SPARC

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

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

ECCV 2026 Spotlight • Malmö, Sweden

*Equal contribution

One question: How can a fast motion forecaster know when its future should be trusted?

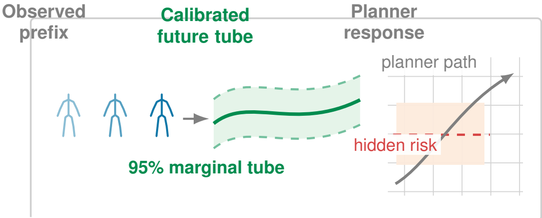


Fig. 1: A calibrated tube exposes risk that a point forecast hides.

DEPLOYMENT NEEDS ALL THREE

The Missing Combination

- **Single-pass:** no MC sampling or ensemble loop at test time [2, 5].
- **Structured:** uncertainty must couple future time steps and body joints.
- **Calibrated:** nominal 95% tubes should cover about 95% under the stated assumptions [6, 7].

Gap: No existing recipe gives all three at once.

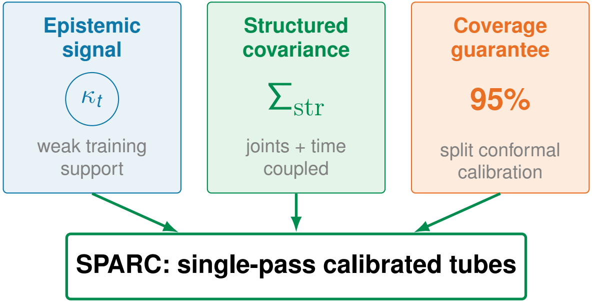
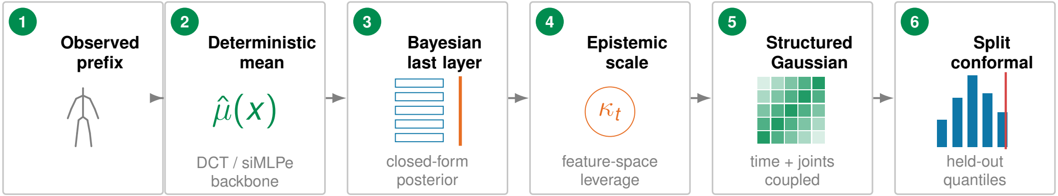


Fig. 2: SPARC targets epistemic sensitivity, structured uncertainty, and finite-sample calibration together.

THE COMPLETE PIPELINE

SPARC in One Forward Pass



$\kappa_t(x)\,\Sigma_{\mathrm{str},t}(x)$ **scales uncertainty while preserving correlation structure**

Fig. 3: A deterministic forecaster becomes a structured, calibrated density without stochastic test-time sampling.

Backbone mean [4]; closed-form last-layer posterior [1]; SPARC changes only the uncertainty layer.

TU Clausthal
Clausthal University of Technology

From Feature Leverage to Valid Tubes

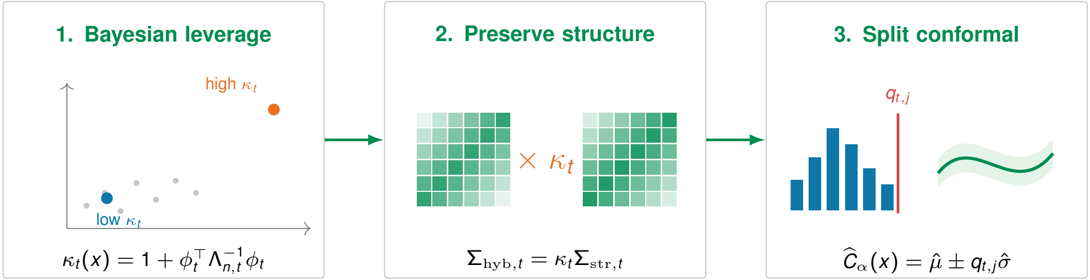


Fig. 4: Analytic leverage scales a structured covariance; calibration converts the scale into a target-coverage tube.

Guarantee: 95% finite-sample **marginal** coverage for exchangeable data [6, 7].

Scope: not conditional, trajectory-wise, or arbitrary-shift coverage [3].

TU Clausthal
Clausthal University of Technology

Evidence Across Nine Dataset/Protocol Blocks

1.00 NLL rank

2.69 MPJPE+NLL

2.50 W95(CP)

Model	NLL	MPJPE+NLL
MeanFT	6.11	4.17
Bridge-CQR	3.00	3.81
Deep ensemble	5.56	6.89
CoMusion	6.44	6.25
SPARD	6.56	6.83
SPARC	**1.00**	**2.69**

Table 1: Selected mean ranks; lower is better.

SPARC is not the uniform MPJPE winner. It ranks **first on NLL** and **first on the combined MPJPE+NLL criterion**.

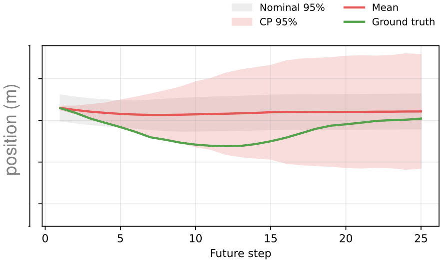


Fig. 5: On a representative 3DPW trajectory, conformal correction yields a tighter 95% tube than the nominal Gaussian band.

Interpretation. Coverage is useful only with informative width; SPARC retains efficient calibrated tubes in one forward pass.

DEPLOYMENT TAKEAWAY

The Epistemic Scale Becomes a Risk Signal

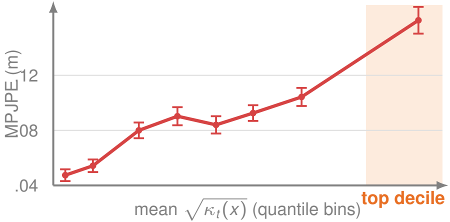


Fig. 6: Human3.6M MPJPE rises across bins of mean $\sqrt{\kappa_t(x)}$.

1.79×

MPJPE in the top-κ decile

- Filtering the top decile reduces MPJPE by about 9%.
- The signal is available after a **single forward pass**.
- Thresholds are held-out policy choices, not formal shift guarantees.

Takeaway: analytic κ_t + structured covariance + split CP = single-pass calibrated motion uncertainty. Full benchmark and ablations at our poster.

TU Clausthal
Clausthal University of Technology

Full Cross-Dataset Mean Ranks

Model	MR MPJPE	MR NLL	MR W95(CP)	MR MPJPE+NLL
siMLPe (base) [4]	7.89	10.67	–	9.28
MeanFT [4]	**2.22**	6.11	–	4.17
AuxTasks	9.06	11.78	–	10.42
Symplectic	16.22	15.78	–	16.00
DeFeeNet	7.61	10.44	–	9.03
GSPS	7.67	9.00	5.44	8.33
HumanMAC	10.22	10.00	–	10.11
SeSGCN	13.39	11.78	–	12.58
Deep ensemble [5]	8.22	5.56	5.39	6.89
Bridge-CQR	4.61	3.00	9.11	3.81
SkeletonDiffusion	15.00	13.67	9.78	14.33
BeLFusion	9.56	8.00	5.89	8.78
TransFusion	12.44	9.78	4.78	11.11
CoMusion	6.06	6.44	3.78	6.25
SPARD	7.11	6.56	3.39	6.83
MotionMap	16.11	16.22	9.44	16.17
SLD-HMP	13.22	15.22	6.50	14.22
SPARC	4.39	**1.00**	**2.50**	**2.69**

Table 2: Mean ranks across nine dataset/protocol blocks.

BACKUP · FROM HELD-OUT RESIDUALS TO VALID TUBES

Split Conformal Calibration

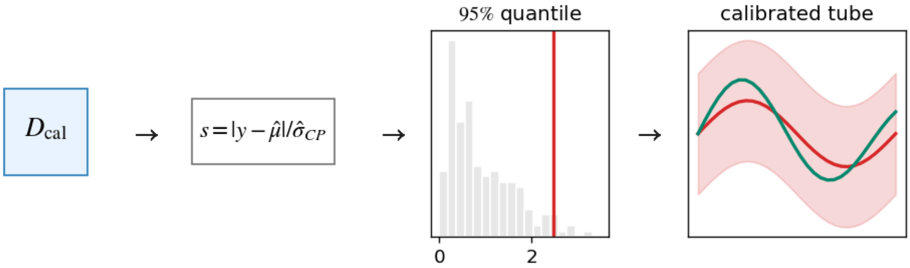


Fig. 7: Held-out residual scores determine a finite-sample quantile that expands the predictive scale into a calibrated tube.

Model frozen: parameters and last-layer statistics unchanged.

Validity: exchangeable held-out marginal scores.

TU Clausthal
Clausthal University of Technology

Component Ablation

Variant	κ_t	Hybrid	GraphJ	Coupled	MR MPJPE+NLL	MR W95(CP)
Base $\kappa(x)$	–	–	–	–	5.28	4.78
Timewise κ_t	✓	–	–	–	4.61	3.56
Hybrid MN	–	✓	–	–	3.94	4.00
Hybrid MN-GraphJ	–	✓	✓	–	**2.47**	4.78
Hybrid MN-GraphJ + κ_t	✓	✓	✓	–	2.53	3.78
SPARC coupled κ_t	✓	✓	✓	✓	3.17	**2.56**

Table 3: The coupled recipe gives the strongest calibrated tube-width rank within the ablation family.

Why this default? Hybrid MN-GraphJ favors MPJPE+NLL; coupled SPARC gives the best calibrated width rank within the complete deployment recipe.

References I

[1] E. Daxberger, A. Kristiadi, A. Immer, R. Eschenhagen, M. Bauer, and P. Hennig. Laplace redux — effortless bayesian deep learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 20089–20103, 2021.

[2] Y. Gal and Z. Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *Proceedings of the 33rd International Conference on Machine Learning (ICML)*, volume 48 of *Proceedings of Machine Learning Research*, pages 1050–1059. PMLR, 2016.

[3] I. Gibbs, J. J. Cherian, and E. J. Candès. Conformal prediction with conditional guarantees, 2023.

[4] W. Guo, Y. Du, X. Shen, V. Lepetit, X. Alameda-Pineda, and F. Moreno-Noguer. Back to MLP: A simple baseline for human motion prediction. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2023. doi: 10.48550/arXiv.2207.01567.

[5] B. Lakshminarayanan, A. Pritzel, and C. Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017. doi: 10.48550/arXiv.1612.01474.

References II

[6] J. Lei, M. G'Sell, A. Rinaldo, R. J. Tibshirani, and L. Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113(523):1094–1111, 2018. doi: 10.1080/01621459.2017.1307116.

[7] V. Vovk, A. Gammerman, and G. Shafer. *Algorithmic Learning in a Random World*. Springer, 2005. doi: 10.1007/b106715.