

SUPPLEMENTARY MATERIALS: Coupling Computer Models through Linking their Statistical Emulators*

Ksenia N. Kzyurova[†], James O. Berger[‡], and Robert L. Wolpert[‡]

SM1. Illustration of a linked GaSP of two computer models with independent designs.

This illustration is the modification of the illustration 2 in subsection 6.1.2. We take the same two functions, $f(x) = \sin(\pi x)$ in the domain $x \in [-1, 1]$ and $g(z) = \cos(5z)$ in the domain $z \in [-1, 1]$.

Model $f(x)$ has been evaluated at 6 training input points (as before in subsection 6.1.2) \mathbf{x} , resulting in $\mathbf{f}(\mathbf{x})$. The difference is that now we choose points \mathbf{z} independently of the design for model f , to make a design for the emulator $g^M(\cdot)$. We take 6 equally spaced desing points from -0.95 to 0.95 . Model g is evaluated at \mathbf{z} . The simulators, along with their designs and constructed type 2 GaSP emulators are shown in Figure SM1.

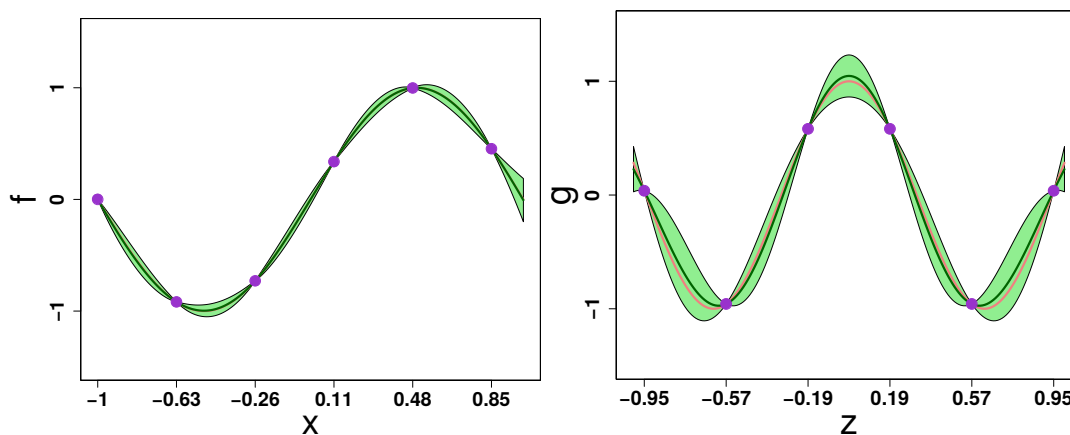


Figure SM1. Independent emulators constructed for two test functions: $f(x)$ on the left panel and $g(x)$ on the right. Each emulator is an interpolator at its design points. The dark green lines are the emulator means. The pink lines are the true functions. The green shaded regions correspond to the emulators areas encompassed by 2.5% and 97.5% quantiles of the emulators. The circles on the plots correspond to the design points which were used to fit the emulators.

The linked GaSP methodology allows to link such models. The resulting linked GaSP is shown on Figure SM2. As can be seen no computer model data are observed from the system of models (thus with no composite emulator to be constructed), however, the (accurate) linked

*Submitted to the editors DATE.

Funding: This research was performed as part of the first author's PhD thesis at Duke University, and supported by NSF grants DMS-12-28317, EAR-1331353, DMS-14-07775 and DMS-16-22467.

[†]CEMSE division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia (ksenia.kzyurova@kaust.edu.sa, <http://www2.stat.duke.edu/~kk194/>).

[‡]Department of Statistical Science, Duke University, Durham, NC, 27708 (berger@stat.duke.edu, wolpert@stat.duke.edu).

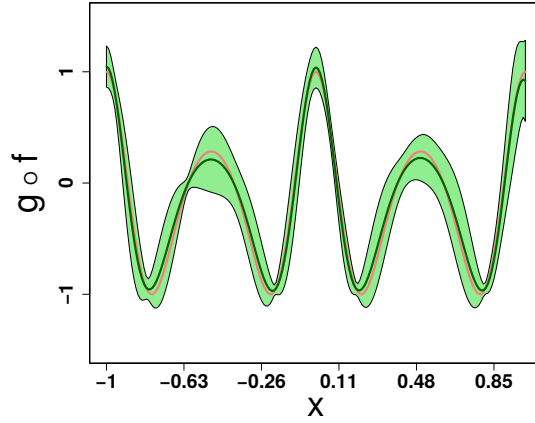


Figure SM2. Linked GaSP of a system of simulators.

The dark green line is the emulator mean. The pink line is the true function. The green shaded region corresponds to the emulator area encompassed by 2.5% and 97.5% quantiles of the emulator. Circles on the plot correspond to sequentially obtained design points $\{\mathbf{x}, (\mathbf{g} \circ \mathbf{f})(\mathbf{x})\}$.

emulator, approximation to a composite model, is constructed.

The composite emulator is not available, since there are no data available to construct such an emulator.

SM2. Linked emulator with multivariate GaSP. GaSP emulator $g^M(\cdot)$ of the model g at any new d -dimensional input given pairs of model observations $\{\mathbf{z}, \mathbf{g}(\mathbf{z})\}$ and GaSP parameters $\theta_{\mathbf{g}}$ is (4).

Suppose linking an emulator $g^M(\cdot, \dots, \cdot)$ with multivariate emulator $f_j^M(\cdot)$ of simulators f_j for coordinates $j \in b, \dots, d$ is of interest. Let first $b - 1$ coordinates are not related to the linking of emulators.

Let a multivariate GaSP for coordinates $j \in b, \dots, d$, a GaSP emulator $f_{b:d}^M(\cdot)$ of simulators f_j

$$(SM1) \quad f_{b:d}^M(\cdot) \sim \text{GaSP}(\mu_{f_{b:d}}^*(\cdot), \Sigma(\cdot, \cdot)).$$

This is such that for any new input $\mathbf{u}^{\mathbf{b:d}}$ to $f_{b:d}^M(\mathbf{u}^{\mathbf{b:d}})$

$$f_{b:d}^M(\mathbf{u}^{\mathbf{b:d}}) \sim N(\mu_{f_{b:d}}^*(\mathbf{u}^{\mathbf{b:d}}), \Sigma(\mathbf{u}^{\mathbf{b:d}}, \mathbf{u}^{\mathbf{b:d}})),$$

where $\mu_{f_{b:d}}^*(\mathbf{u}^{\mathbf{b:d}}) = (\mu_{f_b}^*(\mathbf{u}^{\mathbf{b}}), \dots, \mu_{f_d}^*(\mathbf{u}^{\mathbf{d}}))$ is a vector of means. $\Sigma(\mathbf{u}^{\mathbf{b:d}}, \mathbf{u}^{\mathbf{b:d}})$ is such that $\text{diag}(\Sigma(\mathbf{u}^{\mathbf{b:d}}, \mathbf{u}^{\mathbf{b:d}})) = (\sigma_{f_b}^{*2}(\mathbf{u}^{\mathbf{b}}), \dots, \sigma_{f_d}^{*2}(\mathbf{u}^{\mathbf{d}})) = \sigma_{f_{b:d}}^{*2}(\mathbf{u}^{\mathbf{b:d}})$.

Theorem SM2.1. Let g^M with given parameters $\theta_{\mathbf{g}} = (\beta, \sigma^2, \{\delta_j\}_{j=1, \dots, d}, \eta)$ be a type 1 GaSP emulator of a simulator g exercised at training input points \mathbf{z} , i.e. $g^M(\cdot)$ as in (4). Suppose the mean is linear in the b th coordinate of an input \mathbf{z}^l , so that the mean is $\mathbf{h}(\mathbf{z}^l)\beta = \beta_0 + \beta_1 z_b^l$. Let the $g^M(\cdot)$ GaSP correlation function smoothness parameters α_j of coordinates $j \in b, \dots, d$ be equal to 2. For $j \in b, \dots, d$ let $f_{b:d}^M$ be a multivariate emulator of a multivariate simulator $f_{b:d}$, corresponding to coordinates $j = b, \dots, d$ of the input to the simulator g , i.e.

Table SM1

Range of ratios of twice the inverse of squared range parameters of inputs to model puff to the marginal variances of corresponding emulators of bent outputs.

Ratio	plumeMax	plumeMin	plumeHwidth	AshLogMean	AshLogSdev
Min	1563	2533	3380	8436	2776
Max	3502	7892	11844	4413190	1851116

$f_{b:d}^M(\cdot)$ as in (SM1). Then the mean $E\xi$ and variance $V\xi$ of the linked emulator ξ of the coupled simulator $(g \circ (f_{b:d}))(\mathbf{u})$ are

$$\begin{aligned}
 \text{(SM2)} \quad E\xi &= \beta_0 + \beta_1 \mu_{f_b}^*(\mathbf{u}^b) + \sum_{i=1}^m a_i \prod_{j=1}^{b-1} \exp\left(-\left(\frac{|u_j - z_{ij}|}{\delta_j}\right)^{\alpha_j}\right) \\
 &\quad (2\pi)^{\frac{d-b+1}{2}} \sqrt{\det M} N_{z_{i b:d}}(\mu_{f_{b:d}}^*(\mathbf{u}^{b:d}), M + \Sigma(\mathbf{u}^{b:d}, \mathbf{u}^{b:d})), \\
 \text{(SM3)} \quad V\xi &= \sigma^2(1 + \eta) + \beta_0^2 + 2\beta_0\beta_1\mu_{f_b}^*(\mathbf{u}^b) + \beta_1^2(\sigma_{f_b}^{*2}(\mathbf{u}^b) + (\mu_{f_b}^*(\mathbf{u}^b))^2) - (E\xi)^2 + \\
 &\quad \left((2\pi)^{d-b+1} \det M \sum_{k,l=1}^m (a_l a_k - \sigma^2 \{C_z^{-1}\}_{k,l}) \prod_{j=1}^{b-1} e^{-\left(\left(\frac{|u_j - z_{kj}|}{\delta_j}\right)^{\alpha_j} + \left(\frac{|u_j - z_{lj}|}{\delta_j}\right)^{\alpha_j}\right)} \right) \\
 &\quad N_{z_{k b:d}}(z_{l b:d}, 2M) N_{\mu_{f_{b:d}}^*}(\mathbf{u}^{b:d}) \left(\frac{z_{k b:d} + z_{l b:d}}{2}, \frac{M}{2} + \Sigma(\mathbf{u}^{b:d}, \mathbf{u}^{b:d}) \right) + \\
 &\quad 2(2\pi)^{\frac{d-b+1}{2}} \sqrt{\det M} \sum_{i=1}^m a_i \prod_{j=1}^{b-1} \exp\left(-\left(\frac{|u_j - z_{ij}|}{\delta_j}\right)^{\alpha_j}\right) (\beta_0 + \beta_1 m_1^i) \\
 &\quad N_{z_{i b:d}}(\mu_{f_{b:d}}^*(\mathbf{u}^{b:d}), M + \Sigma(\mathbf{u}^{b:d}, \mathbf{u}^{b:d})),
 \end{aligned}$$

where matrix $M = 2 \cdot \text{diag}(\delta_b^{-2}, \dots, \delta_d^{-2})$ and in general $N_x(\mu, \Sigma)$ denotes a normal density of vector x with mean μ and covariance Σ , $a = (a_1, \dots, a_m)^T = C_z^{-1}(\mathbf{g}^M(\mathbf{z}) - \mathbf{h}(\mathbf{z})\beta)$; $V = (\Sigma(\mathbf{u}^{b:d}, \mathbf{u}^{b:d})^{-1} + M^{-1})^{-1}$ and m_1^i is the first element of the vector $m^i = V(\Sigma(\mathbf{u}^{b:d}, \mathbf{u}^{b:d})^{-1} \mu_{f_{b:d}}^*(\mathbf{u}^{b:d}) + M^{-1} z_{i b:d})$.

SM2.1. Analysis on multivariate linking of bent and puff. If $\Sigma(\cdot, \cdot)$ produces a diagonal matrix, multivariate model reduces to an independent scenario, keeping the same marginal variances. Analysis of expressions for the mean shows that the difference in dependent and independent cases is in the following set of distributions $N_{z_{i b:d}}(s_{b:d}, M + \Sigma)$ (dependent) and $N_{z_{i b:d}}(s_{b:d}, M + \text{diag}(\Sigma))$ (independent), where $\text{diag}(\Sigma)$ is the diagonal matrix of Σ .

In bent and puff the ratios of $\frac{\text{diag}(M)}{\text{diag}\Sigma}$, i.e. ratios of twice the inverse of squared range parameters of model puff to the marginal variances of corresponding bent emulators, where $\text{diag}\Sigma$ are variances of predictive distributions obtained from independent bent GaSPs of outputs plumeMax, plumeMin, plumeHwidth, AshLogMean and AshLogSdev, are summarized in Table SM1. For all of the 1000 test points, the described ratio is calculated and the range of corresponding ratios are presented.

According to the expressions for the linked mean (SM2) and linked variance (SM3), Σ has

not much weight compared to M . The expressions would be “disturbed” very little through changes in the densities of corresponding distributions. This provides insight on possible performance of the multivariate linking compared to the independent scenario. Numerical differences in the multivariate and independent linking of bent and puff should be expected negligible.