Kalman Filter and Box-Jenkins Techniques for Monthly and Annual Streamflows Prediction in Northern Algeria

Boukharouba, K.; Kettab, A.

Kalman Filter and Box-Jenkins Techniques for Monthly and Annual Streamflows Prediction in Northern Algeria

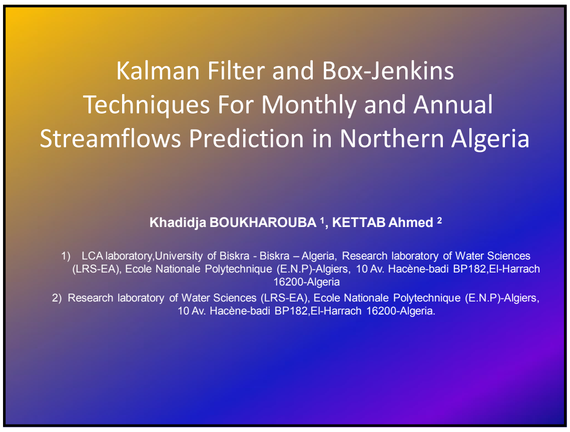
K. Boukharouba[1,*], A. Kettab[2]

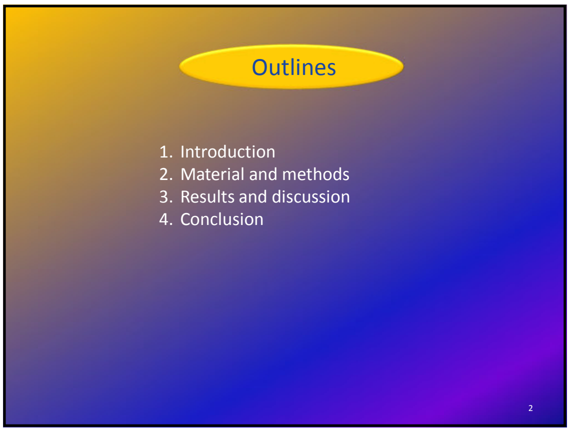
[1]University of Biskra, Faculty of SESNV, department of Biology, LCA laboratory, Biskra, Algeria
[2]National Polytechnic School (ENP) of Algiers, Laboratory of water sciences LRS-EAU, 10 Av. Hacene Badi, 16200 El harrache, Algiers, Algeria
[*]Corresponding author

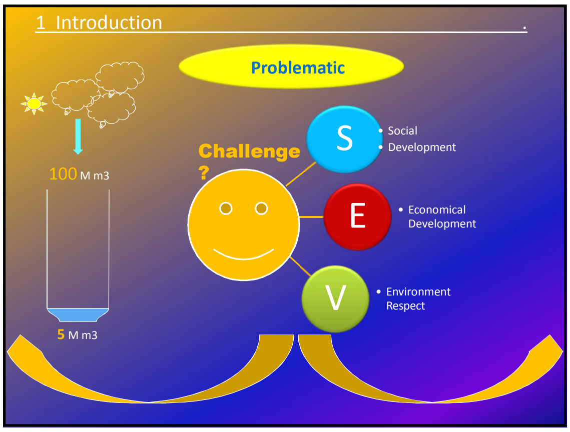
Email: b_malikadz@yahoo.com

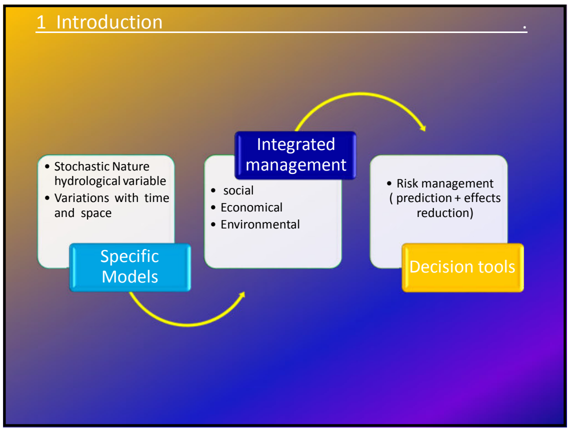
Keywords: Kalman Filter, Box-Jenkins, Stream Flows, Prediction, Algeria

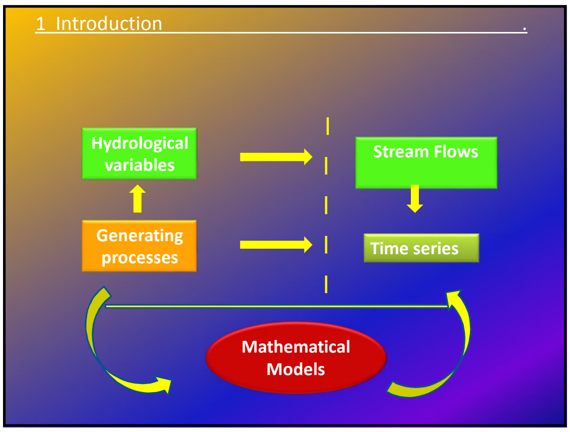
Analyzing stream flow records can give significant ideas for both past and future characteristics of stream flows. Therefore, recording and analyzing stream flow measurements have highly important roles in water resource management, planning and design. Unfortunately, hydrological processes, such as stream flows, are very complex because they are dependent of many hydro-physiographic parameters, additionally they vary with time and space. To cope with these difficulties, two different techniques have been applied to the modeling and prediction of the monthly and annual stream flow data in Northern Algeria: 1) the conventional time series approach of Box-Jenkins (BJ), a class of stochastic processes that is known by its good modeling of hydrological time series, so as stream flows, and many other environmental relating variables. Such time series are generally characterized by a strong temporal dependence between successive observations. In addition, they can be represented by some linear models, which means that the concerned time series can be considered as the output of a linear system whose input is a discrete white noise. Unfortunately, this class of stochastic processes presents the limitation of working under the restrictive assumptions of stationary and normality, in addition it needs large amount of available data, 2) the adaptive prediction approach of Kalman filtering (KF) is considered to be one of the most well-known and often-used significant mathematical tools that can be used for stochastic estimation from noisy measurements. It is as an optimal recursive data processing algorithm, which is constituted essentially by a set of 5 mathematical equations that implement a predictor-corrector type estimator that is optimal, in the sense that it minimizes the estimated error covariance, when some presumed conditions are met. Those equations are recursive and present the great advantage to provide accurately with the prediction error. The obtained results show the superiority of Kalman filter. The best model is sought using the performance measures including mean square error and determination coefficient.

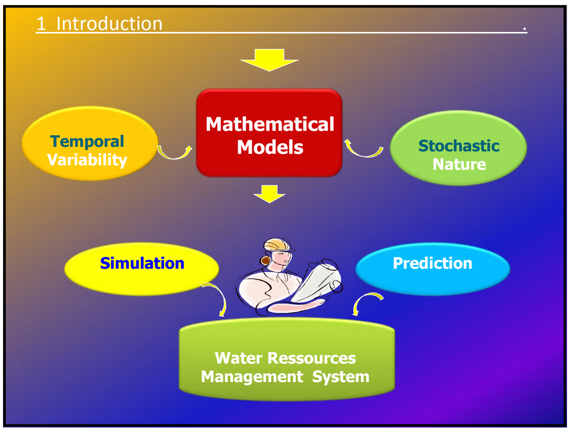

Kalman Filter and Box-Jenkins Techniques For Monthly and Annual Streamflows Prediction in Northern Algeria
Khadidja BOUKHAROUBA [1], KETTAB Ahmed [2]
1) LCA laboratory,University of Biskra - Biskra – Algeria, Research laboratory of Water Sciences (LRS-EA), Ecole Nationale Polytechnique (E.N.P)-Algiers, 10 Av. Hacène-badi BP182,El-Harrach 16200-Algeria
2) Research laboratory of Water Sciences (LRS-EA), Ecole Nationale Polytechnique (E.N.P)-Algiers, 10 Av. Hacène-badi BP182,El-Harrach 16200-Algeria.


Outlines
1. Introduction
2. Material and methods
3. Results and discussion
4. Conclusion
2


1 Introduction
Problematic
100 M m3
5 M m3
Challenge ?
S
• Social
• Development
E
• Economical Development
V
• Environment Respect


1 Introduction
• Stochastic Nature hydrological variable
• Variations with time and space
Specific Models
Integrated management
• social
• Economical
• Environmental
• Risk management (prediction + effects reduction)
Decision tools


1 Introduction
Hydrological variables
Stream Flows
Generating processes
Time series
Mathematical Models


1 Introduction
Temporal Variability
Mathematical Models
Stochastic Nature
Simulation
Prediction
Water Ressources Management System

2 Material and method

a) Material

Stream flows

(monthly and annual) observed at 10 hydro stations in northern Algeria (1968-1992)

Fig. 1 : Location of the hydrometric Stations in Northern Algeria.
(*Source: ANRH*)

2 Material and method

b) Methods

Linear Stochastiques Models

BOX-JENKINS (ARMA)

KALMAN FILTER

2 Material and method

BOX –JENKINS method

- Time serie Z_t with μ mean is considered as the output of a linear system whos input is a discrete white Gaussian noise ε_t .

$$\varepsilon_t \longrightarrow \boxed{\text{Linear System}} \longrightarrow Z_t$$

$$Z_t = \mu + \varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \cdots$$

$$B^k\varepsilon_t = \varepsilon_{t-k}$$

$$Z_t = \mu + (1 + \psi_1 B + \psi_2 B^2 + \cdots)\varepsilon_t$$

$$Z_t = \mu + \psi(B)\varepsilon_t$$

Look for the serie ε_t and the parameters ψ_i allowing to go from ε_t to Z_t

2 Material and method

Models

1) **AUTOREGRESSIVE model of p order AR(p)**

The centred output Z_t of the system is the weighted finite sum of p precedant values plus a random term ε_t .

$$Z_t = (\phi_1 Z_1 + \phi_2 Z_2 + \cdots + \phi_p Z_p) + \varepsilon_t$$

$$Z_t = (\phi_1 B^1 + \phi_2 B^2 + \cdots + \phi_q B^q)Z_t + \varepsilon_t$$

$$(1 - \phi_1 B^1 - \phi_2 B^2 - \cdots - \phi_q B^q)Z_t = \varepsilon_t$$

$$\boxed{\phi(B)Z_t = \varepsilon_t}$$

$\phi(B)$ is the operator AR(p) which is a polynôme in B with ordre p that converges for $|\phi i| < 1$, to insure the stationarity condition of the process.

2 Material and method

2) **MOVING AVERAGE model of q order MA(q)**

The centred output Z_t is the weighted sum of q+1 precedant values of a white noise ε_t .

$$Z_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots - \theta_q\varepsilon_{t-q}$$

$$Z_t = (1 - \theta_1 B^1 - \theta_2 B^2 - \cdots - \theta_q B^q)\varepsilon_t$$

$$\boxed{Z_t = \theta(B)\varepsilon_t}$$

$\theta(B)$ is the operator MA(q) which is a polynôm in B of ordre q that converges for $|\theta i| < 1$, to insure the invertibility condition of the process.

2 Material and method

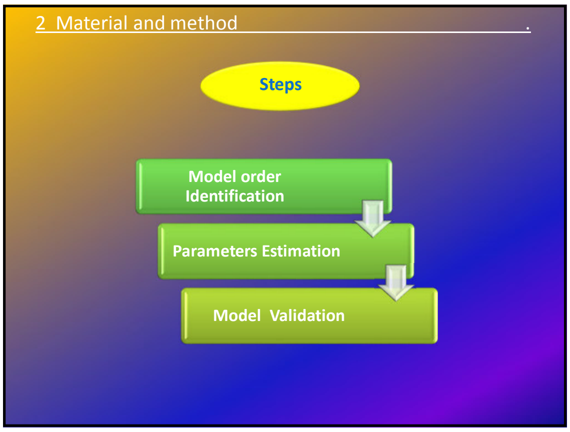
3) **AUTOREGRESSIVE and MOVING AVERAGE ARMA(p,q)**

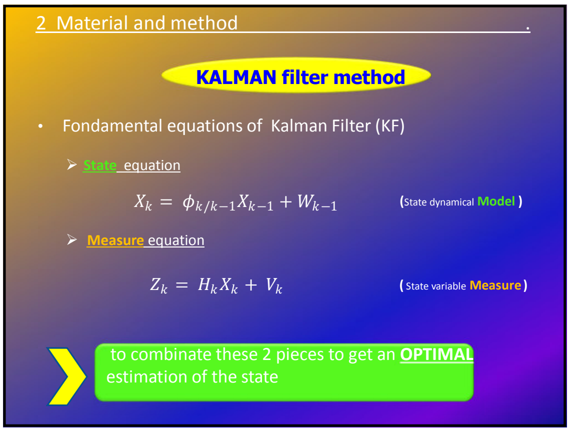
a mixture of AR(p) andMA(q) models

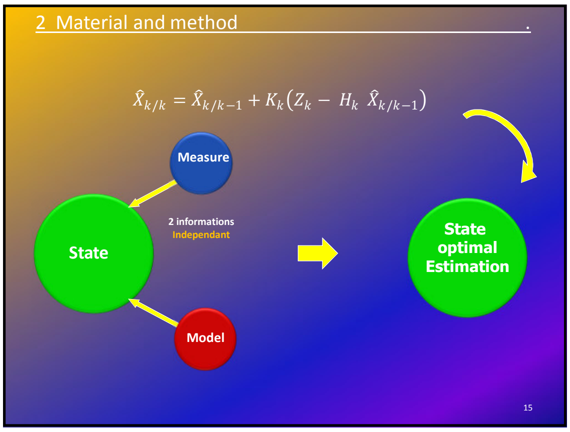
$$\boxed{\phi(B)Z_t = \theta(B)\varepsilon_t}$$

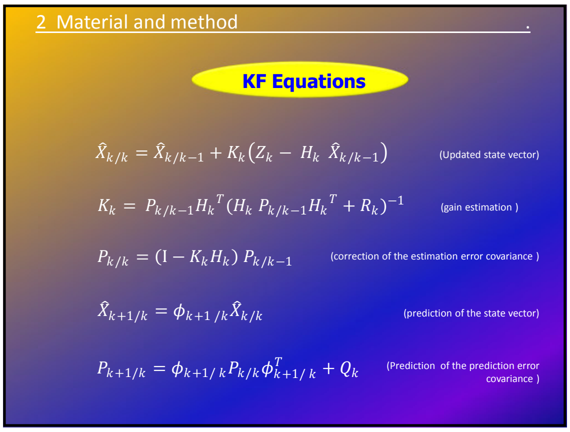
$$Z_t = [\theta(B)/\phi(B)]\ \varepsilon_t$$

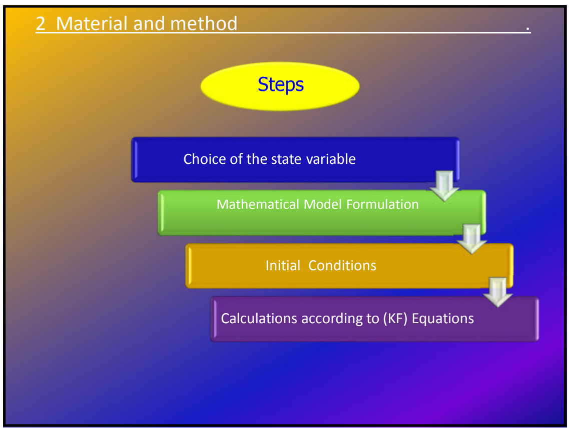
$\theta(B)/\ \phi(B)$ is the operator ARMA(p,q) which converges for $|\theta i| < 1$, and $|\phi i| < 1$, to insure the conditions of stationarity and invertibility.

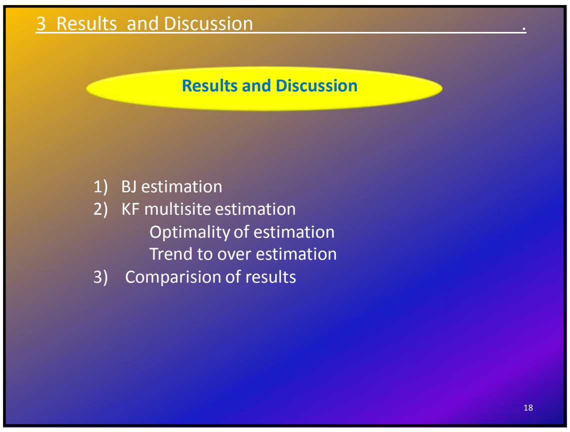

2 Material and method
Steps
Model order Identification
Parameters Estimation
Model Validation


2 Material and method
KALMAN filter method
Fondamental equations of Kalman Filter (KF)
State equation
$X_k = \phi_{k/k-1} X_{k-1} + W_{k-1}$
(State dynamical Model)
Measure equation
$Z_k = H_k X_k + V_k$
(State variable Measure)
to combinate these 2 pieces to get an OPTIMAL estimation of the state


2 Material and method
$\hat{X}_{k/k} = \hat{X}_{k/k-1} + K_k(Z_k - H_k \hat{X}_{k/k-1})$
Measure
2 informations
Independant
State
Model
State optimal Estimation
15


2 Material and method
KF Equations
$\hat{X}_{k/k} = \hat{X}_{k/k-1} + K_k(Z_k - H_k \hat{X}_{k/k-1})$
(Updated state vector)
$K_k = P_{k/k-1} H_k{}^T (H_k P_{k/k-1} H_k{}^T + R_k)^{-1}$
(gain estimation)
$P_{k/k} = (I - K_k H_k) P_{k/k-1}$
(correction of the estimation error covariance)
$\hat{X}_{k+1/k} = \phi_{k+1/k} \hat{X}_{k/k}$
(prediction of the state vector)
$P_{k+1/k} = \phi_{k+1/k} P_{k/k} \phi^T_{k+1/k} + Q_k$
(Prediction of the prediction error covariance)


2 Material and method
Steps
Choice of the state variable
Mathematical Model Formulation
Initial Conditions
Calculations according to (KF) Equations


3 Results and Discussion
Results and Discussion
1) BJ estimation
2) KF multisite estimation
Optimality of estimation
Trend to over estimation
3) Comparision of results
18

3 Results and Discussion — BJ Estimation .

Table 1: BJ Annual modeling results

N°	Station	N	Transf	VT	Model	VR (est)	VR (cal)	VE (%)	TEM	TEV
01	Ain berda	31	d=0	119,08	- -	--	--	--		
02	Béni bahdel	69	d=1	1304,74	ARIMA(0,1,2)	1220,21	1030 ,72	21,00	OK	OK
03	Bouchegouf	34	d=0	5016,59	- -	- -	--	- -		
04	Bouhnifia	59	d=0	4049,08	ARIMA(1,0,3)	3031,81	3015,51	25,00	OK	OK
05	Cheffia	27	d=0	8735,7	- -	- -	- -	- -		
06	Ksob	27	d=0	509,51	ARIMA(2,0,0)	487,31	307,92	39,50	OK	OK
07	Mefrouche	50	d=0	84,47	ARIMA(2,0,3)	71,67	73,28	13,24	NO	OK
08	Mirebeck	27	d=0	72451	- -	- -	- -	- -		
09	Pierre du chat	42	d=1	15644,7	- -	- -	- -	- -		
10	Remchi	43	$\sqrt{\ }$	15,42	ARIMA(1,0,3)	10,83	9,83	36,25	OK	OK
	Average			**1192,64**			**887,45**	**26,99**		

Notation : N: number of observations; **Transf**: transformation; **VT**: total variance; **VR(est)**: estimated residual variance; **VR(cal)**: calculated residual variance; **VE(%)**: explained variance in %; **TEM**: Average equality test ; **TEV**: variance equality test.

3 Results and Discussion — BJ Estimation .

Table 2: BJ Annual modeling errors

N	Station	Model	MSE		MAE		ME	
			estimation	validation	estimation	validation	estimation	validation
01	Béni bahdel	ARIMA(0,1,2)	1170,67	809,39	27,48	23,77	3,58	-21,52
02	Bouhnifia	ARIMA(1,0,3)	3118,66	4164,25	42,05	56,52	31,56	113,57
03	Ksob	ARIMA(2,0,0)	477,72	191,00	13,878	12,40	-0,97	-3,46
04	Mefrouche	ARIMA(2,0,3)	73,41	88,74	5,97	8,38	0,15	0,15
05	Remchi	ARIMA(1,0,3)	5211,81	2379,47	53,66	42,35	6,02	37,92
	Average		**2010,45**	**1526,57**	**28,61**	**28,68**	**8,07**	**25,33**

Notation : MSE : Mean square error ; **MAE** : mean absolute error **; ME** : mean error.

3 Results and Discussion — BJ Estimation .

Table 3: BJ monthly modeling results

N°	Station	N	Diff (d,D,S)	VT	Model	VR	VE(%)	TEM	TEV
01	Ain berda	372	(0,1,12)	10,00	SARIMA(1,0,0) x (0,1,1)12	4,17	58,3	OK	NO
02	Béni bahdel	828	(1,1,12)	103,73	SARIMA(1,1,1) x (2,1,1)12	39,37	62	OK	NO
03	Bouchegouf	408	(0,1,12)	339,98	SARIMA(1,0,0) x (0,1,1)12	158,28	53,4	NO	OK
04	Bouhnifia	708	(1,1,12)	443,50	SARIMA(1,1,1) x (0,1,1)12	128,29	71	OK	NO
05	Cheffia	324	(1,1,12)	849,90	SARIMA(1,1,1) x (0,1,1)12	291,35	65,7	OK	OK
06	Ksob	324	(0,1,12)	29,34	SARIMA(0,0,3) x (0,1,1)12	12,83	56,2	OK	OK
07	Mefrouche	600	(0,1,12)	5,06	SARIMA(1,0,0) x (0,1,1)12	2,80	44,6	OK	OK
08	Mirebeck	324	(0,1,12)	4802,36	SARIMA(2,0,0) x (0,1,1)12	2020,58	58	OK	NO
09	Pierre du chat	504	(0,1,12)	709,00	SARIMA(1,0,0) x (1,1,1)12	443,3	37,4	OK	NO
10	Remchi	516	(0,1,12)	351,96	SARIMA(0,0,1) x (0,1,1)12	203,67	42	OK	NO
	Moyenne			**764,48**		**330,48**	**54,86**		

Notation :
N: number of observations; **Diff (d,D,S):** differenciation of ordre (p,D,S); **VT**: total variance; **VR**:residual variance ; **VE(%)**: explained variance in %; **TEM**: Average equality test ; **TEV**: variance equality test.

3 Results and Discussion — BJ Estimation .

Table 4: BJ monthly modeling errors

N	Station	EQM		EAM		EM	
		estimation	validation	estimation	validation	estimation	validation
01	Ain berda	4,74	2,79	1,05	0,96	0,009	0,17
02	Béni bahdel	47,34	15,85	3,30	2,46	0,13	-0,39
03	Bouchegouf	177,65	103,25	7,16	6,00	-0,05	-0,97
04	Bouhnifia	155,90	64,03	5,57	4,13	1,20	-0,10
05	Cheffia	301,22	246,25	9,91	9,70	0,23	-1,33
06	Ksob	13,01	12,98	2,41	2,13	-0,31	0,83
07	Mefrouche	3,59	1,18	0,99	0,66	0,05	-0,23
08	Mirebeck	2428,46	1240,27	24,39	19,36	-0,55	6,00
09	Pierre du chat	615,17	200,86	12,18	9,54	2,46	-5,49
10	Remchi	314,07	39,60	7,71	4,28	1,39	-2,48
	Average	**406,11**	**192,70**	**7,46**	**5,92**	**0,45**	**-0,39**

Notation : MSE : Mean square error ; **MAE** : mean absolute error **; ME** : mean error.

3 Results and Discussion — KF Estimation .

Predictions in Time dimension (ANNUAL)

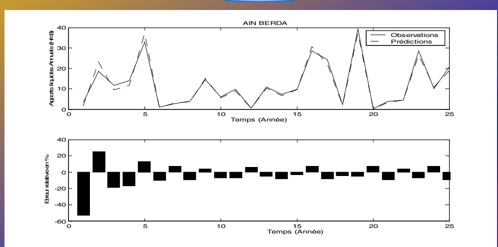


Fig.2 Annual Observations and KF obtained estimation at **AIN BERDA** (1968-1992)

3 Results and Discussion — KF Estimation .

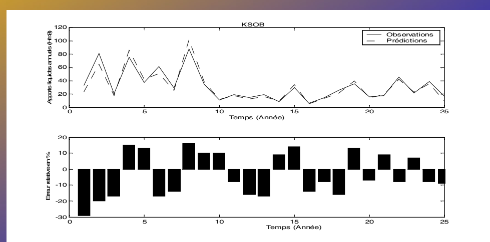


Fig.3 Annual Observations and KF obtained estimation at **KSOB** (1968-1992)

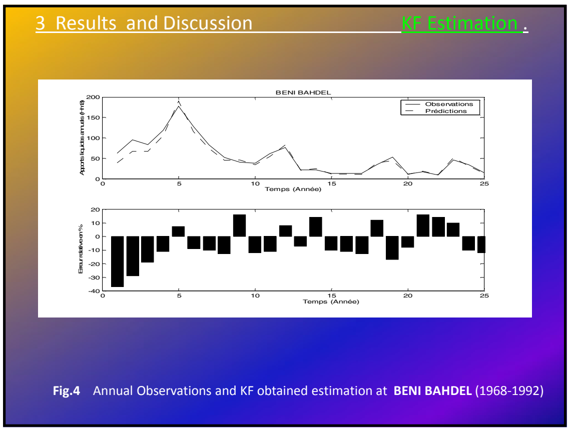


Fig.4 Annual Observations and KF obtained estimation at **BENI BAHDEL** (1968-1992)

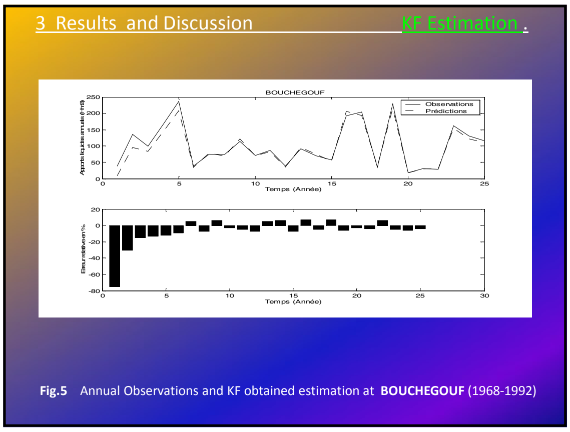


Fig.5 Annual Observations and KF obtained estimation at **BOUCHEGOUF** (1968-1992)

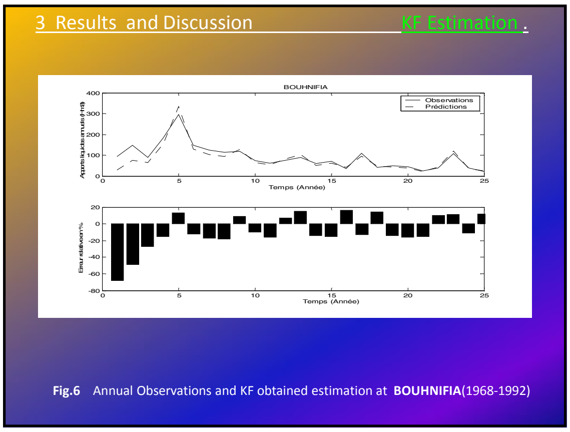


Fig.6 Annual Observations and KF obtained estimation at **BOUHNIFIA**(1968-1992)

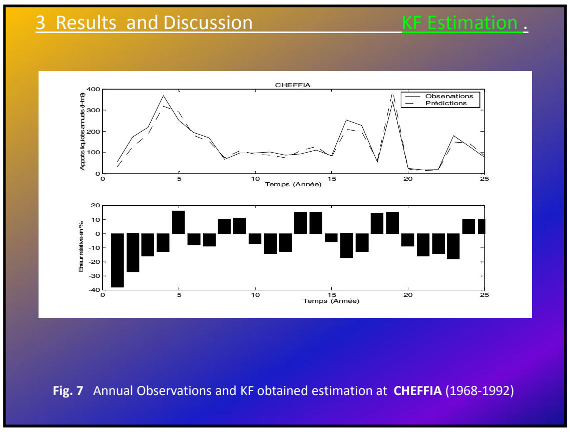


Fig. 7 Annual Observations and KF obtained estimation at **CHEFFIA** (1968-1992)

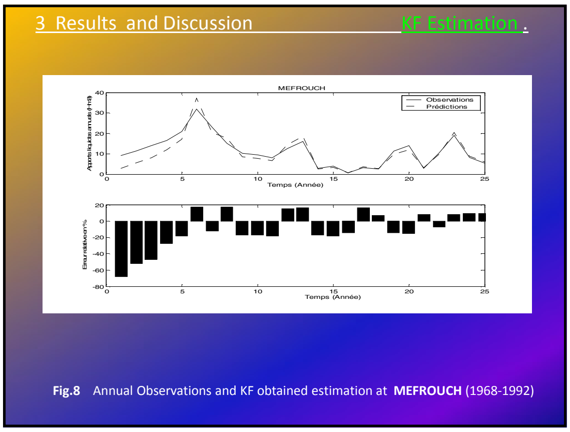


Fig.8 Annual Observations and KF obtained estimation at **MEFROUCH** (1968-1992)

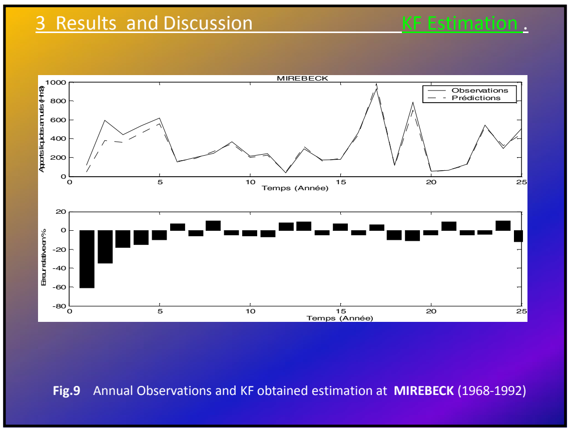


Fig.9 Annual Observations and KF obtained estimation at **MIREBECK** (1968-1992)

3 Results and Discussion — KF Estimation .

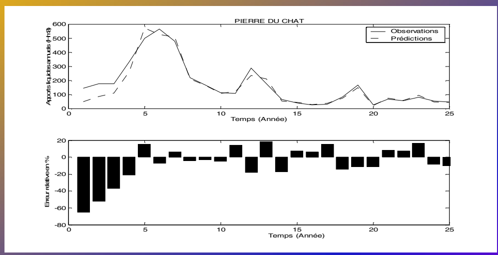


Fig.10 Annual Observations and KF obtained estimation at **PIERRE DU CHAT** (1968-1992)

3 Results and Discussion — KF Estimation .

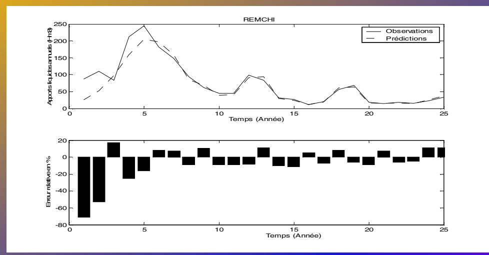


Fig.11 Annual Observations and KF obtained estimation at **REMCHI** (1968-1992)

3 Results and Discussion — KF Estimation .

Predictions in Space Dimension

Beginning of calculations (1968) — End of calculations (1992)

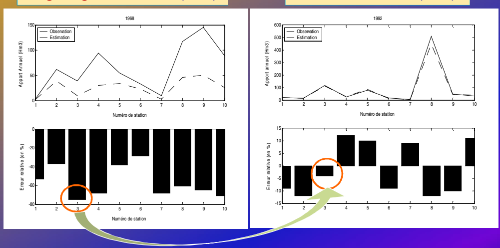


Fig.12 Calculations comparison in terms of relative error between 1968 and 1992.

3 Results and Discussion — KF Estimation .

Predictions in Time dimension (MONTHLY)

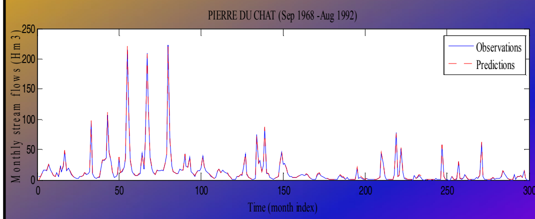


Fig.13 Monthly Observations and predictions at **PIERRE du CHAT** (Sep 1968 – Aug. 1992).

3 Results and Discussion — KF Estimation .

Predictions in Time dimension (MONTHLY)

Table 5: Mean and Standart Deviation of KF prediction error for monthly stream flows.

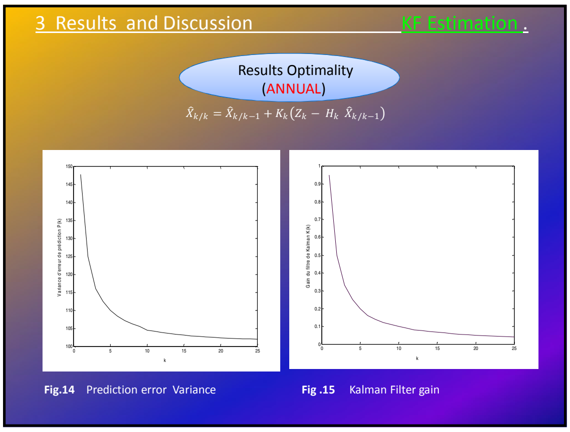
Predictions	Period	Mean	St.Dev
Temporal	Observation	101,34	75,95
	Estimation	100,11	70,03
	Err.Rel (%)	1.21	7.79
Spatial	Observation	101,34	111,26
	Estimation	93,53	68,63
	Err.Rel (%)	7.70	38.31

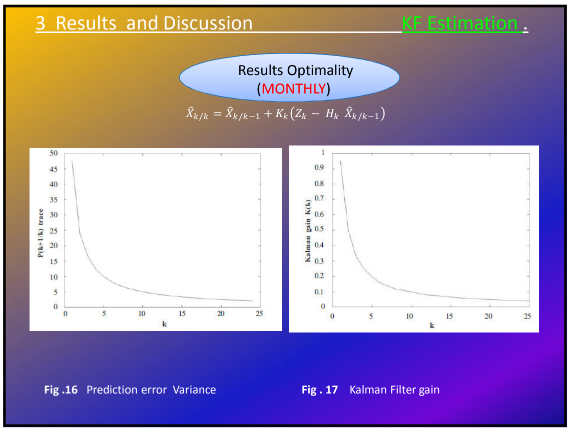
3 Results and Discussion — KF Estimation .

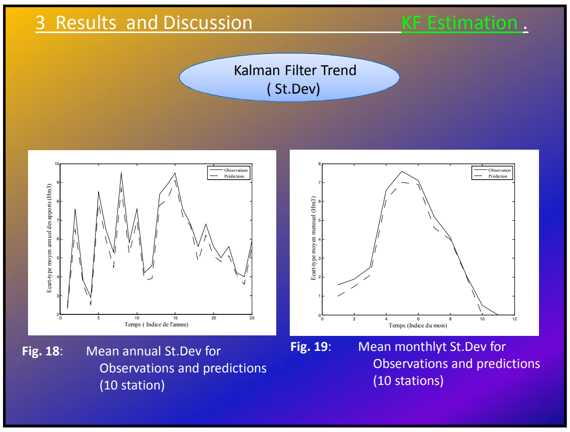
Results Optimality

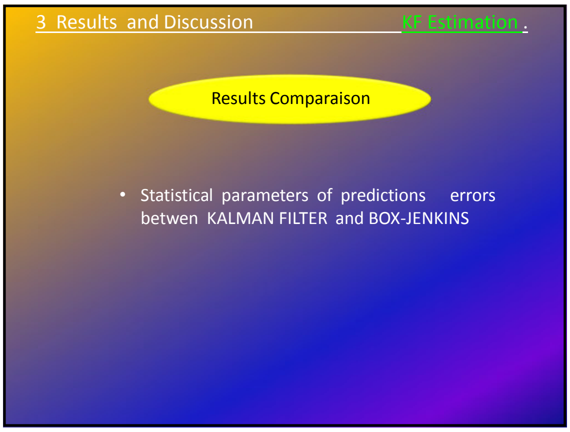
$$P_{k/k} = (I - K_k H_k)\, P_{k/k-1}$$

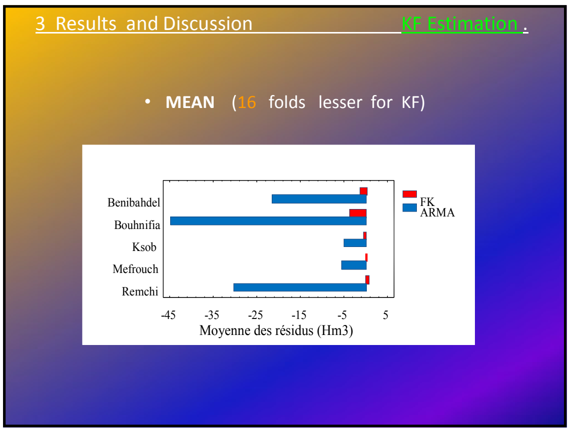
$$K_k = P_{k/k-1} H_k^{\,T} (H_k\, P_{k/k-1} H_k^{\,T} + R_k)^{-1}$$

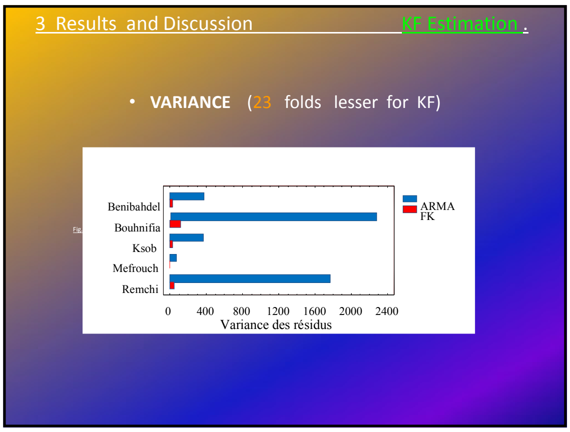

3 Results and Discussion
KF Estimation .
Results Optimality
(ANNUAL)
$\hat{X}_{k/k} = \hat{X}_{k/k-1} + K_k(Z_k - H_k\ \hat{X}_{k/k-1})$
Fig.14 Prediction error Variance
Fig .15 Kalman Filter gain

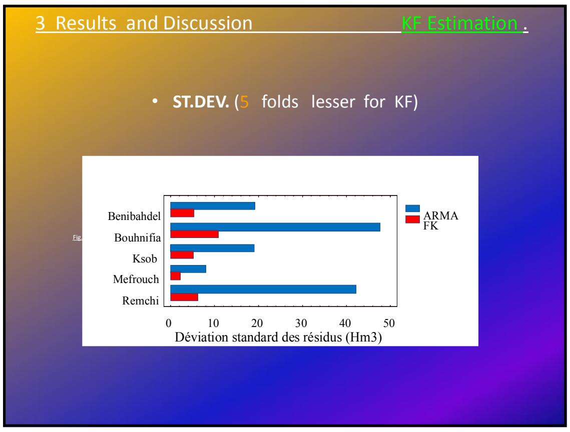

3 Results and Discussion
KF Estimation .
Results Optimality
(MONTHLY)
$\hat{X}_{k/k} = \hat{X}_{k/k-1} + K_k(Z_k - H_k\ \hat{X}_{k/k-1})$
P(k+1/k) trace
Kalman gain K(k)
k
Fig .16 Prediction error Variance
Fig . 17 Kalman Filter gain

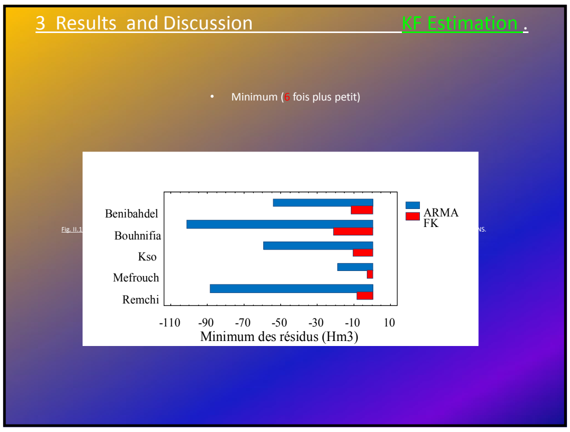

3 Results and Discussion
KF Estimation .
Kalman Filter Trend
(St.Dev)
Fig. 18: Mean annual St.Dev for Observations and predictions (10 station)
Fig. 19: Mean monthlyt St.Dev for Observations and predictions (10 stations)

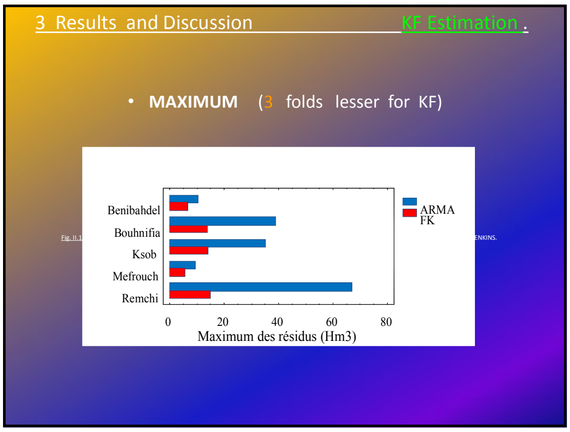

3 Results and Discussion
KF Estimation .
Results Comparaison
Statistical parameters of predictions errors betwen KALMAN FILTER and BOX-JENKINS

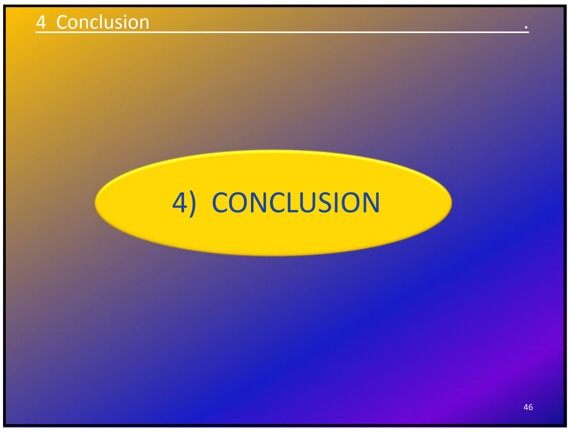

3 Results and Discussion
KF Estimation .
MEAN (16 folds lesser for KF)
Benibahdel
Bouhnifia
Ksob
Mefrouch
Remchi
FK
ARMA
-45 -35 -25 -15 -5 5
Moyenne des résidus (Hm3)

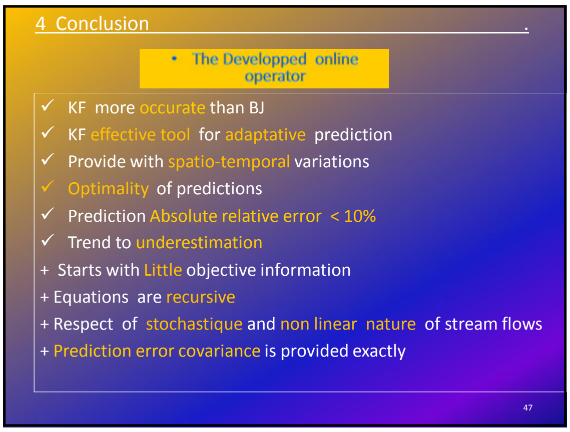

3 Results and Discussion
KF Estimation .
VARIANCE (23 folds lesser for KF)
Benibahdel
Bouhnifia
Ksob
Mefrouch
Remchi
ARMA
FK
0 400 800 1200 1600 2000 2400
Variance des résidus


3 Results and Discussion
KF Estimation .
ST.DEV. (5 folds lesser for KF)
Benibahdel
Bouhnifia
Ksob
Mefrouch
Remchi
ARMA
FK
0 10 20 30 40 50
Déviation standard des résidus (Hm3)


3 Results and Discussion
KF Estimation .
Benibahdel
Bouhnifia
Kso
Mefrouch
Remchi
ARMA
FK
-110 -90 -70 -50 -30 -10 10
Minimum des résidus (Hm3)


3 Results and Discussion
KF Estimation .
MAXIMUM (3 folds lesser for KF)
Benibahdel
Bouhnifia
Ksob
Mefrouch
Remchi
ARMA
FK
0 20 40 60 80
Maximum des résidus (Hm3)


4 Conclusion
4) CONCLUSION
46


4 Conclusion
The Developped online operator
KF more occurate than BJ
KF effective tool for adaptative prediction
Provide with spatio-temporal variations
Optimality of predictions
Prediction Absolute relative error < 10%
Trend to underestimation
+ Starts with Little objective information
+ Equations are recursive
+ Respect of stochastique and non linear nature of stream flows
+ Prediction error covariance is provided exactly
47


Thank you