CRPS-Based Targeted Sequential Design with Application in Chemical Space

Lea Friedli

Engineering Risk Analysis Group

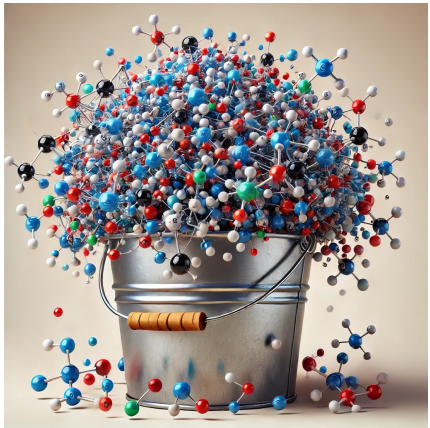
Technical University of Munich

Joint work with A. Gautier, A. Broccard and D. Ginsbourger

mODa14

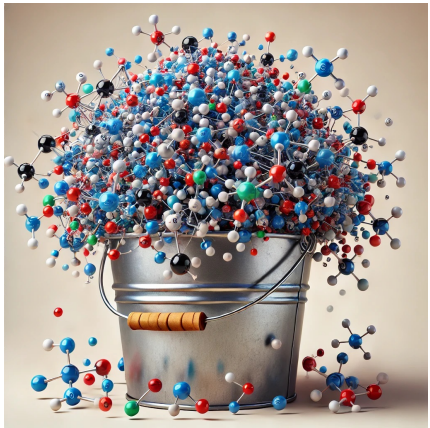
June 17, 2026

Navigating molecular spaces



- Identify molecules with desired properties

Navigating molecular spaces



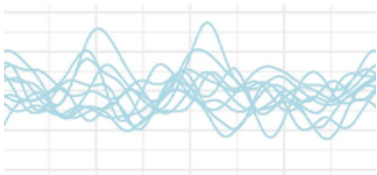
- Identify molecules with desired properties
- Predictive models to prioritize promising candidates

GP modeling

Model function $f : \mathcal{X} \to \mathbb{R}$ based on finite number of pointwise evaluations

GP modeling

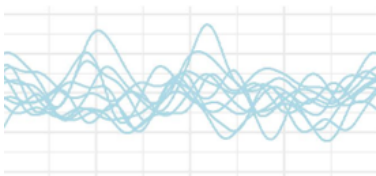
Model function $f : \mathcal{X} \to \mathbb{R}$ based on finite number of pointwise evaluations



- Assume f is realization of GP $\xi = (\xi(\mathbf{x}))_{\mathbf{x}\in\mathcal{X}} \sim \mathrm{GP}(m, k)$

GP modeling

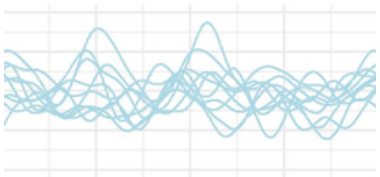
Model function $f : \mathcal{X} \to \mathbb{R}$ based on finite number of pointwise evaluations

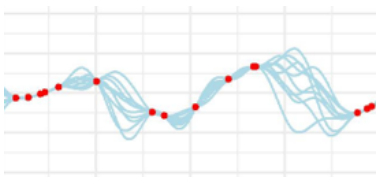


- Assume f is realization of GP $\xi = (\xi(\boldsymbol{x}))_{\boldsymbol{x} \in \mathcal{X}} \sim \mathrm{GP}(m, k)$
- Noisy observation event $\mathcal{A}_n = \{\boldsymbol{Z}_n = \boldsymbol{z}_n\}$ with $Z_i = f(\boldsymbol{x}_i) + \varepsilon_i$

GP modeling

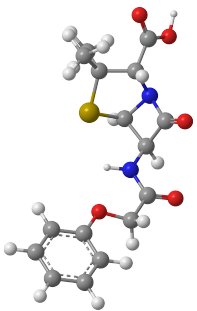
Model function $f : \mathcal{X} \to \mathbb{R}$ based on finite number of pointwise evaluations



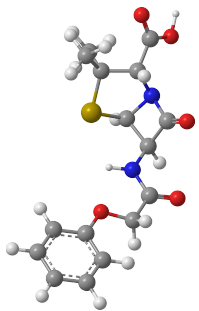


- Assume f is realization of GP $\xi = (\xi(\boldsymbol{x}))_{\boldsymbol{x}\in\mathcal{X}} \sim \mathrm{GP}(m, k)$
- Noisy observation event $\mathcal{A}_n = \{\boldsymbol{Z}_n = \boldsymbol{z}_n\}$ with $Z_i = f(\boldsymbol{x}_i) + \varepsilon_i$
- ξ conditional on $\mathcal{A}_n$: $\xi_n = (\xi_n(\boldsymbol{x}))_{\boldsymbol{x}\in\mathcal{X}} \sim \mathrm{GP}\,(m_n, k_n)$

Kernels for molecules



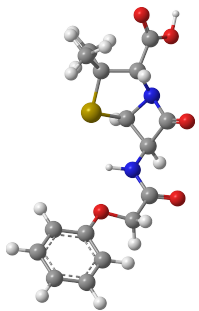
Kernels for molecules



Fingerprints

Presence and absence of substructures $\rightsquigarrow (1, 0, \cdots, 1)$

Kernels for molecules



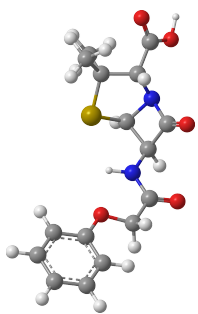
Fingerprints

Presence and absence of substructures $\rightsquigarrow (1, 0, \cdots, 1)$

Tanimoto kernel [4]

Counts the number of common present features normalized by the total number of features present.

Kernels for molecules



Fingerprints

Presence and absence of substructures $\rightsquigarrow (1, 0, \cdots, 1)$

Tanimoto kernel [4]

Counts the number of common present features normalized by the total number of features present.

For $\boldsymbol{x}, \boldsymbol{x'} \in \{0,1\}^d, d \geq 1$:

$$k(\boldsymbol{x}, \boldsymbol{x'}) := \begin{cases} \sigma_k^2 & \text{if } \langle \boldsymbol{x}, \boldsymbol{x} \rangle = \langle \boldsymbol{x'}, \boldsymbol{x'} \rangle = 0, \\ \sigma_k^2 \frac{\langle \boldsymbol{x}, \boldsymbol{x'} \rangle}{||\boldsymbol{x}||^2 + ||\boldsymbol{x'}||^2 - \langle \boldsymbol{x}, \boldsymbol{x'} \rangle} & \text{else,} \end{cases}$$

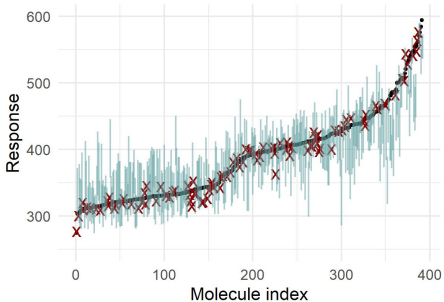
where $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = \boldsymbol{x}^\top \boldsymbol{x}$ and $\sigma_k^2 > 0$.

Photoswitch dataset [6]

- Molecules undergoing reversible transformation when exposed to light
- Input: Morgan fingerprint, response: transition wavelength

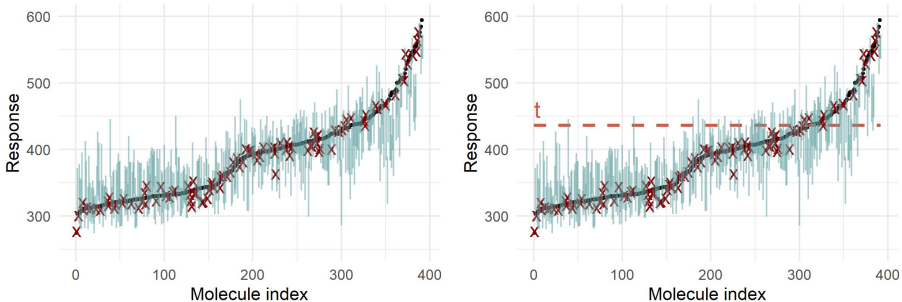
Photoswitch dataset [6]

- Molecules undergoing reversible transformation when exposed to light
- Input: Morgan fingerprint, response: transition wavelength



Photoswitch dataset [6]

- Molecules undergoing reversible transformation when exposed to light
- Input: Morgan fingerprint, response: transition wavelength



- Excursion set of f above $t \in \mathbb{R}$: $\Gamma := \{\boldsymbol{x} \in \mathcal{X} : f(x) \geq t\}$

Scoring Rules

- A scoring rule $\mathcal{S} : \mathcal{F} \times \Omega \to \mathbb{R}$ represents the **loss** when the **predictive distribution** $F \in \mathcal{F}$ is issued and the observation $y \in \Omega$ materializes

Scoring Rules

- A scoring rule $\mathcal{S} : \mathcal{F} \times \Omega \to \mathbb{R}$ represents the **loss** when the **predictive distribution** $F \in \mathcal{F}$ is issued and the observation $y \in \Omega$ materializes
- Scoring rules evaluate both **calibration** and **sharpness** of the prediction

Scoring Rules

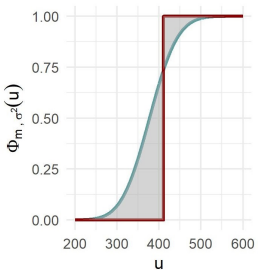
- A scoring rule $\mathcal{S} : \mathcal{F} \times \Omega \to \mathbb{R}$ represents the **loss** when the **predictive distribution** $F \in \mathcal{F}$ is issued and the observation $y \in \Omega$ materializes
- Scoring rules evaluate both **calibration** and **sharpness** of the prediction
- A scoring rule is **proper** relative to $\mathcal{F}$ if no alternative forecast distribution $F \in \mathcal{F}$ can yield a lower expected score than the true data-generating distribution $Q \in \mathcal{F}$ [2]:

$$\mathbb{E}_{Y \sim Q}\left[\mathcal{S}(Q, Y)\right] \leq \mathbb{E}_{Y \sim Q}\left[\mathcal{S}(F, Y)\right]$$

Continuous Ranked Probability Score

- **Continuous Ranked Probability Score** [CRPS; 5] for CDF $F(\cdot)$ and y:

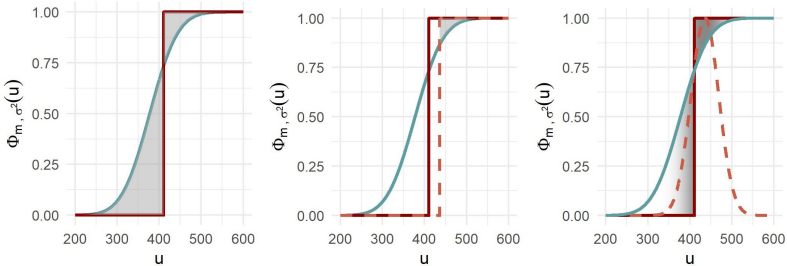
$$\mathrm{CRPS}(F, y) = \int_{-\infty}^{\infty} [F(u) - \mathbb{1}\{y \leq u\}]^2 \, \mathrm{d}u.$$



Threshold-weighted CRPS

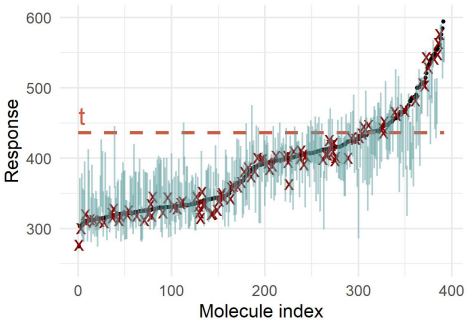
- Scoring rules [2]: Alignment predictive distribution and materialized value
- **Threshold-weighted CRPS** [5, 3] with weighting measure ν:

$$\mathrm{CRPS}_\nu(F, y) = \int_{-\infty}^{\infty} [F(u) - \mathbb{1}\{y \leq u\}]^2 \, \mathrm{d}\nu(u).$$

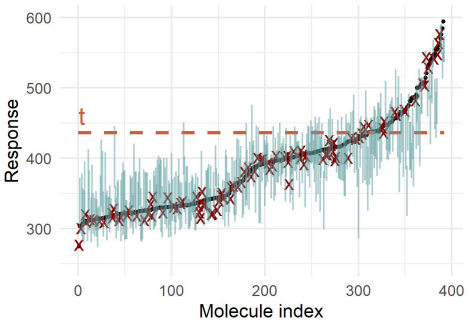


$$\mathrm{d}\nu_1(u) = \mathbb{1}\{u \geq t\}\mathrm{d}\lambda(u) \quad \text{and} \quad \mathrm{d}\nu_2(u) = \tfrac{1}{\sigma_\nu}\phi\left(\tfrac{u-t}{\sigma_\nu}\right)\mathrm{d}\lambda(u)$$

From scoring to sequential design



From scoring to sequential design



One-step look-ahead (myopic) strategy:

- Current GP $\boldsymbol{\xi_n} \sim \mathrm{GP}(m_n, k_n)$ with inputs $\boldsymbol{x_1}, \ldots, \boldsymbol{x_n} \in \mathcal{X}$
- Apply selection criterion to choose $\boldsymbol{x_{n+1}}$
- Collect Z_{n+1} and update GP to $\boldsymbol{\xi_{n+1}} \sim \mathrm{GP}(m_{n+1}, k_{n+1})$

Targeted sequential design

- Select next $\boldsymbol{x_{n+i}}$ such that it reduces the uncertainty of the future GP $\boldsymbol{\xi_{n+i}}$, especially in (proximity to) the excursion set Γ

Targeted sequential design

- Select next $\boldsymbol{x}_{n+i}$ such that it reduces the uncertainty of the future GP $\boldsymbol{\xi}_{n+i}$, especially in (proximity to) the excursion set Γ

  $\rightarrow$ Molecule with high $\mathrm{CRPS}_\nu(F, y)$

Targeted sequential design

- Select next $\boldsymbol{x}_{n+i}$ such that it reduces the uncertainty of the future GP $\boldsymbol{\xi}_{n+i}$, especially in (proximity to) the excursion set Γ

  $\rightarrow$ Molecule with high $\mathrm{CRPS}_{\nu}(F, y)$

- During acquisition no materialized values y for candidates

Targeted sequential design

- Select next $\boldsymbol{x}_{n+i}$ such that it reduces the uncertainty of the future GP $\boldsymbol{\xi}_{n+i}$, especially in (proximity to) the excursion set Γ

  $\rightarrow$ Molecule with high $\mathrm{CRPS}_\nu(F, y)$

- During acquisition no materialized values y for candidates

- Expected version of the threshold-weighted CRPS [also called entropy; 2]

$$\mathrm{CRPS}_\nu(F) = \int_{-\infty}^{\infty} \mathrm{CRPS}_\nu(F, y') \, \mathrm{d}F(y') = \int F(u) \, (1 - F(u)) \, \mathrm{d}\nu(u),$$

  with weighting measure $\nu(\cdot)$ and predictive CDF $F(\cdot)$.

Expected threshold-weighted CRPS

For $F = \Phi_{\mu,\sigma^2}$, $\mathrm{d}\nu_1(u) = \mathbb{1}\{t \leq u\}\mathrm{d}\lambda(u)$ and $\widetilde{t} = (t-\mu)/\sigma$, one obtains:

$$\mathrm{CRPS}_{\nu_1}(\Phi_{\mu,\sigma^2}) = \sigma\left[\widetilde{t}\Phi(\widetilde{t})^2 - \widetilde{t}\Phi(\widetilde{t}) + \frac{1}{\sqrt{\pi}} - \frac{1}{\sqrt{\pi}}\Phi(\widetilde{t}\sqrt{2}) + 2\phi(\widetilde{t})\Phi(\widetilde{t}) - \phi(\widetilde{t})\right]$$

Expected threshold-weighted CRPS

For $F = \Phi_{\mu,\sigma^2}$, $\mathrm{d}\nu_1(u) = \mathbb{1}\{t \leq u\}\mathrm{d}\lambda(u)$ and $\widetilde{t} = (t-\mu)/\sigma$, one obtains:

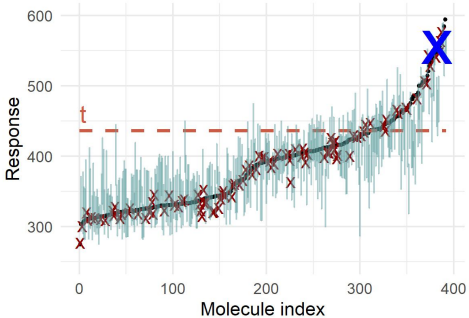
$$\mathrm{CRPS}_{\nu_1}(\Phi_{\mu,\sigma^2}) = \sigma\left[\widetilde{t}\Phi(\widetilde{t})^2 - \widetilde{t}\Phi(\widetilde{t}) + \frac{1}{\sqrt{\pi}} - \frac{1}{\sqrt{\pi}}\Phi(\widetilde{t}\sqrt{2}) + 2\phi(\widetilde{t})\Phi(\widetilde{t}) - \phi(\widetilde{t})\right]$$

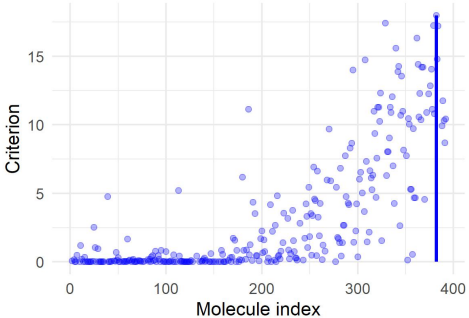
For $\mathrm{d}\nu_2(u) = \frac{1}{\sigma_\nu}\phi\left(\frac{u-t}{\sigma_\nu}\right)\mathrm{d}\lambda(u)$, we get:

$$\mathrm{CRPS}_{\nu_2}(\Phi_{\mu,\sigma^2}) = \Phi\left(\begin{pmatrix}0\\0\end{pmatrix}; \begin{pmatrix}\mu - t\\ t - \mu\end{pmatrix}, \begin{pmatrix}\sigma_\nu^2 + \sigma^2 & -\sigma_\nu^2\\ -\sigma_\nu^2 & \sigma_\nu^2 + \sigma^2\end{pmatrix}\right)$$

Select the next molecule

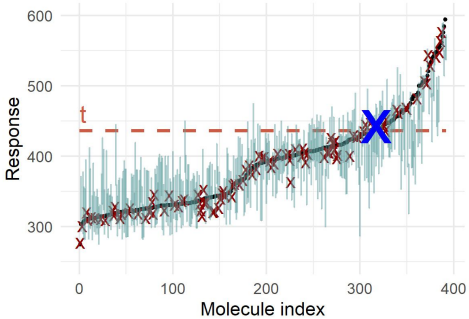
Pointwise CRPS_{ν_1}

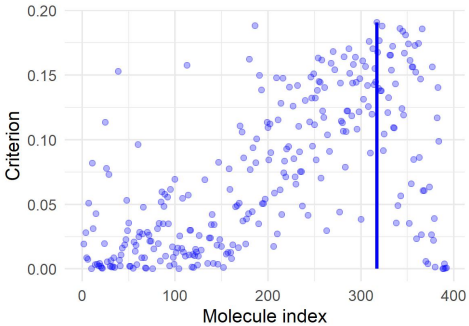




Select the next molecule

Pointwise CRPS_{ν_2}





Sports break: Decathlon

Day 1

- 100 meters
- Discus throw
- Pole vault
- Javelin throw
- 400 meters

Day 2

- 100/110 meters hurdles
- Long jump
- Shot put
- High jump
- 1500 meters

Sports break: Decathlon

Day 1

- 100 meters
- Discus throw
- Pole vault
- Javelin throw
- 400 meters

WHAT SHOULD YOU TRAIN NEXT?

Day 2

- 100/110 meters hurdles
- Long jump
- Shot put
- High jump
- 1500 meters

Sports break: Decathlon

Day 1

- 100 meters
- Discus throw
- Pole vault
- Javelin throw
- 400 meters

Day 2

- 100/110 meters hurdles
- Long jump
- Shot put
- High jump
- 1500 meters

WHAT SHOULD YOU TRAIN NEXT?

Pole vault?

Sports break: Decathlon

Day 1

- 100 meters
- Discus throw
- Pole vault
- Javelin throw
- 400 meters

Day 2

- 100/110 meters hurdles
- Long jump
- Shot put
- High jump
- 1500 meters

WHAT SHOULD YOU TRAIN NEXT?

Pole vault?

$\rightarrow$ **Pointwise criteria**
Marginal distributions $\xi_n(\boldsymbol{x})$

Sports break: Decathlon

Day 1

- 100 meters
- Discus throw
- Pole vault
- Javelin throw
- 400 meters

Day 2

- 100/110 meters hurdles
- Long jump
- Shot put
- High jump
- 1500 meters

WHAT SHOULD YOU TRAIN NEXT?

Pole vault?

$\rightarrow$ **Pointwise criteria**
Marginal distributions $\xi_n(\boldsymbol{x})$

Running?

Sports break: Decathlon

Day 1

- 100 meters
- Discus throw
- Pole vault
- Javelin throw
- 400 meters

Day 2

- 100/110 meters hurdles
- Long jump
- Shot put
- High jump
- 1500 meters

WHAT SHOULD YOU TRAIN NEXT?

Pole vault?

$\rightarrow$ **Pointwise criteria**
Marginal distributions $\xi_n(\boldsymbol{x})$

Running?

$\rightarrow$ **Stepwise Uncertainty Reduction** [SUR; 1]
Uncertainty over entire domain

SUR criteria

Given a GP model ξ_n at time n, select next acquisition point by minimizing

$$\boldsymbol{x} \mapsto J_n(\boldsymbol{x}) = \mathbb{E}_{n,\boldsymbol{x}}[\mathcal{H}(\xi_{n+1})]$$

SUR criteria

Given a GP model ξ_n at time n, select next acquisition point by minimizing

$$\boldsymbol{x} \mapsto J_n(\boldsymbol{x}) = \mathbb{E}_{n,\boldsymbol{x}}[\mathcal{H}(\xi_{n+1})]$$

- $\xi_{n+1} \sim \mathrm{GP}(m_{n+1}, k_{n+1})$ is the future GP conditioned on $\mathcal{A}_{n+1}$

SUR criteria

Given a GP model ξ_n at time n, select next acquisition point by minimizing

$$\boldsymbol{x} \mapsto J_n(\boldsymbol{x}) = \mathbb{E}_{n,\boldsymbol{x}}[\mathcal{H}(\xi_{n+1})]$$

- $\xi_{n+1} \sim \mathrm{GP}(m_{n+1}, k_{n+1})$ is the future GP conditioned on $\mathcal{A}_{n+1}$
- $\mathcal{H}(\cdot)$: uncertainty functional mapping the whole (updated) GP to real-value

SUR criteria

Given a GP model ξ_n at time n, select next acquisition point by minimizing

$$\boldsymbol{x} \mapsto J_n(\boldsymbol{x}) = \mathbb{E}_{n,\boldsymbol{x}}[\mathcal{H}(\xi_{n+1})]$$

- $\xi_{n+1} \sim \mathrm{GP}(m_{n+1}, k_{n+1})$ is the future GP conditioned on $\mathcal{A}_{n+1}$
- $\mathcal{H}(\cdot)$: uncertainty functional mapping the whole (updated) GP to real-value
- $\mathbb{E}_{n,\boldsymbol{x}}(\cdot)$ denotes the conditional expectation with respect to ξ_n when $\boldsymbol{x}_{n+1} = \boldsymbol{x}$

SUR criteria with CRPS_ν

$$J_n(\boldsymbol{x}) = \mathbb{E}_{n,\boldsymbol{x}}\left[\int \mathrm{CRPS}_\nu(\xi_{n+1}(\boldsymbol{x'}))\mathbb{P}_{\mathcal{X}}(\mathrm{d}\boldsymbol{x'})\right] = \int \mathbb{E}_{n,\boldsymbol{x}}\left[\mathrm{CRPS}_\nu(\xi_{n+1}(\boldsymbol{x'}))\right]\mathbb{P}_{\mathcal{X}}(\mathrm{d}\boldsymbol{x'})$$

SUR criteria with CRPS_ν

$$J_n(\boldsymbol{x}) = \mathbb{E}_{n,\boldsymbol{x}}\left[\int \mathrm{CRPS}_\nu(\xi_{n+1}(\boldsymbol{x}'))\mathbb{P}_{\mathcal{X}}(\mathrm{d}\boldsymbol{x}')\right] = \int \mathbb{E}_{n,\boldsymbol{x}}\left[\mathrm{CRPS}_\nu(\xi_{n+1}(\boldsymbol{x}'))\right]\mathbb{P}_{\mathcal{X}}(\mathrm{d}\boldsymbol{x}')$$

- Fix $\boldsymbol{x}$ and $\boldsymbol{x}'$: marginal Gaussians $\xi_{n+1}(\boldsymbol{x}') \sim \mathcal{N}(m_{n+1}(\boldsymbol{x}'), k_{n+1}(\boldsymbol{x}', \boldsymbol{x}'))$,

SUR criteria with CRPS_ν

$$J_n(\boldsymbol{x}) = \mathbb{E}_{n,\boldsymbol{x}}\left[\int \mathrm{CRPS}_\nu(\xi_{n+1}(\boldsymbol{x}'))\mathbb{P}_{\mathcal{X}}(\mathrm{d}\boldsymbol{x}')\right] = \int \mathbb{E}_{n,\boldsymbol{x}}\left[\mathrm{CRPS}_\nu(\xi_{n+1}(\boldsymbol{x}'))\right]\mathbb{P}_{\mathcal{X}}(\mathrm{d}\boldsymbol{x}')$$

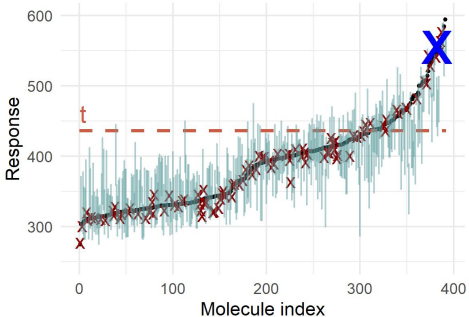
- Fix $\boldsymbol{x}$ and $\boldsymbol{x}'$: marginal Gaussians $\xi_{n+1}(\boldsymbol{x}') \sim \mathcal{N}(m_{n+1}(\boldsymbol{x}'), k_{n+1}(\boldsymbol{x}', \boldsymbol{x}'))$,
- With $\sigma^2_{n+1} = k_{n+1}(\boldsymbol{x}', \boldsymbol{x}')$, $\mu_n = m_n(\boldsymbol{x}')$, $\alpha_n = \frac{k_n(\boldsymbol{x}, \boldsymbol{x}')}{\sqrt{k_n(\boldsymbol{x}, \boldsymbol{x}) + \tau^2}}$ and $N, \widetilde{N}, V \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, we obtain:

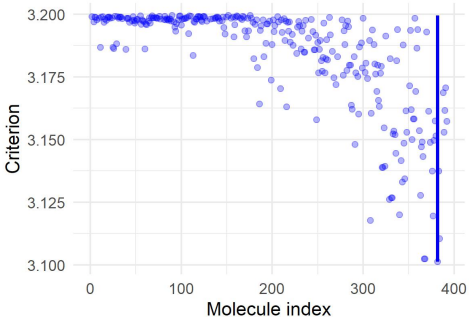
$$\mathbb{E}_{n,\boldsymbol{x}}\left[\mathrm{CRPS}_{\nu_1}(\xi_{n+1}(\boldsymbol{x}'))\right] = \sigma_{n+1}\,\mathbb{E}\left[\left(N - \max\left(\widetilde{N}, \frac{t - \mu_n - \alpha_n V}{\sigma_{n+1}}\right)\right)_+\right]$$

$$\mathbb{E}_{n,\boldsymbol{x}}\left[\mathrm{CRPS}_{\nu_2}(\xi_{n+1}(\boldsymbol{x}'))\right] = \Phi\left(\begin{pmatrix}0\\0\end{pmatrix}; \begin{pmatrix}\mu_n - t\\ t - \mu_n\end{pmatrix}, \begin{pmatrix}\sigma^2_{n+1} + \alpha^2_n + \sigma^2_\nu & -\alpha^2_n - \sigma^2_\nu\\ -\alpha^2_n - \sigma^2_\nu & \sigma^2_{n+1} + \alpha^2_n + \sigma^2_\nu\end{pmatrix}\right)$$

Select the next molecule

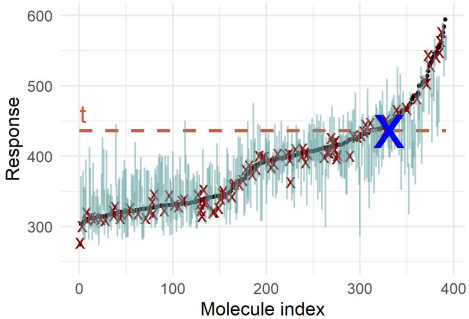
SUR CRPS_{ν_1}

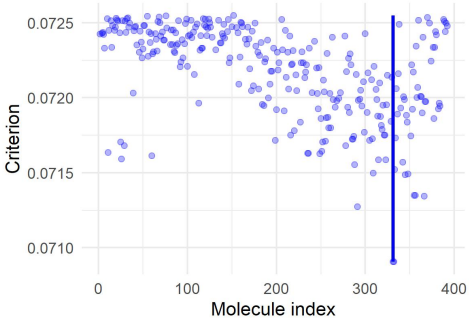




Select the next molecule

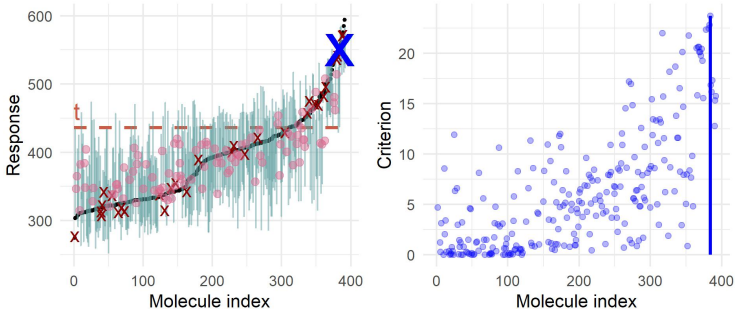
SUR CRPS_{ν_2}



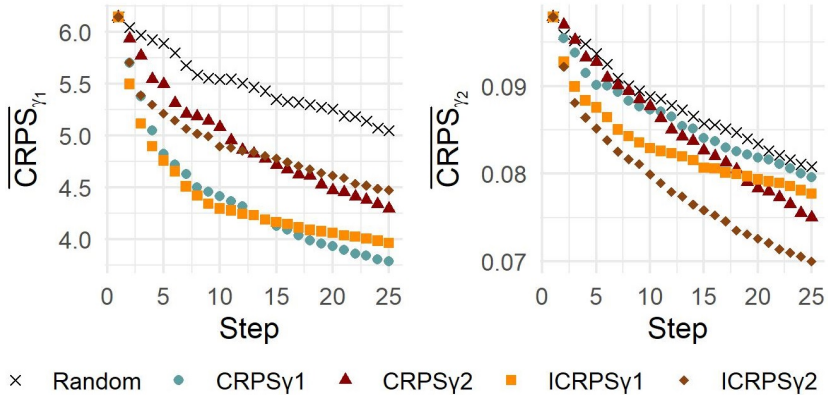


Results Photoswitch

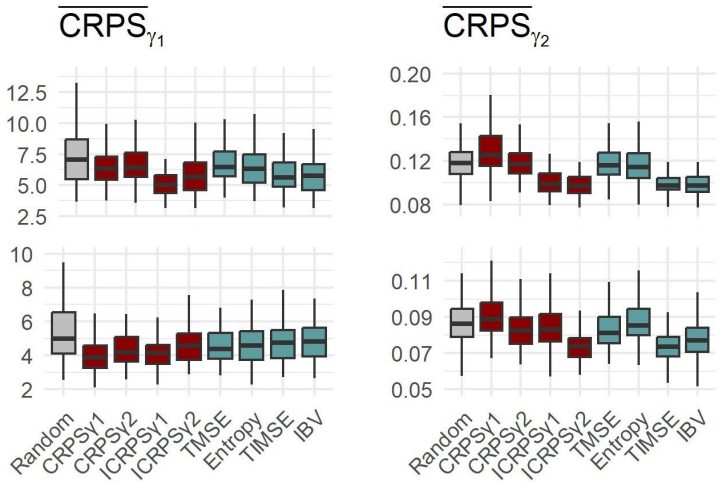
- 100 validation molecules (predictive mean as pink points)
- 30 initial molecules (red crosses), 50 added (blue crosses), 262 candidate molecules (predictive mean $\pm$ 1 SD as turquoise bars)
- Pointwise CRPS_{ν_1}:



Results ($n = 30$ and $\widetilde{n} = 25$)



Comparison ($n = 5$ and $\widetilde{n} = 10, 35$)



Conclusions and outlook

- Introduced new CRPS-based criteria for targeted sequential design
- Promising performance in molecular test case

Conclusions and outlook

- Introduced new CRPS-based criteria for targeted sequential design
- Promising performance in molecular test case
- Applicable to a wide range of fields (e.g., agriculture)

Conclusions and outlook

- Introduced new CRPS-based criteria for targeted sequential design
- Promising performance in molecular test case
- Applicable to a wide range of fields (e.g., agriculture)
- Targeted sequential design based on scoring rules

Conclusions and outlook

- Introduced new CRPS-based criteria for targeted sequential design
- Promising performance in molecular test case
- Applicable to a wide range of fields (e.g., agriculture)
- Targeted sequential design based on scoring rules

Preprint and code (including benchmarks on continuous functions):

- CRPS-Based Targeted Sequential Design with Application in Chemical Space (L.Friedli, A. Gautier, A. Broccard and D. Ginsbourger)
- GitHub/LeaFrie/GP_sequentialCRPS

References I

[1] Julien Bect, Franccois Bachoc, and David Ginsbourger. A supermartingale approach to Gaussian process based sequential design of experiments. Bernoulli, 25(4A):2883–2919, 2016.

[2] Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association, 102(477):359–378, 2007.

[3] Tilmann Gneiting and Roopesh Ranjan. Comparing density forecasts using threshold-and quantile-weighted scoring rules. Journal of Business & Economic Statistics, 29(3):411–422, 2011.

[4] John C. Gower. A general coefficient of similarity and some of its properties. Biometrics, 27(4):857–871, 1971. ISSN 0006341X, 15410420. URL http://www.jstor.org/stable/2528823.

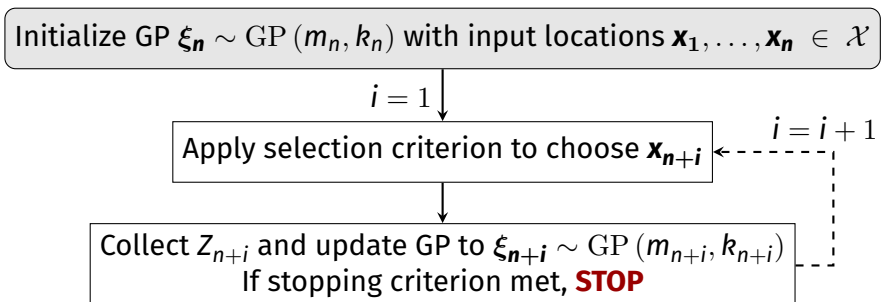
References II

[5] James E Matheson and Robert L Winkler. Scoring rules for continuous probability distributions. Management Science, 22(10):1087–1096, 1976.

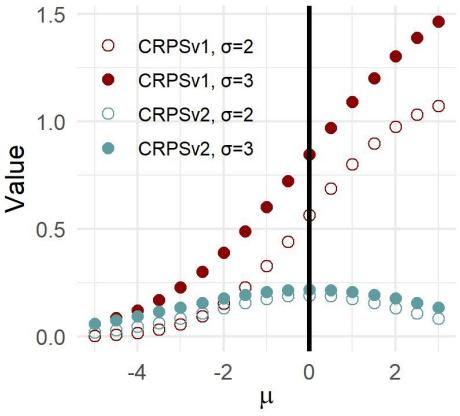
[6] Aditya R. Thawani, Ryan-Rhys Griffiths, Arian Jamasb, Anthony Bourached, Penelope Jones, William McCorkindale, Alexander A. Aldrick, and Alpha A. Lee. The Photoswitch Dataset: A molecular machine learning benchmark for the advancement of synthetic chemistry. ChemRxiv, doi:10.26434/chemrxiv.12609899.v1, 2020. This content is a preprint and has not been peer-reviewed.

Targeted sequential design

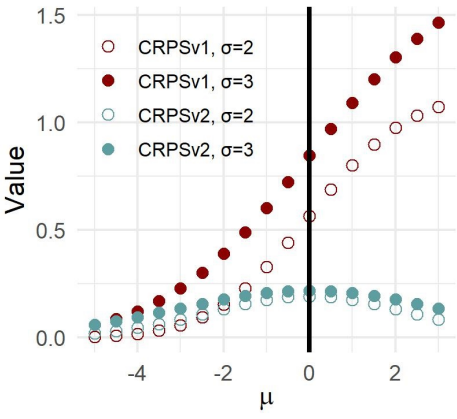
- Targeted scoring of GP predictions **for Sequential Design of Experiments**
- One-step look-ahead (myopic) strategy

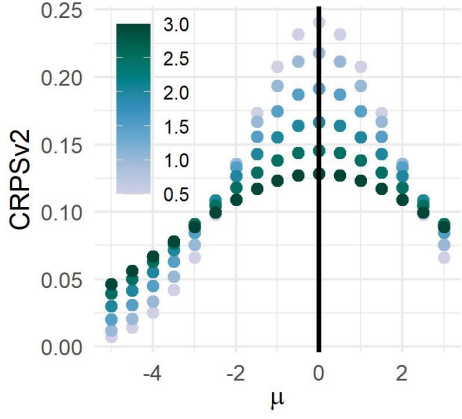


Expected CRPS_{ν_1} and CRPS_{ν_2} for Φ_{μ,σ^2} and $t = 0$



Expected CRPS_{ν_1} and CRPS_{ν_2} for Φ_{μ,σ^2} and $t = 0$





Results Photoswitch

Pointwise CRPS_{ν_2}: