UPDATING OF GEOSPATIAL DATA: A THEORETICAL FRAMEWORK

J. Raul Ramirez, PhD
The Ohio State University Center for Mapping, Columbus, Ohio, USA

ABSTRACT

Geospatial data are fundamental for the analysis and visualization of the environment, but the currency of the data is critical. Out-of-date geospatial data have historical value, but if the data are used for analysis of current events, the results could be contaminated to a point where the results are incorrect or completely wrong.

There are two possible solutions to the problem of out-of-date geospatial data. The first solution is to collect the data again and then generate new maps or any other type of geospatial data visualization. This approach is very costly, especially for countries of medium to large size.

The second solution is the identification of changes to the landscape and the replacement of these changes by new data. This paper presents the theoretical framework for the updating of geospatial data. Two major issues are discussed: (1) how to detect changes in the data, and (2) how to correct the changes.

INTRODUCTION

The need for up-to-date geospatial data is increasing due to the growth of Geographic Information System (GIS) technology and the needs of homeland security. More and more people are using GIS for analysis and visualization of geospatial data. The quality of GIS results depends on the quality of the geospatial data. Positional and attribute accuracy, completeness, and consistency, affect the data quality. The quality component “completeness” is understood here as having all the information needed for the analysis and visualization for a given moment in time. Completeness may be affected the most by the process of geospatial data updating.

Geospatial data updating is needed because the surface of the Earth is a dynamic surface in a constant state of change, subject to natural forces and human actions. The understanding of these changes, their magnitude, and their causes is a major part of the theory of geospatial data updating. This theory is fundamental to improving current map updating methods. Specific questions to be answered by this theory are:

What are the sources of geospatial changes?
How do these sources change the surface of the Earth?
What are the magnitude and frequency of these changes and over what period of time?
Which of these changes are significant for day-to-day geospatial applications?
What changes on the surface of the Earth are routinely recorded?
What changes should be recorded, and by whom?
How can these changes be forecast?

These questions must be answered in the context of mapping. From this viewpoint, we also need to consider mapping quality standards. Mapping quality standards provide us with guidelines about the magnitude and type of changes allowed in a geospatial dataset at a particular map scale. For example, the National Map Accuracy Standard (NMAS) 1947 (American Society of Photogrammetry, 1980) states, “For maps on publication scale of 1:20,000 or smaller, no more that 10 percent of the points tested shall be in error by no more than 1/50 inch, measured on the publication scale.” Therefore, for maps at scale 1:24,000 the maximum positional error allowed is 40 feet.

The Geospatial Positional Accuracy Standards Part 3: National Standard for Spatial Data Accuracy (NSSDA) 1998 states, “The NSSDA uses root-mean-square error (RMSE) to estimate positional accuracy. RMSE is the square root of the average of the set of square differences between dataset coordinate values and coordinate values from an independent source of higher accuracy for identical points.” This standard “does not define threshold accuracy values. Agencies are encouraged to establish thresholds for their product specifications and applications and for contracting purposes.” Therefore, once a particular agency sets the accuracy thresholds, then changes in the geo-spatial dataset smaller than those thresholds are unimportant and do not need to be considered as part of the updating process.
At The Ohio State University Center for Mapping we have been addressing technical issues related to the questions mentioned above and have developed answers to the first two, “What are the sources of geospatial changes?” and, “How do each of these sources change the surface of the Earth?” These issues are discussed in the following sections.

GEOSPATIAL DATA UPDATING FRAMEWORK

Ramirez (1998) defines geospatial data updating as: “improving and correcting the content of existing data to obtain a current representation of the terrain in agreement with a predefined purpose.” In other words, the purpose of spatial data updating is to revise the information of existing data from a time \( t_1 \) in the past to a time \( t_2 \) in the present.

The obvious reason why geospatial data (or maps) are updated is because they represent a dynamic surface, the surface of the earth. The surface of the earth is constantly changing. We are only interested in the subset of changes in features traditionally represented in topographic maps, including relief. Changes are due to two major sources: (1) natural forces, and (2) human actions.

Natural forces, in general, generate two types of changes: systematic and abrupt. Systematic changes are those continuous changes on the surface of the earth generated by the forces of gravity, wind, life cycle, and others. Systematic changes may include a change in the course of a river, in a coastline, in sand dunes, and in vegetation. Some of these changes are very slow and can only be appreciated after long periods of time, while others can be appreciated sooner. However, all of them are predictable (we know they will happen and affect the surface of the earth) and require, in general, a long time interval to alter the currency of the geospatial data representation.

Abrupt changes are caused by the forces of nature and immediately affect the currency of geospatial data. Examples of these changes are those caused by earthquakes, flooding, forest fires, and landslides. Abrupt changes are unpredictable and in a very short time interval affect the currency of the spatial data representation.

Human actions also generate abrupt changes. The use of explosives causes the most common type of these changes. Sometimes explosives are used with the specific purpose of changing the landscape, as part of an engineering project, but many times they are used as the result of national or international conflicts. War, unfortunately, is one of the major reasons why geospatial data need to be updated.

Humans also generate predictable and unpredictable changes over time. For example, construction of many man-made features such as housing projects, shopping malls, and new roads, may be done over a larger time period, but in these cases the outcome is known and predictable. Unpredictable or random changes are the result of unforeseen circumstances. For example, a road may be permanently closed because of changes in the demographics of a region; or because of economic hardship a city park may be permanently closed; or, because of a new baby, a homeowner may add a room to his/her house. Unpredictable changes are also those changes, such as open-field mining and logging, whose outcomes are unknown at the time and are only evident later on. Human actions can be summarized as resulting in two types of changes: predictable and unpredictable. Ramirez (1996) classifies geospatial changes as shown in Table 1.

<table>
<thead>
<tr>
<th>Origin</th>
<th>Frequency</th>
<th>Magnitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systematic</td>
<td>Constant</td>
<td>Small</td>
</tr>
<tr>
<td>Abrupt</td>
<td>Low</td>
<td>Large</td>
</tr>
<tr>
<td>Predictable</td>
<td>High</td>
<td>Large</td>
</tr>
<tr>
<td>Unpredictable</td>
<td>Medium</td>
<td>Medium</td>
</tr>
</tbody>
</table>

TIME AND GEOSPATIAL DATA UPDATING

Geospatial data updating is time-dependent. There are three epochs of interest when we are dealing with updating geospatial data (see Figure 1). The epoch \( T_F \) is when the data for generation of the existing map (geospatial data)
was collected. Next, there is the epoch \( T_N \) of collection of the data for updating of the geospatial data. Finally, there is the current epoch \( T_C \). This epoch is constantly changing.

The goal of geospatial data updating is to revise the terrain representation in such a way that:

\[
\begin{align*}
T_N - T_P &\sim 0 \quad \text{(amount of changes)} \\
T_C - T_N &\sim 0 \quad \text{(closeness to reality)}
\end{align*}
\] (1)

There are two reasons for this. First, we would like geospatial data to be as current as possible. Second, the cost of updating would decrease as a function of the time difference. We need to remember that updating is an ongoing process that goes on forever.

![Timeline of Geospatial Data Updating](image1)

There are other epochs of interest: the epoch of publication of the map \( T_{PP} \), the date of publication of the revised geospatial dataset \( T_{NP} \), and the always changing time \( T_{C1}, T_{C2} \). We may represent these times as follows:

![Expanded Timeline Representation](image2)

The age of a geospatial dataset (how old is the representation of the ground), is a function of \( T_F \) and \( T_C \) and is given by the expression:

\[
\text{Age} = T_C - T_F
\]

But generally, we may associate the age of a geospatial dataset with the expression:

\[
\text{Age}_M = T_C - T_{PP}
\]

which indicates how long ago a map (geospatial dataset) was published.

Similar considerations and expression can be derived for up-to-date (revised) geospatial data. Of course, in all these expressions, the results are valid for the instant of time they are computed.

Why are these time issues important? To derive a strategy for updating geospatial data, cost is a very important consideration because updating is a never-ending process. Therefore, we need to consider all the facts that will affect the updating cost. The first one is the fact that updating geospatial data cannot be stopped once it is started. The reason for this is very simple. If you stop updating your geospatial dataset, it will become out-of-date, and then you will have lost all the money you spent in the past on updating the dataset. This means that you need to develop approaches to updating that keep the data up-to-date as long as possible, because this will minimize the cost of updating. Because the age of a geospatial data set depends on the epoch in which the images of the ground and other data sources were acquired, we would like to use images and other data sources collected as close as possible to the current epoch \( T_C \).

The types of updating as a function of time are not discussed in great detail in the mapping literature. We present next a brief discussion of the two types of updating: continuous and cyclical. Continuous updating is the type of revision of geospatial datasets that is done on demand. If there is a change worthy to be recorded, it is incorporated in the geospatial database as soon as it occurs. Continuous updating is used mostly with nautical charts, rather than with topographic maps.

Cyclical updating is based on the premise that geospatial datasets should be revised every “\( x \)” number of years. The interval of time “\( x \)” between updating is what it is called the “revision cycle.” Van Zuylen and Shearer (1970) indicate that the revision cycle is a function of the nature of the areas mapped. For example, the revision cycle may be five years for urban areas, 10 years for rural areas, and 15 years for natural areas with little or no cultural development. Of course, the revision cycle is related to the acquisition of the data sources.
Updating time must also be considered. Updating time is the time interval necessary to revise the area of interest. The revision of the whole area must be accomplished in a reasonable period of time. Therefore, it is important to minimize the difference: \( T_{NP} - T_N \).

In other words, the ideal situation is to start revising the geospatial datasets as soon as the images of the ground are collected and processed.

On the other hand, if the updating of the whole area lasts several years (for example, five years), then image acquisition must be subdivided into different projects. Ideally, each year you would like to acquire only those images for the area to be revised during that year. This, of course, will increase the image acquisition cost, but will preserve the currency of the data longer.

**BASIC FORMULATION**

Using Set Theory and features such as buildings, rivers, lakes, hills, and mountains as the basic units, the digital datasets for times \( T_P \) and \( T_N \) can be expressed as:

\[
M_P = \{m_P | m_P \text{ is the digital representation of a feature, at a date } T_P, \text{ and scale } S, \text{ for purpose } P\},
\]

and,

\[
M_N = \{m_N | m_N \text{ is the digital representation of a feature, at a date } T_N, \text{ and scale } S, \text{ for purpose } P\},
\]

where \( M_P \) is the out-of-date geo-spatial dataset and \( M_N \) is the up-to-date dataset.

The set \( M \) representing the terrain for the date \( T_P \), in the past, can be expressed at the date \( T_N \) in the near past as:

\[
M_P = \{D, C, U\},
\]

where:

\[
D = \{d | d \text{ is the digital representation of a feature that no longer exists}\},
\]

\[
C = \{c | c \text{ is the digital representation of a feature before change}\},
\]

\[
U = \{u | u \text{ is the digital representation of an unchanged feature}\}.
\]

Two additional sets need to be considered for the date \( T_N \): the set \( N \) of new features to be shown in the terrain representation (in agreement with the scale and purpose of the representation), and the set \( H \) of the features that have changed. Sets \( N \) and \( H \) are defined as:

\[
N = \{n | n \text{ is the digital representation of a new feature}\},
\]

\[
H = \{h | h \text{ is the new digital representation of a feature that has changed}\}.
\]

The revised geospatial dataset may be expressed as:

\[
M_N = (M_P - \{D, C\}) \cup N \cup H,
\]

A closer representation of what may be realistically expected is:

\[
M_N = U \cup N \cup \alpha C
\]

where:

\[
H = \alpha C
\]

and,

\[
H = \alpha C = \{\alpha_1 c_1, \alpha_2 c_2, \alpha_3 c_3, \ldots\},
\]

where, \( \alpha_i \) is the modification operator, which converts the previous representation of feature \( c_i \) into the new representation \( h_i \).

Identification and manipulation of the sets \( D \), \( U \), and \( N \), and the transformation of set \( C \) into the set \( H \) are the goals of geospatial data updating.
DETECTION OF CARTOGRAPHIC CHANGES

Revision of geospatial data requires identifying changes on the ground, acquiring data reflecting those changes, and incorporating these changes into the geospatial database. Change detection requires comparing the current geospatial data (map) with new data of the same area. Space and airborne images of the ground are becoming easy to acquire and are less expensive. Commercial satellites, for example, with 1 m resolution or better are collecting images of the Earth at a reasonable cost. These images could be used to revise most of the national map series used today. It may be feasible to revise any part of a geospatial dataset several times in a year, if it is needed, independent of the cost. Unfortunately, cost is and always will be a major consideration. Therefore, the questions that need to be answered from the viewpoint of developing a strategy for updating of geospatial databases are, “What are the features that need to be revised?” And, “What datasets, of all those available, should we acquire and use in the revision?”

Let us try to clarify why answering these questions is important. To do that, let us use an example. There are 3,172 digital orthophoto quarter-quadrangles at scale 1:12,000 covering the State of Ohio, USA. Conceptually, these quarter-quadrangles can be generated from the 1m resolution satellite images. Therefore, we could have new digital orthophoto quarter-quadrangles made several times in a year. The size of each image is about 50 megabytes, for a total of 158,600 megabytes for the whole state. To detect changes, we could use one of two traditional approaches. We could compare each one of the images of the previous set of digital orthophoto quarter-quadrangles with the corresponding new image, or we could overlay the Digital Line Graph (DLG) vector data from the current geospatial database on top of the new images and look for changes. Either process is time consuming and costly.

Let us look at some possible results of these comparisons: (1) Let us assume that there has not been any major change in the State of Ohio, since the geospatial database was modified the last time. Because we do not know that fact, after exhaustive comparison we will find out that there was no reason to compare the previous and the new datasets. In other words, we have wasted time and money. (2) Let us assume that only one major change has occurred (for example, a new road was constructed). In this particular case we will need, again, to look at all 3,172 images to find the only change. Even if we find the change in the first image, we will need to continue checking the rest of the images because there is no way for us to know that this was the only change. (3) Let us assume that there is one change per image. Again, we would need to look at all the images in their total extension to locate all the changes.

From the situations described above, it is obvious that to locate all the feature changes in a geospatial database, using only the information provided by new images and the existing digital geospatial database (or the images used for its creation), we need to perform an exhaustive comparison. This comparison of the new images to the images used for the generation of the existing geospatial database or to the existing geospatial database must be exhaustive and, therefore, this will be time consuming and costly. A sampling approach is not recommended because it may generate incomplete results. To decrease the time and cost of detecting changes on the ground, different approaches must be found.

A DETERMINISTIC SOLUTION

As indicated earlier, changes on the surface of the Earth are due to human actions and natural forces. Let us consider in some detail these two sources of changes. Some of the major human actions are construction and exploitation of the Earth’s resources. Construction, in general, requires the issue of permits by the appropriate authority. The issue of a construction permit only indicates the intent to build (generate a change). Exploitation of the Earth’s resources, in some cases, also requires permits (for example, open field mining). Again, the issue of an exploitation permit only indicates the desire of exploiting a specific Earth’s resource. Some construction projects also require an inspection, once construction is completed. Inspection certificates indicate that construction took place (there is a change).

Natural forces change the surface of the Earth. As indicated by the USA Federal Emergency Management Agency (FEMA, 1999) “DISASTER. It strikes anytime, anywhere. It takes many forms—a hurricane, an earthquake, a tornado, a flood, a fire, or a hazardous spill, an act of nature or an act of terrorism. It builds over days or weeks, or hits suddenly, without warning.” The mission of FEMA is “to reduce loss of life and property and protect our nation’s critical infrastructure from all types of hazards through a comprehensive, risk-based, emergency management program of mitigation, preparedness, response and recovery.” In short, FEMA deals with major natural disasters and some man-made disasters. Therefore, there is a centralized agency in the United States where information about natural disasters (and some man-made) can be obtained.
Based on the fact that most changes on the surface of the Earth in the United States are recorded as a permit (and in some cases an inspection), or by FEMA as a response to a disaster, a deterministic approach to the location of feature changes in geospatial databases may be possible. This solution may go as follows:

Let us assume that for a given area we can obtain copies of all construction permits (and corresponding inspection reports, if required) and the location and nature of all natural disasters (plus those disasters caused by humans). Let us assume that we receive this information in a systematic fashion and in digital form. Let us call the construction permit dataset by P and the disaster information by DI. Let us also assume that we can obtain appropriate Earth images for any time epoch. Let us call these images I. The digital geospatial dataset to be updated is MP.

The first step will be the geocoding of the information from the sets P and DI. Most permits require a set of plans showing the location and dimensions of the proposed constructions. These plans or similar information will be used to generate from the set P a closed polygon covering the area affected by the construction. Also, information about the type of construction and if the construction has been inspected and approved will be stored. Similarly, from FEMA’s information, we will geocode a set of closed polygons covering the areas affected by disasters and the nature of these disasters. All this information will be stored in the dataset CH. The dataset CH, therefore, will consist of several closed polygons with attributes. The coordinate values of these polygons will be expressed in the same type of coordinates as the set MP (same datum, same projection system, same type of units).

**Figure 2. Major Steps for Updating of Geospatial Data**

The second step will be the georeferencing of the dataset I. The set I will contain only those images of the surface of the earth that cover all the closed polygons included in the dataset CH. These images will be warped into the geospatial space covered by the set MP. The outcome will be three datasets (CH, I, and MP) that could be overlaid on the same geospatial space.

The third step is the updating of the geospatial dataset MP. This process will be similar to the process described by Ramirez (1998). In this particular case, the process will start in Step 3 of the diagram shown in Figure 2. From the closed polygons of file CH, we will generate georeferenced buffers (Step 3) and continue with the remaining steps. This will be done for every layer of the geospatial database, and the final output will be a consistent up-to-date geospatial dataset MN.
A PROBABILISTIC DETECTION OF CHANGES

In the past we have used Belief or Bayesian networks to compute the conditional probabilities of building structures as extracted from digital orthophoto quarter-quadrangles for updating purposes (Ramirez et al. 1998). We propose to use Bayesian networks for the detection of changes on the ground. The following discussion of Bayesian networks is taken from Ramirez et al. (1998).

As indicated by Pearl (1986, 1988) Bayesian networks are a directed acyclic graph, in which the nodes represent multi-valued variables, comprising a collection of mutually exclusive and exhaustive hypothesis. The labels of these nodes are either given or have to be assessed and the relationships between the nodes are expressed in terms of conditional probabilities. Each variable can be labeled from a corresponding finite set of labels with a certain degree of confidence. The input data may concern any of the nodes. Once they are fed into the right node, a mechanism is provided for the propagation of their information to all other nodes of the system. After the propagation is over, the network will provide the conditional probabilities of each label of each of the variables that occur, given the pieces of evidence observed. Figure 3 (taken from Ramirez et al., 1998) is an example of a Bayesian network.

![Figure 3. Parents and Children of a Typical Bayesian Network](image)

In this example X is a typical node having m children, Y, and n parents, U, as shown in Figure 3 (the node variables are denoted by capital letters and the parent and children also have subscripts to distinguish among them). The purpose of the Bayesian network is to give a belief in each possible labeling for each node after some evidence arrives. To estimate the belief of a node we need the information sent by the parents (causal), the information sent by the children (diagnostic) and the conditional probability matrices. The messages that communicate this information obtained by the parents are denoted by $\pi$ and the messages that carry information from the children are denoted by $\lambda$. The $\pi$ and $\lambda$ messages are depicted in Figure 3.

An element of the conditional probability matrix, $P(x|u)$ gives the probability of state $i$ for node $X$ conditioned on the states of its parent nodes. In addition to the above, each node, except the nodes below $X$, has a $\pi$ vector that can be communicated to its children. Also, each node, except the nodes above $X$, carries a $\lambda$ vector, which can be communicated to its parents. Upon receiving evidence (in the form of a $\pi$ or $\lambda$ message), the node will update its belief and send the corresponding $\pi$ and $\lambda$ message to all its children and parents, respectively, except for the one it had just received information from. Each of the parents and children will then recursively follow the same procedure, until there are no more messages to be sent (i.e., all evidence is absorbed) and the network then reaches equilibrium. The final belief of each node will be the actual posterior probability, i.e., the probability of each node assuming each of its states, given all evidence observed.

In a Bayesian network we represent by nodes all the variables, which influence, either directly or indirectly, the hypothesis. The connections (arcs and their directions) of the network are designed using domain knowledge and form a significant part in designing Bayesian networks. In this case, we propose to design one Bayesian network per information layer (buildings, roads, vegetation). The hypothesis for each case is the facts that determine changes on the ground for each specific layer. For example, in the case of the building layer, zoning regulations, urban growth trends, population trends, commercial development trends, and construction permits, are some of the facts that may be used to predict changes. For each one of these nodes, we need to design how they are connected and assign a conditional probability. Figure 4 shows an example of a Bayesian network for the building layer based on the nodes mentioned above.
A COMBINED APPROACH FOR DETECTION OF CHANGES

We have discussed two possible alternatives for detection of cartographic changes: one deterministic, and the other probabilistic. Both alternatives present many advantages, but also disadvantages. For example, there are two major problems with the deterministic approach: (1) the collection of all the permits and inspection certifications for a specific region get more difficult with the size of the region. This is also true with the information from FEMA. For example, in the case of the State of Ohio, there are 88 counties and many cities, towns, and villages. Each one of them, at some level is responsible for issuing permits and inspection certificates. Therefore, to develop an approach that will guarantee that the entity responsible for the updating of the base map for Ohio will receive all this information accurately and on-time is not an easy task. Perhaps the only way to assure this is to have a law that requires that all those entities provide copies of this information to the revision entity. To have that law in place may be very difficult. (2) The other fundamental problem of the deterministic approach is that it does not account for systematic changes caused by the natural forces. For example, the changes in the bank of a river usually are happening constantly but only after a usually long period of time are these changes large enough to be applied to the geospatial database. Permits or FEMA information generally do not reflect these types of changes.

The probabilistic approach also has two major problems: (1) you need to identify and model for each geospatial layer all the sources that contribute to changes to forecast new changes on that layer. This is not an easy task because we do not have at this time a complete understanding of those sources and their interactions. (2) You need to assign the initial conditional probabilities. We do not have yet a scientific approach to accomplish that. Therefore, at this time personal judgment will be used.

We believe the best we can do today is to work on a solution that uses part of the principles of the deterministic and the probabilistic solutions described earlier. This solution fundamentally will use all the deterministic information available to improve a Bayesian network model of changes. In a case at one extreme, if no deterministic information is available, then the model will be equivalent to the probabilistic model described earlier. In a case at the other extreme, if all changes but the systematic ones are known from permits and FEMA data, then the probabilistic model will be used only to forecast systematic changes. Any other situation between these extreme cases will be accommodated by this solution.

Conceptually, we will start by collecting as much information about geospatial changes as possible for the area of interest. This means that all entities responsible for issuing permits and inspection certifications will be contacted and a request for this information for the interval of time \((T_c - T_P)\) will be issued. In parallel with this activity, we will develop the models for the Bayesian network. These models will cover each geospatial layer. A third activity will be the collection of samples of natural systematic changes. These samples will be collected from historical ground images (for example, NAPP photos of the last 15 years).

Next, the network will be initialized. As indicated by Ramirez et al. (1998), “before any propagation commences we initialize the network. Every node is assigned a vector \(\lambda\), a vector \(\pi\) and a vector \(\text{BEL}\). These vectors have as many elements as there are possible states for this node. To initialize the network, we set all elements of all \(\lambda\) vectors to 1. We also set all elements of the \(\pi\) vectors of the root nodes equal to the prior probabilities for the corresponding states. All the values of the remaining variables of the network can be computed now from the above initialized quantities and the elements of the conditional probability matrices.” This is a crucial part of the
process. Initial values will be tested using permits, inspection certificates, and historical data (especially for systematic natural changes). The outcome of the process will be a set of conditional probabilities for sub-regions of the area of interest. These probabilities will reflect, based on the information available, where changes may have occurred. Those sub-regions with the highest probability, together with those with permits and inspection certificates will constitute the set of possible changes to be investigated during the updating process.

The strength of this approach is based in the fact that it could be implemented immediately without waiting for any change in the law regarding construction permits. This approach is based on volunteer cooperation among the agencies responsible for issuing permits and inspection certificates, FEMA, and the updating agency. The expectation is that this approach will result in an immediate save in the cost of the revision process, and that this fact will increase the amount of volunteer cooperation to a point where only systematic change prediction will be needed.

UPDATING STRATEGY

The goal of this strategy is to maintain an up-to-date geospatial database in a cost-efficient manner. We will assume that the existing geospatial database is many years out-of-date, perhaps, 30 years or so, as is the case for the USA base map. We also assume that revision will be an ongoing activity that will be done in part from images of the ground. Finally, we will assume that cost is an important consideration and that funding is limited.

The first updating of the digital geospatial database will be exhaustive. It will be done using the most current data sources. Ground images no older than one year will be used. Therefore, if the overall revision cannot be accomplished in one year, image acquisition and processing will be scheduled in such a way that the one-year old constraint can be enforced. Updating will be done following the quality thresholds set by the appropriate authorities. Once these thresholds are set, available data sources appropriate for those thresholds will be identified for the area of interest. Updating will be limited to those features whose position or attributes do not comply with those thresholds. The outcome of the first updating will be an up-to-date geospatial database of the area of interest at the same scale as the original one. It is expected that the first revision will be the costliest.

We will adopt the continuous revision approach. As indicated earlier, this means that changes will be incorporated in the geospatial database as soon as possible. To accomplish this, parallel with the first revision, the Bayesian networks for the different geospatial layers will be designed. Also, deterministic information in the form of construction permits, inspection certificates, and natural or man-made disaster reports will be collected from any source willing to provide it. Finally, a set of historical images will be used to study continuous natural changes. The goal is to have a prototype-forecasting model in place at the end of the first year of revision.

Assuming the first revision cannot be accomplished in one year, then during the second year two different revision activities will take place. The first activity is the continuation of the first exhaustive revision for the area of interest. Of course if the exhaustive revision can be accomplished in one year, only the second activity to be discussed next will take place. The second activity is the continuous revision of the portion of the area of interest revised during the first year. The prototype-forecasting model will be used (the prototype-forecasting model includes the Bayesian networks and deterministic information) and appropriate data (including ground images) will be acquired for those places where changes have been forecasted. Revision will be performed in those places and the results will be used to improve the forecasting model.

![Figure 5. A Continuous Revision Process](image-url)

The process described in the previous paragraph will be repeated until the entire exhaustive revision of the area of interest is completed. Of course, each additional year the forecasting model will be used for a larger area.
Once the exhaustive revision is completed, only the forecasting model will be used as part of the continuous revision process. A major effort in this strategy is to increase the volunteer participation of those entities collecting deterministic information about geospatial changes. Ideally, we would like to use the forecasting model only for natural systematic changes.

By adopting this revision strategy, we will be able to answer the two questions posed earlier, “What are the features that need to be revised?” And “What datasets, of all those available should we acquire and use in the revision?” The deterministic information and the forecasting model provide the answer to the first question. The answer to the second question is obtained from the dates of the deterministic information and the forecast. Only datasets acquired after those dates should be used. A major saving in this approach is the fact that only a limited number of images and portions of other datasets are needed (only datasets for the areas where the possible changes are located are needed). This together with our knowledge about dates will maximize the efficiency in the selection of datasets.

CONCLUSIONS

We have presented a strategy for revision of geospatial datasets that introduces the idea of continuous revision for topographic maps. Figure 5 is a conceptualization of a continuous revision process. Emphasis placed on selecting the appropriate data sources from the viewpoint of time. This is important to keep the revision costs low. This strategy also uses data sources different from ground images to forecast geospatial changes and, therefore, to restrict the search for these changes to a minimum. A combination of deterministic and probabilistic approaches for forecasting geospatial changes is suggested. This approach is flexible enough to allow different degrees of deterministic information. The model and the strategy have not been implemented. We hope to do so in the State of Ohio.

REFERENCES


References from BOOKS;