# Diagnostics for Gaussian Process Emulators

Leo Bastos  Jeremy Oakley  Tony O'Hagan

Department of Probability & Statistics
University of Sheffield

Managing Uncertainty in Complex Models

The University Of Sheffield.

# Outline

1. Introduction
   - Motivation
   - Notation
   - Gaussian Process Emulator
2. Diagnostics
   - Individual prediction errors
   - Mahalanobis distance
   - Uncorrelated prediction errors
3. Conclusion and work in progress

The University Of Sheffield.

# Motivation

- Emulators are stochastic approximations of computer models
- Non-valid emulators can induce wrong conclusions
- There is little research into validating emulators
- Validation often means little more than:
  "*We compared some predictions with some observations and look, they fit quite well*"
- Present some diagnostics checking whether an emulator is valid or not, taking account all the associated uncertainty.

The University Of Sheffield.

# Motivation

- Emulators are stochastic approximations of computer models
- Non-valid emulators can induce wrong conclusions
- There is little research into validating emulators
- Validation often means little more than:
  "*We compared some predictions with some observations and look, they fit quite well*"
- Present some diagnostics checking whether an emulator is valid or not, taking account all the associated uncertainty.

The University Of Sheffield.

# Motivation

- Emulators are stochastic approximations of computer models
- Non-valid emulators can induce wrong conclusions
- There is little research into validating emulators
- Validation often means little more than:
  "*We compared some predictions with some observations and look, they fit quite well*"
- Present some diagnostics checking whether an emulator is valid or not, taking account all the associated uncertainty.

# Motivation

- Emulators are stochastic approximations of computer models
- Non-valid emulators can induce wrong conclusions
- There is little research into validating emulators
- Validation often means little more than:
  "*We compared some predictions with some observations and look, they fit quite well*"
- Present some diagnostics checking whether an emulator is valid or not, taking account all the associated uncertainty.

# Motivation

- Emulators are stochastic approximations of computer models
- Non-valid emulators can induce wrong conclusions
- There is little research into validating emulators
- Validation often means little more than:
  "*We compared some predictions with some observations and look, they fit quite well*"
- Present some diagnostics checking whether an emulator is valid or not, taking account all the associated uncertainty.

The University Of Sheffield.

# Notation

- **Simulator** is a complex deterministic mathematical function

$$\eta(\cdot) \qquad \eta : \mathcal{X} \in \Re^p \to \Re$$

- **Gaussian process emulator**

$$\eta(\cdot)|\beta, \sigma^2, \psi \sim GP\left(m_0(\cdot), V_0(\cdot, \cdot)\right),$$

where

$$\begin{aligned} m_0(\mathbf{x}) &= h(\mathbf{x})^T \beta \\ V_0(\mathbf{x}, \mathbf{x}') &= \sigma^2 C(\mathbf{x}, \mathbf{x}'; \psi) \end{aligned}$$

# Notation

- **Simulator** is a complex deterministic mathematical function

$$\eta(\cdot) \qquad \eta : \mathcal{X} \in \Re^p \to \Re$$

- **Gaussian process emulator**

$$\eta(\cdot)|\beta, \sigma^2, \psi \sim GP\left(m_0(\cdot), V_0(\cdot, \cdot)\right),$$

where

$$\begin{aligned} m_0(\mathbf{x}) &= h(\mathbf{x})^T \beta \\ V_0(\mathbf{x}, \mathbf{x}') &= \sigma^2 C(\mathbf{x}, \mathbf{x}'; \psi) \end{aligned}$$

# Gaussian Process Emulators

- **Training data** $(\mathbf{y}, \mathbf{X}) \quad : \quad y_i = \eta(\mathbf{x}_i), \quad i = 1, \ldots, n$
- **Predictive Gaussian Process Emulator**

$$\eta(\cdot)|\mathbf{y}, \mathbf{X}, \psi = \hat{\psi} \sim \text{Student-Process}\left(n - q, m_1(\cdot), V_1(\cdot, \cdot)\right),$$

where

$$\begin{aligned} m_1(x) &= h(x)^T\widehat{\beta} + t(x)^T\mathbf{A}^{-1}(\mathbf{y} - H\widehat{\beta}), \\ V_1(x, x') &= \hat{\sigma}^2\Big[C(x, x'; \psi) - t(x)^T\mathbf{A}^{-1}t(x') + \left(h(x) - t(x)^T\mathbf{A}^{-1}H\right) \\ &\times \left(H^T\mathbf{A}^{-1}H\right)^{-1}\left(h(x') - t(x')^T\mathbf{A}^{-1}H\right)^T\Big]. \end{aligned}$$

- Plug-in estimator for the correlation parameters $\psi = \hat{\psi}$.

# Gaussian Process Emulators

- **Training data** $(\mathbf{y}, \mathbf{X}) \quad : \quad y_i = \eta(\mathbf{x}_i), \quad i = 1, \ldots, n$
- **Predictive Gaussian Process Emulator**

$$\eta(\cdot)|\mathbf{y}, \mathbf{X}, \psi = \hat{\psi} \sim \text{Student-Process}\left(n-q, m_1(\cdot), V_1(\cdot,\cdot)\right),$$

where

$$\begin{aligned} m_1(x) &= h(x)^T\widehat{\beta} + t(x)^T\mathbf{A}^{-1}(\mathbf{y} - H\widehat{\beta}), \\ V_1(x,x') &= \hat{\sigma}^2\Big[C(x,x';\psi) - t(x)^T\mathbf{A}^{-1}t(x') + \left(h(x) - t(x)^T\mathbf{A}^{-1}H\right) \\ &\times \quad (H^T\mathbf{A}^{-1}H)^{-1}\left(h(x') - t(x')^T\mathbf{A}^{-1}H\right)^T\Big]. \end{aligned}$$

- Plug-in estimator for the correlation parameters $\psi = \hat{\psi}$.

The University Of Sheffield.

# Gaussian Process Emulators

- **Training data** $(\mathbf{y}, \mathbf{X}) \quad : \quad y_i = \eta(\mathbf{x}_i), \quad i = 1, \ldots, n$
- **Predictive Gaussian Process Emulator**

$$\eta(\cdot)|\mathbf{y}, \mathbf{X}, \psi = \hat{\psi} \sim \text{Student-Process}\left(n-q, m_1(\cdot), V_1(\cdot,\cdot)\right),$$

where

$$\begin{aligned} m_1(x) &= h(x)^T\widehat{\beta} + t(x)^T\mathbf{A}^{-1}(\mathbf{y} - H\widehat{\beta}), \\ V_1(x,x') &= \hat{\sigma}^2\Big[C(x,x';\psi) - t(x)^T\mathbf{A}^{-1}t(x') + \Big(h(x) - t(x)^T\mathbf{A}^{-1}H\Big) \\ &\times \quad (H^T\mathbf{A}^{-1}H)^{-1}\Big(h(x') - t(x')^T\mathbf{A}^{-1}H\Big)^T\Big]. \end{aligned}$$

- Plug-in estimator for the correlation parameters $\psi = \hat{\psi}$.

# Diagnostics

- Our diagnostics are based on a set of new simulator runs
  **Validation data** $(\mathbf{y}^*, \mathbf{X}^*) \quad : \quad y_i^* = \eta(\mathbf{x}_i^*), \quad i = 1, \ldots, m$
- Using a diagnostic $D(\cdot)$ , we compare $D(\mathbf{y}^*)$ with the referenced distribution of $D(\eta(\mathbf{X}^*))|\mathbf{y}, \mathbf{X}, \psi$.
- Proposal diagnostics:
  - Individual prediction errors
  - Mahalanobis distance
  - Uncorrelated prediction errors
  - Graphical methods

The University Of Sheffield.

# Diagnostics

- Our diagnostics are based on a set of new simulator runs
  **Validation data** $(\mathbf{y}^*, \mathbf{X}^*) \quad : \quad y_i^* = \eta(\mathbf{x}_i^*), \quad i = 1, \ldots, m$

- Using a diagnostic $D(\cdot)$ , we compare $D(\mathbf{y}^*)$ with the referenced distribution of $D(\eta(\mathbf{X}^*))|\mathbf{y}, \mathbf{X}, \psi$.

- Proposal diagnostics:
  - Individual prediction errors
  - Mahalanobis distance
  - Uncorrelated prediction errors
  - Graphical methods

The University Of Sheffield.

# Diagnostics

- Our diagnostics are based on a set of new simulator runs
  **Validation data** $(\mathbf{y}^*, \mathbf{X}^*) \quad : \quad y_i^* = \eta(\mathbf{x}_i^*), \quad i = 1, \ldots, m$
- Using a diagnostic $D(\cdot)$ , we compare $D(\mathbf{y}^*)$ with the referenced distribution of $D(\eta(\mathbf{X}^*))|\mathbf{y}, \mathbf{X}, \psi$.
- Proposal diagnostics:
  - Individual prediction errors
  - Mahalanobis distance
  - Uncorrelated prediction errors
  - Graphical methods

# Diagnostics

- Our diagnostics are based on a set of new simulator runs
  **Validation data** $(\mathbf{y}^*, \mathbf{X}^*) \quad : \quad y_i^* = \eta(\mathbf{x}_i^*), \quad i = 1, \ldots, m$
- Using a diagnostic $D(\cdot)$ , we compare $D(\mathbf{y}^*)$ with the referenced distribution of $D(\eta(\mathbf{X}^*))|\mathbf{y}, \mathbf{X}, \psi$.
- Proposal diagnostics:
  - Individual prediction errors
  - Mahalanobis distance
  - Uncorrelated prediction errors
  - Graphical methods

The University Of Sheffield.

# Diagnostics

- Our diagnostics are based on a set of new simulator runs
  **Validation data** $(\mathbf{y}^*, \mathbf{X}^*) \quad : \quad y_i^* = \eta(\mathbf{x}_i^*), \quad i = 1, \ldots, m$
- Using a diagnostic $D(\cdot)$ , we compare $D(\mathbf{y}^*)$ with the referenced distribution of $D(\eta(\mathbf{X}^*))|\mathbf{y}, \mathbf{X}, \psi$.
- Proposal diagnostics:
  - Individual prediction errors
  - Mahalanobis distance
  - Uncorrelated prediction errors
  - Graphical methods

The University Of Sheffield.

# Diagnostics

- Our diagnostics are based on a set of new simulator runs
  **Validation data** $(\mathbf{y}^*, \mathbf{X}^*) \quad : \quad y_i^* = \eta(\mathbf{x}_i^*), \quad i = 1, \ldots, m$
- Using a diagnostic $D(\cdot)$ , we compare $D(\mathbf{y}^*)$ with the referenced distribution of $D(\eta(\mathbf{X}^*))|\mathbf{y}, \mathbf{X}, \psi$.
- Proposal diagnostics:
  - Individual prediction errors
  - Mahalanobis distance
  - Uncorrelated prediction errors
  - Graphical methods

# Diagnostics

- Our diagnostics are based on a set of new simulator runs
  **Validation data** $(\mathbf{y}^*, \mathbf{X}^*) \quad : \quad y_i^* = \eta(\mathbf{x}_i^*), \quad i = 1, \ldots, m$
- Using a diagnostic $D(\cdot)$ , we compare $D(\mathbf{y}^*)$ with the referenced distribution of $D(\eta(\mathbf{X}^*))|\mathbf{y}, \mathbf{X}, \psi$.
- Proposal diagnostics:
  - Individual prediction errors
  - Mahalanobis distance
  - Uncorrelated prediction errors
  - Graphical methods

The University Of Sheffield.

# Toy example

- $\eta(\mathbf{x})$ is a Gaussian process $\left(\mathbf{x}^T\beta,\ \sigma^2 \exp\{\sum_k (\frac{x_k - x'_k}{\psi_k})^2\}\right)$
- 100 training data points (Maximin LHS design)
- 50 validation points (validation design*)
- 3 scenarios: (Varying the plug-in correlation parameters)

The University Of Sheffield.

# Toy example

- $\eta(\mathbf{x})$ is a Gaussian process $\left(\mathbf{x}^T\beta,\ \sigma^2 \exp\{\sum_k (\frac{x_k - x_k'}{\psi_k})^2\}\right)$
- 100 training data points (Maximin LHS design)
- 50 validation points (validation design*)
- 3 scenarios: (Varying the plug-in correlation parameters)
  1. $\hat{\psi} \approx \psi \Longrightarrow V_1(\eta(x)|\hat{\psi}) \approx V_1(\eta(x)|\psi)$
  2. $\hat{\psi} > \psi \Longrightarrow V_1(\eta(x)|\hat{\psi}) < V_1(\eta(x)|\psi)$
  3. $\hat{\psi} < \psi \Longrightarrow V_1(\eta(x)|\hat{\psi}) > V_1(\eta(x)|\psi)$

# Toy example

- $\eta(\mathbf{x})$ is a Gaussian process $\left(\mathbf{x}^T\beta,\ \sigma^2\,\text{exp}\{\sum_k(\frac{x_k - x'_k}{\psi_k})^2\}\right)$
- 100 training data points (Maximin LHS design)
- 50 validation points (validation design*)
- 3 scenarios: (Varying the plug-in correlation parameters)
  1. $\hat{\psi} \approx \psi \Longrightarrow V_1(\eta(x)|\hat{\psi}) \approx V_1(\eta(x)|\psi)$
  2. $\hat{\psi} > \psi \Longrightarrow V_1(\eta(x)|\hat{\psi}) < V_1(\eta(x)|\psi)$
  3. $\hat{\psi} < \psi \Longrightarrow V_1(\eta(x)|\hat{\psi}) > V_1(\eta(x)|\psi)$

The University Of Sheffield.

# Toy example

- $\eta(\mathbf{x})$ is a Gaussian process $\left(\mathbf{x}^T\beta,\ \sigma^2 \exp\{\sum_k (\frac{x_k - x'_k}{\psi_k})^2\}\right)$
- 100 training data points (Maximin LHS design)
- 50 validation points (validation design*)
- 3 scenarios: (Varying the plug-in correlation parameters)
  1. $\hat{\psi} \approx \psi \Longrightarrow V_1(\eta(x)|\hat{\psi}) \approx V_1(\eta(x)|\psi)$
  2. $\hat{\psi} > \psi \Longrightarrow V_1(\eta(x)|\hat{\psi}) < V_1(\eta(x)|\psi)$
  3. $\hat{\psi} < \psi \Longrightarrow V_1(\eta(x)|\hat{\psi}) > V_1(\eta(x)|\psi)$

# Toy example

- $\eta(\mathbf{x})$ is a Gaussian process $\left(\mathbf{x}^T\beta,\ \sigma^2 \exp\{\sum_k (\frac{x_k - x'_k}{\psi_k})^2\}\right)$
- 100 training data points (Maximin LHS design)
- 50 validation points (validation design*)
- 3 scenarios: (Varying the plug-in correlation parameters)
  1. $\hat{\psi} \approx \psi \implies V_1(\eta(x)|\hat{\psi}) \approx V_1(\eta(x)|\psi)$
  2. $\hat{\psi} > \psi \implies V_1(\eta(x)|\hat{\psi}) < V_1(\eta(x)|\psi)$
  3. $\hat{\psi} < \psi \implies V_1(\eta(x)|\hat{\psi}) > V_1(\eta(x)|\psi)$

# Toy example

- $\eta(\mathbf{x})$ is a Gaussian process $\left(\mathbf{x}^T\beta,\ \sigma^2 \exp\{\sum_k (\frac{x_k - x'_k}{\psi_k})^2\}\right)$
- 100 training data points (Maximin LHS design)
- 50 validation points (validation design*)
- 3 scenarios: (Varying the plug-in correlation parameters)
  1. $\hat{\psi} \approx \psi \Longrightarrow V_1(\eta(x)|\hat{\psi}) \approx V_1(\eta(x)|\psi)$
  2. $\hat{\psi} > \psi \Longrightarrow V_1(\eta(x)|\hat{\psi}) < V_1(\eta(x)|\psi)$
  3. $\hat{\psi} < \psi \Longrightarrow V_1(\eta(x)|\hat{\psi}) > V_1(\eta(x)|\psi)$

The University Of Sheffield.

# Diagnostics

## Individual standardized prediction errors

$$D_i(\mathbf{y}^*) = \frac{(y_i^* - E[\eta(\mathbf{x}_i^*)])}{\sqrt{V[\eta(\mathbf{x}_i^*, \mathbf{x}_i^*)]}}, i = 1, \ldots, n$$

$$D_i(\eta(\mathbf{x}_i^*)|\mathbf{y}, \mathbf{X}, \psi \sim \text{Student-t}(n-q)$$

# Diagnostics

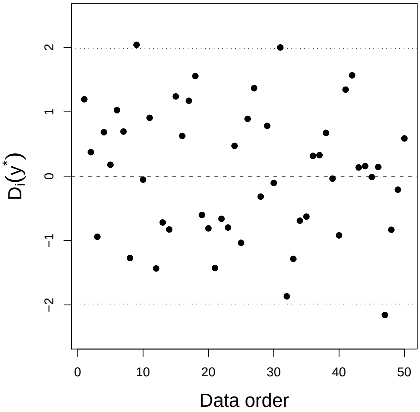
## Individual standardized prediction errors

$$D_i(\mathbf{y}^*) = \frac{(y_i^* - E[\eta(\mathbf{x}_i^*)])}{\sqrt{V[\eta(\mathbf{x}_i^*, \mathbf{x}_i^*)]}}, i = 1, \ldots, n$$

$$D_i(\eta(\mathbf{x}_i^*)|\mathbf{y}, \mathbf{X}, \psi \sim \text{Student-t}(n - q)$$

The University Of Sheffield.

# Individual Standardized Prediction errors

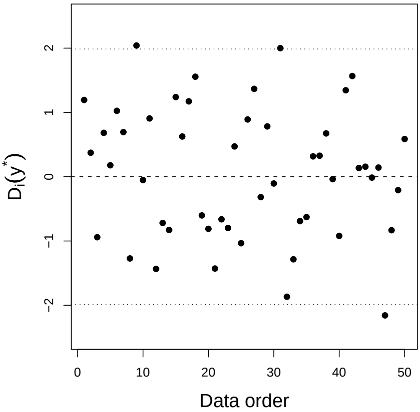



- The dotted lines indicate the 95% credible limits
- A large number of extreme errors might indicate a stationarity problem
- Too many **large errors** suggest **under-estimation** of the uncertainty
- Too many **small errors** suggest **over-estimation** of the uncertainty

However, these errors are not independent!

The University Of Sheffield.

# Individual Standardized Prediction errors

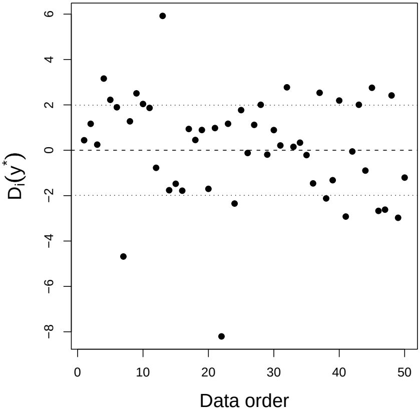



- The dotted lines indicate the 95% credible limits
- A large number of extreme errors might indicate a stationarity problem
- Too many **large errors** suggest **under-estimation** of the uncertainty
- Too many **small errors** suggest **over-estimation** of the uncertainty

However, these errors are not independent!

The University Of Sheffield.

# Individual Standardized Prediction errors

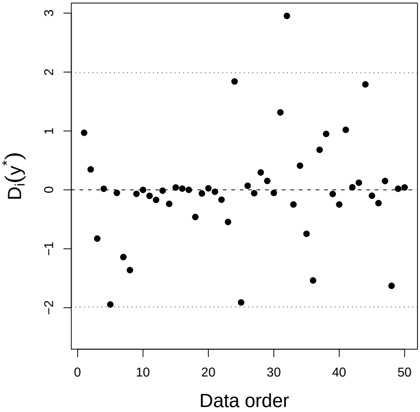



- The dotted lines indicate the 95% credible limits
- A large number of extreme errors might indicate a stationarity problem
- Too many **large errors** suggest **under-estimation** of the uncertainty
- Too many **small errors** suggest **over-estimation** of the uncertainty

However, these errors are not independent!

The University Of Sheffield.

# Individual Standardized Prediction errors

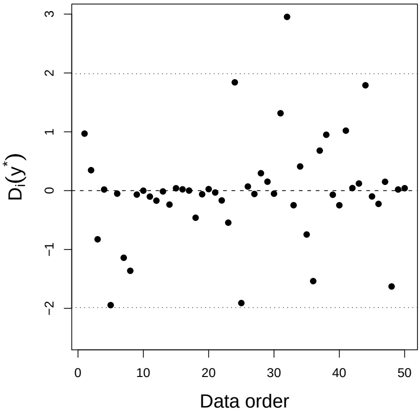



- The dotted lines indicate the 95% credible limits
- A large number of extreme errors might indicate a stationarity problem
- Too many **large errors** suggest **under-estimation** of the uncertainty
- Too many **small errors** suggest **over-estimation** of the uncertainty

However, these errors are not independent!

The University Of Sheffield.

# Individual Standardized Prediction errors




- The dotted lines indicate the 95% credible limits
- A large number of extreme errors might indicate a stationarity problem
- Too many **large errors** suggest **under-estimation** of the uncertainty
- Too many **small errors** suggest **over-estimation** of the uncertainty

However, these errors are not independent!

## Mahalanobis distance

$$D_{MD}(\mathbf{y}^*) = (\mathbf{y}^* - m_1(\mathbf{X}^*))^T \, V_1(\mathbf{X}^*)^{-1} \, (\mathbf{y}^* - m_1(\mathbf{X}^*))$$

- Large values for $D_{MD}(\mathbf{y}^*)$ are associated with
  - Large individual errors
  - Small uncertainty
- Small values for $D_{MD}(\mathbf{y}^*)$ are associated with
  - Small individual errors
  - Large uncertainty

$$\frac{(n-q)}{m(n-q-2)} D_{MD}(\eta(\mathbf{X}^*)) \sim F_{m,n-q}$$

## Mahalanobis distance

$$D_{MD}(\mathbf{y}^*) \;=\; (\mathbf{y}^* - m_1(\mathbf{X}^*))^T\, V_1(\mathbf{X}^*)^{-1}\, (\mathbf{y}^* - m_1(\mathbf{X}^*))$$

- Large values for $D_{MD}(\mathbf{y}^*)$ are associated with
  - Large individual errors
  - Small uncertainty
- Small values for $D_{MD}(\mathbf{y}^*)$ are associated with
  - Small individual errors
  - Large uncertainty

$$\frac{(n-q)}{m(n-q-2)} D_{MD}(\eta(\mathbf{X}^*)) \sim F_{m,n-q}$$

## Mahalanobis distance

$$D_{MD}(\mathbf{y}^*) = (\mathbf{y}^* - m_1(\mathbf{X}^*))^T V_1(\mathbf{X}^*)^{-1} (\mathbf{y}^* - m_1(\mathbf{X}^*))$$

- Large values for $D_{MD}(\mathbf{y}^*)$ are associated with
  - Large individual errors
  - Small uncertainty
- Small values for $D_{MD}(\mathbf{y}^*)$ are associated with
  - Small individual errors
  - Large uncertainty

$$\frac{(n-q)}{m(n-q-2)} D_{MD}(\eta(\mathbf{X}^*)) \sim F_{m,n-q}$$

## Mahalanobis distance

$$D_{MD}(\mathbf{y}^*) = (\mathbf{y}^* - m_1(\mathbf{X}^*))^T V_1(\mathbf{X}^*)^{-1} (\mathbf{y}^* - m_1(\mathbf{X}^*))$$

- Large values for $D_{MD}(\mathbf{y}^*)$ are associated with
  - Large individual errors
  - Small uncertainty
- Small values for $D_{MD}(\mathbf{y}^*)$ are associated with
  - Small individual errors
  - Large uncertainty

$$\frac{(n-q)}{m(n-q-2)} D_{MD}(\eta(\mathbf{X}^*)) \sim F_{m,n-q}$$

## Mahalanobis distance

$$D_{MD}(\mathbf{y}^*) = (\mathbf{y}^* - m_1(\mathbf{X}^*))^T V_1(\mathbf{X}^*)^{-1} (\mathbf{y}^* - m_1(\mathbf{X}^*))$$

- Large values for $D_{MD}(\mathbf{y}^*)$ are associated with
  - Large individual errors
  - Small uncertainty
- Small values for $D_{MD}(\mathbf{y}^*)$ are associated with
  - Small individual errors
  - Large uncertainty

$$\frac{(n-q)}{m(n-q-2)} D_{MD}(\eta(\mathbf{X}^*)) \sim F_{m,n-q}$$

## Mahalanobis distance

$$D_{MD}(\mathbf{y}^*) = (\mathbf{y}^* - m_1(\mathbf{X}^*))^T V_1(\mathbf{X}^*)^{-1} (\mathbf{y}^* - m_1(\mathbf{X}^*))$$

- Large values for $D_{MD}(\mathbf{y}^*)$ are associated with
  - Large individual errors
  - Small uncertainty
- Small values for $D_{MD}(\mathbf{y}^*)$ are associated with
  - Small individual errors
  - Large uncertainty

$$\frac{(n-q)}{m(n-q-2)} D_{MD}(\eta(\mathbf{X}^*)) \sim F_{m,n-q}$$

## Mahalanobis distance

$$D_{MD}(\mathbf{y}^*) \quad = \quad (\mathbf{y}^* - m_1(\mathbf{X}^*))^T \, V_1(\mathbf{X}^*)^{-1} \, (\mathbf{y}^* - m_1(\mathbf{X}^*))$$

- Large values for $D_{MD}(\mathbf{y}^*)$ are associated with
  - Large individual errors
  - Small uncertainty
- Small values for $D_{MD}(\mathbf{y}^*)$ are associated with
  - Small individual errors
  - Large uncertainty

$$\frac{(n-q)}{m(n-q-2)} D_{MD}(\eta(\mathbf{X}^*)) \sim F_{m,n-q}$$

The University Of Sheffield.

# Mahalanobis distance

## Mahalanobis distance

$$D_{MD}(\mathbf{y}^*) = (\mathbf{y}^* - m_1(\mathbf{X}^*))^T \, V_1(\mathbf{X}^*)^{-1} \, (\mathbf{y}^* - m_1(\mathbf{X}^*))$$

- Large values for $D_{MD}(\mathbf{y}^*)$ are associated with
  - Large individual errors
  - Small uncertainty
- Small values for $D_{MD}(\mathbf{y}^*)$ are associated with
  - Small individual errors
  - Large uncertainty

$$\frac{(n-q)}{m(n-q-2)} D_{MD}(\eta(\mathbf{X}^*)) \sim F_{m,n-q}$$

The University Of Sheffield.

# Numerical diagnostics

## Uncorrelated prediction errors

$$D^G(\mathbf{y}^*) = (\mathbf{G}^{-1})^T (\mathbf{y}^* - m_1(\mathbf{X}^*))$$

where $V_1(\mathbf{X}^*) = \mathbf{G}^{\mathbf{T}}\mathbf{G}$.

- Cholesky decomposition: $\mathbf{G}$ is a triangular matrix
- Eigen decomposition: $\mathbf{G} = \Lambda^{1/2} E^T$
- Pivoted Cholesky decomposition: $\mathbf{G} = \mathbf{P}\mathbf{R}^T$, $\mathbf{P}$ is a permutation matrix

$$D^G(\mathbf{y}^*) \sim \text{Student-t}_m(n-q, \mathbf{0}, \mathcal{I}_m)$$

$$D^G(\mathbf{y}^*)^T D^G(\mathbf{y}^*) = D_{MD}(\mathbf{y}^*)$$

# Numerical diagnostics

## Uncorrelated prediction errors

$$D^G(\mathbf{y}^*) = (\mathbf{G}^{-1})^T (\mathbf{y}^* - m_1(\mathbf{X}^*))$$

where $V_1(\mathbf{X}^*) = \mathbf{G}^\mathbf{T}\mathbf{G}$.

- Cholesky decomposition: **G** is a triangular matrix
- Eigen decomposition: $\mathbf{G} = \Lambda^{1/2} E^T$
- Pivoted Cholesky decomposition: $\mathbf{G} = \mathbf{P}\mathbf{R}^T$, **P** is a permutation matrix

$$D^G(\mathbf{y}^*) \sim \text{Student-t}_m(n - q, \mathbf{0}, \mathcal{I}_m)$$

$$D^G(\mathbf{y}^*)^T D^G(\mathbf{y}^*) = D_{MD}(\mathbf{y}^*)$$

# Numerical diagnostics

## Uncorrelated prediction errors

$$D^G(\mathbf{y}^*) = (\mathbf{G}^{-1})^T (\mathbf{y}^* - m_1(\mathbf{X}^*))$$

where $V_1(\mathbf{X}^*) = \mathbf{G}^\mathbf{T}\mathbf{G}$.

- Cholesky decomposition: **G** is a triangular matrix
- Eigen decomposition: $\mathbf{G} = \Lambda^{1/2} E^T$
- Pivoted Cholesky decomposition: $\mathbf{G} = \mathbf{P}\mathbf{R}^T$, **P** is a permutation matrix

$$D^G(\mathbf{y}^*) \sim \text{Student-t}_m(n-q, \mathbf{0}, \mathcal{I}_m)$$

$$D^G(\mathbf{y}^*)^T D^G(\mathbf{y}^*) = D_{MD}(\mathbf{y}^*)$$

The University Of Sheffield.

## Uncorrelated prediction errors

$$D^G(\mathbf{y}^*) = (\mathbf{G}^{-1})^T (\mathbf{y}^* - m_1(\mathbf{X}^*))$$

where $V_1(\mathbf{X}^*) = \mathbf{G^T G}$.

- Cholesky decomposition: $\mathbf{G}$ is a triangular matrix
- Eigen decomposition: $\mathbf{G} = \Lambda^{1/2} E^T$
- Pivoted Cholesky decomposition: $\mathbf{G} = \mathbf{P}\mathbf{R}^T$, $\mathbf{P}$ is a permutation matrix

$$D^G(\mathbf{y}^*) \sim \text{Student-t}_m(n - q, \mathbf{0}, \mathcal{I}_m)$$

$$D^G(\mathbf{y}^*)^T D^G(\mathbf{y}^*) = D_{MD}(\mathbf{y}^*)$$

The University Of Sheffield.

## Uncorrelated prediction errors

$$D^G(\mathbf{y}^*) \;=\; (\mathbf{G}^{-1})^T (\mathbf{y}^* - m_1(\mathbf{X}^*))$$

where $V_1(\mathbf{X}^*) = \mathbf{G}^\mathbf{T}\mathbf{G}$.

- Cholesky decomposition: **G** is a triangular matrix
- Eigen decomposition: $\mathbf{G} = \Lambda^{1/2} E^T$
- Pivoted Cholesky decomposition: $\mathbf{G} = \mathbf{P}\mathbf{R}^T$, **P** is a permutation matrix

$$D^G(\mathbf{y}^*) \sim \text{Student-t}_m(n - q, \mathbf{0}, \mathcal{I}_m)$$

$$D^G(\mathbf{y}^*)^T D^G(\mathbf{y}^*) = D_{MD}(\mathbf{y}^*)$$

# Numerical diagnostics

## Uncorrelated prediction errors

$$D^G(\mathbf{y}^*) \quad = \quad (\mathbf{G}^{-1})^T(\mathbf{y}^* - m_1(\mathbf{X}^*))$$

where $V_1(\mathbf{X}^*) = \mathbf{G^T G}$.

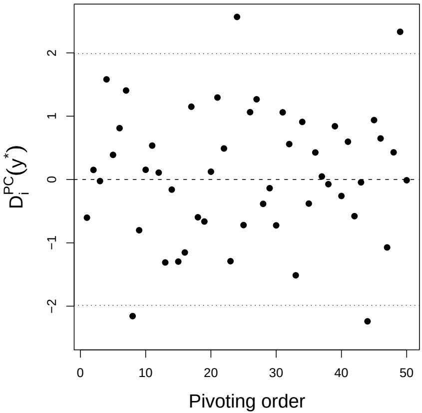
- Cholesky decomposition: **G** is a triangular matrix
- Eigen decomposition: $\mathbf{G} = \Lambda^{1/2}E^T$
- Pivoted Cholesky decomposition: $\mathbf{G} = \mathbf{P}\mathbf{R}^T$, **P** is a permutation matrix

$$D^G(\mathbf{y}^*) \sim \text{Student-t}_m(n - q, \mathbf{0}, \mathcal{I}_m)$$

$$D^G(\mathbf{y}^*)^T D^G(\mathbf{y}^*) = D_{MD}(\mathbf{y}^*)$$

The University Of Sheffield.

# Uncorrelated prediction errors against pivoting order

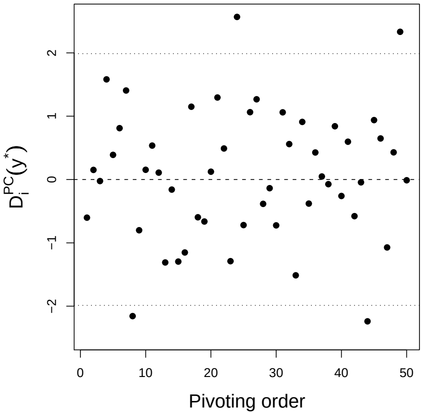



$$D_{MD}(\mathbf{y}^*) = 51.727$$

- Mahalanobis distance close to its expected value (50.0)
- Theoretical 95% CI (29.618, 78.767)
- Errors should be randomly distributed around zero
- Suggesting a good emulator

The University Of Sheffield.

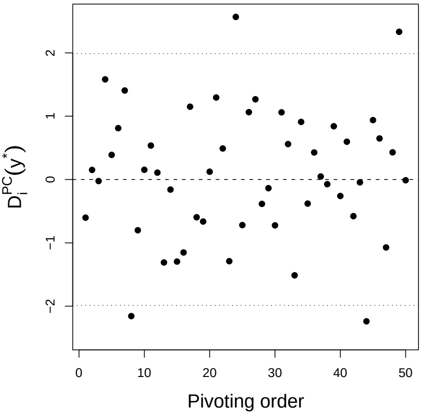
# Uncorrelated prediction errors against pivoting order




$$D_{MD}(\mathbf{y}^*) = 51.727$$

- Mahalanobis distance close to its expected value (50.0)
- Theoretical 95% CI (29.618, 78.767)
- Errors should be randomly distributed around zero
- Suggesting a good emulator

# Uncorrelated prediction errors against pivoting order

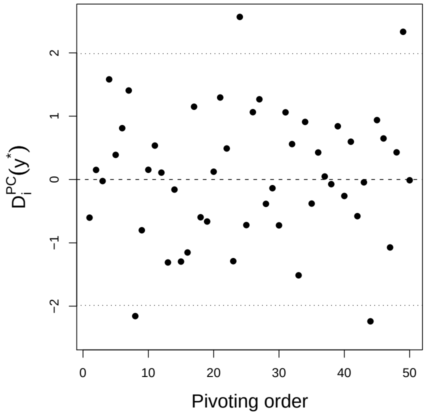



$$D_{MD}(\mathbf{y}^*) = 51.727$$

- Mahalanobis distance close to its expected value (50.0)
- Theoretical 95% CI (29.618, 78.767)
- Errors should be randomly distributed around zero
- Suggesting a good emulator

The University Of Sheffield.

# Uncorrelated prediction errors against pivoting order

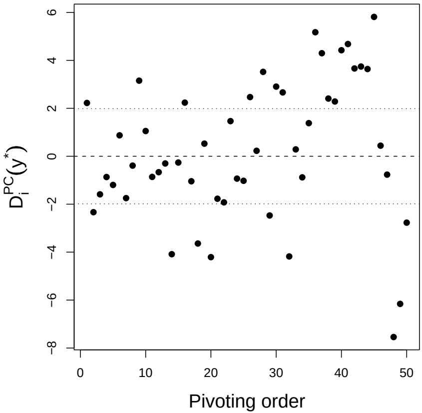



- Mahalanobis distance close to its expected value (50.0)
- Theoretical 95% CI (29.618, 78.767)
- Errors should be randomly distributed around zero
- Suggesting a good emulator

$$D_{MD}(\mathbf{y}^*) = 51.727$$

The University Of Sheffield.

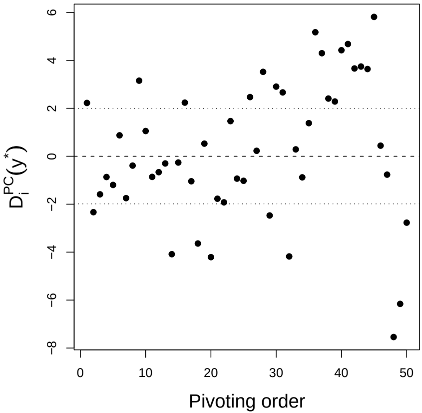
# Uncorrelated prediction errors against pivoting order




- Large mahalanobis distance
- Theoretical 95% CI (29.618, 78.767)
- Non-stationarity
- Large errors at the end indicate under estimation of the correlation parameters

$$D_{MD}(\mathbf{y}^*) = 432.171$$

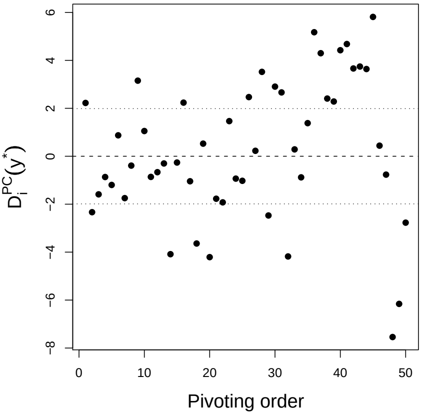
# Uncorrelated prediction errors against pivoting order




- Large mahalanobis distance
- Theoretical 95% CI (29.618, 78.767)
- Non-stationarity
- Large errors at the end indicate under estimation of the correlation parameters

$$D_{MD}(\mathbf{y}^*) = 432.171$$

# Uncorrelated prediction errors against pivoting order

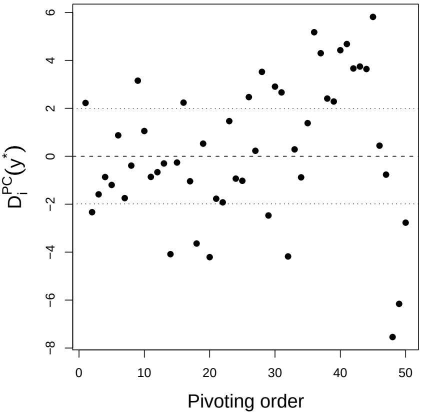



- Large mahalanobis distance
- Theoretical 95% CI (29.618, 78.767)
- Non-stationarity
- Large errors at the end indicate under estimation of the correlation parameters

$$D_{MD}(\mathbf{y}^*) = 432.171$$

The University Of Sheffield.

# Uncorrelated prediction errors against pivoting order

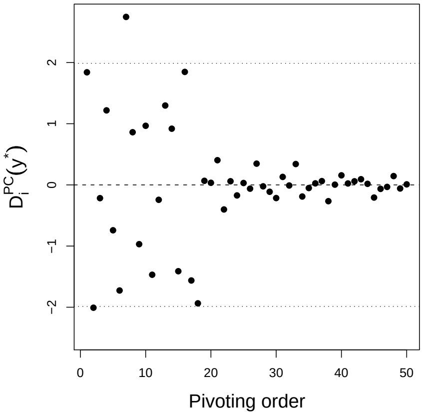



- Large mahalanobis distance
- Theoretical 95% CI (29.618, 78.767)
- Non-stationarity
- Large errors at the end indicate under estimation of the correlation parameters

$$D_{MD}(\mathbf{y}^*) = 432.171$$

The University Of Sheffield.

# Uncorrelated prediction errors against pivoting order

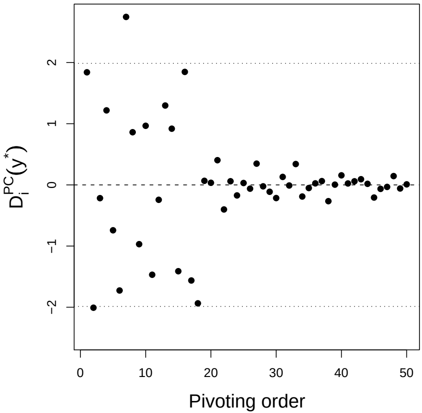



$$D_{MD}(\mathbf{y}^*) = 24.143$$

- Small mahalanobis distance
- Theoretical 95% CI (29.618, 78.767)
- Non-stationarity
- Small errors at the end indicate over estimation of the correlation parameters

The University Of Sheffield.

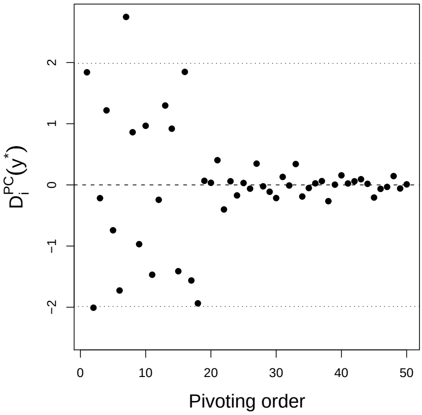
# Uncorrelated prediction errors against pivoting order




$$D_{MD}(\mathbf{y}^*) = 24.143$$

- Small mahalanobis distance
- Theoretical 95% CI (29.618, 78.767)
- Non-stationarity
- Small errors at the end indicate over estimation of the correlation parameters

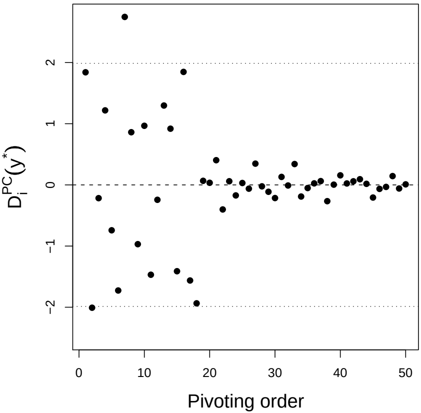
# Uncorrelated prediction errors against pivoting order




$$D_{MD}(\mathbf{y}^*) = 24.143$$

- Small mahalanobis distance
- Theoretical 95% CI (29.618, 78.767)
- Non-stationarity
- Small errors at the end indicate over estimation of the correlation parameters

# Uncorrelated prediction errors against pivoting order




- Small mahalanobis distance
- Theoretical 95% CI (29.618, 78.767)
- Non-stationarity
- Small errors at the end indicate over estimation of the correlation parameters

$$D_{MD}(\mathbf{y}^*) = 24.143$$

The University Of Sheffield.

# More graphical methods

- Individual errors against emulator's predictions
- QQ-plots of the uncorrelated standardized errors
- Individual or (pivoted) Cholesky errors against inputs

# More graphical methods

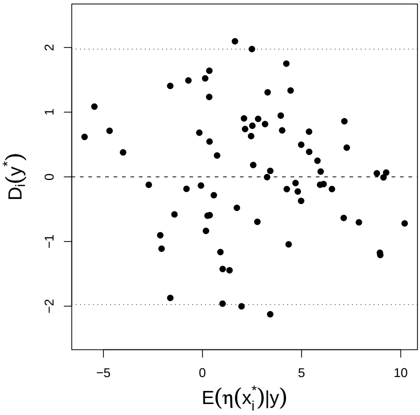
- Individual errors against emulator's predictions
- QQ-plots of the uncorrelated standardized errors
- Individual or (pivoted) Cholesky errors against inputs

The University Of Sheffield.

# More graphical methods

- Individual errors against emulator's predictions
- QQ-plots of the uncorrelated standardized errors
- Individual or (pivoted) Cholesky errors against inputs

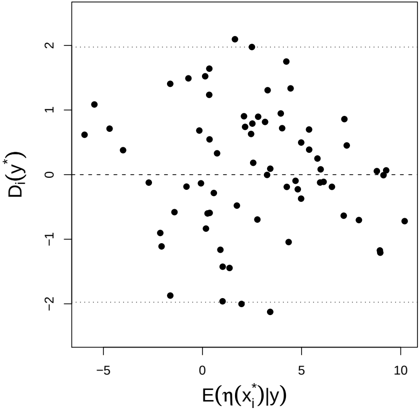
# Individual errors against emulator’s predictions




Possible problems:

- Patterns:
  Misspecification of $m_0(\cdot)$
- Small or large errors:
  Uncertainty poorly estimated
- Heterocedasticity:
  Non-stationarity

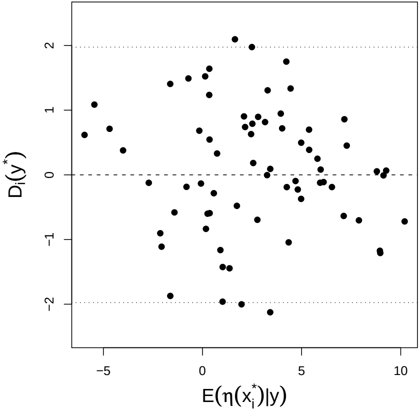
# Individual errors against emulator's predictions




Possible problems:

- Patterns:
  Misspecification of $m_0(\cdot)$
- Small or large errors:
  Uncertainty poorly estimated
- Heterocedasticity:
  Non-stationarity

# Individual errors against emulator's predictions



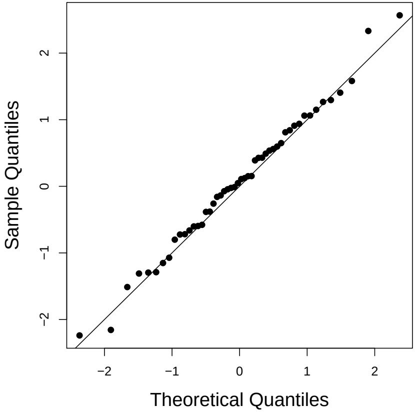

Possible problems:

- Patterns:
  Misspecification of $m_0(\cdot)$
- Small or large errors:
  Uncertainty poorly estimated
- Heterocedasticity:
  Non-stationarity

The University Of Sheffield.

# QQ-plot for the uncorrelated errors

$$V_1(\eta(x)|\hat{\psi}) \approx V_1(\eta(x)|\psi)$$

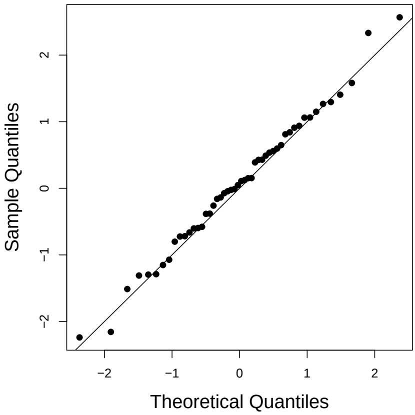



Possible problems:

- Non-Gaussianity
- Non-stationarity
- Predictive variance poorly estimated
- Uncertainty under estimated
- Uncertainty over estimated

# QQ-plot for the uncorrelated errors

$$V_1(\eta(x)|\hat{\psi}) \approx V_1(\eta(x)|\psi)$$

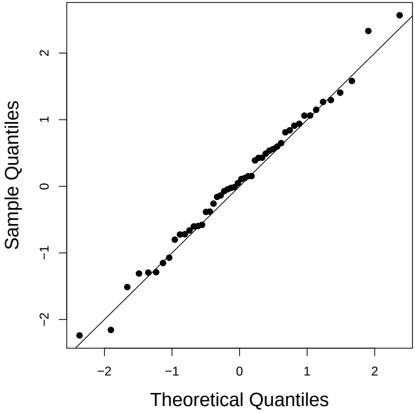



Possible problems:

- Non-Gaussianity
- Non-stationarity
- Predictive variance poorly estimated
- Uncertainty under estimated
- Uncertainty over estimated

# QQ-plot for the uncorrelated errors

$$V_1(\eta(x)|\hat{\psi}) \approx V_1(\eta(x)|\psi)$$



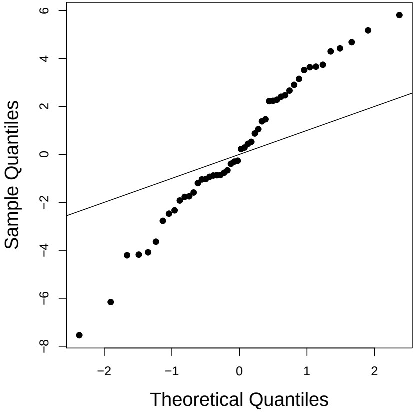

Possible problems:

- Non-Gaussianity
- Non-stationarity
- Predictive variance poorly estimated
- Uncertainty under estimated
- Uncertainty over estimated

The University Of Sheffield.

# QQ-plot for the uncorrelated errors

$$V_1(\eta(x)|\hat{\psi}) < V_1(\eta(x)|\psi)$$

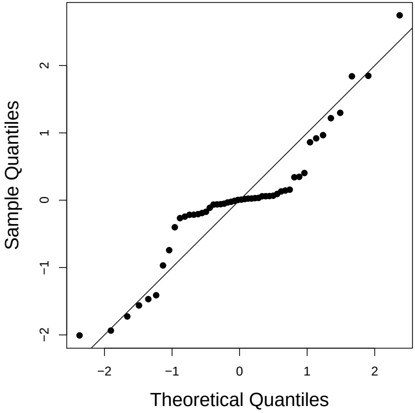



Possible problems:

- Non-Gaussianity
- Non-stationarity
- Predictive variance poorly estimated
- Uncertainty under estimated
- Uncertainty over estimated

# QQ-plot for the uncorrelated errors

$$V_1(\eta(x)|\hat{\psi}) > V_1(\eta(x)|\psi)$$

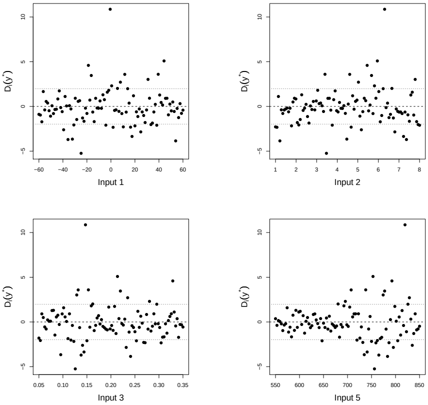



Possible problems:

- Non-Gaussianity
- Non-stationarity
- Predictive variance poorly estimated
- Uncertainty under estimated
- Uncertainty over estimated

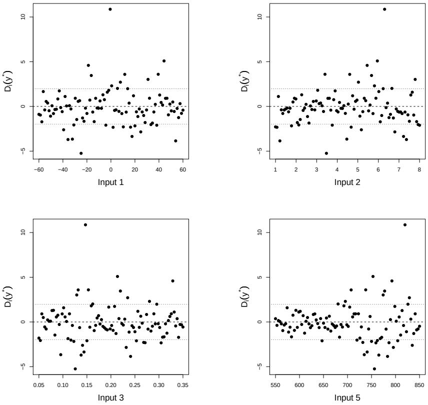
The University Of Sheffield.

# Predictions errors agains inputs




Possible problems:

- Patterns not included in the mean function
- Non-stationarity (problems in sub regions of th input space)
- $D_i(\mathbf{y}^*)$'s against Input 5 suggests a nonlinearity.

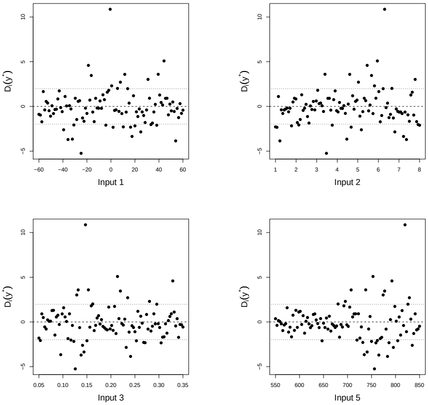



Possible problems:

- Patterns not included in the mean function
- Non-stationarity (problems in sub regions of th input space)
- $D_i(\mathbf{y}^*)$'s against Input 5 suggests a nonlinearity.

The University Of Sheffield.

# Predictions errors agains inputs




Possible problems:

- Patterns not included in the mean function
- Non-stationarity (problems in sub regions of th input space)
- $D_i(\mathbf{y}^*)$'s against Input 5 suggests a nonlinearity.

# Conclusion and work in progress

- Validating an emulator is a important before using the emulator as a surrogate for the computer model.
- Our diagnostics are useful tools in the process of validating an emulator
- Cross-validation methods were not mentioned
- Next
  - Design for validation
  - Diagnostics for multiple output emulators and dynamical emulators
  - Diagnostics for GP model discrepancy function (when no emulation is needed)

# Conclusion and work in progress

- Validating an emulator is a important before using the emulator as a surrogate for the computer model.
- Our diagnostics are useful tools in the process of validating an emulator
- Cross-validation methods were not mentioned
- Next
  - Design for validation
  - Diagnostics for multiple output emulators and dynamical emulators
  - Diagnostics for GP model discrepancy function (when no emulation is needed)

The University Of Sheffield.

# Conclusion and work in progress

- Validating an emulator is a important before using the emulator as a surrogate for the computer model.
- Our diagnostics are useful tools in the process of validating an emulator
- Cross-validation methods were not mentioned
- Next
  - Design for validation
  - Diagnostics for multiple output emulators and dynamical emulators
  - Diagnostics for GP model discrepancy function (when no emulation is needed)

The University Of Sheffield.

# Conclusion and work in progress

- Validating an emulator is a important before using the emulator as a surrogate for the computer model.
- Our diagnostics are useful tools in the process of validating an emulator
- Cross-validation methods were not mentioned

- Next
  - Design for validation
  - Diagnostics for multiple output emulators and dynamical emulators
  - Diagnostics for GP model discrepancy function (when no emulation is needed)

The University Of Sheffield.

# Conclusion and work in progress

- Validating an emulator is a important before using the emulator as a surrogate for the computer model.
- Our diagnostics are useful tools in the process of validating an emulator
- Cross-validation methods were not mentioned
- Next
  - Design for validation
  - Diagnostics for multiple output emulators and dynamical emulators
  - Diagnostics for GP model discrepancy function (when no emulation is needed)

The University Of Sheffield.

# Conclusion and work in progress

- Validating an emulator is a important before using the emulator as a surrogate for the computer model.
- Our diagnostics are useful tools in the process of validating an emulator
- Cross-validation methods were not mentioned

- Next
  - Design for validation
  - Diagnostics for multiple output emulators and dynamical emulators
  - Diagnostics for GP model discrepancy function (when no emulation is needed)

# Conclusion and work in progress

- Validating an emulator is a important before using the emulator as a surrogate for the computer model.
- Our diagnostics are useful tools in the process of validating an emulator
- Cross-validation methods were not mentioned

- Next
  - Design for validation
  - Diagnostics for multiple output emulators and dynamical emulators
  - Diagnostics for GP model discrepancy function (when no emulation is needed)