

Predicción de caudal del Río Uruguay
Tratamiento Estadístico de Señales 2011

Gonzalo Belcredi
4506844-9

Facultad de Ingeniería
Universidad de la República

Contents

1	Introducción	2
2	Modelado del sistema	3
3	Descripción de la cuenca y la red de estaciones consultadas	5
4	Calibración del sistema	8
4.1	Selección de modelo y determinación de parámetros	8
4.2	Modelado en términos de ecuaciones de estado	11
4.3	Algoritmo implementado	11
5	Validación y conclusiones	14

1 Introducción

En este trabajo se analizará un método estadístico para la predicción de caudal de un río, dadas las precipitaciones en la cuenca superior. En este caso nos interesa predecir el caudal en la subcuenca media del Río Uruguay, más específicamente en la ciudad de Paso de los Libres¹ (Provincia de Corrientes, Argentina).

La predicción de caudal es un aspecto importante en la planificación de ciertas actividades humanas asentadas en ríos o arroyos, quizás el aporte más evidente sea la alerta temprana de inundaciones aunque también en la agricultura, el transporte fluvial y la generación de energía hidroeléctrica la evolución del caudal tiene especial relevancia.

En el caso de Uruguay, la represa de Salto Grande representa el 46% de la energía producida en el país por lo que es imprescindible contar con datos confiables de la cota del embalse con días o semanas de anticipación. Una planificación basada en datos confiables de la cota futura del embalse permite incrementar el rendimiento de la central, disminuyendo el costo de producción eléctrica (las otras fuentes relevantes son en general más caras, ya sea por generación termoeléctrica o directamente importación de energía de la región).

El método escogido para la predicción se basa en los aportes de Peter Young [1] y se denomina *Data-based mechanistic rainfall-flow model* cuyo enfoque principal es la búsqueda de un modelo genérico que luego se ajustará en función de los datos, en lugar del enfoque tradicional que se basa en la teoría hidrológica para proponer los modelos. Para Young este nuevo enfoque es más efectivo en cuanto no establece prejuicios a la hora de elegir un modelo, y que por el contrario, un ajuste inductivo puede dar lugar al final del proceso a interpretaciones físicas e hidrológicas.

¹Esta ciudad se ubica a 200 kilómetros al norte de la ciudad de Salto.

2 Modelado del sistema

El flujo de un río depende esencialmente del abastecimiento de agua proveniente de las precipitaciones en la cuenca superior. Tanto los cursos de agua superficiales como los reservorios subterráneos se nutren de estas precipitaciones. El agua subterránea puede asimismo permear en los cursos de agua superficiales. Los factores que influyen en el flujo de un río son entonces: las precipitaciones en la cuenca superior, la permeabilidad del suelo, la capacidad y el tiempo de carga de los reservorios, y el nivel de evaporación que actúa en las aguas superficiales. Es por esta razón que los modelos hidrológicos toman como parámetros el tipo de suelo, la temperatura, las precipitaciones, etc.

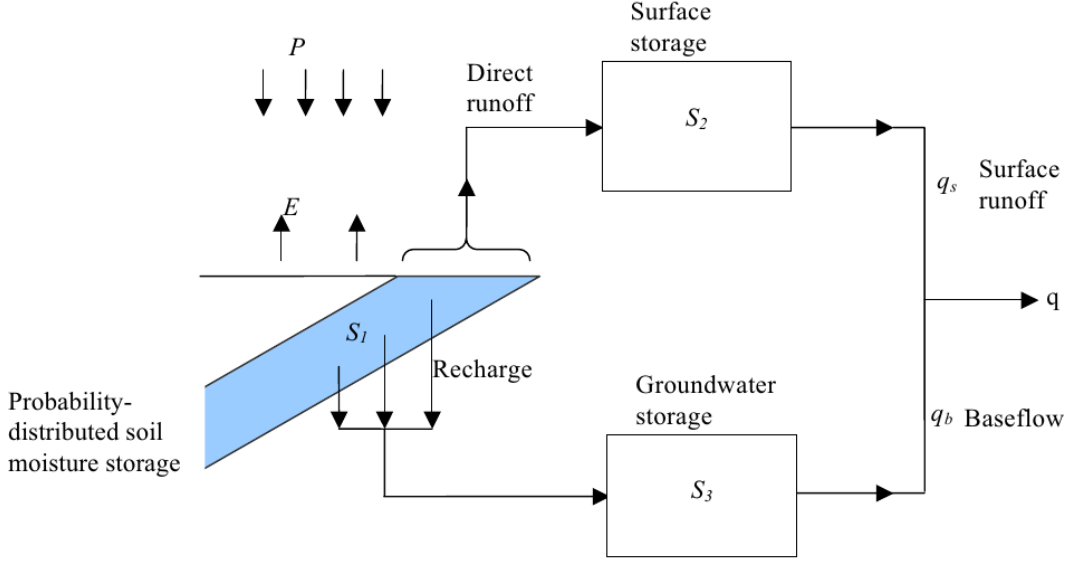


Figure 1: Modelo PDM (Probabilily Distributed Model) rainfall-runoff [2].

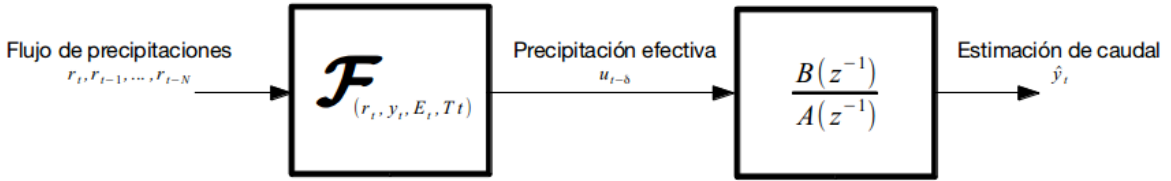


Figure 2: Diagrama de bloques del modelo genérico Función de Transferencia Precipitaciones-Flujo [2]

Young [1] propone un modelo genérico que tiene un componente no lineal $\mathcal{F}(r_t, y_t, E_t, T_t)$, siendo r_t el flujo de precipitaciones en la cuenca, y_t el flujo del río, E_t la evaporación, y T_t la temperatura ambiente en el instante t . La salida de este bloque, denominada precipitación efectiva (u_t) es retardada δ muestras y pasa finalmente por el filtro SLIT $H(z)$:

$$H(z) = \frac{b_0 + b_1 z^{-1} + \dots + b_m z^{-m}}{1 + a_1 z^{-1} + \dots + a_n z^{-n}} \quad (1)$$

El flujo del río está dado entonces por la siguiente ecuación:

$$y(t) = h(n) * u(t - \delta) + \xi(t) \quad (2)$$

La precipitación efectiva $u(t)$ se define mediante la siguiente ecuación no lineal:

$$u(t) = cy(t)^\gamma r(t) \quad (3)$$

En tanto que los errores asociadas a las mediciones y al propio modelo puede modelarse como el siguiente ruido coloreado:

$$\xi(t) = \frac{e_t}{A(z^{-1})}, \quad e_t \sim \mathcal{N}(0, \sigma^2)$$

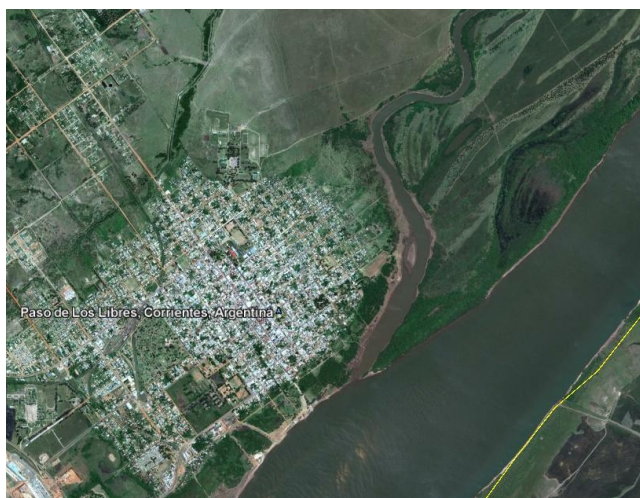
3 Descripción de la cuenca y la red de estaciones consultadas

En este trabajo se intentará predecir el caudal del Río Uruguay en la cuenca media del Río Uruguay, más específicamente en la ciudad de Paso de los Libres (Argentina, $29^{\circ}42'46''\text{S}$ - $57^{\circ}05'12''\text{O}$). Esta localidad fue seleccionada dado que se encuentra entre la cuenca media y la cuenca inferior.

La cuenca media y superior (en adelante la cuenca) del Río Uruguay representa un total de 185.000km^2 y abarca territorios en Brasil y Argentina. La información de precipitaciones en la cuenca proviene de 244 estaciones meteorológicas, 27 ubicadas en Argentina provenientes de la *Base de Datos Hidrológica Integrada* y 217 estaciones en Brasil del SNIRH² de la *Agencia Nacional de Aguas*.



(a) Cuenca Media y Superior del Río Uruguay (en naranja), estaciones meteorológicas consultadas (triángulos)

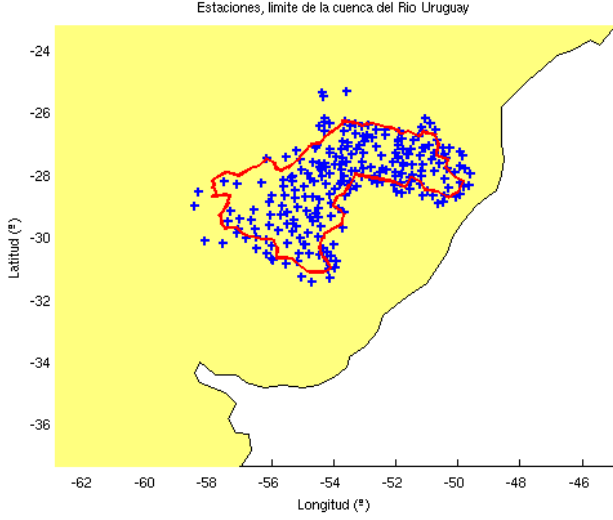


(b) La predicción se realizará a la altura de Paso de los Libres (Corrientes, Argentina), 200 km al norte de Salto.

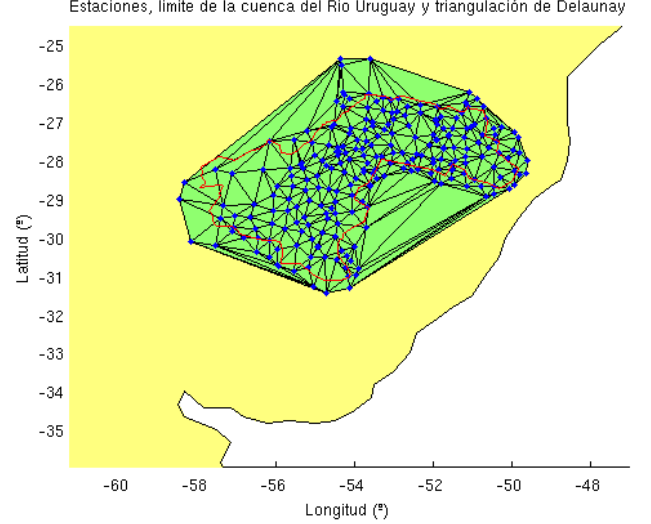
El modelo presentado anteriormente hace uso del flujo diario de precipitaciones en la cuenca ($r(t)[\text{m}^3.\text{s}^{-1}]$), es necesario entonces interpolar la información discreta disponible en la superficie de la cuenca. Para este propósito se elige el método de triangulación de Delaunay, tomando como nodos las estaciones meteorológicas.

El registro de precipitaciones locales diarias es extenso, contando gran parte de las estaciones con datos a partir de la década de 1970. En contrapartida existen datos faltantes, factor que debe tomarse en cuenta a la hora de realizar la serie de precipitaciones en la cuenca. En lugar de realizar técnicas de interpolación para estimar los datos faltantes se opta por realizar la triangulación de Delaunay de forma diaria, contando solo con las estaciones que tienen datos válidos para un día en particular. De esta manera, el número de estaciones consultadas y el grafo de Delaunay generado varían diariamente.

²Sistema Nacional de Informações Sobre Recursos Hídricos



(c) Cuenca Media y Superior del Río Uruguay y estaciones meteorológicas consultadas.



(d) Triangulación de Delaunay para un día determinado.

Una vez construida la malla se procede a asignar un valor de precipitaciones a cada uno de los triángulos que la componen. Siendo C el interior de la cuenca y $x_{i,1}, x_{i,2}, x_{i,3}$ los vértices del triángulo T_i podemos distinguir 3 tipos de triángulos.

- Los triángulos exteriores, $T_{ext} = \{T_i / x_{i,1} \wedge x_{i,2} \wedge x_{i,3} \notin C\}$
- Los triángulos interiores, $T_{int} = \{T_i / x_{i,1} \wedge x_{i,2} \wedge x_{i,3} \subset C\}$.
- Los triángulos fronterizos, $T_f = \{T_i / T_i \notin T_{ext} \wedge T_i \notin T_{int}\}$.

Para asignar un valor de precipitaciones a los triángulos interiores se calcula en primer lugar el baricentro b_i de cada triángulo T_i . Si llamamos A_i el área de T_i , y $r_{i,k}$ la serie de precipitaciones de las estaciones ubicadas en el vértice k -ésimo de T_i , podemos calcular la contribución de cada triángulo al caudal total de la cuenca.

Contribución de los triángulos interiores.

Previamente definimos el peso de cada vértice $P_{i,k}$:

$$P_{i,k} = \frac{1}{\|b_i - r_{i,k}\|}$$

La contribución de cada estación es inversamente proporcional a la distancia que la separa del baricentro, la contribución del triángulo i -ésimo está dado por:

$$c_i = A_i \frac{\sum_{k=1}^3 P_{i,k} r_{i,k}}{\sum_{k=1}^3 P_{i,k}}$$

Contribución de los triángulos fronterizos con dos nodos interiores.

En el caso de los triángulos fronterizos, en primer lugar se calcula el área de intersección de T_i con C , luego se calcula el promedio aritmético de la(s) serie(s) que corresponden a nodos interiores a C , de esta manera:

$$c_i = (T_i \cap C) \cdot \frac{x_{i,k_1} + x_{i,k_2}}{2}$$

Contribución de los triángulos fronterizos con solamente un nodo interior $x_{i,k}$.

$$c_i = (T_i \cap C).x_{i,k}$$

Finalmente el caudal diario de precipitaciones de la cuenca, está dado por:

$$r = \sum_i c_i$$

Los triángulos exteriores no son considerados para construir la serie de precipitaciones de la cuenca, ya que en principio estos no afectan el caudal del río.

4 Calibración del sistema

4.1 Selección de modelo y determinación de parámetros

En esta instancia nos interesa por un lado seleccionar un modelo adecuado (el orden de $H(z)$ y δ) y por otro lado determinar los siguiente parámetros:

- Los coeficientes del filtro H (a_i y b_i)
- El parámetro γ
- La constante c

Determinación de parámetros

El flujo en el río está dado por:

$$\hat{y}_t = - \sum_{k=1}^n a_k y_{t-k} + \sum_{k=0}^m b_k u_{t-\delta} \quad (4)$$

Si reemplazamos con (2), y llamamos $\bar{b}_k = c \cdot b_k$ obtenemos la siguiente estimación:

$$\hat{y}_t = - \sum_{k=1}^n a_k y_{t-k} + \sum_{k=0}^m \bar{b}_k y_{t-\delta}^\gamma r_{t-\delta} \quad (5)$$

Si disponemos de datos de $y(t)$ para valores de $t \leq t - \delta$ (estamos prediciendo con un horizonte de δ días) entonces podemos estimar los parámetros definidos en la ecuación anterior mediante el método de mínimos cuadrados no lineal.³

Luego pasamos a calcular la constante de normalización c , que está definida de la siguiente manera:

$$c = \frac{\sum_N y_t}{\sum_N y_t^\gamma r_t}$$

Donde N es el número de muestras de la serie de precipitaciones. Finalmente, los coeficientes b_k están dados por:

$$b_k = \frac{\bar{b}_k}{c}$$

Si bien estamos en condiciones de determinar los parámetros del sistema, es necesario estimar el orden del filtro y el retardo δ .

Selección de modelo

Dado que no tenemos a priori n, m y δ para definirlos es necesario tener en cuenta que buscamos por un lado minimizar el error cuadrático medio del modelo (generalmente implica sistemas más complejos con mayor resolución) pero por otro privilegiar sistemas menos complejos (en este caso buscamos que $p = n + m$ sea chico) de forma de tener una menor cantidad de picos espúreos que pueden dar lugar a errores locales importantes.[4]⁴

³Se implementó mediante la función *lsqnonlin* en Matlab.

⁴Hayes p.445 cap.8

Existen diversos criterios de selección de modelo, como el AIC (Akaike Information Criterion) o el BIC (Bayesian Information Criterion). Por ejemplo el AIC se define como:

$$AIC = 2p - 2\ln(L_p)$$

Siendo p la complejidad del modelo (en nuestro caso $p = n + m$) y L_p el valor maximizado de la función de máxima verosimilitud. Dado que estos indicadores pueden dar lugar a distintas selecciones y que en nuestro caso resulta difícil calcularlos con los datos que disponemos, se intentará deducir los modelos que mejor ajustan las observaciones a través de la gráfica de la figura 1.

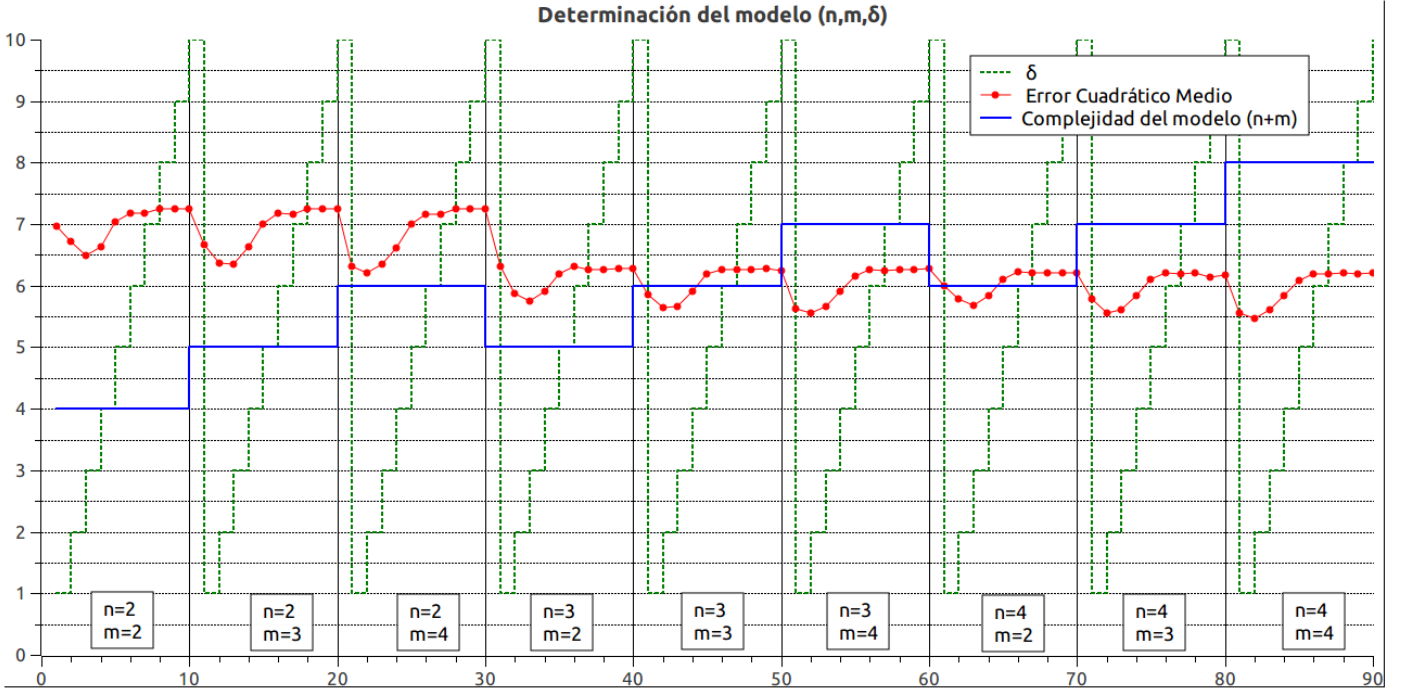


Figure 3: Selección de modelo

Podemos observar en la gráfica por un lado el error cuadrático medio (ε_m) del modelo⁵ en función de n y m , y por otro el valor de δ . Se ve claramente que la mejor estimación de δ es 2 o 3, y además que a partir de $p > 5$ no disminuye significativamente el ε_m al aumentar la complejidad del modelo. Con estas observaciones podemos seleccionar los siguientes modelos:

$[n, m, \delta]$	ε_m
$[3, 2, 3]$	$1449m^3 \cdot s^{-1}$
$[3, 3, 2]$	$1541m^3 \cdot s^{-1}$
$[3, 4, 2]$	$1704m^3 \cdot s^{-1}$

Table 1: Modelos seleccionados

En tanto que los parámetros asociados son:

⁵En realidad estos valores se encuentran fuera de escala de forma de visualizar mejor los parámetros del modelo.

$[n, m, \delta]$	$[a_1, \dots, a_n]$	$[b_0, \dots, b_{m-1}]$	c	γ	ε_m
[3, 2, 3]	$[-2.1561, 1.5995, -0.4243]$	$[0.0087, 0.0082]$	0.2751	0.3451	$1704m^3.s^{-1}$
[3, 3, 2]	$[-2.1348, 1.5735, -0.4162]$	$[0.0094, 0.0057, 0.0058]$	0.3553	0.2529	$1541m^3.s^{-1}$
[3, 4, 2]	$[-2.0852, 1.5082, -0.3957]$	$[0.0058, 0.0079, 0.0064, 0.0061]$	0.2516	0.3559	$1449m^3.s^{-1}$

Table 2: Parámetros de los modelos seleccionados

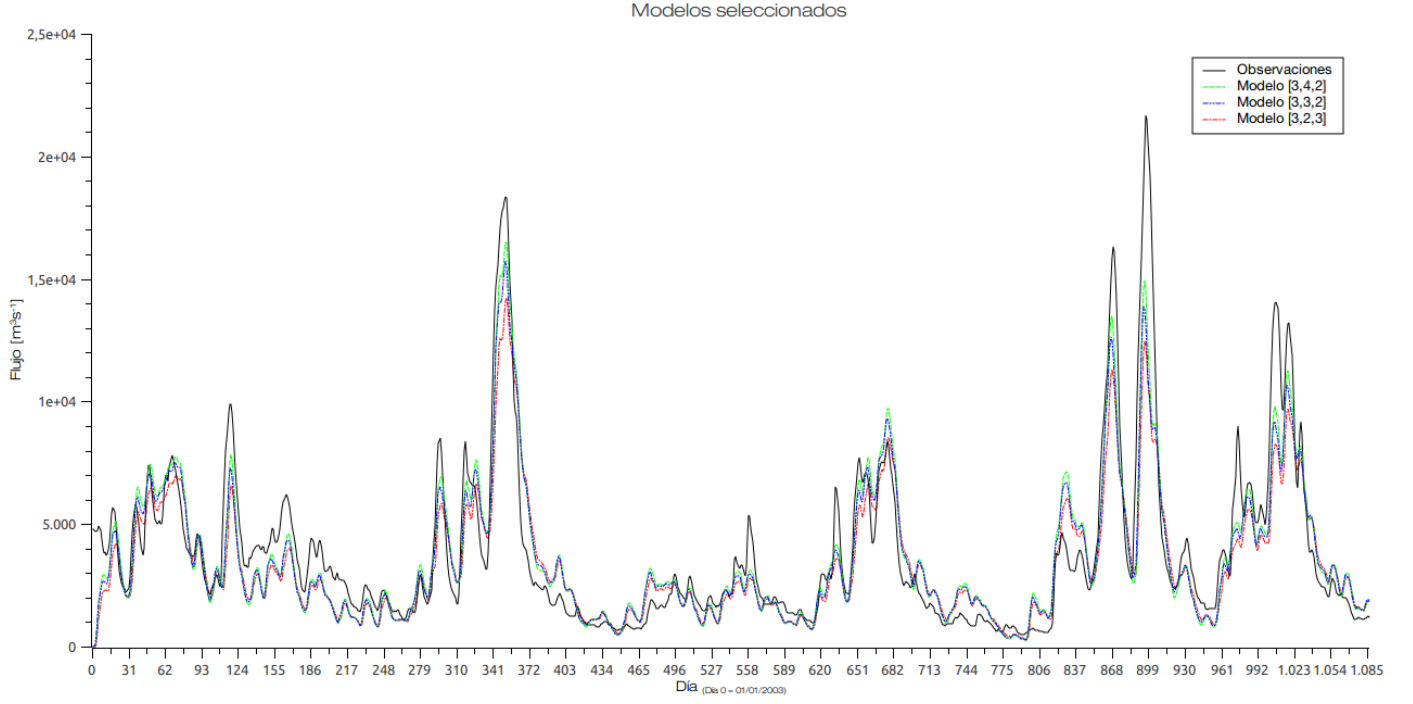


Figure 4: Modelos seleccionados y observaciones (1087 días).

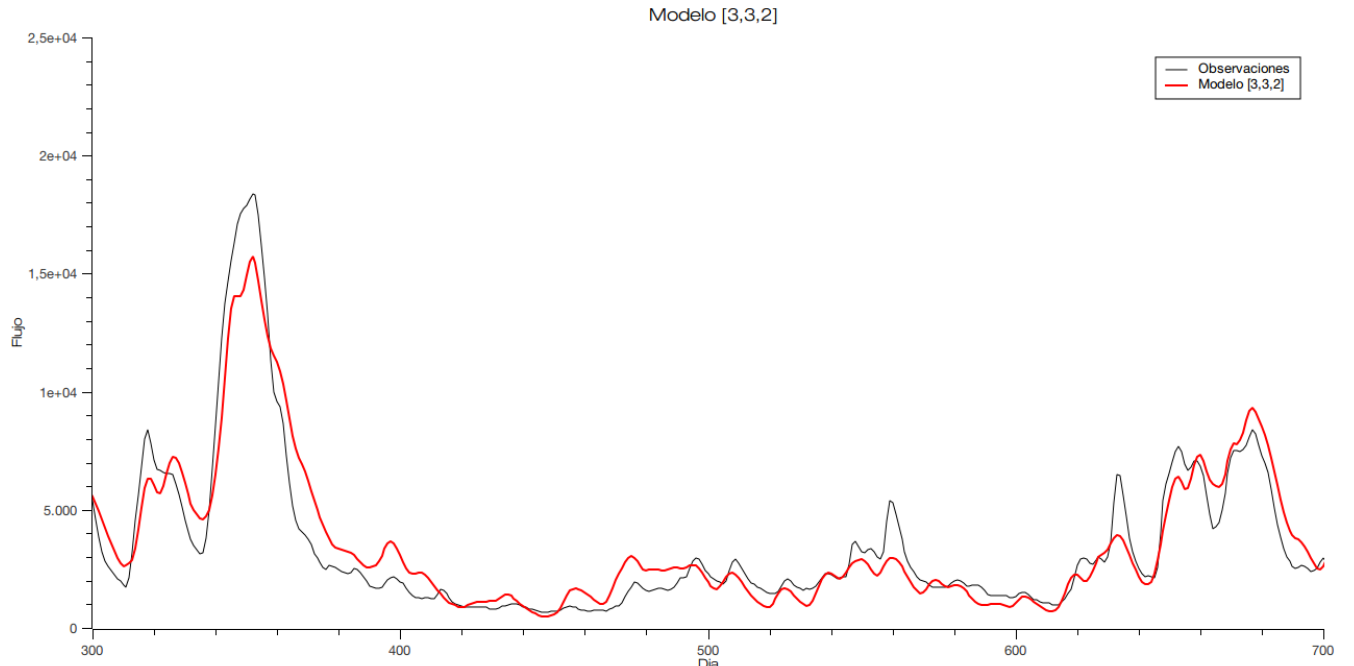


Figure 5: Modelo [3,3,2] y observaciones (400 días).

Filtrado Kalman

4.2 Modelado en términos de ecuaciones de estado

El flujo de la cuenca esá dado por:

$$y_t = -a_1 y_{t-1} - a_2 y_{t-2} - \dots - a_n y_{t-n} + b_0 u_{t-\delta} + b_1 u_{t-\delta-1} + \dots + b_m u_{t-\delta-m} \quad (6)$$

Se trata de un proceso ARMA(p,q) y puede escribirse de la siguiente manera como modelo de estados:

Ecuación de transición

$$\mathbf{x}_t = \begin{pmatrix} -a_1 & 1 & 0 & \dots & 0 \\ -a_2 & 0 & 1 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ -a_{p-1} & 0 & \dots & 0 & 1 \\ -a_p & 0 & \dots & 0 & 0 \end{pmatrix} \mathbf{x}_{t-1} + \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-1} \\ b_p \end{pmatrix} u_{t-\delta} \quad (7)$$

Donde $p = \max(n, m)$.

Ecuación de observación

$$\mathbf{y}_t = (1, 0, \dots, 0)^T \mathbf{x}_t$$

Si al modelo le incluimos ruido de observación $e(t)$ y ruido de proceso $\zeta(t)$, entonces tenemos que:

$$\begin{aligned} x_t &= F x_{t-1} + G u_{t-\delta} + \zeta(t) \\ y_t &= h^T x_t + e(t) \end{aligned}$$

Las matrices de covarianza del ruido quedan definidas por:

$$\begin{aligned} e(t) &\sim \mathcal{N}(0, \sigma^2) \\ \zeta(t) &\sim \mathcal{N}(0, Q_r) \end{aligned}$$

4.3 Algoritmo implementado

Dada la representación de estados estamos en condiciones de implementar un filtrado de Kalman de forma de minimizar el error cuadrático medio en la predicción. Para ello consideramos las siguientes matrices de covarianza para los ruidos de modelo y de observación:

$$Q = \sigma_q^2 \mathcal{I}_{p \times p}, \quad R = \sigma^2$$

El algoritmo implementado se basa en el propuesto por Young [1].

Predicción a priori.

$$\begin{aligned}\hat{x}_{t|t-1} &= F\hat{x}_{t-1} + Gu_{t-\delta} \\ P_{t|t-1} &= FP_{t-1}F^T + \sigma^2 Q_r \\ \hat{y}_{t|t-1} &= h^T \hat{x}_{t|t-1}\end{aligned}$$

Corrección a posteriori.

$$\begin{aligned}\hat{x}_t &= \hat{x}_{t|t-1} + K[y_{t-\delta} - \hat{y}_{t-\delta|t-\delta-1}] \\ K &= P_{t|t-1}h[\sigma^2 + h^T P_{t|t-1}h]^{-1} \\ P_t &= P_{t|t-1} - Kh^T P_{t|t-1} \\ \hat{y}_t &= h^T \hat{x}_t\end{aligned}$$

Los parámetros del ruido σ^2 y σ_q^2 son calibrados manualmente de forma de minimizar el error cuadrático medio. A continuación graficamos la predicción mediante filtrado de Kalman:

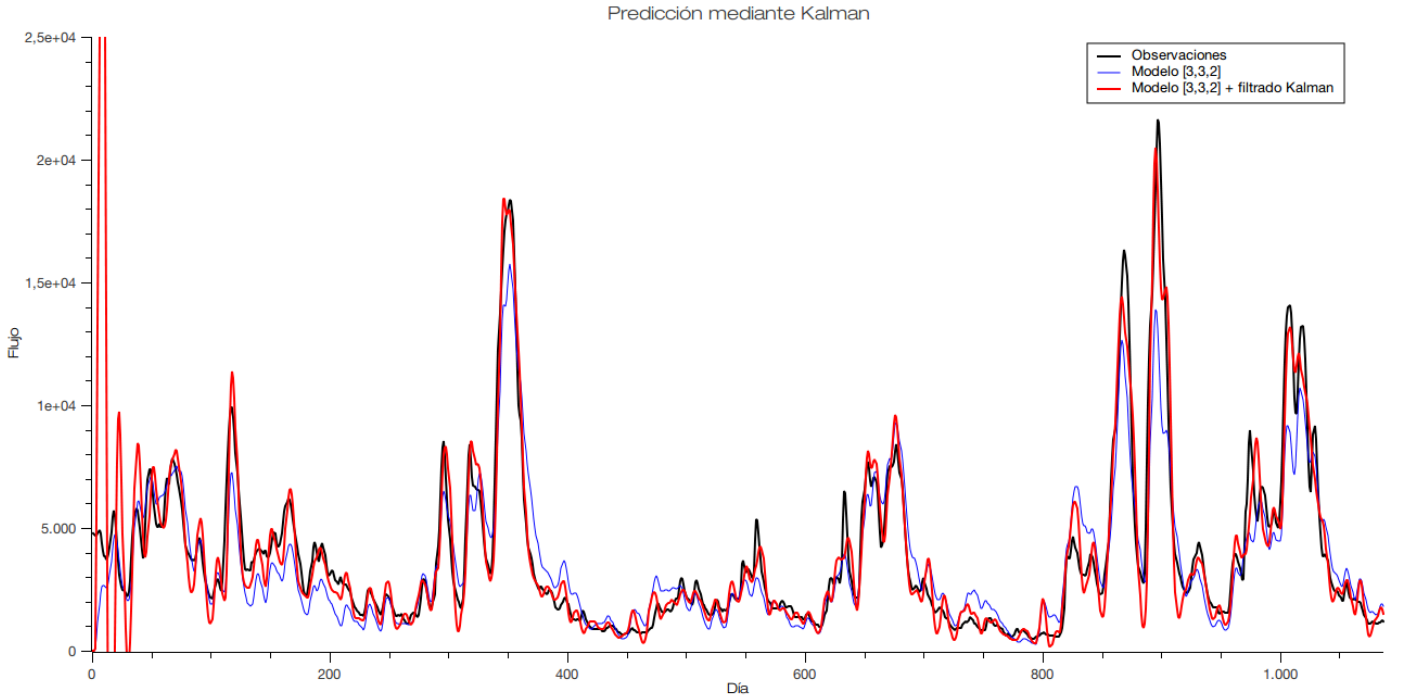


Figure 6: Predicción mediante filtrado de Kalman (1087 días).

En la siguiente tabla incluimos los parámetros del ruido de modelo y de observación, y el error cuadrático medio que resulta del filtrado de Kalman:

$[n, m, \delta]$	σ_q^2	σ^2	$\varepsilon_m \text{modelo}$	$\varepsilon_m \text{Kalman}$
$[3, 3, 2]$	0.002	0.001	$1541m^3.s^{-1}$	$1045m^3.s^{-1}$

Table 3: Modelos seleccionados

Podemos apreciar que el ε_m disminuye considerablemente al utilizar la predicción basada en filtrado Kalman.

5 Validación y conclusiones

Para validar los resultados es necesario estimar el flujo del río utilizando una serie temporal que no esté correlacionada con la serie utilizada en la calibración. Mientras que para la calibración se utilizó una serie de precipitaciones de los años 2003 a 2005, para la validación se utilizarán las series de precipitaciones de los años 2008 y 2009.

A continuación graficamos las observaciones de flujo y la estimación correspondiente:

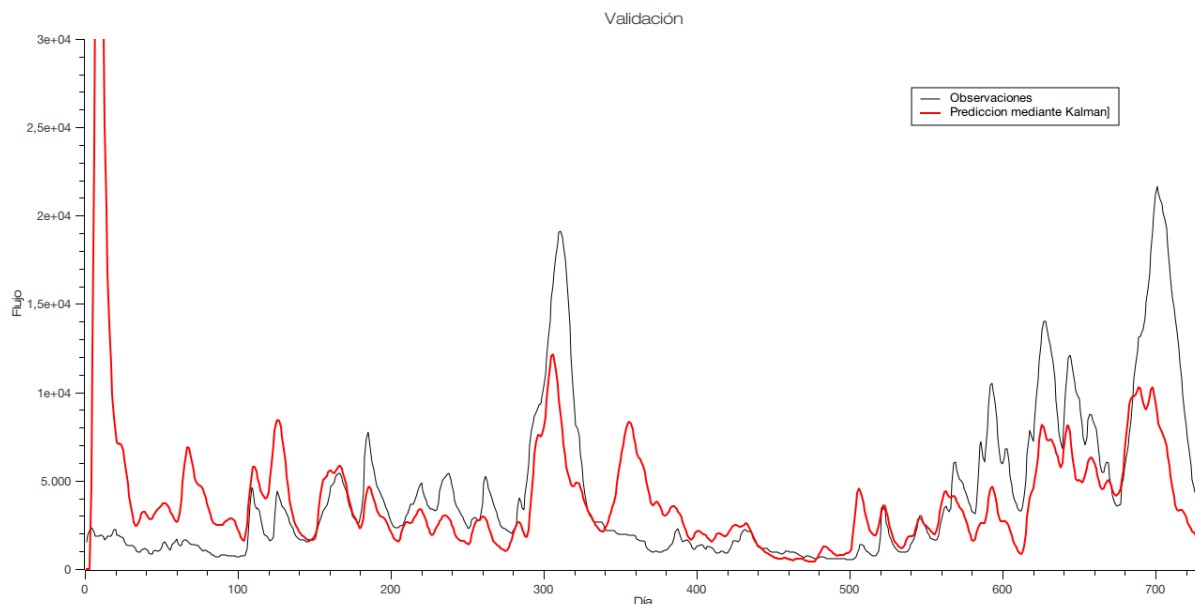


Figure 7: Validación de la predicción (731 días).

El ε_m es de $3038 m^3 \cdot s^{-1}$, vemos que la predicción en este caso tiene una baja performance en relación a la predicción realizada en la etapa de calibración (el error cuadrático medio se multiplica aproximadamente por 3), esto puede explicarse fundamentalmente a un error de modelado.

El error de modelado puede deberse por un lado al gran tamaño de la cuenca (aproximadamente $200.000 km^2$, mientras que la predicción de Young se realiza en una cuenca de $500 km^2$) y por otro a la utilización de un único bloque SLIT, en la bibliografía consultada propone adicionalmente colocar otro bloque SLIT en paralelo. Esto permite colocar filtros con distintos días de retardo, lo que podría mejorar significativamente la performance de la predicción.

En relación al tamaño de la cuenca, puede ser conveniente subdividir la cuenca en otras de menor tamaño y colocar filtros en cascada. De esta manera los parámetros del modelo se adaptarían mejor a las especificidades de las regiones (por ejemplo zonas con distintos tipo de suelo), por otra parte el retardo entre las precipitaciones y el flujo en determinada altura del río tiene relación con la distancia que existe entre éstos, factor que podría justificar la subdivisión de la cuenca.

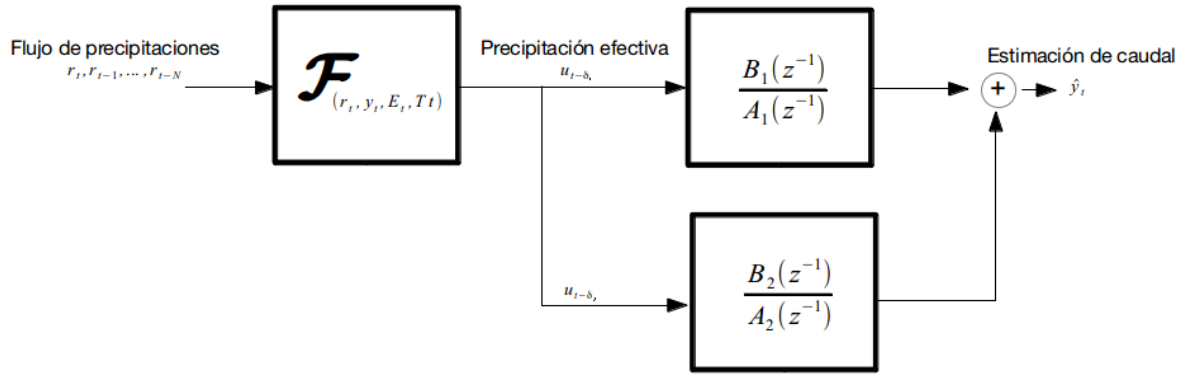


Figure 8: Incorporación de un bloque SLIT en paralelo, los bloques difieren en la entrada por el retardo de la precipitación efectiva.

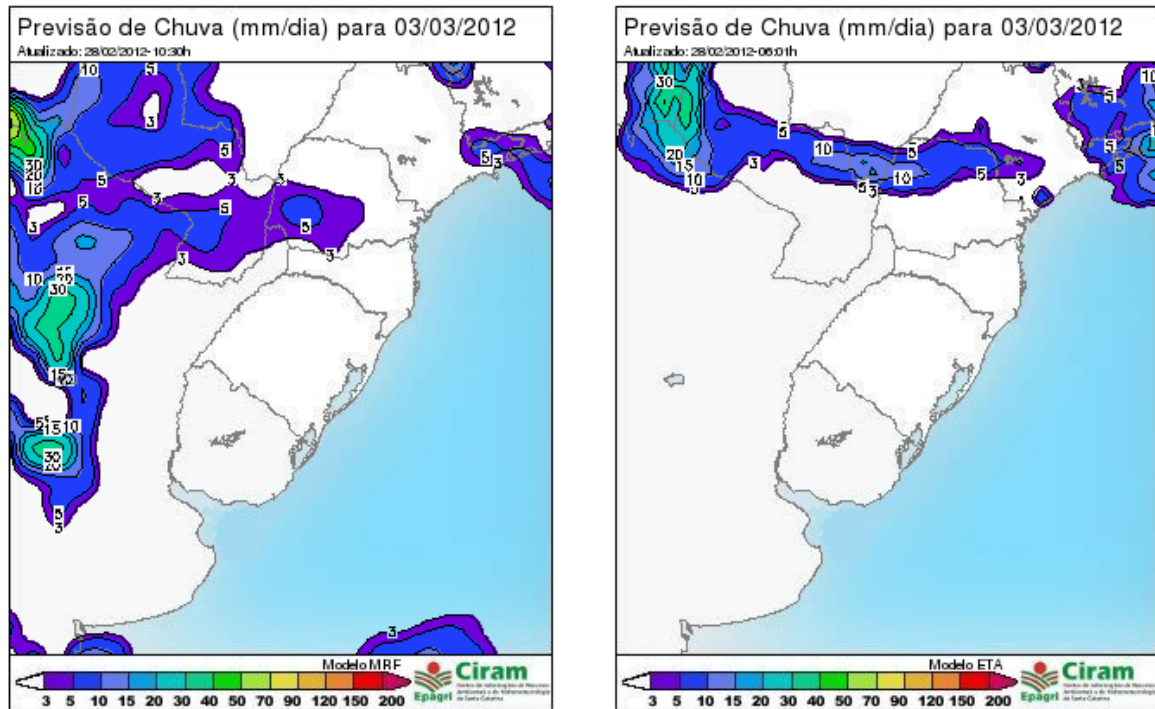


Figure 9: Incorporación de un bloque SLIT en paralelo, los bloques difieren en la entrada por el retardo de la precipitación efectiva.

Si con estas mejoras se logra una predicción confiable es posible luego aumentar el horizonte de predicción usando el pronóstico a t_p días de las precipitaciones en la cuenca, el nuevo horizonte sería $t_h = \delta + t_p$. A diferencia de la serie histórica, no sería necesario interpolar los datos de las estaciones sino que basta con procesar mapas que tengan el pronóstico de precipitaciones diarias, por ejemplo pueden utilizarse las imágenes producidas por el INPE ⁶ disponibles en la web del CIRAM que tienen pronósticos a 5 días.

⁶Instituto de Pesquisas Espaciais

References

- [1] **P. Young.** *Advances in Real-Time Flood Forecasting.* Centre for Research on Environmental Systems and Statistics (CRES), Lancaster University, 2002.
- [2] **R. Moore** *The PDM rainfall-runoff model.* Centre for Ecology and Hydrology, Wallingford,UK, 2007.
- [3] **C. Zotelo, S. Martín, I. Camilloni.** *Estimación del tiempo de retardo de la onda de crecida en la cuenca superior del Río Uruguay.* Universidad de Buenos Aires, 2006.
- [4] **M. Hayes.** *Statistical Difital Signal Processing and Modeling.* Georgia Institute Of Technology, 1996.