



HAL
open science

Infering meteorological information at different scales from several sources of data

Didier Josselin, Matthieu Vignal, Nicolas Viaux, Delphine Blanke, Céline
Lacaux

► **To cite this version:**

Didier Josselin, Matthieu Vignal, Nicolas Viaux, Delphine Blanke, Céline Lacaux. Infering meteorological information at different scales from several sources of data. ECTQG'2023, Université de Braga (Portugal), Sep 2023, Braga, Portugal. 2 p. hal-04351525

HAL Id: hal-04351525

<https://hal.science/hal-04351525v1>

Submitted on 18 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ECTQ2023: Braga (Portugal) 14 -17 September 2023

UNIVERSIDADE D
COIMBRA

Infering meteorological information at different scales from several sources of data

Didier Josselin*, Matthieu Vignal*, Nicolas Viaux*, Delphine Blanke**, Céline Lacaux**



* *UMR ESPACE, CNRS, France*
** *Laboratoire de Mathématique d'Avignon, France*

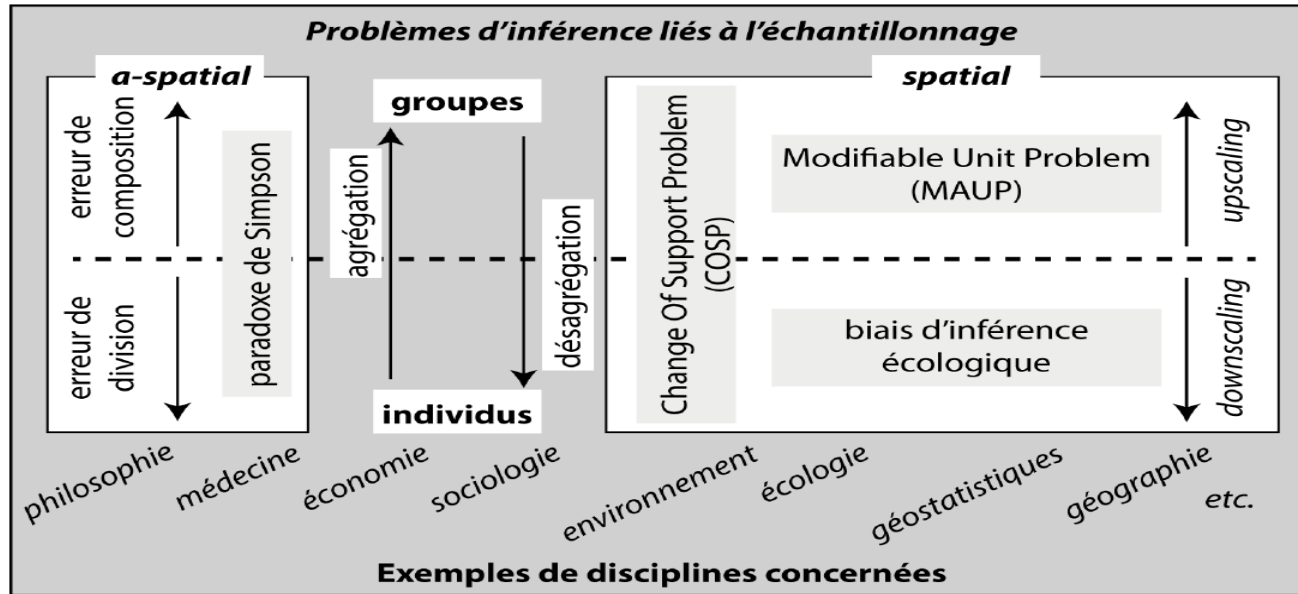


Outline

- Effect of the support on spatial statistics
- Data aggregation through scales
- Eliminating the spatial support effect by resampling and *Relative Scalar Deviation* calculation
- Conclusion

Effect of the spatial support

Spatial (dis)aggregation



[Openshaw, 1974]

[Yule, 1911
Theil, 1972]

[King et al., 2004,
Josselin et al., 2004]

[Robinson, 1950,
Goodman, 1953,
King, 1997]

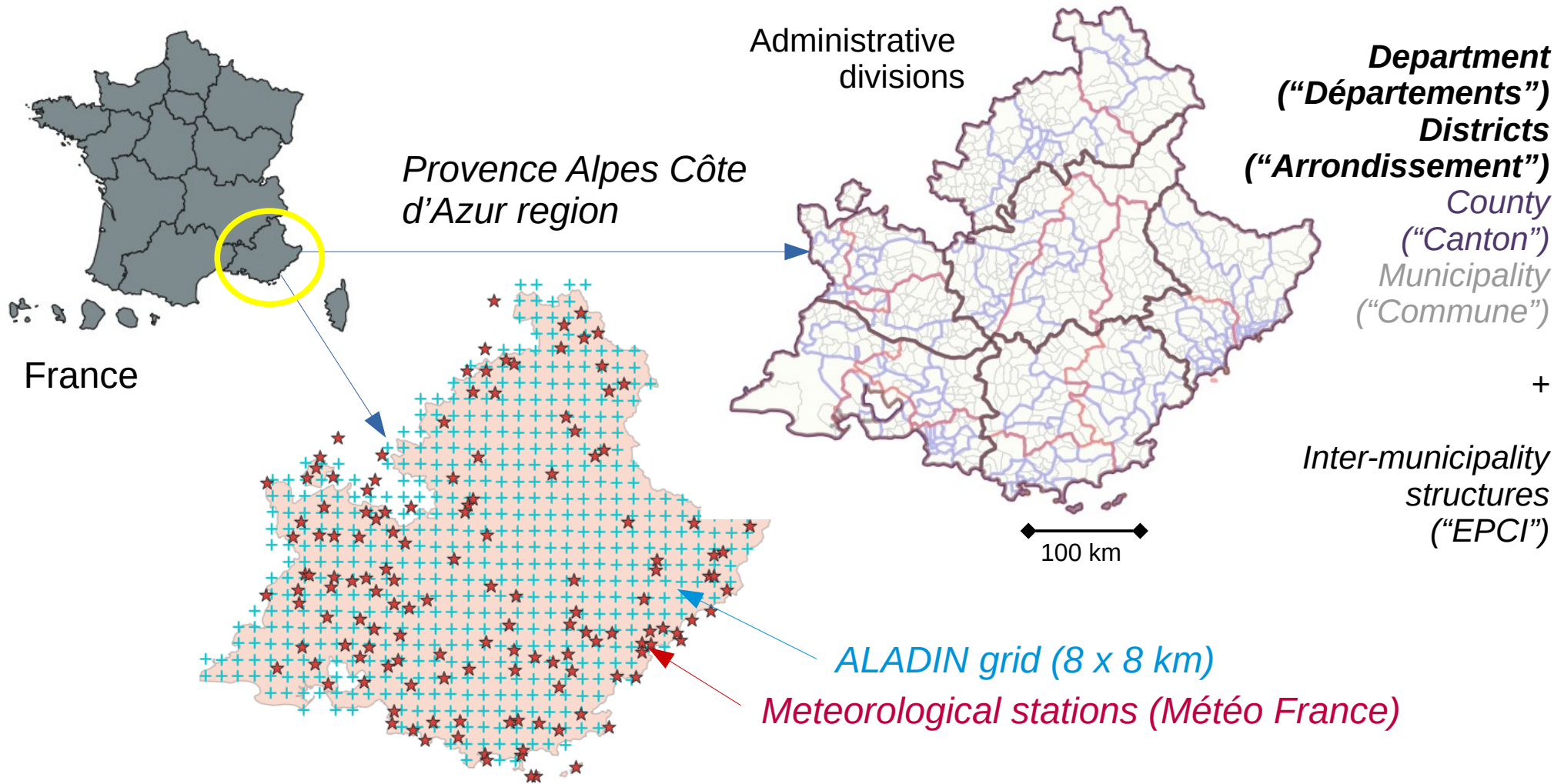
Josselin & Louvet, 2016

Objectives of the research

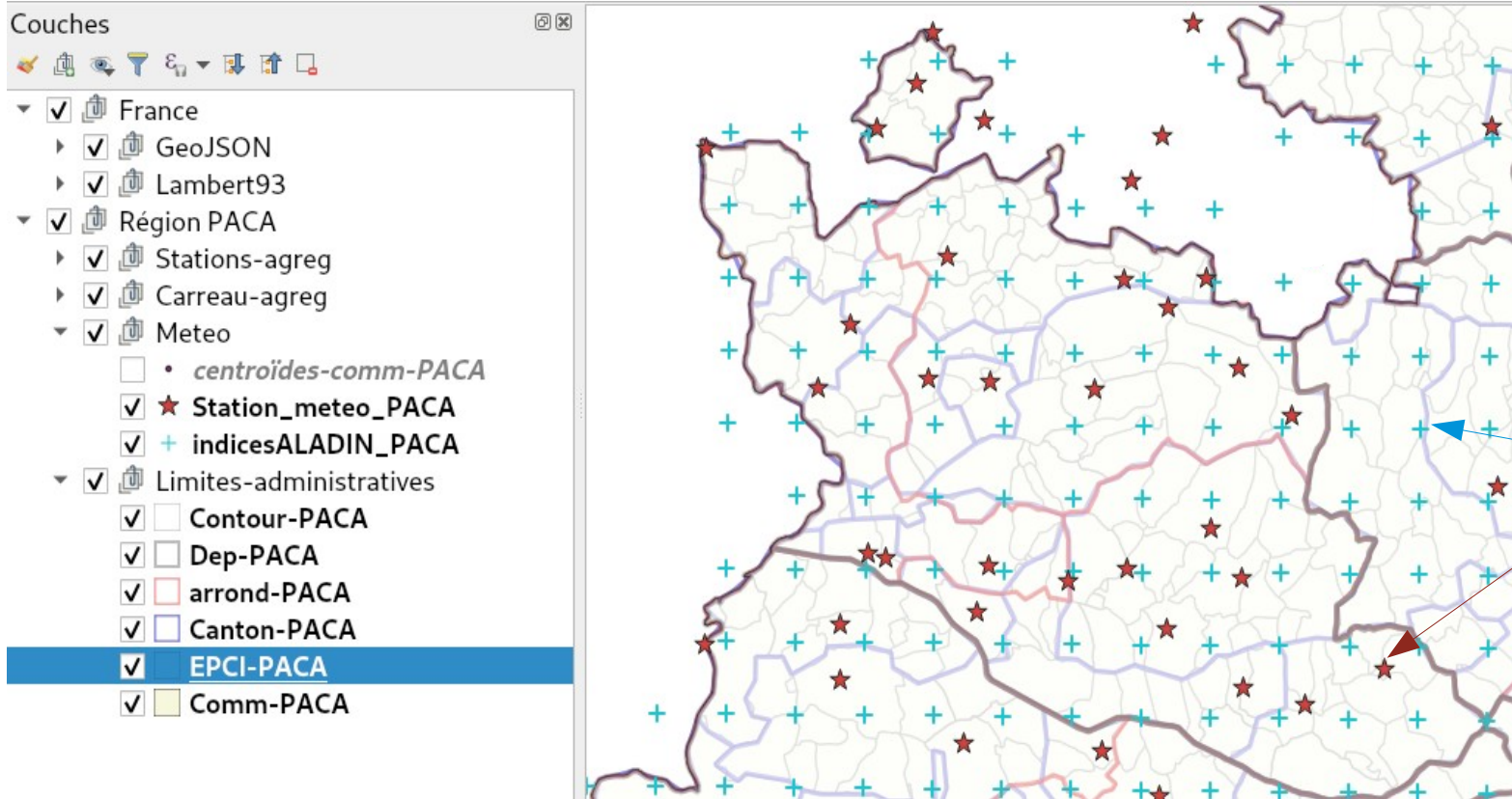
- Studying the relationship between climatic statistical data and aggregation scales
- Finding an accurate and relevant scale for climatic data, at a given level (*i.e.* administrative division)
- Generalizing a method of resampling to eliminate the spatial support effect in rescaling procedures

Data aggregation through scales

Data used (over 30 years)



Aggregation process

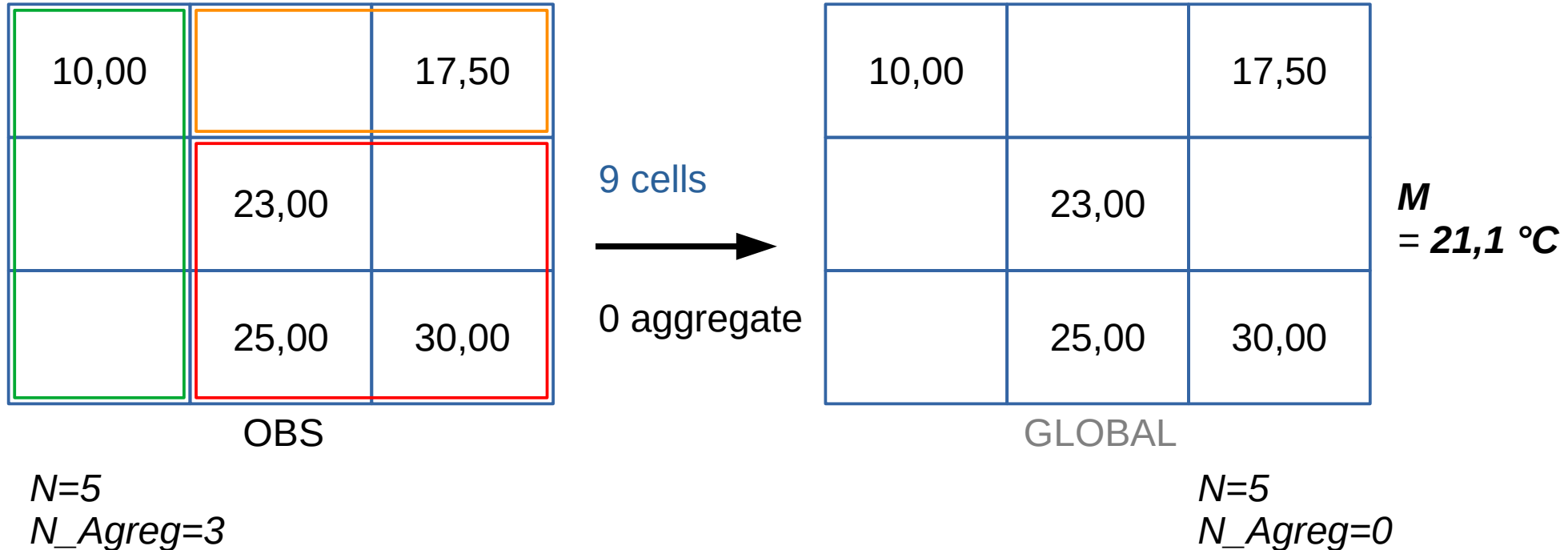


For each aggregate of a given spatial partition, we calculate central values and dispersion statistics of *mean temperature* on both series of:

the points on the grid
the local stations

Calculation without partition (e.g. global average M)

No resampling, we compute the clue on observed data without spatial partition



Calculation considering the partition (e.g. aggregated average M)

We compute the clue on the observed data for each aggregate, and then aggregate them

9 cells
in 3 aggregates

Aggregates
1, 2 et 3

10,00		17,50
	23,00	
	25,00	30,00

OBS

$N=5$

$N_Agreg=3$

Temperature

Average T° of the aggregate 3
= 26 °C

Example :

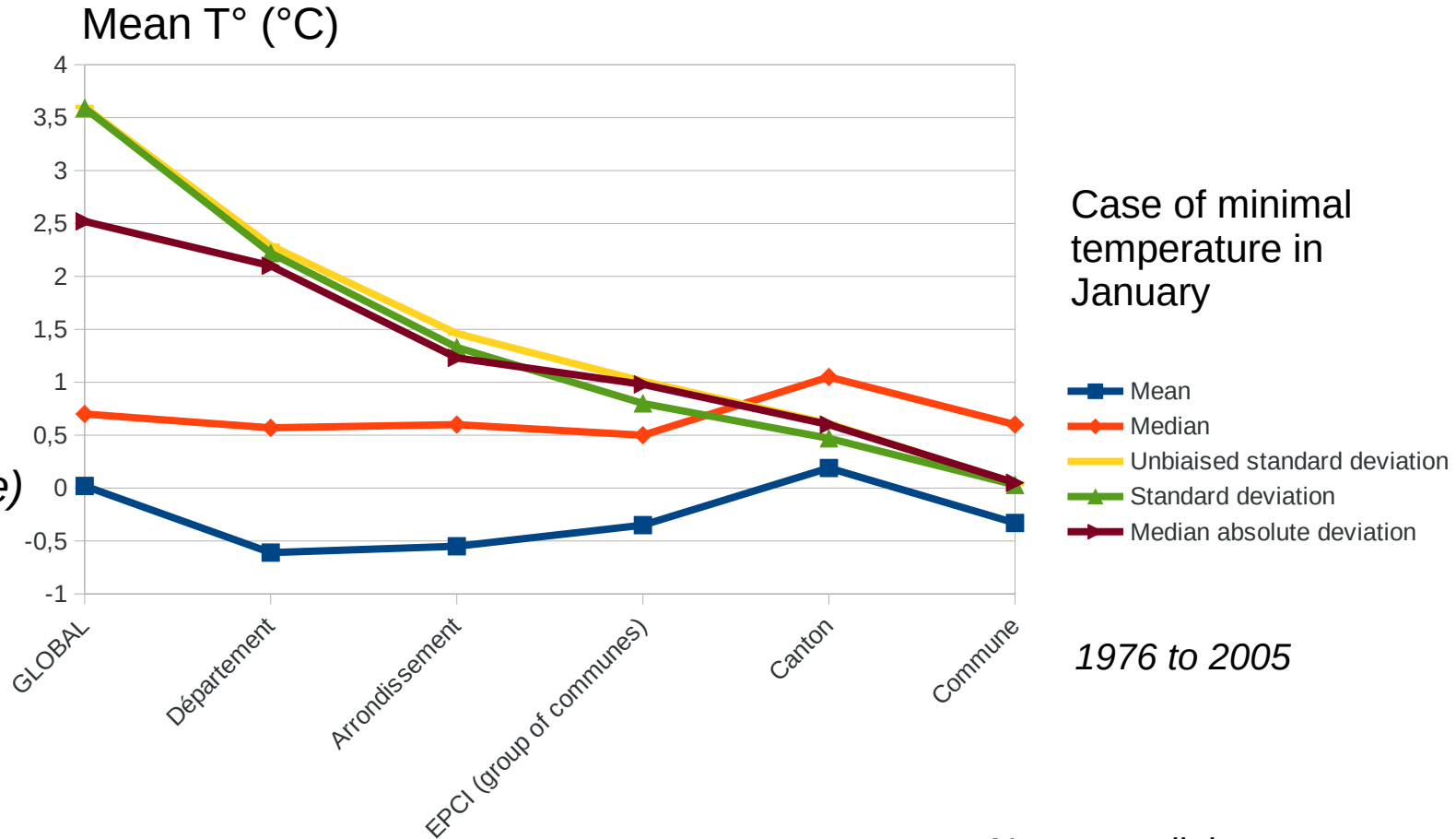
Mean of the average T° :

$$M = (10,00 + 17,50 + 26,00) / 3 = \mathbf{17,83 \text{ } ^\circ\text{C}}$$

Scalogram of temperatures



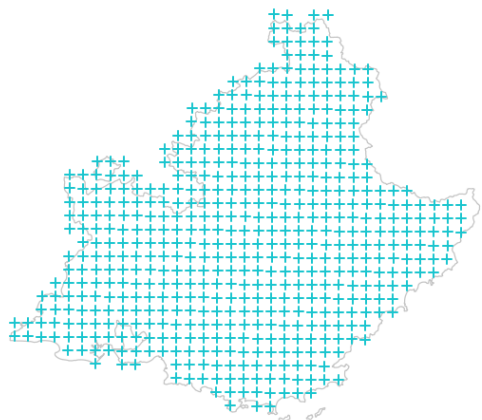
*Stations (Meteo France)
In the region
Provence Alpes
Côte d'Azur*



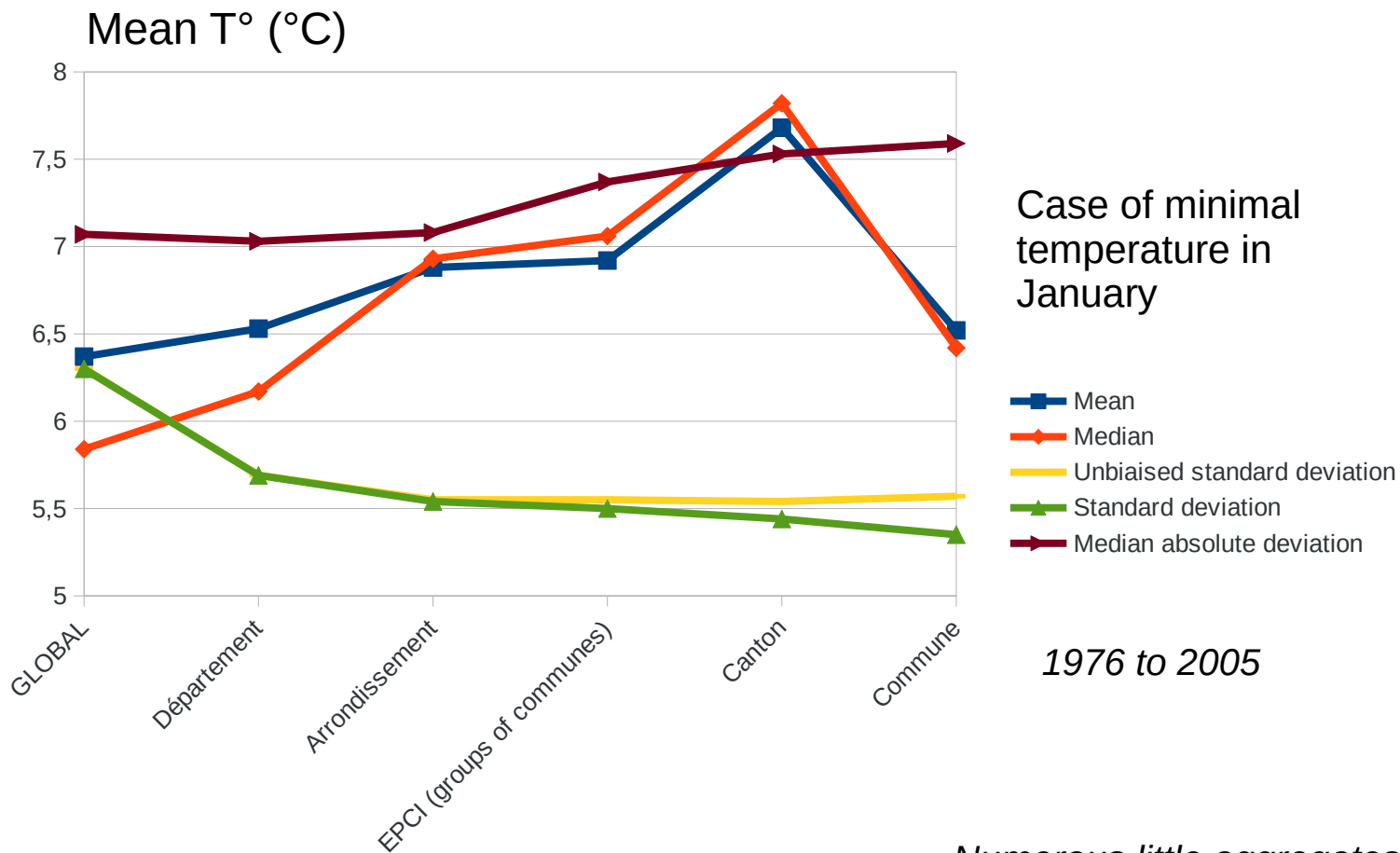
A few large aggregates

Numerous little aggregates

Scalogram of temperatures



*ALADIN climatic model
from CRNM
in region
Provence Alpes
Côte d'Azur*



A few large aggregates ←

→ *Numerous little aggregates*

Observations

- There are significant gaps of mean temperatures according to partitions (up to 3.5 or 8 °C)
- Statistical dispersion decreases when number of aggregates increase
- We observe a peak where mean T° is maximal
- *But: we need a reference to allow comparison between partitions*

Eliminating the spatial support effect
by resampling and Relative
Deviation calculation
(Josselin et al., 2012, 2023)

Resampling procedure

- We randomly permute N times the observed temperatures without changing the spatial partition
- We re-compute statistical clues for each partition
- *This is our “control tube” that draws a random spatial distribution of the temperature, without spatial autocorrelation any more*

10,00	OBS	17,50
	23,00	
	25,00	30,00

$N=5$
 $N_Agreg=3$

9 cells



Aggregates
1, 2 et 3

23,00	RAND	25,00
	10,00	
	30,00	17,50

Permutation 1

$M_1 = 22,38\text{ °C}$



30,00	RAND	10,00
	17,50	
	23,00	25,00

Permutation 2

$M_2 = 20,61\text{ °C}$

Distribution spatiale des valeurs différentes

25,00	RAND	30,00
	23,00	
	10,00	17,50

Permutation 3

$M_3 = 19,77\text{ °C}$

Mean of the M permutations
= 21,05 °C

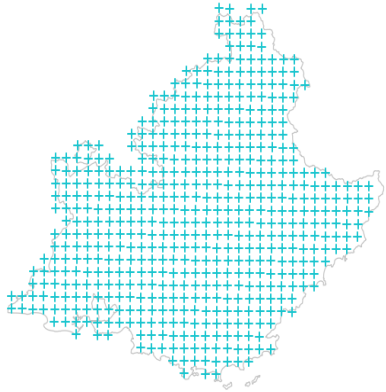
Spatial structure of measures does not change, but values are permuted (drawing without replacement)

Relative Scalar Deviation

$$RSD(\%) = 100 * \frac{T_{obs} - T_{rand}}{T_{rand}}$$

- Due to permutations, the **normalized difference between the observed and the randomized clue allows to eliminate the change of support problem because its effect is similar in both cases** (for a given scale/partition)
- **The Relative Scalar Deviation reflects the effective part of the geography in the measured values (e.g. Temperature), because the random process deleted all the spatial autocorrelation**

Scalogram with Relative Scalar Deviation

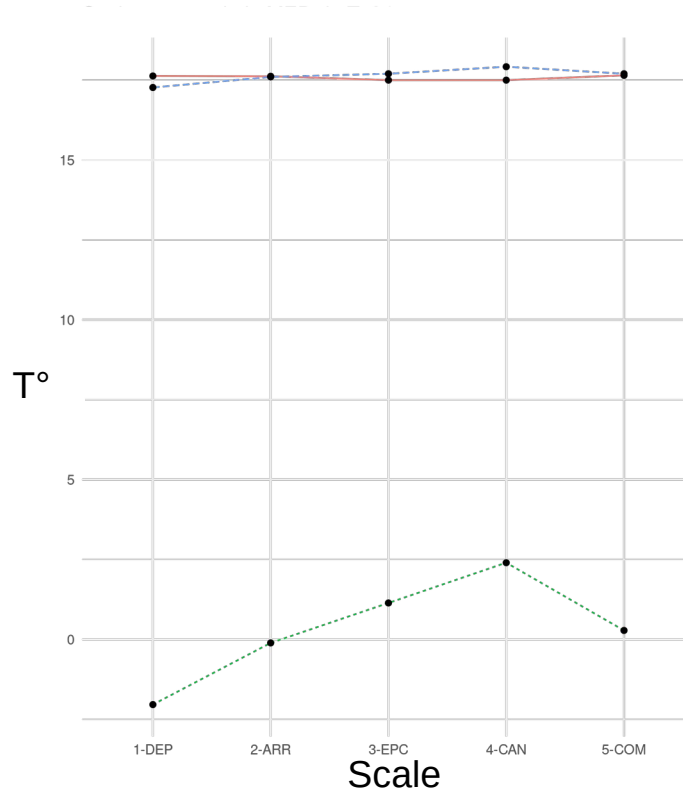


ALADIN climatic model from CRNM in region Provence Alpes Côte d'Azur

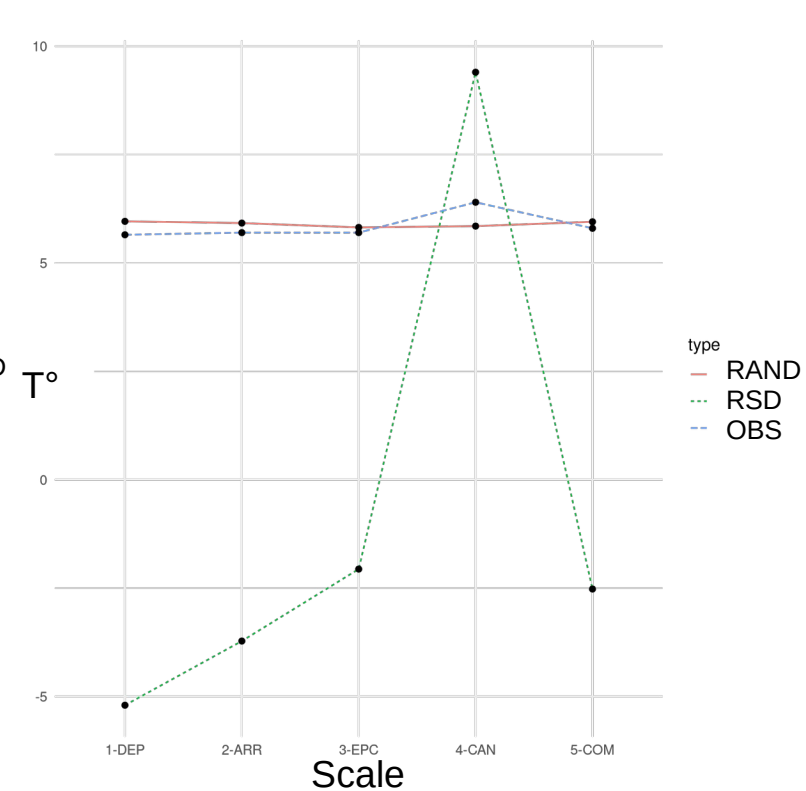
1976 to 2005

Median Temperature (°C)

Maximal in April



Minimal in April



Scalogram with Relative Scalar Deviation



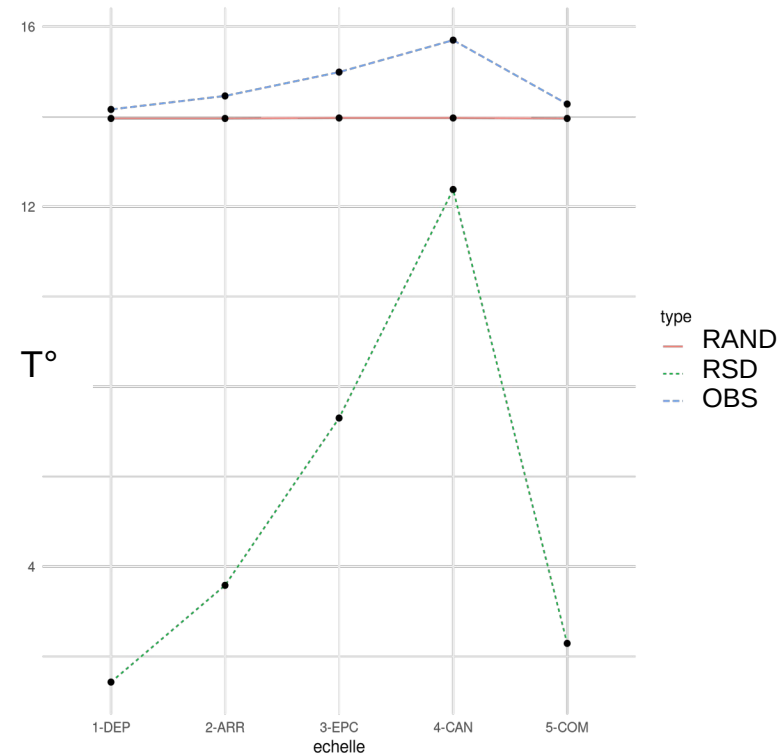
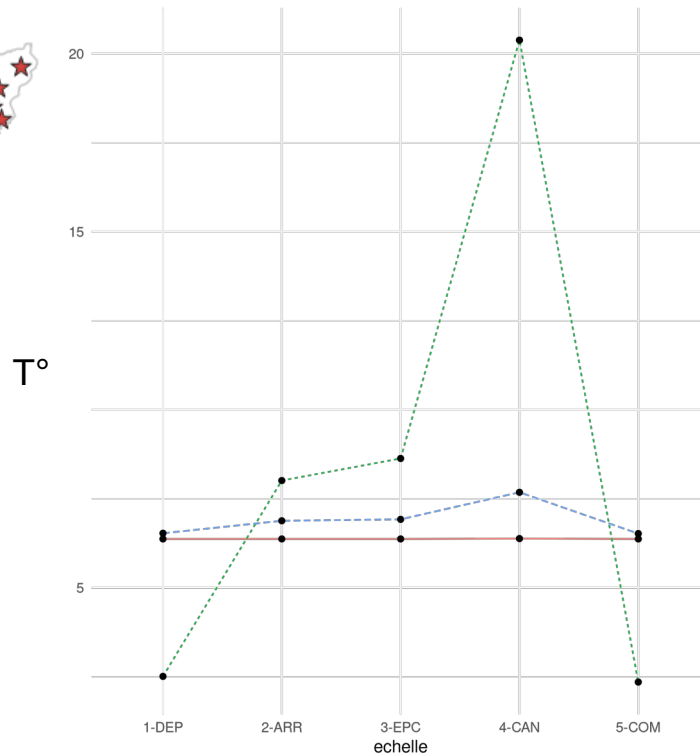
Stations (Meteo France)
en région
Provence Alpes
Côte d'Azur

1976 to 2005

Mean Temperature (°C)

Minimal during the year

Maximal during the year



Conclusion

- A proposition to eliminate the Change of Spatial Support Problem
- The County scale seems to be the partition the furthest from a random distribution of mean T°
- It was shown that different scales can appear as relevant depending on the tackled topic
- The Relative Scalar Deviation being generalized

Josselin *et al.*, 2023, Uncertainties related to real estate price estimation scales, in *Geographic Data Imperfection 2* (Eds.: Batton-Hubert & Pinet) ISTE Wiley

ECTQ2023: Braga (Portugal) 14 -17 September 2023

UNIVERSIDADE D
COIMBRA

Thank you for your attention

Didier.josselin@univ-avignon.fr



Simpson paradox

ONE SAMPLE

	Sick due to pesticide	Not sick	TOTAL
Orsini Viper	200	800	1000
Apollon Butterfly	50	950	1000
TOTAL	250	1750	2000

Probability to be sick for vipers: $200/1000 = 0,20 = 20 \%$

Probability to be sick for butterflies: $50/1000 = 0,05 = 5 \%$

Relative Risk = $0,20/0,05 = 4$

(4 times more for vipers)

Simpson paradox

TWO
SEPARATED SAMPLES

<u>Sample 1</u>	Sick	OK	TOTAL
Viper	193	224	417
Butterly	39	45	84
TOTAL	232	269	501

$$\text{Relative Risk} = (193/417) / (39/84) = 1$$

<u>Sample 2</u>	Sick	OK	TOTAL
Viper	7	576	583
Butterfly	11	905	916
TOTAL	18	1481	1499

$$\text{Relative Risk} = (7/583) / (11/916) = 1$$

Simpson paradox

TWO
SEPARATED SAMPLES

<u>Sample 1</u>	Sick	OK	TOTAL
Viper	193	224	417
Butterfly	39	45	84
TOTAL	232	269	501

$$\text{Relative Risk} = (193/417) / (39/84) = 1$$

<u>Sample 2</u>	Sick	OK	TOTAL
Viper	7	576	583
Butterfly	11	905	916
TOTAL	18	1481	1499

$$\text{Relative Risk} = (7/583) / (11/916) = 1$$

Aggregated Relative Risk = 1 ←

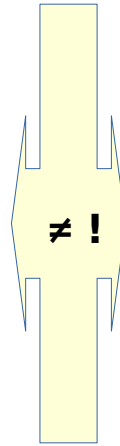
(same risk)

Simpson paradox

TWO
SEPARATED SAMPLES

	Sick due to pesticide	Not sick	TOTAL
Orsini Viper	200	800	1000
Apollon Butterfly	50	950	1000
TOTAL	250	1750	2000

ONE SAMPLE



<u>Sample 1</u>	Sick	OK	TOTAL
Viper	193	224	417
Butterfly	39	45	84
TOTAL	232	269	501

$$\text{Relative Risk} = (193/417) / (39/84) = 1$$

<u>Sample 2</u>	Sick	OK	TOTAL
Viper	7	576	583
Butterfly	11	905	916
TOTAL	18	1481	1499

$$\text{Relative Risk} = (7/583) / (11/916) = 1$$

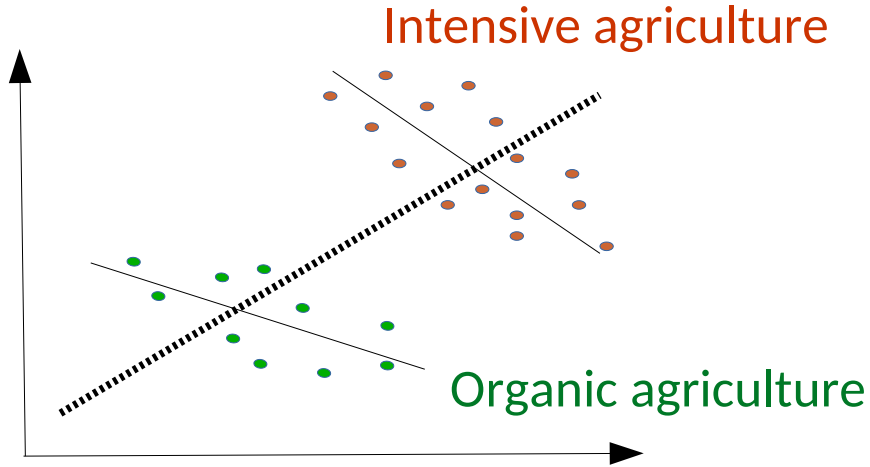
Aggregated Relative Risk = 1 ←

(same risk)

(Spatial) aggregation bias

2 parcels of vine
described by V1 et V2

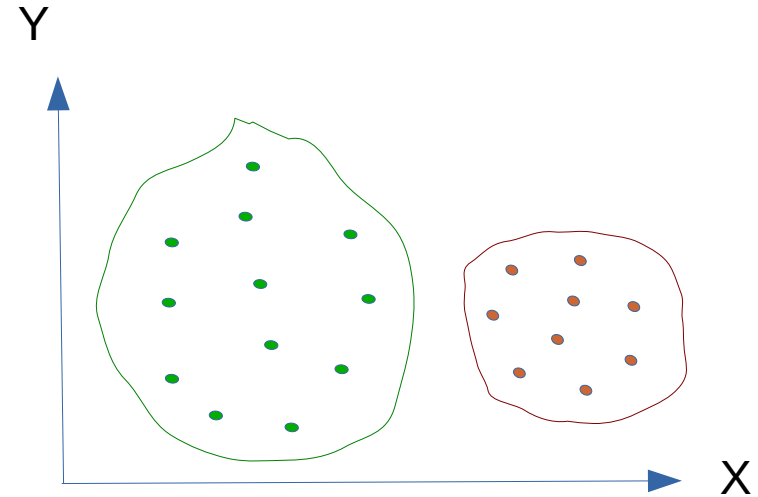
V2 : cost / ha



V1 : production / ha

In a point plot

Located measures
on a map



In geographical space