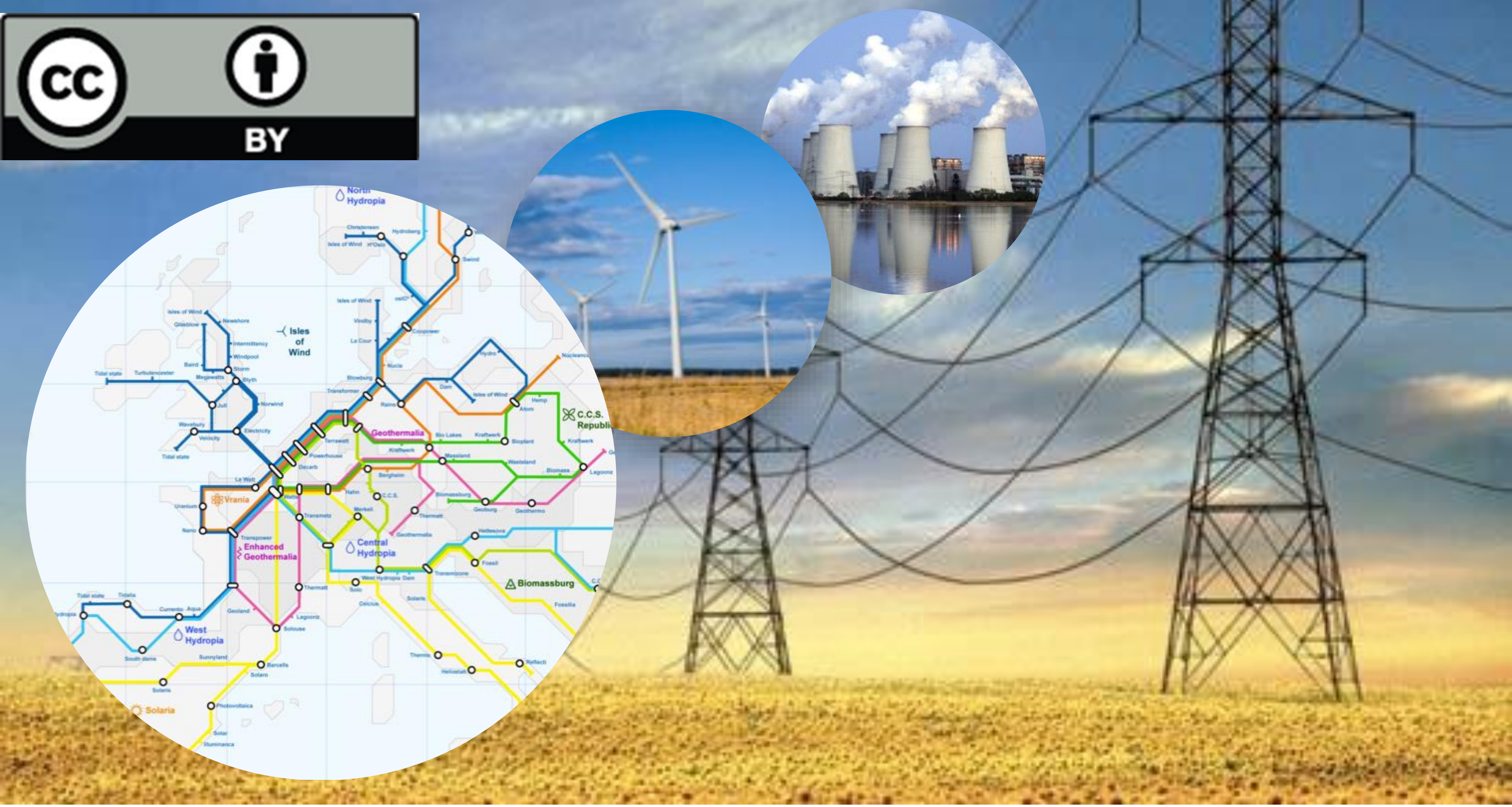
CONSISTENT BIAS CORRECTION / DOWNSCALING OF DIFFERENT CLIMATE VARIABLES

Léonard Dekens[1], Mathilde Grandjacques[2], Sylvie Parey[1], Didier Dacunha-Castelle[2]
1 EDF R&D, Chatou, France
2 Orsay University, PARIS 11

EMS 2016, Trieste

EUROPEAN CLIMATIC ENERGY MIXES

A Copernicus Sectoral Information System



UEA University of East Anglia
EDF
University of Reading
Met Office
MINES ParisTech
ARMINES
ENEA

PROPOSED METHODOLOGY

- **Generalization of quantile mapping approaches**
  - Distribution correction
  - Multivariate distributions

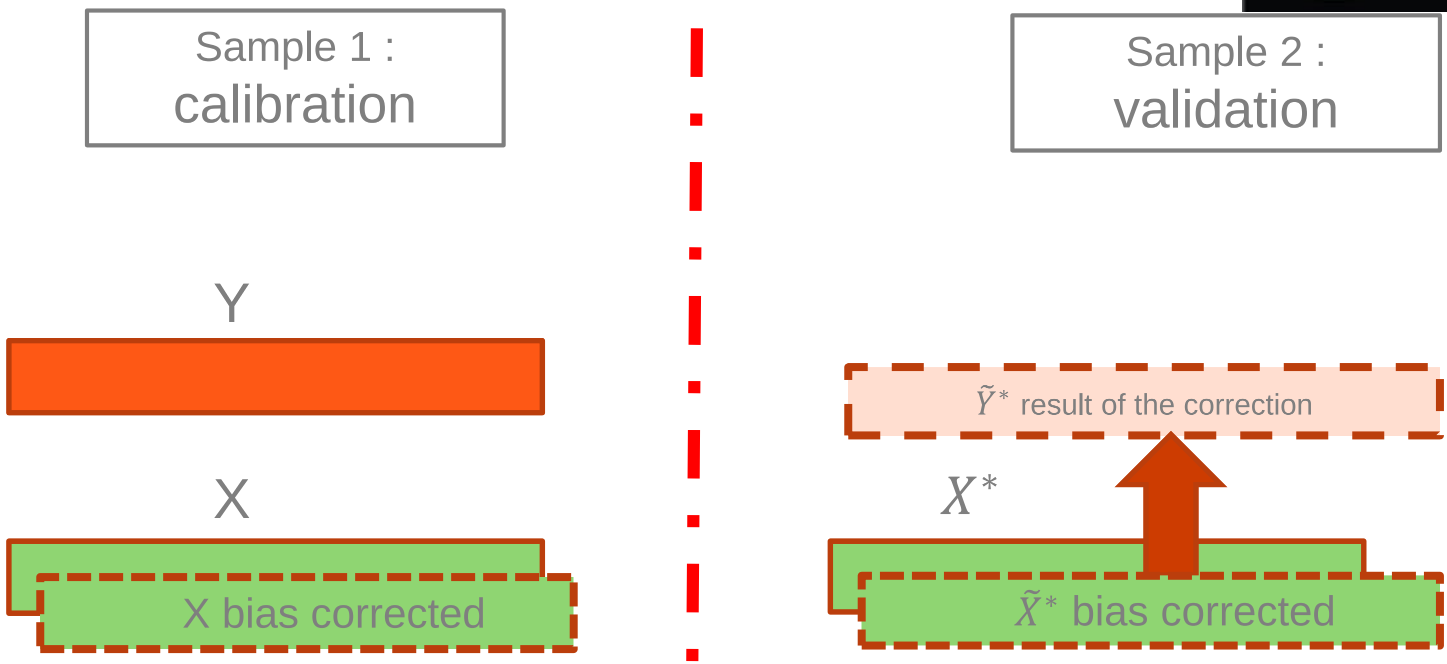
- **Principle**
  - $X = (X_1,X_2)$ large scale data $Y = (Y_1,Y_2)$ local observations
  - Correction => X distribution ≡ Y distribution

- **Mathematical background**
  - Rosenblatt theorem
    - $Z=T_X(X)$

$$\begin{cases} z^{(1)} = G_{X^{(1)}}(x^{(1)}) \\ z^{(2)} = G_{X^{(2)}|X^{(1)}=x^{(1)}}(x^{(2)}) \\ \dots \\ z^{(p)} = G_{X^{(p)}|X^{(1)}=x^{(1)},\dots,X^{(p-1)}=x^{(p-1)}}(x^{(p)}) \end{cases}$$

    - Z uniform and independent marginal distributions
  - Consequence: $\exists\ T_{X,Y}$ allowing to match X distribution with Y distribution

ILLUSTRATION



From sample 1: estimation of CDFs
$F_{Y_1}, G_{X_1}, F_{Y_2|Y_1=y_1}, G_{X_2|X_1=x_1}$

From sample 2 : bias correction of X* using CDFs estimated from sample 1 :
$F_{Y_1}, G_{X_1}, F_{Y_2|Y_1=y_1}, G_{X_2|X_1=x_1}$

1) **estimate** $\widetilde{X^{1,*}}$
2) **estimate** $\widetilde{X^{2,*}}$ conditioned by $\widetilde{X^{1,*}}$

DIFFERENT TESTS

- **Notations**
  - X large scale, Y local obs., ^ estimated, ~ bias corrected/downscaled, * validation period
- **Univariate bias correction**
  - Quantile matching $\widetilde{x}_t^* = \widehat{F}_Y^{-1}(\widehat{G}_X(x_t^*))$ (Hyp: X distribution ≈ X* distribution)
  - CDFt: $T(G_X(z)) = F_Y(z)$ and $T(G_{X^*}) = F_{Y^*}$ so $T(u)=F_Y(G_X^{-1}(u))$ and $F_{Y^*} = F_Y \circ G_X^{-1} \circ G_{X^*}$

$$\widetilde{x^*}_t = \hat{F}_{Y^*}^{-1}(\hat{G}_{X^*}(x_t^*)) = \hat{G}_{X^*}^{-1}(\hat{G}_X(\hat{F}_Y^{-1}(\hat{G}_{X^*}(x_t^*))))$$

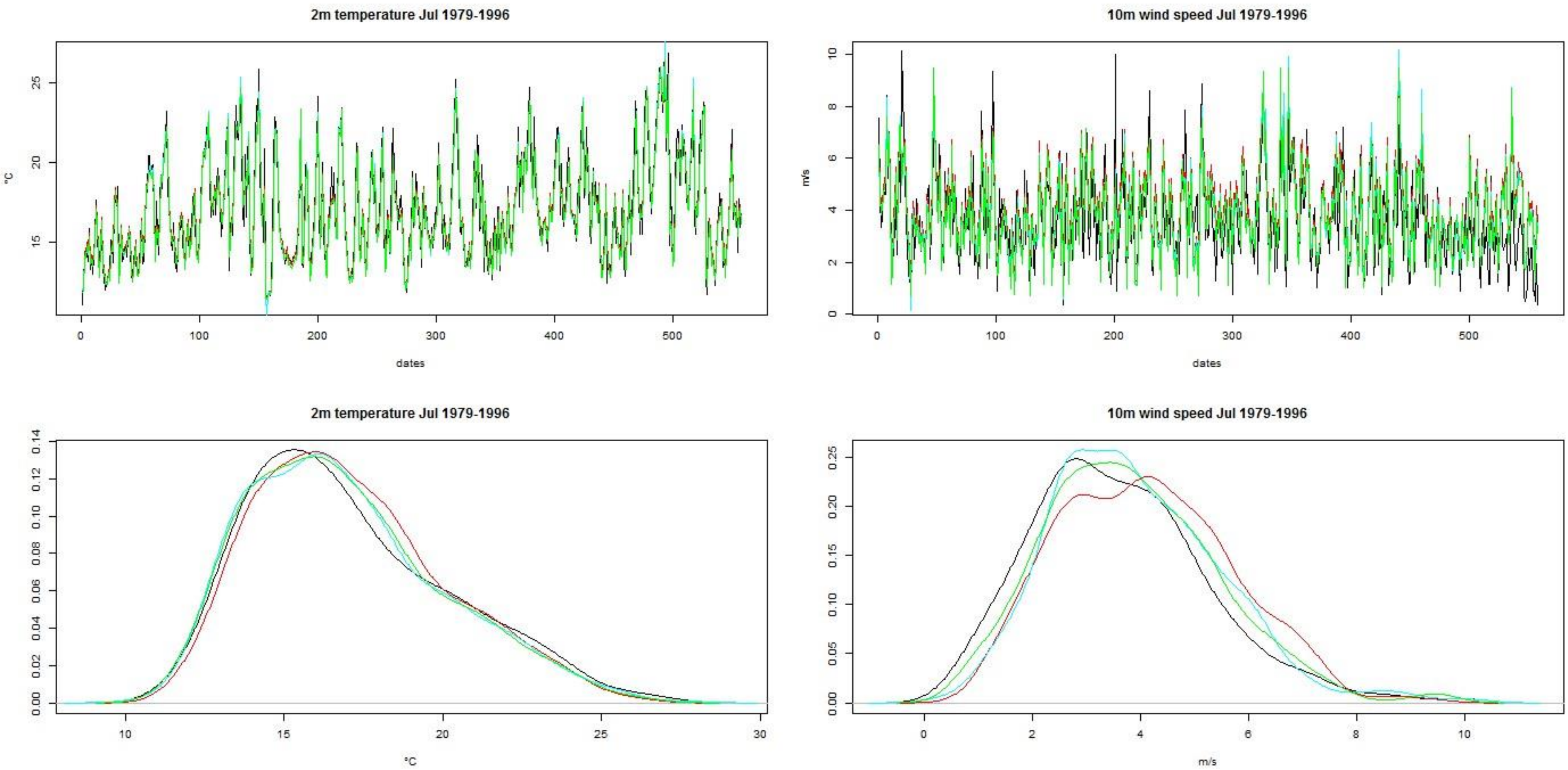
(Hyp: X and X* distribution not necessarily ≈; $T_{Y,Y^*} \approx T_{X,X^*}$ => X and Y similar distributions)

- **Bivariate correction**
  - Classes on 1st variable
    - Optimal class length?
    - Generalization to more than 2 variables?
  - Kernel density estimation: ∃ R codes allowing the selection of optimal parameter
- **Applied month by month**

DATA

- **Variables**
  - Daily means
  - 10m wind speed and 2m temperature
- **Observations**
  - Temperature: E-OBS
  - Wind speed: HadISD database
- **Large scale**
  - ERA-Interim 0.5°
- **Period**
  - Total: 1979-2014
  - Calibration: 1997-2014 Validation: 1979-1996
- **2 locations:** Hamburg, Paris-Orly

TEMPERATURE FIRST



observations

ERA-Interim

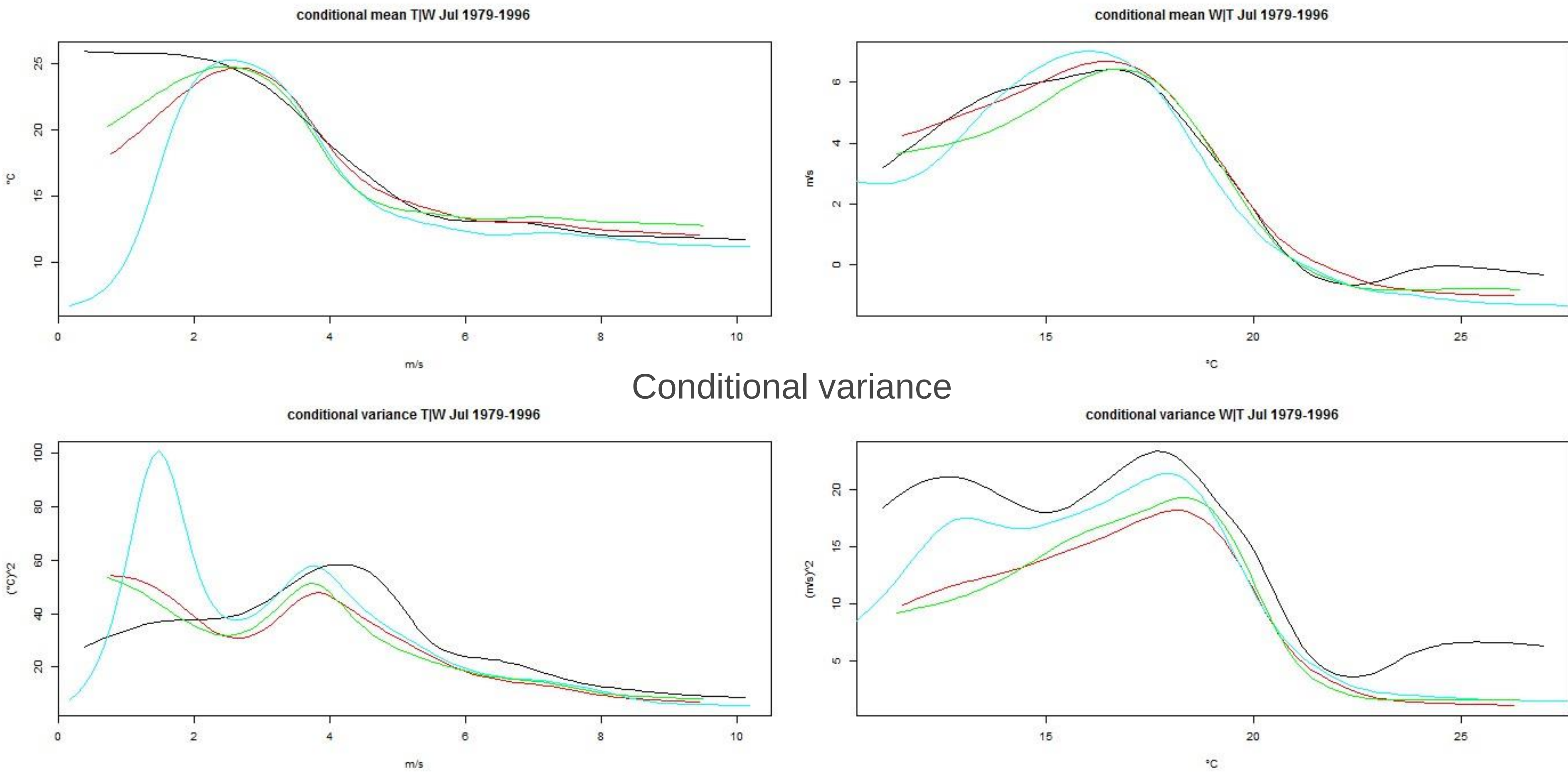
Univariate BC

Bivariate BC

TEMPERATURE FIRST

T | W Conditional mean W | T

Conditional variance



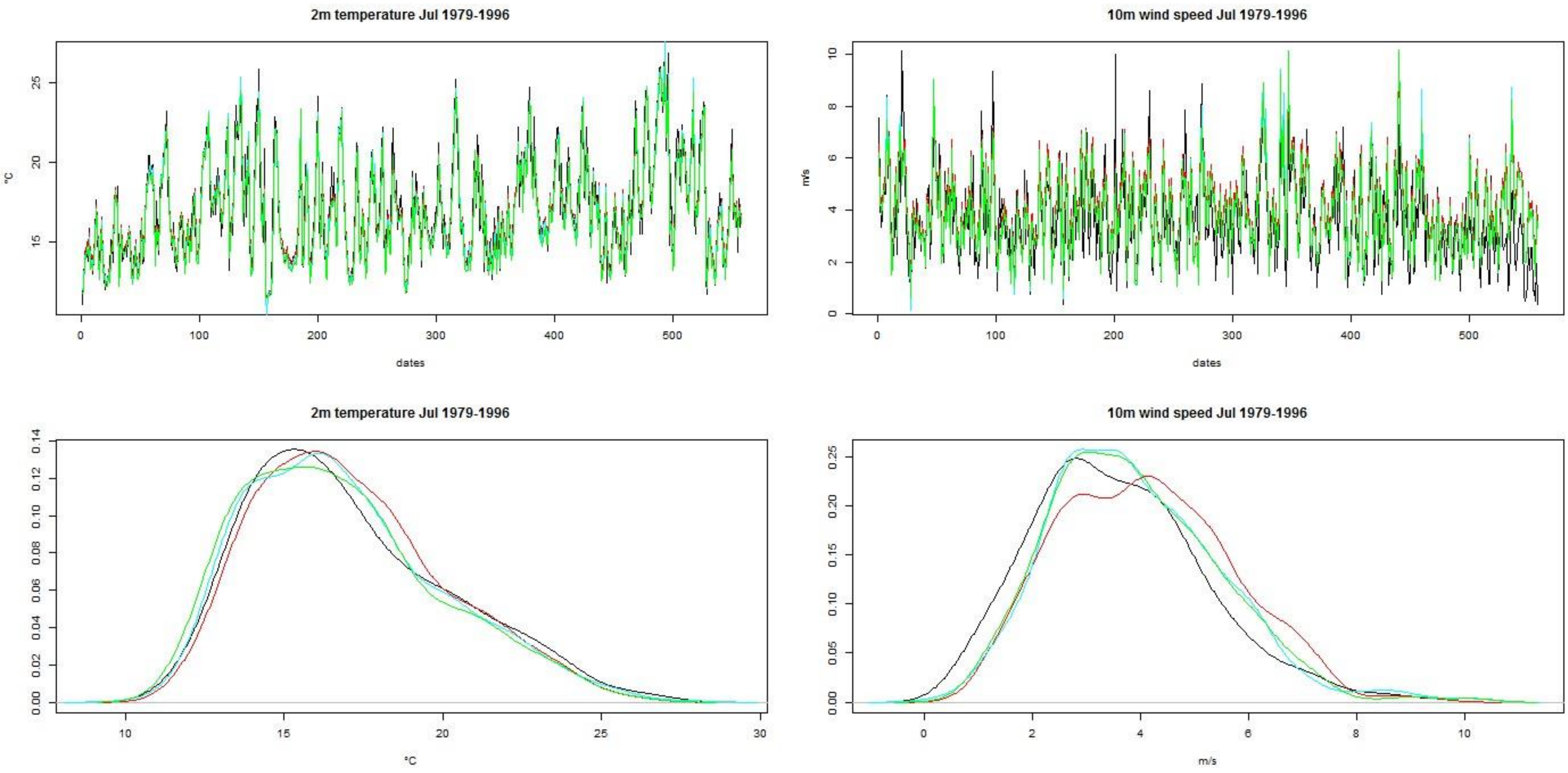
observations

ERA-Interim

Univariate BC

Bivariate BC

WIND FIRST



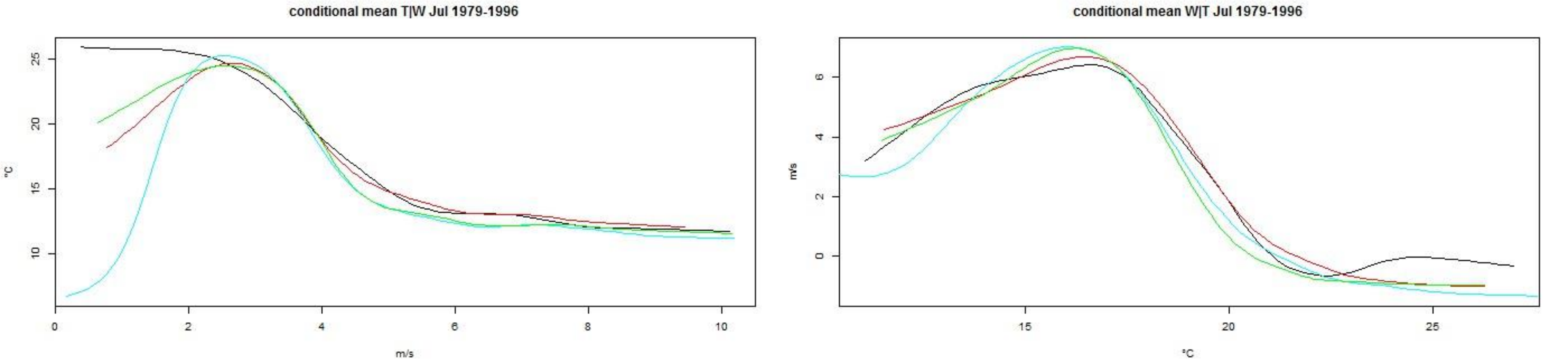
observations
ERA-Interim
Univariate BC
Bivariate BC

WIND FIRST

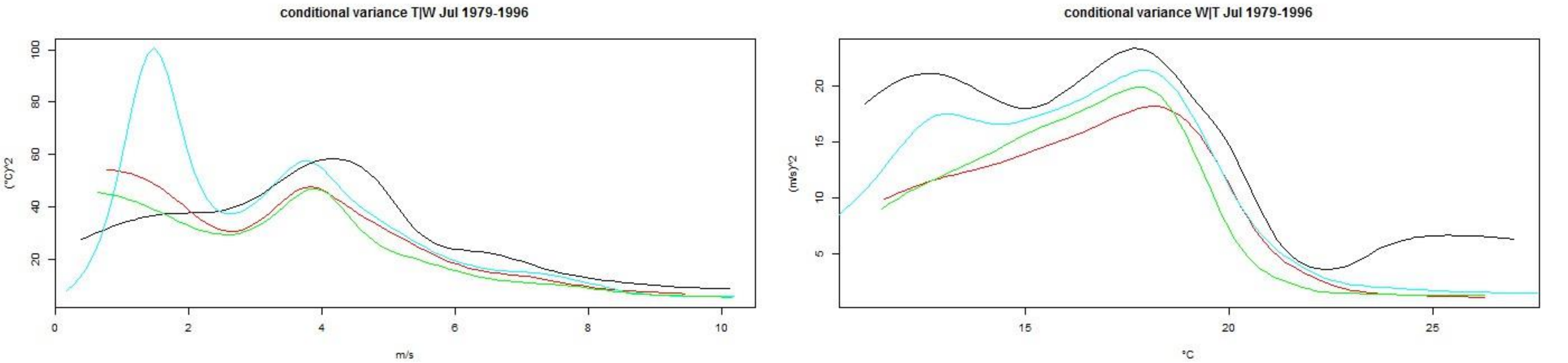
Conditional mean

T | W

W | T



Conditional variance



observations

ERA-Interim

Univariate BC

Bivariate BC

BIVARIATE DISTRIBUTIONS: CDFt

- **L1 norm between the bivariate distributions**

month	Without correction	Univariate correction	Bivariate correction	
			Temperature 1	Wind 1
January	0.0154	0.0122	0.0120	0.0254
February	0.0143	0.0171	0.0171	0.0201
March	0.0128	0.0125	0.0127	0.0102
April	0.0199	0.0166	0.0179	0.0163
May	0.0149	0.0156	0.0145	0.0165
June	0.0239	0.0111	0.0098	0.0126
July	0.0370	0.0220	0.0162	0.0230
August	0.0400	0.0180	0.0161	0.0137
September	0.0472	0.0209	0.0193	0.0176
October	0.0433	0.0164	0.0161	0.0161
November	0.0260	0.0133	0.0113	0.0212
December	0.0192	0.0142	0.0131	0.0316

BIVARIATE DISTRIBUTIONS: QUANTILE MATCHING

- **L1 norm between the bivariate distributions**

month	Without correction	Univariate correction	Bivariate correction	
			Temperature 1	Wind 1
January	0.0155	0.0122	0.0101	0.0127
February	0.0144	0.0181	0.0131	0.0178
March	0.0130	0.0111	0.0108	0.0110
April	0.0206	0.0212	0.0197	0.0186
May	0.0150	0.0174	0.0165	0.0167
June	0.0231	0.0101	0.0129	0.0111
July	0.0385	0.0212	0.0196	0.0225
August	0.0395	0.0222	0.0203	0.0173
September	0.0487	0.0188	0.0184	0.0159
October	0.0433	0.0191	0.0151	0.0144
November	0.0252	0.0133	0.0102	0.0111
December	0.0190	0.0165	0.0147	0.0210

DISCUSSION

- **Bivariate bias correction**
  - Generally better than univariate correction
  - In this case, quantile matching better than CDFt
  - Temperature first seems better

- **Future work**
  - Is it worth doing it?
  - Design a toy model allowing the control of
    - Variable correlation
    - Differences between large scale representation and local observations
  - Adapt Kernel density to extremes behavior?

- **Extension to > 2 variables**
  - Possible, optimal Kernel density parameter identification not necessarily available

Thank you