dynardo | presented at the 16th Weimar Optimization and Stochastic Days 2019 |
Source: www.dynardo.de/en/library

PROBABILISTIC INTELLIGENCE

Deep Gaussian Covariance Network

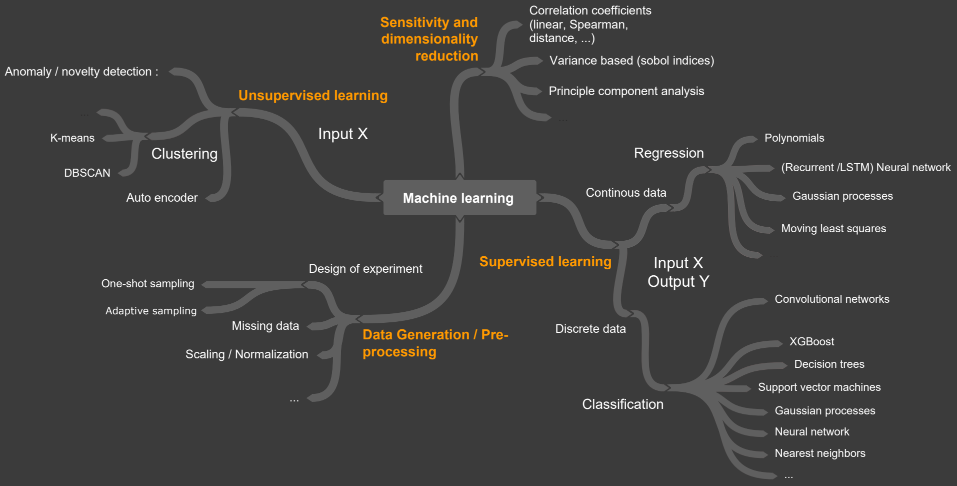
Machine Learning based on Probabilistic Intelligence

Kevin Cremanns[1], Can Bogoclu[1], Prof. Dr.-Ing. Dirk Roos[2]

WOST 2019 - 06/07.06.2019

[1]PI Probaligence GmbH
[2]Niederrhein University of Applied Sciences -
IMH Institute of Modelling and High-Performance Computing

Overview machine learning algorithms

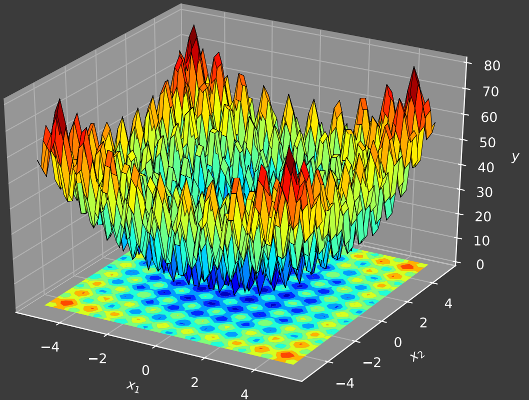


Requirements

What should a machine learning algorithm be capable of?

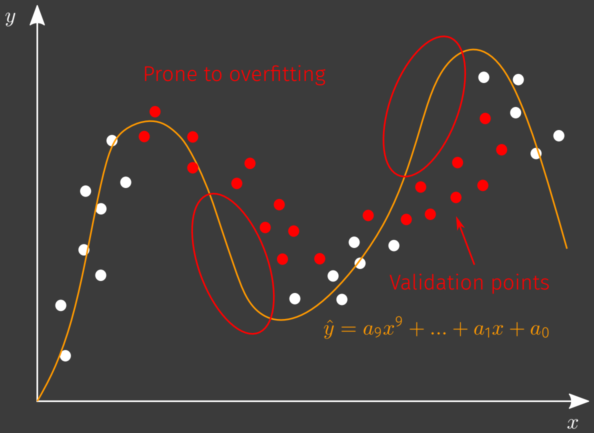
Requirements

- Maximum flexibility



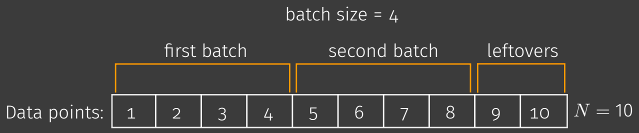
Requirements

- Maximum flexibility
- Good generalization



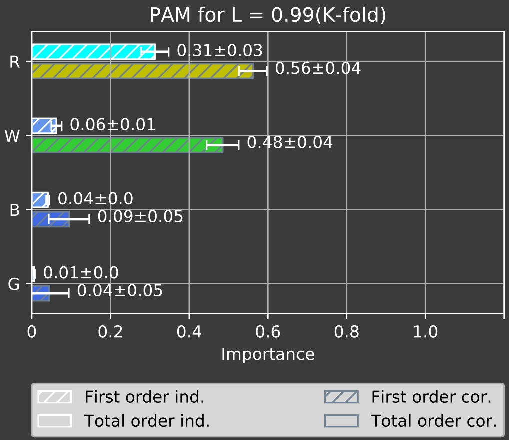
Requirements

- Maximum flexibility
- Good generalization
- Scalability



Requirements

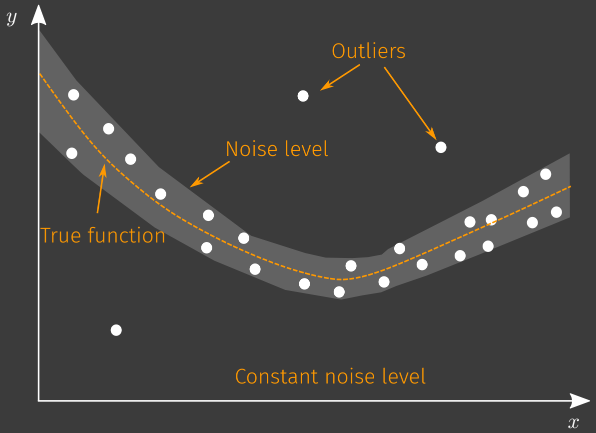
- Maximum flexibility
- Good generalization
- Scalability
- Automatic design space reduction



PAM = Predictive accuracy of meta model

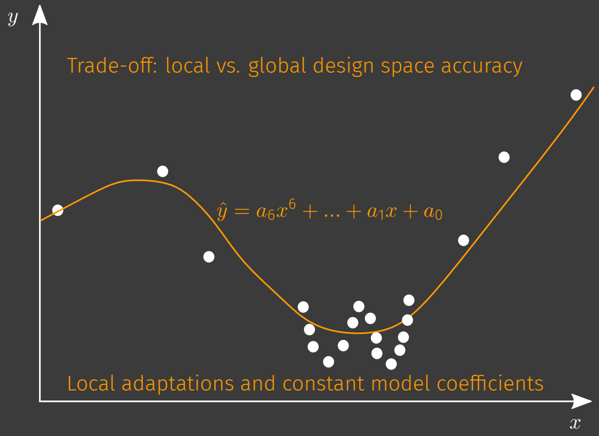
Requirements

- Maximum flexibility
- Good generalization
- Scalability
- Automatic design space reduction
- Robust against noise and outliers



Requirements

- Maximum flexibility
- Good generalization
- Scalability
- Automatic design space reduction
- Robust against noise and outliers
- Non-stationarity



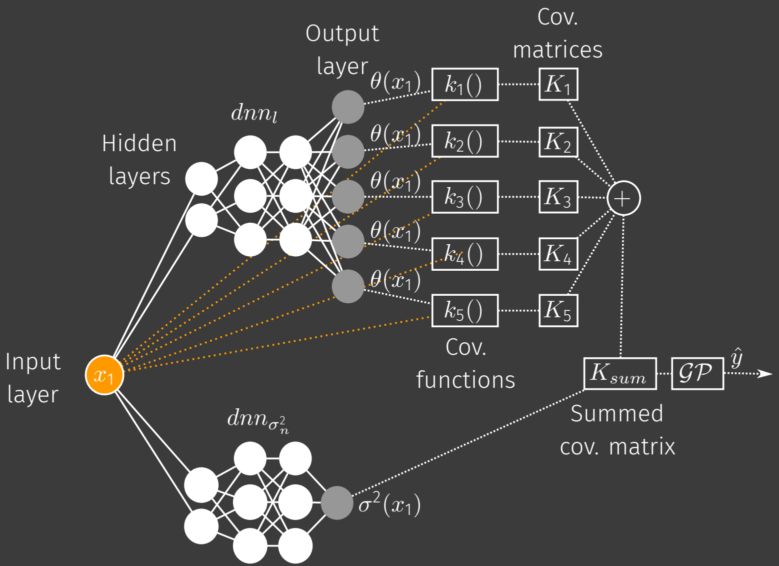
Requirements

- Maximum flexibility
- Good generalization
- Scalability
- Automatic design space reduction
- Robust against noise and outliers
- Non-stationarity
- Easy to use

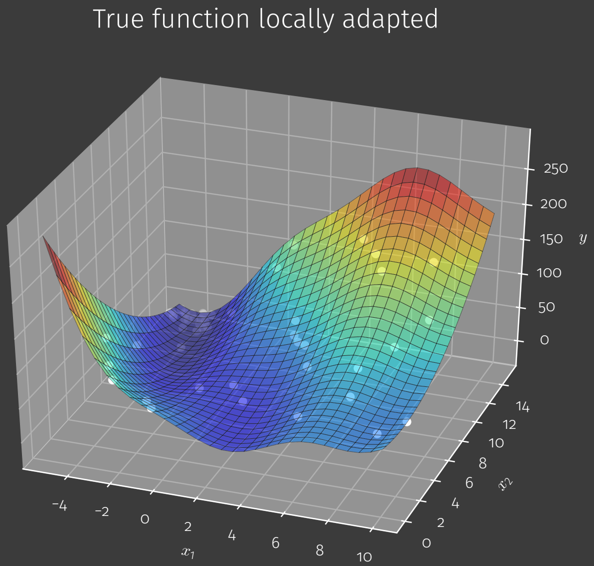
No user settings / hyperparameters!

Deep Gaussian covariance network

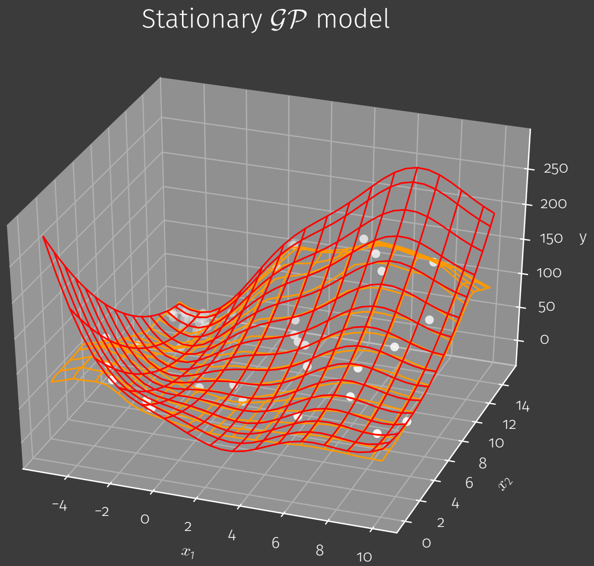
Deep Gaussian covariance network - *DGCN*



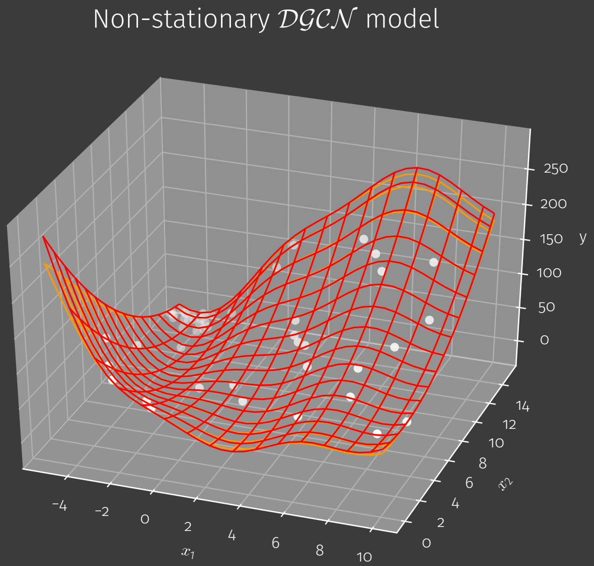
Example non-stationary length-scales



Example non-stationary length-scales

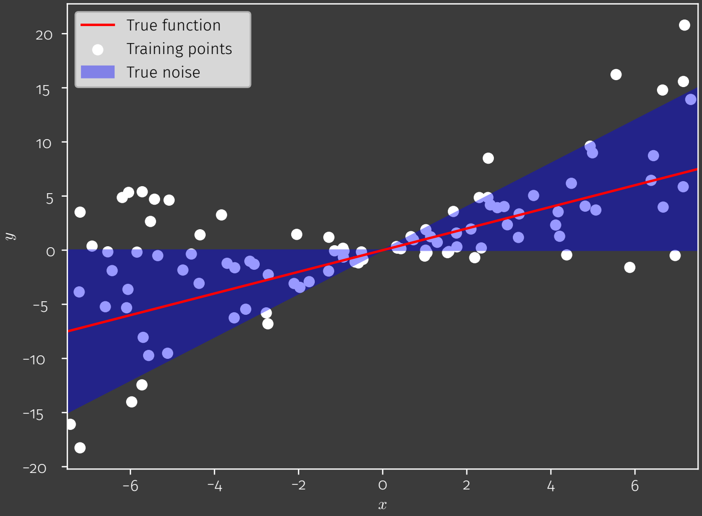


Example non-stationary length-scales



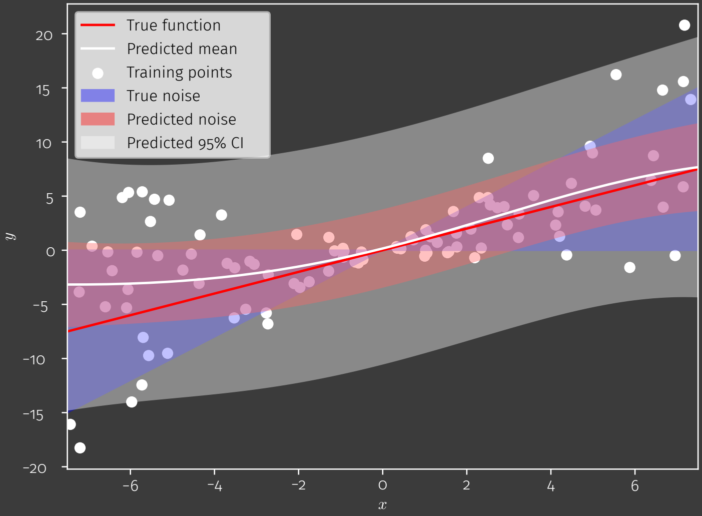
Noisy data and $\mathcal{DGCN}$

True function:



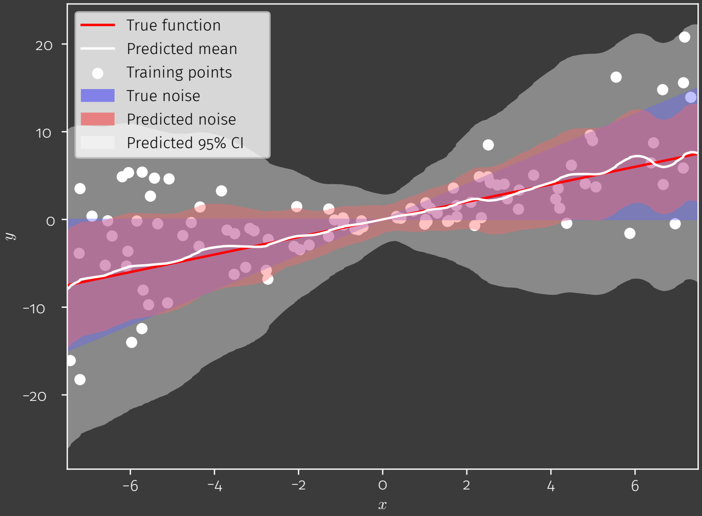
Noisy data and $\mathcal{DGCN}$

Stationary $\mathcal{GP}$ with constant noise:



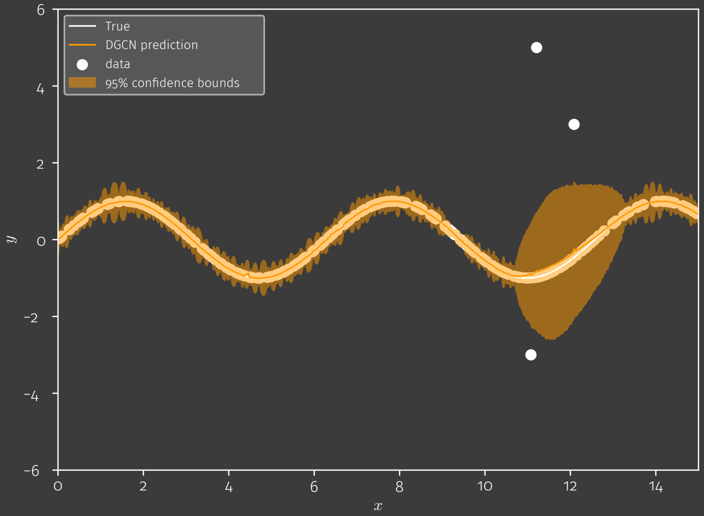
Noisy data and $\mathcal{DGCN}$

Non-stationary $\mathcal{DGCN}$ with point dependent noise:



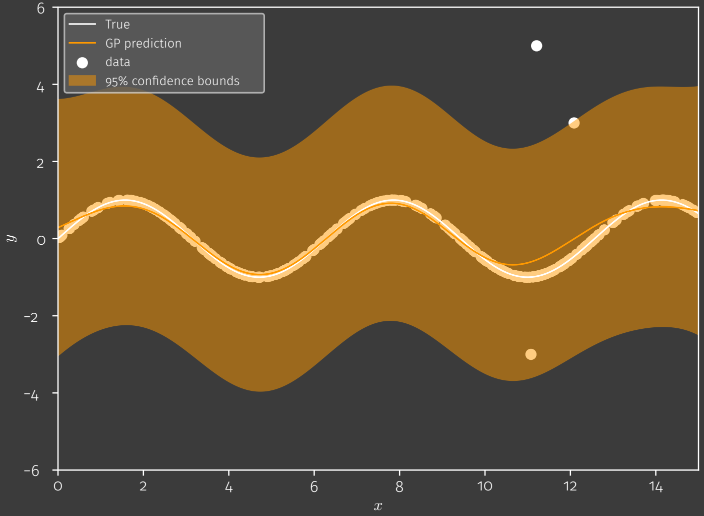
Noisy data and $\mathcal{DGCN}$

Automatic outlier avoidance with dynamic noise level ($\mathcal{DGCN}$)



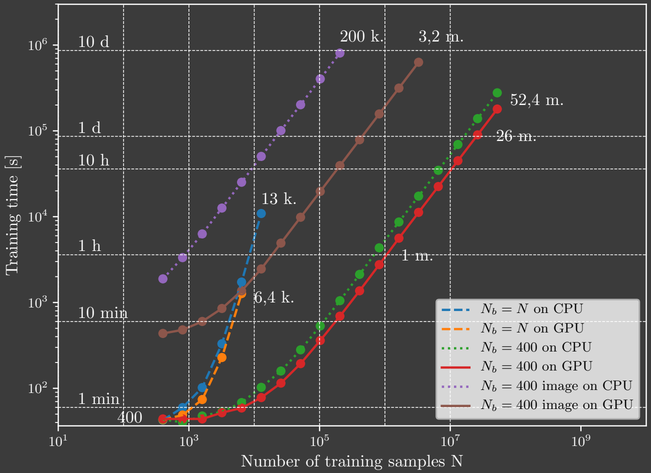
Noisy data and $\mathcal{DGCN}$

Automatic outlier avoidance with constant noise level ($\mathcal{GP}$)



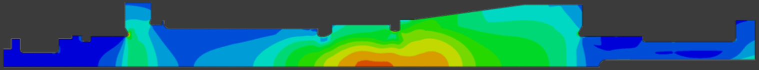
Scalability

- $\mathcal{DGCN}$ can use batch / online learning for high scalability.
- Training on CPU / GPU or even distributed on clusters.

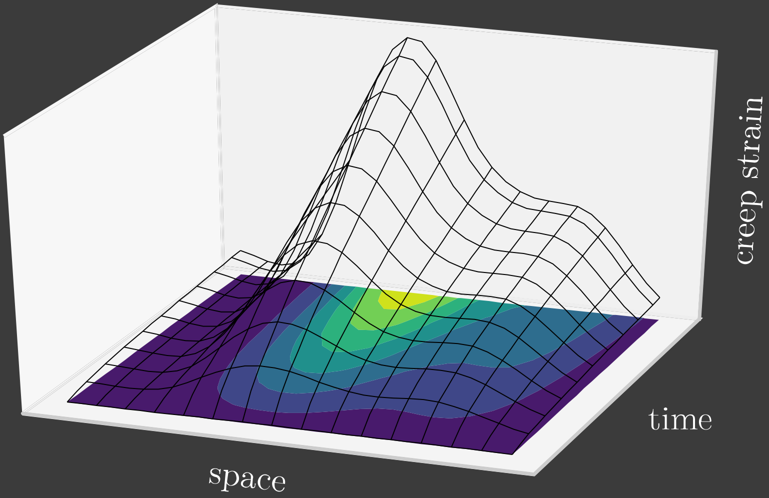


Creep strain estimation (1/4) - problem description

- Creep strain of a material depends on the temperature and pressure history over time $\rightarrow$ expensive to simulate.
- 300 simulations were used to train a $\mathcal{DGCN}$ model. This model should be utilized for probabilistic assessments with about 14,000 evaluations.
- The results of 31 finite element nodes in the rotor center over 34 time steps (1,054 outputs) should be analyzed $\rightarrow$ correlation over space and time.

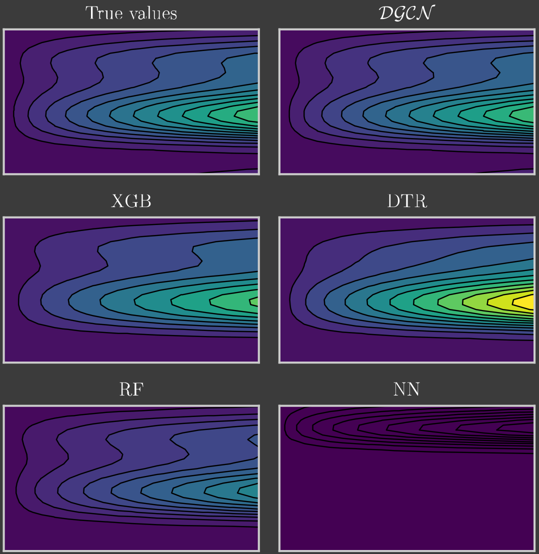


Creep strain estimation (2/4) - one example



Creep strain estimation (3/4) - visual validation

- $\mathcal{DGCN}$=Deep Gaussian Covariance Network
- XGB=XGBoost
- DTR=Decision trees
- RF=Random forest
- NN=Neural network

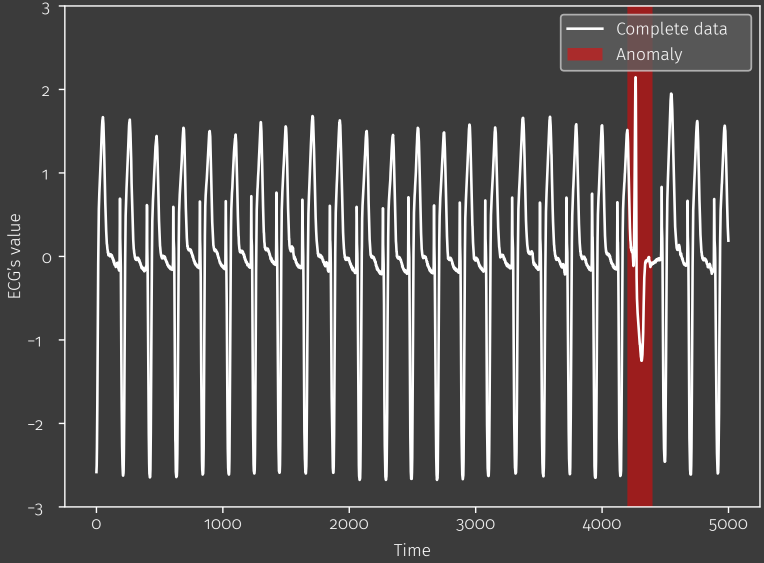


Creep strain estimation (4/4) - numerical validation

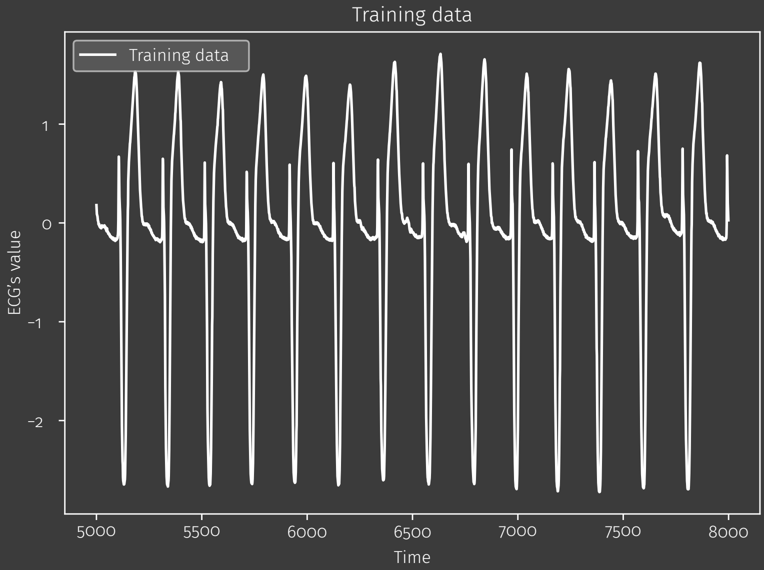
Algorithm	MMAPE with PCA	MMAPE without PCA
$\mathcal{DGCN}$	1.71	-
XGB	36	31
RF	289	253
DTR	396	1265
NN	218065	121321

Table 1: Comparison of $\mathcal{DGCN}$, neural networks (NN), random forests (RF), XGboost (XGB), decision trees (DTR) with PCA and without PCA for 21 test cases of the creep strain approximation. The mean MAPE (MMAPE) metric in [%] over all 21 multi-output sequences are estimated.

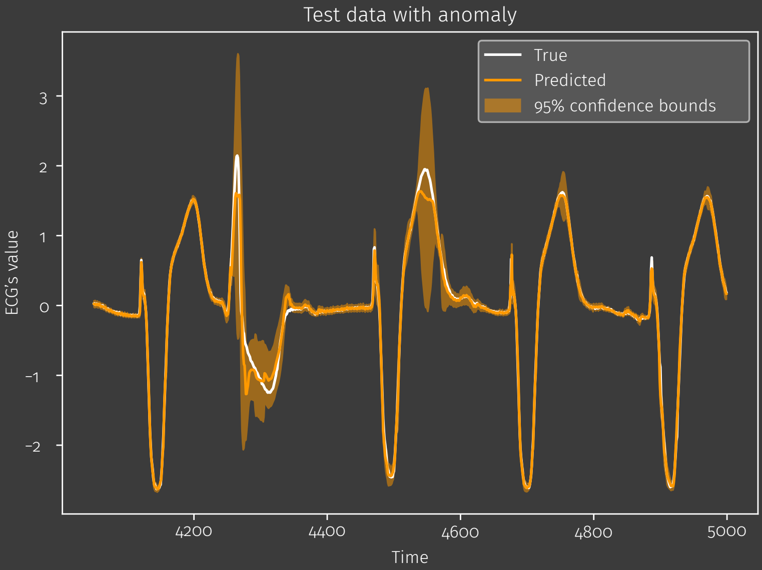
ECG anomaly detection (1/3)



ECG anomaly detection (2/3)



ECG anomaly detection (3/3)



Bayesian optimization

Adaptive design of experiment

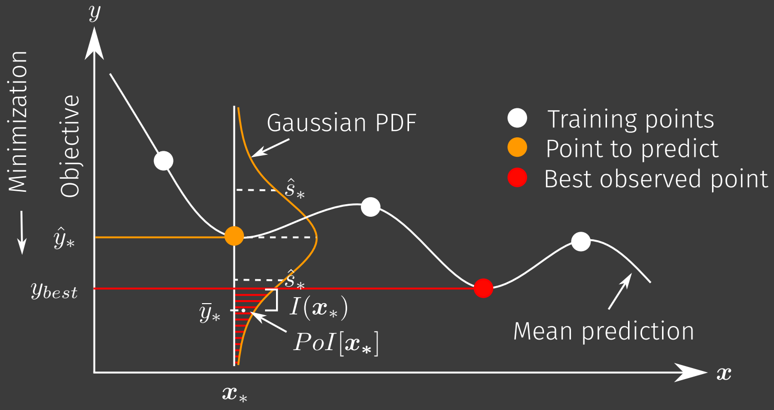
The ability to adapt locally to the design space makes $\mathcal{DGCN}$ particularly suitable for adaptive sampling methods like:

- Expected improvement (EI) for single or multi-objective optimization (not contrary).
- Expected volume improvement (EVI) for multi-objective Pareto optimization.
- Variance based global model improvement (Var).

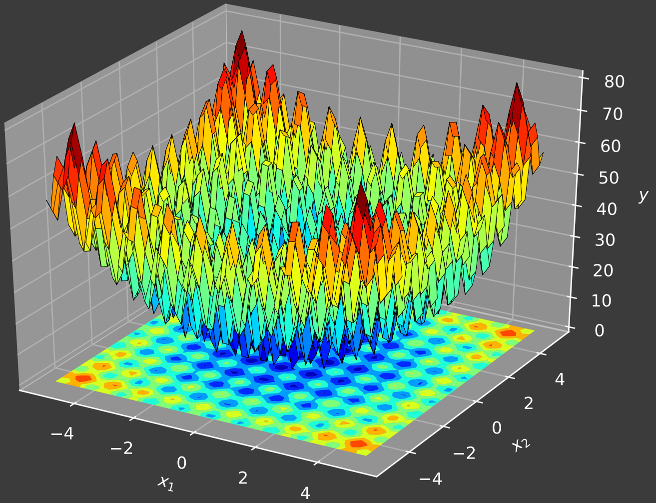
Expected improvement

$$EI(\boldsymbol{x}_*) = \int_{-\infty}^{y_{best}} I(\boldsymbol{x}_*) PoI(\boldsymbol{x}_*) dy_* = \dots$$

$$= (y_{best} - \hat{y}_*)\, \Phi\left(\frac{y_{best} - \hat{y}_*}{\hat{s}_*}\right) + \hat{s}_* \phi \left(\frac{y_{best} - \hat{y}_*}{\hat{s}_*}\right)$$

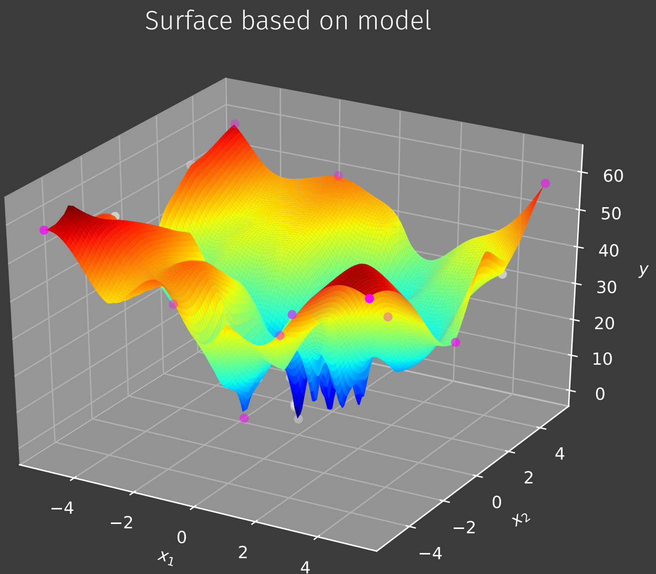


EI - Rastrigin example (1/3)



EI - Rastrigin example (2/3)

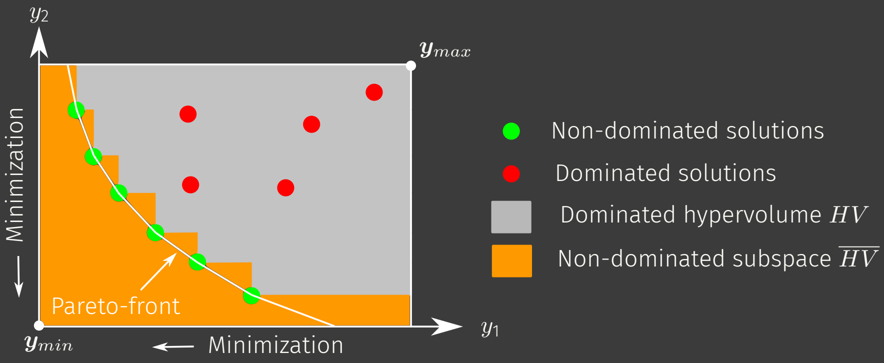
EI - Rastrigin example (3/3)



EI - avoidance of unfeasible areas

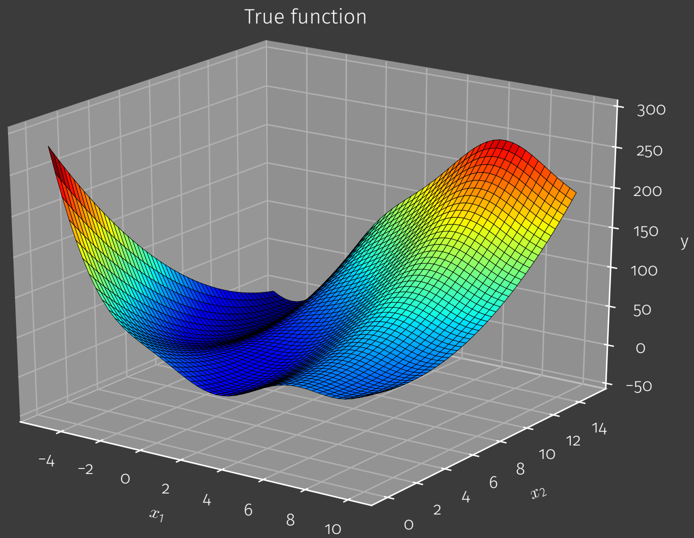
Expected volume improvement

$$EI_{hyp}(\boldsymbol{x}_*) = \int_{\boldsymbol{y} \in \overline{HV}} I_{hyp}(\boldsymbol{x}_*) PDF_{\boldsymbol{x}}(\boldsymbol{y}) d\boldsymbol{y} \prod_{i=1}^{n_{con}} \Phi\left(\frac{\hat{y}_{i*}}{\hat{s}_{i*}}\right)$$

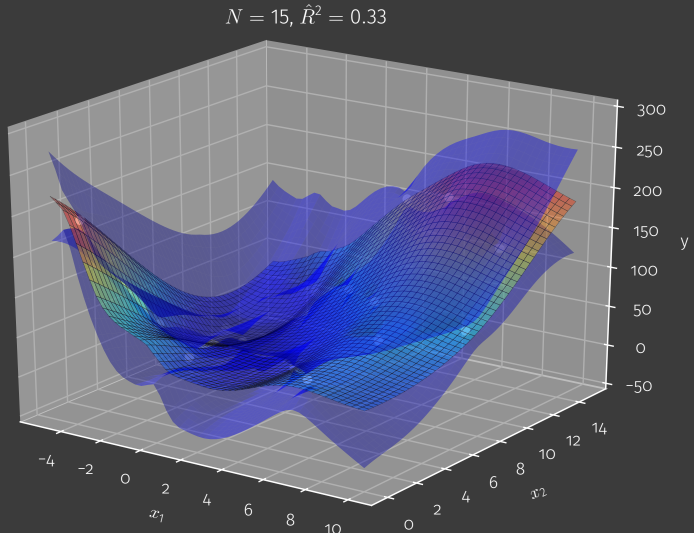


Expected hypervolume improvement - example

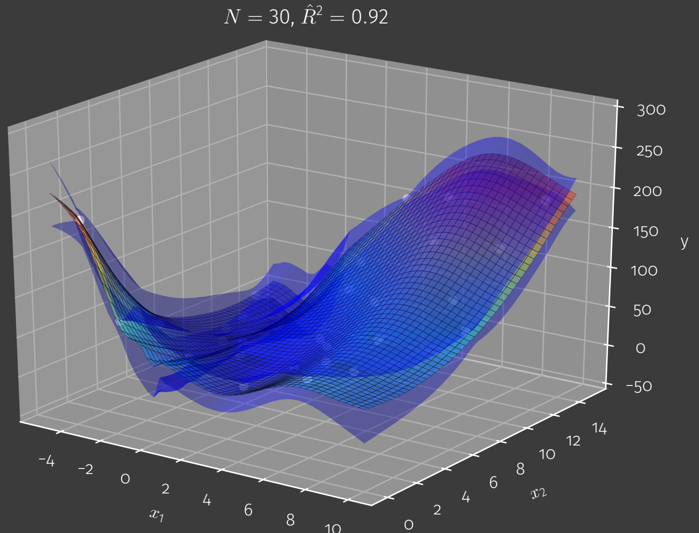
Model variance improvement



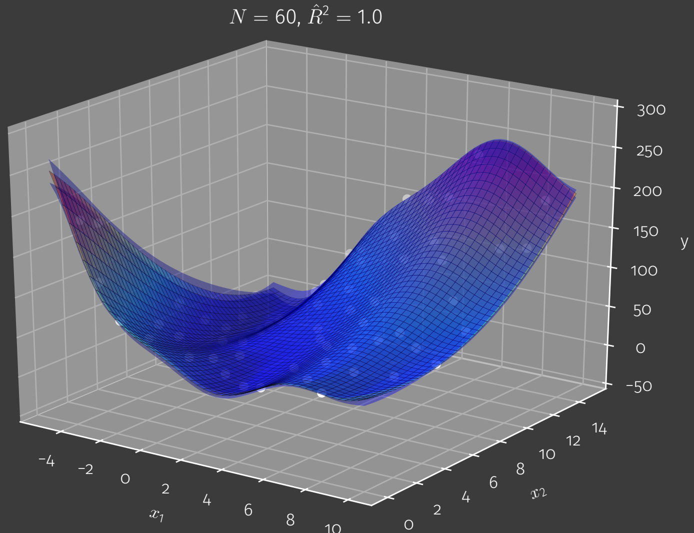
Model variance improvement



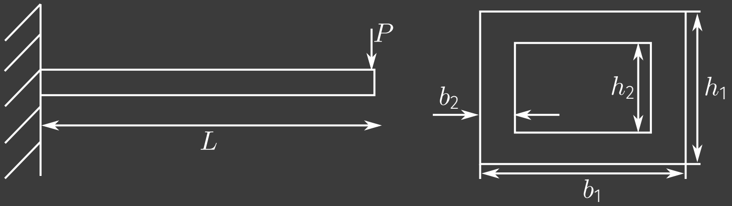
Model variance improvement



Model variance improvement

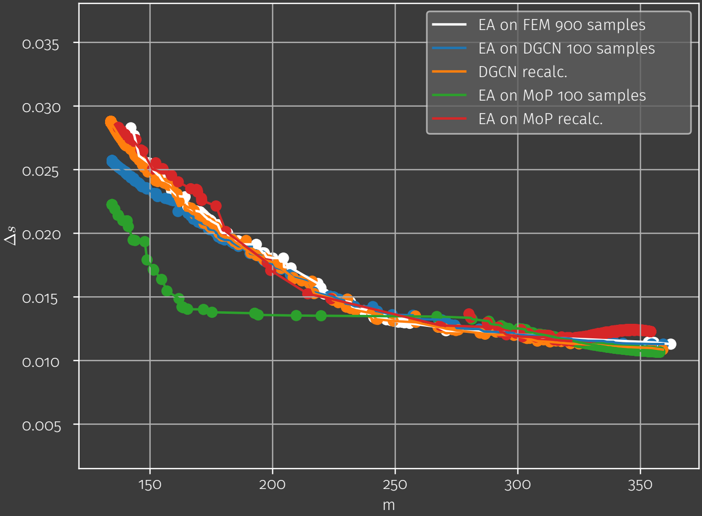


- 7 parameters (b_1, b_2, h_1, h_2, L, E, ρ)
- 5 constraints (stress, reaction force, Eigenfrequency, Δs, m)
- 2 objectives (Δs, m)
- Solved via FEM



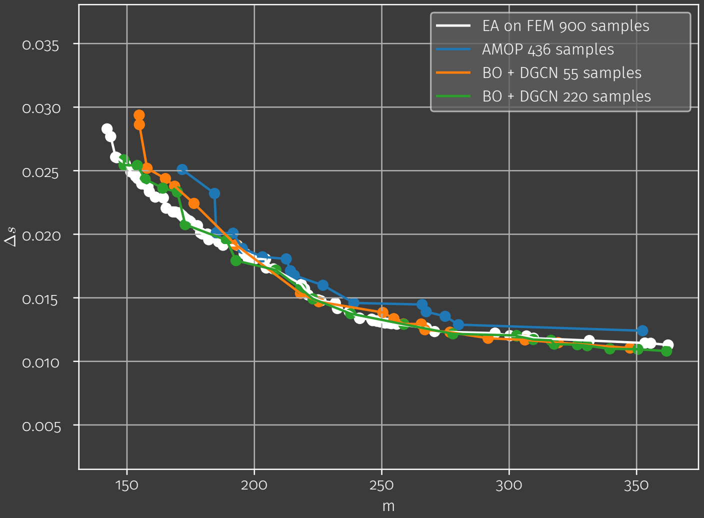
Beam example (2/3)

Optimization via EA ($N = 10,000$) on MOP / $\mathcal{DGCN}$ models trained with 100 samples, reference EA on FEM simulation ($N = 900$).



Beam example (3/3)

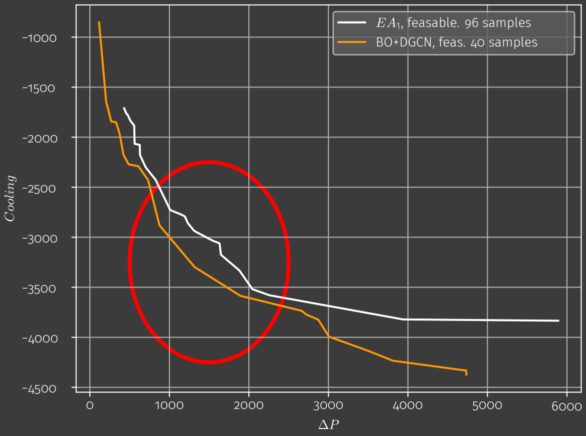
Adaptive optimization via AMOP ($N = 436$) / BO+$\mathcal{DGCN}$ ($N = 55/220$), reference EA on FEM simulation ($N = 900$)



Pareto optimization

- Electronic chip design optimization.
- 5 optimization parameters.
- 30 dependent parameters.
- Increase cooling performance.
- Decrease pressure loss.

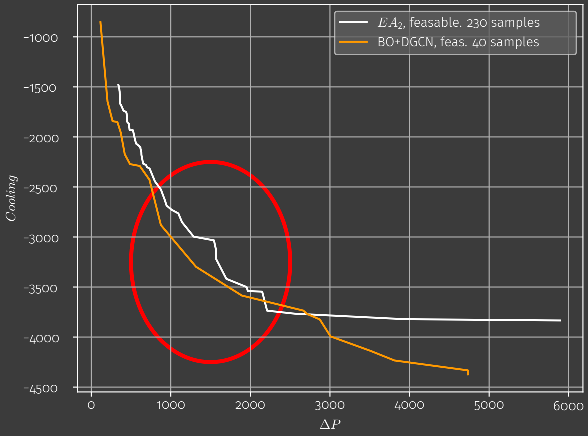
96 num. opt. vs. 40 BO+$\mathcal{DGCN}$ samples.



Pareto optimization

- Electronic chip design optimization.
- 5 optimization parameters.
- 30 dependent parameters.
- Increase cooling performance.
- Decrease pressure loss.

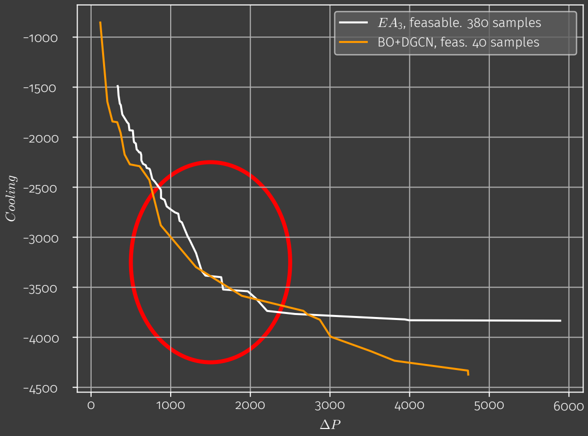
230 num. opt. vs. 40 BO+$\mathcal{DGCN}$ samples.



Pareto optimization

- Electronic chip design optimization.
- 5 optimization parameters.
- 30 dependent parameters.
- Increase cooling performance.
- Decrease pressure loss.

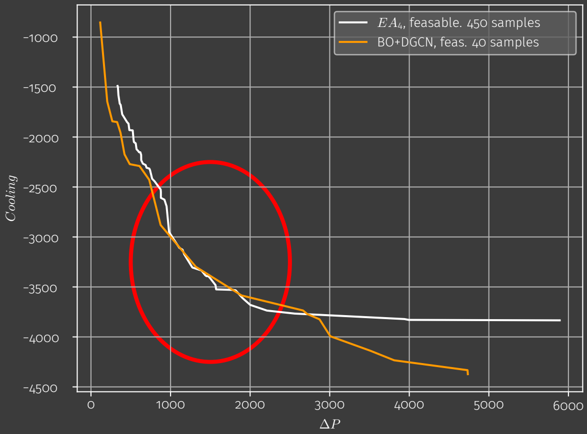
380 num. opt. vs. 40 BO+$\mathcal{DGCN}$ samples.



Pareto optimization

- Electronic chip design optimization.
- 5 optimization parameters.
- 30 dependent parameters.
- Increase cooling performance.
- Decrease pressure loss.

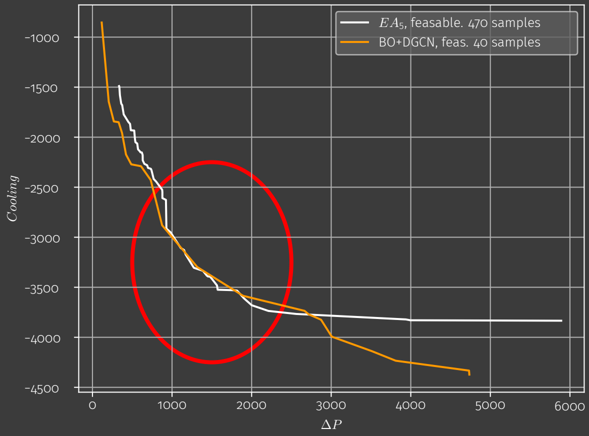
450 num. opt. vs. 40 BO+$\mathcal{DGCN}$ samples.



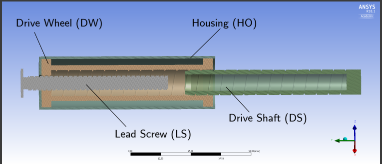
Pareto optimization

- Electronic chip design optimization.
- 5 optimization parameters.
- 30 dependent parameters.
- Increase cooling performance.
- Decrease pressure loss.

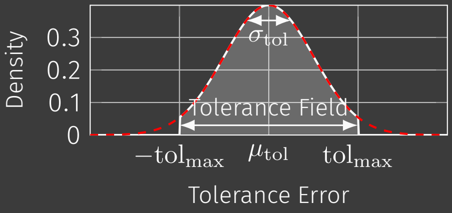
470 num. opt. vs. 40 BO+$\mathcal{DGCN}$ samples.



Reliability-based tolerance optimization of an insulin pen



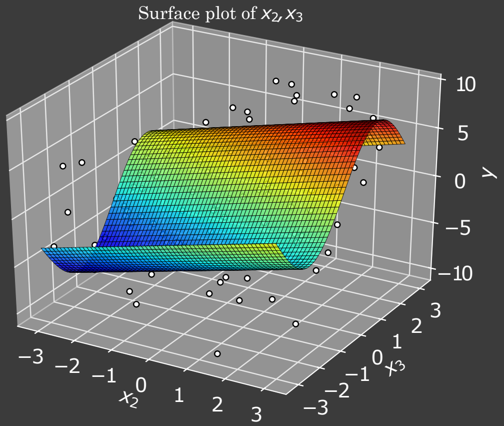
- Total of 40 stochastic model parameters.
- Failure defined as higher than 1% misdosing.
- Tolerance field is to be maximized.
- Reliability-based stochastic design optimization on $\mathcal{DGCN}$.
- 90 training samples.
- Optimization with FEM approx. 7,000 calculations.



Result	$\sum_i \text{tol}.i$	$P(\mathcal{F})$
Reference	0 mm	0
Optimum (Approx.)	3.963 mm	$2.2 \cdot 10^{-7}$
Optimum (Val.)[a]	3.963 mm	0

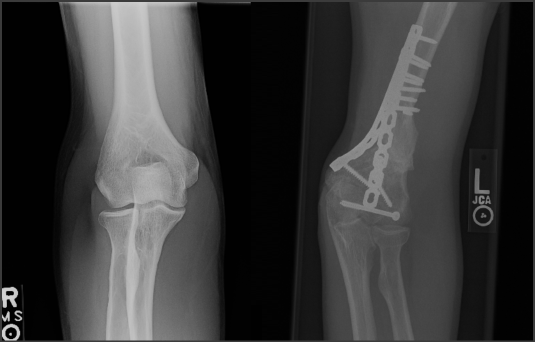
[a] 70 designs with Form for validation

- (Image) Regression.



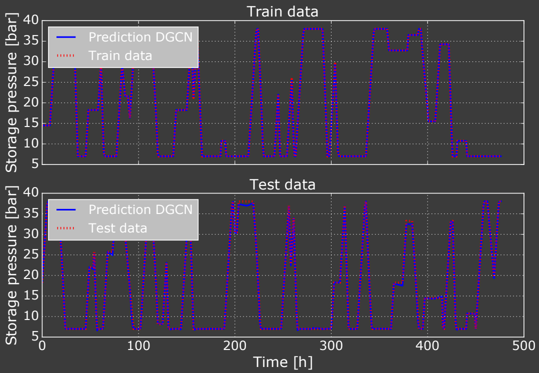
$\mathcal{DGCN}$ - capabilities

- (Image) Regression.
- (Image) Classification.



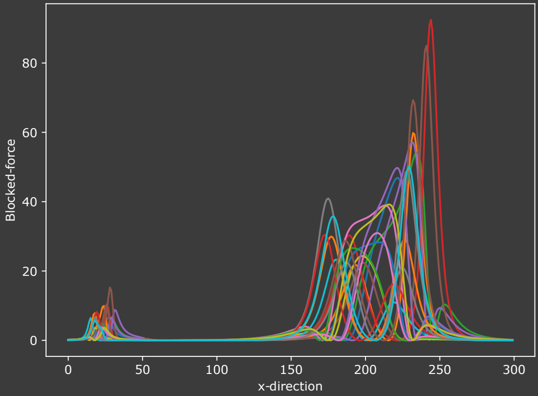
$\mathcal{DGCN}$ - capabilities

- (Image) Regression.
- (Image) Classification.
- Sequential dependent output.



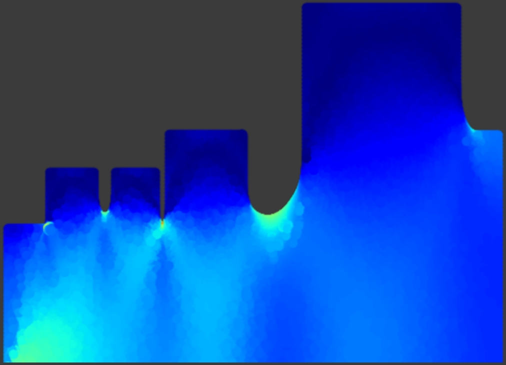
$\mathcal{DGCN}$ - capabilities

- (Image) Regression.
- (Image) Classification.
- Sequential dependent output.
- Multi-output.



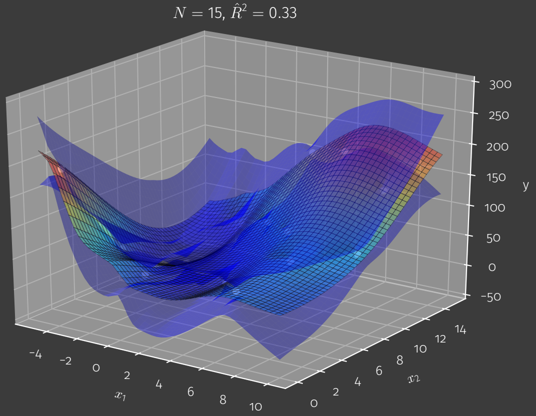
DGCN - capabilities

- (Image) Regression.
- (Image) Classification.
- Sequential dependent output.
- Multi-output.



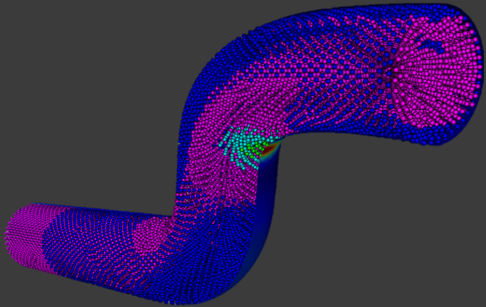
DGCN - capabilities

- (Image) Regression.
- (Image) Classification.
- Sequential dependent output.
- Multi-output.
- Confidence interval of prediction.

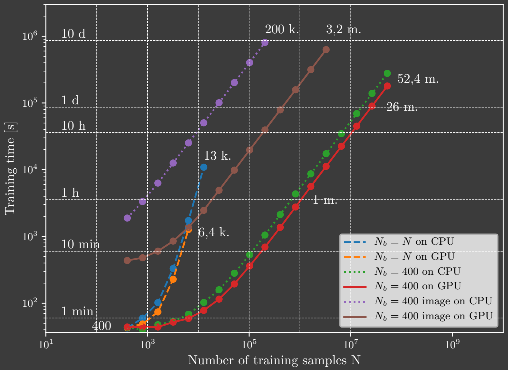


$\mathcal{DGCN}$ - capabilities

- (Image) Regression.
- (Image) Classification.
- Sequential dependent output.
- Multi-output.
- Confidence interval of prediction.
- Gradients of prediction.

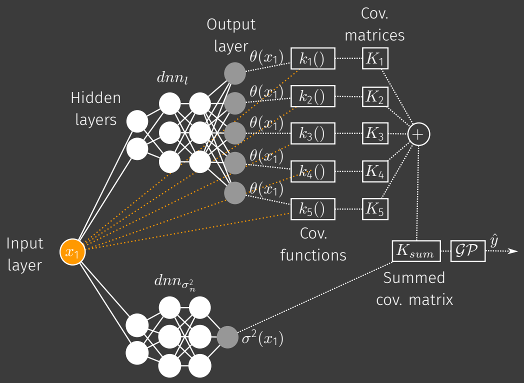


- (Image) Regression.
- (Image) Classification.
- Sequential dependent output.
- Multi-output.
- Confidence interval of prediction.
- Gradients of prediction.
- Big data (image based on a single workstation).

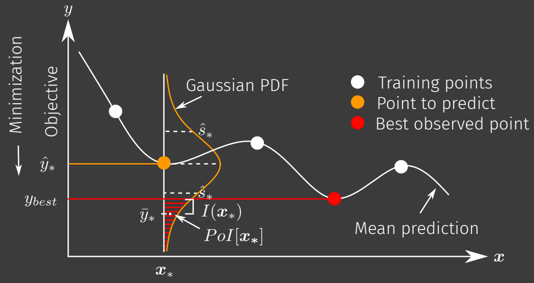


$\mathcal{DGCN}$ - capabilities

- (Image) Regression.
- (Image) Classification.
- Sequential dependent output.
- Multi-output.
- Confidence interval of prediction.
- Gradients of prediction.
- Big data (image based on a single workstation).
- No adjustments to be made by the user.

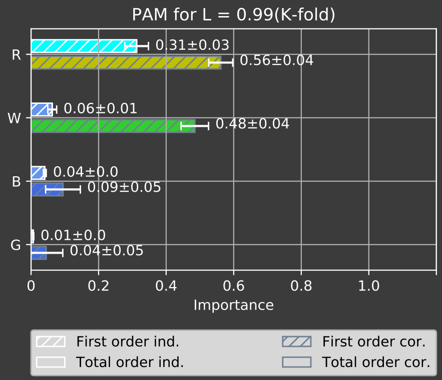


- Bayesian optimization (expected improvement).



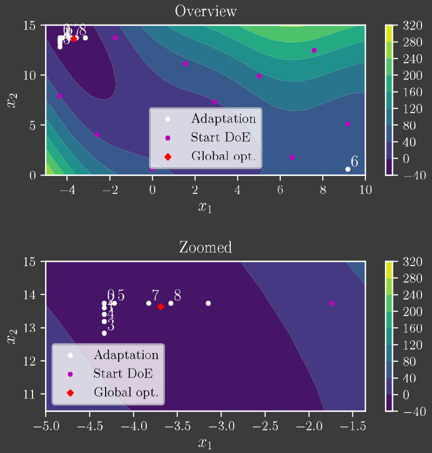
STOCHOS - capabilities

- Bayesian optimization (expected improvement).
- Sensitivity analysis (even for correlated input parameters).



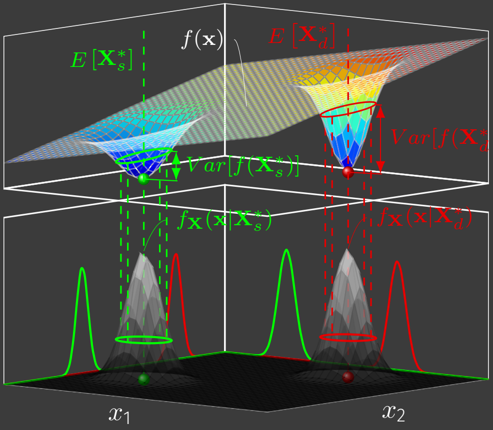
STOCHOS - capabilities

- Bayesian optimization (expected improvement).
- Sensitivity analysis (even for correlated input parameters).
- Adaptive sampling.

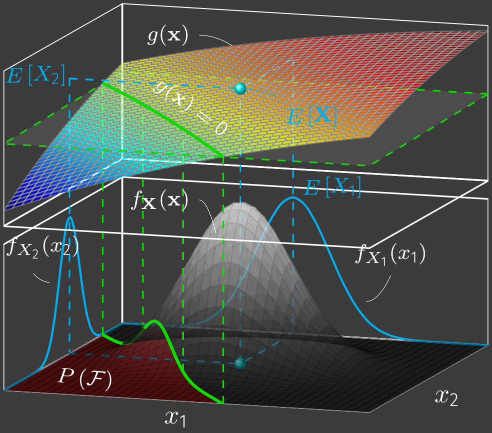


STOCHOS - capabilities

- Bayesian optimization (expected improvement).
- Sensitivity analysis (even for correlated input parameters).
- Adaptive sampling.
- Robust design optimization.

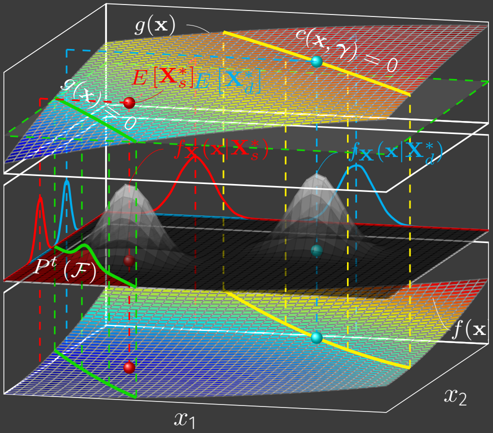


- Bayesian optimization (expected improvement).
- Sensitivity analysis (even for correlated input parameters).
- Adaptive sampling.
- Robust design optimization.
- Reliability analysis.



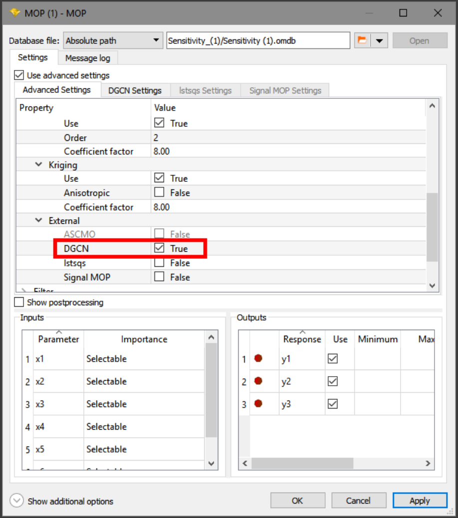
STOCHOS - capabilities

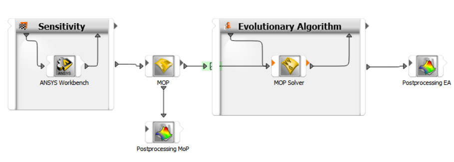
- Bayesian optimization (expected improvement).
- Sensitivity analysis (even for correlated input parameters).
- Adaptive sampling.
- Robust design optimization.
- Reliability analysis.
- Reliability-based robust design optimization.



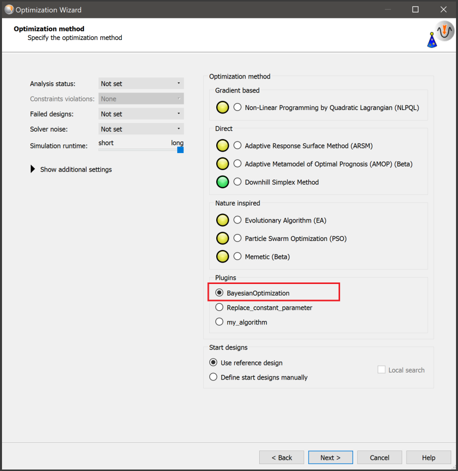
Integration in optiSLang & Outlook

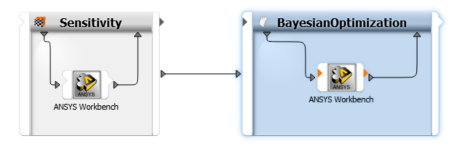
Integration status in optiSLang: MOP





Integration status in optiSLang: optimization





Integration status in optiSLang:

Implemented in optiSLang:

- $\mathcal{DGCN}$ for scalar outputs.
- Bayesian optimization (EI / EVI / Var) for scalar outputs.

Next possible steps within optiSlang:

- Support for field and sequential dependent output.
- RDO-based Bayesian optimization.

Methodological future developments in $\mathcal{STOCHOS}$:

- Optimal control of running processes via reinforcement learning.

Contact

Kevin Cremanns
info@probaligence.com

PROBABILISTIC INTELLIGENCE

PI Probaligence GmbH