SPARC: Single-Pass Scaling for Motion Forecasting with Conformal Bayesian Last Layers

Sakif Hossain*, Julian Teusch*, Jörg P. Müller

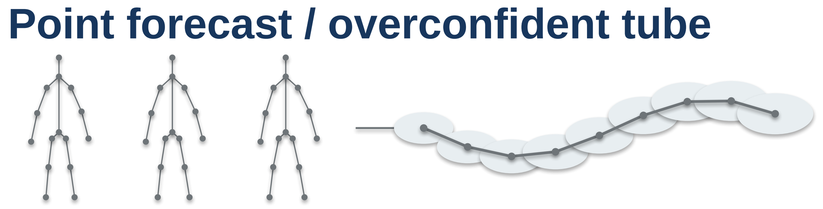
Clausthal University of Technology (TU Clausthal), Germany · * Equal contribution

ECCV 2026 · Malmö, Sweden · Paper 12880

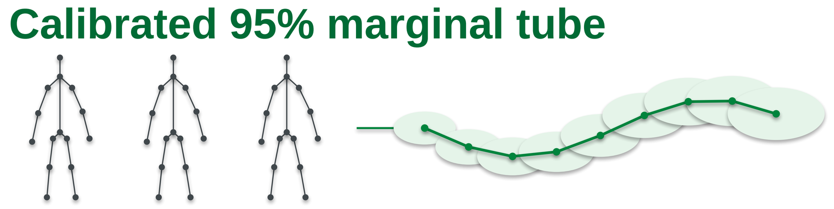
Why uncertainty?

Human motion forecasting predicts future 3D joint trajectories from a short observed prefix. A point forecast hides intrinsic ambiguity and failures outside familiar training support [2,11].

Point forecast / overconfident tube



Calibrated 95% marginal tube



Research question. Can a fast point forecaster become a structured, calibrated density without stochastic test-time sampling?

Why existing tools fall short

- **Repeated inference.** MC dropout and deep ensembles require stochastic passes [3,4].
- **Factorized outputs.** Analytic last-layer methods are efficient, but commonly use diagonal or isotropic residual models [5].
- **Calibration alone.** Conformal prediction gives marginal validity, but not an epistemic signal or trajectory covariance [8,9].

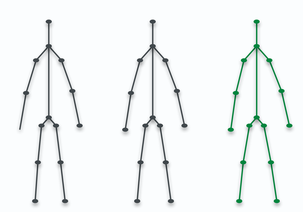
What SPARC contributes

1. **Exact factorization.** $\kappa_t(x)\Sigma_{str,t}(x)$ preserves the learned correlation structure.
2. **Unified recipe.** Bayesian leverage, graph-temporal covariance, and split CP in one layer.
3. **Broad evidence.** Nine dataset/protocol blocks with density, point, and tube metrics.
4. **Risk monitoring.** High-κ windows concentrate larger errors; thresholds remain held-out policies.

SPARC: structured scale + conformal calibration in one pass

Starting from the DCT/siMLPe backbone [1], SPARC keeps the deterministic mean unchanged and adds analytic Bayesian leverage, graph-temporal covariance, and post-hoc split conformal calibration.

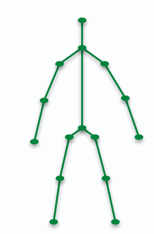
1 Observed prefix
3D poses up to t_0



$x_{1:T_{obs}}$

Short observed sequence supplied to the backbone

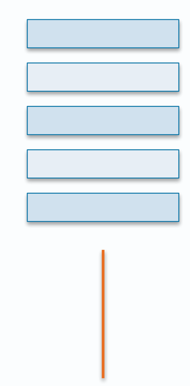
2 Deterministic mean
DCT/siMLPe [1]



$\hat{\mu}(x)$

Strong deterministic mean; no stochastic sampling

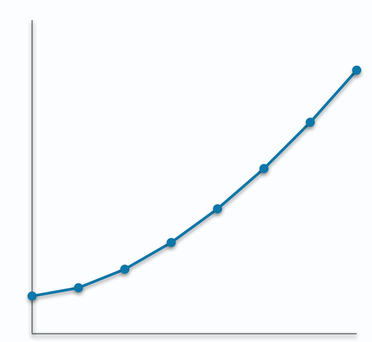
3 Bayesian last layer
conjugate posterior [5,6]



$B \mid \text{data}$

Conjugate update fitted offline on the training split

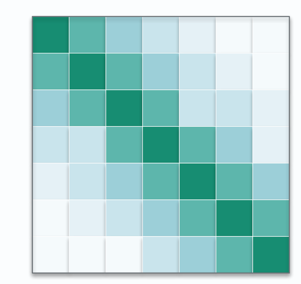
4 Epistemic scale
feature leverage



$\kappa_t(x) = 1 + \varphi_t \Lambda_{n,t}^{-1} \varphi_t$

Input-dependent, horizon-wise leverage signal

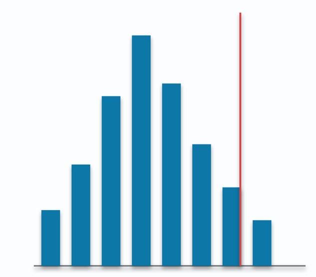
5 Structured covariance
matrix-normal + GMRF [6,7]



$\Sigma_{str,t}(x)$

Graph coupling + temporal smoothness preserve trajectory correlations

6 Split conformal
split CP [8,9]



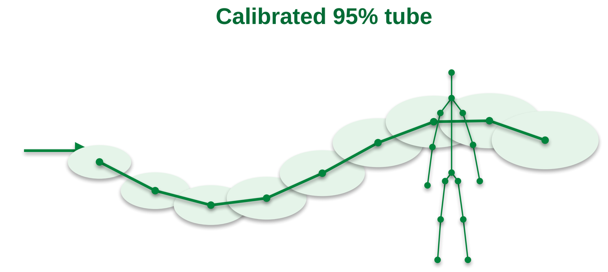
$q_{t,j}$

Standardized residual quantiles turn scales into 95% marginal tubes

Proposition 1: exact structured-covariance scaling

$$\kappa_t(x) \cdot \Sigma_{str,t}(x)$$

Under matrix-normal conjugacy, integrating the last-layer weights changes covariance magnitude, not correlations; the result holds for any positive-definite $\Sigma_{str,t}$ [6].



Offline fit + calibration

1. Train: fit the mean, Bayesian last layer, and graph-temporal covariance.
2. Calibrate: store standardized residual quantiles on a disjoint split.
3. Freeze all parameters before the final evaluation split.

Online: one forward pass

1. Predict the mean and backbone features.
2. Compute κ_t and scale the structured covariance.
3. Apply the stored quantile to obtain a 95% tube.

No sampling or test-time optimization.

Results across 9 dataset/protocol blocks

1.00	2.69	2.50	1.79×
NLL mean rank	MPJPE+NLL mean rank	W95(CP) mean rank	MPJPE in top-κ decile

SPARC is not the uniform point-error winner. It instead leads density quality and the combined point+density trade-off while retaining efficient, calibrated uncertainty tubes.

Evaluation protocol. H36M plus seven cross-domain datasets; typically T=50 and H=25. Train, tuning, calibration (n=512), and evaluation (n=1024) are disjoint; α=.05. Mean ranks normalize heterogeneous units.

NLL mean rank (↓ better)

SPARC	**1.00**
Bridge-CQR [12]	3.00
Deep ensemble [4]	5.56
MeanFT [1]	6.11
CoMusion [13]	6.44
SPARD [14]	6.56

MPJPE + NLL mean rank (↓ better)

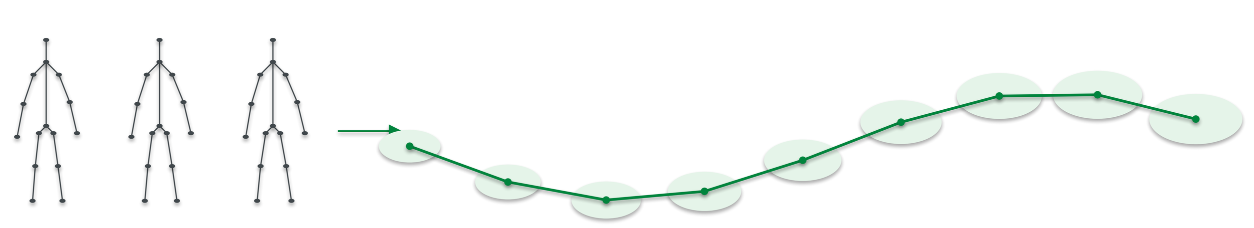
SPARC	**2.69**
Bridge-CQR [12]	3.81
MeanFT [1]	4.17
CoMusion [13]	6.25
SPARD [14]	6.83
Deep ensemble [4]	6.89

Ready for deployment

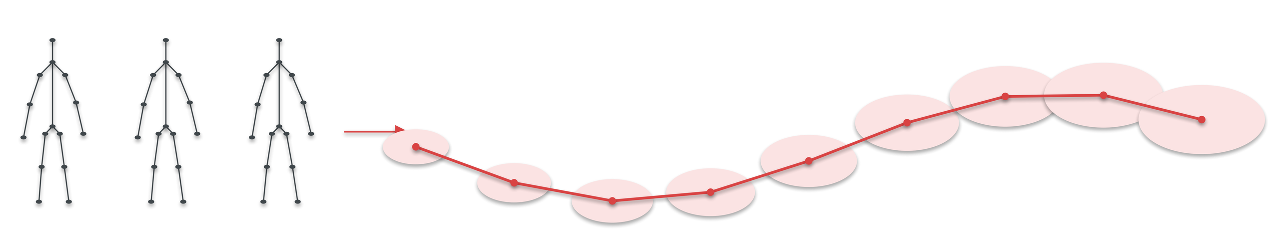
- **1 — One forward pass.** No MC or diffusion sampling loop.
- **95% — Calibrated tubes.** Finite-sample marginal validity [8,9].
- **κ — Risk monitor.** Top-decile filtering lowers MPJPE by ≈9%.

Qualitative intervals (3DPW)

Low-κ example: familiar feature region



High-κ example: weak feature support



Low/high κ marks familiar/weak feature support. The structured density expands before both cases are conformalized to the same nominal marginal coverage.

Guarantee

$$\Pr\{Y_{t,j} \in \hat{C}_{t,j}(X)\} \geq 1 - \alpha$$

For a fresh exchangeable sample and each joint j / horizon t, split conformal gives finite-sample marginal validity [8,9].

Calibration is post hoc and disjoint: no model or covariance parameter is updated.

Scope and limitations

- Marginal per joint/horizon, not simultaneous over a full trajectory.
- Strong conditional guarantees require extra assumptions [10].
- κ is representation-dependent: learned features determine leverage.
- Arbitrary shift can violate exchangeability [11].

Use κ as a held-out policy signal, not as a theorem-backed alarm.

Paper & contact

Paper · contact · · supplementary

sakif.hossain & julian.teusch@tu-clausthal.de

[1] Guo et al., WACV 2023 (siMLPe). [2] Kendall & Gal, NeurIPS 2017. [3] Gal & Ghahramani, ICML 2016. [4] Lakshminarayanan et al., NeurIPS 2017. [5] Daxberger et al., NeurIPS 2021. [6] Dawid, Biometrika 1981. [7] Rue & Held, Gaussian Markov Random Fields, 2005. [8] Vovk et al., Algorithmic Learning in a Random World, 2005. [9] Lei et al., JASA 2018. [10] Gibbs et al., 2023. [11] Ovadia et al., NeurIPS 2019. [12] Romano et al., NeurIPS 2019 (CQR). [13] Sun & Chowdhary, ECCV 2024 (CoMusion). [14] Zhang et al., AAAI 2026 (SPARD).