Nonlinear Thinking and Methods – Getting Started

STEPHEN MERRILL
MARQUETTE UNIVERSITY

MARK SHELHAMER
JOHNS HOPKINS UNIVERSITY

25th Annual SCTPLS Conference -- U of Florida
7/29/2015

What is “linear”

- Proportional response in rates, probabilities, or something else – solutions to linear differential equations are exponential and trigonometric
- The mathematics tends to be straightforward and in elementary courses. As an undergrad – one thinks the world “linear”
- Constants in models likely to have units that are meaningful and measureable
- Associated data can be analyzed by familiar statistical techniques (e.g. linear regression)
- Linear tends to be easy to approximate well over long times or quite large ranges in states
- Simulation tends to be straightforward to accomplish – results can be easily described

What is “nonlinear”? NOT Linear

- NOT Proportional response in rates, probabilities, or something else
- The mathematics tends to NOT be straightforward
- Constants likely to have units that are NOT meaningful and measureable
- Associated data can NOT be analyzed by familiar statistical techniques (e.g. regression)
- Linear may NOT approximate data well over even short times or small ranges in states
- Simulation might NOT be easy to accomplish – results are often NOT easy to interpret

How do you know that a process is NOT linear?

- Linear models do not “fit”
- The time series tests out as chaotic (not possible for a linear process)

Testing for nonlinearity in time series: the method of surrogate data

J Theiler, S Eubank, A Longtin, B Galdrikian, D. Farmer... - Physica D: Nonlinear ..., 1992 - Elsevier

Abstract We describe a statistical approach for identifying nonlinearity in time series. The method first specifies some linear process as a null hypothesis, then generates **surrogate** data sets which are consistent with this null hypothesis, ... construct a statistical test.

How do you know that what you are observing is random and not nonlinear?

- What of a process is deterministic and what is random? (Most "deterministic" processes have random aspects – either within the dynamics or as measurement noise.)
- Also of interest in a signal is when it is primarily stochastic (random) but has a nonrandom (nonlinear) structure within.
- There is also a large literature on nonlinear time series.

A good example to study here is standard (linear) ARIMA time series models when the randomness is within the dynamics. It is a regression problem otherwise.

Nonlinear thinking often uses models to express them. What is a model?

- Fashion model– a person who is employed to display, advertise and promote commercial products (notably fashion clothing)
- Civil Engineering – scale model to test design aspects of a structure.
- Biology – animal model which mimics important aspects of the human in disease or drug response
- Mathematics – mathematical models employing gross simplifications but hopefully retaining important aspects. “First principles” builds the model
- Simulation models – models nearing the complexity of the system being studied. Used as a way to describe the data in this system. Data validates the model.
- Statistical models – model descriptions created from regression (for instance). Data builds the model.

More “models”

- Hybrid models – combinations of the above; e.g., a mouse with a human immune system or a simulation model with some mathematical models embedded
- Physical model – physical representation of an object often for visualization
- 3D modelling – 3D polygonal representation of an object, usually displayed with a computer, e.g., FEM.
- Model building – hobby centered around the construction of material replicas, usually scale models.
- Solid modelling – study of very accurate representations of the solid parts of an object, as in CAD.
- Conceptual Models – analogies used to aid understanding (my definition)
- Role models and Thesis directors – an example for others
- Crash Test Dummy – see above
- Flight simulator – to provide repetitions and situations to increase experience

Some common themes for models

- A real system, or collection of real things that are of interest.

  For instance, it may be "all bridges that could be built using a particular design"

- Context of use (purpose and setting for which it was constructed). The nature of the specific use varies with the model.
- The "model" represents idealized and simplified version of the real system. The level of simplification is dictated by the type of model.
- The model allows uses not possible in the real system.

  For instance, testing of a device under different failure modes.

- A test (sometimes heuristic, usually data driven) to determine if the model is sufficient at the current stage of development. If not, "improvements" are made and the test reapplied. This could be more than a validation.
- An understanding that the model is not the system – the nature of its differences and what difficulties those differences may cause.

The model is often confused with the real system in models that have been used for a long time.

Choosing a type of model

- Each type of model has its strengths and weaknesses.

For instance, a simulation model is at nearly the complexity of the system it was designed to describe (which may have been too complicated to understand).

- Whether the model is a "good" one depends on the intended use. If one can address the questions that motivated the model – it is a good model. Note that it need not reproduce data perfectly as a criteria (unless that is a requirement).
- Limitations of data place real limits on the complexity of a model. Data gaps can force too many arbitrary decisions or parameter values to make the results believable for a complex model.
- The type of model that is appropriate for a given situation is often determined by the system under study and the nature of the questions – not the "bias of training."

Other aspects

- Legacy models – created in the past, found useful, and continue to be used. Often associated with large projects.

  Often associated with confusing the system with the model. e.g. chemistry as described in CHEM001 where they are actually teaching a model of the real interactions.

- Single use (disposable) models – created to answer a specific question – it did its job, and now is no longer used.

  An example would be competing models created to see which one gave best fit.

- Deterministic model – every simulation with the same parameters and initial conditions should result in the same result.

- Stochastic model – randomness, usually in addition to deterministic aspects, plays a role in the results.

  A good example is (again) autoregressive (AR) time series models.

Some examples illustrating these ideas

- Clinical course in autoimmune thyroiditis (Hashimoto's)

  Limited data, individual differences, interesting question
- Heart rate variability classification using Markov Chains

  Method to identify patterns and changes in patterns (like stats model)
- Early stages in HIV infection

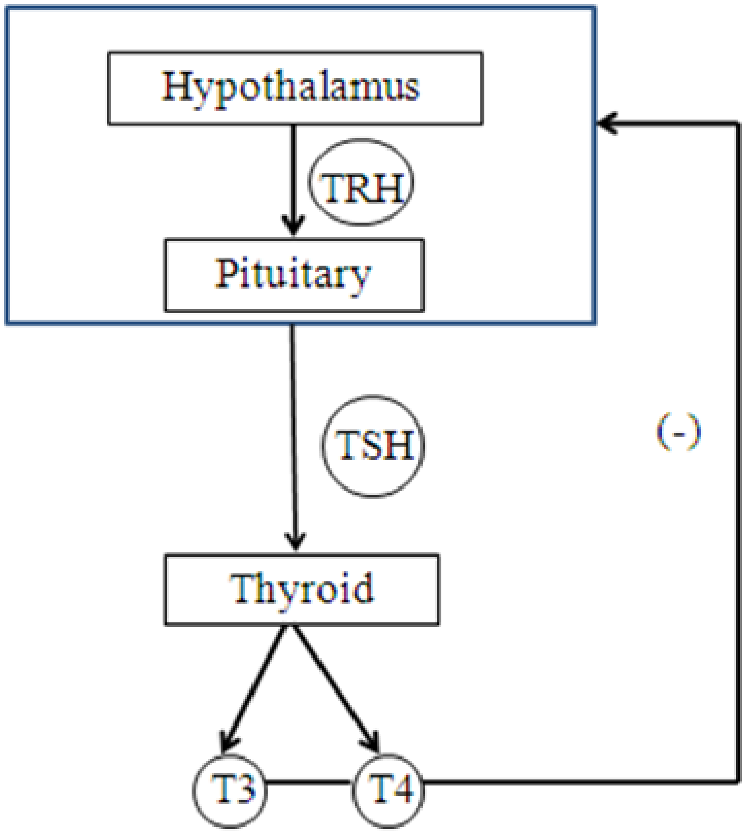
  Stochastic model – identify key parameters
- Dynamics of engraftment in hematopoietic stem cell transplants

  Data in the context of a model – discovery of dynamical surprise
- Spatial variation in cDNA microarray

  simple description – complicated result

Hashimoto's – autoimmune Destruction of the thyroid

HPT Axis



The Question

- Can we determine if the patient will eventually develop chronic hypothyroidism?

(If so, when do we start treatment to minimize effects of the disease)

The Data

- 119 patients with autoimmune antibody in Sicily. Each patient has 2-7 measurements of TSH and free T4 at irregular intervals over years.
- Although there are models of the HPT axis, none exist for this situation where the response of the thyroid is disrupted.

Possible Responses

- Impossible – need more data
- Impossible – individual differences reduce the usefulness of even the little data available
- Maybe a simple model can tell us something

$$\frac{dFT4}{dt} = \frac{k_3\, T\, TSH}{k_d + TSH} - k_4 FT4$$

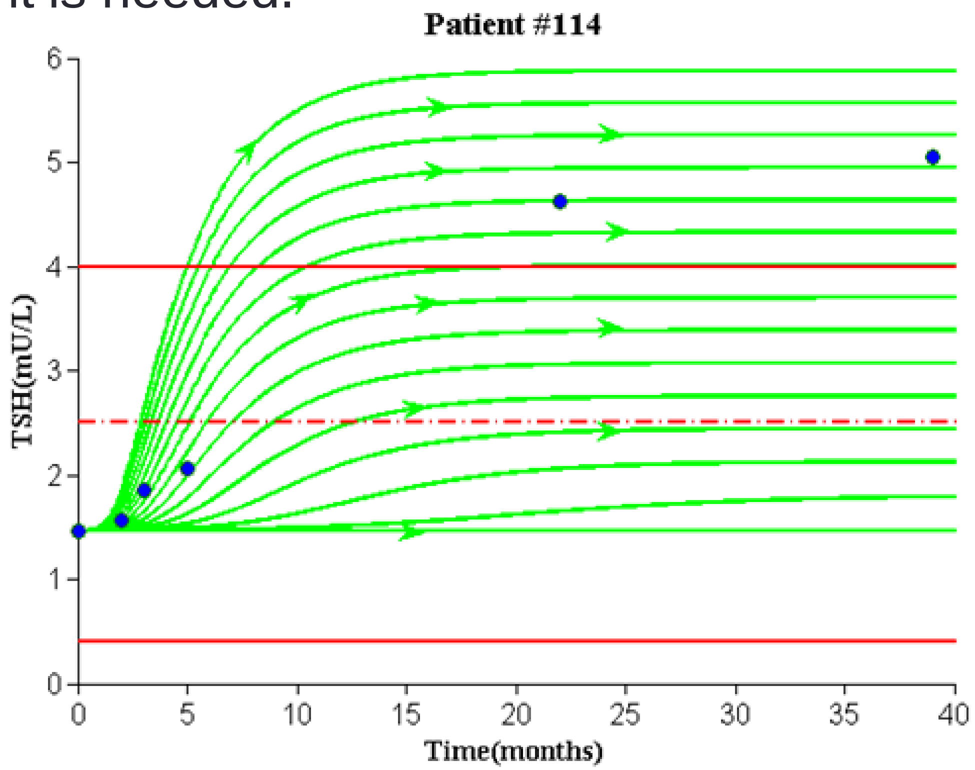
$$\frac{dTSH}{dt} = k_1 - \frac{k_1\, FT4}{k_a + FT4} - k_2\, TSH$$

$$\frac{dT}{dt} = k_5\left(\frac{TSH}{T} - N\right) - k_6 Ab\, T$$

$$\frac{dAb}{dt} = k_7 Ab\, T - k_8\, Ab$$

Results

1. Dynamics are simple – depending on only a few parameters
2. For each patient, an approach to the equilibrium can be determined
3. The position of the equilibrium will determine if (and when) treatment is needed.



HRV using Markov chains

- Heart Rate Variability (HRV) describes the beat-to-beat variation in the time interval between beats as seen on ECG (R-R interval).
- It is described by many different indices.
- The variability is due to several different control mechanisms in the systems
- Operation of the controls are affected by drugs (specifically here, anesthesia)

The Question

- Design a real-time monitor to detect a patient heading toward sudden cardiac arrest

The Data

- pediatric patients

undergoing surgery

- Patient 29

  2.5 years old

- Patient 55

  7.5 years old

Early with halothane

Late with atropine

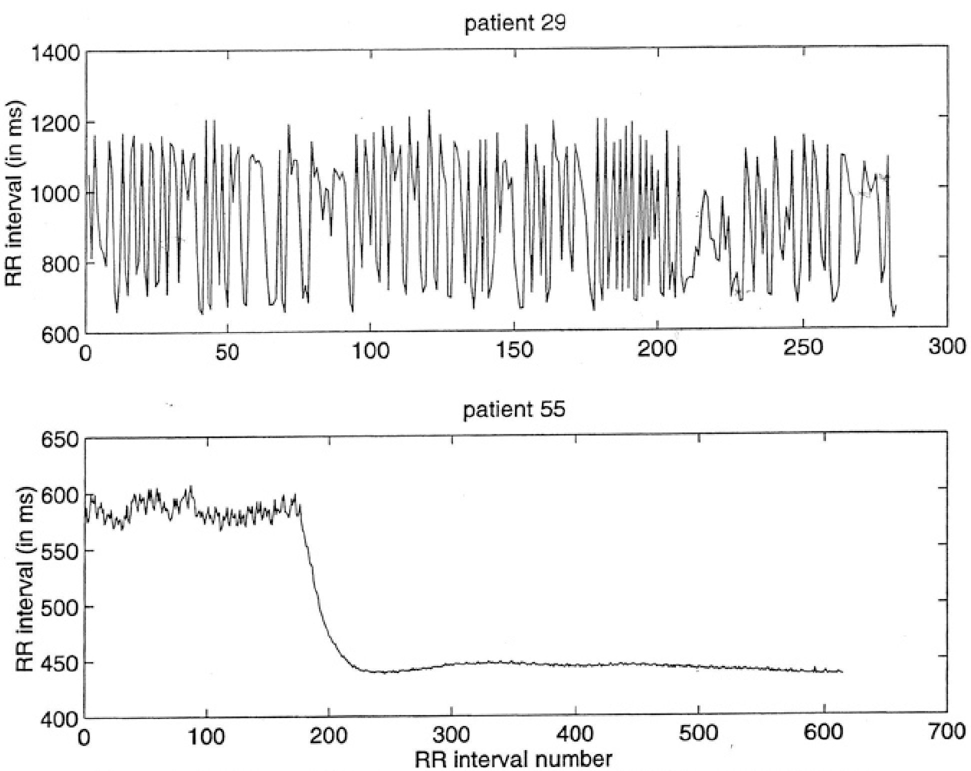


Figure 1 RR interval data from Patient 29 and Patient 55. In the analysis, the second set will be divided into an early and late phase.

Lag 1 maps

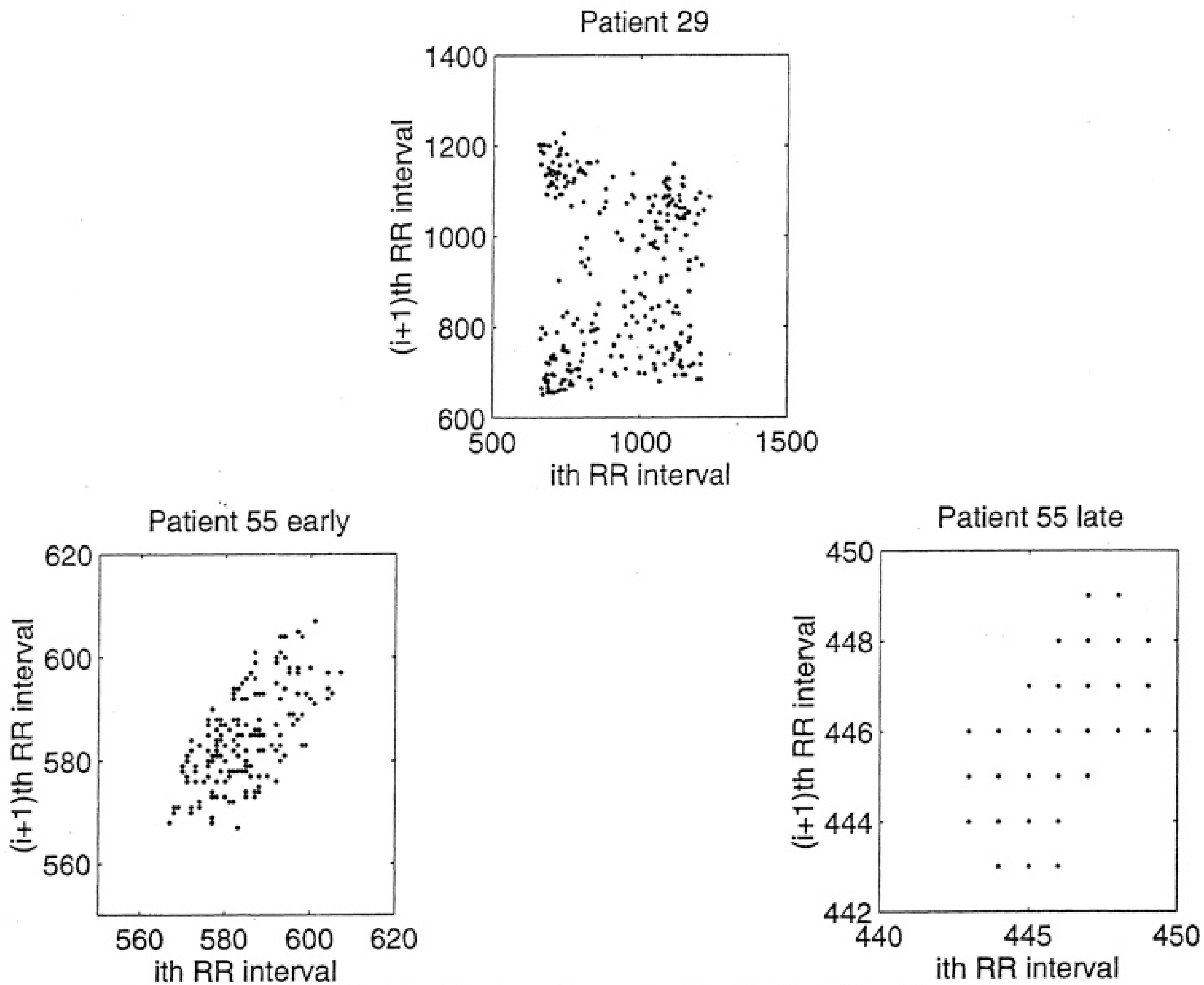


Figure 3 Lag 1 plots for the three data sets. The Patient 29 pattern has been described previously as a "complex pattern" (Woo et al. [1992]). Patient 55 early and late data plots would be classified as "torpedo patterns" by Woo et al. [1992].

Model of R-R interval data

- Create an empirical Markov chain. Data is in the form of sequence of numbers and lag 1 maps indicate first order structure.
- Need to define the bin size corresponding to the length of the data set (usual number of bins used was 10). Then estimating transition probabilities to get a transition matrix. Note that many possible transitions are not observed.
- Transient aspects of the chain are of interest (not asymptotic behavior). Characterization of the dynamics (or the resulting matrix) is desired.
- Basic idea is to use properties of the matrix (such as eigenvalues) to distinguish between cases.

Eigenvalue maps

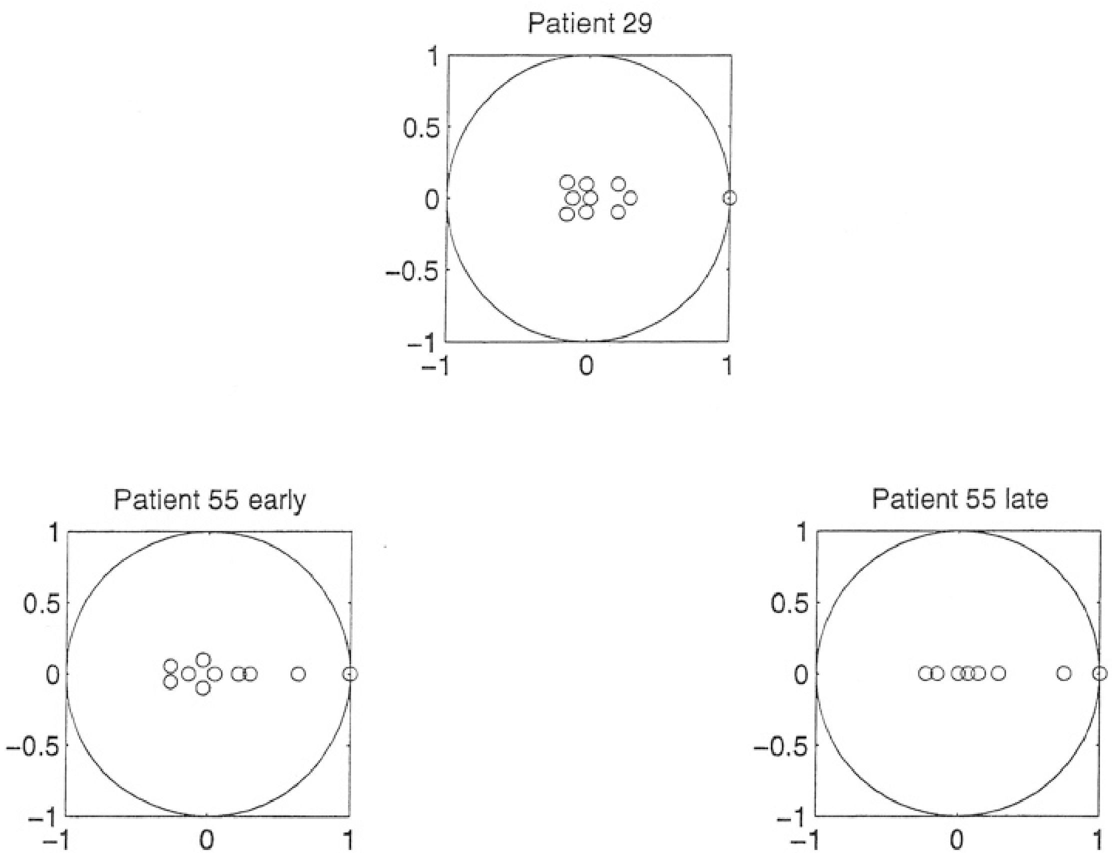


Figure 5 Eigenvalues for each of the three transition matrices are shown in relation to the unit circle. Besides the nearness of the "non-1" eigenvalues to the unit circle, notice the differences in the number of complex eigenvalues.

Early HIV infection

- Long term time-course of the infection depends on the “set point” -- related to the state of the infection at the time the immune response controls the initial acute infection.

- Interested in computing the incubation-time distribution (defined as the time from infection to a fixed clinical marker such as seroconversion -- the appearance of anti-HIV antibodies).

The Question

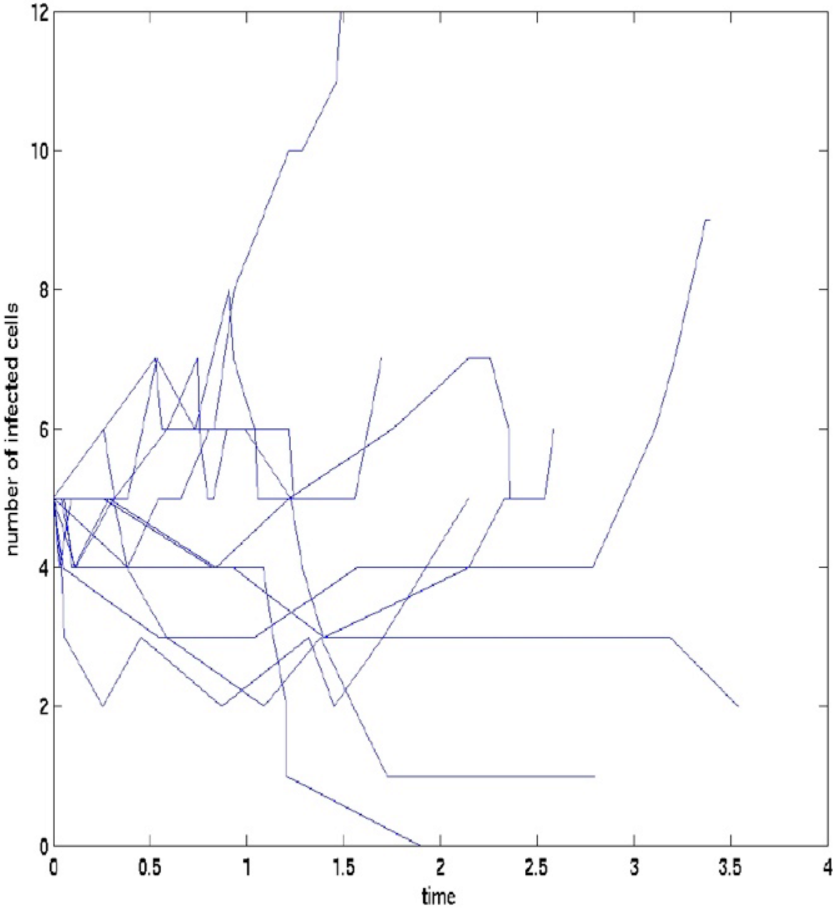
- What is the nature of this distribution (for instance its mean) and what are the critical parameters.

The Data

- Existing statistical descriptions of the distribution built from large numbers of patients (censoring is a problem here)

The Model

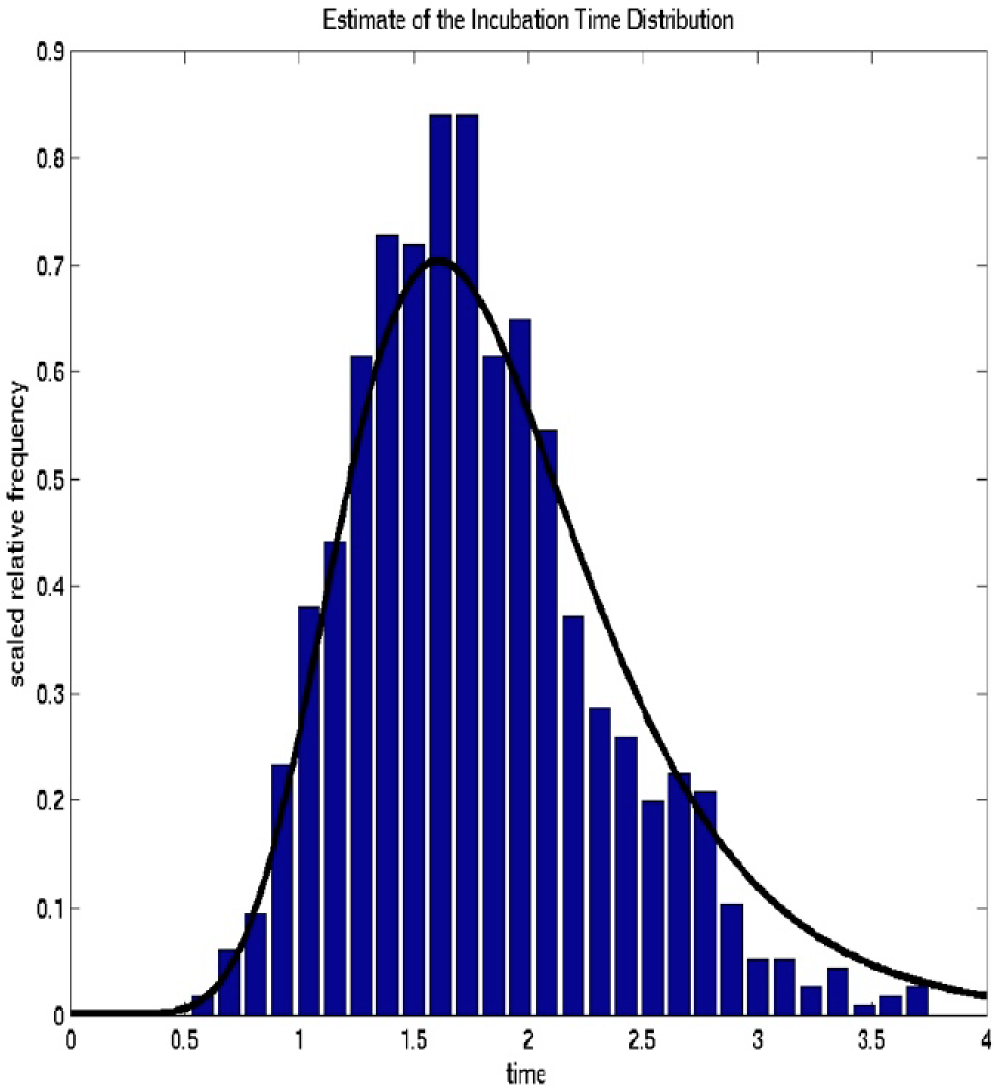
- Branching process with immigration. The basic quantity tracked is the number of infected T cells.
- Only data that exists is the distributional information
- Model designed to see what happens in the initial stages.



Sample Paths of branching model -- no immigration

Simulations

1000 simulations of model, recording the time at which the number of infected cells reach some fixed Value (stopping time distribution)



Dynamics of Engraftment

- Hematopoietic stem cells can be collected from blood (or bone marrow) for later infusion (transplantation) after high-dose chemotherapy.
- In autologous transplants, no rejection is present.
- Interested in monitoring engraftment (return to normal levels) of each cell type – primarily leukocytes (WBC in early counts), lymphocytes, platelets, and red cells.

The Question

- From daily blood counts, estimate "time to engraftment" and detect possible problems before they occur.

The Data

- Daily counts from 32 women following transplantation

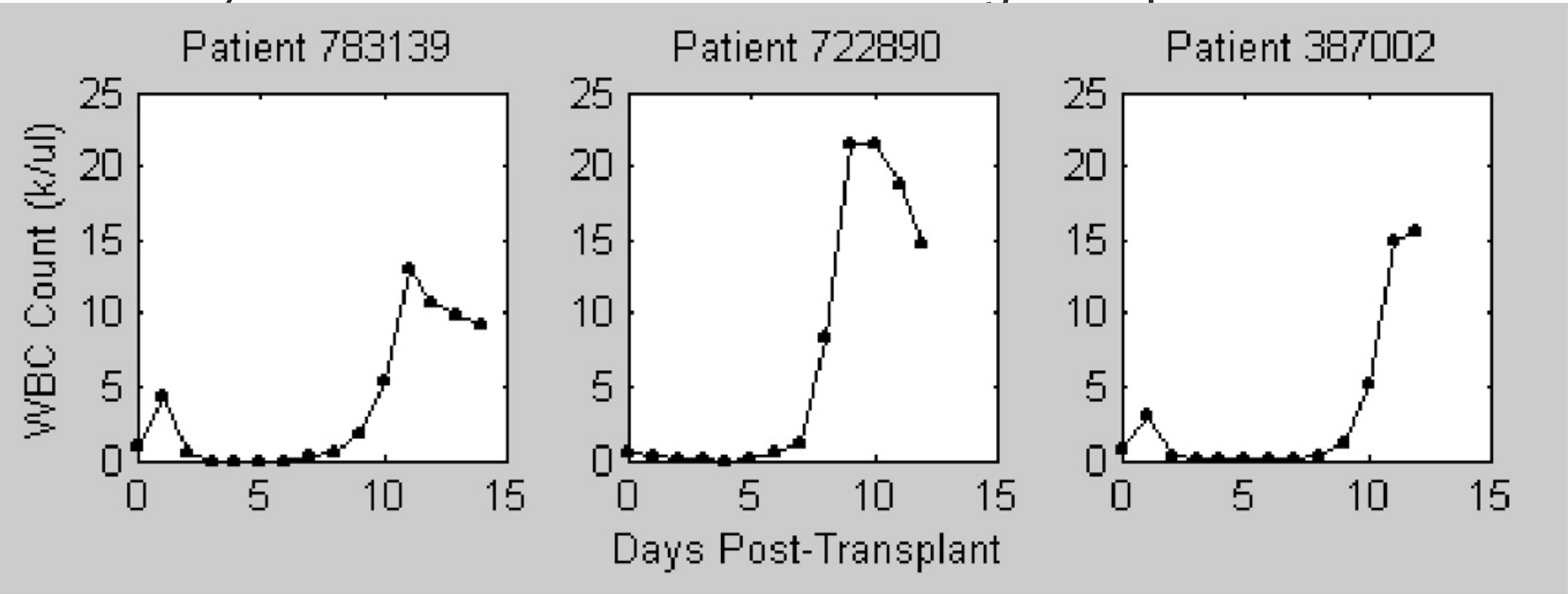
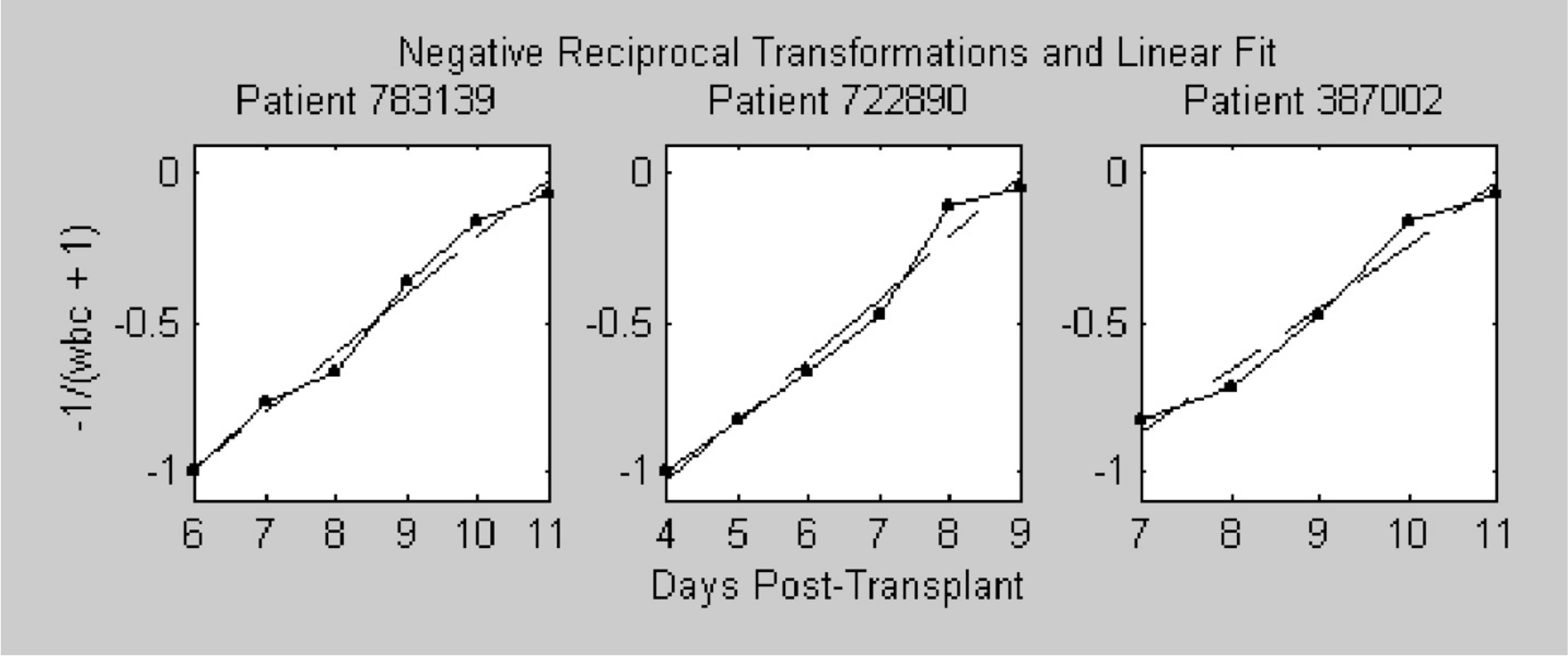


Figure 3.1. Typical WBC plots

The Model

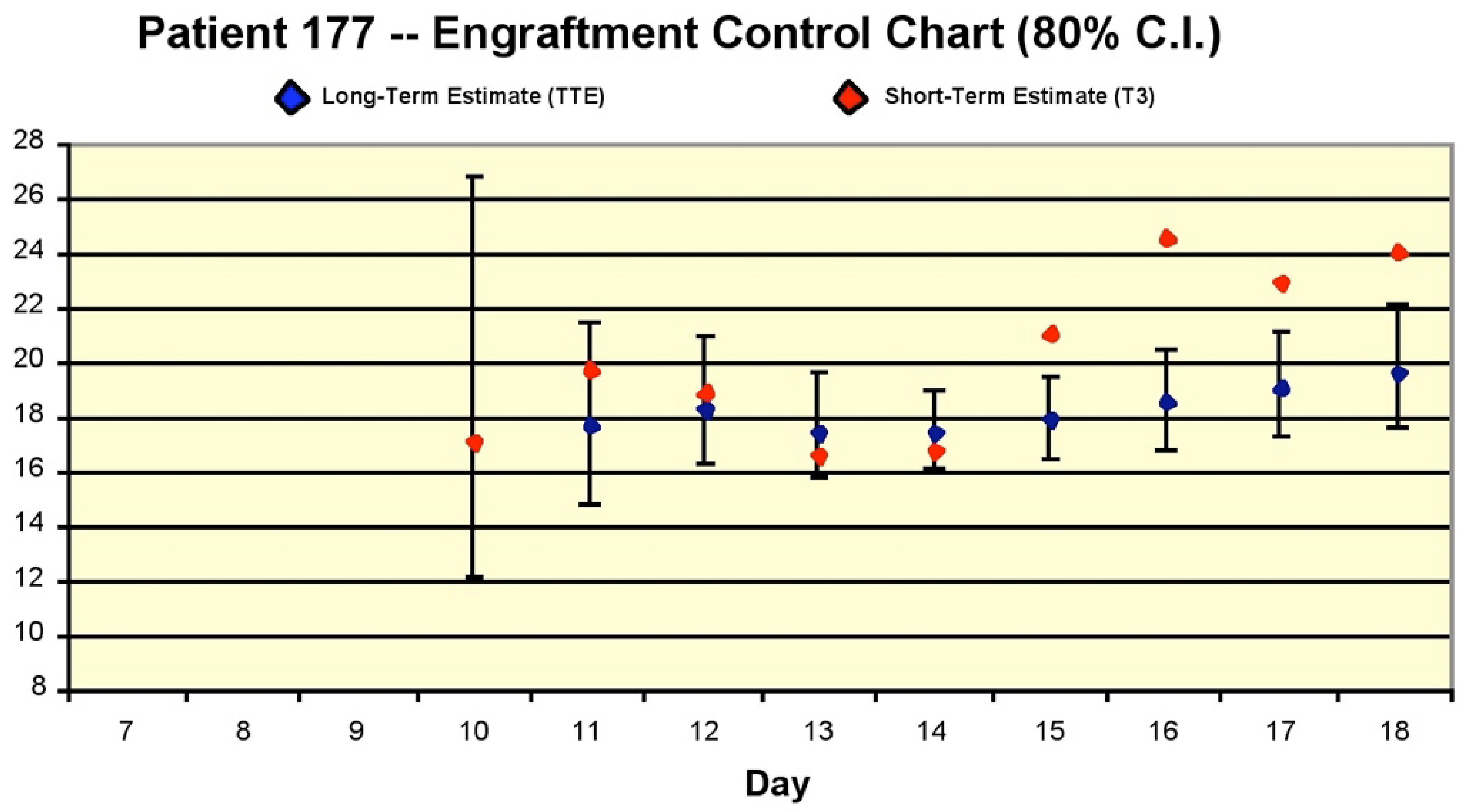
- Reciprocal plot shows hyperbolic growth $r^2 = .94$



Results

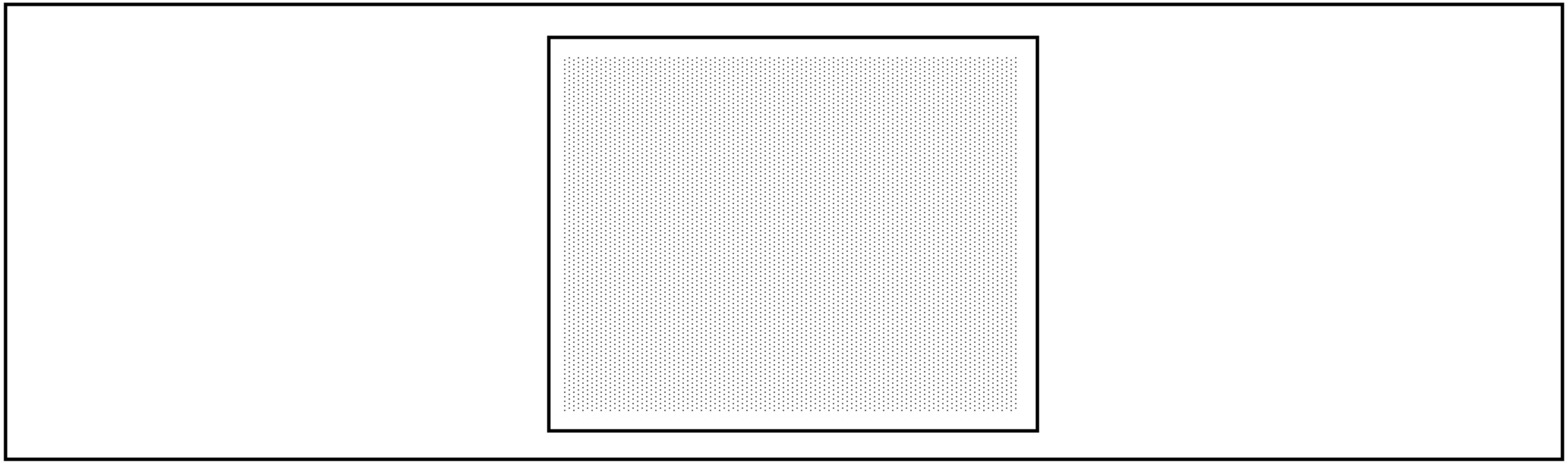
- Estimating the position of the asymptote (and as it changes with each days data) allows the estimation of the time to engraftment -- and resulting release from the hospital.
- Changes in the estimates indicate problems, represented in a control chart.
- Similar results for lymphocytes. For platelets, polynomial growth was observed.

Final Result – a Control Chart



Spatial variation in the microarray

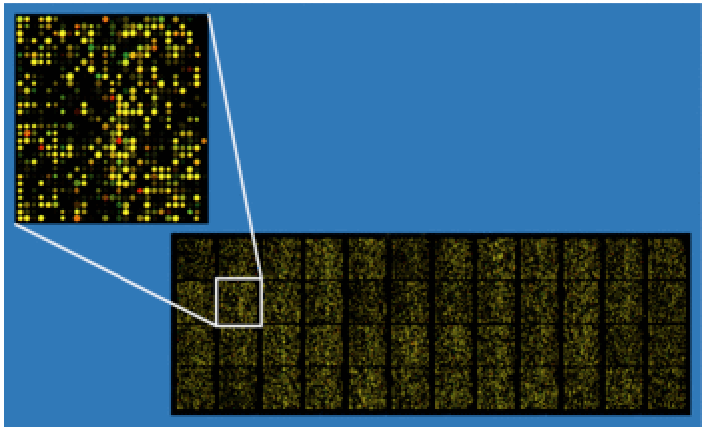
- cDNA microarray used to identify genes that are differentially under or over expressed in a sample (as seen through mRNA).



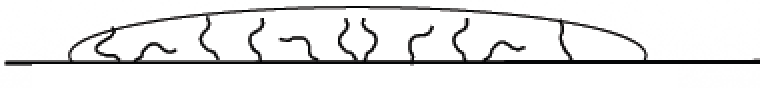
The Question

- Why are my edges bright? Or what “normalization” is the right one to use for this process.
- What is the source of the variability – how should replicates be done (and what statistics should be used)

The Data



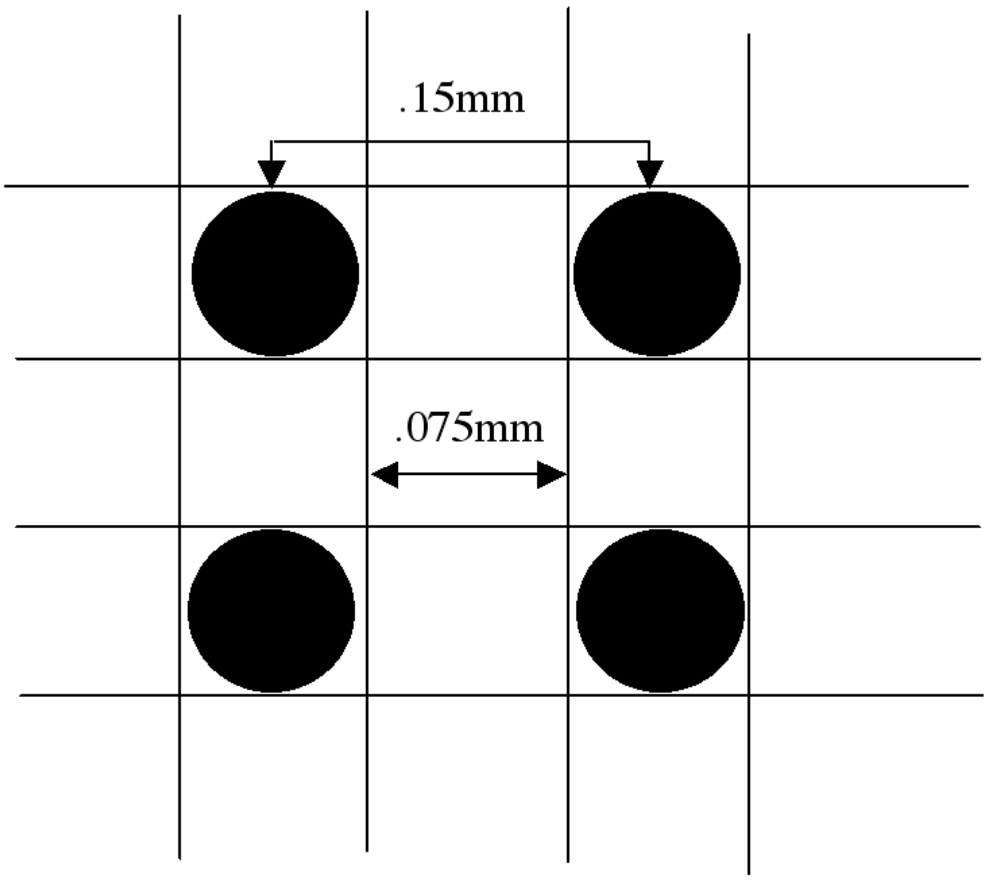
A bit of the process – one version

- the slide or chip is printed with a library of genes including those of special interest 
- collect mRNA under two different conditions. Using RT and two different fluorescent dies, samples of labeled (“red” and “green”) DNA are produced.
- incubate the samples with the slide under a cover slip.
- scan the result to measure the amount of red and green fluorescence at each spot to measure the relative amount of mRNA present in the two samples.

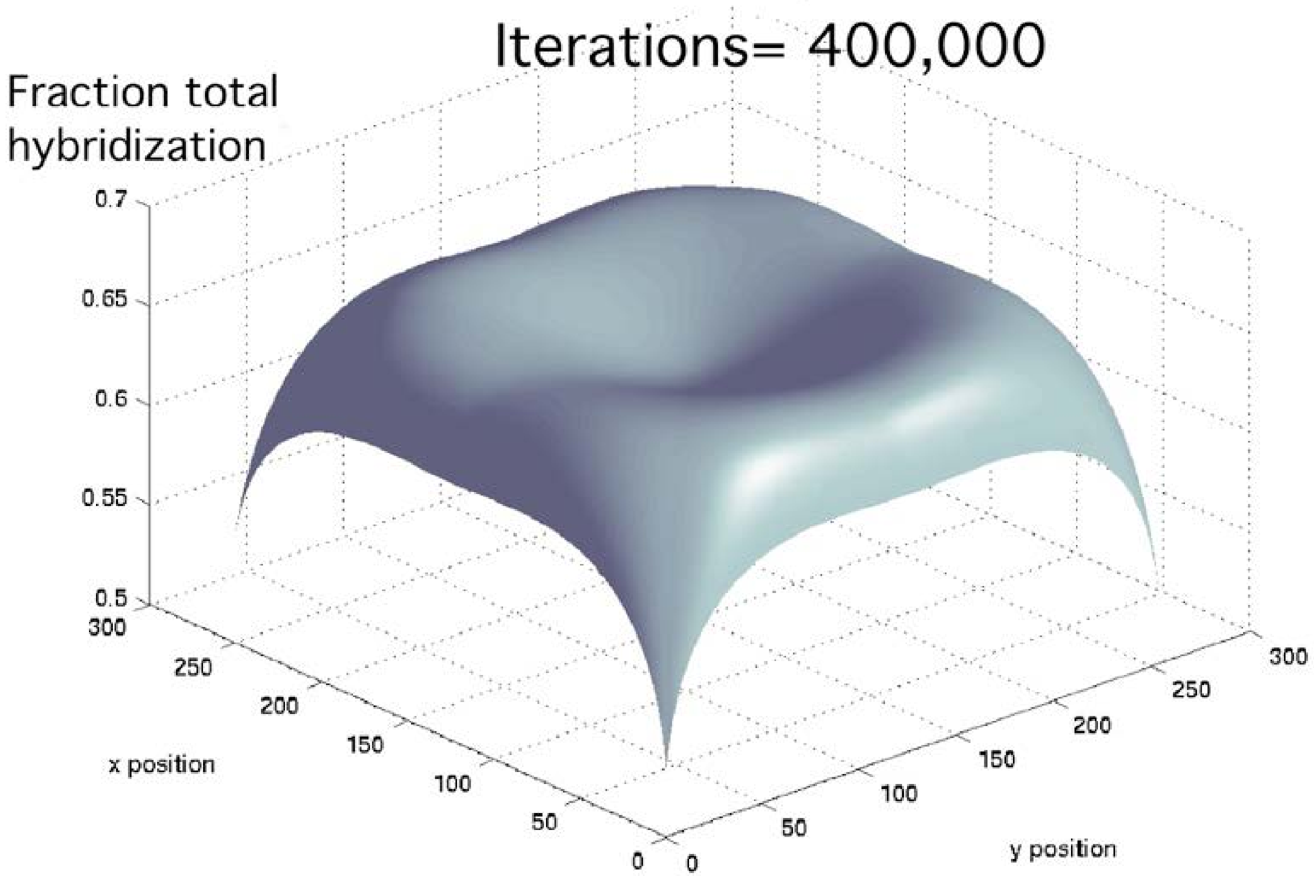
The Model of microarray hybridization

- Using the natural grid of positions on a slide, a Markov corresponding to each of the 16,000 dots is constructed. The goal being to compute the probability of absorption as a function of the transition number. Assume all dots are same.

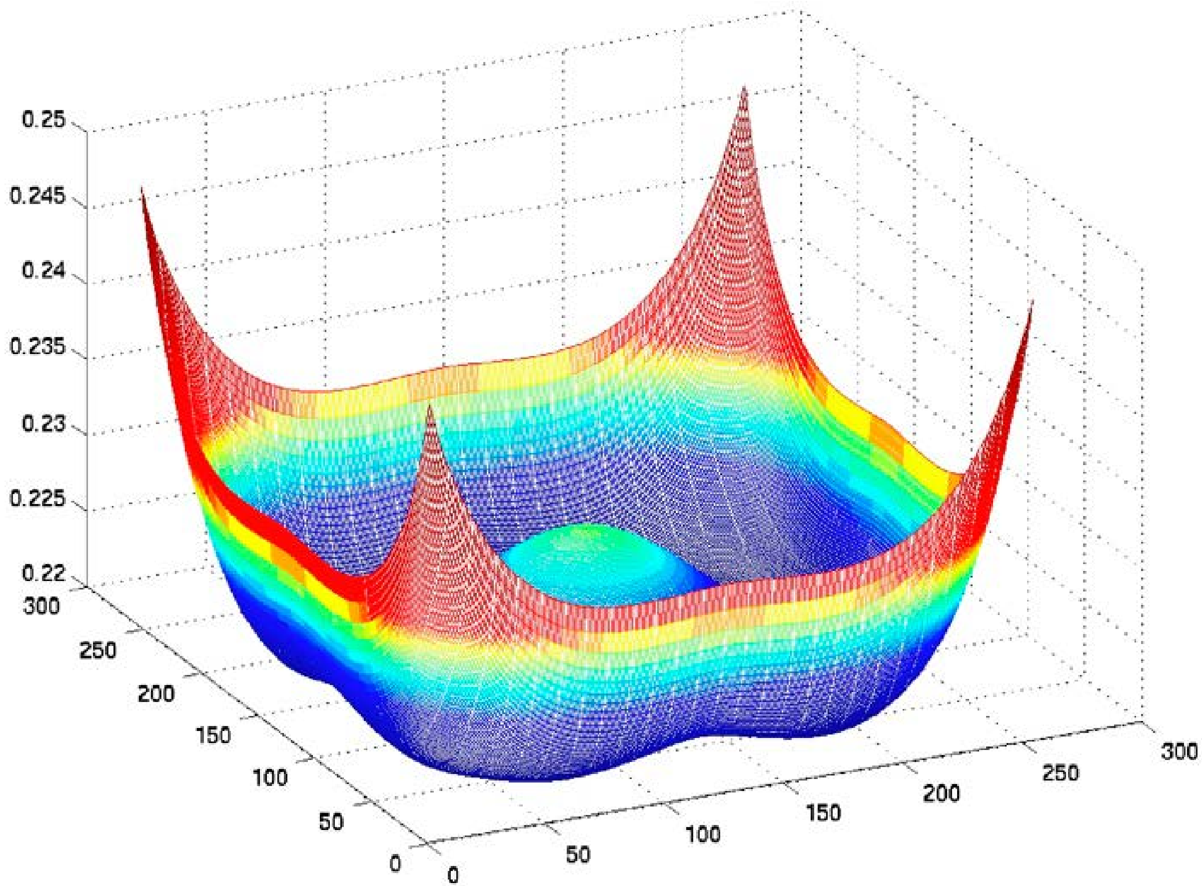
- The transition probabilities are based on the “taxi-cab” metric on the grid.



12 hours of hybridization



Variance after 12 hours



Conclusions

- The idea of “model” is complicated.
- Important aspects:
  1. Definition of the Real System (context of use of the model) – what does it cover and what does it not cover.
  2. Specific use – questions that the model is designed to address
  3. Tests – validation and refinement
  4. If the model is to be used over a period of time, software design principles should be used.
- Questions and nature of the system dictate the model form
- It is often a great surprise that (simple) models tell us anything about the real system
- The process of modeling (and solving problems in general) is a complicated one.

References

- S.J. Merrill & J.R. Cochran (1997) Markov chain methods in the analysis of heart rate variability, *Fields Institute Comm.* **11**:241-252.
- M. Adibuzzaman, G. Kramer, L. Galeotti, S. Merrill, D. Strauss, C.G. Scully, The mixing rate of the arterial blood pressure waveform Markov Chain is correlated with Shock Index during hemorrhage in anesthetized swine, Proceedings of Engineering in Medicine and Biology Society (EMBC), August 2014, 3268-3271.
- S.J. Merrill & B.M. Murphy (2002) Detection of autocatalytic dynamics in data modeled by a compartmental model, *Math. Biosci.* **180**:255-262.
- S.J. Merrill, S. Nelson, & C.A. Struble (2003) Spatial dependence of hybridization in the cDNA microarray, *Can. Appl. Math. Quart.* **11**: 321-337.
- S.J. Merrill (2005) The stochastic dance of early HIV infection, *J. Comp. Appl. Math.* **184**:242-257.
- B. Pandiyan, S.J. Merrill, and S. Benvenga, A patient-specific model of the negative-feedback control of the hypothalamus–pituitary–thyroid (HPT) axis in autoimmune (Hashimoto's) thyroiditis, *Mathematical Medicine and Biology* **31**(2013) 226-258.

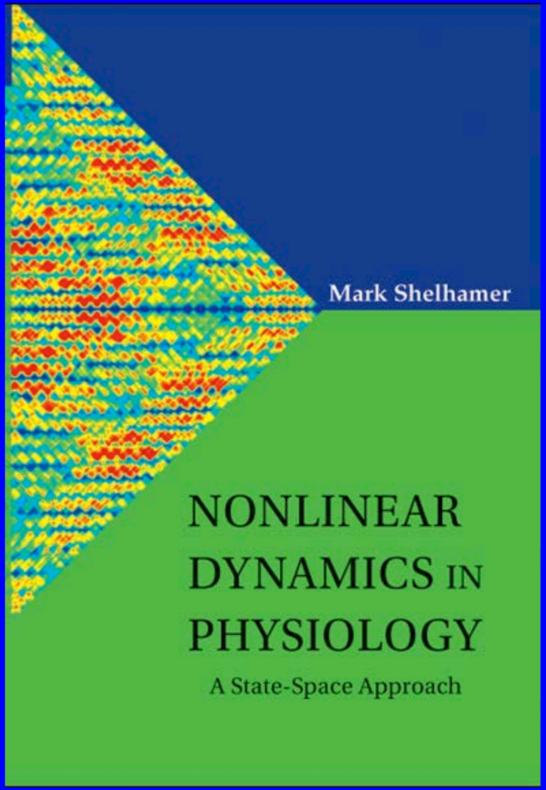
Introduction to Some NLD Techniques

Mark Shelhamer
Otolaryngology – Head & Neck Surgery
Biomedical Engineering
The Johns Hopkins University School of Medicine
mjs@dizzy.med.jhu.edu

Outline

- Descriptions of dynamic systems
  - Time, frequency, parametric
  - State variables, state space
- Time-delay reconstruction
  - Justification and benefits
  - Topology
  - Practical aspects
- Recurrence
- Dimensions
  - Definitions, examples
  - Topology, types of dimensions
  - Correlation dimension
  - Example, interpretation
- Surrogate data
  - Necessity, interpretation
- Physiological example
  - Reconstruction
  - Dimension
  - Surrogates

Figures from:



Dynamical Systems

- System whose behavior evolves over time
  - Typically subject to differential equations
- Random, deterministic
- Linear, nonlinear
- Behaviors
  - Random, fixed point, periodic, quasi-periodic, chaotic

Cross-correlation,Autocorrelation

- Correlation coefficient *r* is a measure of the degree of (linear) correlation between two random variables.
- Generalization to measure correlation between two signals, when one is shifted in time relative to the other
  - Cross-correlation function between two functions $x(t)$ and $y(t)$:

$$R_{xy}(\tau) = E[x(t)y(t+\tau)]$$

- Autocorrelation function:
  - Cross-correlation of signal $x(t)$ with itself:

$$R_{xx}(\tau) = E[x(t)x(t+\tau)]$$

Frequency Spectrum

- Fourier transform converts to frequency domain

$$X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-i\omega t} dt = \int_{-\infty}^{\infty} x(t) \exp(-i\omega t) dt$$

- Magnitude (power) spectrum

$$|X(\omega)| = \sqrt{\mathrm{Re}[X(\omega)]^2 + \mathrm{Im}[X(\omega)]^2}, \quad PSD = |X(\omega)|^2$$

- Phase spectrum

$$\angle X(\omega) = \mathrm{Im}[X(\omega)] / \mathrm{Re}[X(\omega)]$$

- Power spectrum is also the Fourier transform of the autocorrelation function
  - Contains no phase information

Time, Frequency Domain

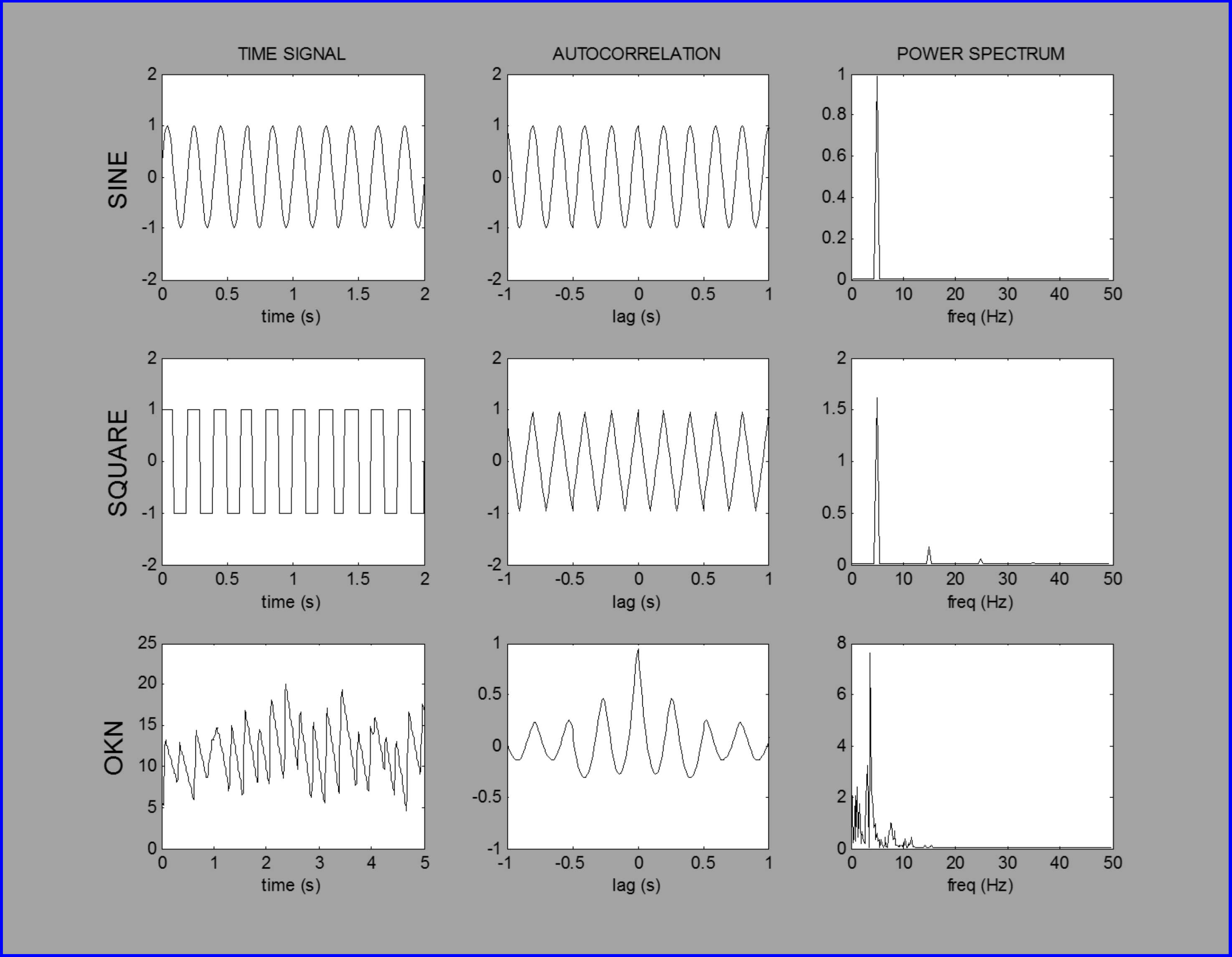


Figure 3.4.1a

Time, Frequency Domain

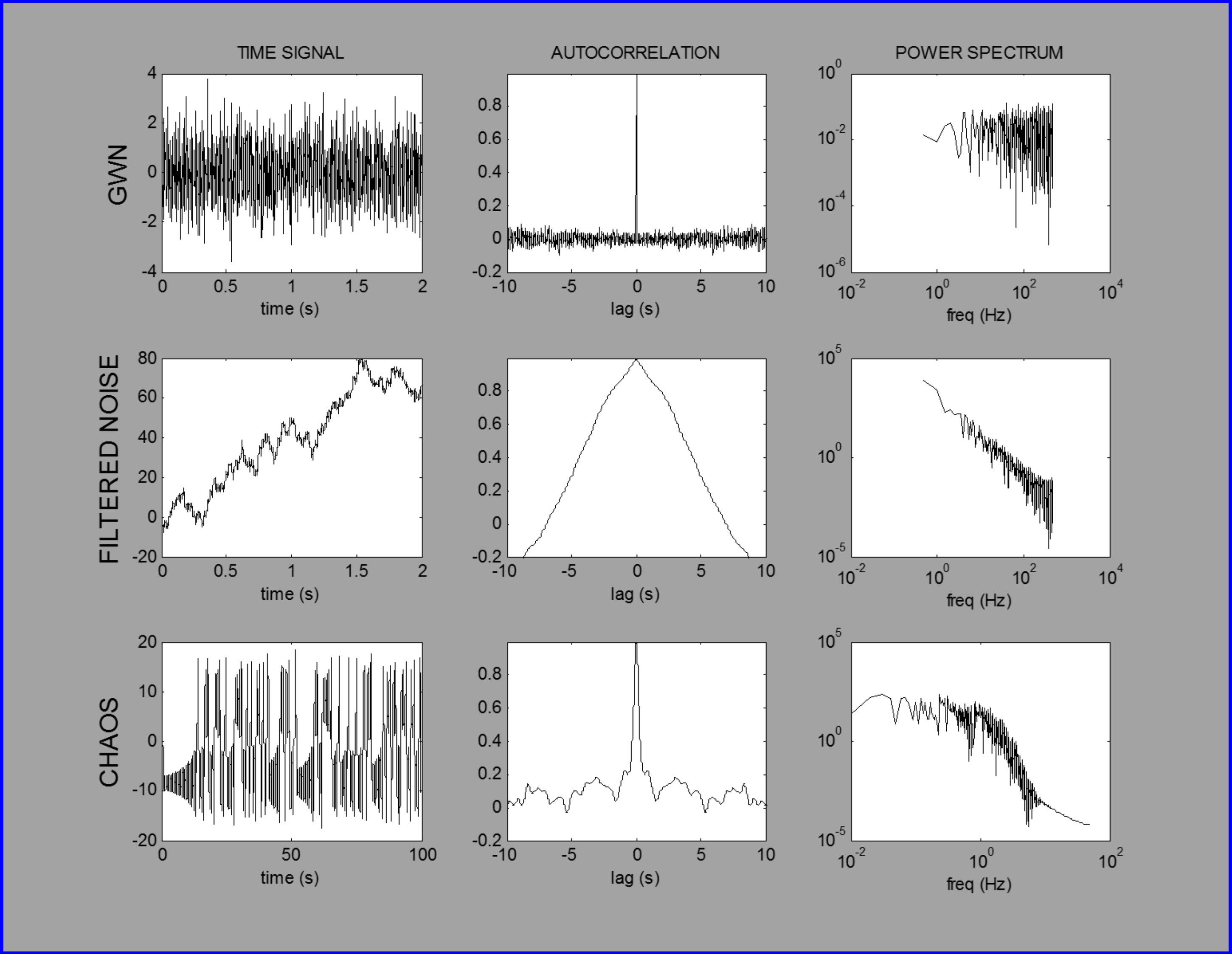
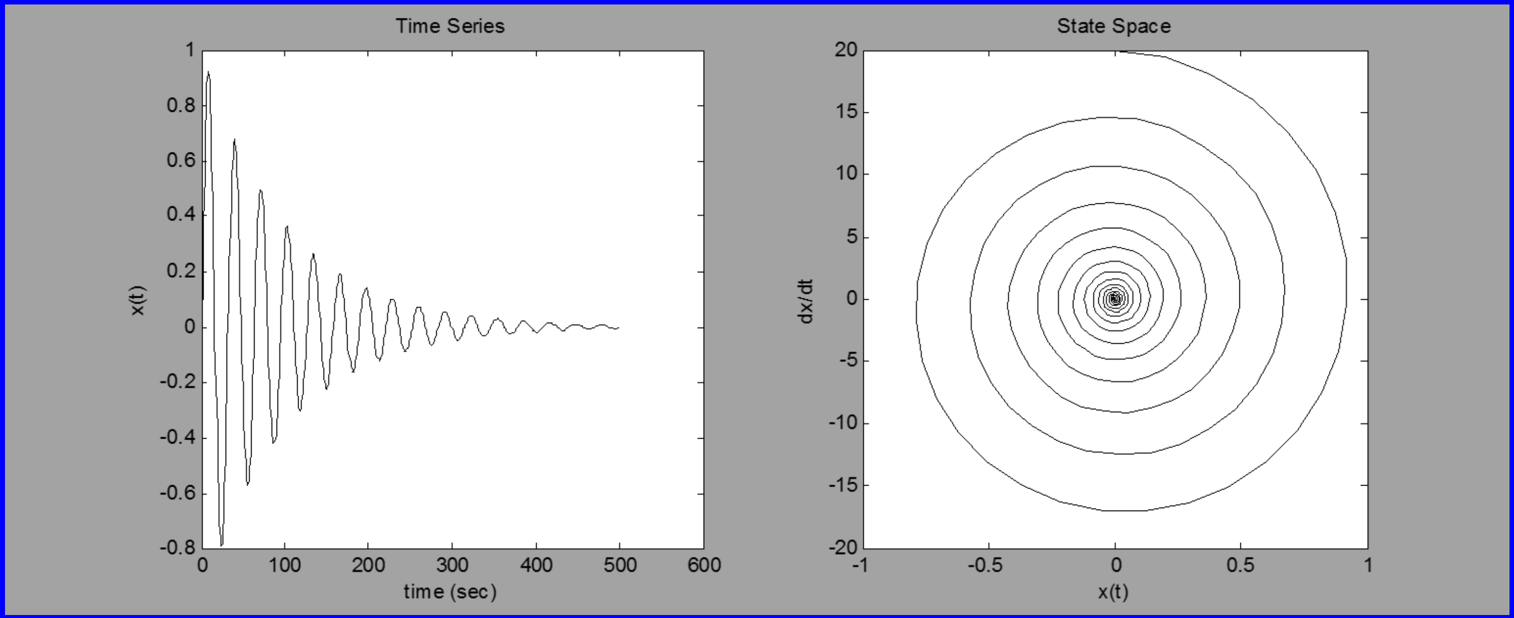


Figure 3.4.1b

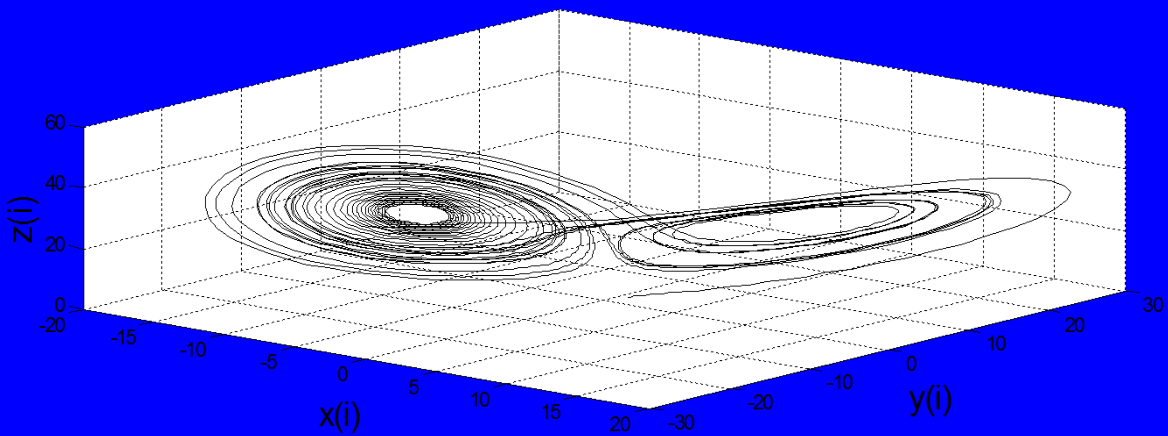
Phase Space Examples



Damped oscillator

Figure 4.1.1

Phase Space Examples



Lorenz attractor

Figure 4.7.1

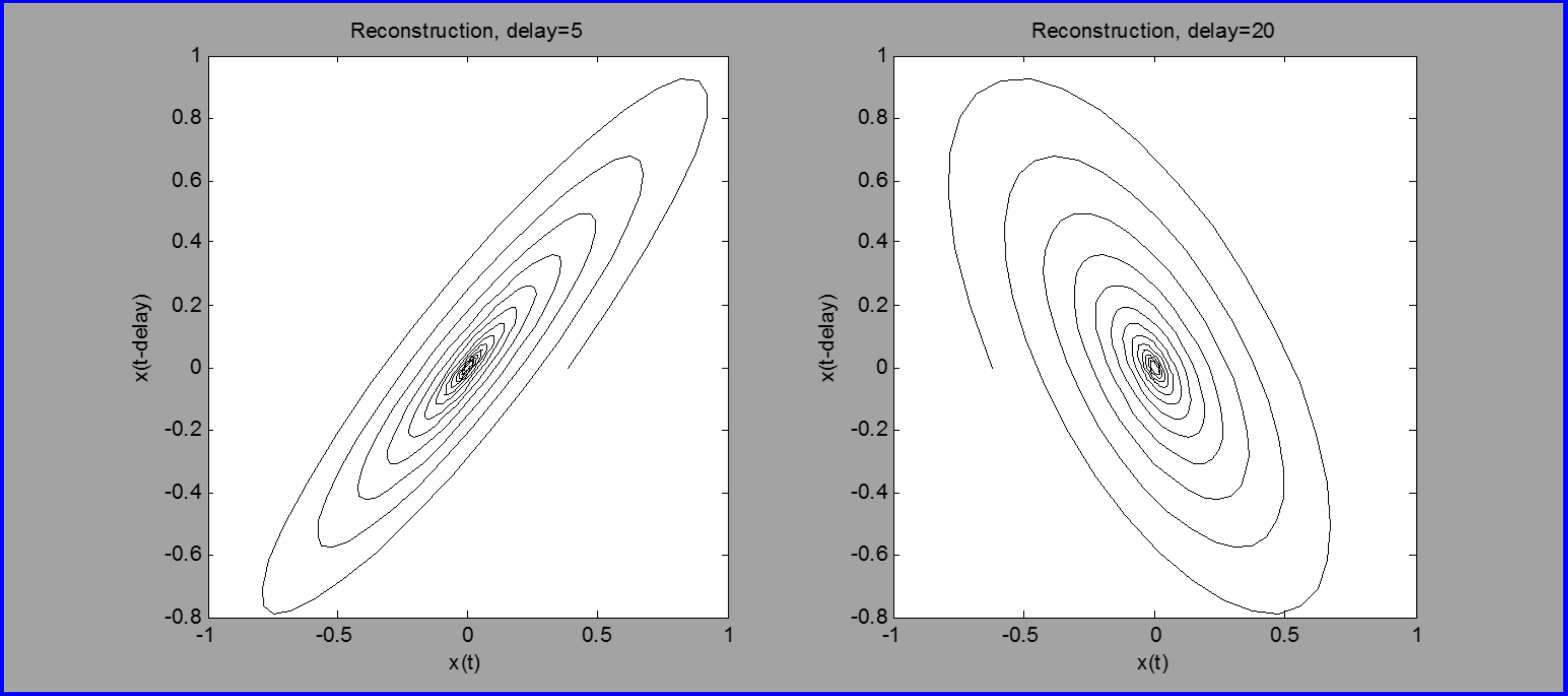
Time-delay reconstruction

- Advantages to phase-space description
  - Intuition
  - Utilize known information
  - Clear depiction of dynamical behaviors
- But what are the state variables?
  - What if they are unknown or not measurable?

Time-delay reconstruction

- Miracle cure: time-delay reconstruction
  - Do not have to know the state variables (but they must be dynamically related!)
  - Do not require high-order differentiation (position, velocity, acceleration,…)
  - Straightforward extension to high dimensions
- Each point in M-dimensional phase space is formed from M time-delayed values of signal x(t)
  - y(1) = [x(1) x(1+L) x(1+2L) … x(1+(M-1)L)]
  - y(2) = [x(2) x(2+L) x(2+2L) … x(2+(M-1)L)]

Time-delay reconstruction



Why does it work?

- Intuitive answer:
  - Time delay approximates differentiation

$$f' = \frac{df(x)}{dt} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

- Rigorous answer:
  - Topology

Example – Lorenz Attractor

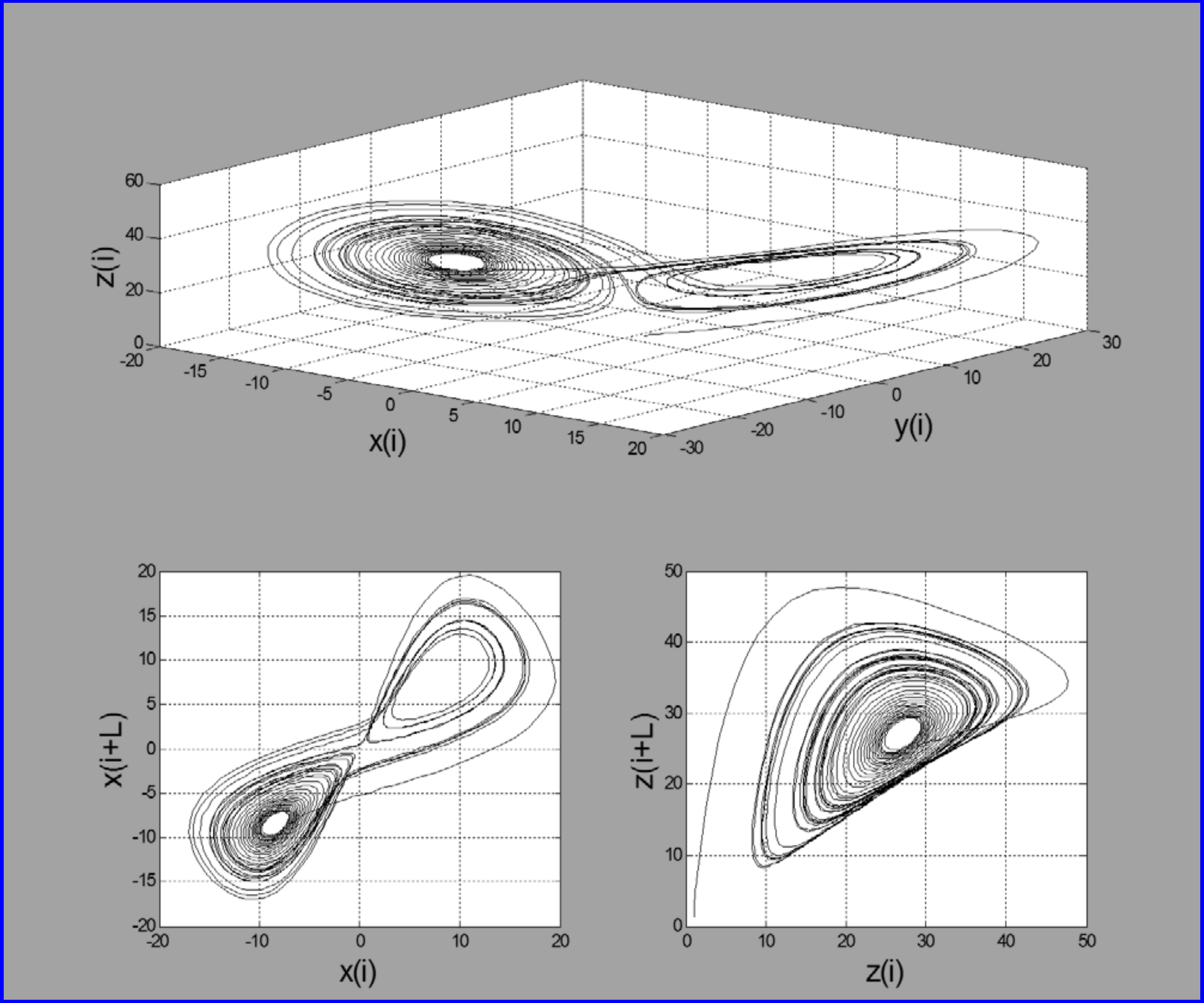


Figure 4.7.1

How to select parameters

- The dirty little secret (fudge factors): selection of embedding parameters
- Time delay L
  - Mutual information
  - Return time
  - Small multiple of correlation time
- Embedding dimension M
  - Saturation of dimension measurement
  - False nearest neighbors (FNN)

False Nearest Neighbors

- Concept:
  - If M is too small, attractor is not fully unfolded
    - Points not actually close together may appear close together (false neighbors)
- Approach:
  - Find nearest neighbors in dimension M
  - Find increase in distance between neighbors in dimension M+1
  - If distance increase is large, M is too small

False nearest neighbors

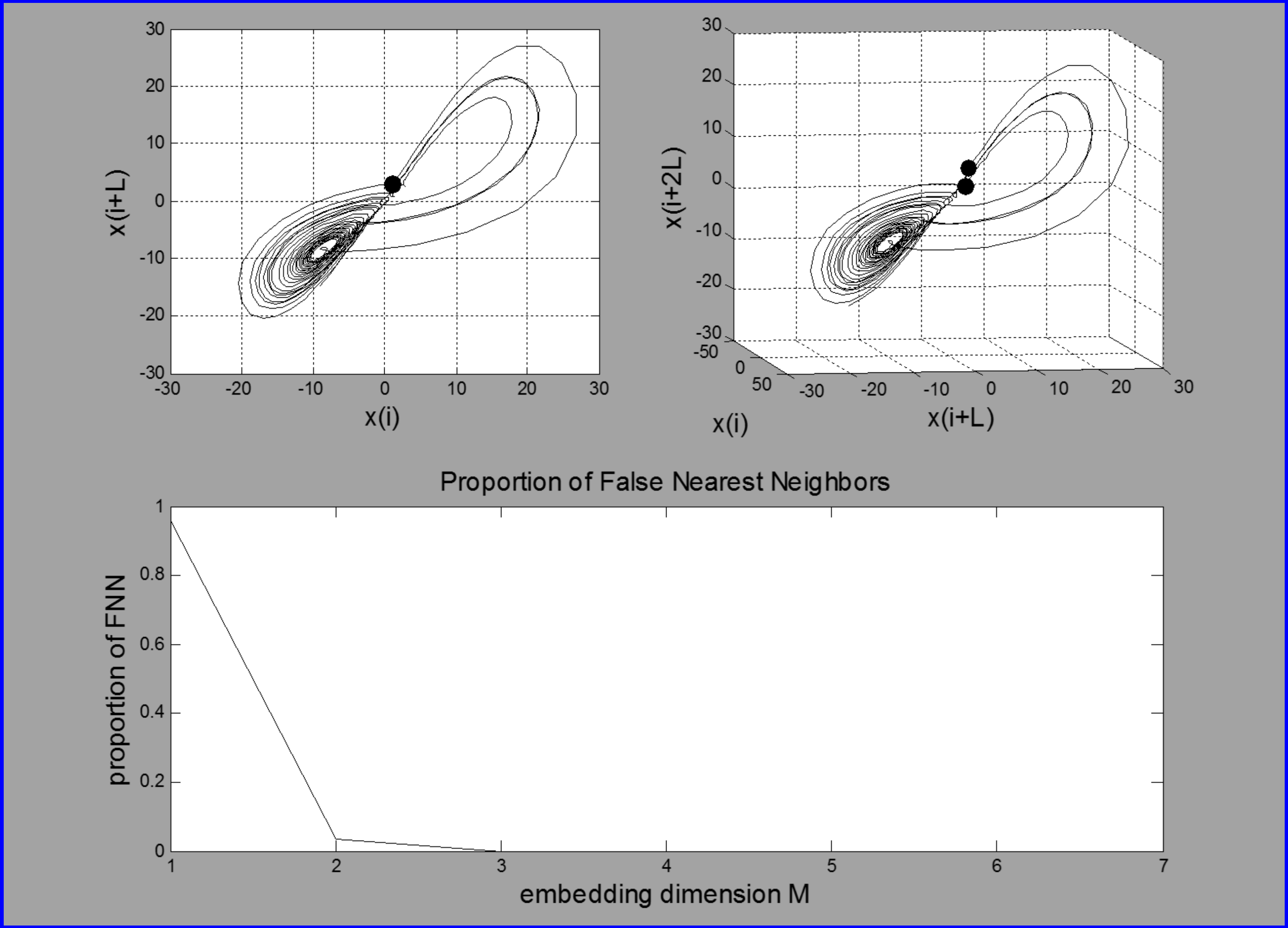


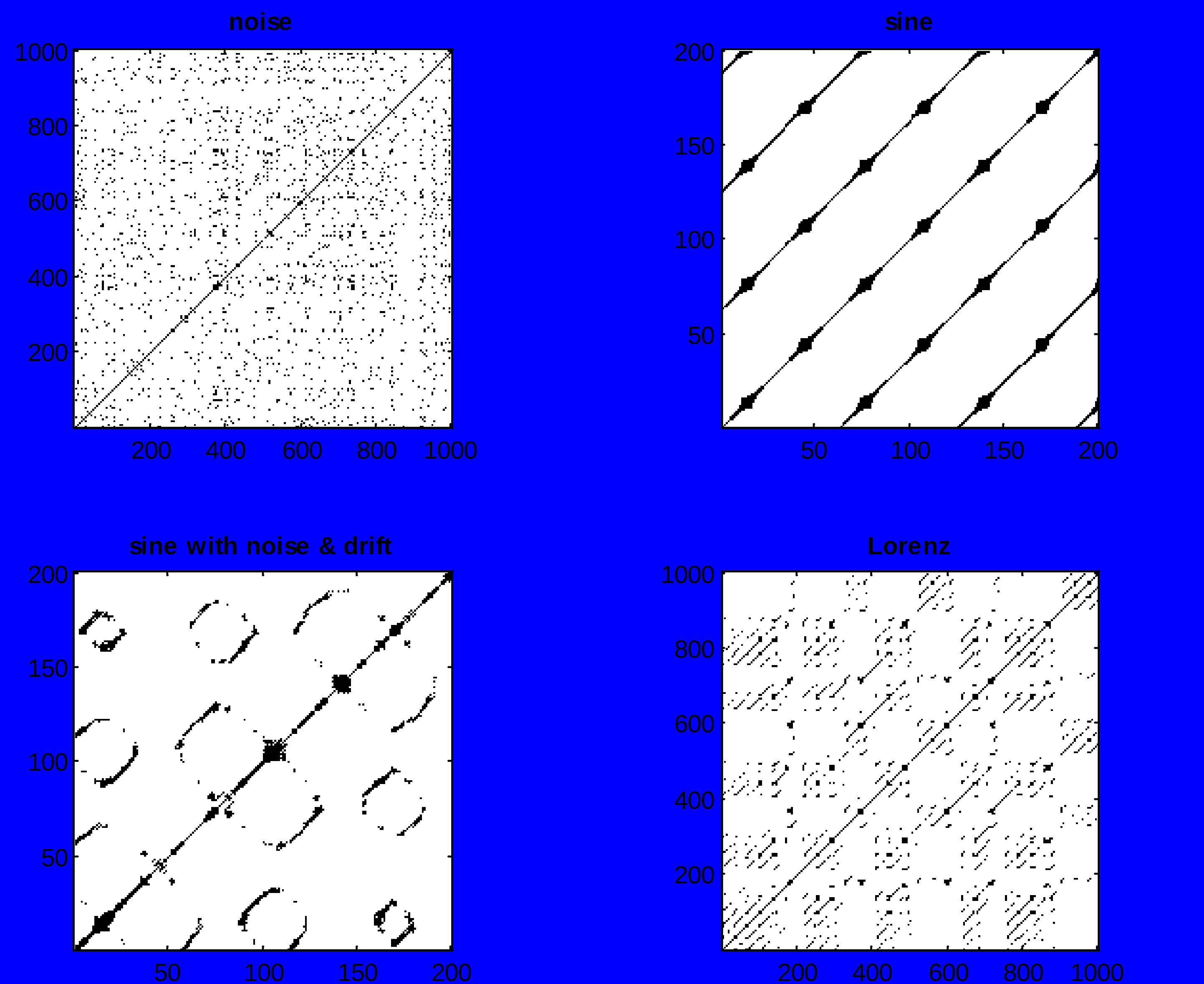
Figure 4.4.1

Recurrence Analysis

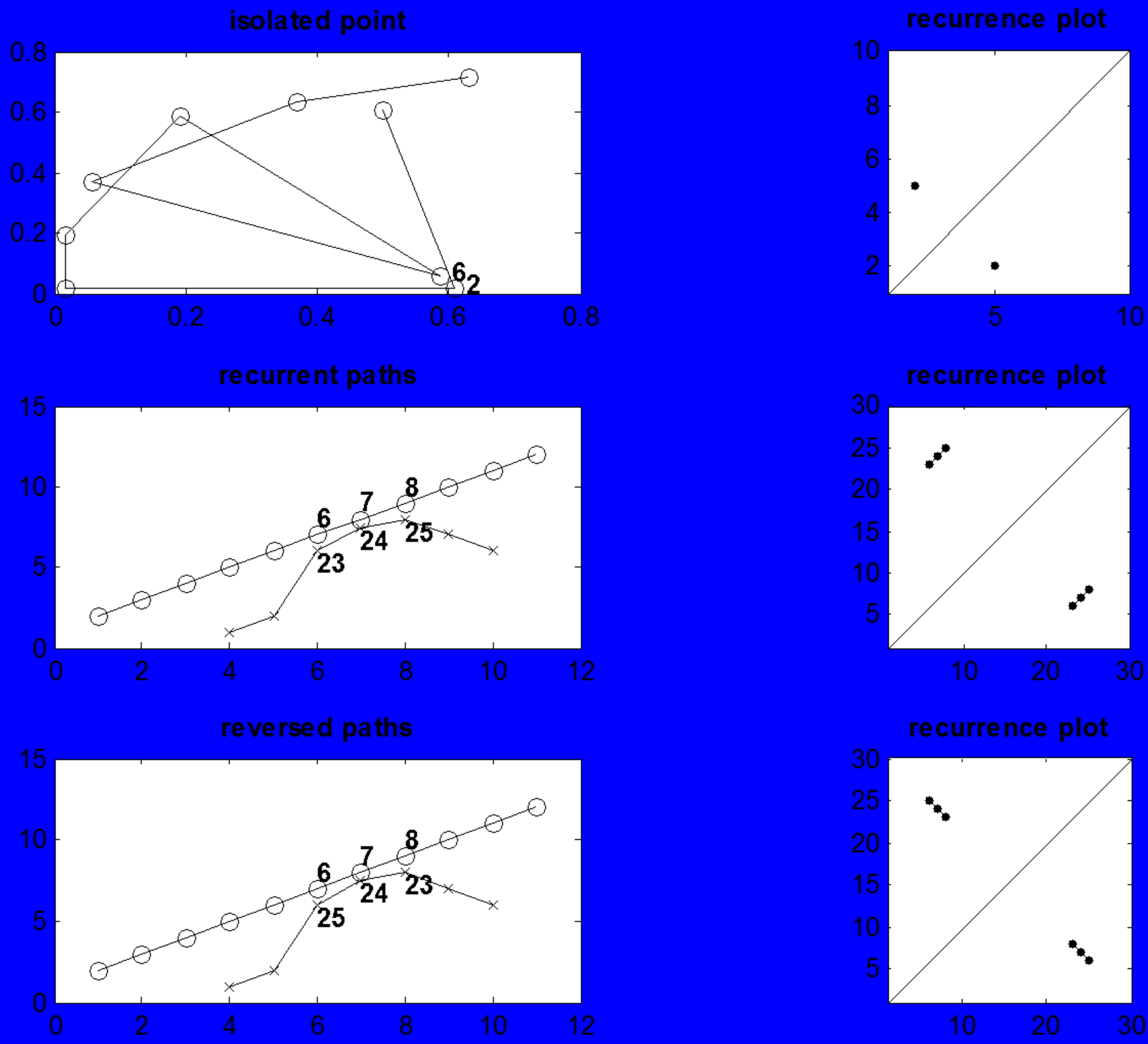
- Recurrence plot
  - graphical display of spatial correlations in an attractor
  - in terms of relative time
  - demonstrates periodicities, nonstationarity, and determinism

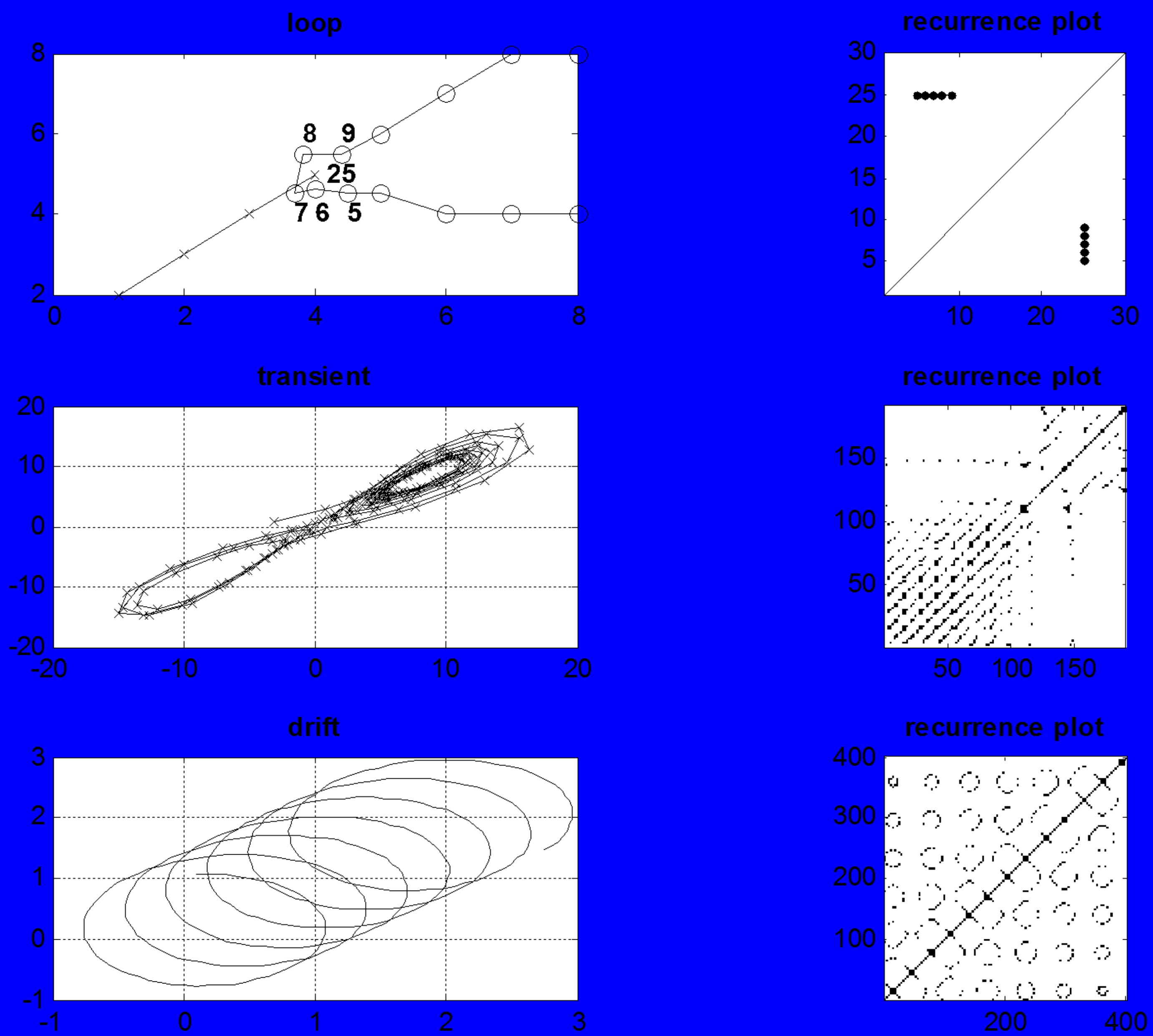
Recurrence Plot

- reconstruct the attractor reconstruction
- select a reference point $x(i)$
- center a ball of radius r on that point
- if $x(j)$ is within the ball (within distance r of $x(i)$)
  - place a dot at coordinates (i,j) on the recurrence plot



noise
sine
sine with noise & drift
Lorenz


isolated point
recurrence plot
recurrent paths
recurrence plot
reversed paths
recurrence plot
6
2
6
7
8
23
24
25
6
7
8
25
24
23
0
0.2
0.4
0.6
0.8
0
5
10
12
15
20
25
30


loop
8
9
25
7 6
5
recurrence plot
transient
recurrence plot
drift
recurrence plot

- Isolated recurrent points can arise at random.
- Periodic structure is evident in repeating patterns.
- Diagonal line segments (parallel to the main diagonal) represent recurring paths, as in a deterministic system.
- Diagonals orthogonal to the main diagonal represent nearby paths that move in opposite directions with time.
- For chaotic systems, line segments are short due to rapid divergence of trajectories.
- Horizontal and vertical lines result from a path that is recurrent for several consecutive points with a single point at some other time, as in looping around a point.
- Nonstationarity or drift is reflected in decreasing density away from the main diagonal.
- A transient results in recurrence gaps.

RQA

- Percent recurrence.
  - percentage of points that are defined as recurrent
  - periodic components should lead to more recurrent points
  - care must be exercised in using comparable attractor reconstructions (*M*, *L*, *etc*.) and recurrence plot constructions (in particular selection of the threshold distance)
- Percent determinism.
  - percentage of recurrent points that are in line segments parallel to the main diagonal
  - normalized representation of the number of recurrent points that are derived from trajectory paths that follow each other closely, as in a deterministic system
- Entropy.
  - measure of "complexity"
  - quantifies the distribution of the lengths of deterministic line segments (those parallel to the main diagonal)
  - Shannon entropy is defined as: $E = -\sum_i P_i \log_2(P_i)$
    - *P*i is probability of a line segment having a length that falls into bin *i* of a histogram of segment lengths
- Ratio.
  - ratio of percent determinism to percent recurrence.
- Trend
  - reduction in density away from the main diagonal, due to attractor drift

Dimensions

- Euclidean dimension
  - Number of independent coordinates needed to specify location
- Topological dimension
  - Point: dimension 0
  - Others: dimension is one greater than the dimension of (finite number) of other objects that can cut it into pieces

Topological Definitions

- Topological Dimension is a Topological Invariant
  - Not altered under a homeomorphic transformation

Dimension - Scaling

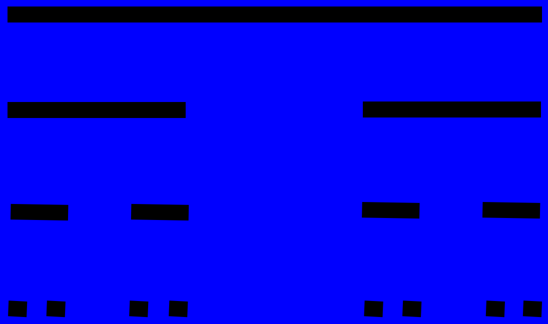
- Coastline of Britain
- Power-law scaling
  - $N \propto \varepsilon^{-D}$
  - $\log(N) = \log(K\varepsilon^{-D}) = \log(K) - D \log(\varepsilon)$
  - Plot of $\log(N)$ vs. $\log(\varepsilon) \rightarrow$ slope = D

Dimension - Scaling

- Line of length L
  - Cover N with segments of length ε
  - Let ε get very small
  - $N(\varepsilon) = L/\varepsilon = L\varepsilon^{-D} \propto \varepsilon^{-D}$, with D=1
- Plane of area A
  - Cover with segments of area ε^2
  - $N(\varepsilon) = A/\varepsilon^2 = A\varepsilon^{-D} \propto \varepsilon^{-D}$, with D=2

Dimension - Scaling

- Middle-thirds Cantor set



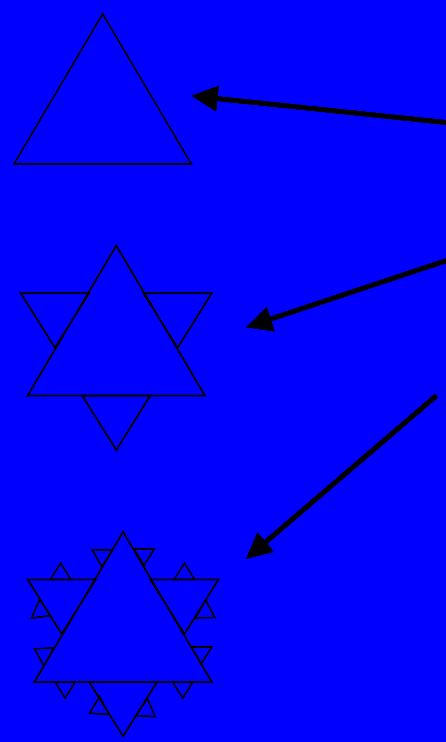
step (n)	number of segments	segment length	total length
0	1	1	1
1	2	1/3	2/3
2	4	1/9	4/9
3	8	1/27	8/27
N	2^N	$(1/3)^N$	$(2/3)^N$
∞	∞	0	0

- D = 0.6309…

Figure 5.2.1

Dimension - Scaling

- Koch curve



step (n)	number of segments	segment length	total length
0	3	1	3
1	12	1/3	12/3
2	48	1/9	48/9
3	192	1/27	192/27
N	3×4^N	$(1/3)^N$	$3 \times (4/3)^N$
∞	∞	0	∞

- D = 1.2619…

Box-counting Dimension

- Number of ε-boxes needed to cover object

$$N(\varepsilon) \propto (1/\varepsilon)^D \quad as \;\; \varepsilon \to 0$$

$$\log(N) = \log(k) + D\log(1/\varepsilon) \quad as \;\; \varepsilon \to 0$$

$$D = \frac{\log(N) - \log(k)}{\log(1/\varepsilon)} \quad as \;\; \varepsilon \to 0$$

$$D = \lim_{\varepsilon \to 0} \frac{\log(N) - \log(k)}{\log(1/\varepsilon)} = \lim_{\varepsilon \to 0} \frac{\log(N)}{\log(1/\varepsilon)}$$

Spectrum of Dimensions

- Box-counting dimension
  - Based on number of small boxes visited by attractor
- Information dimension
  - Based on length of time attractor visits a given box
- Correlation dimension
  - Based on distances between attractor points

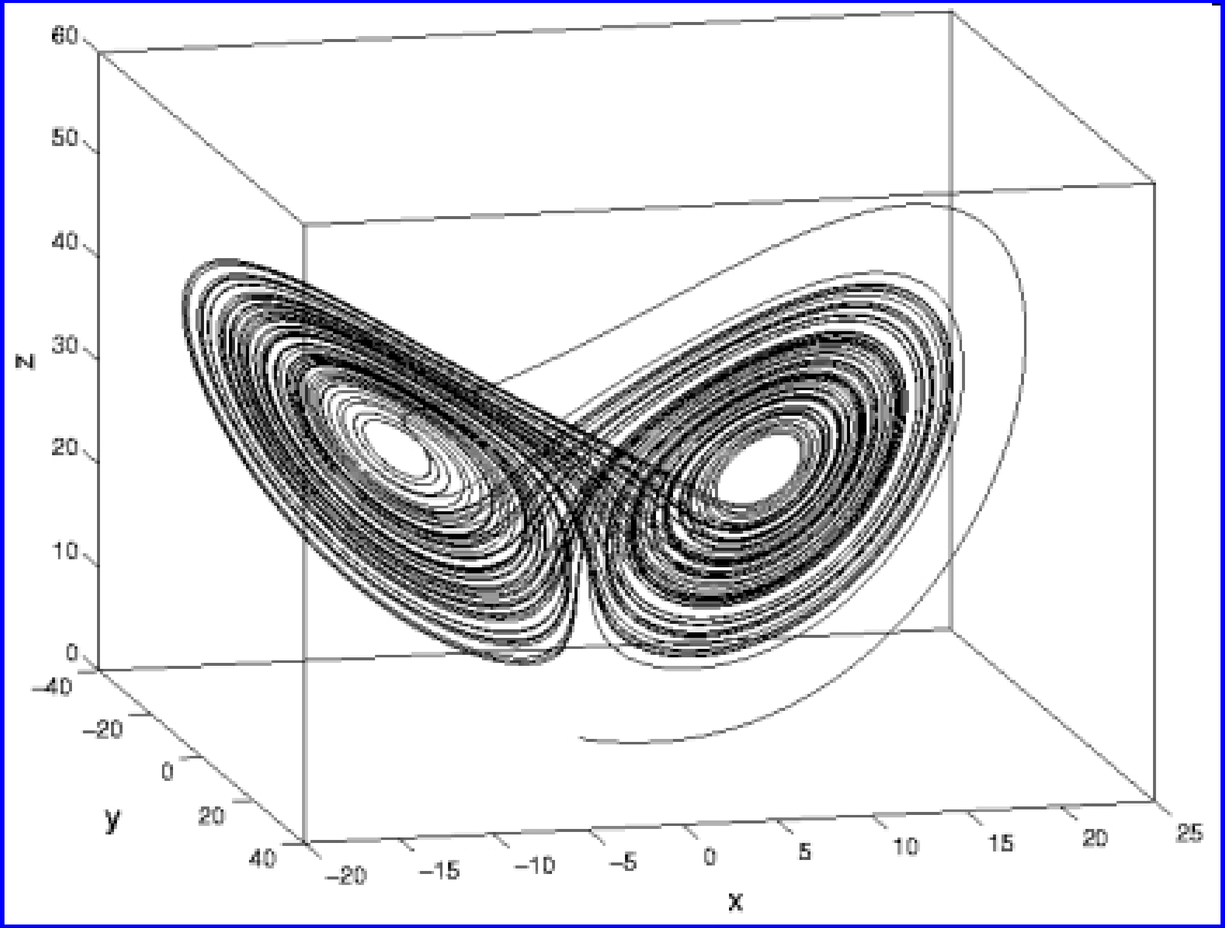
Computing D_c

- Find slope of correlation integral C(r)

$$D_2 = \lim_{r \to 0} \frac{\log[C(r)]}{\log(r)}$$

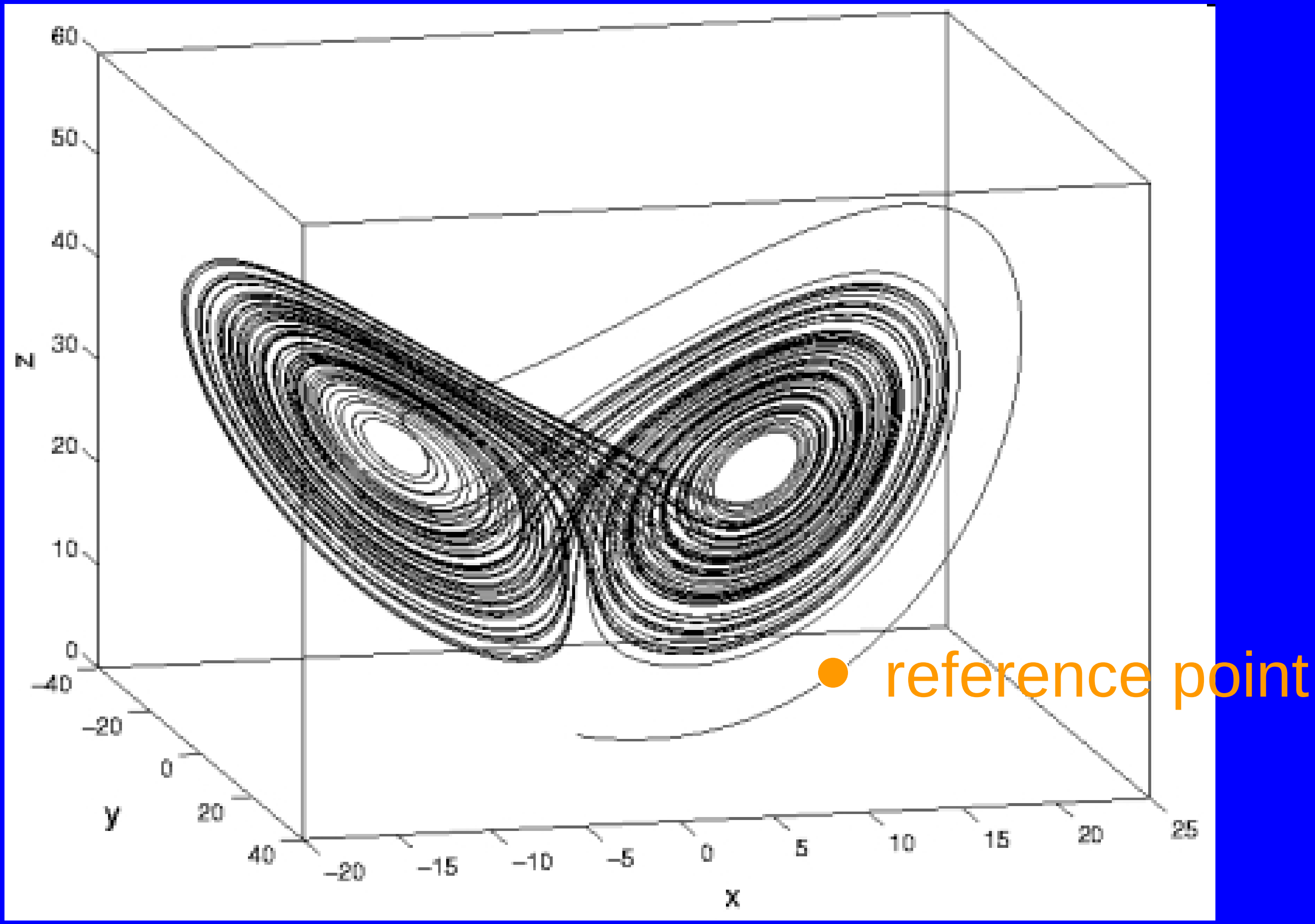
$$C(r) = \frac{1}{N(N-1)} \sum_i \sum_j \mathbf{U}(r, |y_i - y_j|) \quad (i \neq j)$$

Computing D_c



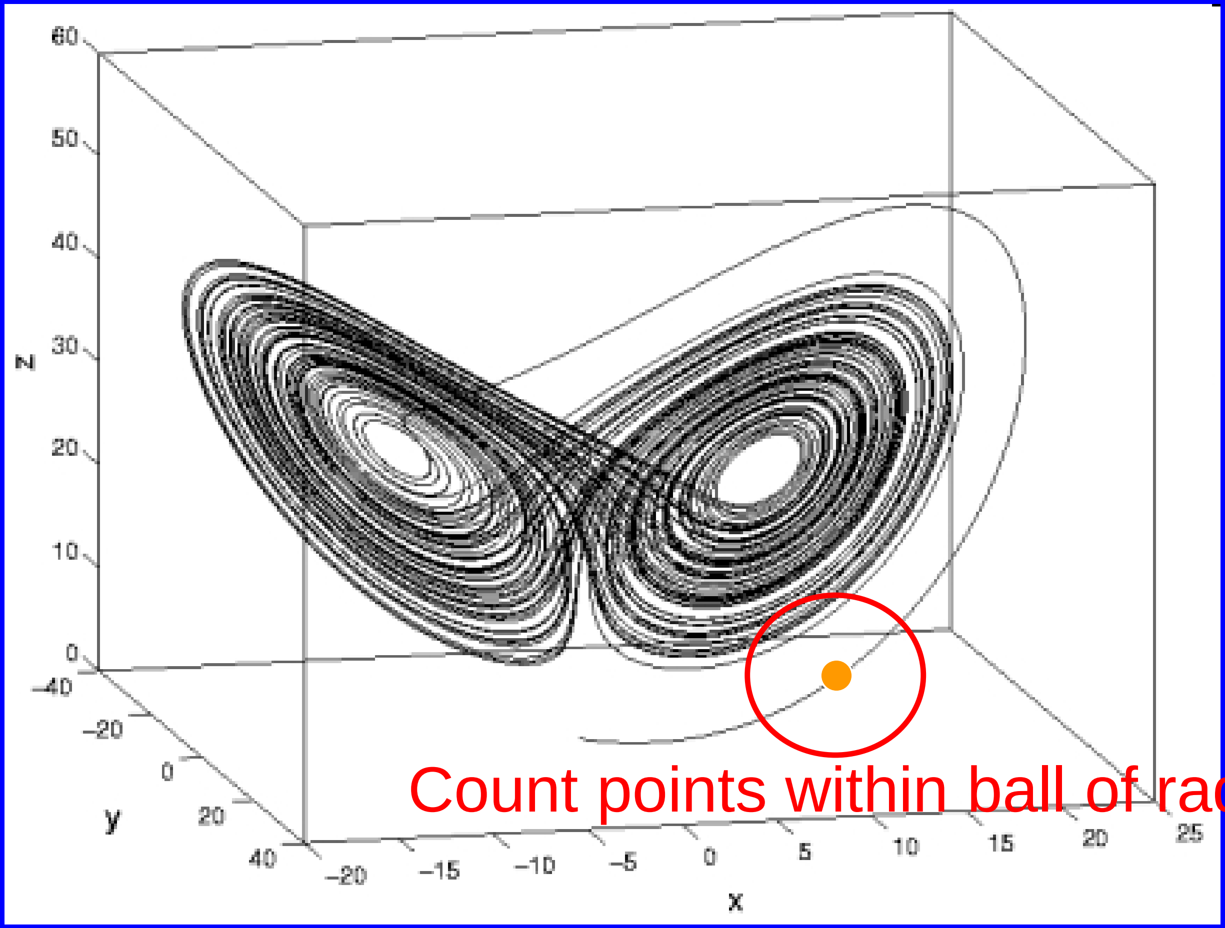
http://complex.upf.es/~josep/Chaos.html

Computing D_c



http://complex.upf.es/~josep/Chaos.html

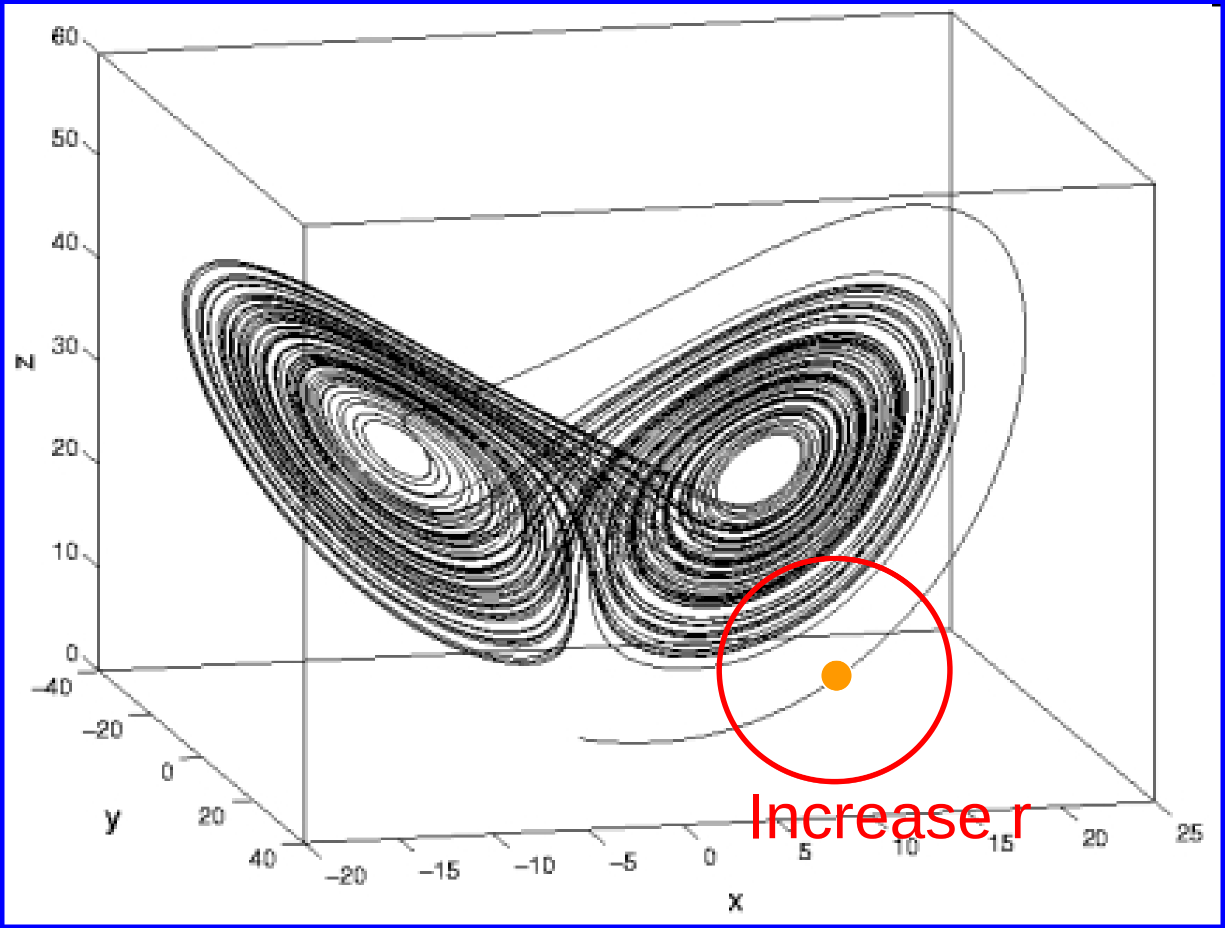
Computing D_c



Count points within ball of radius r

http://complex.upf.es/~josep/Chaos.html

Computing D_c



Correlation Integral = C(r) = sum of pairs of points within distance r

http://complex.upf.es/~josep/Chaos.html

Recipe for Computing D_c

- $C(r)$ increases as a power-law function of r.
- $C(r)$ versus r on log-log plot is straight line.

- Reconstruct attractor in M-dimensional embedding space.
- Choose reference distance r.
- Select a *reference point y(i)*.
- For every other point *y(j)*, find distance between this point and reference point
- If distance < r, add 1to the correlation integral.
- Choose next point as reference, and repeat.
- Divide the summation, the accumulated number of point pairs that are closer than r, by $N(N-1)$.
- Repeat for another reference distance r.
- Plot $\log[C(r)]$ versus $\log(r)$.
- Slope is the dimension D_c.

Example - Lorenz

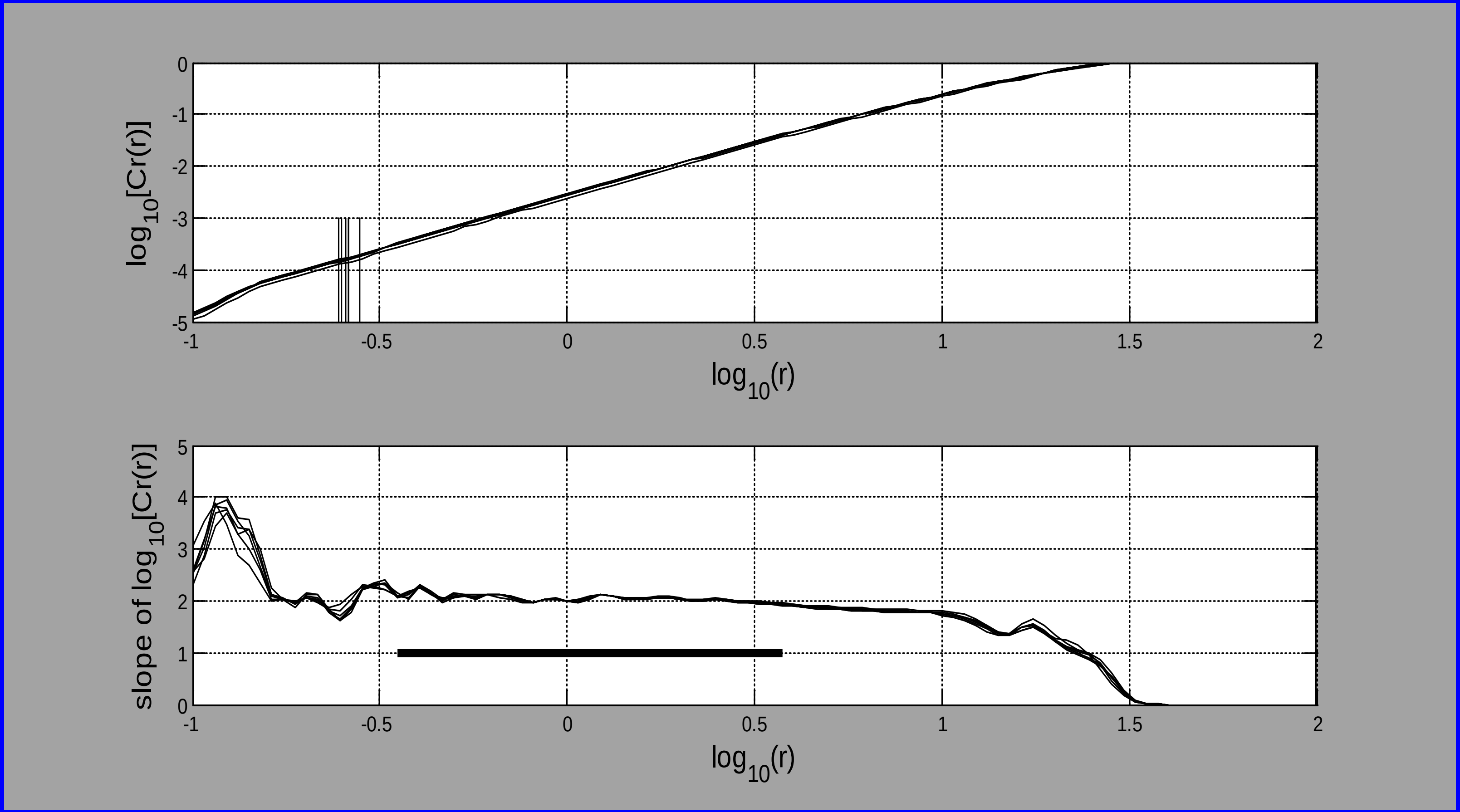


Figure 5.4.1

Example - Lorenz

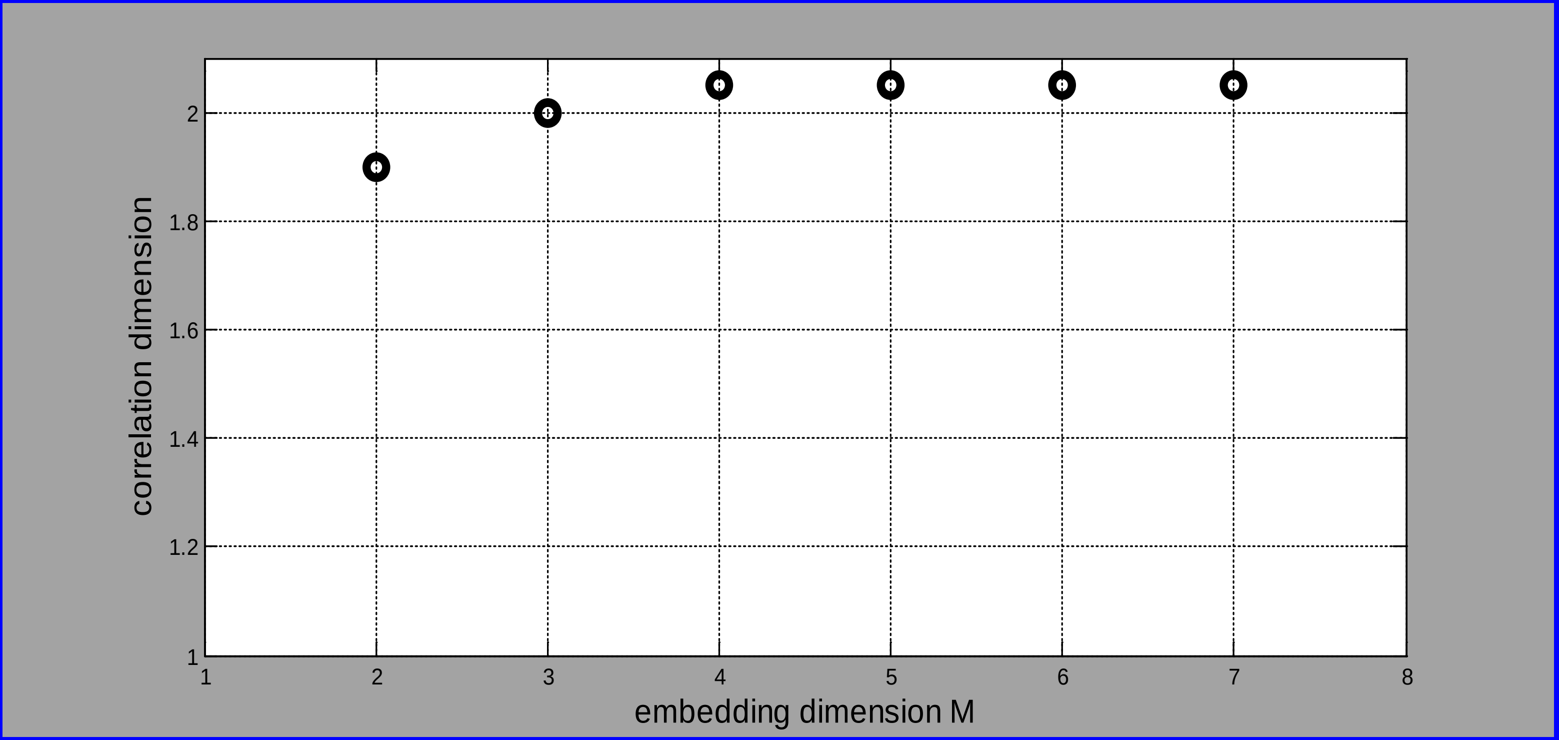


Figure 5.4.2

Interpretation

- Number of state variables
- Complexity
- Useful as a comparative measure

Error Bars for Dimensions

- How reliable is the dimension estimate?
- Can we put error bars on it?
  - Need to know statistics of the data, propagate the error through the computations
  - Not feasible
- Alternate approach
  - Surrogate data
  - Generate data sets from the original data, based on a null hypothesis
  - Place error bars on surrogates
  - Dimension estimate from original data should be "significantly different" from that of surrogates to reject null hypothesis

Surrogates as a Form of Statistical Hypothesis Testing

- Surrogates are *random* signals
  - Generated by randomizing some aspect of the original signal
  - Noise has infinite dimension
  - ➔ Surrogates should have increased dimensions
- How confident are we that the surrogates are different from the original?
  - Generate *n*-1 surrogate signals (total of *n* dimension values)
  - *n*! ways to arrange *n* dimension values
  - In 1/*n* arrangements, original dimension will be less than all surrogates
    - Totally by chance, original dimension will be less than all others with probability 1/*n*
  - If original dimension is less than all others, this could have happened by chance with probability 1/*n*
    - ➔ statistical significance level
  - Example: 39 surrogates, *n*=40, significance level is 1/40=0.025
  - This is a one-tailed test (expect original dimension < surrogates)
    - If we just want original dimension to be outside the range of the surrogates, significance level is doubled.

Random Surrogates

- Null hypothesis: data come from random system with same distribution of values
  - Shuffle the data
- Null hypothesis: data come from Gaussian random system with same first-order statistics
  - Draw values from Gaussian pdf

Phase-Randomization Surrogate

- Null hypothesis: data come from random system with same autocorrelation (linear correlations, power spectrum)
  - Test for nonlinearity
  - Autocorrelation: linear correlation structure
  - Power spectrum = Fourier transform of autocorrelation
  - Randomize signal but retain *linear correlations*
    - Shuffle the phase spectrum of the data

AAFT Surrogate

Amplitude Adjusted Fourier Transform

- Null hypothesis: data come from random system with same autocorrelation as the original (linear correlations, power spectrum) passed through a *static monotonic* nonlinearity

$$y(t) \xrightarrow{\quad h(\cdot) \quad} x(t)$$

AAFT Surrogate

$$y(t) \xrightarrow{\quad h(\cdot) \quad} x(t)$$

- Note that $h(\cdot)$ does not change the rank ordering of $y(t)$
  - $y(t)$ "follows" $x(t)$ in amplitude
  - If $x(t)$ is generated by a Gaussian signal through $h(\cdot)$, as hypothesized, then so will the surrogate $x'(t)$
    - ➔$x'(t)$ has same amplitude distribution as $x(t)$
    - ➔we can generate $x'(t)$ by reordering $x(t)$

--- But what determines the new ordering? ---

AAFT Surrogate

$$y(t) \xrightarrow{\quad h(\cdot) \quad} x(t)$$

- New ordering of $x'(t)$ is determined by a phase-randomized Gaussian signal $y'(t)$
  - (since this is the hypothesized underlying signal that produced $x(t)$)

- Therefore need to create phase-randomized signals $y'(t)$, which have same autocorrelation as $x(t)$ before it went through $h(\cdot)$.
- Get this by reordering a GWN signal so that it "follows" $x(t)$.
  - This represents $x(t)$ after going through $h(\cdot)$:
    - It's Gaussian (as is underlying $y(t)$)
    - It follows $x(t)$ and can be considered as $x(t)$ passed through a static monotonic function (the inverse of $h(\cdot)$)

AAFT Surrogate

the gory details

1. Generate Gaussian signal (GWN) $y(t)$.
2. Reorder the values of $y(t)$ to match the rank order of $x(t)$. This produces $y_R(t)$, which "follows" $x(t)$ but is Gaussian. This is meant to represent the hypothesized "original" signal, call it $x_{pre}(t)$, *before it was rescaled by h to become x(t).*
3. At this point, $y_R(t)=x_{pre}(t)=h^{-1}[x(t)]$.
4. Now take the Fourier transform of the signal $x_{pre}(t)$ or $y_R(t)$, randomize the phases, and take the inverse Fourier transform to get a random time series $y'(t)$. This can be done any number of times to generate as many surrogates as desired. Each $y'(t)$ is a random signal that represents a Gaussian linearly correlated time series before passing through $h(\cdot)$. (Note that yR(t) and $y'(t)$ are linearly correlated Gaussian signals with identical autocorrelation functions and power spectra but different phase spectra.)
5. Finally, reorder the values of $x(t)$ so that they follow the rank order of the values of $y'(t)$. This produces the surrogate $x'(t)$, which has the same amplitude distribution as the original $x(t)$ but mimics passing a linear Gaussian signal through $h(\cdot)$.

Pseudo-periodic Surrogate

- Null hypothesis: data can be modeled as a periodic process driven by uncorrelated noise (may change periodicity from cycle to cycle)
  - Pseudo-periodic surrogate
  - Generate *M*-dimensional surrogate *attractor:*
    - match original attractor in large scale (periodic and near-periodic orbits)
    - disrupt small-scale structure with noise

Pseudo-periodic Surrogate

1. Randomly select starting point on the original attractor
2. Find a close neighbor to that point
3. Project neighbor one step ahead, this is next point on surrogate attractor
4. Repeat

- Maintains large-scale flow
- Disrupts correlations between nearby points

Multivariate Surrogate

- Null hypothesis:
  - Each signal can be modeled as linearly correlated Gaussian noise (autocorrelation function)
  - Only linear correlations between the signals (described by cross-correlation functions)
- Retain linear correlations *within* $x_i(t)$
  - Retain power spectrum
  - Randomize phases in frequency spectrum
- Retain linear correlations *between* $x_i(t)$ and $x_j(t)$
  - Retain the cross-spectrum for all pairs (i,j)
- Add the *same* random set of phases to the phase spectrum of each of the $x_i(t)$

Multivariate Surrogate

Why does this work?

- Power spectrum $A_i(f)$
- Phase spectrum $\phi_i(f)$
- Cross-spectrum:

$$|S_{x_i x_j}(f)| = A_i(f) A_j(f)$$

$$\angle S_{x_i x_j}(f) = \phi_j(f) - \phi_i(f).$$

Surrogate Examples – Time Domain

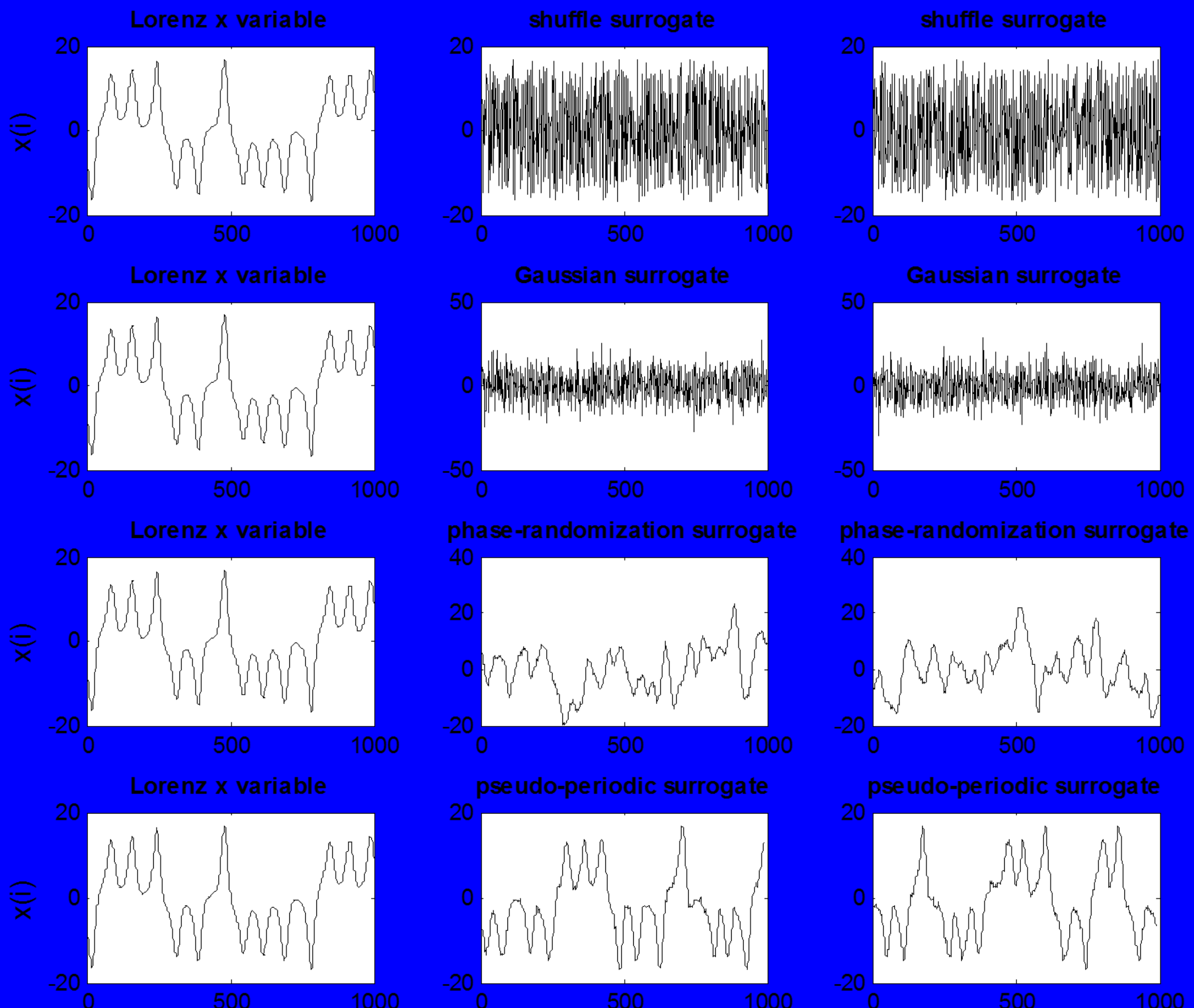
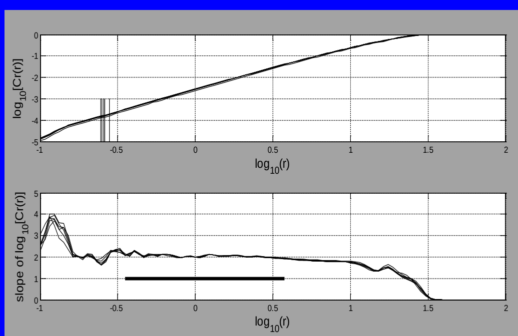


Figure 6.11.1

Surrogate data

Surrogate Examples – Correlation Integral



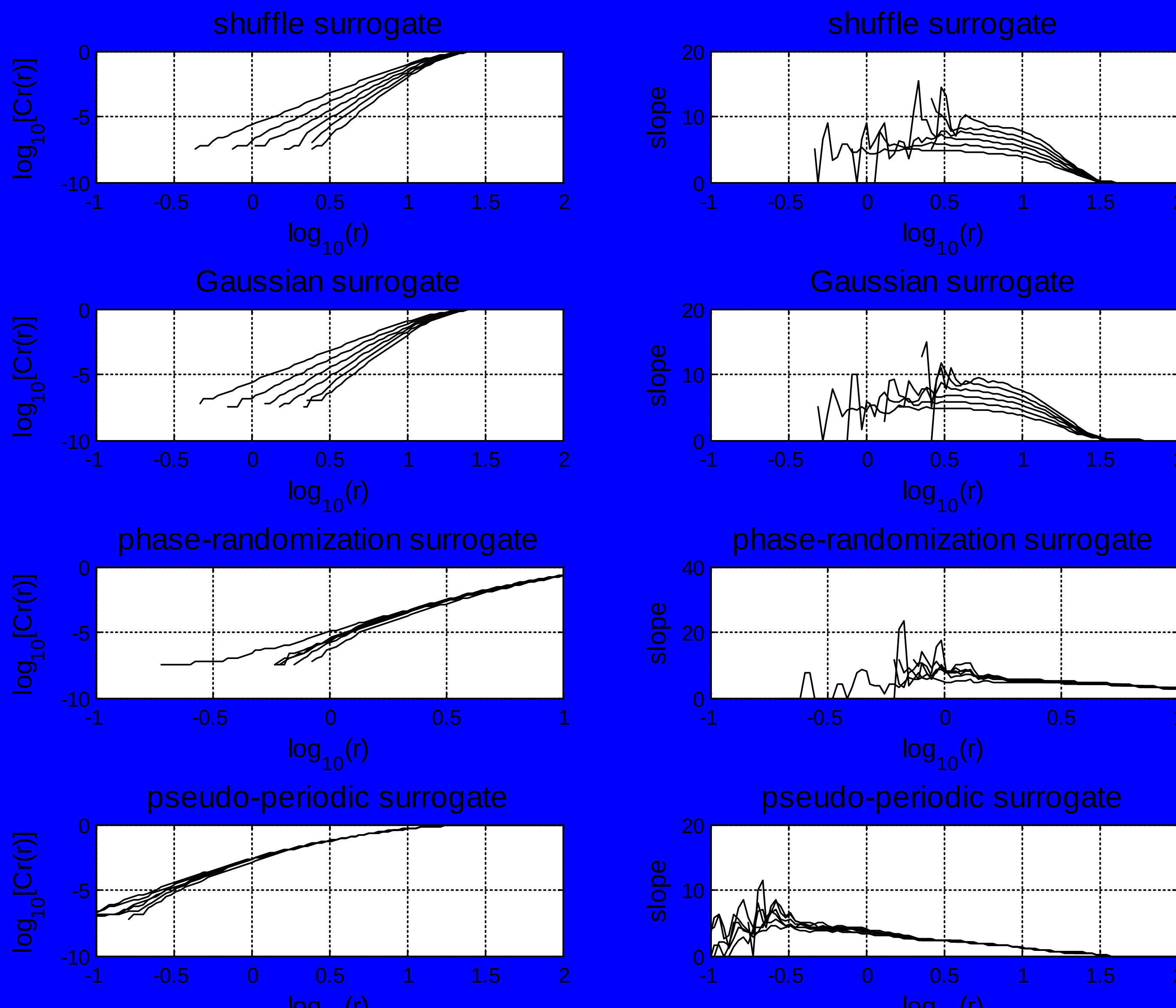


Figure 6.11.2

Eye-movement example

Example - OKN

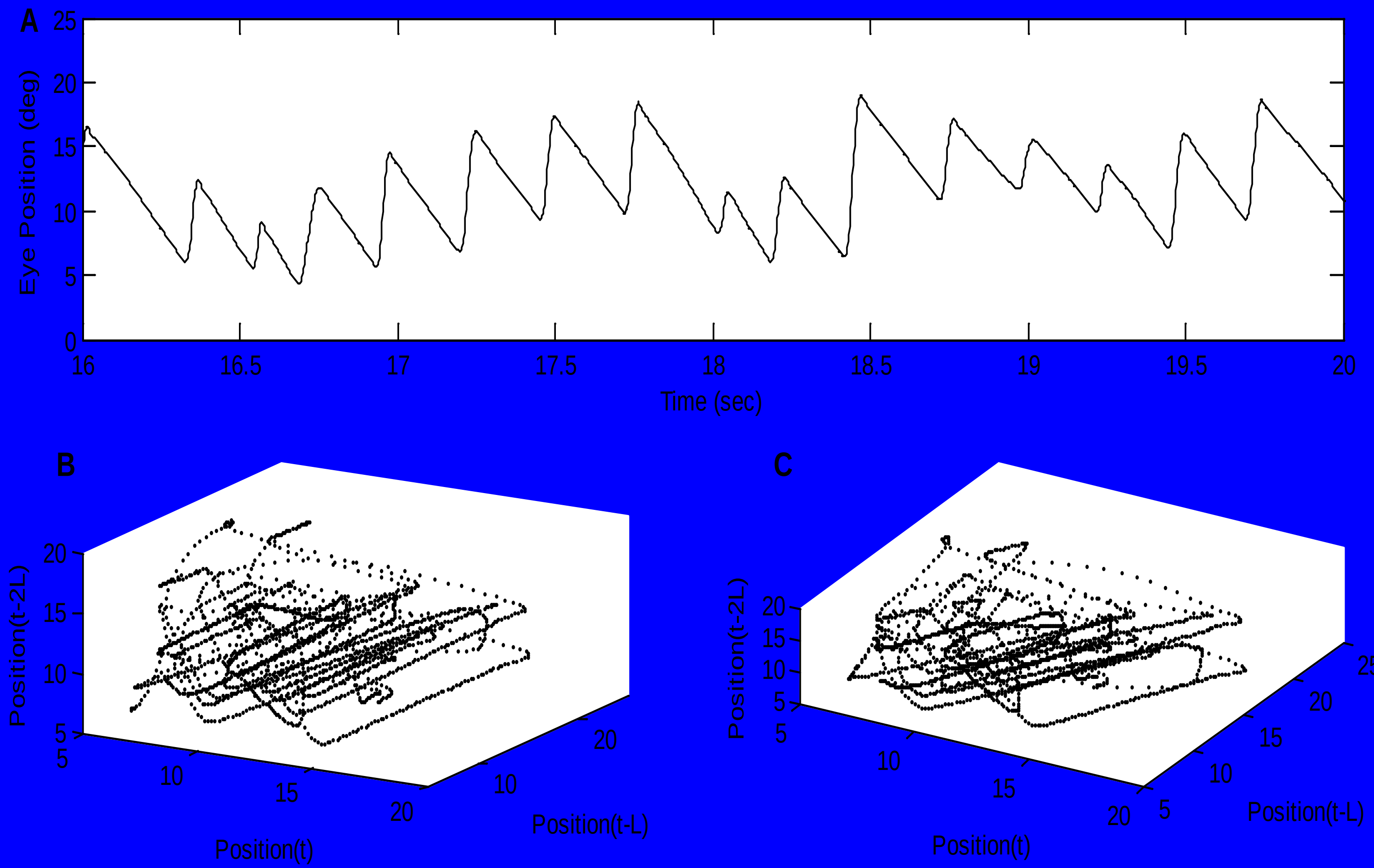
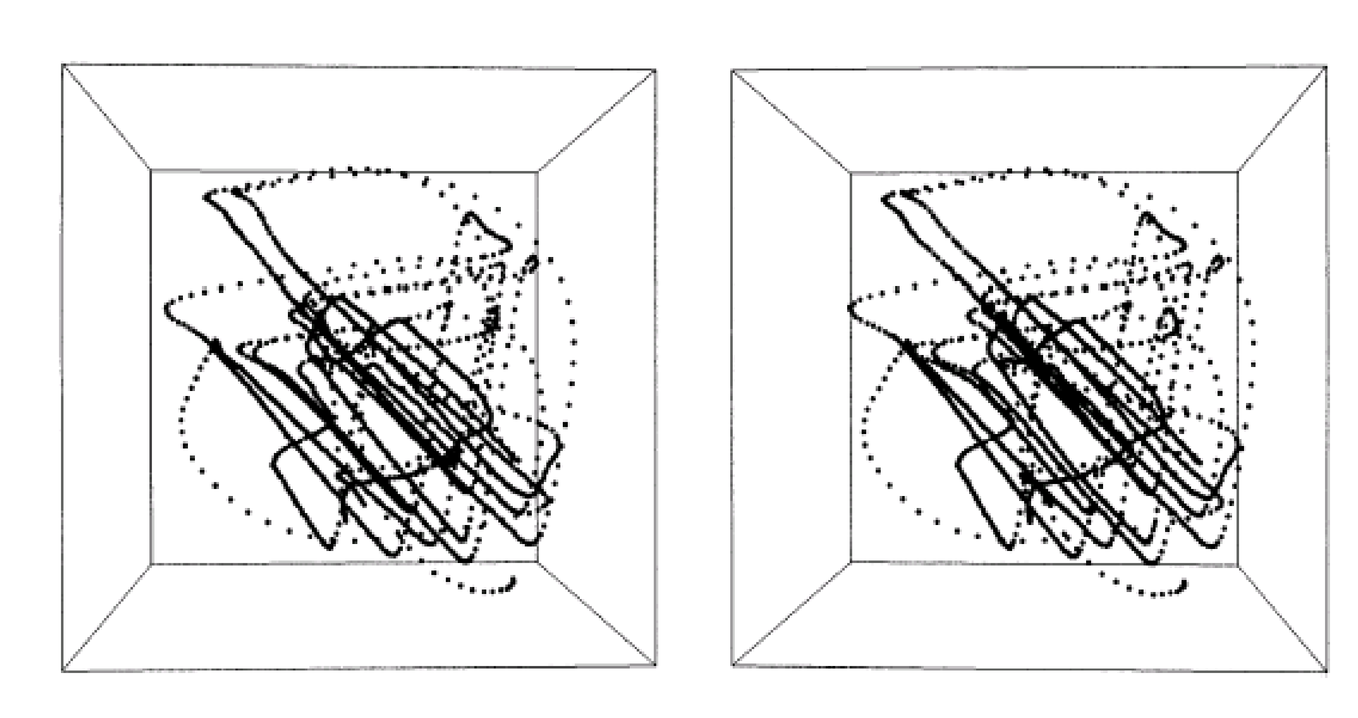


Figure 4.5.1

Attractor Reconstruction



- Stereo view of OKN attractor
  - N=2000 pts
  - L=0.3 sec
  - M=3
  - Sampled at 500 Hz

Figure 15.1.1

Eye-movement example

Correlation Dimension

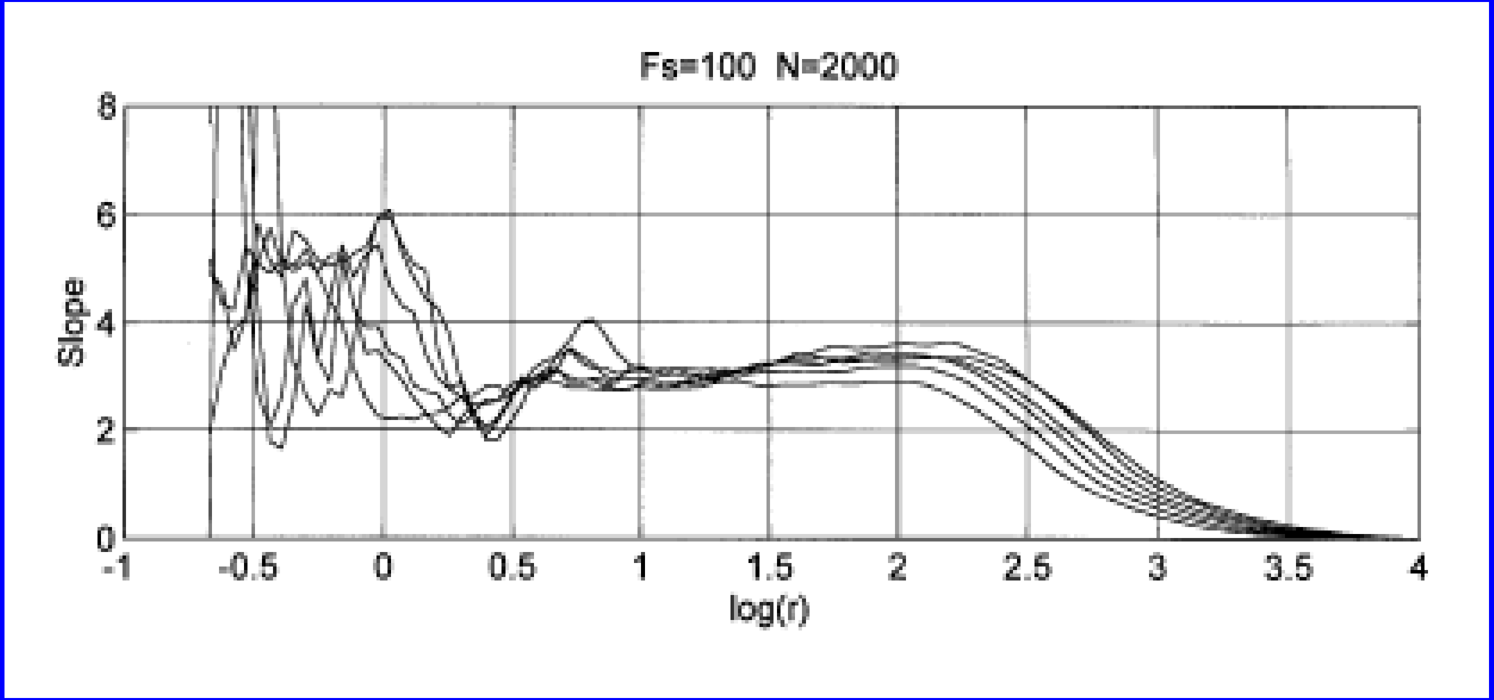


Fig. 4. Slopes of log(correlation integral)/log(distance) for 2000 samples of OKN data sampled at 100 Hz, using a time window of 0.3 s in the attractor reconstruction. Curves for embedding dimensions from 5 to 10 are shown

Shelhamer, Biol. Cybern. 76: 237 (1997)

Eye-movement example

Dimension – change over time

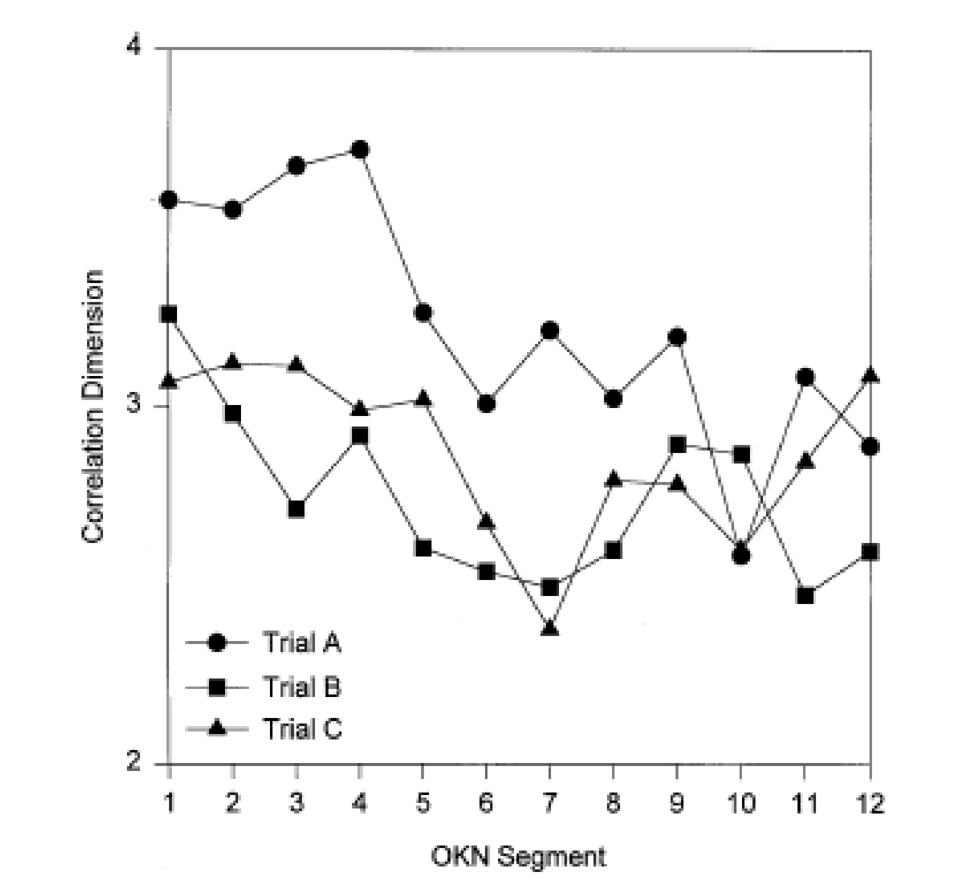


Fig. 6. Nonstationarity of the correlation dimension. Correlation dimensions of 12 separate 10-s segments of OKN data, for three trials. (Trial A supplied the data for the other analyses in this paper.) The dimensions generally decrease over time within a single trial

Shelhamer, Biol. Cybern. 76: 237 (1997)

Dimension – effects of filtering

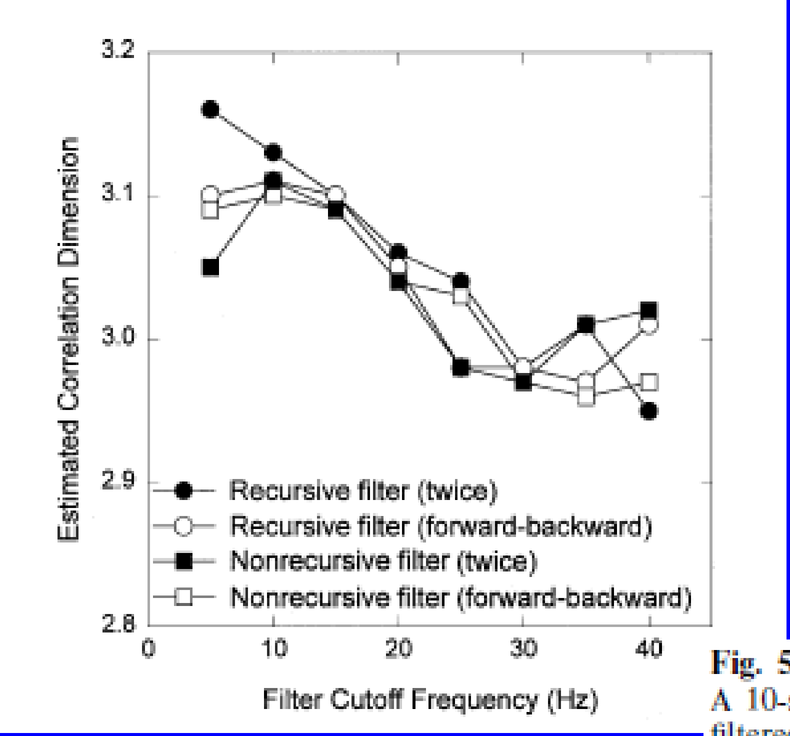


Fig. 5. Effect of digital low-pass filtering on the correlation dimension. A 10-s segment of OKN, sampled at 100 Hz ($N = 1000$ points) was filtered with the cutoff frequencies indicated along the *abscissa*, before computing correlation dimensions. At each frequency, a recursive filter and a nonrecursive filter were used. Each filter was applied in two different ways: in one the data were simply filtered two times (labeled *twice*), in the other the data were filtered once in the forward direction and the filtered data were then filtered in the backward direction, to cancel time delays introduced by the forward filtering (labeled *forward-backward*). Note the overall increase in dimension as the cutoff frequency decreases

Shelhamer, Biol. Cybern. 76: 237 (1997)

Eye-movement example

Surrogates

- Test specific hypothesis about the structure of OKN
  - Random?
  - Linear?
  - Fast-slow couples?
  - Slow-fast couples?
  - Population of slow & fast?

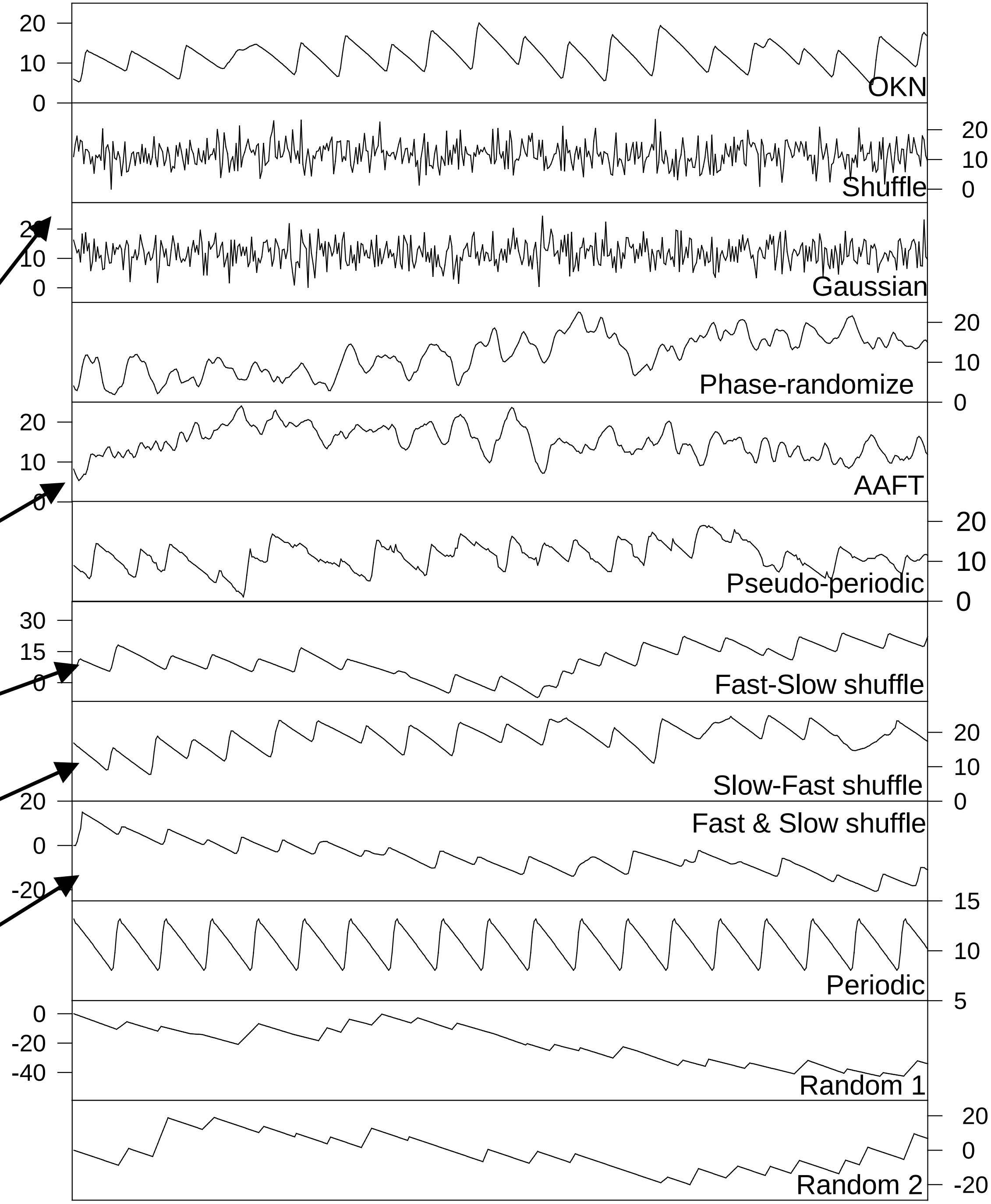


Figure 15.1.5

Surrogates

Surrogate	Correlation dimension (mean±sd)	Magnitude relative to original OKN
OKN	3.46	
Shuffle	—	—
Gaussian	—	—
Phase-randomize	5.10±0.21	10/10 > 3.46
AAFT	—	—
Pseudo-periodic	4.15±0.13	10/10 > 3.46
Fast-Slow shuffle	3.43±0.08	3/10 > 3.46
Slow-Fast shuffle	3.47±0.06	6/10 > 3.46
Fast & Slow shuffle	3.31±0.06	10/10 < 3.46
Periodic	—	—
Random 1	2.88±0.11	10/10 < 3.46
Random 2	2.69±0.16	10/10 < 3.46

Table 15.1.2

OKN Conclusions

- Dimension ~2.8
- Decreases with continued stimulation (decreased volitional component?)
- Increases with filtering
- Surrogates
  - Slow phase does not determine next fast phase
  - Fast phase does not determine next slow phase
  - Population of fast & slow phases important

Summary

- Phase space
  - Visualization of system dynamics
  - Intuition about system behavior
- Time-delay reconstruction
  - Generate phase space without knowing s.v.
  - Allows work in high dimension
  - Use care in parameter selection
- Dimension
  - Scaling process (bulk as function of measurement scale)
  - Requires care in computations
  - Interpret as number of state variables, comparative level of complexity
- Surrogates
  - Error bounds for dimension estimates
  - Testing of hypotheses:
    - Dynamical: random, deterministic, linear, nonlinear
    - Physiological

References

JB Bassingthwaighte, LS Liebovitch, BJ West (1994) Fractal Physiology. Bethesda MD: American Physiological Society.

M Ding, C Grebogi, E Ott, T Sauer, JA Yorke (1993) Plateau onset for correlation dimension: when does it occur? *Physical Review Letters* 70:3872-3875.

J-P Eckmann, D Ruelle (1992) Fundamental limitations for estimating dimensions and Lyapunov exponents in dynamical systems. *Physica D* 56:185-187.

JD Farmer, E Ott, JA Yorke (1983) The dimension of chaotic attractors. *Physica D* 7:153-180.

P Grassberger, I Procaccia (1983) Measuring the strangeness of strange attractors. *Physica D* 9:189-208.

MB Kennel, R Brown, HD Abarbanel (1992) Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A* 45:3403-3411.

B Mandelbrot (1967) How long is the coast of Britain? Statistical self-similarity and fractional dimension. *Science* 156:636-638.

BB Mandelbrot (1983) The Fractal Geometry of Nature. New York: WH Freeman and Co.

E Ott, T Sauer, JA Yorke (eds) (1994) Coping with Chaos: Analysis of Chaotic Data and the Exploitation of Chaotic Systems. New York: Wiley-Interscience.

N Packard, J Crutchfield, J Farmer, R Shaw (1980) Geometry from a time series. *Physical Review Letters* 45:712.

D Prichard, J Theiler (1994) Generating surrogate data for time series with several simultaneously measured variables. *Physical Review Letters* 73:951-954.

PE Rapp (1993) Chaos in the neurosciences: cautionary tales from the frontier. *Biologist* 40:89-94.

PE Rapp, AM Albano, ID Zimmerman, MA Jiménez-Montaño (1994) Phase-randomized surrogates can produce spurious identifications of non-random structure. *Physics Letters A* 192:27-33.

D Ruelle (1990) Deterministic chaos: the science and the fiction. *Proceedings of the Royal Society of London A* 427:241-248.

T Sauer, JA Yorke, M Casdagli (1991) Embedology. *Journal of Statistical Physics* 65:579-616.

SJ Schiff, T Chang (1992) Differentiation of linearly correlated noise from chaos in a biologic system using surrogate data. *Biological Cybernetics* 67:387-393.

T Schreiber, A Schmitz (1996) Improved surrogate data for nonlinearity tests. *Physical Review Letters* 77:635-638.

T Schreiber, A Schmitz (2000) Surrogate time series. *Physica D* 142:346-382.

M Shelhamer (1992) Correlation dimension of optokinetic nystagmus as evidence of chaos in the oculomotor system. *IEEE Transactions on Biomedical Engineering* 39:1319-1321.

M Shelhamer (1997) On the correlation dimension of optokinetic nystagmus eye movements: computational parameters, filtering, nonstationarity, and surrogate data. *Biological Cybernetics* 76:237-250.

M Small, D Yu, RG Harrison (2001) Surrogate test for pseudoperiodic time series data. *Physical Review Letters* 87:188101-1:4.

M Small, CK Tse (2002) Applying the method of surrogate data to cyclic time series. *Physica D* 164:187-201.

CJ Stam, JPM Pijn, WS Pritchard (1998) Reliable detection of nonlinearity in experimental time series with strong periodic components. *Physica D* 112:361-380.

G Sugihara, RM May (1990) Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature* 344:734-741.

F Takens (1981) Detecting strange attractors in turbulence. In: Lecture notes in mathematics, Vol.898: Dynamical systems and turbulence. Page 366. Berlin: Springer.

J Theiler (1986) Spurious dimension from correlation algorithms applied to limited time-series data. *Physical Review A* 34:2427-2432.

J Theiler (1990) Estimating fractal dimension. *Journal of the Optical Society of America A* 7:1055-1073.

J Theiler, S Eubank, A Longtin, B Galdrikian, JD Farmer (1992) Testing for nonlinearity in time series: the method of surrogate data. *Physica D* 58:77-94.

Fractals in Motor Control

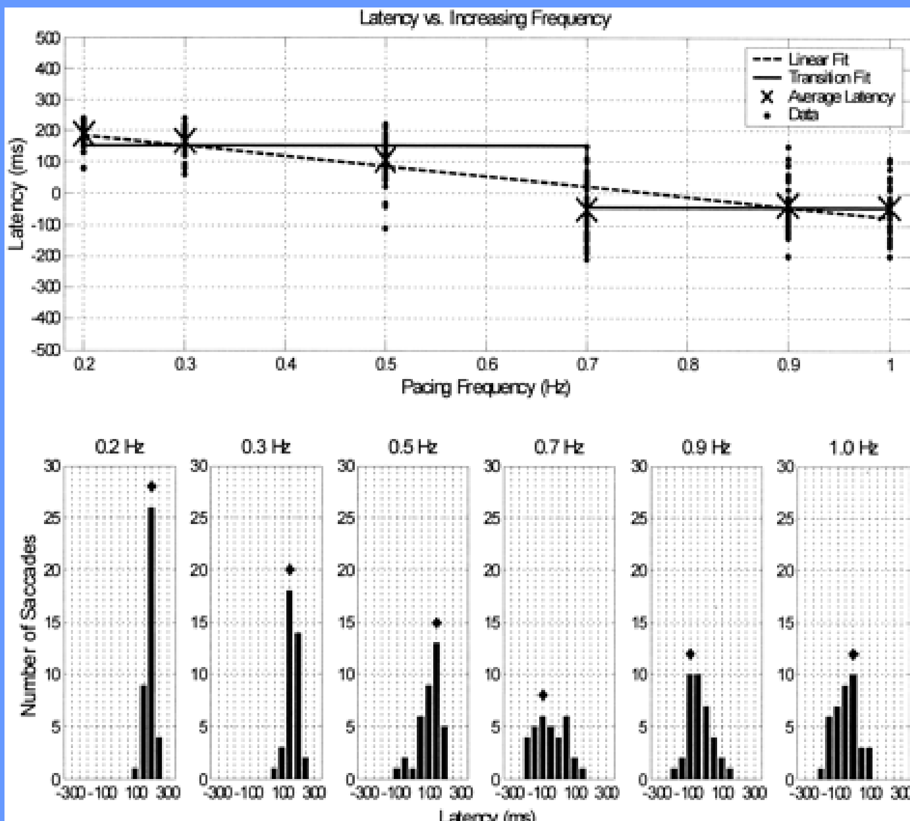
Mark Shelhamer
Departments of Otolaryngology and Biomedical Engineering
The Johns Hopkins University School of Medicine
Baltimore MD

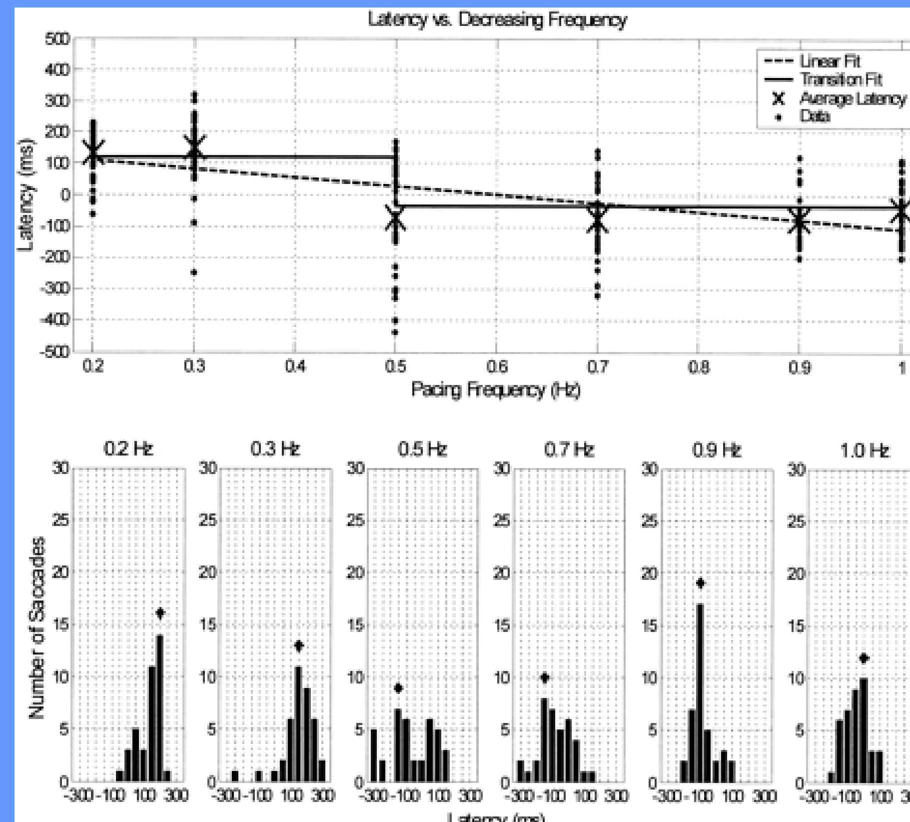
Overview

- Timing of predictive saccades
  - Inter-trial correlations
- Fractals
  - Math
  - Physiology
- Saccade amplitudes
  - Predictive amplitude control
  - Prediction/Adaptation
- VOR
  - Prediction/Adaptation
- Vestibular afferents
  - Correlations over long times

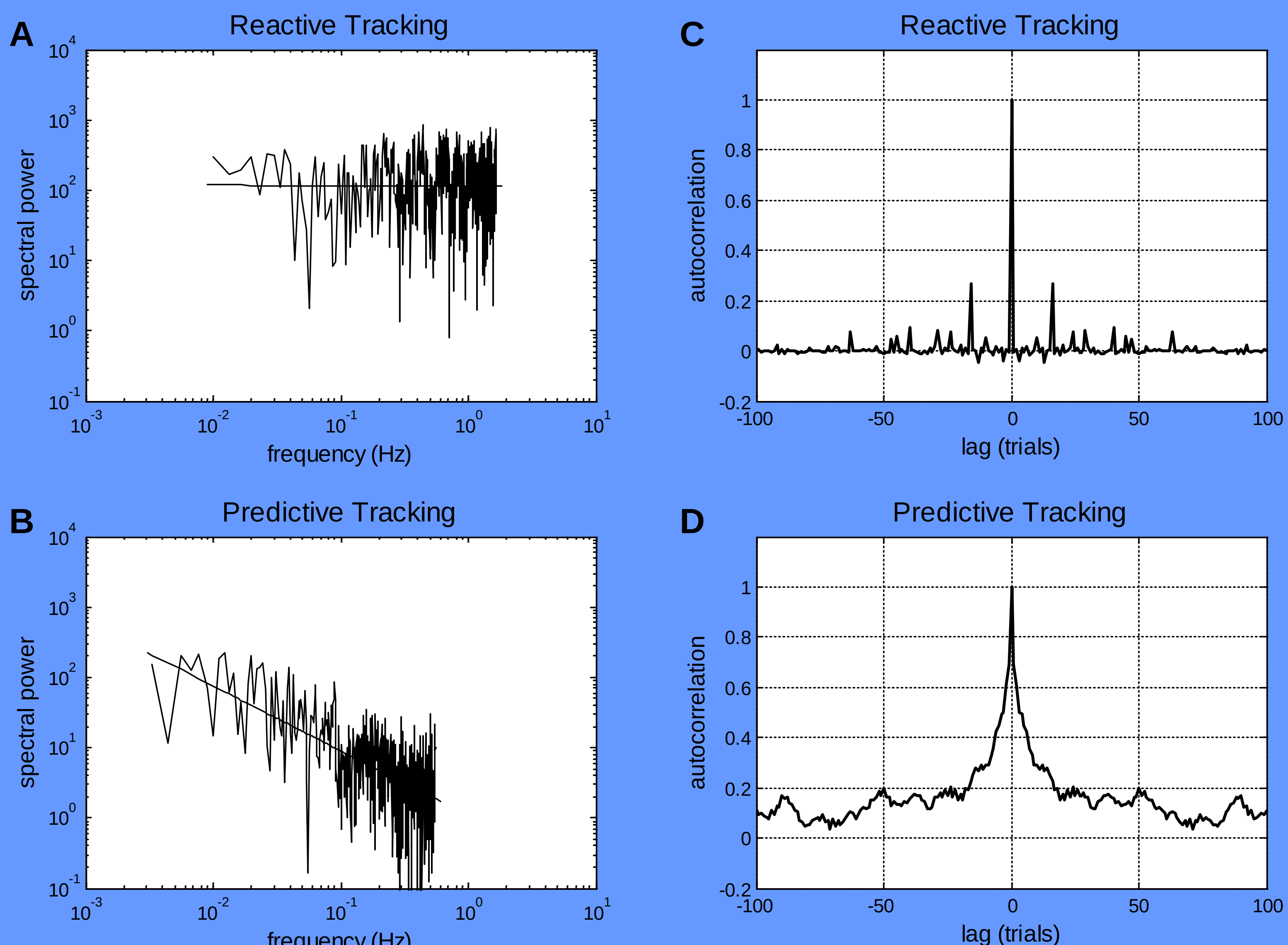
Predictive Saccade Timing

Transition between reactive and predictive tracking is abrupt not smooth





Power Spectra and Autocorrelation Functions for predictive saccade latencies



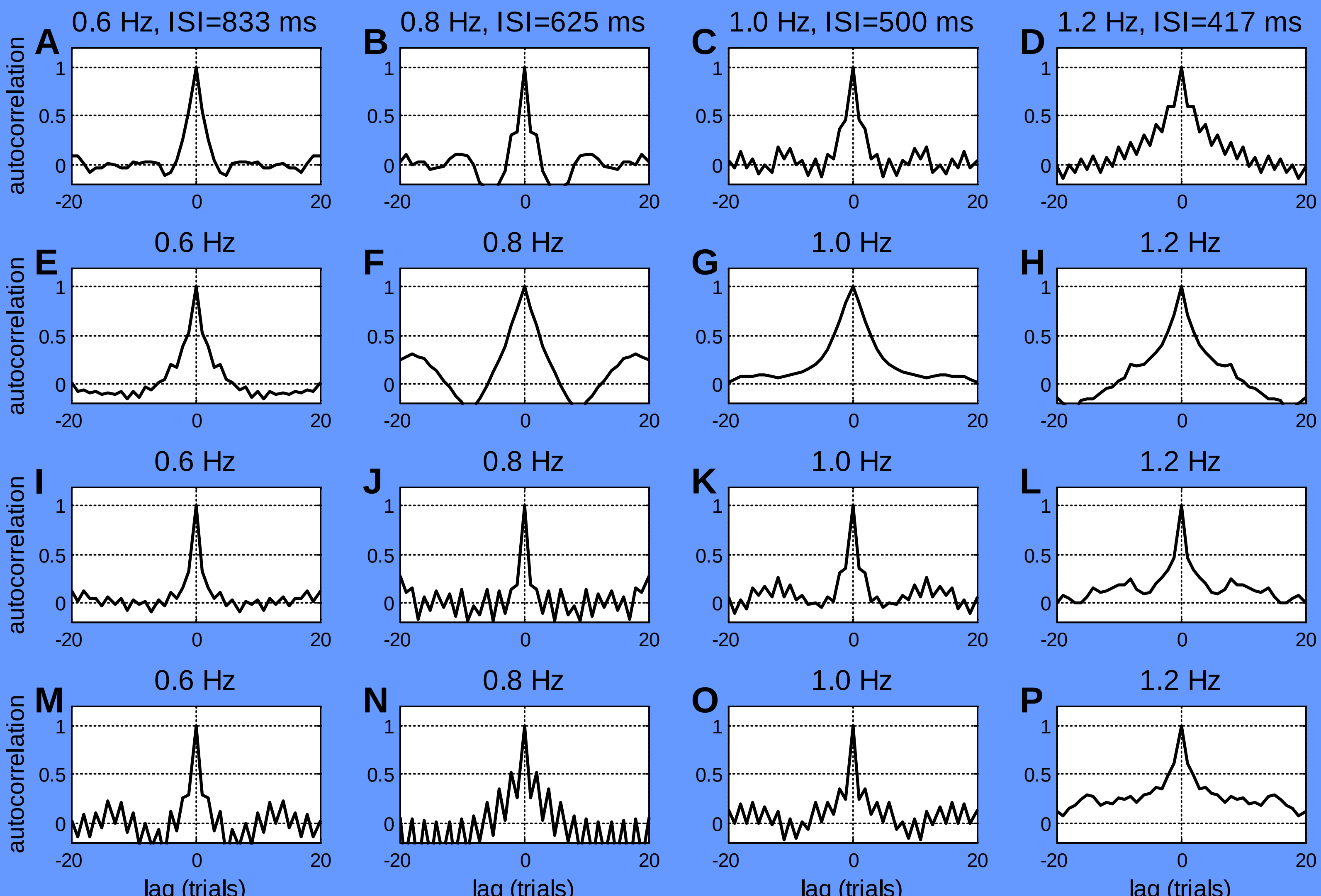
- Latencies of *reactive* saccades are *uncorrelated* (white noise).
- Latencies of *predictive* saccades are *correlated*.
- Inter-trial correlations imply a parametric feedback mechanism
  - Performance of previous saccades influences programming of future saccades

---

Are the correlations constant in terms of **time** or in terms of **trials**?

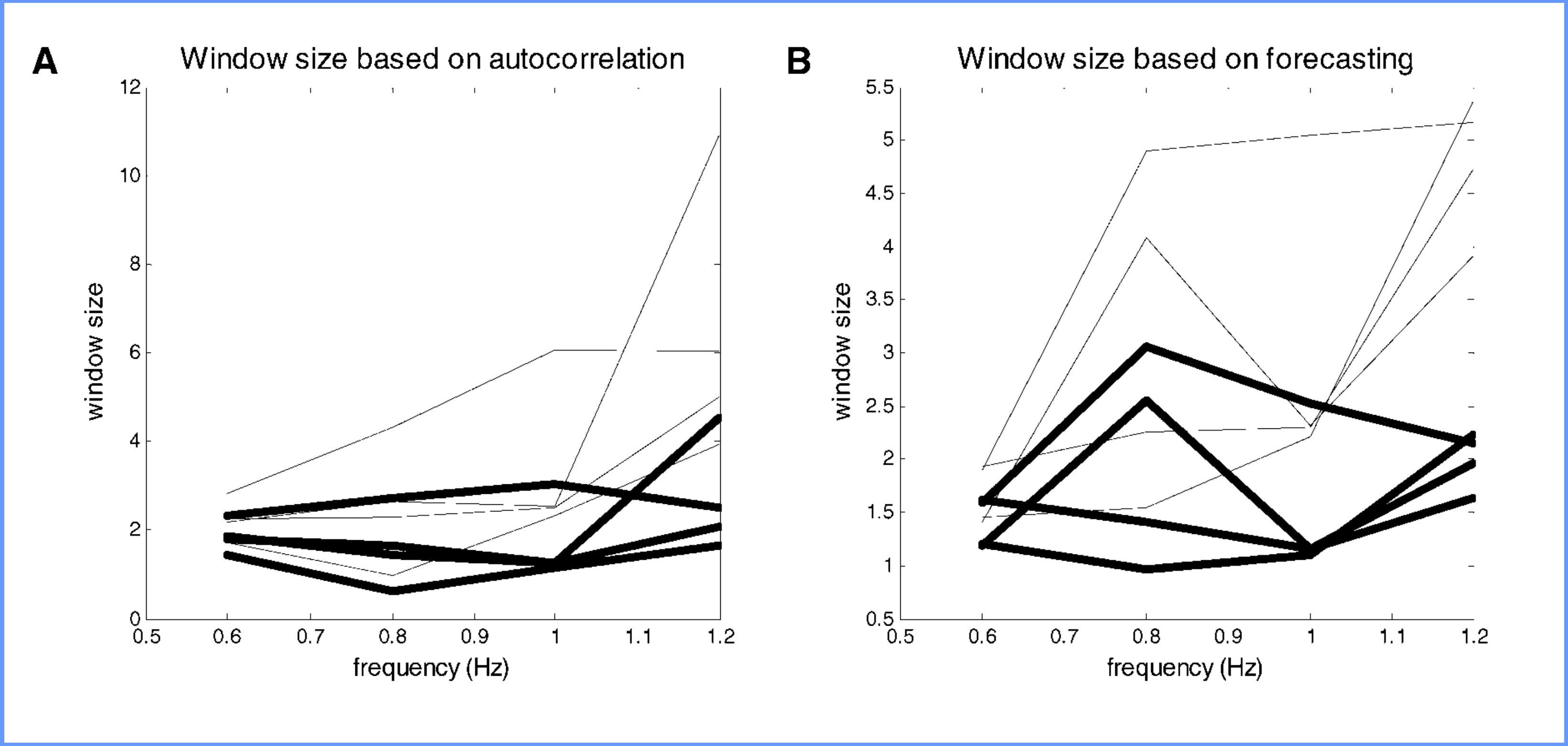
Autocorrelations

Define correlation window as time over which correlations are ≥ 0.2.



Correlation window is constant in terms of time not trials.

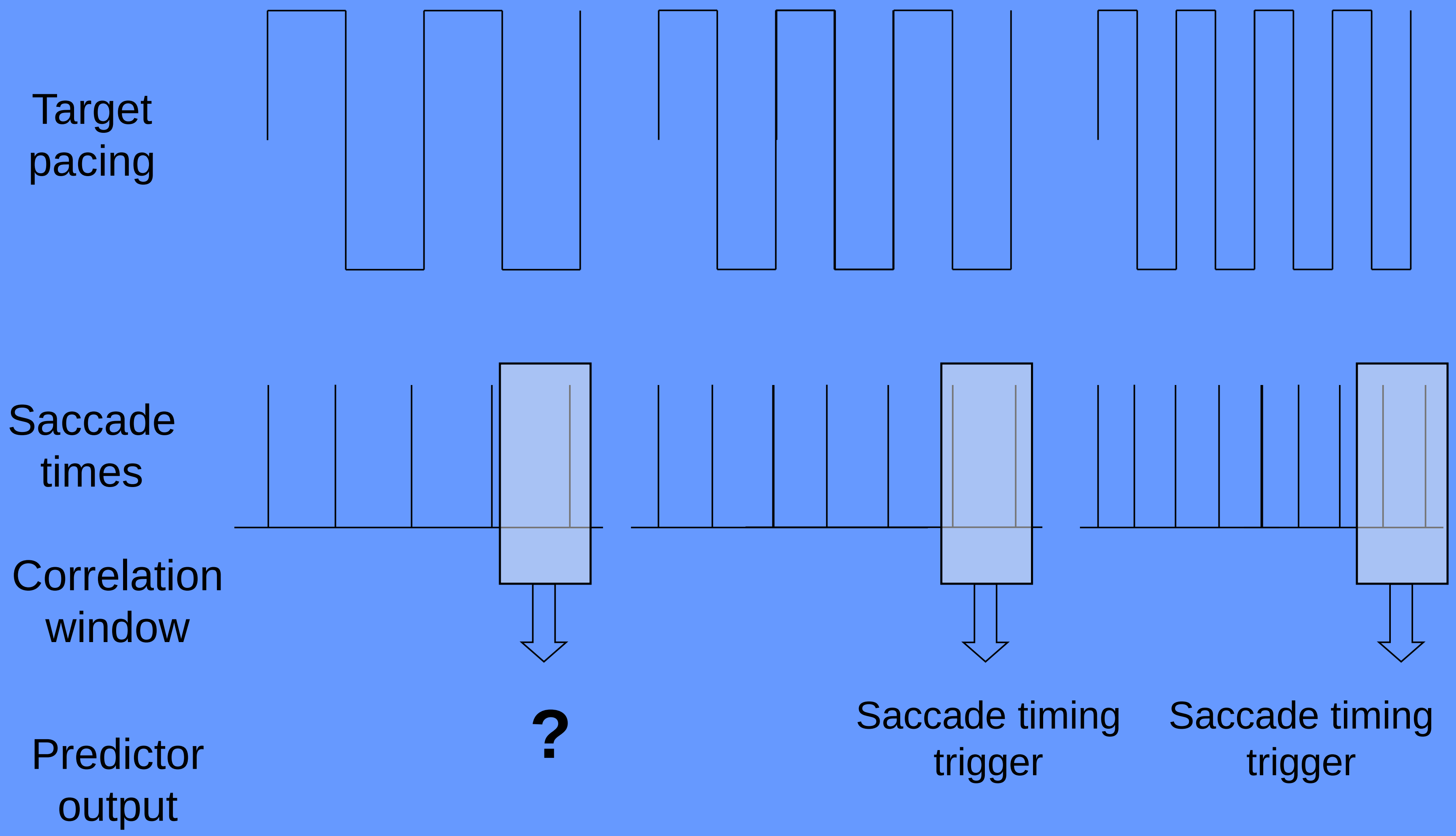
Two-second correlation window for previous saccades



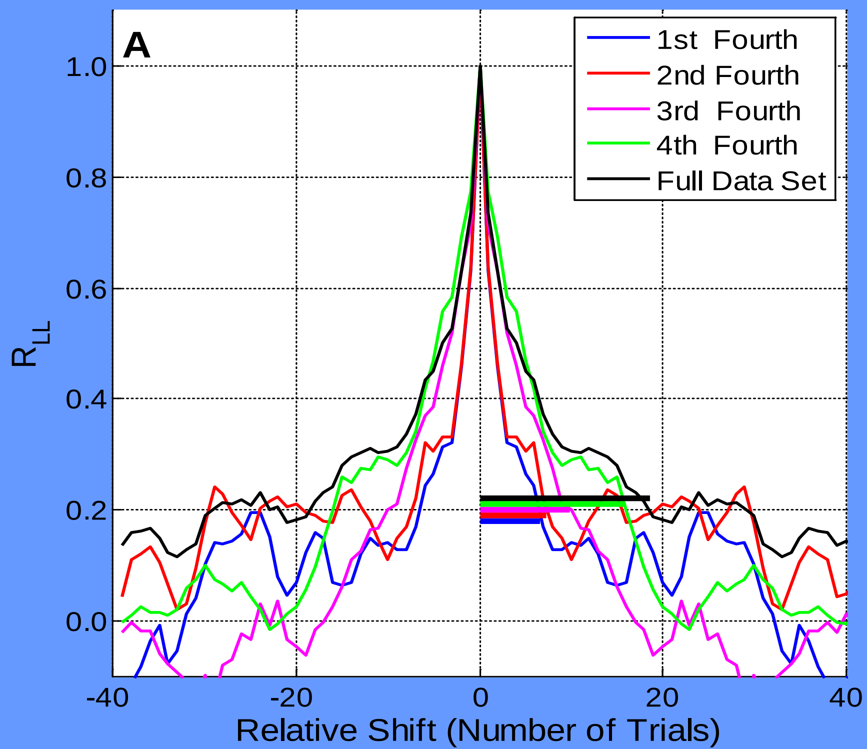
Thin lines: window in terms of trials.
Thick lines: window in terms of time (sec).

Outline of a model for predictive saccades

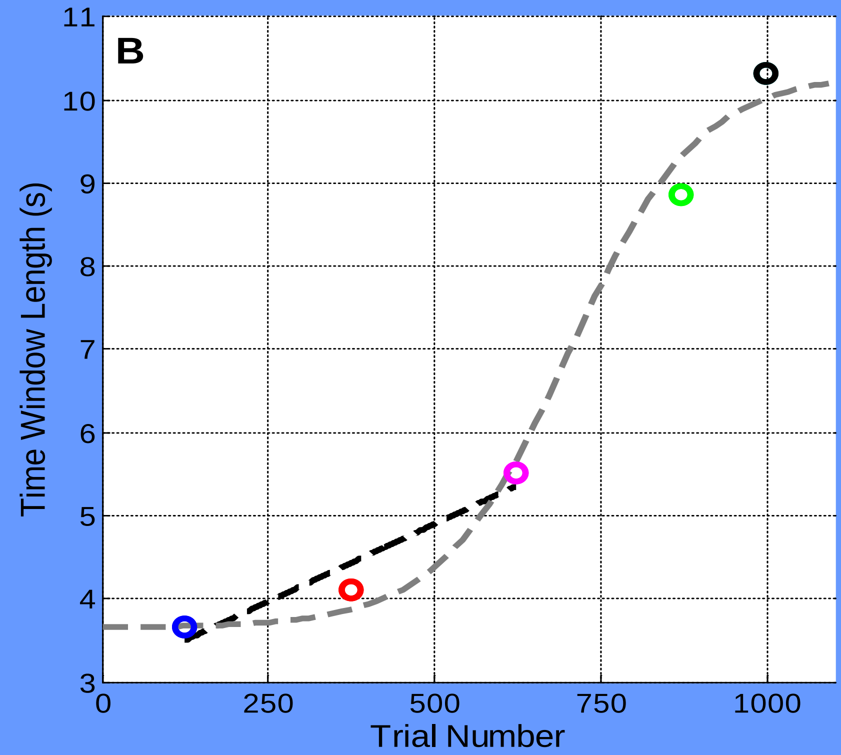
Predictive saccades are made when two or more saccades fall into preceding 2-second window



Extended Tracking Increases Correlations



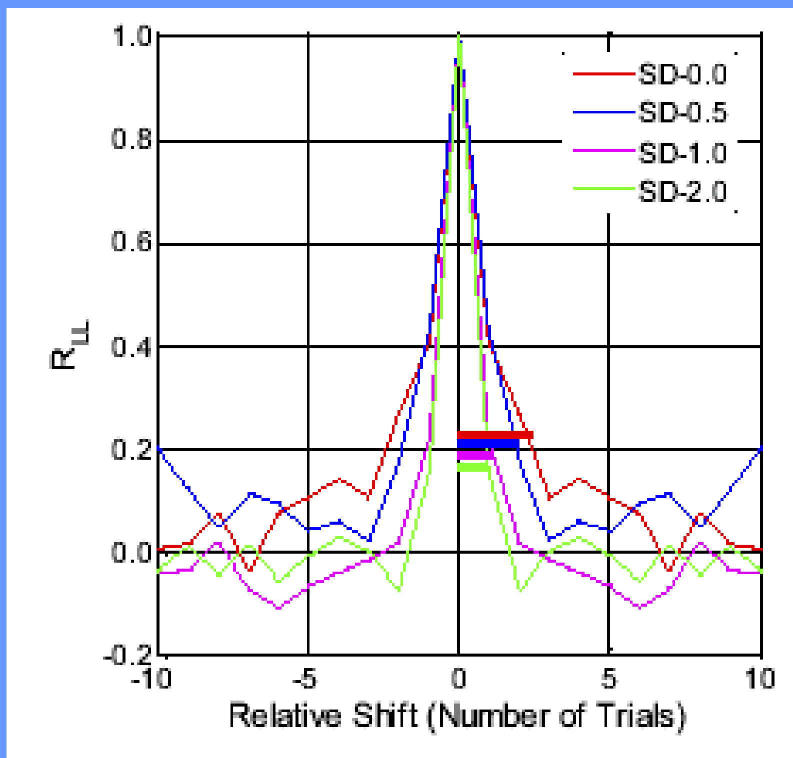
Autocorrelation functions for four non-overlapping quarters (250 points each) from 1000 consecutive predictive saccade latencies.

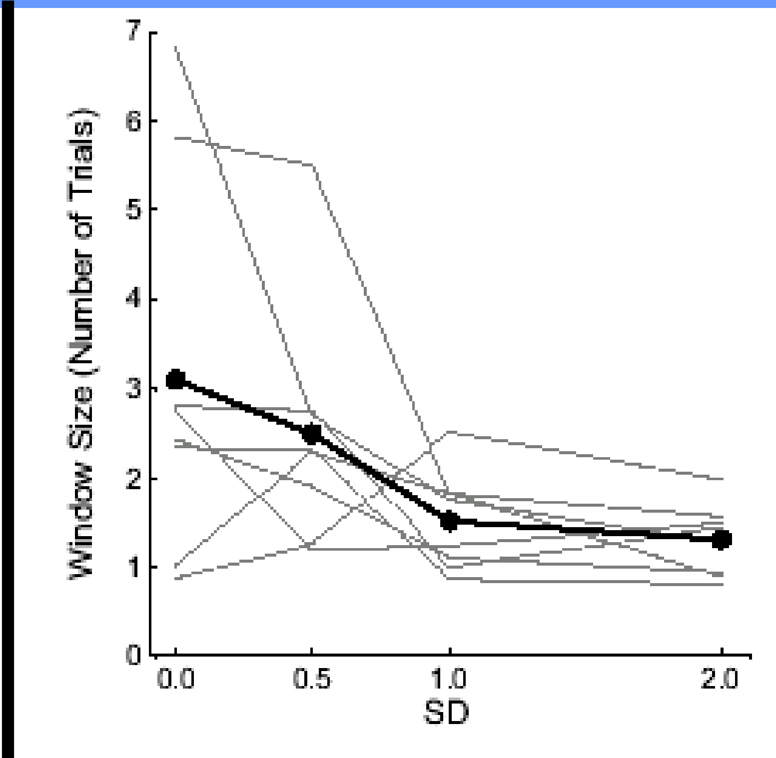


Correlation window sizes. Correlation window increases as tracking continues.

Variable timing

Response to different levels of stimulus (ISI) variability.





Autocorrelations and correlation windows (horizontal lines) as ISI variability increases.

Correlation window as ISI variability increases, for all subjects (mean in bold).

What is the nature of the correlations between predictive saccades?

- $1/f^{\alpha}$ power spectrum
- Slow decay of autocorrelation
  - Power-law ($\tau^{-\beta}$) rather than exponential ($e^{-a\tau}$)
- Nonlinear forecasting decay
  - Ability to forecast future values decays exponentially fast
- Hurst exponent consistent with spectral decay

Fractals

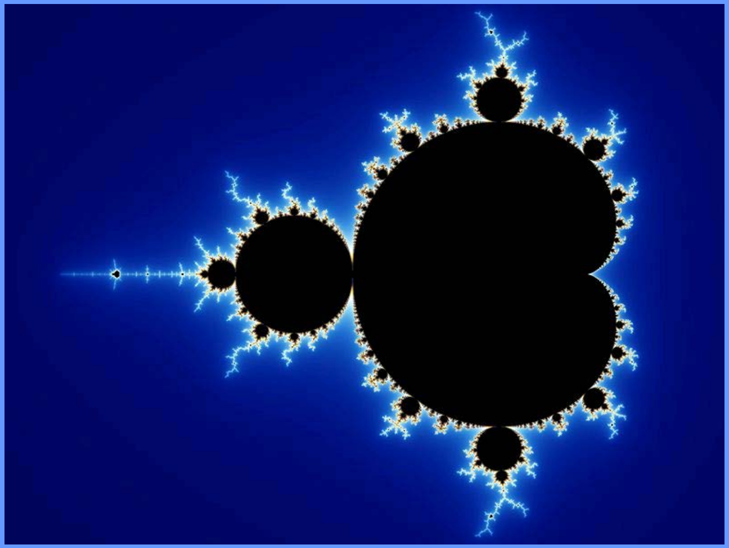
Overview – Fractals in Neurophysiology

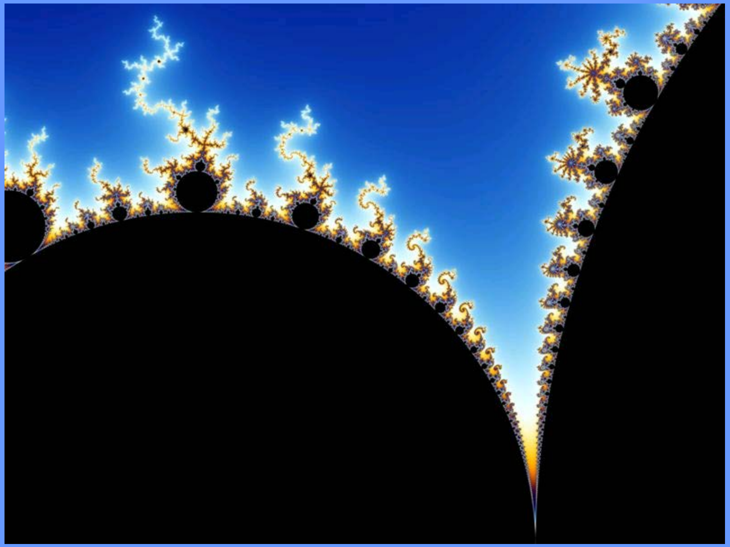
Physiological relevance and functional role of fractal structure is not known.

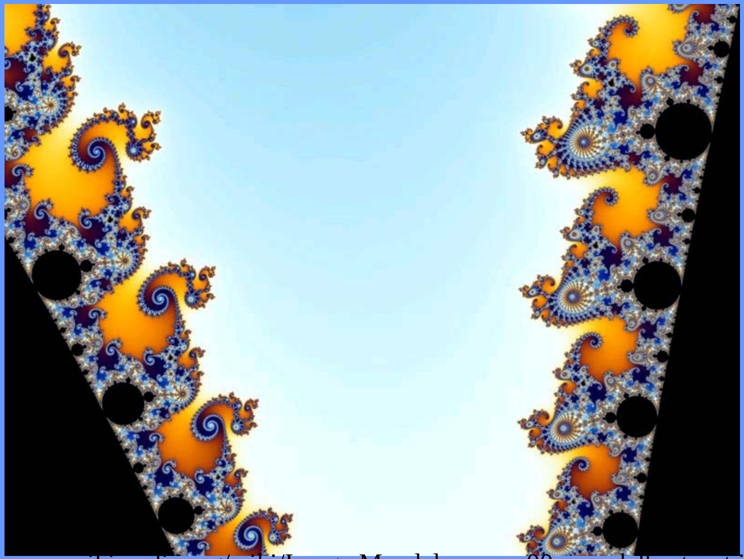
- Fractal behavior has been observed in
  - sequences of reaction times [Wing et al. 2004]
  - stride intervals in gait [Hausdorff 2007, Peng et al. 2000]
  - sequences of predictive saccades
  - visual search [Aks et al. 2002, Stephen & Anastas 2011]
- Distributed processing provides a plausible mechanism for generation of correlations over multiple time scales.
- Fractal structure (correlations) breaks down in pathology and disease [see West 2006].
- Epiphenomenon of fractal ion channel dynamics [Bassingthwaighte et al. 1994]?

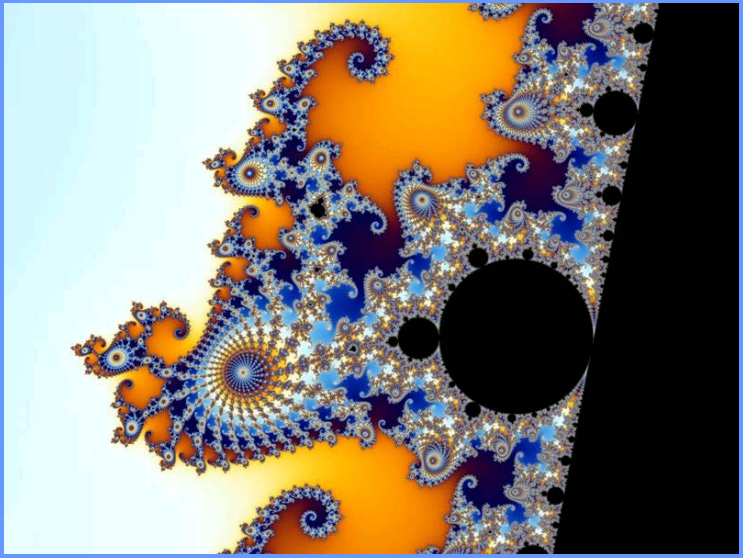
Fractals

A deterministic fractal – the Mandelbrot Set



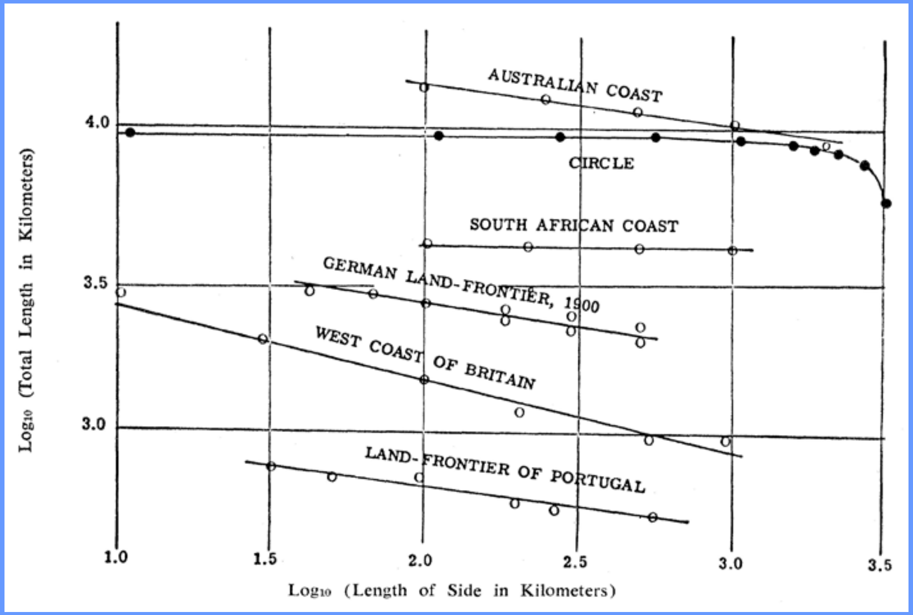






http://commons.wikimedia.org/wiki/Image:Mandel_zoom_00_mandelbrot_set.jpg

Random Fractals - Coastline



- Coastline is a *fractal* – variation on many length scales [Mandelbrot].
- Measured length changes as length of “ruler” changes:
  - – No characteristic length scale produces the “correct” answer.
- Change in measured length with ruler size described by a *scaling law*.
- Scaling law yields fractal exponent (slope on log-log plot), indicating *power-law* relationship:

$$\textbf{(Measured Length) = (Ruler Size)}^{-\alpha}$$

B Mandelbrot (1967) How Long Is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension. *Science* 156:636-638.

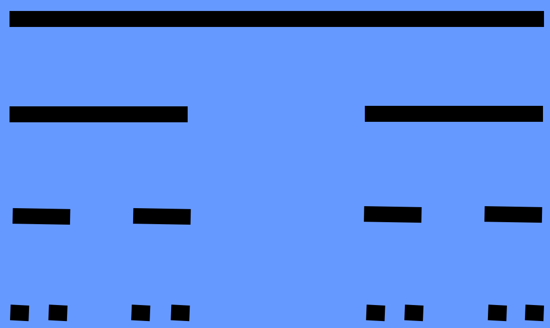
http://commons.wikimedia.org/wiki/Image:Britain-fractal-coastline-combined.jpg

Dimension - Scaling

- Power-law scaling
  - Length $\propto$ (measurement unit)$^{-D}$
  - $L \propto \varepsilon^{-D}$

Fractal Scaling

- Middle-thirds Cantor set

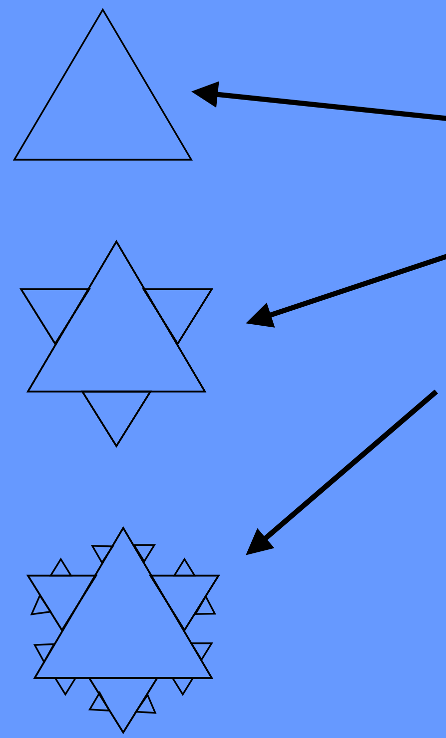


step (n)	number of segments	segment length	total length
0	1	1	1
1	2	1/3	2/3
2	4	1/9	4/9
3	8	1/27	8/27
N	2^N	$(1/3)^N$	$(2/3)^N$
∞	∞	0	0

- D = 0.6309…

Fractal Scaling

- Koch curve



step (n)	number of segments	segment length	total length
0	3	1	3
1	12	1/3	12/3
2	48	1/9	48/9
3	192	1/27	192/27
N	3×4^N	$(1/3)^N$	$3 \times (4/3)^N$
∞	∞	0	∞

- D = 1.2619…

Quantifying Fractal Scaling

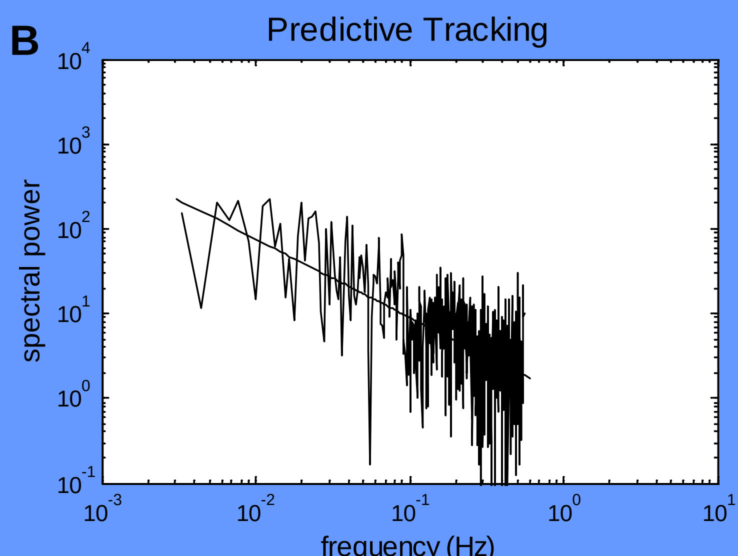
Spectral Analysis

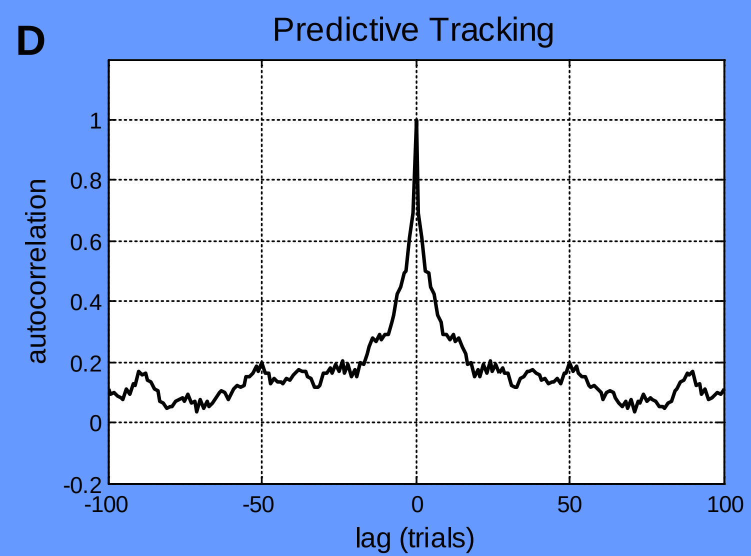
- The power spectrum $S_{xx}(f)$:
  - Fourier transform of autocorrelation function $R_{xx}(\tau)$
- If autocorrelation decays as a power law:

$$R_{xx}(\tau) \sim \tau^{-\beta}$$

power spectrum will have the form:

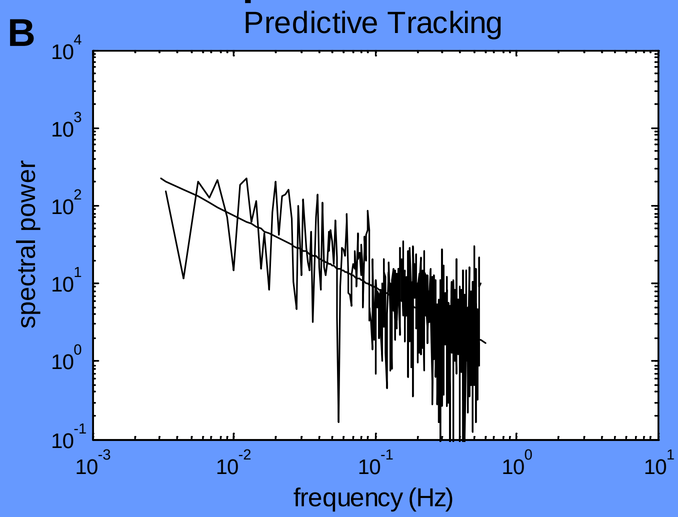
$$S_{xx} \sim f^{-\alpha}$$

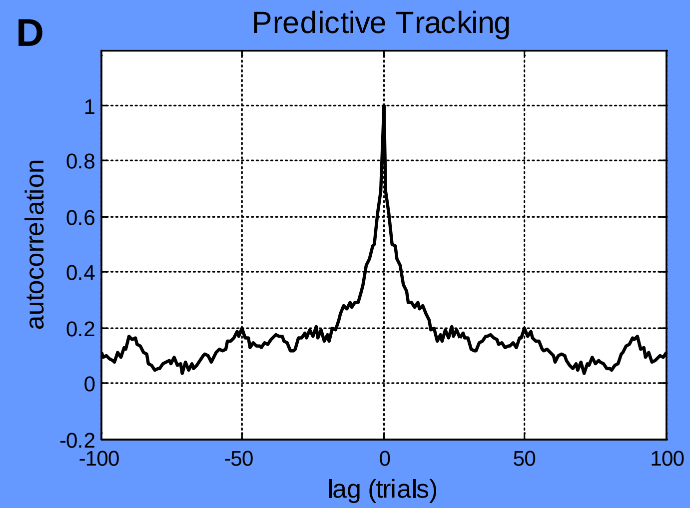




Aspects of power-law scaling

Power-law decay of spectrum ⟷ Power-law decay of autocorrelation

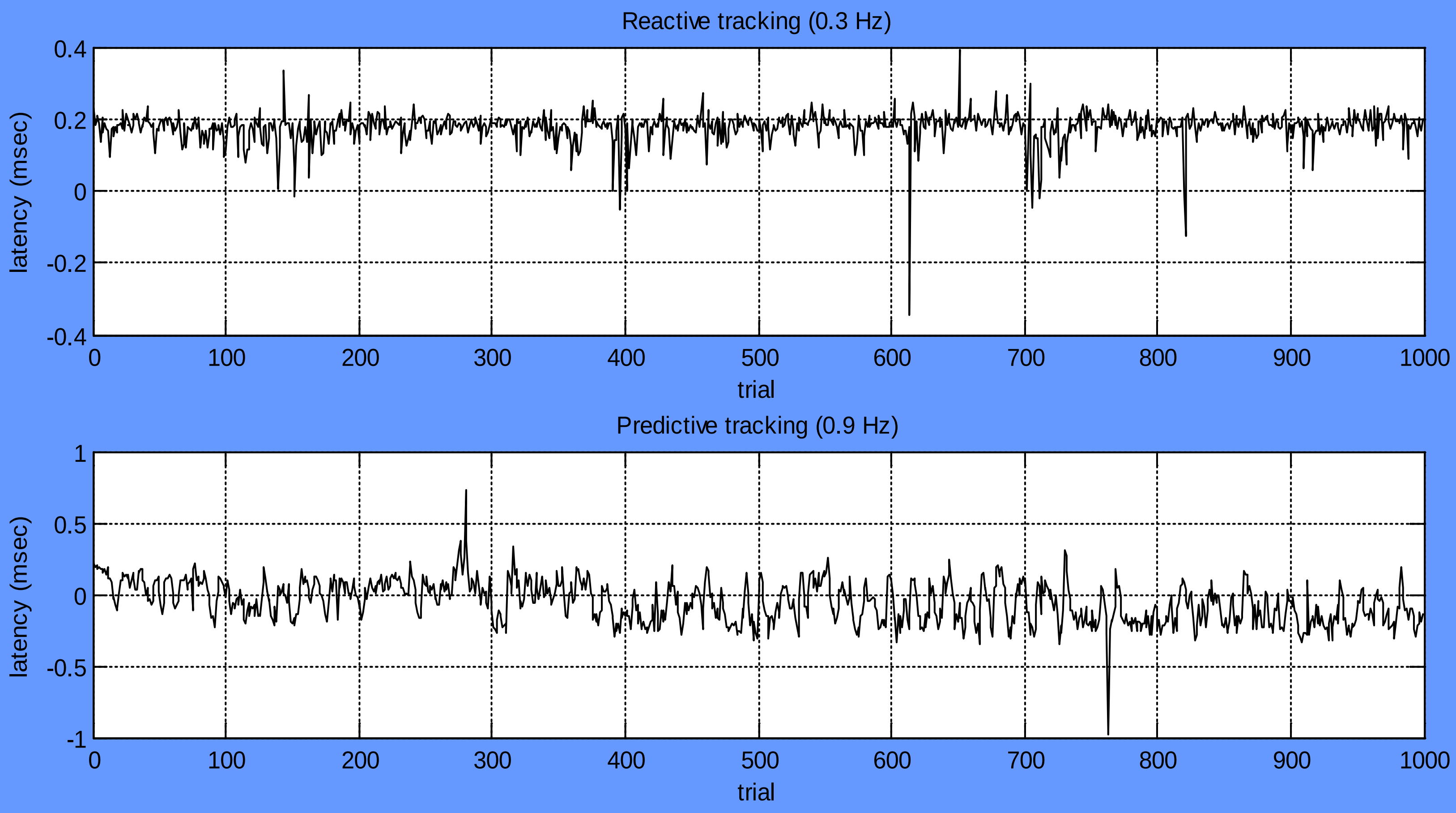




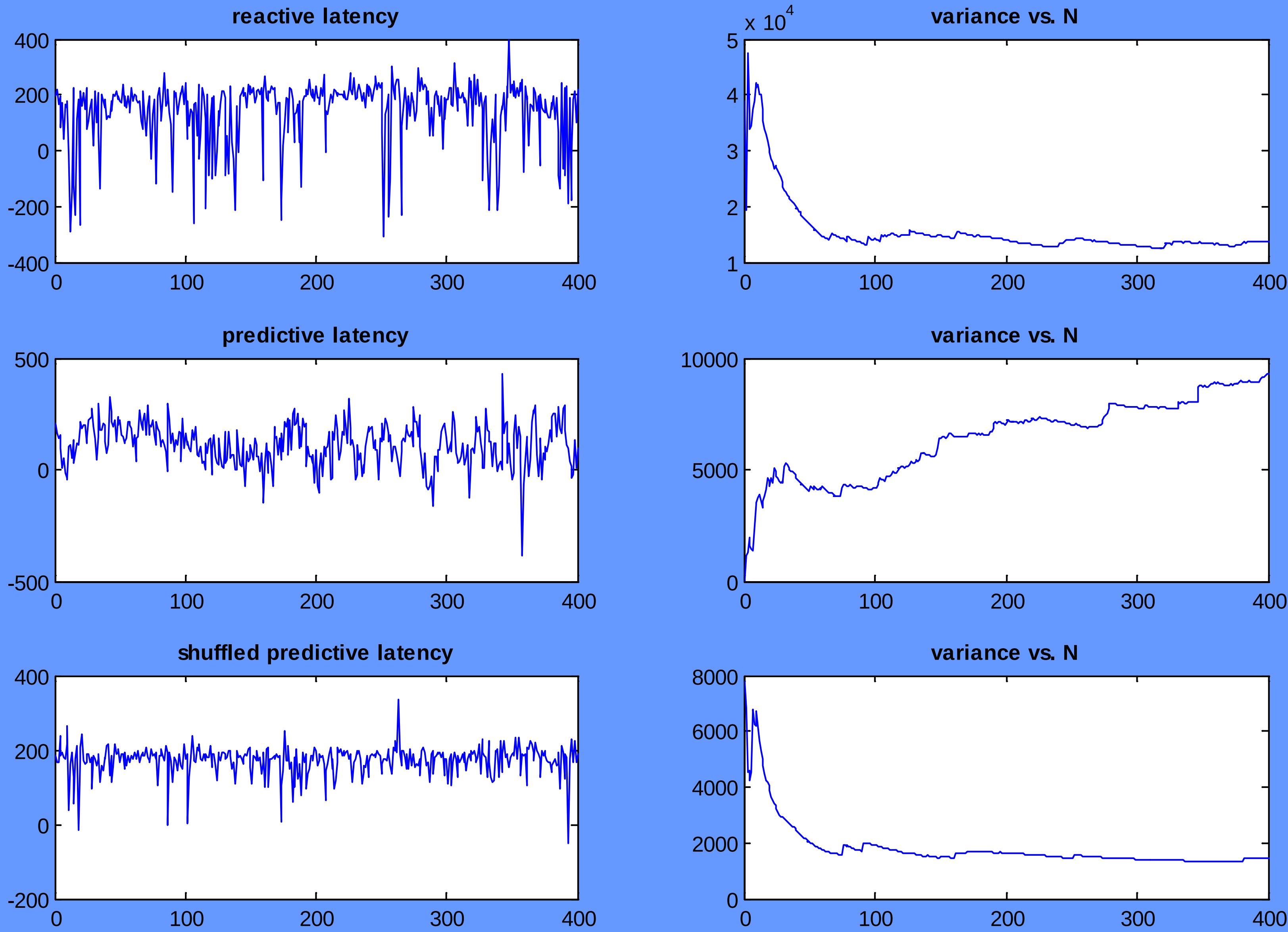
- Variability changes with size (duration) of measurement window.
- Reflected in frequency domain as power-law decay:
  - $S_{xx}(f) \sim 1/f^{\alpha}$
  - shift in frequency by Δf changes power by $(\Delta f)^{-\alpha}$
- Non-integrable in either domain: $\int R_{xx}(\tau) \to \infty \qquad \int S_{xx}(f) \to \infty$
- Substantial low-frequency energy:
  - implies long-term correlations.
  - makes parameter measurement difficult.

Back to Saccades

Reactive and Predictive Latencies



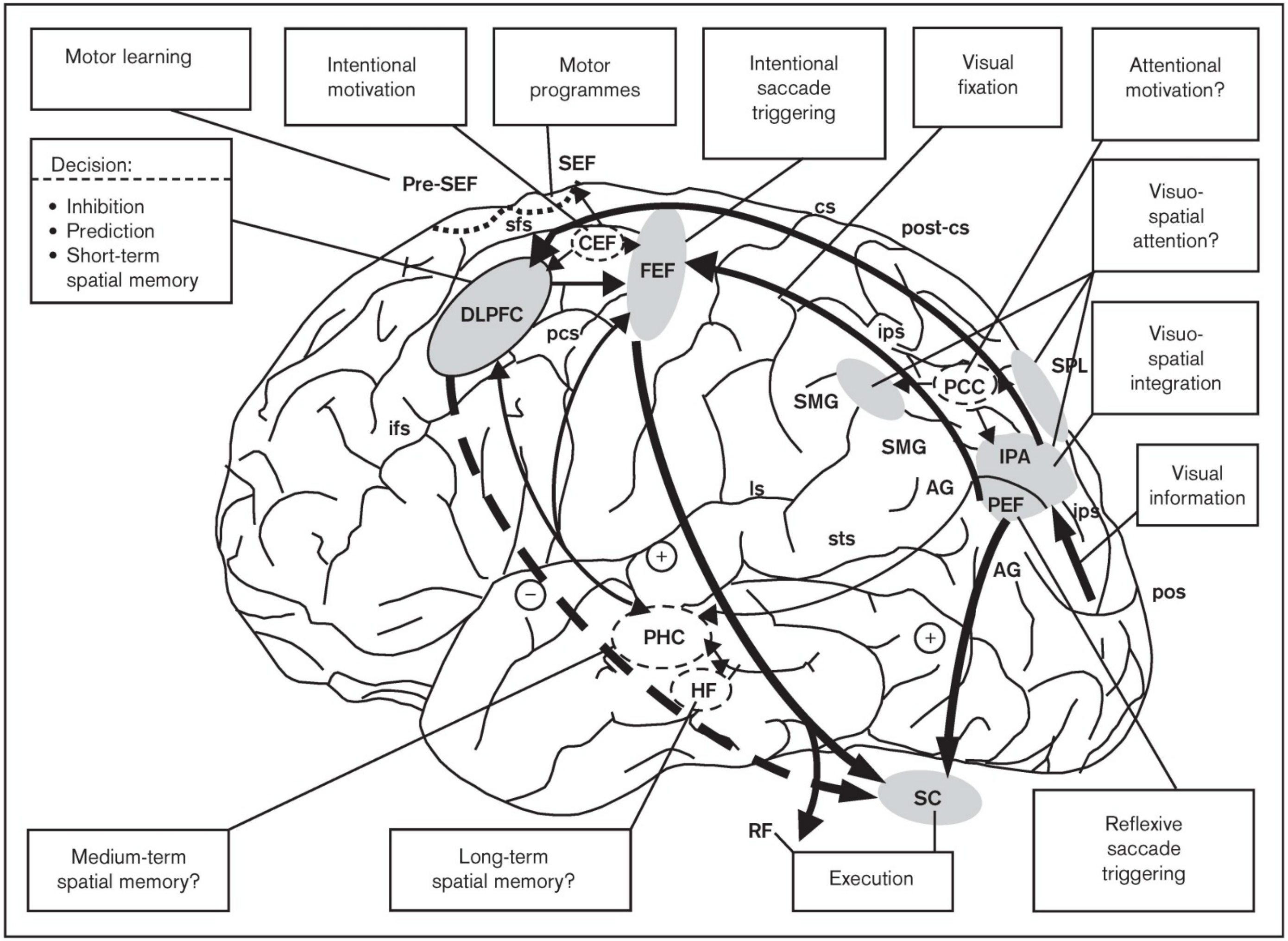
Variance is not well-behaved for predictive saccades



Signal Paths and Fractal Scaling

- Do varied and shifting neural pathways produce fractal responses?
- Time constants of eyeball dynamics 0.012 to 7.8 sec
- Variation on short time scales: trial-to-trial monitoring of performance
- Variation on longer time scales: monitoring of stimulus structure and history
- Cerebellum: timing, hundreds of milliseconds
- Basal ganglia: timing, several seconds
- Cortex: learning & memory, several minutes

Cortical areas involved in the control of saccades



SEF, supplementary eye field; sfs, superior frontal sulcus; CEF, cingulate eye field; cs, central sulcus; DLPFC, dorsolateral prefrontal cortex; pcs, precentral sulcus; FEF, frontal eye field; ips, intraparietal sulcus; ifs, inferior frontal sulcus; SMG, supramarginal gyrus; PCC, posterior cingulate cortex; SPL, superior parietal lobule; IPA, intraparietal areas; ls, lateral sulcus; AG, angular gyrus; PEF, posterior eye field; sts, superior temporal sulcus; pos, parieto-occipital sulcus; PHC, parahippocampal cortex; HF, hippocampal formation; SC, superior colliculus; RF, reticular formations.

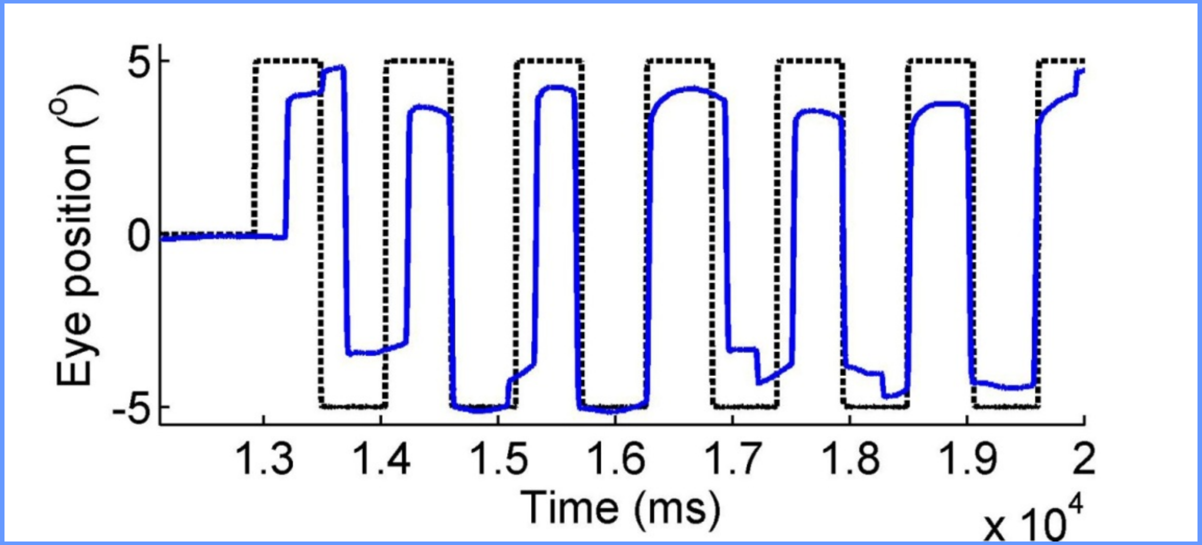
C Pierrot-Deseilligny, D Milea, R Müri (2004) Eye movement control by the cerebral cortex. *Current Opinion in Neurology* 17:17-25.

What about saccade amplitudes?

- Predictive saccades must anticipate not only *when* but *where* a target will appear
- Predictive amplitudes must be controlled
- More biologically relevant than timing?

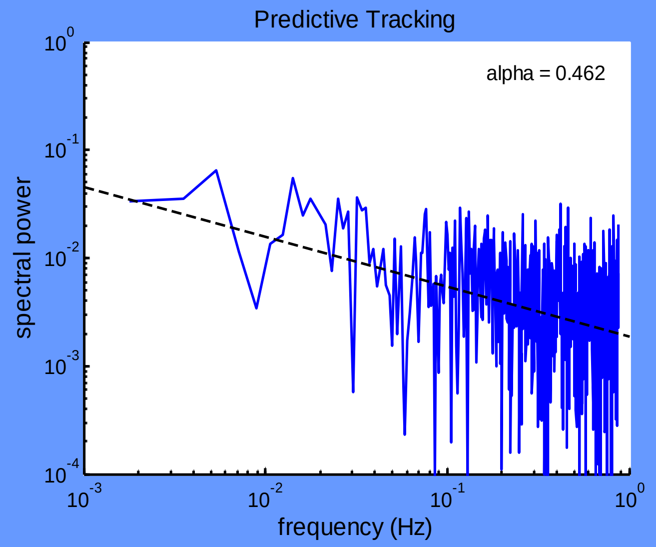
Predictive saccades

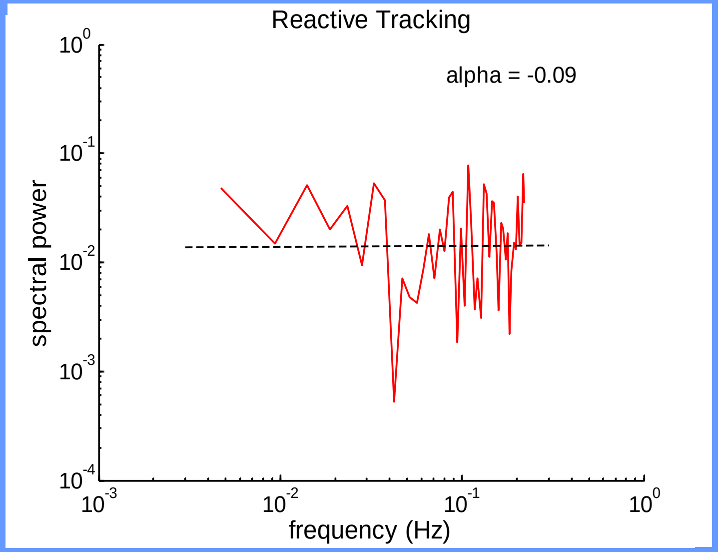
- Form of motor learning
  - – Automatic, non-cued phenomenon
  - – Example of error-based learning
  - – Stationary process, no induced gain change



Long-term Adjustment

- Long-term, inter-trial correlations
  - Wide correlation window
  - Power-law decay of power spectrum
  - Verified by Hurst exponent and ARFIMA modeling
  - Bootstrap analysis: correlations may extend as many as ~70 trials in the past

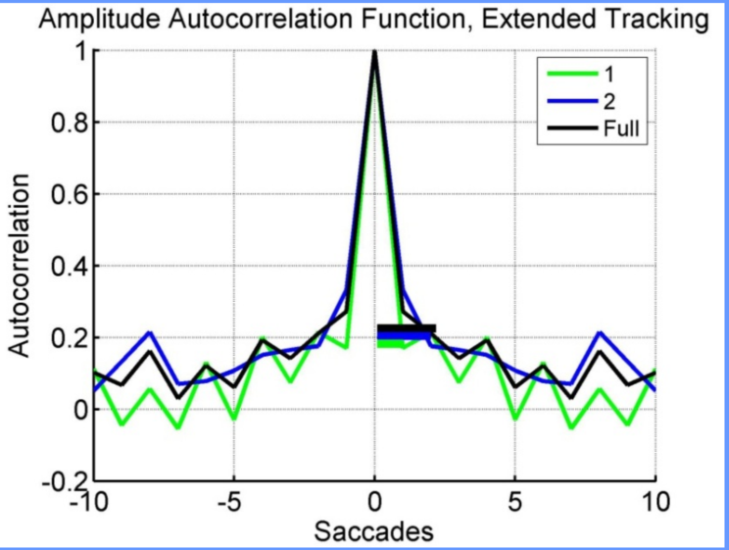


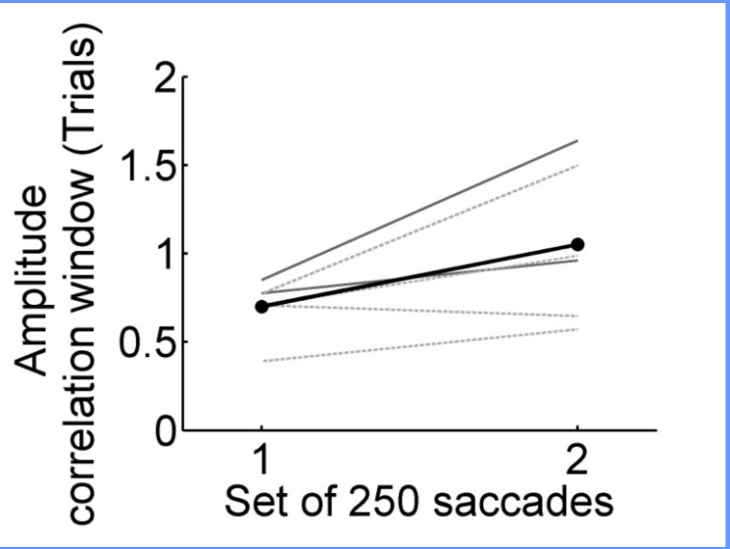


Wong and Shelhamer, under review

Correlation window modulates with stimulus reliability

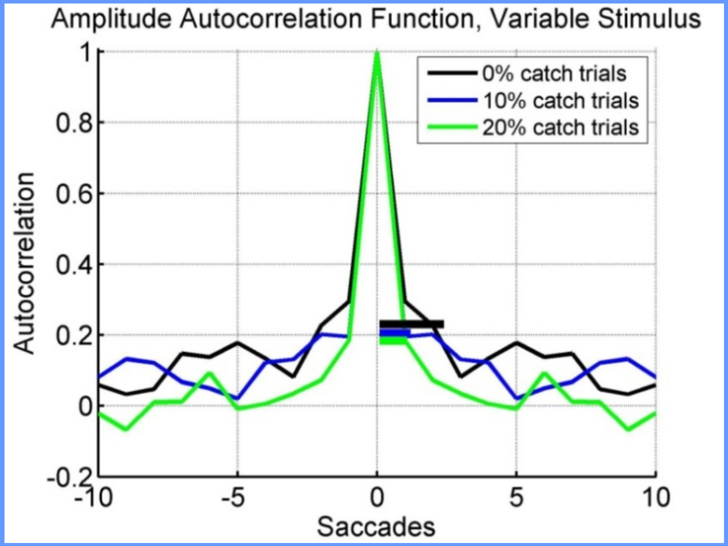
- Extended tracking increases correlations



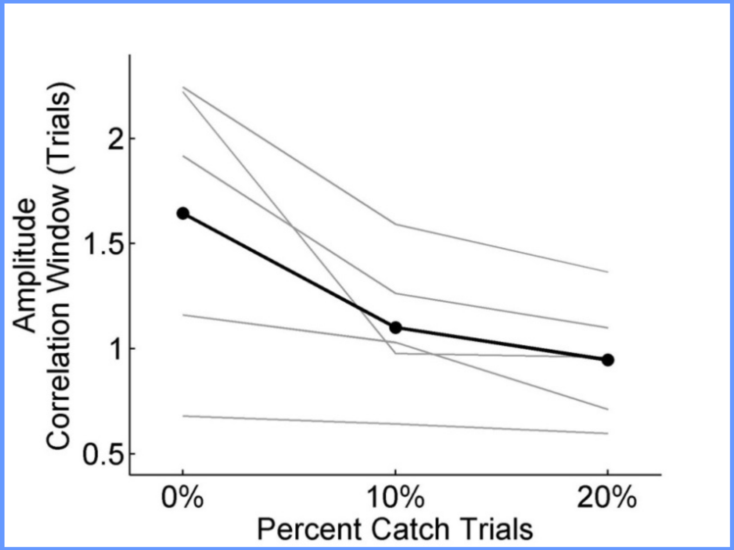


Correlation window modulates with stimulus reliability

- Stimulus variability decreases correlations



- Consistent with Bayesian optimization

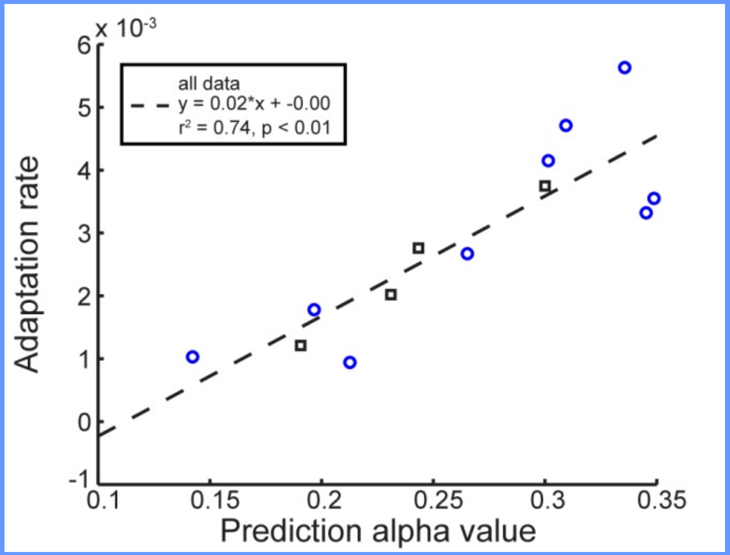


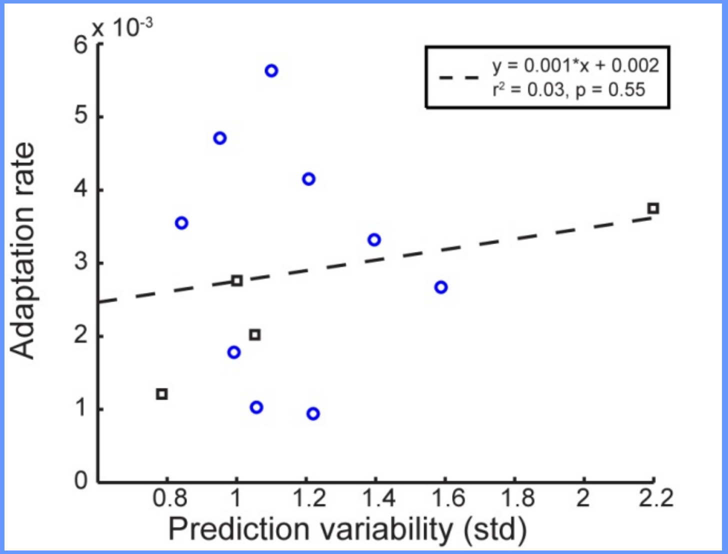
So What?

- Do these fractal fluctuations have any value or meaning?
- Model suggests fractal structure is only apparent.
- Prevailing paradigm for 1/f noise:
  - System on *edge of criticality*
  - Maximizes flexibility while maintaining organization
  - Can we test this?

Prediction and Adaptation

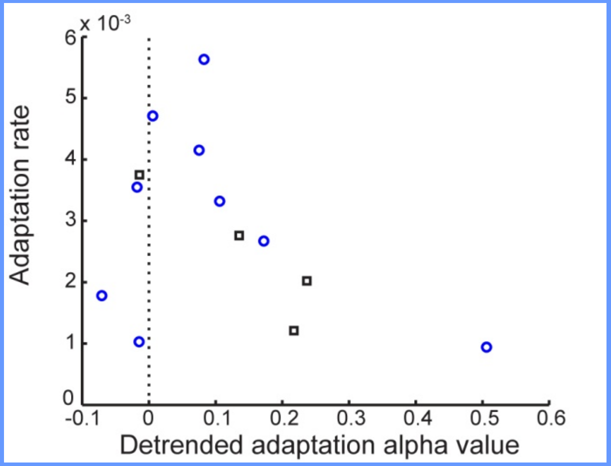
- In the same session, subjects performed a predictive-saccade and adaptation task
- Prediction α values and adaptation rates are tightly correlated
- Adaptation rate is not related to variability per se, but to the structure of the fluctuations





What is the fractal structure *during* adaptation?

- Remove linear adaptation trend:
  - Fractal structure of residuals increases as adaptation rate decreases
  - Adaptation faster if more of the underlying fractal process is engaged



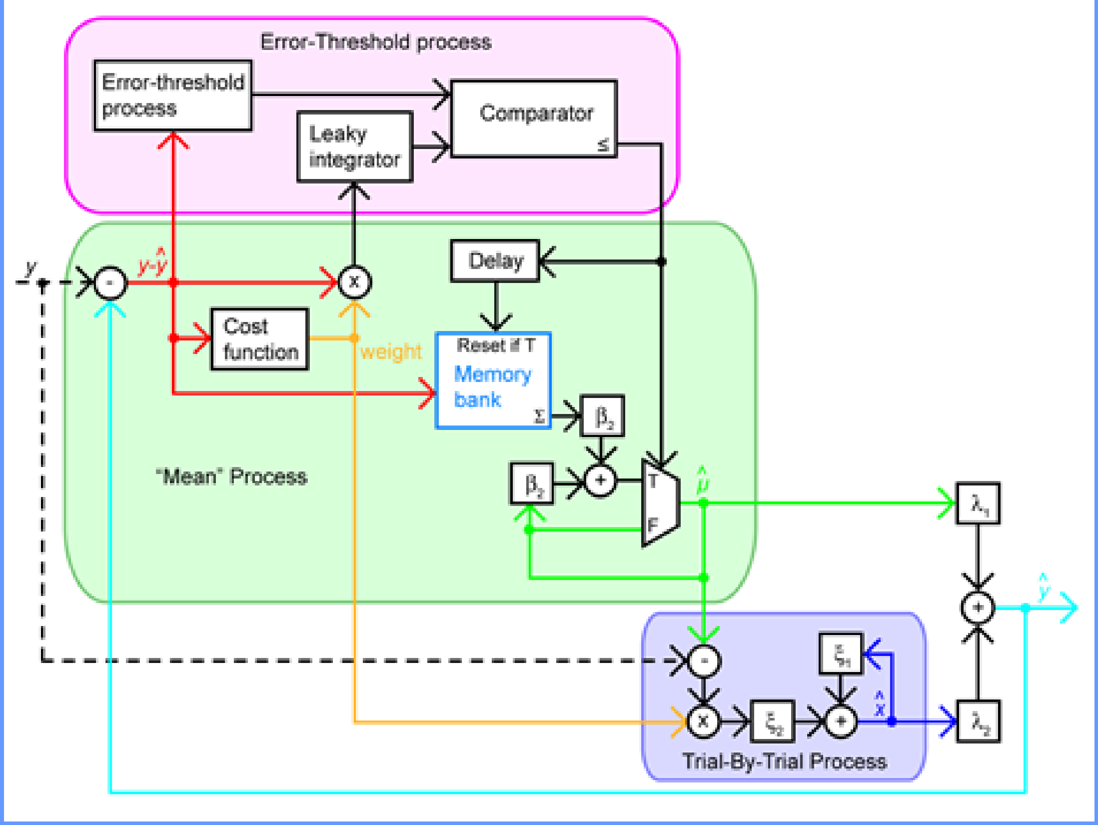
Possible meaning of fractal correlations in motor control

- Optimize "new information" with each trial, constrained by the need to retain information between trials.
- "…1/f noise represents an optimal compromise between efficient transfer of information (maximized by white noise) and immunity to error." – Gilden et al. 1995

How to Model this?

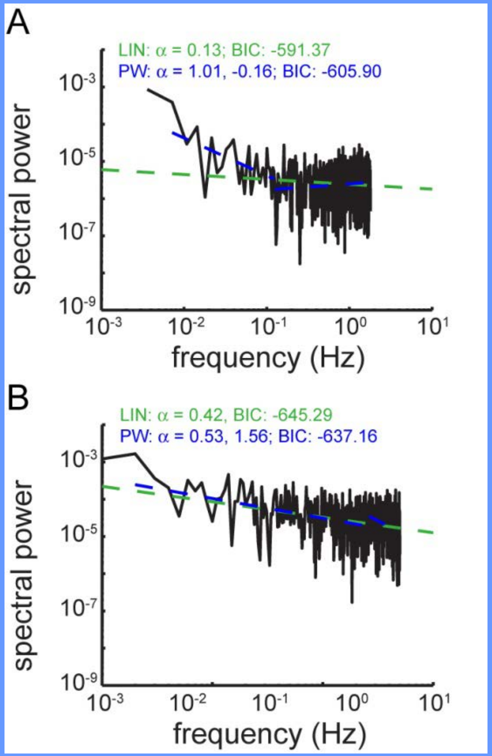
- True long-memory implies storage over very long time periods
  - Not so realistic for biological systems
- Other modeling approaches:
  - Sum of simple linear systems with wide range of time constants.
  - Fast (trial-to-trial) process with learning parameter modulated by slower process.

Model for Amplitude Control



Comparison to Two-State Model

- Popular model for motor learning:
  - Slow process: adapts slowly, forgets slowly
  - Fast process: adapts quickly, forgets quickly
  - Explains savings
  - Smith et al., Interacting adaptive processes with different timescales underlie short-term motor learning. PLoS Biol 4: e179, 2006.
- Does <u>not</u> reproduce 1/f spectrum

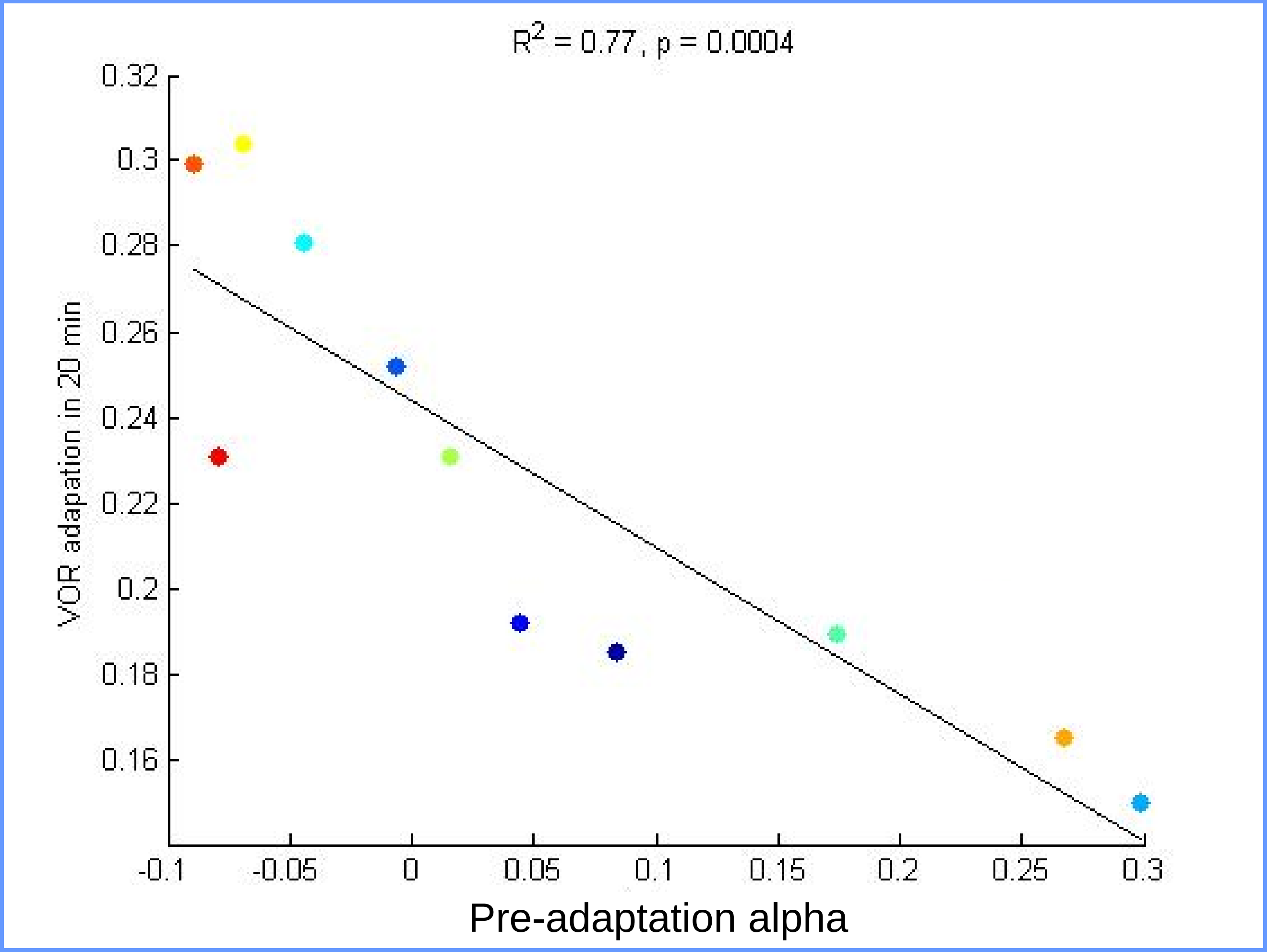


Vestibulo-Ocular Reflex

VOR Paradigm

- Before adaptation
  - 420 active head yaw movements
  - Viewing a point target
  - Gain measured for each movement
  - Gains form a time series. Find slope of power spectrum.
- Adaptation
  - 0.5X lenses, 20 minutes
  - Active sinusoidal head rotations, 1.5 Hz, in the light
  - Adaptation extent = difference between baseline and post-adaptation gains

Inter-trial Correlations Predict VOR Adaptation



Differences from Saccades

- Association is in the “wrong” direction
  - Weaker correlations correspond to better VOR adaptation
- Unlike saccades, VOR error (retinal slip) is available constantly and in real time
  - No need to store error from one movement to the next
- Best strategy may be to forget previous error as quickly as possible since it represents outdated information
  - Make maximum use of current information
  - → Reduced correlations are better

Vestibular Afferents

Teich and colleagues, 1980:

- Fractal scaling in spontaneous firing activity of cat *auditory* afferents
- Not in *vestibular* afferents.
- Source ?
  - Ion channels in cell membranes open and close in a fractal pattern
  - Resulting ion currents are fractal, leading to fractal firing patterns.

- Vestibular afferents
  - Mice
  - Resting activity
  - Fractal behavior only on longer time scales

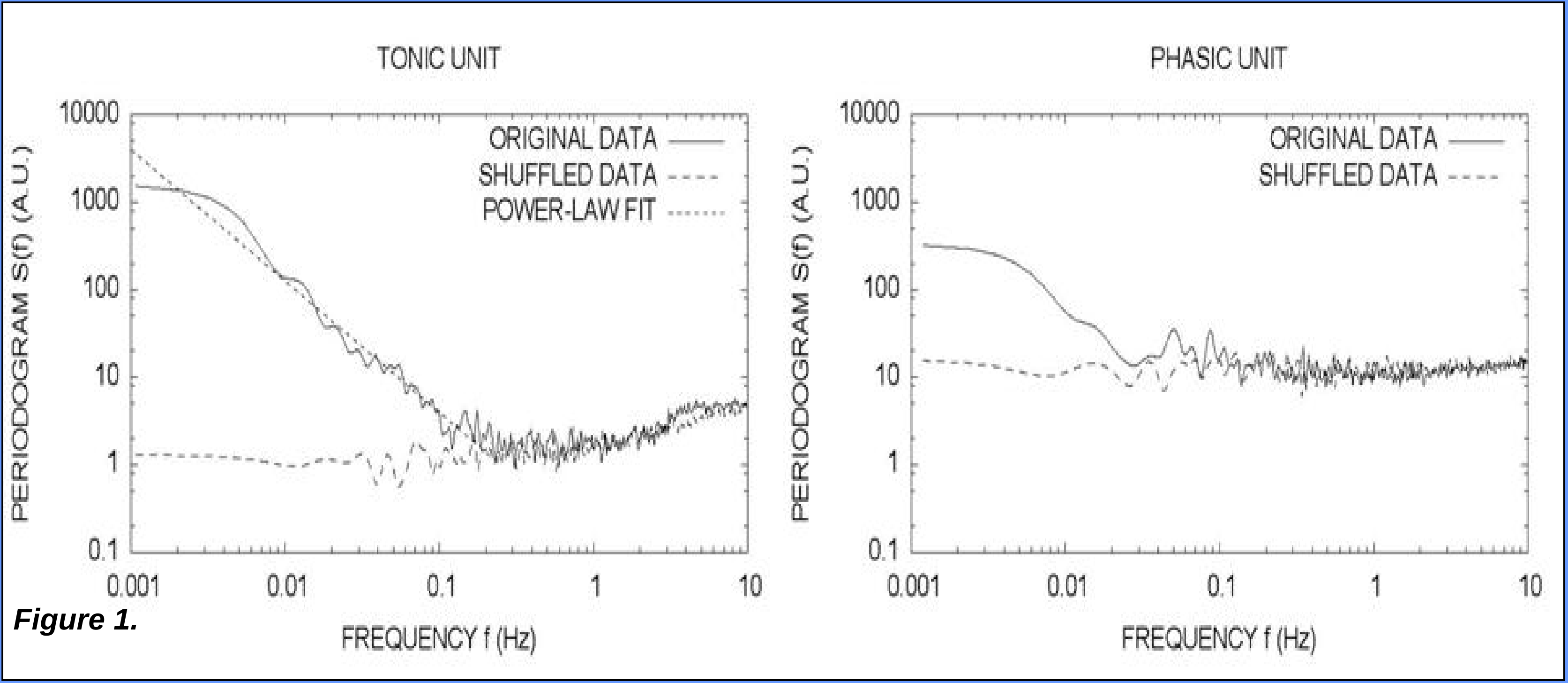
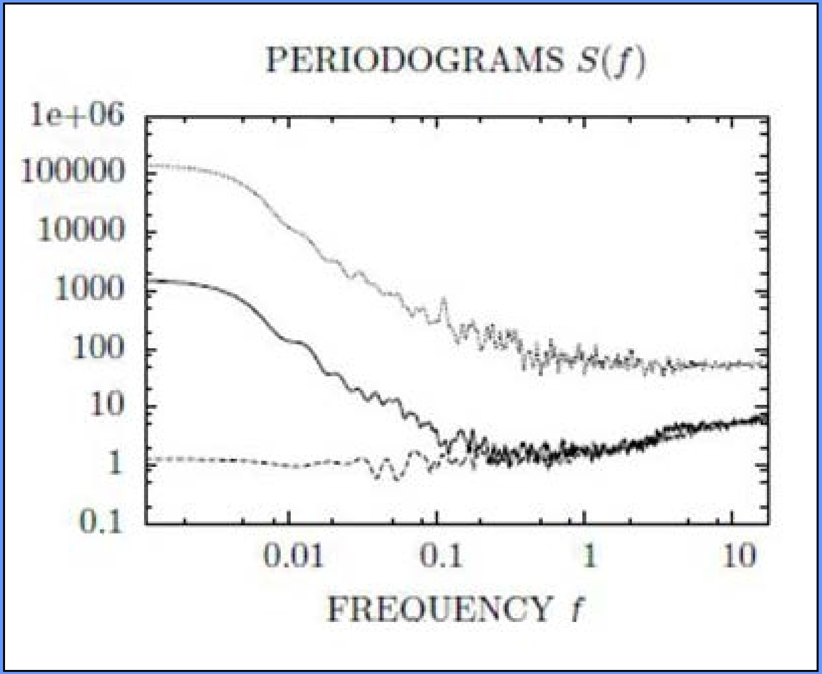


Figure 1.



Surrogate analysis

- Solid lines are from the afferent data.
- Lower dotted lines: randomized-order surrogates, (null hypothesis: point process, independent intervals).
- Upper dotted lines: random values from exponential distribution, retaining temporal ordering of ranks of intervals (null hypothesis: results arise from ordering of intervals, regardless of distribution).

Functional relevance

- Fractal structure apparent only <~1.0 Hz, >1.0 sec
    - beyond range of normal head movements

- Fractal structure is "permitted" outside the range of normal movements.

- Value of fractal structure might be in long-term behavior – several seconds or longer. →

Velocity storage

• Cupula and afferent time constant (~6 sec) extended to time constant of the VOR (~20 sec).

• Afferent information relevant to velocity storage (1-10 sec or more) within the fractal regime: “partially integrated.”

• Balance between quick action for transient head motion and perseveration (storage) for longer movements.

Compensation for unilateral deactivation

• Acute imbalance in tonic activity, corrected minutes - days.

• Long-range correlations might help track long-term activity to maintain balance.

Neural integrator function

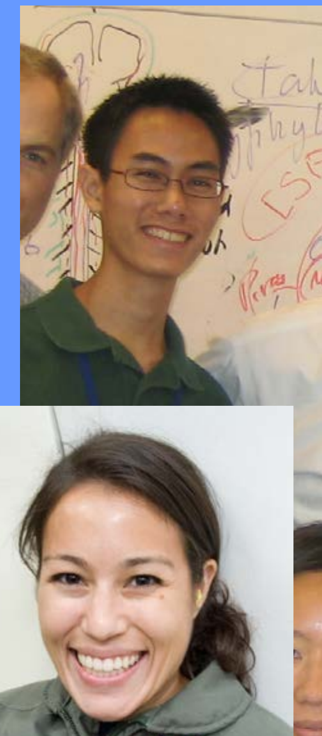
• Velocity information in premotor neurons transformed to a signal with position component to drive motoneurons.

• Integrator must avoid integrating baseline firing rate.

• Correlations could help integrator track baseline rate.

Acknowledgements

David Zee, Dale Roberts, Adrian Lasker,
Andrew Zorn, Michael Schubert, Steve Lowen



Wilsaan Joiner

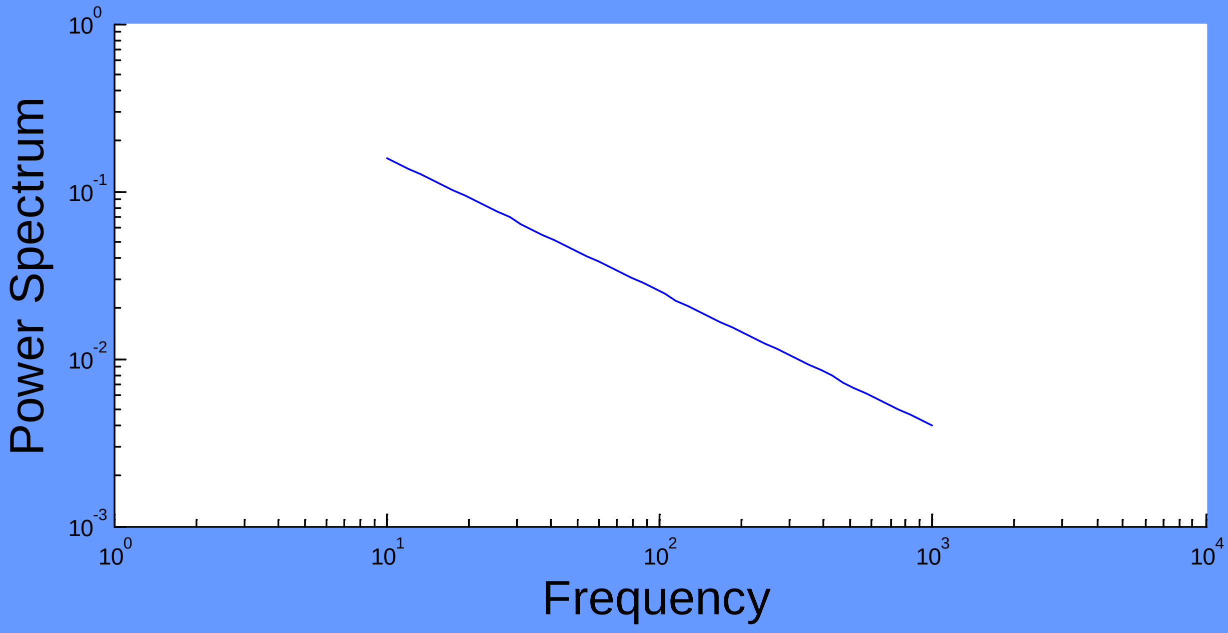
Aaron Wong

Kara Beaton

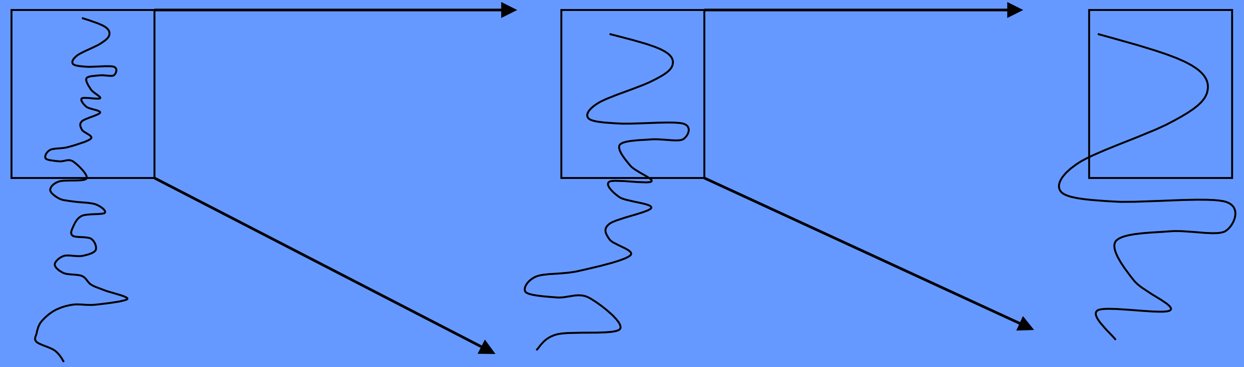
NIH grant EY15193, NSF grant BCS-0615106
Society for Neuroscience Minority Fellowship Program

How to measure variability in a time series

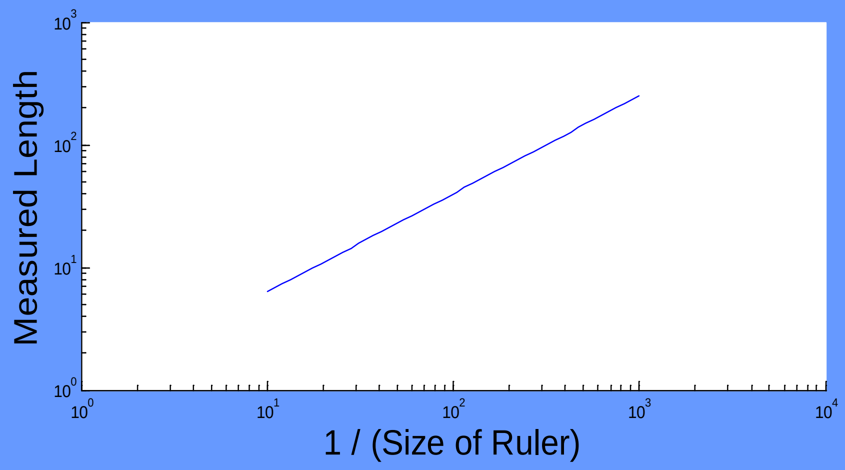
- Fractal time series has variability that changes with size (duration) of measurement window – this leads to a *statistical* rather than *deterministic* fractal.
- Reflected in frequency domain as power-law decay:
  - $S_{xx}(f) \sim 1/f^{\alpha}$
  - shift in frequency by Δf changes power by $(\Delta f)^{-\alpha}$
- Substantial low-frequency energy:
  - implies long-term correlations.
  - makes parameter measurement difficult.



How to measure the length of a coastline



- Coastline is a *fractal* – variation on many length scales [Mandelbrot].
- Measured length changes as length of “ruler” changes:



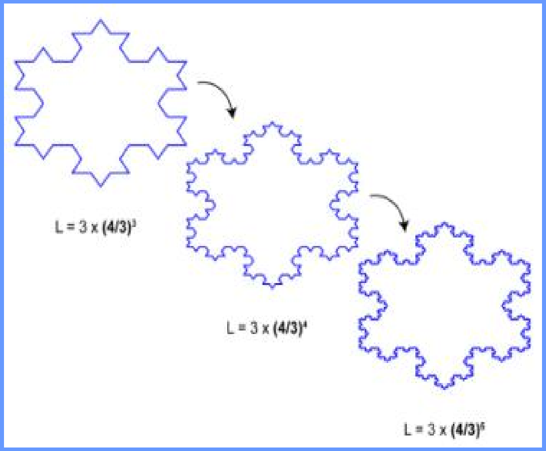
- No characteristic length scale produces the “correct” answer.
- Variation in measured coastline length with ruler size is best described by a *scaling law*.
- Scaling law yields fractal exponent (slope on log-log plot), indicating *power-law* relationship:

(Measured Length) = (Ruler Size)$^{-\alpha}$

What is a fractal?

A fractal is self-similar:

repeated pattern at different scales

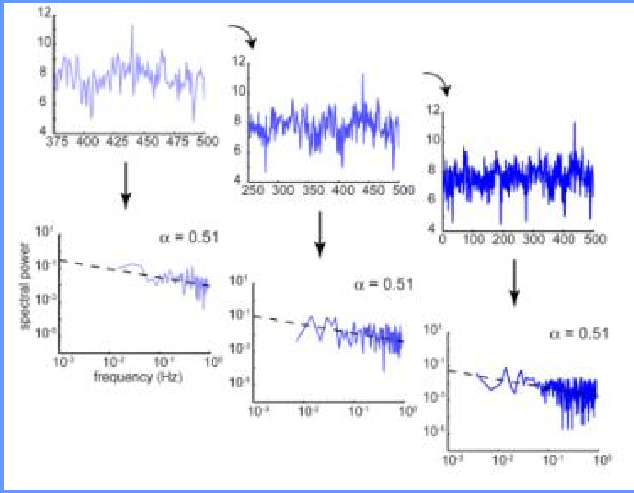


Geometric self-similarity

Scale invariance (power-law scaling relationship)

$$f(\lambda x) = \lambda^k \cdot f(x)$$

$$L \propto \varepsilon^{-D}$$



Statistical self-similarity

Power spectrum scales as a power law

$$Sxx(f) \sim f^{-(2H+1)}$$

Dimension - Scaling

- Power-law scaling
  - $N \propto \varepsilon^{-D}$
  - $\log(N) = \log(K\varepsilon^{-D}) = \log(K) - D \log(\varepsilon)$
  - Plot of $\log(N)$ vs. $\log(\varepsilon) \rightarrow$ slope = D

CONCLUSION

Horizontal saccades: dot versus vertical line

Vertical line exaggerates persistence in vertical direction

- Less certainty about orthogonal "error" affects inter-trial learning
- This increases residual information remaining in the orthogonal endpoint errors
- No change in primary saccade accuracy requirements
- Subjects exhibit a greater difference between primary and orthogonal endpoint-error behavior

Oblique saccades

- Primary and orthogonal directions show similar trend (N/S).
- Due to the increased variance of oblique saccades?

Another form of scaling

ANALYSIS IN THE TIME DOMAIN

- **Hurst exponent *H***

Quantify variability over different time scales

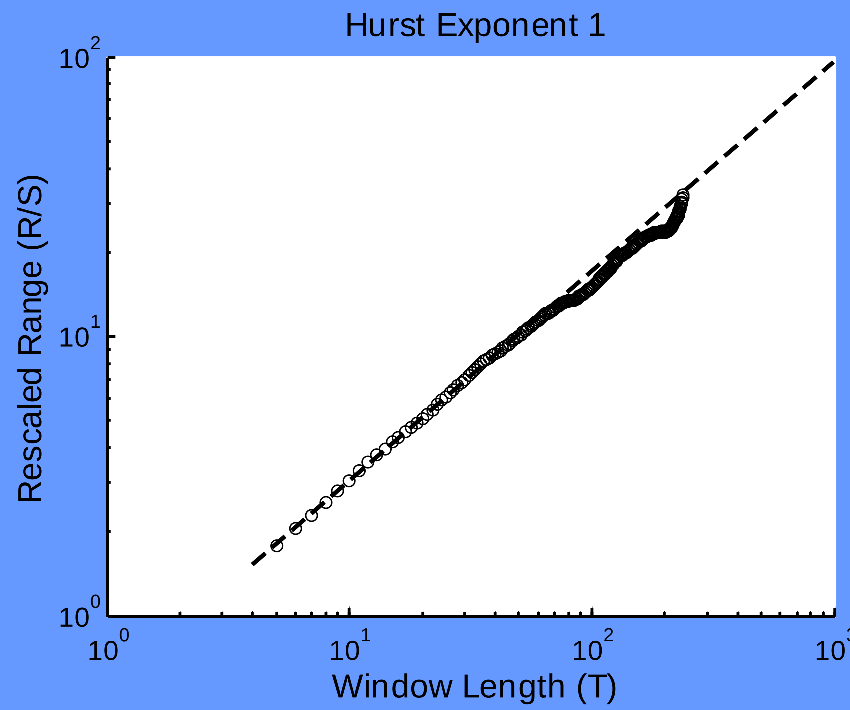
(Variability) = (ΔT)H

Variability measured in an unusual way

R e sa c n = a g l e = R/S a D Rn/S c

H and *α* should agree

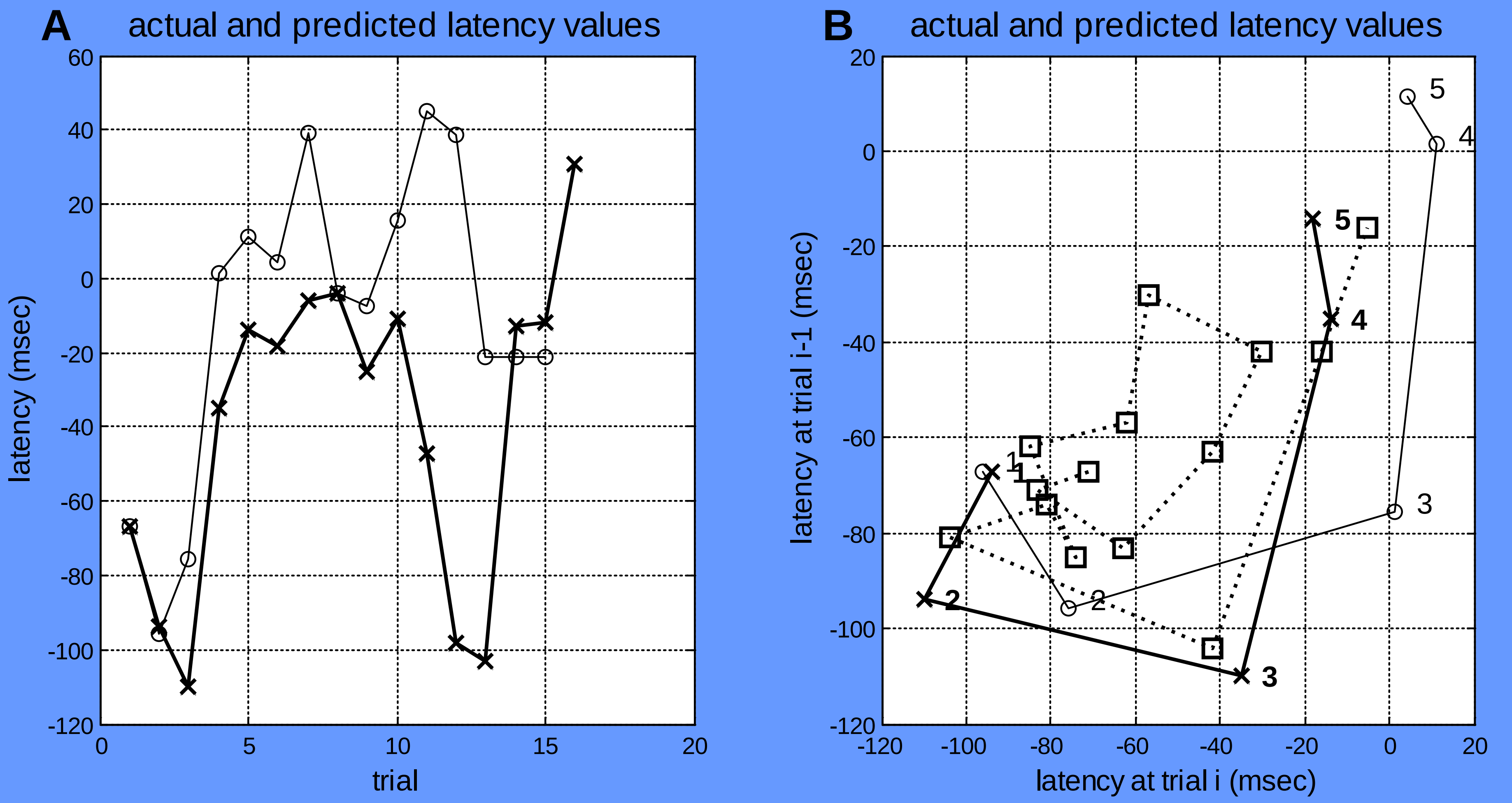
$$H = \frac{1+\alpha}{2}$$



Summary *α* AND *H*

	Horizontal saccades *Target: dot*	*Vertical saccades* *Target: dot*	*Horizontal saccades* *Target: vertical line*	*Oblique saccades* *Target: dot*
Subject	***α* primary / *α* orthogonal** ***H* primary / *H* orthogonal**	***α* primary / *α* orthogonal** ***H* primary / *H* orthogonal**	***α* primary / *α* orthogonal** ***H* primary / *H* orthogonal**	***α* primary / *α* orthogonal** ***H* primary / *H* orthogonal**
1	0.002 / 0.264 * 0.664 /0.681	-0.006 / 0.469 * 0.605 / 0.693	0.052 / 0.506 * 0.664 / 0.707	0.237 / 0.369 * 0.633 / 0.692
2	0.018 / 0.158 * 0.648 / 0.747	0.079 / 0.215 * 0.637 / 0.698	0.276 / 0.456 * 0.652 / 0.822	0.077 / 0.326 * 0.592 / 0.691
3	0.071 / 0.551 * 0.615 / 0.824	0.425 / 0.155 * 0.718 / 0.690	-0.044 / 0.717 * 0.640 / 0.750	0.476 / 0.178 * 0.684 / 0.629
4	0.219 / 0.275 * 0.687 /0.711	-0.001 /0.198 * 0.643 / 0.665	0.137 / 0.725 * 0.651 / 0.785	0.210 / 0.386 0.802 / 0.651
5	0.286 / 0.346 0.643 /0.622	0.183 /0.208 * 0.569 / 0.627	0.324 / 0.610 * 0.714 / 0.838	0.017 / 0.204 * 0.651 / 0.737
6	-0.009 / 0.417 * 0.585 /0.722	0.303 / 0.484 * 0.643 / 0.665	0.139 / 0.550 * 0.660 / 0.780	0.339 / 0.315 * 0.675 / 0.654
7	0.423 / 0.218 * 0.697 / 0.694	0.088 / 0.631 * 0.701 / 0.861	0.408 /0.856 * 0. 724/ 0.874	0.588 / 0.426 0.734 / 0.754
9	-0.077 / 0.624 * 0.590 / 0.790			
10	0.163 / 0.488 * 0.582 / 0.744			

Nonlinear forecasting

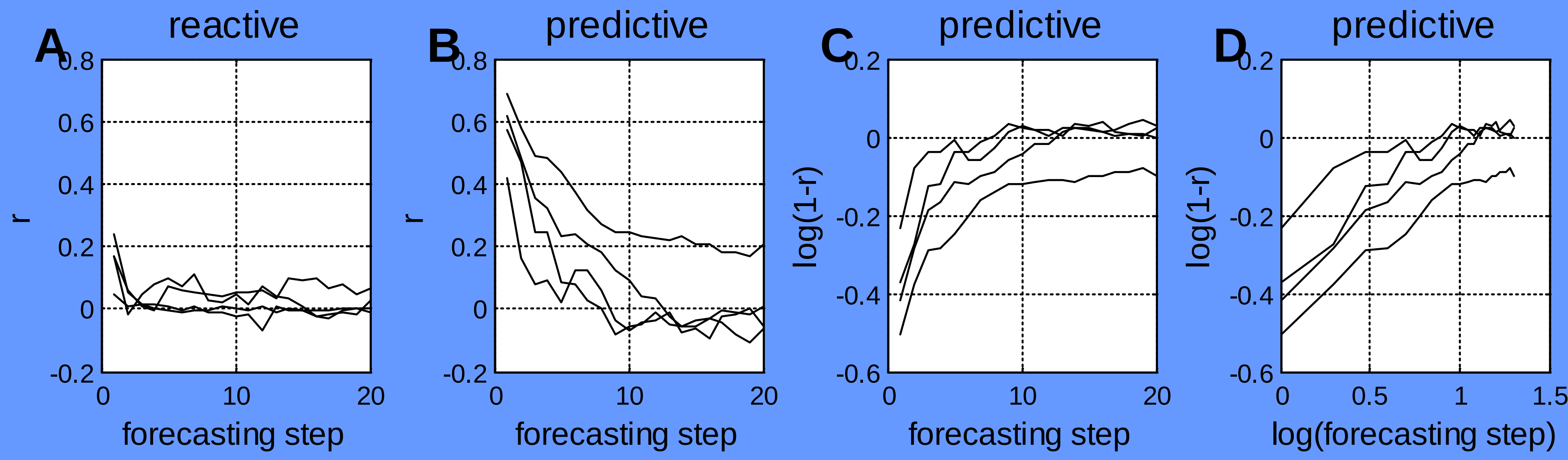


Forecasting suggests that scaling is fBm

For fBm:

• log(1-r) vs. log(forecasting step) is linear

• AA Tsonis, JB Elsner (1992) Nonlinear prediction as a way of distinguishing chaos from random fractal sequences. *Nature* 358: 217-220.

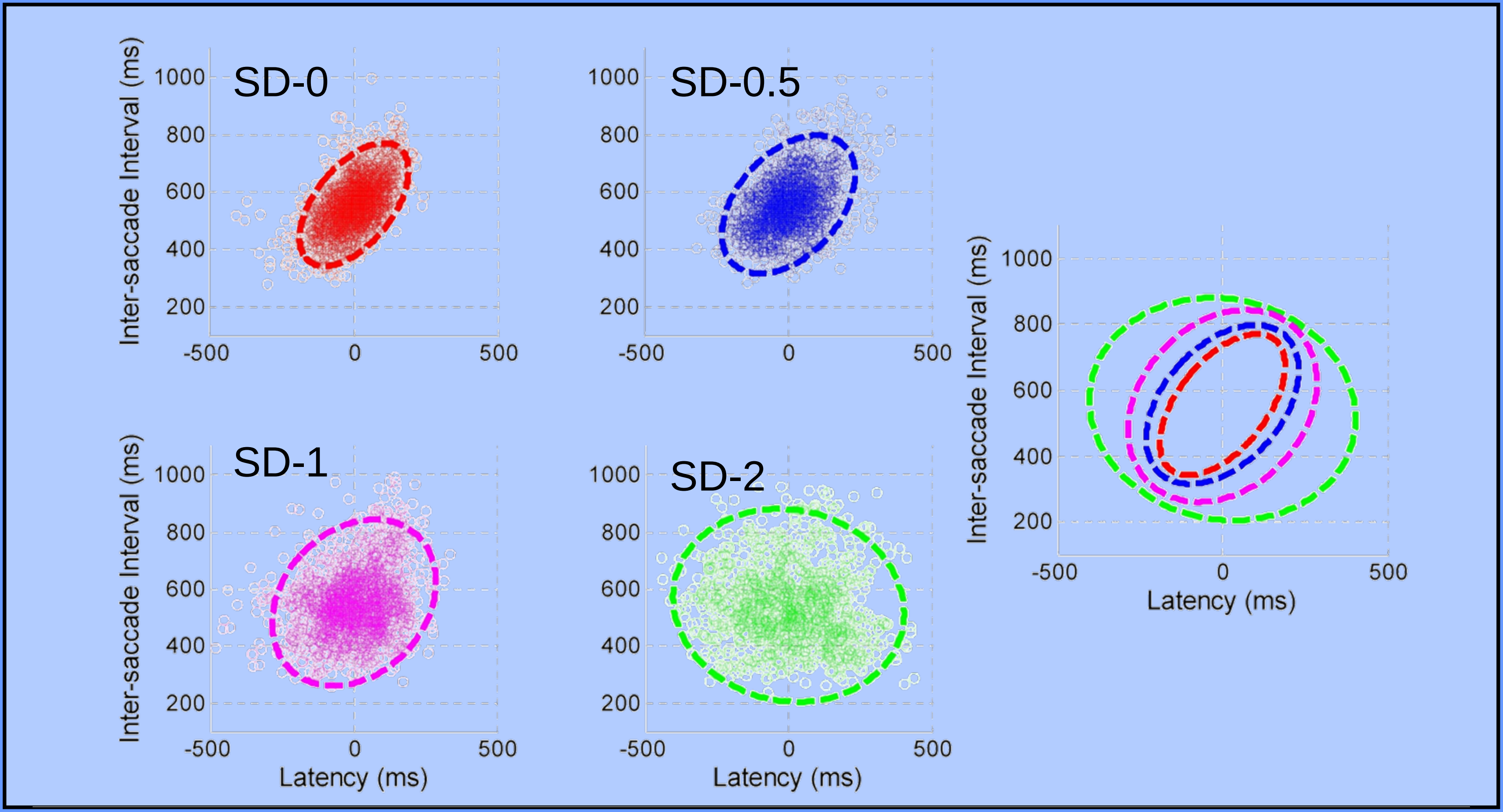


No-Correction Zone

- Range of intervals and latencies which are considered **adequate**.
- When most recent intervals and latencies are in this **no-correction-zone** there is no effect of error on timing of subsequent saccade.
- Size of **no-correction-zone** increases with stimulus-timing variability.

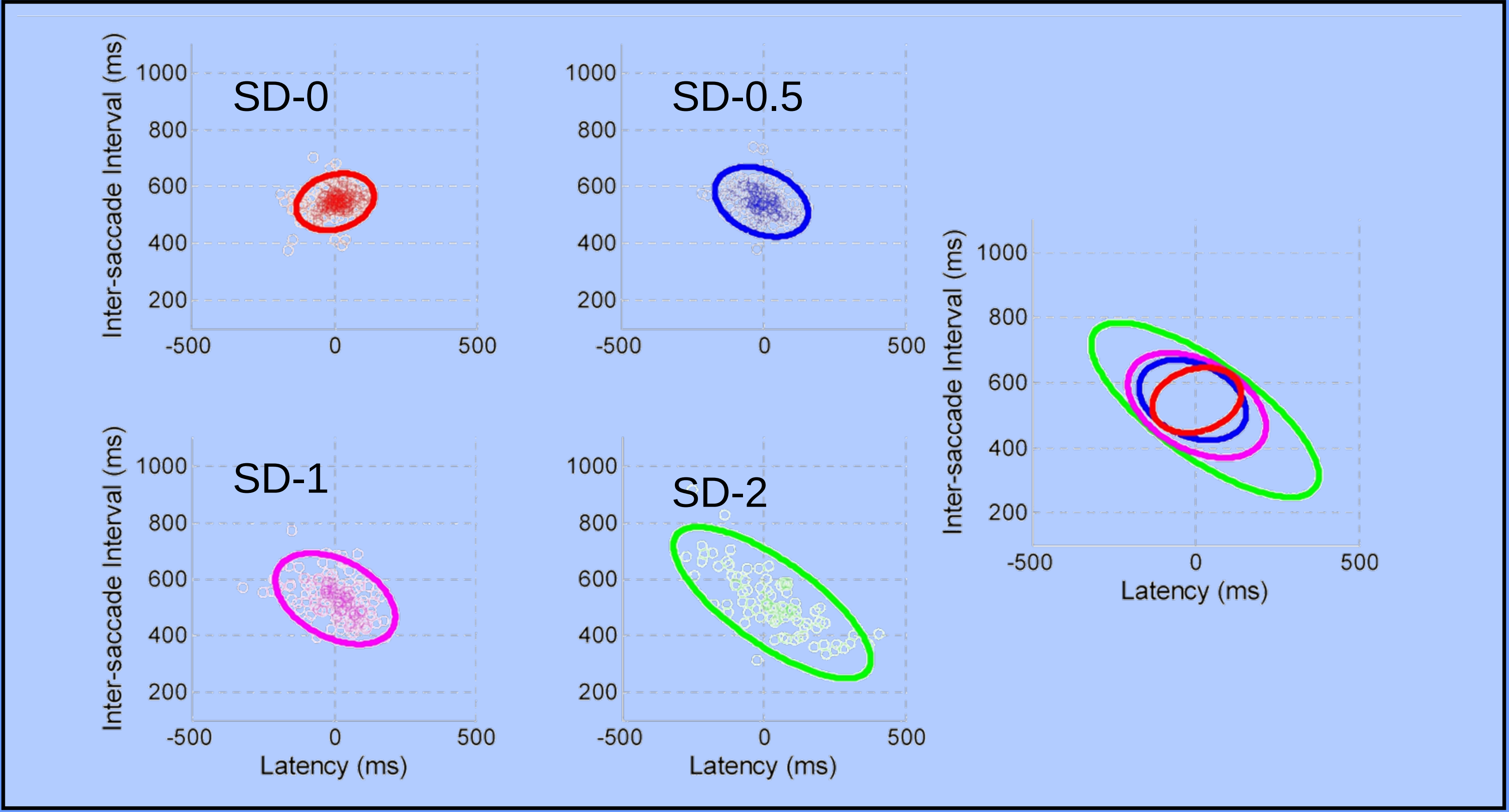
Latency vs. Preceding Inter-Saccade Interval

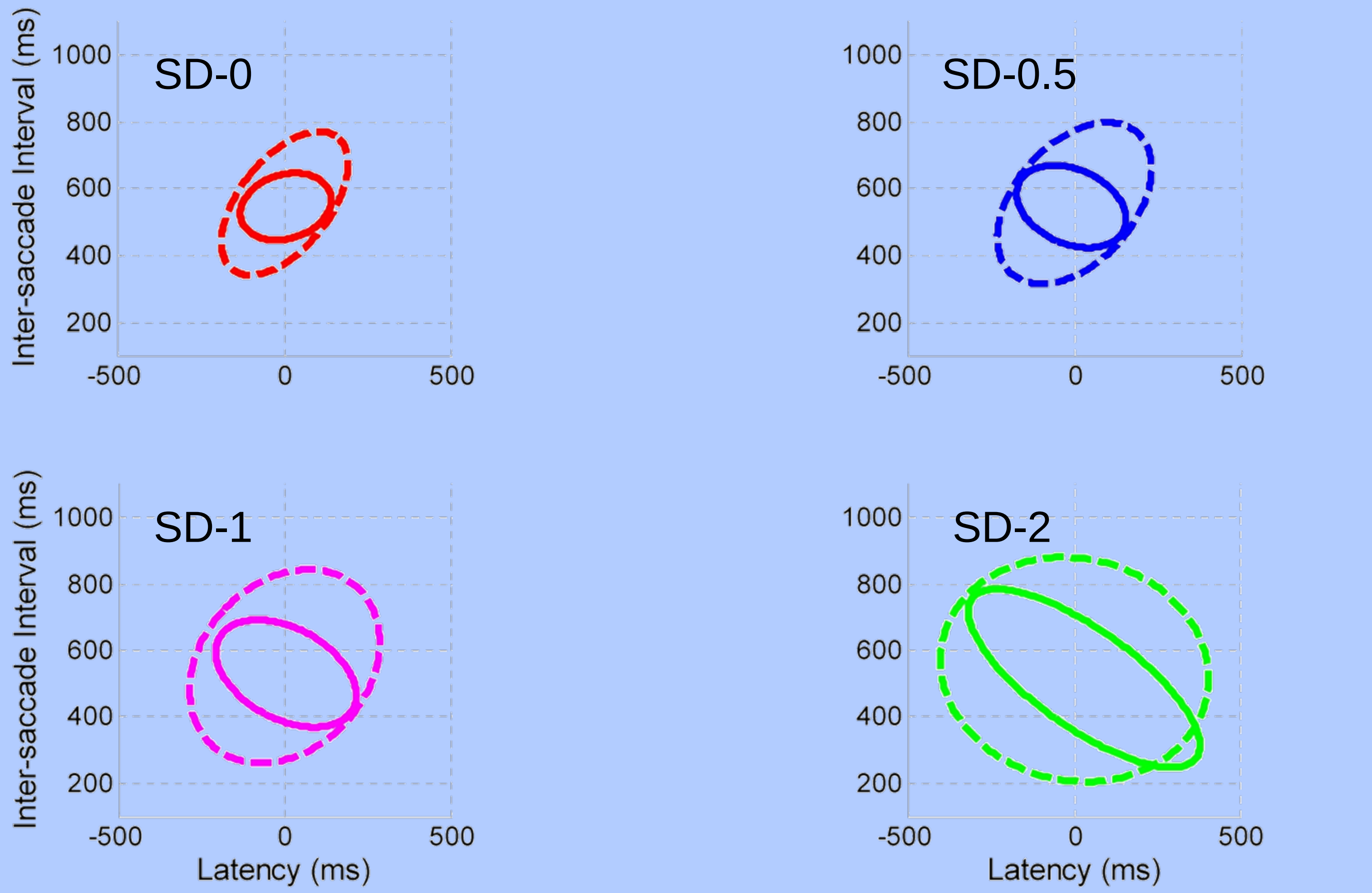
- all data
- four different levels of stimulus variability



Latency vs. Preceding Inter-Saccade Interval

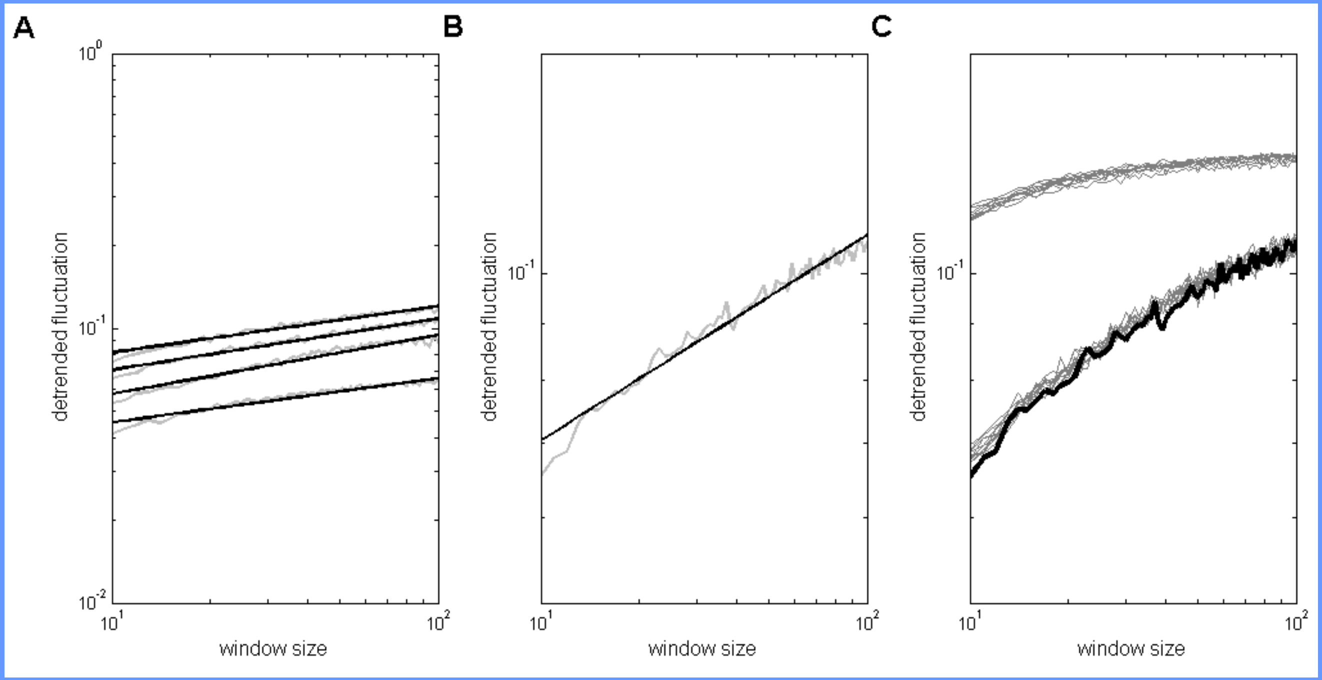
- only trials where there is little or no interval correction on subsequent trial
- four different levels of stimulus variability




SD-0
SD-0.5
SD-1
SD-2
Inter-saccade Interval (ms)
Latency (ms)
1000
800
600
400
200
-500
0
500

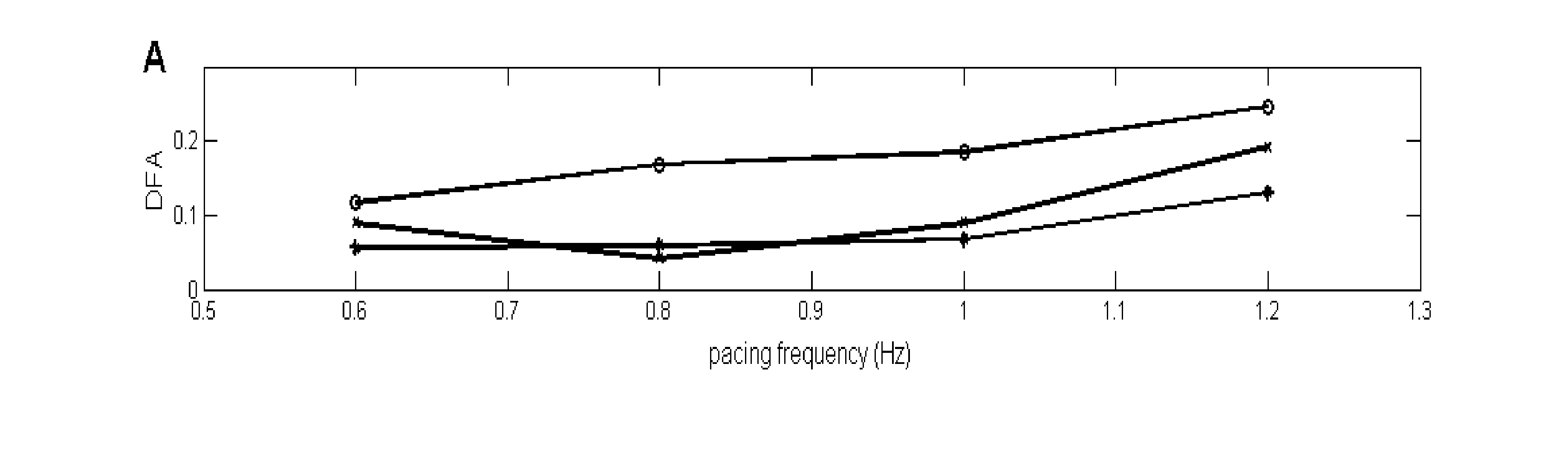
Detrended Fluctuation Analysis (DFA)

- DFA: measure of variability on different time scales
  - integrate data $x(n)$ in window of size k
  - fit line to integrated data $y(k)$
  - determine fluctuations $F(n)$, for a given window size n (length of data series), by RMS variation about the fitted line
  - slope of $F(n)$ versus n, on log-log plot, yields scaling exponent α.
- DFA on saccade latency increments.
- *Shuffle* surrogates and *phase-randomization* surrogates (C).

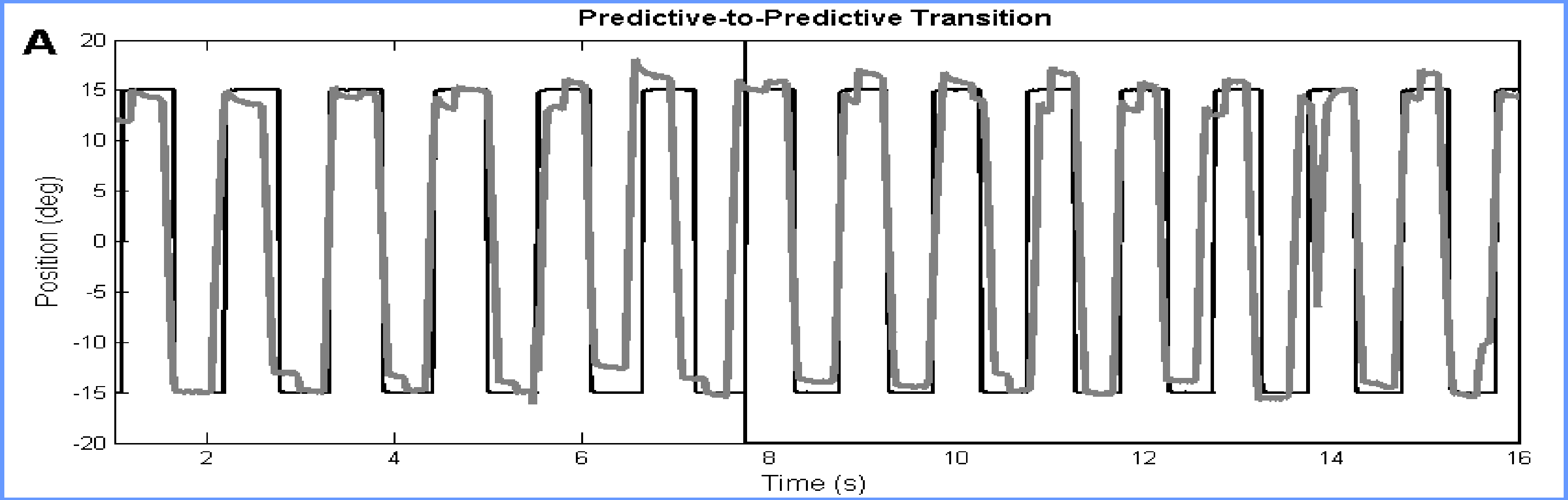


DFA on latencies

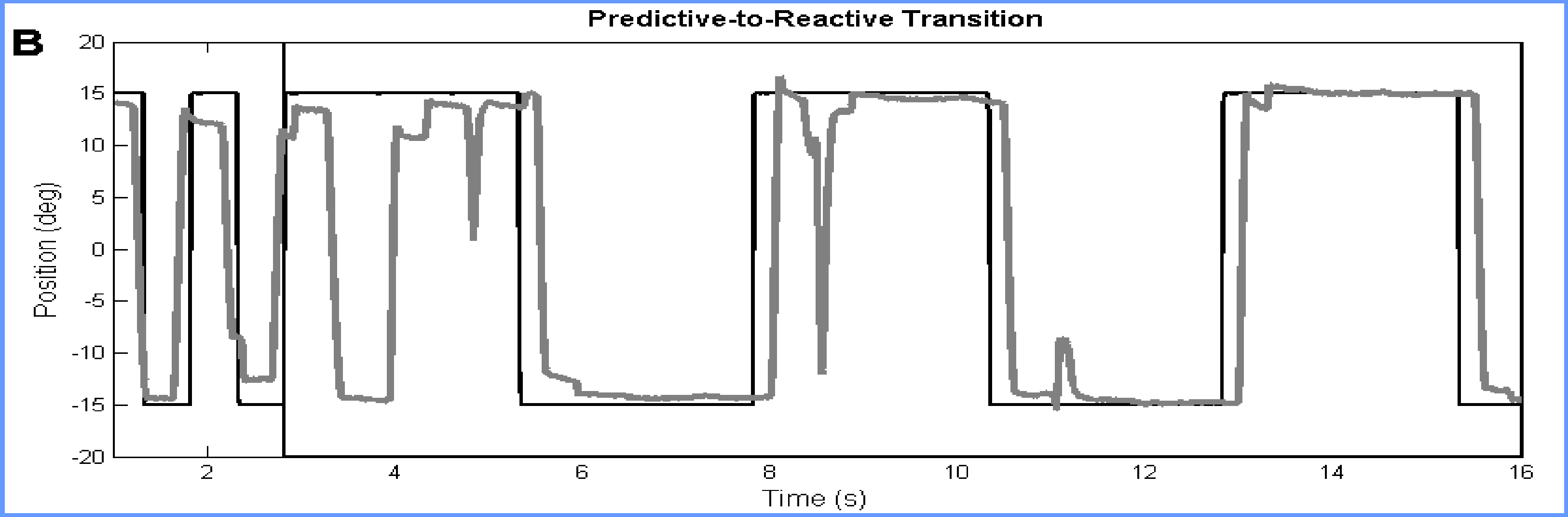
- Three subjects at four frequencies (0.6, 0.8, 1.0, 1.2 Hz).
- Scaling exponents increase with increasing frequency.
  - inter-trial correlations become stronger, and extend over more trials, as pacing frequency increases.
  - increase in correlations at higher frequencies suggests that the predictive tracking system can adjust parameters as necessary for stimulus conditions.



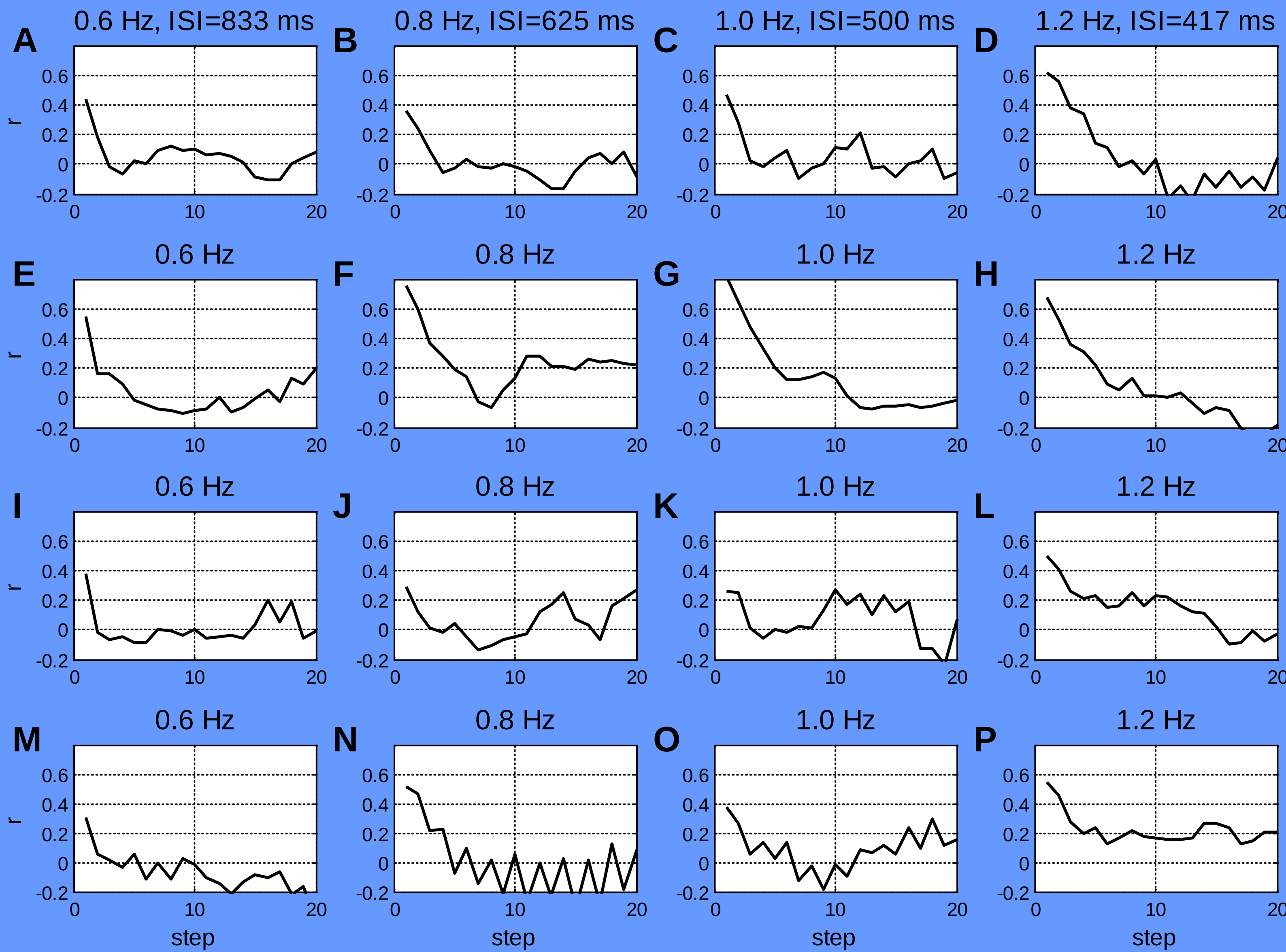
Predictive saccadic tracking: Pacing Perturbation



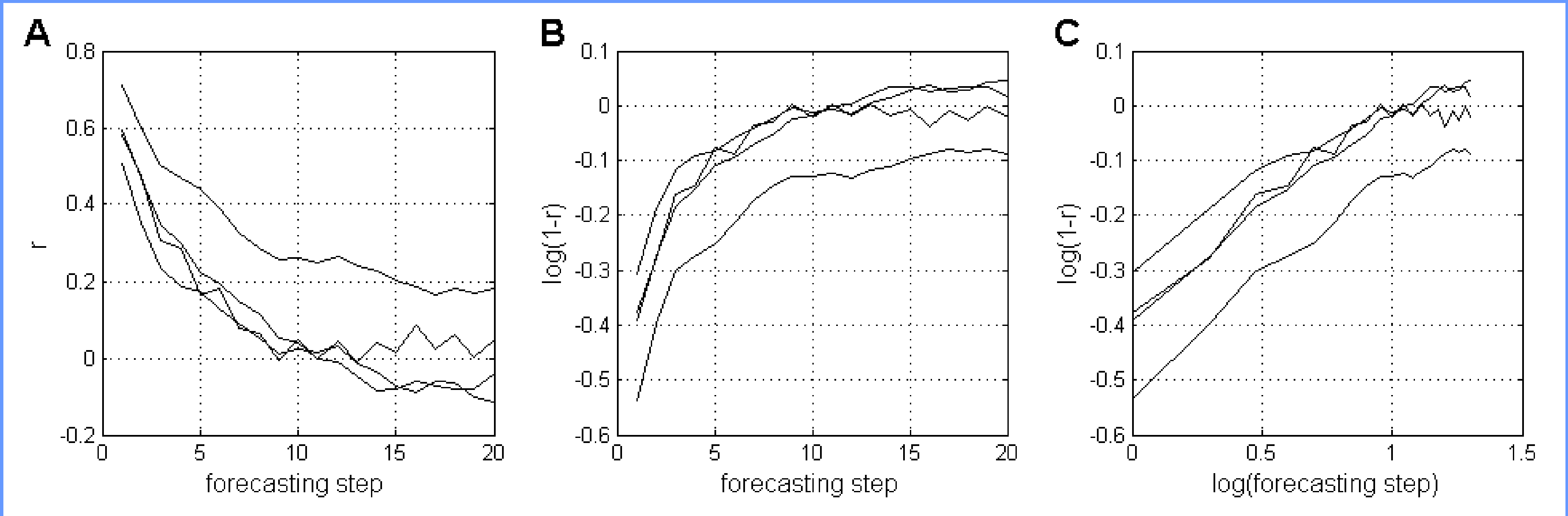
- Predictive→Predictive transition (top, 0.9 → 1.0 Hz)
- Predictive→Reactive transition(bottom, 1.0 → 0.2 Hz)



Forecasting Results

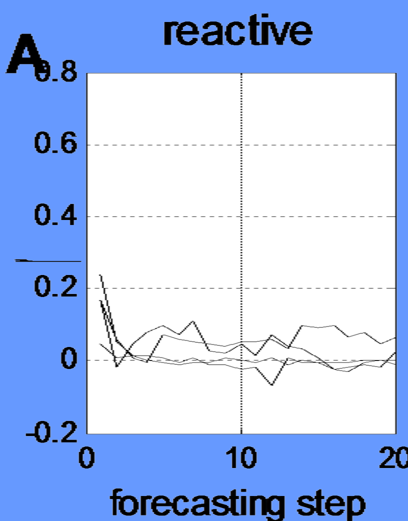


Forecasting of predictive-saccade latencies.

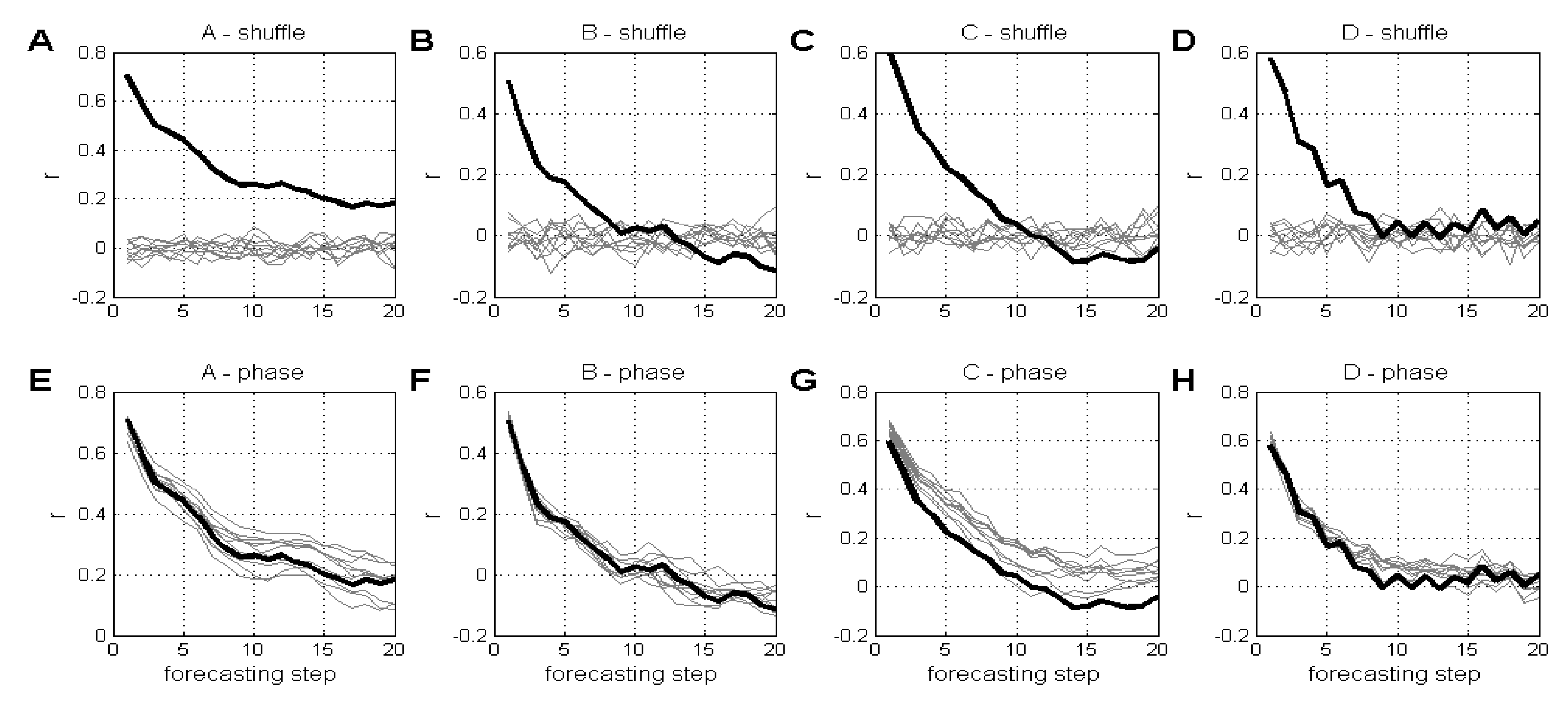


Linear decay of forecasting on log-log plot
- confirms power-law loss of information
- indicated fBm

Unable to forecast reactive saccades ➔



Forecasting of predictive-saccade latencies & surrogates



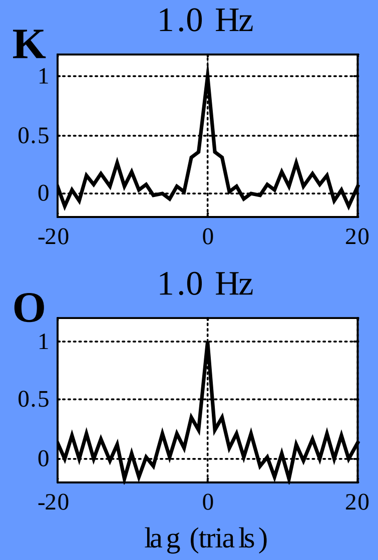
- Forecasting of *shuffle* surrogates (A-D, gray)
  - worse than forecasting of original latencies.
- Forecasting of *phase-randomization* surrogates (E-H, gray)
  - similar to forecasting of original latencies.

➔ Temporal correlations create pattern of forecasting ability

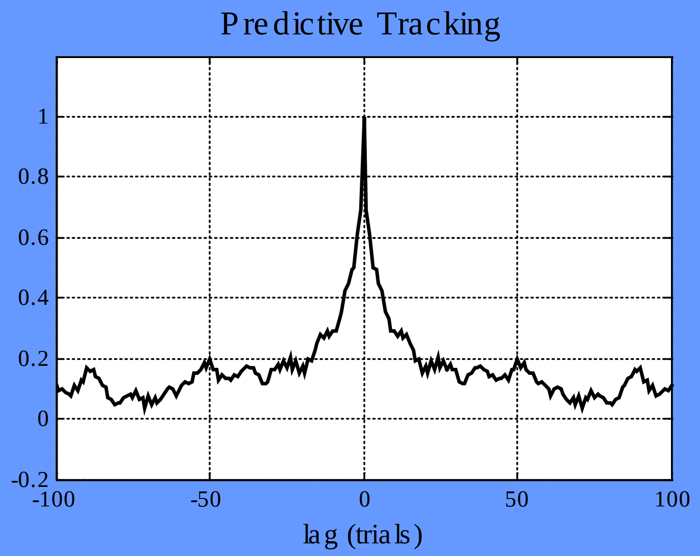
Modulation of the correlation window

- Inter-trial correlations extend further in time as pacing duration increases.
- Reflects increasing confidence in past behavior.
- Variable time course interacting with trial-to-trial error (fBm).

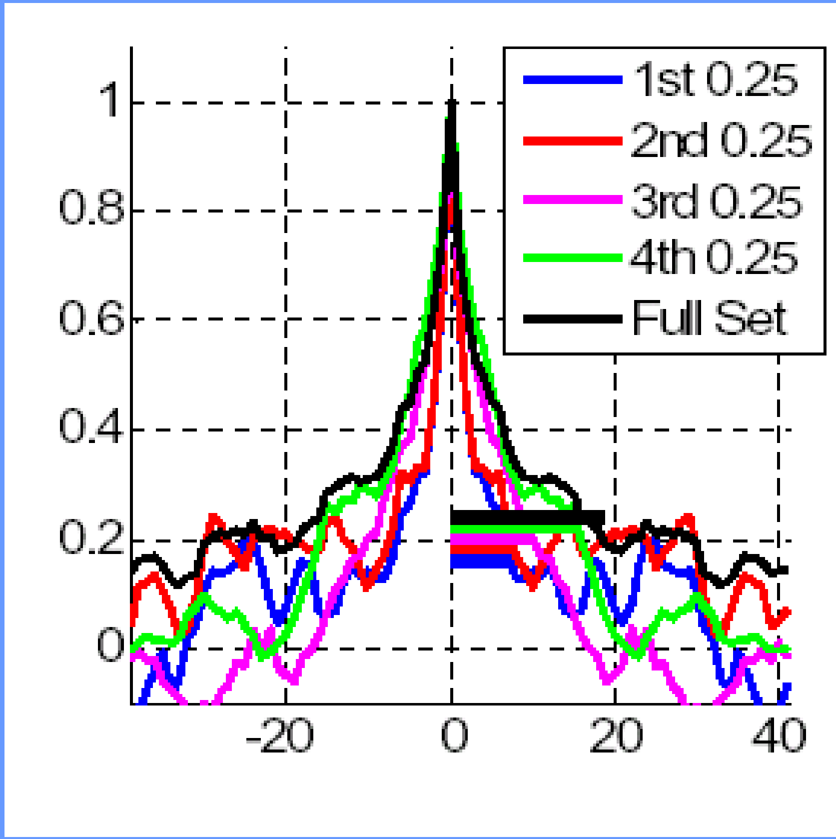
150 trials at 1.0 Hz (75 sec)
significant correlations
over 5-10 trials (2.5-5 sec)



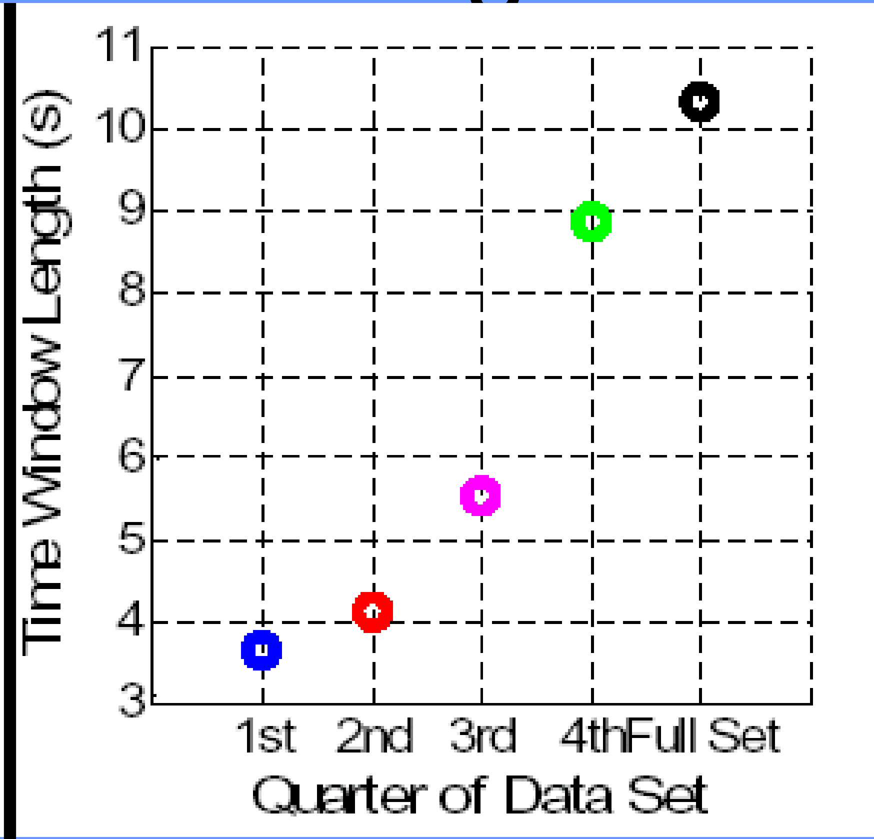
1000 trials at 0.9 Hz (556 sec)
significant correlations
over ~20 trials (11 sec)



Extended tracking

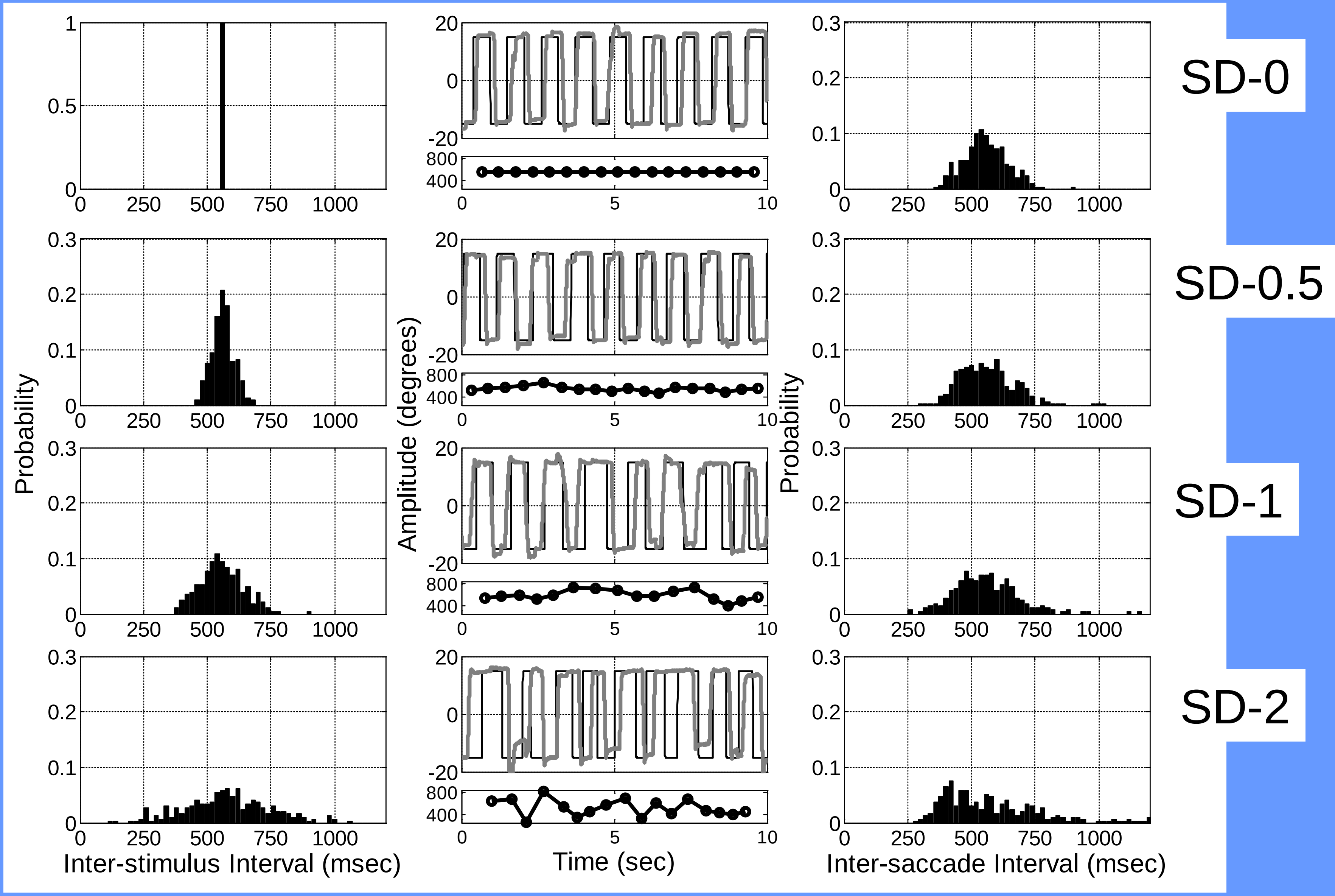


Autocorrelation functions for four non-overlapping quarters (250 points each) drawn from a set of 1000 consecutive predictive saccade latencies.



Correlation window sizes. Correlation window increases as tracking continues.

Variable Timing

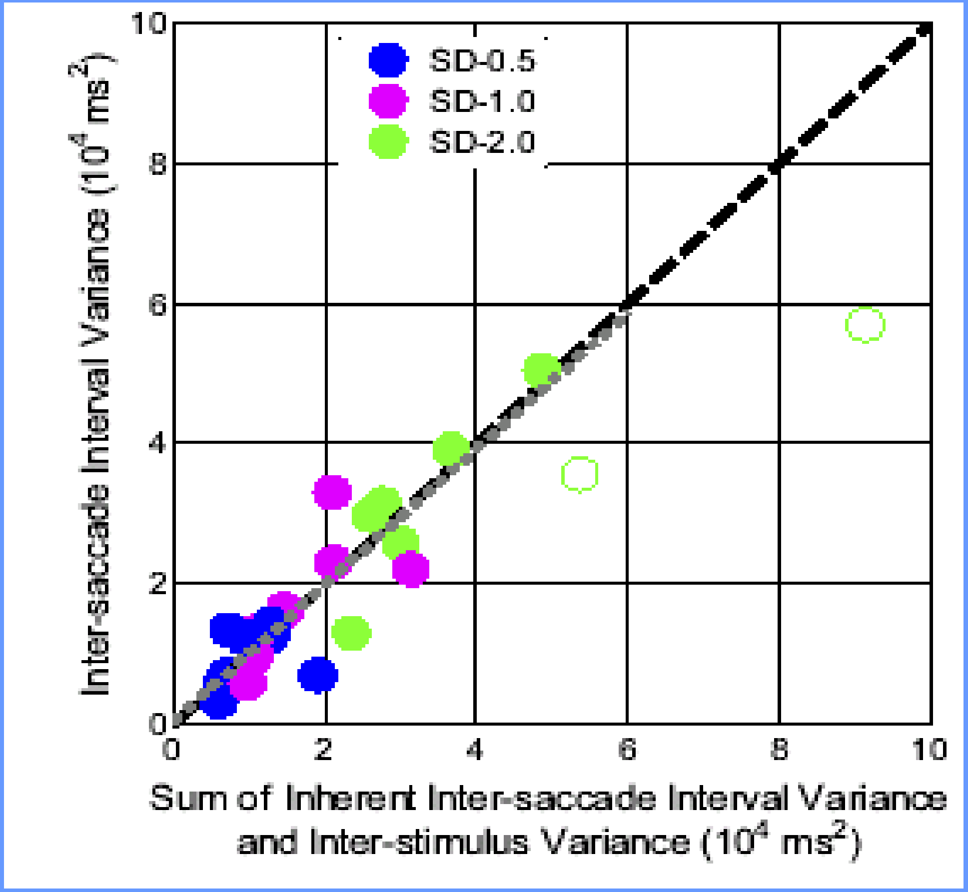


Implications of fBm

- Also found in:
  - Cardiac inter-beat intervals
  - Stride intervals while walking
- Correlations break down under pathology

Variable timing

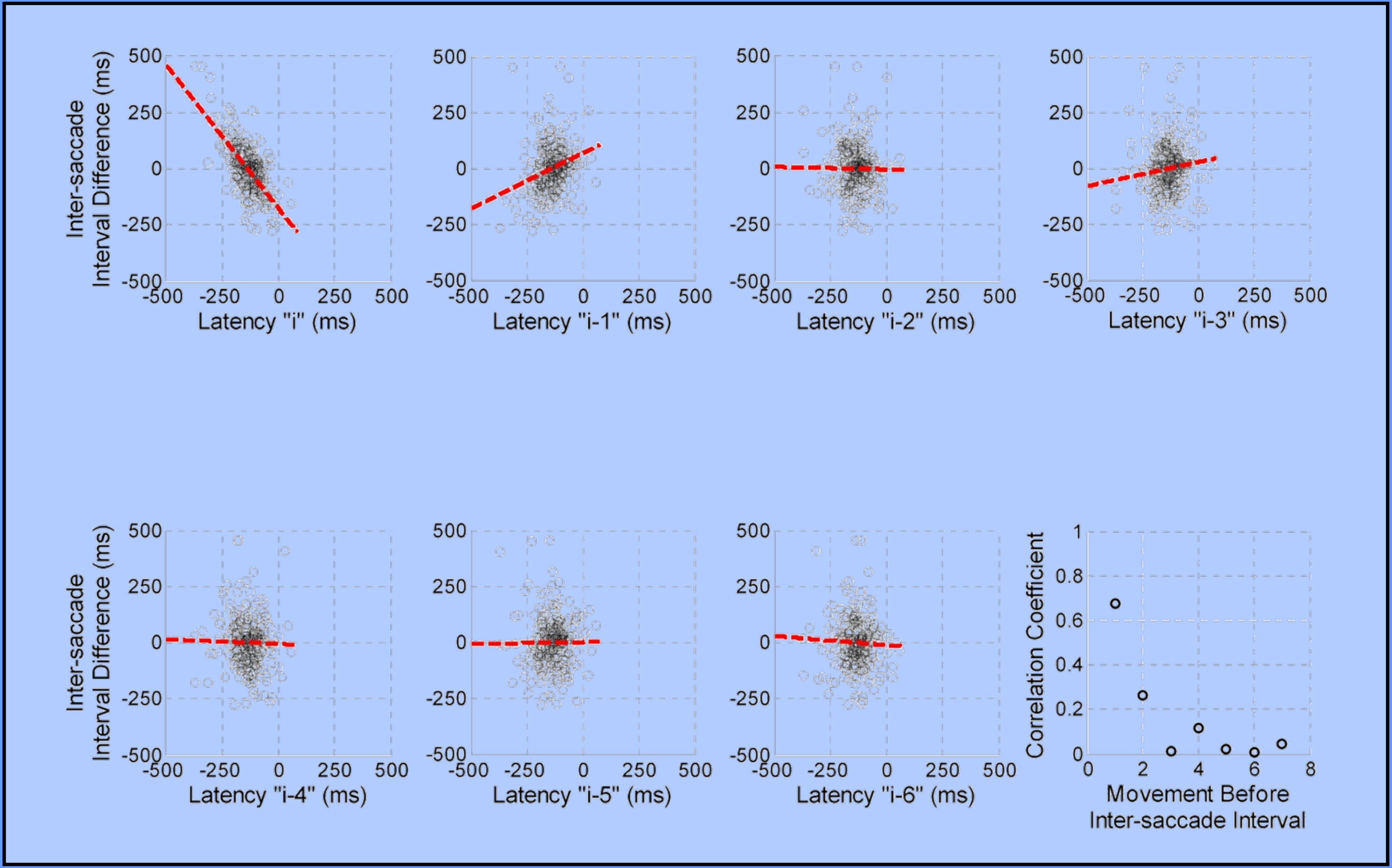
Variance of inter-saccade intervals (ordinate) is the sum of the inherent variance and the ISI variance.



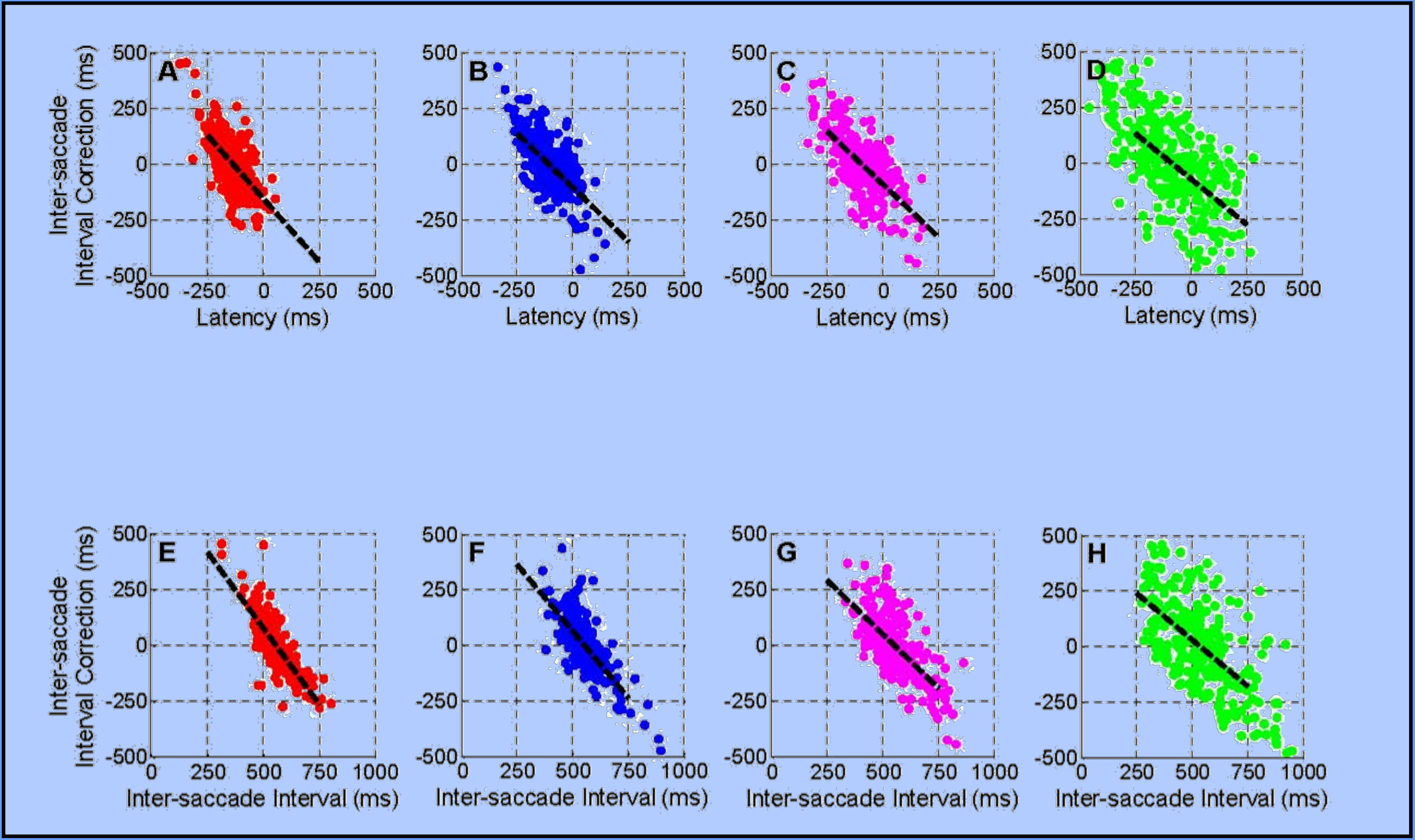
Self-Organized Criticality?

- Is this an SOC system?
- Characteristics of SOC:
  - Self-organization: no external tuning of parameters
  - Criticality:
    - heavily interconnected
    - extensive interactions
  - Scale-invariance
  - Self-similarity
  - 1/*f* spectral signature

Interval Correction is based on only most recent latency



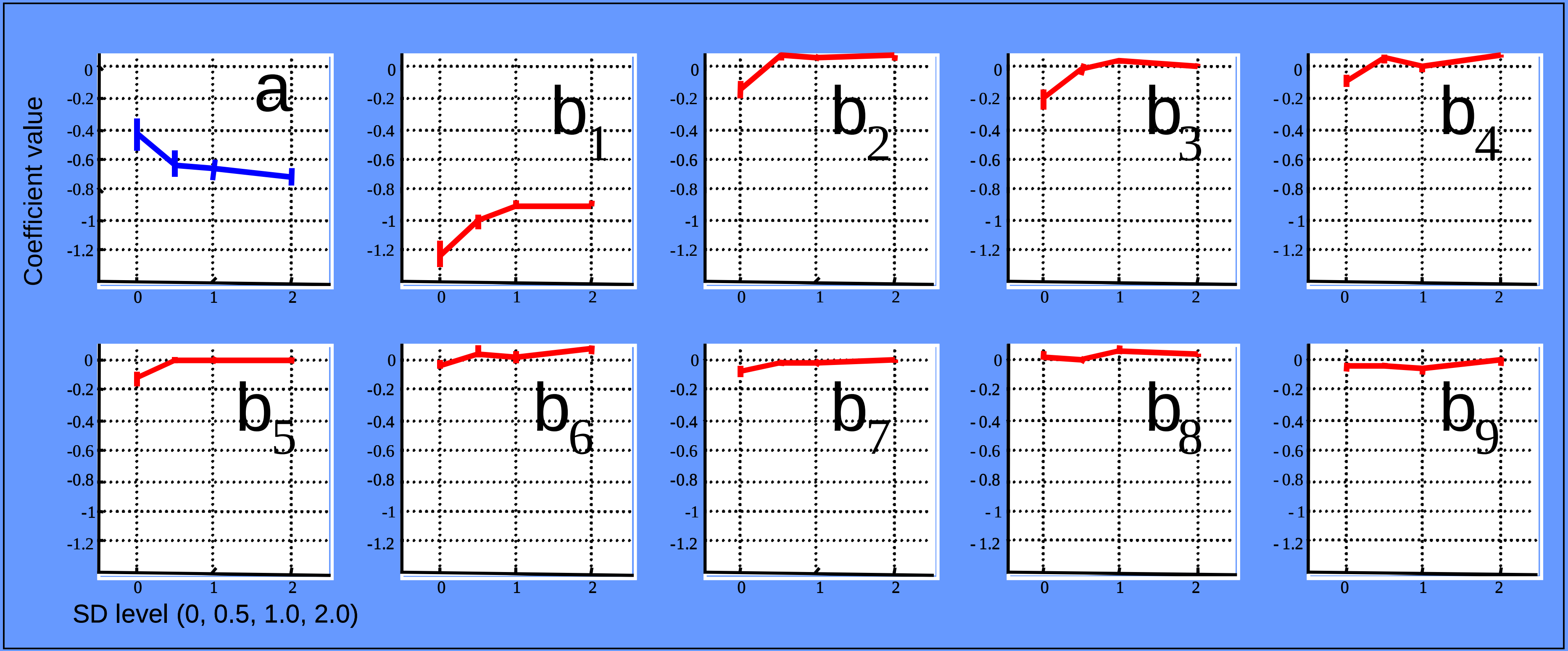
Effect of latency and interval errors on next interval decreases with variability



Simplified trial-to-trial model of predictive timing

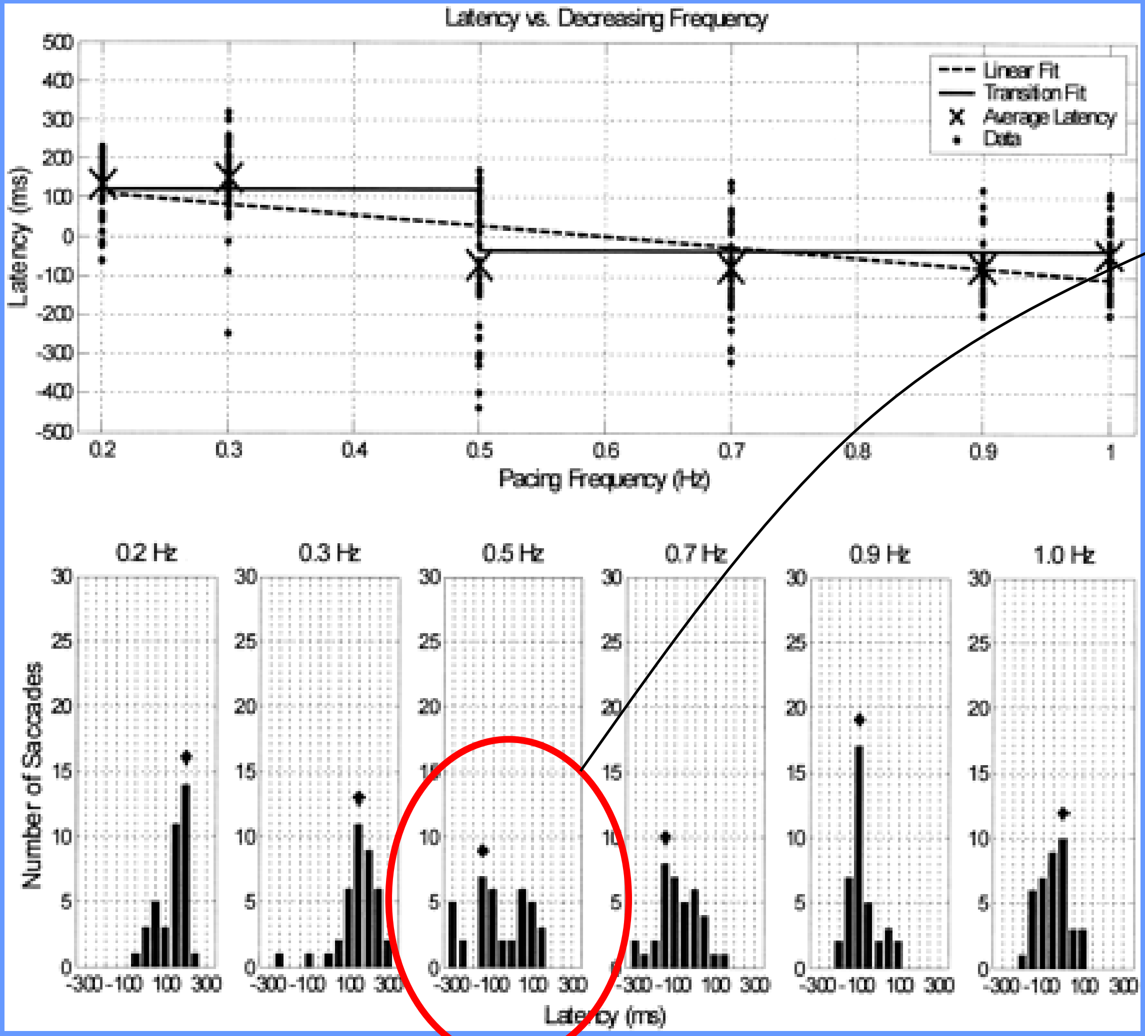
$$I(i) = a\, L(i-1) + b_1\, I(i-1) + b_2\, I(i-1) + b_3\, I(i-1) + b_4\, I(i-1) + \ldots.$$

L=latency, I =interval



Tuning to criticality

- Spontaneous transitions

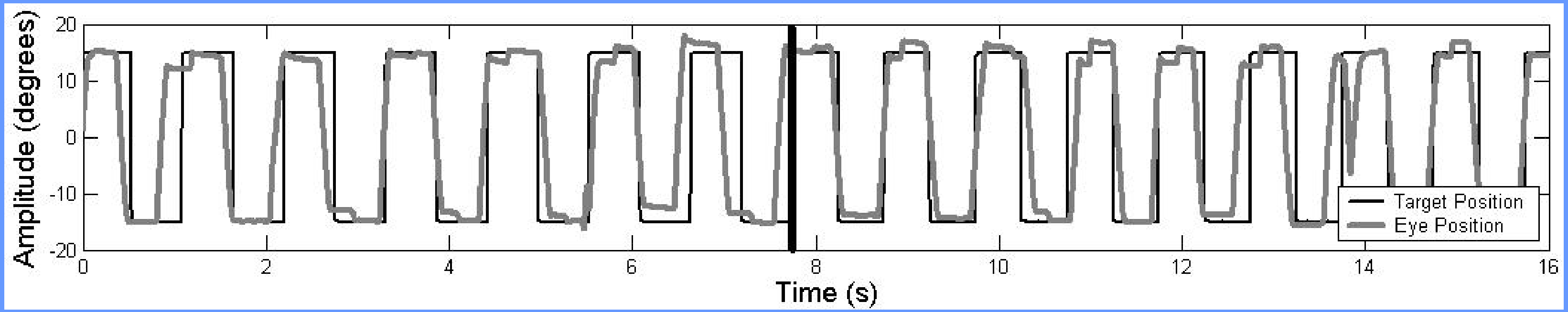


Spontaneous transitions between predictive and reactive states

Example of self-organization in prediction

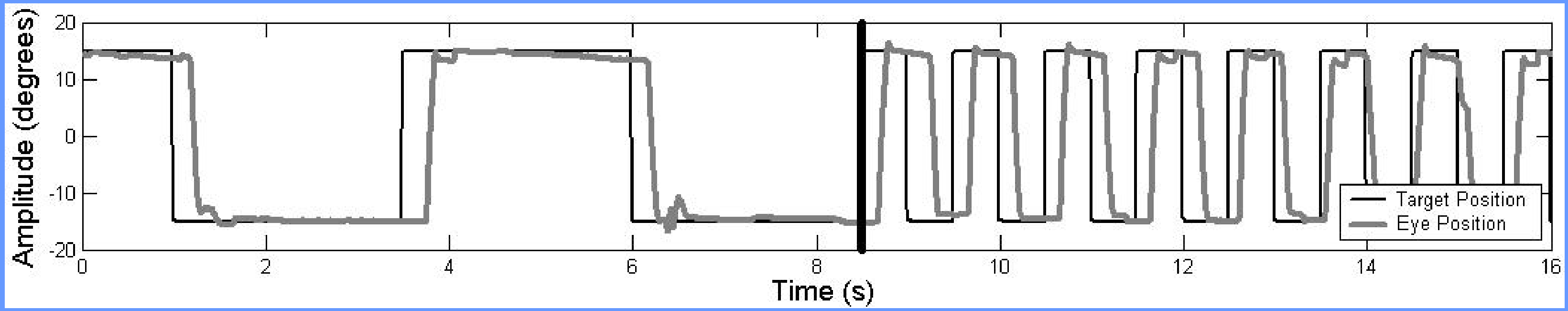
Rapid transitions into and between predictive states.

0.9 Hz 1.0 Hz



Predictive-predictive transition

0.2 Hz 1.0 Hz

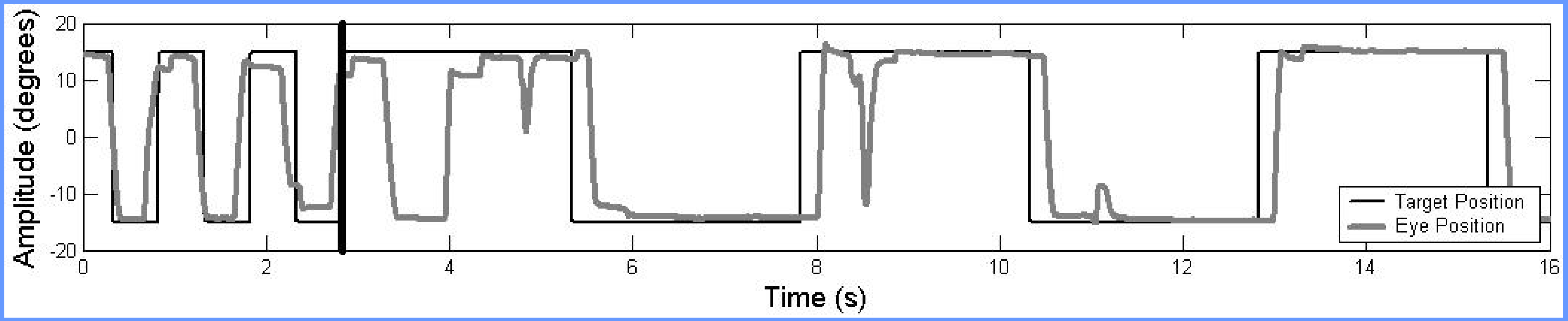


Reactive-predictive transition

Example of self-organization in prediction

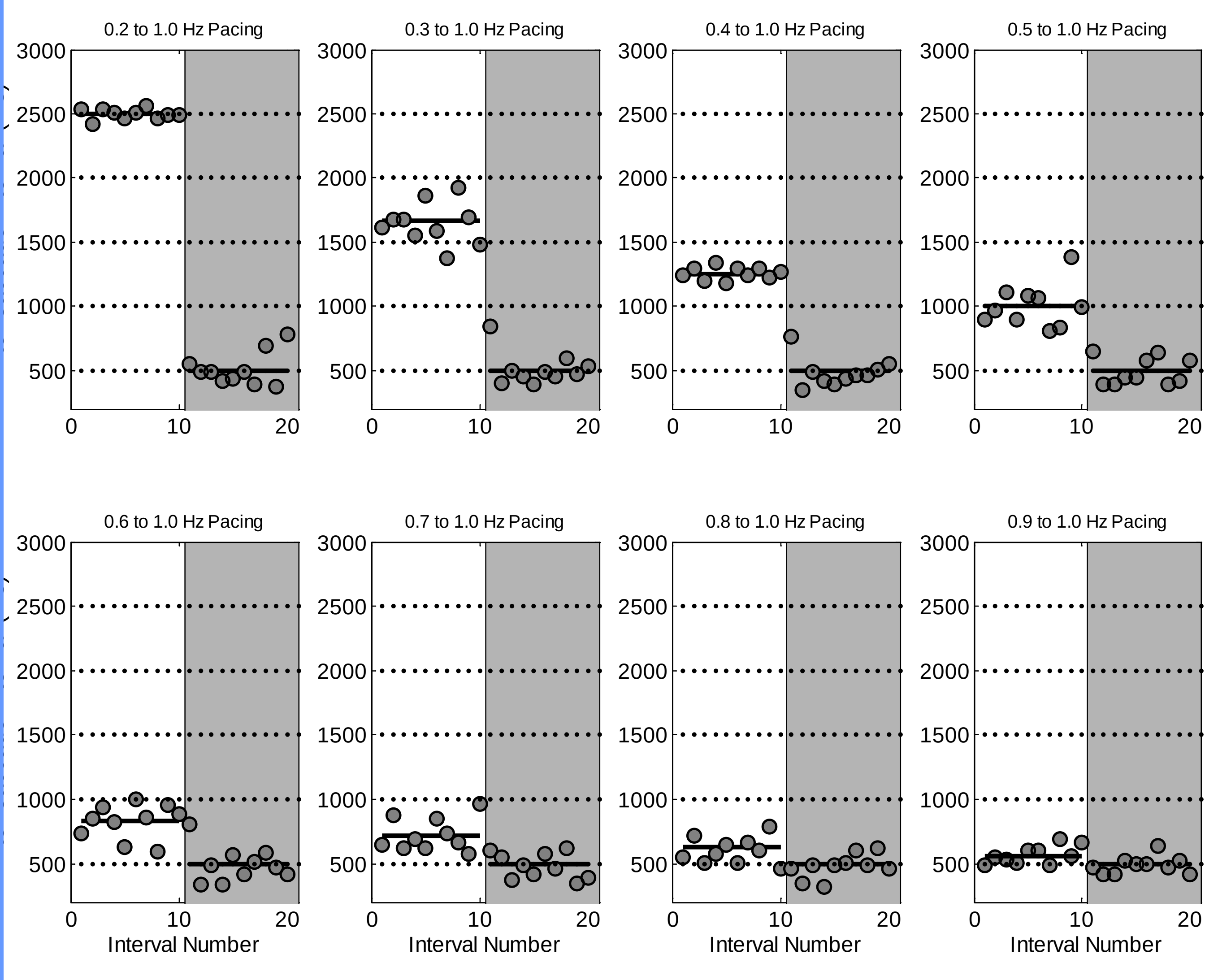
Preference for predictive state:
Predictive tracking continues on transition from predictive to reactive pacing.

1.0 Hz   0.2 Hz



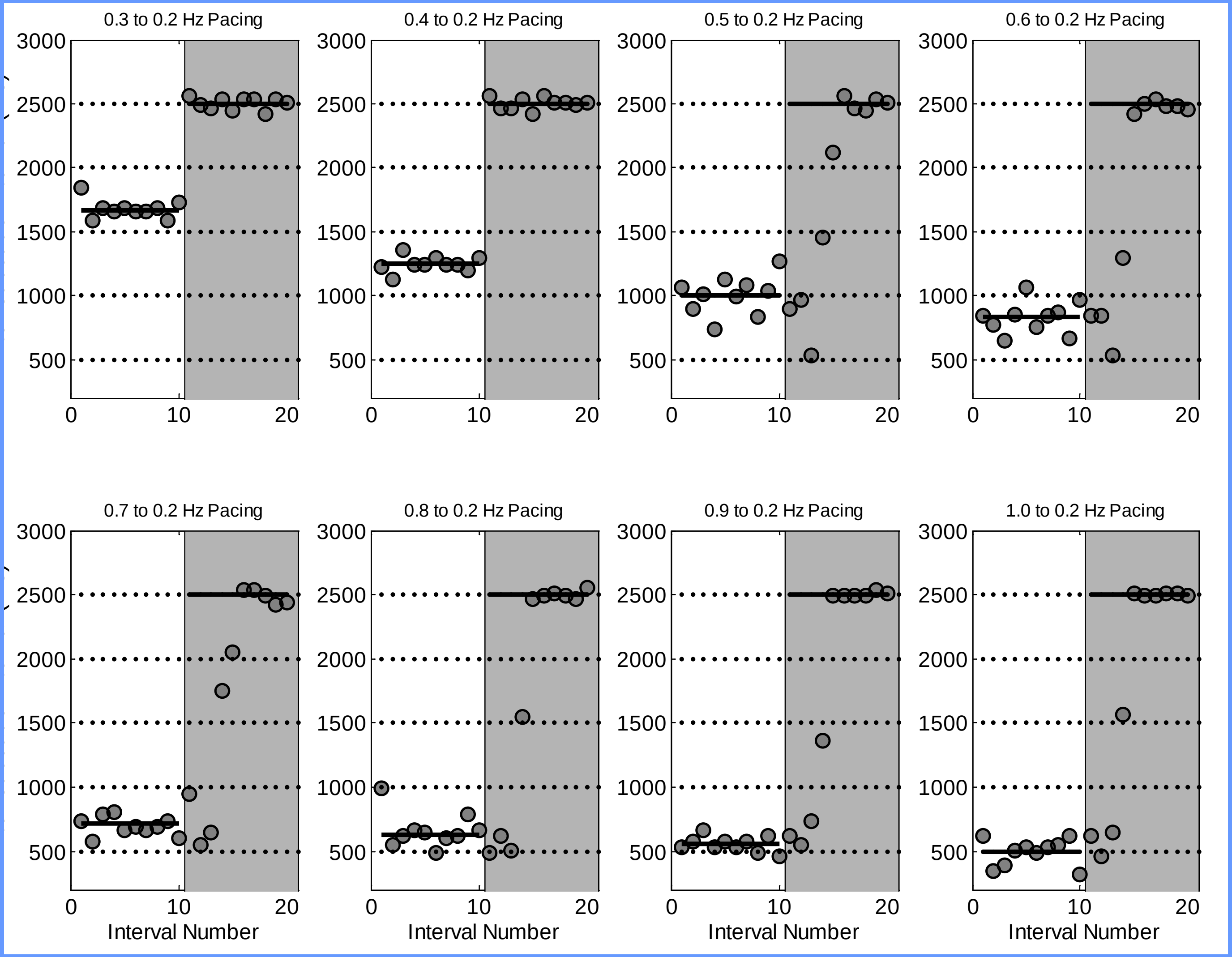
Perturbations into Predictive Mode

- inter-saccade intervals
- pacing abruptly changes to predictive rate (1.0 Hz)



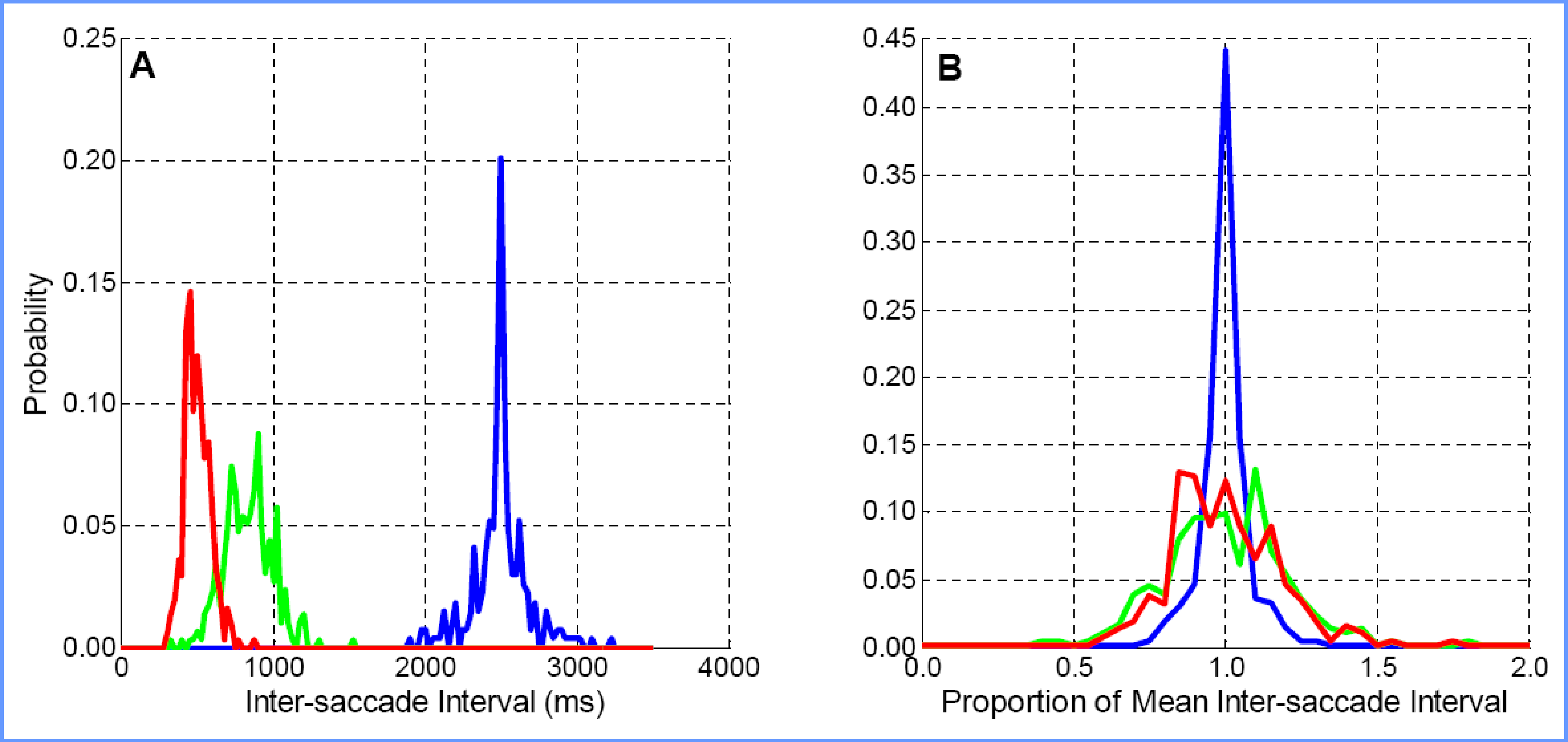
Perturbations into Reactive Mode

- inter-saccade intervals
- pacing abruptly changes to reactive rate (0.2 Hz)



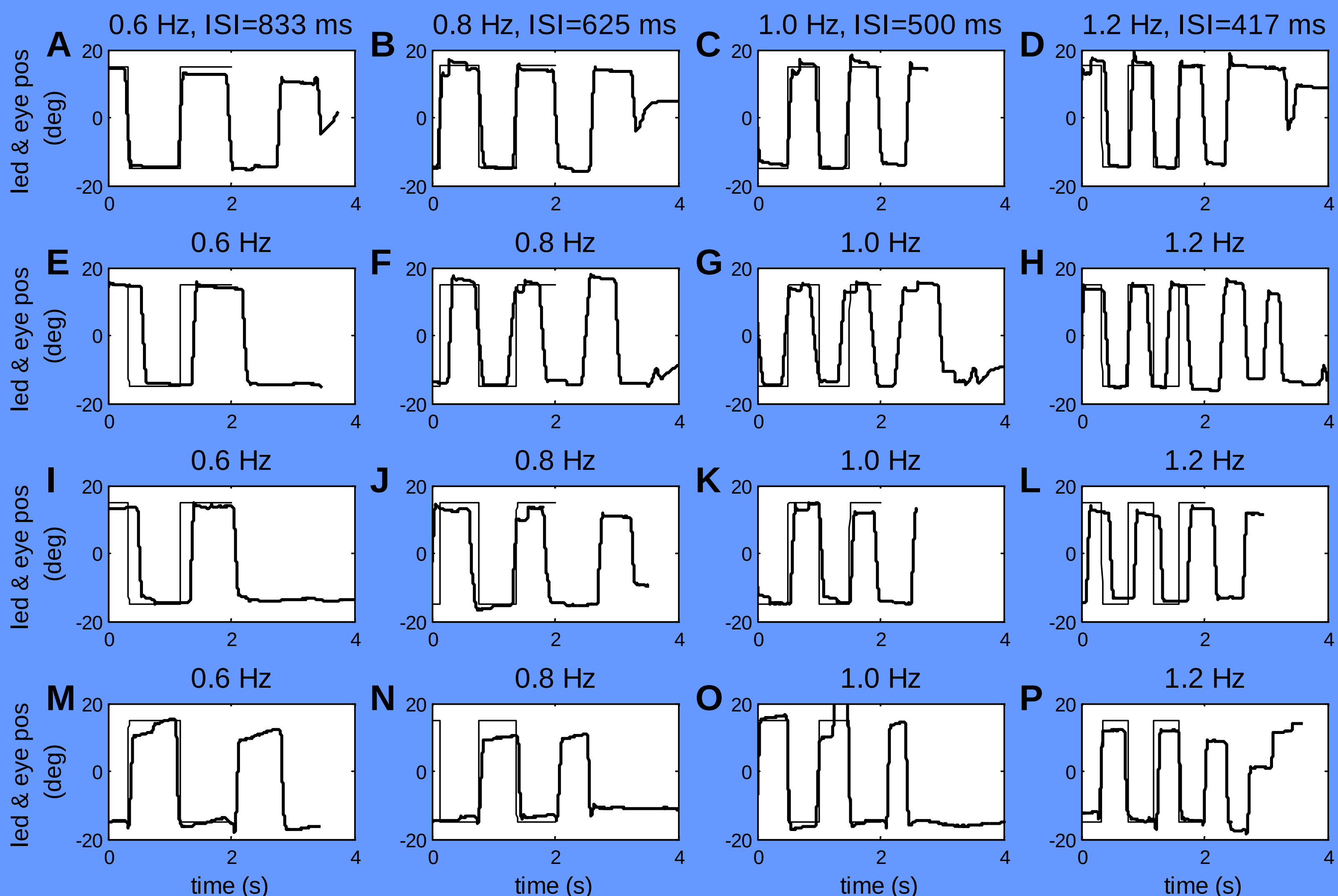
Clock

Scalar Property. Distributions of inter-saccade intervals at predictive pacing rates vary around the ISI and the variance increases with interval length (A). Predictive distributions (red & green) overlap when divided by mean interval (B).



Saccades continue after halt in pacing

Further demonstration of 2-sec window

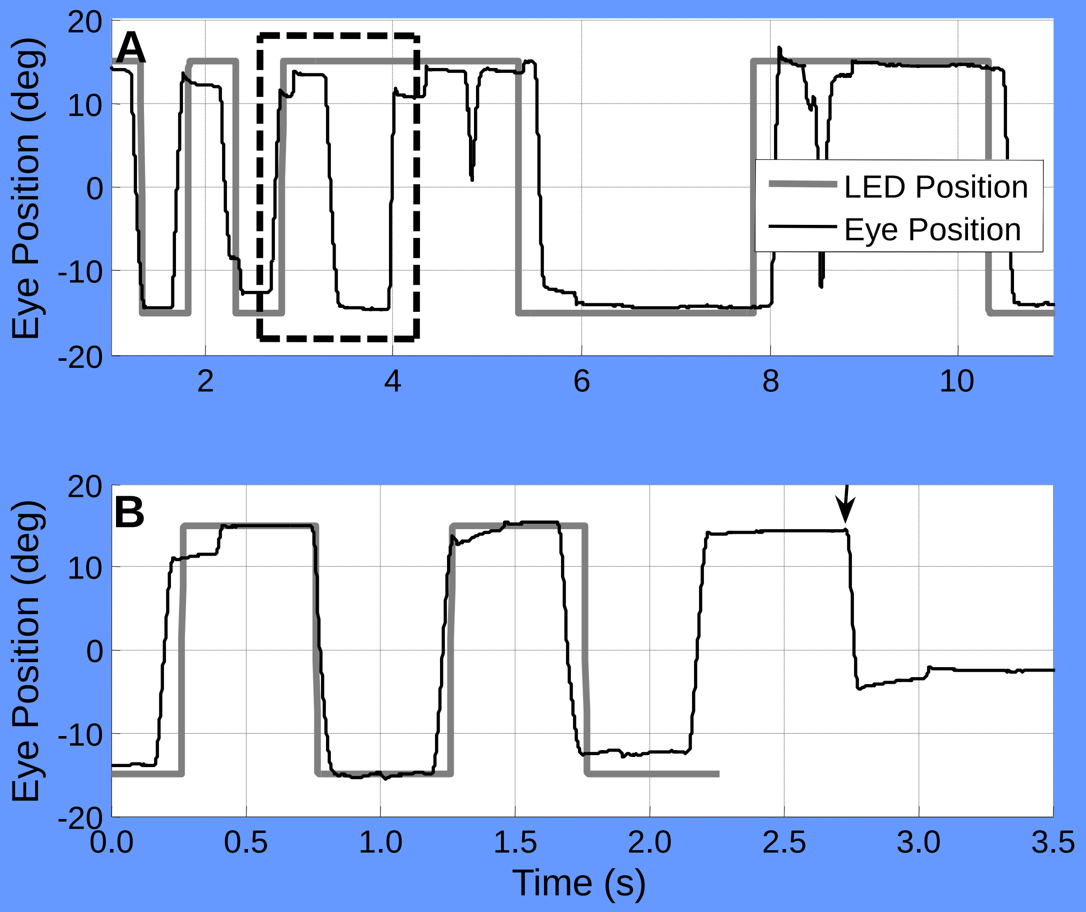


Clock

- This model results in a neural clock for generating inter-saccade intervals during predictive tracking.
- Further evidence for a clock:

Persistence of tracking after perturbations:

(A) Change in target timing, (B) extinguishing targets.



A Mathematical Model for Predictive Saccade Sequences

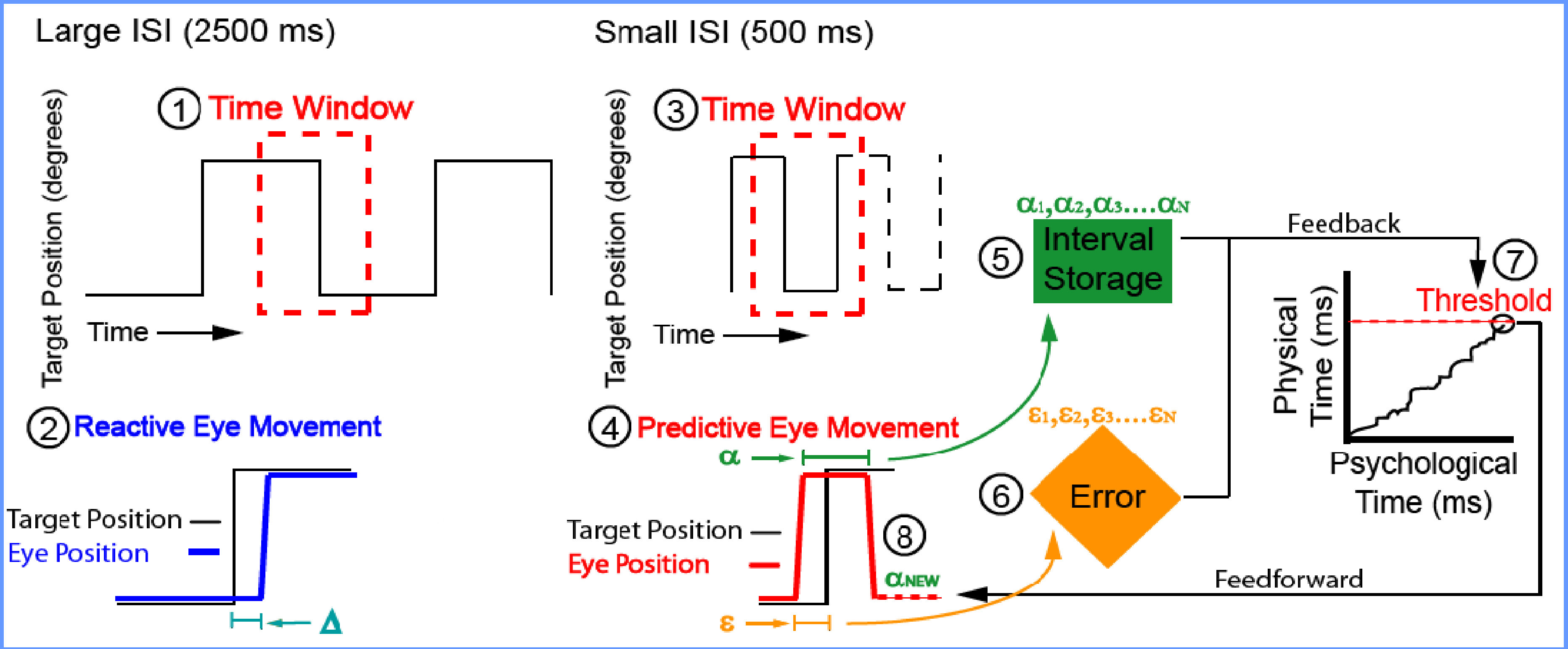
Goal – develop a model that reproduces observed behaviors:

- Reactive and predictive tracking
- Abrupt transition to predictive tracking above ~0.5 Hz
- Hysteresis in reactive-predictive transition
- Correlated predictive saccades
- Scalar property
- Extended tracking increases correlations
- Continue predictive tracking after perturbation

Basis:

- Integrate-to-threshold neural activity (LATER model, Carpenter et al.)
- Feedback of timing errors within a specific window adjusts the threshold of the integrator

Model

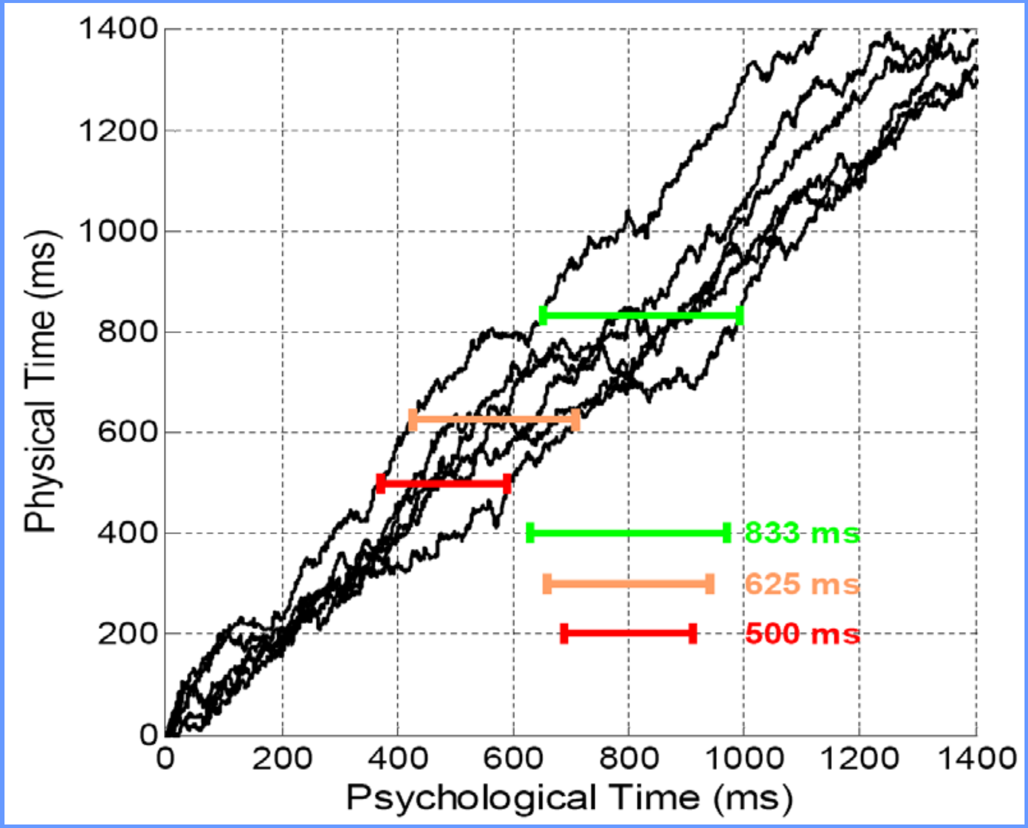


(1)Large ISI: consecutive target jumps not in window → reactive saccade (2).
(3) Small ISI: consecutive target jumps in window → predictive saccade (4).

Feedback of intervals (α_i) (5) and latencies (ε_i) (6) within correlation window adjusts threshold (7) of integrate-to-threshold system.

Timing of next saccade α_{NEW} (8) is determined by integration of noisy neural signal to threshold.

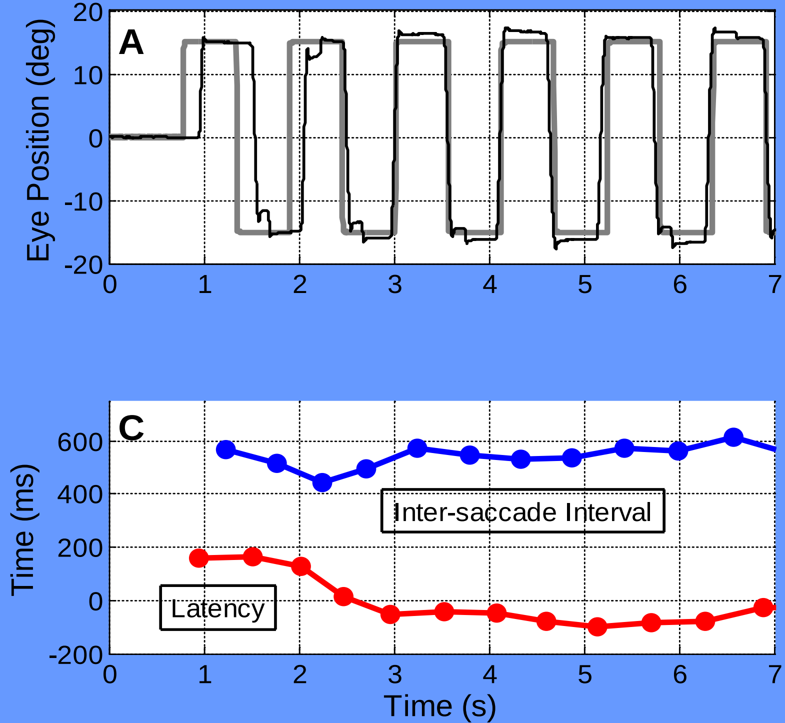
Scalar Property of Time Estimation



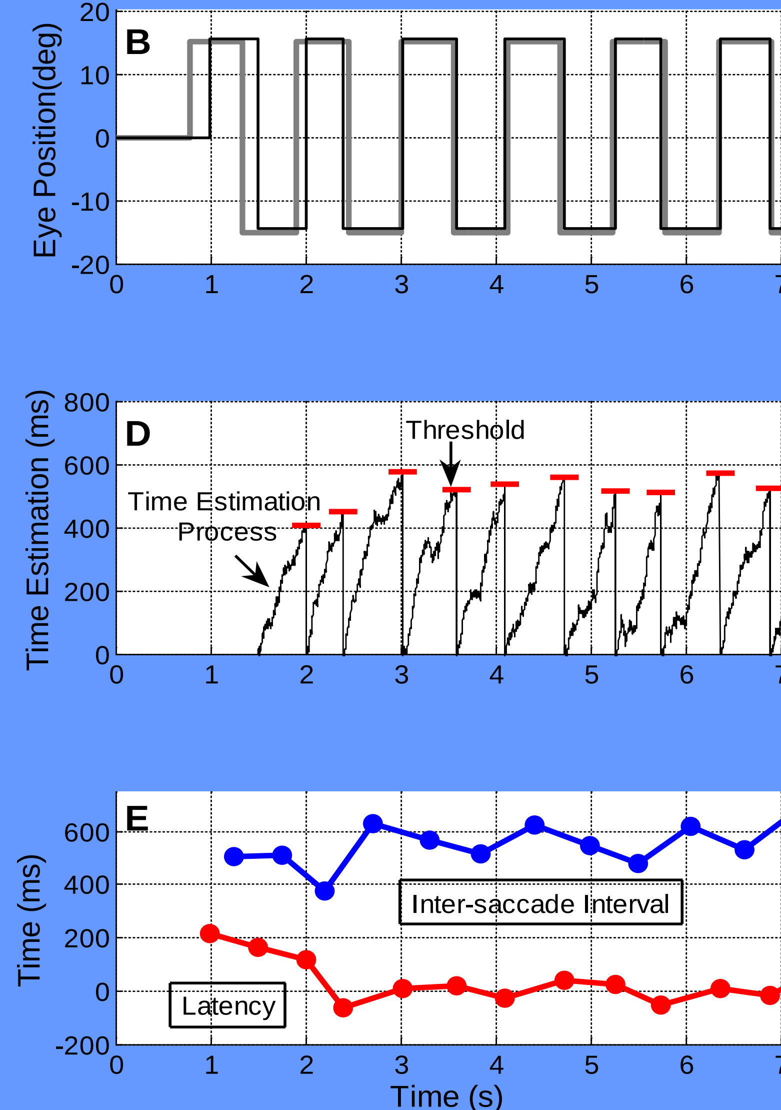
- Integration of a noisy signal to threshold.
- Longer interval, longer integration of noisy signal.
- Variability of interval estimate proportional to ISI.
- Scalar Property

The Model in Action

Behavioral data



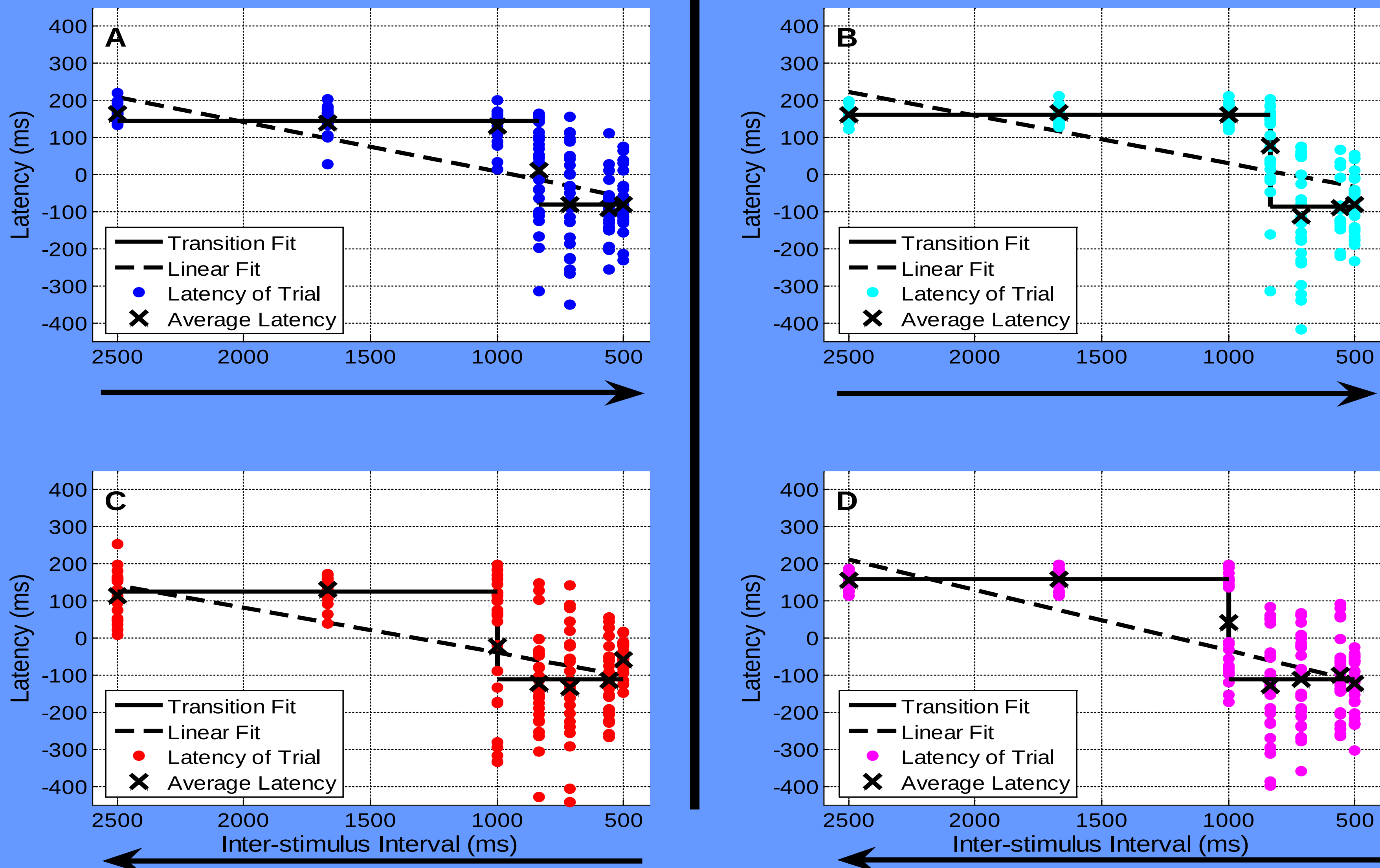
Model data



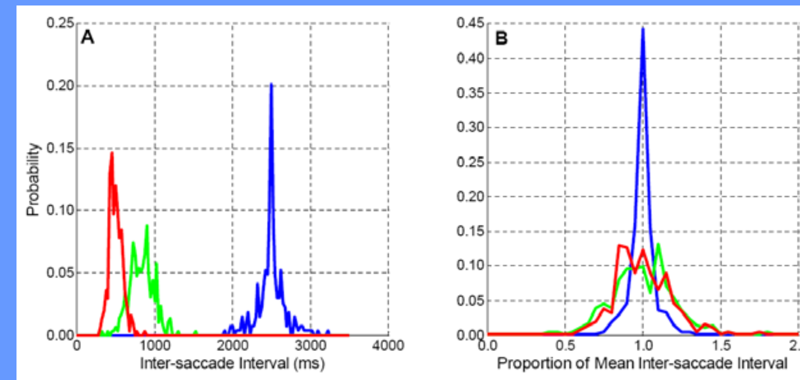
Simulation Results

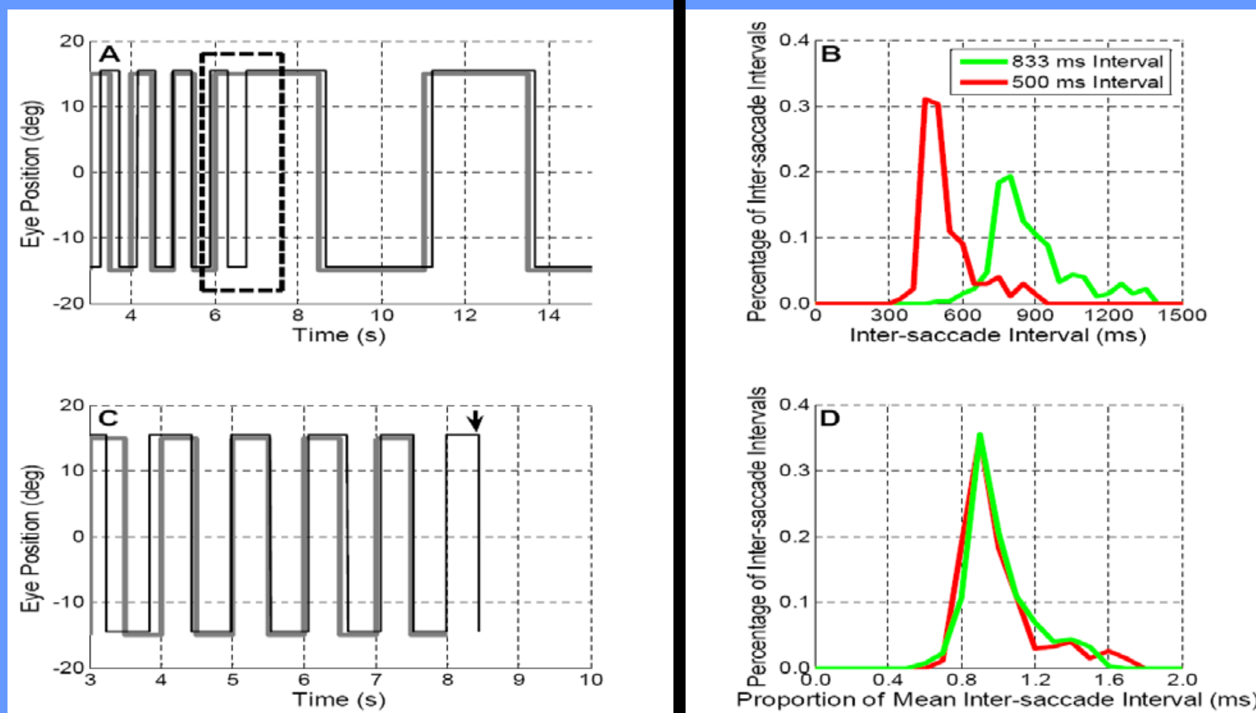
Behavioral data

Model data



Simulation Results



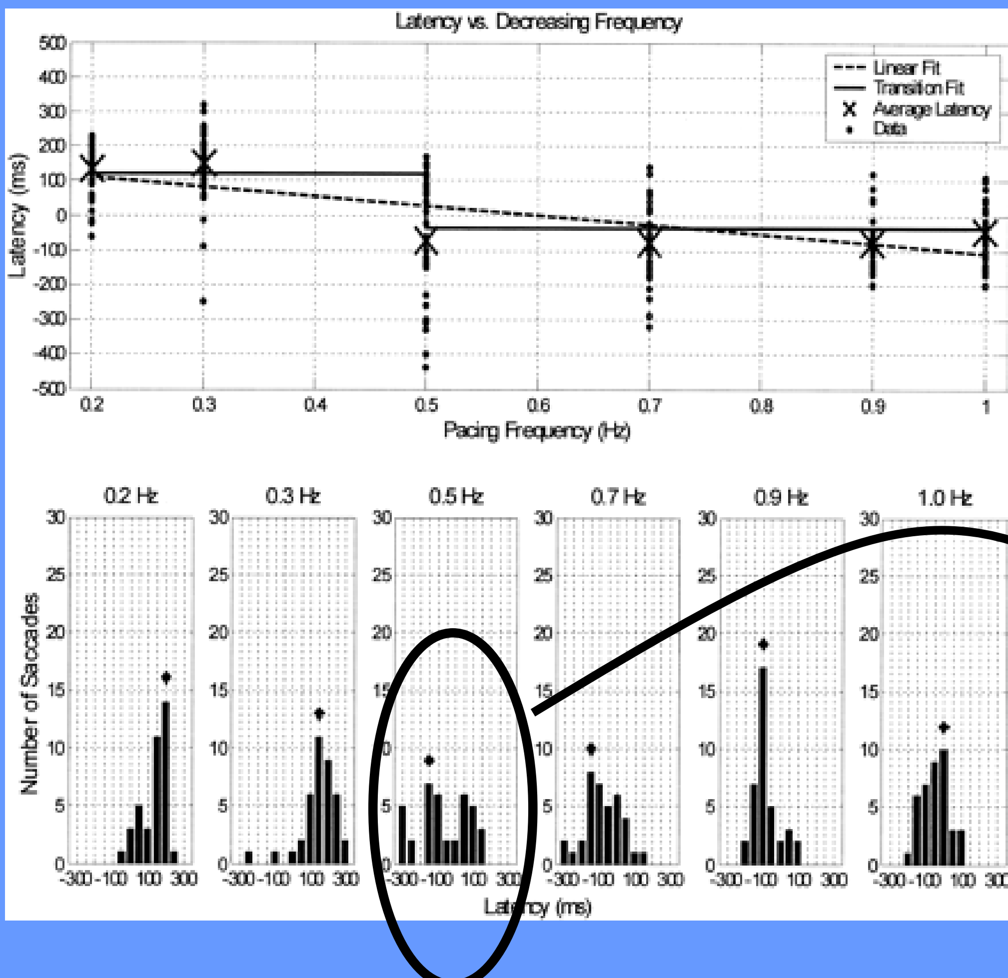


- Tracking continues when stimulus timing changes (A) or stops (C).

- Variability increases with rate (B).
- Predictive distributions overlap.

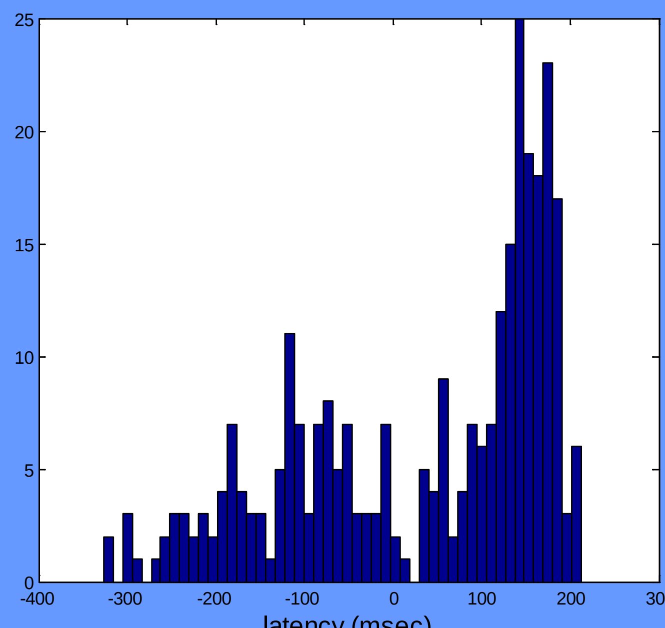
Latency Distributions in Transition Range

Behavioral data



At ~0.5 Hz, tracking switches between reactive and predictive

Model data

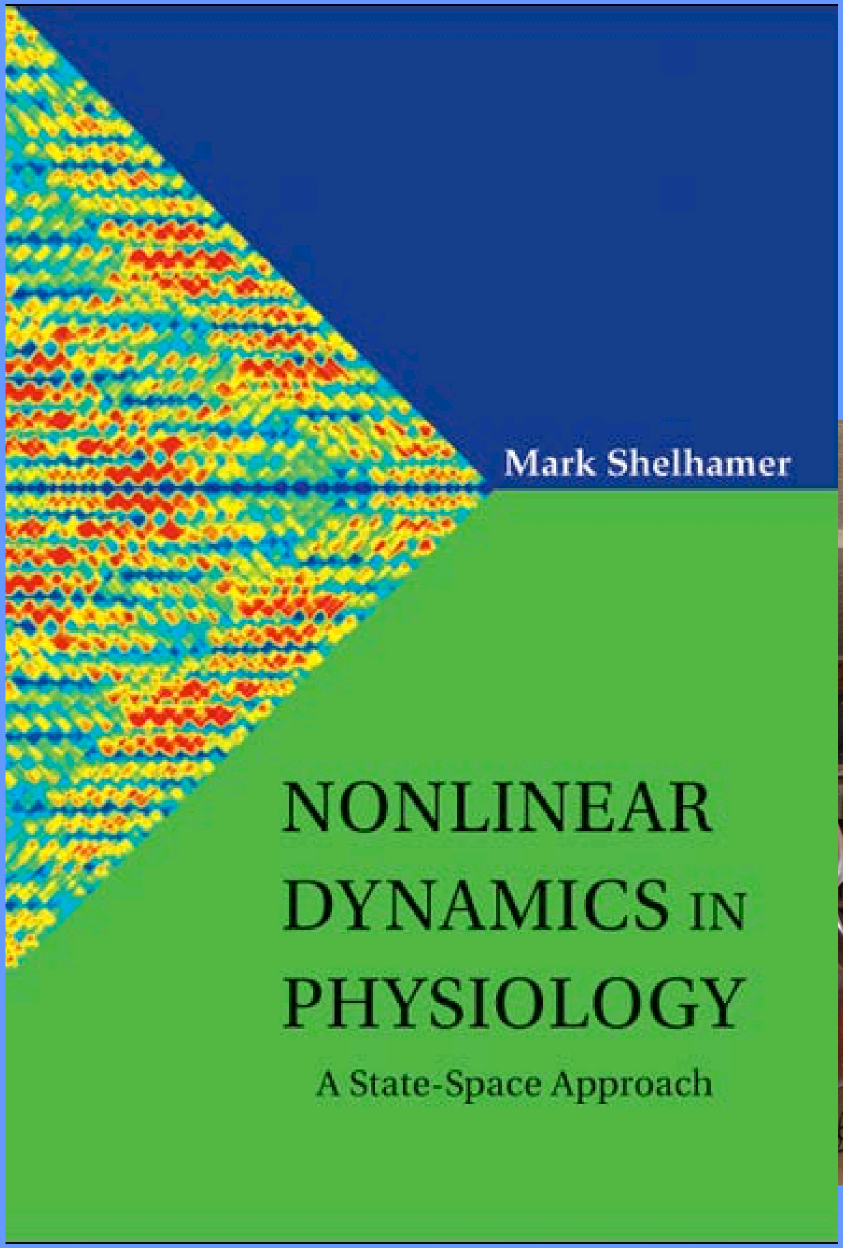


Model Summary

- Model implements internal clock with integrate-to-threshold mechanism (Buhusi and Meck, 2005).
- Internal clock is internal representation of stimulus timing, modified by feedback from previous inter-saccade intervals and latencies.
- Feedback occurs within a time window estimated from inter-saccade intervals during reactive-to-predictive transition in tracking.
- Model reproduces:
    - Phase transition and hysteresis.
    - Interval/latency adjustments.
    - Bimodal tracking in transition range.
    - Continuation despite perturbations.
    - Scalar Property.

Implications of fBm

- "Long-term correlation"
  - Gradual decay of inter-trial interactions
- Slow decay of power spectrum
  - Substantial low-frequency components
  - Variations on many time scales
- Suggests "self-organized criticality" (SOC)
  - System organizes itself to be stable yet quickly responsive to changes


Mark Shelhamer
NONLINEAR DYNAMICS IN PHYSIOLOGY
A State-Space Approach


Publications

M Shelhamer, W Joiner (2003) Saccades exhibit abrupt transition between reactive and predictive, predictive saccade sequences have long-term correlations. *J Neurophysiol* 90:2763-2769.

M Shelhamer (2005) Sequences of predictive saccades are correlated over a span of ~2s, and produce a fractal time series. *J Neurophysiol* 93:2002-2011.

M Shelhamer (2005) Phase transition between reactive and predictive eye movements is confirmed with nonlinear forecasting and surrogates. *Neurocomp* 65-66:769-776.

M Shelhamer (2005) Sequences of predictive eye movements form a fractional Brownian series – implications for self-organized criticality in the oculomotor system. *Biol Cybern* 93:43-53.

WM Joiner, M Shelhamer, YH Ying (2005) Cerebellar influence in oculomotor phase-transition behavior. *Ann NY Acad Sci* 1039:536-539.

WM Joiner, M Shelhamer (2006) An internal clock generates repetitive predictive saccades. *Exp Brain Res* 175:305-320.

WM Joiner, M Shelhamer (2006) Responses to noisy periodic stimuli reveal properties of a neural predictor. *J Neurophysiol* 96:2121-2126.

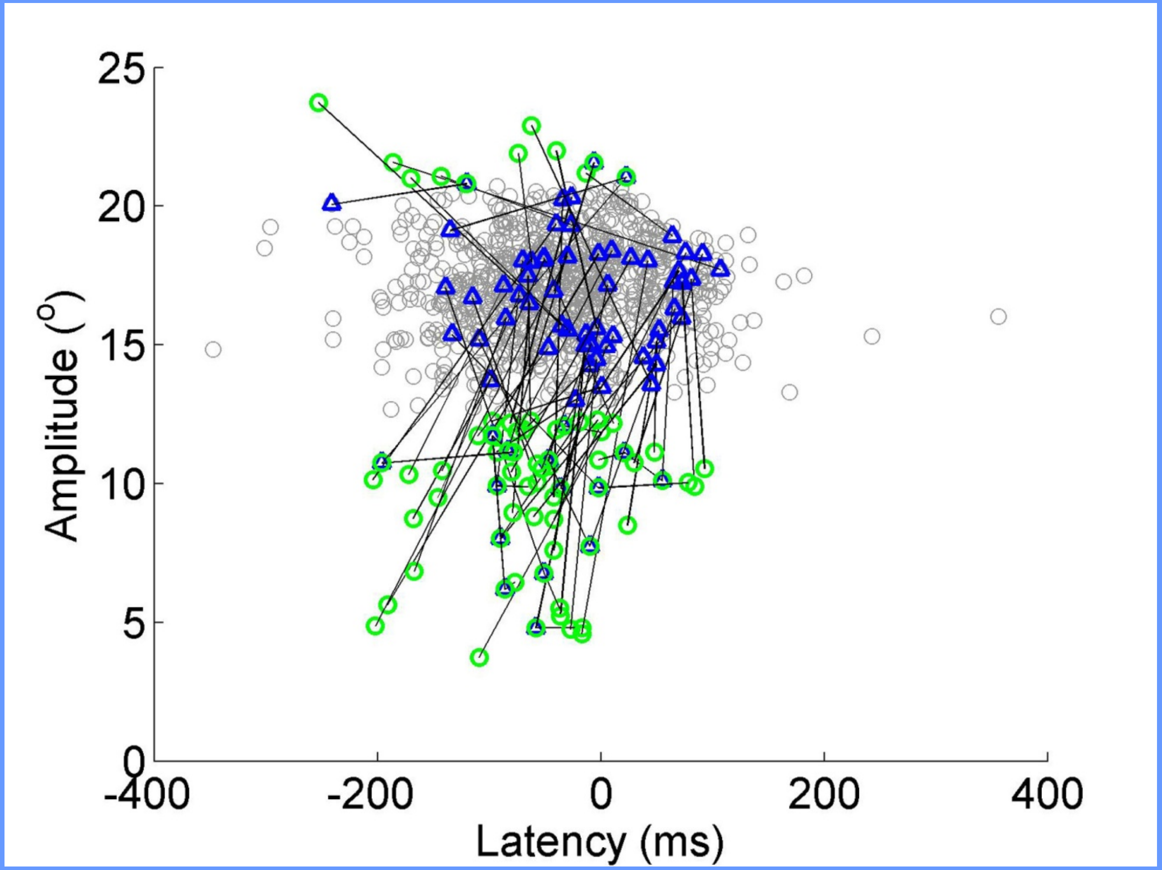
WM Joiner, J-E Lee, M Shelhamer (2007) Behavioral analysis of predictive saccade tracking as studied by countermanding. *Exp Brain Res* 181:307-320.

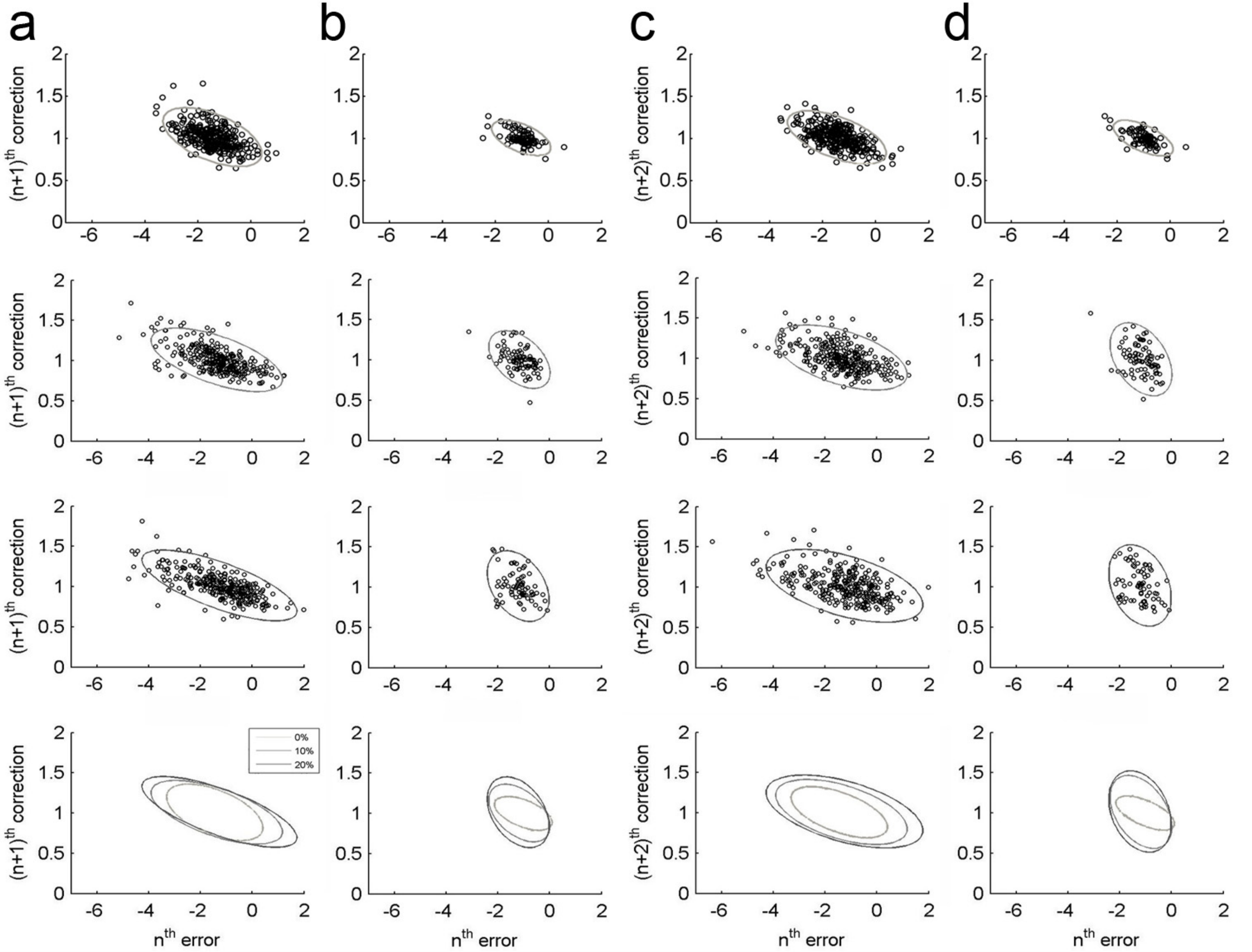
WM Joiner, J-E Lee, AG Lasker, M Shelhamer (2007) An internal clock for predictive saccades is established identically by auditory or visual information. *Vision Res* 47:1645-1654.

M Shelhamer, W Joiner (2008) Stability of predictive oculomotor tracking: an exploration of the relationships between dynamical systems, information theory, and probability. In: DA De Jong (ed) Progress in Biological Cybernetics Research. Hauppauge NY: Nova Science Publishers.

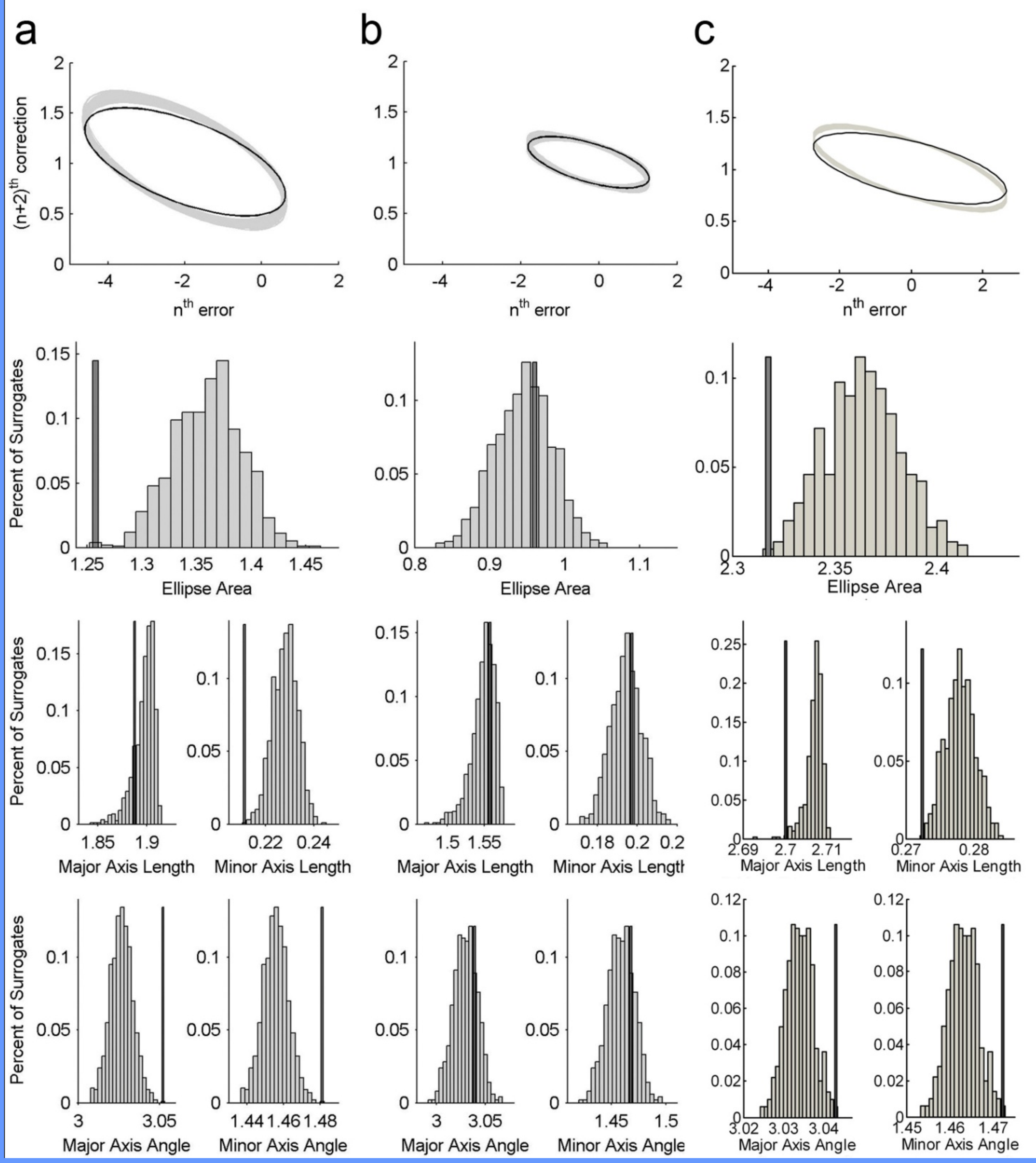
WM Joiner, M Shelhamer (2008) A model of time estimation and error feedback in predictive timing behavior. *J Comput Neurosci* online doi: 10.1007/s10827-008-0102-x.

Lack of correspondence between predictive-saccade timing and amplitude

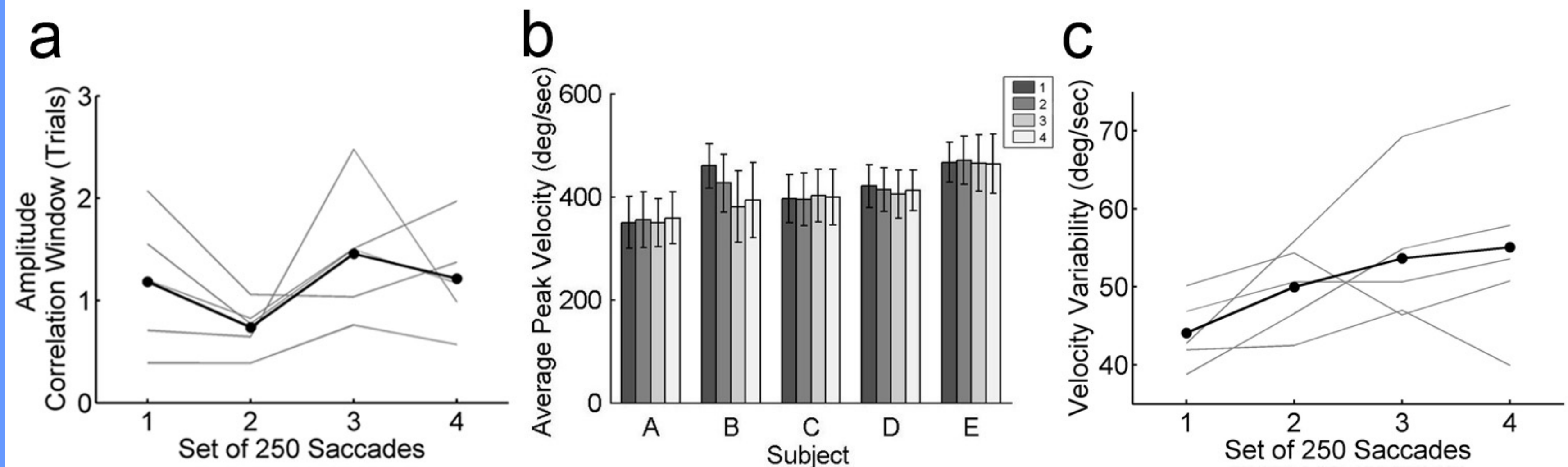


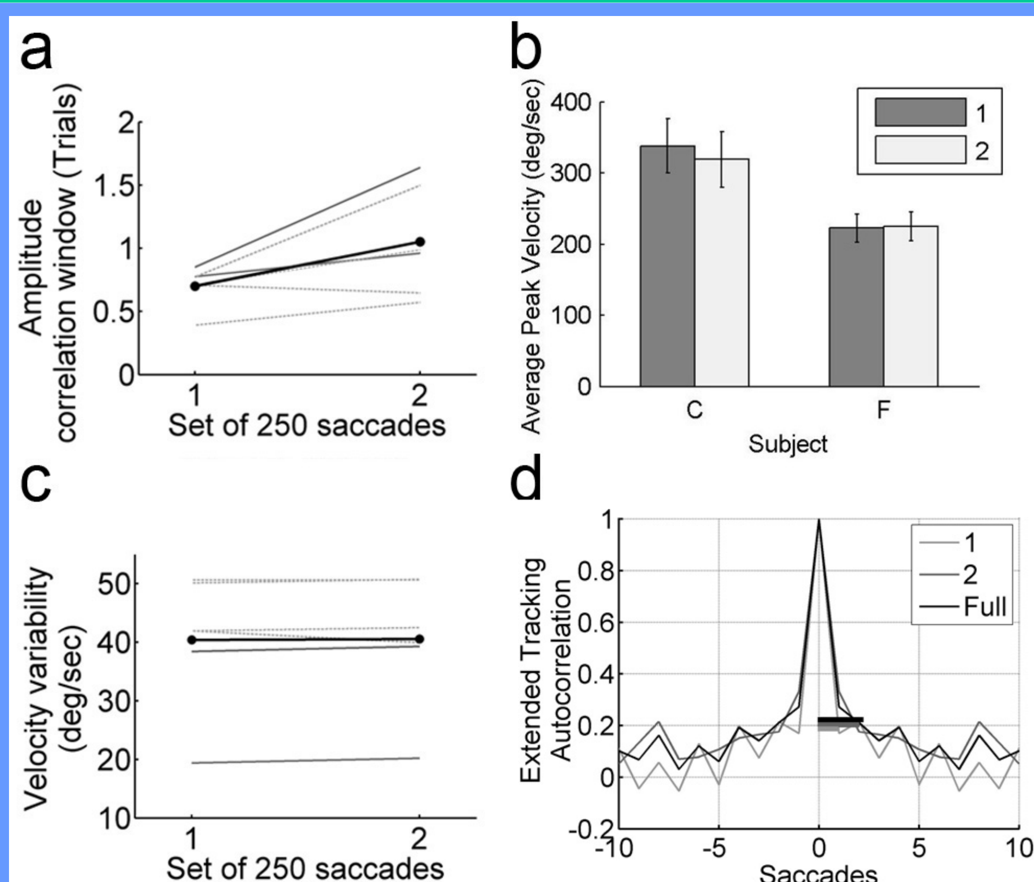


a
b
c
d
(n+1)th correction
(n+2)th correction
nth error
0%
10%
20%



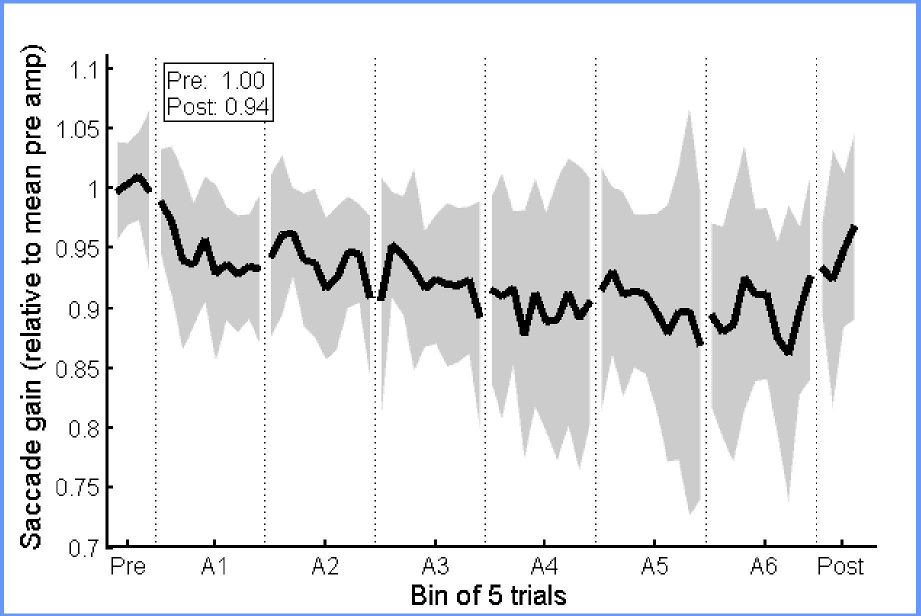
a
b
c
(n+2)th correction
nth error
Percent of Surrogates
Ellipse Area
Major Axis Length
Minor Axis Length
Major Axis Angle
Minor Axis Angle

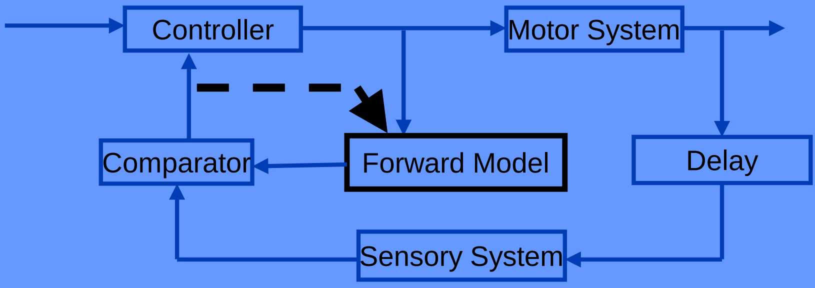

a
Amplitude Correlation Window (Trials)
Set of 250 Saccades
b
Average Peak Velocity (deg/sec)
Subject
A B C D E
1 2 3 4
c
Velocity Variability (deg/sec)
Set of 250 Saccades


a
Amplitude correlation window (Trials)
Set of 250 saccades
b
Average Peak Velocity (deg/sec)
Subject
C F
1 2
c
Velocity variability (deg/sec)
Set of 250 saccades
d
Extended Tracking Autocorrelation
Saccades
1
2
Full

An *adaptable* forward model

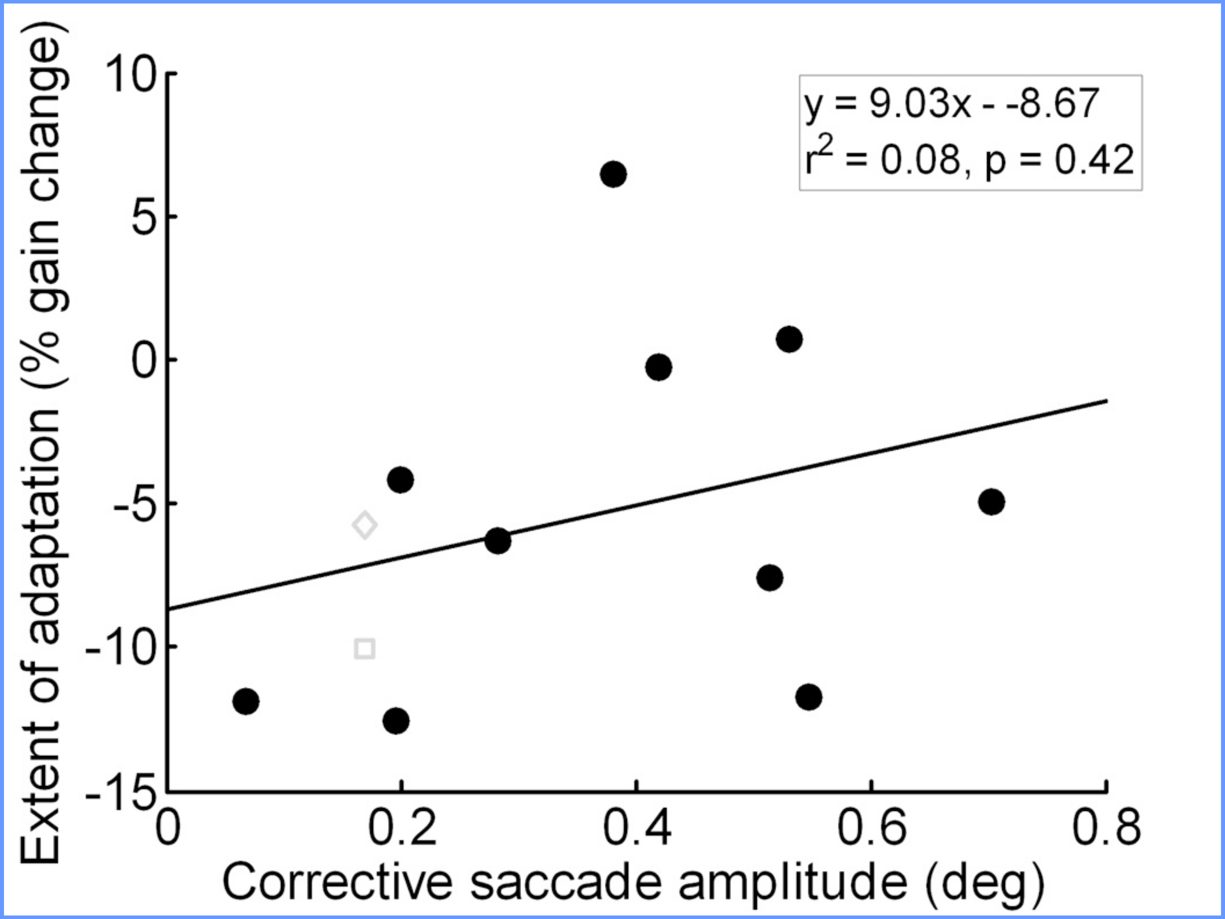
- Adaptation reaches an asymptote, but the retinal error size does not change
- This can be explained by an adaptable forward model (predictions incorporate the target behavior)



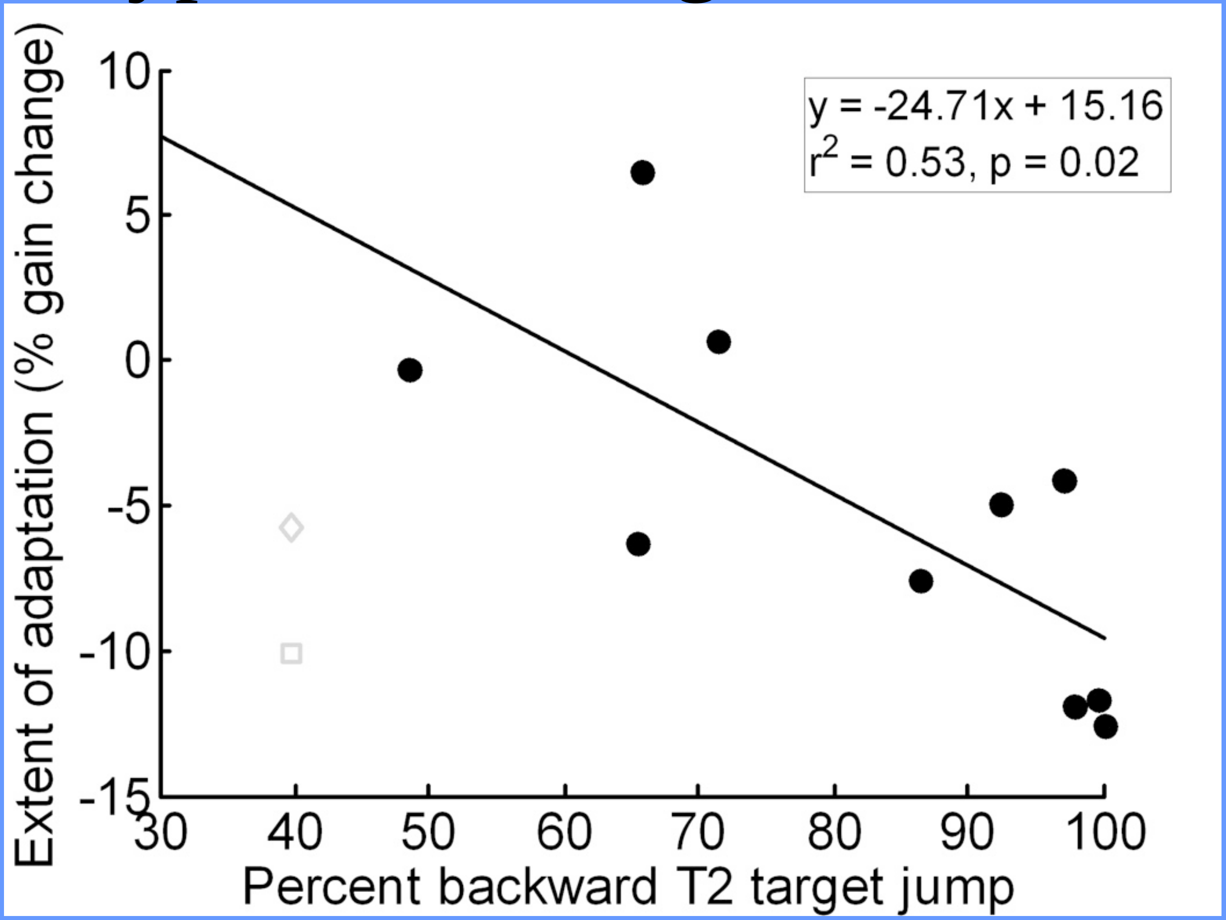


Wong and Shelhamer

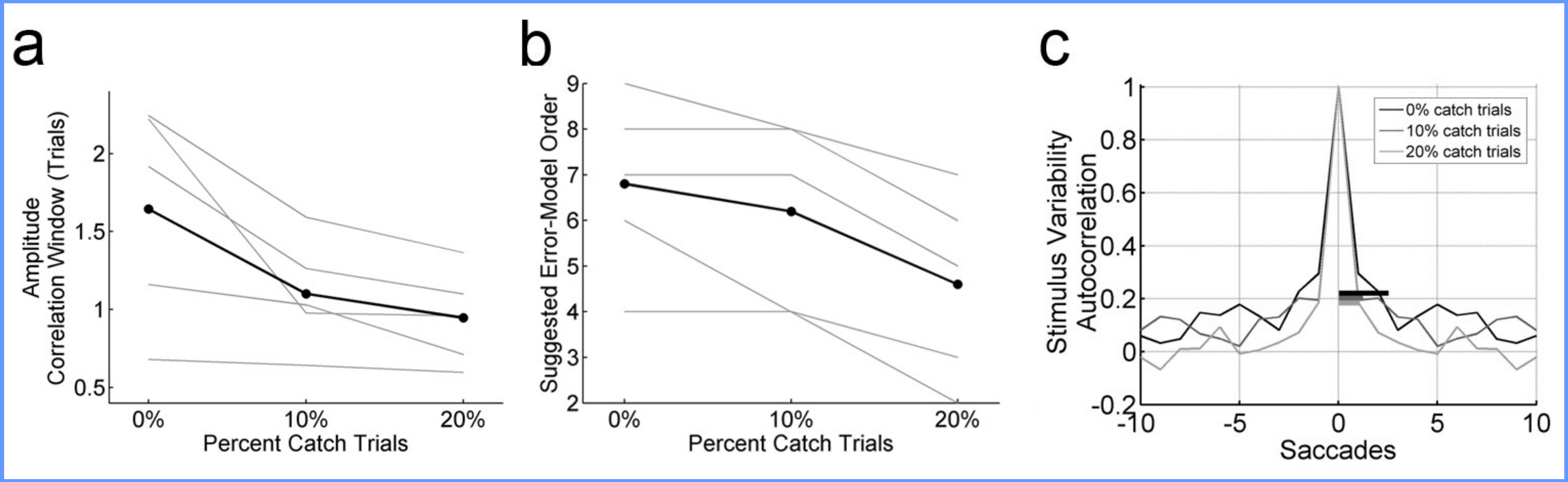
Constant-Error Adaptation Hypothesis: Motor Correction



Constant-Error Adaptation Hypothesis: Target Behavior

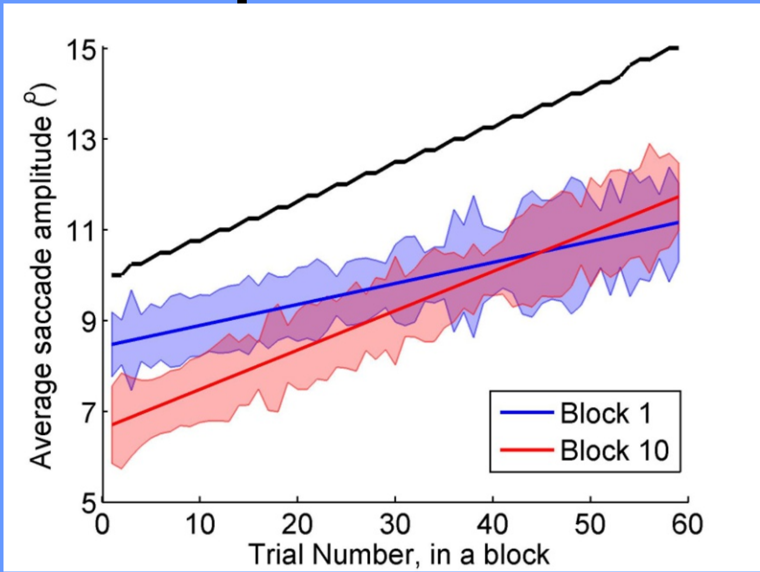


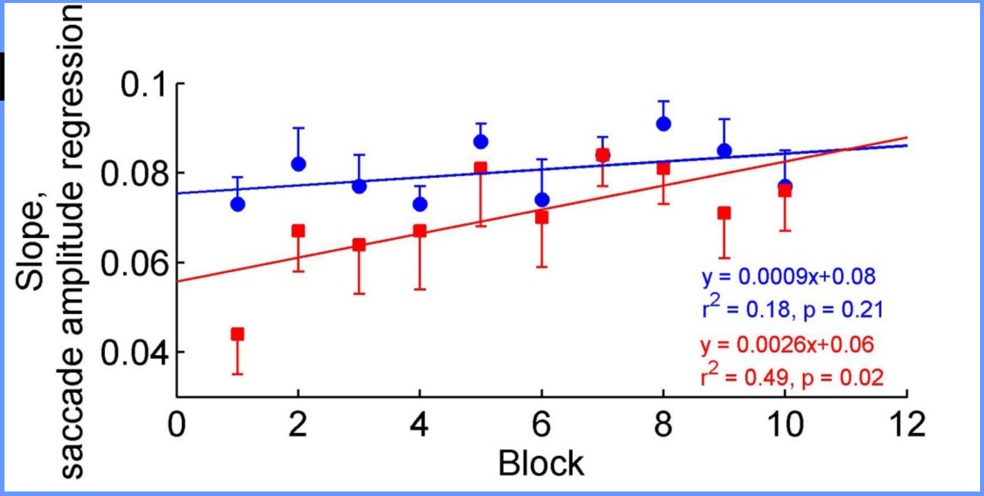
Stimulus Variability and the Correlation Window



SCA6: Learning a slope?

- For SCA6, slope in block 1 is less steep than in block 10
- In controls, slope is already close to ideal
- Initially SCA6 behavior poorly matches target behavior, but improves with training
- This is achieved by decreasing saccade gain (easier?)

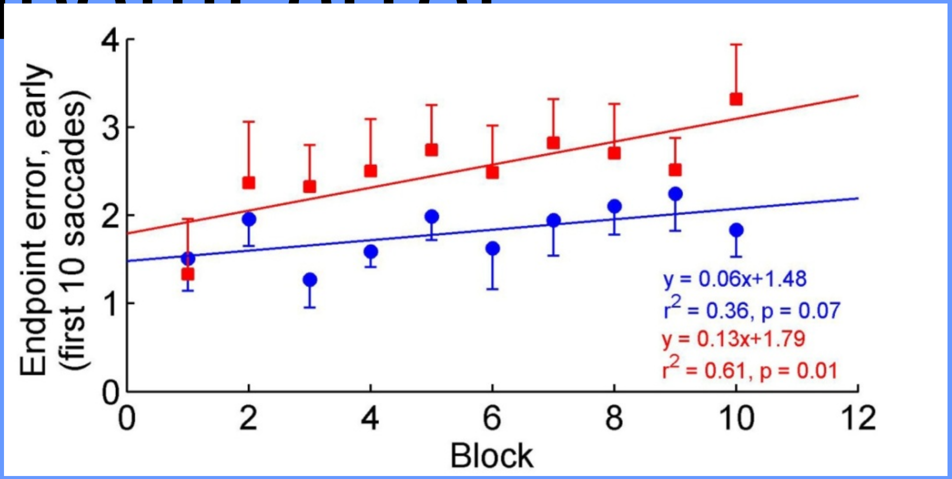


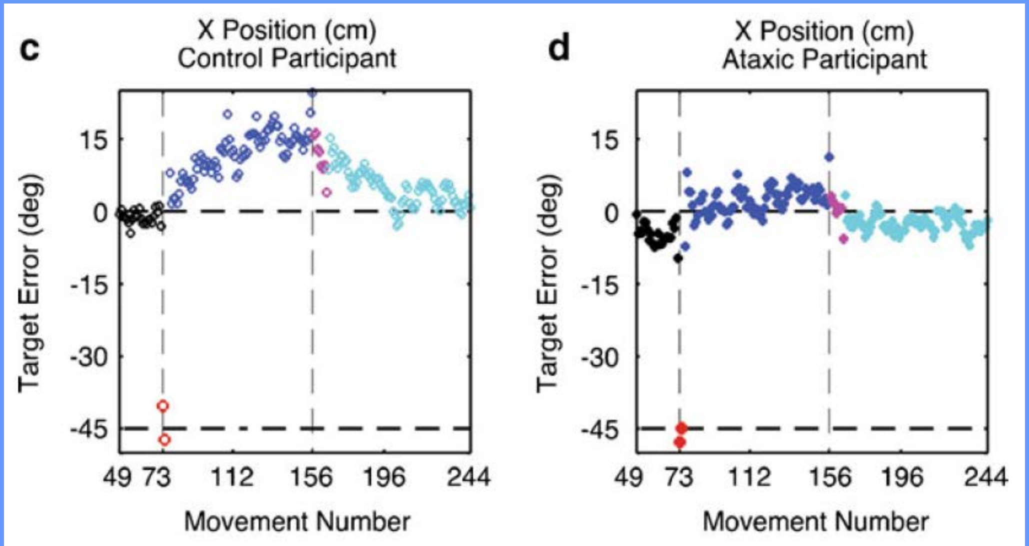


Plotted:

Early endpoint error

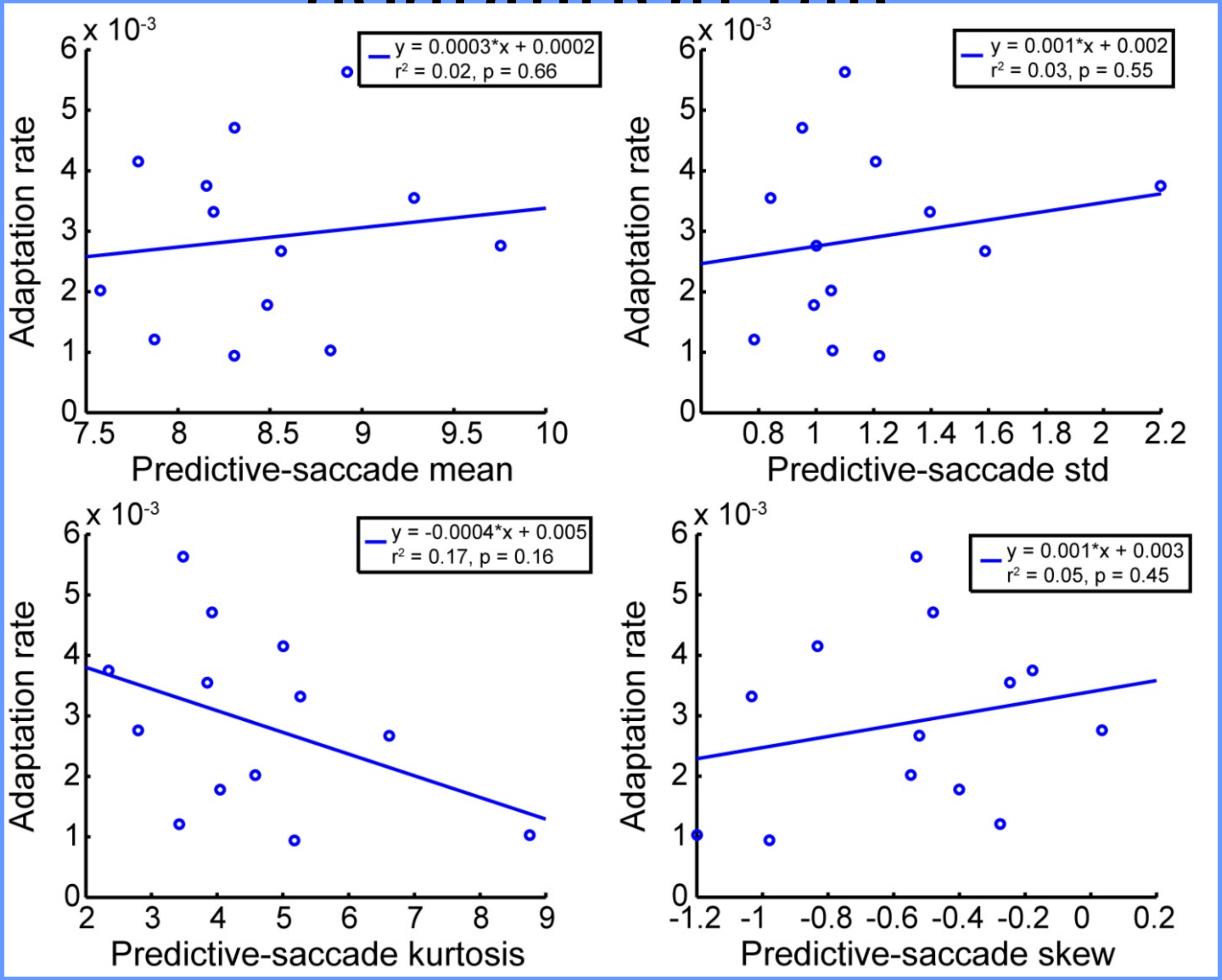
- Slope corrections arise from decreasing gain early, not increasing gain late
- Slope-matching may arise as a strategy to complete an implicit task, which can supersede adaptation





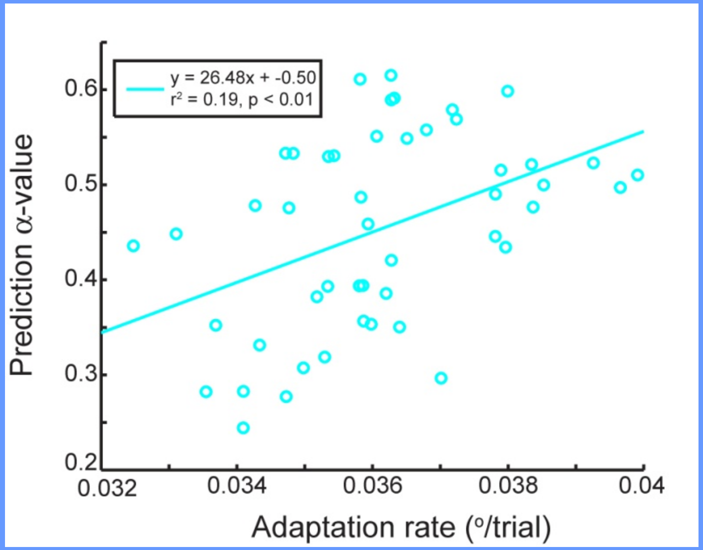
Taylor et al. 2010; Mazzoni and

Prediction variability versus adaptation rate



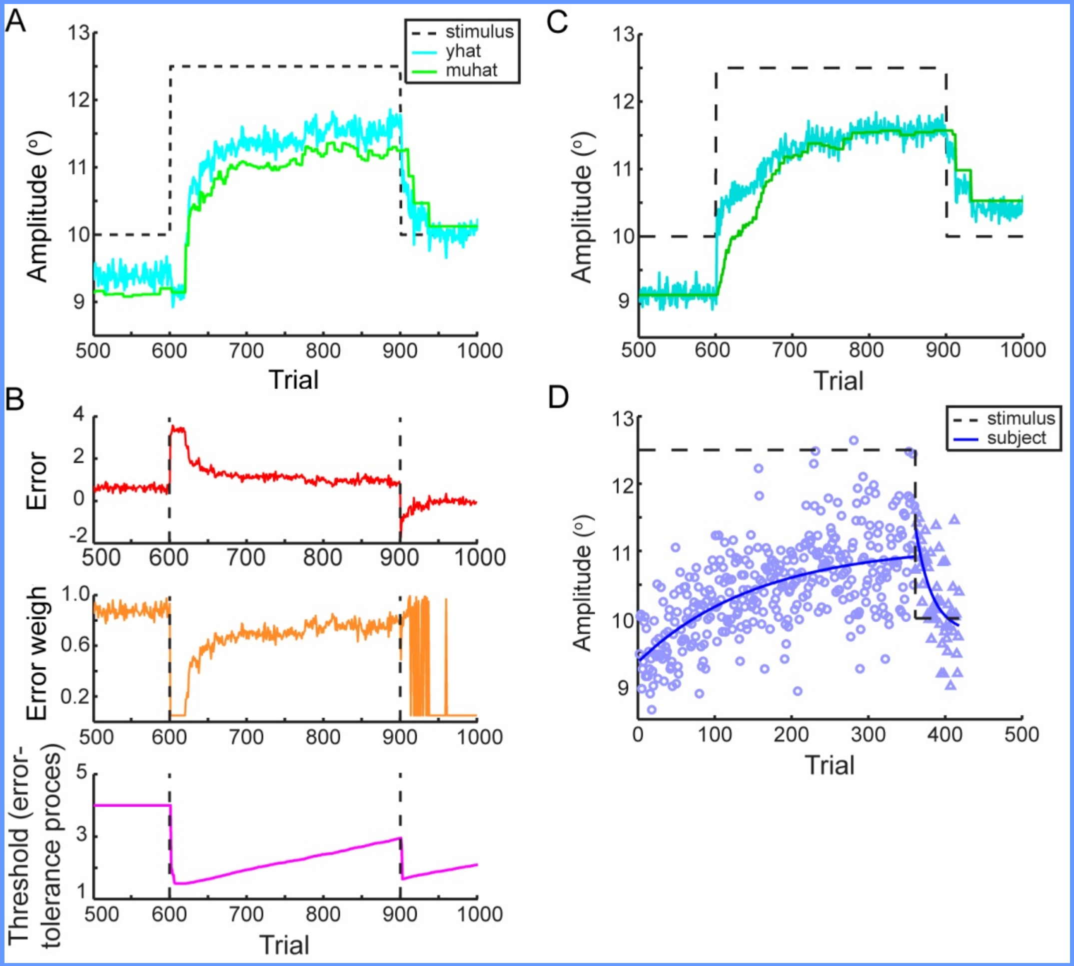
Model prediction: Relationship between Prediction and Adaptation

- As parameters vary, both prediction (α value) and adaptation (rate) change
- These changes describe the differential use of current and prior information to drive future performance
- This predicts a relationship between motor prediction and adaptation tasks in human subjects



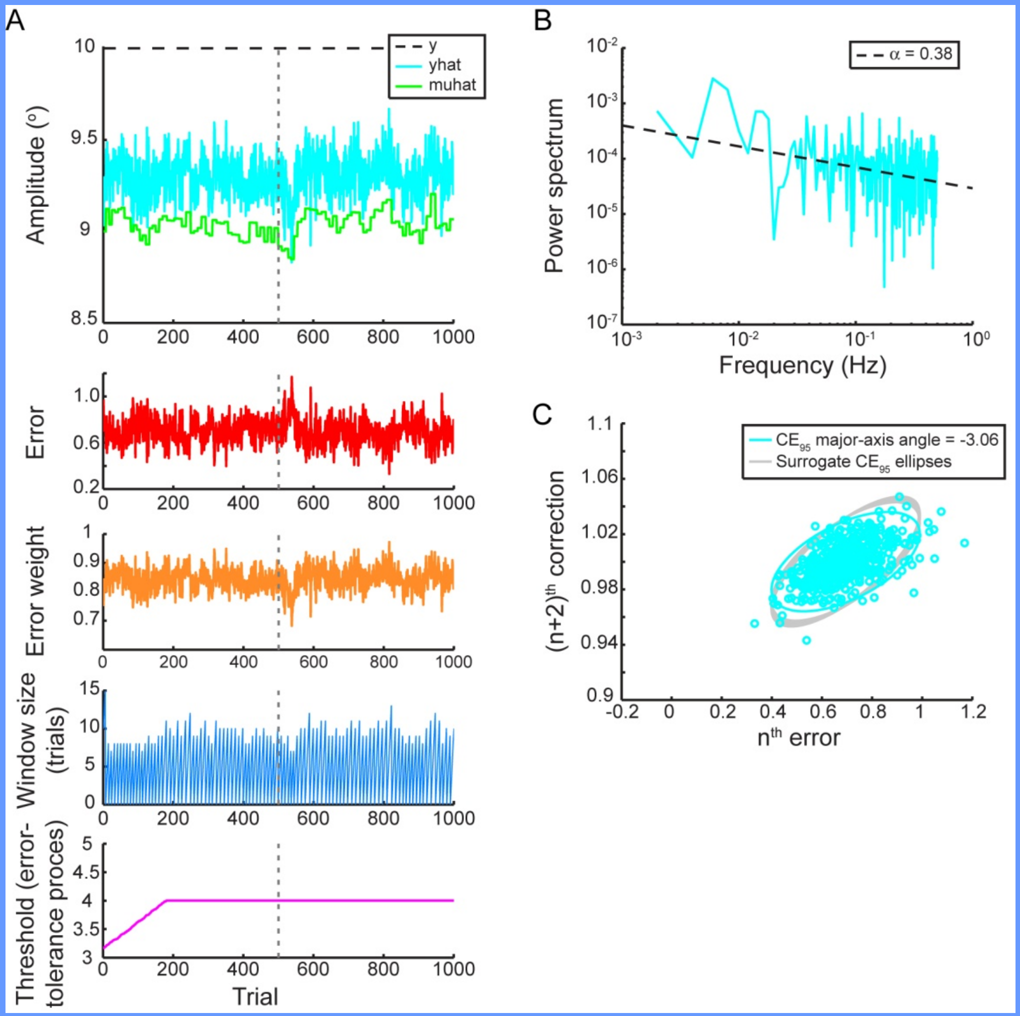
Adaptation Simulation

- Simulation of a conventional adaptation paradigm
- Model responds reasonably well
  - Exponential learning curve
  - Smooth but rapid washout



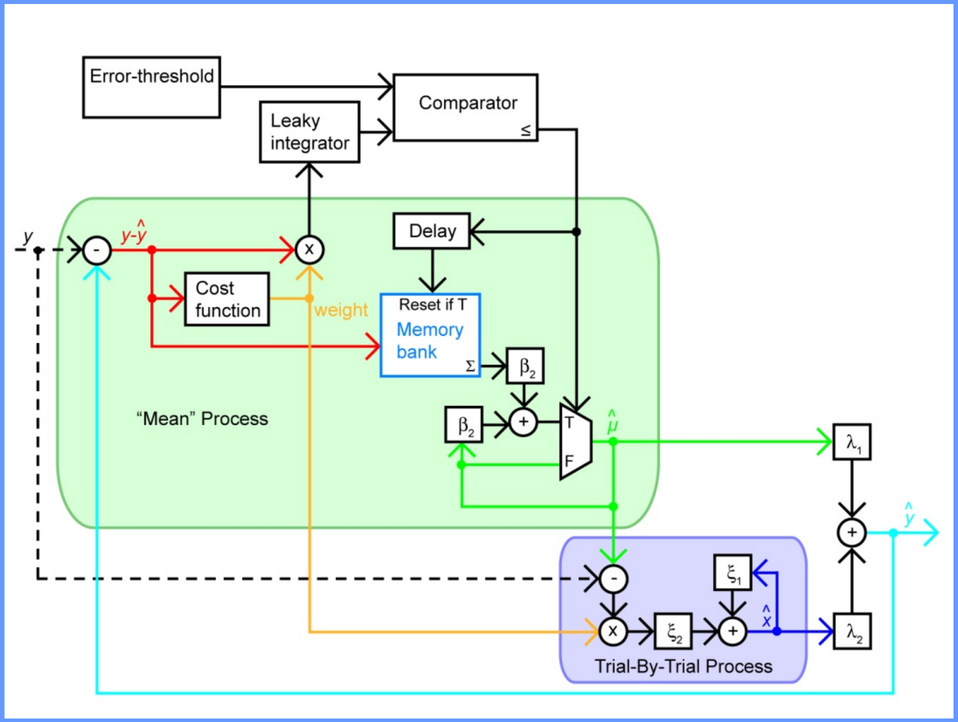
Simulation of predictive saccades

- Model output could match four defining characteristics:
  - Long-memory α value
  - Quality of trial-by-trial corrections
  - Appropriate average model output
  - Response to stimulus variability (α decreases)

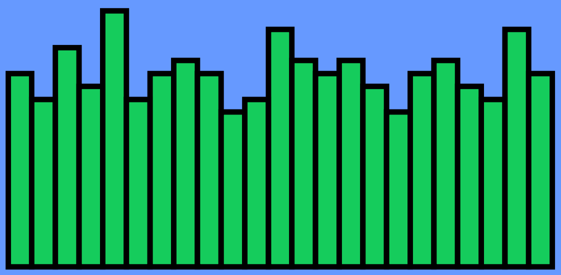


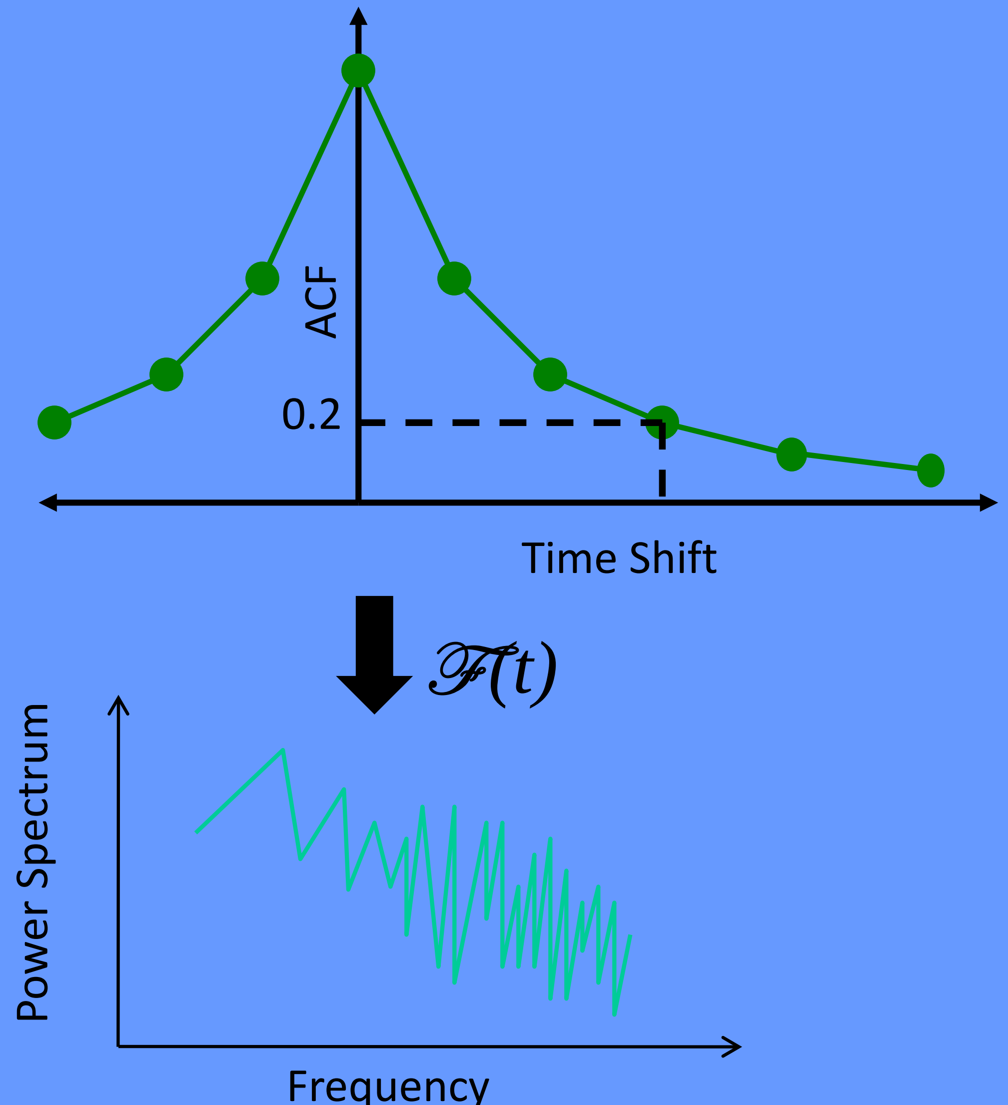
The predictor model

- Trial-by-trial process corrects for immediate errors
- Mean process estimates average model output according to past performance



Analyzing Dynamics Beyond Trial-to-Trial





Initial Motivation:
Phase transition in bimanual movements

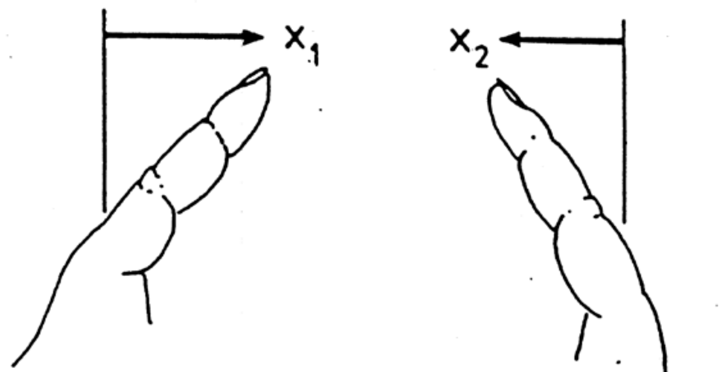


Fig. 2. The displacements of x_1 and x_2 of the finger tips of the left and right hand in the symmetrical (in-phase, homologous) mode

Spontaneous transition from out-of-phase to in-phase as pacing rate increases:

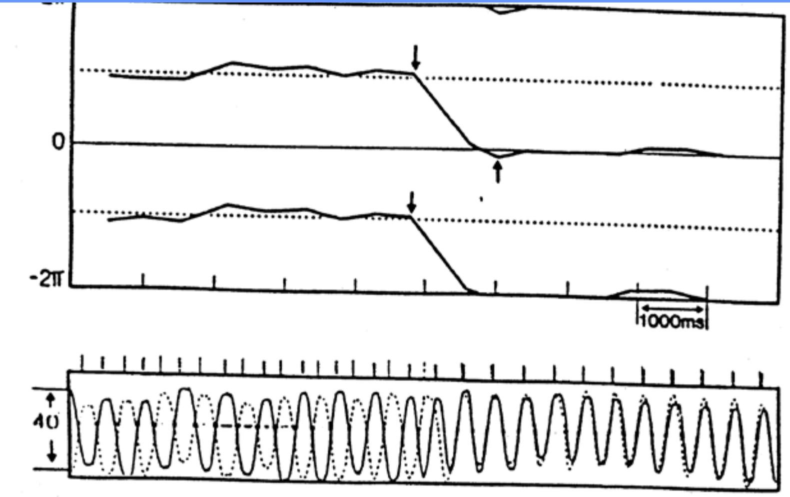
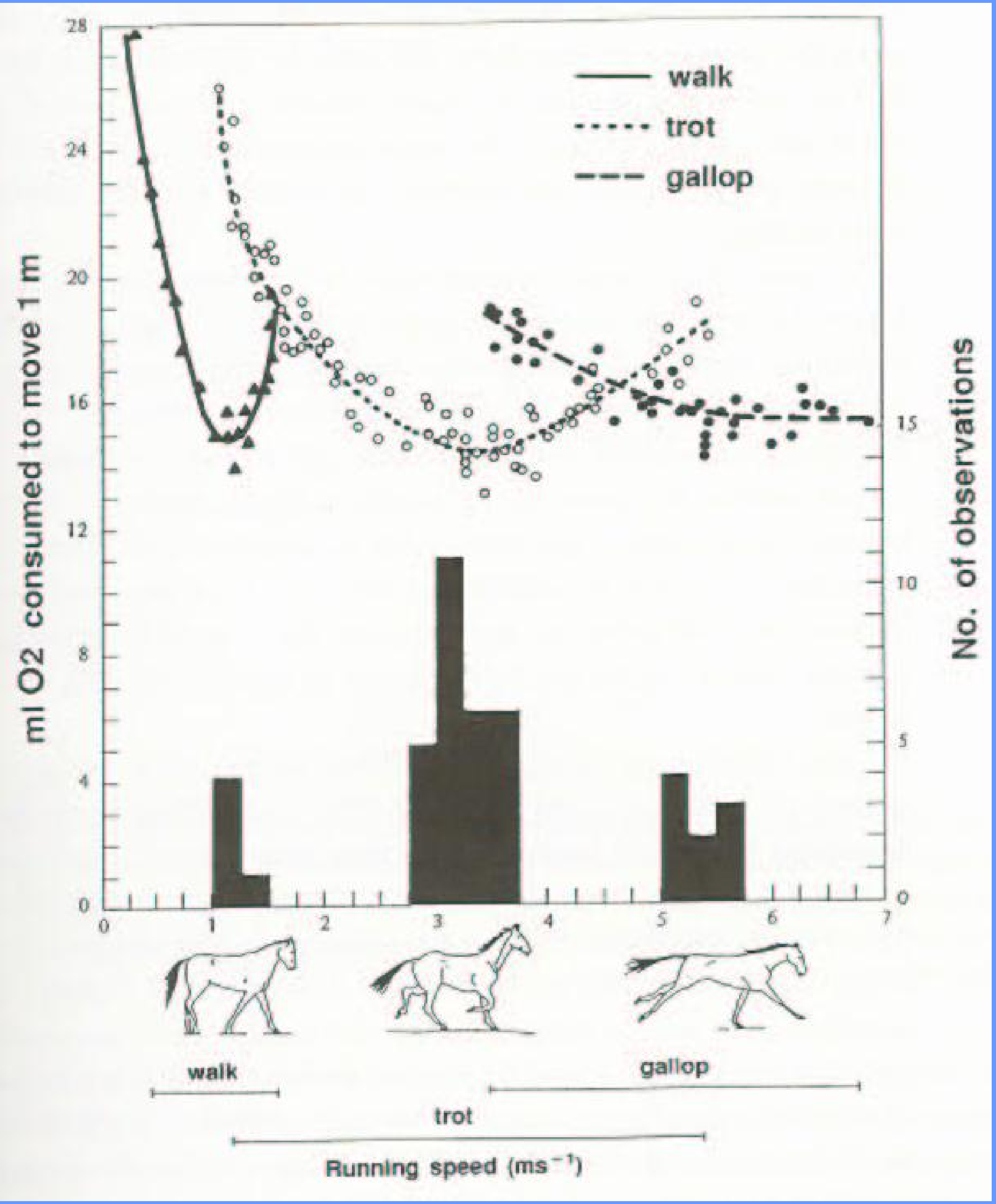


Fig. 1. *Bottom:* Displacements over time of left (solid line) and right (dashed line) hands. The subject is simply increasing cycling frequency in an antisymmetric mode in response to a verbal cue from the experimenter. *Top:* Phase relationship between the two hands. The peaks of one hand movement act as a "target" file and their phase position is calculated continuously relative to the peak-to-peak period of the other "reference" file. The graphic display repeats the phase curve so that phase lags and leads can be noted

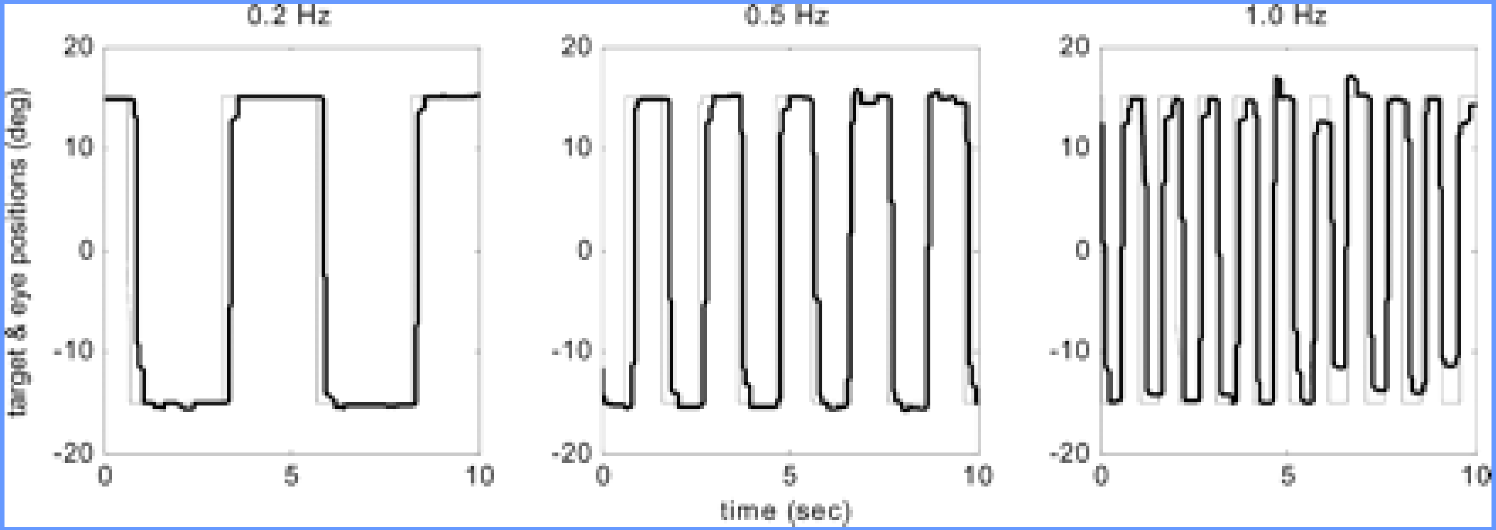
Haken, Kelso, Bunz (1985) A theoretical model of phase transition in human hand movements. Biological Cybernetics 51:347.

Phase Transitions

Abrupt transition from walk to trot to gallop.



Saccade tracking at different pacing frequencies



0.2 Hz Pacing: **reactive** response; eyes lag the target

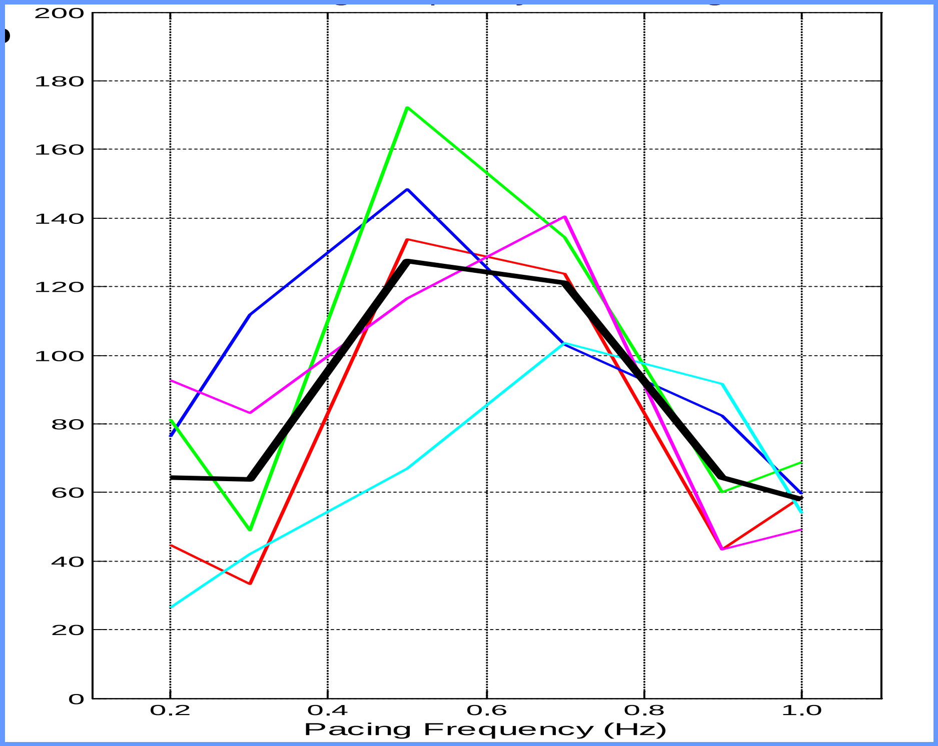
1.0 Hz Pacing: **predictive** response; eyes anticipate the target

0.5 Hz Pacing: mixed response; behavior switches between lagging the target and anticipating the target

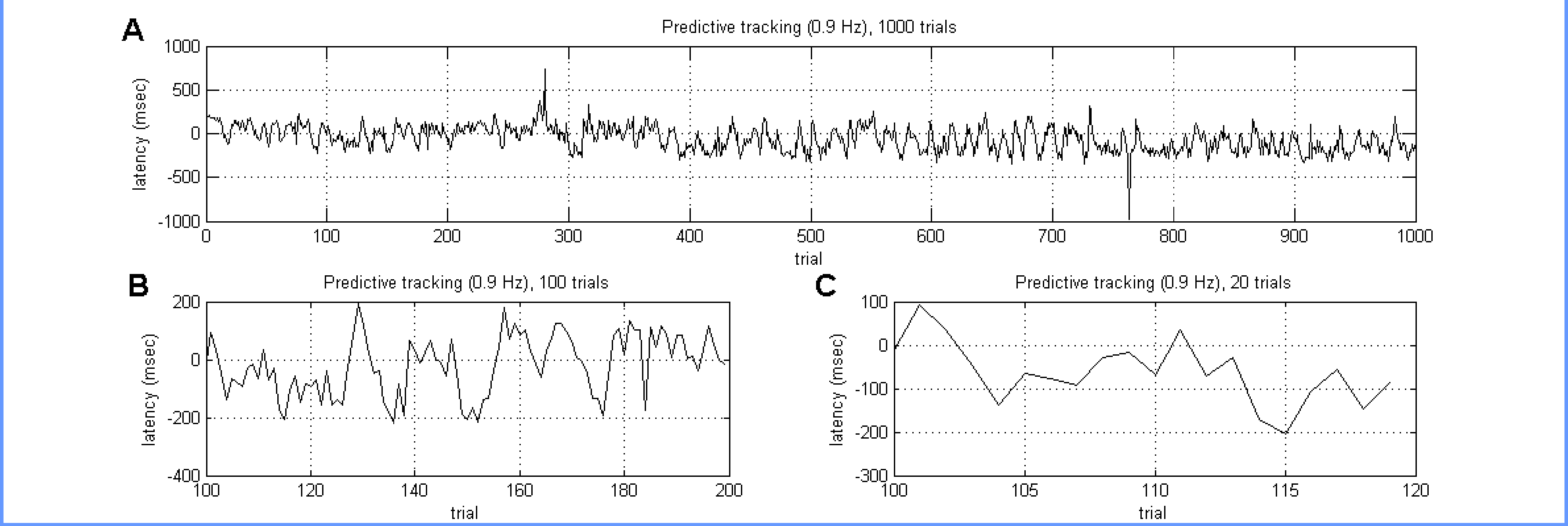
Bistability

Do reactive and predictive states reflect local minima of a "potential function"?

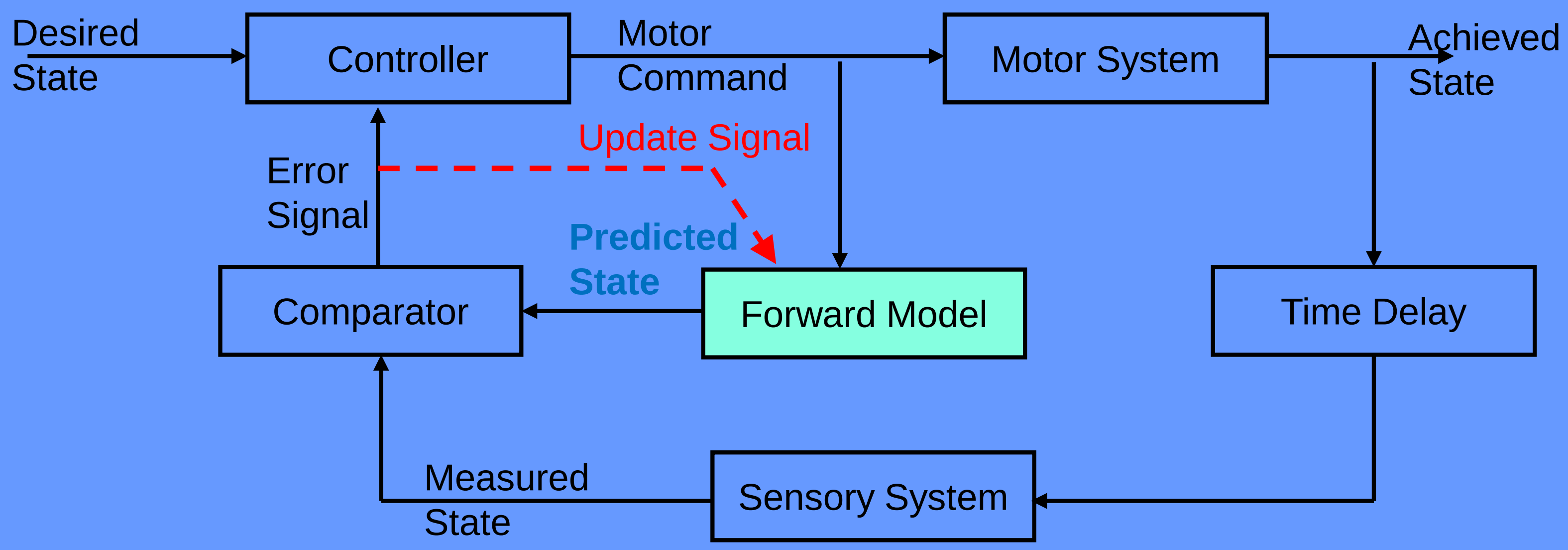
Mean pacing error: Σ|interval error|



Consecutive saccade latencies showing variations on different time scales.



Forward model hypothesis:
Linking prediction to adaptation



➔ Adaptation is based on predicted outcomes

What is the error signal for adaptation?

- Is it related to prediction?

- Test two alternatives:
  - Adaptation is driven by observed (retinal) error
  - Adaptation is driven by difference between *actual* and *expected* movement outcomes

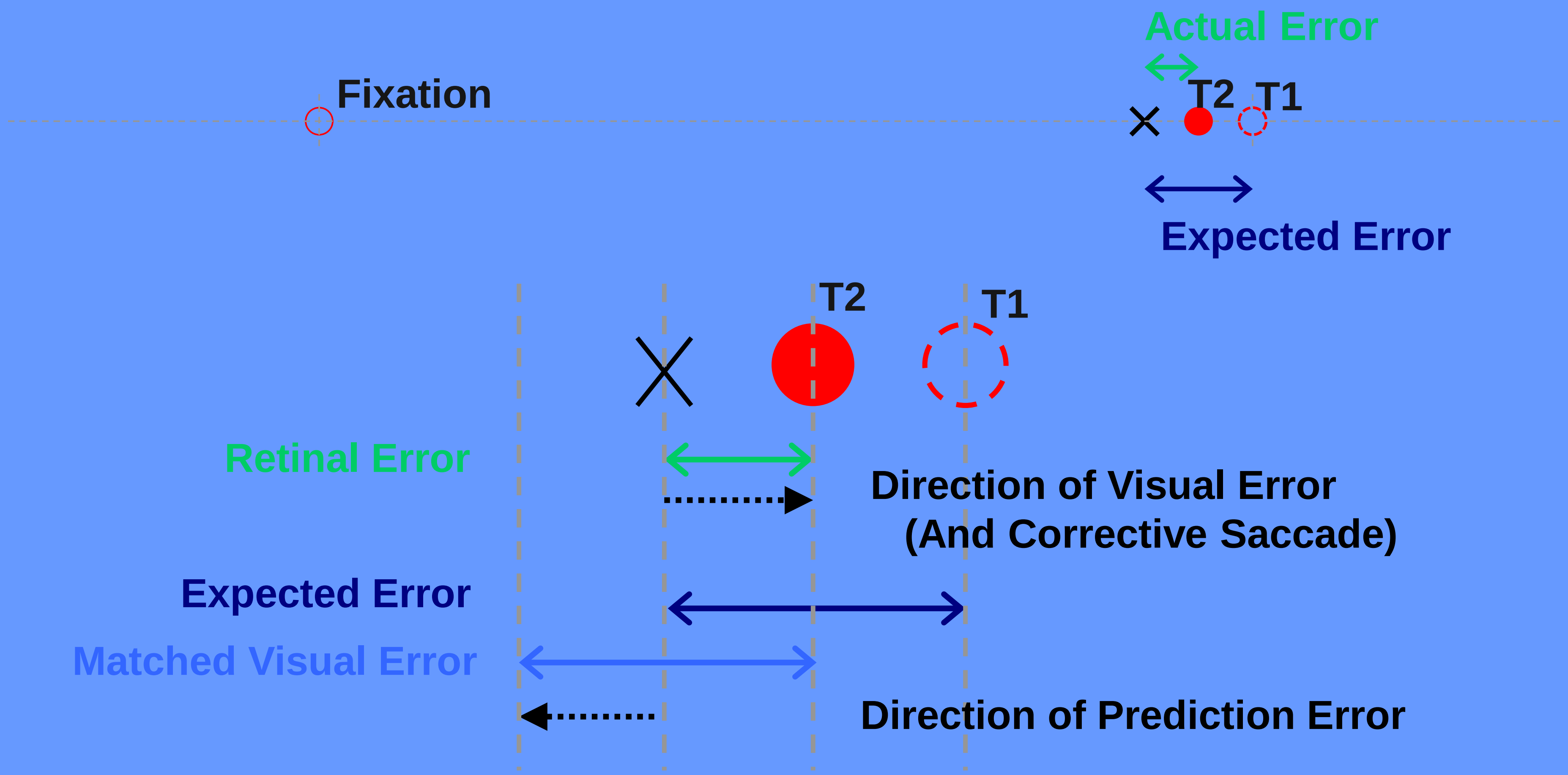
Saccade Adaptation

- Saccades are *adaptable*
  - Move target during saccade – saccades adaptively adjust to go to the new target position
- Saccades are *hypometric*
  - Typically fall short of target by 10%

If they are adaptable, why aren’t they more accurate?

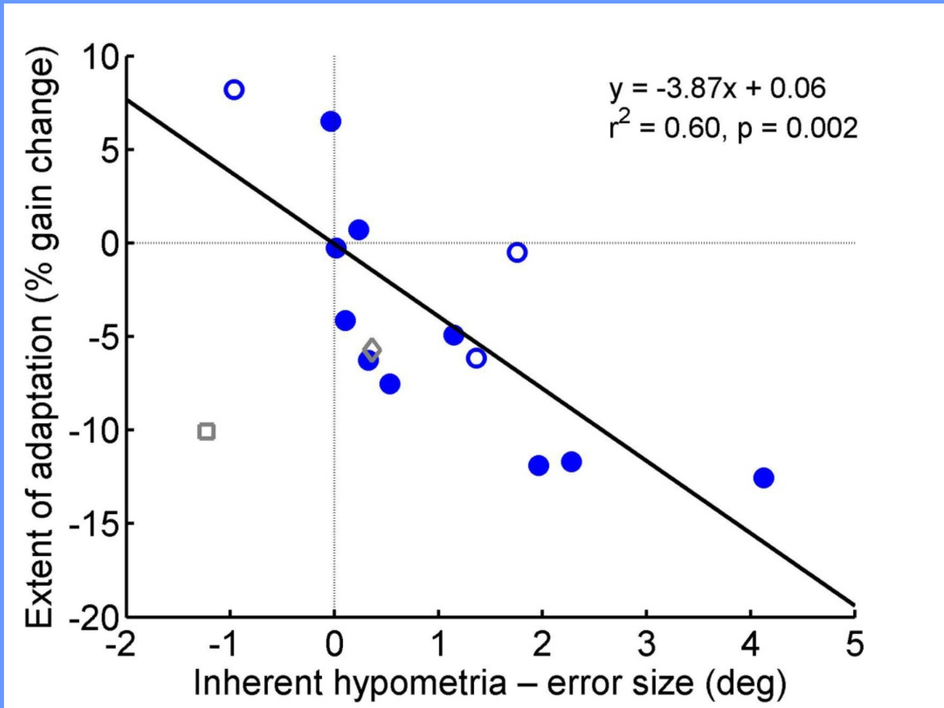
Do they deliberately fall short?

Assessing Adaptation Error Signal



Evidence for Predictive Error Signal

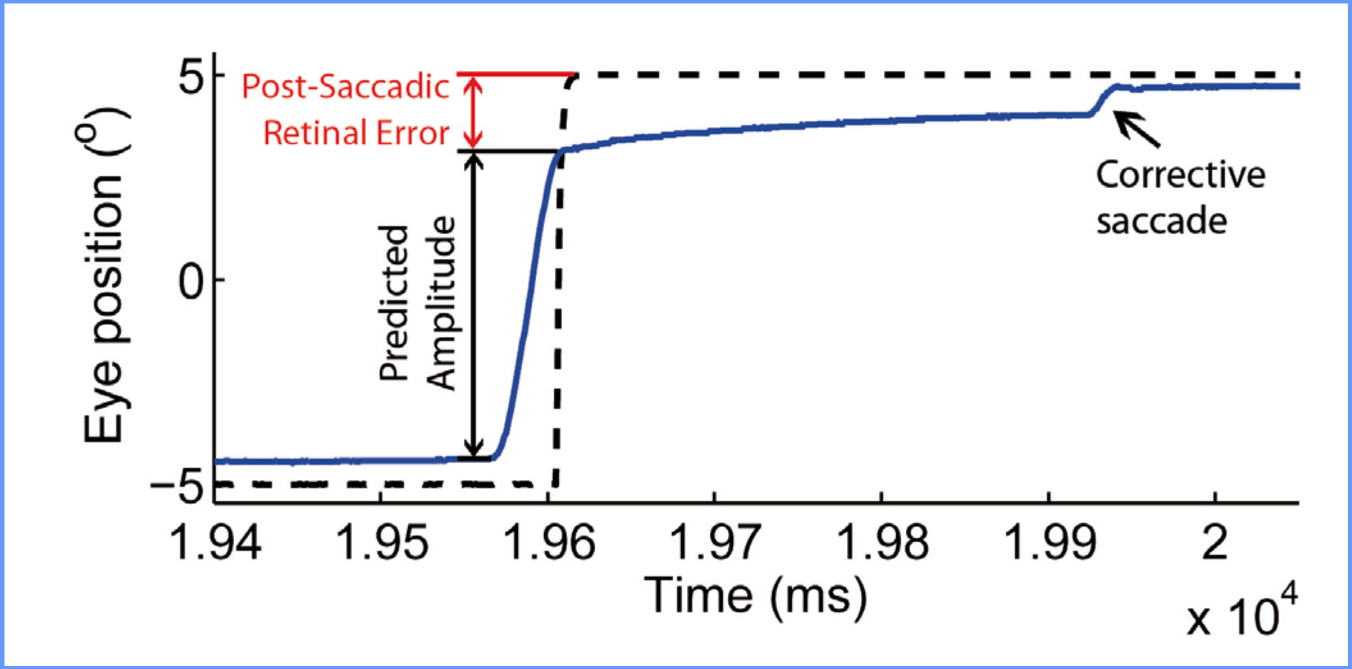
- Adaptation is driven by difference between expected and observed movement outcomes
- Adaptation requires a predictive process



Wong and Shelhamer 2011a

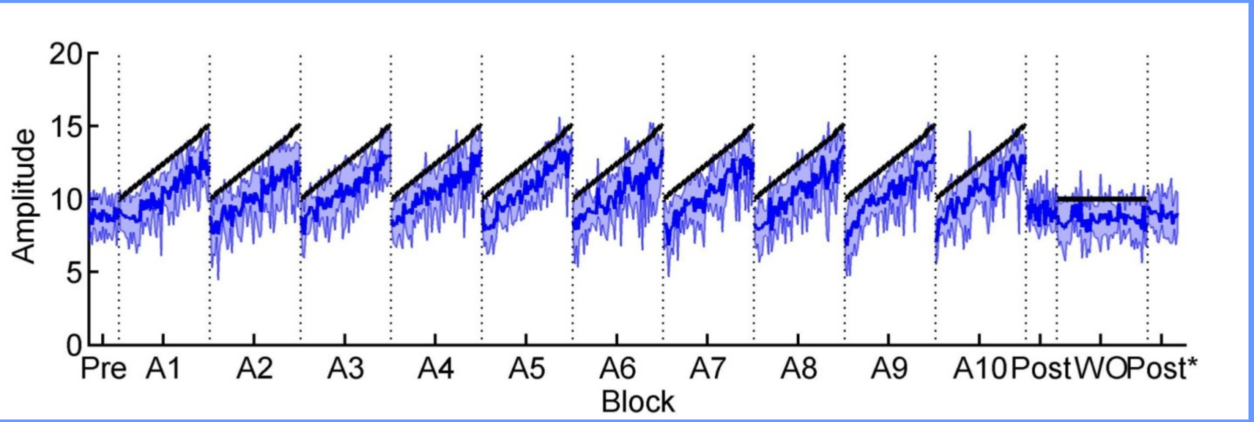
Predictive Saccades

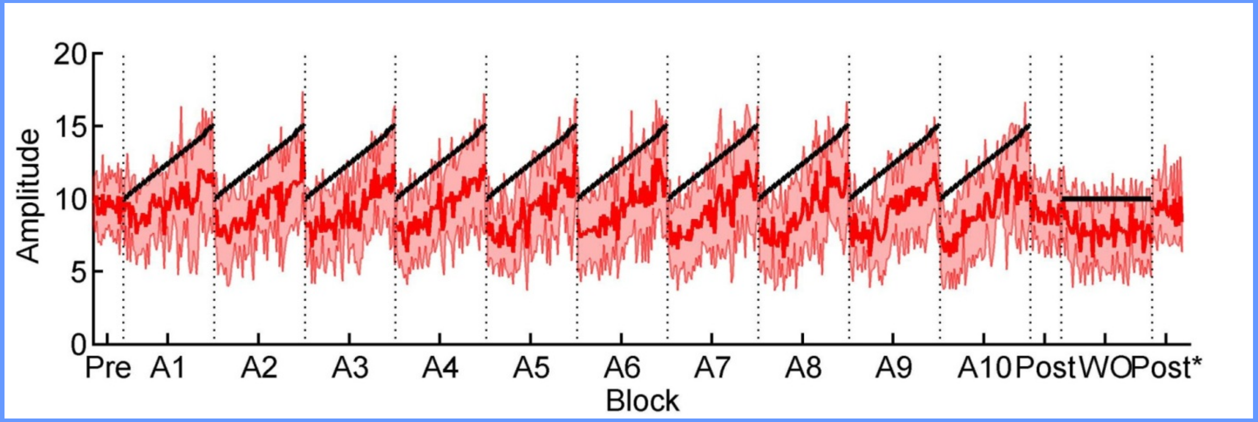
- Error signal = Prediction - Observation
- Can predictive saccades drive adaptation?
  - Is the error signal shared?



Prediction-Driven Adaptation

- 10 blocks, 60 saccades, increasing amplitude
- Controls: gain increase (3.93%)
- SCA6: gain decrease (-5.75%)



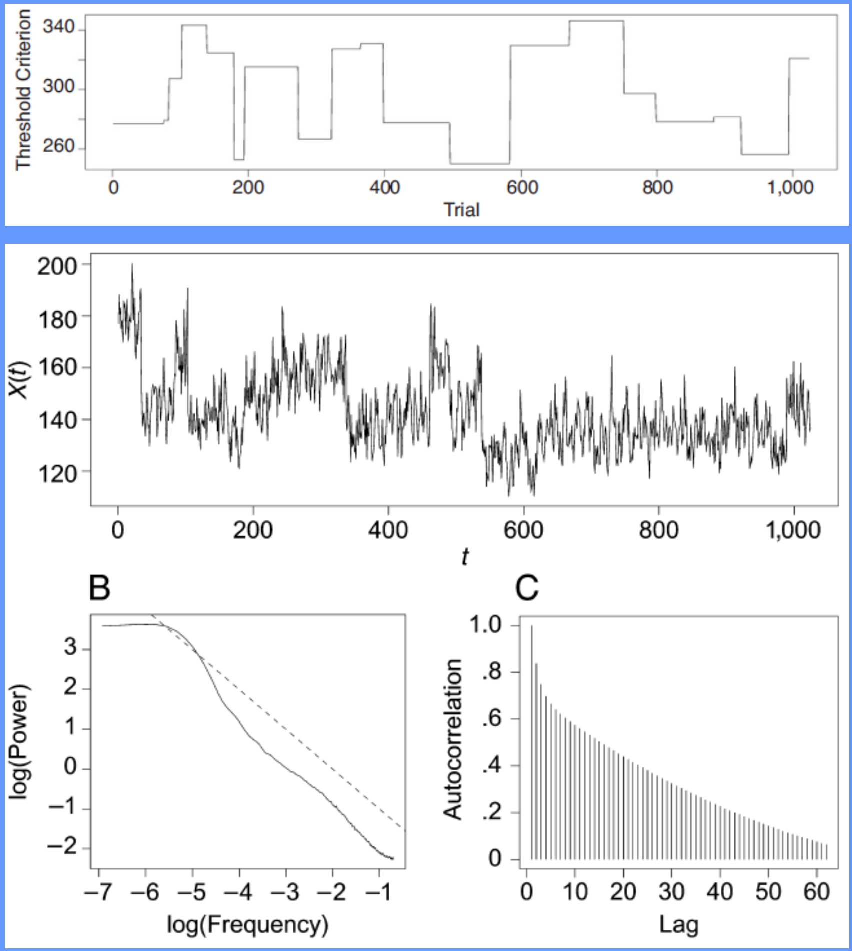


Can we model this?

- Predictive process (internal model)
- Drives adaptation
- Generates fractal fluctuations
  - Long memory

Model, Phase I: Predictor

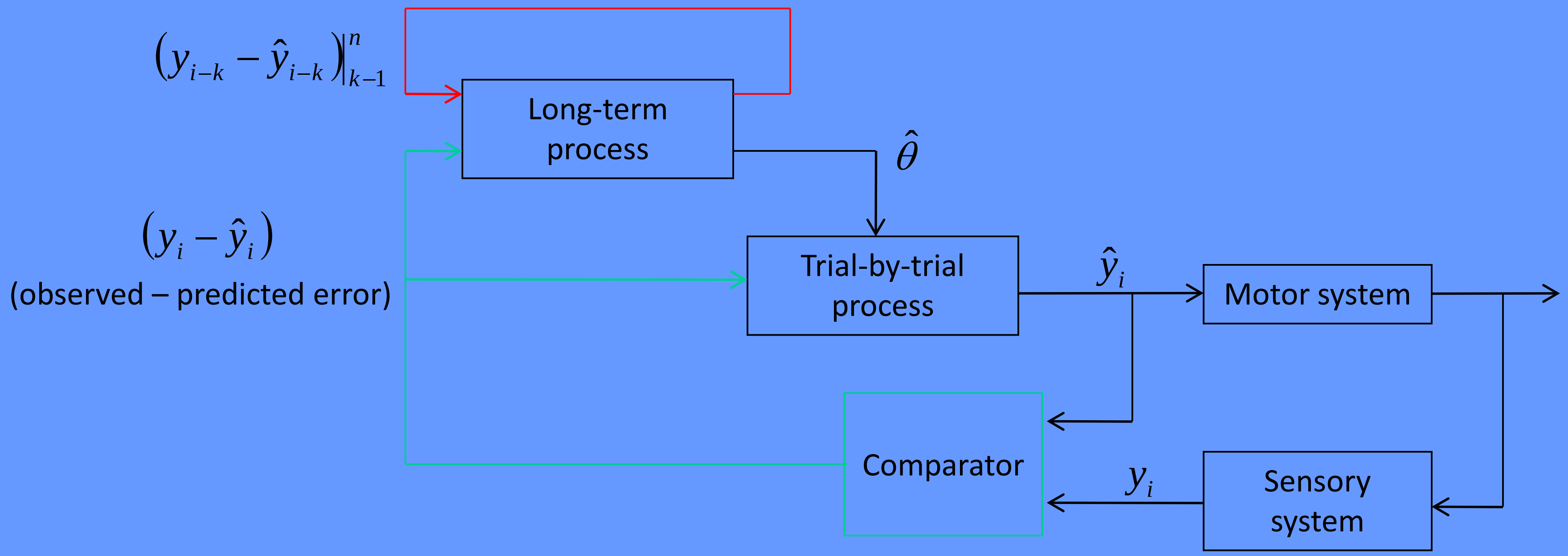
- Long-memory may be mimicked (Diebold & Inoue 2001) by:
  - Aggregate of short-memory processes with different time scales
  - Regime-switching (switching of parameter values based on…)



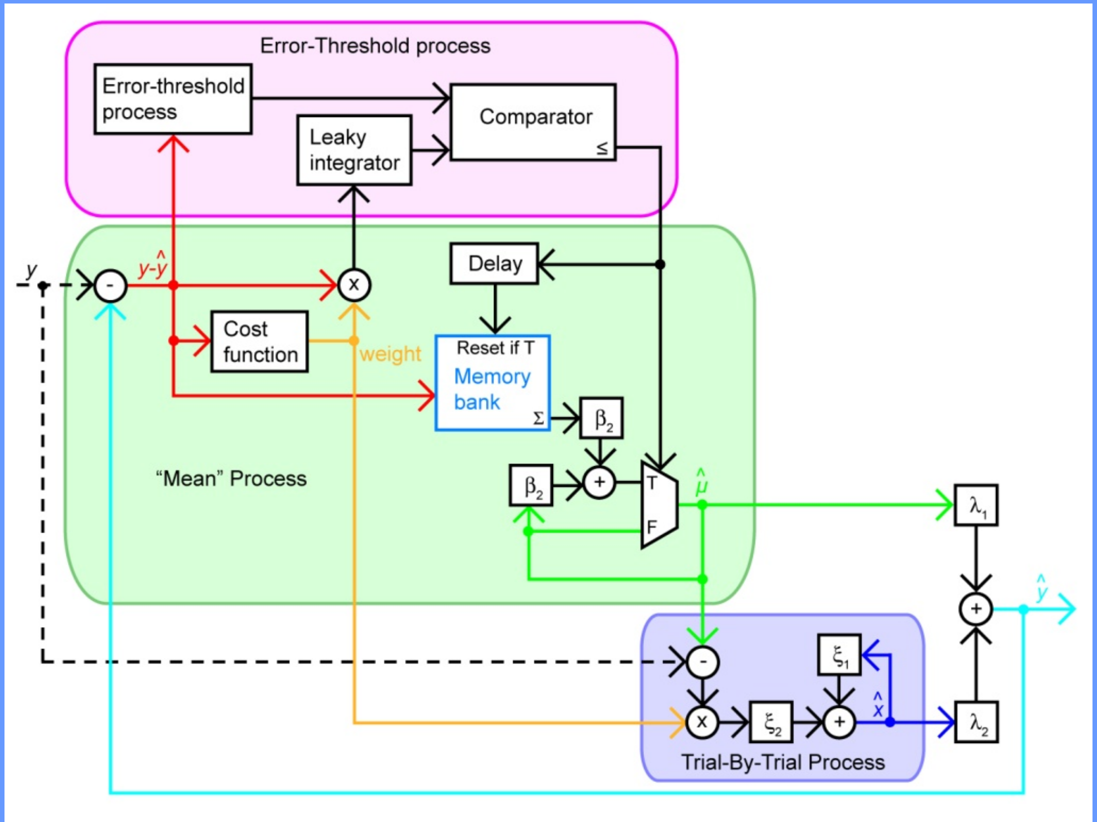
Wagenmakers et al. 2004

Model, Phase I: Predictor

- Trial-by-trial process corrects for immediate errors
- Slow process modulates error-corrector parameters based on prior performance



Component process	Time scale	Responsibility
trial-by-trial process	Fast	Rapidly adjust model output to address single-trial variability (errors and noise)
mean process	Moderate	When error-accumulation reaches threshold, re-estimate "mean" of the process by averaging over prior model outputs (only including values since the last regime switch; at most ~20 trials)
error-tolerance process	Slow	Adjust cumulative-error threshold used to determine when regime switching occurs, based on prior performance

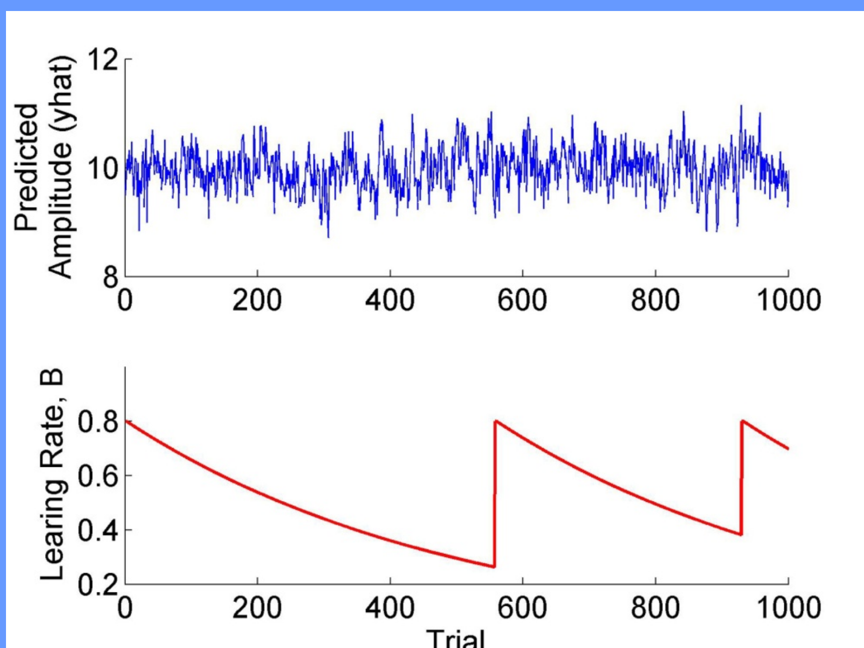


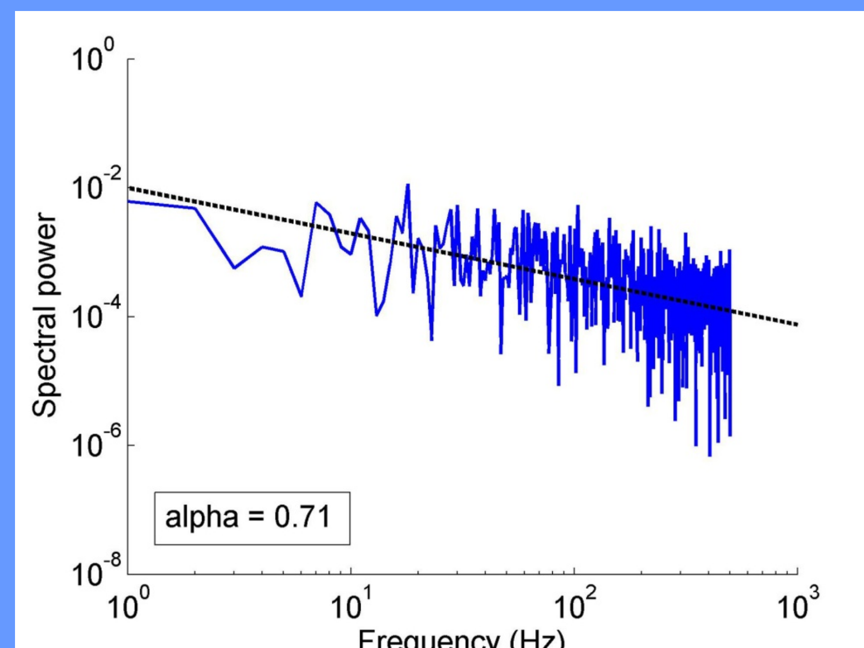
Model, Phase I: Predictor

- Initial simulation results

$$\hat{y}_{i+1} = \hat{y}_i + B_i\left(y_i - \hat{y}_i\right) + \varepsilon_i \qquad \varepsilon_i \sim N\left(0, \sigma_r{}^2\right)$$

$$B_i = \begin{cases} B_o & \sigma_{j=i-n:i}\left(y_j - \hat{y}_j\right) \geq k \\ B_o \mathrm{e}^{\mathrm{C_o}\,(1-(i-\tau))} & \textit{otherwise}; \quad \tau = \text{last } B_i \text{ reset} \end{cases}$$





Saccade Endpoints

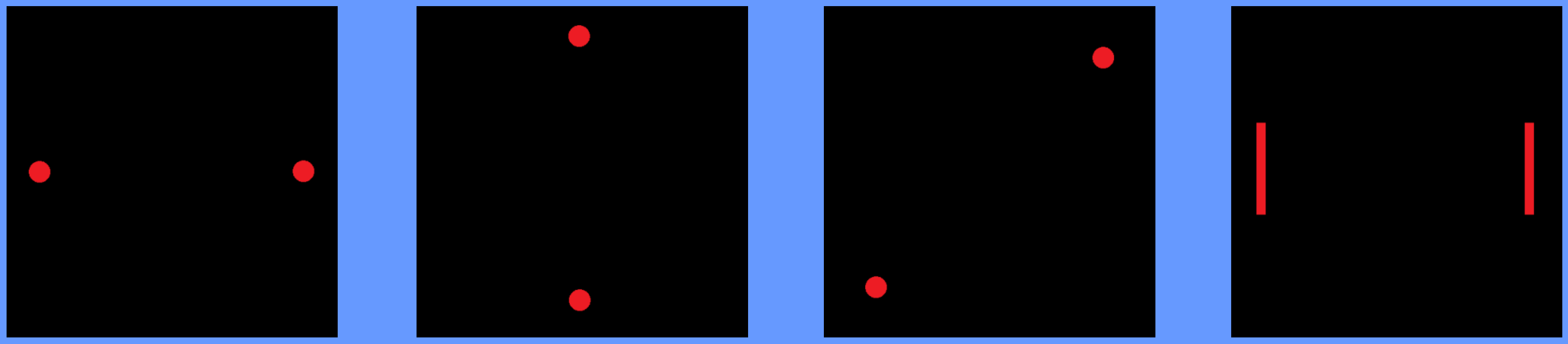
Is position controlled differently along the direction of the saccade?

We posit stronger correlations along the primary saccade direction then along the orthogonal saccade direction, because control along the primary axis is more important.

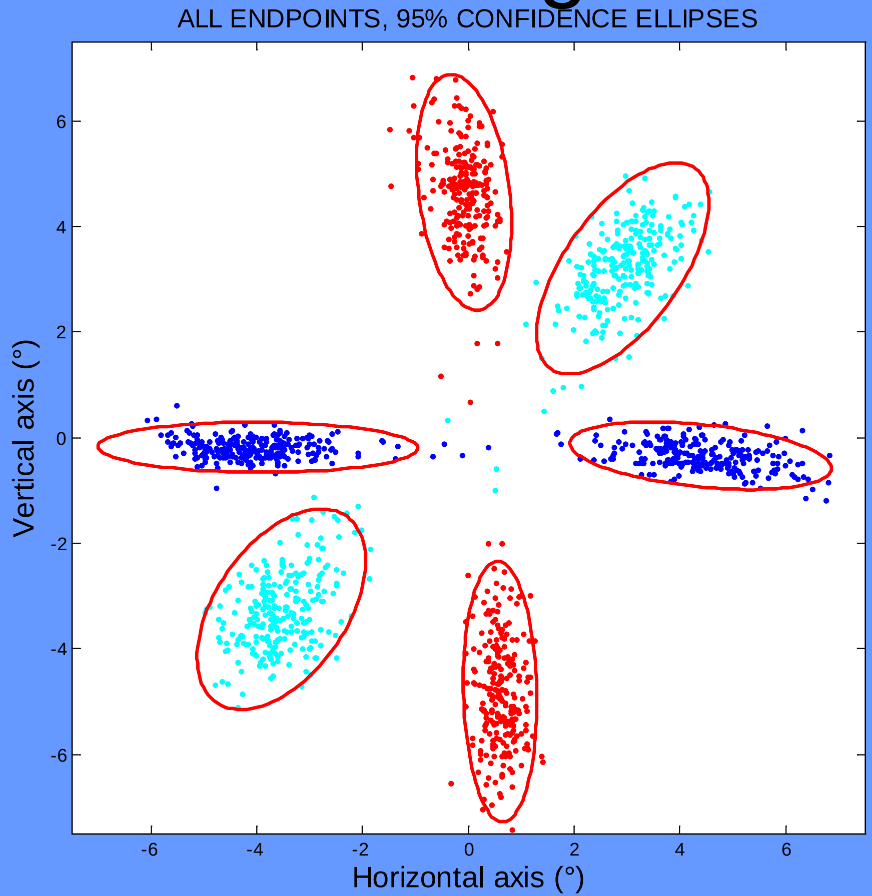
EXPERIMENTAL SETUP

Four conditions

Measured saccade endpoints along direction of targets, and orthogonal



Saccade Endpoints, all Conditions, Point Targets



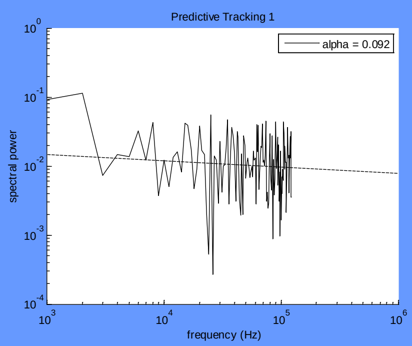
- Figure shows saccade endpoints for one subject, and 95% confidence ellipses
- Most variability lies in primary direction

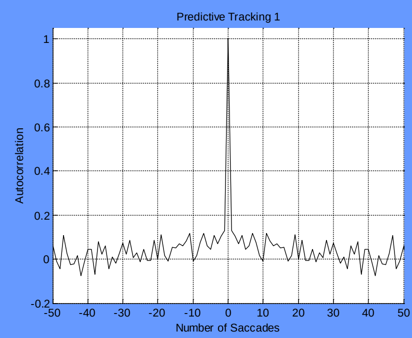
Horizontal saccades – Point Targets

endpoint error along the PRIMARY saccade direction

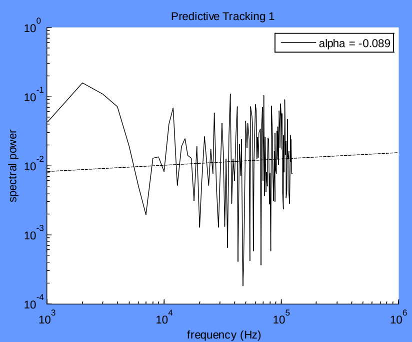
endpoint error along the ORTHOGONAL saccade direction

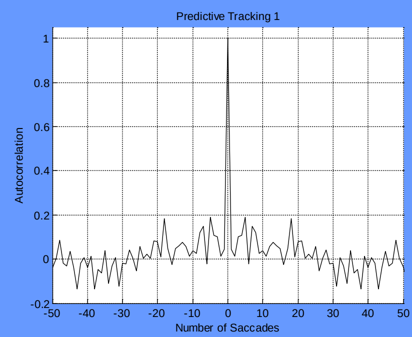
Rightward saccades





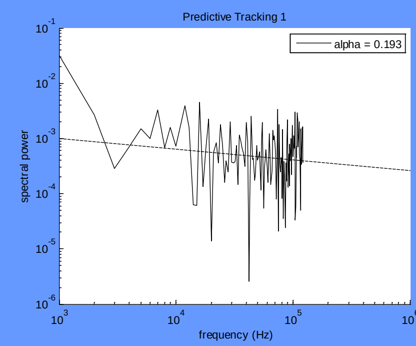
Leftward saccades

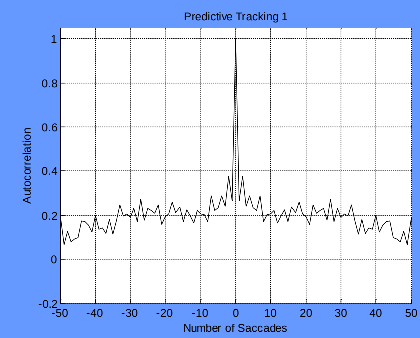




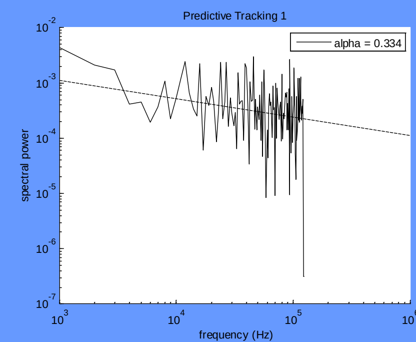
α primary direction= 0.121±0.194

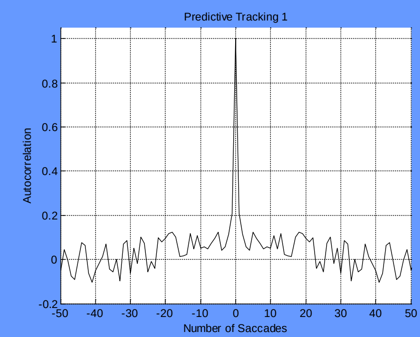
Rightward saccades





Leftward saccades





α orthogonal direction = 0.371±0.193

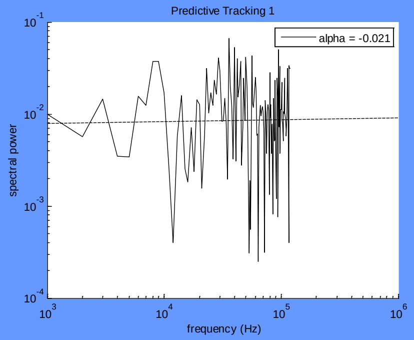
Scaling exponents (*α* values) of endpoint error along orthogonal direction significantly larger than those along primary direction (P=0.004)

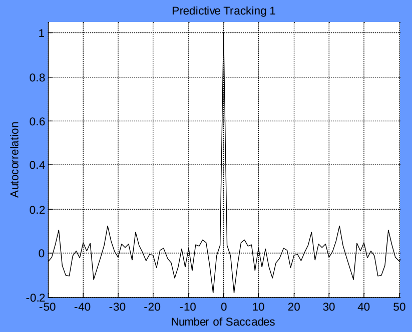
Vertical saccades – Point Targets

endpoint error along the PRIMARY saccade direction

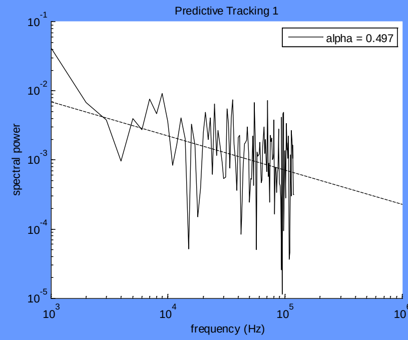
endpoint error along the ORTHOGONAL saccade direction

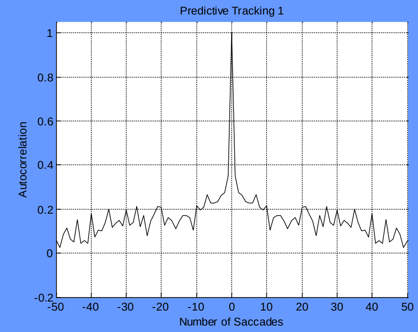
Upward saccades



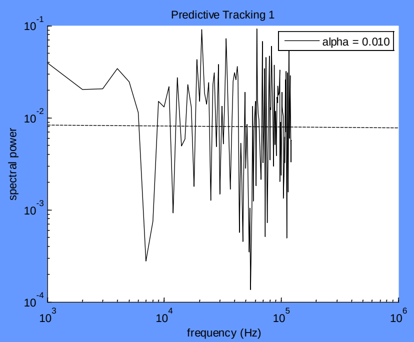


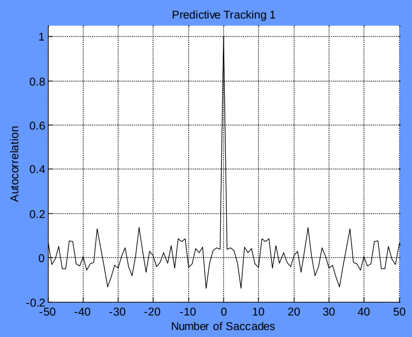
Upward saccades



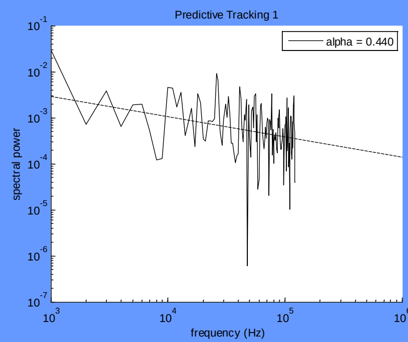


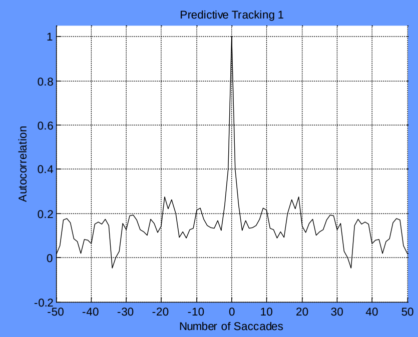
Downward saccades





Downward saccades





α primary direction= 0.154±0.186

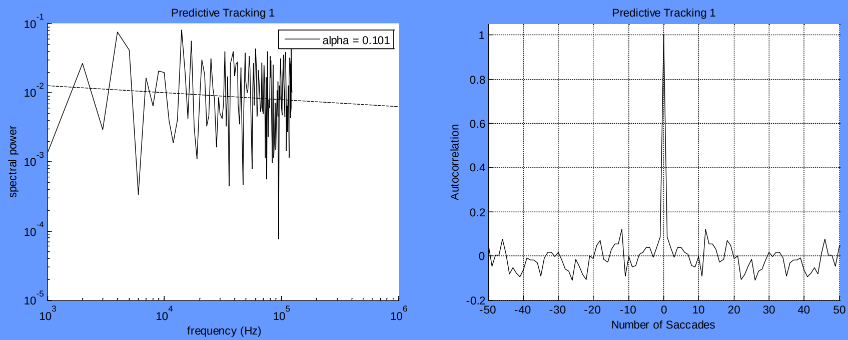
α orthogonal direction = 0.338 ±0.197

Scaling exponents (*α* values) of endpoint error along orthogonal direction significantly larger than those along primary direction (P=0.035)

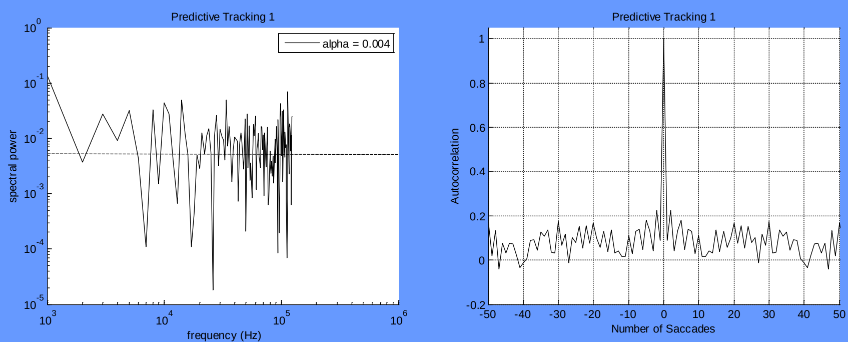
Horizontal saccades – Line Targets

endpoint error along the PRIMARY saccade direction

Rightward saccades

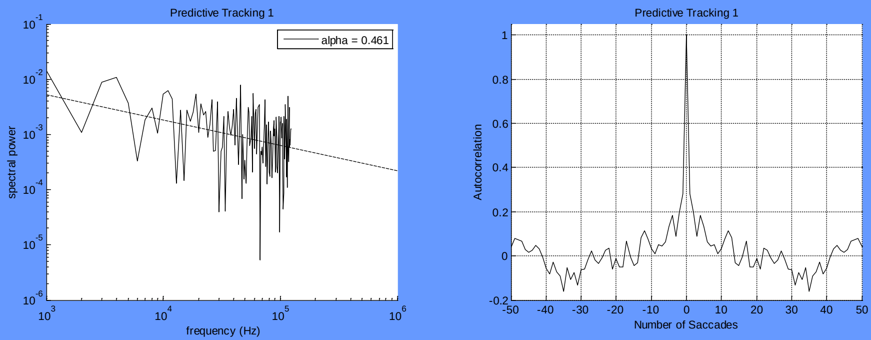


Leftward saccades

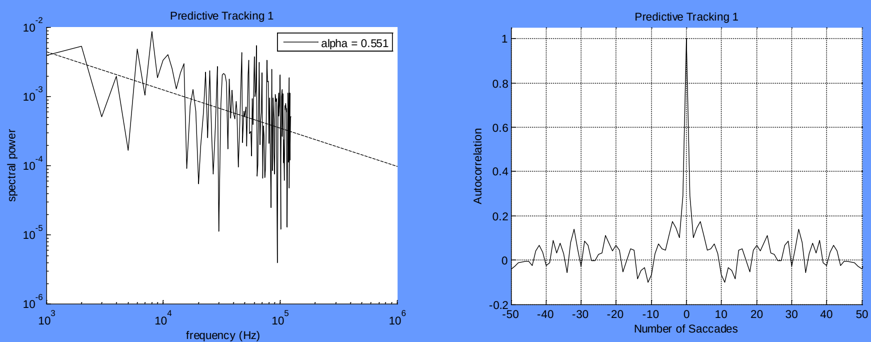


endpoint error along the ORTHOGONAL saccade direction

Rightward saccades



Leftward saccades



α primary direction= 0.185±0.196

α orthogonal direction = 0.632 ±0.191

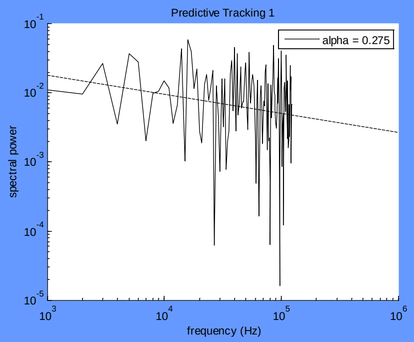
Scaling exponents (*α* values) of endpoint error along orthogonal direction significantly larger than those along primary direction (P=0.001)

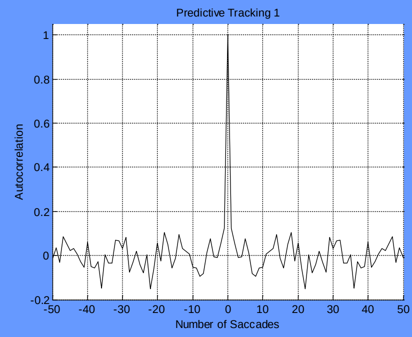
Oblique saccades – Point Targets

endpoint error along PRIMARY saccade direction

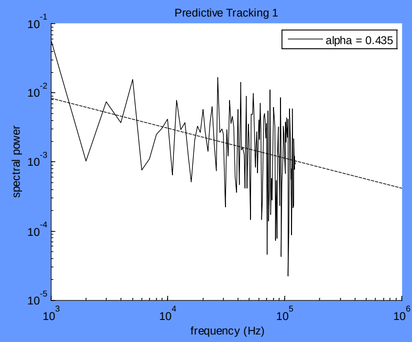
endpoint error along ORTHOGONAL saccade direction

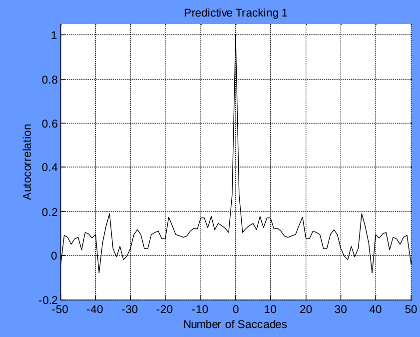
Rightward saccades (primary)



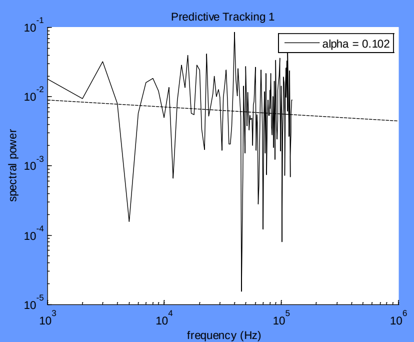


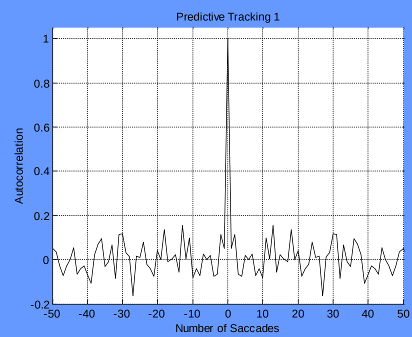
Rightward saccades (orthogonal)



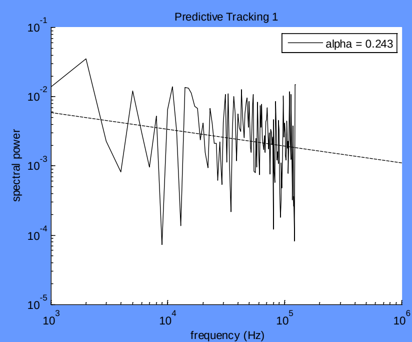


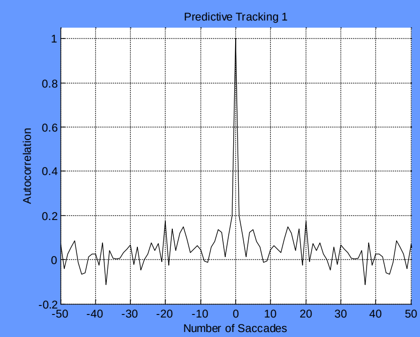
Leftward saccades (primary)





Leftward saccades (orthogonal)





α primary direction= 0.2780.221)

α orthogonal direction = 0.315±0.130

Scaling exponents (*α* values) of endpoint error along orthogonal direction tend to be larger than along primary direction

Summary

	Horizontal saccades *Target: dot*	*Vertical saccades* *Target: dot*	*Horizontal saccades* *Target: vertical line*	*Oblique saccades* *Target: dot*
Subject	**α primary/α orthogonal**	**α primary/α orthogonal**	**α primary/α orthogonal**	**α primary α orthogonal**
1	0.002 / 0.264 *	-0.006 / 0.469 *	0.052 / 0.506 *	0.237 / 0.369 *
2	0.018 / 0.158 *	0.079 / 0.215 *	0.276 / 0.456 *	0.077 / 0.326 *
3	0.071 / 0.551 *	0.425 / 0.155	-0.044 / 0.717 *	0.476 / 0.178
4	0.219 / 0.275 *	-0.001 /0.198 *	0.137 / 0.725 *	0.210 / 0.386 *
5	0.286 / 0.346 *	0.183 /0.208 *	0.324 / 0.610 *	0.017 / 0.204 *
6	-0.009 / 0.417 *	0.303 / 0.484 *	0.139 / 0.550 *	0.339 / 0.315
7	0.423 / 0.218	0.088 / 0.631 *	0.408 /0.856 *	0.588 / 0.426
9	-0.077 / 0.624 *			
10	0.163 / 0.488 *			

CONCLUSION

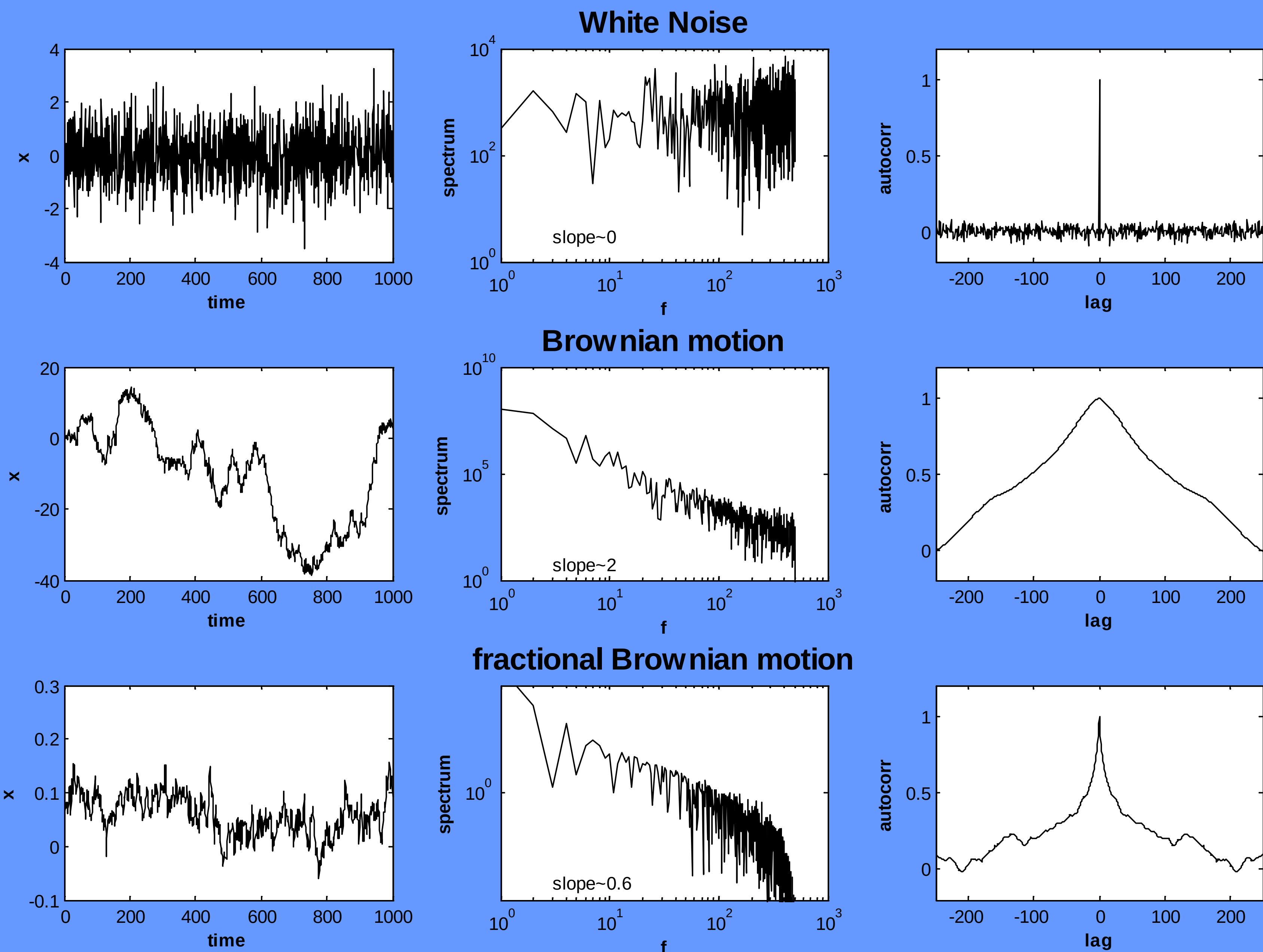
Why is scaling greater in the orthogonal direction?

- Endpoints are errors

  Strong scaling (large value of α) means that errors have information in them (information persists between trials).

- Most effective use of error information in primary direction leads to uncorrelated errors (residuals).

Examples: time, frequency, autocorrelation



Thoughts on predictive timing

- System exhibits a *preference for prediction.*
- "Prediction is hard, especially about the future." - Niels Bohr or Yogi Berra.
- If you are clever, you can design an experiment to eliminate one of the defining features of human behavior – prediction. But you have to be very clever.

Trial-by-trial adjustment

- Errors are corrected on the “next” saccade:
  - (n+1): return primary saccade
  - (n+2): primary saccade in same direction
- Errors are not fully corrected
  - Surrogate data
  - An additional process improves performance