Transformation models for the compressibility properties of Finnish clays using a multivariate database

Monica Susanne Löfman and Leena Katariina Korkiala-Tanttu

Department of Civil Engineering, Aalto University, Espoo, Finland

ABSTRACT

The compression index $C_c$, which is defined as the decrease in void ratio per tenfold increase in vertical stress, is often used to estimate the compression of clayey subsoil. $C_c$ is obtained using an oedometer test performed on an undisturbed specimen. However, as oedometer tests require good-quality specimens and an incremental loading oedometer test usually lasts several days, compressibility parameters are rarely available when the preliminary geotechnical design is being done. Consequently, many researchers have presented empirical regression equations (i.e., transformation models) to estimate $C_c$ from other properties such as water content ($w_L$) or liquid limit ($w_n$) (Skempton 1944; Helenelund 1951; Hough 1957; Moran and Mueser 1958; Nishida 1956; Cozzolino 1961; Terzaghi and Peck 1967; El-Naggar and Krizek 1970; Azzouz, Krizek, and Corotis 1976; Ogawa and Matsumoto 1978; Mayne 1980; Koppula 1981; Saarelainen 1981; Al-Khafaji and Andersland 1992; Kootah and Moradi 2017; Park and Lee 2011; Di Buò et al. 2019).

Regarding Finnish fine-grained soils, very few transformation models are available for compressibility. Most Finnish geotechnical handbooks present only two models: Helenelund’s (1951) non-linear empirical correlation for compressibility index using $w_L$ and Janbu’s (1988) model, which links the modulus number to $w_n$. However, neither of these transformation models were derived specifically for Finnish clays, and they lack the statistical measures required for use in reliability analyses. Because these transformation models for compressibility are characterised by a rather large transformation uncertainty, it is essential that they can be easily defined in probabilistic terms (e.g., Phoon 2017). Recent studies on the transformation models for compressibility of Finnish clays have been conducted, but they are characterised by a rather small number of distinct sites or a small geographical area. For instance, Saarelainen (1978, 1981) created linear regression models for municipal data (Helsinki capital area) and Di Buò et al. (2019) studied the empirical correlations for the compressibility of soft sensitive clays from four sites in Finland.

Updated transformation models for the compressibility of Finnish clayey soils are clearly needed. Thus, the main objectives of this paper are: (i) to compile a large and versatile database of compressibility of Finnish clays and gyttja soils; (ii) to compare the existing transformation models with the compiled database; and (iii) to derive new transformation models for the...
compressibility of Finnish clayey soils (namely, \( C_c \), compression ratio and swelling index).

In this paper a regional database of Finnish clayey soils consisting of various geotechnical properties is presented. The compiled database contains 856 oedometer test results from over 30 sites. The database is partially multivariate, including up to 14 parameters of interest; for instance, almost half of the oedometer tests were accompanied by the undrained shear strength of an adjacent specimen. The database is called FI-CLAY/14/856, following the name format recommended by TC304 (http://140.112.12.21/issmge/tc304.htm) and Phoon (2019). To the authors’ knowledge, this is the largest database of oedometer tests on Finnish clays to date. Another extensive multivariate database of Finnish clays was compiled by D’Ignazio et al. (2016), but it does not include compressibility properties. One practical benefit of multivariate databases is that they can be used to derive multivariate distributions. Effectively, the uncertainty in one geotechnical parameter is reduced when one or more additional parameters are available (Müller, Larsson, and Spross 2014; Ching and Phoon 2014a; Phoon et al. 2016).

This paper is organised as follows. Firstly, a brief overview of the existing transformation models for compressibility is presented. Secondly, the compiled database of Finnish clays FI-CLAY/14/856 is described and compared to existing transformation models. Next, any potential outliers are removed and new transformation models for compressibility are derived for FI-CLAY/14/852 using polynomial regression applied to the natural logarithms of the soil properties. Finally, the transformation uncertainty is evaluated using statistical measures.

2. Overview of the existing transformation models for compressibility

2.1. Measures of compressibility

\( C_c \) is commonly used to calculate one-dimensional vertical strain, which is defined as the change in soil layer thickness divided by the layer thickness. In the case of a normally consolidated (NC) clay layer, in which preconsolidation pressure \( \sigma_p \) equals effective in-situ stress \( \sigma_{vo} \), the strain \( \varepsilon_{NC} \) is calculated using Equation (1):

\[
\varepsilon_{NC} = \frac{C_c}{1 + e_0} \log_{10} \left( \frac{\sigma_{vo} + \Delta \sigma_s}{\sigma_{vo}} \right)
\]

where \( \Delta \sigma_s \) is the added vertical load and \( e_0 \) is the initial void ratio. The term \( C_c/(1+e_0) \), known as the compression ratio (CR), is considered by some researchers to be a better measure of compressibility than \( C_c \) (e.g. Janbu 1988; Wesley 1988). Similarly, for overconsolidated (OC) clays (i.e. \( \sigma_p' \) greater than \( \sigma_{vo}' \)), the strain \( \varepsilon_{OC+NC} \) is calculated using Equation (2):

\[
\varepsilon_{OC+NC} = \frac{C_c}{1 + e_0} \log_{10} \left( \frac{\sigma_p'}{\sigma_{vo}'} \right) + \frac{C_c}{1 + e_0} \log_{10} \left( \frac{\sigma_{vo}' + \Delta \sigma_s}{\sigma_p'} \right)
\]

where \( C_c \) is the swelling index, which is defined as the slope of the swelling line (or recompression line). By dividing \( C_c \) and \( C_s \) by \( \ln(10) \) (\( \approx 2.3 \)), one acquires approximations of the Cam-Clay parameters \( \lambda \) and \( \kappa \), respectively (\( \ln[10] \) accounts for conversion to natural logarithm). The typical shape of oedometer curve for soft Finnish clays together with the interpretations for both \( C_s \) and \( C_c \) is shown in Figure 1. The stress–strain behaviour of many marine Finnish clays is so non-linear in semi-logarithmic scales that the \( C_c \) method cannot be applied without precautions (e.g. Tanaka 2002; Di Buò et al. 2019). The common practice is to determine \( C_c \) using two load increments after the preconsolidation pressure, which usually corresponds to the relevant stresses in a practical geotechnical design.

Due to the potential limitations of \( C_c \), the Ohdhe-Janbu method is often used in Finland instead. In this model, the strain \( \varepsilon_{NC} \) caused by the stress increase from preconsolidation stress \( \sigma_p' \) to final stress \( (\sigma_{vo}' + \Delta \sigma_s) \) was calculated using Equation (3) (Janbu 1963; Helenelund 1974):

\[
\varepsilon_{NC} = \begin{cases} 
\frac{1}{m_1 \beta_1} \left[ \left( \frac{\sigma_{vo}' + \Delta \sigma_s}{\sigma_{ref}} \right)^{\beta_1} - \left( \frac{\sigma_p'}{\sigma_{ref}} \right)^{\beta_1} \right] & (\text{for } \beta_1 \neq 0) \\
\frac{1}{m_1} \ln \left( \frac{\sigma_{vo}' + \Delta \sigma_s}{\sigma_p'} \right) & (\text{for } \beta_1 = 0) 
\end{cases}
\]

In Equation (3), modulus number \( m_1 \) and stress exponent \( \beta_1 \) are the parameters defined from oedometer tests by means of curve fitting and \( \sigma_{ref} \) is the reference...
stress (100 kPa). If $\beta_1 = 0$, the stress–strain curve is linear in a semi-logarithmic scale, as in the $C_c$ concept. As a result, $C_c$ equals $(1+e_0)\ln(10)$ divided by $m_t$. Equation (3) can also be used to calculate the strain in an overconsolidated state $e_{OC}$ (stress states from $\sigma_{vo}$ to $\sigma_{p}$), in which case the parameters are usually marked with subscript “2” ($m_2$ and $\beta_2$ are defined using the unloading- or reloading-curve of the oedometer test). Although Ohde–Janbu is usually more accurate for modelling the compressibility of soft Finnish clays (due to closer replication of the oedometer curve, see Figure 1), its parameters cannot be used as the target parameters of simple transformation models due to their correlation; the Pearson’s coefficient of correlation is $\rho = 0.40$ for $m_1$ and $\beta_1$ (see Table S2 in the online supplementary material).

### 2.2. Existing transformation models for compressibility

Most of the existing transformation models for compressibility of clays are simple linear regression models with a single predictor. The earliest models for $C_c$ were constructed using $w_L$ as the predictor (Skempton 1944). However, subsequent studies showed that using the initial void ratio $e_0$ as the predictor leads to a higher $R^2$ value (coefficient of determination) (e.g. Azzouz, Krizek, and Corotis 1976; Al-Khafaji and Andersland 1992; Koo- tahi and Moradi 2017). Similarly, $w_n$ is associated with high $R^2$, although Al-Khafaji and Andersland (1992) pointed out that many studies have reported $w_n$ values defined for partially saturated soil specimens (rather than fully saturated), which in turn can lead to weaker model accuracy. In Finland, the specimens used in an oedometer test are almost always fully saturated; consequently, $w_n$ is often selected as the predictor for $C_c$ (Hele- nelund 1951; Saarelainen 1981, 1978; Di Buò et al. 2019). Indeed, Di Buò et al. (2019) observed that the relationship between $C_c$ and $w_L$, plasticity index $PI$, or clay content is characterised by larger scatter compared to water content $w_{nr}$. Likewise, Koo- tahi and Moradi (2017) argued that single-predictor models incorporating index properties (i.e. $w_L$ or $PI$) are characterised by larger scatter due to greater measurement uncertainty compared to physical properties such as $w_n$ or $e_0$.

Many researchers recommend using a two-predictor model with void ratio $e_0$ (or $w_n$) combined with $w_L$ (or $PI$) to account for both the structure and mineralogical effects on clay compressibility (e.g. Al-Khafaji and Andersland 1992; Koo- tahi and Moradi 2017). That is, the mineralogical composition and particle-surface characteristics can be represented by intrinsic index properties such as $w_L$ or $PI$, whereas the physical properties ($e_0$ and $w_n$) represent the structure induced by the sedimentation environment. Hence, depending on the clay deposit characteristics, adding $w_L$ or $PI$ as an additional predictor may increase the model accuracy. On the other hand, Azzouz, Krizek, and Corotis (1976) concluded that including $w_L$ as an additional predictor did not significantly increase the accuracy of $C_c$ or $CR$ predictions.

Besides $w_L$ or $PI$, other additional predictors has been suggested as well. For instance, Leroueil, Tavenas, and Bihan (1983) suggested adding sensitivity $S_t$ to the relation between $e_0$ and $C_c$. Al-Khafaji and Andersland (1992) argued that the multiple parameter models are more accurate especially for organic clays, as earlier studies (e.g. Al-Khafaji 1979) have showed that the $C_c$ of organic soils is highly dependent on organic content and preconsolidation pressure. Some existing transformation models for compressibility (i.e. $C_c$, CR and $C_r$) are presented later in Table 2.

### 3. Compiled clay database (FI-CLAY/14/856)

#### 3.1. Collected multivariate data

The database of Finnish clays (FI-CLAY/14/856) compiled for this study contains mostly glaciomarine fine-grained soils with some specimens from glaciolacustrine origins. The full database can be found in the online supplementary material (Table S1). The compiled database was based on two sources: Subset A (Table A1 in the Appendix), which consists of multivariate data from literature, and Subset B (Table A2), which includes data collected for this study specifically (35% of all data). Subset A consists of three sites from Gardemeister (1973) (3% of all data), data compiled by the City of Vantaa (Pajunen 1976) (13% of all data), and the RITA database (Rekonen and Lojander 1999) (49% of all data). The RITA database consists of oedometer tests performed and analysed by the geotechnical research group at the Helsinki University of Technology (now Aalto University).

Most sites in the compiled database FI-CLAY/14/856 represent small areas or test embankments; in other words, the distance between sampling profiles is in the scale of 3–30 m. At the Suurelpo (Subset B) and Vantaa sites (Subset A; Pajunen 1976) however, the maximum distance between the sampling profiles was approximately 1–3 kilometres. The Söderkulla-Nikkilä (RITA database) specimens were collected as a part of a road renovation project, so their maximum separation distance is up to 10 kilometres.

Basic statistics for the parameters of interest (14 soil properties) are presented in Table 1. Each observation (table row) in the FI-CLAY/14/856 database represents
soil properties at a specific depth and location. Each data point \((n = 856)\) has at least three soil properties: \(w_{\text{r}}\), \(e_0\) and \(C_c\). Other data points have additional soil properties to a varying extent; in other words, the database is partly multivariate. For example, Table 1 shows that 413 data points contained both the undrained shear strength \(s_u\) (determined from a fall cone test) and \(C_c\) (i.e. the oedometer test result). Table 1 also shows the corresponding soil property statistics for a global database of marine fine-grained soils compiled by Kootahi and Moradi (2017). These databases are also compared in Figure 6. Table 1 shows that the main differences between the regional and global databases are the following: the global database contains higher values of \(w_{\text{r}}\), \(e_0\) and PI whereas the FI-CLAY/14/856 contains higher values of \(C_c\), CR and \(S_f\). These Finnish sensitive clays may be one explanation for higher values of \(C_c\) (e.g. Leroueil, Tavenas, and Bihan 1983; Di Buò et al. 2019).

Data from Gardemeister (1973) and Pajunen (1976) were originally fully multivariate, and the RITA database was partially multivariate. For the remaining data, multivariate data points were interpreted for this study. The oedometer test results and other properties (classification and strength properties) were combined if both of the following applied: (i) the vertical distance between specimens taken from the same sampling point was not more than 100 mm; and (ii) the \(w_{\text{r}}\) of the oedometer test specimen and the other specimen did not differ more than 10 percentage points. It should be noted that as some of the oedometer tests were performed on specimens within a 100 mm section, there are a few multivariate rows which include duplicate values of the other property (e.g. plastic limit).

### 3.2. Description of the database

#### 3.2.1. Soil types

In the FI-CLAY/14/856 database, the specimens were classified according to the Finnish geotechnical classification system (Korhonen, Gardemeister, and Tammimie 1974). In this system, the grain size distribution and organic content defined the soil name, meaning that plasticity was not considered and thus many of the data points lack consistency limits. The naming convention based on clay fraction (grain size smaller than 0.002 mm) and organic content are presented in Figure 2(a), together with the number of each soil type in the database. The pie chart in Figure 2(a) shows that most of the specimens were clays. For some specimens (i.e. “fine-grained soils”), the soil type was not reported.

The specimens in the database which were accompanied with consistency limits and complete soil classification are plotted in the plasticity chart in Figure 2(b). According to the Unified Soil Classification System (ASTM D 2487-06), most of the specimens (Fat clays, Organic clays, Organic silts, and Gyttja) lie above the A-line, i.e. they represent CH or OH types. On the

---

**Table 1.** Basic statistics of the database FI-CLAY/14/856 and comparison to a global database.

| Soil property          | FI-CLAY/14/856 | Global database Abb | N | Min | 25% | Median | 75% | Max | Mean | COV |
|------------------------|----------------|---------------------|---|-----|-----|--------|-----|-----|------|-----|
| \(w_r\) (%)            | 856            | 500                 | 23.0 | 60.0 | 85.0 | 104    | 187 | 84.5 | 333  |
| \(e_0\), Global database Abb | 500 | 10 | 41.0 | 66.5 | 97.0 | 384    | 76.1 | 0.63 |
| \(e_0\) Global database Abb | 856 | 0.520 | 1.632 | 2.302 | 2.850 | 5.200 | 2.309 | 0.36 |
| \(S_f\), Global database Abb | 500 | 0.280  | 1.11 | 1.800 | 2.583 | 8.80  | 1.992 | 0.59 |
| \(F\) (%)              | 206            | 40.5               | 59.0 | 67.6 | 84.0 | 197.5 | 71.3 | 0.40 |
| \(w_l\) (%)            | 354            | 19.0              | 49.7 | 69.0 | 86.0 | 203.9 | 69.7 | 0.39 |
| \(w_l\), Global database Abb | 500 | 19.0  | 43.0 | 68.0 | 91.3 | 395.0 | 76.6 | 0.61 |
| \(w_w\) (%)            | 370            | 12.0              | 23.7 | 26.6 | 30.0 | 55.6  | 27.3 | 0.22 |
| \(P\) (%)              | 354            | 1.0               | 25.6 | 41.5 | 57.0 | 153.7 | 42.5 | 0.54 |
| \(P\), Global database Abb | 500 | 2.00 | 19.0 | 37.0 | 54.5 | 218.0 | 41.4 | 0.69 |
| \(\gamma\) (kN/m³)     | 799            | 12.0              | 14.2 | 15.0 | 16.3 | 20.6  | 15.3 | 0.10 |
| \(\sigma_0\) (kPa)     | 369            | 0.0               | 0.0  | 0.8  | 1.6  | 7.8   | 1.2  | 1.26 |
| \(\sigma_0\) Abb       | 263            | 12.7              | 40.9 | 54.5 | 73.0 | 95.0  | 56.6 | 0.36 |
| \(C_r\) (%)            | 413            | 5.2               | 15.6 | 22.5 | 31.5 | 240.0 | 102  | 1.01 |
| \(S_p\) (%)            | 281            | 1.7               | 11.3 | 17.0 | 29.7 | 163.0 | 23.2 | 0.84 |
| \(S_p\), Global database Abb | N/A | 1.0  | N/A  | N/A  | N/A  | N/A   | 8.0  | N/A |
| \(\sigma_0\) (kPa) Abb | 843            | 6.0               | 30.0 | 49.0 | 83.0 | 604.1 | 66.7 | 0.93 |
| OCR                    | 817            | 0.3               | 1.0  | 1.4  | 2.1  | 41.3  | 2.2  | 1.32 |
| OCR, Global database Abb | N/A | 0.7  | N/A  | N/A  | N/A  | N/A   | 2.4  | N/A |
| \(C_c\) (%)            | 856            | 0.026             | 0.657 | 1.287 | 1.977 | 9.284 | 1.417 | 0.70 |
| \(C_c\), Global database Abb | 500 | 0.080  | 0.380 | 0.730 | 1.150 | 4.500 | 0.847 | 0.73 |
| CR                     | 856            | 0.014             | 0.242 | 0.378 | 0.519 | 2.088 | 0.394 | 0.56 |
| CR, Global database Abb | 500 | 0.052 | 0.177 | 0.260 | 0.329 | 0.495 | 0.254 | 0.39 |
| \(C_j\)                | 656            | 0.002             | 0.064 | 0.101 | 0.148 | 0.611 | 0.109 | 0.55 |

aSample percentiles.

bGlobal database of marine soils compiled by Kootahi and Moradi (2017) for model building.

Notes: \(n\) = number of data points; \(\text{COV}\) = coefficient of variation; \(w_{\text{r}}\) = natural water content (%); \(e_0\) = initial void ratio; \(w_l\) = liquid limit (%); \(F\) = fineness number (%); \(w_p\) = plastic limit (%); \(\gamma\) = unit weight (kN/m³); \(\sigma_0\) = preconsolidation pressure (kPa); OCR = over-consolidation ratio; \(C_c\) = compression index; \(C_r\) = compression ratio; \(C_j\) = swelling index.
other hand, the majority of Lean clays and Clayey silts represent CL or OL types, and a few Clayey silts fall into the CL-ML type.

Finally, Figure 2(c) represents the $C_r(w_n)$ relationship for different soil types. Gyttja specimens (including Clayey Gyytja) behaved somewhat differently compared to other soil types, but there was also overlap with other soil types (i.e. it was not necessary to remove the Gyttja specimens from the data).

### 3.2.2. Oedometer test types

Most of the data ($n = 723$) represent incrementally loaded (IL) standard oedometer tests with a 24-hour load sequence and loading which is doubled for each increment. However, the data also include some speciality oedometer test types, which are shown in the $C_r(w_n)$ graph presented in Figure 3: “horizontal” tests were performed on specimens rotated 90 degrees, “permeability” IL tests included a falling head test performed at selected load increments and “creep” tests contained longer-than-usual load increments. However, for most of these creep tests, the strains at 24 h were utilised instead of the final values. Besides 24-hour values, most of the recent oedometer test curves were constructed using the end-of-primary (EOP) values; those of the normal compression line (Figure 1) were estimated by means of consolidation time analysis (e.g. the square root time method) in order to minimise the effects of creep on the interpreted parameters. Figure 4 illustrates these EOP results ($n = 128$) in a $C_r(w_n)$ graph. All oedometer tests in Subset B sites except the Otaniemi test site and the POKO site represented EOP results.

Figure 3 also illustrates the constant rate of strain (CRS) oedometer tests ($n = 133$). Three sites contained

---

**Table 2. Existing models for compressibility and their calibration results for FI-CLAY/14/856.**

| Reference | Applicability/soils included in the study | Transformation model equation | Calibration results |
|-----------|----------------------------------------|-------------------------------|--------------------|
| Helenelund (1951) | Finnish and Danish clays, silts, and organic soils ($a = 0.85$ suggested from range $a = 0.5 - 1.5$) | $C_r = a \left(\frac{w_n}{100}\right)^{1.5} ; a = 0.85$ | $n = 856$ $b = 1.97$ $\delta = 0.44$ |
| Saarelainen (1981) | Finnish (Helsinki area) silts, clays, and organic soils | $C_r = a \left(\frac{w_n}{100}\right)^{1.5} ; a = 1.50$ | $n = 856$ $b = 1.68$ $\delta = 0.47$ |
| Di Buò et al. (2019) | Finnish soft sensitive clays (for stress ranges preconsolidation pressure +10/20/50 kPa) | $C_{r(CM+M)} = 0.000102(w_n)^{2.284}$ | $n = 856$ $b = 5.04$ $\delta = 0.51$ |
| Moran and Mueser (1958) | Organic soils, peats, organic silts and clays | $C_r = 0.0115 w_n$ | $n = 856$ $b = 1.34$ $\delta = 0.49$ |
| Nakase, Kamei, and Kusakabe (1988) | Reconstituted Japanese marine clays | $C_r = 0.501(w_n - 0.33) + 0.012(w_n - 8)$ | $n = 354$ $b = 1.50$ $\delta = 0.64$ |
| Nishida (1956) | All clays | $C_r = 0.0155(w_n - 19)$ | $n = 354$ $b = 1.43$ $\delta = 0.80$ |
| Ogawa and Matsumoto (1978) | Japanese marine clays | $C_r = 0.020(PI + 0.65)$ | $n = 354$ $b = 0.73$ $\delta = 0.43$ |
| Andersen (2012) | Danish clays, silts and organic soils | $CR(%) = 28.9 \log_{10}\left(\frac{w_n}{11.4}\right)$ | $n = 354$ $b = 1.16$ $\delta = 0.46$ |
| Janbu (1988) | All normally consolidated clays | $CR = \frac{\ln 10}{700} \cdot w_n$ | $n = 856$ $b = 1.40$ $\delta = 0.42$ |
| Elnaggar and Krizek (1979) | Various clays and silts | $CR = 0.156e_0 + 0.0107$ | $n = 856$ $b = 1.04$ $\delta = 0.42$ |
| Di Buò et al. (2019) | Finnish soft sensitive clays | $C_r = 0.0113(w_n)^{0.67}$ | $n = 656$ $b = 0.49$ $\delta = 0.44$ |
| Nakase, Kamei, and Kusakabe (1988) | Reconstituted Japanese marine clays | $C_r = 0.001934(PI - 4.6)$ | $n = 258$ $b = 1.99$ $\delta = 1.14$ |
| Kootahi (2017) | Various overconsolidated fine-grained soils | $C_r = 0.0014(w_n - 10)$ | $n = 656$ $b = 1.09$ $\delta = 0.43$ |
| Kulhawy and Mayne (1990) | All clays | $C_r = 0.0540(e_0 - 0.3)$ | $n = 656$ $b = 1.05$ $\delta = 0.41$ |
| Tiwari and Ajmera (2011) | Reconstituted montmorillonite dominated soils | $C_r = 0.0017(w_n - 21)$ | $n = 265$ $b = 1.68$ $\delta = 1.13$ |
| Kootahi (2017) | Various overconsolidated fine-grained soils | $C_r = 0.0025(PI - 3)$ | $n = 264$ $b = 1.64$ $\delta = 1.47$ |
| Tiwari and Ajmera (2011) | Reconstituted montmorillonite dominated soils | $C_r = 0.0210(e_0 + 0.06w_n - 1.60)$ | $n = 267$ $b = 1.16$ $\delta = 0.52$ |
| Kulhawy and Mayne (1990) | All clays | $C_r = 0.0330(e_0 + 0.04PI - 0.55)$ | $n = 266$ $b = 1.05$ $\delta = 0.44$ |
| Tiwari and Ajmera (2011) | Reconstituted montmorillonite dominated soils | $C_r = 0.0027PI^2$ | $n = 267$ $b = 1.18$ $\delta = 0.87$ |
| Kulhawy and Mayne (1990) | All clays | $C_r = 0.00087(e_0)^2 + 0.031(e_0)$ | $n = 656$ $b = 0.96$ $\delta = 0.43$ |
| Tiwari and Ajmera (2011) | Reconstituted montmorillonite dominated soils | $C_r = 6 \cdot 10^{-9}(PI)^2 + 0.0026(PI)$ | $n = 267$ $b = 1.15$ $\delta = 0.93$ |

Notes: $n =$ number of samples; $b =$ mean bias; $\delta =$ COV of the bias values; $C_r =$ compression index; $CR =$ compression ratio $C_r/(1 + e_0)$; $C_s =$ Swelling index; $w_n =$ natural water content ($\%$); $e_0 =$ initial void ratio; $w_s =$ liquid limit ($\%$); $PI =$ plasticity index.
CRS tests: Haarajoki \((n = 34)\), Söderkulla-Nikkilä \((n = 52)\) and Perniö test embankment \((n = 47)\). The applied strain rate varied greatly due to rate-dependency studies \((0.03–15\%/hour)\). The OCR values of CRS test specimens were calculated using the apparent preconsolidation pressure (i.e. no rate correction was applied). Figures 3 and 4 demonstrate that none of these non-standard oedometer tests formed their own distinct populations and therefore they may be used to derive the transformation models for compressibility. It should be noted that Figure 4 also illustrates specimens with low OCR which are further discussed in “Removal of outliers” section.

### 3.2.3. Applied approximations

As explained above, in Finland the Ohde-Janbu method is generally preferred over \(C_v\), so many of the oedometer tests lack interpretations of \(C_v\). For these tests \(C_v\) was estimated using the Ohde-Janbu parameters \(m_1\) and

![Figure 2](image-url)  
*Figure 2.* (a). Pie chart of the various soil types. (b). Plasticity chart and different soil types. (c). The \(C_v(w_a)\) relationship of different soil types.

![Figure 3](image-url)  
*Figure 3.* The different oedometer test types.
This approximation was performed using the method proposed by Löfman and Korkiala-Tanttu (2019) and applied to the Subset A data and all the CRS data from the RITA database ("$C_\varepsilon$ estimated" in Figure 4).

In Gardemeister’s (1973) data, $e_0$ was approximated by assuming full saturation and density of solid particles $\rho_s = 2.7 \text{ g/cm}^3$ (i.e. $e_0 = 2.7(\omega_n/100)$). Because these specimens were taken from depths below the groundwater level, the estimation based on $\omega_n$ is quite accurate. Figure S1 in the online supplementary material shows that the relationship between $\omega_n$ and $e_0$ is linear with little scatter. Accordingly, the Pearson’s coefficient of correlation for $\omega_n$ and $e_0$ was found to be very high ($\rho = 0.98$) (see Table S2 in the online supplementary material).

Finally, some data points contained fineness number $F$. This parameter was determined from the fall cone test on remolded specimens, and the result was taken as the approximation of $\omega_L$ determined using the Casagrande test (e.g. Gardemeister 1975). For many specimens, both parameters were defined, and it was verified that their relationship was linear and almost non-biased (see Figure S2 in the online supplementary material). Likewise, the $\rho$ for $\omega_L$ and $F$ was found to be as high as 0.98 (see Table S2 in the online supplementary material).

4. Comparison of existing transformation models to FI-CLAY/14/856

In this section, the database is compared to existing transformation models for compressibility. The bias factors and transformation uncertainties are evaluated for the selected existing models. The bias $b_i$ of ith data point can be defined as the actual target value ($C_\sigma$, $CR$, or $C_s$) divided by the predicted target value. The bias factor $b$ is the sample mean of all biases $b_i$. Thus, if $b = 1$, the model is unbiased. The sample coefficient of variation of values $b_i$ is the COV value of the transformation model, denoted by $\delta$. According to of Ching and Phoon (2014b), the variability term $\varepsilon$ for a transformation model can calculated using Equation (4):

$$\varepsilon = \frac{b \cdot \text{predicted target value}}{\text{unbiased prediction}}$$

The mean of $\varepsilon$ is 1 and its sample coefficient of variation describes the deviation between the actual target value and the unbiased prediction. The COV of $\varepsilon$ corresponds to $\delta$. If $\delta = 0$, there is zero data scatter about the transformation model, i.e. there is zero transformation uncertainty.

The bias factors $b$ and $\delta$ (COV of $\varepsilon$) for the selected existing models are collected to Table 2. The selected models are also illustrated in Figures 5–9 together with the database values. Transformation models which have been found to be inapplicable to marine fine-grained soils (Kootahi and Moradi 2017) have been omitted from the calibration. In addition, predicted target values, which were either negative or smaller than the observed minimum value, were removed from calibration data. Calibration results in Table 2 show that most of the existing transformation models compressibility had bias factors greater than 1, meaning that the target value was generally underestimated.
However, models with multiple predictors had smaller biases on average. Next, each major predictor for compressibility is discussed in detail.

Figure 5 illustrates that the existing $C_c(w_m)$-transformation models performed well for soil specimens with low $w_m$ (less than 75%), whereas all the existing models except Di Buò et al.’s (2019) models for Finnish sensitive clays underestimate $C_c$ for specimens with higher $w_m$. Furthermore, the upper range reported by Helenelund (1951) overlapped the data. The overestimation by Di Buò et al.’s (2019) models is due to different interpretation method applied ($C_c$ interpreted from small stress increments after the preconsolidation pressure in a CRS test) and due to high sensitivity of the studied clay specimens. In fact, as discussed later, some of the high $C_c$ values in FI-CLAY/14/856 represent sensitive Perniö clay included in both of these datasets (i.e. different tests results for the same clay layer).

A similar under-prediction of $C_c$ was observed using transformation models with $e_0$ (Figure 6) and $w_L$ (Figure 7). Figure 7 also includes the above-mentioned fineness number $F$, which can be considered as the approximation of $w_L$. Moreover, the relationship between $C_c$ and $w_L$ is characterised by significant scatter, which also explains the high transformation...
uncertainties (δ in Table 2). Table 2 shows that neither \( w_t \) nor \( PI \) should be used as the single predictor of compressibility for FI-CLAY/14/856 due to high values of δ. Finally, Figure 8 shows the data together with the selected transformation models for \( CR \). As observed above, the existing models performed well at lower \( w_n \) values but their biases increased with \( w_n \).

The existing models for estimating \( C_s \) with \( w_n \) are illustrated in Figure 9, which shows that the existing model for Finnish sensitive clays provided by Di Buò et al. (2019) overestimated \( C_s \). The reason for this overestimation could be the testing conditions, as Di Buò et al. (2019) actually defined the initial recompression index (prior surpassing the preconsolidation pressure) rather than \( C_s \) at unloading-reloading. Moreover, they observed that the coefficient of determination between \( w_n \) and the recompression index was very weak (\( R^2 = 0.16 \)). Similarly, the \( C_s \) data (FI-CLAY/14/856) is marked with greater scatter compared to \( C_s \) and \( CR \). Nevertheless, the model suggested by Kootahi (2017) fits rather well with the data, and the bias factor \( b \) is approximately 1 (Table 2). On contrary, the relationship between \( PI \) and \( C_s \) was characterised by larger scatter (see Figure S3 in the supplement); accordingly, calibration yielded very high δ values.

5. New transformation models for compressibility (FI-CLAY/14/822)

5.1. The polynomial regression model

The calibration studies above demonstrated that most of the existing transformation models underestimate the compressibility of Finnish clays. Moreover, as the scatter increases with predictor value (e.g. Figure 5), a simple linear regression model cannot provide a robust definition of the transformation uncertainty. Hence, a 2-degree polynomial regression model applied to the natural logarithms of the soil properties (i.e. the target and the predictors) was selected. The polynomial model, which is an extension of the linear regression, was fitted using the ordinary least squares (OLS) method within the Scikit-learn library for Python (Pedregosa et al. 2011). The transformation model (a paraboloid) was defined using Equation (5):

\[
\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + b_3 (X_1)^2 + b_4 X_1 X_2 + b_5 (X_2)^2 + \epsilon
\]  

(5)

where \( \hat{Y} \) is the predicted ln-transformed target (such as ln\([C_s]\)), \( X_1 \) is the ln-transformed primary predictor (such as ln\([C_s]\)) and \( X_2 \) is the ln-transformed secondary predictor. The \( b \) terms are regression coefficients. In the case of single-predictor model, the coefficients for variables containing \( X_2 \) would be zero. The reason for using a natural logarithm (ln-transformation) is explained below. The transformation error \( \epsilon \) is a zero-mean normal random variable with standard deviation \( \sigma_\epsilon \). For a single data point, this transformation error \( \epsilon_i \) was quantified as the residual (error) term of \( i \)th observation, i.e. the actual value \( Y_i \) defined from the oedometer test result minus the predicted value \( \hat{Y}_i \). The standard deviation of the transformation error \( \sigma_\epsilon \) can then be estimated from these individual residual errors using Equation (6) (e.g. Ang and Tang 2006):

\[
\sigma_\epsilon = \sqrt{\frac{1}{n-v} \sum_{i=1}^{n} [Y_i - \hat{Y}_i]^2}
\]

(6)

where \( n \) is the number of samples used in the linear regression and \( v \) is the number of degrees of freedom (that is, the number of estimated regression coefficients).

The transformation model presented above was selected in order to satisfy two assumptions usually made in OLS linear regressions:

(1) Homoscedasticity, which means that the standard deviation of the transformation error \( \sigma_\epsilon \) (Equation (6)) should be equal across the regression line. After applying the ln-transformations and fitting the polynomial regression model, the scatter became almost constant along the predictor’s value (e.g. Figure 12(a)).

(2) The normality of the residuals \( \epsilon_i \). After applying the abovementioned transformations and polynomial fit, the residuals became normally distributed (see Figure S4 in the online supplement).

Another assumption in OLS to be considered is the independence of the predictors: If the predictors are strongly correlated, a phenomenon known as multicollinearity can lead to unstable regression coefficient estimates. Predictors such as \( w_n \) and \( e_0 \) were thus not combined as predictors in this paper.

The standard deviation of error \( \sigma_\epsilon \) provides one measure of the transformation model uncertainty but it cannot be used to compare models for different target parameters. Hence, the definition by Ching and Phoon (2014b) was used to derive another measure of transformation uncertainty, i.e. the coefficient of variation \( \delta \) of the transformation error (Equation 4). Hence, the value of \( \delta \) was derived similarly as in the calibration analysis. Even though \( \delta \) was originally intended for calibration studies, its value is an indication of the transformation model quality (Phoon 2017). In addition,
the mean and median of bias values \( b_i \) were calculated for each model to confirm that the newly derived models are unbiased.

5.2. Removal of outliers

Potential outliers have to be detected and removed for two reasons: (1) they may distort the fitting of the transformation models, thus resulting in a biased model; and (2) the estimated transformation uncertainty should include a minimal number of other sources of uncertainty, such as measurement error or inherent variability (Phoon and Kulhawy 1999). Outliers related to measurement error can result from factors such as corrupted data (e.g. human error in data transfer) or specimen disturbance. On the other hand, the uncertainty related to inherent variability is caused by the very nature of multivariate databases; only a limited number of laboratory tests can be performed on a single soil specimen, and no two soil specimens are perfectly identical. Finally, the data point may come from another population entirely; for example, soils with very high organic content might behave like outliers due to their extreme physical properties.

First, under-consolidated and potentially disturbed specimens are detected using the physical constraints of OCR (over-consolidation ratio, \( \sigma'_p/\sigma'_v \)). Specifically, data points with an OCR smaller than 0.7 were removed as outliers (29 tests were conducted in total; see Figure 4). The threshold value of 0.7 was selected because the OCR for normally consolidated Finnish clays varies between 0.8 and 1.5 (Korhonen, Gardemeister, and Tammirinne 1974). Finally, the threshold value of 0.7 (instead of 0.8) was selected because the value of OCR is also affected by uncertainty in determining the preconsolidation pressure from incrementally loaded odometer tests.

Next, the extreme values of \( C_c \) were identified. A histogram of \( C_c \) (Figure 10) showed that there were five oedometer tests which produced very high \( C_c \). In particular, with a bin width of 0.5, all other bins \( (C_c < 5) \) contained at least two records. Based on this observation, tests with \( C_c \geq 5 \) (five tests in total) were considered rare cases and thus outliers. In fact, these tests were all CRS tests on highly sensitive Perniö clay (with the same sampling point and the same depth) and the \( C_c \) slope was interpreted as the tangent positioned right after the preconsolidation pressure, resulting in an almost vertical slope. To summarise, in total 34 data points (4% of all data) were considered outliers and removed prior to deriving the transformation models for \( C_c \) and \( C_r \). In the case of \( C_s \) a similar extreme-value criterion was applied and an additional three records \( (C_s > 0.4) \) were removed as outliers (Figure 10).

Finally, another outlier detection was applied during each transformation model fitting separately to account for factors such measurement error and inherent variability. This detection was based on the standard deviation method with a 3\( \sigma \)-threshold. In other words, those data points in which the residual error \( \varepsilon_i \) was more than three standard deviations \( (\sigma_e) \) apart from the zero-mean, were detected as outliers and hence removed before defining the final regression coefficients. The number of outliers removed in this last stage are presented in Tables 3 and 4 and in Figure 12(a), Figures 13–16(b). In addition, the figures illustrate the fitted model and adjusted R-squared \( (R^2_{adj}) \) also for the datasets including these outliers.

5.3. Leave-one-group-out analysis

Because some of the sites in the database contained significantly greater numbers of data points than others, a leave-one-group-out (LOGO) analysis was performed in order to verify that no single site had too great an influence on the transformation model. A simple single-variable linear model for \( \ln(C_c) \) with \( \ln(w_n) \) was selected for the LOGO study. In the LOGO analysis, one group (here one site) is left out of the training data each time; that is, the linear regression coefficients are estimated (fitted) using all the remaining data except that one group. The LOGO analysis recognised three sites (Söderkulla-Nikkilä, Haarajoki and Suurpelto).
that caused the most significant changes to the regression coefficients. $C_c(w_n)$-transformation models (polynomial regression, Equation (5)) were then derived for all data and for the datasets without each of these extreme cases. Figure 11 shows that the difference in the models is hardly noticeable. At higher values of $w_n$, the difference in the models was greater, but so was the scatter in $C_c$. This indicated that all the sites can be included in the fitting of the transformation models without the risk of creating a biased model due to a large sample from another population.

5.4. New transformation models and their transformation uncertainty

The most suitable predictors for the transformation models for $C_c$ and $CR$ were identified using a correlation matrix (see Table S2 in the online supplementary material). The three most suitable predictors were found to be $w_n$, $e_0$ and unit weight $γ$. The estimated regression coefficients and the model metric for these single-predictor models are presented in Table 3. The corresponding scatter plots for models with $w_n$ are presented in Figure 12(a,b) and Figure 13(a, b). The fitted models for $CR$ and $C_c$ are also illustrated in Figures 8 and 11, respectively. All the other scatter plots are presented in Figures S6–S7 in the online supplementary material. For both $C_c$ and $CR$, the models with $e_0$ had the smallest transformation uncertainty; for $C_c(e_0)$ model, the $R^2_{adj}$ was 0.83, the standard deviation of error $σ_e$ was 0.346, and the coefficient of variation $δ$ was 0.35. Using $w_n$ as the predictor resulted in very similar model metrics. On the other hand, the models with $w_L$ and $PI$ as single predictors resulted in clearly higher transformation errors, and hence these results have been omitted from Table 3. For example, for $C_c(w_L)$ model, the $R^2_{adj}$ was only 0.38.

Table 4 presents the regression coefficients and model metrics for two-predictor models for $C_c$ and $CR$. Void ratio $e_0$ was selected to be the primary predictor due to the lowest transformation uncertainty observed amongst single-predictor models. The four most suitable secondary predictors were found by studying the correlation between the potential secondary predictors $X_2$ and the errors $ε_i$ of the $C_c(e_0)$ model. A high correlation would effectively imply that this another predictor could increase the accuracy of the model if added as a secondary predictor (Ching and Phoon 2014b). The highest $ρ$ values were found with ratio ($s_i/σ_i$) ($ρ = −0.29$), preconsolidation pressure $ρ_p$ ($ρ = 0.20$), plastic limit $w_p$ ($ρ = −0.21$), and organic content $Or$ ($ρ = −0.19$). On the other hand, the corresponding correlations with $w_L$ and $PI$ were $ρ = −0.03$ and $ρ = 0.01$, respectively. Similar correlation results were acquired for the errors of the $CR(e_0)$ model. Models with three predictors were also derived, but the decrease in transformation uncertainty was considered to be too

![Figure 11. New $C_c(w_n)$ model and the models fitted to data without three most influential sites.](image)

Table 3. Single-predictor transformation models for compressibility of Finnish clays.

| Model variables | Regression model: $Y = b_0 + b_1X_1 + b_2X_1^2 + ε$ | Bias statistics | MD statistics |
|-----------------|---------------------------------|-----------------|---------------|
| $Y$ & $X_1$ & Outliers & $b_0$ & $b_1$ & $b_2$ & $σ_i$ & $n$ & $R^2_{adj}$ & MD & $δ$ |
| ln($C_c$) & ln($e_0$) & 6 & −1.5863 & 2.861 & −0.720 & 0.346 & 816 & 0.83 & 1.01 & 1.06 & 0.35 |
| ln($w_n$) & 9 & −21.567 & 8.227 & −0.743 & 0.357 & 813 & 0.82 & 1.00 & 1.07 & 0.37 |
| ln($γ$) & 7 & −87.951 & 71.341 & −14.307 & 0.407 & 758 & 0.76 & 1.03 & 1.08 & 0.42 |
| ln($CR$) & ln($e_0$) & 6 & −2.279 & 2.353 & −0.831 & 0.343 & 816 & 0.69 & 1.01 & 1.06 & 0.35 |
| ln($w_n$) & 8 & −20.748 & 7.986 & −0.791 & 0.350 & 814 & 0.67 & 1.00 & 1.06 & 0.36 |
| ln($γ$) & 6 & −100.74 & 77.334 & −14.941 & 0.374 & 759 & 0.63 & 1.02 & 1.07 & 0.38 |
| ln($C_c$) & ln($e_0$) & 9 & −3.253 & 1.480 & −0.317 & 0.435 & 619 & 0.49 & 1.10 & 1.09 & 0.38 |
| ln($w_n$) & 9 & −12.339 & 3.618 & −0.302 & 0.440 & 619 & 0.48 & 1.09 & 1.09 & 0.40 |
| ln($γ$) & 11 & −37.999 & 29.734 & −6.102 & 0.453 & 608 & 0.43 & 1.10 & 1.09 & 0.38 |

Notes: $C_c =$ compression index; $CR =$ compression ratio $C_c/(1 + e_0)$; $C_s =$ Swelling index; $n =$ number of samples; $R^2_{adj} =$ adjusted $R^2$-squared; $σ_i =$ standard deviation of the transformation error; $MD =$ median; $δ =$ COV of the bias values; $w_n =$ natural water content (%); $e_0 =$ initial void ratio; $γ =$ unit weight (kN/m$^3$).
small to outweigh the potential uncertainties brought by added complexity of the model.

Graphs comparing the predicted target values and the actual values are illustrated for one secondary predictor: $C_c(e_0, w_P)$-model is presented in Figure 14 and $CR(e_0, w_P)$ in Figure 15. Adding these secondary parameters resulted in a notable decrease in transformation uncertainty. For example, for $C_c(e_0, w_P)$ model, the adjusted $R^2_{adj}$ was 0.88, $\sigma_e$ was 0.317, and $\delta$ was 0.32. (For the same data, the metrics for the $C_c(e_0)$ model were $R^2_{adj} = 0.85$, $\sigma_e = 0.350$, and $\delta = 0.36$.) The smallest transformation uncertainty was reached by using the ratio $(s_u/\sigma_p)$ as the secondary predictor. On the other hand, using $w_L$ as the secondary predictor improved the model metrics only slightly ($R^2_{adj} = 0.85$, $\sigma_e = 0.350$, and $\delta = 0.36$), and hence this model was omitted from Table 4.

For swelling index $C_s$, the transformation uncertainties regarding the single-predictor models were greater compared to models for $C_c$ and $CR$ (Table 3). Overall, $R^2_{adj}$ was less than 0.50 for $C_s$ models. The scatter plots for $C_s(w_n)$ model are shown in Figure 16(a,b), and the
Table 4. Two-predictor transformation models for compressibility of Finnish clays.

| Model variables | Regression model: | Bias statistics |
|-----------------|------------------|----------------|
|                 | $Y = b_0 + b_1X_1 + b_2X_2 + b_3(X_1)^2 + b_4X_1X_2 + b_5(X_2)^2 + \varepsilon$ | $\delta$ |
| $Y$             | $X_1$            | $X_2$          | Outliers | $b_0$ | $b_1$ | $b_2$ | $b_3$ | $b_4$ | $b_5$ | $\sigma$ | $n$ | $R_{adj}^2$ | MD | Mean | $\delta$ |
| $\ln(C_c)$     | $\ln(e_0)$      | $\sigma$      | 5      | $-1.600$ | $1.885$ | $-0.150$ | $-0.319$ | $-0.545$ | $-0.045$ | $0.293$ | $392$ | $0.89$ | $1.01$ | $1.04$ | $0.30$ |
| $\ln(C_c)$     | $\ln(e_0)$      | $\sigma$      | 10     | $-2.089$ | $1.667$ | $0.201$ | $-0.376$ | $0.242$ | $-0.020$ | $0.302$ | $799$ | $0.87$ | $1.01$ | $1.05$ | $0.30$ |
| $\ln(C_c)$     | $\ln(w_l)$      | $\sigma$      | 5      | $-7.516$ | $1.911$ | $4.429$ | $-0.365$ | $0.248$ | $-0.807$ | $0.317$ | $357$ | $0.88$ | $1.02$ | $1.05$ | $0.32$ |
| $\ln(C_c)$     | $\ln(w_l)$      | $\sigma$      | 4      | $-1.559$ | $2.593$ | $-0.015$ | $-0.432$ | $-0.043$ | $-0.001$ | $0.335$ | $358$ | $0.84$ | $1.01$ | $1.06$ | $0.34$ |
| $\ln(C_c)$     | $\ln(\sigma_p)$ | $\sigma$      | 5      | $-2.294$ | $1.381$ | $-0.151$ | $-0.433$ | $-0.543$ | $-0.044$ | $0.293$ | $392$ | $0.76$ | $1.01$ | $1.04$ | $0.30$ |
| $\ln(C_c)$     | $\ln(\sigma_p)$ | $\sigma$      | 10     | $-2.785$ | $1.165$ | $0.201$ | $-0.488$ | $0.241$ | $-0.020$ | $0.302$ | $799$ | $0.76$ | $1.01$ | $1.05$ | $0.30$ |
| $\ln(C_c)$     | $\ln(w_p)$      | $\sigma$      | 5      | $-8.193$ | $1.395$ | $4.420$ | $-0.481$ | $0.252$ | $-0.806$ | $0.317$ | $357$ | $0.78$ | $1.02$ | $1.05$ | $0.32$ |
| $\ln(C_c)$     | $\ln(w_p)$      | $\sigma$      | 4      | $-2.253$ | $2.087$ | $-0.015$ | $-0.546$ | $-0.043$ | $-0.001$ | $0.334$ | $358$ | $0.72$ | $1.01$ | $1.06$ | $0.34$ |

Notes: $C_c$ = compression index; $CR$ = compression ratio $C_c/(1+e_0)$; $n$ = number of samples; $R_{adj}^2$ = adjusted R-squared; $\sigma$ = standard deviation of the transformation error; MD = median; $\delta$ = COV of the bias values; $e_0$ = initial void ratio; $\sigma'_p$ = preconsolidation pressure (kPa); $w_p$ = plastic limit (%); $\sigma_u$ = undrained shear strength (kPa) (fall cone); $Org$ = organic content (%).

The fitted model is also illustrated in Figure 9. The remaining scatter plots can be found from the online supplement (Figure S8). Regarding the potential secondary predictors for the $C_c(e_0)$ model, the highest $\rho$ value was found for clay content ($\rho = 0.14$). However, adding this secondary predictor did not notably increase the model metric, and hence two-predictor models for $C_c$ are omitted from Table 4.

Considering the results discussed above, the following equations can be recommended for estimating compressibility of Finnish fine-grained soils:

$$C_c = \exp \{-1.586 + 2.861 \ln(e_0) - 0.720 \ln(e_0)^2\}$$ \hspace{1cm} (7)

$$C_c = \exp \{-21.567 + 8.227 \ln(w_p) - 0.743 \ln(w_p)^2\}$$ \hspace{1cm} (8)

$$C_c = \exp \left\{-7.516 + 1.991 \ln(e_0) + 4.429 \ln(w_p) - 0.365 \ln(e_0)^2 + 0.248 \ln(e_0) \ln(w_p) - 0.807 \ln(w_p)^2\right\}$$ \hspace{1cm} (9)

$$CR = \exp \{-20.748 + 7.986 \ln(w_p) - 0.791 \ln(w_p)^2\}$$ \hspace{1cm} (10)

$$C_s = \exp \{-12.339 + 3.618 \ln(w_p) - 0.302 \ln(w_p)^2\}$$ \hspace{1cm} (11)

6. Discussion

Contrary to previous findings on multiple predictor models for clay compressibility (e.g. Kootahi and Moradi 2017), adding $w_l$ or $PI$ as the secondary predictor did not notably decrease the transformation uncertainty in estimating $C_c$ of Finnish fine-grained soils. As suspected by Di Buò et al. (2019), the physical properties (such as $w_p$) seem to better explain the variability of $C_c$ for Finnish clays. On the other hand, adding the plastic limit $w_p$ as the secondary predictor did result in a notable decrease in transformation uncertainty for $C_c$. 
As such, for Finnish clays, w_P may be a better representation of the non-structural effects (such as mineralogy and particle-surface characteristics) to compressibility. Kootahi and Moradi (2017) also used w_P as the secondary predictor of C_c for marine fine-grained soils, but the increase in model metrics was greater when w_L or PI was used instead.

The smallest transformation uncertainty was reached by using the ratio (σ_p / σ_p') or σ_p' as the secondary predictor. In a preliminary design stage, σ_p' is usually not available, and hence these transformation models are intended to be used in Bayesian approaches. For instance, one site-specific oedometer test is enough to compute the prior distribution of C_c for Bayesian updating via the reported standard deviation of error σ_e. Alternatively, the engineer can utilise the limited site-specific data to verify whether the predicted C_c is within the 95% confidence limits as instructed by Ching (2017).

In general, the transformation uncertainty was found to be moderate at δ = 0.30–0.43. On the other hand, transformation models for the undrained shear strength of clays exhibit somewhat smaller transformation uncertainty (e.g. δ = 0.25–0.30 for the most reliable models derived by D'Ignazio et al. [2016]). Higher transformation uncertainty for compressibility is to be expected, since the interpretation of C_c from the oedometer curve of soft clays is not explicit. In practice, the 95% confidence interval for a settlement prediction can be approximated by considering the ±2σ limits of the predicted C_c' for a five metres thick clay layer with void ratio of e_0 = 1, a two-metre high embankment (corresponding to 40 kPa) on normally consolidated clay would result in a settlement of 77–308 mm (assuming effective in-situ stress of 40 kPa). The prediction’s reliability decreases with increasing values of e_0 and w_n. On the other hand, a ground-supported foundation is not usually considered a feasible option for highly compressive soft clays, hence reducing the need for accurate settlement predictions.

The transformation uncertainty related to swelling index C_s was notably greater compared to C_c and CR. The greater scatter about the trend could be explained by the stress-dependency of C_s; since the unloading-reloading loop of Finnish clays is often marked with hysteresis, the value of C_s depends on the stress path applied in the oedometer test (e.g. Vipulanandan et al. 2008).

The statistics of the new models’ bias values (Tables 3 and 4) show that the mean bias is systematically slightly larger than 1 (1.04–1.09). Median, on the other hand, is close to 1. This pattern could be explained by the fact that the bias (i.e. the actual target value divided by the predicted value) is log-normally distributed. In fact, the transformation error is often assumed to be log-normal (e.g. Ching 2017). Indeed, the biases for C_c(w_n) model seemed to be log-normally distributed based on the histogram’s shape (see Figure S4 in the online supplementary material). Other possible explanations for the average bias other than 1 are inaccuracies related to the transformation model or factors such as outliers which remained undetected.

Lastly, it can be concluded that the selected 3σ threshold is a rather strict outlier criteria for geotechnical data, because inherent variability and measurement error also contribute to the observed transformation uncertainty.
uncertainty. However, since the standard deviation method is not the most robust outlier detection method (because the statistics which define the outlier criteria are calculated with the very same data which contains the potential outliers), a stricter threshold was selected for this study. Nevertheless, it can be concluded that the apparent transformation uncertainty would be smaller if one would use a less strict (e.g. 2σ threshold) outlier criteria instead.

Finally, it should be noted that the recommended new transformation models (Equations [7]–[11]) should not be used for soil specimens with very high water content (\( w_n > 150\% \)). Figure 8 illustrates how the trend curve for the CR\( (w_n) \) model plateaus at highest \( w_n \) values. Similarly, Figure 13(b) depicts how the predicted CR has a maximum value. Hence, at very high \( w_n \) values, the transformation model becomes biased. However, as previously discussed, the accuracy of the model is more essential for clays with moderate to low compressibility (i.e. clays with lower \( w_n \)).

### 7. Conclusions

The following conclusions can be drawn:

- An extensive, partly multivariate clay database FI-CLAY/14/856 was compiled and compared to existing transformation models for compressibility of clays.
- New transformation models for compressibility of Finnish clays were derived via 2-degree polynomial regression and their transformation uncertainty was evaluated.
- The best predictors of compressibility were void ratio and water content. For compression index, the adjusted \( R^2 \) was 0.82–0.83, whereas models for compression ratio and swelling index had lower \( R^2 \) values (0.67–0.69 and 0.48–0.49, respectively).
- When the void ratio was combined with a secondary predictor, such as the ratio between undrained shear strength and preconsolidation pressure or plastic limit, the transformation uncertainty decreased notably for compression index and compression ratio. For compression index, the adjusted \( R^2 \) increased to 0.88–0.89. For swelling index, adding a secondary predictor did not notably decrease the transformation uncertainty.

### Acknowledgements

The authors would like to acknowledge Finnish Transport Infrastructure Agency for funding this research and Rita Rekonen and Matti Lojander for compiling the RITA database. Gratitude is also directed to the Reviewers for providing valuable comments on the manuscript.

### Disclosure statement

No potential conflict of interest was reported by the author(s).

### Funding

This work was supported by Finnish Transport Infrastructure Agency [research contract number 0133U-0065705].

### ORCID

Monica Susanne Löfman © http://orcid.org/0000-0003-3086-1031
Leena Katariina Korkiala-Tanttu © http://orcid.org/0000-0003-0894-7218

### References

Al-Khafaji, A. W. N. 1979. “Decomposition Effects on Engineering Properties of Fibrous Organic Soils.” (PhD diss.). Michigan State University.

Al-Khafaji, A. W. N., and O. B. Andersland. 1992. “Equations for Compression Index Approximation.” ASCE Journal of Geotechnical Engineering 118 (1): 148–153.

Andersen, Jan Dannemand. 2012. “Prediction of Compression Ratio for Clays and Organic Soils.” In Proceedings of the 16th Nordic Geotechnical Meeting (NGM 2012), Copenhagen, 9–12 May 2012, Danish Geotechnical Society.

Ang, Alfredo H.-S., and Wilson H. Tang. 2006. Probability Concepts in Engineering: Emphasis on Applications to Civil and Environmental Engineering. New York, USA: Wiley.

Azzouz, Amr S., Raymond J. Krizek, and Ross B. Corotis. 1976. “Regression Analysis of Soil Compressibility.” Soils and Foundations 16 (2): 19–29.

Ching, Jianye. 2017. “Transformation Models and Multivariate Soils Databases.” Chap. 1 in Joint TC205/TC304 Working Group on “Discussion of statistical/reliability methods for Eurocodes” – Final Report (September 2017). Accessed March 9, 2020. http://140.112.12.21/issmg/TC205_304_reports/All%20Chapters%20(TC205_TC304).pdf.

Ching, Jianye, and Kok-Kwang Phoon. 2014a. “Correlations Among Some Clay Parameters – The Multivariate Distribution.” Canadian Geotechnical Journal 51 (6): 686–704. doi:10.1139/cgj-2013-0353.

Ching, Jianye, and Kok-Kwang Phoon. 2014b. “Transformations and Correlations Among Some Clay Parameters – The Global Database.” Canadian Geotechnical Journal 51 (6): 663–685. doi:10.1139/cgj-2013-0262.

Coffolino, V. M. 1961. “Statistical Forecasting of Compression Index.” In Proceedings of the Fifth International Conference on Soil Mechanics and Foundation Engineering. Paris, France, 51–53.

Di Buò, Bruno, Tim Länssivaa, Andrea Franzia, and Paul Mayne. 2019. “Compressibility of Finnish Sensitive Clay.”
Appendix

Tables A1 and A2 describe the database’s Subset A and B, respectively.

### Table A1. Subset A: site descriptions and references.

| Reference | Site name                  | City        | n  | Sampling year(s) | Dominant soil type(s)   |
|-----------|----------------------------|-------------|----|------------------|-------------------------|
| Gardemeister (1973) | Lahti          | Lahti       | 12 | 1968             | Clay and Silt           |
| Riihimäki |               | Riihimäki   | 6  | 1967             |                         |
| Rytylä    |               | Hausjärvi   | 10 | 1970             |                         |
| Keravanjoen laaksoalue | Vantaa | Keravanjoen laaksoalue | 36 | 1976             | Clay and Clayey silt    |
| Korso     |               |             | 6  | 1976             |                         |
| Kylmäoja  |               |             | 4  | 1976             |                         |
| Länsi-Vantaa |             |             | 43 | 1976             |                         |
| Pohjois-Vantaa |             |             | 14 | 1976             |                         |
| Vantaanportti |             |             | 7  | 1976             |                         |
| City of Vantaa (Pajunen 1976) | Riihimäki | Riihimäki   | 6  | 1967             |                         |
| Maaria    |               |             | 3  | 1990s            | Fat clay                |
| Turku     |               |             | 3  | 1990s            | Fat clay                |
| Maarinranta, Otaniem | Espoo    |             | 38 | 1998             | Clay                     |
| Muro test embankmentb | Seinäjoki |             | 30 | 1993             | Organic silt            |
| Sompio    |               |             | 11 | 1992             | Clay and Clayey silt    |
| Söderkulla-Nikkilä | Sipo       |             | 116| 1997–1999        | Clay, Organic clay and Gyttja |
| Vaasa test embankment | Vaasa     |             | 9  | 1989             | Organic silt            |
| RITA database (Rekonen and Lojander 1999, unpublished internal report) | Haapa-Kimolaa | Iitti | 79 | 1980             | Fat clay and Lean clay  |
| Haarajoki test embankment | Järvenpää |             | 66 | 1995             | Fat clay                |
| Karviainen | Aura         |             | 3  | 1990s            | Fat clay                |
| Kaukalhti | Espoo        |             | 4  | 1990s            | Fat clay                |
| Kellokoski | Tuusula      |             | 16 | 1996             | Fat clay                |
| Kyrö      | Pöytäjä      |             | 3  | 1990s            | Fat clay                |
| Loimaa    | Loimaa       |             | 3  | 1990s            | Fat clay                |
| Maarja    | Turku        |             | 3  | 1990s            | Fat clay                |
| Maarinranta, Otaniem | Espoo |             | 38 | 1998             | Clay                     |
| Murro test embankmentb | Seinäjoki |             | 30 | 1993             | Organic silt            |
| Sompio    | Kerava       |             | 11 | 1992             | Clay and Clayey silt    |
| Söderkulla-Nikkilä | Sipo       |             | 116| 1997–1999        | Clay, Organic clay and Gyttja |
| Vaasa test embankment | Vaasa     |             | 9  | 1989             | Organic silt            |
| Partly reported by Lojander (1980) and Loukola (1996). |

*Notes: n = number oedometer test samples.*

### Table A2. Subset B (collected for this study): site descriptions and references.

| Reference | Site name                  | City   | n  | Sampling year(s) | Dominant soil type(s)   |
|-----------|----------------------------|--------|----|------------------|-------------------------|
| Kimola canal |                              | Kuopio | 20 | 2017             | Lean clay               |
| Partly reported by Löfman and Korkiala-Tanttut (2019) | Kujal test embankment 1 (4 m) | Lahti | 6  | 2017             | Lean clay               |
| Partly reported by Löfman and Korkiala-Tanttut (2019) | Kujal test embankment 2 (5.5 m) | Lahti | 13 | 2017             | Lean clay               |
| Ossinlampi (Otaniem) | Espoo | 5°+12 | 2009*, 2012 | Fat clay       |
| Partly reported by Koskinen (2000) | Ossinlampi (Otaniem) | Espoo | 40 | 1986             | Clay                    |
| Pernio test embankment | Salo | 94  | 2009–2010 | Clay       |
| POKO (Porvoo-Koskenkilä) | Porvoo | 36 | 1999 | Clay, Organic clay and Gyttja |
| Suurpelto | Espoo        | 32  | 2005, 2008 | Clay and Gyttja |
| Östersundom test embankment | Espoo | 42 | 2013 | Clay       |

*Partly reported by Koskinen (2014).*

*Notes: n = number oedometer test samples.*

---

Suhonen, Kati. 2010. “Creep of Soft Clay.” (Master thesis). Aalto University School of Science and Technology.

Tanaka, H. 2002. “A Comparative Study on Geotechnical Characteristics of Marine Soil Deposits Worldwide.” *International Journal of Offshore and Polar Engineering* 12 (2): 81–88.

Terzaghi, Karl, and Ralph B. Peck. 1967. *Soil Mechanics in Engineering Practice.* New York: John Wiley & Sons.

Tiwari, Binod, and Beena Ajmera. 2011. “Consolidation and Swelling Behavior of Major Clay Minerals and Their Mixtures.” *Applied Clay Science* 54 (2011): 264–273.

Vipulanandan, C., Y. J. Ahossin Guezo, Ö Bilgin, S. Yin, and M. Khan. 2008. “Recompression Index (Cr) for Overconsolidated Soft Clay Soils.” Geocongress 2008: Characterization, Monitoring, and Modeling of GeoSystems, ASCE: 68–75.

Wesley, Laurence D. 1988. “Compression Index: Misleading Parameter?” *Journal of Geotechnical Engineering* 114 (6): 718–723.