2019/10/9

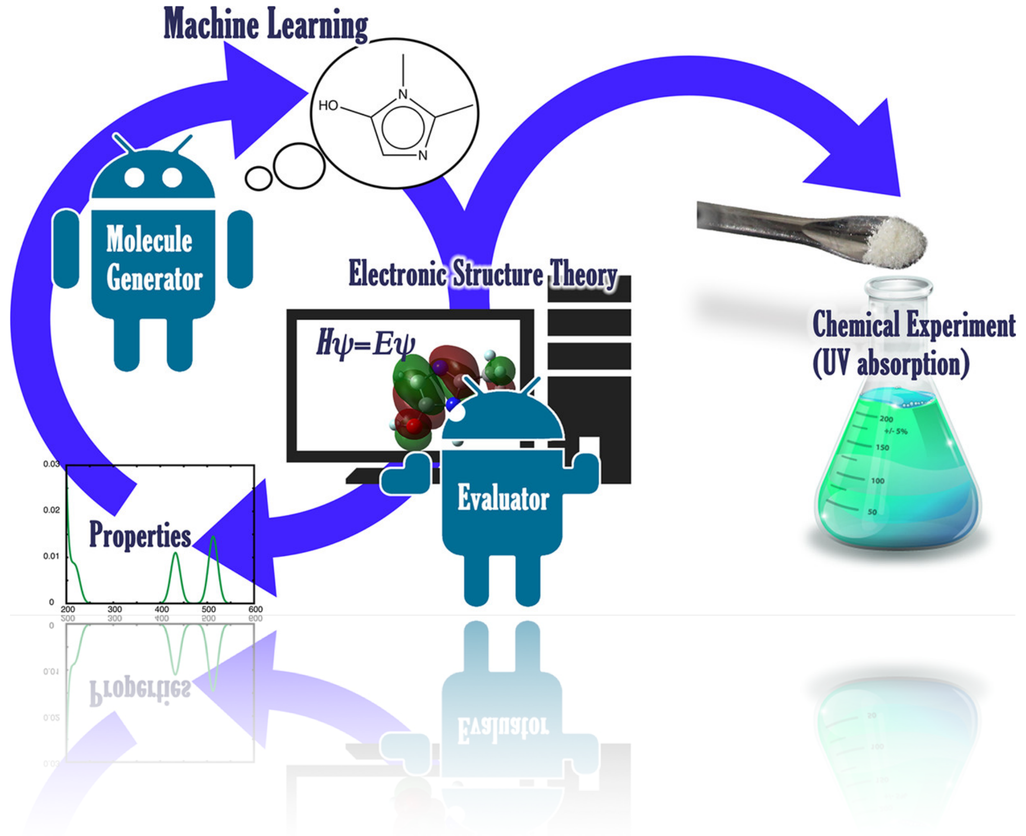
The 3rd Innovation Camp for Computational Materials Science

# Effective model estimation by machine learning and Bayesian optimization

NIMS/U. Tokyo Ryo Tamura

# Current my works

## Organic

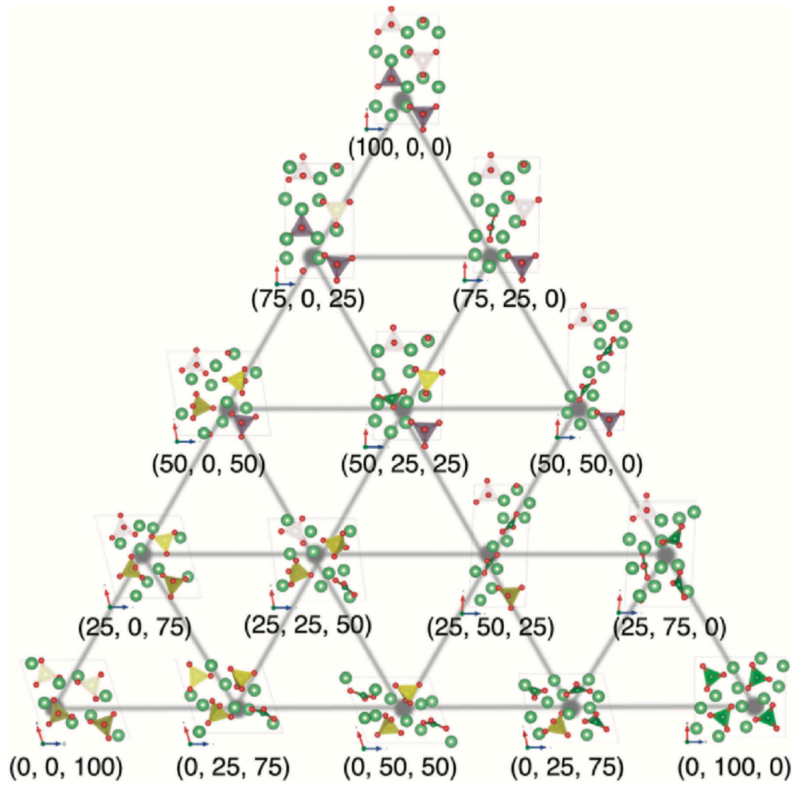



ACS Cent. Sci. 4, 1126 (2018)

## Smells Sensor

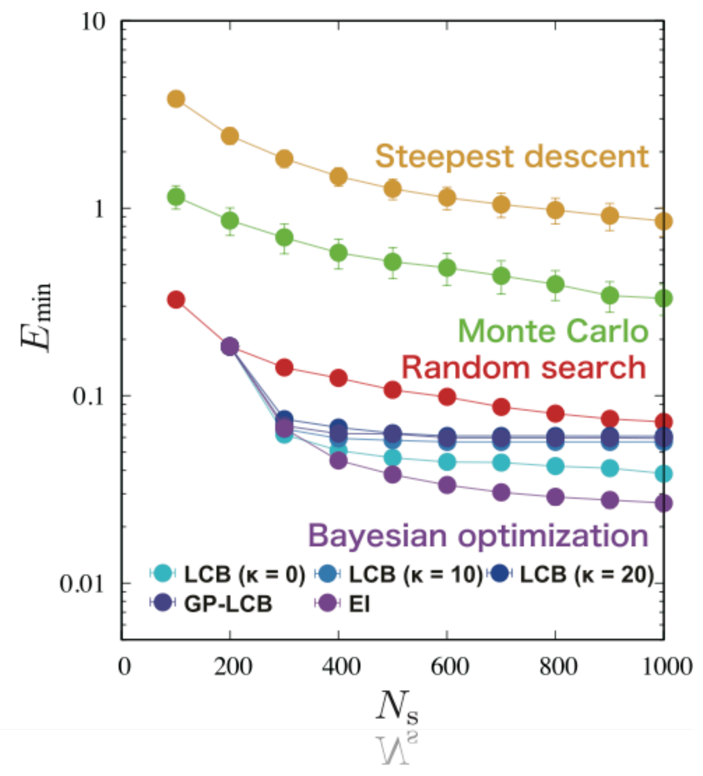

Sci. Rep. 7, 3661 (2017)
ACS Sensors 3, 1592 (2018)

## Li-ion conductivity



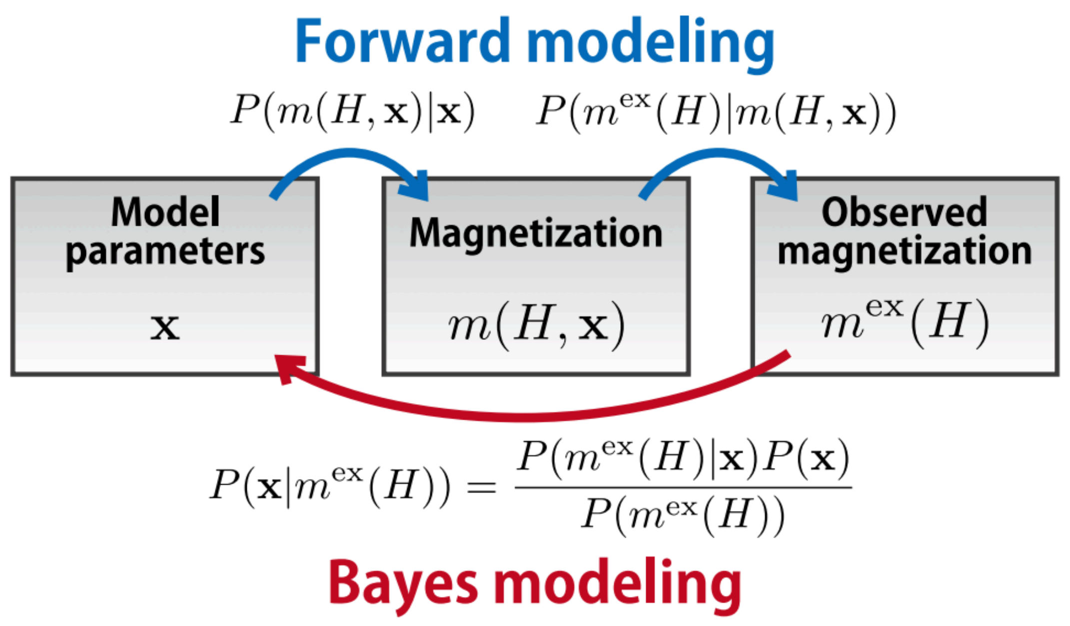

BCSJ 92, 1100 (2019)

## Bayesian Opt.



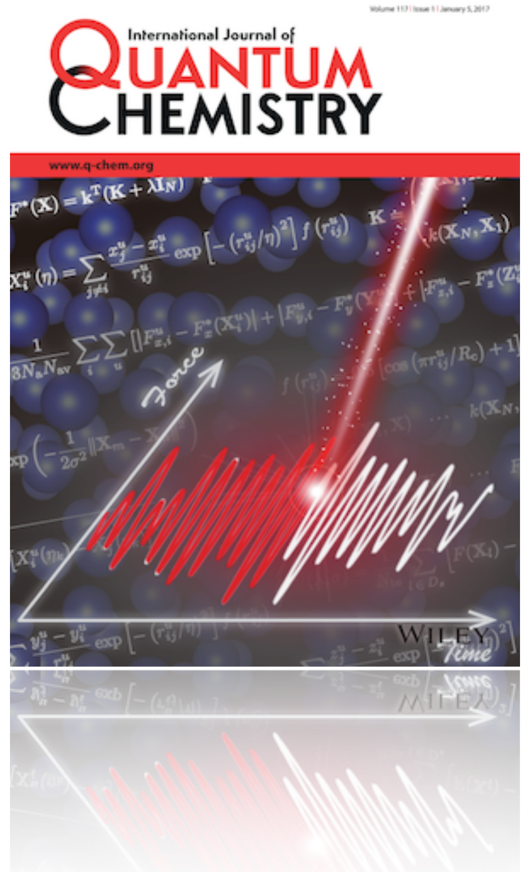

PLoS ONE 13, e0193785 (2018)

## Magnet



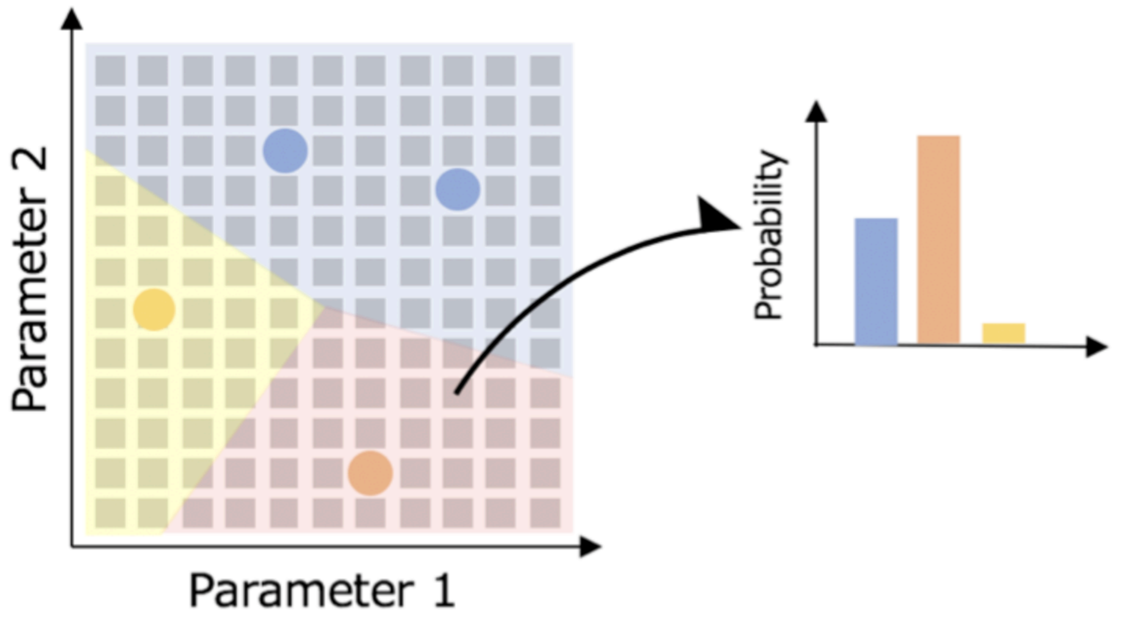

Phys. Rev. B 95, 064407 (2017)

## Atomic Force



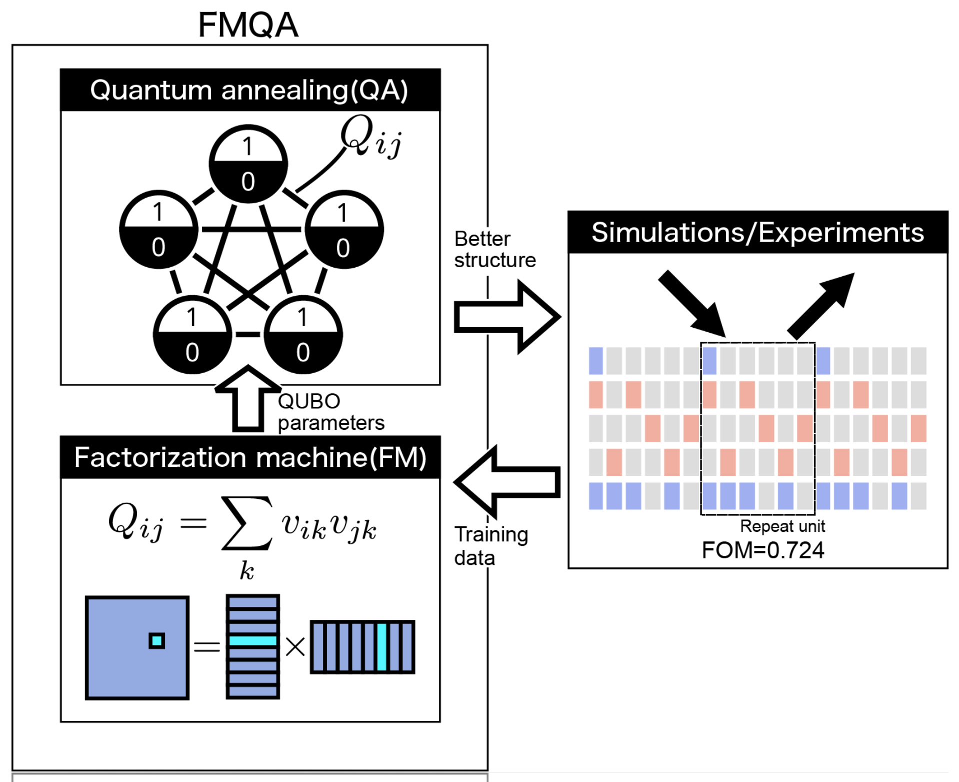

IJQC. 117, 33 (2017)
JPSJ. 88, 044601 (2019)

## Phase Diagram




Phys. Rev. Mat. 3, 033802 (2019)

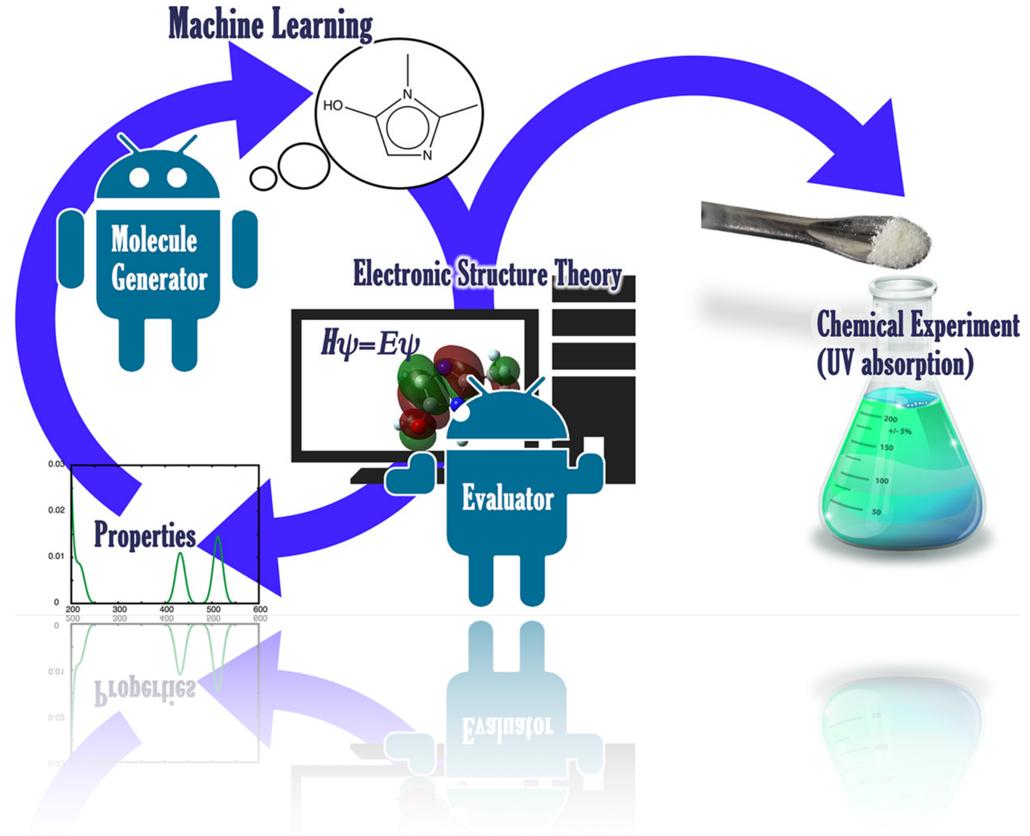
## Metamaterial




arXiv: 1902.06573 (2019)

# Current my works

## Organic

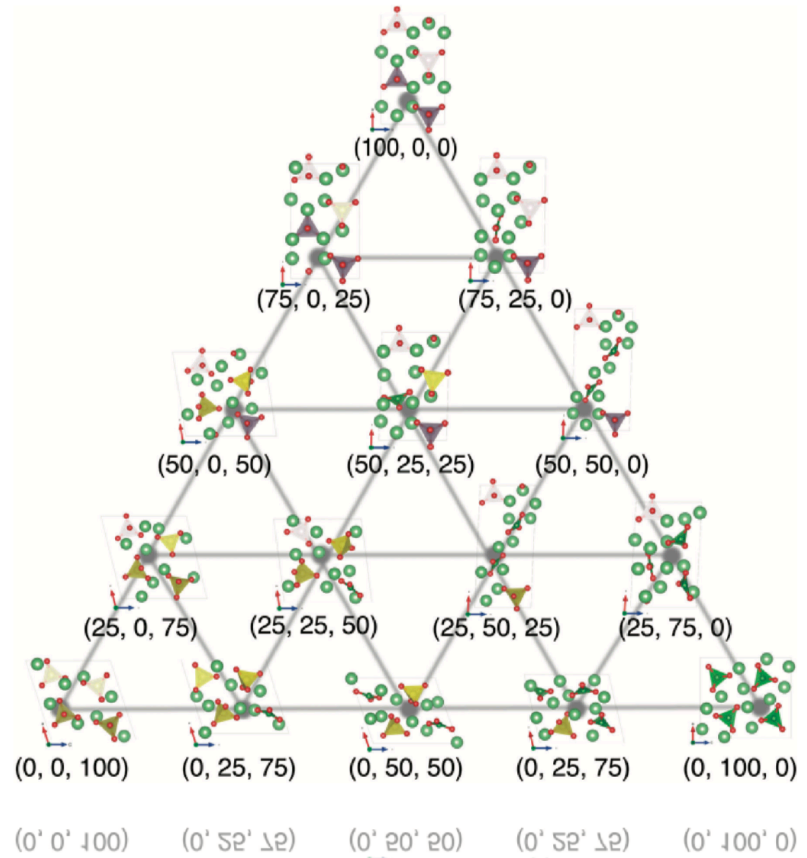



ACS Cent. Sci. 4, 1126 (2018)

## Smells Sensor

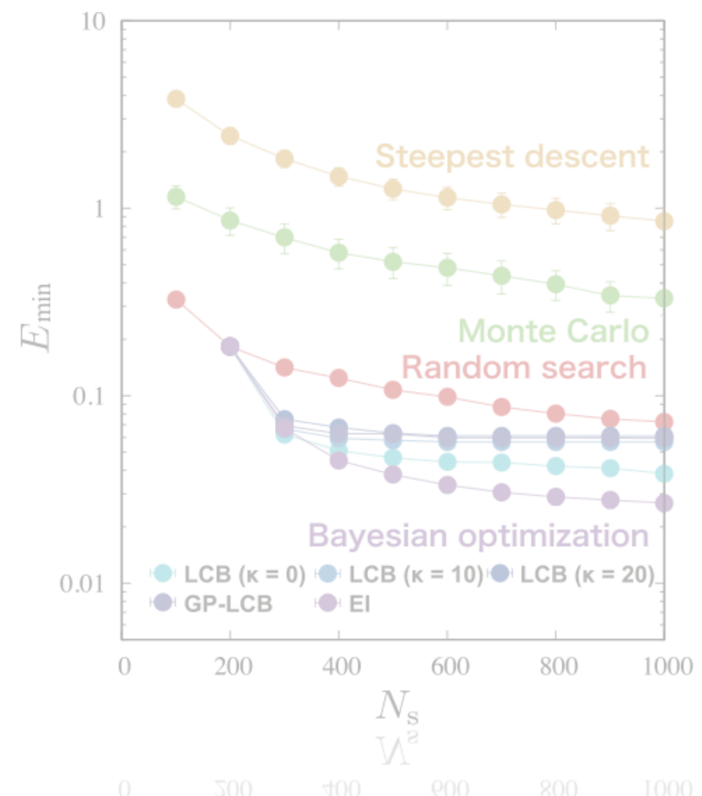

Sci. Rep. 7, 3661 (2017)
ACS Sensors 3, 1592 (2018)

## Li-ion conductivity



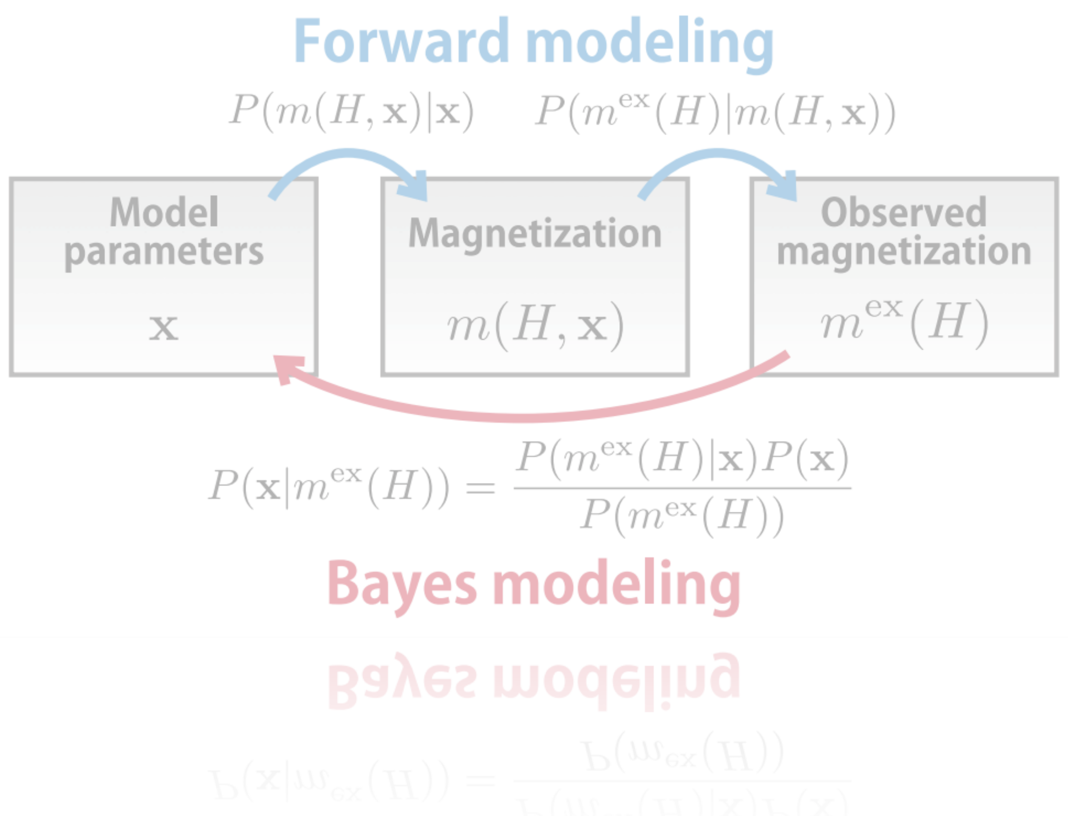

BCSJ 92, 1100 (2019)

## Bayesian Opt.



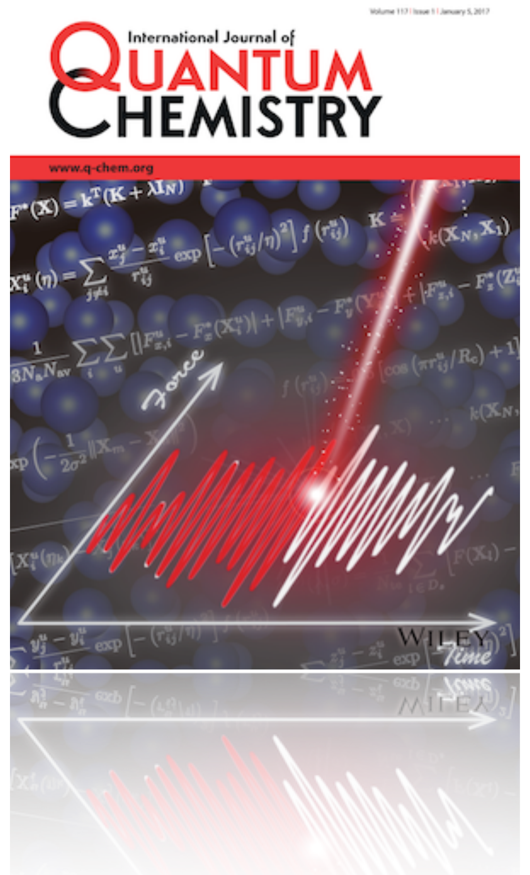

PLoS ONE 13, e0193785 (2018)

## Magnet



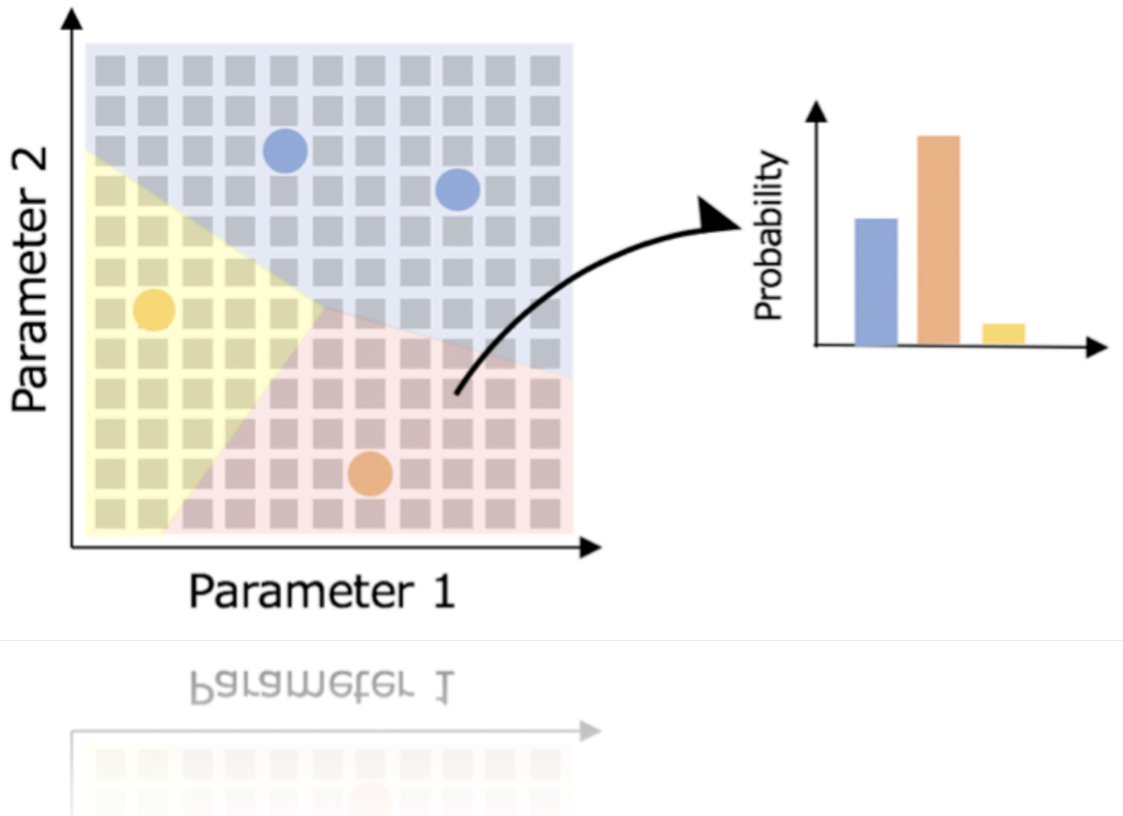

Phys. Rev. B 95, 064407 (2017)

## Atomic Force



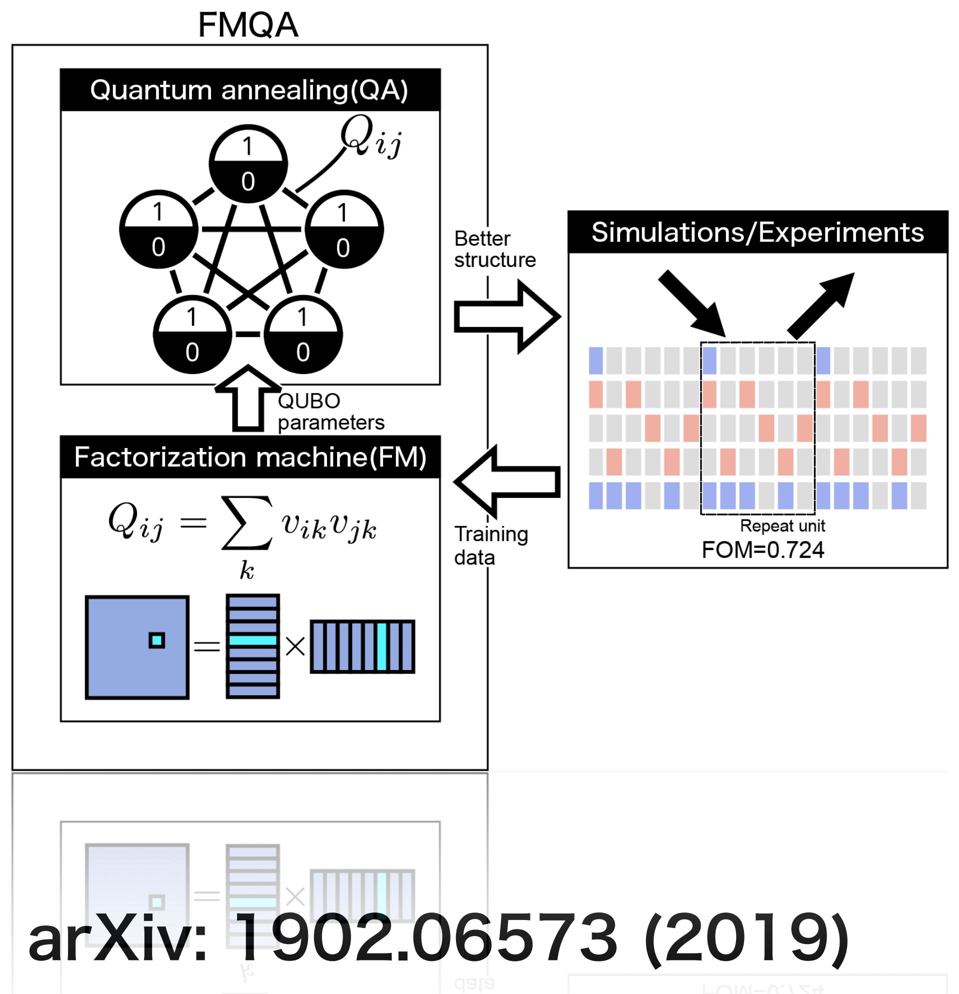

IJQC. 117, 33 (2017)
JPSJ. 88, 044601 (2019)

## Phase Diagram




Phys. Rev. Mat. 3, 033802 (2019)

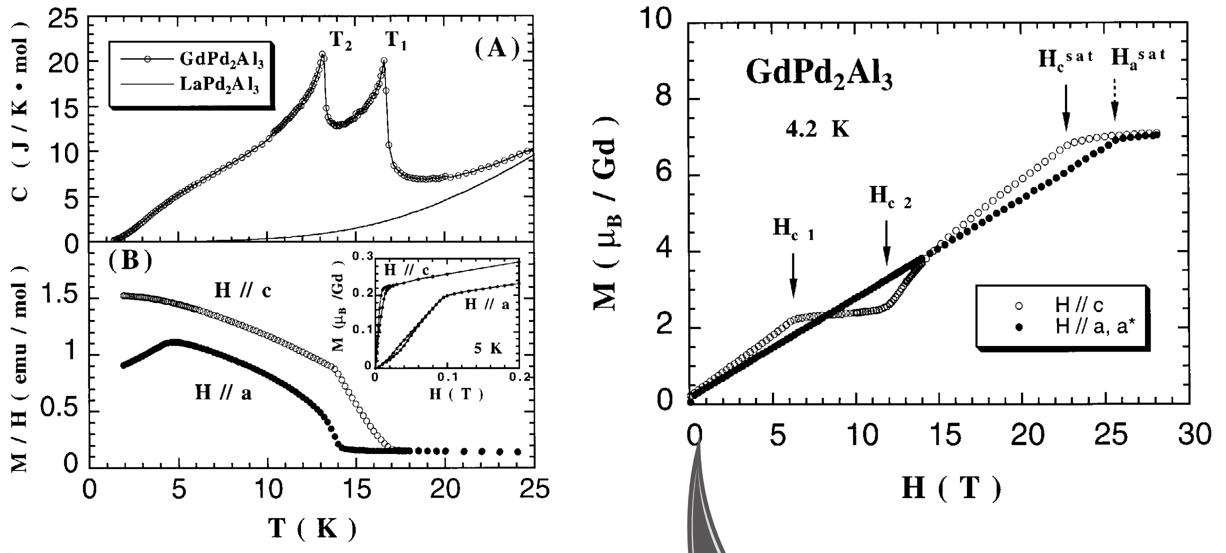
## Metamaterial




arXiv: 1902.06573 (2019)

# Motivation

## Experimental results




## Candidate models

$$b_{ij}(\mathbf{s}_i\cdot\mathbf{s}_j)^2$$

$$J_{ij}\mathbf{s}_i\cdot\mathbf{s}_j \quad \mathbf{d}_{ij}\cdot[\mathbf{s}_i\times\mathbf{s}_j] \quad D_i(s_i^z)^2$$

$$\frac{\mathbf{s}_i\cdot\mathbf{s}_j}{r_{ij}^3}-3\frac{(\mathbf{s}_i\cdot\mathbf{r}_{ij})(\mathbf{s}_j\cdot\mathbf{r}_{ij})}{r_{ij}^5}$$

input

input

## Automatic model estimation

***Bayesian statistics***

L1 regularization
L2 regularization
Full search
+
Cross validation
Elbow method

output

**Plausible effective model for experimental results**

# Forward modeling and Bayes modeling

e.g. magnetization is target input property

## Forward modeling

$P(m(H,\mathbf{x})|\mathbf{x})$

$P(m^{\mathrm{ex}}(H)|m(H,\mathbf{x}))$

$\mathbf{x} =$ {model parameters}

$$\mathcal{H} = -\sum_{i,j} J_{ij}\mathbf{s}_i \cdot \mathbf{s}_j + \cdots$$

magnetization

$$m(H,\mathbf{x})$$

observed magnetization

$$m^{\mathrm{ex}}(H)$$

$$P(\mathbf{x}|m^{\mathrm{ex}}(H)) = \frac{P(m^{\mathrm{ex}}(H)|\mathbf{x})P(\mathbf{x})}{P(m^{\mathrm{ex}}(H))}$$

## Bayes modeling

$P(B|A)$ : Conditional probability of event $B$ given event $A$
(Posterior distribution)

# Forward modeling and Bayes modeling

## Forward modeling

$P(m(H,\mathbf{x})|\mathbf{x})$ $\quad$ $P(m^{\mathrm{ex}}(H)|m(H,\mathbf{x}))$

$\mathbf{x} =$ {model parameters} → magnetization → observed magnetization

$\mathcal{H} = -\sum_{i,j} J_{ij}\mathbf{s}_i \cdot \mathbf{s}_j + \cdots$ → $m(H,\mathbf{x})$ → $m^{\mathrm{ex}}(H)$

Definition of magnetization as thermal average of spins

$$\langle \mathbf{s}_i \rangle_{H,\mathbf{x}} = \frac{\mathrm{Tr}\mathbf{s}_i e^{-\beta\mathcal{H}}}{\mathrm{Tr}e^{-\beta\mathcal{H}}} \longrightarrow m(H,\mathbf{x}) = \left| \frac{1}{N|\mathbf{s}|}\sum_{i=1}^{N} \langle \mathbf{s}_i \rangle_{H,\mathbf{x}} \right|$$

Conditional probability of $m(H,\mathbf{x})$ given $\mathbf{x}$

$$P(m(H,\mathbf{x})|\mathbf{x}) = \delta\left( m(H,\mathbf{x}) - \left| \frac{1}{N|\mathbf{s}|}\sum_{i=1}^{N} \langle \mathbf{s}_i \rangle_{H,\mathbf{x}} \right| \right)$$

Magnetization is uniquely obtained when the model parameters are given.

# Observation noise

## Forward modeling

$P(m(H,\mathbf{x})|\mathbf{x})$ $\quad$ $P(m^{\mathrm{ex}}(H)|m(H,\mathbf{x}))$

$\mathbf{x} =$ {model parameters}

$$\mathcal{H} = -\sum_{i,j} J_{ij}\mathbf{s}_i \cdot \mathbf{s}_j + \cdots$$

magnetization

$$m(H,\mathbf{x})$$

observed magnetization

$$m^{\mathrm{ex}}(H)$$

Existence of observation noise in $m^{\mathrm{ex}}(H)$

$$m^{\mathrm{ex}}(H) = m(H,\mathbf{x}) \underbrace{+\,\varepsilon}_{\text{observation noise}}$$

Assumption : $P(\varepsilon) \propto \exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right)$

Conditional probability of $m^{\mathrm{ex}}(H)$ given $m(H,\mathbf{x})$

$$P(m^{\mathrm{ex}}(H)|m(H,\mathbf{x})) \propto \exp\left(-\frac{1}{2\sigma^2}(m^{\mathrm{ex}}(H) - m(H,\mathbf{x}))^2\right)$$

# Conditional probability

## Forward modeling

$P(m(H,\mathbf{x})|\mathbf{x})$ $\quad$ $P(m^{\mathrm{ex}}(H)|m(H,\mathbf{x}))$

$\mathbf{x}=$ {model parameters}

$$\mathcal{H}=-\sum_{i,j}J_{ij}\mathbf{s}_i\cdot\mathbf{s}_j+\cdots$$

magnetization

$$m(H,\mathbf{x})$$

observed magnetization

$$m^{\mathrm{ex}}(H)$$

Conditional probability of $m^{\mathrm{ex}}(H)$ given $\mathbf{x}$

$$P(m^{\mathrm{ex}}(H)|\mathbf{x})\propto\int dm(H,\mathbf{x})P(m^{\mathrm{ex}}(H)|m(H,\mathbf{x}))P(m(H,\mathbf{x})|\mathbf{x})$$

$$\propto\exp\left[-\frac{1}{2\sigma^2}\left(m^{\mathrm{ex}}(H)-\left|\frac{1}{N|\mathbf{s}|}\sum_{i=1}^{N}\langle\mathbf{s}_i\rangle_{H,\mathbf{x}}\right|\right)^2\right]$$

$m^{\mathrm{ex}}(H)$ where $P(m^{\mathrm{ex}}(H)|\mathbf{x})$ is maximize. $\rightarrow$ Observed magnetization

# Bayes modeling

## Forward modeling

$P(m(H,\mathbf{x})|\mathbf{x})$ $\quad$ $P(m^{\mathrm{ex}}(H)|m(H,\mathbf{x}))$

$\mathbf{x} =$ {model parameters}

$$\mathcal{H} = -\sum_{i,j} J_{ij}\mathbf{s}_i \cdot \mathbf{s}_j + \cdots$$

magnetization

$$m(H,\mathbf{x})$$

observed magnetization

$$m^{\mathrm{ex}}(H)$$

$$P(\mathbf{x}|m^{\mathrm{ex}}(H)) = \frac{P(m^{\mathrm{ex}}(H)|\mathbf{x})P(\mathbf{x})}{P(m^{\mathrm{ex}}(H))}$$

: Posterior distribution

## Bayes modeling

$\mathbf{x}$ where $P(\mathbf{x}|m^{\mathrm{ex}}(H))$ is maximize. → Plausible model parameters

# Summary of effective model estimation

**We search the maximizer of the posterior distribution when the measured physical quantities are inputted.**

Posterior distribution

$$P(\underbrace{\mathbf{x}}_{\text{Model parameters}}|\{y^{\mathrm{ex}}(g_l)\}_{l=1,\cdots,L}) = \exp[-E(\mathbf{x})]$$

Energy function

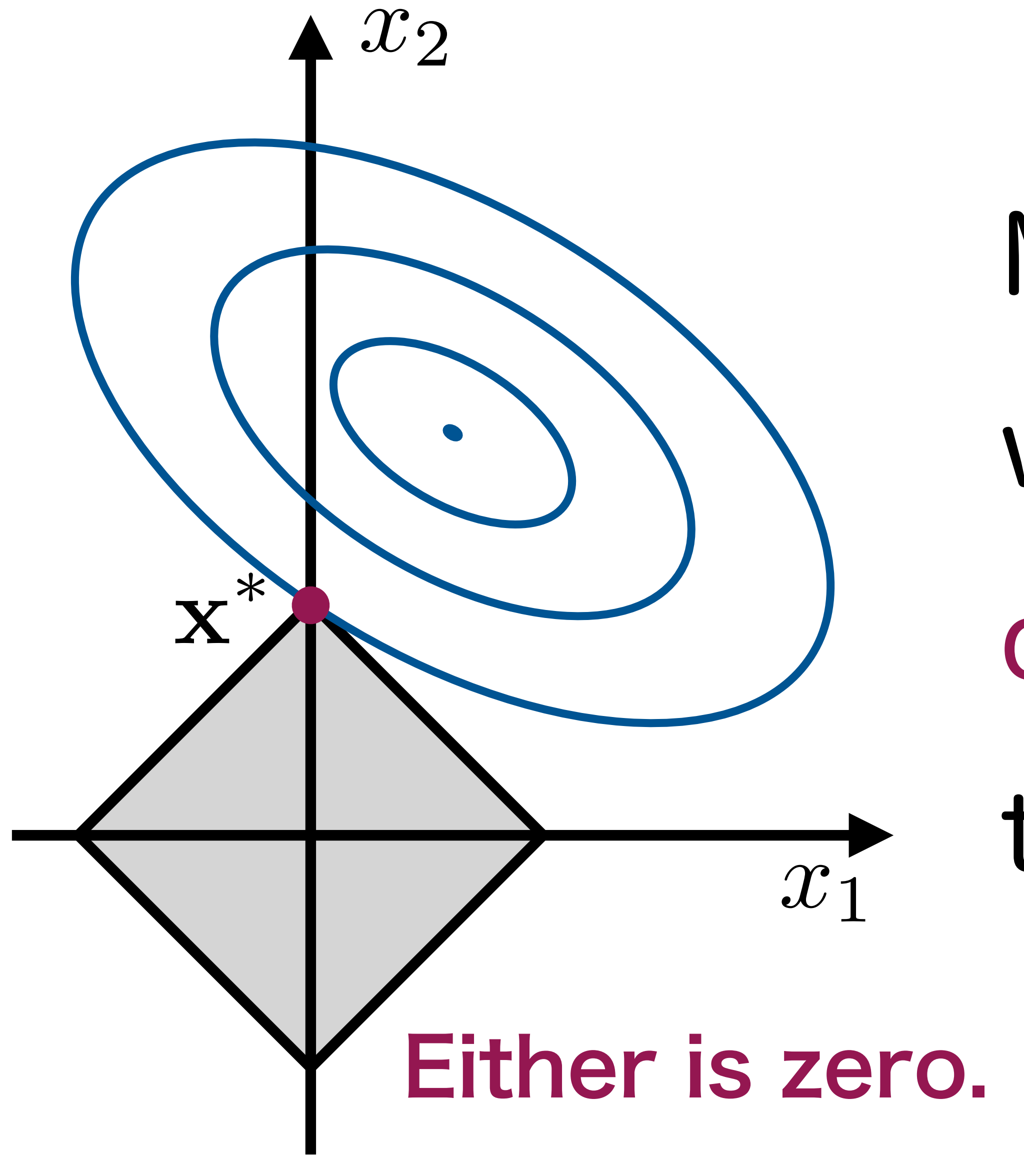
$$E(\mathbf{x}) = \frac{1}{2\sigma^2}\sum_{l=1}^{L}\Big[\underbrace{y^{\mathrm{ex}}(g_l)}_{\text{Input physical quantities}} - \underbrace{y^{\mathrm{cal}}(g_l,\mathbf{x})}_{\text{Calculated physical quantities by effective model}}\Big]^2 - \log \underbrace{P(\mathbf{x})}_{\text{Prior distribution}}$$

R. Tamura and K. Hukushima, Phys. Rev. B **95**, 064407 (2017).

# Prior distribution

- **L1 reguralization**

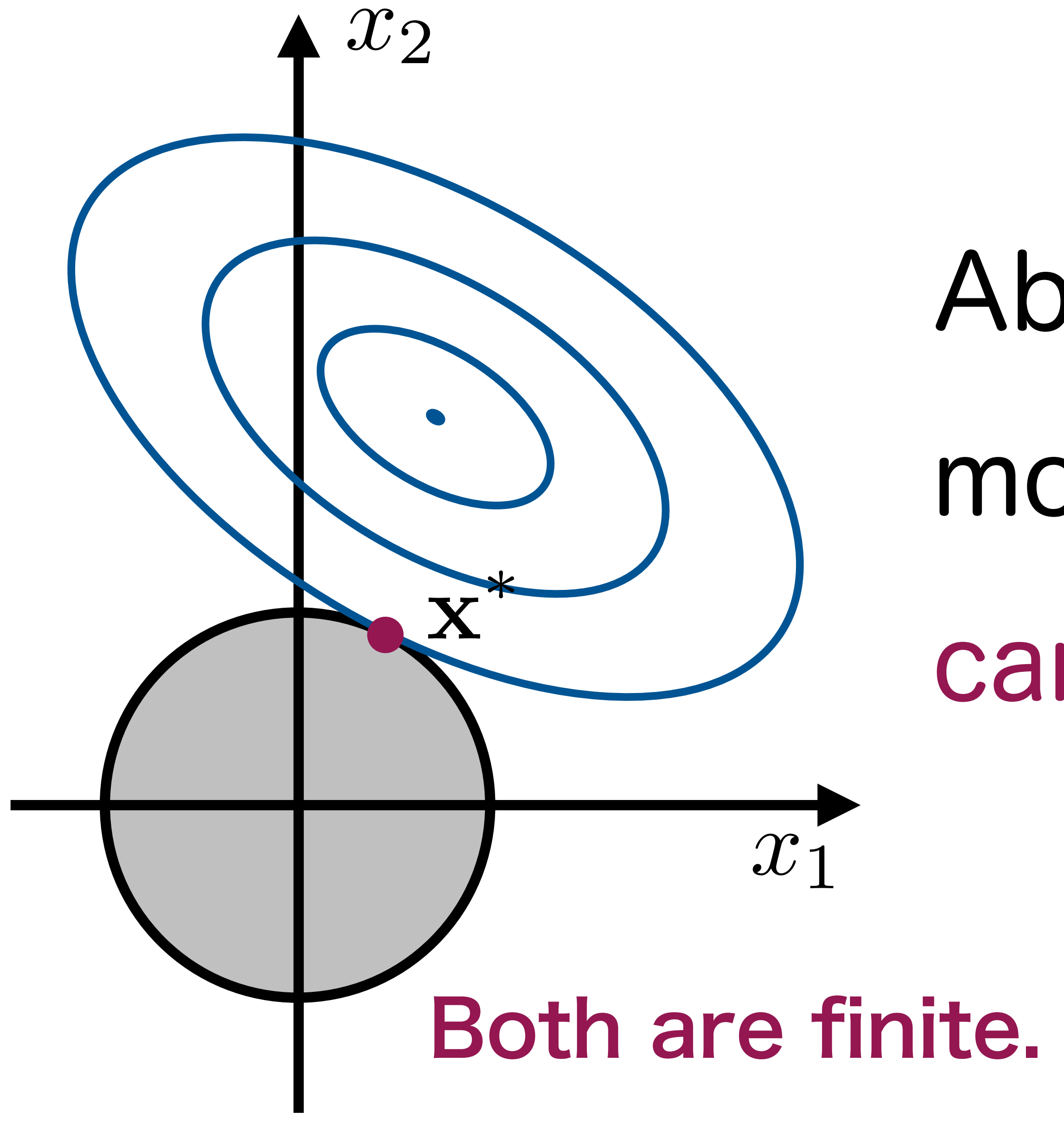
$$P(\mathbf{x}) \propto \exp\left(-\lambda|\mathbf{x}|\right)$$




Model parameters with large contributions can be selected based on the feature selection.

- **L2 reguralization**

$$P(\mathbf{x}) \propto \exp\left(-\lambda||\mathbf{x}||^2\right)$$




Absolute values of model parameters can be suppressed.

**Depending on the situation, it is necessary to select a proper prior distribution.**

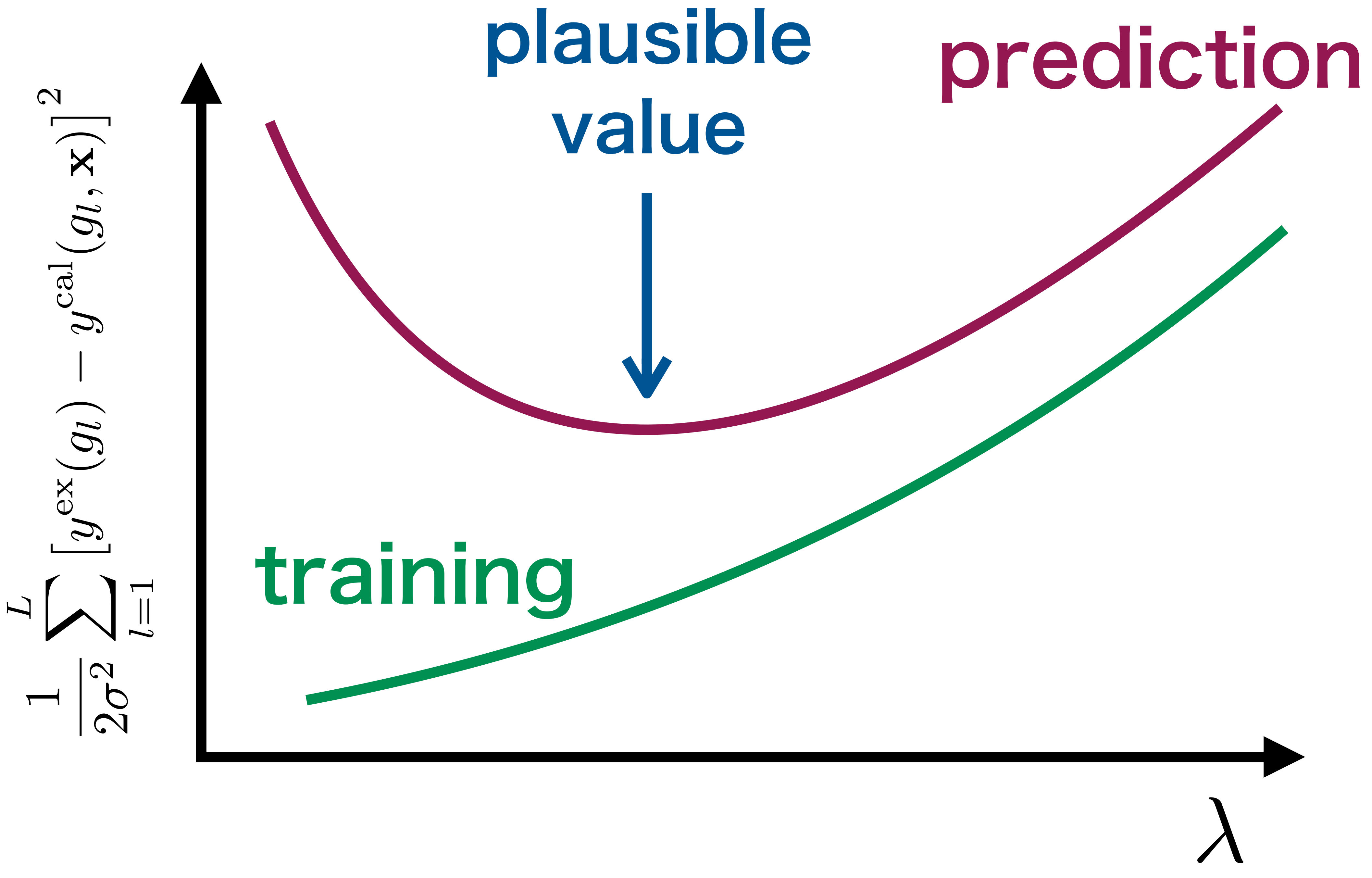
# Determination of hyperparameter

- **Cross validation**

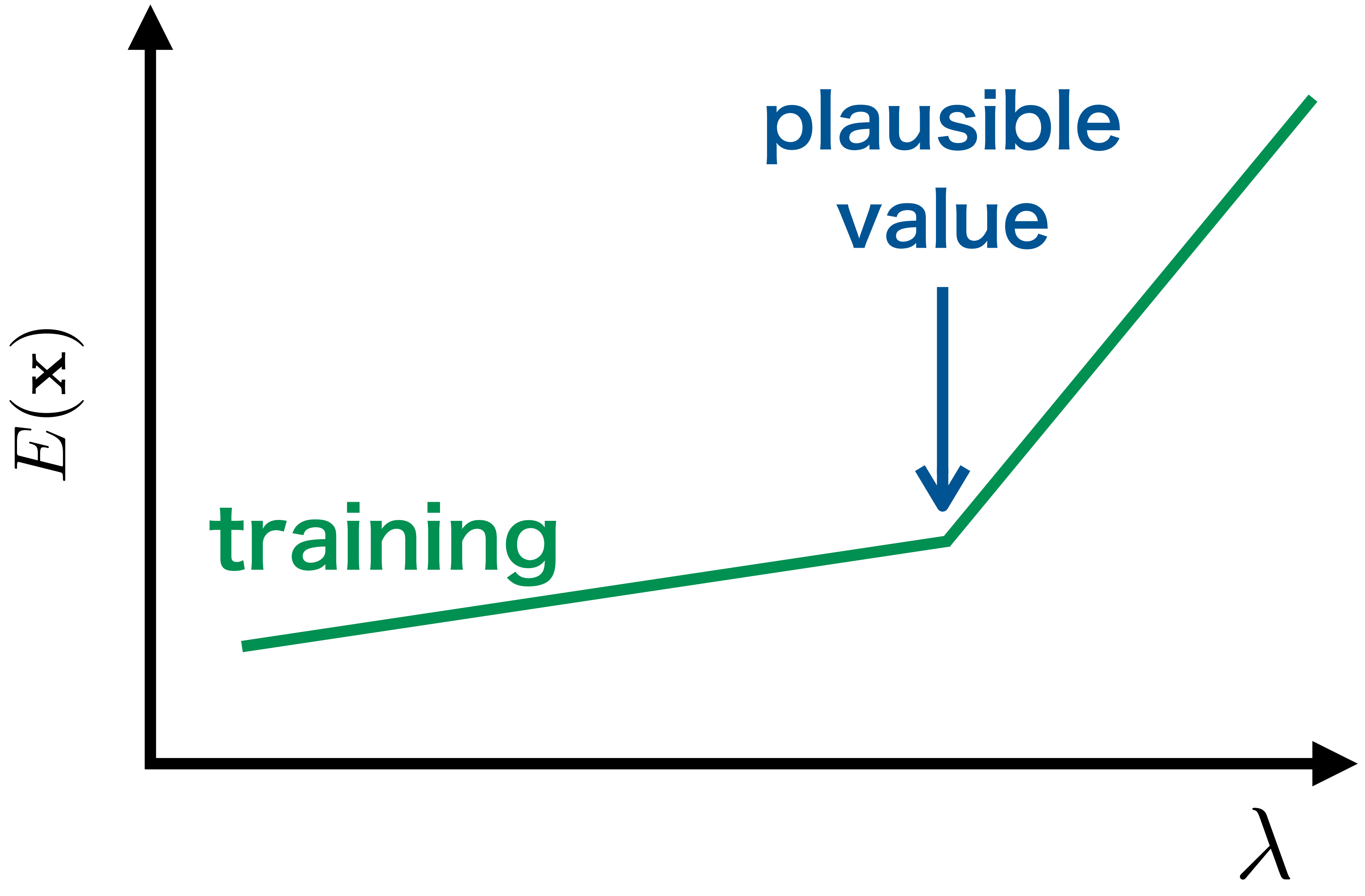
  Plausible value is determined at the point where prediction error is minimized.

- **Elbow method**

  Plausible value is determined by the large change point in energy function.




useful for L1 regularization
(overfitting occurs)




useful for L2 regularization
(overfitting does not occur)

# Demonstration 1: Theoretical model

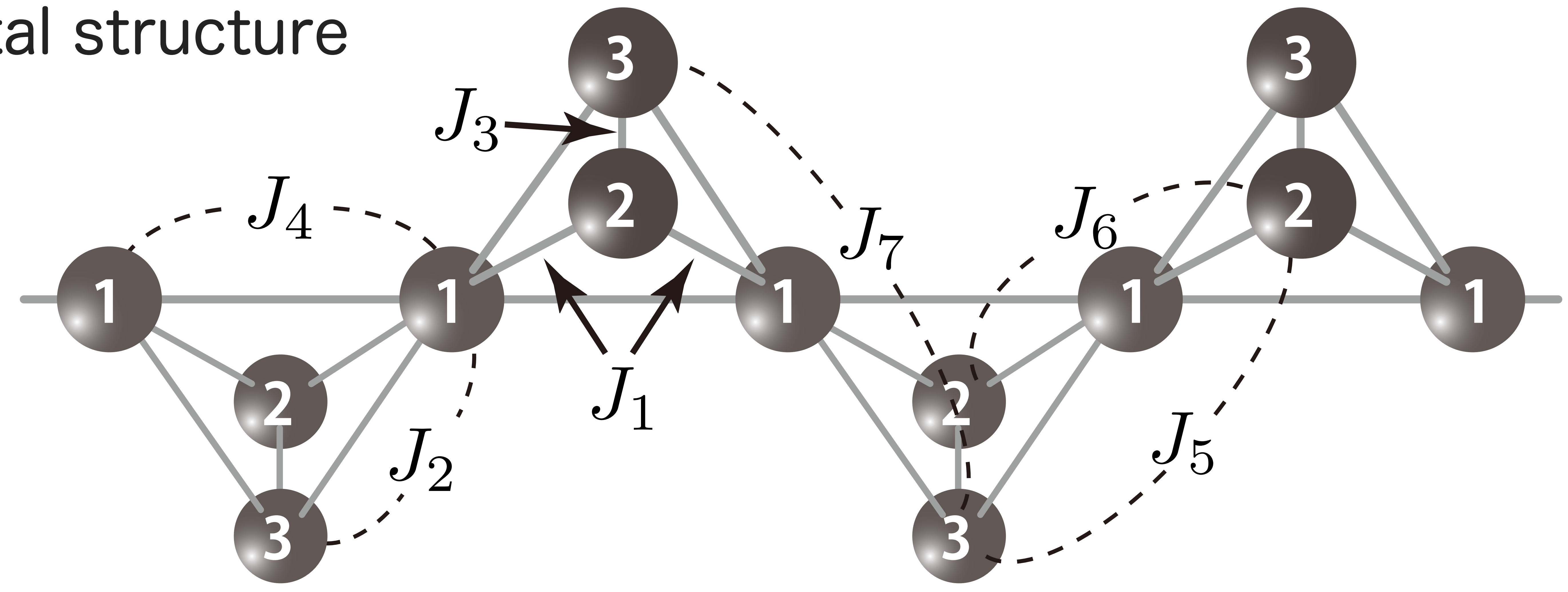
Target classical Heisenberg model with biquadratic interactions

(magnetization plateau is appeared)

$$\mathcal{H} = \sum_{\langle i,j \rangle} J_{ij} \left[ \mathbf{s}_i \cdot \mathbf{s}_j - b_{ij} (\mathbf{s}_i \cdot \mathbf{s}_j)^2 \right] - H \sum_i s_i^z \qquad b_{ij} = bJ_{ij}$$

$\mathbf{s}_i$ : Classical Heisenberg spin (S=1/2)

Crystal structure



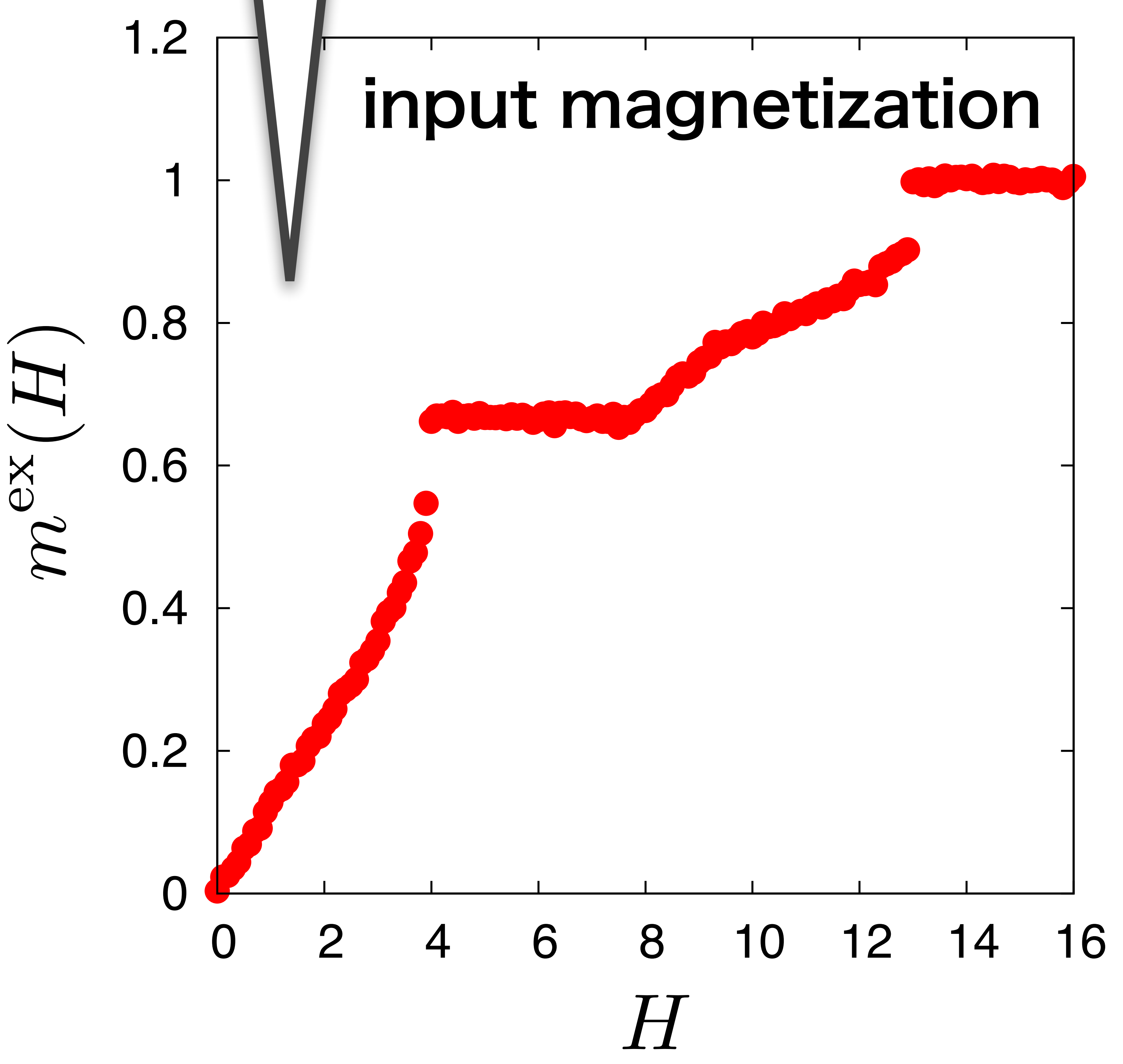

**We tried:**
**L1 regularization**
**&**
**cross validation**

model parameters : $\mathbf{x} = \{J_1, J_2, J_3, J_4, J_5, J_6, J_7, b\}$

# Estimation results

Zero temperature magnetization curve for $J_1 = 1, J_2 = 4, J_3 = 5, J_4 = 6, b = 0.1$, $\boxed{J_5 = J_6 = J_7 = 0}$ + Gaussian noise

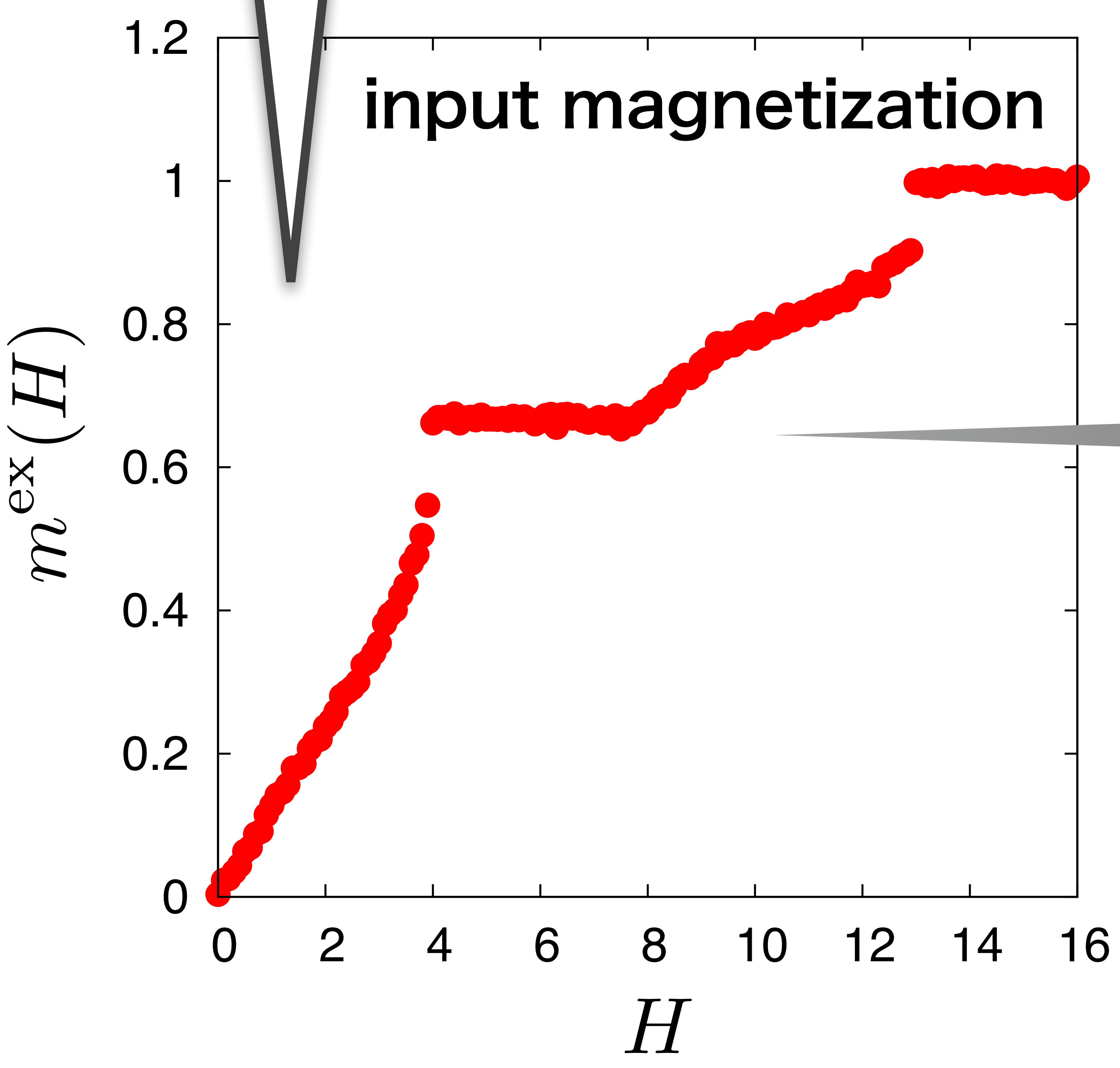
Magnetization is calculated by the steepest descent method.



# Estimation results

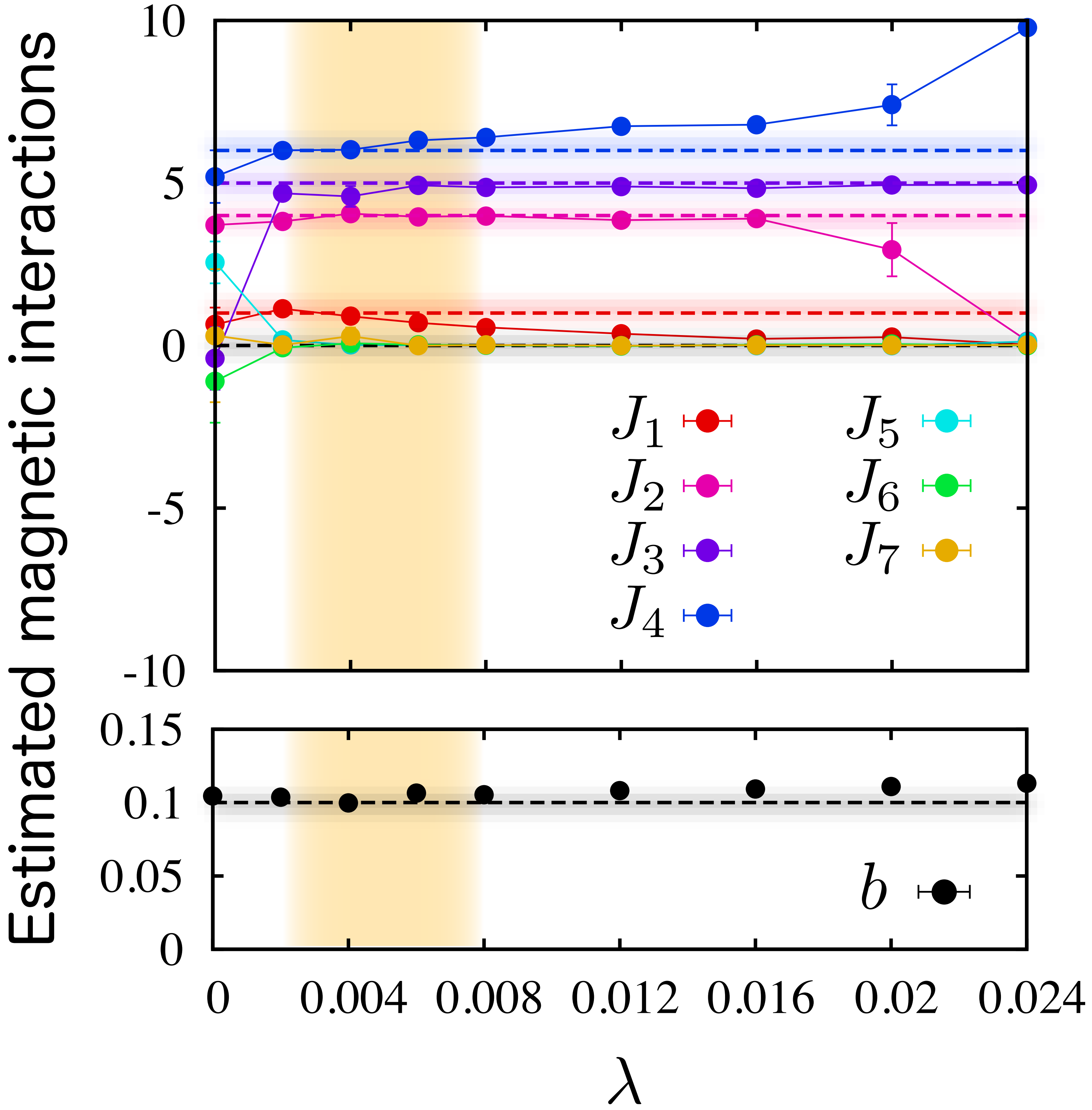
Zero temperature magnetization curve for $J_1 = 1, J_2 = 4, J_3 = 5, J_4 = 6, b = 0.1$ $\boxed{J_5 = J_6 = J_7 = 0}$ + Gaussian noise

Magnetization is calculated by the steepest descent method.

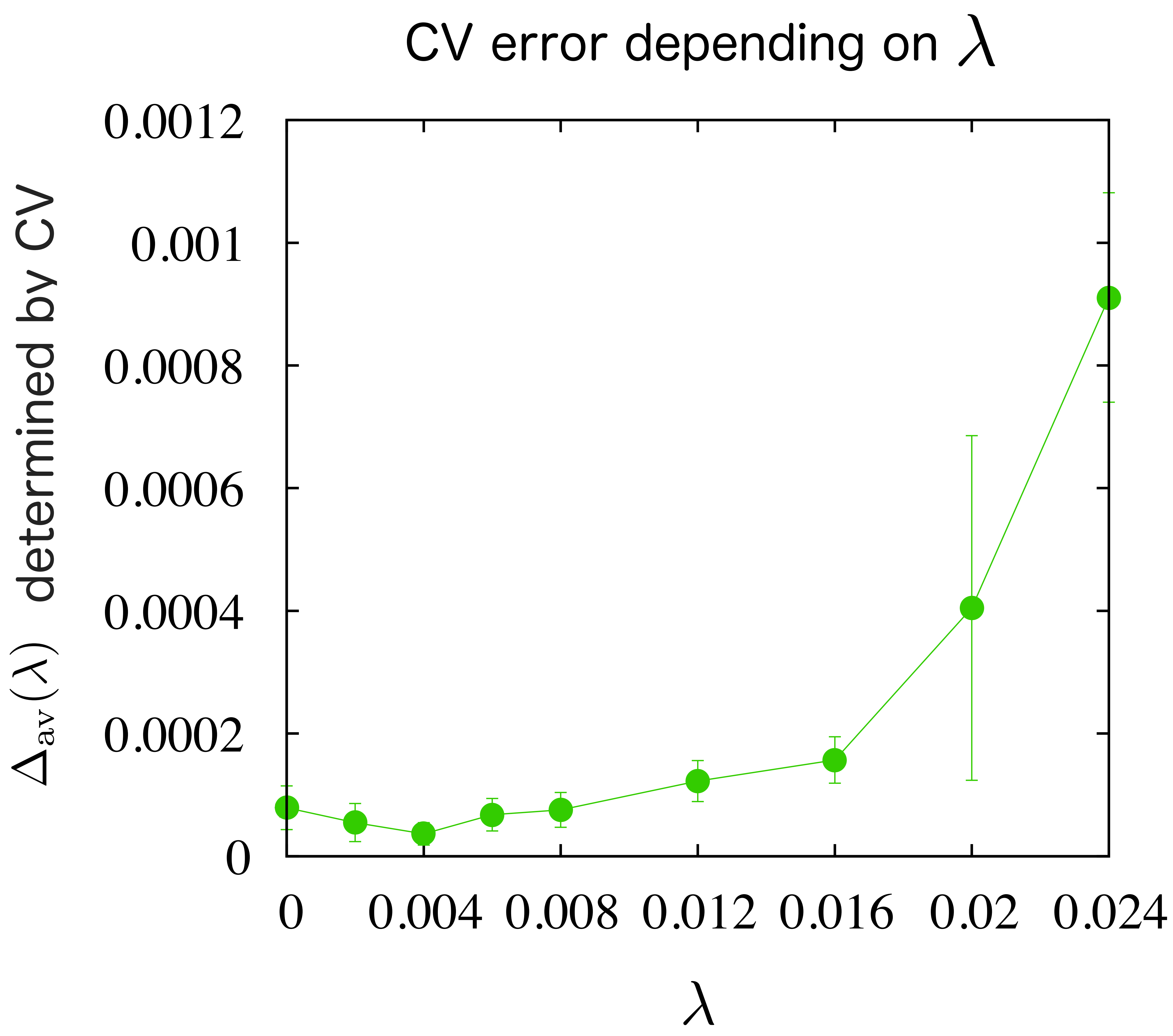



Estimation results

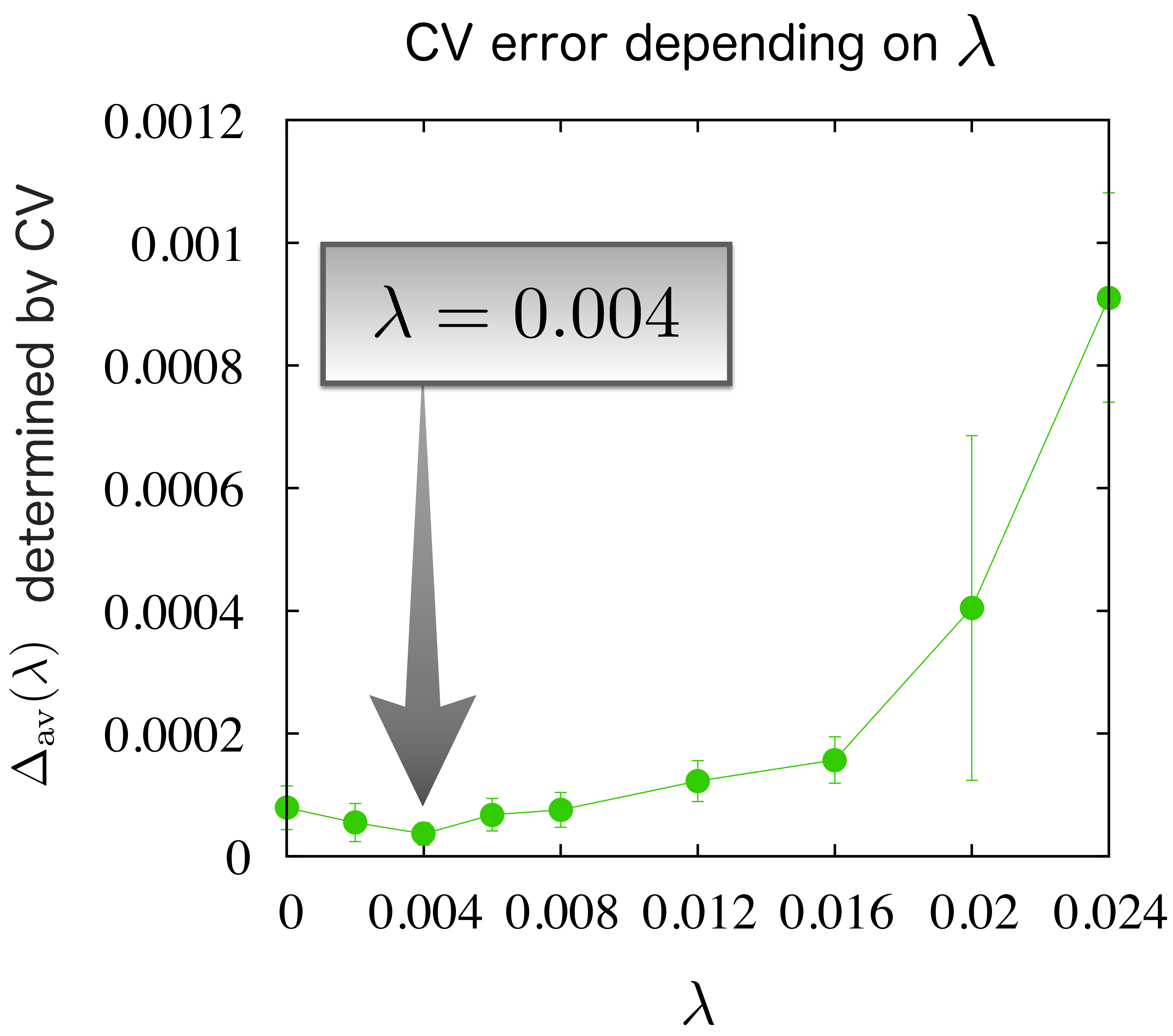
Prior distribution: L1 regularization



# Estimated results and prediction

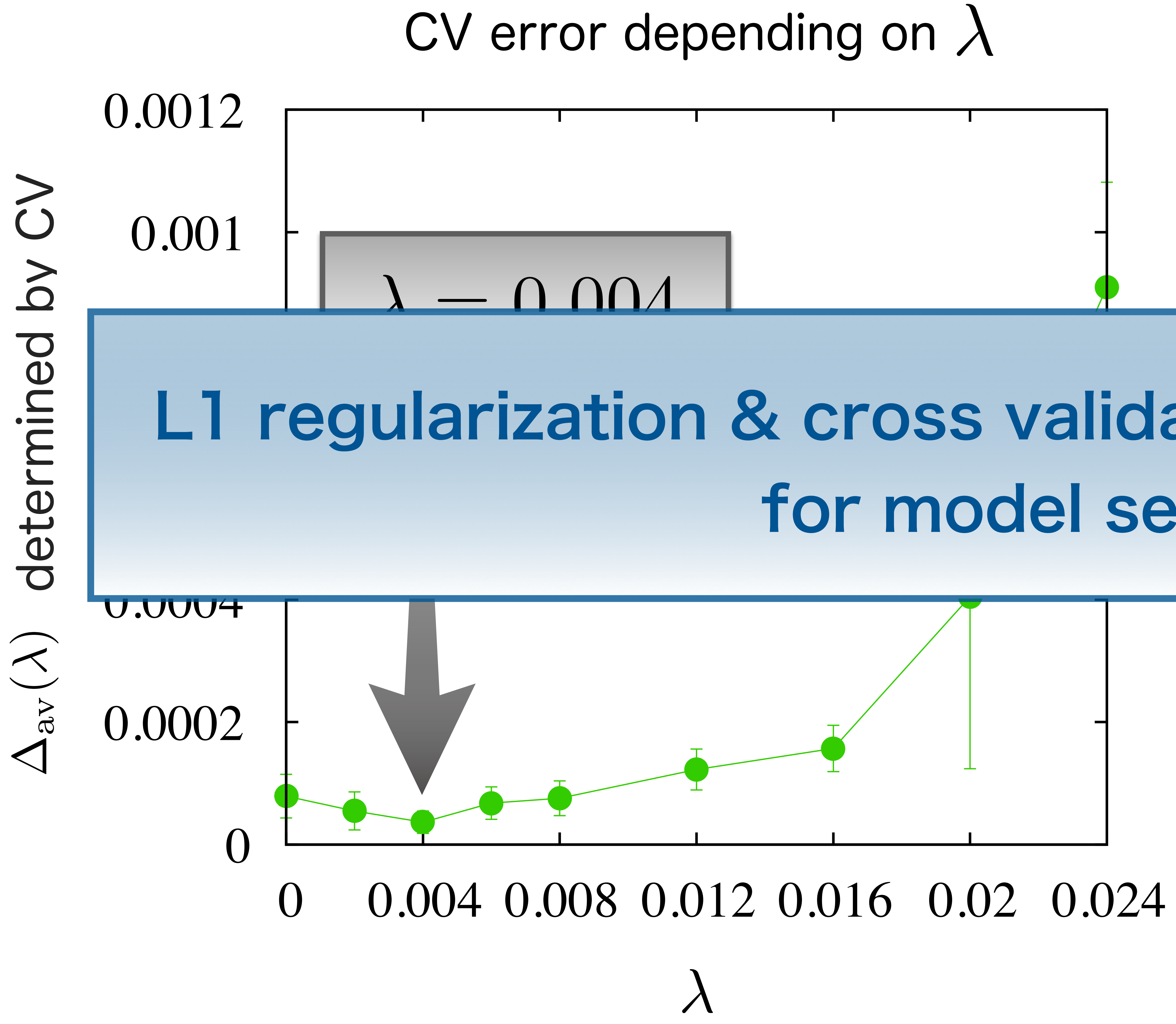


# Estimated results and prediction




| | Estimated | Correct |
|---|---|---|
| $J_1$ | 1.074 | 1.000 |
| $J_2$ | 3.850 | 4.000 |
| $J_3$ | 5.012 | 5.000 |
| $J_4$ | 6.051 | 6.000 |
| $J_5$ | 0.011 | 0.000 |
| $J_6$ | -0.051 | 0.000 |
| $J_7$ | 0.002 | 0.000 |
| $b$ | 0.102 | 0.100 |

# Estimated results and prediction




| | Estimated | Correct |
|---|---|---|
| $J_1$ | 1.074 | 1.000 |
| $J_2$ | 3.850 | 4.000 |
| $J_5$ | 0.011 | 0.000 |
| $J_6$ | -0.051 | 0.000 |
| $J_7$ | 0.002 | 0.000 |
| $b$ | 0.102 | 0.100 |

**L1 regularization & cross validation is effectively worked for model selection.**

# Time consuming problem

Energy function in effective model estimation

$$E(\mathbf{x}) = \frac{1}{2\sigma^2}\sum_{l=1}^{L}\left[y^{\mathrm{ex}}(g_l) - y^{\mathrm{cal}}(g_l,\mathbf{x})\right]^2 - \log P(\mathbf{x})$$

**We should calculate the thermal average of physical quantity by Monte Carlo, exact diagonalization, DMRG, mean-field, etc.**

# Time consuming problem

Energy function in effective model estimation

$$E(\mathbf{x}) = \frac{1}{2\sigma^2}\sum_{l=1}^{L}\left[y^{\mathrm{ex}}(g_l) - y^{\mathrm{cal}}(g_l, \mathbf{x})\right]^2 - \log P(\mathbf{x})$$

**We should calculate the thermal average of physical quantity by Monte Carlo, exact diagonalization, DMRG, mean-field, etc.**

Searching procedure of maximizer of posterior distribution

Values of model parameters are changed. → Thermal average is calculated. → One point value on posterior distribution is obtained.

# Time consuming problem

Energy function in effective model estimation

$$E(\mathbf{x}) = \frac{1}{2\sigma^2} \sum_{l=1}^{L} \left[ y^{\mathrm{ex}}(g_l) - y^{\mathrm{cal}}(g_l, \mathbf{x}) \right]^2 - \log P(\mathbf{x})$$

**We should calculate the thermal average of physical quantity by Monte Carlo, exact diagonalization, DMRG, mean-field, etc.**

Searching procedure of maximizer of posterior distribution

Values of model parameters are changed. → Thermal average is calculated. → One point value on posterior distribution is obtained.

***Time consuming !***

**We want to reduce the number of calculations of thermal averages.** → **Bayesian optimization**

# Bayesian optimization

✔ We want to search the maximizer in M candidates.

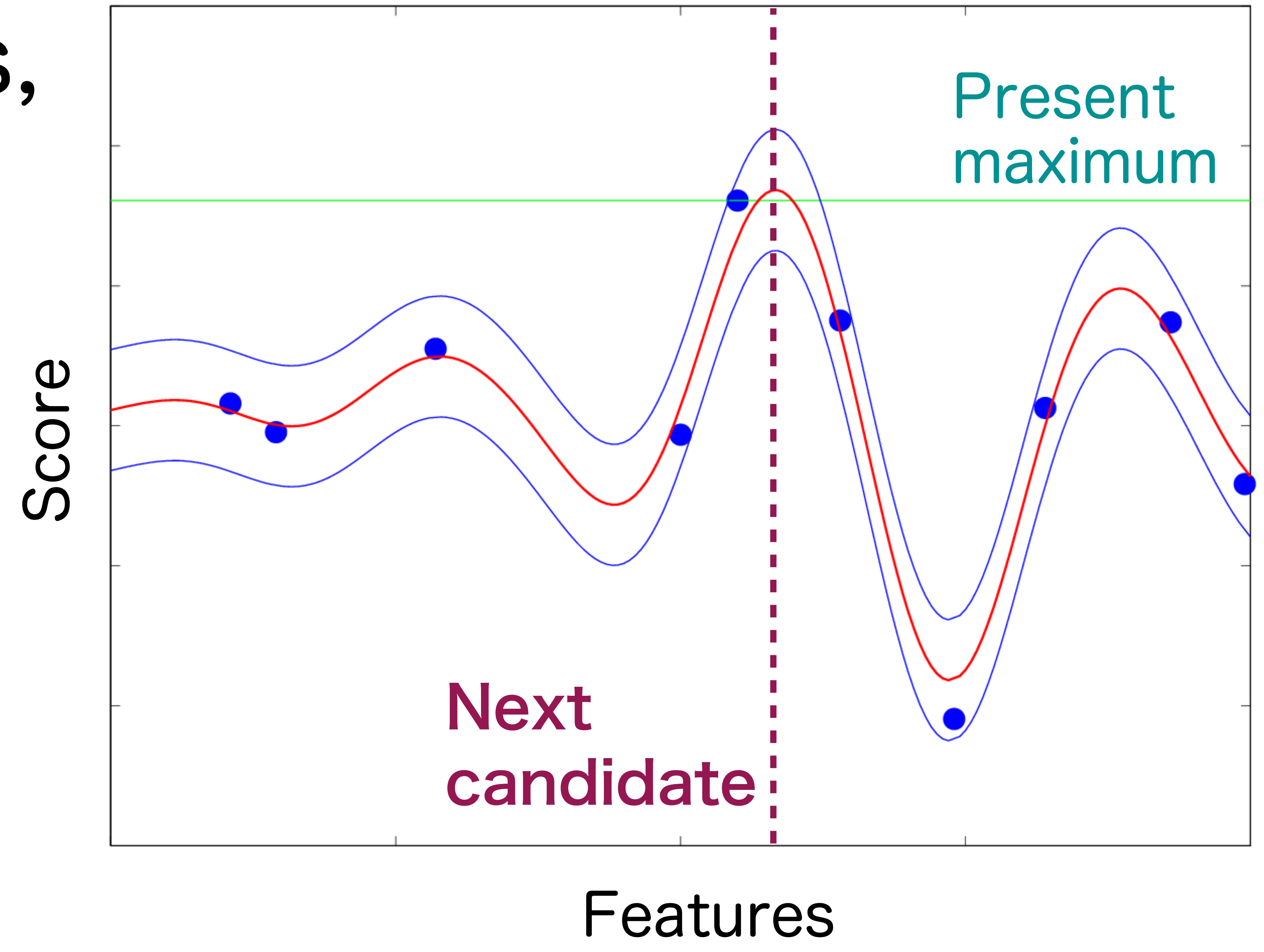
✔ We want to reduce the number of measurements or synthesis.

✔ We finished evaluations of N candidates. M-N candidates are remaining.

✔ We want to select next candidate, effectively.

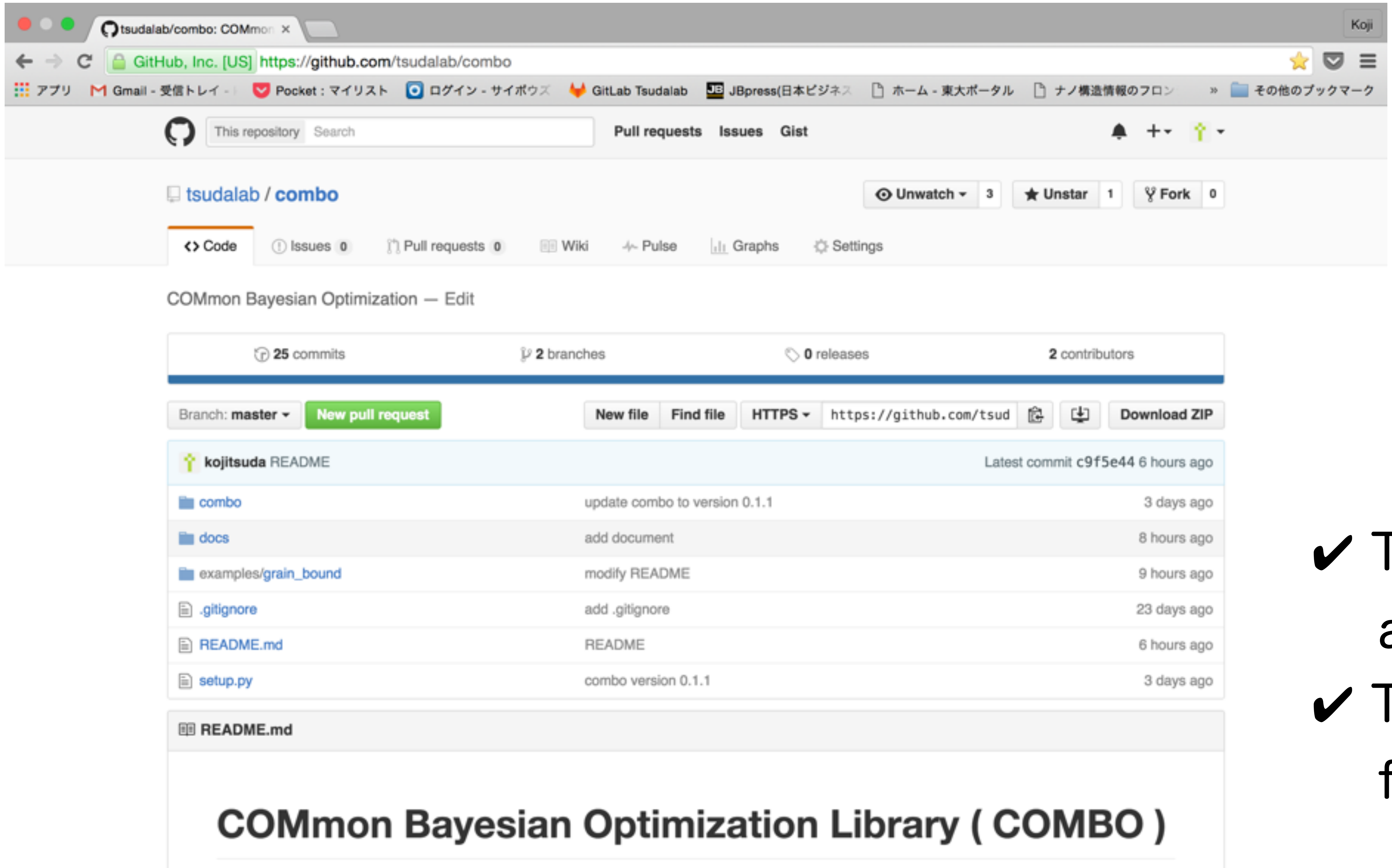
✔ We train Gaussian process using N candidates, and scores of M-N candidates are predicted.

✔ We chose the next candidate with high score.




T. Ueno, T. D. Rhone, Z. Hou, T. Mizoguchi and K. Tsuda, Materials Discovery, **4**, 18, 2016.

# COMBO



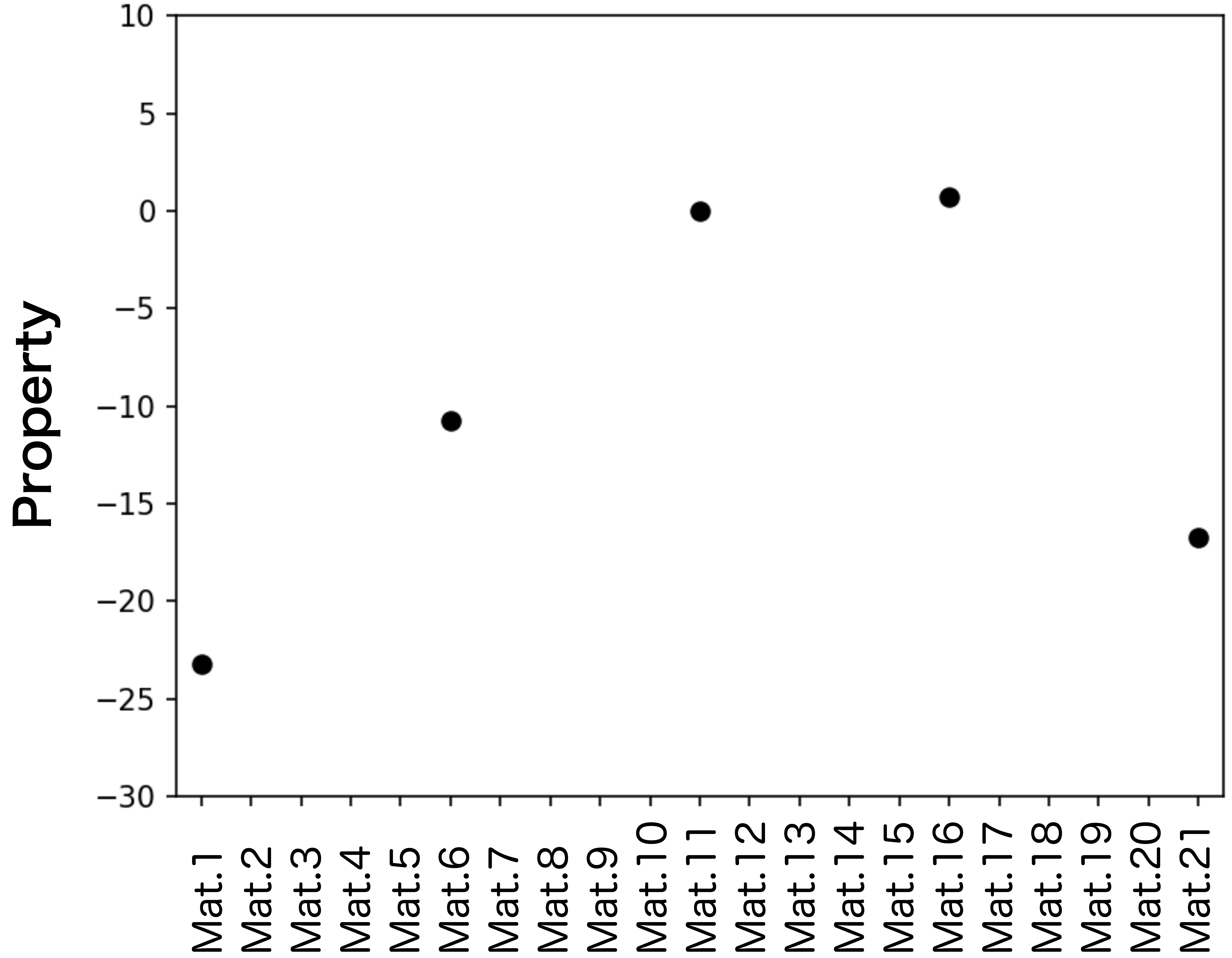

✔ The hyperperameters are automatically determined.

✔ The order N calculation time for training data is realized.

https://github.com/tsudalab/combo

# Bayesian optimization

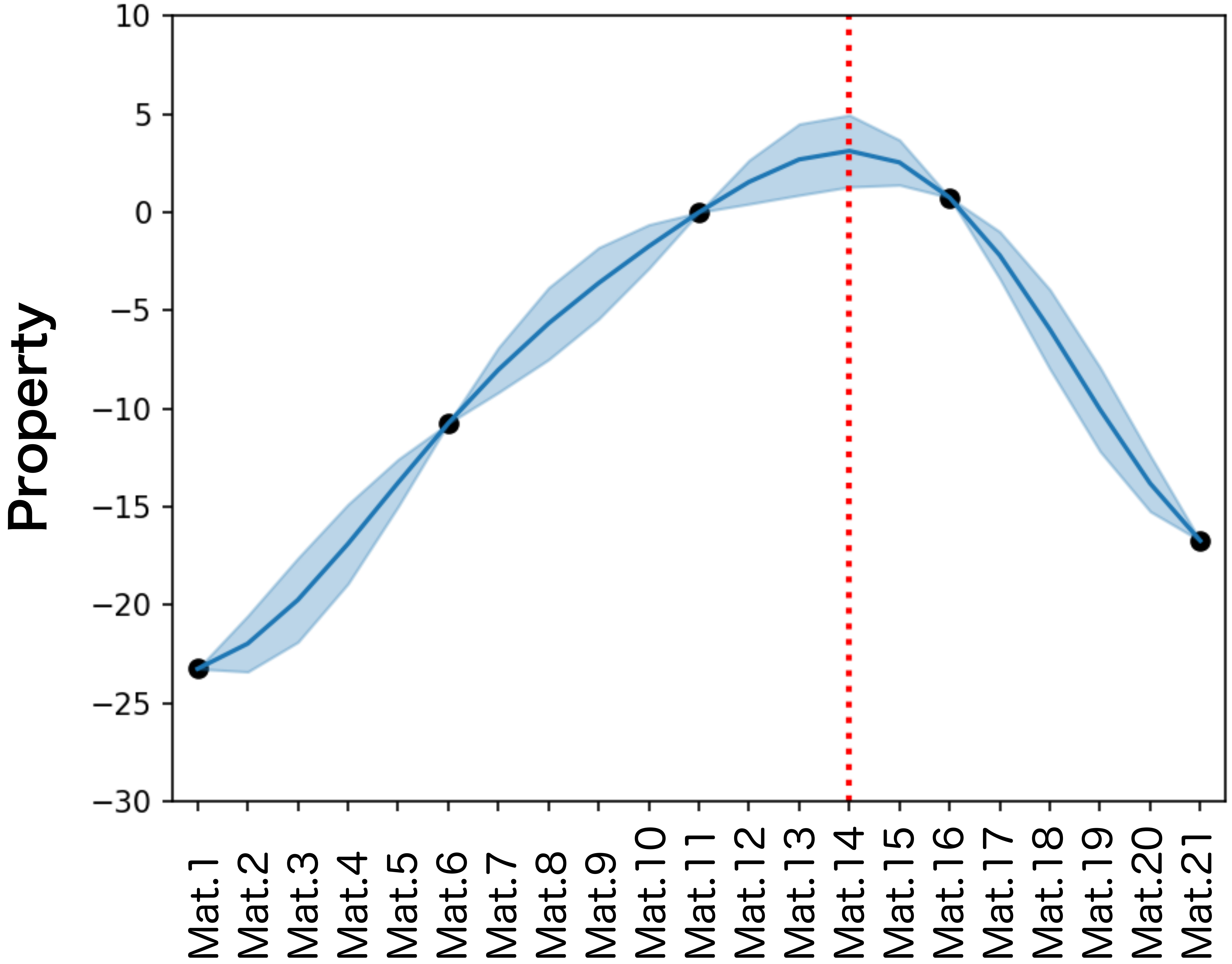
## Prediction model: Gaussian process




We chose the next candidate by mean and variance.

# Bayesian optimization

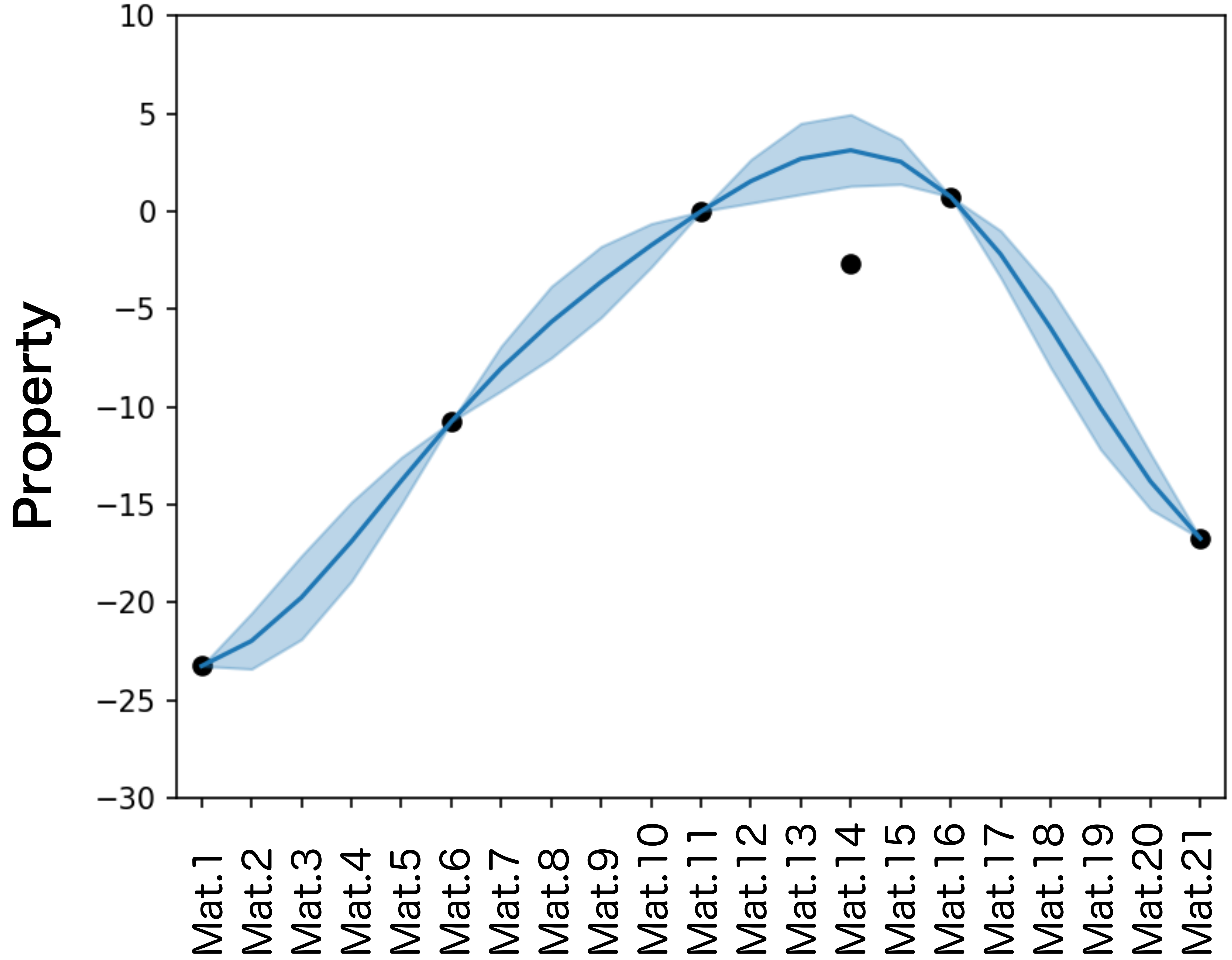
## Prediction model: Gaussian process




We chose the next candidate by mean and variance.

# Bayesian optimization

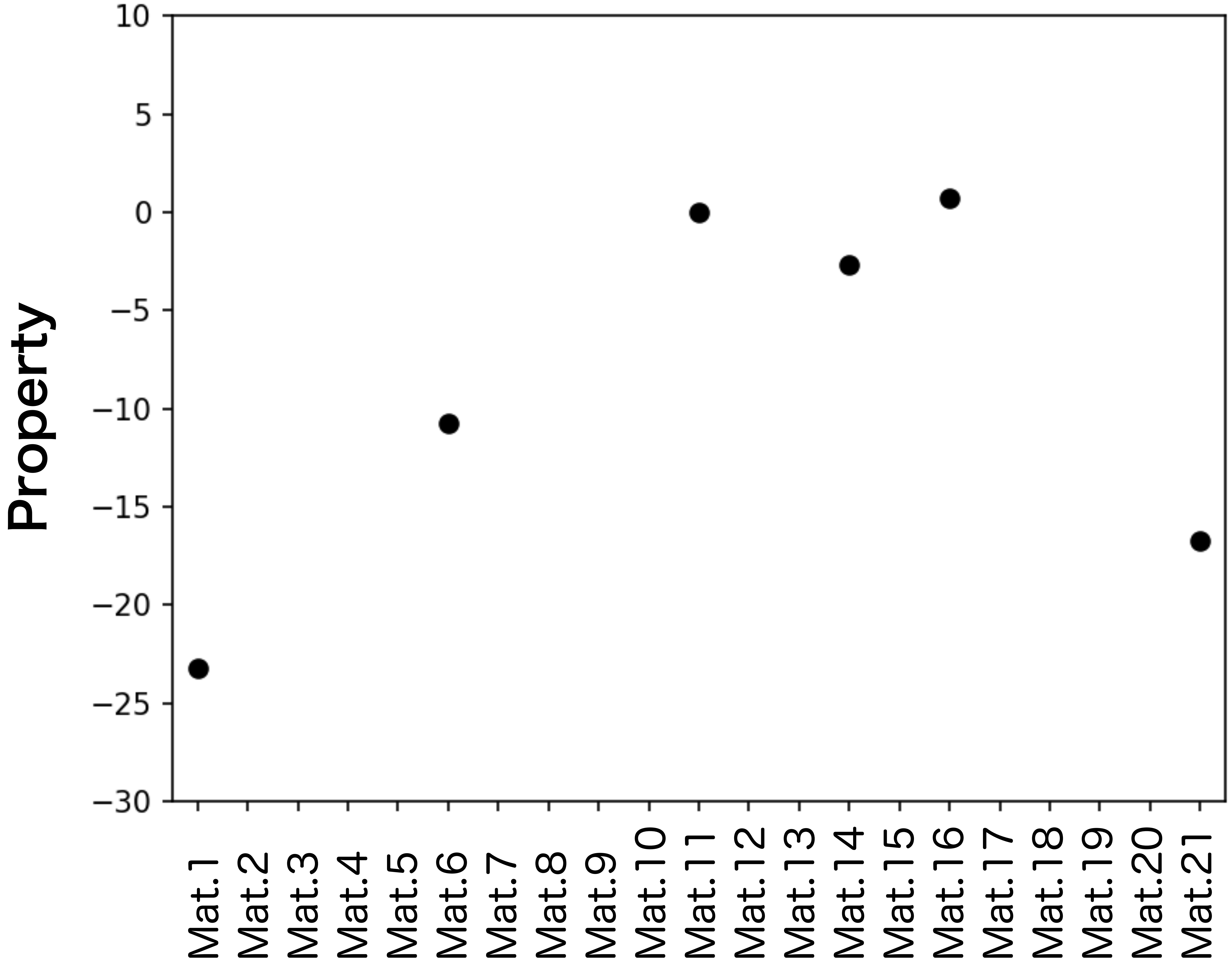
## Prediction model: Gaussian process




We chose the next candidate by mean and variance.

# Bayesian optimization

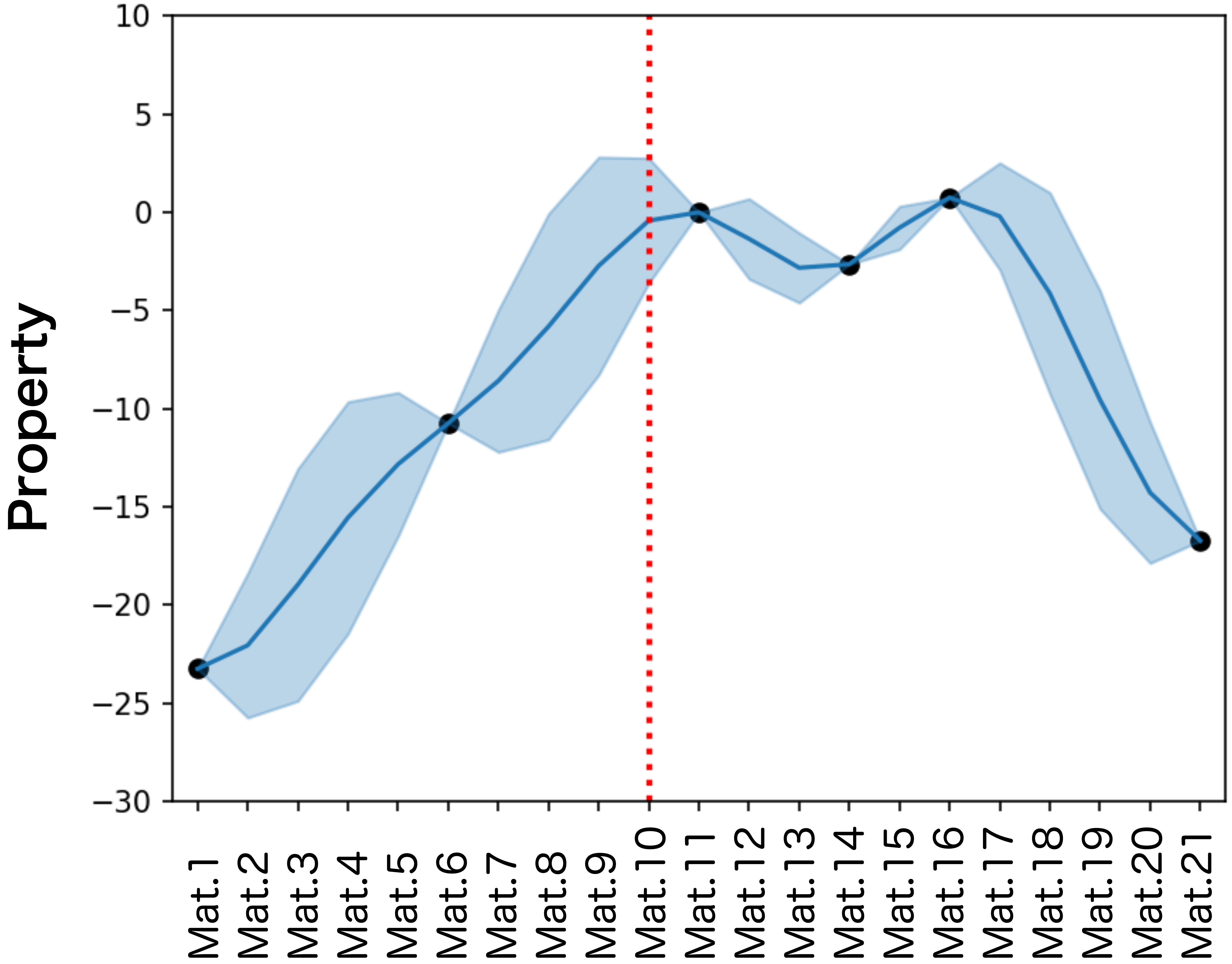
## Prediction model: Gaussian process




We chose the next candidate by mean and variance.

# Bayesian optimization

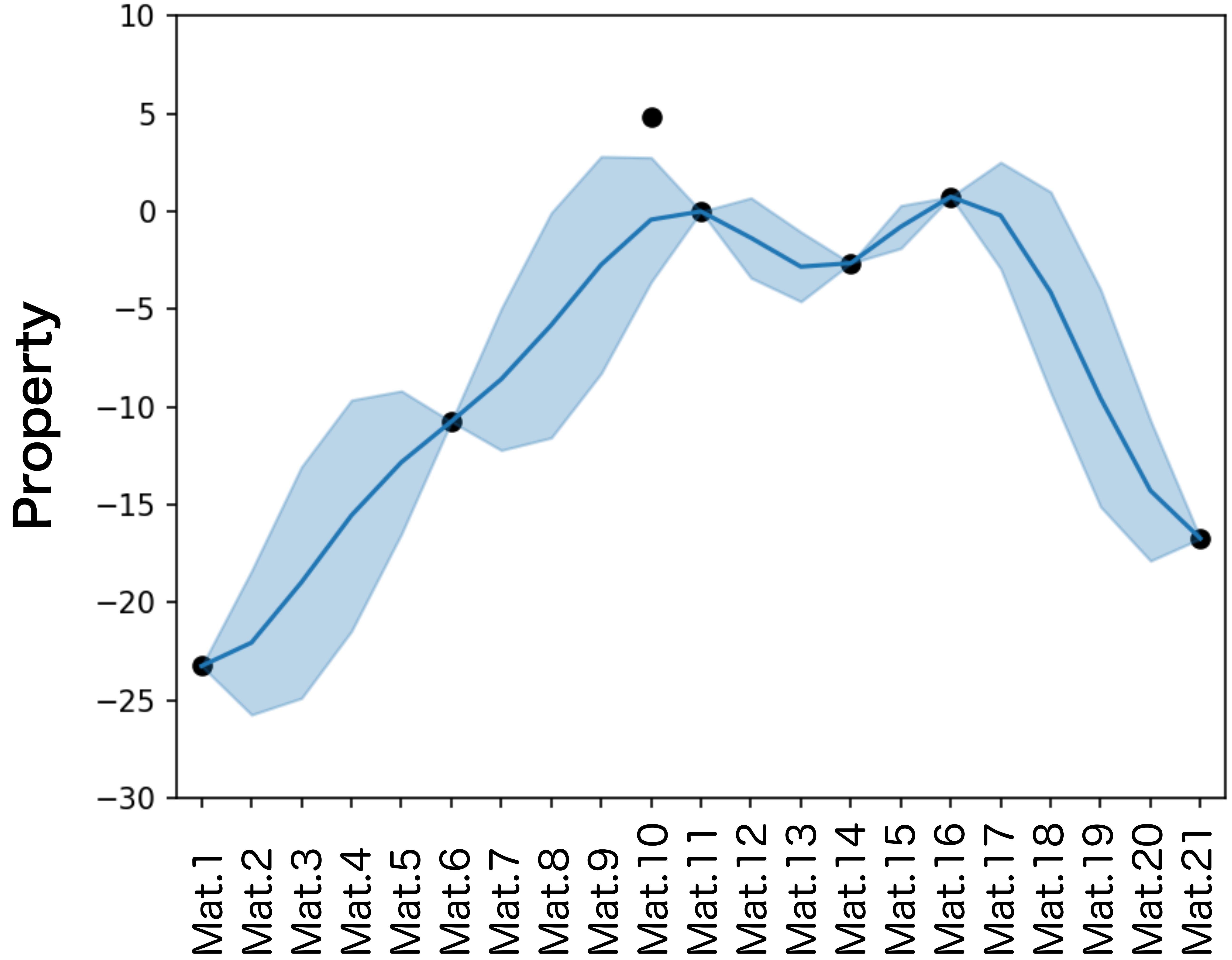
## Prediction model: Gaussian process




We chose the next candidate by mean and variance.

# Bayesian optimization

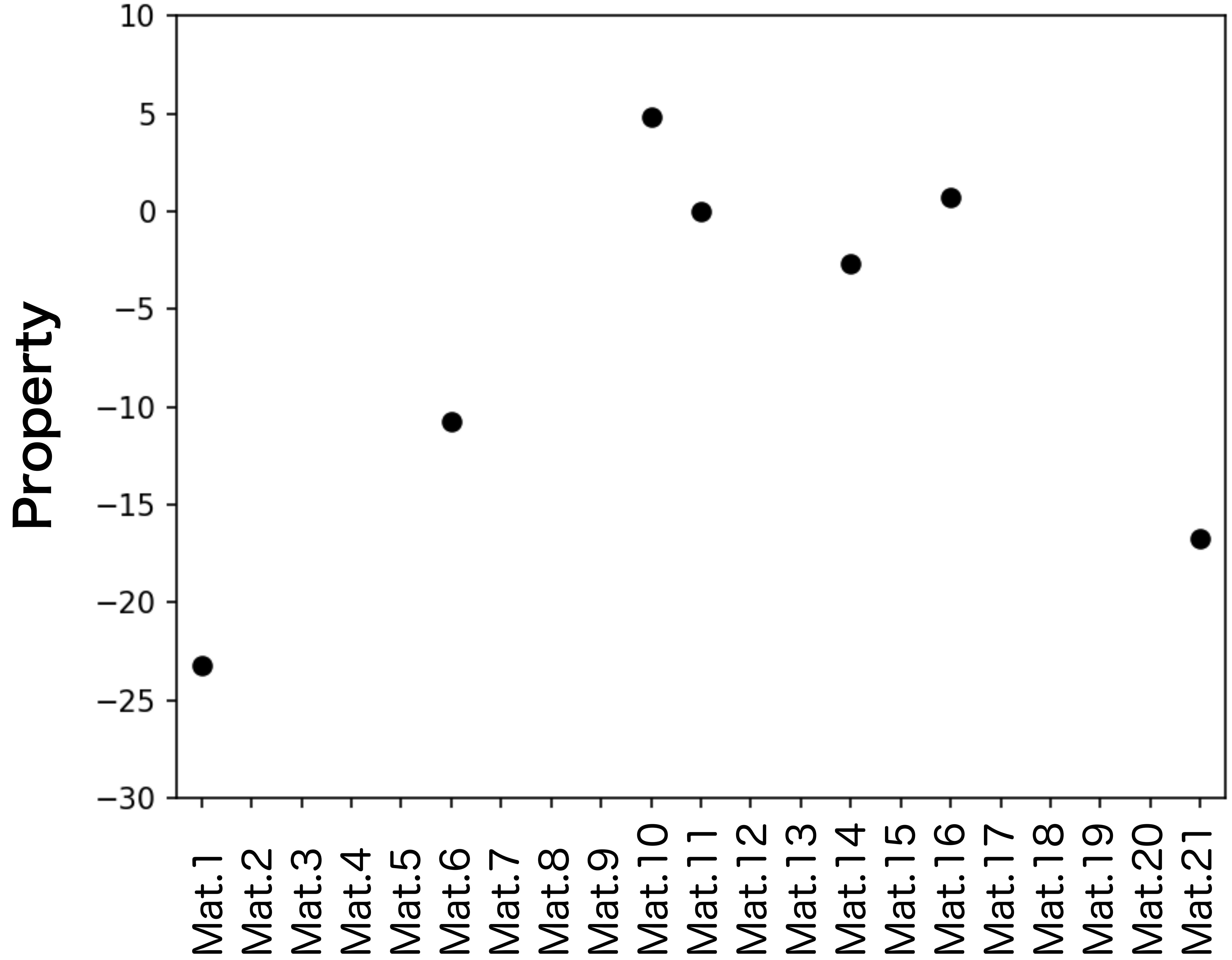
## Prediction model: Gaussian process




We chose the next candidate by mean and variance.

# Bayesian optimization

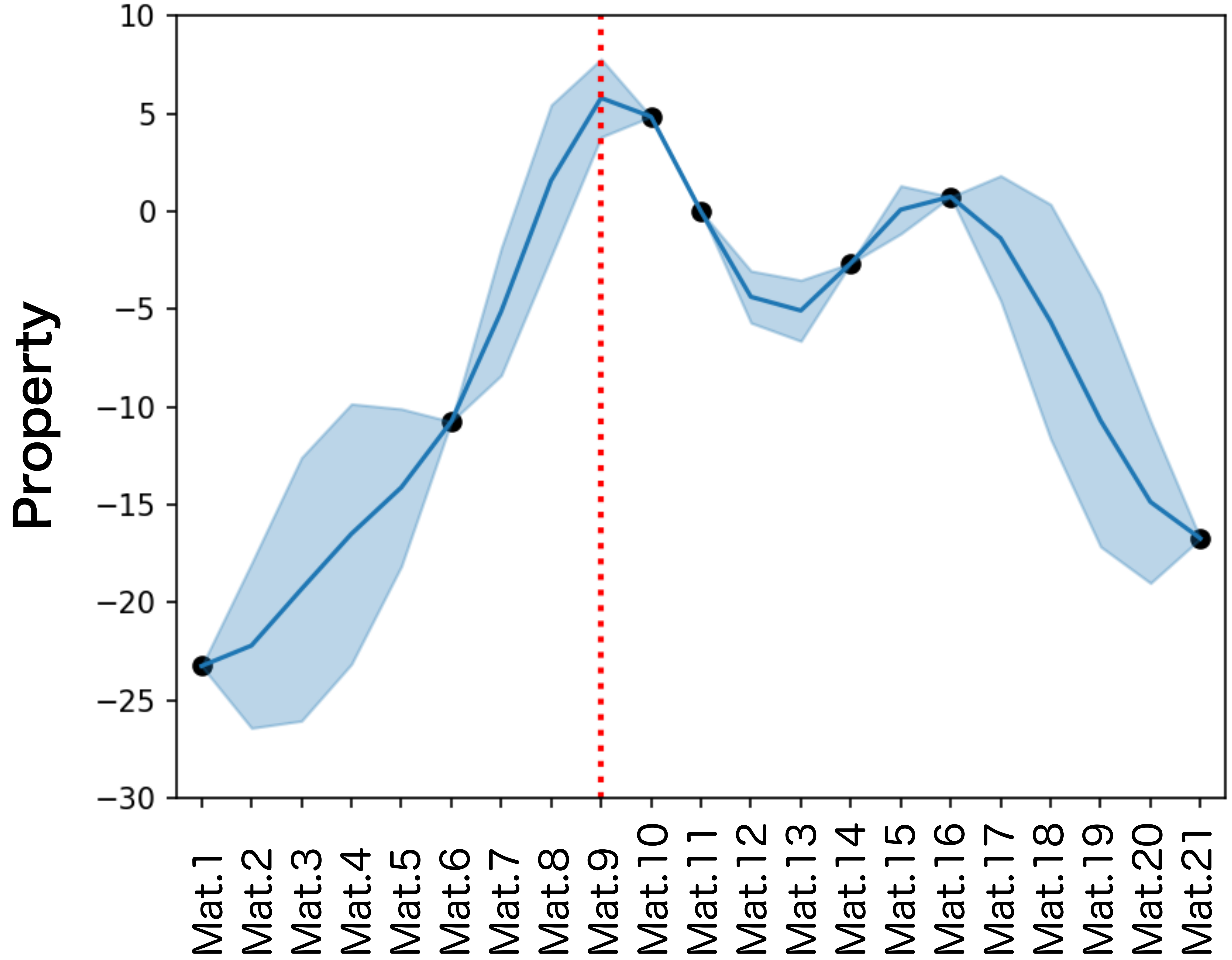
## Prediction model: Gaussian process




We chose the next candidate by mean and variance.

# Bayesian optimization

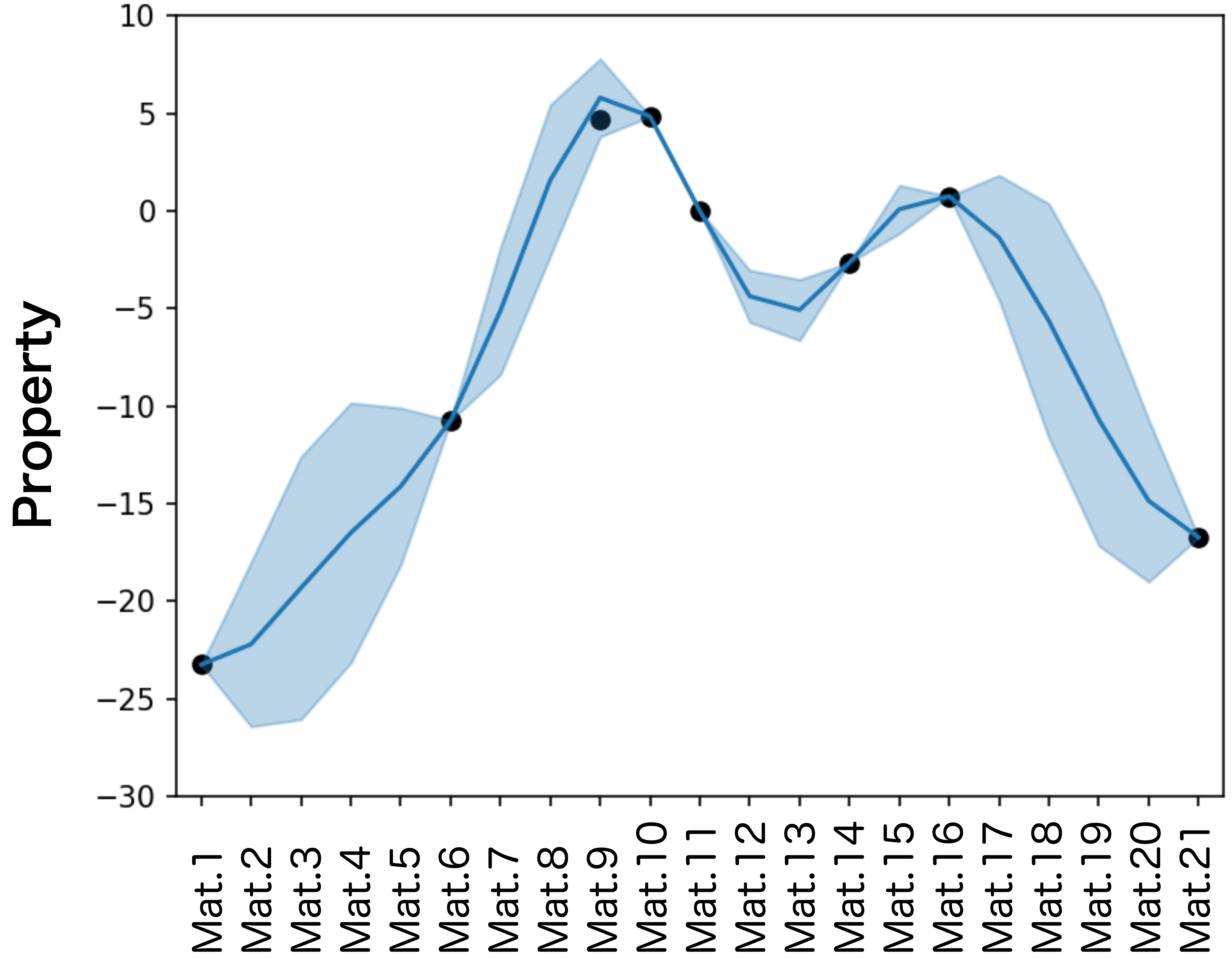
## Prediction model: Gaussian process




We chose the next candidate by mean and variance.

# Bayesian optimization

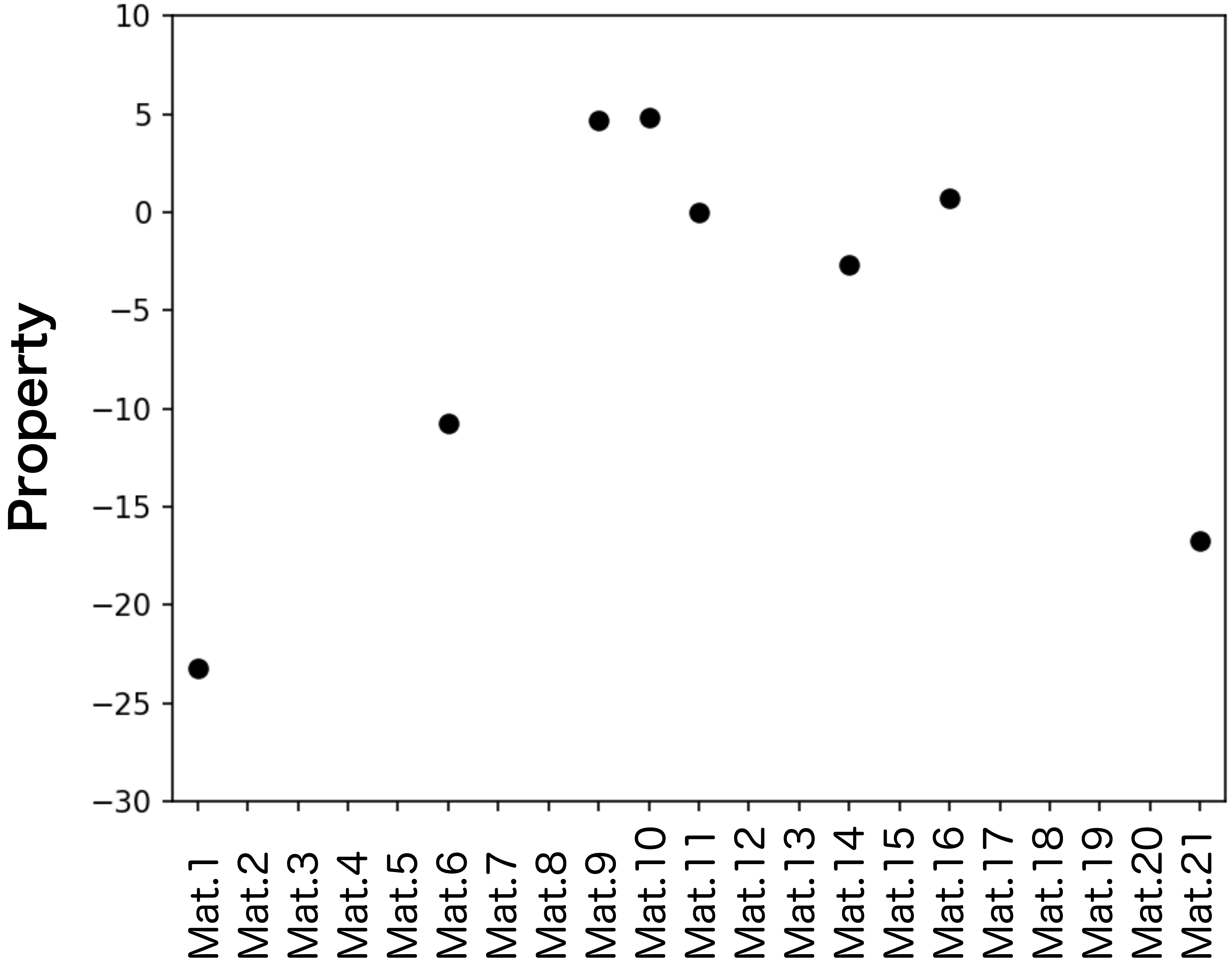
## Prediction model: Gaussian process




We chose the next candidate by mean and variance.

# Bayesian optimization

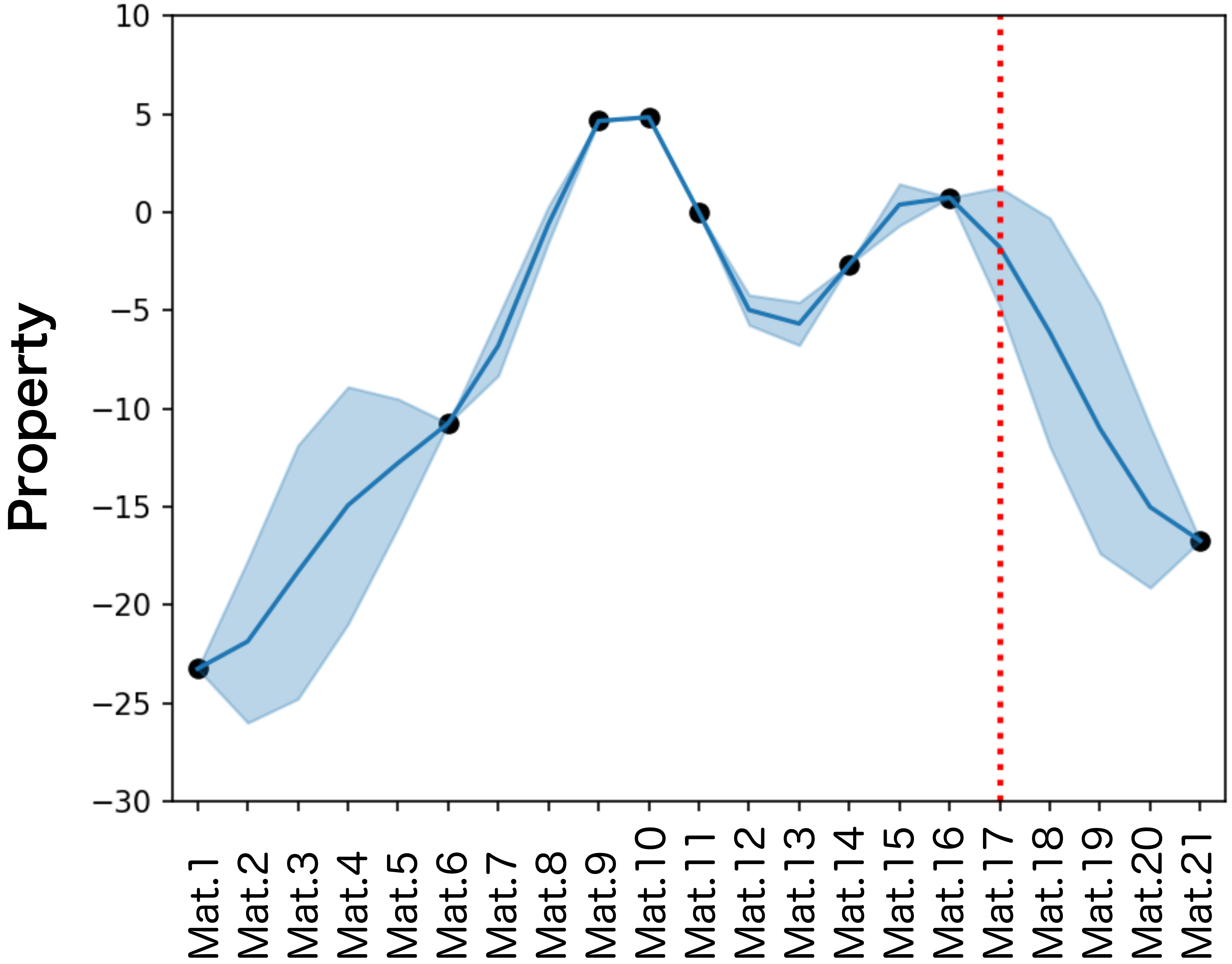
## Prediction model: Gaussian process




We chose the next candidate by mean and variance.

# Bayesian optimization

## Prediction model: Gaussian process




We chose the next candidate by mean and variance.

# Acquisition functions

✔ Maximum Probability of Improvement (PI)

–Probability to exceed current max

✔ Maximum Expected Improvement (EI)

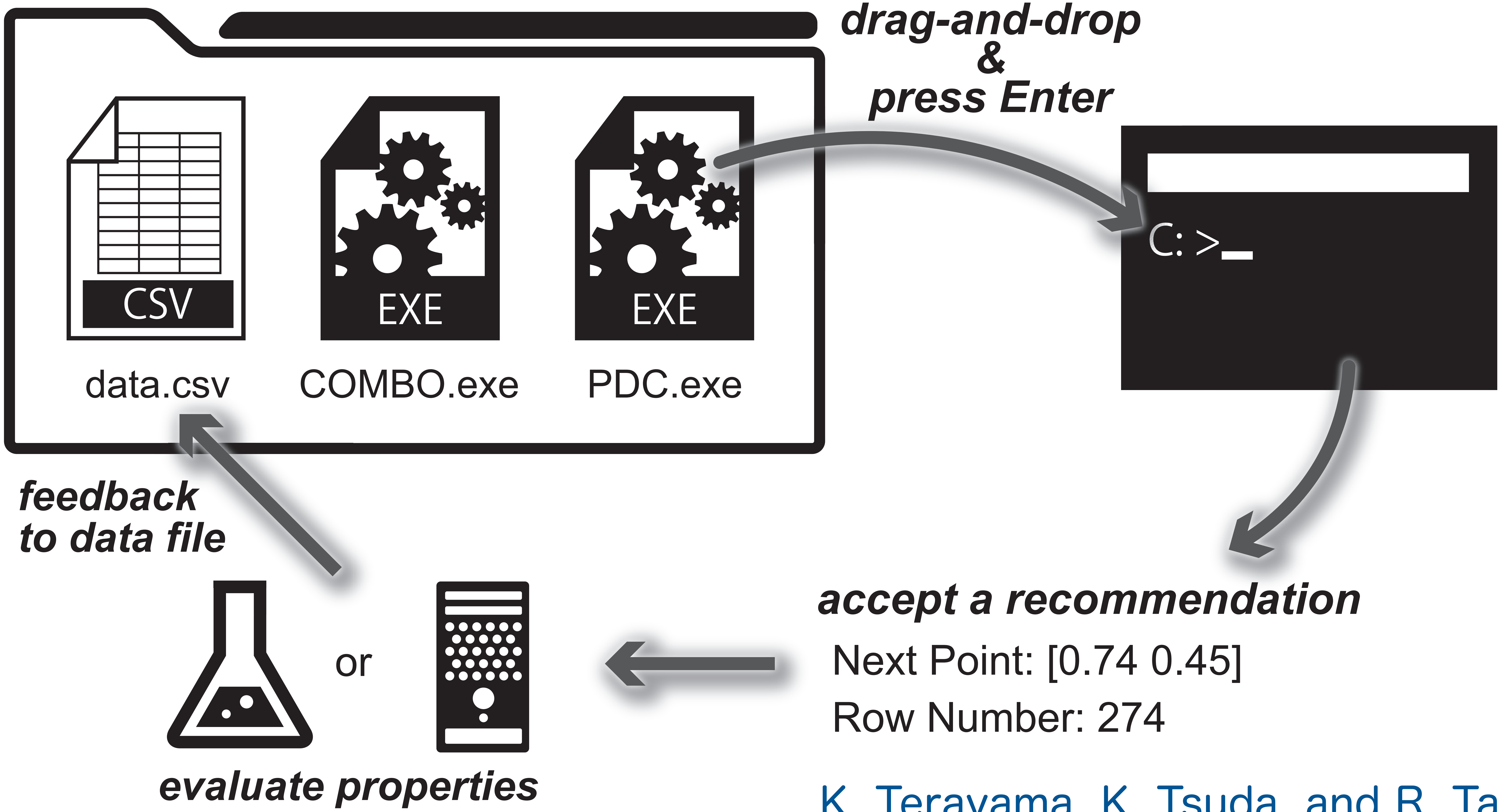
–Expected value of
(measurement value - current max)

✔ Thompson Sampling (TS)

–By using conditional probability,
sampling is performed.

# For Windows users

## Executable files of COMBO and PDC are developed.




K. Terayama, K. Tsuda, and R. Tamura, Jpn. J. Appl. Phys. **58**, 098001 (2019).

# Data in Bayesian optimization

data.csv

Explanations of each column (or blanks) → row 1

| 1 | objective | parameter1 | parameter2 | … |
|---|---|---|---|---|
| 2 | | 0.0 | 10 | … |
| 3 | 8.2 | 0.2 | 9.7 | … |
| 4 | | 0.4 | 9.4 | … |
| 5 | | 0.6 | 9.1 | … |
| 6 | | 0.8 | 8.8 | … |
| 7 | 4.5 | 1.0 | 8.5 | … |
| 8 | | 1.2 | 8.2 | … |
| 9 | 2.0 | 1.4 | 7.9 | … |
| 10 | 3.7 | 1.6 | 7.6 | … |
| 11 | | 1.8 | 7.3 | … |
| 12 | | 2.0 | 7.0 | … |
| 13 | 10.1 | 2.2 | 6.7 | … |
| 14 | ⋮ | ⋮ | ⋮ | |

Recomend a next candidate from blanks

COMBO.exe
Objective
as real value

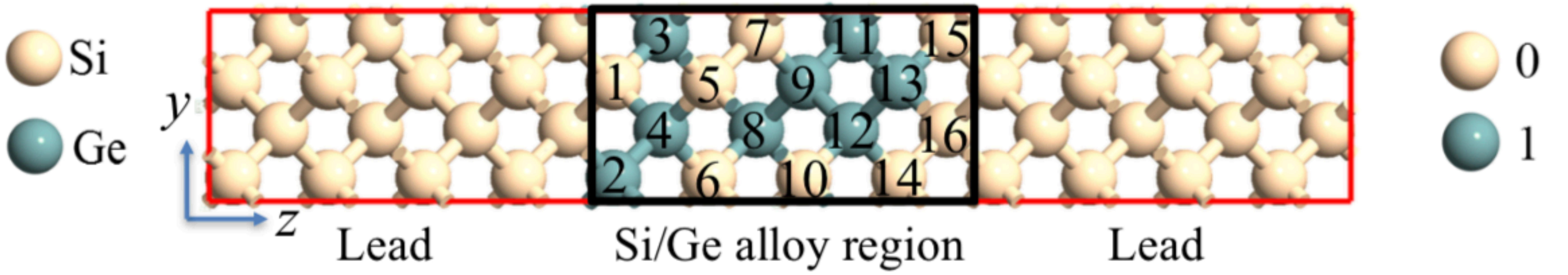
PDC.exe
Index of phases
as integer value

Non filled objective variables (1D)

Filled explanatory variables (any dimension)

# Optimization of thermal conductivity

**Question:** How to organize 16 alloy atoms (Si: 8, Ge: 8) to obtain the largest and smallest interfacial thermal conductance?




**Descriptors:** $C_{16}^{8} = 12{,}870$

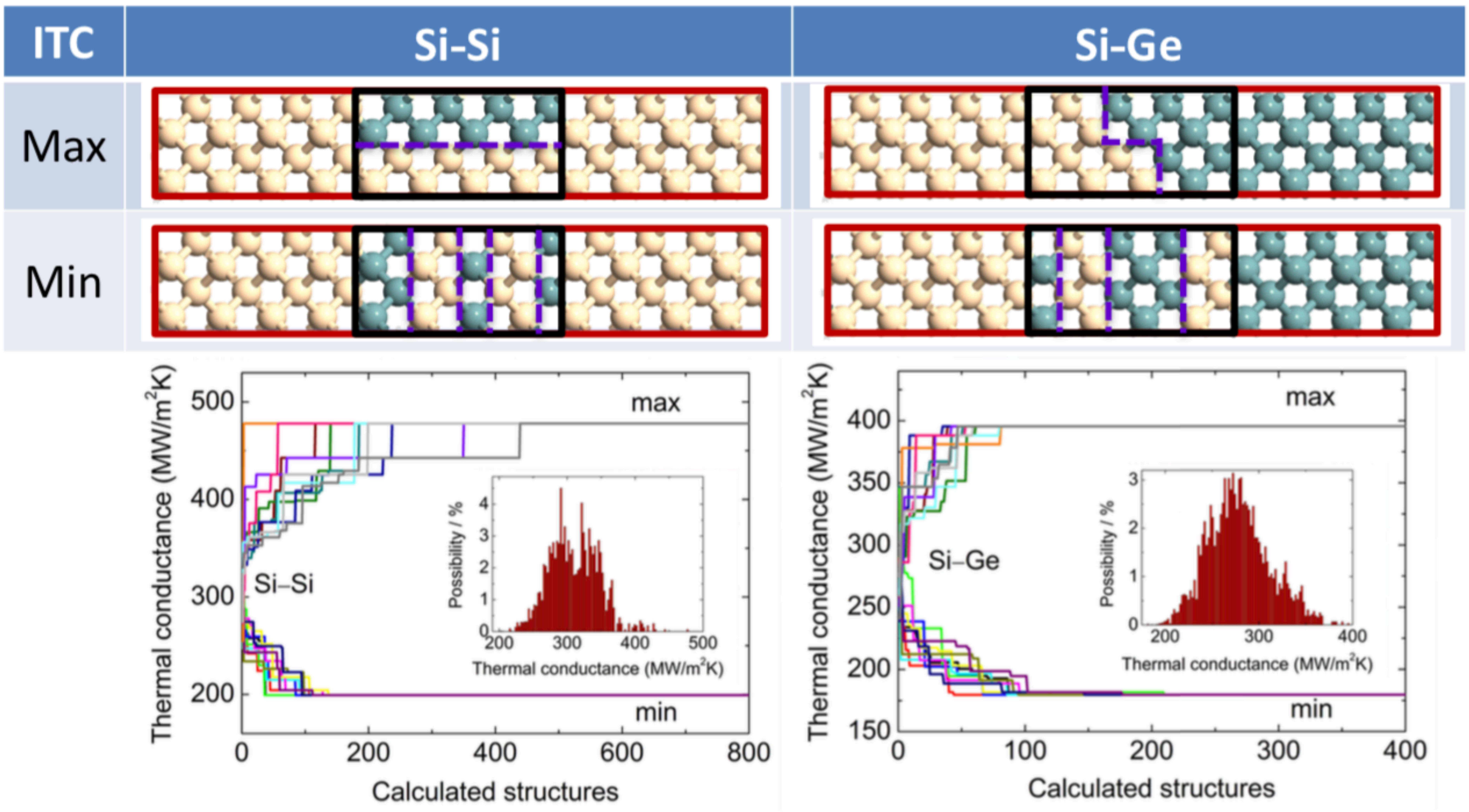
| Case | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … |

**Calculator:** Atomistic Green's Function (AGF): Phonon transmission

**Evaluator:** Interfacial Thermal Conductance (ITC)

**Optimization method:** Thompson Sampling (Bayesian Optimization)

# Optimization of thermal conductivity

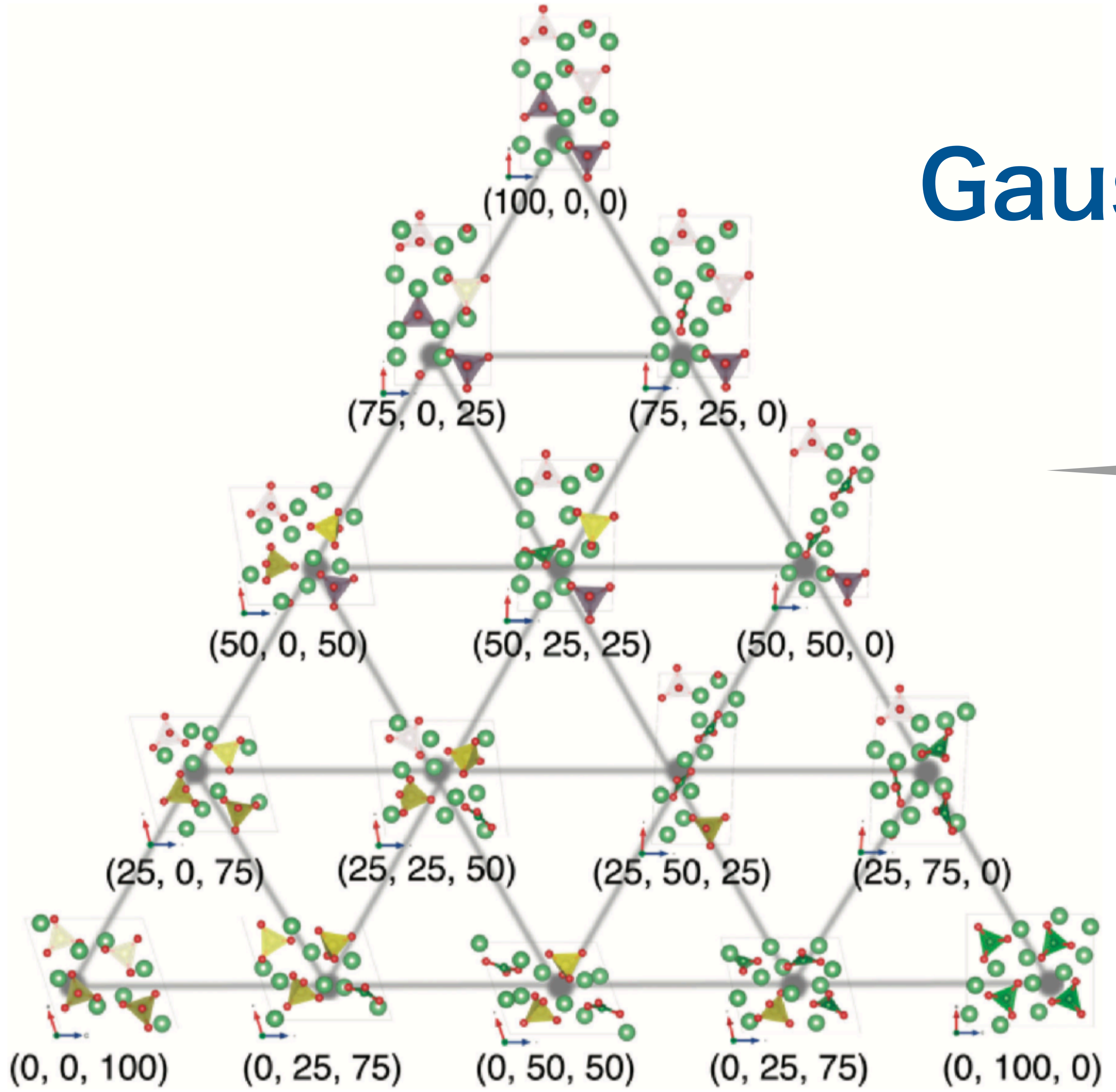



S. Ju, K. Tsuda, J. Shiomi, et al, Phys. Rev. X **7**, 021024 (2017).

# Compositional optimization

## Li-ion conductivity of ternary mixed system

$Li_3PO_4$



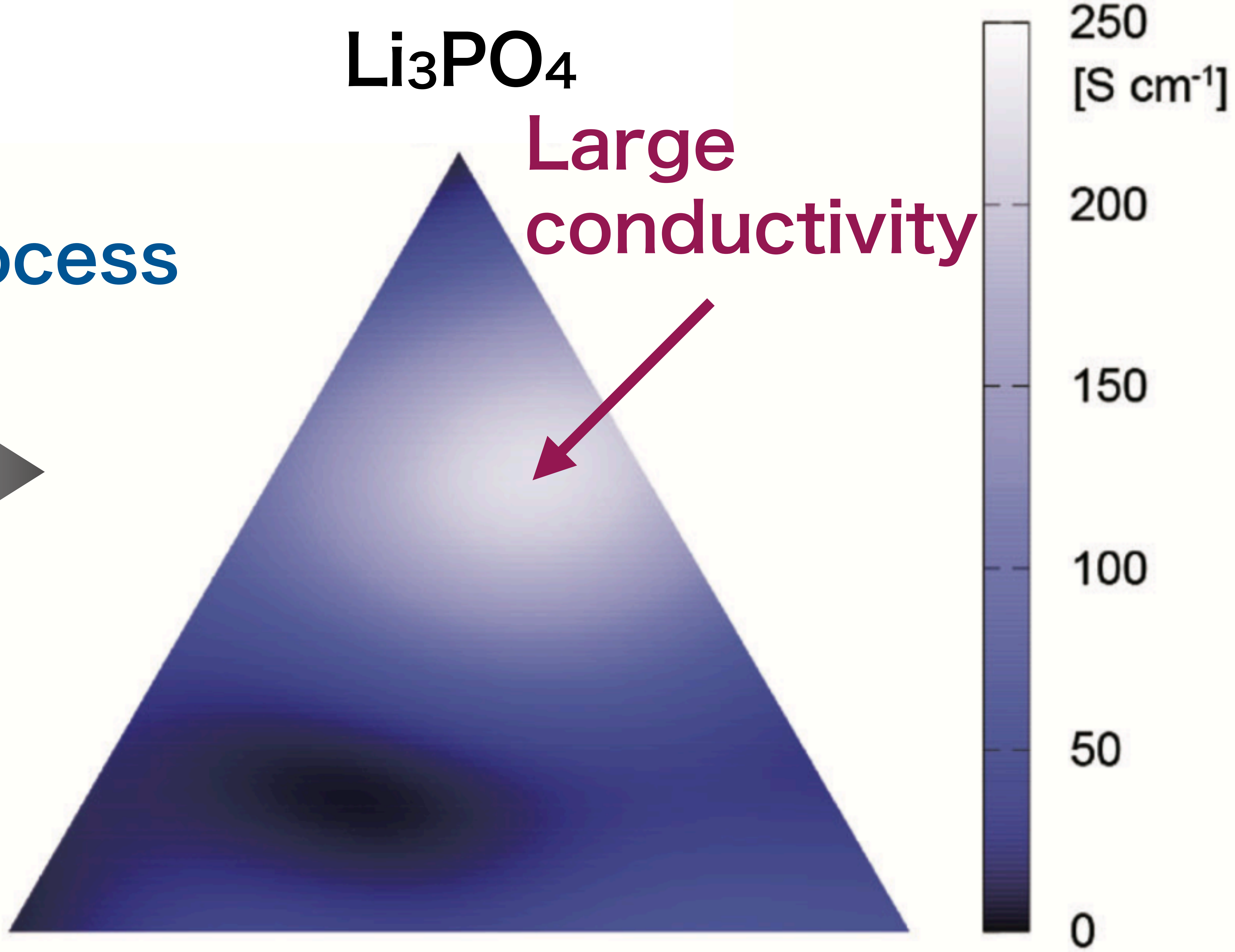

$Li_2SO_4$ $Li_3BO_3$

Gaussian process
COMBO

only
15 data

$Li_3PO_4$

Large conductivity




$Li_2SO_4$ $Li_3BO_3$

M. Sumita, R. Tamura, K. Tsuda, et al. Bull. Chem. Soc. Jpn. **92**, 1100 (2019).

# Compositional optimization

## Li-ion conductivity of ternary mixed system

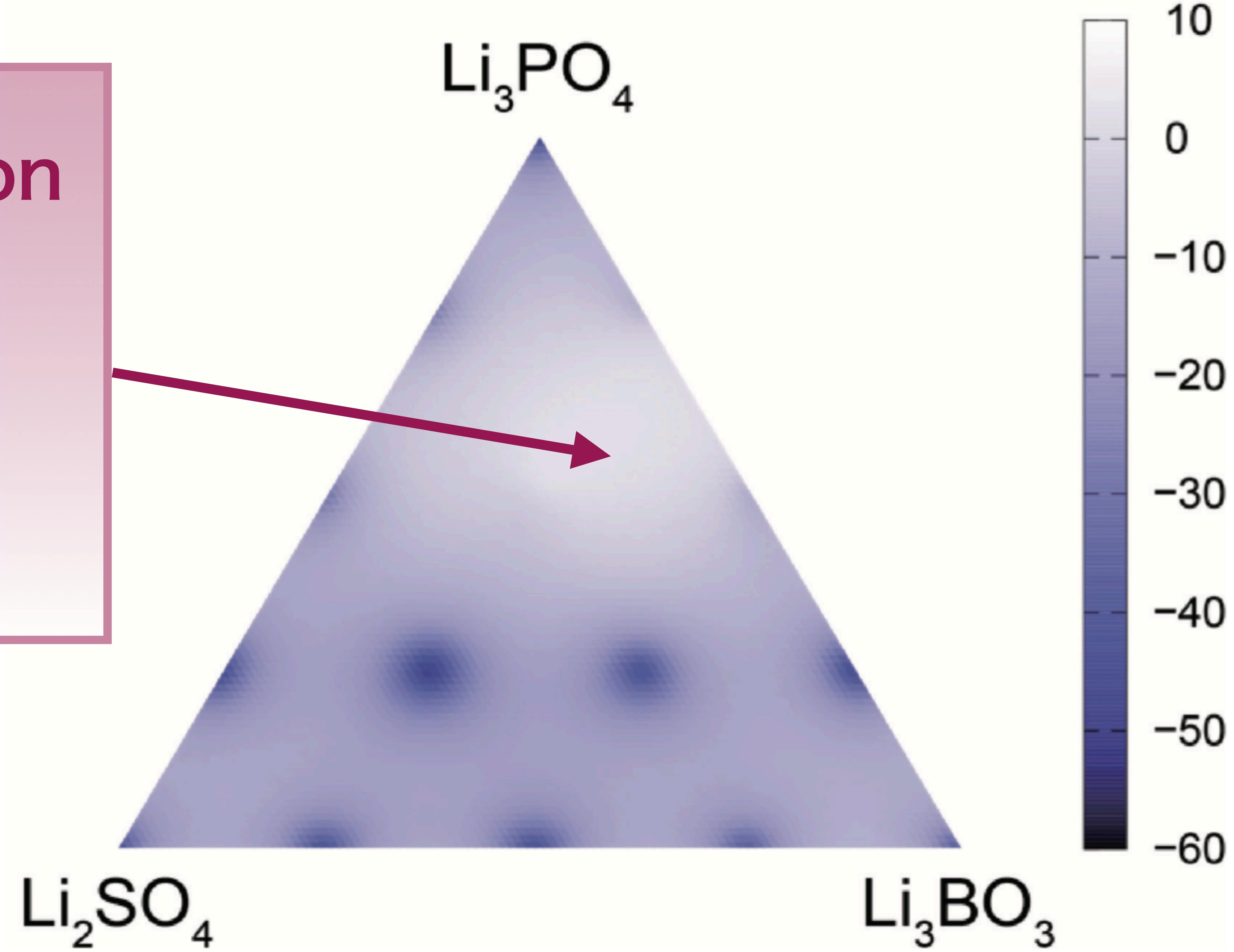
z-score : Approximation of PI

Optimal composition

$Li_3PO_4$ : 55%

$Li_3BO_3$ : 26%

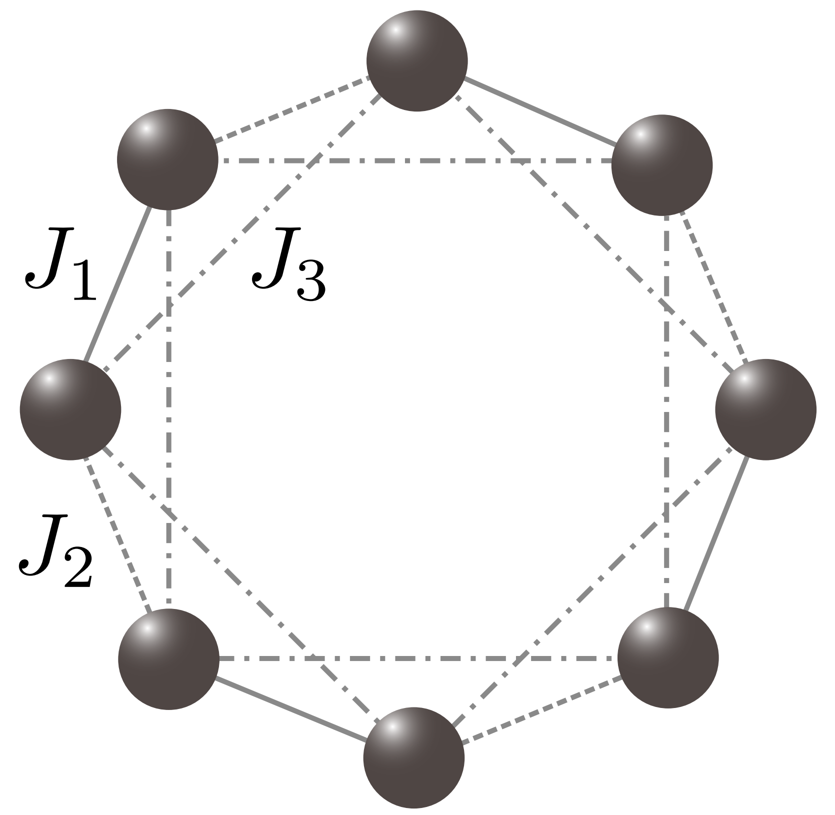
$Li_2SO_4$ : 19%

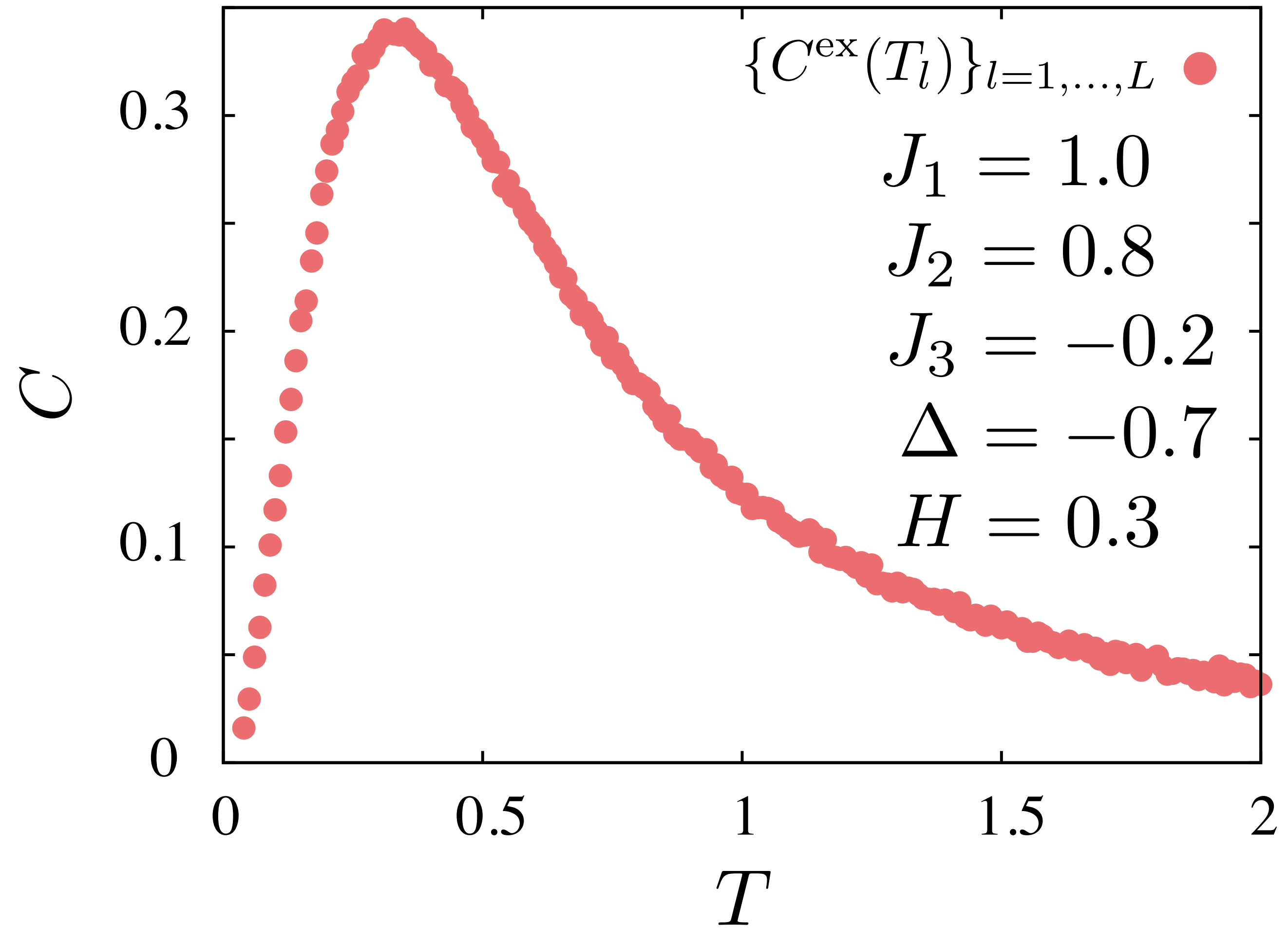


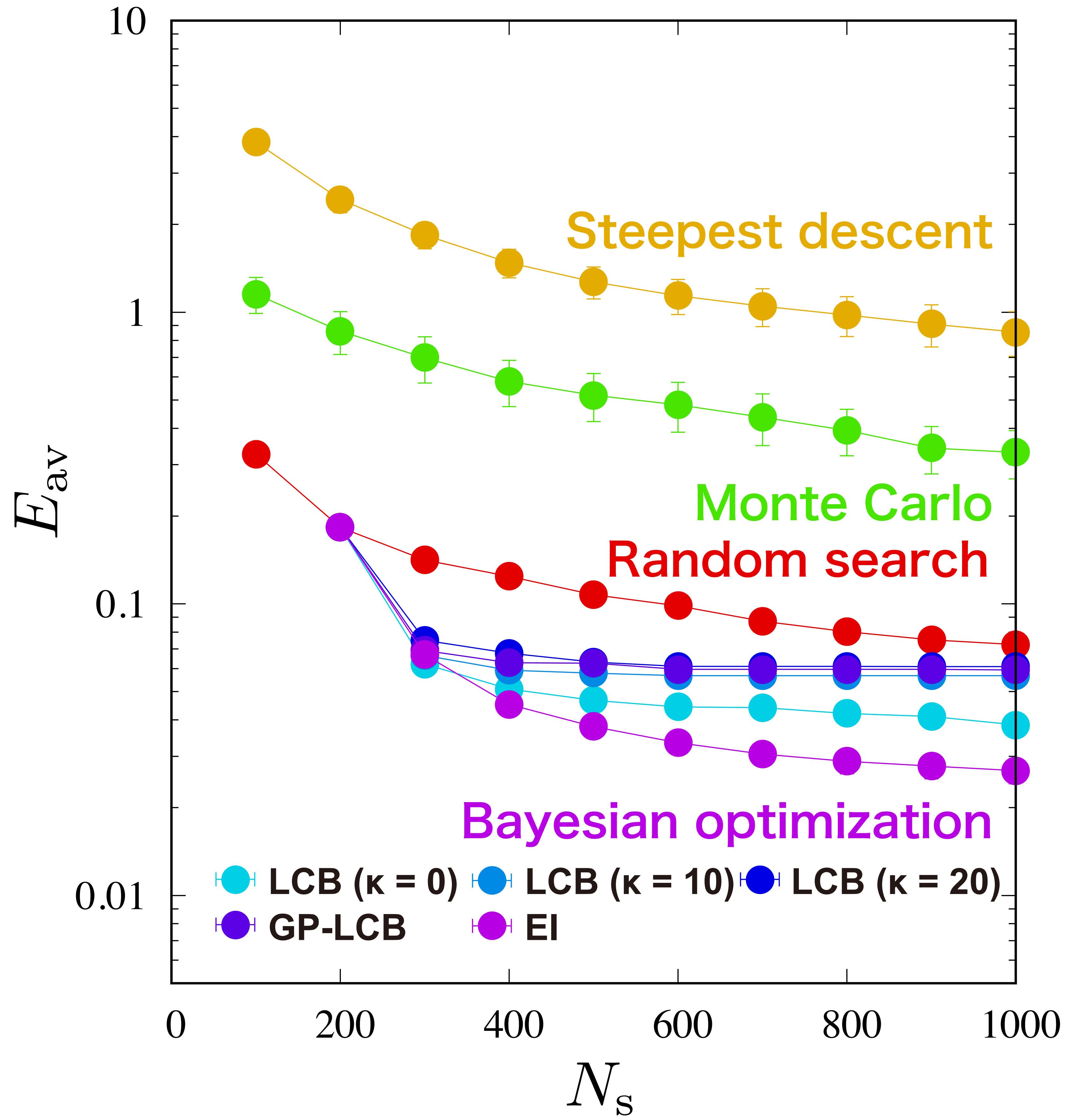

M. Sumita, R. Tamura, K. Tsuda, et al. Bull. Chem. Soc. Jpn. **92**, 1100 (2019).

# Bayesian optimization

Quantum Heisenberg model : $\mathcal{H} = -\sum_{i,j} J_{ij}\left[s_i^x s_j^x + s_i^y s_j^y + \Delta s_i^z s_j^z\right] - H\sum_i s_i^z$




R. Tamura and K. Hukushima, PLoS ONE 13, e0193785 (2018).

**Time consuming problem will be overcome by using Bayesian optimization.**

# Explanation of COMBO code

ICCMS/2019/2019-10-08/tamura/combo/combo_tutorial.py

# hphi-modeling

ICCMS/2019/2019-10-08/tamura/hphi-modeling/model_estimation.py