Machine Learning for Healthcare

HST.956, 6.S897

Lecture 6: Physiological time-series

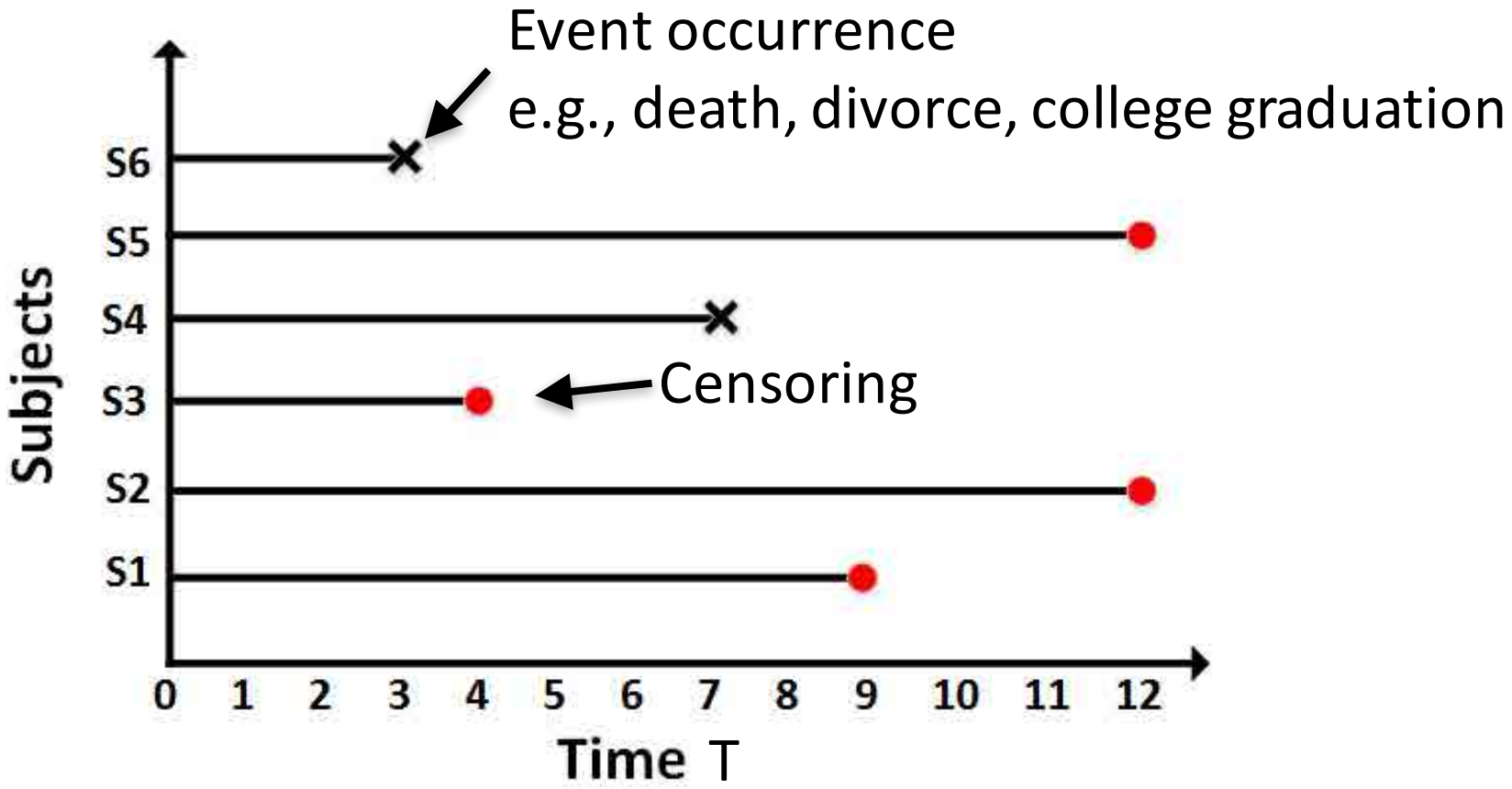
David Sontag

HEALTH SCIENCES
& TECHNOLOGY

Outline of today's lecture

1. **Recap of risk stratification**
2. Physiological time-series
   - Monitoring babies in neonatal ICUs
   - Detecting atrial fibrillation

Survival modeling with right-censored data



[Wang, Li, Reddy. Machine Learning for Survival Analysis: A Survey. 2017]

Notation and formalization

- f(t) = be the probability of death at time t
- Survival function: $S(t) = P(T > t) = \int_t^{\infty} f(x)dx$

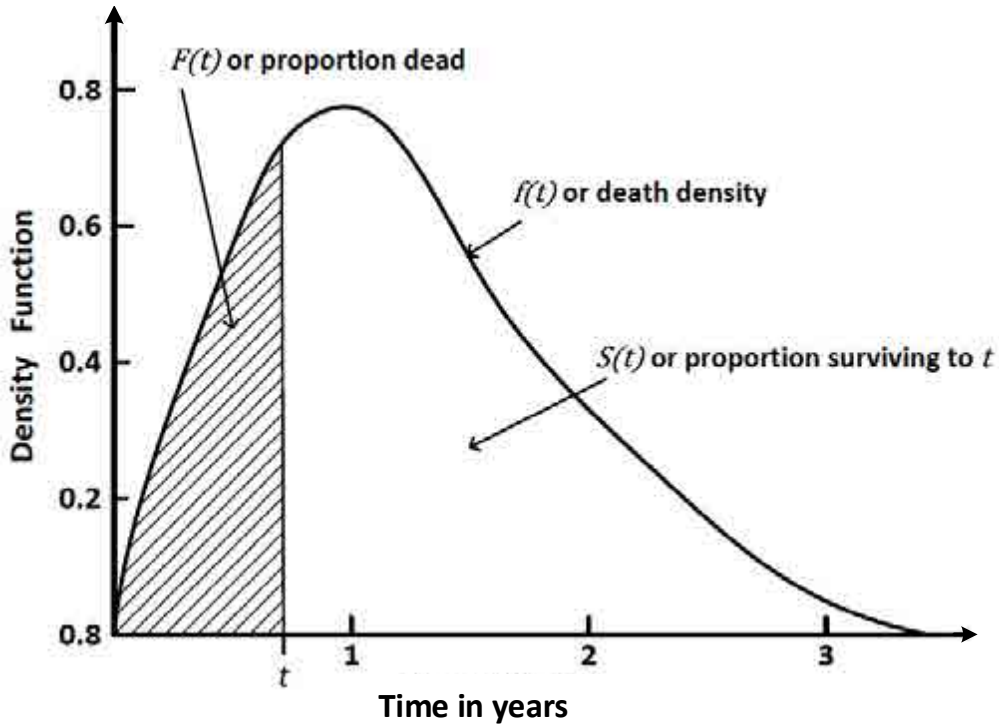


Fig. 2: Relationship among different entities $f(t)$, $F(t)$ and $S(t)$.

[Wang, Li, Reddy. Machine Learning for Survival Analysis: A Survey. 2017]

[Ha, Jeong, Lee. Statistical Modeling of Survival Data with Random Effects. Springer 2017]

Maximum likelihood estimation

- Two kinds of observations: censored and uncensored

Uncensored likelihood

$$p_{\boldsymbol{\theta}}(T = t \mid \mathbf{x}) = f(t)$$

Censored likelihood

$$p_{\boldsymbol{\theta}}^{\text{censored}}(t \mid \mathbf{x}) = p_{\boldsymbol{\theta}}(T > t \mid \mathbf{x}) = S(t)$$

- Putting the two together, we get:

$$\sum_{i=1}^{n} b_i \log p_{\boldsymbol{\theta}}^{\text{censored}}(t \mid \mathbf{x}) + (1 \quad b_i) \log p_{\boldsymbol{\theta}}(t \mid \mathbf{x})$$

Optimize via gradient or stochastic gradient ascent!

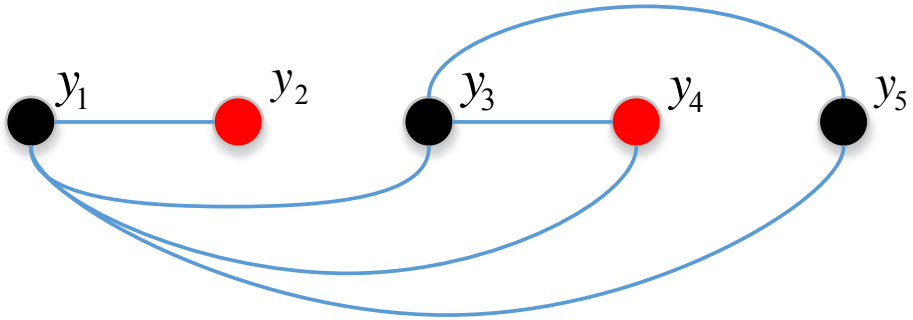
Evaluation for survival modeling

- Concordance-index (also called C-statistic): look at model's ability to predict *relative* survival times:

$$\hat{c} = \frac{1}{num} \sum_{i:\delta_i = 0} \sum_{j:y_i < y_j} I[S(\hat{y}_j|X_j) > S(\hat{y}_i|X_i)]$$

- Illustration – blue lines denote pairwise comparisons:

Black = uncensored
Red = censored



- Equivalent to AUC for binary variables and no censoring

[Wang, Li, Reddy. Machine Learning for Survival Analysis: A Survey. 2017]

Final thoughts on survival modeling

- Could also evaluate:
  - Mean-squared error for uncensored individuals
  - Held-out (censored) likelihood
  - Derive binary classifier from learned model and check calibration

- Partial likelihood estimators (e.g. for cox-proportional hazards models) can be much more data efficient

Dealing with non-stationarity

- Baseline: Retrain the model with most recent data
- How to best use historical data?
  - Impute or transform historical data to look like current data (e.g., Ganin et al., JMLR ‘16)
  - Reweight historical data to look like current data (see e.g. Sugiyama and Kawanabe, ‘12)
  - Online algorithm that adapts quickly (see e.g. Shen et al. AI Stats ‘18)

Recap of risk stratification

- Classification vs. survival modeling (regression)
- Causal interpretation of predictive features
- Imputation of missing data

Outline of today's lecture

1. Recap of risk stratification
2. **Physiological time-series**
   - Monitoring babies in neonatal ICUs
   - Detecting atrial fibrillation

Physiological time-series

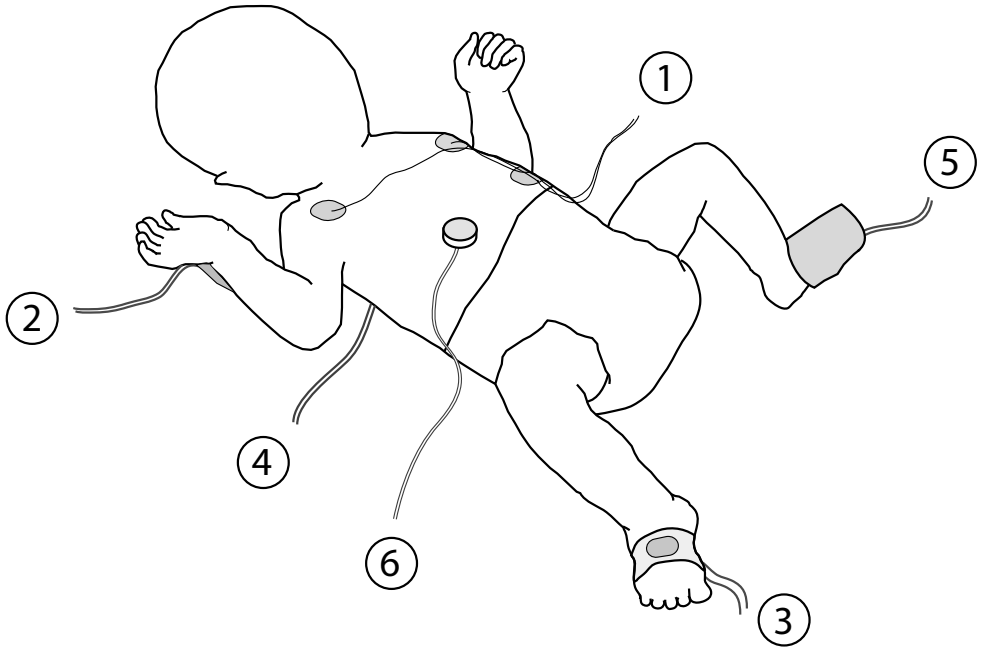


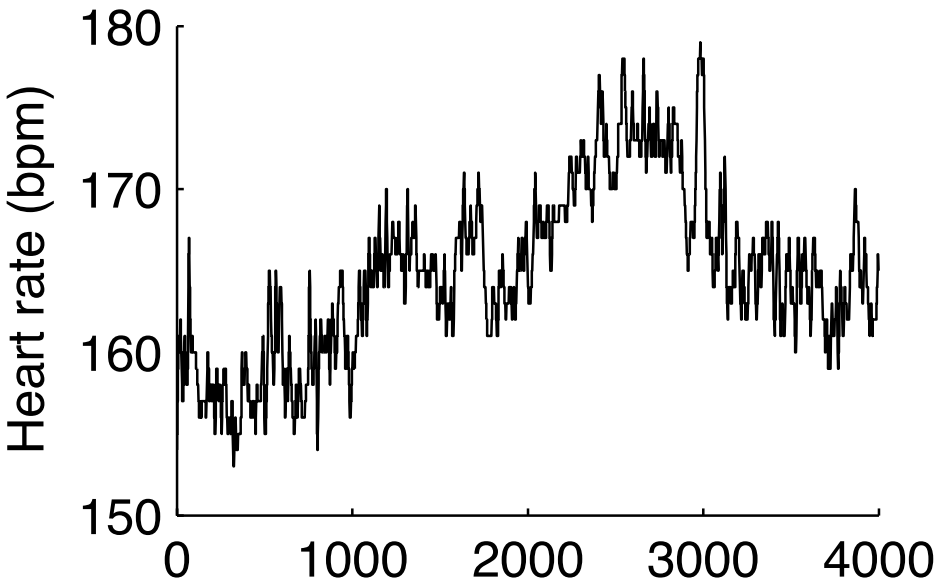
Fig. 4. Probes used to collect vital signs data from an infant in intensive care. 1) Three-lead ECG, 2) arterial line (connected to blood pressure transducer), 3) pulse oximeter, 4) core temperature probe (underneath shoulder blades), 5) peripheral temperature probe, 6) transcutaneous probe.

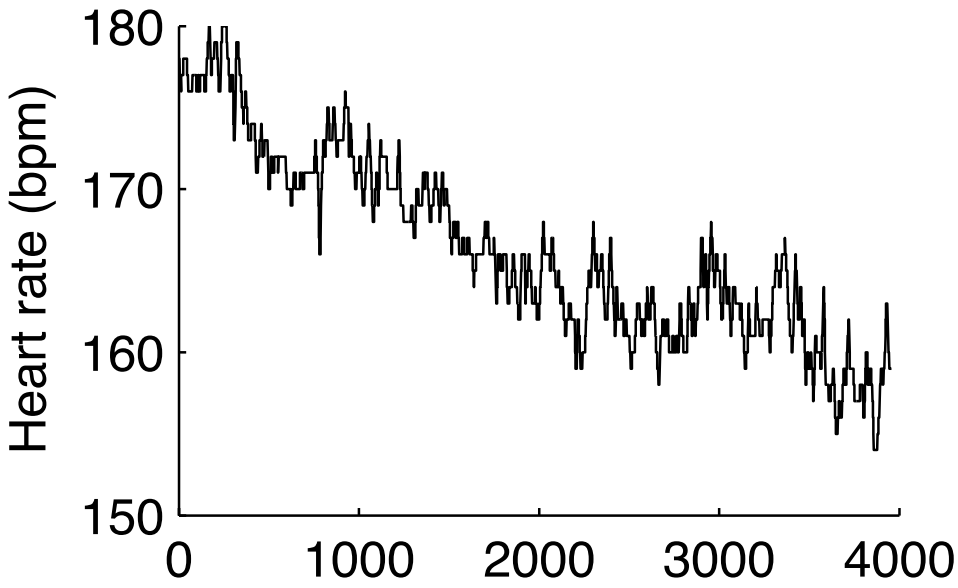
(Quinn et al., TPAMI 2008)

Physiological time-series

- Typical use cases:
  1. Infer true physiological signal from noisy observations
  2. Risk stratification, e.g. predict clinical deterioration, or diagnosis
- Approach taken depends on:
  - Is labeled data available?
  - Do we have a good mechanistic/statistical model?
  - How much training data is there?

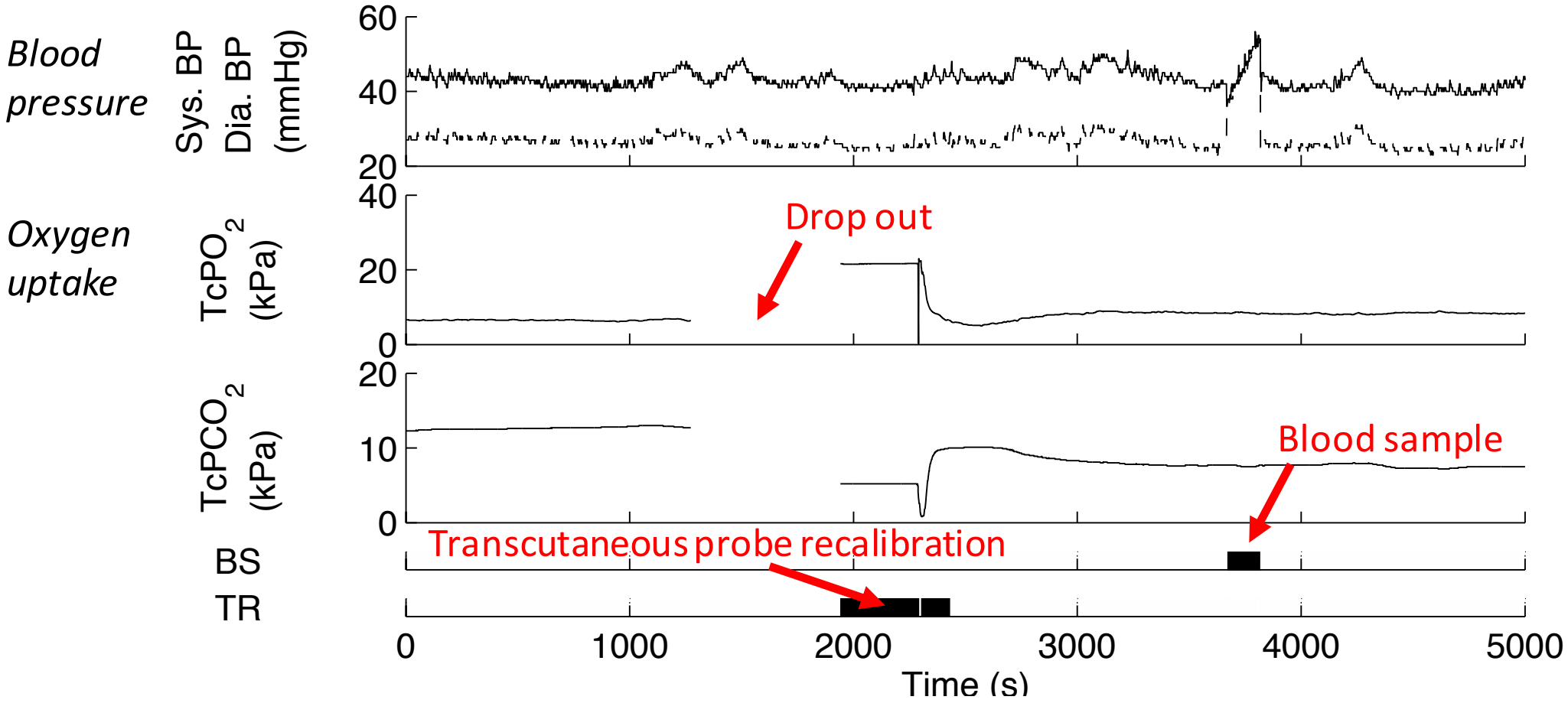
Two very different trajectories





(Quinn et al., TPAMI 2008)

Problem: measurements confounded by interventions & measurement errors



(Quinn et al., TPAMI 2008)

Can we identify the artifactual processes?

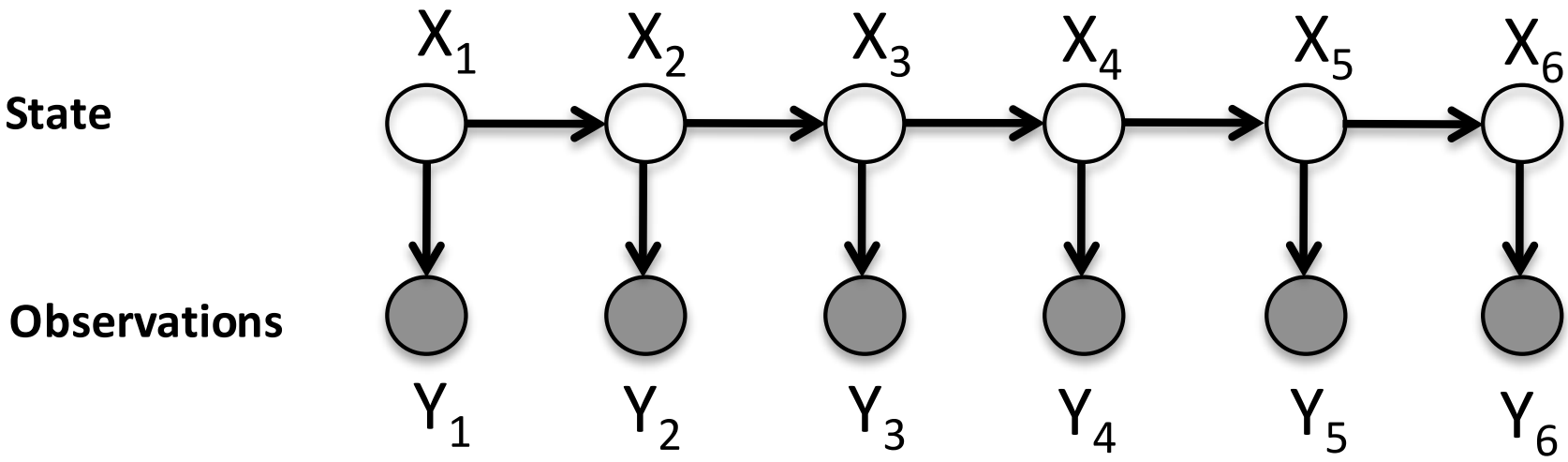
- Once identified, can remove for use in downstream predictive tasks (must deal with missing data)
- Can help mitigate **alarm fatigue** by not alerting the clinicians when unnecessary
- More broadly, can we maintain beliefs about the true physiological values of a patient?

(Switching) linear dynamical systems

- Conditioned on s_t, linear Gaussian state-space models (Kalman filters):

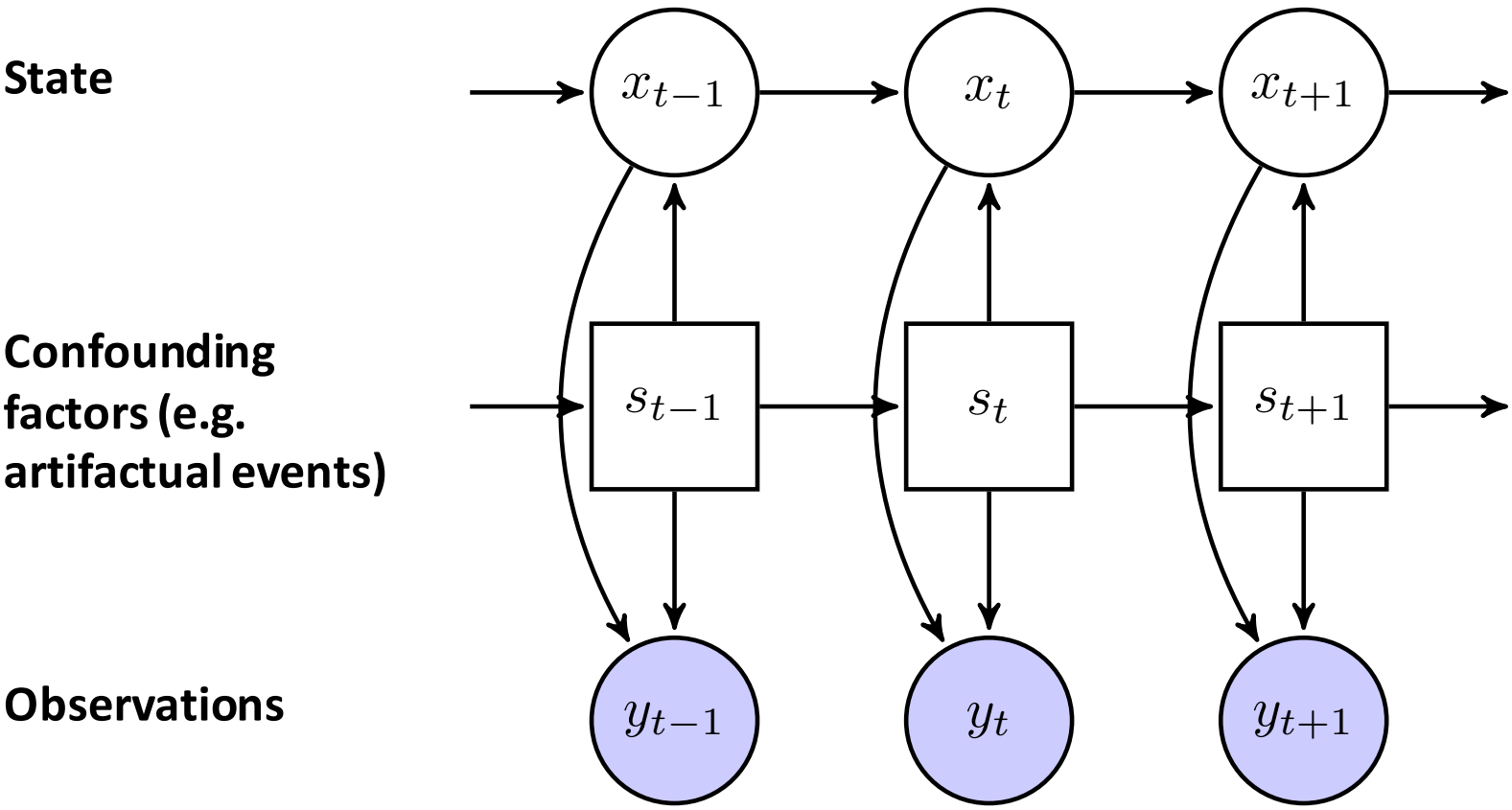
$$\mathbf{x}_t \sim \mathcal{N}\left(\mathbf{A}^{(s_t)}\mathbf{x}_{t-1} + \mathbf{d}^{(s_t)}, \mathbf{Q}^{(s_t)}\right)$$

$$\mathbf{y}_t \sim \mathcal{N}\left(\mathbf{C}^{(s_t)}\mathbf{x}_t, \mathbf{R}^{(s_t)}\right)$$



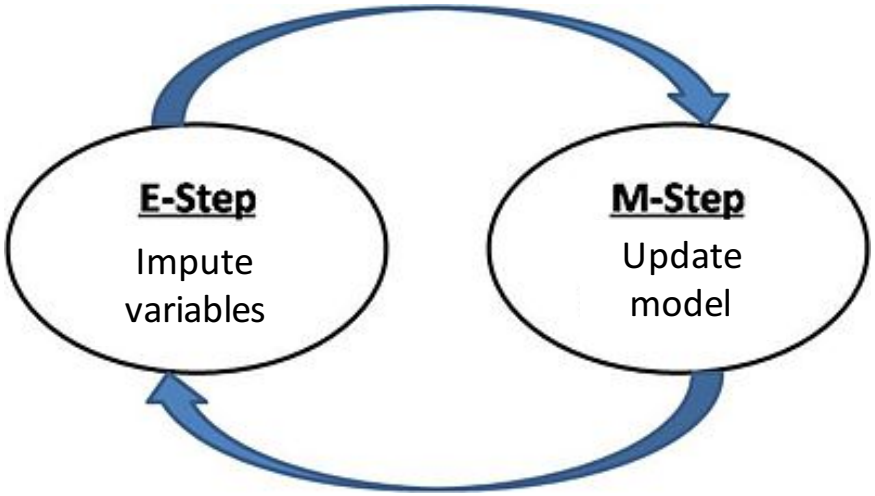
(Switching) linear dynamical systems

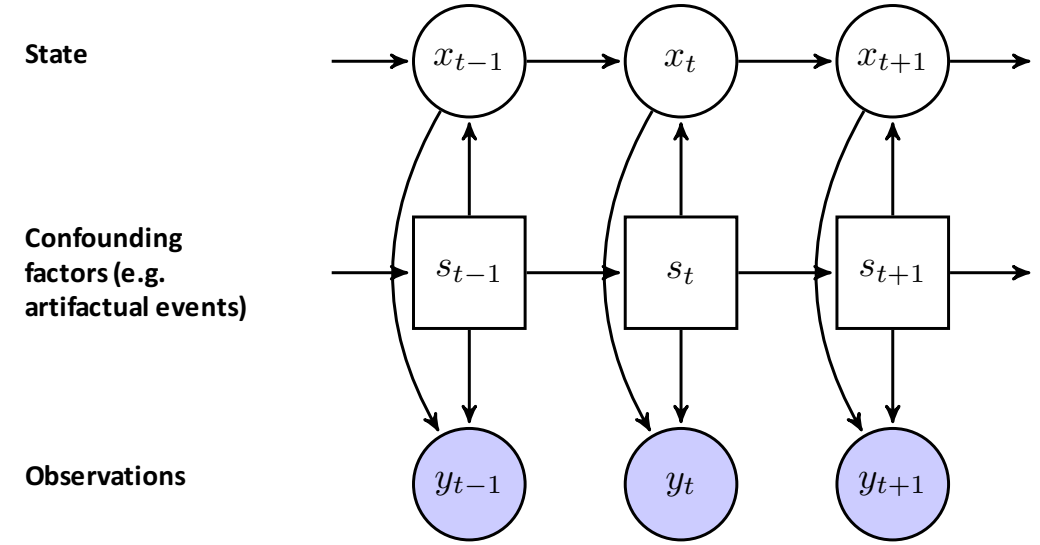
- Full model:



Learning SLDS models

- Assume some labeled training data {**s,y**}
- *True state **x** assumed to never be observed*
- Learn using expectation maximization





Parameterizing model

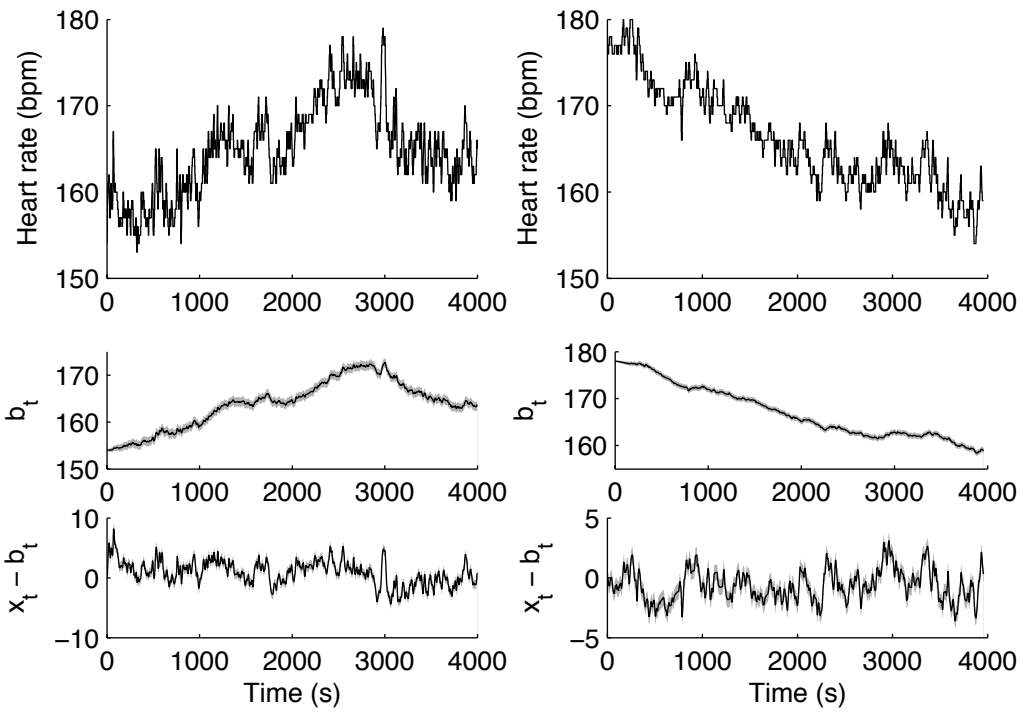
- Normal heart rate dynamics are well-modeled using an autoregressive process, e.g.

$$x_t \quad b_t \quad \sim \quad \mathcal{N} \quad \sum_{k=1}^{p_1} \left(\alpha_k (x_{t-k} \quad b_{t-k}), \eta_1 \right)$$

$$b_t \quad \sim \quad \mathcal{N} \quad \sum_{k=1}^{p_2} \left({}_k b_{t-k}, \eta_2 \right) \left(,\right.$$

Baseline process (smooth) →

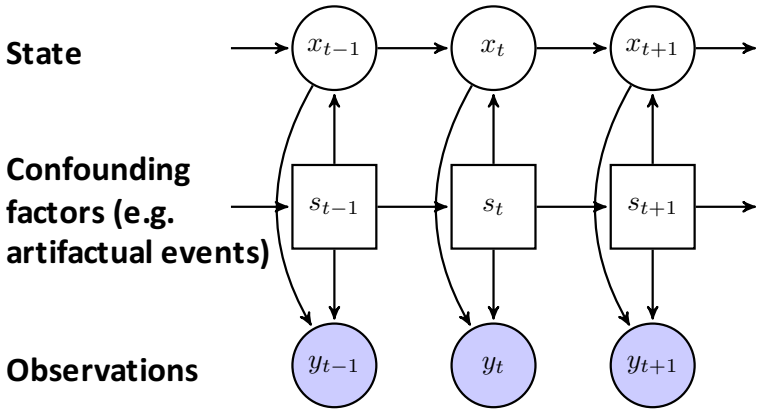
Zero-mean, high frequency →



(Quinn et al., TPAMI 2008)

Parameterizing model

- One can use domain knowledge to specify parts of the artifacts model
  - Probe dropouts modeled by removing dependence of observation y_t on patient state x_t
  - Temperature probe disconnection: exponential decay to room temperature



Evaluation

- 3-fold cross validation, where for each fold train on 10 babies and test on 5
- 24-hours of data for each baby
- Normal dynamics refit for test babies using a 30-minute section near the start

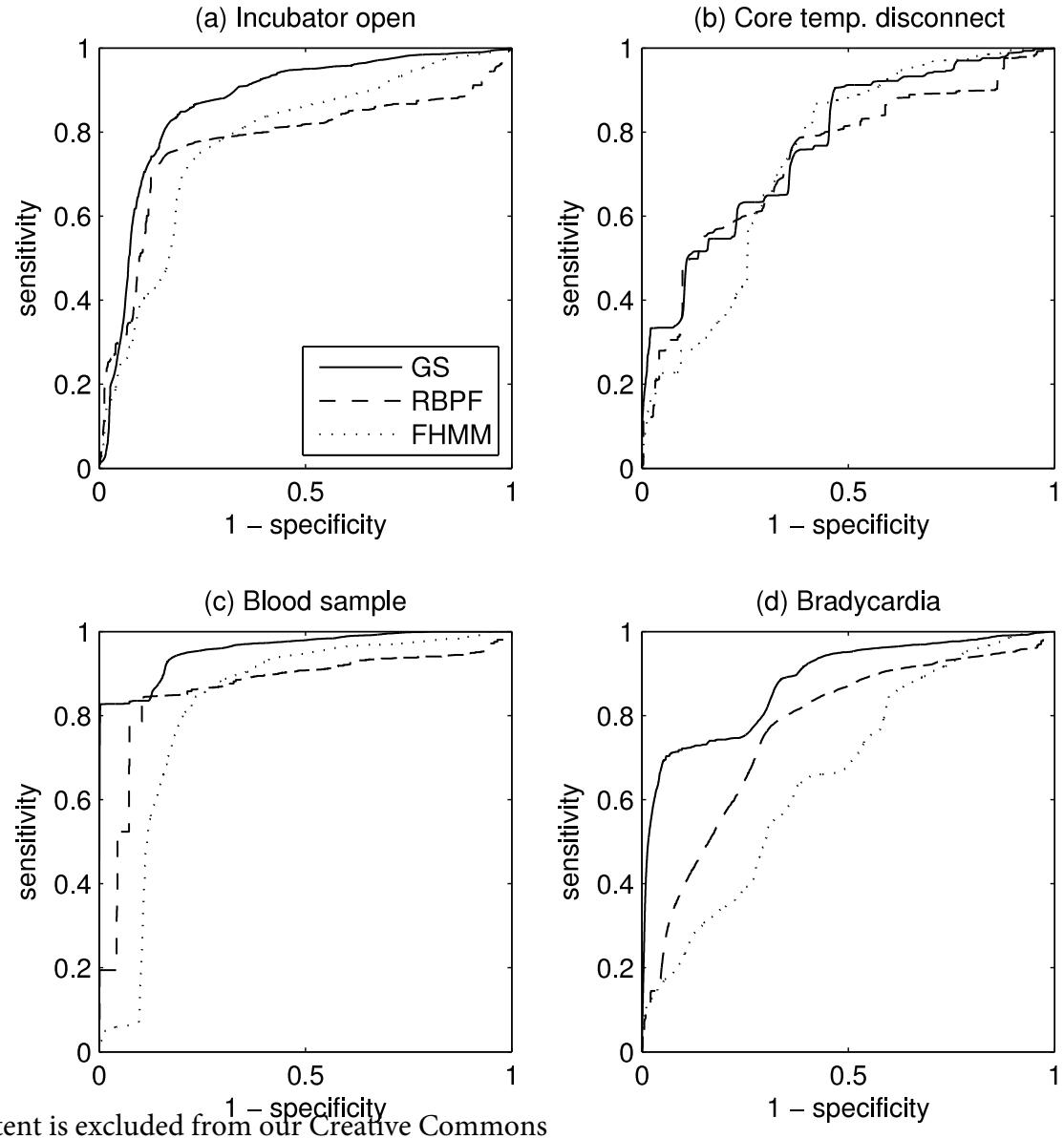
(Quinn et al., TPAMI 2008)

Evaluation

GS = Gaussian-sum approximation (used for inference)

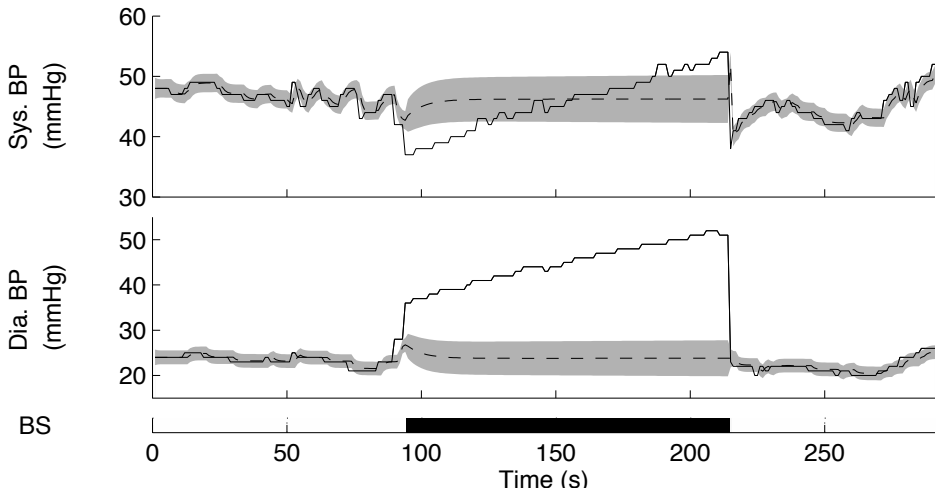
RBPF = Rao-Blackwellized particle filtering approximation (used for inference)

FHMM = Factorial HMM (simpler model which does not model normal physiological dynamics)

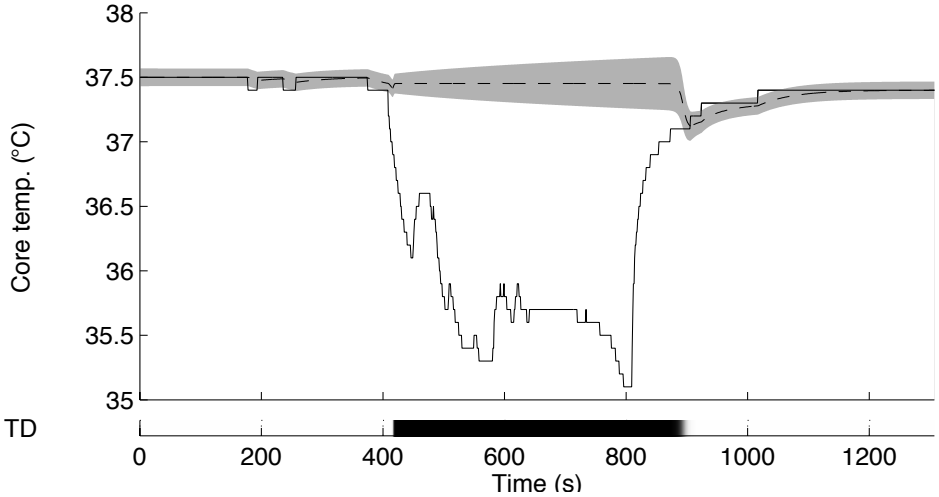


(Quinn et al., TPAMI 2008)

Inference of physiological state



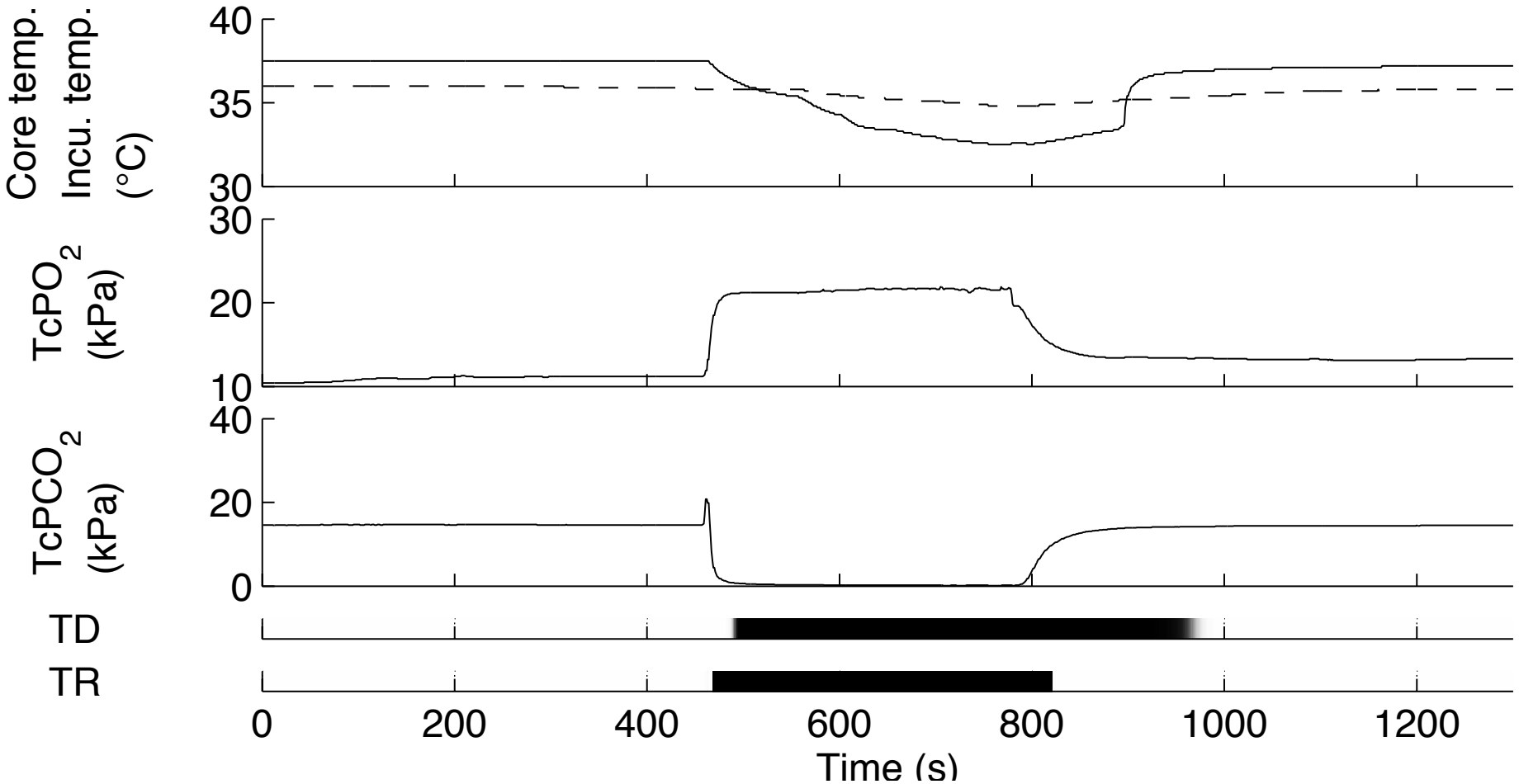
Blood sample draw



Temperature probe disconnection

(Quinn et al., TPAMI 2008)

Inferred switch settings



TD= core temperature probe disconnection

TR = recalibration

Outline of today's lecture

1. Recap of risk stratification
2. Physiological time-series
   - Monitoring babies in neonatal ICUs
   - **Detecting atrial fibrillation**

Detecting atrial fibrillation

AliveCore ECG device

ECG = electrocardiogram

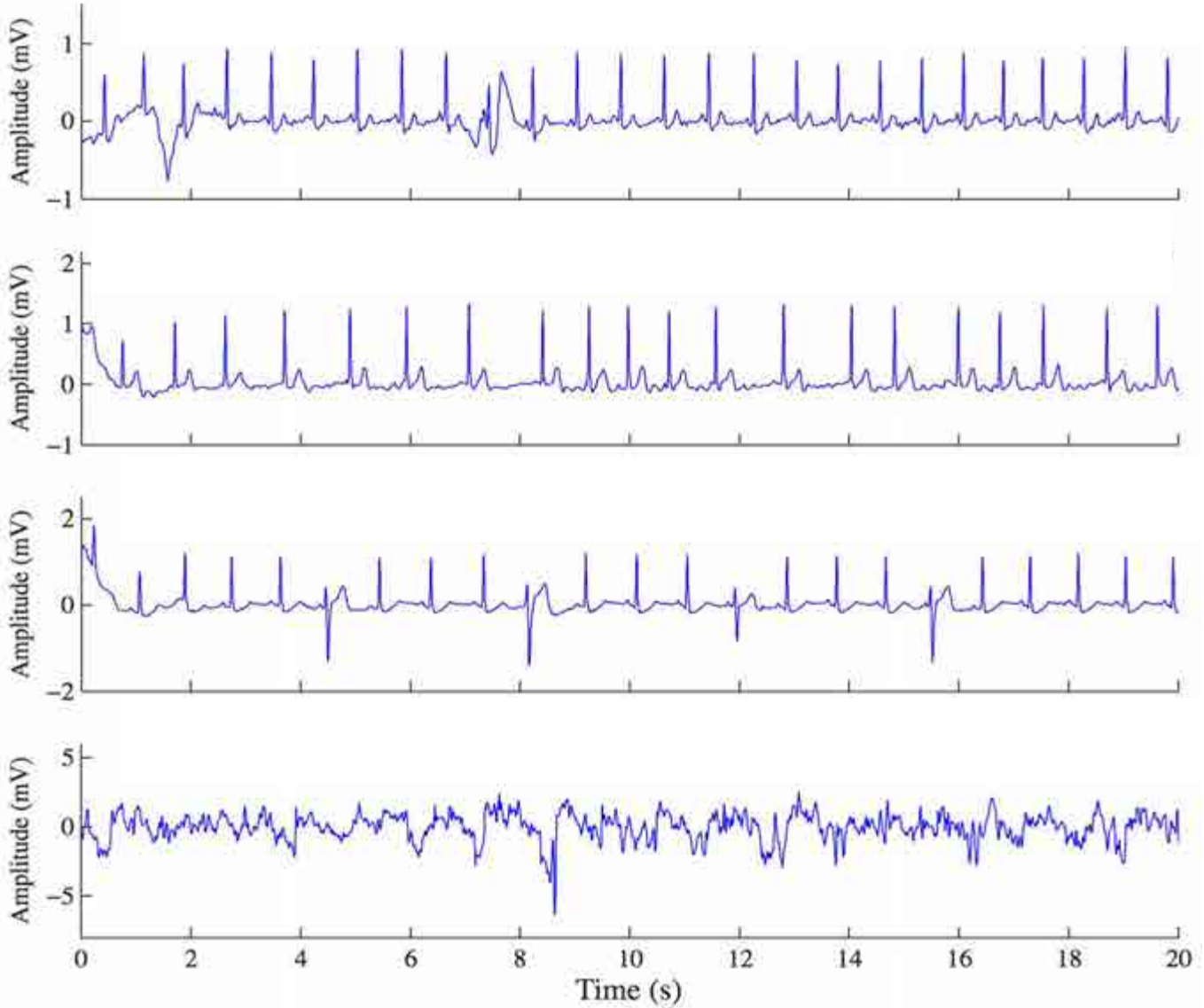
What type of heart rhythm?

Normal rhythm

AF rhythm

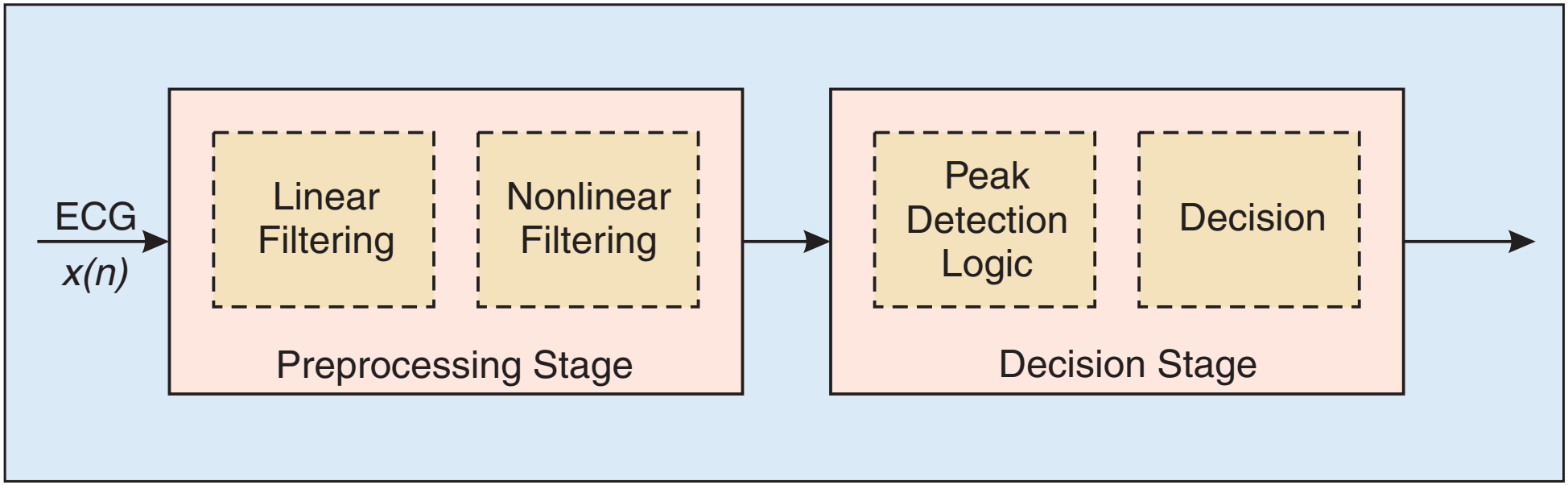
Other rhythm

Noisy recording



[Clifford, Liu, Moody, Mark. PhysioNet Computing in Cardiology Challenge 2017]

Traditional approach

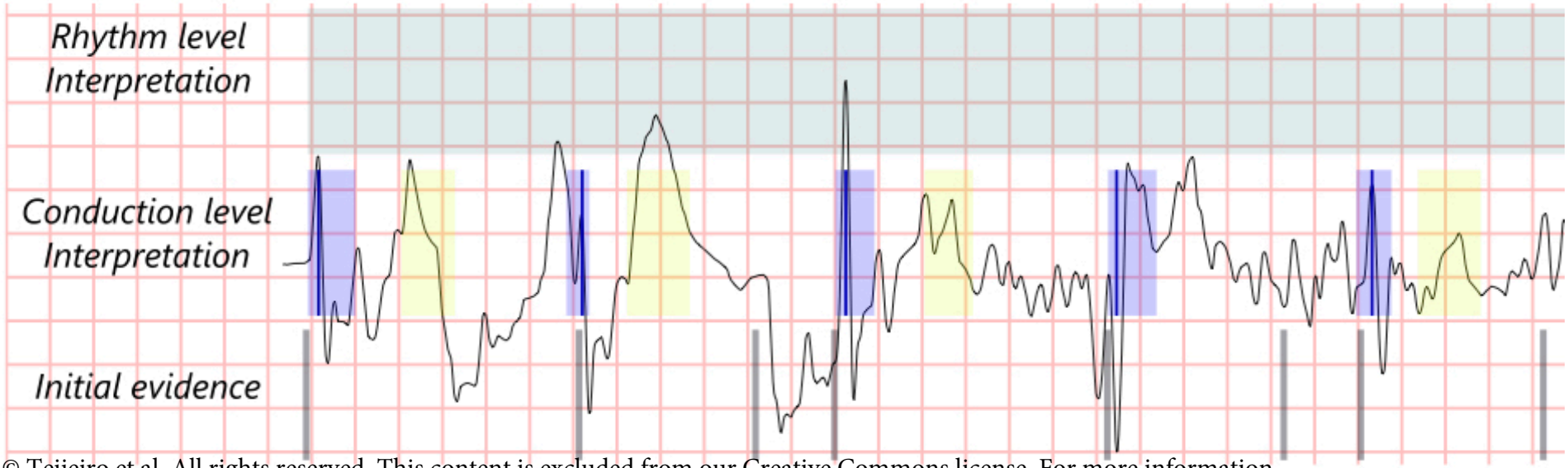


2. Common structure of the QRS detectors.

[Kohler, Hennig, Orglmeister. The Principles of Software QRS Detection, IEEE Engineering in Medicine & Biology, 2002]

Winning approach

- Training data in 2017 Physionet challenge: ~8500 ECGs
- Best algorithms use a combination of expert-derived features and machine learning



[Teijeiro, Garcia, Castro, Felix. arXiv:1802.05998, 2018]

Not enough data for deep learning? Wrong architectures?

"However, the fact that a standard random forest with well chosen features performed as well as more complex approaches, indicates that perhaps a set of 8,528 training patterns was not enough to give the more complex approaches an advantage. With so many parameters and hyperparameters to tune, the search space can be enormous and significant overtraining was seen…"

[Clifford et al. AF Classification from a Short Single Lead ECG Recording: the PhysioNet/Computing in Cardiology Challenge, Computing in Cardiology 2017]

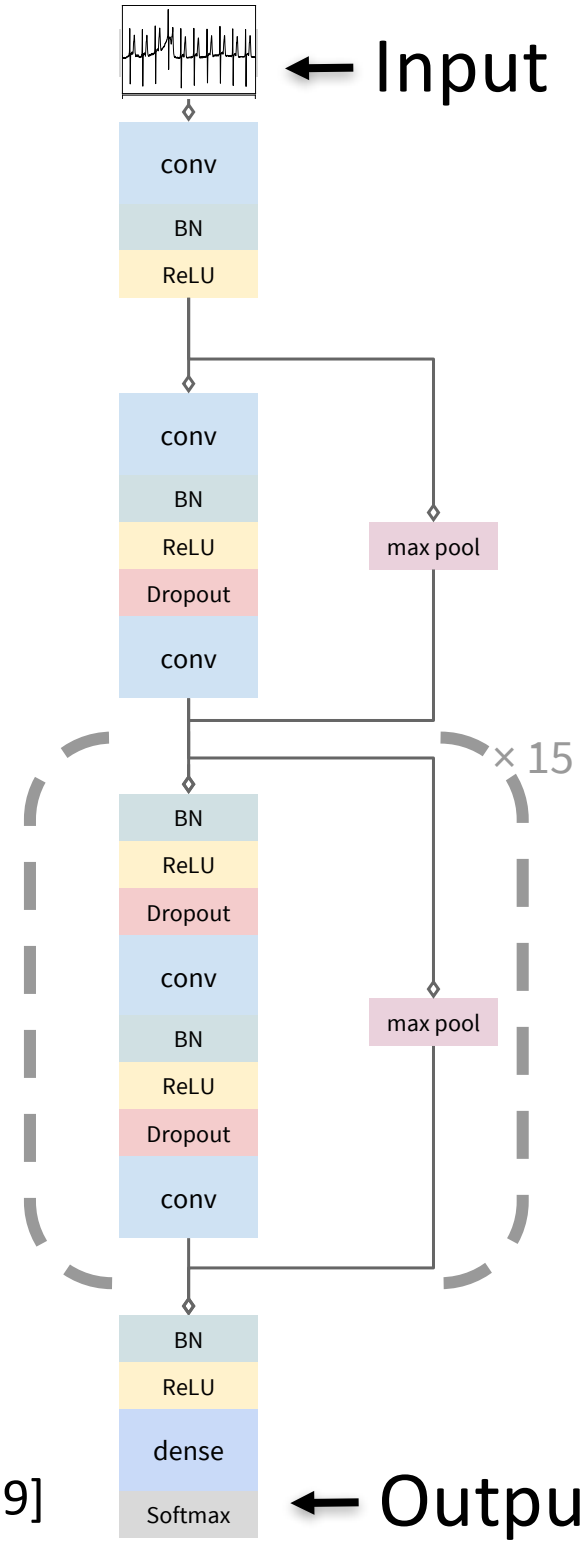
Differences with previous work

- Sensor is a Zio patch – conceivably much less noisy:

- ~90K ECG records annotated (from ~50K patients)
- Identify 12 heart arrhythmias, sinus rhythm and noise for a total of 14 output classes

Deep convolutional network

- 1-D signal sampled at 200Hz, labeled at 1 sec intervals
- 34 layers
- Shortcut connections (ala residual networks) with max-pooling
- Subsampled every other layer (2^8 in total)



[Rajpurkar et al., arXiv:1707.01836, 2017; Nature Medicine '19]

Summary

- We are nearly always in realm of “not enough data”
- Modeling and incorporating prior knowledge is critical to good performance
- Design principles
  - Model the distribution of physiological dynamics
  - Derive features using existing clinical knowledge
  - Start from the simplest possible model
  - Share statistical strength across tasks