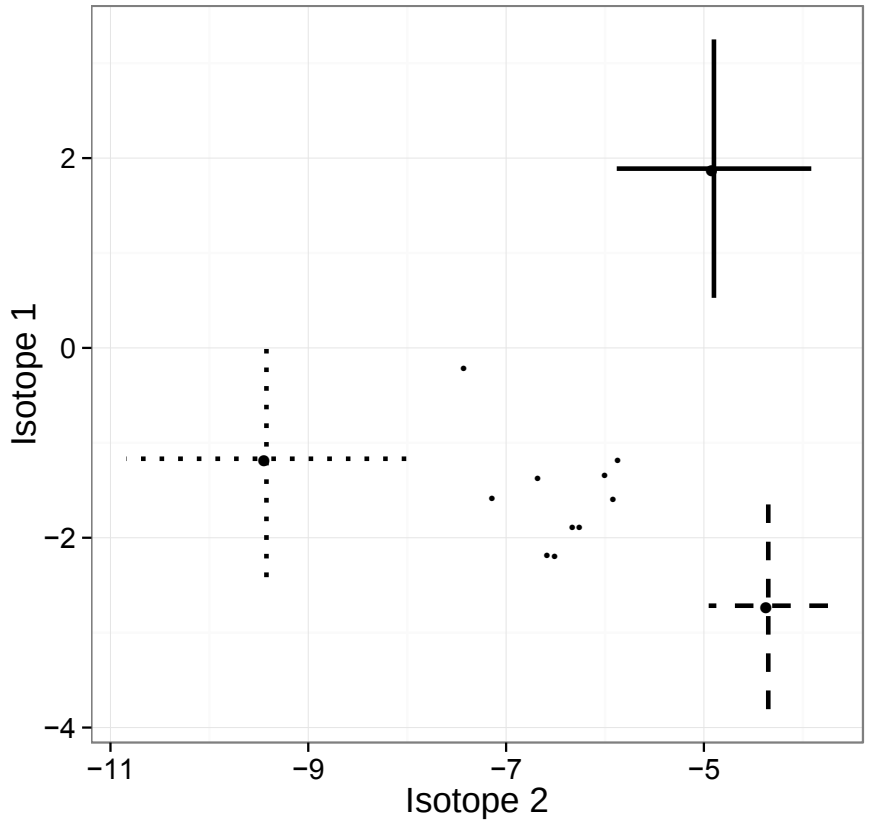
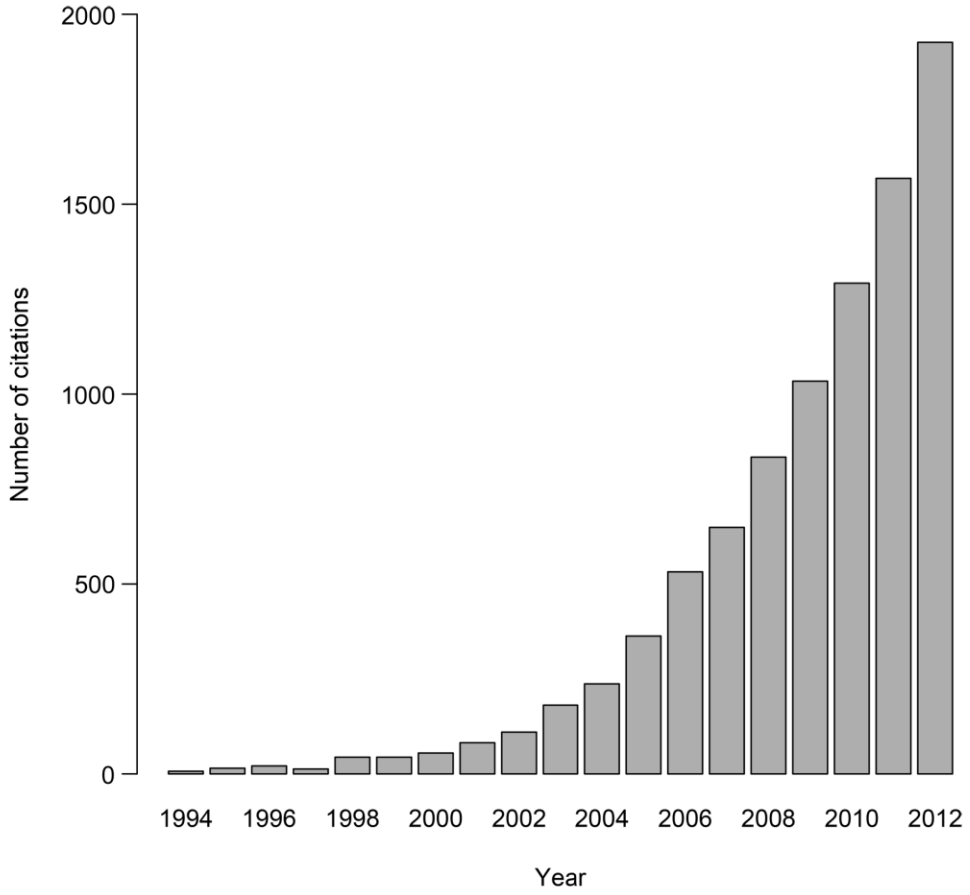
Use and Abuse of Mixing Models

BRIAN STOCK, BRICE SEMMENS, ERIC WARD, ANDREW PARNELL, ANDREW JACKSON, DONALD PHILLIPS, STUART BEARHOP, RICHARD INGER

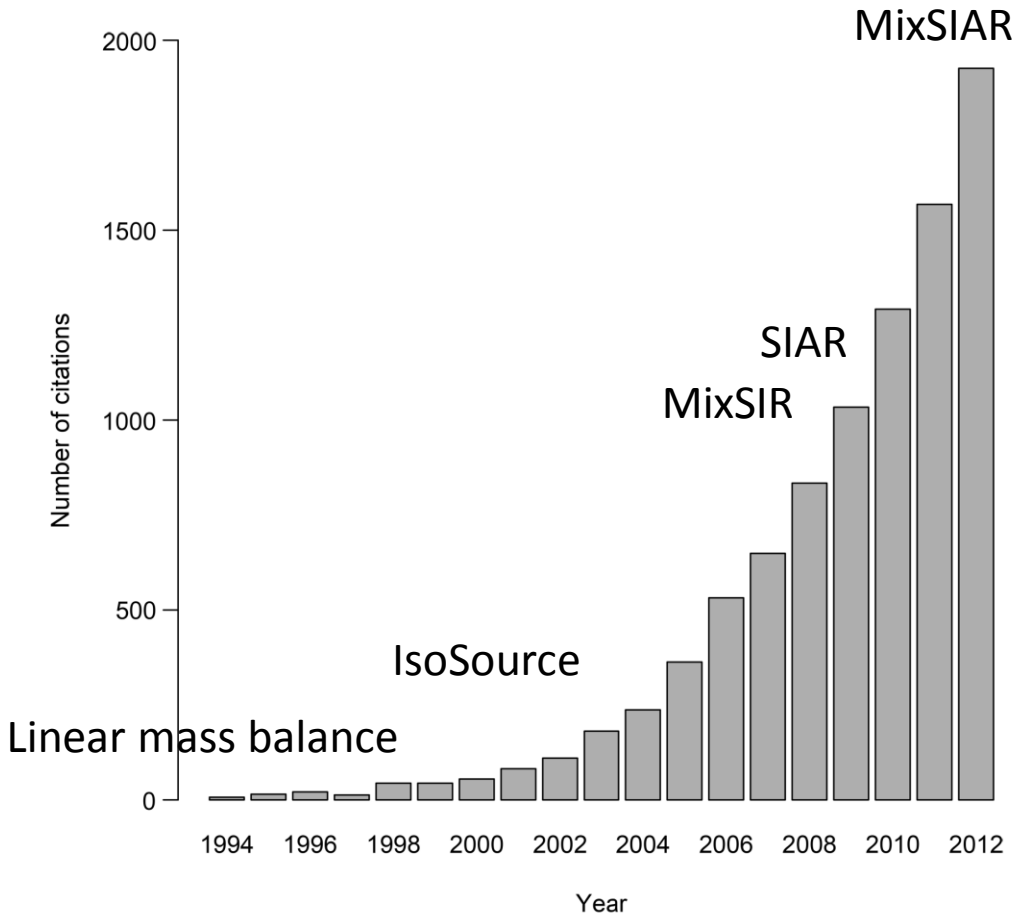
People use mixing models a lot



People use mixing models a lot



People use mixing models a lot



Pros of Bayesian mixing models

Firm statistical foundation

- True probability distributions
- Uncertainty in consumer, source, and TDF data

Biological complexity

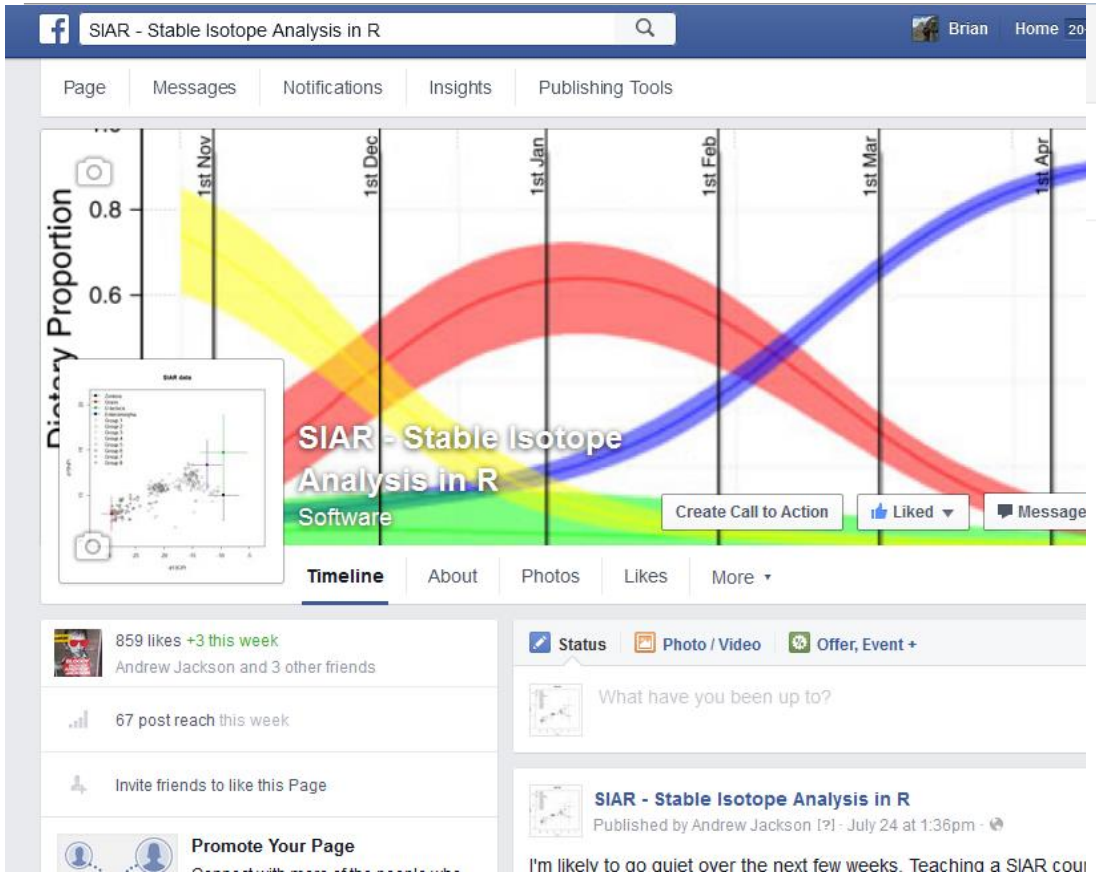
- Differences due to covariates (e.g. sex, region, size)
- Non-biotracer data as priors (e.g. stomach/fecal contents, prey abundance)

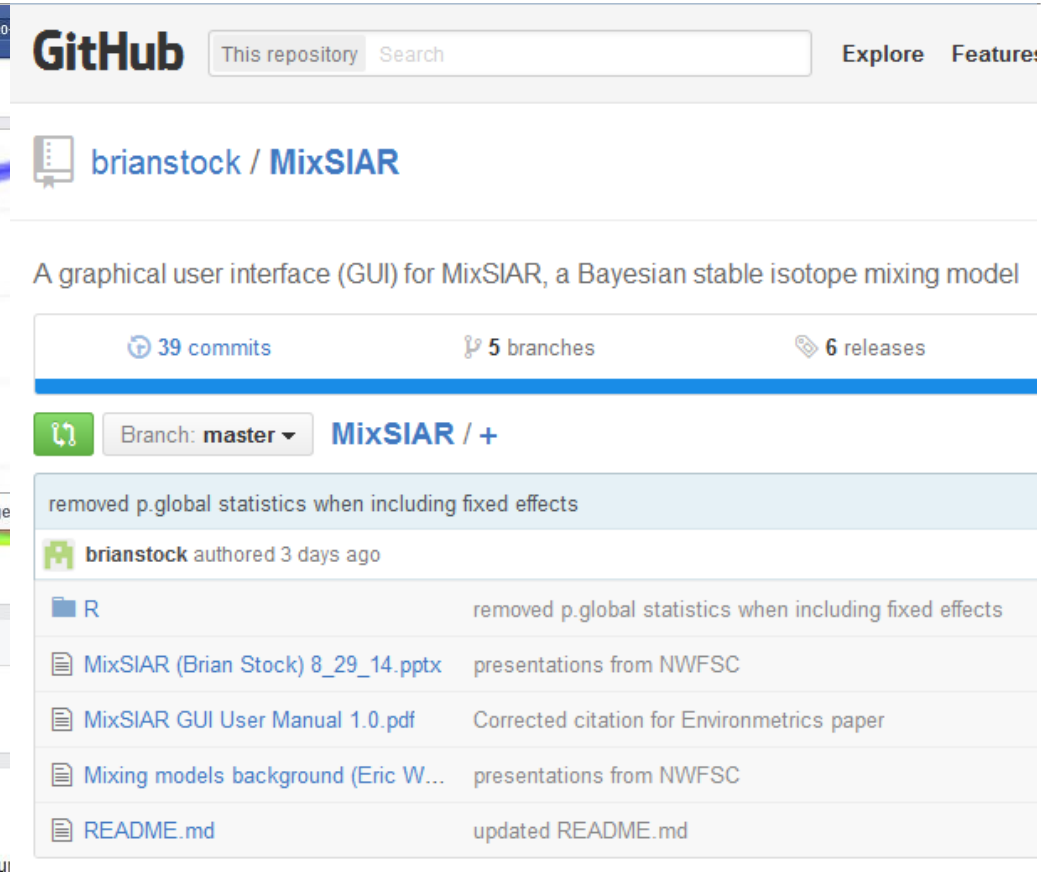
Cons of Bayesian mixing models

They’re more complex and prone to abuse

Garbage in, garbage out (ex. many sources, 2 tracers)

Lots of questions





Pitfalls and misconceptions

1. Source geometry
2. MCMC convergence
3. Effect of priors
4. Error structures
5. Source lumping/splitting
6. How to include covariates
7. Application to biotracers other than stable isotopes

Pitfalls and misconceptions

1. Source geometry
2. MCMC convergence
3. **Effect of priors**
4. Error structures
5. Source lumping/splitting
6. How to include covariates
7. **Application to biotracers other than stable isotopes**

Effect of priors/ “Bayesian mixing models are biased”

REDUCE THE INFLUENCE OF THE GENERALIST PRIOR

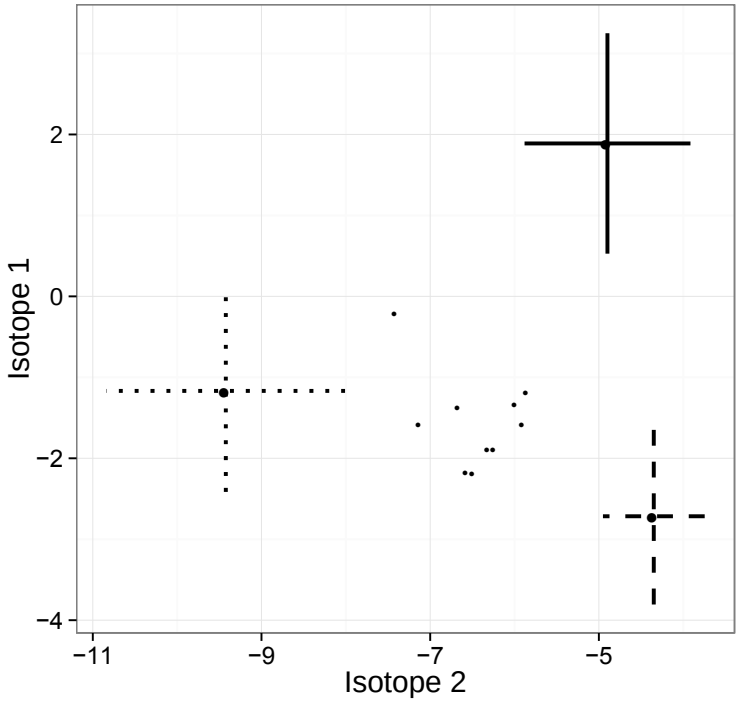
0. What is a prior?

$$Pr(\theta|data) \propto Pr(\theta) * Pr(data|\theta)$$

Posterior Prior Likelihood

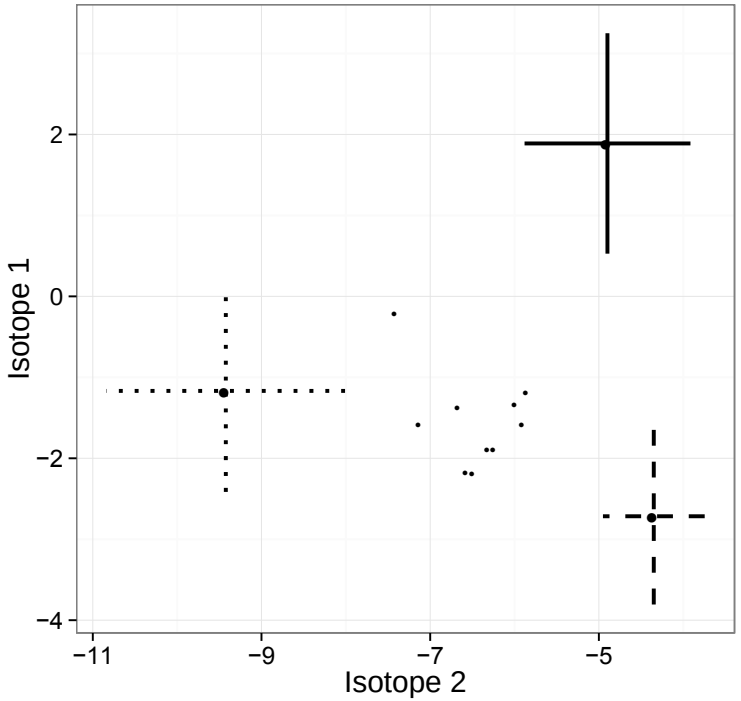
"From a Bayesian perspective, the principle of unbiasedness is reasonable in the limit of large samples, but otherwise it is potentially misleading."

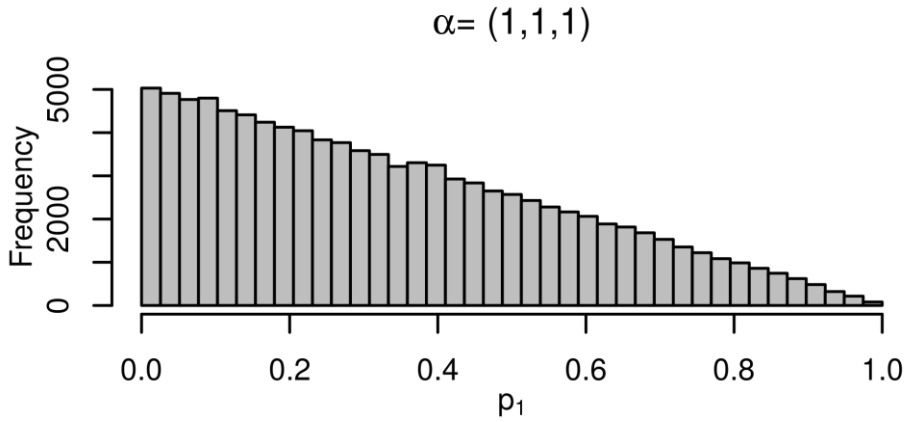
1. There is no “uninformative” prior



1. There is no “uninformative” prior

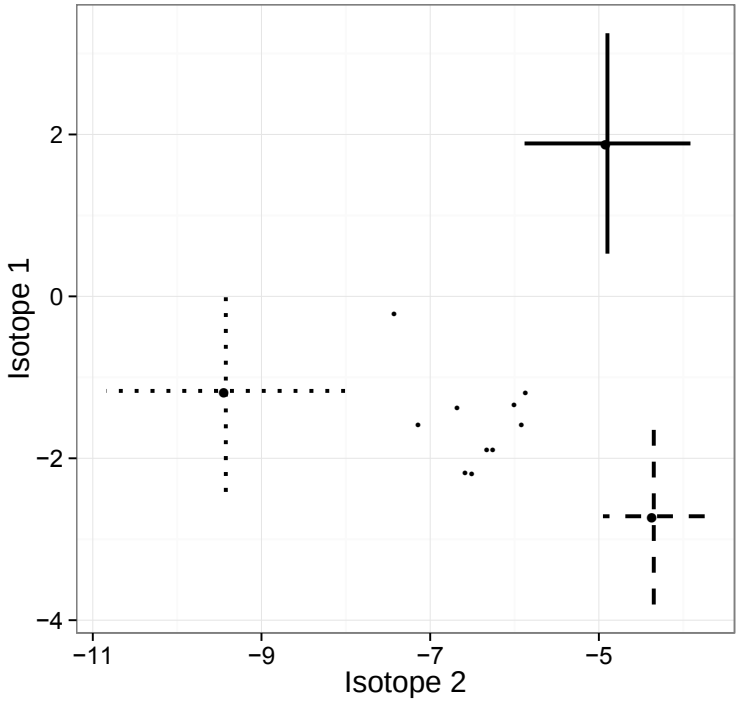
Problem: proportions are not independent!

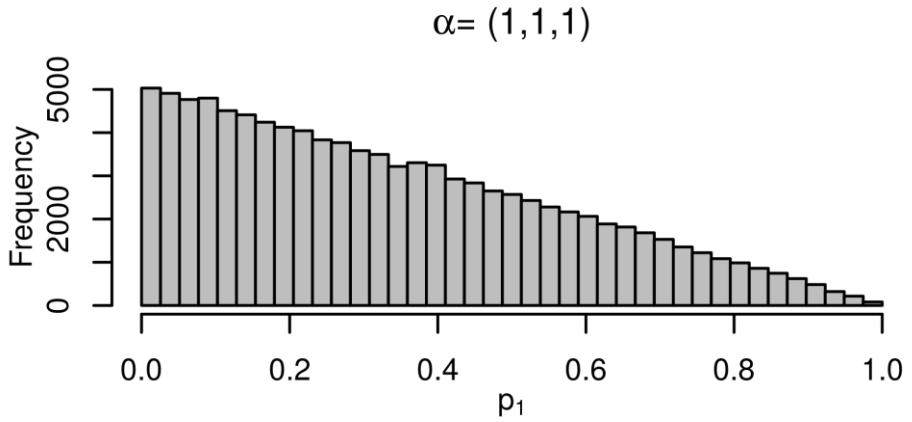


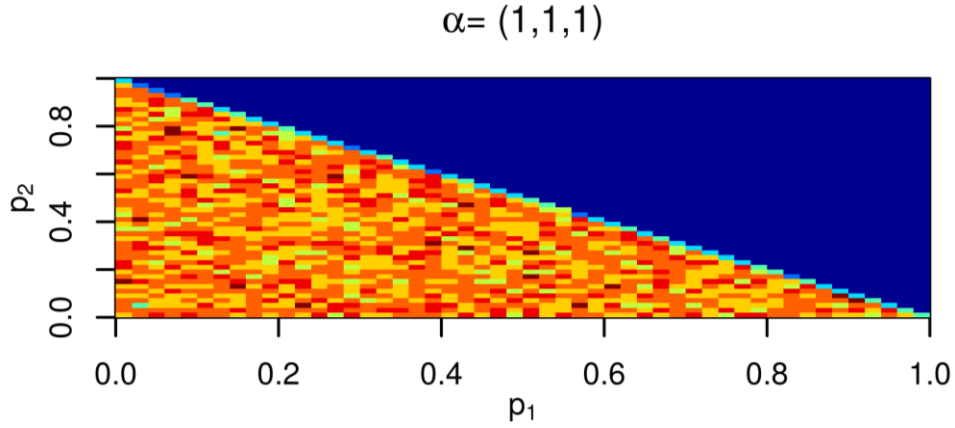


1. There is no “uninformative” prior

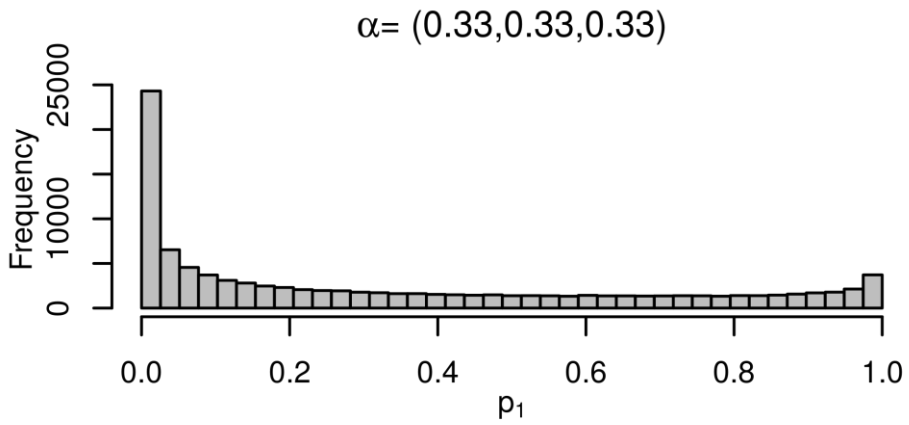
Problem: proportions are not independent!

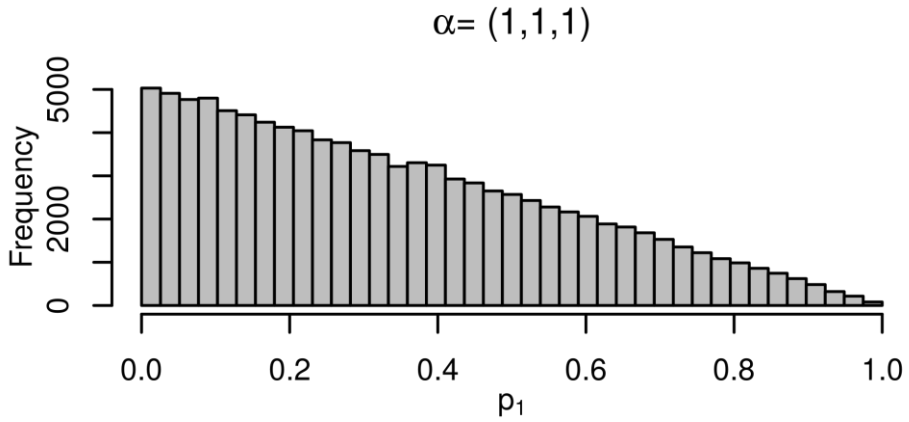


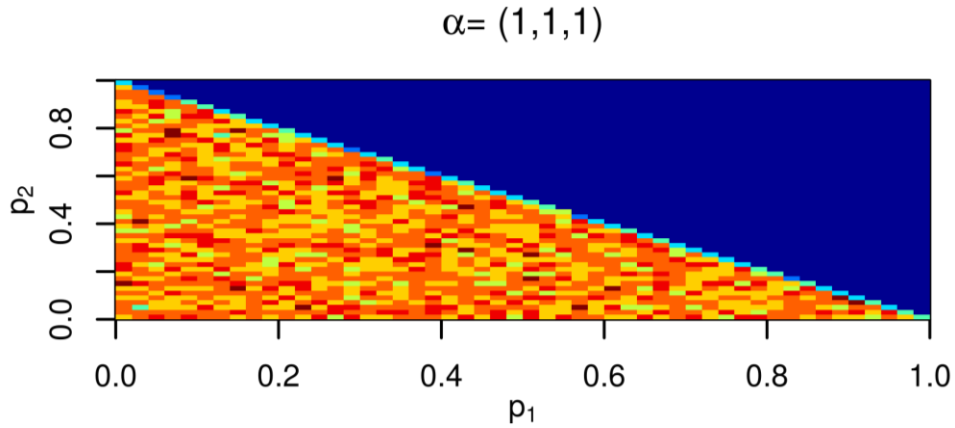




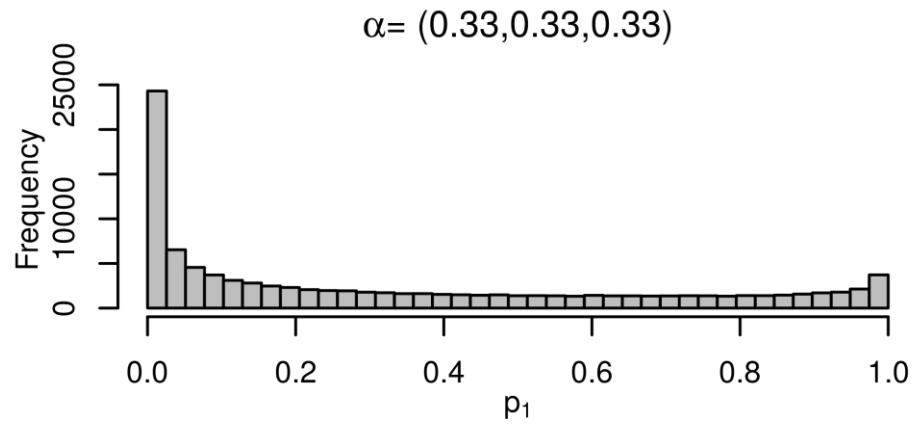
1. There is no “uninformative” prior

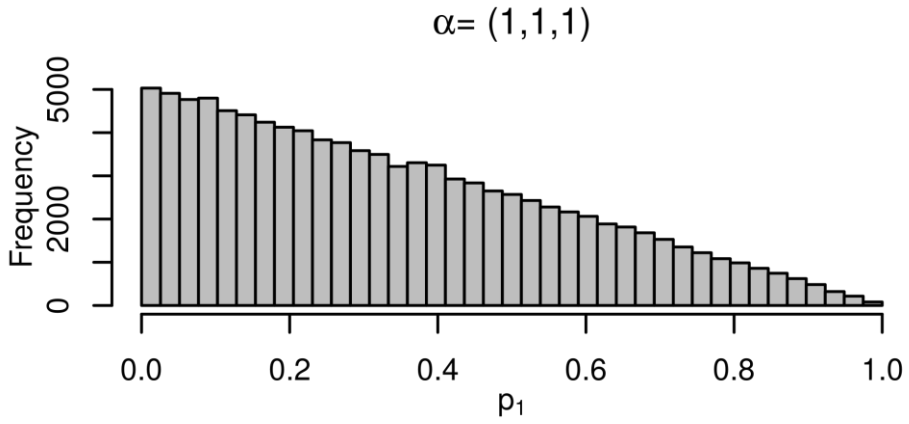


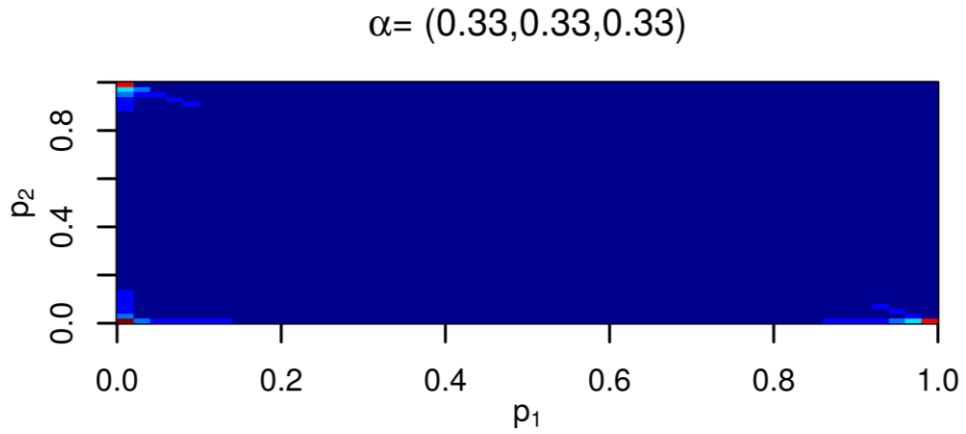


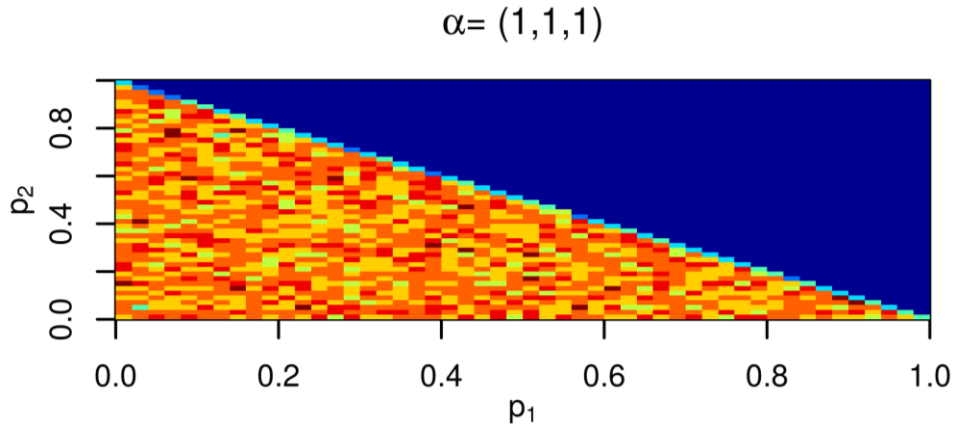


1. There is no “uninformative” prior



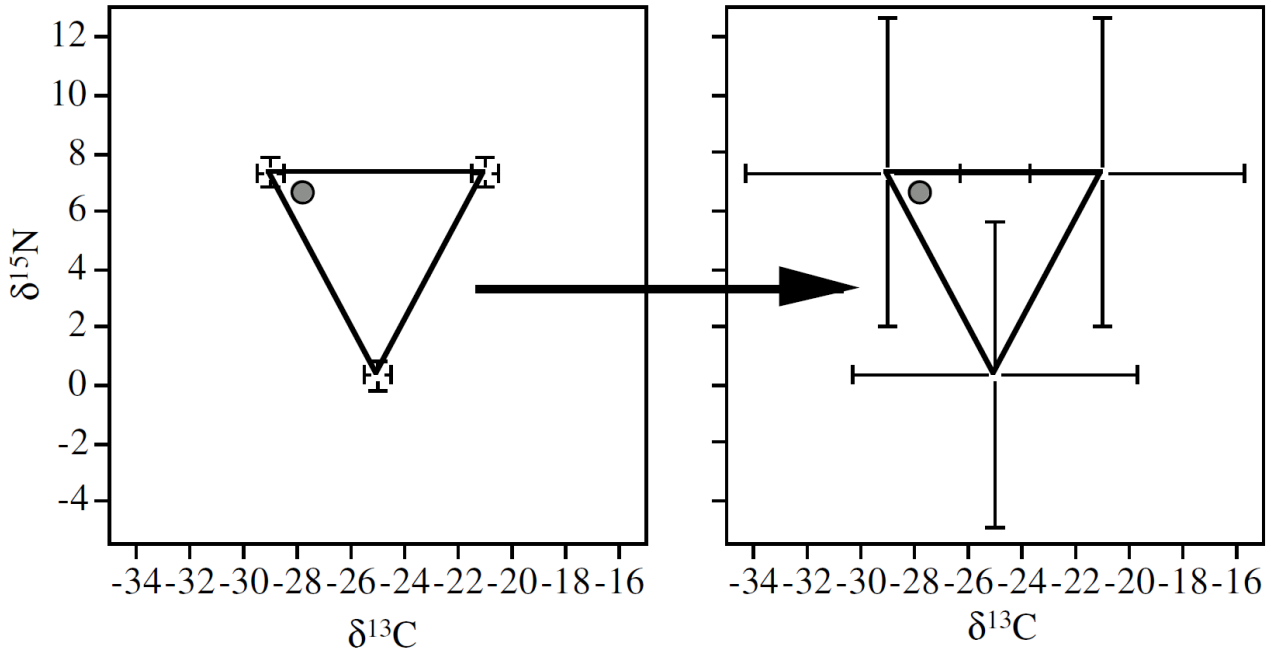






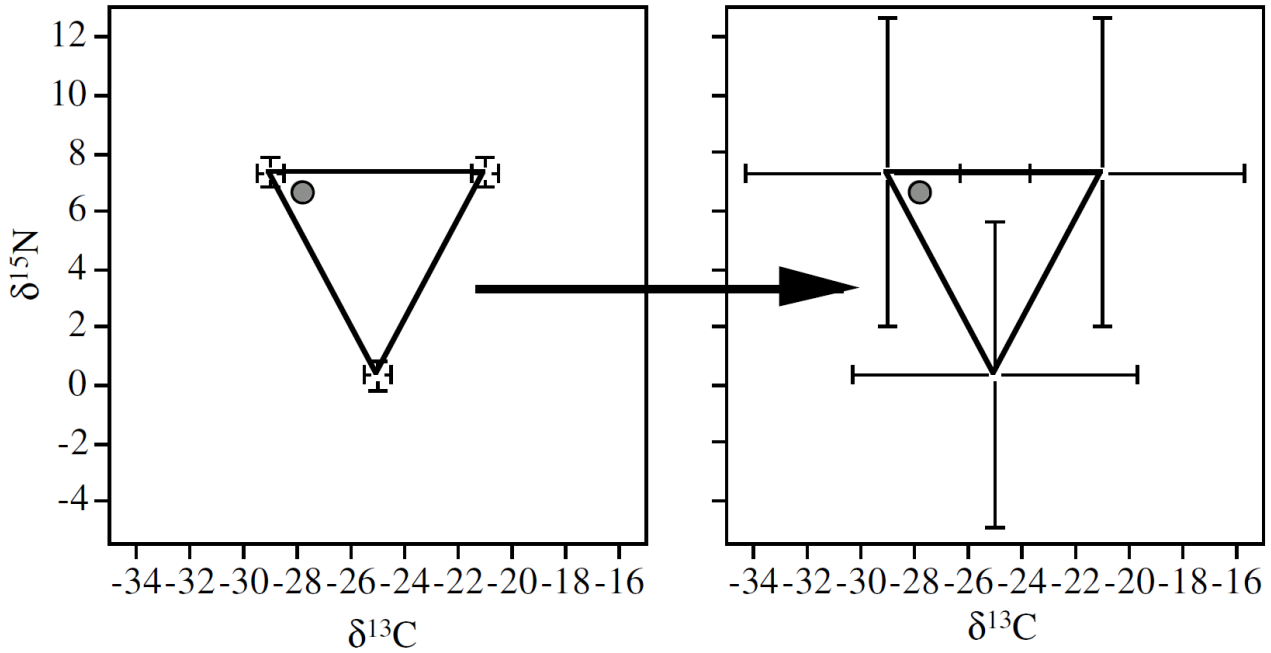
2. Effect of the “uninformative” prior

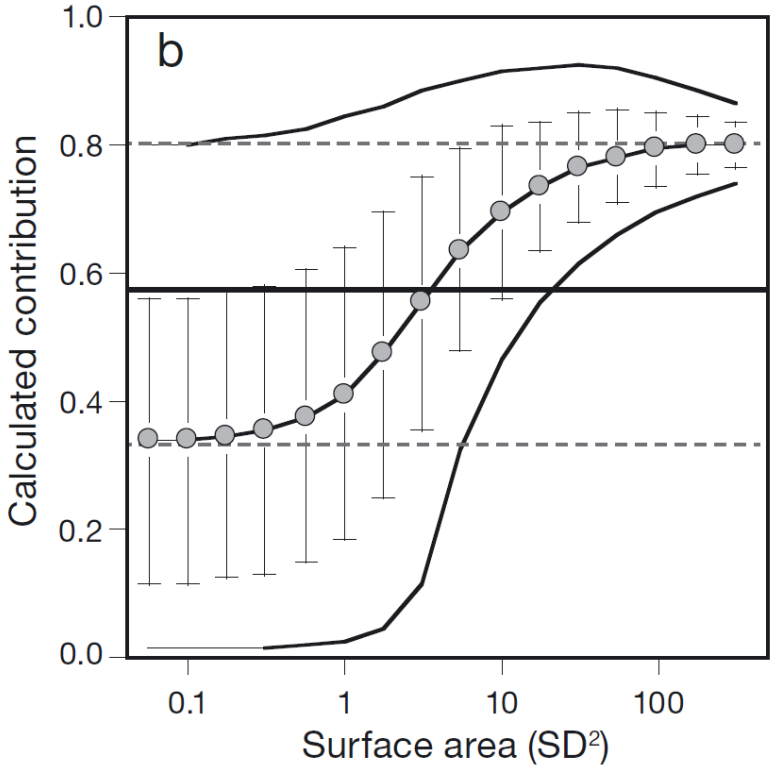
1. How good is your data?



2. Effect of the “uninformative” prior

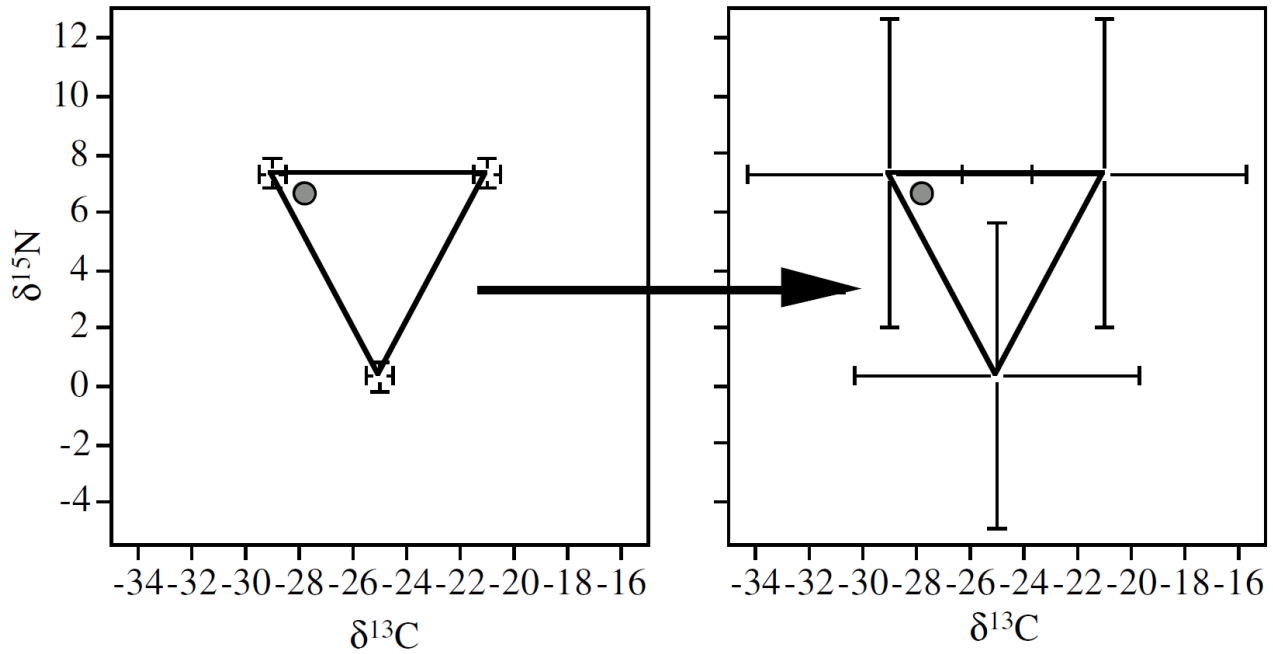
1. How good is your data?



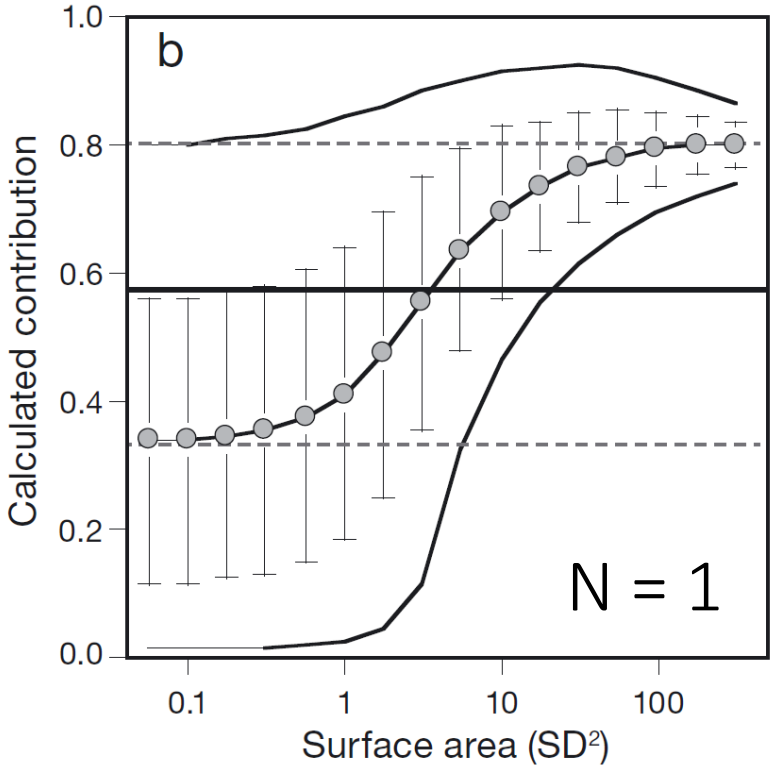


2. Effect of the “uninformative” prior

1. How good is your data?

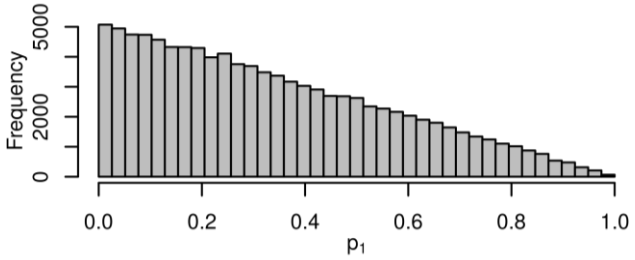


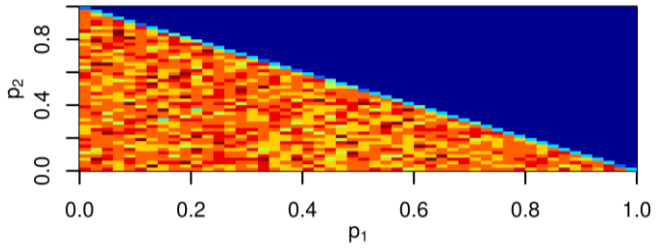
2. How much data do you have?



3. Constructing informative priors

You control the mean proportions AND the variance (“informativeness”)

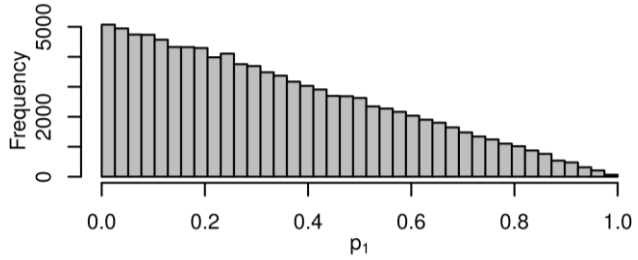


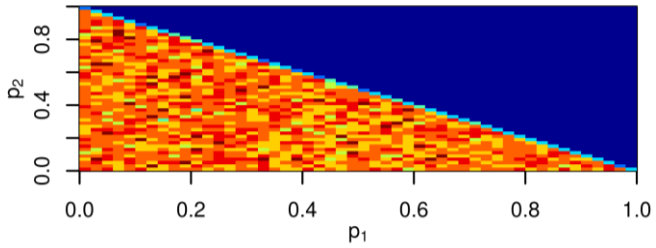


$\alpha = (1, 1, 1)$

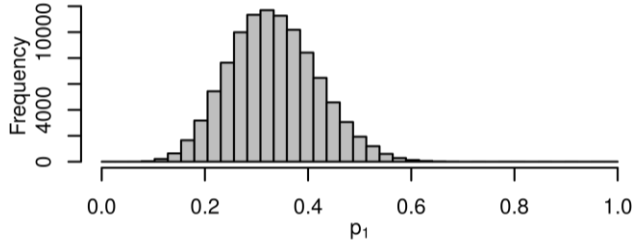
3. Constructing informative priors

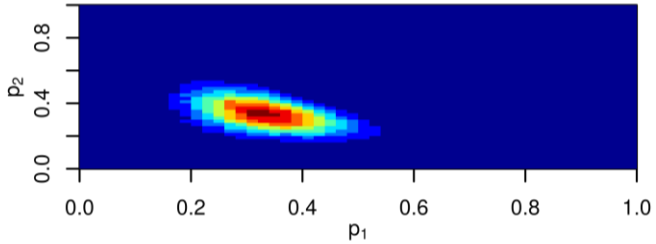
You control the mean proportions AND the variance (“informativeness”)



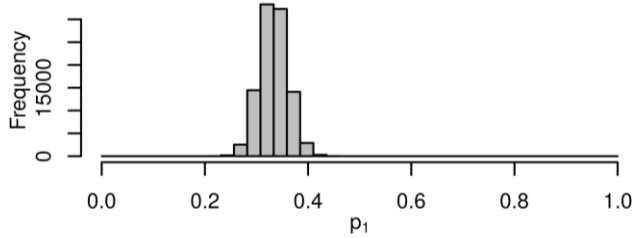


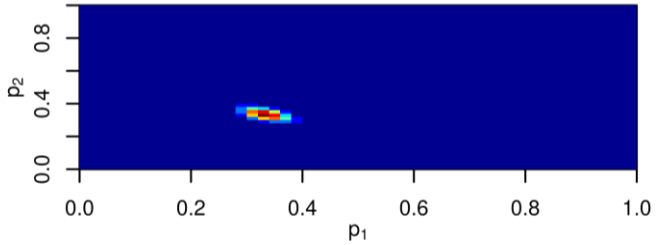
α = (1, 1, 1)





α = (10, 10, 10)





α = (100, 100, 100)

3. Constructing informative priors

You control the mean proportions AND the variance (“informativeness”)

30 

8 

25 

3. Constructing informative priors

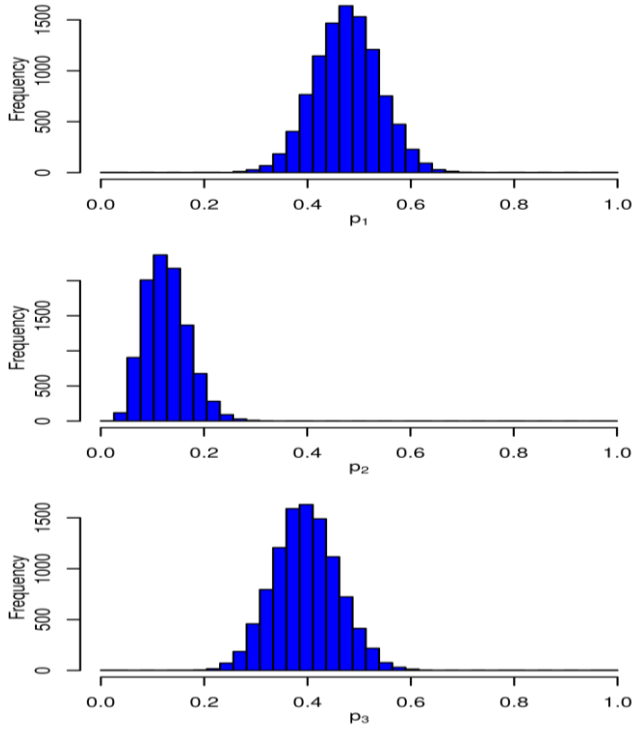
You control the mean proportions AND the variance (“informativeness”)

30

8

25

$$\boldsymbol{\alpha} = (\mathbf{30}, \mathbf{8}, \mathbf{25})$$



3. Constructing informative priors

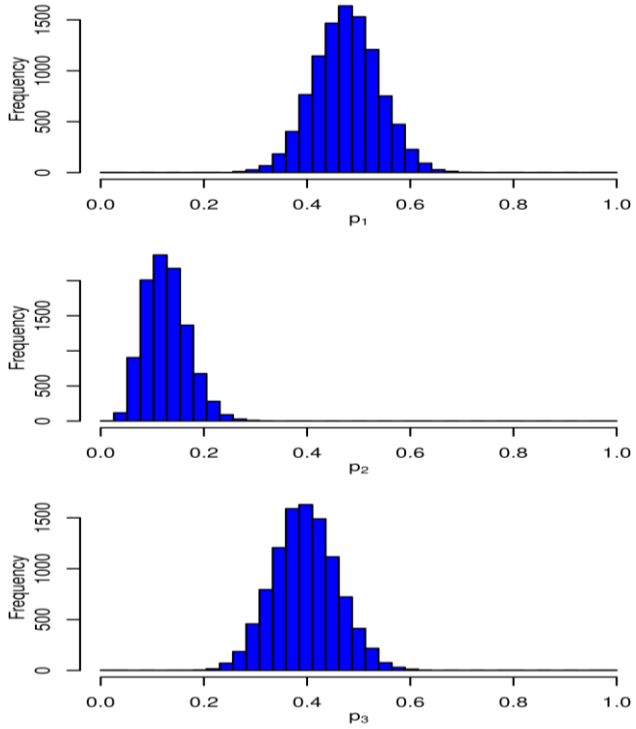
You control the mean proportions AND the variance (“informativeness”)

30

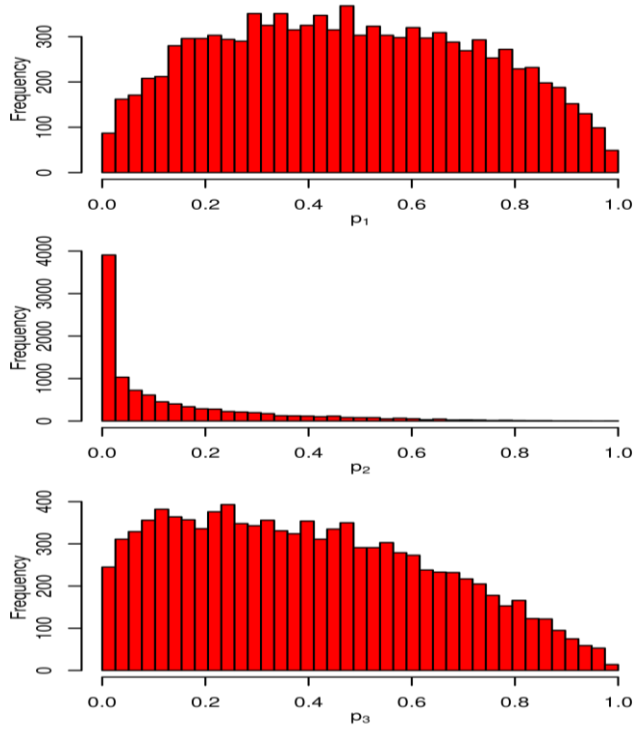
8

25

$$\boldsymbol{\alpha} = (\mathbf{30}, \mathbf{8}, \mathbf{25})$$



$$\boldsymbol{\alpha} = \frac{\mathbf{3} * (\mathbf{30}, \mathbf{8}, \mathbf{25})}{\mathbf{63}}$$



Effect of priors/ “Bayesian mixing models are biased”

REDUCE THE INFLUENCE OF THE GENERALIST PRIOR:

1. COLLECT MORE DATA (SOURCE AND CONSUMER)
2. SPECIFY A NON-GENERALIST PRIOR

Application to other biotracers
“Stable isotope” mixing models

Application to other biotracers

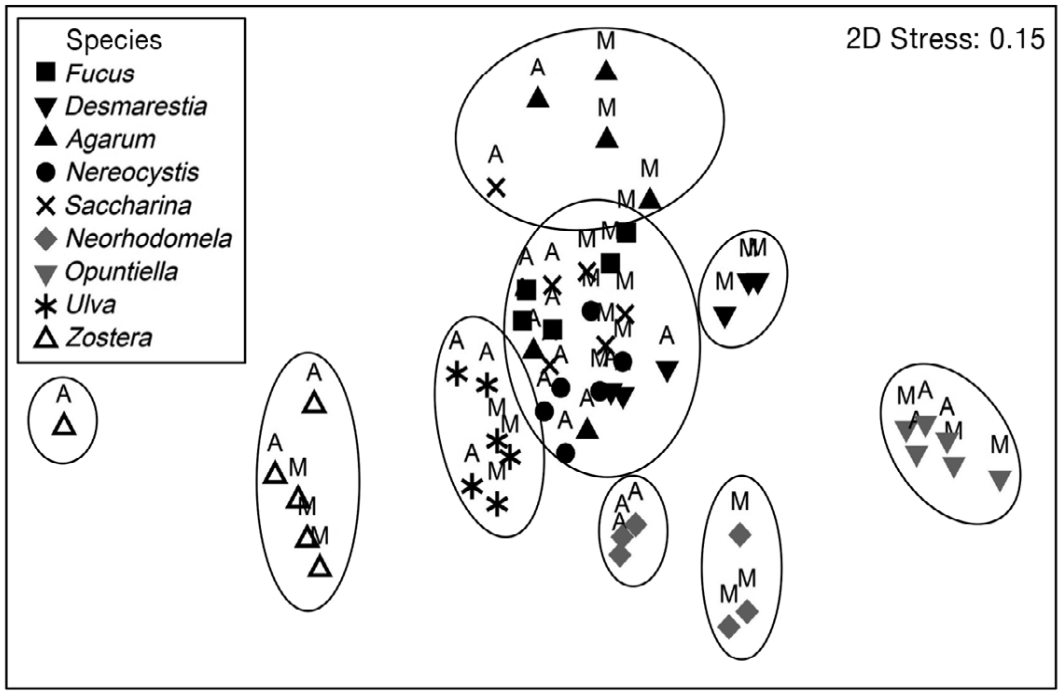
- Fatty acids
- Compound-specific stable isotopes
- Element concentrations
- Sediment color

Application to other biotracers

- Fatty acids
- Compound-specific stable isotopes
- Element concentrations
- Sediment color

Great promise!

Application to other biotracers



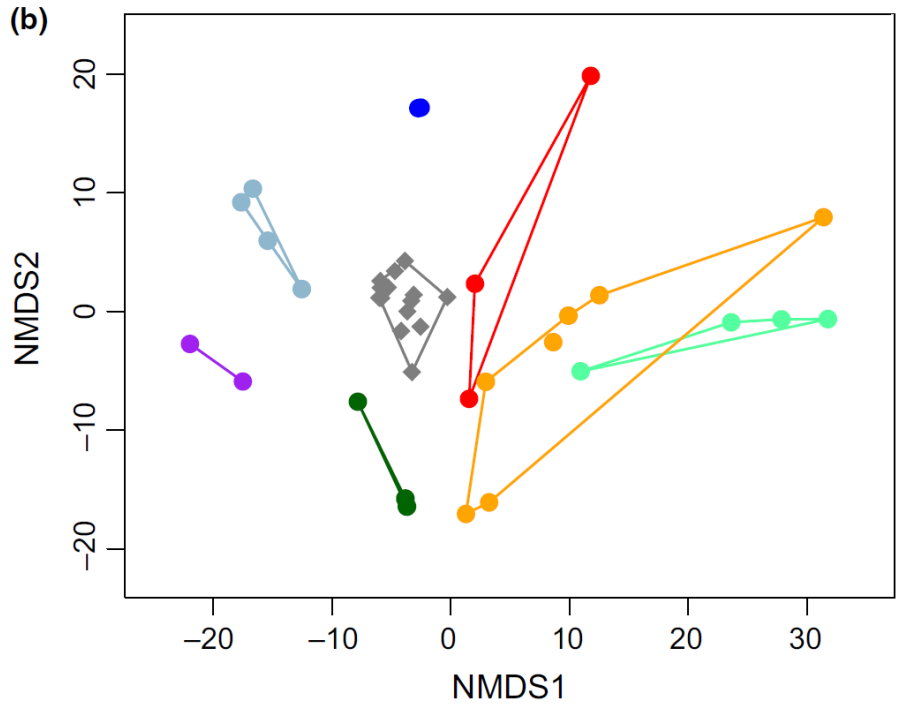
Theoretical consumer from source data

- 3 stable isotopes
- 6 essential fatty acids
- 19 non-essential fatty acids

SIAR

More biotracers = better performance

Application to other biotracers



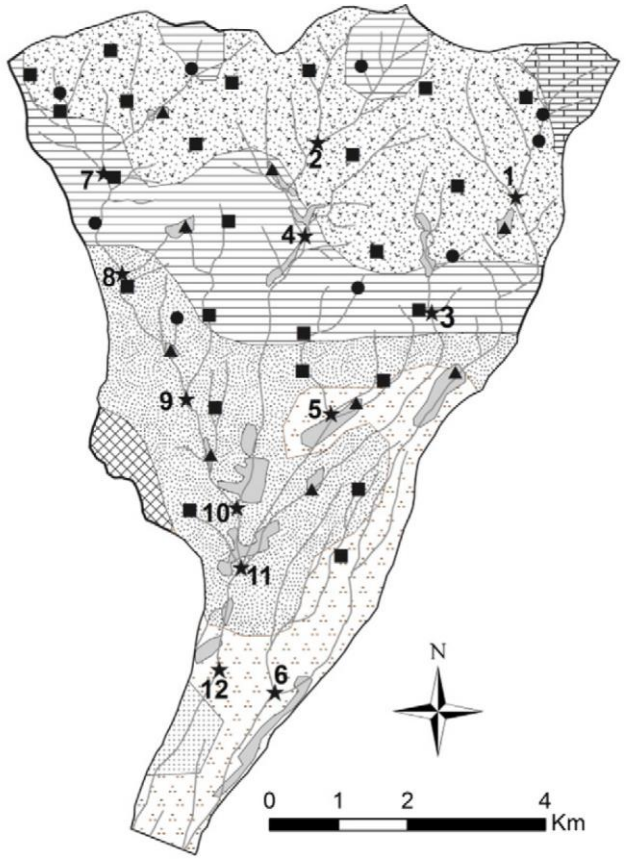
Source$_i$ data are consumers fed source *i*

- TDF = 0

22 fatty acids

MixSIR

Application to other biotracers



Mixtures are sediment samples

- TDF = 0

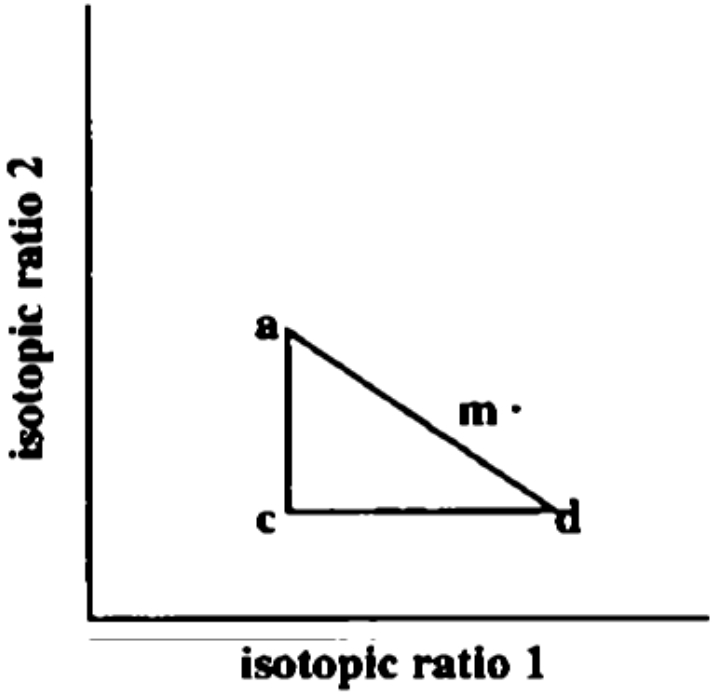
28 element concentrations

Other sediment datasets:

- element concentrations (n = 56)
- color variables (n = 15)
- CSSI (n = 20)

1. Testing mix/source geometry

Less obvious if mix data is inside source hypervolume



Standard in sediment fingerprinting:

- Check each dimension separately

2. Check for normality

Fatty acid profiles are *proportions*

$$t_j \sim N(alr(\tau_j), \Sigma_\tau)$$

$$\tau_j = C\left\{\sum_{s}^{n}\left(\pi_{j,s}\Phi_s\right)\left(\kappa_s \otimes \phi_{j,s}\right)\right\}$$

3. Multiple data types

Fatty acid profiles (n = 25):

$$t_j \sim N(alr(\tau_j), \Sigma_\tau)$$

$$\tau_j = C\left\{\sum_s^n \left(\pi_{j,s}\Phi_s\right)\left(\kappa_s \otimes \phi_{j,s}\right)\right\}$$

Stable isotopes (n = 3):

$$t_r^{SI} = \sum_s^n \pi_{r,s}\left(y_{q,r} + \gamma_s\right)$$

$$clr(\pi_r) \sim N(\Pi, \Sigma_\Pi)$$

$$\gamma_{s,SI} \sim N(\nu_{SI}, \sigma_{SI}).$$

Weight each data type equally?
Weight by number of tracers?

4. Selecting biotracers

Are more biotracers always better?

Discriminant function analysis (DFA) to choose “optimum subset”

Biotracer selection within mixing model?

Application to other biotracers
“Stable isotope” mixing models

TEST MIX/SOURCE GEOMETRY

CHECK FOR NORMALITY

MULTIPLE DATA TYPES

SELECTING BIOTRACERS

Pitfalls and misconceptions

1. Source geometry
2. MCMC convergence
3. **Effect of priors**
4. Error structures
5. Source lumping/splitting
6. How to include covariates
7. **Application to biotracers other than stable isotopes**

Acknowledgements

MixSIAR team:

- Brice Semmens
- Eric Ward
- Andrew Parnell
- Andrew Jackson
- Donald Phillips
- Stuart Bearhop
- Richard Inger

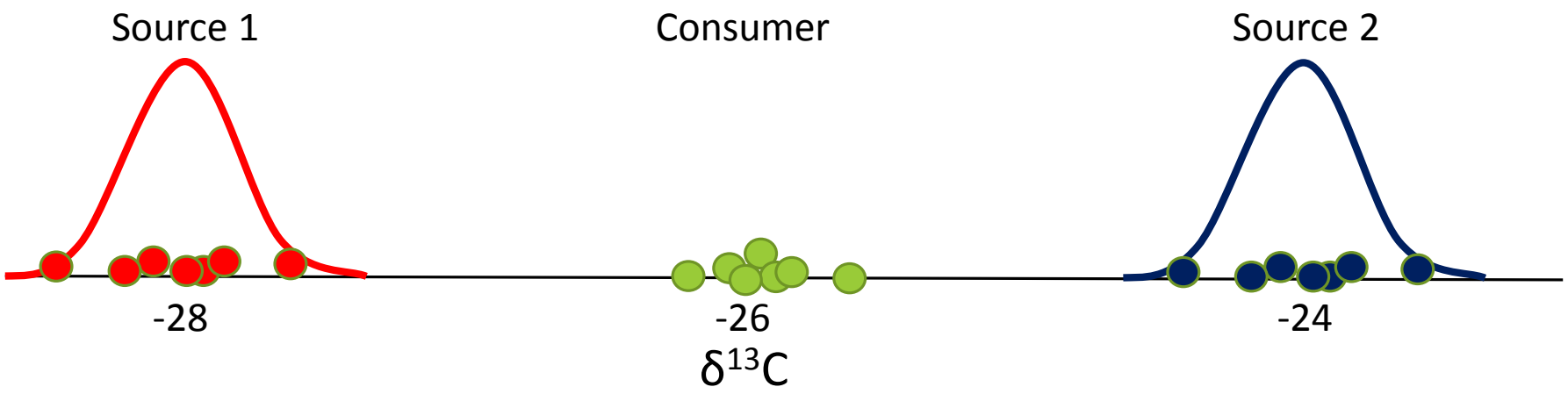
Support:

XSEDE

Extreme Science and Engineering Discovery Environment

Error structures

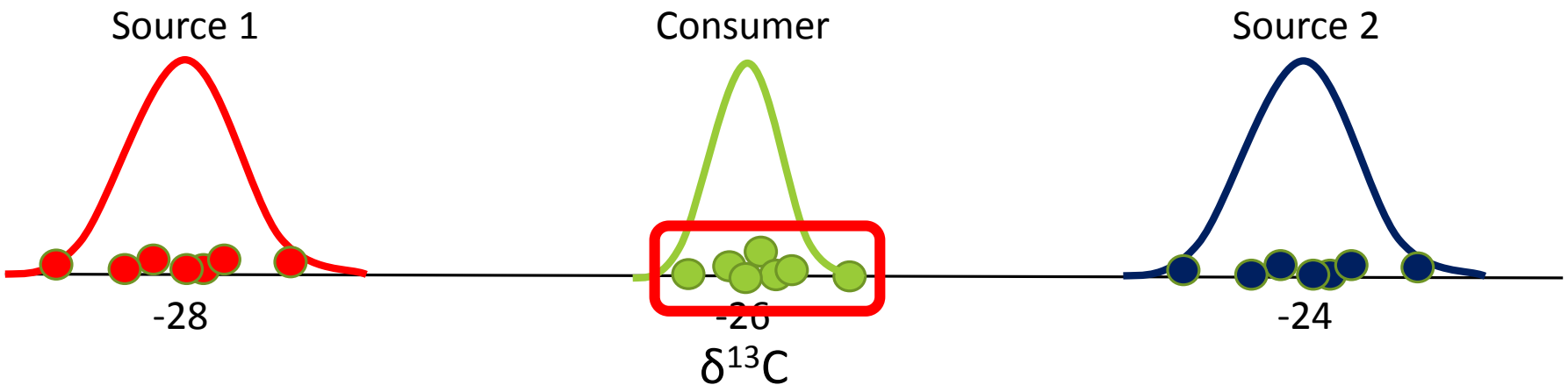
Error structures



Error structures

MixSIR

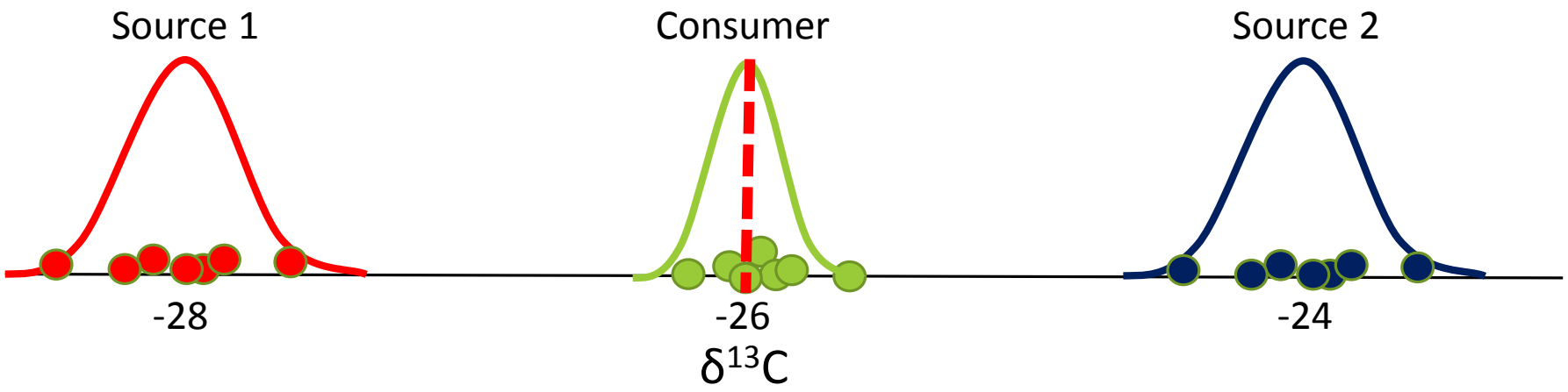
$$X_{ij} \sim N\left(\sum_{k=1}^{K} p_k(\mu_{jk} + \lambda_{jk}), \left[\sum_{k=1}^{K} p_k^2(\omega_{jk}^2 + \tau_{jk}^2)\right]\right)$$



Error structures

MixSIR

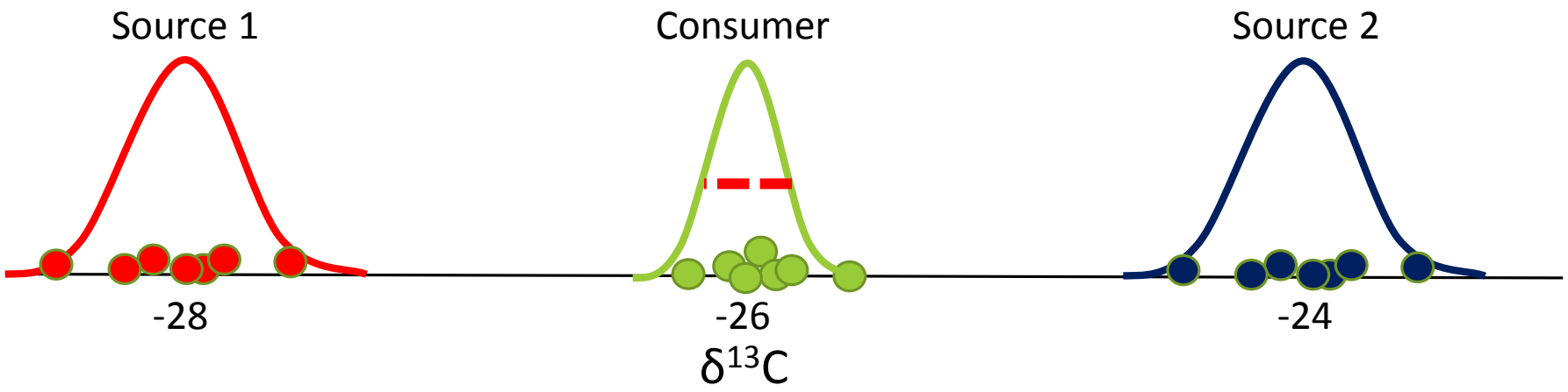
$$X_{ij} \sim N\left(\sum_{k=1}^{K} p_k(\mu_{jk}+\lambda_{jk}),\left[\sum_{k=1}^{K} p_k^2(\omega_{jk}^2+\tau_{jk}^2)\right]\right)$$



Error structures

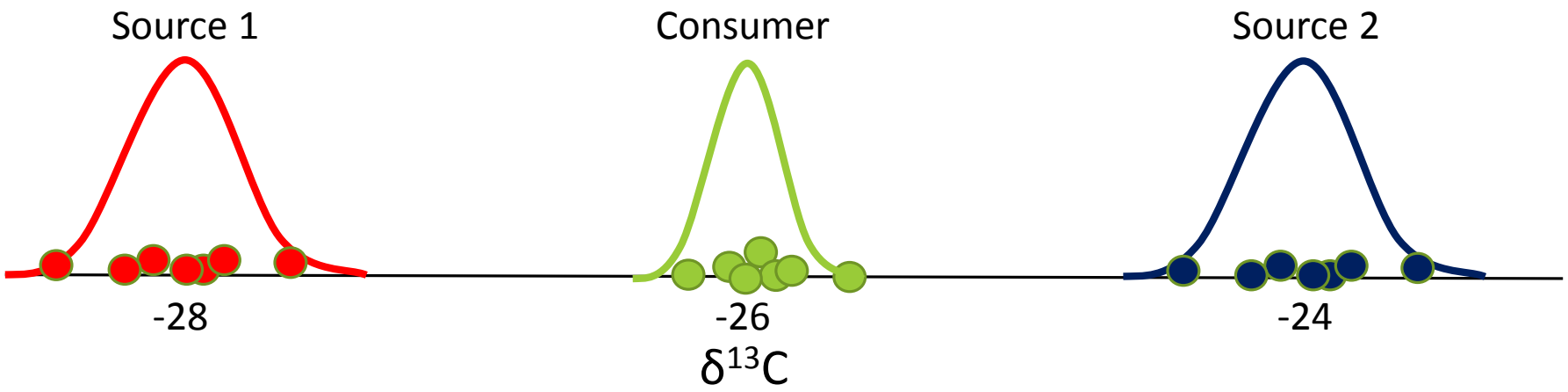
MixSIR

$$X_{ij} \sim N\left(\sum_{k=1}^{K} p_k(\mu_{jk} + \lambda_{jk}), \sum_{k=1}^{K} p_k^2(\omega_{jk}^2 + \tau_{jk}^2)\right)$$



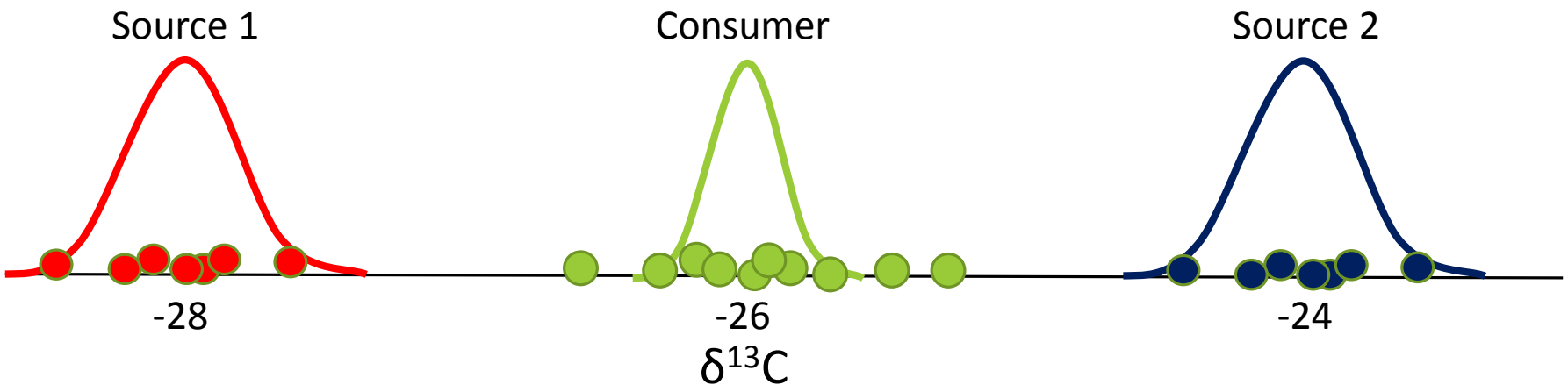
Error structures

MixSIR $\sigma^2_{process}$ Medium variance



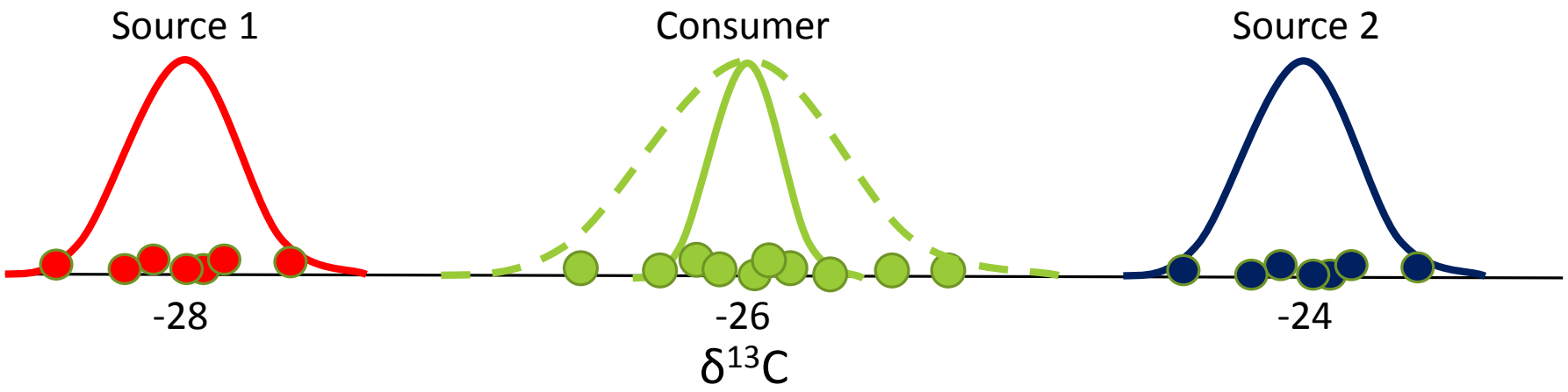
Error structures

MixSIR $\sigma^2_{process}$ High variance



Error structures

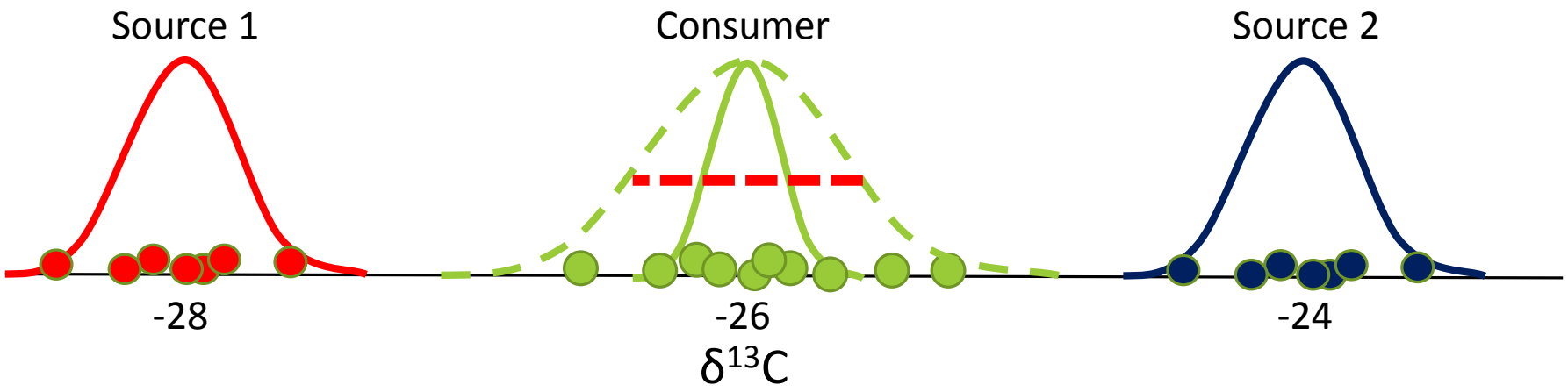
SIAR $\sigma^2_{process} + \sigma^2_{resid}$ High variance Note: cannot fit N = 1



Error structures

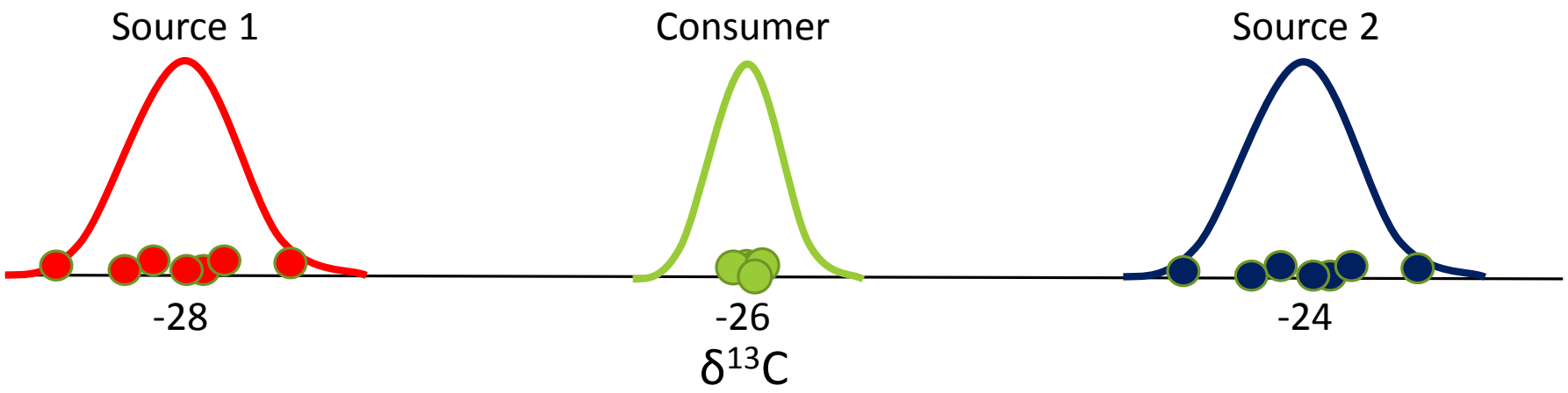
SIAR

$$X_{ij} \sim N\left(\sum_{k=1}^{K} p_k(\mu_{jk} + \lambda_{jk}), \left[\sum_{k=1}^{K} p_k^2(\omega_{jk}^2 + \tau_{jk}^2)\right] + \sigma_j^2\right)$$



Error structures

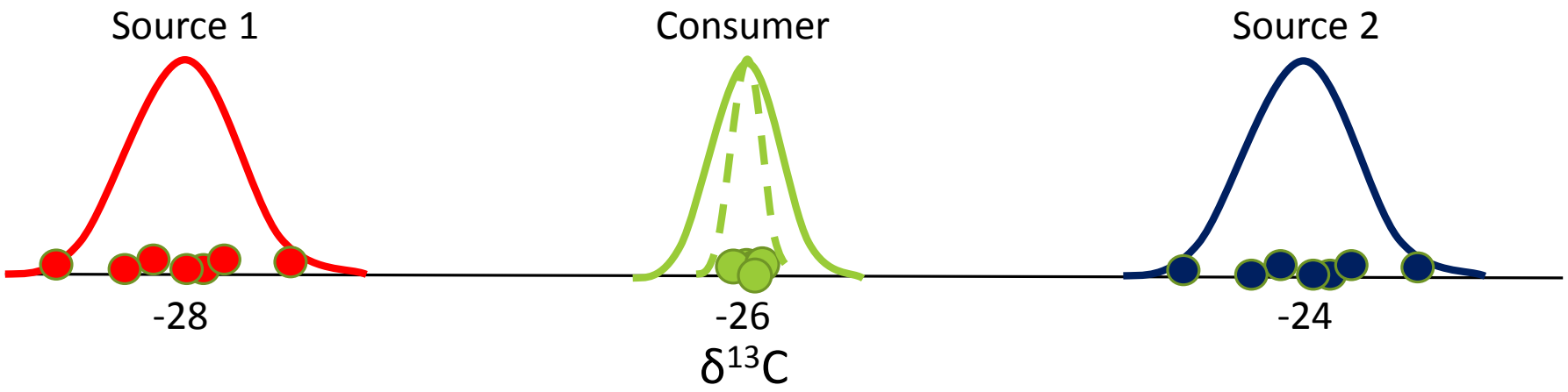
SIAR $\sigma^2_{process} + \sigma^2_{resid}$ Low variance



Error structures

$\sigma^2_{process} * \varepsilon_{resid}$ Low variance

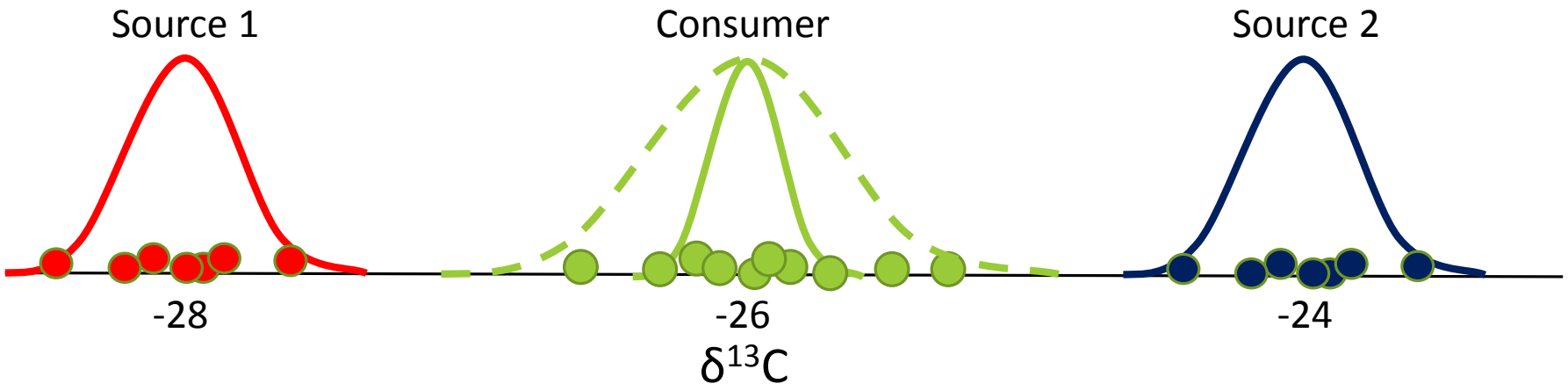
$\varepsilon < 1$



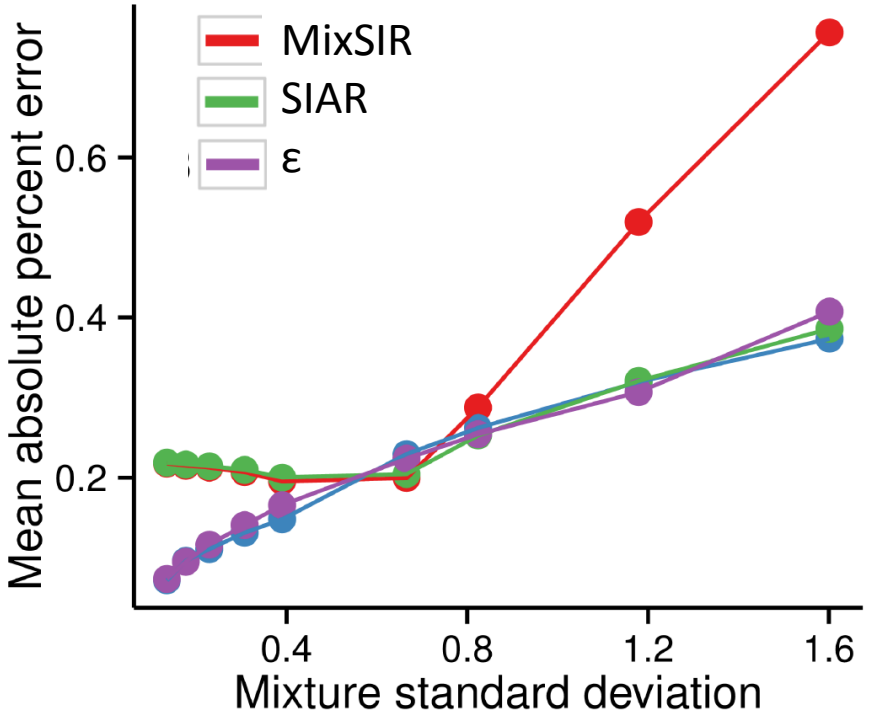
Error structures

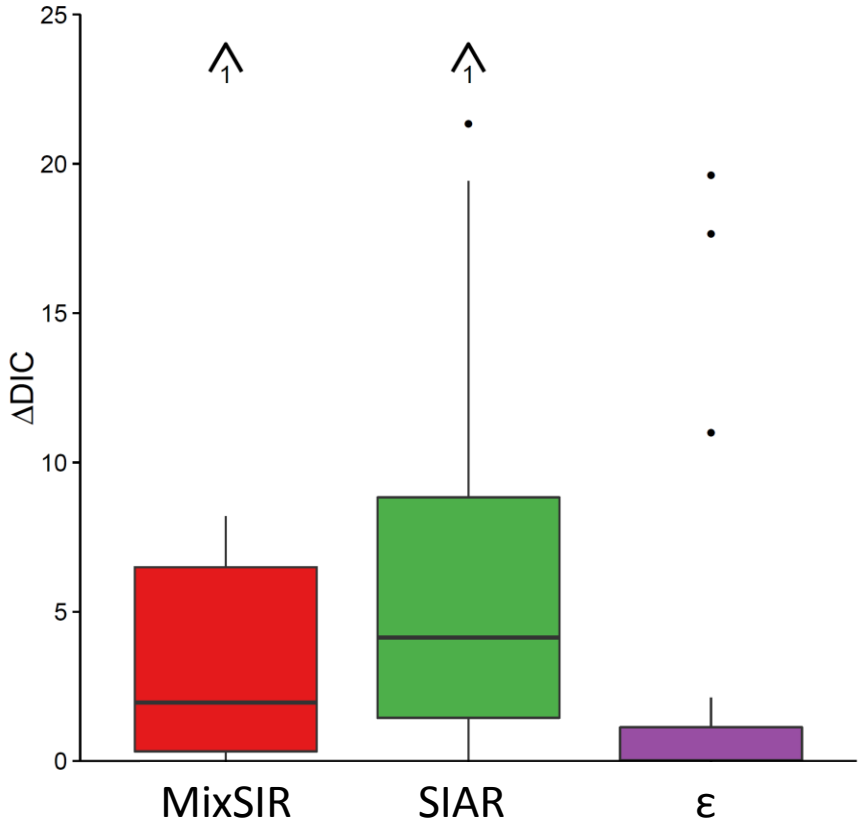
$\sigma^2_{process} * \varepsilon_{resid}$ High variance

$\varepsilon > 1$



Error structures





Error structures

MIXSIR (N = 1)

SIAR

MULTIPLICATIVE RESIDUAL ERROR