Small Area Estimation via Spatio-Temporal M-Quantile Modeling

Francesco Schirripa Spagnolo[1] Nicola Salvati[1] María Bugallo[2] Domingo Morales[2]

[1]Dipartimento di Economia e Management, Università di Pisa
francesco.schirripa@unipi.it
[2]Center of Operations Research, Miguel Hernández University of Elche

Motivation of work

- Introduces a new approach to capture both spatial and temporal dependencies in a robust and flexible way
- Classical spatio-temporal mixed models involve a large number of parameters, rely on strong distributional assumptions, and pose considerable computational challenges.
- Extend M-Quantile to *spatio-temporal* settings to borrow strength across areas and times approach that combine robustness, interpretability, and computational efficiency, capturing local spatio-temporal variations without relying on imposed random effects structures

Overview

- Introduction
- Data
- Methodology
  - M-quantile
  - Temporal M-quantile model
  - Spatio-Temporal M-Quantile (STMQ) model
- Simulation experiment
- Empirical application: Median AQI prediction using STMQ
- Final remarks

Introduction

Why AQI matters

- **Air Quality Index (AQI)** is a fundamental environmental and public health indicator
- The U.S. Environmental Protection Agency publishes the Air Quality Index (AQI) annually as a composite measure of several pollutants (such as $PM_{2.5}$, NO_2, SO_2 and O_3), which is widely used for policy decisions and health risk assessments
- **Decision need:** reliable prediction of AQI values at fine territorial levels (e.g., counties)
- **Challenge:** Sparse observations across *space* (counties) and *time* (years) $\Rightarrow$ classic Small Area Estimation (SAE) setting with out-of-sample counties/periods.
- **Goal:** Robust, flexible method that borrows strength across space & time to predict county AQI.

Why not just standard mixed models?

- Parametric spatio-temporal mixed models can be heavy (many parameters), sensitive to misspecification, and prone to boundary fits.
- Environmental data often contain outliers (events, fires, instrumentation), calling for robust methods.
- We need a flexible and robust alternative to spatio-temporal mixed models : **M-quantile regression**.

Data

Data descriptio

- The data derived from the annual summary reports of the AQI collected from outdoor monitoring stations across the U.S:
  - Years: $T = 8$ (2016–2023);
  - States: $D = 47$;
  - K_d the number of counties within state d, for $d = 1, \ldots, D$
    Counties total: 1014 (not all observed each year).
- Target: y_{dtk} = median AQI in county k of state d at year t.
- Auxiliaries (standardized) variables available at the county, state, and year levels: $\mathbf{x}'_{dtk} = (x_{dtk1}, \ldots, x_{dtkP})$
- AQI (or any environmental signal) shows temporal and spatial autocorrelation

Observed vs out-of-sample counties

- The set of observed subregions is defined as: $\mathcal{K}_d(t) = \{k \in \{1, \ldots, K_d\} : y_{dtk} \text{ is observed}\}$, with $K_{dt} = \text{card}(\mathcal{K}_d(t))$ and $\text{card}(\mathcal{B})$ denotes the cardinality of a set $\mathcal{B}$
- We predict both sampled and out-of-sample y_{dtk} using covariates and spatio-temporal borrowing of strength.
- Distances between county centroids via Haversine (curvature-aware) and then normalized for weights.

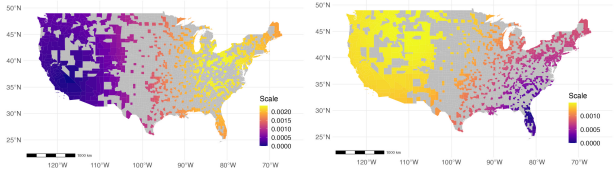

Figure 1: Studied region: Map of U.S. observed counties in 2023 by normalized distance to Los Angeles, California (left) and Palm Beach, Florida (right).

Methodology

Methodology

M-quantile

M-quantile: an alternative to QR

M-quantile (MQ) was proposed by Breckling and Chambers (1988) and it can be defined like a *'quantile-like 'generalization of regression based on influence functions (M-regression).*

In the linear case the M-quantile regression model is:

$$MQ_q(y_{dtk}; \sigma_q, \psi | \mathbf{x}_{dtk}) = \mathbf{x}'_{dtk}\boldsymbol{\beta}_\psi(q)$$

- ψ denotes an asymmetric influence function;
- $\boldsymbol{\beta}_{\psi q}$ is the vector of the regression parameters for each $q \in (0, 1)$
- Homoscedasticity: $\sigma_q = \text{var}^{1/2}(e_{\psi,dtk}(q)) = \sigma_\psi(\boldsymbol{\beta}_\psi(q))$

M-quantile: an alternative to QR (cont.)

Generally, ψ is the Huber proposal 2 influence function:

$$\psi(r) = r\, I(-c \leq r \leq c) + c\, \mathrm{sgn}(r) I(|r| > c),$$

c is a *tuning constant* bounded away from zero.

By setting the value of c it is possible to trade robustness for efficiency:

- $c \to 0$: ↑ Robustness ↓ Efficiency (we move towards QR)
- $c \to \infty$: ↑ Efficiency ↓ Robustness (we move towards ER)

A common value of c is $c = 1.345$: it produces 95% efficiency when the errors are normal, and still offers protection against outliers.

Estimation of the regression parameter $\beta_\psi(q)$ is obtained using iterative weighted least square (IWLS) algorithm.

MQ approach in SAE

- LMM: Residual between-group variability even after allowing for covariate $\rightarrow$ random effects specification where the area effects characterise the heterogeneity between groups.
- Chambers and Tzavidis (2006) characterise the variability across the population of interest by introducing the idea of MQ coefficients (q_{dtk})
- If the grouping structure explains part of the variability, after accounting for the effect of covariates, units within the same small area are expected to have similar MQ coefficients
- The small area-specific MQ coefficient, denoted by q_d, identifies the most characteristic MQ regression line for that small area and it is computed by averaging the M-quantile coefficients of the units that belong to that small area

Group heterogeneity through MQ

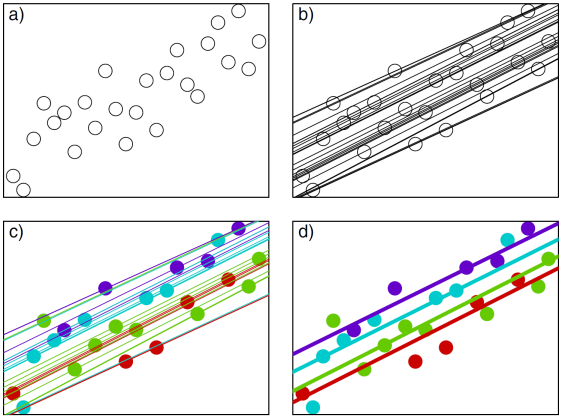

Figure 2: (a) raw data; (b) ensemble MQ regression model; (c) q_{ij} grouped by colour; (d) q_i calculated for each group, the fitted lines shown corresponding to the MQ fits with indices equal to these q_i

Methodology

Temporal M-quantile model

Time-Weighted M-Quantile models

- To capture temporal correlation, Bugallo et al. (2025) introduced the **Temporal Weighted M-Quantile (TWMQ)** models.
- TWMQ relaxes the independence assumption and introduces inter-period weights $\boldsymbol{w}_t$, linking adjacent years.

For each quantile level $0 < q < 1$, and for all regions $d = 1, \ldots, D$, times $t = 1, \ldots, T$, subregions $k = 1, \ldots, K_d$:

$$y_{dtk} = \boldsymbol{x}_{dtk}^{\top} \boldsymbol{\beta}_{\psi}(q, \boldsymbol{w}_t) + e_{\psi,dtk}(q, \boldsymbol{w}_t),$$

where

- $\boldsymbol{\beta}_{\psi}(q, \boldsymbol{w}_t)$: regression vector depending on order q and time weights $\boldsymbol{w}_t$,
- $\boldsymbol{w}_t = (w_{t1}, \ldots, w_{tT})'$: non-negative inter-period weights, reflecting correlation between periods,
- $e_{\psi,dtk}(q, \boldsymbol{w}_t)$: model errors such that $Q_q(e_{\psi,dtk} | \boldsymbol{x}_{dtk}) = 0$.

Inter-region variability through MQ coefficients

$$\hat{q}_{dtk} : \ \hat{Q}_q(y_{dtk}; \hat{\sigma}_q, \psi | \boldsymbol{x}_{dtk}) = y_{dtk}, \qquad \hat{\theta}_d = \frac{1}{TK_{d\cdot}} \sum_{t=1}^{T} \sum_{k=1}^{K_{dt}} \hat{q}_{dtk}.$$

- $\hat{\theta}_d$ summarizes the region's position in the conditional distribution—analogous to a random effect.
- Provides the region-specific quantile level used for prediction.

Constructing inter-period weights

Define a vector of asymmetric inter-period weights $\boldsymbol{w}_t = (w_{t1}, \ldots, w_{tT})'$ that captures close past dependencies. The idea is as follows:

- Compute MQ residuals

$$r_{\psi,dt} = \frac{1}{K_{dt}} \sum_{k=1}^{K_{dt}} \left(y_{dtk} - \boldsymbol{x}_{dtk}^{\top} \hat{\boldsymbol{\beta}}_{\psi}(\hat{\theta}_d) \right).$$

- Fit a seasonal autoregressive model $\mathrm{AR}(Q)_D$:

$$z_i = \phi_0 + \phi_1 z_{i-D} + \cdots + \phi_Q z_{i-QD} + a_i,$$

  where $\{a_i\}$ are independent $N(0, \sigma_a^2)$ errors and ϕ_j measures the "carry-over" of the past j years into the present

- Derive normalized weights for period t:

$$w_{ti} = \begin{cases} \dfrac{|\phi_{t+1-i}|}{S_t}, & \tilde{t} \le i \le t, \\ 0, & \text{otherwise}, \end{cases} \qquad S_t = \sum_{j=0}^{t} |\phi_j|, \quad \mathcal{T}_t = \{\tilde{t}, \ldots, t\}.$$

Estimation via weighted equations (IRLS)

The regression vector $\beta_\psi(q, \boldsymbol{w}_t)$ is obtained as the solution to the equations:

$$\sum_{d=1}^{D} \sum_{i \in \mathcal{T}_t} w_{ti} \sum_{k=1}^{K_{di}} \psi_q\left(y_{dik} - \boldsymbol{x}_{dik}^\top \beta_\psi(q, \boldsymbol{w}_t), \hat{\sigma}_{qt}\right) x_{dikp} = 0, \quad p = 1, \ldots, P.$$

- ψ_q is based on the Huber influence function.
- The IRLS algorithm iteratively updates residuals, weights, and coefficients until convergence.

Prediction under TWMQ

Plug-in predictor for region d, subregion k, time t

$$\hat{y}_{dtk}^{tmq} = \boldsymbol{x}_{dtk}^{\top}\hat{\boldsymbol{\beta}}_{\psi}(\hat{\theta}_d, \boldsymbol{w}_t).$$

If region d has no observed data over time

$$\hat{y}_{dtk}^{tmq} = \boldsymbol{x}_{dtk}^{\top}\hat{\boldsymbol{\beta}}_{\psi}(0.5, \boldsymbol{w}_t),$$

which corresponds to a median-based temporal M-quantile fit.

- TWMQ improves robustness and efficiency by capturing temporal dependence.
- It does not yet include spatial effects—this motivates the next step: **Spatio-Temporal M-Quantile (STMQ)** models.

Methodology

Spatio-Temporal M-Quantile (STMQ) model

From TWMQ to STMQ

- TWMQ captures temporal dependence but ignores spatial correlation.
- The **Spatio-Temporal M-Quantile (STMQ)** model integrates both:
  - Spatial proximity (through geographic distances),
  - Temporal dependence (through inter-period weights).
- Extends Salvati et al. (2012)'s Geographically Weighted M-Quantile (MQ-GWR) approach.

STMQ model formulation

For $0 < q < 1$, regions $d, g = 1, \ldots, D$, time periods $t, i = 1, \ldots, T$, and subregions $k = 1, \ldots, K_d$, $l = 1, \ldots, K_g$:

$$y_{gil} = \boldsymbol{x}_{gil}^{\top} \boldsymbol{\beta}_{\psi}(q, \boldsymbol{w}_{dtk}) + e_{\psi,gil}(q, \boldsymbol{w}_{dtk}),$$

where:

- $\boldsymbol{\beta}_{\psi}(q, \boldsymbol{w}_{dtk})$: regression coefficients depending on space-time weights,
- $\boldsymbol{w}_{dtk}$: vector of known nonnegative spatio-temporal weights,
- $Q_q(e_{\psi,gil}; \sigma_{qdtk}, \psi | \boldsymbol{x}_{gil}) = 0$ and $\sigma_{qdtk} = \mathrm{var}^{1/2}(e_{\psi,gil})$.

Estimation

The coefficients are estimated by solving:

$$\sum_{g=1}^{D}\sum_{i\in\mathcal{T}_t}\sum_{l=1}^{K_g} w_{dtk,gil}\,\psi_q\big(y_{gil}-\boldsymbol{x}_{gil}^{\top}\boldsymbol{\beta}_{\psi}(q,\boldsymbol{w}_{dtk}),\hat{\sigma}_{qdtk}\big)x_{gilp}=0,\quad p=1,\ldots,P,$$

where ψ is the Huber function.

- Solved via an IRLS algorithm adapted to spatio-temporal weighting.
- Fitted regression surface:

$$\hat{Q}_q(y_{dtk})=\boldsymbol{x}_{dtk}^{\top}\hat{\boldsymbol{\beta}}_{\psi}(q,\boldsymbol{w}_{dtk}).$$

Spatio-temporal weights: concept

Weights capture similarity between observations across both space and time:

$$w_{dtk,gil} = w_{dk,gl} \cdot w_{ti}.$$

Spatial component

$$w_{dk,gl}(b) = \exp\big(-b\, d^{HV}_{dk,gl}\big),$$

where $d^{HV}_{dk,gl}$ is the Haversine distance between counties k and l, and b is the bandwidth controlling decay (chosen by minimising cross-validation criteria: Smaller b: smoother surface (more global borrowing); Larger b: higher local sensitivity).

Temporal component

Weights w_{ti} obtained from AR$(Q)_D$ model on MQ residuals (as in TWMQ).

Modeling inter-region variability

- Region-, subregion-, and time-specific M-quantile coefficients:

$$q_{gil}(\boldsymbol{w}_{dtk}) = \text{solution of } Q_q(y_{gil}; \sigma_q(\boldsymbol{w}_{dtk}), \psi | \boldsymbol{x}_{gil}) = y_{gil}.$$

- Predicted coefficients:

$$\hat{q}_{gil}(\boldsymbol{w}_{dtk}) = \text{solution of } \hat{Q}_q(y_{gil}; \hat{\sigma}_q(\boldsymbol{w}_{dtk}), \psi | \boldsymbol{x}_{gil}) = y_{gil}.$$

- Regional averages:

$$\hat{\theta}_g(\boldsymbol{w}_{dtk}) = \frac{1}{TK_{g\cdot}} \sum_{i=1}^{T} \sum_{l=1}^{K_{gi}} \hat{q}_{gil}(\boldsymbol{w}_{dtk}).$$

Spatio-temporal prediction

- **Case (0):** observed subregion k in region d at time t:

$$\hat{y}^{stmq}_{dtk} = \boldsymbol{x}^{\top}_{dtk}\hat{\boldsymbol{\beta}}_{\psi}\left(\hat{\theta}_d(\boldsymbol{w}_{dtk}), \boldsymbol{w}_{dtk}\right).$$

- **Case (1):** previously observed same subregion — use past $\boldsymbol{w}_{dik}$.
- **Case (2):** not observed, use spatially nearest subregion m with available time i.
- **Case (3):** no data at all — use median-based model:

$$\hat{y}^{stmq}_{dtk} = \boldsymbol{x}^{\top}_{dtk}\hat{\boldsymbol{\beta}}_{\psi}(0.5).$$

Robust bias-corrected STMQ (BSTMQ)

To improve robustness, bias correction is added:

$$\hat{y}^{bstmq}_{dtk} = \hat{y}^{stmq}_{dtk} + \frac{\hat{\sigma}_{\hat{\theta}_d(\boldsymbol{w})}}{K_{(t)}} \sum_{g=1}^{D} \sum_{i \in \mathcal{T}_t} \sum_{l=1}^{K_{gi}} \phi\Big(\hat{\sigma}^{-1}_{\hat{\theta}_d(\boldsymbol{w})} \hat{\boldsymbol{e}}_{\psi,gil}(\hat{\theta}_d(\boldsymbol{w}), \boldsymbol{w}) \Big),$$

where ϕ is an influence function (e.g., Huber) with robustness constant c_ϕ.

- $K_{(t)}$: number of relevant observed data at time t.
- c_ϕ optimized by minimizing analytical MSE.

MSE estimation

- Following Chambers and Tzavidis (2006), a pseudo-linearization approach is used to derive a first-order approximation of

$$MSE(\hat{y}_{dtk}^{stmq}) = \text{var}\left(\hat{y}_{dtk}^{stmq} - y_{dtk}\right) + \text{bias}^2\left(\hat{y}_{dtk}^{stmq}\right).$$

- To estimate the variance two approaches are considered:
  1. *Median approach*: the variance estimators are based on the median model
  2. *Area quantile coefficient approach*: the variance estimators are calculated based on the representative probability of the region g to which the subregion belongs
- These give two first–order MSE estimators:

$$mse_{1,dtk}^{stmq}, \qquad mse_{2,dtk}^{stmq}.$$

- For the BSTMQ predictor, the corresponding "naive" MSE estimators include an additional robust correction term

Simulation experiment

Simulation design

- Purpose: evaluate predictive performance of STMQ and BSTMQ vs existing methods (LMM, MQ, TWMQ).
- Spatial layout: $D = 25$ regions in a 2×2 grid, each with $K = 4$ subregions, observed over $T = 8$ years.
- Model:

$$y_{dtk} = 1 + 5x_{dtk} + u_d + e_{dtk}, \quad e_{dtk} = \rho e_{d,t-1,k} + \eta_{dtk}, \quad \rho = 0.5.$$

- Spatial correlation: $\Sigma_{dtk,gtl} = \exp(-\chi^2 \|\boldsymbol{s}_{dk} - \boldsymbol{s}_{gl}\|^2)$, with $\chi \in \{0.5, 2, 4\}$.
- Outlier scenarios:
  1. $[0, 0]$ no outliers
  2. $[e, 0]$ unit-level outliers
  3. $[e, u]$ unit and area-level outliers
- Sampling: observed proportion $p \in \{0, 0.15, 0.3, 0.45\}$, 100 replications.

Performance metrics and predictors

- Predictors compared:

$$\{\hat{y}^{lmm}, \hat{y}^{mq}, \hat{y}^{tmq}, \hat{y}^{stmq}, \hat{y}^{bstmq}\}.$$

- Metrics:

$$\text{ARBIAS} = \frac{1}{DTK} \sum_{d,t,k} \left| \frac{\bar{\hat{y}}_{dtk} - \bar{y}_{dtk}}{\bar{y}_{dtk}} \right|, \quad \text{ARRMSE} = \frac{1}{DTK} \sum_{d,t,k} \frac{\sqrt{(\hat{y}_{dtk} - y_{dtk})^2}}{\bar{y}_{dtk}}.$$

- Relative improvement:

$$\uparrow \text{ARRMSE\%} = 100 \cdot \frac{\text{ARRMSE}(\hat{y}) - \text{ARRMSE}(\hat{y}^{bstmq})}{\text{ARRMSE}(\hat{y}^{bstmq})}.$$

Results: scenario [o, o]

		$\chi = 0.5$			$\chi = 2$			$\chi = 4$		
		ARBIAS	ARRMSE	↑ ARRMSE%	ARBIAS	ARRMSE	↑ ARRMSE%	ARBIAS	ARRMSE	↑ ARRMSE%
$p = 0$	$\hat{y}^{lmm}_{dtk}$	**1.054**	14.449	1.475	1.388	16.644	1.923	1.390	16.715	2.020
	$\hat{y}^{mq}_{dtk}$	1.238	15.562	9.291	1.398	16.708	2.315	1.400	16.753	2.252
	$\hat{y}^{tmq}_{dtk}$	1.124	14.977	5.183	1.387	16.593	1.611	1.389	16.639	1.556
	$\hat{y}^{stmq}_{dtk}$	1.060	14.264	0.176	1.366	16.335	0.031	**1.369**	16.388	0.024
	$\hat{y}^{bstmq}_{dtk}$	1.058	**14.239**		**1.365**	**16.330**		**1.369**	**16.384**	
$p = 0.15$	$\hat{y}^{lmm}_{dtk}$	**1.057**	14.507	1.590	1.387	16.689	1.837	1.394	16.754	1.922
	$\hat{y}^{mq}_{dtk}$	1.239	15.587	9.153	1.397	16.729	2.081	1.399	16.775	2.050
	$\hat{y}^{tmq}_{dtk}$	1.125	15.012	5.126	1.389	16.643	1.556	1.392	16.686	1.509
	$\hat{y}^{stmq}_{dtk}$	1.060	14.307	0.189	**1.367**	16.393	0.031	**1.371**	16.442	0.024
	$\hat{y}^{bstmq}_{dtk}$	1.058	**14.280**		**1.367**	**16.388**		**1.371**	**16.438**	

ARBIAS/ARRMSE (%) — Scenario [o,o]. Best values in bold.

Results: scenario [o, o] (cont.)

		$\chi = 0.5$			$\chi = 2$			$\chi = 4$		
		ARBIAS	ARRMSE	↑ ARRMSE%	ARBIAS	ARRMSE	↑ ARRMSE%	ARBIAS	ARRMSE	↑ ARRMSE%
$p = 0.3$	$\hat{y}^{lmm}_{dtk}$	**1.061**	14.561	1.598	1.391	16.735	1.658	1.397	16.808	1.817
	$\hat{y}^{mq}_{dtk}$	1.241	15.603	8.868	1.400	16.758	1.798	1.401	16.807	1.811
	$\hat{y}^{tmq}_{dtk}$	1.134	15.052	5.024	1.396	16.706	1.482	1.393	16.751	1.472
	$\hat{y}^{stmq}_{dtk}$	1.066	14.356	0.167	1.374	16.468	0.036	1.371	16.512	0.024
	$\hat{y}^{bstmq}_{dtk}$	1.065	**14.332**		**1.373**	**16.462**		**1.370**	**16.508**	
$p = 0.45$	$\hat{y}^{lmm}_{dtk}$	1.072	14.668	1.833	1.401	16.816	1.393	1.406	16.878	1.522
	$\hat{y}^{mq}_{dtk}$	1.240	15.637	8.560	1.410	16.815	1.387	1.408	16.854	1.377
	$\hat{y}^{tmq}_{dtk}$	1.130	15.097	4.811	1.415	16.823	1.435	1.402	16.844	1.317
	$\hat{y}^{stmq}_{dtk}$	1.060	14.426	0.153	1.399	16.590	0.030	1.386	16.628	0.018
	$\hat{y}^{bstmq}_{dtk}$	**1.059**	**14.404**		**1.398**	**16.585**		**1.385**	**16.625**	

ARBIAS/ARRMSE (%) — Scenario [o, o]. Best values in bold.

Results: scenario [e, o]

Relative increase in ARRMSE (%) vs BSTMQ — Scenario [e,o]

	$\chi = 0.5$				$\chi = 2$				$\chi = 4$			
	$\hat{y}^{lmm}$	$\hat{y}^{mq}$	$\hat{y}^{tmq}$	$\hat{y}^{stmq}$	$\hat{y}^{lmm}$	$\hat{y}^{mq}$	$\hat{y}^{tmq}$	$\hat{y}^{stmq}$	$\hat{y}^{lmm}$	$\hat{y}^{mq}$	$\hat{y}^{tmq}$	$\hat{y}^{stmq}$
$p = 0$	1.16	6.51	3.52	0.16	1.55	1.94	1.32	0.04	1.70	1.98	1.33	0.04
$p = 0.15$	1.34	6.38	3.49	0.15	1.52	1.76	1.29	0.05	1.60	1.78	1.28	0.04
$p = 0.30$	1.56	6.22	3.44	0.14	1.48	1.57	1.26	0.04	1.54	1.60	1.25	0.04
$p = 0.45$	1.81	5.84	3.26	0.12	1.26	1.22	1.17	0.03	1.49	1.51	1.22	0.03

Results: scenario $[e, u]$

Relative increase in ARRMSE (%) vs BSTMQ — Scenario [e,u]

	$\chi = 0.5$				$\chi = 2$				$\chi = 4$			
	$\hat{y}^{lmm}$	$\hat{y}^{mq}$	$\hat{y}^{tmq}$	$\hat{y}^{stmq}$	$\hat{y}^{lmm}$	$\hat{y}^{mq}$	$\hat{y}^{tmq}$	$\hat{y}^{stmq}$	$\hat{y}^{lmm}$	$\hat{y}^{mq}$	$\hat{y}^{tmq}$	$\hat{y}^{stmq}$
$p = 0$	1.40	7.86	4.28	0.29	1.02	3.14	2.27	0.13	1.01	3.12	2.36	0.13
$p = 0.15$	1.29	7.46	4.20	0.28	0.70	2.63	2.15	0.14	0.45	2.08	2.14	0.12
$p = 0.30$	1.30	7.10	4.18	0.32	0.64	2.18	2.04	0.13	0.47	2.09	2.13	0.12
$p = 0.45$	0.84	6.04	3.78	0.26	0.32	0.97	1.80	0.13	0.29	2.01	1.91	0.13

Performance of RMSE estimators

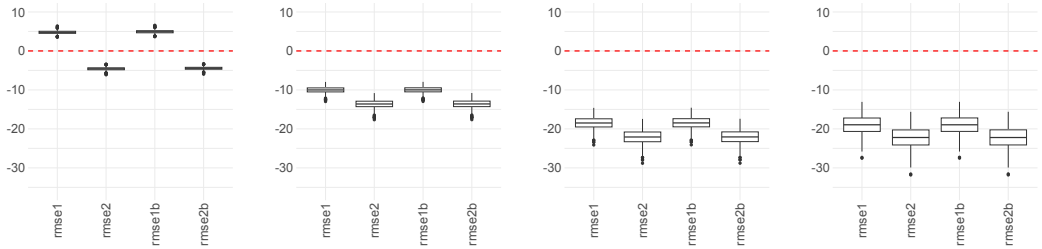

Figure 3: Relative Bias (RBIAS, %) for $p = 0,\ 0.15,\ 0.30,\ 0.45$ (left→right), $\rho = 0.5$, $\chi = 2$.

Performance of RMSE estimators

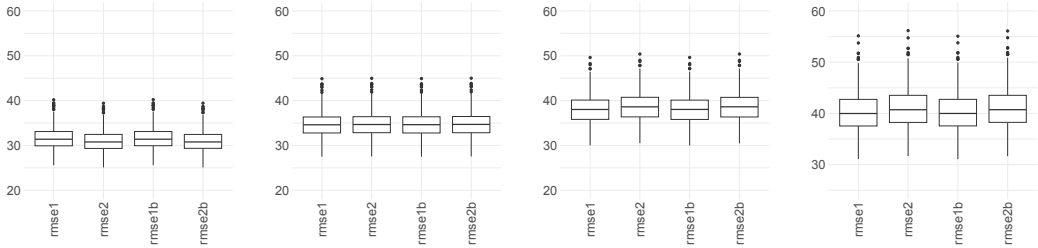

Figure 4: Relative RMSE (RRMSE, %) for $p = 0,\ 0.15,\ 0.30,\ 0.45$ (left→right), $\rho = 0.5$, $\chi = 2$.

Summary of simulation results

Performance of point predictors

- STMQ and BSTMQ predictors perform almost identically and show clear advantages under low sampling rates and strong spatial dependence.
- The BSTMQ predictor consistently achieves the lowest variability (lowest ARRMSE), confirming its robustness and stability.

Performance of proposed RMSE estimators

- The RRMSE demonstrates uniformly strong performance across all evaluated scenarios, indicating reliable predictive accuracy in practical applications.
- For both the STMQ and BSTMQ predictors, the most reliable RMSE estimator is the one based on the *median approach*.

Empirical application: Median AQI prediction using STMQ

Application setup & weight calculation

Goal: predict median AQI over counties (in/out-of-sample) with STMQ. **Spatio-temporal weights**

- Spatial weights: exponential decay with Haversine distance; bandwidth b via CV (grid $[0, 50]$).
- Temporal weights: seasonal $AR(Q)_D$ on MQ residuals $\Rightarrow$ here $Q = 2$ (current & previous year).

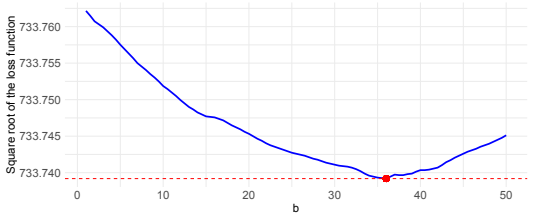

Figure 5: Square root loss for CV bandwidth selection; optimum $\hat{b} = 36.7$: moderate spatial smoothing.

STMQ model fitting — coefficients over time

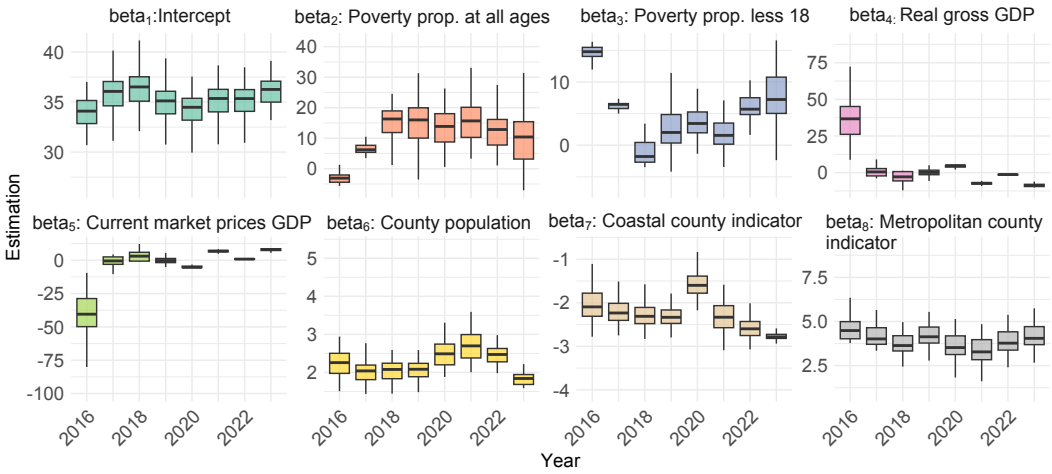

Figure 6: Boxplots of estimated regression coefficients by year (2016–2023).

Predictions over time — STMQ vs BSTMQ

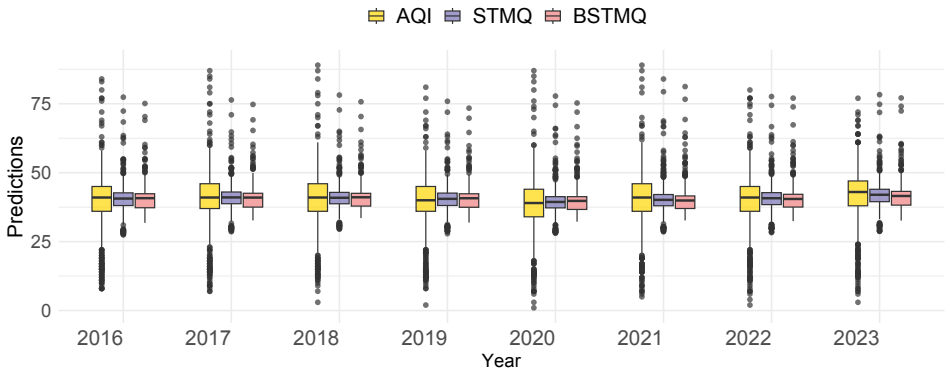

Figure 7: Observed vs plug-in STMQ and robust bias-corrected BSTMQ predictions (2016–2023).

Error measures (BSTMQ)

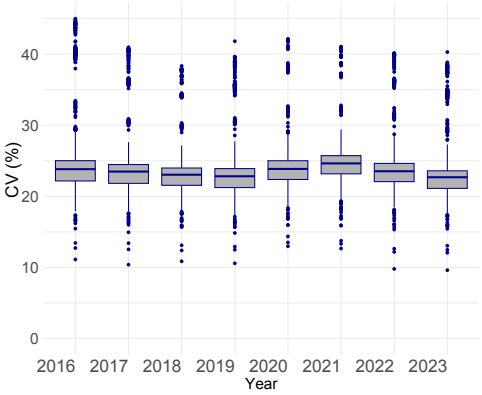

Figure 8: CV from MSE estimators by year.

Mapping predictions for 2023

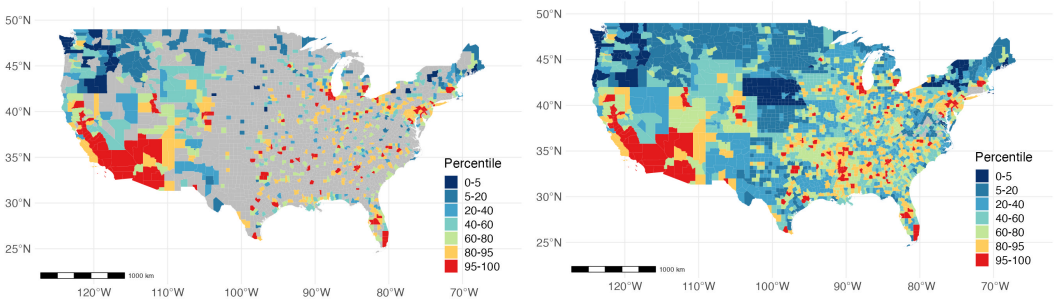

Figure 9: Predicted AQI percentiles in 2023 — observed counties (left) vs out-of-sample (right), BSTMQ.

Final remarks

Conclusions and future directions

Main contributions

- Proposed a new **Spatio-Temporal M-Quantile (STMQ)** framework for robust small area estimation.
- Unified spatial and temporal extensions of M-quantile regression within a single model.
- Developed analytical MSE estimators and robust bias-corrected predictors (BSTMQ).

Advantages over classical SAE or mixed models

- Robust to outliers at both unit and area levels.
- Fewer distributional assumptions, no need for random-effect specification.
- Flexible weighting: borrows information in space and time simultaneously.
- Computationally efficient and interpretable.

Limitations and future developments

Current limitations

- First-order MSE approximation — second-order unbiased estimators not yet derived.
- Fixed bandwidth and robustness parameters (b, c_ϕ); no adaptive selection.
- Only continuous outcomes modeled; discrete/categorical responses not yet integrated.
- Temporal weights limited to short memory.

Future research directions

- Data-driven and dynamic tuning of robustness parameters by region and year.
- Extension to non-linear, count, or categorical responses.
- Development of second-order MSE estimators.

References

Bianchi, A. and Salvati, N. (2015). Asymptotic properties and variance estimators of the M-quantile regression coefficients estimators. *Communications in Statistics – Theory and Methods*, **44**, 813–827.

Breckling, P. and Chambers, R. (1988). M-quantiles. *Biometrika*, **75**(4), 761–771.

Bugallo, M., Morales, D., Salvati, N., and Schirripa Spagnolo, F. (2025). Temporal M-quantile models and robust bias-corrected small area predictors. *Journal of Survey Statistics and Methodology*, forthcoming.

Chambers, R. and Tzavidis, N. (2006). M-quantile models for small area estimation. *Biometrika*, **93**(2), 255–268.

Chambers, R., Chandra, H., Salvati, N., and Tzavidis, N. (2014). Outlier robust small area estimation. *Journal of the Royal Statistical Society: Series B*, **76**(1), 47–69.

Fotheringham, A. S., Brunsdon, C., and Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley.

Holland, P. W. and Welsch, R. E. (1977). Robust regression using iteratively reweighted least squares. *Communications in Statistics – Theory and Methods*, **6**(9), 813–827.

Marchetti, S., Tzavidis, N., and Chambers, R. (2018). M-quantile models for multilevel data. *Statistical Methods and Applications*, **27**(4), 667–696.

Marhuenda, Y., Molina, I., and Morales, D. (2013). Small area estimation with spatio-temporal Fay–Herriot models. *Computational Statistics and Data Analysis*, **58**, 308–325.

Rao, J. N. K. and Molina, I. (2015). *Small Area Estimation*. Wiley, 2nd ed.

Salvati, N., Tzavidis, N., Pratesi, M., and Chambers, R. (2012). Small area estimation via M-quantile geographically weighted regression. *Computational Statistics and Data Analysis*, **56**, 288–300.

Francesco Schirripa Spagnolo

Dipartimento Economia e Management
Università di Pisa

francesco.schirripa@unipi.it

This work was supported by the project *PRIN – Quantification in the Context of Dataset Shift (QuaDaSh), Grant P2022TB5JF.*