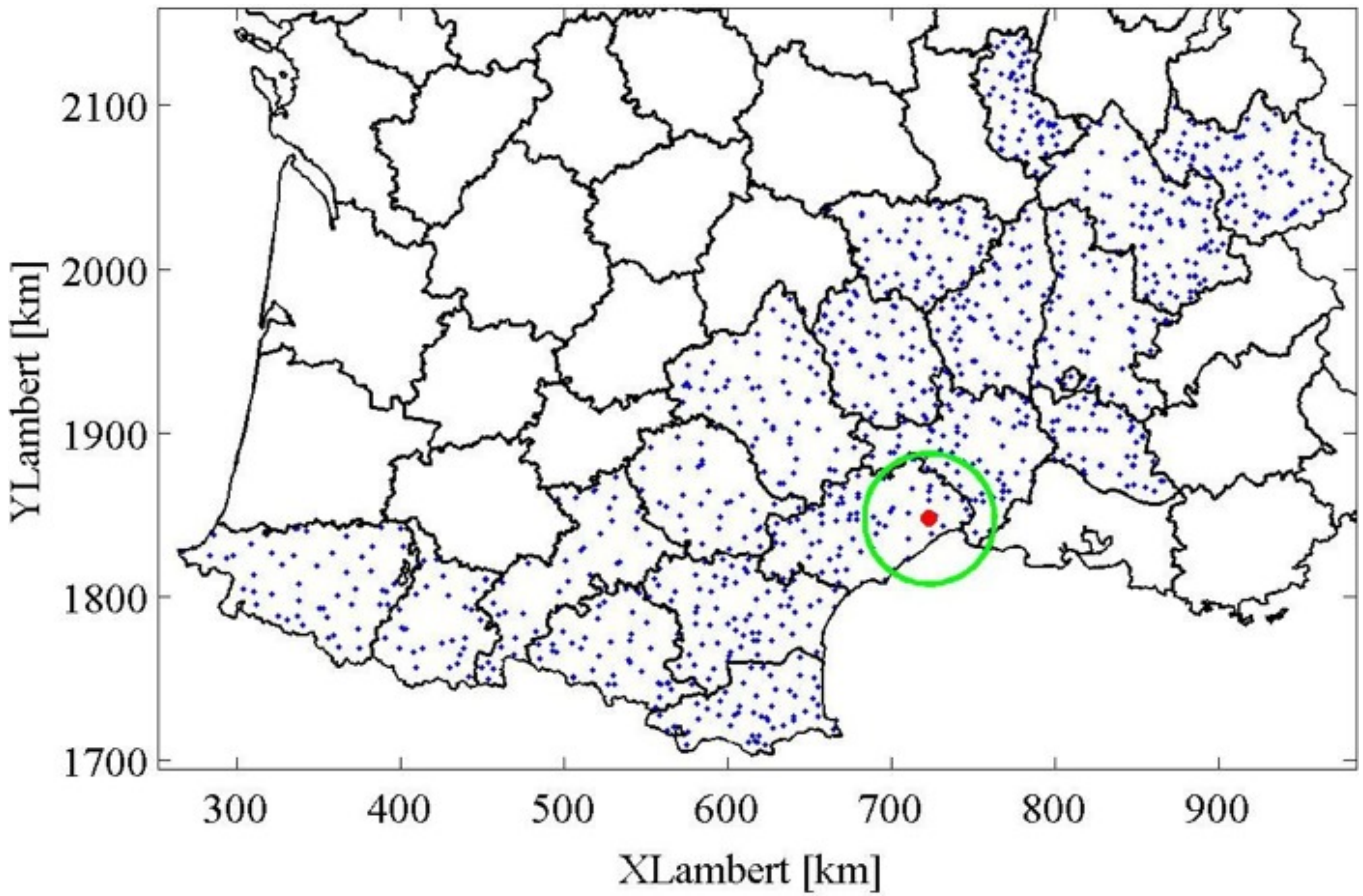
Sources of Variability in Regional Frequency Analysis of Annual Maximum Rainfall in the South of France

Extraflo ANR project

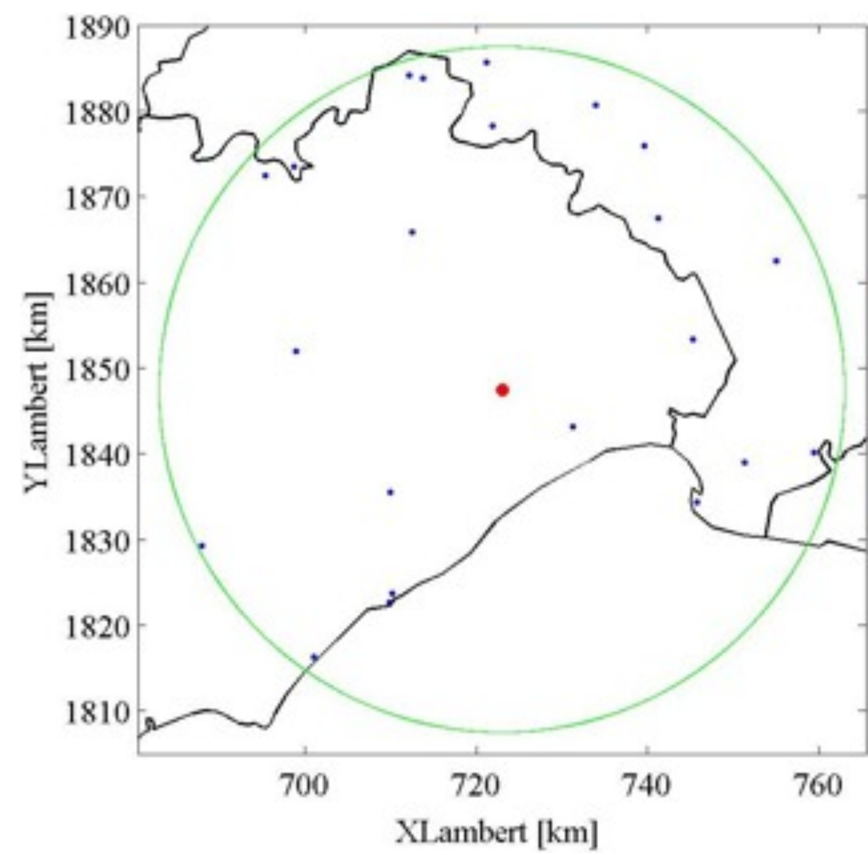
Julie Carreau (julie.carreau@univ-montp2.fr) & Luc Neppel
HydroSciences Montpellier, France
PEPER workshop, Aussois 2013

Regional Frequency Analysis

40 km circular neighborhood on Montpellier



Regional Frequency Analysis



Stations : $\{S_1, \ldots, S_k\}$

Annual Maxima : $S_i = \{X_1^i, \ldots, X_{n_i}^i\}$

Index values : $m_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_j$

Normalized Data : $Y_j^i = \frac{X_j^i}{m_i}$

Regional Sample : $S_{\text{regional}} = \{Y_1^1, \ldots, Y_{n_1}^1, \ldots, Y_1^k, \ldots, Y_{n_k}^k\}$

Regional Frequency Analysis

Assumptions

1. Homogeneity : IID of regional sample

$$S_{\text{regional}} = \{Y_1^1, \ldots, Y_{n_1}^1, \ldots, Y_1^k, \ldots, Y_{n_k}^k\}$$

2. $Y \sim GEV(\mu_R, \sigma_R, \xi_R)$ with $E[Y] = 1$

3. $X^i \sim GEV(\mu_R\, m_i, \sigma_R\, m_i, \xi_R)$

Regional Frequency Analysis

Assumptions

1. Homogeneity : IID of regional sample

Spatial declustering

$$S_{\text{regional}} = \{Y_1^1, \ldots, Y_{n_1}^1, \ldots, Y_1^k, \ldots, Y_{n_k}^k\}$$

2. $Y \sim GEV(\mu_R, \sigma_R, \xi_R)$ with $E[Y] = 1$

3. $X^i \sim GEV(\mu_R\, m_i, \sigma_R\, m_i, \xi_R)$

Regional Frequency Analysis

Assumptions

1. Homogeneity : IID of regional sample

Spatial declustering

$$S_{\text{regional}} = \{Y_1^1, \ldots, Y_{n_1}^1, \ldots, Y_1^k, \ldots, Y_{n_k}^k\}$$

Estimation of index values

2. $Y \sim GEV(\mu_R, \sigma_R, \xi_R)$ with $E[Y] = 1$

3. $X^i \sim GEV(\mu_R\, m_i, \sigma_R\, m_i, \xi_R)$

Regional Frequency Analysis

Assumptions

1. Homogeneity : IID of regional sample

Spatial declustering

$$S_{\text{regional}} = \{Y_1^1, \ldots, Y_{n_1}^1, \ldots, Y_1^k, \ldots, Y_{n_k}^k\}$$

2. $Y \sim GEV(\mu_R, \sigma_R, \xi_R)$ with $E[Y] = 1$

Estimation of index values

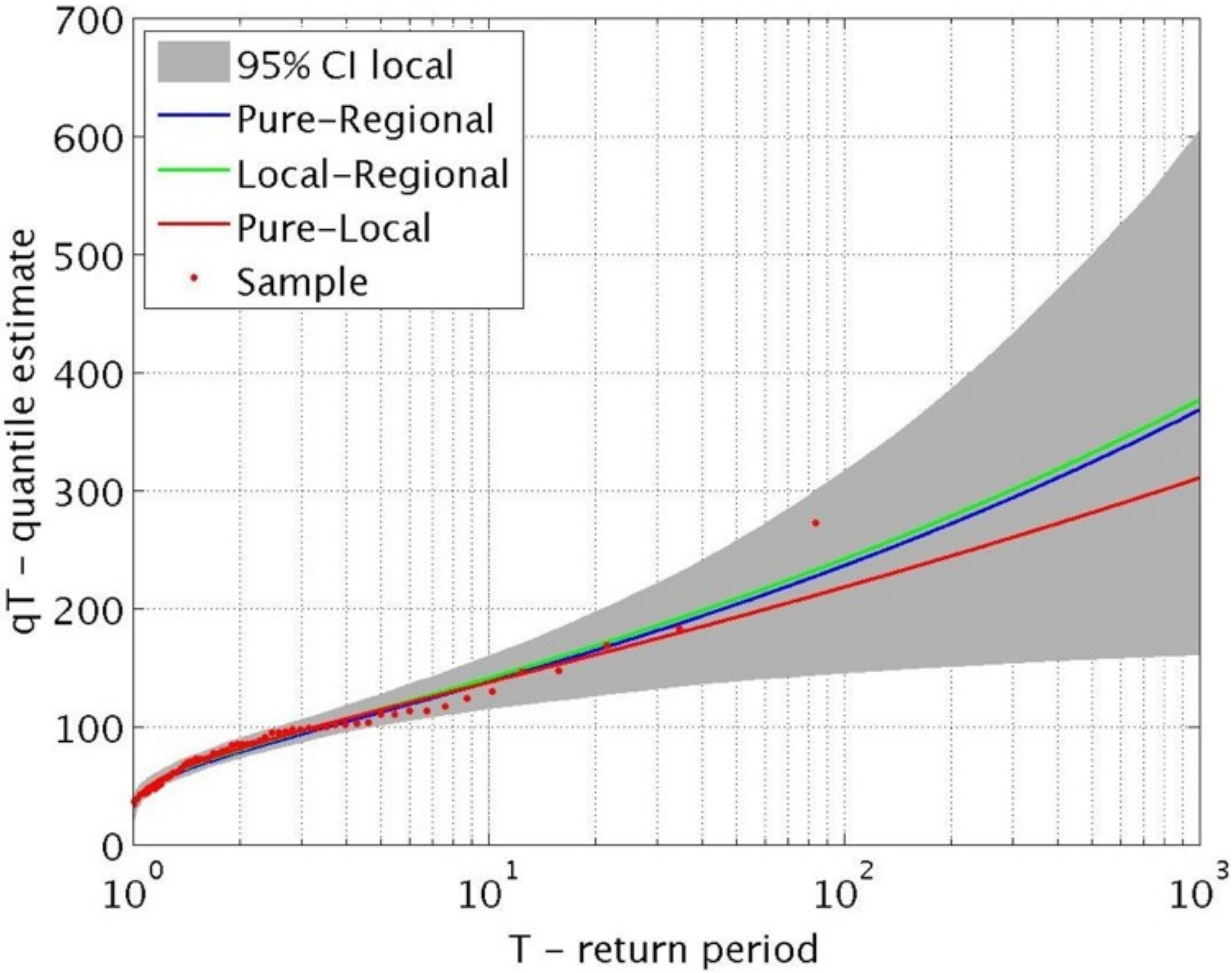
3. $X^i \sim GEV(\mu_R \, m_i, \sigma_R \, m_i, \xi_R)$

Annual maxima of hydrological year

Sources of Variability

- Sources of variability in the choice of the neighborhood **IID assumption**
- Variance of the L-Moment estimator of (μ_R, σ_R, ξ_R)
- Variability introduced by the interpolation of the index values **at ungauged sites**

Sources of Variability



Study Area : South of France

Calibration set : 609 stations with 2-34 maxima

Validation set : 411 stations with 46-58 maxima

Study Area : South of France

Calibration set : 609 stations with 2-34 maxima

Validation set : 411 stations with 46-58 maxima

Goal

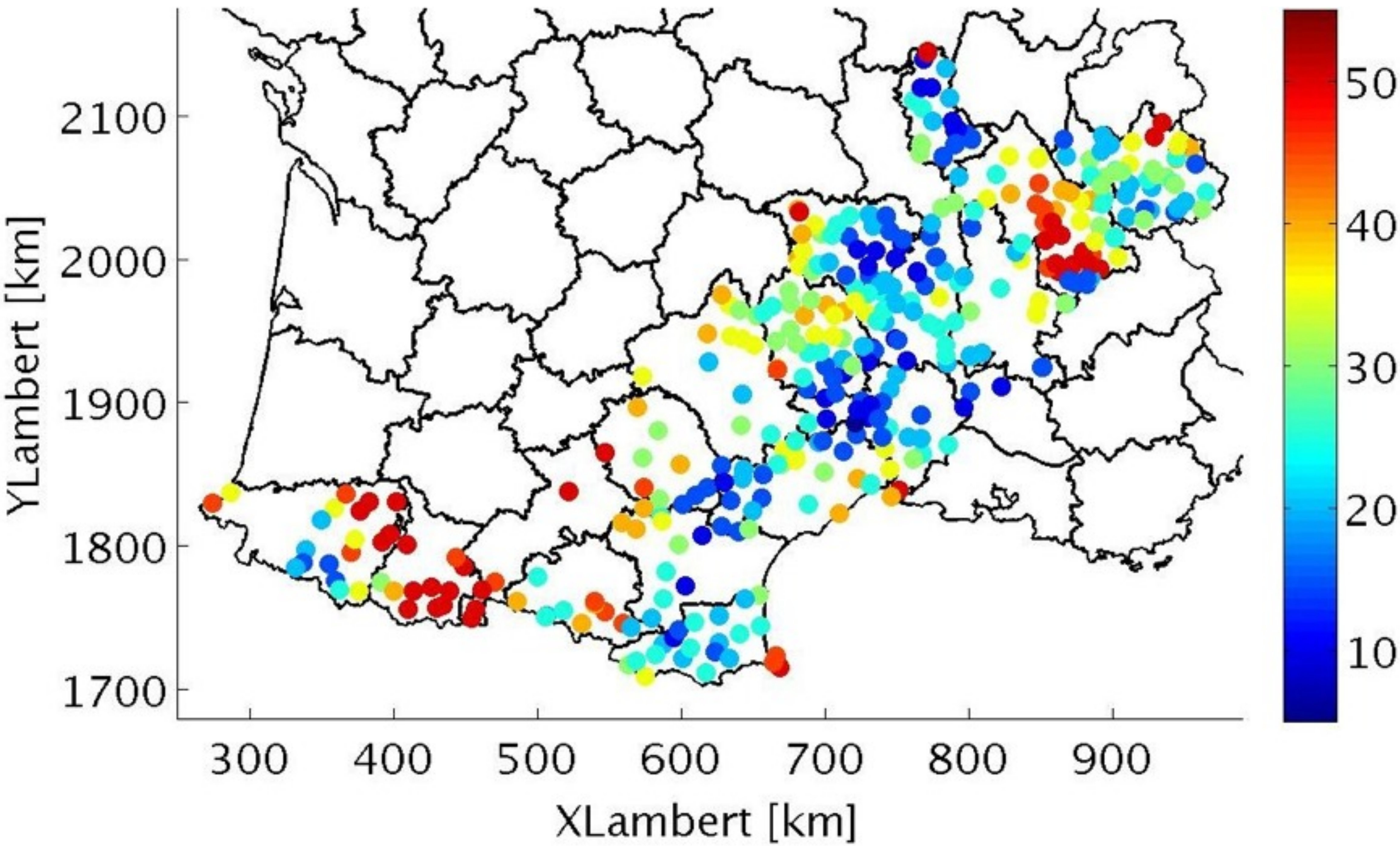
- Compute the regional estimate at each validation station
- Compare with the local estimate

Choice of neighborhood

1. **Homogeneous circular** neighborhood (Hosking & Wallis, Anderson & Darling)
2. **Fixed 50 km radius** circular neighborhood
3. **Fixed regional sample** size (> 100)
4. **Random radius circular** neighborhood (10-50km)
5. **10 stations at random** **Benchmark**

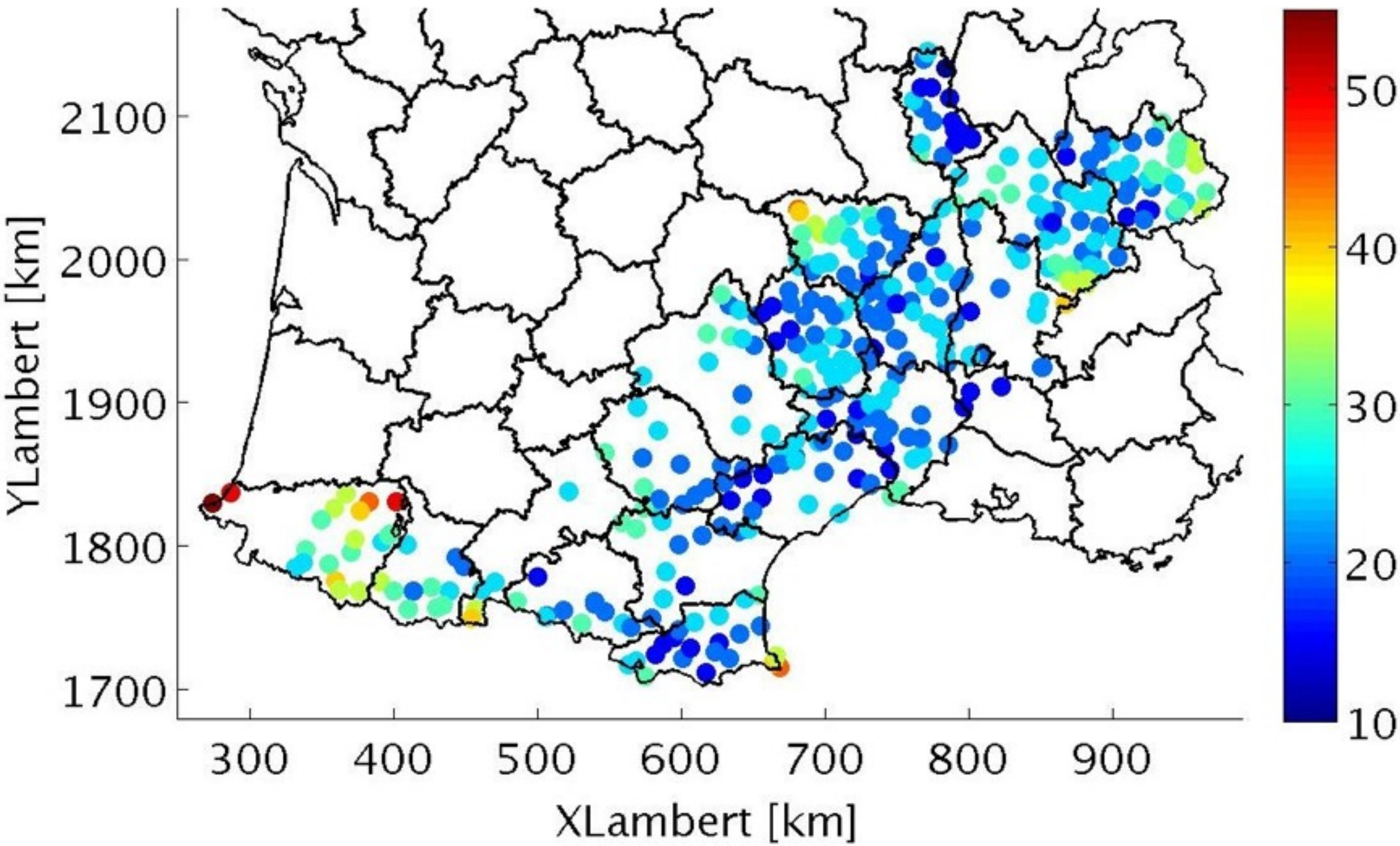
Size of neighborhood

Homogeneous circular neighborhood



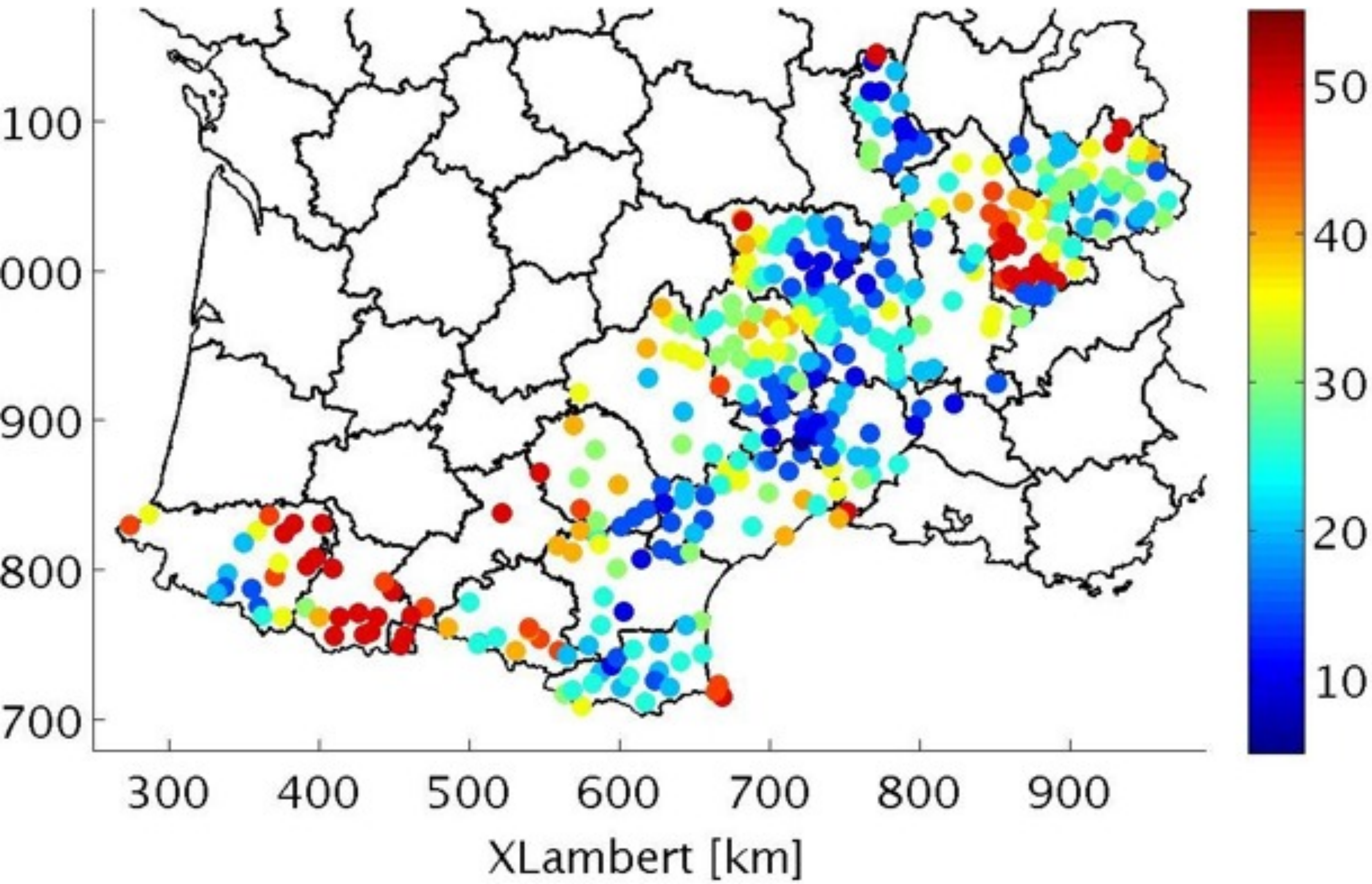
Size of neighborhood

100 - fixed size regional sample

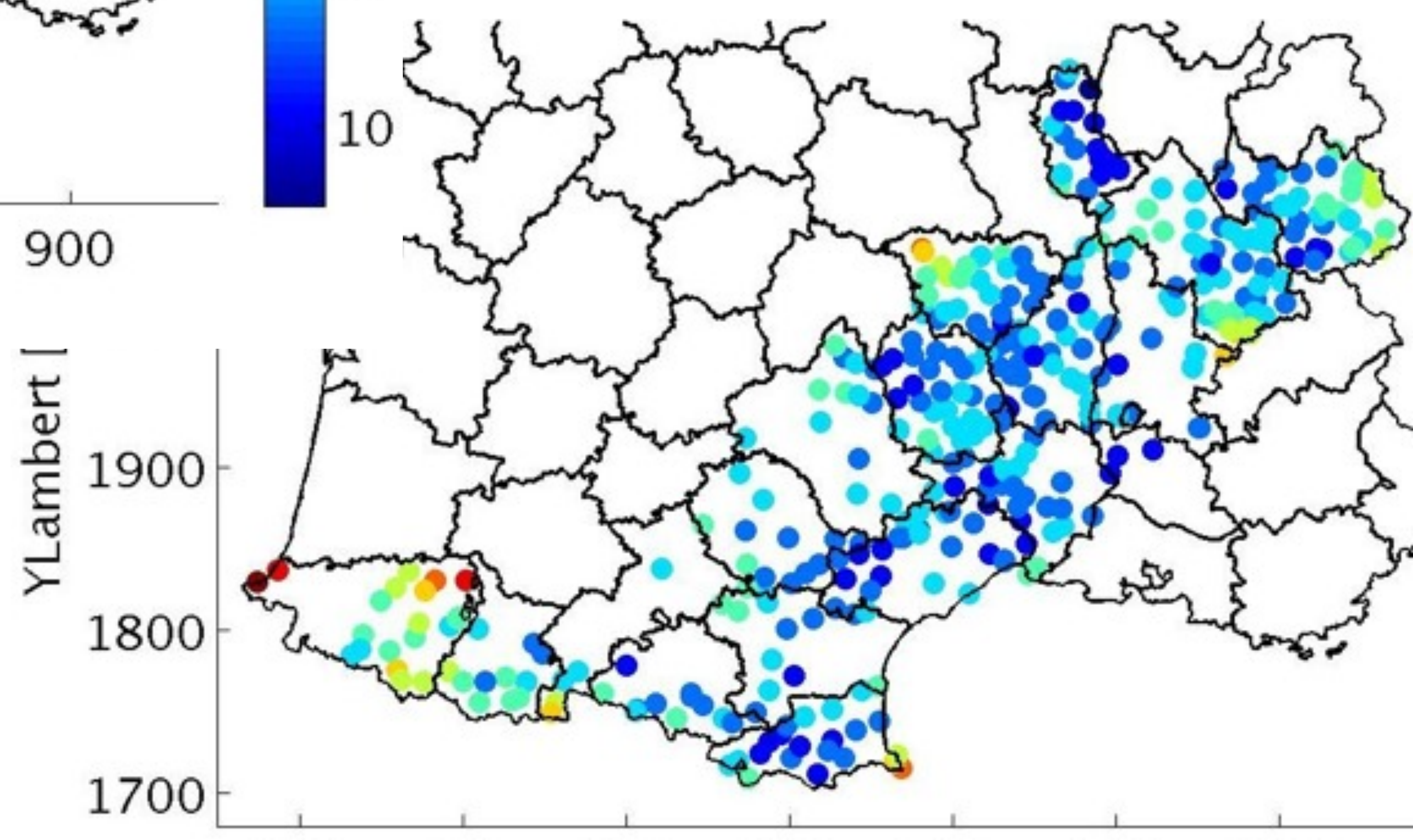


Size of neighborhood

Homogeneous circular neighborhood

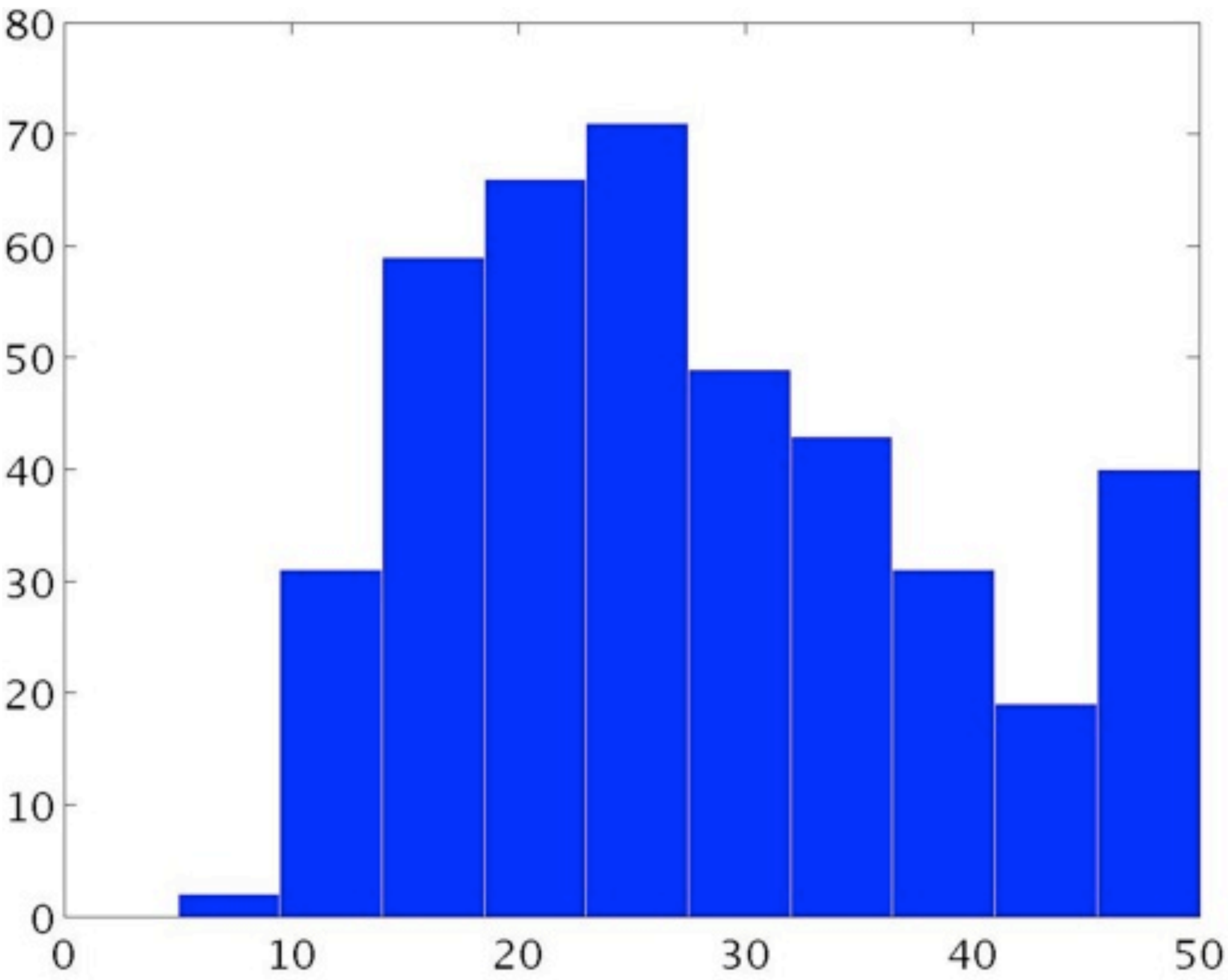


fixed size regional sample

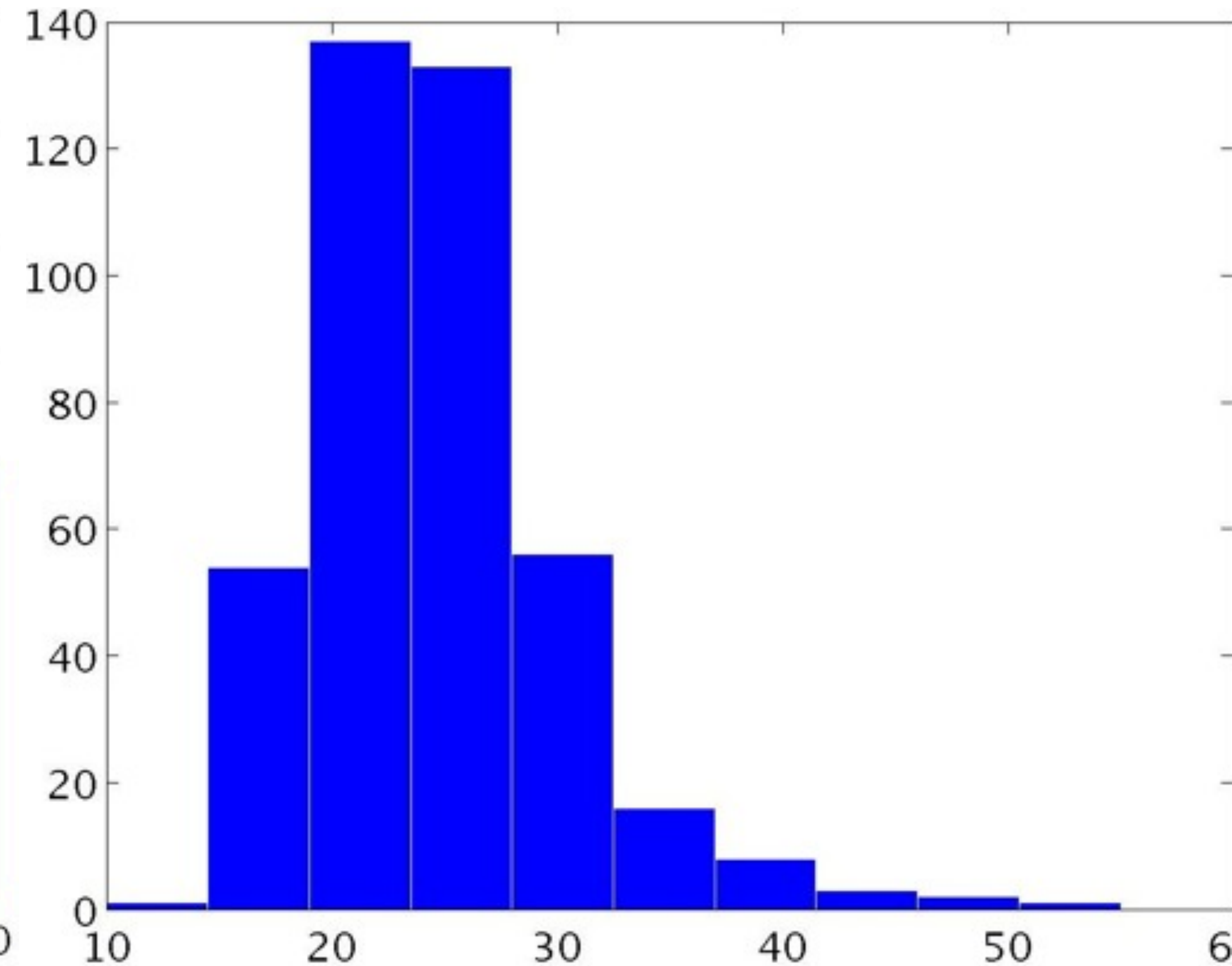


Size of neighborhood

Homogeneous circular neighborhood

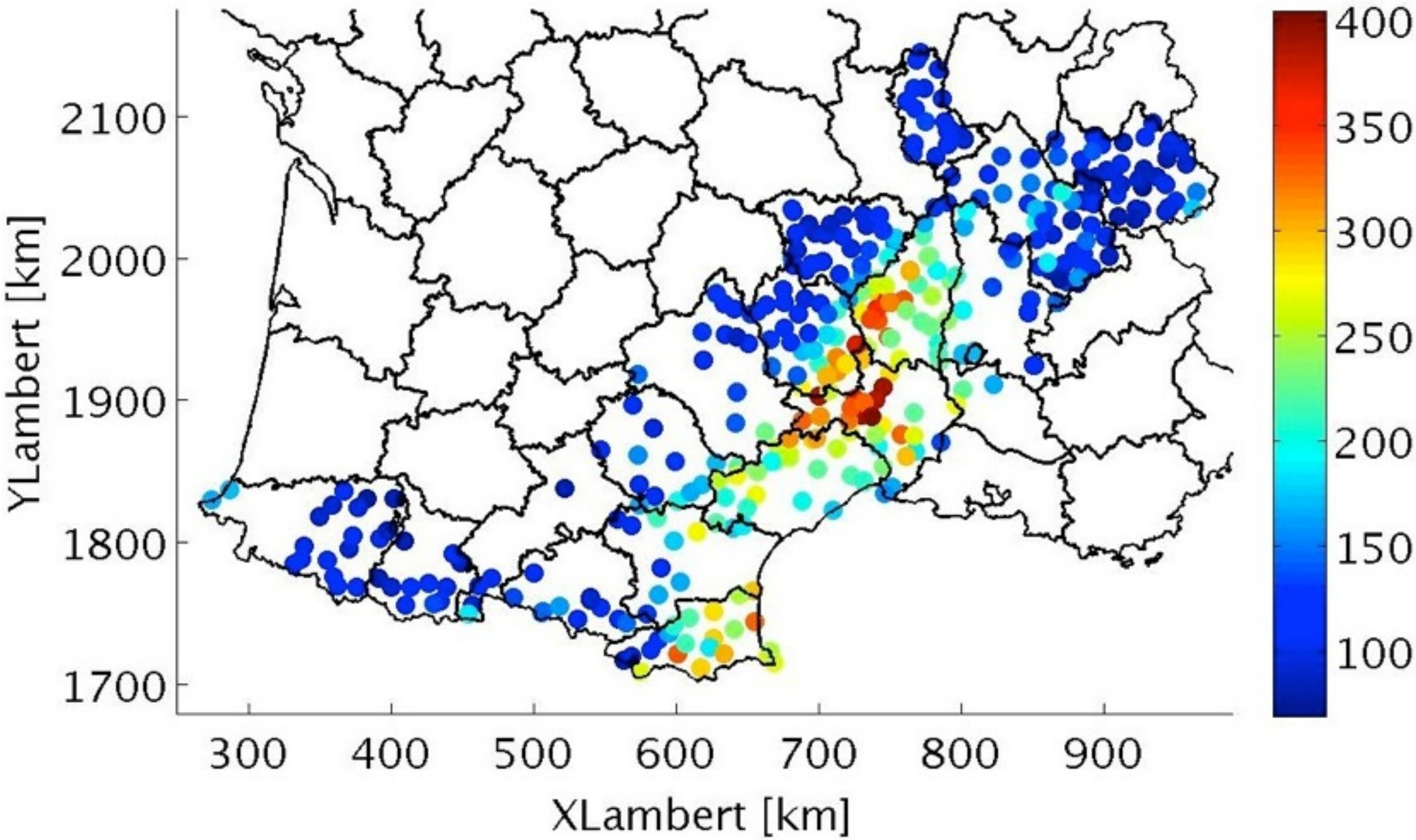


100 - Fixed size regional sample



Local Quantile Estimates

Level 99 %



Relative Quantile Error

$$1 - \frac{\hat{y}_p}{y_p^{\mathrm{loc}}} \begin{cases} = 0 & \Leftrightarrow \hat{y}_p = y_p^{\mathrm{loc}} \\ > 0 & \Leftrightarrow \hat{y}_p < y_p^{\mathrm{loc}} \\ < 0 & \Leftrightarrow \hat{y}_p > y_p^{\mathrm{loc}} \end{cases}$$

Perfect estimation

Under-estimation

Over-estimation

Relative Quantile Error

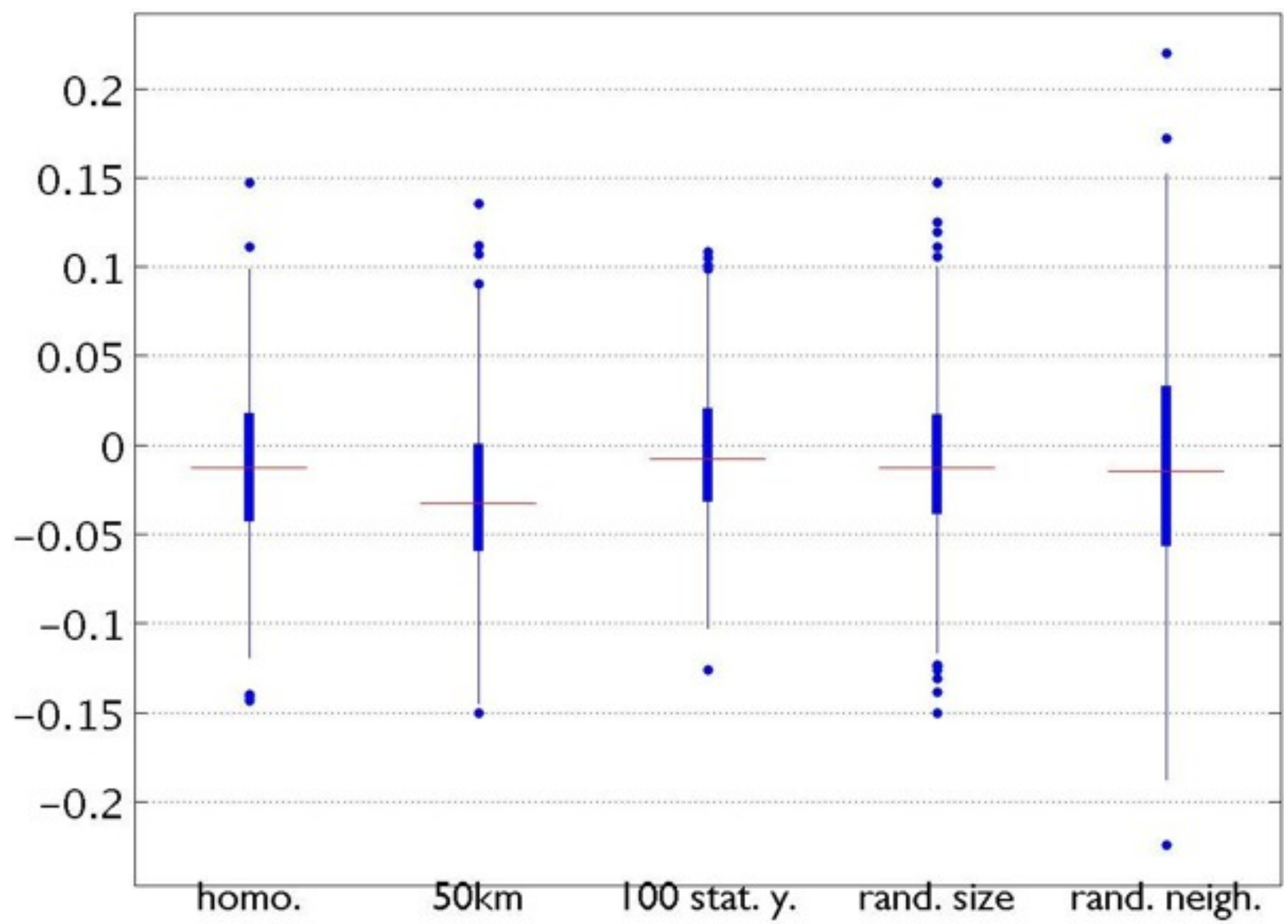
$$1 - \frac{\hat{y}_p}{y_p^{\mathrm{loc}}} \begin{cases} = 0 & \Leftrightarrow \hat{y}_p = y_p^{\mathrm{loc}} \\ > 0 & \Leftrightarrow \hat{y}_p < y_p^{\mathrm{loc}} \\ < 0 & \Leftrightarrow \hat{y}_p > y_p^{\mathrm{loc}} \end{cases}$$

Perfect estimation

Under-estimation

Over-estimation

Level 90 %



Relative Quantile Error

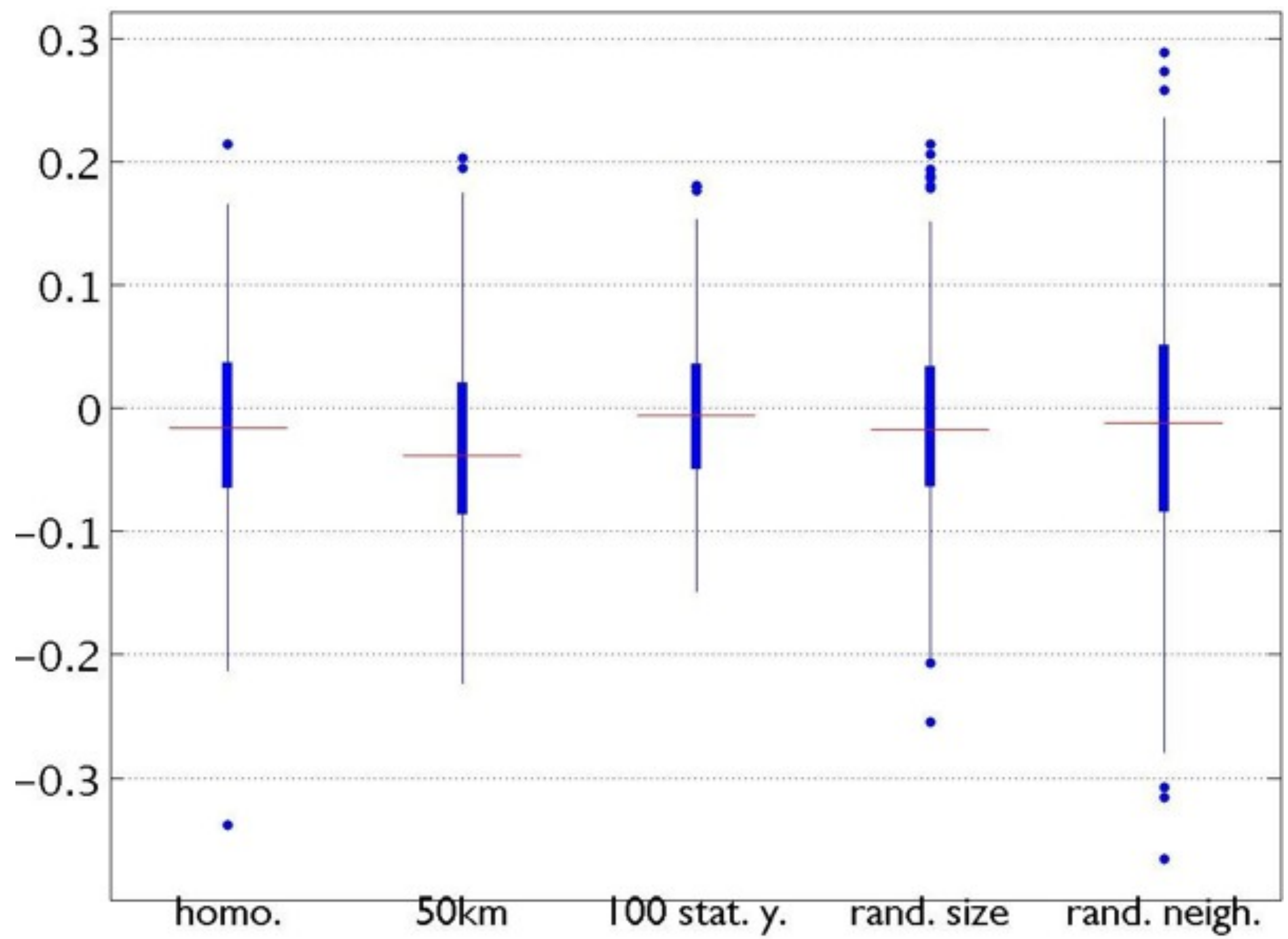
$$1 - \frac{\hat{y}_p}{y_p^{\text{loc}}} \begin{cases} = 0 & \Leftrightarrow \hat{y}_p = y_p^{\text{loc}} \\ > 0 & \Leftrightarrow \hat{y}_p < y_p^{\text{loc}} \\ < 0 & \Leftrightarrow \hat{y}_p > y_p^{\text{loc}} \end{cases}$$

Perfect estimation

Under-estimation

Over-estimation

Level 95 %



Relative Quantile Error

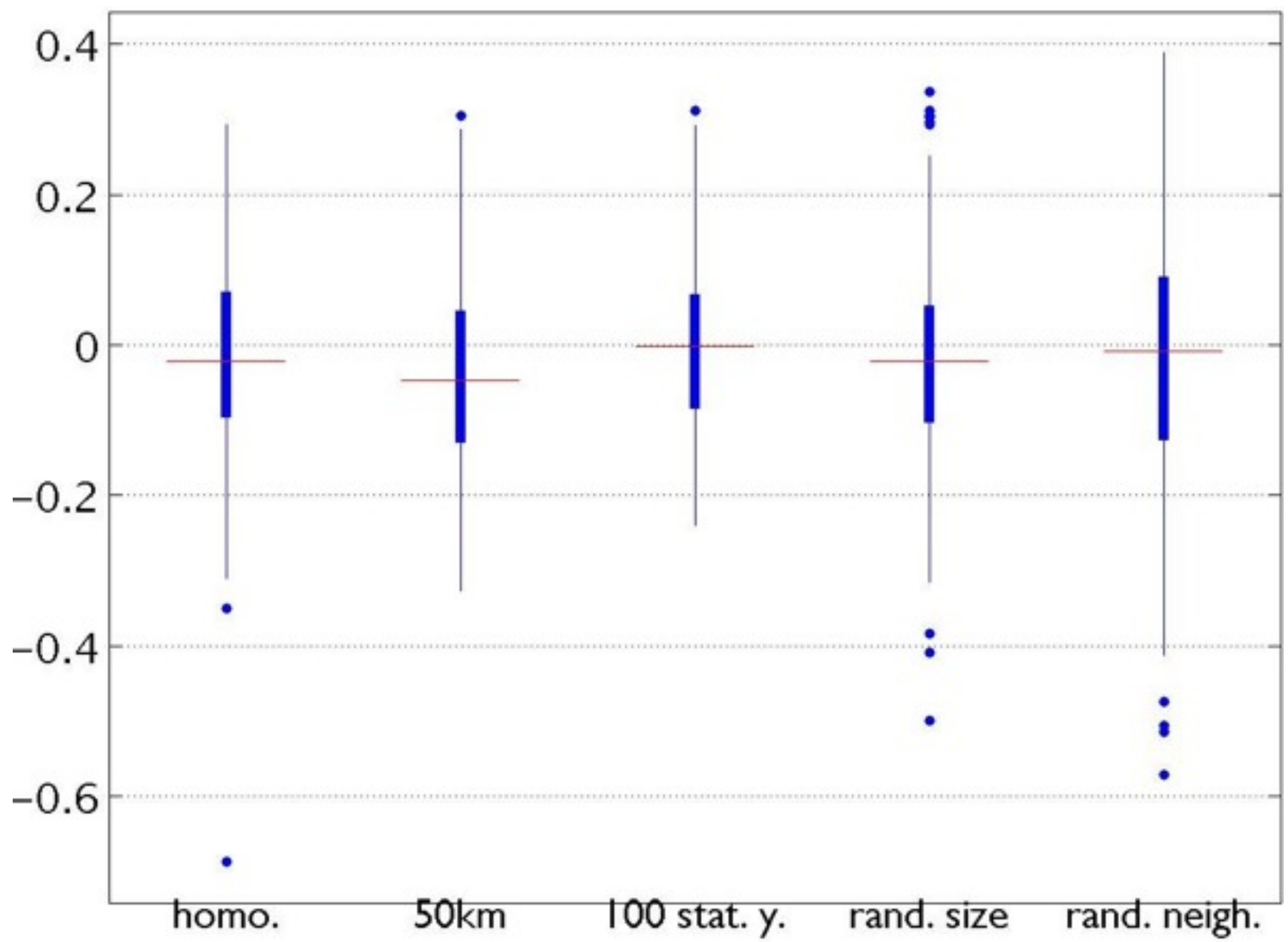
$$1 - \frac{\hat{y}_p}{y_p^{\text{loc}}} \begin{cases} = 0 & \Leftrightarrow \hat{y}_p = y_p^{\text{loc}} \\ > 0 & \Leftrightarrow \hat{y}_p < y_p^{\text{loc}} \\ < 0 & \Leftrightarrow \hat{y}_p > y_p^{\text{loc}} \end{cases}$$

Perfect estimation

Under-estimation

Over-estimation

Level 98 %



Relative Quantile Error

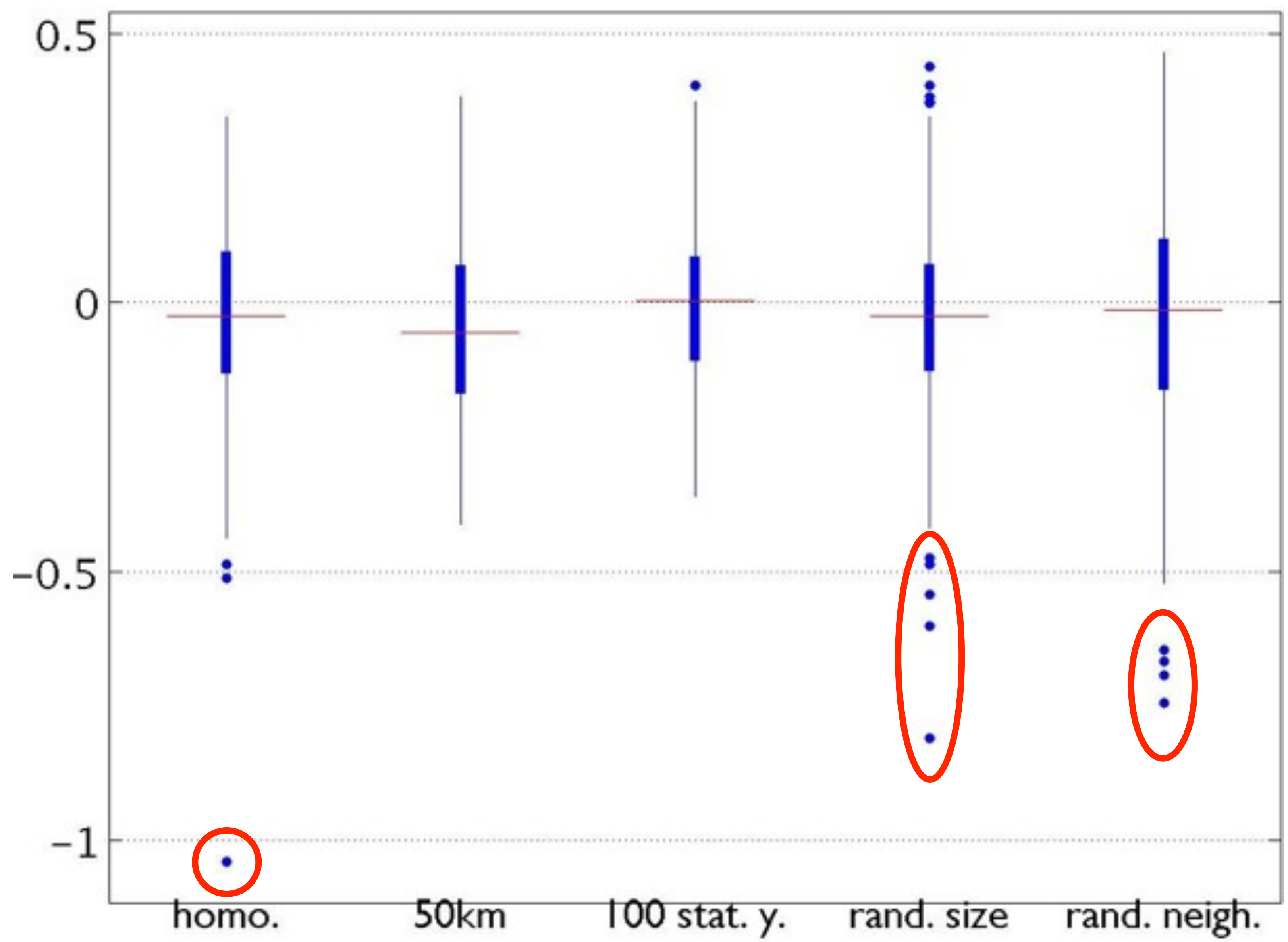
$$1 - \frac{\hat{y}_p}{y_p^{\text{loc}}} \begin{cases} = 0 & \Leftrightarrow \hat{y}_p = y_p^{\text{loc}} \\ > 0 & \Leftrightarrow \hat{y}_p < y_p^{\text{loc}} \\ < 0 & \Leftrightarrow \hat{y}_p > y_p^{\text{loc}} \end{cases}$$

Perfect estimation

Under-estimation

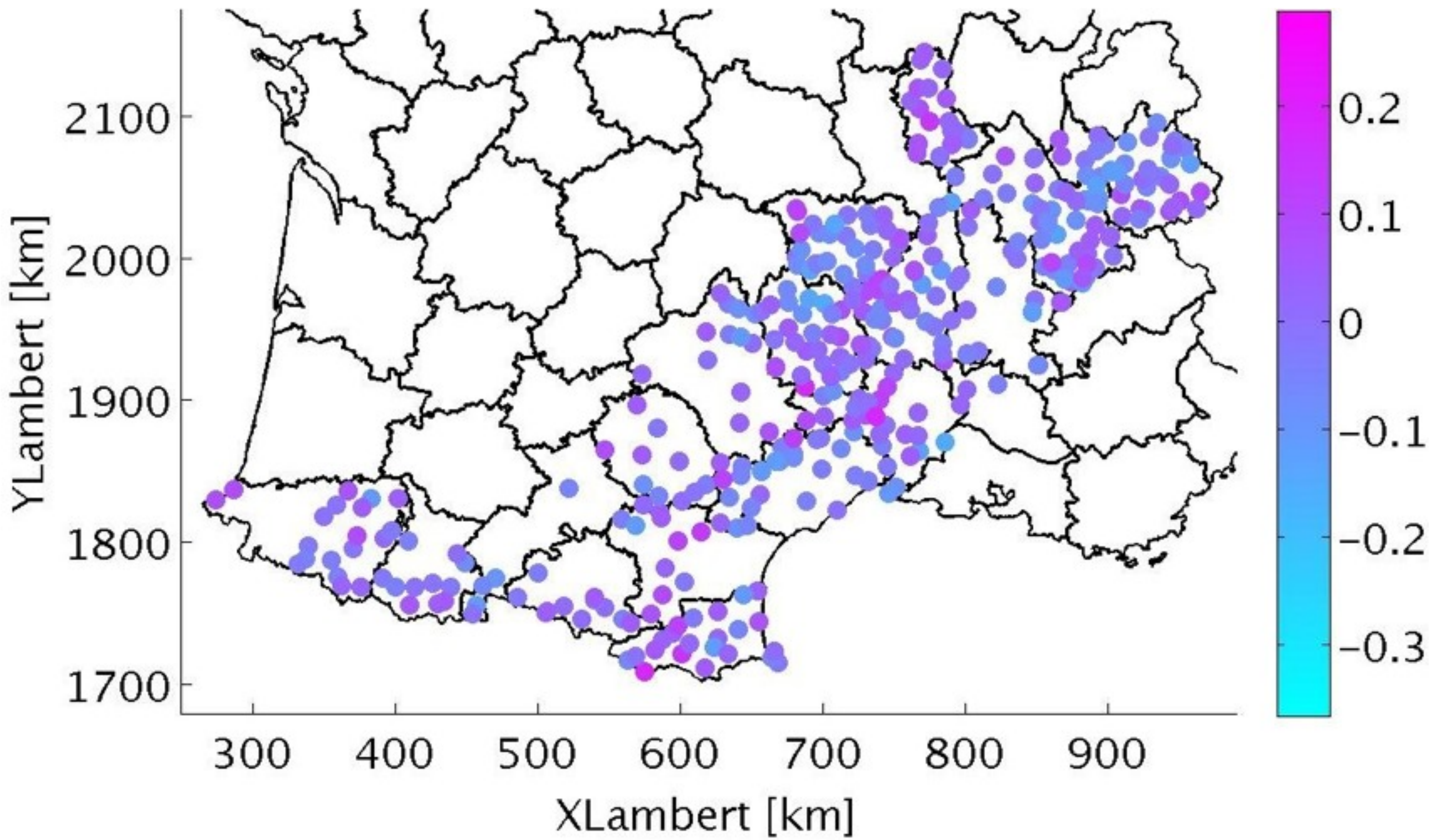
Over-estimation

Level 99 %



Spatial Organization

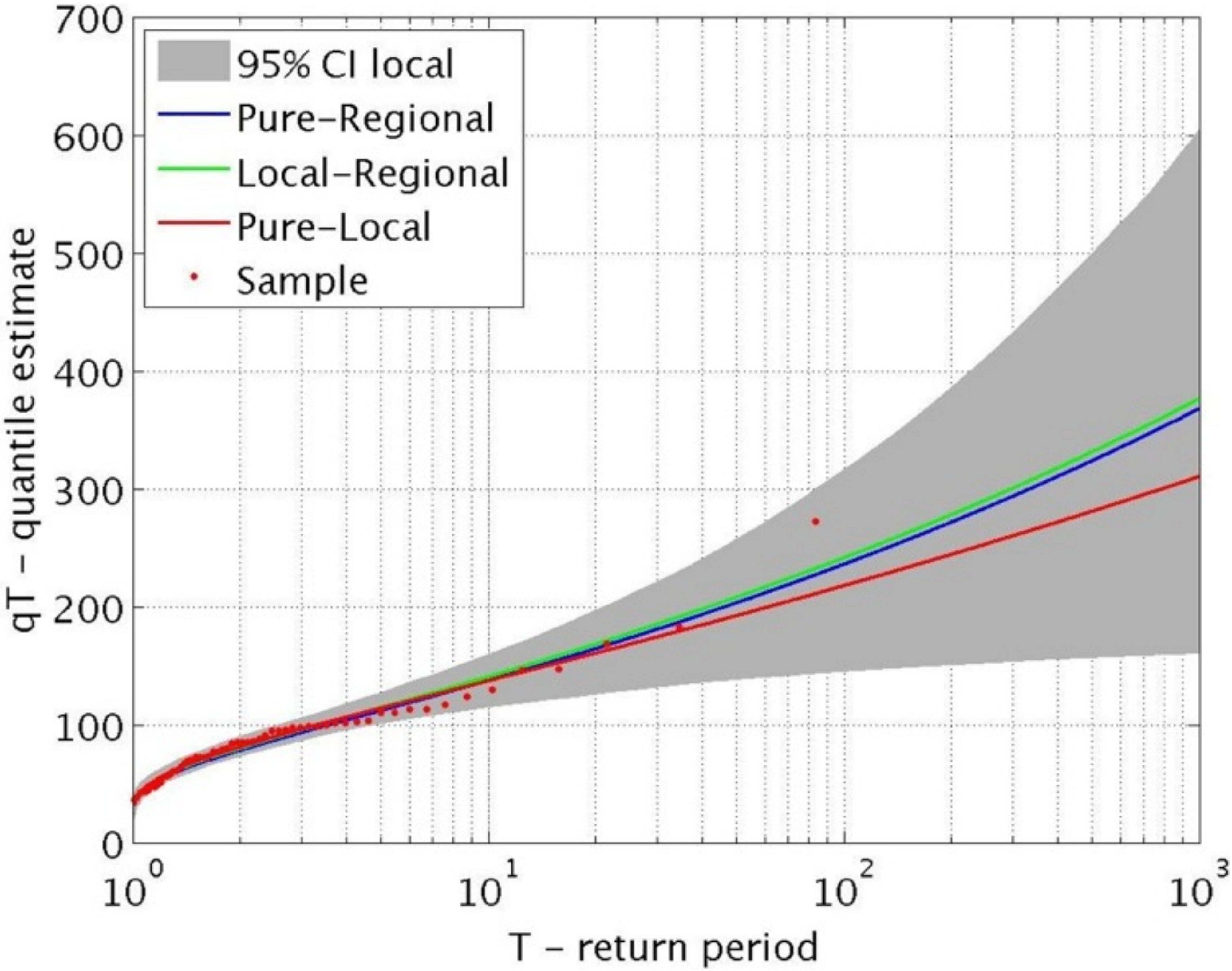
Relative Quantile Error Level 95 %



L-Moment Estimator Variance

- 1000 bootstrap samples
- Compute L-Moment local GEV estimates on each sample
- Get an empirical distribution for a range of return level
- Compute 95 % confidence interval for each return level

L-Moment Estimator Variance



Local Confidence Interval

Percentage of Regional Quantile estimates within the 95 % local confidence interval

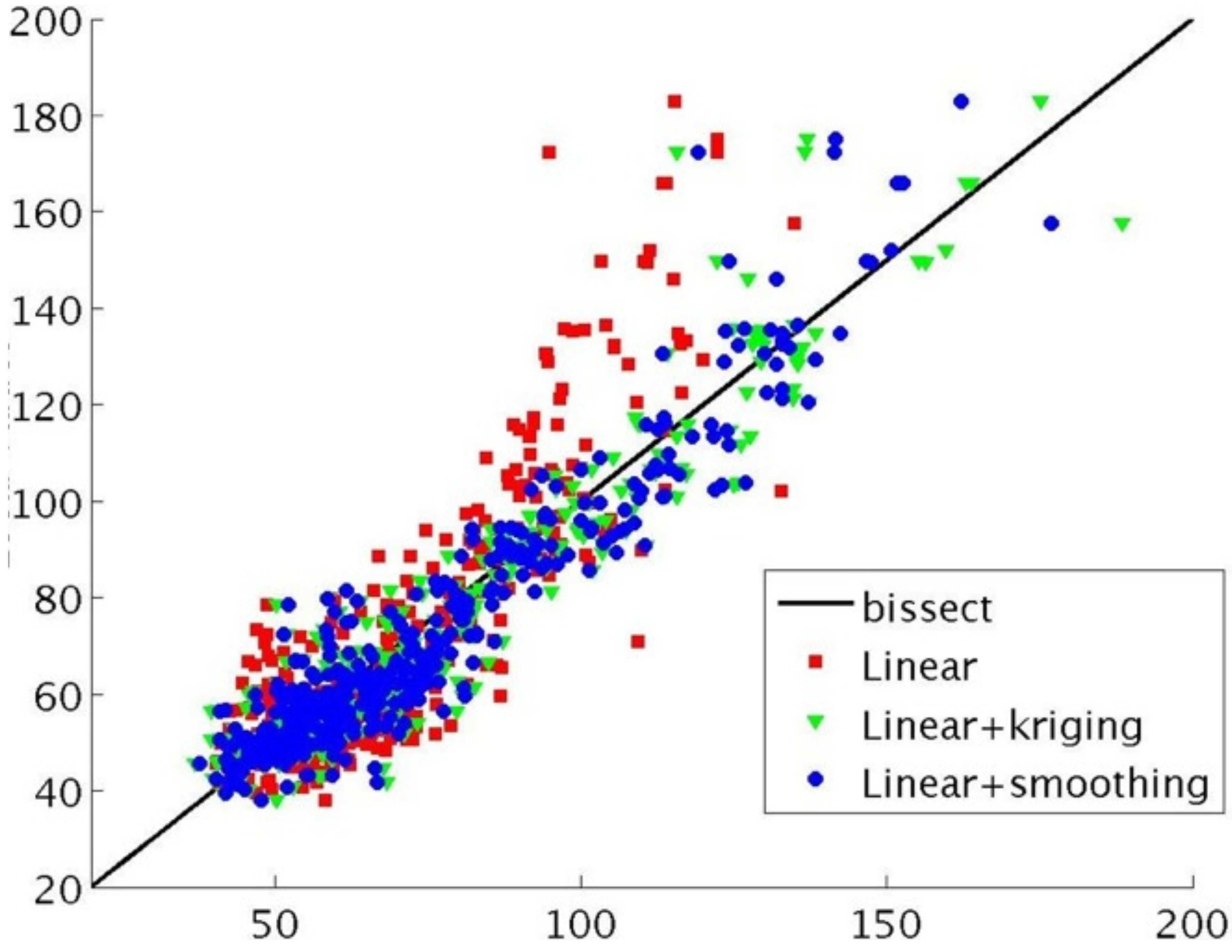
	90 %	95 %	98 %	99 %
Homogeneous	97	89	82	80
Fixed 50 km	92	84	78	77
100 sample	**98**	**92**	**87**	**85**
Random size	90	84	79	76
Random stations	87	81	78	76

Interpolation of Index Value

1. Linear regression
2. Linear regression **+ kernel smoothing**
3. Linear regression **+ kriging**
4. Multilayer perceptron (MLP) *or artificial neural networks == non-linear non-parametric regression*
5. MLP **+ kernel smoothing**
6. MLP **+ kriging**

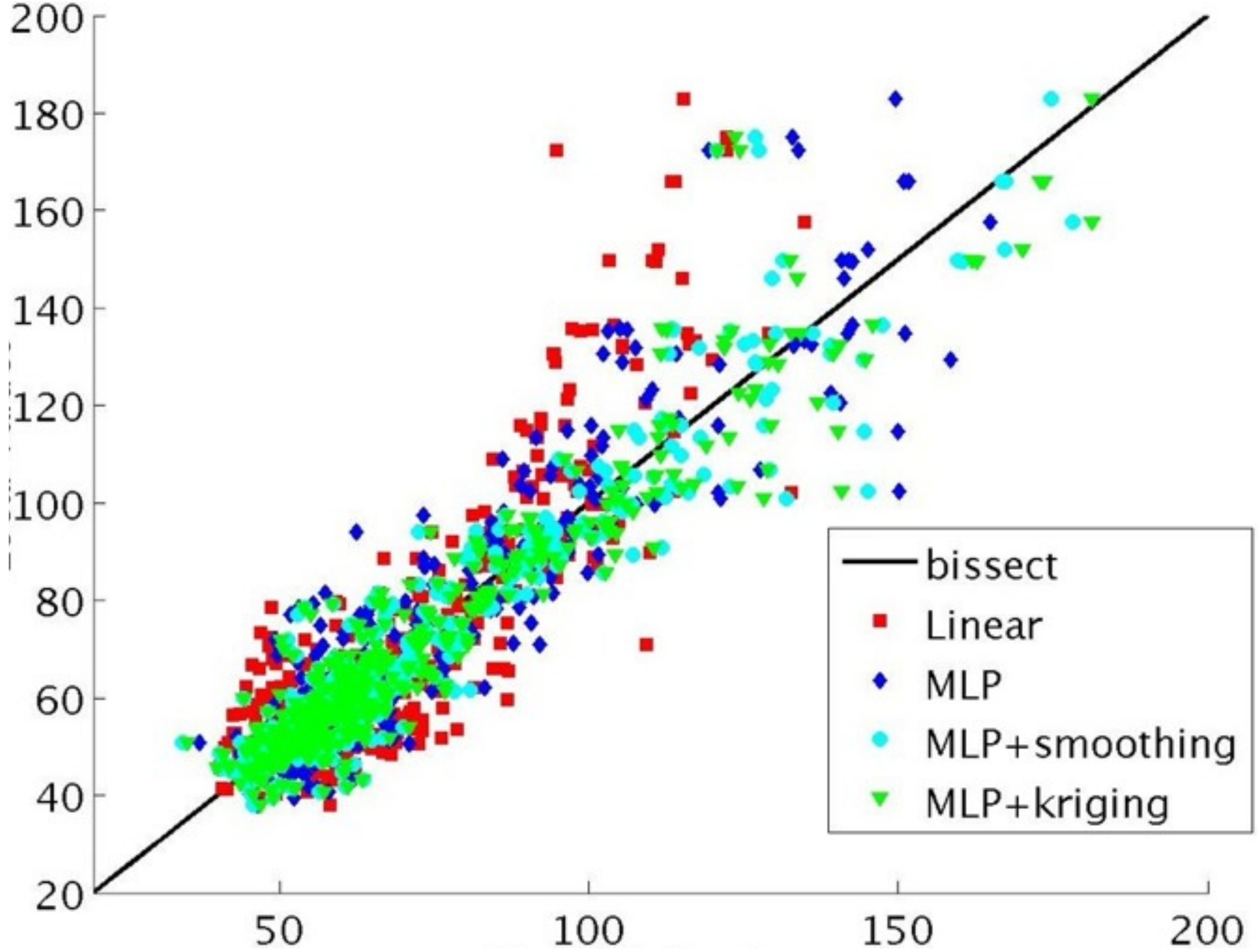
Comparison of Interpolators

Interpolator (X) vs Local Estimates (Y) on Validation Set



Comparison of Interpolators

Interpolator (X) vs Local Estimates (Y) on Validation Set



Comparison of Interpolators

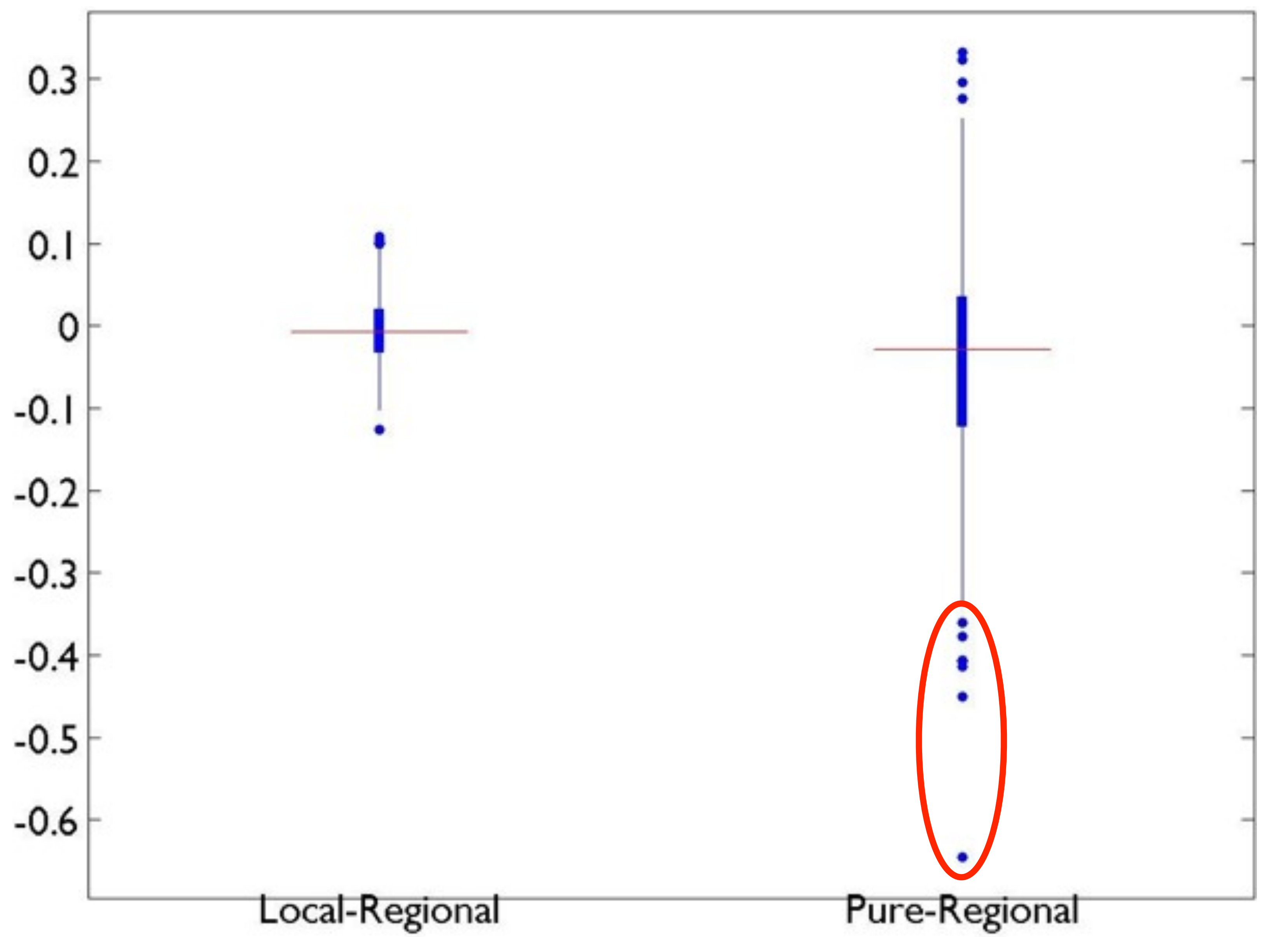
Interpolator (X) vs Local Estimates (Y) on Validation Set

RMSE		+ smoothing	+ kriging
Linear	14.62	**8.88**	9.01
MLP	10.55	9.15	9.05

Relative Quantile Error

$$1 - \frac{\hat{y}_p}{y_p^{\text{loc}}} \begin{cases} = 0 & \Leftrightarrow \hat{y}_p = y_p^{\text{loc}} \\ > 0 & \Leftrightarrow \hat{y}_p < y_p^{\text{loc}} \\ < 0 & \Leftrightarrow \hat{y}_p > y_p^{\text{loc}} \end{cases}$$

Level 90 %



Local Confidence Interval

Percentage of Regional Quantile estimates within the 95 % local confidence interval

	90 %	95 %	98 %	99 %
Local Regional	98	92	86	85
Pure Regional	**66**	**70**	76	75

Classification of Estimates

Local Confidence intervals

- Regional GEV estimate is good if

$$\hat{\xi}_R \in [\hat{\xi}_L^{0.025}, \hat{\xi}_L^{0.975}]$$

- Index Value interpolation is good if

$$\widehat{m}_{\text{interp}} \in [\widehat{m}_L^{0.025}, \widehat{m}_L^{0.975}]$$

Classification of Estimates

Local Confidence intervals

- Regional GEV estimate is good if

$$\hat{\xi}_R \in [\hat{\xi}_L^{0.025}, \hat{\xi}_L^{0.975}]$$

88 %

- Index Value interpolation is good if

$$\widehat{m}_{\mathrm{interp}} \in [\widehat{m}_L^{0.025}, \widehat{m}_L^{0.975}]$$

62 %

Classification of Estimates

Local Confidence intervals

- Regional GEV estimate is good if

$$\hat{\xi}_R \in [\hat{\xi}_L^{0.025}, \hat{\xi}_L^{0.975}]$$

88 %

- Index Value interpolation is good if

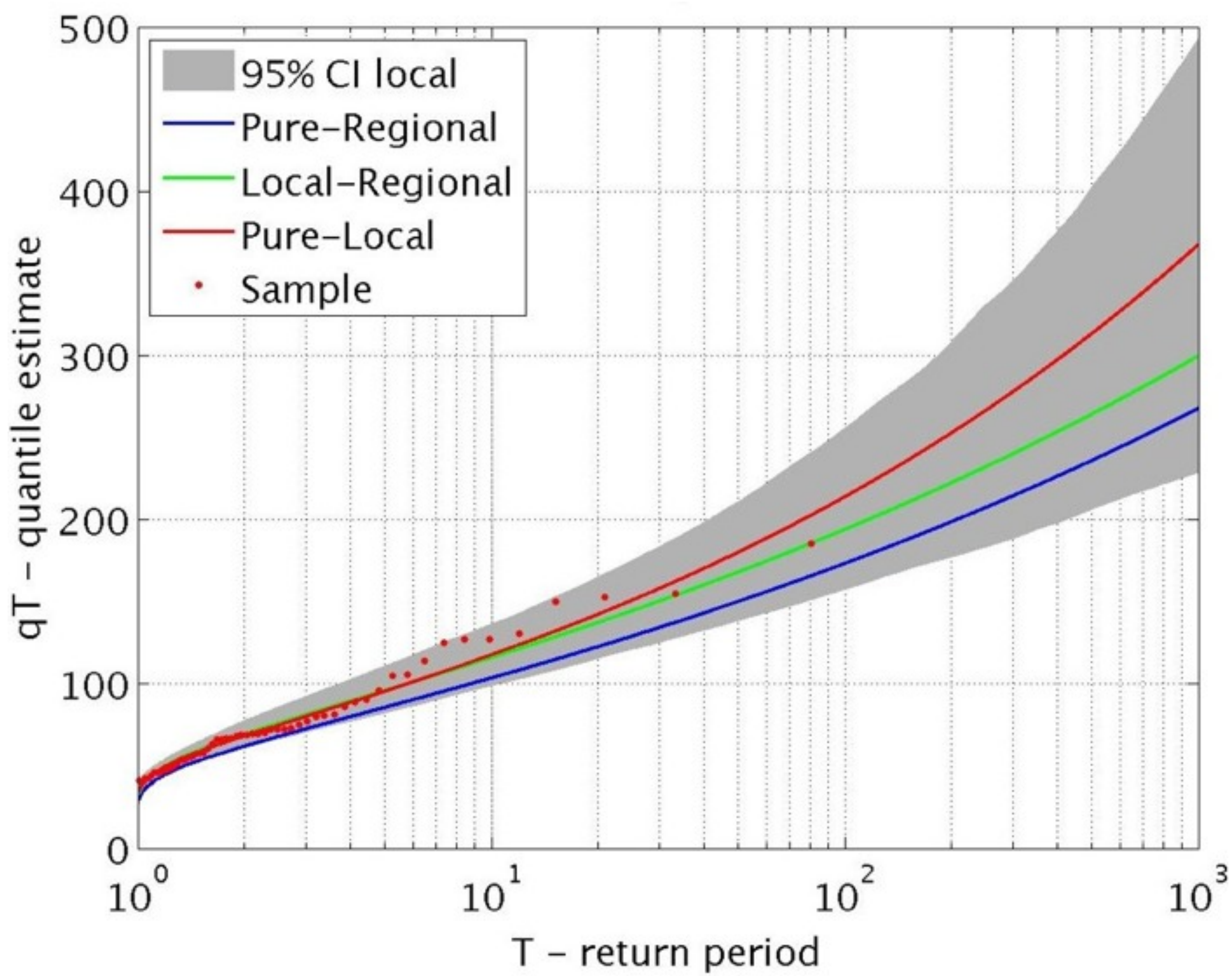
$$\widehat{m}_{\text{interp}} \in [\widehat{m}_L^{0.025}, \widehat{m}_L^{0.975}]$$

62 %

	90 %	95 %	98 %	99 %
Local Regional	98	92	86	**85**
Pure Regional	**66**	70	76	75

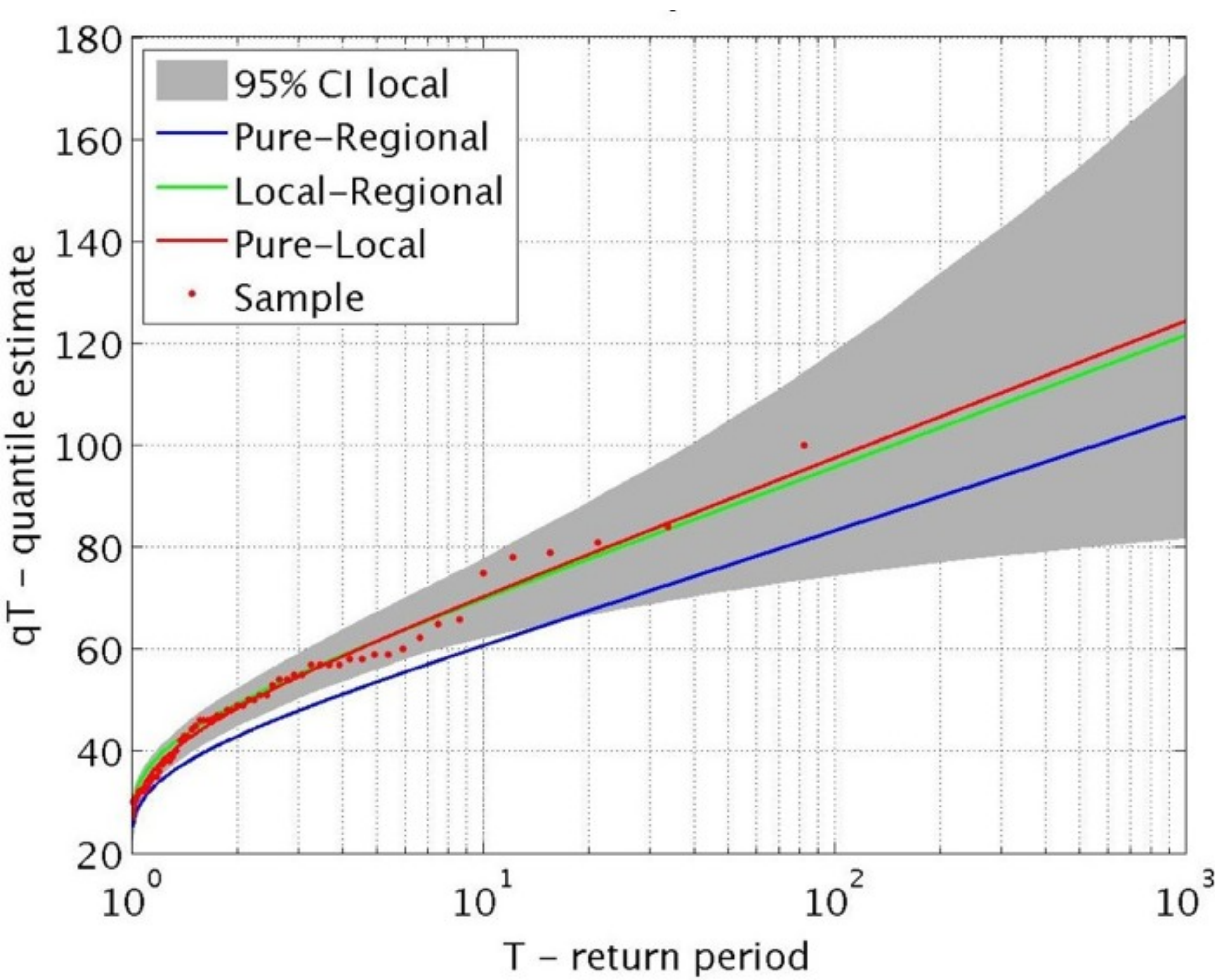
Classification Examples

Good Regional GEV and Interpolated Index Value



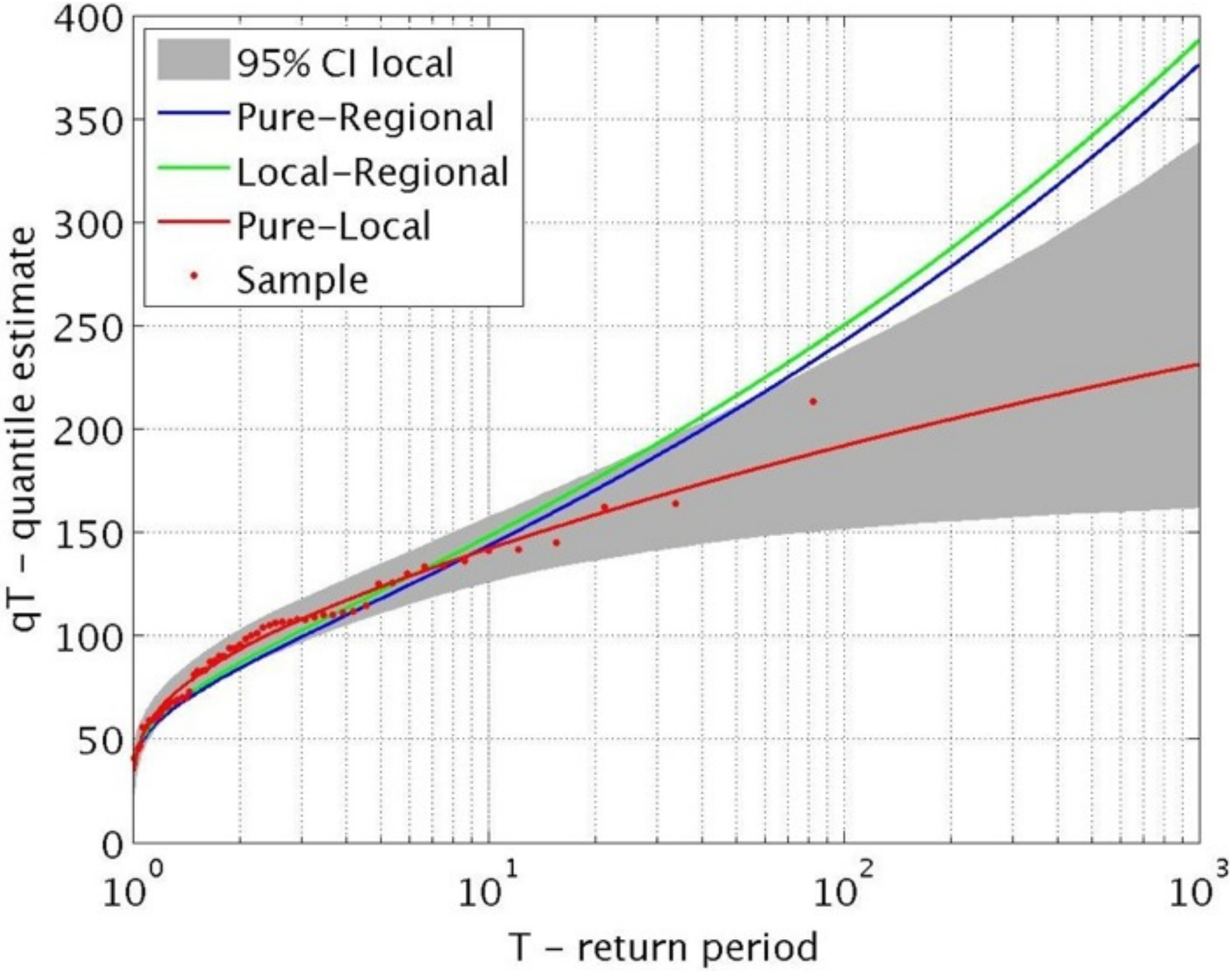
Classification Examples

Good Regional GEV and bad Interpolated Index Value



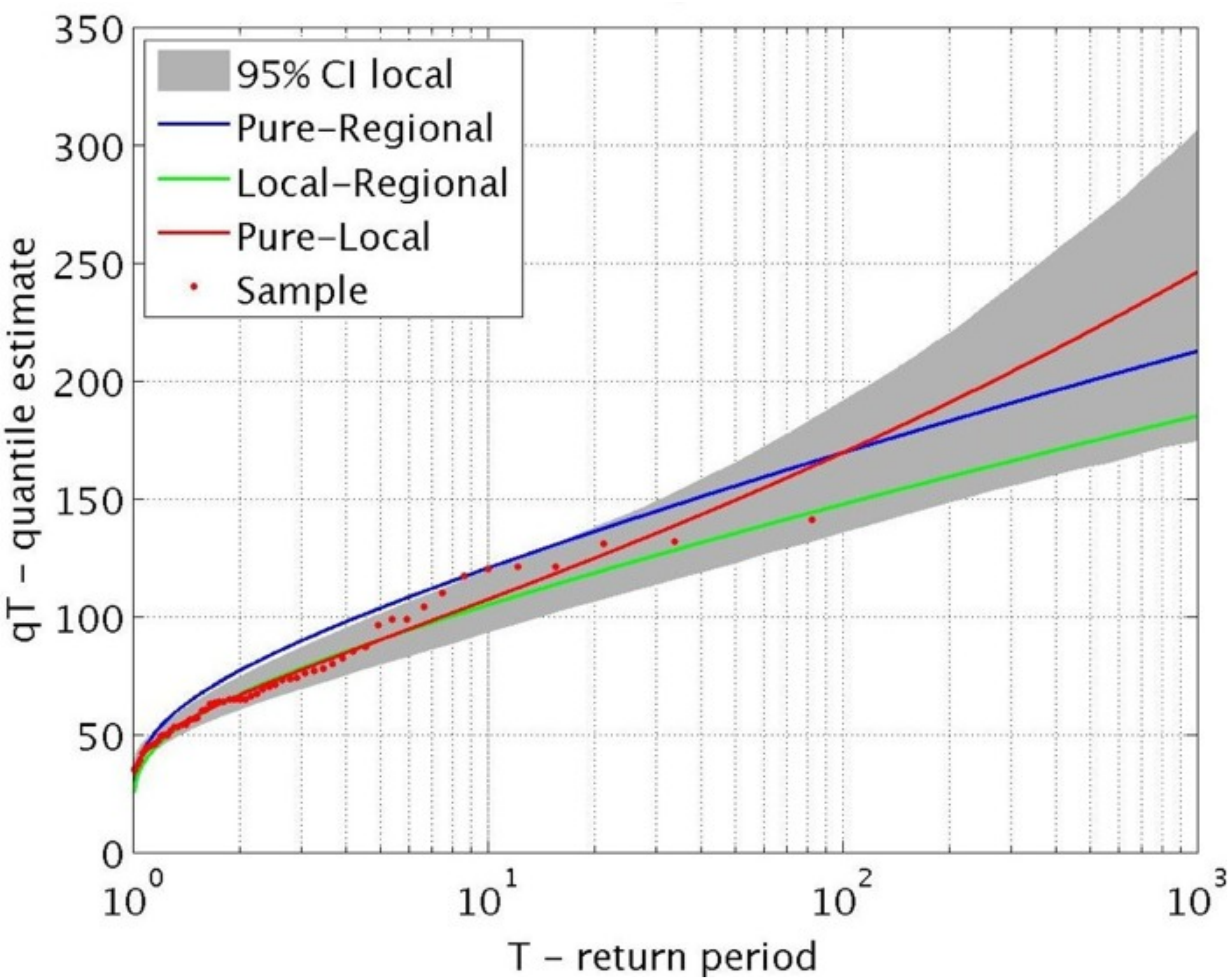
Classification Examples

Bad Regional GEV and good Index Value



Classification Examples

Bad Regional GEV and Index Value



Conclusion

1. **Homogeneous neighborhood :** reduce bias more outliers

2. **Power of homogeneity tests** might be too low

3. **100 - regional sample** is a good rule for regional neighborhood
   (depends on sampling)

4. No **spatial organization** of error

Conclusion

1. **Homogeneous neighborhood :** reduce bias more outliers

2. **Power of homogeneity tests** might be too low

3. **100 - regional sample** is a good rule for regional neighborhood *(depends on sampling)*

4. No **spatial organization** of error

5. **Variability of L-Moment** greater than uncertainty of neighborhood choice

6. **Interpolation of Index Value** affects first extreme quantile estimates (90% 95%)

7. **Focus on the estimation** of the index values, specially for lower quantiles

Thank you for your attention !