

Detección de *mismodelling* en base a un *test* de independencia definido según una estimación de información mutua *data-driven*

Experimentos preliminares modelos de ruido aditivo

Camilo A. Ramírez Canales

En el presente documento se presentan resultados de experimentos relacionados a la estimación de información mutua, el *test* de independencia asociado, y los tiempos de detección de no independencia; con respecto a entradas y residuos de modelos de ruido aditivo.

Los modelos de ruido aditivo siguen en general la siguiente regla:

$$Y|X = \eta_{\delta_1, \delta_2}(X) + h(W),$$

en donde X corresponde a una variable aleatoria que representa una entrada potencialmente multidimensional de distribución arbitraria, W corresponde a otra variable aleatoria que representa un ruido potencialmente multidimensional de distribución arbitraria, y además $\eta_{(\delta_1, \delta_2)}$ y h son funciones medibles de dimensionalidad adecuada, en donde $\delta \doteq (\delta_1, \delta_2)$ representa el grado de *mis-modelling* del sistema, respecto a la condición nominal dada por $\delta = (0, 0) \doteq \mathbf{0}$.

Para los sistemas expresados, el residuo está dado por la siguiente expresión:

$$r_{\eta_0}(X, Y) \doteq r_{\eta}(X, Y) \doteq R = \eta_{\delta_1, \delta_2}(X) - \eta_0(X) + h(W).$$

Uno de los elementos registrados en los experimentos corresponde a la estimación de información mutua de la entrada, o componentes de la misma, con el residuo; esto se puede denotar $\hat{I}(\mathbf{Z}_n)$, en donde \mathbf{Z}_n corresponde a un arreglo de n muestras de $(U, r_{\eta}(X, Y))$, siendo U la entrada o alguna de sus componentes. El segundo elemento a registrar en los experimentos desprende del *test* de independencia, para estos efectos se aplica la siguiente regla de decisión $\phi(\mathbf{Z}_n) = 1_{\{0\}}(\hat{I}(\mathbf{Z}_n))$, en donde $\phi(\mathbf{Z}_n) = 0$ implica no rechazar la hipótesis de buen modelo ($H_0 : \delta = \mathbf{0}$) y $\phi(\mathbf{Z}_n) = 1$ implica rechazar la hipótesis de buen modelo (hipótesis alternativa $H_1 : \delta \neq \mathbf{0}$); en esta línea, el tiempo de detección de *mismodelling* corresponde a lo siguiente:

$$\mathcal{T}_U = \sup\{n \in \mathbb{N} : \phi(\mathbf{Z}_n) = 0\}.$$

Debido a la carga computacional de la obtención de tiempos de detección, esta es aproximada y se explora hasta una cota máxima de cantidad de muestras. Alcanzada esta cota, se presume indetectable el *mismodelling* para el valor de δ asociado.

En las siguientes breves secciones se exponen los modelos explorados y sus resultados asociados.

Modelo lineal en los parámetros

Este modelo está determinado por la siguiente regla:

$$Y|X = (c_1 + \delta_1)X_1 + (c_2 + \delta_2)X_2 + W,$$

en donde $X = (X_1, X_2) \sim \mathcal{N}(\mu_X, \Sigma_X)$ y $W \sim \mathcal{N}(\mu_W, \sigma_W^2)$, con valores numéricos correspondientes a $\mu_X = \mathbf{0}$, $\Sigma_X = \text{diag}((1, 1))$, $\mu_W = 0$, $\sigma_W = 0.5$, $c_1 = 5$ y $c_2 = -5$. El residuo está dado por:

$$R = \delta_1 X_1 + \delta_2 X_2 + W.$$

Con este modelo se calcula la EMI para valores $\delta \in \delta_{\text{máx}}^{\text{grid}} = \{-\delta_{\text{máx}}, -0.99\delta_{\text{máx}}, \dots, \delta_{\text{máx}}\}^2$, en donde $\delta_{\text{máx}} \in \{0.1, 1, 10\}$, con 2000 muestras sobre los pares (X, R) , (X_1, R) y (X_2, R) , esto se muestra en la figura 1. Por otro lado se estima \mathcal{T}_X con una cota de muestras de 10^6 , esto se muestra en la figura 2. Para todos los experimentos, se consideran los datos $\mathbf{Z}_n = ((u[i], r[i]))_{i=1}^n$, el cual es un arreglo de muestras i.i.d. de la variable (U, R) .

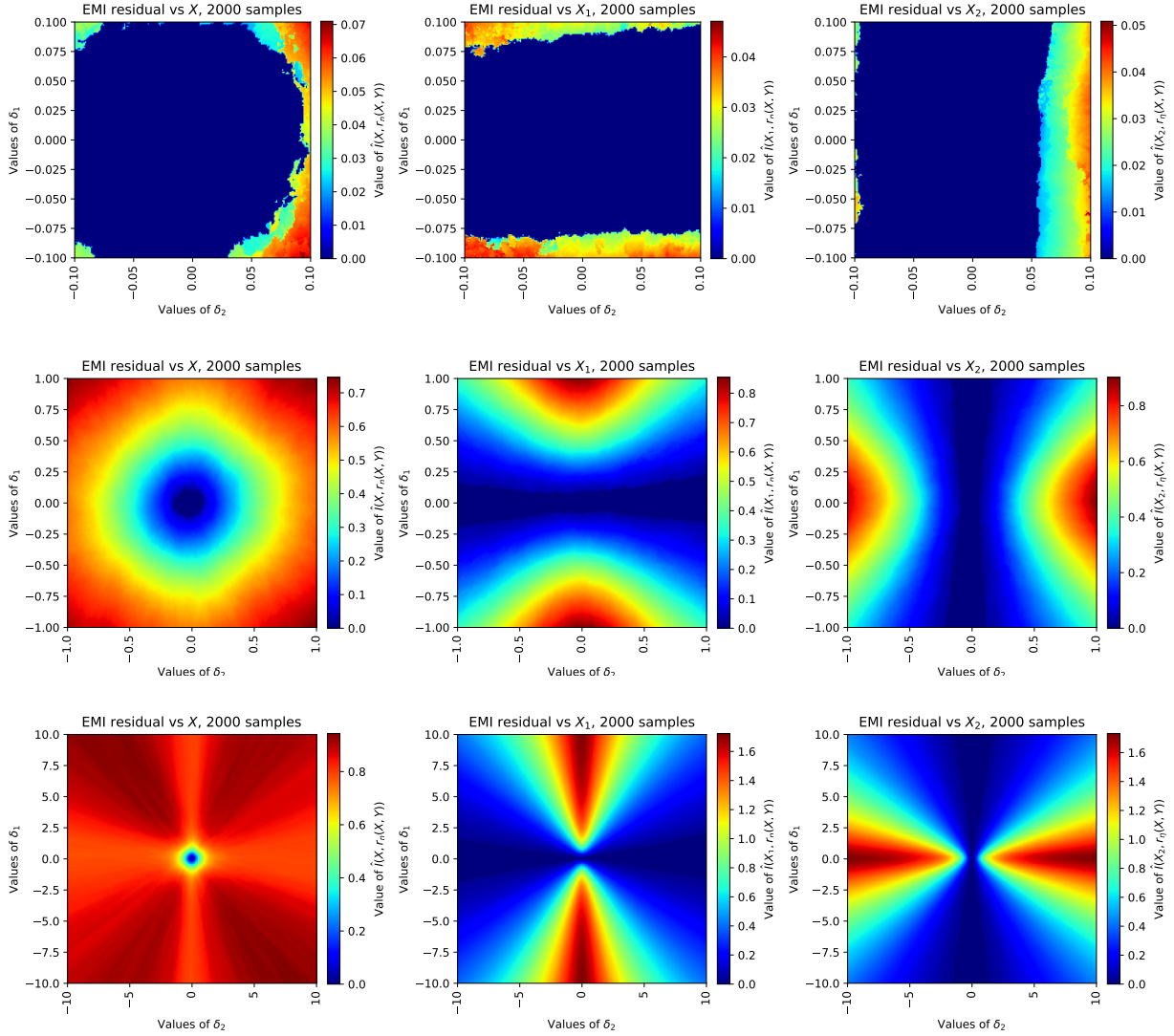


Figura 1: EMI en *mismodelling* para modelo lineal en los parámetros

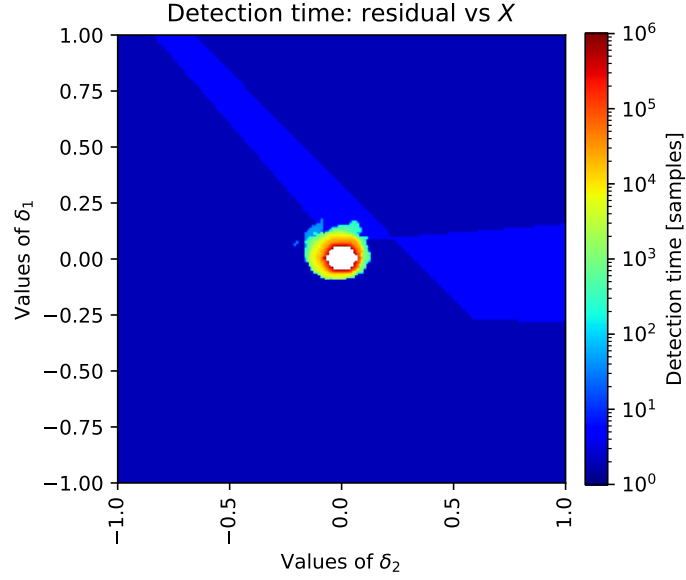


Figura 2: Matriz de tiempos de detección de *mismodelling* para modelo lineal en los parámetros

Modelo polinomial en los parámetros

Este modelo está determinado por la siguiente regla:

$$Y|X = (c_1 + \delta_1)X_1^2 + (c_2 + \delta_2)X_2^3 + W,$$

en donde $X = (X_1, X_2) \sim \mathcal{N}(\mu_X, \Sigma_X)$ y $W \sim (\mu_W, \sigma_W^2)$, con valores numéricos correspondientes a $\mu_X = \mathbf{0}$, $\Sigma_X = \text{diag}((1, 1))$, $\mu_W = 0$, $\sigma_W = 0.5$, $c_1 = 5$ y $c_2 = -5$. El residuo está dado por:

$$R = \delta_1 X_1^2 + \delta_2 X_2 + W.$$

Con este modelo se calcula la EMI para valores $\delta \in \delta_{\text{máx}}^{\text{grid}} = \{-\delta_{\text{máx}}, -0.99\delta_{\text{máx}}, \dots, \delta_{\text{máx}}\}^2$, en donde $\delta_{\text{máx}} \in \{0.1, 1, 10\}$, con 2000 muestras sobre los pares (X, R) , (X_1, R) y (X_2, R) , esto se muestra en la figura 3. Por otro lado se estima \mathcal{T}_X con una cota de muestras de 10^6 , esto se muestra en la figura 4. Para todos los experimentos se consideran los datos $\mathbf{Z}_n = ((u[i], r[i]))_{i=1}^n$, el cual es un arreglo de muestras i.i.d. de la variable (U, R) .

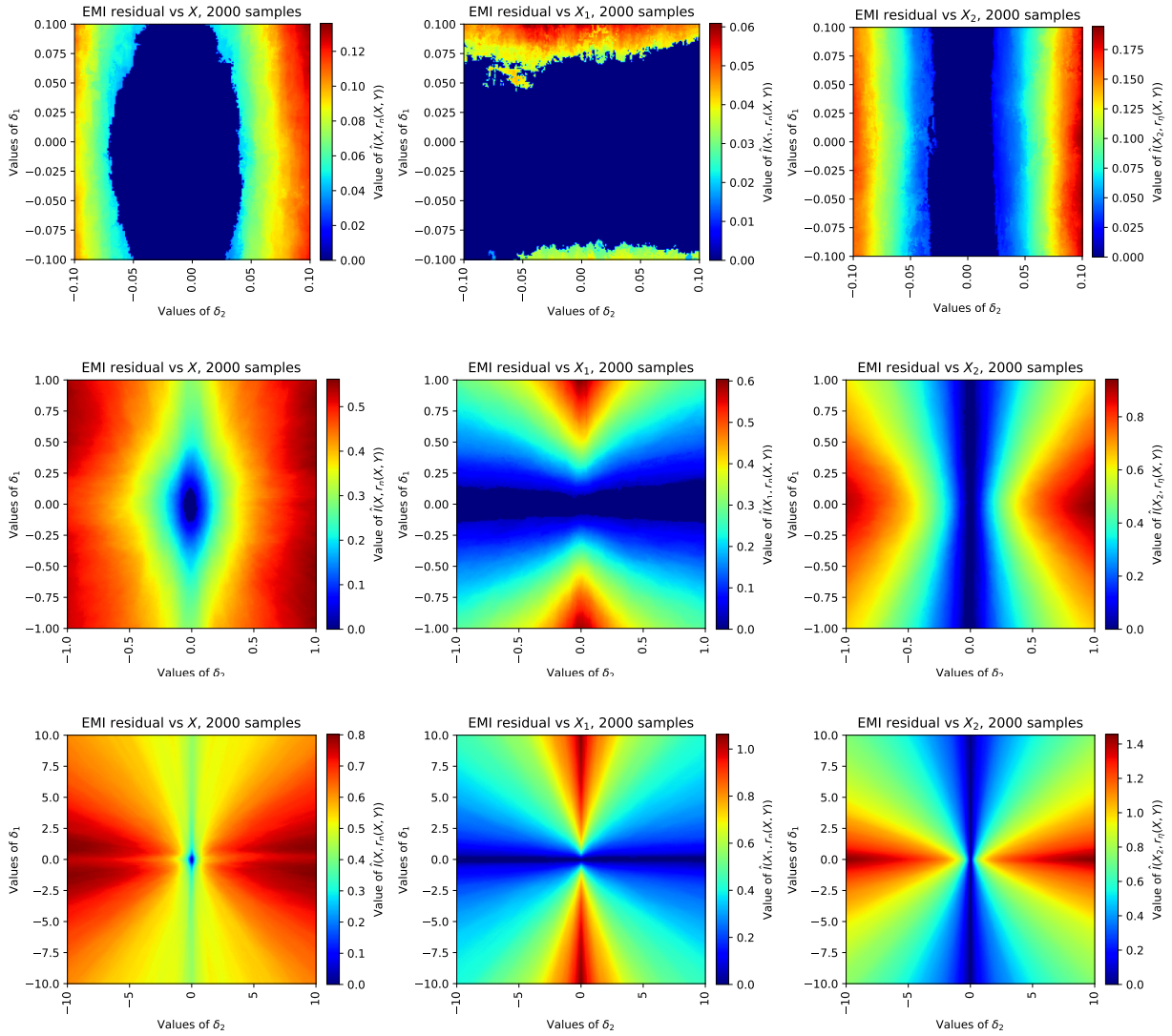


Figura 3: EMI en *mismodelling* para modelo polinomial en los parámetros

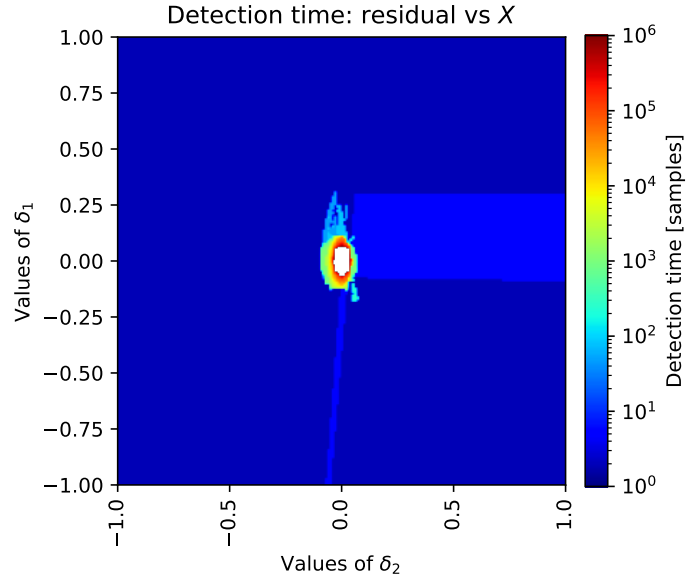


Figura 4: Matriz de tiempos de detección de *mismodelling* para modelo polinomial en los parámetros

Modelo no lineal en sus parámetros

Este modelo está determinado por la siguiente regla:

$$Y|X = (c_1 + \delta_1)X_1X_2 + (1 + X_1)^{(c_2 + \delta_2)X_2} + kW_1W_2,$$

en donde $X = (X_1, X_2)$, $W = (W_1, W_2)$, con componentes $X_1 \sim \exp(\lambda_X)$, $X_2 \sim \mathcal{N}(\mu_X, \sigma_X^2)$, $W_1 \sim \mathcal{N}(\mu_W, \sigma_W^2)$ y $W_2 \sim \text{uniforme}([a_W, b_W])$ todas independientes entre sí. Los valores numéricos corresponden a $\lambda_X = 1$, $\mu_X = 0$, $\sigma_X = 2$, $\mu_W = -1$, $\sigma_W = 5$, $a_W = -1$, $b_W = 2$ y $c_1 = c_2 = k = 1$. El residuo está dado por:

$$R = \delta_1X_1X_2 + (1 + X_1)^{(c_2 + \delta_2)X_2} - (1 + X_1)^{\delta_2X_2} + kW_1W_2.$$

Con este modelo se calcula la EMI para valores $\delta \in \delta_{\text{máx}}^{\text{grid}} = \{-\delta_{\text{máx}}, -0.99\delta_{\text{máx}}, \dots, \delta_{\text{máx}}\}^2$, en donde $\delta_{\text{máx}} \in \{0.1, 1, 10\}$, con 2000 muestras sobre los pares (X, R) , (X_1, R) y (X_2, R) , esto se muestra en la figura 5. Por otro lado se estima \mathcal{T}_X con una cota de muestras de 10^6 , esto se muestra en la figura 6. Para todos los experimentos se consideran los datos $\mathbf{Z}_n = ((u[i], r[i]))_{i=1}^n$, el cual es un arreglo de muestras i.i.d. de la variable (U, R) .

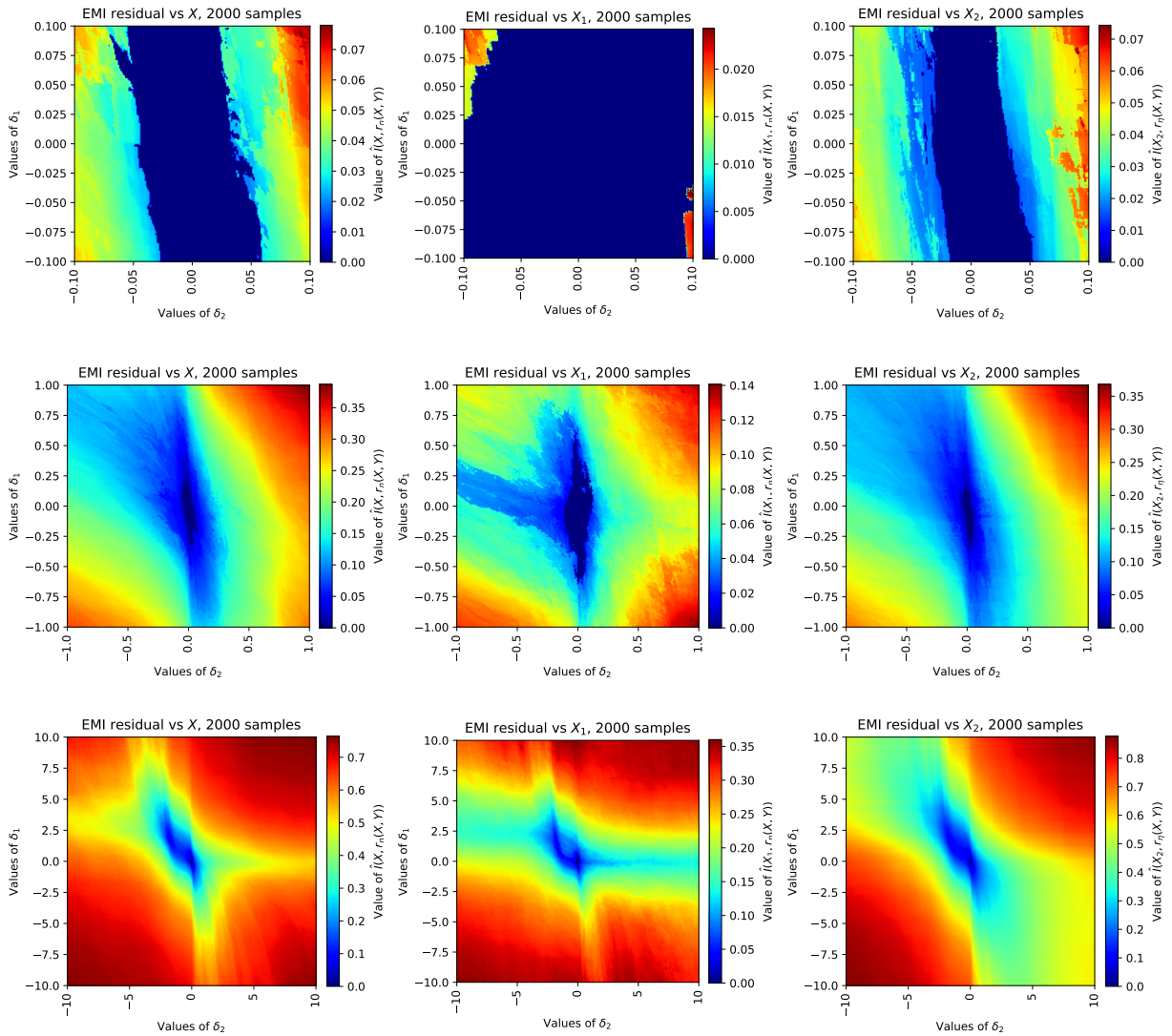


Figura 5: EMI en *mismodelling* para modelo no lineal en sus parámetros

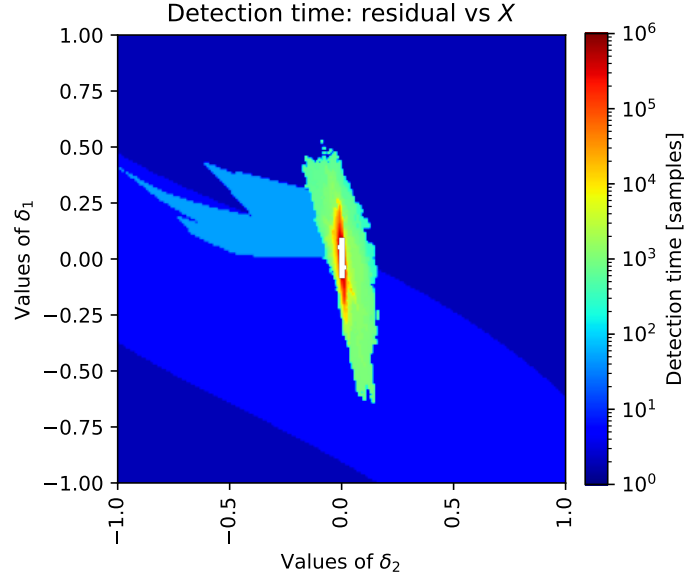


Figura 6: Matriz de tiempos de detección de *mismodelling* para modelo no lineal

Modelo lineal autorregresivo

Este modelo está determinado por la siguiente regla $\forall k \in \mathbb{N}$:

$$\begin{aligned} d_k &= (c_1 + \delta_1)d_{k-1} + (c_2 + \delta_2)u_{k-1} + v_k, \\ y_k &= d_k + w_k, \end{aligned}$$

en donde $\{u_k\}_{k \in \mathbb{N}}$, $\{v_k\}_{k \in \mathbb{N}}$ y $\{w_k\}_{k \in \mathbb{N}}$ son conjuntos de muestras i.i.d. de las variables aleatorias $u \sim \mathcal{N}(\mu_U, \sigma_U^2)$, $v \sim \mathcal{N}(\mu_V, \sigma_V^2)$ y $w \sim \mathcal{N}(\mu_W, \sigma_W^2)$ respectivamente; los coeficientes numéricos corresponden a $\mu_U = 1$, $\mu_V = \mu_W = 0$, $\sigma_U = 1$, $\sigma_V = 0.64$, $\sigma_W = 0.2$, $d_{-1} = u_{-1} = 0$, $c_1 = 0.6$ y $c_2 = 1$. El residuo está dado por:

$$r_k = \delta_1 d_{k-1} + \delta_2 u_{k-1} + v_k + w_k, \forall k \in \mathbb{N}.$$

Con este modelo se calcula la EMI para valores $\delta \in \delta_{\text{máx}}^{\text{grid}} = \{-\delta_{\text{máx}}, -0.99\delta_{\text{máx}}, \dots, \delta_{\text{máx}}\}^2$, en donde $\delta_{\text{máx}} \in \{0.1, 0.8\}$; además de $\delta \in \{-1.6, -1.59, \dots, 0.4\} \times \{-1, -0.99, \dots, 1\}$. Estas EMIs consideran como entrada $x_k = (d_{k-1}, u_{k-1})$, generándose $n = 2000$ muestras con las cuales se puede calcular:

$$\begin{aligned} \hat{I}(((x_k, r_k))_{k=1}^n) &\approx I(X, R), \\ \hat{I}(((d_{k-1}, r_k))_{k=1}^n) &\approx I(d[k-1], R), \\ \hat{I}(((u_{k-1}, r_k))_{k=1}^n) &\approx I(u[k-1], R). \end{aligned}$$

Las EMIs calculadas se pueden ver en la figura 7. Por otro lado, se estima \mathcal{T}_X , donde X es la variable aleatoria cuyas muestras son $(x_k)_{i=1}^n$, con una cota de muestras de 10^6 , esto se puede ver en la figura 8.

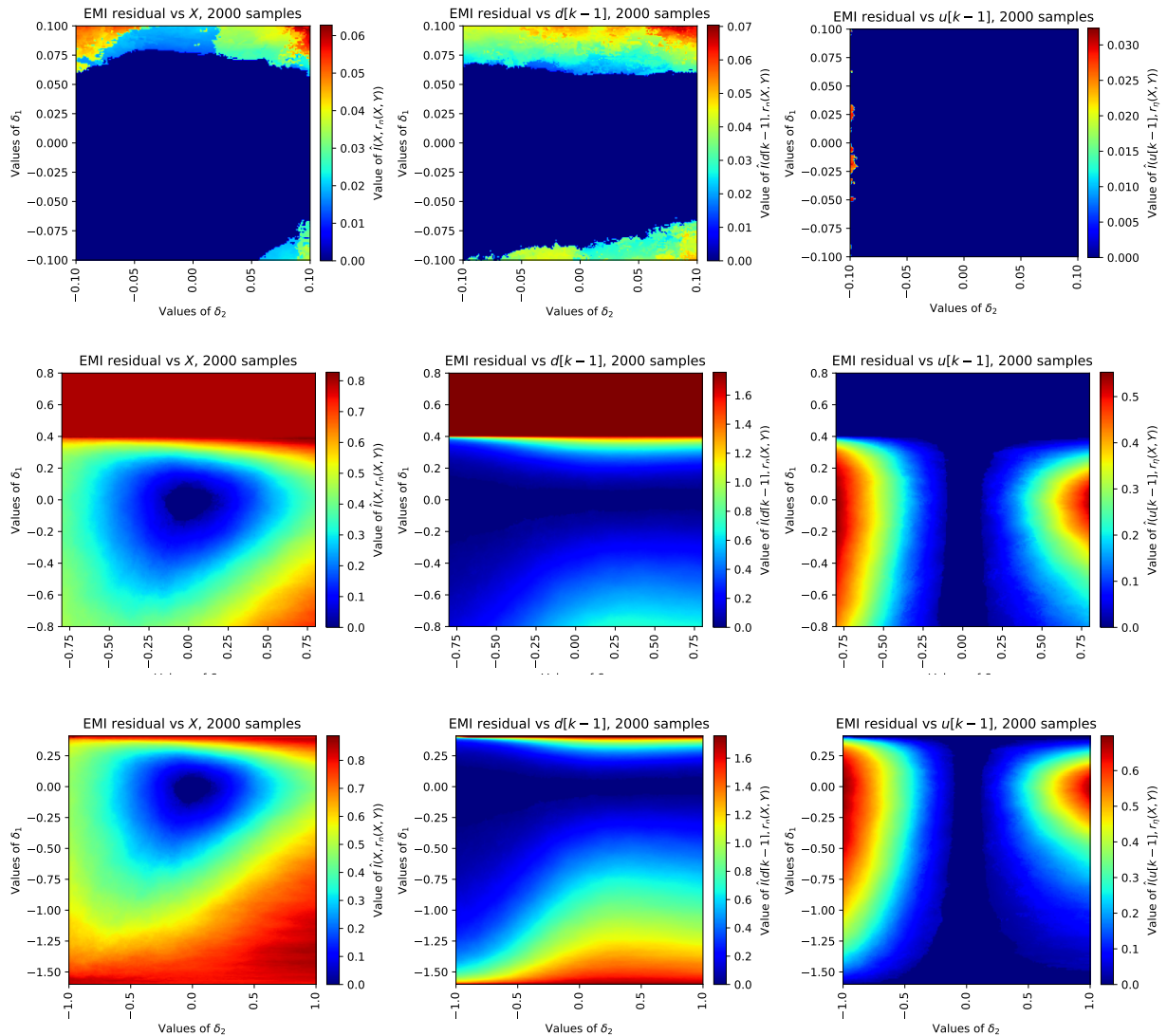


Figura 7: EMI en *mismodelling* para modelo lineal autorregresivo

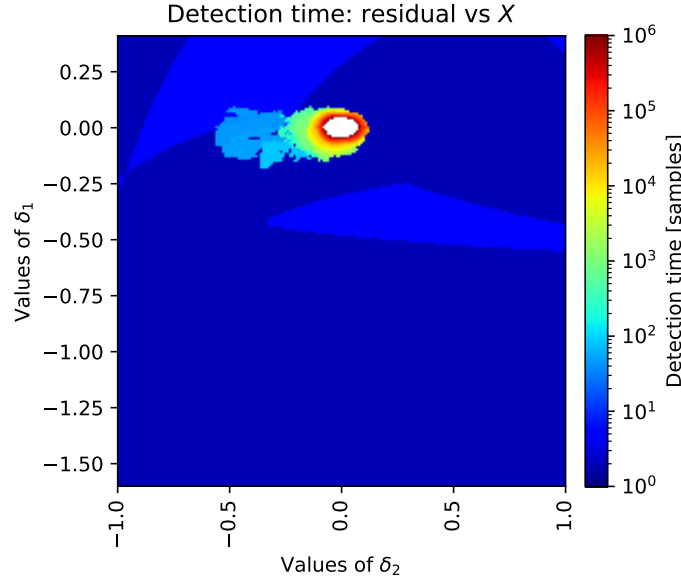


Figura 8: Matriz de tiempos de detección de *mismodelling* para modelo lineal autorregresivo

Modelo no lineal autorregresivo

Este modelo está determinado por la siguiente regla $\forall k \in \mathbb{N}$:

$$d_k = (c_1 + (c_2 + \delta_1) \exp(-d_{k-1}^2)) d_{k-1} + (c_3 + (c_4 + \delta_2) \exp(-d_{k-1}^2)) d_{k-2} + c_5 u_{k-1} + c_6 u_{k-2} \\ + c_7 u_{k-1} u_{k-2} + c_8 \exp(-d_{k-1}^2) + v_k, \\ y_k = d_k + w_k,$$

en donde $\{u_k\}_{k \in \mathbb{N}}$, $\{v_k\}_{k \in \mathbb{N}}$ y $\{w_k\}_{k \in \mathbb{N}}$ son conjuntos de muestras i.i.d. de las variables aleatorias $u \sim \mathcal{N}(\mu_U, \sigma_U^2)$, $v \sim \mathcal{N}(\mu_V, \sigma_V^2)$ y $w \sim \mathcal{N}(\mu_W, \sigma_W^2)$ respectivamente; los coeficientes corresponden a $\mu_U = 1$, $\mu_V = \mu_W = 0$, $\sigma_U = 1$, $\sigma_W = 0.64$, $\sigma_V = 0.2$, $d_{-1} = d_{-2} = u_{-1} = u_{-2} = 0$, $c_1 = 0.8$, $c_2 = -0.5$, $c_3 = -0.3$, $c_4 = -0.9$, $c_5 = 1$, $c_6 = 0.2$, $c_7 = 0.1$ y $c_8 = 0.5$. El residuo está dado por:

$$r_k = (\delta_1 d_{k-1} + \delta_2 d_{k-2}) \exp(-d_{k-1}^2) + v_k + w_k, \forall k \in \mathbb{N}.$$

Con este modelo se calcula la EMI para valores $\delta \in \delta_{\text{máx}}^{\text{grid}} = \{-\delta_{\text{máx}}, -0.99\delta_{\text{máx}}, \dots, \delta_{\text{máx}}\}^2$, en donde $\delta_{\text{máx}} \in \{0.1, 1, 10, 100\}$. Estas EMIs consideran como entrada $x_k = (d_{k-1}, d_{k-2}, u_{k-1}, u_{k-2})$, generándose $n = 2000$ muestras con las cuales se puede calcular:

$$\begin{aligned} \hat{I}((x_k, r_k))_{k=1}^n &\approx I(X, R), \\ \hat{I}((d_{k-1}, r_k))_{k=1}^n &\approx I(d[k-1], R), \\ \hat{I}((d_{k-2}, r_k))_{k=1}^n &\approx I(d[k-2], R), \\ \hat{I}((u_{k-1}, r_k))_{k=1}^n &\approx I(u[k-1], R), \\ \hat{I}((u_{k-2}, r_k))_{k=1}^n &\approx I(u[k-2], R). \end{aligned}$$

Las EMIs calculadas se pueden ver en las figuras 9 – 12. Por otro lado, se estiman \mathcal{T}_X y $\mathcal{T}_{d[k-1]}$, donde X y $d[k-1]$ son las variables aleatorias cuyas muestras son $(x_k)_{k=1}^n$ y $(d_{k-1})_{k=1}^n$ respectivamente, con una cota de muestras de 10^6 sobre valores de $\delta_{\text{máx}} \in \{10, 100\}$, esto se puede ver en la figura 13.

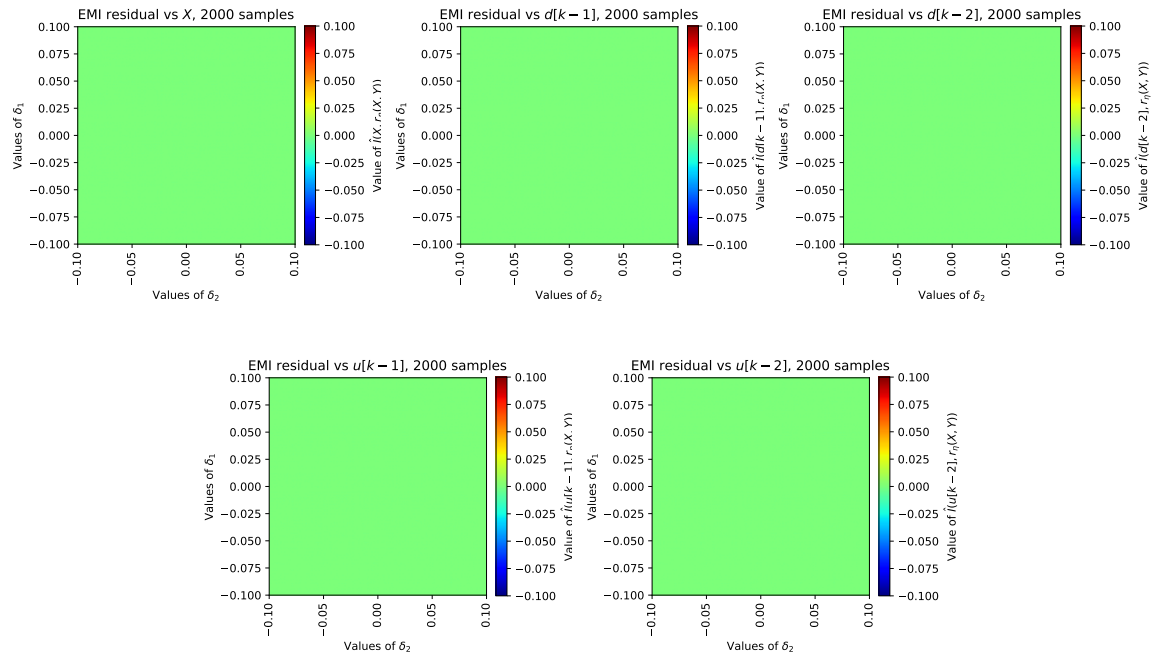


Figura 9: EMI en *mismodelling* para modelo no lineal autorregresivo ($\delta_{\text{máx}} = 0.1$)

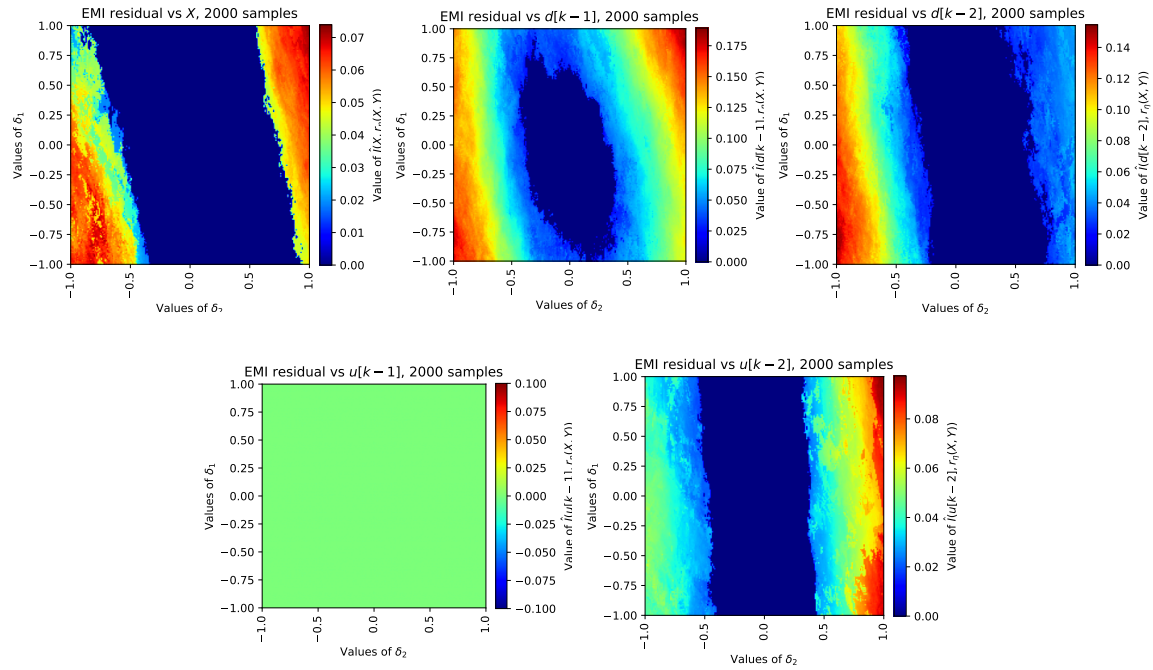


Figura 10: EMI en *mismodelling* para modelo no lineal autorregresivo ($\delta_{\text{máx}} = 1$)

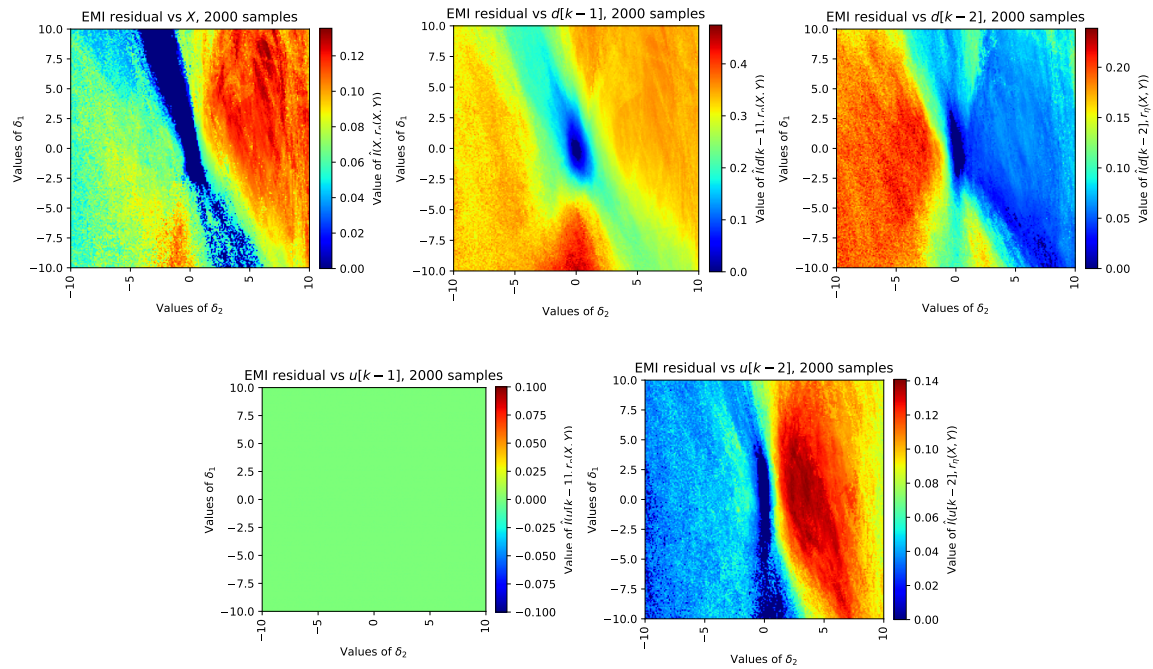


Figura 11: EMI en *mismodelling* para modelo no lineal autorregresivo ($\delta_{\text{máx}} = 10$)

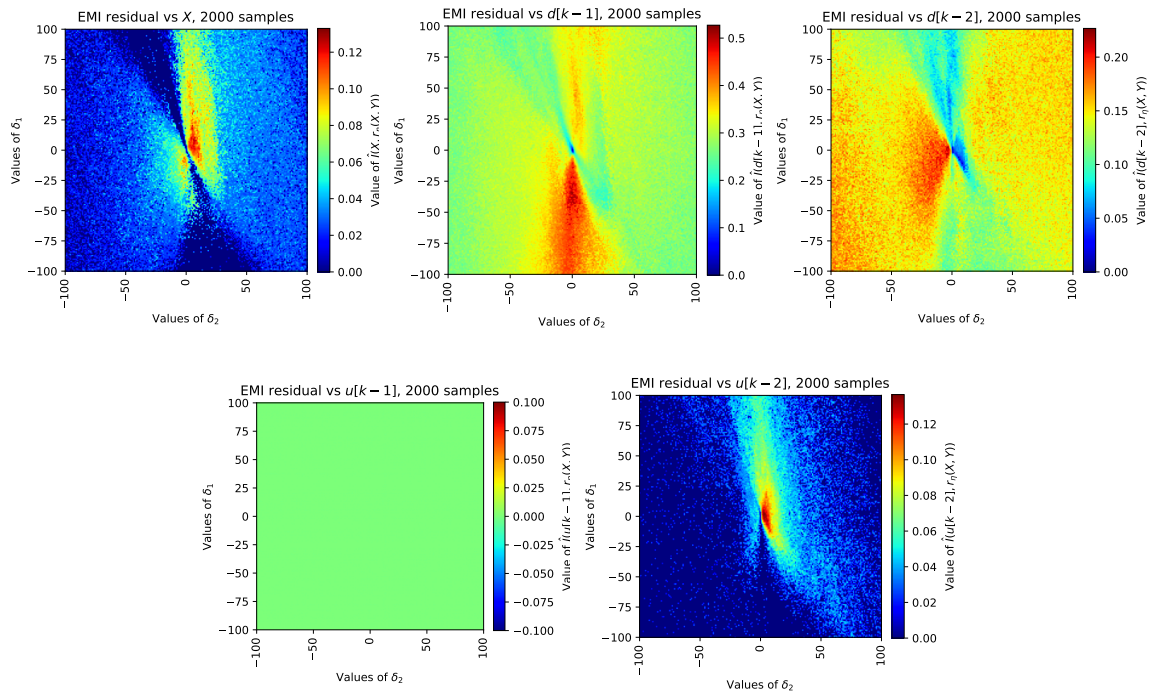


Figura 12: EMI en *mismodelling* para modelo no lineal autorregresivo ($\delta_{\text{máx}} = 100$)

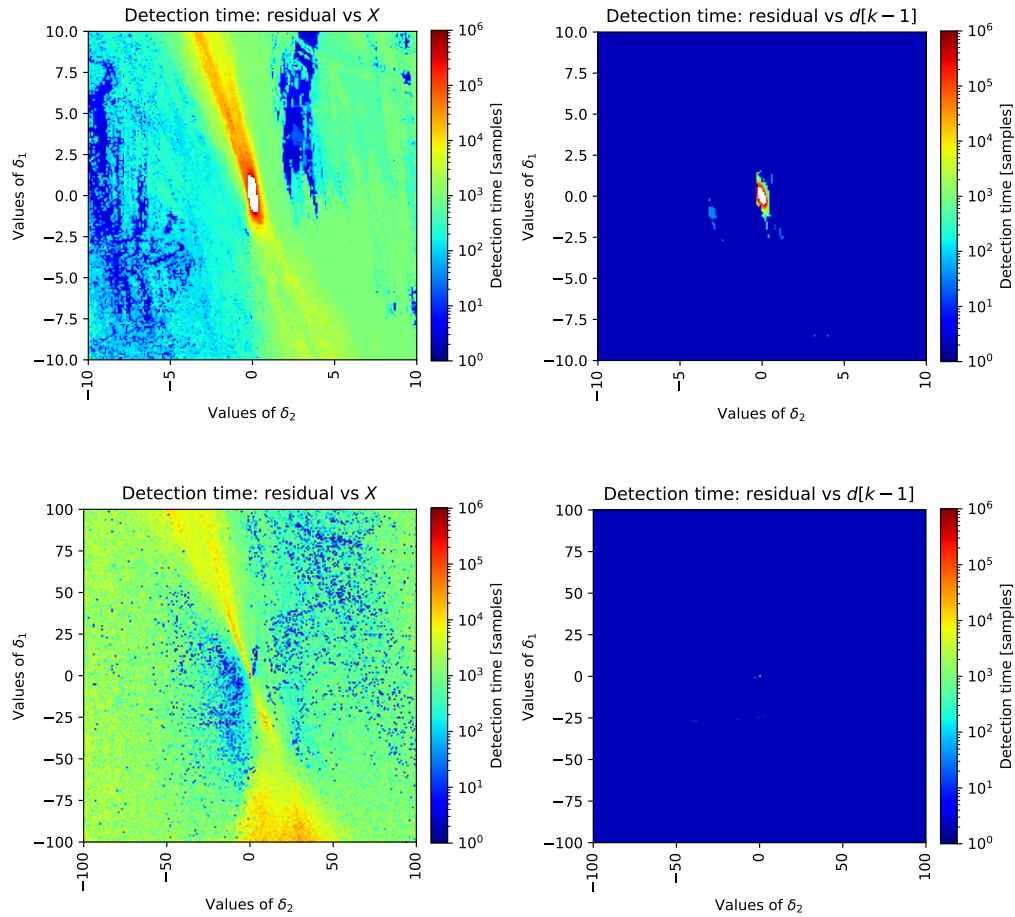


Figura 13: Matrices de tiempos de detección de *mismodelling* para modelo no lineal autorregresivo

Modelo de estanque cónico¹

La versión a tiempo finito de este modelo está determinada por la siguiente regla $\forall k \in \mathbb{N}$:

$$d_k = \frac{(c_1 + \delta_1)|x_{k-1}| + c_2 - (c_3 + \delta_2)\sqrt{d_{k-1}}}{c_4 d_{k-1}^2 + c_5 d_{k-1} + c_6} + d_{k-1},$$

$$y_k = d_k + w_k,$$

en donde $\{x_k\}_{k \in \mathbb{N}}$ y $\{w_k\}_{k \in \mathbb{N}}$ son conjuntos de muestras i.i.d. de las variables aleatorias dadas por $x \sim \mathcal{N}(\mu_X, \sigma_X^2)$ y $w \sim \mathcal{N}(\mu_W, \sigma_W^2)$ respectivamente; los coeficientes corresponden a $\mu_X = 35$, $\sigma_X = 0.1$, $\mu_W = 0.0$, $\sigma_W = 0.2$, $y_{-1} = 25$, $c_1 = 5.43$, $c_2 = -78.23$, $c_3 = 20.21$, $c_4 = 0.63$, $c_5 = 11.4$ y $c_6 = 17.1$. El residuo está dado por:

$$r_k = \frac{\delta_1|x_{k-1}| - \delta_2\sqrt{d_{k-1}}}{c_4 d_{k-1}^2 + c_5 d_{k-1} + c_6} + w_k, \forall k \in \mathbb{N}.$$

Con este modelo se calcula la EMI para valores δ en los siguientes conjuntos:

$$\begin{aligned} &\{-0.1, -0.999, \dots, 0.999, 1.0\}^2, \\ &\{-1, -0.99, \dots, 0.99, 1\}^2, \\ &\{-5, -4.9, \dots, 14.9, 15\} \times \{-10, -9.9, \dots, 9.9, 10\}, \\ &\{-5, -4, \dots, 194, 195\} \times \{-20, -19, \dots, 179, 180\}. \end{aligned}$$

Estas EMIs consideran como entrada $u_k = (x_{k-1}, y_{k-1})$, generándose $n = 2000$ muestras con las cuales se puede calcular:

$$\begin{aligned} \hat{I}((u_k, r_k))_{k=1}^n &\approx I(U, R) \\ \hat{I}((x_{k-1}, r_k))_{k=1}^n &\approx I(x[k-1], R) \\ \hat{I}((y_{k-1}, r_k))_{k=1}^n &\approx I(y[k-1], R) \end{aligned}$$

Las EMIs calculadas se pueden ver en la figura 14. Adicional a lo mencionado, debido a que el residuo está compuesto de dos términos, uno ruidoso y otro no ruidoso, donde el primero domina al segundo en órdenes de magnitud, se repite el experimento, pero esta vez ajustando $\sigma_W \in \{10^0, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$; los resultados de estos experimentos se pueden ver en la figura 15, en esta figura los mapas de la fila i -ésima se calculan con $\sigma_W = 10^{-(i-1)}$. Finalmente, en la figura 16 se muestra un experimento efectuado para $\sigma_W = 10^{-4}$ y $\delta \in \{-0.1, -0.999, \dots, 0.999, 0.1\}^2$.

¹Jáuregui Salinas, C. (2016). Evaluación de estrategias de sintonización de controladores fraccionarios para planta no lineal: sistema de estanques. Disponible en <https://repositorio.uchile.cl/handle/2250/140963>

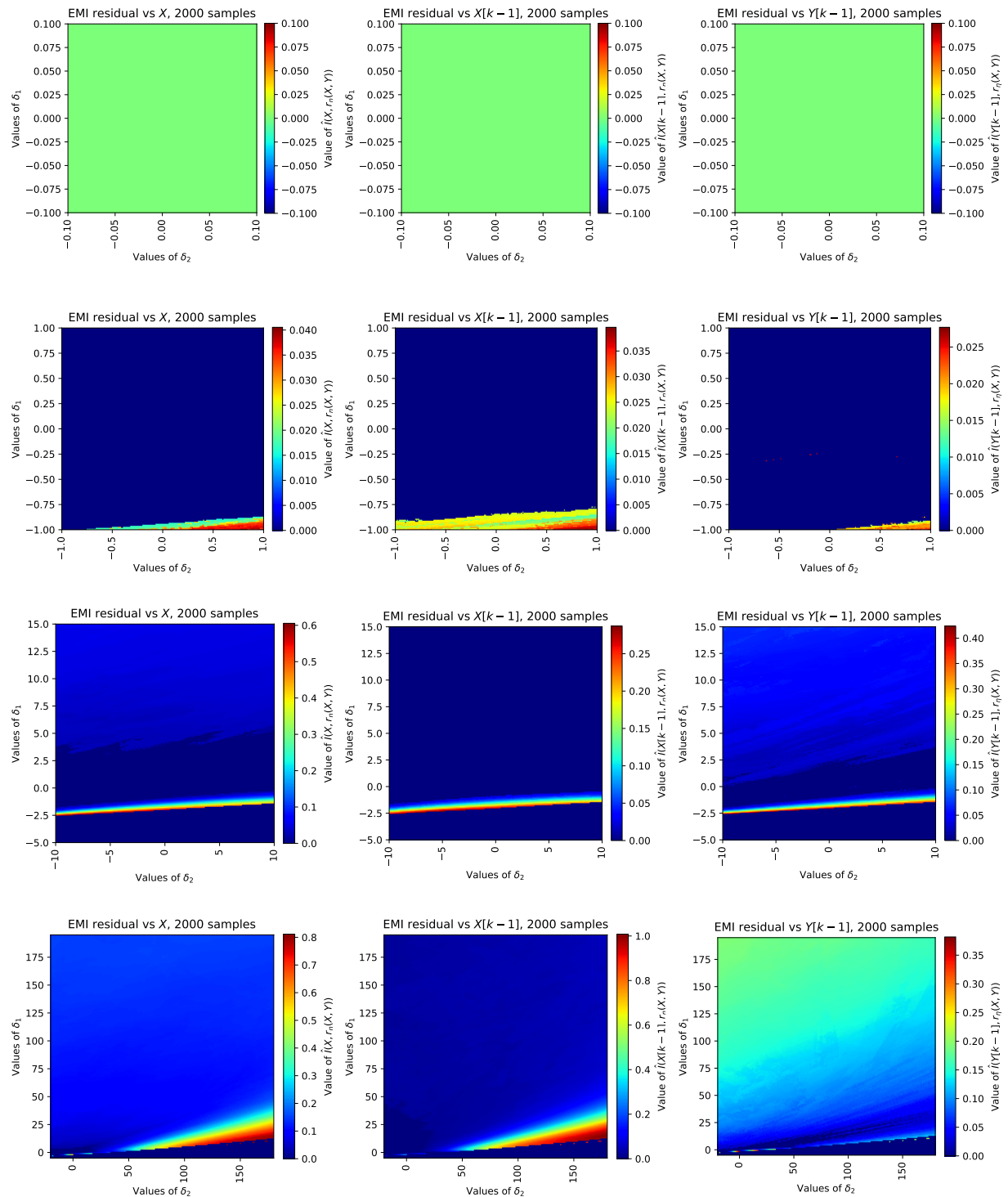


Figura 14: EMI en *mismodelling* para modelo de estanque cónico

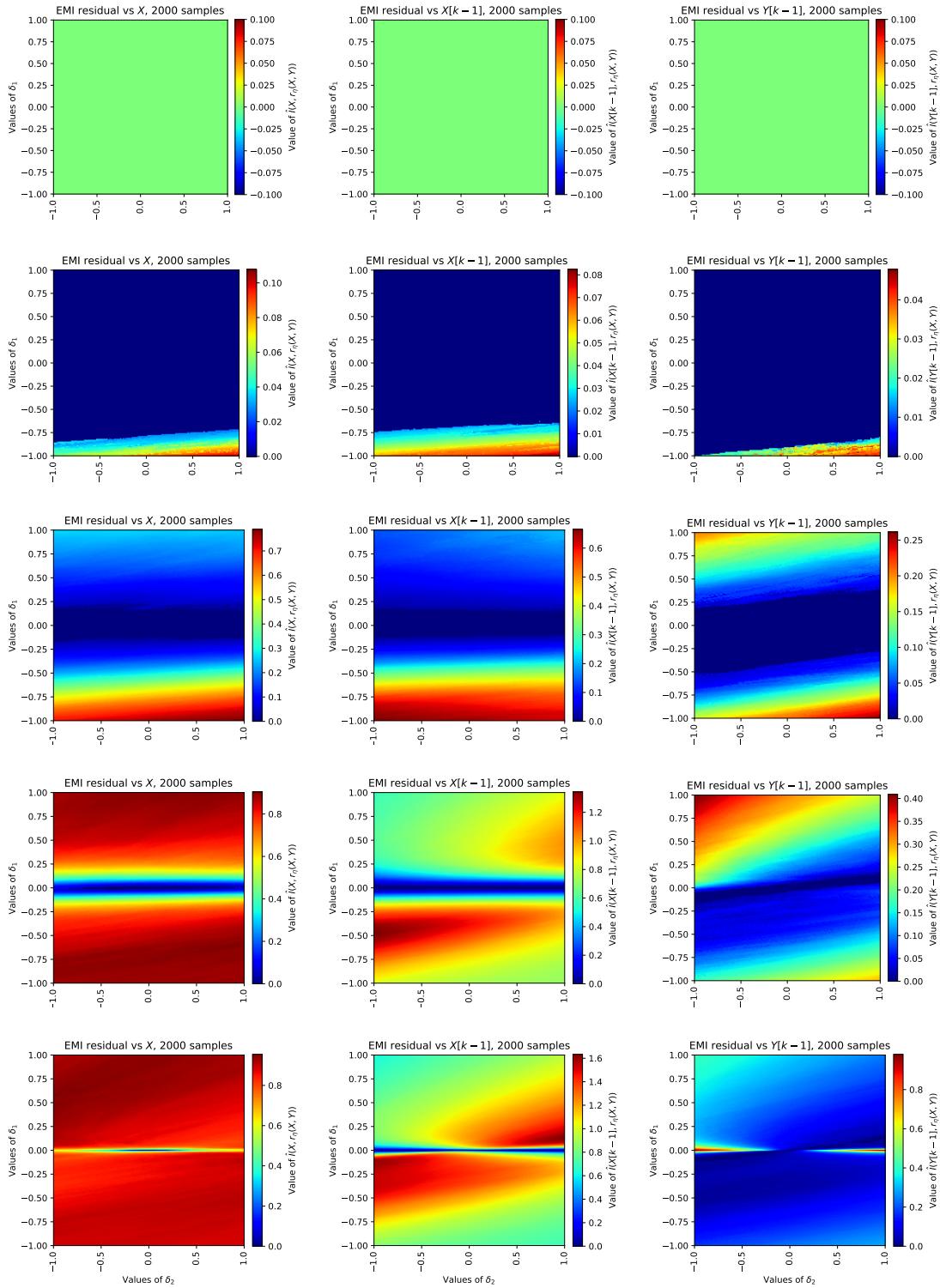


Figura 15: EMI en *mismodelling*, estanque cónico; $\sigma_W = 10^{-(i-1)}$, i el número de fila del mapa

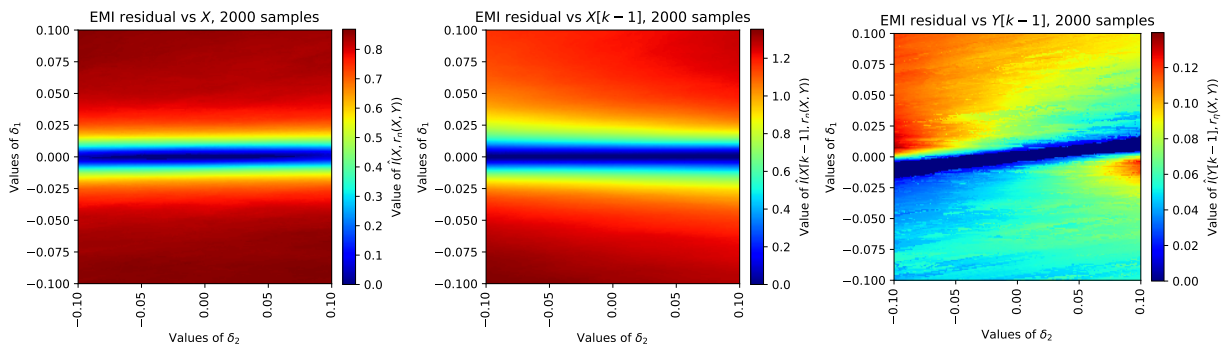


Figura 16: EMI en *mismodelling* para modelo de estanque cónico con $\sigma_W = 10^{-4}$, valores finos de δ