

紙面地盤情報電子化技術を利用する高密度  
地盤情報DB構築手法に関する研究

茨城大学工学部都市システム工学科  
講師 村上 哲

平成19年9月

## 目 次

第1章 序論	1
第2章 紙面の地盤情報を電子化する方法と解決すべき課題	3
第3章 既存の紙面地盤情報を効率的にデジタル化する技術開発	5
3.1 はじめに	5
3.2 土質柱状図の特徴と認識効率化	6
3.3 紙面地盤情報の電子化フロー	7
3.4 既存面地盤情報の電子化技術開発	8
3.5 既存面地盤情報の電子化技術の適用	12
第4章 電子化された情報の信頼度の判定方法	14
4.1 はじめに	14
4.2 対象地域と地盤情報DB	15
4.3 土質に着目した鉛直方向のN値のトレンドモデル	15
4.4 調査結果との比較による検討	18
4.5 電子化された情報の信頼度の判定方法の提案	19
第5章 紙面地盤情報電子化技術を利用する高密度地盤情報DB構築手法の提案	20
5.1 構築手法の提案	20
5.2 今後の課題と方向性	21
謝 辞	23
参 考 文 献	24
助成研究者紹介	25

## 第1章 序論

柱状図や各種土質試験をデジタル化した地盤情報データベース（以下、地盤情報 DB と略称する）が構築され、建設時における地盤調査計画、構造物設計などの業務支援や地域の防災および環境保全のための貴重な情報源と位置づけられている。わが国における地盤情報 DB は、とりわけ主要都市部において構築・公開が進んでおり、これは近年の情報電子化による DB への移行の容易さと地盤調査の機会が多さなどに依存している。これに対し、地方都市においては、建設頻度の低さから、調査地点が疎な地盤情報 DB と成らざるを得ない。

しかし、過去の地盤調査結果や地盤柱状図集といった紙面情報が存在することが多く、これらを効率的かつ十分な信頼度を持ってデジタル化することができれば、地盤情報収集が困難な地域の地盤情報 DB 構築の大きなドライビングフォースとなりうるとともに、地域の建設業務支援や、防災・環境保全のための貴重な情報整備を支援できる。また、社会的経済的な点と DB の整備状況から見た地域格差の是正も期待できる。

そこで、本研究では、地盤調査報告書や地盤柱状図集といった既存の紙面地盤情報を、デジタルスキニング技術を利用し、個々の情報を効率的に電子化するとともに、集積したデジタル地盤情報から個々の情報の信頼度を地盤統計学的手法に基づき判定するとともに、高密度地盤情報データベースを構築する手法を開発することが目的である。

本研究の主要な部分は、①地盤調査報告書や地盤柱状図集といった既存の紙面地盤情報を効率的にデジタル化する技術開発、および、②電子化された情報の信頼度の判定方法の確立である。

これらを達成するために、まず、既存の紙面情報の印刷形式について調査を行い、スキャニングによる情報認識のためのカテゴリー分けを行う。次に、情報化すべき項目について優先順位を決定し、カテゴリー項目に応じた電子化プロセスについて検討を行う。この検討では、ある程度規格化されている柱状図書式を最大限に活用すること、および、効率化と信頼度の2軸と処理時間の3つの軸を念頭におく電子化プロセスを構築する。以上により、①の技術開発を行う。

次に、②では、地盤情報 DB に入力された地盤情報の信頼度の判定方法の確立について行う。データベース化された個々の情報の信頼度は、情報源である調査そのものの信頼度、紙面情報の記載ミス、本電子化プロセスにおける誤認識などに依存する。この信頼度はデータベース全体の信頼度を左右させるため、本研究では、個々の情報の信頼度評価手法の1つとして、地盤統計学に基づく手法の適用性を検討する。これにより、信頼度の低いデータの点検・修正、破棄といったプロセスを組み込むことができると考える。

以上の2つのパートを有機的に結合させることによって紙面柱状図を効率的にまた大量に電子化することによる高密度地盤情報データベース構築手法を提案する。

## 第2章 紙面の地盤情報を電子化する方法と解決すべき課題

紙面の地盤情報を電子化する方法としては、現在、紙面の情報を直接、人の手によって入力する方法に頼らざるを得ない。この方法は、逐次、入力者が確認しながら作業を進めるので、入力データは、元の情報とほとんど相違無いものが得られると考えられる。一方、この方法によると、入力すべき紙面情報の量に依存し、例えば、国や地方自治体が管理保管している情報を入力する場合を想定すると、多大な時間と費用を要することとなる。このため、紙面情報を電子化し、その情報資源を有効活用する重要性は認知されているものの、財政面の問題から着手できないのが一般である。

一方、昨今のコンピュータ技術の発展とパーソナルコンピュータの普及により、これらの機材、周辺機器の低価格化、高性能化が進んでいる。文書情報をイメージスキャナにより画像ファイルとして保存し、整理することは、近年、日常的に行えるようになってきた。また、単に画像ファイルとして保存するだけでなく、その画像データから、光学文字認識（OCR：Optical Character Recognition）を実行し、文字情報を付加した PDF（Portable Document Format）ファイルとして保存できるようになっている。文書情報をそのまま、電子化された画像ではない文書情報として完全復元するまでには至っていないが、かなり

の文字を認識できるようになってきた。この情報化技術を、これまで多大な時間と費用を要していた紙面の地盤情報の電子化に利用できるのではないかと考えたのが、この技術開発の発想である。

このデジタルスキニング技術を援用し、紙面の土質柱状図を、例えば、JACIC 仕様のボーリングデータ交換用フォーマットの XML 形式で電子化するところを考えてみる。この時、以下の課題が挙げられる。

- ・ 調査地点、土質区分と深度、標準貫入試験結果など多様な情報が記載されている土質柱状図から、適切に各読み取り位置を決定しうるか。
- ・ 手書きからワープロ打ちまである文字を正確に認識することができるか。
- ・ 画像化する際、高解像度にすることによって文字認識の精度向上が期待できるが、そのトレードオフとして、処理時間やコンピュータの性能に依存してしまう。利用に資する解像度で文字認識が可能か。

さらに、高機能高性能により、紙面情報を完全に電子化することが可能となっても、費用の面が解決できなければ、この方法は時期早々という判断をせざるを得ない。したがって、紙面情報を電子化する技術を、導入しやすい投資の範囲で開発することと位置づけた。

この目的を達成するために、

- ① 地盤調査報告書や地盤柱状図集といった既存の紙面地盤情報の書式や情報の性質に着目することによって、効率的に高い整合性を持ってデジタル化する技術開発と、
- ② 地盤の空間的な連続性や土質特性といった地盤工学的な知見を利用する電子化された情報の信頼度の判定方法と地盤情報 DB への入力フローの確立

に取り組むことが必要である。これらの2つのパートを有機的に結合させることによって紙面地盤情報を効率的にまた大量に電子化できると考える。以下では、この技術開発において、もっとも核心となる①の既存紙面地盤情報のデジタル化技術について、詳しく述べる。

## 第3章 既存の紙面地盤情報を効率的にデジタル化する技術開発

### 3.1 はじめに

土質柱状図には、土質区分と深度、標準貫入試験結果など多様な記載情報が記載されており、この土質柱状図画像から適切に各種情報を読み取ることが必要である。土質柱状図の様式は、昭和62年の「地質調査成果資料整理要領(案)」<sup>1)</sup>以降、ほぼ統一されているが、それ以前は、調査業者独自の様式が使用されている。すなわち、土質柱状図情報の個々の読み取り位置を一般化することは不可能に近いと考えられる。しかし、一冊の報告書を手にとってみると、これはある調査業者によってとりまとめられたものであり、報告書内に収められている土質柱状図の様式は、統一化されている。また、国土交通省地方整備局、都道府県、市町村などの管理単位で考えると、地盤調査を実施している業者には限りがあると予想されることから、ここで開発する技術を用いる際には、土質柱状図の様式を数種

類に限定しうるものと思われる。また、柱状図のヘッダ部に相当する調査名、事業工事名などは、同一の報告書内では同じである。さらに、深度方向の標尺、標高、深度、柱状図、N 値など各種情報は、同一の報告書に収められていたとしても、調査位置によっては異なるのが通常であるが、ここでの記載箇所は、横一線に並べて記載されるのが一般である。このような柱状図特有の性質を利用した読み取り位置決定が有効であると考ええる。

本章では、土質柱状図の特徴を整理し、その特徴を踏まえた上で、デジタル化する手順を構築し、それに沿ったソフトウェアの開発を行ったことについて述べる。

### 3.2 土質柱状図の特徴と認識効率化

土質柱状図の記載事項としては、紙面の上部に位置する調査位置、調査件名、発注機関などの標題情報と、地表面から深さ方向に関する土質岩種区分、色調、観察記事、標準貫入試験などのいわゆるコア情報から構成される。

また、認識すべき情報とその種別、および、認識しなくてもよい項目名や罫線などを区別分類することによって、認識率の向上が図られると思われる。例えば、標高値には数字と小数点、符号の数値しか現れない、土質区分には漢字などの文字で記載されているなどである。

次に、読み込んだ情報で、相互に整合性を持つべき情報（例えば、標高値、深度、層厚の数値の変化量、土質名称には限られた文字、“年度”という文字は無いが“粘土”はある、など）に着目し、認識結果に対して整合性をチェックすることもソフトウェアとして処理できると考える。



### 3.3 紙面地盤情報の電子化フロー

以上のような土質柱状図の特徴を生かした認識を実施し、項目ごとの情報を電子化することができる。このような考えに基づいて、基本的な電子化の流れを示すと図 3.1 のような手順となる。

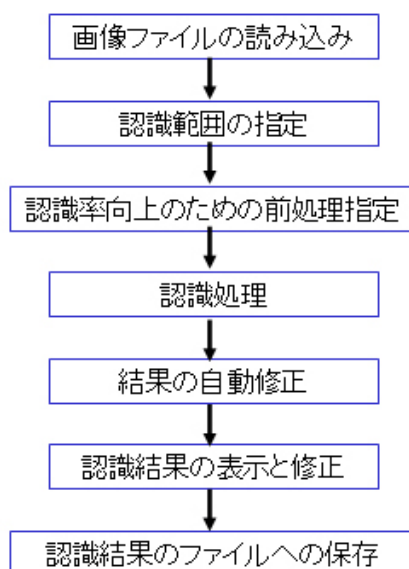


図 3.1 電子化の流れ

デジタルスキャナを用いて、紙面の柱状図を電子ファイルとしての画像データに変換する。次に、柱状図に記載されている標題情報および土質区分、標準貫入試験などのコア情報が画像ファイルの中で、どの位置を占めているかを指定する。その後、認識向上のための前処理を実施する。この前処理では、画像データの傾きや濃さに関する画像全体の補正、数値・文字などの認識文字タイプ、枠線の有無や特に標準貫入試験における記載ルールの区別、認識開始位置の指定を行うことによって、次に行う認識処理の読み取り精度を向上させる。認識処理では、既存の文字認識エンジン（「エー・アイ・ソフト 読ん de!! ココ活字文書OCRライブラリ Ver.5」）を使用）により、画像を数値や文字として認識する。更に、変換された数値文字情報の整合性を高めるために、土質柱状図の特徴を活かした修正を行う。この結果を認識結果とし、画面上で画像データと認識結果を比較し確認および修正を行うことによって、最終的な電子ファイルとして保存する。

以上の手順により、紙面地盤情報の電子化を実施できるアプリケーションソフトウェアを開発する。

### 3.4 既存面地盤情報の電子化技術開発

上述した土質柱状図の記載特徴に着目した電子化手順を実施できるアプリケーションソフトウェア「柱状図認識システム SRS」を開発した。このソフトウェアは Window 上で動作するものであり、一般的に利用されているパーソナルコンピュータでの利用を想定している。

まず、イメージスキャナにより紙面柱状図をビットマップ形式の画像ファイルとして所定のフォルダに保存する。通常、1つの報告書には複数の柱状図情報が収められており、オートシートフィーダーや連続読み取り装置が付属しているイメージスキャナを使うと効率的である。また、柱状図様式が同一のものをひとつのフォルダに収めることによって、認識位置の抽出は最初の1枚目に指定することによって、2枚目以降この作業は不要となるように工夫した。

ソフトウェアを起動し、画像ファイルが保存されているフォルダを指定し、一覧表示する。ひな型となる1つのファイルを選択して、認識対象範囲の指定を行う。図 3.2 はこの時の操作画面を示している。認識対象範囲は抽出したい情報に応じて複数指定することが可能である。なお、ここで用いたひな型となる画像ファイルにおいて、認識範囲の座標指定や認識率向上のための情報設定が行われる、また複数ファイルの位置あわせのため、画像領域の左上隅の点（基準点）が指定され、画像として表示された柱状図の罫線方向とパソコン画面の上下左右方向が一致するように補正する機能を持っている。

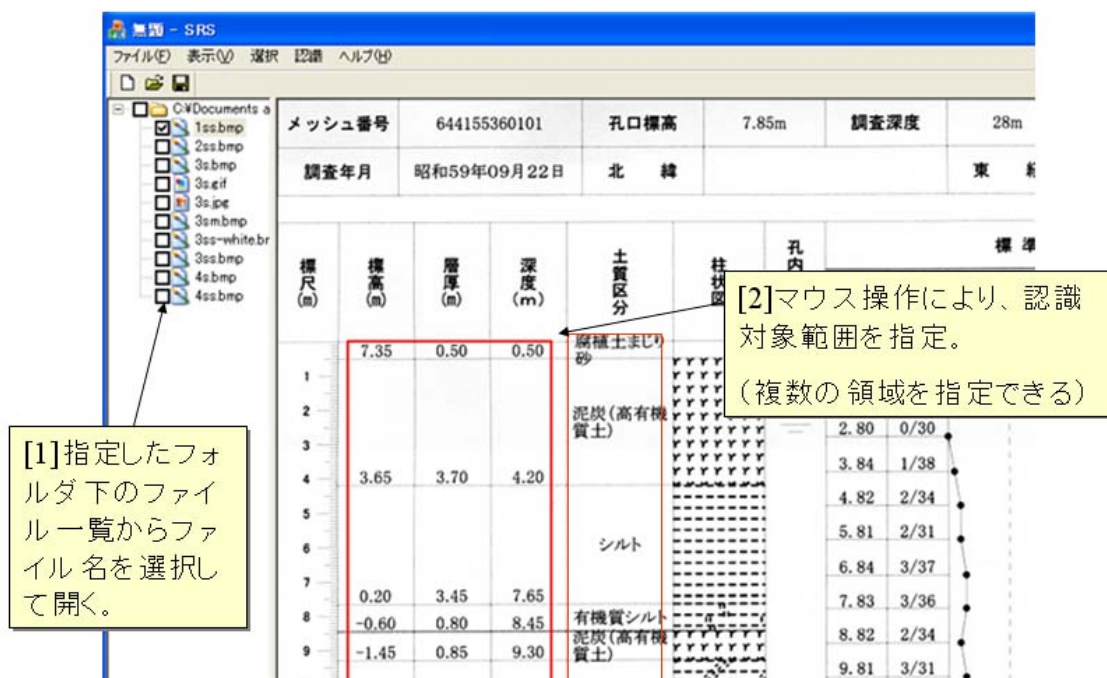


図 3.2 画像ファイルの表示と認識対象範囲の設定

次に、認識率向上のための前処理として、まず、画像全体の補正を行う。複数の柱状図を処理することを念頭に置くと、電子化された柱状図画像データファイルは、厳密には、パソコン画面の上下左右方向と、柱状図表の罫線が一致せず、各々異なる傾きを有する。この傾きを補正するために、表の四隅の角および罫線に着目し、傾き補正が行われる。次に、読み取られる紙面柱状図情報の用紙の種類や印字の程度は様々であるため、これを統一するために濃度補正を行う。以上、2つの補正と罫線強調、枠線と数値を含めた文字を分離する前処理を行うことにより認識度を高めている。(図 3.3 参照)

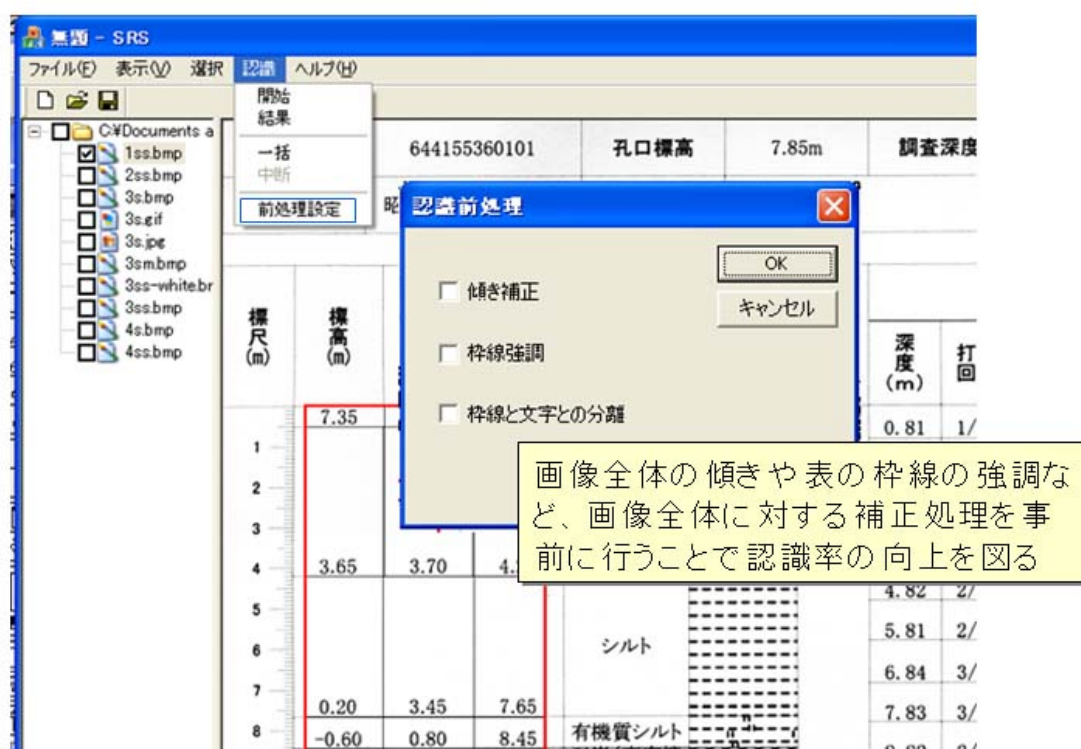


図 3.1 認識率向上のための前処理 (画像全体の補正)

さらに、認識タイプ・文字種別等の指定を実行することで、認識率向上のための前処理が終わる。また、抽出領域ごとに、表かテキストかの認識対象の種別と、数字記号かそれ以外の文字を含むものかなどの文字種別の指定を行うことで、さらに読み取り精度を向上させることができるようになる。(図 3.4 参照)

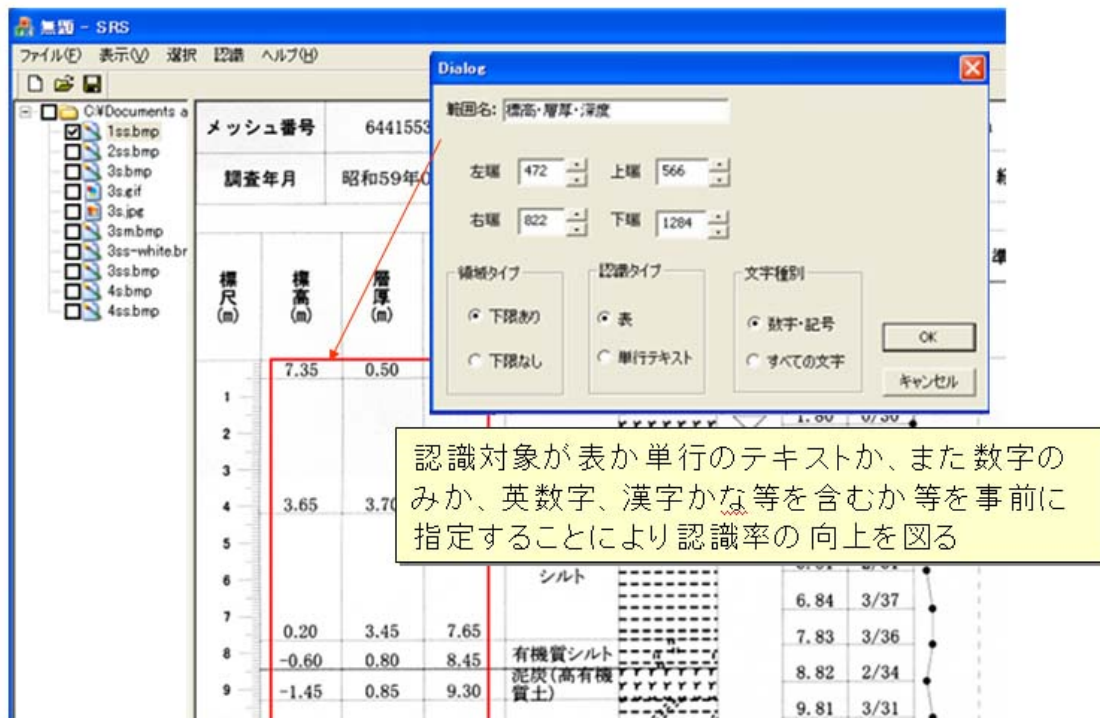


図 3.4 認識率向上のための前処理（認識タイプ・文字種別等の指定）

以上の前処理をひな型の画像ファイルで行った後、全ての画像ファイルに対して、指定領域の認識が開始される。認識した結果は、次のチェックを行い、自動修正や修正候補の表示が行えるようになっている。（図 3.5 参照）

- (1) 表内の数値の整合性チェック
- (2) 記号や文字等の整合性チェック
- (3) 土質区分のように、現れる単語の候補が決まっている場合は、認識結果が候補の中に含まれるかのチェック・類似度の高い単語への置き換え等

認識終了後、元画像と認識結果を並べて表示し、認識結果の修正が行えるようになっている。この時、認識結果中で確信度が低いものは、赤い文字で強調表示される。また、エディットボックス上で文字を選択すると、画像上の対応する箇所が強調表示される。これらの操作により、認識結果を最終確定する。

最後に、図 3.6 に示すように、認識結果を項目ごとのテキストとしてファイルへの保存できる。

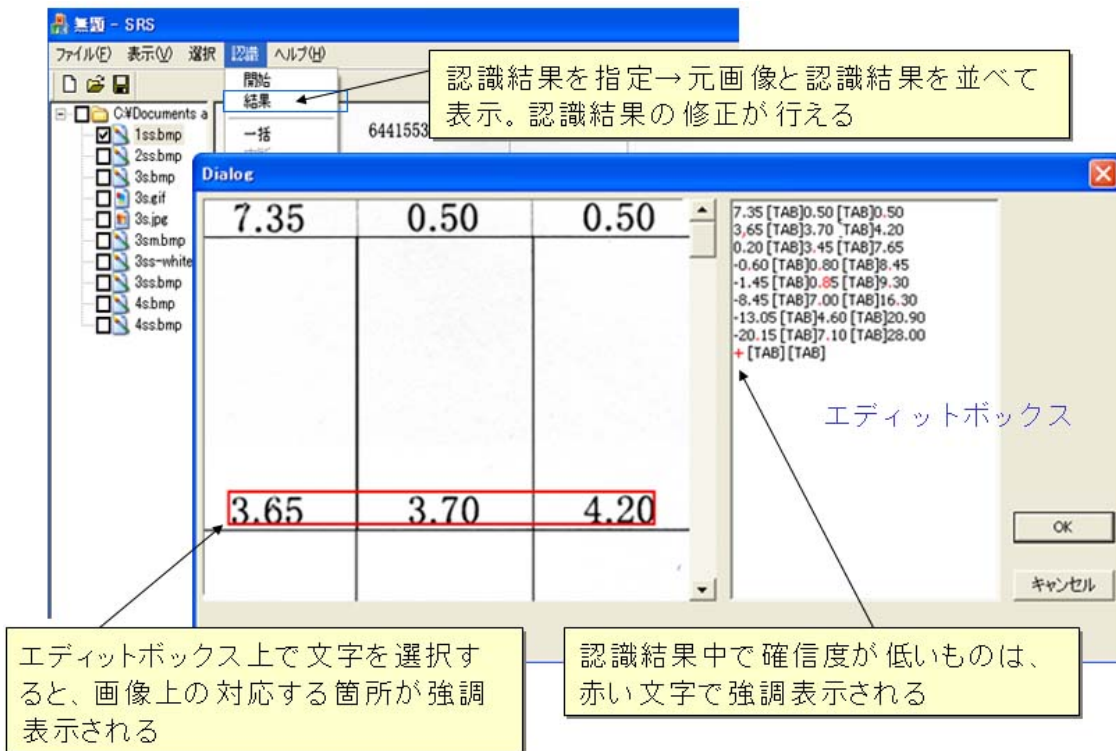
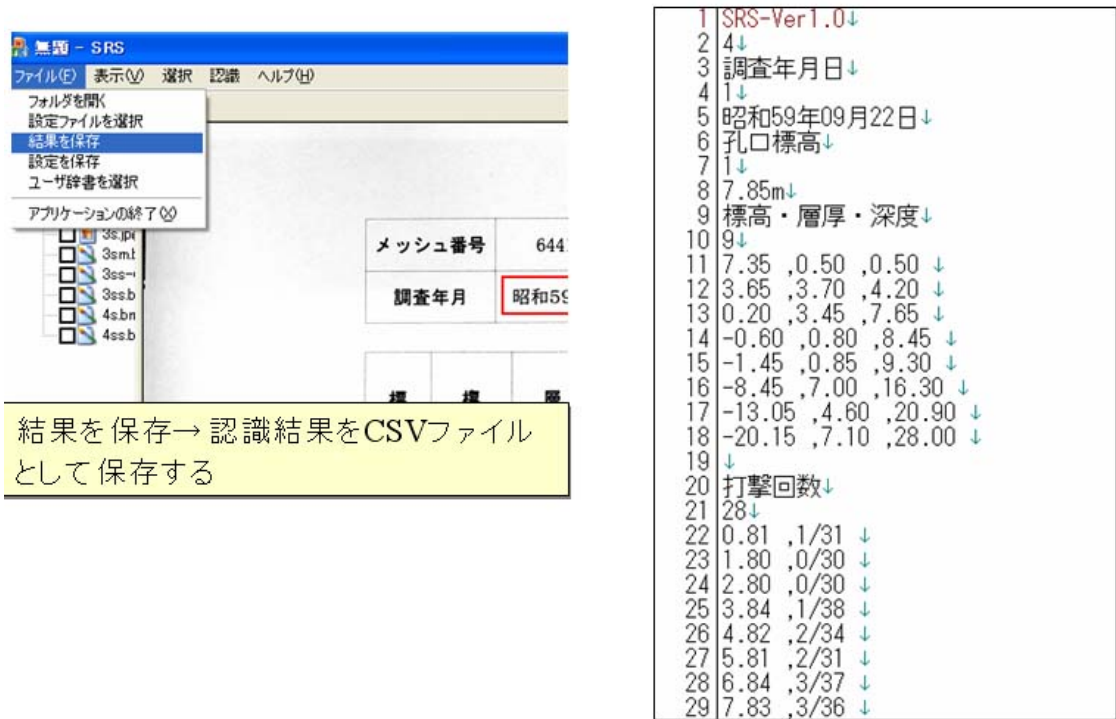


図 3.5 認識結果の表示と修正



CSVファイル例

図 3.6 認識結果のファイルへの保存

### 3.5 既存面地盤情報の電子化技術の適用

本研究により開発した土質柱状図の記載特徴に着目した電子化手順を実施できるアプリケーションソフトウェア「柱状図認識システム SRS」について、既存の紙面地盤情報を用いて電子化する試みを行い、その仕様について明らかにした。用いた紙面地盤情報は、長崎県土木部技術情報室により提供いただいた既存の地盤調査報告書（既に画像データとしてPDF化されたもの）、および、新潟県県民生活環境部より提供いただいた地盤調査報告書である。表-3.1 は、長崎県より提供いただいた地盤調査報告書の内容を調査年度、本数、体裁についてまとめたものである。

表 3.1 入手した紙面地盤情報の特徴

調査年度	柱状図本数	コメント
S40	6	手書き・汚れ
S42	3	手書き・一部汚れ
S44	5	手書き・文字かすれ
S46	4	手書き
S48	13	手書き
S50	8	手書き
S51	4	ワープロ
S52	26	手書き
S54	13	手書き
S56	24	手書き
S57-01	14	ワープロ
S57-02	26	手書き
S60	11	手書き
S62	7	手書き
H01	4	ワープロ
H03	6	ワープロ
H05	5	ワープロ
H07-01	11	ワープロ
H07-02	2	ワープロ
H09	3	ワープロ
H11	6	ワープロ
H13	6	ワープロ
H15	7	ワープロ
H17	7	PDF

表 3.1 を見ると、昭和時代まで手書きの柱状図が多いことが分かる。ワープロ打ちの印字による柱状図情報は平成元年以降のようである。コメントにおいて昭和 40 年代で汚れとあるのは、これは青焼きによる資料を意味している。これらの多様な種類の紙面柱状図情報からいくつかをピックアップし、認識できる紙面地盤情報について検討を行った。その

結果を表 3.2 に示す。

表 3.2 柱状図認識システム SRS の認識性能

動作環境	適用
OS	Windows XP Second Edition (日本語版)
メモリ	256MB 以上
対象文書	
認識可能な紙面情報	活字で記載されたボーリング柱状図 (手書き数字も対応可)
画像化する際のファイル形式	BMP、J P E G、T I F 形式
画像解像度	300dpi 以上
紙面情報の文字サイズ	5 ~ 4 5 ポイント (ただし、5 ~ 6 ポイントは画像解像度が 600dpi 取り込み時)

本ソフトウェアは、Windows XP Second Edition (日本語版) のオペレーティングシステム、メモリは 256MB で動作可能である。この点は、一般に利用されているパーソナルコンピュータで十分動作可能であることが確かめられた。

認識できる文書については、活字で記載されたボーリング柱状図であり、一部手書き数字も対応できるようになっている。画像化する際のファイル形式は、非圧縮のファイル形式 BMP と T I F の 2 種類と、よく知られている J P E G である。画像化する際の解像度は、300dpi 以上 (ただし、5 ~ 6 ポイントは画像解像度が 600dpi が必要) であり、汎用のスキャニング装置でも十分対応できるものとなっている。認識画像はカラーでなく、モノクロで行うので、この解像度でも十分対応できる。

さらに、認識精度を向上させるためには、以下の条件を満たすものが望ましいことが分かった。

- ・文字サイズが 10 ~ 11 ポイント
- ・300 ~ 400 dpi で取り込み
- ・コントラストがはっきりしている (白黒 2 値画像よりグレースケールが良い)
- ・傾きやゆがみが少ない、汚れやごみが少ない

## 第4章 電子化された情報の信頼度の判定方法

### 4.1 はじめに

データベース化された個々の情報の信頼度は、情報源である調査そのものの信頼度、紙面情報の記載ミス、本電子化プロセスにおける誤認識などに依存する。この信頼度はデータベース全体の信頼度を左右させる。個々の情報の信頼度評価手法の1つとして、統計的手法に基づく手法の適用性を検討する。これにより、信頼度の低いデータの点検・修正、破棄といったプロセスを組み込むことができると考える。

多くの地盤調査とその記録である土質柱状図には、深度方向の土質区分と標準貫入試験が実施されている。地盤情報データベースに格納される個々のデータは、深度、土質区分、標準貫入試験結果であるN値は独立に入力される。先に示した電子化技術においても同様である。しかし、軟弱な粘土であれば一般的に低いN値であり、一方、砂礫層ではそれに比較して高いN値となる。すなわち、独立に入力された情報であるが、地盤の特性を考慮すると、深度、土質、N値に相関が見えてくる可能性がある。この相関を明らかにすることができれば、深度、土質、N値の組み合わせにおいて、異常と思われる記録を抽出する



ことが可能となると考える。そこで、本研究では、まず、既存のデータベースを用いて、深度、土質、N値の組み合わせに相関が現れるかどうかを検討する。次に、その相関を用いた、地盤情報の信頼度判定方法を提案する。

## 4.2 対象地域と地盤情報DB

本研究で対象とした地域は図 4.1 に示す北海道札幌市とその周辺である。この地域の地盤に関する情報は、地盤工学会北海道支部が発行している北海道地盤情報データベース 1)があり、本研究ではこれを用いた。この地盤情報 DB には、地盤調査位置と調査で得られた土質区分、標準貫入試験の結果などが収められている。この地盤情報 DB に収められている全調査地点数約 13,000 本から、深度とその土質および N 値がある結果を全て抽出し、91,454 組の深度、土質、N 値のデータセットを作成した。なお、このデータセットを作成するに当たり、N 値は全て 30cm 貫入の打撃回数に換算している。このデータセットからさらに土質種別ごとに分類したデータセットを作成した。

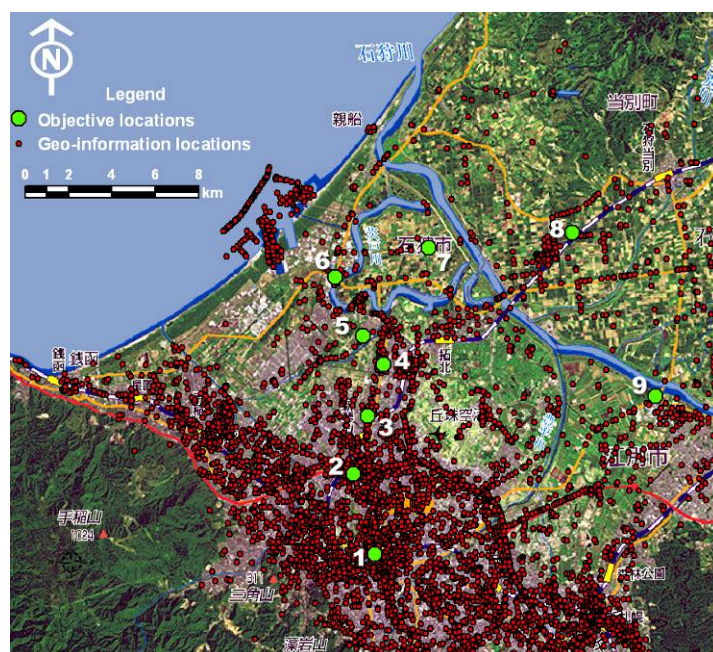


図 4.1 対象領域と地盤調査地点

## 4.3 土質に着目した鉛直方向のN値のトレンドモデル

図-4.2～4.4 に示した 3 つの図中の小さいプロットは、作成したデータセットの内、砂、シルト質砂、砂質シルトについて深度と N 値の関係を示したものである。ここで、 $N \geq 50$  のデータは除いた処理を以降で行っている。3 つの土質とも比較的浅い位置から N 値が広

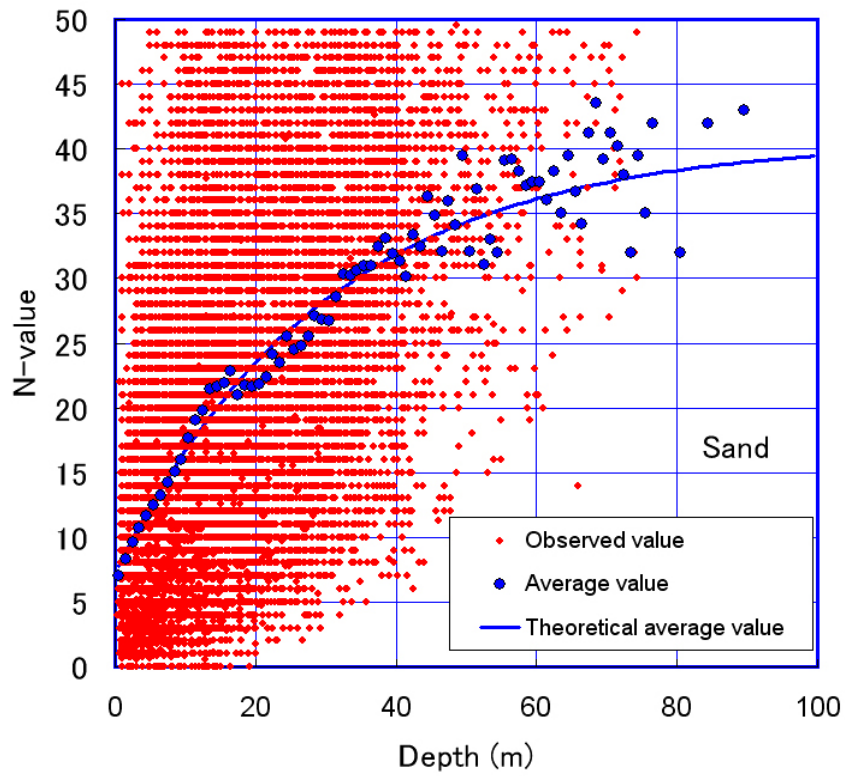


図 4.2 砂における N 値、区間平均 N 値、理論平均 N 値と深度の関係

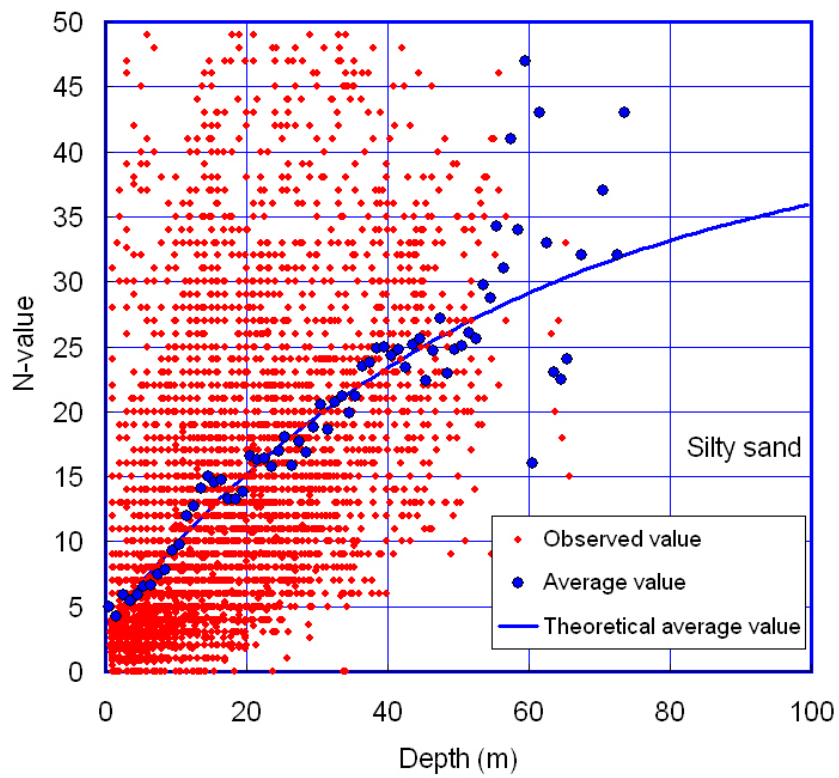


図 4.3 シルト質砂における N 値、区間平均 N 値、理論平均 N 値と深度の関係

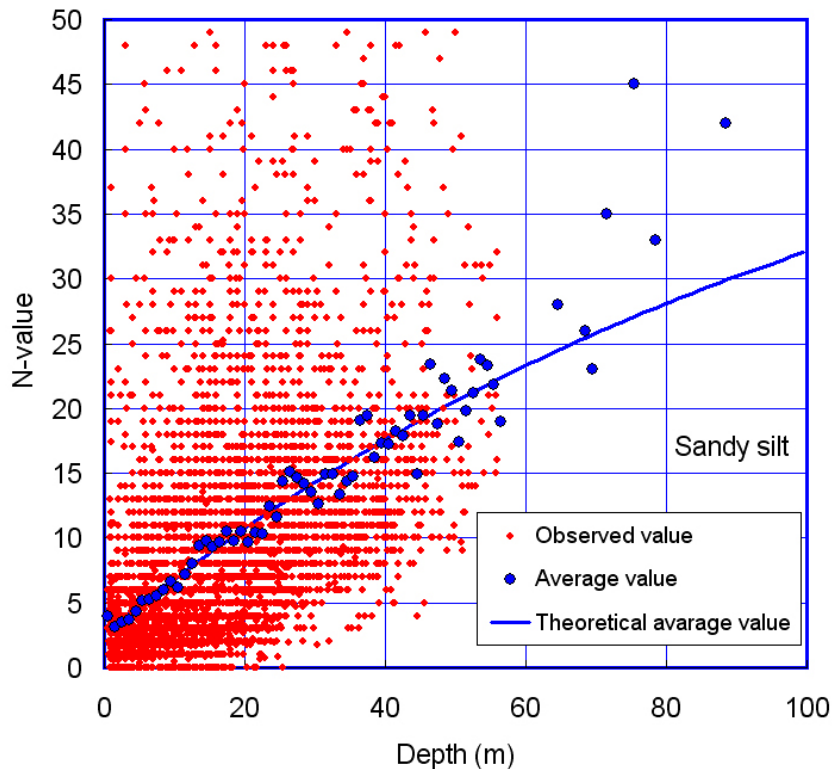


図 4.4 砂質シルトにおける N 値、区間平均 N 値、理論平均 N 値と深度の関係

く分布していることが分かる。これは、単に“砂”と分類されていても、密な砂・ゆるい砂、せん断抵抗が大きい砂・小さい砂など、多様な“砂”と多様な状態が存在することから、このように広く分布する結果となったと考えられる。

このように観測された N 値とその深度の関係をそのままプロットするとばらつきの大きい関係と見えるが、深度毎に平均値を求めてみると、ある傾向が見えてくる。図中の大きいプロットは深度 1m 区間ごとに、その区間に存在する N 値の平均値（以下、区間平均 N 値と呼ぶ）を示したものである。このように区間平均 N 値は、深度の増加に伴って次第に大きくなる傾向が分かる。これは、均質な土質で構成される地盤を想像したとき、N 値は深度とともに増加すると考えると、区間平均することによって、多種多様な状態で存在する土質の 1 つ（例えば“砂”）が平均化された状態であると解釈することもできる。すなわち、対象領域で観測される同一の土質区分における N 値の鉛直方向のトレンド成分と考える。図中の実線は、次の関数で、区間平均 N 値と距離の関係を近似したものである。

$$N = N_0 + (N_{\max} - N_0) \left\{ 1 - \exp\left(-\frac{z}{m}\right) \right\} \quad (4.1)$$

ここで、 $N_0$ ,  $N_{\max}$ ,  $m$  は土質区分によるパラメータであり、これらのパラメータは区間平均 N 値と距離  $z$  の関係から重み付き最小二乗近似により決定した。この式を用いて求められた N

値を理論平均 N 値と呼ぶことにし、この理論平均 N 値と実測結果との比較により、実際のトレンド成分と考えられるかどうかを確かめる。

#### 4.4 調査結果との比較による検討

理論平均 N 値が実地盤のトレンド成分を表現できるかどうかを確かめるため、北海道地盤情報 DB から、札幌市およびその周辺に存在する比較的深い深度まで調査された 9 地点 (図-1 参照) を選定し、理論平均 N 値と比較した。図 4.5 は、比較結果である。図中には、粘性土、砂質土といった大分類での土質区分も同時に示している。土質の変化に伴う N 値の増減について、理論平均 N 値はよく表現できている。したがって、N 値の鉛直方向のトレンドはある程度表現できていると考えられる。しかし、その差異は、9 の地点でそれぞればらつきがある。No.2, No.6 地点では、全体的に過小評価となっているが、No.8 ではやや過大評価、No.3, No.4, No.5, No.7 では深度の上部、下部においてその評価が違ってきている。これは、理論平均 N 値が多種多様な状態で存在する土質の 1 つを平均化したものであると考えれば、ここで見えてくる差異は、調査地点の堆積構造や土の状態や種類によるものと考えられる。したがって、この鉛直方向のトレンド成分に加え、その差異の成分を水平方向の状況から判断することによって、実測値の変動を説明できると考える。

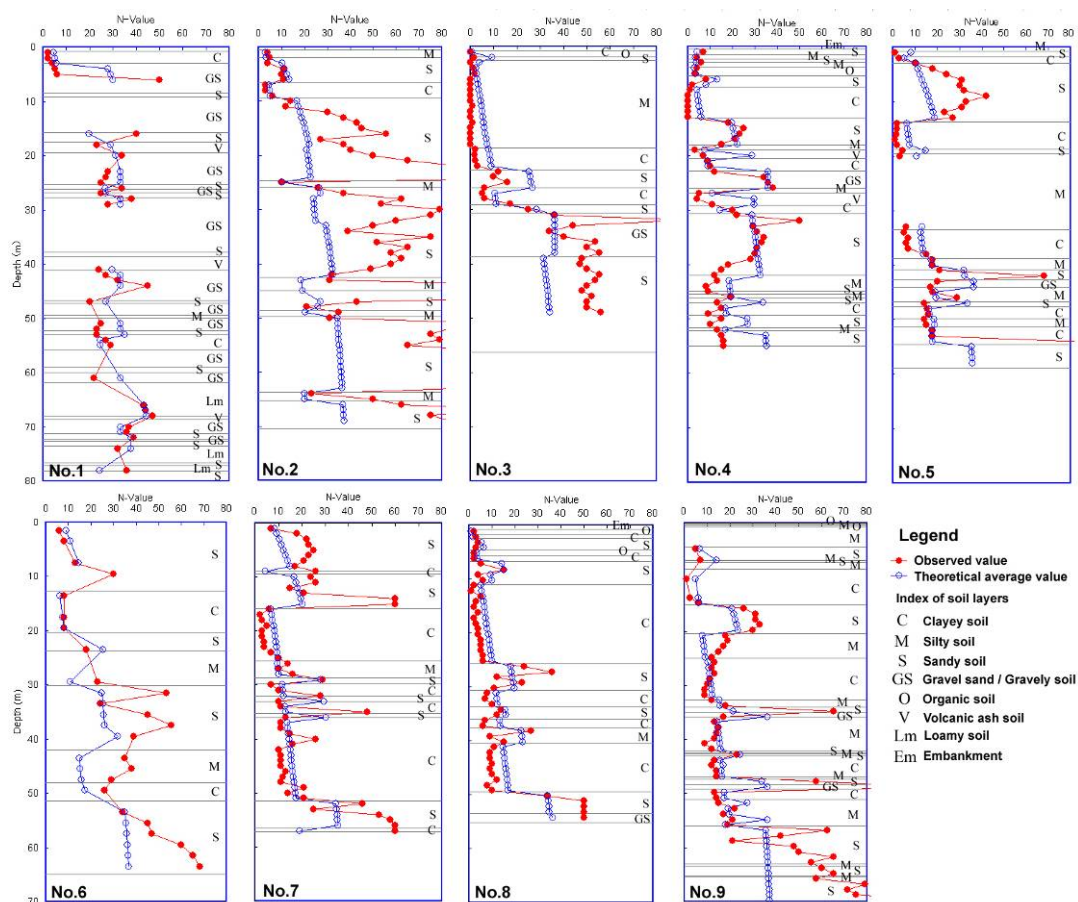


図 4.6 札幌市とその周辺の調査結果と理論平均 N 値との比較

#### 4.5 電子化された情報の信頼度の判定方法の提案

データベース化された個々の情報の信頼度は、情報源である調査そのものの信頼度、紙面情報の記載ミス、本電子化プロセスにおける誤認識などに依存する。この信頼度はデータベース全体の信頼度を左右させる。個々の情報の信頼度評価手法の1つとして、統計的手法に基づく手法の適用性を検討した。その結果、深度、N値、土質区分には、深度方向にトレンドがあることが分かった。このトレンドの関係式のパラメータは地盤の堆積環境に依存すると思われるので、この傾向を地域ごとに確かめることによって、このトレンドより大きく異なる情報が含まれる場合には、信頼度が低いデータと判断することが可能となると考える。

## 第 5 章 紙面地盤情報電子化技術を利用する高密度 地盤情報 DB 構築手法の提案

### 5.1 構築手法の提案

本研究では、地盤調査報告書や地盤柱状図集といった既存の紙面地盤情報を、デジタルスキャニング技術を利用し、個々の情報を効率的に電子化するとともに、集積したデジタル地盤情報から個々の情報の信頼度を地盤統計学的手法に基づき判定するとともに、高密度地盤情報データベースを構築する手法を開発する試みを行った。

本研究の主要な部分は、①地盤調査報告書や地盤柱状図集といった既存の紙面地盤情報を効率的にデジタル化する技術開発、および、②電子化された情報の信頼度の判定方法の確立である。

これらを達成するために、まず、第 2 章において、紙面の地盤情報を電子化する方法と解決すべき課題についてとりまとめた後、第 3 章において、既存の紙面地盤情報を効率的にデジタル化する技術開発を行い、紙面地盤情報を電子化できるソフトウェアとその手順について説明した、第 4 章では、電子化された情報の信頼度の判定方法について既存の地盤情報 DB を用いた検討を行い、信頼度判定方法を提案した。

以上の成果を有機的に結合させることによって紙面柱状図を効率的にまた大量に電子化することによる高密度地盤情報データベース構築できる。その手法を図 5.1 に示す。

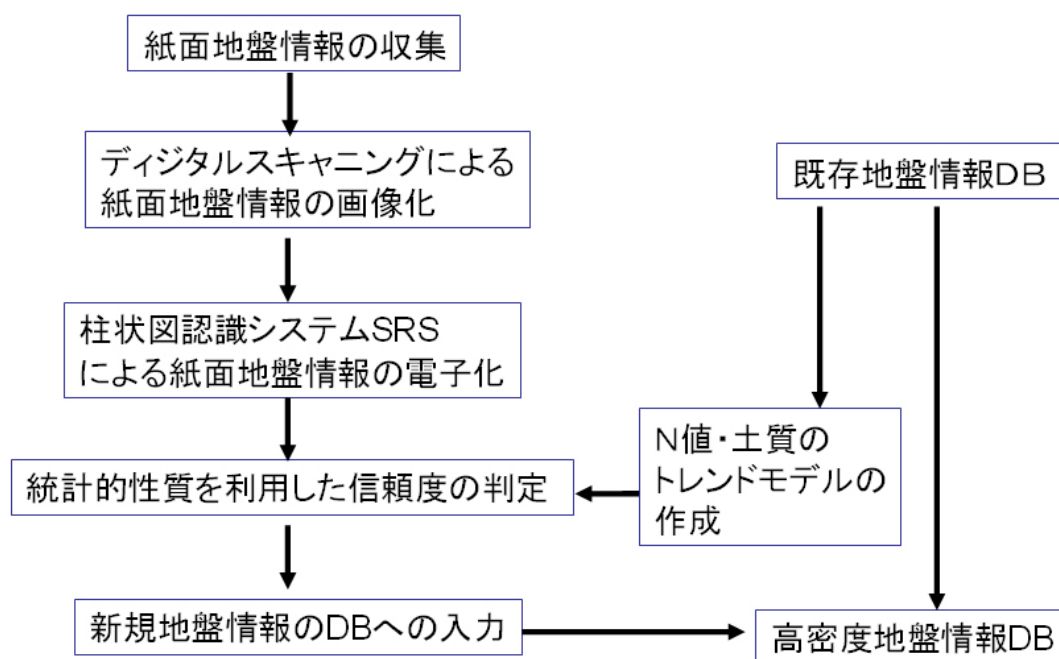


図 5.1 紙面地盤情報電子化技術を利用する高密度地盤情報 DB 構築手法

## 5.2 今後の課題と方向性

柱状図情報についてその特性とデジタルスキャンニングおよびOCR技術を活用して電子化する技術開発を行った。現在は、ワープロなどで印字された文字の認識を主として行っているが、古い年代の柱状図には手書きの文字のものもある。手書き文字をどのように認識し確信度を上げるかの工夫が必要である。さらに、手書きと同時に、いわゆる“青焼き”の紙面情報も存在する。この種の紙面情報は認識すべき情報である文字の潰れ、汚れ擦れなのかあるいは文字なのかの判別が困難になるものと思われる。今後は、どれくらいの年代までの紙面柱状図であれば電子化できるのかを示していきたい。

また、地盤情報 DB に収めるためには、その調査位置として緯度経度情報を含めなければならない。これについては次の技術開発であると考えているが、昨今の GIS（地理情報システム）の発展と高機能化、電子地図情報の整備と公開の状況を踏まえて実施すれば、比較的効率的に情報化ができると考えている。

さらに、現在平衡して実施している地盤の空間的な連続性や土質特性といった地盤工学的な知見を利用することによる電子化された情報の信頼度の判断と地盤情報 DB への入力、

元情報との再確認といった地盤工学的知見を考慮することによって、より確からしい地盤情報 DB の構築が可能となると考える。

以上のことが達成されれば、従来、紙面情報として取り扱われていた地盤情報が、電子化された地盤情報 DB として構築、または、追加されることによる情報の高密度化が期待できる。



## 謝辞

本研究を進めるにあたり、長崎県土木部技術情報室、新潟県県民生活環境部にご協力いただいた。付記して謝意を表します。

#### 参考文献

- 1) 前田伊瑞実：地盤情報の標準化の現状とデータ利用、建設関連業月報、Vol.26、No.295、pp.5-8, 2006.
- 2) 地盤工学会北海道支部北海道地盤情報データベース化委員会：北海道地盤情報データベース Ver.2003, 地盤工学会北海道支部, 2003.

助成研究者紹介

むらかみ さとし

村上 哲

現職：茨城大学工学部講師（工学博士）

うえだ よしかず

上田 賀一

現職：茨城大学工学部准教授（工学博士）

# DIGITAL SCANNING TECHNOLOGY FOR UPGRADING GEO- INFORMATION DATABASE

Murakami, S.<sup>1</sup> Ueda, Y.<sup>1</sup>  
<sup>1</sup>Ibaraki University

Geo-information database is available for prevention and mitigation against natural geo-disasters and planning for construction. Recently, almost data of site investigation of ground is recorded by means of digital media and the data is collected in a geo-information database. However, there are a lot of data of past site-investigation of ground recorded on reports. Therefore, it is important for soil geo-information database to develop an efficient method for converting paper geo-information into digital one. This study developed and proposed the method for geo-information database by using digital scanning technology. And it also proposed the method for evaluation of the quality of an adding new geo-information.

**KEYWORDS:** *Geo-information database, Digital scanning technology, Boring log*

## 研 究 成 果 の 要 約

助成番号	助成研究名	研究者・所属
第2006-1号	紙面地盤情報電子化技術を利用する高密度地盤情報DB構築手法に関する研究	村上 哲・茨城大学工学部
<div style="display: flex; justify-content: space-between;"> <div style="width: 48%;"> <p>柱状図や各種土質試験をデジタル化した地盤情報データベース（以下、地盤情報 DB と略称する）が構築され、建設時における地盤調査計画、構造物設計などの業務支援や地域の防災および環境保全のための貴重な情報源と位置づけられている。わが国における地盤情報 DB は、とりわけ主要都市部において構築・公開が進んでおり、これは近年の情報電子化による DB への移行の容易さと地盤調査の機会が多さなどに依存している。これに対し、地方都市においては、建設頻度の低さから、調査地点が疎な地盤情報 DB と成らざるを得ない。</p> <p>しかし、過去の地盤調査結果や地盤柱状図集といった紙面情報が存在することが多く、これらを効率的かつ十分な信頼度を持ってデジタル化することができれば、地盤情報収集が困難な地域の地盤情報 DB 構築の大きなドライバーフォースとなりうるとともに、地域の建設業務支援や、防災・環境保全のための貴重な情報整備を支援できる。また、社会的経済的な点と DB の整備状況から見た地域格差の是正も期待できる。</p> <p>そこで、本研究では、地盤調査報告書や地盤柱状図集といった既存の紙面地盤情報を、デジタルスキニング技術を利用し、個々の情報を効率的に電子化するとともに、集積したデジタル地盤情報から個々の情報の信頼度を地盤統計学的手法に基づき判定するとともに、高密度地盤情報データベースを構築する手法を開発することが目的である。</p> <p>本研究の主要な部分は、①地盤調査報告書や地盤柱状図集といった既存の紙面地盤情報を効率的にデジタル化する技術開発、および、②電子化された情報の信頼度の判定方法の確立である。</p> </div> <div style="width: 48%;"> <p>これらを達成するために、まず、既存の紙面情報の印刷形式について調査を行い、スキニングによる情報認識のためのカテゴリー分けを行った。次に、情報化すべき項目について優先順位を決定し、カテゴリー項目に応じた電子化プロセスについて検討を行う。この検討では、ある程度規格化されている柱状図書式を最大限に活用すること、および、効率化と信頼度の2軸と処理時間の3つの軸を念頭におく電子化プロセスを構築した。この構築プロセスに基づいて、一般的に利用されているパーソナルコンピュータでも利用できる「土質柱状図認識システム SRS」開発した。具体的な適用を行うことによって、</p> <ul style="list-style-type: none"> <li>認識可能な紙面情報</li> <li>画像化する際のファイル形式</li> <li>画像解像度</li> <li>紙面情報の文字サイズ</li> </ul> <p>を明らかにするとともに、もっとも認識精度の良い状態を示した。</p> <p>次に、②では、地盤情報 DB に入力された地盤情報の信頼度の判定方法の確立について、既存の地盤情報 DB を用いた検討を行った。特に、N値、深度、土質の組み合わせにおいて、統計的性質と深度方向のトレンドモデルを提案した。これを用いることによって、個々の情報の信頼度の評価法を提案した。</p> <p>以上の結果を有機的に結合させることによって紙面柱状図を効率的にまた大量に電子化することによる高密度地盤情報データベース構築手法を提案した。</p> </div> </div>		