RESULTS OF EXPERIMENTS ON AUTOMATED MATCHING OF NETWORKS AT DIFFERENT SCALES

Sébastien Mustière

COGIT Laboratory, Institut Géographique National, 2 av. Pasteur, 94160 Saint-Mandé – France,
sebastien.mustiere@ign.fr

KEY WORDS: Data Matching, Networks, Geometrical and topological matching, Inconsistency, Level of detail

ABSTRACT:

Many geographical databases of the same area are produced and maintained. In order to remove inconsistencies between those databases, and in order to facilitate the updating process, a close integration is required. This paper reports the results of experiments of data matching between the networks of two IGN databases at different scales (road, electric, hydrographical, railway and hiking routes networks). We illustrate the main results of these experiments through four different aspects. The first aspect is the identification of data that can not be matched because they only appear in one database. Some of these differences between the contents of the databases are clearly explained by the specifications: they reflect the difference between points of views. Some other differences are just discovered in the data: they reflect the different sources used to build databases and inconsistencies due to errors. The second aspect is the analysis of differences and inconsistencies between databases when corresponding objects are identified. In particular, we identified differences between attribute values, geometric descriptions, but also topological relationships between objects. A third studied aspect is the degree of automation of the matching process. We managed to automatically match from 90% to 100% of objects in networks, depending of the complexity of the networks. Finally, the fourth considered aspect concern the interactive checking of results. We identified ergonomic difficulty to visualize data and results of matching. We thus propose some solutions to overcome these difficulties.

1. INTRODUCTION

Many geographical databases of the same area exist. Even a single data producer usually maintains and updates several geographical databases. In order to facilitate this maintenance a close database integration of is required [Sheth and Larson 1990, Parent and Spaccapietra 2000]. Integration is useful for data producers, but also for data users who require up-to-date and rich data [Kilpeläinen 2000, Hampe and Sester 2002, Sheeren et al. 2004].

Integration first helps to propagate updates from one database to others. For data producers, this is important to decrease the cost of updating data: updates may be integrated once and propagated, at least semi-automatically, to different representations of a same geographic phenomenon. For data users, integration could help to easily integrate updates coming from data producers, without loosing their own enrichment of data.

Integration is also useful to perform some quality analysis and identify inconsistencies. For data managers, evaluation of quality is important to control and increase the quality of their data. For data users, evaluation of quality is important to assess the fitness for use of data.

Finally, integration is useful to increase the potentiality of applications using these databases. For data users, the use of data coming from different databases would be much easier if these data were well integrated. For data producers, the development of new databases including data coming from existing databases would also be facilitated.

Database integration requires some works at the schema level and some works at the data level, especially for geographic data. First, schema integration is necessary to identify corresponding classes in the schemas [Devogele et al 1998, Rodriguez et al. 1999, Uitermark, 2001, Stonykova 2003]. Then data matching is necessary to actually identify corresponding objects in the data [Devogele 1997, Walter and Fritsh 1999, Mantel and Lipeck 2004, Haunert 2005, Zhang et al. 2005]. Data matching in geographic databases require tools comparing the geometrical and semantic information of data, because there usually does not exist any universal identifier to guide the matching.

With the intention of evaluating how data from IGN-France databases can be integrated, and how far this integration can be automated, we conducted several experiments of automated data matching with two databases at different scales. We especially wanted to answer to the following questions: which data can be matched? What are the typical inconsistencies detected? How far do we succeed to automate the process? And how can we interactively check and alter the results? This paper reports on the results of these experiments concerning the different networks of the databases: road, electric, hydrographical, railway and hiking routes networks.

The next part presents the data used in the experiments. Then, part 3 introduces the matching process used on these experiments. Part 4 presents the results, first in terms of matching possibilities (which objects do we encounter in both databases), second in term of inconsistencies (where do we encounter abnormal differences), third in term of the results of automation (are the results similar to an interactive matching), and finally in term of ergonomic aspects.

2. USED DATA

Our experiments concern two different databases made from different sources and with different purposes.

The less detailed database used is named BDCarto. Its precision is of several decametres and is varying according the considered themes (see figure 1 and 2, on right). It is used at IGN, for example, to make maps at 1:100,000 scale or 1:250,000 scale. It is also typically used to make geographic studies, concerning whole France or a French department. This database is dedicated to some network studies, thus attention has been paid to ensure a topologically correctly connected network. BDCarto is also used as a reference database for some users, who add their own information on it, especially on the road network. BDCarto mainly originates from the digitalisation of 1:50,000 scale maps.

The more detailed database, BDTopo, has a metric precision (see figure 1 and 2, on left). It is used at IGN, for example, to make some maps at 1:25,000 scale. It is typically used to make some local studies, like environmental impact assessments. It is used has reference data to precisely locate extra information. Attention is especially paid on the precise topographic description of the world. BDTopo mainly originates from stereoscopic aerial images.

Figure 1. Road network of BDTopo (left) and BDCarto (right).

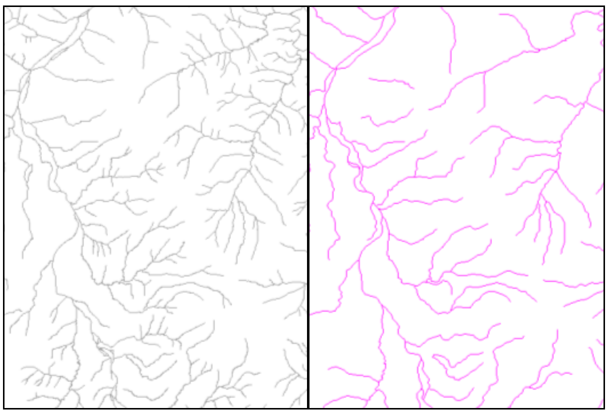

Figure2. Hydrographic network of BDTopo (left) and BDCarto (right).

If we roughly compare the networks encountered in these databases, we can say first that BDCarto is less detailed than BDTopo, and thus has fewer details regarding the geometric shape of arcs, and the organisation of connections.

Another important characteristic is that BDCarto concentrates on the topological organisation of the networks, and BDTopo concentrate on the precise topographic description of it. A typical example of it is in the electric network around the transforming stations (see figure 3). BDTopo describes the transforming station (the gray surface) and how electric lines enter on it, while BDCarto directly connect lines arriving in the same transforming station. This difference is of course due to the difference of levels of details, but not only: it reflects that BDTopo describes what we see on the ground, and that BDCarto describes how things are connected.

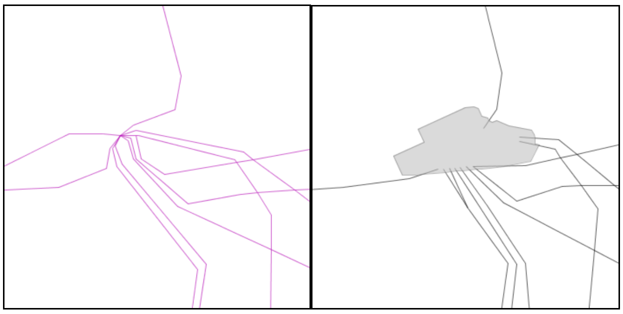

Figure 3. Electric network of BDTopo (left) and BDCarto (right), around a transforming station.

Another typical example is that BDCarto sometimes contains underground part of the network (underground pipes, underground electric lines), while BDTopo does not contains all of them. This is for example the case of the missing lines in BDTopo on figure 3.

3. AUTOMATIC MATCHING PROCESS

Matching of networks already received some attention in the literature [Devogele 1997, Walter and Fritsh 1999, Zhang et al. 2005]. Our experiments are based on a two-steps process: we first transform one or both networks to make them closer, we then match networks with an algorithm inspired from the approach of T. Devogele [1997].

3.1 Pre-treatment

If the networks have very different structures, it is hard to directly compare them. We thus perform a pre-treatment of them, in order to give them a similar structure and to prepare the matching.

For both networks, our algorithm requires arcs and nodes, and topological relationships between them. But some of the networks only contain arcs and no explicit nodes. In this case, a first required pre-treatment is the creation of the graph structure (figure 4). Different strategies are possible. The simplest one is to just create nodes at extremities of arcs and to compute topological relationships. In some cases, it may also be useful to fusion distinct but very close nodes coming from different arcs, in order to overcome topological problems in the data. Sometimes, it is also useful to build a planar graph from the arcs.

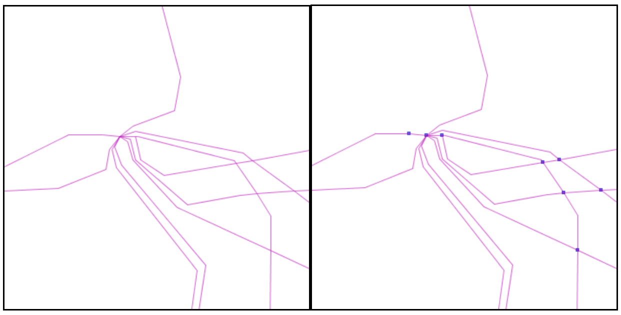

Figure 4. Nodes are added at the extremities of electric lines of BDCarto.

Another typical required pre-treatment is the transformation of connections. For example, modelling of connections of electric networks are very different (see figure 3). We thus first transform the connections of BDTopo to make them similar to the connections of BDCarto. In this case, a node is created in the centre of each transforming station, and all lines entering in the station are connected to this node (figure 5). Matching will after be done with this transformed network.

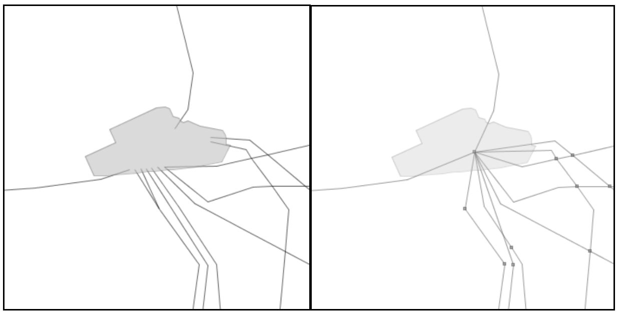

Figure 4. Lines of BDTopo arriving in the transforming station are connected before matching.

In some cases, there also may exist important differences in the exact location of the extremities of the network. For example, this appends for hydrographical and railway networks in our databases. In these cases we split arcs and add extra nodes to networks, by projecting extremities of one network on the other one (figure 6).

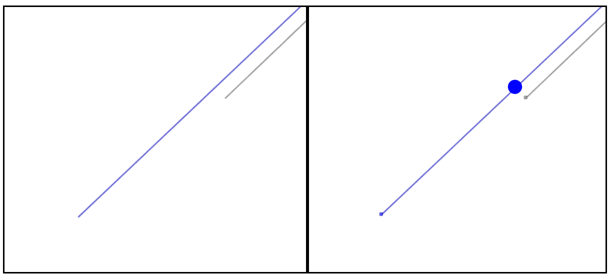

Figure 6. Networks may not stop at the same place: extra nodes are added to facilitate matching.

Finally, a pre-treatment is also necessary to enrich databases and explicit some useful implicit phenomena in the data. For example, roundabouts are not explicitly in the road network of BDTopo, but they appear as a set of arcs. We thus detect them through the analysis of their shape: they are round faces of the graph (figure 7). These roundabouts are very useful to guide the matching because we know that one node of BDCarto may correspond to a fully detailed roundabout in BDTopo, but not to only a part of it. This information will be used during the matching process.

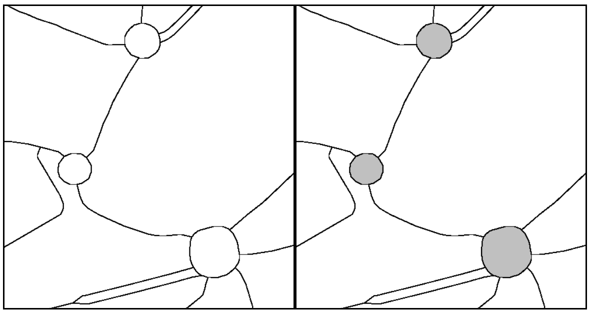

Figure 7. Roundabouts are detected in BDTopo according to their shape.

3.2 Matching

Once networks have been prepared, they can be matched. Describing precisely the algorithm used is out of the scope of this paper. We briefly describe here its main steps.

The first step is a *pre-matching of nodes*. For each node of the less detailed database (BDCarto), we look for close nodes in BDTopo candidate for matching. The distance determining how far we look for candidates is determined according to semantic values of nodes of BDCarto (figure 8). For example, the distance is quite big if the node is known representing a complex interchange. But the distance is quite small if we know that the node is representing a small junction.

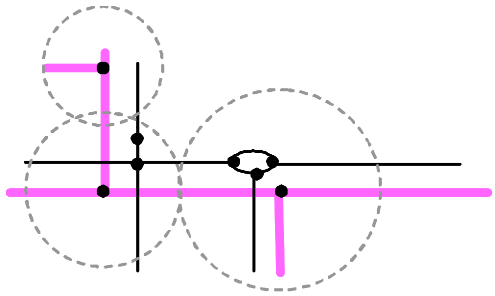

Figure 8. Pre-matching of nodes of BDTopo (narrow lines) and BDCarto (wide lines).

The second step is a pre-matching of arcs. For each arc of BDTopo, we look for close arcs in BDCarto, candidate for matching (figure 9). This pre-matching is based on geometric criteria like the Hausdorf distance between lines. It is also based on semantic criteria. For example we prevent pre-matching highways with secondary roads.

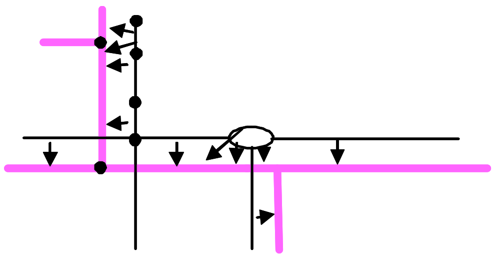

Figure 9. Pre-matching of nodes of BDTopo (narrow lines) and BDCarto (wide lines).

The third and most complex step is the actual matching of nodes. It is based on the combined analysis of the results of the two pre-matching steps.

We first look for 1-1 matching, i.e. one node of BDCarto corresponding to one node of BDTopo. For each pair of nodes candidates for matching, we look if their respective connected arcs are also candidate for matching. When nodes and arcs matching are fully consistent, they are matched. For example, in figure 10 the node of BDCarto Nc is matched to the node Nt of BDTopo because: 1/ these two nodes are candidate for matching, and 2/ all the 3 arcs of BDCarto connected to Nc are candidate for matching with one arc of BDTopo connected to Nt. This is the only case like that in the example.

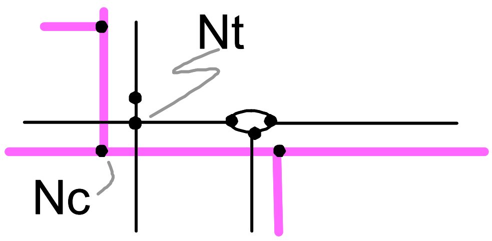

Figure 10. A 1-1 node matching.

We then look for 1-n matching, i.e. one node of BDCarto corresponding to several nodes and arcs of BDTopo. Without detailing this step, let's say that its principle is to group nodes and lines of BDTopo. These groups are then considered as hyper-nodes in the network and managed like the nodes: we look for groups where pre-matching of nodes and arcs are consistent (figure 11).

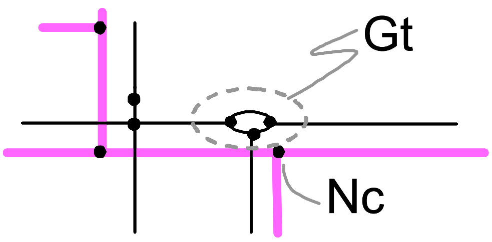

Figure 11. A 1-n node matching.

Finally, the last step concern line matching. Once nodes of BDCarto have been matched, we consider arcs of BDCarto one by one, and each arc is matched to a set of arcs of BDTopo. Roughly speaking, this arc of BDCarto is matched to the shortest path in BDTopo linking the nodes matched to the extremities of the arc of BDCarto (figure 12). Some semantic criteria are also used to calculate this shortest path. Typically, when matching uncovered road of BDCarto, we favour path going through hiking trail or secondary roads in BDTopo rather than paths going through primary roads.

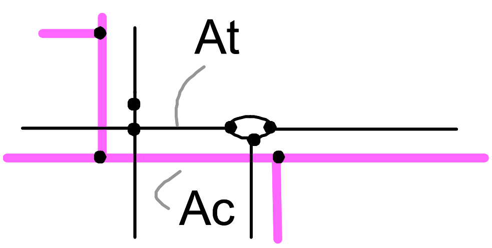

Figure 12. Two arcs matched

4. MAIN RESULTS

4.1 Matching possibilities

The first results of the experiments concern the data, without taking into account the problem of automation. We wondered which information is present in the less detailed DB BDCarto, but not in the more detailed DB BDTopo. These information are very problematic for integrating these databases. They can not be matched, and they will require a special updating process.

Of course, these results are very specific to the studied database. But they are quite characteristic of typical differences that can be found between databases with different point of view.

First we identified some *types of objects* in BDCarto with no homologous object in BDTopo. Most of these missing objects are explained by the difference between the points of view, and in particular by the fact that BDTopo especially describes what is seen on the ground and BDCarto describes how things are connected. For example we miss in BDTopo some underground objects like electric lines or water pipes, and some virtual object with no ground prints like boat connections in the sea.

Second, we identified some *objects* in BDCarto with no homologous object in BDTopo, without being able to classify these missing objects according to some particular attribute value. This can be explained by the fact that the databases have different sources. It mainly concerns secondary objects: important objects of BDCarto always have a corresponding object in BDTopo, except in very rare cases. For example, the selection of some temporary rivers or secondary roads is different, and their selection of BDCarto is not strictly a sub-selection of BDTopo.

Third, we identified some *attributes of objects* in BDCarto with no homologous information in BDTopo. Once again, this mainly originate from the different point of views of the databases. For example, BDCarto describes more in detail tourist information and traffic directions.

Some of these differences are explained clearly by the specifications of the databases; they reflect the difference of point of views. Some other differences are just discovered in the data; they reflect the different sources and the different inconsistencies due to errors.

4.2 Differences and inconsistencies

Another result of the matching experiments is to detect and characterise differences and inconsistencies between the databases.

The main difference concerns of course the level of detail. This difference can be seen at different levels: the number of represented objects is different and the connections can be differently described (see figure 13), the geometric detail of shapes and the planimatric accuracy are different, and the attribute value can be different. These differences are not identical for all networks: while road networks are very different, relatively straight networks like electric lines and railways are much closer.

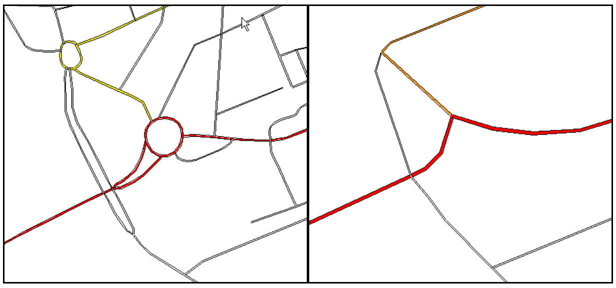

Figure 13. Different levels of detail between road networks of BDTopo (left) and BDCarto (right).

When the differences can be explained by the specifications, they are normal. But when they can not be explained by the specifications, we consider them as inconsistencies [Sheeren et al 2004], and one of the important interest of matching data is to detect and correct them.

[Sheeren et al 2004, Sheeren 2005] propose a methodology to detect these errors, based on the analysis of data specifications and on the use of automatic machine learning. [Egenhofer et al 1994, Paiva 1998] also propose methodologies to detect inconsistencies, concentrating on topological inconsistencies. During our experiments, we identified inconsistencies through the analysis of our automatic matching process, either when it has been efficient or not. Some of the inconsistencies have been identified once the matching has been done, for example by automatically comparing attributes values of surely well matched data. Other consistencies have been identified because they resulted in abnormal unmatched data.

The first inconsistencies we identified during our experiments are inconsistencies between attribute values. Some of them are clearly due to errors: for example two matched roads may have different numbers, two matched rivers may have different names, or two matched electric lines may have very different powers (for example “63kV” matched with “250kV”). But some other differences are not so clearly errors of the data, they come from some confusions or imprecisions in the databases specifications and in the ground itself. For example, we identified an important number of electric lines with the power “63kV” matched with lines with the power “90kV”. Are these differences inconsistencies or just different (but similarly correct) interpretations of the world and the database specifications? We consider that the answer to this question is one of the most important issue to be solved, in order to better automate matching and inconsistency correction.

Then, we identified inconsistencies in term of the presence or not of some objects. For example, one hiking route of BDCarto may have no corresponding road or path in BDTopo, which emphasize certainly a lack of data in BDTopo (figure 14). As another example, one electric line of BDCarto may have no corresponding object in BDTopo, even if it is not underground, which certainly emphasize in this case a difference of uptodateness (figure 15).

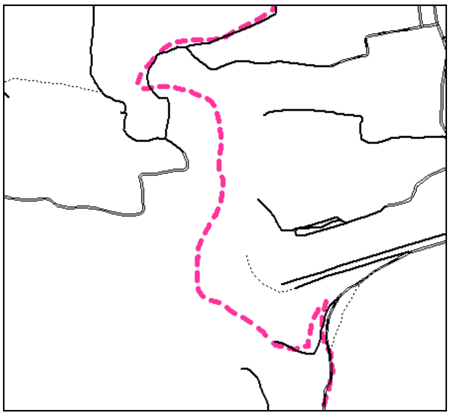

Figure 14. Inconsistency: part of a hiking route of BDCarto (dashed line) has no corresponding object in BDTopo (gray)

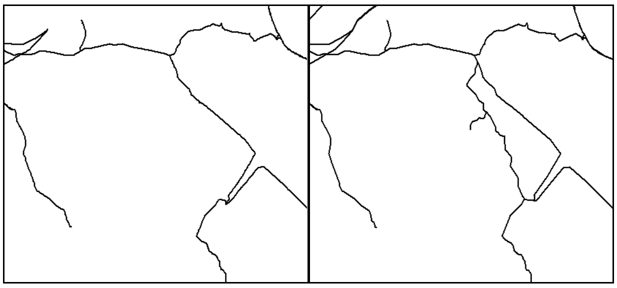

Figure 15. Inconsistency: an electric line of BDCarto (right) has no corresponding object in BDTopo (left)

Finally we identified topological inconsistencies between data. Sometimes, this problem is internal to one database: for example two very close road junctions are not connected in one database and this is typically an integrity error of this database. But in some cases, we identify important differences in the topological organisation of networks, without being able to identify the correct one (figure 16 and 17). These inconsistencies are usually due to differences in uptodateness, but it is most of the time difficult to ensure it.

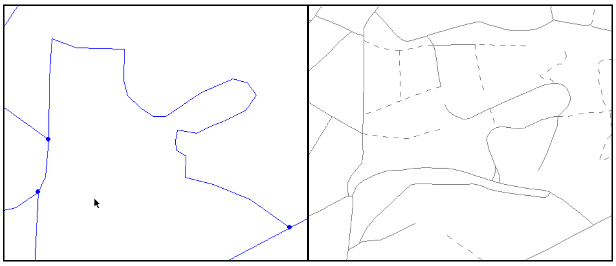

Figure 16. Inconsistency (center of image): the roads are connected in BDCarto (left) and not in BDTopo (right): Missing data in BDTopo or wrong generalisation in BDCarto? Uptodateness difference or error during capture?

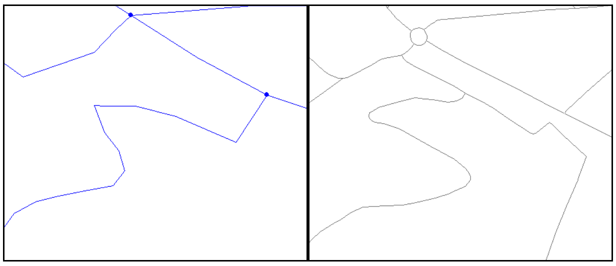

Figure 17. Inconsistency: the roads do not connect the same way in BDCarto (left) and BDTopo (right): Missing data in both BDCarto and BDTopo?

4.3 Results of automation

An important purpose of the experiments was to know how far we can go on the automation of the matching process.

As a global evaluation, the matching process has been efficient for 90% to 100% of the objects, depending on the network studied. Let us precise that "being efficient" does not mean "successfully finding an object to match", but "matching, or not matching, exactly as we would do interactively". In order to illustrate the results, let us first emphasize some special cases where the process is efficient.

First, it is important to notice that the process does not necessary match together the closest objects. Figure 18 illustrates this on the road network. BDCarto (upper left) and BDTopo (upper right) are displayed, and the result of automatic matching is represented below: segments connect matched arcs, and the node of BDTopo corresponding to the node of BDCarto is circled. We can see in this example that the node of BDCarto is close to two BDTopo nodes, and is (efficiently) matched to the more far one. This has been possible thanks to the comparison of topological connections between nodes and lines.

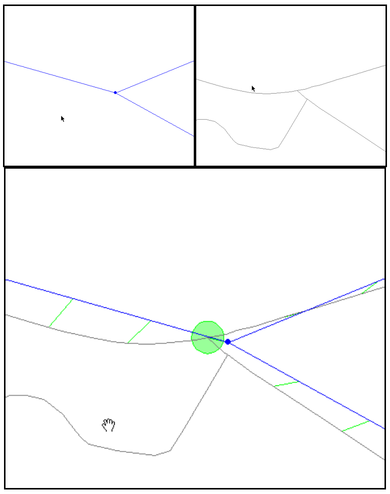

Figure 18. BDCarto (upper left), BDTopo (upper right) and result of matching (down) for a road connection.

Figure 19 shows a similar but more complex case for railways. The node of BDCarto is efficiently matched to the node of BDTopo on the far right of the image, which is not the closest one. This has been possible thanks to the comparison of node and line topological connections, but also to the comparison of semantic values.

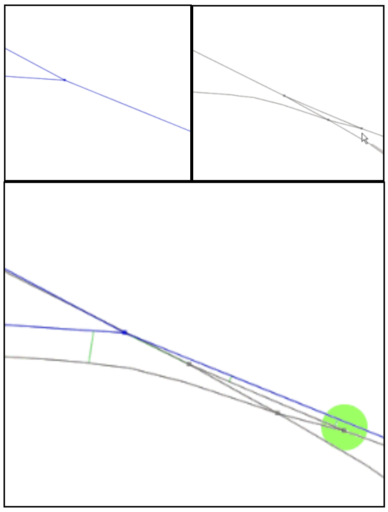

Figure 19. BDCarto (upper left), BDTopo (upper right) and result of matching for a railway connection.

Figure 20 illustrate similar considerations for arcs: the BDCarto electric line on the right is efficiently not matched to the closest one. This has been possible thanks to the comparison of arcs extremities (not shown in the image).

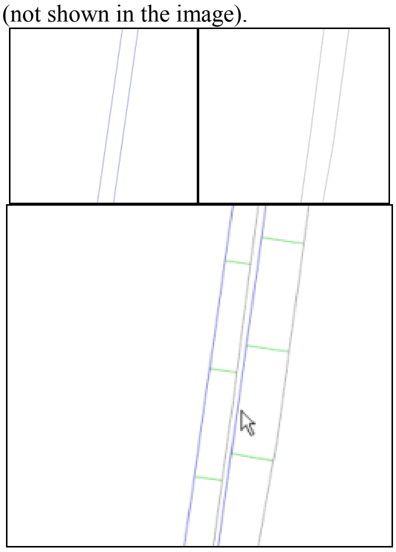

Figure 20. BDCarto (upper left), BDTopo (upper right) and result of matching for electric lines.

Second, when one node of BDCarto corresponds to a detailed but relatively simple connection of BDTopo, the result is generally efficient. Figure 21 and 22 illustrates this on road matching. In Figure 21, one node of BDCarto is matched to two separated nodes of BDCarto. In figure 22, one node of BDCarto is matched to a full roundabout in BDTopo.

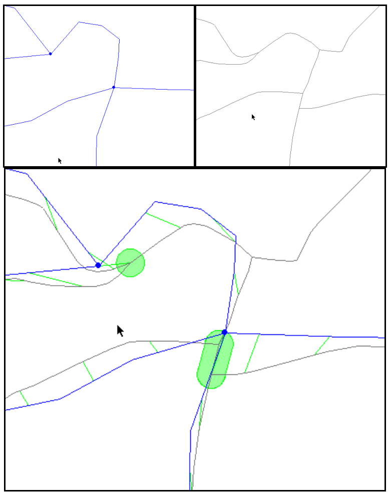

Figure 21. BDCarto (upper left), BDTopo (upper right) and result of matching for a road connection.

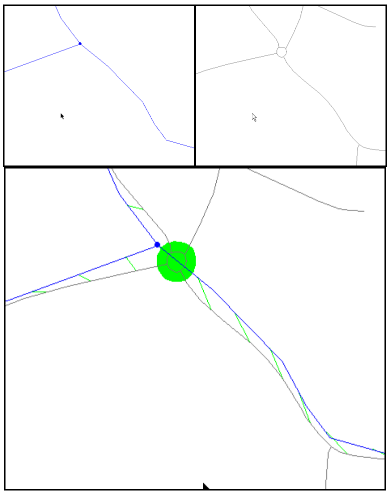

Figure 22. BDCarto (upper left), BDTopo (upper right) and result of matching for a road connection (roundabout).

1-n matching also occurs for lines: one arc of BDCarto is usually matched to several consecutive arcs of BDTopo. Figure 23 illustrates this on road networks (the orange means that the automatic process is uncertain of the result and requires interactive checking).

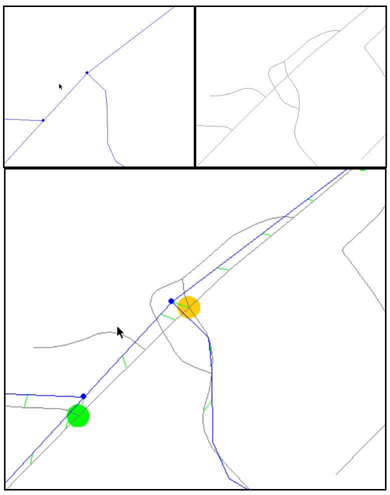

Figure 23. BDCarto (upper left), BDTopo (upper right) and result of matching for a road connection (roundabout).

Unfortunately the result is not always so efficient. First, the process may be mistaken in some cases. For example, Figure 24 illustrates a typical error: the BDCarto node is wrongly matched to a full line of BDTopo. This kind of problem is especially current when the connections are very complex, like around highways.

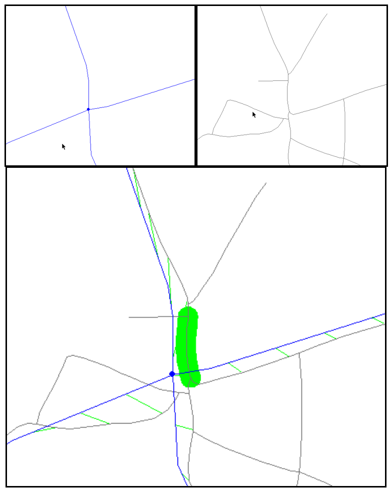

Figure 24. BDCarto (upper left), BDTopo (upper right) and wrong result of matching for a road connection.

4.4 Ergonomic considerations

Even if these experiments demonstrate that the matching process can be highly automated, some researches must be carried on to improve the level of automation. But we believe that a fully automated process is almost impossible for the most complex cases.

Thus, in a production context, the issue of interactive checking and correction is an important one. In order to analyse the result, one must display the two different databases as well as the links between them. Figures in this paper do illustrate the ergonomic difficulty of it: it is sometimes hard to understand the geometric relative configuration of objects among all this displayed information. Figure 25 illustrates this particularly on a complex junction.

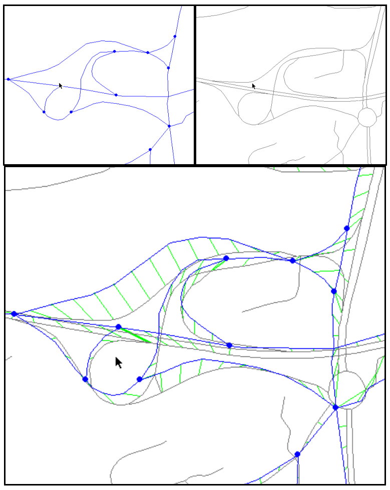

Figure 25. Arc matching: is it right or wrong?

The development of an adequate interactive environment raises some important human-computer interaction issues. But two directions may be explored to guide it.

The first direction takes advantage of the duality between arcs and nodes. Usually, when line matching is efficient, node matching is efficient, and conversely. This is particularly true with our process as line matching fully rely on a previous node matching. Our experiments showed that we where much more able to check results of node matching rather than results of line matching. For example, we considered figure 26 much easier to interpret than figure 25.

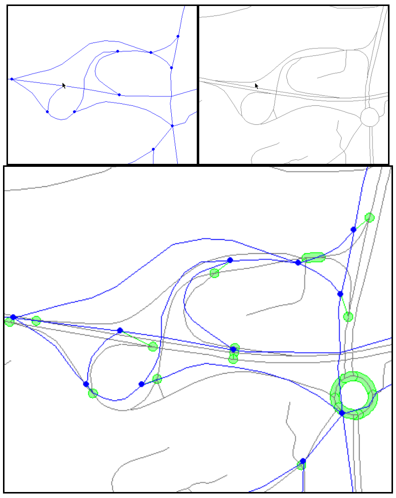

Figure 26. Nodes matching emphasizes some problems

The second direction takes advantage of the different levels of detail. Analysing directly links (the segments between databases) is difficult, but simply highlighting the parts of the more detailed DB matched to the other DB usually emphasizes errors of the process (figure 27): we identify differences between the less detailed database (upper left) and the highlighted part of the network of the more detailed one (bottom).

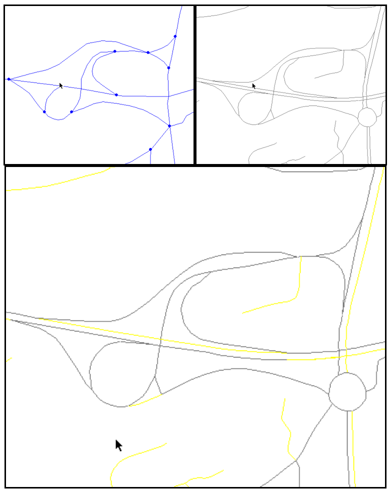

Figure 27. Highlighting the matched part of the detailed network emphasizes some problems.

5. CONCLUSION

The first purpose of these experiments concerned directly the involved databases: we wanted to evaluate which data can be matched and if this can be automated. We thus identified data that can be matched and the types of inconsistencies encountered between the data. Beyond the special case of BDCarto and BDTopo, we believe that these types of inconsistencies may be encountered in a lot of cases. In term of automation, we managed to match from 90% to 100% of these relatively different networks. We thus consider that a highly automated matching of networks at different scales is possible.

The second purpose of these experiments was to identify some of the locks requiring some researches in order to improve data matching. These issues define our forthcoming research agenda.

One issue to be studied concerns the use of semantic information simultaneously to geometric and topologic information. Like many matching processes, our process is not holistic enough and considers in turn these different kinds of information. This should be improved.

Another issue concerns the generality of the process. For each of the studied network (road, electric, hydrographical…), we adapted the process according to our knowledge concerning the data. We hope that this study pave the way to the development of a more generic approach, with this external knowledge as a kind of parameter of the system.

Finally, we believe that database issues (how to efficiently model, store and retrieve matching links) [Vangenot et al. 2002, Friis-Christensen 2003, Mustière and van Smaalen 2006] and ergonomic issues (how to display results of matching) are also of prime importance for an actual matching in a production context.

ACKNOWLEDGMENT

The author whish to thank Eric Grosso who made tests on the hydrographical network, and who improved the environment of the research platform GéOxygène especially for data matching. The author also thanks Guillaume Touya and Cécile Duchêne who made a study on the generalisation of the same networks simultaneously to these experiments, and who helped to better understand the data.

REFERENCES

Devogele T. 1997. Processus d'intégration et d'appariement de bases de données Géographiques. Application à une base de données routières multi-échelles, PhD Thesis, University of Versailles (in French).

Devogele T., Parent C. and Spaccapietra S. 1998. On spatial database integration. *International Journal of Geographical Information Science*, 12(4), pp.335-352.

Egenhofer M.J., Clementini E., and Di Felice P. 1994. Evaluating inconsistencies among multiple representations, In *Proceedings of the Sixth International Symposium on Spatial Data Handling*, Edinburgh, Scotland, pp. 901-920.

Friis-Christensen A. 2003. *Issues in the Conceptual Modeling of Geographic Data*. PhD Dissertation, Aalborg University, Denmark.

Hampe M. and Sester M. 2002. Real-time integration and generalization of spatial data for mobile applications. *Geowissenschaftliche Mitteilungen, Maps and the Internet*, Wien, Heft(60), pp.167-175.

Haunert J.-H. 2005. Link based Conflation of Geographic Datasets. *8th ICA workshop on generalisation and multiple representation*, A Coruña, Spain, July 2005.

Kilpelaïnen T. 2000. Maintenance of Multiple Representation Databases of Topographic Data. *The Cartographic Journal*, 37(2), December 2000, pp.101-107.

Mantel D. and Lipeck U.W. 2004. Matching Cartographic Objects in Spatial Databases. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences : ISPRS'2004*, Vol. 35, Part B4, 172-176.

Mustière S. and van Smaalen J. 2006. Databases Requirements for Generalisation and Multiple Representations. To be published in *The Generalisation of Geographic Information : Models and Applications*, W. Mackaness, A. Ruas et T. Sarjakoski (eds), Elsevier.

Paiva J.A. 1998. *Topological equivalence and similarity in multi-representation geographic databases*. PhD Thesis, University of Maine.

Parent C. and Spaccapietra S. 2000. Database Integration: the Key to Data Interoperability. *Advances in Object-Oriented Data Modeling*, Papazoglou M., Spaccapietra S. and Tari Z. (eds), the MIT Press.

Rodriguez M.A., Egenhofer M.J. and Rugg R.D. 1999. Assessing Semantic Similarities Among Geospatial Feature Class Definitions. *Interoperating Geographic Information Systems: Interop '99,* Lecture Notes in Computer Science, Vol. 1580, pp.189-202, Springer-Verlag.

Sheeren D., Mustière S. and Zucker J.-D. 2004. Consistency Assessment Between Multiple Representations of Geographical Databases: a Specification-Based Approach. *International Symposium on Spatial Data Handling (SDH)*, pp. 617-627.

Sheeren D. 2005. *Méthodologie d'évaluation de la cohérence inter-représentations pour l'intégration de bases de données spatiales – Une approche combinant l'utilisation de métadonnées et l'apprentissage automatique*. PhD dissertation, University Paris 6.

Sheth A. and Larson J. 1990. Federated database systems for managing distributed, heterogeneous and autonomous databases. *ACM Computing Surveys*, 22(3), pp. 183-236.

Stonykova A. 2003. Design and Implementation of Federation of Spatio-Temporal Databases: Methods and Tools. Final Project Report of AMBER.

Uitermark H. 2001. *Ontology-Based Geographic Data Set Integration*. PhD thesis, Universiteit Twente, the Netherlands

Vangenot C., Parent C. and Spaccapietra S. 2002, Modelling and manipulating multiple representations of spatial data. *10th International Symposium on Spatial Data Handling*, pp.81-93.

Walter V. and Fritsch D. 1999. Matching Spatial Data Sets: a Statistical Approach. *International Journal of Geographical Information Science*, 13(5), 445-473.

Zhang M., Shi W. and Meng L. A generic matching algorithm for line networks of different resolutions. *8th ICA workshop on generalisation and multiple representation*, A Coruña, Spain, July 2005.