Statistica Sinica Preprint No: SS-2016-0217R1	
Title	Computer Experiments: Prediction Accuracy, Sample Size and Model Complexity Revisited
Manuscript ID	SS-2016-0217R1
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202016.0217
Complete List of Authors	Ofir Harari Derek Bingham Angela Dean and David Higdon
Corresponding Author	Ofir Harari
E-mail	oharari@sfu.ca

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0217

Proof. First note that, from (1)

$$\operatorname{cov}\left(y\left(\boldsymbol{x}\right),\widehat{y}\left(\boldsymbol{x}\right)\right)=\boldsymbol{r}\left(\boldsymbol{x}\right)^{\mathsf{T}}\boldsymbol{R}^{-1}\operatorname{cov}\left(y\left(\boldsymbol{x}\right),\boldsymbol{y}\right)=\sigma^{2}\boldsymbol{r}\left(\boldsymbol{x}\right)^{\mathsf{T}}\boldsymbol{R}^{-1}\boldsymbol{r}\left(\boldsymbol{x}\right)=\mathbb{V}\mathrm{ar}\left\{\widehat{y}\left(\boldsymbol{x}\right)\right\}$$

and, since $\mathbb{V}\mathrm{ar}\left\{y\left(\boldsymbol{x}\right)\right\}=\sigma^{2}$, we have

$$\rho^{2}\left(y\left(\boldsymbol{x}\right),\widehat{y}\left(\boldsymbol{x}\right)\right)=\frac{\left\{\sigma^{2}\boldsymbol{r}\left(\boldsymbol{x}\right)^{\mathsf{T}}\boldsymbol{R}^{-1}\boldsymbol{r}\left(\boldsymbol{x}\right)\right\}^{2}}{\sigma^{2}\cdot\sigma^{2}\boldsymbol{r}\left(\boldsymbol{x}\right)^{\mathsf{T}}\boldsymbol{R}^{-1}\boldsymbol{r}\left(\boldsymbol{x}\right)}=\boldsymbol{r}\left(\boldsymbol{x}\right)^{\mathsf{T}}\boldsymbol{R}^{-1}\boldsymbol{r}\left(\boldsymbol{x}\right).$$

The result follows from (2) and (4). □

We can interpret $\overline{\rho}^{2}\left(y,\widehat{y}\right)$ in (5) as the average squ[illegible] the simulator responses and predicted responses at unsampled inputs. Proposition 1 [illegible]ds new light on the interpretation of IMSPE-optimal designs (see e.g. Sacks e[illegible]). The proposition demonstrates that minimizing the IMSPE is equi[illegible]lent to maximizing the average, squared, out-of-sample correlation betwe[illegible]n $y\left(\boldsymbol{x}\right)$ and $\widehat{y}\left(\boldsymbol{x}\right)$. E[illegible]mizing the IMSPE we can expect to improve the predictive ability of the Kr[illegible]ng pr[illegible]or.

Definition 1.

The *Root Average Unexp[illegible]d [illegible]ri[illegible]ty* (RAUV) of predictor $\widehat{y}$, evaluated at design $\mathcal{D}\subset$ [illegible]

$$\left.\boldsymbol{\theta}\right)=\left(\frac{\mathcal{J}\left(\widehat{y},\mathcal{D};\boldsymbol{\theta}\right)}{\sigma^{2}}\right)^{1/2}=\left(\int_{[0,1]^{d}}\frac{\mathbb{V}\mathrm{ar}\left\{y\left(\boldsymbol{x}\right)\left|\mathcal{D}\right.\right\}}{\mathbb{V}\mathrm{ar}\left\{y\left(\boldsymbol{x}\right)\right\}}\mathrm{d}\boldsymbol{x}\right)^{1/2}.$$

We propose RAUV as a measure of expected prediction error on designing a computer experiment. Its indicated magnitude is relative to the prior standard deviation, and the choice of the square root scale is in line with similar measures used by Loeppky et al. (2009) and Chen et al. (2016). It is common practice to measure model performance by its