

Identifiability: construction error as apparent soil stiffness in back-analysis

Raul Fuentes¹, Eden Almog-Goldreich¹

¹ Institute of Geomechanics and Underground Technology, RWTH Aachen, Aachen

Abstract

Geotechnical back-analysis usually estimates ground parameters while treating the construction process as known. Because ground and construction jointly generate the stress path, an error in the assumed construction state can be absorbed by the calibrated soil parameters without degrading the fit. This letter writes the inverse problem as $y = \mathcal{M}(\mathcal{F}(\theta, c) + \delta) + \varepsilon$, separating ground parameters θ from the mechanically effective construction state c , and distinguishes confounding, when c is estimated jointly with θ , from information loss, when an uncertain recorded value is fixed. Both are quantified with a verified semi-analytical shield-tunnelling model applied to the paired TULIP sections. With section specific grout pressure records, one site investigation stiffness reproduces both measured settlement peaks; replacing the records by a common nominal pressure creates an apparent stiffness contrast of about 40% with equally good fits. Because operational grout records do not directly measure the effective boundary pressure, settlement data cannot decide between these interpretations, and additional settlement points do not help. Lining deformation, which responds oppositely to grout pressure, would restore most of the lost information. Goodness of fit alone is therefore not evidence that calibrated soil parameters are correct.

Keywords: back-analysis; identifiability; tunnels & tunnelling; numerical modelling; monitoring; construction process

Introduction

Monitoring data are routinely used to calibrate constitutive parameters and update predictions (Finno and Calvello, 2005). Non-uniqueness and parameter correlation in such analyses are well established (Honjo *et al.*, 1994; Ledesma *et al.*, 1996a, 1996b; Calvello and Finno, 2004; Levasseur *et al.*, 2009) and, in current terminology, are problems of practical identifiability (Lam *et al.*, 2022). Almost invariably, however, the construction history entering the forward model is treated as perfectly known. Yet the stress path is generated jointly by the ground and by construction, and the planned, recorded, modelled and mechanically effective construction states need not coincide.

Benchmark evidence shows that discrepancy between the model and the data cannot be attributed to soil parameters alone. In the COST C7 tunnel benchmark, prescribed geometry and parameters still produced about 20% scatter in maximum settlement because of modelling choices (Schweiger, 2002; Potts, 2003), and in the TULIP class A prediction exercise substantial scatter remained even when a common Hardening Soil parameter set was imposed (Bourgeois *et al.*, 2023).

Identifiability can be improved by adding non-redundant observations or by reducing, through better records, the uncertainty in competing construction variables. Optimal experimental design has addressed the first route (Hölter *et al.*, 2018); the second, and the interaction between the two, have rarely been quantified.

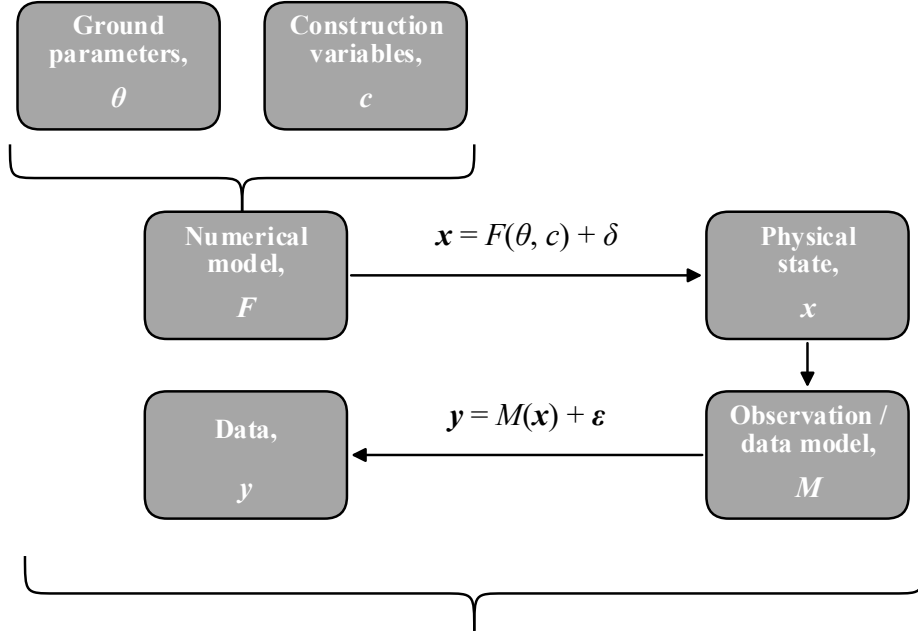
The statistical methods used here are standard (Cox and Reid, 1987; Kennedy and O’Hagan, 2001) but used to show something previously not thoroughly presented. This letter shows, with a verified forward model and a field anchored case, how large construction parameter compensation is in practice and which observations break it. The paired TULIP sections, previously analysed with the Advanced Displacement Confinement Method (ADCM) by Almog-Goldreich and Fuentes (2026), are used because full access to the model inputs and code is needed to compute internally consistent sensitivities.

Construction aware inverse problem

Following the calibration framework of Kennedy and O’Hagan (2001), the back-analysis problem is written as (Fig. 1)

$$y = \mathcal{M}(\mathcal{F}(\theta, c) + \delta) + \varepsilon \quad (1)$$

where θ contains ground/model parameters, c is the mechanically effective construction state (e.g., face and grout pressures, shield gap, temporary geometry), $\mathcal{F}$ is the forward model, δ is model discrepancy, $\mathcal{M}$ is the observation operator (sensor location, baseline, data processing) and ε is observation error with covariance $\mathbf{R}$. The key step is separating c from θ : c is the construction state that actually acts on the ground, not automatically the value written in a log.



$$\mathbf{y} = M(F(\theta, c) + \delta) + \varepsilon$$

Figure 1. Construction aware formulation of back-analysis. Ground/model parameters θ and the mechanically effective construction state c enter the forward model $\mathcal{F}$; δ is model discrepancy. The observation operator $\mathcal{M}$ maps the physical state $\mathbf{x}$ to the recorded data $\mathbf{y}$, with observation error ε .

Linearising about a reference state, with δ conditioned on the adopted forward model,

$$\Delta \mathbf{y} = \mathbf{J}_\theta \Delta \theta + \mathbf{J}_c \Delta c \quad (2)$$

Each column of $\mathbf{J}_\theta$ and $\mathbf{J}_c$ is a sensitivity fingerprint in observation space. If $\mathbf{J}_\theta \Delta \theta + \mathbf{J}_c \Delta c \approx 0$ for non-zero perturbations, a construction error can be absorbed by θ without changing the fit.

Two distinct situations follow.

Joint inference. If c is estimated together with θ , it acts as a nuisance quantity. For scalar θ and c , the profile information remaining on θ is (Cox and Reid, 1987)

$$I_\theta^{prof} = I_{\theta\theta}(1 - \rho^2), \quad \rho = \frac{j_\theta^T R^{-1} j_c}{\sqrt{(j_\theta^T R^{-1} j_\theta)(j_c^T R^{-1} j_c)}} \quad (3)$$

where ρ measures the similarity of the two fingerprints after weighting by the observation covariance. As $|\rho| \rightarrow 1$, $I_\theta^{prof} \rightarrow 0$.

Uncertain record. If c is instead fixed at a recorded value c_0 , but $c = c_0 + \Delta c$ with variance σ_c^2 , the construction uncertainty enters observation space through $\mathbf{J}_c$. For one scalar observation with residual scale σ_y , the information on a scalar θ is

$$I_\theta^{c0} = \frac{J_\theta^2}{\sigma_y^2 + J_c^2 \sigma_c^2} \quad (4)$$

Equation (3) reflects collinearity of fingerprints and Eq. (4) broadening of the model–data comparison; both reduce information on θ , by different mechanisms with different remedies.

Application to the TULIP paired sections

Forward model and verification

ADCM links TBM operation to ground confinement, ground convergence, lining response and volume loss, which is converted to surface settlement via a separate method such as O'Reilly and New (1983) trough (Almog-Goldreich and Fuentes, 2026). In the notation of Eq. (1), θ includes constitutive parameters E_{50}^{ref} , $E_{\text{ur}}^{\text{ref}}$, ϕ' , c' , m and K_{TBM} ; c includes P_g , face pressure and shield gap Δ ; and the trough parameter of the settlement curve K_i belongs to $\mathcal{M}$. The original research implementation was re-executed through a non-interactive wrapper without altering the governing equations, but using 720 angular nodes, with no recalibration. It reproduces the published confinement ratios, shield pressures, volume losses and plotted peak settlements for TULIP Cases 1a and 1b presented in Almog-Goldreich and Fuentes (2026), to within about 3% (Table 1). Because E_{50}^{ref} controls shear hardening whereas $E_{\text{ur}}^{\text{ref}}$ controls elastic unloading–reloading and the ground stiffness entering the lining model, stiffness is varied throughout as a single scale k with $E_{\text{ur}}^{\text{ref}} = 3E_{50}^{\text{ref}}$, the TULIP ratio.

The two sections presented in TULIP share geometry, ground parameters and operating conditions except the recorded grout pressure (Table 1). $E_{50}^{\text{ref}} = 20$ MPa and $E_{\text{ur}}^{\text{ref}} = 60$ MPa were selected from the site investigation stiffness data summarised by Michalski *et al.* (2024), not calibrated to the SMRN or SMRS records. With them and the recorded pressures, the model reproduces the measured peak settlements of 8.6 and 12.1 mm (Michalski *et al.*, 2024) with a two point root-mean-square error of 0.20 mm.

Table 1. TULIP paired sections: inputs, verification of the re-executed ADCM implementation (interpretation A), and the alternative interpretation B that fits the measured peak settlements equally well. Published values from Almog-Goldreich and Fuentes (2026); published peak settlements are read from the calculated troughs in its Figs 9–10 (± 0.1 mm); L_s is measured from the cutting face to the shield tail, i.e. the 11 m shield body plus the cutting wheel, about 1 m (Abdallah *et al.*, 2026, Fig. 2); u_{max} is the predicted maximum lining deformation.

Quantity	SMRN (Case 1a)	SMRS (Case 1b)
Common inputs		
Face pressure (bar) / shield gap Δ (mm) / shield length L_s (m), cutting face to tail	1.8 / 35 / 12	1.8 / 35 / 12
ϕ' (°) / c' (kPa) / K_{TBM}	33 / 10 / 0.45	33 / 10 / 0.45
Measured		
S_{max} (mm)	8.6	12.1

A: section specific recorded pressures, common stiffness $E_{50}^{\text{ref}} = 20$ MPa, $E_{\text{ur}}^{\text{ref}} = 60$ MPa

Quantity	SMRN (Case 1a)	SMRS (Case 1b)
P_g (bar)	2.40	2.20
λ_{TBM} crown, published / re-executed	0.64 / 0.640	0.73 / 0.729
P_{SHIELD} springline (bar), published / re-executed	2.3 / 2.32	2.0 / 2.03
Volume loss (%), published / re-executed	0.26 / 0.258	0.37 / 0.377
S_{max} (mm), published (plotted) / re-executed	≈ 8.2 / 8.34	≈ 11.9 / 12.22
u_{max} (mm), re-executed	3.52	3.02
B: common nominal effective pressure, stiffness fitted to each section ($E_{\text{ur}}^{\text{ref}} = 3E_{50}^{\text{ref}}$)		
P_g (bar)	2.30	2.30
E_{50}^{ref} (MPa)	24.1	17.1
S_{max} (mm)	8.60	12.10
u_{max} (mm)	2.80	3.67

Uncertain grout pressure record

Central differences about the midpoint $P_g = 2.30$ bar give $\partial S_{\text{max}} / \partial P_g = -19.82$ mm/bar, and the response over 2.20–2.40 bar is close to linear ($R^2 = 0.9996$; Fig. 2). Grout pressure is logged at the injection lines, where it fluctuates through each injection cycle and varies around the ring (Bezuijen *et al.*, 2004; Dias and Bezuijen, 2017). The pressure distribution transmitted to the ground was not measured in TULIP (Abdallah *et al.*, 2026). A logged value is therefore information about the effective boundary pressure, not a direct measurement of it. Taking $\sigma_y = 1$ mm (the ADCM validation error, a model–data comparison scale rather than instrument noise), Eq. (4) gives Table 2: $\sigma_{P_g} = 0.05$ bar halves the information on any scalar θ carried by peak settlement, and 0.2 bar leaves about 6%.

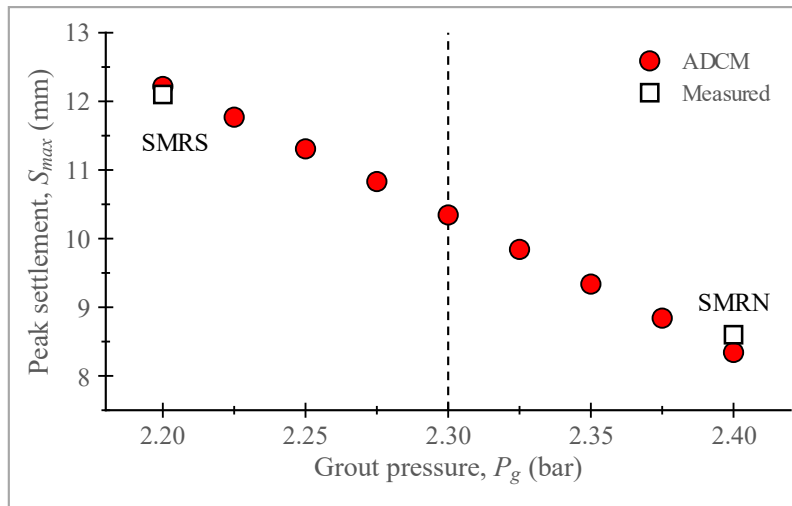


Figure 2. ADCM grout pressure sweep for the TULIP reference state ($E_{50}^{\text{ref}} = 20$ MPa, $E_{\text{ur}}^{\text{ref}} = 60$ MPa), with the measured peak settlements of the two sections. The dashed line marks the 2.30 bar midpoint at which the local derivative is evaluated.

Table 2. Information remaining, I_{θ}^{c0} relative to the construction perfect case ($\sigma_c = 0$) in Eq. (4), for uncertainties in the mechanically effective grout pressure (not sensor error) and three comparison scales σ_y .

σ_{Pg} (bar)	Induced σ_s (mm)	$\sigma_y = 0.5$ mm	$\sigma_y = 1$ mm	$\sigma_y = 2$ mm
0.02	0.40	61%	86%	96%
0.05	0.99	20%	50%	80%
0.10	1.98	6%	20%	50%
0.20	3.96	2%	6%	20%

Joint inference: construction error becomes apparent stiffness

With one scalar observation, two parameters are trivially unidentifiable; the practical question is how strongly they compensate. At the midpoint, $\partial S_{\max}/\partial E_{50}^{\text{ref}} = -0.518$ mm/MPa (with $E_{\text{ur}}^{\text{ref}} = 3E_{50}^{\text{ref}}$), so a constant settlement direction has $dE_{50}^{\text{ref}}/dP_g \approx -38$ MPa/bar. This means a 0.05 bar pressure error is absorbed by a 10% change in stiffness and a 0.10 bar error by 19%. The full nonlinear model confirms this compensation ridge (Fig. 3): holding settlement constant requires $E_{50}^{\text{ref}} = 23.6$ MPa at $P_g = 2.20$ bar and 16.13 MPa at 2.40 bar.

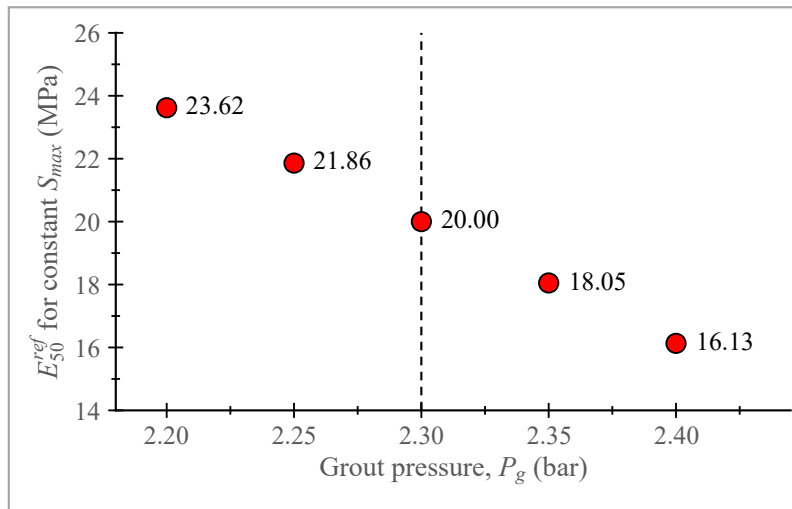


Figure 3. ADCM compensation ridge: combinations of grout pressure (c) and stiffness (θ ; $E_{\text{ur}}^{\text{ref}} = 3E_{50}^{\text{ref}}$) that give the same peak settlement as the reference state at 2.30 bar (dashed line).

Interpretation B in Table 1 shows the consequence. If the section specific records are replaced by a common nominal effective pressure $P_g = 2.30$ bar, separate back-analyses reproduce the measured settlements to within 0.01 mm, a closer fit than interpretation A (0.20 mm), but require $E_{50}^{\text{ref}} = 24.1$ and 17.1 MPa: an apparent stiffness contrast of 41% (relative to SMRS) between geotechnically similar sections.

Settlement data cannot decide between common stiffness with different effective pressures (A) and different stiffness with a common effective pressure (B), and the operational records cannot either: even for $\sigma_{Pg} = 0.10$ bar the recorded difference is only 1.4 standard deviations, and

realistic uncertainties are larger. The TULIP benchmark accordingly specified a single site wide grout pressure (Bourgeois *et al.*, 2023), which is the assumption behind B. A is favoured mainly by the expectation of geological similarity, corroborated by the measured settlement difference (3.5 mm) being close to the 3.9 mm predicted from the records. A linearised joint inference of section specific E_{50}^{ref} and P_g , using the records as prior means ($\sigma_{P_g} = 0.05\text{--}0.5$ bar) and a section-to-section stiffness coefficient of variation of 0.1–1.0, gives stiffness contrasts below 5%; the 41% arises only when a common pressure is treated as exact. The ambiguity is therefore real, and resolving it requires an independent observable.

Which observations separate θ and c ?

An additional observation increases identifiability only if it adds a non-proportional row to $\mathbf{J}$. A second ordinate of the same ADCM settlement trough does not: with K_i fixed, $S(x) = g(x)S_{\text{max}}$, so every ordinate's fingerprint is proportional to that of S_{max} and the $[E_{50}^{\text{ref}}, P_g]$ Jacobian remains rank one. Estimating K_i would make further ordinates informative about trough width, but not about this pair, which enters only through the volume loss.

Maximum lining deformation u_{max} behaves differently. ADCM computes it with the Duddeck and Erdmann (1985) solution using the unload–reload modulus at the excavation boundary as ground stiffness, so it is insensitive to E_{50}^{ref} alone. With stiffness scaled as a whole, the normalised fingerprints for $\mathbf{y} = [S_{\text{max}}, u_{\text{max}}]^T$ are $\mathbf{j}_k = [-1.00, -0.79]^T$ and $\mathbf{j}_{P_g} = [-4.41, 1.75]^T$: stiffer ground reduces both quantities, whereas higher grout pressure reduces settlement but increases lining deformation. With equal relative weighting, $|\rho|$ falls from 1 to 0.50 and $1 - \rho^2 = 0.75$, so u_{max} restores three quarters of the information on stiffness (Fig. 4); weighting u_{max} half or twice as heavily gives 0.29 or 0.99. Table 1 shows the practical consequence: the two interpretations predict opposite orderings of lining deformation, with SMRN 0.49 mm larger than SMRS under interpretation A and 0.86 mm smaller under B. Lining deformation was not measured at TULIP, so this is an experimental design result: a lining convergence measurement at the two sections would have discriminated directly between the interpretations. This work shows clearly the potential contradiction.

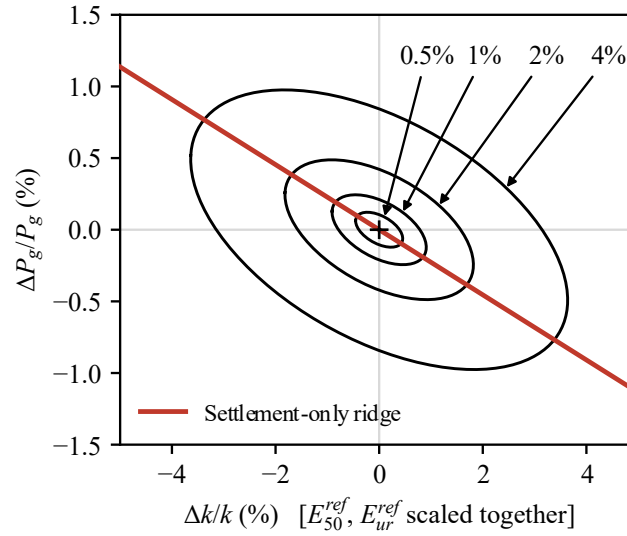


Figure 4. Local information geometry for the stiffness scale k (θ ; E_{50}^{ref} and $E_{\text{ur}}^{\text{ref}}$ scaled together) and P_g (c). The red line is the settlement only compensation ridge; closed contours (values indicated by arrows) show combined normalised changes of 0.5–4% in peak settlement and maximum lining deformation (equal relative weighting).

Ranking influence by illustrative uncertainty

Normalised sensitivities from $\pm 1\%$ perturbations compare inputs at equal fractional change, which are not equally plausible. Table 3 therefore weights each sensitivity by an illustrative engineering uncertainty, not a measured statistic, giving the settlement standard deviation $|\partial S_{\text{max}}/\partial q| \sigma_q$ contributed by each input. The grout pressure contribution (1.0–4.0 mm) is of the same order as, or larger than, that of stiffness (2.1–3.1 mm) and exceeds that of every other ground parameter. Similarly, for ADCM Lyon Case 4 (Almog-Goldreich and Fuentes, 2026), $\partial S_{\text{max}}/\partial \Delta = 0.54$ mm/mm at the published 15 mm shield gap, so a 2–3 mm gap uncertainty contributes 1.1–1.6 mm. The ranking, not the precise numbers, is the point.

Table 3. Uncertainty weighted contributions to TULIP peak settlement uncertainty at $P_g = 2.30$ bar (reference $S_{\text{max}} = 10.34$ mm, $u_{\text{max}} = 3.25$ mm). Normalised sensitivities are $(q/y)\partial y/\partial q$ from $\pm 1\%$ ADCM perturbations; uncertainty ranges are illustrative engineering values.

Input q	Role	Normalised sensitivity, S_{max}	Normalised sensitivity, u_{max}	Indicative uncertainty	Contribution to σ_s (mm)
P_g	c	−4.41	1.75	$\sigma = 0.05\text{--}0.20$ bar	1.0–4.0
P_{face}	c	1.53	−0.61	$\sigma = 0.05\text{--}0.10$ bar	0.4–0.9
K_i	M	−1.00	0	CoV = 0.10	1.0
Stiffness k (E_{50}^{ref} , $E_{\text{ur}}^{\text{ref}}$)	θ	−1.00	−0.79	CoV = 0.20–0.30	2.1–3.1
K_{TBM}	θ	0.71	−0.66	CoV = 0.10	0.7
m	θ	−0.46	−0.21	CoV = 0.15	0.7
φ'	θ	−0.48	−0.03	$\sigma = 2^\circ$	0.3

Discussion

The decomposition in Eq. (1) is useful because each non- θ term has a different remedy. Uncertainty in c calls for better records of the mechanically effective construction state, such as direct pressure measurement at the extrados, a defensible σ_c , or joint inference where the data support it. Model discrepancy δ calls for improved physics rather than recalibration of θ ; $\mathcal{M}$ must reproduce sensor position, baseline and processing; and ε belongs in R .

For the observational method (Peck, 1969), predicted ranges are meaningful only if model and observations refer to the same construction and observation states (Hashash and Finno, 2008; Spross and Johansson, 2017). Construction data have an identifiable information value: reducing σ_c increases what can be learned about θ without adding a sensor, while a well-chosen observable can separate θ and c without changing the construction process. Case histories should therefore record lining convergence alongside surface settlement.

Construction uncertainty and model form error δ can produce similar effects in observation space and compete for the same apparent stiffness adjustment. The TULIP results are conditional on ADCM: they demonstrate the identifiability mechanism and its engineering scale, not that grout pressure uniquely caused the difference between the sections. The analysis is local and restricted to one parameter pair.

Conclusions

(1) Construction process fidelity is an integral part of the inverse problem: joint estimation of c removes information on θ according to fingerprint similarity, and fixing an uncertain record c_0 broadens the model–data comparison.

(2) For TULIP, peak settlement trades grout pressure against E_{50}^{ref} (with $E_{\text{ur}}^{\text{ref}} = 3E_{50}^{\text{ref}}$) at about 38 MPa/bar. Replacing the section specific records by a common nominal pressure turns a common stiffness into an apparent 41% contrast with an equally good, indeed closer, fit; goodness of fit is therefore not evidence that calibrated parameters are correct.

(3) Neither settlement data nor operational grout records can uniquely discriminate between the interpretations; geological similarity favours a common stiffness, but confirming it requires an independent observable.

(4) Measurement count is not information dimension: further settlement ordinates may identify trough width but cannot separate stiffness from grout pressure, whereas lining deformation restores most of the information on stiffness (three quarters at equal weighting).

(5) In the TULIP and Lyon cases, and for the illustrative uncertainties adopted, construction variables contribute settlement uncertainty of the same order as, or larger than, soil stiffness.

(6) The findings of this Letter were proven against a specific case study, but it shows that increasing construction process complexity will exacerbate the identifiability problem, if more accurate records are not kept.

Data availability

Case inputs are given in Almog-Goldreich and Fuentes (2026). The derived ADCM results and the data plotted in all figures are provided as supplementary material. The core ADCM implementation cannot be released publicly because of employer intellectual property requirements; the method is fully described in Almog-Goldreich and Fuentes (2026).

References

- Abdallah, A., Branque, D., Berthoz, N. & Rallu, A. (2026). Steady-state and transient pressure conditions around the EPB-TBM: numerical modelling based on experimental data. *Tunnelling and Underground Space Technology* **172**, 107512. <https://doi.org/10.1016/j.tust.2026.107512>.
- Almog-Goldreich, E. & Fuentes, R. (2026). The advanced displacement-confinement method for design of mechanized shield tunnels in soft ground. *Tunnelling and Underground Space Technology* **174**, 107668.
- Bezuijen, A., Talmon, A. M., Kaalberg, F. J. & Plugge, R. (2004). Field measurements of grout pressures during tunnelling of the Sophia Rail Tunnel. *Soils and Foundations* **44**, No. 1, 39–48.
- Bourgeois, E., Berthoz, N., Mohamad, W., Szymkiewicz, F., Le Kouby, A., Branque, D., Michalski, A., Kreziak, C. & Soyez, L. (2023). Results of a benchmark exercise of prediction of tunnel–pile interaction: the TULIP project. *Proceedings of the 10th European Conference on Numerical Methods in Geotechnical Engineering*. <https://doi.org/10.53243/NUMGE2023-82>.
- Calvello, M. & Finno, R. J. (2004). Selecting parameters to optimize in model calibration by inverse analysis. *Computers and Geotechnics* **31**, No. 5, 410–424.
- Cox, D. R. & Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society: Series B* **49**, No. 1, 1–18.
- Dias, T. G. S. & Bezuijen, A. (2017). Grout pressure distribution during TBM tunnelling. In *Geotechnical aspects of underground construction in soft ground*, pp. 153–160. London: CRC Press.
- Duddeck, H. & Erdmann, J. (1985). On structural design models for tunnels in soft soil. *Underground Space* **9**, No. 5–6, 246–259.
- Finno, R. J. & Calvello, M. (2005). Supported excavations: observational method and inverse modeling. *Journal of Geotechnical and Geoenvironmental Engineering* **131**, No. 7, 826–836.
- Hashash, Y. M. A. & Finno, R. J. (2008). Development of new integrated tools for predicting, monitoring, and controlling ground movements due to excavations. *Practice Periodical on Structural Design and Construction* **13**, No. 1, 4–10.

254 Hölter, R., Zhao, C., Mahmoudi, E., Alimardani Lavasan, A., Datcheva, M., König, M. & Schanz, T.
 255 (2018). Optimal measurement design for parameter identification in mechanized tunneling.
 256 *Underground Space* **3**, No. 1, 34–44.

257 Honjo, Y., Liu, W.-T. & Sakajo, S. (1994). Application of Akaike information criterion statistics to
 258 geotechnical inverse analysis: the extended Bayesian method. *Structural Safety* **14**, No. 1–2, 5–
 259 29.

260 Kennedy, M. C. & O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the*
 261 *Royal Statistical Society: Series B* **63**, No. 3, 425–464.

262 Lam, N. N., Docherty, P. D. & Murray, R. (2022). Practical identifiability of parametrised models: a
 263 review of benefits and limitations of various approaches. *Mathematics and Computers in*
 264 *Simulation* **199**, 202–216.

265 Ledesma, A., Gens, A. & Alonso, E. E. (1996a). Estimation of parameters in geotechnical
 266 backanalysis—I. Maximum likelihood approach. *Computers and Geotechnics* **18**, No. 1, 1–27.

267 Ledesma, A., Gens, A. & Alonso, E. E. (1996b). Estimation of parameters in geotechnical
 268 backanalysis—II. Application to a tunnel excavation problem. *Computers and Geotechnics* **18**,
 269 No. 1, 29–46.

270 Levasseur, S., Malécot, Y., Boulon, M. & Flavigny, E. (2009). Statistical inverse analysis based on
 271 genetic algorithm and principal component analysis: method and developments using synthetic
 272 data. *International Journal for Numerical and Analytical Methods in Geomechanics* **33**, No. 12,
 273 1485–1511.

274 Michalski, A., Branque, D., Berthoz, N., Rallu, A., Mohamad, W., Szymkiewicz, F., Le Kouby, A. &
 275 Bourgeois, E. (2024). Greenfield response to EPBM tunneling in Paris and relations with TBM
 276 operation variables. *Journal of Geotechnical and Geoenvironmental Engineering* **150**, No. 10,
 277 05024011.

278 O'Reilly, M. P. & New, B. M. (1983). Settlement above tunnels in the United Kingdom—their
 279 magnitude and prediction. *International Journal of Rock Mechanics and Mining Sciences &*
 280 *Geomechanics Abstracts* **20**, No. 1, A18.

281 Peck, R. B. (1969). Advantages and limitations of the observational method in applied soil mechanics.
 282 *Géotechnique* **19**, No. 2, 171–187.

283 Potts, D. M. (2003). Numerical analysis: a virtual dream or practical reality? *Géotechnique* **53**, No. 6,
 284 535–573.

285 Schweiger, H. F. (2002). Results from numerical benchmark exercises in geotechnics. *Proceedings of*
 286 *the 5th European Conference on Numerical Methods in Geotechnical Engineering*, Paris, pp.
 287 305–314.

288 Spross, J. & Johansson, F. (2017). When is the observational method in geotechnical engineering
 289 favourable? *Structural Safety* **66**, 17–26.