A Location-Mixture Autoregressive Model for Online Forecasting of Lung Tumor Motion

Dan Cervone, Natesh Pillai, Debdeep Pati, Ross Berbeco, John H. Lewis

Harvard Statistics Department and Harvard Medical School

August 6, 2014

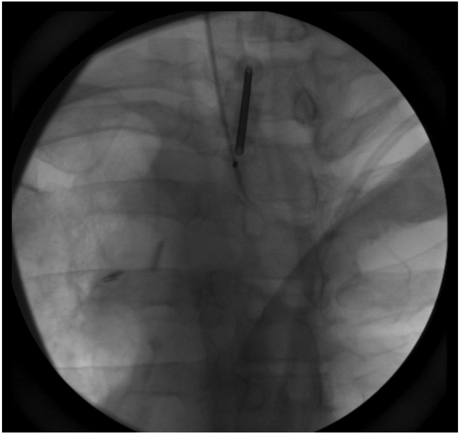
Introduction

External beam radiotherapy:

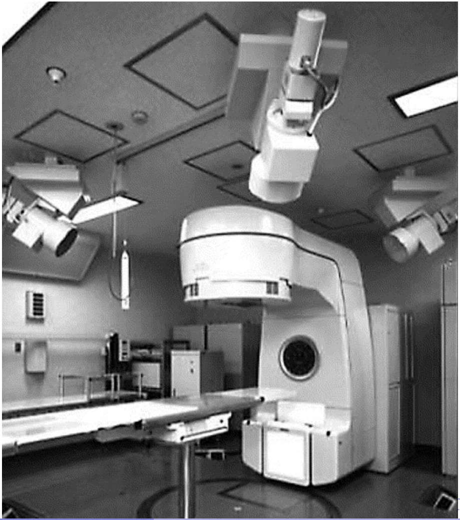
- Lung tumor patients are given an implant (fiducial) at the location of their tumor.
- X-ray tomography can reveal location of the fiducial, thus the tumor.
- Radiotherapy is applied to the tumor location in a narrow beam, minimizing exposure within healthy tissue.

Introduction

Fiducial



External beam radiotherapy



Introduction

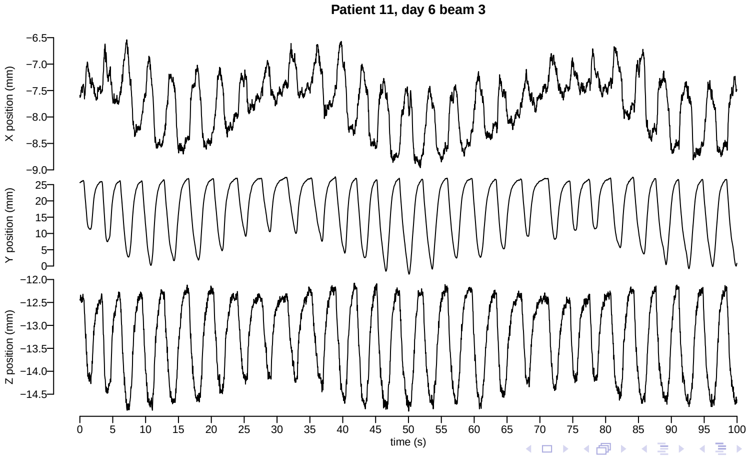
Our problem:

- Patient's respiration means the tumor is in constant motion.
- Tracking the fidual lags 0.1-2s behind, depending on specific machinery used.
- **We need to forecast the location of the tumor to overcome this latency and ensure concentrated, accurate radiothrapy.**

The data

We consider data from 11 patients:

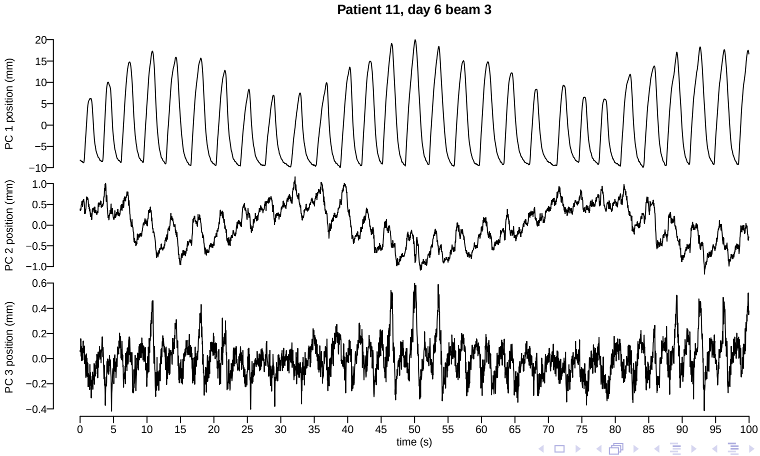
- multiple days receiving treatment.
- multiple radiotherapy sessions ("beams") per treatment.



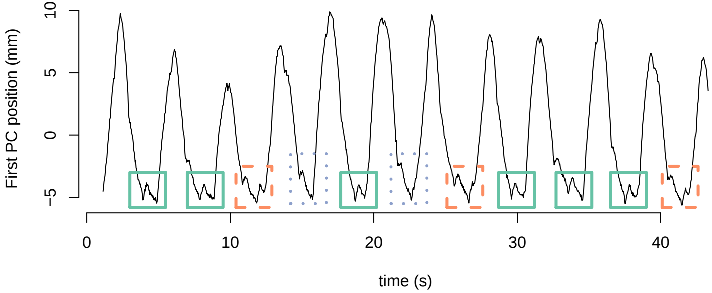
The data

We consider data from 11 patients:

- multiple days receiving treatment.
- multiple radiotherapy sessions ("beams") per treatment.



Semiperiodic features



- Signals at different frequencies.
- Fluctuations in location/amplitude/periodicity.
- Repeated motifs.

Location-mixture autoregressive process

Assume Y_i, $i = 0, \ldots, n$ is a time series in $\mathbb{R}$, and for some p,

$$Y_n | Y_{n-1}, Y_{n-2}, \ldots \sim \sum_{j=1}^{d_n} \alpha_{n,j} N(\mu_{n,j}, \sigma^2)$$

$$\mu_{n,j} = \tilde{\mu}_{n,j} + \sum_{\ell=1}^{p} \gamma_\ell Y_{n-\ell}$$

Location-mixture autoregressive process

Assume Y_i, $i = 0, \ldots, n$ is a time series in $\mathbb{R}$, and for some p,

$$Y_n | Y_{n-1}, Y_{n-2}, \ldots \sim \sum_{j=1}^{d_n} \alpha_{n,j} N(\mu_{n,j}, \sigma^2)$$

$$\mu_{n,j} = \tilde{\mu}_{n,j} + \sum_{\ell=1}^{p} \gamma_\ell Y_{n-\ell}$$

- $\alpha_{n,j}$ are mixture weights: $\sum_{j=1}^{d_n} \alpha_{n,j} = 1$.
- Mixture means have autoregressive component $\sum_{\ell=1}^{p} \gamma_\ell Y_{n-\ell}$.
- Mixture means have location shift component $\tilde{\mu}_{n,j}$.

Location-mixture autoregressive process

Let $\mathbf{Y}_{i-1:p}$ be the subseries $(Y_{i-1} \ \dots \ Y_{i-p})'$, and define mixture components:

$$\alpha_{i,j} = \frac{\exp\left[-\frac{1}{2}(\mathbf{Y}_{i-1:p} - \mathbf{Y}_{i-j-1:p})'\Sigma^{-1}(\mathbf{Y}_{i-1:p} - \mathbf{Y}_{i-j-1:p})\right]}{\sum_{\ell \leq i-p} \exp\left[-\frac{1}{2}(\mathbf{Y}_{i-1:p} - \mathbf{Y}_{i-\ell-1:p})'\Sigma^{-1}(\mathbf{Y}_{i-1:p} - \mathbf{Y}_{i-\ell-1:p})\right]}$$

$$\mu_{i,j} = Y_{i-j} - \gamma'\mathbf{Y}_{i-j-1:p} + \gamma'\mathbf{Y}_{i-1:p}$$

Or, using latent variables,

$$Y_i | M_i = j, Y_{i-1}, \dots, Y_0 \sim N(Y_{i-j} + \gamma'(\mathbf{Y}_{i-1:p} - \mathbf{Y}_{i-j-1:p}), \sigma^2)$$

$$P(M_i = \ell | Y_{i-1}, \dots, Y_0) \propto \exp\left[-\frac{1}{2}(\mathbf{Y}_{i-1:p} - \mathbf{Y}_{i-\ell-1:p})'\Sigma^{-1}(\mathbf{Y}_{i-1:p} - \mathbf{Y}_{i-\ell-1:p})\right]$$

Location-mixture autoregressive process

Latent variables M_i instantiate time series motifs: let

$$\Omega = \begin{pmatrix} \tau^2 & \gamma' \\ \gamma & \Sigma \end{pmatrix},$$

Location-mixture autoregressive process

Latent variables M_i instantiate time series motifs: let

$$\Omega = \begin{pmatrix} \tau^2 & \gamma' \\ \gamma & \Sigma \end{pmatrix},$$

then $M_i = j$ implies

- $(\mathbf{Y}_{i-0:p} - \mathbf{Y}_{i-j-0:p})'\Omega^{-1}(\mathbf{Y}_{i-0:p} - \mathbf{Y}_{i-j-0:p}) =$ small.

Location-mixture autoregressive process

Latent variables M_i instantiate time series motifs: let

$$\Omega = \begin{pmatrix} \tau^2 & \gamma' \\ \gamma & \Sigma \end{pmatrix},$$

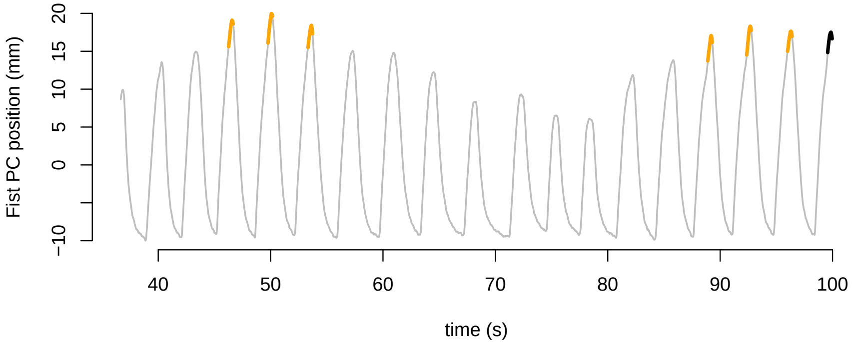
then $M_i = j$ implies

- $(\mathbf{Y}_{i-0:p} - \mathbf{Y}_{i-j-0:p})'\Omega^{-1}(\mathbf{Y}_{i-0:p} - \mathbf{Y}_{i-j-0:p}) =$ small.
- Assuming

$$\mathbf{Y}_{i-0:p} | M_i = j, \mathbf{Y}_{i-j-0:p} \sim N(\mathbf{Y}_{i-j-0:p}, \Omega)$$

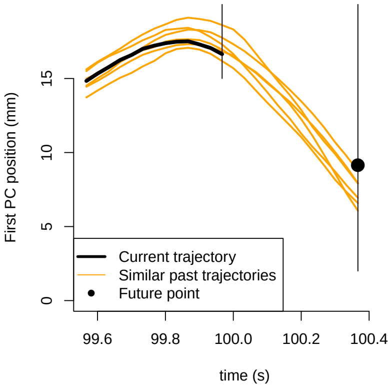
yields the LMAR predictive distribution for $Y_i | M_i = j, Y_{i-1}, \ldots$

Illustration of time series motifs



- $\mathbf{Y}_{i-1:p}$ most recent p values of times series.
- $\mathbf{Y}_{i-j-1:p}$ previous instances of this motif.

Illustration of time series motifs



- $\mathbf{Y}_{i-1:p}$ most recent p values of times series.
- $\mathbf{Y}_{i-j-1:p}$ previous instances of this motif.

Parameter estimation

We focus on estimating Ω:

- 1-1 correspondence between Ω and LMAR parameters γ, Σ, σ^2.

Latent variables M_i yield estimating equation:

$$h(\Omega) = \sum_{i \geq p} -\frac{1}{2}\log(|\Omega|) - \frac{1}{2}\sum_{j \leq i-p} \mathbf{1}[M_i = j](\mathbf{Y}_{i-0:p} - \mathbf{Y}_{i-j-0:p})'\Omega^{-1}(\mathbf{Y}_{i-0:p} - \mathbf{Y}_{i-j-0:p})$$

- $\hat{\Omega} = \text{argmax}(h(\Omega))$ computable quickly using EM algorithm.
- Not loglikelihood, but $\hat{\Omega}$ enjoys same large-sample properties as MLE.

k step ahead predictive distributions

Given a point estimate $\hat{\Omega}$, k-step ahead predictive distributions can be *approximated* by marginalizing over future steps 1 to $(k-1)$:

$$Y_{i+k} | Y_i, Y_{i-1}, \ldots \sim \sum_j \alpha_j^k N(\mu_j^k, \sigma_k^2).$$

k step ahead predictive distributions

Given a point estimate $\hat{\Omega}$, k-step ahead predictive distributions can be *approximated* by marginalizing over future steps 1 to $(k-1)$:

$$Y_{i+k}|Y_i, Y_{i-1}, \ldots \sim \sum_j \alpha_j^k N(\mu_j^k, \sigma_k^2).$$

- Mixture components $\alpha_j^k, \mu_j^k, \sigma_k^2$ available cheaply, analytically.
- No need for sampling or Monte Carlo.

Predictive performance on clinical data

We tested predictive performance using the following procedure for all clinical observations in our data set.

1. Train model on first 35s of data
2. Fit prediction model in under 5 seconds
3. Simulate predictions for the next 40 seconds of data using model fit
4. Compare predictions with observed values

Predictive performance on clinical data

We tested predictive performance using the following procedure for all clinical observations in our data set.

1. Train model on first 35s of data
2. Fit prediction model in under 5 seconds
3. Simulate predictions for the next 40 seconds of data using model fit
4. Compare predictions with observed values

- Predictions must be made in real time (30Hz)!

Predictive performance on clinical data

We tested predictive performance using the following procedure for all clinical observations in our data set.

1. Train model on first 35s of data
2. Fit prediction model in under 5 seconds
3. Simulate predictions for the next 40 seconds of data using model fit
4. Compare predictions with observed values

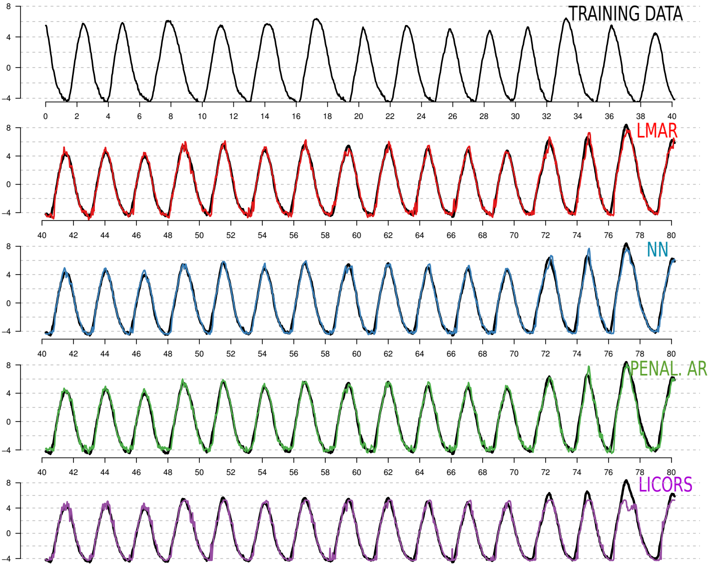
- Predictions must be made in real time (30Hz)!

Alternative prediction models considered:

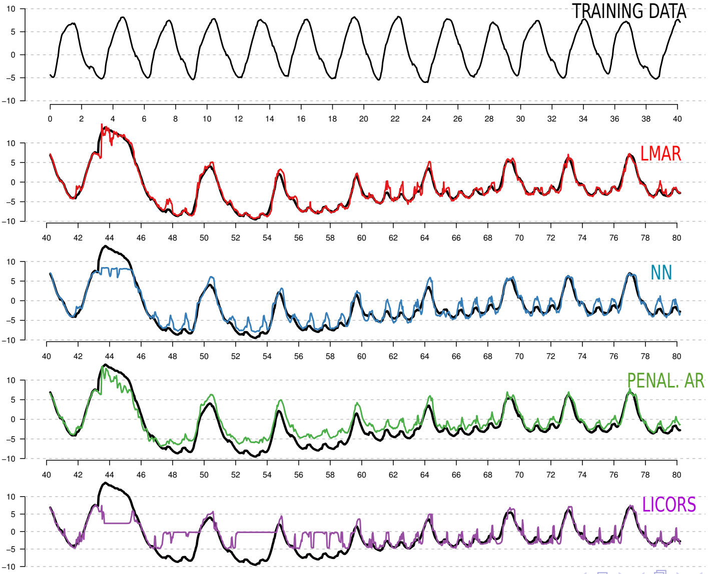
- Our model: LMAR
- Neural networks
- Penalized AR model
- LICORS (Georg and Shalizi, 2012)

For all methods, any tuning parameters were set to patient-independent optimal values, using separate data.

Example 1 ($k = 6$)



Example 2 ($k = 6$)


TRAINING DATA
LMAR
NN
PENAL. AR
LICORS
10
5
0
−5
−10
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40
40 42 44 46 48 50 52 54 56 58 60 62 64 66 68 70 72 74 76 78 80

Quantitative summaries (point prediction)

For all methods, predictive performance varies by patient, but some examples:

		0.2s forecast ($k = 6$)		0.4s forecast ($k = 12$)		0.6s forecast ($k = 18$)	
Pat.	Method	RMSE	MAE	RMSE	MAE	RMSE	MAE
9	LMAR	**0.58**	**0.22**	**1.29**	**0.52**	**2.03**	**0.90**
	NNs	0.73	0.32	1.69	0.64	2.45	0.92
	Ridge	0.81	0.34	1.68	0.73	2.42	0.98
	LICORS	1.35	0.53	2.20	0.98	2.64	1.19
10	LMAR	**0.88**	**0.36**	**1.73**	**0.77**	**2.55**	**1.19**
	NNs	1.09	0.44	2.16	0.93	2.98	1.35
	Ridge	0.95	0.45	1.84	0.94	2.67	1.41
	LICORS	1.62	0.61	2.20	1.10	3.25	1.56
11	LMAR	**1.13**	**0.44**	**2.59**	**1.06**	**3.70**	**1.49**
	NNs	1.24	0.50	2.95	1.19	3.99	1.70
	Ridge	1.19	0.63	2.69	1.51	3.99	2.40
	LICORS	1.64	0.57	3.04	1.09	4.21	1.65

Quantitative summaries (interval prediction)

Summaries of interval/distributional predictions:

Patient	Method	0.2s Forecast ($k = 6$) Coverage	0.2s Forecast ($k = 6$) Log PS	0.4s Forecast ($k = 12$) Coverage	0.4s Forecast ($k = 12$) Log PS	0.6s Forecast ($k = 18$) Coverage	0.6s Forecast ($k = 18$) Log PS
9	LMAR	**0.89**	**0.87**	**0.90**	**1.65**	**0.92**	**2.07**
	NNs	0.86	1.02	0.78	2.20	0.80	2.77
	Ridge	0.81	1.54	0.81	2.21	0.81	2.64
	LICORS	0.86	1.62	0.81	1.98	0.79	2.31
10	LMAR	**0.86**	**1.18**	**0.88**	**1.94**	**0.91**	2.33
	NNs	0.84	1.23	0.76	2.25	0.79	2.65
	Ridge	0.83	1.35	0.84	2.03	0.84	2.44
	LICORS	**0.86**	1.61	0.82	2.02	0.81	**2.31**
11	LMAR	0.85	**1.38**	**0.87**	2.13	**0.91**	2.36
	NNs	0.87	1.50	0.80	2.70	0.83	2.91
	Ridge	0.86	1.63	0.85	2.44	0.85	2.84
	LICORS	**0.88**	1.56	0.83	**1.99**	0.82	**2.25**

Conclusions

Future improvements:

- Hierarchical models; information sharing between patients.
- Code optimizations (current pipeline in R/C++)

Further reading:

- D. Cervone, N. Pillai, D. Pati, R. Berbeco, J. H. Lewis, "A Location-Mixture Autoregressive Model for Online Forecasting of Lung Tumor Motion". *Annals of Applied Statistics* (to appear). arXiv:1309.4144.