Application of the Bootstrap Method for Optimal Sensor Location

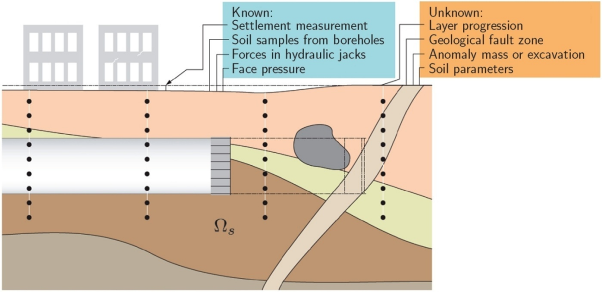
Raoul Hölter, Elham Mahmoudi, Tom Schanz & Markus König

Department of Civil and Environmental Engineering
Ruhr Universität Bochum, Germany

Motivation

Overall objectives

- Development of optimised measurement concepts:
  "Design of Experiment" applied to geotechnics in particular
- Identify optimal set-up including:
  time, position, measurement accuracy, amount of sensors, type of measurement device
- Application of the bootstrap method for DoE



Concept "Design of Experiment"

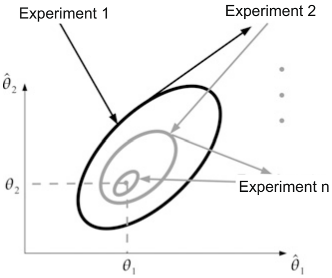
Concept

How to design an "Experiment"(Monitoring), to gain the most reliable measurements for an inverse analysis?

- Objective: identify model parameters θ, using inverse analysis: $\theta = f^{-1}(\tilde{y})$
- The more precise θ is identified, the smaller the discrepancy between model response y and measurements $\tilde{y}$
- Problem: inverse analysis $\theta = f^{-1}(\tilde{y})$ is being falsified by:
  - Model uncertainty
  - Measurement errors
  - Inhomogeneity of subsoil

  $\Longrightarrow$ Parameter are identified, but affected with errors(variance, COV)
- Model/system response is depending on soil parameters and "experiment"/monitoring: $y = f(\theta, X)$

Which X allows $f^{-1}(\tilde{y})$, in a way that $COV(\theta) = \min$?

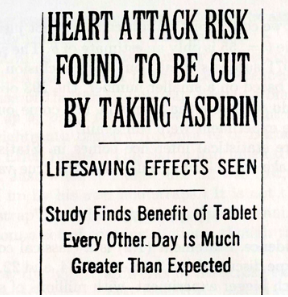


Concept of DoE-process for a two-dimensional parameters space (Schenkendorf et al., 2009)

"Bootstrap"-Approach

Concept

- Initially developed as resampling method to augment the information content of a given statistical sample (Efron, 1979)
- How to increase the accuracy without additional data?
- Create new populations from existing data with same distribution
- Identify statistics as means of large number of samples
  $\Longrightarrow$ Increase accuracy without increase of database



HEART ATTACK RISK FOUND TO BE CUT BY TAKING ASPIRIN

LIFESAVING EFFECTS SEEN

Study Finds Benefit of Tablet Every Other Day Is Much Greater Than Expected

New York Times, 27.01.1987

	Subjects	heart attacks	heart strokes
Aspirin group	11037	104	119
Placebo group	11034	189	98

"Bootstrap"-Approach

Concept

- Initially developed as resampling method to augment the information content of a given statistical sample (Efron, 1979)
- How to increase the accuracy without additional data?
- Create new populations from existing data with same distribution
- Identify statistics as means of large number of samples
  $\Longrightarrow$ Increase accuracy without increase of database

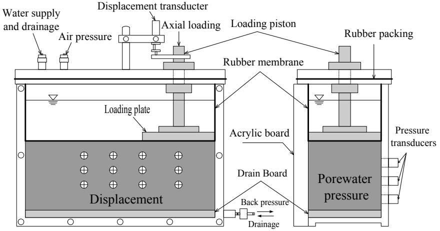
	Subjects	heart attacks	heart strokes
Aspirin group	11037	104	119
Placebo group	11034	189	98

- Evaluation heart attacks:
  Estimate value
  $\frac{104/11037}{189/11034} = 0.55 = \hat{\mu} \neq \mu$
  0.95 confidence interval:
  $0.43 < \hat{\mu} < 0.70$
- Evaluation heart strokes:
  Estimate value
  $\frac{119/11037}{98/11034} = 1.21 = \hat{\mu} \neq \mu$
  0.95 confidence interval:
  $0.93 < \hat{\mu} < 1.59$
- After drawing of 1000 samples with replacement from initial population:
  - 119 "ones", 10918 "zeros"
  - 98 "ones", 10936 "zeros"
- Confidence interval based on new data: $1.04 < \hat{\mu} < 1.38$

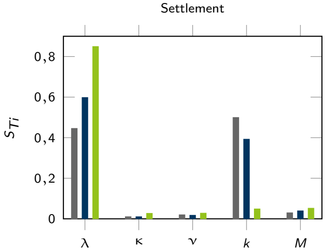
Inroducing the reference experiment

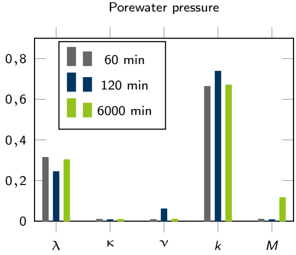
Procedure

- Soft clay sample undergoes stepwise loading
- Limited drainage possibilities provoke long time consolidation behaviour
- Extensive measurement set-up allows parameter identification
- GSA allows identification of relevant parameters

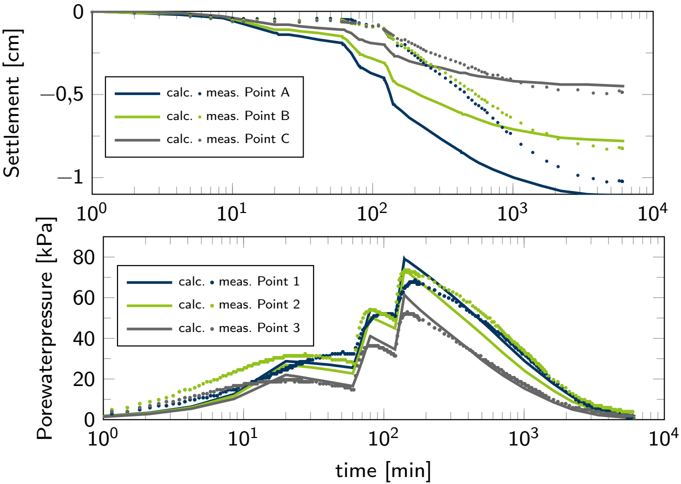


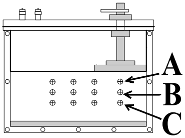
experimental setup

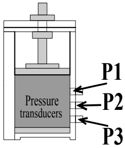




Optimised parameters, assuming closed boundaries





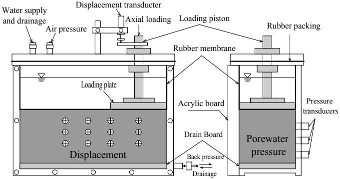


Employment of Bootstrap method:

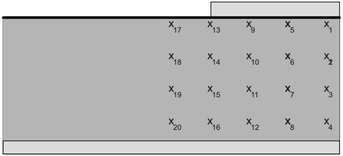
“Using sampling from the sample to model sampling from the population”

Initial situation

- Numerical model with “known” parameters: From GSA, λ and k are of interest
- Primary design definition: Three sensors for pore water pressure and for three displacements
- General specification of possible measurement positions: 20 possible positions assumed $\Rightarrow \binom{20}{3}^2 = 1,299,600$ combinations
- Known distribution type and variance of measurement devices: Gaussian white noise, COV $= 20\%$



Initial experimental setup



Distribution of possible measurement points

Application on current subject

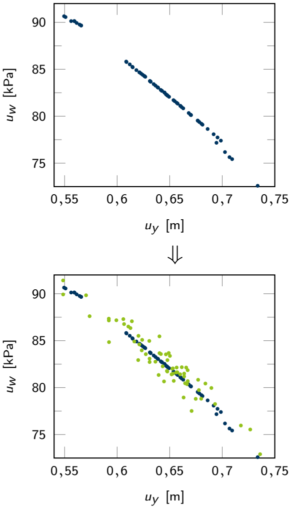
Proceeding

1. 20 Positions $P_j(X, Y)$ are chosen as possible locations for the three sensors
2. Metamodel g creates output values $U(u_y, u_w)$ out of input parameters $\theta(k, \lambda)$
3. 100 new random samples within the former ranges of U (inappropriate, "wrong ", noisy)
4. PI is performed on each samples for different positions:

$$\theta_i = g^{-1}(U_i, P_j) \Longrightarrow \overline{\theta} = \sum_{N}^{i=0} \theta_i$$

5. Covariance matrix of the identified parameters is calculated:

$$C_\theta = \frac{1}{N-1} \sum_{N}^{i=0} (\theta_i - \overline{\theta}) \cdot (\theta_i - \overline{\theta})^T$$

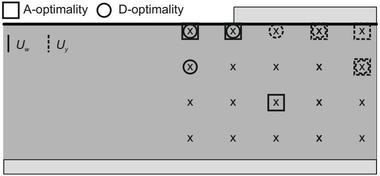


Results of numerical simulation and artificial noisy values

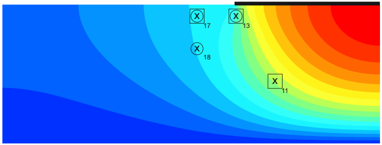
Evaluation I

Identification of Sensor positions:

- C_θ is identified for each possible combination of P_j
- To quantify the quality of the results, an optimality criterion is applied on each C_θ:
  A-optimal design $\Phi_A(C_\theta) = \mathrm{tr}(C_\theta)$
  D-optimal design $\Phi_D(C_\theta) = \det(C_\theta)$
- The smaller the criterion is, the more optimal the considered combination
- The most informative sensor locations for settlements are suggested below the loading plate
- It is recommended to measure pore water pressures in the midfield of the device



Suggested measurement points for $\Phi(C_\theta) = \min$

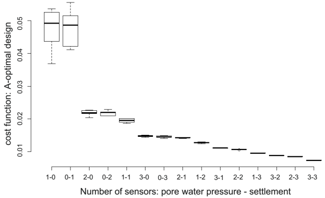


Suggested pwp-sensors with pwp-distribution

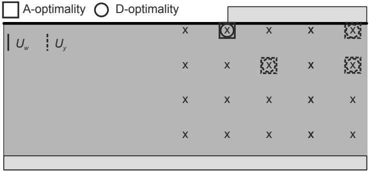
Evaluation II

How many sensors should be used?

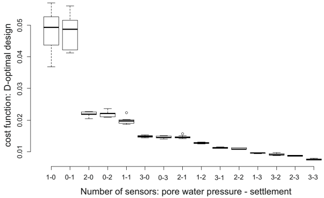
- Variation of the number of sensors, from [0, 0] to [3, 3]
- Identification of optimal positions in each case
- Normalised cost function $\Phi(C_\theta)$ allows comparison of different set-ups



Cost function values, using $\Phi_A(C_\theta)$



Suggested measurement points for set-up 1-3



Cost function values, using $\Phi_D(C_\theta)$

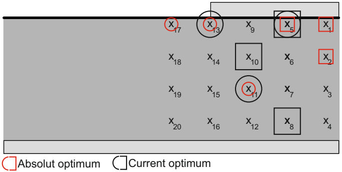
Reducing the computational effort

Approach

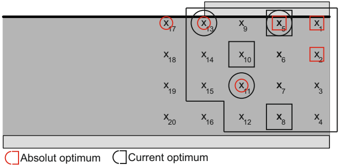
- Computational costs are bottleneck of each statistical application
- Iterative exploration of design space (inspired by subset simulation)

Procedure

- Randomly run 400 samples
- Accuracy estimation: $f_r = \frac{c_{f,Best} - c_{f,Opt}}{c_{f,Worst} - c_{f,Opt}}$
- Best set-up is selected, new search area defined in near field, two further iterations
- After third iteration, all identified positions from earlier steps are considered
- In the final confined area, all set-ups are tested
- 1900 instead of 1.2 M combinations tested



Best set-up of random sampling $f_r = 0.1513$



Reduced area to consider settlements

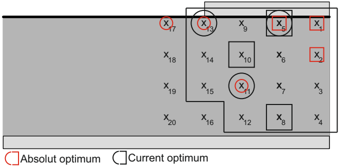
Reducing the computational effort

Approach

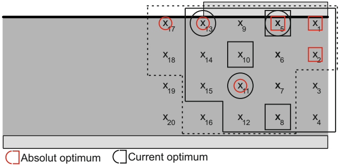
- Computational costs are bottleneck of each statistical application
- Iterative exploration of design space (inspired by subset simulation)

Procedure

- Randomly run 400 samples
- Accuracy estimation: $f_r = \frac{c_{f,Best} - c_{f,Opt}}{c_{f,Worst} - c_{f,Opt}}$
- Best set-up is selected, new search area defined in near field, two further iterations
- After third iteration, all identified positions from earlier steps are considered
- In the final confined area, all set-ups are tested
- 1900 instead of 1.2 M combinations tested



Reduced area to consider settlements



Reduced areas to consider both outputs

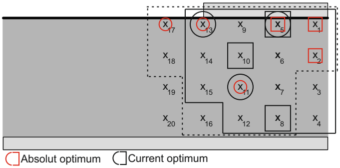
Reducing the computational effort

Approach

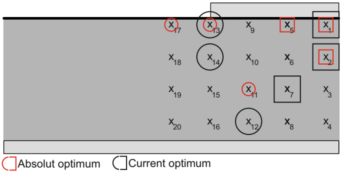
- Computational costs are bottleneck of each statistical application
- Iterative exploration of design space (inspired by subset simulation)

Procedure

- Randomly run 400 samples
- Accuracy estimation: $f_r = \frac{c_{f,Best} - c_{f,Opt}}{c_{f,Worst} - c_{f,Opt}}$
- Best set-up is selected, new search area defined in near field, two further iterations
- After third iteration, all identified positions from earlier steps are considered
- In the final confined area, all set-ups are tested
- 1900 instead of 1.2 M combinations tested



Reduced areas to consider both outputs



Best set-up after second random sampling $f_r = 0.1401$

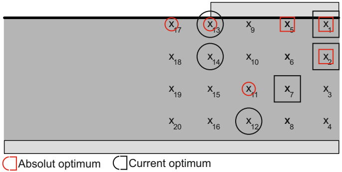
Reducing the computational effort

Approach

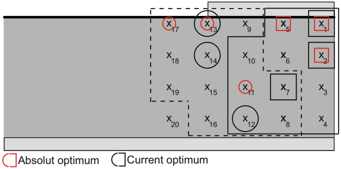
- Computational costs are bottleneck of each statistical application
- Iterative exploration of design space (inspired by subset simulation)

Procedure

- Randomly run 400 samples
- Accuracy estimation: $f_r = \frac{c_{f,Best} - c_{f,Opt}}{c_{f,Worst} - c_{f,Opt}}$
- Best set-up is selected, new search area defined in near field, two further iterations
- After third iteration, all identified positions from earlier steps are considered
- In the final confined area, all set-ups are tested
- 1900 instead of 1.2 M combinations tested



Best set-up after second random sampling $f_r = 0.1401$



Reduced areas for third sampling

RUB

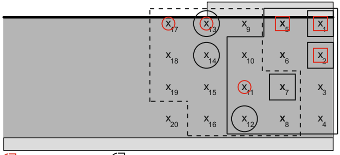
Reducing the computational effort

Approach

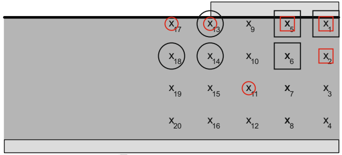
- Computational costs are bottleneck of each statistical application
- Iterative exploration of design space (inspired by subset simulation)

Procedure

- Randomly run 400 samples
- Accuracy estimation: $f_r = \frac{c_{f,Best} - c_{f,Opt}}{c_{f,Worst} - c_{f,Opt}}$
- Best set-up is selected, new search area defined in near field, two further iterations
- After third iteration, all identified positions from earlier steps are considered
- In the final confined area, all set-ups are tested
- 1900 instead of 1.2 M combinations tested



Reduced areas for third sampling



Best set-up after third random sampling $f_r = 0.1256$

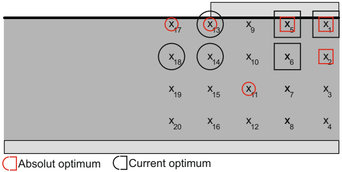
Reducing the computational effort

Approach

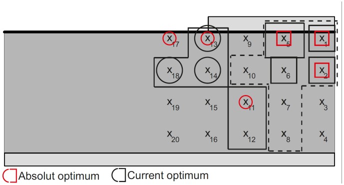
- Computational costs are bottleneck of each statistical application
- Iterative exploration of design space (inspired by subset simulation)

Procedure

- Randomly run 400 samples
- Accuracy estimation: $f_r = \frac{c_{f,Best} - c_{f,Opt}}{c_{f,Worst} - c_{f,Opt}}$
- Best set-up is selected, new search area defined in near field, two further iterations
- After third iteration, all identified positions from earlier steps are considered
- In the final confined area, all set-ups are tested
- 1900 instead of 1.2 M combinations tested



Best set-up after third random sampling $f_r = 0.1256$



Area for final optimisation run

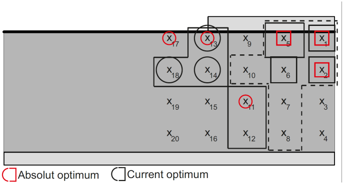
Reducing the computational effort

Approach

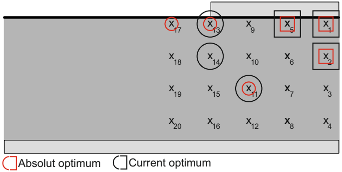
- Computational costs are bottleneck of each statistical application
- Iterative exploration of design space (inspired by subset simulation)

Procedure

- Randomly run 400 samples
- Accuracy estimation: $f_r = \frac{c_{f,Best} - c_{f,Opt}}{c_{f,Worst} - c_{f,Opt}}$
- Best set-up is selected, new search area defined in near field, two further iterations
- After third iteration, all identified positions from earlier steps are considered
- In the final confined area, all set-ups are tested
- 1900 instead of 1.2 M combinations tested



Area for final optimisation run



Finally identified set-up $f_r = 0.0145$

Outlook

Conclusion

- Approach to reduce uncertainties in geotechnical investigation by creating a rational measurement design
- Application to a well documented experiment
- Considerable reduction of computational effort

Next steps

- Further consideration of measurement uncertainties (higher order uncertainty)
- Application of Sequential Bayesian DoE or Bayesian learning for DoE
- Application to 3D-Tunneling problems $\Rightarrow$ time dependency
- Further improvement and validation of statistical methods to reduce computational efforts

Thank you for your attention!

Raoul Hölter
Department of Civil and Environmental Engineering
Ruhr-Universität Bochum (Germany)
Raoul.Hoelter@rub.de