

Informe sobre la deducción de reglas de gestión a partir de las observaciones y la inclusión de estas en SASER

Proyecto HUMID¹, Entregable 6a.

Autor(es)	Anaïs Barella Ortiz Pere Quintana Seguí Luis Garrote
Fecha versión inicial	04/09/2019
Fecha versión final	08/10/2019

Resumen

Entre los objetivos del proyecto HUMID, se encuentra mejorar la cadena de modelización SASER. Con este fin se ha acoplado un módulo de simulación de embalses a partir de reglas de gestión. El presente documento se centra en la deducción de estas reglas a partir de la generación de modelos mensuales utilizando la técnica de la regresión lineal múltiple.

¹ HUMID es un proyecto financiado por el Ministerio de Ciencia, Innovación y Universidades del Gobierno de España (código CGL2017-85687-R).

1. Introducción

Uno de los objetivos del proyecto HUMID es la mejora de SASER (SAFRAN & SURFEX & Eau-Dyssée & RAPID), un modelo hidrológico distribuido y de base física para la península ibérica. Se introducirán los embalses en el modelo con el fin de analizar su impacto en los procesos de sequía. Para ello, se elaborarán modelos que reproduzcan las reglas de gestión de algunos de los principales embalses de la cuenca del Ebro.

1.1. Estado del arte

En primer lugar, se hace referencia a artículos que resumen el estado del arte en lo referente a modelos operativos de gestión y los distintos métodos empleados. [Yeh (1985)] divide estos métodos en cuatro grupos dependiendo de las características del embalse, la disponibilidad de datos, los objetivos y las limitaciones. Estos grupos son:

- Programación Lineal (PL)
- Programación Dinámica (PD)
- Programación No Lineal (PNL)
- Simulación

[Labadie (2004)] presenta el problema de la optimización de sistemas con la función objetivos y las limitaciones. A la hora de revisar los distintos métodos, los clasifica según:

- Optimización Estocástica Implícita (OEI)
- Optimización Estocástica Explícita (OEE)
- “Real Time Control with Forecasting”, que se podría traducir por Control en Tiempo Real Con Pronóstico
- Modelos de programación heurística.

En el último grupo trata las Redes Neuronales Artificiales (RNA) y la modelización utilizando las “Fuzzy Rules”. [Labadie (2004)] destaca que las RNA pueden ser una buena alternativa a Análisis de Regresión Múltiple (ARM), además de la capacidad de las Fuzzy Rules de traducir a un lenguaje matemático una descripción lingüística de reglas de gestión.

[Lin and Rutten,(2016)] exponen métodos de PL, PD y Programación cuadrática (PC) utilizados para la optimización de embalses. También comentan modelos de simulación de caudales para evaluar la respuesta de un sistema a una serie de reglas de operación y control. Por último, describen el “Model Predictive Control”, un método más para la gestión de embalses.

En segundo lugar, expondremos algunos artículos sobre el cálculo de reglas de operación de embalses de interés por el método empleado o el sistema de gestión a simular. [Solemaini et al.,(2016)] presentan un método de PD estocástica que tiene en cuenta la naturaleza estocástica de los flujos de entrada a embalses y la variabilidad de la demanda de agua en la agricultura. Analizan el resultado de este método aplicado a cuatro casos distintos en los que consideran la incertidumbre de los flujos de entrada y los de demanda tanto por separado como juntas. También emplean PNL. [Mousavi et al.,(2007)] utilizan FR y el método ANFIS, que es el acrónimo en inglés para “Adaptive-Network Fuzzy Inference System”, para deducir reglas de operación de embalses en un modelo a largo plazo y otro a corto plazo. En este caso, las FR están representadas por reglas del estilo “if-then”. Por ejemplo: si el volumen es x y el flujo de entrada es z , entonces el flujo de salida es la suma de ambos donde cada uno está multiplicado por un determinado parámetro que se estima con una red neuronal.

[Coerver (2015)] deduce reglas de operación de embalses a partir de series históricas de volumen, flujos de entrada y de salida, utilizando ANFIS y redes neuronales. Expone tres casos en los que el objetivo es simular flujos de salida. En el primero, los predictores son el flujo de entrada junto con el volumen en el periodo actual. En el segundo, también tiene en cuenta el volumen en el periodo anterior. En el tercer caso, además de las variables anteriores utiliza el “Time of Year” (ToY) que corresponde al momento del año, hecho que permite tener en cuenta la estacionalidad a la hora de deducir las reglas de gestión. Evalúa las reglas obtenidas en once embalses, mostrando que el primer caso ofrece buenos resultados en la mayoría de ellos y que el hecho de añadir variables de entrada puede mejorar la deducción de reglas de gestión en algunos casos.

Finalmente, en el séptimo capítulo de la tesis de Raül Marcos [Marcos (2016)], se estudia el potencial del pronóstico estacional para predecir anomalías mensuales de volumen, flujos de entrada y de salida. Para ello, se desarrollan modelos mensuales de Regresión Lineal Múltiple (RLM) basados en combinaciones de distintos indicadores de anomalías (precipitación y temperatura máxima y mínima entre otros). Se identifica el grado de predicción de los indicadores para cada una de las variables estudiadas. Por ejemplo, el mejor predictor para el flujo de entrada son las anomalías de precipitación, mientras que para el volumen es la anomalía del flujo acumulado.

2. Datos

Se han utilizado datos de volumen, caudal de entrada y de salida del anuario de aforos del Ministerio para la Transición Ecológica. El Cuadro 1 muestra información sobre los embalses seleccionados. De los cinco listados, se descartó Rialb por disponer de datos para un periodo inferior al del resto de embalses.

Los datos de volumen y caudal de salida fueron proporcionados por el Dr. Luis Garrote. Los de caudal de entrada corresponden a la suma de los datos de caudal de las estaciones de aforos listadas en el Cuadro 1.

3. Metodología

3.0.1. Regresión lineal y regresión lineal múltiple

Se decide emplear la RL o RLM (en caso de utilizar varios predictores) para la deducción de reglas de gestión de los embalses listados en la Tabla 1. El fin de la RL es conocer la relación lineal entre dos variables. La variable dependiente o respuesta (Y) y la variable independiente o predictor (X). Se obtiene la línea que, siguiendo el criterio de mínimos cuadrados, minimiza las distancias cuadráticas de los puntos con la línea. La ecuación de la línea es la siguiente:

$$\hat{y} = a + bx \quad (1)$$

donde \hat{y} corresponde al valor pronosticado de y . Las ecuaciones para el cálculo de la ordenada en el origen (a) y la pendiente (b) son las siguientes:

$$a = \bar{y} - b\bar{x} \quad (2)$$

$$b = \frac{\sum_{i=1}^n [(x_i - \bar{x})(y_i - \bar{y})]}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (3)$$

Donde \bar{x} y \bar{y} corresponden a los valores medios de las series de x y y .

En el caso de la regresión lineal múltiple, el objetivo es obtener una relación lineal entre un conjunto de variables independientes con una variable dependiente:

$$\hat{y} = a + b_1x_1 + b_2x_2 + b_3x_3 \dots + b_nx_n \quad (4)$$

3.0.2. Análisis realizados

Se han realizado modelos mensuales de RLM utilizando las siguientes variables:

- Volumen (mes actual y anterior)
- Caudal de salida (mes actual y anterior)
- Caudal de entrada (mes actual)

Los predictores varían de un mes a otro y de un embalse a otro debido principalmente a la meteorología, la estacionalidad y los diferentes usos que se le den a cada embalse. Se han seguido varios criterios para seleccionar qué predictores son los adecuados para cada modelo: En primer lugar, el p-value ha de ser inferior a 0.05, indicando que el predictor es significativo y aporta valor al modelo. En segundo lugar, el R^2 ajustado ha de ser lo más cercano posible a 100 % para así obtener una varianza elevada.

RLM Se han llevado a cabo los siguientes análisis:

1. Análisis 1:

- Calcular modelos mensuales de RL simulando el caudal de salida utilizando como predictor el volumen del mes actual.

2. Análisis 2:

- Calcular modelos mensuales de RLM simulando el caudal de salida utilizando como predictores el resto de los listados previamente.
- Seleccionar aquellas variables cuyos P-value < 0.05 .
- Calcular modelos mensuales de RLM con la selección previa.

3. Análisis 3:

- Calcular modelos mensuales de RL simulando el caudal de salida utilizando como predictor cada uno de los listados previamente (excepto el caudal de salida).
- Seleccionar aquellas variables cuyos P-value < 0.05 .
- Calcular modelos mensuales de RLM con la selección previa.

En un principio se decidió que el caudal de salida del mes actual sería la variable dependiente a simular, pero ésta fue sustituida por el volumen del mes actual por resultar más operativo de cara a forzar el modelo SASER (SAFRAN & Surfex & Eau-Dyssée & RAPID). Es por ello que el “Análisis 3” se repitió utilizando como variable independiente el volumen.

3.0.3. Validación

Se han efectuado tests de validación de los modelos mensuales:

- Global Stat: Analiza si las relaciones entre los predictores y la variable dependiente a simular son lineales.
- Skewness: Analiza si la asimetría de la distribución. Si ésta está sesgada positiva o negativamente, sería necesario transformarla para alcanzar la condición de normalidad.
- Kurtosis: Analiza el pico y las colas de la distribución. De no ser mesocúrtica (distribución normal), sería necesario transformarla para alcanzar la condición de normalidad.
- Link Function: Analiza si los predictores son categóricos. De no cumplirse esta asunción, debería considerarse una opción alternativa al modelo lineal.
- Heteroscedasticity: Analiza si la varianza de los residuos es constante a lo largo del rango del predictor. Si los residuos son heteroscedásticos, la precisión de los pronósticos del modelo varía a lo largo del citado rango.

- Leave-one-out cross validation: En la validación cruzada (cross-validation) se repite y calcula la media aritmética resultante de las medidas de evaluación sobre diferentes particiones. Tiene como fin comprobar la independencia entre los datos utilizados para entrenar el modelo y los de prueba. En la validación cruzada dejando uno fuera (leave-one-out cross-validation) la repartición entre datos de entrenamiento y de prueba es tal que hay una sola muestra para datos de prueba y el resto corresponde a los de entrenamiento. Este test sólo se ha realizado para uno de los análisis. Proporciona los siguientes datos:
 - RMSE (Root Mean Squared Error). Éste es posteriormente normalizado.
 - Rsquared
 - MAE (Mean Absolute Error). Éste es posteriormente normalizado.

3.1. Código

Para deducir las reglas de gestión a través de RL y RLM, se han elaborado una serie de programas utilizando el lenguaje R (la versión empleada se muestra en la Tabla 2. A continuación se listan las principales funciones utilizadas junto con las referencias proporcionadas por www.rdocumentation.org

- *lm* [Chambers (1992), Wilkinson and Rogers,(1973)]: Sirve para ajustar modelos lineales y se ha empleado para el cálculo de RL y RLM
- *gvlma* [Peña and Slate,(2006)]: Se emplea para la validación global de asunciones de modelos lineales.
- *loocv* [Torgo (2010)]: Realiza un test de validación cruzada dejando uno fuera (leave one out cross validation, *loocv*) de un sistema con un conjunto de datos dado.

Los programas elaborados son los siguientes:

- *mlr_dams_monthly.r* : Ordena los datos en series mensuales y calcula RL (utilizando como predictor el volumen) y RLM (utilizando n predictores citados anteriormente).
- *mlr_dams_monthly_input – selection.R* : Calcula RL o RLM en función del número de predictores que hayan sido identificados como significativos.
- *dam_manag_rules_out – storage.R* : Elaborado para la realización del “Análisis 3” tomando el volumen como variable dependiente. En este programa se incluye el test de *loocv*. Genera dos documentos:
 - Documento de texto detallando la RL o RLM realizada, los resultados de los tests de asunciones obtenidos con la función *gvlma* de R y los del test de *loocv*. Un ejemplo de este documento se muestra en la Fig. 2.
 - Cuadro resumen de los tests realizados (Cuadros 15 a 18).

4. Resultados

4.1. Análisis 1

Esta sección muestra los resultados obtenidos de la realización del “Análisis 1” detallado en la Sección 3. Los Cuadros (11 - 14 muestran los p-value, los R^2 ajustado y el resultado de los análisis de la función *gvlma* de R para cada uno de los embalses. Se ha incluido una escala de color para facilitar la lectura de los cuadros. Las celdas en verde indican que el p-value (del predictor o de la RLM) es inferior a 0.05. Las celdas de la columna “R2-adj.” azul indican que la varianza explicada supera el 50 %.

4.2. Análisis 2

Esta sección muestra los resultados obtenidos de la realización del “Análisis 2” detallado en la Sección 3. En los gráficos de la Fig. 1 se pueden ver el p-value y el coeficiente R^2 ajustado de los modelos mensuales de RLM realizados para cada uno de los cuatro embalses.

A continuación se muestran los Cuadros (7 - 10) conteniendo los p-value, los R^2 ajustado y el resultado de los análisis de la función *gvlma* de R para cada uno de los embalses. Se ha incluido la misma escala de color que en la sección anterior. Además, si ninguno de los predictores mostrara un p-value inferior a 0.05, los caracteres de la fila se han marcado en rojo. En estos casos, se ha calculado una RL utilizando el volumen del mes actual como predictor.

4.3. Análisis 3

Esta sección muestra los resultados obtenidos de la realización del “Análisis 3” detallado en la Sección 3. Primero, los modelos mensuales de RLM se calcularon definiendo como variable independiente el caudal de salida. Posteriormente, se decidió definir como variable dependiente el volumen del mes actual.

4.3.1. Análisis 3 (variable dependiente = caudal de salida)

Los Cuadros (11 - 14 muestran los p-value, los R^2 ajustado y el resultado de los análisis de la función *gvlma* de R para cada uno de los embalses. Se ha incluido la misma escala de color que en la Sección 4.1.

El caudal de entrada (inflow) es el predictor más significativo (p-value < 0.05) en todos los embalses. La varianza explicada es superior al 75 % en los modelos de enero, febrero, marzo, septiembre y octubre en el embalse de Oliana, que tiene una varianza media del 63 %. Los modelos de agosto y septiembre del embalse de Yesa también superan el 75 % de varianza explicada y tiene una varianza media del 55 %. Los embalses de Mequinenza y Ribarroja tienen una varianza media de 35 % y 40 % respectivamente.

4.3.2. Análisis 3 (variable dependiente = volumen)

Los Cuadros (15 - 18 muestran nRMSE, R^2 , nMAE y el resultado de los análisis de la función *gvlma* de R para cada uno de los embalses.

5. Conclusiones y próximos pasos

Para la deducción de reglas de gestión en la cuenca del Ebro, se han calculado modelos mensuales de RL y RLM para cuatro embalses, siguiendo diferentes análisis y realizando tests de validación. En general, el método que ha dado mejores resultados es el “Análisis 3” (Sección 4.3). En dichos análisis se pueden observar efectos estacionales que también pueden estar relacionados con los usos a los que está destinado cada embalse. En algunos casos se ha obtenido un modelo que cumple todas las asunciones analizadas en los tests de validación. En aquellos casos en los que esto no ocurre, se deberá valorar el peso de éstas y la varianza explicada con el fin de escoger aquel que mejor represente el comportamiento del embalse en ese determinado mes.

Los siguientes pasos serán:

- Simular los volúmenes de salida.
- Alimentar la cadena de modelización SASER con los volúmenes simulados.
- Obtener y validar caudales simulados en régimen antrópico.

Referencias

- [Chambers (1992)] Chambers, J. M.: Linear models, Chapter 4 of Statistical Models in S eds J. M. Chambers and T. J. Hastie, Wadsworth and Brooks/Cole, 1992
- [Coerver (2015)] Coerver, H.M.: Modelling of Reservoir Operations using Fuzzy Logic and Artificial Neural Networks, Master Thesis TU Delft, <http://resolver.tudelft.nl/uuid:3cb0d217-01b7-4d7e-a37a-7bade4bcca6a>, 2015
- [Labadie (2004)] Labadie, JW: Optimal Operation of Multireservoir Systems: State-of-the-Art Review, Journal of Water Resources Planning and Management, 130(2):93-111, doi=10.1061/(ASCE)0733-9496(2004)130:2(93), 2004
- [Lin and Rutten,(2016)] Lin, N.M, and Rutten, M.: Optimal Operation of a Network of Multi-purpose Reservoir: A Review, Procedia Engineering, 154:1376 - 1384, doi = <https://doi.org/10.1016/j.proeng.2016.07.504>, 2016
- [Marcos (2016)] Marcos, R.: Improvement of seasonal forecasting techniques applied to water resources and forest fires, Universitat de Barcelona, 2016.

- [Mousavi et al.,(2007)] Mousavi, S.J., Ponnambalam, K., and Karray, F.: Inferring operating rules for reservoir operations using fuzzy regression and ANFIS, *Fuzzy Sets and Systems*, 158:1064-1082, 2007.
- [Peña and Slate,(2006)] Peña, E. A., and Slate, E. H.: Global Validation of Linear Model Assumptions, *Journal of the American Statistical Association*, 101(473), 341. doi:10.1198/016214505000000637, 2006
- [Solemaini et al.,(2016)] Solemaini, S., Bozorg-Haddad, O., Loáiciga, H.A.: Reservoir Operation Rules with Uncertainties in Reservoir Inflow and Agricultural Demand Derived with Stochastic Dynamic Programming, *Journal of Irrigation and Drainage Engineering*,12(11), doi:10.1061/(ASCE)IR.1943-4774.0001065, 2016.
- [Torgo (2010)] Torgo, L.: *Data Mining using R: learning with case studies*, CRC Press (ISBN: 9781439810187), 2010
- [Wilkinson and Rogers,(1973)] Wilkinson, G. N. and Rogers, C. E.: Symbolic descriptions of factorial models for analysis of variance, *Applied Statistics*, 22, 392-399, 10.2307/2346786, 1973
- [Yeh (1985)] Yeh, W.W.G.: Reservoir Management and Operations Models' A State-of-the-Art Review, *Water Resources Research*, 12(12):1797-1818, 1985.

6. Cuadros y figuras

6.1. Cuadros

6.1.1. Cuadros de la Sec. 2

Cuadro 1: Embalses seleccionados

Embalse	Extensión datos	Uso	Caudal entrada (estaciones aforo)
Oliana	1980 - 2013	Electricidad, riego y abastecimiento	9111, 9148
Rialb	1999 - 2013		9083
Yesa	1980 - 2013		9063, 9170
Mequinenza		Electricidad y riego	9014, 9099, 9112, 9231
Ribarroja		Electricidad, riego y abastecimiento	9017, 9025, 9803, 9176, 9177

Cuadro 2: Información sobre la versión de R

lenguaje	R
versión	R version 3.5.3 Patched (2019-03-11 r76221)
alias	Great Truth

month	P-value		R ² -adj	Assumptions				
	storage	intercept		Global Stat	Skewness	Kurtosis	Link Function	Heteroscedasticity
jan	1.6E-04	9.5E-03	34%					
feb	3.6E-04	2.7E-01	31%					
mar	8.6E-05	6.9E-02	37%					
apr	2.2E-01	3.0E-01	2%					
may	6.7E-02	8.5E-01	7%					
jun	6.2E-02	5.3E-01	8%					
jul	1.7E-03	4.2E-01	25%					
aug	1.8E-02	3.3E-02	14%					
sep	1.1E-05	6.6E-01	44%					
oct	4.5E-06	5.0E-02	47%					
nov	1.0E-04	1.9E-02	37%					
dec	6.3E-04	5.9E-02	29%					

Cuadro 3: Análisis 1 (Oliana) RLM: p-value, R^2 ajustado y resultados de validación utilizando la función *gvlma* (1980-2013).

month	P-value		R ² -adj	Assumptions				
	storage	intercept		Global Stat	Skewness	Kurtosis	Link Function	Heteroscedasticity
jan	6.9E-06	4.1E-02	47%					
feb	2.1E-03	2.6E-01	24%					
mar	2.7E-03	2.8E-01	23%					
apr	9.7E-06	3.6E-02	46%					
may	6.7E-05	7.2E-02	39%					
jun	6.5E-03	1.9E-01	19%					
jul	2.3E-02	1.4E-03	13%					
aug	8.8E-03	6.1E-07	18%					
sep	8.6E-04	4.9E-05	28%					
oct	8.5E-04	3.2E-01	28%					
nov	1.4E-04	3.8E-01	36%					
dec	3.1E-04	2.1E-01	33%					

Cuadro 4: Análisis 1 (Yesa) RLM: p-value, R^2 ajustado y resultados de validación utilizando la función *gvlma* (1980-2013).

6.1.2. Cuadros de la Sec. 3.1

6.1.3. Cuadros de la Sec. 4.1

6.1.4. Cuadros de la Sec. 4.2

6.1.5. Cuadros de la Sec. 4.3.1

6.1.6. Cuadros de la Sec. 4.3.2

6.2. Figuras

month	P-value		R ² -adj	Assumptions				
	storage	intercept		Global Stat	Skewness	Kurtosis	Link Function	Heteroscedasticity
jan	3.9E-02	4.0E-01	10%					
feb	1.7E-02	2.5E-01	14%					
mar	3.2E-02	6.7E-01	11%					
apr	8.4E-02	5.3E-01	6%					
may	6.5E-02	7.2E-01	8%					
jun	1.8E-01	8.4E-01	3%					
jul	9.5E-02	9.9E-01	6%					
aug	9.7E-01	2.4E-01	-3%					
sep	6.6E-01	5.2E-02	-3%					
oct	1.1E-01	8.1E-01	5%					
nov	8.3E-04	2.4E-01	28%					
dec	2.5E-02	4.8E-01	12%					

Cuadro 5: Análisis 1 (Mequinensa) RLM: p-value, R^2 ajustado y resultados de validación utilizando la función *gulma* (1980-2013).

month	P-value		R ² -adj	Assumptions				
	storage	intercept		Global Stat	Skewness	Kurtosis	Link Function	Heteroscedasticity
jan	1.3E-02	7.2E-04	16%					
feb	4.0E-01	4.8E-02	-1%					
mar	1.8E-01	1.3E-02	3%					
apr	8.6E-01	1.6E-01	-3%					
may	2.9E-01	1.1E-02	0%					
jun	5.4E-02	3.1E-03	9%					
jul	7.0E-02	3.1E-03	7%					
aug	4.8E-01	3.3E-02	-2%					
sep	1.0E-02	4.3E-05	17%					
oct	3.0E-01	3.7E-02	0%					
nov	2.1E-02	9.9E-04	14%					

Cuadro 6: Análisis 1 (Ribarroja) RLM: p-value, R^2 ajustado y resultados de validación utilizando la función *gulma* (1980-2013).

month	P-value				R ² -adj	P-value	Assumptions					
	storage	storage (m-1)	release (m-1)	inflow			intercept	Global Stat	Skewness	Kurtosis	Link Function	Heteroscedasticity
jan	NA	NA	NA	5.4E-16	6.1E-01	87%	5.4E-16					
feb	NA	NA	8.9E-03	3.0E-06	5.6E-02	75%	1.4E-10					
mar	6.3E-02	NA	5.9E-01	1.9E-07	1.7E-01	73%	2.7E-09					
apr	NA	NA	9.1E-03	9.0E-05	9.9E-01	45%	3.7E-05					
may	3.0E-02	NA	2.1E-01	NA	5.8E-01	9%	8.5E-02					
jun	NA	NA	4.5E-01	3.9E-05	8.3E-01	40%	1.5E-04					
jul	NA	1.5E-05	1.6E-02	NA	1.6E-01	57%	7.5E-07					
aug	NA	NA	NA	9.9E-05	3.1E-07	36%	9.9E-05					
sep	1.8E-01	NA	1.1E-01	1.6E-05	1.7E-01	70%	1.5E-08					
oct	NA	NA	4.2E-03	8.0E-12	9.3E-02	79%	1.1E-11					
nov	1.0E-04	NA	NA	1.9E-02	NA	37%	1.0E-04					
dec	NA	NA	1.7E-02	1.2E-06	1.7E-01	62%	1.9E-07					

Cuadro 7: Análisis 2 (Oliana) RLM: p-value, R^2 ajustado y resultados de validación utilizando la función *gulma* (1980-2013).

month	P-value					R ² -adj	P-value	Assumptions					
	storage	storage (m-1)	release (m-1)	inflow	intercept			Global Stat	Skewness	Kurtosis	Link Function	Heteroscedasticity	
jan	NA	NA	NA	3.9E-04	1.7E-01	32%	3.9E-04						
feb	NA	NA	2.1E-03	5.0E-03	6.7E-02	42%	1.1E-04						
mar	NA	NA	2.2E-04	NA	1.7E-01	34%	2.2E-04						
apr	NA	NA	NA	8.4E-02	1.3E-04	6%	8.4E-02						
may	6.7E-05	NA	NA	NA	7.2E-02	39%	6.7E-05						
jun	NA	NA	1.4E-02	NA	3.6E-01	15%	1.4E-02						
jul	5.1E-04	8.2E-03	1.2E-05	1.2E-02	5.7E-01	56%	1.6E-05						
aug	NA	NA	1.7E-05	NA	2.6E-02	44%	1.7E-05						
sep	5.5E-04	1.8E-05	NA	NA	1.9E-01	60%	3.8E-07						
oct	2.0E-01	7.4E-03	NA	4.7E-03	3.7E-01	51%	4.1E-05						
nov	NA	NA	8.1E-05	NA	3.5E-01	38%	8.1E-05						
dec	NA	NA	NA	1.1E-05	2.4E-01	45%	1.1E-05						

Cuadro 8: Análisis 2 (Yesa) RLM: p-value, R^2 ajustado y resultados de validación utilizando la función *gvlma* (1980-2013).

month	P-value					R ² -adj	P-value	Assumptions					
	storage	storage (m-1)	release (m-1)	inflow	intercept			Global Stat	Skewness	Kurtosis	Link Function	Heteroscedasticity	
jan	NA	NA	9.0E-02	NA	8.4E-01	6%	9.0E-02						
feb	7.5E-04	1.0E-03	6.2E-03	NA	3.0E-01	41%	3.5E-04						
mar	NA	NA	2.0E-01	NA	9.7E-01	2%	2.0E-01						
apr	NA	NA	3.5E-01	NA	7.6E-01	0%	3.5E-01						
may	NA	NA	2.4E-01	NA	7.8E-01	1%	2.4E-01						
jun	NA	NA	1.9E-01	NA	8.5E-01	2%	1.9E-01						
jul	1.2E-01	NA	1.6E-01	NA	3.3E-01	9%	9.4E-02						
aug	NA	NA	7.7E-01	NA	5.5E-02	-3%	7.7E-01						
sep	NA	NA	5.2E-01	NA	5.9E-01	-2%	5.2E-01						
oct	NA	NA	1.1E-02	NA	1.8E-01	17%	1.1E-02						
nov	NA	1.5E-05	2.3E-04	NA	6.9E-03	64%	5.9E-08						
dec	NA	NA	1.6E-02	NA	3.9E-01	14%	1.6E-02						

Cuadro 9: Análisis 2 (Mequinensa) RLM: p-value, R^2 ajustado y resultados de validación utilizando la función *gvlma* (1980-2013).

month	P-value					R ² -adj	P-value	Assumptions					
	storage	storage (m-1)	release (m-1)	inflow	intercept			Global Stat	Skewness	Kurtosis	Link Function	Heteroscedasticity	
jan	NA	NA	1.7E-02	NA	8.4E-04	15%	1.7E-02						
feb	NA	NA	4.3E-01	2.1E-05	1.1E-01	45%	6.4E-05						
mar	NA	NA	1.9E-01	1.0E-04	2.6E-02	40%	2.0E-04						
apr	8.6E-01	NA	NA	NA	1.6E-01	-3%	8.6E-01						
may	NA	NA	2.7E-01	8.1E-04	4.2E-03	28%	3.0E-03						
jun	5.4E-02	NA	NA	NA	3.1E-03	9%	5.4E-02						
jul	NA	NA	1.0E-01	NA	5.0E-03	6%	5.0E-03						
aug	2.3E-02	6.2E-05	2.5E-02	NA	4.5E-01	39%	5.0E-04						
sep	NA	NA	8.4E-03	NA	3.7E-05	18%	8.4E-03						
oct	NA	NA	5.2E-01	NA	6.8E-02	-2%	5.2E-01						
nov	NA	NA	4.2E-02	1.7E-01	4.6E-03	14%	4.1E-02						
dec	NA	NA	3.7E-02	NA	4.4E-03	11%	3.7E-02						

Cuadro 10: Análisis 2 (Ribarroja) RLM: p-value, R^2 ajustado y resultados de validación utilizando la función *gvlma* (1980-2013).

month	P-value					R ² -adj	P-value	Assumptions				
	storage	release (m-1)	storage (m-1)	inflow	intercept			Global Stat	Skewness	Kurtosis	Link Function	Heteroscedasticity
jan	1.9E-01	5.7E-02	1.2E-01	8.1E-12	3.7E-01	91%	2.7E-15					
feb	7.1E-02	1.3E-03	4.8E-01	1.3E-02	1.4E-01	83%	1.4E-11					
mar	9.9E-03	1.2E-02	2.9E-01	4.0E-05	1.5E-01	78%	7.3E-10					
apr	NA	3.6E-02	4.1E-01	1.2E-03	3.4E-01	51%	1.9E-05					
may	NA	NA	NA	2.2E-03	7.8E-03	23%	2.2E-03					
jun	NA	2.3E-03	NA	2.7E-05	4.0E-01	55%	1.7E-06					
jul	2.8E-01	8.7E-05	2.7E-02	2.4E-01	4.1E-01	57%	7.3E-06					
aug	4.5E-01	1.5E-01	8.2E-02	4.4E-03	3.3E-01	49%	7.7E-05					
sep	4.2E-02	3.3E-04	4.2E-01	3.8E-04	6.2E-03	80%	1.4E-10					
oct	7.4E-01	2.2E-03	NA	9.5E-08	2.4E-01	81%	1.2E-11					
nov	9.0E-02	4.1E-01	7.2E-01	4.1E-01	6.2E-02	34%	2.9E-03					
dec	7.6E-01	4.8E-03	6.8E-01	1.9E-06	9.7E-01	70%	9.1E-08					

Cuadro 11: Análisis 3 (Oliana) RLM: p-value, R^2 ajustado y resultados de validación utilizando la función *gvlma* (1980-2013).

month	P-value					R ² -adj	P-value	Assumptions				
	storage	release (m-1)	storage (m-1)	inflow	intercept			Global Stat	Skewness	Kurtosis	Link Function	Heteroscedasticity
jan	5.9E-01	5.4E-01	5.2E-01	1.0E-02	2.5E-02	55%	1.8E-05					
feb	1.4E-01	3.8E-03	1.2E-01	1.4E-03	8.6E-01	59%	6.7E-06					
mar	9.2E-01	2.9E-03	3.0E-01	NA	4.7E-01	49%	4.1E-05					
apr	8.5E-02	4.7E-01	8.5E-01	NA	6.8E-02	44%	1.9E-04					
may	1.3E-01	3.7E-01	9.3E-01	2.3E-01	2.4E-01	37%	1.8E-03					
jun	2.6E-01	4.2E-02	6.1E-01	NA	5.1E-01	25%	9.5E-03					
jul	1.4E-02	1.4E-01	2.3E-04	NA	8.3E-01	46%	9.4E-05					
aug	6.2E-01	4.2E-04	1.3E-01	NA	4.9E-01	77%	4.1E-10					
sep	7.0E-03	2.6E-01	1.0E-04	NA	2.3E-01	76%	1.1E-09					
oct	2.8E-02	7.5E-01	1.7E-04	7.7E-06	5.8E-01	70%	1.5E-07					
nov	2.4E-01	2.3E-02	4.4E-01	7.6E-02	9.5E-03	58%	7.9E-06					
dec	5.6E-01	6.5E-02	8.0E-01	4.9E-03	8.4E-02	63%	1.6E-06					

Cuadro 12: Análisis 3 (Yesa) RLM: p-value, R^2 ajustado y resultados de validación utilizando la función *gvlma* (1980-2013).

month	P-value					R ² -adj	P-value	Assumptions				
	storage	release (m-1)	storage (m-1)	inflow	intercept			Global Stat	Skewness	Kurtosis	Link Function	Heteroscedasticity
jan	4.8E-02	2.8E-04	NA	NA	2.5E-01	41%	1.5E-04					
feb	3.6E-02	2.3E-02	NA	NA	2.1E-01	26%	4.2E-03					
mar	2.0E-01	5.5E-04	NA	NA	7.1E-01	39%	2.5E-04					
apr	NA	6.0E-03	NA	NA	1.7E-01	19%	6.0E-03					
may	NA	4.4E-04	NA	NA	6.8E-04	31%	4.4E-04					
jun	NA	1.3E-02	NA	1.9E-01	5.8E-01	25%	4.8E-03					
jul	NA	8.3E-05	NA	NA	3.8E-04	38%	8.3E-05					
aug	NA	8.0E-04	NA	NA	1.0E-02	29%	8.0E-04					
sep	NA	3.6E-05	NA	NA	7.8E-02	41%	3.6E-05					
oct	NA	NA	1.1E-02	NA	1.8E-01	17%	1.1E-02					
nov	6.9E-02	1.1E-05	3.2E-02	NA	2.6E-03	67%	6.4E-08					
dec	4.7E-02	1.1E-05	8.1E-02	NA	6.4E-01	53%	1.1E-05					

Cuadro 13: Análisis 3 (Mequinensa) RLM: p-value, R^2 ajustado y resultados de validación utilizando la función *gvlma* (1980-2013).

month	P-value					R ² -adj	P-value	Assumptions				
	storage	release (m-1)	storage (m-1)	inflow	intercept			Global Stat	Skewness	Kurtosis	Link Function	Heteroscedasticity
jan	6.8E-01	1.6E-03	9.2E-01	NA	6.0E-02	38%	9E-04					
feb	NA	1.8E-02	NA	4.2E-05	4.0E-02	54%	5E-06					
mar	NA	5.6E-04	NA	7.2E-03	4.2E-02	57%	1E-06					
apr	NA	3.1E-02	NA	4.1E-02	2.4E-01	29%	2E-03					
may	NA	1.6E-03	NA	1.6E-03	1.3E-03	46%	4E-05					
jun	NA	6.9E-02	NA	1.5E-01	1.8E-01	24%	6E-03					
jul	NA	2.5E-05	NA	NA	1.1E-02	42%	3E-05					
aug	NA	4.2E-04	NA	NA	3.2E-03	31%	4E-04					
sep	4.9E-01	5.1E-06	3.1E-01	NA	2.2E-04	58%	3E-06					
oct	NA	8.9E-03	NA	NA	2.3E-01	18%	9E-03					
nov	2.5E-01	1.2E-03	3.6E-01	NA	1.2E-02	40%	5E-04					
dec	NA	1.7E-04	4.6E-01	NA	4.1E-01	44%	9E-05					

Cuadro 14: Análisis 3 (Ribarroja) RLM: p-value, R^2 ajustado y resultados de validación utilizando la función *gvlma* (1980-2013).

Month	nRMSE	R2 (%)	nMAE	G. Stat	Skewness	Kurtosis	Link F.	Heterosc.
Jan	0.09	82	0.07	yes	yes	yes	yes	yes
Feb	0.24	24	0.15	no	no	no	no	no
Mar	0.17	30	0.11	no	yes	yes	no	yes
Apr	0.11	11	0.09	yes	yes	yes	yes	yes
May	0.11	9	0.10	yes	yes	yes	yes	yes
Jun	0.08	25	0.06	no	no	yes	no	yes
Jul	0.09	56	0.07	yes	yes	yes	yes	yes
Aug	0.24	36	0.20	no	yes	yes	no	yes
Sep	0.20	69	0.16	yes	yes	yes	yes	yes
Oct	0.20	61	0.17	no	yes	yes	no	yes
Nov	0.13	76	0.11	yes	yes	yes	no	yes
Dec	0.13	59	0.10	no	yes	yes	no	yes

Cuadro 15: Análisis 3 (Oliana) RLM: nRMSE, R^2 nMAE y resultados de validación utilizando la función *gvlma* (1980-2013).

Month	nRMSE	R2 (%)	nMAE	G. Stat	Skewness	Kurtosis	Link F.	Heterosc.
Jan	0.11	89	0.08	no	yes	yes	no	yes
Feb	0.09	89	0.07	no	no	no	no	no
Mar	0.10	83	0.07	no	no	no	yes	yes
Apr	0.08	81	0.06	no	yes	yes	no	yes
May	0.07	73	0.05	yes	yes	yes	no	yes
Jun	0.07	68	0.05	yes	yes	yes	yes	yes
Jul	0.06	89	0.04	yes	yes	yes	yes	yes
Aug	0.09	89	0.06	yes	yes	yes	no	yes
Sep	0.11	90	0.08	yes	yes	yes	yes	yes
Oct	0.25	64	0.19	yes	yes	yes	yes	yes
Nov	0.26	59	0.20	no	no	no	yes	yes
Dec	0.17	79	0.13	no	yes	yes	no	yes

Cuadro 16: Análisis 3 (Yesa) RLM: nRMSE, R^2 nMAE y resultados de validación utilizando la función *gvlma* (1980-2013).

Month	nRMSE	R2 (%)	nMAE	G. Stat	Skewness	Kurtosis	Link F.	Heterosc.
Jan	0.20	5	0.11	no	yes	no	no	yes
Feb	0.08	74	0.06	no	yes	yes	no	yes
Mar	0.21	13	0.11	no	no	no	yes	no
Apr	0.20	1	0.10	no	no	no	yes	yes
May	0.20	1	0.11	no	no	no	yes	yes
Jun	0.20	0	0.11	no	no	no	yes	yes
Jul	0.21	0	0.13	no	no	no	yes	yes
Aug	0.16	70	0.13	yes	yes	yes	yes	yes
Sep	0.11	73	0.08	yes	yes	yes	yes	yes
Oct	0.11	72	0.09	yes	yes	yes	yes	no
Nov	0.24	31	0.15	no	no	no	yes	yes
Dec	0.11	85	0.09	yes	yes	yes	yes	yes

Cuadro 17: Análisis 3 (Mequinzenza) RLM: nRMSE, R^2 nMAE resultados de validación utilizando la función *gvlma* (1980-2013).

Month	nRMSE	R ² (%)	nMAE	G. Stat	Skewness	Kurtosis	Link F.	Heterosc.
Jan	0.04	93	0.02	no	no	no	no	no
Feb	0.03	97	0.02	no	no	no	yes	yes
Mar	0.02	99	0.01	no	yes	yes	no	yes
Apr	0.20	29	0.07	no	no	no	yes	yes
May	0.21	5	0.10	no	yes	no	no	no
Jun	0.02	99	0.01	yes	yes	yes	yes	yes
Jul	0.01	99	0.01	no	yes	yes	no	yes
Aug	0.01	99	0.01	no	yes	yes	no	yes
Sep	0.02	99	0.01	no	yes	yes	no	yes
Oct	0.08	71	0.04	no	no	no	yes	no
Nov	0.02	99	0.01	no	yes	no	no	no
Dec	0.04	95	0.02	no	no	no	no	no

Cuadro 18: Análisis 3 (Ribarroja) RLM: nRMSE, R^2 nMAE y tests de asunciones (1980-2013).

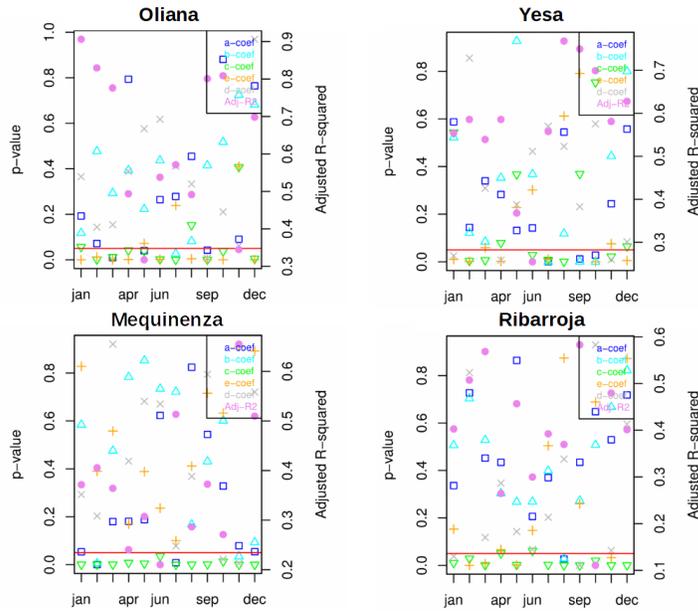


Figura 1: RLM: p-value y R^2 ajustado (1980-2013). Los coeficientes indicados en la leyenda corresponden a los siguientes predictores: a = volumen (mes actual), b = volumen (mes anterior), c = caudal de salida (mes anterior), d = caudal de entrada, e = ordenada en el origen

```

Oliana (month =1): mlr(output = storage) & loocv & gvlma
[1] "4 significant input variables for dam Oliana in month 1."

Call:
lm(formula = .outcome ~ ., data = dat)

Residuals:
    Min       1Q   Median       3Q      Max
-13.4077  -3.0449   0.3198   2.9467  14.5351

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.29218    5.27496   1.193  0.24260
release     -0.20873    0.15648  -1.334  0.19261
release_1   -0.11674    0.07998  -1.460  0.15516
storage_1    0.86287    0.08857   9.742 1.19e-10 ***
inflow       0.55456    0.17433   3.181  0.00348 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.522 on 29 degrees of freedom
Multiple R-squared:  0.8588, Adjusted R-squared:  0.8393
F-statistic: 44.1 on 4 and 29 DF,  p-value: 6.325e-12

Linear Regression

34 samples
 4 predictor

No pre-processing
Resampling: Leave-One-Out Cross-Validation
Summary of sample sizes: 33, 33, 33, 33, 33, 33, ...
Resampling results:

    RMSE      Rsquared    MAE
 5.794759  0.8197535  4.44454

Tuning parameter 'intercept' was held constant at a value of TRUE

Call:
lm(formula = datain$storage ~ ., data = datain)

Coefficients:
(Intercept)      release      release_1      storage_1      inflow
   6.2922      -0.2087      -0.1167      0.8629      0.5546

ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
Level of Significance = 0.05

Call:
gvlma(x = lm(datain$storage ~ ., data = datain))

      Value p-value      Decision
Global Stat  1.938199  0.7471 Assumptions acceptable.
Skewness    0.006109  0.9377 Assumptions acceptable.
Kurtosis    1.777226  0.1825 Assumptions acceptable.
Link Function  0.066797  0.7961 Assumptions acceptable.
Heteroscedasticity 0.088066  0.7666 Assumptions acceptable.

```

Figura 2: Ejemplo de documento resumen de la RL o RLM y los tests loocv y gvlma.