Supplementary Material for: Single-Pass Scaling for Motion Forecasting with Bayesian Last Layers

Anonymous ECCV 2026 Submission

Paper ID #00000

A Human3.6M detailed results

Table 1: Human3.6M (H=25; seed=304, n_{cal}=512, n_{eval}=1024, α=0.05): common baselines vs. ConjGraph-CP. Cov95(CP)/W95(CP) are split-conformal calibrated 95% marginal tubes when available; deterministic methods do not output $\hat{\sigma}_{t,c}(x)$ and thus do not report CP tube metrics.

Model	MPJPE/FDE↓	NLL↓	Cov95(CP)	W95(CP)↓
siMLPe (base)	0.1012 / 0.1535	12.84	–	–
MeanFT (siMLPe fine-tune)	**0.0720**	5.24	–	–
AuxTasks (50/50)	0.0999	12.63	–	–
Symplectic (50/50)	0.1070 / 0.2151	13.08	–	–
DeFeeNet (50/50)	0.0981	12.31	–	–
GSPS (50/50)	0.0960 / 0.1540	10.96	0.945	0.313
HumanMAC (50/50)	0.0982	7.83	–	–
SeSGCN (teacher)	0.1119 / 0.1505	9.85	–	–
Deep ensemble	0.0971 / 0.1515	9.12	0.946	0.337
Bridge-CQR	0.0721 / 0.1189	1.13	0.995	1.238
SkeletonDiffusion	0.1242 / 0.1662	13.58	0.947	0.406
BeLFusion	0.1055 / 0.1512	6.22	0.948	0.316
TransFusion	0.1060 / 0.1449	7.18	0.948	0.295
CoMusion	0.0944 / 0.1409	4.49	0.945	0.287
SPARD	0.0928 / 0.1364	4.24	0.947	0.250
MotionMap	0.2504 / 0.2530	29.09	**0.951**	0.718
SLD-HMP	0.1170 / 0.1620	26.15	0.947	0.368
ConjGraph-CP†	0.0721 / 0.1189	**-1.66**	0.946	**0.248**

Legend: bold = best; underline = second-best (ties included). For Cov95(CP), best/second-best are defined by closeness to the 0.95 target. † (MN-GraphJ, time-coupled κ_t, shrink).

Table 2: Human3.6M (H=25; seed=304, n_{cal}=512, n_{eval}=1024, α=0.05): component isolation under a fixed mean predictor. The structured covariance models (MatrixNormal, MN-GraphJ, and ConjGraph-CP) share the same strong MeanFT mean predictor (freeze-mean), so differences among them are driven by uncertainty modeling. Conformal calibration (CP) moves coverage toward the 0.95 target with a modest width adjustment, providing a distribution-free marginal calibration layer under exchangeability even under likelihood misspecification.

Model	MPJPE↓	NLL↓	Cov95	W95↓	Cov95(CP)	W95(CP)↓
MeanFT (deterministic) + fixed σ_{obs}	0.0721	5.27	0.715	0.067	–	–
DCT-Bridge (diagonal Gaussian)	0.0721	-1.03	0.903	0.184	0.945	0.257
MatrixNormal (freeze-mean)	0.0721	-1.26	0.925	0.216	0.945	0.253
MatrixNormal-GraphJ (freeze-mean)	0.0721	-1.66	0.868	0.137	0.945	0.251
ConjGraph-CP[†]	0.0721	-1.66	0.868	0.137	0.946	0.248

[†] (MN-GraphJ, time-coupled κ_t, shrink).

B Cross-dataset detailed results

Table 3: Baselines vs. ConjGraph-CP† across datasets (H=25). Bold marks the best MPJPE/FDE, NLL, and W95(CP) per dataset among the shown methods (ties included); for Cov95(CP), bold marks the value closest to the 0.95 target. Cov95(CP)/W95(CP) are split-conformal calibrated 95% marginal tubes (coverage/width averaged over coordinates). External multimodal baselines include SkeletonDiffusion [2], BeLFusion [1], TransFusion [5], CoMusion [4], SPARD [7], MotionMap [3], and SLD-HMP [6].

	AMASS				LaFAN1				LaFAN1-HF			
Model	MPJPE/FDE↓	NLL↓	Cov95(CP)	W95(CP)↓	MPJPE/FDE↓	NLL↓	Cov95(CP)	W95(CP)↓	MPJPE/FDE↓	NLL↓	Cov95(CP)	W95(CP)↓
siMLPe (base)	0.0954 / 0.1230	13.55	–	–	0.5453 / 0.7066	319.27	–	–	0.4046 / 0.6400	172.36	–	–
MeanFT (siMLPe fine-tune)	**0.0593 / 0.0840**	3.16	–	–	0.5358 / 0.6729	303.33	–	–	**0.3093**	99.80	–	–
AuxTasks (50/50)	0.0949 / 0.1253	12.39	–	–	0.5827 / 0.6938	353.83	–	–	0.4299	184.26	–	–
Symplectic (50/50)	0.1424 / 0.4436	39.70	–	–	1.4504	3203.83	–	–	0.4756	280.68	–	–
DeFeeNet (50/50)	0.0909 / 0.1547	12.15	–	–	0.5666	341.82	–	–	0.3992 / 0.6459	169.44	–	–
GSPS (50/50)	0.1000 / 0.1318	14.03	**0.950**	0.311	0.5408 / 0.6939	282.97	0.953	1.945	0.3833 / 0.6384	148.57	**0.950**	1.225
HumanMAC (50/50)	0.0722 / 0.0935	4.34	–	–	0.5564	290.09	–	–	0.4477	166.64	–	–
SeSGCN (teacher)	0.1029 / 0.1095	9.48	–	–	0.5872 / 0.6457	314.76	–	–	0.4061 / 0.5435	138.80	–	–
Deep ensemble	0.0954 / 0.1248	2.60	**0.950**	0.283	0.5388 / 0.7121	43.45	0.953	1.856	0.3894 / 0.6260	80.18	**0.950**	1.167
Bridge-CQR	**0.0593 / 0.0840**	0.15	0.954	0.384	0.5358 / 0.6729	139.07	**0.950**	1.907	0.3103 / 0.4719	45.42	0.962	1.251
SkeletonDiffusion (CVPR 2025)	0.0971 / 0.1049	9.02	**0.950**	0.334	0.6028 / 0.6769	280.23	0.951	4.209	1.8071 / 1.8613	3357.02	0.959	32.988
BeLFusion (ICCV 2023)	0.0792 / 0.1015	3.84	0.952	0.235	0.5618 / 0.7022	251.70	0.952	1.894	1.0006 / 1.6484	520.35	0.961	3.049
TransFusion (RA-L 2024)	0.0790 / 0.0972	3.76	0.951	0.198	0.6496 / 0.7432	324.90	**0.950**	1.847	1.4215 / 2.0418	1222.72	0.956	3.015
CoMusion (ECCV 2024)	0.0617 / 0.0836	1.49	0.952	**0.174**	**0.5268 / 0.6421**	219.66	0.949	1.590	0.3604 / 0.5383	92.62	**0.950**	1.242
SPARD (AAAI 2026)	0.0638 / 0.0880	1.51	0.953	0.175	0.5682 / 0.7141	292.27	0.955	1.759	0.3677 / 0.5393	105.70	0.954	1.078
MotionMap (CVPR 2025)	0.2038 / 0.2012	16.42	0.949	0.536	0.7463 / 0.7614	351.70	**0.950**	2.006	1.4691 / 1.5391	470.72	0.953	4.645
SLD-HMP (ECCV 2024)	0.1120 / 0.1252	22.45	0.949	0.308	0.6091 / 0.7049	491.37	0.955	1.936	0.4801 / 0.6722	401.14	0.952	1.347
ConjGraph-CP†	**0.0593 / 0.0840**	**-2.37**	0.952	0.211	0.5358 / 0.6729	**-0.84**	0.954	**1.580**	0.3103 / 0.4719	**-0.05**	0.952	**0.940**

	CMU-MoCap				Human3.6M				3DPW			
Model	MPJPE/FDE↓	NLL↓	Cov95(CP)	W95(CP)↓	MPJPE/FDE↓	NLL↓	Cov95(CP)	W95(CP)↓	MPJPE/FDE↓	NLL↓	Cov95(CP)	W95(CP)↓
siMLPe (base)	2.6560	8726.23	–	–	0.1012 / 0.1535	12.84	–	–	0.0709 / 0.1021	5.08	–	–
MeanFT (siMLPe fine-tune)	2.6709	8820.46	–	–	**0.0720**	5.24	–	–	**0.0572 / 0.0832**	1.71	–	–
AuxTasks (50/50)	**2.5429**	8292.99	–	–	0.0999	12.63	–	–	0.0745 / 0.1029	5.27	–	–
Symplectic (50/50)	6.8024 / 10.4453	54096.87	–	–	0.1070 / 0.2151	13.08	–	–	0.1042 / 0.2794	20.62	–	–
DeFeeNet (50/50)	2.6316	8633.05	–	–	0.0981	12.31	–	–	0.0701 / 0.0996	4.99	–	–
GSPS (50/50)	2.6321 / 3.7290	8307.55	0.952	9.785	0.0960 / 0.1540	10.96	0.945	0.313	0.0741 / 0.1138	4.96	**0.949**	0.232
HumanMAC (50/50)	4.8433	21062.22	–	–	0.0982	7.83	–	–	0.0707	2.68	–	–
SeSGCN (teacher)	4.9092 / 5.1552	20996.24	–	–	0.1119 / 0.1505	9.85	–	–	0.0952 / 0.1100	5.32	–	–
Deep ensemble	2.7084 / 3.7452	1310.38	0.951	9.855	0.0971 / 0.1515	9.12	0.946	0.337	0.0690 / 0.1004	3.90	**0.951**	0.255
Bridge-CQR	2.7454 / 3.9290	4183.52	0.953	9.923	0.0721 / 0.1189	1.13	0.995	1.238	0.0580 / 0.0841	-0.45	0.955	0.483
SkeletonDiffusion (CVPR 2025)	8.0957 / 8.3124	49441.83	**0.950**	381.727	0.1242 / 0.1662	13.58	0.947	0.406	0.0822 / 0.0988	4.91	0.948	0.256
BeLFusion (ICCV 2023)	3.6186 / 4.4319	12261.03	0.954	14.329	0.1055 / 0.1512	6.22	0.948	0.316	0.0771 / 0.1059	2.13	**0.949**	0.207
TransFusion (RA-L 2024)	8.0839 / 8.4094	44901.52	0.952	25.699	0.1060 / 0.1449	7.18	0.948	0.295	0.0787 / 0.0997	2.72	0.945	0.187
CoMusion (ECCV 2024)	8.6798 / 8.6127	54190.65	0.953	24.961	0.0944 / 0.1409	4.49	0.945	0.287	0.0690 / 0.1001	1.19	0.945	**0.174**
SPARD (AAAI 2026)	5.5562 / 5.7267	25823.22	0.952	37.934	0.0928 / 0.1364	4.24	0.947	0.250	0.0703 / 0.1035	1.25	0.946	**0.174**
MotionMap (CVPR 2025)	5.3103 / 5.6501	21490.65	0.951	18.384	0.2504 / 0.2530	29.09	**0.951**	0.718	0.1841 / 0.1851	9.58	0.948	0.476
SLD-HMP (ECCV 2024)	2.7547 / 3.8620	16866.43	0.951	11.966	0.1170 / 0.1620	26.15	0.947	0.368	0.0846 / 0.1055	15.79	0.948	0.235
ConjGraph-CP†	2.7207 / 3.9083	**1.85**	0.954	**8.219**	0.0721 / 0.1189	**-1.66**	0.946	**0.248**	0.0580 / 0.0841	**-1.94**	0.947	0.188

	CHICO				HA4M				AnDy-onePerson			
Model	MPJPE@10/25f (mm)↓	NLL↓	Cov95(CP)	W95(CP)↓	MPJPE@10/25f (mm)↓	NLL↓	Cov95(CP)	W95(CP)↓	MPJPE@10/25f (mm)↓	NLL↓	Cov95(CP)	W95(CP)↓
siMLPe (base)	43.2 / 75.2	0.66	–	–	**46.1 / 57.0**	-1.08	–	–	28.0 / 70.3	2.00	–	–
MeanFT (siMLPe fine-tune)	42.7 / 64.8	-0.35	–	–	53.6 / 57.7	-1.00	–	–	**27.4 / 55.7**	-0.27	–	–
AuxTasks (50/50)	44.6 / 77.7	0.88	–	–	91.5 / 70.6	0.30	–	–	32.5 / 70.7	2.22	–	–
Symplectic (50/50)	46.6 / 91.2	0.77	–	–	79.5 / 364.9	24.42	–	–	30.2 / 105.5	2.62	–	–
DeFeeNet (50/50)	42.7 / 75.2	0.69	–	–	93.5 / 77.2	3.11	–	–	26.1 / 72.6	1.75	–	–
GSPS (50/50)	42.2 / 74.2	0.28	0.947	0.163	42.8 / 73.8	-0.22	0.929	**0.131**	27.9 / 73.4	1.89	0.951	0.182
HumanMAC (50/50)	52.0 / 70.7	0.27	–	–	189.5 / 196.7	35.25	–	–	55.7 / 80.0	1.98	–	–
SeSGCN (teacher)	66.2 / 77.7	1.43	–	–	150.9 / 171.5	21.59	–	–	55.7 / 81.3	2.05	–	–
Deep ensemble	41.2 / 76.4	-0.38	0.946	0.162	78.0 / 192.6	0.19	0.942	0.236	25.8 / 69.8	1.03	0.952	0.188
Bridge-CQR	42.7 / 64.8	-1.47	0.955	0.523	103.5 / 193.9	5.48	0.972	0.843	**27.4 / 55.7**	-1.44	0.965	0.866
SkeletonDiffusion (CVPR 2025)	57.7 / 76.8	1.41	0.952	0.214	196.7 / 189.9	29.06	0.958	0.778	63.3 / 89.7	4.60	0.949	0.265
BeLFusion (ICCV 2023)	48.4 / 69.7	-1.04	0.949	0.150	144.4 / 141.6	8.99	0.951	0.344	38.7 / 68.5	-0.14	**0.950**	0.150
TransFusion (RA-L 2024)	48.6 / 66.4	-0.81	0.949	0.133	145.5 / 170.8	5.43	0.944	0.296	54.7 / 82.1	0.07	0.949	0.142
CoMusion (ECCV 2024)	44.8 / 65.7	-1.28	**0.950**	0.131	152.7 / 168.8	22.95	0.951	0.547	32.4 / 59.4	-1.31	0.948	0.110
SPARD (AAAI 2026)	**44.5 / 64.5**	-1.39	0.949	**0.129**	152.2 / 176.1	22.71	**0.950**	0.473	32.9 / 61.0	-1.40	0.951	**0.107**
MotionMap (CVPR 2025)	161.2 / 165.5	3.15	0.953	0.435	173.2 / 166.5	50.54	0.944	0.431	179.3 / 190.2	15.52	0.951	0.532
SLD-HMP (ECCV 2024)	54.3 / 80.1	4.17	0.948	0.162	58.4 / 80.4	7.34	0.936	0.155	52.0 / 76.7	7.91	0.948	0.193
ConjGraph-CP†	42.7 / 64.8	**-2.19**	0.949	**0.129**	103.5 / 193.9	**-2.34**	0.984	0.255	**27.4 / 55.7**	**-3.63**	0.951	0.114

Legend: bold = best; underline = second-best (ties included). For Cov95(CP), best/second-best are defined by closeness to the 0.95 target. † (MN-GraphJ, time-coupled κ_t, shrink). Dataset-specific protocol-only baselines (e.g., CHICO-only SeSGCN protocol variants) are omitted to avoid redundant all-dataset rows.

C Full FC-Out ablation (per-dataset)

Table 4: ConjGraph-CP family component ablation matrix. Columns mark which components are active in each variant.

Variant	DCT pinv	Timewise κ_t	Hybrid Σ_{ale}	GraphJ	κ_t coupling	$\mathrm{MR}_{\mathrm{MPJPE+NLL}}\downarrow$
ConjGraph-CP family (base $\kappa(x)$)	–	–	–	–	–	5.33
ConjGraph-CP family (timewise κ_t)	–	✓	–	–	–	4.67
ConjGraph-CP family (DCT pinv)	✓	–	–	–	–	6.06
ConjGraph-CP family (hybrid MN)	–	–	✓	–	–	4.00
ConjGraph-CP family (hybrid MN-GraphJ)	–	–	✓	✓	–	**2.53**
ConjGraph-CP family (hybrid MN-GraphJ, timewise κ_t)	–	✓	✓	✓	–	2.58
ConjGraph-CP (hybrid MN-GraphJ, time-coupled κ_t)	–	✓	✓	✓	✓	2.83

Legend: bold = best; underline = second-best.

Table 5: Mean rank (MR; ↓ better) across datasets within the ConjGraph-CP ablation family. $\mathrm{MR}_{\mathrm{MPJPE+NLL}}$ is the mean of MPJPE and NLL ranks.

Variant	$\mathrm{MR}_{\mathrm{MPJPE}}\downarrow$	$\mathrm{MR}_{\mathrm{NLL}}\downarrow$	$\mathrm{MR}_{W95(\mathrm{CP})}\downarrow$	$\mathrm{MR}_{\mathrm{MPJPE+NLL}}\downarrow$
ConjGraph-CP family (base $\kappa(x)$)	5.78	4.89	–	5.33
ConjGraph-CP family (timewise κ_t)	5.78	3.56	–	4.67
ConjGraph-CP family (DCT pinv)	5.89	6.22	–	6.06
ConjGraph-CP family (hybrid MN)	**2.33**	5.67	–	4.00
ConjGraph-CP family (hybrid MN-GraphJ)	2.50	2.56	–	**2.53**
ConjGraph-CP family (hybrid MN-GraphJ, timewise κ_t)	2.50	2.67	–	2.58
ConjGraph-CP (hybrid MN-GraphJ, time-coupled κ_t)	3.22	**2.44**	**1.00**	2.83

Legend: bold = best; underline = second-best (ties included).

Table 6: Full ablation of ConjGraph-CP family variants across datasets (H=25). Bold/underline indicate best/second-best per dataset within this ablation family (ties included). For Cov95(CP), ranking is by distance to 0.95.

Variant	MPJPE↓	NLL↓	Cov95(CP)	W95(CP)↓
AMASS				
ConjGraph-CP family (base $\kappa(x)$)	0.0549	-1.84	–	–
ConjGraph-CP family (timewise κ_t)	0.0551	-2.04	–	–
ConjGraph-CP family (DCT pinv)	0.0605	-1.63	–	–
ConjGraph-CP family (hybrid MN)	**0.0542**	-1.64	–	–
ConjGraph-CP family (hybrid MN-GraphJ)	**0.0542**	-1.85	–	–
ConjGraph-CP family (hybrid MN-GraphJ, timewise κ_t)	0.0542	-1.86	–	–
ConjGraph-CP (hybrid MN-GraphJ, time-coupled κ_t)	0.0593	**-2.37**	**0.952**	**0.211**
LaFAN1				
ConjGraph-CP family (base $\kappa(x)$)	3.5196	2.37	–	–
ConjGraph-CP family (timewise κ_t)	0.6858	1.02	–	–

Continued on next page

Table 6: Full ablation of ConjGraph-CP family variants across datasets (H=25). Bold/underline indicate best/second-best per dataset within this ablation family (ties included). For Cov95(CP), ranking is by distance to 0.95. (continued)

Variant	MPJPE↓	NLL↓	Cov95(CP)	W95(CP)↓
ConjGraph-CP family (DCT pinv)	3.4217	2.13	–	–
ConjGraph-CP family (hybrid MN)	0.5498	0.28	–	–
ConjGraph-CP family (hybrid MN-GraphJ)	0.5498	-0.38	–	–
ConjGraph-CP family (hybrid MN-GraphJ, timewise κ_t)	0.5498	-0.29	–	–
ConjGraph-CP (hybrid MN-GraphJ, time-coupled κ_t)	**0.5358**	**-0.84**	**0.954**	**1.580**
LaFAN1-HF				
ConjGraph-CP family (base $\kappa(x)$)	0.4995	1.03	–	–
ConjGraph-CP family (timewise κ_t)	0.5271	3.72	–	–
ConjGraph-CP family (DCT pinv)	0.4417	1.10	–	–
ConjGraph-CP family (hybrid MN)	0.3093	3.58	–	–
ConjGraph-CP family (hybrid MN-GraphJ)	0.3093	**-0.07**	–	–
ConjGraph-CP family (hybrid MN-GraphJ, timewise κ_t)	**0.3093**	0.02	–	–
ConjGraph-CP (hybrid MN-GraphJ, time-coupled κ_t)	0.3103	-0.05	**0.952**	**0.940**
CMU-MoCap				
ConjGraph-CP family (base $\kappa(x)$)	4.4830	2.40	–	–
ConjGraph-CP family (timewise κ_t)	2.7509	2.12	–	–
ConjGraph-CP family (DCT pinv)	21.8957	3.19	–	–
ConjGraph-CP family (hybrid MN)	**2.6549**	3.67	–	–
ConjGraph-CP family (hybrid MN-GraphJ)	2.6836	2.18	–	–
ConjGraph-CP family (hybrid MN-GraphJ, timewise κ_t)	**2.6549**	2.77	–	–
ConjGraph-CP (hybrid MN-GraphJ, time-coupled κ_t)	2.7207	**1.85**	**0.954**	**8.219**
Human3.6M				
ConjGraph-CP family (base $\kappa(x)$)	0.0729	-1.42	–	–
ConjGraph-CP family (timewise κ_t)	0.0730	**-1.72**	–	–
ConjGraph-CP family (DCT pinv)	0.0730	-1.18	–	–
ConjGraph-CP family (hybrid MN)	**0.0720**	-1.29	–	–
ConjGraph-CP family (hybrid MN-GraphJ)	**0.0720**	-1.69	–	–
ConjGraph-CP family (hybrid MN-GraphJ, timewise κ_t)	**0.0720**	-1.69	–	–
ConjGraph-CP (hybrid MN-GraphJ, time-coupled κ_t)	0.0721	-1.66	**0.946**	**0.248**
3DPW				
ConjGraph-CP family (base $\kappa(x)$)	0.0599	-1.86	–	–
ConjGraph-CP family (timewise κ_t)	0.0599	-2.06	–	–
ConjGraph-CP family (DCT pinv)	0.0595	-1.90	–	–
ConjGraph-CP family (hybrid MN)	0.0581	-1.87	–	–
ConjGraph-CP family (hybrid MN-GraphJ)	0.0581	-2.48	–	–
ConjGraph-CP family (hybrid MN-GraphJ, timewise κ_t)	0.0581	**-2.48**	–	–
ConjGraph-CP (hybrid MN-GraphJ, time-coupled κ_t)	**0.0580**	-1.94	**0.947**	**0.188**
CHICO				
ConjGraph-CP family (base $\kappa(x)$)	48.6 / 74.1	**-2.31**	–	–

Continued on next page

Table 6: Full ablation of ConjGraph-CP family variants across datasets (H=25). Bold/underline indicate best/second-best per dataset within this ablation family (ties included). For Cov95(CP), ranking is by distance to 0.95. (continued)

Variant	MPJPE↓	NLL↓	Cov95(CP)	W95(CP)↓
ConjGraph-CP family (timewise κ_t)	49.1 / 74.0	-2.30	–	–
ConjGraph-CP family (DCT pinv)	46.4 / 69.6	-2.07	–	–
ConjGraph-CP family (hybrid MN)	43.7 / 65.5	-1.68	–	–
ConjGraph-CP family (hybrid MN-GraphJ)	43.7 / 65.5	-2.21	–	–
ConjGraph-CP family (hybrid MN-GraphJ, timewise κ_t)	43.7 / 65.5	-2.23	–	–
ConjGraph-CP (hybrid MN-GraphJ, time-coupled κ_t)	**42.7 / 64.8**	-2.19	**0.949**	**0.129**
HA4M				
ConjGraph-CP family (base $\kappa(x)$)	57.7 / 130.4	-1.28	–	–
ConjGraph-CP family (timewise κ_t)	41.3 / 83.9	-1.09	–	–
ConjGraph-CP family (DCT pinv)	56.0 / 148.6	-0.86	–	–
ConjGraph-CP family (hybrid MN)	**42.0 / 76.7**	-1.36	–	–
ConjGraph-CP family (hybrid MN-GraphJ)	**42.0 / 76.7**	**-2.55**	–	–
ConjGraph-CP family (hybrid MN-GraphJ, timewise κ_t)	42.0 / 76.7	-2.41	–	–
ConjGraph-CP (hybrid MN-GraphJ, time-coupled κ_t)	103.5 / 193.9	-2.34	**0.984**	**0.255**
AnDy-onePerson				
ConjGraph-CP family (base $\kappa(x)$)	55.2 / 83.5	-1.44	–	–
ConjGraph-CP family (timewise κ_t)	59.3 / 129.6	-2.17	–	–
ConjGraph-CP family (DCT pinv)	48.3 / 75.7	-0.81	–	–
ConjGraph-CP family (hybrid MN)	30.7 / 58.6	-1.65	–	–
ConjGraph-CP family (hybrid MN-GraphJ)	30.7 / 58.6	-2.69	–	–
ConjGraph-CP family (hybrid MN-GraphJ, timewise κ_t)	30.7 / 58.6	-2.69	–	–
ConjGraph-CP (hybrid MN-GraphJ, time-coupled κ_t)	**27.4 / 55.7**	**-3.63**	**0.951**	**0.114**

Standard protocol: seed=304, $n_{\text{cal}} = 512$, $n_{\text{eval}} = 1024$, $\alpha = 0.05$. "–" denotes missing runs for a variant/dataset.

D Epistemic scale diagnostics

We retain only diagnostics that directly support decision-making: how epistemic scaling inflates uncertainty (Fig. 1(a)), whether κ stratifies error in a monotone way (Fig. 1(b)), the cross-dataset behavior of the two conformal variants enabled by κ and the structured covariance (Table 7), and a simple post-hoc selective-deployment analysis based on thresholding κ (Table 8), which does not affect the main benchmark ranks.

Table 7: ConjGraph-CP conformal variants across datasets (target 95%). CP denotes split conformal marginal tubes; Mondrian denotes κ-conditioned split conformal tubes; Set denotes the trajectory-level Mahalanobis conformal ellipsoid. $\log\mathrm{Vol}/d$ is the mean log-volume per dimension (lower is tighter).

Dataset	Cov95(CP)	W95(CP)↓	Cov95(Mondrian)	W95(Mondrian)↓	Cov95(Set)	$\log\mathrm{Vol}_{\mathrm{tube}/}$
AMASS	0.952	0.211	0.956	0.209	0.958	-1.

Continu

Table 7: ConjGraph-CP conformal variants across datasets. (continued)

Dataset	Cov95(CP)	W95(CP)↓	Cov95(Mondrian)	W95(Mondrian)↓	Cov95(Set)	$\log\mathrm{Vol}_{\mathrm{tube}/}$
LaFAN1	0.954	1.580	0.956	1.627	0.964	0.
LaFAN1-HF	0.952	0.940	0.953	0.927	0.938	-0.
CMU-MoCap	0.954	8.219	0.958	8.562	0.950	1.
Human3.6M	0.946	0.248	0.948	0.248	0.954	-1.
3DPW	0.947	0.188	0.949	0.183	0.951	-2.
CHICO	0.949	0.129	0.952	0.131	0.957	-2.
HA4M	0.984	0.255	0.966	0.242	1.000	-1.
AnDy-onePerson	0.951	0.114	0.951	0.110	0.961	-2.

Table 8: Selective prediction using the analytic epistemic scale $\bar{\kappa}(x)$ on Human3.6M (H=25; seed=304, n_{eval}=1024). We sort evaluation sequences by mean $\sqrt{\kappa_t(x)}$ and report MPJPE on subsets induced by abstention/fallback thresholds (keep lowest-κ vs. worst highest-κ). This is a post-hoc analysis (point predictions are unchanged) and is not used in the main ranking tables; it provides no formal guarantee, but illustrates how κ can prioritize difficult windows. Pearson correlation between mean $\sqrt{\kappa_t(x)}$ and MPJPE is r=0.52.

Subset	Fraction	#Seq.	MPJPE↓	Rel.
All (baseline)	100%	1024	0.0721	1.00
Keep lowest-κ	90%	922	0.0658	0.91
Keep lowest-κ	80%	820	0.0622	0.86
Keep lowest-κ	70%	717	0.0605	0.84
Keep lowest-κ	60%	615	0.0596	0.83
Keep lowest-κ	50%	512	0.0580	0.80
Worst highest-κ	20%	205	0.1118	1.55
Worst highest-κ	10%	103	0.1292	1.79

Table 9: Human3.6M: κ-conditioned (Mondrian) split conformal for marginal tubes. We calibrate on a held-out subset (n=512) and evaluate on a disjoint test subset (n=1024); target miscoverage is $\alpha = 0.05$. Sequences are binned into 3 quantile bins by mean $\sqrt{\kappa_t(x)}$ on the calibration subset. CP uses joint-wise quantiles $q_{t,j}$ (pooled over the 3 coordinates of each joint); Mondrian calibrates $q_{t,j,b}$ per κ-bin b. We report marginal coverage (Cov95) and mean interval width (W95).

$\sqrt{\kappa}$ bin	#Seq.	Cov95(CP)	W95(CP)↓	Cov95(Mondrian)	W95(Mondrian)↓
0–33% [1.000175,1.001171)	290	0.971	0.248	0.944	0.187
33–67% [1.001171,1.003041)	370	0.958	0.248	0.950	0.231
67–100% [1.003041,1.012778)	364	0.912	0.249	0.950	0.315
All	1024	0.946	0.248	0.948	0.248

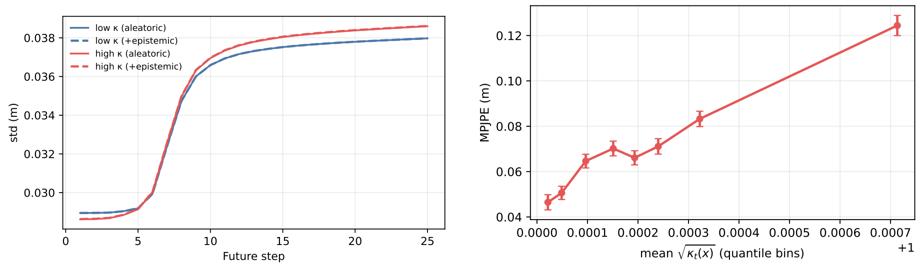

Fig. 1: Epistemic scale diagnostics. **(a)** How κ_t inflates uncertainty: we illustrate how epistemic scaling inflates a structured aleatoric marginal standard deviation (mean over coordinates) for a low-κ and a high-κ example. **(b)** Actionability of the epistemic scale: MPJPE increases monotonically across quantile bins of mean $\sqrt{\kappa_t(x)}$ (error bars: standard error), supporting κ as a usable risk stratifier beyond average coverage metrics.

Table 10: Human3.6M: trajectory-level conformal ellipsoid from the structured hybrid covariance. We calibrate on a held-out subset (n=512) and evaluate on a disjoint test subset (n=1024); $\alpha = 0.05$. The tube is the axis-aligned CP rectangle (marginal coverage); the ellipsoid uses the structured Mahalanobis/whitened-residual score under $\Sigma_T^{\text{hyb}} \otimes \Sigma_C$ and is conformalized to achieve trajectory-level coverage. $\log\text{Vol}/d$ reports mean log-volume per dimension (lower is tighter).

Set	Coverage	$\log\text{Vol}/d \downarrow$
Marginal tube (CP)	0.946	-1.714
Trajectory ellipsoid (Mahalanobis CP)	0.954	-1.316

E Controlled synthetic uncertainty benchmark

To complement real-dataset evaluation, we run a controlled synthetic benchmark with known uncertainty structure and test whether ConjGraph-CP recovers the intended epistemic/aleatoric components. Given observed prefixes, we generate future trajectories as

$$y_t = \mu_t^\star(x) + \varepsilon_t, \qquad \varepsilon \sim \mathcal{N}\big(0,\ \kappa_t^\star(x)\, \Sigma_T^\star \otimes \Sigma_C^\star\big),$$

where $\Sigma_T^\star$ is first-order autoregressive (AR(1)) temporal covariance, $\Sigma_C^\star$ is graph-coupled coordinate covariance, and $\kappa_t^\star(x)$ is an input-dependent heteroscedastic scale. We use a protocol-aligned split (n_{train}=16,000, n_{cal}=512, n_{eval}=1024) plus a shifted evaluation split. Figure 2 first summarizes what the synthetic benchmark encodes: in-distribution vs. shifted trajectory families, the induced uncertainty-regime shift in $\sqrt{\kappa_t^\star}$, and horizon-wise uncertainty profiles.

Table 11: Before/after synthetic tuning summary for ConjGraph-CP under the fixed protocol split (n_{cal}=512, n_{eval}=1024, α=0.05). "Scale" denotes the fixed multiplier applied to the analytic Bayesian quadratic form in κ_t. Cov95(CP) uses joint-wise split conformal quantiles $q_{t,j}$ (pooled over the 3 coordinates of each joint).

Setting	r_κ ID ↑	r_κ shift ↑	Temp MAE ↓	Joint MAE ↓	Cov95(CP) ID	Cov95(CP)
Legacy (no feat-std, scale 1)	0.475	0.820	**0.022**	0.103	0.950	0.667
Tuned (feat-std, scale 200)	**0.499**	**0.848**	0.022	**0.100**	0.949	0.847

Table 11 compares the legacy synthetic fit to our tuned setup (feature-standardized Bayesian precision fit + calibrated epistemic sensitivity scale). The tuned setup improves epistemic tracking (r_κ: 0.475 → 0.499 in-distribution; 0.820 → 0.848 under shift), where r_κ is the Pearson correlation between the predicted vs. true sequence-level mean $\sqrt{\kappa_t^\star}$, maintains temporal component recovery (temporal-correlation mean absolute error (MAE) 0.022 → 0.022) while improving joint-correlation MAE (0.103 → 0.100), and substantially improves shifted conformal coverage (0.667 → 0.847) while keeping in-distribution Cov95(CP) close to target ($\approx$ 0.949). Figure 3 visualizes these recovered components: top-left/top-middle preserve the AR-style temporal correlation geometry, top-right shows tighter true-vs.-learned $\sqrt{\kappa}$ alignment (with stronger separation under shift), and the bottom row shows improved graph-coupled correlation recovery and risk-profile matching over horizon.

κ-conditioned (Mondrian) conformal tubes. Beyond global split conformal, we consider a κ-conditioned (Mondrian) variant that bins sequences by mean $\sqrt{\kappa_t(x)}$ and calibrates group-wise quantiles, yielding risk-adaptive tube scaling. Table 12 reports Cov95/W95 by κ-bin for both unconditioned CP and the Mondrian variant.

Table 12: Synthetic ConjGraph benchmark: κ-conditioned (Mondrian) split conformal for marginal tubes. We split the saved ID evaluation set (n=1024) into calibration/test halves (512/512) and bin sequences by mean $\sqrt{\kappa_t(x)}$ on the calibration subset (5 quantile bins). CP uses joint-wise quantiles $q_{t,j}$ (pooled over the 3 coordinates of each joint); Mondrian calibrates $q_{t,j,b}$ per κ-bin b. Target miscoverage is $\alpha = 0.05$. We report marginal coverage (Cov95) and mean interval width (W95).

$\sqrt{\kappa}$ bin	#Seq.	Cov95(CP)	W95(CP)↓	Cov95(Mondrian)	W95(Mondrian)↓
In-distribution (ID) test					
0–20% [1.03,1.05)	107	0.959	0.146	0.958	0.146
20–40% [1.05,1.07)	133	0.954	0.149	0.957	0.151
40–60% [1.07,1.09)	93	0.954	0.152	0.961	0.158
60–80% [1.09,1.13)	87	0.946	0.156	0.952	0.162
80–100% [1.13,1.65)	92	0.953	0.175	0.954	0.176
All	512	0.954	0.155	0.957	0.158
Shifted evaluation					
0–20% [1.03,1.05)	2	0.938	0.147	0.932	0.147
20–40% [1.05,1.07)	9	0.962	0.149	0.966	0.151
40–60% [1.07,1.09)	18	0.963	0.152	0.970	0.158
60–80% [1.09,1.13)	39	0.948	0.156	0.955	0.162
80–100% [1.13,1.65)	956	0.844	0.281	0.846	0.283
All	1024	0.851	0.272	0.853	0.275

Trajectory-level conformal sets. We also construct a trajectory-level conformal ellipsoid using a structured Mahalanobis/whitened-residual score under $\Sigma_T^{\text{hyb}} \otimes \Sigma_C$ as an alternative to axis-aligned tubes. Table 13 compares coverage and set volume against marginal CP rectangles.

Table 13: Synthetic ConjGraph benchmark: trajectory-level conformal set from the structured hybrid covariance. We split the saved ID evaluation set (n=1024) into calibration/test halves (512/512); $\alpha = 0.05$. The trajectory set uses the whitened Frobenius/Mahalanobis score $\|(L_T^{\text{hyb}})^{-1}(y-\mu)L_C^{-\top}\|_F$ under $\Sigma_T^{\text{hyb}} \otimes \Sigma_C$ and outputs an ellipsoid; the tube is the axis-aligned CP rectangle. $\log\text{Vol}/d$ reports mean log-volume per dimension (lower is tighter). Shift results (n=1024) are shown for reference (no exchangeability guarantee).

Set	Cov95 ID	Cov95 shift	$\log\text{Vol}/d$ ID↓	$\log\text{Vol}/d$ shift↓
Marginal tube (CP)	0.954	0.851	-1.870	-1.357

Continued on next page

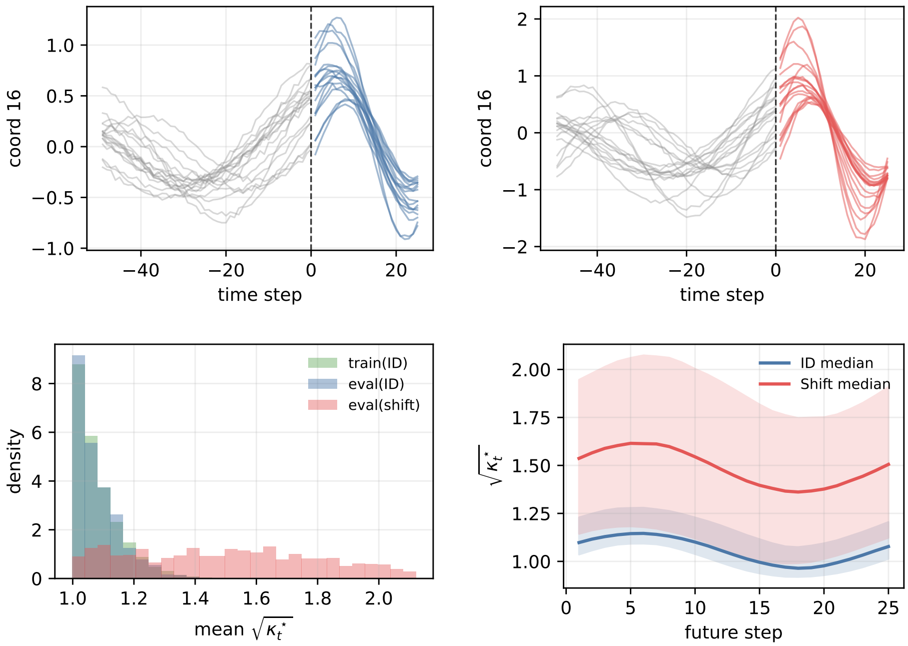

Fig. 2: Synthetic dataset overview used for ConjGraph-CP stress testing. **(a)** Representative in-distribution trajectory slices (same coordinate) with observed prefix and future continuation. **(b)** Representative shifted trajectory slices. **(c)** Distribution shift of the true mean epistemic scale $\sqrt{\kappa_t^\star}$. **(d)** Horizon-wise 10–50–90% profiles of $\sqrt{\kappa_t^\star}$ for in-distribution vs. shifted splits.

Table 13: Synthetic ConjGraph benchmark: trajectory-level conformal set from the structured hybrid covariance. (continued)

Set	Cov95 ID	Cov95 shift	logVol/d ID↓	logVol/d shift↓
Trajectory ellipsoid (Mahalanobis CP)	0.967	0.275	-2.546	-2.032

Overall, this synthetic stress test confirms that ConjGraph-CP can (i) preserve and recover structured spatio-temporal correlations while (ii) tracking an input-dependent epistemic scale that responds strongly under shift. The conformal variants further demonstrate how κ and the structured covariance can be leveraged for risk-adaptive calibration and coherent trajectory-level sets.

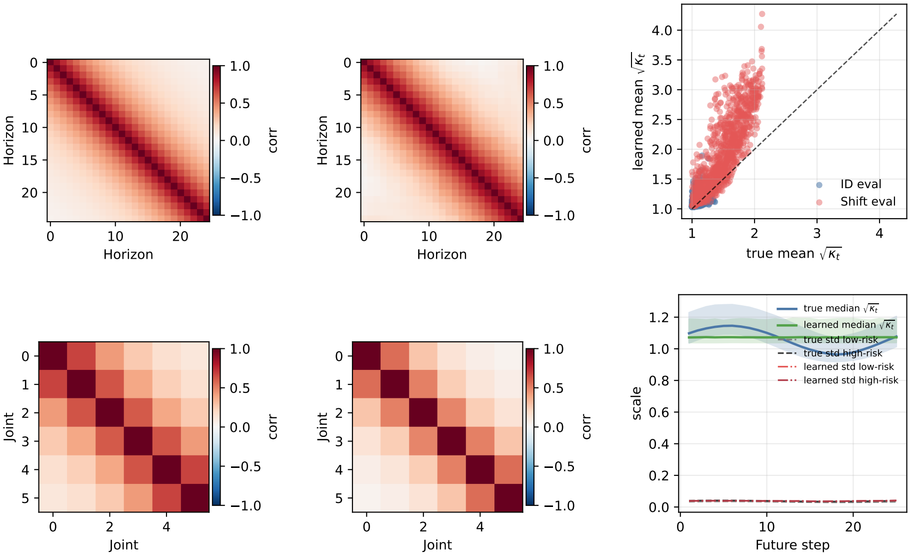

Fig. 3: Synthetic component-recovery benchmark for ConjGraph-CP. **(a)** True temporal correlation. **(b)** Learned temporal correlation. **(c)** True vs. learned epistemic scale ($\sqrt{\kappa}$) on in-distribution and shifted splits. **(d)** True graph-coupled joint correlation. **(e)** Learned joint correlation. **(f)** Risk-profile recovery over horizon.

F Compute and storage overhead

To contextualize the practical cost of uncertainty layers, Table 15 reports parameter counts, additional stored non-parameter state (for the analytic κ statistics), and single-sample inference latency on a representative Human3.6M test sequence (batch size 1, horizon H=25). Absolute times depend on hardware and implementation details; the intent is to quantify the order-of-magnitude overhead relative to the deterministic backbone and to contrast analytic ConjGraph-CP against sampling-based epistemic baselines. For stochastic baselines we measure end-to-end cost to form predictive mean/std from K samples (reported in the table); for the heaviest external baselines we only report GPU timings.

Table 14: Single-sample inference latency for structured uncertainty heads (Human3.6M, H=25, batch size 1).

Model	CPU (ms)	GPU (ms)	GPU peak (MB)
siMLPe (MeanFT)	1.21	2.36	9.8
MatrixNormal	1.42	2.83	9.9
MatrixNormalGraph	1.44	2.91	9.9
GMRF	1.31	2.67	9.8

Table 15: Compute and storage overhead (Human3.6M, H=25, batch size 1). We report the number of samples K used to form predictive mean/std (stochastic baselines), total parameter count, additional stored non-parameter state (e.g., ConjGraph-CP: Λ^{-1} statistics), and single-sample inference latency. External baselines: SkeletonDiffusion [2], BeLFusion [1], TransFusion [5], CoMusion [4], SPARD [7], MotionMap [3], SLD-HMP [6].

Model	K	Params (M)	Extra state (MB)	CPU (ms)	GPU (ms)	GPU pe
siMLPe (base)	1	0.14	0.00	1.23	2.38	
MeanFT (siMLPe fine-tune)	1	0.14	0.00	1.22	2.39	
AuxTasks (50/50)	1	0.14	0.00	1.23	2.39	
Symplectic (50/50)	1	0.14	0.00	1.34	2.63	
DeFeeNet (50/50)	1	0.16	0.00	1.46	2.79	
GSPS (50/50)	7	0.16	0.00	1.32	2.49	
HumanMAC (50/50)	1	0.29	0.00	1.52	2.36	
SeSGCN (teacher)	1	0.44	0.00	2.45	0.64	
Deep ensemble	5	0.69	0.00	6.41	12.08	
Bridge-CQR	1	0.15	0.00	1.29	2.48	
MatrixNormal	1	0.17	0.00	1.43	2.86	
MatrixNormalGraph	1	0.16	0.00	1.45	2.92	
GMRF	1	0.14	0.00	1.32	2.68	
ConjGraph-CP	1	0.16	0.45	2.81	5.49	
SkeletonDiffusion (CVPR 2025)	20	0.31	0.00	–	19.13	
BeLFusion (ICCV 2023)	20	0.71	0.00	–	0.95	
TransFusion (RA-L 2024)	20	15.73	0.00	–	36.59	
CoMusion (ECCV 2024)	20	16.16	0.00	–	16.05	
SPARD (AAAI 2026)	20	32.03	0.00	–	35.52	
MotionMap (CVPR 2025)	20	2.23	0.50	–	3.72	
SLD-HMP (ECCV 2024)	20	0.03	0.00	–	6.85	

References

1. Barquero, G., Escalera, S., Palmero, C.: BeLFusion: Latent diffusion for behavior-driven human motion prediction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 2317–2327 (2023). https://doi.org/10.48550/arXiv.2211.14304, https://arxiv.org/abs/2211.14304 3, 12
2. Curreli, C., Muhle, D., Saroha, A., Ye, Z., Marin, R., Cremers, D.: Nonisotropic gaussian diffusion for realistic 3D human motion prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1871–1882 (2025). https://doi.org/10.48550/arXiv.2501.06035, https://arxiv.org/abs/2501.06035 3, 12
3. Hosseininejad, R., Shukla, M., Saadatnejad, S., Salzmann, M., Alahi, A.: MotionMap: Representing multimodality in human pose forecasting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2025). https://doi.org/10.48550/arXiv.2412.18883, https://arxiv.org/abs/2412.18883 3, 12
4. Sun, J., Chowdhary, G.: CoMusion: Towards consistent stochastic human motion prediction via motion diffusion. In: Proceedings of the European Conference on Computer Vision (ECCV) (2024). https://doi.org/10.48550/arXiv.2305.12554, https://arxiv.org/abs/2305.12554 3, 12
5. Tian, S., Zheng, M., Liang, X.: TransFusion: A practical and effective transformer-based diffusion model for 3D human motion prediction. IEEE Robotics and Automation Letters **9**(7), 6232–6239 (2024). https://doi.org/10.1109/LRA.2024.3401116 3, 12

6. Xu, G., Tao, J., Li, W., Duan, L.: Learning semantic latent directions for accurate and controllable human motion prediction (2024). https://doi.org/10.48550/arXiv.2407.11494, https://arxiv.org/abs/2407.11494 3, 12
7. Zhang, Y., Han, B., Li, X., Pang, W., Giunchiglia, F., Feng, X., Guan, R.: SPARD: Single-step inference with adaptive sampling in residual diffusion for human motion prediction. In: Proceedings of the AAAI Conference on Artificial Intelligence (2026), https://github.com/ZYM-JLU/SPARD, accepted 3, 12