Gaussian Process Regression for the Prediction of the Solid-State Transformation Rate

Anna Grebennikova
Supervised by: Clémentine Prieur
University of Grenoble Alpes

May 30, 2025

Motivation

- **Flexible Modeling:** Ideal for building nonlinear models with adjustable complexity.
- **Data Efficiency:** Performs well even with limited experimental data.
- **Low Computational Cost:** Efficient for small-to-moderate datasets.
- **Field Relevance:** Highly suitable for physics problems where complex dependencies exist.
- **Versatility:** Applicable across various fields, including material science, engineering, and environmental modeling.
- **Uncertainty Quantification:** Provides confidence intervals and quantifies prediction uncertainty.

Contents

Gaussian Vectors

Definition:

- A Gaussian vector is defined as:

$$\mathbf{V} = (X_1, X_2, \ldots, X_n)^T \sim \mathcal{N}(\mu, \Sigma),$$

where Σ is symmetric and positive semi-definite.

Joint Probability Density Function:

$$f_{\mathbf{V}}(\mathbf{X}) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\left(-\frac{1}{2}(\mathbf{X} - \mu)^T \Sigma^{-1} (\mathbf{X} - \mu)\right)$$

Gaussian Conditional Theorem (GCT) for $\mathbf{V} = (X_1, X_2)$:

$$\mathbb{E}(X_2 \mid X_1 = x_1) = \mu_2 + \Sigma_{1,2}^\top \Sigma_1^{-1}(x_1 - \mu_1),$$

$$\mathrm{Cov}(X_2 \mid X_1 = x_1) = \Sigma_2 - \Sigma_{1,2}^\top \Sigma_1^{-1} \Sigma_{1,2}.$$

Gaussian Process

Definition:

$$
\begin{aligned}
Z : (\Omega, \mathcal{A}, P) \times [0,1]^d &\to \mathbb{R}, \\
(\omega, \boldsymbol{x}) &\mapsto Z(\omega, \boldsymbol{x}).
\end{aligned}
$$

- Z is a Gaussian process if for all $\boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_n^T \in [0,1]^d$, the function $\omega \mapsto (Z(\omega, \boldsymbol{x}_1), \ldots, Z(\omega, \boldsymbol{x}_n))$ is a Gaussian vector.

Key Components:

- **Mean function:**

$$\mu : [0,1]^d \to \mathbb{R}, \quad \boldsymbol{x} \mapsto \mathbb{E}(Z(\boldsymbol{x})).$$

- **Covariance function (Kernel):**

$$K : [0,1]^d \times [0,1]^d \to \mathbb{R}, \quad (\boldsymbol{x}, \boldsymbol{y}) \mapsto \mathrm{Cov}(Z(\boldsymbol{x}), Z(\boldsymbol{y})).$$

Prediction Using the GCT

Gaussian Conditional Theorem:

$$\begin{pmatrix} Y^{(n)} \\ Y(\boldsymbol{x}) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_y \\ \mu(\boldsymbol{x}) \end{pmatrix}, \begin{pmatrix} R & r(\boldsymbol{x}) \\ r(\boldsymbol{x})^\top & K(\boldsymbol{x}, \boldsymbol{x}) \end{pmatrix} \right), \quad \text{where} \quad Y^{(n)} = \begin{pmatrix} Y(\boldsymbol{x}_1) \\ \vdots \\ Y(\boldsymbol{x}_n) \end{pmatrix}$$

Assumption: $f(\boldsymbol{x})$ which is a realization of $Y^{(n)}$ is a Gaussian Process.

Conditional Expectation:

$$\mathbb{E}[Y(\boldsymbol{x}) \mid Y^{(n)} = y^{(n)}] = \mu(\boldsymbol{x}) + r(\boldsymbol{x})^\top R^{-1}(y^{(n)} - \mu_y)$$

Metamodel:

- The function $\hat{Y}(\boldsymbol{x}) = \mathbb{E}[Y(\boldsymbol{x}) \mid Y^{(n)} = y^{(n)}]$ is a metamodel of $f(\boldsymbol{x})$.
- **Key Advantage:** Computing the metamodel is computationally inexpensive compared to running a full simulation to evaluate $f(\boldsymbol{x})$.

Learning with Gaussian Processes

Key Steps in Learning with GP:

1. **Choosing a Kernel:** Select a covariance function that captures the underlying data structure.
2. **Parameter Estimation:** Optimize kernel parameters using Maximum Likelihood Estimation.
3. **Model Validation:** Assess the model's performance with metrics such as Q^2-score and Leave-One-Out cross-validation.
4. **Optimal Training Data Selection:** Strategically choose data points to train a metamodel.

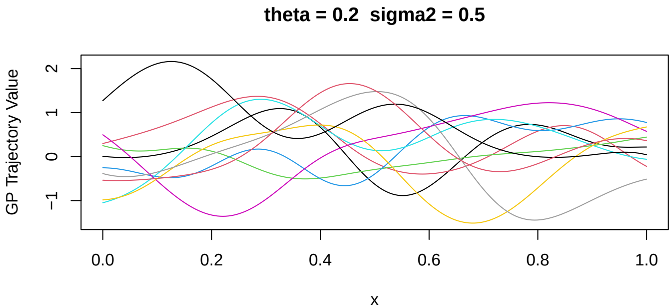
Choosing a Kernel

Gaussian Kernel:

$$k(x, x') := \sigma^2 \exp\left(-\frac{h^2(x, x')}{2\theta^2}\right) \in \mathcal{C}^\infty$$

Key Properties:

- **Smoothness:** Produces infinitely differentiable functions.
- **Applications:** Suitable for modeling smooth and continuous processes.



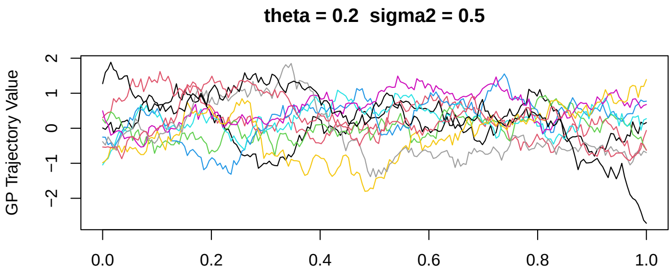
Choosing a Kernel

Exponential Kernel:

$$k(x, x') := \sigma^2 \exp\left(-\frac{|h(x, x')|}{\theta}\right) \in \mathcal{C}^0$$

Key Properties:

- **Smoothness:** Produces less smooth functions, once differentiable *almost everywhere*.
- **Applications:** Suitable for modeling processes with less smooth behavior.



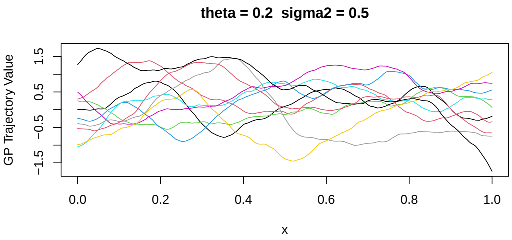
Choosing a Kernel

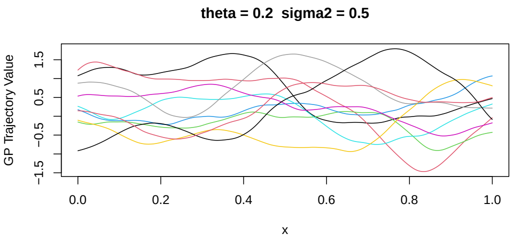
Matern Kernel $\nu = \frac{3}{2}$:

$$k(x,x') := \sigma^2 \left(1 + \frac{\sqrt{3}|h(x,x')|}{\theta}\right) \exp\left(-\frac{\sqrt{3}|h(x,x')|}{\theta}\right) \in \mathcal{C}^1$$

Matern Kernel $\nu = \frac{5}{2}$:

$$k(x,x') := \sigma^2 \left(1 + \frac{\sqrt{5}|h(x,x')|}{\theta} + \frac{5h^2(x,x')}{3\theta^2}\right) \exp\left(-\frac{\sqrt{5}|h(x,x')|}{\theta}\right) \in \mathcal{C}^2$$





Parameter Estimation

The likelihood function is defined as:

$$L(\theta) = \prod_{i=1}^{m} f(X_i; \theta).$$

For Gaussian data with mean μ and covariance $\Sigma(\sigma, \theta)$, the log-likelihood is given by:

$$\ell(\mu, \Sigma) = \sum_{i=1}^{m} \left(-\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln \det(\Sigma) - \frac{1}{2} (X_i - \mu)^T \Sigma^{-1} (X_i - \mu) \right).$$

The parameters μ, σ, and θ are estimated by maximizing the log-likelihood:

$$\hat{\mu}, \hat{\sigma}, \hat{\theta} \in \arg\max_{\mu,\sigma,\theta} \ell(\mu, \Sigma(\sigma, \theta)).$$

Model Validation

The Q^2-score is defined as:

$$Q^2 = 1 - \frac{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}.$$

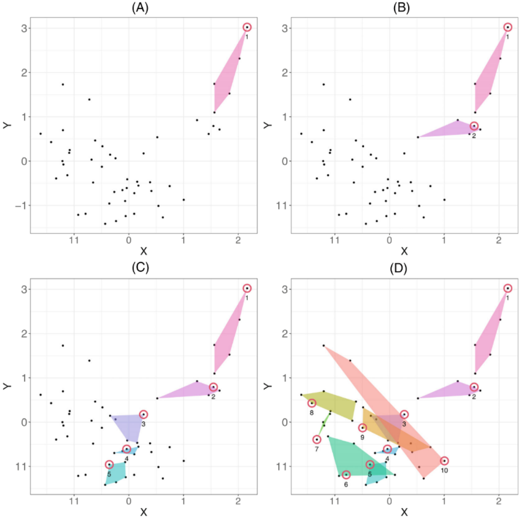
Validation methods:

- Train-test validation.
- Leave-One-Out: each point was excluded, predicted, and the Q^2-score calculated.

Twinning Sampling

Energy Distance (to be minimized):

$$
\begin{aligned}
\overline{\mathbb{ED}}_{n,N-n} = {} & \frac{2}{n(N-n)} \sum_{i=1}^{n} \sum_{j=1}^{N-n} \| \mathrm{U}_i - \mathrm{V}_j \|_2 \\
& - \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \| \mathrm{U}_i - \mathrm{U}_j \|_2 \\
& - \frac{1}{(N-n)^2} \sum_{i=1}^{N-n} \sum_{j=1}^{N-n} \| \mathrm{V}_i - \mathrm{V}_j \|_2 .
\end{aligned}
$$



Adaptive Learning

Concept: The point with the highest Mean Squared Error (MSE) is identified, and an additional point is added around it. The model is then retrained iteratively.

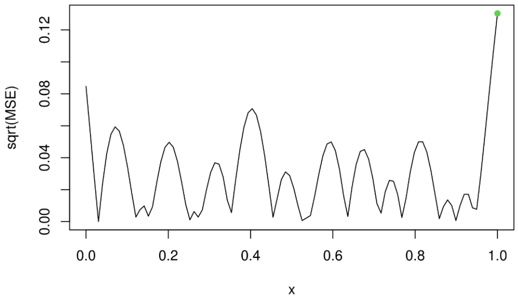


Figure: First Iteration

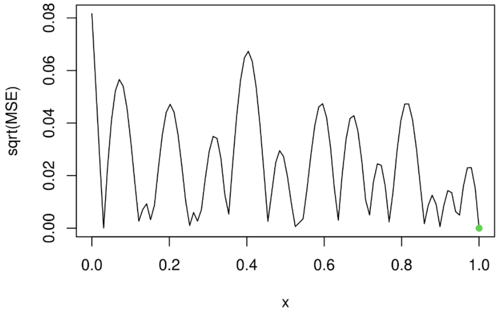


Figure: Second Iteration

Latent Variable Gaussian Process

We consider a Gaussian Process (GP) with $\boldsymbol{x} = (x_1, ..., x_p)$ as quantitative variables and $\boldsymbol{t} = (t_1, ..., t_q)$ as qualitative factors.
Qualitative variables are mapped to a latent space:

$$(\boldsymbol{x}, \boldsymbol{t}) \to (\boldsymbol{x}, \boldsymbol{z}(\boldsymbol{t}))$$

Mapping to Latent Space:
$\boldsymbol{t} \to \boldsymbol{z}(\boldsymbol{t}) \in$ latent space $\mathbb{R}^d$

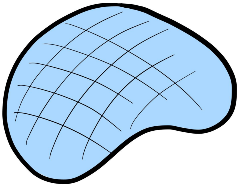


Figure: Latent Variable Space

For $d = 2$, the mapping produces m points in the 2D latent variable space:

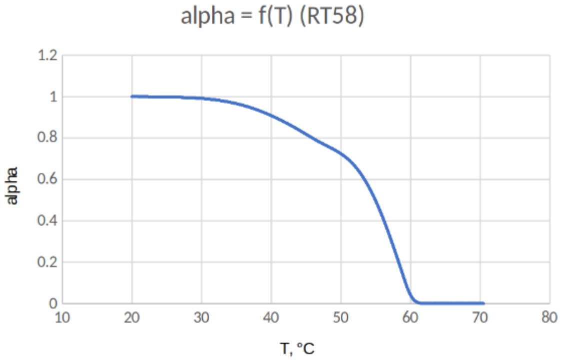
$$\{\boldsymbol{z}(1) = (z_1(1), z_2(1)), \ldots, \boldsymbol{z}(m) = (z_1(m), z_2(m))\}$$

This latent representation captures qualitative differences as continuous features.

Experiment Description

- Data provided by Gilles Fraisse from the Laboratory of Energy Processes and Building Materials (LOCIE).
- Data include t (time), T (temperature), and α (solid-state fraction), collected using a scanning calorimeter.
- The solid-state transformation rate $d\alpha/dt$ was the primary focus of analysis.

It is useful for determining heat capacity during phase transitions, particularly for building materials.



Reference Formula

$$\frac{d\alpha}{dt} = -\beta^{n_4} \cdot \mathrm{sign}(\beta) \cdot \left\{ k_1 \frac{\alpha^{m_1}(\alpha_{\mathrm{max}} - \alpha)^{n_1}}{\alpha_{\mathrm{max}}^{m_1+n_1}} + k_2 \frac{(\alpha - \alpha_{\mathrm{min}})^{m_2}(1-\alpha)^{n_2}}{(1-\alpha_{\mathrm{min}})^{m_2+n_2}} \right\} \cdot \left(1 - k_3 \frac{d\beta}{dt}\right)^{n_3}$$

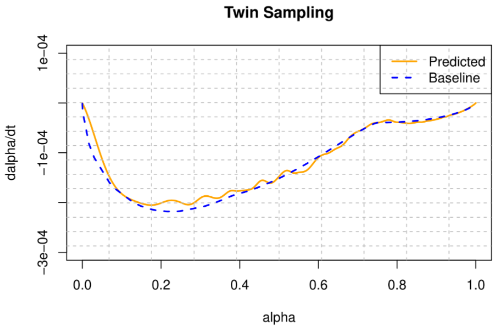
Where: $\beta = \frac{dT}{dt}$

- Parameters of the formula: $(n_1, n_2, n_3, k_1, k_2, k_3, m_1, m_2, \alpha_{\mathrm{min}}, \alpha_{\mathrm{max}})$
- Obtained using Excel Solver's *GRG Nonlinear* method by Gilles Fraisse.

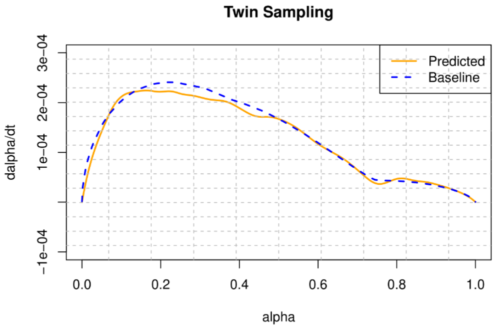
Twinning Sampling Results

The models to be built are as follows:

$$\left(\alpha, \beta, \frac{d\beta}{dt}, \text{sign}(\beta) = 1\right) \rightarrow \frac{d\alpha}{dt} \quad \text{and} \quad \left(\alpha, \beta, \frac{d\beta}{dt}, \text{sign}(\beta) = -1\right) \rightarrow \frac{d\alpha}{dt}$$

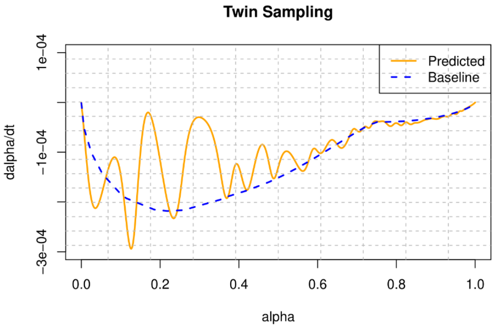


275 observations; $\beta > 0$
$Q^2 = 0.9980,\ Q^2_{\mathrm{LOO}} = 0.9424$

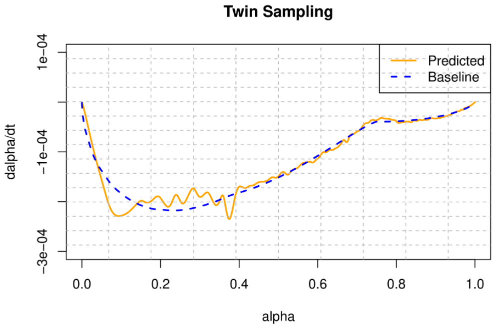


417 observations; $\beta < 0$
$Q^2 = 0.9986,\ Q^2_{\mathrm{LOO}} = 0.9965$

Twinning Sampling Simulation Challenges



138 observations; $\beta > 0$
$Q^2 = 0.8813,\ Q^2_{\mathrm{LOO}} = 0.8381$

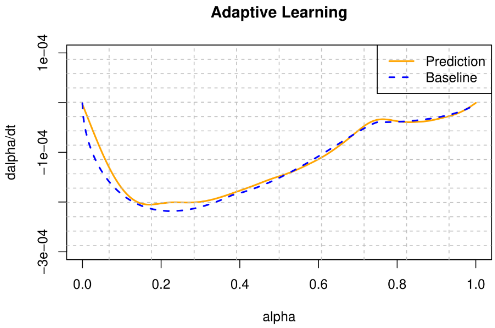


412 observations; $\beta > 0$
$Q^2 = 0.9901,\ Q^2_{\mathrm{LOO}} = 0.9886$

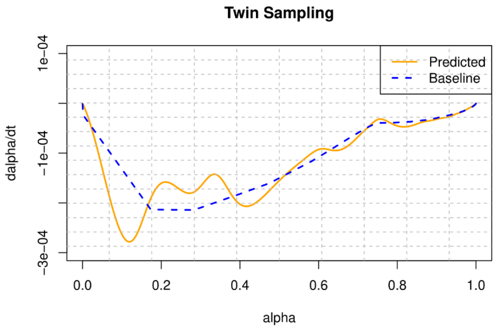
Figure: Examples of less accurate simulations with fixed seed `set.seed(123)` for $\beta > 0$.

Adaptive Learning Results

The process began with 15 points, with 35 points added iteratively.



50 observations; $\beta > 0$
$Q^2 = 0.9955$, $Q^2_{\text{LOO}} = 0.9529$.

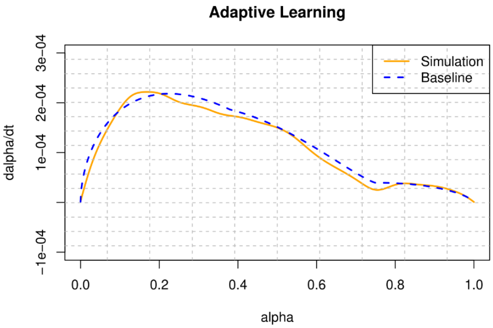


50 observations; $\beta > 0$
$Q^2 = 0.9377$, $Q^2_{\text{LOO}} = 0.8471$.

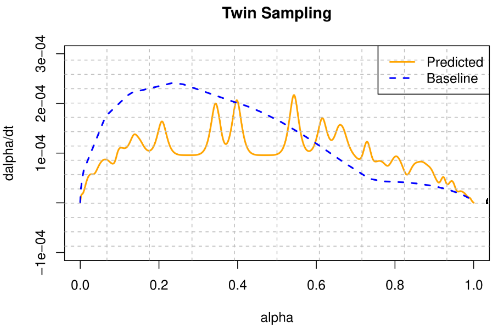
Figure: Comparison of adaptive learning and twinning sampling; $\beta > 0$; `set.seed(123)`; 50 observations.

Adaptive Learning Results

The process began with 15 points, with 65 points added iteratively.



80 observations; $\beta < 0$
$Q^2 = 0.9980$, $Q^2_{\mathrm{LOO}} = 0.9973$

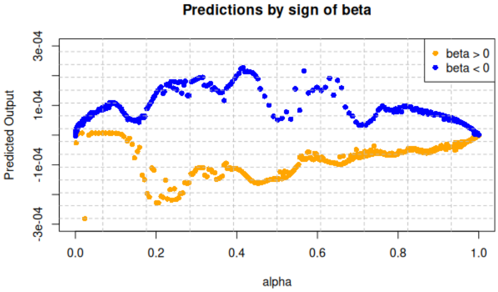


80 observations; $\beta < 0$
$Q^2 = 0.8315$, $Q^2_{\mathrm{LOO}} = 0.8498$.

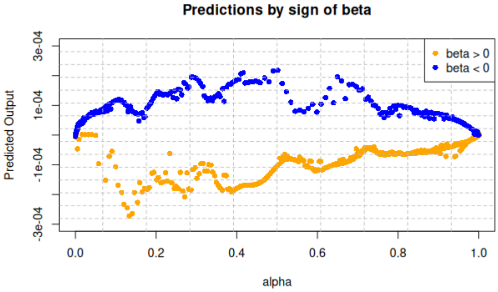
Figure: Comparison of adaptive learning and twinning sampling; $\beta < 0$; `set.seed(123)`; 80 observations.

Latent Variable GP Results

Several experiments were conducted using varying numbers of points selected via twinning sampling to train the model, followed by testing on 1500 points.

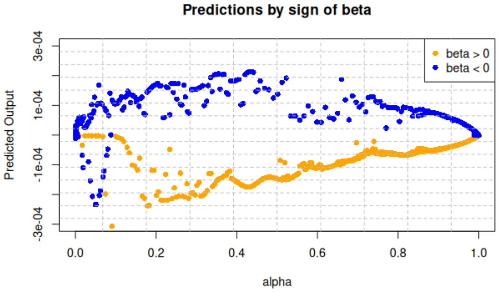


Model prediction with 208 training observations; Training time: 6 min 49 sec; $Q^2 = 0.7871$

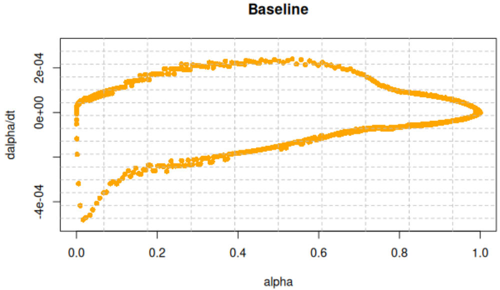


Model prediction with 277 training observations; Training time: 9 min 8 sec; $Q^2 = 0.9323$

Latent Variable GP Results



Model prediction with 415 training observations;
Training time: 20 min 34 sec; $Q^2 = 0.8668$



Results computed with Gilles' formula

Oscillations may appear as a result of noisy data when errors are correlated within two groups of β.

Conclusion

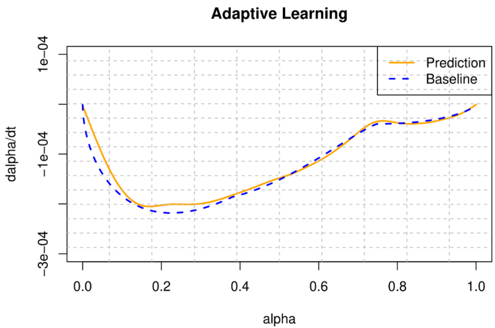
- Twinning Sampling is sensitive to the number of training points used.
- Adaptive Learning provides more reliable results, though it is computationally more expensive.
- The LVGP model with Twinning Sampling yields oscillations in the results, which could be a consequence of error correlation within the model.
- The model obtained via Adaptive Learning demonstrates potential for application in studying more complex materials.

Perspectives

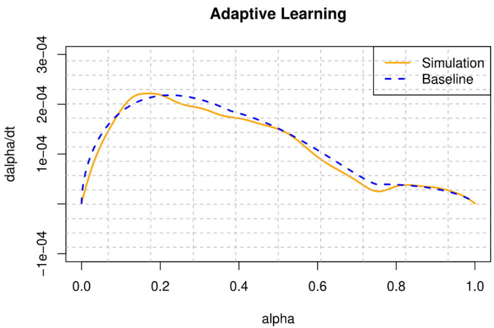
- Implement adaptive learning using the LVGP model.
- Estimate the nugget effect in the LVGP model to address oscillations.
- Deliver the obtained results to colleagues at LOCIE.
- Apply the methodology to more complex materials.
- Develop a custom parameter estimation function to build a model incorporating binary variables, as an alternative to the LVGP model.

Questions and Discussion

Best results obtained via adaptive learning:



50 observations; $\beta > 0$
$Q^2 = 0.9955,\ Q^2_{\mathrm{LOO}} = 0.9529.$



80 observations; $\beta < 0$
$Q^2 = 0.9980,\ Q^2_{\mathrm{LOO}} = 0.9973.$

Thank you for your attention!