



# Interpolation und Extrapolation von korrelierten physiologischen Zeitreihen

Bachelorarbeit  
Klara-Maria Hörschläger  
Betreuer: Dr. Florian Reiterer

# Wiederholung Thema

- Daten-Lücken in Zeitreihen mehrerer simultan gemessener Größen rekonstruieren
- Anwendung auf den MIMIC-III Datensatz
- Vergleich Interpolation, Imputation und M-RNN

# Größenordnungen MIMIC-III

- 23160 Patienten in Metavision
  - 2992 verschiedene Aufzeichnungsgrößen (Notizen, Medikamente, Diagnosen, Messungen, ... )
- 
- Nur Aufzeichnungsgrößen die Zahlenwerte sind brauchbar (Einheit mmHg, °F, ...)
  - 30 ‚Routine Vital Signs‘ - 17718 Patienten

# Struktur Datensatz

	Stream1:	$x_{t-1}^1$ ,	$x_t^1$ ,	$x_{t+1}^1$
	Stream2:	$x_{t-1}^2$ ,	$x_t^2$ ,	$x_{t+1}^2$
X	Stream3:	$x_{t-1}^3$ ,	$x_t^3$ ,	$x_{t+1}^3$
	Stream4:	$x_{t-1}^4$ ,	$x_t^4$ ,	$x_{t+1}^4$
	Stream5:	$x_{t-1}^5$ ,	$x_t^5$ ,	$x_{t+1}^5$
		t-1	t	t+1

- mittels Matlab stündliche Timebuckets erstellt, um simultan gemessene Größen zuordnen zu können
- aufgrund Übersichtlichkeit und Handling für weitere Anwendung beschränkt auf 3 Größen (Blutdruck systolisch und diastolisch, Herzfrequenz) und 25 Patienten

# Interpolation

$$D = \{\{x_1, y_1\}, \{x_2, y_2\}, \dots, \{x_n, y_n\}\} \quad \{x_i, y_i^{int}\}$$

$$\{\{x_m, y_m\}, \{x_{m+1}, y_{m+1}\}, \dots, \{x_{m+d}, y_{m+d}\}\} \subseteq D:$$

$$x_m < x_i < x_{m+d}$$

Parameter P der Funktion f(x) werden über die Stützstellen und Zusatzbedingungen berechnet

$$\{x_i, f(x_i)\}$$

### Lineare Interpolation

Stützstellen	$\{x_1, x_2\}$
Funktion	$f(x) = a \cdot x + b$
Parameter	$\{a, b\}$
Bedingungen	$f(x_1) = y_1, f(x_2) = y_2$

### Polynomiale Interpolation

Stützstellen	$n, \{x_1, \dots, x_n\}$
Funktion	$f(x) = a_{n-1} \cdot x^{n-1} + a_{n-2} \cdot x^{n-2} + \dots + a_0 \cdot x^0$
Parameter	$n, \{a_{n-1}, \dots, a_0\}$
Bedingungen	$f(x_i) = y_i \forall i$

### Cubische Spline Interpolation

Stützstellen	$n, \{x_1, \dots, x_n\}$
Funktion	$f_i(x) = a_i \cdot x^3 + b_i \cdot x^2 + c_i \cdot x + d_i, x_i \leq x \leq x_{i+1}, 1 \leq i \leq n-1$
Parameter	$4 \cdot n, \{a_i, b_i, c_i, d_i\}$
Bedingungen	$f_i(x_i) = y_i, f_i(x_{i+1}) = y_{i+1}, f_i'(x_{i+1}) = f_{i+1}'(x_{i+1}), f_i''(x_{i+1}) = f_{i+1}''(x_{i+1})$

# Imputation

$$\begin{aligned}
 X &= \{ \{x_{11}, x_{22}, \dots, x_{1n}\}, \{x_{21}, x_{22}, \dots, x_{2n}\}, \{x_{m1}, x_{m2}, \dots, x_{mn}\} \} \\
 &= \{x_{1.}, x_{2.}, \dots, x_{m.}\} \\
 &= \{x_{.1}, x_{.2}, \dots, x_{.n}\}
 \end{aligned}$$

Schätzwert für  $x_{tk}$  auf Basis der zeitgleichen Messwerte  $x_{t.}$  generieren

„nearest neighbour“ – Spalte:

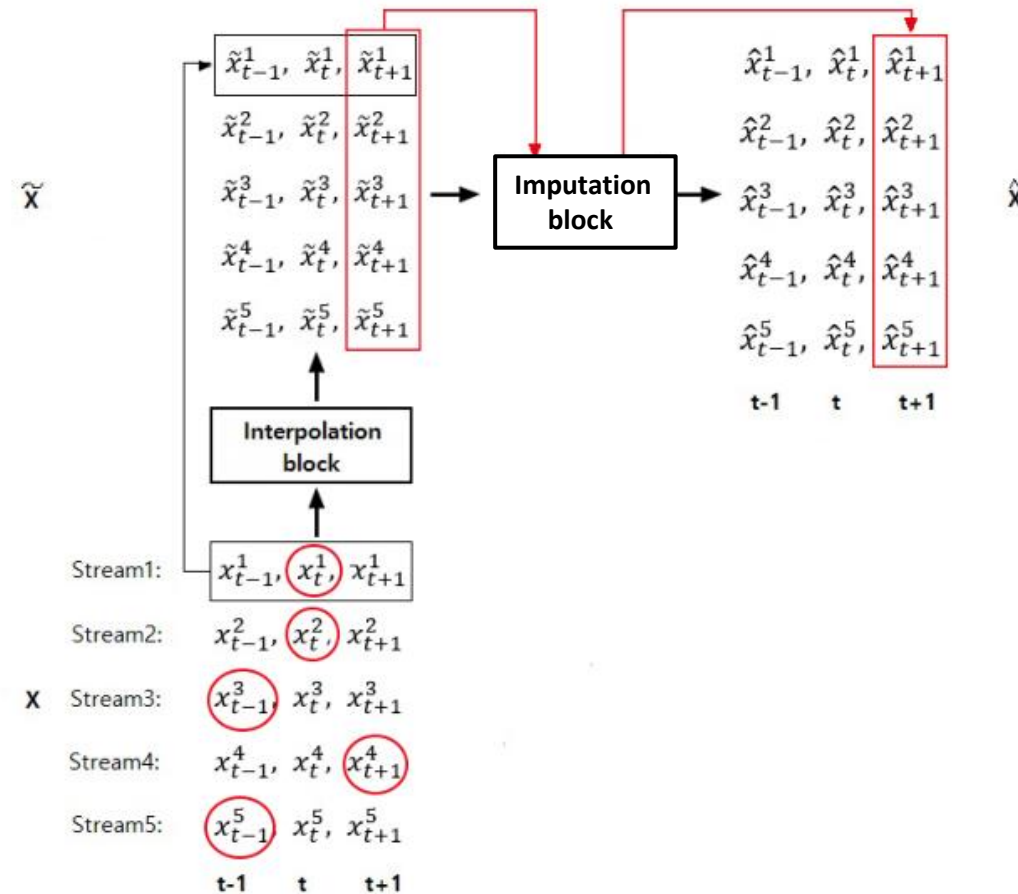
$$d_{ij}(x_{.k}, x_{.p}) < d_{ij}(x_{.i}, x_{.k}) \quad \forall i \notin \{p, k\}$$

Cityblock Metrik:

$$d_{s,t} = \sum_{j=1}^T |x_{s,j} - x_{t,j}|$$

# Multidimensional Recurrent Neural Network

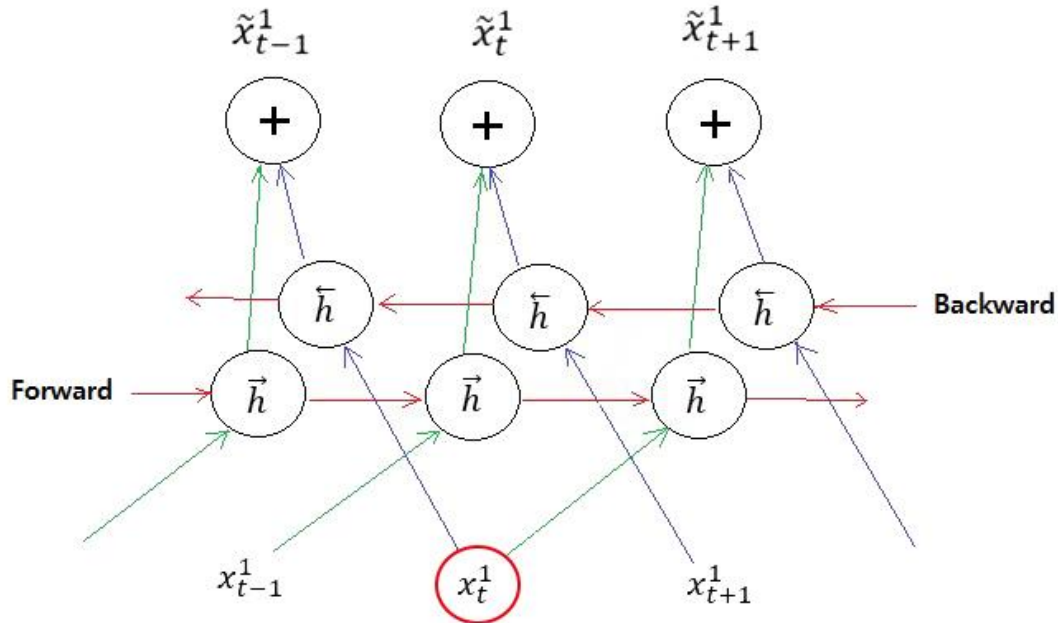
## Struktur des M-RNN





# Interpolationsblock des M-RNN

- Bi – RNN
- Single hidden layer



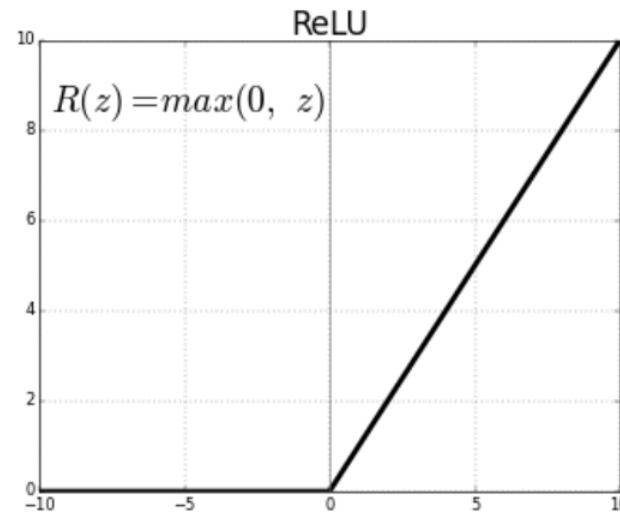
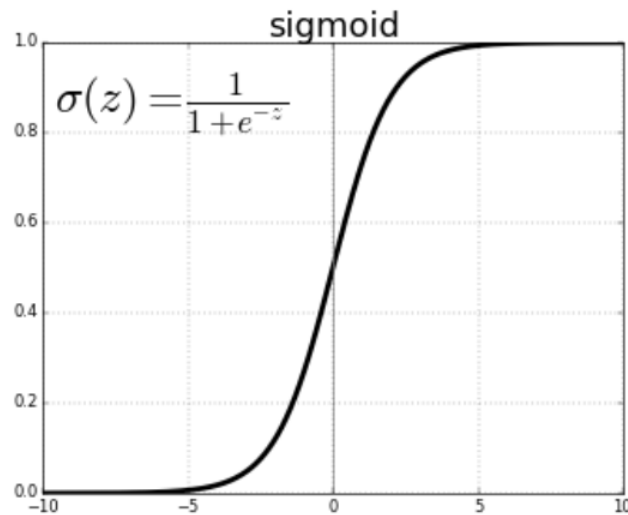
$$\vec{h}_t^d = f(\vec{W}^d \vec{h}_{t-1}^d + \vec{V}^d \mathbf{x}_{t-1}^d + \vec{c}^d)$$

$$\overleftarrow{h}_t^d = f(\overleftarrow{W}^d \overleftarrow{h}_{t+1}^d + \overleftarrow{V}^d \mathbf{x}_{t+1}^d + \overleftarrow{c}^d)$$

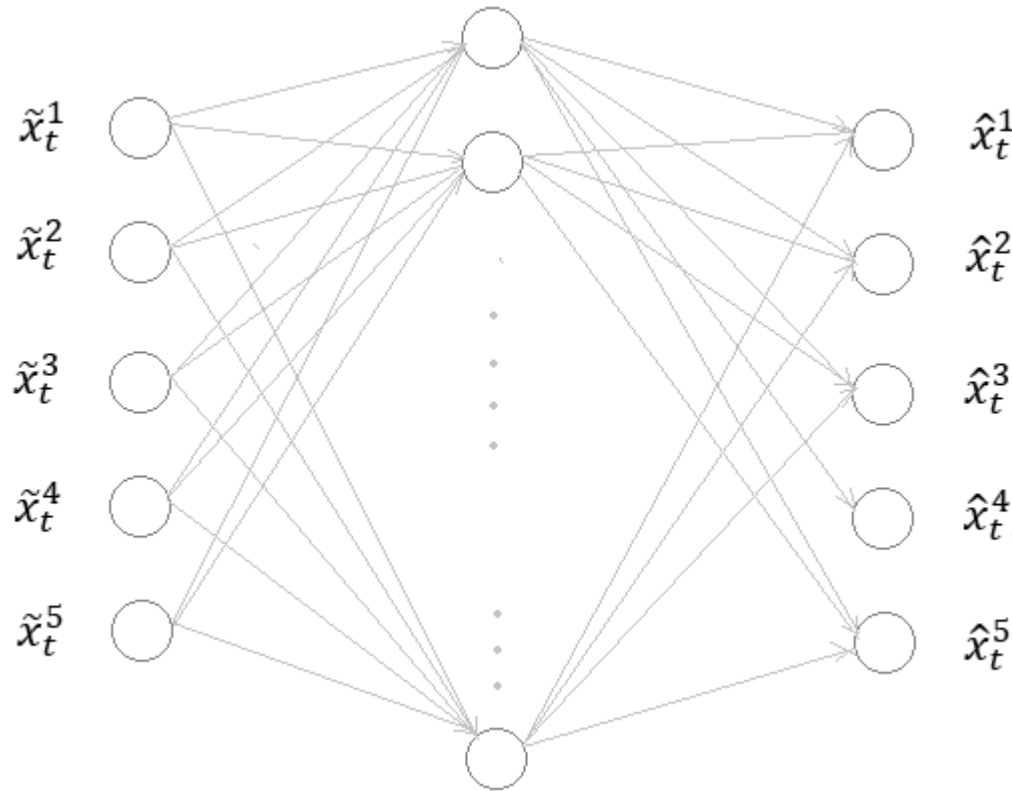
$$\tilde{x}_t^d = g(U^d[\vec{h}_t^d, \overleftarrow{h}_t^d] + \mathbf{c}_o^d) = g(\vec{U}^d \vec{h}_t^d + \overleftarrow{U}^d \overleftarrow{h}_t^d + \mathbf{c}_o^d)$$

# Aktivierungsfunktion

$$net_j = \sum_{i=1}^n (x_i w_{ij}) + \alpha_j$$



# Imputation



- Feed forward NN
- Fully connected Layer
- Backpropagation
- Sigmoid Activationfunktion

$$\hat{x}_t = \sigma(W \mathbf{h}_t + \alpha)$$

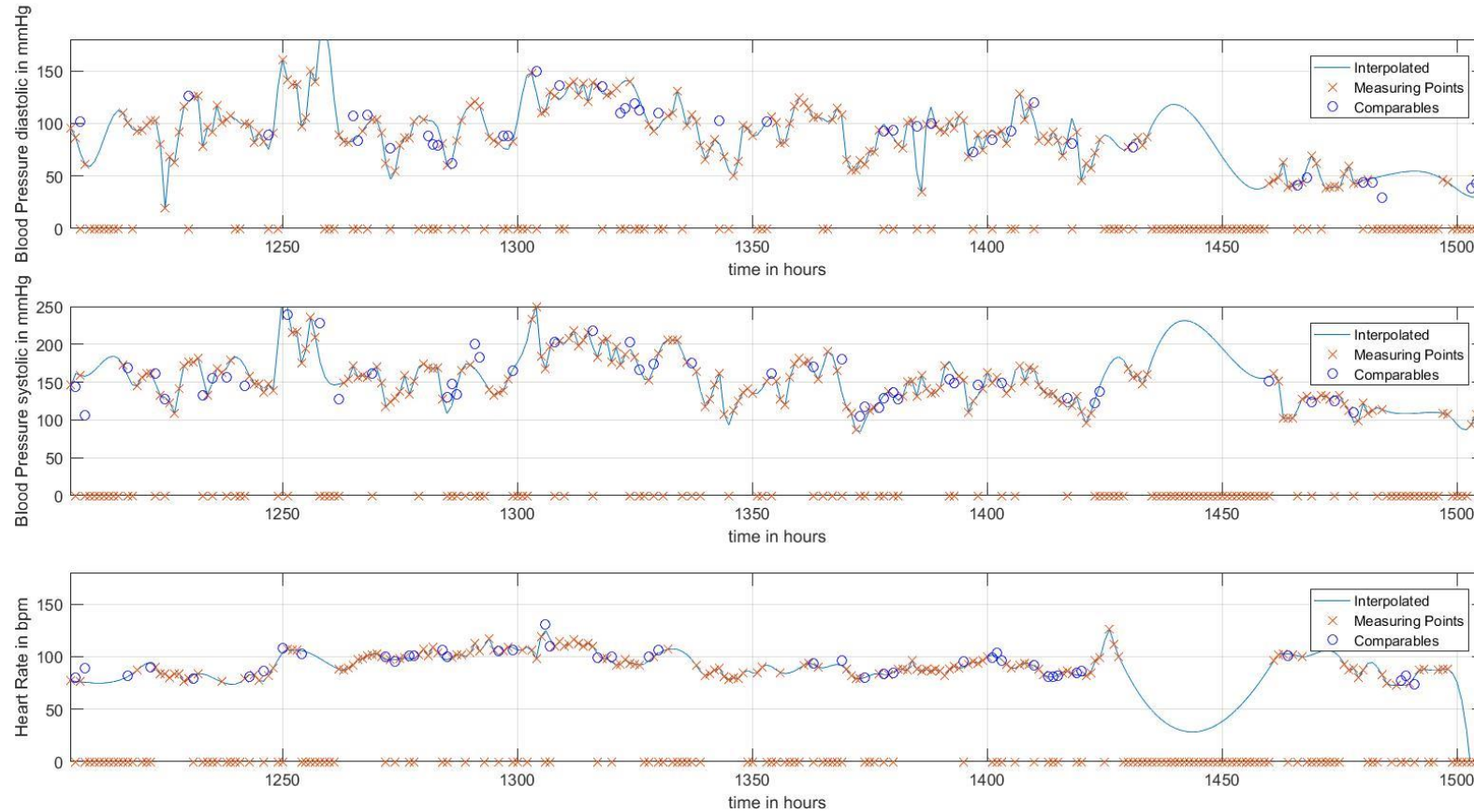
$$\mathbf{h}_t = \phi(U \mathbf{x}_t + \beta)$$

# Root Mean Square Error

$$E = \sqrt{\sum_{t=1}^T \sum_{d=1}^D (1 - m_t^d) (x_t^d - \hat{x}_t^d)^2}$$

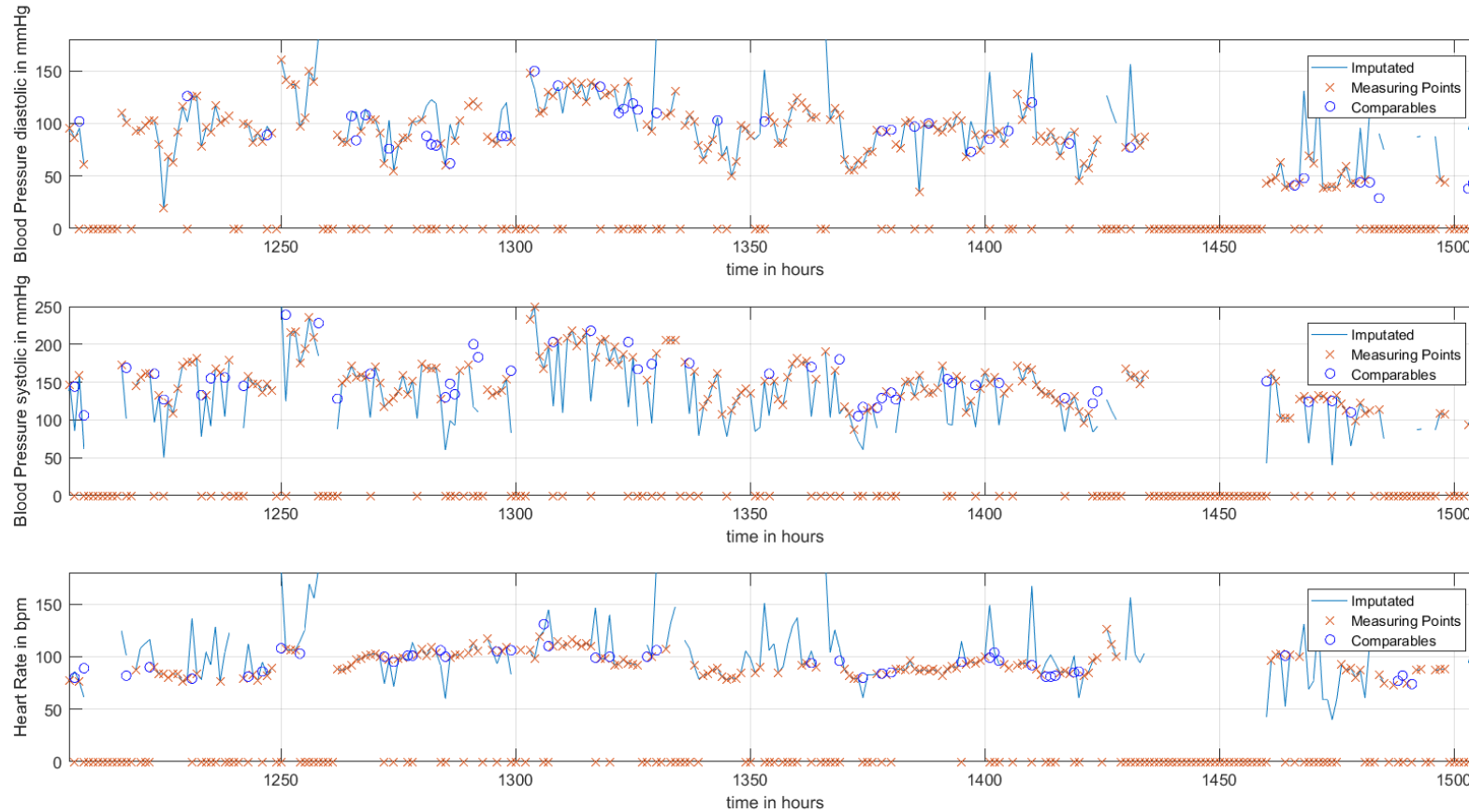
Missing Rate	RMSE linear	RMSE cubic	RMSE spline	RMSE cityblock	RMSE M-RNN
0.7162	0.0508	0.0468	0.0601	0.3246	0.0163

# Physiologische Validität



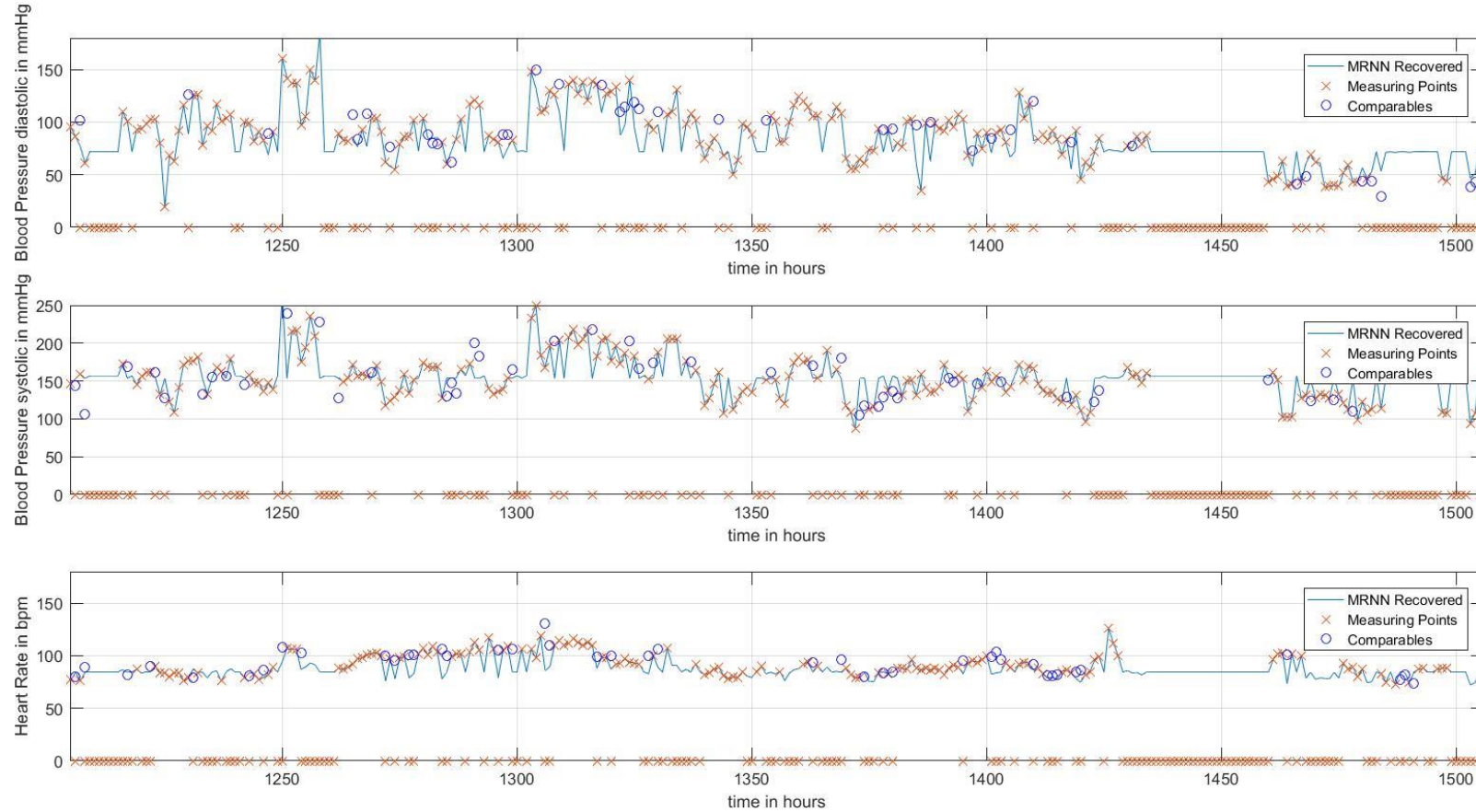
Spline Interpolation, Patient #24

# Physiologische Validität



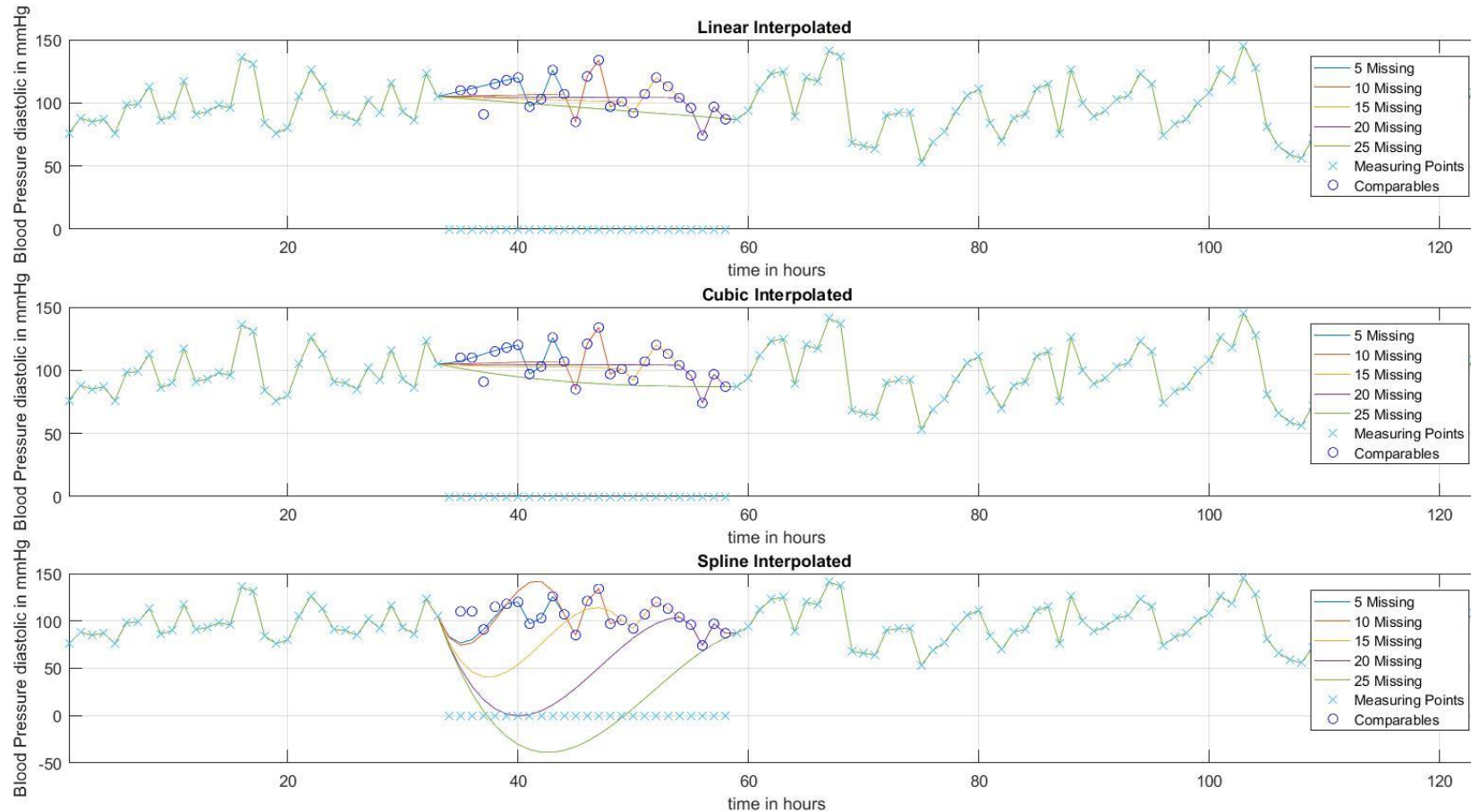
Imputation, Patient #1

# Physiologische Validität



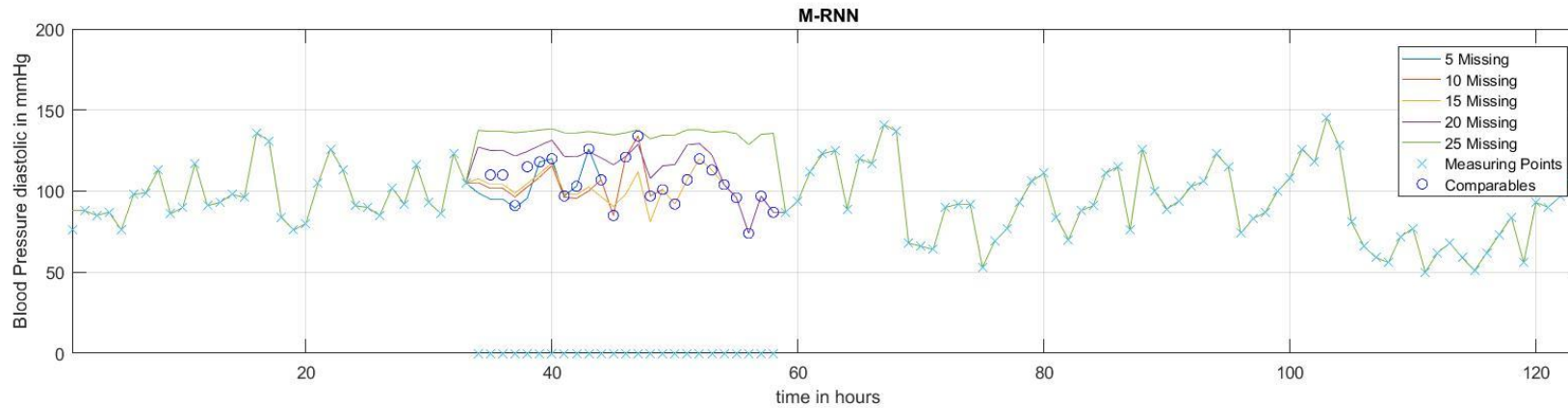
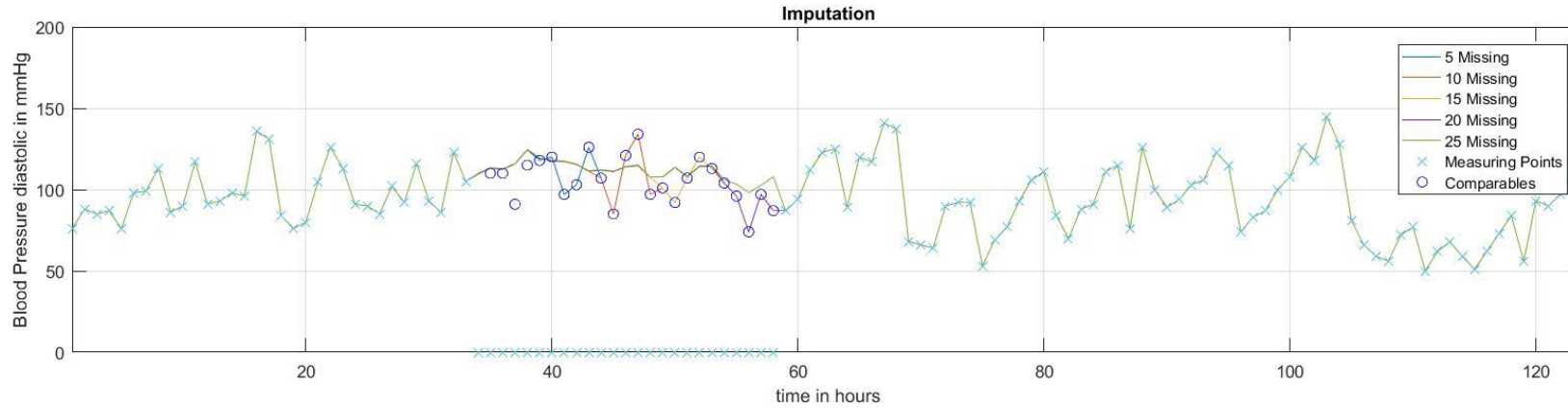
M-RNN, Patient #1

# Verhalten für wachsende Lückenlängen





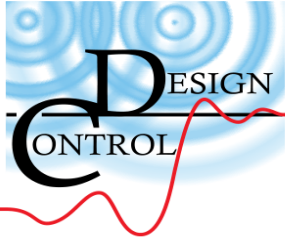
# Verhalten für wachsende Lückenlängen



# Resümee

Missing Rate	RMSE linear	RMSE cubic	RMSE spline	RMSE cityblock	RMSE M-RNN
0.7162	0.0508	0.0468	0.0601	0.3246	0.0163

- Verwendete Imputationsmethode sehr reduziert in Ausführung – Fehler könnte geringer Werden durch Erweiterung
- RMSE des M-RNN eindeutig am besten
- RMSE der Interpolationsmethoden akzeptabel
  
- Imputation nur anwendbar, wenn simultan gemessene Werte vorhanden
- Physiologische Validität linearer und kubischer Interpolation gegeben, spline Interpolation nicht
- Physiologische Validität der M-RNN Signale gegeben



Institut für Design und Regelung mechatronischer Systeme

Vielen Dank für die Betreuung der Arbeit  
und Ihre Aufmerksamkeit.