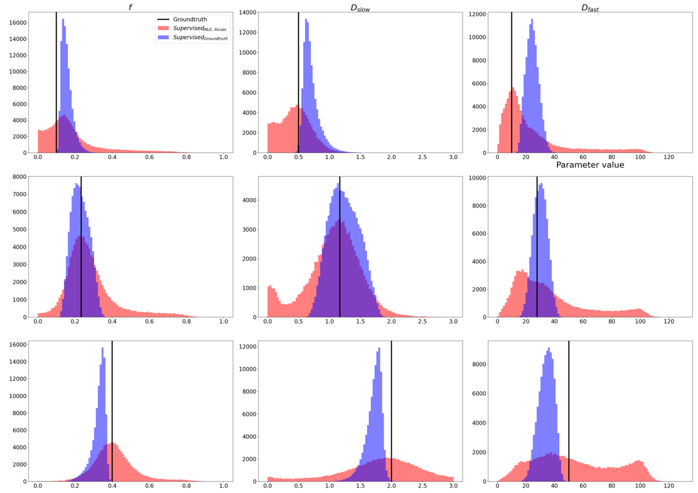
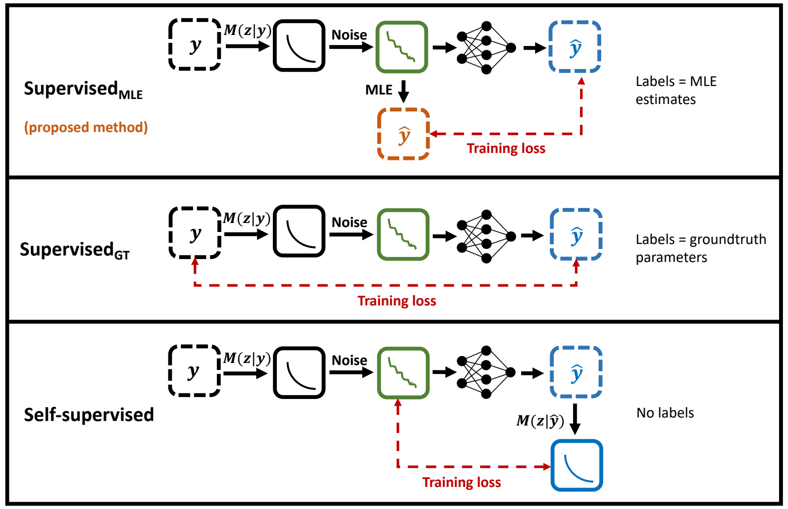
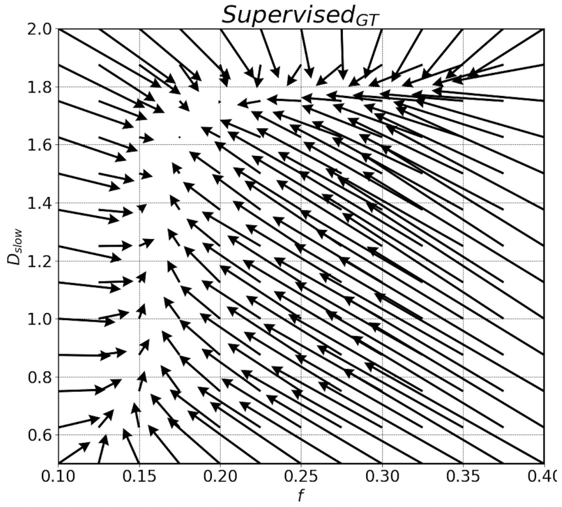
Choice of Training Label Matters: Deep Learning for Quantitative MRI Parameter Estimation

Sean Epstein

Centre for Medical Image Computing, UCL







What am I talking about today?

Acquisition scheme
(e.g. b-value)

Tissue property
(e.g. diffusivity)

Modelled signal
(e.g. ADC)

$$S(z, \epsilon | y) = \mathrm{M}(z | y) + \epsilon$$

Measured
signal

Noise

Noise instantiation

What am I talking about today?

Acquisition scheme (e.g. b-value)

Tissue property (e.g. diffusivity)

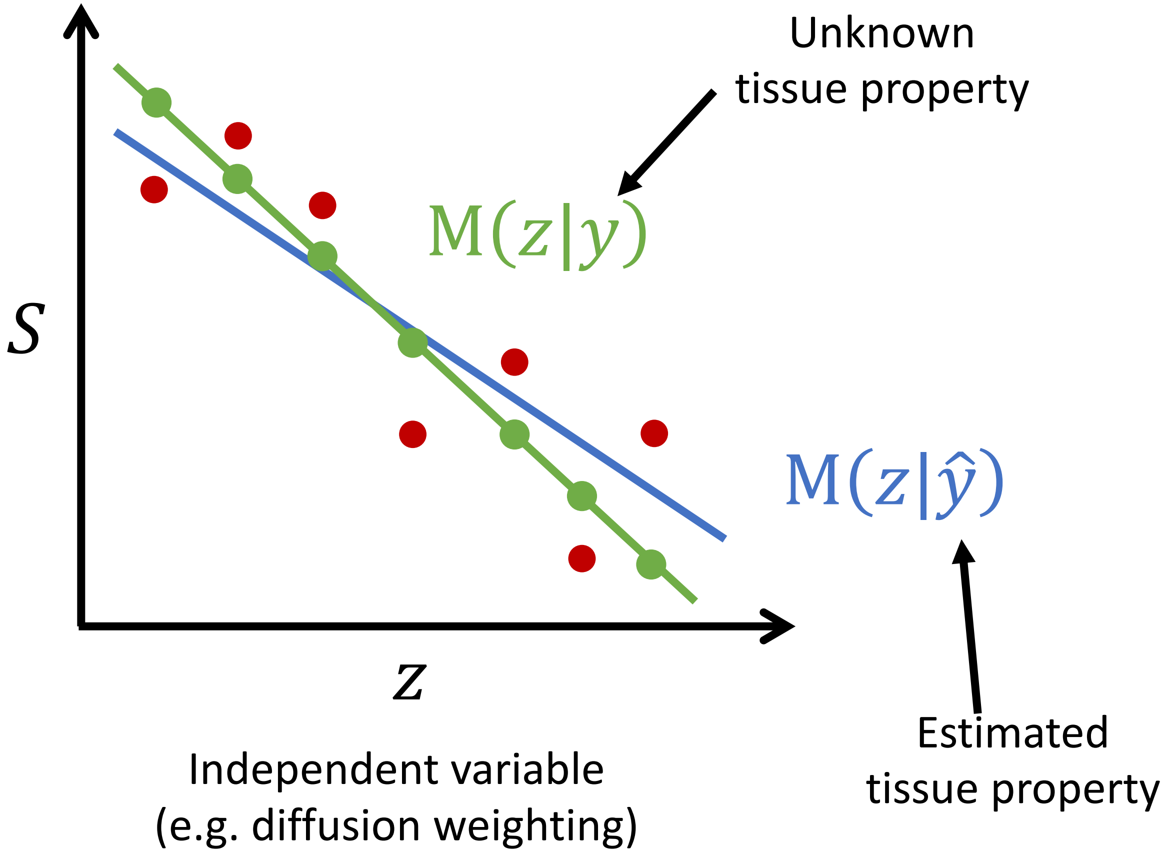
Modelled signal (e.g. ADC)

$$S(z, \epsilon|y) = \mathrm{M}(z|y) + \epsilon$$

Measured signal

Noise

Noise instantiation



How do we solve this inverse problem?

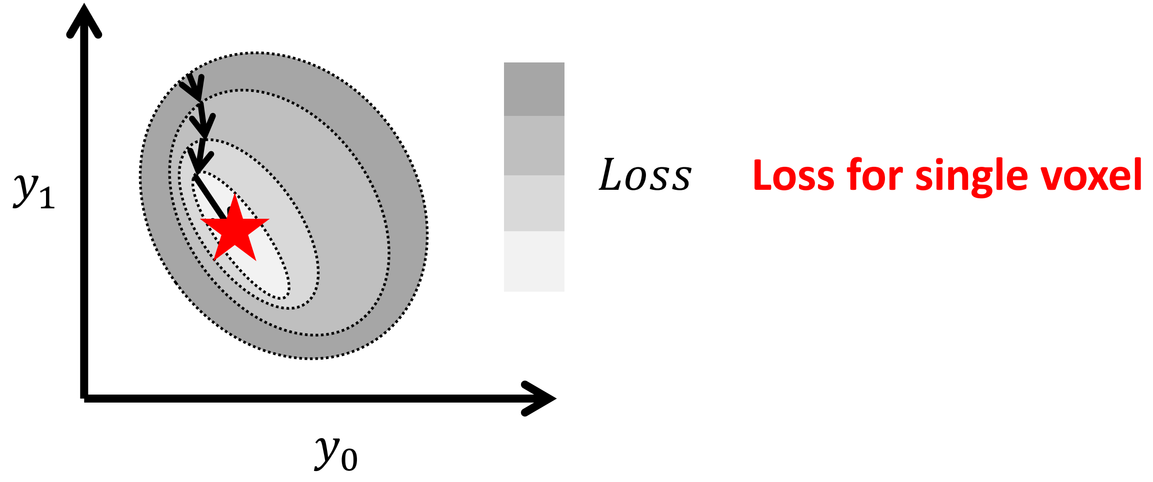
For a given noisy S, we want the corresponding y

Traditional (MLE): $\min_{y} Loss(y|S)$

Tackle the problem **directly**

Optimise over **variable of interest** y

Repeat optimisation for every signal



How do we solve this inverse problem?

For a given noisy S, we want the corresponding y

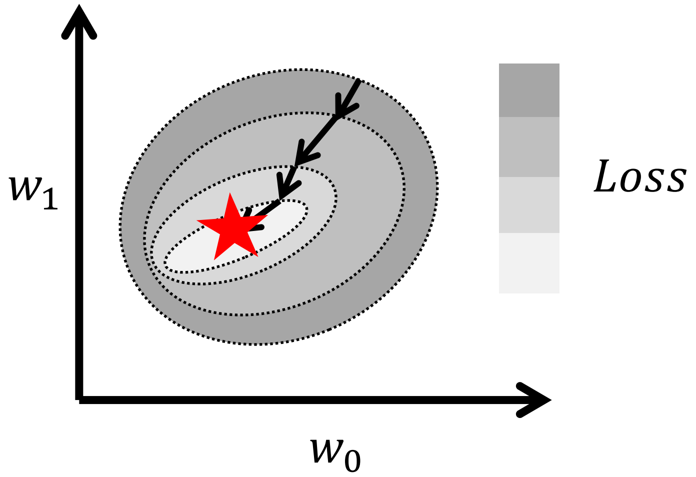
Machine learning (ML): $\min_{w} \sum_{i=1}^{N_{train}} Loss(w|S_i)$

Tackle the problem **indirectly**

Find **general mapping** from any S to corresponding y

Optimise over **latent variable** w

Perform optimisation **only once** (training)



Loss averaged over all training voxels

How do we solve this inverse problem?

For a given noisy S, we want the corresponding y

Traditional (MLE): $\min_{y} Loss(y|S)$

Tackle the problem **directly**

Optimise over **variable of interest** y

Repeat optimisation for every signal

Machine learning (ML): $\min_{w} \sum_{i=1}^{N_{train}} Loss(w|S_i)$

Tackle the problem **indirectly**

Find **general mapping** from any S to corresponding y

Optimise over **latent variable** w

Perform optimisation **only once** (training)

Limitations

- **Cost**: expensive, scales linearly with newly acquired data
- **Performance**: suffers from local minima; each optimisation solved in isolation

Solutions

- **Cost**: frontloaded (training), then negligible
- **Performance**: leverages patterns across training voxels

How is ML parameter estimation implemented today?

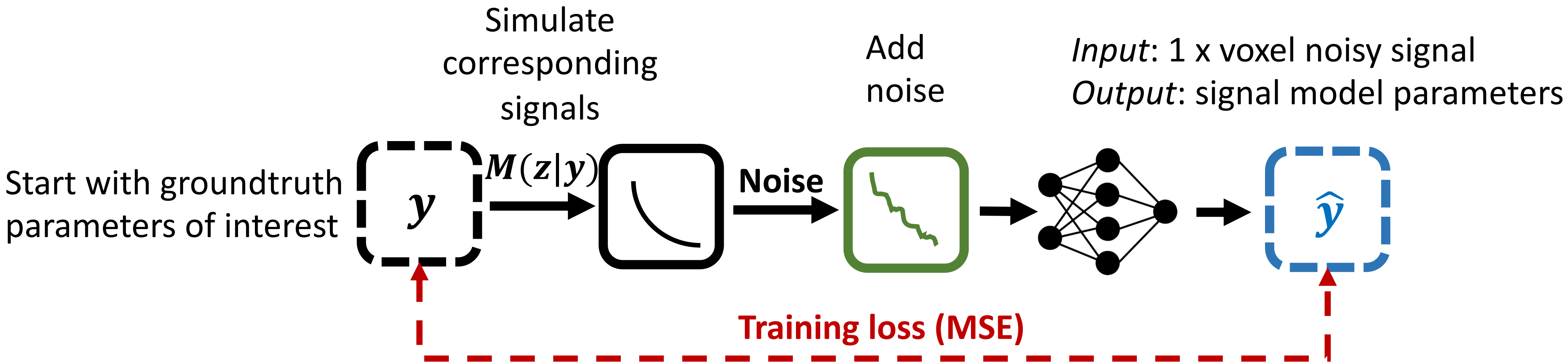
Supervised methods

e.g. Bertleff 2017, Gyori 2022

Synthetic training data

Self-supervised ("unsupervised") methods

e.g. Barbieri 2019, Kaandorp 2021

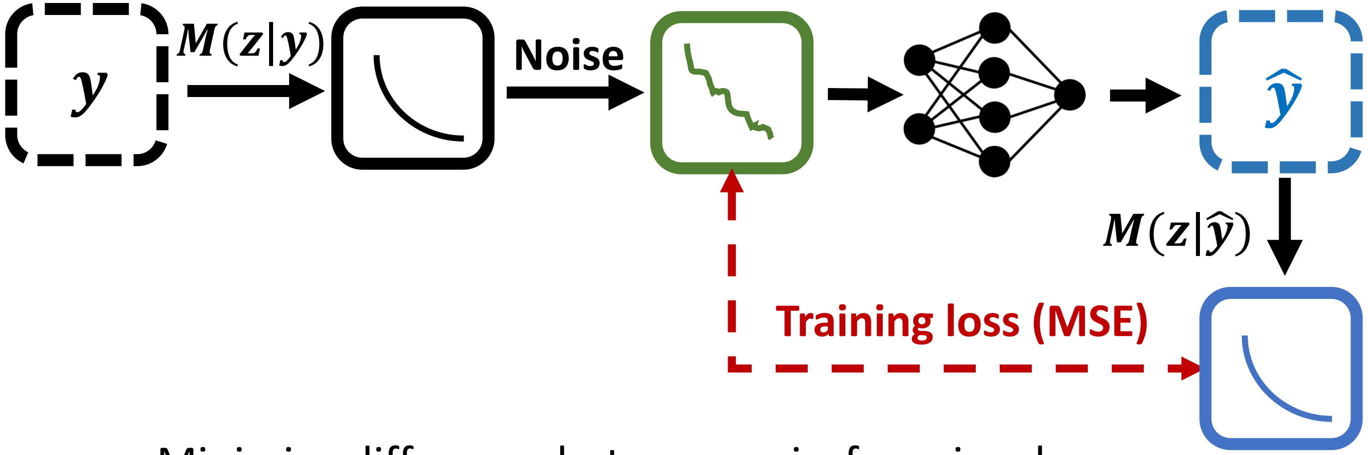


Minimise difference between output and groundtruth value (label)

How is ML parameter estimation implemented today?

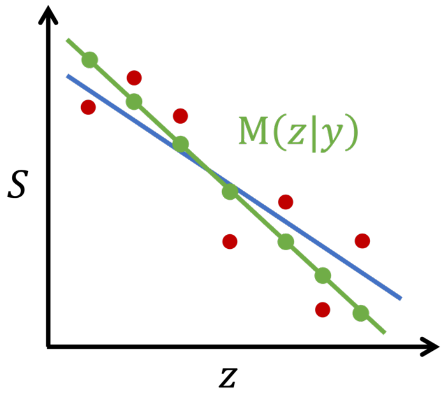
Supervised methods
e.g. Bertleff 2017, Gyori 2022

Self-supervised ("unsupervised") methods
e.g. Barbieri 2019, Kaandorp 2021



Minimise difference between noisefree signal representation of output and noisy input

Replicating traditional MLE, regularised over a large training dataset



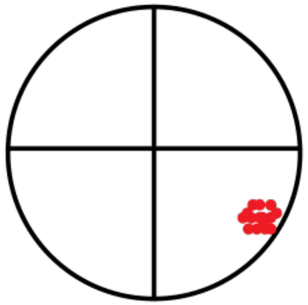
How do these methods compare?

Supervised methods

e.g. Bertleff 2017, Gyori 2022

Low variance: consistent parameter estimates under noise repetition

High bias: estimates biased away from groundtruth



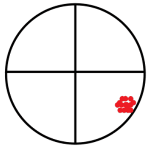
e.g. Gyori 2022, Grussu 2021

Low information content: bias depends on groundtruth, i.e. different groundtruths indistinguishable

How do these methods compare?

Supervised methods

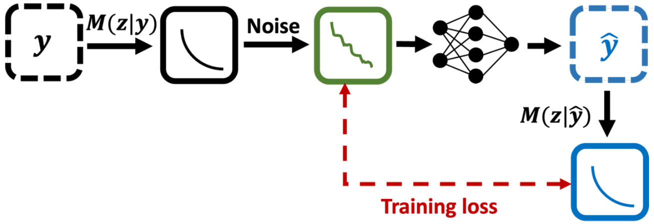
e.g. Bertleff 2017, Gyori 2022



e.g. Gyori 2022, Grussu 2021

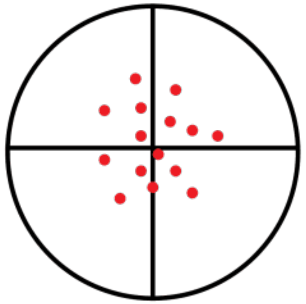
Self-supervised methods

e.g. Barbieri 2019, Kaandorp 2021



Lower bias: mean parameter estimates closer to groundtruth

High variance: wide parameter estimation distribution under noise



e.g. Grussu 2021, Barbieri 2019

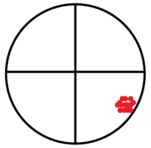
Loss calculated in signal-space:

- requires differentiable loss formulation (i.e. MSE, Gaussian noise assumption; limits signal models);
- relative parameter loss weighting limited by acquisition protocol z

Which method to use?

Supervised methods

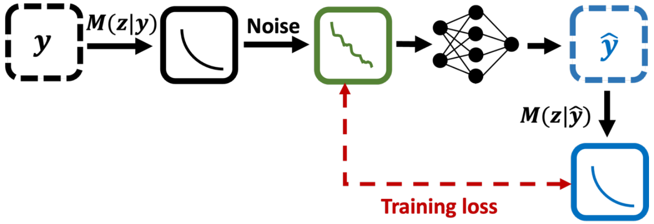
e.g. Bertleff 2017, Gyori 2022

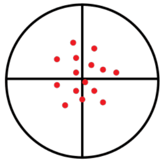


e.g. Gyori 2022, Grussu 2021

Self-supervised methods

e.g. Barbieri 2019, Kaandorp 2021





e.g. Grussu 2021, Barbieri 2019

Spectrum: bias/variance trade-off

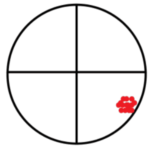
Self-supervised (low bias) has practical limitations

Would like to address these limitations and **move along this bias/variance spectrum**

Supervised training with a change of label

Supervised methods

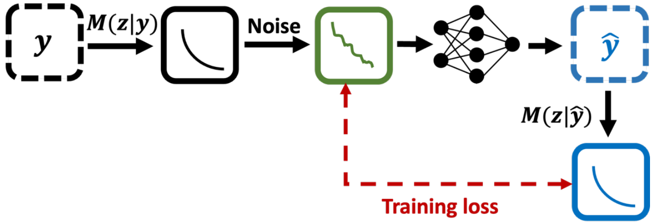
e.g. Bertleff 2017, Gyori 2022

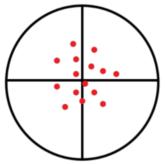


e.g. Gyori 2022, Grussu 2021

Self-supervised methods

e.g. Barbieri 2019, Kaandorp 2021

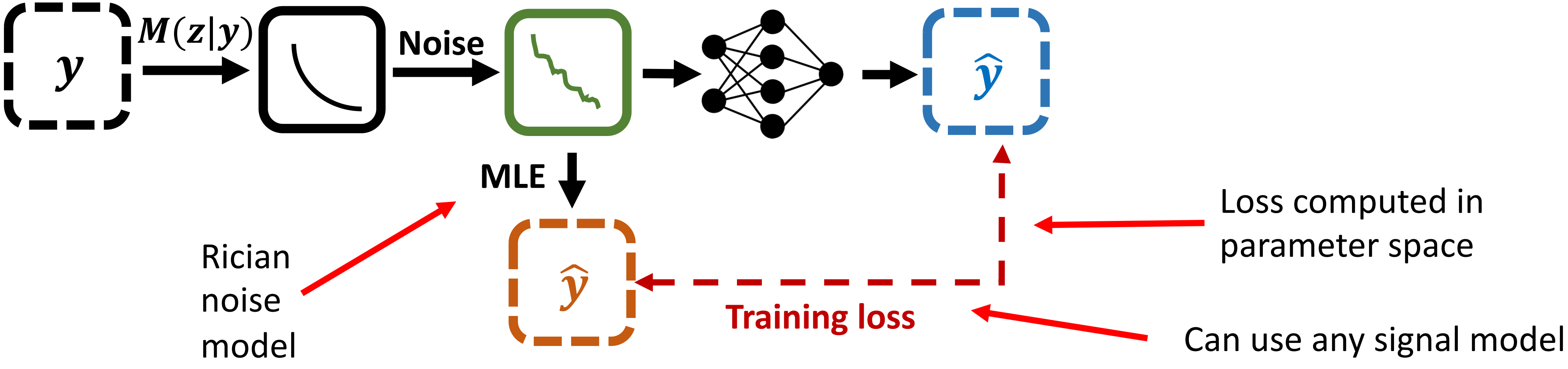




e.g. Grussu 2021, Barbieri 2019

Supervised methods (MLE labels)

Epstein 2022

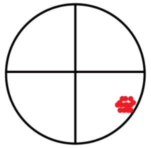


$$\text{Hybrid loss} = \alpha \cdot \text{Supervised}_{\text{MLE}}\ \text{loss} + (1-\alpha) \cdot \text{Supervised}_{\text{GT}}\ \text{loss}$$

Summary of methods

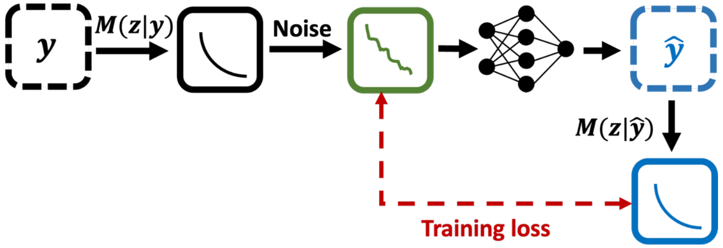
Supervised methods (groundtruth labels)

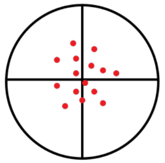
e.g. Bertleff 2017, Gyori 2022



Self-supervised methods

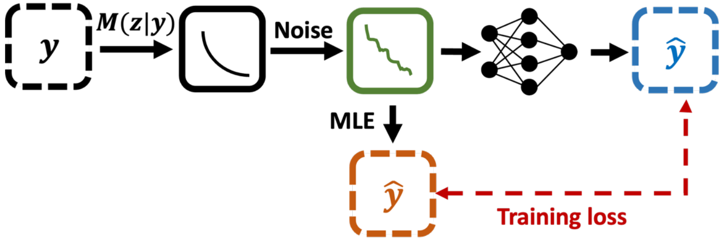
e.g. Barbieri 2019, Kaandorp 2021

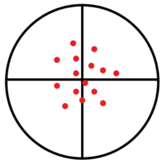




Supervised methods (MLE labels)

e.g. Epstein 2022



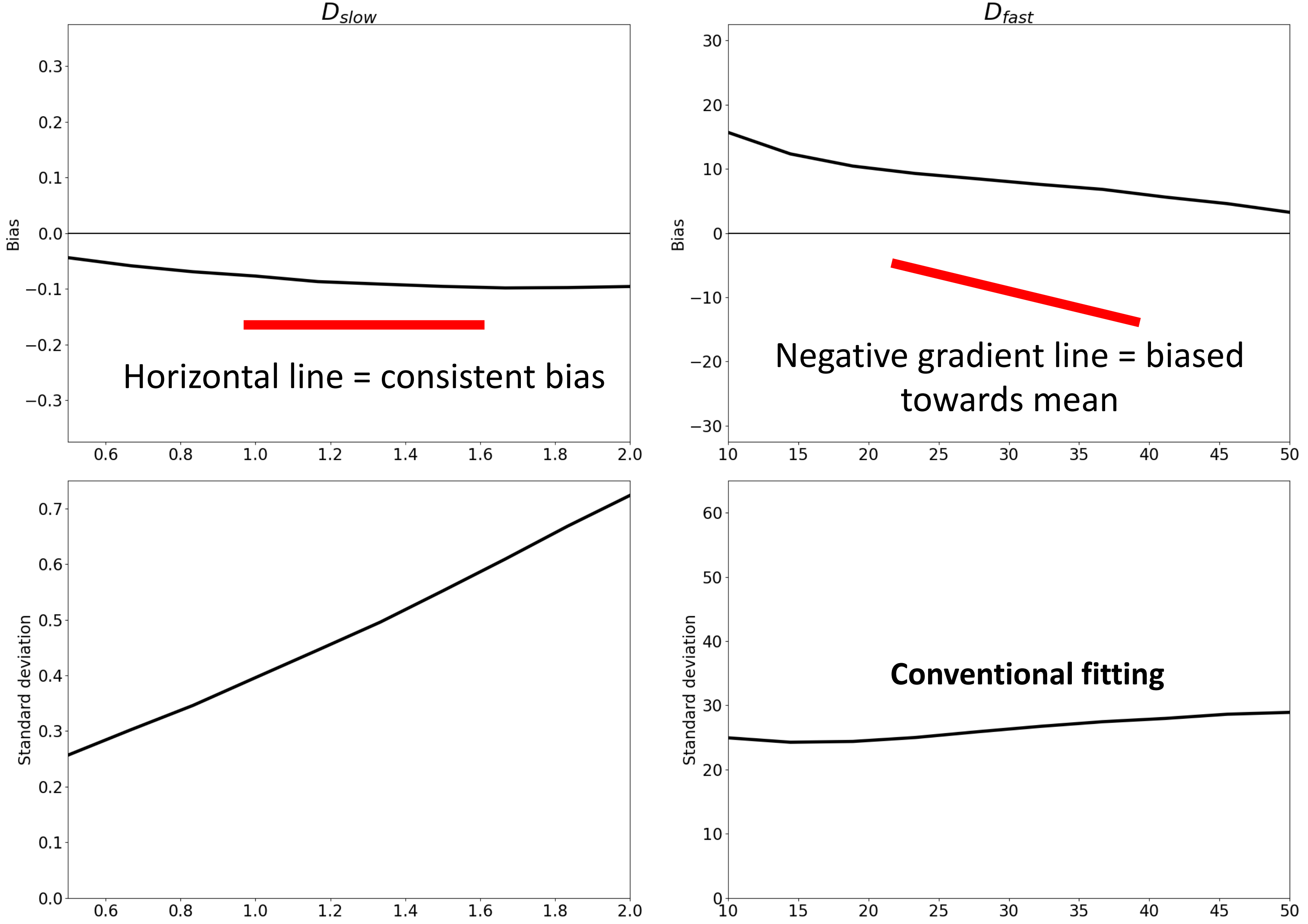


y_1: easy to fit

y_2: difficult to fit

BIAS

VARIANCE

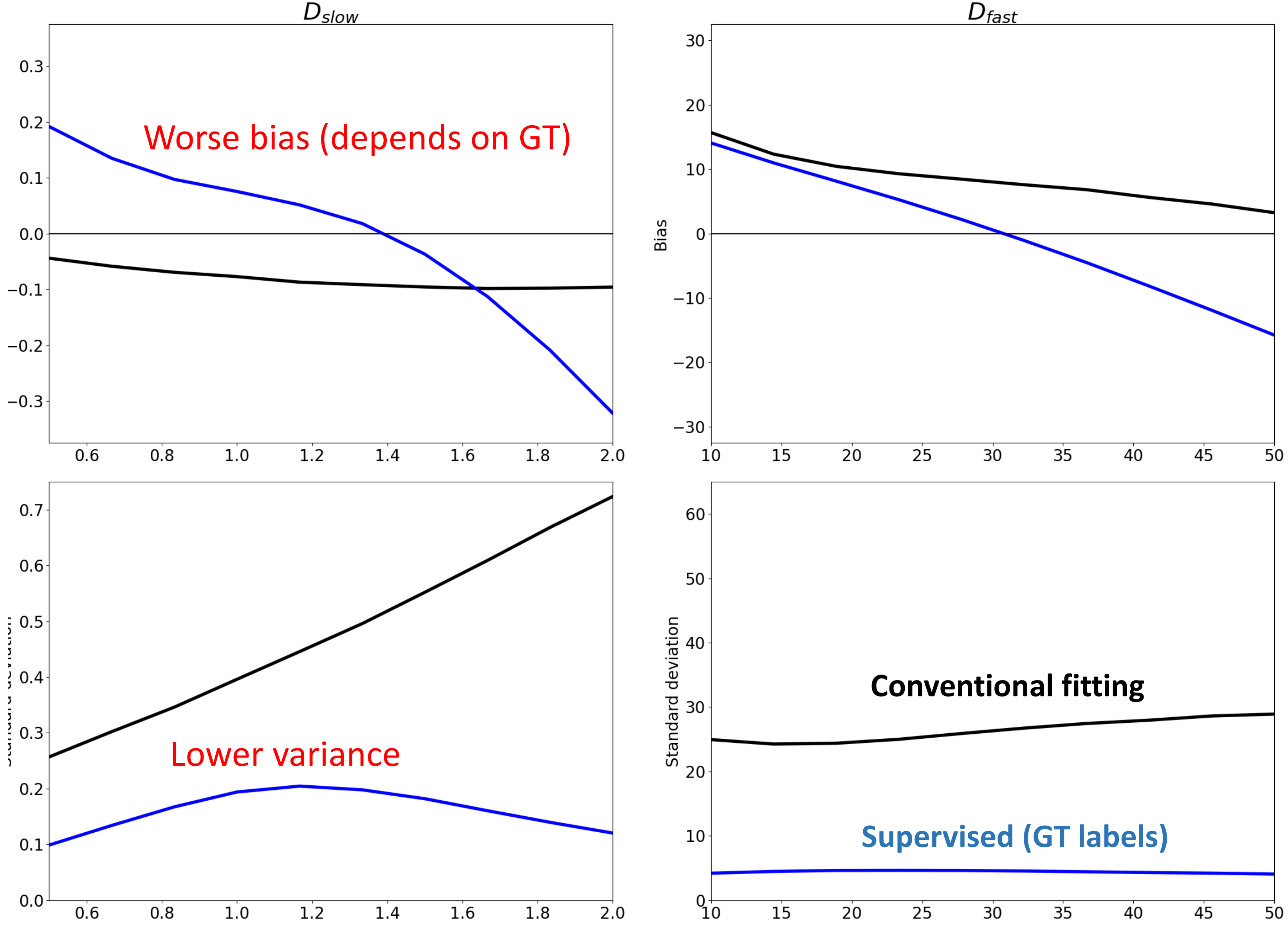


y_1: easy to fit

y_2: difficult to fit

BIAS

VARIANCE

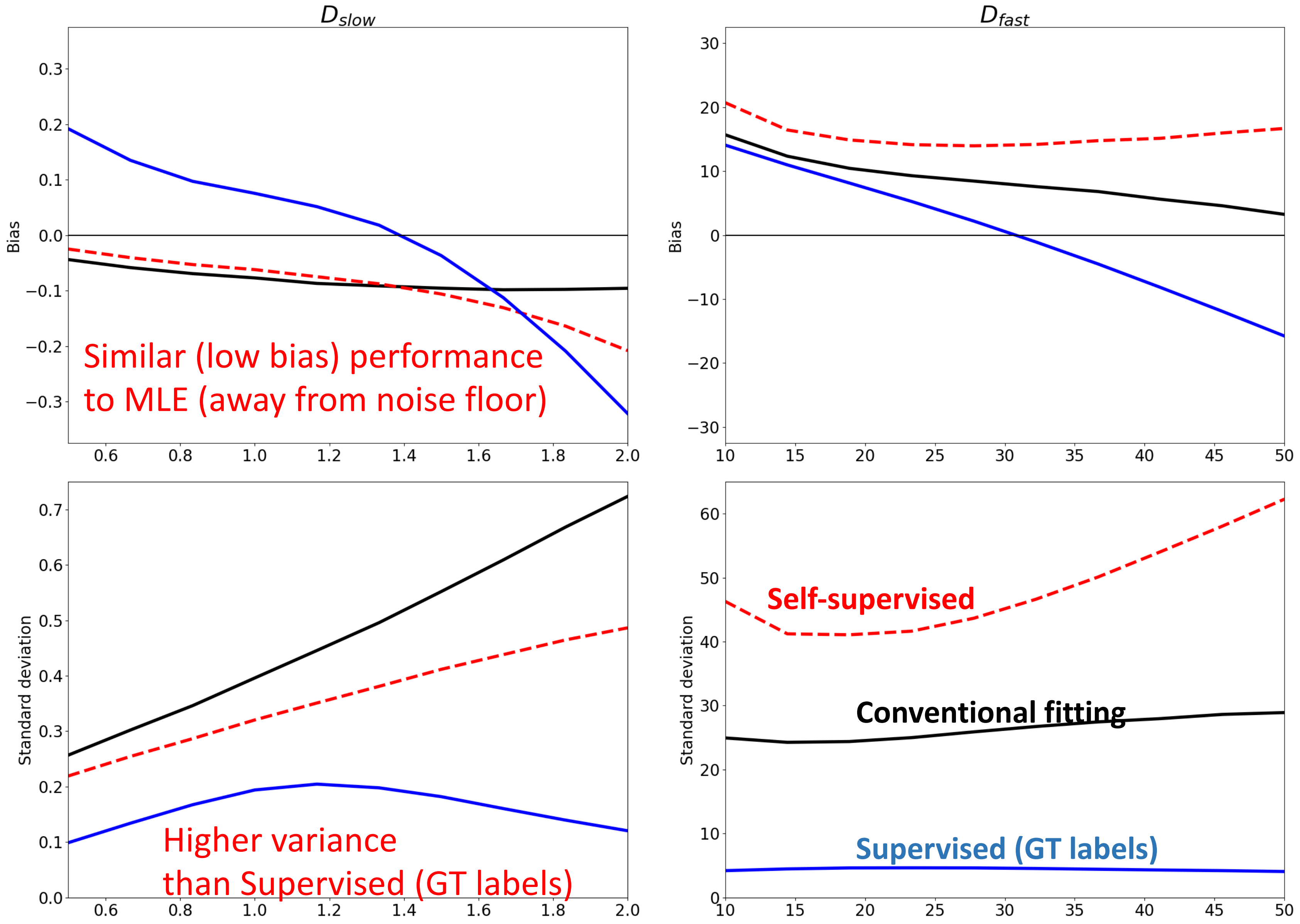


y_1: easy to fit

y_2: difficult to fit

BIAS

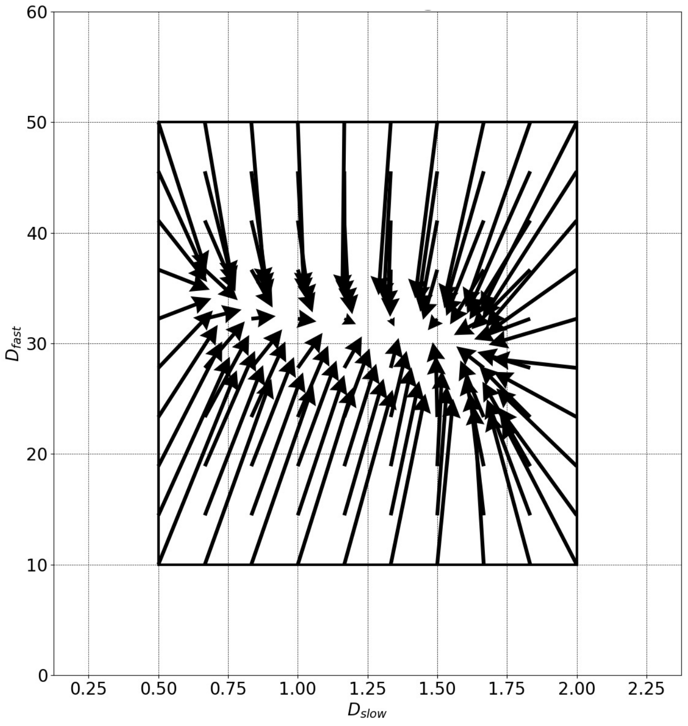
VARIANCE



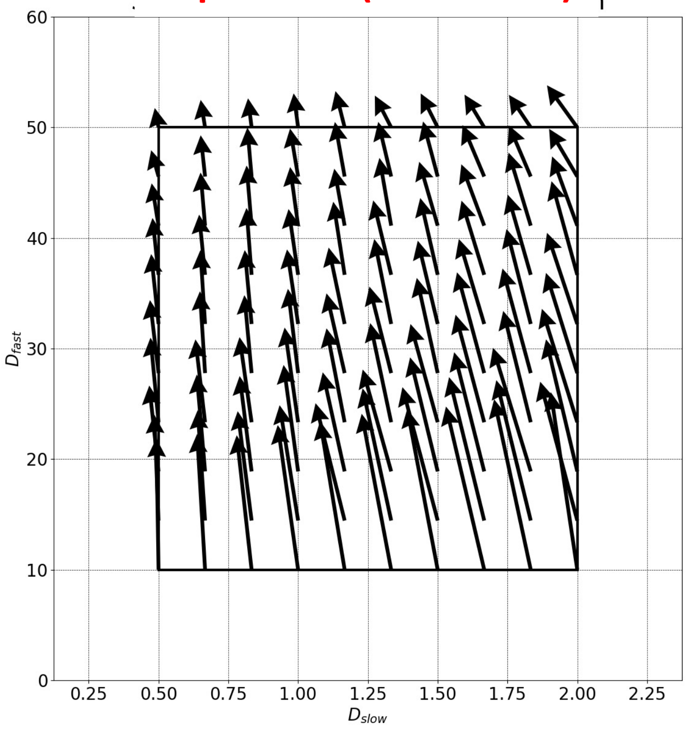
Worse y_2 performance: underweight in signal space

y_1: easy to fit
y_2: difficult to fit
BIAS
VARIANCE
D_{slow}
D_{fast}
Bias
Standard deviation
Similar bias
Better performance on y_2
(equal weighting)
Similar variance
Self-supervised
Conventional fitting
Supervised (MLE labels)
Supervised (GT labels)
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
30
20
10
0
−10
−20
−30
15
25
35
40
45
50
0.7
0.5
0.4
60

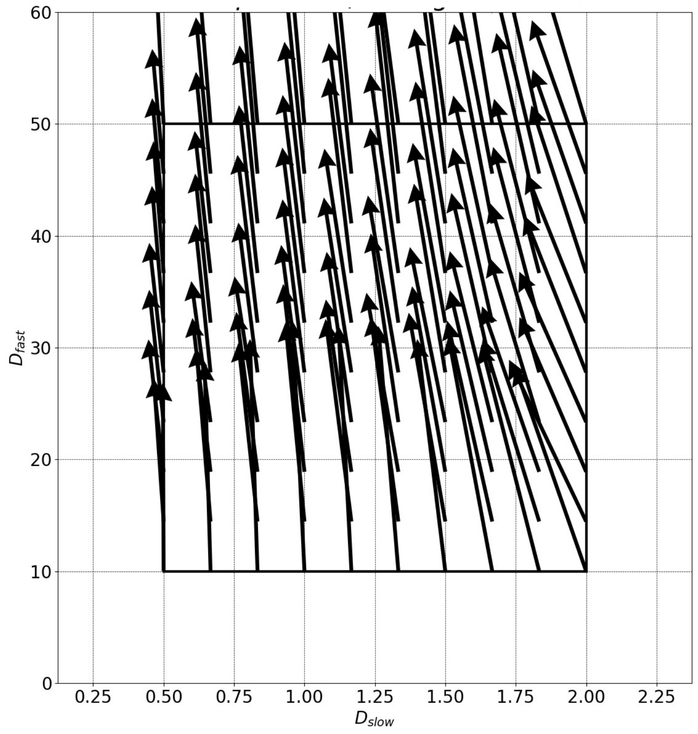
Supervised (GT labels)

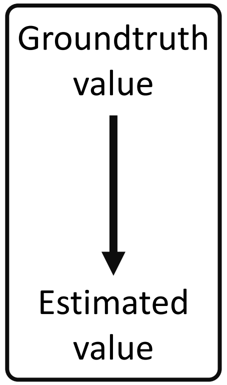


Supervised (MLE labels)



Self-supervised

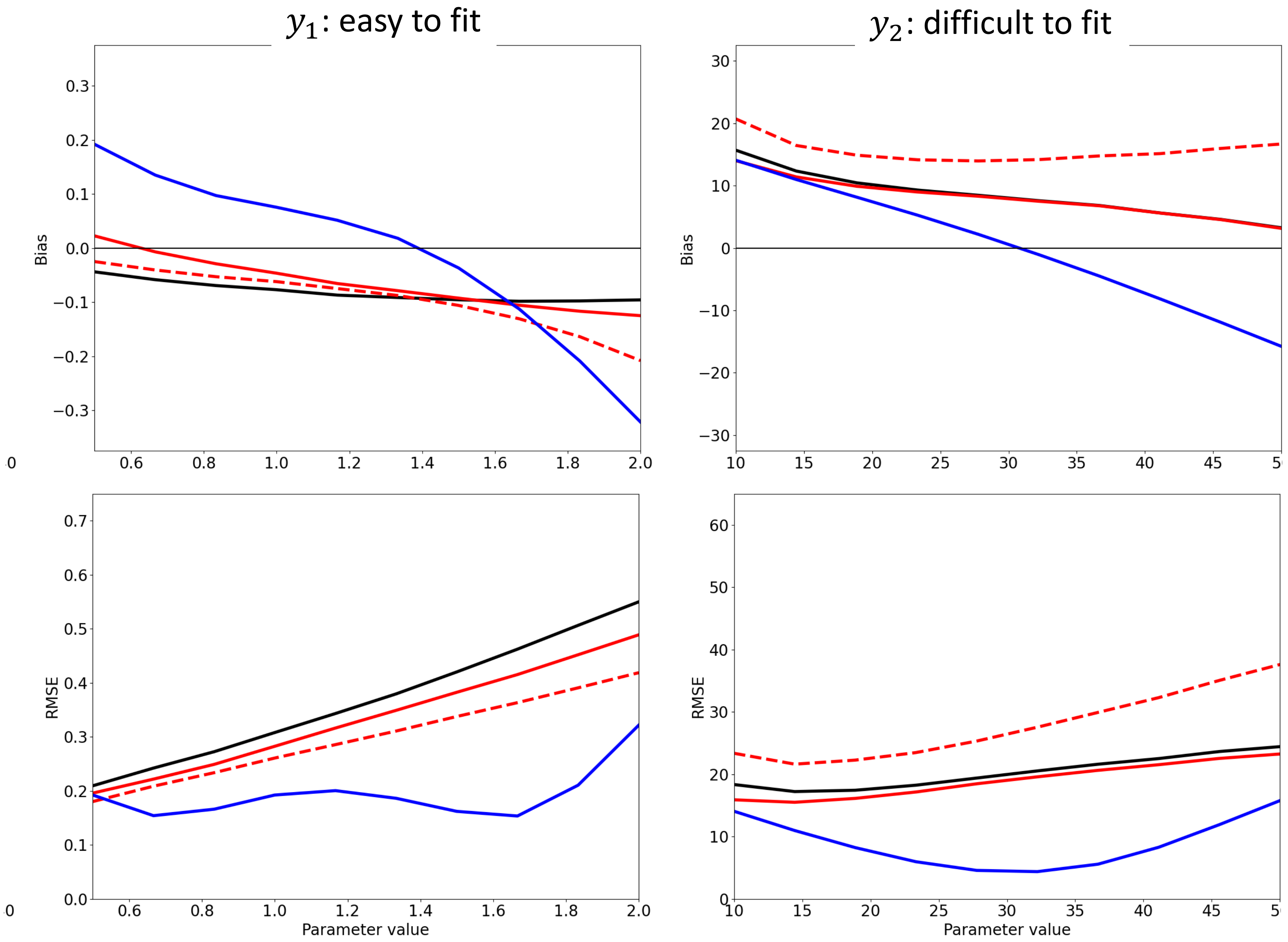




see Gyori 2022 for more examples

BIAS

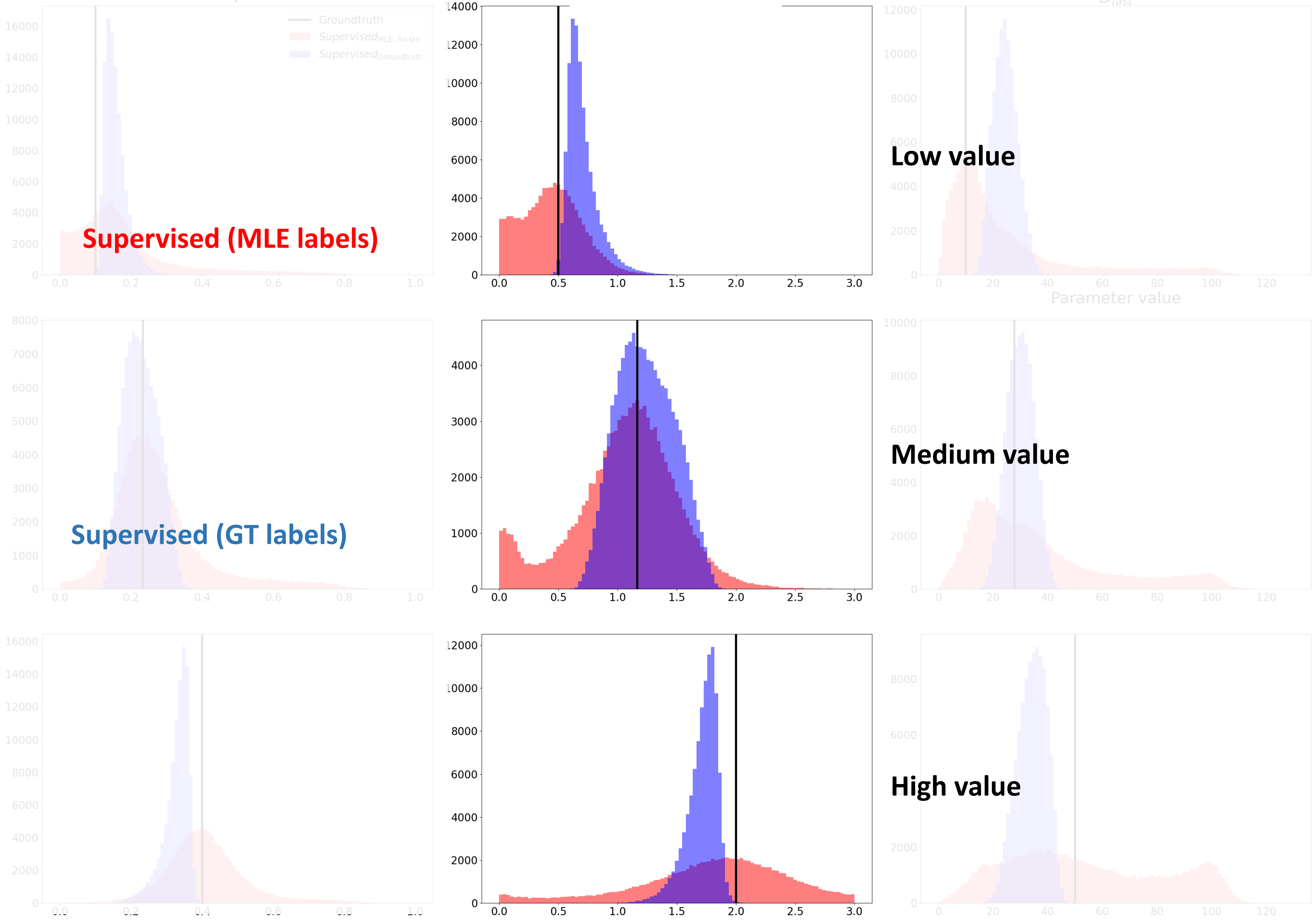
RMSE



Bias worse

RMSE better

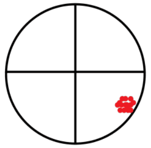
y_1: easy to fit



Is there a middle ground?

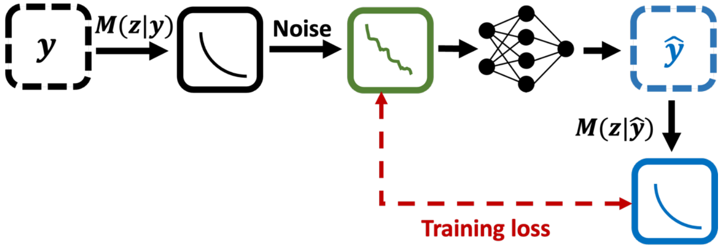
Supervised methods (groundtruth labels)

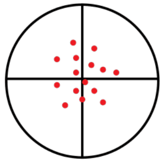
e.g. Bertleff 2017, Gyori 2022



Self-supervised methods

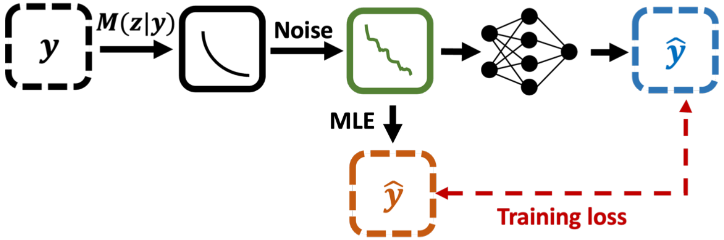
e.g. Barbieri 2019, Kaandorp 2021

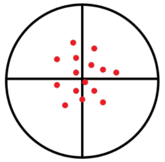




Supervised methods (MLE labels)

e.g. Epstein 2022

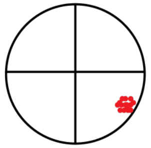




Is there a middle ground?

Supervised methods (groundtruth labels)

e.g. Bertleff 2017, Gyori 2022

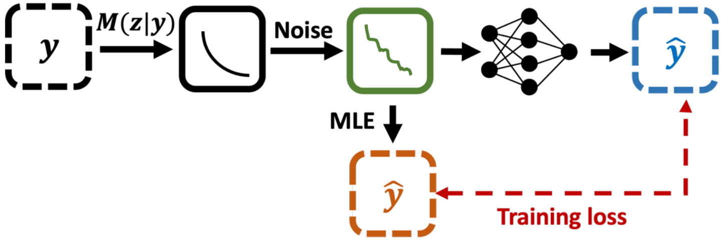


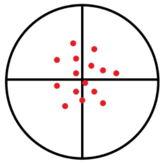
Hybrid loss function during training:

$$\text{Hybrid loss} = \alpha \cdot \text{Supervised}_{\text{MLE}} \text{ loss} + (1 - \alpha) \cdot \text{Supervised}_{\text{GT}} \text{ loss}$$

Supervised methods (MLE labels)

e.g. Epstein 2022



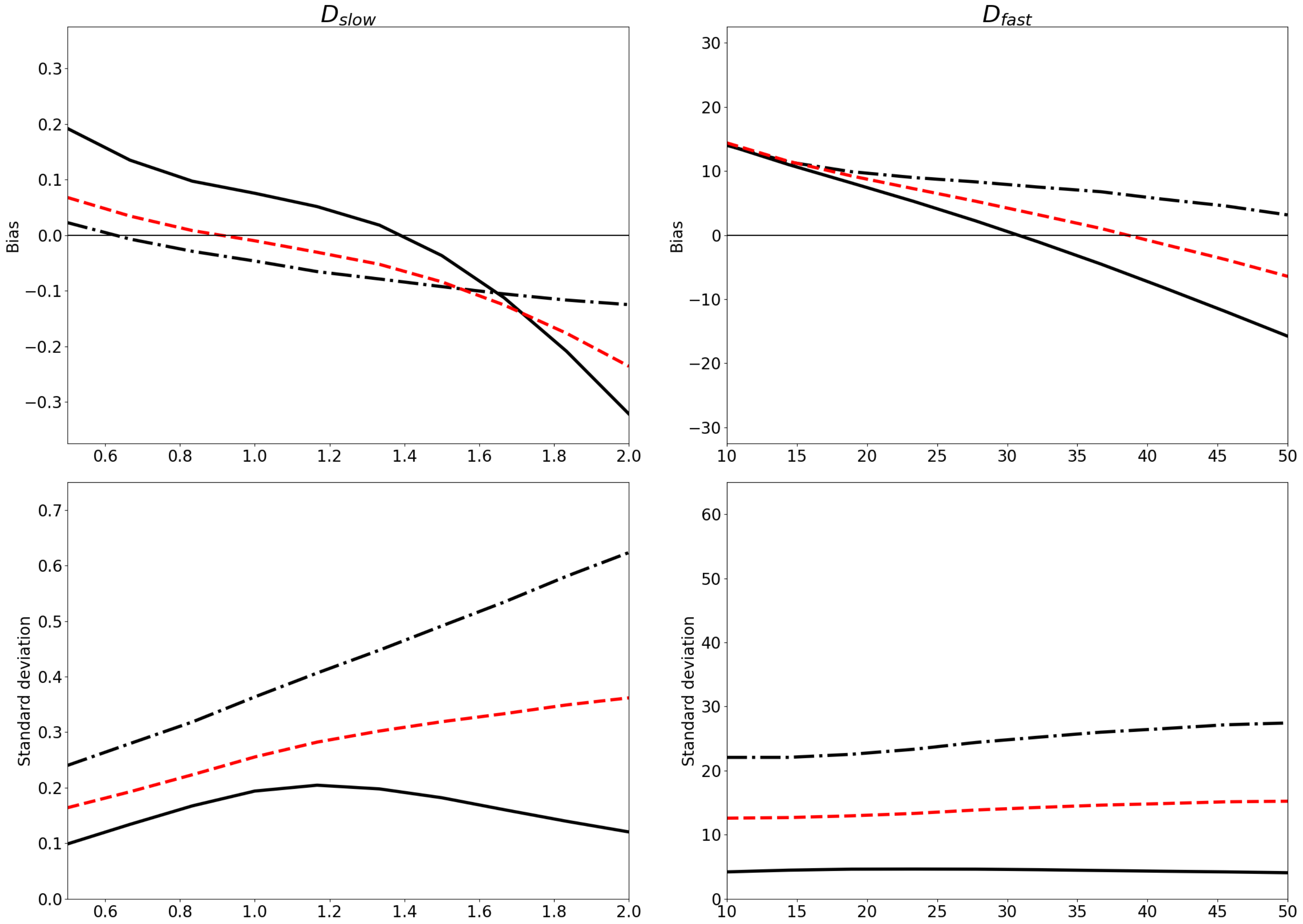


y_1: easy to fit

y_2: difficult to fit

BIAS

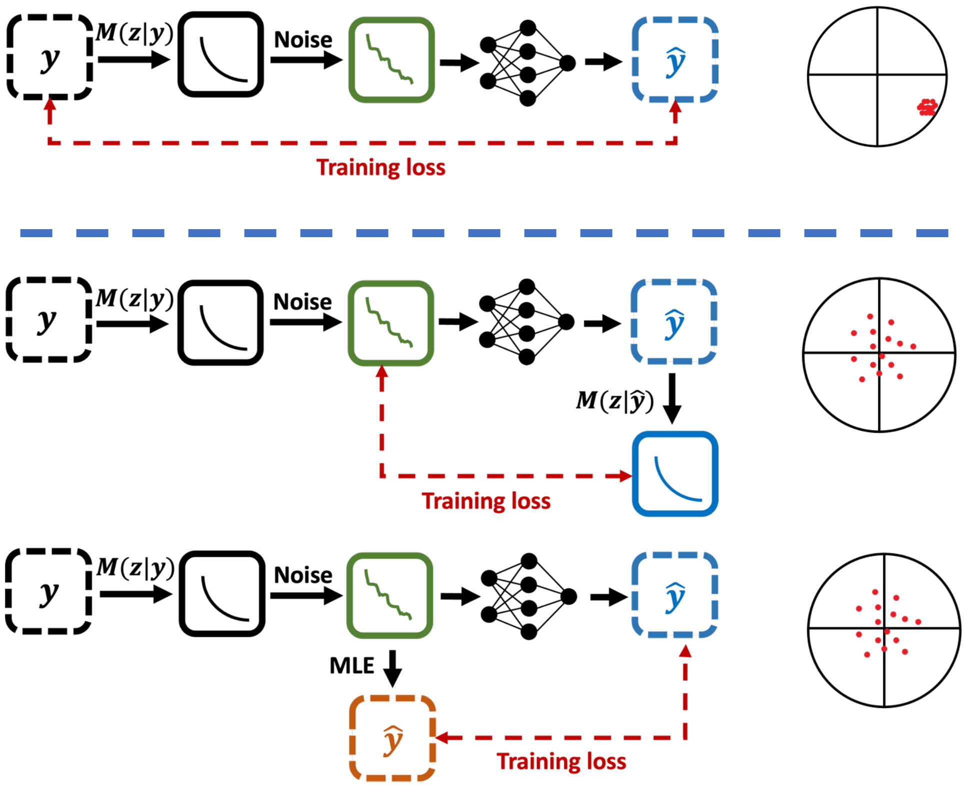
VARIANCE



Supervised methods (groundtruth labels)
e.g. Bertleff 2017, Gyori 2022

Self-supervised methods
e.g. Barbieri 2019, Kaandorp 2021

Supervised methods (MLE labels)
Epstein 2022



- ***Bias/variance tradeoff***
- **Look beyond RMSE**: misleading summary metric
- Supervised training has **practical advantages** over self-supervised
- **Don't always use GT labels** – even if you have access to them
- ***Can adjust network performance by tailoring contribution of different labels***

Tim Bray

Margaret Hall-Craggs

Gary Zhang

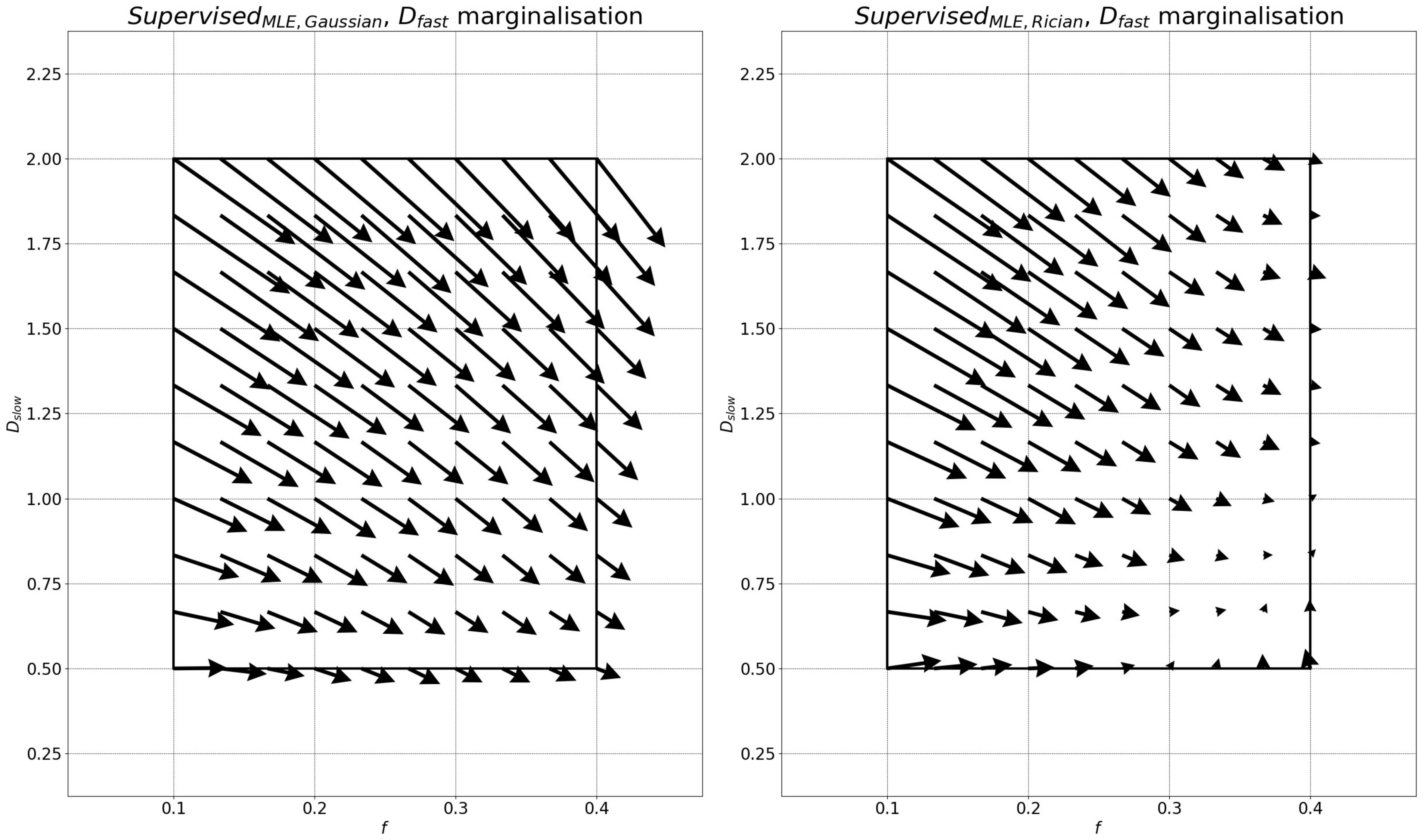
arXiv:2205.05587

Choice of training label matters: how to best use deep learning for quantitative MRI parameter estimation

NIHR | University College London Hospitals Biomedical Research Centre

i4health
intelligent integrated imaging

UKRI Engineering and Physical Sciences Research Council



$Supervised_{MLE,Gaussian}$, D_{fast} marginalisation
$Supervised_{MLE,Rician}$, D_{fast} marginalisation
D_{slow}
f
2.25
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.1
0.2
0.3
0.4