A Additional Results for Real Dataset

We observed that our method's performance gains over baseline for the Real dataset were more moderate on the Simulated dataset, perhaps due to noise in the carbohydrate signal, as explored in section 5.5. We report results on the Real dataset here, where we see similar trends as observed in the simulated data, but to a lesser degree. **Figure 8** shows how SIV usage varies as a function of SIV magnitude and position. **Figure 9** depicts how improvements over baseline vary across individuals, and **Table 3** contains ablation results for the Real dataset.

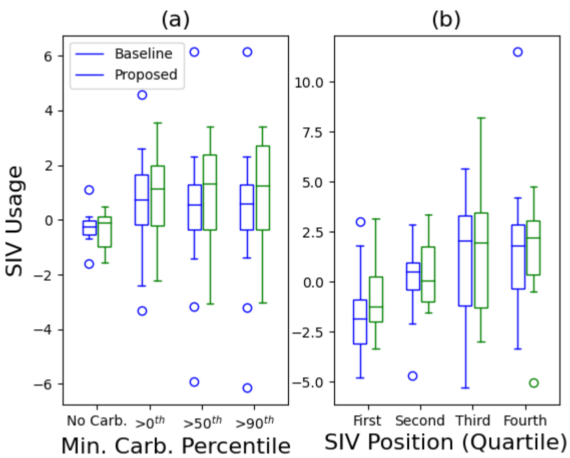

Figure 8: (a) Our architecture and baseline encoder/decoder SIV usage for windows without carbohydrates, windows with any carbohydrates, and windows with carbohydrates above the 50^{th} and 90^{th} percentiles for the Real dataset, all individuals. We see that in general usage and improvement over baseline increase when larger carbs are present. (b) SIV usage for windows who's final non-zero SIV value occur in the first, second, third and fourth quartiles of the input time series for Real data, all individuals. Usage peaks in the third quartile, which is the point at which carbohydrates and boluses most effect the prediction window, given their delay in effects. This effect is less pronounced than in the Simulated dataset.

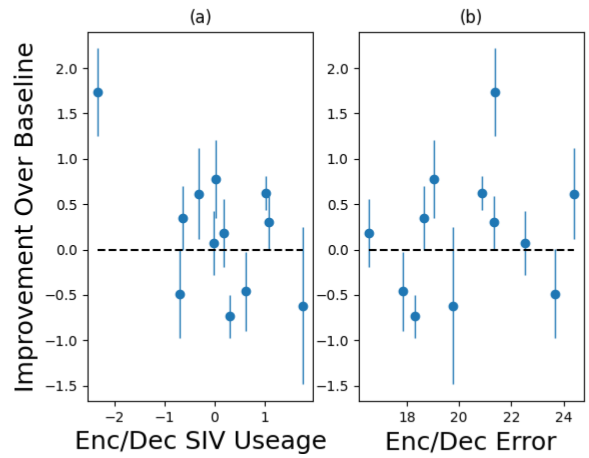

Figure 9: (a) Our architecture's improvement over the encoder/decoder baseline vs baseline SIV usage for the Real dataset. Our method's benefit increases as baseline SIV usage decreases. (b) Improvement over baseline vs baseline prediction error for Real data, for each individual in the simulated dataset. Improvement over baseline is not correlated with baseline error.

Table 3: rMSE and MAE, with SIV usage, for each ablation. Outcomes are reported as: Error [95% confidence interval] (SIV Usage). Confidence intervals were calculated from bootstraps with 1,000 re-samples.

Model	rMSE [95%CI] (Usage)	MAE [95%CI] (Usage)
	Real	
No Gating	20.34,[19.48,21.27] (0.11)	14.77,[14.22,15.33] (0.13)
No Restriction	20.38,[19.48,21.31] (0.06)	14.73,[14.16,15.32] (0.18)
No Dec. SIV Input	20.71,[19.81,21.64] (-0.27)	14.97,[14.41,15.55] (-0.07)
Only Dec. SIV Input	20.52,[19.62,21.47] (-0.08)	14.70,[14.14,15.29] (0.20)
Proposed	20.16,[19.28,21.06] (0.28)	14.64,[14.09,15.20] (0.27)

B Impact of carry-forward approach

Utilizing the Carry-forward approach improves performance on both datasets for both the baseline encoder/decoder and our proposed approach (**Figure 4**).

Table 4: Forecasting Error and SIV usage for both datasets, examining our primary baseline and proposed approach with and without our carry-forward approach. Outcomes are reported as: Error [95% confidence interval] (SIV Usage). Both methods benefit from utilizing the carry-forward approach on both datasets. Confidence intervals were calculated from bootstraps with 1,000 re-samples.

Model	rMSE [95%CI] (Usage)	MAE [95%CI] (Usage)
	Simulated- Carry Forward	
Encoder/Decoder	15.63,[14.08,16.89] (11.13)	12.42,[11.14,13.59] (6.63)
Proposed	**13.07,[11.77,14.16] (13.69)**	**10.45,[9.37,11.37] (8.61)**
	Simulated- NO Carry Forward	
Encoder/Decoder	16.46,[14.64,17.84] (10.30)	12.97,[11.53,14.13] (6.09)
Proposed	16.08,[14.46,17.37] (10.68)	12.80,[11.43,13.91] (6.25)
	Real- Carry Forward	
Encoder/Decoder	20.36,[19.46,21.30] (0.08)	14.67,[14.11,15.24] (0.24)
Proposed	**20.16,[19.28,21.06] (0.28)**	**14.64,[14.09,15.20] (0.27)**
	Real- NO Carry Forward	
Encoder/Decoder	20.64,[19.74,21.56] (-0.20)	14.98,[14.43,15.56] (-0.08)
Proposed	20.41,[19.53,21.35] (0.03)	14.85,[14.28,15.41] (0.05)