VaSE: A Dynamic Perspective On Gaussian Processes

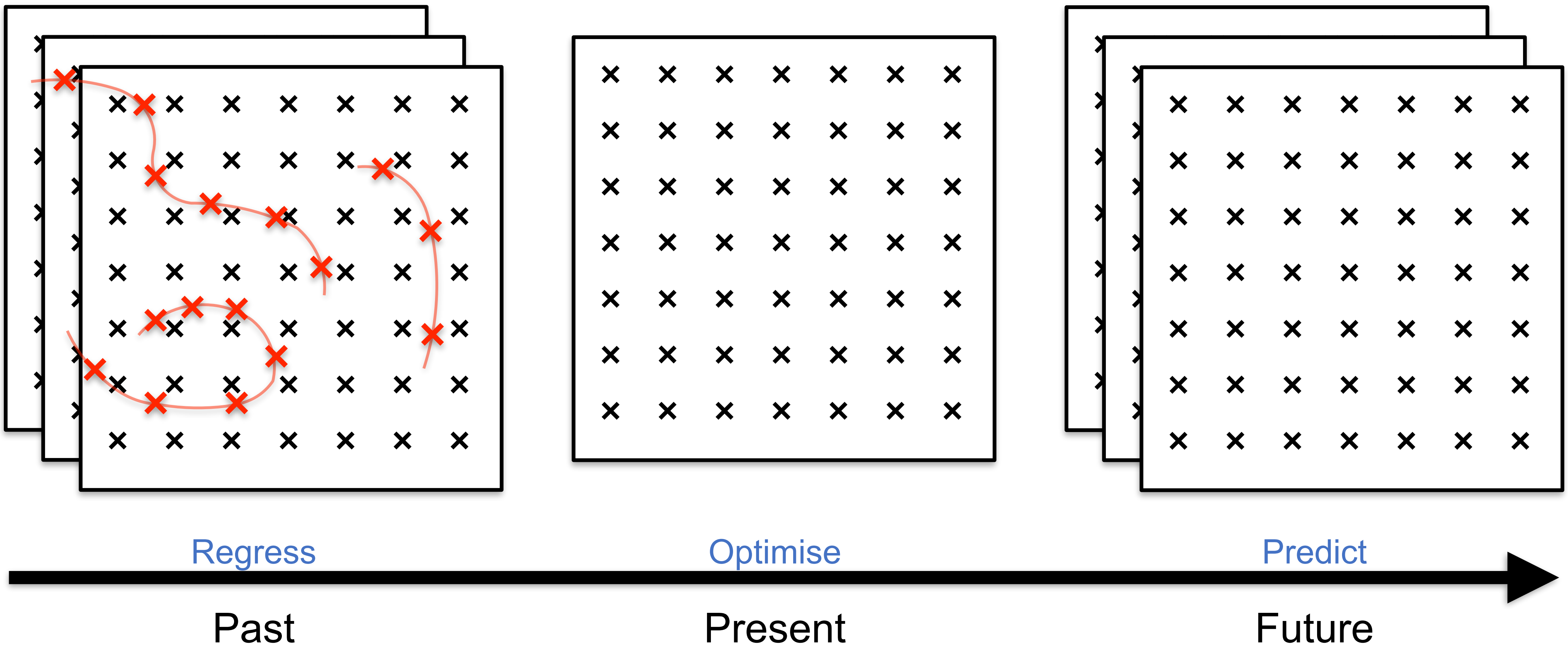
CSML Reading Group

30 Oct 2025

Rui-Yang Zhang

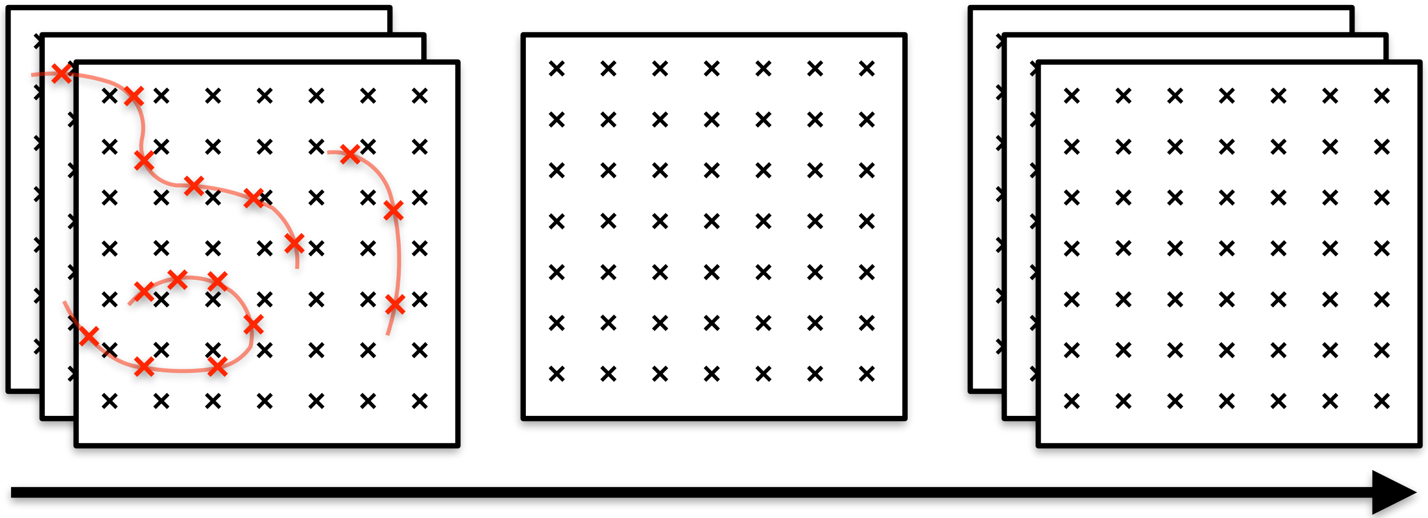
Zhang, R.-Y., Moss, H. B., Astfalck, L., Cripps, E. and Leslie, D. S. (2025). BALLAST: Bayesian Active Learning with Look-ahead Amendment for Sea-drifter Trajectories under Spatio-Temporal Vector Fields, *arXiv preprint* arXiv:2509.26005 .

Spatio-Temporal Inference with Disjoint Spatial Locations



Zhang, R.-Y., Moss, H. B., Astfalck, L., Cripps, E. and Leslie, D. S. (2025). BALLAST: Bayesian Active Learning with Look-ahead Amendment for Sea-drifter Trajectories under Spatio-Temporal Vector Fields, *arXiv preprint* arXiv:2509.26005 .

Spatio-Temporal Inference with Disjoint Spatial Locations



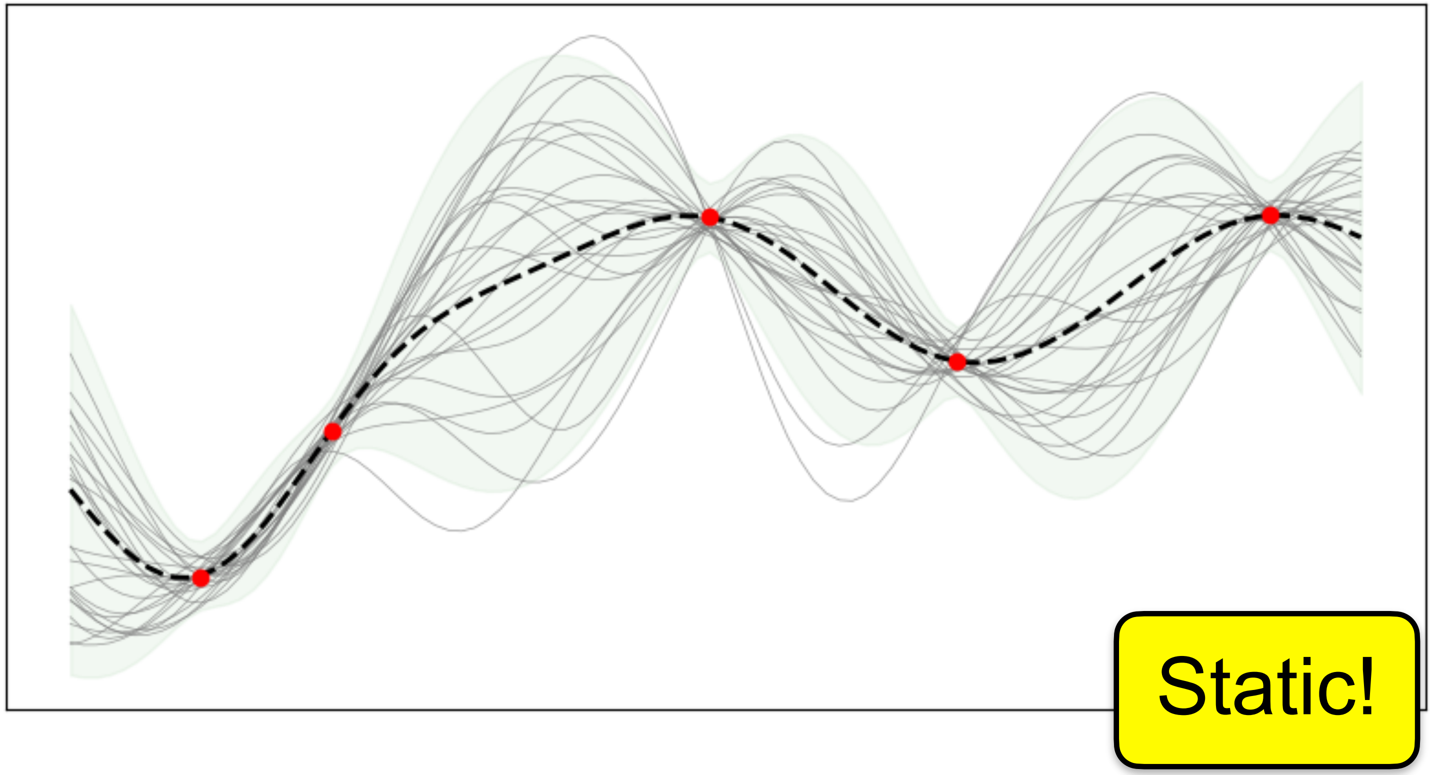
Vanilla GP
- Cheap to optimise
- Expensive to predict

SPDE-GP
- Cheap to predict once regressed
- Expensive to regress

Gaussian Processes [Static]

Lancaster University

$$f \sim GP(\mu, k)$$



Weight-Space View

$$f(x) = \phi(x)^T w, \quad w \sim N(0, \Sigma_p)$$

Function-Space View

$$f(x) \sim N(\mu(x), k(x, x))$$

Posterior Predictive

$$y_* \,|\, x_*, \mathcal{D}, f \sim N(\mu_{y_*|\mathcal{D}}, K_{y_*|\mathcal{D}})$$

$$\mu_{y_*|\mathcal{D}} = \mu(X_*) + K_*^T (K + \sigma^2 I)^{-1} (y - \mu(X))$$

$$K_{y_*|\mathcal{D}} = K_*^T (K + \sigma^2 I)^{-1} K_*$$

Williams, C. K. and Rasmussen, C. E. (2006). *Gaussian Processes for Machine Learning*, MIT press, Cambridge, MA.

Gaussian Processes [Dynamic]

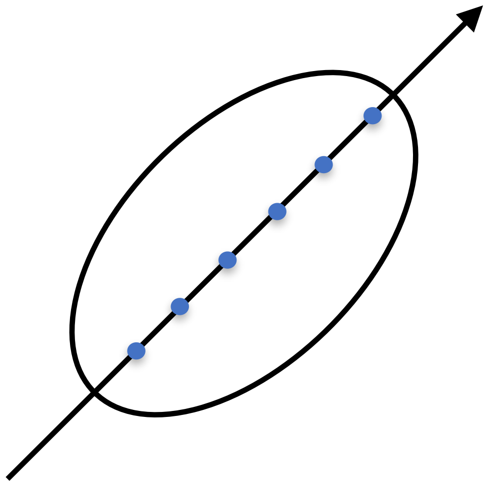
JOURNAL ARTICLE

ON STATIONARY PROCESSES IN THE PLANE

Get access >

P. WHITTLE

Biometrika, Volume 41, Issue 3-4, 3 December 1954, Pages 434–449,
https://doi.org/10.1093/biomet/41.3-4.434
Published: 03 December 1954



$$L(T)f_t := [-aT^{-1} + T^0 - bT]f_t$$
$$= -af_{t-1} + f_t - bf_{t+1} = \varepsilon_t$$

$$f_t = L^{-1}(T)\varepsilon_t$$

White Noise ε $\xrightarrow{\text{Filter } L^{-1}}$ Target Process f

Filter L yields the transfer function $L(e^{i\omega}) = \sum_{j=-\infty}^{\infty} a_j e^{i\omega j}$.

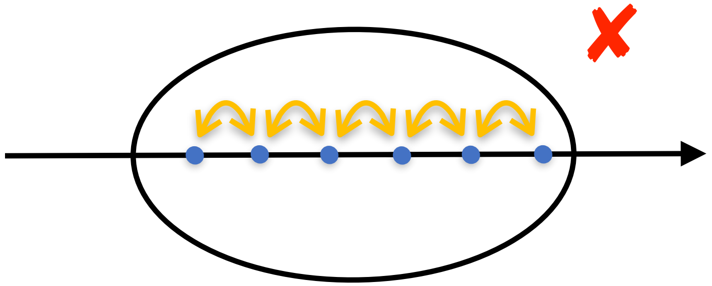
So the spectral density is $S_f(\omega) = \dfrac{\sigma^2}{L(e^{i\omega})L(e^{-i\omega})}$.

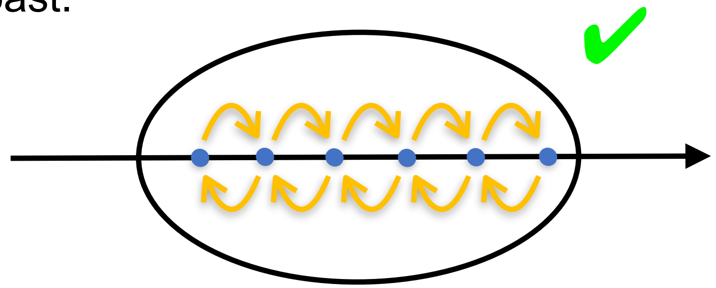
Whittle, P. (1954). On stationary processes in the plane, *Biometrika*. pp. 434–449.

Hartikainen, J. and Särkkä, S. (2010). Kalman filtering and smoothing solutions to temporal Gaussian process regression models, *2010 IEEE international workshop on machine learning for signal processing*, IEEE, pp. 379– 384.

Gaussian Processes [Dynamic]

<ins>Goal</ins>: Find a filter L that
(1) matches with desired spectrum,
(2) is causal, i.e. future only depends on past.





Kalman filtering and smoothing solutions to temporal Gaussian process regression models

Publisher: IEEE

Cite This | PDF

Jouni Hartikainen ; Simo Särkkä **All Authors**

The filter $(\lambda + \nabla)^{p+1}$ is causal and yields Matérn-$(p + 1/2)$ spectrum.

Example (Matérn-3/2): Solve for $f''(t) + 2\lambda f'(t) + \lambda^2 f(t) = \varepsilon(t)$.

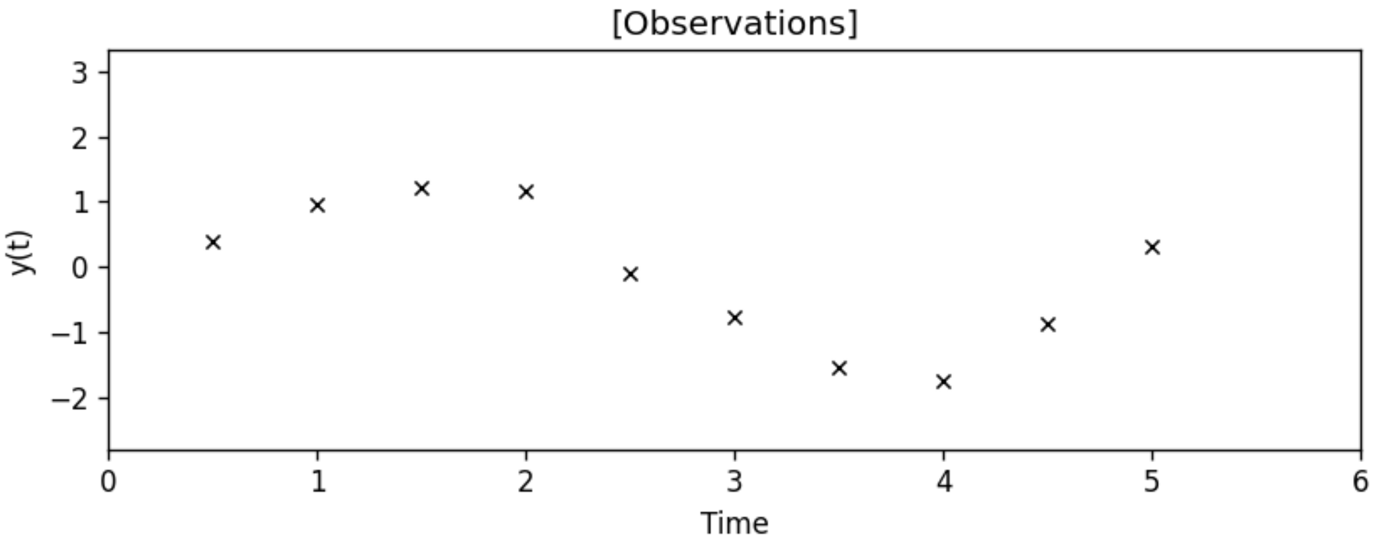
Hartikainen, J. and Särkkä, S. (2010). Kalman filtering and smoothing solutions to temporal Gaussian process regression models, *2010 IEEE international workshop on machine learning for signal processing*, IEEE, pp. 379– 384.

Gaussian Processes [Dynamic]

Lancaster University

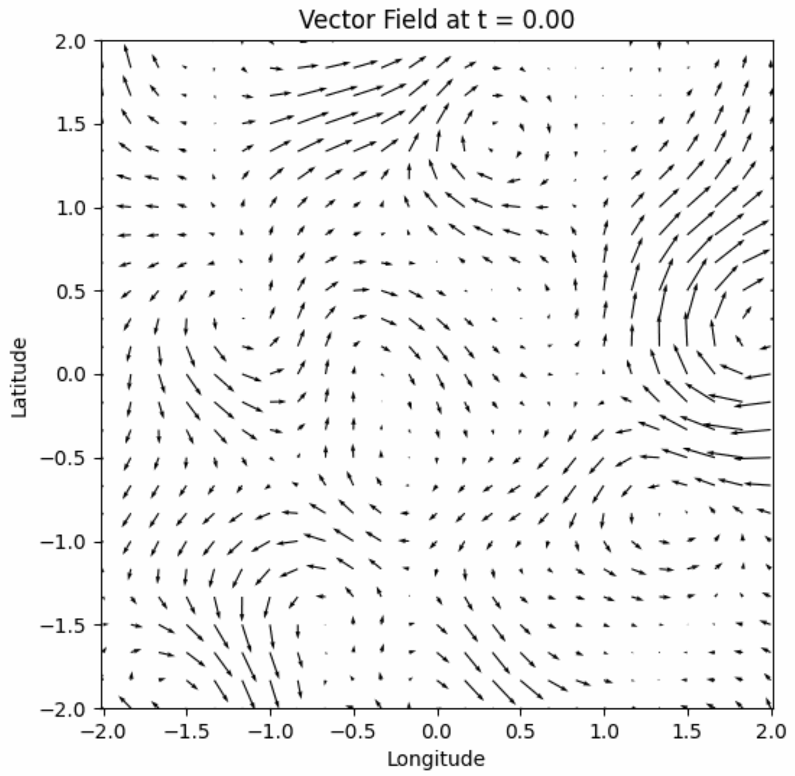
$$f''(t) + 2\lambda f'(t) + \lambda^2 f(t) = \varepsilon(t)$$

Linear SDE
- produce exact transition density
- observations are linear emission
- admit sequential inference via Kalman filter & smoother



Sarkka, S., Solin, A. and Hartikainen, J. (2013). Spatio-Temporal learning via infinite-dimensional Bayesian filtering and smoothing: A look at Gaussian process regression through Kalman filtering, *IEEE Signal Processing Magazine* 30(4): 51–61.

Gaussian Processes [Dynamic]



$_{ie}$, the spectral density is the product $S(\omega_1, \omega_2) = S_{space}(\omega_1)S_{time}(\omega_2)$.

different white noise

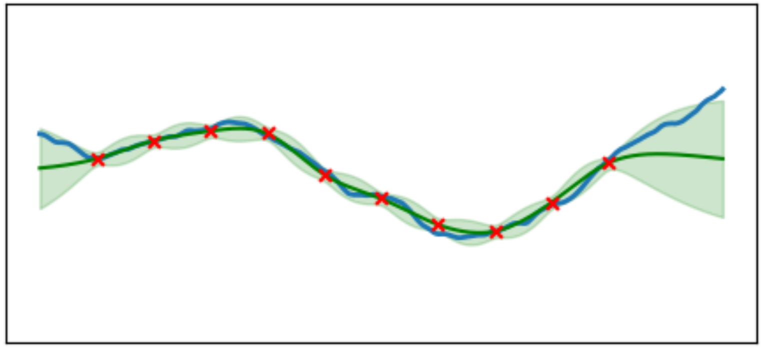
the SDE form, we have $S(\omega_1, \omega_2) = \dfrac{\sigma^2 S_{space}(\omega_1)}{L(e^{i\omega_2})L(e^{-i\omega_2})}$.

same dynamics

$$_t\, \boldsymbol{f}(\boldsymbol{x}, t) = F_{full}\boldsymbol{f}(\boldsymbol{x}, t)dt + L_{full}d\boldsymbol{w}(x, t)$$

Vanilla-SPDE Exchange (VASE)

Static



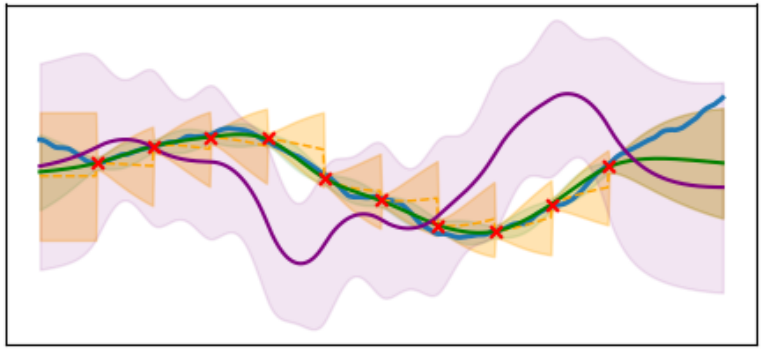
$$\boldsymbol{f} = [f, \nabla_t f]^T$$

State Extension

State Extraction

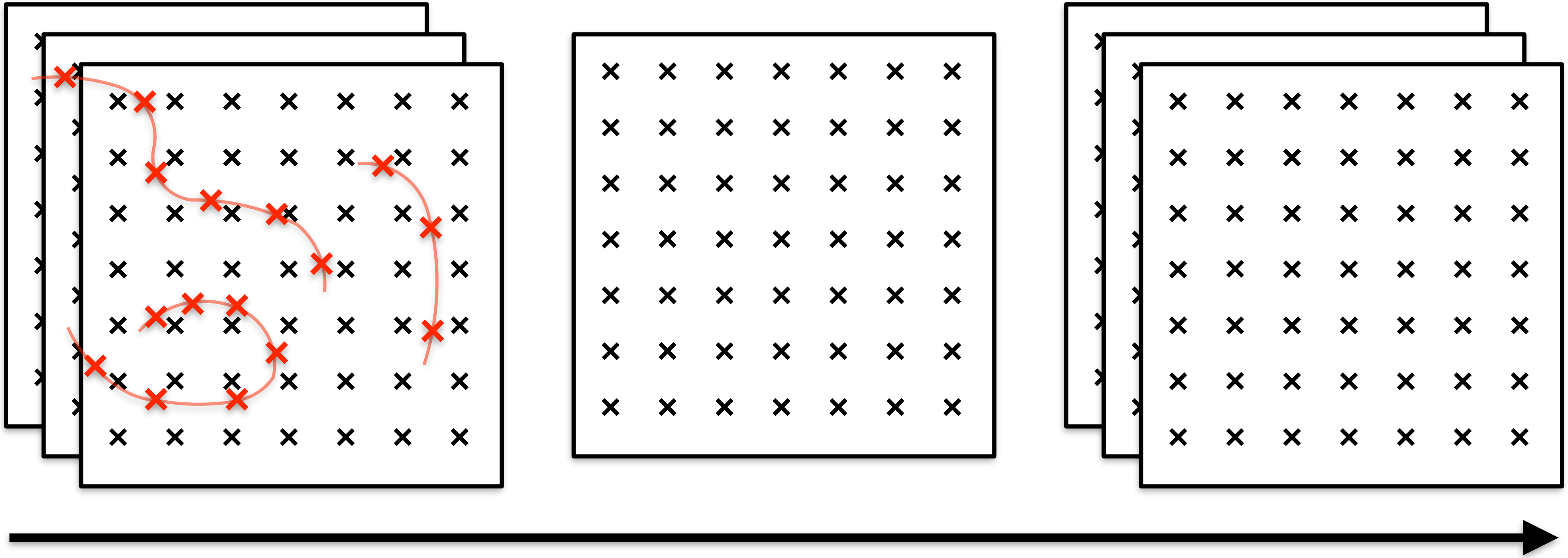
$$f = G \cdot \boldsymbol{f}$$

Dynamic



Zhang, R.-Y., Moss, H. B., Astfalck, L., Cripps, E. and Leslie, D. S. (2025). BALLAST: Bayesian Active Learning with Look-ahead Amendment for Sea-drifter Trajectories under Spatio-Temporal Vector Fields, *arXiv preprint* arXiv:2509.26005 .

Spatio-Temporal Inference with Disjoint Spatial Locations



Regress using static GP with extended state space

[Vanilla-SPDE Exchange]

Predict using dynamic GP and extract original states

Zhang, R.-Y., Moss, H. B., Astfalck, L., Cripps, E. and Leslie, D. S. (2025). BALLAST: Bayesian Active Learning with Look-ahead Amendment for Sea-drifter Trajectories under Spatio-Temporal Vector Fields, *arXiv preprint* arXiv:2509.26005 .

Spatio-Temporal Inference with Disjoint Spatial Locations

Lancaster University

Denote the number of spatial grid points as N_s for each time slice, and the number of time slices we wish to sample into the future is N_t. So, the total number of test points of posterior prediction is $N_t N_s$. The observation number N_{obs}, which we, for simplicity, assume to be made at distinct locations over $N_{\text{obs,t}}$ time slices. We remark also that the size of both N_t and N_s would often be much larger than N_{obs}, while N_t and $N_{\text{obs,t}}$ are of the same magnitude.

Method	Regression	Sampling	Total
Vanilla	$O(N_{\text{obs}}^3)$	$O(N_s^3 N_t^3)$	$O(N_{\text{obs}}^3 + N_s^3 N_t^3)$
SPDE	$O((N_s + N_{\text{obs}})^3 N_{\text{obs,t}})$	$O(N_s^3 + N_s^2 N_t)$	$O((N_s + N_{\text{obs}})^3 N_{\text{obs,t}} + N_s^3 + N_s^2 N_t)$
VASE	$O(N_{\text{obs}}^3 + N_s^2 N_{\text{obs}} + N_s N_{\text{obs}}^2)$	$O(N_s^3 + N_s^2 N_t)$	$O(N_{\text{obs}}^3 + N_s^2 N_{\text{obs}} + N_s N_{\text{obs}}^2 + N_s^3 + N_s^2 N_t)$

'commission fee' of the exchange

Table 1: Computational cost summary table of different methods. The green indicates the lowest cost in each column, while the red indicates the highest.

Vanilla-SPDE Exchange (VASE)