Environmental Modelling & Software 62 (2014) 240–252

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

A data porting tool for coupling models with different discretization needs

S. Ion [a], D. Marinescu [a, *], S.G. Cruceanu [a], V. Iordache [b]

[a] *"Gheorghe Mihoc-Caius Iacob" Institute of Statistical Mathematics and Applied Mathematics, Romanian Academy, Calea 13 Septembrie No. 13, Bucharest 050711, Romania*
[b] *University of Bucharest, Research Center for Ecological Services, Romania*

ARTICLE INFO

Article history:
Received 25 June 2014
Received in revised form
9 September 2014
Accepted 10 September 2014
Available online

Keywords:
Raster data
Interpolation
Cubic polynomials
Hexagonal raster
Hydrological modelling
Water routing

ABSTRACT

The presented work is part of a larger research program dealing with developing tools for coupling biogeochemical models in contaminated landscapes. The specific objective of this article is to provide researchers with a data porting tool to build hexagonal raster using information from a rectangular raster data (e.g. GIS format). This tool involves a computational algorithm and an open source software (written in C). The method of extending the reticulated functions defined on 2D networks is an essential key of this algorithm and can also be used for other purposes than data porting. The algorithm allows one to build the hexagonal raster with a cell size independent from the geometry of the rectangular raster. The extended function is a bi-cubic spline which can exactly reconstruct polynomials up to degree three in each variable. We validate the method by analyzing errors in some theoretical case studies followed by other studies with real terrain elevation data. We also introduce and briefly present an iterative water routing method and use it for validation on a case with concrete terrain data.

Software availability

Name of software:: Asterix Porting Data 1.0.
Developers:: Stelian Ion, Dorin Marinescu, Stefan G. Cruceanu.
Email:: dorin.marinescu@ima.ro
Year first available:: 2013.
Program Language:: C.
Software Required:: 32 bit Linux operating system, gcc compiler, SDL libraries.
Program size:: 2 MB.
Availability:: free downloadable program sources at http://www.ima.ro/software/asterix_porting_data.tar.xz

1. Introduction

Predicting the long distance effects of local environmental changes requires a coupling between local and regional models of ecological and abiotic processes. Examples include the integration of local vegetation processes with regional transport models for heavy metals (e.g. Iordache et al. (2009, 2011, 2012) for reviews and general methodological steps), or the local resources development with movement of animal species (e.g. Holmes et al. (1994) for review of Partial Differential Equations (PDEs) used in spatial interactions and population dynamics, and Gough and Rushton (2000) for an application using Geographic Information Systems (GIS)).

Direct coupling of the models is technically possible, but allows less flexibility for further model development. Alternatively, more common and flexible geographical objects are used as an interface among models of processes occurring at different space-time scales. One specific problem with this integration strategy is that the kind and properties of the geographical objects used as input to or output from local and regional models are usually different (see Voinov and Shugart (2013) for some examples of problems raised by integrated modelling). The differences arise due to methodological constraints related to the measurements of the space

---

Abbreviations: CRS, Catmull-Rom Spline; DEM, Digital Elevation Model; DP, Data Porting; DTM, Digital Terrain Model; ENO, Essentially Non-Oscillating; GIS, Geographic Information System; LiDAR, Light Detection And Ranging; OF, Outlier Filtering; PDEs, Partial Differential Equations.

* Corresponding author. Tel./fax: +40 213182439.
E-mail addresses: stelian.ion@ima.ro (S. Ion), dorin.marinescu@ima.ro, marinescu.dorin@gmail.com (D. Marinescu), stefan.cruceanu@ima.ro (S.G. Cruceanu), virgil.iordache@bio.unibuc.ro (V. Iordache).

http://dx.doi.org/10.1016/j.envsoft.2014.09.012
1364-8152/

variables and to the modelling techniques. Table 1 summarizes the types of geographical objects which can occur (summarized and developed from Goodchild (1992)).

It can be seen from the inspection of this table that raster data and vector data are not basic terms in geographical ontology (see also Galton (2004)). For instance a raster refers in principle to a square tessellation of the plane with a constant value of the space variable function inside each square. Although raster data models are not a primary information source, in practical modelling they are frequently the only available primary data source scales (e.g. in digital terrain models). A large volume of GIS raster data is nowadays collected and used for scientific research purposes, as well as for different practical applications, and a wide variety of software has been developed over time for reading and processing it Hillier (2011); Chambers (2008) (one may also see Varekamp et al. (1996) for a practical guide to use public domain geostatistical and GIS software). The data of a raster are organized into rows and columns and structured as a matrix. Any information of interest is characterized by a unique value in each cell which maybe "null" if no data is available. This information may represent continuous data, as elevation, temperature, rainfall intensity or discrete (thematic, nominal) data, as land use or soil category, Hillier (2011). The fidelity of a raster data with respect to the real information is a challenging issue Eberly et al. (November 2004); MacKay (1992); Mitas and Mitasova (1999); Moore et al. (1991); Smith et al. (2004) that must be taken into consideration since the errors propagate on raster data outputs Pelletier (2010); Vazquez and Feyen (2007); Walker and Willgoose (1999). Usually, large cells entail poor accuracy. Some mathematical models of environmental phenomena require information at a subgrid-scale. For example, the input data for partial differential equations must satisfy a certain degree of smoothness. Also, the domain discretization may be better suited to configurations other than that of common square cells. Such situations may appear when one uses finite volume or finite element methods to numerically solve PDEs, as well as when one uses models based on hexagonal cellular automata Baetens et al. (2013). With the development of laser altimetry in geography and ecology (Kumhálová et al., 2013; Srinivasan et al., 2014), point clouds tend to be more often used for direct analysis within a GIS. Hexagonal lattices can be used for the extraction of knowledge by clustering techniques from such data (e.g. Jiang, 2004; Hagenauer and Helbich, 2013). Using the rasters obtained by interpolations of Light Detection And Ranging (LiDAR) data, Moreira et al. (2011) have pointed out the important influence of Digital Terrain Model (DTM) resolution on the estimated soil loss by erosion modelling. This influence could be explored also by erosion models using hexagonal lattices if the porting software would be available. In this context, one can say that there is a need to build spatial interpolation algorithms for the purpose of porting rectangular raster information to other networks with cells of a different size or geometry, i.e. a method of *data porting* (DP). These algorithms should be developed in such a way to preserve as much as possible the original measured space variables. In this article we tackle the particular situation of variables treated by a field type approach, observable at a scale much smaller than the derived geographical objects and the empirical precision of geographical location, having the size of the derived and not empirically constrained polygons. In particular, we present a DP method to construct a hexagonal raster using a rectangular raster data input. To the best of our knowledge and according to Wei et al. (2012), there are currently no databases provided in "hexagonal raster format".

1.1. Data porting methods

A rough classification divides DP into one stage and multi-stage methods. In a one stage method, the data value corresponding to a cell c_i of a new grid is given by the direct inspection of the values corresponding to the old grid cells that intersect c_i. The methods in this class (e.g. the *nearest neighbor*, Dodgson (1992), *area weighted mean*, Wei et al. (2012), *and kriging methods*, Cressie (1993)) approximate the discrete raster function. In Gardner et al. (2008), the authors develop a rescaling method also applicable to the transformation of grid configuration (from rectangular to triangular or hexagonal). The method is based on sampling points on the original grid located around the point in the center of a pixel of the rescaled grid. The kriging methods equally work on regular or irregular sample points, but they are not effective on regular and dense grids. Moreover, they require a set of information concerning the correlation functions. Such information must be known in advance or can be inferred by a proper process from the existing data (see also Matheron, 1981; Dubrule, 1983). The multi-stage class methods Dodgson (1992); Lee et al. (1997) assume the existence of one or more intermediate (everywhere defined) functions from which the cell values in the new grid are sampled. In this article, we propose a two stage method similar to one widely used in image resampling, Dodgson (1992); Lee et al. (1997). The basic assumption is that the raster values represent the point values of an everywhere defined function that models a certain physical property. In most cases, the function is not analytically known and the raster values are calculated from the values of

Table 1
Types of geographical objects occurring directly and indirectly in the modelling of coupled environmental process (reconstructed and adapted from Goodchild (1992)). Field type approach allows a rigorous description of the error and empirical verification, while discrete type approach does not allow a good treatment of errors, usually involves a filtering of empirical data.

Approach	Measurement, observation	Primary ("real") geographical objects	Derived ("methodological") geographical objects (resulted from data modelling and plane discretization)
Field type (properties with relation of spatial location)	Variables z − observable at a scale much smaller than the derived geographical objects and the empirical precision of geographical location Variables z − observable at a scale larger than the potentially derivable geographical objects and the empirical precision of geographical location	Tuples $<x,y,z_1,\ldots,z_n>$ with space variables z and location (x,y). Tuples with (x,y) on the observation polygon or on the transect line (e.g the center of the 50 × 50 m plot). Interpolated field (the infinite set of tuples).	Substrates with attributes (polygons, contour lines, regularly distributed points as centers of a plane discretization). The polygons are characterized by a variation of the variables inside them (a constant in the simplest case). The size of the polygons is not empirically constrained; Substrates with attributes. It makes no natural sense to have the size of the discretization units smaller than the observation scale of the spatial variables.
Discrete type (substrate located in space with properties)	The observation scale does not influence the approach. Usually this approach makes use of old geographical maps or methodological objects resulted from field approach.	Points, polygons, lines filling and empty geographical space.	Field obtained by planar enforcement (points with a certain value inside the discrete object, by class, and a constant value in the empty space).