地盤定数データベースの構築と港湾の粘性土の圧縮指数推定について

片山　遥平[1]・笠間　清伸[2]

[1]正会員　五洋建設株式会社　技術研究所（〒329-2746 栃木県那須塩原市四区町 1534-1）
E-mail: youhei.katayama@mail.penta-ocean.co.jp (Corresponding Author)
[2]正会員　九州大学大学院教授　工学研究院社会基盤部門（〒819-0395 福岡県福岡市西区元岡 744）
E-mail: kasama@civil.kyushu-u.ac.jp

圧密沈下量の高精度な予測には地盤の圧縮指数が必要となる．圧縮指数は圧密試験によって得られるが，長時間を要する試験であるため，実務では液性限界や間隙比から推定されることがある．様々な経験式が提案されているが，その多くは地域ごとの地盤特性を考慮しない統一的なものであり，安易に適用すると大きな推定誤差が生じる恐れがある．そこで本研究では，地域ごとの地盤特性を考慮した圧縮指数の推定方法を構築することを目的として検討を実施した．まずWeb上で公開されているボーリング資料などをもとに，テーブルデータ形式の地盤定数データベースを構築した．このデータを国内の主要な港湾ごとに分割し，重回帰分析および機械学習によって推定モデルを構築し，圧縮指数を推定した．

Key Words: *compression index, database, multiple regression analysis, machine learning*

1. はじめに

日本の沿岸部には軟弱な粘性土地盤が広く堆積しており，構造物の建設に伴い圧密沈下を生じる可能性がある．設計条件を決定するため圧密沈下量の事前予測を行い，施工中の観測によって事前設計を適宜修正していくのが一般的である．ただし，事前予測が大きく外れると経済性や工期に影響するため，これには精度が求められる．圧密沈下量の事前予測手法として，一次元圧密理論に基づいたΔe法，m_v法およびC_c法などがある[例えば1)]．これらは少ないパラメータから簡易に圧密沈下量を求めることができるが，地盤条件や載荷条件が複雑な場合には適用が難しい．そのため実務では，有限要素法による数値解析が行われることが多い．様々な理論や地盤構成則が提案されているが，共通して粘性土の圧縮指数が必要となる．圧縮指数は圧密試験から入手することができるが，長時間を要する試験であるため，他のパラメータから推定されることがある．液性限界や間隙比などを用いる様々な経験式が提案されており，いくつかの例を式(1a)～(1d)に示す[2),3),4)]．しかし，土の特性は地域によって大きく異なるため，安易にこれらの経験式を適用すると大きな推定誤差が生じる恐れがある．

$$C_c = 0.009(w_L - 10) \quad (1a)$$

$$C_c = 0.015(w_L - 19) \quad (1b)$$

$$C_c = 0.545(e_0 - 0.756) \quad (1c)$$

$$C_c = 0.490(e_0 - 0.410) \quad (1d)$$

ここに，C_c：圧縮指数，w_L：液性限界，e_0：自然間隙比である．

ここで，日本各地で実施されたボーリング調査資料などを取りまとめている地盤情報データベースが数多く存在する．代表的なものを**表-1**に示す．

データベースを活用した研究は多くある．田中らは，北海道を除く全国の港湾地域で実施されたボーリングデータが登録されたデータベースを用いて，沿岸域における土の物理特性と圧密特性の統計解析を行っている[5)]．田中は，港湾における土質データベースを利用して，地盤の液状化判定プログラムと三次元可視化プログラムを開発し，設計業務の効率化を図っている[6)]．山本らは，大阪湾地盤情報データベースから解析対象地の地盤特性を抽出し，洪積粘土層の沈下予測を行っている[7)]．北岡らは，関西圏地盤情報データベースから地盤調査データを整理し，人工ニューラルネットワークを用いて三軸圧縮試験から求められる内部摩擦角の推定を行っている[8)]．このように，これまでに様々な目的でデータベースが活用されている．

本研究では，はじめに全国の様々な特性を持つ地盤の物性値をとりまとめた地盤定数データベース（以下，

表-1　主要なデータベース

データベース名	作成主体	内容	地域	使用料
国土地盤情報検索サイト KuniJiban	国土交通省 国立研究開発法人土木研究所 国立研究開発法人港湾空港技術研究所	国土交通省のボーリング柱状図， 土質試験結果など	全国の道路・河川・港湾など	無償
ジオ・ステーション	国立研究開発法人防災科学技術研究所	防災科学技術研究所，産業技術総合研究所， 土木研究所および9つの自治体が 提供しているデータ	全国（主に関東）	無償
北海道地盤情報データベース	北海道土木技術会 地盤工学会北海道支部	ボーリング柱状図など	北海道	有償 （絶版）
とうほく地盤情報システム「みちのくGIDAS」	東北地盤情報システム研究会	デジタル地質図，ボーリングデータ および災害履歴データ	東北	有償 （閲覧は無償）
ほくりく地盤情報システム	北陸地盤情報活用協議会	ボーリング柱状図など	北陸	有償
新・関東の地盤	地盤工学会関東支部	ボーリング柱状図，土質試験結果など	関東	有償 （書籍および付録DVD）
関西圏地盤情報データベース	関西圏地盤情報協議会 関西圏地盤DB運営機構 関西圏地盤研究会	ボーリング柱状図，土質試験結果など	近畿	有償 （閲覧は無償）
四国地盤情報データベース	四国地盤情報活用協議会	ボーリング柱状図，土質試験結果など	四国	有償 （CD-ROM）
九州地盤共有データベース	地盤工学会九州支部	ボーリング柱状図，土質試験結果など	九州	有償 （CD-ROM）

地盤定数 DB）を構築した．そしてデータ整理を行ったうえで港湾ごとのデータを抽出し，入手が比較的容易なパラメータから圧縮指数を精度良く推定する手法について検討を行った．検討には重回帰分析と機械学習の 2 手法を用いた．

2.　地盤定数データベースの構築

(1)　利用するデータベース

表-1 に示したデータベースのうち，無償かつ自由に利用することができ，全国の調査データを収集している「国土地盤情報検索サイト KuniJiban[9]」（以下，KuniJiban）を利用することとした．

(2)　KuniJiban からのデータの入手

KuniJiban からは国土交通省管内のボーリング柱状図，土質試験結果一覧表および土性図等を入手できる．拡張子は pdf と XML の 2 種類があり，テキスト形式で取り扱いが容易な後者を使用した．本研究を開始した 2022 年 12 月時点では，ボーリングと土質試験データのペアは全国で約 7 万個あり，これらを手作業でダウンロードするには大きな労力を要するため，プログラミング言語の一種である Python を使用してスクレイピングを行った．この際，万が一にも DDOS 攻撃とみなされないよう，アクセス頻度は 3 秒に 1 回とした．

(3)　テーブルデータへの変換

XML 形式のままでは統計的な分析が行えないため，必要な情報のみを抽出してテーブルデータとする必要がある．前述のとおりデータ数が膨大で手作業では困難であるため，Python を活用し自動化した．この際データ記述方法が他のものと著しく乖離しているものは除外し，最終的には 50,959 地点分のデータを用いて地盤定数 DB を構築した．今回構築した地盤定数 DB に含まれるパラメータのうち，データ数が 1 万個を超えるパラメータとその個数を**表-2** に示す．ここでデータ数 1 はボーリングデータと土質試験データを組み合わせた場合，データ数 2 は土質試験データのみを用いた場合の個数である．ボーリングデータと組み合わせる場合は，N 値を取得している深度のデータのみを抽出する都合上，N 値以外のパラメータの個数は少なくなる．今回の検討では，データ数をできる限り多くするため，土質試験データのみを用いて構築した地盤定数 DB を検討に使用した．

3.　データの整理

(1)　データの前処理

地盤定数 DB の構築に使用したデータの数は膨大であり，一つ一つのデータの精査は実施できていない．そのため検討に使用する前にデータの前処理が必須となる．

まずは圧縮指数が空の行を削除する．次に負の値などの異常値を除外する．次に，今回の検討では粘性土の圧縮指数を求めることを目的としているため，細粒分含有率が 50 %未満のデータを砂質土・礫として除外した．次に外れ値を除外する．例えば粘性土の土粒子密度が 1.0 g/cm^3 以下や 4.0 g/cm^3 以上となることは考えにくい．

このようなデータを除外するため，四分位数IQRによる外れ値の除去を行った．下限値と上限値を決定する式を式(2a), (2b)に示す．

$$lower\ limit = Q_1 - 1.5 \times (Q_3 - Q_1) \quad (2a)$$

$$upper\ limit = Q_3 + 1.5 \times (Q_3 - Q_1) \quad (2b)$$

ここに，Q_1：データの下位 25 %値，Q_3：データの上位 75 %値である．なお，$Q_3 - Q_1$ を IQR と呼ぶ．

(2)　各パラメータと圧縮指数の相関の確認

表-3 に前処理後の各パラメータと圧縮指数の相関係数を示す．相関係数とデータ数をもとに，圧縮指数の推定に用いる説明変数を①湿潤密度，②間隙比，③含水比，④粘性土分，⑤液性限界および⑥塑性限界に決定した．実務適用を考えると，できるだけ使用する説明変数を少なくしたいという思惑も含まれている．これら 6 つの説明変数と圧縮指数のペアプロットを**図-1** に示す．この図の作成には Python ライブラリの seaborn-analyzer[10]を使用した．下三角部は 2 パラメータのプロット，対角部は各パラメータのヒストグラム，上三角部は 2 パラメータ間の相関係数を示している．なお相関係数は，絶対値が大きいほどフォントサイズが大きくなる仕様となっている．

ここで湿潤密度，間隙比および含水比間の相関係数の絶対値はいずれも 0.95 以上と非常に大きく，多重共線性が疑われる．6つの説明変数で VIF[11]を算出した結果を**表-4** に示す．多重共線性を認める VIF の値には諸説あるが，一般的に採用される 10 を閾値とすると，湿潤密度，間隙比および含水比のうち 2 つを除外すべきという考えになる．ただし本研究では，多重共線性が圧縮指数の推定にどう影響するかを確認するため，6 変数とも使用することとした．

4.　港湾データの抽出

(1)　検討対象とした港湾

日本には多くの港湾が存在するが，このうち国際戦略港湾の 5 港湾と国際拠点港湾の 18 港湾を検討対象と設定した．これらの港湾を**図-2**[12]に加筆，**表-5** に示す．

(2)　港湾ごとのデータの抽出

地盤定数 DB には調査地点の緯度・経度が含まれているため，これを利用して前述の港湾ごとにデータを抽出する．各港湾の領域は港湾計画図[例えば 13]より設定した．本研究では港湾計画図の色が塗られている範囲（港湾区域を含む）をその港湾の領域と仮定した．次に国土地理院の基盤地図情報ダウンロードサービス [14]から各港湾の海岸線および道路縁の XML データ（以下，地図データ）を入手した．港湾計画図と地図データを用いた港湾領域

表-2　データ数が 1 万個を超えるパラメータとその個数

分類	パラメータ	データ数1	データ数2
調査データ	調査点数	42,306	50,712
	地下水位	692,779	-
	標高	771,844	363,448
	深度	771,870	367,593
標準貫入試験	N値	771,837	-
密度	土粒子密度	236,941	328,819
	湿潤密度	32,052	65,965
	乾燥密度	30,668	62,199
	間隙比	30,642	61,922
含水状態	含水比	233,901	326,072
	飽和度	30,365	61,328
粒度	礫分	257,799	354,461
	砂質土分	258,420	355,405
	シルト分	193,134	265,379
	粘性土分	184,708	254,818
	最大粒径	256,009	351,678
	均等係数	217,241	292,155
	D50	118,965	161,009
	D10	101,083	137,101
コンシステンシー	液性限界	110,817	156,616
	塑性限界	110,569	156,193
圧密試験	圧縮指数	16,190	31,855
	圧密降伏応力	16,173	31,813
	破壊ひずみ	9,439	23,000
三軸圧縮試験	せん断応力（全応力）	18,022	33,462
	内部摩擦角（全応力）	18,034	33,080
強度	一軸圧縮強さ	10,488	26,238

表-3　各パラメータと圧縮指数の相関係数

パラメータ	相関係数	データ数
深度	0.033	15,637
土粒子密度	-0.332	15,827
湿潤密度	-0.823	14,736
乾燥密度	-0.838	14,389
間隙比	0.861	14,691
含水比	0.857	15,844
飽和度	0.109	14,603
礫分	-0.481	25
砂質土分	-0.206	15,468
シルト分	-0.273	15,868
粘性土分	0.350	15,868
最大粒径	0.103	15,507
均等係数	-0.104	866
D50	-0.162	7,605
D10	-0.242	341
液性限界	0.729	15,524
塑性限界	0.632	15,523
圧密降伏応力	-0.270	15,696
破壊ひずみ	-0.403	7,959
せん断応力（全応力）	-0.021	6,348
内部摩擦角（全応力）	0.068	4,216
一軸圧縮強さ	-0.155	8,898

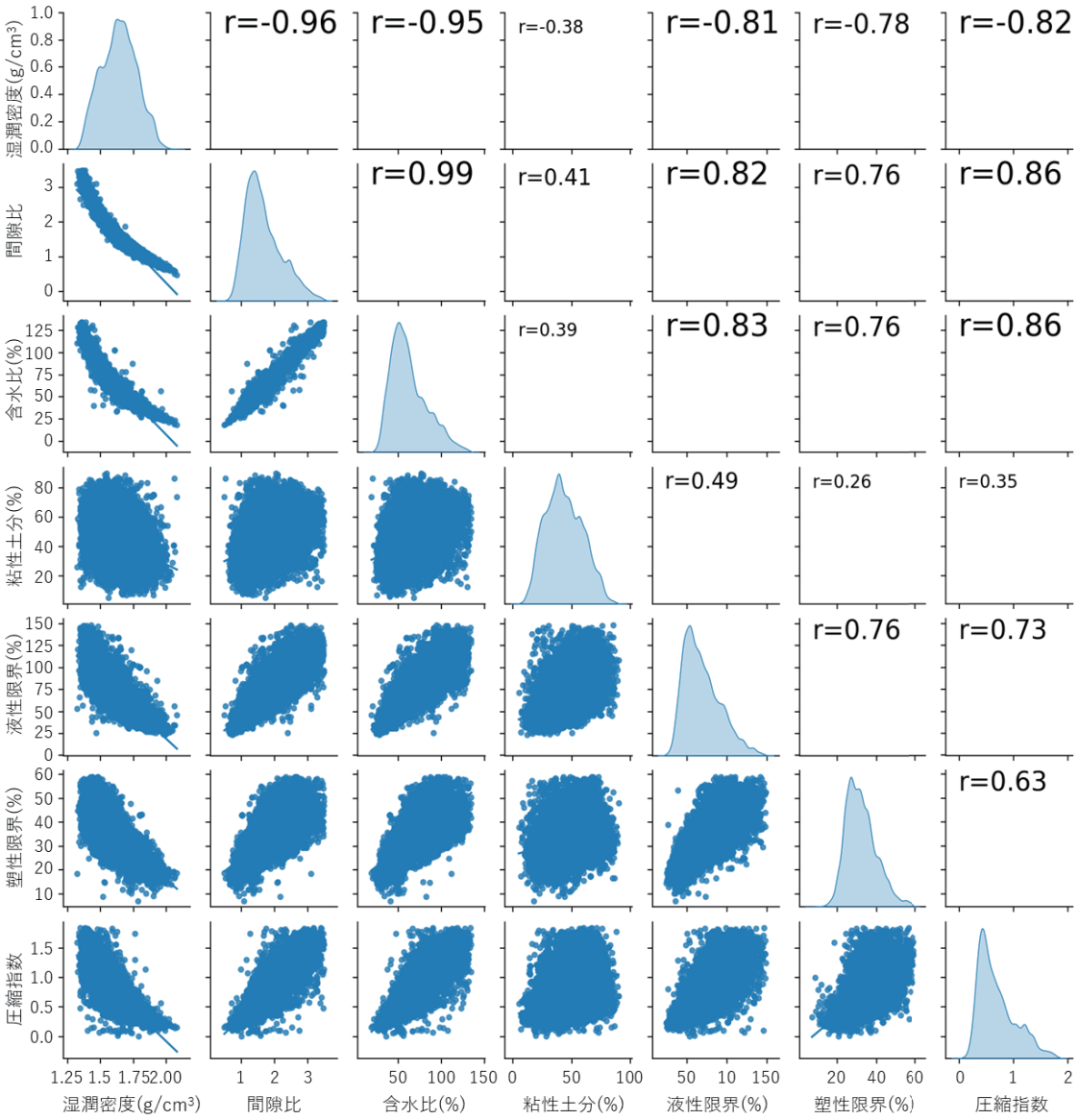


図-1　各パラメータと圧縮指数の相関係数

の設定手順を下記に示す．

①地理情報システムの QGIS[15]で地図データを描画する．

②QGIS のジオリファレンサ機能から画像変換した港湾計画図を開く．

③地図データと港湾計画図で特徴的な共通地点を複数選択することで，これらを重ね合わせる．

④港湾計画図の塗りつぶし範囲のうち，最も離隔が大きいと思われる 2 点を目視で設定する．

⑤この 2 点を楕円の長軸としたとき，楕円から最も離れていると思われる地点を複数選択する．

⑥Python ライブラリの shapely[16]を用いて，⑤の点全てを内側に収めることのできる最小の楕円を求める．

東京港 [13]の例を**図-3** に示す．形状が歪んでいるのは，港湾計画図と地図データを重ね合わせる際の誤差の補正によるものである．地盤定数 DB のうち，この楕円内に入っているデータ（**図-3** 中の白丸）をその港湾のデータと設定した．

表-4　説明変数の VIF

変数の数	湿潤密度	間隙比	含水比	粘性土分	液性限界	塑性限界
6	13.74	55.77	50.99	1.42	4.20	2.99
5	-	49.33	50.59	1.42	4.19	2.83
5	12.15	-	12.21	1.36	4.16	2.99
5	13.63	13.36	-	1.38	4.01	2.99
4	3.64	-	-	1.36	3.86	2.99
4	-	3.57	-	1.37	3.98	2.83
4	-	-	3.66	1.36	4.16	2.79

(3)　港湾ごとの傾向の確認

7 変数の箱ひげ図を**図-4** に示す．港湾名の下の数値はデータ数を示す．なおデータ数の都合から，前述の 23 港湾のうち 13 港湾に絞って検討した．まず目的変数である圧縮指数について着目すると，和歌山下津港，徳山下松港および下関港は範囲が非常に狭くなった．説明変

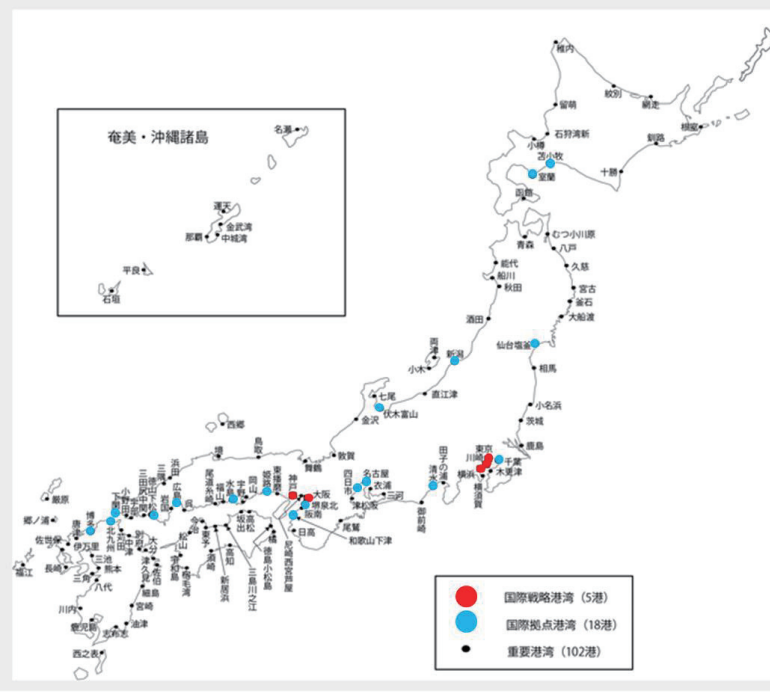


図-2　対象港湾 [12]に加筆

表-5　対象港湾

国際戦略港湾	国際拠点港湾
東京港	苫小牧港
横浜港	室蘭港
川崎港	仙台塩釜港
大阪港	千葉港
神戸港	新潟港
	伏木富山港
	清水港
	名古屋港
	四日市港
	堺泉北港
	和歌山下津港
	姫路港
	水島港
	広島港
	徳山下松港
	下関港
	北九州港
	博多港

数に関しては，和歌山下津港および徳山下松港が他の港湾の傾向と大きく異なるように見える．また北九州港はデータの前処理済みであるにも関わらず，粘性土分を除いて外れ値が多く，ばらつきが大きいことがわかる．

5.　重回帰分析による推定

(1)　推定手法

第 4 章で得られたデータを対象に，比較的実装が容易な重回帰分析による推定を実施した．損失関数へ正則化項を導入しない，いわゆる通常の線形回帰（以下，線形回帰）と，L1 と L2 の 2 つの正則化項を導入した ElasticNet[17]の 2 手法により検討を実施した．両者の損失関数を式(3a), (3b)に示す．

$$L = \sum_{i}^{N} \left(\left(\sum_{j}^{D} a_j x_j + b \right) - y_i \right)^2 \tag{3a}$$

$$L = \sum_{i}^{N} \left(\left(\sum_{j}^{D} a_j x_j + b \right) - y_i \right)^2 + L1 + L2 \tag{3b}$$

ここに，L：損失関数，N：データ数，D：説明変数の次元数，a：係数，x：説明変数，b：切片，y：目的変数の実測値，$L1$：$L1$正則化，$L2$：$L2$正則化である．

推定式は線形式，指数式の 2 通りとした．両者を式(4a), (4b)に示す．後者は変数の非線形性や変数間の相互作用を考慮できることを期待して採用した．なお，対数変換することで線形回帰の形に変換することができる．

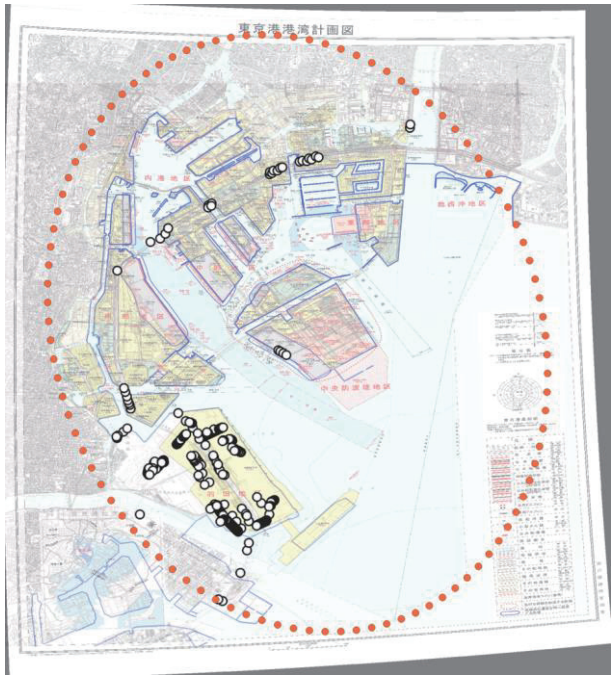


図-3　東京港の領域とデータ [13]に加筆

$$y = \sum_{j}^{D} a_j x_j + b \tag{4a}$$

$$y = \prod_{j}^{D} {x_j}^{a_j} \times b \tag{4b}$$

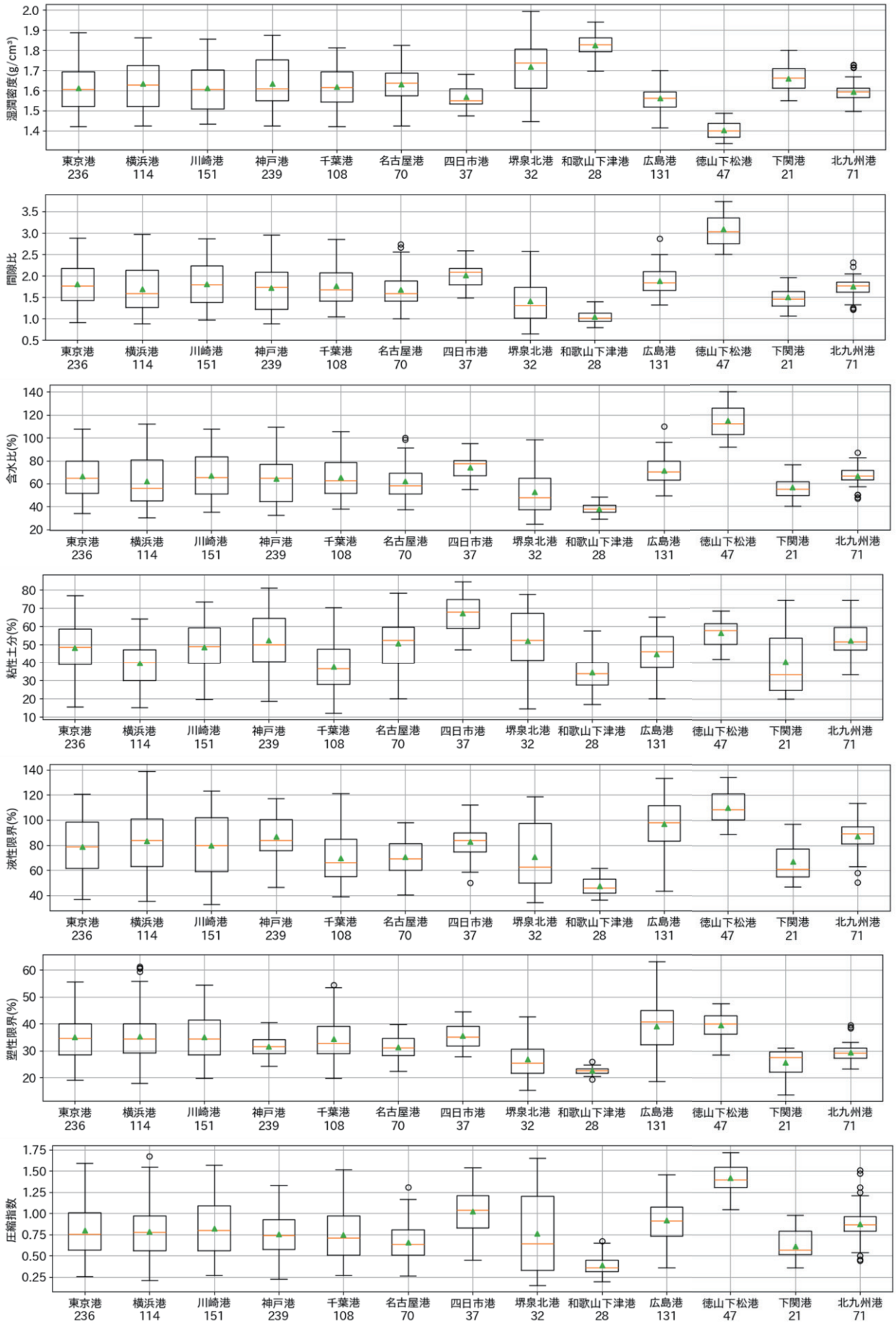

湿潤密度(g/cm³)
間隙比
含水比(%)
粘性土分(%)
液性限界(%)
塑性限界(%)
圧縮指数
東京港 236
横浜港 114
川崎港 151
神戸港 239
千葉港 108
名古屋港 70
四日市港 37
堺泉北港 32
和歌山下津港 28
広島港 131
徳山下松港 47
下関港 21
北九州港 71

図-4　7変数の箱ひげ図

線形回帰では 6 つの説明変数の全ての組合せ（=63 通り）で計算を行い，最も良い説明変数の組合せを探索することにした．ElasticNet に関しては，目的変数への寄与度が小さい説明変数を自動的に判断して，その係数を小さくすることで過学習を防止する機能が備わっている．そのため説明変数の組合せを検討する必要はないが，alpha および l1_ratio という 2 つのハイパーパラメータの調整が必要となるため，線形回帰同様に全探索を実施した．alpha は 10^{-4}～10^{2} の範囲で対数的に 10 分割，l1_ratio は 0.0～1.0 の範囲で線形的に 10 分割，合計 100 通りとした．

また ElasticNet では説明変数の標準化が必須であるため，線形回帰でも同様の処理を実施した．線形回帰の場合は標準化の有無によって圧縮指数の推定結果が変わることはない．

(2)　データの分割

未知の地盤への推定精度を評価するため，ダブルクロスバリデーションを採用した．概要図を**図-5** に示す．

まず手持ちのデータを 5 分割し，うち 4 つを学習データ，残り 1 つをテストデータとする．次に 4 つの学習データをさらに 5 分割し，うち 4 つを学習データ，残り 1 つを検証データとする．データを入れ替えながら，学習データで各説明変数の係数・切片を決定，および検証データで推定精度を確認する作業を 5 回繰り返し，最も良い説明変数またはハイパーパラメータの組合せ（以下，最適パラメータ）を決定した．これを内側のクロスバリデーション（以下，CV）と呼ぶ．学習データと検証データを再結合したデータに最適パラメータを適用して係数・切片を決定し，最初に分割したテストデータにて推定精度を検証した．このテストデータは学習には一切関与していないため，これにより未知の地盤への推定精度が得られる．これを全てのデータが 1 回はテストデータとして使用されるよう 5 回繰り返す．これを外側の CV と呼ぶ．

なお，この手法では最適パラメータが異なる可能性のある 5 つのモデルが作成されることになる．実務適用の際は，最も精度の良いモデルを採用するか，5 つのモデルにて推定を行い，その結果を平均するアンサンブル学習を行うことなどが考えられる．

(3)　検討ケース

実務を想定すると，下記の 3 つの状況が考えられる．

a)　対象港湾のデータが少ない場合

この場合は対象港湾のデータのみを用いると，データ数の少なさゆえに傾向を十分に学ぶことができないため，他の港湾のデータも用いて推定式を構築するのが良いと考えられる．

具体的には，前述のデータ分割を港湾ごとに行い，13 港湾分の学習データを結合してモデルを構築する．検証データおよびテストデータの推定は港湾ごとに行い，港湾ごとのデータ数および圧縮指数のレンジを用いて，正規化 RMSE の加重平均を算出した．計算式を式(5)に示す．

$$Weighted\ NRMSE = \sqrt{\frac{\sum_i \left\{ n_i \cdot \left(\frac{RMSE_i}{range_i} \right)^2 \right\}}{\sum_i n_i}} \qquad (5)$$

ここに，n：港湾ごとの推定データの個数，$RMSE$：港湾ごとの RMSE，$range$：港湾ごとの圧縮指数の最大値−最小値である．

なおレンジは検証データから求めている．これを内側の CV の評価に使用した．

b)　対象港湾に十分なデータが存在する場合

この場合は対象港湾のデータのみを用いることで，その港湾に特化した推定式を構築することができ，精度が高くなると考えられる．

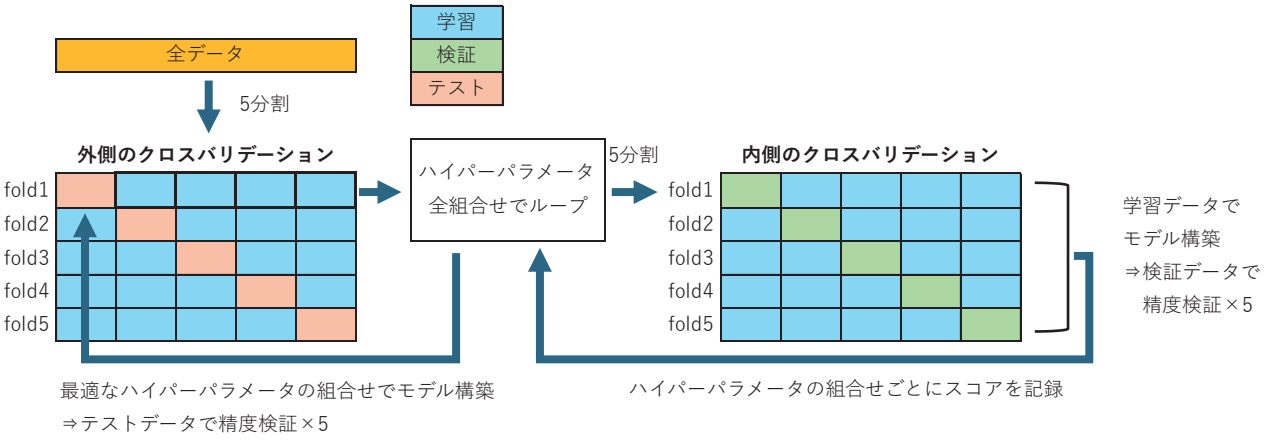


図-5　ダブルクロスバリデーションの概要

c)　対象港湾のデータが全くない場合

この場合は他の港湾のデータを用いて推定式を構築する以外にない．データの総数ができる限り等しくなるよう 13 港湾を 5 つのグループに分け，ダブル CV を実施した．グループの組合せを**表-6** に示す．内側の CV の評価にはケース a と同様に正規化 RMSE の加重平均を使用した．

表-6　グループの組合せ

港湾			データ数
東京	和歌山下津	-	264
横浜	千葉	堺泉北	254
川崎	名古屋	四日市	258
神戸	下関	-	260
広島	徳山下松	北九州	249

(4)　結果の整理

線形回帰および ElasticNet による推定結果を**表-7〜12** に示す．**表-7, 8** はケース a，**表-9〜11** はケース b，**表-12** はケース c の結果を示している．なお推定結果は，港湾ごとの全データのレンジを用いて算出した正規化 RMSE を示している．

a)　対象港湾のデータが少ない場合

本ケースでは全ての港湾を共通の推定式で推定している．まずは**表-7** の線形回帰において，圧縮指数の推定に採用された説明変数の組合せに着目する．線形式，指数式ともに 3〜5 の説明変数が採用されている．これは，異なる特性を持つ全国の地盤を精度良く推定するために，より多くの情報を必要としたものと考えられる．液性限界と塑性限界は線形式，指数式の 5 回の CV 全てに採用されており，間隙比に関しても指数式の 1 回目以外には採用されている．これは，これらの変数が圧縮指数の推定に特に重要であることを示唆している．

ここで，液性限界や塑性限界と比較して圧縮指数との相関係数が大きい湿潤密度と含水比の採用回数が少ないのは，多重共線性が原因と考えられる．**図-1** や**表-4** に示したように湿潤密度，含水比および間隙比には多重共線性があり，圧縮指数との相関係数が最も大きい間隙比のみで推定に十分寄与できるものと考えられる．

次に各説明変数の係数に着目する．表中の係数平均は係数の 5 回分の平均値を取ったものである．各説明変数は標準化されているため，係数の絶対値が重要度の指標となる．間隙比と液性限界の係数は比較的大きいが，塑性限界の係数は小さく，採用回数とは必ずしも一致しない結果となった．

ここで，間隙比と比較して湿潤密度と含水比の係数が小さいのは，採用回数と同様に多重共線性が原因と考えられる．ただし指数式では，湿潤密度の係数は比較的大きな値となっている．**図-1** を見ると，湿潤密度と圧縮指数の間には若干の非線形性が見受けられ，指数式を用いることでこの非線形性を考慮することができたものと考えられる．

次に**表-8** の ElasticNet について見ていく．ElasticNet では全説明変数を使用しているが，一部で含水比の係数が 0 となり，実質 5 変数での推定となる場合があった．係数の傾向は線形回帰とほぼ同様であった．

表-7, 8 に赤色で示した数値は，各港湾で誤差が最小のものを示している．線形回帰は 5/13 港湾，ElasticNet は 10/13 港湾でベストモデルとなっており（同率一位を含む），わずかな差ではあるが ElasticNet が有効であることを確認した．また**表-7** には式(1a)〜(1d)に示す従来式のうち，最も小さい推定誤差を併記している．また理論値とは，各港湾の全てのデータおよび全ての説明変数を用いて線形回帰の線形式（いわゆる通常の重回帰分析）により算出した推定誤差である．理論値には及ばないが，堺泉北港を除いた 12 港湾では，いずれの手法でも従来式よりも高い精度で推定することができた．

b)　対象港湾に十分なデータが存在する場合

本ケースでは港湾ごとに専用の推定式を構築している．**表-9〜11** において，各説明変数の係数やハイパーパラメータを記載すると冗長になるため割愛している．**表-9, 10** を見ると，採用された説明変数が 2 つ以下の場合が多く見られる．これはケース a と大きく異なる点で，単独の港湾であれば，少ない情報で精度良く推定できるということになる．また**表-9, 10** の右側の数字は，5 回の CV のうち各説明変数が採用された回数を示している．港湾によって採用される変数に大きな違いがあり，地盤特性の多様性を示している．線形式と指数式で推定誤差に大きな差はない．

ベストモデルとなる回数は，線形回帰は 5/13 港湾，ElasticNet は 8/13 港で，本ケースにおいても ElasticNet の方が相対的に高精度であり，有効であることを確認した．

ここで，データ数の非常に少ない港湾（下関港，和歌山下津港，堺泉北港および四日市港など）では，その港湾のデータのみを用いるより，他の港湾のデータも用いる方が高精度になると考えていたが，下関港を除いてそのような結果にはならなかった．これはやはり，港湾ごとの特色が非常に強いことを裏付けている．本研究では地理や地史を踏まえた考察はできていないが，今後はこのような観点からも検討を行い，地域ごとの地盤特性を明らかにしていきたい．

c)　対象港湾のデータが全くない場合

ケース b よりもケース a の方が高精度であった港湾は東京港，川崎港，千葉港および下関港である．**表-12** を見るとこれら 4 つの港湾は，それぞれのデータを全く使わずに推定式を構築した本ケースでも良好な精度が得ら

表-7　ケースa・線形回帰（線形式・指数式）

港湾	データ数	Cc min	Cc max	線形回帰						従来式		理論値
				線形	説明変数	係数平均	指数	説明変数	係数平均			
東京	236	0.260	1.590	0.102	1:湿間含－液塑 2:－間－－液塑 3:－間含－液塑 4:－間－－液塑 5:－間－粘液塑	湿潤密度 -0.032 間隙比 0.153 含水比 0.013 粘性土分 0.030 液性限界 0.099 塑性限界 0.020	0.100	1:湿－含粘液塑 2:－間含粘液塑 3:－間－－液塑 4:－間－－液塑 5:－間－粘液塑	湿潤密度 -0.148 間隙比 0.212 含水比 0.037 粘性土分 0.041 液性限界 0.128 塑性限界 0.023	0.139	鬼塚	0.098
横浜	114	0.210	1.670	0.144			0.151			0.160	Skempton	0.124
川崎	151	0.270	1.566	0.098			0.098			0.141	小川	0.095
神戸	239	0.230	1.329	0.112			0.110			0.145	Skempton	0.086
千葉	108	0.275	1.520	0.142			0.141			0.160	小川	0.128
名古屋	70	0.262	1.310	0.164			0.164			0.169	鬼塚	0.140
四日市	37	0.450	1.540	0.224			0.230			0.223	小川	0.169
堺泉北	32	0.153	1.648	0.163			0.165			0.099	小川	0.071
和歌山下津	28	0.200	0.674	0.148			0.149			0.222	Skempton	0.096
広島	131	0.359	1.460	0.133			0.128			0.207	Skempton	0.111
徳山下松	47	1.050	1.718	0.243			0.255			0.273	小川	0.177
下関	21	0.362	0.977	0.153			0.155			0.194	鬼塚	0.110
北九州	71	0.440	1.510	0.162			0.168			0.243	Skempton	0.140

表-8　ケースa・ElasticNet（線形式・指数式）

港湾	ElasticNet							
	線形	alpha	l1_ratio	係数平均	指数	alpha	l1_ratio	係数平均
東京	0.101	1:$10^{-2.00}$ 2:$10^{-3.33}$ 3:$10^{-4.00}$ 4:$10^{-2.67}$ 5:$10^{-2.67}$	1:0.778 2:1.000 3:0.000 4:0.778 5:0.333	湿潤密度 -0.027 間隙比 0.122 含水比 0.008 粘性土分 0.027 液性限界 0.083 塑性限界 0.024	0.099	1:$10^{-1.33}$ 2:$10^{-1.33}$ 3:$10^{-2.00}$ 4:$10^{-2.67}$ 5:$10^{-1.33}$	1:0.000 2:0.000 3:0.000 4:0.333 5:0.000	湿潤密度 -0.075 間隙比 0.099 含水比 0.039 粘性土分 0.039 液性限界 0.111 塑性限界 0.029
横浜	0.146				0.151			
川崎	0.098				0.098			
神戸	0.110				0.107			
千葉	0.139				0.138			
名古屋	0.170				0.163			
四日市	0.214				0.223			
堺泉北	0.156				0.165			
和歌山下津	0.146				0.144			
広島	0.130				0.128			
徳山下松	0.257				0.260			
下関	0.159				0.152			
北九州	0.163				0.169			

表-9　ケースb・線形回帰（線形式）

港湾	線形回帰											
	線形	説明変数						湿	間	含	粘	液	塑
東京	0.107	－間－粘液－	－間－－液－	湿間－粘液塑	－間含－－塑	－間－－－塑		1	5	1	3	3	2
横浜	0.136	－－－－液塑	－－－－液－	－間含粘液－	湿－－粘液－	－間含－液－		1	2	2	5	1	2
川崎	0.102	湿－－粘液塑	－－含－液－	－－含－液－	－－含粘液－	－－含－－塑		1	0	4	4	2	2
神戸	0.088	－間－粘液塑	－－含粘液－	－－含粘液－	湿－－粘液－	－－含粘液－		1	1	3	5	1	5
千葉	0.151	湿間－粘－－	－間含粘液塑	湿間－粘－－	－間－粘－塑	－間－粘－－		2	5	1	1	2	5
名古屋	0.150	－－含－－塑	湿－含－－塑	－－含－－塑	－間－－－塑	湿－含－－塑		2	1	4	0	5	0
四日市	0.192	－間含－液－	湿間－－液塑	湿間含－－－	湿間－－－－	－間含－－－		3	5	3	2	1	0
堺泉北	0.092	－間－粘液塑	－間－粘液塑	－－含粘液塑	－間－粘液塑	－間含－液－		0	4	2	5	4	4
和歌山下津	0.128	－－含－－－	－－含粘－塑	－－含－－－	－－含－液－	－－含－－－		0	0	5	1	1	1
広島	0.115	－間－粘－塑	－間－粘－－	－間含粘－－	－間－粘－－	－間－粘－－		0	5	1	0	1	5
徳山下松	0.193	－間含－－－	湿間含－－－	－間－－－－	湿間－－－－	－間－－－－		2	5	2	0	0	0
下関	0.192	湿－含－－－	－間含－－－	湿－含－－－	－－含－－塑	－間－粘液塑		2	2	4	1	2	1
北九州	0.153	－－含粘－－	－－含粘－塑	－間－粘－塑	湿間－粘－塑	湿－含粘－－		2	2	3	0	3	5

れている．一方で残りの9港湾では目に見えて精度が低下する結果となった．これら9港湾は独自の地盤特性を有していることを確認した．

ベストモデルとなる回数は，線形回帰は 6/13 港湾，ElasticNet は 10/13 港で（同率一位を含む），本ケースでもElasticNetの有効性を確認した．ベストモデルの回数で

表-10　ケースb・線形回帰（指数式）

港湾	線形回帰											
	指数	説明変数					湿	間	含	粘	液	塑
東京	0.103	湿－－粘－塑	－－含－－塑	湿－－粘－塑	湿－含粘－塑	湿－－粘－塑	4	0	2	4	0	5
横浜	0.134	－－－－液塑	－－－－液塑	湿間含粘液－	－－－粘液－	湿間含－液塑	2	2	2	2	5	3
川崎	0.100	湿－－粘液－	湿－－－液－	湿－－－液－	－間－－液－	湿－－－液－	4	1	0	1	5	0
神戸	0.089	湿－含粘液塑	湿間含粘－塑	－－含粘液－	－－含粘液－	－－含粘液－	2	1	5	5	4	2
千葉	0.147	－間－粘液－	－間含粘液塑	－間－粘－－	湿－－粘－塑	－間－粘－－	1	4	1	5	2	2
名古屋	0.150	－－含－－塑	湿－含－－塑	湿－含－－塑	－間－－－塑	－－含－－塑	2	1	4	0	0	5
四日市	0.209	－間含－－－	湿－含－－塑	－間含－－－	－－含－液塑	－間含－－－	1	3	5	0	1	2
堺泉北	0.099	湿間－粘液塑	湿間－粘液塑	湿間－－液－	湿間－粘液－	－間含－－塑	4	5	1	3	4	3
和歌山下津	0.124	－－含－－－	－－含粘－塑	－－含－－－	－－含－－塑	－－含－－－	0	0	5	1	0	2
広島	0.120	－間－粘液－	－間－粘－－	湿間－粘－－	－間－粘－－	－間－粘－塑	1	5	0	5	1	1
徳山下松	0.193	－間含－－－	湿間含－－－	－間－－－－	－間含－－－	－間－－－－	1	5	3	0	0	0
下関	0.201	湿－含－－－	－間含－－－	－－含－－－	－－含－液－	－間－粘液塑	1	2	4	1	2	1
北九州	0.152	－－含粘液－	－－含粘液－	湿－－粘液－	－間－粘－塑	－間含粘－－	1	2	3	5	3	1

表-11　ケースb・ElasticNet（線形式・指数式）

港湾	ElasticNet	
	線形	指数
東京	0.101	0.101
横浜	0.139	0.135
川崎	0.099	0.100
神戸	0.088	0.088
千葉	0.140	0.140
名古屋	0.149	0.148
四日市	0.207	0.214
堺泉北	0.094	0.089
和歌山下津	0.127	0.124
広島	0.117	0.119
徳山下松	0.193	0.195
下関	0.185	0.208
北九州	0.147	0.150

表-12　ケースc・線形回帰・ElasticNet（線形式・指数式）

港湾	線形回帰				ElasticNet	
	線形		指数		線形	指数
東京	0.103	－間－－液－	0.100	－間－粘液塑	0.102	0.100
横浜	0.154	－間含－液－	0.154	湿間－粘液塑	0.150	0.154
川崎	0.100	－間－－液－	0.101	湿－－－液－	0.101	0.098
神戸	0.126	湿間－粘液－	0.123	－間－粘液－	0.121	0.109
千葉	0.144	－間含－液－	0.139	湿間－粘液塑	0.140	0.138
名古屋	0.168	－間－－液－	0.164	湿－－－液－	0.173	0.164
四日市	0.237	－間－－液－	0.259	湿－－－液－	0.212	0.226
堺泉北	0.169	－間含－液－	0.174	湿間－粘液塑	0.172	0.180
和歌山下津	0.149	－間－－液－	0.151	－間－粘液塑	0.143	0.153
広島	0.144	湿間－粘液塑	0.136	湿－－粘液塑	0.141	0.137
徳山下松	0.339	湿間－粘液塑	0.295	湿－－粘液塑	0.351	0.309
下関	0.172	湿間－粘液－	0.153	－間－粘液－	0.170	0.150
北九州	0.181	湿間－粘液塑	0.180	湿－－粘液塑	0.180	0.188

は ElasticNet の指数式，誤差の合計値の小ささでは ElasticNet の線形式が最も優秀という結果であった．

6.　機械学習による推定

次に機械学習による推定を検討した．テーブルデータを対象とする機械学習手法は多く存在し，小林はディープニューラルネットワークを用いて圧縮指数などのパラメータの推定を行っている[18]．本研究では，実装の容易さや処理速度の速さなどを理由として，勾配ブースティング決定木を採用した．ライブラリは処理が高速な LightGBM[19]（以下，LGBM）を使用した．

機械学習では ElasticNet と同様にハイパーパラメータのチューニングが必須となる．LGBM には多くのハイパーパラメータが存在するが，**表-13** に示す 9 つを調整することとした．一部のハイパーパラメータ間には相関があるため，それを考慮して設定し，合計 729 通りとした．LGBM では変数の標準化は必要ないが，重回帰分析と条件を揃えるため標準化したデータを使用した．

(1)　検討ケース

第 5 章に示した 3 つのケースのうち，ケース a と c を対象とした．ケース b は港湾によっては使用できるデータが非常に少なく，機械学習は適さないためである．

(2)　結果の整理

LGBM による推定結果を**表-14, 15** に示す．それぞれの表において，重回帰分析の 4 手法より精度の良かったものを赤字で示している．

ケース a では 5/13 港湾で精度が向上した．ある程度の有効性は確認できたが，**図-1** を見ると，今回使用した 6 つの説明変数と圧縮指数はおおよそ線形的な関係にあり，非線形性を考慮できる機械学習の強みをうまく活かせな

表-13　LightGBM のハイパーパラメータ

max_ depth	num_ leaves	min_child_ samples	learning_ rate	n_ estimators	reg_ alpha	reg_ lamda	colsample_ bytree	subsample
3	0.5×2^{max_depth}	0.5×num_leaves	0.01	10/learning_rate	num_leaves≧128：0.1 else：0.0		0.6	1-colsample_bytree
5	0.7×2^{max_depth}	1.0×num_leaves	0.05	20/learning_rate			0.8	
8	1.0×2^{max_depth}	1.5×num_leaves	0.10	30/learning_rate			1.00	

表-14　ケース a・LGBM

港湾	RMSE	max_ depth	num_ leaves	min_child_ samples	learning_ rate	n_ estimators	reg_ alpha	reg_ lamda	colsample_ bytree	subsample
東京	0.101									
横浜	0.152									
川崎	0.093	1:3	$1:0.5\times2^3$	$1:0.5\times:0.5\times2^3$	1:0.05	1:10/0.05	1:0.0	1:0.0	1:0.8	1:0.8
神戸	0.100									
千葉	0.138	2:5	$2:0.5\times2^5$	$2:0.5\times0.5\times2^5$	2:0.05	2:10/0.05	2:0.0	2:0.0	2:0.6	2:1.0
名古屋	0.176									
四日市	0.210	3:3	$3:0.5\times2^3$	$3:1.0\times0.5\times2^3$	3:0.01	3:10/0.01	3:0.0	3:0.0	3:0.6	3:1.0
堺泉北	0.152									
和歌山下津	0.162	4:3	$4:0.5\times2^3$	$4:1.5\times0.5\times2^3$	4:0.01	4:10/0.01	4:0.0	4:0.0	4:0.6	4:1.0
広島	0.137									
徳山下松	0.275	5:3	$5:0.5\times2^3$	$5:1.0\times0.5\times2^3$	5:0.05	5:10/0.05	5:0.0	5:0.0	5:0.6	5:1.0
下関	0.155									
北九州	0.162									

表-15　ケース c・LGBM

港湾	RMSE	max_ depth	num_ leaves	min_child_ samples	learning_ rate	n_ estimators	reg_ alpha	reg_ lamda	colsample_ bytree	subsample
東京	0.098	3	0.5×2^3	$0.5\times0.5\times2^3$	0.05	10/0.05	0.0	0.0	0.6	1.0
横浜	0.167	3	0.5×2^3	$0.5\times0.5\times2^3$	0.05	10/0.05	0.0	0.0	0.6	1.0
川崎	0.092	3	0.5×2^3	$1.0\times0.5\times2^3$	0.05	10/0.05	0.0	0.0	0.6	1.0
神戸	0.128	3	0.5×2^3	$0.5\times0.5\times2^3$	0.01	10/0.01	0.0	0.0	0.6	1.0
千葉	0.141	3	0.5×2^3	$0.5\times0.5\times2^3$	0.01	10/0.01	0.0	0.0	1.0	0.6
名古屋	0.167	3	0.5×2^3	$1.0\times0.5\times2^3$	0.05	10/0.05	0.0	0.0	0.6	1.0
四日市	0.238	3	0.5×2^3	$1.0\times0.5\times2^3$	0.05	10/0.05	0.0	0.0	0.6	1.0
堺泉北	0.174	3	0.5×2^3	$0.5\times0.5\times2^3$	0.01	10/0.01	0.0	0.0	0.6	1.0
和歌山下津	0.155	3	0.5×2^3	$0.5\times0.5\times2^3$	0.05	10/0.05	0.0	0.0	0.6	1.0
広島	0.164	5	0.7×2^5	$1.5\times0.7\times2^5$	0.01	10/0.01	0.0	0.0	1.0	0.6
徳山下松	0.430	5	0.7×2^5	$1.5\times0.7\times2^5$	0.01	10/0.01	0.0	0.0	1.0	0.6
下関	0.160	3	0.5×2^3	$0.5\times0.5\times2^3$	0.01	10/0.01	0.0	0.0	0.6	1.0
北九州	0.187	5	0.7×2^5	$1.5\times0.7\times2^5$	0.01	10/0.01	0.0	0.0	1.0	0.6

かったものと思われる．また**表-14** を見ると，選ばれたハイパーパラメータは狭い範囲に固まっており，探索範囲が不十分であったという懸念がある．本研究ではハイパーパラメータの探索範囲を人為的に設定したが，ハイパーパラメータ最適化フレームワークの Optuna[20]を使用するなどして，より効率的に適切なハイパーパラメータを設定できる可能性がある．Optuna ではベイズ最適化をはじめとした様々な探索アルゴリズムを利用できる．その他，異なる機械学習手法の採用や，使用する説明変数の見直し，およびデータ数の増加などにより，機械学習による推定精度向上の余地はあると考えている．

ケース c で精度向上したのは東京港と川崎港のみであった．これは，一般的に機械学習が外挿データの推定を苦手としているためであると考えられる．

7. おわりに

地域ごとの地盤特性を考慮した圧縮指数の推定方法を構築するため，地盤定数データベースを構築し，港湾ごとに分割したうえで，重回帰分析と機械学習により検討を行った．その成果を下記に示す．

(1) 国土地盤情報検索サイト KuniJiban からデータスクレイピングにより地盤調査データを収集し，テーブルデータ形式の地盤定数データベースを構築した．

(2) 圧縮指数との相関係数を確認し，データ数も考慮したうえで，湿潤密度，間隙比，含水比，粘性土分，液性限界および塑性限界の 6 つを説明変数とした．

(3) 地理情報システムの QGIS 上で港湾計画図と地図データを重ね合わせ，港湾計画図の色付き範囲を網羅できる楕円を作図し，各港湾の領域を設定した．これに基づき地盤定数データベースから港湾ごとのデータを抽出した．

(4) 正則化項を導入しない通常の線形回帰と ElasticNet により重回帰分析を行った．対象港湾のデータが少ない場合，対象港湾に十分なデータが存在する場合，および対象港湾のデータが全くない場合に分けて検討した．ほとんどの場合で，単一の説明変数を用いる従来式よりも高い精度で圧縮指数を推定することができた．また，通常の線形回帰と比較して ElasticNet が有効であることを確認した．

(5) 機械学習手法の一つである LightGBM による検討を行った．本研究の範囲では ElasticNet ほどの有効性は確認できなかった．

推定精度向上のための今後の課題を下記に示す．

(1) 目的変数との相関係数だけでなく，説明変数間の VIF を考慮し，多重共線性のない最適な説明変数の組合せを検討する．

(2) 本研究では国際戦略港湾と国際拠点港湾のデータのみを用いて検討を行った．これに他港湾や陸上のデータを結合し，データの全体的な傾向を有効活用できる階層ベイズモデルを検討する．

(3) ベイズ最適化などの様々な探索アルゴリズムを利用できるハイパーパラメータ最適化フレームワークの Optuna を使用するなどし，ハイパーパラメータの最適化を検討する．

(4) KuniJiban 以外の地盤情報データベースも活用し，データ数を増加させる．

(5) LightGBM 以外の機械学習手法を検討する．

(6) 地理や地史を踏まえた考察を行い，地域ごとの地盤特性を明らかにする．

最後に，本研究で構築した地盤定数データベースは，圧縮指数の推定に限らず様々な用途で活用することができる．今後，本データベースをより実務的なタスクに適用していきたいと考えている．

参考文献

1) 公益社団法人日本港湾協会：港湾の施設の技術上の基準・同解説，pp.318-319，2018.
2) 公益社団法人日本港湾協会：港湾の施設の技術上の基準・同解説，p.325，2018.
3) 渡辺進，井本美晴：初期間隙比から圧密沈下量を求める方法，土と基礎，Vol.10，No.9，pp.7-11，1962.
4) 鬼塚克忠，吉武茂樹：有明粘土の圧縮指数と物理的性質の相関，土と基礎，Vol.33，No.4，pp.57-59，1985.
5) 田中洋行，榊原基生：港湾地域における土の一次性質の統計解析，港湾技研資料，No.719，1991.
6) 田中政典：港湾における土質データベースの開発とその活用について，地質ニュース，No.675，pp.19-24，2010.
7) 山本浩司，濱田晃之，三村衛：地盤情報データベースの地盤変形解析への適用例，第 39 回地盤工学研究発表会，pp.185-186，2004.
8) 北岡貴文，山本雄平，水谷未来，小林泰三：地盤データベースの利活用における ANN による三軸圧縮試験の内部摩擦角推定の試み，AI・データサイエンス論文集，Vol.3，No.2，pp.17-22，2022.
9) 国土交通省，国立研究開発法人土木研究所，国立研究開発法人港湾空港技術研究所：国土地盤情報検索サイト KuniJiban，https://www.kunijiban.pwri.go.jp/jp/index.html（2025 年 6 月 1 日確認）
10) seaborn-analyzer : https://pypi.org/project/seaborn-analyzer/（2025 年 6 月 1 日確認）
11) O'Brien, R. M.: A Caution Regarding Rules of Thumb for Variance Inflation Factors, Quality & Quantity, Vol.41(5), pp.673-690, 2007.
12) 国土交通白書 2020 資料編：https://www.mlit.go.jp/hakusyo/mlit/r01/hakusho/r02/data/html/ns012020.html（2025 年 8 月 15 日確認）
13) 東京港港湾計画図：https://www.kouwan.metro.tokyo.lg.jp/documents/d/kouwan/22636_kouwan_keikakuzu（2025 年 6 月 1 日確認）
14) 基盤地図情報ダウンロードサービス：https://service.gsi.go.jp/kiban/（2025 年 6 月 1 日確認）
15) QGIS：https://qgis.org/（2025 年 6 月 1 日確認）
16) Shapely：https://shapely.readthedocs.io/en/stable/（2025 年 8 月 17 日確認）
17) Zhou, H. and Hastie, T.: Regularization and variable selection via the ElasticNet, Journal of the Royal Statistical Society, Series B, Vol.67, Issue 2, pp.301-320, 2005.
18) 小林泰三：深層学習で土質特性に迫る，地盤工学会誌，Vol.70，No.7，Ser.No.774，pp.5-12，2022.
19) Guolin, Ke., Qi, Meng., Thomas, Finely., Taifeng, Wang., Wei, Chen., Weidong, Ma., Qiwei, Ye. and Tie-Yan, Liu.: LightGBM: A Highly Efficient Gradient Boosting Decision Tree, Advances in Neural Information Processing Systems 30, 2017.
20) Optuna：https://optuna.org/（2025 年 8 月 17 日確認）

(Received June 30, 2025)
(Accepted August 31, 2025)

Construction of soil database and estimation of compression index of cohesive soil deposited in coastal area

Youhei KATAYAMA and Kiyonobu KASAMA

Compression index of the soil is necessary for accurate estimation of consolidation settlement. Compression index can be obtained by a consolidation test, but in practice, it is often estimated from liquid limit or void ratio because it is a time-consuming test. Various empirical formulas have been proposed to estimate compression index, but most of them do not take into account the geotechnical characteristics of each region, which may lead to large estimation errors if applied easily. This study aims to construct an estimation method for compression index that takes into account the geotechnical characteristics of each region. First, a soil database in the form of table data was constructed based on boring data available on the Web. The data was divided by major ports in Japan, and an estimation model was constructed by multiple regression analysis and machine learning to estimate compression index.