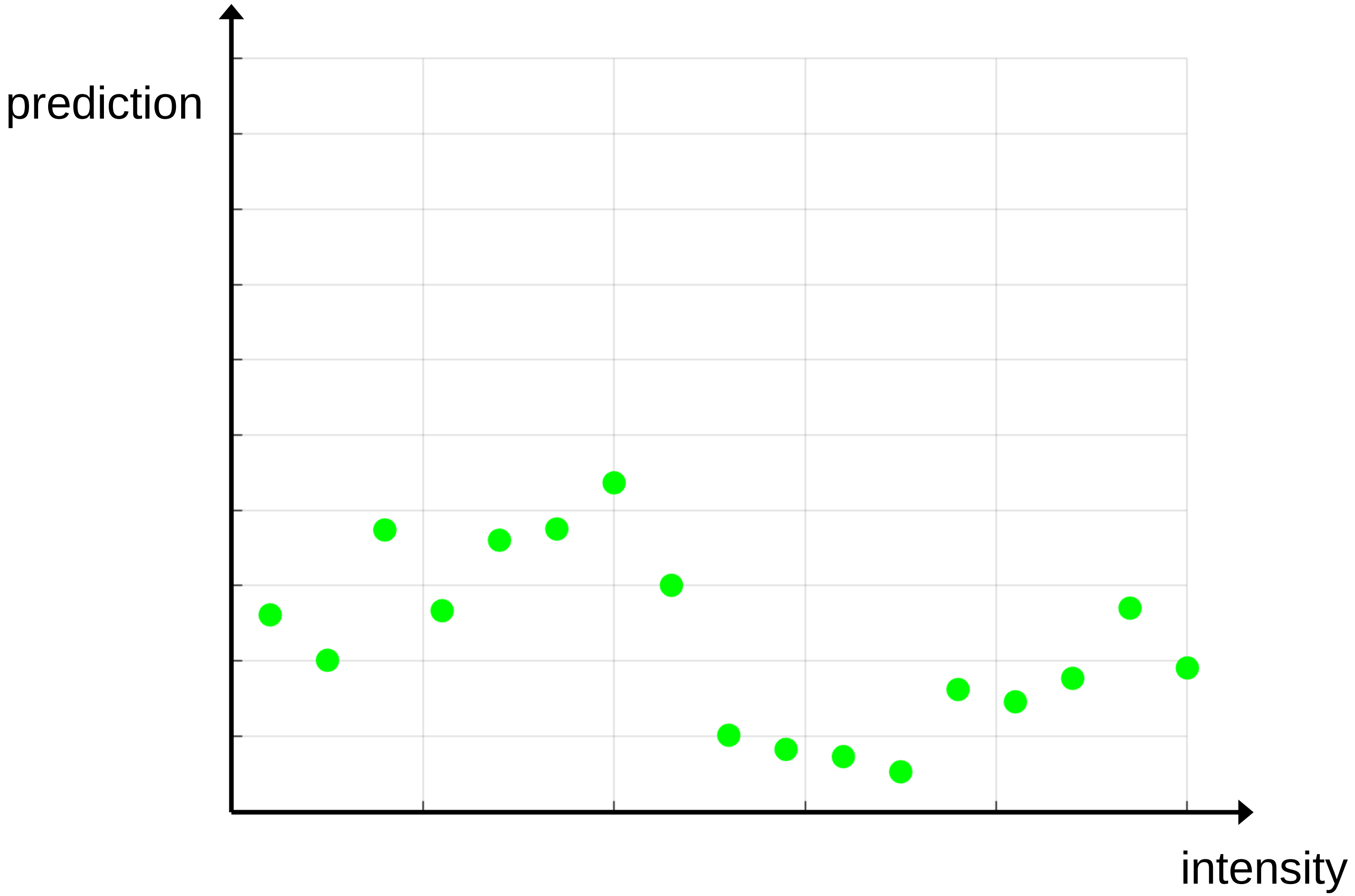
Input and Output Uncertainty in Medical Image Analysis

Koen Van Leemput

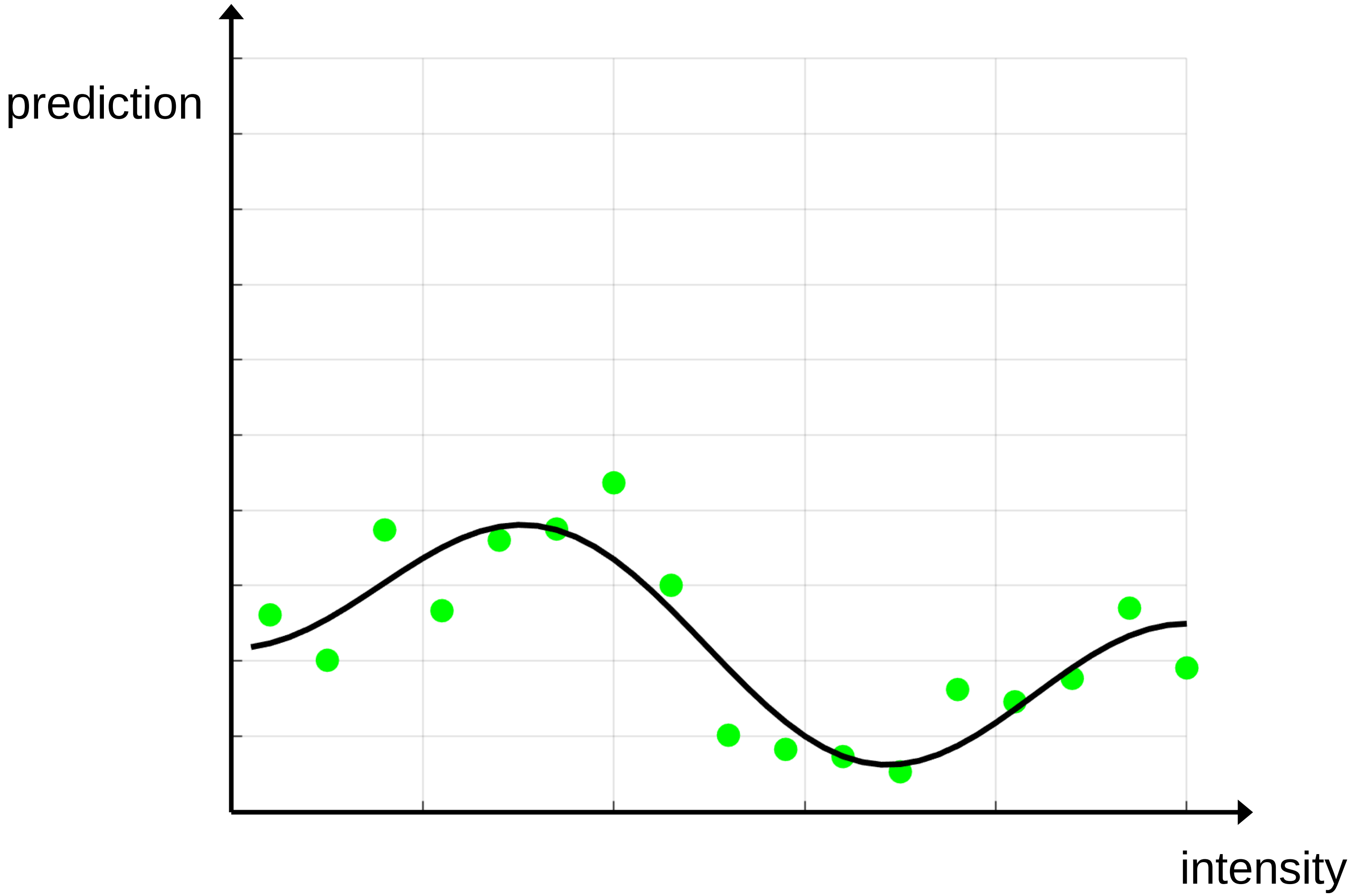
October 17, 2019
MICCAI UNSURE Workshop
Shenzhen, China

MGH/HST Athinoula A. Martinos Center for Biomedical Imaging
MASSACHUSETTS GENERAL HOSPITAL
Harvard-MIT Health Sciences & Technology
DTU Technical University of Denmark

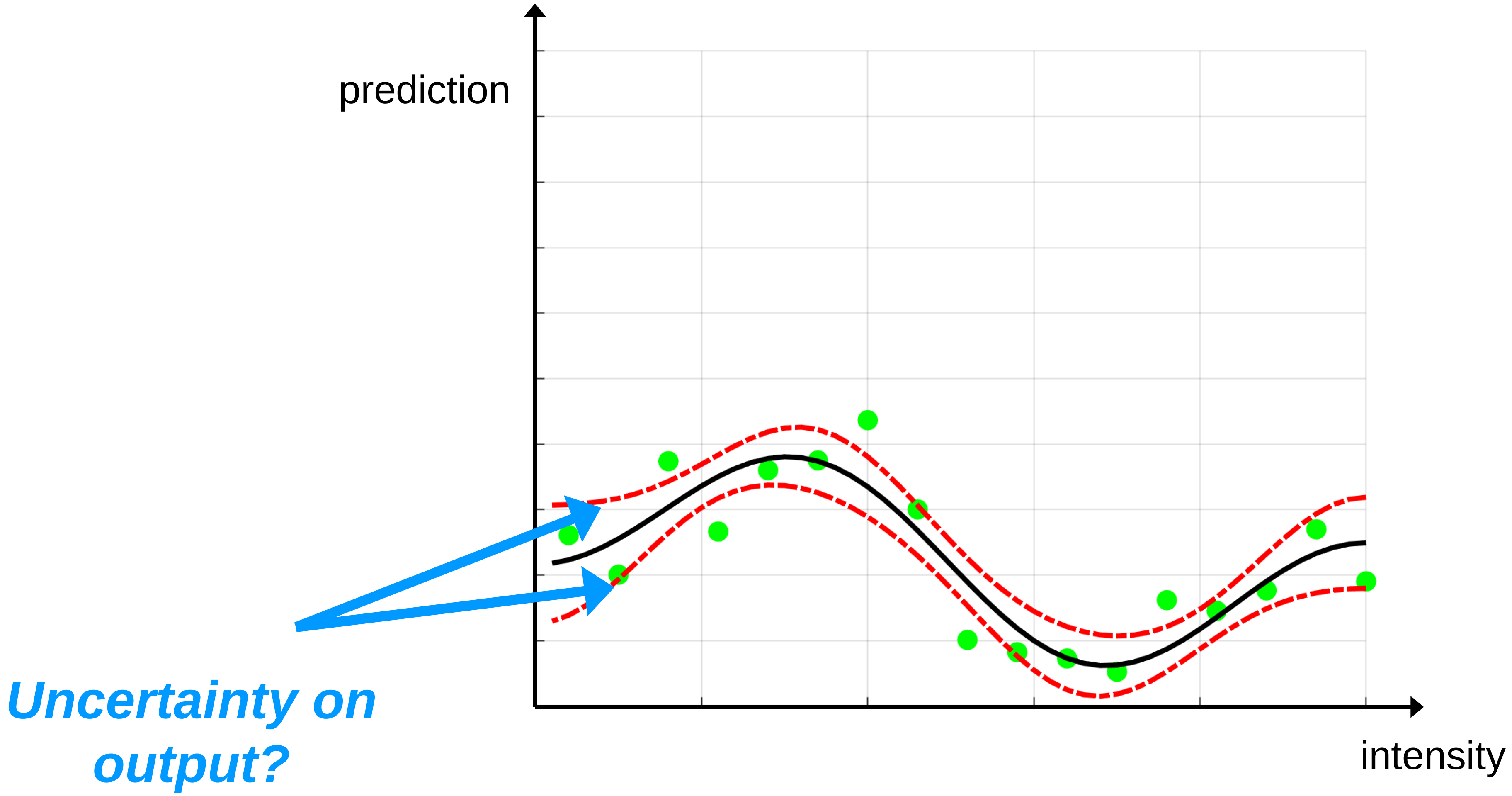
Uncertainty in input and output?



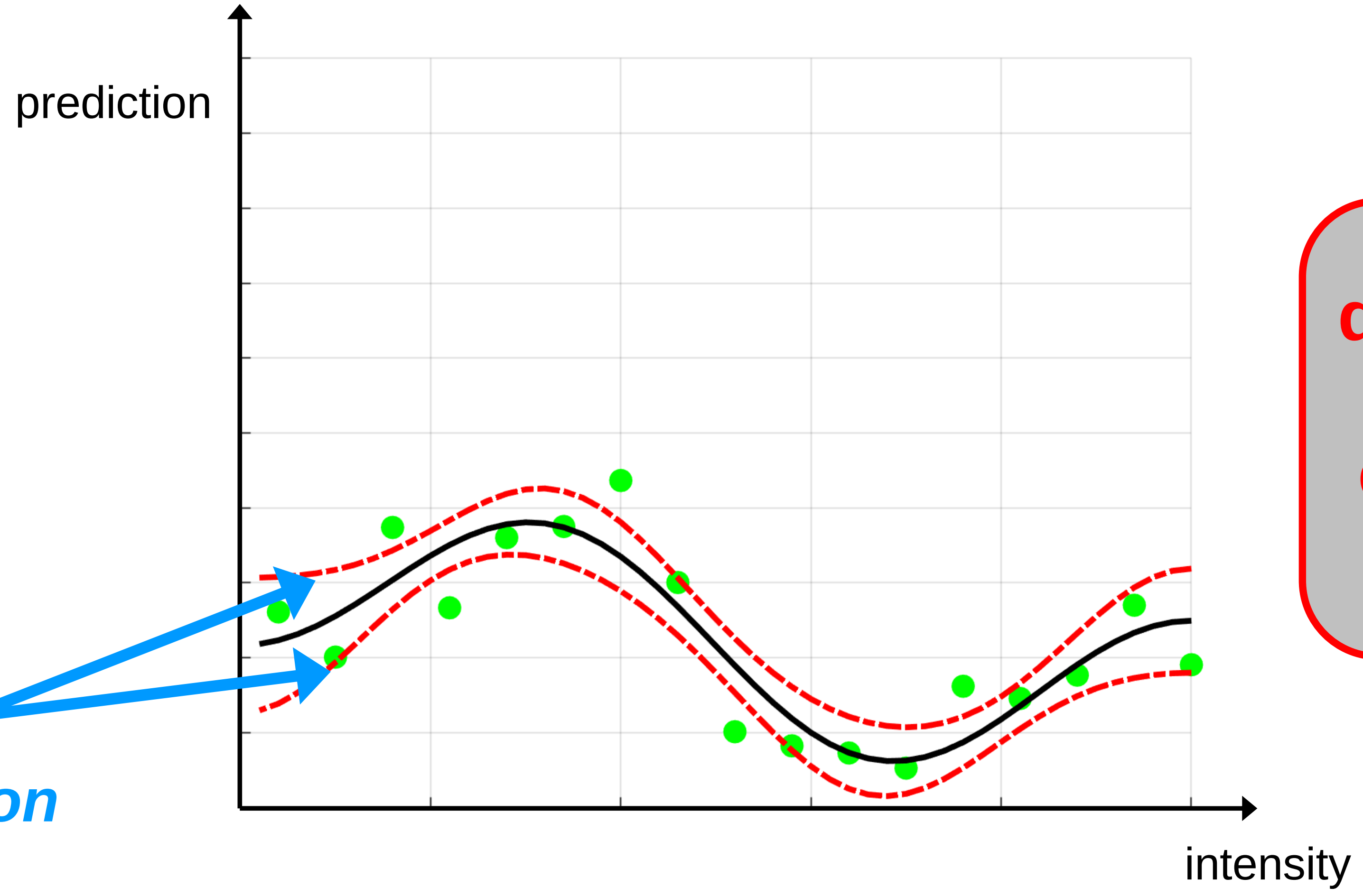
Uncertainty in input and output?



Uncertainty in input and output?



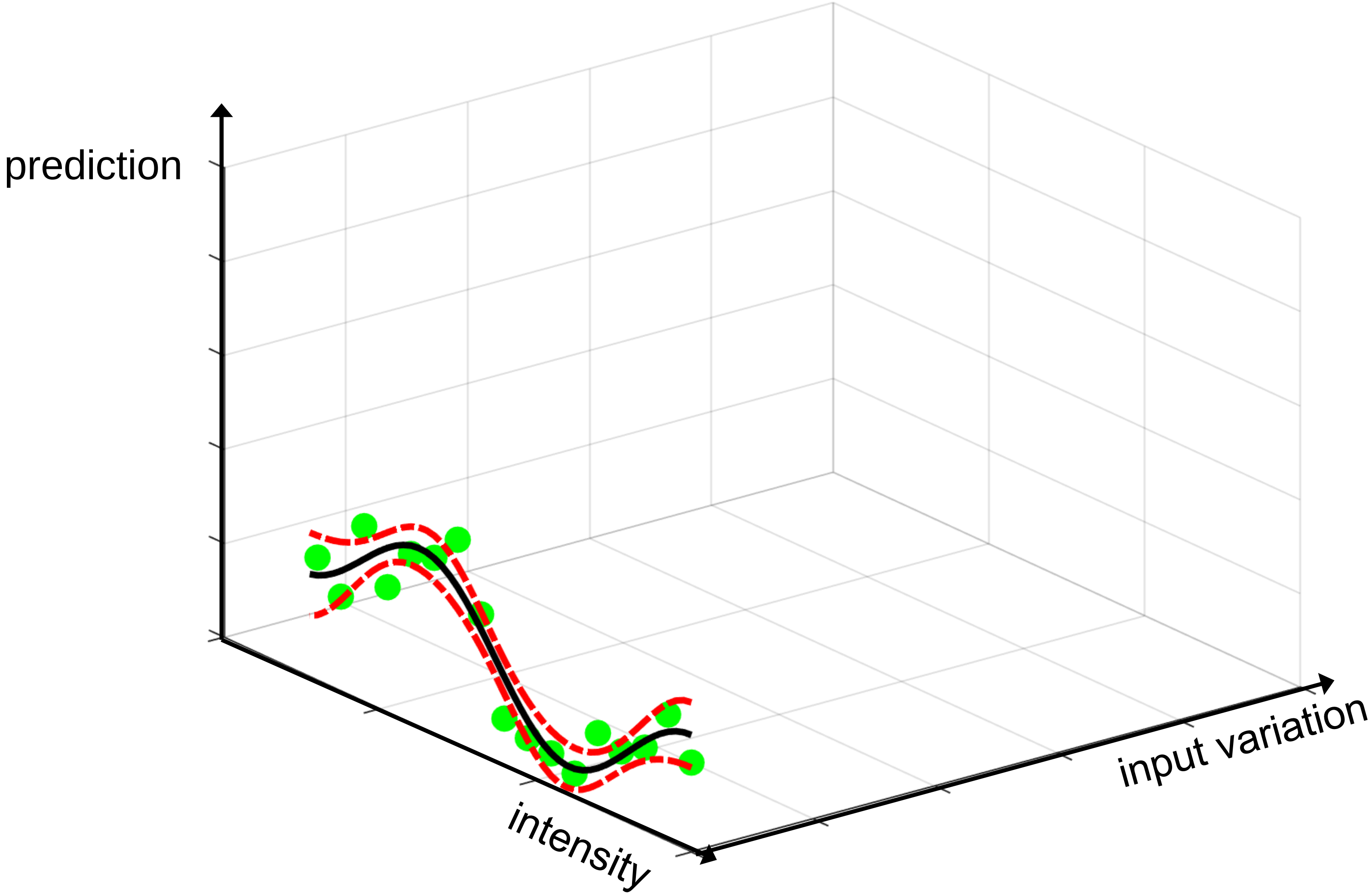
Uncertainty in input and output?



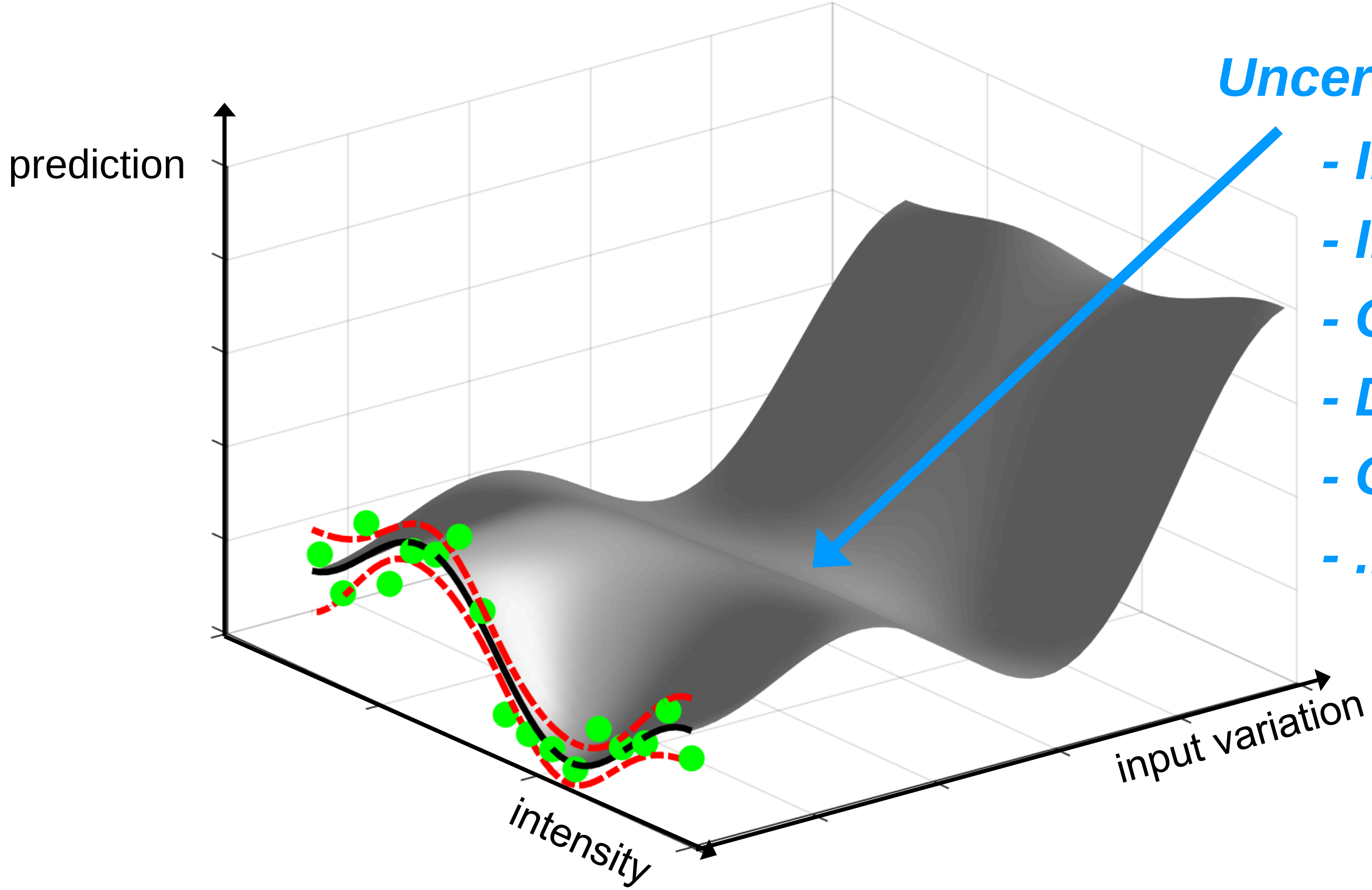
Uncertainty on output?

discriminative models:
CNNs + dropout at test time

Uncertainty in input and output?



Uncertainty in input and output?

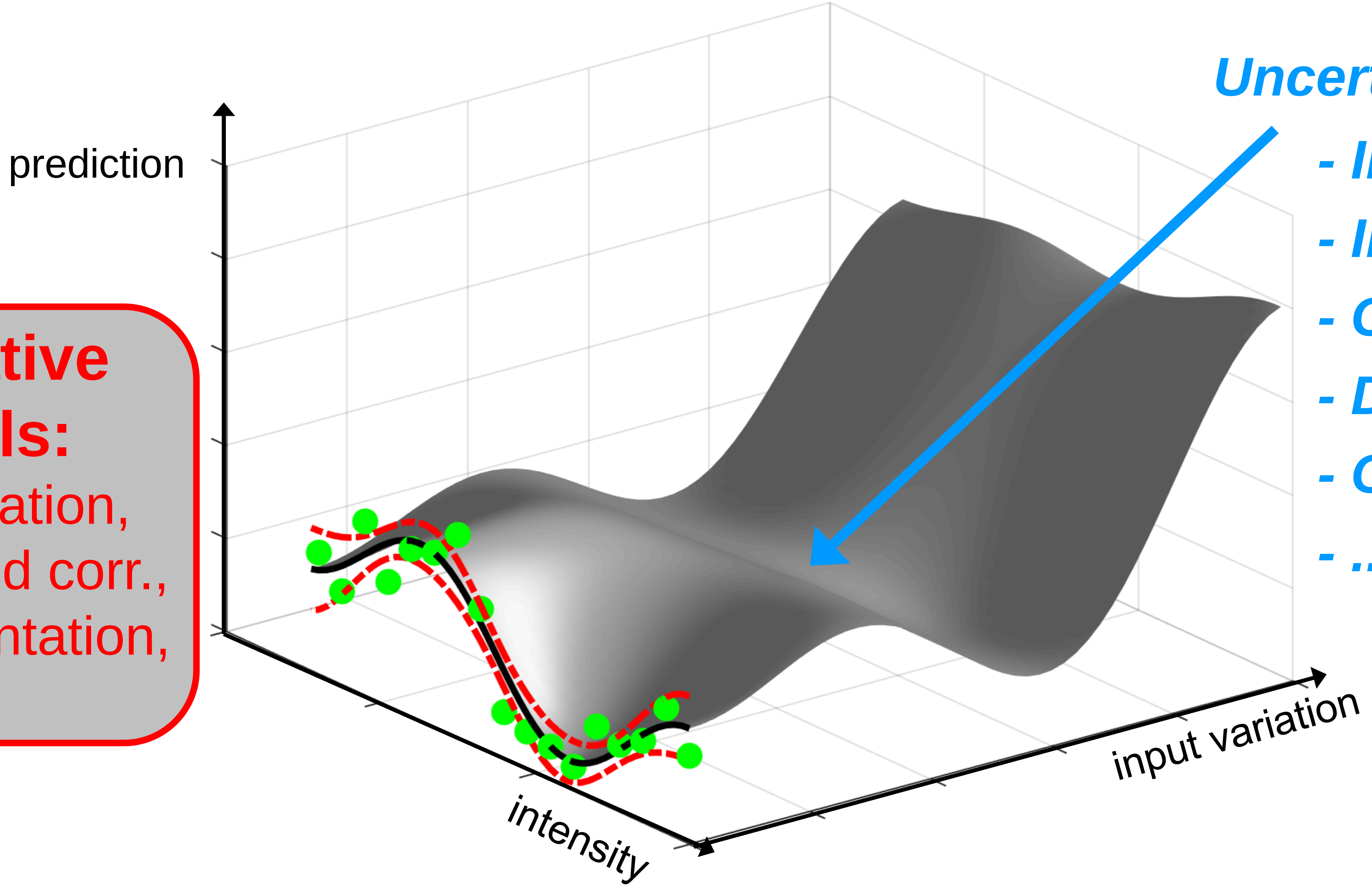


Uncertainty on input

- ***Imaging hardware?***
- ***Imaging protocol?***
- ***Contrast agent?***
- ***Disease?***
- ***Organ?***
- ***...?***

Uncertainty in input and output?

generative models:
MI registration,
N3 bias field corr.,
EM segmentation,
…



Uncertainty on input

- ***Imaging hardware?***
- ***Imaging protocol?***
- ***Contrast agent?***
- ***Disease?***
- ***Organ?***
- ***...?***

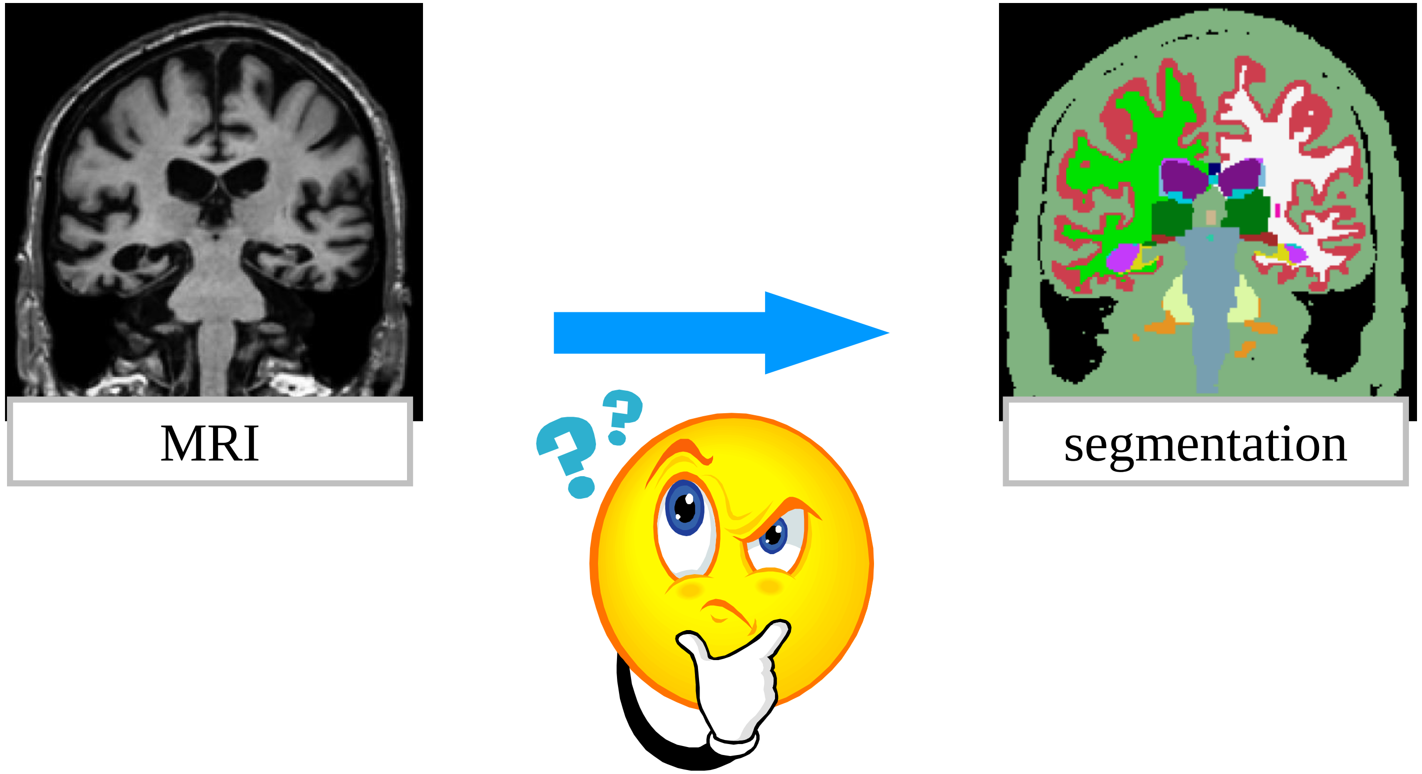
Overview

Uncertainty on input using generative models

Uncertainty on output using Monte Carlo sampling

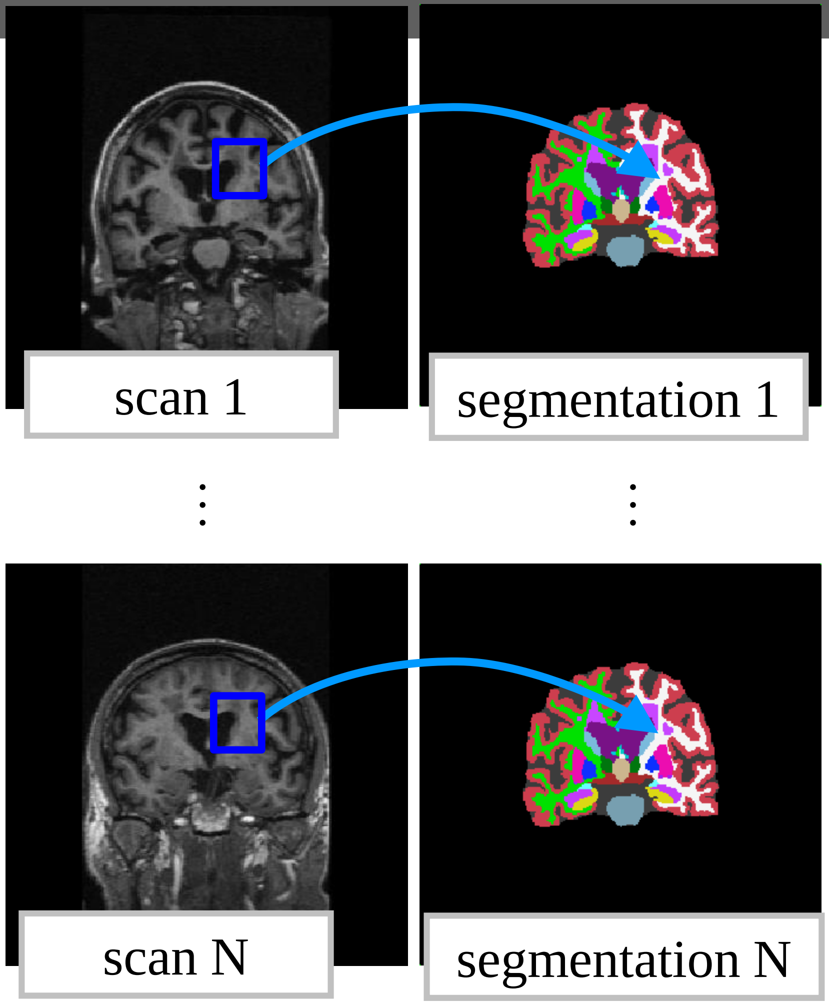
Discussion and conclusion

Brain segmentation in the clinic



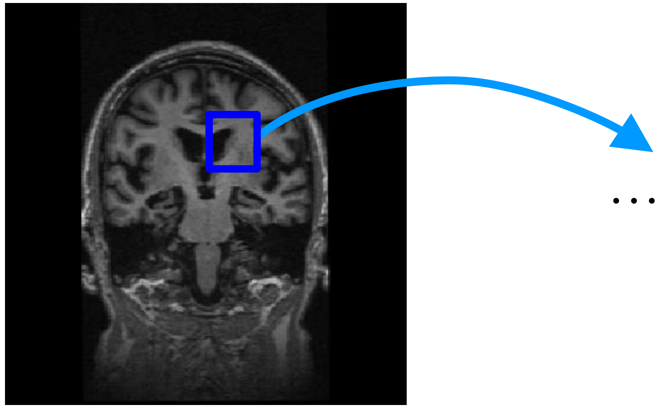
Focus: translation into clinical settings!

Discriminative methods?



Training data

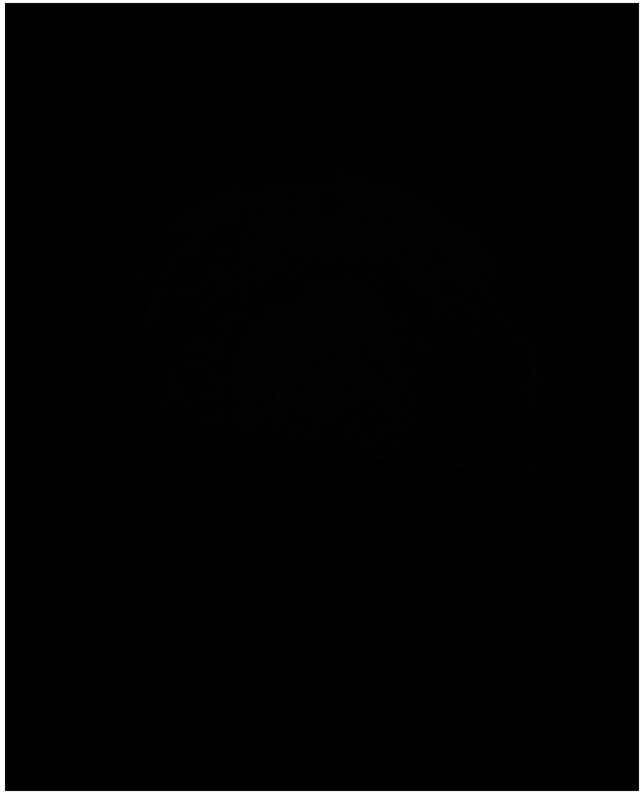
Directly learn to predict segmentation labels from intensities

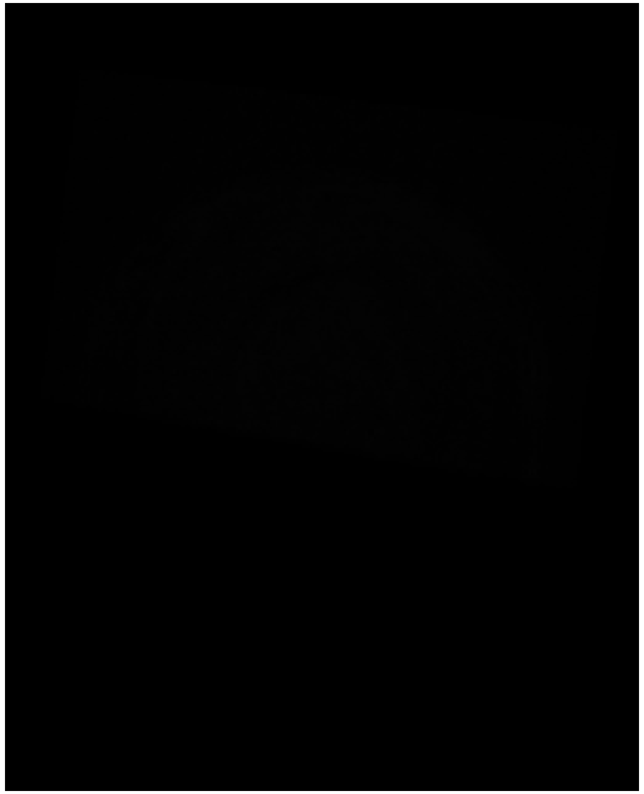


“Monkey see, monkey do!”

Discriminative methods?

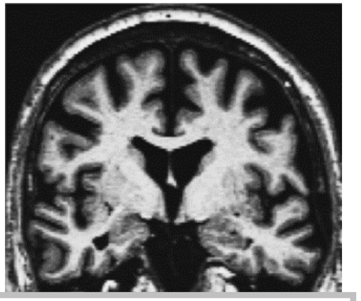
How long to manually label one MRI scan?





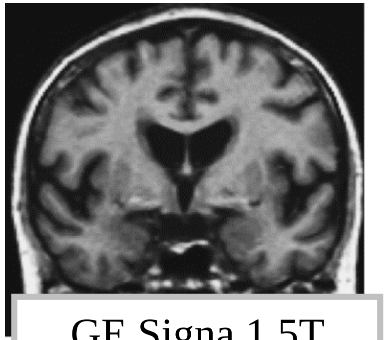
Discriminative methods?

Tissue contrast depends on hardware and pulse sequence



Siemens Vision 1.5T MPRAGE

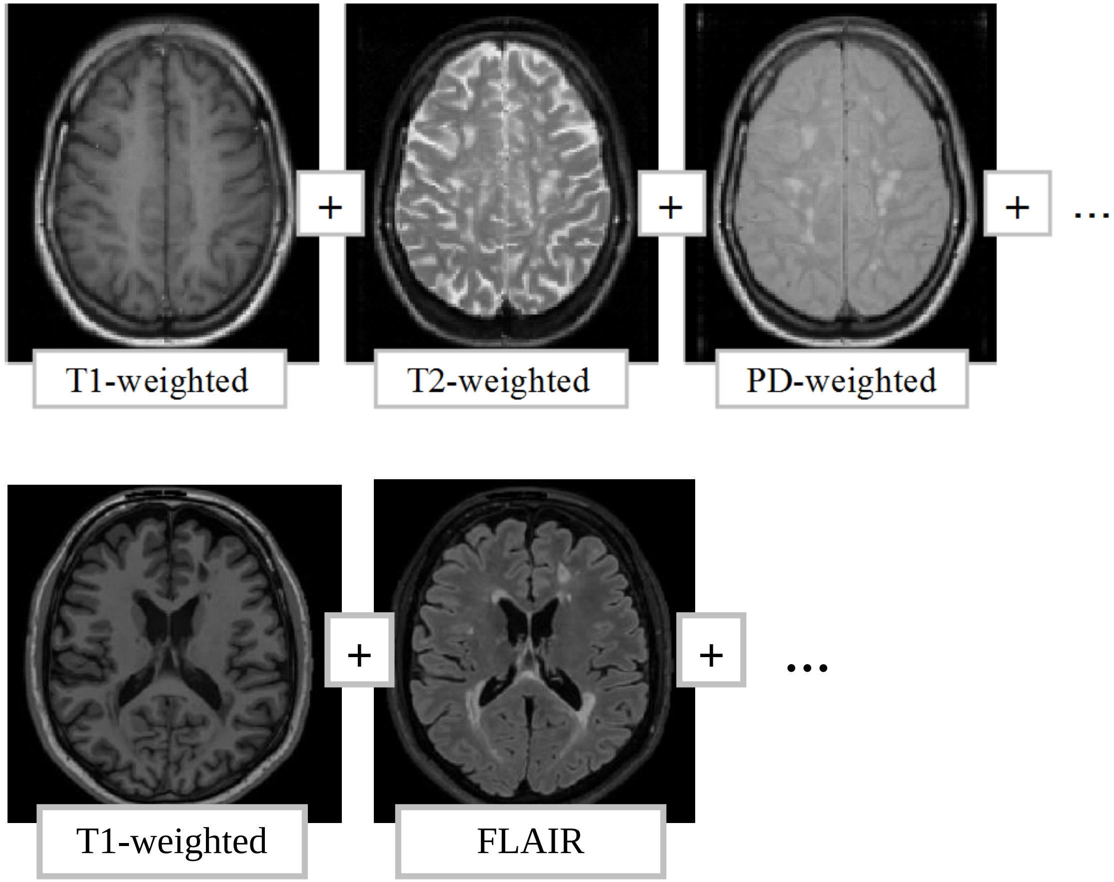
Siemens Tim Trio 3T MPRAGE



GE Signa 1.5T SPGR

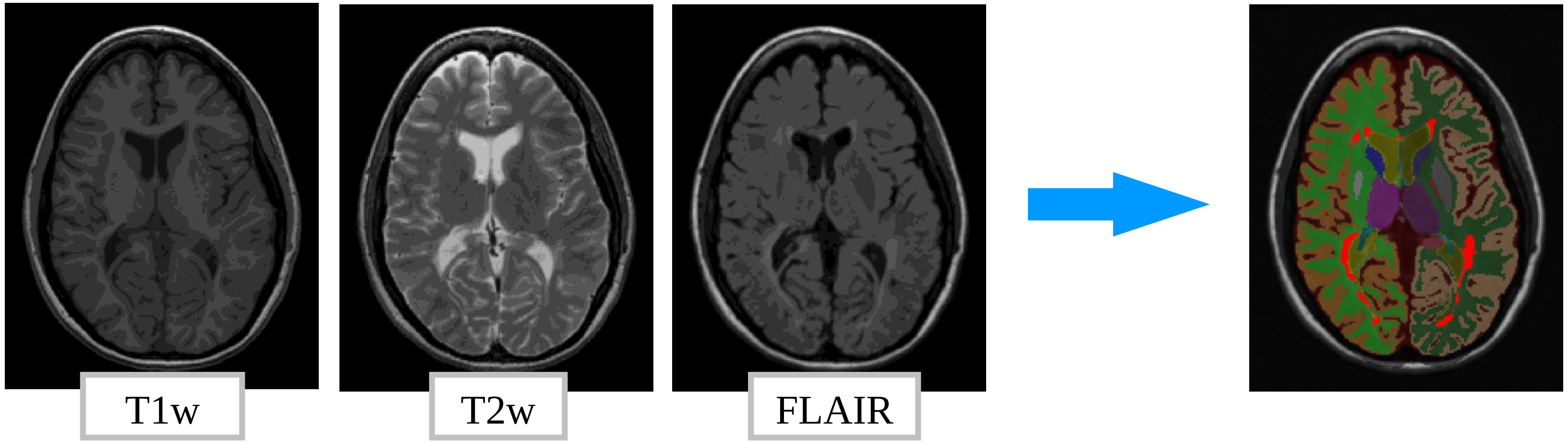
Discriminative methods?

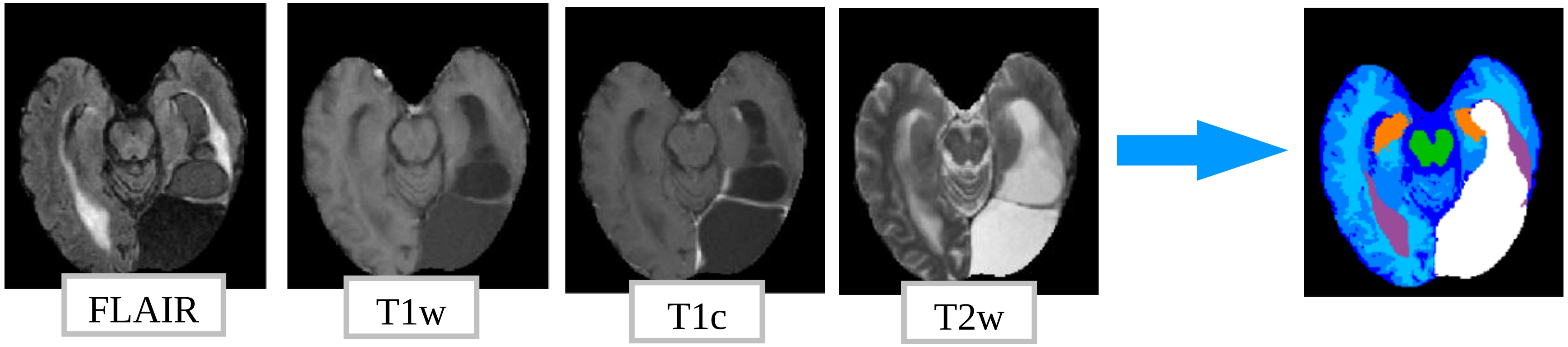
No standard protocol for clinical imaging (multi-contrast)



Discriminative methods?

Many different brain diseases





Generative methods: Gaussian mixture model

– Voxel-wise classification: maximize $p(l|d)$

label (3 classes)

intensity

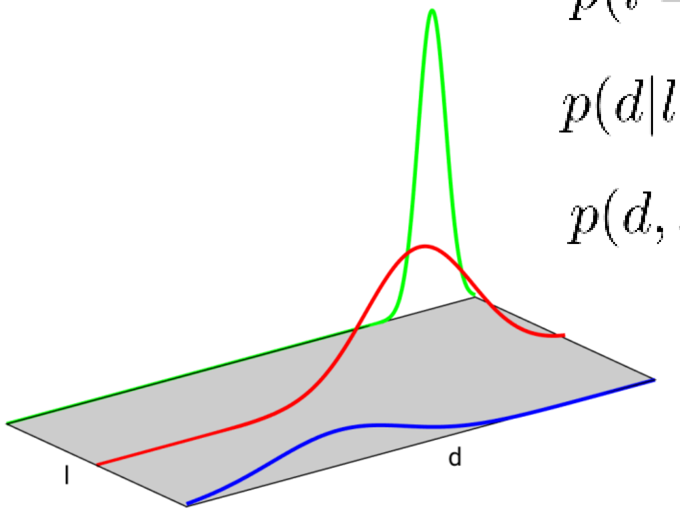
Generative methods: Gaussian mixture model

– Voxel-wise classification: maximize $p(l|d)$

label (3 classes)

intensity

– Idea: model the joint distribution $p(d, l)$ instead



$$p(l = k) = \pi_k$$

$$p(d|l = k) = \mathcal{N}(d|\mu_k, \sigma_k^2)$$

$$p(d, l) = p(d|l)p(l)$$

Generative methods: Gaussian mixture model

– Voxel-wise classification: maximize $p(l|d)$

label (3 classes)

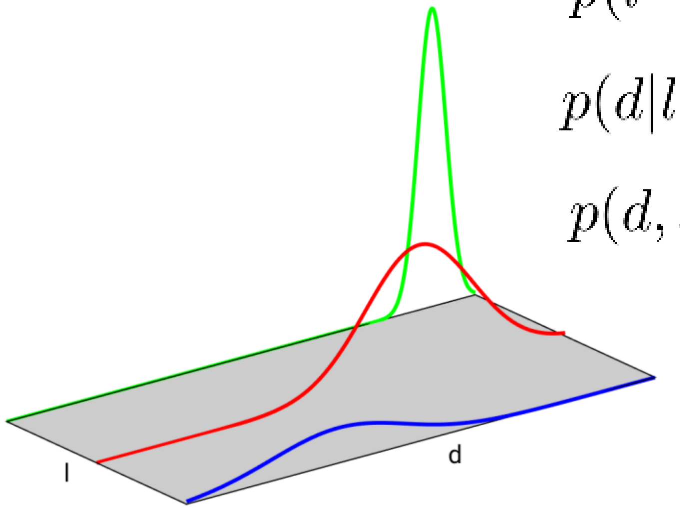
intensity

– Idea: model the joint distribution $p(d, l)$ instead

$$p(l = k) = \pi_k$$

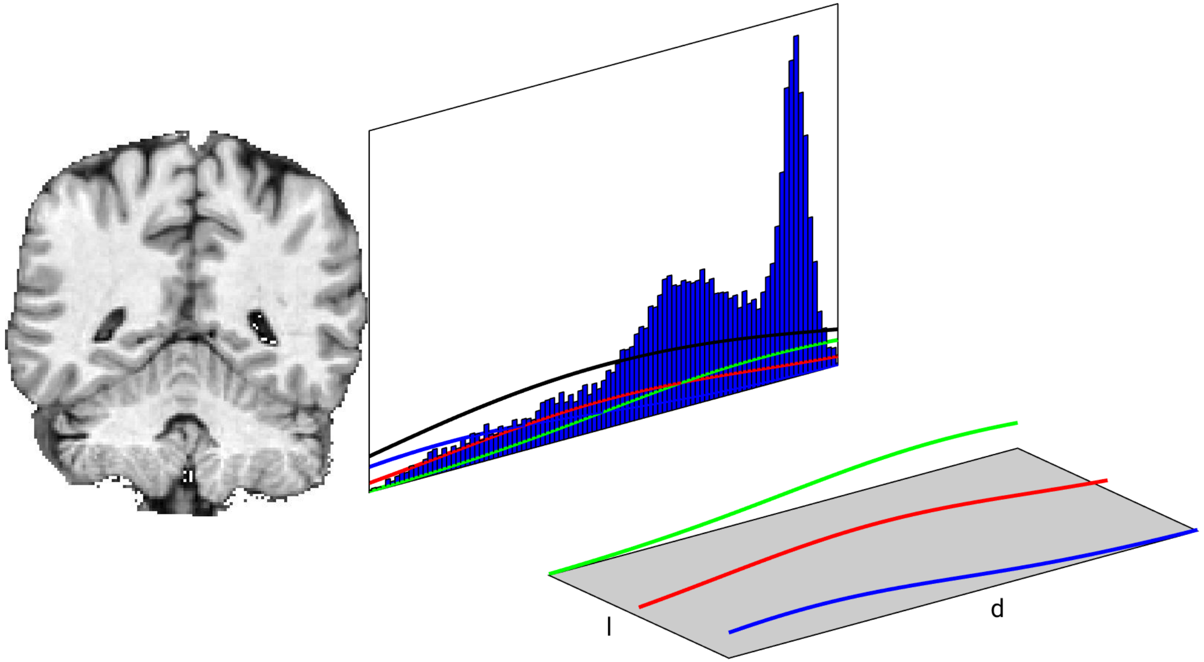
$$p(d|l = k) = \mathcal{N}(d|\mu_k, \sigma_k^2)$$

$$p(d, l) = p(d|l)p(l)$$

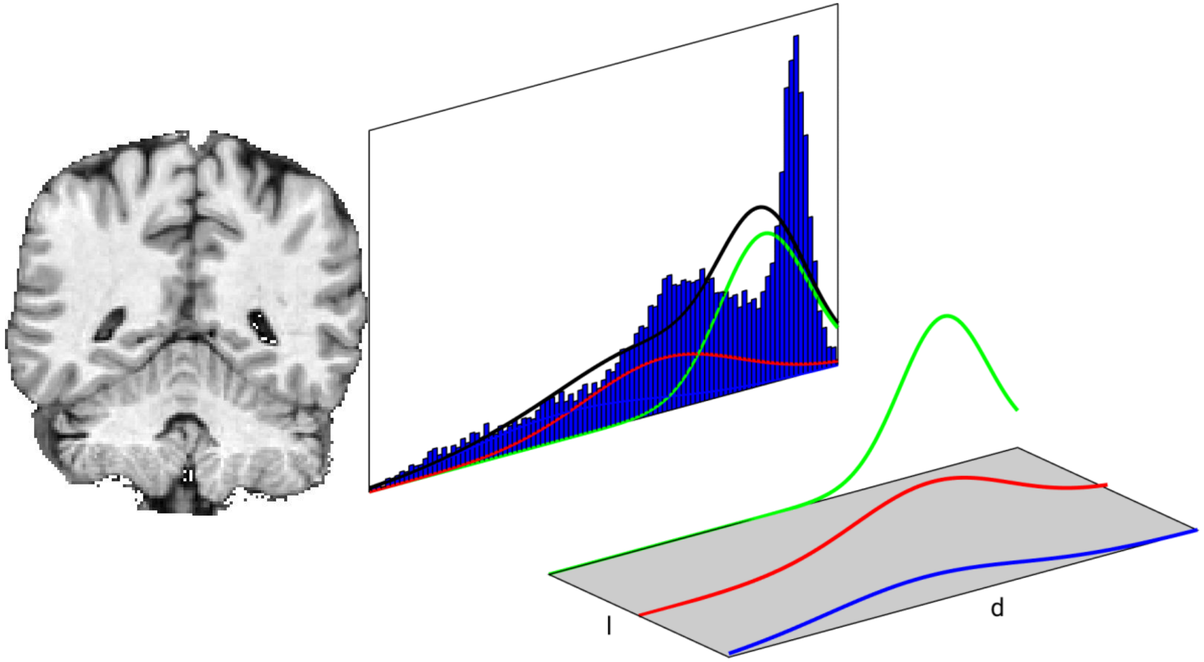


model parameters

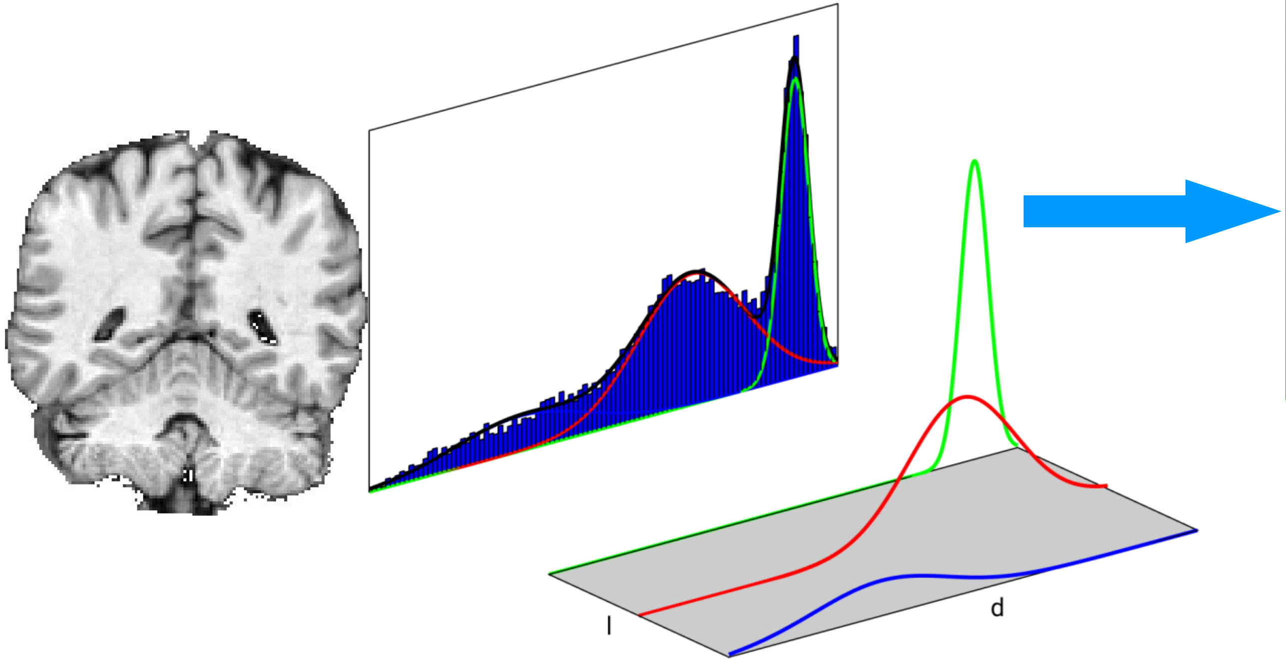
Generative methods: Gaussian mixture model

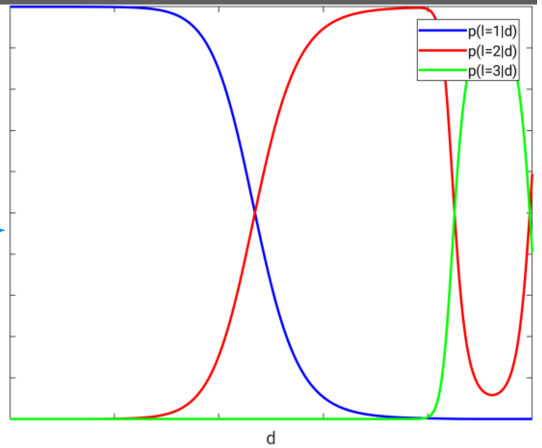


Generative methods: Gaussian mixture model



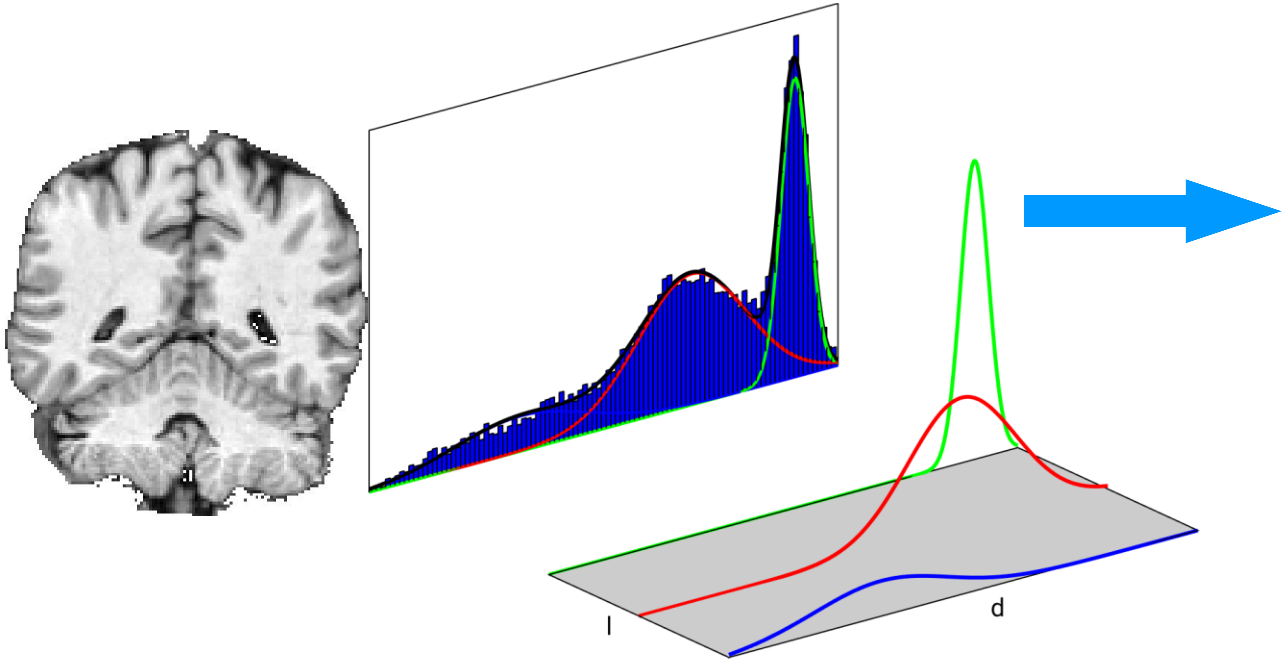
Generative methods: Gaussian mixture model

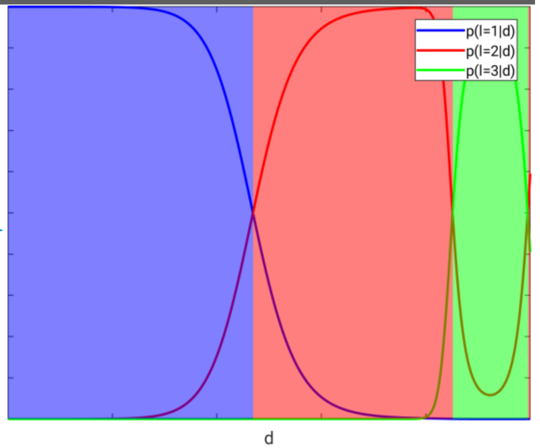




$$p(l|d) = \frac{p(d|l)p(l)}{p(d)}$$

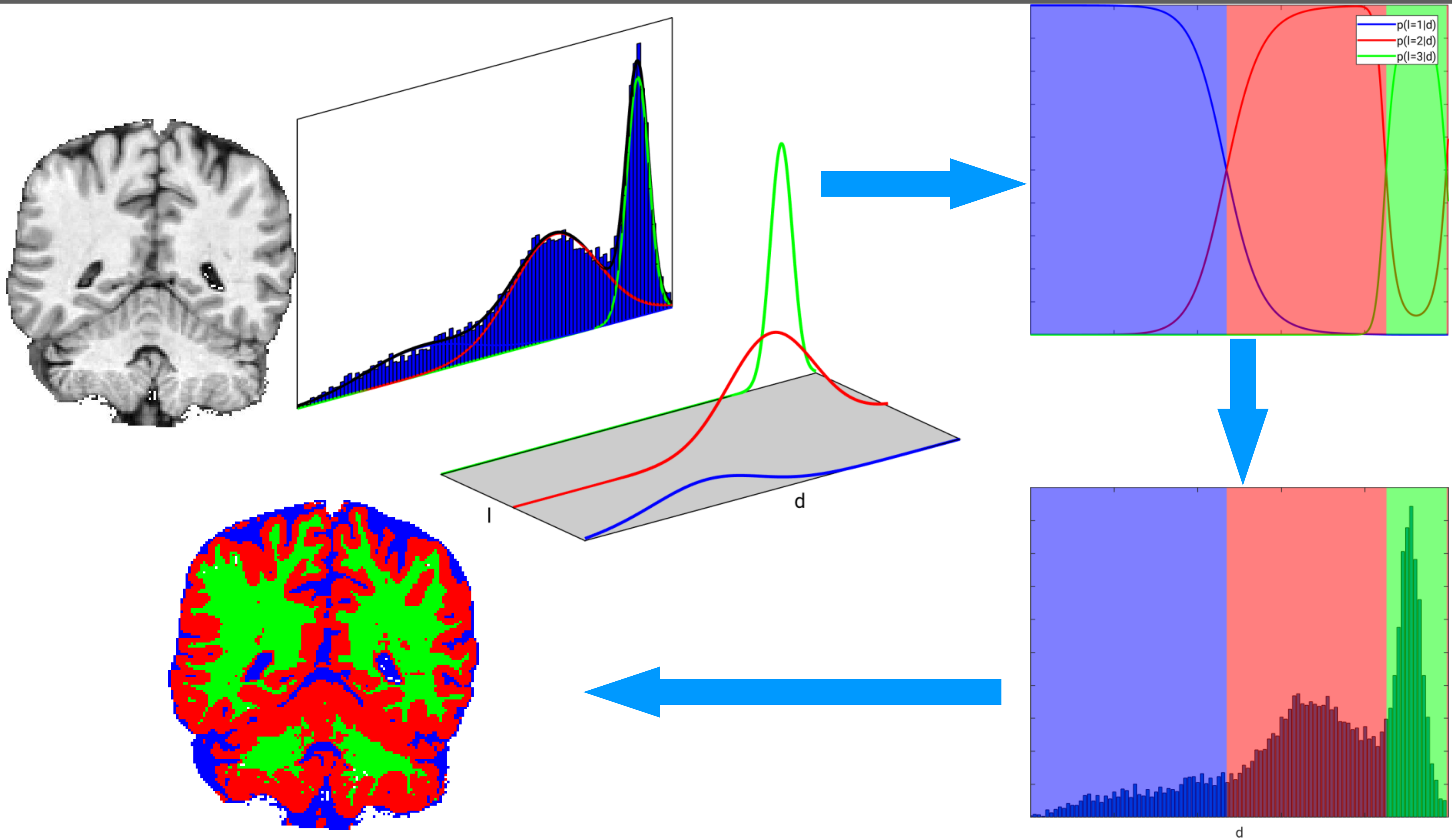
Generative methods: Gaussian mixture model



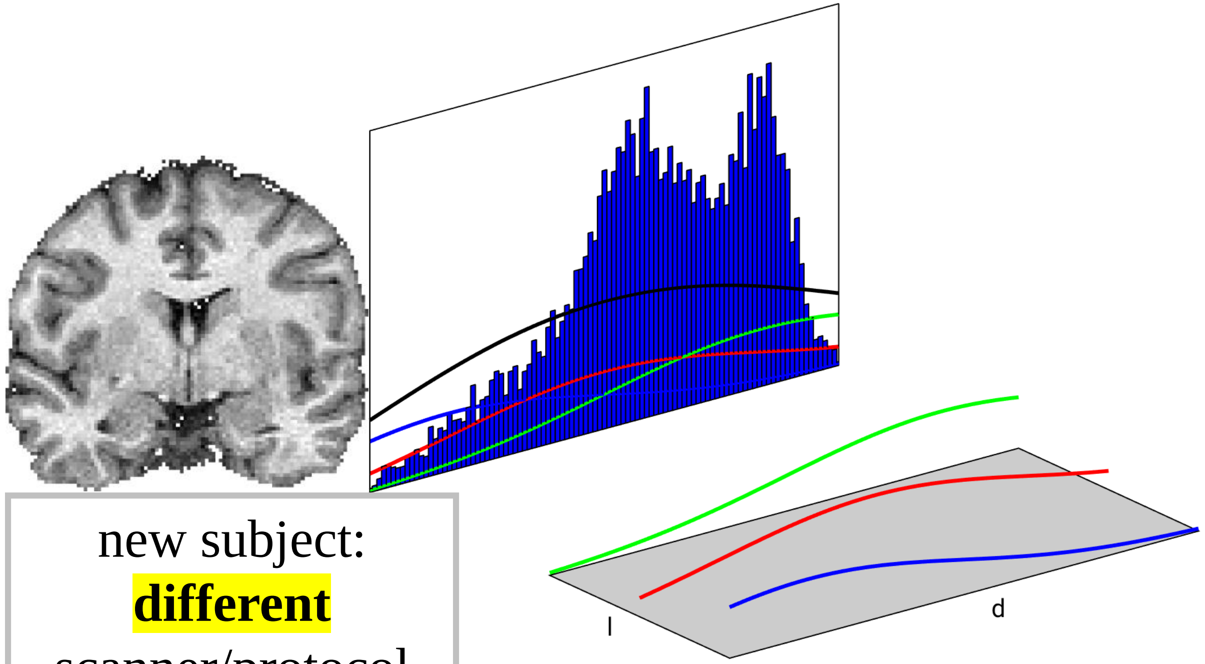


$$p(l|d) = \frac{p(d|l)p(l)}{p(d)}$$

Generative methods: Gaussian mixture model

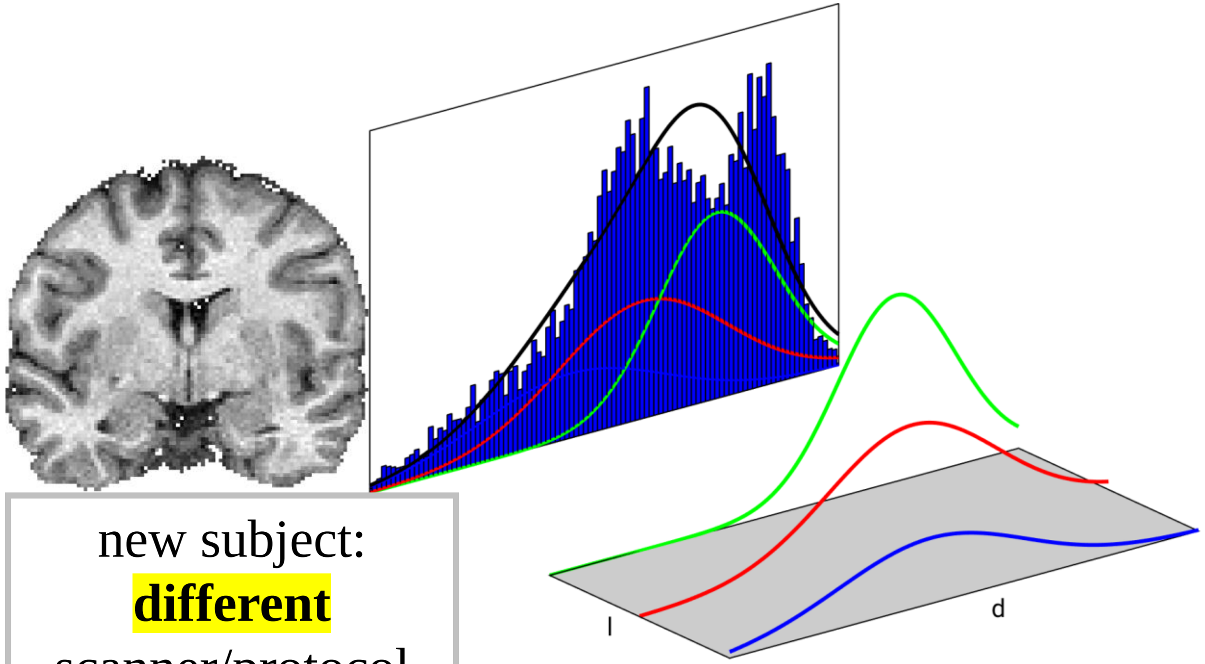


Generative methods: Gaussian mixture model



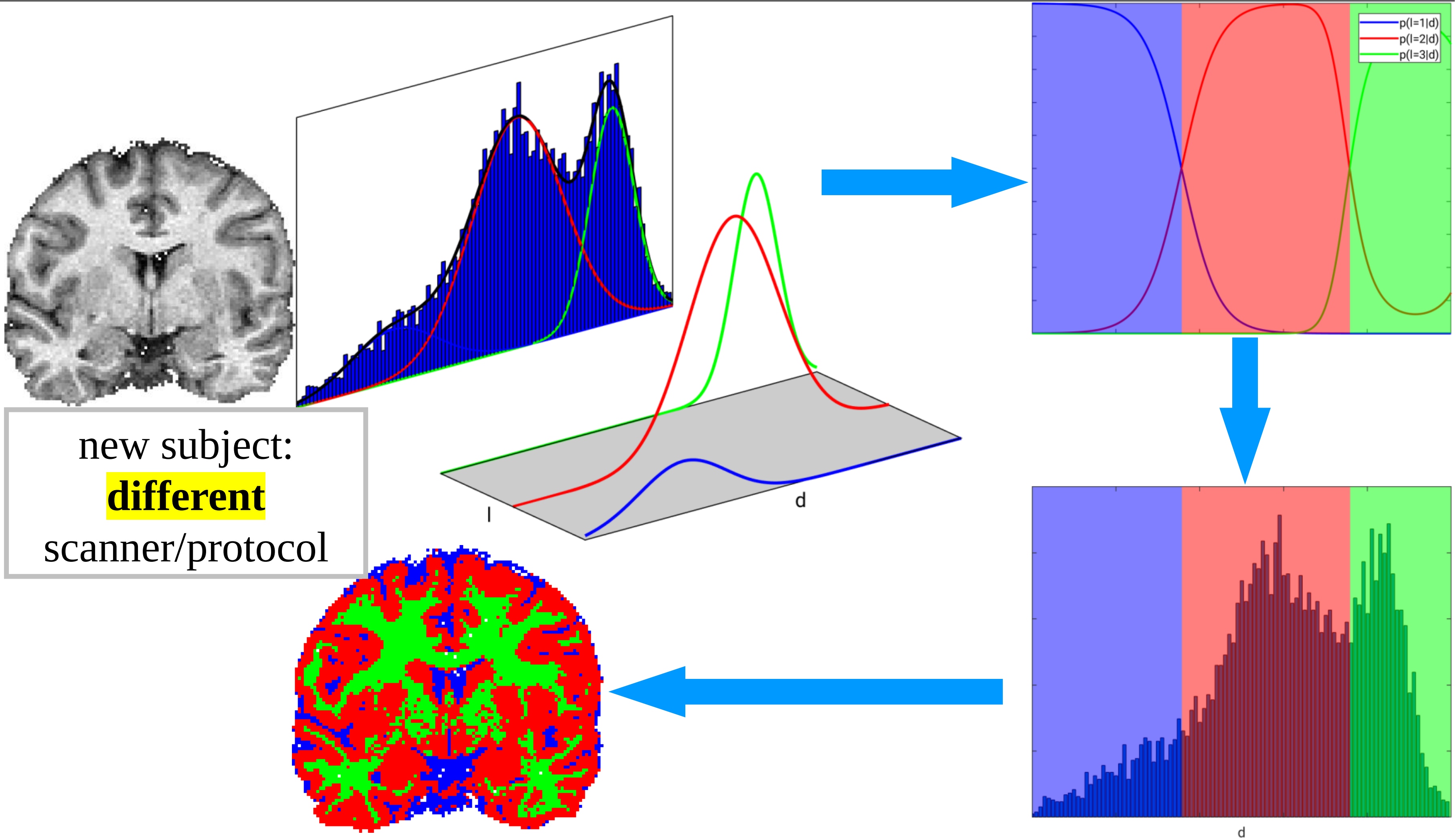
new subject:
different
scanner/protocol

Generative methods: Gaussian mixture model



new subject:
different
scanner/protocol

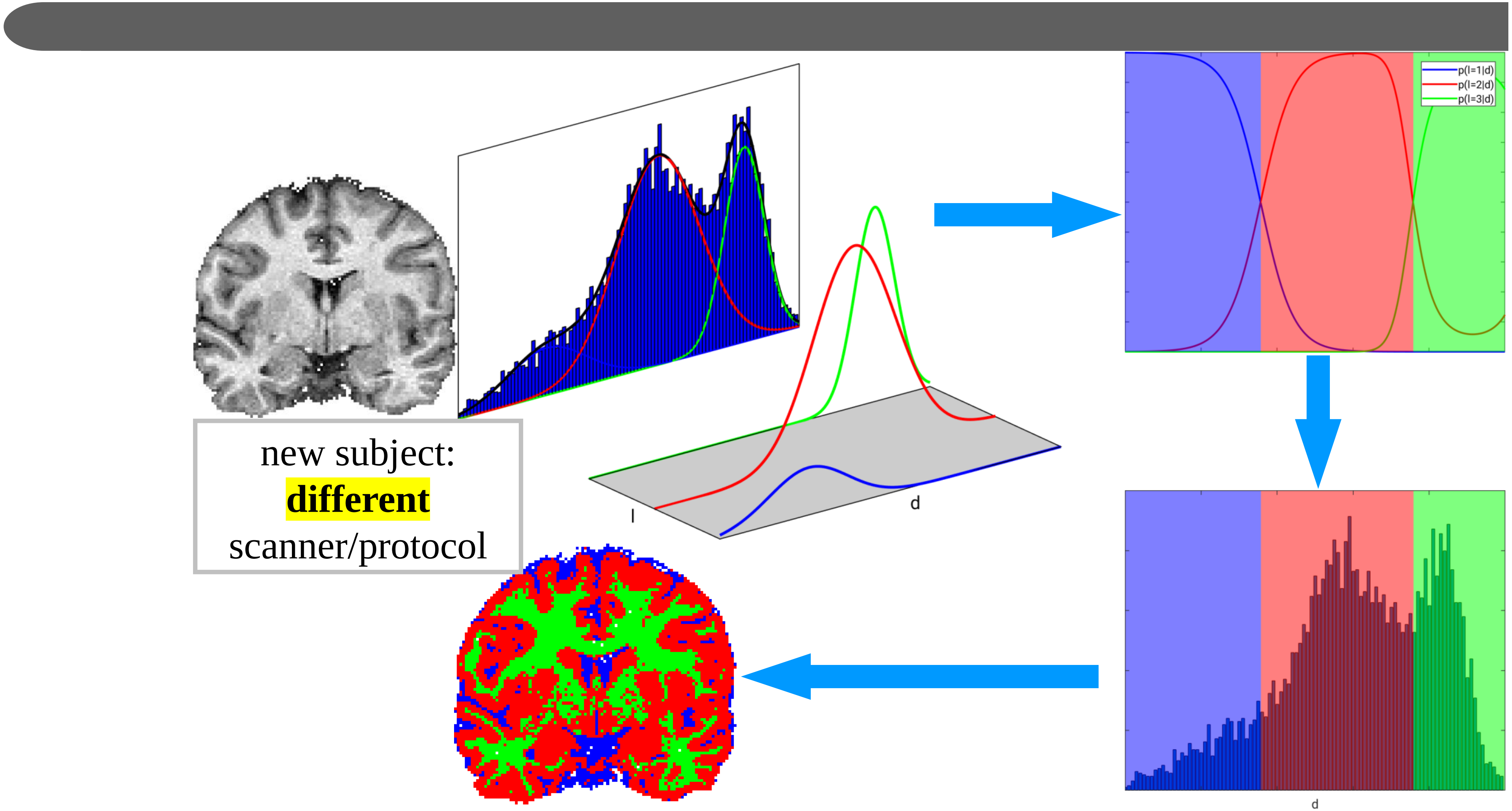
Generative methods: Gaussian mixture model



new subject:
different
scanner/protocol

Adaptive!
No matching training data needed!

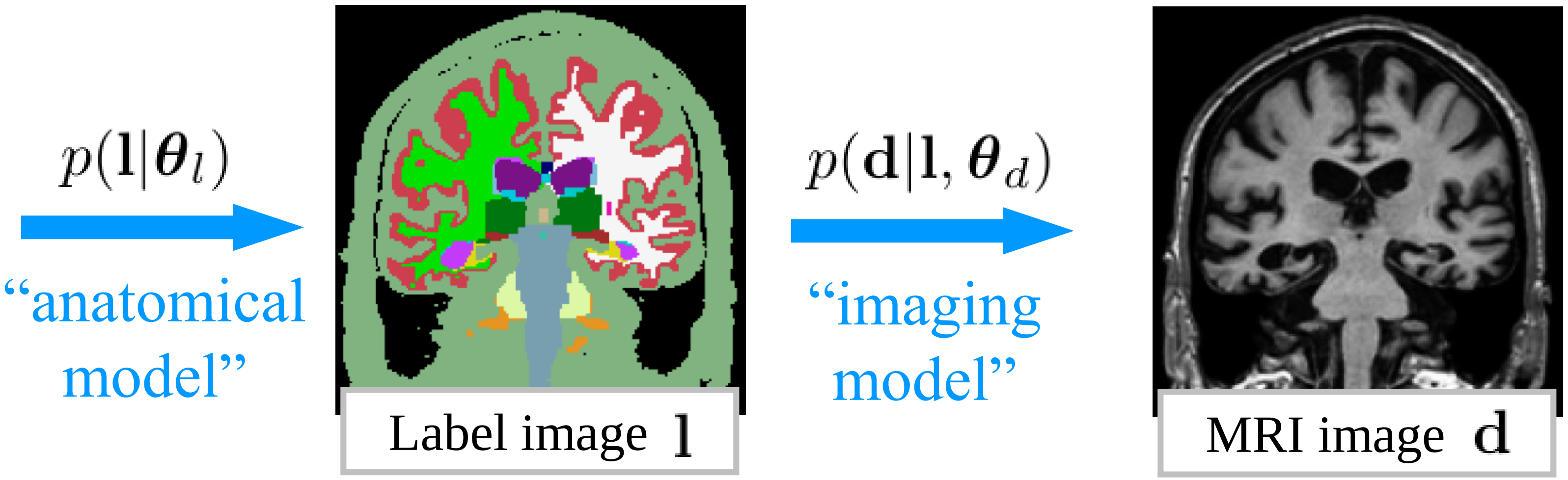
Slow because of image-specific optimization



new subject:
different
scanner/protocol

Generative models

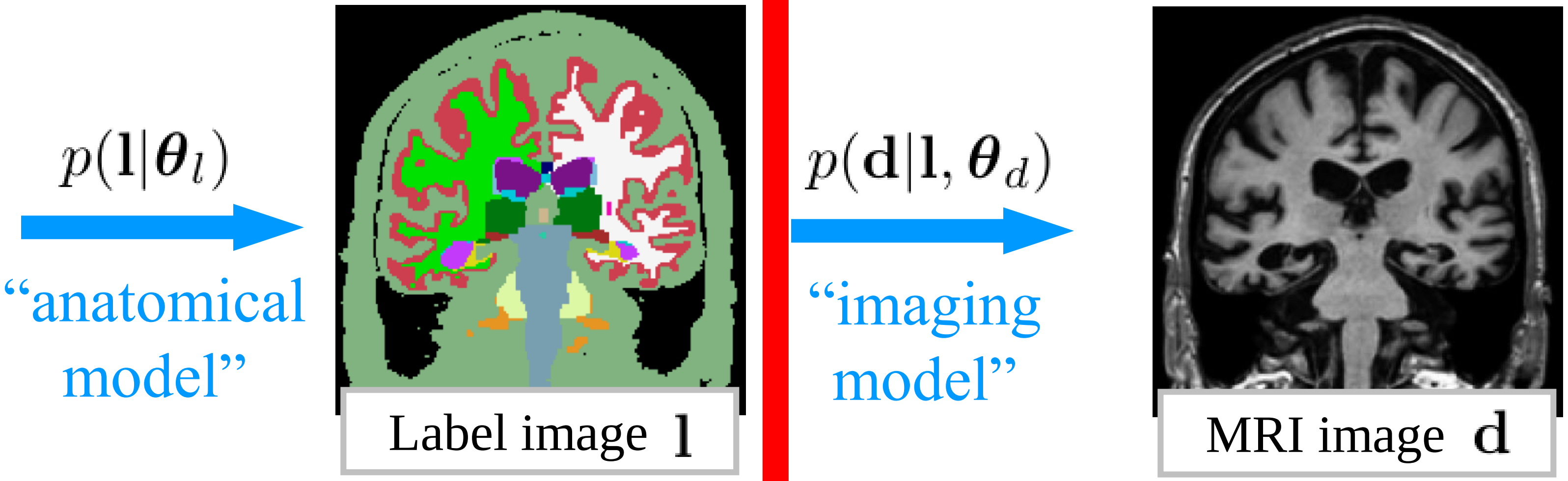
- A probabilistic model of how an MRI image is formed



- The model depends on some parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_l^{\mathrm{T}}, \boldsymbol{\theta}_d^{\mathrm{T}})^{\mathrm{T}}$
- Appropriate values $\hat{\theta}$ can be estimated

Generative models

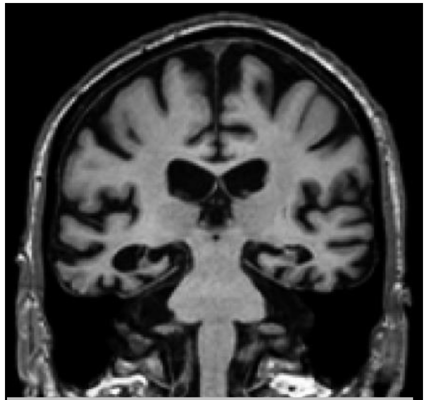
– A probabilistic model of how an MRI image is formed



– The model depends on some parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_l^{\mathrm{T}}, \boldsymbol{\theta}_d^{\mathrm{T}})^{\mathrm{T}}$

– Appropriate values $\hat{\boldsymbol{\theta}}$ can be estimated

Segmentation = inverse problem



MRI image $\mathbf{d}$

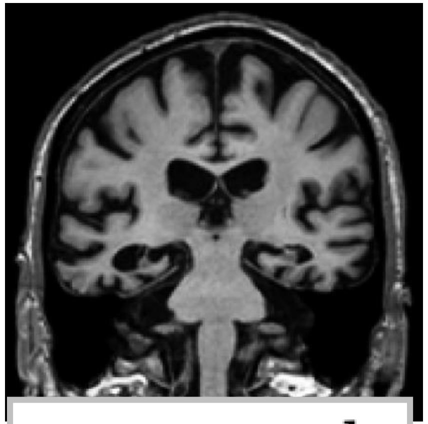
Label image $\mathbf{l}$

$$\hat{\mathbf{l}} = \arg\max_{\mathbf{l}} p(\mathbf{l}|\mathbf{d}, \hat{\boldsymbol{\theta}})$$

The posterior distribution $p(\mathbf{l}|\mathbf{d}, \hat{\boldsymbol{\theta}})$ is given by Bayes rule:

$$p(\mathbf{l}|\mathbf{d}, \hat{\boldsymbol{\theta}}) = \frac{p(\mathbf{d}|\mathbf{l}, \hat{\boldsymbol{\theta}}_d)p(\mathbf{l}|\hat{\boldsymbol{\theta}}_l)}{p(\mathbf{d}|\hat{\boldsymbol{\theta}})}$$

Segmentation = inverse problem



MRI image $\mathbf{d}$

Label image $\mathbf{l}$

$$\hat{\mathbf{l}} = \arg\max_{\mathbf{l}} p(\mathbf{l}|\mathbf{d}, \hat{\boldsymbol{\theta}})$$

The posterior distribution $p(\mathbf{l}|\mathbf{d}, \hat{\boldsymbol{\theta}})$ is given by Bayes rule:

imaging model

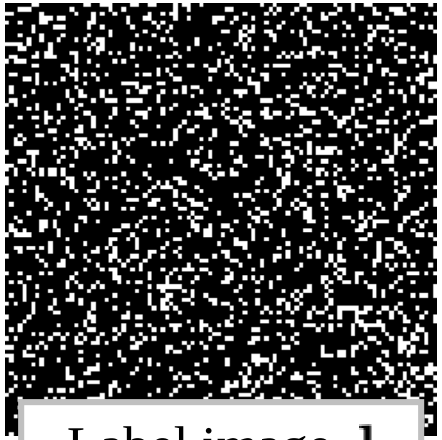
$$p(\mathbf{l}|\mathbf{d}, \hat{\boldsymbol{\theta}}) = \frac{p(\mathbf{d}|\mathbf{l}, \hat{\boldsymbol{\theta}}_d)\, p(\mathbf{l}|\hat{\boldsymbol{\theta}}_l)}{p(\mathbf{d}|\hat{\boldsymbol{\theta}})}$$

anatomical model

Gaussian mixture model

$p(\mathbf{l}|\boldsymbol{\theta}_l)$

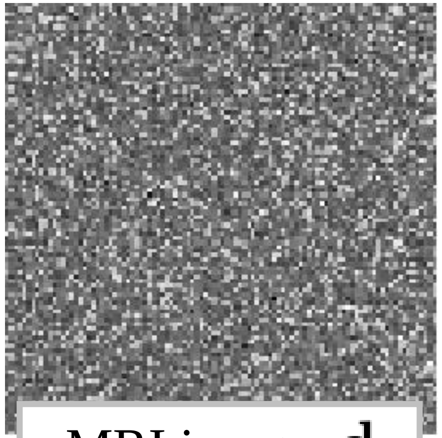
"anatomical model"



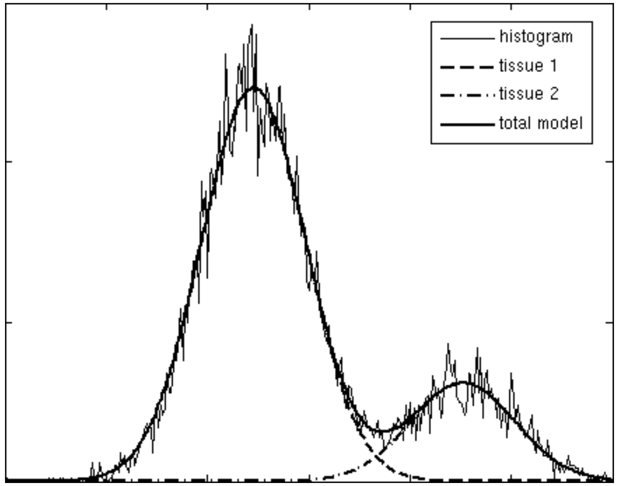
Label image $\mathbf{l}$

$p(\mathbf{d}|\mathbf{l}, \boldsymbol{\theta}_d)$

"imaging model"



MRI image $\mathbf{d}$



You can't segment 40+ brain structures with this…

Anatomical model

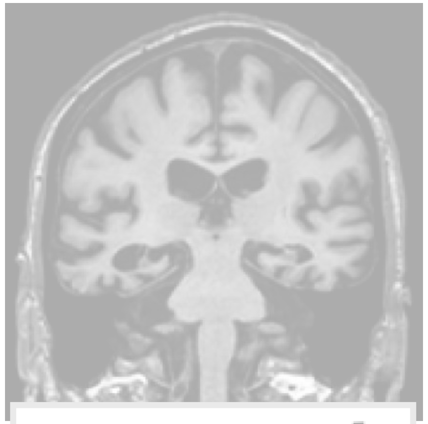
$p(\mathbf{l}|\boldsymbol{\theta}_l)$

"anatomical model"

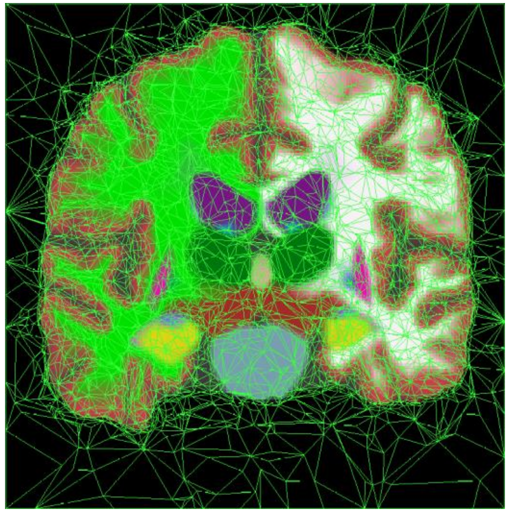
Label image $\mathbf{l}$

$p(\mathbf{d}|\mathbf{l}, \boldsymbol{\theta}_d)$

"imaging model"



MRI image $\mathbf{d}$



– The parameters $\boldsymbol{\theta}_l$ are the location of mesh nodes in an atlas

– Deformation prior $p(\boldsymbol{\theta}_l)$

[Van Leemput TMI 2009]

Imaging model

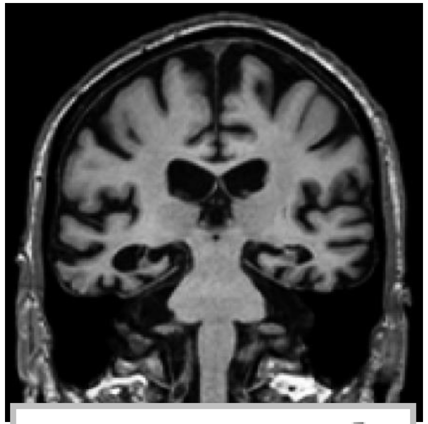
$p(\mathbf{l}|\boldsymbol{\theta}_l)$

"anatomical model"

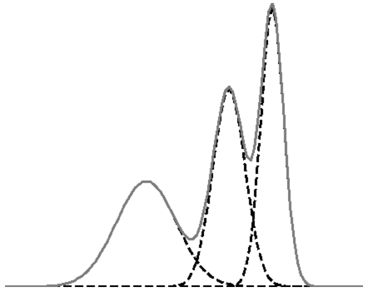
Label image $\mathbf{l}$

$p(\mathbf{d}|\mathbf{l}, \boldsymbol{\theta}_d)$

"imaging model"

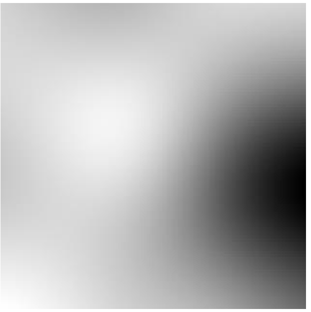


MRI image $\mathbf{d}$



Gaussian mixture model

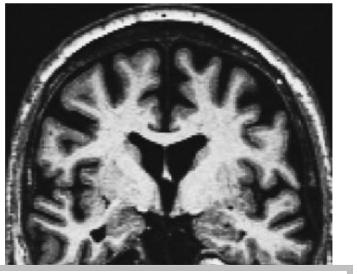
+



parametric bias field model

http://freesurfer.net/

"Inverting" the model

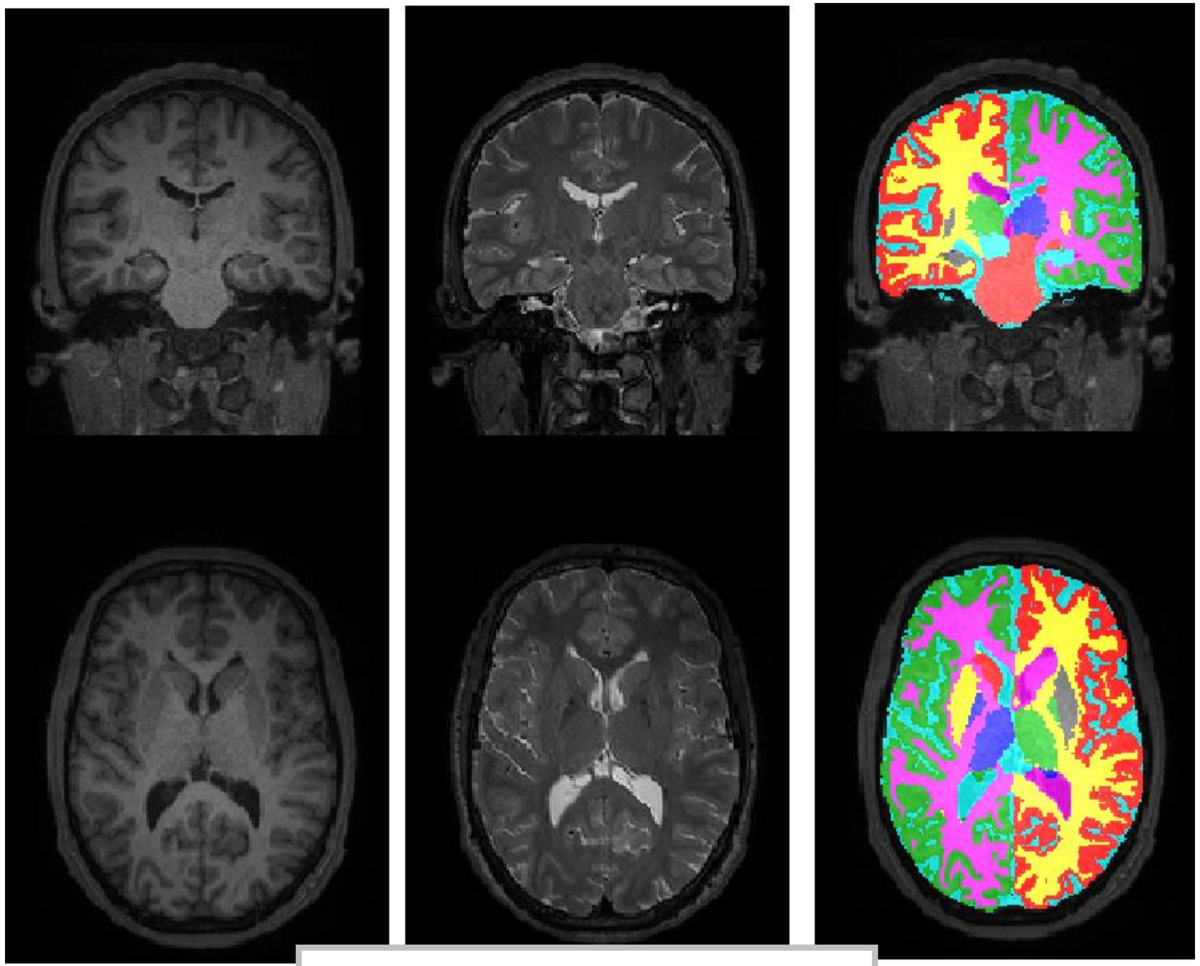


1.5T T1-weighted scans

with manual delineations

(20 subjects)

~10min (CPU)

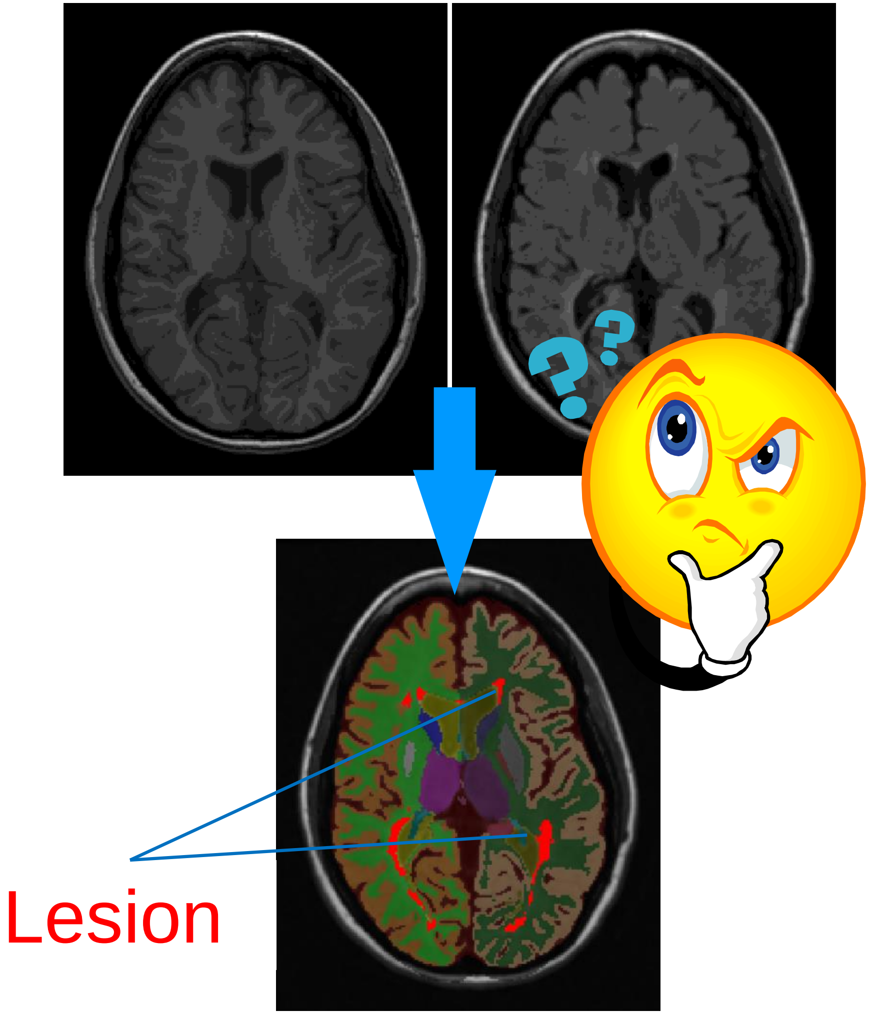


3T multi-contrast scan

Training data

[Puonti, NeuroImage 2016]

Adding pathology into the mix…



- Tissue contrast depends on hardware and pulse sequence
- No standard protocol for clinical imaging (multimodal)

- **Additional difficulty:** Disease-specific lesions!

Idea: alter the generative model

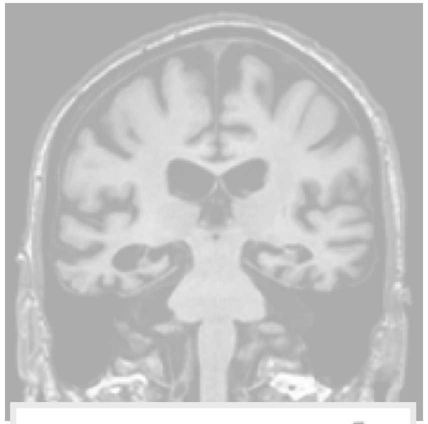
$p(\mathbf{l}|\boldsymbol{\theta}_l)$

"anatomical model"

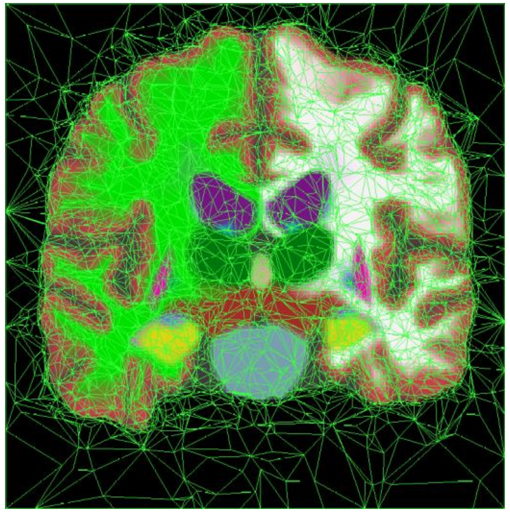
Label image $\mathbf{l}$

$p(\mathbf{d}|\mathbf{l}, \boldsymbol{\theta}_d)$

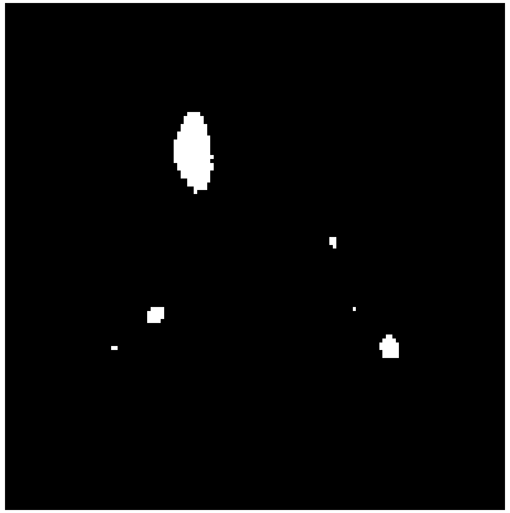
"imaging model"



MRI image $\mathbf{d}$



+

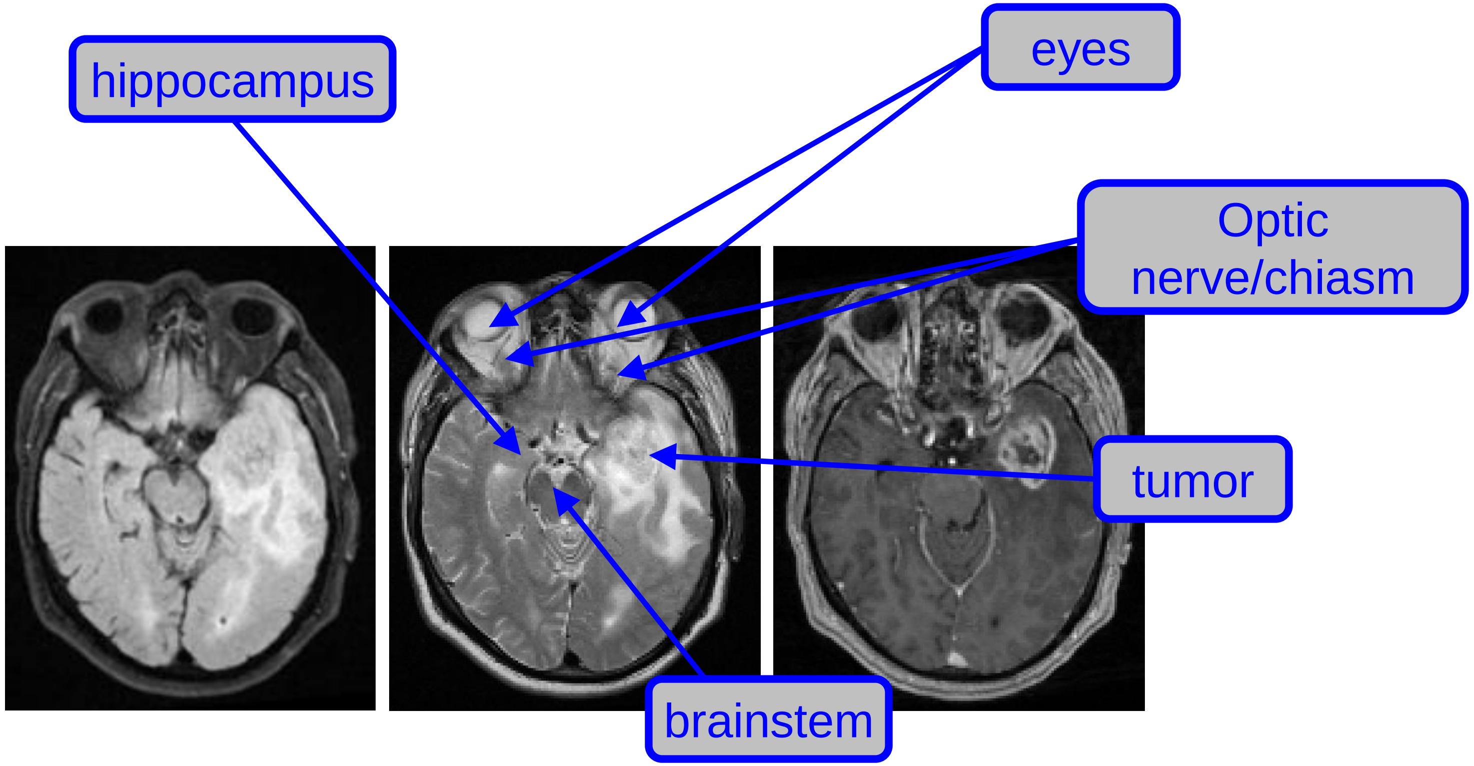


Use two anatomical models that are trained on different subjects!

Radiation therapy planning

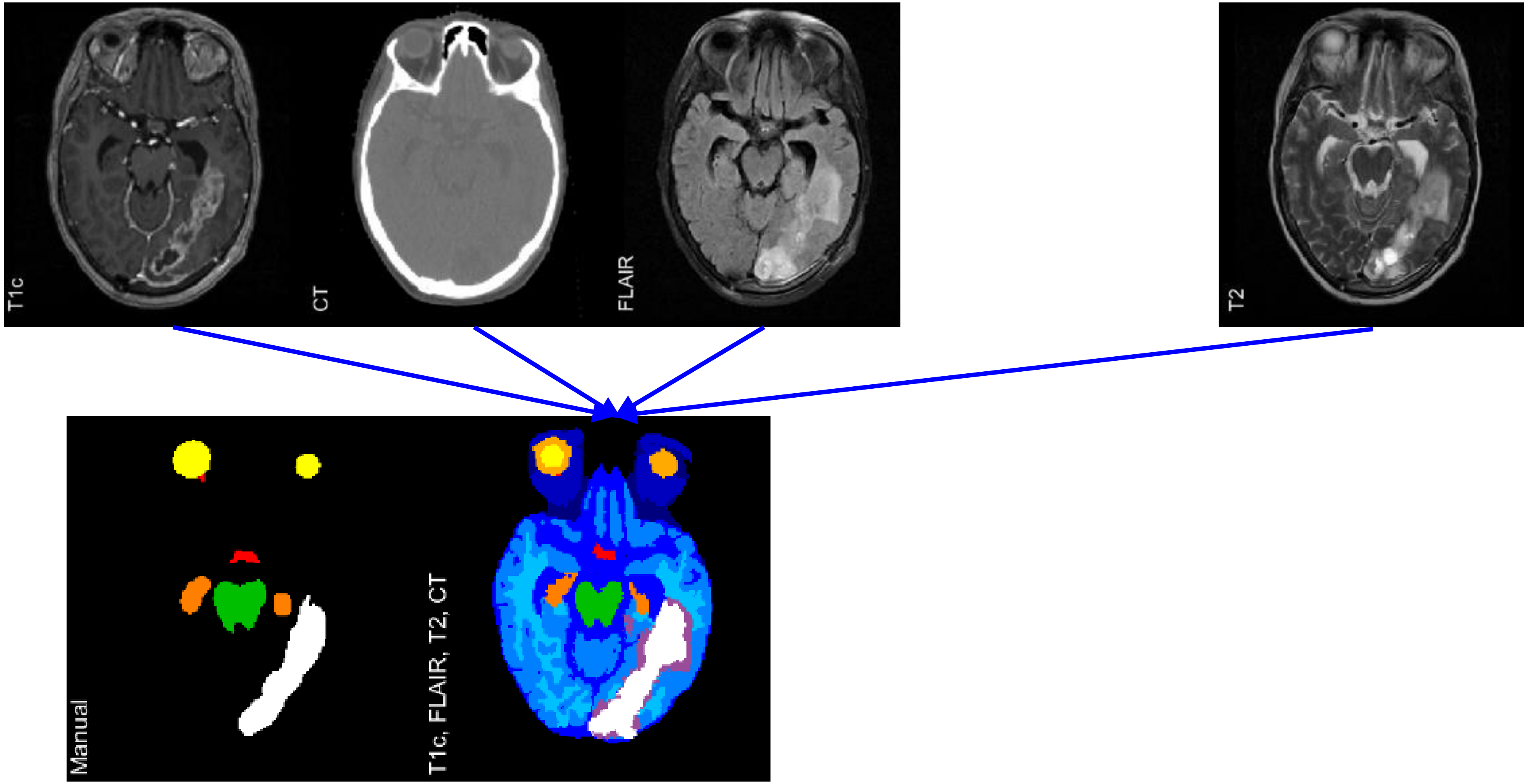
– Radiate a brain tumor while sparing organs-at-risk

– Need to know where these structures are



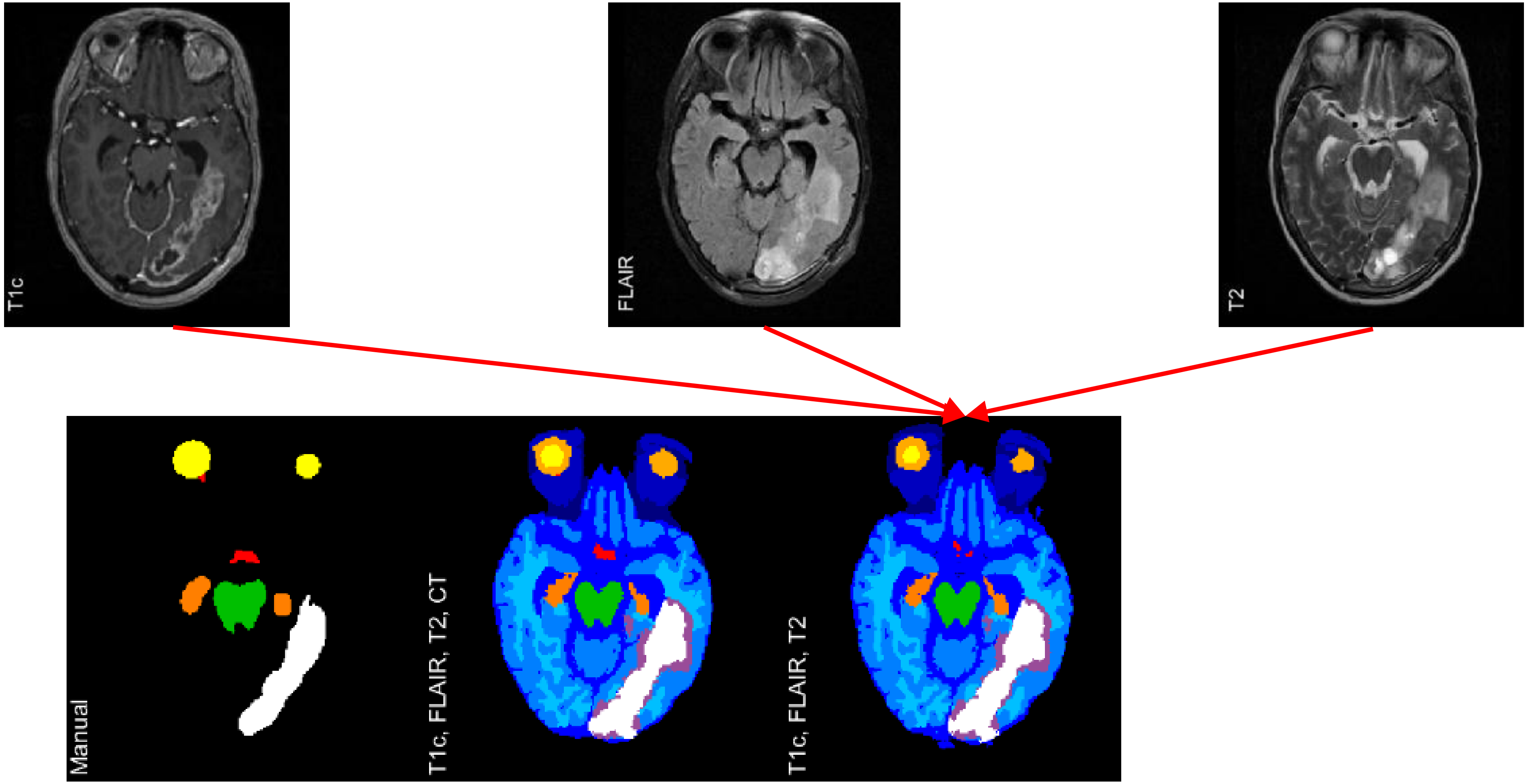
[Agn et al., Medical Image Analysis 2019]

Radiation therapy planning



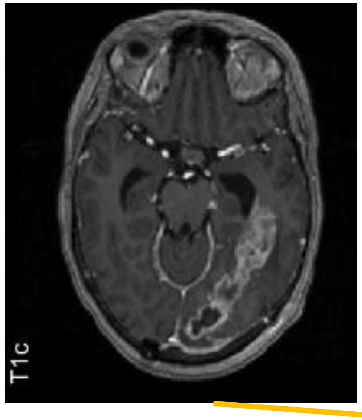
[Agn et al., Medical Image Analysis 2019]

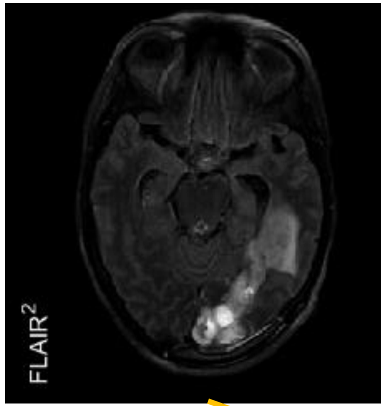
Radiation therapy planning

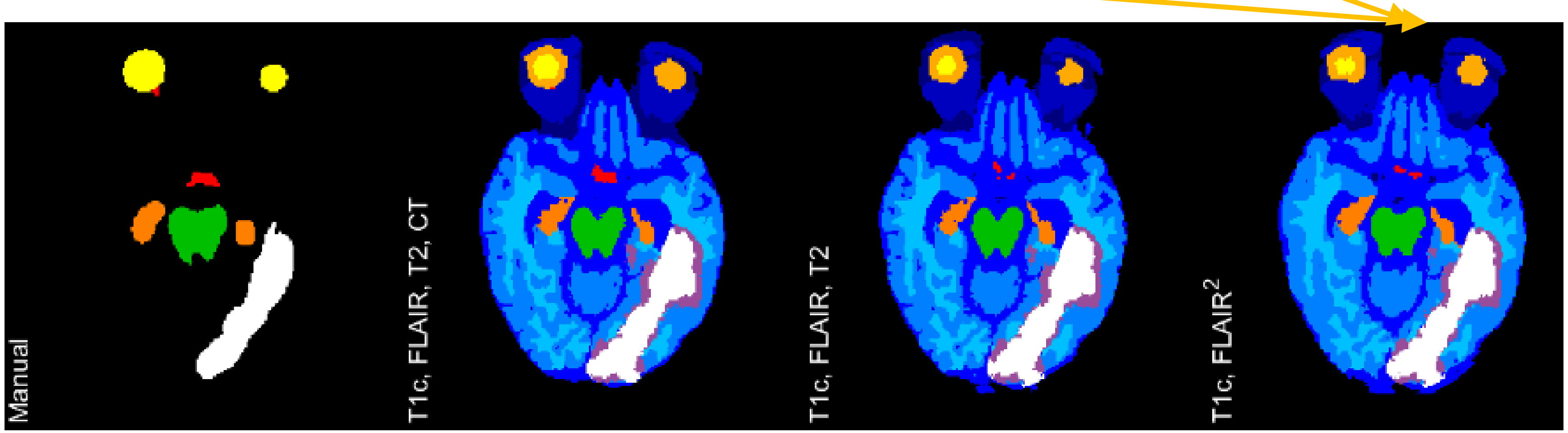


[Agn et al., Medical Image Analysis 2019]

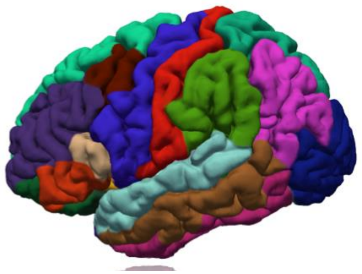
Radiation therapy planning





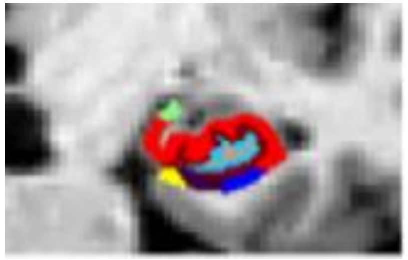


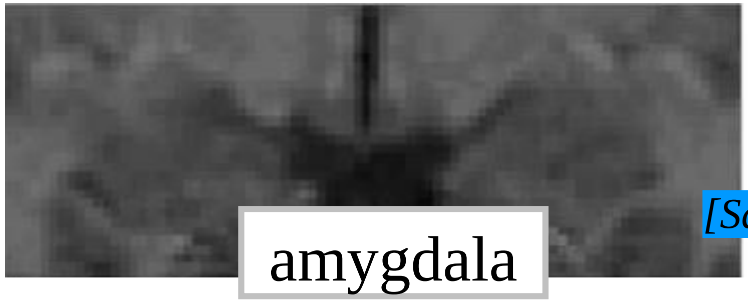
Tiny brain structures



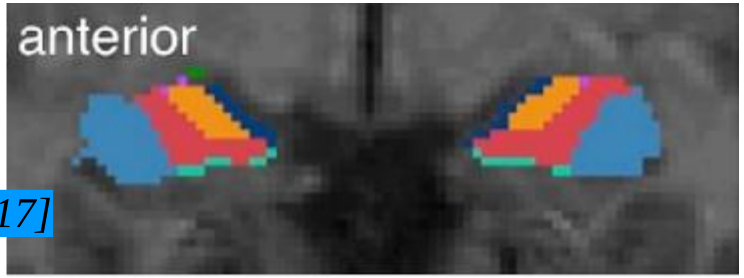
http://freesurfer.net/

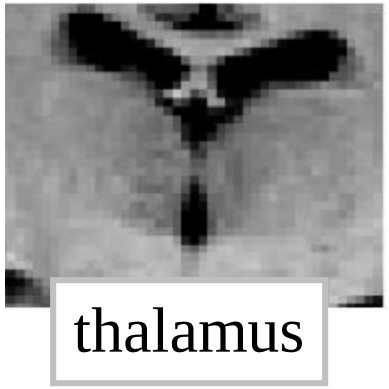
[Iglesias NeuroImage 2015]



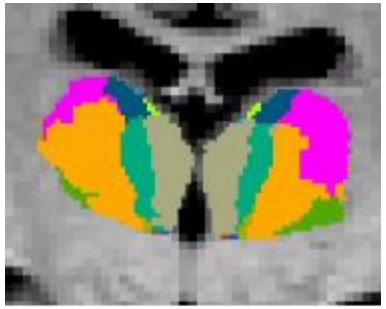


[Saygin NeuroImage 2017]

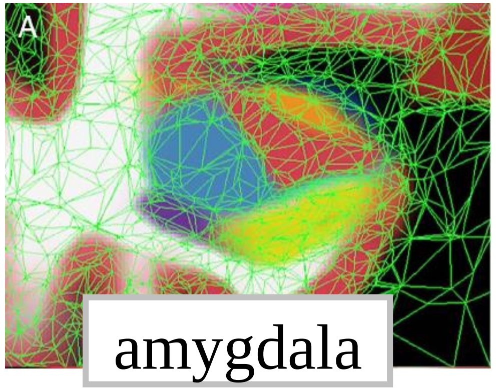


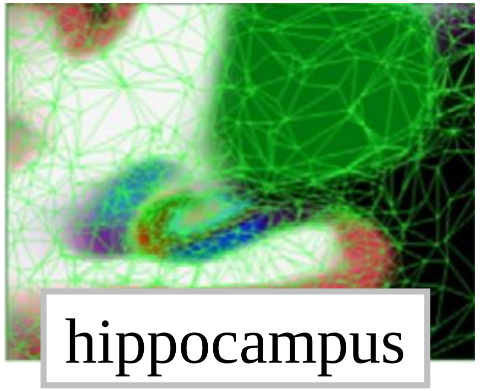


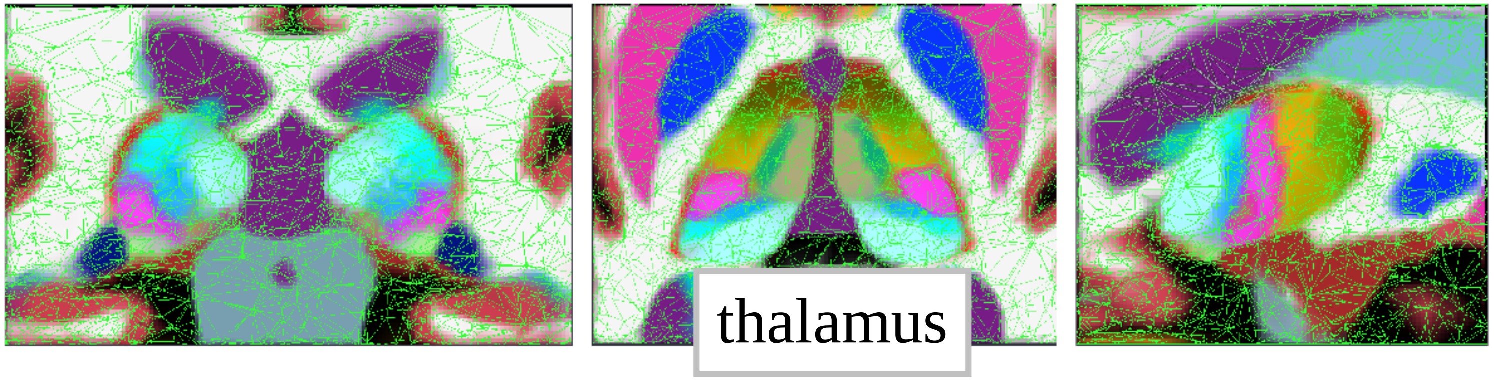
[Iglesias NeuroImage 2018]



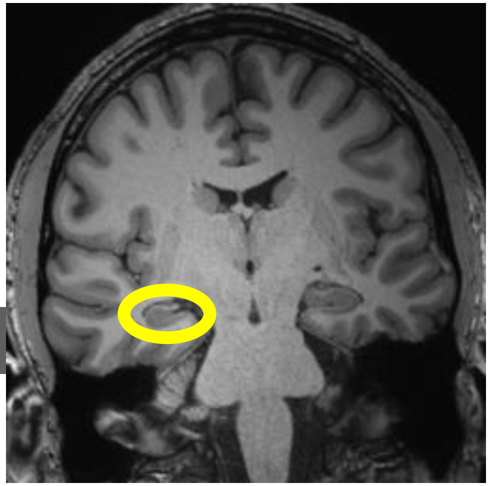
Segmentation method same as before

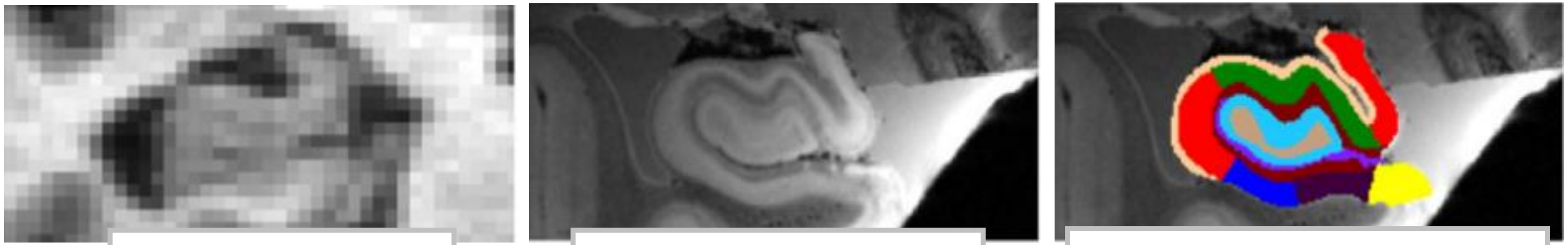






Discriminative methods?





in vivo
1mm isotropic

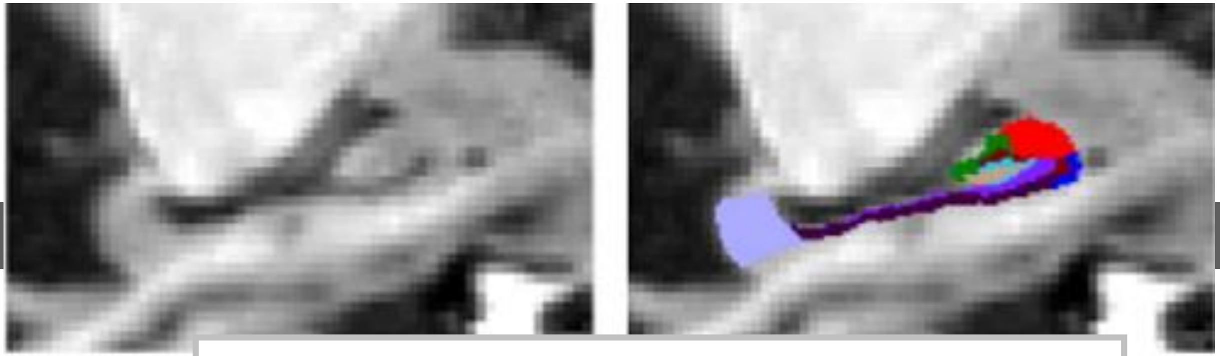
ex vivo
0.13mm isotropic

ex vivo
expert segmentation

Manual delineations on high-resolution *ex vivo* scans:

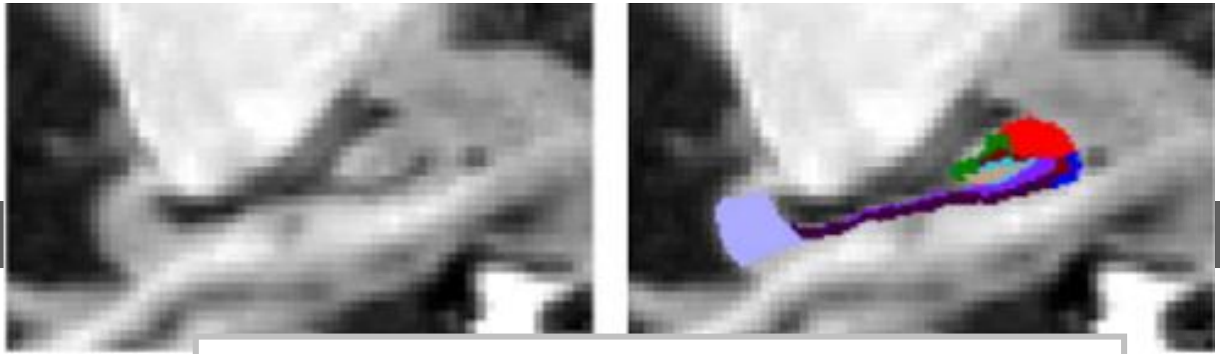
- Intrinsic MRI contrast changes
- Only ~10 training subjects

[Iglesias et al., NeuroImage 2015]

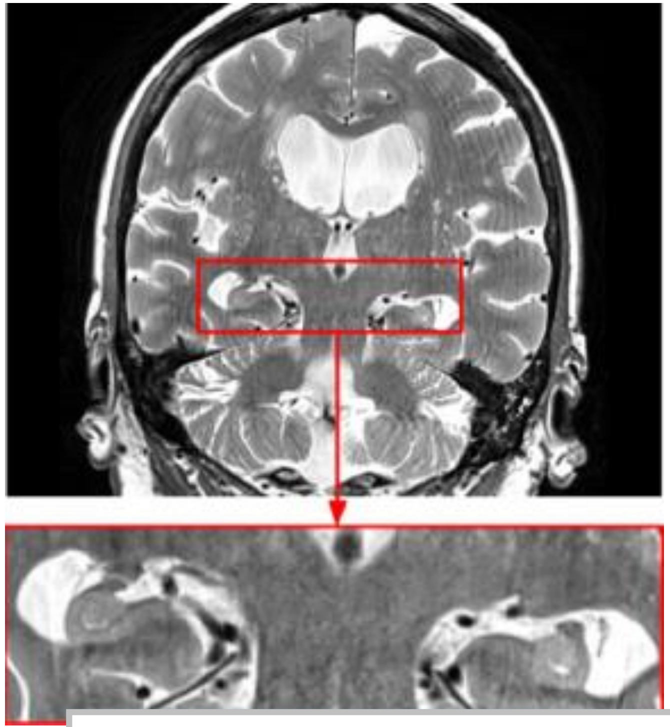


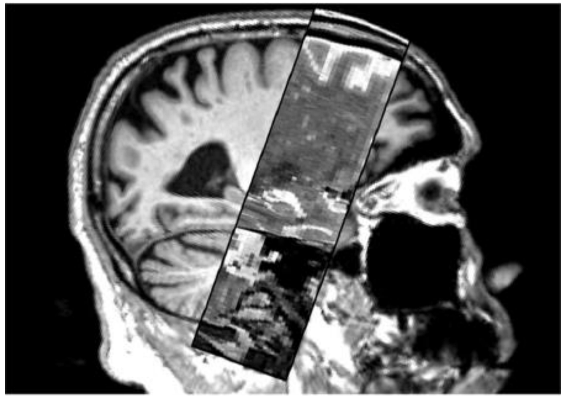
T1w only (1mm isotropic)

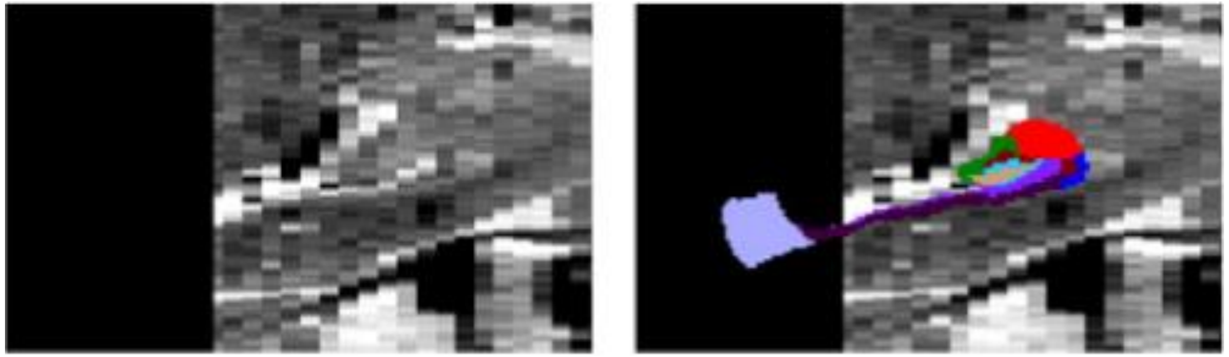
[Iglesias et al., NeuroImage 2015]



T1w only (1mm isotropic)







Additional T2w (0.38 x 0.38 x 0.8 mm)

Overview

Uncertainty on input using generative models

Uncertainty on output using Monte Carlo sampling

Discussion and conclusion

Uncertainty papers from me/my group…

variational approximation

- Mean-field approximation of MRFs (TMI 1999)
- Metropolis sampling for partial volume models (TMI 2003)
- HMC for visualizing atlas warp priors (TMI 2009)
- HMC + Gibbs sampling for segmentation uncertainty (MEDIA 2013)
- Gibbs sampling for analyzing STAPLE (MICCAI 2014)
- HMC for PET reconstruction uncertainty (MICCAI 2015)
- Gibbs sampling for radiotherapy segmentations (MEDIA 2019)
- Gibbs sampling for MI-based registration uncertainty (UNSURE 2019)

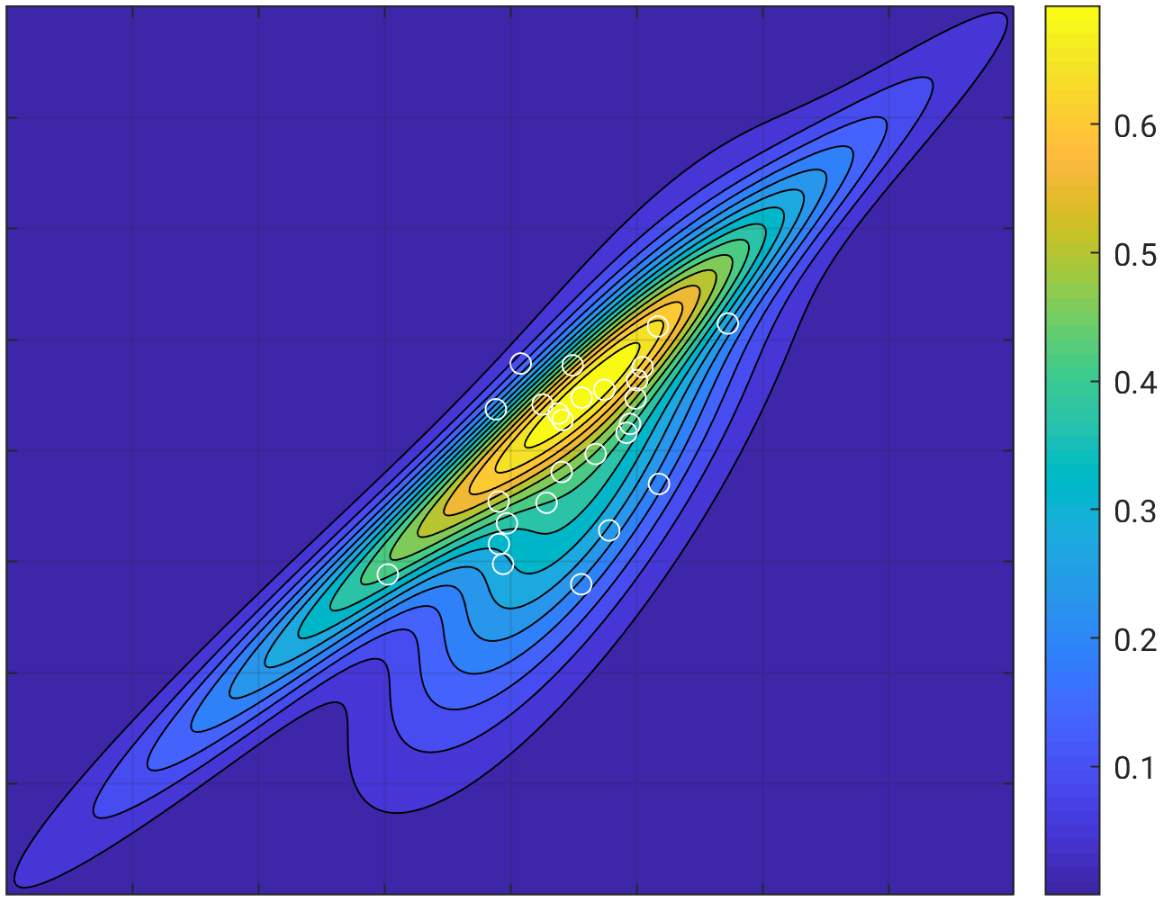
Monte Carlo sampling

Monte Carlo (MC) sampling

- Obtain samples of $p(\mathbf{x})$: $\{\mathbf{x}^{(s)}\}_{s=1}^{S}$
  - Samples from marginal distributions are obtained automatically
  - Integrals over nuisance variables automagically "disappear"!

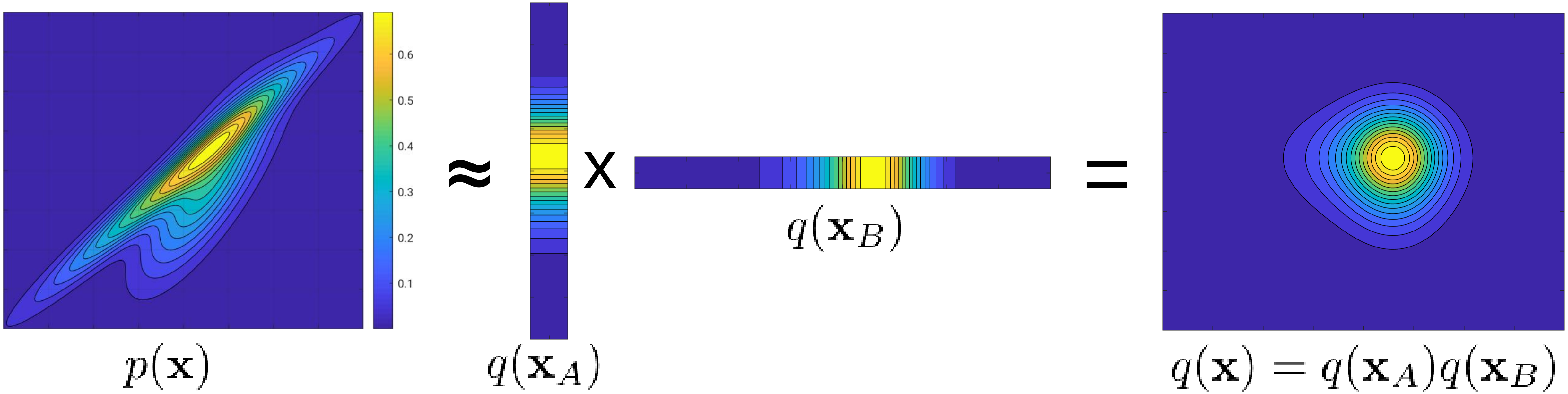
- Approximate expectations:

$$\mathbb{E}[\phi(\mathbf{x})] = \int_{\mathbf{x}} \phi(\mathbf{x}) p(\mathbf{x})\, \mathrm{d}\mathbf{x}$$
$$\simeq \frac{1}{S} \sum_{s=1}^{S} \phi(\mathbf{x}^{(s)})$$



Why Monte Carlo sampling?

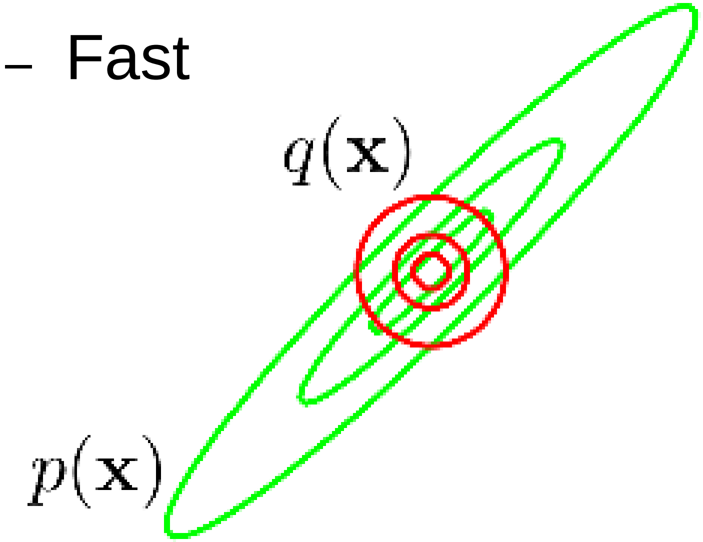
– Variational methods (mean-field approximation): $\mathbf{x} = (\mathbf{x}_A^{\mathrm{T}}, \mathbf{x}_B^{\mathrm{T}})^{\mathrm{T}}$



– Minimize the Kullback-Leibler divergence between $q(\mathbf{x})$ and $p(\mathbf{x})$

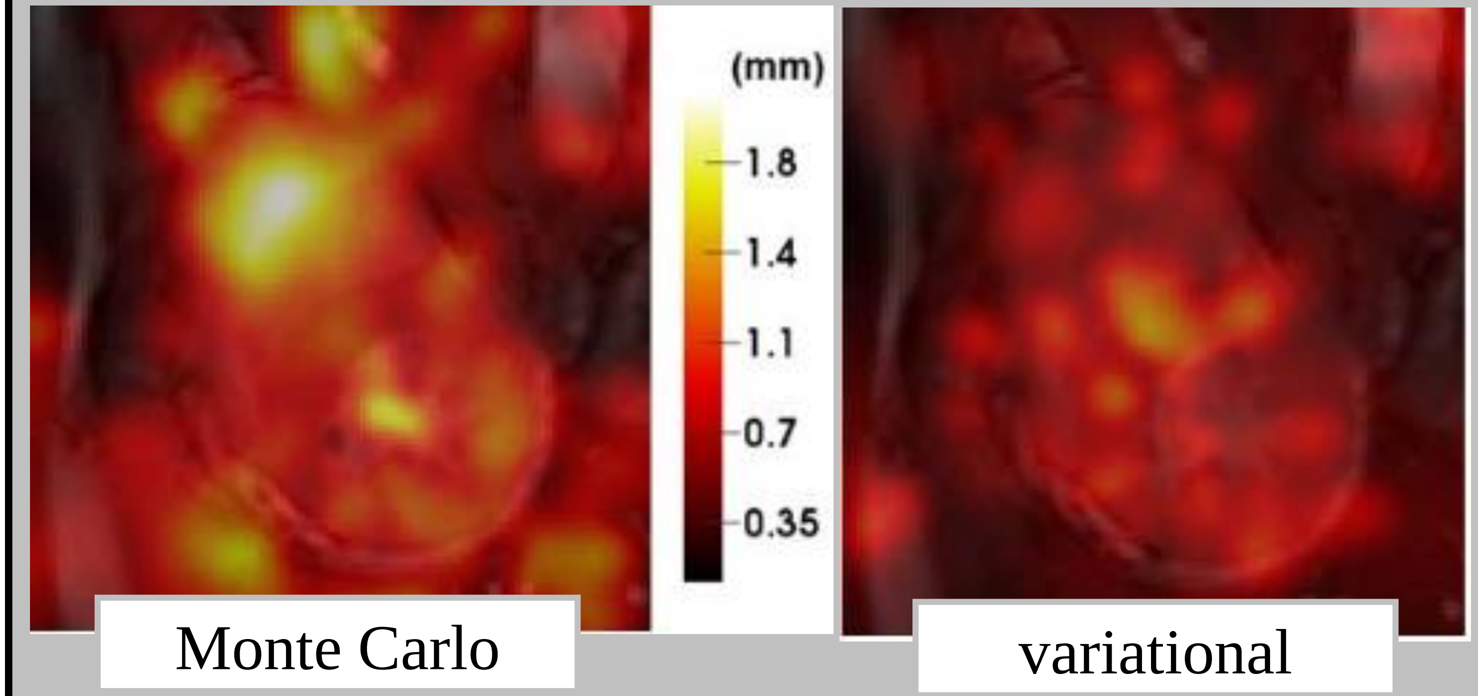
Variational approximations (mean-field)?

– Fast



[Bishop PRML 2016]

Registration uncertainty *[Le Folgoc TMI 2017]*



– Can be hugely inaccurate (for strong dependencies)

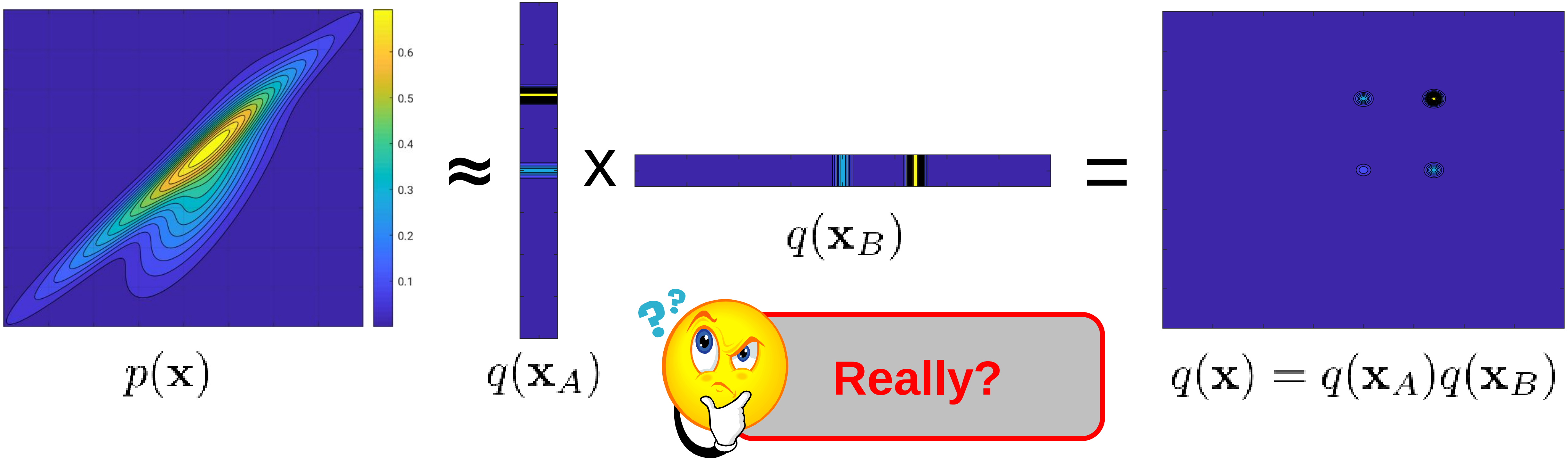
– No red flags raised

- OK for increasing e.g., Dice scores

- Dubious for anything else

BTW: dropout at test time?

- Randomly remove some units in a neural network *[Gal & Ghahramani ICML 2016]*
- Popular in medical image analysis



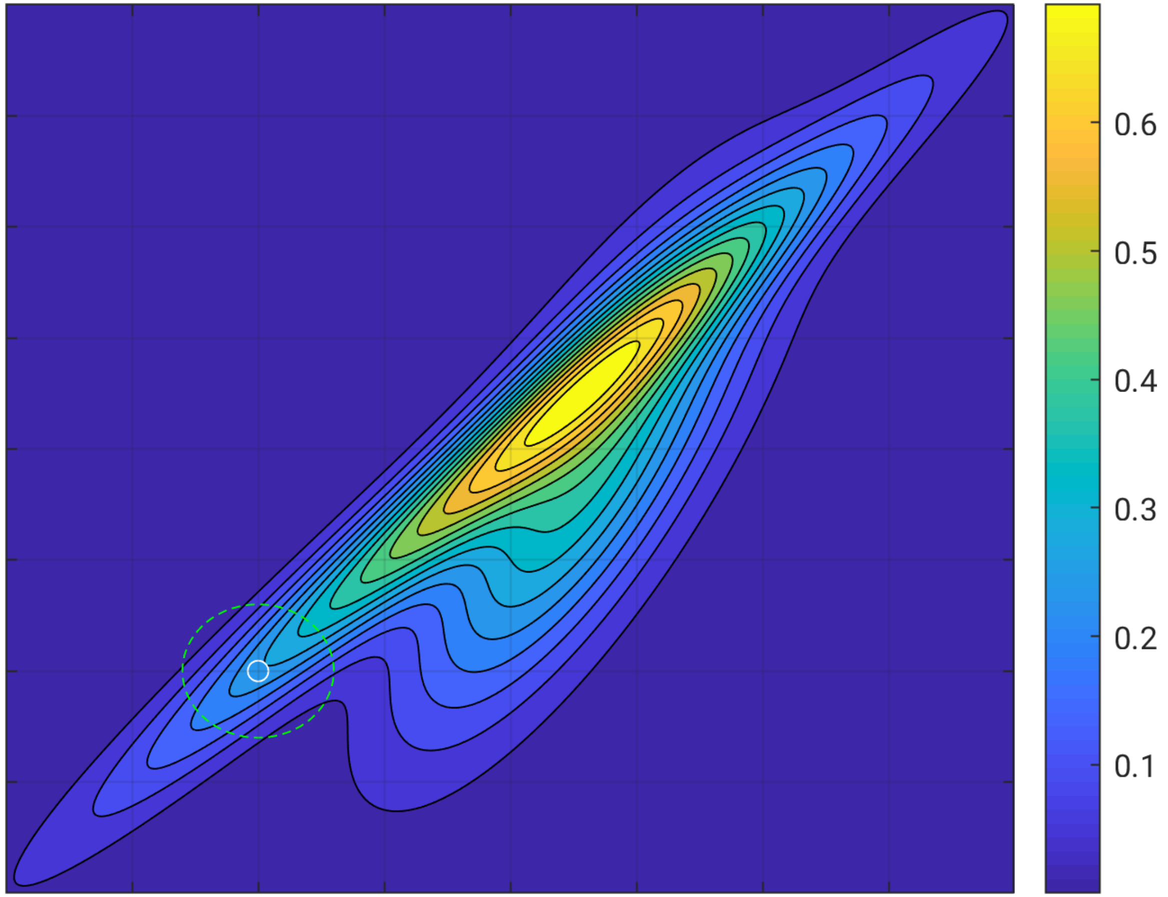
Monte Carlo sampling in high dimensions

Markov chain Monte Carlo (MCMC):

- Propose a perturbation $\mathbf{x}^*$ of a current state $\mathbf{x}^{(\tau)}$ using a proposal distribution $q(\mathbf{x}^*|\mathbf{x}^{(\tau)})$
- Accept $\mathbf{x}^*$ with probability

$$\min\left(1, \frac{p(\mathbf{x}^*)q(\mathbf{x}^{(\tau)}|\mathbf{x}^*)}{p(\mathbf{x}^{(\tau)})q(\mathbf{x}^*|\mathbf{x}^{(\tau)})}\right)$$

[Metropolis-Hastings]



Monte Carlo sampling in high dimensions

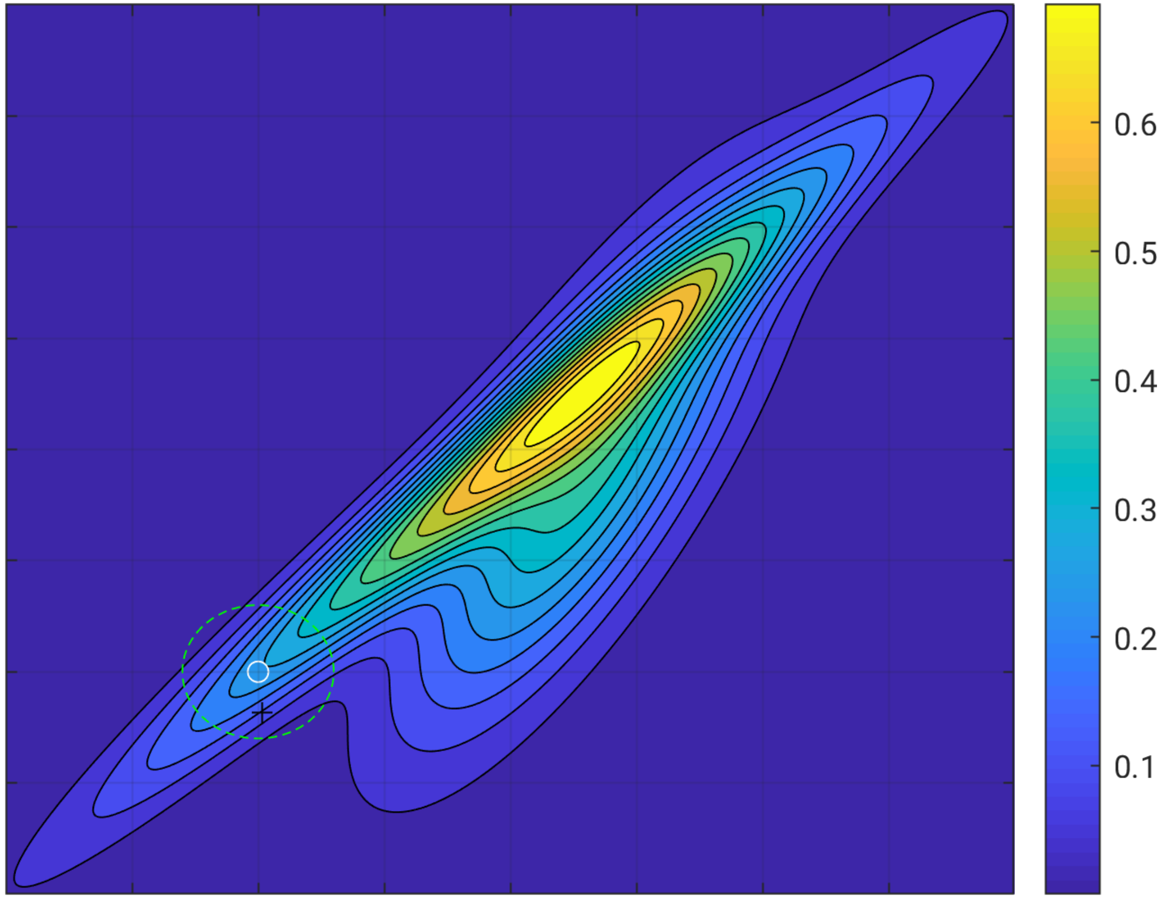
Markov chain Monte Carlo (MCMC):

- Propose a perturbation $\mathbf{x}^*$ of a current state $\mathbf{x}^{(\tau)}$ using a proposal distribution $q(\mathbf{x}^*|\mathbf{x}^{(\tau)})$

- Accept $\mathbf{x}^*$ with probability

$$\min\left(1, \frac{p(\mathbf{x}^*)q(\mathbf{x}^{(\tau)}|\mathbf{x}^*)}{p(\mathbf{x}^{(\tau)})q(\mathbf{x}^*|\mathbf{x}^{(\tau)})}\right)$$

[Metropolis-Hastings]



Monte Carlo sampling in high dimensions

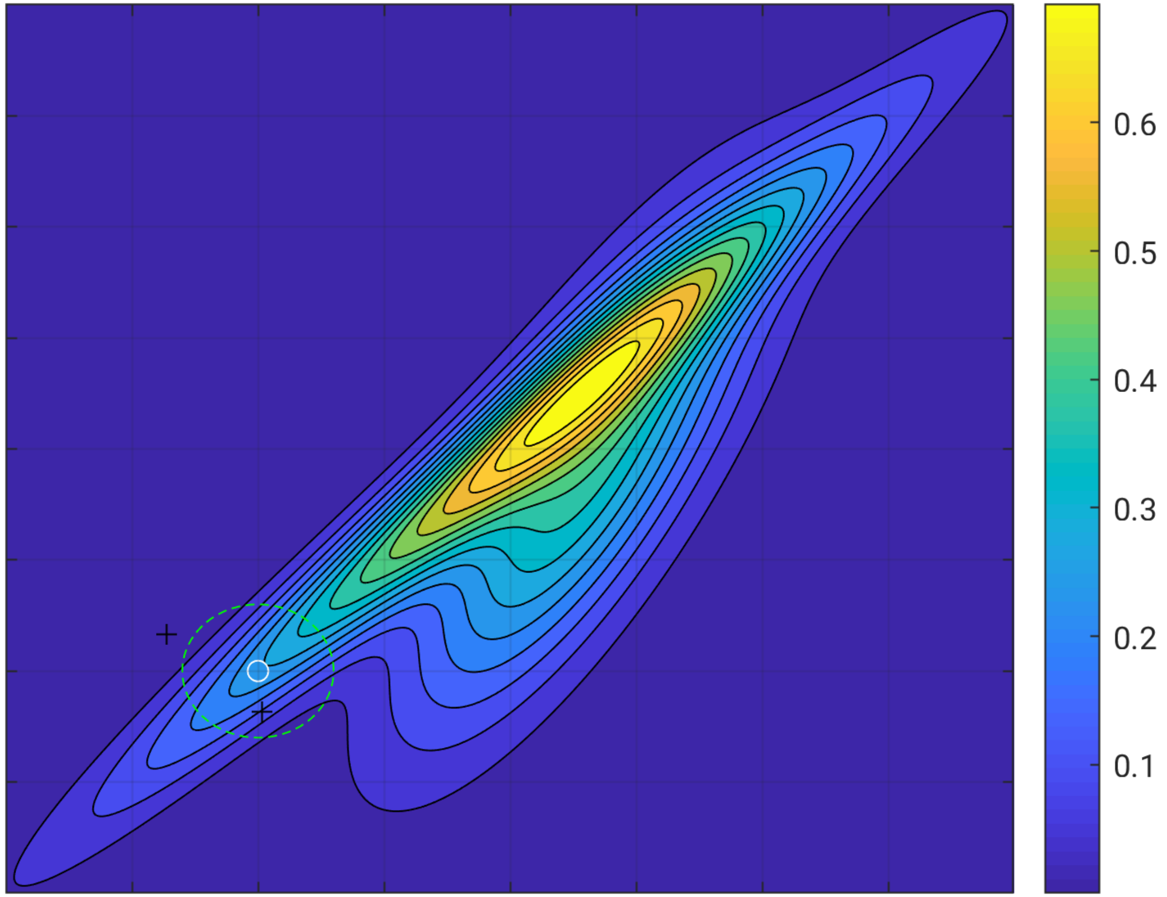
Markov chain Monte Carlo (MCMC):

- Propose a perturbation $\mathbf{x}^*$ of a current state $\mathbf{x}^{(\tau)}$ using a proposal distribution $q(\mathbf{x}^*|\mathbf{x}^{(\tau)})$

- Accept $\mathbf{x}^*$ with probability

$$\min\left(1, \frac{p(\mathbf{x}^*)q(\mathbf{x}^{(\tau)}|\mathbf{x}^*)}{p(\mathbf{x}^{(\tau)})q(\mathbf{x}^*|\mathbf{x}^{(\tau)})}\right)$$

[Metropolis-Hastings]



Monte Carlo sampling in high dimensions

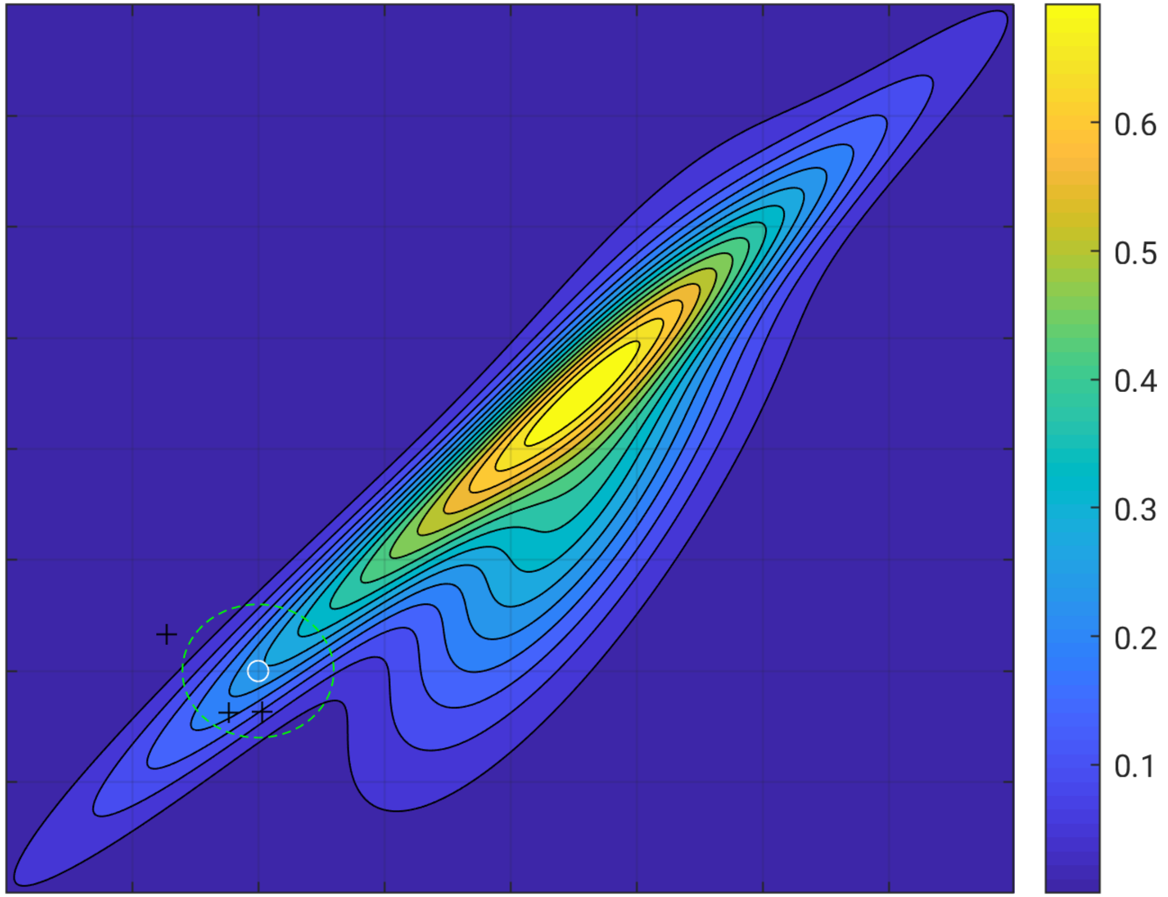
Markov chain Monte Carlo (MCMC):

- Propose a perturbation $\mathbf{x}^*$ of a current state $\mathbf{x}^{(\tau)}$ using a proposal distribution $q(\mathbf{x}^*|\mathbf{x}^{(\tau)})$

- Accept $\mathbf{x}^*$ with probability

$$\min\left(1, \frac{p(\mathbf{x}^*)q(\mathbf{x}^{(\tau)}|\mathbf{x}^*)}{p(\mathbf{x}^{(\tau)})q(\mathbf{x}^*|\mathbf{x}^{(\tau)})}\right)$$

[Metropolis-Hastings]



Monte Carlo sampling in high dimensions

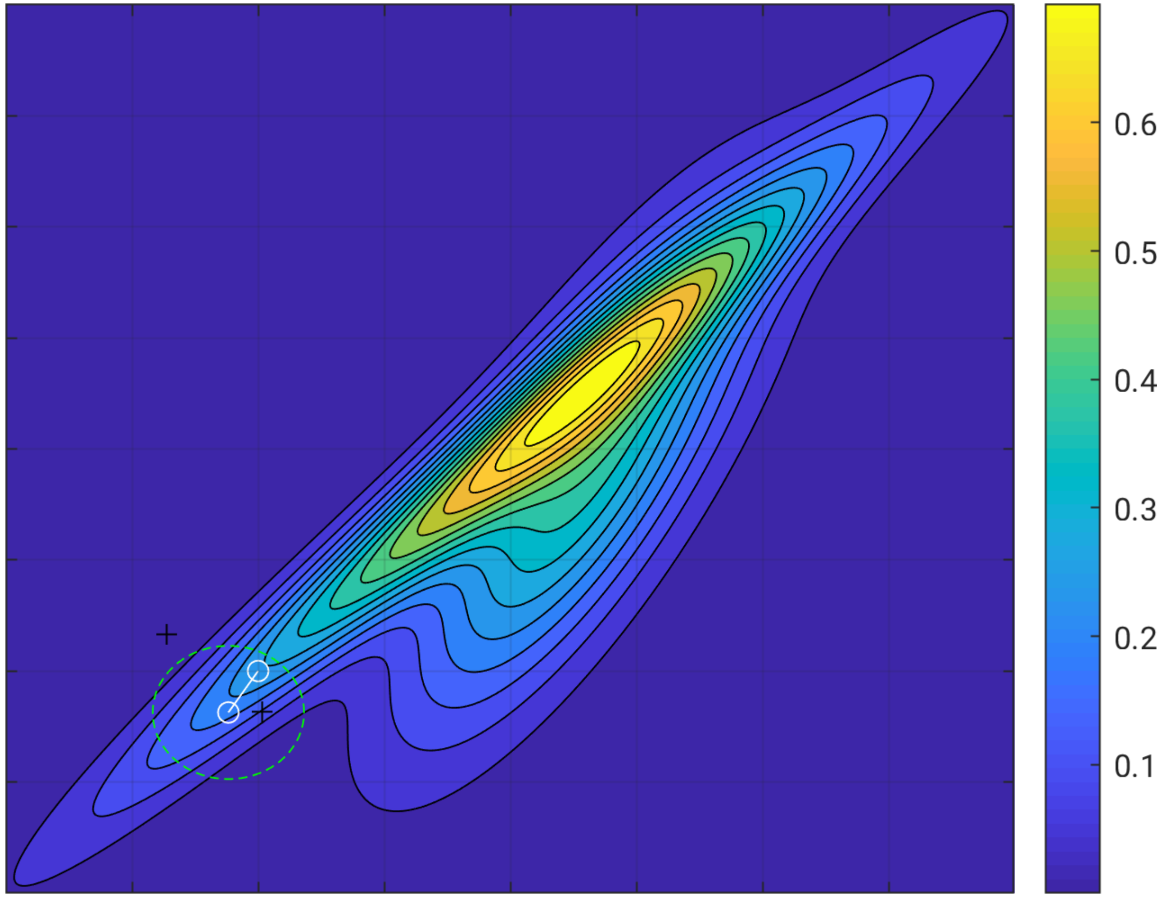
Markov chain Monte Carlo (MCMC):

- Propose a perturbation $\mathbf{x}^*$ of a current state $\mathbf{x}^{(\tau)}$ using a proposal distribution $q(\mathbf{x}^*|\mathbf{x}^{(\tau)})$

- Accept $\mathbf{x}^*$ with probability

$$\min\left(1, \frac{p(\mathbf{x}^*)q(\mathbf{x}^{(\tau)}|\mathbf{x}^*)}{p(\mathbf{x}^{(\tau)})q(\mathbf{x}^*|\mathbf{x}^{(\tau)})}\right)$$

[Metropolis-Hastings]



Monte Carlo sampling in high dimensions

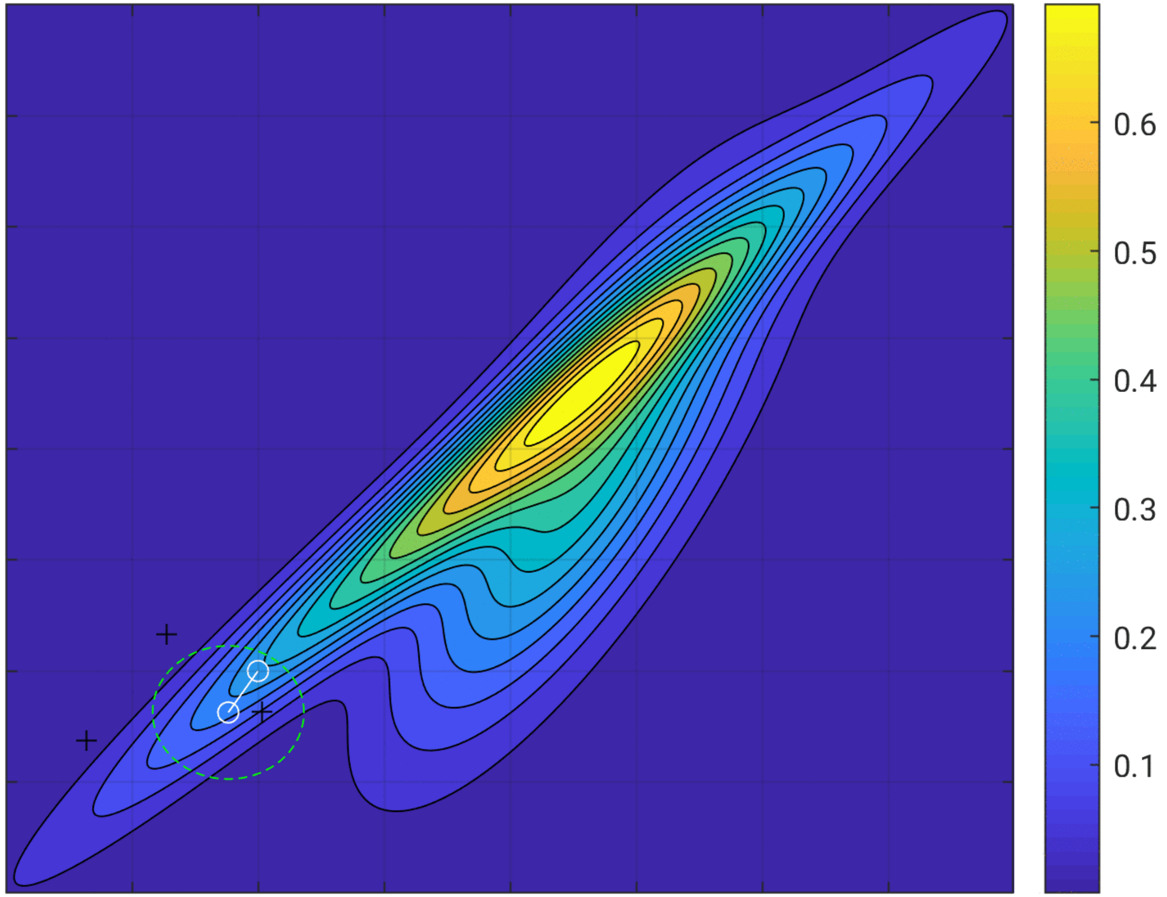
Markov chain Monte Carlo (MCMC):

- Propose a perturbation $\mathbf{x}^*$ of a current state $\mathbf{x}^{(\tau)}$ using a proposal distribution $q(\mathbf{x}^*|\mathbf{x}^{(\tau)})$

- Accept $\mathbf{x}^*$ with probability

$$\min\left(1, \frac{p(\mathbf{x}^*)q(\mathbf{x}^{(\tau)}|\mathbf{x}^*)}{p(\mathbf{x}^{(\tau)})q(\mathbf{x}^*|\mathbf{x}^{(\tau)})}\right)$$

[Metropolis-Hastings]



Monte Carlo sampling in high dimensions

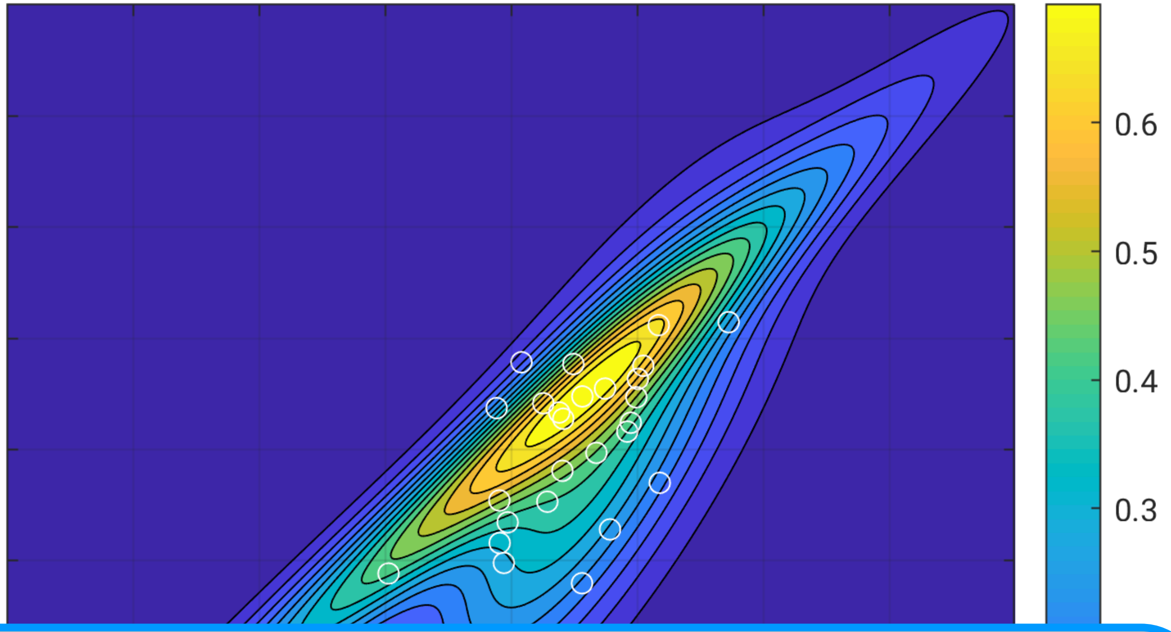
Markov chain Monte Carlo (MCMC):

– Propose a perturbation $\mathbf{x}^*$ of a current state $\mathbf{x}^{(\tau)}$ using a proposal distribution $q(\mathbf{x}^*|\mathbf{x}^{(\tau)})$

– Accept $\mathbf{x}^*$ with probability

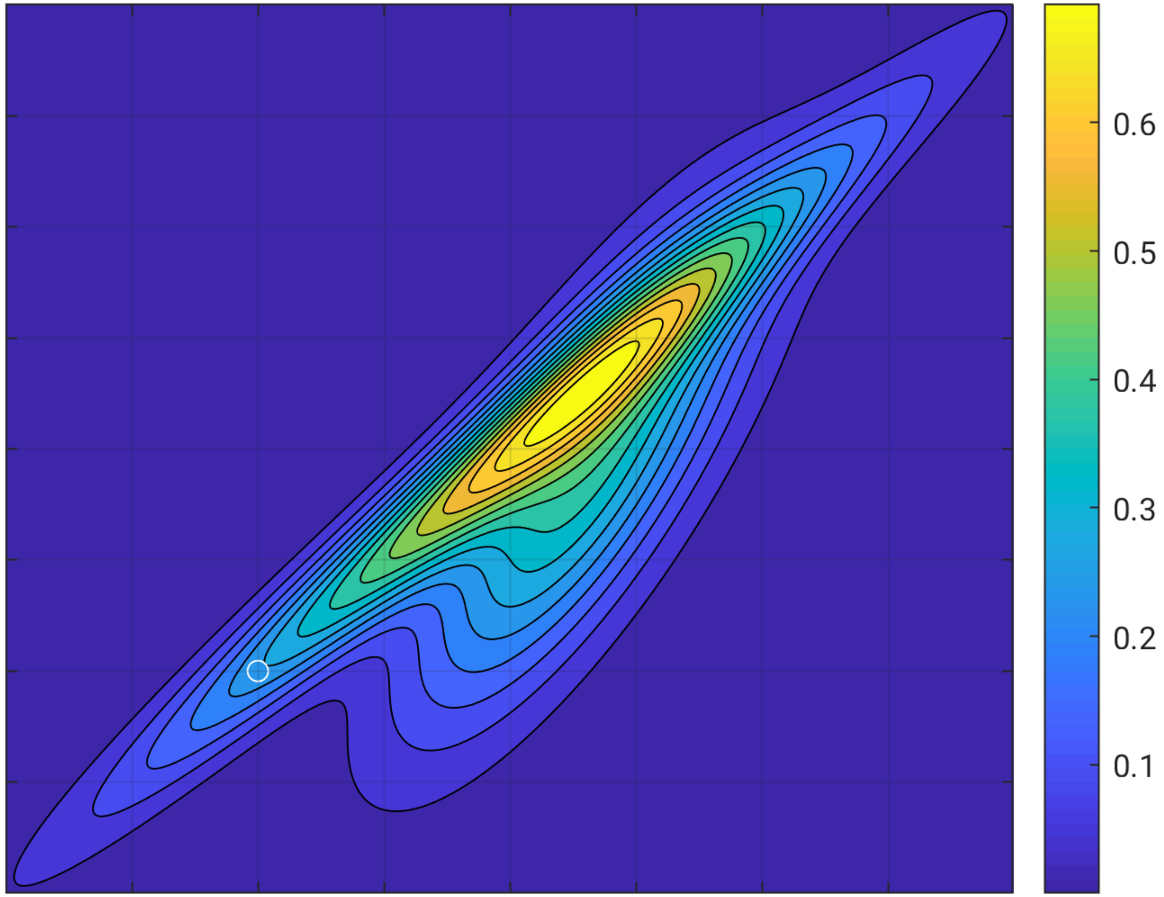
$$\min\left(1, \frac{p(\mathbf{x}^*)q(\mathbf{x}^{(\tau)}|\mathbf{x}^*)}{p(\mathbf{x}^{(\tau)})q(\mathbf{x}^*|\mathbf{x}^{(\tau)})}\right)$$

[Metropolis-Hastings]



- ***Great approximations***
- ***Dependencies can make things super slow***
- ***Burn-in period?***

Gibbs sampling

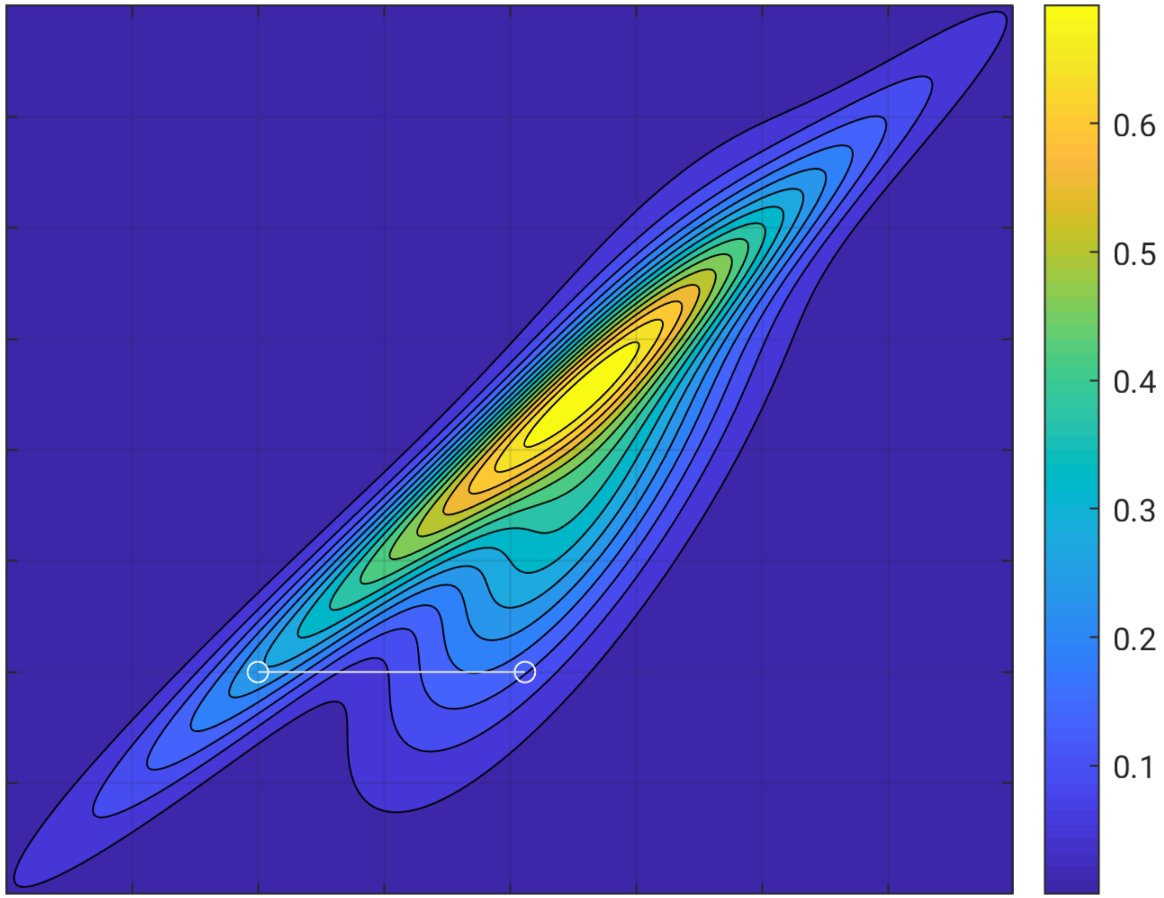


- Proposal distributions are conditional distributions:

$$
\begin{aligned}
\mathbf{x}_A^{(\tau+1)} &\sim p(\mathbf{x}_A | \mathbf{x}_B^{(\tau)}) \\
\mathbf{x}_B^{(\tau+1)} &\sim p(\mathbf{x}_B | \mathbf{x}_A^{(\tau+1)})
\end{aligned}
$$

- Proposals always accepted
- No tuning needed!
- Cf. optimization:
  - Coordinate descent
  - EM algorithm (95% same code!)

Gibbs sampling

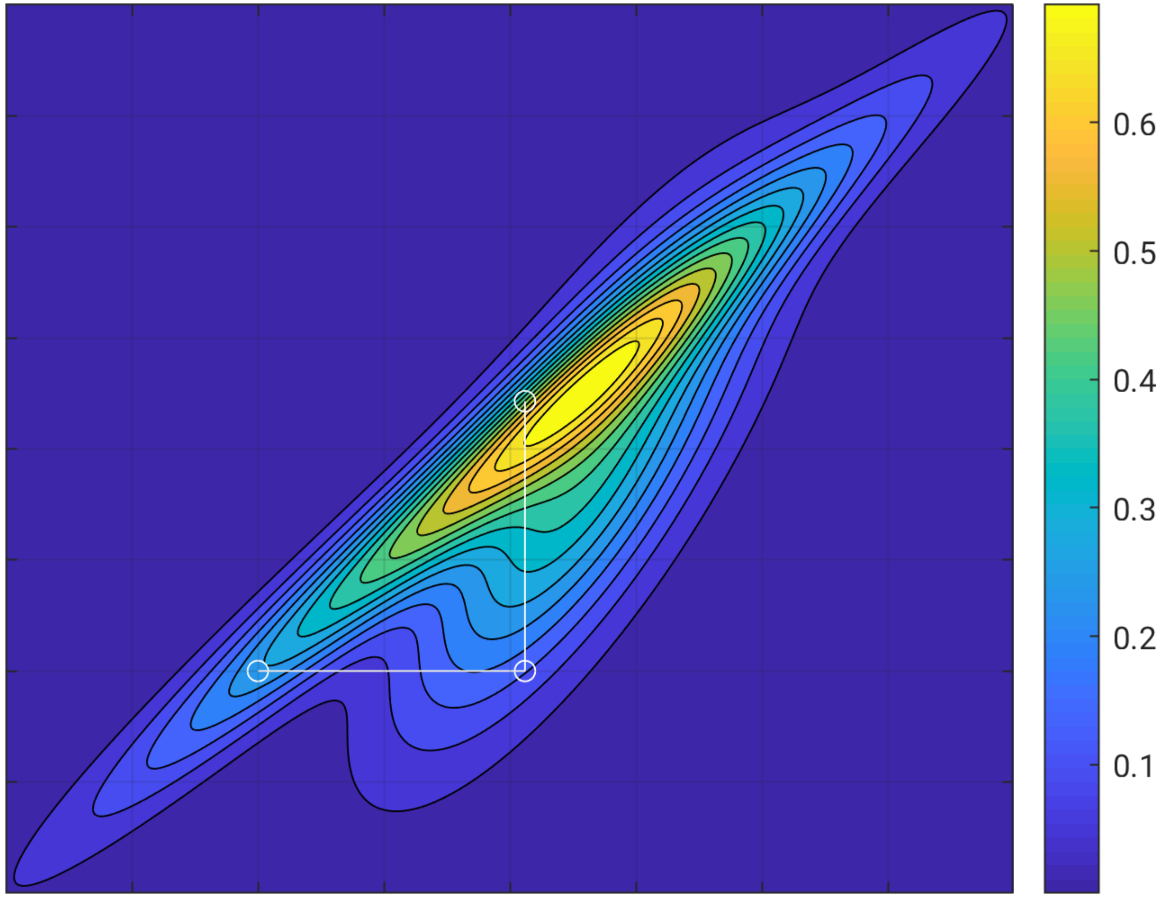


- Proposal distributions are conditional distributions:

$$\begin{aligned}\mathbf{x}_A^{(\tau+1)} &\sim p(\mathbf{x}_A | \mathbf{x}_B^{(\tau)}) \\ \mathbf{x}_B^{(\tau+1)} &\sim p(\mathbf{x}_B | \mathbf{x}_A^{(\tau+1)})\end{aligned}$$

- Proposals always accepted
- No tuning needed!
- Cf. optimization:
  - Coordinate descent
  - EM algorithm (95% same code!)

Gibbs sampling

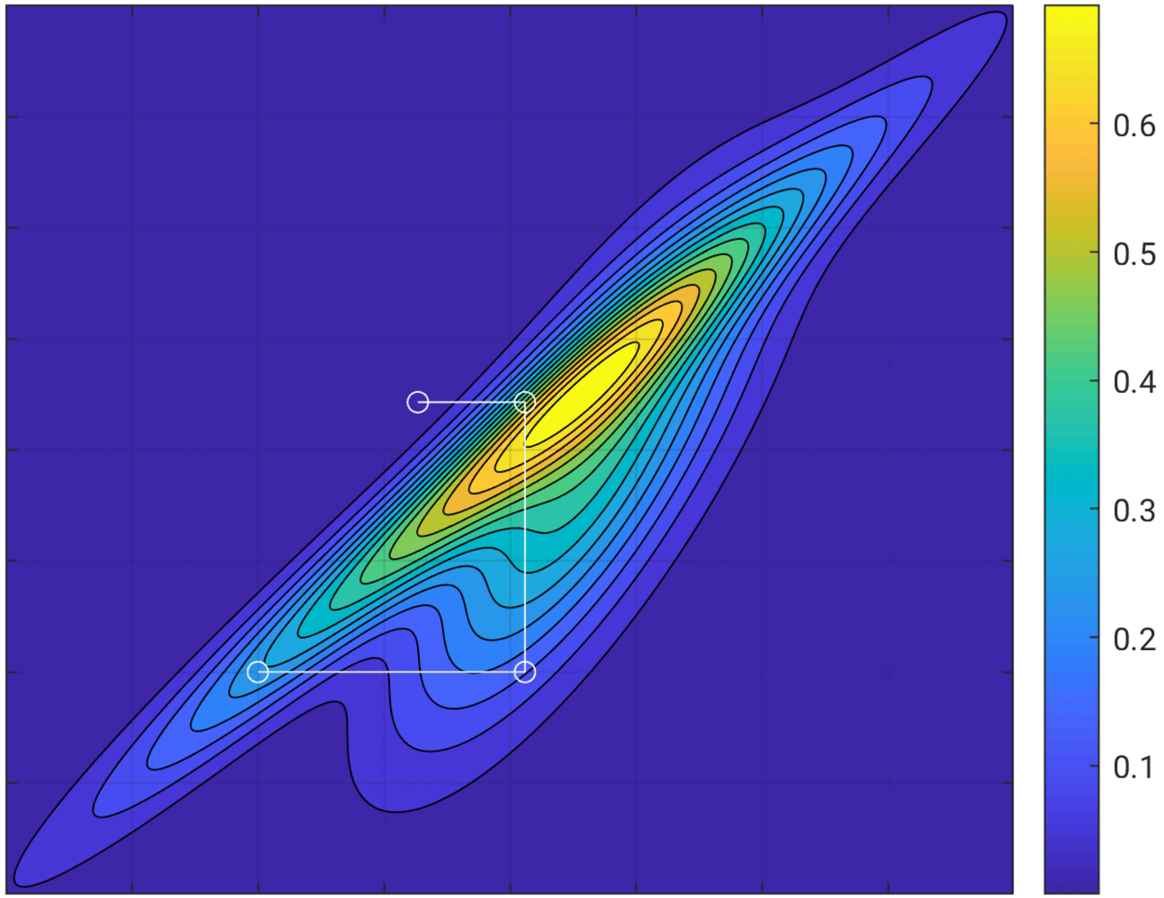


- Proposal distributions are conditional distributions:

$$\begin{aligned} \mathbf{x}_A^{(\tau+1)} &\sim p(\mathbf{x}_A|\mathbf{x}_B^{(\tau)}) \\ \mathbf{x}_B^{(\tau+1)} &\sim p(\mathbf{x}_B|\mathbf{x}_A^{(\tau+1)}) \end{aligned}$$

- Proposals always accepted
- No tuning needed!
- Cf. optimization:
  - Coordinate descent
  - EM algorithm (95% same code!)

Gibbs sampling



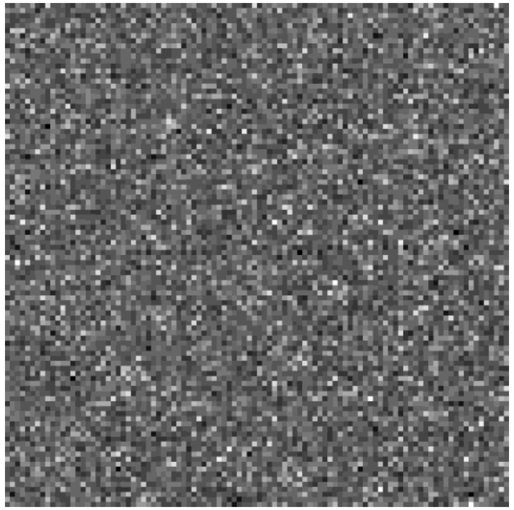
- Proposal distributions are conditional distributions:

$$\mathbf{x}_A^{(\tau+1)} \sim p(\mathbf{x}_A | \mathbf{x}_B^{(\tau)})$$
$$\mathbf{x}_B^{(\tau+1)} \sim p(\mathbf{x}_B | \mathbf{x}_A^{(\tau+1)})$$

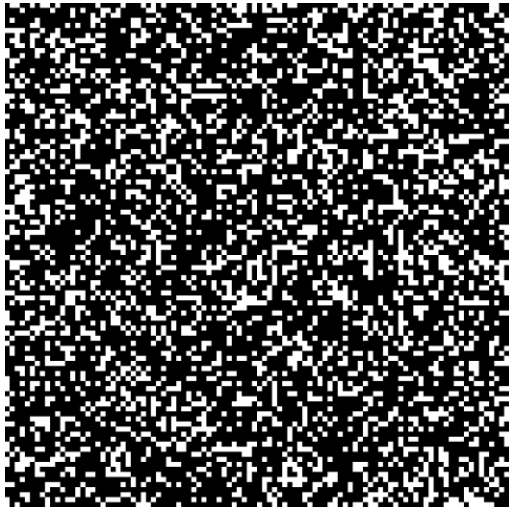
- Proposals always accepted
- No tuning needed!
- Cf. optimization:
  - Coordinate descent
  - EM algorithm (95% same code!)

Gibbs sampling: Gaussian mixture model

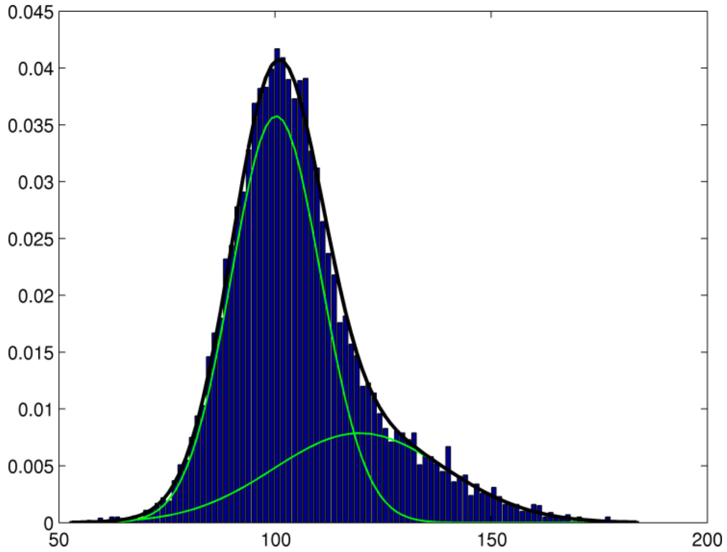
Sampling from $p(\mathbf{l}|\mathbf{d}) = \int_{\boldsymbol{\theta}} p(\mathbf{l}, \boldsymbol{\theta}|\mathbf{d})\, \mathrm{d}\boldsymbol{\theta}$



$\mathbf{d}$



$\mathbf{l}^{(\tau+1)} \sim p(\mathbf{l}|\mathbf{d}, \boldsymbol{\theta}^{(\tau)})$



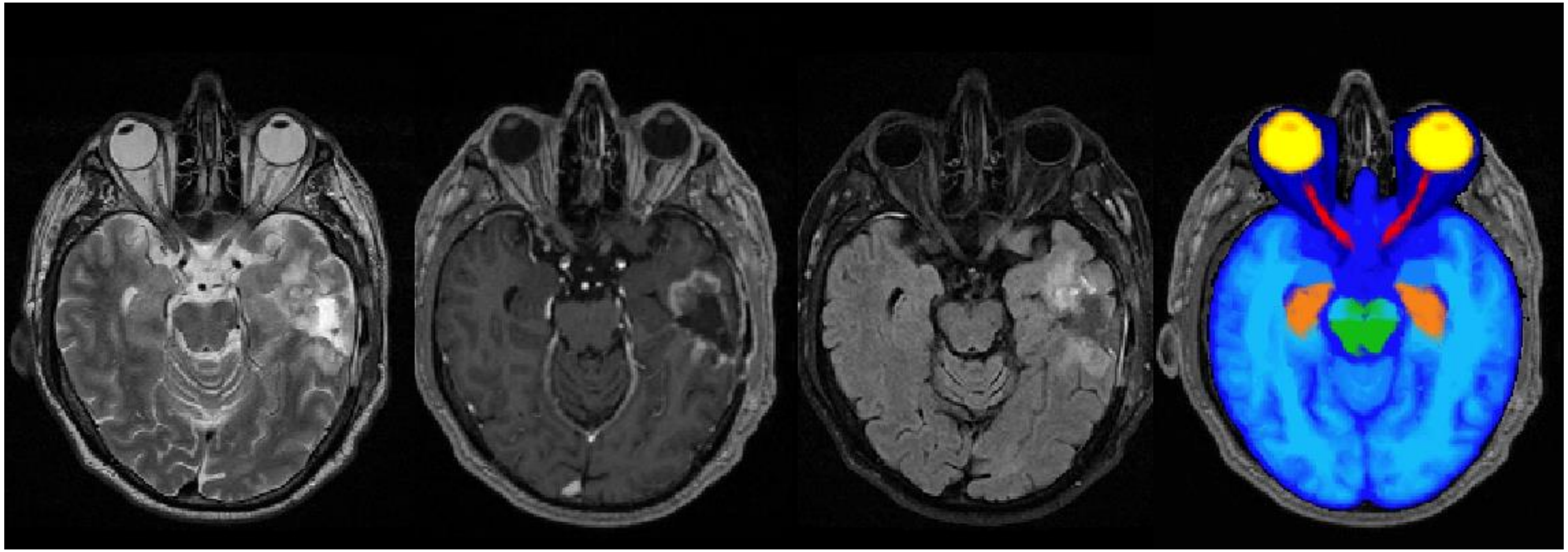
$\boldsymbol{\theta}^{(\tau+1)} \sim p(\boldsymbol{\theta}|\mathbf{d}, \mathbf{l}^{(\tau+1)})$

$\boxed{\text{sample from } p(\mathbf{l}, \boldsymbol{\theta}|\mathbf{d}) \text{ but only keep samples of } \mathbf{l}}$

Gibbs sampling: radiation therapy application

- Optimization as initialization
- Then Gibbs sampling from everything except atlas warp

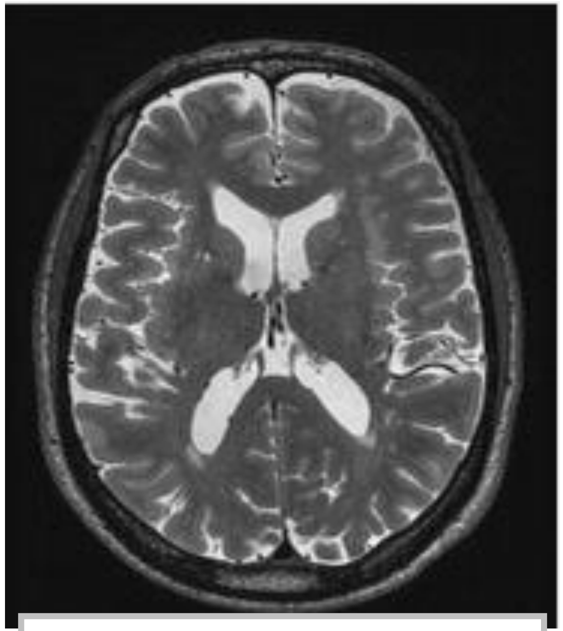
(segmentation – bias field – Gaussian mixture parameters – latent shape model parameters)



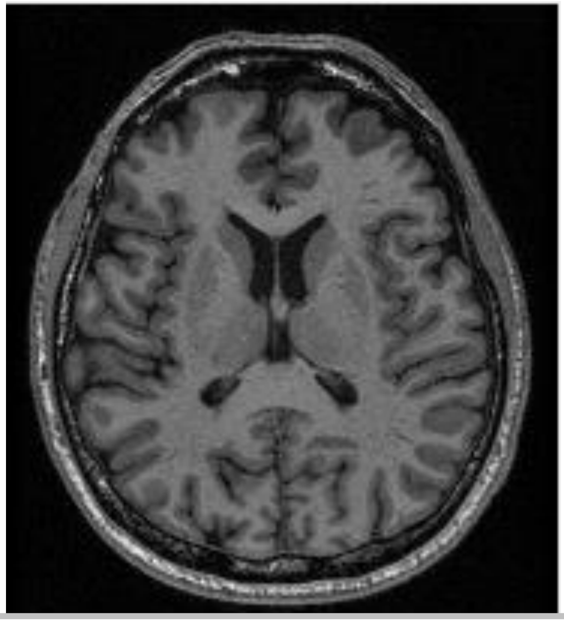
[Agn et al., Medical Image Analysis 2019]

Gibbs sampling: MI-based registration

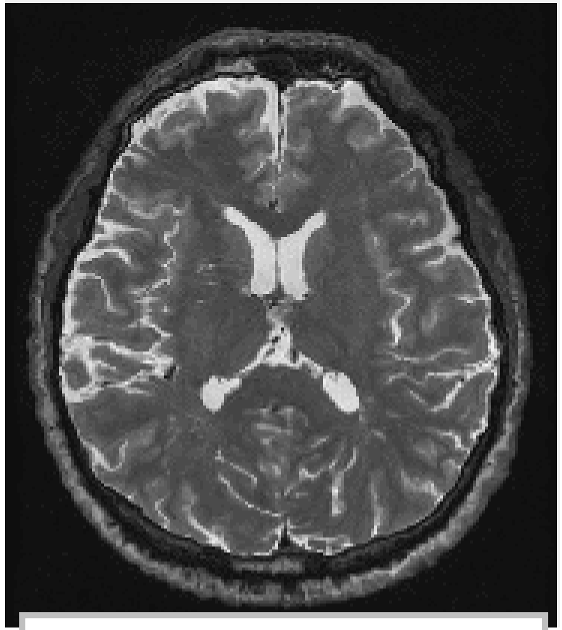
- Mutual Information (MI) rephrased as a generative model
- EM optimization as initialization, then Gibbs sampling (code re-use!)



T2w



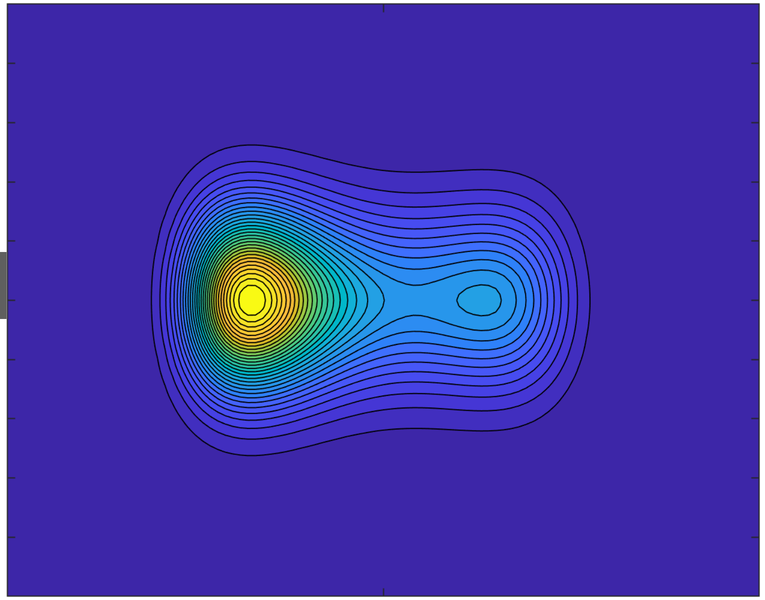
T1w different subject



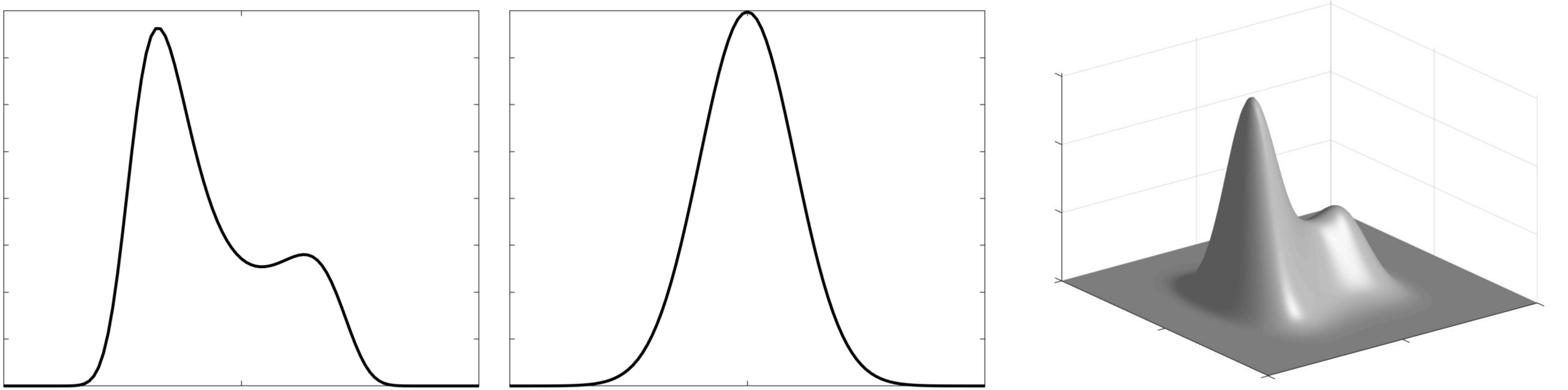
deformed T2w (samples)

[Agn and Van Leemput, UNSURE 2019]

Hamiltonian Monte Carlo (HMC)



- Introduce virtual “momentum” variables $\mathbf{m}$
- Sample from $p(\mathbf{x}, \mathbf{m}) = p(\mathbf{x}) \cdot \mathcal{N}(\mathbf{m}|\mathbf{0}, \mathbf{\Sigma})$, and ignore momentum samples



$p(\mathbf{x})$ $\qquad$ $\mathcal{N}(\mathbf{m}|\mathbf{0}, \mathbf{\Sigma})$ $\qquad$ $p(\mathbf{x}, \mathbf{m}) = p(\mathbf{x}) \cdot \mathcal{N}(\mathbf{m}|\mathbf{0}, \mathbf{\Sigma})$

Hamiltonian Monte Carlo (HMC)

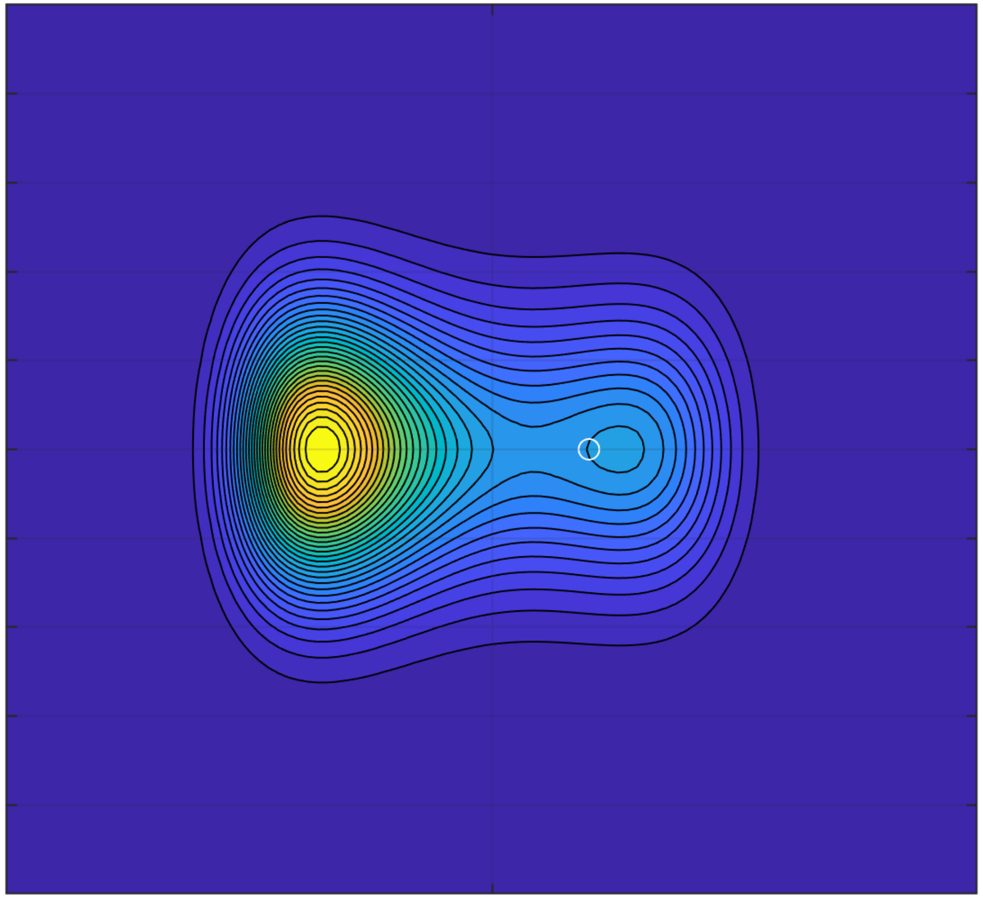
(1) Sample momentum from $\mathcal{N}(\mathbf{m}|\mathbf{0}, \mathbf{\Sigma})$

(2) Follow isocontour of $p(\mathbf{x}, \mathbf{m})$

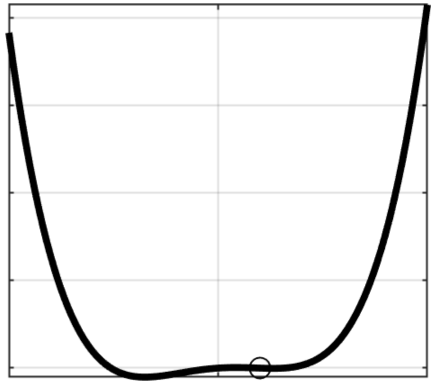
(3) Repeat (1) and (2)

- *Efficiency by using gradient information* $\nabla \log p(\mathbf{x})$

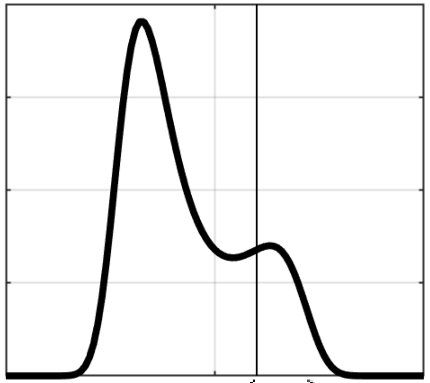
- *Cf. optimization*



$p(\mathbf{x}, \mathbf{m})$



$-\log p(\mathbf{x})$



$p(\mathbf{x})$

Hamiltonian Monte Carlo (HMC)

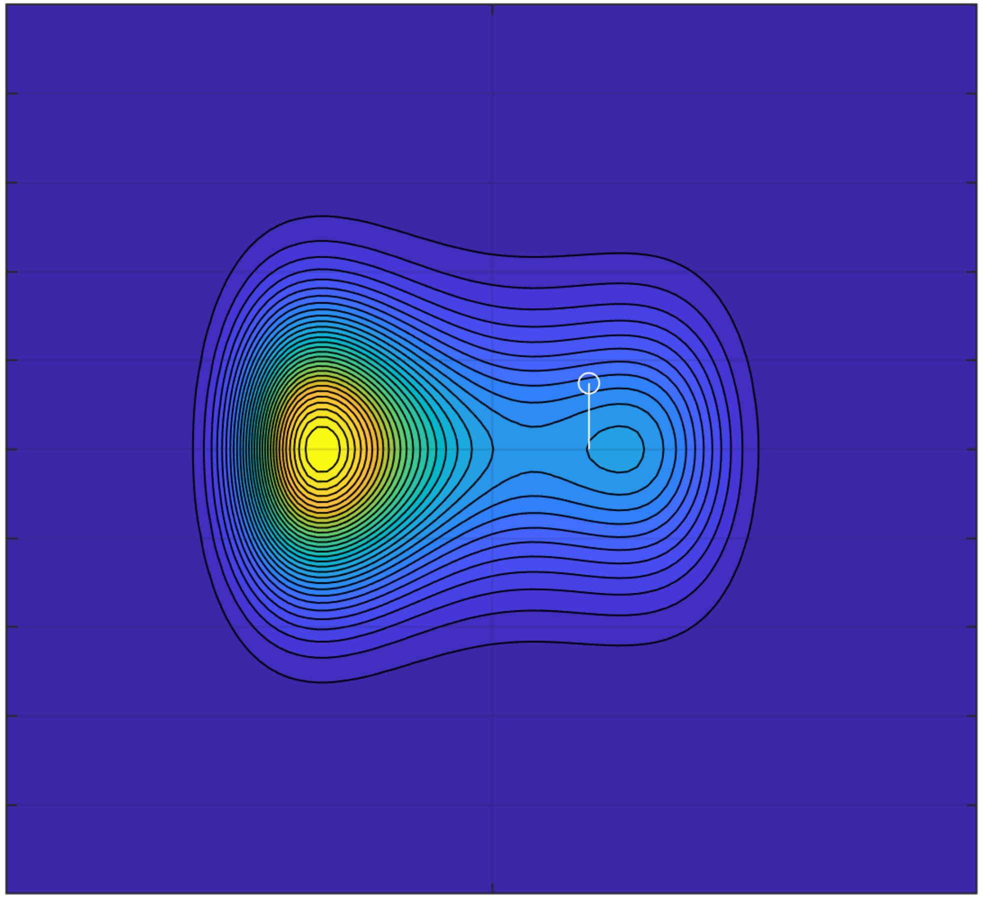
(1) Sample momentum from $\mathcal{N}(\mathbf{m}|\mathbf{0}, \mathbf{\Sigma})$

(2) Follow isocontour of $p(\mathbf{x}, \mathbf{m})$

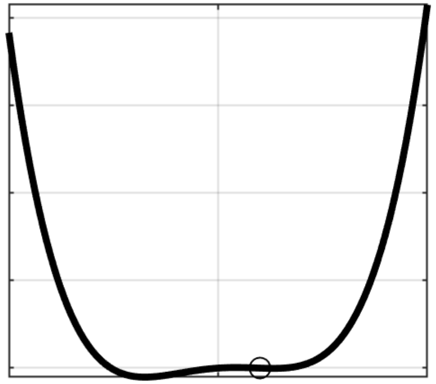
(3) Repeat (1) and (2)

- *Efficiency by using gradient information* $\nabla \log p(\mathbf{x})$

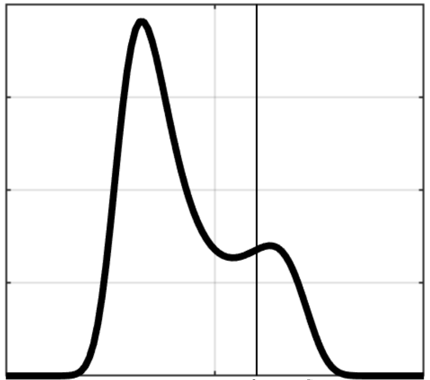
- *Cf. optimization*



$p(\mathbf{x}, \mathbf{m})$



$-\log p(\mathbf{x})$



$p(\mathbf{x})$

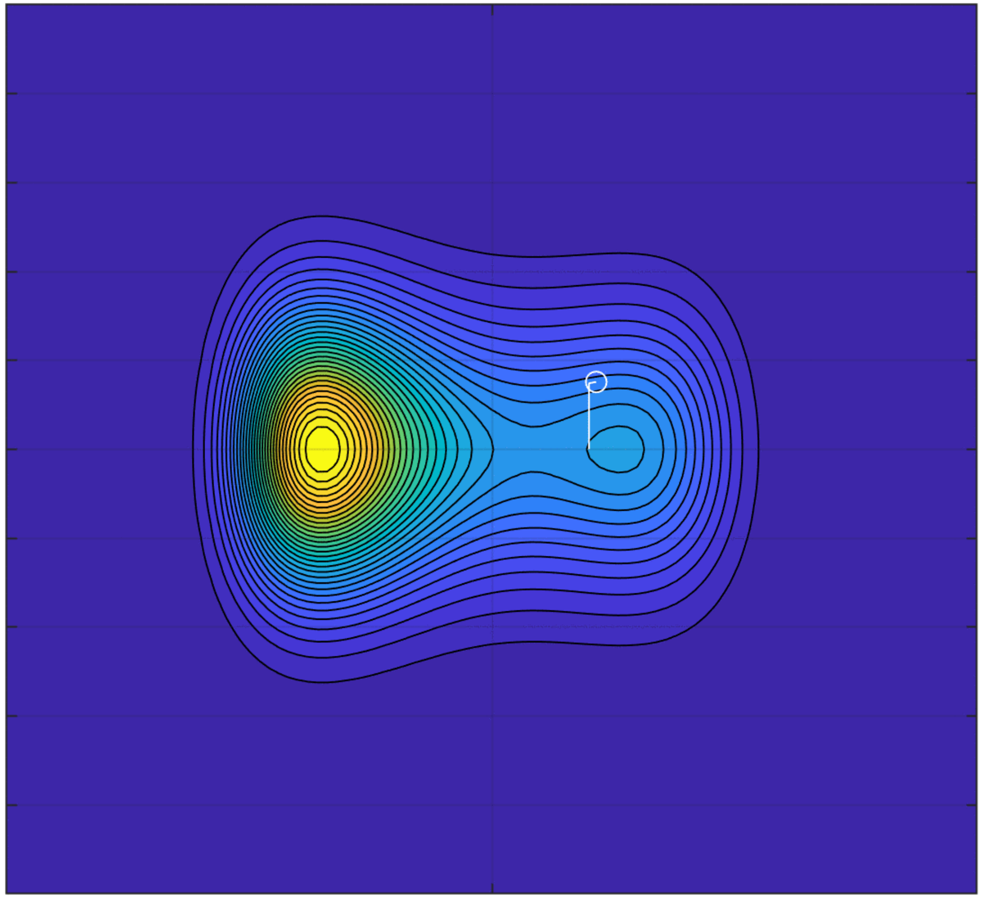
Hamiltonian Monte Carlo (HMC)

(1) Sample momentum from $\mathcal{N}(\mathbf{m}|\mathbf{0}, \mathbf{\Sigma})$

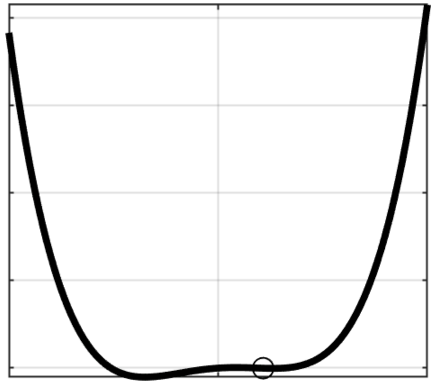
(2) Follow isocontour of $p(\mathbf{x}, \mathbf{m})$

(3) Repeat (1) and (2)

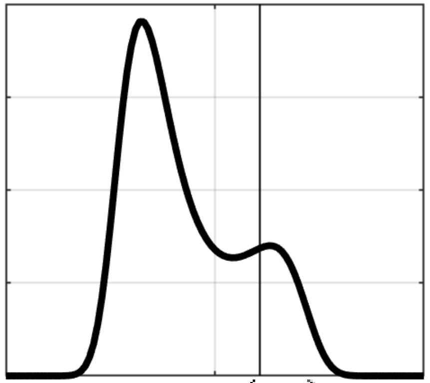
- *Efficiency by using gradient information* $\nabla \log p(\mathbf{x})$
- *Cf. optimization*



$p(\mathbf{x}, \mathbf{m})$



$-\log p(\mathbf{x})$



$p(\mathbf{x})$

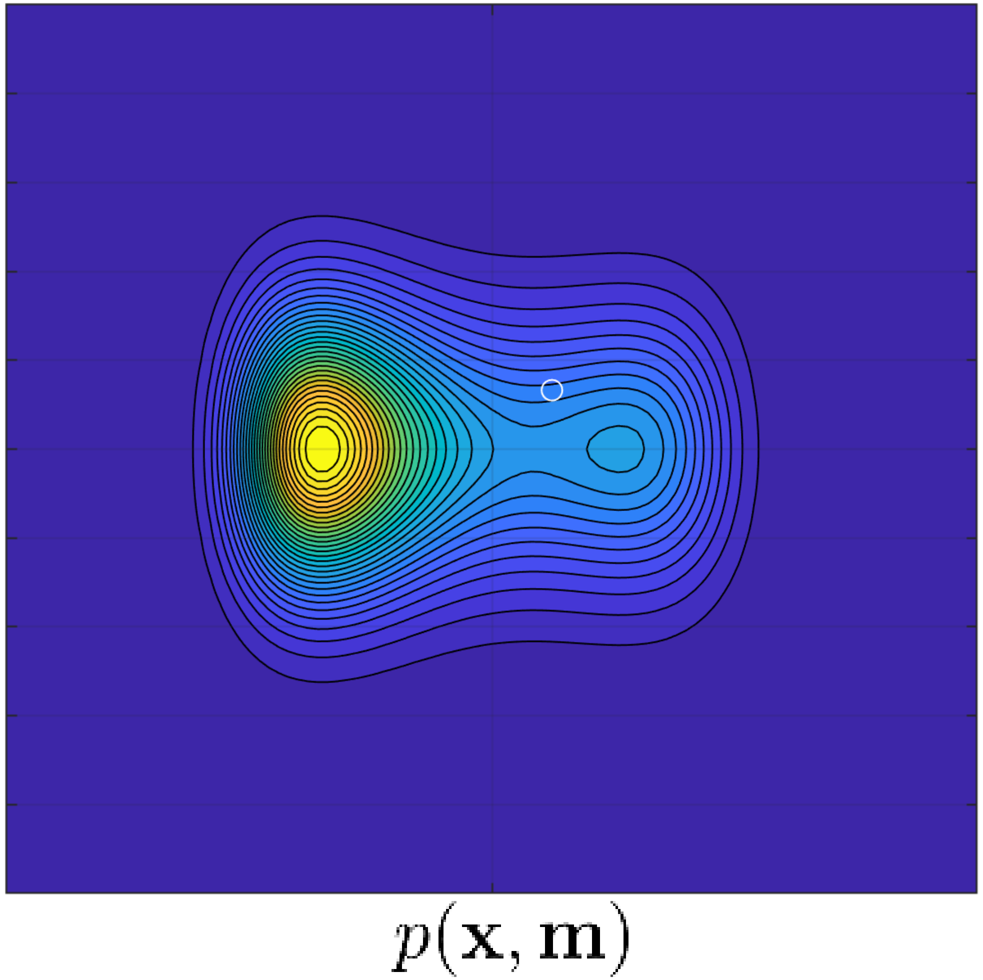
Hamiltonian Monte Carlo (HMC)

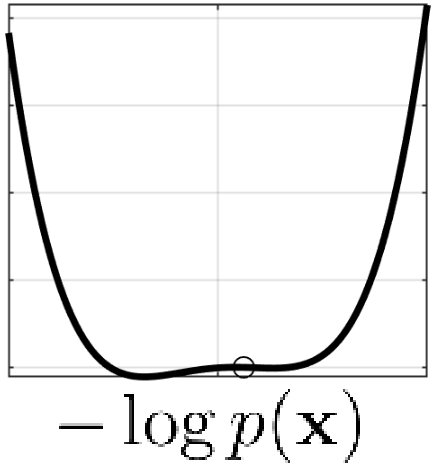
(1) Sample momentum from $\mathcal{N}(\mathbf{m}|\mathbf{0},\mathbf{\Sigma})$

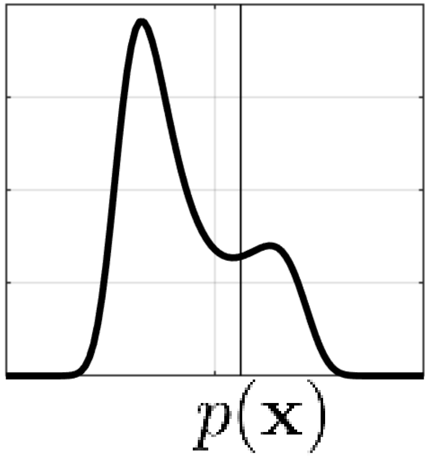
(2) Follow isocontour of $p(\mathbf{x},\mathbf{m})$

(3) Repeat (1) and (2)

- *Efficiency by using gradient information* $\nabla \log p(\mathbf{x})$
- *Cf. optimization*







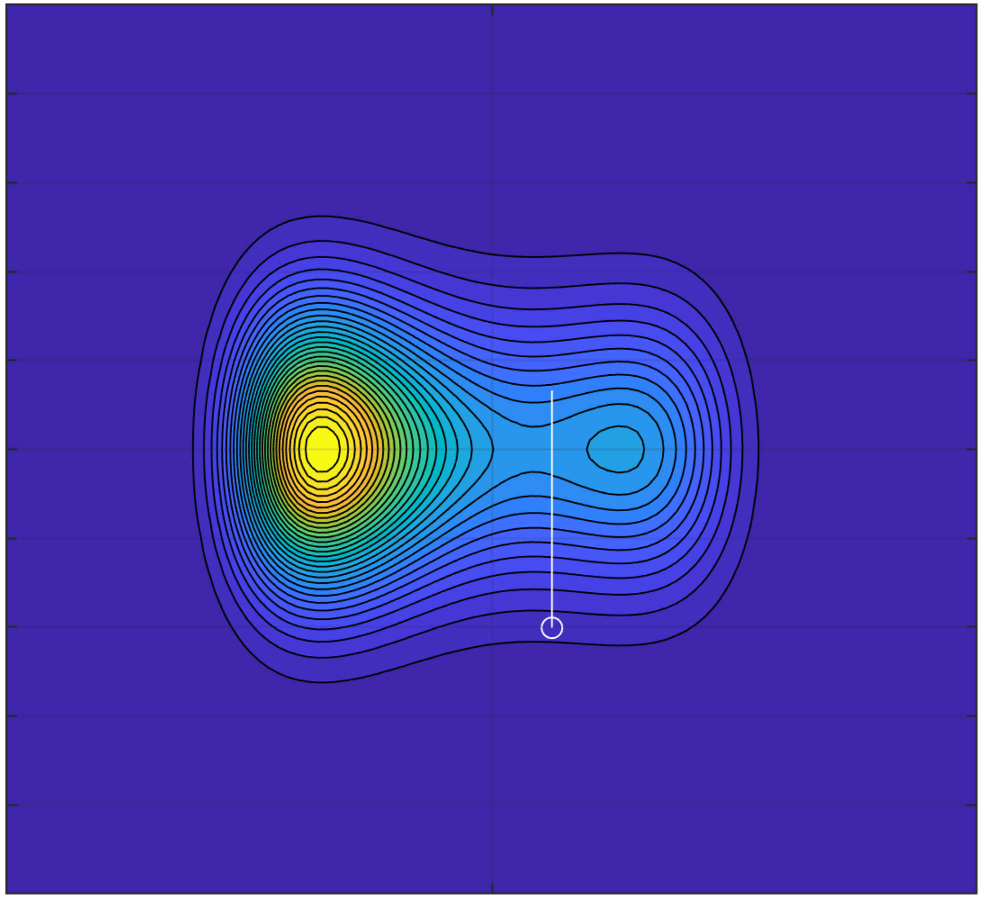
Hamiltonian Monte Carlo (HMC)

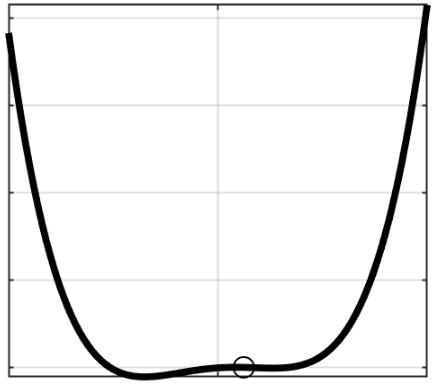
(1) Sample momentum from $\mathcal{N}(\mathbf{m}|\mathbf{0}, \mathbf{\Sigma})$

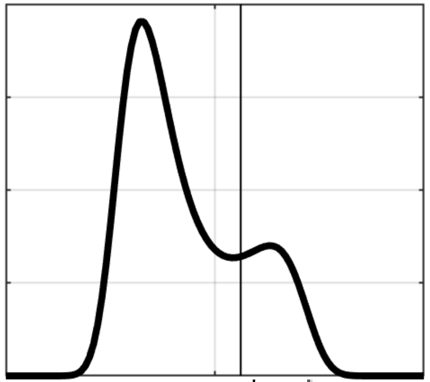
(2) Follow isocontour of $p(\mathbf{x}, \mathbf{m})$

(3) Repeat (1) and (2)

- *Efficiency by using gradient information* $\nabla \log p(\mathbf{x})$
- *Cf. optimization*







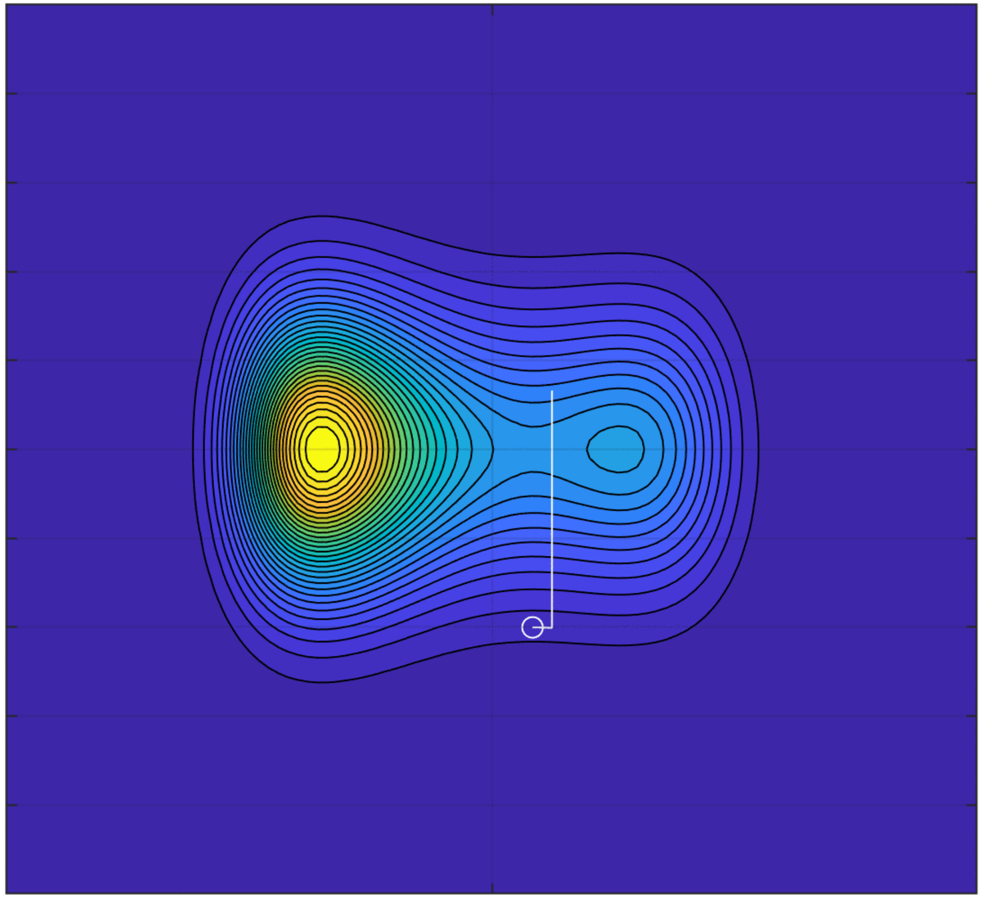
Hamiltonian Monte Carlo (HMC)

(1) Sample momentum from $\mathcal{N}(\mathbf{m}|\mathbf{0}, \mathbf{\Sigma})$

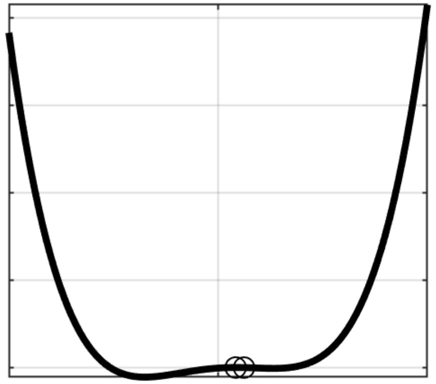
(2) Follow isocontour of $p(\mathbf{x}, \mathbf{m})$

(3) Repeat (1) and (2)

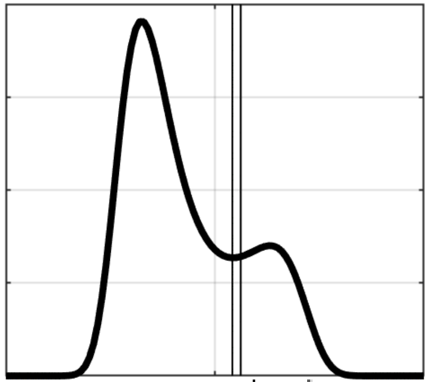
- *Efficiency by using gradient information* $\nabla \log p(\mathbf{x})$
- *Cf. optimization*



$p(\mathbf{x}, \mathbf{m})$

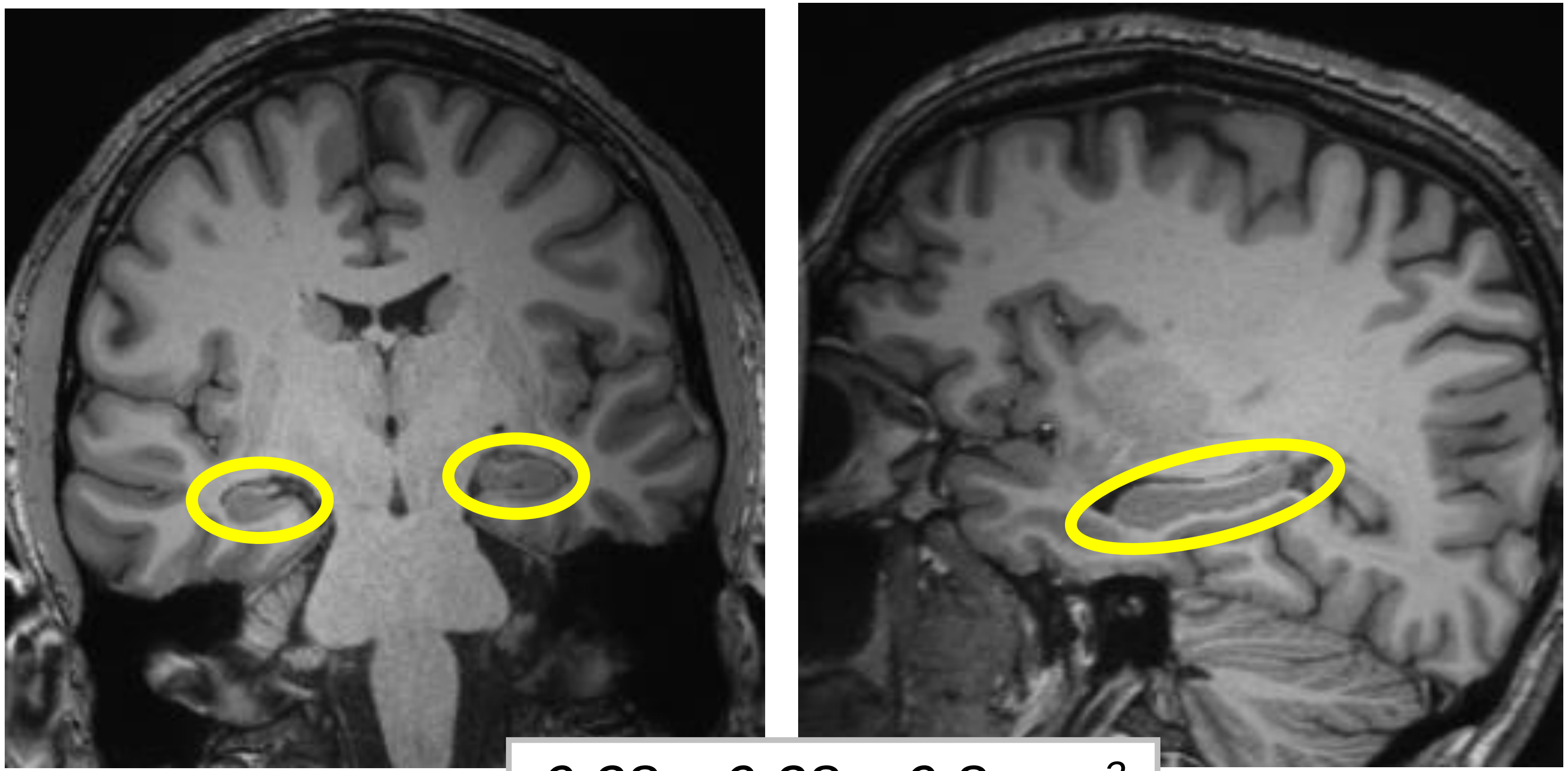


$-\log p(\mathbf{x})$



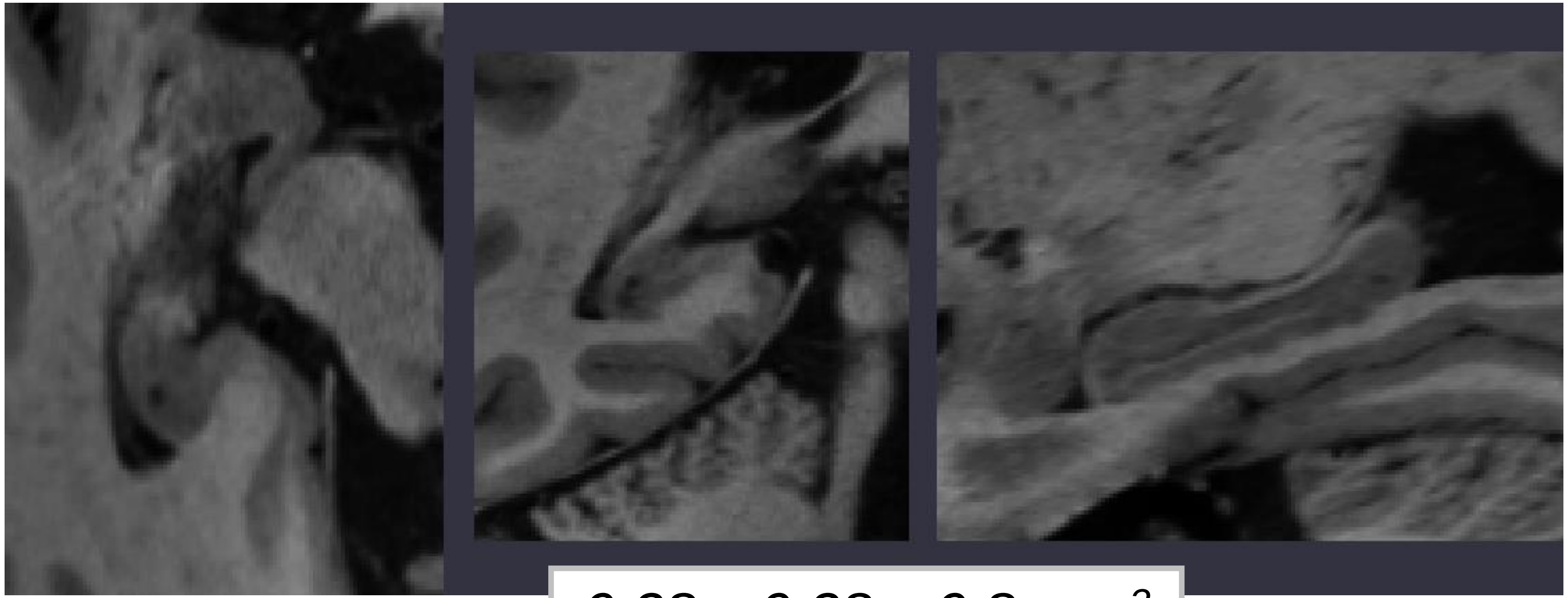
$p(\mathbf{x})$

HMC: hippocampal subfield segmentation



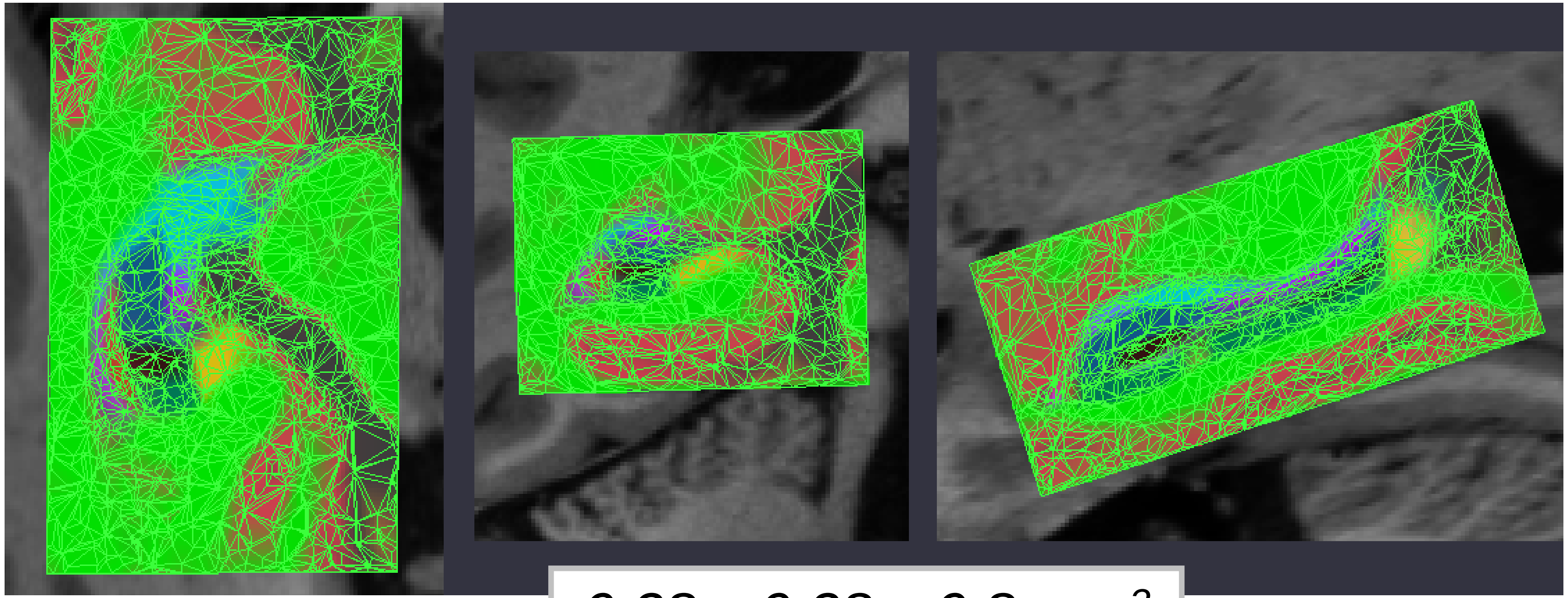
0.38 x 0.38 x 0.8 mm^3

HMC: hippocampal subfield segmentation



0.38 x 0.38 x 0.8 mm^3

HMC: hippocampal subfield segmentation



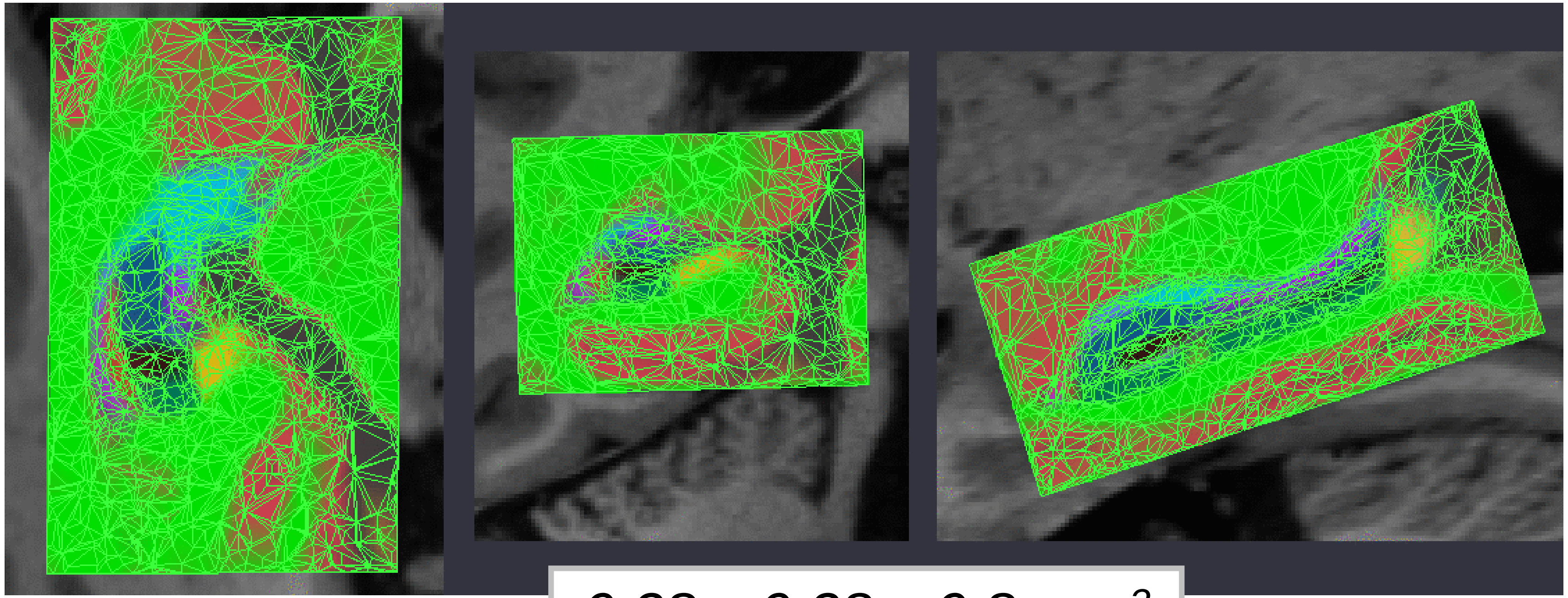
0.38 x 0.38 x 0.8 mm³

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} p(\underline{\boldsymbol{\theta}}|\mathbf{d})$$

mesh node positions;

Gaussian means and variances

HMC: hippocampal subfield segmentation



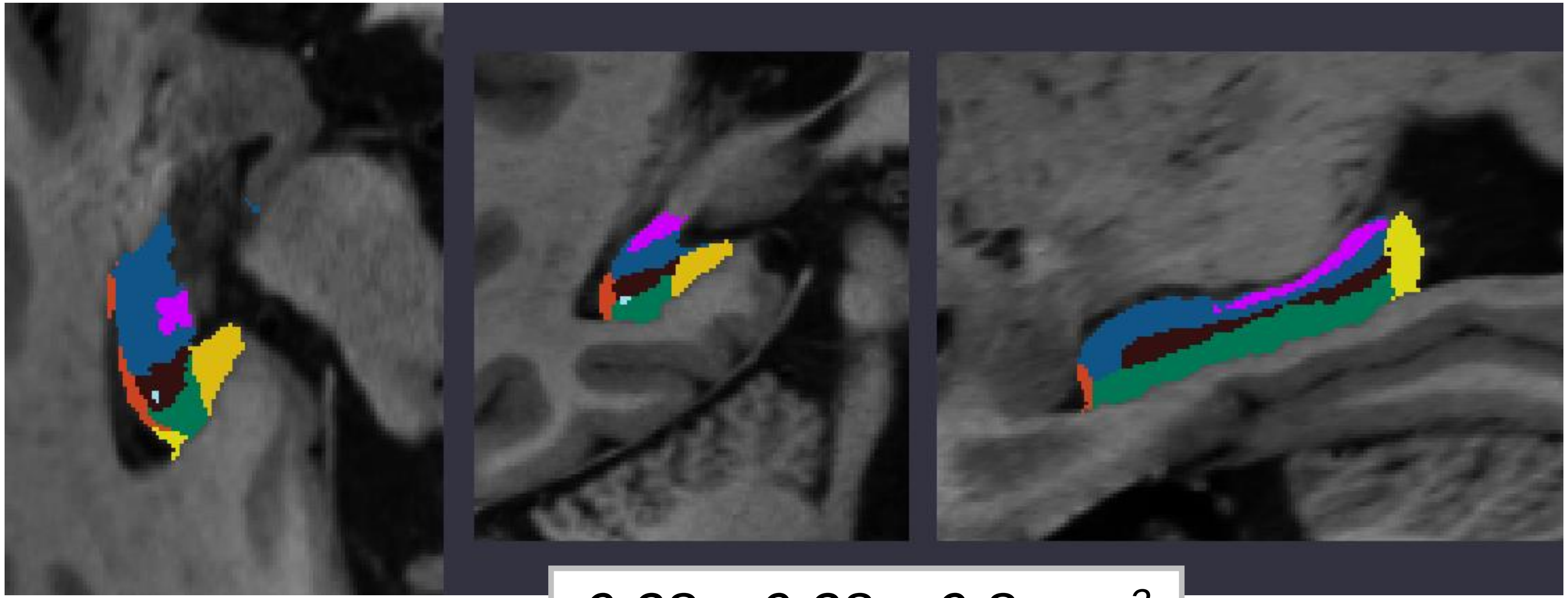
0.38 x 0.38 x 0.8 mm³

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} p(\underline{\boldsymbol{\theta}}|\mathbf{d})$$

mesh node positions;

Gaussian means and variances

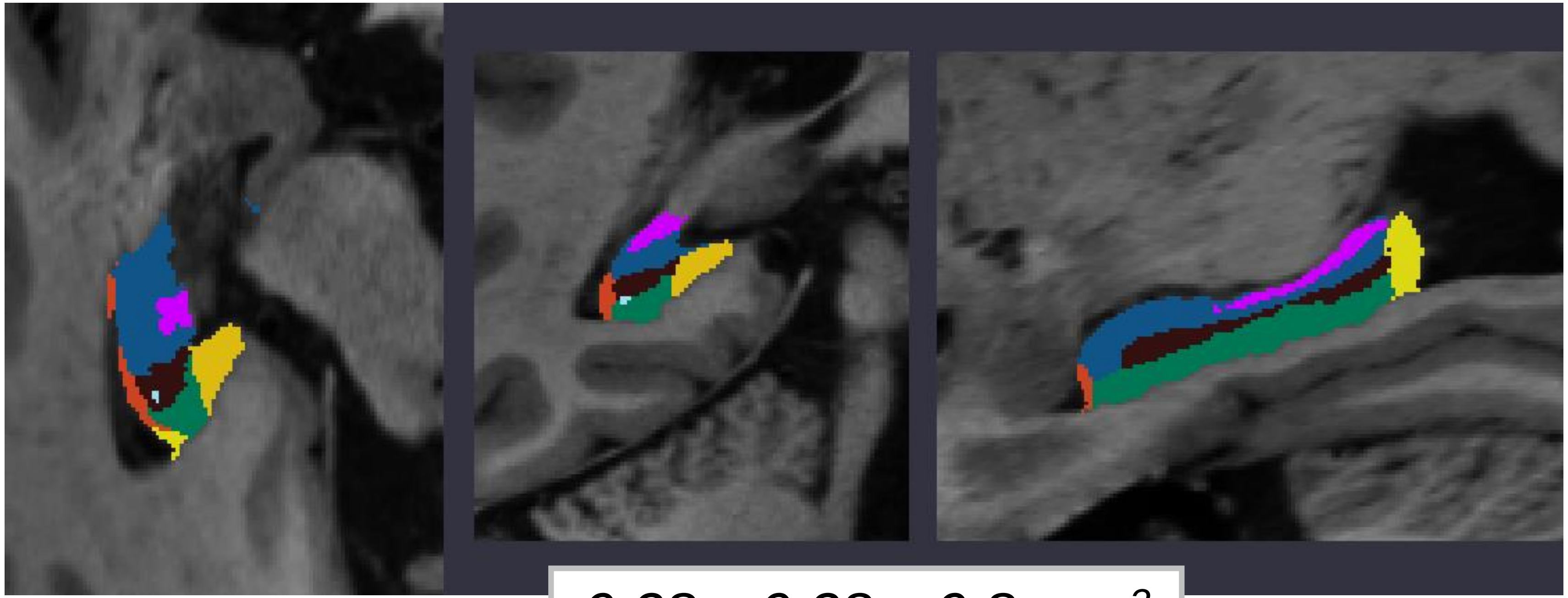
HMC: hippocampal subfield segmentation



0.38 x 0.38 x 0.8 mm³

$$\hat{\mathbf{l}} = \arg\max_{\mathbf{l}} p(\mathbf{l}|\mathbf{d}, \hat{\boldsymbol{\theta}})$$

HMC: hippocampal subfield segmentation



0.38 x 0.38 x 0.8 mm³

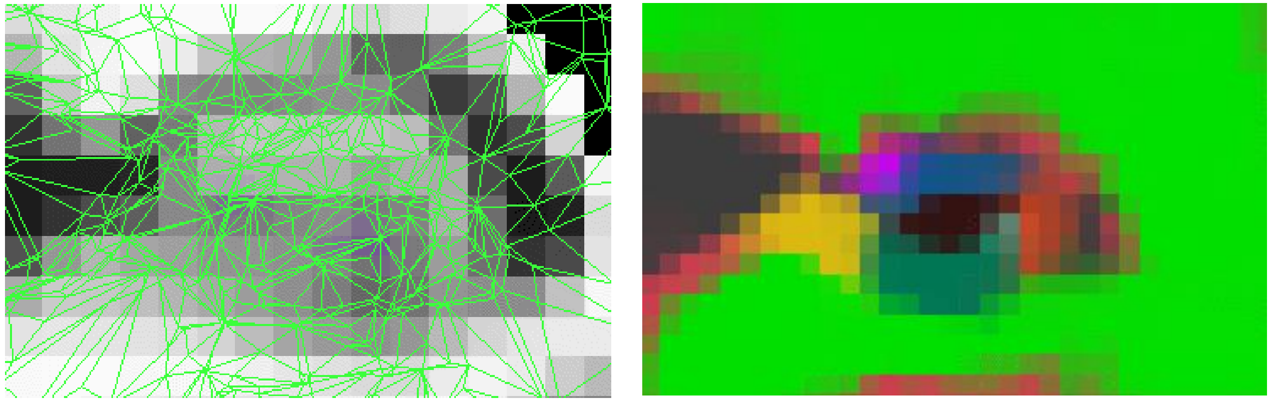
$$\hat{\mathbf{l}} = \arg\max_{\mathbf{l}} p(\mathbf{l}|\mathbf{d}, \hat{\boldsymbol{\theta}})$$

$$\boxed{p(\mathbf{l}|\mathbf{d}) = \int_{\boldsymbol{\theta}} p(\mathbf{l}, \boldsymbol{\theta}|\mathbf{d})\, \mathrm{d}\boldsymbol{\theta}}$$

HMC: hippocampal subfield segmentation

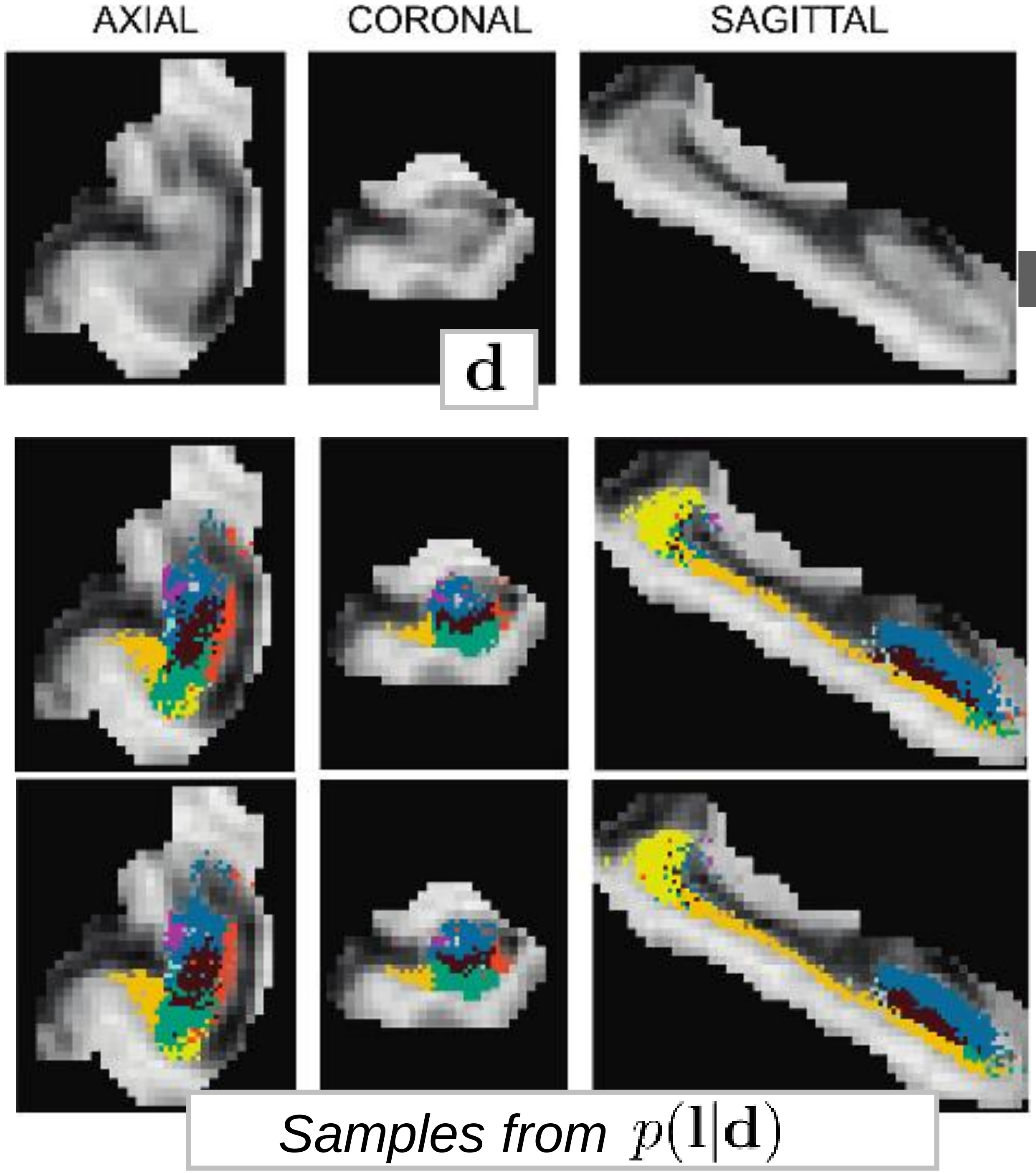
– *Sample from* $p(\mathbf{l}|\mathbf{d}) = \int_{\boldsymbol{\theta}} p(\mathbf{l}, \boldsymbol{\theta}|\mathbf{d})\, \mathrm{d}\boldsymbol{\theta}$ *by sampling from* $p(\mathbf{l}, \boldsymbol{\theta}|\mathbf{d})$

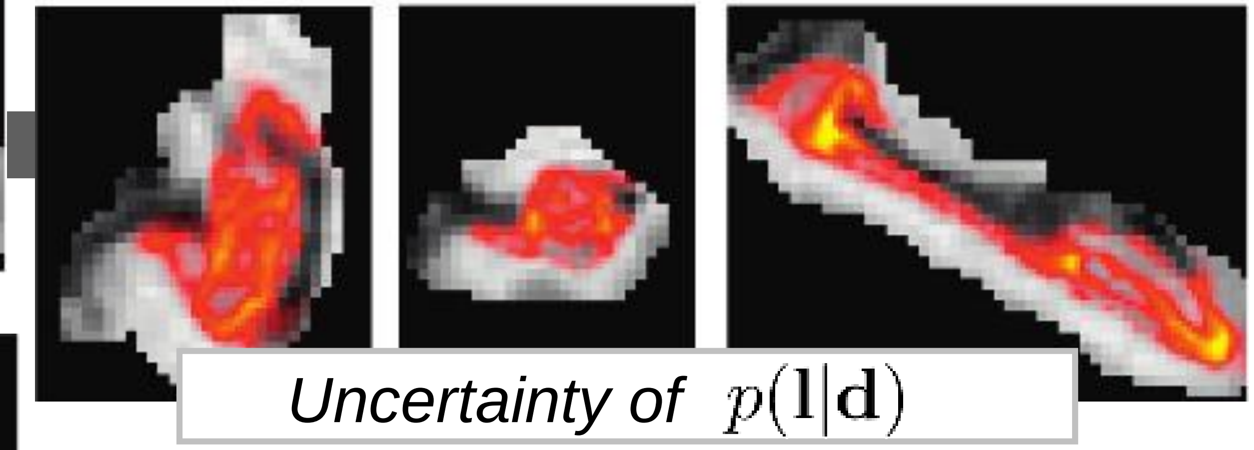
– *HMC for atlas deformation; Gibbs sampling for the rest*

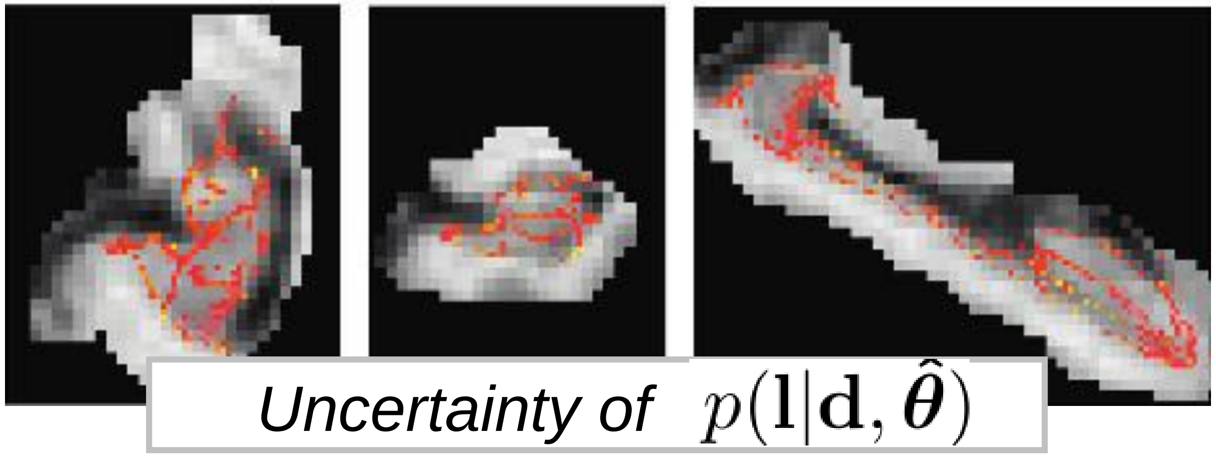


1x1x1 mm^3

[Iglesias et al., Medical Image Analysis 2013]







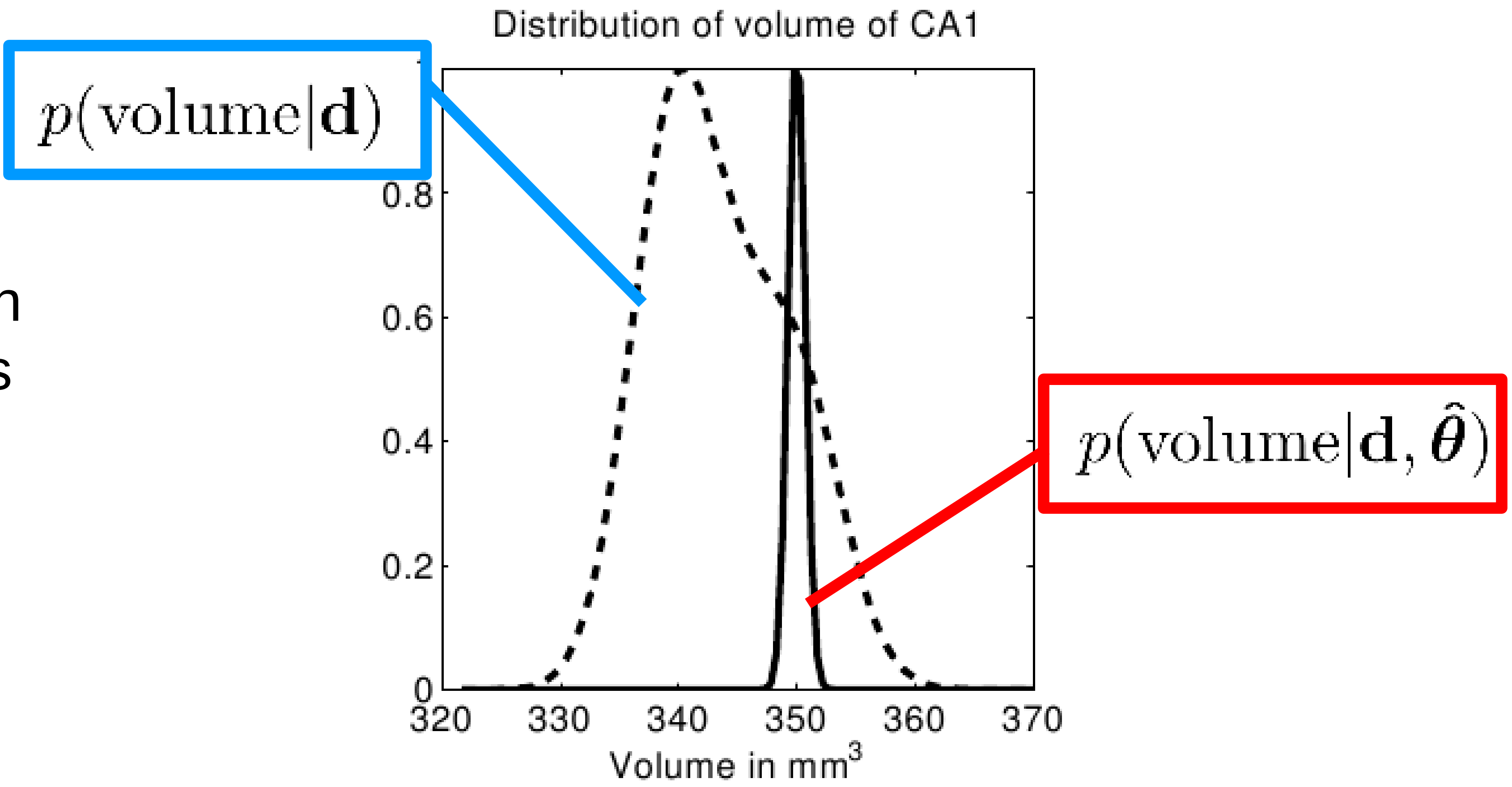
[Iglesias et al., Medical Image Analysis 2013]

HMC: hippocampal subfield segmentation

Volumes of the subfields:

– Better “error bars”

– 2% boost in classification accuracy AD vs. controls



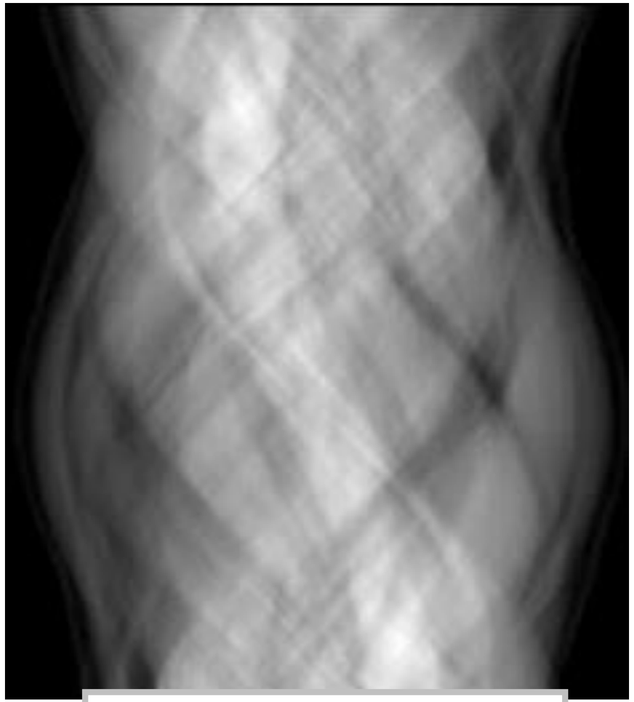
[Iglesias et al., Medical Image Analysis 2013]

HMC: PET reconstruction

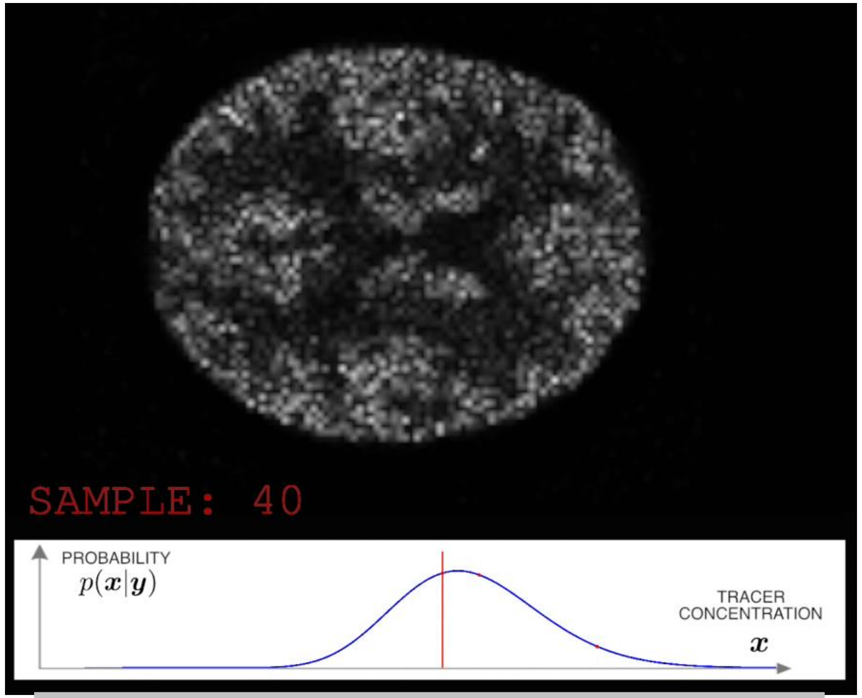
– Reconstruction = inverting a generative model

– HMC: $p(\mathbf{x}, \mathbf{m}) = p(\mathbf{x}) \cdot \mathcal{N}(\mathbf{m}|\mathbf{0}, \boldsymbol{\Sigma})$

chosen "wisely" (using FFTs)



projection data



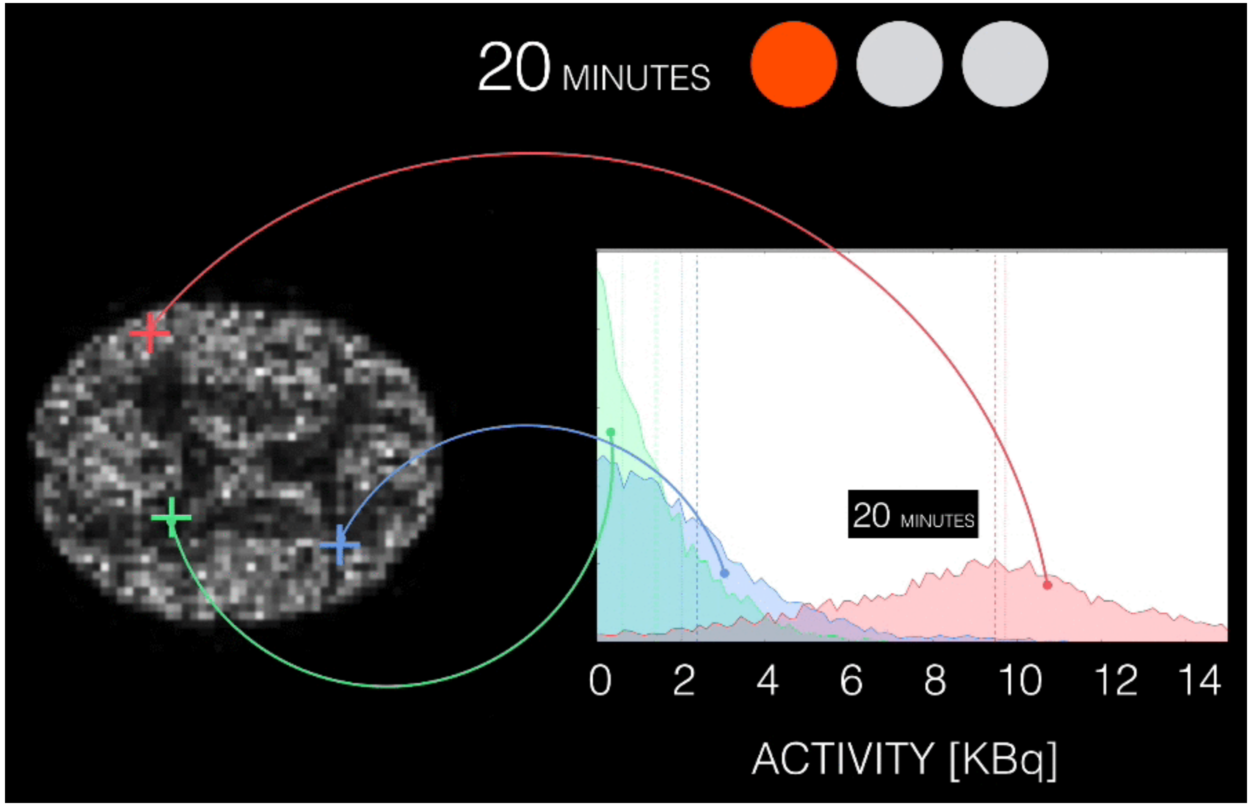
reconstruction samples

[Pedemonte and Van Leemput, MICCAI 2015]

HMC: PET reconstruction

- Reconstruction = inverting a generative model
- HMC: $p(\mathbf{x}, \mathbf{m}) = p(\mathbf{x}) \cdot \mathcal{N}(\mathbf{m}|\mathbf{0}, \boldsymbol{\Sigma})$

chosen "wisely" (using FFTs)



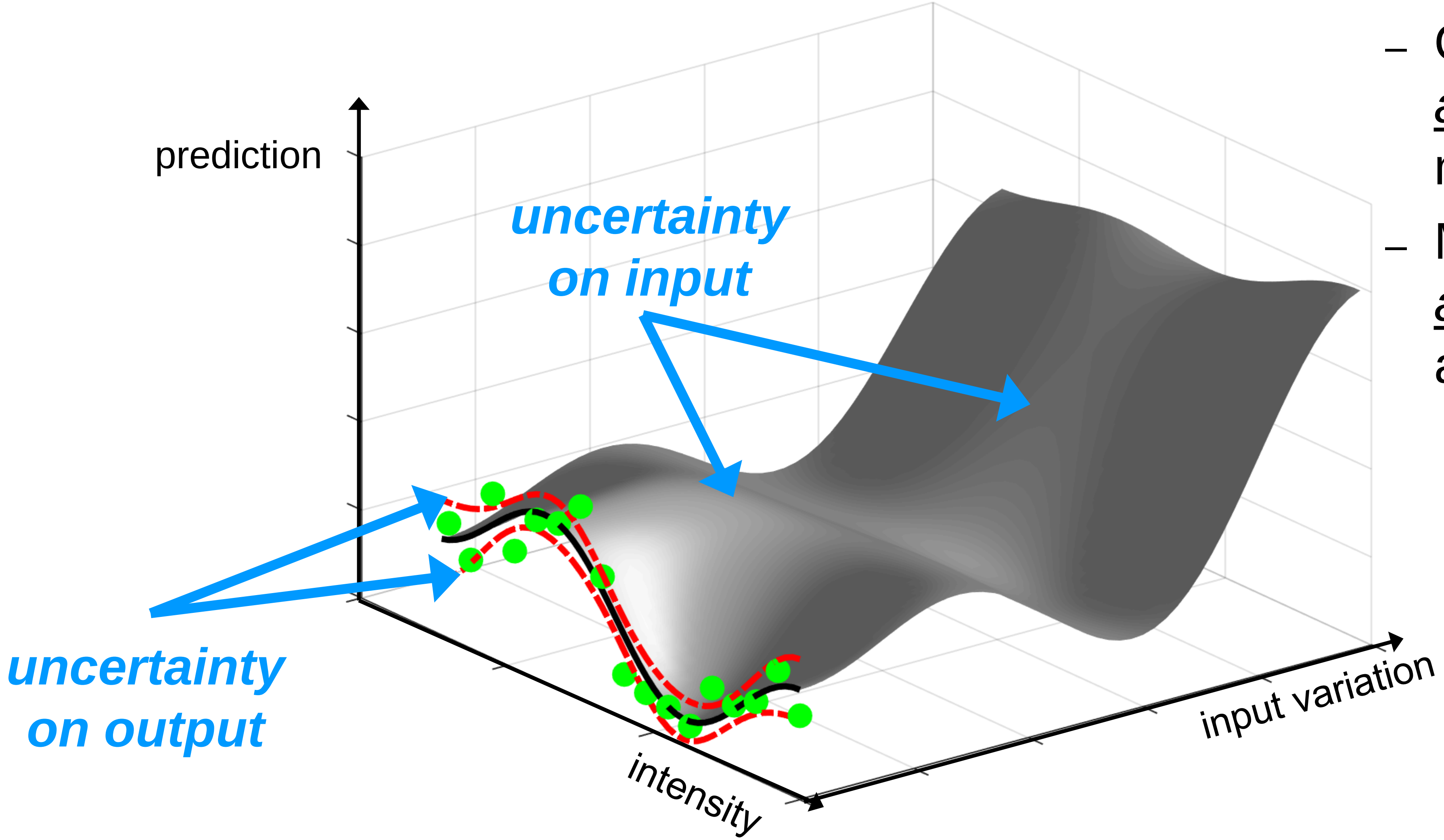
Overview

Uncertainty on input using generative models

Uncertainty on output using Monte Carlo sampling

Discussion and conclusion

Discussion and Conclusion



– Generative models <u>*in addition to*</u> discriminative models
– Monte Carlo sampling <u>*in addition*</u> to variational approximations

Main drawback is computational speed, but sometimes it's worth it

Acknowledgements

Laboratory for Computational Neuroimaging

We're hiring!

(postdoc in Boston)

Eugenio Iglesias

Oula Puonti

Mikael Agn

Stefano Pedemonte