Best Practices and Process Improvement of IFSAR DEM Production

Jessie Yan

IntermapTechnologies Corp.
200- 2 Gurdwara Road
Nepean, Ontario K2E 1A2
Canada
jyan@intermap.com

Abstract

Intermap is working its way into the second year of a multiyear project to produce 5-meter posted digital elevation models of the entire continental U.S.A. The substantial amount of data to be processed and shipped each day poses a tremendous challenge for both operations and management. Intermap is constantly optimizing its work processes to maximize throughput and efficiency. This paper will focus on how Intermap has provided automation to the most critical and time consuming phases of the workflow, three-dimensional (3D) interactive editing and value added vector extraction. A complicated toolset in stream network checking, data finishing, data management and quality control will be discussed. This toolset is currently being expanded into a larger enterprise-wide production control system, in which one of the functions is to use selected public data in predictive analytics to assist in knowledge-based decision making, scheduling, planning and management.

Introduction

The dazzling advances in information technology along with the explosive growth in the availability of geospatial data have created many new business opportunities. More and more applications and business processes operate in a spatial context and include location-specific capabilities. With Google's and Yahoo's recent integration of maps into their search engines and portal pages, it can be anticipated that the ability to widely spread and wisely use geospatial data will continue to gain popularity. Undoubtedly, a highly accurate and low-cost Digital Elevation Models (DEMs) is one of the most important components in geospatial datasets. Driven by the growing market demand and equipped with state-of-the-art airborne Interferometric Synthetic Aperture Radar (IFSAR) mapping technology, Intermap Technologies is uniquely positioned to meet this growing need. As part of its NEXTMap program, Intermap is currently engaged in mapping the entire continental U.S.of some 7.8 million square kilometers at 5-meter posting resolution.

Since the launch of the NEXTMap USA program, Intermap has experienced an unprecedented increase in data acquisition, processing, delivery and management. With the business model having migrated from a project-based to a production-based environment, Intermap now maintains a large data inventory base in its Web store. This large database puts great pressure on the existing production system, resulting in a need to find better, more efficient and less costly ways to conduct daily business. This paper aims to share the commercial experience and good practice guidelines of Intermap's NEXTMap program. It also highlights the collective role that ArcGIS technology played in supporting the decision making process and its benefits on day-to-day operations. A few selected scenarios will be discussed to share practical ideas and expertise with peers in the geographic information system (GIS) community. Finally, some promising areas for future expansion will be identified.

Core Business Review

Intermap Technologies' core business is defined as collecting, processing and providing industry standard digital representations of the earth's surface using cutting-edge IFSAR technology. Intermap's core products include Digital Surface Models (DSMs), bare-earth Digital Terrain Models (DTMs) and Orthorectified Radar Images (ORIs). Various experts have published articles and papers about IFSAR technology, its commercial implementation and the validity of Intermap's core product specifications and accuracies. These subjects are beyond the scope of this paper. Detailed information can be obtained by visiting Intermap's Web site www.intermap.com.

Successful DEM production requires a cooperative team effort. Due to the advanced technology employed and the complexity of the end-to-end work flow, the system has to be able to adapt to different product specifications and be able to produce customized data formats and sizes. Over the past several years, Intermap has been worked diligently to improve its production facility, optimize its workflows and monitor performance metrics. Figure 1 shows the high-level view of the end-to-end IFSAR DEM generation process.

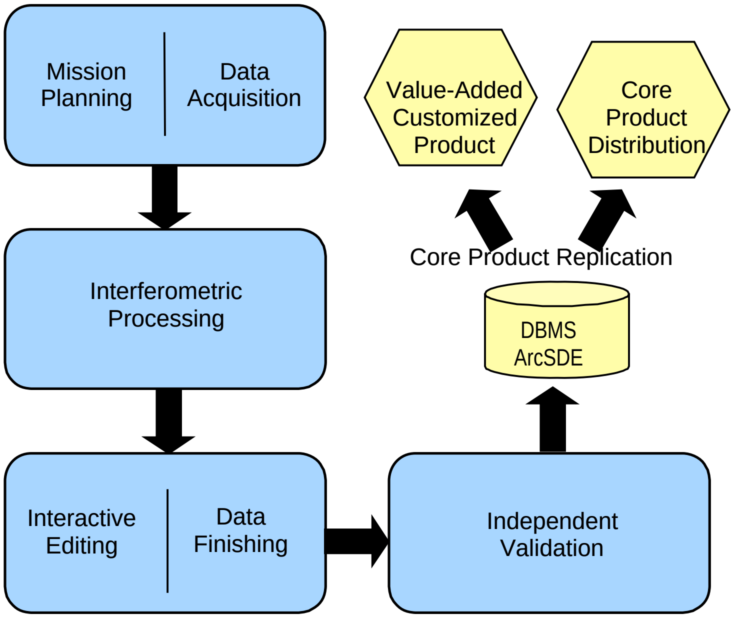

Figure1. IFSAR production work flow process diagram

While Intermap's DEM production facility relies heavily on its proprietary Interferometric Editing System (IES), ArcGIS is frequently called upon to work in parallel with IES to assist in decision making, data evaluation and analysis. Due to the repetitive nature of these tasks, numerous automated routines have been developed to handle them. The next section demonstrates a few snapshots of the usage of ArcGIS technology in the context of Intermap's NEXTMap program evolution.

Integration Challenge

At the beginning of the NEXTMap program, a new in-house interactive DEM editing system known as IES was introduced to replace the legacy SOCET Set photogrammetry solution. To facilitate the smooth transition and integration into the rest of the production system, ArcGIS was used as third-party software to validate the IES-produced data. These data were validated against the core product specification as defined in the Intermap Core Product Handbook; ArcGIS was also used to aid in the identification of any ambiguities and quality issues.

The very first dataset - 9 tiles in West Virginia edited by IES to the core product specification - was thoroughly evaluated, and the findings and suggestions were used to assist in process optimization. The following is a selected list of major assessments:

- *DEM and imagery co-registration*. Ambiguity in raster reference is not uncommon in the geospatial industry, especially when dealing with data of different resolutions. Intermap's DEMs are structured at 5-meter posting, and their corresponding images are at 1.25-meter posting. Georeferencing and conversion tools are used to examine BIL, GeoTIFF and ArcGrid formats to ensure proper alignment - center of the upper left pixel. As a result, tile geographic extents are unambiguously documented both in accompanying metadata files and in the Intermap Core Product Handbook.

- *Edge matching mechanism*. This is a common bottleneck in many elevation/raster datasets and can also exist within the IES process. Edge issues sometimes consume many production resources. IES map sheets are designed such that a one pixel overlap is reserved on the right and bottom of the tile for edge matching purposes. An AML routine was written to read each elevation value for the overlapping pixels and compare the difference. Statistical analysis follows to determine the distribution (mean, max, min and standard deviation, etc.) of the data analyzed and further evaluate the work of the edge matching mechanism.

- *DSM and DTM difference*. Theoretically all DTM elevation values should be lower or equal to the corresponding point in the DSM, but due to the nature of the bald-earth algorithm and subsequent local/global filter processing, a small range and percentage of elevation difference is allowed in the final product. Raster calculation in ArcGIS is utilized to compute the maximum and minimum difference between the DTM and DSM, and several ranges (2 meter, 5 meter,10 meter, 20 meter, and so on) were presented. Analysis of these data yields some interesting insights into the nature of the errors in each DEM and the overall data quality.

- *Spike and well detection.* Several approaches were employed in the effort to find elevation anomalies. Contours were generated at a 5-meter interval to visually display any potential anomaly, but the large dataset and small interval make this approach computation-intensive. At the same time, neighborhood and focal functions were used in convoluting the desired neighborhood to find extreme values. Again, ranges of difference were presented for further statistical interpretation.

- *Stream cross section***.** The stream network is one of the most important features to be edited because river monotonicity and water confinement are crucial to many applications, such as flood modeling. The ArcGIS profile tool proved to be useful in the examination of river cross sections and course profiles. The same tools were used on other features where surface elevations need to be edited such as roads, railways and airport runways.

The above assessments of conformity served as an important prerequisite for acceptance of a new editing system and precise data description in the Intermap Core Product Handbook. As a result, corresponding editing procedures were introduced or modified to deal with the issues detected during evaluation.

Rule-Based Target Mask

Rule-based raster classification utilizes multiple spatial datasets and recognizes that the distinct geographic patterns of each layer provide valuable information that can be used to draw “rules” for identifying spatial features within the target layer. A rule is considered to be a series of conditional statements that identify the range of values that meet the criteria in the input data layers.

Problem Statement

Due to Radar’s inherent characters, relatively poor signal returns can be expected from dense tree cover and high-relief terrain resulting in deteriorated data quality. In line with the increased production cycle and need for effective editing, a solution was sought to concentrate editing activities mainly on high-quality areas, which we defined as the targeted area. The targeted area receives a full edit according to established core product edit rule, while the lower quality areas are masked out and receive a reduced editing effort, focusing on major radar blunders and connectivity of features such as major streams and roads.

Spatial Data Layers

Two public data sources were used and a tool was developed in an ArcInfo workstation environment to produce the 8-bit GeoTIFF format edit mask. The following steps highlight the major functions of the tool:

1. Regroup the popular 21 classes of National Land Cover Data (NLCD) at 30-meter resolution into 5 categories based on their contribution to the editing effort.

    - Cat 1: Open Water - NLCD class 11,12
    - Cat 2: Urban Area - NLCD class 21,22,23
    - Cat 3: Barren Surface - NLCD class 31,32,33
    - Cat 4: Forested/Shrubland - NLCD class 41,42,43,51
    - Cat 5: Others - NLCD class 61,71,81,82,83,83,85,91,92

2. The slope factor is calculated from the Shuttle Radar Topography Mission (SRTM) DEM data. A threshold value of 10 degrees is established based on IFSAR quality and product requirements.

Rule Generation

By combining the land cover and slope information, a simplified decision tree model is generated to produce a new raster layer representing the targeted/non-targeted area. An example of the rule used is:

IF {Cat 1 OR Cat 2 OR (Cat 3 OR Cat 5 AND slope < 10)} THEN targeted area
ELSE IF {Cat 4 OR (Cat 3 OR Cat5 AND slope > 10)} THEN Non targeted area

The above criteria were established in the context of editing radar data of California, as illustrated in figure 2. As Intermap collects data from other parts of the country where terrain may not be a significant factor, the rules will be adjusted, but the programming logic will remain the same. For example, target areas in Florida and Mississippi valley are controlled more by drainage patterns and proximity to inhabited areas than by terrain and vegetation.

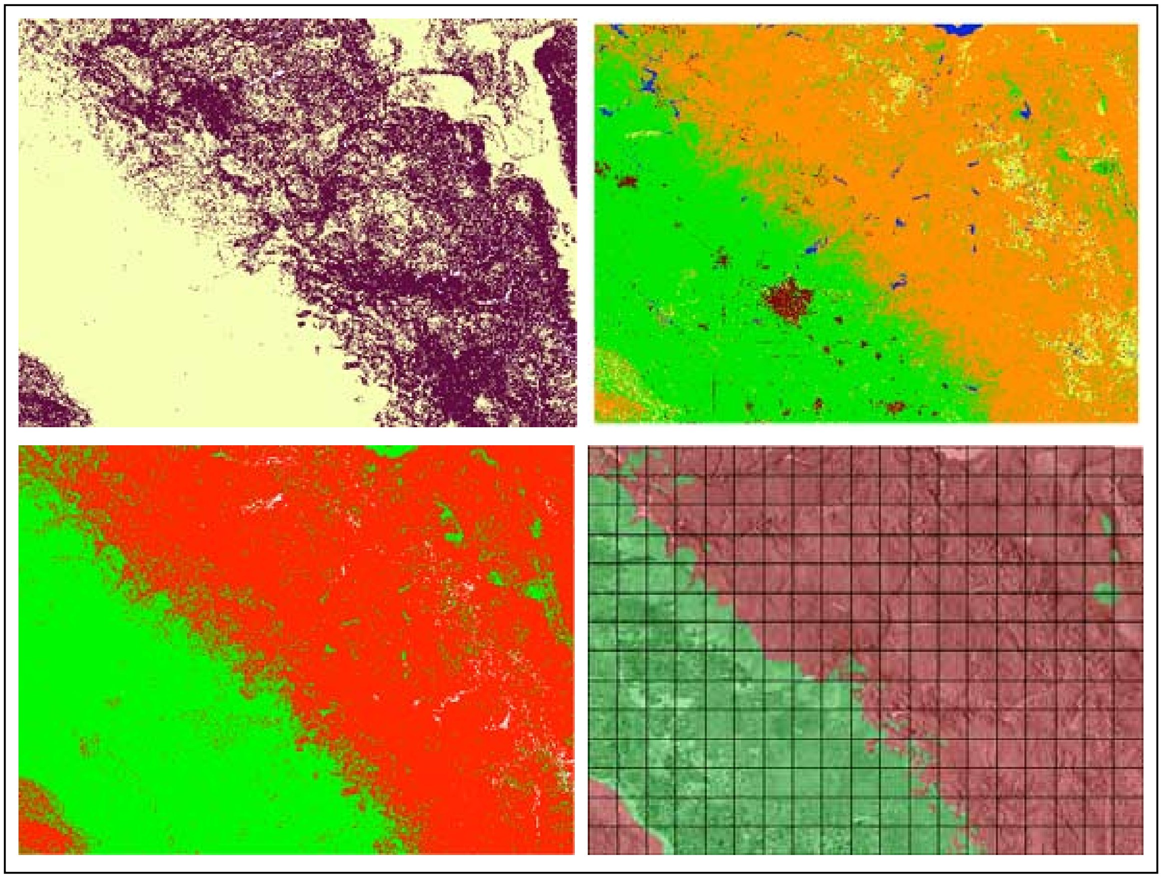

Figure 2. Target mask generation process
Upper left: slope derived from STRM DEM. Upper right: reclassified NLCD data.
Lower left: automatically generated target mask. Lower right: manually generalized mask overlays with ORI.

Data Finishing and Management

An efficient way to eliminate redundancies and potential human errors, as well as increase productivity, is to automate frequently performed tasks, such as those which occur during the data finishing stage. Although IES allows for the classification and primary edit of the DEM, the edited data need to be finished, formatted and completed (in terms of accompanying metadata). The data are then ready to be uploaded into Intermap's central database for inclusion in the company Web store for dissemination to the end customer.

Process Automation

Typically, data finishing deals with large amounts of data in a short timeframe. Therefore, process automation becomes a high priority in this stage. Various tools were developed to locate spatial data on dedicated server platforms, and basic statistics and file system checks are performed on each dataset. The following routines are used frequently as a batch process and are sometimes scheduled as a night time job.

- Convert data from IES native format to ArcSDE format ready for Web store upload or delivery in a customer-specified format.
- Check data consistency, essential statistics (min/max elevation and difference map) to ensure data conform to product specifications.
- Check data content of original data and data copied to delivery media.

These automated processes have provided qualitative improvements in operations and contributed to achieving the corporate goal of maximizing value, minimizing risk and optimizing performance.

Metadata Generation

Metadata generation has been part of Intermap's daily routine for many years and has become an important component of spatial data sharing and exchange. Intermap uses in-house utilities with USGS mp parser to generate FGDC-compliant metadata in three formats: ASCII text, HTML and XML. Several implementations coexist nowadays in order to satisfy different levels of customer requirements. The ArcGIS metadata program version includes information about which flight line segments a specific tile interacts with. Other definitions in the metadata include acquisition-related information such as flight direction, time, and base station, which provide the user with a better understanding of the dataset's characteristics. Figure 3 highlights the interrelationship between finished tiles and their flight line segments.

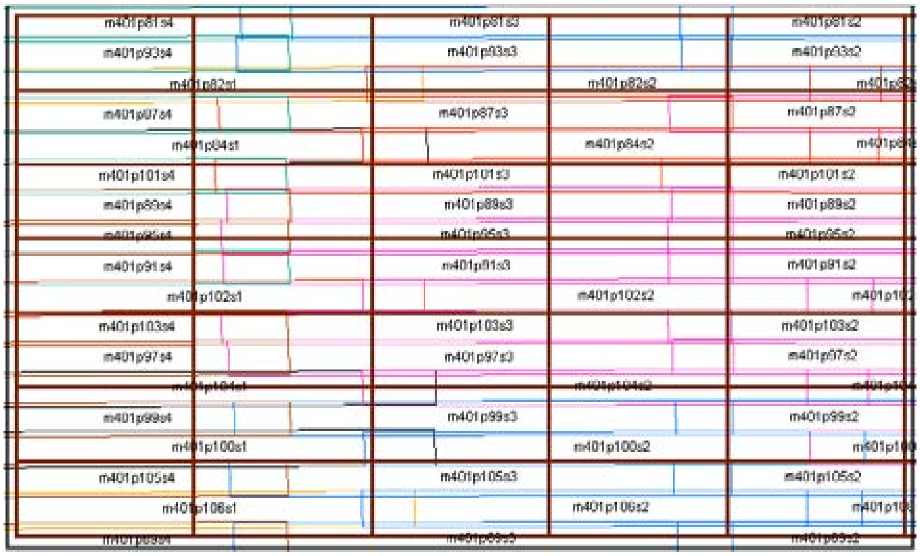

Figure 3. Data acquisition flight line information and tile boundaries

Value-Added 3D Vector Extraction

Value-added products are those built from standard datasets to which information is added to enable customers to use them for particular applications. Since commercial off-the-shelf geoprocessing functionalities can not always satisfy the requirements for building value-added products, Intermap uses COM-based ArcObjects to take advantage of modular software components. In this case, 3D vector stream networks are compiled from IFSAR base data using Intermap's legacy system. ArcObjects is deployed to quality check the elevation values along the river course. A customized code is added into the ArcMap interface so that suspicious points can be rooted out and corrected accordingly. The following shows the code logic for this QC tool:

```
For each iteration - each layer in ArcMap doc /or user selected layer
{
        Ffor every FeatureCursor in FeatureLayer
         {
                Check Z-aware – make sure it is a 3D polyline feature
                For each Feature's point collection
                {
                 Compare current vertex z value with previous one
                 Output bad points where $elevation_{downstream} > elevation_{upstream}$
                }
         }
}
```

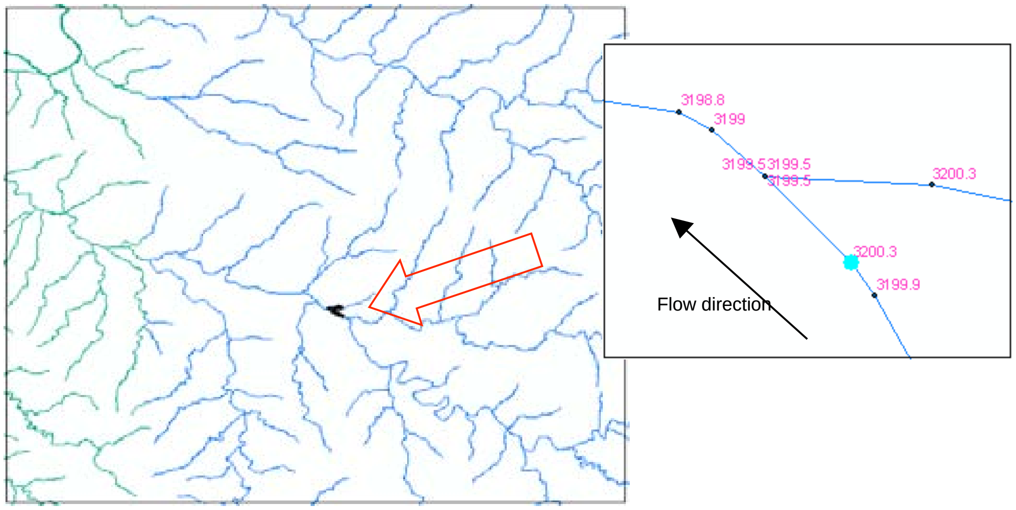

Figure 4. Vertex z value checking in 3D shape file.

In addition, ArcObject progressDialogue is used to display the program's progress; this is helpful information particularly when multiple files are processed. An example message reads:

```
Processing (layer Name) out of <total number of Layers>
Currently at Feature <number> out of <total number of features>
```

Towards Enterprise Solution

All of the above-mentioned tools and guidelines have played an important role in the Intermap operational environment and are considered as critical infrastructures in a broader context. They lay the foundation for building a larger, more complex and high-performance enterprise system. An enterprise-wide business process is a popular trend in today's mainstream information technology and GIS strategy. Apart from presenting technology challenges, it requires tremendous resources and long-term commitment to build a successful system. Such transformations will not happen overnight, but rather incrementally, element by element. An enterprise-wide GIS implementation is under construction within Intermap. The current initiative of predicting editing level of effort is clear evidence of a shift towards the long-term corporate strategy of centralized data sharing, administration and coordinated activities.

Goals and Objectives of Prototyping Level of Effort (LoE)

- Explore the possibility of mining existing IES and supporting datasets to extract meaningful information for resource management and effective scheduling.
- Explore the possibility of predicting the editing LoE based on publicly available data prior to IFSAR acquisition to facilitate middle- and long-term strategic planning.

Methodology - Statistical Approach

By establishing the relationship between terrain features affecting major editing activities and the actual editing effort used, a multiple criteria regression model can be built using historical data as a sample set to predict the future LoE. The prototype based on the California dataset will be used in other parts of the country and eventually integrated into an enterprise-wide business intelligence solution.

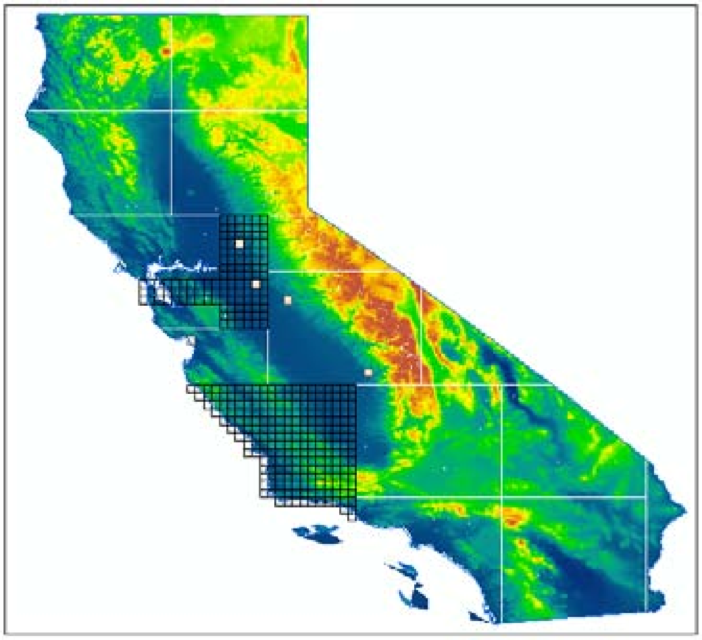

Figure 5. Terrain characters and sample tile locations

Existing Data Sources (Input Layers)

- The USGS National Land Cover Database (NLCD92)
- The U.S. Census Bureau TIGER LINE feature (roads, rails, hydrography)
- The Shuttle Radar Topography Mission (SRTM) DEM
- NEXTMap USA acquisition blocks and tile boundaries
- IES database edit time records

- IES system XML report pertaining to edit activities
- Production Excel index indicating editing time used

A rich array of data exists in Intermap's current production system. The challenge is to make use of these data by obtaining meaningful information to assist in daily operation, milestone scheduling and high-level decision making.

Reasoning Process

The power of GIS again manifests itself in geoprocessing by calculating all explanatory variables construct the prediction model. By consulting domain experts and intuition identification, the following independent variables were determined as relevant, and a set of Python and AML scripts have been developed to extract each of the variables at tile level:

- Tile area in square kilometers
- Targeted editing area in square kilometers
- Area of forest in square kilometers
- Area of urban in square kilometers
- Area of open water in square kilometers
- Length of Double Line Drainage (DLD)
- Length of Single Line Drainage (SLD)
- DLD/SLD adjacent to forest/urban
- Length of road/railway
- Road/railway adjacent to forest/urban
- Percentage of tile in which slope is classified every 5 degrees up to 25 degrees
- Terrain roughness index

After the calculation of the above variables, correlations between each variable and the dependent variable (edit time) are calculated. The correlation coefficient between variable *x* and *y* which we label r_{xy} takes the form:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}}$$

Note that variable selection is a critical and iterative process; special considerations have been given to issues like spatial correlation, multi-collinearity, outliers and nonlinear relationships. Several combinations of intuitive variables that appear as explanatory variables on the right hand side of the regression equation are carried out and subsequent ANOVA and R^2 values are analyzed.

$$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_n x_n + \varepsilon$$

Where *y* is the predicted value of the dependent variable, *x* is independent variable values for *n* observations, b_0 is intercept, *b* is partial regression coefficients for the corresponding *x*; _ is the error term reflected in the residuals.

Preliminary Findings and Observations

An expected strong correlation between established predictor and editing effort was not particularly well formed during the prototyping stage. A further investigation into why the desired R^2 was not present indicated several issues in the current process configuration. The following factors were determined to be the major affecting elements:

- Quality and consistency of historical sample data in terms of editing effort measured by hours
- Operator variance through time and different proficiency levels
- Consistency in edited features
- Certain unquantifiable activities; i.e., radar artifact editing and processing fix-ups

In addition to the above dynamic factors, continually evolving edit rules and technical enhancement of the software environment further complicate the estimation process. After careful review and deliberation, Intermap arrived at several realizations about the complexity and sophistication of business conditions in a factory setting. This exercise provides an excellent opportunity to systematically and quantifiably understand the established edit process and paves the road towards more informed, knowledge-based decision making rather than intuitive management. The adoption of the ESRI framework serves as a solid step forward, and future work will benefit greatly from ESRI's ambitious and comprehensive enterprise implementation.

Conclusion and Looking Ahead

Intermap believes that the future demand for high-resolution, low-cost geospatial data is enormous. The success of an efficient end-to-end production system is absolutely critical for Intermap to be successful at meeting this demand. The general guidelines and practical examples discussed in this paper are just a few of the many requirements for optimally configuring processes to serve better and larger productions ahead. The best-practice criteria and constant process improvements have received a great deal of attention since the inception of the NEXTMap program, and significant progress has been made along the way both in engineering and in practice. The current successful operation establishes a foundation for future expansion.

Although the toolset discussed was primarily developed out of necessity, the emphasis is always given to the generality and reusability of the code so that as business requirements change, this toolset can be repackaged as loosely coupled components in newer and larger enterprise-wide solution strategies. Intermap believes that the most effective solutions build on lessons learned. Moving forward, much work remains to ensure the success of such a system, especially in extending the system functionality and integrating it with existing subsystems.

Acknowledgements

The author gratefully acknowledges the support of Ian K. Isaacs Intermap director NEXTMap programs, Dan Lynch, Intermap production manager, Denis Charland, Intermap shift manager, John Michael, Intermap senior project manager who kindly provided valuable comments on an earlier version. Special thanks also go to Carol Thorpe, director, NEXTMap product promotion and Nick Allan for their support during the writing of this paper.

References

Zeiler, M. (2001) ESRI, Exploring ArcObjects.
Peters, D. (April 2004) ESRI, System Design Strategies.
Rogerson, P.A. (2001) Statistical Methods for Geography, SAGE Publications Ltd.

Author Information

Jessie Yan

Intermap Technologies Corp.
Suite 200, 2 Gurdwara Road
Nepean, Ontario
Canada K2E 1A2

Tel: (613) 226-5442
Fax: (613) 226-5529
jyan@intermap.com