Bayesian evidence for visualizing model selection uncertainty

Gordon L. Kindlmann
glk@uchicago.edu

THE UNIVERSITY OF CHICAGO
COMPUTER SCIENCE

Computation Institute

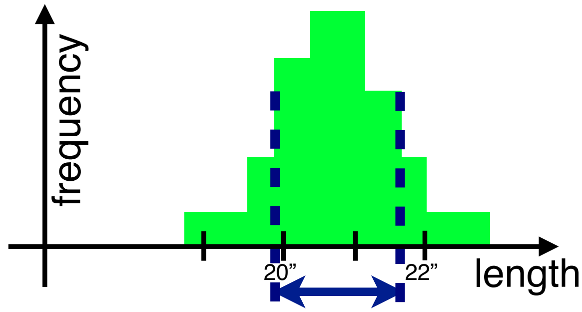
Basic message

- There many different kinds of uncertainty
  (Rheingans “Ways of not knowing”, Monday)
- There’s at least one more kind of uncertainty
  Based on an **empirical** view of data & visualization
  (not sure if relevant for simulation studies?)
- Goal here is to:
  Define this kind of uncertainty
  Argue for its relevance
  Get feedback from you

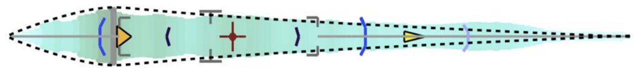
Work initiated by Maxwell Shron (now data scientist for OkCupid); collaboration with Thomas Schultz

One flavor of uncertainty

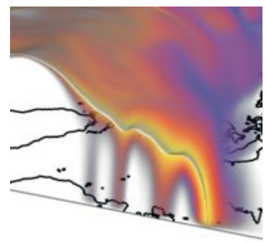
Variance in measurement of scalar (e.g. length)



uncertainty in length due to the measurement process itself



Potter et al. "Visualizing Summary Statistics and Uncertainty". Eurovis 2010 pp. 823–832

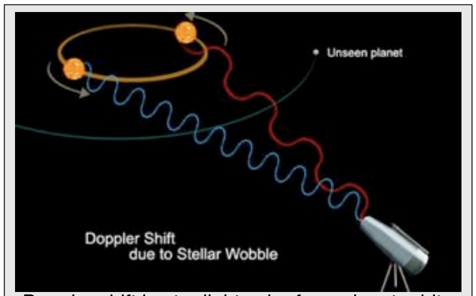


Pöthkow et al. "Probabilistic Marching Cubes". Eurovis 2011 pp. 931-940

Second flavor of uncertainty

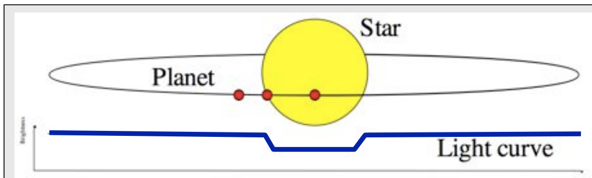
Variance in parameters of **model** of data

Model: mathematical representation of a hypothesis, parameterized by physically meaningful degrees-of-freedom, to predict measurements



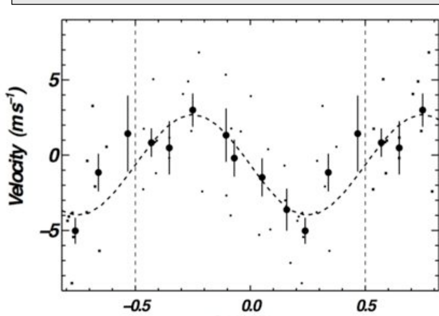
Doppler shift in star light color from planet orbit

Finding planets around other stars

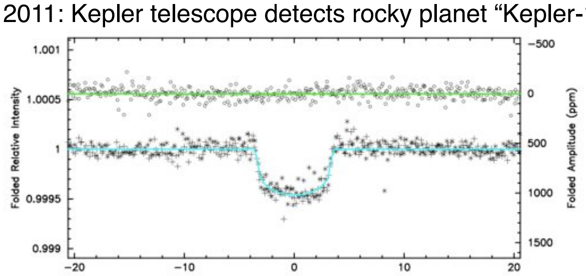


Decrease in star light intensity due to planet transit

http://en.wikibooks.org/wiki/General_Astronomy/Extrasolar_Planets



Jan 2011: Kepler telescope detects rocky planet “Kepler-10b”



http://www.nasa.gov/pdf/509370main_Batalha_N_Kepler-10b.pdf

Second flavor of uncertainty

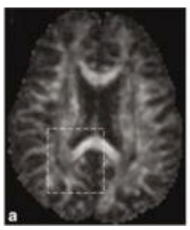
Variance in parameters of **model** of data

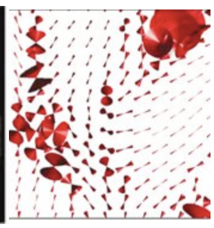
...

otherwise Sun-like Main Sequence star with $T_{\rm eff} = 5627 \pm 44$ K, $M_\star = 0.895 \pm 0.060$ $M_\odot$, $R_\star = 1.056 \pm 0.021$ $R_\odot$. Physical models simultaneously fit to the transit light curves and the precision Doppler measurements yielded tight constraints on the properties of Kepler-10b that speak to its rocky composition: $M_{\rm P} = 4.56^{+1.17}_{-1.29}$ $M_\oplus$, $R_{\rm P} = 1.416^{+0.033}_{-0.036}$ $R_\oplus$, and $\rho_{\rm P} = 8.8^{+2.1}_{-2.9}$ g cm^{-3}. Kepler-10b is the smallest transiting exoplanet discovered to date.

http://www.nasa.gov/pdf/509370main_Batalha_N_Kepler-10b.pdf

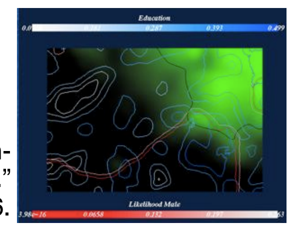
=> **Parameters** of **models** of empirical data are fundamental quantities in the scientific method





DK Jones, “Determining and visualizing uncertainty in estimates of fiber orientation from diffusion tensor MRI” Magnetic Resonance in Medicine 49:7-12 (2003)

Rheingans & desJardins, “Visualizing High-Dimensional Predictive Model Quality.” Visualization 2000, pp, 493-496.



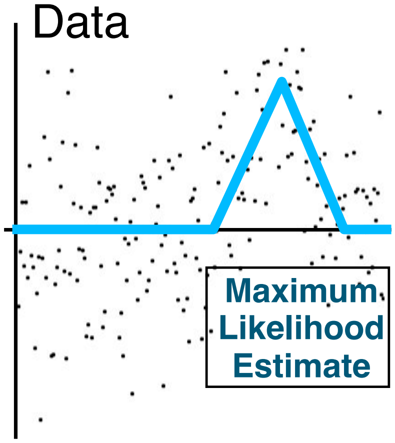
Second flavor of uncertainty

Variance in parameters of **model** of data

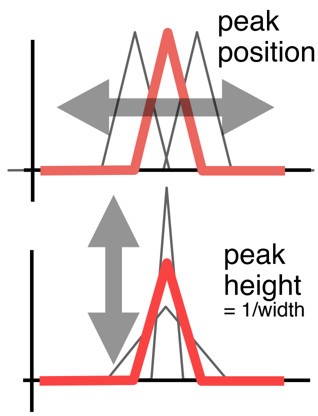
Synthetic data and low-DOF model

Know something about noise => compute likelihood **L** of the data given particular model parameters

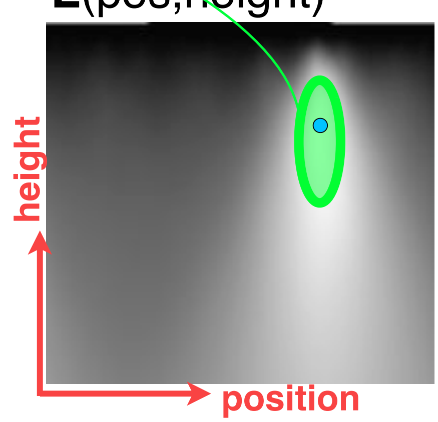
Data



Model (2 DOF)

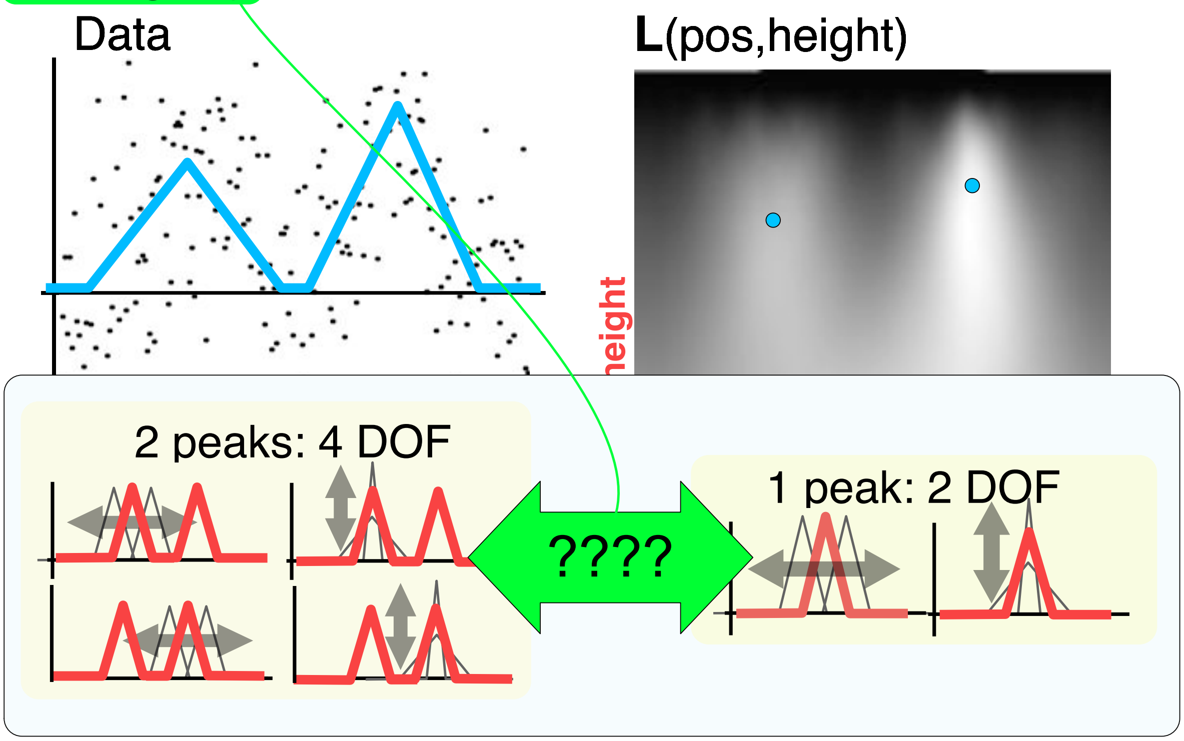


L(pos,height)



Proposed new flavor of uncertainty

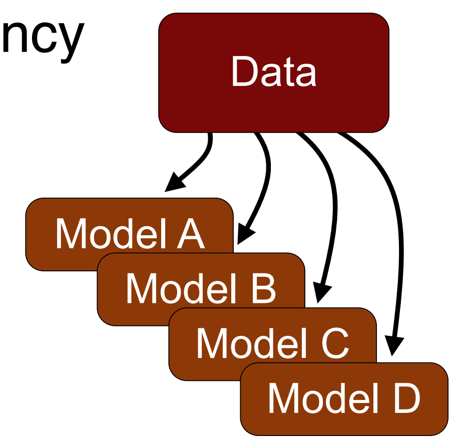
Ambiguity in **choice of model** to describe data



Goals of visualizing model ambiguity

1) Understand quality/sufficiency of set of models in novel and complex imaging modalities

- Modern imaging produces multiple values per-voxel
- Discover spatial/anatomic structure of where models are descriptive, and where not

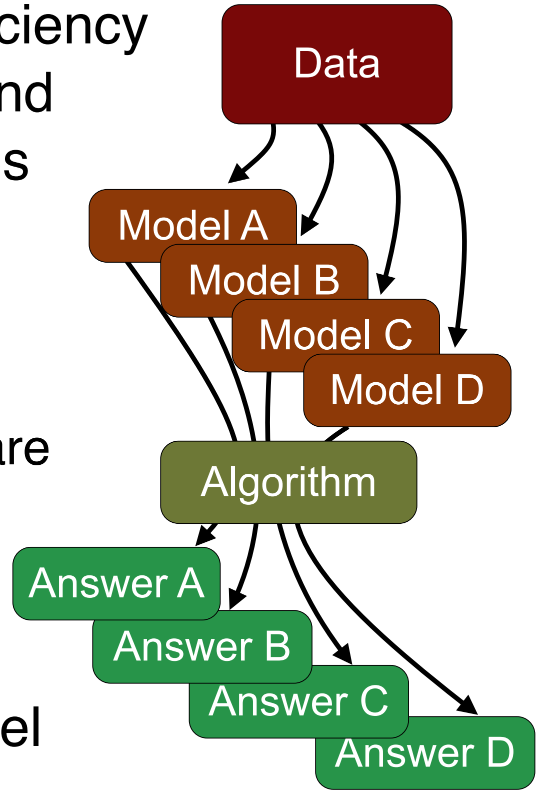


Goals of visualizing model ambiguity

1) Understand quality/sufficiency of set of models in novel and complex imaging modalities

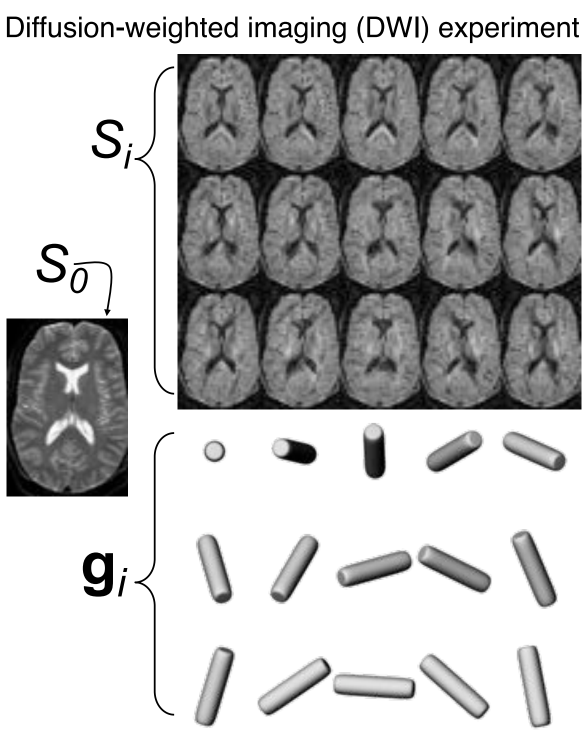
- Modern imaging produces multiple values per-voxel
- Discover spatial/anatomic structure of where models are descriptive, and where not

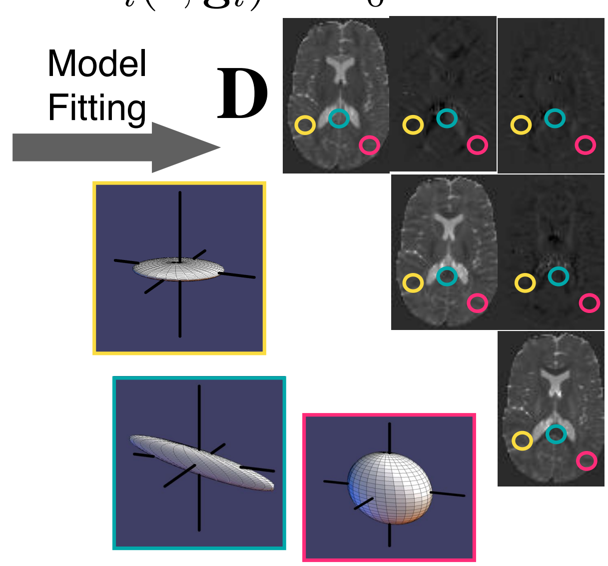
2) Characterize stability of visualization/analysis with respect to changes in model



Diffusion Tensors from DWI data

Single Tensor Model (Basser et al. 1994) : $S_i(b, \mathbf{g}_i) = S_0 e^{-b\mathbf{g}_i \cdot \mathbf{D}\mathbf{g}_i}$



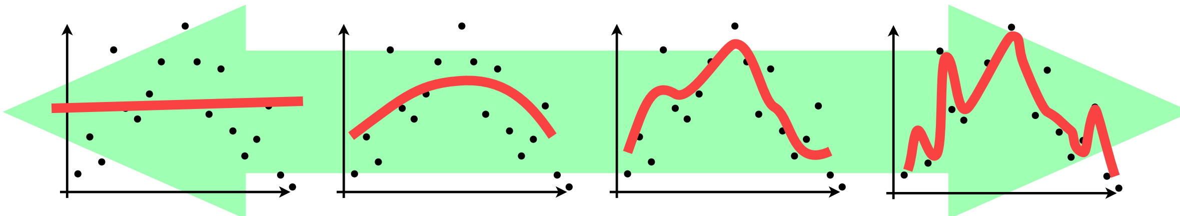


Scalar invariants of tensor:
- Bulk mean diffusivity “MD”
- Fractional Anisotropy “FA”

How to quantify model quality?

Can use error (residual) in fit

- More DOF => better fit, but less explanatory



- Various schemes for penalizing high DOF

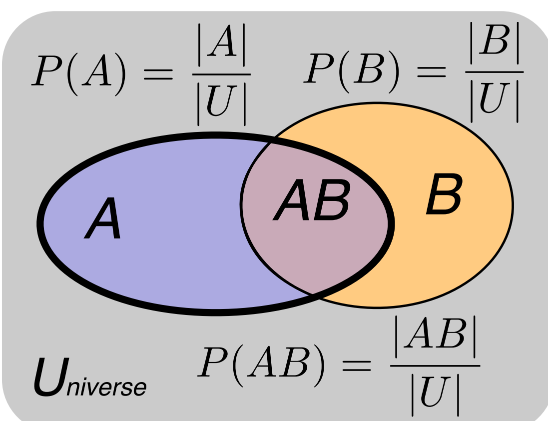
One method: **Bayesian Model Inference**

Naturally implements Occam's Razor

Cleanly includes probabilistic noise model

Bayes Visualized

http://oscarbonilla.com/2009/05/visualizing-bayes-theorem/



$$P(A) = \frac{|A|}{|U|} \qquad P(B) = \frac{|B|}{|U|} \qquad P(AB) = \frac{|AB|}{|U|}$$

$$\boxed{P(A|B) = \frac{P(B|A)P(A)}{P(B)}}$$

$$\Uparrow$$

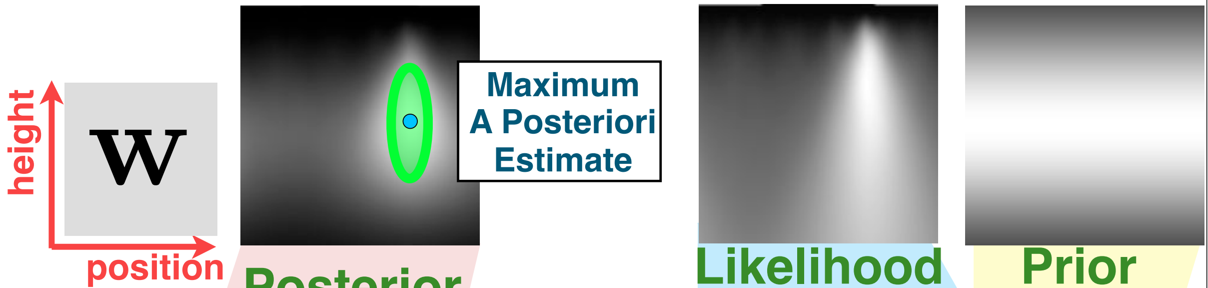
$$P(A|B)P(B) = P(B|A)P(A)$$

$$P(B|A) = \frac{|AB|}{|A|} = \frac{|AB|/|U|}{|A|/|U|} = \frac{P(AB)}{P(A)}$$

$$P(A|B) = \frac{|AB|}{|B|} = \frac{|AB|/|U|}{|B|/|U|} = \frac{P(AB)}{P(B)}$$

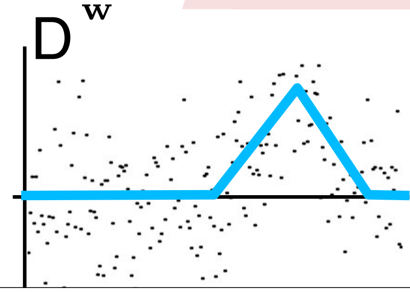
Bayes for fitting (one, fixed) model

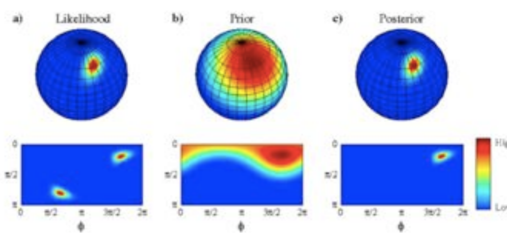
Find parameters **w** of model M_i that maximize posterior probability of the measured data D



Posterior — Likelihood — Prior

$$\underset{\mathbf{w}}{\arg\max}\, \underbrace{P(\mathbf{w}|D,\mathcal{M}_i)}_{\text{Posterior}} = \underset{\mathbf{w}}{\arg\max}\, \frac{\overbrace{P(D|\mathbf{w},\mathcal{M}_i)}^{\text{Likelihood}}\overbrace{P(\mathbf{w}|\mathcal{M}_i)}^{\text{Prior}}}{P(D|\mathcal{M}_i)}$$





Friman et al. "A Bayesian approach for stochastic white matter tractography" IEEE TMI 25:965-978 (2006)

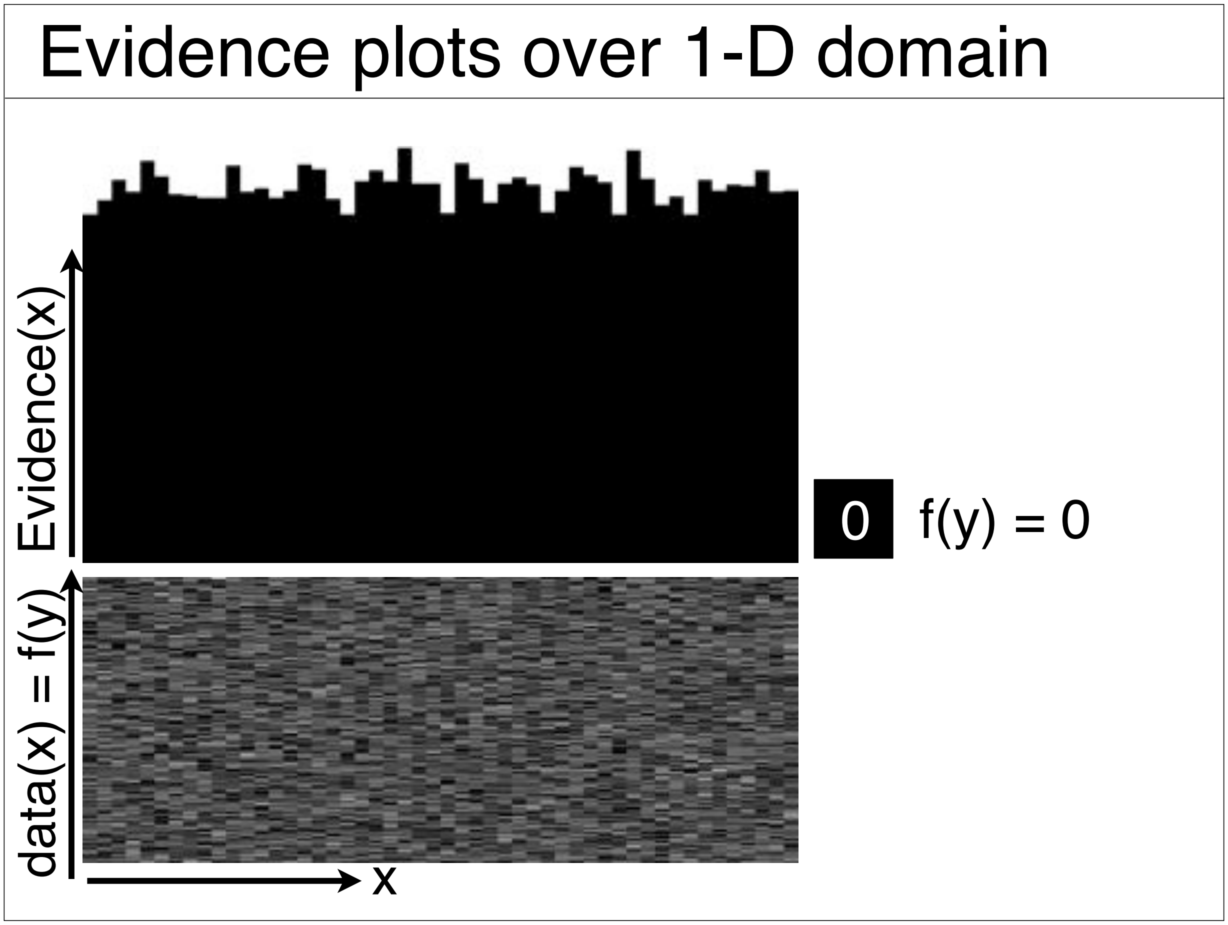
Bayes for evaluating a model

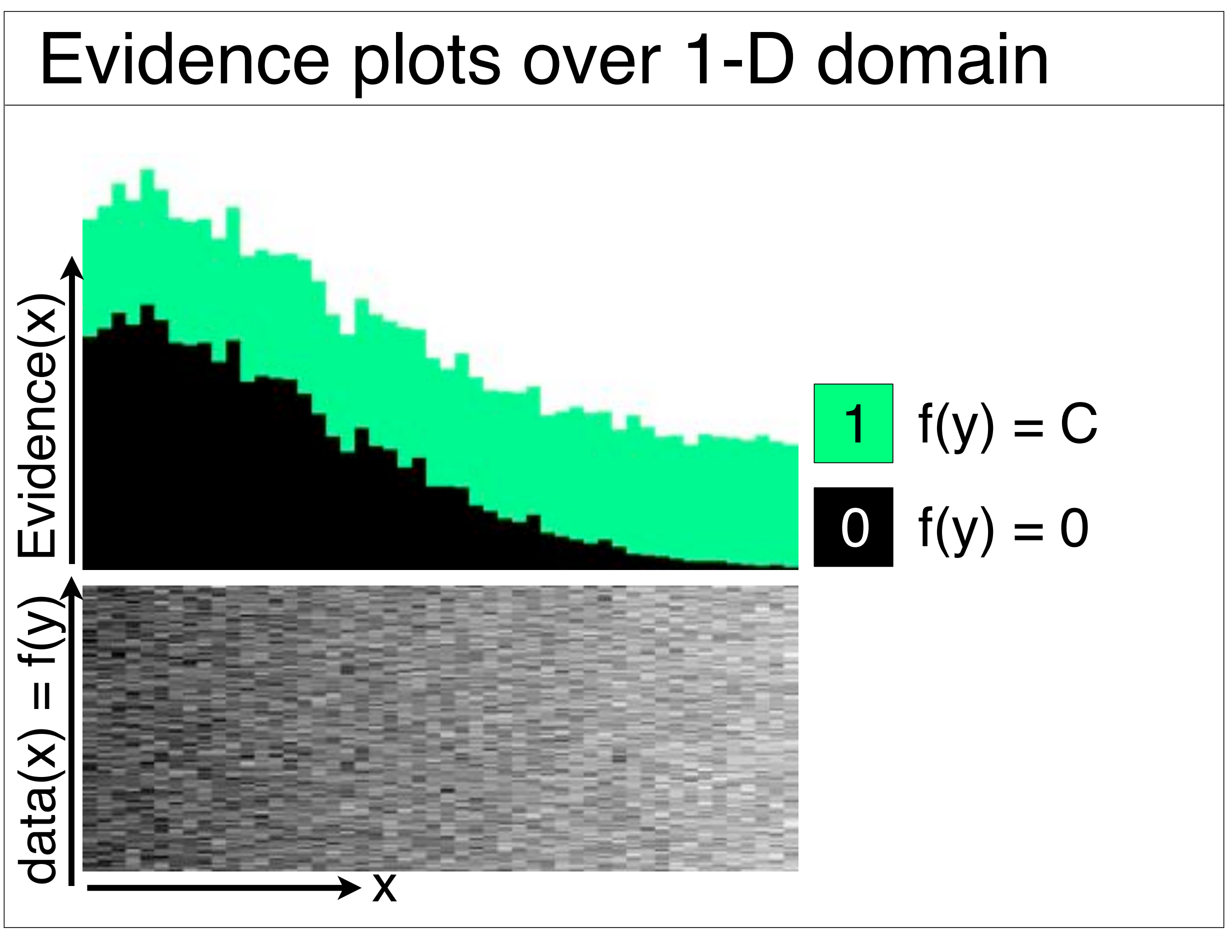
Want to quantify: how plausible is model M_i, as a whole, given the data?

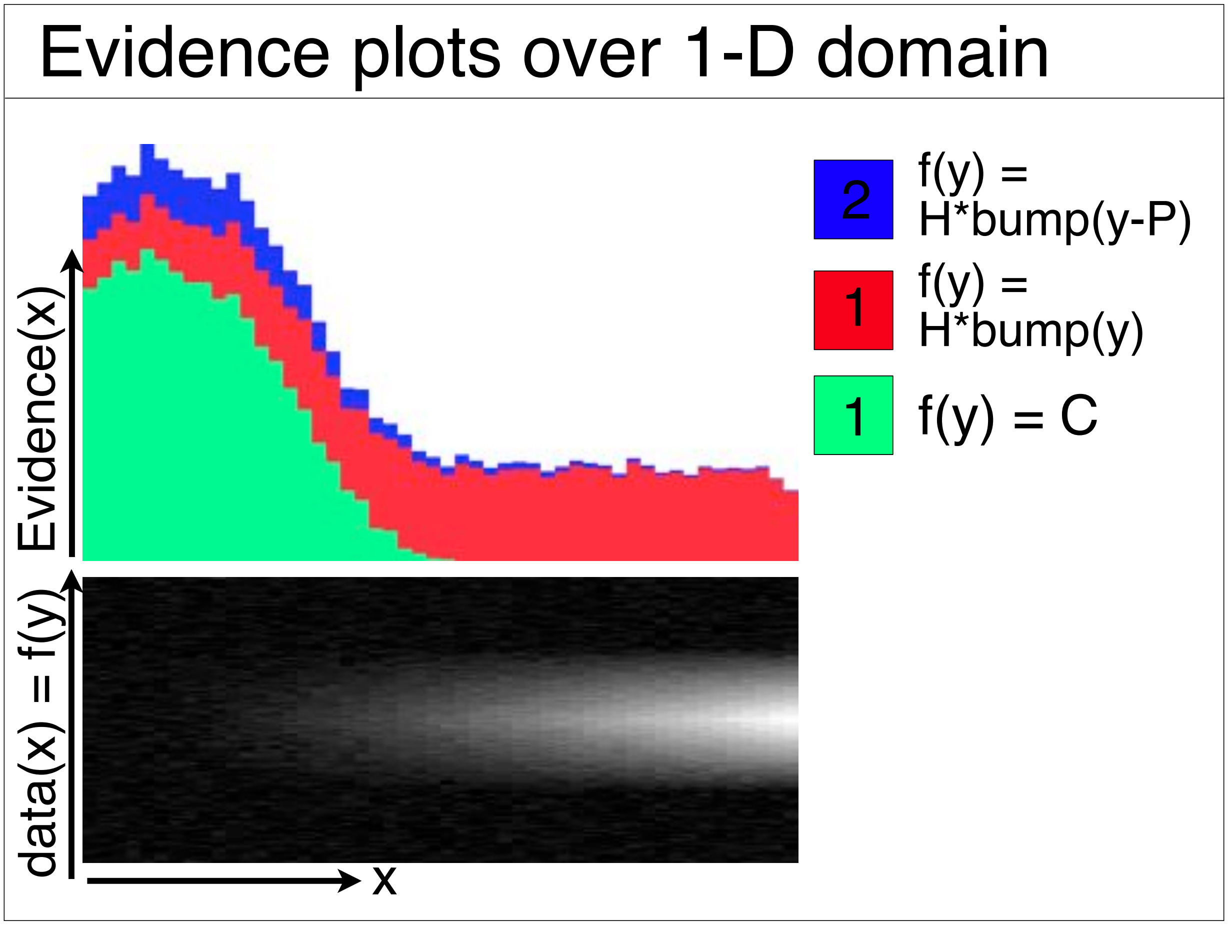
$$P(\mathcal{M}_i|D) = \frac{\overbrace{P(D|\mathcal{M}_i)}^{\text{Evidence}}\overbrace{\cancel{P(\mathcal{M}_i)}}^{\text{Prior}\ \to\ \text{constant}}}{P(D)} = \frac{P(D|\mathcal{M}_i)P(\mathcal{M}_i)}{\boxed{\sum_i P(D|\mathcal{M}_i)}P(\mathcal{M}_i)}$$

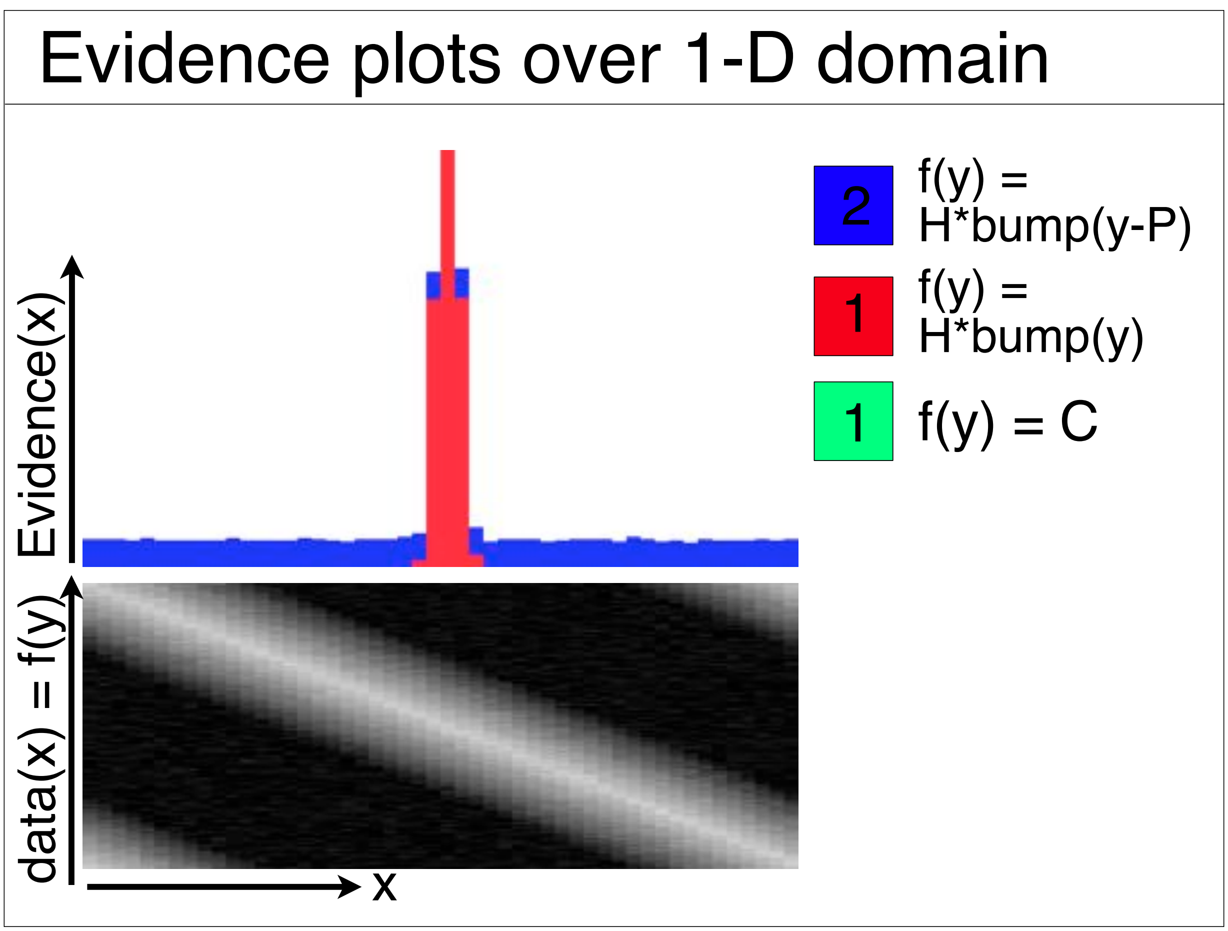
$$\overbrace{P(D|\mathcal{M}_i)}^{\text{Evidence}} = \int \overbrace{P(D|\mathbf{w},\mathcal{M}_i)}^{\text{Likelihood}} P(\mathbf{w}|\mathcal{M}_i)\,d\mathbf{w}$$

- Evidence: integral of data likelihood over model's entire parameter space (not specific to single fit)
- Uncertainty in model choice (of M_1 vs M_2) is quantified by comparing evidence(M_1), evidence(M_2)
- Perhaps Sum(evidence) shows gaps in models?


Evidence plots over 1-D domain
Evidence(x)
data(x) = f(y)
x
0
f(y) = 0


Evidence plots over 1-D domain
Evidence(x)
data(x) = f(y)
x
1
f(y) = C
0
f(y) = 0


Evidence plots over 1-D domain
Evidence(x)
data(x) = f(y)
x
2
f(y) =
H*bump(y-P)
1
f(y) =
H*bump(y)
1
f(y) = C


Evidence plots over 1-D domain
Evidence(x)
data(x) = f(y)
x
2
f(y) =
H*bump(y-P)
1
f(y) =
H*bump(y)
1
f(y) = C

Discussion

New opportunity for visualization to help understand statistics & add scientific value

- Data & models complex; can **see** how data can be explained, and where it can't
- Can show exactly when simple models suffice, backed up with statistical theory

Open research question: how to visualize the multiple scalar fields generated by computing per-pixel model evidences?

Thank you! glk@uchicago.edu