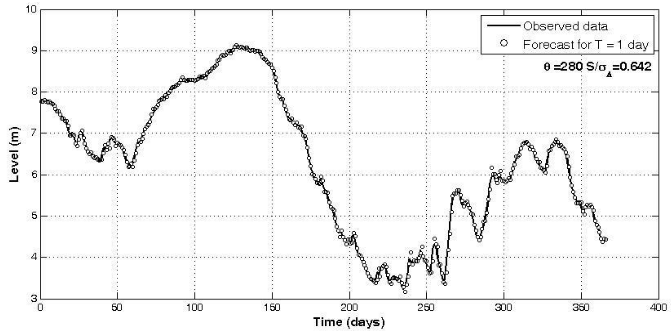
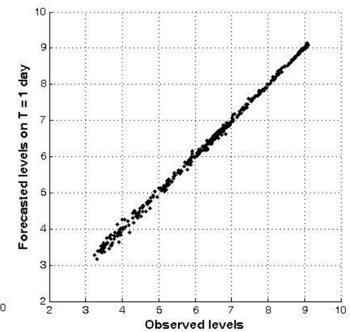
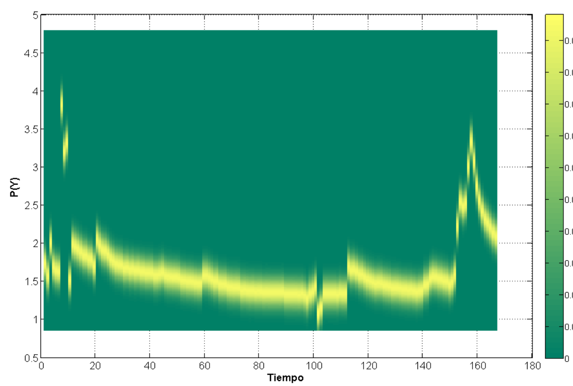
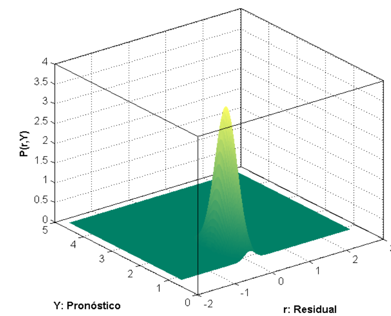
Pontificia Universidad JAVERIANA Bogotá

Adaptive linearization of phase space. A hydrological case study









Héctor Angarita, MSc

Efraín Domínguez, PhD*

Pontificia Universidad Javeriana

Environmental and Rural Studies Faculty

**Presenting author,*

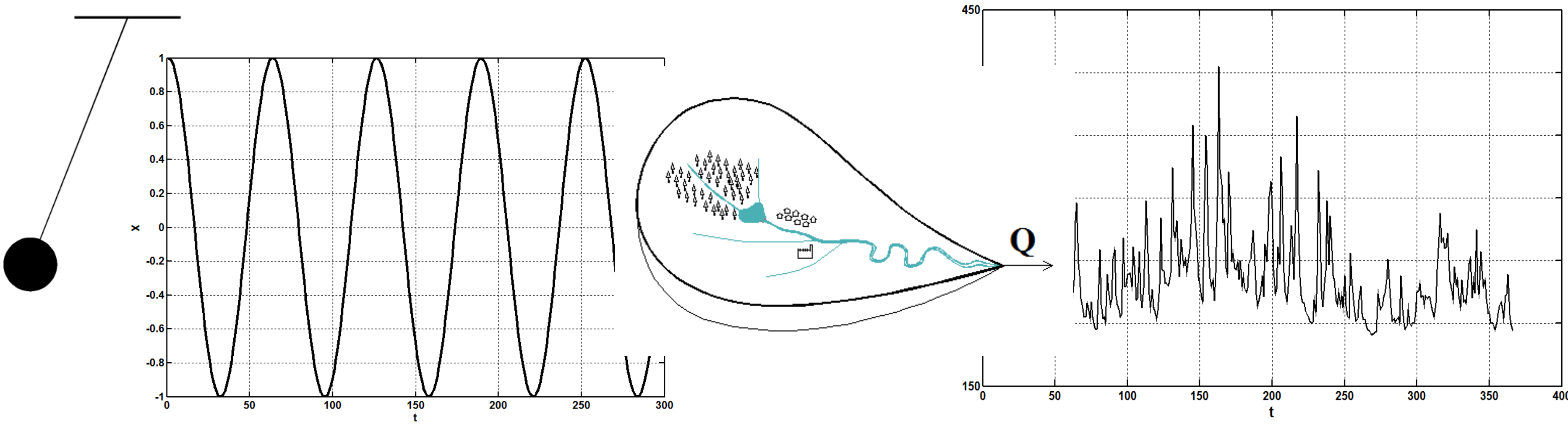
e-mail: e.dominguez@javeriana.edu.co

European Geosciences Union General Assembly 2013

Vienna | Austria | 07 – 12 April 2013

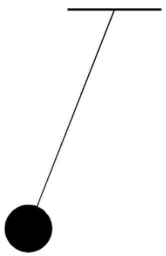
INTRODUCTION

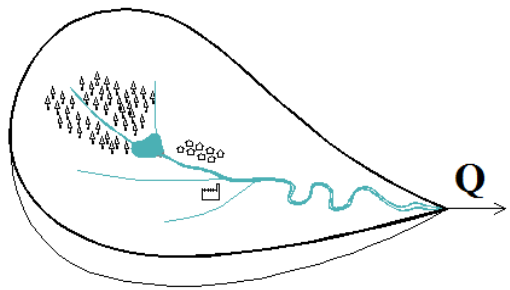
- A key issue in the study of natural systems is how to identify patterns or structures in data sequences in order to reduce the uncertainty about non-observed states. :



INTRODUCTION

- *Algorithmic complexity* of a system(Chaitin 1987): for an observer, a process will be more 'complex' as requiring more information to describe the observed trajectories.





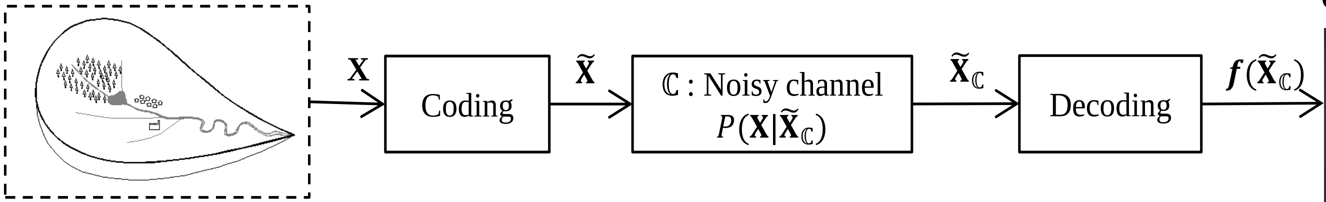
Minimum algoritmic complexity → **SAC System** Significant algoritmic complexity → Maximum algoritmic complexity

- Kolmogorov, Chaitín: The complexity can be measured as the minimum number of bits to be "transmitted" to communicate the entire sequence without any ambiguity.

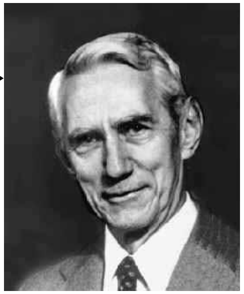
INTRODUCTION

- From the perspective of information theory, the observed randomness of a signal can be explained, to some degree, by the limited observation capabilities of the actual process:

Hydrological system:



Observer:



$\mathbf{X}$: Set of actual system signals

$\tilde{\mathbf{X}}$: Set of observables

$\tilde{\mathbf{X}}_{\mathbb{C}}$: Observed signals at channel capacity

$P\left(\mathbf{X}|\tilde{\mathbf{X}}_{\mathbb{C}}\right)$: Probability of system signal given the available observations

- The capability to completely capture the signal dynamics is related to the channel bandwidth, measured as the maximum sampling frequency – f, which, if higher than a given threshold – f_S, the signal can be completely recovered with virtually zero probability of error for any given signal to noise ratio

Pontificia Universidad JAVERIANA Bogotá

Methods

- In a SAC system, is necessary to adopt a formalization of the uncertainty in the data and its transition properties, i.e. an stochastic differential equation (Langevin):

$$\frac{d\mathrm{X}_i}{dt} = g_i(\mathbb{X}, t) + h_i(\mathbb{X}, t) * \Gamma(t)$$

- Where
  - $g_i(\mathbb{X}, t)$ is the deterministic kernel of the vector field
  - $\Gamma(t)$ is a non-periodic and irregular signal (noise) and
  - $h_i(\mathbb{X}, t)$ is a transformation that adjusts the noise dynamic influence.
  - In this equation, each X_i is a random process. Consequently, the solution of the equation is the evolution of the probability density curve $P(\mathbb{X}, t)$ given a set of initial conditions and the uncertainty of the system.

Methods

From the previous equation, If assumed that for some type of systems, local observations at sampling frequency f, contain enough information to estimate unobserved states within a lead time interval - T, a simple *adaptive linearization method* can be derived to iteratively calculate/update optimal vector field estimation from incoming data:

$$\check{x}_{i,t+T} = \sum_{k=1}^{m} \sum_{j=1}^{\rho} c_{i,j,k}(\mathrm{t}) \mathrm{W}_{f_{j,k}}(t) + \gamma_i(t) + \varepsilon(t) \qquad [1]$$

Where:

- $\mathrm{W}_f(t)$ is a $m \times \rho$ matrix containing a subset of discrete observations in the interval $[t - \rho, t]$ at frequency f
- ε is a random variable distributed as $P(r|X = \widehat{X}_{t+T})$.
- Each $\mathrm{c}_{i,j,k}$ and γ_i are model coefficients

Identifying the optimal:

- Sampling frequency
- System state dimensionality: Shape and size of $\mathrm{W}_f(t)$ the $m \times \rho$ matrix containing a subset of discrete observations

$$\mathrm{W}_f(t) = \begin{bmatrix} \tilde{x}_{1,tt} & \cdots & \tilde{x}_{m,t-\rho} \\ \vdots & \ddots & \vdots \\ \tilde{x}_{1,t-\rho} & \cdots & \tilde{x}_{m,t-\rho} \end{bmatrix}$$

- Parameterization window length – θ: is defined as the span of local states of the system in the interval $[t-\theta, t]$: $\mathbf{W}_f(t)$, $\mathbf{W}_f(t-1/f)$, ..., $\mathbf{W}_f(t-\theta/f)$ to determine c_{ijk} and γ_i at instant t

Pontificia Universidad JAVERIANA Bogotá

Optimality criteria

- To a solution at the time t, if the prediction error defined as the following differences:

$$\Delta_i = \hat{x}_{i_{t+T}} - \tilde{x}_{i_{t+T}}$$

- Usually, an exhaustive search algorithm using the objective function $\min \rightarrow (S/\sigma_\Delta)$, defined as the ratio between the standard deviation of square of the prediction error S and standard deviation of the signal increments, is able to identify optimal state dimensionality – $\mathbf{W}_f(t)$, Parametrization window length – θ and sampling frequency – n/f in a reasonable time:

$$S = \sqrt{\sum(\Delta)^2}/N$$

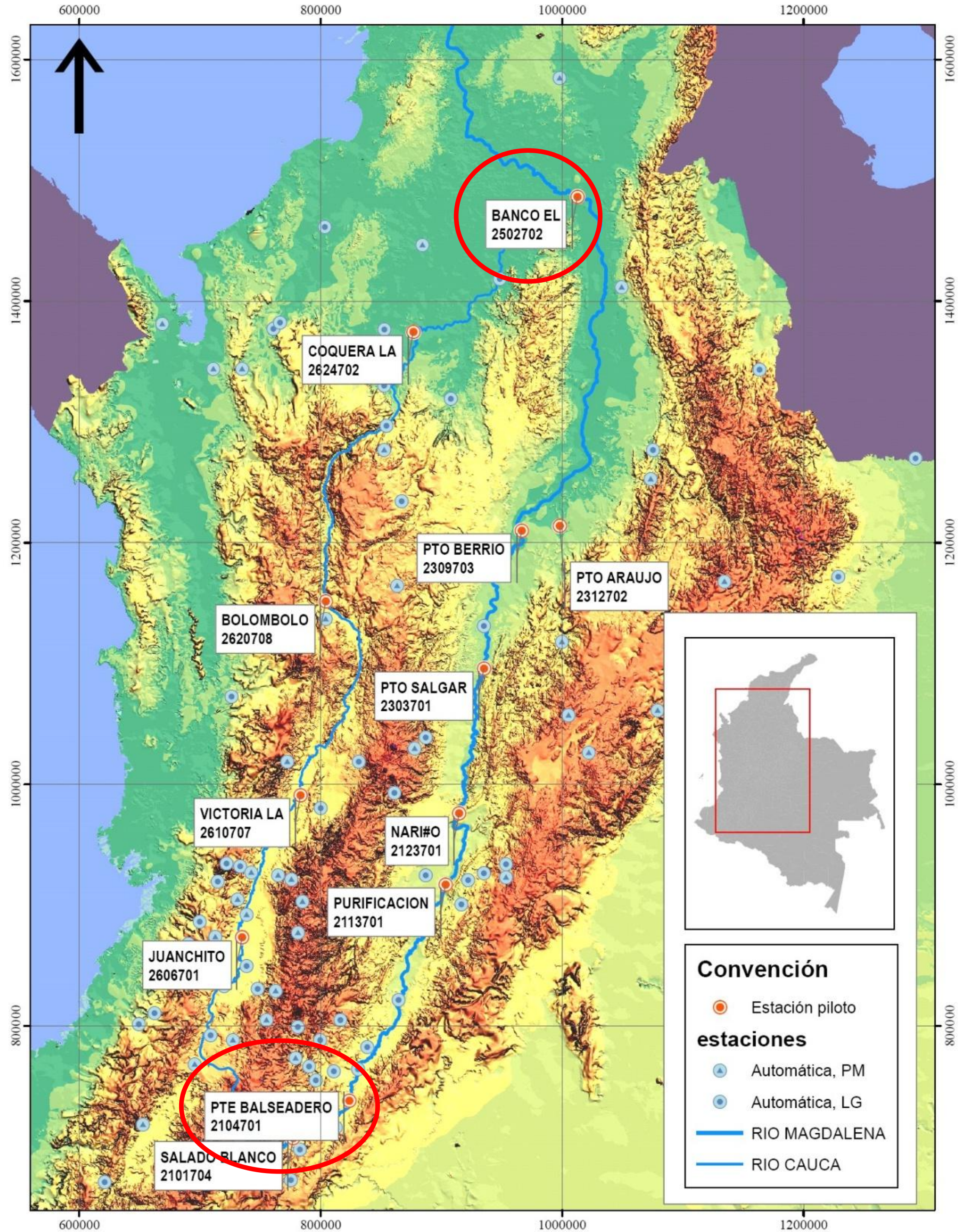
$$\sigma_\Delta = \sqrt{\sum\left(\tilde{x}_{i_{t+T}} - \tilde{x}_{i_t}\right)^2/N}$$

- According to the Russian Hydrometeorological Center (Appolov 1974), the value of S/σ_Δ must be less than 0.8 to accept that the model exceeds the naïve forecast.

Pontificia Universidad JAVERIANA Bogotá

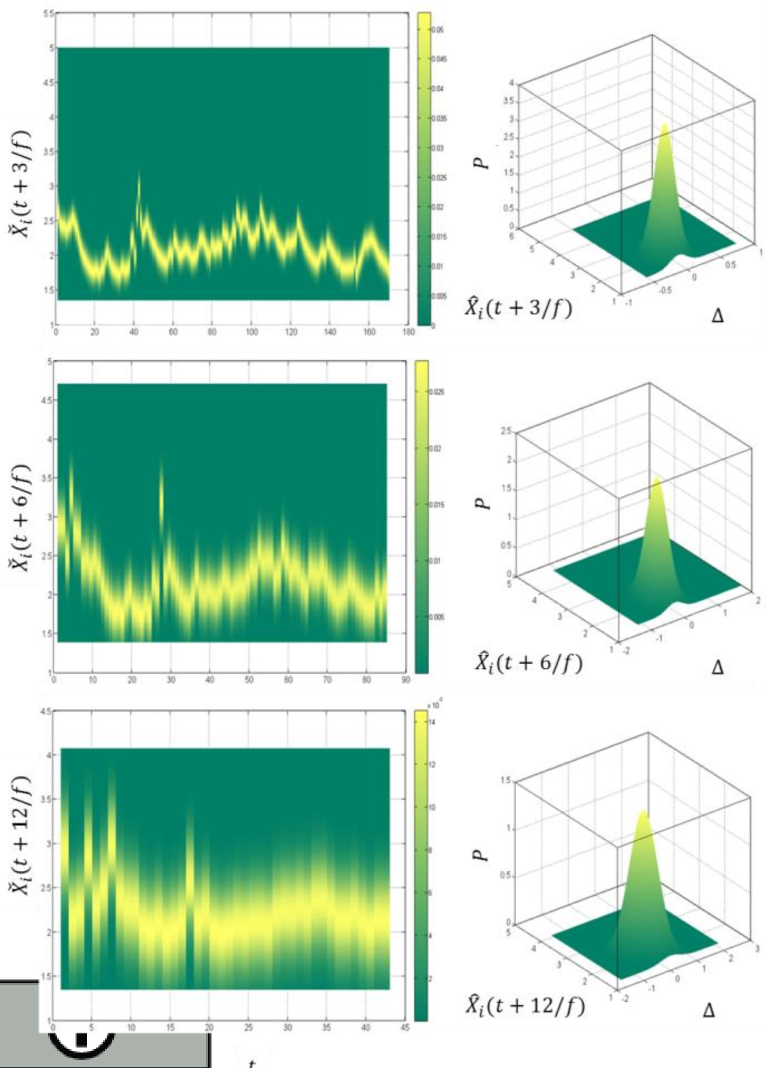
Aplication 1
Hydrological forecasting

The eq. (1) serves as a mathematical operator feed by real or near-real-time data for forecasting the future state of water levels at lead-time T

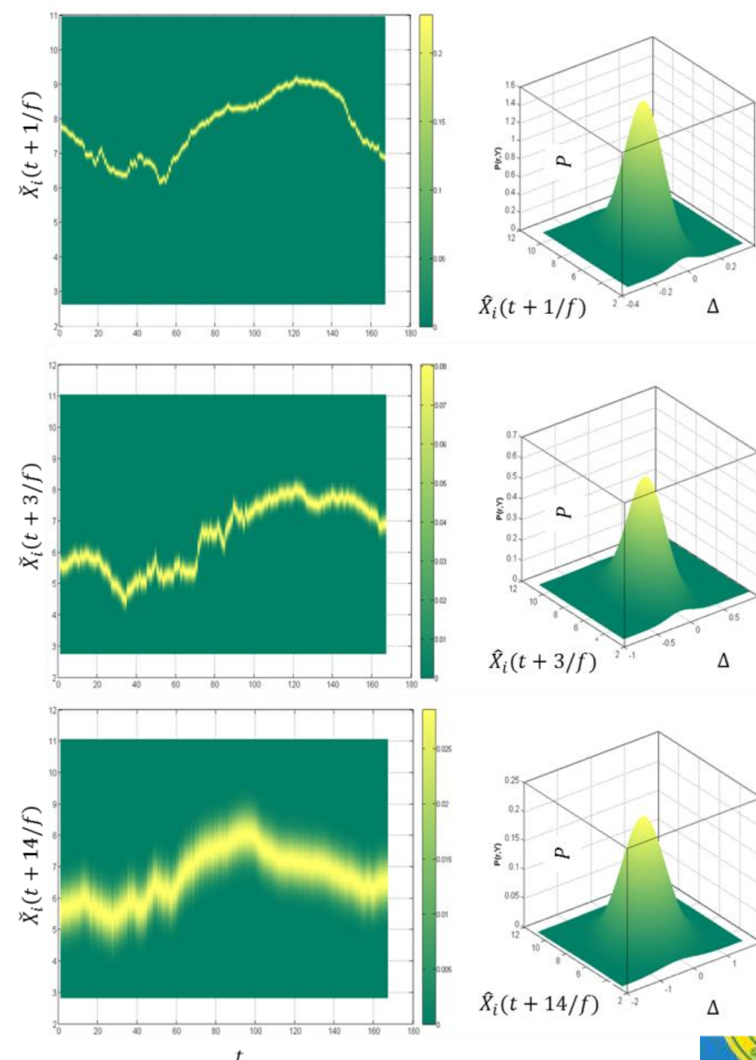


Examples of optimal models output at different sampling frequencies:

Pte Balseadero



El Banco

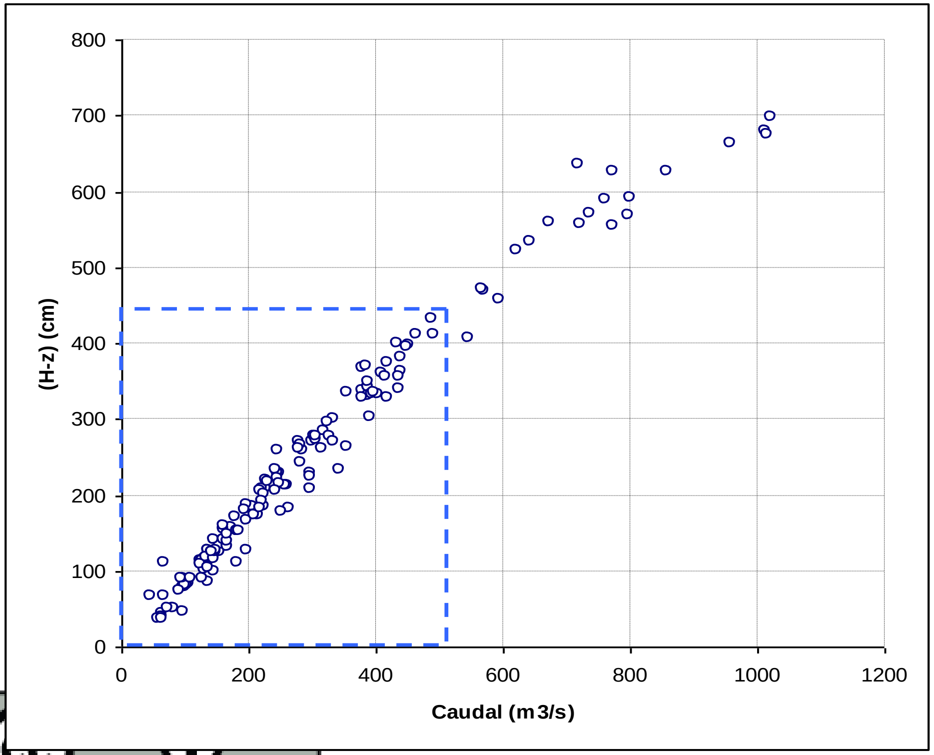


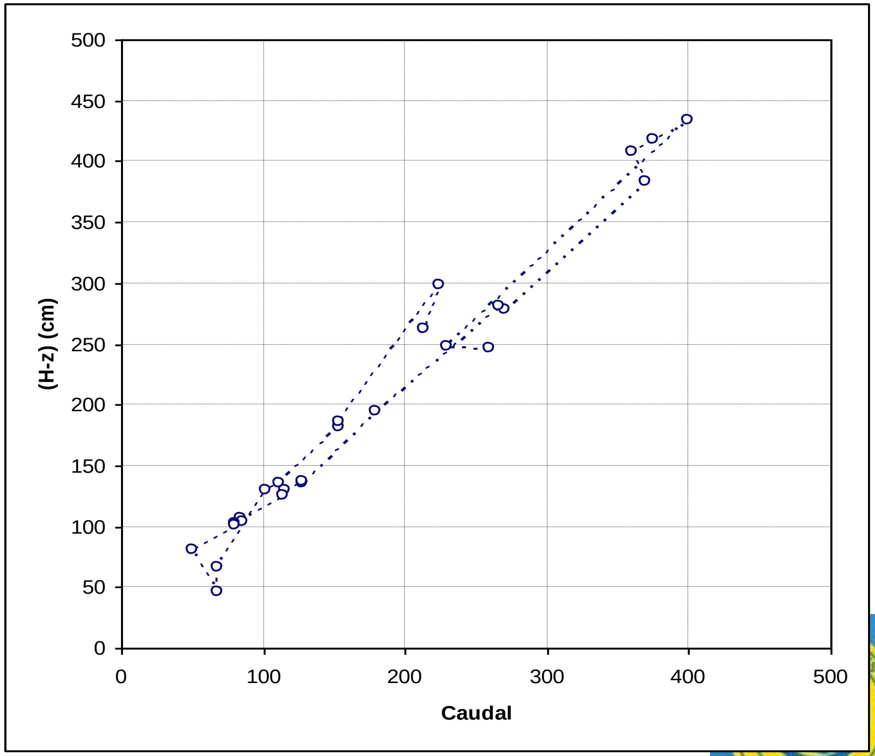
Model comparison:

Estación	AMI				S/σΔ				Forecast feasibility			
	Lead Time (Hours)				Forecast Lead Time (Hours)				Forecast Lead Time (Hours)			
	1	3	6	12	1	3	6	12	1	3	6	12
Bolombolo (2620708)	0.99	0.97	0.92		0.768	0.955	0.816		Viable			
La coquera (2624702)	1.00	0.98	0.95	0.91	0.759	0.762	0.855	0.857	Viable	Viable		
El Banco (2502702)	1.00	1.00	1.00	1.00	0.913	0.703	0.656	0.616		Viable	Viable	Viable
Narino (2123701)	1.00	0.99	0.94		0.500	0.628	0.858			Viable		
Pte. Balseadero (2104701)	0.90	0.95	0.82	0.48	0.988	0.774	0.828	0.919		Viable		
PtoAraujo (2312702)	1.00	0.94	0.97	0.93	0.738	0.953	0.845	0.790	Viable			
Pto. Berrio (2309703)	0.90	0.95	0.82	0.48	0.812	0.725	0.874	0.818		Viable		
Pto. Salgar (2303701)	1.00	0.99	0.98		0.436	0.533	0.520		Viable	Viable	Viable	
Purificación (2502702)	1.00	0.97	0.86	0.72	0.553	0.878	0.917	0.925	Viable			
Salado Blanco (2101704)	0.98	0.93	0.83	0.66	1.107	1.116	0.915	0.858				
La Victoria (2610707)	1.00	1.00	0.99	0.97	1.033	0.849	0.923	0.964				

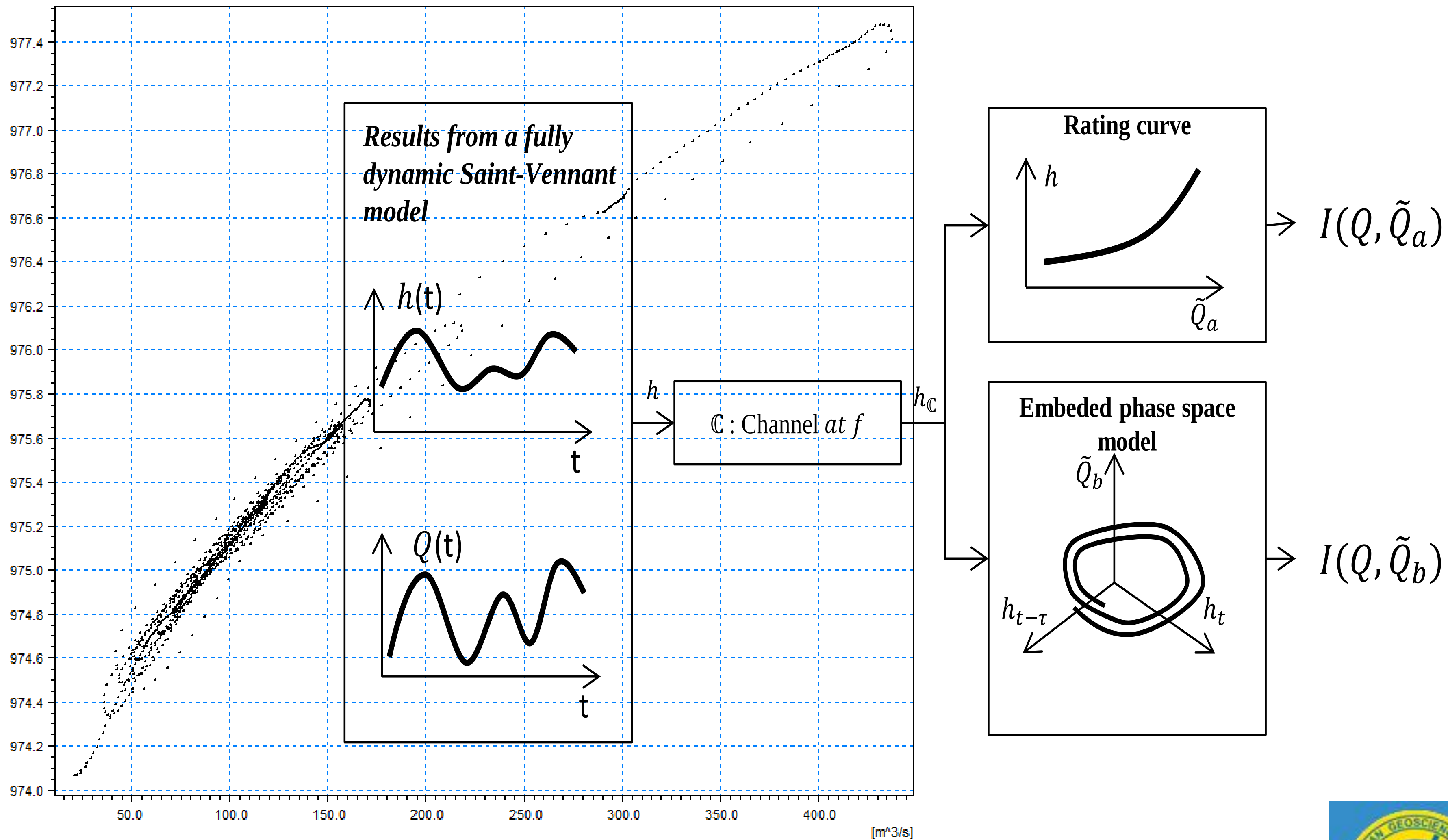
Aplication 2: Phase Space Base Stage-Discharge Decoding

- Conventionally the stage-discharge decoding function is a bi-univocal relationship between “real” flow and water level at the point where the measurement is performed.
- Usually is determined by fitting a curve to known points of level and flow records obtained in gauging campaigns.





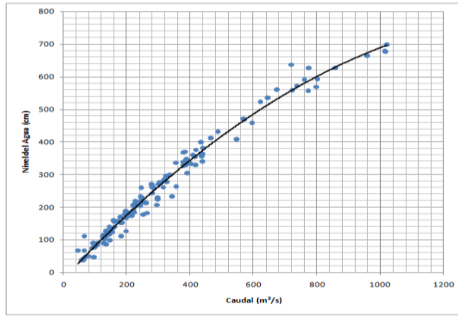
Aplication 2: Phase Space Base Stage-Discharge Decoding



Pontificia Universidad JAVERIANA Bogotá

Conventional :

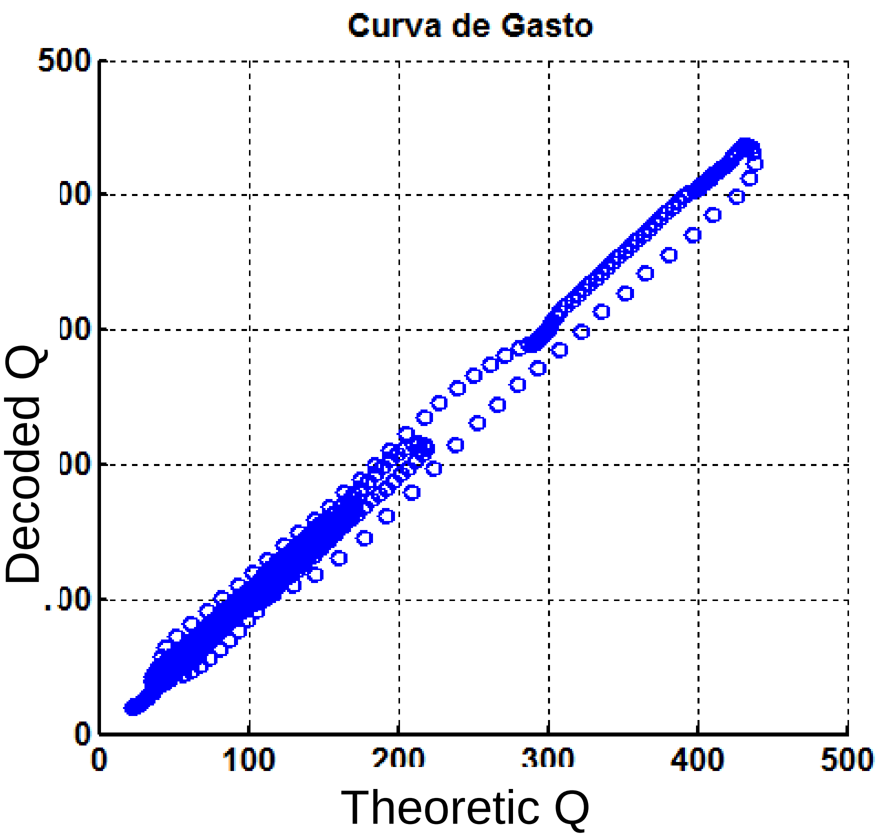
Q(t) = f(h(t))

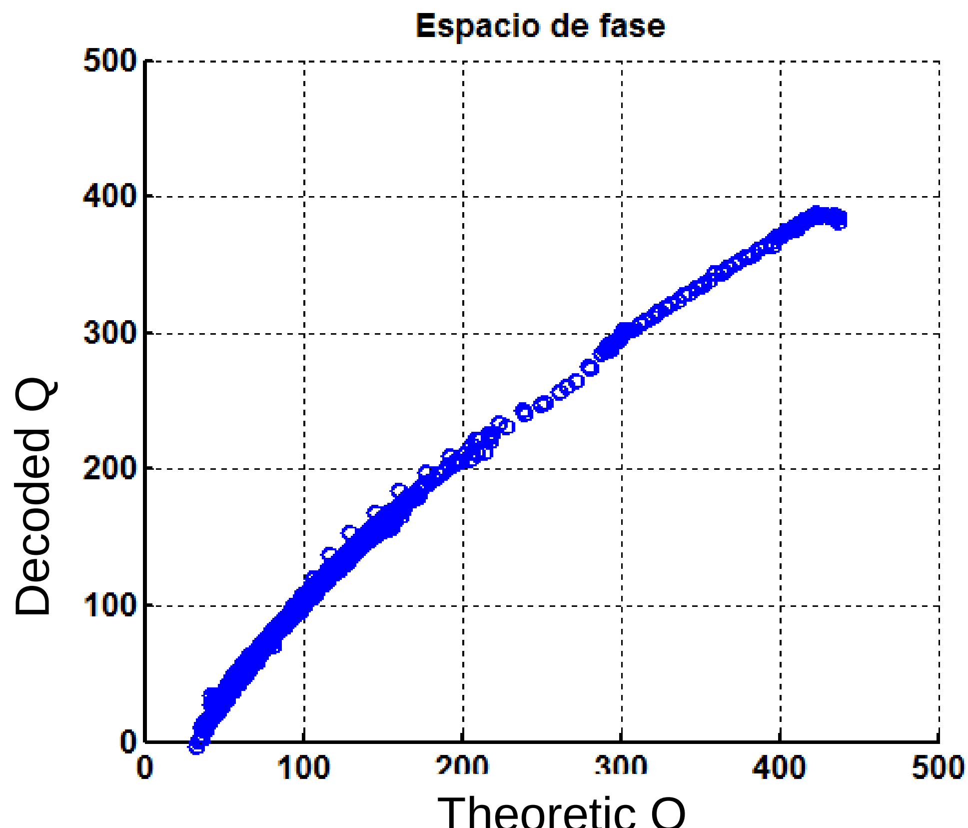


Phase space decoder:

$$[Q_t] = \{L_1[\mathbf{W}_1(t)]\} = B \cdot \mathbf{Wt} + c_{te}$$

$$\mathbf{Wt} = \begin{bmatrix} h_t \\ h_{t-1} \\ h_{t-2} \end{bmatrix}$$





Conclusions (1)

- Here is proposed a simple adaptive linearization method to characterize signals from SAC systems, with an empirical approach to define optimum sampling frequency. This approach is associated with the channel capacity theorem proposed by Shannon, in particular, the sampling frequency relation with the ability to accurately predict the state transitions of a system.

- The method leads to computationally simple adaptive models. This feature offers some interesting attributes for forecasting applications in real time. The first and most important is the ability to continually adjust and unsupervised nonstationary dynamics of the system or even build new models in case of failures or assimilation processes of information transmission.

Conclusions (2)

- For example, in the application of the method in forecasting levels and flows in the seasons studied, the implementation of comprehensive process of identifying optimal operator took on average less than 2 minutes for forecast horizons of hours and up to 14 minutes in forecasts 14 days.

- With regard to the implementation of a stage discharge decoder, a technique was proposed for estimating functional relationships based on linearized phase space operators. It is generally found that according to the performance criterion proposed (AMI> 0.95), the technique is applicable and leads to a reduction of the uncertainty in the flow decoding. However, the results show quadratic bias in the decoded values.

Pontificia Universidad JAVERIANA Bogotá

Poster Presentation

Attendance Fri, 12 Apr, 10:30–12:00 / Red Posters

EGU2013-3574

The FAST-T approach for operational, real time, short term hydrological forecasting: Results from the Betania Hydropower Reservoir case study

- Efraín Domínguez, Hector Angarita, Thomas Rosmann, Zulma Mendez, and Gustavo Angulo