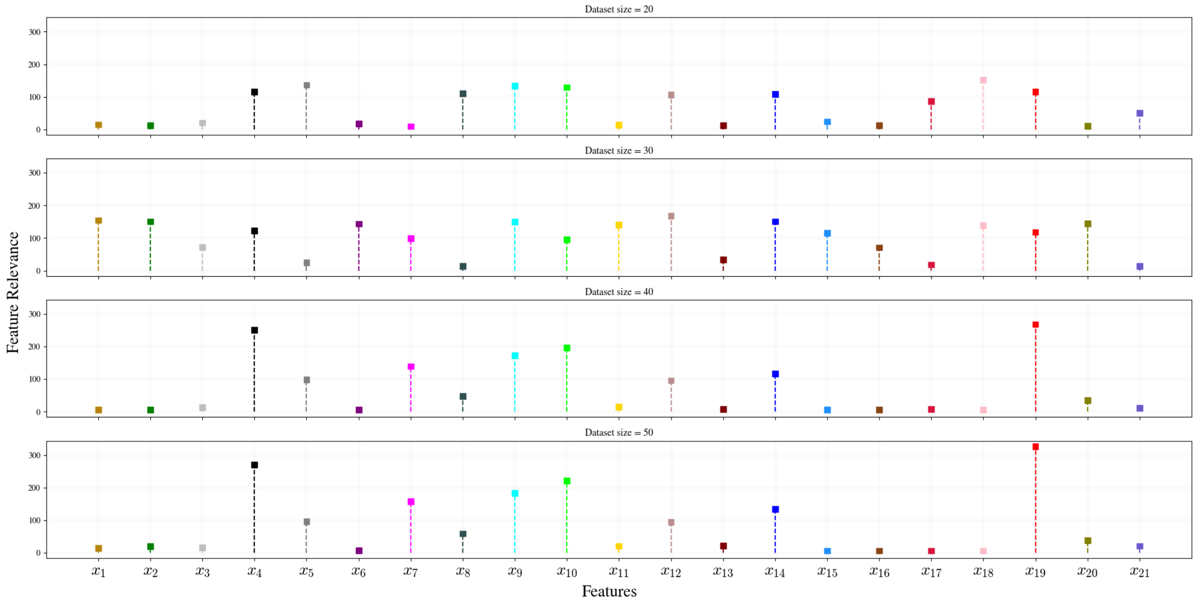
Sensitivity analysis guided Active Learning



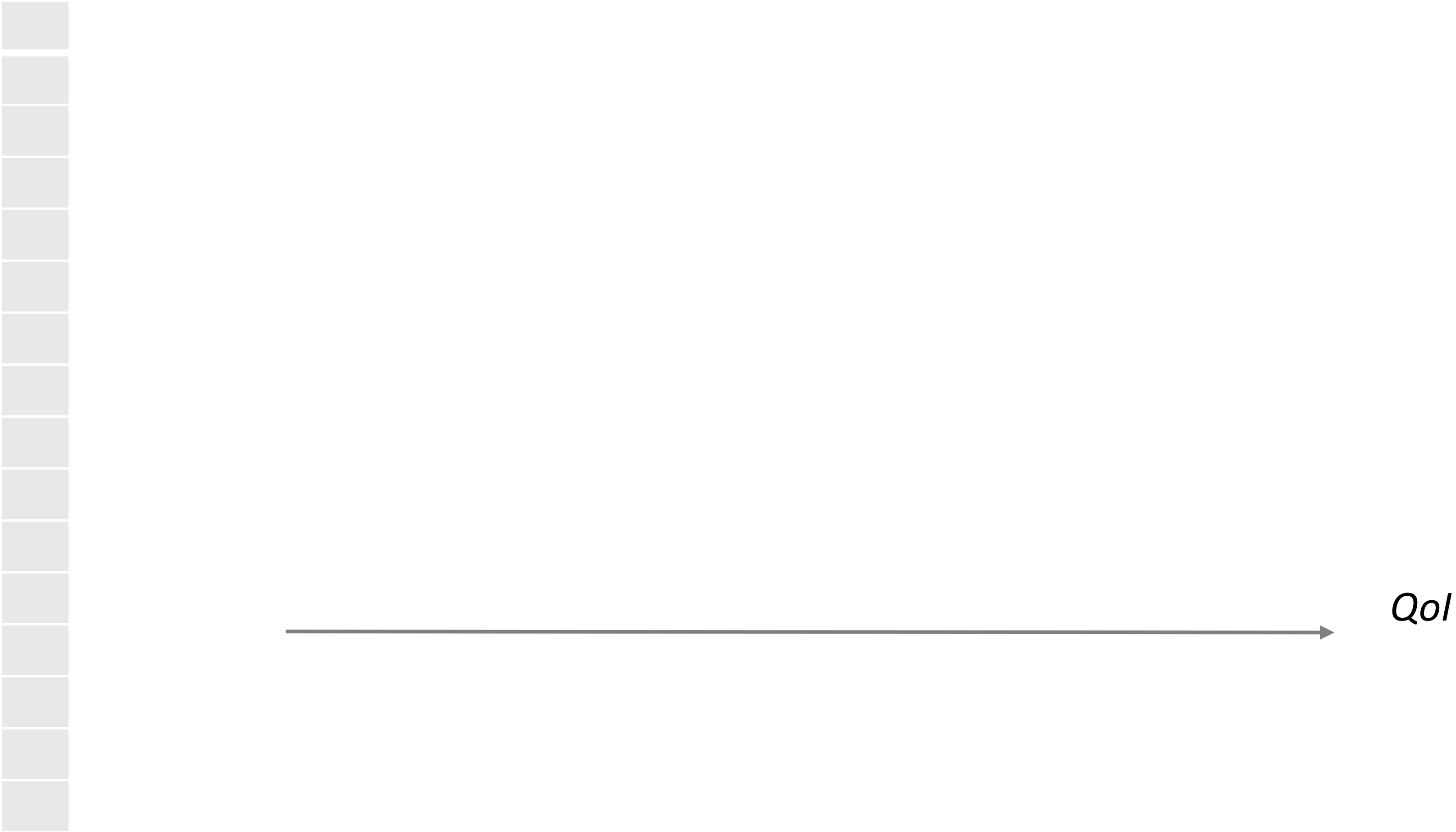
Albert Turon, Anbazhagan Subramani, Aravind Sasikumar, Dai Kandil

What is the problem ?

What is the problem ?

QoI

What is the problem ?



What is the problem ?

E_{11}

E_{22}

G_{12}

G_{23}

X_T

f_{XT}

t

v_{12}

v_{23}

f_{GT}

X_C

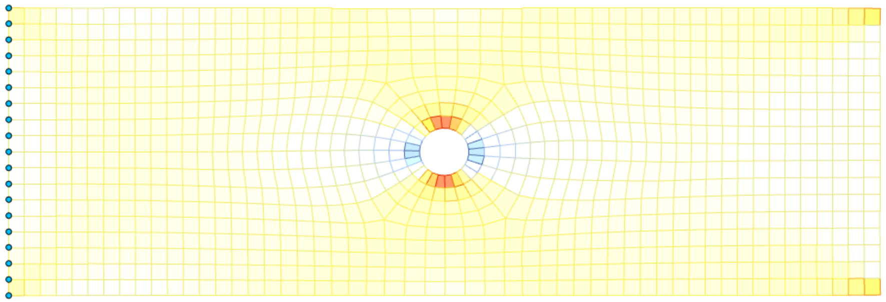
.....

.....

......

f_{XC}

a_0



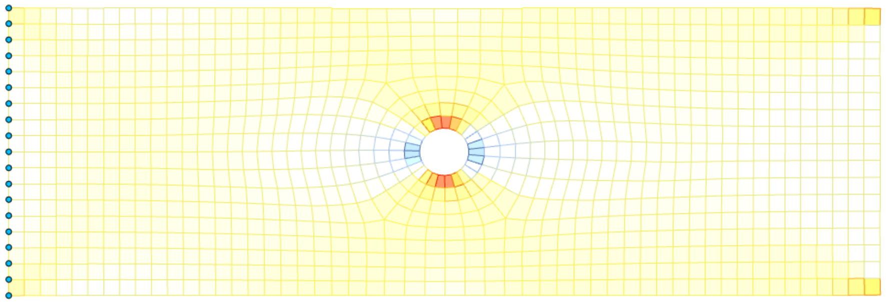
QoI

σ

What is the problem ?

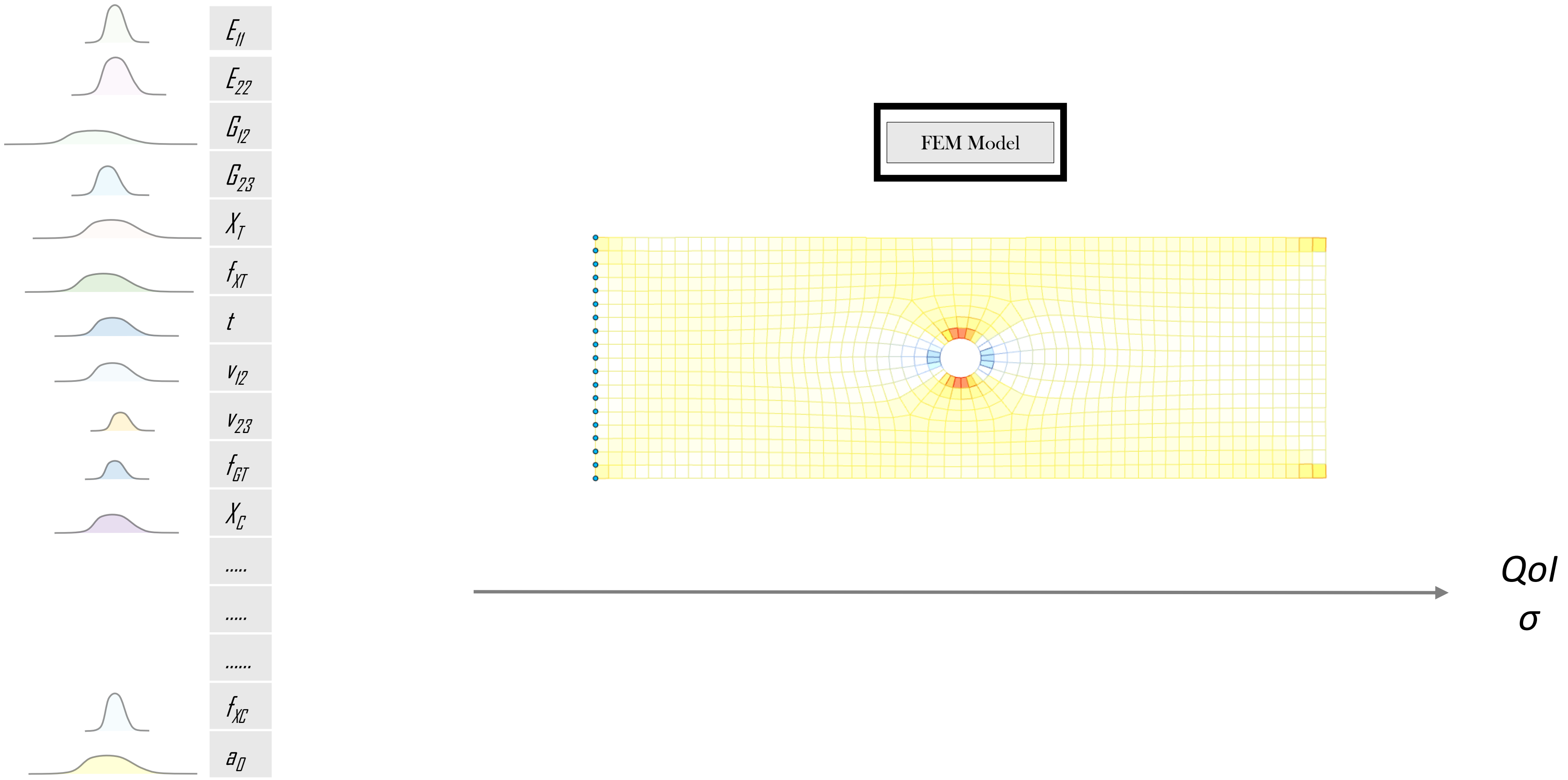
E_{11}
E_{22}
G_{12}
G_{23}
X_T
f_{XT}
t
v_{12}
v_{23}
f_{GT}
X_C
.....
.....
......
f_{XC}
a_0

FEM Model

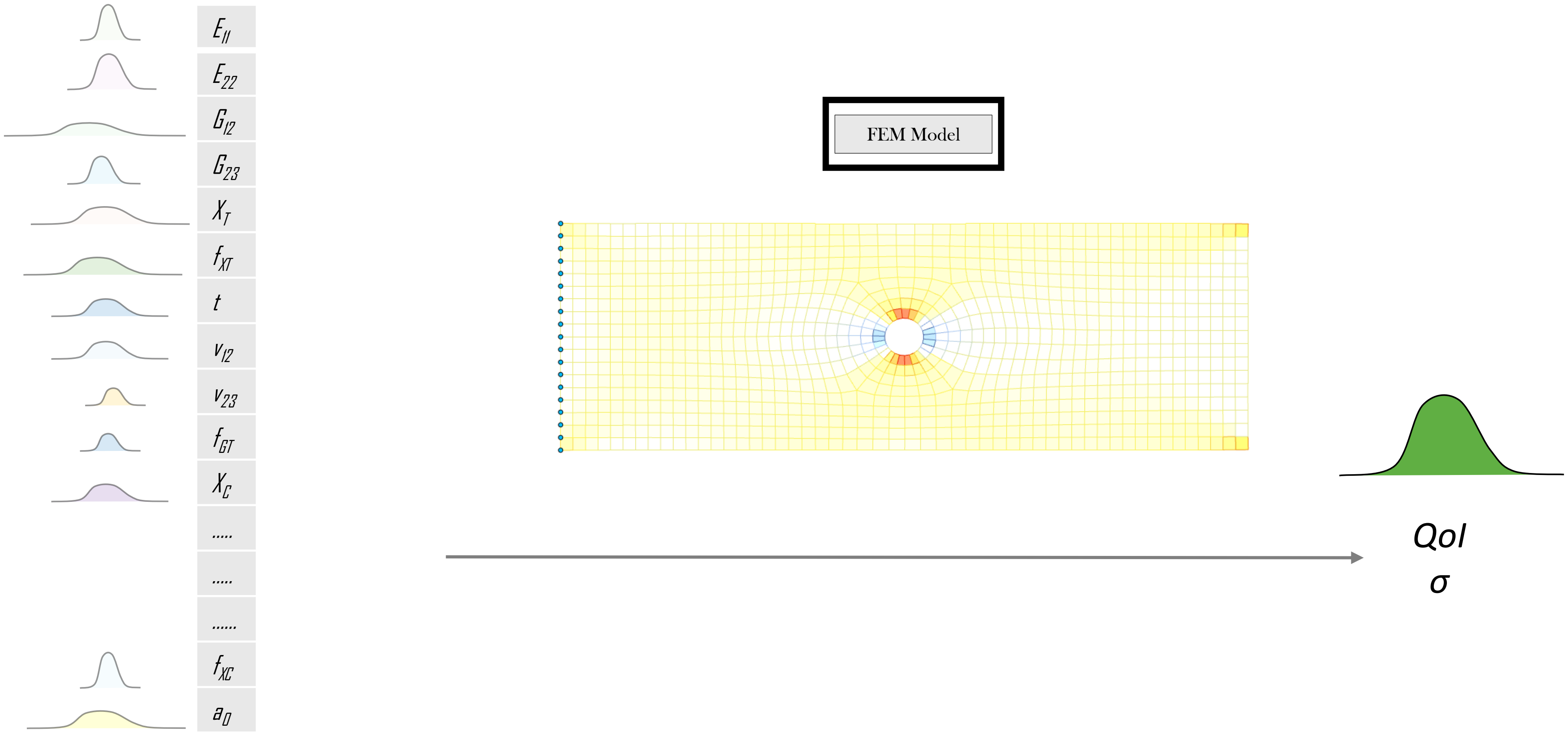


QoI
σ

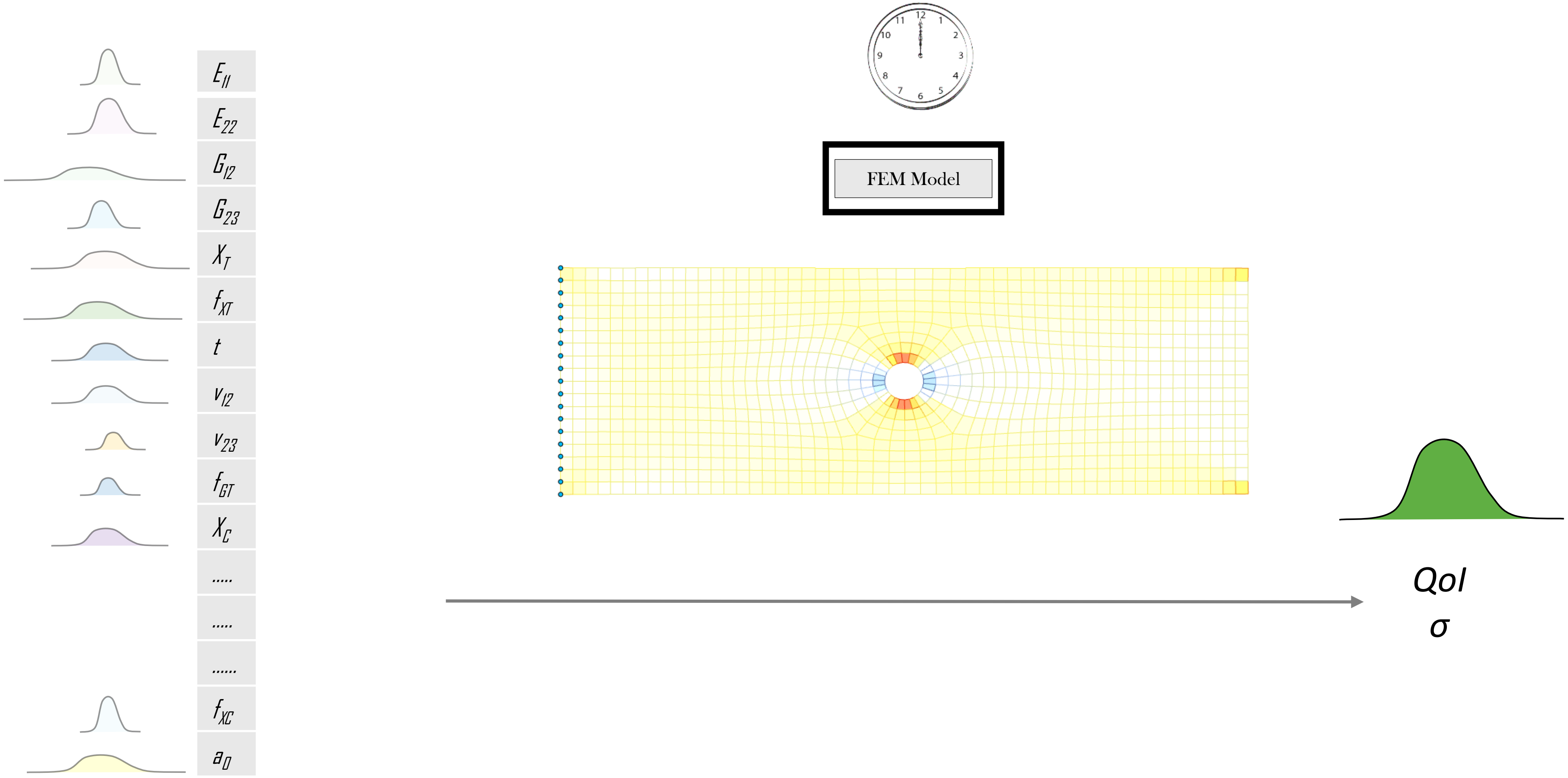
What is the problem ?



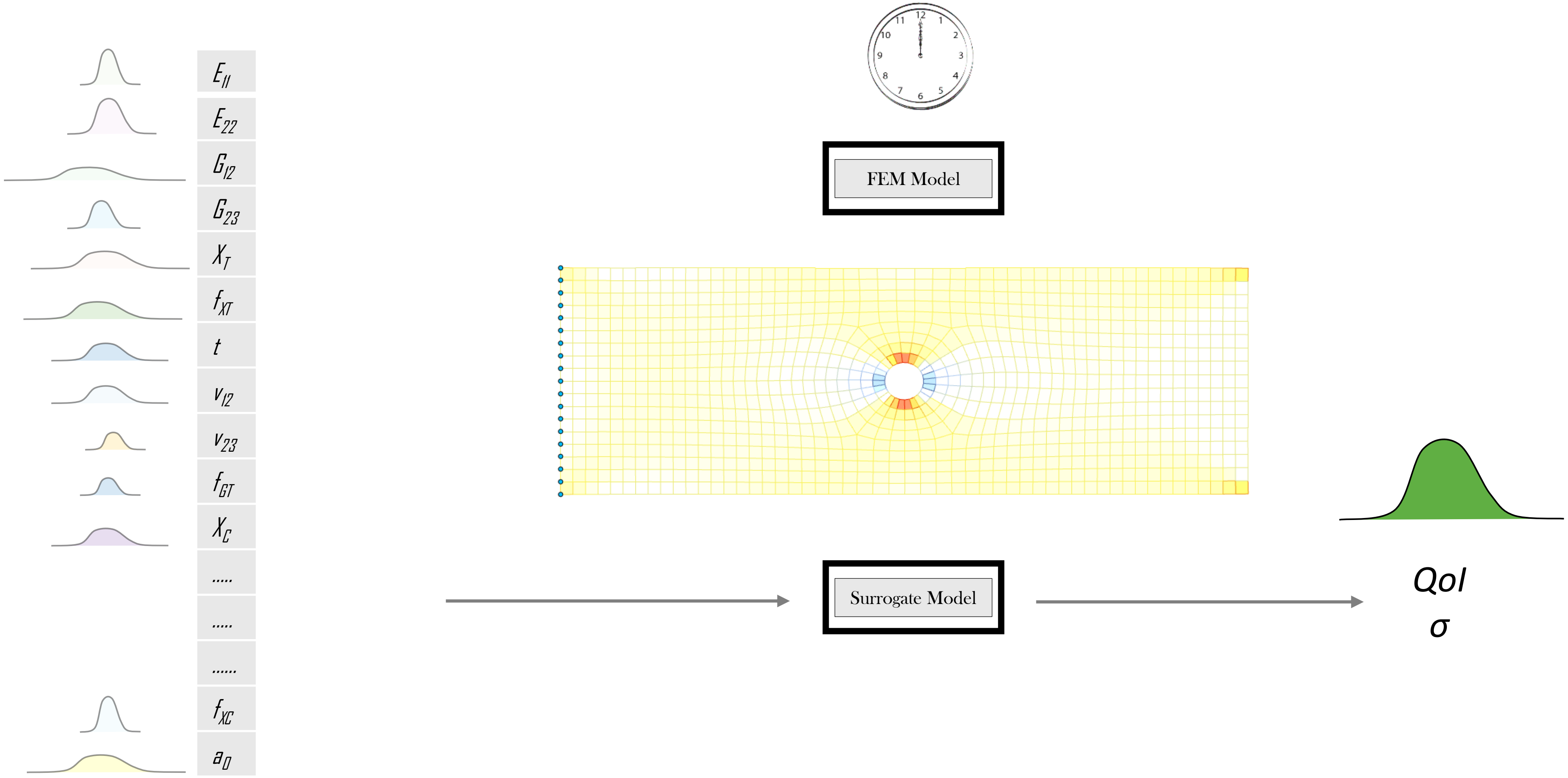
What is the problem ?



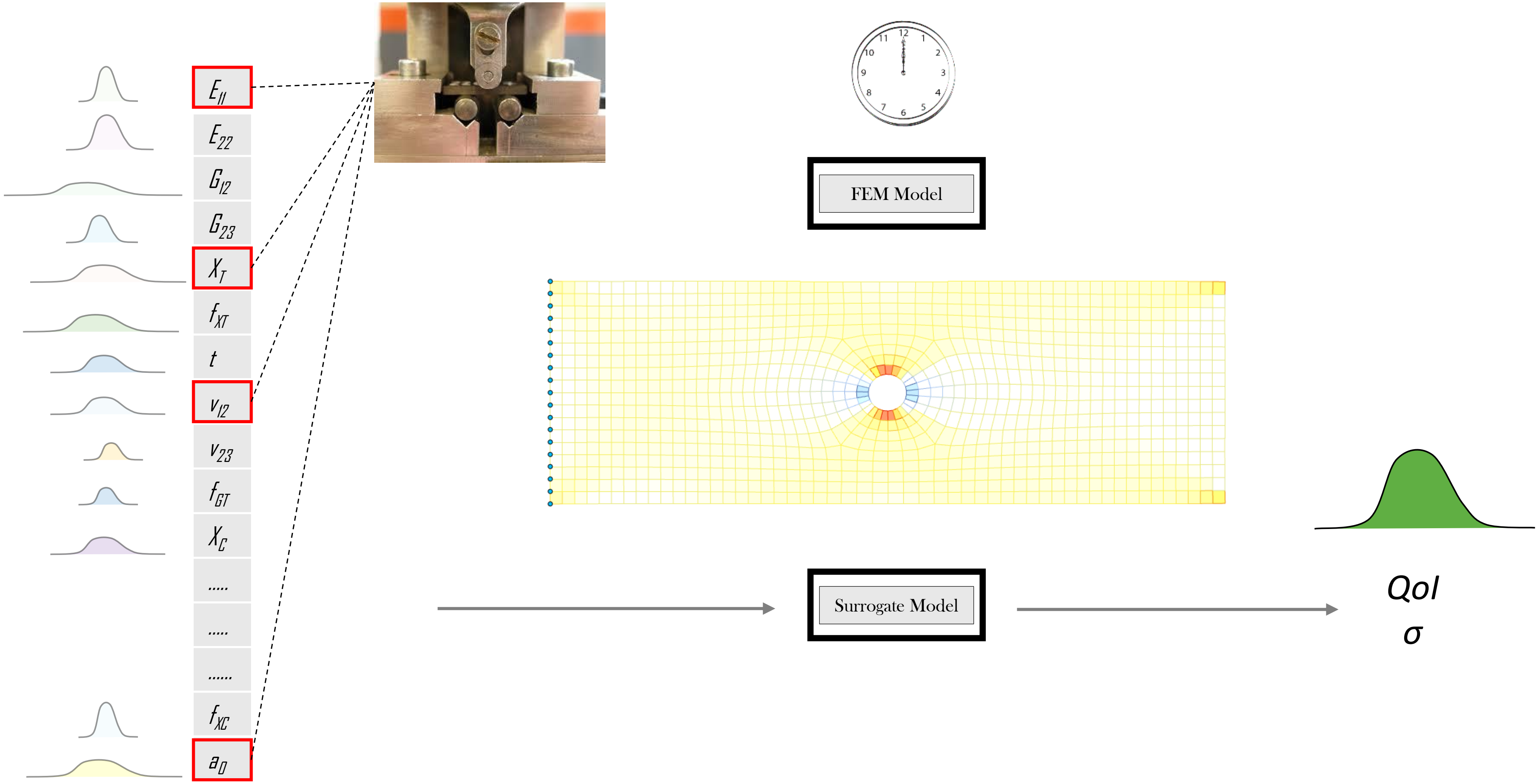
What is the problem ?



What is the problem ?

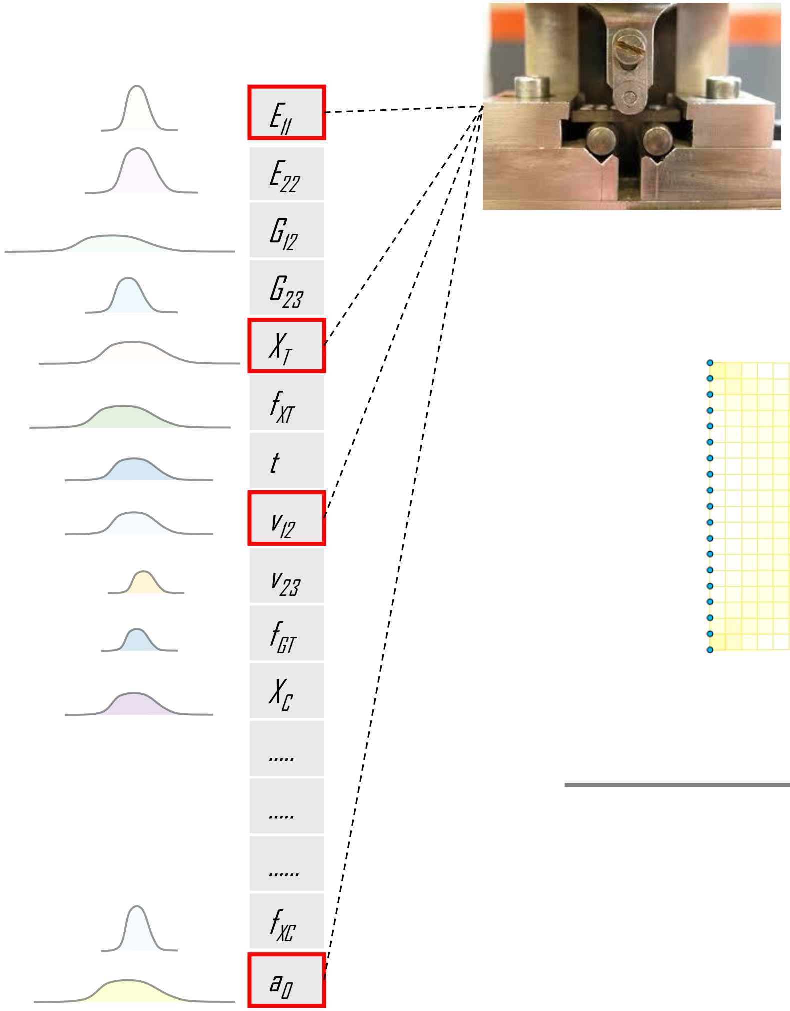


What is the problem ?

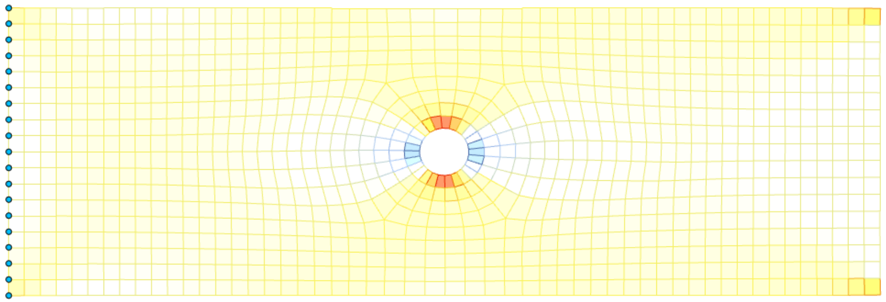


What is the problem ?

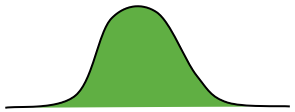
(1) Surrogate Model



FEM Model



Surrogate Model



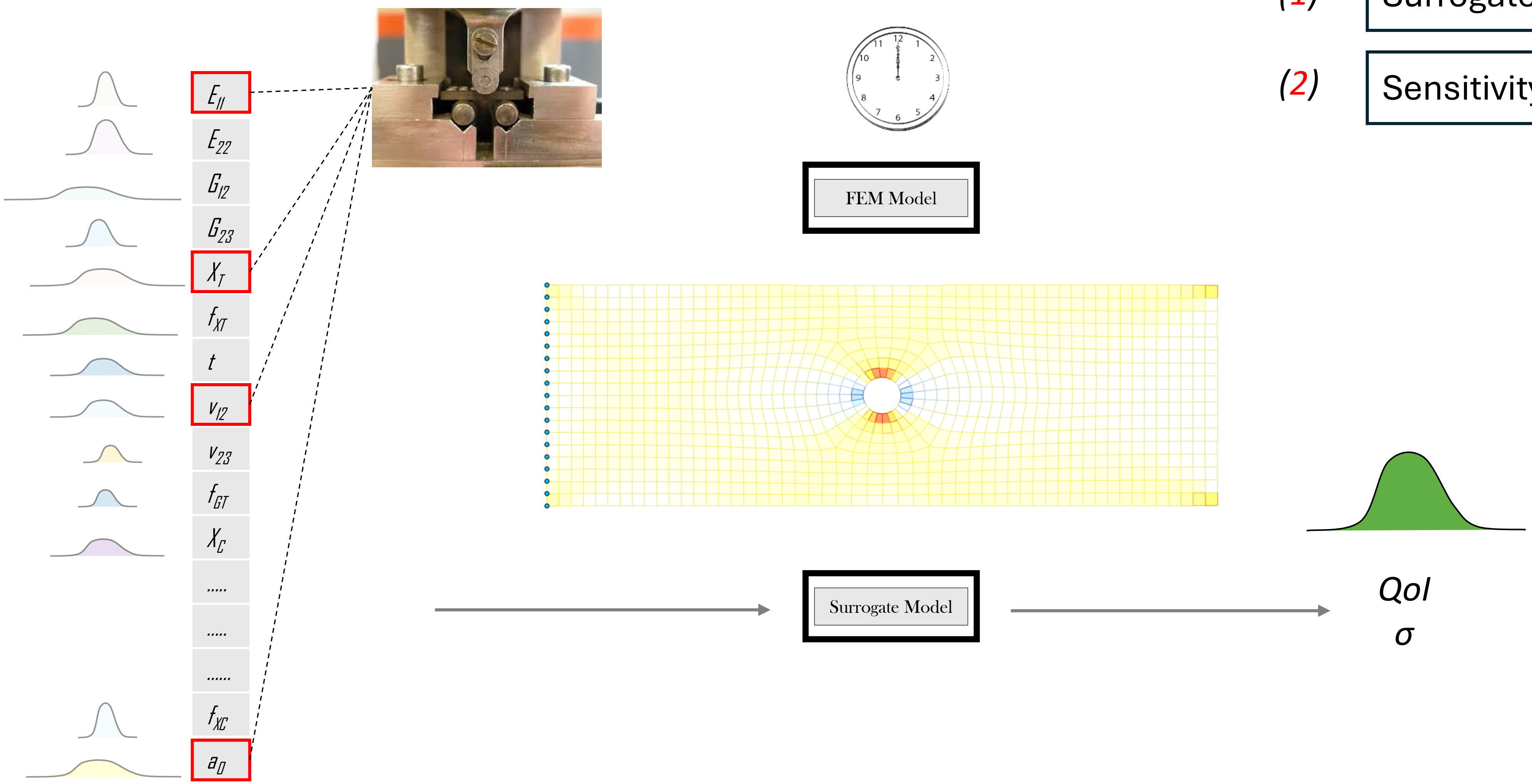
QoI

σ

What is the problem ?

(1) Surrogate Model

(2) Sensitivity Analysis

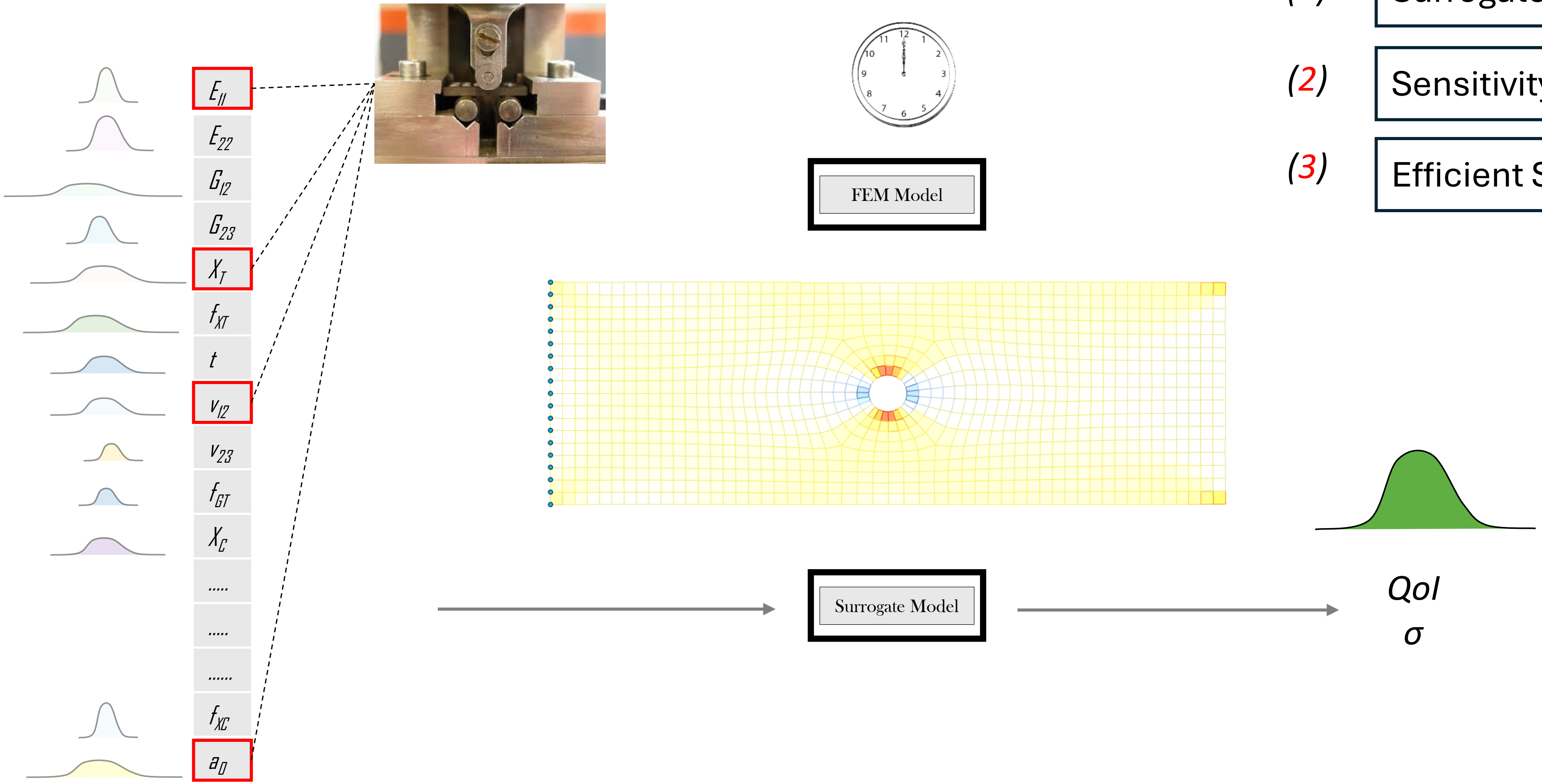


What is the problem ?

(1) Surrogate Model

(2) Sensitivity Analysis

(3) Efficient Sampling



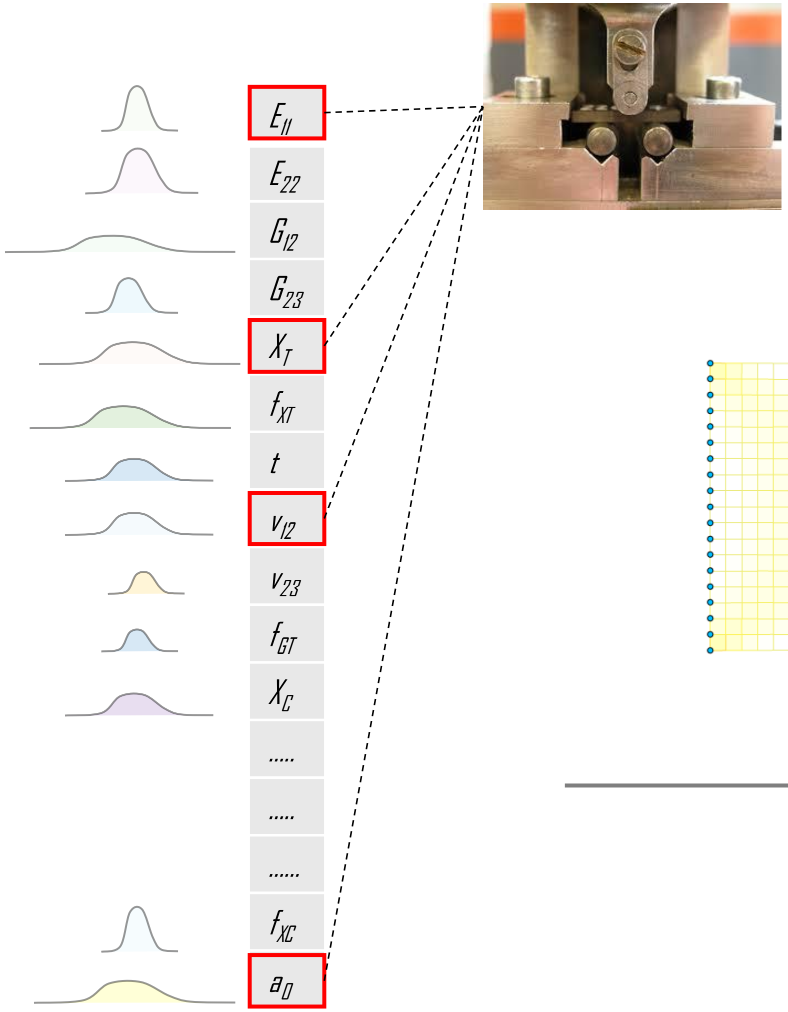
What is the problem ?

Collecting data to train the surrogate and figure out the sensitive inputs

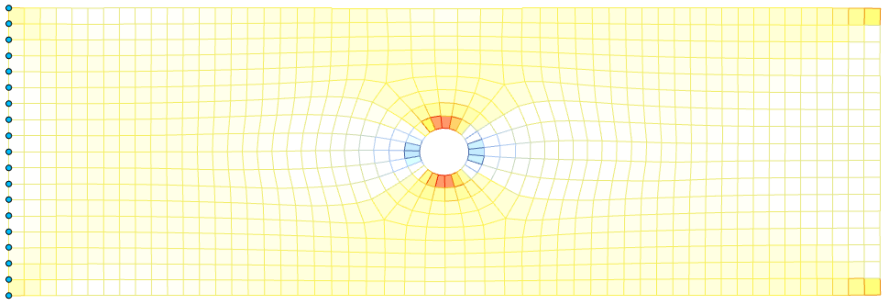
(1) Surrogate Model

(2) Sensitivity Analysis

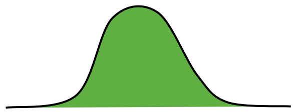
(3) Efficient Sampling



FEM Model



Surrogate Model



QoI

σ

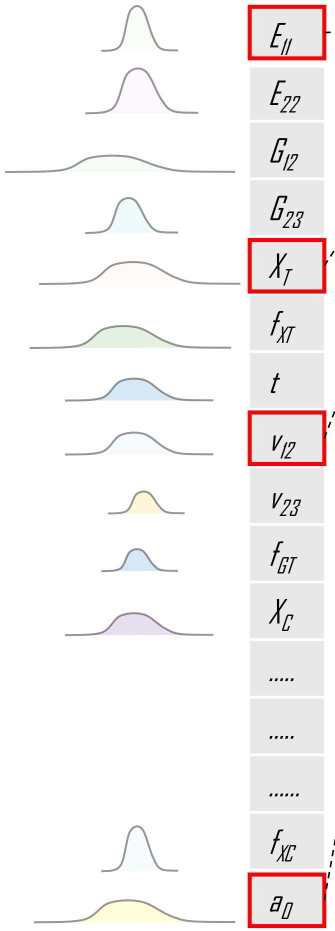
What is the problem ?

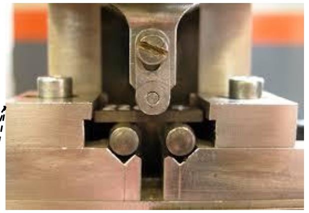
Collecting data to train the surrogate and figure out the sensitive inputs

(1) Surrogate Model

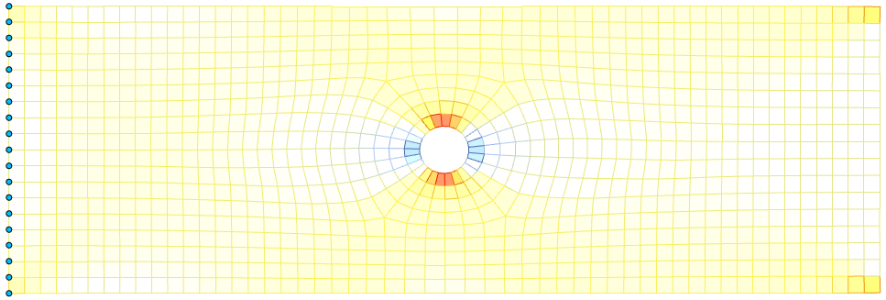
(2) Sensitivity Analysis

(3) Efficient Sampling

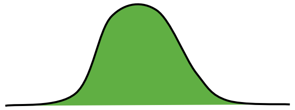




FEM Model



Surrogate Model



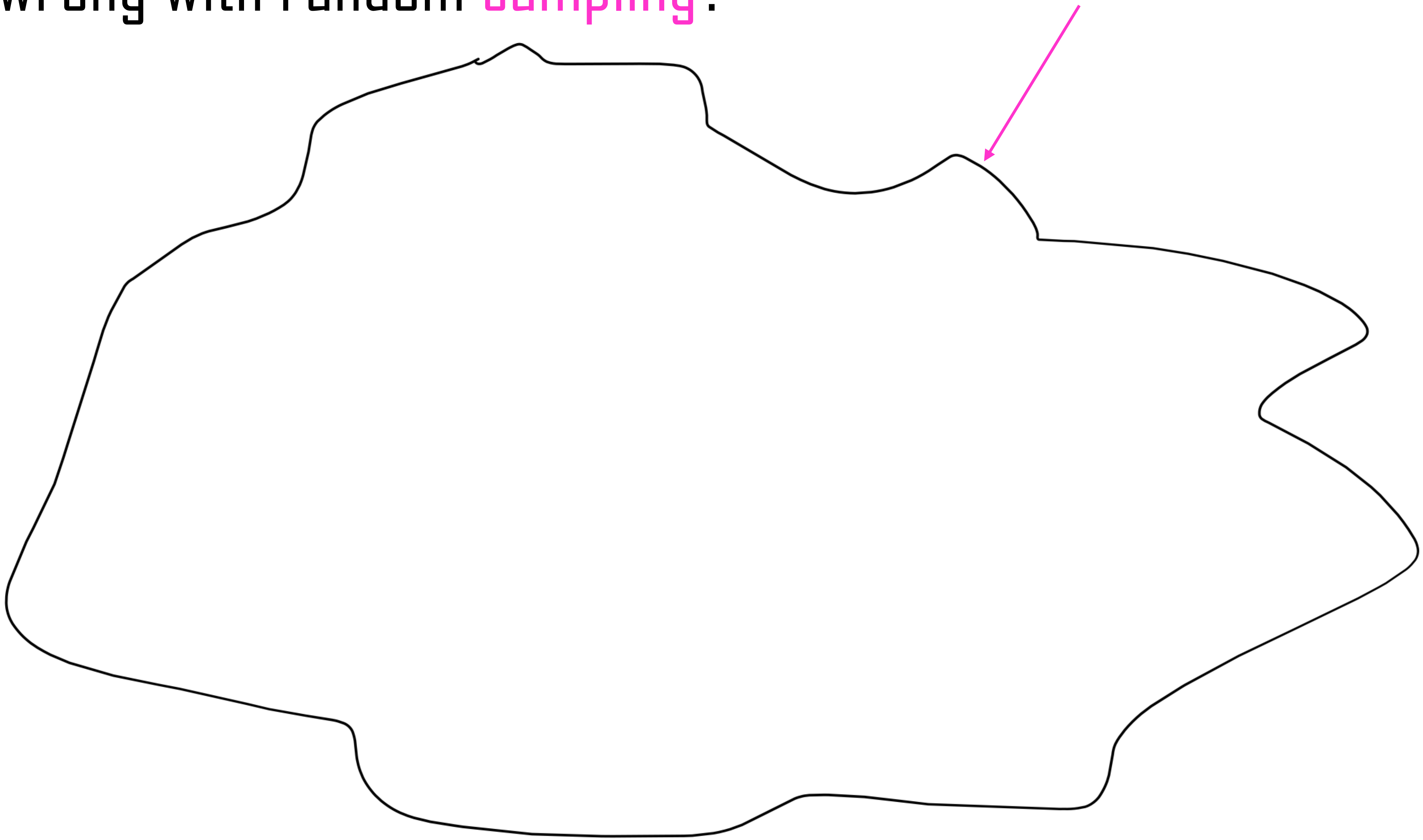
QoI

σ

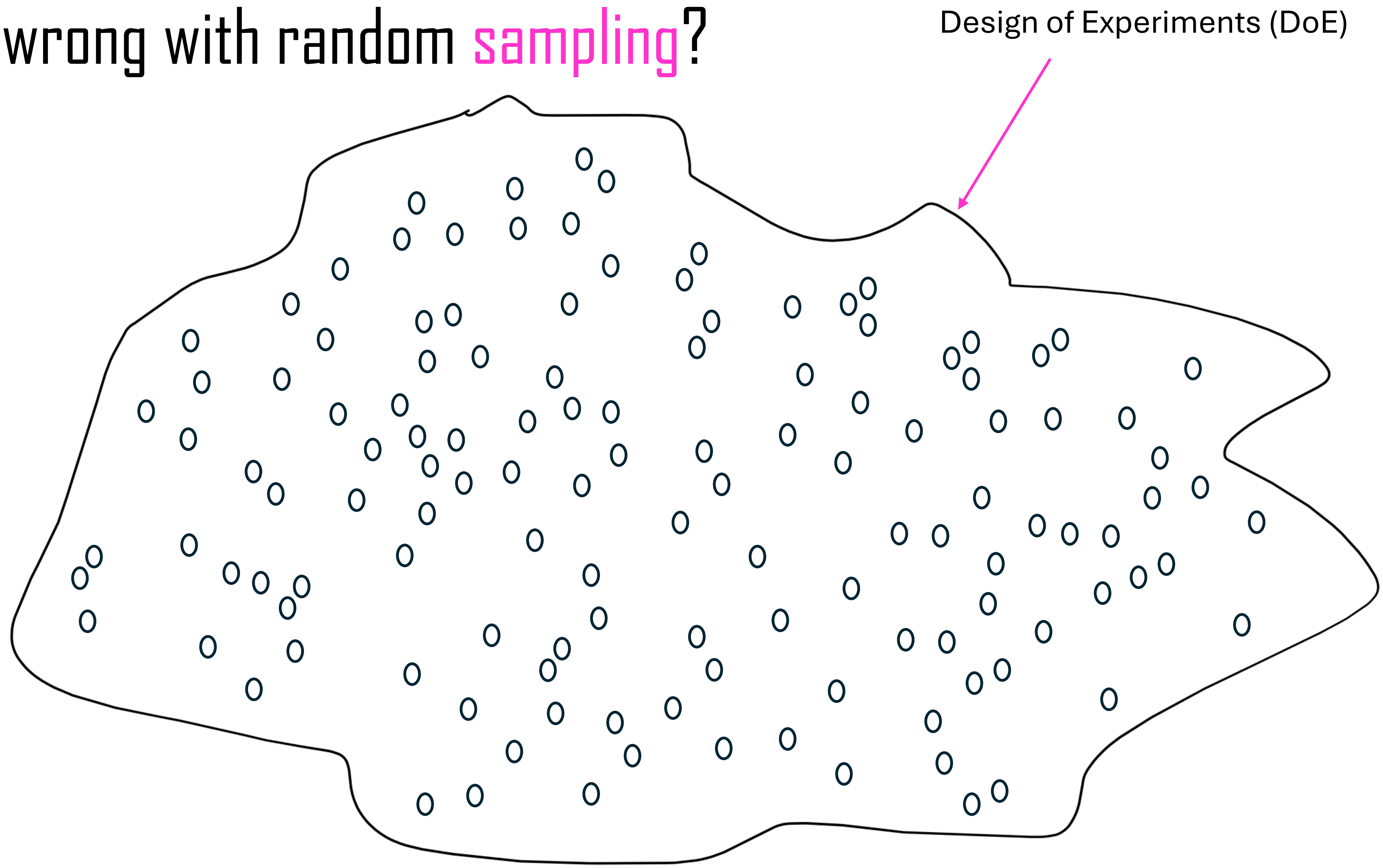
Approaches

Sampling

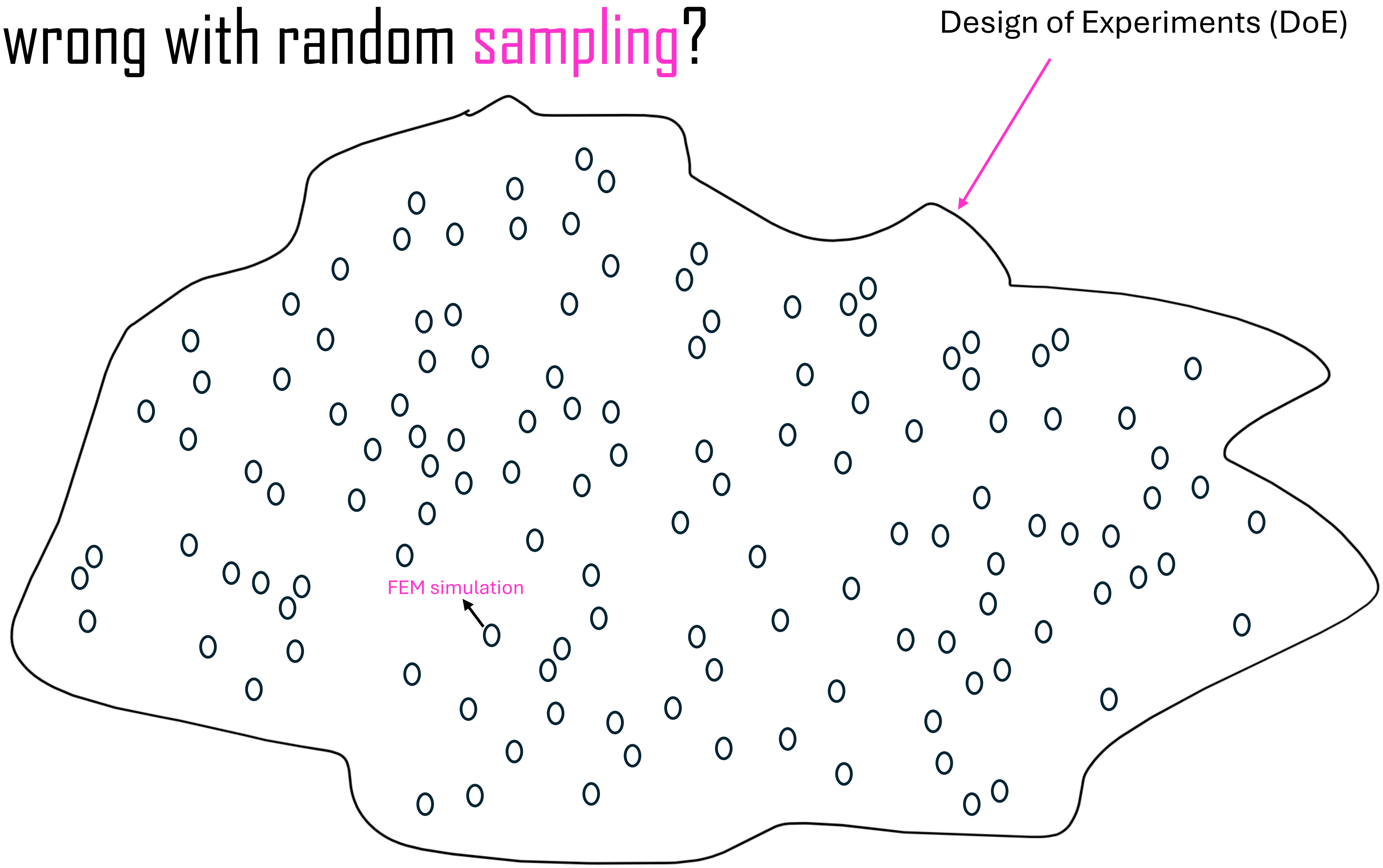
What is wrong with random sampling?



What is wrong with random sampling?

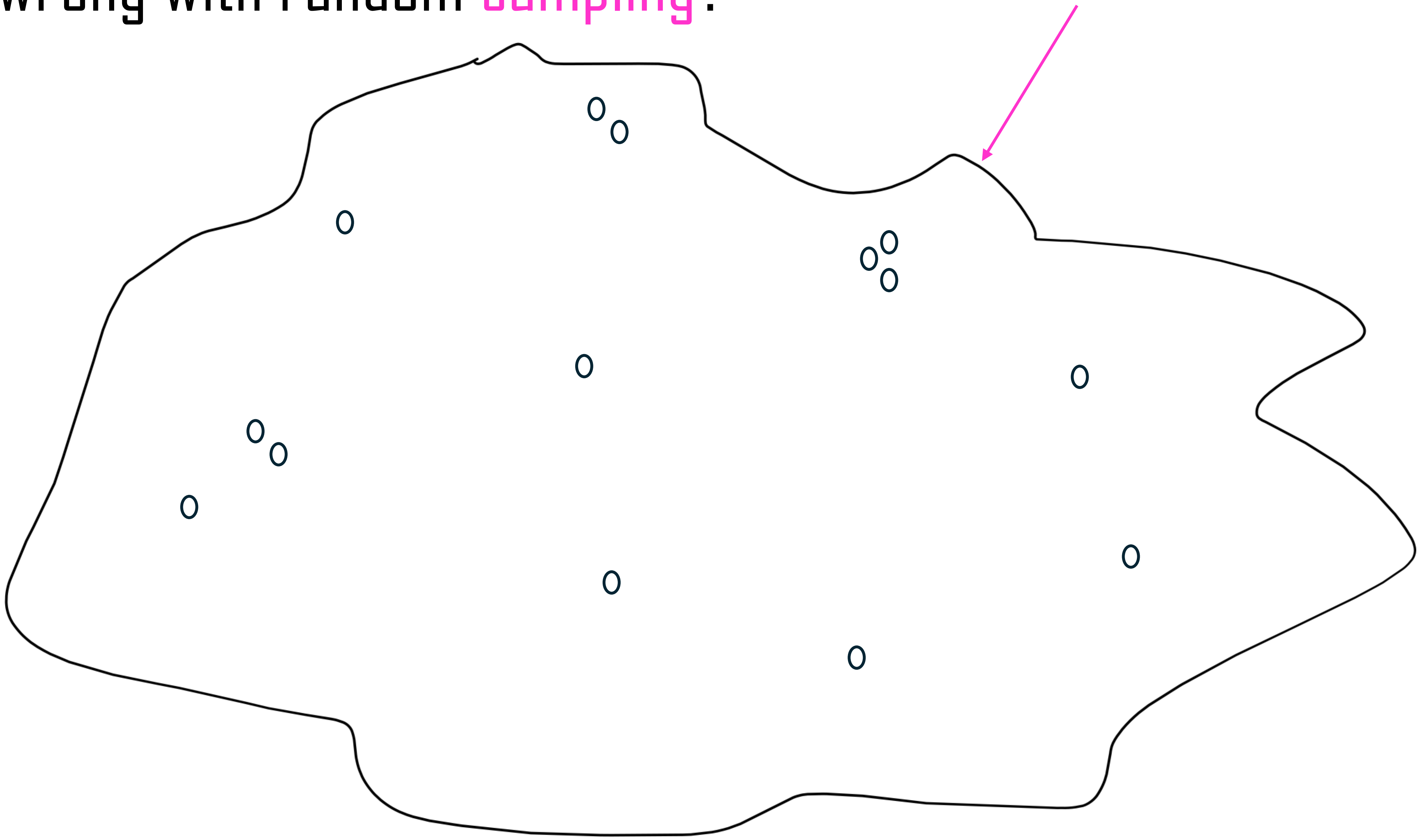


What is wrong with random sampling?

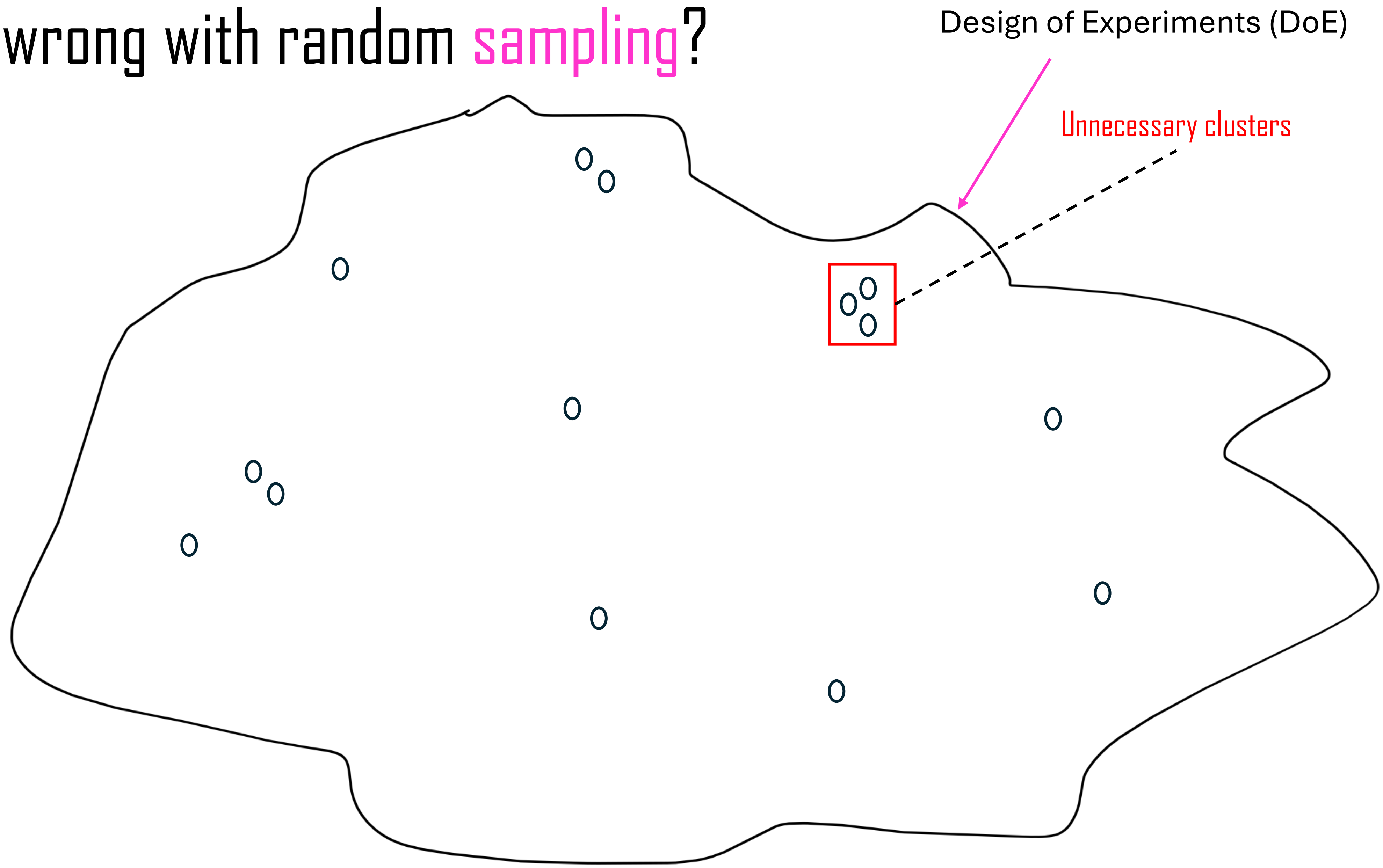


What is wrong with random sampling?

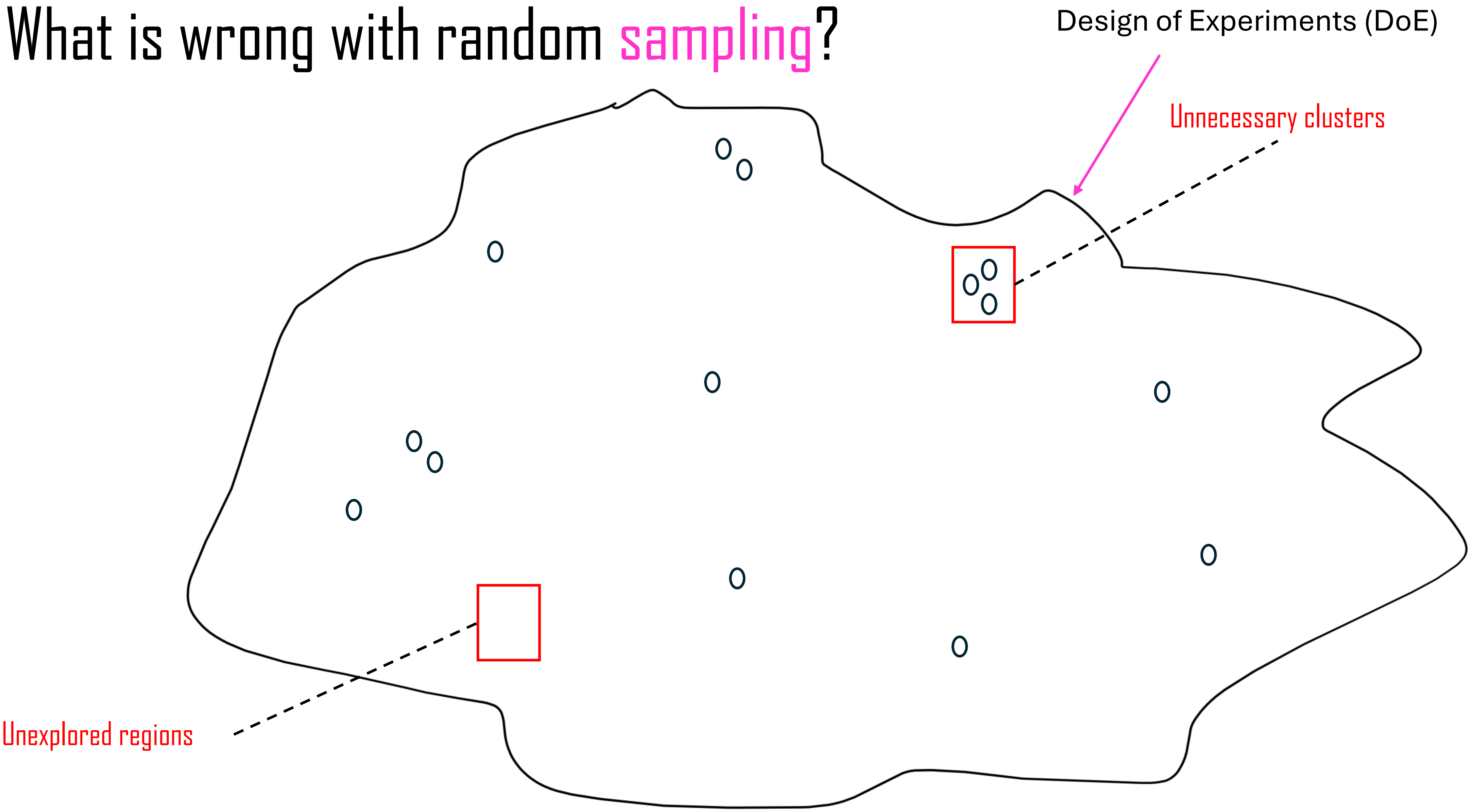
Design of Experiments (DoE)



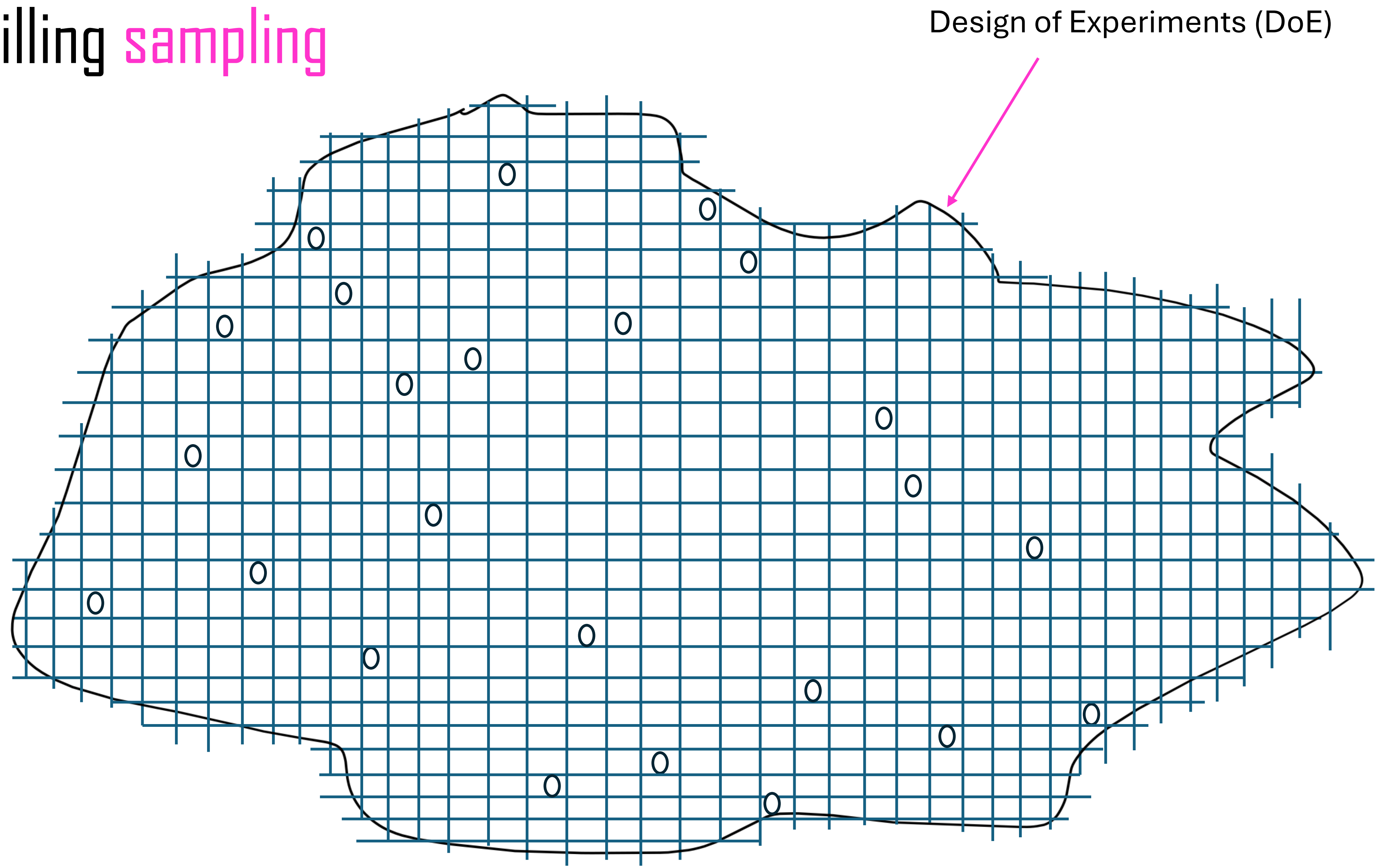
What is wrong with random sampling?



What is wrong with random sampling?



Space Filling sampling



Sensitivity Analysis

Morris Sensitivity Analysis

One-factor-at-a-time

FEM model

random trajectories

- E_{11}
- E_{22}
- G_{12}
- G_{23}
- X_T
- f_{XT}
- t
- v_{12}
- v_{23}
- f_{GT}
- X_C
- .....
- .....
- ......
- f_{XC}
- a_0

$$EE_i = \frac{f(x_1, x_2, ..., x_i + \Delta_i, ..., x_n) - f(x_1, x_2, ..., x_i, ..., x_n)}{\Delta_i}$$

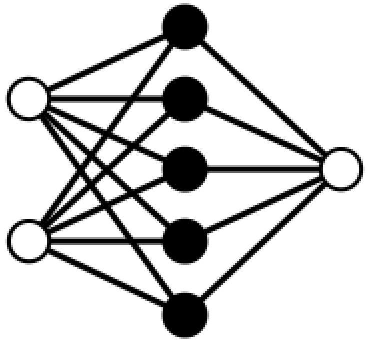
$$\mu_k^\star = \frac{1}{r} \sum_{i=1}^{r} |EE_k^r|$$

E_{11}	E_{22}	G_{12}	G_{23}	X_T	f_{XT}	t	v_{12}	v_{23}	f_{GT}	X_C	....	....	....	f_{XC}	a_0
$X_{E11} + \Delta_i$	X_{E22}	X_{G12}	X_{G23}	X_{XT}	X_{fXT}	X_t	X_{v12}	X_{v23}	X_{fGT}	X_{XC}	....	....	....	X_{fXC}	X_{a0}
X_{E11}	$X_{E22}+\Delta_j$	X_{G12}	X_{G23}	X_{XT}	X_{fXT}	X_t	X_{v12}	X_{v23}	X_{fGT}	X_{XC}	....	....	....	X_{fXC}	X_{a0}
X_{E11}	X_{E22}	$X_{G12}+\Delta_k$	X_{G23}	X_{XT}	X_{fXT}	X_t	X_{v12}	X_{v23}	X_{fGT}	X_{XC}	....	....	....	X_{fXC}	X_{a0}
X_{E11}	X_{E22}	X_{G12}	$X_{G23}+\Delta_l$	X_{XT}	X_{fXT}	X_t	X_{v12}	X_{v23}	X_{fGT}	X_{XC}	....	....	....	X_{fXC}	X_{a0}
X_{E11}	X_{E22}	X_{G12}	X_{G23}	$X_{XT}+\Delta_m$	X_{fXT}	X_t	X_{v12}	X_{v23}	X_{fGT}	X_{XC}	....	....	....	X_{fXC}	X_{a0}
X_{E11}	X_{E22}	X_{G12}	X_{G23}	X_{XT}	$X_{fXT} + \Delta_n$	X_t	X_{v12}	X_{v23}	X_{fGT}	X_{XC}	....	....	....	X_{fXC}	X_{a0}
X_{E11}	X_{E22}	X_{G12}	X_{G23}	X_{XT}	X_{fXT}	X_t	X_{v12}	X_{v23}	X_{fGT}	X_{XC}	....	....	....	X_{fXC}	X_{a0}

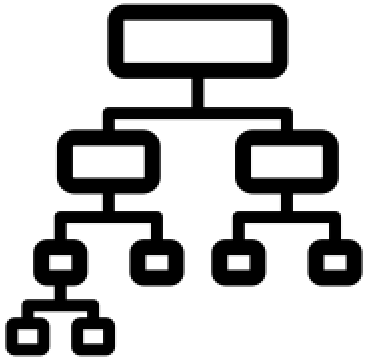
Factorial Sampling Plans for Preliminary Computational Experiments
Max D. Morris

Surrogate Models

What about Surrogates

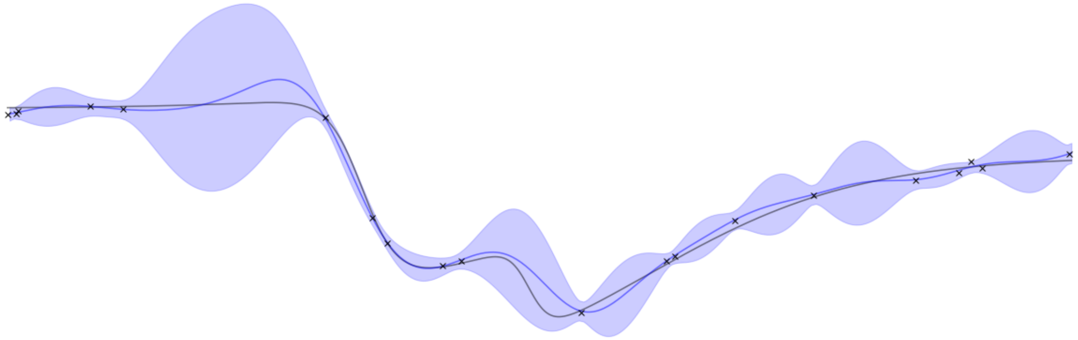


Neural Networks



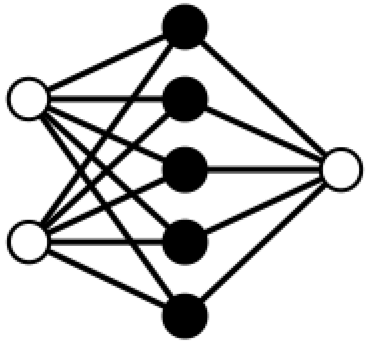
Decision trees

Linear Regression

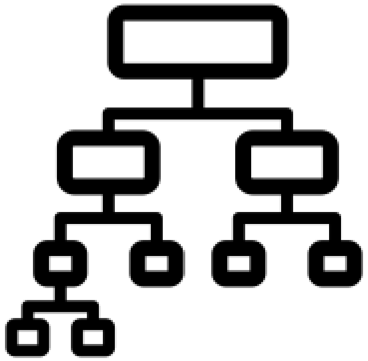


Gaussian Processes

What about Surrogates

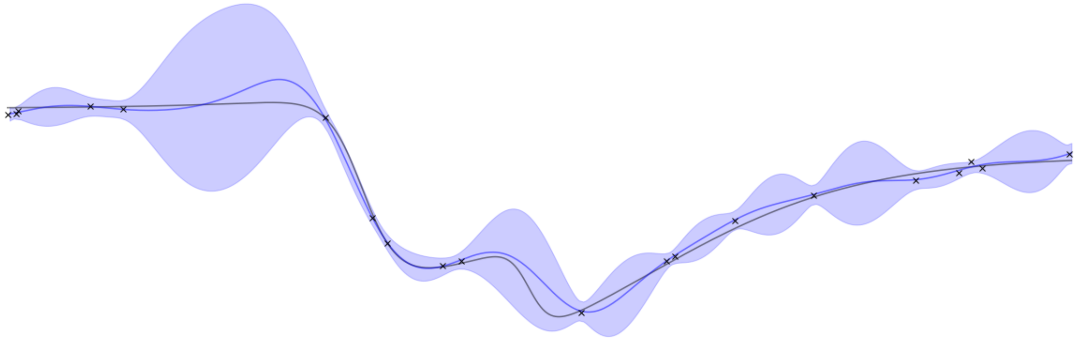


Neural Networks

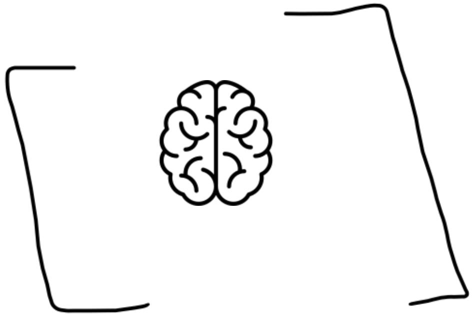


Decision trees

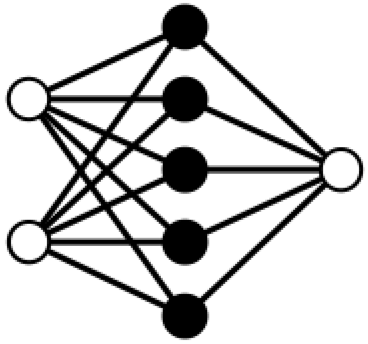
Linear Regression



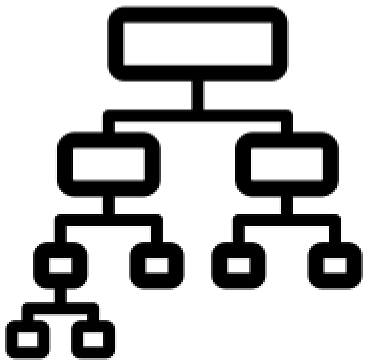
Gaussian Processes



What about Surrogates

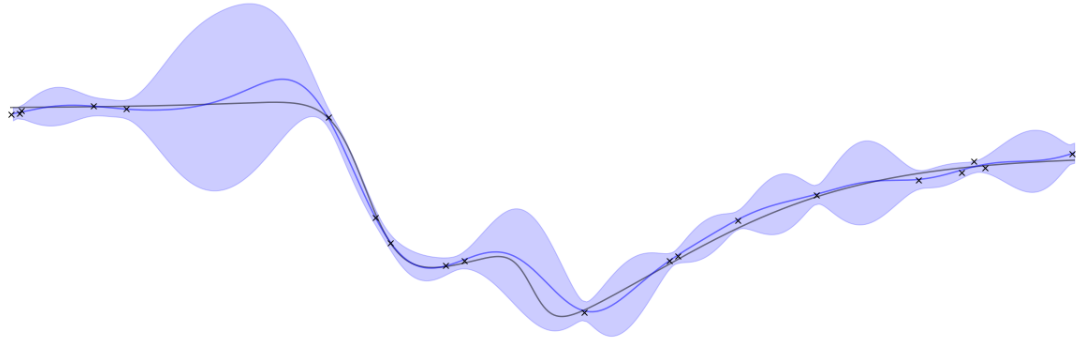


Neural Networks

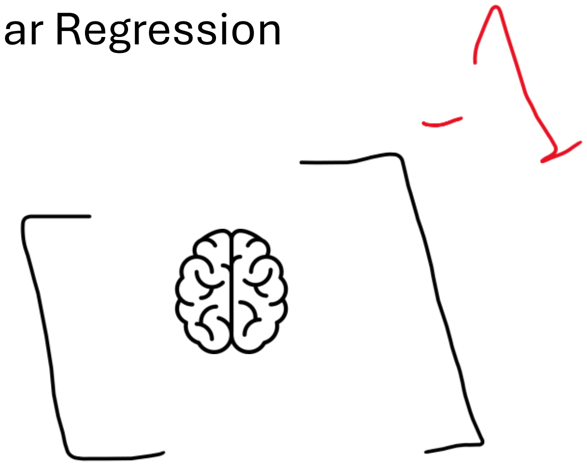


Decision trees

Linear Regression



Gaussian Processes



Powerful Techniques

Surrogate helps Sensitivity analysis

Projection Predictive

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	...	...	X_n

Full model (All parameters)

Projection Predictive → Best Technique Ever...

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	...	...	x_n

Full model (All parameters)

PROJECTION PREDICTIVE MODEL SELECTION FOR GAUSSIAN PROCESSES
Juho Piironen, Aki Vehtari

Projection Predictive → Best Technique Ever...

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	...	...	X_n

Full model (All parameters)

↓ Train a GP

θ Gold Standard

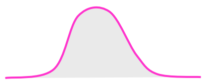
PROJECTION PREDICTIVE MODEL SELECTION FOR GAUSSIAN PROCESSES
Juho Piironen, Aki Vehtari

Projection Predictive → Best Technique Ever…

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	…	…	x_n

Full model (All parameters)

Train a GP



θ Gold Standard

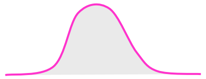
PROJECTION PREDICTIVE MODEL SELECTION FOR GAUSSIAN PROCESSES
Juho Piironen, Aki Vehtari

Projection Predictive ⟶ Best Technique Ever…

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	…	…	X_n

Full model (All parameters)

↓ Train a GP



θ Gold Standard —Project→ θ projected

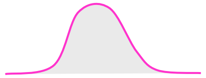
PROJECTION PREDICTIVE MODEL SELECTION FOR GAUSSIAN PROCESSES
Juho Piironen, Aki Vehtari

Projection Predictive → Best Technique Ever...

X_1 | X_2 | X_3 | X_4 | X_5 | X_6 | X_7 | X_8 | X_9 | X_{10} | ... | ... | X_n

Full model (All parameters)

Train a GP



θ Gold Standard

Project

θ projected

Reduced model (parameters sub-set)

X_2 | X_3 | X_4 | X_7 | X_8 | X_9

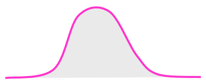
PROJECTION PREDICTIVE MODEL SELECTION FOR GAUSSIAN PROCESSES
Juho Piironen, Aki Vehtari

Projection Predictive → Best Technique Ever...

x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_8 x_9 x_{10} ... ... x_n

Full model (All parameters)

↓ Train a GP



θ Gold Standard —Project→ θ projected

↑

Reduced model (parameters sub-set)

x_2 x_3 x_4 x_7 x_8 x_9

Not only for sensitivity analysis,

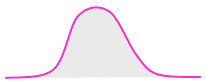
PROJECTION PREDICTIVE MODEL SELECTION FOR GAUSSIAN PROCESSES
Juho Piironen, Aki Vehtari

Projection Predictive → Best Technique Ever…

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	…	…	x_n

Full model (All parameters)

Train a GP



θ Gold Standard —Project→ **θ projected**

x_2	x_3	x_4	x_7	x_8	x_9

Reduced model (parameters sub-set)

<u>Not only for sensitivity analysis</u>, but to train the model for available inputs at training time but absent or expensive to collect at production phase of the model.

PROJECTION PREDICTIVE MODEL SELECTION FOR GAUSSIAN PROCESSES
Juho Piironen, Aki Vehtari

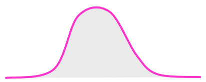
Projection Predictive → Best Technique Ever...

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	...	...	x_n

x_2	x_3	x_4	x_7	x_8	x_9

Full model (All parameters)

↓ Train a GP



θ Gold Standard → Project → **θ projected**

↑

Reduced model (parameters sub-set)

<u>Not only for sensitivity analysis</u>, but to train the model for available inputs at training time but absent or expensive to collect at production phase of the model.

But how many possible combinations of reduced sets?

PROJECTION PREDICTIVE MODEL SELECTION FOR GAUSSIAN PROCESSES
Juho Piironen, Aki Vehtari

Projection Predictive → Best Technique Ever...

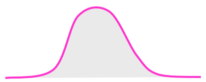
X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	...	...	X_n

X_2	X_3	X_4	X_7	X_8	X_9

Full model (All parameters)

Reduced model (parameters sub-set)

Train a GP



θ Gold Standard → Project → **θ** projected

Out of **30** parameters, there are **593,775** **6** different possible combinations....

Not only for sensitivity analysis, but to train the model for available inputs at training time but absent or expensive to collect at production phase of the model.

But how many possible combinations of reduced sets?

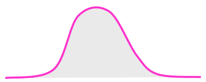
PROJECTION PREDICTIVE MODEL SELECTION FOR GAUSSIAN PROCESSES
Juho Piironen, Aki Vehtari

Projection Predictive → Best Technique Ever...

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	...	...	X_n

Full model (All parameters)

Train a GP



θ Gold Standard

Project

θ projected

X_2	X_3	X_4	X_7	X_8	X_9

Reduced model (parameters sub-set)

Heuristic Search

Out of **30** parameters, there are **593,775** **6** different possible combinations....

Not only for sensitivity analysis, but to train the model for available inputs at training time but absent or expensive to collect at production phase of the model.

But how many possible combinations of reduced sets?

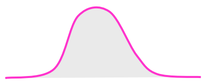
PROJECTION PREDICTIVE MODEL SELECTION FOR GAUSSIAN PROCESSES
Juho Piironen, Aki Vehtari

Projection Predictive → ~~Best Technique Ever...~~

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	...	...	X_n

Full model (All parameters)

↓ Train a GP



θ Gold Standard → Project → θ projected

↑

X_2	X_3	X_4	X_7	X_8	X_9

Reduced model (parameters sub-set)

Heuristic Search

Out of **30** parameters, there are **593,775** **6** different possible combinations....

Not only for sensitivity analysis, but to train the model for available inputs at training time but absent or expensive to collect at production phase of the model.

But how many possible combinations of reduced sets?

PROJECTION PREDICTIVE MODEL SELECTION FOR GAUSSIAN PROCESSES
Juho Piironen, Aki Vehtari

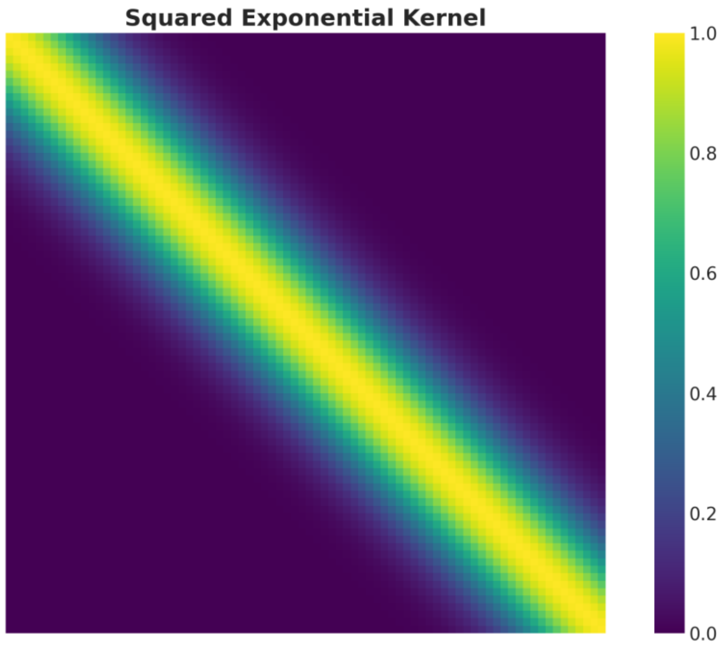
Automatic Relevance Determination (ARD)

Use the Kernel used to train the GP

$$k_{se-ard}(\mathbf{x}_i, \mathbf{x}_j') = \sigma_f^2 exp\Big(-\frac{1}{2}\sum_{d=1}^{D} \frac{(x_{i,d} - x_{j,d})^2}{l_d^2} \Big)$$

data

latent variables



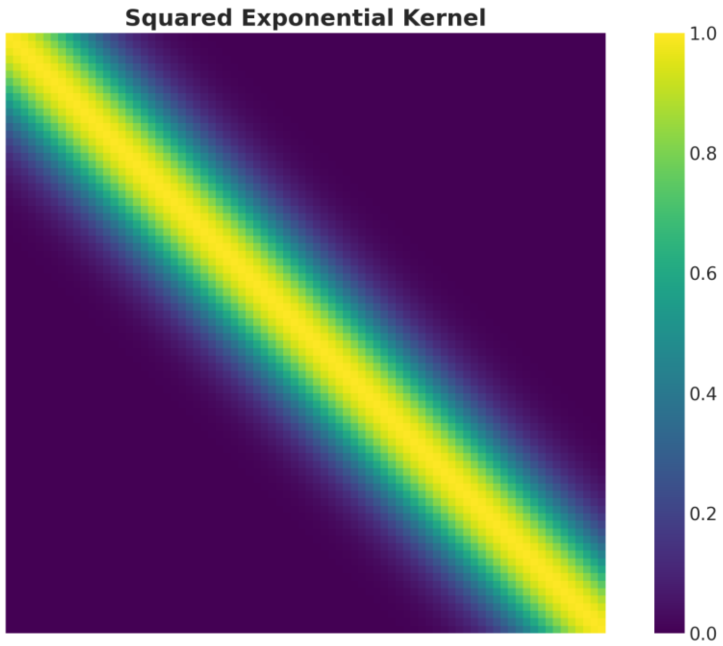
Automatic Relevance Determination (ARD)

Use the Kernel used to train the GP

$$k_{se-ard}(\mathbf{x}_i, \mathbf{x}_j') = \sigma_f^2 exp\Big(-\frac{1}{2}\sum_{d=1}^{D}\frac{(x_{i,d} - x_{j,d})^2}{l_d^2}\Big)$$

data

latent variables



Very computationally cheap

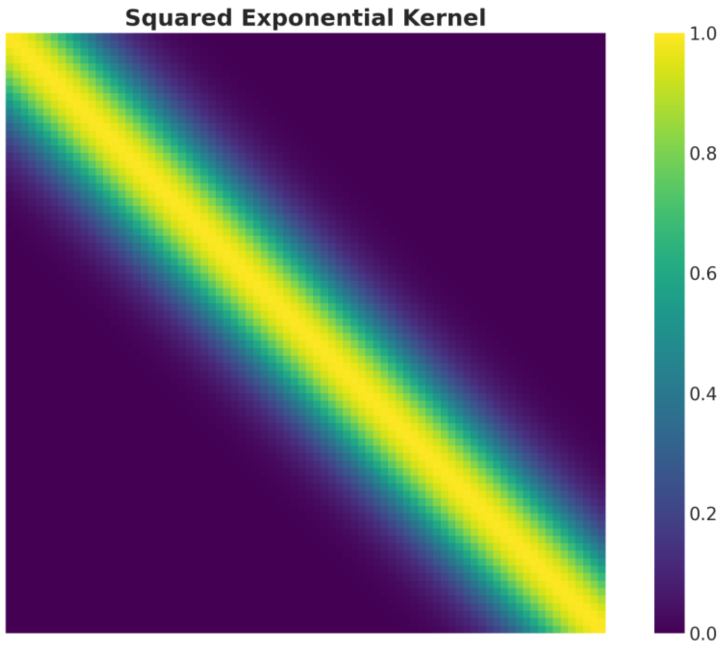
Automatic **R**elevance **D**etermination (ARD)

Use the Kernel used to train the GP

$$k_{se-ard}(\mathbf{x}_i, \mathbf{x}_j') = \sigma_f^2 exp\Big(-\frac{1}{2}\sum_{d=1}^{D} \frac{(x_{i,d} - x_{j,d})^2}{l_d^2} \Big)$$

data

latent variables



Very computationally cheap

It is usually used

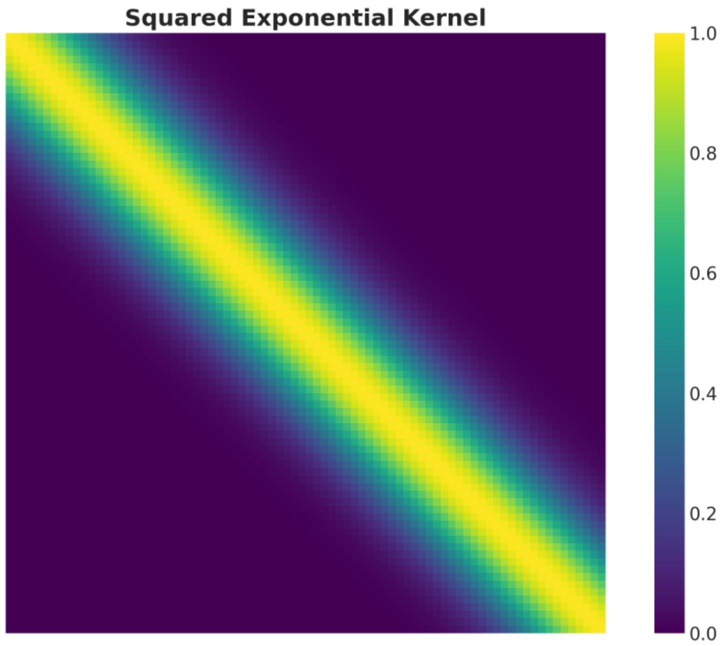
Automatic Relevance Determination (ARD)

Use the Kernel used to train the GP

$$k_{se-ard}(\mathbf{x}_i, \mathbf{x}_j') = \sigma_f^2 exp\left(-\frac{1}{2}\sum_{d=1}^{D}\frac{(x_{i,d} - x_{j,d})^2}{l_d^2}\right)$$

data

latent variables



Very computationally cheap

It is usually used

It is working

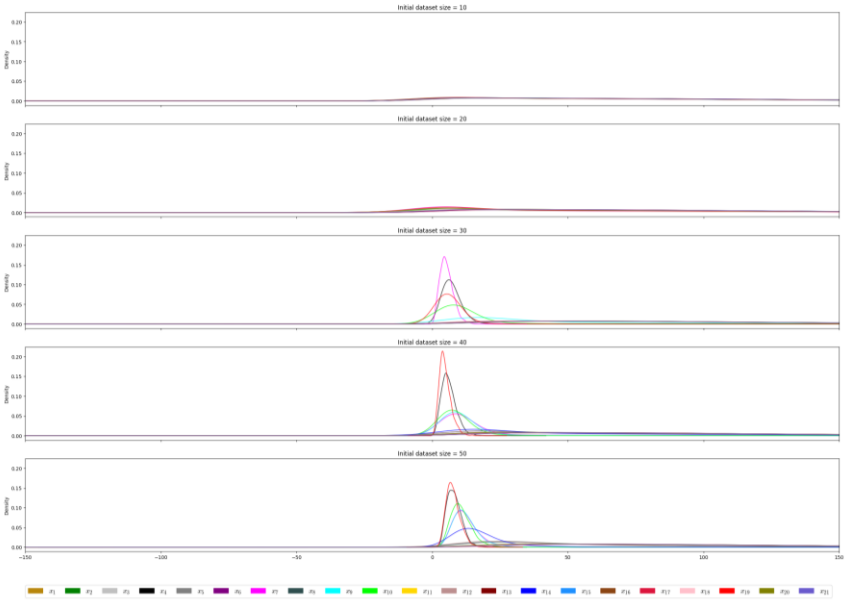
Automatic Relevance Determination (ARD)

Use the Kernel used to train the GP

$$k_{se-ard}(\mathbf{x}_i, \mathbf{x}'_j) = \sigma_f^2 exp\Big(-\frac{1}{2}\sum_{d=1}^{D}\frac{(x_{i,d} - x_{j,d})^2}{l_d^2}\Big)$$

data

latent variables



Very computationally cheap

It is usually used

It is working *

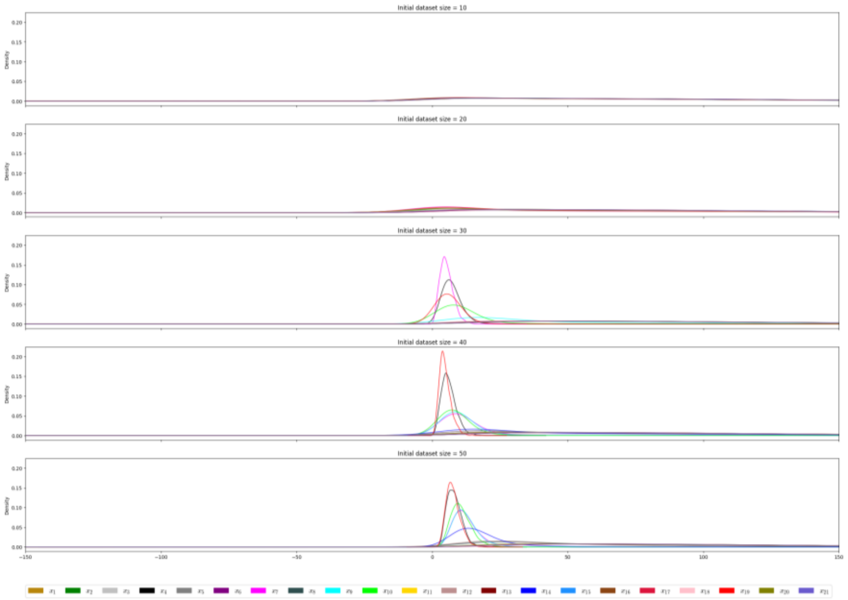
Automatic **R**elevance **D**etermination (ARD)

Use the Kernel used to train the GP

$$k_{se-ard}(\mathbf{x}_i, \mathbf{x}'_j) = \sigma_f^2 exp\Big(-\frac{1}{2}\sum_{d=1}^{D}\frac{(x_{i,d} - x_{j,d})^2}{l_d^2}\Big)$$

data

latent variables



Very computationally cheap

It is usually used

It is working *

* But MISLEADING → Favors **nonlinear** parameters over **linear** ones

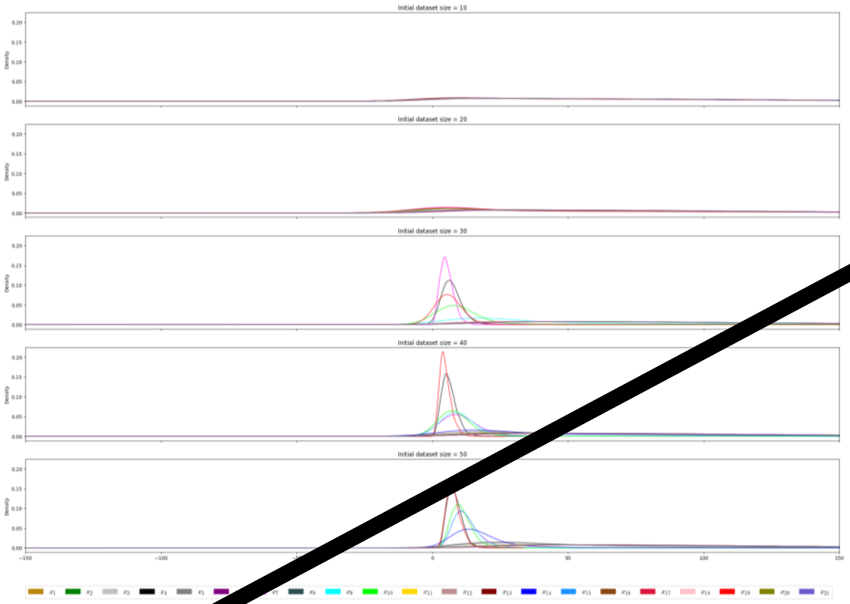
Automatic **R**elevance **D**etermination (ARD)

Use the Kernel used to train the GP

$$k_{se-ard}(\mathbf{x}_i, \mathbf{x}'_j) = \sigma_f^2 exp\left(-\frac{1}{2}\sum_{d=1}^{D}\frac{(x_{i,d}-x_{j,d})^2}{l_d^2}\right)$$

data

latent variables



Very computationally cheap

It is usually used

It is working *

* But MISLEADING → Favors **nonlinear** parameters over **linear** ones

Kullback-Leibler Divergence as a Measure of Predictive Relevance

$$d(p \,||\, q) = \sqrt{2\,\mathcal{D}_{\mathrm{KL}}(p \,||\, q)}$$

Variable selection for Gaussian processes via sensitivity analysis of the posterior predictive distribution
Topi Paananen, Juho Piironen, Michael Riis Andersen, Aki Vehtari

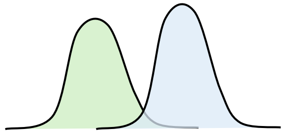
Kullback-Leibler Divergence as a Measure of Predictive Relevance

$$D_{\mathrm{KL}}(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$

$$d(p \,||\, q) = \sqrt{2\, \mathcal{D}_{\mathrm{KL}}(p \,||\, q)}$$

Variable selection for Gaussian processes via sensitivity analysis of the posterior predictive distribution
Topi Paananen, Juho Piironen, Michael Riis Andersen, Aki Vehtari

Kullback-Leibler Divergence as a Measure of Predictive Relevance

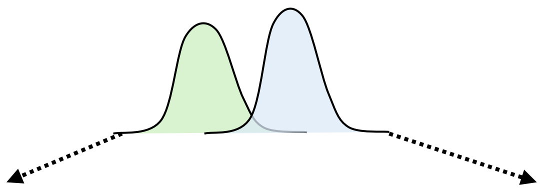


$$D_{\mathrm{KL}}(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$

$$d(p \,||\, q) = \sqrt{2\, \mathcal{D}_{\mathrm{KL}}(p \,||\, q)}$$

Variable selection for Gaussian processes via sensitivity analysis of the posterior predictive distribution
Topi Paananen, Juho Piironen, Michael Riis Andersen, Aki Vehtari

Kullback-Leibler Divergence as a Measure of Predictive Relevance

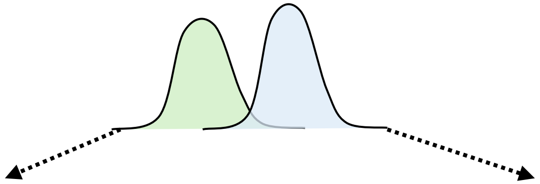


$$D_{\mathrm{KL}}(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x)\ \log\left(\frac{P(x)}{Q(x)}\right)$$

$$d(p\,||\,q) = \sqrt{2\,\mathcal{D}_{\mathrm{KL}}(p\,||\,q)}$$

Variable selection for Gaussian processes via sensitivity analysis of the posterior predictive distribution
Topi Paananen, Juho Piironen, Michael Riis Andersen, Aki Vehtari

Kullback-Leibler Divergence as a Measure of Predictive Relevance



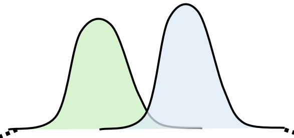
$$D_{\mathrm{KL}}(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x) \, \log\left(\frac{P(x)}{Q(x)}\right)$$

$$d(p \,||\, q) = \sqrt{2\, D_{\mathrm{KL}}(p \,||\, q)}$$

$$r(i, j, \Delta) = \frac{d(p(y_* | \mathbf{x}^{(i)}, \mathbf{y}) \,||\, p(y_* | \mathbf{x}^{(i)} + \Delta_j, \mathbf{y}))}{\Delta}$$

Variable selection for Gaussian processes via sensitivity analysis of the posterior predictive distribution
Topi Paananen, Juho Piironen, Michael Riis Andersen, Aki Vehtari

Kullback-Leibler Divergence as a Measure of Predictive Relevance



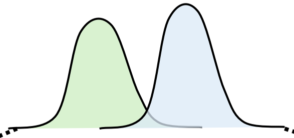
$$D_{\mathrm{KL}}(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$

$$d(p \,||\, q) = \sqrt{2\, D_{\mathrm{KL}}(p \,||\, q)}$$

$$r(i, j, \Delta) = \frac{d\left(p(y_* | \mathbf{x}^{(i)}, \mathbf{y}) \,||\, p(y_* | \mathbf{x}^{(i)} + \Delta_j, \mathbf{y})\right)}{\Delta}$$

Variable selection for Gaussian processes via sensitivity analysis of the posterior predictive distribution
Topi Paananen, Juho Piironen, Michael Riis Andersen, Aki Vehtari

Kullback-Leibler Divergence as a Measure of Predictive Relevance



$$D_{\mathrm{KL}}(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$

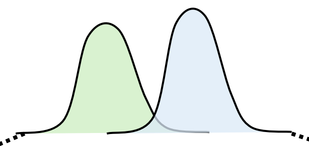
$$d(p \,||\, q) = \sqrt{2\, D_{\mathrm{KL}}(p \,||\, q)}$$

Surrogate Model

$$r(i, j, \Delta) = \frac{d\left(p(y_* | \mathbf{x}^{(i)}, \mathbf{y}) \,||\, p(y_* | \mathbf{x}^{(i)} + \Delta_j, \mathbf{y})\right)}{\Delta}$$

Variable selection for Gaussian processes via sensitivity analysis of the posterior predictive distribution
Topi Paananen, Juho Piironen, Michael Riis Andersen, Aki Vehtari

Kullback-Leibler Divergence as a Measure of Predictive Relevance



$$D_{\mathrm{KL}}(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$

$$d(p\,||\,q) = \sqrt{2\,\mathcal{D}_{\mathrm{KL}}(p\,||\,q)}$$

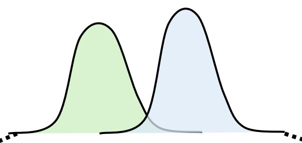
Surrogate Model

Parameter index

$$r(i, j, \Delta) = \frac{d\left(p(y_* | \mathbf{x}^{(i)}, \mathbf{y})\,||\,p(y_* | \mathbf{x}^{(i)} + \Delta_j, \mathbf{y})\right)}{\Delta}$$

Variable selection for Gaussian processes via sensitivity analysis of the posterior predictive distribution
Topi Paananen, Juho Piironen, Michael Riis Andersen, Aki Vehtari

Kullback-Leibler Divergence as a Measure of Predictive Relevance



Predictive distribution of training data

Predictive distribution of perturbed training data

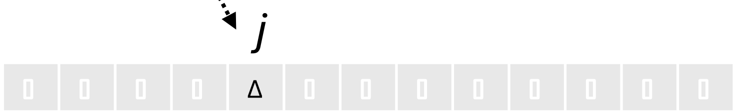
$$D_{\mathrm{KL}}(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$

$$d(p \,||\, q) = \sqrt{2\, D_{\mathrm{KL}}(p \,||\, q)}$$

Surrogate Model

Parameter index

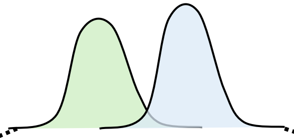
$$r(i, j, \Delta) = \frac{d\left(p(y_* | \mathbf{x}^{(i)}, \mathbf{y}) \,||\, p(y_* | \mathbf{x}^{(i)} + \Delta_j, \mathbf{y})\right)}{\Delta}$$



Variable selection for Gaussian processes via sensitivity analysis of the posterior predictive distribution

Topi Paananen, Juho Piironen, Michael Riis Andersen, Aki Vehtari

Kullback-Leibler Divergence as a Measure of Predictive Relevance



Predictive distribution of training data

Predictive distribution of perturbed training data

$$D_{\mathrm{KL}}(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$

$$d(p\,||\,q) = \sqrt{2\,\mathcal{D}_{\mathrm{KL}}(p\,||\,q)}$$

Surrogate Model

Parameter index

$$r(i, j, \Delta) = \frac{d\left(p(y_* | \mathbf{x}^{(i)}, \mathbf{y}) \,||\, p(y_* | \mathbf{x}^{(i)} + \Delta_j, \mathbf{y})\right)}{\Delta}$$

$$\mathrm{KL}_j = \frac{1}{n} \sum_{i=1}^{n} r(i, j, \Delta)$$

j

Variable selection for Gaussian processes via sensitivity analysis of the posterior predictive distribution

Topi Paananen, Juho Piironen, Michael Riis Andersen, Aki Vehtari

Proposed Algorithm

Proposed Algorithm

Explore ... Then, Exploit what you explored

(1) Surrogate Model

(2) Sensitivity Analysis

(3) Efficient Sampling

Proposed Algorithm

Explore … Then, Exploit what you explored

(1) Surrogate Model

(2) Sensitivity Analysis

(3) Efficient Sampling

Proposed Algorithm

Part A

Active leaning guides Sensitivity Analysis

Explore ... Then, Exploit what you explored

(1) Surrogate Model

(2) Sensitivity Analysis

(3) Efficient Sampling

Proposed Algorithm

Part A

Active leaning guides Sensitivity Analysis

Pure Exploration

Explore ... Then, Exploit what you explored

(1) Surrogate Model

(2) Sensitivity Analysis

(3) Efficient Sampling

Proposed Algorithm

Part A

Active leaning guides Sensitivity Analysis

Pure Exploration

Part B

Sensitivity Analysis guides Active Learning

Explore … Then, Exploit what you explored

(1) Surrogate Model

(2) Sensitivity Analysis

(3) Efficient Sampling

Proposed Algorithm

Part A

Active leaning guides Sensitivity Analysis

Pure Exploration

Part B

Sensitivity Analysis guides Active Learning

Exploitation with wise Exploration

Case Study

$$y = 12x_1 + 20x_2 - 13x_3 + 122x_4 - 44x_5 + 0.01x_6 - 77x_7 - 33x_8 + 90x_9 - 102x_{10} + 11x_{11} + 44x_{12} + 11x_{13} - 66x_{14}$$
$$+ 8x_{15} - 0.001x_{16} + 0.05x_{17} - 0.8x_{18} + 145x_{19} - 22x_{20} + 13x_{21}$$

Case Study

$$y = 12x_1 + 20x_2 - 13x_3 + 122x_4 - 44x_5 + 0.01x_6 - 77x_7 - 33x_8 + 90x_9 - 102x_{10} + 11x_{11} + 44x_{12} + 11x_{13} - 66x_{14} + 8x_{15} - 0.001x_{16} + 0.05x_{17} - 0.8x_{18} + 145x_{19} - 22x_{20} + 13x_{21}$$

Case Study

$$y = 12x_1 + 20x_2 - 13x_3 + 122x_4 - 44x_5 + 0.01x_6 - 77x_7 - 33x_8 + 90x_9 - 102x_{10} + 11x_{11} + 44x_{12} + 11x_{13} - 66x_{14} + 8x_{15} - 0.001x_{16} + 0.05x_{17} - 0.8x_{18} + 145x_{19} - 22x_{20} + 13x_{21}$$

Case Study

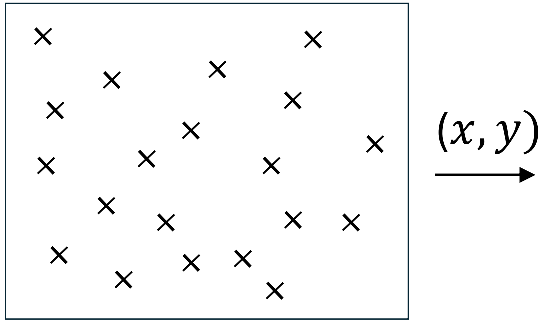
$$y = 12x_1 + 20x_2 - 13x_3 + 122x_4 - 44x_5 + 0.01x_6 - 77x_7 - 33x_8 + 90x_9 - 102x_{10} + 11x_{11} + 44x_{12} + 11x_{13} - 66x_{14} + 8x_{15} - 0.001x_{16} + 0.05x_{17} - 0.8x_{18} + 145x_{19} - 22x_{20} + 13x_{21}$$

Case Study

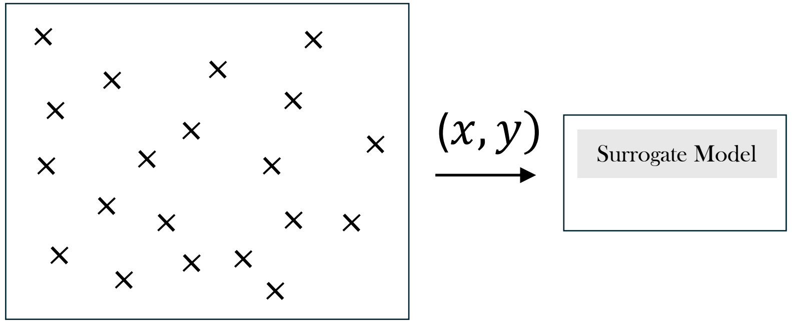
$$y = 12x_1 + 20x_2 - 13x_3 + 122x_4 - 44x_5 + 0.01x_6 - 77x_7 - 33x_8 + 90x_9 - 102x_{10} + 11x_{11} + 44x_{12} + 11x_{13} - 66x_{14}$$
$$+ 8x_{15} - 0.001x_{16} + 0.05x_{17} - 0.8x_{18} + 145x_{19} - 22x_{20} + 13x_{21}$$

Case Study

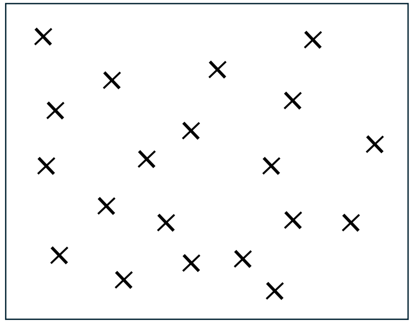
$$y = 12x_1 + 20x_2 - 13x_3 + 122x_4 - 44x_5 + 0.01x_6 - 77x_7 - 33x_8 + 90x_9 - 102x_{10} + 11x_{11} + 44x_{12} + 11x_{13} - 66x_{14} + 8x_{15} - 0.001x_{16} + 0.05x_{17} - 0.8x_{18} + 145x_{19} - 22x_{20} + 13x_{21}$$



LHS $n_{initial}$=20



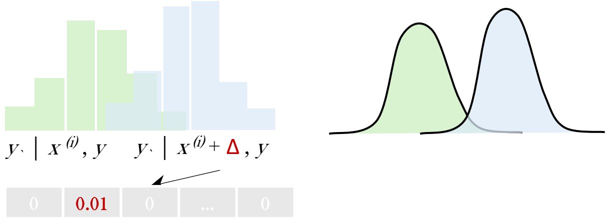
LHS $n_{initial}$=20



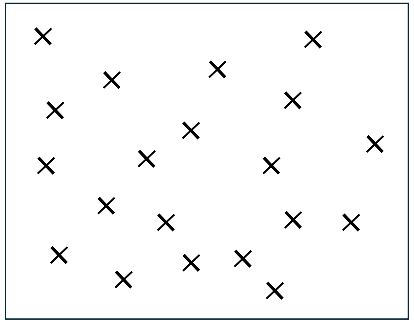
LHS $n_{initial}$=20

(x, y) → Surrogate Model →

$$r(i,j,\Delta) = \frac{d(p(y_*|\mathbf{x}^{(i)},\mathbf{y}) \,||\, p(y_*|\mathbf{x}^{(i)}+\Delta_j,\mathbf{y}))}{\Delta}$$



For each input parameter ,j, and training point, i

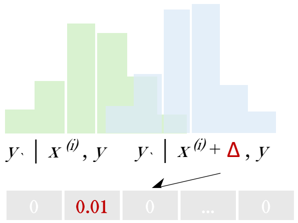


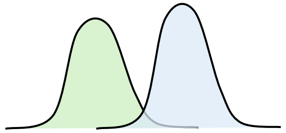
LHS $n_{initial}$=20

(x, y)

Surrogate Model

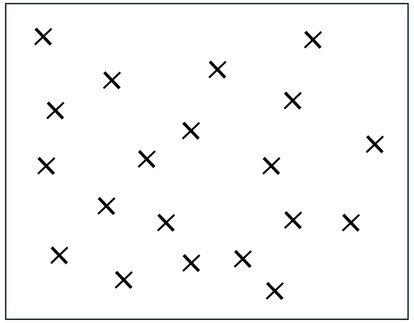
$$r(i,j,\Delta) = \frac{d(p(y_*|\mathbf{x}^{(i)},\mathbf{y})\,||\,p(y_*|\mathbf{x}^{(i)}+\Delta_j,\mathbf{y}))}{\Delta}$$





For each input parameter ,j, and training point, i

$$\mathrm{KL}_j = \frac{1}{n}\sum_{i=1}^{n} r(i,j,\Delta)$$

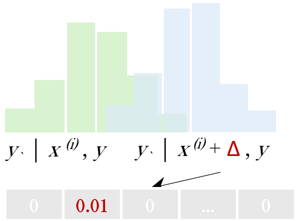


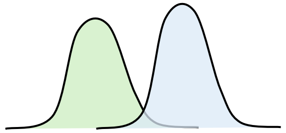
LHS $n_{initial}$=20

(x, y)

Surrogate Model

$$r(i, j, \Delta) = \frac{d\left(p(y_* | \mathbf{x}^{(i)}, \mathbf{y}) \,||\, p(y_* | \mathbf{x}^{(i)} + \Delta_j, \mathbf{y})\right)}{\Delta}$$





For each input parameter ,j, and training point, i

$$\mathrm{KL}_j = \frac{1}{n} \sum_{i=1}^{n} r(i, j, \Delta)$$

Rank the parameters according to their relevance

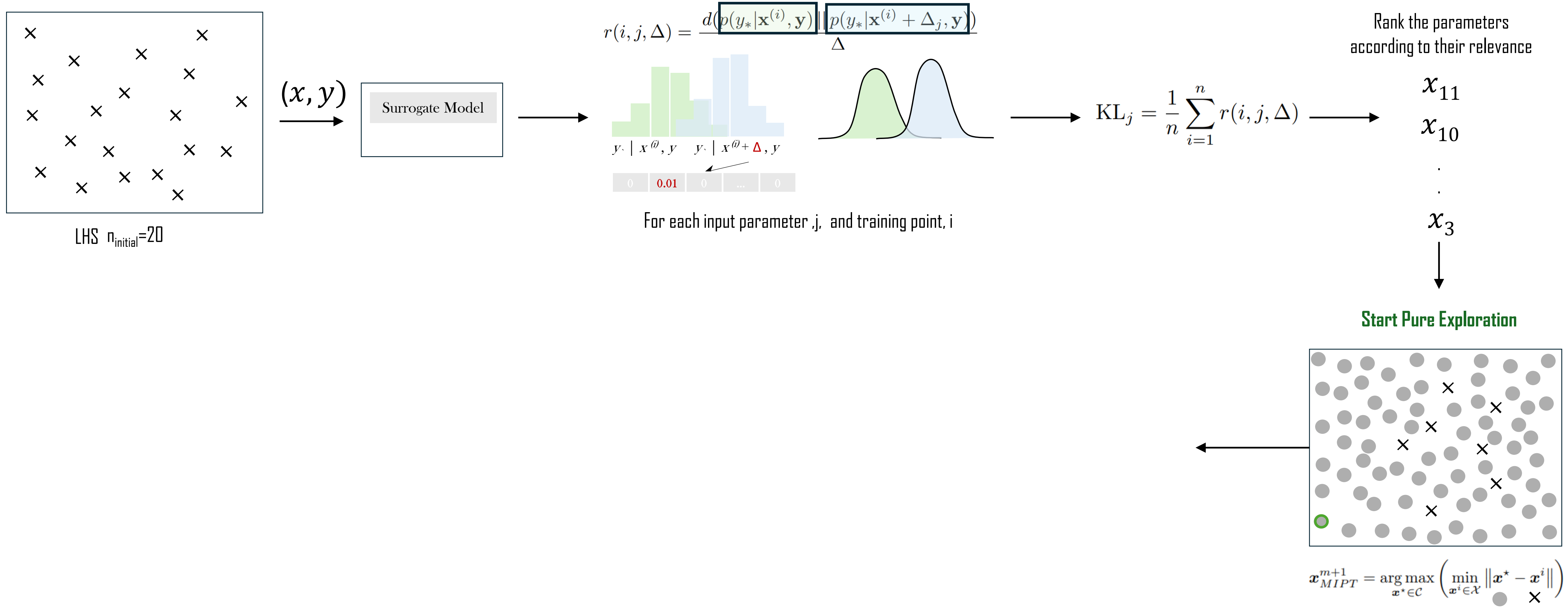
x_{11}

x_{10}

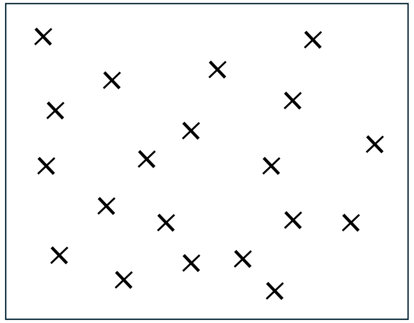
.

.

x_3



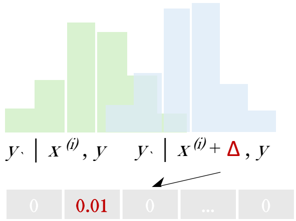
LHS $n_{initial}$=20
(x, y)
Surrogate Model
$r(i, j, \Delta) = \frac{d(p(y_*|\mathbf{x}^{(i)}, \mathbf{y}) || p(y_*|\mathbf{x}^{(i)} + \Delta_j, \mathbf{y}))}{\Delta}$
$y_* | x^{(i)}, y$
$y_* | x^{(i)} + \Delta, y$
0
0.01
0
...
0
For each input parameter ,j, and training point, i
$\mathrm{KL}_j = \frac{1}{n} \sum_{i=1}^{n} r(i, j, \Delta)$
Rank the parameters according to their relevance
x_{11}
x_{10}
.
.
x_3
Start Pure Exploration
$\boldsymbol{x}_{MIPT}^{m+1} = \arg\max_{\boldsymbol{x}^\star \in \mathcal{C}} \left(\min_{\boldsymbol{x}^i \in \mathcal{X}} \|\boldsymbol{x}^\star - \boldsymbol{x}^i\| \right)$

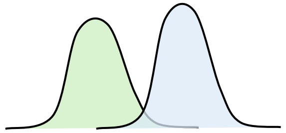


LHS $n_{initial}$=20

(x, y) → Surrogate Model →

$$r(i,j,\Delta) = \frac{d\left(p(y_*|\mathbf{x}^{(i)},\mathbf{y}) \,||\, p(y_*|\mathbf{x}^{(i)}+\Delta_j,\mathbf{y})\right)}{\Delta}$$





For each input parameter ,j, and training point, i

$$\mathrm{KL}_j = \frac{1}{n}\sum_{i=1}^{n} r(i,j,\Delta)$$

Rank the parameters according to their relevance

x_{11}

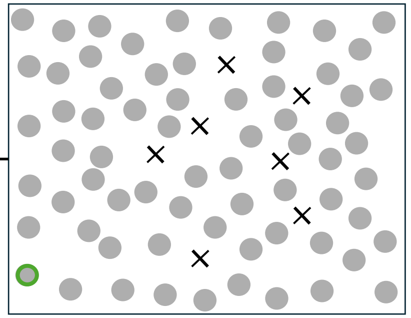
x_{10}

.

.

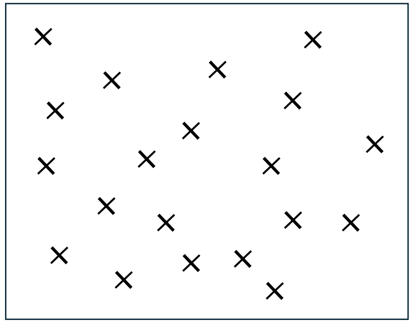
x_3

Start Pure Exploration



$$\boldsymbol{x}^{m+1}_{MIPT} = \arg\max_{\boldsymbol{x}^\star \in \mathcal{C}} \left(\min_{\boldsymbol{x}^i \in \mathcal{X}} \left\| \boldsymbol{x}^\star - \boldsymbol{x}^i \right\| \right)$$

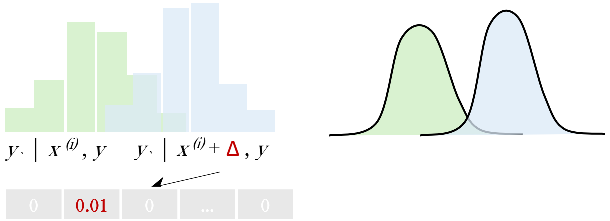
Have Parallel Resources ? Run 10 simulations rather than 1



LHS $n_{initial}$=20

(x, y) → Surrogate Model →

$$r(i,j,\Delta) = \frac{d\left(p(y_*|\mathbf{x}^{(i)},\mathbf{y}) \,||\, p(y_*|\mathbf{x}^{(i)}+\Delta_j,\mathbf{y})\right)}{\Delta}$$



For each input parameter ,j, and training point, i

$$\mathrm{KL}_j = \frac{1}{n}\sum_{i=1}^{n} r(i,j,\Delta)$$

Rank the parameters according to their relevance

x_{11}

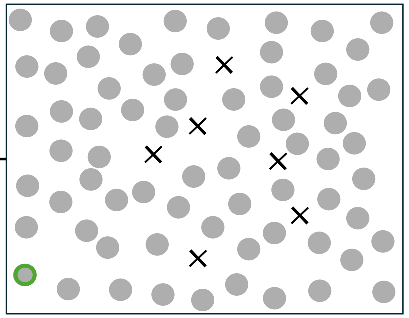
x_{10}

.

.

x_3

Start Pure Exploration

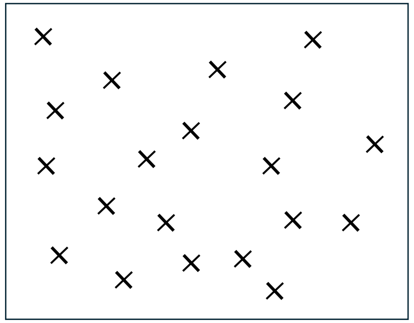


$$\boldsymbol{x}^{m+1}_{MIPT} = \arg\max_{\boldsymbol{x}^\star \in \mathcal{C}} \left(\min_{\boldsymbol{x}^i \in \mathcal{X}} \|\boldsymbol{x}^\star - \boldsymbol{x}^i\| \right)$$

Have Parallel Resources ? Run 10 simulations rather than 1

Can we take the top 10 points from this equation instead of the argmax ?

$$\boldsymbol{x}^{m+1}_{MIPT} = \arg\max_{\boldsymbol{x}^\star \in \mathcal{C}} \left(\min_{\boldsymbol{x}^i \in \mathcal{X}} \|\boldsymbol{x}^\star - \boldsymbol{x}^i\| \right)$$

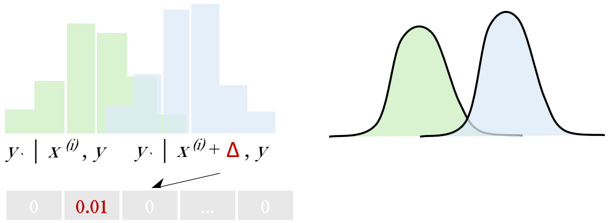


LHS $n_{initial}$=20

(x, y)

Surrogate Model

$$r(i,j,\Delta) = \frac{d\left(p(y_*|\mathbf{x}^{(i)},\mathbf{y}) \,\|\, p(y_*|\mathbf{x}^{(i)}+\Delta_j,\mathbf{y})\right)}{\Delta}$$



For each input parameter ,j, and training point, i

$$\mathrm{KL}_j = \frac{1}{n}\sum_{i=1}^{n} r(i,j,\Delta)$$

Rank the parameters according to their relevance

x_{11}

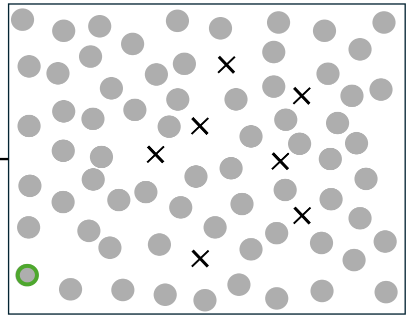
x_{10}

.

.

x_3

Start Pure Exploration



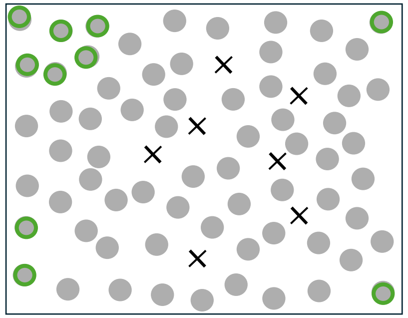
$$\boldsymbol{x}_{MIPT}^{m+1} = \arg\max_{\boldsymbol{x}^\star \in \mathcal{C}} \left(\min_{\boldsymbol{x}^i \in \mathcal{X}} \|\boldsymbol{x}^\star - \boldsymbol{x}^i\| \right)$$

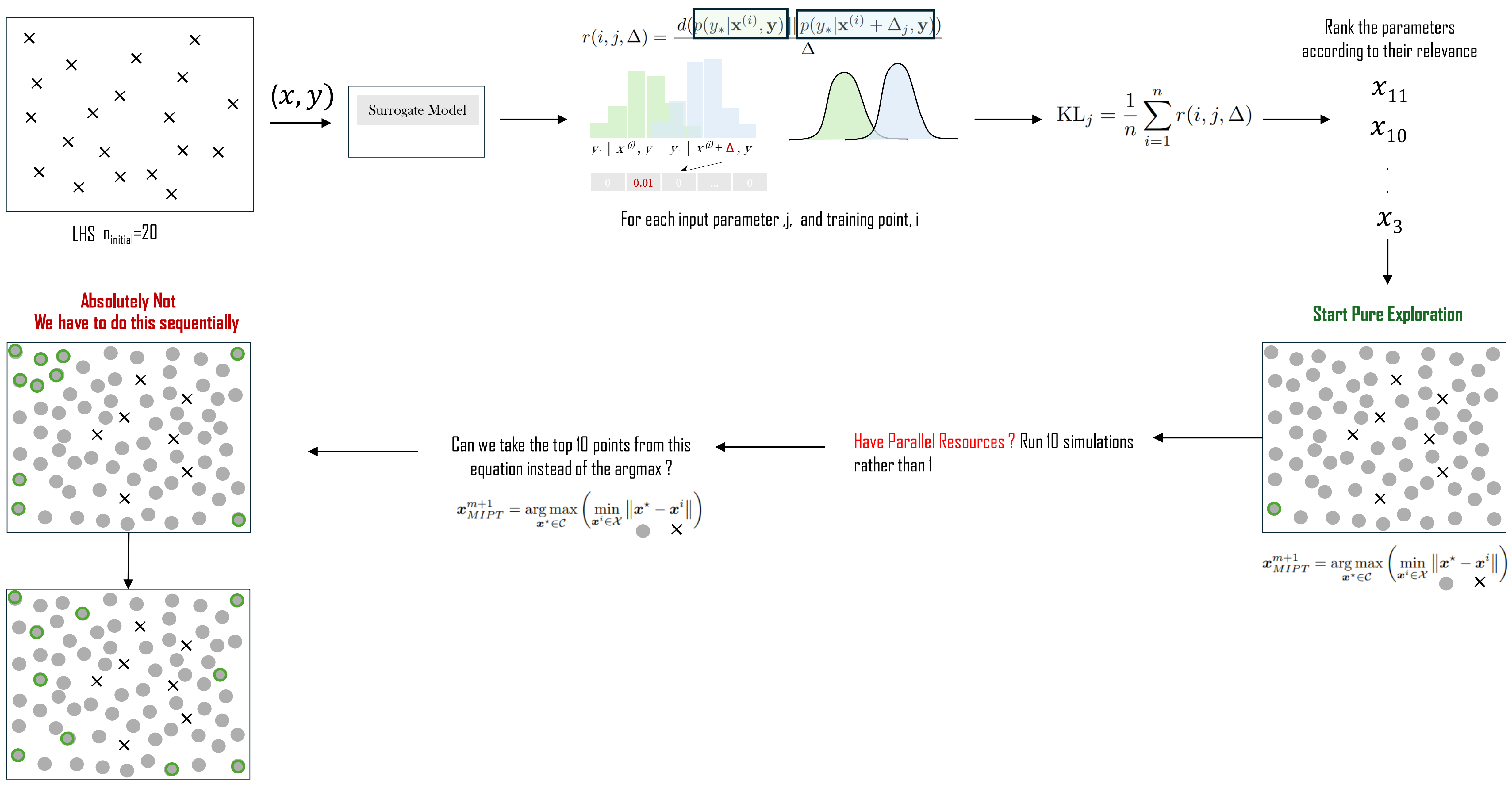
Have Parallel Resources ? Run 10 simulations rather than 1

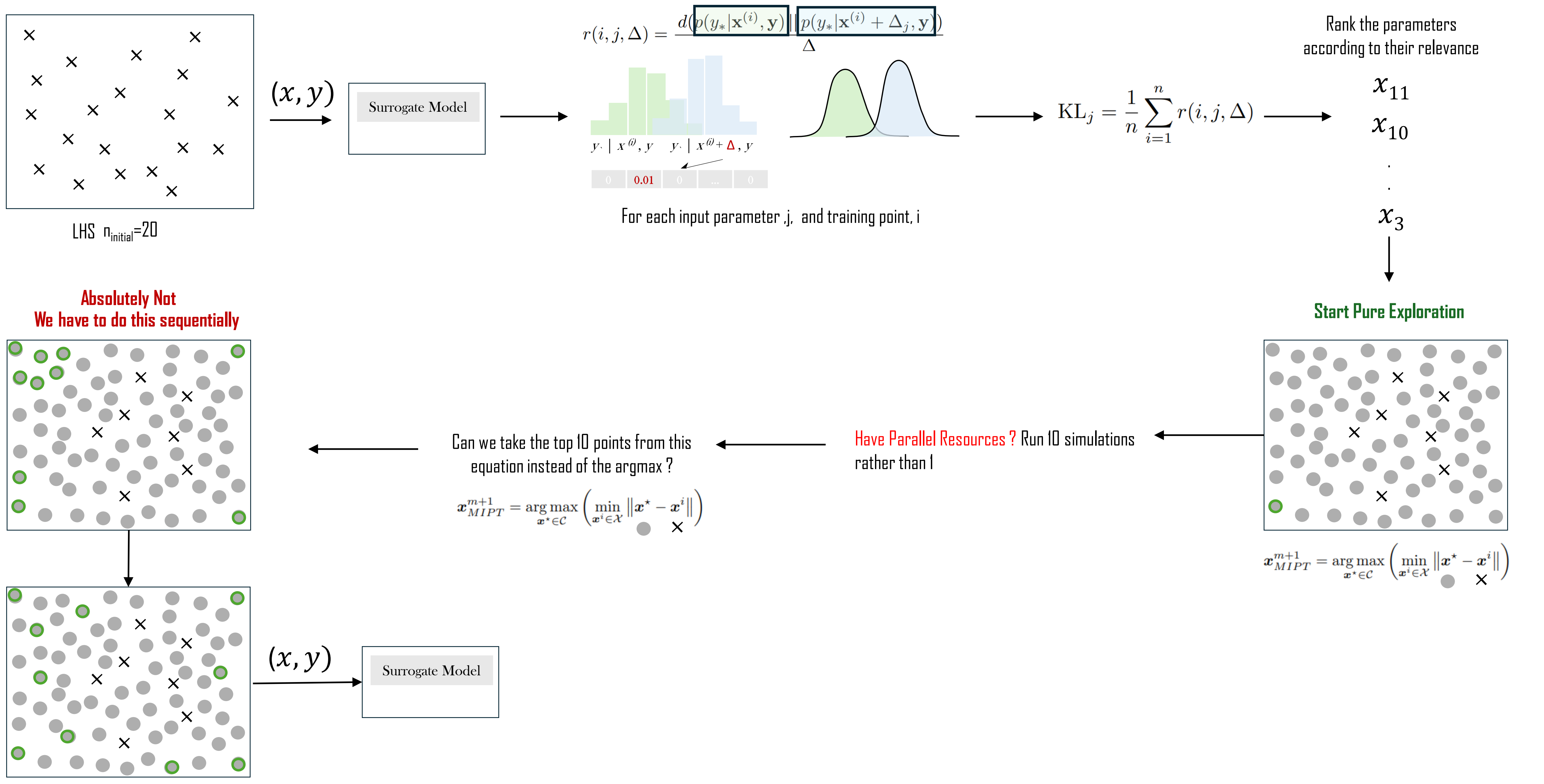
Can we take the top 10 points from this equation instead of the argmax ?

$$\boldsymbol{x}_{MIPT}^{m+1} = \arg\max_{\boldsymbol{x}^\star \in \mathcal{C}} \left(\min_{\boldsymbol{x}^i \in \mathcal{X}} \|\boldsymbol{x}^\star - \boldsymbol{x}^i\| \right)$$

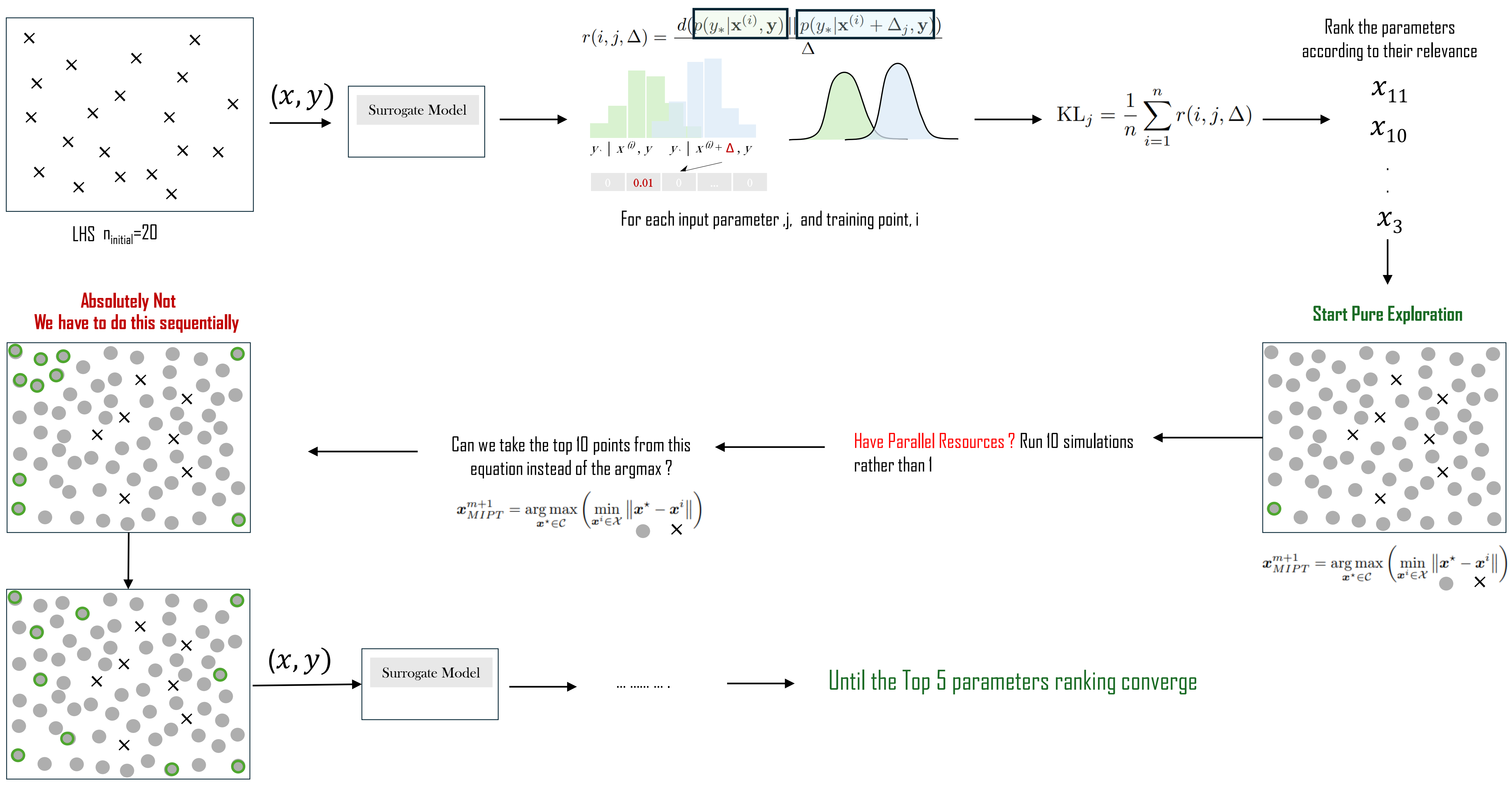
Absolutely Not
We have to do this sequentially



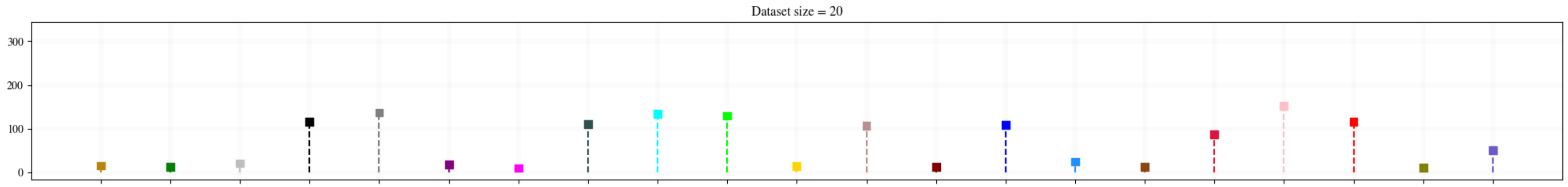

LHS $n_{initial}$=20
(x, y)
Surrogate Model
$r(i,j,\Delta) = \frac{d(p(y_*|\mathbf{x}^{(i)},\mathbf{y}) \| p(y_*|\mathbf{x}^{(i)}+\Delta_j,\mathbf{y}))}{\Delta}$
0
0.01
0
...
0
For each input parameter ,j, and training point, i
$\mathrm{KL}_j = \frac{1}{n}\sum_{i=1}^{n} r(i,j,\Delta)$
Rank the parameters according to their relevance
x_{11}
x_{10}
x_3
Start Pure Exploration
$\boldsymbol{x}_{MIPT}^{m+1} = \arg\max_{\boldsymbol{x}^\star \in \mathcal{C}} \left(\min_{\boldsymbol{x}^i \in \mathcal{X}} \|\boldsymbol{x}^\star - \boldsymbol{x}^i\| \right)$
Have Parallel Resources ? Run 10 simulations rather than 1
Can we take the top 10 points from this equation instead of the argmax ?
$\boldsymbol{x}_{MIPT}^{m+1} = \arg\max_{\boldsymbol{x}^\star \in \mathcal{C}} \left(\min_{\boldsymbol{x}^i \in \mathcal{X}} \|\boldsymbol{x}^\star - \boldsymbol{x}^i\| \right)$
Absolutely Not
We have to do this sequentially



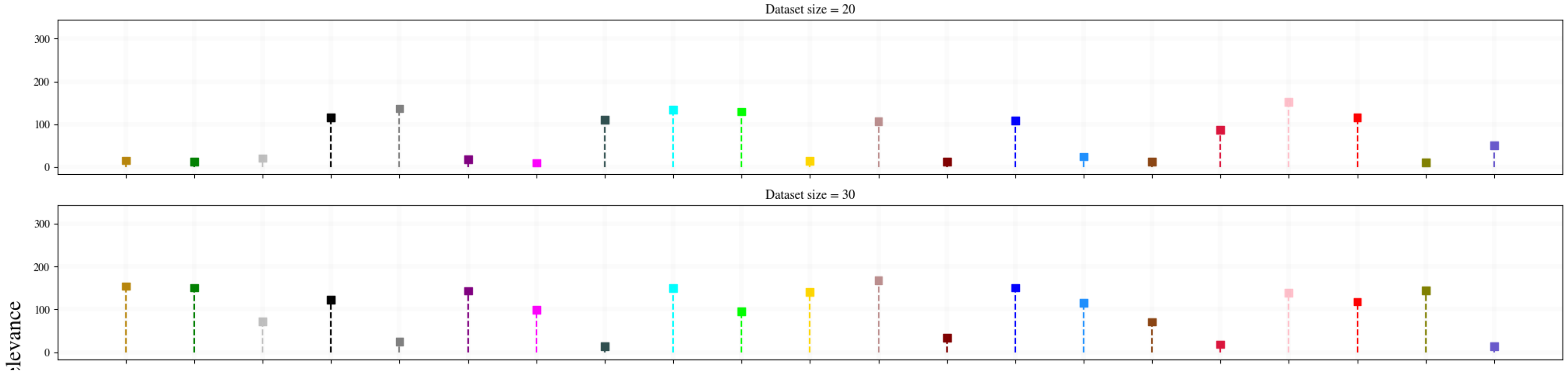
LHS $n_{initial}$=20
(x, y)
Surrogate Model
$r(i, j, \Delta) = \frac{d(p(y_*|\mathbf{x}^{(i)}, \mathbf{y}) \| p(y_*|\mathbf{x}^{(i)} + \Delta_j, \mathbf{y}))}{\Delta}$
0 0.01 0 ... 0
For each input parameter ,j, and training point, i
$\mathrm{KL}_j = \frac{1}{n} \sum_{i=1}^{n} r(i, j, \Delta)$
Rank the parameters according to their relevance
x_{11}
x_{10}
x_3
Start Pure Exploration
$\boldsymbol{x}^{m+1}_{MIPT} = \arg\max_{\boldsymbol{x}^\star \in \mathcal{C}} \left(\min_{\boldsymbol{x}^i \in \mathcal{X}} \|\boldsymbol{x}^\star - \boldsymbol{x}^i\| \right)$
Have Parallel Resources ? Run 10 simulations rather than 1
Can we take the top 10 points from this equation instead of the argmax ?
$\boldsymbol{x}^{m+1}_{MIPT} = \arg\max_{\boldsymbol{x}^\star \in \mathcal{C}} \left(\min_{\boldsymbol{x}^i \in \mathcal{X}} \|\boldsymbol{x}^\star - \boldsymbol{x}^i\| \right)$
Absolutely Not
We have to do this sequentially
(x, y)
Surrogate Model


LHS $n_{initial}$=20
(x, y)
Surrogate Model
$r(i,j,\Delta) = \frac{d(p(y_*|\mathbf{x}^{(i)},\mathbf{y}) \| p(y_*|\mathbf{x}^{(i)}+\Delta_j,\mathbf{y}))}{\Delta}$
0
0.01
0
...
0
For each input parameter ,j, and training point, i
$\mathrm{KL}_j = \frac{1}{n}\sum_{i=1}^{n} r(i,j,\Delta)$
Rank the parameters according to their relevance
x_{11}
x_{10}
x_3
Start Pure Exploration
$\boldsymbol{x}^{m+1}_{MIPT} = \arg\max_{\boldsymbol{x}^\star \in \mathcal{C}} \left(\min_{\boldsymbol{x}^i \in \mathcal{X}} \|\boldsymbol{x}^\star - \boldsymbol{x}^i\| \right)$
Have Parallel Resources ? Run 10 simulations rather than 1
Can we take the top 10 points from this equation instead of the argmax ?
$\boldsymbol{x}^{m+1}_{MIPT} = \arg\max_{\boldsymbol{x}^\star \in \mathcal{C}} \left(\min_{\boldsymbol{x}^i \in \mathcal{X}} \|\boldsymbol{x}^\star - \boldsymbol{x}^i\| \right)$
Absolutely Not
We have to do this sequentially
(x, y)
Surrogate Model
... .
Until the Top 5 parameters ranking converge

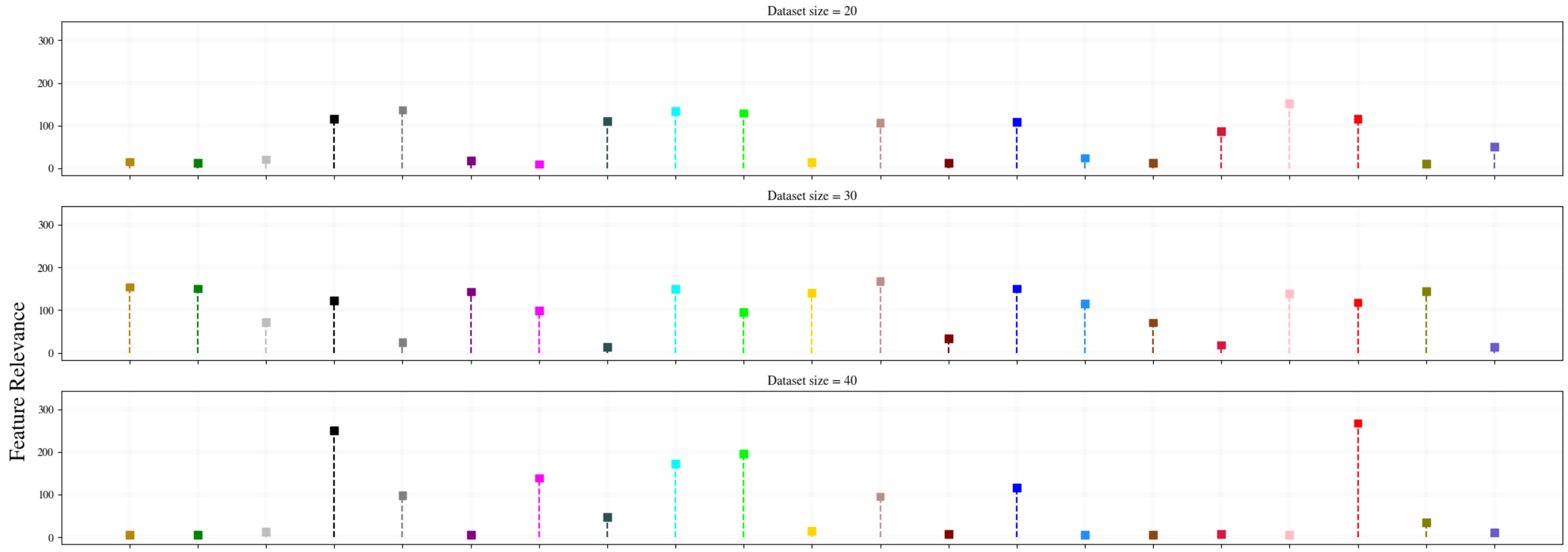
$$y = 12x_1 + 20x_2 - 13x_3 + 122x_4 - 44x_5 + 0.01x_6 - 77x_7 - 33x_8 + 90x_9 - 102x_{10} + 11x_{11} + 44x_{12} + 11x_{13} - 66x_{14} + 8x_{15} - 0.001x_{16} + 0.05x_{17} - 0.8x_{18} + 145x_{19} - 22x_{20} + 13x_{21}$$



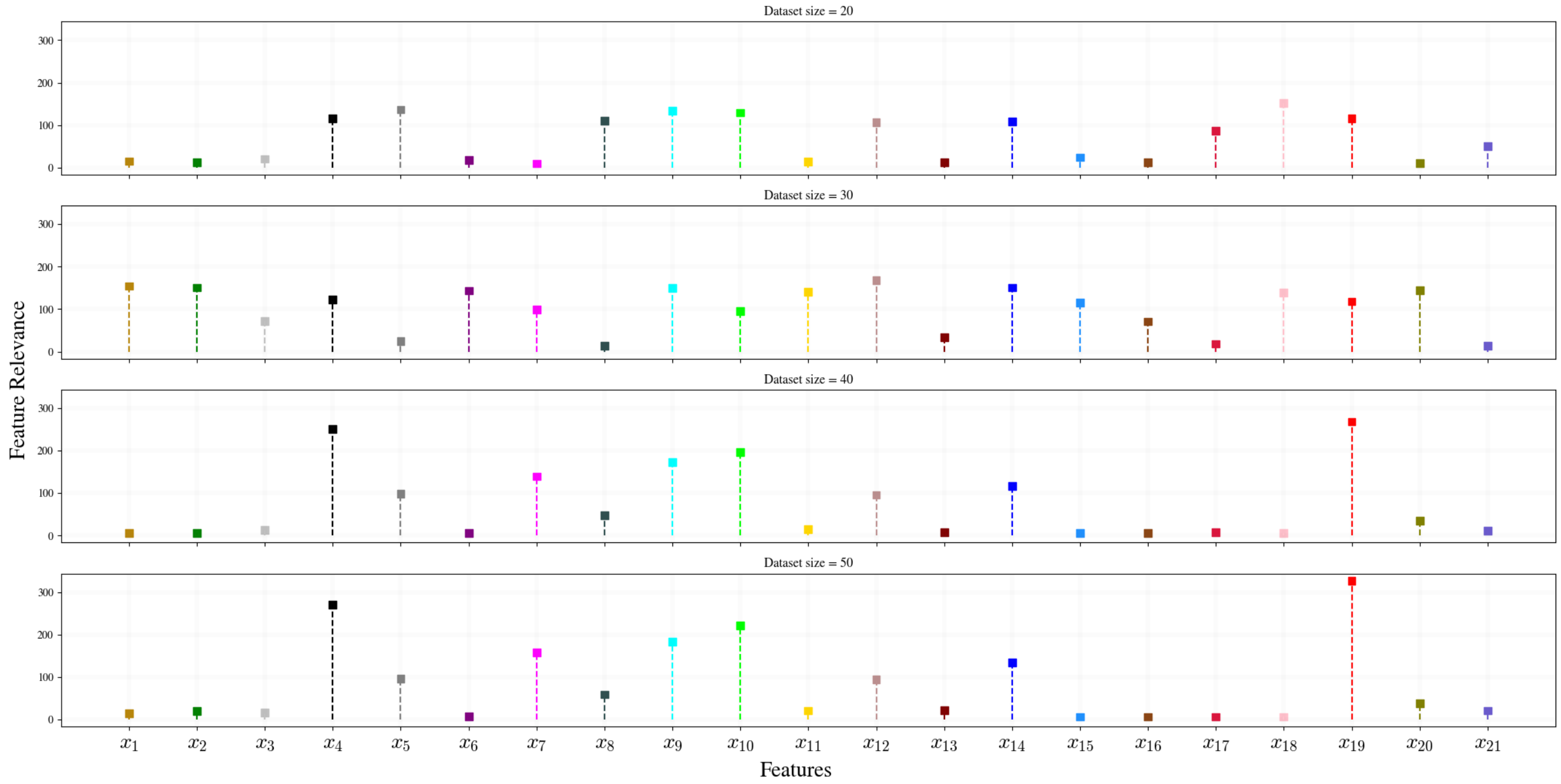
$$y = 12x_1 + 20x_2 - 13x_3 + 122x_4 - 44x_5 + 0.01x_6 - 77x_7 - 33x_8 + 90x_9 - 102x_{10} + 11x_{11} + 44x_{12} + 11x_{13} - 66x_{14} + 8x_{15} - 0.001x_{16} + 0.05x_{17} - 0.8x_{18} + 145x_{19} - 22x_{20} + 13x_{21}$$



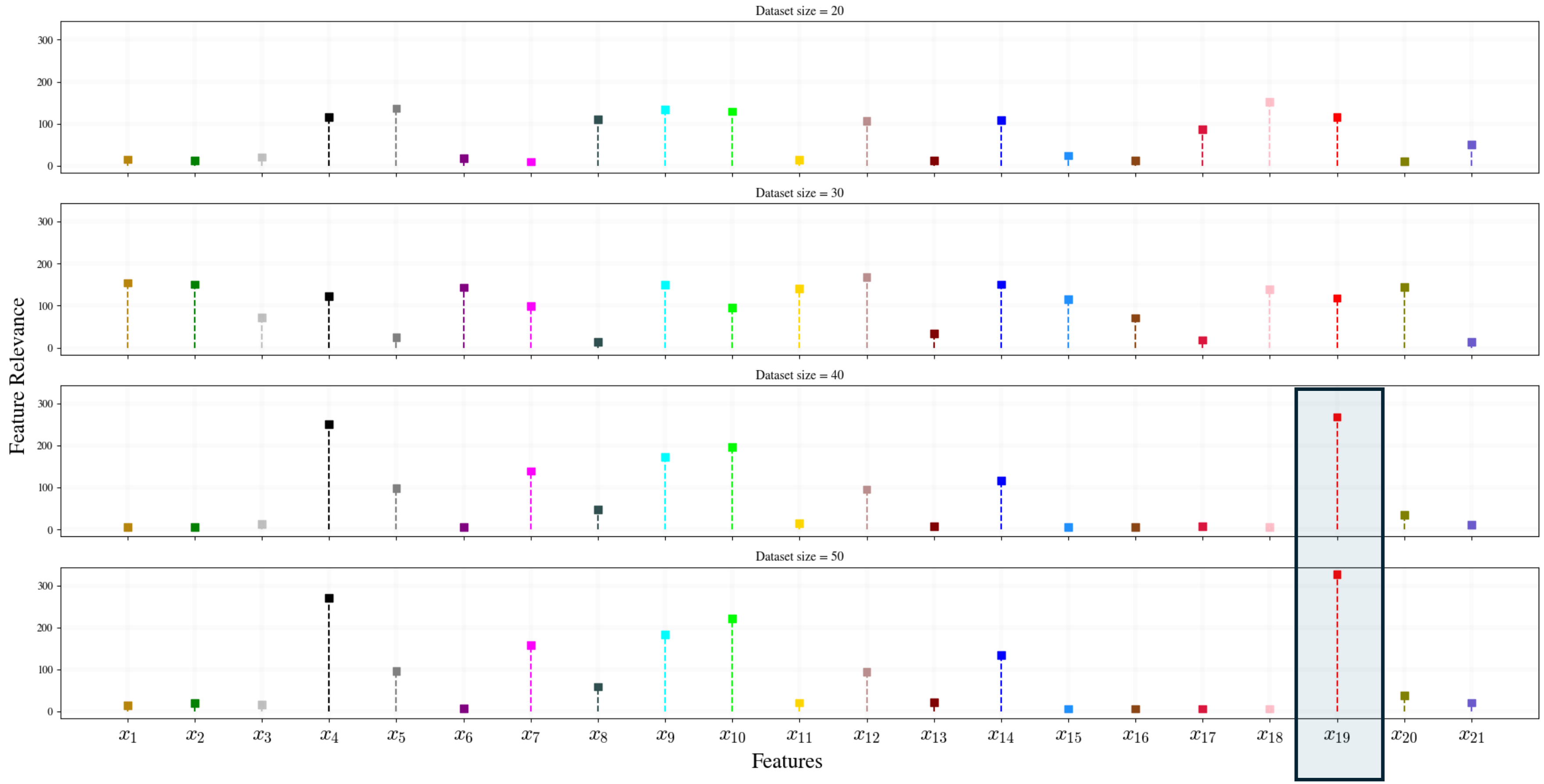
$$y = 12x_1 + 20x_2 - 13x_3 + 122x_4 - 44x_5 + 0.01x_6 - 77x_7 - 33x_8 + 90x_9 - 102x_{10} + 11x_{11} + 44x_{12} + 11x_{13} - 66x_{14} + 8x_{15} - 0.001x_{16} + 0.05x_{17} - 0.8x_{18} + 145x_{19} - 22x_{20} + 13x_{21}$$



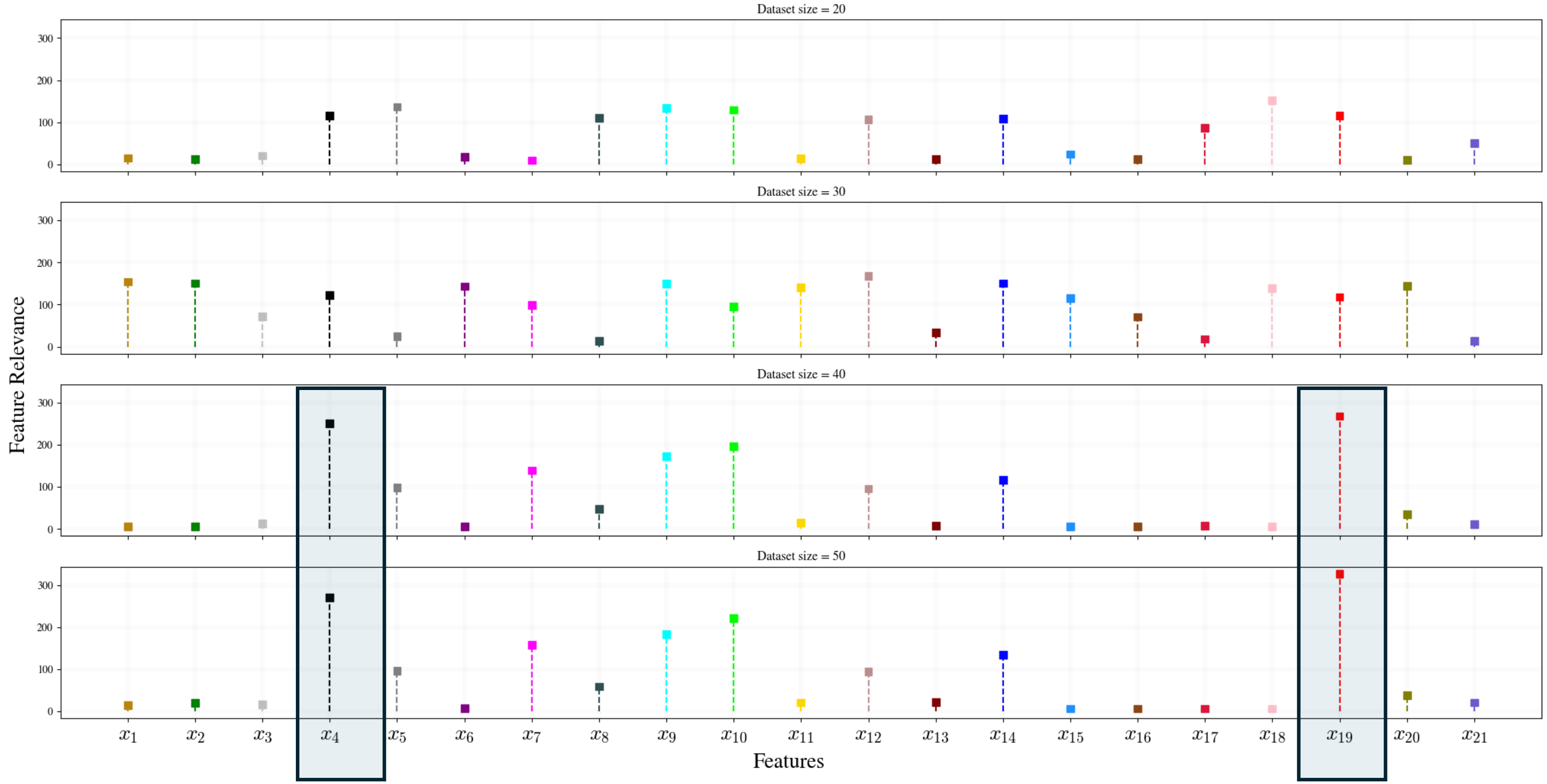
$$y = 12x_1 + 20x_2 - 13x_3 + 122x_4 - 44x_5 + 0.01x_6 - 77x_7 - 33x_8 + 90x_9 - 102x_{10} + 11x_{11} + 44x_{12} + 11x_{13} - 66x_{14} + 8x_{15} - 0.001x_{16} + 0.05x_{17} - 0.8x_{18} + 145x_{19} - 22x_{20} + 13x_{21}$$



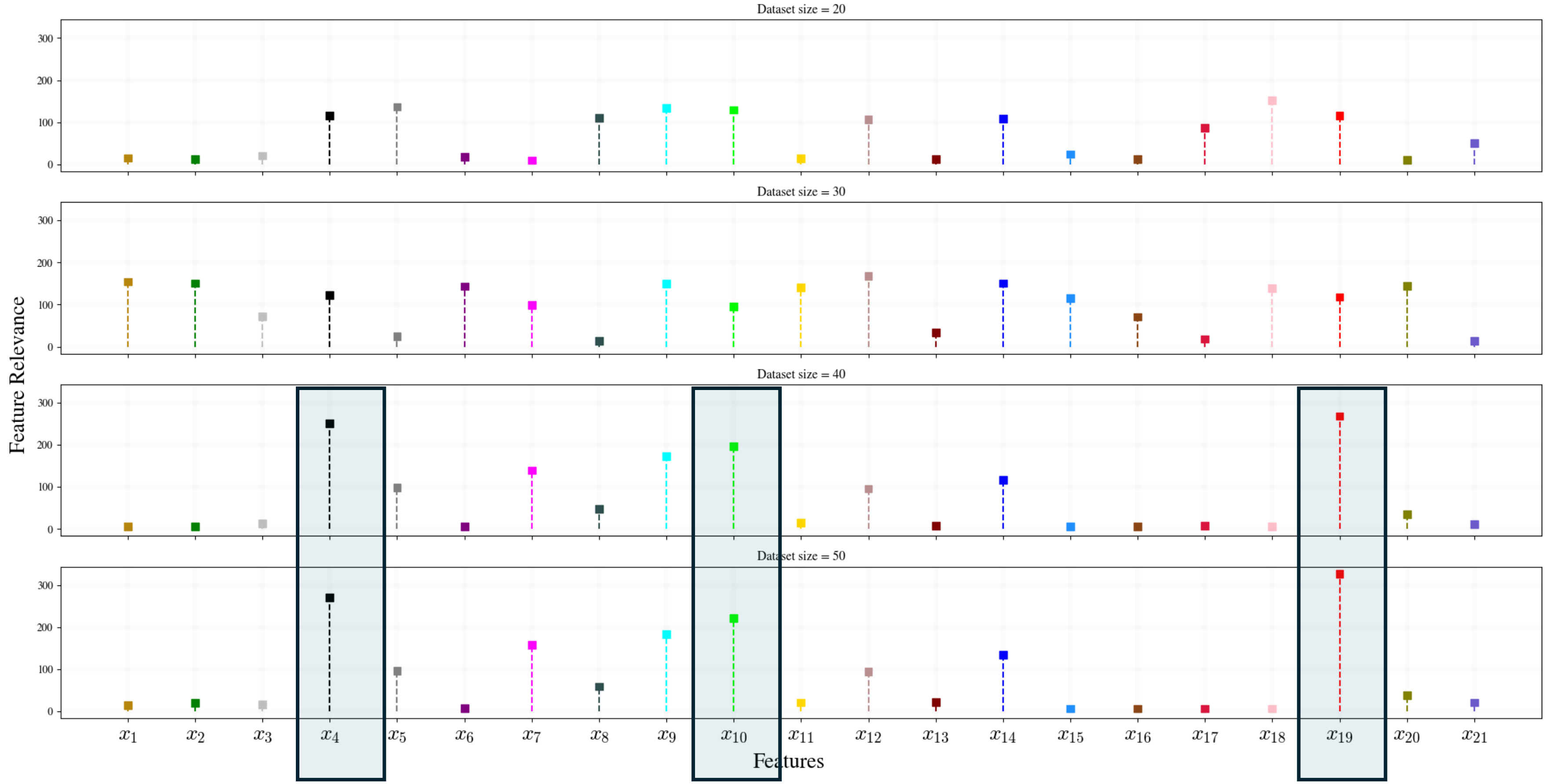
$$y = 12x_1 + 20x_2 - 13x_3 + 122x_4 - 44x_5 + 0.01x_6 - 77x_7 - 33x_8 + 90x_9 - 102x_{10} + 11x_{11} + 44x_{12} + 11x_{13} - 66x_{14} + 8x_{15} - 0.001x_{16} + 0.05x_{17} - 0.8x_{18} + 145x_{19} - 22x_{20} + 13x_{21}$$



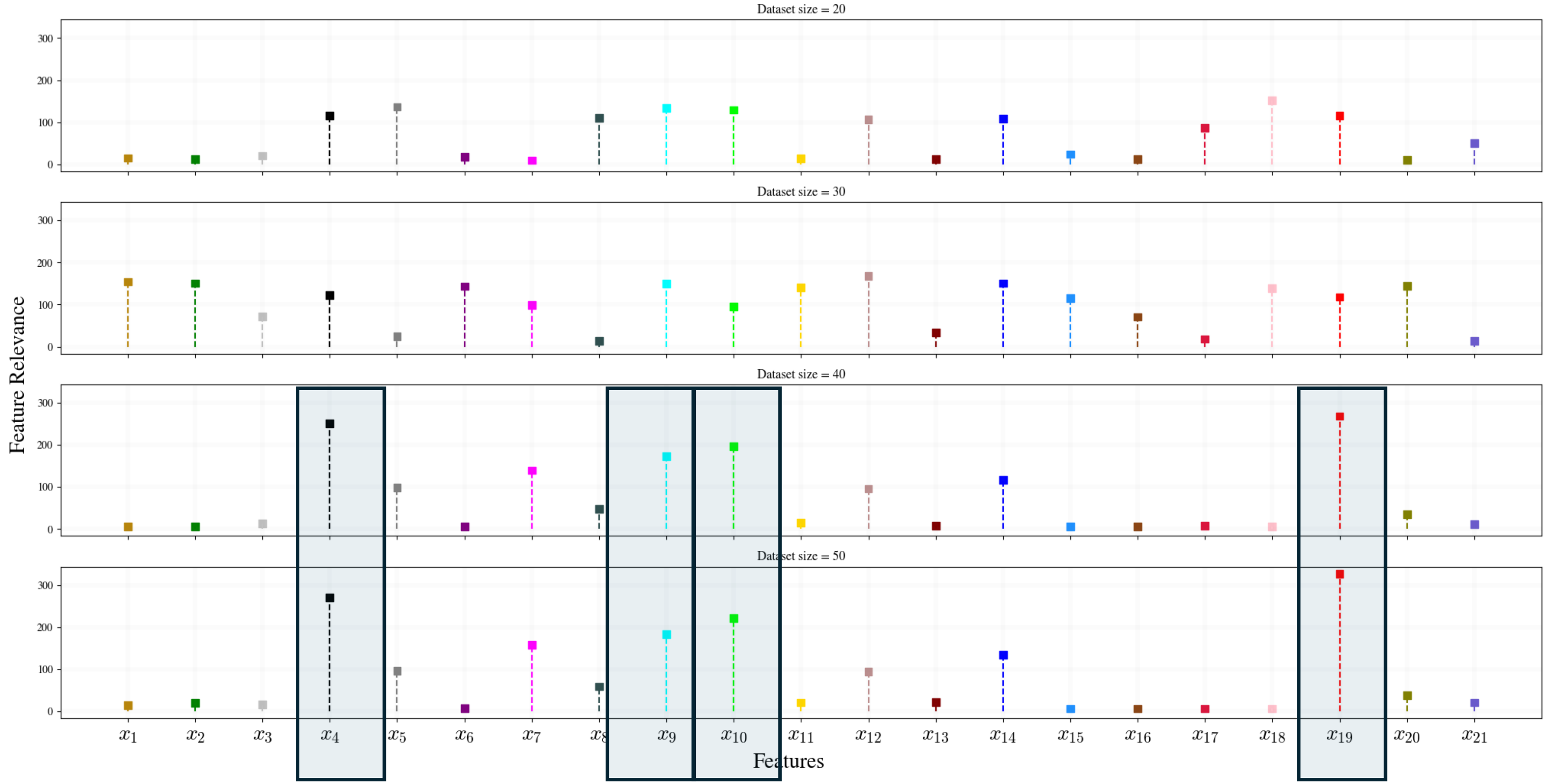
$$y = 12x_1 + 20x_2 - 13x_3 + 122x_4 - 44x_5 + 0.01x_6 - 77x_7 - 33x_8 + 90x_9 - 102x_{10} + 11x_{11} + 44x_{12} + 11x_{13} - 66x_{14} + 8x_{15} - 0.001x_{16} + 0.05x_{17} - 0.8x_{18} + 145x_{19} - 22x_{20} + 13x_{21}$$



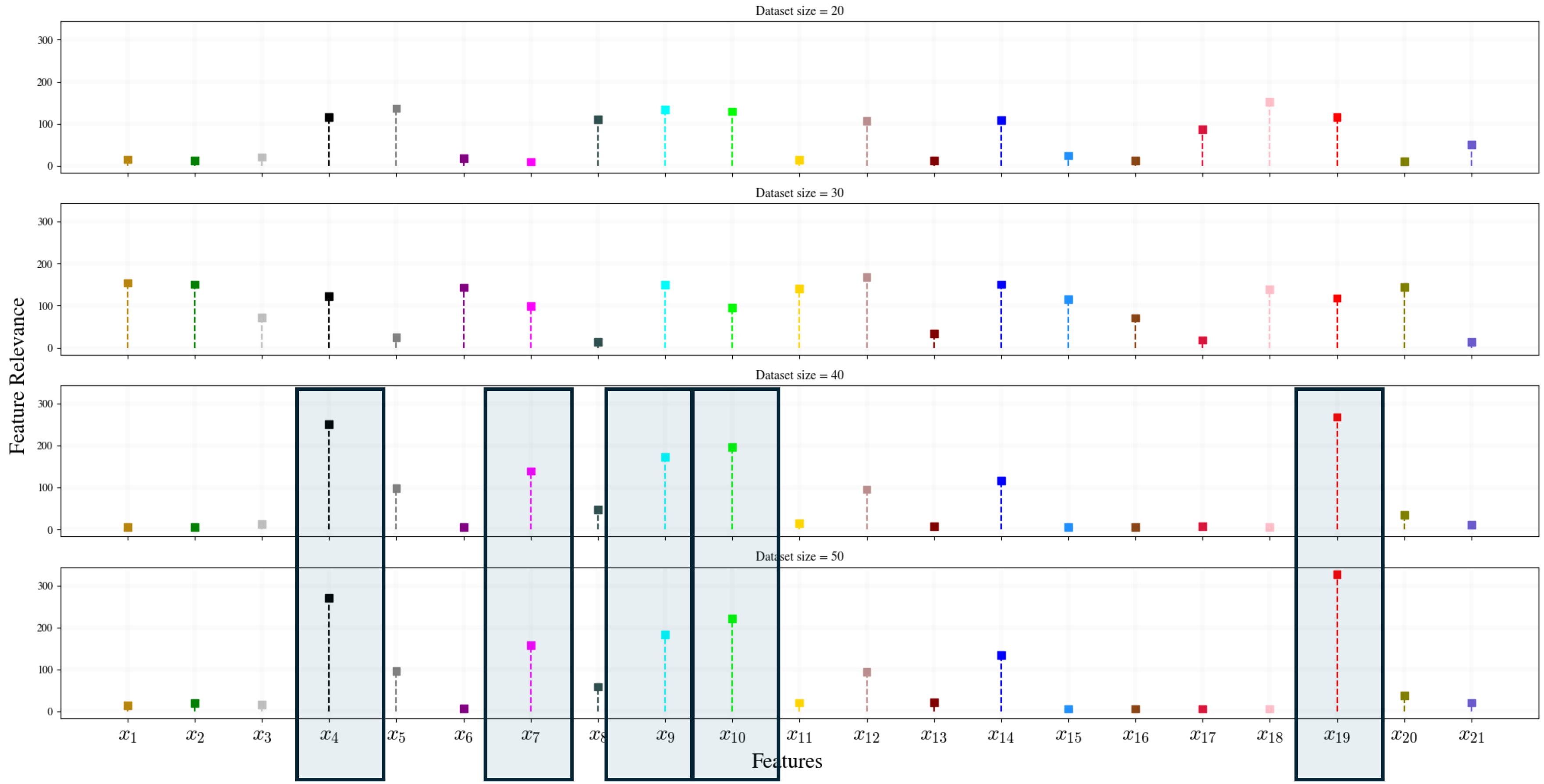
$$y = 12x_1 + 20x_2 - 13x_3 + 122x_4 - 44x_5 + 0.01x_6 - 77x_7 - 33x_8 + 90x_9 - 102x_{10} + 11x_{11} + 44x_{12} + 11x_{13} - 66x_{14} + 8x_{15} - 0.001x_{16} + 0.05x_{17} - 0.8x_{18} + 145x_{19} - 22x_{20} + 13x_{21}$$



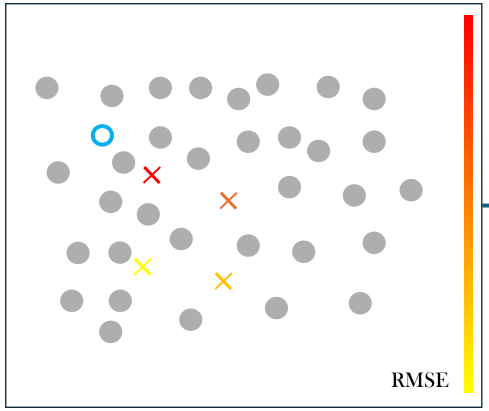
$$y = 12x_1 + 20x_2 - 13x_3 + 122x_4 - 44x_5 + 0.01x_6 - 77x_7 - 33x_8 + 90x_9 - 102x_{10} + 11x_{11} + 44x_{12} + 11x_{13} - 66x_{14} + 8x_{15} - 0.001x_{16} + 0.05x_{17} - 0.8x_{18} + 145x_{19} - 22x_{20} + 13x_{21}$$



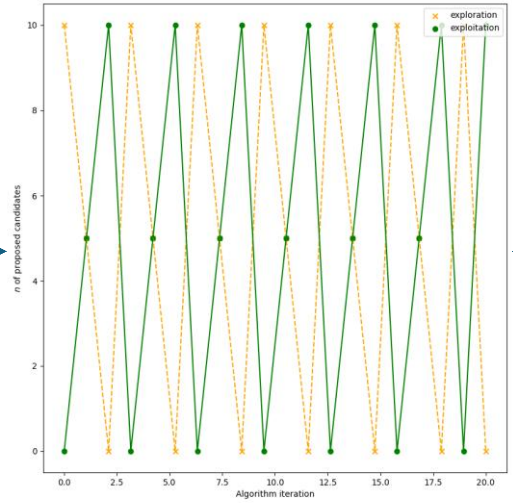
$$y = 12x_1 + 20x_2 - 13x_3 + 122x_4 - 44x_5 + 0.01x_6 - 77x_7 - 33x_8 + 90x_9 - 102x_{10} + 11x_{11} + 44x_{12} + 11x_{13} - 66x_{14} + 8x_{15} - 0.001x_{16} + 0.05x_{17} - 0.8x_{18} + 145x_{19} - 22x_{20} + 13x_{21}$$



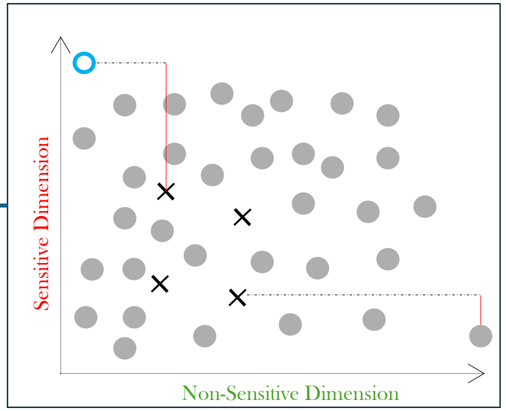
Surrogate models guided Exploitation



$$x^{m+1} = \text{argmax}_{x*\in C} \sum_{i=1}^{m} e_{\text{loocv}} \exp\left(-||x^i - x^*||\right),$$
$$\text{s.t } \min_{i\in m} ||xi - x^*|| \geq S_{th}$$



Sensitivity analysis guided Exploration



$$x^{m+1} = \arg\max_{x*\in C} \left(min_{i\in m} w^T || x^i - x^*||\right)$$

$X_1^* X_2^* X_3^* \; ..... \; X_d^*$

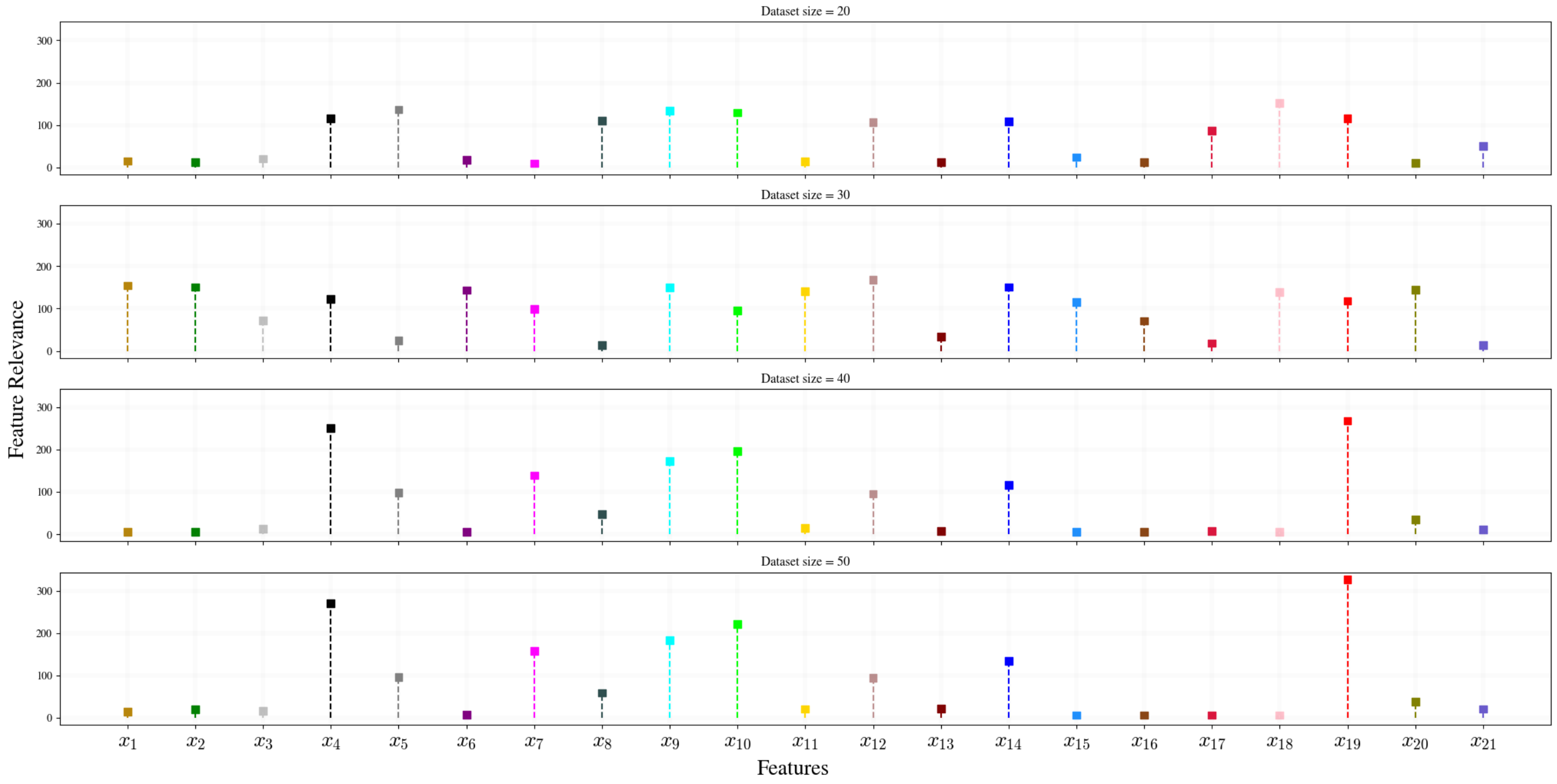
Surrogate Model	Surrogate Model	… … … … …	Surrogate Model

$y \cdot y \cdot y \cdot$

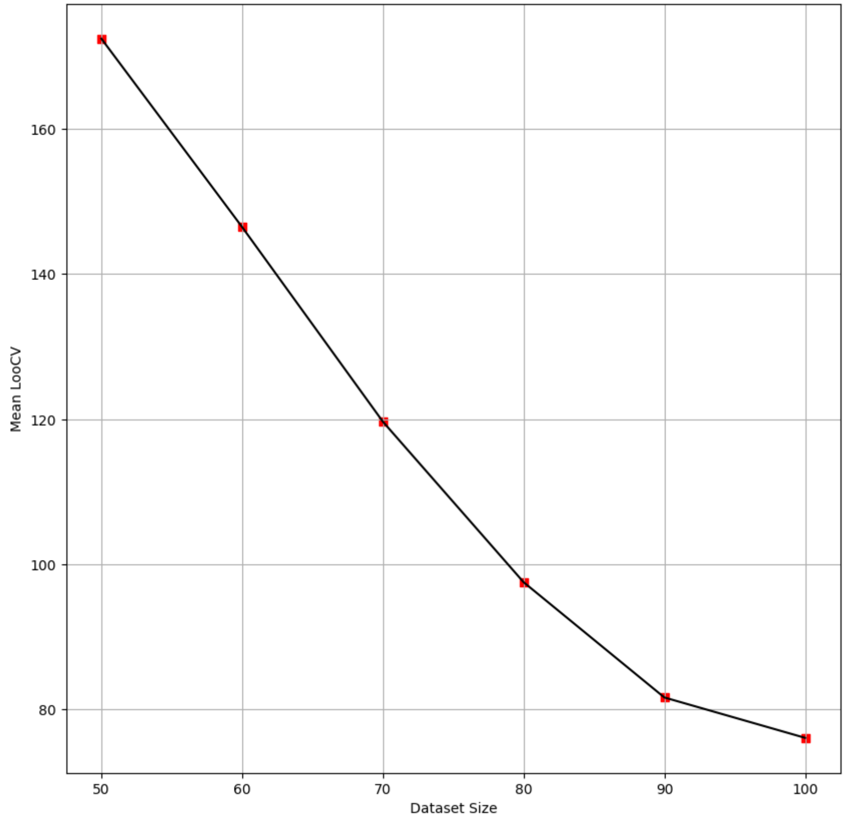
Estimate RMSE

Check Convergence

Pure Exploration



Exploitation with wise Exploration



Problem | Approaches | Powerful Techniques | Proposed Algorithm