# Recommendations for Baselines and Benchmarking Approximate Gaussian Processes

***Sebastian W. Ober, David R. Burt, Artem Artemev, Mark van der Wilk***

**Motivation:** approximate GP methods are often compared in a wide range of settings, making it difficult to determine what works best.

**Our goal**: provide a training procedure that ensures that sparse GPs are a strong baseline and an experimental procedure for comparing Gaussian process approximations.

## Contributions

- **Concrete recommendations** for how approximate GP methods should be assessed
- A **recommended training procedure** for the SGPR method of Titsias (2009)
- We characterise and illustrate **two approximation regimes** for SGPR
- An **experimental procedure** for comparing approximate GP methods

## Recommendations

→ **Approximation quality.** Assess the quality of the *approximation* to the exact GP, for both hyperparameter selections and posterior quality.

→ **Recommended procedure.** For a new method, a recommended training procedure should be given, and assessed for 1) various compute budgets, 2) how much compute is needed for desired performance

→ **Dataset suitability.** Compare to mean prediction and linear model to avoid trivial solutions and that the dataset is not too simple

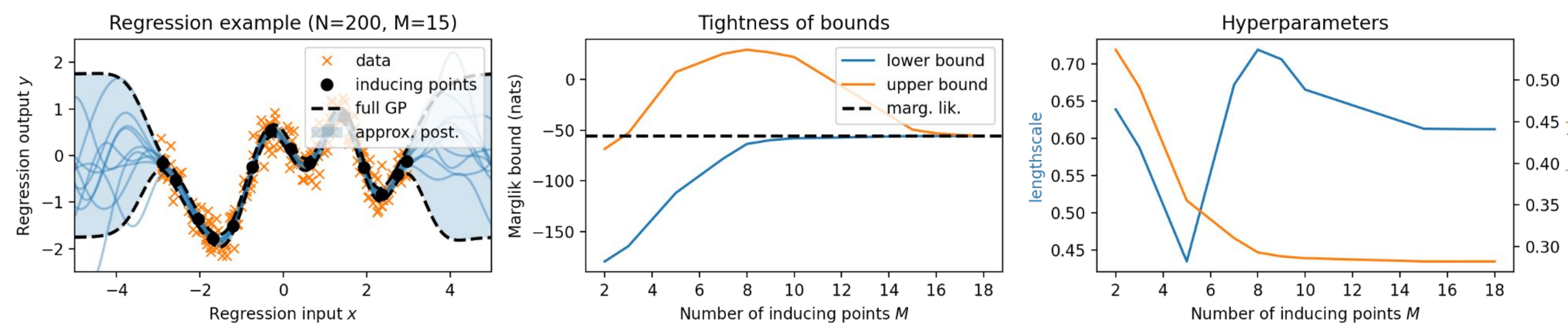
→ **Near-exact regime.** Compare the compute required to achieve near-exactness

→ **Non-exact regime.** Compare how many datasets methods can achieve near-exactness on

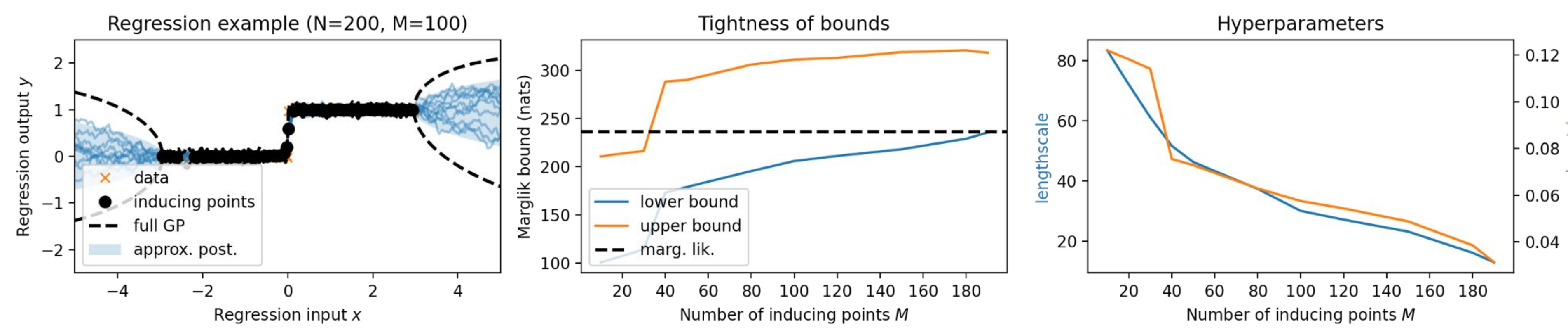
→ **Timed performance.** Run each method for an extended amount of time, comparing at multiple time points

## Near-Exact vs. Non-Sparse Regimes for SGPR

→ In the **near-exact regime**, an approximation can get close to the exact GP:
- Compare the compute required
- For SGPR, few inducing points are needed



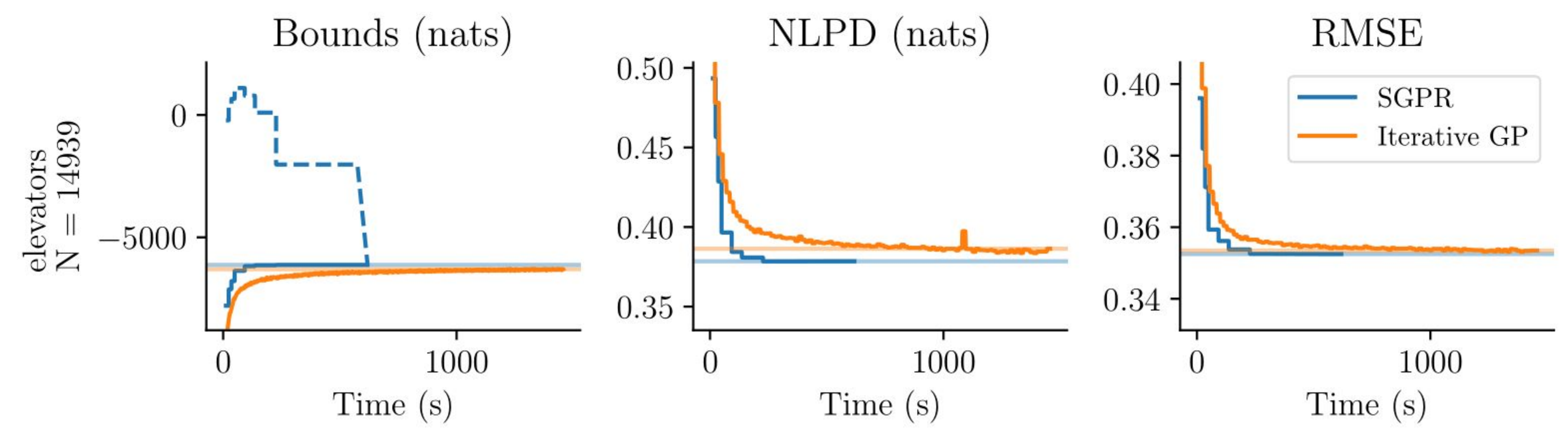

→ For **non-sparse datasets:**
- ~N inducing points are needed for exact performance
- One cause is *model misspecification*




## Timed Performance Evaluation

**Training procedure:**

→ Iterate:
- Choose inducing points using procedure from Foster et al. (2009)/Burt et al. (2020)
- Learn hyperparameters with *fixed* inducing points




**Results:**

→ For elevators, both methods perform near-exactly, with SGPR being faster

→ kin40k is non-sparse, so that the Iterative GP performs better