FUNCTION FACTORIZATION USING WARPED GAUSSIAN PROCESSES

UNIVERSITY OF
CAMBRIDGE

MIKKEL N. SCHMIDT

Computational and Biological Learning

1 ABSTRACT

Function factorization: A new approach to non-linear regression.

- Approximate complicated function on high-dimensional space by product-sum of simpler functions on low-dimensional subspaces.
- Appropriate when data has factorial structure.
- Generalizes matrix and tensor factorization.
- Non-parametric Bayesian approach: Priors over the simple functions are warped Gaussian processes [4].
- Inference using Hamiltonian Markov chain Monte Carlo [2].
- Superior predictive performance compared to Gaussian process regression and tensor factorization.

2 MODEL

Problem: Non-linear regression $y(\boldsymbol{x}) : \mathcal{X} \to \mathbb{R}$
Model: Function is modelled by a product sum of simpler functions defined on subspaces of the input space

Components

Factors

$$y(\boldsymbol{x}) = \sum_{k=1}^{K} \prod_{i=1}^{I} f^{(i,k)}(\boldsymbol{x}_i)$$

Output
$y \in \mathbb{R}$

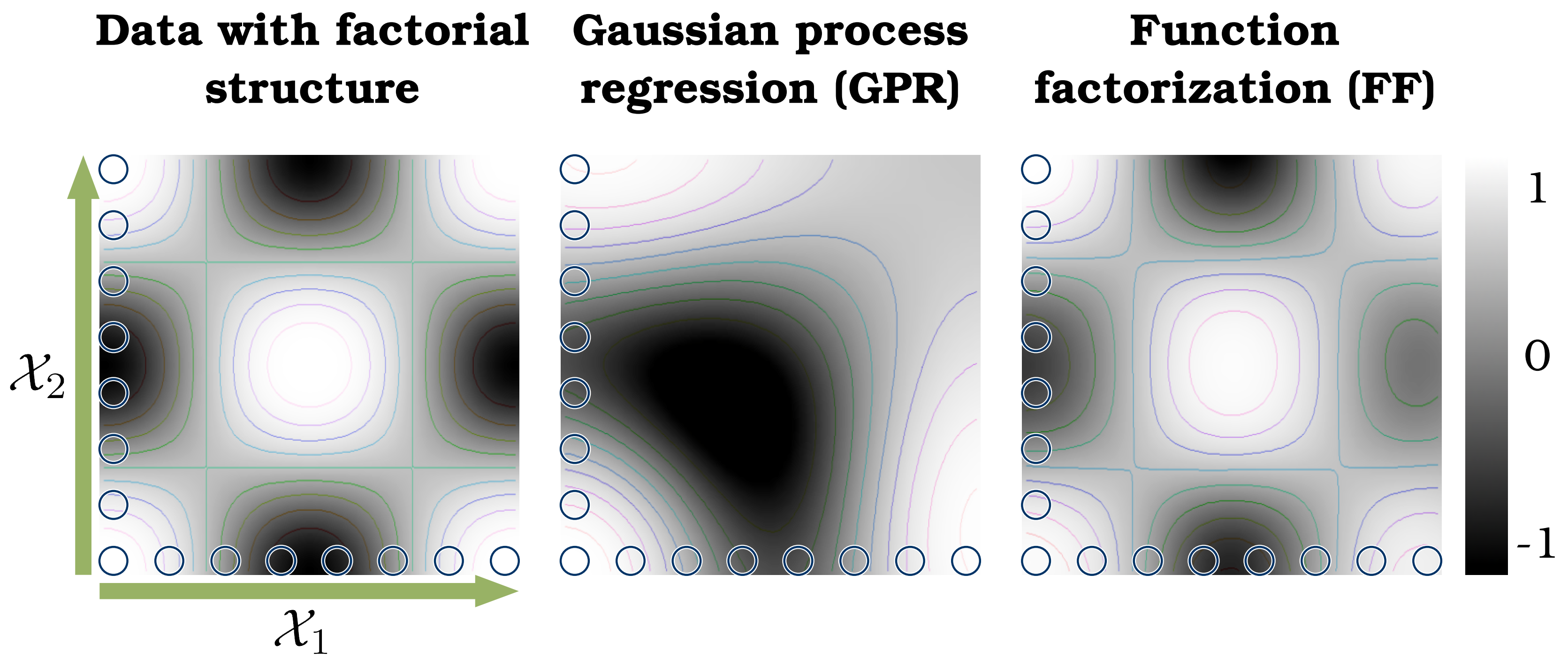
Input
$\boldsymbol{x} \in \mathcal{X}$

Set of functions
$f^{(i,k)} : \mathcal{X}_i \to \mathbb{R}$

Input in subspace
$\boldsymbol{x}_i \in \mathcal{X}_i$

Inference: Make predictions by integrating out the functions.

3 COMPARISON TO
GAUSSIAN PROC. REGRESSION

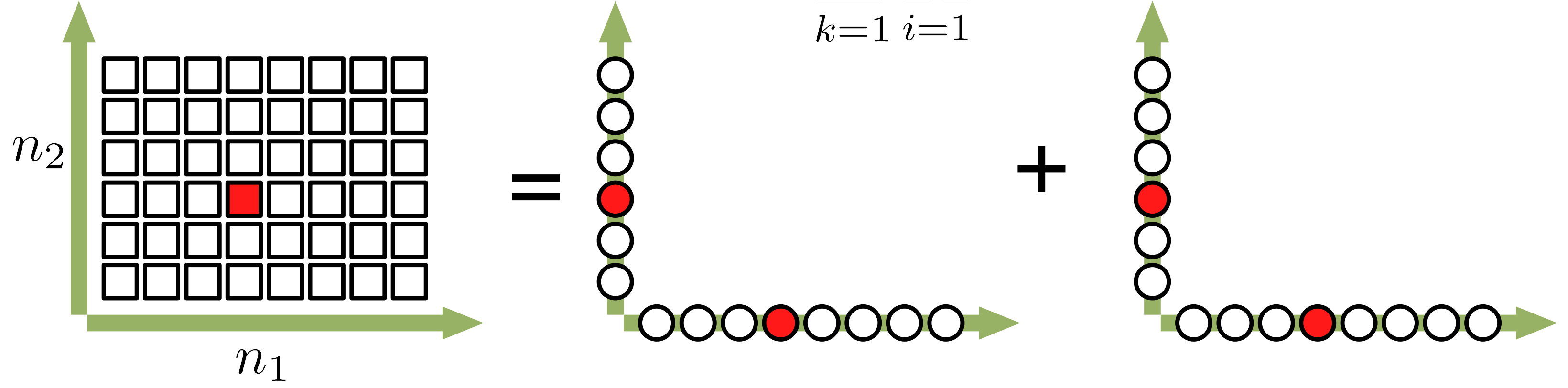


Data with factorial structure | **Gaussian process regression (GPR)** | **Function factorization (FF)**

Data has a factorized structure, and observations are taken at the indicated points. **GPR**: Fits well in the region close to observations and reverts to its mean away from observations. **FF**: Fits well in the entire region because of the correct assumption of factorized structure.

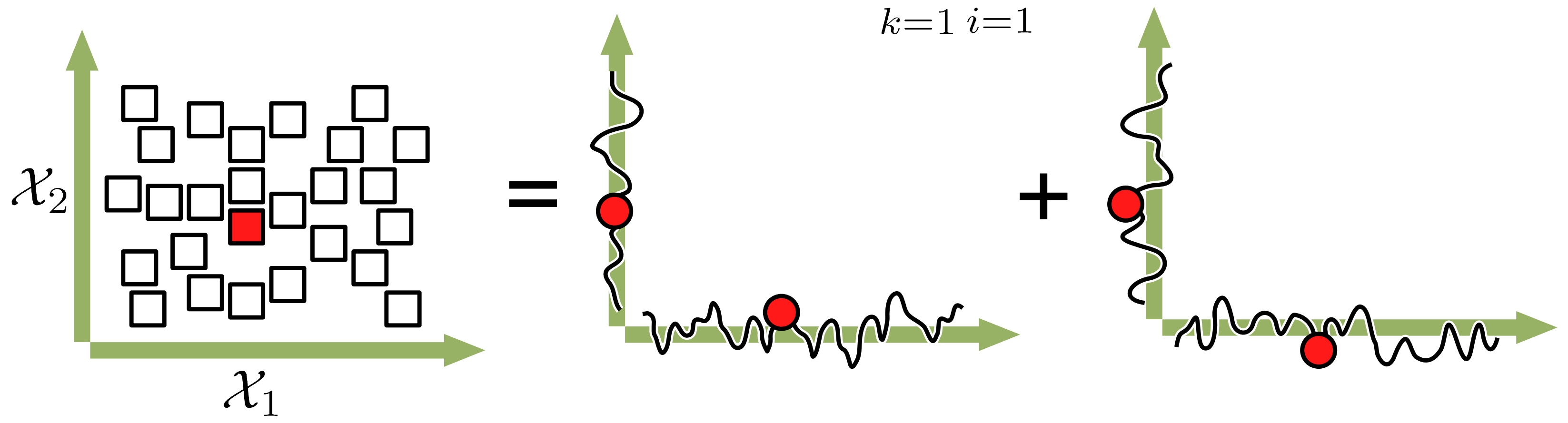
4 COMPARISON TO
MATRIX FACTORIZATION

Matrix factorization: $y(n_1, n_2) = \sum_{k=1}^{2} \prod_{i=1}^{2} f_{n_i}^{(i,k)}$ $\quad \boldsymbol{Y} = \boldsymbol{F}_1^\top \boldsymbol{F}_2$



Data is modelled as a sum of outer products of *vectors*.

Function factorization: $y(x_1, x_2) = \sum_{k=1}^{2} \prod_{i=1}^{2} f^{(i,k)}(x_i)$



Data is modelled as a sum of outer products of *functions*.

5 PRIORS OVER FUNCTIONS

Parametric functions with priors over the parameters.
Gaussian processes (GP) are flexible non-parametric distributions over functions, limited by the assumption of joint Gaussianity.
Warped Gaussian processes (WGP) [4] overcome this limitation by passing a GP through a non-linear warp function.

$$f(\boldsymbol{x}) = h(g(\boldsymbol{x})) \qquad g(\boldsymbol{x}) \sim \mathcal{GP}\big(m(\boldsymbol{x}), c(\boldsymbol{x}, \boldsymbol{x}')\big)$$

Non-linear
warp function

Gaussian
process

Mean
function

Covariance
function

6 INFERENCE

Hamiltonean Markov chain Monte Carlo [2]
- Avoids random walk associated with other MCMC methods.
- Requires gradients of the log-posterior wrt. all parameters.

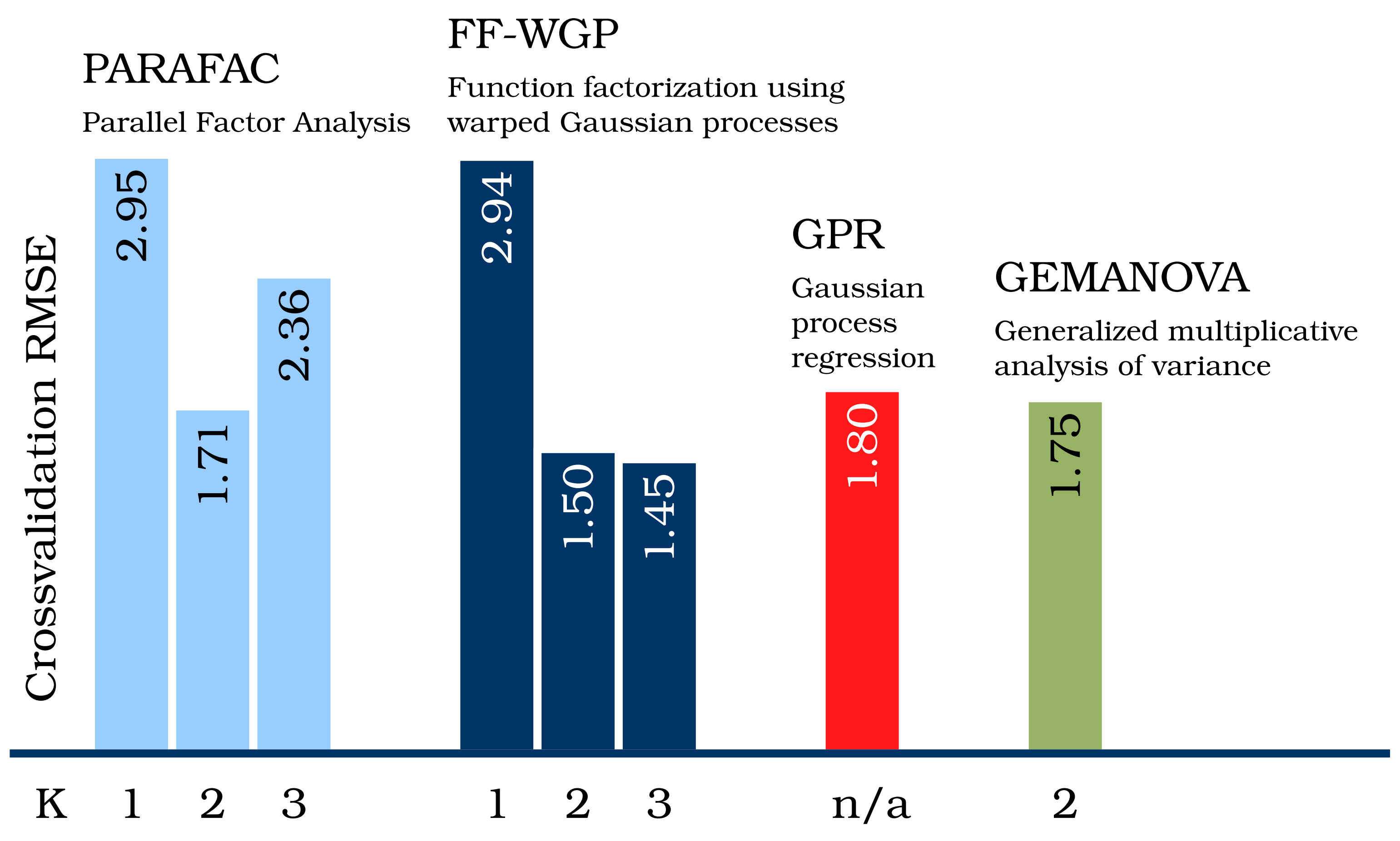
Numerically integrate out all parameters, including
- Likelihood function (e.g. noise variance)
- GP latent variables
- Covariance function
- Link function

7 EXPERIMENTS

Data: Measurements of the color of fresh beef as it changes during storage under different conditions [1]: Storage time, temperature, oxygen content, and light exposure.

- Reduced factorial experimental design: Measurements taken at a subset of the possible combinations of input variables.
- Data represented as a 5-D tensor with 60 pct. missing values.
- Input data points lie on a regular grid: Required for tensor factorization; not required for function factorization.

Results



8 CONCLUSIONS

- New approach to non-linear, non-parametric regression.
- Combines the ideas of a factorized model with non-parametric Bayesian regression.
- Outperforms tensor factorization and Gaussian process regression in terms of predictive performance.

REFERENCES

[1] Bro, R., & Jakobsen, M. (2002). Exploring complex interactions in designed data using GEMANOVA. Color changes in fresh beef during storage. Chemometrics, Journal of, 16, 294–304.
[2] Duane, S., Kennedy, A. D., Pendleton, B. J., & Roweth, D. (1987). Hybrid Monte Carlo. Physics Letters B, 195, 216–222.
[3] Rasmussen, C. E., & Williams, C. K. I. (2006). Gaussian processes for machine learning. MIT Press.
[4] Snelson, E., Rasmussen, C. E., & Ghahramani, Z. (2004). Warped gaussian processes. Neural Information Processing Systems, Advances in (NIPS) (pp. 337–344).