TU Clausthal
Clausthal University of Technology

SPARC

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

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

[*]Equal contribution

European Conference on Computer Vision (ECCV) 2026
Malmö, Sweden, September 8–12, 2026

Spotlight presentation

Motion Forecasting Needs More Than Points

- Input: a short observed 3D pose sequence [16, 30].
- Output: a plausible future trajectory.
- Safety-critical HRI needs predictive envelopes, not only means [21, 23, 37].
- Target: uncertainty that is efficient, calibrated, and structured.

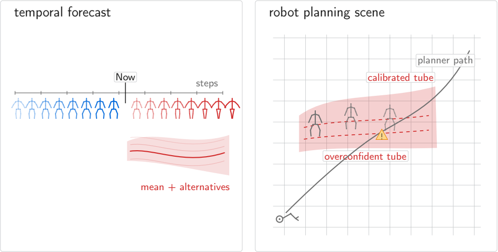

Fig. 1: Observed motion, forecast uncertainty, and a robot path motivate calibrated tubes.

PROBLEM

Three Requirements Rarely Meet

- **Epistemic:** detect weak training support [20].
- **Aleatoric:** represent genuine motion ambiguity [20].
- **Calibrated:** reach target coverage, e.g., 95% [1, 15].

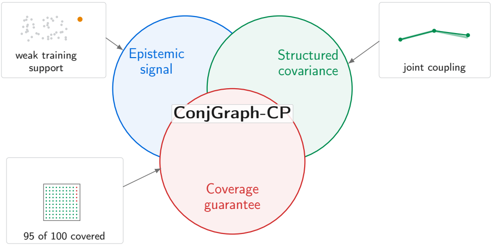

Fig. 2: The method combines epistemic signal, structured covariance, and conformal coverage.

RELATED WORK

Existing Methods Face a Deployment Trade-off

- MC dropout and ensembles often require repeated stochastic passes [13, 22].
- Generative forecasters use samples, hypotheses, or diffusion loops [2, 38, 47].
- Calibration-only methods provide coverage but no epistemic risk signal [1, 35].
- Factorized Gaussian heads lose temporal and skeletal dependence [3, 10, 36].

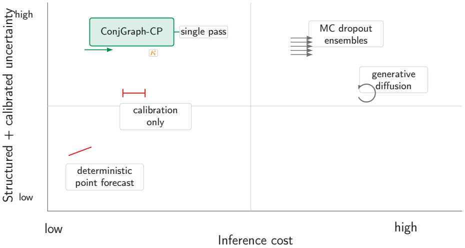

Fig. 3: Deployment methods trade inference cost against structured calibrated uncertainty.

ONE-SLIDE SUMMARY

SPARC as an Uncertainty Layer

1. Deterministic siMLPe/DCT backbone $\rightarrow$ mean forecast [16].
2. Bayesian last layer $\rightarrow$ $\kappa_t(x)$[11, 27].
3. Graph-temporal covariance $\rightarrow$ structured tubes [3, 10, 36].
4. Split conformal calibration $\rightarrow$ 95% marginal coverage [1, 26, 43].

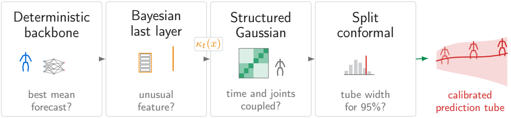

Fig. 4: SPARC wraps a deterministic forecaster with Bayesian scaling and split conformal calibration.

CONTRIBUTIONS

Paper Contributions

- **Structured Bayesian scaling:** $\kappa_t(x)\Sigma_{\mathrm{str},t}(x)$ preserves covariance structure.
- **Single-pass uncertainty:** no MC sampling and no ensembles.
- **Cross-dataset gains:** best mean NLL rank and best combined MPJPE+NLL rank.
- **Risk monitoring:** on H36M, the top-κ decile has $1.79\times$ average MPJPE.

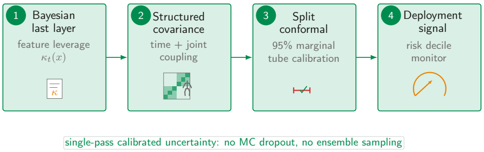

Fig. 5: The contribution chain turns feature leverage into single-pass calibrated uncertainty.

Task and Notation

- Observed prefix: $x_{1:T} \in \mathbb{R}^{T \times C}$.
- Future trajectory: $y_{1:H} \in \mathbb{R}^{H \times C}$.
- $C = 3J$ coordinates for J joints.
- Training uses displacements relative to the last observed pose x_T.

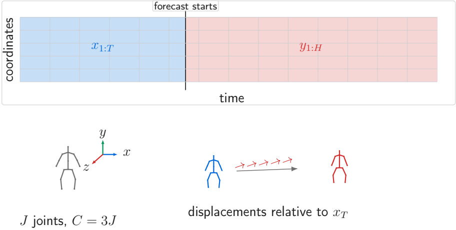

Fig. 6: Observed prefixes and future targets are represented as time-by-joint trajectory blocks.

CALIBRATED MARGINAL INTERVALS

Desired Output: Prediction Tubes

$$\widehat{C}_\alpha(x) = \prod_{t,j,d} \left[\hat{\mu}_{t,j,d}(x) \pm q_{t,j}\hat{\sigma}^{\mathrm{CP}}_{t,j,d}(x)\right].$$

- $q_{t,j}$ is a split-conformal factor [26, 43].
- The paper uses $\alpha = 0.05$ for 95% coverage.
- The guarantee is marginal, not automatically trajectory-level [12, 25].

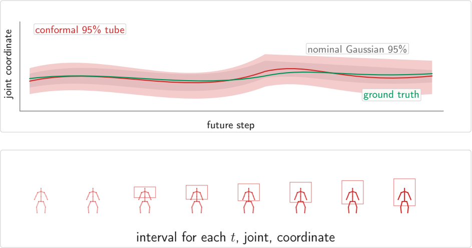

Fig. 7: Prediction tubes wrap the mean forecast with conformal marginal intervals.

DETERMINISTIC MEAN FORECASTER

Backbone: siMLPe/DCT

- The time axis is represented by DCT coefficients [16].
- An MLP processes coefficient-domain features.
- iDCT maps outputs back to the time domain.
- MeanFT provides a strong deterministic forecast.

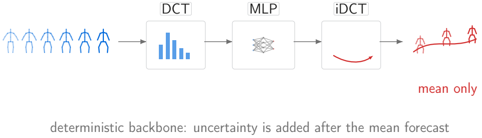

Fig. 8: The siMLPe-style DCT backbone maps pose histories to a deterministic future mean.

Time-Domain Design Vector

$$\hat{\mu}_t(x) = B_t^\top \phi_t(x), \qquad \phi_t(x) = \begin{bmatrix} g_t(x) \\ s_t \end{bmatrix}.$$

$$g_t(x) = \sum_{n=1}^{N} A_{t,n} h_n(x), \qquad s_t = \sum_{n=1}^{N} A_{t,n}.$$

- One design vector per forecast horizon step.
- The leverage of $\phi_t(x)$ later determines $\kappa_t(x)$.

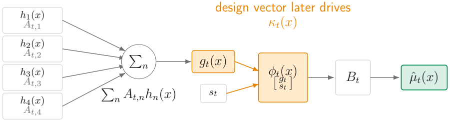

Fig. 9: DCT-domain features are aggregated into time-specific design vectors.

Epistemic Uncertainty as Feature Leverage

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

- Low leverage: test features lie in a dense training region.
- High leverage: weak feature-space training support.
- This is representation-dependent, not a universal OOD detector [33].

Bayesian last-layer background: [11, 27, 34, 44].

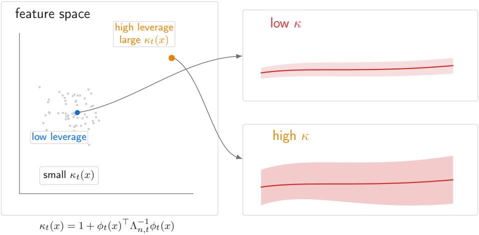

Fig. 10: Feature leverage separates dense training support from higher epistemic risk regions.

EXACT STRUCTURED-COVARIANCE SCALING

Proposition 1: Core Formula

Model:

$$y_t = B_t^\top \phi_t(x) + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \Sigma_{\mathrm{str},t}(x)).$$

Matrix-normal prior [10]:

$$B_t \mid \Sigma_{\mathrm{str},t} \sim \mathrm{MN}(B_{0,t}, \Lambda_0^{-1}, \Sigma_{\mathrm{str},t}).$$

Result:

$$y_t \mid x, D \sim \mathcal{N}(\mu_{\mathrm{post},t}(x), \kappa_t(x)\Sigma_{\mathrm{str},t}(x)).$$

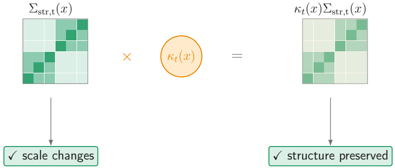

Fig. 11: Bayesian last-layer scaling enlarges covariance while preserving its structure.

Scale Epistemics, Preserve Structure

Weak alternative

- Diagonal or isotropic uncertainty treats coordinates as largely independent.

SPARC

$$\Sigma_{\mathrm{hyb},t}(x) = \kappa_t(x)\,\Sigma_{\mathrm{str},t}(x).$$

- Scale increases; correlations are preserved.

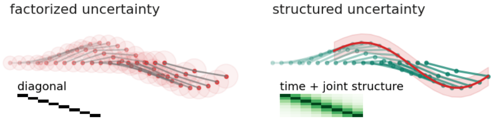

Fig. 12: Structured scaling preserves coherent skeletal uncertainty compared with diagonal inflation.

Offline Fit and Single-Pass Inference

Offline

$$S_{xx,t} = \sum_i \phi_t(x_i)\phi_t(x_i)^\top.$$

$$\Lambda_{n,t}^{-1} = (S_{xx,t} + \lambda_0 I)^{-1}.$$

Inference

- one forward pass,
- a quadratic form for $\kappa_t(x)$,
- no MC sampling and no ensembles.

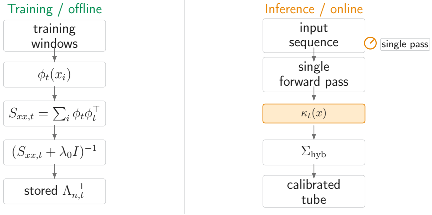

Fig. 13: Offline statistics enable one-pass online inference without sampling loops.

Stabilizing κ

$$\bar{\kappa}(x) = \frac{1}{H}\sum_{t=1}^{H}\left(1 + \phi_t(x)^\top \Lambda_{\mathrm{glob}}^{-1}\phi_t(x)\right).$$

$$\tilde{\kappa}_t(x) = (1-\rho)\kappa_t(x) + \rho\bar{\kappa}(x).$$

- Shrinkage smooths horizon-wise scales.
- Per-joint $\kappa_j(x)$ scales CP tubes, not the Gaussian NLL core.

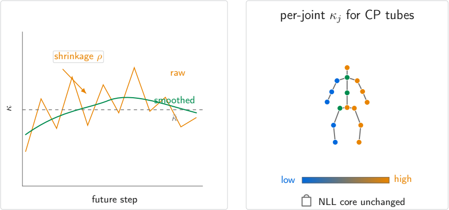

Fig. 14: Shrinkage stabilizes horizon-wise leverage while keeping the Gaussian NLL core unchanged.

Structured Aleatoric Covariance

Matrix-normal residual model [10]

$$R(x) = y_{1:H} - \hat{\mu}_{1:H}(x),$$
$$R(x) \sim \mathrm{MN}(0, \Sigma_T(x), \Sigma_C(x)).$$

Joint-graph precision [3, 36]

$$Q_J(x) = \tau(x) L_{\text{joint}} + \epsilon(x) I_J,$$
$$\Sigma_C(x) = (Q_J(x) \otimes I_3)^{-1}.$$

Σ_T captures time; Σ_C captures skeletal coupling.

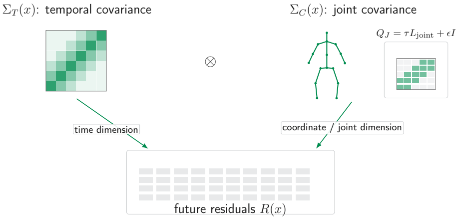

Fig. 15: MN-GraphJ combines temporal covariance with skeletal joint coupling.

TU Clausthal
Clausthal University of Technology

Hybridization

$$D_\kappa(x) = \mathrm{diag}\left(\sqrt{\tilde{\kappa}_t(x)}\right)_{t=1}^{H}.$$

$$\Sigma_T^{\mathrm{hyb}}(x) = D_\kappa(x)\Sigma_T(x)D_\kappa(x)^\top.$$

$$\Sigma^{\mathrm{hyb}}(x) = \Sigma_T^{\mathrm{hyb}}(x) \otimes \Sigma_C(x).$$

- Correlations remain intact.
- Variance increases where $\tilde{\kappa}_t$ is high.

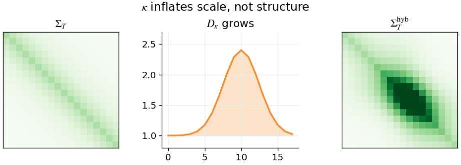

Fig. 16: Hybridization inflates temporal variance according to κ while retaining correlations.

Split Conformal Calibration

Calibration scores [1, 26, 43]:

$$s_{i,t,j,d} = \frac{|y_{i,t,j,d} - \hat{\mu}_{t,j,d}(x_i)|}{\hat{\sigma}^{\mathrm{CP}}_{t,j,d}(x_i)}.$$

Quantile:

$$q_{t,j} = Q_{1-\alpha}\left(\{s_{i,t,j,d}\}_{i,d}\right).$$

Tube:

$$\hat{\mu}_{t,j} \pm q_{t,j}\hat{\sigma}_{t,j}.$$

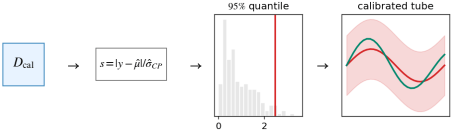

Fig. 17: Split conformal calibration converts residual scores into 95 percent prediction tubes.

VALIDITY SCOPE

What the Guarantee Says – and Does Not Say

Guaranteed under exchangeability

[26, 43]

$$\Pr\{y_{t,j} \in \widehat{C}_{t,j}(X)\} \geq 1 - \alpha.$$

Not guaranteed

- strong conditional coverage [14],
- shift-robust guarantees [33],
- automatic trajectory-level coverage.

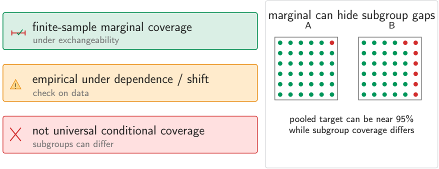

Fig. 18: Marginal validity can hold even when subgroup coverage differs.

ALGORITHM

SPARC in Nine Steps

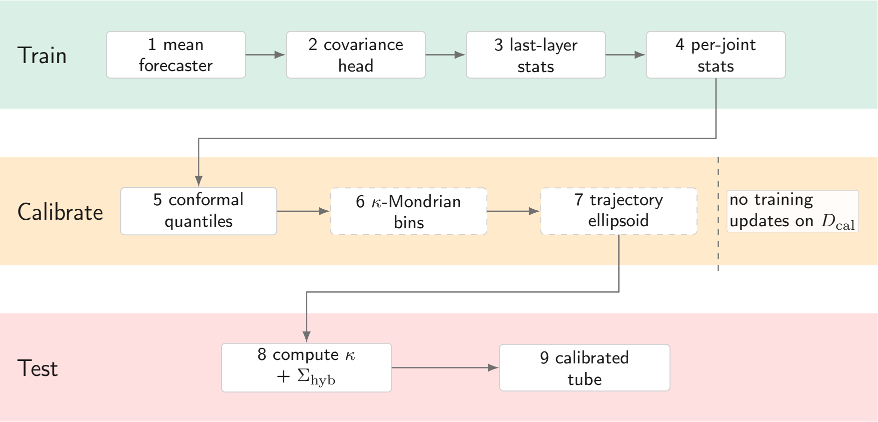

Fig. 19: The algorithm separates training, calibration, and deployment without updating on calibration data.

Optional Extensions

- κ-Mondrian CP: separate quantiles for risk bins [4, 42].
- Trajectory CP: structured Mahalanobis or whitened-residual score.
- Goal: risk-adaptive or more coherent set-valued prediction [24, 32, 48].

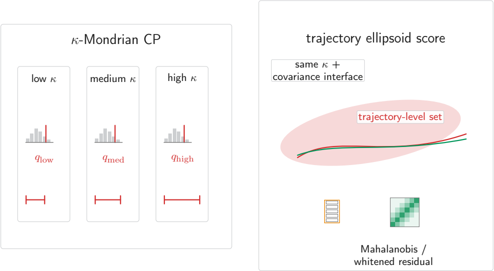

Fig. 20: Optional extensions reuse the same interface for Mondrian bins and trajectory-level scores.

Datasets

- Primary benchmark: Human3.6M [19].
- Additional motion/MoCap datasets: AMASS, LaFAN1, CMU-MoCap, 3DPW [5, 17, 28, 41].
- HRI-style datasets: CHICO, HA4M, AnDy [8, 31, 37].
- Non-H36M protocol: $T = 50$, $H = 25$.

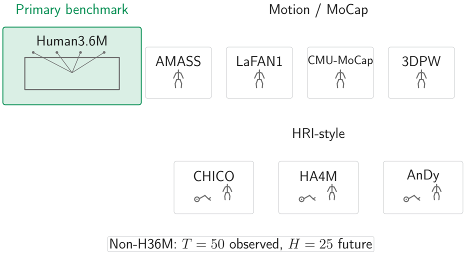

Fig. 21: The evaluation spans controlled MoCap, in-the-wild motion, and HRI-style datasets.

Metrics

- MPJPE: mean per-joint position error.
- FDE: final-horizon MPJPE.
- NLL: per-element Gaussian negative log-likelihood.
- Cov95/W95: nominal coverage and interval width.
- Cov95(CP)/W95(CP): conformalized coverage and width.

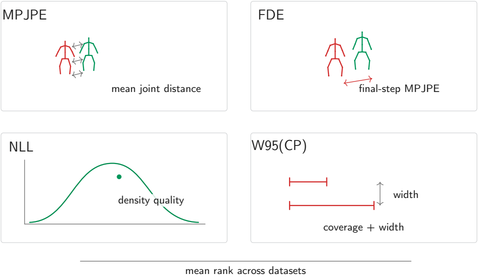

Fig. 22: Metrics compare point accuracy, density quality, empirical coverage, and tube width.

TABLE 1

Main Result: Mean Ranks

SPARC

- $\mathrm{MR}_{\mathrm{NLL}} = 1.00$.
- $\mathrm{MR}_{\mathrm{MPJPE+NLL}} = 2.69$.
- $\mathrm{MR}_{\mathrm{W95(CP)}} = 2.50$.
- $\mathrm{MR}_{\mathrm{MPJPE}} = 4.39$.

Lower rank is better. SPARC is strongest on density and the combined point+density trade-off.

Model	MPJPE	NLL	W95(CP)	MPJPE+NLL
MeanFT	**2.22**	6.11	–	4.17
Bridge-CQR	4.61	3.00	9.11	3.81
Deep ensemble	8.22	5.56	5.39	6.89
CoMusion	6.06	6.44	3.78	6.25
SPARD	7.11	6.56	3.39	6.83
SPARC	4.39	**1.00**	**2.50**	**2.69**

Table 1: Selected mean-rank rows; the next slide gives the full cross-dataset table.

LOWER IS BETTER

Full Table 1: Cross-Dataset Mean Ranks

Model	MR MPJPE	MR NLL	MR W95(CP)	MR MPJPE+NLL
siMLPe (base) [16]	7.89	10.67	–	9.28
MeanFT (siMLPe fine-tune) [16]	**2.22**	6.11	–	4.17
AuxTasks (50/50) [45]	9.06	11.78	–	10.42
Symplectic (50/50) [6]	16.22	15.78	–	16.00
DeFeeNet (50/50) [39]	7.61	10.44	–	9.03
GSPS (50/50) [29]	7.67	9.00	5.44	8.33
HumanMAC (50/50) [7]	10.22	10.00	–	10.11
SeSGCN (teacher)	13.39	11.78	–	12.58
Deep ensemble [22]	8.22	5.56	5.39	6.89
Bridge-CQR [35]	4.61	3.00	9.11	3.81
SkeletonDiffusion (CVPR 2025) [9]	15.00	13.67	9.78	14.33
BeLFusion (ICCV 2023) [2]	9.56	8.00	5.89	8.78
TransFusion (RA-L 2024) [40]	12.44	9.78	4.78	11.11
CoMusion (ECCV 2024) [38]	6.06	6.44	3.78	6.25
SPARD (AAAI 2026) [49]	7.11	6.56	3.39	6.83
MotionMap (CVPR 2025) [18]	16.11	16.22	9.44	16.17
SLD-HMP (ECCV 2024) [46]	13.22	15.22	6.50	14.22
SPARC	4.39	**1.00**	**2.50**	**2.69**

Table 2: Mean ranks across dataset/protocol blocks; MPJPE+NLL averages the MPJPE and NLL ranks.

INTERPRETATION

Coverage Must Not Come From Huge Tubes

- Best density quality in the benchmark.
- Efficient calibrated tubes: coverage at moderate width.
- Single forward pass.
- κ remains interpretable as an epistemic risk signal for deployment-oriented HRI monitoring [21, 23, 37].

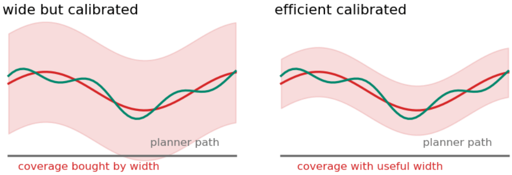

Fig. 23: Efficient calibration seeks target coverage without unnecessarily wide intervals.

Ablation: Component Contributions

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 + timewise κ_t	✓	✓	✓	–	<u>2.53</u>	3.78
SPARC coupled κ_t	✓	✓	✓	✓	3.17	**2.56**

Table 3: Ablation over leverage, hybrid covariance, graph coupling, and coupled per-joint calibration.

FIGURE 3

Qualitative Uncertainty

- Gray: nominal 95% intervals, $\pm 1.960\sigma$.
- Red: split-conformal 95% intervals, $\pm q_{t,j}\sigma$.
- Green: ground truth.
- Higher mean κ creates visible epistemic inflation.

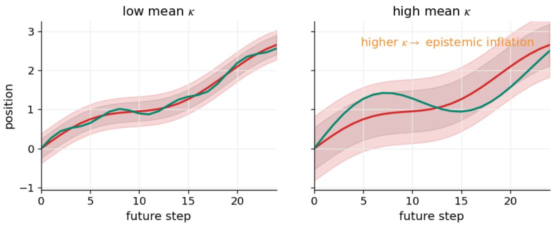

Fig. 24: Higher mean κ visibly expands calibrated intervals around the trajectory.

κ as a Risk Monitor

H36M diagnostic reported in the paper:

- Highest κ decile: $1.79\times$ average MPJPE.
- Filtering that decile reduces MPJPE by about 9%.
- The diagnostic supports κ as a deployment-time risk signal, not as a formal guarantee.
- Thresholds are tuned on held-out data and should be treated as policy parameters.

Why Single-Pass Matters

Inference requires:

- one backbone forward pass,
- horizon-wise quadratic forms for κ_t,
- a structured Gaussian summary,
- conformal quantile lookup.

No MC dropout loop [13], deep ensemble loop [22], or diffusion sampling loop [2, 38].

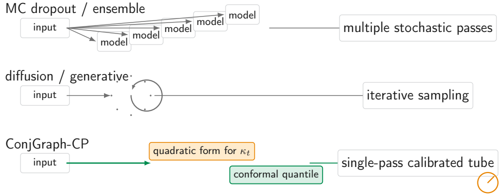

Fig. 25: Single-pass inference avoids repeated stochastic or generative sampling loops.

Takeaways and Limitations

Takeaways

- Analytic epistemic scaling via $\kappa_t(x)$.
- Structured trajectory covariance via $\Sigma_{\mathrm{str},t}(x)$.
- Split conformal calibration gives 95% marginal tubes [26, 43].
- The method is strongest on density quality and the combined point+density trade-off.

Limitations

- κ is representation-dependent.
- Coverage is marginal, not strongly conditional [14].
- The graph covariance is a modeling choice.
- Online updates and coherent trajectory-level sets remain open.

Strict Split Separation

- Deterministic window permutation with seed 304.
- $n_{\mathrm{cal}} = 512$ for split conformal calibration.
- $n_{\mathrm{eval}} = 1024$ for final metrics.
- $\alpha = 0.05$ for 95% target coverage.
- Hyperparameters selected on held-out training windows.

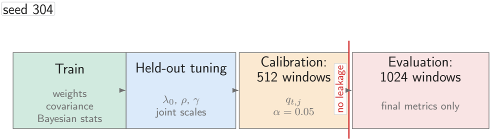

Fig. 26: Strict splits keep training, calibration, and evaluation windows separated.

TU Clausthal
Clausthal University of Technology

References I

[1] A. N. Angelopoulos and S. Bates. Conformal prediction: A gentle introduction. *Foundations and Trends® in Machine Learning*, 2023. doi: 10.1561/2200000101.

[2] G. Barquero, S. Escalera, and C. Palmero. BeLFusion: Latent diffusion for behavior-driven human motion prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2317–2327, 2023. doi: 10.48550/arXiv.2211.14304. URL `https://arxiv.org/abs/2211.14304`.

[3] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225, 1974. doi: 10.1111/j.2517-6161.1974.tb00999.x. URL `https://doi.org/10.1111/j.2517-6161.1974.tb00999.x`.

[4] H. Boström and U. Johansson. Mondrian conformal regressors. In *Proceedings of the Ninth Symposium on Conformal and Probabilistic Prediction and Applications*, volume 128 of *Proceedings of Machine Learning Research*, pages 114–133. PMLR, 2020. URL `https://proceedings.mlr.press/v128/bostrom20a.html`.

[5] Carnegie Mellon University. CMU graphics lab motion capture database. Online, 2026. URL `https://mocap.cs.cmu.edu/`. Accessed: 2026-02-09.

TU Clausthal
Clausthal University of Technology

References II

[6] H. Chen, K. Lyu, Z. Liu, Y. Yin, X. Yang, and Y. Lyu. Rethinking human motion prediction with symplectic integral. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2134–2143, 2024. doi: 10.1109/CVPR52733.2024.00208. URL `https://openaccess.thecvf.com/content/CVPR2024/html/Chen_Rethinking_Human_Motion_Prediction_with_Symplectic_Integral_CVPR_2024_paper.html`.

[7] L.-H. Chen, J. Zhang, Y. Li, Y. Pang, X. Xia, and T. Liu. HumanMAC: Masked motion completion for human motion prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9510–9521, 2023. doi: 10.1109/ICCV51070.2023.00875.

[8] G. Cicirelli, R. Marani, L. Romeo, M. García Domínguez, J. Heras, A. G. Perri, and T. D'Orazio. The HA4M dataset: Multi-modal monitoring of an assembly task for human action recognition in manufacturing. *Scientific Data*, 9:745, 2022. doi: 10.1038/s41597-022-01843-z. URL `https://doi.org/10.1038/s41597-022-01843-z`.

[9] C. Curreli, D. Muhle, A. Saroha, Z. Ye, R. Marin, and D. Cremers. Nonisotropic gaussian diffusion for realistic 3D human motion prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1871–1882, 2025. doi: 10.48550/arXiv.2501.06035. URL `https://arxiv.org/abs/2501.06035`.

[10] A. P. Dawid. Some matrix-variate distribution theory: Notational considerations and a Bayesian application. *Biometrika*, 68(1):265–274, 1981. doi: 10.1093/biomet/68.1.265.

References III

[11] 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. URL `https://proceedings.neurips.cc/paper/2021/file/a7c9585703d275249f30a088cebba0ad-Paper.pdf`.

[12] J. Diquigiovanni, M. Fontana, and S. Vantini. Conformal prediction bands for multivariate functional data, 2021. URL `https://arxiv.org/abs/2106.01792`.

[13] 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. URL `https://proceedings.mlr.press/v48/gal16.html`.

[14] I. Gibbs, J. J. Cherian, and E. J. Candès. Conformal prediction with conditional guarantees, 2023. URL `https://arxiv.org/abs/2305.12616`.

[15] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger. On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, volume 70 of *Proceedings of Machine Learning Research*, pages 1321–1330. PMLR, 2017. URL `https://proceedings.mlr.press/v70/guo17a.html`.

References IV

[16] 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. URL https://openaccess.thecvf.com/content/WACV2023/html/Guo_Back_to_MLP_A_Simple_Baseline_for_Human_Motion_Prediction_WACV_2023_paper.html.

[17] F. G. Harvey, M. Yurick, D. Nowrouzezahrai, and C. J. Pal. Robust motion in-betweening. *ACM Transactions on Graphics*, 39(4):60, 2020. doi: 10.1145/3386569.3392480. URL https://doi.org/10.1145/3386569.3392480.

[18] R. Hosseininejad, M. Shukla, S. Saadatnejad, M. Salzmann, and A. Alahi. MotionMap: Representing multimodality in human pose forecasting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025. doi: 10.48550/arXiv.2412.18883. URL https://arxiv.org/abs/2412.18883.

[19] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu. Human3.6M: Large scale datasets and predictive methods for 3D human sensing in natural environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36 (7):1325–1339, 2014. doi: 10.1109/TPAMI.2013.248.

[20] A. Kendall and Y. Gal. What uncertainties do we need in Bayesian deep learning for computer vision? In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 30, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/2650d6089a6d640c5e85b2b88265dc2b-Abstract.html.

TU Clausthal
Clausthal University of Technology

References V

[21] T. Kruse, A. K. Pandey, R. Alami, and A. C. Kirsch. Human-aware robot navigation: A survey. *Robotics and Autonomous Systems*, 61(12):1726–1743, 2013. doi: 10.1016/j.robot.2013.05.007.

[22] 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. URL `https://arxiv.org/abs/1612.01474`.

[23] P. A. Lasota, T. Fong, and J. A. Shah. A survey of methods for safe human-robot interaction. *Foundations and Trends® in Robotics*, 5(4):261–349, 2017. doi: 10.1561/2300000052.

[24] J. Lee, C. Xu, and Y. Xie. Flow-based conformal prediction for multi-dimensional time series. In *The Fourteenth International Conference on Learning Representations (ICLR)*, 2026. URL `https://openreview.net/forum?id=Uv3efQiPBZ`.

[25] J. Lei, A. Rinaldo, and L. Wasserman. A conformal prediction approach to explore functional data, 2013. URL `https://arxiv.org/abs/1302.6452`.

[26] 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. URL `https://doi.org/10.1080/01621459.2017.1307116`.

References VI

[27] D. J. C. MacKay. Bayesian interpolation. *Neural Computation*, 4(3):415–447, 1992. doi: 10.1162/neco.1992.4.3.415.

[28] N. Mahmood, N. Ghorbani, N. F. Troje, G. Pons-Moll, and M. J. Black. AMASS: Archive of motion capture as surface shapes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5442–5451, 2019. doi: 10.1109/ICCV.2019.00554. URL `https://doi.org/10.1109/ICCV.2019.00554`.

[29] W. Mao, M. Liu, and M. Salzmann. Generating smooth pose sequences for diverse human motion prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 13289–13298, 2021. doi: 10.1109/ICCV48922.2021.01306.

[30] J. Martinez, M. J. Black, and J. Romero. On human motion prediction using recurrent neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. doi: 10.48550/arXiv.1705.02445. URL `https://arxiv.org/abs/1705.02445`.

[31] P. Maurice, A. Malaisé, S. Ivaldi, O. Rochel, C. Amiot, N. Paris, G.-J. Richard, and L. Fritzsche. Andydata-lab-oneperson. Zenodo, 2018. URL `https://doi.org/10.5281/zenodo.1471975`.

TU Clausthal
Clausthal University of Technology

References VII

[32] S. Nag, U. Ghosh, C.-K. Ta, S. Bose, J. Li, and A. K. Roy-Chowdhury. Conformal prediction and MLLM-aided uncertainty quantification in scene graph generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11676–11686, 2025. URL https://openaccess.thecvf.com/content/CVPR2025/html/Nag_Conformal_Prediction_and_MLLM_aided_Uncertainty_Quantification_in_Scene_Graph_CVPR_2025_paper.html.

[33] Y. Ovadia, E. Fertig, J. Ren, Z. Nado, D. Sculley, S. Nowozin, J. Dillon, B. Lakshminarayanan, and J. Snoek. Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/8558cb408c1d76621371888657d2eb1d-Paper.pdf.

[34] C. Riquelme, G. Tucker, and J. Snoek. Deep Bayesian bandits showdown: An empirical comparison of bayesian deep networks for thompson sampling. In *International Conference on Learning Representations (ICLR)*, 2018. doi: 10.48550/arXiv.1802.09127. URL https://arxiv.org/abs/1802.09127.

[35] Y. Romano, E. Patterson, and E. Candès. Conformalized quantile regression. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019. doi: 10.48550/arXiv.1905.03222. URL https://arxiv.org/abs/1905.03222.

References VIII

[36] H. Rue and L. Held. *Gaussian Markov Random Fields: Theory and Applications*. Chapman and Hall/CRC, 2005. doi: 10.1201/9780203492024. URL `https://doi.org/10.1201/9780203492024`.

[37] P. Sampieri, X. Alameda-Pineda, J. Solhugot, V. Lepetit, and F. Moreno-Noguer. Pose forecasting in industrial human-robot collaboration. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 51–69, 2022. doi: 10.1007/978-3-031-19839-7_4. URL `https://doi.org/10.1007/978-3-031-19839-7_4`.

[38] J. Sun and G. Chowdhary. CoMusion: Towards consistent stochastic human motion prediction via motion diffusion. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2024. doi: 10.48550/arXiv.2305.12554. URL `https://arxiv.org/abs/2305.12554`.

[39] X. Sun, H. Sun, B. Li, D. Wei, W. Li, and J. Lu. DeFeeNet: Consecutive 3D human motion prediction with deviation feedback. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5527–5536, 2023. doi: 10.1109/CVPR52729.2023.00535.

[40] S. Tian, M. Zheng, and X. Liang. TransFusion: A practical and effective transformer-based diffusion model for 3D human motion prediction. *IEEE Robotics and Automation Letters*, 9(7):6232–6239, 2024. doi: 10.1109/LRA.2024.3401116.

References IX

[41] T. von Marcard, R. Henschel, M. J. Black, B. Rosenhahn, and G. Pons-Moll. Recovering accurate 3D human pose in the wild using IMUs and a moving camera. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 614–631, 2018. doi: 10.1007/978-3-030-01249-6_37. URL `https://doi.org/10.1007/978-3-030-01249-6_37`.

[42] V. Vovk. Conditional validity of inductive conformal predictors. In *Proceedings of the Asian Conference on Machine Learning*, volume 25 of *Proceedings of Machine Learning Research*, pages 475–490. PMLR, 2012. URL `https://proceedings.mlr.press/v25/vovk12.html`.

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

[44] J. Watson, J. A. Lin, P. Klink, J. Pajarinen, and J. Peters. Latent derivative bayesian last layer networks. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 130 of *Proceedings of Machine Learning Research*, pages 1198–1206. PMLR, 2021. URL `https://proceedings.mlr.press/v130/watson21a.html`.

[45] C. Xu, R. T. Tan, Y. Tan, S. Chen, X. Wang, and Y. Wang. Auxiliary tasks benefit 3D skeleton-based human motion prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9475–9486, 2023. doi: 10.1109/ICCV51070.2023.00872.

References X

[46] G. Xu, J. Tao, W. Li, and L. Duan. Learning semantic latent directions for accurate and controllable human motion prediction, 2024. URL https://arxiv.org/abs/2407.11494.

[47] Y. Yuan and K. M. Kitani. DLow: Diversifying latent flows for diverse human motion prediction. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 346–364, 2020. doi: 10.1007/978-3-030-58545-7_20.

[48] B. Zhang, S. Li, and O. Bastani. Conformal structured prediction. In *The Thirteenth International Conference on Learning Representations (ICLR)*, pages 1–21, 2025. URL https://openreview.net/forum?id=2ATD8a8P3C.

[49] Y. Zhang, B. Han, X. Li, W. Pang, F. Giunchiglia, X. Feng, and R. Guan. SPARD: Single-step inference with adaptive sampling in residual diffusion for human motion prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2026. URL https://github.com/ZYM-JLU/SPARD. Accepted.