university of groningen / faculty of science and engineering

Predictive glucose concentration modeling in Type 1 Diabetes patients by application of Reservoir Computing

Bachelor's Project Thesis

Rafael Bankosegger, s2776758, rafael@bankosegger.at,
Supervisors: Dr. Marco Wiering, Dr. Sietse van Netten

Abstract: Predictive modeling of glucose concentration has the potential to positively influence Diabetes Mellitus therapy by alerting to hyper- or hypoglycemic events and by providing patients with a reasoning framework for short-term treatment decisions. Literature describes the successful application of data-driven techniques such as feed-forward neural networks and support vector regression. In this study, the viability of Reservoir Computing as a novel approach is investigated. This is motivated by the fact that in some instances of time series modeling Reservoir Computing has yielded more accurate results than the above mentioned techniques. The proposed model, an Echo State Network, is based on subcutaneous blood glucose, carbohydrate intake, bolus insulin intake and time of day. Rolling-origin evaluation for forecasting horizons up to 120 minutes was performed on data collected from three patients in free-living conditions. Results show that the Echo State Network consistently outperformed the control model (assuming no change in glucose concentration) by achieving lower error rates. Although these results are promising, the examined model was unable to outperform the control model in terms of clinical accuracy, which was attributed to the network's inability to predict fast changes in glucose concentration.

1 Introduction

The human pancreas produces insulin, a hormone responsible for the human body to process sugar in the blood stream. The pancreas of a diabetes patient has partially or fully lost the ability to produce insulin. The body is unable to absorb glucose. The glucose stays in the blood, and therefore the blood glucose level (BGL) increases. A too high BGL can lead to serious complications for the patient, ranging from organ failure to nerve damage and heart problems. In order to decrease a diabetic's blood glucose level, the patient needs to manually dose and inject insulin, however, injecting an inappropriate amount of insulin may lead to hypoglycemia, meaning the glucose concentration is too low. Hypoglycemia in turn leads to another series of complications, mostly light-headedness, unclear thinking and general weakness. Thus, keeping the glucose concentration in the Euglycemic (healthy) range affords constant self-therapy, monitoring of the blood glucose level, injecting insulin and eating food when needed. As of now, self-therapy can be facilitated by the use of continuous glucose measurement (CGM) devices which measure subcutaneous blood glucose levels with a sample rate of up to 1/5 samples per minute (1 sample every 5 minutes).

However, CGM does not eliminate the constant cognitive load inflicted on the patient by the task of self-monitoring. This load could be lifted by providing patients with tools which are able to predict the short-term glucose concentration trend, thus informing the patient early and accurately enough in order to prevent hyper- and hypoglycemic events.

So far, a series of efforts have been made in order to solve the problem of predicting glucose concentration. In a first investigation of BGL time series data for predictive properties it is shown that there is a significant statistical dependence between individual glycemic measurements. However, the non-

stationarity of the data makes it hard to exploit that statistical dependence for predictions (Bremer and Gough, 1999). Several attempts have been made to exploit that dependence with autoregressive models. More recent research includes (Gani et al., 2009, 2010; Sparacino et al., 2007). In a similar context, a multilayer perceptron was used in (Pérez-Gandía et al., 2010).

Further, it was shown that the inclusion of additional features, such as insulin kinetics, glucose adsorbtion after food consumption, as well as external factors such as exercise data and even time can have a positive influence on the prediction accuracy (Georga et al., 2013). Thus the most promising approach as of now seems to be to apply compartmental, physiological models such as (Dalla Man et al., 2007) for oral glucose absorbtion and (Tarin et al., 2005) for insulin kinetics in combination with some data-driven modeling technique. This has been done by (Georga et al., 2013), where the application of Support Vector Regression has yielded the most promising results found by the authors. A more extensive review of the literature can be found in (Georga et al., 2011).

A model which the author found not yet represented in the literature is the Echo State Network (ESN) as first proposed by (Jaeger, 2001). Echo State Networks have shown to outperform some of the above mentioned algorithms. One of the most prominent examples is the test on the Mackey-Glass series as performed in (Jaeger, 2001), where the Echo State Network is compared to the previous results achieved using a Multilayer Perceptron. Further results can be found for the "Figure 8 problem" and a Japanese vowel classification task (Jaeger et al., 2007) and for the task of equalizing a wireless communication channel (Jaeger and Haas, 2004). Also, Echo State Networks can be combined with any sort of readout algorithm. In other words, even if the network alone does not yield better results, the combination of a reservoir with one of the above mentioned algorithms for readout training might do so. A summary of readout methods can be found in (Lukoševičius and Jaeger, 2009).

The focus of this thesis will be to evaluate the viability of using an ESN for predictive glucose concentration modeling: CGM data, as well as Carbohydrate intake and insulin injections of three subjects under free-living conditions were provided for this study by the company mySugr GmbH, with the consent of the subjects. Based on that, ESNs were trained and evaluated using rolling-origin-recalibration evaluation. The Root Mean Square Error (RMSE) is reported for several combinations of features and a continuous glucose-error grid analysis (CG-EGA) is performed in order to test for clinical accuracy. The model's ability to predict change is tested by investigating the correlation between the prediction error and the actual change in glucose concentration.

	Subject 1	Subject 2	Subject 3
Gender	male	male	female
Age	38	42	34
BMI	21.9	36.7	19.6
HbA1C	8.4	7.6	6.4

Table 2.1: Statistics of subjects. BMI in kg/m^2, HBA1C in % (estimate based on mean BGL).

2 Methods

2.1 Data collection

Three subjects provided their data voluntarily for this research, see Table 2.1 for details. The available data was produced by subjects using the mobile application mySugr over a time period of two years (2015-2017). The data consists of manual log entries including self-monitored blood glucose level (SMBG), carbohydrate intake from meals, injected insulin units and subcutaneous glucose concentration samples that were automatically measured by continuous glucose measurement (CGM) devices.

2.2 Data processing

First, the CGM time series were split into subsets, with individual measurements being no further than 15 minutes apart from each other within the subsets. Given the sample rate of one sample every five minutes (12 samples per hour), subsets thus contain gaps of maximally two samples missing. Those missing samples were artificially filled in using linear interpolation of the previous sample and the next sample.

Daily event data, such as meal carbohydrates and bolus insulin intake have a much lower sample rate. Those samples thus had to be resampled to fit the

	Subject 1	Subject 2	Subject 3
N	9000	86940	16320
N_{sub}	375 ± 316	318 ± 242	281 ± 132
CGM	155 ± 49	164 ± 64	141 ± 46
INS	2.4 ± 1.8	9.7 ± 7.0	4.5 ± 3.9
CARB	46 ± 27	60 ± 32	34 ± 21

Table 2.2: Statistics of the subsets after filtering zero-values for insulin and food entries.

CGM time series. This was done by assigning each of the entries to their closest available CGM sample. In case the distance between the entry and the closest CGM sample was more than 5 minutes, the entry was discarded. In case there were multiple entries of the same type having the same closest sample, their sum was taken in order to account for the possibility of having multiple food / insulin entries in a short time period. The result of this is a multivariate time series containing the variables CGM, meal carbohydrates and bolus insulin intake with a sample rate of 12 samples per hour. Empty entries for meal carbohydrates and insulin were set to zero.

Next, the subsets were split into chunks of 20 samples (100 minutes). For each chunk, it was checked whether it contained non-zero values for food or insulin after applying the later described models for food digestion and insulin injection. If not, the chunk was removed since allowing those zero-values would corrupt the training process of insulin- and carb-based models. Then, temporally adjacent chunks were joined together into new subsets.

The dataset summary for the subsets after filtering can be seen in Table 2.2. N describes the total number of available samples. N_{sub} summarizes the number of samples per subset. CGM, INS, CARB summarize the variables for subcutaneous glucose concentration (mg/dl), insulin intake (insulin units IU) and carbohydrate intake (g), respectively. Descriptions are of the form: mean $\pm$ standard deviation.

2.3 Feature engineering

2.3.1 Insulin injection model

(Tarin et al., 2005) developed a pharmacokinetic model, which was taken as a basis for the model used here. The study identified two sub-models: the absorption model $I_{ex}(t)$, describing the exogenous insulin flow into the blood stream and the insulin model $i_p(t)$, describing the concentration-time evolution of the exogenous plasma insulin. The model has a considerable amount of parameters that change with the type of used insulin agent. In this study, the insulin agent used by the patients were unknown and needed to be fitted for the patients. In order to simplify that fitting procedure, an approximation of the original formula was deployed, which allowed the modeling of differences in insulin agent by changing just one parameter:

$$I_n^{\text{msk}} = \frac{e^{-n*p_0^I} - e^{-n*p_0^I*p_1^I}}{1 + e^{(n-p_2^I)*p_3^I}} \tag{2.1}$$

$$\hat{\boldsymbol{I}}^{\text{ex}} = \boldsymbol{I}^{\text{in}} * \frac{\boldsymbol{I}^{\text{msk}}}{|\boldsymbol{I}^{\text{msk}}|} \tag{2.2}$$

$$p_1^I = 1.5,\ p_2^I = 10,\ p_3^I = 0.2 \tag{2.3}$$

$\boldsymbol{I}^{\text{msk}}$ was based on the solution of a differential equation and was used as a mask for convolution ($*$) with $\boldsymbol{I}^{\text{in}}$, the time series of insulin intakes as logged by the patient. A sigmoid function was applied in order to ensure that the mask ultimately reached zero again.

The parameters p_1^I, p_2^I and p_3^I were fixed. The parameter p_0^I was designed to vary with the used insulin agent and thus was treated as an optimizable hyper-parameter. Figure 2.1 shows the effect of changing p_0^I.

Note that similarly simple models were also discussed by other authors (Cescon, Johansson, Renard, and Maran, 2014; Jørgensen, Huusom, and Sin, 2012) as cited in (Kirchsteiger, Jørgensen, Renard, and Del Re, 2016, p.7) which however had no influence on the development of the model used in this research because they were only found after the research discussed here was already completed.

2.3.2 Food digestion model

As with the insulin model, the glucose rate of appearance ***Ra*** after a meal was modeled in a simplified manner. The deployed model, based on the model of oral glucose absorption described in (Dalla Man et al., 2007), was inspired by the solution of a differential equation and approximated

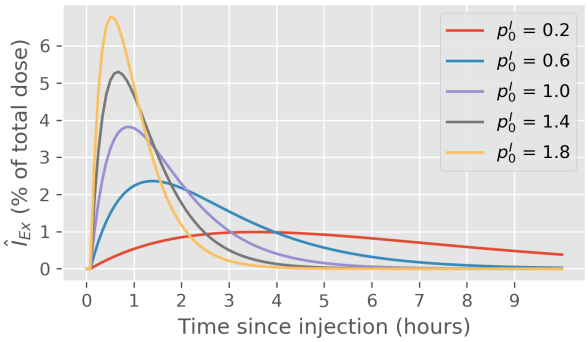

Figure 2.1: Unit impulse response of the insulin injection model output $\hat{\boldsymbol{I}}_{\boldsymbol{ex}}$ when using different values for p_0^I.

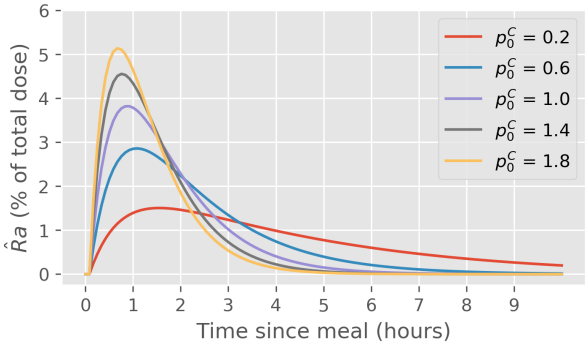

Figure 2.2: Unit impulse response of the food digestion model output $\hat{\boldsymbol{Ra}}$ when using different values for p_0^C.

the model with less parameters and using convolution. A sigmoid function was applied in order to ensure that the mask became zero for large n.

$$Ra_n^{\text{msk}} = \frac{e^{-n*p_0^C} - e^{-n*p_1^C}}{1 + e^{(n-p_2^C)*p_3^C}} \tag{2.4}$$

$$\hat{\boldsymbol{Ra}} = \boldsymbol{G}^{\text{in}} * \frac{\boldsymbol{Ra}^{\text{msk}}}{|\boldsymbol{Ra}^{\text{msk}}|} \tag{2.5}$$

$$p_1^C = 6,\ p_2^C = 4,\ p_3^C = 0.4 \tag{2.6}$$

The parameters p_1^C, p_2^C and p_3^C were fixed. p_0^C was treated as an optimizable hyper-parameter. Figure 2.2 shows the effect of changing p_0^C.

2.3.3 Time model

The circadian rhythm as well as habits could also influence the blood glucose level, for example by consistent alcohol consumption on Friday nights or fixed meal times on workdays. The time model described in equations 2.7 to 2.9 was used in order to take those influences into account.

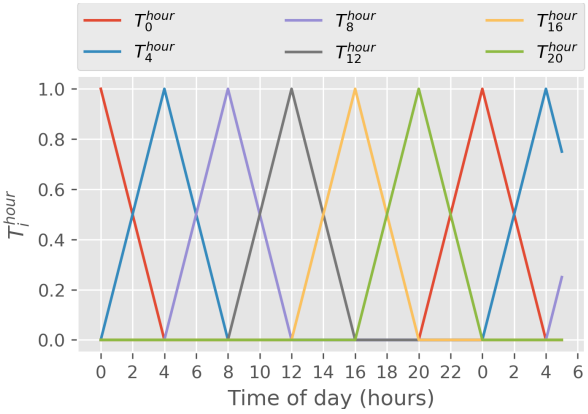

Figure 2.3: The six outputs of the time model that depend on the hour of day. Every fourth hour, a distinct output reaches its peak at 1.0 while the others become 0.0. The plot also illustrates how each output repeats after 24 hours.

For the model, $h\,(mod\ 24)$ was defined as the time of day in hours of any given CGM sample and $d\,(mod\ 7)$ was defined as the day of week of any given CGM sample. Both h and d are floating-point numbers and precise to the minute. Note the special case of $T_0^{\text{hour}}(h)$, which needed to be nonzero whenever $h < 4$ and $h > 20$. The solution was to subtract 24 such that $-4 \leq h \leq 4$. The special case $T_0^{\text{day}}(d)$ was similarly handled by subtracting 7. Figures 2.3 and 2.4 show how the 13 dimensions of $\boldsymbol{T}(n)$ behave over time.

$$T_{h_0}^{\text{hour}}(h) = \begin{cases} \max(0, 1 + \frac{h-h_0}{4}) & \text{if } h \leq h_0 \\ \max(0, 1 - \frac{h-h_0}{4}) & \text{otherwise} \end{cases} \tag{2.7}$$

$$T_{d_0}^{\text{day}}(d) = \begin{cases} \max(0, 1 + \frac{d-d_0}{1}) & \text{if } d \leq d_0 \\ \max(0, 1 - \frac{d-d_0}{1}) & \text{otherwise} \end{cases} \tag{2.8}$$

$$\boldsymbol{T}(h, d) = [\boldsymbol{T}_{(0,4,8,12,16,20)}^{\text{hour}}; \boldsymbol{T}_{(0,1,2,3,4,5,6)}^{\text{day}}] \tag{2.9}$$

2.4 Proposed prediction model

The basis for the model used in this study was the Echo State Network (ESN) using Leaky Integrator Neurons as explored in (Jaeger et al., 2007). The

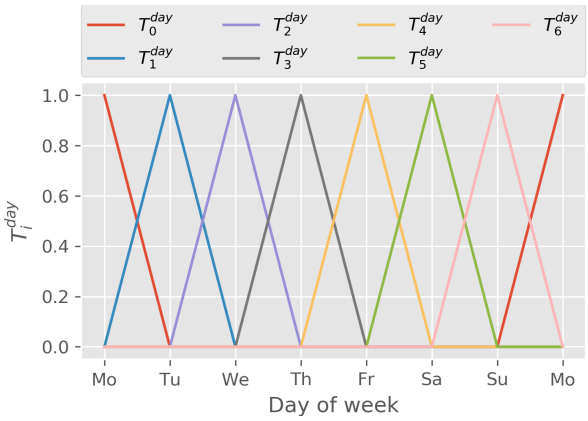

Figure 2.4: The seven outputs of the time model that depend on the day of week. Every day, a distinct output reaches its peak at 1.0 while the other dimensions become 0.0. The plot also illustrates how each output repeats after seven days. Monday to Sunday are mapped to the numbers 0 to 6.

goal of the proposed model is to predict not only time series subcutaneous glucose concentration, but also all time series for above described features.

$$\boldsymbol{x'}_{n+1} = f(\boldsymbol{W^{in}}\boldsymbol{u}_{n+1} + \boldsymbol{W}\boldsymbol{x}_n + \boldsymbol{W^{fb}}\boldsymbol{y}_n) \quad (2.10)$$

$$\boldsymbol{x}_{n+1} = (1 - \frac{a\delta}{c})\boldsymbol{x}_n + \frac{\delta}{c}\boldsymbol{x'}_{n+1} \quad (2.11)$$

$$\boldsymbol{y}_n = g(\boldsymbol{W^{out}}[\boldsymbol{x}_n; \boldsymbol{u}_{n\delta}]) \quad (2.12)$$

Some modifications were made to that model. First, (Jaeger et al., 2007) assume a uniform leaking rate for simplicity. However, given the highly dynamic and unstationary nature of the problem, introducing a non-uniform leaking rate might improve the modeling capabilities of the ESN. Thus, the leaking rate (originally denoted as a) was changed to be a vector $\boldsymbol{a}$ where the leaking rate of each individual neuron a_i was drawn from a uniform distribution ranging from a_{min} to a_{max}. Leakage of the previous state is determined by element-wise multiplication ($\odot$) of $\boldsymbol{a}$ and $\boldsymbol{x}_n$.

Also, the original model made a distinction between the external input $\boldsymbol{u}_n$ and the output vector $\boldsymbol{y}_n$. In this study, the model was used in a way where the model output is a prediction of the model's next input. Thus, $\boldsymbol{u}_n$ was not necessary and could be discarded, along with the input weights W^{in}.

f was set to the identity activation function which, as suggested in (Millea, 2014), delivers better accuracy. g was set to be the identity function as well, however, values below zero or above one were cropped. Since the time series is neither oversubsampled, $\delta = 1$. Finally, a bias was added to the input signal. After applying those changes, the final model can be summarized as follows:

$$\boldsymbol{x'}_{n+1} = \boldsymbol{W}\boldsymbol{x}_n + \boldsymbol{W^{fb}}\boldsymbol{y}_n \quad (2.13)$$

$$\boldsymbol{x}_{n+1} = (\boldsymbol{1} - \frac{1}{c}\boldsymbol{a}) \odot \boldsymbol{x}_n + \frac{1}{c}\boldsymbol{x'}_{n+1} \quad (2.14)$$

$$\boldsymbol{y}_{n+1} = g(\boldsymbol{W^{out}}[\boldsymbol{1}; \boldsymbol{y}_n; \boldsymbol{x}_{n+1}]) \quad (2.15)$$

$$g(\boldsymbol{y'_i}) = \min(1, \max(0, \boldsymbol{y'_i})) \quad (2.16)$$

2.4.1 Weight initialization

In order to construct the reservoir weight matrix W, a two-dimensional matrix of the size $N \times N$ (where N is the number of neurons in the reservoir) was drawn from a uniform distribution of the interval $[-0.5, 0.5]$. From the resulting matrix, a linearly independent matrix of the same size was constructed as described in (Millea, 2014). The same study also stated that, more than the echo state property, the input-weight to reservoir-weight ratio was vital to the network's success. To account for that, the matrix W was multiplied by a scaling coefficient. Given that the scaling coefficient needed to be optimized anyway, the division by the largest eigenvalue as done by (Jaeger et al., 2007) was disregarded. Also, a matrix of the same size, drawn from a binomial distribution (with entries being either zero or one) was element-wise multiplied in order to lower connectivity, the amount of nonzero weights in W (Millea, 2014). This introduced three parameters: the reservoir size, the reservoir weight scaling and the connectivity.

In order to construct W^{fb}, a matrix of the size $(N \times M)$ (where M is the output size) was drawn from a uniform distribution of the interval $[-1, 1]$. Every element of W^{fb} was then individually scaled, depending on a scale parameter set for the input type that element corresponds to. For example, elements corresponding to time input are scaled by a time scale parameter, but elements corresponding to food intake are scaled by a food scale parameter. This introduced five new parameters for scaling the

bias, cgm, glucose rate of appearance after meals, exogenous insulin flow after injections and time.

2.4.2 Training the output weights

W^{out} needed to be determined such that $y_n \approx t_n$. In this study, Ridge regression as defined in (Lukoševičius, 2012) seemed to be the default method of training ESNs and thus was used in this study as well, see equation 2.17.

To determine the state and target matrices $\boldsymbol{X}$ and $\boldsymbol{Y}^{\text{target}}$, teacher forcing was used as described in (Lukoševičius, 2012). Accordingly, equation 2.13 was replaced by equation 2.18, for which $\boldsymbol{t}_n$ was defined as the actual observed values for the features at time step n. β was defined as the regularization coefficient and I as the identity matrix.

The network weights were initialized and the network was run on a given training set $\boldsymbol{t_0} \dots \boldsymbol{t_{j-1}}$, yielding the state vectors $\boldsymbol{x_1} \dots \boldsymbol{x_j}$. The first i state vectors were discarded as initialization period. This was needed to wash out the ESN' initial state $\boldsymbol{x}_0 = \boldsymbol{0}$. The individual rows of $\boldsymbol{X}$ and $\boldsymbol{Y}$ were constructed from the remaining states, as described in equations 2.19 and 2.20

The teacher signal $\boldsymbol{t}_n$ was created by concatenating all used features and a bias of 1 into a vector. In order to explore the effectiveness of certain features, multiple combinations of features were considered, the teacher signal for each combination can be seen in table 2.3.

Features	$\boldsymbol{t}_n =$	$M =$
None	$[1; t_n^{\text{cgm}}]$	2
Time	$[1; t_n^{\text{cgm}}; \boldsymbol{t}_n^{\text{time}}]$	15
Carbs	$[1; t_n^{\text{cgm}}; t_n^{\text{carbs}}]$	3
Bolus	$[1; t_n^{\text{cgm}}; t_n^{\text{bolus}}]$	3
All	$[1; t_n^{\text{cgm}}; t_n^{\text{carbs}}; t_n^{\text{bolus}}; \boldsymbol{t}_n^{\text{time}}]$	17

Table 2.3: Teacher signals, composed of different features. The signals correspond to the feature engineering models as follows: $t_n^{\textbf{carbs}} = Ra_n$, $t_n^{\textbf{bolus}} = I_n^{ex}$ and $\boldsymbol{t}_n^{\textbf{time}} = \boldsymbol{T}(h_n, d_n)$. M denotes the number of dimensions for each teacher signal.

$$W^{out} = \boldsymbol{Y}^{\text{target}} \boldsymbol{X}^T (\boldsymbol{X}\boldsymbol{X}^T + \beta \boldsymbol{I})^{-1} \tag{2.17}$$

$$\boldsymbol{x}'_{n+1} = \boldsymbol{W}\boldsymbol{x}_n + \boldsymbol{W^{fb}}\boldsymbol{t}_n \tag{2.18}$$

$$\boldsymbol{X}_n = [\boldsymbol{1}; \boldsymbol{t}_{n-1}; \boldsymbol{x}_n] \tag{2.19}$$

$$\boldsymbol{Y}_n^{\text{target}} = [\boldsymbol{t}_n] \tag{2.20}$$

3 Results

3.1 Validation method

The literature usually reports cross validation as evaluation technique for glucose concentration predictors. However, external validity should be ensured by making new predictions only with old samples: those that appeared before the prediction is performed. Cross validation does not sustain that chronological order. Data from the future may be used to train for a prediction done in the present. Instead, rolling-origin-recalibration evaluation as mentioned in (Tashman, 2000) and (Bergmeir and Benítez, 2012) was used here. It ensures external validity of future predictions by only allowing samples in the past to influence future predictions. Furthermore, leaving out one set between train-sets and test set introduces a time gap which makes sure that train and test data are not autocorrelated, which further increases generalizability.

All subsets of a single subject were ordered by time (earliest first). Then, the first subset is taken as train-set. Reservoir states are collected and output weights are trained as described in section 2.4.2. Next, the third subset is taken as test set and a rollout is generated for each sample $n_{\text{test}} \in [i, N - 24]$ in it, where i refers to the initialization period and N refers to the length of the subset. For each rollout, the network is run with teacher forcing for samples $0 \dots n_{\text{test}}$. Then the actual equations of the ESN are used to generate the predictions $y_{n_{\text{test}}+1} \dots y_{n_{test}+24}$. Thus, the rollout spans a prediction horizon of 5 min up to 120 min.

After all possible rollouts of the third subset are created, the second subset is added to the train-set and the fourth subset is used to generate the next set of rollouts. Next, the network is retrained on subsets one, two and three and tested on subset five. This pattern is repeated until all but the last two subsets are included in the train set and the last subset is used for rollout generation.

3.2 Optimized parameters

Using the root mean squared error (RMSE) as cost function, the above described hyper-parameters were optimized via a random sweep. Table 3.1 shows the hyper-parameters with the best fit, i.e. the ones that achieved the lowest RMSE scores, see next section.

3.3 Comparing RMSE

The obtained predictions are compared against the target signal. The RMSE is assessed for the prediction horizons of 15, 30, 60 and 120 minutes. As a baseline for comparison, a control model predicting no change, i.e. $y_{n_{\text{test}}+1\ldots24} = y_{n_{\text{test}}}$, was deployed.

The RMSE for all predictions are shown in table 3.2. As a baseline, the control model results are quite similar between subjects, ranging from $13mg/dl$ to $14mg/dl$ for a 15 minutes prediction horizon and reaching a maximum RMSE of $58.86mg/dl$ for 120 minutes. The ESN generally achieves lower RMSE scores than the control model. Predictions of 15 minutes for all ESN-based predictions show errors of approximately $10mg/dl$ for subject 1 and $11mg/dl$ for subjects 2 and 3. For higher prediction horizons, the difference in prediction errors between subjects becomes less strong. Across all prediction observed horizons, the prediction error difference between feature sets is negligible, i.e. their errors are about equal.

3.4 CG-EGA

While RMSEs provide a measurement of the systems's general accuracy, it does not a give clear insight into clinical accuracy. A continuous glucose-error grid analysis (CG-EGA) as in (Clarke et al., 2008) is applied to account for that. The analysis is performed for the control model, the ESN with no features and the one with all features. Results per model were obtained by combining the results for all patients. Tables 3.3, 3.4, 3.5 show the results of the prediction error grid analysis.

Predictions are classified as either AR, BE or ER. AR is a clinically accurate prediction. BE is a prediction with benign error: the prediction does not lead to the clinically correct action but also does not lead to any dangerous situations. ER is an erroneous prediction: a prediction that leads to or causes missing of some action and thus causes a potentially dangerous situation. The tables show the relative occurrence of each prediction class across clinically relevant events (hypoglycemia, euglycemia and hyperglycemia) and several prediction horizons.

As a baseline, the control model performs remarkably well, with 94.06% of the predictions in euglycemic range and 83.64% of the predictions in hyperglycemic range being classified as clinically accurate for a prediction horizon of 120 minutes. For hypoglycemic events, 15.58% are classified as clinically accurate for a 120 minute prediction horizon. Comparable accuracy is achieved for collection 1 (the unfiltered data).

The results for the ESN using no features and for the ESN using all features are again very similar, with some variation: the model without features is slightly more accurate in predicting hyperglycemic events (85.75%) while the model with all features is slightly more accurate in predicting hypoglycemic events (3.77%).

However, comparing both ESNs to the baseline, the control model outperforms both ESNs in predicting long-term hyperglycemia. For euglycemia and hyperglycemia the ESN predictions are generally slightly better than the baseline.

3.5 Recognizing fast changes

The similar accuracy of the control model and the ESNs suggests that most of the correctly predicted data contains little change in glucose concentration. In order to investigate how the models perform with regard to fast changes in the data, such as sudden hypoglycemic events, the Pearson correlation between prediction error and rate of change was investigated. The correlation coefficient for 30 minutes is $r^{30min}_{\text{None}} = 0.843$ for the model without features and $r^{30min}_{\text{All}} = 0.847$ for the model with all features included. For 60 minutes, $r^{60min}_{\text{None}} = 0.868$ and $r^{60min}_{\text{All}} = 0.866$

Thus there is strong correlation between change in glucose concentration and size of error.

4 Discussion

In this thesis, a study on predictive glucose concentration modeling in type 1 diabetes patients was

	None	Carbs	Bolus	Time	All
Food model p_0^{C}	-	1.74	-	-	0.80
Insulin model p_0^{I}	-	-	0.17	-	0.08
Maximum leaking rate a_{max}	0.74	0.32	0.96	0.32	0.59
Minimum leaking rate a_{min}	0.92	0.63	0.97	0.63	0.62
Connectivity	0.39	0.11	0.01	0.11	0.82
Time scale $\frac{1}{c}$	1.35	1.88	0.76	1.88	2.29
Regularization coefficient β	0.21	0.11	1.52	0.11	0.04
Weight scale reservoir	0.86	0.20	1.46	0.20	0.29
Weight scale input bias	0.14	1.80	0.98	1.80	2.28
Weight scale input cgm	0.53	0.28	2.13	0.28	0.96
Weight scale input food	-	0.17	-	-	2.77
Weight scale input insulin	-	-	1.63	-	2.36
Weight scale input times	-	-	-	0.17	1.97

Table 3.1: Optimized parameters for features (same across all patients)

performed. The Echo State Network (ESN) was proposed as a novel approach to this problem and several feature combinations were evaluated.

The results show that the ESNs generally outperform the control model predicting no changes at all by achieving lower errors. However, the addition of features did not yield significant improvements in accuracy. This is contrary to other findings in the literature where these features are reported as having a significant effect.

The CG-EGA showed that the large majority of predictions in the euglycemic and hyperglycemic range lead to clinically correct predictions even after two hours. However, this accuracy needs to be attributed to the small amount of clinically relevant change in the dataset as proven by similar results for the control model. The ESNs are generally not able to give clinically more accurate predictions than the control model.

An investigation into how well the algorithm performs with drastic blood glucose changes revealed that the ESN error is correlated with the rate of change. In other words, the more the glucose concentration changes, the higher the error. Thus, the model is generally unable to predict fast changes in blood glucose concentration.

This begs the question whether the Echo State Network is unsuitable for this task in general. To investigate this, the assumption that the hyperparameters do not change over time and across patients needs to be questioned. The assumption was crated in order to ensure external validity, but in hindsight this assumption should be reconsidered. Given the non-stationarity of the glucose concentration time series as mentioned in (Bremer and Gough, 1999), it is improbable that the parameters do not change over time.

To account for that an on-line parameter optimization approach similar to the gradient descent algorithm used by (Jaeger et al., 2007) may be more appropriate. Similarly, it may be more appropriate to use a sliding window of training data instead of using all data from the past, at least when using one-shot training methods such as Ridge Regression.

In order to address the fact that event data had no influence, it has to be asked whether the data is simply not accurate enough in order to have impact on the prediction. This needs to be asked since daily event data was logged by the patients themselves, introducing a potential source of error. A similar question was investigated in (Bunescu et al., 2013) by asking physicians to perform predictions comparing their performance to the baseline. The utility of the physiological features in that study was confirmed and proven to yield significant improvements in RMSE.

That being said, it is more probable that the data is accurate enough but not used optimally. A possible way to use the data more appropriately would be to design separate echo state networks for each feature. This is because every feature has its own inherent dynamics and putting them all into just one echo state network probably exceeds the mod-

Features	Subject	15min	30min	60min	120min
Control	Subj. 1	13.02	23.33	39.31	58.86
	Subj. 2	13.19	22.79	37.55	57.50
	Subj. 3	13.98	24.25	38.58	54.14
None	Subj. 1	9.81	18.82	32.83	46.68
	Subj. 2	11.01	19.73	33.25	49.78
	Subj. 3	11.14	20.48	33.15	43.56
Time	Subj. 1	9.92	19.13	33.65	48.67
	Subj. 2	11.01	19.74	33.31	49.82
	Subj. 3	11.19	20.63	33.57	44.68
Carbs	Subj. 1	9.83	18.87	33.05	47.04
	Subj. 2	10.98	19.65	33.16	49.81
	Subj. 3	11.10	20.34	32.84	43.48
Bolus	Subj. 1	9.92	18.92	33.05	47.03
	Subj. 2	11.03	19.77	33.34	49.89
	Subj. 3	11.28	20.65	33.39	43.94
All	Subj. 1	9.96	19.35	33.95	48.87
	Subj. 2	11.01	19.72	33.72	55.60
	Subj. 3	11.19	20.68	33.98	46.16

Table 3.2: RMSE (mg/dl) across subjects and time

	Hypoglycemia			Euglycemia			Hyperglycemia		
	AR (%)	BE (%)	ER (%)	AR (%)	BE (%)	ER (%)	AR (%)	BE (%)	ER (%)
15 min	83.70	1.72	14.58	93.90	6.10	0.00	92.42	2.95	4.63
30 min	56.65	1.56	41.79	93.91	6.09	0.00	92.28	2.92	4.81
60 min	31.03	1.01	67.97	94.02	5.98	0.00	90.49	2.74	6.77
90 min	20.93	0.51	78.56	94.07	5.93	0.00	87.26	2.63	10.12
120 min	15.58	0.39	84.03	94.06	5.94	0.00	83.64	2.58	13.77

Table 3.3: CG-EGA control model (no change)

eling capability of a single ESN. A next step thus may be to build an ensemble of ESNs with each only using one feature as input.

Finally, some tweaks to the feature models may increase performance, although it is unlikely that these tweaks are sufficient for fixing the problem of predicting fast changes in glucose concentration. This research assumes that food and insulin model parameters do not change across patients and over time. This assumption is not accurate. For example, the effect of the insulin model depends upon others on the used insulin agent and the weight of the subject, which are of course different per person.

5 Conclusion

In this study, using Echo State Networks for predictive glucose concentration modeling was proposed and evaluated for the first time in the literature, including feature exploration of external event data for food and insulin intake as well as time. The results show that our proposed method yields better results than the baseline in terms of RMSE. However, the model generally is unable to predict fast changes such as transitions between hyperglycemic, euglycemic and hypoglycemic phases, resulting in low clinical accuracy especially for predicting hypoglycemic phases. External event data does not have relevant effects on the performance. However, we do not fully reject the Echo State Network as a possible solution to the problem due to the be-

	Hypoglycemia			Euglycemia			Hyperglycemia		
	AR (%)	BE (%)	ER (%)	AR (%)	BE (%)	ER (%)	AR (%)	BE (%)	ER (%)
15 min	86.97	1.47	11.55	95.95	3.89	0.16	94.40	1.94	3.66
30 min	52.94	1.11	45.94	95.39	4.58	0.03	93.38	1.63	5.00
60 min	14.63	0.71	84.65	95.09	4.82	0.09	91.36	1.59	7.05
90 min	4.90	0.30	94.80	94.68	5.09	0.22	87.30	1.93	10.77
120 min	3.77	0.43	95.80	94.44	5.36	0.20	82.16	2.04	15.79

Table 3.4: CG-EGA all features (ESN)

	Hypoglycemia			Euglycemia			Hyperglycemia		
	AR (%)	BE (%)	ER (%)	AR (%)	BE (%)	ER (%)	AR (%)	BE (%)	ER (%)
15 min	87.14	1.64	11.22	95.87	4.00	0.13	94.45	1.97	3.58
30 min	53.56	1.32	45.12	95.41	4.58	0.01	93.65	1.63	4.72
60 min	11.32	0.34	88.34	95.52	4.48	0.00	92.17	1.46	6.37
90 min	1.19	0.00	98.81	95.39	4.61	0.00	89.42	1.49	9.09
120 min	0.04	0.00	99.96	95.20	4.80	0.00	85.75	1.70	12.55

Table 3.5: CG-EGA no features (ESN)

lieve that adjustments to the hyper-parameter optimization strategy and by deploying an ensemble of ESNs, each optimized for just one feature, might yield more accurate results.

References

Christoph Bergmeir and José M Benítez. On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191:192–213, 2012.

Troy Bremer and David A Gough. Is blood glucose predictable from previous values? a solicitation for data. *Diabetes*, 48(3):445–451, 1999.

Razvan Bunescu, Nigel Struble, Cindy Marling, Jay Shubrook, and Frank Schwartz. Blood glucose level prediction using physiological models and support vector regression. In *Machine Learning and Applications (ICMLA), 2013 12th International Conference on*, volume 1, pages 135–140. IEEE, 2013.

Marzia Cescon, Rolf Johansson, Eric Renard, and Alberto Maran. Identification of individualised empirical models of carbohydrate and insulin effects on T1DM blood glucose dynamics. *Int. J. Control*, 87:1438–1453, 2014.

William L Clarke, Stacey Anderson, and Boris Kovatchev. Evaluating clinical accuracy of continuous glucose monitoring systems: Continuous glucose-error grid analysis (CG-EGA). *Current diabetes reviews*, 4(3):193–199, 2008.

Chiara Dalla Man, Robert A Rizza, and Claudio Cobelli. Meal simulation model of the glucose-insulin system. *IEEE Transactions on biomedical engineering*, 54(10):1740–1749, 2007.

Adiwinata Gani, Andrei V Gribok, Srinivasan Rajaraman, W Kenneth Ward, and Jaques Reifman. Predicting subcutaneous glucose concentration in humans: data-driven glucose modeling. *IEEE Transactions on Biomedical Engineering*, 56(2):246–254, 2009.

Adiwinata Gani, Andrei V Gribok, Yinghui Lu, W Kenneth Ward, Robert A Vigersky, and Jaques Reifman. Universal glucose models for predicting subcutaneous glucose concentration in humans. *IEEE Transactions on Information Technology in Biomedicine*, 14(1):157–165, 2010.

Eleni I Georga, Vasilios C Protopappas, and Dimitrios I Fotiadis. Glucose prediction in type 1 and type 2 diabetic patients using data driven techniques. In *Knowledge-oriented applications in data mining*. InTech, 2011.

Eleni I Georga, Vasilios C Protopappas, Diego Ardigò, Michela Marina, Ivana Zavaroni, Demosthenes Polyzos, and Dimitrios I Fotiadis. Multivariate prediction of subcutaneous glucose concentration in type 1 diabetes patients based on support vector regression. *IEEE journal of biomedical and health informatics*, 17(1):71–81, 2013.

Herbert Jaeger. The echo state approach to analysing and training recurrent neural networks-with an erratum note. *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report*, 148(34): 13, 2001.

Herbert Jaeger and Harald Haas. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, 304 (5667):78–80, 2004.

Herbert Jaeger, Mantas Lukoševičius, Dan Popovici, and Udo Siewert. Optimization and applications of echo state networks with leaky-integrator neurons. *Neural networks*, 20 (3):335–352, 2007.

John Bagterp Jørgensen, Jakob Kjøbsted Huusom, and Grkan Sin. Control of blood glucose for people with type 1 diabetes: an in vivo study. In *Proceedings of the 17th Nordic Process Control Workshop*, pages 133–140. Technical University of Denmark (DTU), 2012. ISBN 978-87-643-0946-1.

Harald Kirchsteiger, John Bagterp Jørgensen, Eric Renard, and Luigi Del Re. *Prediction methods for blood glucose concentration : design, use and evaluation*. Springer, 2016.

Mantas Lukoševičius. A practical guide to applying echo state networks. In *Neural networks: tricks of the trade*, pages 659–686. Springer, 2012.

Mantas Lukoševičius and Herbert Jaeger. Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3 (3):127–149, 2009.

A Millea. Explorations in echo state networks (master's thesis). *Department of Artificial Intelligence, University of Groningen, Netherlands*, 2014.

Carmen Pérez-Gandía, A Facchinetti, G Sparacino, C Cobelli, EJ Gómez, M Rigla, Alberto de Leiva, and ME Hernando. Artificial neural network algorithm for online glucose prediction from continuous glucose monitoring. *Diabetes technology & therapeutics*, 12(1):81–88, 2010.

Giovanni Sparacino, Francesca Zanderigo, Stefano Corazza, Alberto Maran, Andrea Facchinetti, and Claudio Cobelli. Glucose concentration can be predicted ahead in time from continuous glucose monitoring sensor time-series. *IEEE Transactions on biomedical engineering*, 54(5):931–937, 2007.

Cristina Tarin, Edgar Teufel, Jesús Picó, Jorge Bondia, and H-J Pfleiderer. Comprehensive pharmacokinetic model of insulin glargine and other insulin formulations. *IEEE Transactions on Biomedical Engineering*, 52(12):1994–2005, 2005.

Leonard J Tashman. Out-of-sample tests of forecasting accuracy: an analysis and review. *International journal of forecasting*, 16(4):437–450, 2000.