PROWLER.io

Variational Fourier Features for Gaussian Processes

James Hensman

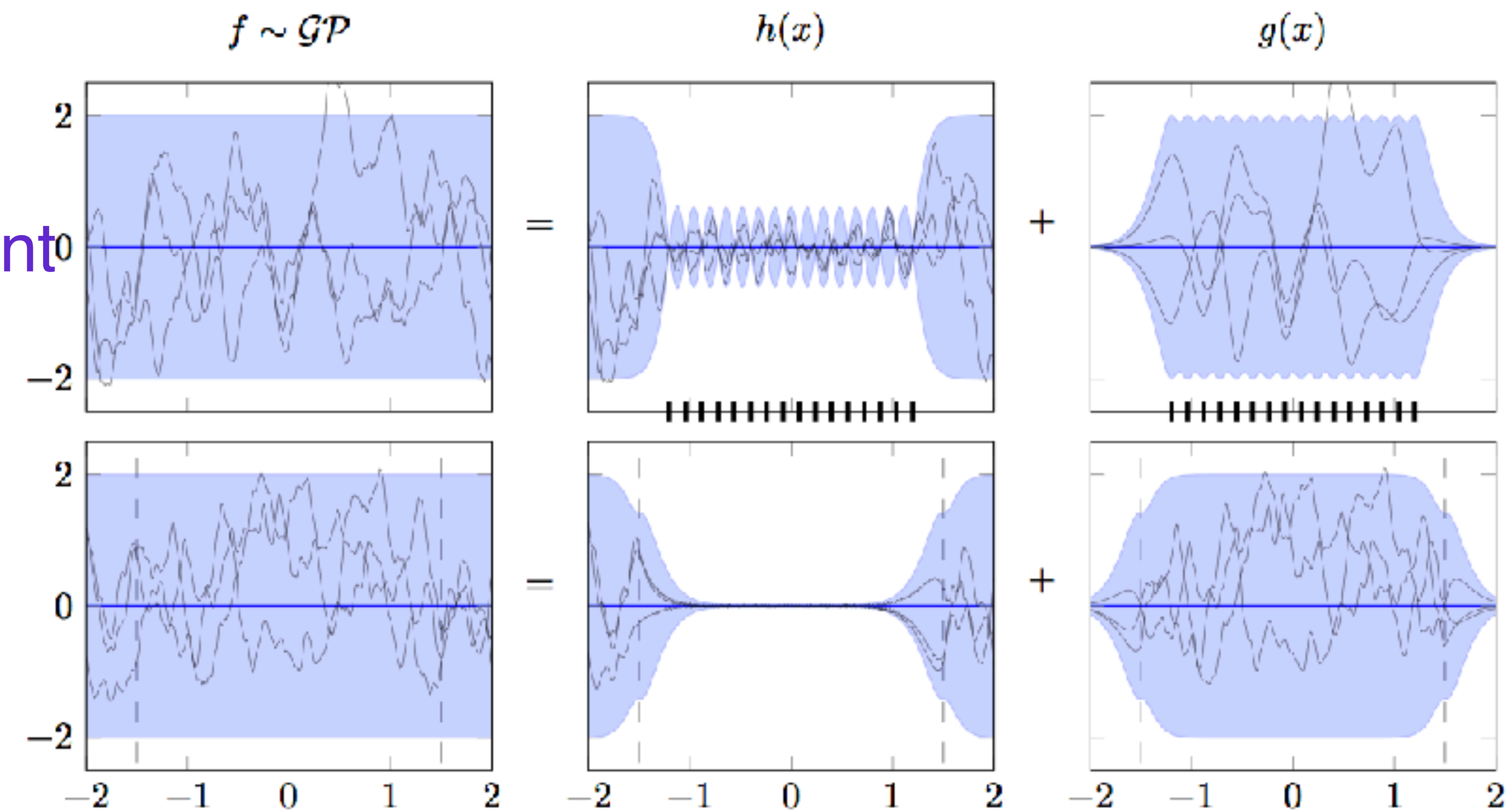
joint work with Nicolas Durrande, Arno Solin

Variational framework for GP approximation

$$\mathrm{KL}[q(f(\cdot))||p(f(\cdot)|\mathcal{D})]$$

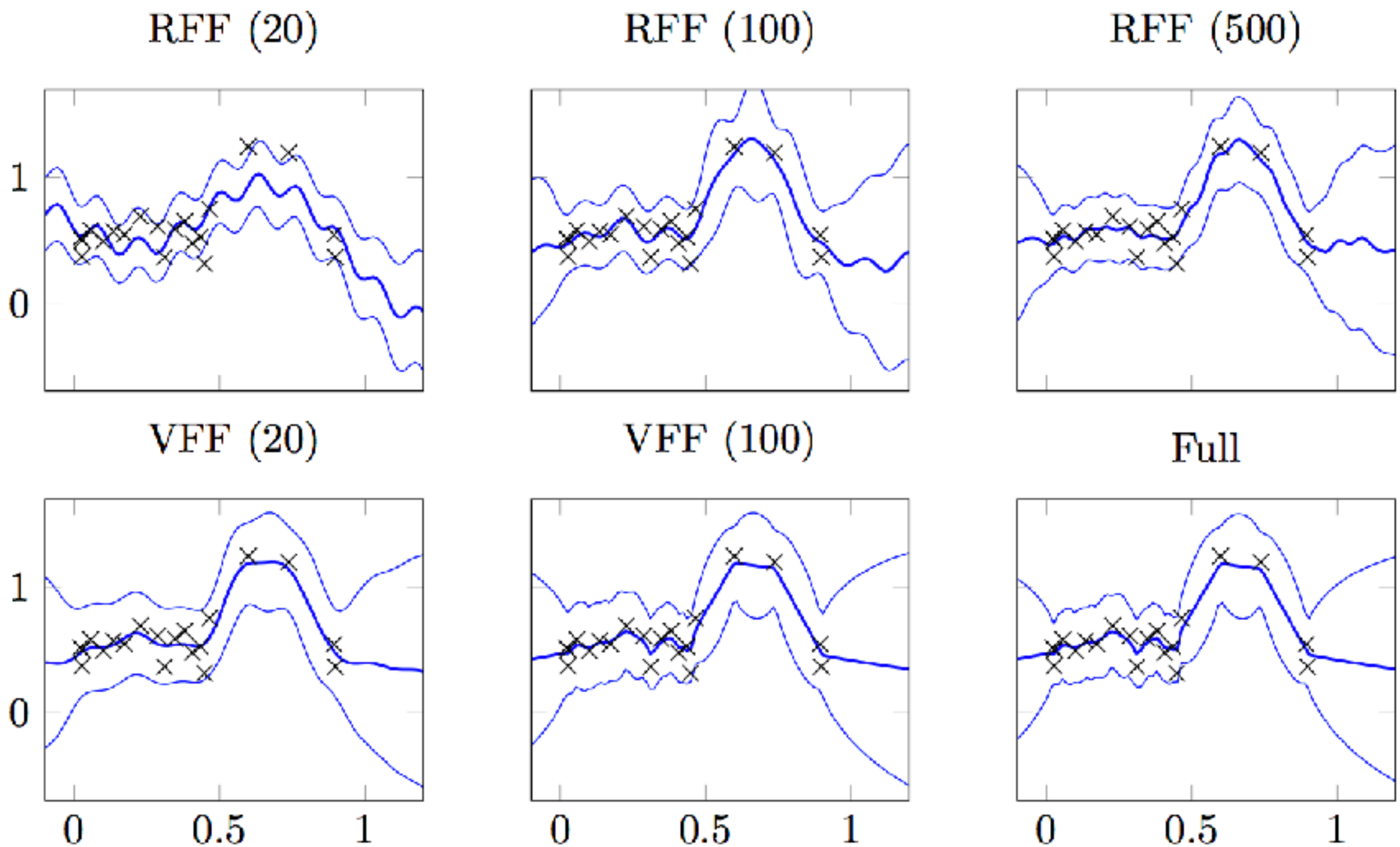
pseudo-point

Fourier



Random Fourier features

$$k(r) \approx \tilde{k}(r) = \frac{4\pi\sigma^2}{M} \sum_{m=1}^{M} \cos(\omega_m r)$$


RFF (20)
RFF (100)
RFF (500)
VFF (20)
VFF (100)
Full
1
0
1
0
0
0.5
1
0
0.5
1
0
0.5
1

The trouble with Fourier features

$$u_m = \int_{-\infty}^{\infty} e^{-\omega_m x} f(x) \mathrm{d}x$$

Our solution: truncate

$$u_m = \int_a^b e^{-\omega_m x} f(x) \mathrm{d}x$$

We use RKHS inner product

$$u_m = \langle e^{-\omega_m x},\ f(x) \rangle_{\mathcal{H}_{[a,b]}}$$

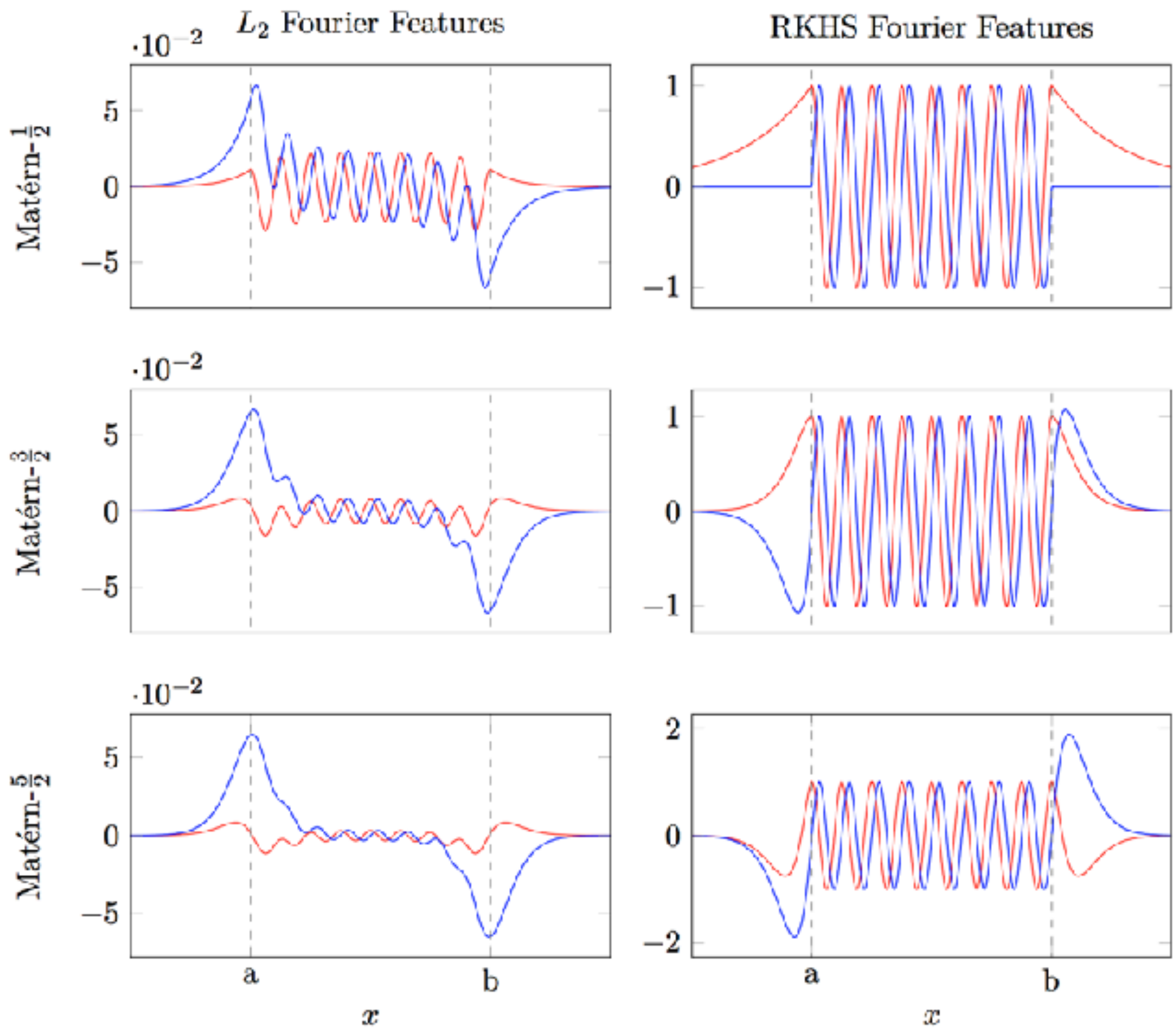
Kfu and Kuu are nice!

$$\text{cov}(f(\cdot), u_m) = \langle e^{-\omega_m x},\ k(x, \cdot)\rangle_{\mathcal{H}_{[a,b]}} = e^{-\omega_m x}$$

$$\text{cov}(u_m, u_{m'}) = \langle e^{-\omega_m x},\ e^{-\omega_{m'} x}\rangle_{\mathcal{H}_{[a,b]}} \approx \text{diag}(S(\omega_m))$$

cov(u_m, f(.))

L2 vs RKHS features



Beyond 1 dimension

$$k(\mathbf{x}, \mathbf{x}') = \sum_i k_i(x_i, x_i')$$

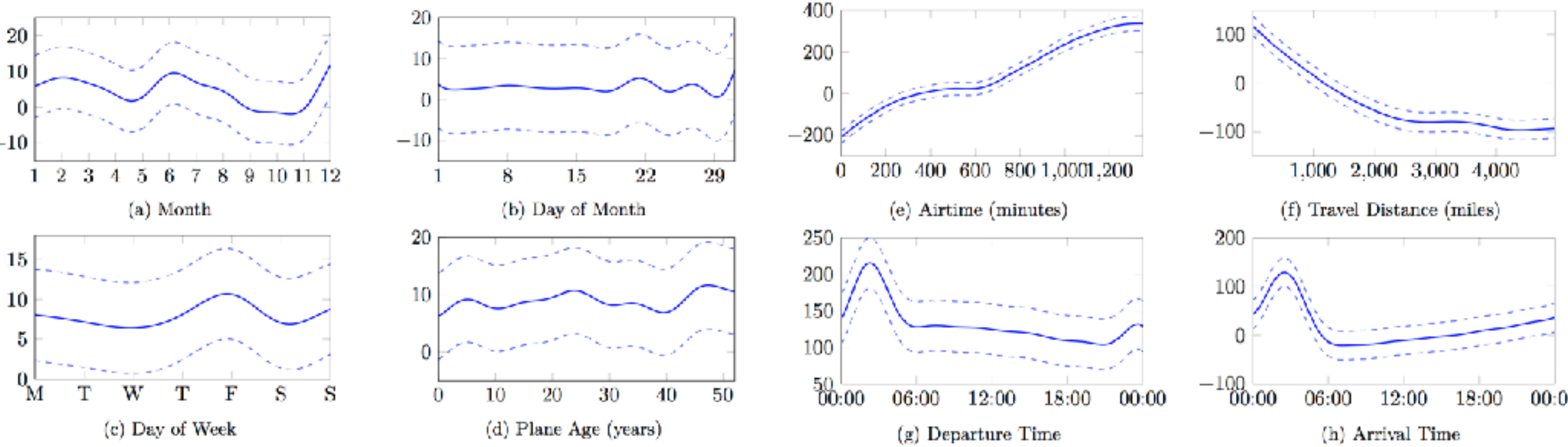
$$k(\mathbf{x}, \mathbf{x}') = \prod_i k_i(x_i, x_i')$$

Additive airline results

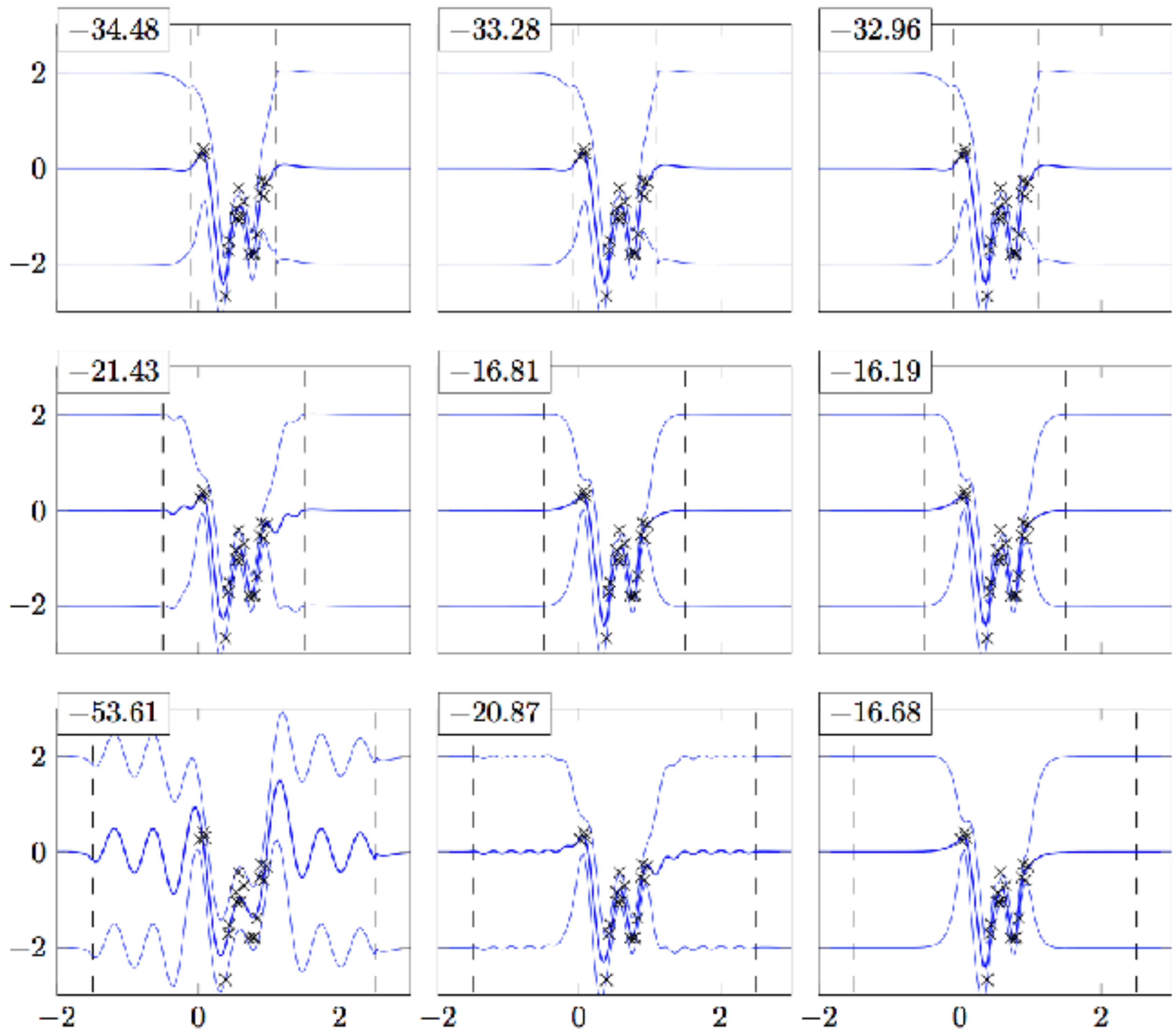
N	10,000		100,000		1,000,000		5,929,413	
	MSE	NLPD	MSE	NLPD	MSE	NLPD	MSE	NLPD
VFF	0.89 ± 0.15	1.362 ± 0.091	0.82 ± 0.05	1.319 ± 0.030	0.83 ± 0.01	1.326 ± 0.008	0.827 ± 0.004	1.324 ± 0.003
Full-RBF	0.89 ± 0.16	1.349 ± 0.098	N/A	N/A	N/A	N/A	N/A	N/A
Full-additive	0.89 ± 0.16	1.362 ± 0.096	N/A	N/A	N/A	N/A	N/A	N/A
SVIGP	0.90 ± 0.14	1.354 ± 0.096	0.81 ± 0.04	1.299 ± 0.033	0.83 ± 0.01	1.301 ± 0.009	0.83 ± 0.01	1.300 ± 0.003
String GP†	1.03 ± 0.10	N/A	0.93 ± 0.03	N/A	0.93 ± 0.01	N/A	0.90 ± 0.01	N/A
rBCM†	1.06 ± 0.10	N/A	1.04 ± 0.04	N/A	N/A	N/A	N/A	N/A

Dramatically faster than competing methods! 5min vs 91 hours

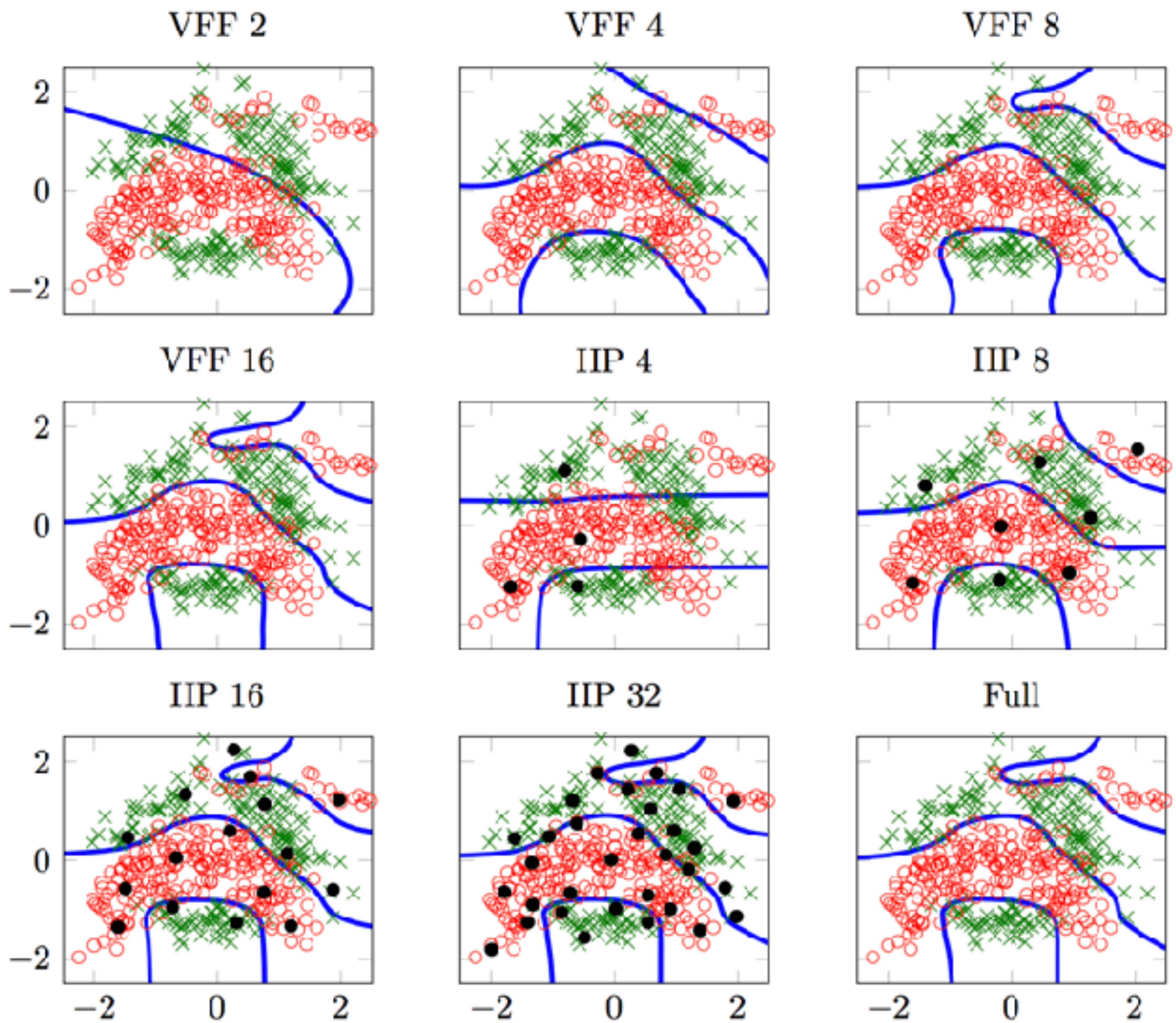
Additive airline results



Changing the limits

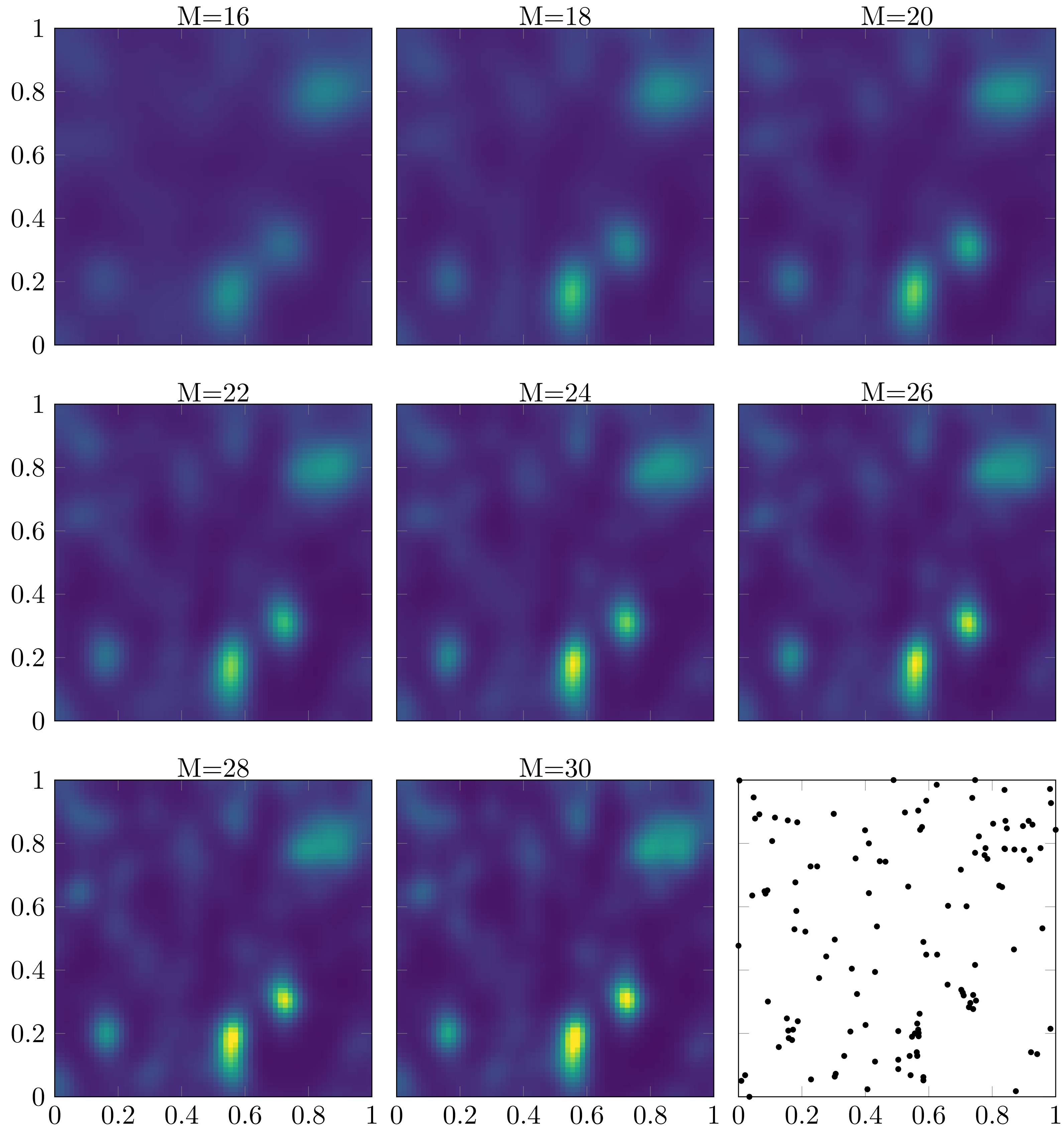


Banana dataset



Log Gaussian Cox processes

(with MCMC)



Comments

- All the goodness of the variational method
- Global, not local features
- Exceptional scaling in N
- Horrible scaling in D for Kronecker structures
- Additive models help scaling in D: appropriate?
- Might be good for LVMS, deeper structures with low D
- NUFFT might make it even faster

Thank you!

github.com/jameshensman/vff

arxiv.org/1611.06740

Email: james@prowler.io

Tweet: @jameshensman

PROWLER.io