02457 Non-linear signal processing

2017 - Lecture 12

Technical University of Denmark
DTU Compute, Kgs Lyngby, Denmark

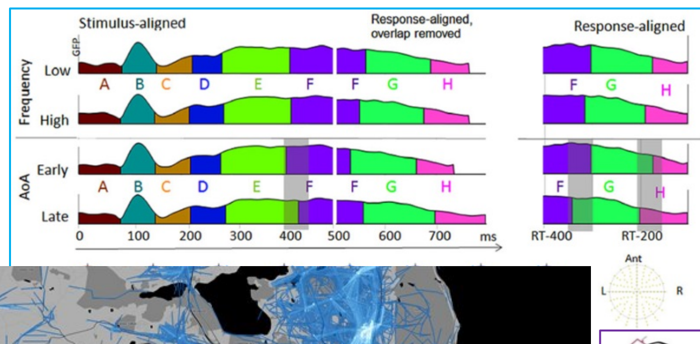
Outline lecture 12

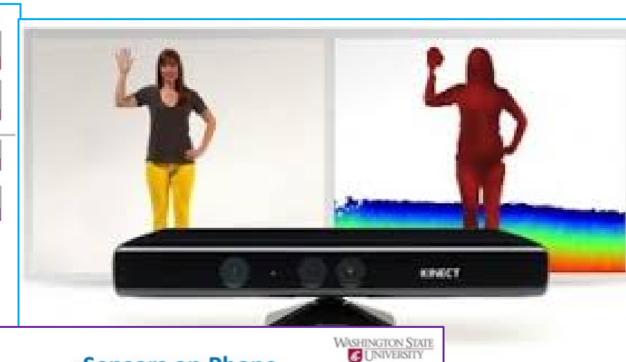
- Real time systems - search engines
  - EEG: Smartphone brain scanner
  - Audio, Speech production, symbolic representations
  - Music recommendation
  - Castsearch, spoken documents
- Markov models
  - Non-stationary Markov models
- Bayesian linear models
- Dynamic linear models
  - Forward recursion, dynamic updates of the posterior

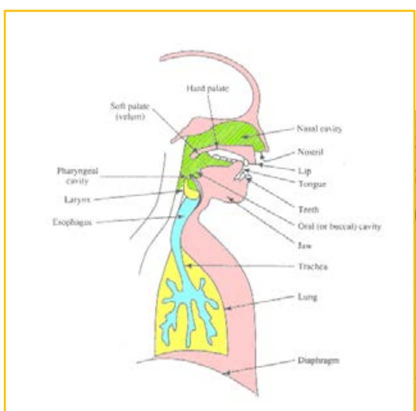
Real time machine learning

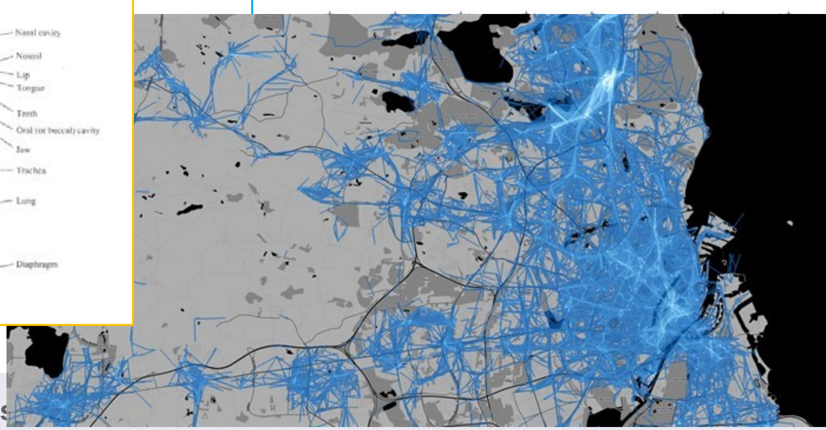
Many real life applications of machine learning are **real time**

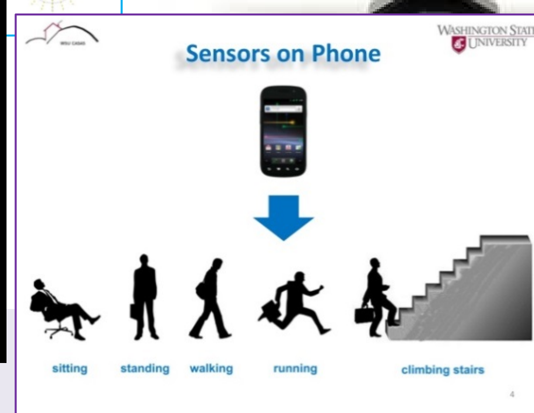
EEG: Neurofeedback, decoding, brain-computer interface...
Speech: Recognition, speaker, diagnostics,...
Text: search, sentiment, argument,...
Activity recognition: wellness, rehabilitation, social modeling,...











Smartphone brain scanner

A. Stopczynski, C. Stahlhut, M.K. Petersen, J.E. Larsen, C.F. Jensen, M.G. Ivanova, T.S. Andersen, L.K. Hansen. *Smartphones as pocketable labs: Visions for mobile brain imaging and neurofeed-back.* International Journal of Psychophysiology, (2014).
A. Stopczynski, C. Stahlhut, J.E. Larsen, M.K. Petersen, L.K. Hansen.
The Smartphone Brain Scanner: A Portable Real-Time Neuroimaging System. PloS one 9 (2), e86733, (2014)

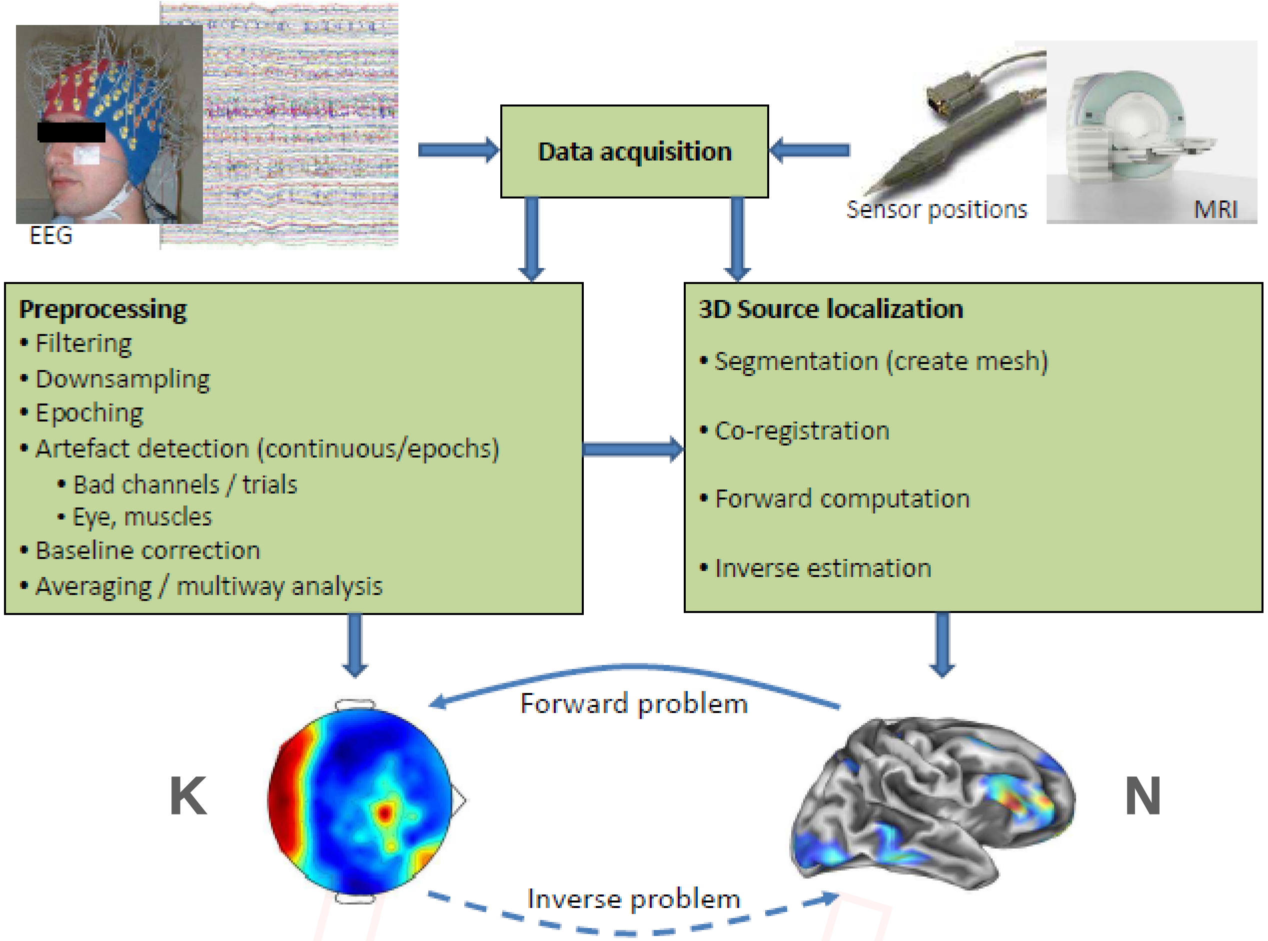
EEG imaging

Linear ill-posed inverse problem

X: N x T
Y: K x T
A: K x N

N >> K

Need priors to solve!



$$Y_{k,t} = \sum_{n=1}^{N} A_{k,n} X_{n,t} + E_{k,t}.$$

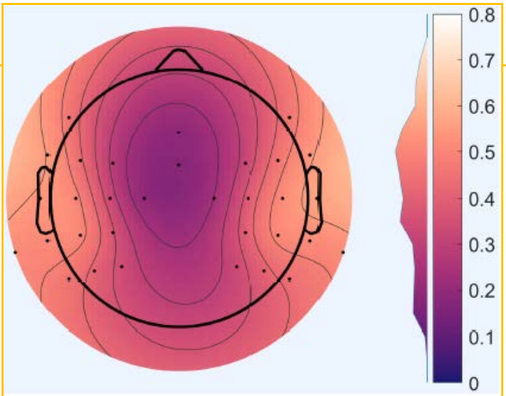
Neurotech for 24/7 monitoring: EarEEG / Widex

A discreete, <u>non-invasive</u> solution for long time recording in the wild

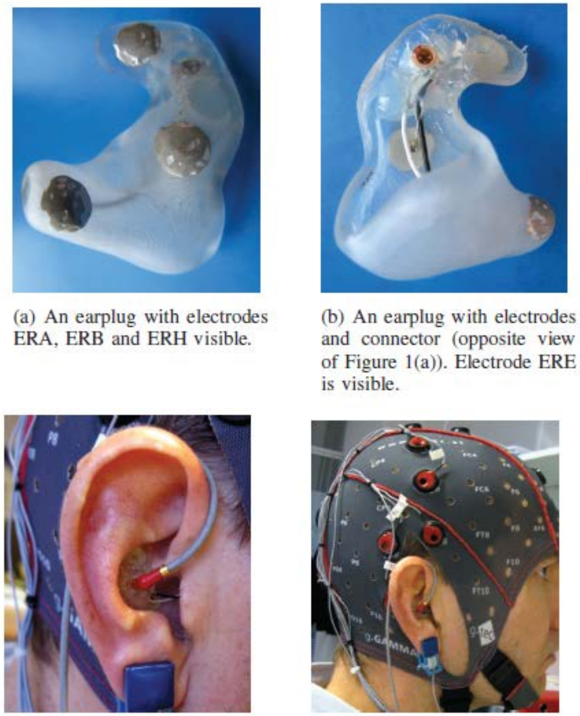
EarEEG is a well-established technology

Classical EEG reproduced: Sustained and event related responses to audio and visual stimulus

High mutual information between ear and scalp EEG



(c) Coupled ears, with common reference.



(a) An earplug with electrodes ERA, ERB and ERH visible.

(b) An earplug with electrodes and connector (opposite view of Figure 1(a)). Electrode ERE is visible.

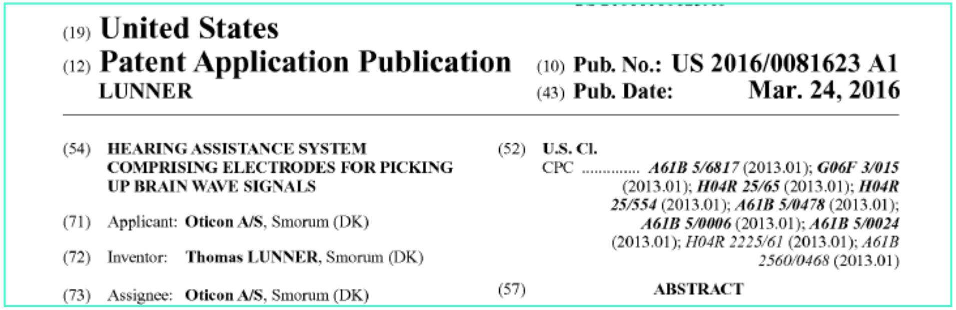
(c) Right ear with earplug.

(d) Side view of test subject showing the recording setup.

Fig. 1. View of a right ear earplug and the Ear-EEG recording setup.

Kidmose, Preben, et al. "Ear-EEG from generic earpieces: A feasibility study." *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*. IEEE, 2013.

Oticon mobile EEG device


(19) United States
(12) Patent Application Publication
LUNNER

(10) Pub. No.: US 2016/0081623 A1
(43) Pub. Date: Mar. 24, 2016

(54) HEARING ASSISTANCE SYSTEM COMPRISING ELECTRODES FOR PICKING UP BRAIN WAVE SIGNALS

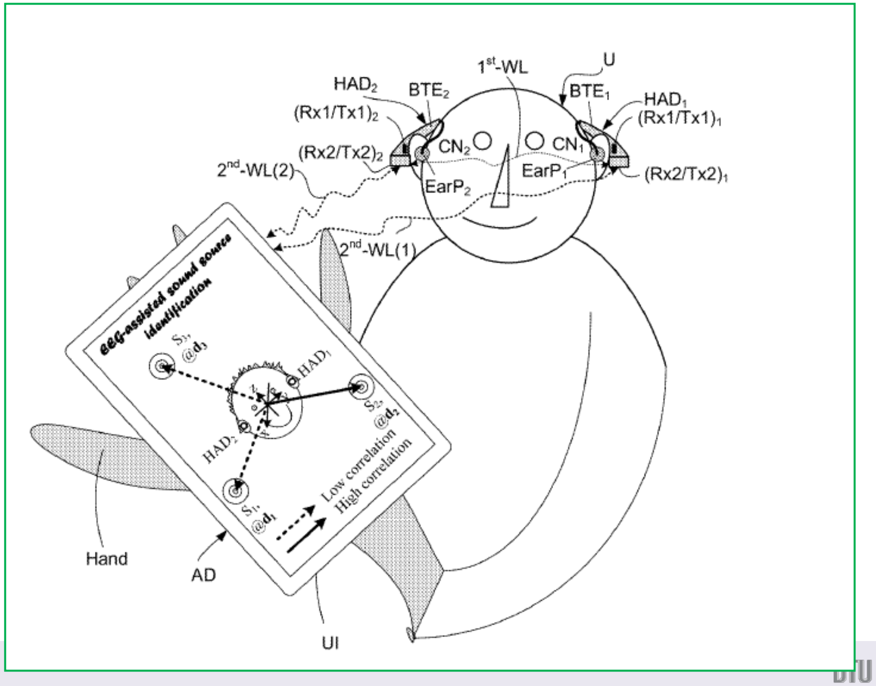
(71) Applicant: **Oticon A/S**, Smorum (DK)

(72) Inventor: **Thomas LUNNER**, Smorum (DK)

(73) Assignee: **Oticon A/S**, Smorum (DK)

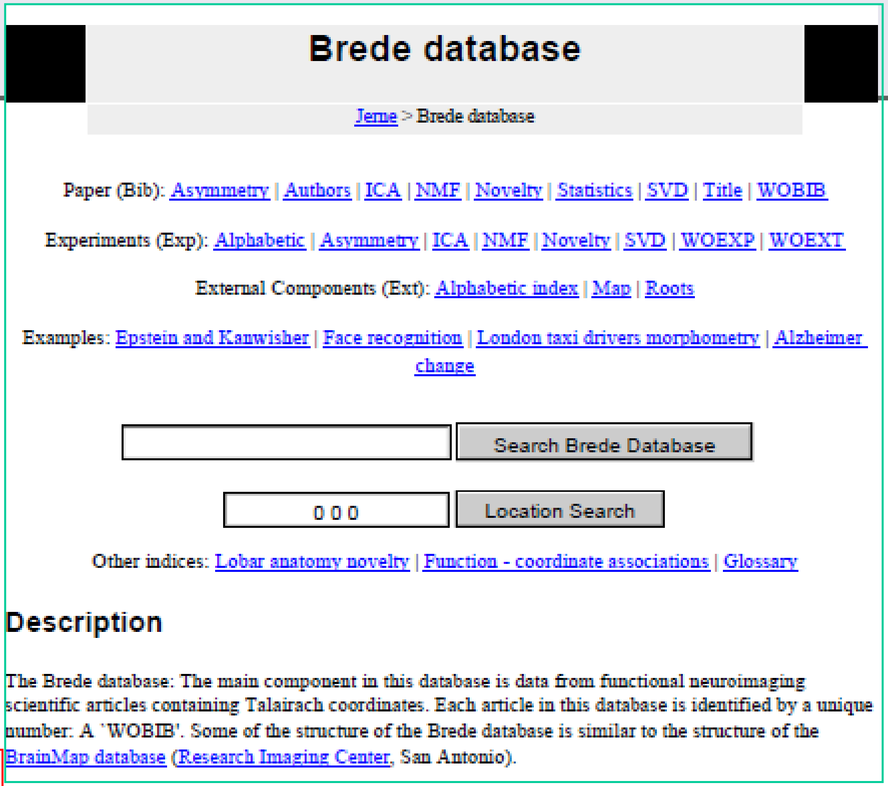
(52) U.S. Cl.
CPC *A61B 5/6817* (2013.01); *G06F 3/015* (2013.01); *H04R 25/65* (2013.01); *H04R 25/554* (2013.01); *A61B 5/0478* (2013.01); *A61B 5/0006* (2013.01); *A61B 5/0024* (2013.01); *H04R 2225/61* (2013.01); *A61B 2560/0468* (2013.01)

(57) ABSTRACT



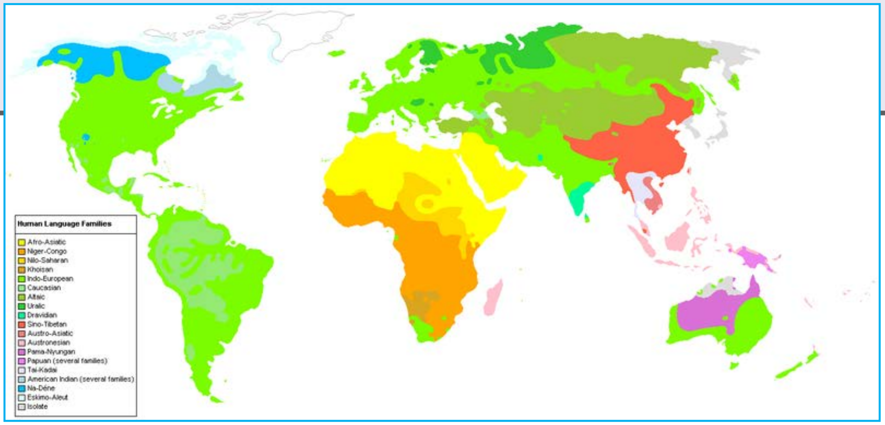
Generic Search engines

Specialized search



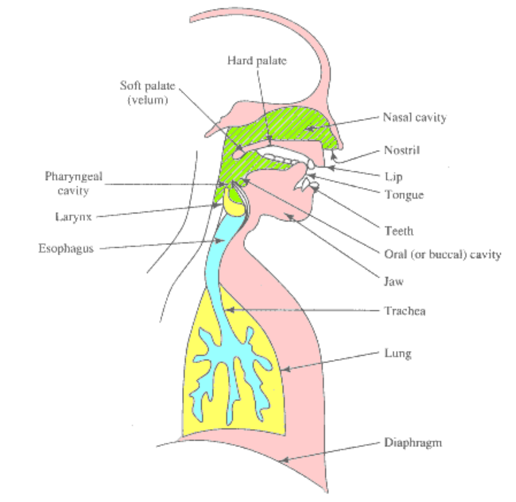
http://hendrix.imm.dtu.dk/services/jerne/brede/brede.html

SPEECH –some definitions



- Speech signals are sequences of sounds
- The basic sounds and the transitions between them serve as symbolic representation of information - **semantics**
- The arrangement of these sounds (symbols) is governed by the rules of ***language***
- The study of these rules and their implications in human communication is called ***linguistics***
- The study of and classification of the sounds of speech is called ***phonetics***

SPEECH PRODUCTION



Speech is produced by the human vocal tract

The vocal tract is excited either by short burst of periodic stimulus or white noise

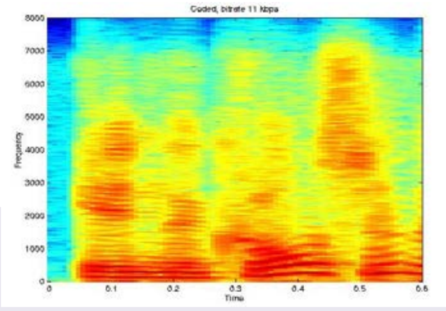
Voiced sounds are produced by an airflow through tight vocal cords.

Unvoiced sounds are produced by turbulent flow.

Sounds are classified broadly into phonemes: Vowels and consonants
More complex structures include diphthongs, semi-vowels.

Formants are peaks in the power spectrum modulation (envelopes).

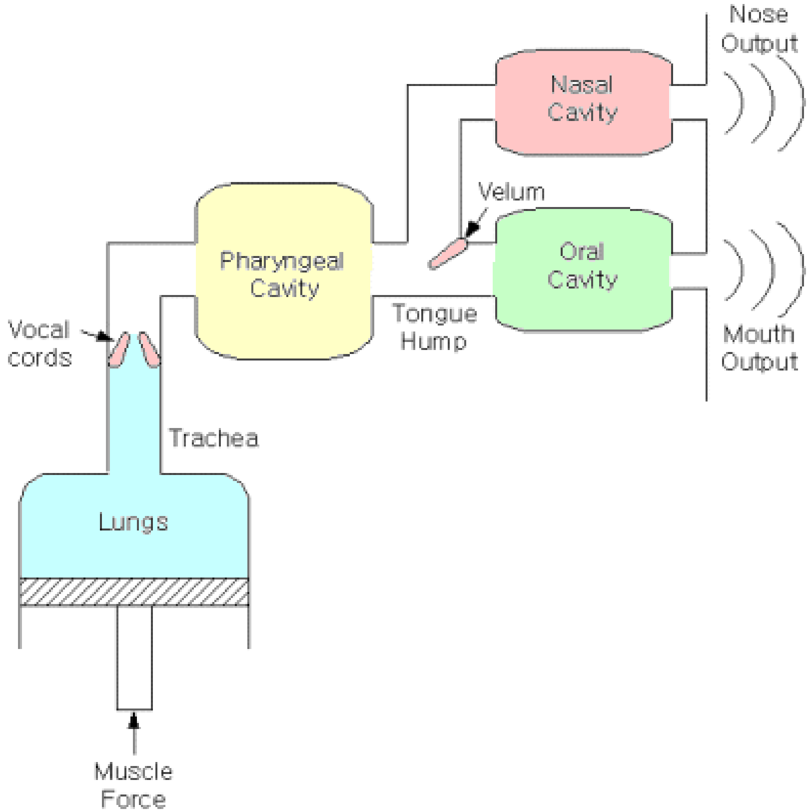
Frequencies in the range:
F1 (270-730 Hz), F2 (840-2290 Hz), and F3 (1690-3010 Hz).



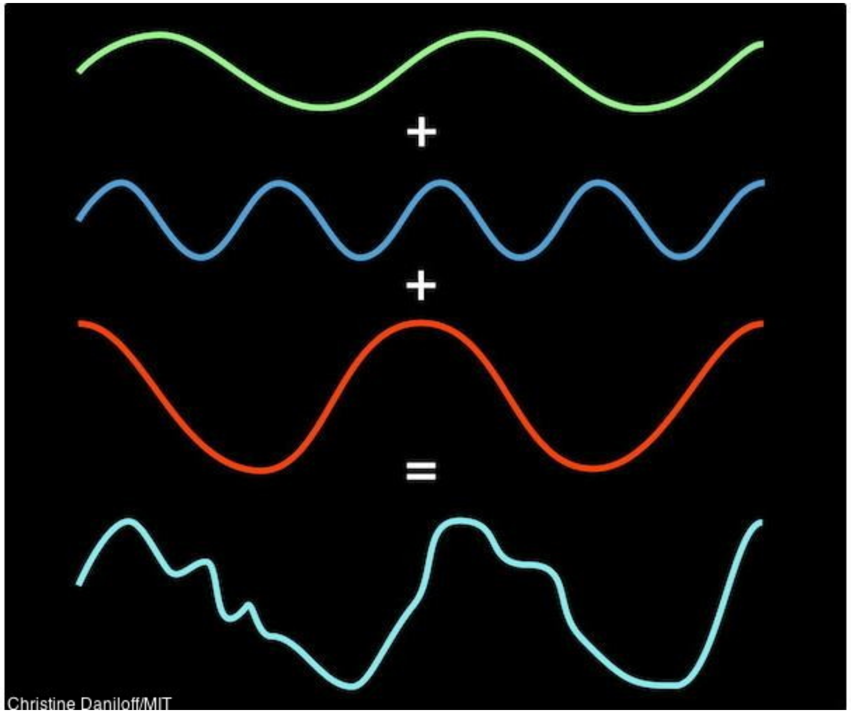
The speech signal is modelled as a linear time variant system excited by a high-frequency signal (periodic og random)

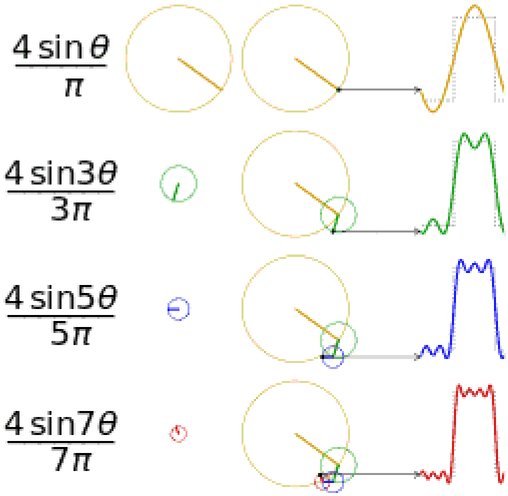
- Linear system model (IIR model)

$$x(n) = \sum_{j=1}^{p} w_j x(n-j) + \epsilon(n)$$



Fourier transform

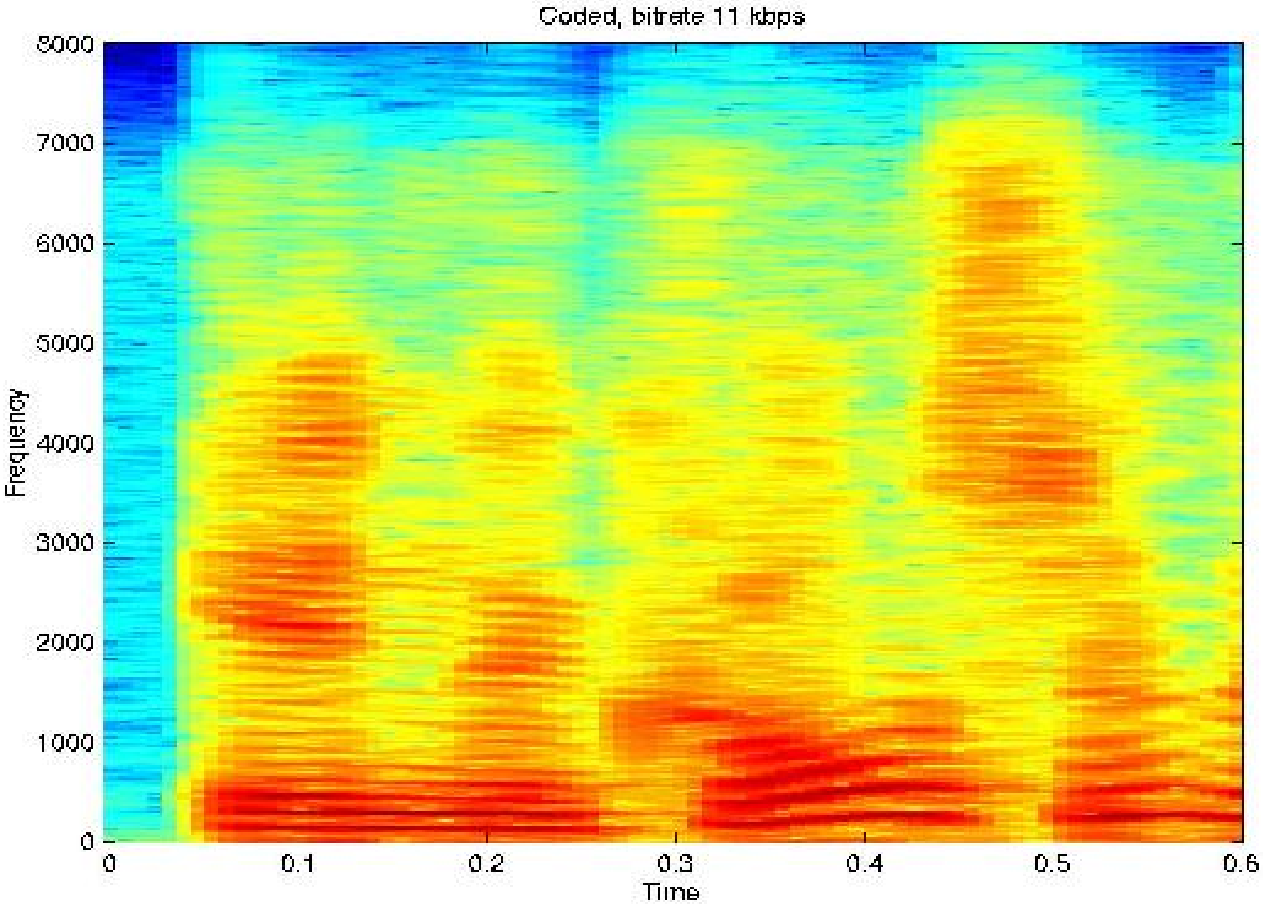




https://en.wikipedia.org/wiki/Fourier_series

http://en.wikipedia.org/wiki/Discrete_Fourier_transform

Speech spectrogram



CEPSTRAL COEFFICIENTS

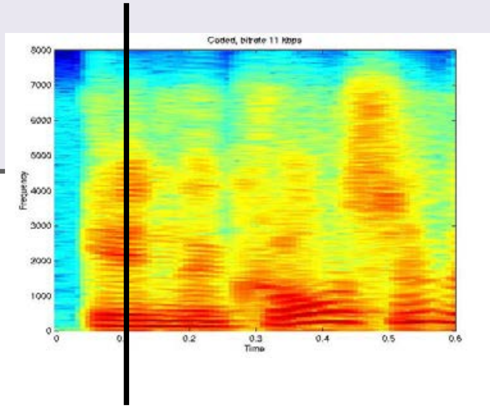
- Linear system model (IIR model)

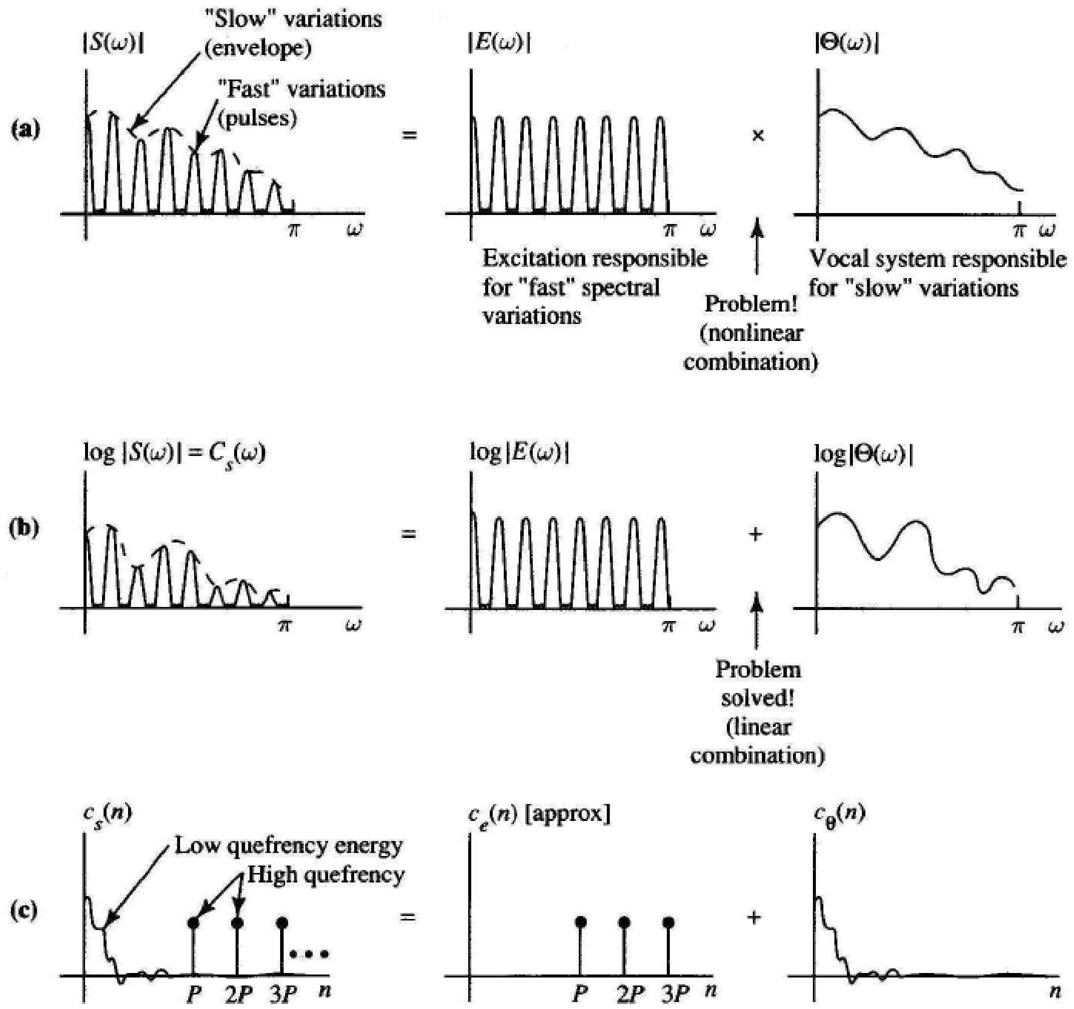
$$
\begin{aligned}
s(n) &= \sum_{j=1}^{p} w(j) \cdot s(n-j) + \epsilon(n) \\
s(n) &= \sum_{i} \theta(i) \cdot \epsilon(n-i)
\end{aligned}
$$

- In the Fourier domain

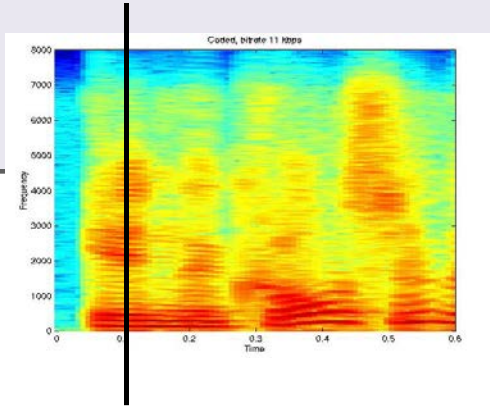
$$
\begin{aligned}
S(\omega) &= \Theta(\omega) \cdot E(\omega) \\
\log|S(\omega)| &= \log|\Theta(\omega)| + \log|E(\omega)|
\end{aligned}
$$

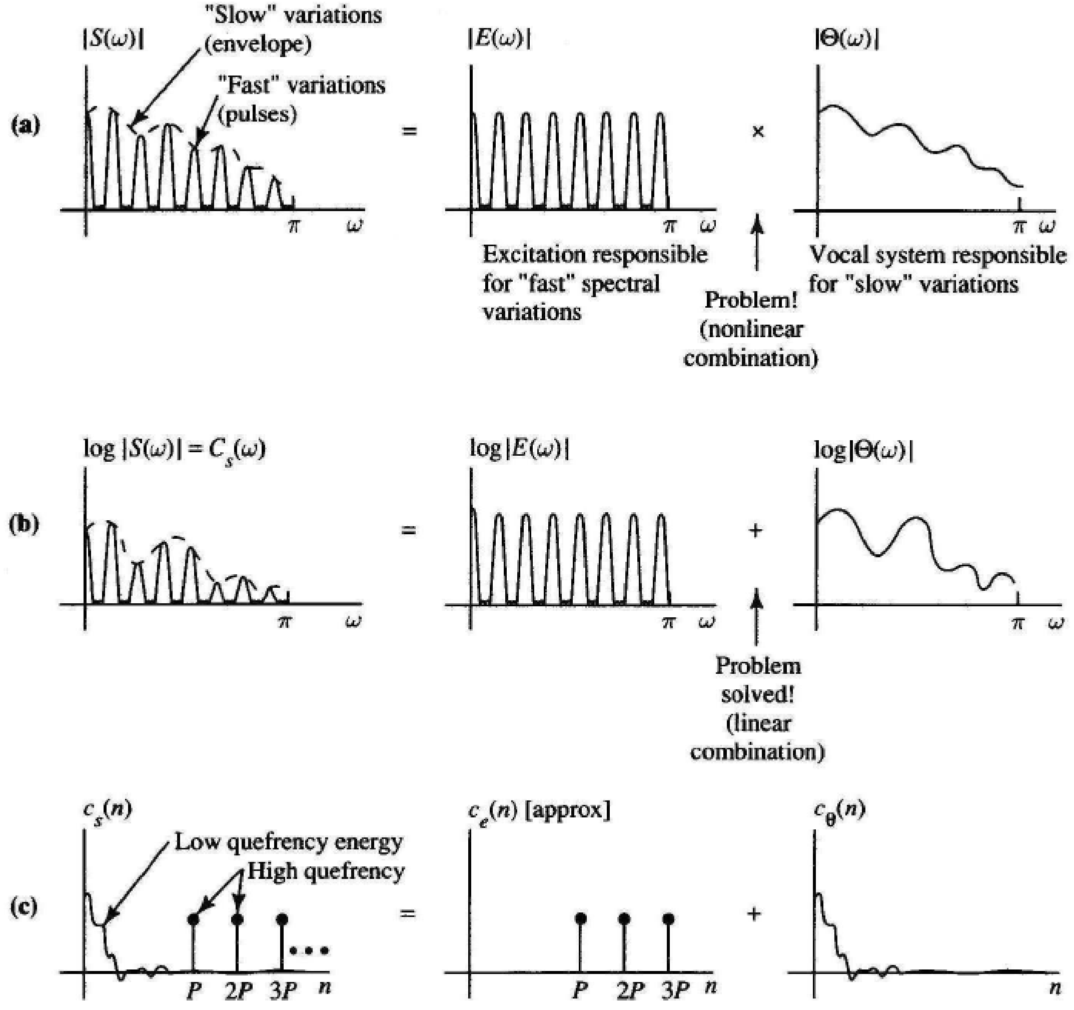
Cepstral liftering





Cepstral liftering



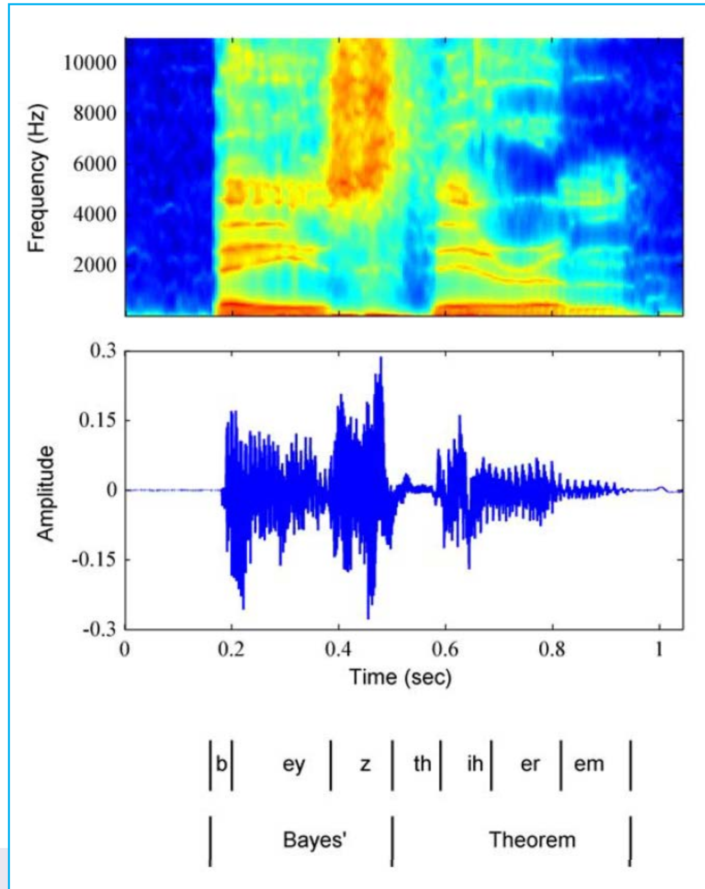


Phoneme grouping and sequence recognition

Provides discrete symbols for identification of word.

The audio stream is segmented into a string of symbols (~phonemes), by assigning windows to most likely phoneme using K-means clustering.

Cepstral analysis → Clustering of features → Recognition of word sequences



Markov process / chain

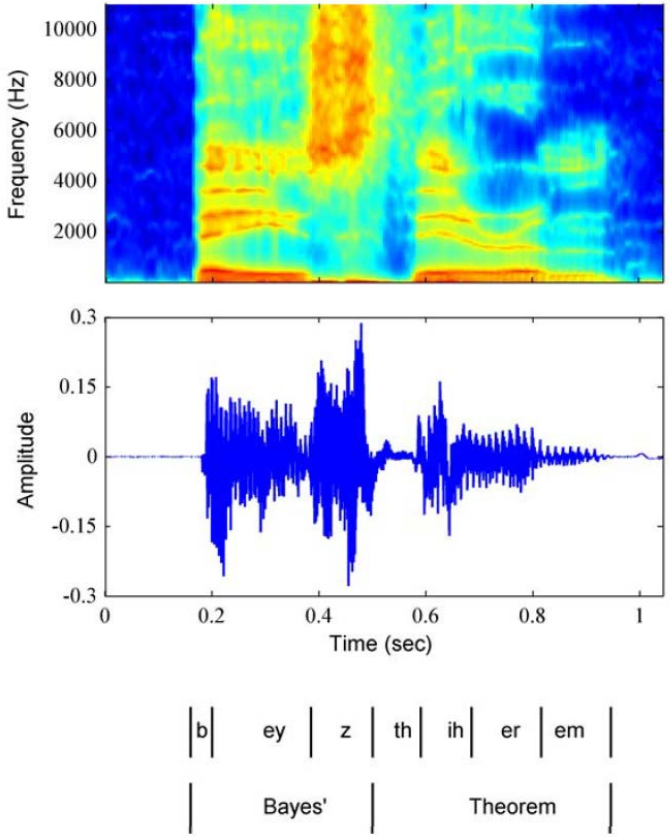
Andrey Markov (1906)

A string of symbols 'y_n' is a Markov chain if

$$p(y_n \mid y_1, y_2, \ldots y_{n-2}\, y_{n-1}) = p(y_n \mid y_{n-L}, \ldots y_{n-2}\, y_{n-1})$$

A string of symbols is 1'st order Markovian if

$$p(y_n \mid y_1, y_2, \ldots y_{n-2}\, y_{n-1}) = p(y_n \mid y_{n-1})$$



SIMPLE MARKOV MODEL

Let y_n be a sequence of symbols with K states

Let $a_{j,j'}$ be the probability of jumping from j to j'

i.e., $a_{j,j'} = p(y_n = j' | y_{n-1} = j)$

Matrix $a_{j,j'}$ is a stochastic matrix $\Sigma_{j'} a_{j,j'} = 1$

a can be estimated by maximum likelihood

Analysis of Markov chains - The ensemble picture

Consider a large number of 'parallel' Markov chains $(y^i)_n$
Each chain makes random moves according to a common a-matrix (stationarity)
The probability over states, at time step n, obeys the Kolmogorov Chapman equation

$$P_{n+1}(j') = \sum_{j=1}^{K} P_n(j) a_{j,j'} \qquad \mathbf{P}_{n+1} = \mathbf{P}_n \mathbf{A}$$

The stationary distribution (if exists) is a (left) eigenvector, with unit eigenvalue

$$P_*(j') = \sum_{j=1}^{K} P_*(j) a_{j,j'}. \qquad \mathbf{P}_* = \mathbf{P}_* \mathbf{A}$$

Detection based on sequences

$$P(\mathrm{B}|\{y_n\}) = \frac{P(\{y_n\}|\mathrm{B})P(\mathrm{B})}{P(\{y_n\})}.$$

$$\boxed{\begin{aligned} P(\{y_n\}|a) &= P(y_1)\prod_{n=2}^{N} P(y_n|y_{n-1},a) \\ &= P(y_1)\prod_{j,j'} (a_{j,j'})^{\mathrm{n_{j,j'}}} \end{aligned}}$$

Markov chain estimation by maximum likelihood

The likelihood can be written, introducing the observed number of transitions j => j' : $n_{j,j'}$

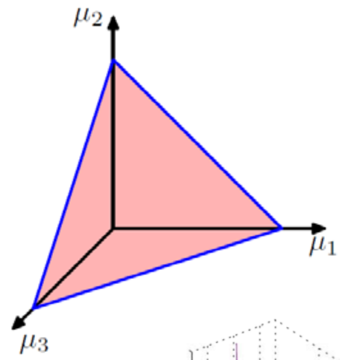
$$
\begin{aligned}
P(\{y_n\}|a) &= P(y_1)\prod_{n=2}^{N} P(y_n|y_{n-1},a) \\
&= P(y_1)\prod_{j,j'} (a_{j,j'})^{\mathrm{n}_{\mathrm{j,j'}}}
\end{aligned}
$$

Estimator

$$
\boxed{\widehat{a}_{j,j'} = \frac{n_{j,j'}}{\sum_{j''} n_{j,j''}}}
$$

MAP estimate with Dirichlet priors on $a_{j,:}$

The Dirichlet distribution over three variables μ_1, μ_2, μ_3 is confined to a simplex (a bounded linear manifold) of the form shown, as a consequence of the constraints $0 \leqslant \mu_k \leqslant 1$ and $\sum_k \mu_k = 1$.



$$\boxed{\mathrm{Dir}(\boldsymbol{\mu}|\boldsymbol{\alpha}) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_K)} \prod_{k=1}^{K} \mu_k^{\alpha_k - 1}}$$

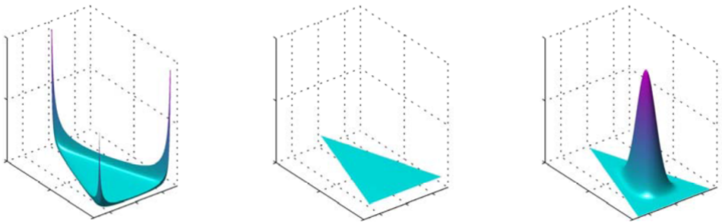


Figure 2.5 Plots of the Dirichlet distribution over three variables, where the two horizontal axes are coordinates in the plane of the simplex and the vertical axis corresponds to the value of the density. Here $\{\alpha_k\} = 0.1$ on the left plot, $\{\alpha_k\} = 1$ in the centre plot, and $\{\alpha_k\} = 10$ in the right plot.

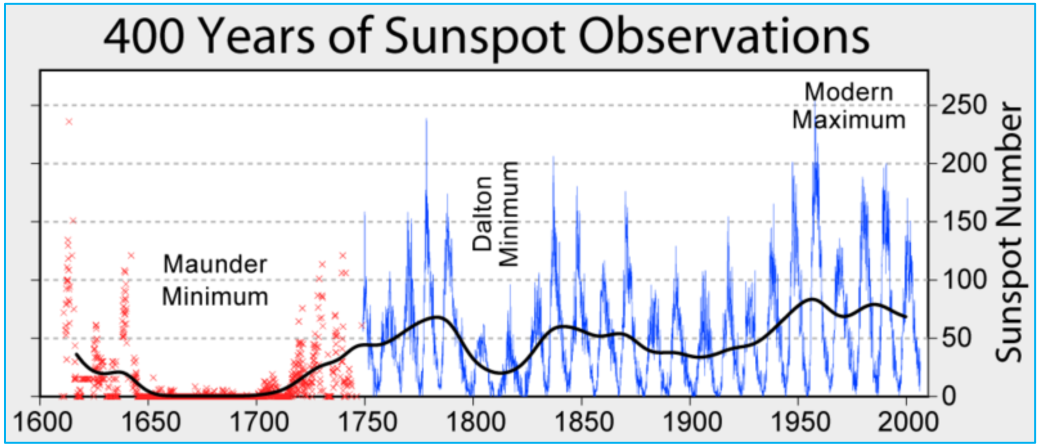
$$\widehat{a}_{j,j'} = \frac{n_{j,j'} + \alpha_{j,j'} - 1}{\sum_{j''} n_{j,j''} + \alpha_{j,j''} - 1}$$

What if data is non-stationary?

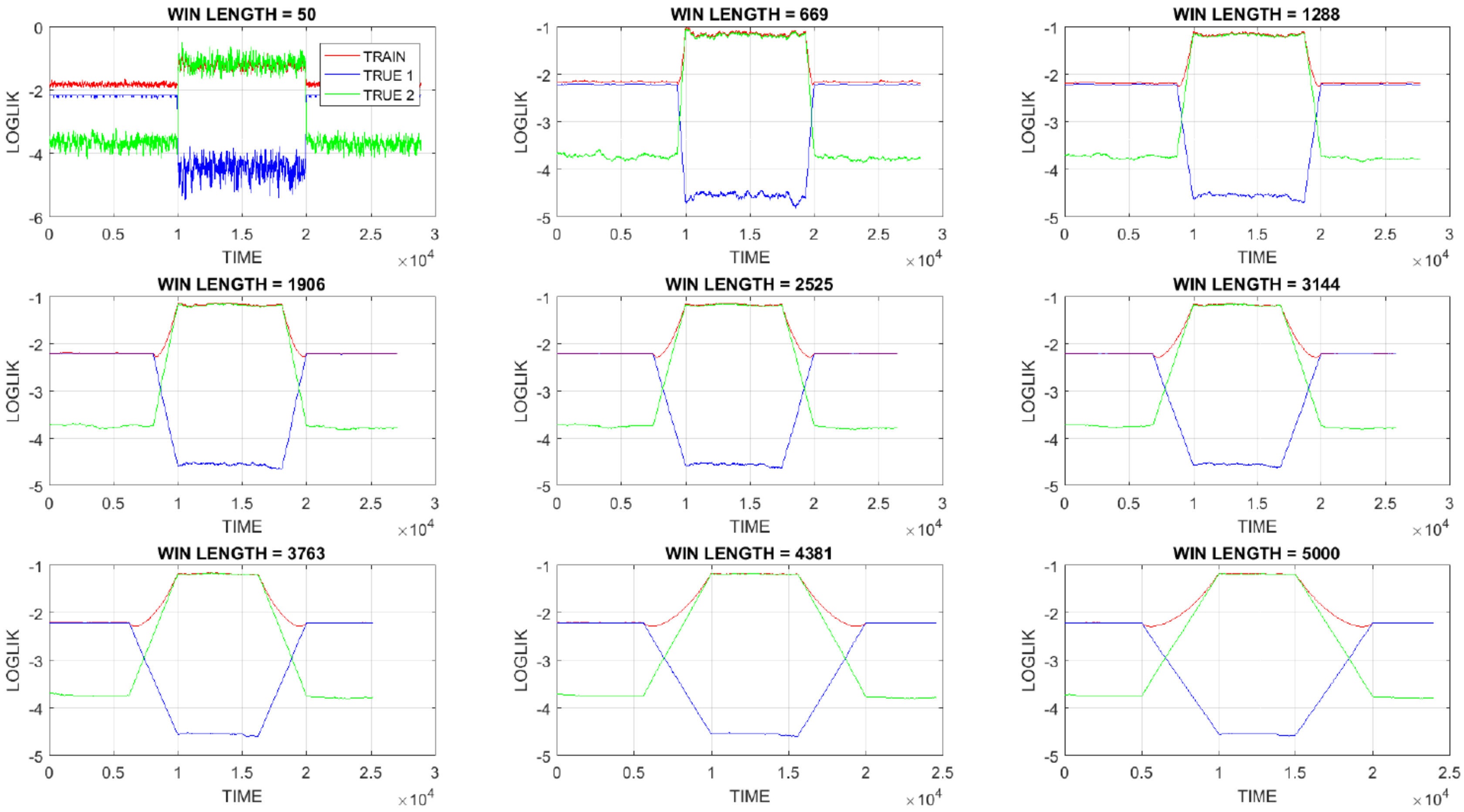
Two issues

How to track the changes in the model?

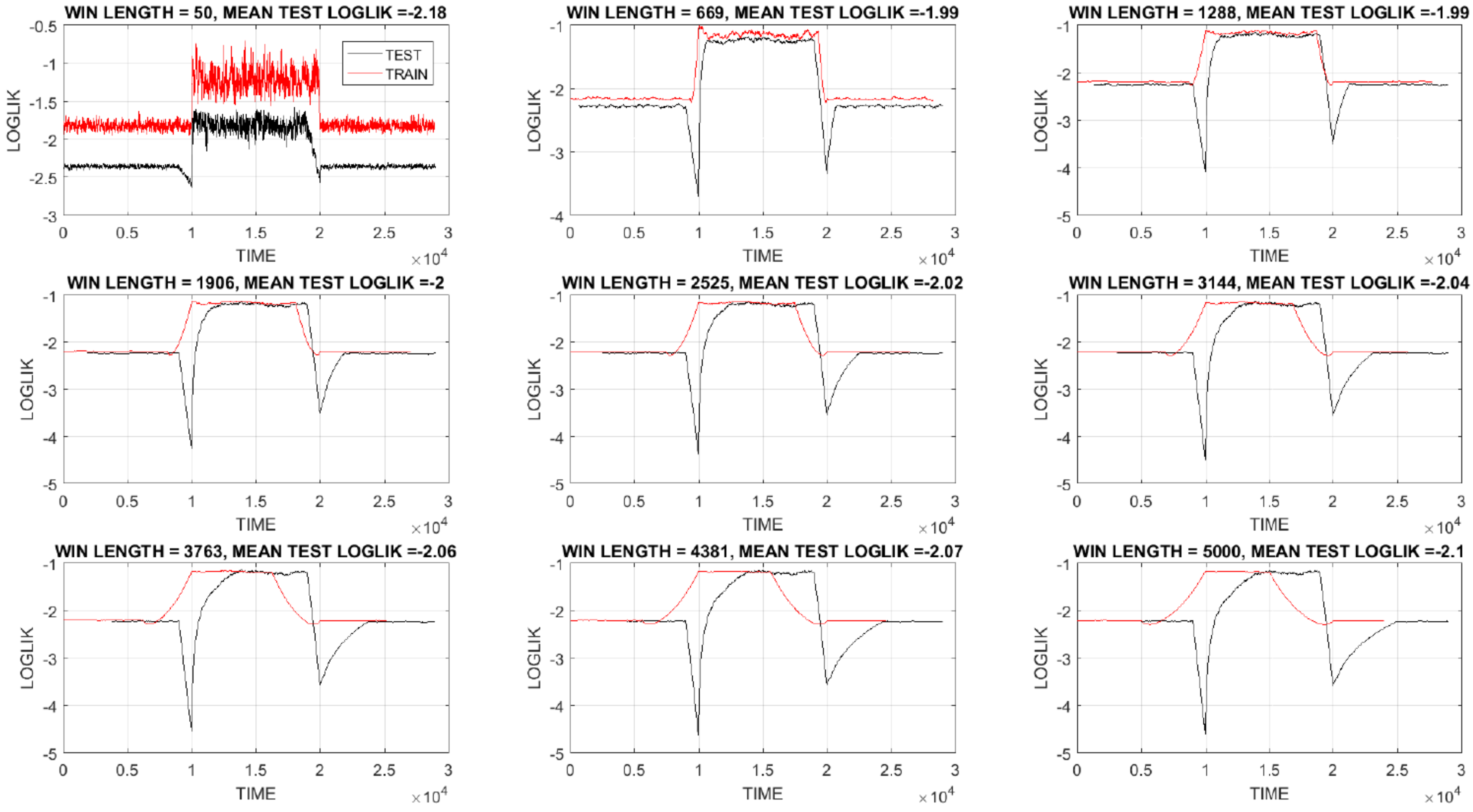
How to evaluate the error?



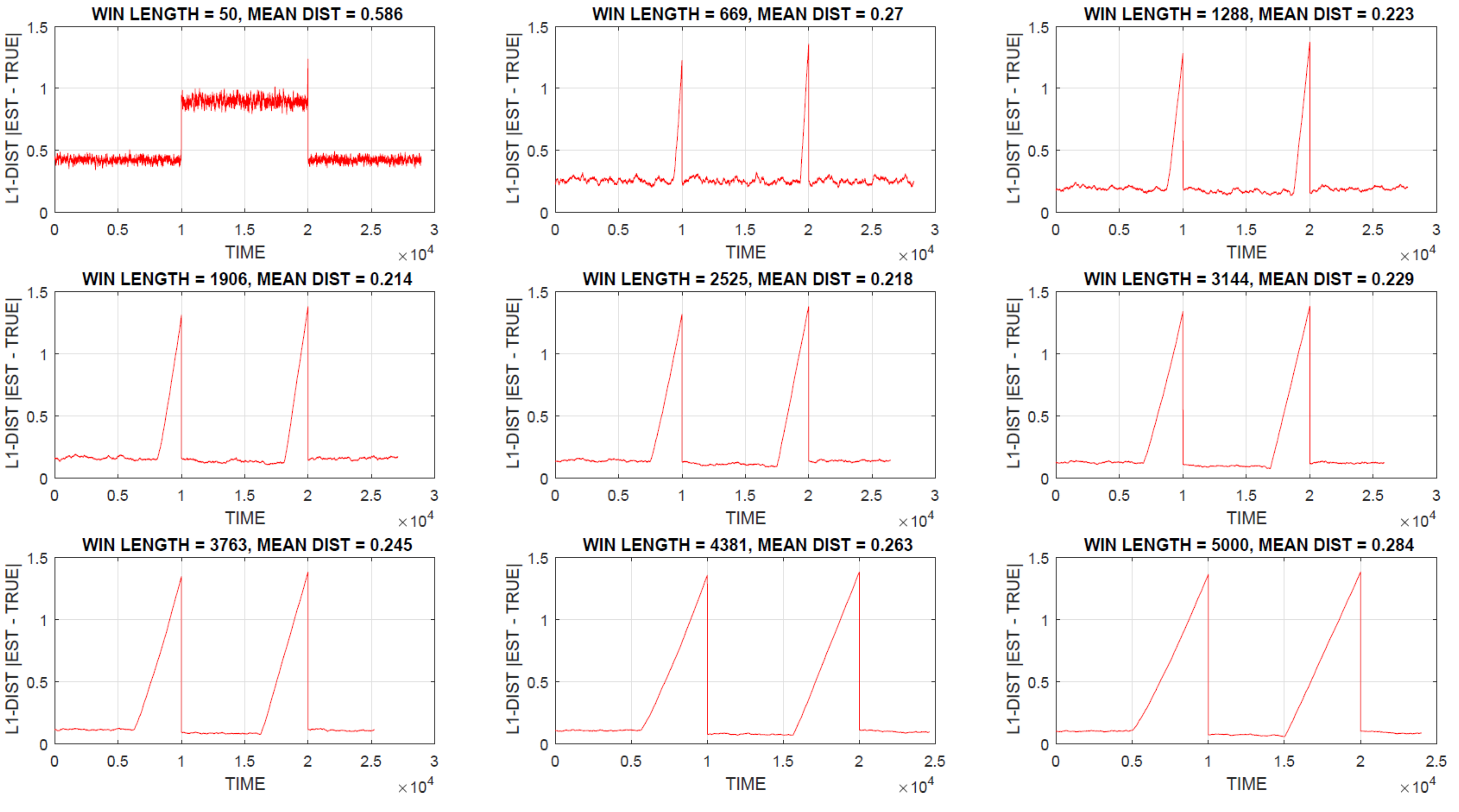
Window-based estimation of Markov model



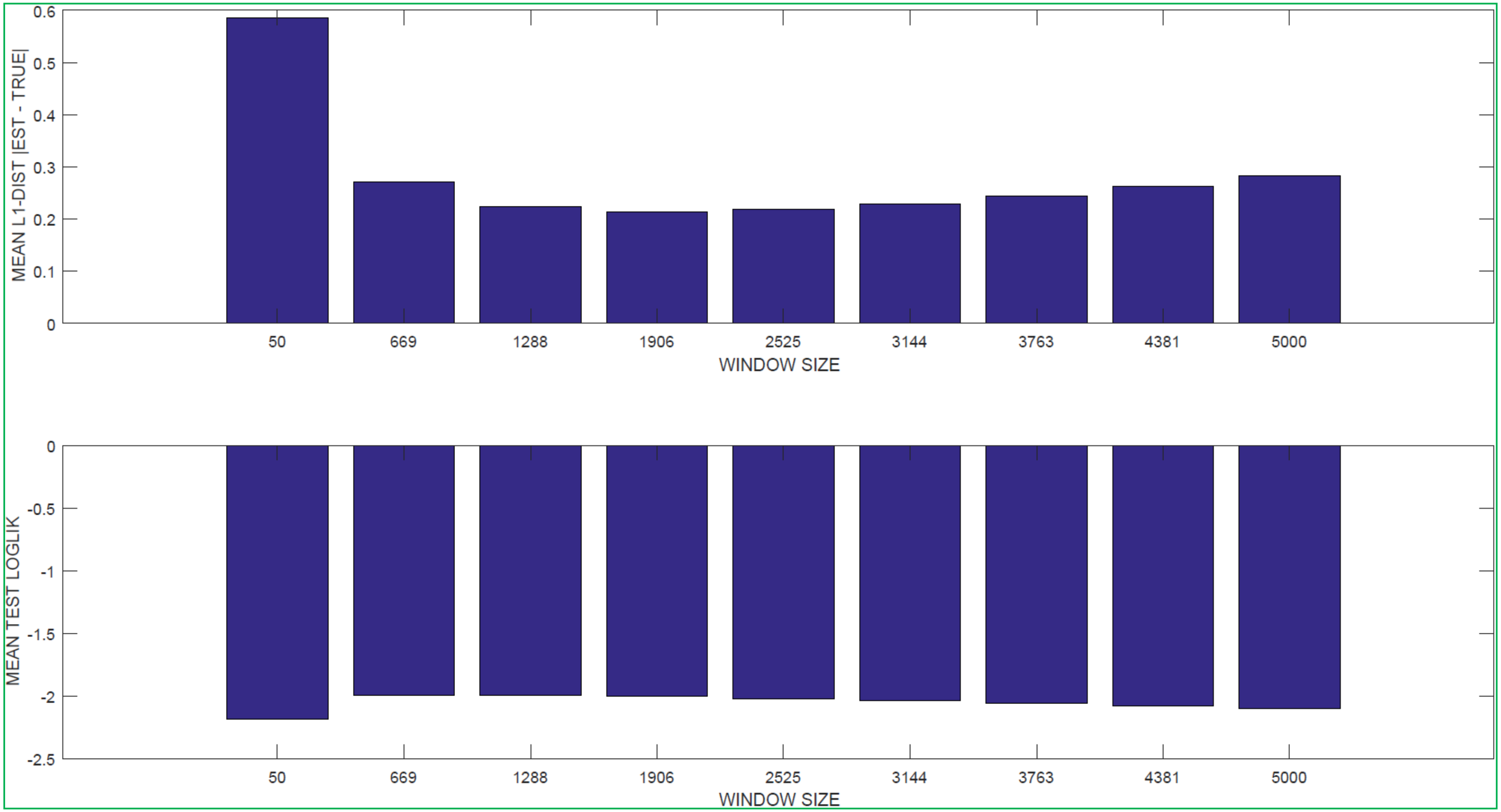
Window-based estimation of Markov model ..future windows based test MSE



Window-based estimation of Markov model
L1 distance to true model



Window-based estimation of Markov model
Evaluation of performance – L1 distance, MSE

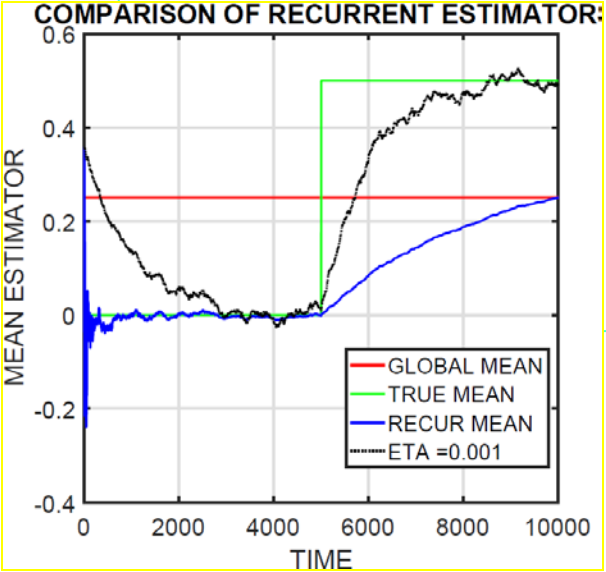


Parallel evaluation – ensemble voting
Martinez-Ramon, M., Arenas-Garcia, J., Navia-Vázquez, A. and Figueiras-Vidal, A.R., 2002.
An adaptive combination of adaptive filters for plant identification. In *Digital Signal Processing, 2002. 14th International Conference on* (Vol. 2, pp. 1195-1198). IEEE.

Dynamic estimator of the mean

Dynamic updates for stream of data $\{x_1, x_2, ..., x_N\}$, $\mu = \frac{1}{N}\sum_{n=1}^{N} x_n$

$$
\begin{aligned}
\mu_N &= \frac{1}{N} x_N + \frac{1}{N}\sum_{n=1}^{N-1} x_n \\
&= \frac{1}{N} x_N + \frac{N-1}{N}\frac{1}{N-1}\sum_{n=1}^{N-1} x_n \\
&= \frac{1}{N} x_N + \frac{N-1}{N}\mu_{N-1} \\
&= \mu_{N-1} + \frac{1}{N}(x_N - \mu_{N-1})
\end{aligned}
$$

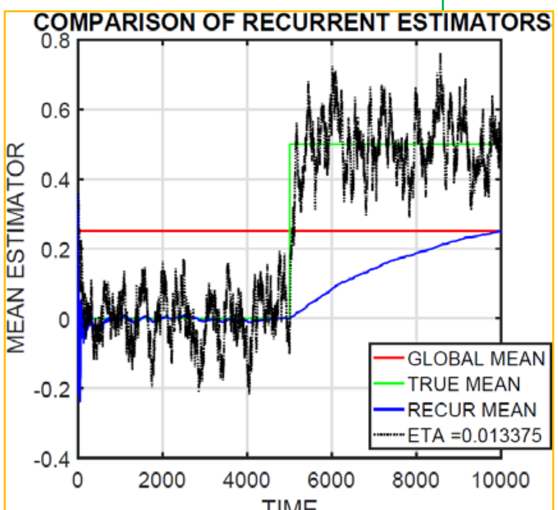


Non-stationarity, stochastic gradient

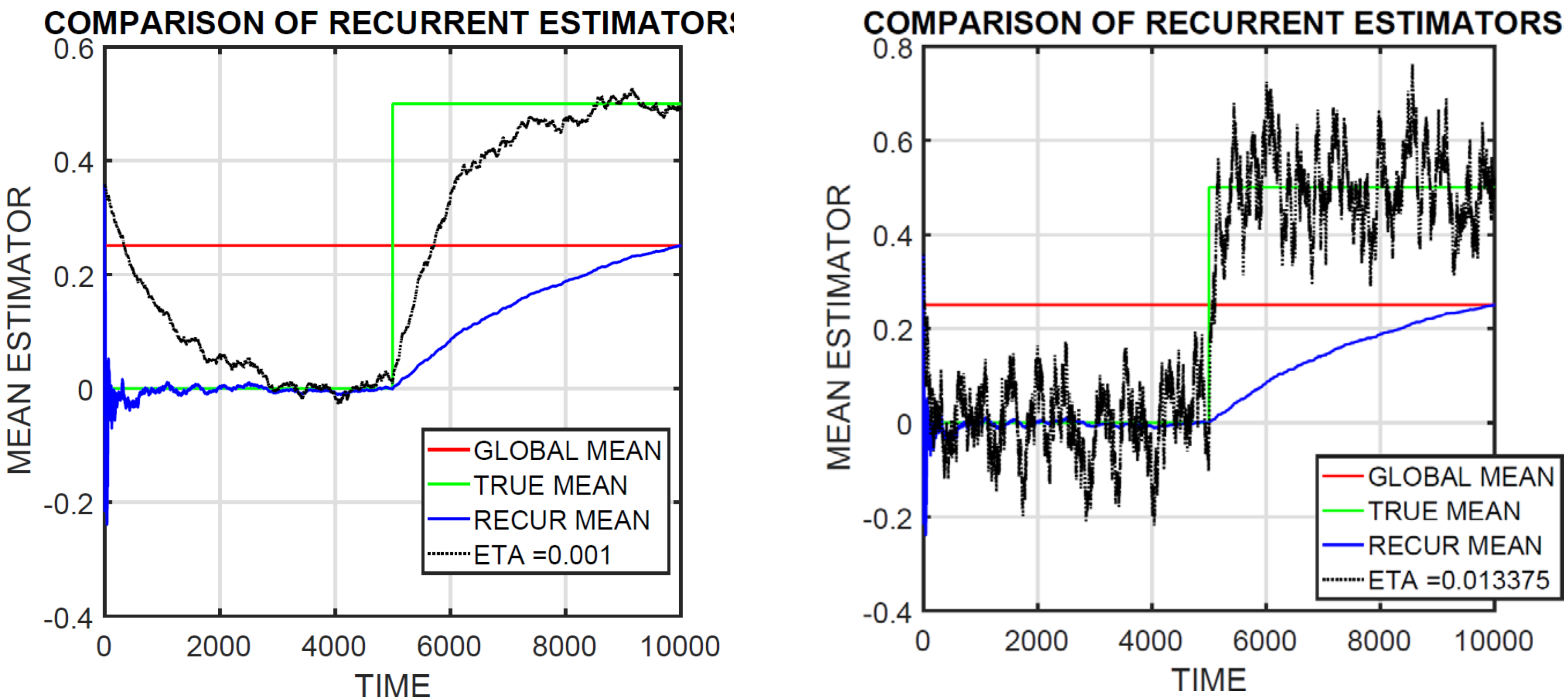
$$
\begin{aligned}
\mu_N &= \mu_{N-1} - \eta \frac{\partial}{\partial \mu}\left[\frac{1}{2}(x_N - \mu)^2\right]_{N-1} \\
&= \mu_{N-1} + \eta(x_N - \mu_{N-1})
\end{aligned}
$$

Difference equation has explicit solution

$$
\begin{aligned}
\mu_N &= \sum_{n=1}^{N} \eta(1-\eta)^{N-n} x_n \\
&= \eta \sum_{n=1}^{N} \exp\left((N-n)\log(1-\eta)\right) x_n \\
&\approx \eta \sum_{n=1}^{N} \exp\left(-(N-n)\eta\right) x_n \\
&\approx \eta \sum_{q=1}^{N} \exp(-q\eta) x_{N-q} \\
&\approx \frac{1}{W} \sum_{q=1}^{W} x_{N-q}
\end{aligned}
$$



Compare dynamic estimators in non-stationary data



Bayesian linear learning

Let $\mathcal{D} = \{(\mathbf{x}_1, t_1), (\mathbf{x}_2, t_2), ..., (\mathbf{x}_N, t_N)\}$ be a data set of N samples with $\mathbf{x} \in \mathbb{R}^d$.

$$\boxed{t = \mathbf{w}^\top \mathbf{x} + \epsilon}$$

$$p(\mathcal{D}|\mathbf{w}, \beta) = \left(\sqrt{\frac{\beta}{2\pi}}\right)^N \exp\left(-\frac{\beta}{2}\sum_{n=1}^{N}(t_n - \mathbf{w}^\top \mathbf{x}_n)^2\right)$$

assign a standard Gaussian prior to the weights $\mathbf{w} \sim \mathcal{N}(0, \alpha^{-1}\mathbf{I})$

$$\begin{aligned} p(\mathbf{w}|\alpha, \beta, \mathcal{D}) &= \frac{p(\mathcal{D}|\mathbf{w}, \beta)p(\mathbf{w}|\alpha)}{p(\mathcal{D}|\alpha, \beta)} \\ &\propto \left(\sqrt{\frac{\beta}{2\pi}}\right)^N \exp\left(-\frac{\beta}{2}\sum_{n=1}^{N}(t_n - \mathbf{w}^\top \mathbf{x}_n)^2\right) \left(\sqrt{\frac{\alpha}{2\pi}}\right)^d \exp\left(-\frac{\alpha}{2}||\mathbf{w}||^2\right) \end{aligned}$$

The product between $\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $\mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$, is proportional to $\mathcal{N}(\boldsymbol{\mu}_p, \boldsymbol{\Sigma}_p)$ with mean vector and covariance matrix given by,

$$\begin{aligned} \boldsymbol{\mu}_p &= \left(\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1}\right)^{-1} \left(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2\right) \\ \boldsymbol{\Sigma}_p &= \left(\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1}\right)^{-1}. \end{aligned}$$

Bayesian linear learning

The product between $\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $\mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$, is proportional to $\mathcal{N}(\boldsymbol{\mu}_p, \boldsymbol{\Sigma}_p)$ with mean vector and covariance matrix given by,

$$\begin{aligned} \boldsymbol{\mu}_p &= \left(\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1}\right)^{-1} \left(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2\right) \\ \boldsymbol{\Sigma}_p &= \left(\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1}\right)^{-1}. \end{aligned}$$

$$\begin{aligned} p(\mathbf{w}|\alpha, \beta, \mathcal{D}) &= \frac{p(\mathcal{D}|\mathbf{w}, \beta) p(\mathbf{w}|\alpha)}{p(\mathcal{D}|\alpha, \beta)} \\ &\propto \left(\sqrt{\frac{\beta}{2\pi}}\right)^N \exp\left(-\frac{\beta}{2}\sum_{n=1}^{N}(t_n - \mathbf{w}^\top \mathbf{x}_n)^2\right) \left(\sqrt{\frac{\alpha}{2\pi}}\right)^d \exp\left(-\frac{\alpha}{2}||\mathbf{w}||^2\right) \end{aligned}$$

In this case the prior is given by $\boldsymbol{\mu}_2 \equiv \boldsymbol{\mu}_{\text{prior}} = \mathbf{0}$ and $\boldsymbol{\Sigma}_2 \equiv \boldsymbol{\Sigma}_{\text{prior}} = \alpha^{-1}\mathbf{I}$. For the likelihood a bit of algebra leads to,

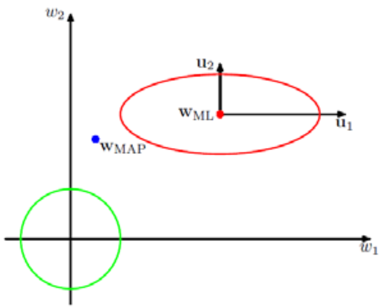
$$\begin{aligned} \boldsymbol{\mu}_1 &\equiv \left(\sum_{n=1}^{N} \mathbf{x}_n \mathbf{x}_n^\top\right)^{-1} \sum_{n=1}^{N} \mathbf{x}_n t_n \\ \boldsymbol{\Sigma}_1 &\equiv \left(\beta \sum_{n=1}^{N} \mathbf{x}_n \mathbf{x}_n^\top\right)^{-1}. \end{aligned} \qquad (3)$$

Bayesian linear learning

Hence the posterior mean vector and covariance matrix are found as

$$
\begin{aligned}
\boldsymbol{\mu}_p &\equiv \left(\alpha \mathbf{I} + \sum_{n=1}^{N} \mathbf{x}_n \mathbf{x}_n^\top\right)^{-1} \sum_{n=1}^{N} \mathbf{x}_n t_n \\
\boldsymbol{\Sigma}_p &\equiv \left(\alpha \mathbf{I} + \beta \sum_{n=1}^{N} \mathbf{x}_n \mathbf{x}_n^\top\right)^{-1} .
\end{aligned}
$$

Figure 3.15 Contours of the likelihood function (red) and the prior (green) in which the axes in parameter space have been rotated to align with the eigenvectors $\mathbf{u}_i$ of the Hessian. For $\alpha = 0$, the mode of the posterior is given by the maximum likelihood solution $\mathbf{w}_{\mathrm{ML}}$, whereas for nonzero α the mode is at $\mathbf{w}_{\mathrm{MAP}} = \mathbf{m}_N$. In the direction w_1 the eigenvalue λ_1, defined by (3.87), is small compared with α and so the quantity $\lambda_1/(\lambda_1 + \alpha)$ is close to zero, and the corresponding MAP value of w_1 is also close to zero. By contrast, in the direction w_2 the eigenvalue λ_2 is large compared with α and so the quantity $\lambda_2/(\lambda_2 + \alpha)$ is close to unity, and the MAP value of w_2 is close to its maximum likelihood value.



The predictive density is computed as

$$
p(t_{N+1}|\mathbf{x}_{N+1}, \mathcal{D}) = \int p(t_{N+1}|\mathbf{x}_{N+1}, \mathbf{w}) p(\mathbf{w}|\mathcal{D}) d\mathbf{w}
$$

This is again a normal distribution. We note

$$
t_{N+1} = \mathbf{w}_N^\top \mathbf{x}_{N+1} + \epsilon_{N+1}, \quad \text{and} \quad \mathbf{w}_N \sim \mathcal{N}(\boldsymbol{\mu}_p, \boldsymbol{\Sigma}_p),
$$

which leads to the predictive mean and variance,

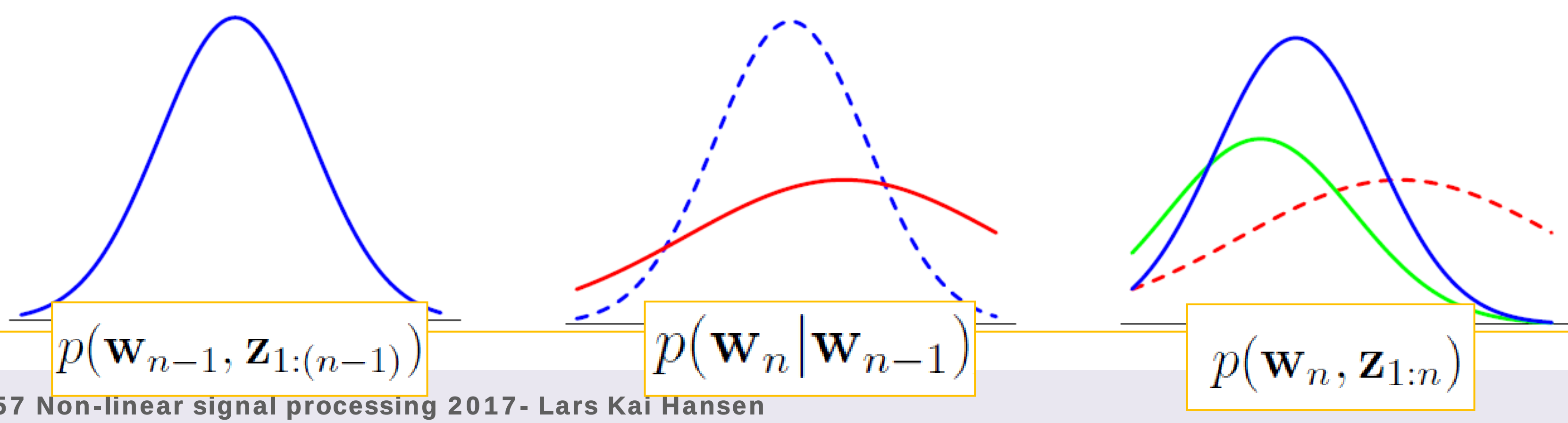
$$
\begin{aligned}
\boldsymbol{\mu}_{t_{N+1}} &= \boldsymbol{\mu}_p^\top \mathbf{x}_{N+1}, \\
\sigma_{t_{N+1}}^2 &= \beta^{-1} + \mathbf{x}_{N+1}^\top \boldsymbol{\Sigma}_p \mathbf{x}_{N+1}.
\end{aligned}
$$

Dynamic linear learning: Markov prior on $\mathbf{w}_n$

A natural prior is then the Markovian random walk $\mathbf{w}_n = \mathbf{w}_{n-1} + \boldsymbol{\nu}_n$ with $\boldsymbol{\nu}_n \sim \mathcal{N}(\mathbf{0}, \alpha^{-1}\mathbf{I})$,

$$p(\mathbf{w}_n|\mathbf{w}_{n-1}, \alpha) \quad = \quad \left(\sqrt{\frac{\alpha}{2\pi}}\right)^d \exp\left(-\frac{\alpha}{2}||\mathbf{w}_n - \mathbf{w}_{n-1}||^2\right),$$

$$p(\mathbf{w}_n|\mathbf{w}_{n-1}, \alpha) \quad = \quad \left(\sqrt{\frac{\alpha}{2\pi}}\right)^d \exp\left(-\frac{\alpha}{2}||\mathbf{w}_n - \mathbf{w}_{n-1}||^2\right)$$



Forward recursion

$$\begin{aligned}
p(\mathbf{w}_n, \mathbf{z}_{1:n}) &= \int p(\mathbf{w}_n, \mathbf{w}_{n-1}, \mathbf{z}_{1:n}) d\mathbf{w}_{n-1} \\
&= \int p(\mathbf{w}_n, \mathbf{w}_{n-1}, \mathbf{z}_n, \mathbf{z}_{1:(n-1)}) d\mathbf{w}_{n-1} \\
&= \int p(\mathbf{z}_n | \mathbf{w}_n, \mathbf{w}_{n-1}, \mathbf{z}_{1:(n-1)}) p(\mathbf{w}_n, \mathbf{w}_{n-1}, \mathbf{z}_{1:(n-1)}) d\mathbf{w}_{n-1} \\
&= p(\mathbf{z}_n | \mathbf{w}_n) \int p(\mathbf{w}_n | \mathbf{w}_{n-1}, \mathbf{z}_{1:(n-1)}) p(\mathbf{w}_{n-1}, \mathbf{z}_{1:(n-1)}) d\mathbf{w}_{n-1} \\
&= p(\mathbf{z}_n | \mathbf{w}_n) \int p(\mathbf{w}_n | \mathbf{w}_{n-1}) p(\mathbf{w}_{n-1}, \mathbf{z}_{1:(n-1)}) d\mathbf{w}_{n-1}.
\end{aligned}$$

The posterior distribution of $\mathbf{w}_n$, in turn, can be obtained by normalization,

$$p(\mathbf{w}_n | \mathbf{z}_{1:n}) = \frac{p(\mathbf{w}_n, \mathbf{z}_{1:n})}{\int p(\mathbf{w}_n, \mathbf{z}_{1:n}) d\mathbf{w}_n}.$$

Linear model

For the linear model we analyzed above, we get specifically,

$$
\begin{aligned}
p(\mathbf{z}_n|\mathbf{w}_n,\beta) &= p(t_n|\mathbf{w}_n,\mathbf{x}_n,\beta)p(\mathbf{x}_n) = \sqrt{\frac{\beta}{2\pi}}\exp\left(-\frac{\beta}{2}(t_n-\mathbf{w}^\top\mathbf{x}_n)^2\right)p(\mathbf{x}_n) \\
p(\mathbf{w}_n|\mathbf{w}_{n-1},\alpha) &= \left(\sqrt{\frac{\alpha}{2\pi}}\right)^d \exp\left(-\frac{\alpha}{2}||\mathbf{w}_n-\mathbf{w}_{n-1}||^2\right).
\end{aligned}
$$

Initialization

$$p(\mathbf{w}_1,\mathbf{z}_1) \propto p(\mathbf{z}_1|\mathbf{w}_1)p(t_1|\mathbf{x}_1,\mathbf{w}_1)p(\mathbf{x}_1)$$

$$p(\mathbf{w}_2,\mathbf{z}_{1:2}) = p(\mathbf{z}_2|\mathbf{w}_2,\beta)\int p(\mathbf{w}_2|\mathbf{w}_1,\alpha)p(\mathbf{w}_1,\mathbf{z}_1)d\mathbf{w}_1.$$

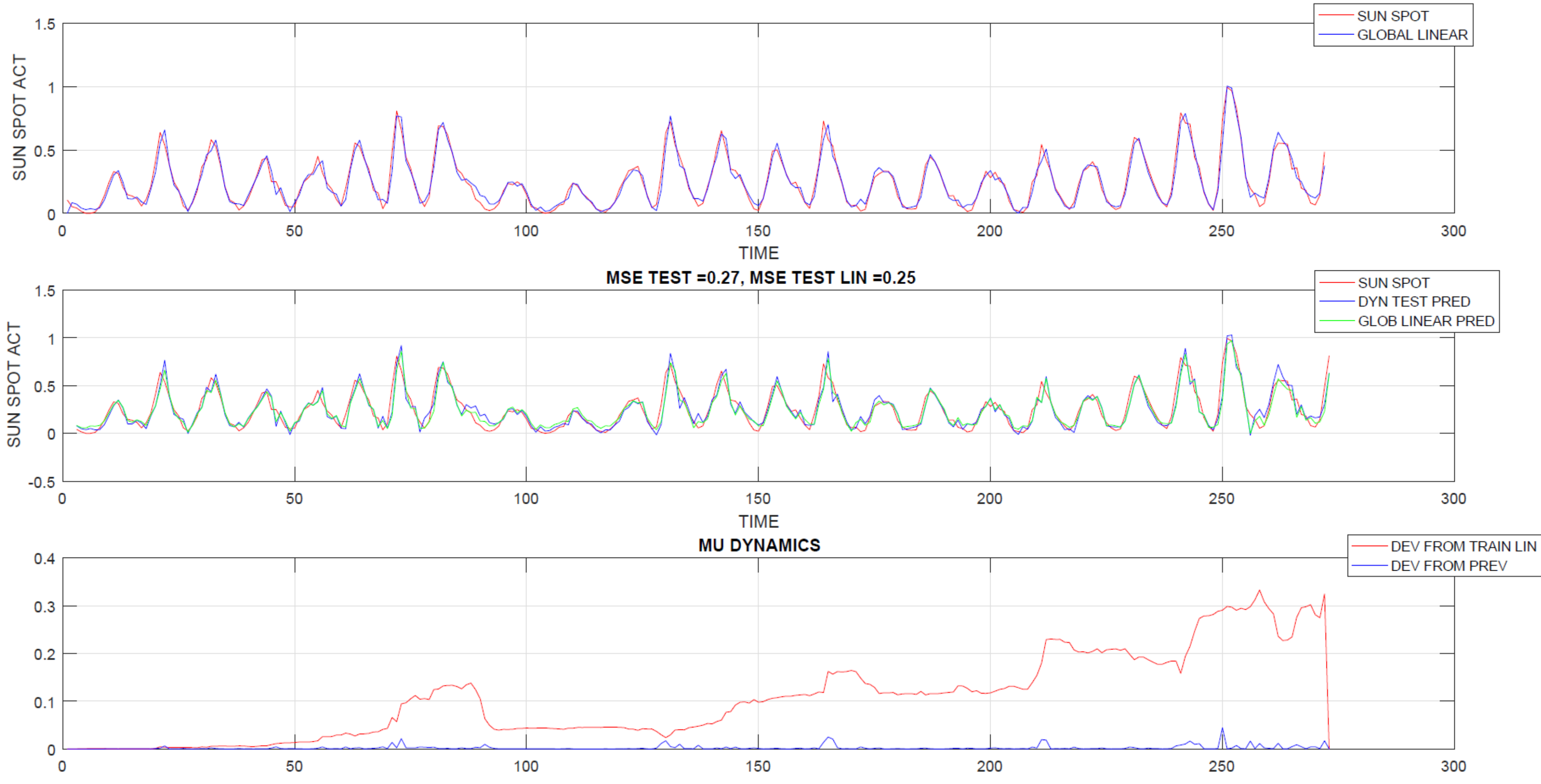
Dynamic linear model – message passing

$$p(\mathbf{w}_2, \mathbf{z}_{1:2}) = p(\mathbf{z}_2|\mathbf{w}_2, \beta) \int p(\mathbf{w}_2|\mathbf{w}_1, \alpha) p(\mathbf{w}_1, \mathbf{z}_1) d\mathbf{w}_1.$$

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{w},n} &= \left(\left(\boldsymbol{\Sigma}_{\mathbf{w},n-1}^{-1} + \alpha^{-1} \right)^{-1} + \beta \mathbf{x}_n \mathbf{x}_n^\top \right)^{-1} \left(\boldsymbol{\Sigma}_{\mathbf{w},n-1}^{-1} \boldsymbol{\mu}_{\mathbf{w},n-1} + \beta t_n \mathbf{x}_n \right) \\ \boldsymbol{\Sigma}_{\mathbf{w},n} &= \left(\left(\boldsymbol{\Sigma}_{\mathbf{w},n-1}^{-1} + \alpha^{-1} \right)^{-1} + \beta \mathbf{x}_n \mathbf{x}_n^\top \right)^{-1} \end{aligned}$$

α determines the effective window

Dynamic linear model: sun spots



PLEASE BRING YOUR LAPTOPS THURSDAY….