### S.1 Proof of (14)

Recall that the model parameters are $\boldsymbol{\theta} = (\boldsymbol{\theta}_l, \boldsymbol{\theta}_h)$. We express $p(y_h(\mathbf{x}_0)|\mathbf{y}_l, \mathbf{y}_h)$ as

$$\begin{aligned} p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \mathbf{y}_l, \mathbf{y}_h) &= \int_{\boldsymbol{\theta}_l}\int_{\boldsymbol{\theta}_h} p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \mathbf{y}_l, \mathbf{y}_h, \boldsymbol{\theta}_l, \boldsymbol{\theta}_h) p(\boldsymbol{\theta}_l, \boldsymbol{\theta}_h|\mathbf{M}_k, \mathbf{y}_l, \mathbf{y}_h) d\boldsymbol{\theta}_h d\boldsymbol{\theta}_l \\ &= \int_{\boldsymbol{\theta}_l}\int_{\boldsymbol{\theta}_h} p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \mathbf{y}_h, \boldsymbol{\eta}_l, \boldsymbol{\theta}_h) p(\boldsymbol{\theta}_h|\mathbf{M}_k, \boldsymbol{\eta}_l(\mathbf{X}_l), \mathbf{y}_h) d\boldsymbol{\theta}_h p(\boldsymbol{\eta}_l|\mathbf{y}_l, \boldsymbol{\theta}_l) p(\boldsymbol{\theta}_l|\mathbf{y}_l) d\boldsymbol{\theta}_l \\ &= \int_{\boldsymbol{\theta}_l} p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \mathbf{y}_h, \boldsymbol{\eta}_l) p(\boldsymbol{\eta}_l|\mathbf{M}_k, \mathbf{y}_l, \boldsymbol{\theta}_l) p(\boldsymbol{\theta}_l|\mathbf{y}_l) d\boldsymbol{\theta}_l \end{aligned} \tag{23}$$

where

$$\begin{aligned} p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \boldsymbol{\eta}_l, \mathbf{y}_h) &\equiv \int_{\boldsymbol{\theta}_h} p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \mathbf{y}_h, \boldsymbol{\eta}_l, \boldsymbol{\theta}_h)\ p(\boldsymbol{\theta}_h|\mathbf{M}_k, \boldsymbol{\eta}_l, \mathbf{y}_h) d\boldsymbol{\theta}_h \\ &= \int_{\boldsymbol{\alpha}, \sigma_e^2, \boldsymbol{\lambda}} p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \boldsymbol{\eta}_l, \boldsymbol{\alpha}, \sigma_e^2, \boldsymbol{\lambda})\ p(\boldsymbol{\alpha}, \sigma_e^2, \boldsymbol{\lambda}|\mathbf{M}_k, \boldsymbol{\eta}_l, \mathbf{y}_h)\ d\boldsymbol{\alpha}\ d\sigma_e^2\ d\boldsymbol{\lambda} \\ &= \int_{\boldsymbol{\lambda}}\int_{\sigma_e^2, \boldsymbol{\alpha}} p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \boldsymbol{\eta}_l, \boldsymbol{\alpha}, \sigma_e^2, \boldsymbol{\lambda})\ p(\boldsymbol{\alpha}, \sigma_e^2|\mathbf{M}_k, \boldsymbol{\lambda}, \mathbf{y}_h, \boldsymbol{\eta}_l)\ d\boldsymbol{\alpha}\ d\sigma_e^2\ p(\boldsymbol{\lambda}|\mathbf{M}_k, \boldsymbol{\eta}_l, \mathbf{y}_h) d\boldsymbol{\lambda} \end{aligned} \tag{24}$$

In order to get the expression of $p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \boldsymbol{\eta}_l, \mathbf{y}_h)$, perform the integration in (24) in the following two steps:
(i) Integrate out $\boldsymbol{\alpha}$ and $\sigma_e^2$;
(ii) Integrate out $\boldsymbol{\lambda}$.

Step (i) integrate out $\boldsymbol{\alpha}$ and $\sigma_e^2$. We denote the inner integration in (24) by $p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \mathbf{y}_h, \boldsymbol{\eta}_l, \boldsymbol{\lambda})$, that is,

$$\begin{aligned} p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \mathbf{y}_h, \boldsymbol{\eta}_l, \boldsymbol{\lambda}) &\equiv \int_{\sigma_e^2, \boldsymbol{\alpha}} p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \boldsymbol{\eta}_l, \boldsymbol{\alpha}, \sigma_e^2, \boldsymbol{\lambda})\ p(\boldsymbol{\alpha}, \sigma_e^2|\mathbf{M}_k, \boldsymbol{\lambda}, \mathbf{y}_h, \boldsymbol{\eta}_l)\ d\boldsymbol{\alpha}\ d\sigma_e^2 \\ &\propto \int_{\sigma_e^2, \boldsymbol{\alpha}} p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \boldsymbol{\eta}_l, \boldsymbol{\alpha}, \sigma_e^2, \boldsymbol{\lambda}) p(\mathbf{y}_h|\mathbf{M}_k, \boldsymbol{\eta}_l, \boldsymbol{\alpha}, \sigma_e^2, \boldsymbol{\lambda})\ p(\boldsymbol{\alpha}, \sigma_e^2)\ d\boldsymbol{\alpha}\ d\sigma_e^2 \end{aligned}$$

Given the kernel width $\boldsymbol{\lambda}$, the linkage model can be considered as a linear regression model $\mathbf{y}_h = \mathbf{F}_{\boldsymbol{\lambda}}\boldsymbol{\alpha} + \boldsymbol{\epsilon}_h$. Recall that $\boldsymbol{\epsilon}_h \sim N(\mathbf{0}, \sigma_e^2\mathbf{I})$. Therefore,

$$\begin{aligned} (\mathbf{y}_h|\mathbf{M}_k, \boldsymbol{\eta}_l, \boldsymbol{\alpha}, \sigma_e^2, \boldsymbol{\lambda}) &\sim N(\mathbf{F}_{\boldsymbol{\lambda}}\boldsymbol{\alpha}, \sigma_e^2\mathbf{I}) \\ (y_h(\mathbf{x}_0)|\mathbf{M}_k, \boldsymbol{\eta}_l, \boldsymbol{\alpha}, \sigma_e^2, \boldsymbol{\lambda}) &\sim N(\mathbf{F}_{\boldsymbol{\lambda}}(\mathbf{x}_0)\boldsymbol{\alpha}, \sigma_e^2) \end{aligned}$$

These are the same results as in (10) and (11). Given that the prior distribution of $\boldsymbol{\alpha}$ and $\sigma_e^2$ is $p(\boldsymbol{\alpha}, \sigma_e^2) \propto \sigma_h^{-2}$, Gelman et al. (2003, Page 359) stated that under this priors, the posterior predictive distribution of $y_h(\mathbf{x}_0)$, conditioned on the data and kernel width $\boldsymbol{\lambda}$, is

$$(y_h(\mathbf{x}_0)|\mathbf{M}_k, \mathbf{y}_h, \boldsymbol{\eta}_l(\mathbf{X}_0), \boldsymbol{\lambda}) \sim t_{m_h-2}(\mathbf{F}_{\boldsymbol{\lambda}}(\mathbf{x}_0)\hat{\boldsymbol{\alpha}}, s^2(1 + \mathbf{F}_{\boldsymbol{\lambda}}(\mathbf{x_0})(\mathbf{F}_{\boldsymbol{\lambda}}^T\mathbf{F}_{\boldsymbol{\lambda}})^{-1}\mathbf{F}_{\boldsymbol{\lambda}}(\mathbf{x_0})^T)).$$

where $\hat{\boldsymbol{\alpha}} = (\mathbf{F}_{\boldsymbol{\lambda}}^T\mathbf{F}_{\boldsymbol{\lambda}})^{-1}\mathbf{F}_{\boldsymbol{\lambda}}^T\mathbf{y}_h$ and $s^2 = \frac{1}{m_h-2}(\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}}\hat{\boldsymbol{\alpha}})^T(\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}}\hat{\boldsymbol{\alpha}})$. This is how (14) is obtained. Consequently, after $\boldsymbol{\alpha}$ and $\sigma_e^2$ are integrated out, (24) becomes

$$p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \boldsymbol{\eta}_l, \mathbf{y}_h) = \int_{\boldsymbol{\lambda}} p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \mathbf{y}_h, \boldsymbol{\eta}_l(\mathbf{X}_0), \boldsymbol{\lambda})\ p(\boldsymbol{\lambda}|\mathbf{M}_k, \boldsymbol{\eta}_l, \mathbf{y}_h) d\boldsymbol{\lambda}. \tag{25}$$

Step (ii), integrate out $\boldsymbol{\lambda}$. Recall that $\boldsymbol{\lambda}$ has a discrete distribution. Thus, the integration in (25) can be written as a summation (and (15) is obtained):

$$p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \boldsymbol{\eta}_l, \mathbf{y}_h) = \sum_{\substack{\lambda_1=1,2,\ldots,\lambda_0 \\ \vdots \\ \lambda_d=1,2,\ldots,\lambda_0}} p(y_h(\mathbf{x}_0)|\mathbf{M}_k, \mathbf{y}_h, \boldsymbol{\eta}_l, \boldsymbol{\lambda})\ p(\boldsymbol{\lambda}|\mathbf{M}_k, \mathbf{y}_h, \boldsymbol{\eta}_l),$$

where

$$p(\boldsymbol{\lambda}|\mathbf{M}_k, \mathbf{y}_h, \boldsymbol{\eta}_l) \propto p(\boldsymbol{\lambda})p(\mathbf{y}_h|\mathbf{M}_k, \boldsymbol{\eta}_l, \boldsymbol{\lambda}). \tag{26}$$

The marginal distribution of the high-resolution data given the inputs $\boldsymbol{\eta}_l$ and the kernel width $\boldsymbol{\lambda}$ is as follows

$$\begin{aligned}
p(\mathbf{y}_h|\mathbf{M}_k, \boldsymbol{\eta}_l, \boldsymbol{\lambda}) &= \int_{\sigma_e^2, \boldsymbol{\alpha}} p(\mathbf{y}_h, \boldsymbol{\alpha}, \sigma_e^2|\mathbf{M}_k, \boldsymbol{\eta}_l, \boldsymbol{\lambda})\ d\boldsymbol{\alpha}\ d\sigma_e^2 \qquad (27)\\
&= \int_{\sigma_e^2, \boldsymbol{\alpha}} p(\mathbf{y}_h|\mathbf{M}_k, \boldsymbol{\eta}_l, \boldsymbol{\alpha}, \sigma_e^2, \boldsymbol{\lambda})\ p(\boldsymbol{\alpha}, \sigma_e^2)\ d\boldsymbol{\alpha}\ d\sigma_e^2 \\
&= \int_{\sigma_e^2} \int_{\boldsymbol{\alpha}} (2\pi)^{-\frac{m_h}{2}} (\sigma_e^2)^{-\frac{m_h}{2}}\ exp\left\{-\frac{1}{2\sigma_e^2}(\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}}\boldsymbol{\alpha})^T(\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}}\boldsymbol{\alpha})\right\}\ d\boldsymbol{\alpha}\ \sigma_h^{-2} d\sigma_e^2 \\
&= \int_{\sigma_e^2} \int_{\boldsymbol{\alpha}} exp\left\{-\frac{1}{2\sigma_e^2}[(\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}}\hat{\boldsymbol{\alpha}})^T(\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}}\hat{\boldsymbol{\alpha}}) + (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})^T\mathbf{F}_{\boldsymbol{\lambda}}^T\mathbf{F}_{\boldsymbol{\lambda}}(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})]\right\}\ d\boldsymbol{\alpha} \\
&\quad \cdot (2\pi)^{-\frac{m_h}{2}} (\sigma_e^2)^{-\frac{m_h}{2}-1} d\sigma_e^2 \\
&= \int_{\sigma_e^2} \int_{\boldsymbol{\alpha}} exp\left\{-\frac{1}{2\sigma_e^2}(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})^T\mathbf{F}_{\boldsymbol{\lambda}}^T\mathbf{F}_{\boldsymbol{\lambda}}(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})\right\}\ d\boldsymbol{\alpha} \\
&\quad \cdot (2\pi)^{-\frac{m_h}{2}} exp\left\{-\frac{1}{2\sigma_e^2}(\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}}\hat{\boldsymbol{\alpha}})^T(\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}}\hat{\boldsymbol{\alpha}})\right\} (\sigma_e^2)^{-\frac{m_h}{2}-1} d\sigma_e^2 \\
&= \int_{\sigma_e^2} 2\pi|\mathbf{F}_{\boldsymbol{\lambda}}^T\mathbf{F}_{\boldsymbol{\lambda}}|^{-\frac{1}{2}}\sigma_e^2(2\pi)^{-\frac{m_h}{2}} exp\left\{-\frac{1}{2\sigma_e^2}(\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}}\hat{\boldsymbol{\alpha}})^T(\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}}\hat{\boldsymbol{\alpha}})\right\} (\sigma_e^2)^{-\frac{m_h}{2}-1} d\sigma_e^2 \\
&=(2\pi)^{-\frac{m_h}{2}+1}|\mathbf{F}_{\boldsymbol{\lambda}}^T\mathbf{F}_{\boldsymbol{\lambda}}|^{-\frac{1}{2}} \int_{\sigma_e^2} (\sigma_e^2)^{-(\frac{m_h}{2}-1+1)} exp\left\{-\frac{(\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}}\hat{\boldsymbol{\alpha}})^T(\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}}\hat{\boldsymbol{\alpha}})}{2\sigma_e^2}\right\} d\sigma_e^2 \\
&=(2\pi)^{-\frac{m_h}{2}+1}|\mathbf{F}_{\boldsymbol{\lambda}}^T\mathbf{F}_{\boldsymbol{\lambda}}|^{-\frac{1}{2}}\Gamma\left(\frac{m_h}{2} - 1\right) \left[\frac{(\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}}\hat{\boldsymbol{\alpha}})^T(\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}}\hat{\boldsymbol{\alpha}})}{2}\right]^{-\frac{m_h}{2}+1}.
\end{aligned}$$

Note that in the second line of the above derivation, we utilize that (11) which specifies the distribution of $p(\mathbf{y}_h|\boldsymbol{\eta}_l, \boldsymbol{\alpha}, \sigma_e^2, \boldsymbol{\lambda})$. Given the above, (26) can now be written as

$$p(\boldsymbol{\lambda}|\mathbf{y}_h, \boldsymbol{\eta}_l) \propto p(\boldsymbol{\lambda})|\mathbf{F}_{\boldsymbol{\lambda}}^T\mathbf{F}_{\boldsymbol{\lambda}}|^{-\frac{1}{2}} \left[\frac{(\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}}\hat{\boldsymbol{\alpha}})^T(\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}}\hat{\boldsymbol{\alpha}})}{2}\right]^{-\frac{m_h}{2}+1}.$$

This shows how (14) is obtained.

## S.2 Proof of (20)

To prove (20), we just integrate (27) over $\lambda$. As $\lambda$ takes discrete values, we end up with a summation over possible values for $\lambda$ as in (20):

$$p(\mathbf{y}_h|\mathbf{M}_k, \boldsymbol{\eta}_l) = \sum_{\substack{\lambda_1=1,2,\ldots,\lambda_0 \\ \vdots \\ \lambda_d=1,2,\ldots,\lambda_0}} (2\pi)^{-\frac{m_h}{2}+1} |\mathbf{F}_{\boldsymbol{\lambda}}^T \mathbf{F}_{\boldsymbol{\lambda}}|^{-\frac{1}{2}} \Gamma(\frac{m_h}{2} - 1)\big[\frac{(\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}}\hat{\boldsymbol{\alpha}})^T(\mathbf{y}_h - \mathbf{F}_{\boldsymbol{\lambda}}\hat{\boldsymbol{\alpha}})}{2}\big].$$