Online Resource 2: Supplementary tables

Article: GlucoKoopman-UQ: Spectrally stable controlled Koopman dynamics for multi-horizon glucose forecasting, glycaemic-risk estimation, and empirical prediction intervals

Journal: Journal of Healthcare Informatics Research

Authors: Jean Marie Vianney Nyirimana, Archana Mathur, Alexander Ngenzi

Corresponding author: Jean Marie Vianney Nyirimana, African Centre of Excellence in Data Science (ACE-DS), College of Business and Economics, University of Rwanda, Kigali, Rwanda; anne4505@gmail.com

Tables S1–S2 describe the source archive; S3 gives event prevalence; S4–S7 give full results for every model at each horizon; S8 gives the Friedman tests; S9 the 15-minute contrasts; S10–S11 participant- and cohort-level results; S12–S13 the completion status of ablations and planned sensitivity runs; S14 training summaries; and S15 model size and inference time. All values come from the evidence files in Online Resource 1.

Table S1 Raw OhioT1DM archive event counts and coverage.

Event group	Events	Participants	Participant coverage (%)	Files	File coverage (%)
acceleration	290,430	6	50.0	12	50.0
basal	1,293	12	100.0	24	100.0
basis_air_temperature	89,381	6	50.0	12	50.0
basis_gsr	379,634	12	100.0	24	100.0
basis_heart_rate	90,020	6	50.0	12	50.0
basis_skin_temperature	379,802	12	100.0	24	100.0
basis_sleep	7,535	12	100.0	24	100.0
basis_steps	90,203	6	50.0	12	50.0
bolus	3,733	12	100.0	24	100.0
exercise	221	10	83.3	19	79.2
finger_stick	4,762	12	100.0	24	100.0
glucose_level	166,533	12	100.0	24	100.0
hypo_event	131	10	83.3	14	58.3
illness	14	8	66.7	10	41.7
meal	2,168	12	100.0	23	95.8
sleep	501	11	91.7	21	87.5
stressors	7	4	33.3	5	20.8
temp_basal	572	12	100.0	21	87.5
work	214	8	66.7	16	66.7

Table S2 Participant-level raw-variable availability.

Table S3 Glycaemic-event prevalence in the held-out test windows.

Horizon (min)	Event	Events	Windows	Prevalence (%)
15	hypoglycaemia	754	28,190	2.67
15	hyperglycaemia	10,041	28,190	35.62
30	hypoglycaemia	753	28,190	2.67
30	hyperglycaemia	10,043	28,190	35.63
45	hypoglycaemia	759	28,190	2.69
45	hyperglycaemia	10,039	28,190	35.61
60	hypoglycaemia	767	28,190	2.72
60	hyperglycaemia	10,020	28,190	35.54

Variable	Mean	SD	Minimum	Maximum
acceleration	0.323	0.355	0.000	0.893
basal	0.007	0.003	0.002	0.012
basis_air_temperature	0.466	0.487	0.000	0.959
basis_gsr	0.789	0.184	0.484	0.958
basis_heart_rate	0.467	0.488	0.000	0.961
basis_skin_temperature	0.789	0.185	0.484	0.959
basis_sleep	0.000	0.000	0.000	0.000
basis_steps	0.122	0.131	0.000	0.305
bolus	0.019	0.007	0.012	0.029
exercise	0.021	0.026	0.000	0.081
finger_stick	0.023	0.012	0.009	0.044
glucose_level	0.887	0.072	0.708	0.967
hypo_event	0.001	0.001	0.000	0.002
illness	8.01e-05	0.000	0.000	0.001
meal	0.011	0.005	0.001	0.018
sleep	0.236	0.112	0.000	0.342
stressors	3.38e-05	8.19e-05	0.000	0.000
temp_basal	0.032	0.047	0.000	0.174
work	0.125	0.102	0.000	0.262

Table S4 Complete participant-level performance at 15 minutes; mean [participant-cluster bootstrap 95% CI].

A. Point accuracy and clinical agreement

Model	MAE	RMSE	MAPE	Clarke A+B (%)
Circadian persistence	57.933 [51.231, 64.724]	75.714 [67.920, 83.717]	40.863 [35.597, 46.439]	84.517 [81.242, 87.643]
DLinear	8.121 [7.598, 8.653]	11.811 [10.844, 12.834]	5.549 [5.003, 6.058]	99.722 [99.615, 99.821]
GlucoKoopman-UQ	7.559 [7.080, 8.062]	11.189 [10.292, 12.177]	5.053 [4.569, 5.518]	99.812 [99.735, 99.887]
KNN-ARX	7.622 [7.096, 8.168]	11.220 [10.283, 12.182]	5.114 [4.587, 5.618]	99.735 [99.628, 99.833]
Persistence	9.555 [8.946, 10.145]	13.546 [12.599, 14.445]	6.372 [5.746, 6.922]	99.688 [99.555, 99.818]
Ridge-ARX	7.396 [6.887, 7.951]	10.990 [10.029, 12.020]	4.993 [4.493, 5.482]	99.764 [99.676, 99.845]

B. Glycaemic-event performance

Model	Hypo recall	Hypo AUPRC	Hyper recall	Hyper AUPRC
Circadian persistence	0.033 [0.005, 0.066]	0.080 [0.036, 0.139]	0.334 [0.255, 0.421]	0.400 [0.325, 0.488]
DLinear	0.537 [0.382, 0.675]	0.627 [0.471, 0.769]	0.928 [0.916, 0.940]	0.975 [0.968, 0.982]
GlucoKoopman-UQ	0.623 [0.508, 0.727]	0.475 [0.339, 0.607]	0.924 [0.910, 0.938]	0.958 [0.949, 0.969]
KNN-ARX	0.580 [0.395, 0.739]	0.614 [0.436, 0.773]	0.926 [0.914, 0.939]	0.967 [0.957, 0.977]
Persistence	0.553 [0.384, 0.704]	0.605 [0.461, 0.741]	0.907 [0.893, 0.922]	0.967 [0.957, 0.975]
Ridge-ARX	0.605 [0.439, 0.743]	0.663 [0.507, 0.799]	0.929 [0.917, 0.942]	0.979 [0.972, 0.985]

C. Prediction-interval performance

Model	PICP	MPIW (mg/dL)
Circadian persistence	0.874 [0.842, 0.905]	234.000 [234.000, 234.000]
DLinear	0.921 [0.908, 0.935]	40.197 [40.197, 40.197]
GlucoKoopman-UQ	0.925 [0.912, 0.936]	39.463 [39.463, 39.463]
KNN-ARX	0.923 [0.909, 0.937]	38.996 [38.996, 38.996]
Persistence	0.917 [0.903, 0.931]	46.000 [46.000, 46.000]
Ridge-ARX	0.924 [0.909, 0.937]	37.804 [37.804, 37.804]

MAE, RMSE, and MPIW are in mg/dL. MAPE and Clarke A+B are percentages. Values preserve the complete confidence intervals from the manuscript result export.

Table S5 Complete participant-level performance at 30 minutes; mean [participant-cluster bootstrap 95% CI].

A. Point accuracy and clinical agreement

Model	MAE	RMSE	MAPE	Clarke A+B (%)
Circadian persistence	58.563 [52.128, 65.736]	75.938 [68.446, 84.355]	41.251 [35.947, 47.359]	84.445 [81.037, 87.582]
DLinear	14.980 [13.978, 15.975]	20.804 [19.257, 22.370]	10.276 [9.191, 11.278]	98.619 [97.942, 99.176]
GlucoKoopman-UQ	13.195 [12.315, 14.103]	18.709 [17.378, 20.047]	8.882 [7.927, 9.801]	99.330 [98.991, 99.593]
KNN-ARX	13.763 [12.805, 14.767]	19.448 [17.987, 20.941]	9.227 [8.236, 10.181]	99.052 [98.720, 99.393]
Persistence	16.834 [15.762, 17.944]	23.225 [21.683, 24.806]	11.218 [10.115, 12.288]	99.059 [98.721, 99.383]
Ridge-ARX	13.793 [12.906, 14.731]	19.354 [17.982, 20.780]	9.345 [8.396, 10.242]	98.992 [98.626, 99.357]

B. Glycaemic-event performance

Model	Hypo recall	Hypo AUPRC	Hyper recall	Hyper AUPRC
Circadian persistence	0.024 [0.000, 0.056]	0.078 [0.034, 0.137]	0.323 [0.243, 0.413]	0.397 [0.321, 0.485]
DLinear	0.281 [0.153, 0.412]	0.411 [0.268, 0.558]	0.865 [0.844, 0.889]	0.926 [0.905, 0.946]
GlucoKoopman-UQ	0.529 [0.406, 0.661]	0.398 [0.260, 0.542]	0.889 [0.870, 0.908]	0.927 [0.909, 0.947]
KNN-ARX	0.351 [0.201, 0.513]	0.404 [0.247, 0.569]	0.855 [0.831, 0.880]	0.897 [0.871, 0.923]
Persistence	0.379 [0.225, 0.536]	0.374 [0.236, 0.522]	0.836 [0.809, 0.864]	0.910 [0.883, 0.934]
Ridge-ARX	0.344 [0.204, 0.478]	0.436 [0.292, 0.586]	0.851 [0.827, 0.880]	0.936 [0.918, 0.954]

C. Prediction-interval performance

Model	PICP	MPIW (mg/dL)
Circadian persistence	0.874 [0.839, 0.904]	234.000 [234.000, 234.000]
DLinear	0.917 [0.901, 0.931]	71.705 [71.705, 71.705]
GlucoKoopman-UQ	0.922 [0.908, 0.936]	67.585 [67.585, 67.585]
KNN-ARX	0.922 [0.907, 0.935]	69.500 [69.500, 69.500]
Persistence	0.913 [0.896, 0.929]	80.000 [80.000, 80.000]
Ridge-ARX	0.922 [0.907, 0.935]	68.094 [68.094, 68.094]

MAE, RMSE, and MPIW are in mg/dL. MAPE and Clarke A+B are percentages. Values preserve the complete confidence intervals from the manuscript result export.

Table S6 Complete participant-level performance at 45 minutes; mean [participant-cluster bootstrap 95% CI].

A. Point accuracy and clinical agreement

Model	MAE	RMSE	MAPE	Clarke A+B (%)
Circadian persistence	59.057 [52.638, 65.819]	76.195 [68.811, 84.323]	41.568 [36.319, 47.346]	84.295 [80.984, 87.540]
DLinear	21.001 [19.534, 22.515]	28.379 [26.326, 30.458]	14.653 [13.048, 16.208]	97.236 [95.743, 98.433]
GlucoKoopman-UQ	18.239 [16.914, 19.620]	25.243 [23.535, 26.999]	12.387 [10.977, 13.842]	98.688 [98.100, 99.197]
KNN-ARX	19.098 [17.736, 20.538]	26.379 [24.492, 28.372]	12.889 [11.486, 14.323]	98.084 [97.307, 98.788]
Persistence	22.973 [21.526, 24.558]	31.294 [29.217, 33.480]	15.393 [13.881, 16.977]	97.994 [97.327, 98.622]
Ridge-ARX	19.425 [18.144, 20.793]	26.477 [24.673, 28.372]	13.287 [11.921, 14.678]	97.659 [96.518, 98.618]

B. Glycaemic-event performance

Model	Hypo recall	Hypo AUPRC	Hyper recall	Hyper AUPRC
Circadian persistence	0.027 [0.003, 0.056]	0.077 [0.035, 0.134]	0.317 [0.238, 0.406]	0.394 [0.319, 0.481]
DLinear	0.126 [0.059, 0.201]	0.293 [0.175, 0.419]	0.801 [0.767, 0.839]	0.864 [0.825, 0.899]
GlucoKoopman-UQ	0.439 [0.286, 0.592]	0.330 [0.199, 0.471]	0.851 [0.823, 0.878]	0.885 [0.856, 0.913]
KNN-ARX	0.229 [0.106, 0.375]	0.291 [0.157, 0.436]	0.788 [0.753, 0.827]	0.826 [0.786, 0.863]
Persistence	0.295 [0.157, 0.443]	0.260 [0.143, 0.402]	0.784 [0.746, 0.819]	0.845 [0.799, 0.884]
Ridge-ARX	0.191 [0.096, 0.298]	0.323 [0.201, 0.453]	0.777 [0.736, 0.822]	0.882 [0.850, 0.912]

C. Prediction-interval performance

Model	PICP	MPIW (mg/dL)
Circadian persistence	0.873 [0.840, 0.903]	234.000 [234.000, 234.000]
DLinear	0.911 [0.892, 0.928]	96.715 [96.715, 96.715]
GlucoKoopman-UQ	0.920 [0.903, 0.935]	90.670 [90.670, 90.670]
KNN-ARX	0.917 [0.900, 0.933]	93.431 [93.431, 93.431]
Persistence	0.910 [0.892, 0.927]	108.000 [108.000, 108.000]
Ridge-ARX	0.919 [0.901, 0.934]	92.605 [92.605, 92.605]

MAE, RMSE, and MPIW are in mg/dL. MAPE and Clarke A+B are percentages. Values preserve the complete confidence intervals from the manuscript result export.

Table S7 Complete participant-level performance at 60 minutes; mean [participant-cluster bootstrap 95% CI].

A. Point accuracy and clinical agreement

Model	MAE	RMSE	MAPE	Clarke A+B (%)
Circadian persistence	59.396 [52.897, 66.285]	76.386 [68.742, 84.702]	41.786 [36.566, 47.634]	84.168 [80.814, 87.426]
DLinear	26.075 [24.246, 27.982]	34.501 [32.113, 37.019]	18.636 [16.491, 20.711]	96.058 [94.062, 97.704]
GlucoKoopman-UQ	22.592 [20.908, 24.333]	30.837 [28.739, 33.035]	15.485 [13.642, 17.318]	97.906 [97.029, 98.668]
KNN-ARX	23.667 [22.011, 25.377]	32.203 [29.965, 34.501]	16.150 [14.344, 17.945]	97.049 [95.841, 98.117]
Persistence	28.185 [26.439, 30.125]	38.020 [35.557, 40.680]	19.035 [17.139, 20.981]	96.877 [95.877, 97.848]
Ridge-ARX	24.315 [22.773, 26.010]	32.463 [30.394, 34.703]	16.829 [15.018, 18.596]	96.412 [94.667, 97.869]

B. Glycaemic-event performance

Model	Hypo recall	Hypo AUPRC	Hyper recall	Hyper AUPRC
Circadian persistence	0.025 [0.003, 0.053]	0.077 [0.035, 0.134]	0.313 [0.232, 0.404]	0.392 [0.314, 0.477]
DLinear	0.067 [0.018, 0.128]	0.231 [0.136, 0.334]	0.747 [0.704, 0.796]	0.801 [0.745, 0.853]
GlucoKoopman-UQ	0.383 [0.253, 0.524]	0.273 [0.153, 0.400]	0.815 [0.780, 0.849]	0.840 [0.802, 0.879]
KNN-ARX	0.168 [0.069, 0.282]	0.210 [0.099, 0.331]	0.729 [0.682, 0.780]	0.761 [0.709, 0.814]
Persistence	0.212 [0.096, 0.338]	0.198 [0.102, 0.315]	0.721 [0.673, 0.769]	0.782 [0.718, 0.837]
Ridge-ARX	0.103 [0.039, 0.175]	0.249 [0.145, 0.358]	0.706 [0.649, 0.766]	0.824 [0.775, 0.870]

C. Prediction-interval performance

Model	PICP	MPIW (mg/dL)
Circadian persistence	0.872 [0.839, 0.903]	234.000 [234.000, 234.000]
DLinear	0.912 [0.892, 0.930]	118.333 [118.333, 118.333]
GlucoKoopman-UQ	0.915 [0.897, 0.932]	109.729 [109.729, 109.729]
KNN-ARX	0.914 [0.895, 0.931]	112.821 [112.821, 112.821]
Persistence	0.908 [0.889, 0.925]	130.000 [130.000, 130.000]
Ridge-ARX	0.917 [0.898, 0.933]	112.972 [112.972, 112.972]

MAE, RMSE, and MPIW are in mg/dL. MAPE and Clarke A+B are percentages. Values preserve the complete confidence intervals from the manuscript result export.

Table S9 Prespecified 15-minute accuracy contrasts retained for transparent reporting.

Metric	Comparator	Mean paired Δ	Holm p	Direction
RMSE	DLinear	-0.621	0.0024	GlucoKoopman-UQ
RMSE	KNN-ARX	-0.030	0.7910	GlucoKoopman-UQ
RMSE	Ridge-ARX	0.200	0.1846	Ridge-ARX
MAE	DLinear	-0.562	0.0029	GlucoKoopman-UQ
MAE	KNN-ARX	-0.063	0.5186	GlucoKoopman-UQ
MAE	Ridge-ARX	0.163	0.0322	Ridge-ARX

Δ = GlucoKoopman-UQ minus comparator in mg/dL. The Ridge-ARX MAE advantage is small but statistically significant; its RMSE advantage is not. No comparator was removed on the basis of performance.

Table S8 Complete Friedman omnibus tests across the six models.

Horizon (min)	Metric	Participants	Q_F	p	Significant
15	RMSE	12	52.429	4.41e-10	Yes
15	MAE	12	52.905	3.52e-10	Yes
15	Hypo AUPRC	12	34.190	2.18e-06	Yes
15	Hypo recall	12	27.694	4.18e-05	Yes
15	Hyper AUPRC	12	54.619	1.56e-10	Yes
15	Hyper recall	12	36.146	8.88e-07	Yes
15	Clarke A+B	12	36.667	6.99e-07	Yes
15	PICP	12	11.238	0.0469	Yes
30	RMSE	12	53.619	2.51e-10	Yes
30	MAE	12	56.714	5.79e-11	Yes
30	Hypo AUPRC	12	23.762	0.0002	Yes
30	Hypo recall	12	41.927	6.09e-08	Yes
30	Hyper AUPRC	12	51.190	7.91e-10	Yes
30	Hyper recall	12	43.696	2.67e-08	Yes
30	Clarke A+B	12	41.523	7.36e-08	Yes
30	PICP	12	23.524	0.0003	Yes
45	RMSE	12	55.952	8.31e-11	Yes
45	MAE	12	56.524	6.34e-11	Yes
45	Hypo AUPRC	12	24.619	0.0002	Yes
45	Hypo recall	12	38.365	3.19e-07	Yes
45	Hyper AUPRC	12	50.571	1.06e-09	Yes
45	Hyper recall	12	43.857	2.48e-08	Yes
45	Clarke A+B	12	45.263	1.28e-08	Yes
45	PICP	12	12.637	0.0270	Yes
60	RMSE	12	55.857	8.70e-11	Yes
60	MAE	12	56.286	7.10e-11	Yes
60	Hypo AUPRC	12	20.667	0.0009	Yes
60	Hypo recall	12	34.799	1.65e-06	Yes
60	Hyper AUPRC	12	52.714	3.85e-10	Yes
60	Hyper recall	12	42.476	4.72e-08	Yes
60	Clarke A+B	12	44.952	1.48e-08	Yes
60	PICP	12	10.489	0.0625	No

Table S10 Participant-specific GlucoKoopman-UQ RMSE across five seeds, mean (SD), mg/dL.

Participant	15 min	30 min	45 min	60 min
540	11.272 (0.243)	20.814 (0.331)	29.833 (0.379)	37.646 (0.326)
544	9.496 (0.269)	15.613 (0.377)	21.130 (0.511)	26.180 (0.489)
552	9.252 (0.491)	17.320 (0.894)	24.796 (0.987)	30.601 (0.856)
559	10.626 (0.132)	17.987 (0.285)	23.964 (0.380)	29.432 (0.333)
563	12.206 (0.258)	18.537 (0.365)	24.605 (0.386)	29.628 (0.424)
567	11.008 (0.581)	20.325 (1.065)	28.629 (1.162)	35.502 (1.099)
570	9.236 (0.153)	15.176 (0.429)	20.308 (0.638)	24.735 (0.801)
575	15.031 (0.278)	22.648 (0.540)	27.734 (0.766)	32.388 (0.782)
584	13.091 (0.459)	21.718 (0.659)	28.815 (0.799)	34.742 (0.785)
588	10.293 (0.156)	17.305 (0.367)	23.527 (0.502)	29.020 (0.663)
591	12.767 (0.169)	20.838 (0.289)	27.627 (0.549)	33.234 (0.724)
596	9.994 (0.108)	16.225 (0.220)	21.944 (0.428)	26.930 (0.600)

Table S11 Descriptive cohort-stratified GlucoKoopman-UQ performance.

Cohort	Horizon (min)	MAE (mg/dL)	RMSE (mg/dL)	Hypo AUPRC	Hyper AUPRC	PICP
2018	15	7.673	11.693	0.546	0.960	0.924
2018	30	13.026	18.749	0.444	0.934	0.926
2018	45	17.654	24.628	0.351	0.898	0.927
2018	60	21.615	29.739	0.264	0.860	0.925
2020	15	7.445	10.686	0.403	0.957	0.926
2020	30	13.363	18.669	0.351	0.920	0.918
2020	45	18.825	25.858	0.309	0.873	0.912
2020	60	23.569	31.934	0.282	0.821	0.905

Each OhioT1DM release contributed six participants. These are descriptive subgroup summaries, not externally validated cohort comparisons.

Table S12 Required ablation grid and completion status.

Type	Configuration	Intervention	Seeds	Status
Feature	`add_behaviour`	Add behavioural-event channels	5/5	Complete
Feature	`add_limited_wearable`	Add cohort-limited wearable channels	5/5	Complete
Feature	`add_rare_events`	Add sparse illness/stressor/hypo-event channels	5/5	Complete
Feature	`add_weekday`	Add weekday encoding	5/5	Complete
Feature	`core_cgm_only`	Retain core CGM-derived inputs only	5/5	Complete
Architecture	`deterministic_head`	Replace probabilistic head with deterministic output	5/5	Complete
Architecture	`no_koopman_control`	Remove exogenous control injection	5/5	Complete
Architecture	`no_multiresolution`	Remove multiresolution encoder	5/5	Complete
Architecture	`no_patient_embedding`	Remove participant embedding	5/5	Complete
Architecture	`no_risk_multitask`	Remove auxiliary glycaemic-risk objective	5/5	Complete
Feature	`no_shared_wearable`	Remove cross-cohort GSR/skin temperature	5/5	Complete
Feature	`no_treatment`	Remove insulin/carbohydrate treatment features	5/5	Complete
Architecture	`unconstrained_dynamics`	Remove Koopman stability constraint	5/5	Complete

Table S13 Optional sensitivity analyses present in the configuration but not executed in the uploaded export.

Sensitivity	Description	Status
context_120m	120-minute historical context	not_run
context_360m	360-minute historical context	not_run
context_60m	60-minute historical context	not_run
hypo_54	Alternative severe-hypoglycaemia threshold (<54 mg/dL)	not_run
lopo	Strict leave-one-participant-out evaluation	not_run
lopo_calibrated	LOPO with held-out calibration partition	not_run
patient_specific	Participant-specific training	not_run

Table S14 Neural-model training summary for every seed.

Table S15 Model size and inference timing recorded by the evaluation pipeline.

Model	Trainable parameters	Latency (ms/window)	Windows/participant
Circadian persistence	0	0.0003	2349.2
DLinear	300	0.0335	2349.2
GlucoKoopman-UQ	162,505	0.4310	2349.2
KNN-ARX	0	0.7465	2349.2
Persistence	0	0.0004	2349.2
Ridge-ARX	0	0.0026	2349.2

Latency is a software measurement from the CPU evaluation environment and should not be interpreted as a hardware-independent deployment benchmark. Classical estimators report zero neural trainable parameters in this accounting.

Model	Seed	Epochs	Best epoch	Best val. loss	Final train loss	Final val. loss	Spectral radius	Sec./epoch
DLinear	17	160	146	10.355	11.236	10.355	–	6.6
DLinear	41	144	130	10.358	11.229	10.358	–	6.2
DLinear	73	159	139	10.355	11.214	10.355	–	6.6
DLinear	101	150	146	10.356	11.175	10.356	–	6.6
DLinear	137	153	149	10.355	11.198	10.355	–	5.6
GlucoKoopman-UQ	17	26	4	9.238	5.538	13.094	0.934	229.6
GlucoKoopman-UQ	41	27	5	9.221	5.484	13.083	0.934	240.2
GlucoKoopman-UQ	73	28	6	9.267	5.431	13.459	0.939	274.3
GlucoKoopman-UQ	101	27	5	9.242	5.498	13.472	0.936	257.0
GlucoKoopman-UQ	137	26	4	9.259	5.576	12.580	0.935	243.1