Time Series Analysis in AFNI

Outline: 6+ Hours of Edification

- Philosophy (e.g., theory without equations)
- Sample FMRI data
- Theory underlying FMRI analyses: the **HRF**
- "Simple" or "Fixed Shape" regression analysis
  - Theory and Hands-on examples
- "Deconvolution" or "Variable Shape" analysis
  - Theory and Hands-on examples
- Advanced Topics (followed by brain meltdown)

Goals: Conceptual Understanding + Prepare to Try It Yourself

Data Analysis Philosophy

- **Signal** = Measurable response to stimulus
- **Noise** = Components of measurement that interfere with detection of signal
- Statistical detection theory:
  - Understand relationship between stimulus & signal
  - Characterize noise statistically
  - Can then devise methods to distinguish noise-only measurements from signal+noise measurements, and assess the methods' reliability
  - Methods and usefulness depend strongly on the assumptions
    - Some methods are more "robust" against erroneous assumptions than others, but may be less sensitive

FMRI Philosopy: Signals and Noise

- FMRI Stimulus→Signal connection and noise statistics are both complex and poorly characterized
- Result: there is no "**best**" way to analyze FMRI time series data: there are only "**reasonable**" analysis methods
- To deal with data, must make some assumptions about the signal and noise
- Assumptions will be wrong, but must do ***something***
- Different kinds of experiments require different kinds of analyses
  - Since signal models and questions you ask about the signal will vary
  - It is important to **understand** what is going on, so you can select and evaluate "reasonable" analyses

Meta-method for creating analysis methods

- Write down a mathematical model connecting stimulus (or "activation") to signal
- Write down a statistical model for the noise
- Combine them to produce an equation for measurements given signal+noise
  - Equation will have unknown parameters, which are to be estimated from the data
  - N.B.: signal may have zero strength (no "activation")
- Use statistical detection theory to produce an algorithm for processing the measurements to assess signal presence and characteristics
  - e.g., least squares fit of model parameters to data

Time Series Analysis on Voxel Data

- Most common forms of FMRI analysis involve fitting an activation+BOLD model to each voxel's time series ***separately*** (AKA "univariate" analysis)
  - ➢ Some pre-processing steps do include inter-voxel computations; e.g.,
    - o spatial smoothing to reduce noise
    - o spatial registration to correct for subject motion
- Result of model fits is a set of parameters at each voxel, estimated from that voxel's data
  - ➢ e.g., activation amplitude (β), delay, shape
  - ➢ "SPM" = statistical parametric map; e.g., β or t or F
- Further analysis steps operate on individual SPMs
  - ★ e.g., combining/contrasting data among subjects
    - o sometimes called "second level" or "meta" analysis

Some Features of FMRI Voxel Time Series

- FMRI only measures changes due to neural "activity"
  - Baseline level of signal in a voxel means little or nothing about neural activity
  - Also, baseline level tends to drift around slowly (100 s time scale or so; mostly from small subject motions)
- Therefore, an FMRI experiment must have at least 2 different neural conditions ("tasks" and/or "stimuli")
  - Then statistically test for differences in the MRI signal level between conditions
  - Many experiments: one condition is "rest"
- Baseline is modeled separately from activation signals, and baseline model includes "rest" periods
  - In AFNI, that is; in SPM, "rest" is modeled explicitly

Why FMRI Analysis Is Hard

- Don’t know true relation between neural “activity” and BOLD signal:
  - What *is* neural “activity”, anyway?
  - What is connection between “activity” and hemodynamics and MRI signal?

- Noise in data is poorly characterized
  - In space and in time, and in its origin
  - Noise amplitude ≥ BOLD signal
    - Can some of this noise be removed by software?
  - Makes both signal detection and statistical assessment hard
    - Especially with 20,000+ voxels in the brain **=** 20,000+ activation decisions

Why So Many Methods of Analysis?

- Different assumptions about activity-to-MRI signal connection

---

- Different assumptions about noise (≡ signal fluctuations of no interest) properties and statistics

---

- Different experiments and different questions about the results

---

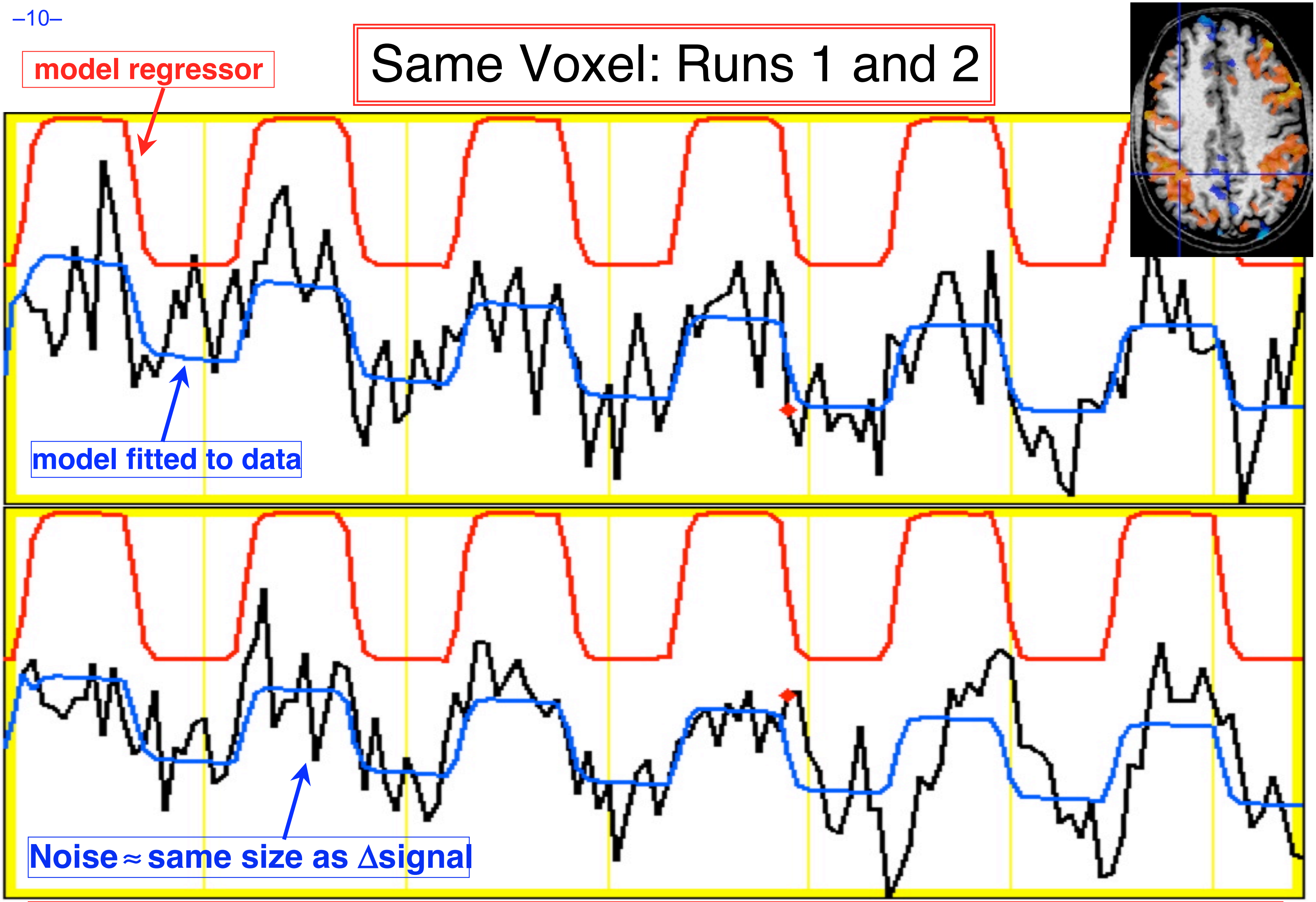
- **Result**: ∃ Many "reasonable" FMRI analysis methods

- Researchers **must** understand the tools (models and software) in order to make choices and to detect glitches in the analysis!!

Some Sample FMRI Data Time Series

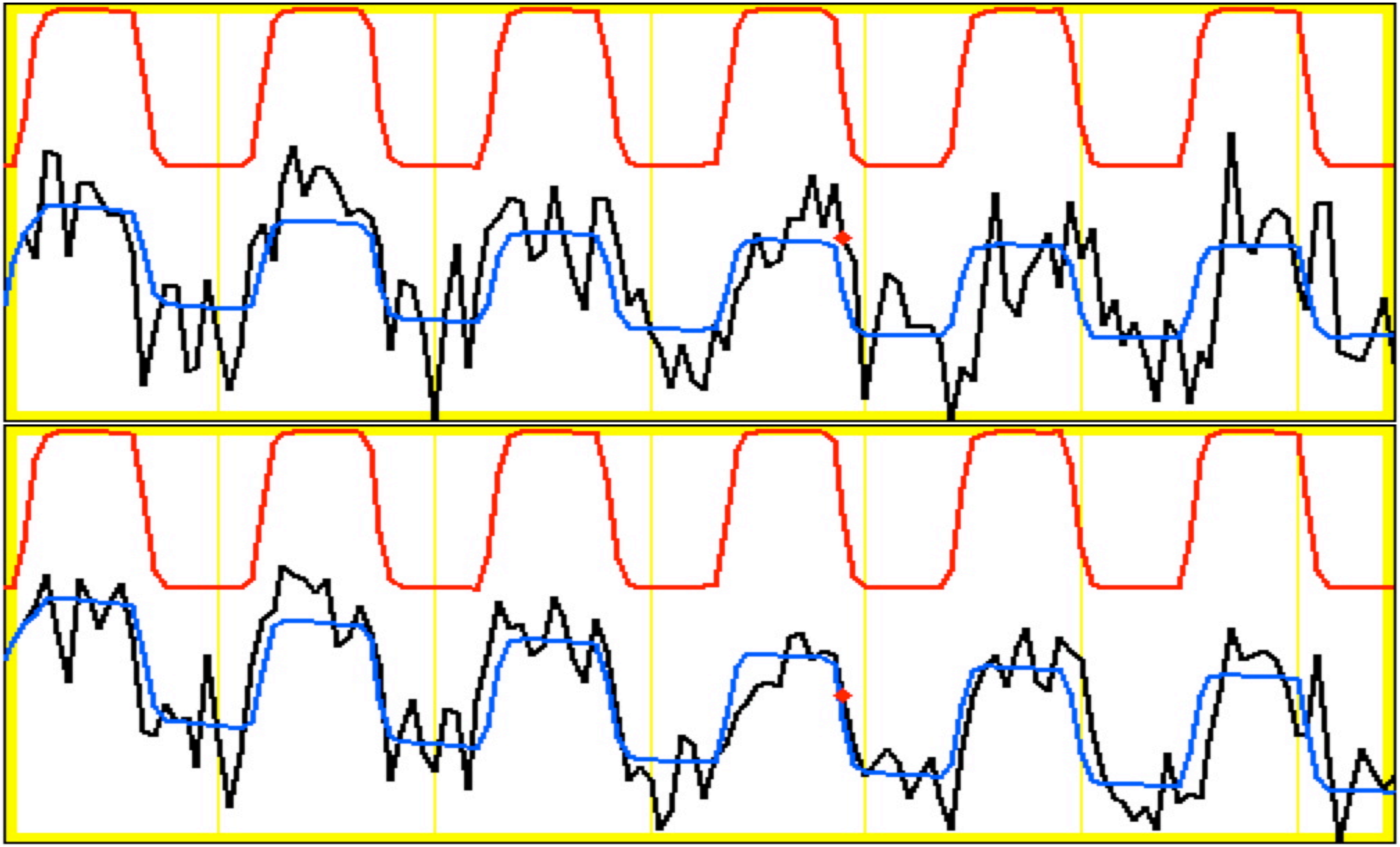
- First sample: Block-trial FMRI data
  - "Activation" occurs over a sustained period of time (say, 10 s or longer), usually from more than one stimulation event, in rapid succession
  - BOLD (hemodynamic) response accumulates from multiple close-in-time neural activations and is large
  - BOLD response is often visible in time series
  - Noise magnitude about same as BOLD response
- Next 2 slides: same brain voxel in 3 (of 9) EPI runs
  - **black curve** (noisy) = data
  - **red curve** (above data) = ideal model response
  - **blue curve** (within data) = model fitted to data
  - somatosensory task (finger being rubbed)

Same Voxel: Runs 1 and 2



Block-trials: 27 s “on” / 27 s “off”; TR=2.5 s; 130 time points/run

Same Voxel: Run 3 and Average of all 9



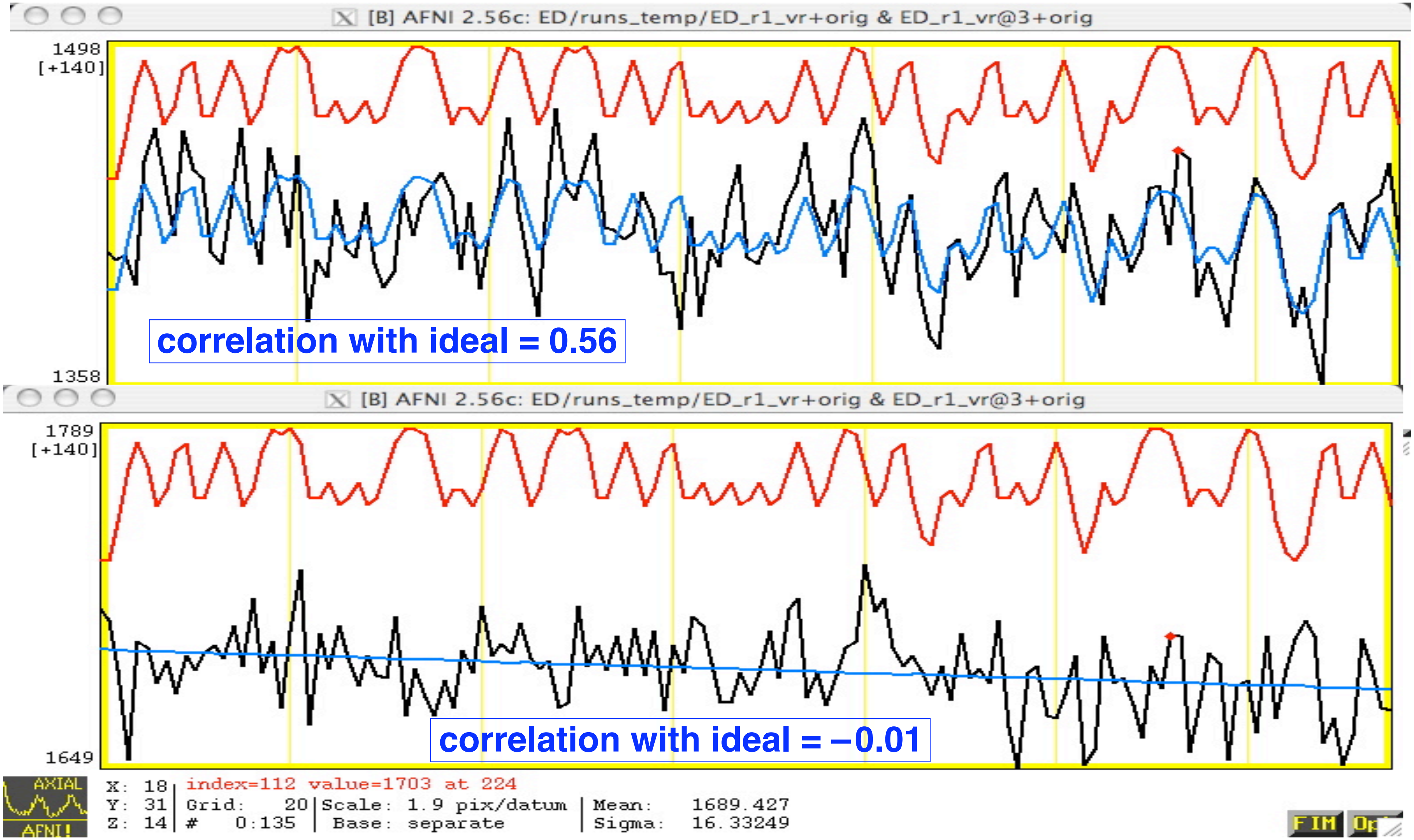
⇒ Activation amplitude & shape vary among blocks! Why???

More Sample FMRI Data Time Series

- Second sample: Event-Related FMRI
  - ➢ "Activation" occurs in single relatively brief intervals
  - ➢ "Events" can be randomly or regularly spaced in time
    - o If events are randomly spaced in time, signal model itself looks noise-like (to the pitiful human eye)
  - ➢ BOLD response to stimulus tends to be weaker, since fewer nearby-in-time "activations" have overlapping signal changes (hemodynamic responses)
- Next slide: Visual stimulation experiment

"Active" voxel shown in next slide

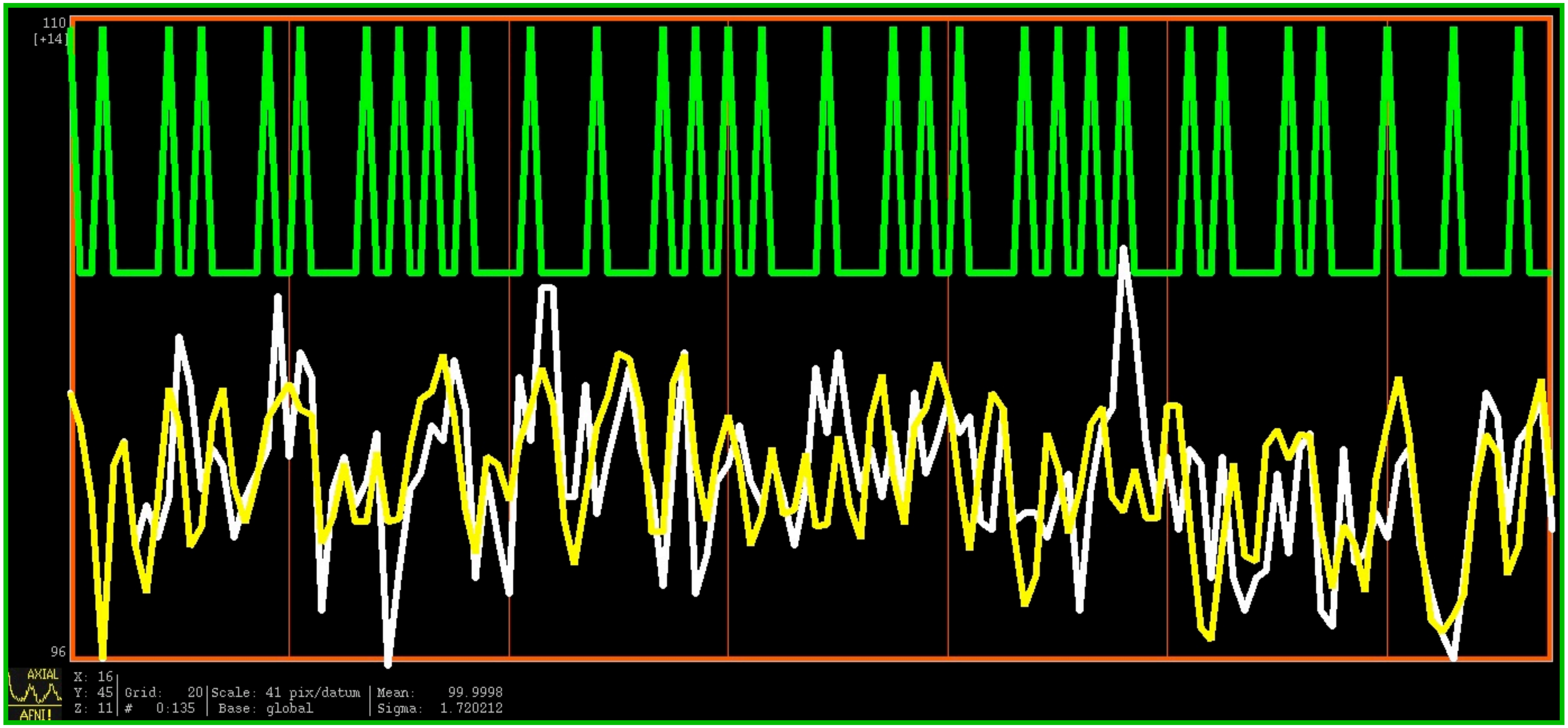
Two Voxel Time Series from Same Run



Lesson: ER-FMRI activation is not obvious via casual inspection

More Event-Related Data

Four different visual stimuli



- White curve = Data (first 136 TRs)
- Orange curve = Model fit (R^2=50%)
- Green = Stimulus timing

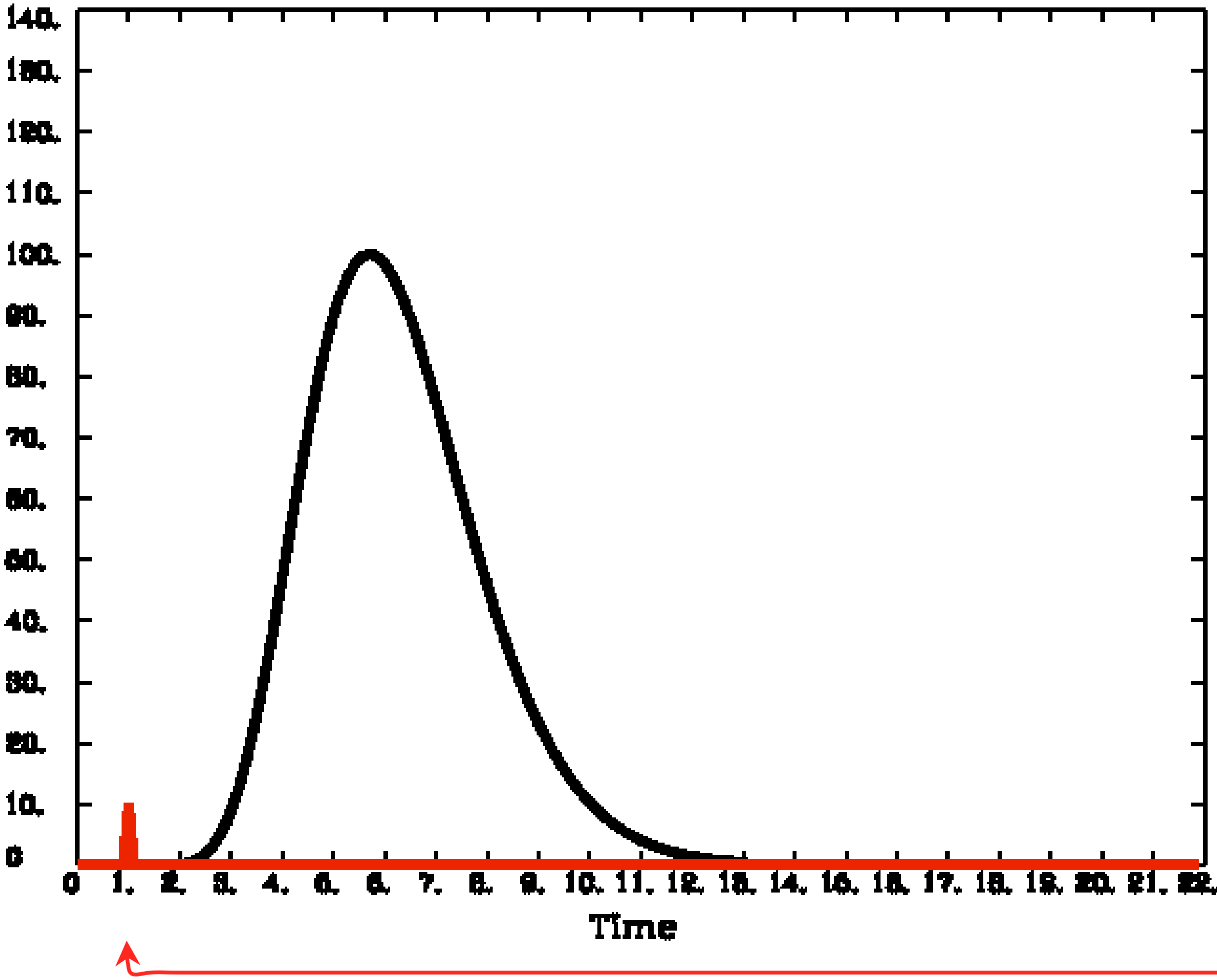
Very good fit for ER data (R^2=10-20% more usual).
Noise is as big as BOLD!

Two Fundamental Principles Underlying Most FMRI Analyses (esp. GLM): HRF ⊗ Blobs

- **H**emodynamic **R**esponse **F**unction
  - *Convolution* model for ***temporal*** relation between stimulus/activity and response
- Activation **Blobs**
  - Contiguous ***spatial*** regions whose voxel time series fit HRF model
  - *e.g.*, Reject isolated voxels even if HRF model fit is good there
  - Not the topic of these talks on time series analysis

Hemodynamic Response Function (HRF)

- **HRF** is the idealization of measurable FMRI signal change responding to a single activation cycle (up and down) from a stimulus in a voxel



Response to brief activation (< 1 s):
- delay of 1-2 s
- rise time of 4-5 s
- fall time of 4-6 s
- model equation:

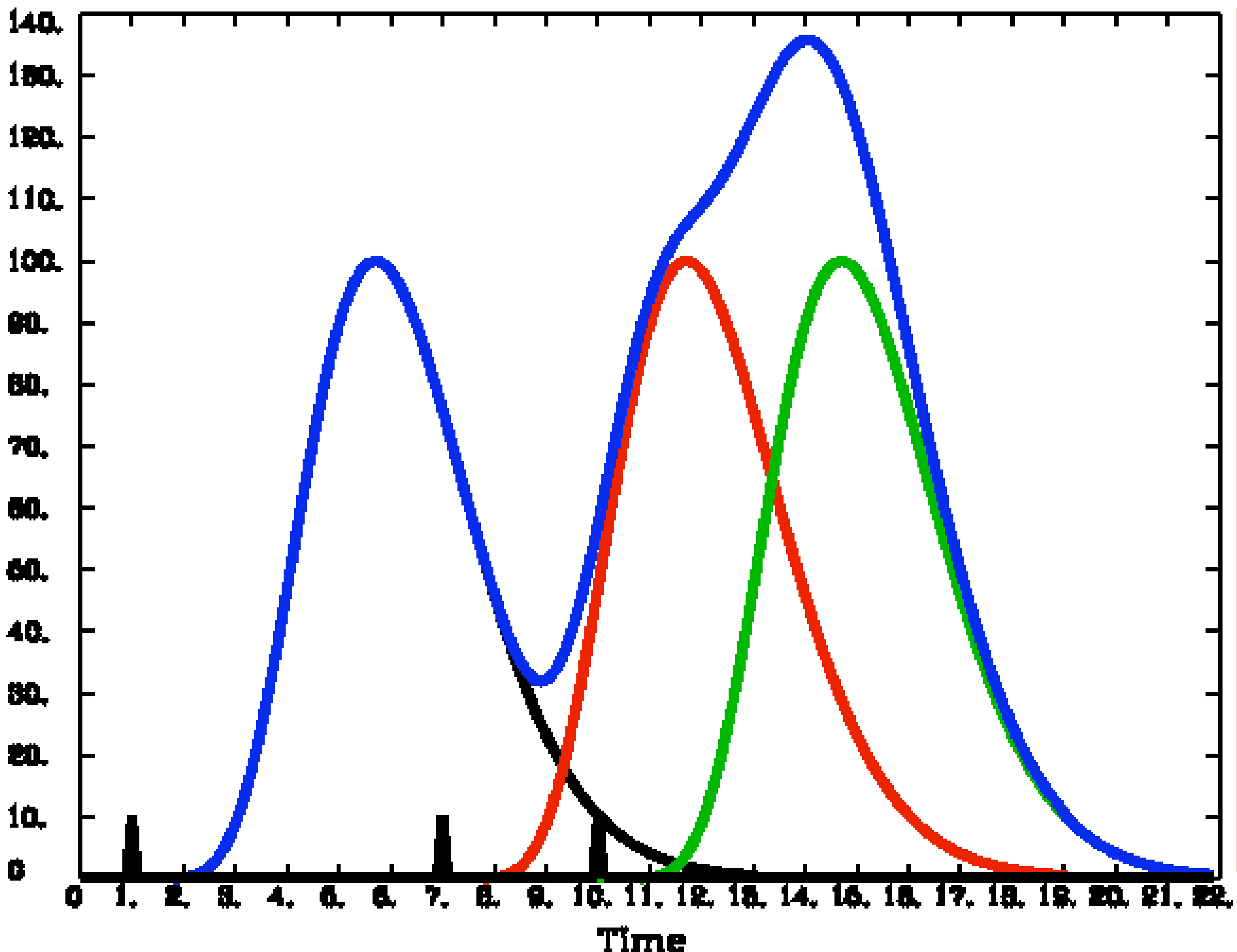
$$h(t) \propto t^b e^{-t/c}$$

- $h(t)$ is signal change t seconds ***after*** activation

1 Brief Activation (Event)

Linearity (Additivity) of HRF

- Multiple activation cycles in a voxel, closer in time than duration of HRF:
  - Assume that overlapping responses add

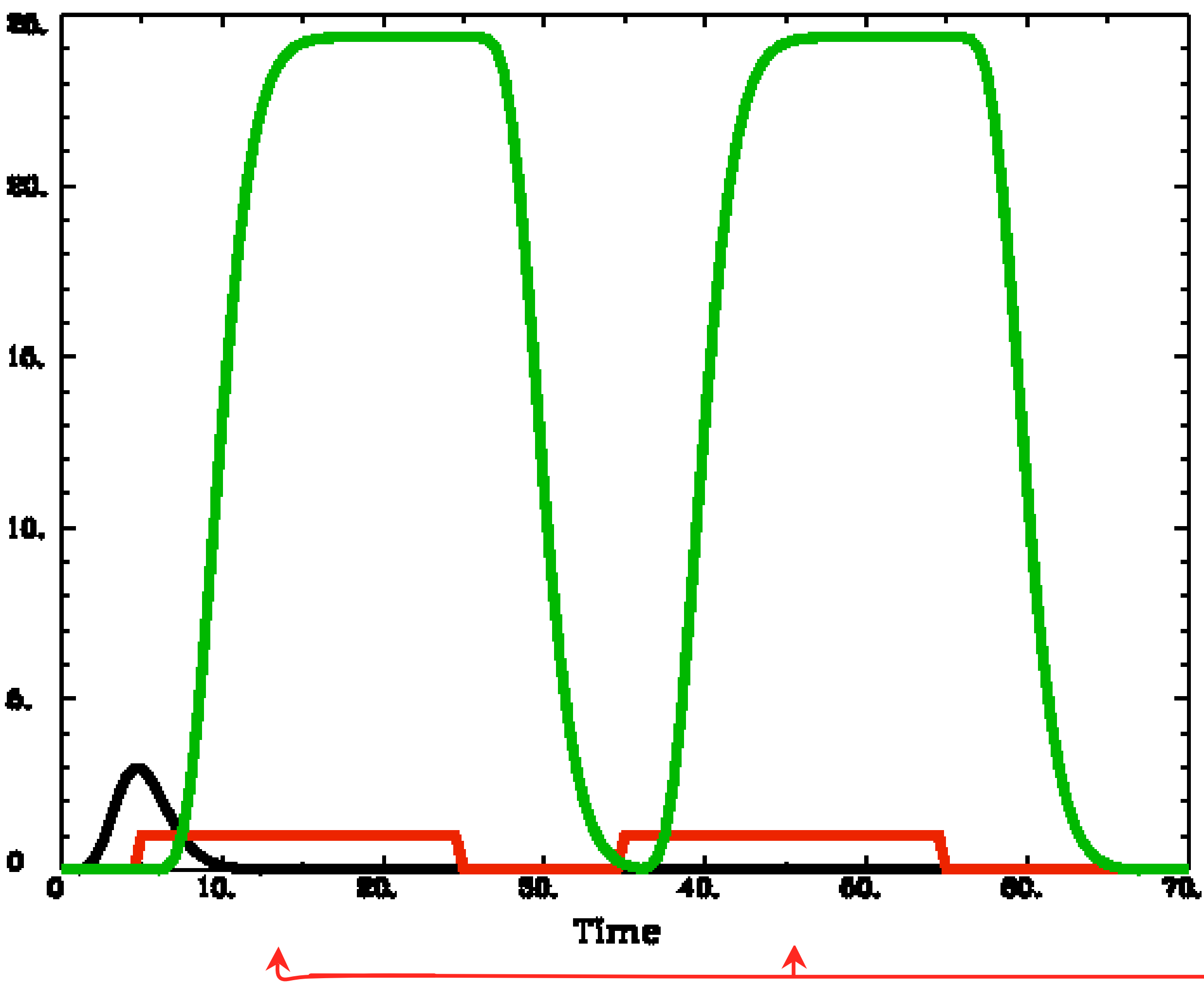


- Linearity is a pretty good assumption
- But not apparently perfect — about 90% correct
- Nevertheless, is widely taken to be true and is the basis for the "general linear model" (GLM) in FMRI analysis

3 Brief Activations

Linearity and Extended Activation

- Extended activation, as in a block-trial experiment:
  - HRF accumulates over its duration (≈ 10 s)

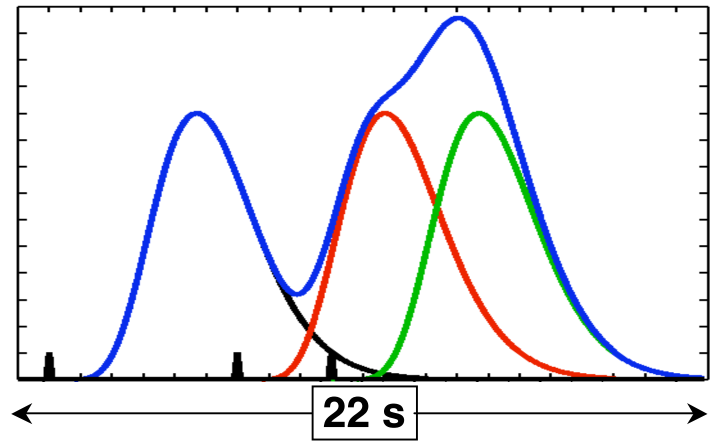


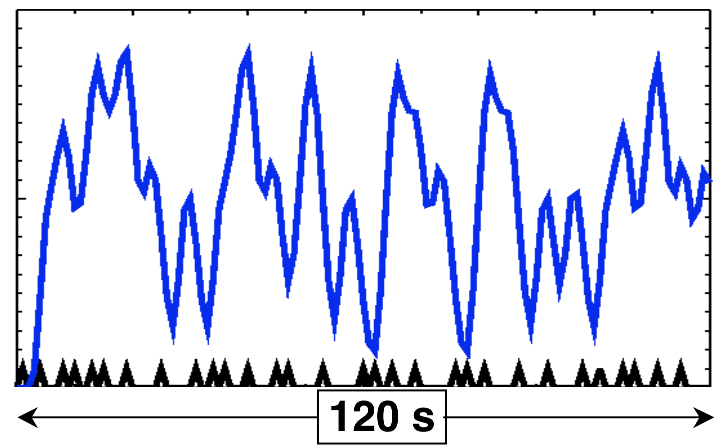
- **Black** curve = response to a single brief stimulus
- **Red** curve = activation intervals
- **Green** curve = summed up HRFs from activations
- Block-trials have larger BOLD signal changes than event-related experiments

2 Long Activations (Blocks)

Convolution Signal Model

- FMRI signal model (in each voxel) is taken as sum of the individual trial HRFs (assumed equal)
  - ➢ Stimulus timing is assumed known (or measured)
  - ➢ Resulting time series (in **blue**) are called the ***convolution*** of the HRF with stimulus timing
  - ➢ Finding HRF = "deconvolution"
  - ➢ AFNI code = **3dDeconvolve** (or its daughter **3dREMLfit**)
  - ➢ Convolution models only the FMRI signal **changes**





- Real data starts at and returns to a nonzero, slowly drifting baseline

Simple Regression Models

- Assume a fixed shape $h(t)$ for the HRF
  - e.g., $h(t) = t^{8.6} \exp(-t/0.547)$ [MS Cohen, 1997]
  - Convolve with stimulus timing to get ideal response (temporal pattern) $r(t) = \sum_{k=1}^{K} h(t - \tau_k) = \text{sum of HRF copies}$
- Assume a form for the baseline (data without activation)
  - e.g., $a + b \cdot t$ for a constant plus a linear trend
- In each voxel, fit data $Z(t)$ to a curve of the form

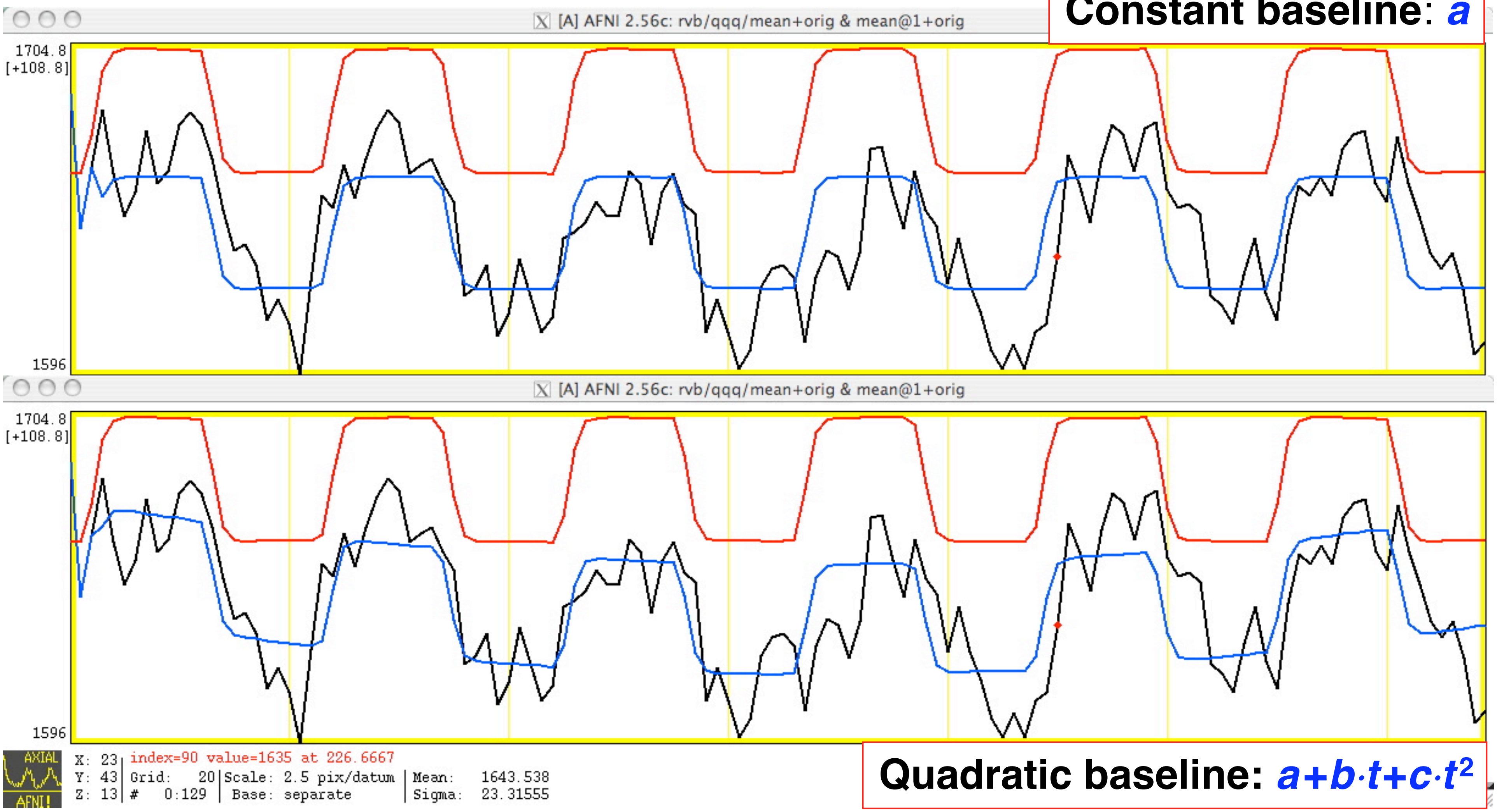
$$Z(t) \approx a + b \cdot t + \beta \cdot r(t)$$

← The signal model!

- a, b, β are unknown values to be found in each voxel
- a, b are "nuisance" parameters
- β is amplitude of $r(t)$ in data = "how much" BOLD
  - In this model, each stimulus assumed to get same BOLD response — in shape and in amplitude

Simple Regression: Sample Fits

Constant baseline: a



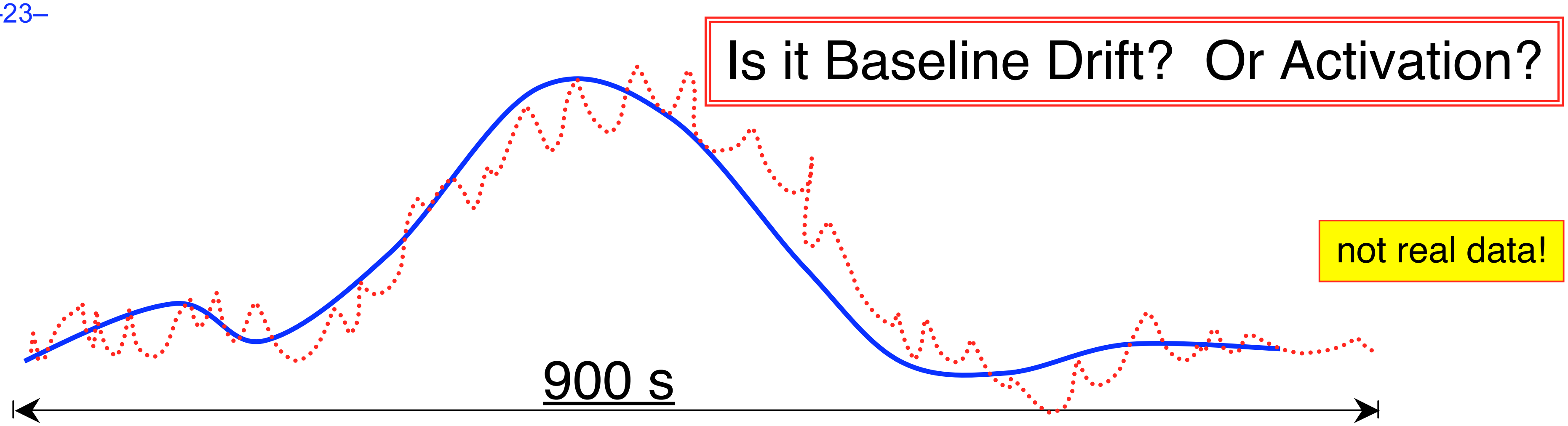
Quadratic baseline: $a+b\cdot t+c\cdot t^2$

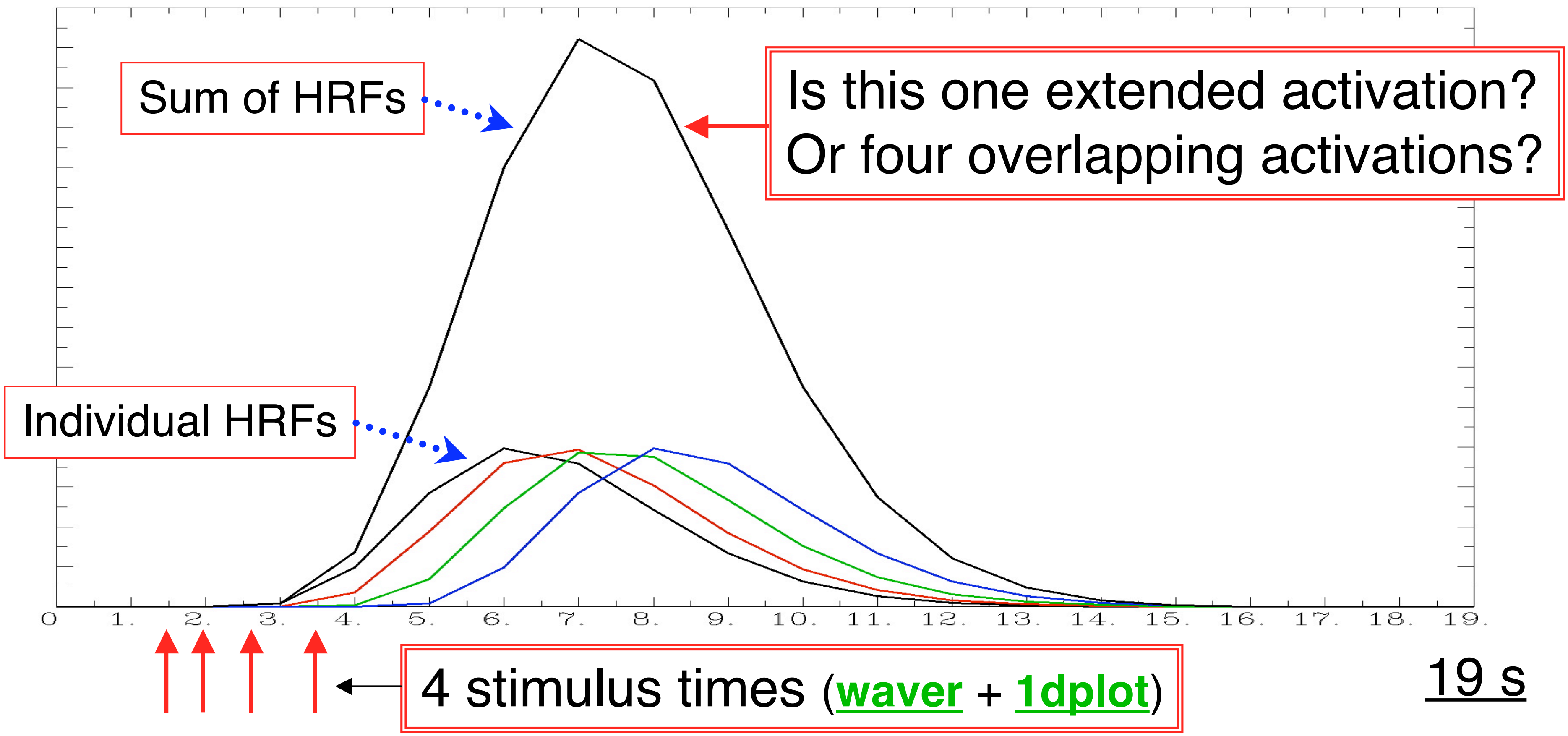
- Necessary baseline model complexity depends on duration of **continuous** imaging — e.g., 1 parameter per ~150 seconds

Duration of Stimuli - Important Caveats

- Slow baseline drift (time scale 100 s and longer) makes doing FMRI with long duration stimuli difficult
  - Learning experiment: where the task is done continuously for ~15 minutes and the subject is scanned to find parts of the brain that adapt during this time interval
  - Pharmaceutical challenge: where the subject is given some psychoactive drug whose action plays out over 10+ minutes (e.g., cocaine, ethanol)

- Multiple very short duration stimuli that are also very close in time to each other are very hard to tell apart, since their HRFs will have 90-95% overlap
  - Binocular rivalry, where percept switches ~ 0.5 s


Is it Baseline Drift? Or Activation?
not real data!
900 s


Sum of HRFs
Is this one extended activation?
Or four overlapping activations?
Individual HRFs
0 1. 2. 3. 4. 5. 6. 7. 8. 9. 10. 11. 12. 13. 14. 15. 16. 17. 18. 19.
4 stimulus times (waver + 1dplot)
19 s

Multiple Stimuli = Multiple Regressors

- Usually have more than one class of stimulus or activation in an experiment
  - e.g., want to see size of "**face activation**" vis-à-vis "**house activation**"; or, "**what**" vs. "**where**" activity
- Need to model each separate class of stimulus with a separate response function $r_1(t)$, $r_2(t)$, $r_3(t)$, ….
  - Each $r_j(t)$ is based on the stimulus timing for activity in class number j
  - Calculate a β_j amplitude = amount of $r_j(t)$ in voxel data time series $Z(t)$ = average BOLD for stim class #j
  - **Contrast** βs to see which voxels have differential activation levels under different stimulus conditions
    - e.g., statistical test on the question $\beta_1 - \beta_2 = 0$?

Multiple Stimuli - Important Caveat

- In AFNI: do ***not*** model baseline ("control") condition
  - e.g., "rest", visual fixation, high-low tone discrimination, or some other simple task
- FMRI can only measure ***changes*** in MR signal levels between tasks
  - So you need some simple-ish task to serve as a reference point
- The baseline model (e.g., $a + b \cdot t$) takes care of the signal level to which the MR signal returns when the "active" tasks are turned off
  - Modeling the reference task explicitly would be redundant (or "collinear", to anticipate a forthcoming concept)

Multiple Stimuli - Experiment Design

- How many distinct stimuli do you need in each class? Our rough recommendations:
  - Short event-related designs: at least 25 events in each stimulus class (spread across multiple imaging runs) — and more is better
  - Block designs: at least 5 blocks in each stimulus class — 10 would be better

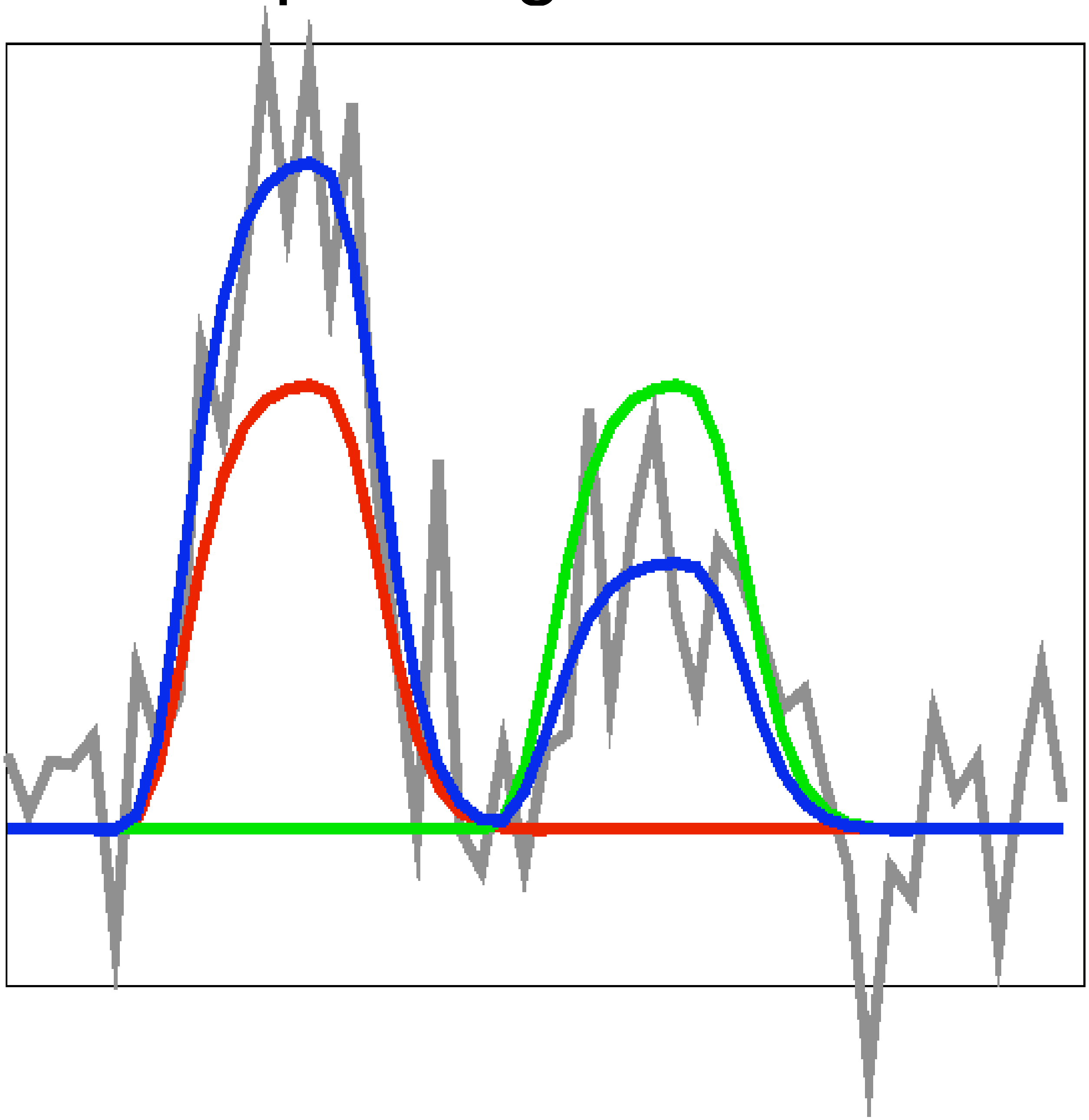
---

- While we're on the subject: **How many subjects?**
  - Several independent studies agree that 20-25 subjects in each category are needed for highly reliable results
  - This number is more than has usually been the custom in FMRI-based studies**!!**

IM Regression - an Aside

- **IM** = **I**ndividual **M**odulation
  - Compute separate amplitude of HRF for each event
    - Instead of the standard computation of the average amplitude of all responses to multiple stimuli in the same class
  - Response amplitudes (βs) for each individual block/event will be highly noisy
    - Can't use individual activation maps for much
    - Must pool the computed βs in some further statistical analysis (*t*-test via **`3dttest`**? inter-voxel correlations in the βs? correlate βs with something?)
  - Further description and examples given in the *Advanced Topics* presentation in this series (**`afni07_advanced`**)

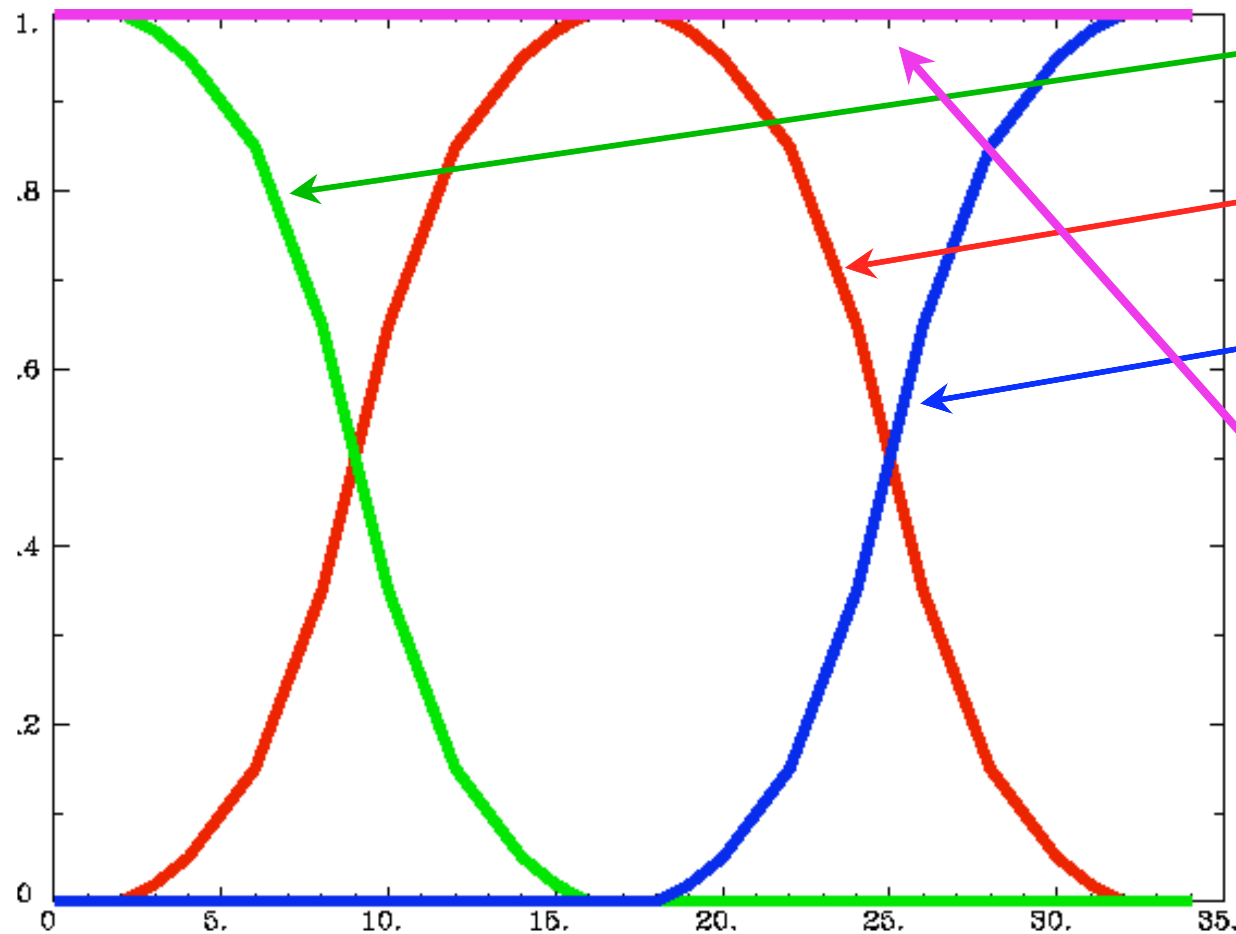
Multiple Regressors: Cartoon Animation



- **Red** curve = signal model for class #1
- **Green** curve = signal model for #2
- **Blue** curve =
  $\beta_1 \cdot \#1 + \beta_2 \cdot \#2$
  where β_1 and β_2 vary from 0.1 to 1.7 in the animation

- Goal of regression is to find β_1 and β_2 that make the blue curve best fit the data time series
- **Gray** curve =
  $1.5 \cdot \#1 + 0.6 \cdot \#2 + \text{noise}$
  = simulated data

Multiple Regressors: Collinearity!!



- **Green** curve = signal model for #1
- **Red** curve = signal model for class #2
- **Blue** curve = signal model for #3
- **Purple** curve = **#1** + **#2** + **#3** which is exactly = 1
- We cannot — ***in principle or in practice*** — distinguish sum of 3 signal models from constant baseline!!

No analysis can distinguish the cases

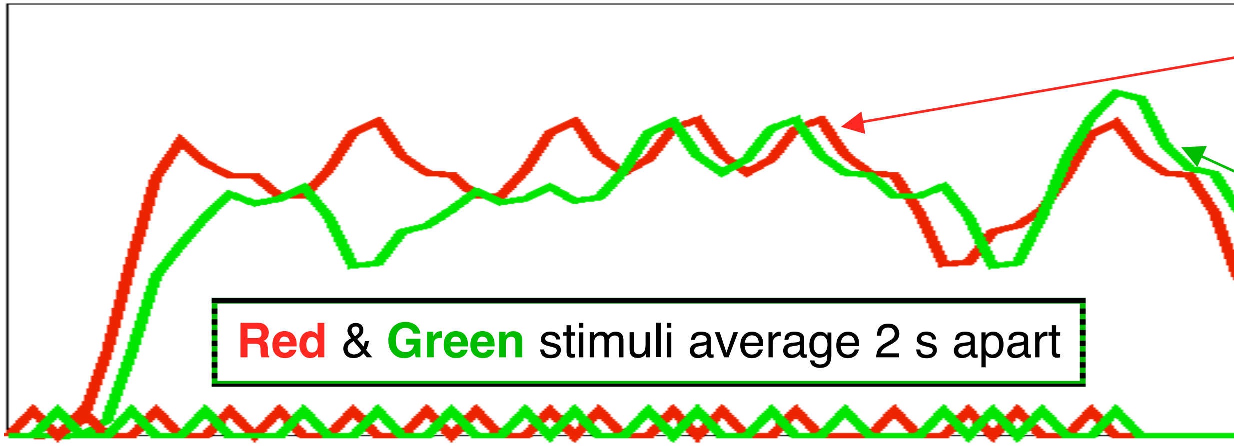
$Z(t)$=10+ 5·**#1** and

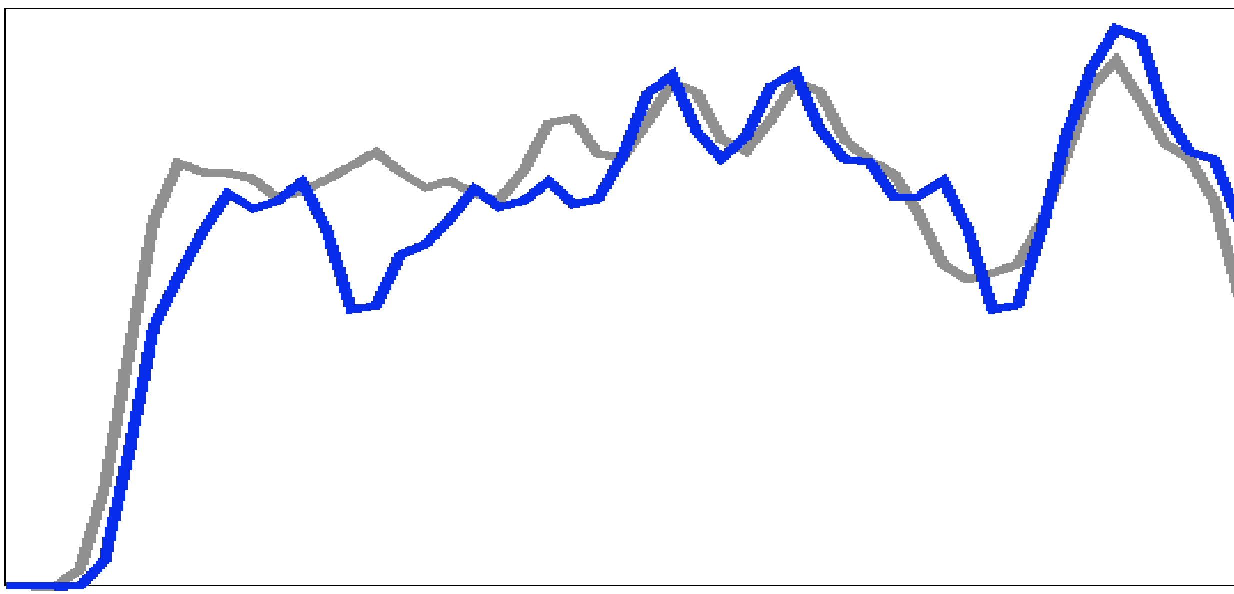
$Z(t)$= 0+15·**#1**+10·**#2**+10·**#3**

and an infinity of other possibilities

Collinear designs are **bad bad bad**!

Multiple Regressors: Near Collinearity

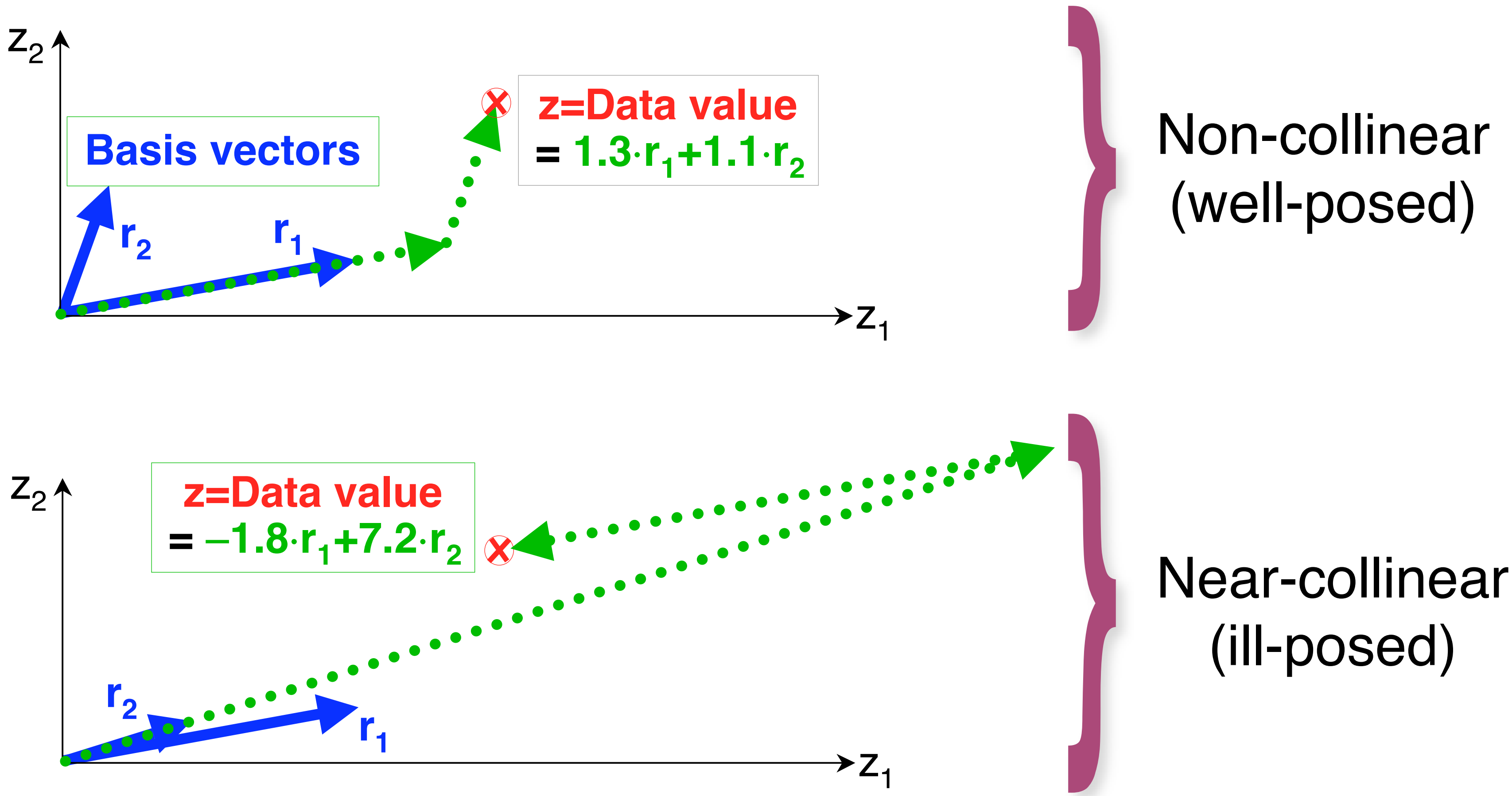




Stimuli are too close in time to distinguish response #1 from #2, considering noise

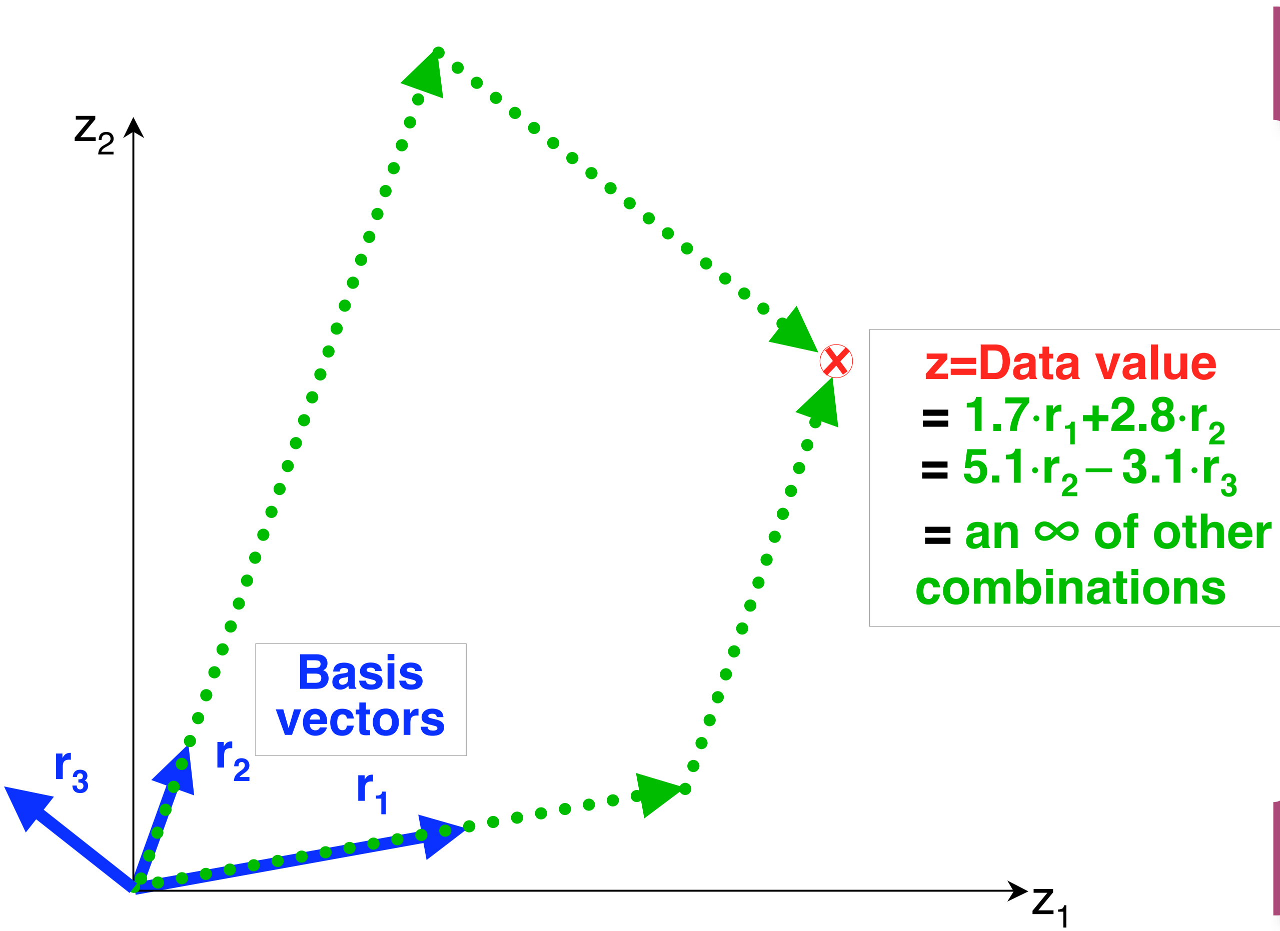
- **Red** curve = signal model for class #1
- **Green** curve = signal model for #2
- **Blue** curve = $\beta_1 \cdot \#1 + (1-\beta_1) \cdot \#2$ where β_1 varies randomly from 0.0 to 1.0 in animation
- **Gray** curve = **0.66·#1+0.33·#2** = simulated data *with no noise*
- Lots of different combinations of **#1** and **#2** are decent fits to gray curve

The Geometry of Collinearity - 1



- Trying to fit data as a sum of basis vectors that are nearly parallel doesn't work well: solutions can be huge
- Exactly parallel basis vectors would be impossible:
  - Determinant of matrix to invert would be zero

The Geometry of Collinearity - 2



Multi-collinear
= more than one solution fits the data
= over-determined

- Trying to fit data with too many regressors (basis vectors) doesn't work: no unique solution

Equations: Notation

- Will approximately follow notation of manual for the AFNI program **3dDeconvolve**
- Time: continuous in reality, but in steps in the data
  - Functions of continuous time are written like $f(t)$
  - Functions of discrete time expressed like $f(\underbrace{n \cdot TR}_{=t_n})$ where n=0,1,2,… and TR=time step
  - Usually use subscript notion f_n as shorthand
  - Collection of numbers assembled in a column is a ***vector*** and is printed in boldface:

$$\left\{\begin{matrix}\text{vector of}\\ \text{length } N\end{matrix}\right\} = \begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ \vdots \\ f_{N-1} \end{bmatrix} = \mathbf{f}$$

$$\begin{bmatrix} A_{00} & A_{01} & \cdots & A_{0,N-1} \\ A_{10} & A_{11} & \cdots & A_{1,N-1} \\ \vdots & \vdots & \ddots & \vdots \\ A_{M-1,0} & A_{M-1,1} & \cdots & A_{M-1,N-1} \end{bmatrix} = \mathbf{A} = \{M \times N \text{ matrix}\}$$

Equations: Single Response Function

- In each voxel, fit data Z_n to a curve of the form

  $$Z_n \approx a + b\cdot t_n + \beta\cdot r_n \quad \text{for } n=0,1,\ldots,N-1 \; (N=\#\text{ time pts})$$

  - a, b, β are unknown parameters to be calculated in each voxel

    $$r_n = \sum_{k=1}^{K} h(t_n - \tau_k) = \text{sum of HRF copies}$$

  - a,b are "nuisance" baseline parameters
  - β is amplitude of $r(t)$ in data = "how much" BOLD
  - Baseline model should be more complicated for long (> 150 s) continuous imaging runs:

    ≈1 param per 150 s

    - $150 < T < 300$ s: $a+b\cdot t+c\cdot t^2$
    - Longer: $a+b\cdot t+c\cdot t^2 + \lceil T/150 \rceil$ low frequency components
      - **`3dDeconvolve`** actually uses Legendre polynomials for baseline
      - Using p^{th} order polynomial analogous to a lowpass cutoff $\approx (p-2)/T$ Hz
    - Often, also include as extra baseline components the estimated subject head movement time series, in order to remove residual contamination from such artifacts (will see example of this later)

Equations: Multiple Response Functions

- In each voxel, fit data Z_n to a curve of the form

$$Z_n \approx [\text{baseline}]_n + \beta_1 \cdot r_n^{(1)} + \beta_2 \cdot r_n^{(2)} + \beta_3 \cdot r_n^{(3)} + \cdots$$

  - β_j is amplitude in data of $r_n^{(j)} = r_j(t_n)$; i.e., "how much" of the j th response function is in the data time series
  - In simple regression, each $r_j(t)$ is derived directly from stimulus timing ***and*** user-chosen HRF model
    - In terms of stimulus times:

$$r_n^{(j)} = \sum_{k=1}^{K_j} h_j(t_n - \tau_k^{(j)}) = \text{sum of HRF copies}$$

    - Where $\tau_k^{(j)}$ is the k^{th} stimulus time in the j^{th} stimulus class
    - These times are input using the **`-stim_times`** option to program **`3dDeconvolve`**

Equations: Matrix-Vector Form

- Express ***known*** data vector as a sum of ***known*** columns with ***unknown*** coefficents:

- Const baseline
- Linear trend
- Response to stim#1
- Response to stim#2

$$\begin{bmatrix} z_0 \\ z_1 \\ z_2 \\ \vdots \\ z_{N-1} \end{bmatrix} \approx \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \cdot a + \begin{bmatrix} 0 \\ 1 \\ 2 \\ \vdots \\ N-1 \end{bmatrix} \cdot b + \begin{bmatrix} r_0^{(1)} \\ r_1^{(1)} \\ r_2^{(1)} \\ \vdots \\ r_{N-1}^{(1)} \end{bmatrix} \cdot \beta_1 + \begin{bmatrix} r_0^{(2)} \\ r_1^{(2)} \\ r_2^{(2)} \\ \vdots \\ r_{N-1}^{(2)} \end{bmatrix} \cdot \beta_2 + \cdots$$

or

$$\begin{bmatrix} z_0 \\ z_1 \\ z_2 \\ \vdots \\ z_{N-1} \end{bmatrix} \approx \begin{bmatrix} 1 & 0 & r_0^{(1)} & r_0^{(2)} & \cdots \\ 1 & 1 & r_1^{(1)} & r_1^{(2)} & \cdots \\ 1 & 2 & r_2^{(1)} & r_2^{(2)} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \\ 1 & N-1 & r_{N-1}^{(1)} & r_{N-1}^{(2)} & \cdots \end{bmatrix} \begin{bmatrix} a \\ b \\ \beta_1 \\ \beta_2 \\ \vdots \end{bmatrix}$$

'≈' means "least squares"

or

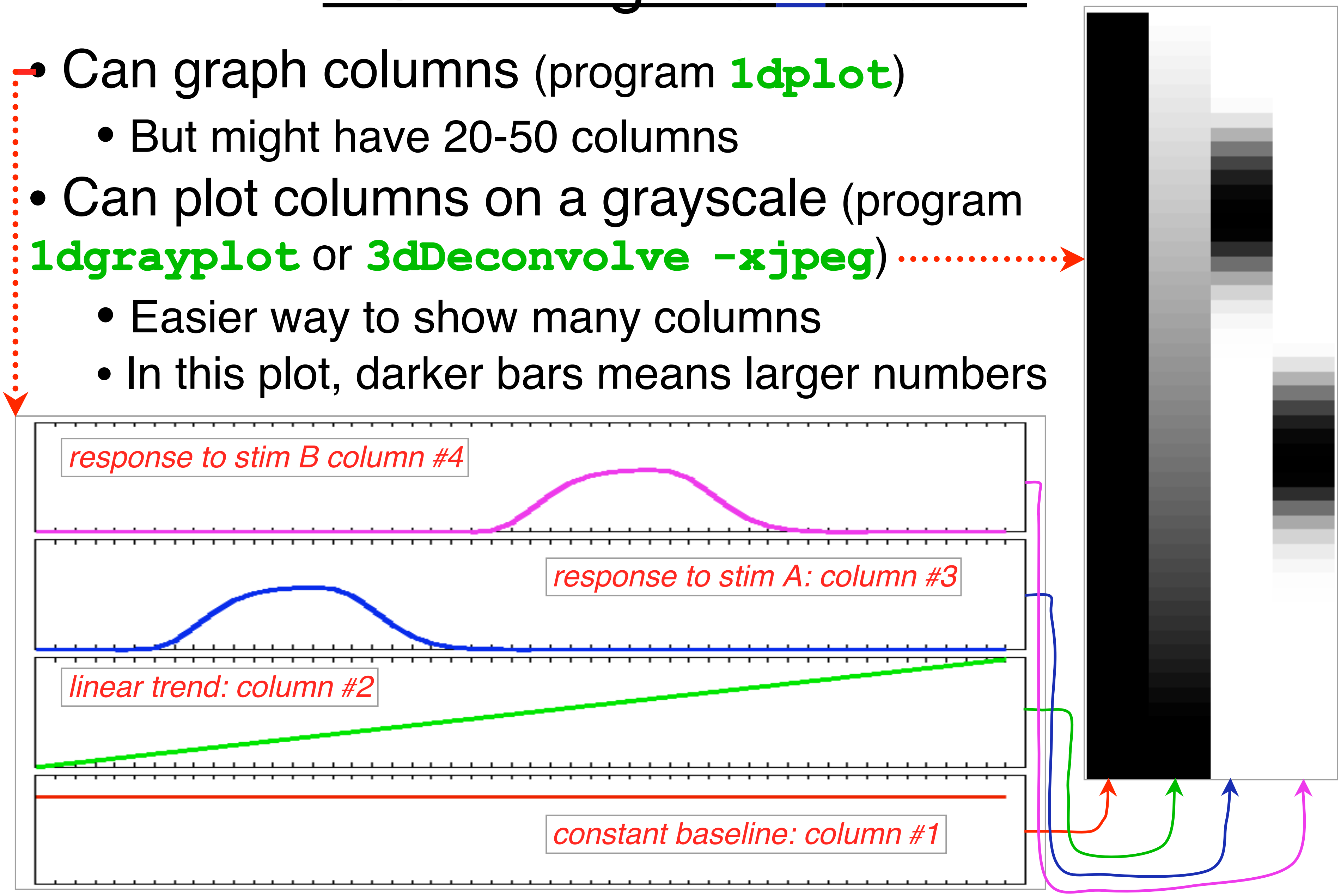
$$\underbrace{\mathbf{z}}_{\text{vector of data}} \approx \underbrace{\mathbf{R}}_{\text{matrix of columns}} \underbrace{\boldsymbol{\beta}}_{\text{vector of coeff}}$$

the "design" matrix; AKA $\mathbf{X}$

$\mathbf{z}$ depends on the voxel; $\mathbf{R}$ doesn't

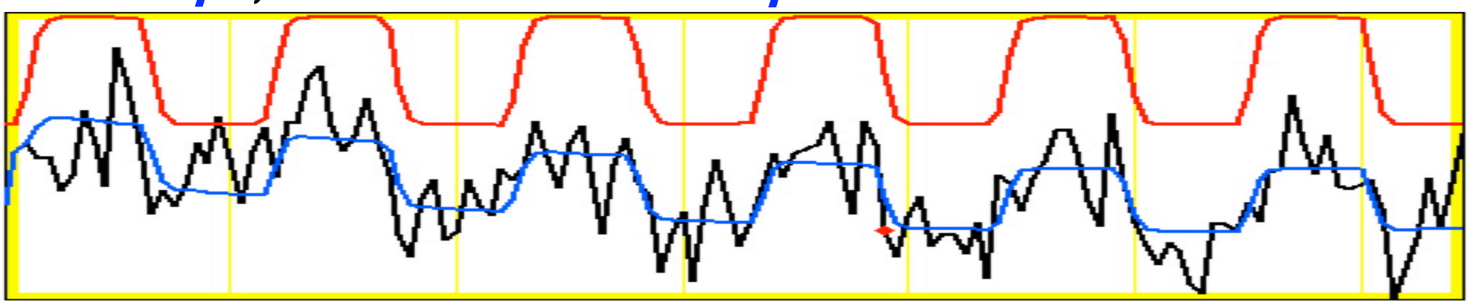
Visualizing the **R** Matrix

- Can graph columns (program **`1dplot`**)
  - But might have 20-50 columns
- Can plot columns on a grayscale (program **`1dgrayplot`** or **`3dDeconvolve -xjpeg`**)
  - Easier way to show many columns
  - In this plot, darker bars means larger numbers



Solving $\mathbf{z} \approx \mathbf{R}\beta$ for β

- Number of equations = number of time points
  - ★ 100s per run, but perhaps 1000s per subject
- Number of unknowns usually in range 5–50
- Least squares solution: $\hat{\beta} = \left[\mathbf{R}^T\mathbf{R}\right]^{-1}\mathbf{R}^T\mathbf{z}$
  - ➢ $\hat{\beta}$ denotes an *estimate* of the true (unknown) β
  - ➢ From $\hat{\beta}$, calculate $\hat{\mathbf{z}} = \mathbf{R}\hat{\beta}$ as the *fitted model*



    - o $\mathbf{z} - \hat{\mathbf{z}}$ is the ***residual time series*** = noise (we hope)
    - o Statistics measure how much each regressor helps reduce residuals
- Collinearity: when matrix $\mathbf{R}^T\mathbf{R}$ can't be inverted
  - ➢ Near collinearity: when inverse exists but is huge

Simple Regression: Recapitulation

- Choose HRF model $h(t)$ [AKA *fixed-model regression*]
- Build model responses $r_n(t)$ to each stimulus class
  - Using $h(t)$ and the stimulus timing
- Choose baseline model time series
  - Constant + linear + quadratic (+ movement?)
- Assemble model and baseline time series into the columns of the $\mathbf{R}$ matrix
- For each voxel time series $\mathbf{z}$, solve $\mathbf{z} \approx \mathbf{R}\beta$ for $\hat{\beta}$
- **Individual subject maps**: Test the coefficients in $\hat{\beta}$ that you care about for statistical significance
- **Group maps**: Transform the coefficients in $\hat{\beta}$ that you care about to Talairach/MNI space, and perform statistics on the collection of $\hat{\beta}$ values across subjects