



Estimación de la producción de energía hidroeléctrica mediante modelamiento estadístico con (GAMLSS)



Edwin Fernando Barrientos Arroyave
Heber Esteban Bermúdez González

Universidad Nacional de Colombia
efbarrientosa@unal.edu.co, hebermudezg@unal.edu.co

Aquí podrás encontrar la base de datos y el código R para replicar los resultados.

Resumen

En este póster se presenta la aplicación del modelamiento estadístico usando la metodología de Modelos Aditivos Generalizados de Localización Escala y Forma **GAMLSS**, para estudiar la generación de energía hidroeléctrica. La metodología es aplicada para investigar como los efectos de variación climática como el fenómeno del niño y la variación de la precipitación pluvial afecta la producción de energía hidroeléctrica.

1. Introduction

En este estudio se comprueba cómo afecta la variación de la precipitación pluvial medida en dos estaciones meteorológicas (**Esta1**, **Esta2**) y las anomalías climáticas como el fenómeno del niño y el fenómeno de la niña dados por el Índice Oceanográfico del Niño (**ONI**), en la generación de energía hidroeléctrica, la base de datos se obtuvo de un estudio observacional de las Empresas Públicas de Medellín (**EPM**) de siete centrales hidroeléctricas desde el año 1996 hasta 2016 y dos estaciones meteorológicas, por otro lado el Índice Oceanográfico del Niño (**ONI**) fue obtenido de la página oficial de la Administración Nacional Oceánica y Atmosférica (**NOAA**), todas las variables aquí involucradas son cuantitativas.

Características de las variables:

Variable	Descripción
Generación	Anual(mayo-abril) en (GWh)
Esta1	Precipitación en (mm)
Esta2	Precipitación en (mm)
ONI	Anomalía en °C

Tabla 1: Resumen de variables involucradas en el estudio

Las dimensiones de la base de datos que se estudio es de (21 x 4), 21 observaciones por cada variable (**Generación**, **Esta1**, **Esta2**, **ONI**), que corresponde a cada año desde 1996 hasta 2016.

2. GAMLSS

A diferencia del modelo de regresión lineal estándar, la metodología **GAMLSS** propuesta por Rigby y Stasinopoulos (2005) permite suponer casi cualquier distribución paramétrica para la variable de respuesta y los parámetros de la distribución pueden variar de acuerdo con las variables explicativas como funciones lineales, no lineales o funciones suaves de estas.

3. Análisis descriptivo

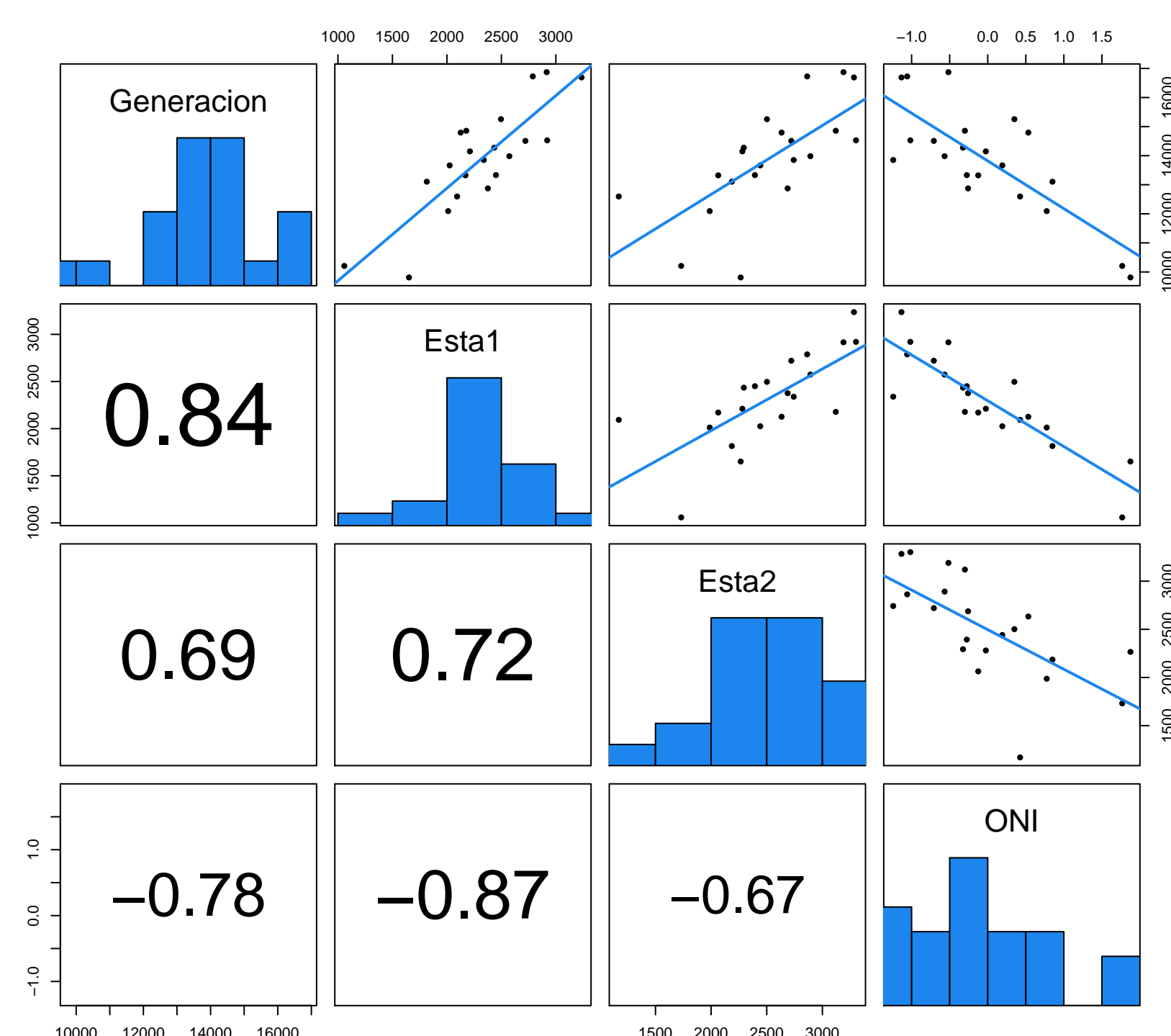


Gráfico 1: Gráfico de dispersión y correlación

En el gráfico se puede apreciar una posible relación lineal positiva entre la generación y la precipitación de las dos estaciones(**Esta1**, **Esta2**), mientras que la posible relación lineal entre la generación y el **ONI** es alta pero negativa. Se realizó un análisis de colinealidad entre las variables haciendo uso de la paquete `mctest()` del programa estadístico **R**, arrojando colinealidad entre las variables involucradas pero se decidió no eliminar ningún regresor para evitar problemas de especificación del modelo.

4. Metodología

En primera instancia mediante una prueba de bondad de ajuste se busca la distribución paramétrica que mejor se ajusta a la distribución marginal de la variable respuesta Generación (GWh), este procedimiento fue desarrollado con el software estadístico **R** haciendo uso de la función `fitDist()` del paquete `gamlss`. Los resultados se muestran a continuación:

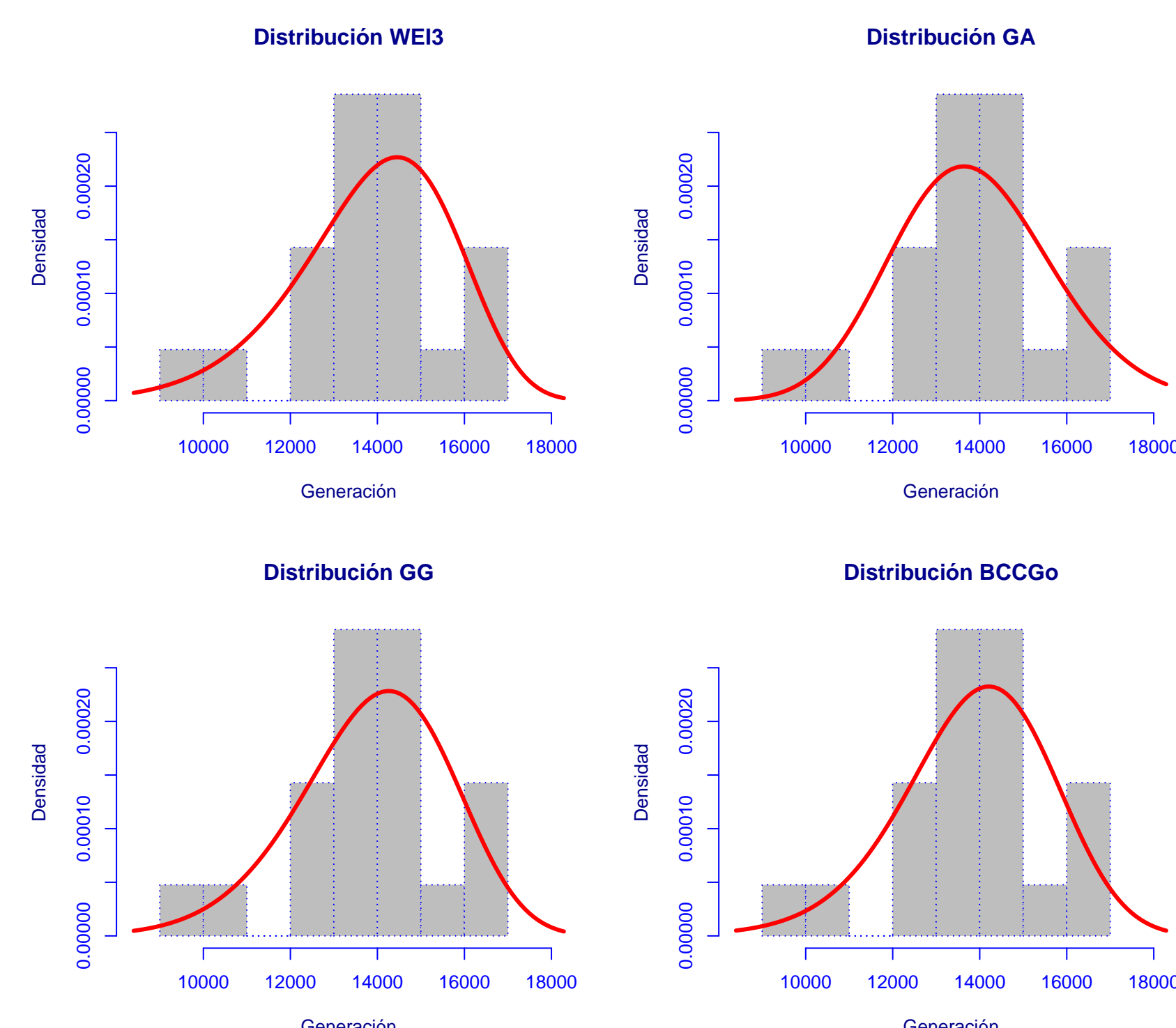


Gráfico 2: Prueba de bondad de ajustes, mejores resultados

Dentro del desarrollo del estudio se propusieron modelos

GAMLSS con las distribuciones obtenidas en la prueba de bondad de ajuste, se propusieron términos lineales, cuadráticos e interacciones entre las variables para luego ser evaluadas mediante un proceso de selección de variables el cual se desarrollo haciendo uso de la función `stepGAICall.A()` de paquete `gamlss`.

Después de seleccionar las variables y corroborar la significancia de las mismas, se ajustan cuatro modelos con las distribuciones mencionadas, los resultados se muestran a continuación:

	Distribución	R ²	MSE	$\rho_{y,\hat{y}}$
modelo 1	WEI3	0.978	735394.0	0.883
modelo 2	GG	0.832	598435.5	0.901
modelo 3	BCCGo	0.947	1141583.4	0.895
modelo 4	GA	0.891	628666.1	0.900

Tabla 2: Resumen, coef. determinación, Error C. medio y correlación

Se uso el Criterio Bayesiano de Schwarz (SBC) (Schwarz Bayesian Criterion) para comparar los modelos construidos, este permitió medir la calidad relativa de los modelos, la expresión para calcular el indicador (SBC) es: $GAIC = -2\hat{l} + \alpha df$ con $\alpha = \log(n)$, donde n es el número de observaciones del modelo, \hat{l} corresponde al valor de log-verosimilitud del modelo y df corresponde al número de parámetros estimados del modelo.

La siguiente tabla muestra el SBC de los cuatro mejores modelos encontrados.

	Distribución	Grados de libertad	SBC	num. Parametros
modelo 1	WEI3	11	326.5119	2
modelo 2	GG	11	345.3574	3
modelo 3	BCCGo	9	354.3411	3
modelo 4	GA	6	355.8511	2

Tabla 3.

5. Resultados

Dado los resultados encontrados en las tablas anteriores y luego de analizar normalidad de los errores para cada modelo se decide descartar aquellos modelos que violan los supuestos de normalidad de los errores y por tanto se elige se elige como mejor modelo el **Weibull** con parametrización 3, dado que el modelo ajustado con esta distribución tiene mayor coeficiente de determinación R^2 , menor Criterio Bayesiano de Schwartz, correcta normalidad de los errores y menor número de parámetros.

A Continuación, se muestra el gráfico de gusano para validar que los residuales del modelo tengan un buen comportamiento, estos estan ubicados en la parte central y no invaden las hipérbolas demarcadas, lo cual indica un buen ajuste del modelo encontrado.

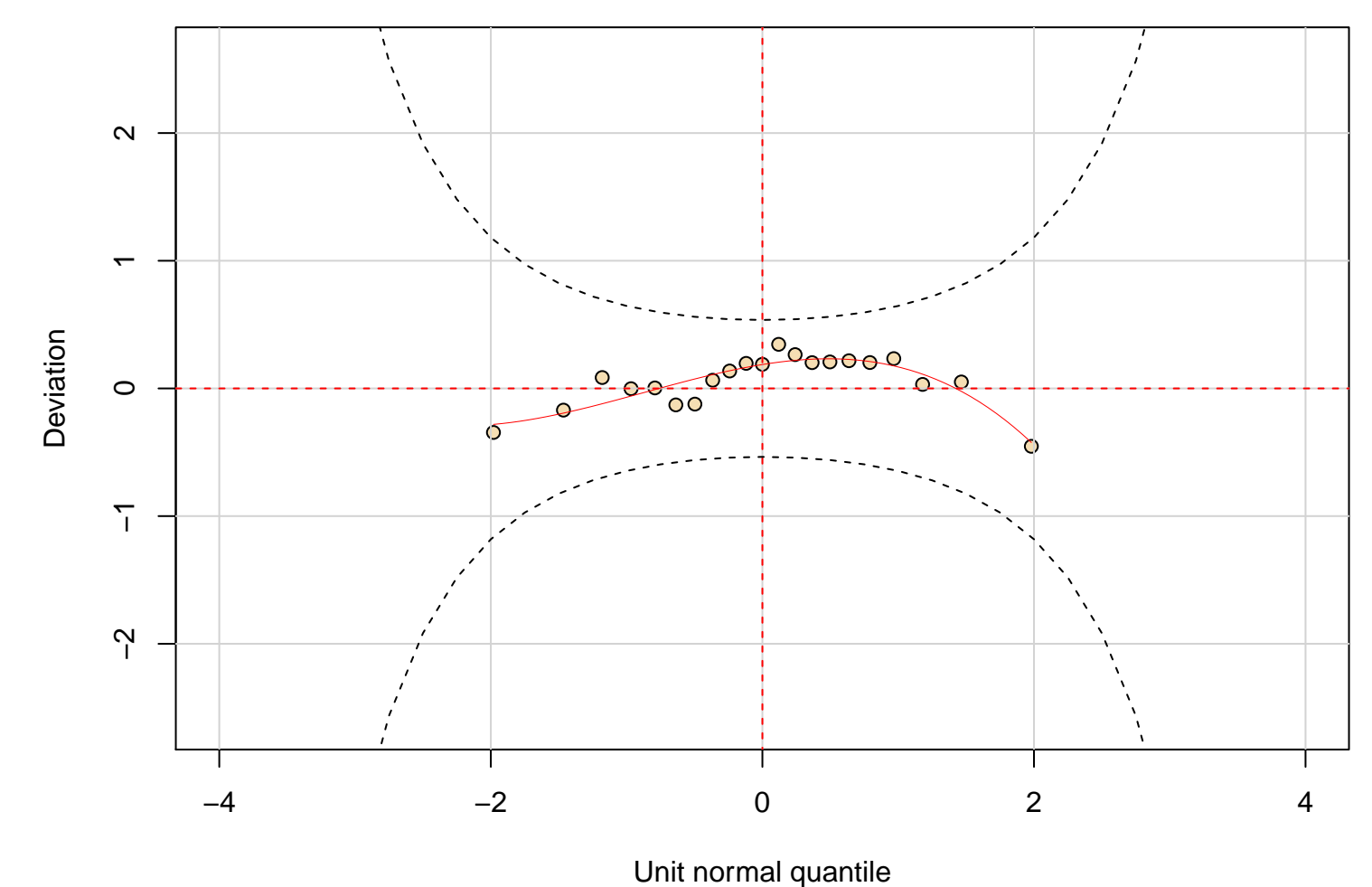


Gráfico de gusano del modelo WEI3

A continuación se presentan los parámetros estimados para el modelo encontrado, el cual tiene una distribución (WEI3), se puede observar que cada variable es significativa a un nivel del 5 %

Modelo para $\log(\mu)$	Estimado	Error estándar	Valor t	Valor-P
Intercepto	9.216	1.266e-02	728.028	<2e-16
Esta1	6.852e-05	7.557e-06	9.067	6.38e-08
Esta2	7.203e-05	2.827e-06	25.481	5.54e-15
(ONI) ³	-3.306e-02	1.730e-03	-19.116	6.26e-13

Tabla 4: estimación de parámetros

Modelo para $\log(\sigma)$	Estimado	Error estándar	Valor t	Valor-P
Intercepto	1.214e+02	9.014	13.463	2.11e-09
Esta2	-1.174e-01	1.246e-02	-9.418	1.95e-07
(Esta2) ²	4.592e-05	5.923e-06	7.754	1.97e-06
(Esta2) ³	-5.913e-09	8.859e-10	-6.674	1.05e-05
(Esta1) ²	-9.769e-06	8.579e-07	-11.386	1.83e-08
(ONI) ³	-6.151e-01	1.290e-01	-4.770	0.000299
(Esta1) ³	2.561e-09	2.285e-10	11.211	2.23e-08

Tabla 5: estimación de parámetros

Las expresiones para la estimación de los parámetros μ y σ son las siguientes considerando la función de enlace de la distribución WEI3 que es $\log()$:

$$\log(\hat{\mu}) = 9.216 + 6.852e^{-05}Esta1 + 7.203e^{-05}Esta2 - 0.03306ONI^3 \quad (1)$$

$$\log(\hat{\sigma}) = 1.214e^{+02} - 0.1174Esta2 + 4.592e^{-05}Esta2^2 - 5.913e^{-09}Esta2^3 - 9.769e^{-06}Esta1^2 - 0.6151ONI^3 + 2.561e^{-09}Esta1^3 \quad (2)$$

6. Conclusiones

La distribución WEI3 con parámetros μ y σ usada en los modelos GAMLSS tiene un valor esperado $\hat{E}(Y) = \mu$ y la varianza $\widehat{Var}(Y) = \hat{\mu}^2 \frac{\Gamma(\frac{1}{\hat{\sigma}} + 1)}{\Gamma(\frac{1}{\hat{\sigma}} + 1)^2} - 1$, De esta forma, el valor esperado y la varianza estimada para el generación de energía se escribe como:

$$\hat{E}(Y) = \hat{\mu} = \exp(9.216 + 6.852e^{-05}Esta1 + 7.203e^{-05}Esta2 - 0.03306ONI^3)$$

$$\widehat{Var}(Y) = \hat{\mu}^2 \frac{\Gamma(\frac{1}{\hat{\sigma}} + 1)}{\Gamma(\frac{1}{\hat{\sigma}} + 1)^2} - 1$$

donde $\hat{\sigma}$ es igual a la exponencial de la ecuación(2) y $\hat{\mu}$ es igual a la exponencial de la ecuación (1).

De las expresiones anteriores se puede observar el valor esperado de la Generación está en función de la variación de la precipitación de las dos estaciones meteorológicas (**Esta1**, **Esta2**) y del índice **ONI** lo cual nos indica que a mayor precipitación pluvial se espera mayor generación de energía hidroeléctrica y dado la estimación del parámetro asociado al índice **ONI** se espera que a mayor **ONI** menor generación de energía hidroeléctrica si se toman las otras variables fijas.

7. Referencias

- Montgomery, D.C ; Peck, E.A; Vining, G.G (2002). *Introducción al análisis de regresión lineal*.
- Rigby, B.; Stasinopoulos, E (2005). *Generalized additive models for location scale and shape. Applied Statistics*
- RStudio Team: (2018). *Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.*