**OBJECTIVES**

Using surrogate model (SM) we seek to mimic the input-output relationship, \( x \rightarrow y \), of computationally expensive simulations in polymer rheology. SM, \( y = \hat{m}(x) \), where \( m : \mathbb{R}^d \rightarrow \mathbb{R}^n \), is considered here, with \( x \) representing the structure of the polymer mixture, and the output \( y \) representing the rheology. The things we considered when building the SM:

1. Computational Cost
2. Prediction Accuracy
3. Functional Output

**GAUSSIAN PROCESS**

A GP \( f(x) \) with \( x \in \mathbb{R} \) is a collection of random variables, any finite number of which have a joint Gaussian distribution\([1]\). GP can be represented as,

\[
f(x) \sim \mathcal{N}(\mu(x), k(x, x'))
\]

Given \( n \) observations \( f = [f_1, \ldots, f_n]^T \) at training points \( X = [x_1, \ldots, x_n] \), we first build a \( n \times n \) co-variance matrix \( K_{ij} = k(x_i, x_j) \). The best hyper-parameters \( \hat{\theta} \) is obtained by maximizing the log-likelihood:

\[
\mathcal{L}(\theta) = -\frac{1}{2} f^T K^{-1} f - \frac{1}{2} \log |K| - \frac{n}{2} \log 2\pi
\]

Given test point \( x_* \), the prediction \( \hat{f}_* \) from the posterior is normally distributed, with mean and variance of this distribution given by:

\[
\hat{f}_* = \mathbb{E}[f(x_*)] = K^T [K(\theta)]^{-1} f
\]

\[
\sigma^2_* = \text{Var}[f(x_*)] = k(x_*, x_*) - K^T [K(\theta)]^{-1} k(x_*)
\]

Because of matrix inversion GP scale as \( O(n^3) \).

**MODEL PIPELINE**

(a) model

\[
x \quad \text{TDD-DR} \quad \phi \quad \text{pyReSpect} \quad h
\]

(b) training

\[
x_1 \ x_2 \ \cdots \ x_n \ \text{dist} \quad \text{minimize} \quad \text{MLE} \quad \hat{\theta}
\]

(c) surrogate model

\[
x' \quad \text{GP-SM(\hat{\theta})} \quad h^*
\]

Figure 2: Schematic diagram showing (a) the true model \( m(x) \) which takes in blend information in the form of \( x = [x_1, x_2, \rho_1, \rho_2, \phi_1] \) and yields the spectrum \( h(s) \) using the TDD-DR model and pyReSpect; (b) training data generated using \( m(x) \) is used to fit GP-SM hyperparameters \( \theta \) using maximum likelihood estimation; (c) the trained GP-SM model can be used in lieu of the forward model at arbitrary input points \( x_*. \)

**REFERENCES**


**FUTURE RESEARCH**

Explore KL expansion, which split the signal into empirical mean and linear combination of eigen-decomposition of co-variance matrix. This method does not require the assumption of stationarity and is more computationally efficient as compared to using separable kernel.

**RESULTS**

Figure 3: (a) GP-SM prediction when \( x_\ast = (31.27, 10.45, 19.1, 38.0, 0.33) \) is selected as one of the training point co-incides with the true model prediction. (b) This is not true for an unseen test point \( x_\ast = (43.0, 18.1, 15.1, 34.0, 0.50) \). The \( \pm 2.5\sigma \) uncertainty interval is shown by the shaded region. (c) The \( \phi(t) \) generated using the relaxation spectra predicted by the GP-SM from (b) is compared with the true \( \phi(t) \).

Figure 4: Variation of (a) MAD, and (b) RMSE with training sample size \( n \) represented using box-plots. The number of test data is fixed to \( n_{\text{test}} = 251 \). The median values of both MAD and RMSE decrease with \( n \) indicating that the predictive ability of the GP-SM improves as it is trained on more data. The diamond corresponds to outliers where the model performed worse.

Figure 5: Variation of MLE estimates \( \hat{\theta} \) with training data size \( n \). Components of \( \theta \) with comparable magnitudes are grouped together.