Accuracy improvement program for VMap1 to Multinational Geospatial Co-production Program (MGCP) using artificial neural networks

Luis Filipe Nunes

Portuguese Army Geographic Institute
Av[a]. Dr. Alfredo Bensaúde
Tel.: + 351 218505313
lnunes@igeoe.pt

Abstract

Deeply committed and involved in international geospatial production military groups, like the Multinational Geospatial Co-production Program (MGCP), the Army Geographic Institute is assessing the spatial accuracy of Vector Smart Level 1 (VMap 1), so that product can serve as the base for the production of the Portuguese national territory, according to the standards approved by all MGCP nations. In order to evaluate the positional, attribute and temporal accuracy of the data, a process was developed to check the suitability in relation to a densification from the base information required by 1: 100k scale MGCP project cells. Simultaneous tests were made in order to determine the uncertainty of positional and attribute elements in the main features of VMap 1, to evaluate the applicability of the data in both products. An expert system using Artificial Neural Networks to improve the accuracy of data from VMap1 database and allowing its integration with MGCP database, in a common environment, is proposed and analyzed.

Keywords: MGCP, VMAP1, positional accuracy, attribute accuracy, artificial neural networks

1 Introduction

The Army Geographic Institute (IGeoE) is the cartographic institution responsible for the Portuguese base map, which is the most used information in GIS for civilian and military purposes in Portugal. Now positioned within the Logistics Command of the Army, IGeoE routinely produces several cartographic products in a variety of analogue and digital formats:

- Topographic Maps at 1:10K, 1:25K, 1:50K and 1:250K scales.
- NATO Maps at 1:50K and 1:250K scales.
- Orthophotomaps at 1:5K and 1:10K scales.
- Satellite Imagery at 1:50K scale.
- Itinerary Maps (Civilian) at 1:250K and 1:500K scales.
- Digital Terrain Models (8 meters grid resolution, DTED1 and DTED2).

IGeoE is certified by ISO 9001 and ISO 14001 standards, and as producer of Digital Geographic Information (DGI), besides the national territory responsibility, one of the endeavours still standing to fulfil the requirements derived from international and North Atlantic Treaty Organization commitments, expressed in co-production groups and projects like the Vector Smart Map Co-production Working Group (VacWG), Digital Geographic Working Group (DGIWG) and more recently the Multinational Geospatial Co-production Program (MGCP).

2 VMap1 and MGCP programs

Deeply committed and involved in international geospatial production military groups, the Army Geographic Institute is assessing the spatial accuracy of Vector Smart Level 1 (VMap 1), in relation to the most up-to-date data from MGCP so that database can serve as the base for the production of Portuguese national territory and other international areas which the production responsibility lays on IGeoE, according to the standards approved by all MGCP nations. Another aspect that has to be taken into account is the concern of data integration among both MGCP data and VMap1 in a common environment, because although the information for the first project is acquired at a major scale and consequent higher precision, the number of cells planed to be produced by all countries doesn't cover the entire Globe, meaning they have to be combined in some areas with the database information at lower scale that covers the entire area of interest (the data from VMap1).

The VMap1 information is characterized by its homogeneity in the acquisition process, and the Portuguese national territory was produced from photogrammetric data acquired by stereo restitution at 1: 25k scale, subject to generalization algorithms, and is available at a 1: 250k scale. However, according to the data source used in each part of the World, classes were established to inform the user about the expected accuracy for that specific region.

The VMAP program is designed to provide vector-based geospatial data at medium (level 1) and high (level 2) resolution. Data is separated into 10 thematic layers consistent throughout the VMAP program. Each layer contains thematically consistent data. A reference library is provided with general information to orient the user. All data is topologically structured. Each coverage contains a set of files that describe the features in that thematic layer. Data volume is tiled at nominal one degree (level 1) and 15 minutes (level 2) expanding in size towards the poles. Cross-tile topology is maintained.

The standard applied to measure absolute horizontal accuracy was derived from cartographic source or generalized from centerline data. Actual map scale and class source of error are contained in the Data Quality coverage. Accuracy is expressed as a circular error at 90% probability.

Table 1 Absolute horizontal accuracy standards used in VMap Level 1 database.

SOURCE CLASS	CIRCULAR ERROR (METERS) VMAP LEVEL 1 1:250 000
A	125
B	250
C	500
D	>500

MGCP requires updated information at an approximate scale of 1: 100k, with a different circular error at 90% probability accuracy (25 meters), with more attributes and feature catalogue levels (MGCP, 2005).

3 Positional accuracy improvement using artificial neural networks

From the conceptualisation phase of this case, the principal task wanted from the expert system to provide, is the determination of coordinates from a knowledge base relative to a set of points (VMap1) that can guarantee the transformation into a more accurate and precise set (MGCP). From an expert systems point of view, this means that the subjacent problem refers to the Knowledge Acquisition area, of the Private type and the learning system form is Induced.

Specifically, the Induced learning seeks knowledge extraction from the chosen samples of points matching in both databases (VMap1 and MGCP), and from that synthesis find a generalization procedure that can be applied to all the surrounding space (the other points in VMap1). Due to the fact that the output values are known (matching MGCP points), then it is possible to estimate the learning error, which means that can be used a Supervised learning process. Using adjustments to the induced knowledge it is tried to reduce the differences between the expected and the obtained results from the learning process.

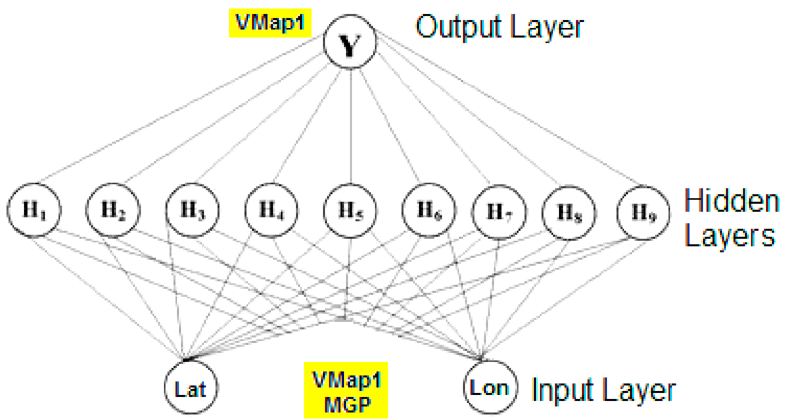

Figure 1 Layers implemented in the Artificial Neural Network (VMap1/MGCP data as Input and VMap1 as output).

Taking into account the above expressed, the most indicated model considered to solve the presented situation was the neural model, which consists of a simulation of biological nervous systems in programs or digital circuits. An Artificial Neural Network (ANN), sometimes referred to as simply a Neural Network, is a computer program designed to model the human brain and its ability to learn tasks (Haykin, 1994). An ANN differs from other forms of computer intelligence in that it is not rule- based, as in a conventional expert system. An ANN is trained to recognize and generalize the relationship between a set of inputs and outputs.Such artificial neural networks are capable of establishing complex relations between sets of data, in a precise form, without any prior information to the networks that relates them. The intelligent behaviour of an ANN derives from the interactions between the processing elements of the network, and their operation was proposed by McCullock and Pitts in 1943.

Neural networks are models that are designed to imitate the human brain through the use of mathematical models. The neural 'network' consists of a series of processing "units" which are collectively “connected” – like synapses in the human brain (Thurston, 2006). The network consists of an input, output and hidden layers. In this case study, the layers were organized according to (Fig 1):

- Input Layer: the coordinates (Latitudes and Longitudes) from the VMap1 themes are the input data;
- Intermediate or Hidden Layers: where the major part of processing is done, using weighted connections;
- Output Layer: the VMap1 coordinates, corrected to best adjust those of MGCP, and separated by Latitudes, Longitudes and themes, which demand different neural networks for each of the parameters to be calculated.

4 Cape Verde NATO exercise area case of study

The chosen study area was the first to be produced for MGCP project, and it was prepared on purpose for the NATO STEDFAST Jaguar exercise. The information of one of the islands of Cape Verde, S. Vicente, was available either on VMap1 and also produced for MGCP. Due to the fact that S. Vicente had sufficient different information, themes and objects, was selected to make a plain trial, and it was believed that could constitute a very representative test-bed.

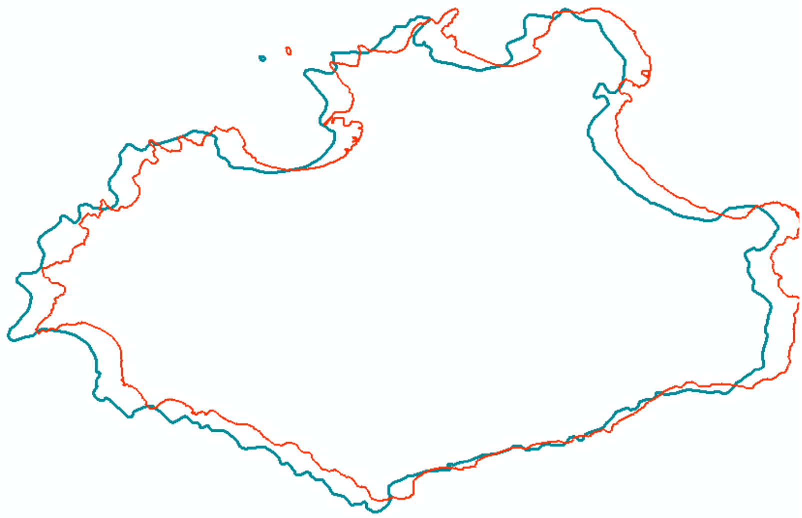

Figure 2 S.Vicente Island, superimposing VMap1 Boundary (in blue) and recently acquired MGCP Boundary (in red).

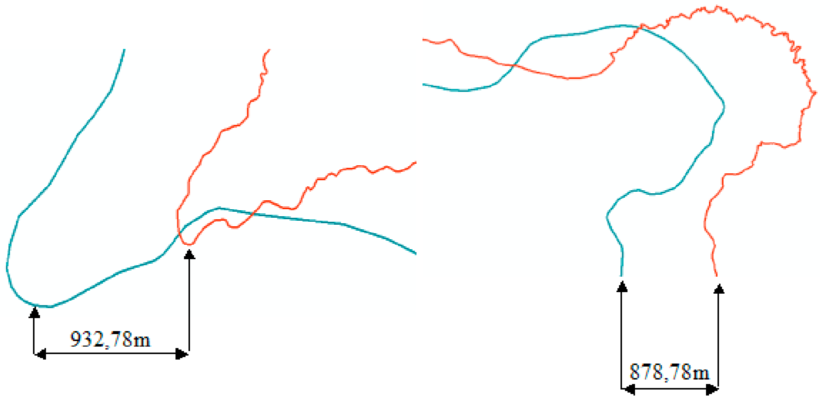

Figure 3 Differences obtained from S. Vicente Island boundaries of VMap1 (in blue) and MGCP (in red).

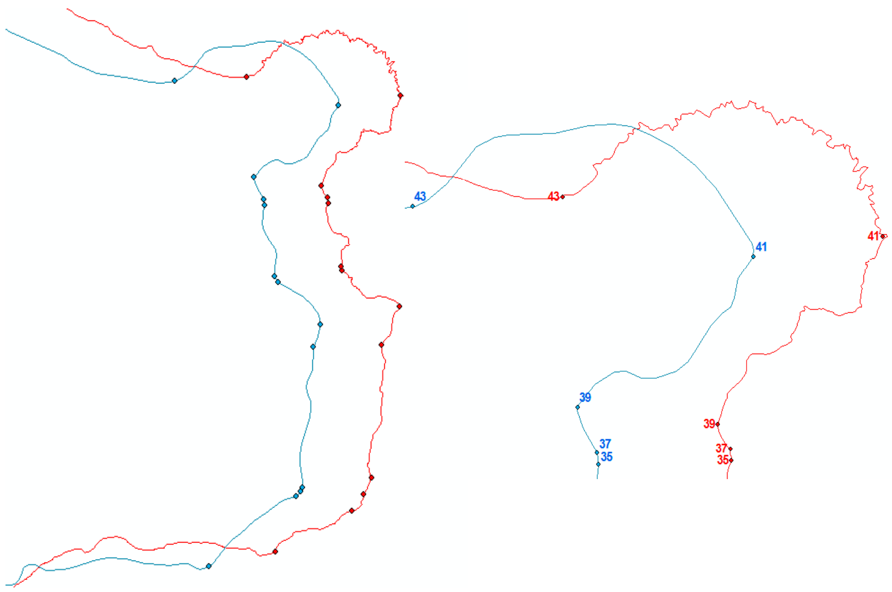

Figure 4 Matching control points acquired in VMap1 Boundary theme, and MGCP corresponding theme.

In a brief analysis, comparing VMap1 and MGCP data it was possible to detect differences in the positioning of different themes (Figure 3) that lead to the conclusion that the S. Vicente island data, was probably, derived from a class D source.

From the two databases (VMap1 and MGCP), 100 points were selected per each main theme, matching each sample point the same place on both products, so that they could serve as knowledge base to the expert system (Figure 4). To accomplish that, VMap1 data format

(Vector Product Format) was transformed to Shapefile, and from shape to simple coordinate text file. MGCP data was transformed from shapefile directly to coordinate text file. From those 100 points, 5% were used to Cross Validation, and another 5% were used to Test, which means that 90 points were used for Training the artificial neural network.

The coordinate system used for both databases was the WGS84 datum with geographic coordinates (Latitude and Longitude in decimal degrees), so the sample points were collected in their native system on both products.

The type of artificial neural network used to calculate the functions was the Multilayer perceptrons (MLPs). MLPs are layered feedforward networks, typically trained with static backpropagation, and in this case, was used the Momentum term. The backpropagation algorithm that minimizes the mean squared error has been used as the training algorithm. These networks have found their way into countless applications requiring static pattern classification, and their main advantage is that they are capable of achieving nonlinear knowledge, and can approximate any input/output map (Suju, 2006). As supporting software it was used the NeuroSolutions For Excel program.

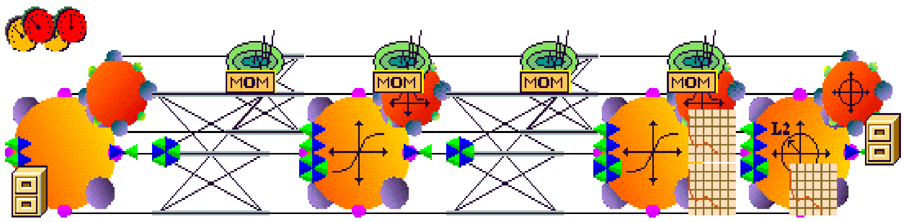

Figure 5 The type of artificial neural network applied was the Multilayer Perceptrons (MLPs) using Backpropagation, with Momentum term.

5 Results exploitation, analysis and application

From a detailed check on the data from selected themes of VMap1 (Boundaries, Road network and Hydrography) and its comparison with the correspondent theme of MGCP data, it was observed that the deviations were not constant and consistent among themes. Even in the same theme there was not a constant pattern of the deviations in the entire area contained by the theme. This conclusion headed to a situation where for each variable wanted to be calculated (Latitudes, Longitudes) a completely new ANN had to be created. Also, for each different theme, had to be chosen different sample matching points in both databases. Although it involved more operation work and computational labour, it was also important, because the results from each artificial neural network produced, had the guarantee of independency from the results obtained to other themes and coordinates (Latitudes and Longitudes).

Table 2 Average of Minimum and Maximum Mean Square Errors, obtained by the artificial neural network, for the Boundary's Latitudes.

All Runs	Training Minimum	Training Standard Deviation	Cross Validation Minimum	Cross Validation Standard Deviation
Average of Minimum MSEs	5.94032E-06	5.4548E-07	1.72623E-06	2.55074E-07
Average of Final MSEs	5.94032E-06	5.4548E-07	1.78762E-06	2.58975E-07

An example of the results obtained, can be checked on Table 2. For the Boundary's theme and specifically the Latitudes calculation, the Average of Minimum and Maximum Mean Square Errors is 5.94032E-6 of decimal degree.

On table 3 it can be seen that the best network results were obtained at the epoch 65000 and 14th run, and both the cross validation and the training Mean Square Errors were according to the expected from table 2.

Table 3 Results of training and cross validation for the best networks obtained for the Boundary's Latitudes.

Best Networks	Training	Cross Validation
Run #	14	14
Epoch #	65000	47108
Minimum MSE	5.90632E-06	1.54088E-06
Final MSE	5.90632E-06	1.70257E-06

From the Figure 6, can be seen the graphical relation between the desired output and the best neural network achieved output, for Boundary's Latitudes, and from the Table 4 can be derived the best performances of the trained and applied ANN for production.

The results obtained were close to each other, in precision and accuracy, on the other themes tested, either in Latitudes as in Longitudes on the ANNs trained. After the process of training the networks, all the points from each of the themes of VMap1 were introduced for production, in the ANNs trained, theme by theme, and separated from Latitudes and Longitudes. That concluded the calculation of the new position for each of the points in VMap1 database.

The application of the best neural networks achieved in Latitudes and Longitudes for the Boundary's theme can be seen in Figure 7. In blue is the original VMap1 Boundary, in red the MGCP Boundary theme, and 100 points of both themes were the knowledge base. The points represented on the figure are the result of the transformation done by the two trained artificial neural networks for VMap 1 Boundary (all points), in Latitudes and Longitudes. It can be concluded that there was an approximation of VMap1 data to the recently produced MGCP data. There can be noticed also some errors and gaps due to the difference in scale of both databases, but mainly it is related with the level of detail, and as ultimate instance can be compared to a generalization process, because one was produced at 1:250k and the other at 1:100k scale.

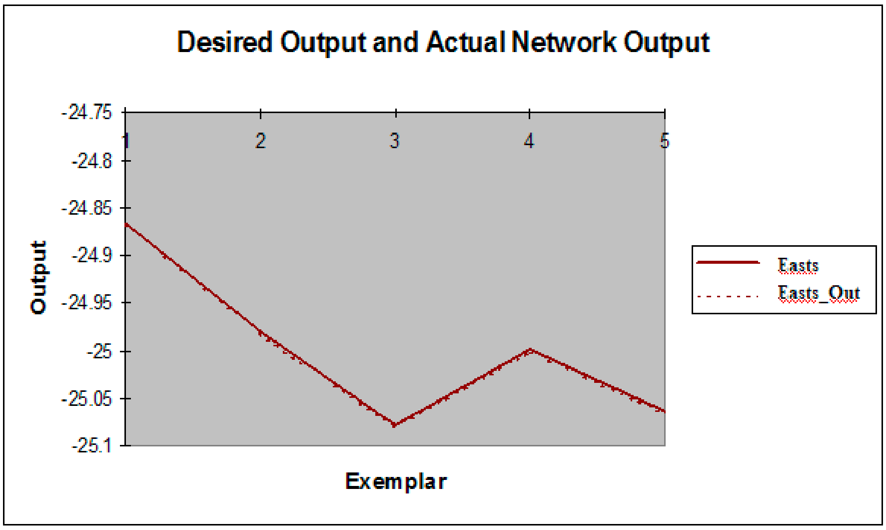

Figure 6 Graphical relation between the desired output and the best neural network achieved output, for Boundary's Latitudes.

Table 4 Performance of the best network obtained for the Boundary's Latitudes. The learning factor (r) was close to 1, which indicates a good knowledge base learning process.

Performance	Boundary Latitudes
MSE	2.18539E-06
NMSE	0.000189703
MAE	0.000325493
Min Abs Error	0.0000545866
Max Abs Error	0.0001238866
r	0.99999878

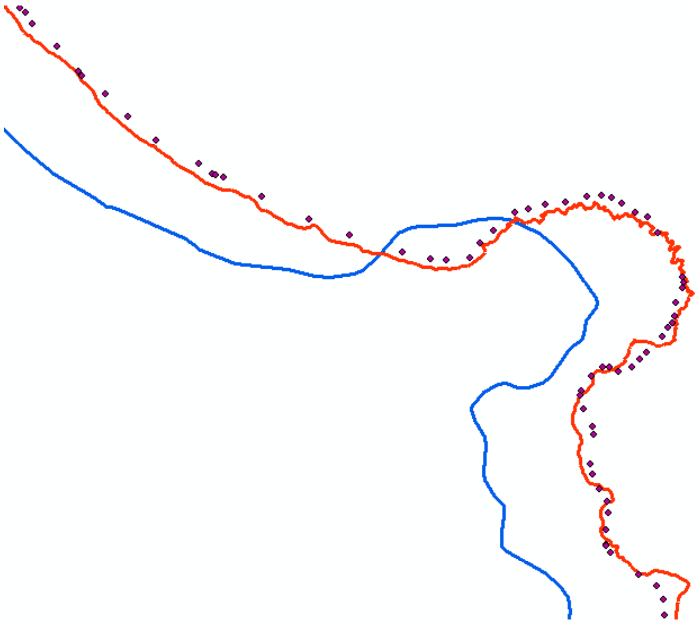

Figure 7 Application of the best neural networks achieved in Latitudes and Longitudes for the Boundary's theme. In blue the original VMap1 Boundary, in red the MGCP Boundary theme, and the points are the result of the transformation done by the neural network of VMap1 Boundary points.

6 Results exploitation, analysis and application

The use of expert systems, in the form of Artificial Neural Networks to improve positional accuracy, is not a new concept or even process, and it is widespread documented, although there aren't so many works on this area related to the geospatial accuracy community.

This investigation had the objective of finding functions induced from the entry information of VMap1 themes and the corresponding output MGCP information respectively, that have a closer approximation to the function that relates both, allowing other themes adjustment and their correspondent coordinates, making possible the improvement of geospatial overall accuracy of VMap1 database, and evaluating the accuracy of the obtained data.

From the lessons learned of this experience, it is possible to conclude that the method can be applied to VMap1 as an Accuracy Improvement Program, using Artificial Neural Networks, to relate that data to a more accurate and precise one, like MGCP.

The Accuracy Improvement Program using Artificial Neural Networks can be a solution, not only to match both databases information in the common areas covered by both VMap1 and MGCP, but also to cross-tiling themes between data derived from both products.

The Accuracy Improvement Program using Artificial Neural Networks permits also the utilization of part of VMap1 themes, improving their accuracy, so they can be inserted as a component of MGCP data, allowing less time consuming efforts in the acquisition process for this program.

References

Suju M. George, P. Ram Babu and P. KhannaGeospatial, Resource Portal, Neural network model for consequence analysis of developmental proposals. Available online at: http://www.gisdevelopment.net/application/miscellaneous/misc017.htm (last accessed 10 May 2006).

Thurston Jeff, GISCafe.com Portal, GIS & Artificial Neural Networks: Does Your GIS Think? Available online at: http://www.integralgis.com/pdf/Neural%20Networks.pdf (last accessed 10 May 2006).

Gordon German, GISCafe.com Portal, Neural Network Classifiers for GIS Data: Improved Search Strategies. Available online at: http://www.geovista.psu.edu/sites/geocomp99/Gc99/093/abs99-093.htm (last accessed 10 May 2006).

Haykin, Simon, 1994. *Neural Networks, A Comprehensive Foundation*. New York: Macmillan College Publishing Company, pp. 696-697.

MGCP, 2005. Image Benchmarking Process v2.3, Multi-national Geospatial Co-production Program, Technical Document.