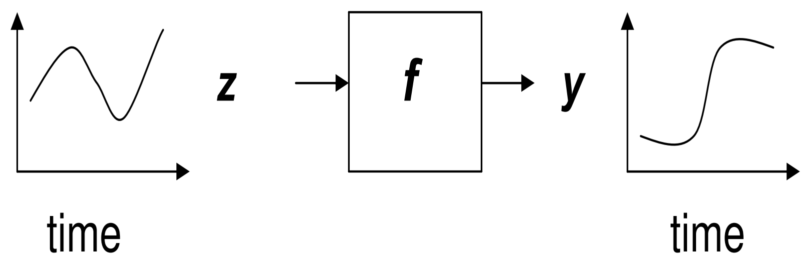
Computer Experiments with Time-Varying Inputs: Gaussian Surrogates and Experimental Designs

Max D. Morris
Department of Statistics
Iowa State University

MASCOT-NUM, St-Etienne

Setting



Examples

- Population growth/diversity as a function of resources
- Material fatigue as a function of stress
- Global climate as a function of greenhouse gas emission

Background & Notation

- Deterministic computer models
- For scalar-valued output and vector-valued input:

$$y_{\mathbf{x}} = f(\mathbf{x}), \quad \mathbf{x} \in \Delta$$

- "*Meta-model*" or "*Surrogate*" based on a prior (pre-data) Gaussian Stochastic Process (GaSP) indexed by input:

$$E(y_{\mathbf{x}}) = \mu \qquad \mathsf{Var}(y_{\mathbf{x}}) = \sigma^2$$

$$\mathsf{Corr}(y_{\mathbf{x}_1}, y_{\mathbf{x}_2}) = e^{-\theta \times D(\mathbf{x}_1, \mathbf{x}_2; \mathbf{w})} = e^{-\theta \sum_i w_i \times d(\mathbf{x}_1^i, \mathbf{x}_2^i)}$$

- View D as a *weighted distance* between $\mathbf{x}$'s; positive correlation decreases as distance increases.

- For:
  - an experimental design: $X = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, ..., \mathbf{x}_N\}$
  - resulting data (outputs): $\mathbf{y}$
  - specified μ, σ^2, θ
- output prediction at $\mathbf{x}_0$ proceeds via the *conditional* GaSP as:

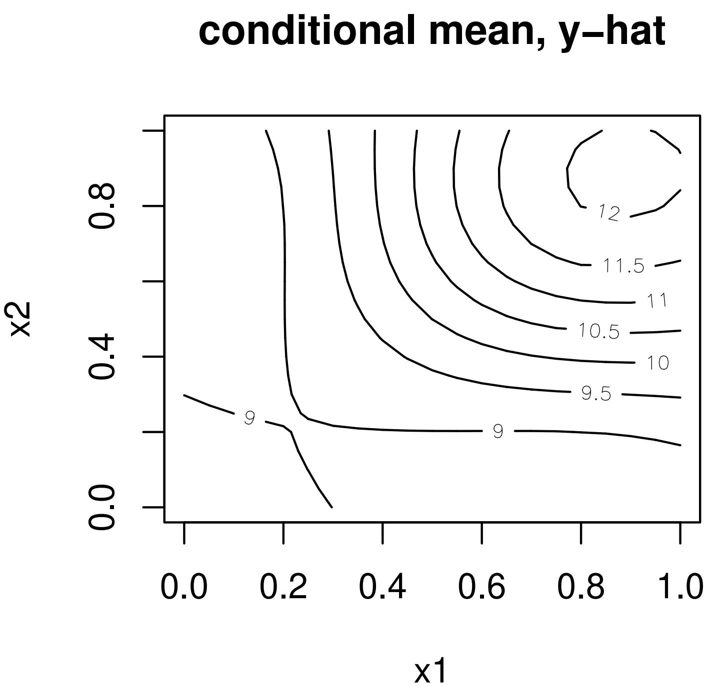
$$\hat{y}_{\mathbf{x}_0} = E(y_{\mathbf{x}_0}|\mathbf{y}) = \mu + \mathbf{r}'_{0X}\mathbf{R}^{-1}_{XX}(\mathbf{y} - \mu\mathbf{1})$$
$$se(\hat{y}_{\mathbf{x}_0}) = \sqrt{\mathsf{Var}(y_{\mathbf{x}_0}|\mathbf{y})} = \sqrt{\sigma^2(1 - \mathbf{r}'_{0X}\mathbf{R}^{-1}_{XX}\mathbf{r}_{0X})}$$

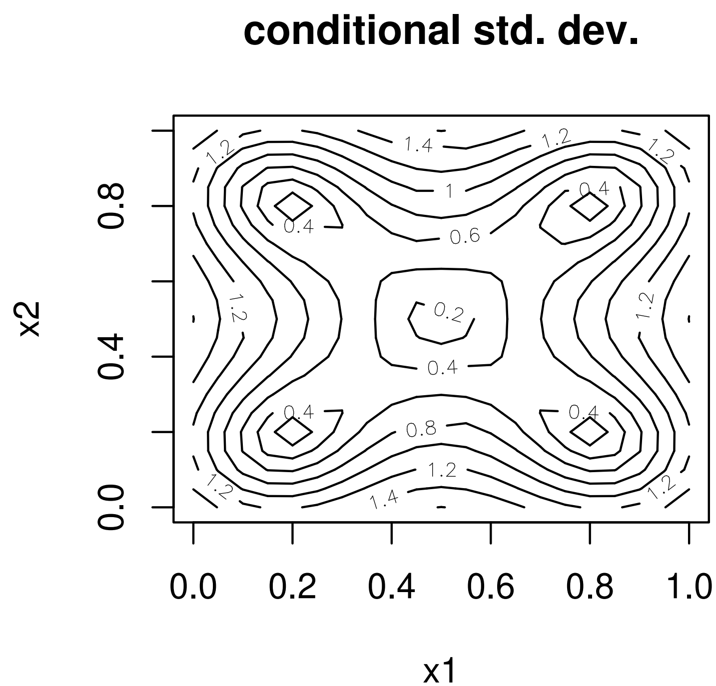
  where $\{\mathbf{r}_{0X}\}_i = \mathsf{Corr}(y_{\mathbf{x}_0}, y_{\mathbf{x}_i})$, and $\{\mathbf{R}_{XX}\}_{ij} = \mathsf{Corr}(y_{\mathbf{x}_i}, y_{\mathbf{x}_j})$
- e.g. Sacks et al. (1989), Currin et al. (1991), Santner et al. (2003).

- Example:

(x^1, x^2)	$(.2, .2)$	$(.2, .8)$	$(.8, .2)$	$(.8, .8)$	$(.5, .5)$
y	9.0	9.0	9.0	12.0	10.0

  - $\mu = 10$, $\sigma^2 = 3$
  - $D(\mathbf{x}_1, \mathbf{x}_2; \mathbf{w}) = \sum_{i=1}^{2} w_i (x_1^i - x_2^i)^2$, $\theta = 1$, $w_1 = w_2 = 1$:





Vector Inputs & Functional Outputs

- Now

$$y_{\mathbf{x}}(t) = f(\mathbf{x}, t), \quad \mathbf{x} \in \Delta, \quad t \in [0, T]$$

- As yesterday, to facilitate things, define a time-grid:

$$G = \{\tau_1, \tau_2, \tau_3, ..., \tau_M\}, \qquad 0 < \tau_1 < \tau_2 < \tau_3 < ... < \tau_M \leq T$$

$$\mathbf{y}_{\mathbf{x}} = f(\mathbf{x}), \quad \mathbf{x} \in \Delta$$

- GaSP: If we restrict the structure to be the same at each $\mathbf{x}$:

$$E(\mathbf{y}_{\mathbf{x}}) = \boldsymbol{\mu} \qquad \text{Var}(\mathbf{y}_{\mathbf{x}}) = \boldsymbol{\Sigma}$$

- Conte and O'Hagan (2011) discuss two approaches to modeling covariances across $\mathbf{x}$-space:

1.) "Multivariate Output" (or MO)

- Cov$(\mathbf{y}_{\mathbf{x}_i}, \mathbf{y}_{\mathbf{x}_j}) = e^{-\theta \times D(\mathbf{x}_1, \mathbf{x}_2; \mathbf{w})} \times \mathbf{\Sigma}$.
- This treats the covariance as *separable*, factoring it into components associated with differences between $\mathbf{x}$ vectors, and output components.
- C & O'H discuss a special case of this, "Time Index" (or TI) that adds structure suggested by outputs that are continuous functions of time:

$$\{\mathbf{\Sigma}\}_{i,j} = \sigma^2 e^{-\phi \times d(t_i, t_j)}$$

- Implications:
  - At any $\mathbf{x}$ and t, the correlation between $y_{\mathbf{x}}(t)$ and $y_{\mathbf{x}}(t+\delta)$ is the same for any fixed δ
  - At any t, the correlation between $y_{\mathbf{x}_i}(t)$ and $y_{\mathbf{x}_j}(t)$ is the same

2.) "Many Single-output ... " (or MS)

- Cov$(\{\mathbf{y}_{\mathbf{x}_i}\}_r, \{\mathbf{y}_{\mathbf{x}_j}\}_s) = \begin{cases} \sigma^2 e^{-\theta \times D(\mathbf{x}_i, \mathbf{x}_j; \mathbf{w}_r)} & r = s \\ 0 & \text{otherwise} \end{cases}$
- Implications:
  - At any $\mathbf{x}$ and t, the correlation between $y_{\mathbf{x}}(t)$ and $y_{\mathbf{x}}(t+\delta)$ is zero for any $\delta \neq 0$ (much stronger assumption than MO/TI)
  - The correlation between $y_{\mathbf{x}_i}(t)$ and $y_{\mathbf{x}_j}(t)$ can be different at different t (weaker assumption than MO/TI)
- In the form given here, TI has only one more parameter than MS.
- Using M output values for each of N model runs, the computational effort for parameter estimation is driven by the order of the correlation matrix:
  - TI: One unified model, kronecker-factors of order M and N
  - MS: M independent models, each of order N

Functional Inputs & Outputs

- Morris (2012), a further development of the MS idea.
- Input function over time:

$$z(t), t \in [0, 1]$$

- Output also a function of time, with y^τ potentially influenced by $z(t)$ with $t \leq \tau$:

$$y_z^\tau = f(z(t), t \in [0, \tau]) \qquad \tau \in [0, T]$$

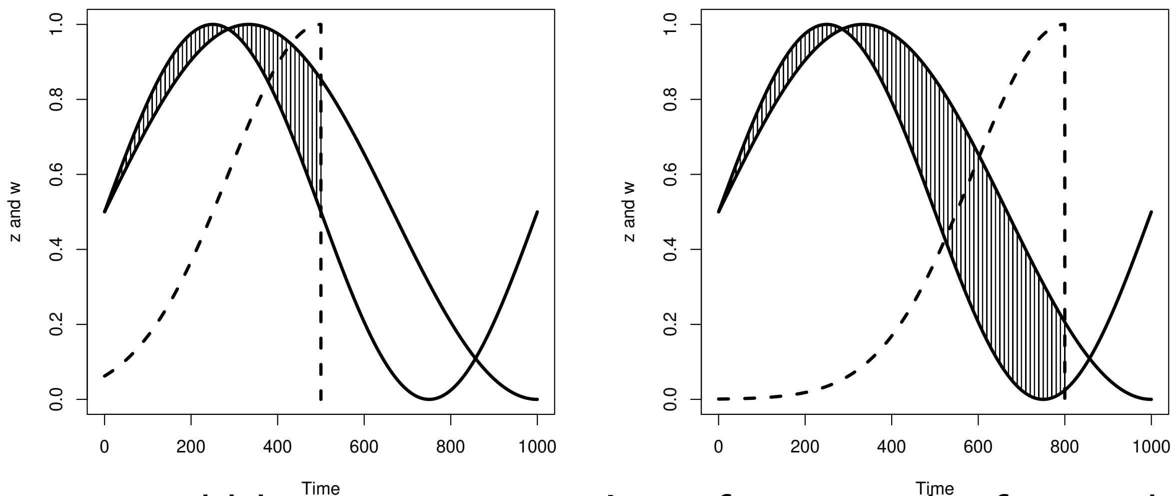
- GaSP:

$$E(y_z^\tau) = \mu(\tau) \qquad \mathsf{Var}(y_z^\tau) = \sigma^2(\tau)$$

$$\begin{aligned} \mathsf{Corr}(y_{z_1}^\tau, y_{z_2}^\tau) &= \mathsf{exp}\{-\theta \int_0^\tau \, w_\tau(\tau - t) \times d(z_1(t), z_2(t)) dt\} \\ &= \mathsf{exp}\{-\theta \times D(z_1, z_2; w_\tau)\} \end{aligned}$$

- Integral generalizes sum in product correlation for vector-valued $\mathbf{x}$; now a weighted distance between functions over $[0, \tau]$.

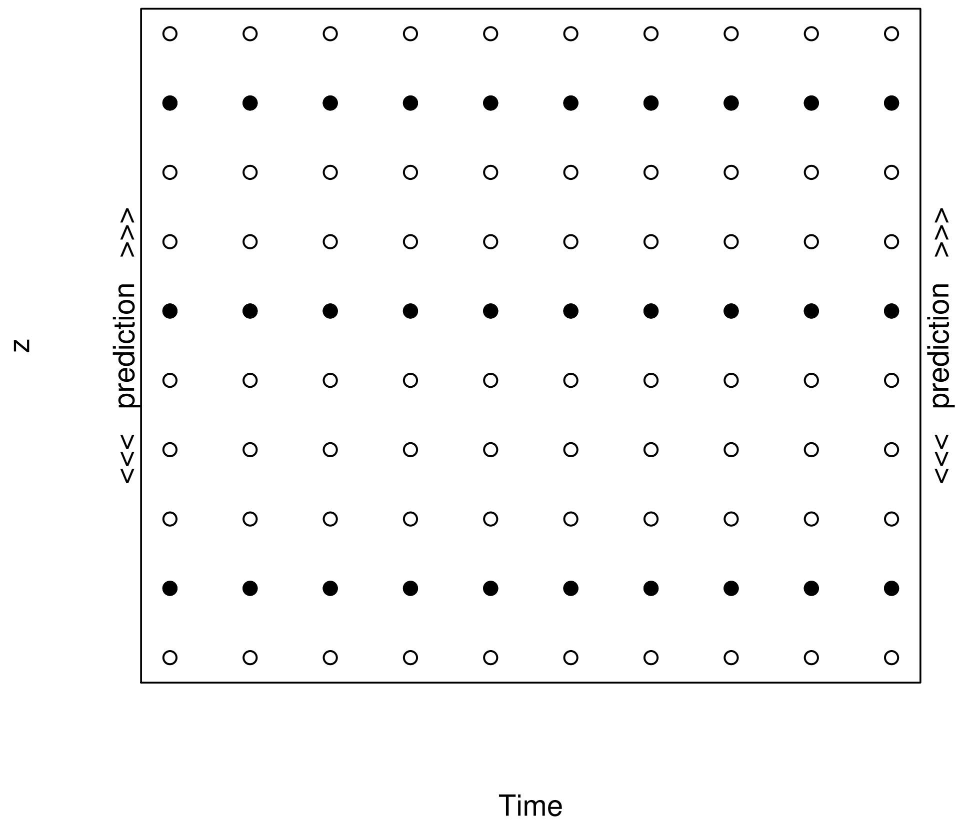
- Here, I'm using $w_\tau(\tau - t) = exp\{-\beta(\tau - t)^2\}$, suggesting a belief that at any time, output is most sensitive to "recent" values of the input function.



- Other forms would be more appropriate, for example, for models in which early inputs are most critical, and the system "solidifies" over time to be less influenced by z (e.g. some chemical reactions).

- In any case, w_τ must be non-zero over $[0, \tau]$ to guarantee non-zero distance between distinct z_1 and z_2.

- As with MS, model $(y_{z_1}^{\tau_1}, y_{z_2}^{\tau_2})$ with $\tau_1 \neq \tau_2$ as independent.



Inference

- Define a time grid for output modeling and prediction:

$$G = \{\tau_1, \tau_2, \tau_3, ..., \tau_M\}, \qquad 0 < \tau_1 < \tau_2 < \tau_3 < ... < \tau_M \le T$$

- Experimental design:

$$Z = \{z_1, z_2, z_3, ..., z_N\}$$

- Resulting data:

$$\begin{array}{llll} \mathbf{y}_1 & \mathbf{y}_2 & ... & \mathbf{y}_N \quad \leftarrow \text{organized by } Z \\ \mathbf{y}^1 & \mathbf{y}^2 & ... & \mathbf{y}^M \quad \leftarrow \text{organized by } G \end{array}$$

- Log likelihood $\propto$:

$$-\sum_{m=1}^{M} \left\{ N \times \ln(\sigma^2(\tau_m)) + N \times \ln(|\mathbf{R}_m|) \right.$$
$$\left. + (\mathbf{y}^m - \mu(\tau_m)\mathbf{1})' \mathbf{R}_m^{-1} (\mathbf{y}^m - \mu(\tau_m)\mathbf{1}) / \sigma^2(\tau_m) \right\}$$

where $\{\mathbf{R}_m\}_{ij} = \exp\{-\theta \times D(z_i, z_j; w_{\tau_m})\}$

- Parameters: θ, and

$$\mu(-) \qquad \sigma^2(-) \qquad w_\tau(-)$$

each over $[0, T]$, assigned a reasonable parametric form.

- For known parameters, output prediction for input z_0 at time τ_m is:

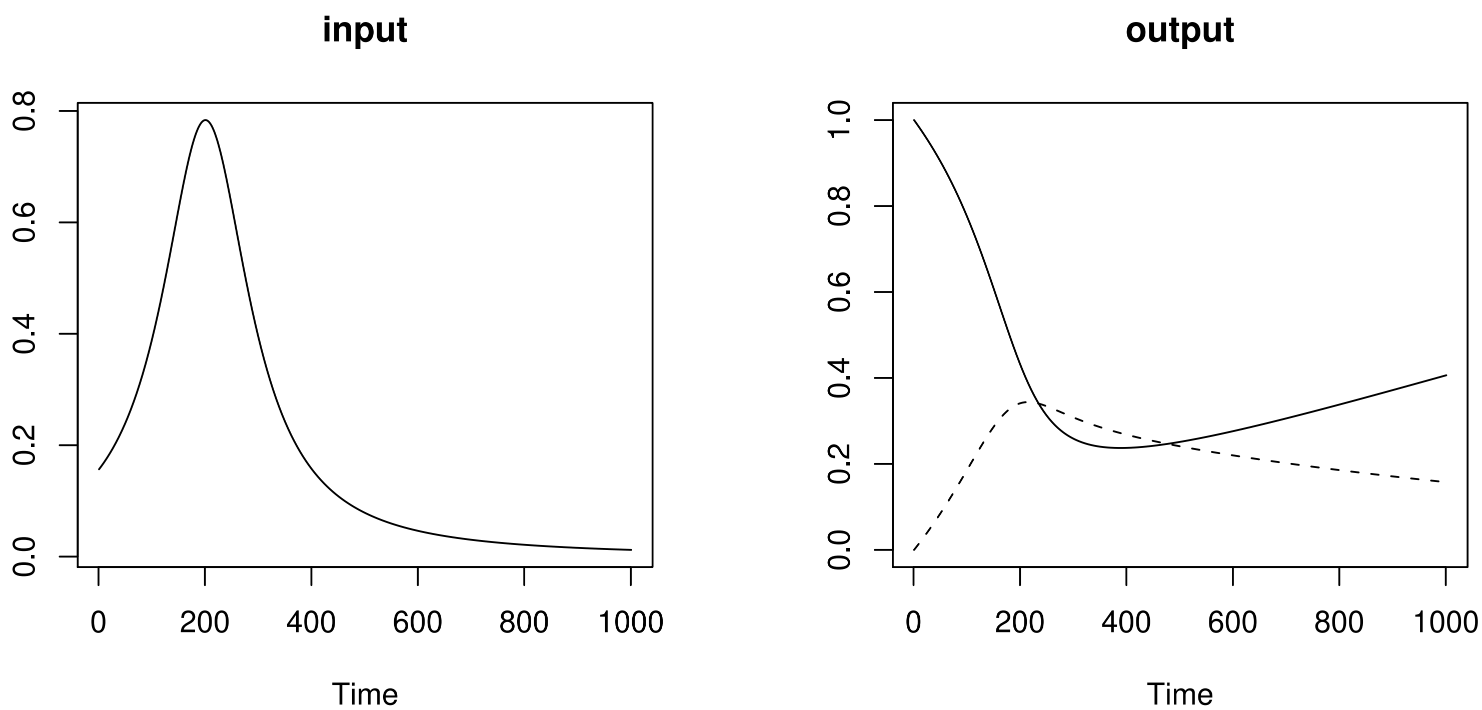
$$E(y_{\chi_0}^{\tau_m}|\mathbf{y}) = \mu(\tau_m) + \mathbf{r}'_{0,m}\mathbf{R}_m^{-1}(\mathbf{y}^m - \mu(\tau_m)\mathbf{1})$$
$$\mathsf{Var}(y_{\chi_0}^{\tau_m}|\mathbf{y}) = \sigma^2(\tau_m)[1 - \mathbf{r}'_{0,m}\mathbf{R}_m^{-1}\mathbf{r}_{0,m}]$$

where $\{\mathbf{r}_{0,m}\}_i = \mathsf{exp}\{-\theta \times D(z_0, z_i; w_{\tau_m})\}$

- For unknown parameters:
  - empirical Bayes: Estimate from data (typically via maximum likelihood) and treat as known
  - full Bayes: Assign priors, incorporate parameter uncertainty

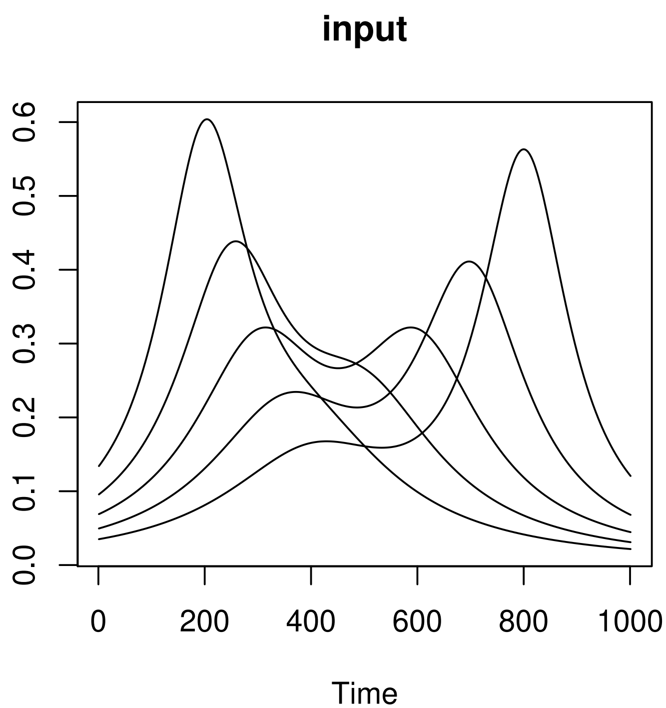
Example: A "Small" Model

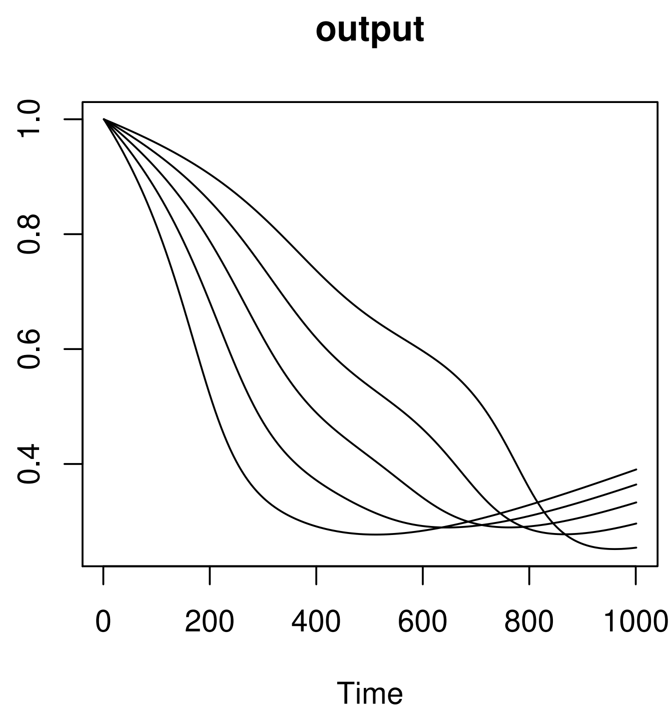
- Model of marrow stem-cells, Jones, Morris & Young (1991):
  - input = time-rate of ionizing radiation exposure
  - output = quantity of normal, injured, and killed cells as functions of time, $t \in [0, 1000]$



Example: Experiment

- $N = 5$ runs of the model and resulting output (normal cells):





- Output prediction at $G = \{400, 450, 500, ..., 1000\}$, with

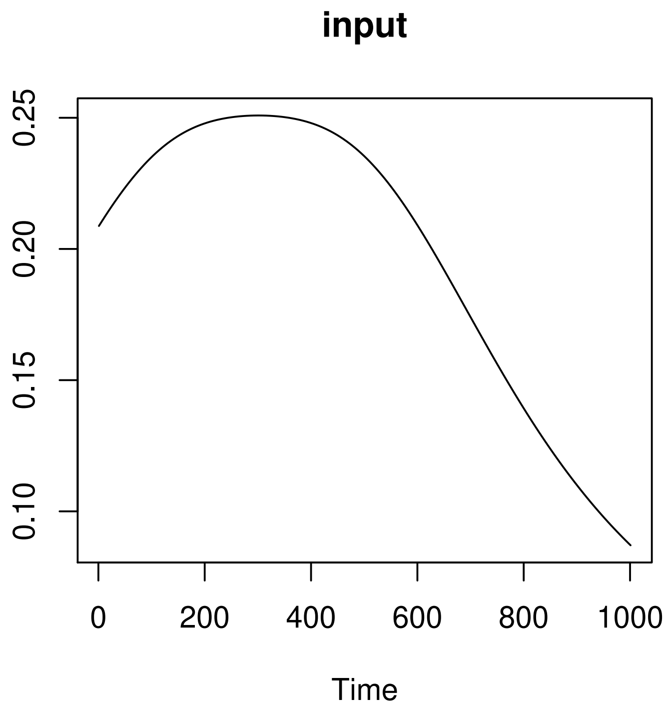
$$w_\tau(\tau - t) = exp\{-\beta(\tau - t)^2\}$$

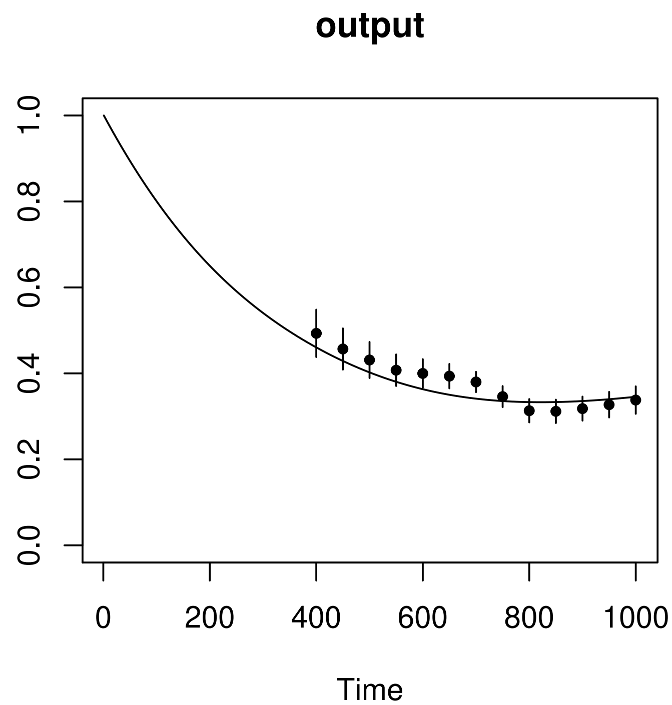
- "Gaussian" correlation form (i.e. weighted L_2 distance between z's):

$$D(z_i, z_j; w_\tau)\} = \int_0^\tau w_\tau(\tau - t)(z_i(t) - z_j(t))^2 dt$$

- Bayesian prediction of y at times in G, using independent priors:
  - $\theta \sim$ Gamma(mean=std.dev.=0.02)
  - $\beta \sim$ Gamma(mean=std.dev.=0.02)
  - at each $\tau \in G$ independently, μ uniform over $(-\infty, \infty)$
  - common σ^2 for all $\tau \in G$, with density inversely proportional to its value

- Predict output for:





Experimental Design

- Select Z so that Var$(y^{\tau}_{z_0}|\mathbf{y})$ is small for all $\tau \in G$ and all z_0 of interest.
- *Predictive D-optimality/Entropy optimality* minimizes a summary measure of this across all $z(t) \notin Z$.
- Johnson, Moore, Ylvisaker (1990) showed that for vector-valued inputs $\mathbf{x}$, as correlations become weak (θ large), *maximin distance designs* are optimal in this sense:

$$\text{Pick } X \text{ to maximize: } \quad \phi = min_{\mathbf{x}_i, \mathbf{x}_j \in X} \;\; D(\mathbf{x}_i, \mathbf{x}_j; \mathbf{w})$$

- In our case, if $\sigma^2(\tau_m) = \sigma^2$, generalization leads to:

$$\text{Pick } Z \text{ to maximize: } \quad \phi = min_{z_i, z_j \in Z} min_{\tau \in G} \;\; D(z_i, z_j; w_\tau)$$

Example: Rerun with Optimal Design

- Input functions of interest: $z^*(t) = \frac{r_1}{s_1^2+(t-t_1)^2} + \frac{r_2}{s_2^2+(t-t_2)^2}$

$$r_1, r_2 = 1, 2, 5$$

$$s_1, s_2 = 100, 200, 500, 1000, 2000, 5000$$

$$t_1, t_2 = 200, 300, 400, ..., 800$$

  each normalized to total dose of 200:

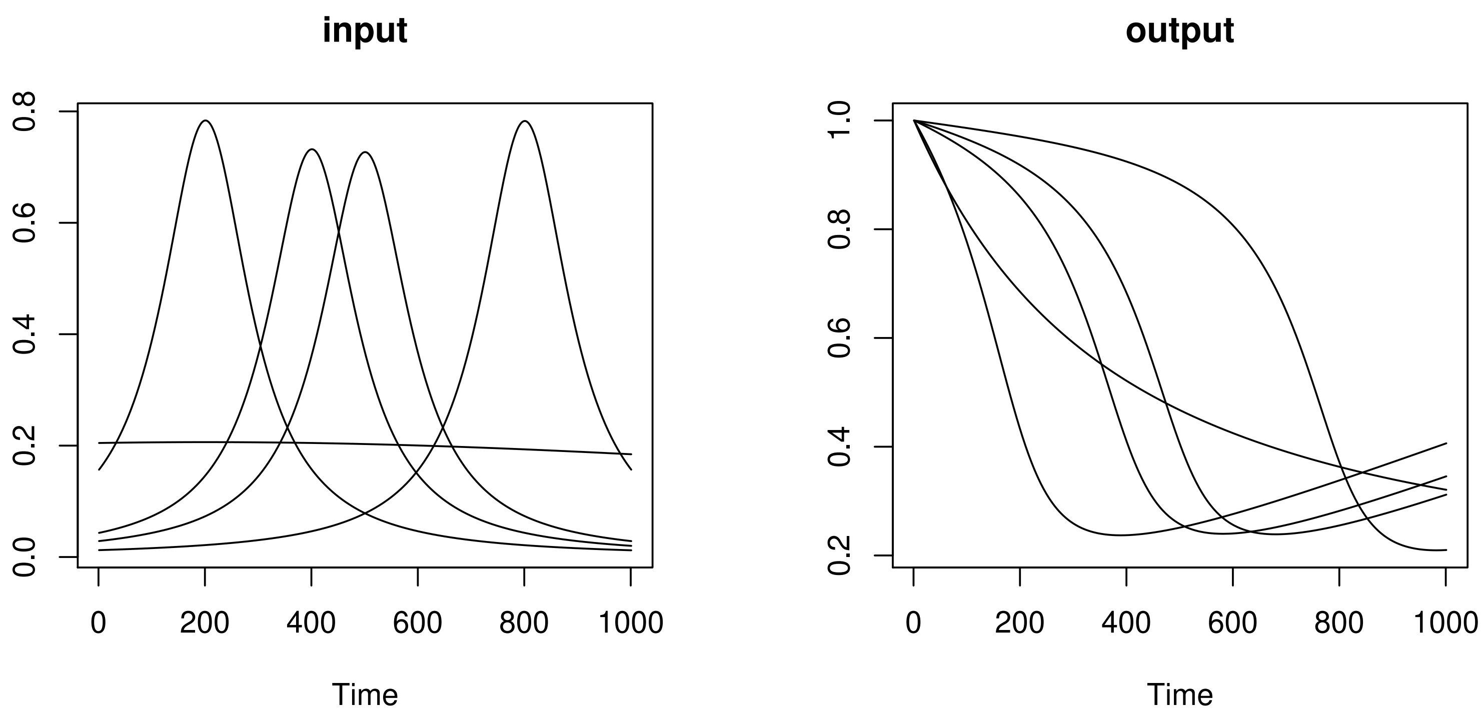
$$z(t) = 200 \times z^*(t) / \int_0^{1000} z^*(u)du$$

- Exposure received along a linear path passing within distances s_1 and s_2, at times t_1 and t_2, of two point sources of relative strength r_1 and r_2.

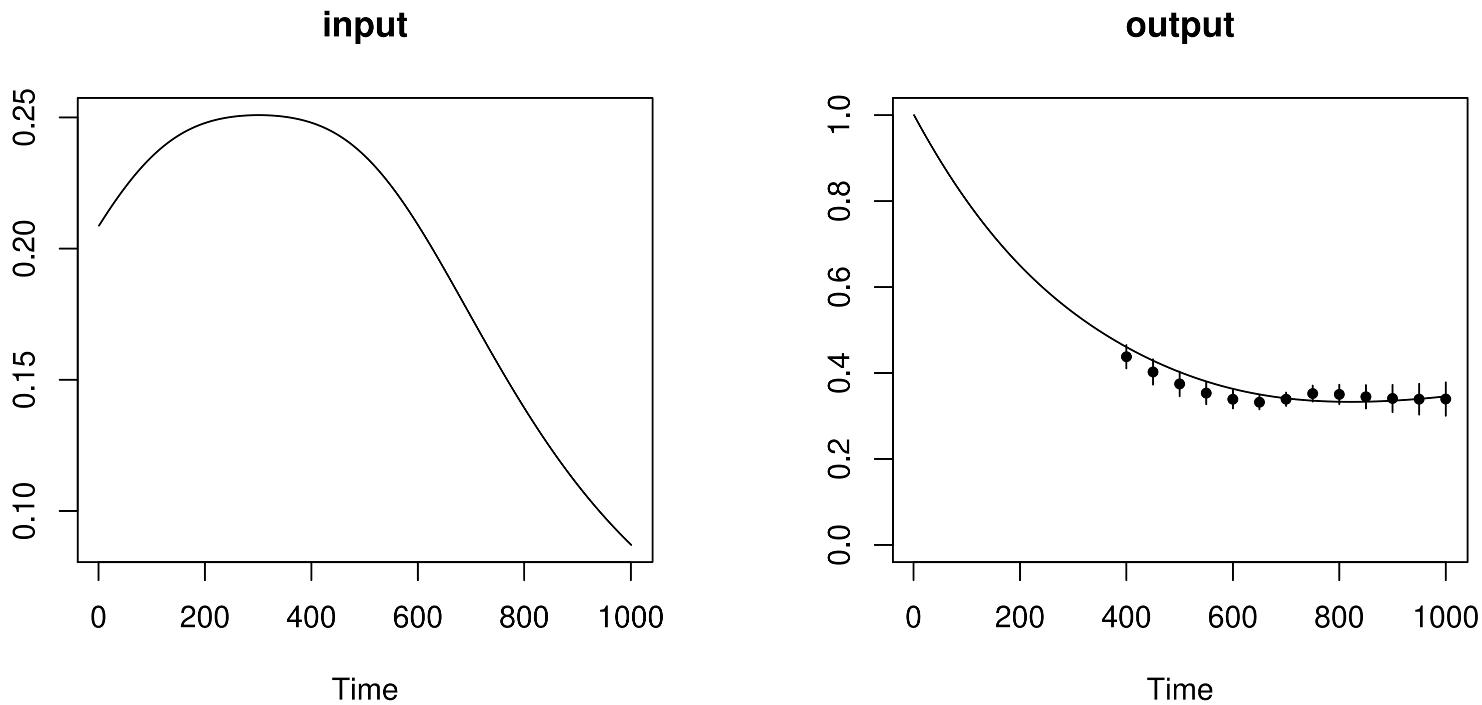
- 9072 $z(t)$'s.

- Construction algorithm: Repeated "backward elimination," from an initial random sample, of z's that are closest to others.

- $N = 5$ runs of the model and resulting output:



- Predictions:

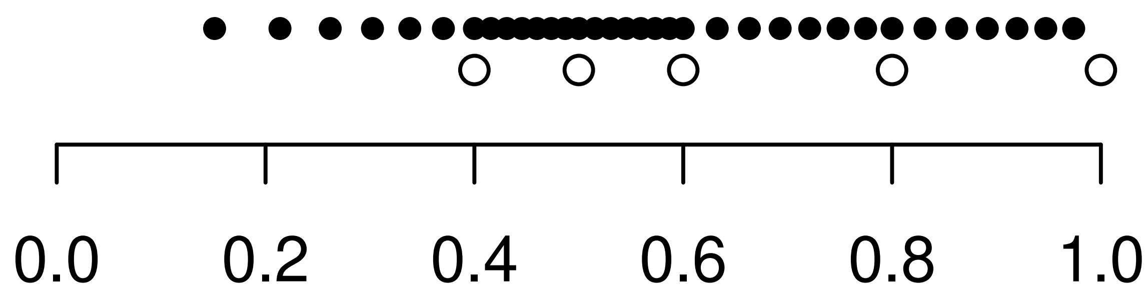


Maximin Distance-Optimal Designs

- Morris (2014)
- $0 < z(t) < 1$
- $t \in [0, 1]$
- For all $\tau \in G$,
  - $w_\tau(\tau - t) > 0$, $\int_0^\tau w_\tau(\tau - t)dt = 1$
  - $D(z_i, z_j; w_\tau) = \int_0^\tau w_\tau(\tau - t)(z_i(t) - z_j(t))^2 dt$
- **Theorem:**
  1. $\underline{N = 2:}$ maximum $\phi = 1$
  2. $\underline{N = 0 \text{ mod } 4:}$ maximum $\phi = \frac{1}{2} \frac{N}{N-1}$
- Proof is by construction, and requires $z(t)$ to jump between 0 and 1 $O(N \times M)$ times! (So the main practical value of this result is the bound, not the construction)

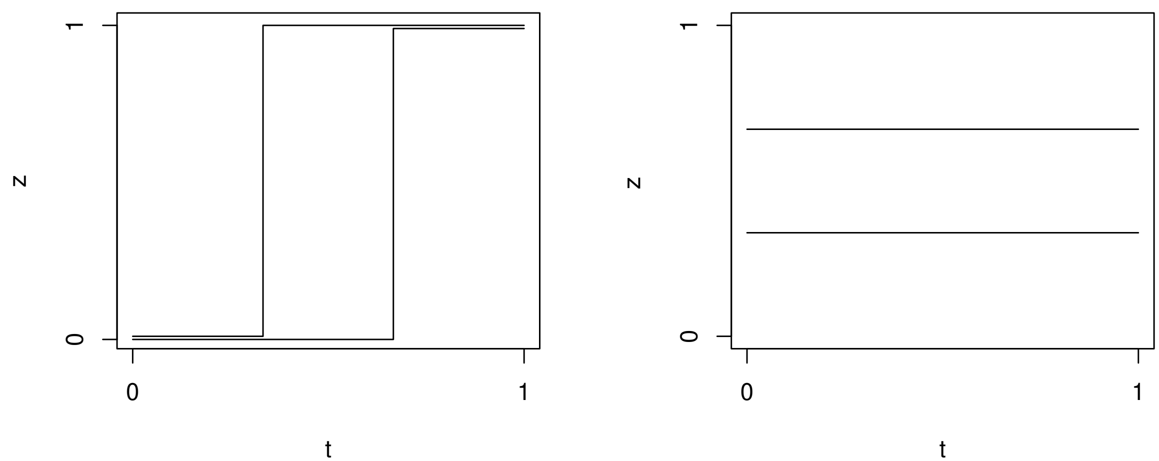
Example

- $G = \{0.4, 0.5, 0.6, 0.8, 1\}$
- $w_\tau(\tau - t) = 2t/\tau^2$
- $N = 8$
- $z_i(t)$ values determined, for example, by regular 2^{7-4} fractional factorial design, with "change points" at:



Concoluding Remarks

- In practice, other distance measure may be more appropriate:



- Still, "distance based" design ideas popular with GaSP models *can* be used.

- The approach easily generalizes to
  - multiple time-series inputs, or mixed time-function and scalar inputs
  - functions of both time and space ...

References:

- Currin, C., T. Mitchell, M. Morris, and D. Ylvisaker, "Bayesian Prediction of Deterministic Functions ...," *JASA* **86**, 1991, 953-963.
- Johnson, M.E., L.M. Moore, and D. Ylvisaker, "Minimax and Maximin Distance Designs," *JSPI* **26**, 1990, 131-148.
- Jones, T.D., M.D. Morris, and R.W. Young, "A Mathematical Model for Radiation-Induced Myelopoiesis," *Radiation Research* **128**, 1991, 258-266.
- Morris, M.D., "Gaussian Surrogates for Computer Models with Time-Varying Inputs and Outpust," *Technometrics* **54**, 2012, 42-50.
- Morris, M.D., "Maximin Distance Optimal Designs for Computer Experiments with Time-Varying Inputs and Outputs," *Journal of Statistical Planning and Inference* **144**, 2014, 63-68.
- Sacks, J., W. Welch, T. Mitchell, and H. Wynn, "Design and Analysis of Computer Experiments," *Stat. Science* **4**, 1989, 409-423.
- Santner, T.J., B.J. Williams, and W.I. Notz, *The Design and Analysis of Computer Experiments*, 2003, Springer.

For N = even, $G = \{\tau_1\}$:

- Find $0 = t_1^0 < t_1^2 < t_1^3 < ... < t_1^{n-2} < t_1^{n-1} = \tau_1$ that evenly divide the integral of w_{τ_1}:

$$\int_0^{t_1^1} w_{\tau_1}(\tau_1 - t)dt = \int_{t_1^1}^{t_1^2} w_{\tau_1}(\tau_1 - t)dt = ... = \int_{t_1^{n-2}}^{\tau_1} w_{\tau_1}(\tau_1 - t)dt = \frac{1}{n-1}$$

- Z such that within each of $[0, t_1^1)$, $[t_1^1, t_1^2]$, ..., $[t_1^{N-2}, \tau_1]$

$$N/2 \text{ of } z_i(t) = 0$$

$$N/2 \text{ of } z_i(t) = 1$$

  maximize *total* inter-z distance:

$$\sum_{i<j} \int_0^{\tau_1} w_{\tau_1}(\tau_1 - t)(z_i(t) - z_j(t))^2 dt = (\tfrac{N}{2})^2$$

- In particular ...

For $N = 0$ mod 4, $G = \{\tau_1\}$:

- Let $\mathbf{Z}$ be the $n \times (N-1)$ design matrix for any balanced, orthogonal, main-effects-saturated, 2-level design, with coding levels 0 and 1, e.g. for $N = 4$

$$\mathbf{Z} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

- For Z s.t. $z_i(t) = \mathbf{Z}_{ij}$ for $t \in [t_1^{j-1}, t_1^j]$ is Mm-optimal with $\phi = \frac{1}{2}\frac{N}{N-1}$

For $N = 0$ mod 4, $G = \{\tau_1, \tau_2, \tau_3, ..., \tau_M\}$:

- Define t_j^m, $m = 2, 3, ..., M$, $j = 1, 2, ..., N-2$ s.t.

$$\sum_{k=1}^{m} \int_{t_k^j}^{t_k^{j+1}} w_{\tau_m}(\tau_m - t)dt = \frac{1}{N-1},\ j = 1, 2, ..., N-1$$

- Extend 0/1 pattern used in $[0, \tau_1]$:

$$Z \text{ s.t. } z_i(t) = \mathbf{Z}_{ij} \text{ for } t \in [t_k^{j-1}, t_k^j],\ k = 2, 3, ..., M$$

- $D(z_i, z_j; w_\tau)$ is the *same* for all pairs of input functions and $\tau \in G$
- $\rightarrow Z$ is Mm-optimal with $\boxed{\phi = \frac{1}{2}\frac{N}{N-1}}$