# Fleet-Wide Policy Iteration using Gaussian Processes

Timothy Verstraeten

VUB – AI Lab & AVRG

In collaboration with:
Pieter J.K. Libin & Jan Helsen & Ann Nowe

VRIJE UNIVERSITEIT BRUSSEL

# Outline

- Fleet Setting
- Reinforcement Learning (RL)
- Gaussian process (GP)
- Fleet GPRL
- Demonstration on mountain car
- Applied on wind farm control case

# Fleet Setting

- **Similar devices, same control task**
- Large scale production / operation




# Fleet Setting

- **Similar devices, same control task**
- Large scale production / operation




# Fleet Setting

- **Similar devices, same control task**
- Large scale production / operation
- High maintenance costs!




# Fleet Setting

- **Similar devices, same control task**
- Large scale production / operation
- High maintenance costs!
- Failure prevention through control




# Fleet Control

- Failures are costly events, how do we learn from them?
  - Sample efficiency is key!
- Group of devices/machines with **same objective & same design**
  - Potential to share knowledge about control task
- Small discrepancies
  - e.g., degradation, production errors

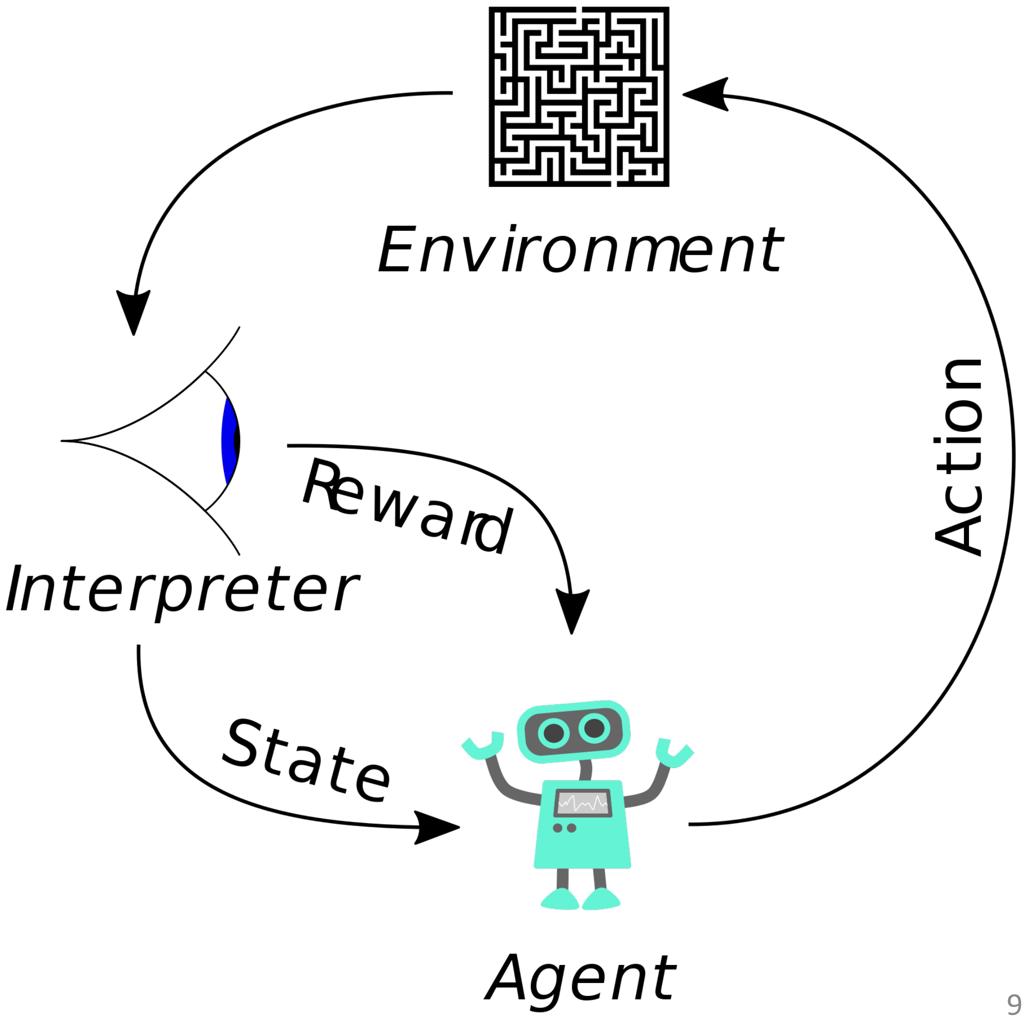
# Sharing Control

≈ wind turbines → ≠ control strategies

Challenge: *How do we share information?*

# Markov Decision Process

- MDP $(S, A, T, \gamma, R)$
  - S, A are possible state and actions
  - T is a transition function
  - R is a reward function
  - $\gamma$ is a discount factor



# Fleet Markov Decision Process

- Fleet MDP $(S, A, \mathbb{T}, \gamma, R)$
  - $\mathbb{T}$ is a set of $M$ transition models $T_m(s, a)$

- How can we detect similarities and transfer knowledge between models?

- Joint Bayesian regression model (i.e., Gaussian process)
  → correlations between fleet members

# Gaussian Process

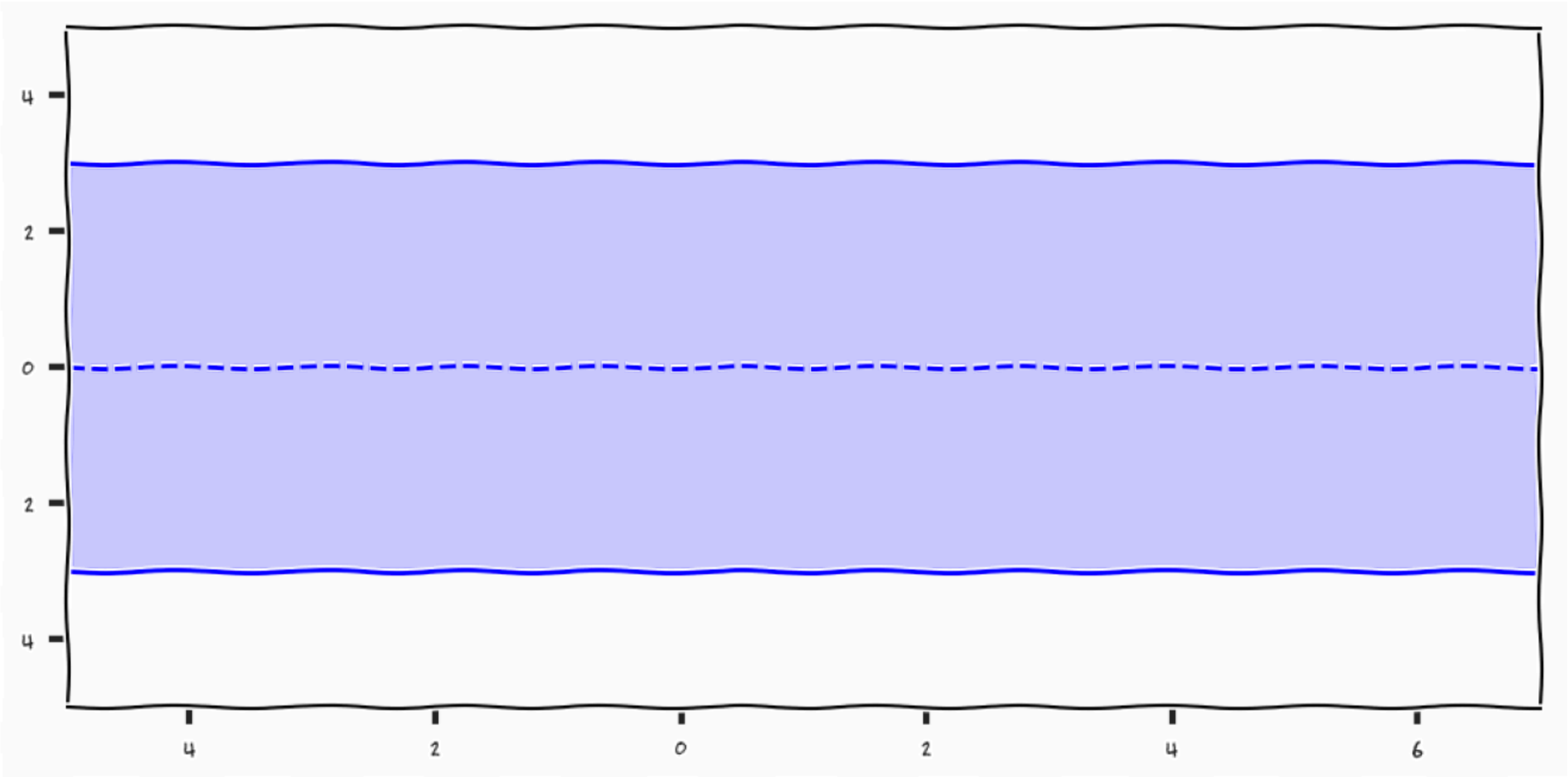





Figure made by Carl Henrik Ek (Gaussian Process Summer School 2018)

# Gaussian Process

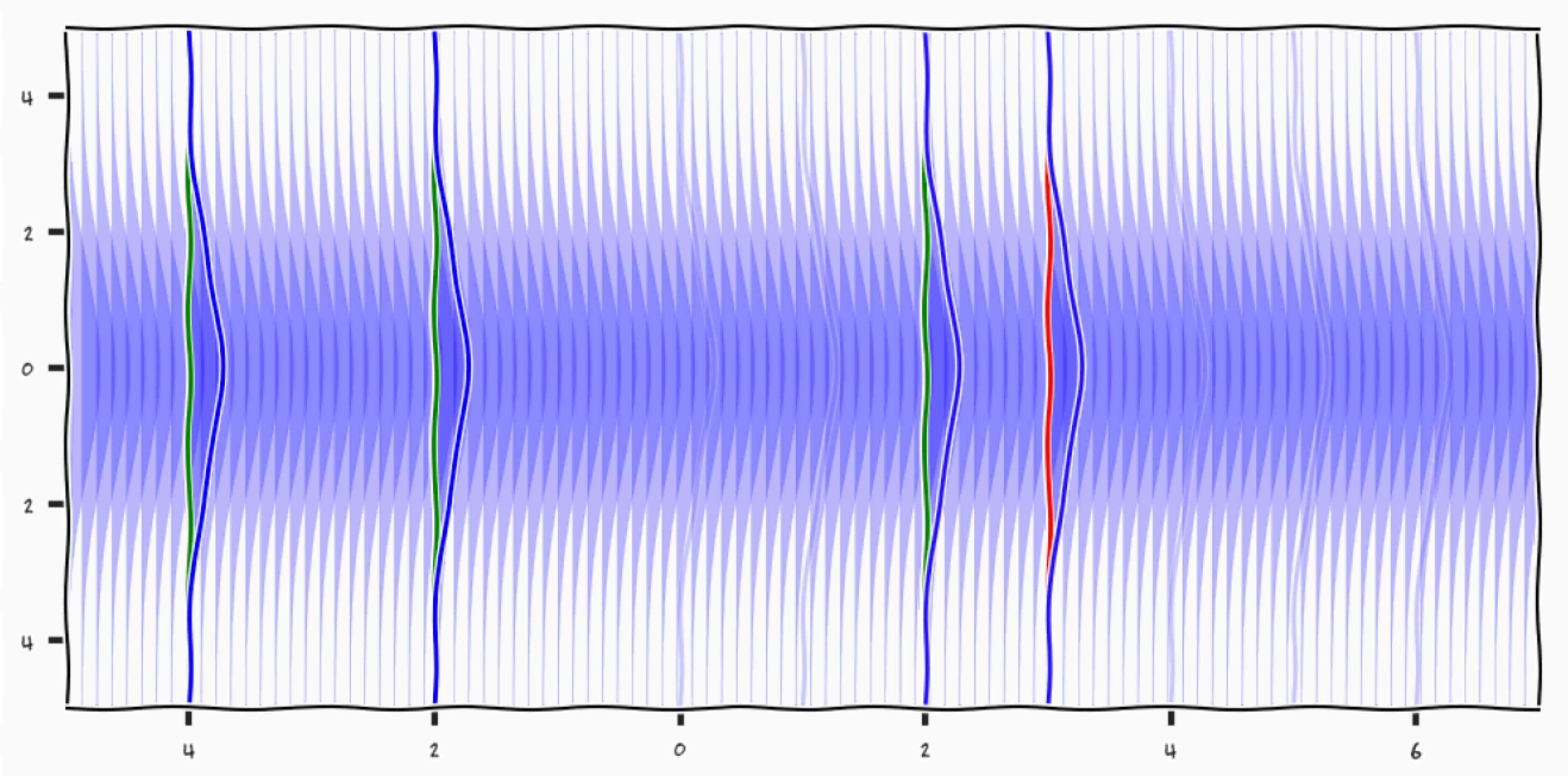





Figure made by Carl Henrik Ek (Gaussian Process Summer School 2018)

# Gaussian Process

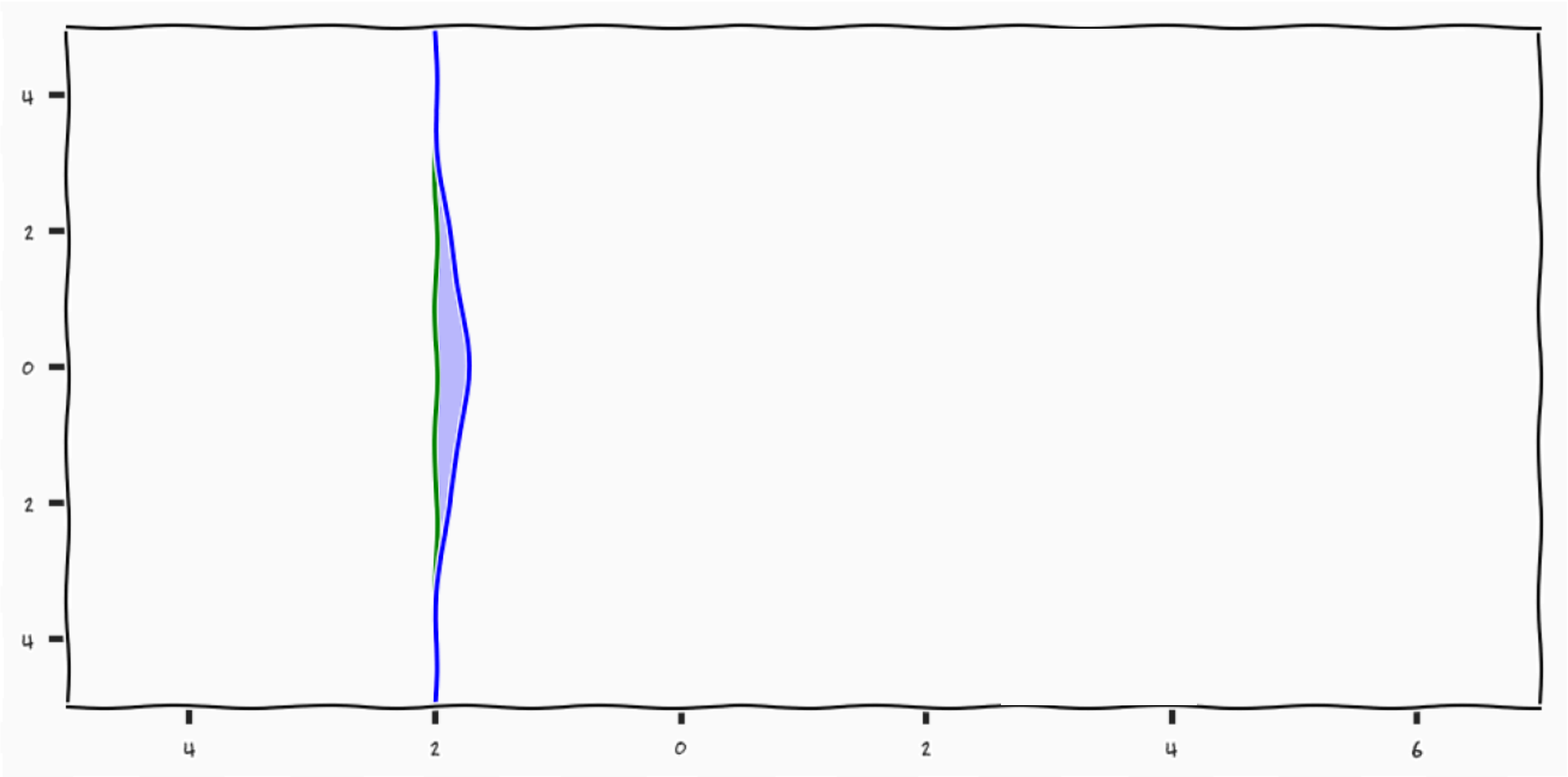





Figure made by Carl Henrik Ek (Gaussian Process Summer School 2018)

# Gaussian Process

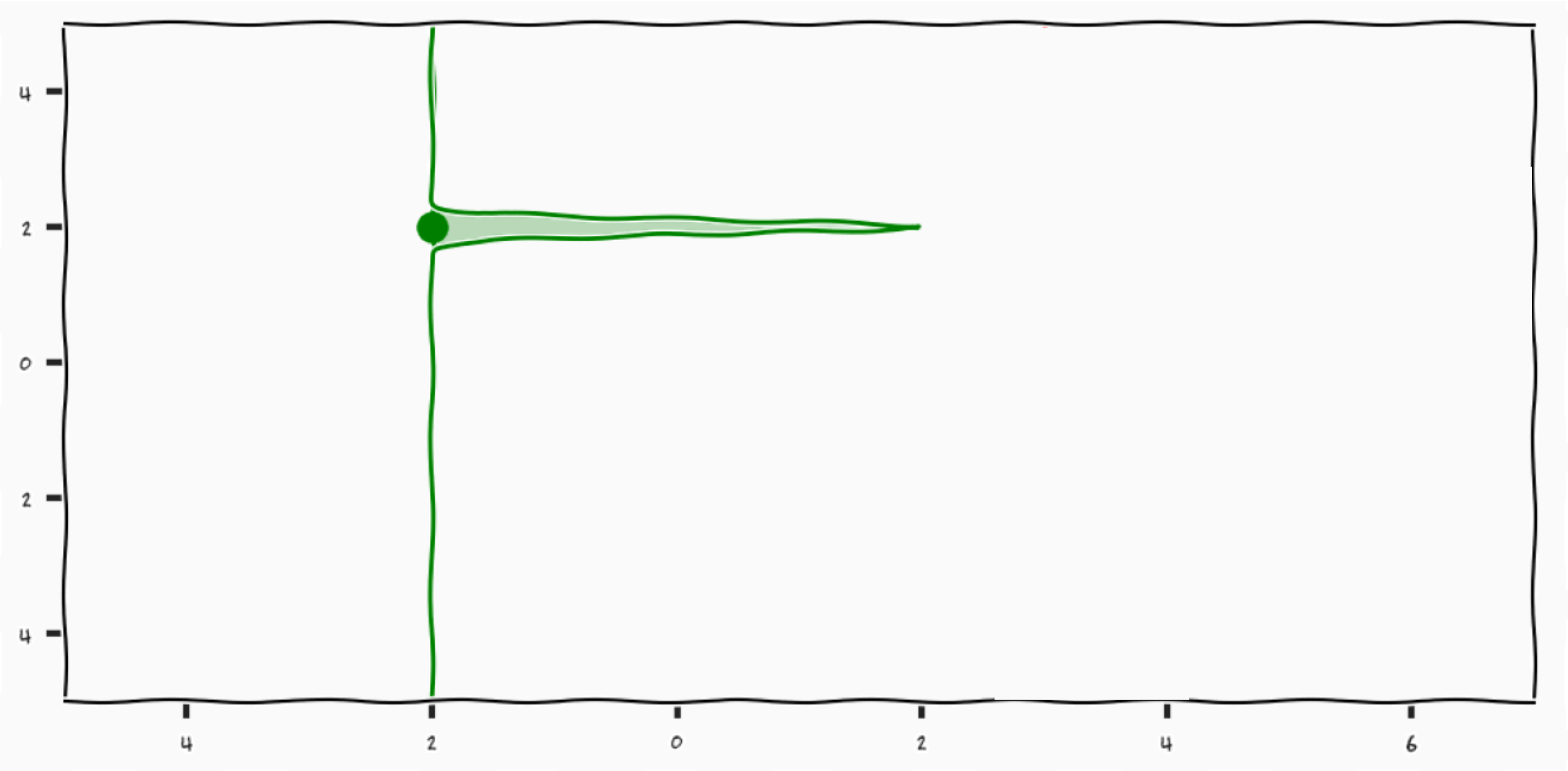





Figure made by Carl Henrik Ek (Gaussian Process Summer School 2018)

# Gaussian Process

$$\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, K\right) \Rightarrow \mathcal{N}\left(\begin{bmatrix}2\\?\end{bmatrix}, \hat{K}\right)$$

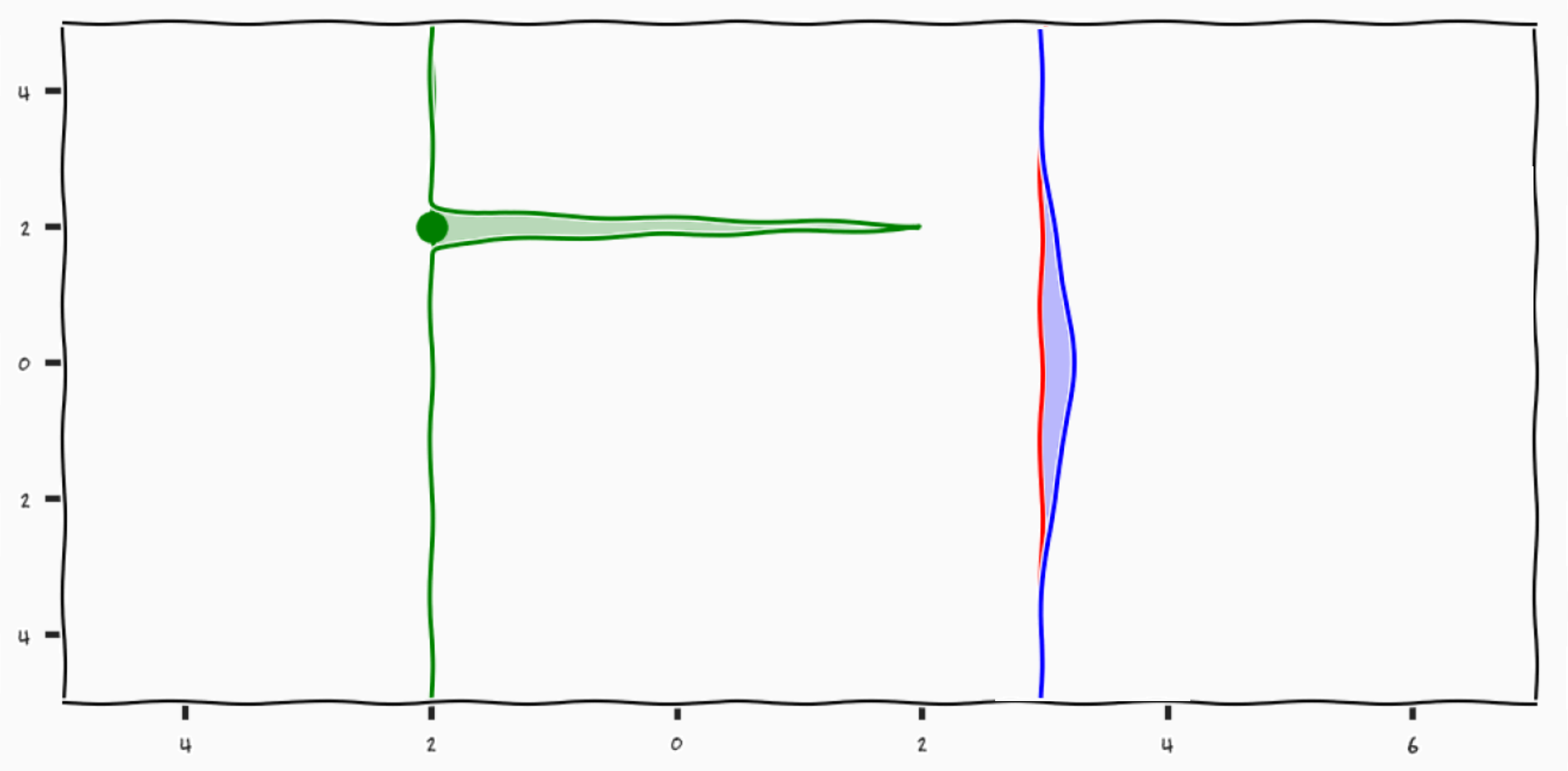





Figure made by Carl Henrik Ek (Gaussian Process Summer School 2018)

# Gaussian Process

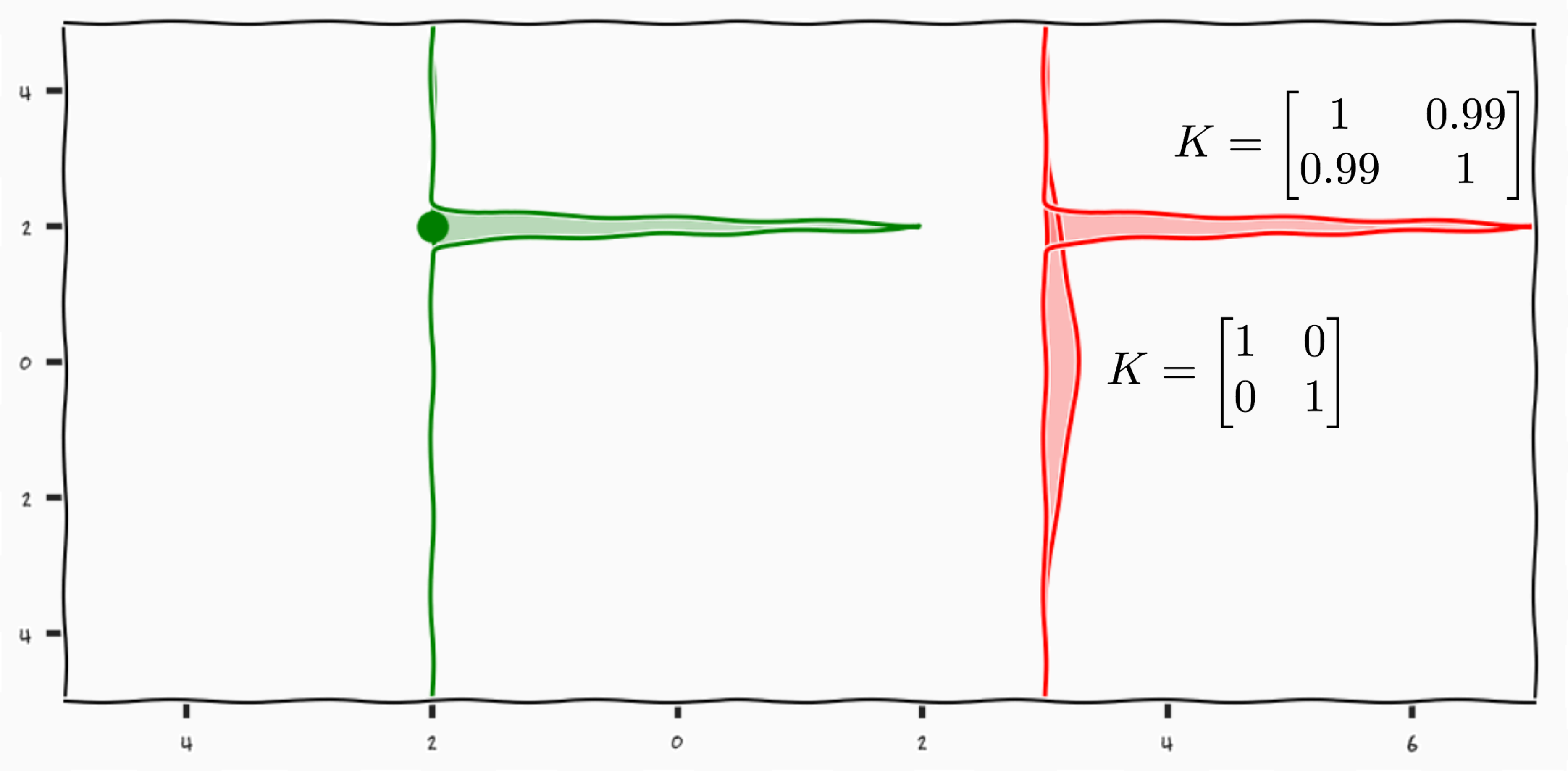





Figure made by Carl Henrik Ek (Gaussian Process Summer School 2018)

# Gaussian Process

$$\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & k(x,x')\\ k(x,x') & 1\end{bmatrix}\right)$$

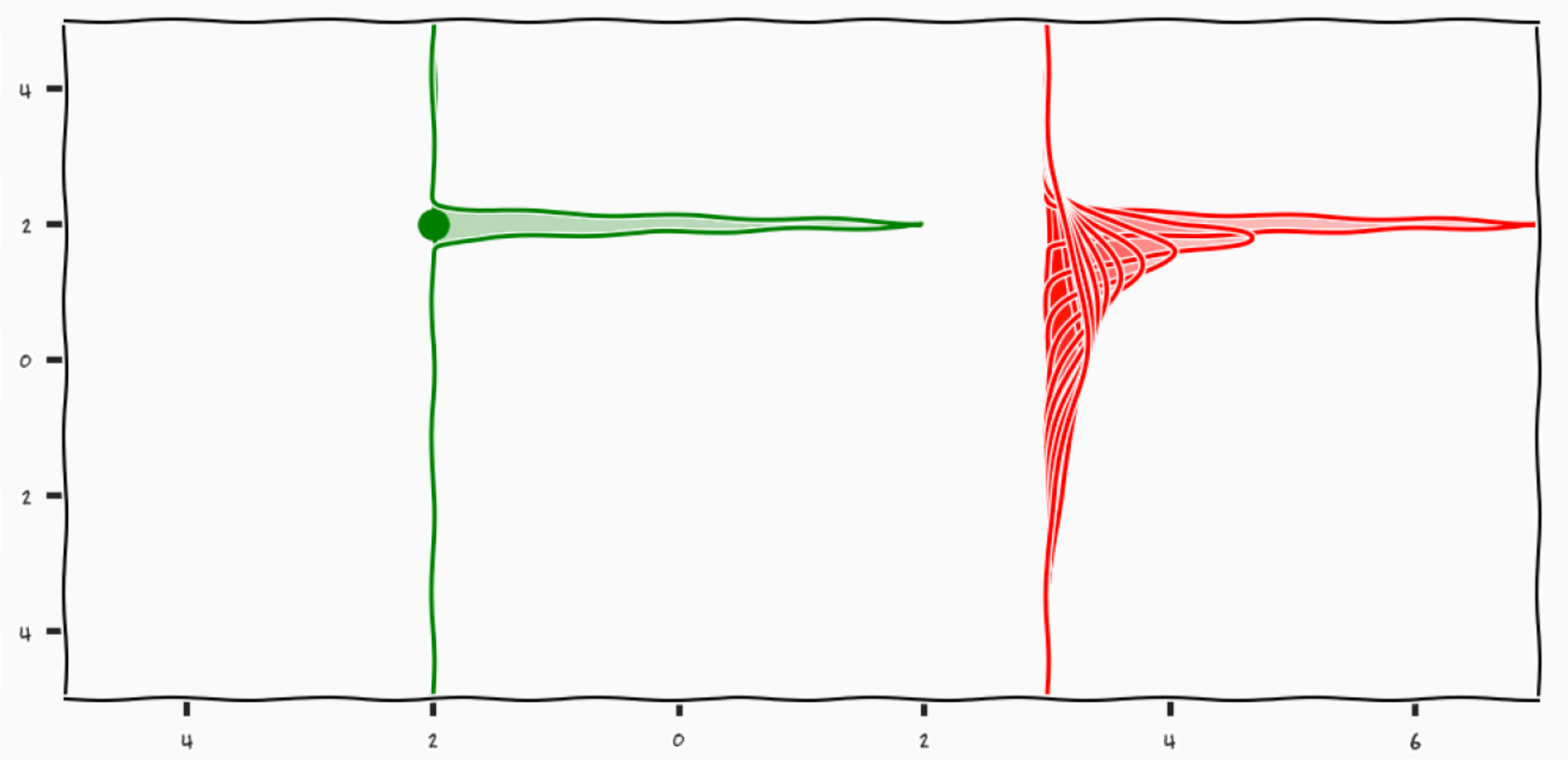





Figure made by Carl Henrik Ek (Gaussian Process Summer School 2018)

# Covariance kernels

Squared-exponential

$$k(x, x') = \exp(-\|x - x'\|^2 / l^2)$$

Linear

$$k(x, x') = x^T x'$$

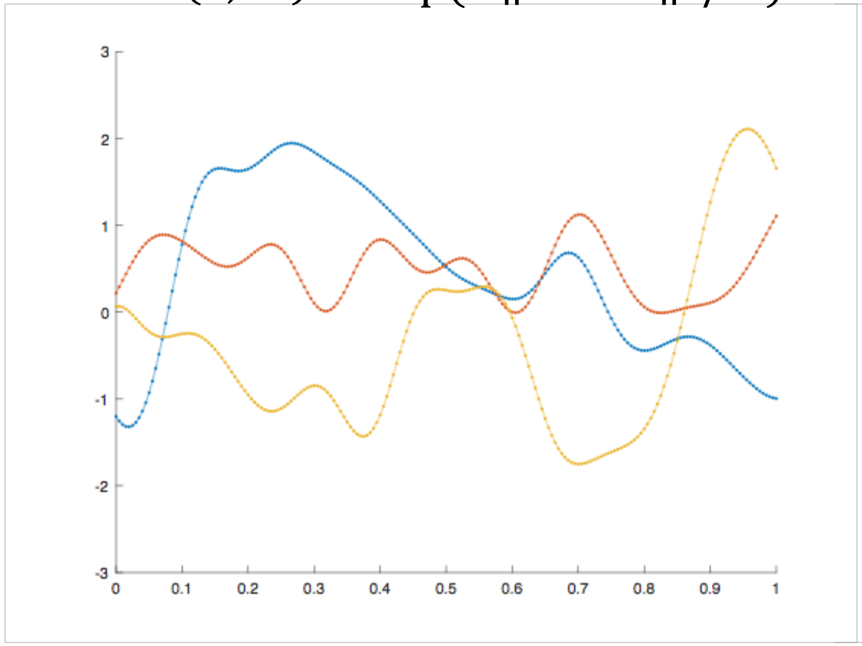
Brownian

$$k(x, x') = \min(x, x')$$

# Sample functions from GPs

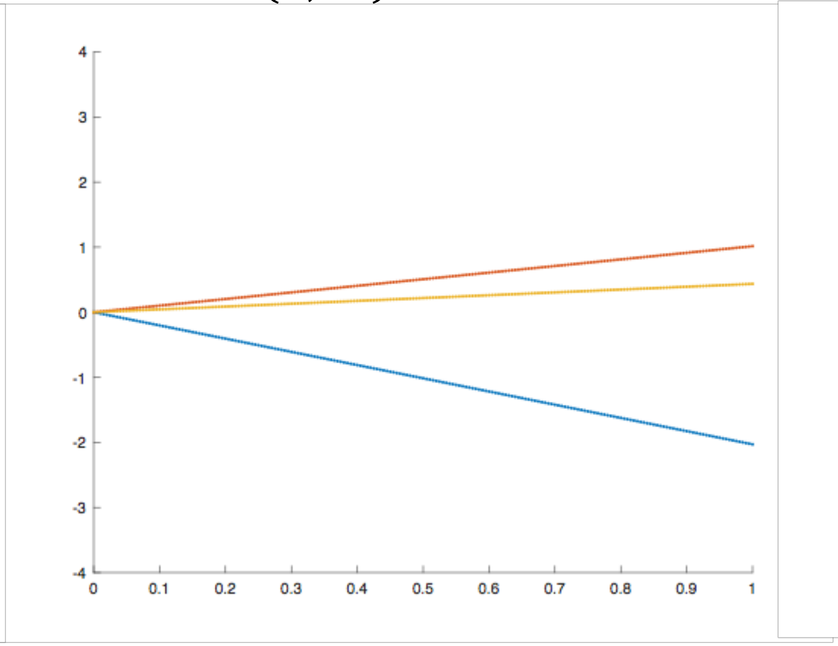
Squared-exponential

$$k(x, x') = \exp(-\|x - x'\|^2 / l^2)$$



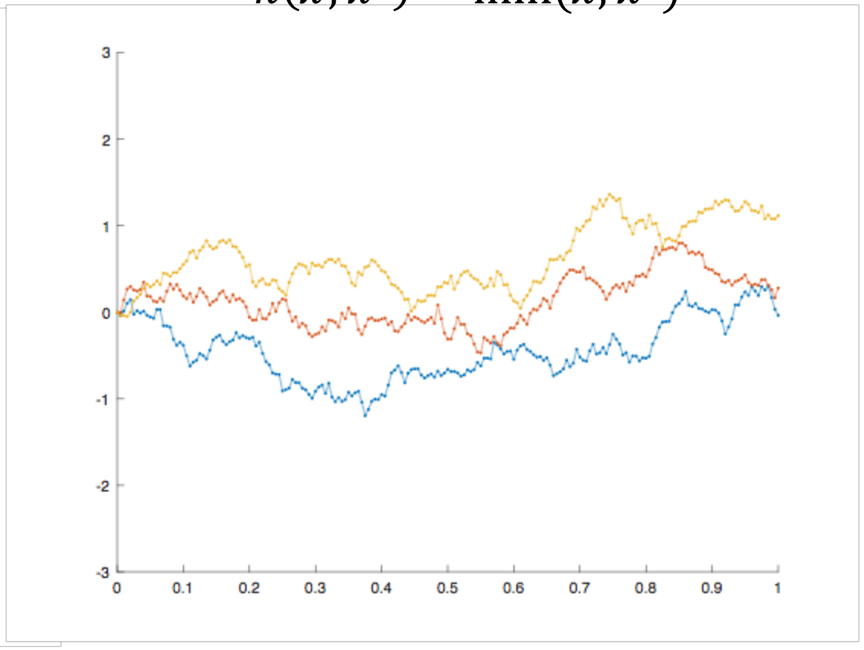

Linear

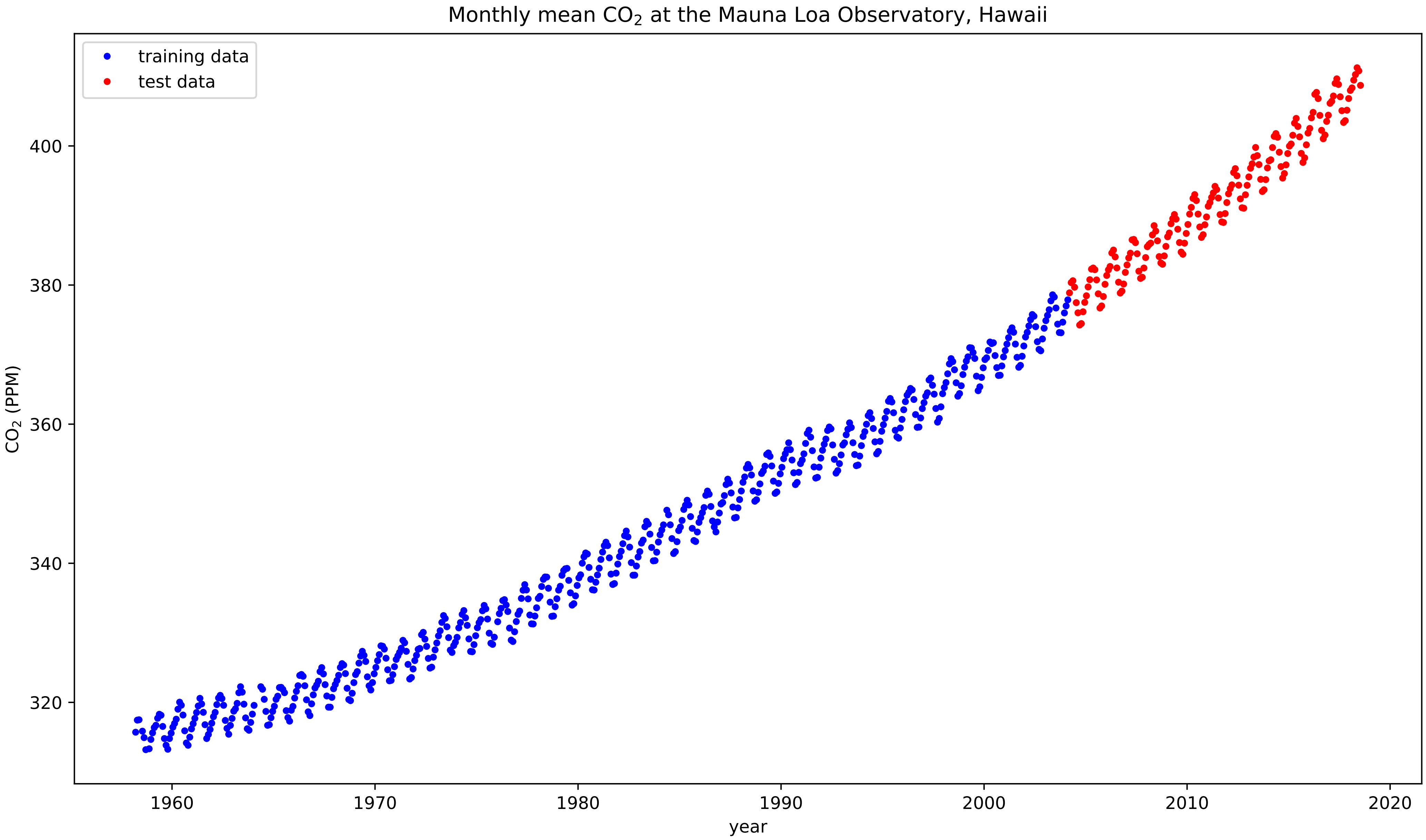
$$k(x, x') = x^T x'$$




Brownian

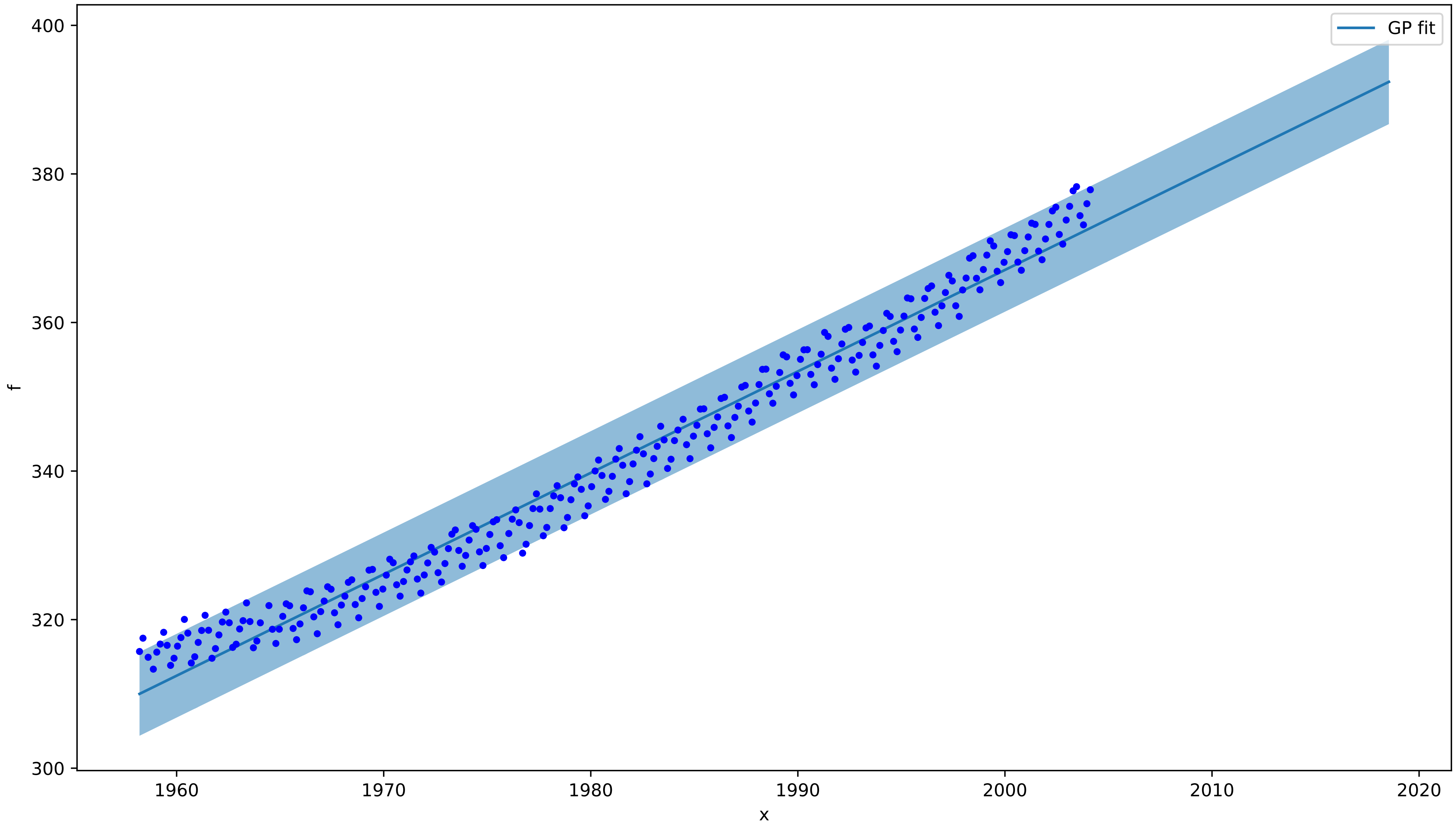
$$k(x, x') = \min(x, x')$$




Monthly mean $CO_2$ at the Mauna Loa Observatory, Hawaii
training data
test data
400
380
360
340
320
$CO_2$ (PPM)
1960
1970
1980
1990
2000
2010
2020
year

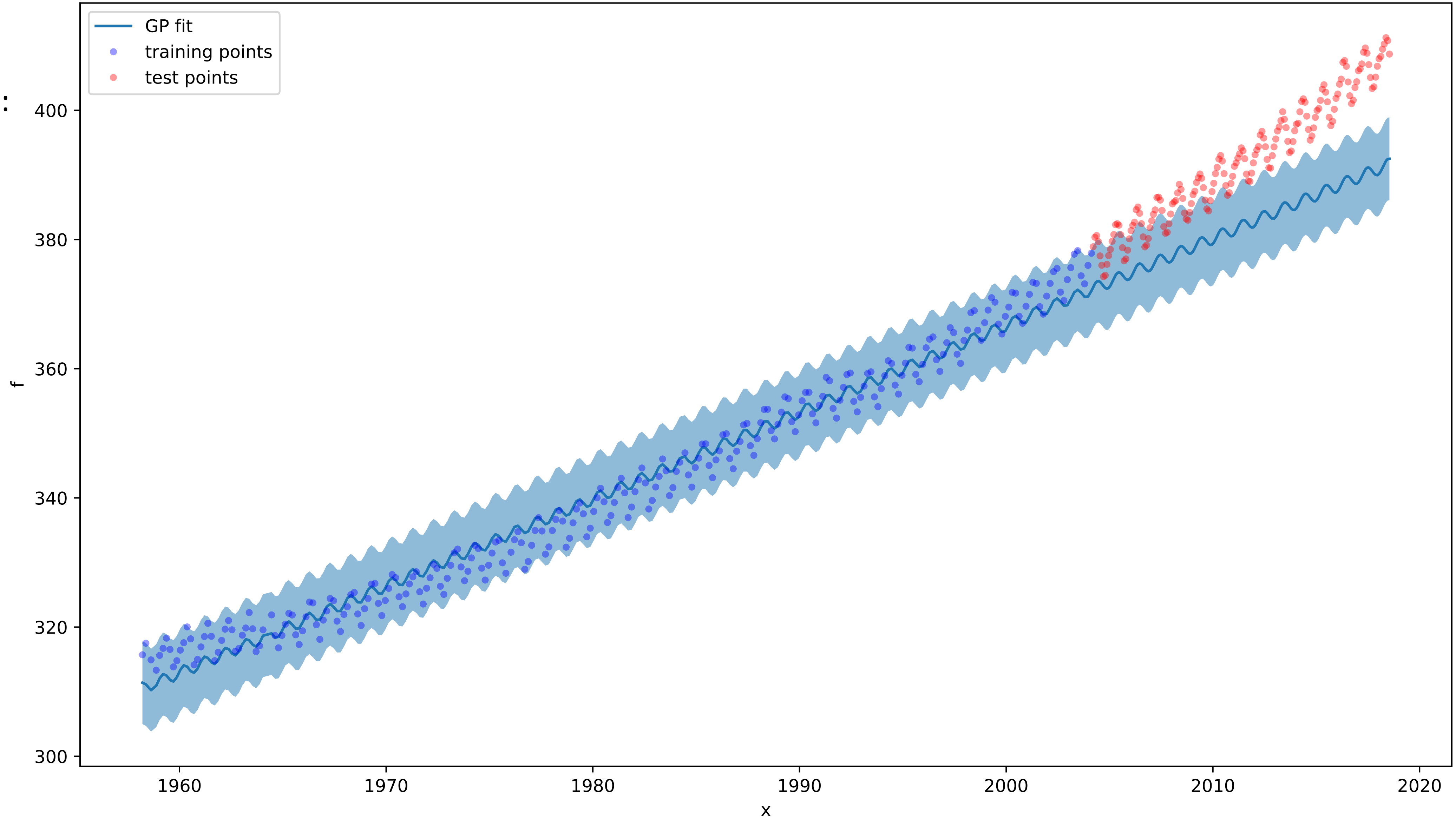
Kernel combination:

- Linear

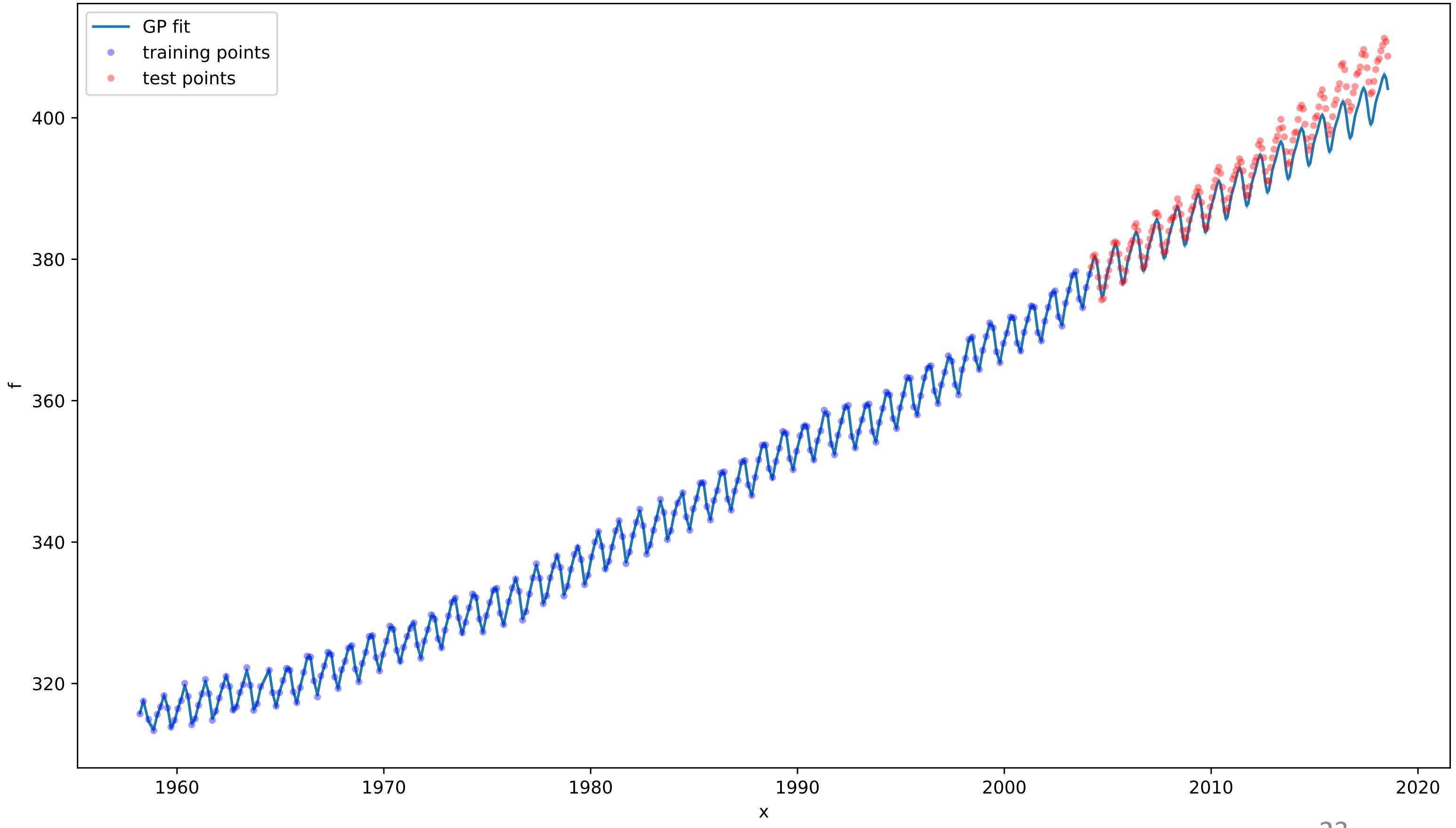


Kernel combination:

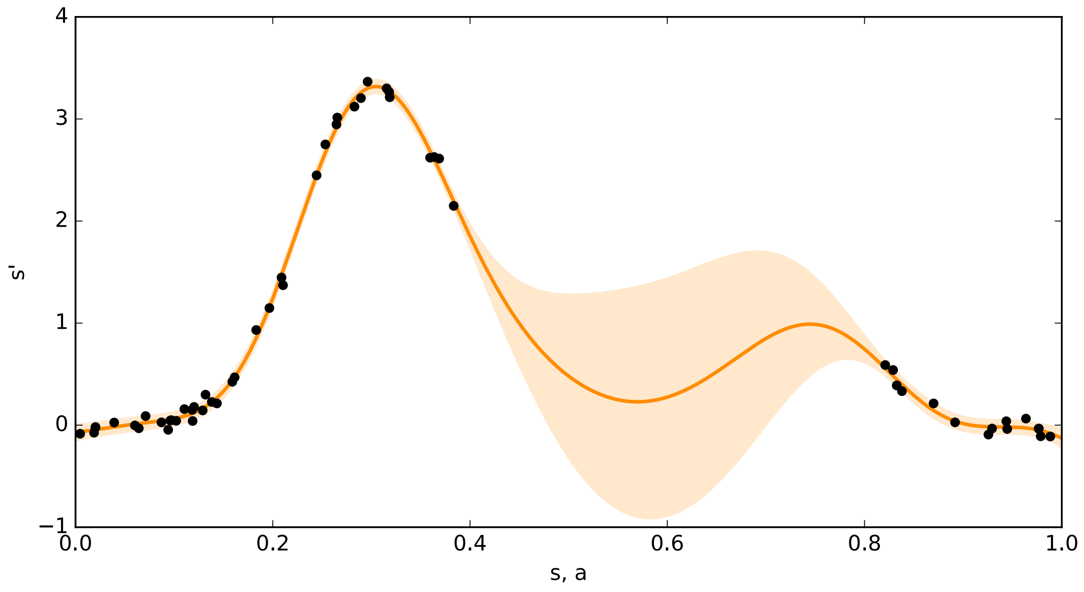
- Linear
- Periodic



| Property | Kernel |
| --- | --- |
| Linear trend | Linear |
| Constant offset | Bias |
| Periodicity (short term) | Periodic |
| Amplitude modulator (long term) | Squared-exponential |
| Non-linearity in overall trend | Exponential |



# Gaussian Process




- Bayesian model:
  - Parameters are considered to be random

- Used for regression:
  1) Describe prior beliefs
  2) Observe data
  3) Update belief (i.e., posterior)

- Generalization through pair-wise correlations:
  - *If $x$ and $x'$ are similar, their outputs $y$ and $y'$ are correlated*
  - E.g., distance-based

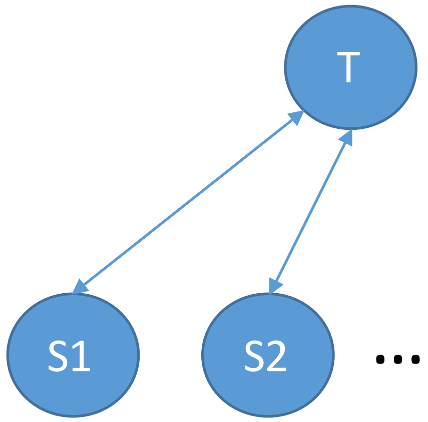
- $f(x) \sim \mathcal{GP}\big(0, k(x, x')\big)$

# Transfer over transition models

1) Adopt "multiple sources – single target" transfer framework
    - Choose target in fleet, the rest are sources

    - Sources have independent components
        - $T_s(x) = w_{s,s} G_s(x) + \alpha_s L_s(x)$

    - Target has an independent component, but is also **dependent on the sources**
        - $T_t(x) = \sum_s w_{t,s} G_s(x) + \alpha_t L_t(x)$




2) The components are sampled from a a zero-mean GP
    - A linear combination of components is also a GP!

# Transfer over transition models

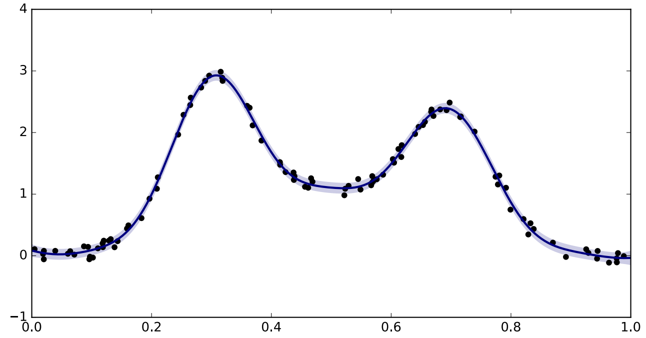
1) $\mathrm{cov}\big(T_t(x),\ T_s(x')\big)$
$= w_{t,s}\ \ w_{s,s}\ \mathrm{cov}\big(G_s(x), G_s(x')\big)$ [independence]
$= w_{t,s}\ \ w_{s,s}\ \mathrm{k}(x, x')$ [covariance kernel]

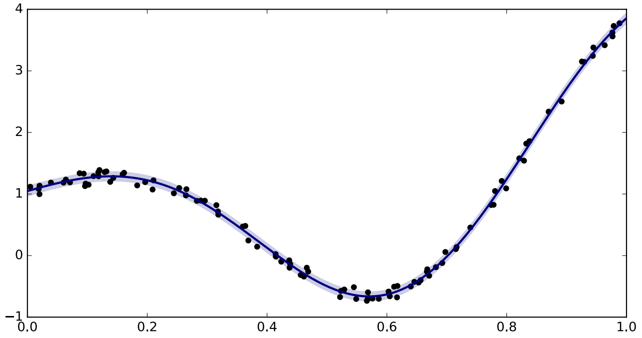
2) New fleet-wide kernel:

- $k_F([x, m], [x', m']) = \ G_{m,m'}\ \mathrm{k}(x, x')$, where G contains the weights

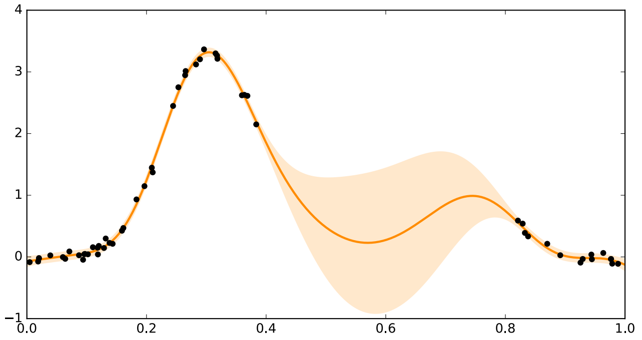
Key insight: Correlate fleet members' transition models

**Sources**



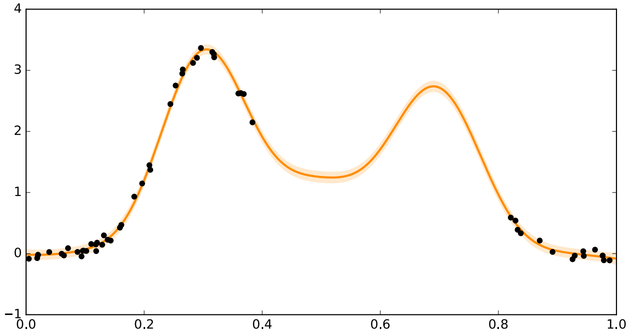

**Target**




| *Correlations* | Member 1 | Member 2 | Member 3 |
|---|---|---|---|
| **Member 1** | 1 | **0.97** | **0.08** |
| **Member 2** | 0.97 | 1 | **0** |
| **Member 3** | 0.08 | **0** | 1 |

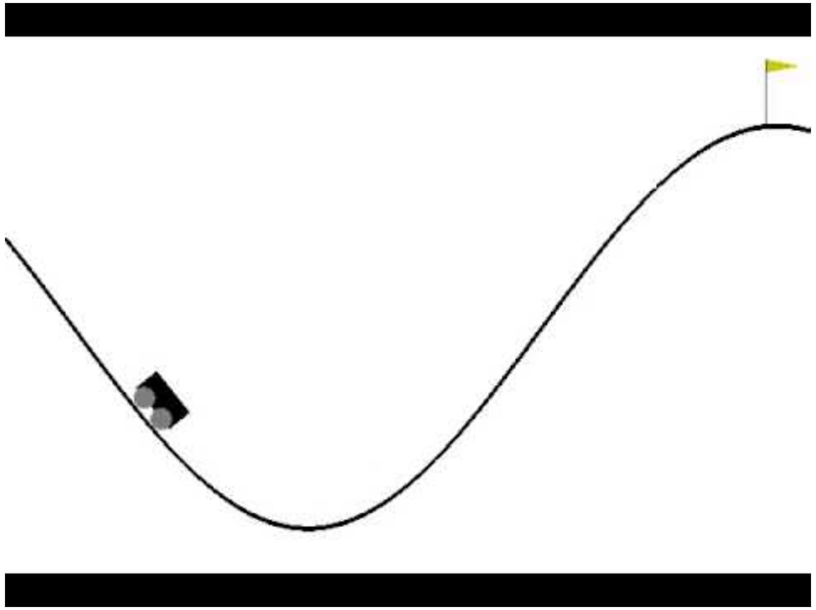
**GP transition model**

*Optimize correlations to maximize model accuracy [maximum likelihood]*



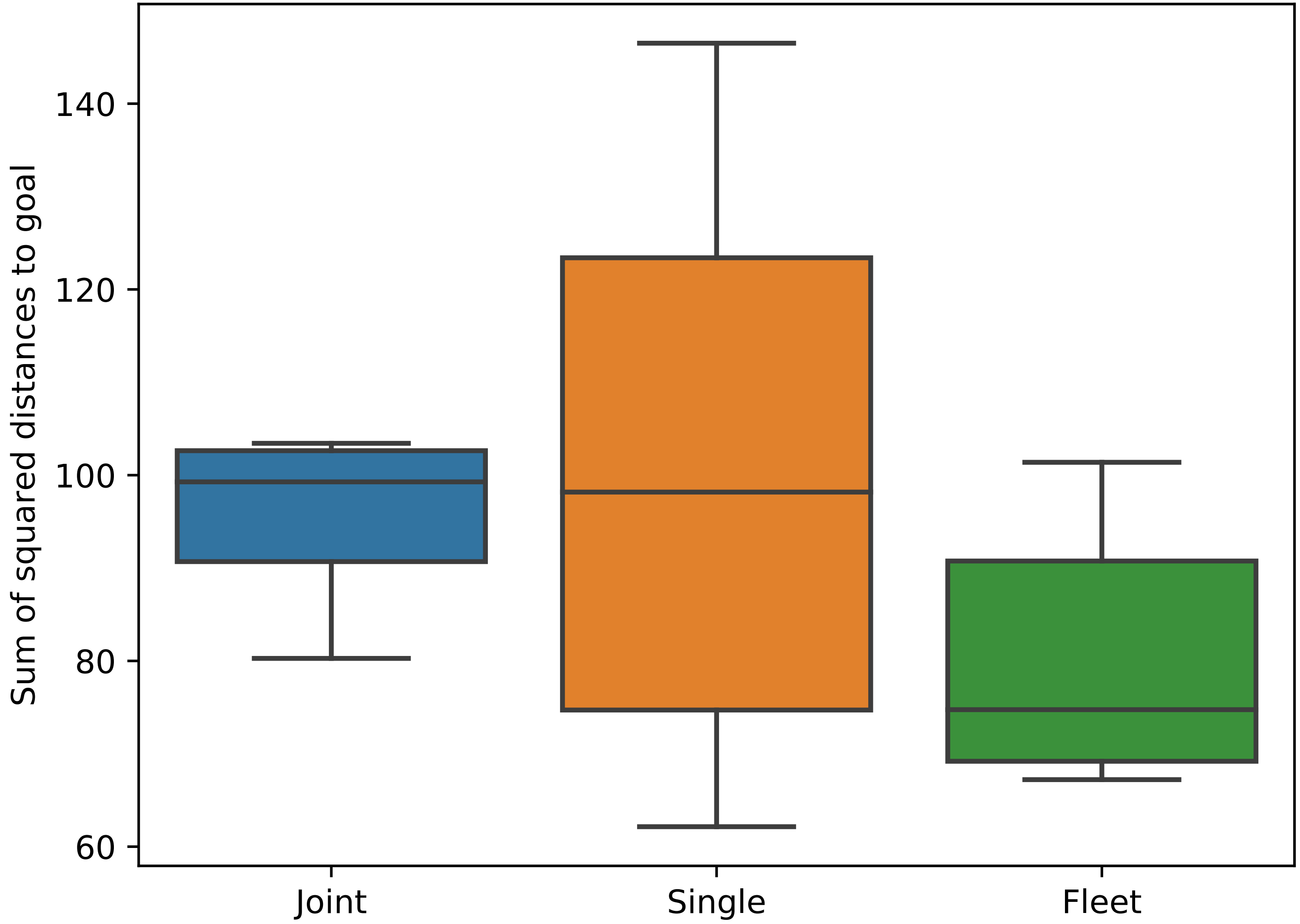

**Policy Iteration (GPRL)**

# Continuous Mountain Car

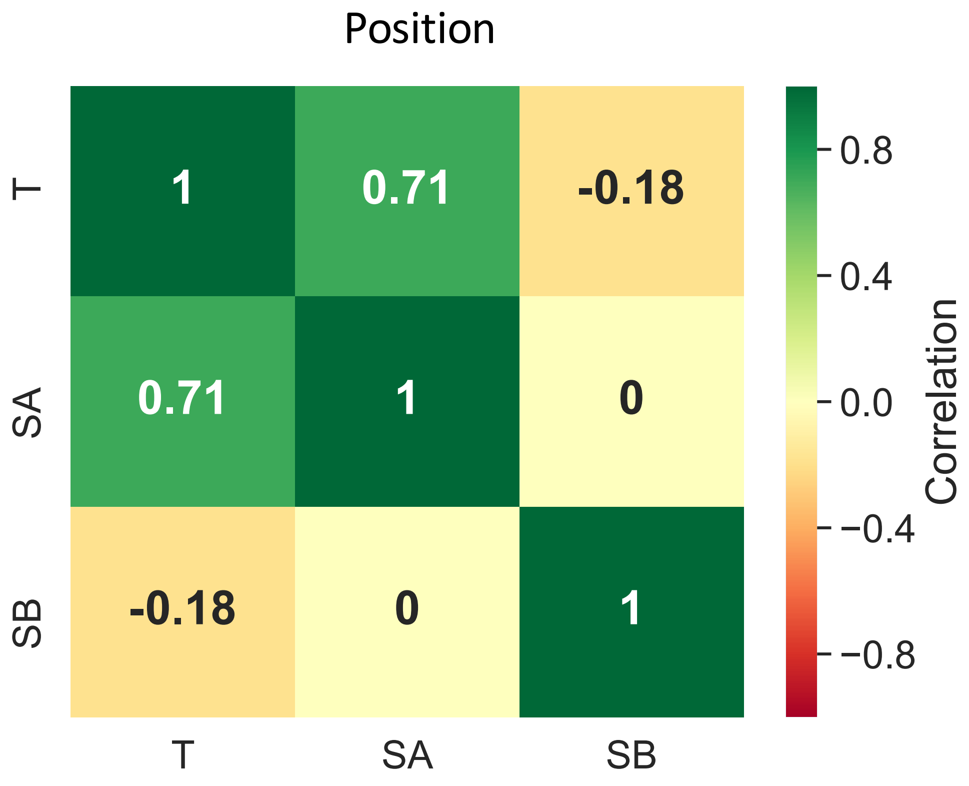


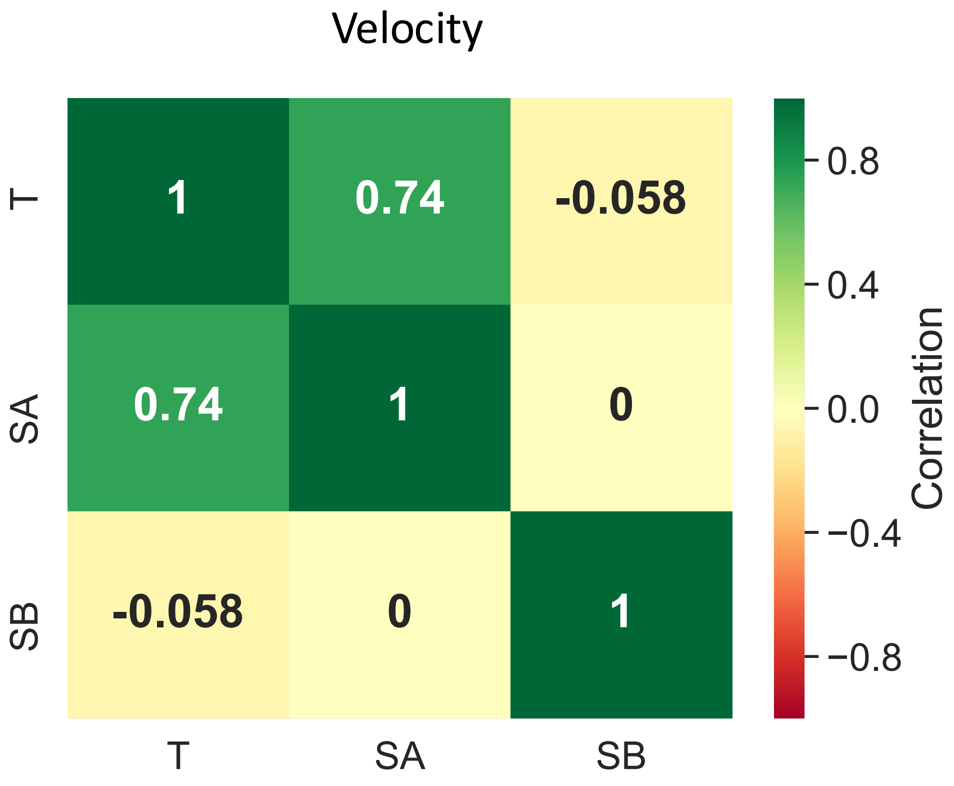
- 1 target with random batch of 20 transitions:
  - Mass 1.0 kg
- 2 sources with random batch of 100 transitions:
  - Source A has mass 1.1 kg
  - Source B has mass 5.0 kg
- Peaked Gaussianly shaped reward is given at the goal (at the flag with a velocity of 0).
- Objective: Reach the goal using sources' and own transition models

# Performance

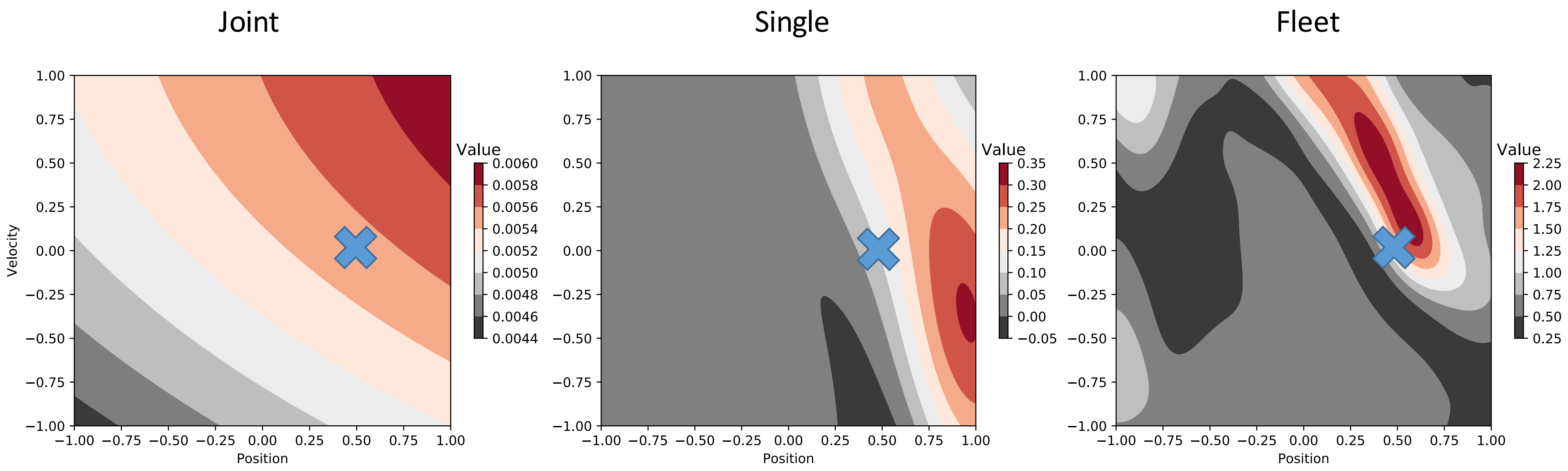


# Learned correlations



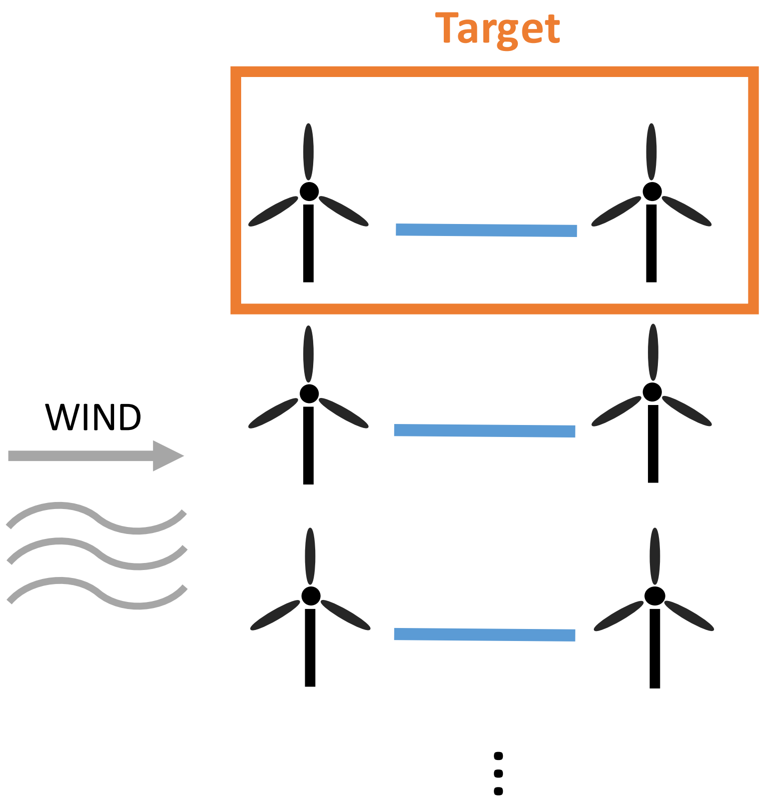



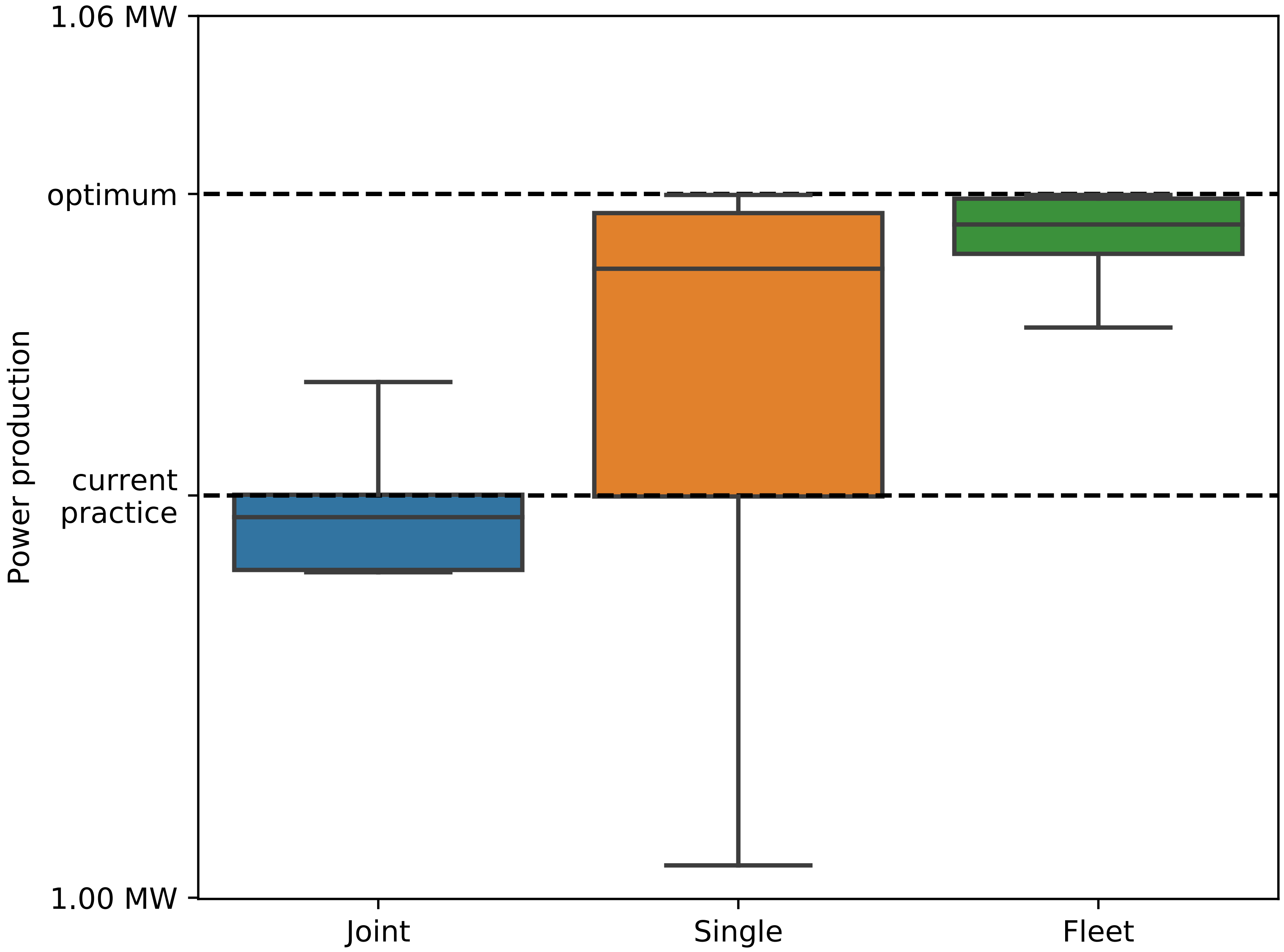
# Value Functions



# Wind Farm Control

- Fleet of 8 agents, 50 samples each
- Members are turbine rows: 1 upstream and 1 downstream turbine
- Non-linear dynamics due to wake
- Vary generator efficiencies due to degradation
- Learn to orient themselves to maximize power production




1.06 MW
optimum
current practice
1.00 MW
Power production
Joint
Single
Fleet

# Conclusion

- Sparse transfer learning model for fleet control
- GPs → sound, efficient framework to deal with correlations between fleet members
- Allows for inclusion of domain knowledge
- URL: https://arxiv.org/abs/1911.10121