

【研究・技術ノート】

更新データの品質検査のための空間データの ID 番号を活用した 差分抽出ツールの開発

渡邊孝三*・岡部篤行**・中村秀至***

Development of a Difference Sampling Tool Utilizing Spatial Data ID Numbers to Evaluate the Quality of Data Updates

Kozo WATANABE, Atsuyuki OKABE and Hideshi NAKAMURA

Abstract: The continued development of GIS requires that spatial data be edited reliably. However, methods to manage the quality of spatial data during updates have not been established, the result being that effort put into the update is not connected to an overall improvement in data quality. Recognizing this problem, a method of evaluating the quality of specific cases is researched, and a tool developed for practical use. While the quality evaluation of the update process involves two steps, the sampling of the update results and the quality evaluation of those results, this research implements a practical method of prioritizing the first of these steps, then categorizing patterns in the differences before and after the update, and presenting the results of the difference sampling with an assigned ID. The tool developed for this research was used on the results of data updated in a local government project, and it was determined that the quality did not achieve the quality requested by the contractor, and that the results of this research could contribute to the overall quality of spatial data as it is updated.

Keywords: Geographic information standard (地理情報標準), Difference sampling (差分抽出), Spatial data quality (空間データ品質), Topological consistency (位相一貫性), Computer aided automatic error detection (自動検査), intersection (交差)

1. はじめに

空間データの作成は、新規に空間データを作成する場合と、すでにある空間データ（以下元データという）に経年変化部分の追加、削除、訂正（以下加除訂正という）を行って空間データ（以下更新データという）を作成する場合がある。いずれの場合でも、地理情報標準プロファイル（JPGIS）（国土地理院，2005）の「品質要求，評価及び報告のための規則」に基づいて作成された品質要求を満たすことが求められる。多くの自治体では、更新データの作成を行う場合、契約書または特記仕様書に元データに対して経年変化の加除訂正を求めている。したがって、契約上の品質責任範囲は、加除訂正範囲と見なすことができる。

ところが、空間データの全域を検査する品質検査ツール（平田，2008）はあるが、加除訂正範囲を抽

出するツールも、加除訂正したデータを品質検査するツールもなかった。

今後、空間データの作成技術が進歩し品質が向上すると、従来、精度不良等で、新規に全空間データを作成していたが、更新で済むようになり、更新データの品質検査の重要性が更に増すことが予想される。そこで、本研究は、受託者が更新データの納品時に、品質保証を行うための品質検査の範囲を特定することに焦点を当てて研究を行った。

多くの自治体は、国土交通省の公共測量作業規程（国土地理院，2004）（以下作業規程という）の、デジタル・マッピング（以下DMという）仕様に基づいてデータを作成している。

ここで、空間上に存在する建物、道路等を地物とよび、空間データ上の面・線・点、属性等を要素とよぶこととすると、DMデータは、地物が重なった場

* 正会員 株式会社 NTT データ信越 第二事業部 (NTT DATA Shinetsu Corporation)
〒135-8671 東京都江東区豊洲 3-3-9 アネックスビル 17F
E-mail : watanabeko@nttdata-shinetsu.co.jp

** 正会員 青山学院大学 総合文化政策学部 (Aoyama Gakuin University)

*** 正会員 株式会社三菱総合研究所 社会システム研究本部 (Mitsubishi Research Institute, Inc.)

合に間断表記をするため、要素を分断、集合させて作成する場合がある。間断表記では、1つの地物が重なる部分と重ならない部分で線分を途切れさせて作成し、他の地物が覆いかぶさった下の地物は破線等の表記をする。このため、DM データから作成している空間データの元データも更新データも重なりのある場合には、地物を複数の要素で作成している。

更新データは、次の4つで構成される。

- (1) 追加要素：元データに追加された要素。
- (2) 訂正要素：元データの形状、属性が修正された要素。
- (3) 削除要素：元データにあり、更新データから消去された要素。
- (4) 無変更要素：経年変化がなく、元データと形状、属性が同じ要素。

このように、更新データは、元データと加除訂正で作成したデータとが混在しているところに特徴がある。

したがって、更新データは、無変更要素のように受託者が品質の責任を負わない要素も含まれていることから、品質責任を負う範囲を明確にするために、効率的かつ正確に加除訂正要素を抽出する方法の確立が重要な課題となる。

加除訂正要素を抽出するためには、元データと更新データの要素、座標、属性を比較して分類抽出する方法（以下差分抽出という）がある。本研究は、この差分抽出を利用して、追加、削除、無変更要素を抽出することとした。

差分抽出も時点の違うデータを比較する時系列管理と考えることができる。これまで、空間データの幾何要素と主題要素に時間要素を付加して、経年変化を捉えようとする時空間データベースの考えがあり、時空間の静的、動的なデータモデルの提案（太田，1999）や、時々刻々変化する道路・交通情報管理について研究されている（星野ほか，1999）。また、数値地図2500のフォーマットをできるだけ変更せず、既存にあるID番号に要素の発生、消滅の情報を付加することで時系列管理を行う方法等が報告されている（門脇ほか，2001）。

町田市共通基盤データ、岐阜県共有空間データで

は、加除訂正部分の抽出を目的として、要素に固有のID番号を附番し、更新の際のID番号の附番方法に時系列情報を意味付けしたものを仕様化している。そこで、この仕様に基づき品質要求に適合する更新データを作成し、加除訂正を行ったデータのみを対象とする品質検査を行うことができるように、差分抽出ツール（以下ツールという）を開発することとした。

ツールは、差分を抽出する必要からデータの論理一貫性の書式一貫性、定義域一貫性、概念一貫性のチェック機能を持たせた。また、追加、訂正、無変更、削除要素の抽出により、それぞれの要素と数量を知ることができる。これを利用して、ツール適用結果と、真値（今回の場合、土地台帳、家屋台帳）及び受託者の品質検査結果を比較することでツールの有効性を検証することとした。

ツールは、追加、削除要素において元データと更新データの同一ID番号の有無によって判定した。また、訂正要素は、元データと更新データの座標比較によって判定するが、訂正要素とする移動量を自由に設定できるようにした。そして、訂正要素とする属性変化は、1文字でも変化がある場合とした。また、位相一貫性を除く論理一貫性と完全性の適合水準は、全数検査で品質要求と同じく0%とした。

本論文は、序説を含めて5節で構成されている。2節では、更新の際に起きる要素の変化を述べる。3節では、元データと更新データの差分抽出を行う際の抽出方法、ID附番の原則、ツールの処理プロセスを述べる。4節では、3節で開発したツールを使って、実際のデータで分析した結果と効果を述べる。5節では、成果のまとめと今後の課題を述べる。

2. 更新の定義

この節では、空間データの加除訂正を行う際に、要素が重なった場合に生ずる変化、すなわち他の要素の影響を受けない独立変化と、影響を受ける関連変化について述べる。

2.1. 加除訂正

加除訂正の修正作業は、新たに要素を加える「追

加], 元データの形状, 属性あるいは両方を修正する「訂正」,そして元データの要素を除く「削除」と, 元データに形状, 属性にまったく手を加えない「無変更」の要素で構成されている。

要素は, 加除訂正で重なりが発生や重なりが解消されると, 他の要素が影響を受けて変化することがある。修正作業で要素が重なると, 例えば, 要素の重なっていない部分を実線表記し, 重なっている部分を非表記または破線表記している。国土基本図図式(国土地理院, 2005)(以下図式という)は, 図表第2に「双方が立体関係にある場合には, 特に定める場合を除き, 下方の記号を間断して表示する」と規定している。

すなわち, データ更新に際し, 重なりと重なりが解消に対応して, 図式表記するために, 個々の要素の修正だけにとどまる場合と, 修正によって他の要素が影響を受け, 形状変更, 分割・集合等を行わなければならない場合が生ずる。

図1は以下に示す作業, 重なり規定, 発生要素の関係を図示したものである。

- (1) 作業: 加除訂正の作業を指し, 追加, 訂正, 削除, 無変更の作業種類がある。
- (2) 重なり規定: 要素が重なっている場合の要素の変化は, 図式に定められた表記の規定により決まる。2.1.1, 2.1.2節で示すように, 修正する要素自身のみ修正する「独立変化」と, 対象要素の修正に関連して他の要素も修正する「関連変化」がある。
- (3) 発生要素: 加除訂正で生じた要素で追加要素, 訂正要素, 削除要素, 無変更要素がある。

以下独立変化と関連変化について述べる。

2.1.1. 独立変化

独立変化とは, 重なりによって他の要素が影響を受けずに, 修正する要素のみが変化することをいう。独立変化は, 他の要素との重なりがない場合と, 図式で重なりを許している要素間で起きる。

図2は, 独立変化の例で, 左から右上に送電線が通っている。送電線は, 他の要素と重ねて表記するので, 追加されても他の要素は影響を与えない。こ

のような要素には, 他に索道, 行政界等がある。

2.1.2. 関連変化

関連変化とは, 重なりで, あるいは重なりが解消で, 影響を受けて他の要素に変化を与えることをいう。

関連変化には, 次の3つのタイプがある。

- (1) 重なりを許さない地物間で, 境界が共有または隣接する場合に, 一方の要素の修正によって他の要素と重なりが発生した場合。
- (2) 重なりを許す地物間で, 要素同士が部分的に重なった場合, または, 部分的重なりが解消によって重なり部分を復元する場合。
- (3) 重なりを許す地物間で, 要素の重なりによって分断した場合。または, 重なりが解消によって分断された要素が集合する場合。

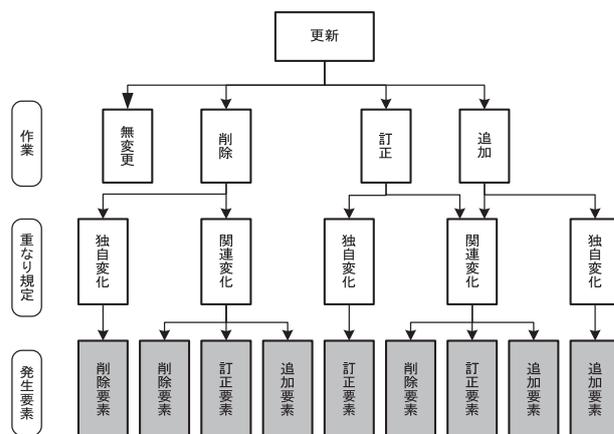


図1 作業と発生要素の関係



図2 独立変化の例

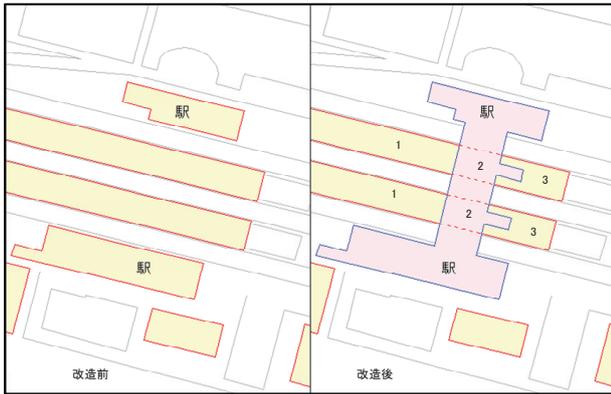


図3 関連変化の例

図3は関連変化の(3)の例である。元データではプラットフォームのみであった。しかし、経年変化でプラットフォームの上に建築物が加わったため、プラットフォームは、建築物の下(破線部分)と、建築物の左右の3つ(図中に示した数字)に分割されている。このように、関連変化は、重なりによって他の要素が影響を受けて変化することをいう。

関連変化は、修正作業の加除訂正によって、新たに追加要素、訂正要素、削除要素が生成される。また、作業方法によって追加要素、訂正要素、削除要素の数が異なってしまう。公共測量作業規程は、修正方法まで規定していないので、要素数の増減を管理したい場合は、作業方法の指定が必要である。

3. 差分抽出ツールの開発

この節では、差分抽出方法、ID附番の原則、処理プロセスの概要について述べる。

3.1. 差分抽出方法

更新データの品質検査は、図4に示すように、元データと更新データの差分である追加、訂正要素と更新データ全体について行う必要がある。そのため、追加、訂正要素の抽出を行う必要がある。

差分抽出の方法には次の2つの方法がある。

- (1) 座標をキーとして変化分を抽出する方法。
- (2) 要素の属性にユニークなID番号を附番する方法。

座標をキーとする方法では、座標列の比較によって追加要素と削除要素の抽出はできる。しかし、要

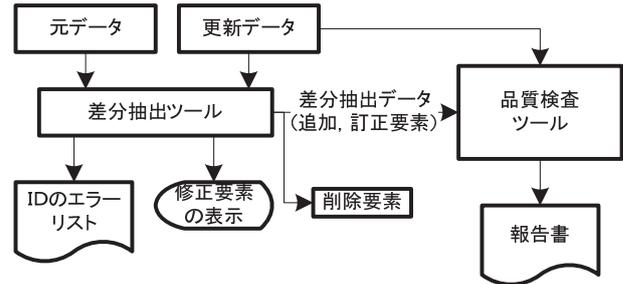


図4 更新データの抽出と品質検査の関係

素の形状が変わらず、位置が全体的に移動した場合、座標列の一部が変更された場合には、追加要素、訂正要素のいずれであるかの識別がつかない。したがって、変化分の抽出はできるが、変化の種類までは識別できない。

一方、ID番号附番方法は、ID番号作成のための複雑さはあるが、ID番号の付番に規則性を持たせることによって追加要素、削除要素、訂正要素の識別を確実に行うことができる。したがって、変化の種類と数の把握に適している。そして、ID番号附番方法は、追加要素、削除要素を座標比較せず、ID番号だけで識別できるので処理時間を短縮できる。また、ツールに、ID付番の文字型の混在や、ID番号の重複等をチェックするための機能を加えることによって、ID番号管理の問題点を補うことができる。そこで、本ツールでは、変化の種類と数の把握に対応するため、ID番号附番方法で差分抽出を行うこととした。

3.2. ID附番の原則

本研究では、追加要素、訂正要素、削除要素、無変更要素の識別を行うため、ID附番の原則を次のとおりとした。

- ① ID番号は、要素単位にユニークな番号を付与。時間属性に追加年月日を付与。
- ② 追加要素のID番号は、最終番号+1で附番。
- ③ 削除要素のID番号は欠番。
- ④ 訂正要素のID番号は変更しない。時間属性に訂正年月日を付与。

表1は、差分抽出の処理および判定の関係を示したものである。処理1では、同じID番号が元データ、

更新データの片方にしかない場合で、元データのみにある場合を削除要素とし、更新データのみにある場合を追加要素と判定する。処理2では、同じID番号が元データ、更新データの両方にあるものに対して、同じID番号の座標列および属性の比較を行って、元データと更新データが異なる場合を訂正要素とし、同じ場合を無変更要素と判定することとした。

表1 差分抽出の処理と判定

処理1		判定1	処理2	判定2
元データ ID 番号	更新データ ID 番号			
1	1	同じ ID 番号	座標, 属性の差分抽出	無変更 or 訂正
2	欠番		削除要素	
3	3	同じ ID 番号	座標, 属性の差分抽出	無変更 or 訂正
4	欠番		削除要素	
5	5	同じ ID 番号	座標, 属性の差分抽出	無変更 or 訂正
.	.			
.	.			
31	31	同じ ID 番号	座標, 属性の差分抽出	無変更 or 訂正
なし	32		追加要素	
なし	33		追加要素	

このように、ツールで元データと更新データから、追加、訂正、削除、無変更の各要素を抽出する。抽出されたデータの追加、訂正要素は、検査対象を決める情報となる。

3.3. 差分抽出ツールの処理プロセス

本研究は、更新時に於ける品質検査を全更新データに対して行うのではなく、元データに加除訂正したデータのみに行うべきであると主張している。この加除訂正データすなわち追加要素、訂正要素、削除要素を抽出するのがツールである。更新時に於ける差分抽出の概要を図5に示す。

なお、このプログラムは、Vivid Solutions社が開発したJavaのオープンソース・デスクトップGISのJumpを活用した。このプログラムには、異なるレイヤー間の座標列が一致しているものを検出する機能がある。そこで、Jumpをベースに差分抽出に必要な機能を追加してツールとした。なお、訂正要素と判定するためには、元データと更新データの対

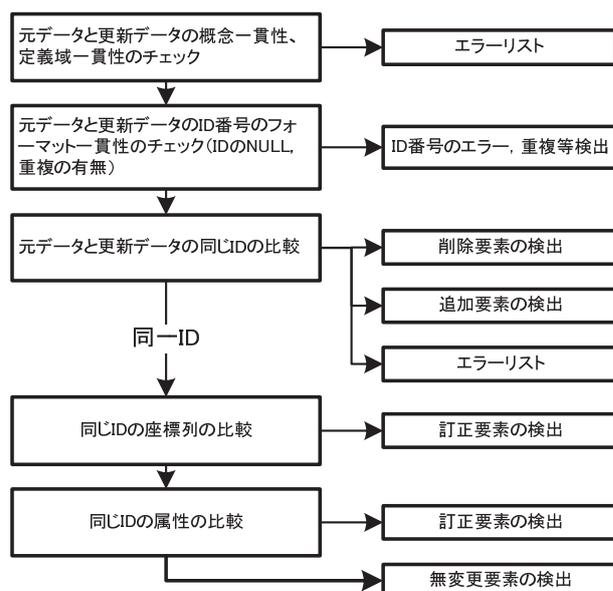


図5 差分抽出の概要

になる座標値が移動していなければならないが、その移動量を現地寸法で設定できるようにした。

4. 本ツールを用いた分析例

この節では、分析に利用したデータ、分析内容、分析の結果を述べる。

4.1. 分析に利用したデータ

事例分析は、町田市の平成17年度町田市共通基盤データの筆界、家屋を元データとし、平成18年度の筆界、家屋を更新データとして行った。

平成18年度の筆界と家屋の更新作業は、それぞれ別の受託者が行った。また、加除訂正は、家屋の改築、増築、筆界の分合筆、区画整理等を変化した要素の元データを削除し、次に、新たに家屋、筆界を追加要素として作成してある。また、加除訂正した筆界に隣接する筆界（以下周辺筆界という）と、加除訂正の筆界との整合を取るための修正は、訂正要素として作成した。同様に、加除訂正した家屋に隣接する家屋（以下周辺家屋という）と、加除訂正した家屋と重なり等を解消する場合は、訂正要素として作成してある。また、加除訂正を行わない無変更要素のIDおよび座標は、変えないことを条件として作成してある。

更新データで追加、削除、訂正の真値が分かる例

は少ない。しかし、筆界と家屋は、土地台帳、家屋台帳で追加と削除を管理しているため、ツールが正しく動作しているかを検証するために採用した。

4.2. 分析内容

分析に使用した更新データは、受託者の社内検査でエラーが0%と報告されたもので、この社内検査結果と差分抽出結果を比較検証することとした。

受入検査では、まず図5に示すツールで、筆界、家屋のID番号の論理性チェックを行い、次に追加要素、削除要素、訂正要素を抽出した。

台帳には、関連変化を生ずる周辺筆界との整合性を取るための修正や、周辺家屋との重なり解消のための位置補正等によって生ずる訂正要素の情報がない。

そこで、新たに元データと更新データの同じID番号の座標、属性を比較し、移動を判定する移動判別プログラムを作成し、訂正要素のツールの有効性を検証することとした。また、座標が移動したと判定する移動量は、特記仕様書により10cm以上に設定した。

4.3. 分析結果

4.3.1. ID番号の論理性

ツールを元データと更新データに適用し、論理一貫性の書式一貫性、概念一貫性、定義域一貫性のチェックを行った。その結果、更新データの筆界、家屋の文字型の誤りを183個抽出した。

次に、更新ファイルの文字型を修正した上で、再びツールを稼働させた。その結果、追加筆界19筆のID番号の重複を見つけた。このように、ツールは、受託者の社内検査で見逃されていた文字型のエラーを容易に見つけ、ID番号の付番の間違いを確実に発見・修正できる。こうした論理性チェックの機能は、修正作業時間の短縮ができるという点で、実用上大きな意味がある。

4.3.2. 筆界の差分抽出

更新データの削除筆界、追加筆界、訂正筆界をツールで抽出し、それぞれの合計数値を算出した。表2

は、土地台帳の筆界数と差分抽出の結果及び受託者の検査結果を比較したものである。なお、追加訂正筆界数は、追加筆界数と訂正筆界の合計である。ただし、土地台帳の筆界数には、訂正筆界数の属性の修正数は管理されているが、図形の修正は管理されないため表では-としている。

表2 抽出筆界数

項目	土地台帳の筆界数	差分抽出の筆数	受託者の筆界数
総筆界数	242,560	242,560	242,560
削除筆界数	3,393	3,393	3,393
追加訂正筆界数	-	14,449	14,677
追加筆界数	6,141	6,141	6,141
訂正筆界数	-	8,308	8,536

そこで、差分抽出の訂正筆界数の結果が正しいかどうかを判断するため、移動判別プログラムで訂正筆界数を検証した。検証の結果、訂正筆界数は、8,308筆であることを確認した。一方、受託者の訂正筆界数は8,536筆で、両者の訂正筆界数に228筆の差が出ている。この原因を調査するため、受託者より8,536筆のIDリストの提供をうけ、これを基に228筆を特定し、移動判別プログラムで228筆が変更されていないことを確認した。追加訂正筆界数14,449筆を母数とすると228筆は、過剰が1.58%である。母数の不確実性はデータ適用範囲に影響を与え、品質評価結果の信頼性を損なうものとなる可能性がある。

原因調査の結果、受託者の更新データを作成する機器（以下生産機器という）は、周辺筆界の整合のために座標列を修正する場合、オペレータが手動で属性に修正履歴のフラグ立てることになっている。今回のケースは、周辺筆界を修正するため修正履歴フラグを立てたが、何らかの理由によって座標列を修正せずに終了したため、上記の差が出たものであった。ツールの訂正筆界抽出は、元データと更新データの座標を比較しているため、このような、人為的なミスも確実に見つけることができる。

4.3.3. 家屋の差分抽出

更新データの削除家屋、追加家屋、訂正家屋は、

ツールで抽出し、それぞれの合計数値を算出した。表3は、家屋台帳の家屋数と差分抽出の結果及び受託者の検査結果を比較したものである。なお、追加訂正家屋数は追加家屋数と訂正家屋数の合計である。ただし、家屋台帳の家屋数の家屋台帳には、訂正家屋数の属性の修正数は管理されているが、図形の修正は管理されないため表では－としている。

表3 抽出家屋数

項目	家屋台帳の家屋数	差分抽出の家屋数	作業機関の家屋数
総家屋数	111,675	111,675	111,675
削除家屋数	6,111	6,111	6,111
追加訂正家屋数	－	13,058	13,036
追加家屋数	12,775	12,775	12,775
訂正家屋数	－	283	261

そこで、差分抽出の訂正筆界数の結果が正しいかどうかを判断するため、移動判別プログラムで訂正家屋数を検証した。検証の結果、訂正家屋数は、283棟であることを確認した。また、受託者の訂正家屋数は、261棟で両者の訂正家屋数に22棟の差が出ている。

この原因を調査するため、261棟のIDリストの提供を受け、これを基に22棟を特定し、移動判別プログラムの結果から22棟の座標が移動していることを確認した。追加訂正家屋数13,058棟を母数とすると22棟は、追加訂正家屋数13,058棟を母数とすると22棟は、漏れが0.168%である。母数の不確実はデータ適用範囲に影響を与え、品質評価結果の信頼性を損なうものとなる可能性がある。

原因調査の結果、受託者は、生産機器で家屋が重なった場合に位置補正を行っている。生産機器は、空間検索機能と集計機能があり、元データと修正後のデータの座標比較を行い、位置補正のため移動させた家屋を訂正家屋として集計している。また、生産機器は、固有のフォーマットで管理されている。一方、元データと更新データのデータ形式は、ともにShape形式である。このため、元データを生産機器用に変換し、更に修正作業後のデータを納品用にと2回変換している。この2回のデータ変換の過程で変換誤差を生じさせている可能性が高い。生産機

器は、生産機器内の元データと更新データを比較しているが、この比較では変換誤差を見つけることができない。このため、変換誤差が見逃されている。

4.3.4. 効果

これまで、加除訂正を行った範囲を特定することが困難であったが、このツールを使うことによって加除訂正を行った要素を抽出することができるようになった。これによって、加除訂正範囲の誤差修正を容易にすることと、加除訂正範囲の品質検査ができるようになったことは、意義が大きいものと考えられる。

また、ツールを使った処理時間は、IBM社製2003サーバー、Pentium4、メモリー4GBで処理した結果、筆界について約25分、家屋について18分であった。

5. おわりに

本研究は、更新データの品質検査に焦点を当て、ID番号の規則性を利用して更新データの構成要素である削除、追加、訂正、無変更の各要素を抽出するツールの開発を行った。そして、そのツールの評価は、資産税の地番図、家屋図を利用して行った。

本研究で得られた主な成果は以下の通りである。

- ① 受託者の品質検査は、更新データ全体に対してプログラム検査を行い、元データに変更を加えない無変更要素のエラーを目視で除外しているのが実態であった。ツールを活用すれば、更新データの元データに対する変化要素を自動で抽出でき、抽出ミスの防止と、作業時間の短縮に効果がある。
- ② 受託者の品質検査は、生産機器内データの全データを検査しており、元データと納品物である更新データを直接比較していない。ツールでは、元データと更新データを直接比較しているため、生産機器で元データから更新データを作成されるまでに行われた種々の座標変換誤差を訂正要素として捉えることができる。

今後の課題として、本研究は、ID番号をキーとして差分抽出をしている。しかし、ID番号の管理

をオペレータが管理することは、品質管理上課題が残る。今後は、空間データ生産機器のID番号を自動的に附番するロジックの研究に取り組みたい。

また、本研究では2.1.2の関連変化における3つのパターンのうち(1)、(2)は解決したが、(3)の重なりを許す地物間での分断と集合が解決していないことがあげられる。分断で複数要素にした場合、集合によって複数要素を1つの要素にした場合の処理に課題が残る。

不連続な複数の要素を1つの要素とすることには、グラフィックグループ(ESRI社ではマルチパート)を活用する方法が考えられる。この機能を利用して、要素の分断、集合のデータを作成し、そのデータに対応するツールを開発することによって解決したいと考えている。そして、複雑な関連変化の生ずる地形図に適用し、ツールをさらに汎用性の高いものにしていくよう研究を深めたい。

参考文献

- 太田守重(1998) GISのための時空間スキーマ,「GIS - 理論と応用」, 7(1), 37-44.
- 門脇利広・中南清晃・小荒井衛(2001) GIS基盤情報(数値地図2500)の時系列管理手法に関する研究,「GIS - 理論と応用」, 9(2), 61-66.
- 国土交通省(2004)『国土交通省公共測量作業規程』, 社団法人日本測量協会.
- 国土地理院(2005)『平成6年国土基本図図式』, 社団法人日本測量協会.
- 国土地理院(2005)『品質要求, 評価及び報告のための規則(JPGIS2.0)』, 国土地理院.
- 星野秀和・久保紀重・飯村威・田宮彰弘・飯田剛輔・平井政二・大伴真吾(1999)空間情報と時系列情報の統合化に関する研究,「地理情報システム学会講演論文集」, Vol.8/1999, 83-88.
- 平田更一・柳田聡(2008)地理情報標準における論理一貫性の検定プログラムの開発(2)「日本写真測量学会平成20年度学術講演会予稿集」2007, 169-172.
- 渡邊孝三・岡部篤行・中村秀至(2008)GISデータの位相検査の交差判定ツールの開発とその適用による有効性の検討,「GIS - 理論と応用」, 16(1), 59-68.
- (2008年5月21日原稿受理, 2010年6月21日採用決定, 2010年7月21日デジタルライブラリ掲載)