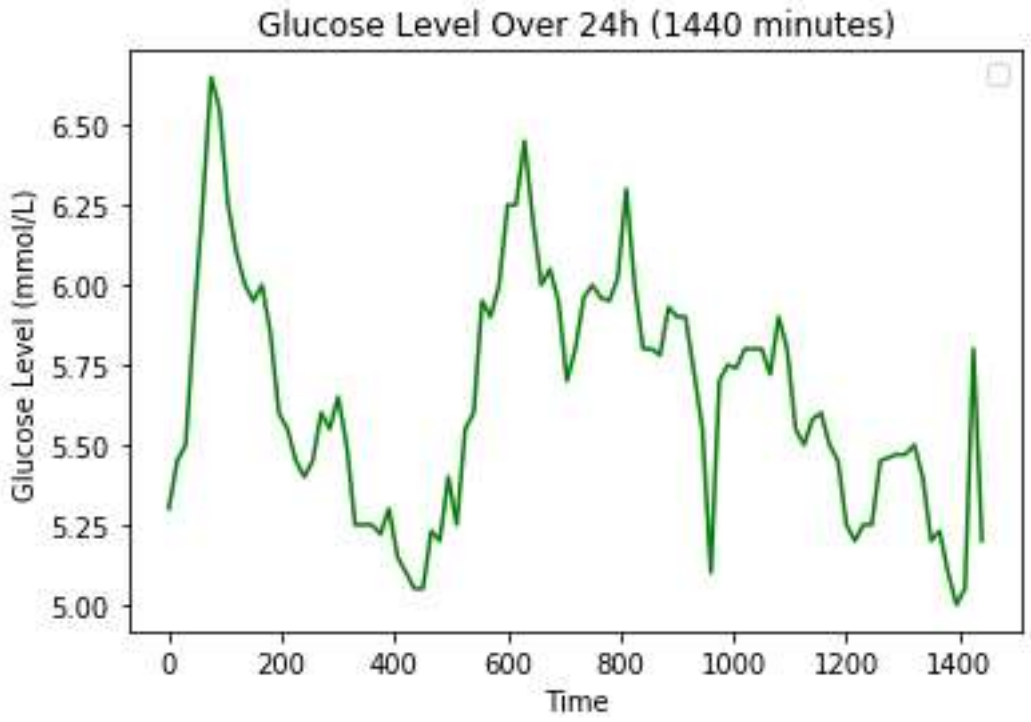
PREDICTION ENGINE

- Introduction
- Goals and requirements
  - Long term prediction
  - Short term predictions
  - Computation complexity
  - Real-time processing

PREDICTION ENGINE

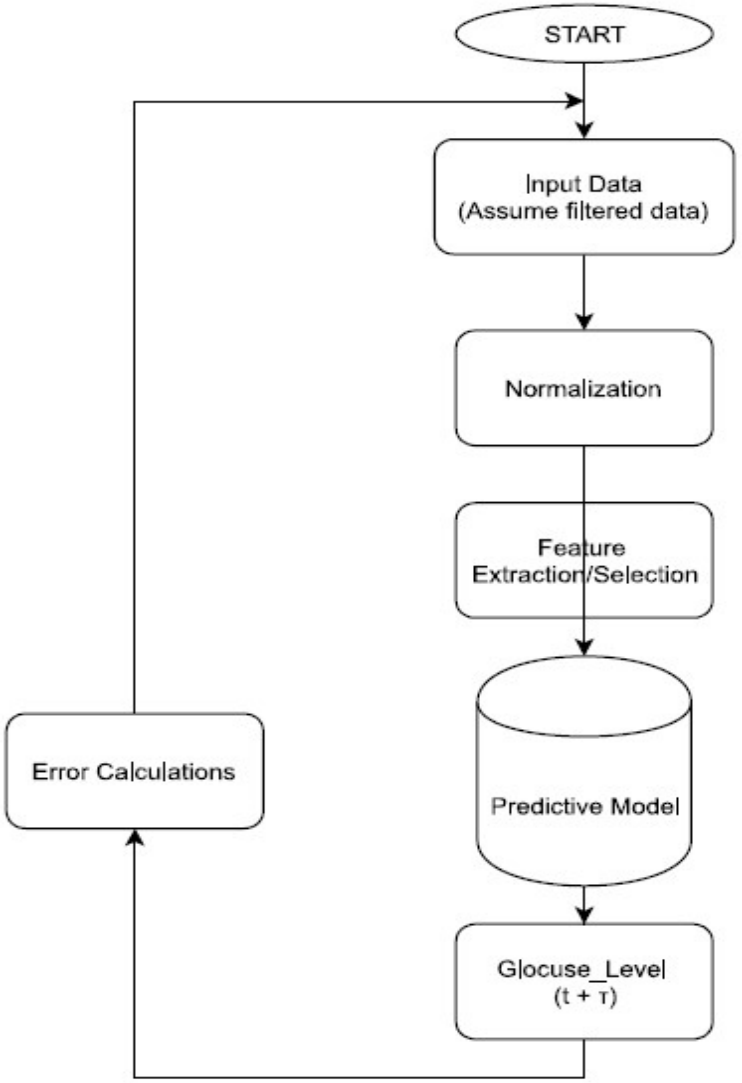
- Introduction
- Goals and requirements
  - Long/short term prediction
  - Computation complexity
  - Real-time processing
- data status



PREDICTION ENGINE

Short Term Prediction

- Prediction framework architecture
  - Filtered data
  - Normalization
  - ~~Feature selection~~
  - Predictive model
  - Output
  - Error calculations



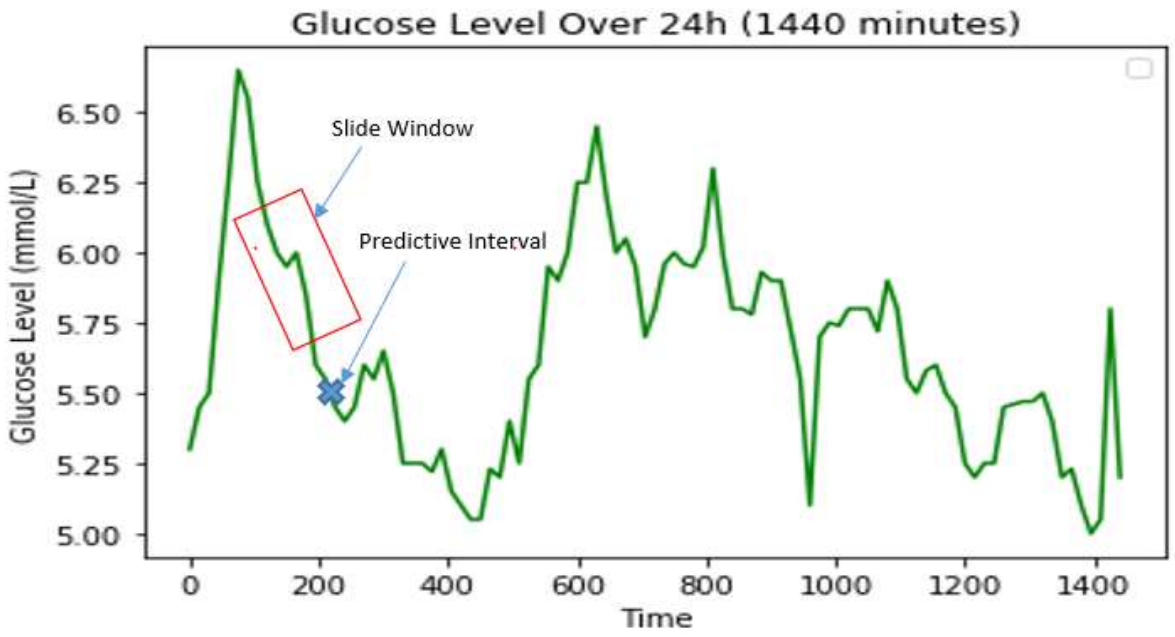
PREDICTION ENGINE

Prototype Model

- Simple linear regression

$$GL(t) = \beta t + \alpha$$

  - Slide window
  - Window size
  - Predictive interval



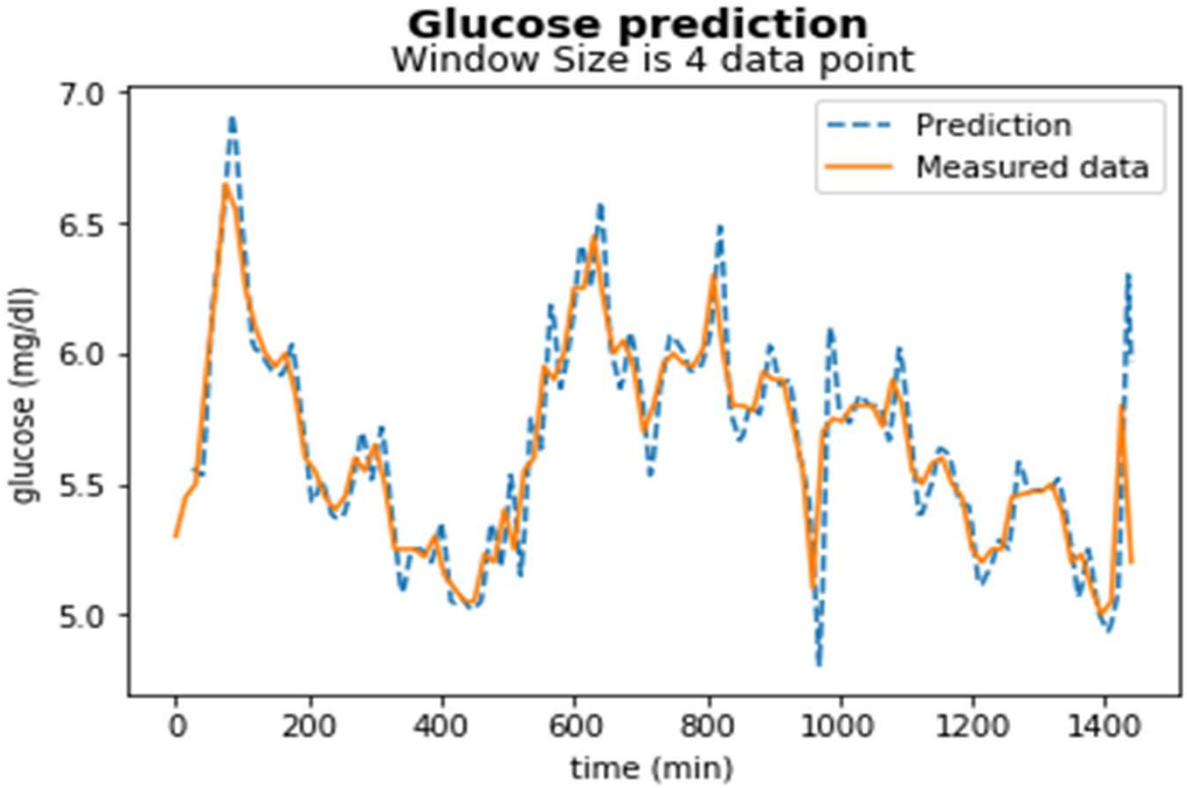
PREDICTION ENGINE

Prototype Model

- Simple linear regression

$$GL(t) = \beta t + \alpha$$

  - Slide window
  - Window size
  - Predictive interval



PREDICTION ENGINE

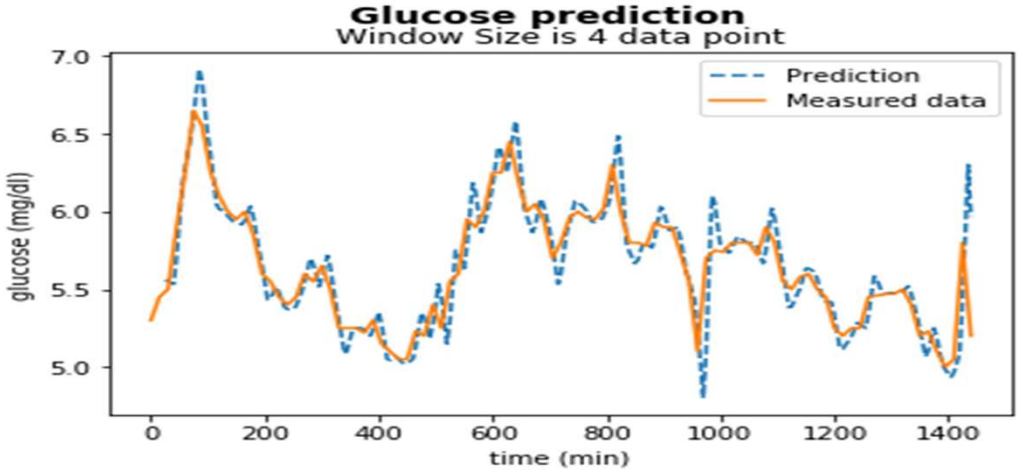
Prototype Model

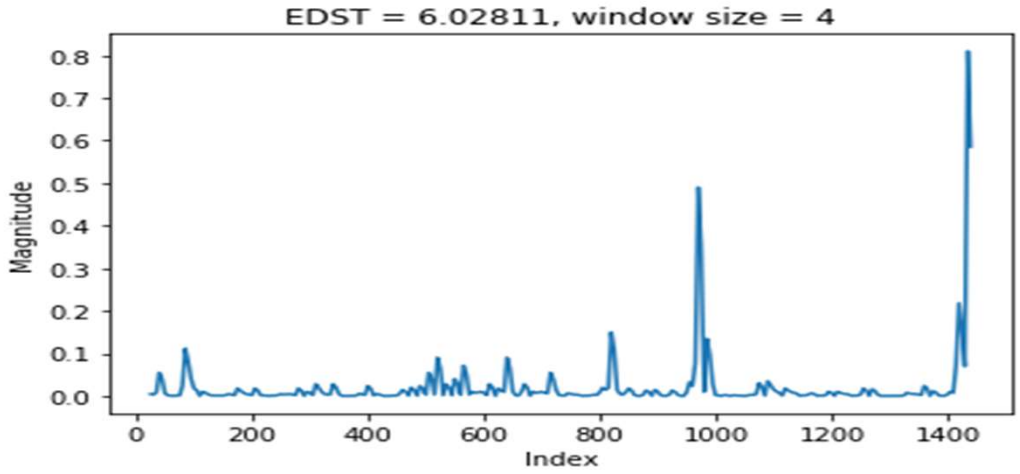
- Simple linear regression

$$GL(t) = \beta t + \alpha$$

  - Slide window
  - Window size
  - Predictive interval

- how good is it?
  - Traditional accuracy
    - Accuracy = (1-mse) = 98.9%
  - Error measurement mechanisms
    - Error difference square (EDS)



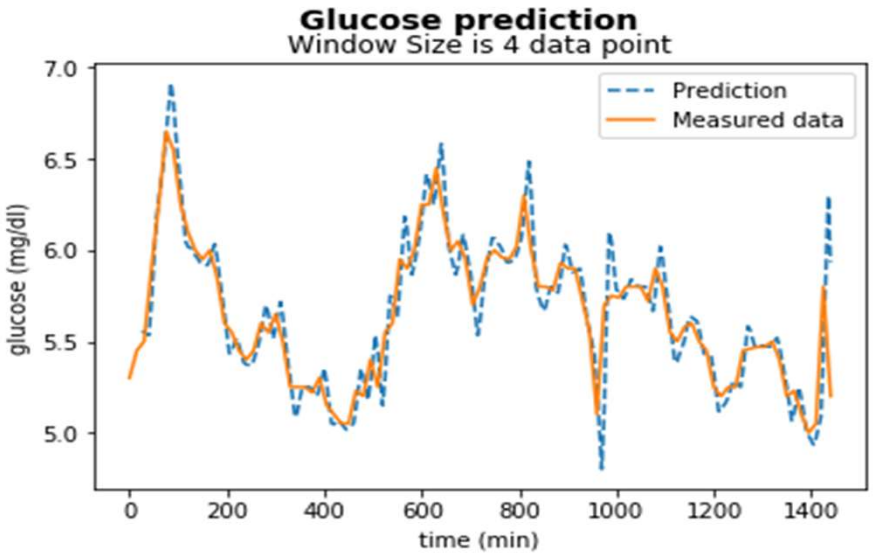


PREDICTION ENGINE

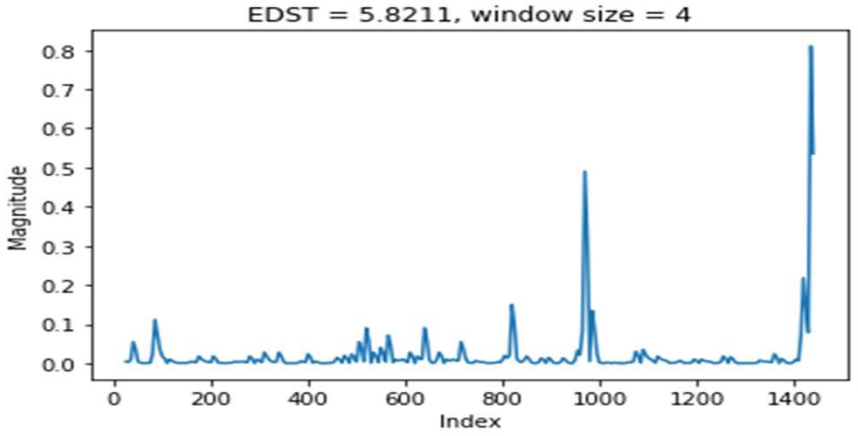
Improve The Model

- Weighted linear regression

$$\sum_{n=0}^{N} w_n \, GL(t) \; = \sum_{n=0}^{N} w_n \beta t + \; \alpha$$



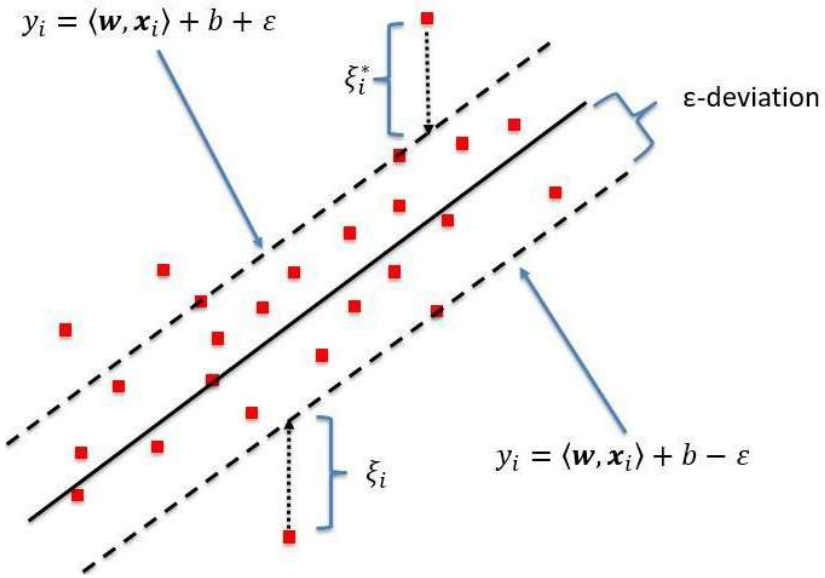
- Compare to prototype
  - Window size
  - Predictive interval
  - Eds



PREDICTION ENGINE

Support Vector Regression Model

- Linear mode



PREDICTION ENGINE

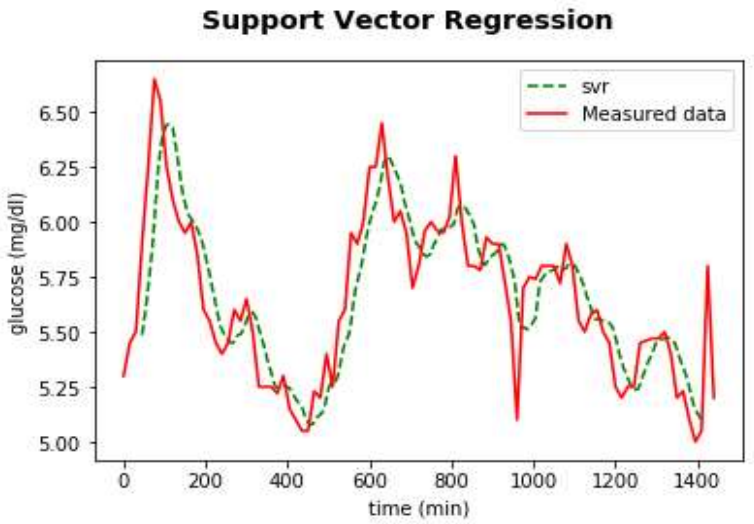
Support Vector Regression Model

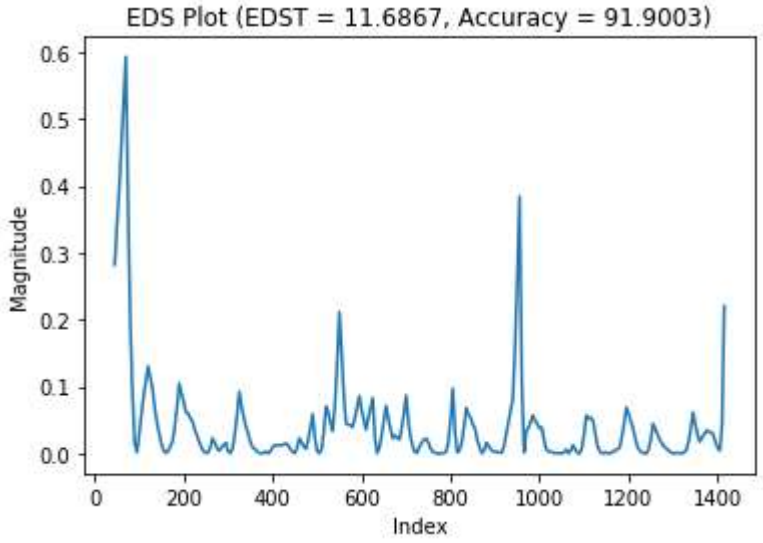
- Linear mode
- Glucose level is not linear
  - Kernel mode

PREDICTION ENGINE

Support Vector Regression Model

- Linear mode
- Glucose level is not linear
  - Kernel mode
  - Increase prediction parameters
    - Window size: 45 minutes
    - Predictive interval: 25 minutes

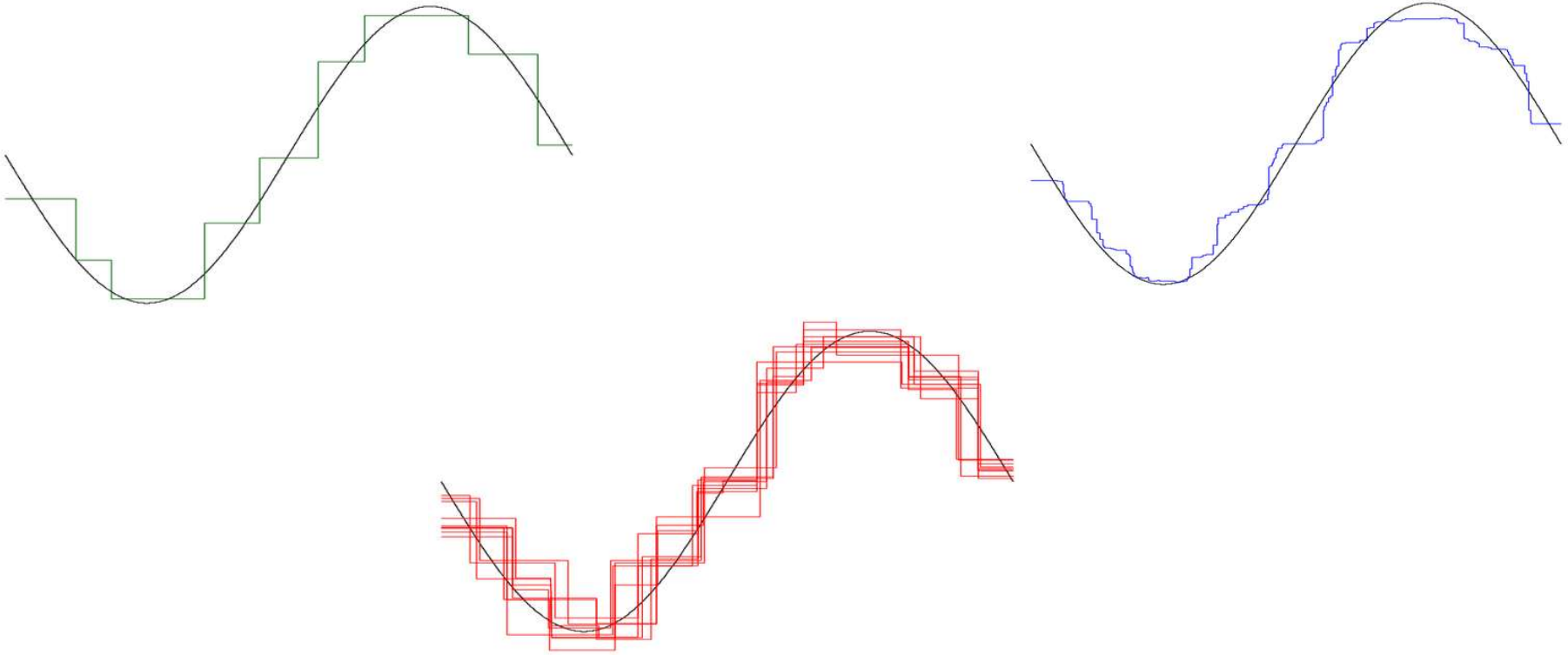




PREDICTION ENGINE

Random Forest Regression Model

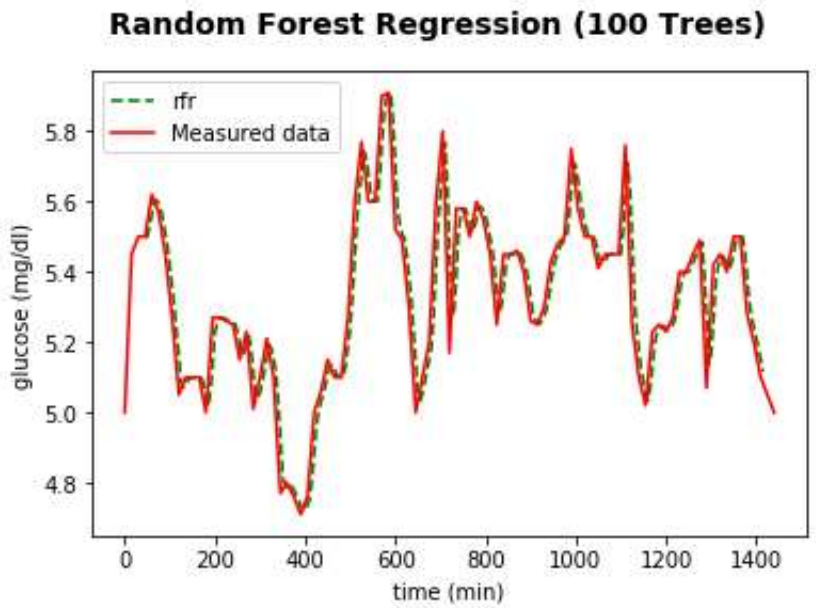
- Multiple of decision trees



PREDICTION ENGINE

Random Forest Regression Model

- Multiple of decision trees
- Window size: 45 minutes
- Predictive interval: 25 minutes
- Too good!
  - Overfitting problem



PREDICTION ENGINE

Compare Models

Window Size Prediction Interval	LR		LR with weights		SVR		RFR	
	EDST	Accuracy (1 - MSE)	EDST	Accuracy (1 - MSE)	EDST	Accuracy (1 - MSE)	EDST	Accuracy (1 - MSE)
ws = 30 minutes P_Int = 20 minutes	17.8525	90.0337	16.9632	90.3019	7.2228	95.3206		
ws = 30 minutes P_Int = 25 minutes	23.3633	87.0602	22.3495	87.3359	7.2128	94.7058		
ws = 30 minutes P_Int = 30 minutes	29.8309	83.8444	28.6904	84.1476	7.1999	94.1226		
ws = 45 minutes P_Int = 20 minutes	15.8998	91.1945	14.4586	91.5922	8.4303	94.9556	3.9872 (100 trees) 4.1051 (20 Trees)	95.9996 (100 trees) 95.9746 (20 Trees)
ws = 45 minutes P_Int = 25 minutes	19.6457	88.8634	18.0539	89.2984	8.4104	94.3795	3.9853 (100 trees) 4.1030 (20 Trees) 4.5620 (5Trees)	95.3117 (100 trees) 95.2915(20 Trees) 95.1504(5Trees)
ws = 45 minutes P_Int = 30 minutes	23.9532	86.3146	22.2187	86.8110	8.3807	93.8433	3.9829 (100 trees) 4.10041(20 Trees)	94.6983 (100 trees) 94.6799(20 Trees)

PREDICTION ENGINE

Conclusion

- Overall look start-to-now
- Time-series analysis

PREDICTION ENGINE

Conclusion

- Overall look start-to-now
- Time-series analysis

$$GL(t) = \beta GL(t)_{i-1} + \alpha$$

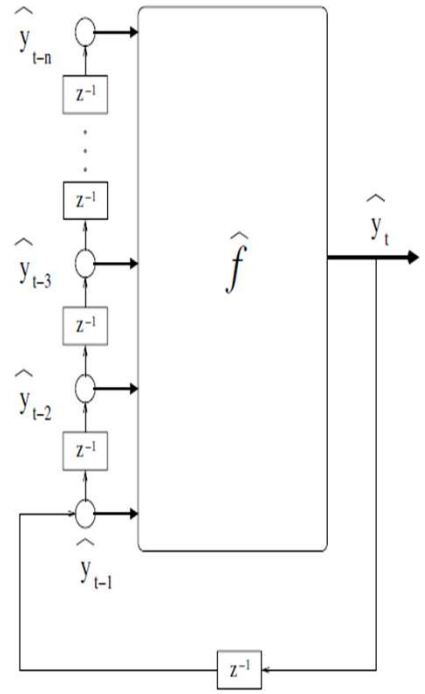


Fig. 3. Iterated prediction. The approximator $\hat{f}$ returns the prediction of the value of the time series at time $t+1$ by iterating the predictions obtained in the previous steps (the rectangular box containing z^{-1} represents a unit delay operator, i.e., $\hat{y}^{t-1} = z^{-1}\hat{y}^t$).

PREDICTION ENGINE

Conclusion

- Overall look start-to-now
- Time-series analysis

$$GL(t) = \beta GL(t)_{i-1} + \alpha$$

y(t-7)	y(t-6)	y(t-5)	y(t-4)	y(t-3)	y(t-2)	y(t-1)	y(t)
5.3	5.35	5.4	5.45	5.4667	5.4833	5.5	5.6333
5.35	5.4	5.45	5.4667	5.4833	5.5	5.6333	5.7667
5.4	5.45	5.4667	5.4833	5.5	5.6333	5.7667	5.9
5.45	5.4667	5.4833	5.5	5.6333	5.7667	5.9	6.0167
5.4667	5.4833	5.5	5.6333	5.7667	5.9	6.0167	6.1333
5.4833	5.5	5.6333	5.7667	5.9	6.0167	6.1333	6.25
5.5	5.6333	5.7667	5.9	6.0167	6.1333	6.25	6.3833
5.6333	5.7667	5.9	6.0167	6.1333	6.25	6.3833	6.5167
5.7667	5.9	6.0167	6.1333	6.25	6.3833	6.5167	6.65
5.9	6.0167	6.1333	6.25	6.3833	6.5167	6.65	6.6167

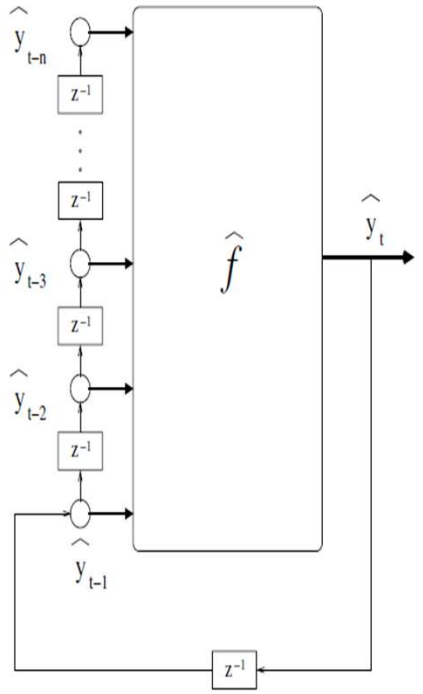


Fig. 3. Iterated prediction. The approximator $\hat{f}$ returns the prediction of the value of the time series at time $t+1$ by iterating the predictions obtained in the previous steps (the rectangular box containing z^{-1} represents a unit delay operator, i.e., $\hat{y}^{t-1} = z^{-1}\hat{y}^t$).

PREDICTION ENGINE

Conclusion

Short term predictions

- Overall look start-to-now
- Time-series analysis

To do in short term predictions

- Time-series models
  - Autoregressive model
- Expert system – model aggregation

Long term predictions

- Recurrent neural network
- Long short term memory
- Feed forward neural network

REFERENCES

For fourth last slide
Bontempi G, Ben Taieb S, Le Borgne Y. ”**Machine Learning Strategies for Time Series Forecasting**.”

SVR slide
http://qetartu.blogspot.com/2016/12/applied-econometrics-statistical.html

Random Forest Regression slide
Kleynhans T, Montanaro M, Gerace A, Kanan C. “Predicting Top-of-Atmosphere Thermal Radiance Using MERRA-2 Atmospheric Data with Deep Learning”, November 2017