Geocomputing Made Simple: Service-Chain Based Automated Geoprocessing for Precision Agriculture

Edward Nash[1], Jans Bobert[2], Karl-Otto Wenkel[2], Wilfried Mirschel[2], Ralf Wieland[2]

[1]Institute for Management of Rural Areas, Rostock University, Justus-von-Liebig-Weg 6, 18059 Rostock, Germany
Telephone: ++49 381 498 3212
Fax: ++49 381 498 3202
Email: edward.nash@uni-rostock.de

[2]Leibniz Centre for Agricultural Landscape Research (ZALF),
Institute of Landscape Systems Analysis, Eberswalder Strasse 84, 15374 Müncheberg, Germany
Telephone ++49 33432 82 161
Fax: ++49 33432 82 334
Email: bobert@zalf.de

1. Introduction

Agricultural fields often have internal variations in their physical properties such as soil or relief. This heterogeneity leads to yield variation across the fields that may justify a site-specific management (precision agriculture, PA) to optimise productivity and minimize ecological damages. PA is a relatively new technique that uses the global positioning system (GPS), non-destructive local or remote sensors and variable rate technology to allocate the optimum input amounts of seeds, fertilisers and pesticides. Because influences and interdependencies of factors determining site-specific yield are complex and not always understood, straightforward approaches for the delineation of zones with similar yield potential, and which can therefore be similarly managed, are a tool for a simple and effective PA. Up to now many approaches for the delineation of management zones (MZ) have been discussed. Whelan and McBratney (2003) categorise five groups of approaches;

1. hand drawn polygons based on yield maps or imagery,
2. classification of remote sensed data,
3. identification of yield stability patterns across seasons at fixed monitoring points,
4. fuzzy multivariate cluster analysis using seasonal yield maps,
5. morphological filters or buffering.

With exception of the first approach, all methods require knowledge and experience with GIS which can not be presupposed from a farmer or even many agricultural consultants. Certainly there is a need for simple, cheap and easy-to-use approaches for the derivation of MZ for farmers and consultants. In this paper we present the use of a service-oriented architecture for automation of this geocomputing task, allowing the complete workflow to be run from simple client software.

2. Algorithm

Most farmers that already practice precision agriculture or are planning to introduce it use yield mapping (Reichardt, 2007). Almost every new combine is equipped with GPS and a yield monitoring system. Consequently the availability of multi-year yield maps as a basis for the delineation of MZ is good. Accurate yield maps depict the influences of site, climate and management factors on yield formation for a specific year. Multi-year yield maps contain valuable information about site-specific yield variability. In this paper we use the “hill climbing” cluster algorithm (Rubin, 1967) to delineate MZ from multi-year yield data. This hill climbing algorithm is an iterative local-search partitioning algorithm that allows the user to define the number of clusters to which the cells of the yield map will be assigned. For each resulting zone, the mean zone value in each source layer and a total variance are also produced.

In order to use the hill-climbing clustering, a certain amount of pre-processing is necessary. As the overall level of yield varies annually, we normalise the input layers using a simple approach of dividing each layer by its mean value. Each layer must also be available in a raster format with the cells aligned and therefore rectified. To produce the results in a suitable format for the farmer, some post-processing is also necessary. The zones will be polygonised, and zones which are too small to be sensibly managed are filtered. This filtering simply merges each undersized zone with an adjacent zone such that the overall perimeter length is minimised, i.e. the resulting zones should be as compact as possible. The complete workflow for producing a MZ map is shown in fig 1.

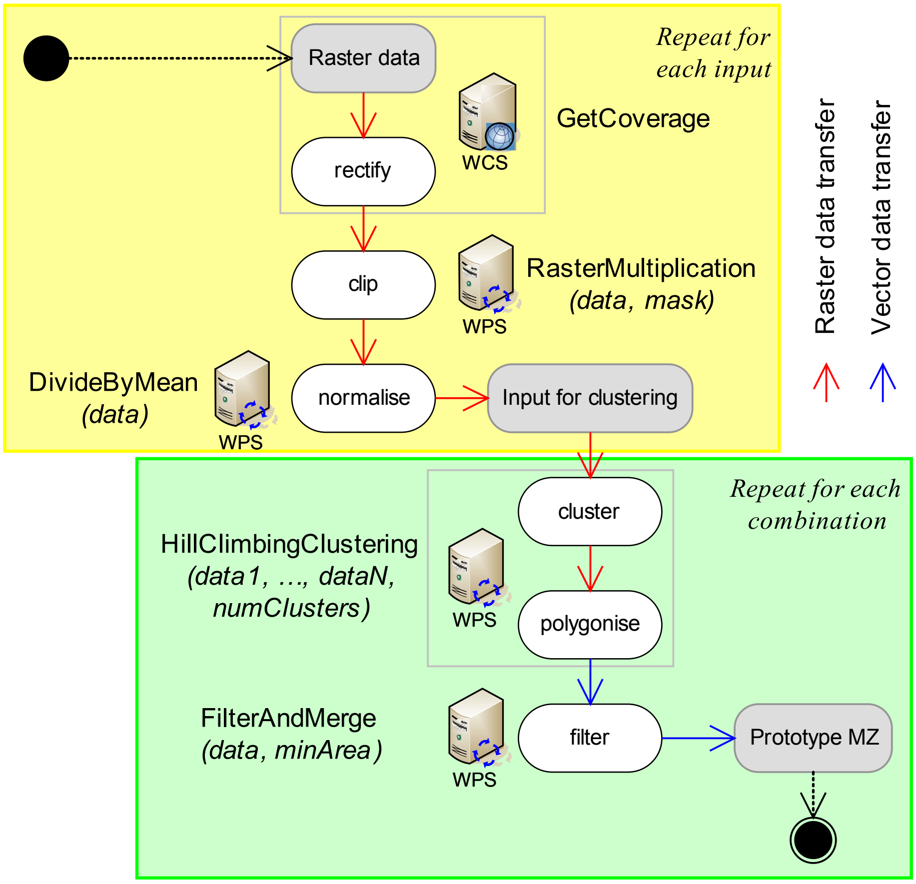


Figure 1. Annotated activity diagram illustrating the workflow and its implementation

3. Technologies

We have implemented the clustering algorithm and the necessary preparation and post-processing using individual web-services. In particular, the interface standards of the Open Geospatial Consortium (OGC) are used to provide access to distributed data and algorithms. The automation of the geoprocessing workflow is accomplished using a service chain implemented as an opaque, aggregated service, with suitable services to fulfil each stage dynamically found through searching a service catalogue. This search queries on keywords and process, input and output names to retrieve compatible services. The aggregated service model was chosen in this case to reflect the limited IT and GIS experience expected in the target audience and enable the use of simple clients, perhaps running on low-powered or handheld devices.

In general, we assume that the datasets to be used as the base for processing are available through a suitable web interface. The current implementation demands a degree of pre-processing (e.g. cleaning, interpolation/kriging, etc.) of the data such that it is directly available in a coverage format. We therefore assume that data providers, resellers or consultants could carry out this pre-processing in order to supply the data appropriately, although in future the workflow should be extended to work with the raw data, perhaps from an agricultural data warehouse as described in Steinberger et al (2006).

3.1 Interfaces

Two interfaces are used for core of the current workflow; the Web Coverage Service (WCS) for delivery of coverage data and the Web Processing Service (WPS) for algorithms. Extension of the workflow to allow 'raw' vector data would necessitate the use of a further interface, the Web Feature Service (WFS). In some cases, extensions to the current draft WPS (version 0.4.0) have been used, such as supporting m…n input cardinality instead of the present 0…1, where it is likely that these may be accepted in future versions.

3.2 Data formats

It is expected that the farmer will wish to view and edit the MZ using a variety of software. The resulting zones may therefore be delivered in a variety of formats, both generic (GoogleEarth KML, zipped ESRI Shapefile, GML) and agriculture-specific (agroXML with extensions for precision farming as described in Steinberger et al., 2007). Communication between services in the chain uses open generic formats; GeoTIFF for coverages and GML for vector data. This reflects the fact that services offering fundamental GIS functionality must accept generic GIS formats, whereas services offering specialised functionality must work with formats relevant to the target audience.

4. Implementation

Each geoprocessing operation has generally been implemented as a separate WPS web service as illustrated in fig 1. There are two exceptions to this:

1. The rectification of the inputs which, assuming data is being requested as a coverage from a web service, is automatically accomplished by setting the appropriate request parameters (coordinate reference system, bounds, width, height) in the request.
2. The polygonisation is integrated with the clustering to allow the statistical information to be given as properties of the resulting features.

A specialised WPS is also implemented to manage the workflow composed of generic services and convert the result into the requested (potentially agriculture-specific) format. This architecture is shown in fig 2. All these services have been realised using open-source software, integrating principally GeoTools for raster handling, deegree for web-service interfaces, JTS for geometry operations (e.g. as part of the polygonisation) and custom code.

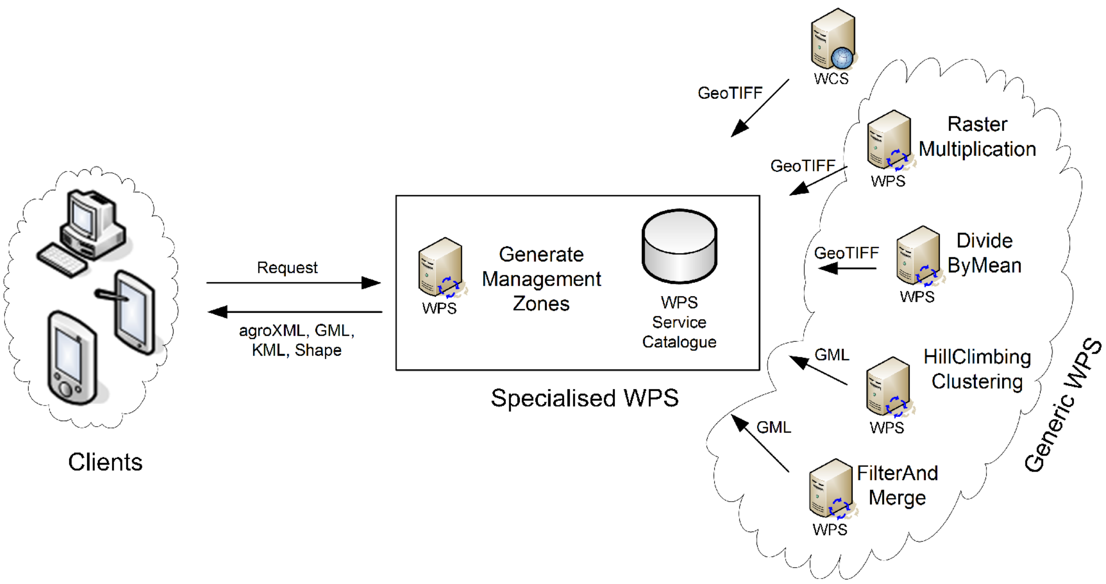
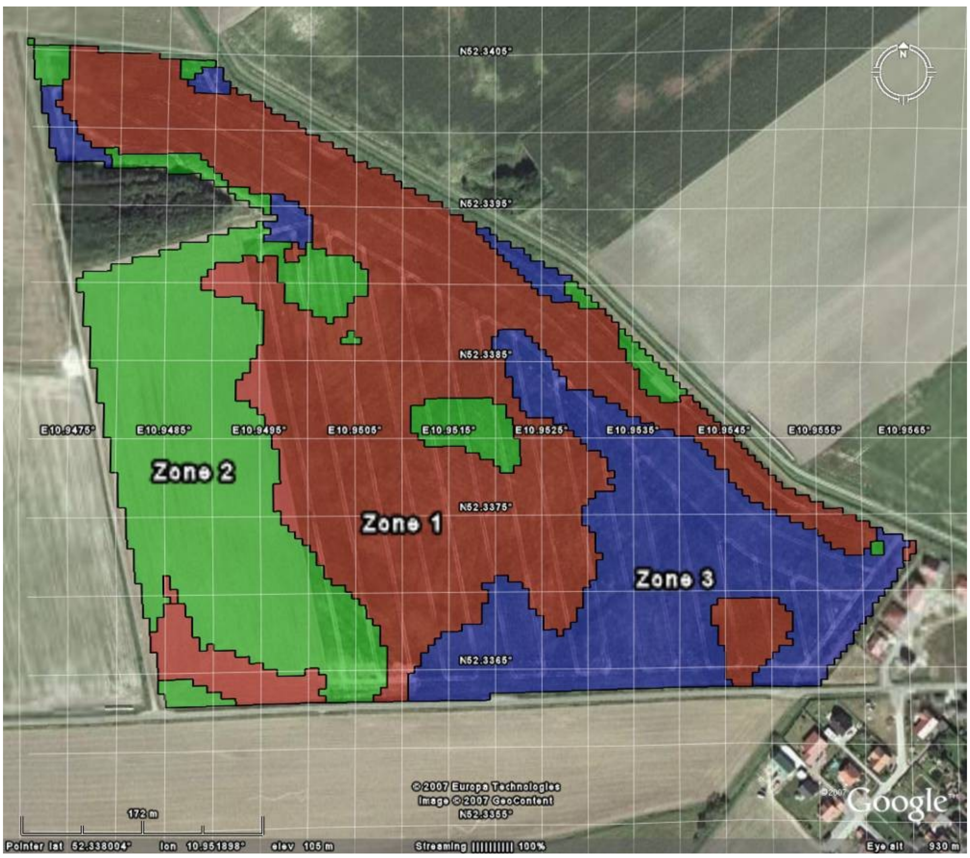


Figure 2. Implemented system architecture comprised of a specialised WPS with a local service catalogue aggregating generic WPS services

5. Results

We tested the procedure on basis of the field “Kamp” located in Lower-Saxony, Germany. We used yield maps from 5 years to delineate MZ. Fig 3 shows a cluster map with 3 zones, generated from yield maps from the years 2001, 2003, 2004, 2005 and 2006 with a minimum zone size of 100m² and displayed in Google Earth.



Background imagery: Google Earth™ mapping service

Figure 3. Result of zoning for field "Kamp" visualised using Google Earth

Table 1 shows the mean yield for each MZ in each observed year (calculated as part of the clustering), the field mean yield and the percentage of area for each MZ.

			mean yield (kg/ha) per year [crop]				
MZ	area (%)	cluster variance	2001 [WW]	2003 [WW]	2004 [WW]	2005 [WR]	2006 [WW]
1	48.27	191.26	5476	4940	7647	2778	6363
2	26.92	137.96	4504	5364	7647	2310	6302
3	24.61	140.58	5067	3576	6756	2596	5746
		mean	5118	4705	7424	2596	6178

Table 1. Mean yields within management zones in input years.

As an example of how the MZ and the statistics may be used to support a farmer's decisions we now show how the required variable amount of nitrogen fertilisation can be estimated. To produce 100kg of wheat grains requires an estimated 2.1kg of nitrogen fertiliser is needed. Table 2 therefore shows the calculated relative nitrogen amount used by the plants of the 3 MU in the 4 observed years where winter wheat was cultivated. The mean value is similar between the MZ 1 and 2, but for MZ 3 around 15% less nitrogen was required in the mean year. The farmer could therefore fertilise around 15% less in the MZ 3 without affecting the expected yield, potentially increasing profits as well as bringing environmental benefits.

	relative nitrogen uptake (MZ 1 = 100.0)				
MZ	2001	2003	2004	2006	mean
1	100.0	100.0	100.0	100.0	100.0
2	82.2	108.6	100.0	99.0	97.5
3	92.5	66.7	88.3	90.3	84.5

Table 2. Calculated relative nitrogen uptake for winter wheat in each zone and year.

6. Discussion

The work presented in this paper is designed to make geoprocessing accessible to an audience of non-specialists using lightweight client software. The use of an aggregated workflow allows flexibility and easy extensibility without requiring changes on the client side, whilst the individual services may also be incorporated in further workflows. Work is still required in ensuring resilience when services do not respond as expected and in ensuring that where parameters are not supplied by the user then sensible defaults can be supplied, as well as in extending the workflow to work with datasets in other formats and those not themselves supplied from web-services.

Further work is also required on the use of this technique with other data sets, such as DTMs and remote sensing data and how the statistics produced could be used to further assist farmers in decision-making beyond the simple example presented here. The visualisation of the degree of influence of each input dataset also requires consideration, as it is possible that e.g. the yield from a year with unusual weather patterns could significantly distort the result.

Acknowledgements

The work presented in this paper was funded by the German Federal Ministry of Education and Research (BMBF) as part of the *preagro* project under grant number 0330663.

References

Reichardt M, 2007, Was halten die Landwirte vom präzisen Landbau? Landtechnik, 62(1):28-29.

Rubin J, 1967, Optimal Classification into Groups: An Approach for Solving the Taxonomy Problem. Journal of Theoretical Biology 15(1):103-144

Steinberger G, Rothmund M and Auernhammer H, 2006, Agricultural Process Data Service (APDS). In: Agricultural Engineering for a Better World: Proceedings of XVI CIGR World Congress. VDI Verlag GmbH Düsseldorf.

Steinberger G, Rothmund M, Martini D, Spietz C, Mallon D, Nash E, 2007, Integration von agroXML in eine landwirtschaftliche Geodateninfrastruktur. Landtechnik, 62(2):2-3.

Whelan B and McBratney A (2003), Definition and interpretation of potential management zones in Australia, In: Solutions for a better environment: Proceedings of the 11th Australian Agronomy Conference, 2-6 Feb. 2003, Geelong, Victoria. Australian Society of Agronomy.