An exchangeable construction to get parcimonious ensemble forecasts post-processing models

Marie Courbariaux, Pierre Barbillon and Éric Parent (AgroParisTech)

Operational mentoring: Joël Gailhard (EDF-DTG) and Luc Perreault (IREQ)

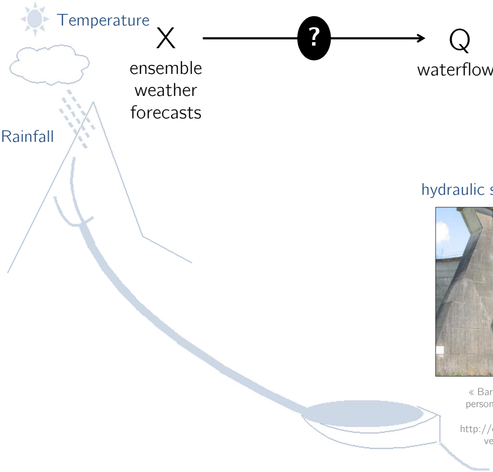
Academic mentoring: Éric Parent and Pierre Barbillon (AgroParisTech/INRA), Anne-Catherine Favre (ENSE3/GINP and LTHE)

HEPEX Workshop 2016

1. Meteorological forecasting for hydrology
2. An exchangeable construction for ensemble forecasts processing
3. Space-time consistency
4. Precipitation : a variable with a discrete component
5. Conclusion and perspectives

1. Meteorological forecasting for hydrology
2. An exchangeable construction for ensemble forecasts processing
3. Space-time consistency
4. Precipitation : a variable with a discrete component
5. Conclusion and perspectives

Goal: Probabilistic hydrological forecasts



X
ensemble weather forecasts

→ ? →

Q
waterflows

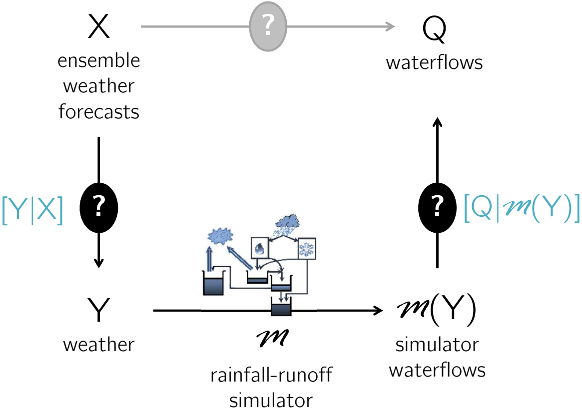
hydraulic structures management

« Barver » par Julien Bertrand — Travail personnel. Sous licence Domaine public via Wikimedia Commons - http://commons.wikimedia.org/wiki/File:Barver.JPG#/media/File:Barver.JPG

2 major uncertainty sources

[Krzysztofowicz (2002)]

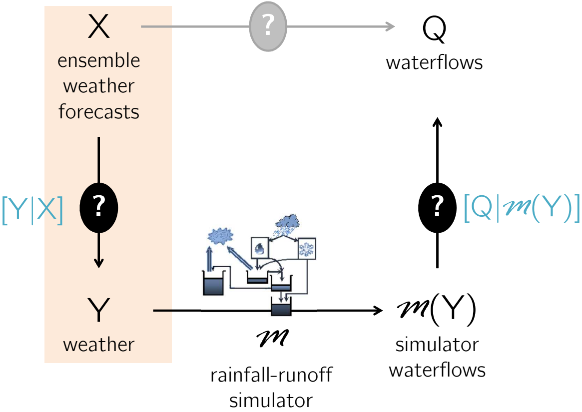
$$[Q|X] = \int [Q|\mathcal{M}(Y)][Y|X]\,dY$$



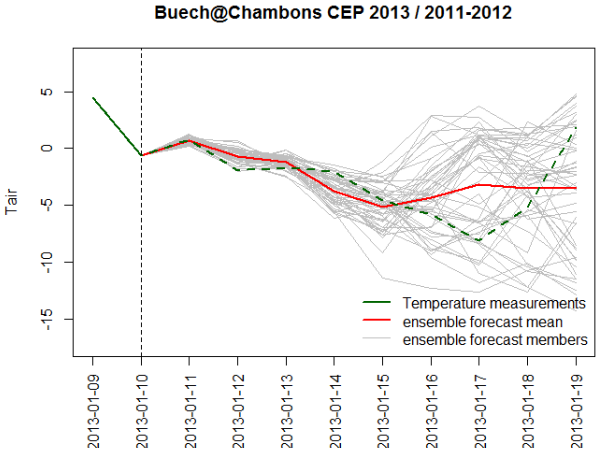
2 major uncertainty sources

[Krzysztofowicz (2002)]

$$[Q|X] = \int [Q|m(Y)][Y|X]\,dY$$



Ensemble weather forecasts



Ensemble weather forecasts

(animation.avi)

1 Meteorological forecasting for hydrology

2 An exchangeable construction for ensemble forecasts processing

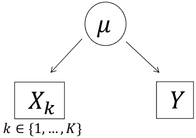
3 Space-time consistency

4 Precipitation : a variable with a discrete component

5 Conclusion and perspectives

An exchangeable construction for ensemble forecasts processing

Hierarchical model with 1 latent variable :

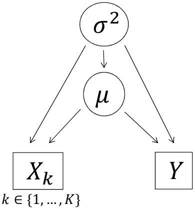


$$\begin{cases} (X_{k,t}|\mu_t) & = a + b\mu_t + c\varepsilon_{k,t} \\ (Y_t|\mu_t) & = a_0 + \mu_t + \varepsilon_{0,t} \\ \mu_t & \underset{t}{\overset{iid}{\sim}} \mathcal{N}\left(0, s^2\right) \\ \varepsilon_{k,t} & \underset{t,k}{\overset{iid}{\sim}} \mathcal{N}\left(0, \sigma^2\right) \end{cases}$$

...For each location, forecast lead time and meteorological variable

An exchangeable construction for ensemble forecasts processing

Hierarchical model with 2 latent variables :

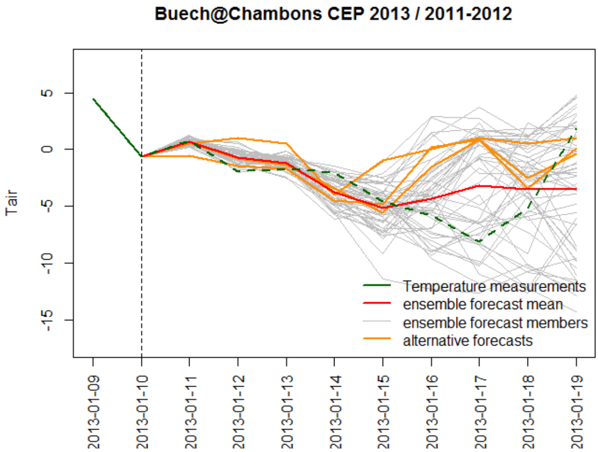


$$\begin{cases} (X_{k,t} | \mu_t) & = a + b\mu_t + c\varepsilon_{k,t} \\ (Y_t | \mu_t) & = a_0 + \mu_t + \varepsilon_{0,t} \\ (\varepsilon_{k,t} | \sigma_t^2) & \overset{iid}{\underset{k}{\sim}} \mathcal{N}\left(0, \sigma_t^2\right) \\ (\mu_t | \sigma_t^2) & \sim \mathcal{N}\left(0, \lambda\sigma_t^2\right) \\ \sigma_t^2 & \overset{iid}{\underset{t}{\sim}} \mathcal{I}\Gamma\left(\alpha, \beta\right) \end{cases}$$

...For each location, forecast lead time and meteorological variable

(With conjugation for inferential convenience)

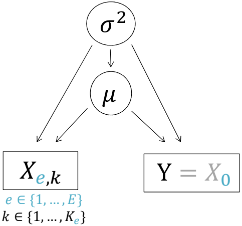
Other high resolution information sources (expertise ?)



→Gain in sharpness ?

An exchangeable structure to combine information sources

Multi-ensemble extension of the hierarchical model :



$$\begin{cases} (X_{e,k,t}|\mu_t) & = a_e + b_e\mu_t + c_e\varepsilon_{e,k,t} \\ (\varepsilon_{e,k,t}|\sigma_t^2) & \overset{iid}{\underset{e,k}{\sim}} \mathcal{N}(0, \sigma_t^2) \\ (\mu_t|\sigma_t^2) & \sim \mathcal{N}(0, \lambda\sigma_t^2) \\ \sigma_t^2 & \overset{iid}{\underset{t}{\sim}} \mathcal{I}\Gamma(\alpha, \beta) \end{cases}$$

Where $Y_t = X_{0,1,t}$ and $b_0 = c_0 = 1$ (identifiability)

Inference : classic EM algorithm

Deviance :

$$D(\alpha, \beta, \lambda, a, b, c) = -2\log\left[X, Y, \mu, \sigma^2 | \alpha, \beta, \lambda, a, b, c\right]$$
$$= \sum_{e=0}^{E} \left\{ \sum_{k=1}^{K_e} (X_{e,k} - b_e\mu - a_e)^2 \sigma^{-2} c_e^{-2} - K_e \log \sigma^{-2} - K_e \log c_e^{-2} \right\}$$
$$+ \mu^2 \lambda^{-1} \sigma^{-2} - \log(\lambda^{-1}) - \log \sigma^{-2} + 2\beta\sigma^{-2} - 2(\alpha - 1)\log \sigma^{-2} - 2\log \frac{\beta^\alpha}{\Gamma(\alpha)}$$

E step : deviance expectation with regard to $\left(\mu, \sigma^2 | X, Y\right)$(using conjugation properties)
M step :
$$(\alpha, \beta, \lambda, a, b, c)^{(i+1)} = \underset{(\alpha,\beta,\lambda,a,b,c)}{\text{argmin}} \ \mathbb{E}_{\left(\mu,\sigma^{-2}|X,Y,(\alpha,\beta,\lambda,a,b,c)^{(i)}\right)} \left\{\sum_t D_t(\alpha, \beta, \lambda, a, b, c)\right\}$$

Inference results

$(X_{e,k,t}|\mu_t) = a_e + b_e\mu_t + c_e\varepsilon_{e,k,t}$
Where $Y_t = X_{0,1,t}$ et $b_0 = c_0 = 1$ (identifiability)

For a 3 days forecasting horizon :

	a	b	c
Y ($= X_0$)	8.75	1.00	1.00
ECMWF ensemble ($= X_1$)	8.68	1.10	0.87
high resolution ECMWF ($= X_2$)	8.40	1.09	0.79

e ensemble member contribution : $\frac{b_e}{c_e^2}$

Forecast

Then :

$(Y_t|X_t) \sim$Student

with

$$\mathbb{E}\left(Y_t|X_t\right) = a_0 + \lambda'' \sum_{e=1}^{E} \frac{b_e}{c_e^2} K_e \left(\bar{X}_e - a_e\right)$$

and

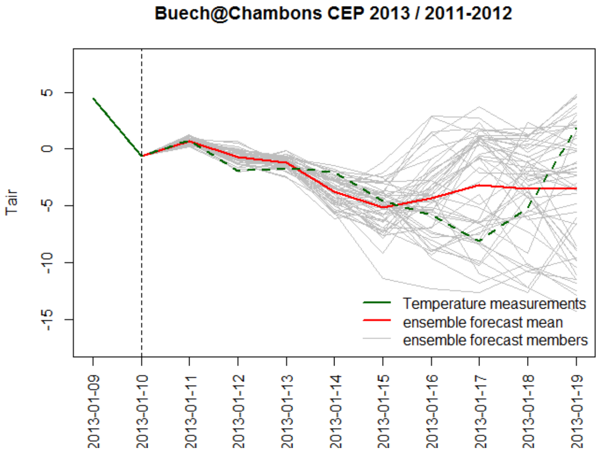
$$\mathbb{V}\left(Y_t|X_t\right) \propto$$

$$\beta + \frac{1}{2}\left(\sum_{e=1}^{E} c_e^{-2} \sum_{k=1}^{K_e} (X_{e,k} - a_e)^2 - \lambda'' \left(\sum_{e=1}^{E} c_e^{-2} b_e K_e \left(\bar{X}_e - a_e\right)\right)^2\right)$$

GEMOS : Generalized EMOS

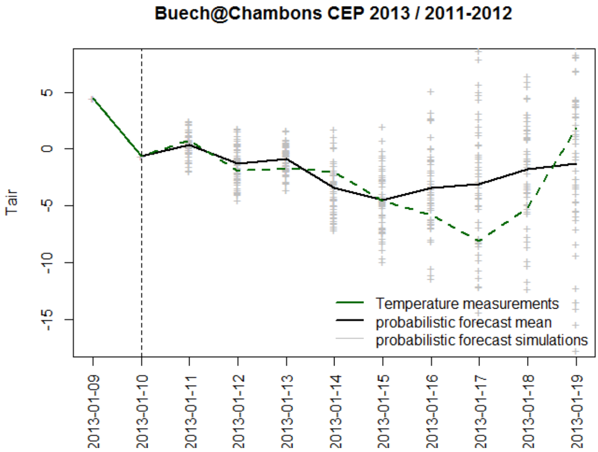
Raw forecast

(50 members)

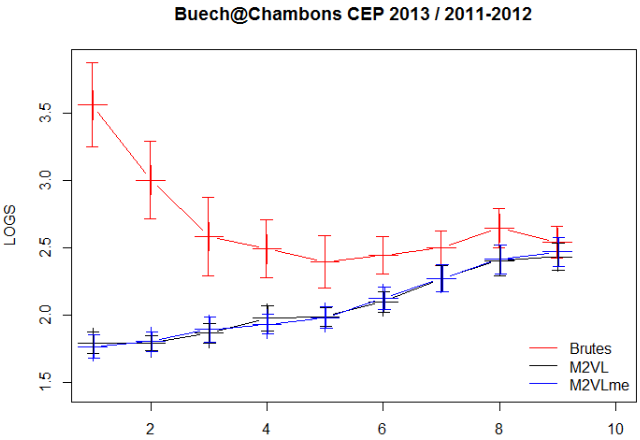


GEMOS forecast with high resolution forecasts

(50 simulations)



Scores


Buech@Chambons CEP 2013 / 2011-2012
LOGS
1.5
2.0
2.5
3.0
3.5
2
4
6
8
10
Brutes
M2VL
M2VLme

1 Meteorological forecasting for hydrology

2 An exchangeable construction for ensemble forecasts processing

3 Space-time consistency

4 Precipitation : a variable with a discrete component

5 Conclusion and perspectives

Space-time consistency

Solution 1 : empirical copulas

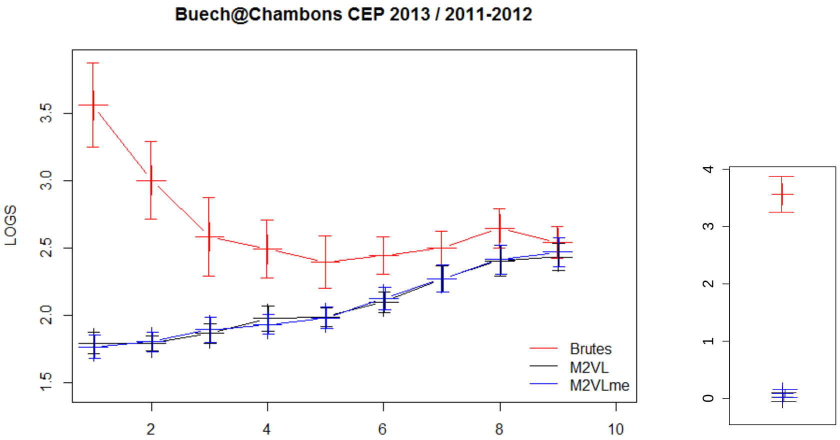
Idea : copy a rank structure (...of past scenarios : Schaake shuffle [Clark et al. (2004)])

Advantage : simultaneously copy inter-variables, space and time structures

GEMOS-Schaake results

(animation2.avi)

Scores


Buech@Chambons CEP 2013 / 2011-2012
LOGS
1.5
2.0
2.5
3.0
3.5
2
4
6
8
10
Brutes
M2VL
M2VLme
0
1
2
3
4

1 Meteorological forecasting for hydrology

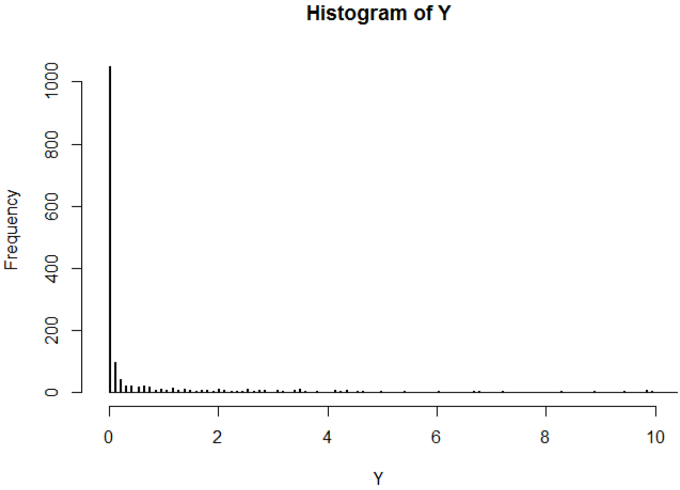
2 An exchangeable construction for ensemble forecasts processing

3 Space-time consistency

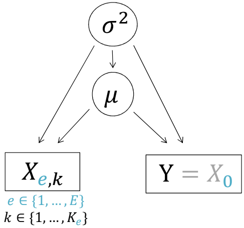
4 Precipitation : a variable with a discrete component

5 Conclusion and perspectives

Precipitation : a variable with a discrete component



GEMOS for mixed variables



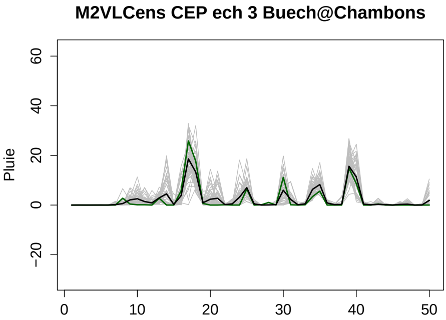
$$\begin{cases} (X_{e,k,t}|\mu_t) & = a_e + b_e \mu_t + c_e \varepsilon_{e,k,t} \\ (X'^{\nu}_{e,k,t}|X) & = \max(0, X_{e,k,t}) \\ (\varepsilon_{e,k,t}|\sigma_t^2) & \underset{e,k}{\overset{iid}{\sim}} \mathcal{N}(0, \sigma_t^2) \\ (\mu_t|\sigma_t^2) & \sim \mathcal{N}(0, \lambda\sigma_t^2) \\ \sigma_t^2 & \underset{t}{\overset{iid}{\sim}} \mathcal{I}\Gamma(\alpha, \beta) \end{cases}$$

Where $Y_t = X_{0,1,t}$ and $b_0 = c_0 = 1$ (identifiability)

$\nu = 1$ as a first try

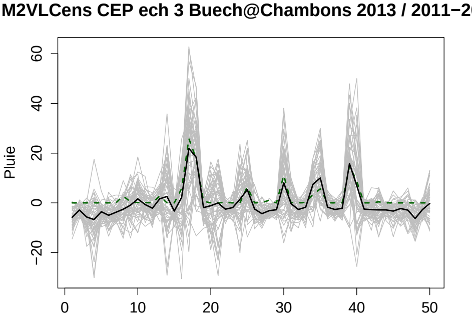
Raw forecasts

(50 simulations, 3-days lead time)



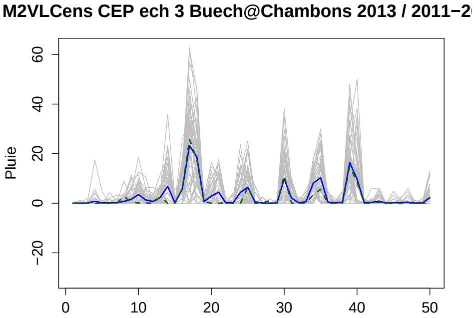
GEMOS forecasts before censoring

(50 simulations, 3-days lead time)

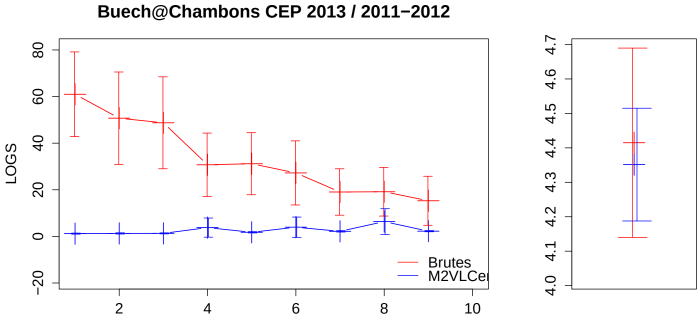


GEMOS forecasts after censoring

(50 simulations, 3-days lead time)



Scores


Buech@Chambons CEP 2013 / 2011−2012
LOGS
80
60
40
20
0
−20
2
4
6
8
10
Brutes
M2VLCe
4.7
4.6
4.5
4.4
4.3
4.2
4.1
4.0

1 Meteorological forecasting for hydrology

2 An exchangeable construction for ensemble forecasts processing

3 Space-time consistency

4 Precipitation : a variable with a discrete component

5 Conclusion and perspectives

Exchangeability virtues

- Exchangeability : an expected property
- If only exchangeable per blocks of members : get back to the multi-ensemble case
- Parsimonious models $\Rightarrow$ shorter training sets are usable (necessary in a changing world)
- Explicit EM algorithm within the Normal framework
- EMOS as a limiting case (in the standard framework)
- Some straightforward extensions : several ensembles, mixed-type variables (precipitations), normal-based multivariate distributions

Perspectives

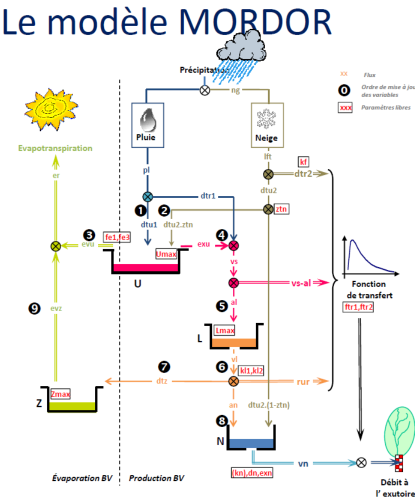
- Precipitation forecasts : work still in progress
- Space-time consistency : use of parametric copulas
- What about multi-modal distributions ?
- To be directly applied to (log-transformed) hydrological ensemble forecasts ?

Thanks for your attention !

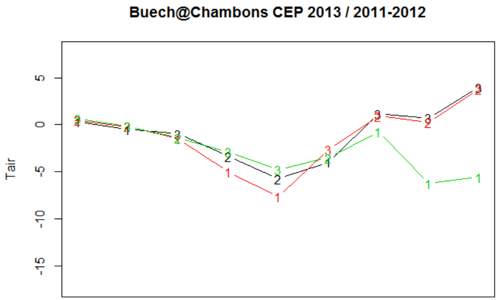
Clark, M., Gangopadhyay, S., Hay, L., Rajagopalan, B., and Wilby, R. (2004). The Schaake shuffle : A method for reconstructing space-time variability in forecasted precipitation and temperature fields. *Journal of Hydrometeorology*, 5(1) :243–262.

Krzysztofowicz, R. (2002). Bayesian system for probabilistic river stage forecasting. *Journal of Hydrology*, 268(1) :16–40.

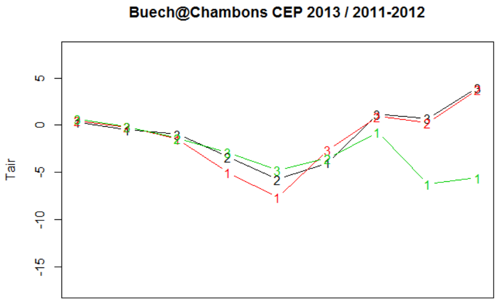
EDF's rainfall-runoff model

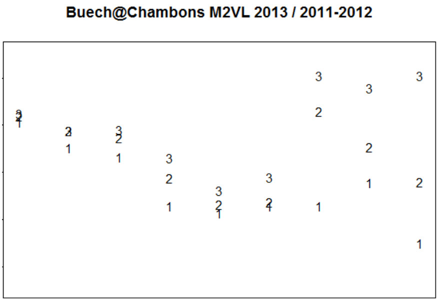


Empirical copula

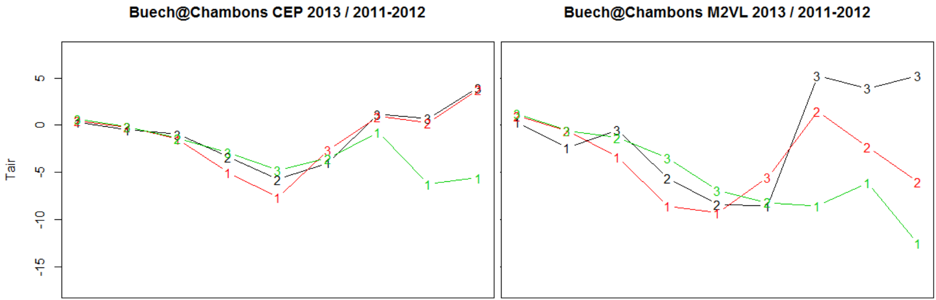


Empirical copula

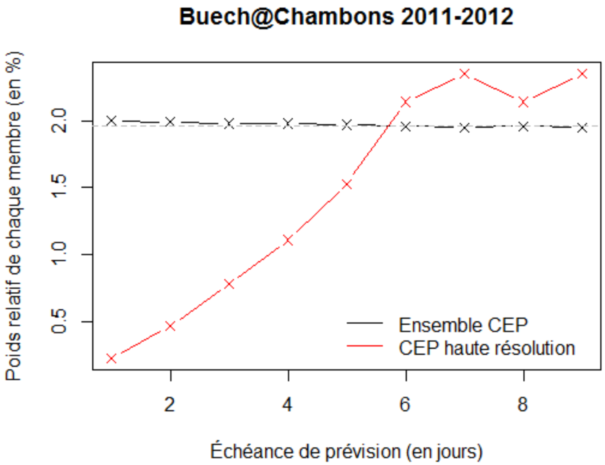




Empirical copula



Information mainly comes from the ECMWF ensemble



Rank histogram of cumulative air temperatures