A. Supplementary Material

A.1. Parameter Estimation Lorenz '63

Ground truth	NPSDE	ESGF	AReS	MaRS
$\theta_0 = 10$	1.28 ± 2.32	$\mathbf{9.97 \pm 0.33}$	7.24 ± 1.08	9.82 ± 0.56
$\theta_1 = 28$	20.69 ± 5.73	$\mathbf{28.00 \pm 0.17}$	28.16 ± 1.08	27.96 ± 0.21
$\theta_0 = 2.667$	1.86 ± 1.08	$\mathbf{2.65 \pm 0.06}$	2.55 ± 0.10	2.64 ± 0.07
$G = \sqrt{10}$	6.51 ± 1.31	$\mathbf{3.03 \pm 0.2}$	3.54 ± 2.45	

Table 2. Median and standard deviation of the 65 best runs of each algorithm. As ESGF crashed in roughly one third of all experiments, we compare only the best 65 runs, where a crash is treated as a complete failure. While this provides somehow a fair comparison, it should be noted that this significantly overestimates the performance of all algorithms.

A.2. Training Times

	NPSDE	VGPA	ESGF	AReS	MaRS
OU Process	48.8 ± 0.9	$\sim (54 \pm 8)$hours	32.2 ± 0.3	321.3 ± 0.8	17.3 ± 0.3
DW Potential	406.9 ± 147.9	$\sim (12 \pm 6)$hours	35.2 ± 0.1	326.0 ± 1.8	17.7 ± 2.0
Lotka-Volterra	1421.8 ± 1.0	/	244.7 ± 1.2	47.6 ± 1.3	19.3 ± 1.1
Lorenz '63	39273.5 ± 8.9	/	670.7 ± 10.7	26274.0 ± 2529.8	721.1 ± 10.7

Table 3. Computational times (in seconds) required for training the different algorithms.

A.3. Densities for Ancestral Sampling of the SDE-Based Model

Given the graphical model in Figure 1a, it is straightforward to compute the densities used in the ancestral sampling scheme in Algorithm 1. After marginalizing out $\dot{\mathbf{z}}$, the joint density described by the graphical model can be written as

$$p(\mathbf{o}, \mathbf{z}, \mathbf{y} | \boldsymbol{\phi}, \mathbf{G}, \boldsymbol{\sigma}) = p(\mathbf{o}|\mathbf{G})p(\mathbf{z}|\boldsymbol{\phi})p(\mathbf{y}|\mathbf{z}, \mathbf{o}, \boldsymbol{\sigma}) \tag{34}$$

Substituting the densities given by Equations (10), (11), (13) and (21) yields

$$p(\mathbf{o}, \mathbf{z}, \mathbf{y} | \boldsymbol{\phi}, \mathbf{G}, \boldsymbol{\sigma}) = \mathcal{N}(\mathbf{o}|\mathbf{0}, \mathbf{B}\boldsymbol{\Omega}\mathbf{B}^T)\mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{C}_\phi)\mathcal{N}(\mathbf{y}|\mathbf{z} + \mathbf{o}, \mathbf{T}). \tag{35}$$

Using a change of variables to simplify notation, we write

$$p(\mathbf{o}, \mathbf{z}, \mathbf{y} | \boldsymbol{\phi}, \mathbf{G}, \boldsymbol{\sigma}) = \mathcal{N}(\mathbf{o}|\mathbf{0}, \tilde{\boldsymbol{\Omega}})\mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{C}_\phi)\mathcal{N}(\mathbf{y}|\mathbf{z} + \mathbf{o}, \mathbf{T}). \tag{36}$$

This equation is now subsequently modified by observing that the product of two Gaussian densities in the same random variable is again a Gaussian density:

$$
\begin{aligned}
p(\mathbf{o}, \mathbf{z}, \mathbf{y} | \boldsymbol{\phi}, \mathbf{G}, \boldsymbol{\sigma}) &= \mathcal{N}(\mathbf{o}|\mathbf{0}, \tilde{\boldsymbol{\Omega}})\mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{C}_\phi)\mathcal{N}(\mathbf{y}|\mathbf{z} + \mathbf{o}, \mathbf{T}) \\
&= \mathcal{N}(\mathbf{o}|\mathbf{0}, \tilde{\boldsymbol{\Omega}})\mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{C}_\phi)\mathcal{N}(\mathbf{z}|\mathbf{y} - \mathbf{o}, \mathbf{T}) \\
&= \mathcal{N}(\mathbf{o}|\mathbf{0}, \tilde{\boldsymbol{\Omega}})\mathcal{N}(\mathbf{y} - \mathbf{o}|\mathbf{0}, \mathbf{C}_\phi + \mathbf{T})\mathcal{N}(\mathbf{z}|\mathbf{m}_z, \mathbf{C}_z) \\
&= \mathcal{N}(\mathbf{o}|\mathbf{0}, \tilde{\boldsymbol{\Omega}})\mathcal{N}(\mathbf{o}|\mathbf{y}, \mathbf{C}_\phi + \mathbf{T})\mathcal{N}(\mathbf{z}|\mathbf{m}_z, \mathbf{C}_z) \\
&= \mathcal{N}(\mathbf{y}|\mathbf{0}, \tilde{\boldsymbol{\Omega}} + \mathbf{C}_\phi + \mathbf{T})\mathcal{N}(\mathbf{o}|\mathbf{m}_o, \mathbf{C}_o)\mathcal{N}(\mathbf{z}|\mathbf{m}_z, \mathbf{C}_z)
\end{aligned} \tag{37}
$$

where

$$\mathbf{m}_z = \mathbf{C}_z(\mathbf{T}^{-1}(\mathbf{y} - \mathbf{o})) \tag{38}$$

$$\mathbf{C}_z = (\mathbf{C}_\phi^{-1} + \mathbf{T}^{-1})^{-1} \tag{39}$$

$$\mathbf{m}_o = \mathbf{C}_o(\mathbf{C}_\phi + \mathbf{T})^{-1}\mathbf{y} \tag{40}$$

$$\mathbf{C}_o = (\tilde{\boldsymbol{\Omega}}^{-1} + (\mathbf{C}_\phi + \mathbf{T})^{-1})^{-1} \tag{41}$$

This formula can be further refined with the Woodbury identity, i.e.

$$\mathbf{C}_z = (\mathbf{C}_\phi^{-1} + \mathbf{T}^{-1})^{-1}$$
$$= \mathbf{C}_\phi - \mathbf{C}_\phi(\mathbf{C}_\phi + \mathbf{T})^{-1}\mathbf{C}_\phi$$
$$= \mathbf{C}_\phi(\mathbf{C}_\phi + \mathbf{T})^{-1}\mathbf{T} \tag{42}$$

which leads to

$$\mathbf{m}_z = \mathbf{C}_\phi(\mathbf{C}_\phi + \mathbf{T})^{-1}(\mathbf{y} - \mathbf{o}) \tag{43}$$

and

$$\mathbf{C}_o = (\tilde{\mathbf{\Omega}}^{-1} + (\mathbf{C}_\phi + \mathbf{T})^{-1})^{-1}$$
$$= \tilde{\mathbf{\Omega}} - \tilde{\mathbf{\Omega}}(\tilde{\mathbf{\Omega}} + \mathbf{C}_\phi + \mathbf{T})^{-1}\tilde{\mathbf{\Omega}}$$
$$= \tilde{\mathbf{\Omega}}(\tilde{\mathbf{\Omega}} + \mathbf{C}_\phi + \mathbf{T})^{-1}(\mathbf{C}_\phi + \mathbf{T}) \tag{44}$$

which leads to

$$\mathbf{m}_o = \tilde{\mathbf{\Omega}}(\tilde{\mathbf{\Omega}} + \mathbf{C}_\phi + \mathbf{T})^{-1}\mathbf{y} \tag{45}$$

Since we observe $\mathbf{y}$, we are interested in calculating the conditional distribution

$$p(\mathbf{o}, \mathbf{z}|\mathbf{y}, \phi, \mathbf{G}, \sigma) = \frac{p(\mathbf{o}, \mathbf{z}, \mathbf{y}|\phi, \mathbf{G}, \sigma)}{p(\mathbf{y}|\phi, \mathbf{G}.\sigma)} \tag{46}$$

Conveniently enough, the marginal density of $\mathbf{y}$ is already factorized out in Equation (37) (compare Equation (22)). Thus, we have

$$p(\mathbf{o}, \mathbf{z}|\mathbf{y}, \phi, \mathbf{G}, \sigma) = \mathcal{N}(\mathbf{o}|\mathbf{m}_o, \mathbf{C}_o)\mathcal{N}(\mathbf{z}|\mathbf{m}_z, \mathbf{C}_z) \tag{47}$$

As $\mathcal{N}(\mathbf{o}|\mathbf{m}_o, \mathbf{C}_o)$ is independent of $\mathbf{z}$, we can employ ancestral sampling by first obtaining a sample of $\mathbf{o}$ through $\mathcal{N}(\mathbf{o}|\mathbf{m}_o, \mathbf{C}_o)$, and then utilizing such sample to get $\mathbf{z}$ through $\mathcal{N}(\mathbf{z}|\mathbf{m}_z, \mathbf{C}_z)$.

A.4. Calculating the GP Posterior for Data-Based Ancestral Sampling

Given the graphical model in Figure 1b, we can calculate the densities used in the ancestral sampling scheme in Algorithm 1. After marginalizing out $\dot{\mathbf{z}}$ and using the variable substitutions introduced in Equation (36), the joint density described by the graphical model can be written as

$$p(\mathbf{o}, \mathbf{z}, \mathbf{y}|\phi, \mathbf{G}, \sigma) = p(\mathbf{o}|\mathbf{G})p(\mathbf{y}|\sigma, \mathbf{o}, \mathbf{z})p(\mathbf{z}|\phi)$$
$$= \mathcal{N}(\mathbf{o}|\mathbf{0}, \tilde{\mathbf{\Omega}})\mathcal{N}(\mathbf{y}|\mathbf{z} + \mathbf{o}, \mathbf{T})\mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{C}_\phi)$$
$$= \mathcal{N}(\mathbf{o}|\mathbf{0}, \tilde{\mathbf{\Omega}})\mathcal{N}(\mathbf{o}|\mathbf{y} - \mathbf{z}, \mathbf{T})\mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{C}_\phi)$$
$$= \mathcal{N}(\mathbf{o}|\mathbf{m}, \mathbf{C})\mathcal{N}(\mathbf{y} - \mathbf{z}|\mathbf{0}, \tilde{\mathbf{\Omega}} + \mathbf{T})\mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{C}_\phi)$$
$$= \mathcal{N}(\mathbf{o}|\mathbf{m}, \mathbf{C})\mathcal{N}(\mathbf{z}|\mathbf{y}, \tilde{\mathbf{\Omega}} + \mathbf{T})\mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{C}_\phi)$$
$$= \mathcal{N}(\mathbf{o}|\mathbf{m}, \mathbf{C})\mathcal{N}(\mathbf{y}|\mathbf{0}, \tilde{\mathbf{\Omega}} + \mathbf{T} + \mathbf{C}_\phi)\mathcal{N}(\mathbf{z}|\mu_z, \mathbf{\Sigma}_z), \tag{48}$$

where

$$\mu_z = \mathbf{\Sigma}_z(\tilde{\mathbf{\Omega}} + \mathbf{T})^{-1}\mathbf{y} \tag{49}$$
$$\mathbf{\Sigma}_z = ((\tilde{\mathbf{\Omega}} + \mathbf{T})^{-1} + \mathbf{C}_\phi^{-1})^{-1}$$
$$= \mathbf{C}_\phi - \mathbf{C}_\phi(\tilde{\mathbf{\Omega}} + \mathbf{T} + \mathbf{C}_\phi)^{-1}\mathbf{C}_\phi$$
$$= (\tilde{\mathbf{\Omega}} + \mathbf{T})(\tilde{\mathbf{\Omega}} + \mathbf{T} + \mathbf{C}_\phi)^{-1}\mathbf{C}_\phi$$
$$= \mathbf{C}_\phi(\tilde{\mathbf{\Omega}} + \mathbf{T} + \mathbf{C}_\phi)^{-1}(\tilde{\mathbf{\Omega}} + \mathbf{T}). \tag{50}$$

After marginalizing out $\mathbf{o}$ and dividing by the marginal of $\mathbf{y}$, we get the conditional distribution

$$p(\mathbf{z}|\mathbf{y}, \phi, \mathbf{G}, \sigma) = \mathcal{N}(\mathbf{z}|\mu_z, \mathbf{\Sigma}_z). \tag{51}$$