended as a web service; however an alternative exists in the

...the applicability of c-squares is not limited to metadata systems...

form of Charles Sturt University's *Map Maker* (life.csu.edu.au/cgi-bin/gis/Map). In both these cases, a simple "c-squares decoder" can be written to obtain (for example) the center point of any supplied c-square, and then send a series of such points (as latitude/longitude pairs) for mapping via the web. A more elegant alternative is to use a mapper which can interpret the c-square codes directly. Such a utility has now been constructed by CSIRO Marine Research (the CMR *c-squares mapper*) and like its precursors mentioned above, can be addressed directly via the web to produce maps to suit the user's own data, along with (in this instance) the option for a map title and legend. Advantages of using such a mapper include removing the requirement to decode the c-squares strings before transmission across the web, and correct scaling of the squares as the map scale is changed.

Development to include depth or the fourth dimension (time) could also theoretically be envisaged…

Currently, the range of base maps on which data footprints can be plotted in this mapper includes topographic maps created from earth observation data at various scales, sourced from the National Geophysical Data Center (NGDC) GLOBE project (www.ngdc.noaa.gov/seg/topo/globe.shtml), as well as modeled sea surface temperature maps provided by CSIRO Division of Atmospheric Research in Australia, and will be expanded further over time. Example output from the c-squares mapper includes the images in Figures 2a-d and 5d, and numerous additional examples can be generated on-line via sources including MarLIN (www.marine.csiro.au/marlin/), FishBase (www.fishbase.org/), or the new OBIS (Ocean Biogeographic Information System) portal (www.iobis.org/). For more information regarding the CMR c-squares mapper and how to link to it, see the page www.marine.csiro.au/csquares/about-mapper.htm on the c-squares website.

Discussion

Aspects of c-squares are not new: indeed, wherever possible, previously established standards have been incorporated so as to avoid creation of unnecessary new notation concepts for users to learn. The global grid selected as the basis for c-squares compares well with Clarke's (2000) list of desirable attributes for a geo-referencing system (such as being universal, authoritative, hierarchical, and tractable, among others), while the "Blue Pages" notation for 5- and 1-degree subdivisions of WMO squares has been adopted as being an ingenious solution for searching at both "primary" and "intermediate" levels, and also as a suitable basis for further extension. The concept of spatial indexing and searching using grid squares is likewise not new—previous examples include the ICES *ROSCOP* database at www.ices.dk/ocean/roscop/,

which uses 5 x 5 degree Marsden Square subdivisions, and Museum Victoria (Australia)'s *Bioinformatics* website at www.museum.vic.gov.au/bioinformatics/ which employs 0.5 x 0.5 degree squares labeled with local mapsheet numbers. Nevertheless, c-squares appears to be novel in the following respects:

- proposing a generalized solution to the limitations of "bounding rectangles" representation of dataset footprints and associated search procedures;
- providing an unlimited, recursive nomenclature for subdividing WMO squares;
- promoting a syntax for general use, which expresses dataset footprints as strings of c-squares codes, suitable for storage as metadata, querying, and metadata exchange; and
- providing dedicated tools for real-world generation and display of the dataset footprints concerned.

Its principal benefits in practice are perceived as being (a) the flexibility of the system to represent a wide variety of shapes and sizes of dataset footprint, (b) the relative simplicity of the storage and query mechanisms for the c-square strings, and (c) the ability to decouple this information from the base data for the purpose of spatial searching and display. Being standard ASCII text, c-squares strings can also easily be included in metadata exchange, and are also well suited for expression in the newer generation Extensible Markup Language (XML) metadata formats—e.g. within a potential new element "*<csquares></csquares>*".

At the same time, certain limitations of c-squares should be mentioned: (1) c-squares are not equal-area, which means that distances and areas cannot readily be computed directly from the c-squares strings; (2) strings can become quite long for large, complex regions (e.g. "World Ocean") at detailed resolutions, even using the "wildcard" notation supported, which may prove an overhead for metadata storage in some situations; (3) a balance needs to be struck (for any one dataset description) between c-square resolution and lengths of the strings which it is feasible to store, since progressively finer resolution permits more detailed searching, but also requires geometrically increasing numbers of codes; and (4) c-squares can be ambiguous at boundaries, i.e., data identified as being present in a square which is traversed by a boundary cannot be further identified as to which side of the boundary they fall without supplementary information.

Certain other restrictions of the system which apply currently are expected to be addressed in the next phase of development of tools for c-squares. For example, at present, c-squares operates on a "presence/absence" (binary) basis—a square is either encoded (present) or not encoded (absent). Future development of the mapper at CSIRO Marine Research is planned which will enable multiple c-square strings (potentially
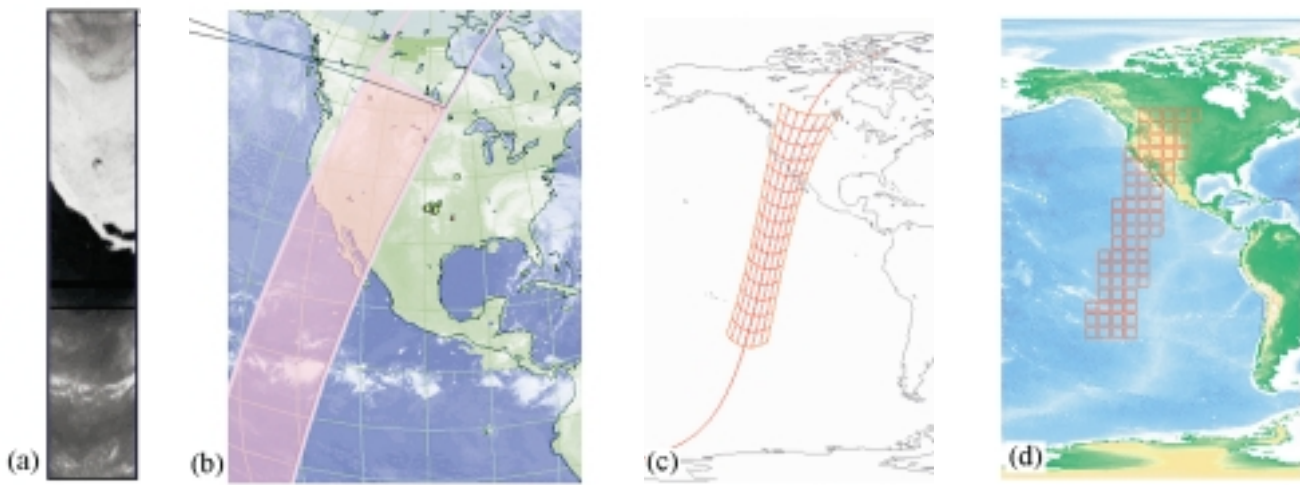
Figure 5. *Potential application of c-squares for indexing swath data from a satellite pass; for detail see text. Images in 5a-c courtesy of Dr. Mike Botts and the "Space-Time Toolkit" (vast.uah.edu/).*

representing a greater variety of data states than just presence/absence) to be plotted. Another planned feature is to be able to return "active maps" to the user, which would then respond to clicking on any individual square by triggering a query to the appropriate database and extracting actual data relevant to that particular square.

While the examples shown in Figure 2 are of vessel-collected oceanographic data, c-squares is equally applicable for indexing many other data types as well. An example use with satellite swath data is shown in Figure 5. A typical swath image (Figure 5a) is of constant width (and follows a linear orbit) on the surface of the globe (Figure 5b), but when plotted on a cylindrical projection (Figure 5c) the path is sinusoidal and the width in degrees varies according to latitude. This would not be well represented by a bounding rectangle; however, representation using c-squares (Figure 5d) is quite straightforward, sufficient to discover whether or not this pass overlaps an inquirer's desired "search" region, at least to the resolution of the squares employed (a "coarse" 5 x 5 degrees in this demonstration plot).

C-Squares Status

Since the c-squares concept is still relatively new, future refinements of the specification or available tools are expected; some potential directions for development to the mapper have already been indicated above. Details of such developments will be posted on the c-squares website as they are introduced. Development to include depth (c-cubes?) or the fourth dimension (time) could also theoretically be envisaged, however, at least for the present, it has been a conscious decision to keep c-squares to two dimensions—

partly so as to be directly analogous to the bounding rectangle and bounding polygon methods of spatial representation for dataset footprints, and also to keep the system to a minimum level of complexity for potential users.

As a final comment, the applicability of c-squares is not limited to metadata systems; while not the only contender in this area, it does offer the potential to be incorporated (as a kind of "global postal code" applicable to both terrestrial and marine data) into any conventional web document or address which it is desired to render identifiable and retrievable according to location on the earth's surface. Indeed, no special technology is required for such a system to be functional, as an internet search -e.g. via "Google" at www.google.com—for a named c-square (such as for the phrases "c-squares" and "3215:495:4") will demonstrate already, for those interested in exploring the possibility of using c-squares as an open, easily implemented spatial indexing system. ▣

References

Note: all WWW addresses (URLs) current as at December 2002

AODC (Australian Oceanographic Data Centre) / ERIN (Environmental Resources Information Network), 1996: *The Marine and Coastal Data Directory of Australia—The Blue Pages.* WWW application, search interface accessible at http://www.marine.csiro.au/marine/mcdd/, logfile (including nomenclature for WMO square subdivisions) at http://www.marine.csiro.au/marine/mcdd/log_analysis.html.

Clarke, K.C., 2000: Criteria and measures for the comparison of global geocoding systems. In: *Proceedings, International Conference on Discrete Global Grids, March 2000.* U.S. National Center for Geographic Information and Analysis, Santa Barbara. WWW document, available via conference website at http://www.ncgia.ucsb.edu/globalgrids/abs.html.

Curry, R., 2001: *HydroBase 2—A Database of Hydrographic Profiles and Tools for Climatological Analysis. Technical Reference, Preliminary Draft, November 2001.* Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, 81 pp. WWW version available at http://www.whoi.edu/science/PO/hydrobase/TechRef_draft.pdf.

FGDC (Federal Geographic Data Committee), 1994: *Content Standards for Digital Spatial Metadata (June 8 draft).* Federal Geographic Data Committee. Washington, D.C. WWW document, available at http://geology.usgs.gov/tools/metadata/standard/metadata.html.

NASA, 1993: *Directory Interchange Format Manual, Version 4.1 April 1993.* World Data Center A for Rockets and Satellites Publication 92-20, NASA Goddard Space Flight Center, Greenbelt, MD. WWW version available at http://nssdca.gsfc.nasa.gov/anon_dir/md_doc/DIFMANUAL.DOC.

NODC (U.S. National Oceanographic Data Center), 1998: *World Ocean Database 1998: Documentation and Quality Control, Version 1.2 (National Oceanographic Data Center Internal Report 14).* Ocean Climate Laboratory, National Oceanographic Data Center (U.S.), Silver Spring, MD, 114 pp. Appendices 10A and 10B (WMO squares for Atlantic/Indian and Pacific Oceans) available on the WWW at http://www.nodc.noaa.gov/OC5/wmoatlind.html and http://www.nodc.noaa.gov/OC5/wmopacific.html.

Rees, T., 2002a: "C-squares"—a new metadata element for improved spatial querying and representation of spatial dataset coverage in metadata records [abstract]. In: *Abstracts, EOGEO 2002 Technical Workshop, May 2002.* European Commission Joint Research Centre, Ispra, Italy. WWW document, available via http://eogeo.jrc.it/ (follow link "EOGEO 2002 Proceedings").

Rees, T., 2002b: C-squares—a new method for representing, querying, displaying and exchanging dataset spatial extents [abstract]. In: *The Colour of Ocean Data—International Symposium on Oceanographic Data and Information Management With Special Attention to Biological Data. Brussels, Belgium, 25–27 November 2002.* M. Brown, M. Costello, C. Heip, S. Levitus, J. Mees, P. Pissiersens and E. Vanden Berghe, eds., Flanders Marine Institute (VLIZ), Oostende, 81.

Rees, T. and K. Finney, 2000: Biological data and metadata initiatives at CSIRO Marine Research, Australia, with implications for the design of OBIS. *Oceanography*, 13(3), 60–65.

Samet, H., 1984: The quadtree and related hierarchical data structures. *ACM Comput. Surv.*, 16(2), 187-260.

Wood, J.W., C.L. Day, P. Lee and R.K. O'Dor, 2000: CephBase: testing ideas for cephalopod and other species-level databases. *Oceanography*, 13(3), 14–20.