Supplementary Materials

A Gaussian process

Gaussian process (GP) is a typical choice for the surrogate model because of its model capacity for complicated black-box functions and uncertainty quantification. Consider, for the time being, a simplified scenario in which we have noise-contaminated observations $\{y_i = g(\mathbf{x}_i) + \epsilon_i\}_{i=1}^N$. In a GP model, a prior distribution is placed over $f(\mathbf{x})$, indexed by $\mathbf{x}$:

$$\eta(\mathbf{x})|\boldsymbol{\theta} \sim \mathcal{GP}\left(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'|\boldsymbol{\theta})\right), \tag{A.1}$$

with mean and covariance functions:

$$\begin{aligned} m_0(\mathbf{x}) &= \mathbb{E}[f(\mathbf{x})], \\ k(\mathbf{x}, \mathbf{x}'|\boldsymbol{\theta}) &= \mathbb{E}[(f(\mathbf{x}) - m_0(\mathbf{x}))(f(\mathbf{x}') - m_0(\mathbf{x}'))], \end{aligned} \tag{A.2}$$

where $\mathbb{E}[\cdot]$ is the expectation and $\boldsymbol{\theta}$ are the hyperparameters that control the kernel function. By centering the data, the mean function may be assumed to be an equal constant, $m_0(\mathbf{x}) \equiv m_0$. Alternative options are feasible, such as a linear function of $\mathbf{x}$, but they are rarely used until previous knowledge of the shape of the function is provided. The covariance function can take several forms, with the automated relevance determinant (ARD) kernel being the most popular.

$$k(\mathbf{x}, \mathbf{x}'|\boldsymbol{\theta}) = \theta_0 \exp\left(-(\mathbf{x} - \mathbf{x}')^T \mathrm{diag}(\theta_1^{-2}, \ldots, \theta_l^{-2})(\mathbf{x} - \mathbf{x}')\right). \tag{A.3}$$

From this point on, we eliminate the explicit notation of $k(x, x')$'s reliance on $\boldsymbol{\theta}$. In this instance, the hyperparameters $\theta_1, \ldots, \theta_l$ are referred to as length-scales. For constant parameter $\mathbf{x}$, $f(\mathbf{x})$ is its random variable. In contrast, a collection of values, $f(\mathbf{x}_i)$, $i = 1, \ldots, N$, is a partial realization of the GP. GP's realizations are functions of x that are deterministic. The primary characteristic of GPs is that the joint distribution of $g(x_i), i = 1, \ldots, N$ is multivariate Gaussian.

Assuming the model deficiency $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ is likewise Gaussian, we can derive the model likelihood using the prior (A.1) and available data.

$$\begin{aligned} \mathcal{L} \triangleq p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) &= \int (f(\mathbf{x}) + \varepsilon)df = \mathcal{N}(\mathbf{y}|m_0\mathbf{1}, \mathbf{K} + \sigma^2\mathbf{I}) \\ &= -\frac{1}{2}\left(\mathbf{y} - m_0\mathbf{1}\right)^T \left(\mathbf{K} + \sigma^2\mathbf{I}\right)^{-1}\left(\mathbf{y} - m_0\mathbf{1}\right) \\ &\quad - \frac{1}{2}\ln|\mathbf{K} + \sigma^2\mathbf{I}| - \frac{N}{2}\log(2\pi), \end{aligned} \tag{A.4}$$

where $\mathbf{K} = [K_{ij}]$ is the covariance matrix, in which $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j), i, j = 1, \ldots, N$. The hyperparameters $\boldsymbol{\theta}$ is often derived from point estimations using the maximum likelihood (MLE) of Eq. (A.4) w.r.t. $\boldsymbol{\theta}$. The joint distribution of $\mathbf{y}$ and $f(\mathbf{x})$ is also a joint Gaussian distribution with mean value $m_0\mathbf{1}$ and covariance matrix.

$$\mathbf{K}' = \left[\begin{array}{c|c} \mathbf{K} + \sigma^2\mathbf{I} & \mathbf{k}(\mathbf{x}) \\ \hline \mathbf{k}^T(\mathbf{x}) & k(\mathbf{x}, \mathbf{x}) + \sigma^2 \end{array}\right], \tag{A.5}$$

where $\mathbf{k}(\mathbf{x}) = (k(\mathbf{x}_1, \mathbf{x}), \ldots, k(\mathbf{x}_N, \mathbf{x}))^T$. Conditioning on $\mathbf{y}$, the conditional predictive distribution at $\mathbf{x}$ is obtained.

$$\begin{aligned} \hat{f}(\mathbf{x})|\mathbf{y} &\sim \mathcal{N}\left(\mu(\mathbf{x}), v(\mathbf{x}, \mathbf{x}')\right), \\ \mu(\mathbf{x}) &= m_0\mathbf{1} + \mathbf{k}(\mathbf{x})^T\left(\mathbf{K} + \sigma^2\mathbf{I}\right)^{-1}\left(\mathbf{y} - m_0\mathbf{1}\right), \\ v(\mathbf{x}) &= \sigma^2 + k(\mathbf{x}, \mathbf{x}) - \mathbf{k}^T(\mathbf{x})\left(\mathbf{K} + \sigma^2\mathbf{I}\right)^{-1}\mathbf{k}(\mathbf{x}). \end{aligned} \tag{A.6}$$

The expected value $\mathbb{E}[f(\mathbf{x})]$ is given by $\mu(\mathbf{x})$ and the predictive variance by $v(\mathbf{x})$. From Eq. (A.5) to Eq. (A.6) is crucial since the prediction posterior of this wake is based on a comparable block covariance matrix.

B From AR to LiFiDE

We first revisit the classic AR and reveal its connection to an ODE's forward Euler solution. Rewrite Eq. (1) as follows,

$$
\begin{aligned}
y(\mathbf{x}, t_T) &= \gamma y(\mathbf{x}, t_0) + v(\mathbf{x}, t_0) \\
y(\mathbf{x}, t_T) - y(\mathbf{x}, t_0) &= (\gamma - 1) y(\mathbf{x}, t_0) + v(\mathbf{x}, t_0). \\
y(\mathbf{x}, t_T) - y(\mathbf{x}, t_0) &= \alpha y(\mathbf{x}, t_0) + v(\mathbf{x}, t_0),
\end{aligned}
\tag{A.7}
$$

where $\alpha \equiv \gamma - 1$. We then divide both sizes by the constant $\Delta(t_T - t_0)$ indicating the fidelity difference and absorb the constant into $(\rho - 1)$ and $u(\mathbf{x}, t_0)$ and write

$$
\frac{y(\mathbf{x}, t_T) - y(\mathbf{x}, t_0)}{\Delta(t_T - t_0)} = \frac{\alpha}{\Delta(t_T - t_0)} y(\mathbf{x}, t_0) + \frac{1}{\Delta(t_T - t_0)} v(\mathbf{x}, t_0).
\tag{A.8}
$$

Since α and $v(\mathbf{x}, t_0)$ are values and function to be estimated, we can absorb the constant $\Delta(t_T - t_0)$ into α and $v(\mathbf{x}, t_0)$ and write

$$
\frac{y(\mathbf{x}, t_T) - y(\mathbf{x}, t_0)}{\Delta(t_T - t_0)} = \beta_0 y(\mathbf{x}, t_0) + u(\mathbf{x}, t_0).
\tag{A.9}
$$

where $\beta_0 \equiv \alpha / \Delta(t_T - t_0)$ and $u(\mathbf{x}, t_0) \equiv v(\mathbf{x}, t_0) / \Delta(t_T - t_0)$. We recognize that is a explicit solution of a different equation. If we take the limit of $\Delta(t_T - t_0) \to 0$, we have

$$
\frac{\mathrm{d}y(\mathbf{x}, t)}{\mathrm{d}t} = \beta(t) y(\mathbf{x}, t) + u(\mathbf{x}, t).
\tag{A.10}
$$

which is in Proposition 1. Notice that in this equation we turn the constant β into a function $\beta(t)$, which allows us to control differential level of information transfer depending on its fidelity level as a general formulation.

Taking a forward Euler solution to solve Eq. (A.10), we have

$$
\begin{aligned}
\frac{y(\mathbf{x}, t + \Delta t) - y(\mathbf{x}, t)}{\Delta t} + \beta(t) y(\mathbf{x}, t) &= u(\mathbf{x}, t) \\
y(\mathbf{x}, t + \Delta t) &= (1 + \Delta t \beta(t)) y(\mathbf{x}, t) + \Delta t u(\mathbf{x}, t).
\end{aligned}
\tag{A.11}
$$

We recognize this is the formulation is extractly same as the classic AR with $\rho = 1 + \Delta t \beta(t)$ with a residual GP $\Delta t u(\mathbf{x}, t)$.

C A General Solution to LiFiDE

We derive the general solution to the derived linear fidelity differential equation (LiFiDE)

$$
\frac{\mathrm{d}y(\mathbf{x}, t)}{\mathrm{d}t} + \beta(t) y(\mathbf{x}, t) = u(\mathbf{x}, t),
\tag{A.12}
$$

which is a standard non-homogeneous first order differential equation. We know that for the homogeneous equation, i.e., $u(\mathbf{x}, t) = 0$, the general solution is

$$
y(\mathbf{x}, t) = C(\mathbf{x}) e^{-B(t)},
\tag{A.13}
$$

where $C(\mathbf{x})$ is a functional of $\mathbf{x}$, and $B(t) = \int^t \beta(\tau) \mathrm{d}\tau$ is the antiderivative of $\beta(t)$. Thus, we assume the general solution of the non-homogeneous equation is of the form $y(\mathbf{x}, t) = v(\mathbf{x}, t) e^{-B(t)}$, where $v(\mathbf{x}, t)$ is a functional of $\mathbf{x}$ and t. Substituting this into Eq. (A.12), we have

$$
\begin{aligned}
\frac{\mathrm{d}v(\mathbf{x}, t)}{\mathrm{d}t} e^{B(t)} - v(\mathbf{x}, t) e^{-B(t)} \beta(t) + \beta(t) v(\mathbf{x}, t) e^{-B(t)} &= u(\mathbf{x}, t) \\
\frac{\mathrm{d}v(\mathbf{x}, t)}{\mathrm{d}t} &= u(\mathbf{x}, t) e^{-B(t)}.
\end{aligned}
\tag{A.14}
$$

Integrating both sides, we have

$$
v(\mathbf{x}, t) = \int^t u(\mathbf{x}, \tau) e^{-B(\tau)} \mathrm{d}\tau + C(\mathbf{x}).
\tag{A.15}
$$

Putting this back into the assumed solution, we have

$$
y(\mathbf{x}, t) = e^{-B(t)} \int^t u(\mathbf{x}, \tau) e^{-B(\tau)} \mathrm{d}\tau + C(\mathbf{x}) e^{-B(t)}.
\tag{A.16}
$$

479   If we assume a constant for $\beta(t)$, i.e., $\beta(t) = \beta$, then we have

$$
\begin{aligned}
y(\mathbf{x}, t) &= e^{-\beta t} \int^{t} u(\mathbf{x}, \tau) e^{\beta \tau} \mathrm{d}\tau + C(\mathbf{x}) e^{-\beta t} \\
&= C(\mathbf{x}) e^{-\beta t} + \int^{t} u(\mathbf{x}, \tau) e^{-\beta(t-\tau)} \mathrm{d}\tau.
\end{aligned}
\tag{A.17}
$$

480   For a practical from where the integral starts from t_0, the lowest-fidelity index, we have

$$
\begin{aligned}
y(\mathbf{x}, t_0) &= C(\mathbf{x}) e^{-\beta t_0} + \int_{0}^{t_0} u(\mathbf{x}, \tau) e^{-\beta(t-\tau)} \mathrm{d}\tau \\
&= C(\mathbf{x}) e^{-\beta t_0},
\end{aligned}
\tag{A.18}
$$

481   where we have assumed $u(\mathbf{x}, t) = 0$ for $t < t_0$ because we are not interested in $t < t_0$. Substituting this into
482   Eq. (A.17), we have

$$
\begin{aligned}
y(\mathbf{x}, t) &= C(\mathbf{x}) e^{-\beta t} + \int_{0}^{t} u(\mathbf{x}, \tau) e^{-\beta(t-\tau)} \mathrm{d}\tau \\
&= y(\mathbf{x}, t_0) e^{\beta t_0} e^{-\beta t} + \int_{0}^{t_0} u(\mathbf{x}, \tau) e^{-\beta(t-\tau)} \mathrm{d}\tau + \int_{t_0}^{t} u(\mathbf{x}, \tau) e^{-\beta(t-\tau)} \mathrm{d}\tau \\
&= y(\mathbf{x}, t_0) e^{-\beta(t-t_0)} + \int_{t_0}^{t} u(\mathbf{x}, \tau) e^{-\beta(t-\tau)} \mathrm{d}\tau.
\end{aligned}
\tag{A.19}
$$

483   This formulation is equivalent to Eq. (A.17). However, it allows for more flexibility in practice as we can give
484   start the model with the lowest-fidelity t_0 and then use the model to predict the higher-fidelity $t > t_0$.

D   General Solutions to LiFiDEs

486   Consider LiFiDEs taking this general form

$$
\frac{d\mathbf{y}(\mathbf{x}, t)}{dt} + \mathbf{B}(t)\mathbf{y}(\mathbf{x}, t) = \mathbf{S}\mathbf{u}(\mathbf{x}, t)
\tag{A.20}
$$

487   where $\mathbf{B}(t)$ and $\mathbf{S}$ are matrices, $\mathbf{y}(\mathbf{x}, t)$ is a vector of unknown functions, and $\mathbf{u}(\mathbf{x}, t)$ is a non-zero vector.

488   The general solution is the sum of the homogeneous solution $\mathbf{y}_h(\mathbf{x}, t)$ and a particular solution $\mathbf{y}_p(\mathbf{x}, t)$.

489   The homogeneous solution comes from the homogeneous equation:

$$
\frac{d\mathbf{y}_h(\mathbf{x}, t)}{dt} + \mathbf{B}(t)\mathbf{y}_h(\mathbf{x}, t) = 0
\tag{A.21}
$$

490   This has the solution:

$$
\mathbf{y}_h(\mathbf{x}, t) = e^{-\int \mathbf{B}(t) dt} \mathbf{c}
\tag{A.22}
$$

491   where $\mathbf{c}$ is a constant vector. The particular solution comes from the non-homogeneous equation and can be
492   obtained using the variation of parameters. The method involves finding a function $\mathbf{v}(\mathbf{x}, t)$ such that:

$$
\mathbf{y}_p(\mathbf{x}, t) = e^{-\int \mathbf{B}(t) dt} \mathbf{v}(\mathbf{x}, t)
\tag{A.23}
$$

493   is a solution to the non-homogeneous system. Substituting $\mathbf{y}_p$ into the non-homogeneous system gives:

$$
\frac{d\mathbf{v}(\mathbf{x}, t)}{dt} = e^{\int \mathbf{B}(t) dt} \mathbf{S}\mathbf{u}(\mathbf{x}, t)
\tag{A.24}
$$

494   which can be integrated to find $\mathbf{v}(\mathbf{x}, t)$. Thus, the particular solution is:

$$
\mathbf{y}_p(\mathbf{x}, t) = e^{-\int \mathbf{B}(t) dt} \int \left[e^{\int \mathbf{B}(t) dt} \mathbf{S}\mathbf{u}(\mathbf{x}, t) \right] dt
\tag{A.25}
$$

The general solution to the non-homogeneous system is the sum of the homogeneous and particular solutions:

$$\mathbf{y}(\mathbf{x},t) = \mathbf{y}_h(\mathbf{x},t) + \mathbf{y}_p(\mathbf{x},t)$$
$$= e^{-\int \mathbf{B}(t)dt}\mathbf{c} + e^{-\int \mathbf{B}(t)dt}\int \left[e^{\int \mathbf{B}(t)dt}\mathbf{Su}(\mathbf{x},t)\right]dt \tag{A.26}$$

which is similar to the solution to LiFiDE in Eq. (A.17). Note that the matrix exponential computation is done by

$$e^{B(t)} = I + B(t) + \frac{(B(t))^2}{2!} + \frac{(B(t))^3}{3!} + \frac{(B(t))^4}{4!} + \cdots \tag{A.27}$$

which is not easy to compute. However, if the matrix $\mathbf{B}(t)$ is a constant matrix $\beta\mathbf{I}$, then the matrix exponential can be computed by

$$e^{B(t)} = e^{\beta I} = Ie^{\beta} = \begin{bmatrix} e^{\beta} & 0 & \cdots & 0 \\ 0 & e^{\beta} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{\beta}, \end{bmatrix} \tag{A.28}$$

which gives us a efficient way to derive a solution to LiFiDEs,

$$\mathbf{y}(\mathbf{x},t) = e^{-\beta t}\mathbf{C} + e^{-\beta t}\int e^{\beta t}\mathbf{Su}(\mathbf{x},t)dt. \tag{A.29}$$

E Proof of Lemma 1

Lemma 1. Autokrigeability in ContinuAR: the particular values of the spatial correlation matrix $\mathbf{H}$ and $\mathbf{SS}^\top$ do not matter in the predictive mean as they will be canceled out.

Proof. The autokrigeability in ContinuAR is easy to derive once we derive the predictive posterior for the subset case in Appendix G and the non-subset case in Appendix H. We can see the the matrix $\mathbf{H}$ and $\mathbf{SS}^\top$ are canceled out in the predictive mean. □

F Joint Likelihood of FiDEs-1

We derive the joint likelihood of FiDEs for observations $\mathbf{Y} = [\vec{\mathbf{y}}^{(0)}; \vec{\mathbf{y}}^{(1)}]$ as an illustrating example. For clarity, we slightly abuse the notations by replacing the previous spatial correlation H in the main paper with $\mathbf{S}^{(0)}$ and $\mathbf{SS}^\top$ with $\mathbf{S}^{(1)}$. The joint probability for $\mathbf{Y}$ is

$$\begin{pmatrix} \vec{\mathbf{y}}^{(0)} \\ \vec{\mathbf{y}}^{(1)} \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{array}{cc} k^0(\mathbf{X}^{(0)}\mathbf{X}^{(0)}) \otimes \mathbf{S}^{(0)} & e^{-\beta(t_1-t_0)}k^0(\mathbf{X}^{(1)},\mathbf{X}^{(0)}) \otimes \mathbf{S}^{(0)} \\ e^{-\beta(t_1-t_0)}k^0(\mathbf{X}^{(0)},\mathbf{X}^{(1)}) \otimes \mathbf{S}^{(0)} & e^{-2\beta(t_1-t_0)}k^0(\mathbf{X}^{(1)},\mathbf{X}^{(1)}) \otimes \mathbf{S}^{(0)} + k^{u_1}(\mathbf{X}^{(1)},\mathbf{X}^{(1)}) \otimes \mathbf{S}^{(1)} \end{array}\right),$$
$$\tag{A.30}$$

where $k^{u_1}(\mathbf{X}^{(1)},\mathbf{X}^{(1)}) = \int_{t_0}^{t_1} e^{-B(t-\tau)} \int_{t_0}^{t_1} e^{-B(t'-\tau')}k^u(\mathbf{X}^{(1)},\tau,\mathbf{X}^{(1)},\tau')d\tau'd\tau$.

Eq. (A.30) reveals that the for a two fidelity problem, ContinuAR is equivalent to AR. Thus, the likelihood decomposition for subset structure data [10] also holds for ContinuAR, which lays out the foundation for our fast inference algorithm for non-subset problems.

G Predictive Posterior With Subset Structure

In order to conduct the proof of decomposition for the likelihood function with non-subset data, we need to derive the form of predictive posterior distribution with multi-fidelity with subset structure i.e., $\mathbf{X}^{(T)} \subset \cdots \subset \mathbf{X}^{(0)}$, first. Here, we take the simplest situation, i.e., $T = 1$ as an example, and the assumption, $\mathbf{X}^{(1)} \subset \mathbf{X}^{(0)}$ holds. To stay consistent with AR, we denote $\rho = e^{-\beta(t_1-t_0)}$ for clarity. Following the format of the covariance matrix

15

in Eq. (A.30) and Eq. (A.6), the mean function and covariance matrix have the following expression,

$$
\bar{\mathbf{y}}(x_*, 1)
$$

$$
= \left(\rho \mathbf{k}_*^{(0)} \otimes \mathbf{S}^{(0)}, \rho^2 \mathbf{k}_*^{(0)}(\mathbf{x}^{(1)}) \otimes \mathbf{S}^{(0)} + \mathbf{k}_{a*}^{(1)} \otimes \mathbf{S}^{(1)} \right) \boldsymbol{\Sigma}^{-1} \left(\begin{array}{c} \vec{\mathbf{y}}^{(0)} \\ \vec{\mathbf{y}}^{(1)} \end{array} \right)
$$

$$
= \left(\rho \mathbf{k}_*^{(0)} \otimes \mathbf{S}^{(0)} \right) \left(\mathbf{K}^{(00)} \otimes \mathbf{S}^{(0)} \right)^{-1} \vec{\mathbf{y}}^{(0)} + \left(\rho \mathbf{k}_*^{(0)} \otimes \mathbf{S}^{(0)} \right) \left(\mathbf{E}^{(1)}(\rho^2 \mathbf{K}^{(00)})^{-1} \left(\mathbf{E}^{(1)} \right)^\top \otimes (\mathbf{S}^{(1)})^{-1} \right) \vec{\mathbf{y}}^{(0)}
$$

$$
- \left(\rho^2 \mathbf{k}_*^{(0)}(\mathbf{x}^{(1)}) \otimes \mathbf{S}^{(0)} \right) \left(\rho \mathbf{K}_a^{(11)}(\mathbf{E}^{(1)})^\top \otimes \mathbf{S}^{(1)} \right)^{-1} \vec{\mathbf{y}}^{(0)}
$$

$$
- \left(\mathbf{k}_*^{(0)} \otimes \mathbf{S}^{(0)} \right) \left(\rho \mathbf{K}_a^{(11)}(\mathbf{E}^{(1)})^\top \otimes \mathbf{S}^{(1)} \right)^{-1} \vec{\mathbf{y}}^{(0)}
$$

$$
- \left(\rho \mathbf{k}_*^{(0)} \otimes \mathbf{S}^{(0)} \right) \left(\mathbf{E}^{(1)}(\rho \mathbf{K}_a^{(11)})^{-1} \otimes \mathbf{S}^{(1)} \right)^{-1} \mathbf{y}_a^{(1)}
$$

$$
+ \left(\rho^2 \mathbf{k}_*^{(0)}(\mathbf{x}^{(1)}) \otimes \mathbf{S} \right) \left(\mathbf{K}_a^{(11)} \otimes \mathbf{S} \right)^{-1} \mathbf{y}_a^{(1)} + \left(\mathbf{k}_{a*}^{(1)} \otimes \mathbf{S}^{(0)} \right) \left(\mathbf{K}_a^{(11)} \otimes \mathbf{S}^{(1)} \right)^{-1} \mathbf{y}_a^{(1)}
$$

$$
= \left(\rho \mathbf{k}_*^{(0)} \otimes \mathbf{S}^{(0)} \right) \left(\mathbf{K}^{(00)} \otimes \mathbf{S}^{(0)} \right)^{-1} \vec{\mathbf{y}}^{(0)} - \left(\mathbf{k}_*^{(0)} \otimes \mathbf{S}^{(0)} \right) \left(\rho \mathbf{K}_a^{(11)}(\mathbf{E}^{(1)})^\top \otimes (\mathbf{S}^{(1)})^{-1} \right) \vec{\mathbf{y}}^{(0)}
$$

$$
+ \left(\mathbf{k}_*^{(0)} \otimes \mathbf{S}^{(0)} \right) \left(\rho \mathbf{K}_a^{(11)}(\mathbf{E}^{(1)})^\top \otimes (\mathbf{S}^{(0)})^{-1} \right) \vec{\mathbf{y}}^{(0)} + \left(\mathbf{k}_{a*}^{(1)} \otimes \mathbf{S}^{(1)} \right) \left(\mathbf{K}_a^{(11)} \otimes \mathbf{S}^{(1)} \right)^{-1} \mathbf{y}_a^{(1)}
$$

$$
= \left(\rho \mathbf{k}_*^{(0)} \left(\mathbf{K}^{(00)} \right)^{-1} \otimes \mathbf{I} \right) \vec{\mathbf{y}}^{(0)} + \left(\mathbf{k}_{a*}^{(1)} \left(\mathbf{K}_a^{(11)} \right)^{-1} \otimes \mathbf{I} \right) \mathbf{y}_a^{(1)},
$$

$$(A.31)$$

$$
\boldsymbol{\Sigma}_*^{(1)}
$$

$$
= \left(\rho^2 \mathbf{k}_*^{(00)} \otimes \mathbf{S}^{(0)} + \mathbf{k}_{a*}^{(11)} \otimes \mathbf{S}^{(1)} \right)
$$

$$
- \left(\rho \mathbf{k}_*^{(0)} \otimes \mathbf{S}^{(0)}, \rho^2 \mathbf{k}_*^{(0)}(\mathbf{x}^{(1)}) \otimes \mathbf{S}^{(0)} + \mathbf{k}_{a*}^{(1)} \otimes \mathbf{S}^{(1)} \right) \boldsymbol{\Sigma}^{-1} \left(\begin{array}{c} \left(\rho \mathbf{k}_*^{(0)} \otimes \mathbf{S}^{(0)} \right)^\top \\ \rho^2 \left(\mathbf{k}_*^{(0)}(\mathbf{x}^{(1)}) \otimes \mathbf{S}^{(0)} \right)^\top + \left(\mathbf{k}_{a*}^{(1)} \otimes \mathbf{S}^{(1)} \right)^\top \end{array} \right)
$$

$$
= \left(\rho^2 \mathbf{k}_*^{(00)} \otimes \mathbf{S}^{(0)} + \mathbf{k}_{a*}^{(11)} \otimes \mathbf{S}^{(1)} \right) - \left(\rho \mathbf{k}_*^{(0)} \otimes \mathbf{S}^{(0)} \right) \left(\mathbf{K}^{(00)} \otimes \mathbf{S}^{(0)} \right)^{-1} \left(\rho \mathbf{k}_*^{(0)} \otimes \mathbf{S}^{(0)} \right)^\top
$$

$$
+ \left(\mathbf{k}_{a*}^{(1)} \otimes \mathbf{S}^{(1)} \right) \left(\rho \mathbf{K}_a^{(11)} \otimes \mathbf{S}^{(1)} \right) \left(\rho \mathbf{k}_*^{(0)} \otimes \mathbf{S}^{(0)} \right)^\top - \left(\mathbf{k}_{a*}^{(1)} \otimes \mathbf{S}^{(1)} \right) \left(\rho \mathbf{K}_a^{(11)} \otimes \mathbf{S}^{(1)} \right) \left(\rho \mathbf{k}_*^{(0)} \otimes \mathbf{S}^{(0)} \right)^\top
$$

$$
- \left(\mathbf{k}_{a*}^{(1)} \otimes \mathbf{S}^{(1)} \right) \left(\mathbf{K}_a^{(11)} \otimes \mathbf{S}^{(1)} \right)^{-1} \left(\mathbf{k}_{a*}^{(1)} \otimes \mathbf{S}^{(1)} \right)^\top
$$

$$
= \rho^2 \left(\mathbf{k}_*^{(00)} - \left(\mathbf{k}_*^{(0)} \right)^\top \left(\mathbf{K}^{(00)} \right)^{-1} \mathbf{k}_*^{(0)} \right) \otimes \mathbf{S}^{(0)} + \left(\mathbf{k}_{a*}^{(11)} - \left(\mathbf{k}_{a*}^{(1)} \right)^\top \left(\mathbf{K}_a^{(11)} \right)^{-1} \mathbf{k}_{a*}^{(1)} \right) \otimes \mathbf{S}^{(1)},
$$

$$(A.32)$$

where $\mathbf{k}_*^{(0)} = k^0(\mathbf{X}_*, \mathbf{X}^{(0)})$, $\mathbf{k}_*^{(00)} = k^0(\mathbf{X}_*, \mathbf{X}_*)$, $\mathbf{K}^{(00)} = k^0(\mathbf{X}^{(0)}, \mathbf{X}^{(0)})$ denotes the covariance in the lowest fidelity, and $\mathbf{k}_{a*}^{(1)} = k^{u_1}(\mathbf{X}_*, \mathbf{X}^{(1)})$, $\mathbf{k}_*^{(11)} = k^{u_1}(\mathbf{X}_*, \mathbf{X}_*)$ and $\mathbf{K}_a^{(11)} = k^{u_1}(\mathbf{X}^{(1)}, \mathbf{X}^{(1)})$.

Notice that Eq. (A.32) decomposes the predictive posterior into two parts that are related to the low- and high-fidelity data, respectively. This lays the foundation for our later derivation for the non-subset structure. In the following section, we will derive the predictive posterior for non-subset structure, and show that the autokrigeability also holds.

H Predictive Posterior With Non-Subset Structure

In this section, we derive the mean function and covariance matrix in the predictive posterior for non-subset structure following the previous two-fidelity setup.

$$
p(\bar{\mathbf{Y}}(\mathbf{x}_*, 2) | \mathbf{Y}^{(2)}, \mathbf{Y}^{(1)}) = \int p(\bar{\mathbf{Y}}(\mathbf{x}_*, 2), \hat{\mathbf{Y}}^{(1)} | \mathbf{Y}^{(2)}, \mathbf{Y}^{(1)}) d\hat{\mathbf{Y}}^{(1)}
$$

$$
= \int \underbrace{p(\bar{\mathbf{Y}}(\mathbf{x}_*, 2) | \mathbf{Y}^{(2)}, \mathbf{Y}^{(1)}, \hat{\mathbf{Y}}^{(1)})}_{\text{fidelity 2 posterior}} \underbrace{p(\hat{\mathbf{Y}}^{(1)})}_{\text{fidelity 1 posterior}} d\hat{\mathbf{Y}}^{(1)}.
$$

$$(A.33)$$

16

As we know, once the $\hat{\mathbf{Y}}^{(1)}$ is decided, the high-fidelity posterior part can be written as the subset posterior distribution, in the following way,

$$p(\bar{\mathbf{Y}}(\mathbf{x}_*, 2)|\mathbf{Y}^{(2)}, \mathbf{Y}^{(1)}, \hat{\mathbf{Y}}^{(1)}) = 2\pi^{-\frac{N_p D}{2}} \times \left|\mathbf{\Sigma}_*^{(2)}\right|^{-\frac{1}{2}} \times \exp - \frac{1}{2}\left[\left(\mathbf{y}_*^{(2)} - \bar{\mathbf{y}}^{(2)}\right)^\top \left(\mathbf{\Sigma}_*^{(2)}\right)^{-1} \left(\mathbf{y}_*^{(2)} - \bar{\mathbf{y}}^{(2)}\right)\right].$$
(A.34)

where

$$\mathbf{\Sigma}_*^{(2)} = \mathbf{\Sigma}(\mathbf{x}_*, 1) + \tilde{\mathbf{\Sigma}}^{(2)},$$

$$\bar{\mathbf{y}}^{(2)} = \left[(\mathbf{k}_*^{(11)})\left(\mathbf{K}^{(11)}\right)^{-1} \otimes \mathbf{I}\right]\begin{pmatrix} \mathbf{Y}^{(1)} \\ \hat{\mathbf{y}}^{(1)} \end{pmatrix} + \left[(\mathbf{k}_{a*}^{(22)})\left(\mathbf{K}_a^{(22)}\right)^{-1} \otimes \mathbf{I}\right]\mathbf{y}^{(2)}.$$
(A.35)

This is depends on the the conclusion in Eq. (A.31), $\mathbf{\Sigma}_*^{(2)}$ and $\bar{\mathbf{y}}^{(2)}$ are posterior covariance matrix and mean function.

At the same time, we know that $\hat{\mathbf{Y}}^{(1)}$ are samples from fidelity 1 model, therefore the probabilityof getting $\hat{\mathbf{Y}}^{(1)}$ can be written as,

$$p(\hat{\mathbf{Y}}^{(1)}) = 2\pi^{-\frac{N_n^{(2)} D}{2}} \times \left|\mathbf{\Sigma}_*^{(1)} \otimes \mathbf{S}^{(1)}\right|^{-\frac{1}{2}} \times \exp\left[-\frac{1}{2}\left(\hat{\mathbf{y}}^{(1)} - \bar{\mathbf{y}}^{(1)}\right)^\top \left(\mathbf{\Sigma}_*^{(1)} \otimes \mathbf{S}^{(1)}\right)^{-1} \left(\hat{\mathbf{y}}^{(1)} - \bar{\mathbf{y}}^{(1)}\right)\right],$$
(A.36)

where the $\mathbf{\Sigma}_*^{(1)} \otimes \mathbf{S}$ means the posterior covariance matrix of sampling $\hat{\mathbf{Y}}^{(1)}$ from lower fidelity model and $\bar{\mathbf{y}}^{(1)}$ are predicted mean for the non-subset data. Therefore the posterior distribution of non-subset data structure is

$$p(\bar{\mathbf{Y}}(\mathbf{x}_*, 2)|\mathbf{Y}^{(2)}, \mathbf{Y}^{(1)})$$

$$= \int \underbrace{p(\bar{\mathbf{Y}}(\mathbf{x}_*, 2)|\mathbf{Y}^{(2)}, \vec{\mathbf{y}}^{(1)}, \hat{\mathbf{Y}}^{(1)})}_{\text{fidelity 2 posterior}} \underbrace{p(\hat{\mathbf{Y}}^{(1)})}_{\text{fidelity 1 posterior}} d\hat{\mathbf{Y}}^{(1)}$$

$$= \int \left\{ 2\pi^{-\frac{N_p D}{2}} \times \left|\mathbf{\Sigma}_*^{(2)}\right|^{-\frac{1}{2}} \times \exp - \frac{1}{2}\left[\left(\mathbf{y}_*^{(2)} - \bar{\mathbf{y}}^{(2)}\right)^\top \left(\mathbf{\Sigma}_*^{(2)}\right)^{-1} \left(\mathbf{y}_*^{(2)} - \bar{\mathbf{y}}^{(2)}\right)\right] \right.$$

$$\left. \times 2\pi^{-\frac{N_n^{(2)} D}{2}} \times \left|\mathbf{\Sigma}_*^{(1)} \otimes \mathbf{S}^{(1)}\right|^{-\frac{1}{2}} \times \exp\left[-\frac{1}{2}\left(\hat{\mathbf{y}}^{(1)} - \bar{\mathbf{y}}^{(1)}\right)^\top \left(\mathbf{\Sigma}_*^{(1)} \otimes \mathbf{S}^{(1)}\right)^{-1} \left(\hat{\mathbf{y}}^{(1)} - \bar{\mathbf{y}}^{(1)}\right)\right] \right\} d\hat{\mathbf{Y}}^{(1)}$$

$$= 2\pi^{-\frac{(N_p + N_n^{(2)}) D}{2}} \times \left|\mathbf{\Sigma}_*^{(2)}\right|^{-\frac{1}{2}} \times \left|\mathbf{\Sigma}_*^{(1)} \otimes \mathbf{S}^{(1)}\right|^{-\frac{1}{2}} \times \exp\left[-\frac{1}{2}\tilde{\mathbf{y}}^\top \left(\mathbf{\Sigma}_*^{(2)}\right)^{-1} \tilde{\mathbf{y}} - \frac{1}{2}\left(\bar{\mathbf{y}}^{(1)}\right)^\top \left(\mathbf{\Sigma}_*^{(1)} \otimes \mathbf{S}^{(1)}\right)^{-1} \bar{\mathbf{y}}^{(1)}\right]$$

$$\times \int \exp\left[\tilde{\mathbf{y}}^\top \left(\mathbf{\Sigma}_*^{(2)}\right)^{-1} \mathbf{\Gamma}\hat{\mathbf{y}}^{(1)} + \left(\bar{\mathbf{y}}^{(1)}\right)^\top \left(\mathbf{\Sigma}_*^{(1)} \otimes \mathbf{S}\right)^{-1} \mathbf{\Gamma}\hat{\mathbf{y}}^{(1)} - \frac{1}{2}\left(\hat{\mathbf{y}}^{(1)}\right)^\top \mathbf{\Gamma}^\top \left(\mathbf{\Sigma}_*^{(2)}\right)^{-1} \hat{\mathbf{y}}^{(1)} \right.$$

$$\left. - \frac{1}{2}\left(\hat{\mathbf{y}}^{(1)}\right)^\top \left(\mathbf{\Sigma}_*^{(1)} \otimes \mathbf{S}^{(1)}\right)^{-1} \hat{\mathbf{y}}^{(1)}\right] d\hat{\mathbf{Y}}^{(1)}$$

$$= 2\pi^{-\frac{N_p D}{2}} \times \left|\mathbf{\Sigma}_*^{(2)}\right|^{-\frac{1}{2}} \times \left|\mathbf{\Sigma}_*^{(1)} \otimes \mathbf{S}^{(1)}\right|^{-\frac{1}{2}} \times \left|\mathbf{\Gamma}^\top \left(\mathbf{\Sigma}_*^{(2)}\right)^{-1} \mathbf{\Gamma} + \mathbf{\Sigma}_*^{(1)}\right|^{-\frac{1}{2}}$$

$$\times \exp\left\{-\frac{1}{2}\tilde{\mathbf{y}}^\top \left[\left(\mathbf{\Sigma}_*^{(2)}\right)^{-1} - \left(\mathbf{\Sigma}_*^{(2)}\right)^{-1} \mathbf{\Gamma} \left(\mathbf{\Gamma}^\top \left(\mathbf{\Sigma}_*^{(2)}\right)^{-1} \mathbf{\Gamma} + \mathbf{\Sigma}_*^{(1)}\right)^{-1} \mathbf{\Gamma}^\top \left(\mathbf{\Sigma}_*^{(2)}\right)^{-1}\right]\tilde{\mathbf{y}}\right\}$$

$$\times \exp\left\{-\frac{1}{2}\left(\bar{\mathbf{y}}^{(1)}\right)^\top \left[\left(\mathbf{\Sigma}_*^{(1)} \otimes \mathbf{S}^{(1)}\right)^{-1} - \left(\mathbf{\Sigma}_*^{(2)}\right)^{-1} \left(\mathbf{\Gamma}^\top \left(\mathbf{\Sigma}_*^{(2)}\right)^{-1} \mathbf{\Gamma} + \mathbf{\Sigma}_*^{(1)}\right)^{-1} \left(\mathbf{\Sigma}_*^{(1)} \otimes \mathbf{S}^{(1)}\right)^{-1}\right]\left(\bar{\mathbf{y}}^{(1)}\right)\right\}$$

$$\times \exp\left\{\tilde{\mathbf{y}}^\top \left[\left(\mathbf{\Sigma}_*^{(2)}\right)^{-1} \mathbf{\Gamma} \left(\mathbf{\Gamma}^\top \left(\mathbf{\Sigma}_*^{(2)}\right)^{-1} \mathbf{\Gamma} + \mathbf{\Sigma}_*^{(1)}\right)^{-1} \left(\mathbf{\Sigma}_*^{(1)} \otimes \mathbf{S}^{(1)}\right)^{-1}\right]\left(\bar{\mathbf{y}}^{(1)}\right)\right\}.$$
(A.37)

where $\tilde{\mathbf{y}}$ and $\mathbf{\Gamma}$ is defined by the following equations,

$$\tilde{\mathbf{y}} = \mathbf{y}_*^{(2)} - \rho\left[(\mathbf{k}_{a*}^{(11)})\left(\mathbf{K}_a^{(11)}\right)^{-1} \otimes \mathbf{I}\right]\begin{pmatrix} \vec{\mathbf{y}}^{(1)} \\ \mathbf{0} \end{pmatrix} - \left[(\mathbf{k}_{a*}^{(22)})\left(\mathbf{K}_a^{(22)}\right)^{-1} \otimes \mathbf{I}\right]\left[\mathbf{y}^{(2)} - \begin{pmatrix} \check{\mathbf{y}}^{(1)} \\ \mathbf{0} \end{pmatrix}\right],$$

$$\mathbf{\Gamma} = \left[\mathbf{k}_{a*}^{(2,2)}\left(\mathbf{K}_a^{(2,2)}\right)^{-1} \left(\mathbf{E}_n^{(2)}\right)^\top \tilde{\mathbf{k}}_{a*}^{(1,1)} \left(\tilde{\mathbf{K}}_a^{(1,1)}\right)^{-1}\right]\left(\mathbf{E}_m^{(2)}\right).$$
(A.38)

where $\mathbf{E}_m^{(2)}$ denotes the selection matrix which select the non-subset parts between two fidelities, $\hat{\mathbf{X}}^{(1)} = \hat{\mathbf{X}}^{(1)} = \left(\mathbf{E}_m^{(2)}\right)^\top \left[\mathbf{X}^{(1)}, \hat{\mathbf{X}}^{(1)}\right]$ and $\mathbf{E}_n^{(2)}$ also denotes the selectio matrix but which selects all inputs for the second fidelity, $\mathbf{X}^{(2)} = \left(\mathbf{E}_n^{(2)}\right)^\top \left[\mathbf{X}^{(1)}, \hat{\mathbf{X}}^{(1)}\right]$.

After that, we can simplify the determenant and the exponential parts by decomposing them into different parts and using the Sherman-Morrison formula to obtain conclusions.

For the determinant part, we can derive

$$
\left|\mathbf{\Sigma}_*^{(2)}\right|^{-\frac{1}{2}} \times \left|\mathbf{\Sigma}_*^{(1)} \otimes \mathbf{S}^{(1)}\right|^{-\frac{1}{2}} \times \left|\mathbf{\Gamma}^\top \left(\mathbf{\Sigma}_*^{(2)}\right)^{-1} \mathbf{\Gamma} + \mathbf{\Sigma}_*^{(1)}\right|^{-\frac{1}{2}}
$$
$$
= \left|\mathbf{\Sigma}_*^{(2)}\right|^{-\frac{1}{2}} \times \left|\mathbf{\Sigma}_*^{(1)} \otimes \mathbf{S}^{(1)}\right|^{-\frac{1}{2}} \times \left|\left(\mathbf{\Sigma}_*^{(2)}\right)^{-1}\right|^{-\frac{1}{2}} \times \left|\left(\mathbf{\Sigma}_*^{(1)} \otimes \mathbf{S}^{(1)}\right)^{-1}\right|^{-\frac{1}{2}} \times \left|\mathbf{\Sigma}_*^{(2)} + \mathbf{\Gamma}^\top \left(\mathbf{\Sigma}_*^{(1)} \otimes \mathbf{S}^{(1)}\right)^{-1} \mathbf{\Gamma}\right|^{-\frac{1}{2}}
$$
$$
= \left|\mathbf{\Sigma}_*^{(2)} + \mathbf{\Gamma}^\top \left(\mathbf{\Sigma}_*^{(1)} \otimes \mathbf{S}^{(1)}\right)^{-1} \mathbf{\Gamma}\right|^{-\frac{1}{2}}.
$$

(A.39)

For the exponential parts in Eq. (A.37),

$$
\left(\mathbf{\Sigma}_*^{(2)}\right)^{-1} - \left(\mathbf{\Sigma}_*^{(2)}\right)^{-1} \mathbf{\Gamma} \left(\mathbf{\Gamma}^\top \left(\mathbf{\Sigma}_*^{(2)}\right)^{-1} \mathbf{\Gamma} + \mathbf{\Sigma}_*^{(1)}\right)^{-1} \mathbf{\Gamma}^\top \left(\mathbf{\Sigma}_*^{(2)}\right)^{-1} = \left(\mathbf{\Sigma}_*^{(2)} + \mathbf{\Gamma}^\top \left(\mathbf{\Sigma}_*^{(1)} \otimes \mathbf{S}^{(1)}\right)^{-1} \mathbf{\Gamma}\right)^{-1}
$$
$$
\left(\mathbf{\Sigma}_*^{(1)} \otimes \mathbf{S}^{(1)}\right)^{-1} - \left(\mathbf{\Sigma}_*^{(2)}\right)^{-1} \left(\mathbf{\Gamma}^\top \left(\mathbf{\Sigma}_*^{(2)}\right)^{-1} \mathbf{\Gamma} + \mathbf{\Sigma}_*^{(1)}\right)^{-1} \left(\mathbf{\Sigma}_*^{(1)} \otimes \mathbf{S}^{(1)}\right)^{-1} = \mathbf{\Gamma}^\top \left(\mathbf{\Sigma}_*^{(2)} + \mathbf{\Gamma}^\top \left(\mathbf{\Sigma}_*^{(1)} \otimes \mathbf{S}^{(1)}\right)^{-1} \mathbf{\Gamma}\right)^{-1} \mathbf{\Gamma}
$$
$$
\left(\mathbf{\Sigma}_*^{(2)}\right)^{-1} \mathbf{\Gamma} \left(\mathbf{\Gamma}^\top \left(\mathbf{\Sigma}_*^{(2)}\right)^{-1} \mathbf{\Gamma} + \mathbf{\Sigma}_*^{(1)}\right)^{-1} \left(\mathbf{\Sigma}_*^{(1)} \otimes \mathbf{S}^{(1)}\right)^{-1} = -\left(\mathbf{\Sigma}_*^{(2)} + \mathbf{\Gamma}^\top \left(\mathbf{\Sigma}_*^{(1)} \otimes \mathbf{S}^{(1)}\right)^{-1} \mathbf{\Gamma}\right)^{-1} \mathbf{\Gamma}.
$$

(A.40)

Therefore, the likelihood of the posterior distribution is

$$
p(\bar{\mathbf{Y}}(\mathbf{x}_*, 2)|\mathbf{Y}^{(2)}, \mathbf{Y}^{(1)})
$$
$$
= 2\pi^{-\frac{N_p D}{2}} \times \left|\mathbf{\Sigma}_*^{(2)} + \mathbf{\Gamma}^\top \left(\mathbf{\Sigma}_*^{(1)} \otimes \mathbf{S}^{(1)}\right)^{-1} \mathbf{\Gamma}\right|^{-\frac{1}{2}} \times \exp\left[-\frac{1}{2} \left(\mathbf{y}_*^{(2)} - \bar{\mathbf{y}}_*^{(2)}\right)^\top \left(\mathbf{\Sigma}_*^{(2)} + \mathbf{\Gamma}^\top \left(\mathbf{\Sigma}_*^{(1)} \otimes \mathbf{S}^{(1)}\right)^{-1} \mathbf{\Gamma}\right)^{-1} \left(\mathbf{y}_*^{(2)} - \bar{\mathbf{y}}_*^{(2)}\right)\right].
$$

(A.41)

From the upper formula, for the non-subset data structure, the posetrior mean and covairance matrix are,

$$
\bar{\mathbf{y}}_*^{(2)} = \left[(\mathbf{k}_{a*}^{(11)}) \left(\mathbf{K}_a^{(11)}\right)^{-1} \otimes \mathbf{I}\right] \left(\begin{array}{c} \vec{\mathbf{y}}^{(1)} \\ \bar{\mathbf{y}}^{(1)} \end{array}\right) + \left[(\mathbf{k}_{a*}^{(22)}) \left(\mathbf{K}_a^{(22)}\right)^{-1} \otimes \mathbf{I}\right] \mathbf{y}^{(2)},
$$
$$
\mathbf{\Sigma}(\mathbf{x}_*, 2) = \mathbf{\Sigma}_*^{(2)} + \mathbf{\Gamma}^\top \left(\mathbf{\Sigma}_*^{(1)} \otimes \mathbf{S}^{(1)}\right)^{-1} \mathbf{\Gamma}.
$$

(A.42)

In Eq. (A.42), we also prove matrix $\mathbf{S}$ does not take part in posterior mean function which means that autokrige-ability again stands also in the non-subset data structure. By recursively applying this conclusion, we can easily extend it to a multi-fidelity problem; we show the detailed application in the main paper.

I Decomposition of Joint Likelihood With Non-Subset Structure

As we have shown that the autokrigeability holds in subset and non-subset data structure, we thus assume an identical spacial correlation, $\mathbf{S} = \mathbf{I}$, for easy computation acceleration (with the cost of losing the accuracy in the predictive variance). Due to the identical spatial correlations, we no longer need vectorization. First, we decompose the joint likelihood $\mathcal{L}$ into several independent parts,

$$
\mathcal{L} = \log p(\mathbf{Y}^{(0:2)}) = \log p(\mathbf{Y}^{(0)}, \mathbf{Y}^{(1)}, \mathbf{Y}^{(2)})
$$
$$
= \log p(\mathbf{Y}^{(0)}) + \log p(\mathbf{Y}^{(1)}|\mathbf{Y}^{(0)}) + \log p(\mathbf{Y}^{(2)}|\mathbf{Y}^{(1)}, \mathbf{Y}^{(0)})
$$
$$
= \log p(\mathbf{Y}^{(0)}) + \log p(\mathbf{Y}^{(1)}|\mathbf{Y}^{(0)}) + \log \int \underbrace{p(\mathbf{Y}^{(2)}|\mathbf{Y}^{(1)}, \hat{\mathbf{Y}}^{(1)})}_{\text{part 1}} \underbrace{p(\hat{\mathbf{Y}}^{(1)}|\mathbf{Y}^{(1)})}_{\text{part 2}} d\hat{\mathbf{Y}}^{(1)}.
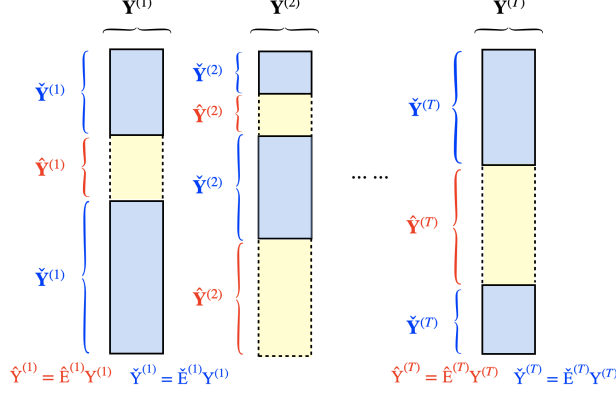$$

(A.43)

18

Figure 6: Index notation explanation.

Once the $\hat{\mathbf{Y}}^{(1)}$ is fixed, the probability of part 1 in Eq. (A.43) can be written as,

$$p(\mathbf{Y}^{(2)}|\mathbf{Y}^{(1)}, \hat{\mathbf{Y}}^{(1)}) = 2\pi^{-\frac{N^{(2)}D}{2}} \times \left|\mathbf{K}_a^{(22)}\right|^{-\frac{1}{2}} \times \exp\left[-\frac{1}{2}(\mathbf{Y}_a^{(2)})^\top \left(\mathbf{K}_a^{(22)}\right)^{-1} \mathbf{Y}_a^{(2)}\right]$$

$$=2\pi^{-\frac{N^{(2)}D}{2}} \times \left|\mathbf{K}_a^{(22)}\right|^{-\frac{1}{2}} \times \exp\left[-\frac{1}{2}\left[\left(\begin{array}{c}\check{\mathbf{Y}}^{(2)} \\ \hat{\mathbf{Y}}^{(2)}\end{array}\right) - \left(\begin{array}{c}\check{\mathbf{Y}}^{(1)} \\ \bar{\mathbf{Y}}^{(1)}\end{array}\right)\right]^\top \left(\mathbf{K}_a^{(22)}\right)^{-1} \left[\left(\begin{array}{c}\check{\mathbf{Y}}^{(2)} \\ \hat{\mathbf{Y}}^{(2)}\end{array}\right) - \left(\begin{array}{c}\check{\mathbf{Y}}^{(1)} \\ \bar{\mathbf{Y}}^{(1)}\end{array}\right)\right]\right].$$

(A.44)

Then is the part 2 in Eq. (A.43), it is based on the posterior distribution of lower fidelity, zu

$$p(\hat{\mathbf{Y}}^{(1)}|\mathbf{Y}^{(1)}) = 2\pi^{-\frac{N_n^{(2)}D}{2}} \times \left|\hat{\boldsymbol{\Sigma}}^{(1)}\right|^{-\frac{1}{2}} \times \exp\left[-\frac{1}{2}\left(\hat{\mathbf{Y}}^{(1)} - \bar{\mathbf{Y}}^{(1)}\right)^\top \left(\hat{\boldsymbol{\Sigma}}^{(1)}\right)^{-1} \left(\hat{\mathbf{Y}}^{(1)} - \bar{\mathbf{Y}}^{(1)}\right)\right].$$

(A.45)

where we use the $N_n^{(2)}$ to denotes the missing point in second fidelity corresponding with fidelity 2, which means the $\hat{\mathbf{Y}}^{(2)}$ parts, yellow part, in Fig. 6.

Therefore, we can combine the Eq. (A.44) with the Eq. (A.45), the integral part in Eq. (A.43) can be written as,

$$\log \int p(\mathbf{Y}^{(2)}|\mathbf{Y}^{(1)}, \hat{\mathbf{Y}}^{(1)})p(\hat{\mathbf{Y}}^{(1)}|\mathbf{Y}^{(1)})d\hat{\mathbf{Y}}^{(0)}$$

$$=\log \int \left\{2\pi^{-\frac{N^{(1)}D}{2}} \times \left|\mathbf{K}_a^{(11)}\right|^{-\frac{1}{2}} \times \exp\left[-\frac{1}{2}\left(\left(\begin{array}{c}\check{\mathbf{Y}}^{(2)} \\ \hat{\mathbf{Y}}^{(2)}\end{array}\right) - \left(\begin{array}{c}\check{\mathbf{Y}}^{(1)} \\ \bar{\mathbf{Y}}^{(1)}\end{array}\right)\right)^\top \left(\mathbf{K}_a^{(22)}\right)^{-1} \left(\left(\begin{array}{c}\check{\mathbf{Y}}^{(2)} \\ \hat{\mathbf{Y}}^{(2)}\end{array}\right) - \left(\begin{array}{c}\check{\mathbf{Y}}^{(1)} \\ \bar{\mathbf{Y}}^{(1)}\end{array}\right)\right)\right]\right.$$

$$\left. \times 2\pi^{-\frac{N_n^{(2)}D}{2}} \times \left|\hat{\boldsymbol{\Sigma}}^{(1)}\right|^{-\frac{1}{2}} \times \exp -\frac{1}{2}\left[\left(\hat{\mathbf{Y}}^{(1)} - \bar{\mathbf{Y}}^{(1)}\right)^\top \left(\hat{\boldsymbol{\Sigma}}^{(1)}\right)^{-1} \left(\hat{\mathbf{Y}}^{(0)} - \bar{\mathbf{Y}}^{(0)}\right)\right]\right\} d\hat{\mathbf{Y}}^{(1)}$$

$$=-\frac{(N^{(2)} + N_n^{(2)})D}{2}\log(2\pi) - \frac{1}{2}\log\left|\mathbf{K}_a^{(11)}\right| - \frac{1}{2}\log\left|\hat{\boldsymbol{\Sigma}}^{(1)}\right|$$

$$+\log \int \left\{\exp\left[-\frac{1}{2}\left(\left(\begin{array}{c}\check{\mathbf{Y}}^{(2)} \\ \hat{\mathbf{Y}}^{(2)}\end{array}\right) - \left(\begin{array}{c}\check{\mathbf{Y}}^{(1)} \\ \bar{\mathbf{Y}}^{(1)}\end{array}\right)\right)^\top \left(\mathbf{K}_a^{(22)}\right)^{-1} \left(\left(\begin{array}{c}\check{\mathbf{Y}}^{(2)} \\ \hat{\mathbf{Y}}^{(2)}\end{array}\right) - \left(\begin{array}{c}\check{\mathbf{Y}}^{(1)} \\ \bar{\mathbf{Y}}^{(1)}\end{array}\right)\right)\right]\right.$$

$$\left. -\frac{1}{2}\left[\left(\hat{\mathbf{Y}}^{(1)} - \bar{\mathbf{Y}}^{(1)}\right)^\top \left(\hat{\boldsymbol{\Sigma}}^{(1)}\right)^{-1} \left(\hat{\mathbf{Y}}^{(1)} - \bar{\mathbf{Y}}^{(1)}\right)\right]\right\} d\hat{\mathbf{Y}}^{(1)}$$

(A.46)

After that, we try to decompose the vector $\mathbf{Y}$ as,

$$\left(\begin{array}{c}\check{\mathbf{Y}}^{(2)} \\ \hat{\mathbf{Y}}^{(2)}\end{array}\right) - \left(\begin{array}{c}\check{\mathbf{Y}}^{(1)} \\ \bar{\mathbf{Y}}^{(1)}\end{array}\right) = \mathbf{E}^{(2)}\check{\mathbf{Y}}^{(2)} - \mathbf{E}^{(2)}\check{\mathbf{Y}}^{(1)} + \hat{\mathbf{E}}^{(2)}\hat{\mathbf{Y}}^{(2)} - \hat{\mathbf{E}}_n^{(2)}\hat{\mathbf{Y}}^{(1)}.$$

(A.47)

The $\mathbf{Y}$ is divided into subset part $\check{\mathbf{Y}}$ and non-subset part $\hat{\mathbf{Y}}$. Then, we consider the data fitting part by substituting Eq. (A.47) into Eq. (A.46), which gives the method to make the Eq. (A.46) calculable,

$$
\log \int p(\mathbf{Y}^{(2)}|\mathbf{Y}^{(1)}, \hat{\mathbf{Y}}^{(1)}) p(\hat{\mathbf{Y}}^{(1)}|\mathbf{Y}^{(1)}) d\hat{\mathbf{Y}}^{(0)}
$$

$$
= -\frac{(N^{(2)} + N_n^{(2)})D}{2} \log(2\pi) - \frac{1}{2} \log \left|\mathbf{K}_a^{(11)}\right| - \frac{1}{2} \log \left|\hat{\mathbf{\Sigma}}^{(1)}\right|
$$
$$
+ \log \int \exp \left\{ -\frac{1}{2} \left(\mathbf{E}^{(2)}\check{\mathbf{Y}}^{(2)} - \mathbf{E}^{(2)}\check{\mathbf{Y}}^{(1)}\right)^{\top} \left(\mathbf{K}_a^{(22)}\right)^{-1} \left(\mathbf{E}^{(2)}\check{\mathbf{Y}}^{(2)} - \mathbf{E}^{(2)}\check{\mathbf{Y}}^{(1)}\right) \right.
$$
$$
- \left(\mathbf{E}^{(2)}\check{\mathbf{Y}}^{(2)} - \mathbf{E}^{(2)}\check{\mathbf{Y}}^{(1)}\right)^{\top} \left(\mathbf{K}_a^{(22)}\right)^{-1} \left(\hat{\mathbf{E}}^{(2)}\hat{\mathbf{Y}}^{(2)} - \hat{\mathbf{E}}^{(2)}\hat{\mathbf{Y}}^{(1)}\right)
$$
$$
- \frac{1}{2} \left(\hat{\mathbf{E}}^{(2)}\hat{\mathbf{Y}}^{(2)} - \hat{\mathbf{E}}^{(2)}\hat{\mathbf{Y}}^{(1)}\right)^{\top} \left(\mathbf{K}_a^{(22)}\right)^{-1} \left(\hat{\mathbf{E}}^{(2)}\hat{\mathbf{Y}}^{(2)} - \hat{\mathbf{E}}^{(2)}\hat{\mathbf{Y}}^{(1)}\right)
$$
$$
\left. - \frac{1}{2} \left(\hat{\mathbf{Y}}^{(1)}\right)^{\top} \left(\hat{\mathbf{\Sigma}}^{(1)}\right)^{-1} \left(\hat{\mathbf{Y}}^{(1)}\right) + \left(\bar{\mathbf{Y}}^{(1)}\right)^{\top} \left(\hat{\mathbf{\Sigma}}^{(1)}\right)^{-1} \left(\bar{\mathbf{Y}}^{(1)}\right) - \frac{1}{2} \left(\bar{\mathbf{Y}}^{(1)}\right)^{\top} \left(\hat{\mathbf{\Sigma}}^{(1)}\right)^{-1} \left(\bar{\mathbf{Y}}^{(1)}\right) \right\} d\hat{\mathbf{Y}}^{(1)}
$$

$$
= -\frac{(N^{(2)} + N_n^{(2)})D}{2} \log(2\pi) - \frac{1}{2} \log \left|\mathbf{K}_a^{(11)}\right| - \frac{1}{2} \log \left|\hat{\mathbf{\Sigma}}^{(1)}\right|
$$
$$
- \frac{1}{2} \left(\mathbf{E}^{(2)}\check{\mathbf{Y}}^{(2)} - \mathbf{E}^{(2)}\check{\mathbf{Y}}^{(1)}\right)^{\top} \left(\mathbf{K}_a^{(22)}\right)^{-1} \left(\mathbf{E}^{(2)}\check{\mathbf{Y}}^{(2)} - \mathbf{E}^{(2)}\check{\mathbf{Y}}^{(1)}\right)
$$
$$
- \left(\mathbf{E}^{(2)}\check{\mathbf{Y}}^{(2)} - \mathbf{E}^{(2)}\check{\mathbf{Y}}^{(1)}\right)^{\top} \left(\mathbf{K}_a^{(22)}\right)^{-1} \left(\hat{\mathbf{E}}^{(2)}\hat{\mathbf{Y}}^{(2)}\right) - \frac{1}{2} \left(\hat{\mathbf{E}}^{(2)}\hat{\mathbf{Y}}^{(2)}\right)^{\top} \left(\mathbf{K}_a^{(22)}\right)^{-1} \left(\hat{\mathbf{E}}^{(2)}\hat{\mathbf{Y}}^{(2)}\right)
$$
$$
- \frac{1}{2} \left(\bar{\mathbf{Y}}^{(1)}\right)^{\top} \left(\hat{\mathbf{\Sigma}}^{(1)}\right)^{-1} \left(\bar{\mathbf{Y}}^{(1)}\right)
$$
$$
+ \log \int \exp \left\{ \left(\mathbf{E}^{(2)}\check{\mathbf{Y}}^{(2)} - \mathbf{E}^{(2)}\check{\mathbf{Y}}^{(1)}\right)^{\top} \left(\mathbf{K}_a^{(22)}\right)^{-1} \left(\hat{\mathbf{E}}^{(2)}\hat{\mathbf{Y}}^{(1)}\right) \right.
$$
$$
+ \left(\hat{\mathbf{E}}^{(2)}\hat{\mathbf{Y}}^{(2)}\right)^{\top} \left(\mathbf{K}_a^{(22)}\right)^{-1} \left(\hat{\mathbf{E}}^{(2)}\hat{\mathbf{Y}}^{(1)}\right) - \frac{1}{2} \left(\hat{\mathbf{E}}^{(2)}\hat{\mathbf{Y}}^{(1)}\right)^{\top} \left(\mathbf{K}_a^{(22)}\right)^{-1} \left(\hat{\mathbf{E}}^{(2)}\hat{\mathbf{Y}}^{(1)}\right)
$$
$$
\left. - \frac{1}{2} \left(\hat{\mathbf{Y}}^{(1)}\right)^{\top} \left(\hat{\mathbf{\Sigma}}^{(1)}\right)^{-1} \left(\hat{\mathbf{Y}}^{(1)}\right) + \left(\bar{\mathbf{Y}}^{(1)}\right)^{\top} \left(\hat{\mathbf{\Sigma}}^{(1)}\right)^{-1} \left(\hat{\mathbf{Y}}^{(1)}\right) \right\}
$$

$$
= -\frac{(N^{(2)} + N_n^{(2)})D}{2} \log(2\pi) - \frac{1}{2} \log \left|\mathbf{K}_a^{(11)}\right| - \frac{1}{2} \log \left|\hat{\mathbf{\Sigma}}^{(1)}\right|
$$
$$
- \frac{1}{2} \left(\mathbf{E}^{(2)}\check{\mathbf{Y}}^{(2)} - \mathbf{E}^{(2)}\check{\mathbf{Y}}^{(1)}\right)^{\top} \left(\mathbf{K}_a^{(22)}\right)^{-1} \left(\mathbf{E}^{(2)}\check{\mathbf{Y}}^{(2)} - \mathbf{E}^{(2)}\check{\mathbf{Y}}^{(1)}\right)
$$
$$
- \left(\mathbf{E}^{(2)}\check{\mathbf{Y}}^{(2)} - \mathbf{E}^{(2)}\check{\mathbf{Y}}^{(1)}\right)^{\top} \left(\mathbf{K}_a^{(22)}\right)^{-1} \left(\hat{\mathbf{E}}^{(2)}\hat{\mathbf{Y}}^{(2)}\right) - \frac{1}{2} \left(\hat{\mathbf{E}}^{(2)}\hat{\mathbf{Y}}^{(2)}\right)^{\top} \left(\mathbf{K}_a^{(22)}\right)^{-1} \left(\hat{\mathbf{E}}^{(2)}\hat{\mathbf{Y}}^{(2)}\right)
$$
$$
- \frac{1}{2} \left(\bar{\mathbf{Y}}^{(1)}\right)^{\top} \left(\hat{\mathbf{\Sigma}}^{(1)}\right)^{-1} \left(\bar{\mathbf{Y}}^{(1)}\right)
$$
$$
+ \log \int \exp \left\{ -\frac{1}{2} \left(\hat{\mathbf{Y}}^{(1)}\right)^{\top} \left((\hat{\mathbf{E}}^{(2)})^{\top}(\mathbf{K}_a^{(22)})^{-1}\hat{\mathbf{E}}^{(2)} + \left(\hat{\mathbf{\Sigma}}^{(1)}\right)^{-1}\right) \hat{\mathbf{Y}}^{(1)} \right.
$$
$$
\left. + \left[\left(\mathbf{E}^{(2)}\check{\mathbf{Y}}^{(2)} - \mathbf{E}^{(2)}\check{\mathbf{Y}}^{(1)}\right)^{\top} \left(\mathbf{K}_a^{(22)}\right)^{-1} \hat{\mathbf{E}}^{(2)} + \left(\hat{\mathbf{E}}^{(2)}\hat{\mathbf{Y}}^{(2)}\right)^{\top} \left(\mathbf{K}_a^{(22)}\right)^{-1} \hat{\mathbf{E}}^{(2)} + \left(\bar{\mathbf{Y}}^{(1)}\right)^{\top} \left(\hat{\mathbf{\Sigma}}^{(1)}\right)^{-1}\right] \hat{\mathbf{Y}}^{(1)} \right\} d\hat{\mathbf{Y}}^{(1)}
$$

$$
= -\frac{N^{(2)}D}{2} \log(2\pi) - \frac{1}{2} \log \left|\mathbf{K}_a^{(11)}\right| - \frac{1}{2} \log \left|\hat{\mathbf{\Sigma}}^{(1)}\right| + \frac{1}{2} \log \left|(\hat{\mathbf{E}}^{(2)})^{\top}(\mathbf{K}_a^{(22)})^{-1}\hat{\mathbf{E}}^{(2)} + \left(\hat{\mathbf{\Sigma}}^{(1)}\right)^{-1}\right|
$$
$$
- \frac{1}{2} \left(\mathbf{E}^{(2)}\check{\mathbf{Y}}^{(2)} - \mathbf{E}^{(2)}\check{\mathbf{Y}}^{(1)}\right)^{\top} \left(\mathbf{K}_a^{(22)}\right)^{-1} \left(\mathbf{E}^{(2)}\check{\mathbf{Y}}^{(2)} - \mathbf{E}^{(2)}\check{\mathbf{Y}}^{(1)}\right)
$$
$$
- \left(\mathbf{E}^{(2)}\check{\mathbf{Y}}^{(2)} - \mathbf{E}^{(2)}\check{\mathbf{Y}}^{(1)}\right)^{\top} \left(\mathbf{K}_a^{(22)}\right)^{-1} \left(\hat{\mathbf{E}}^{(2)}\hat{\mathbf{Y}}^{(2)}\right) - \frac{1}{2} \left(\hat{\mathbf{E}}^{(2)}\hat{\mathbf{Y}}^{(2)}\right)^{\top} \left(\mathbf{K}_a^{(22)}\right)^{-1} \left(\hat{\mathbf{E}}^{(2)}\hat{\mathbf{Y}}^{(2)}\right)
$$
$$
- \frac{1}{2} \left(\bar{\mathbf{Y}}^{(1)}\right)^{\top} \left(\hat{\mathbf{\Sigma}}^{(1)}\right)^{-1} \left(\bar{\mathbf{Y}}^{(1)}\right)
$$
$$
+ \frac{1}{2} \left(\phi^{\top} \left(\mathbf{K}_a^{(22)}\right)^{-1} \hat{\mathbf{E}}^{(2)}\right) \left((\hat{\mathbf{E}}^{(2)})^{\top}(\mathbf{K}_a^{(22)})^{-1}\hat{\mathbf{E}}^{(2)} + \left(\hat{\mathbf{\Sigma}}^{(1)}\right)^{-1}\right)^{-1} \left(\left(\hat{\mathbf{E}}^{(2)}\right) \left(\mathbf{K}_a^{(22)}\right)^{-1} \phi\right)
$$
$$
+ \left(\phi^{\top} \left(\mathbf{K}_a^{(22)}\right)^{-1} \hat{\mathbf{E}}^{(2)}\right) \left((\hat{\mathbf{E}}^{(2)})^{\top}(\mathbf{K}_a^{(22)})^{-1}\hat{\mathbf{E}}^{(2)} + \left(\hat{\mathbf{\Sigma}}^{(1)}\right)^{-1}\right)^{-1} \left(\left(\hat{\mathbf{\Sigma}}\right)^{-1} \bar{\mathbf{Y}}^{(1)}\right)
$$
$$
+ \frac{1}{2} \left(\left(\bar{\mathbf{Y}}^{(1)}\right)^{\top} \left(\hat{\mathbf{\Sigma}}\right)^{-1}\right) \left((\hat{\mathbf{E}}^{(2)})^{\top}(\mathbf{K}_a^{(22)})^{-1}\hat{\mathbf{E}}^{(2)} + \left(\hat{\mathbf{\Sigma}}^{(1)}\right)^{-1}\right)^{-1} \left(\left(\hat{\mathbf{\Sigma}}\right)^{-1} \bar{\mathbf{Y}}^{(1)}\right).
$$
$$
\text{(A.48)}
$$

565 where $\phi = \left(\check{\mathbf{Y}}^{(2)}, \hat{\mathbf{Y}}^{(2)}\right)^{\top} - \left(\check{\mathbf{Y}}^{(1)}, \mathbf{0}\right)^{\top}$. After that we try to simplify (A.48) by different parts with
566 Sherman-Morrison formula.

567 First, we consider the determenant part,

$$
\begin{aligned}
&-\frac{1}{2}\log\left|\mathbf{K}_a^{(11)}\right| - \frac{1}{2}\log\left|\hat{\boldsymbol{\Sigma}}^{(1)}\right| + \frac{1}{2}\log\left|(\hat{\mathbf{E}}^{(2)})^{\top}(\mathbf{K}_a^{(22)})^{-1}\hat{\mathbf{E}}^{(2)} + \left(\hat{\boldsymbol{\Sigma}}^{(1)}\right)^{-1}\right| \\
=&-\frac{1}{2}\log\left|\mathbf{K}_a^{(11)}\right| - \frac{1}{2}\log\left|\hat{\boldsymbol{\Sigma}}^{(1)}\right| + \frac{1}{2}\log\left|\mathbf{K}_a^{(11)}\right| + \frac{1}{2}\log\left|\hat{\boldsymbol{\Sigma}}^{(1)}\right| - \frac{1}{2}\log\left|(\mathbf{K}_a^{(22)})^{-1} + \hat{\mathbf{E}}^{(2)}\left(\hat{\boldsymbol{\Sigma}}^{(1)}\right)^{-1}\left(\hat{\mathbf{E}}^{(2)}\right)^{\top}\right| \\
=&-\frac{1}{2}\log\left|(\mathbf{K}_a^{(22)})^{-1} + \hat{\mathbf{E}}^{(2)}\left(\hat{\boldsymbol{\Sigma}}^{(1)}\right)^{-1}\left(\hat{\mathbf{E}}^{(2)}\right)^{\top}\right|.
\end{aligned}
\tag{A.49}
$$

568 Then, we gather terms with ϕ and simplify them as,

$$
\begin{aligned}
&-\frac{1}{2}\left(\mathbf{E}^{(2)}\check{\mathbf{Y}}^{(2)} - \mathbf{E}^{(2)}\check{\mathbf{Y}}^{(1)}\right)^{\top}\left(\mathbf{K}_a^{(22)}\right)^{-1}\left(\mathbf{E}^{(2)}\check{\mathbf{Y}}^{(2)} - \mathbf{E}^{(2)}\check{\mathbf{Y}}^{(1)}\right) - \frac{1}{2}\left(\mathbf{E}_n^{(2)}\hat{\mathbf{Y}}^{(2)}\right)^{\top}\left(\mathbf{K}_a^{(22)}\right)^{-1}\left(\mathbf{E}_n^{(2)}\hat{\mathbf{Y}}^{(2)}\right) \\
&-\left(\mathbf{E}^{(2)}\check{\mathbf{Y}}^{(2)} - \mathbf{E}^{(2)}\check{\mathbf{Y}}^{(1)}\right)^{\top}\left(\mathbf{K}_a^{(22)}\right)^{-1}\left(\mathbf{E}_n^{(2)}\hat{\mathbf{Y}}^{(2)}\right) \\
&+\frac{1}{2}\left(\phi^{\top}\left(\mathbf{K}_a^{(22)}\right)^{-1}\hat{\mathbf{E}}^{(2)}\right)\left((\hat{\mathbf{E}}^{(2)})^{\top}(\mathbf{K}_a^{(22)})^{-1}\hat{\mathbf{E}}^{(2)} + \left(\hat{\boldsymbol{\Sigma}}^{(1)}\right)^{-1}\right)^{-1}\left(\left(\hat{\mathbf{E}}^{(2)}\right)\left(\mathbf{K}_a^{(22)}\right)^{-1}\phi\right) \\
=&-\frac{1}{2}\phi^{\top}\left[\left(\mathbf{K}_a^{(22)}\right)^{-1} - \left(\mathbf{K}_a^{(22)}\right)^{-1}\hat{\mathbf{E}}^{(2)}\left((\hat{\mathbf{E}}^{(2)})^{\top}(\mathbf{K}_a^{(22)})^{-1}\hat{\mathbf{E}}^{(2)} + \left(\hat{\boldsymbol{\Sigma}}^{(1)}\right)^{-1}\right)^{-1}\left(\hat{\mathbf{E}}^{(2)}\right)\right]\phi \\
=&-\frac{1}{2}\phi^{\top}\left[(\mathbf{K}_a^{(22)})^{-1} + \hat{\mathbf{E}}^{(2)}\left(\hat{\boldsymbol{\Sigma}}^{(1)}\right)^{-1}\left(\hat{\mathbf{E}}^{(2)}\right)^{\top}\right]\phi.
\end{aligned}
\tag{A.50}
$$

569 After that, we consider the interaction part between ϕ and $\bar{\mathbf{Y}}^{(1)}$,

$$
\begin{aligned}
&\left(\phi^{\top}\left(\mathbf{K}_a^{(22)}\right)^{-1}\hat{\mathbf{E}}^{(2)}\right)\left((\hat{\mathbf{E}}^{(2)})^{\top}(\mathbf{K}_a^{(22)})^{-1}\hat{\mathbf{E}}^{(2)} + \left(\hat{\boldsymbol{\Sigma}}^{(1)}\right)^{-1}\right)^{-1}\left(\left(\hat{\boldsymbol{\Sigma}}\right)^{-1}\bar{\mathbf{Y}}^{(1)}\right) \\
=&\phi^{\top}\left[\left(\mathbf{K}_a^{(22)}\right)^{-1} - \left(\mathbf{K}_a^{(22)}\right)^{-1}\hat{\mathbf{E}}^{(2)}\left((\hat{\mathbf{E}}^{(2)})^{\top}(\mathbf{K}_a^{(22)})^{-1}\hat{\mathbf{E}}^{(2)} + \left(\hat{\boldsymbol{\Sigma}}^{(1)}\right)^{-1}\right)^{-1}\right]\bar{\mathbf{Y}}^{(1)} \\
=&\phi^{\top}\left[(\mathbf{K}_a^{(22)})^{-1} + \hat{\mathbf{E}}^{(2)}\left(\hat{\boldsymbol{\Sigma}}^{(1)}\right)^{-1}\left(\hat{\mathbf{E}}^{(2)}\right)^{\top}\right]\hat{\mathbf{E}}^{(2)}\bar{\mathbf{Y}}^{(1)}.
\end{aligned}
\tag{A.51}
$$

570 Finally, we simplify the terms related with $\bar{\mathbf{Y}}^{(1)}$,

$$
\begin{aligned}
&-\frac{1}{2}\left(\bar{\mathbf{Y}}^{(1)}\right)^{\top}\left(\hat{\boldsymbol{\Sigma}}^{(1)}\right)^{-1}\left(\bar{\mathbf{Y}}^{(1)}\right) \\
&+\frac{1}{2}\left(\left(\bar{\mathbf{Y}}^{(1)}\right)^{\top}\left(\hat{\boldsymbol{\Sigma}}\right)^{-1}\right)\left((\hat{\mathbf{E}}^{(2)})^{\top}(\mathbf{K}_a^{(22)})^{-1}\hat{\mathbf{E}}^{(2)} + \left(\hat{\boldsymbol{\Sigma}}^{(1)}\right)^{-1}\right)^{-1}\left(\left(\hat{\boldsymbol{\Sigma}}\right)^{-1}\bar{\mathbf{Y}}^{(1)}\right) \\
=&-\frac{1}{2}\left(\bar{\mathbf{Y}}^{(1)}\right)^{\top}\left[\left(\hat{\boldsymbol{\Sigma}}^{(1)}\right)^{-1} - \left(\hat{\boldsymbol{\Sigma}}\right)^{-1}\left((\hat{\mathbf{E}}^{(2)})^{\top}(\mathbf{K}_a^{(22)})^{-1}\hat{\mathbf{E}}^{(2)} + \left(\hat{\boldsymbol{\Sigma}}^{(1)}\right)^{-1}\right)^{-1}\left(\hat{\boldsymbol{\Sigma}}\right)^{-1}\right]\left(\bar{\mathbf{Y}}^{(1)}\right) \\
=&-\frac{1}{2}\left(\bar{\mathbf{Y}}^{(1)}\right)^{\top}\left(\hat{\mathbf{E}}^{(2)}\right)^{\top}\left[(\mathbf{K}_a^{(22)})^{-1} + \hat{\mathbf{E}}^{(2)}\left(\hat{\boldsymbol{\Sigma}}^{(1)}\right)^{-1}\left(\hat{\mathbf{E}}^{(2)}\right)^{\top}\right]\hat{\mathbf{E}}^{(2)}\left(\bar{\mathbf{Y}}^{(1)}\right).
\end{aligned}
\tag{A.52}
$$

571 Putting all the parts together, the joint likelihood for the $R = 1$ is,

$$
\begin{aligned}
\mathcal{L} &= \log p(\mathbf{Y}^{(0:2)}) = \log p(\mathbf{Y}^{(0)}, \mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}) \\
&= \log p(\mathbf{Y}^{(0)}) + \log p(\mathbf{Y}^{(1)}|\mathbf{Y}^{(0)}) + \log \int p(\mathbf{Y}^{(2)}|\mathbf{Y}^{(1)}, \hat{\mathbf{Y}}^{(1)}) p(\hat{\mathbf{Y}}^{(1)}|\mathbf{Y}^{(1)}) d\hat{\mathbf{Y}}^{(1)} \\
&= \log p(\mathbf{Y}^{(0)}) + \log p(\mathbf{Y}^{(1)}|\mathbf{Y}^{(0)}) + \log \int p(\mathbf{Y}^{(2)}|\mathbf{Y}^{(1)}, \hat{\mathbf{Y}}^{(1)}) p(\hat{\mathbf{Y}}^{(1)}|\mathbf{Y}^{(1)}) d\hat{\mathbf{Y}}^{(0)} \\
&= \log p(\mathbf{Y}^{(0)}) + \log p(\mathbf{Y}^{(1)}|\mathbf{Y}^{(0)}) - \frac{N^{(2)}D}{2} \log(2\pi) - \frac{1}{2} \log \left| (\mathbf{K}_a^{(22)})^{-1} + \hat{\mathbf{E}}^{(2)} \left(\hat{\mathbf{\Sigma}}^{(1)} \right)^{-1} \left(\hat{\mathbf{E}}^{(2)} \right)^{\top} \right| \\
&\quad - \frac{1}{2} \left(\mathbf{Y}_a^{(2)} \right)^{\top} \left[(\mathbf{K}_a^{(22)})^{-1} + \hat{\mathbf{E}}^{(2)} \left(\hat{\mathbf{\Sigma}}^{(1)} \right)^{-1} \left(\hat{\mathbf{E}}^{(2)} \right)^{\top} \right] \mathbf{Y}_a^{(2)}.
\end{aligned}
\tag{A.53}
$$

572 Therefore, we can see that the joint likelihood of three different fidelities with non-subset dataset can be
573 decomposed as three independent model to train. For problems with arbitrary number of fidelity, we can easily
574 apply this conclusion recursively to decompose them into the summation structure, which is scalable to the
575 number of training data and fidelity level and is easy to code up.

J Experiment in Detail

J.1 Canonical PDEs

578 We consider three canonical PDEs: Poisson's equation, Heat equation, and Burger's equation, These PDEs have
579 crucial roles in scientific and technological applications [41–43]. They offer common simulation scenarios, such
580 as high-dimensional spatial-temporal field outputs, nonlinearities, and discontinuities, and are frequently used as
581 benchmark issues for surrogate models [29, 37–39]. x and y denote the spatial coordinates, and t specifies the
582 time coordinate, which contradicts the notation in the main paper. This notation in the appendix serves merely to
583 make the information clear; it has no bearing on or connections to the main article.

Burgers' equation is regarded as a standard nonlinear hyperbolic PDE; it is commonly used to represent a
variety of physical phenomena, including fluid dynamics [42], nonlinear acoustics [48], and traffic flows [49]. It
serves as a benchmark test case for several numerical solvers and surrogate models [50–52] since it can generate
discontinuities (shock waves) based on a normal conservation equation. The viscous version of this equation is
given by

$$
\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = v \frac{\partial^2 u}{\partial x^2},
$$

584 where u indicates volume, x represents a spatial location, t indicates the time, and v denotes the viscosity. We
585 set $x \in [0, 1]$, $t \in [0, 3]$, and $u(x, 0) = \sin(x\pi/2)$ with homogeneous Dirichlet boundary conditions. We
586 uniformly sampled viscosities $v \in [0.001, 0.1]$ as the input parameter to generate the solution field.

587 In the space and time domains, the problem is solved using finite elements with hat functions and backward
588 Euler, respectively. To generate solutions at different fidelities, the solver solve the PDEs based on discretized
589 spatial-temporal domain of regular rectangular meshes with $\{8^2, 16^2, 32^2, 64^2, 128^2\}$ grid points for simulations
590 from low- to high-fidelity.

Poisson's equation is a typical elliptic PDE in mechanical engineering and physics for modeling potential fields,
such as gravitational and electrostatic fields [41]. Written as

$$
\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0.
$$

593 It is a generalization of Laplace's equation [53]. Despite its simplicity, Poisson's equation is commonly
594 encountered in physics and is regularly used as a fundamental test case for surrogate models [37, 54]. In our
595 experiment, we impose Dirichlet boundary conditions on a 2D spatial domain with $\mathbf{x} \in [0, 1] \times [0, 1]$. The
596 input parameters consist of the constant values of the four borders and the center of the rectangular domain,
597 which vary from 0.1 to 0.9 each. We sampled the input parameters equally in order to create the matching
598 potential fields as outputs. Using the finite difference approach with a first-order center differencing scheme and
599 regular rectangular meshes, the PDE is solved. The meshes with $\{8^2, 16^2, 32^2, 64^2, 128^2\}$ grid nodes are used
600 to generate simulations of five different fidelity.

Heat equation is a fundamental PDE that defines the time-dependent evolution of heat fluxes. Despite having
been established in 1822 to describe just heat fluxes, the heat equation is prevalent in many scientific domains,
including probability theory [55, 43] and financial mathematics [56]. Consequently, it is commonly utilized as a
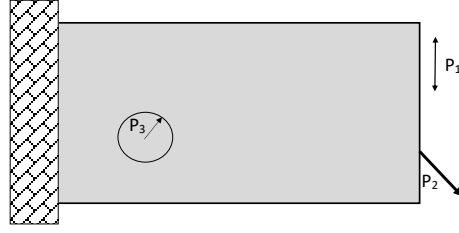
Figure 7: Geometry, boundary conditions, and simulation parameters for cantilever beam

stand-in model. This is the heat equation:

$$\frac{\partial}{\partial x}\left(k\frac{\partial T}{\partial x}\right) + \frac{\partial}{\partial y}\left(k\frac{\partial T}{\partial y}\right) + \frac{\partial}{\partial z}\left(k\frac{\partial T}{\partial z}\right) + q_V = \rho c_p \frac{\partial T}{\partial t}$$

where k is the materials conductivity q_V is the rate at which energy is generated per unit volume of the medium ρ is the density and c_p is the specific heat capacity. The input parameters are the flux rate of the left boundary at $x = 0$ (ranging from 0 to 1), the flux rate of the right boundary at $x = 1$ (ranging from -1 to 0), and the thermal conductivity (ranging from 0.01 to 0.1).

We establish a 2D spatial-temporal domain $x \in [0, 1]$, $t \in [0, 5]$ with the Neumann boundary condition at $x = 0$ and $x = 1$, and $u(x, 0) = H(x - 0.25) - H(x - 0.75)$, where $H(\cdot)$ is the Heaviside step function.

The equation is solved using the finite difference in space and backward Euler in time domains. The spatial-temporal domain is discretized into a 16×16 regular rectangular mesh for the first (lowest) fidelity solver. A refined solver uses a 32×32 mesh for the second fidelity. The result fields are computed on a 100×100 spatial-temporal grid.

The equation is solved using a finite difference in the spatial domain and reverse Euler in the temporal domain. The spatial-temporal domain is discretized into a regular rectangular mesh with $\{8^2, 16^2, 32^2, 64^2, 128^2\}$ nodes to generate simulation results of five different fidelity simulation.

J.2 Multi-Fidelity Fusion for Topology Optimization

We use GAR in a topology structure optimization problem, where the output is the best topology structure (in terms of maximum mechanical metrics like stiffness) of a layout of materials, such as alloy and concrete, given some design parameters like external force and angle. Topology structure optimization is a significant approach in mechanical designs, such as airfoils and slab bridges, especially with recent 3D printing processes in which material is deposited in minute quantities. However, it is well known that topology optimization is computationally intensive due to the gradient-based optimization and simulations of the mechanical characteristics involved. A high-fidelity solution, which necessitates a huge discretization mesh and imposes a significant computing overhead in space and time, makes matters worse.

Utilizing data-driven ways to aid in the process by offering the appropriate structures [57, 32] is subsequently gaining popularity. Here, we investigate the topology optimization of a cantilever beam (shown in the appendix). We employ the rapid implementation [59] to carry out density-based topology optimization by reducing compliance C subject to volume limitations $V \leq \bar{V}$.

The SIMP scheme [60] is used to convert continuous density measurements to discrete, optimal topologies. We set the position of point load $P1$, the angle of point load $P2$, and the filter radius $P3$ [61] as system input. The problems is solved based on a regular mesh with $\{32^2, 40^2, 48^2, 56^2, 64^2\}$ nodes.

J.3 Multi-Fidelity Fusion for Plasmonic nanoparticle arrays

In the final example, we calculate the extinction and scattering efficiencies Q_{ext} and Q_{sc} for plasmonic systems with varying numbers of scatterers using the Coupled Dipole Approximation (CDA) approach. CDA is a method for mimicking the optical response of an array of similar, non-magnetic metallic nanoparticles with dimensions far smaller than the wavelength of light (here 25 nm). Q_{ext} and Q_{sc} are defined as the QoIs in this document. We constructed surrogate models for efficiency with up to three fidelities using our proposed method. We examined particle arrays resulting from Vogel spirals. Since the number of interactions of incident waves from particles influences the magnetic field, the number of nanoparticles in a plasmonic array has a substantial effect on the local extinction field caused by plasmonic arrays. The configurations of Vogel spirals with particle numbers in the set $\{2, 50, 200, 500, 1000\}$ that define five fidelity simulations. $\lambda \in [200, 800]$ nm, $\alpha_{vs} \in [0, 2\pi]$ rad, and

$a_{vs} \in (1, 1500)$ were determined to be the parameter space. These are, respectively, the incidence wavelength, the divergence angle, and the scaling factor. A Sobol sequence was utilized to choose inputs. The computing time required to execute CDA increases exponentially as the number of nanoparticles increases. Consequently, the proposed sampling approach results in significant reductions in computational costs.

The response of a plasmonic array to electromagnetic radiation is calculable using the solution of the local electric fields, $\mathbf{E}_{loc}(\mathbf{r}_j)$, for each nano-sphere. Considering N metallic particles defined by the same volumetric polarizability $\alpha(\omega)$ and situated at vector coordinates $\mathbf{r}_i$, it is possible to calculate the local field $\mathbf{E}_{loc}(\mathbf{r}_j)$ by solving [62] the corresponding linear equation.

$$\mathbf{E}_{loc}(\mathbf{r}_i) = \mathbf{E}_0(\mathbf{r}_i) - \frac{\alpha k^2}{\epsilon_0} \sum_{j=1, j \neq i}^{N} \tilde{\mathbf{G}}_{ij} \mathbf{E}_{loc}(\mathbf{r}_j) \tag{A.54}$$

in which $\mathbf{E}_0(\mathbf{r}_i)$ is the incident field, k is the wave number in the background medium, ϵ_0 denotes the dielectric permittivity of vacuum ($\epsilon_0 = 1$ in the CGS unit system), and $\tilde{\mathbf{G}}_{ij}$ is constructed from 3×3 blocks of the overall $3N \times 3N$ Green's matrices for the ith and jth particles. $\tilde{\mathbf{G}}_{ij}$ is a zero matrix when $j = i$, and otherwise calculated as

$$\tilde{\mathbf{G}}_{ij} = \frac{\exp(ikr_{ij})}{r_{ij}} \left\{ \mathbf{I} - \widehat{\mathbf{r}}_{ij}\widehat{\mathbf{r}}_{ij}^T - \left[\frac{1}{ikr_{ij}} + \frac{1}{(kr_{ij})^2}(\mathbf{I} - 3\widehat{\mathbf{r}}_{ij}\widehat{\mathbf{r}}_{ij}^T) \right] \right\} \tag{A.55}$$

where $\widehat{\mathbf{r}}_{ij}$ denotes the unit position vector from particles j to i and $r_{ij} = |\mathbf{r}_{ij}|$. By solving Eqs. A.54 and A.55, the total local fields $\mathbf{E}_{loc}(\mathbf{r}_i)$, and as a result the scattering and extinction cross-sections, are computed. Details of the numerical solution can be found in [63].

Q_{ext} and Q_{sc} are derived by normalizing the scattering and extinction cross-sections relative to the array's entire projected area. We considered the Vogel spiral class of particle arrays, which is described by [64]

$$\rho_n = \sqrt{n}a_{vs} \quad \text{and} \quad \theta_n = n\alpha_{vs}, \tag{A.56}$$

where ρ_n and θ_n represent the radial distance and polar angle of the n-th particle in a Vogel spiral array, respectively. Therefore, the Vogel spiral configuration may be uniquely defined by the incidence wavelength λ, the divergence angle α_{vs}, the scaling factor a_{vs}, and the number of particles n.

J.4   Detailed Prediction Analysis

To investigate the prediction error in detail, we define the average RMSE field $\mathbf{y}^{(EF)}$ by

$$\mathbf{y}^{(EF)} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\mathbf{y}_i - \tilde{\mathbf{y}}_i)^2},$$

where $\tilde{\mathbf{y}}_i$ is the prediction, $\mathbf{y}_i$ is the ground true value, and the square root is element-wise operation.

The average RMSE field corresponding to the Heat equation, Burger's equation, Poisson's equation, and TopOP problem with a decreasing rate of $\eta = 0.5$, lowest-fidelity training samples,$N^0 = 128$ and 128 testing samples are shown in Fig. 8 for the subset assessment and Fig. 9 for the non-subset assessment. Plasmonic nanoparticle arrays (PNA) have only two output variables and thus we do not show the average RMSE field for it.

For the classic subset assessment in Fig. 8, we can see that clear that ContinuAR outperforms the competitors in all cases with a large margin by showing more blue areas and only tiny red areas. For Heat equation, the error is significantly reduced in most areas except for the bottom area of the domain, where a tiny thin bar of red area is shown. For the Burger's equation, ContinuAR show some checker board pattern in the error field, which is probability caused by the conditional independence assumption. Nonetheless, the error is significantly reduced in most areas the largest error is also reduced. For the Poisson's equation, different method has it own error patterns. ContinuAR show a more blue area in the left part of the domain, while AR has high-error areas everywhere except the center. For the TopOP problem, ContinuAR show a significantly reduced error by a significant reductions of deep red areas. The AR as the deepest blue ares, indicating its good performance. However, it also has a lar areas of red, indicating its poor performance overall.

For the classic non-subset assessment in Fig. 9, the overall conclusion is similar to the subset assessment. ContinuAR outperforms the competitors in all cases with a large margin by showing more blue areas and tiny red areas. The overall error pattern for most methods are similar to the subset assessment with subtle difference. For instance, for the Burger's equation, checker board pattern disappear, indicating a successful improvement by using training data across the input domain. The error for Poisson's equation is also significantly reduced particularly on the right part of the domain. TopOP turns out to be the most stable problem as the error patterns for all methods are almost the same as the subset assessment.
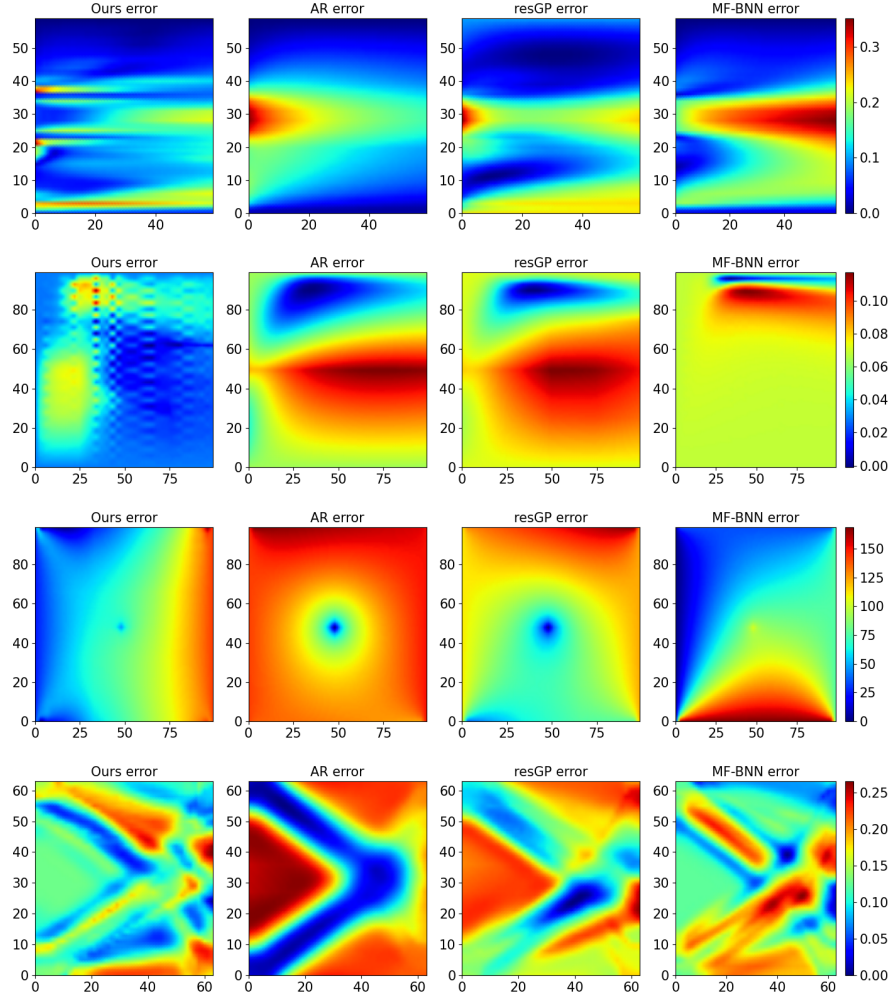
24

Figure 8: Average RMSE fields of the subset evaluation with $\eta = 0.5$ and $128 \ N^0$ training samples for Heat equation(1st row), Burger's equation(2nd row), Poisson's equation(3rd row) and TopOP(4th row).
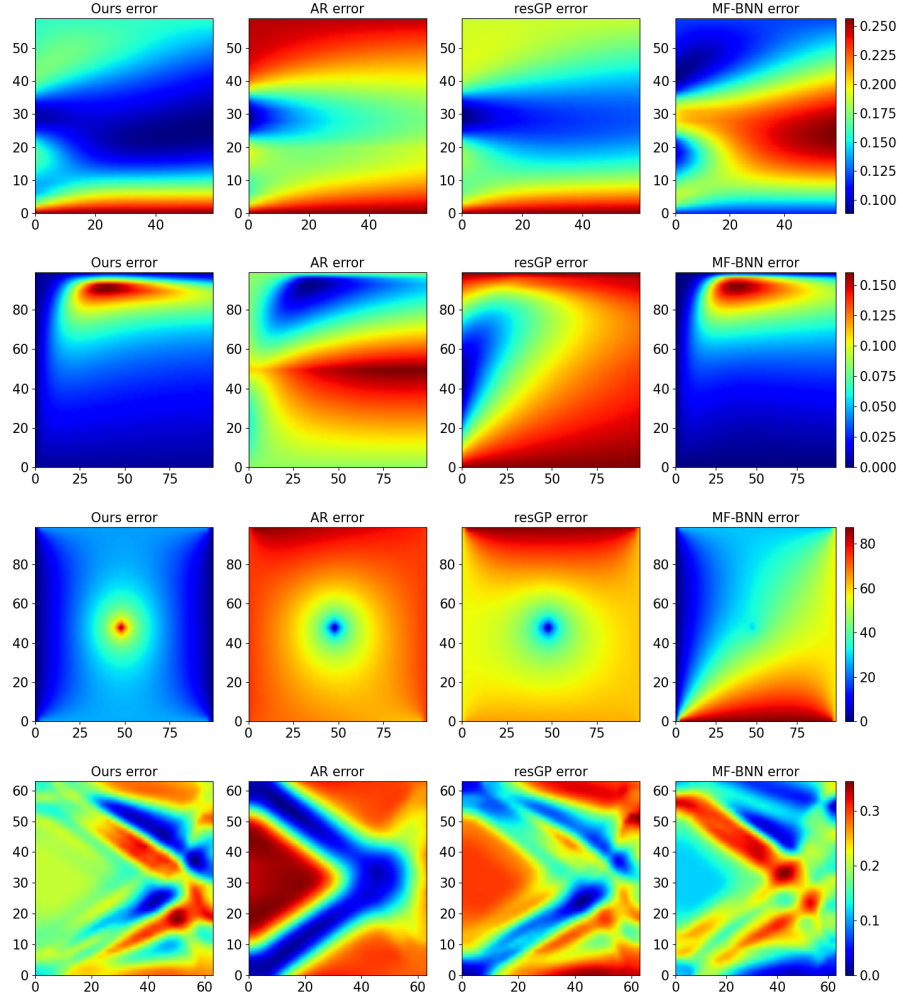
Figure 9: Average RMSE fields of the non-subset evaluation with $\eta = 0.5$ and 128 N^0 training samples for Heat equation(1st row), Burger's equation(2nd row), Poisson's equation(3rd row) and TopOP(4th row).

Supplementary References

[10] Loic Le Gratiet. Bayesian analysis of hierarchical multifidelity codes. SIAM/ASA Journal on Uncertainty Quantification, 1(1):244–269, 2013.

[41] Steven C Chapra, Raymond P Canale, et al. Numerical methods for engineers. Boston: McGraw-Hill Higher Education,.

[42] TJ Chung. Computational fluid dynamics. Cambridge university press.

[43] Krzysztof Burdzy, Zhen-Qing Chen, John Sylvester, et al. The heat equation and reflected brownian motion in time-dependent domains. The Annals of Probability, 32(1B):775–804.

[29] Wei W. Xing, Robert M. Kirby, and Shandian Zhe. Deep coregionalization for the emulation of simulation-based spatial-temporal fields. Journal of Computational Physics, 428:109984, March 2021. ISSN 00219991. doi: 10.1016/j.jcp.2020.109984.

[37] Rui Tuo, C. F. Jeff Wu, and Dan Yu. Surrogate Modeling of Computer Experiments With Different Mesh Densities. Technometrics, 56(3):372–380, July 2014. ISSN 0040-1706, 1537-2723. doi: 10.1080/00401706.2013.842935.

[38] Mehmet Onder Efe and Hitay Ozbay. Proper orthogonal decomposition for reduced order modeling: 2d heat flow. In Proceedings of 2003 IEEE Conference on Control Applications, 2003. CCA 2003., volume 2, pages 1273–1277. IEEE.

[39] Maziar Raissi and George Em Karniadakis. Machine Learning of Linear Differential Equations using Gaussian Processes. Journal of Computational Physics, 348:683–693, November 2017. ISSN 00219991. doi: 10/gbzfgr.

[48] N Sugimoto. Burgers equation with a fractional derivative; hereditary effects on nonlinear acoustic waves. Journal of fluid mechanics, 225:631–653.

[49] Kai Nagel. Particle hopping models and traffic flow theory. Physical review E, 53(5):4655.

[50] S Kutluay, AR Bahadir, and A Özdeç. Numerical solution of one-dimensional burgers equation: explicit and exact-explicit finite difference methods. Journal of Computational and Applied Mathematics, 103(2):251–261.

[51] A. A. Shah, W. W. Xing, and V. Triantafyllidis. Reduced-order modelling of parameter-dependent, linear and nonlinear dynamic partial differential equation models. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 473(2200):20160809, April 2017. ISSN 1364-5021, 1471-2946. doi: 10.1098/rspa.2016.0809.

[52] Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations. arXiv preprint arXiv:1711.10561.

[53] S Persides. The laplace and poisson equations in schwarzschild's space-time. Journal of Mathematical Analysis and Applications, 43(3):571–578.

[54] I.E. Lagaris, A. Likas, and D.I. Fotiadis. Artificial neural networks for solving ordinary and partial differential equations. IEEE Transactions on Neural Networks, 9(5):987–1000, Sept./1998. ISSN 10459227. doi: 10.1109/72.712178.

[55] Frank Spitzer. Electrostatic capacity, heat flow, and brownian motion. Probability theory and related fields, 3(2):110–121.

[56] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. Journal of political economy, 81(3):637–654.

[57] Wei Xing, Shireen Y. Elhabian, Vahid Keshavarzzadeh, and Robert M. Kirby. Shared-Gaussian Process: Learning Interpretable Shared Hidden Structure Across Data Spaces for Design Space Analysis and Exploration. Journal of Mechanical Design, 142(8), August 2020. ISSN 1050-0472, 1528-9001. doi: 10.1115/1.4046074.

[32] Shibo Li, Robert M Kirby, and Shandian Zhe. Deep multi-fidelity active learning of high-dimensional outputs. arXiv preprint arXiv:2012.00901, 2020.

[59] Erik Andreassen, Anders Clausen, Mattias Schevenels, Boyan S. Lazarov, and Ole Sigmund. Efficient topology optimization in matlab using 88 lines of code. Structural and Multidisciplinary Optimization, 43 (1):1–16, Jan 2011. ISSN 1615-1488.

[60] Martin Philip Bendsoe and Ole Sigmund. Topology optimization: Theory, methods and applications. Springer, 2004.

[61] Tyler E. Bruns and Daniel A. Tortorelli. Topology optimization of non-linear elastic structures and compliant mechanisms. Computer Methods in Applied Mechanics and Engineering, 190(26):3443 – 3459, 2001. ISSN 0045-7825.

[62] Charles-Antoine Guérin, Pierre Mallet, and Anne Sentenac. Effective-medium theory for finite-size aggregates. JOSA A, 23(2):349–358, 2006.

[63] Mani Razi, Ren Wang, Yanyan He, Robert M. Kirby, and Luca Dal Negro. Optimization of Large-Scale Vogel Spiral Arrays of Plasmonic Nanoparticles. Plasmonics, 14(1):253–261, February 2019. ISSN 1557-1955, 1557-1963. doi: 10.1007/s11468-018-0799-y.

[64] Aristi Christofi, Felipe A Pinheiro, and Luca Dal Negro. Probing scattering resonances of vogel's spirals with the green's matrix spectral method. Optics letters, 41(9):1933–1936, 2016.