Modellierung von Abhängigkeitsstrukturen in dynamischen Systemen

Ursula Gather
Fachbereich Statistik
Universität Dortmund

Übersicht

Motivation

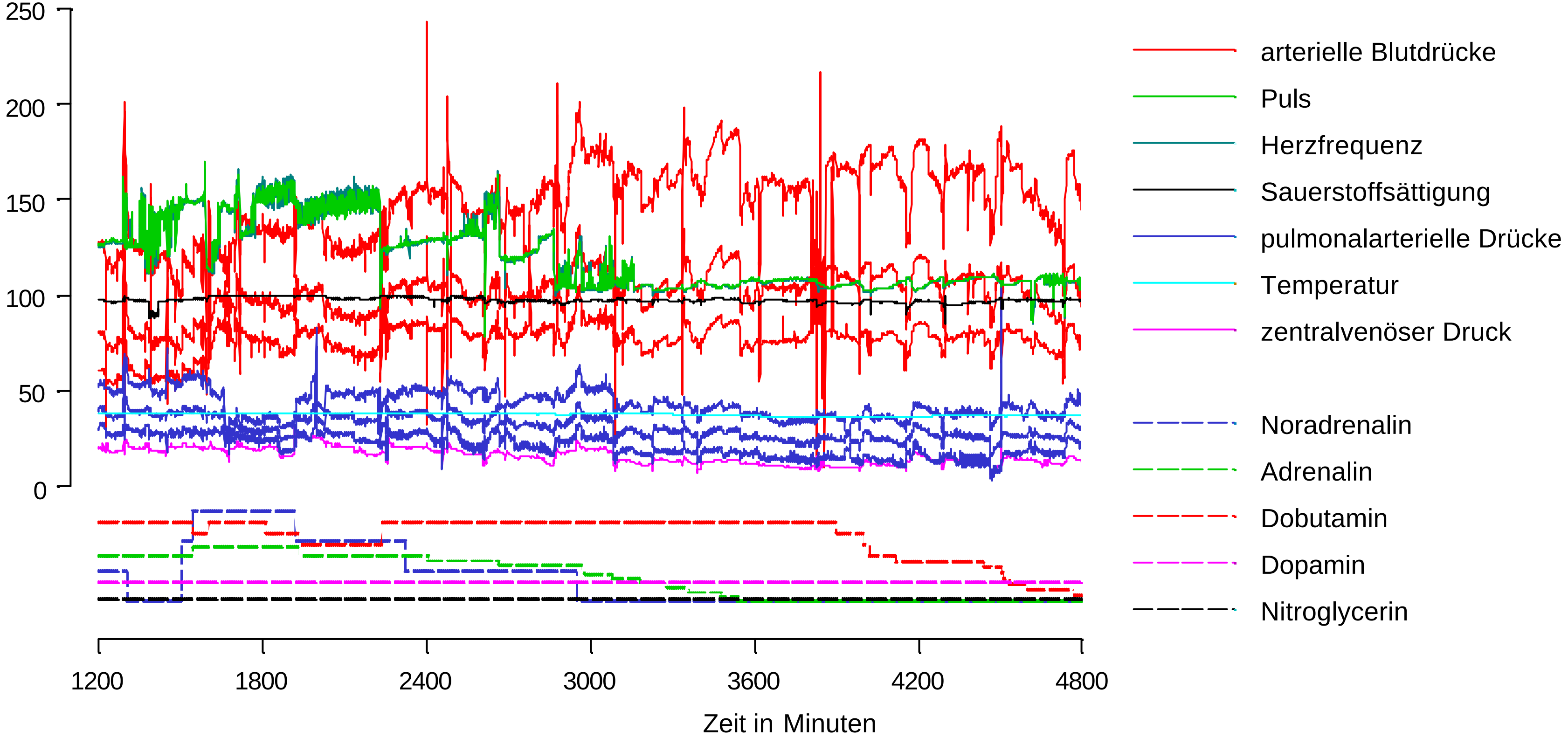
1. Online Monitoring Daten aus der Intensivmedizin

Methoden zur Analyse Dynamischer Strukturen

2. Graphische Modelle
3. Sliced Inverse Regression
4. Dynamische Faktorenanalyse

1) Beispiel: Intensivmedizin

11 Variablen des hämodynamischen Systems
und Medikamentendosen eines Patienten über 60 Stunden



Analyse von intensivmedizinischen Online Monitoring Daten

- hochdimensionale Daten (>> 200 Variablen)
- komplexe Datenstrukturen (Abhängigkeiten zwischen verschiedenen Variablen und Zeitabhängigkeiten)

Statistik

medizinisches Expertenwissen

Statistische Online-Methoden zur Mustererkennung und Entscheidungsfindung

- individuell
- mit schnellen Algorithmen
- einfache Interpretierbarkeit

Ziele

- Dimensionsreduktion
  Verständnis der dynamischen Abhängigkeitsstruktur
  (mit medizinischen Experten)

- Erkennung von Mustern wie
  - Ausreißern
  - Level Shifts
  - Trends

Dimensionsreduktion und Abhängigkeitsstruktur

Sei $\mathbf{Y}(t) \in \mathrm{IR}^k$, $t \in \boldsymbol{Z}$, (stationärer) stochastischer Prozeß

Ziel: extrahiere wenige gemeinsame „Faktoren“,
die die Zeitreihe „antreiben”

Statische Methoden zur Dimensionsreduktion:

- Graphische Modelle
- Sliced Inverse Regression
- Faktorenanalyse
- ...

Zeitreihenkontext $\Longrightarrow$ angemessene Adaptation dieser Verfahren erforderlich

2) Graphische Modelle für multivariate Zeitreihen

(Brillinger 1996, Dahlhaus 2000)

Beobachtet: k-dim. Zeitreihe

$$y(t) = \left(y^1(t), \ldots, y^i(t), \ldots, y^j(t), \ldots, y^k(t) \right), \; t = 1, \ldots, N$$

Ecken: $V = \{1, \ldots, k\}$

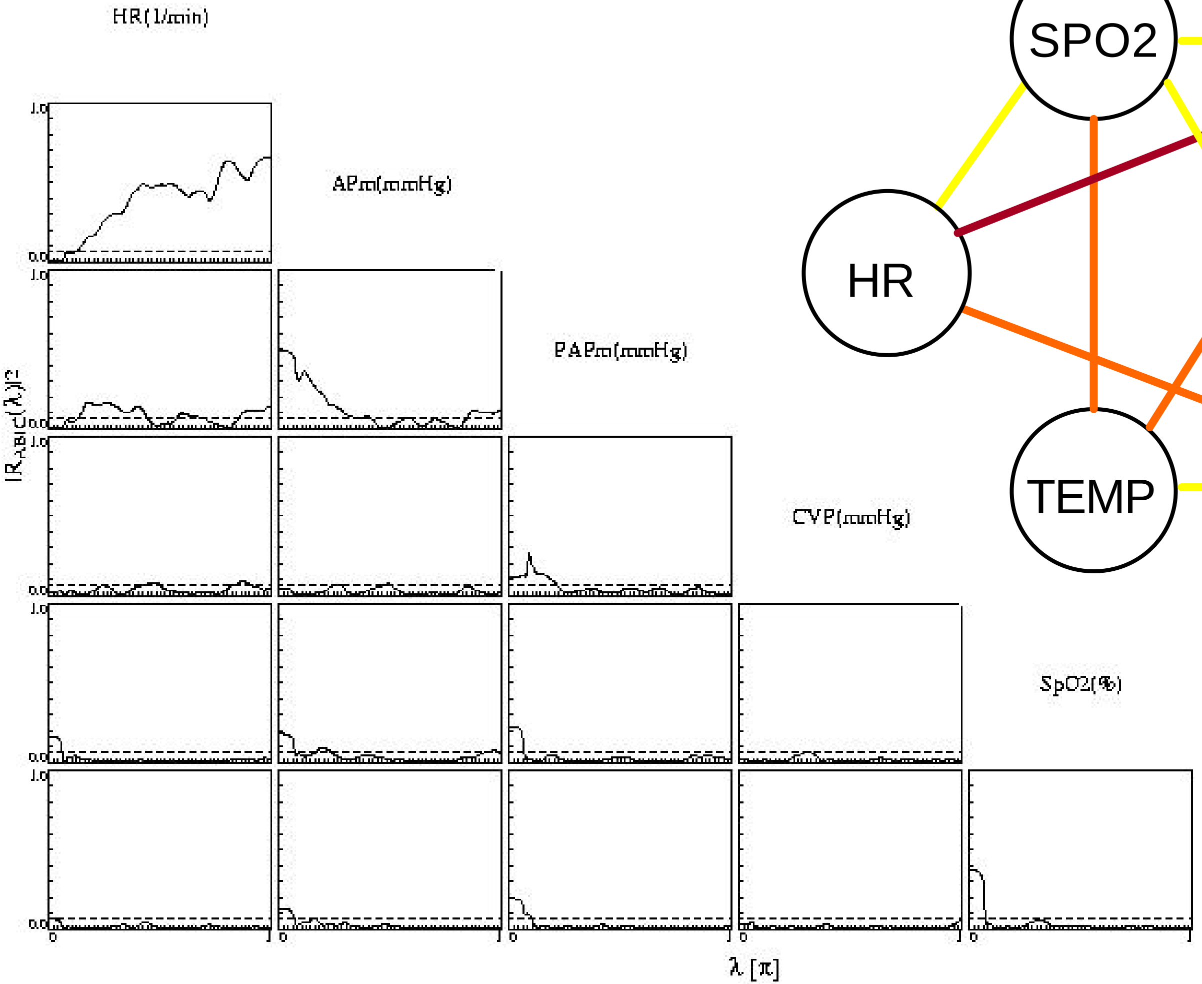
Kanten: $(i,j) \notin E \subset V \times V \quad \Leftrightarrow \quad y^i(t)$ und $y^j(t+h)$ sind unkorreliert
für alle Zeitlags h gegeben
die anderen Variablen

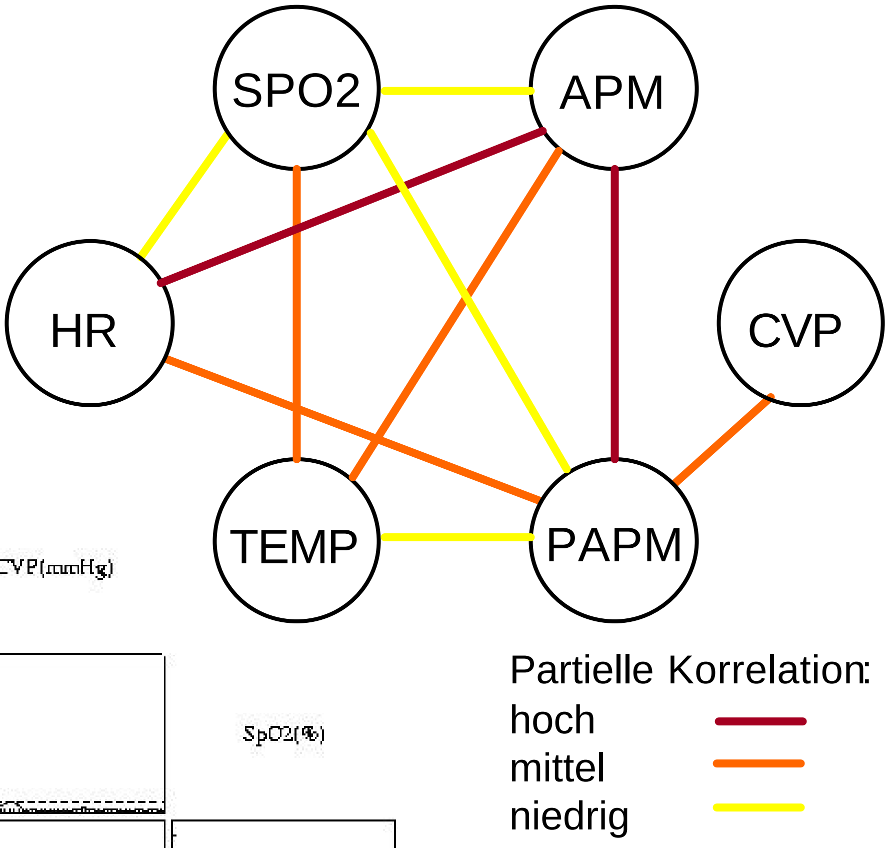
$\Leftrightarrow$ die “partielle spektrale Kohärenz“
ist Null

Test: $H_0 : (i,j) \notin E$
mittels Teststatistik, die auf nichtparametrischem
Schätzer der Kohärenz beruht

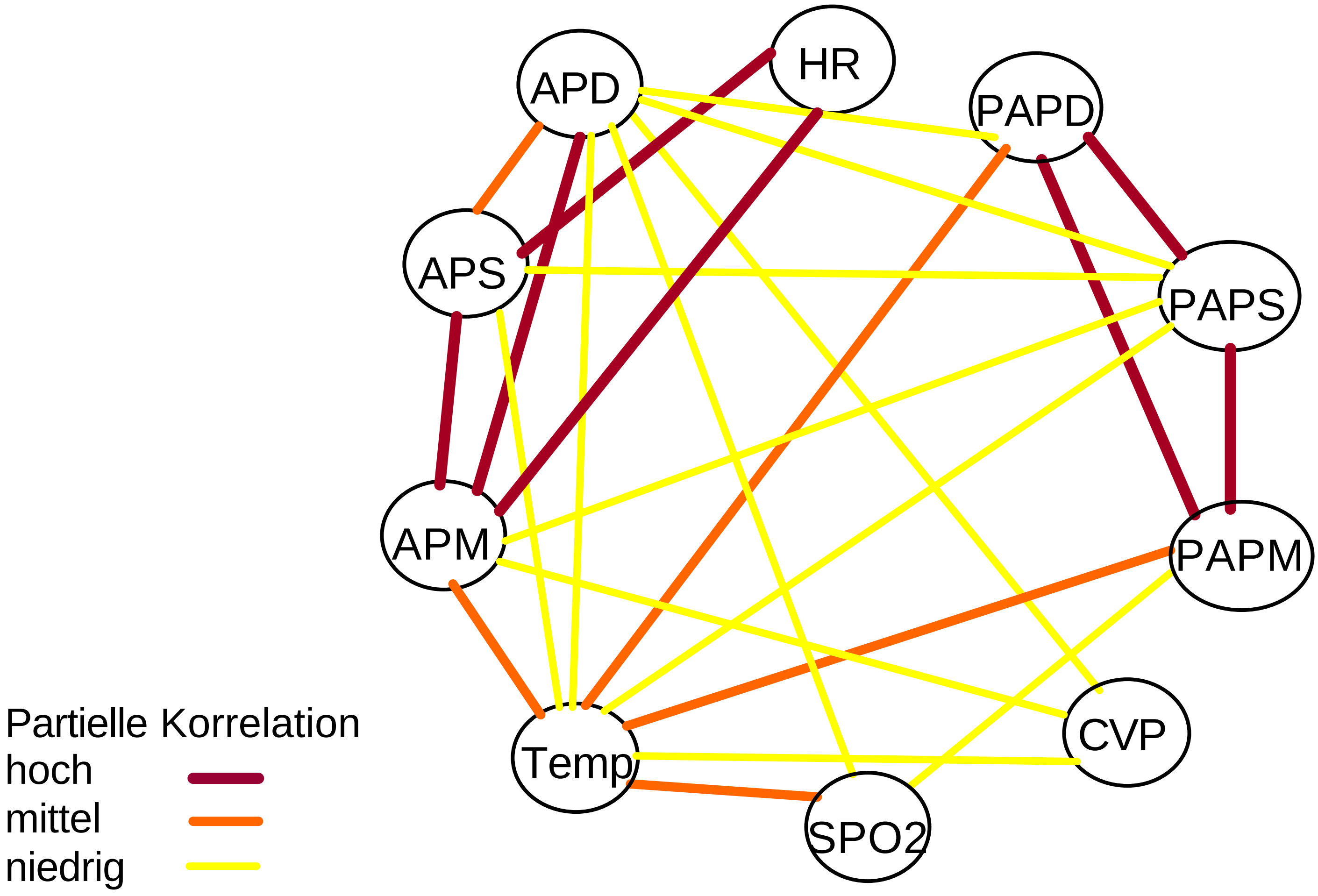
Anwendung auf 6-dim. intensivmedizinische Zeitreihe

Graphische Darstellung der partiellen Korrelationen

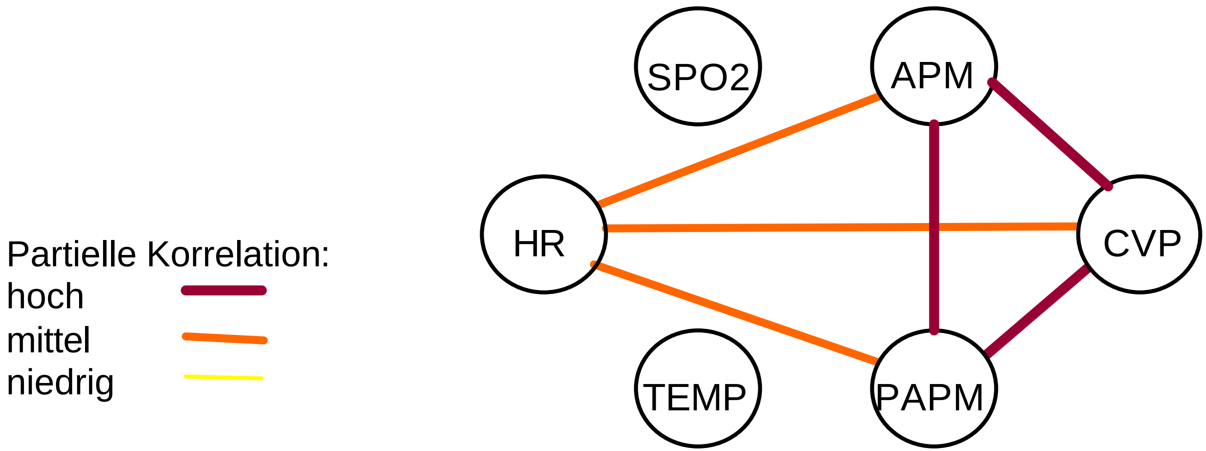




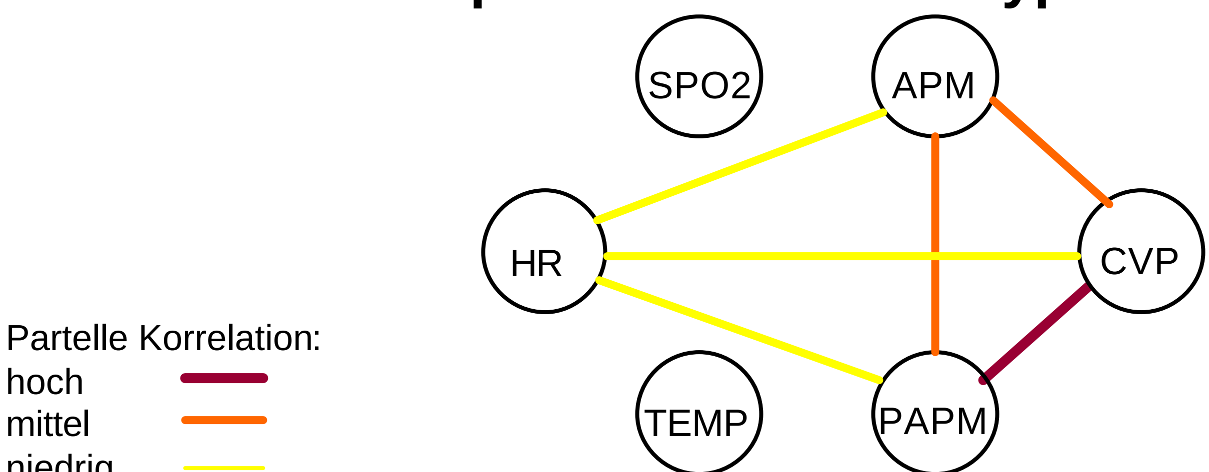
Anwendung auf 10-dimensionale Zeitreihe



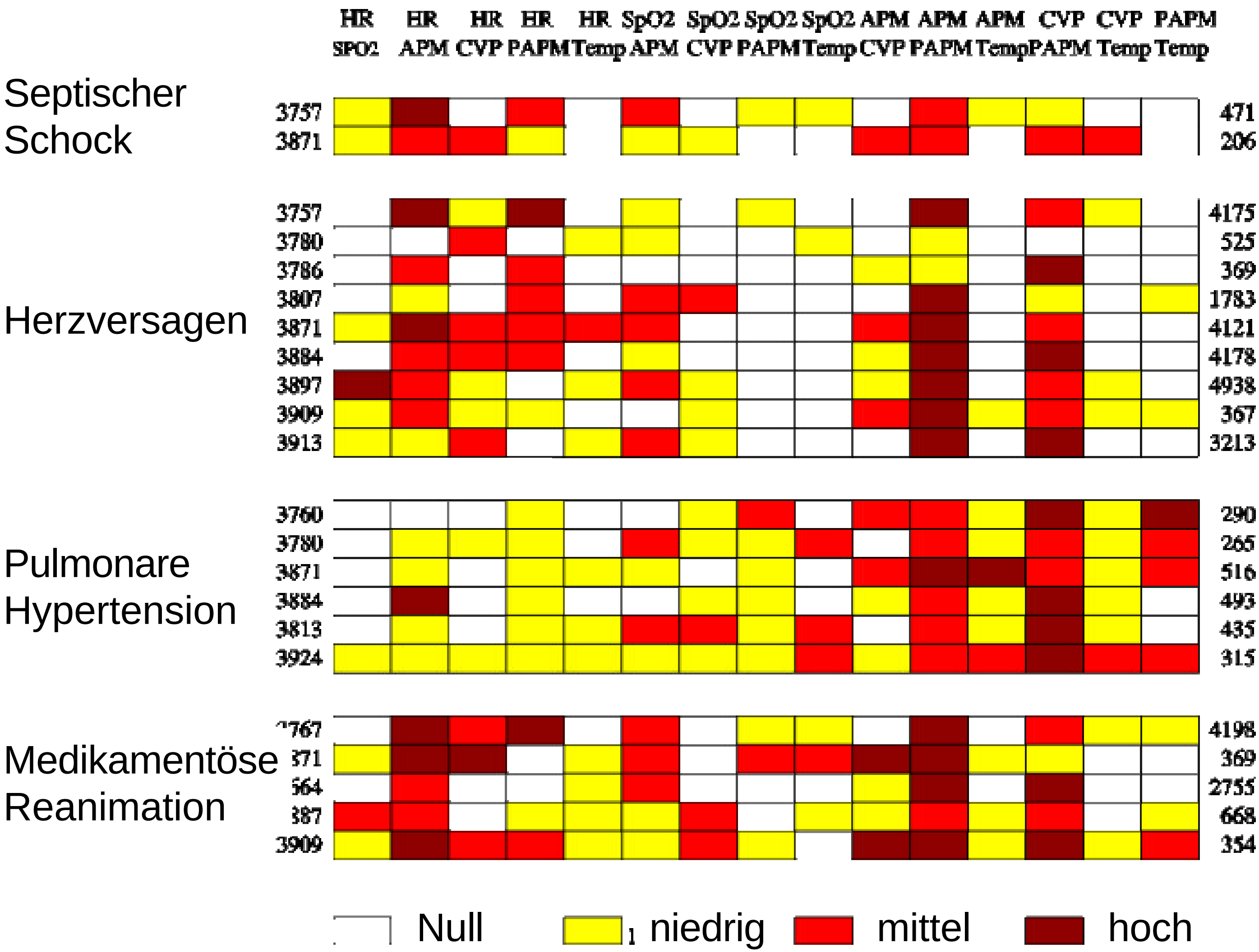
Erwarteter Graph (Normalzustand)



Erwarteter Graph für Pulmonare Hypertension



Partielle Korrelationen für verschiedene klinische Zustände



3) Sliced Inverse Regression SIR

Methode zur Dimensionsreduktion in Regressionsmodellen

$$\begin{tikzcd} \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix} \arrow[rr, "g"] \arrow[dr, "\text{Dimensions-reduktion}"'] & & Y \\ & \boxed{\begin{pmatrix} ß_1^T X \\ \vdots \\ ß_r^T X \end{pmatrix}} \arrow[ur, "f"'] & \end{tikzcd}$$

$$B = \mathrm{span}(\, ß_1 ,\ldots, ß_r)$$

Schätzung des Raums B

Schätzung der Dimension r

Anwendung auf Zeitreihen

$y_t = \left(y_t^1, \ldots, y_t^k\right), \; t = 1, \ldots, N$ Beobachtungen von $\left(Y^1, \ldots, Y^k\right)$

1. Vernachlässigen der zeitlichen Abhängigkeiten

Setze $Y = Y^i, i = 1, \ldots, k$

$$\mathbf{X} = (Y^1, \ldots, Y^{i-1}, Y^{i+1}, \ldots, Y^k)' \in \mathrm{IR}^{k-1}$$

$$Y^i = \tilde{f}(X, \varepsilon) = f(\beta_1' \mathbf{X}, \ldots, \beta_r' \mathbf{X}, \varepsilon)$$

(Annahme: f und $\beta_1, \ldots, \beta_r$ zeitunabhängig)

Für jede Wahl von i schätze k und $\beta_1, \ldots, \beta_r$ mittels

$$\left(y_t, \left(y_t^1, \ldots, y_t^{i-1}, y_t^{i+1}, \ldots, y_t^k\right)\right), \; t = 1, \ldots, N$$

2. Berücksichtigung zeitlicher Abhängigkeiten

"Künstliche" Variablen $Y^{k+1},\ldots,Y^{2k}$: $Y_t^{k+j} = Y_{t-1}^{j},\ j=1,\ldots,k$

Setze $Y = Y^i, i = 1,\ldots,k$

$$\mathbf{X} = (Y^1,\ldots,Y^{i-1},Y^{i+1},\ldots,Y^{2k})' \in IR^{2k-1}$$

$$Y^i = \tilde{f}(X,\varepsilon) = f(\beta_1'\mathbf{X},\ldots,\beta_r'\mathbf{X},\varepsilon)$$

(Annahme: f und $\beta_1,\ldots, \beta_r$ zeitunabhängig)

Für jede Wahl von i schätze r und $\beta_1,\ldots, \beta_r$ mittels

$$\left(y_t, \left(y_t^1,\ldots, y_t^{i-1}, y_t^{i+1},\ldots, y_t^{2k}\right)\right),\ t = 2,\ldots, N$$

Anwendung von SIR auf hämodynamisches System

9-dim. Zeitreihe: arterielle, pulmonalarterielle (syst., diast., Mittel), Zentralvenöser Druck, Herzfrequenz, Temperatur

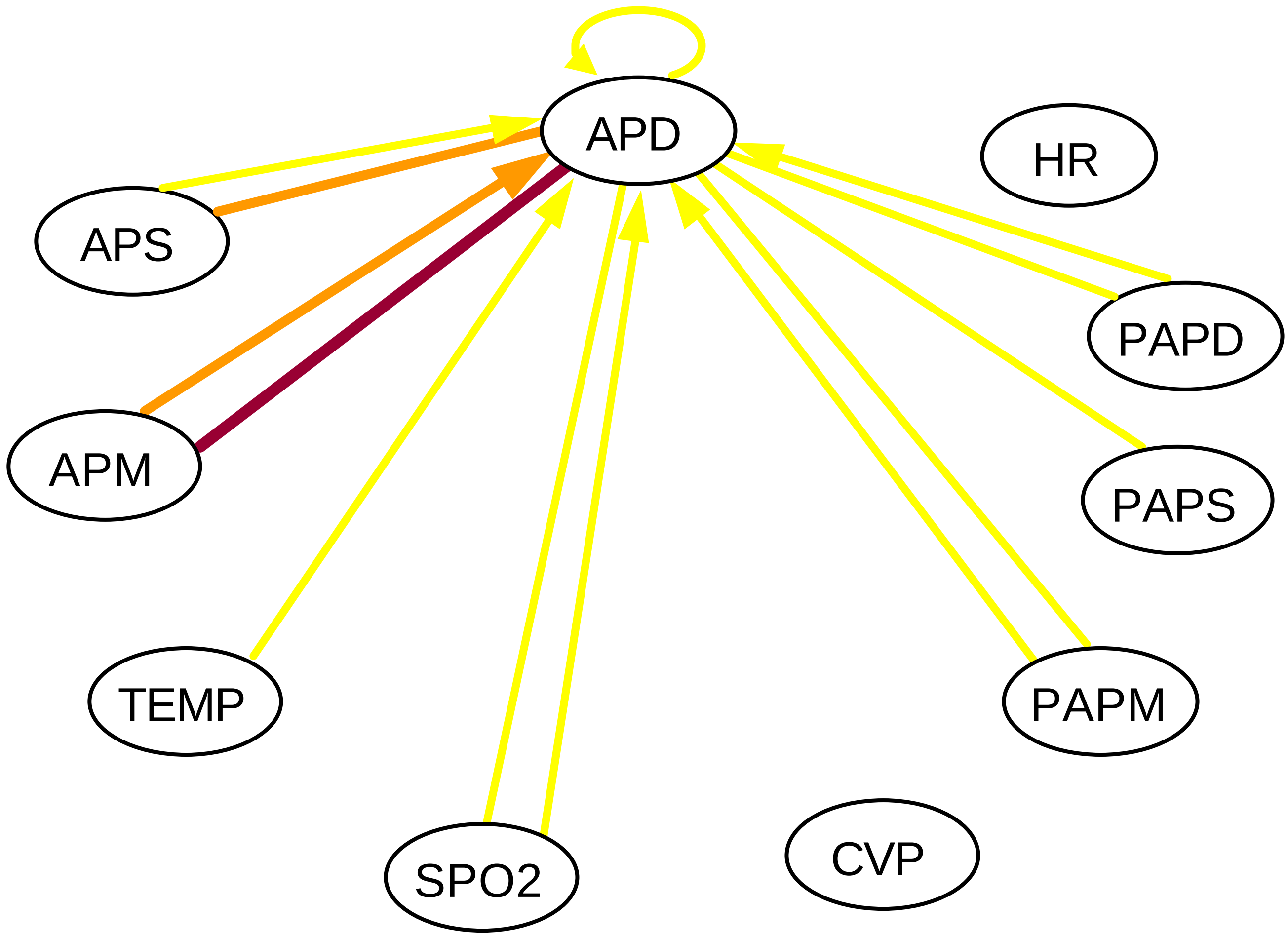
z.B. $Y = HR = f(ß_1 X, \ldots, ß_r X)$, $X =$ alle anderen \ HR

Anzahl der Dimensions-reduzierenden Richtungen:

	APS	APD	APM	Temp	CVP	HR	PAPS	PAPD	PAPM
Vernachl. zeitlicher Abhäng.	3	5	4	6	7	5	3	2	3
Berücks. zeitlicher Abhäng.	2	4	3	5	6	4	3	2	3

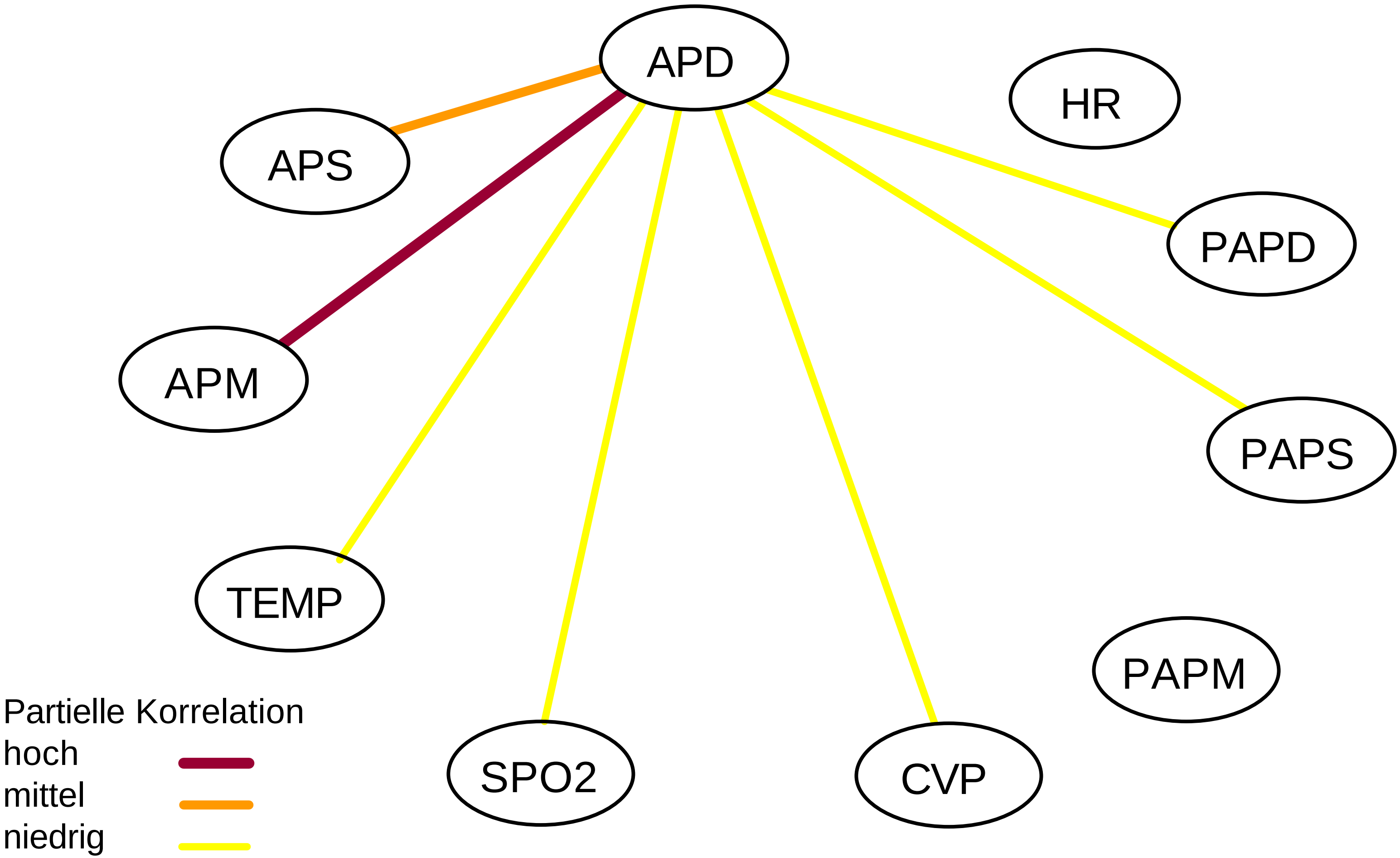
- Leite Zusammenhänge aus Richtungen ab
- Dimensionsreduktion durch Eliminierung von Variablen mit kleinem r

Visualisierung der Einflüsse auf APD



Linien: simultaner Einfluß
Pfeile: Zeit-versetzter Einfluß

Vergleich mit Graphischem Model



4) Dynamische Faktorenanalyse

Definiere dynamisches Faktorenmodell für $\mathbf{Y}(t)$ durch:

$$\mathbf{Y}(t) - m_{\mathbf{Y}} = \sum_{s=-\infty}^{\infty} \Lambda(s)\mathbf{F}(t-s) + e(t),$$

- $\mathbf{F}(t)$ — r-dim. stoch. Prozeß „Faktorreihe" , $r < k$,
- $e(t)$ — k-dim. stoch. Prozeß (Fehler)
- $\Lambda(s)$ — $k \times r$-Faktorladungs-Matrizen

plus weitere Annahmen

häufig sinnvoll:
Einsatz kausaler Filter

$$\mathbf{Y}(t) - m_{\mathbf{Y}} = \sum_{s=0}^{m} \Lambda(s)\mathbf{F}(t-s) + e(t)$$

Dynamisches Faktorenmodell von Peña & Box (1987)

einfache Identifikation latenter Faktoren in multivariaten Zeitreihen

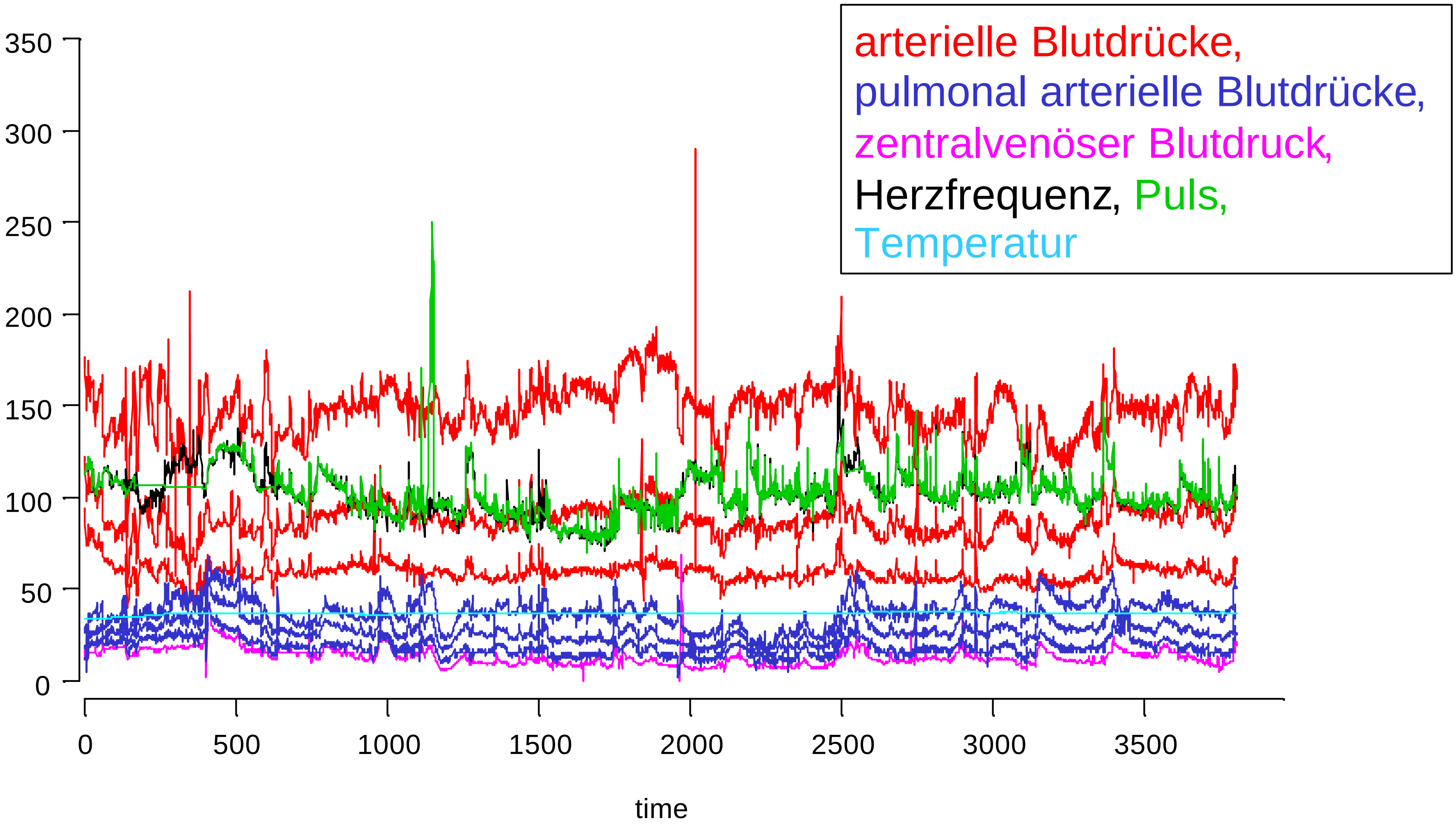
Modell mit unabhängigen Faktoren:

$$\mathbf{Y}(t) - m_{\mathbf{Y}} = \Lambda \mathbf{F}(t) + e(t)$$

Annahmen:

- $\mathbf{F}(t)$ — r-dim., stationärer,invertierbarer
  VARMA(p,q)-Prozeß $\Phi_{\mathrm{F}}(\mathrm{B})\,\mathbf{F}(t) = \Theta_{\mathrm{F}}(\mathrm{B})\,\mathbf{u}(t)$
- Unabhängigkeit der r Faktoren $F_1(t), \ldots, F_r(t)$
  mit diagonalen Matrizen der Filter Φ_{F} und Θ_{F}
- Kovarianzmatrix $\Sigma_{\mathbf{u}}$ diagonal
- $\Lambda' \Lambda = \mathrm{I}_r$

Beispiel: Zeitreihe von 10 Variablen des hämodynamischen Systems eines Patienten auf der Intensivstation



Beispiel: Analyse der standard. Zeitreihe nach Peña und Box

erklärte Varianz

3 Faktoren	**0.7883**
4 Faktoren	**0.8931**

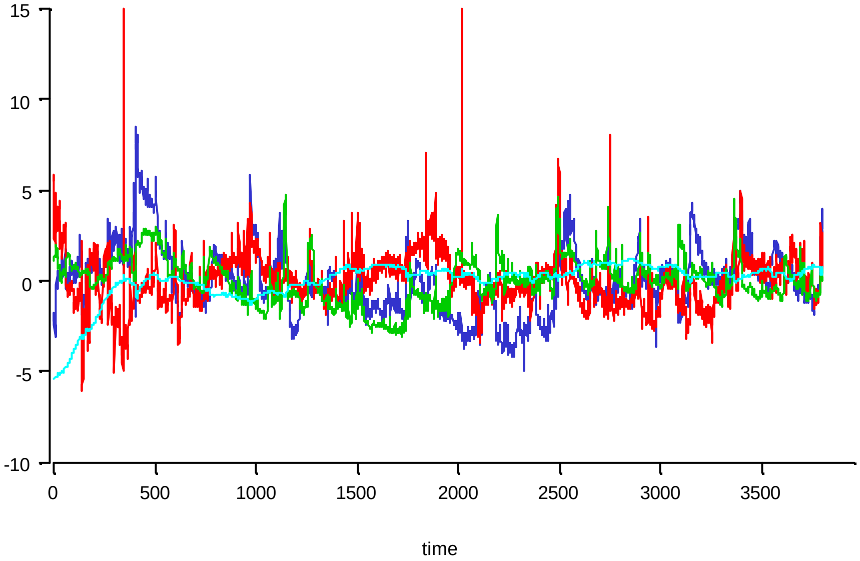
geschätzte Ladungsmatrix nach Varimax-Rotation durch Analyse der geschätzten Kreuzkovarianzmatrizen

Geschätzte Ladungsmatrix

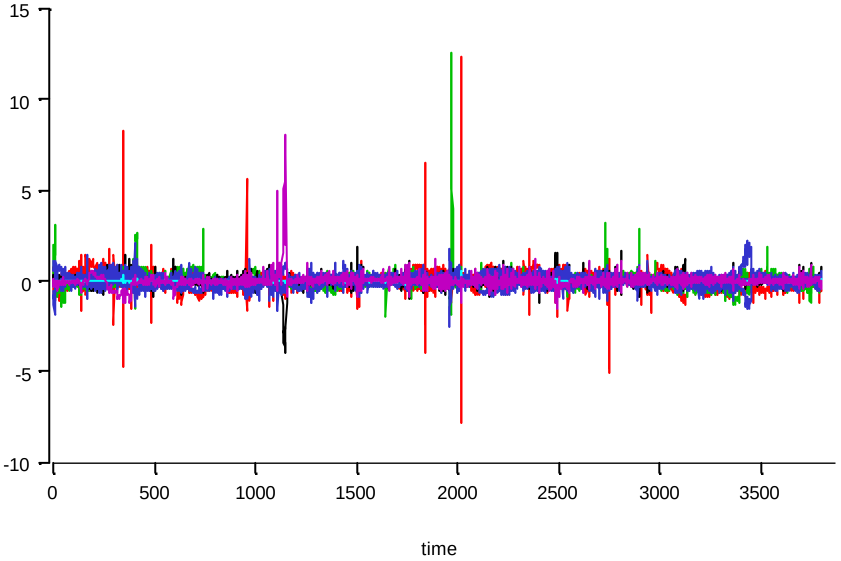
APDIA	0.0076	0.4538	0.0629	-0.1978
APM	0.0336	0.6032	-0.0512	-0.0053
APSYS	-0.0418	0.6528	-0.0030	0.1684
CVP	0.4339	-0.0336	0.0330	-0.2340
HR	0.0082	-0.0208	0.7031	0.0128
PAPDIA	0.4882	0.0104	0.0302	-0.0960
PAPM	0.5389	0.0039	0.0199	0.0752
PAPSYS	0.5301	0.0213	-0.0858	0.2141
TEMP	0.0034	-0.0315	0.0360	0.9038
PULS	0.0015	0.0320	0.6985	-0.0020

Beispiel (Fs.): Analyse der standard. Zeitreihe n. Peña und Box

Schätzung der vier extrahierten Faktorzeitreihen nach Glättung durch Kalman-Filter



Approximationsfehler



Interpretation

pulmonal arterielle Blutdrücke

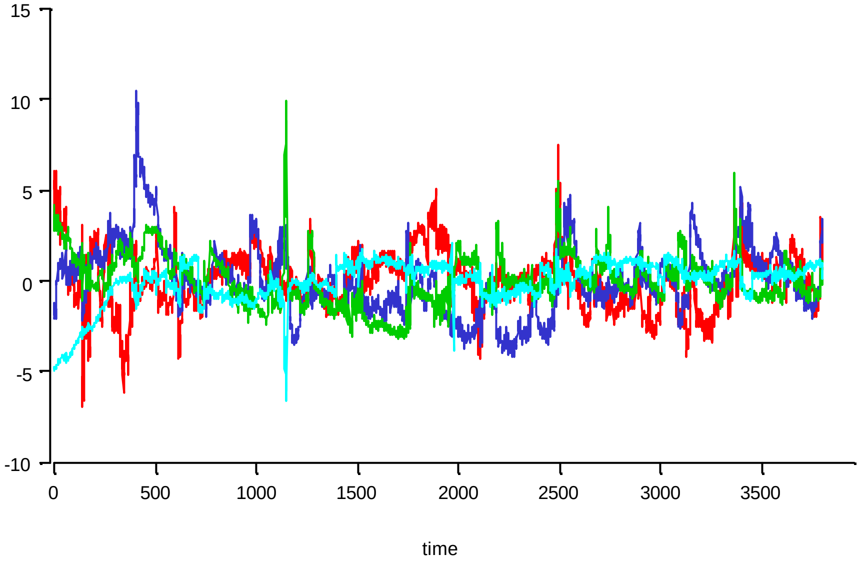
arterielle Blutdrücke

Herzfrequenz / Puls

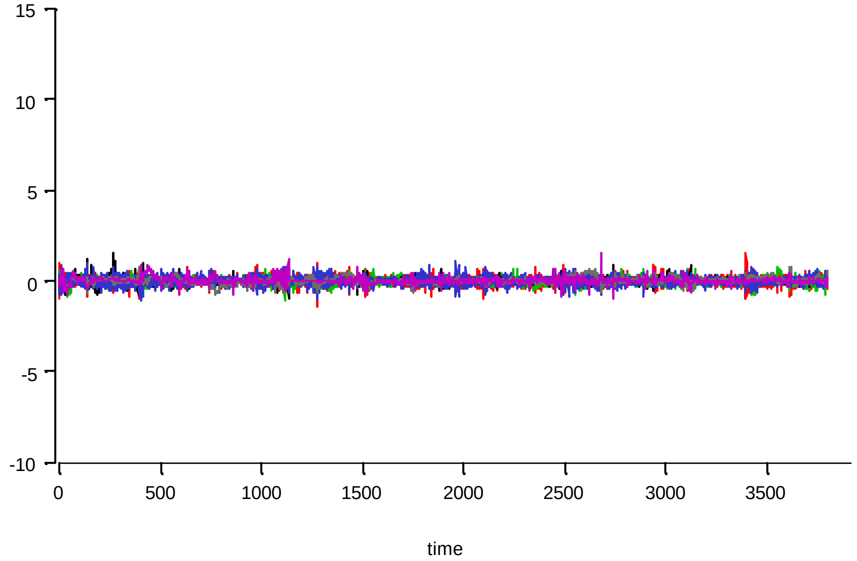
Temperatur

Beispiel (Fs.): Analyse der Median-geglätteten Zeitreihe mit „gleitender” Analyse

Schätzung der vier extrahierten „gleitenden” Faktorenreihen



Approximationsfehler



Interpretation

pulmonal arterielle Blutdrücke

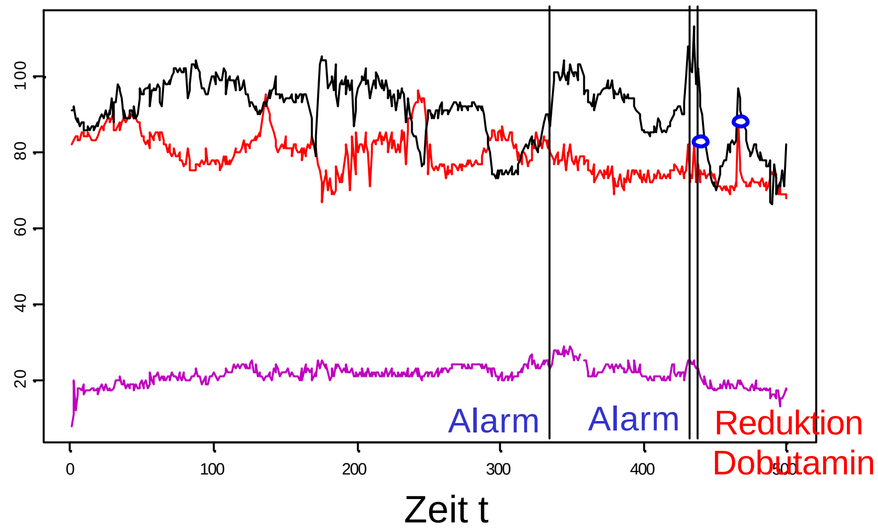
arterielle Blutdrücke

Herzfrequenz / Puls

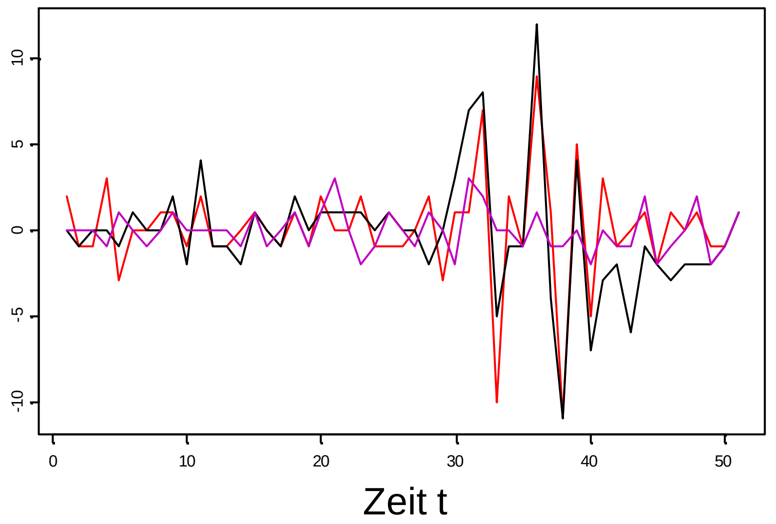
Temperatur

Alarmsystem basierend auf Mustererkennung

Beobachtete Zeitreihe



Differenzierte Zeitreihe, 400-450



Arterieller Druck
Pulmonalarterieller Druck
Herzfrequenz

Zur Anwendung

Besondere Anforderungen auf der Intensivstation
wie in der Prozeßkontrolle:

- ⇨ automatische Verfahren und schnelle Algorithmen
- ⇨ einfache Interpretation ($\rightarrow$ Arzt)
  der Faktoren bzw. der Ursache eines Alarms
- ⇨ Berücksichtigung von Änderungen in der Abhängigkeitssturktur
- ⇨ Umgang mit Ausreißern

Literatur

Becker, C., Fried, R. (2001). Sliced Inverse Regression for High-dimensional Time Series. Technical Report 14/2001, SFB 475, University of Dortmund, Germany.

Brillinger, D.R. (1981). *Time Series. Data Analysis and Theory*. San-Francisco: Holden-Day.

Brillinger, D.R. (1996). Remarks Concerning Graphical Models for Time Series and Point Processes. *Revista de Econometria* **16**, 1-23.

Cox, D.R., Wermuth, N. (1996). *Multivariate Dependencies*. London: Chapman & Hall.

Dahlhaus, R. (2000). Graphical Interaction Models for Multivariate Time Series. *Metrika,* **51**, 157-172

Dahlhaus, R., Eichler, M. (2002). SPECTRUM. C-programme available at http://www.statlab.uni-heidelberg.de/projects

Fried, R., Didelez, V. (2002). Decomposability and Selection of Graphical Models for Multivariate Tim Series. Technical Report 17/2002, SFB 475, University of Dortmund, Germany.

Gather, U., Fried, R., Imhoff, M., Becker, C. (2002). Patterns of Dependencies in Dynamic Multivariat Data. Preprint, Department of Statistics, University of Dortmund, Germany.

Gather, U., Fried, R., Lanius, V., Imhoff, M. (2002). Online Monitoring of High-dimensional Physiological Time Series - a Case-study. *Estadistica*, in press.

Gather, U., Imhoff, M., Fried, R. (2002). Graphical Models for Multivariate Time Series From Intensive Care Monitoring. *Statistics in Medicine*, in press.

Peña, D., Box, G.E.P. (1987). Identifying a Simplifying Structure in Time Series. *JASA* **82**, 836-843.