Material Reconstruction in EPMA as a Bayesian Inverse Problem

Lunch Talk, ACoM

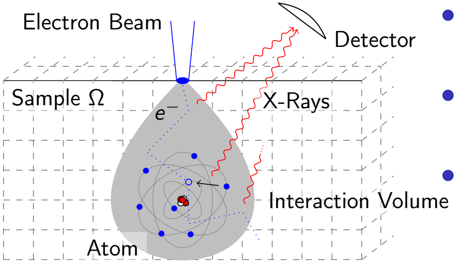
RWTH Aachen University

Tamme Claus

Was once:
Final Project - Stochastic Numerics SS2019
Roxana Pohlmann and Tamme Claus

supervised by: Prof. Raul Tempone, Nadhir Ben Rached and Anamika Pandey

June 2021



- electron beam ionizes atoms
- relaxation by emission of characteristic x ray
- intensity of x-rays (k-ratios) measured by detector

Inverse Problem

→ given measured k-ratios $\boldsymbol{k}^{\text{exp}}$ (normalized x-ray intensities)

What is the chemical composition $\boldsymbol{c}$ that caused those measurements?

Bayesian Inversion - Solution of the Inverse Problem

Classical approach: numerical optimization

- define a forward model $\boldsymbol{k}(\boldsymbol{c})$
- define a cost function $J(\boldsymbol{c}) = ||\boldsymbol{k}^{\text{exp}} - \boldsymbol{k}^{\text{mod}}(\boldsymbol{c})||^2$
- numerical optimization (gradient, regularization, data filtering, ...)

Bayesian Inversion

- treat $\boldsymbol{c}$ as a random variable (random field)
- prior information: $\pi(\boldsymbol{c})$
- likelihood: $\pi(\boldsymbol{k}^{\text{obs}}|\boldsymbol{c})$ (forward model)

Given $\boldsymbol{k}^{obs}$, what do we know about $\boldsymbol{c}$?

Given $\boldsymbol{k}^{obs}$, what do we know about $\boldsymbol{c}$?

joint probability $\quad \pi(\boldsymbol{c}|\boldsymbol{k}^{\text{obs}})\pi(\boldsymbol{k}^{\text{obs}}) = \pi(\boldsymbol{c}, \boldsymbol{k}^{\text{obs}}) = \pi(\boldsymbol{k}^{\text{obs}}|\boldsymbol{c})\pi(\boldsymbol{c})$

Bayes' Theorem

$$\underbrace{\pi(\boldsymbol{c}|\boldsymbol{k}^{obs})}_{\text{posterior}} \propto \underbrace{\pi(\boldsymbol{k}^{obs}|\boldsymbol{c})}_{\text{likelihood}} \underbrace{\pi(\boldsymbol{c})}_{\text{prior}}$$

- the posterior is the 'solution' to the inverse problem

using $\pi(\boldsymbol{c}|\boldsymbol{k}^{obs})$ we can compute:

- expected value of mass fractions $\mathbb{E}(\boldsymbol{c}|\boldsymbol{k}^{\text{obs}})$
- maximum a posteriori (MAP) estimate (maximum likelihood)
- confidence intervals of the estimates

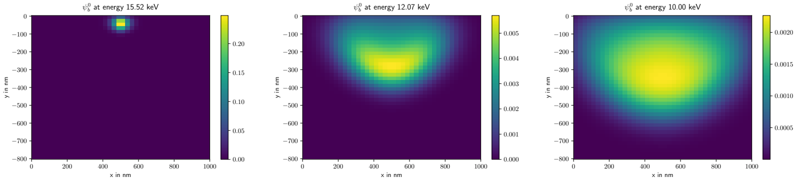
utilizing Monte Carlo methods (here Metropolis-Hastings)

Electron Transport - $\mathcal{M}1$-Model

- given: chemical composition $\boldsymbol{c}(x)$
- solve: $\mathcal{M}1$-Model (Linear Boltzmann, continously-slowing-down approximation, moment expansion, minimum entropy closure)

$$\partial_\epsilon \left(S(\boldsymbol{c},\epsilon) \begin{pmatrix} \psi_b^0(x,\epsilon) \\ \psi_b^1(x,\epsilon) \end{pmatrix} \right) + \nabla_x \begin{pmatrix} \psi_b^1(x,\epsilon) \\ \psi_{ME}^2(\psi_b^0,\psi_b^1) \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & T(\boldsymbol{c},\epsilon) I \end{pmatrix} \begin{pmatrix} \psi_b^0(x,\epsilon) \\ \psi_b^1(x,\epsilon) \end{pmatrix} = 0$$

  - boundary conditions capture the electron beam
  - we use: finite volume method (library: CLAWPACK)



X Ray Intensities

- calculate: k-ratios

$$\boldsymbol{k}_{bi} = \frac{1}{I^i_{std}} \int_\Omega \int_{\epsilon_{min}}^{\epsilon_{max}} e^{-\int_{d(x)} \mu^i(\boldsymbol{c})\,\mathrm{d}y} \omega^i \sigma^i_{ion}(\epsilon) N^i(\boldsymbol{c}) \psi_b^0(\boldsymbol{c}, x, \epsilon)\,\mathrm{d}\epsilon\,\mathrm{d}x$$

here we consider: ionization, fluorescence, absorption, ...

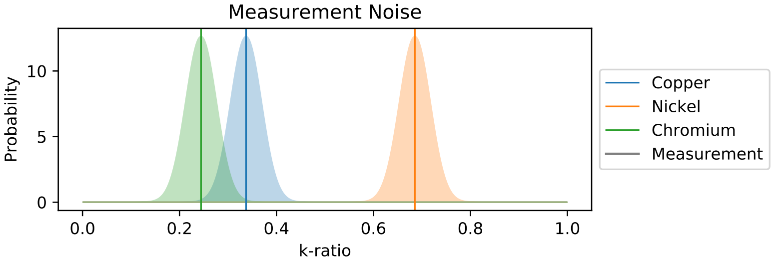
Uncertainty in

- model (e.g. parameter, simplification)
- experimental setup
- ...

assume (academic):

- no model uncertainty
- independent gaussian noise (detector: Poisson point process)

$$\boldsymbol{k}^{\text{exp}} - \boldsymbol{k}^{\text{mod}}(\boldsymbol{c}) \sim N(0, \text{diag}(\sigma^2, \ldots))$$



$\rightarrow$ likelihood $\pi(\boldsymbol{k}^{\text{exp}}|\boldsymbol{c})$

Prior for mass fractions $\boldsymbol{c}$

Prior information

$0 \leq \boldsymbol{c}_{ik} \leq 1 \quad \forall\, {}^{i=1,\dots,n_c}_{k=1,\dots,n_e}$

$\sum_{k=1}^{n_e} \boldsymbol{c}_{ik} = 1 \quad \forall\, i=1,\dots,n_c$

Dirichlet Distribution

$$X \sim \mathrm{Dir}(\alpha) \quad \alpha \in \mathbb{R}^n$$

- pdf: $\frac{1}{B(\alpha)} \prod_{i=1}^{n} x_i^{\alpha_i - 1}$
- normalizing constant $B(\alpha)$
- $\mathbb{E}[X_i] = \frac{\alpha_i}{\alpha_o}$
- $\mathrm{Var}[X_i] = \frac{\frac{\alpha_i}{\alpha_o}(1 - \frac{\alpha_i}{\alpha_o})}{\alpha_o + 1}$
- $\alpha_o = \sum_{i=1}^{n} \alpha_i$

- in each subdomain $\boldsymbol{c}_{i:} \sim \mathrm{Dir}(\alpha_i)$ independent
- we use $\alpha_i = [1, 1, \dots] \, \forall i$
- uniform over the support

$\rightarrow$ prior $\pi(\boldsymbol{c})$

Bayes' Theorem

$$\underbrace{\pi(\boldsymbol{c}|\boldsymbol{k}^{obs})}_{\text{posterior}} \propto \underbrace{\pi(\boldsymbol{k}^{obs}|\boldsymbol{c})}_{\text{likelihood}} \underbrace{\pi(\boldsymbol{c})}_{\text{prior}}$$

Metropolis Hastings: Idea

- construct a Markov Chain (stationary distribution = desired distribution)
- propose a sample (based on the current sample)
- accept / reject the sample (probability based on detailed balance)

Algorithm 1: Metropolis-Hastings: to sample from $\pi(\boldsymbol{c}|\boldsymbol{k})$

Initialize $\boldsymbol{c}^{(0)} \sim$ **prior**
for $j = 1, 2, \ldots$ **do**
  Propose: $\boldsymbol{c}^* \sim q(\cdot|\boldsymbol{c}^{(j-1)})$
  Acceptance Probability:
  $A(\boldsymbol{c}^*|\boldsymbol{c}^{(j-1)}) = \min(1, \frac{q(\boldsymbol{c}^{(j-1)}|\boldsymbol{c}^*)\pi(\boldsymbol{c}^*|\boldsymbol{k}^{obs})}{q(\boldsymbol{c}^*|\boldsymbol{c}^{(j-1)})\pi(\boldsymbol{c}^{(j-1)}|\boldsymbol{k}^{obs})})$
  $u \sim \text{Uni}([0,1])$
  if $u < A$ **then**
    Accept: $\boldsymbol{c}^{(j)} = \boldsymbol{c}^*$
  else
    Reject: $\boldsymbol{c}^{(j)} = \boldsymbol{c}^{(j-1)}$

- How should we choose the proposal $q(\cdot|\boldsymbol{c}^{(i-1)})$?

Metropolis-Hastings: Proposal distribution $q(\cdot|\boldsymbol{c}^{(j-1)})$

<u>Often:</u> $\mathcal{N}(\boldsymbol{c}^{(j-1)}, \Sigma_p)$ as proposal distribution

- with $\boldsymbol{c}^{(j-1)}$ as the mean and a chosen covariance Σ_p
- Not applicable here (conditions on $\boldsymbol{c}$)

$\rightarrow$ Use a proposal distribution which (hopefully) is similar to the posterior

Idea: Use Dirichlet again

In each subdomain i: $\boldsymbol{c}_i^* \sim q(\cdot, \boldsymbol{c}^{(j-1)}) = \text{Dir}(r\boldsymbol{c}_i^{(j-1)}) \quad r \in \mathbb{R}$

- $\mathbb{E}[\boldsymbol{c}_i^*] = \boldsymbol{c}_i^{(j-1)}$
- $\text{Var}[\boldsymbol{c}_{ik}^*] = \frac{\boldsymbol{c}_{ik}^{(j-1)}(1-\boldsymbol{c}_{ik}^{(j-1)})}{r+1}$
- We can control the variance of the proposal with r

Physical Domain

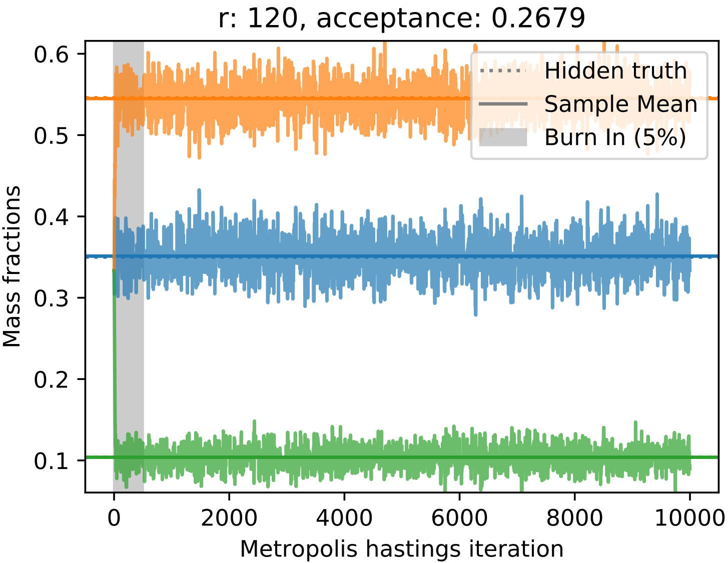
Number of subdomains	1
Number of chemical elements	3

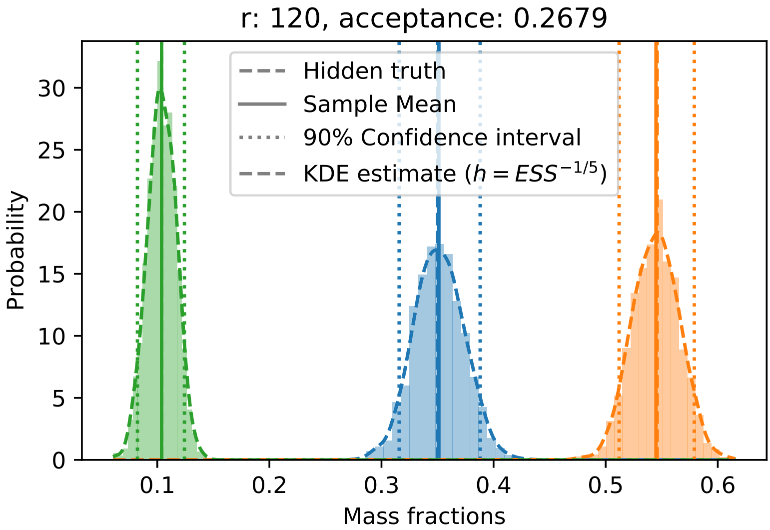
Numerical parameters

Spatial Grid $\bar{x}$	40×40, $(1000 \times 800\text{nm})$
Number of steps ϵ	100, ([10, 17] keV)
Beam energy ϵ_{beam}	16 kV
Measurement noise σ^2	0.001

For the first try, we choose:

Number of MCMC iterations	10000
Proposal variance r	120
Initial mass fraction $\boldsymbol{c}^{(0)}$	$[0.33, 0.33, 0.33]$


r: 120, acceptance: 0.2679
Hidden truth
Sample Mean
Burn In (5%)
Mass fractions
0.1
0.2
0.3
0.4
0.5
0.6
0
2000
4000
6000
8000
10000
Metropolis hastings iteration


r: 120, acceptance: 0.2679
Hidden truth
Sample Mean
90% Confidence interval
KDE estimate ($h = ESS^{-1/5}$)
Probability
0
5
10
15
20
25
30
0.1
0.2
0.3
0.4
0.5
0.6
Mass fractions

Diagnostics - Is the chain converged?

Autocorrelation at lag k

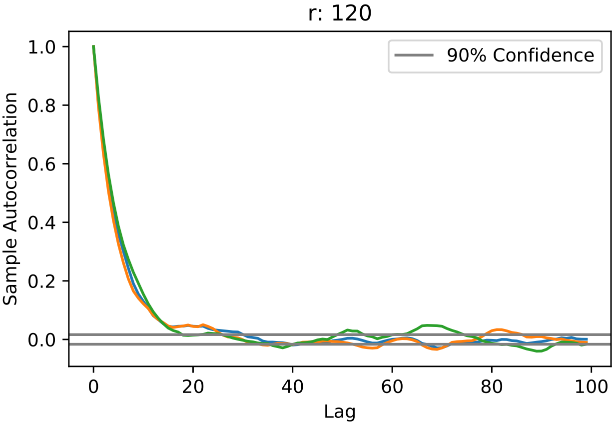
- Correlation of the signal with itself at lag k.

$$\hat{\rho}(k) = \frac{\sum_{i=1}^{T-k}(x_{i+k} - \bar{x})(x_i - \bar{x})}{\sum_{i=1}^{T}(x_i - \bar{x})^2}, \ \bar{x} = \sum_{i=1}^{T} \frac{x_i}{T}$$

Effective Sample Size

- ESS describes the number of weakly correlated samples.
- $\text{ESS} = \frac{N}{1+2\sum_{k=1}^{\infty}\rho(k)}$, with N number of samples and $\rho(k)$ the correlation at lag k
- Best expectation: $\frac{\text{ESS}}{N}$ close to 1.

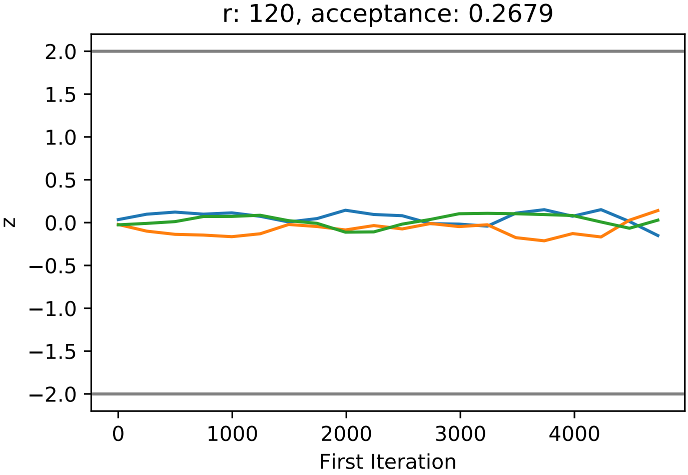
Diagnostics: Autocorrelation and ESS



For $r = 120$, $N = 10000$: $\text{ESS} = \min\,[737.6,\ 732.4,\ 660.2] = 660$

Geweke test

- **Idea** A converged chain has the same expectation in the first and last part
- Say T_1 corresponds to the first 10% of the samples and T_2 to the last 50%
- $z = \frac{\mathbb{E}[T_1] - \mathbb{E}[T_2]}{\sqrt{\text{Var}\, T_1 + \text{Var}\, T_2}}$ should converge to normal distribution

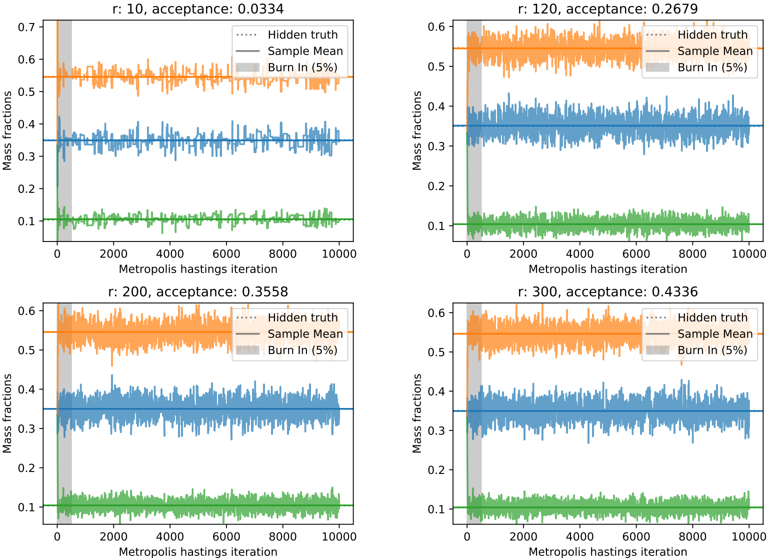

r: 120, acceptance: 0.2679
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
z
0
1000
2000
3000
4000
First Iteration

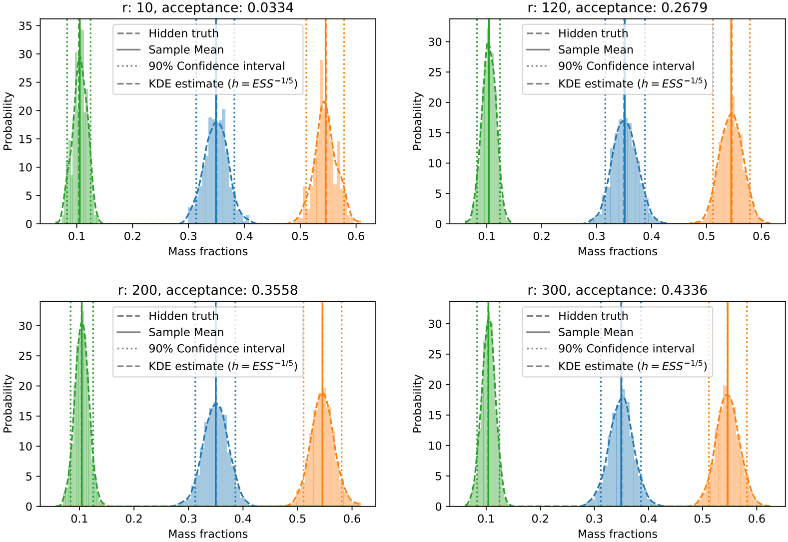
Gelman-Rubin Convergence Diagnostic

- Evaluate MCMC convergence by comparing estimated between-chain and within-chain variance for each parameter
- $\boldsymbol{c}_m^{(j)}$ with $m = 1 \dots M$ different chains and $j = 1 \dots N$ samples
- $\hat{\boldsymbol{c}}_m$ sample mean and $\hat{\sigma}_m^2$ sample variance
- Overall mean $\hat{\boldsymbol{c}} = \frac{1}{M} \sum_{m=1}^{M} \hat{\boldsymbol{c}}_m$
- Between-chain variance $B = \frac{N}{M-1} \sum_{m=1}^{M} (\hat{\boldsymbol{c}}_m - \hat{\boldsymbol{c}})^2$
- Within-chain variance $W = \frac{1}{M} \sum_{m=1}^{M} \hat{\sigma}_m^2$
- Pooled variance $\hat{V} = \frac{N-1}{N} W + \frac{M+1}{MN} B$
- Test: if $\frac{\hat{V}}{W}$ is close to one, the chains have converged

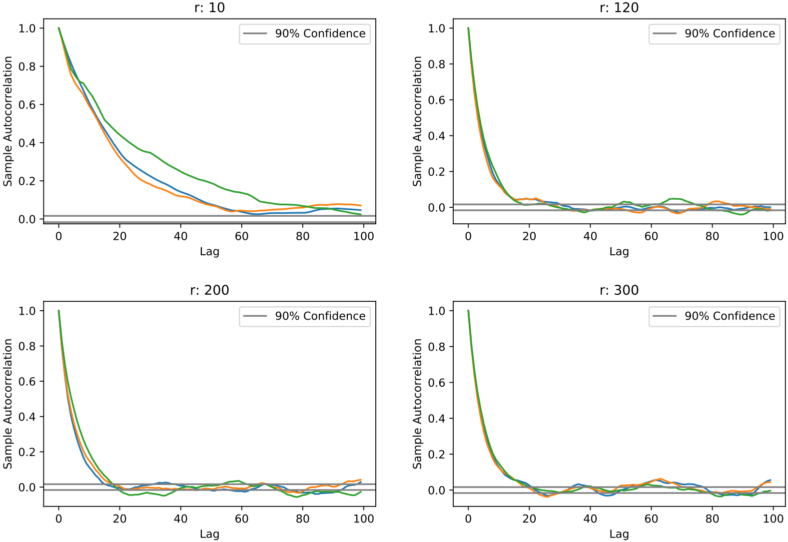
Parameter study (different variances of the proposal r)

Trace Plots: Impact of r

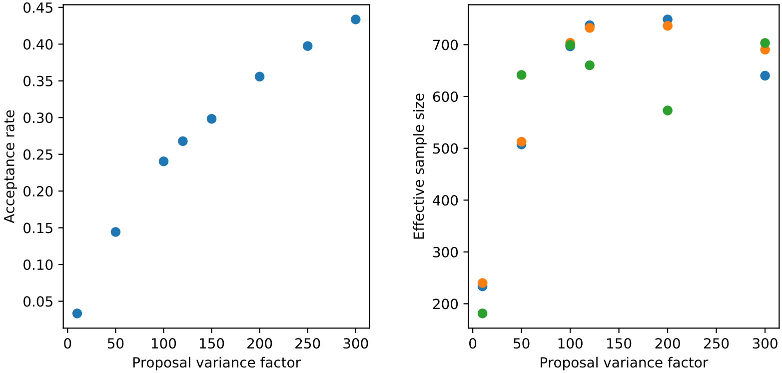



r: 10, acceptance: 0.0334
r: 120, acceptance: 0.2679
r: 200, acceptance: 0.3558
r: 300, acceptance: 0.4336
Hidden truth
Sample Mean
90% Confidence interval
KDE estimate ($h = ESS^{-1/5}$)
Probability
Mass fractions

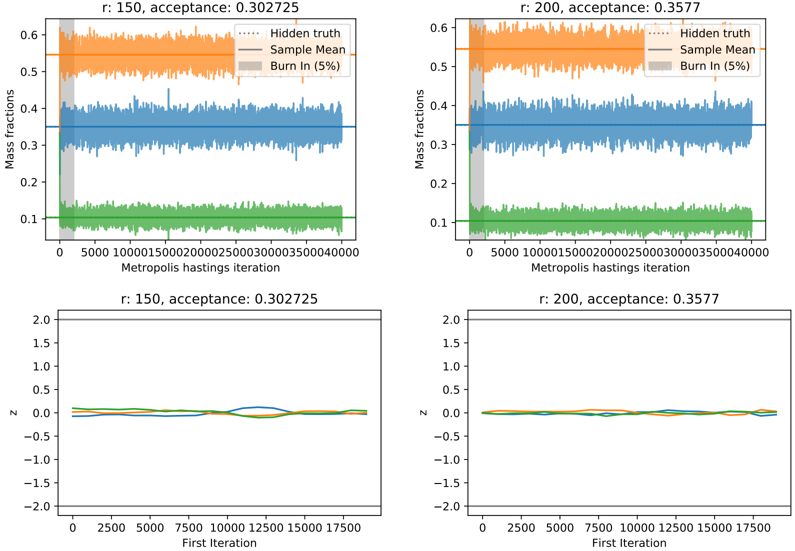
Autocorrelation: Impact of r

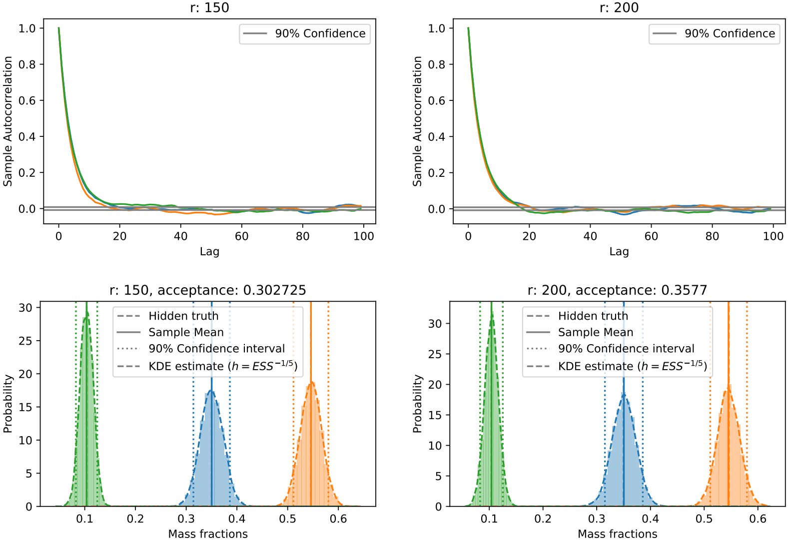


Acceptance rate and ESS



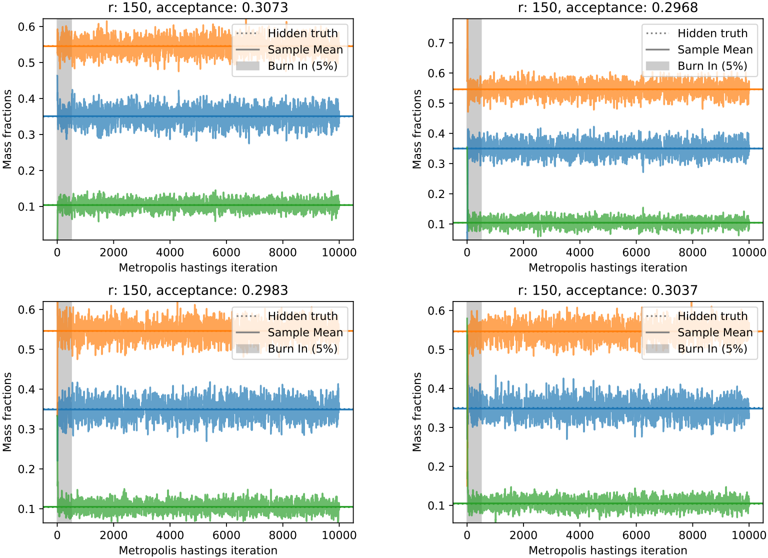
→ run the chain longer: $N = 40000$?





r: 150
r: 200
90% Confidence
Sample Autocorrelation
Lag
r: 150, acceptance: 0.302725
r: 200, acceptance: 0.3577
Hidden truth
Sample Mean
90% Confidence interval
KDE estimate ($h = ESS^{-1/5}$)
Probability
Mass fractions

$\rightarrow$ run with $(M = 6)$ different initial conditions $r = 150$ $N = 10000$



Apply the Gelman-Rubin Diagnostic: $M = 6$, $N = 10000$, $r = 150$ for the three mass fractions (parameters) we get

- $\boldsymbol{c}_0$: $\frac{\hat{V}}{W} = 1.0026$
- $\boldsymbol{c}_1$: $\frac{\hat{V}}{W} = 1.0006$
- $\boldsymbol{c}_2$: $\frac{\hat{V}}{W} = 1.0029$

Conclusion

- simple integration with existing forward models (no gradient information)
- uncertainties of the reconstruction result
- no convergence guarantee (run the chain forever)

Further Investigation

- more subdomains ($\rightarrow$ more paramters, but more realistic)
- parameter tuning for r
- more realistic likelihood (include model uncertainties, other measurement errors)
- more sophisticated proposal distribution

Jonas Buenger, Silvia Richter, and Manuel Torrilhon.
A deterministic model of electron transport for electron probe microanalysis.
IOP Conference Series: Materials Science and Engineering, 304:012004, 01 2018.

Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin.
Bayesian data analysis.
Texts in statistical science. Chapman & Hall/CRC, 2nd ed edition, 2004.

J. Kaipio and E. Somersalo.
Statistical and Computational Inverse Problems.
Applied Mathematical Sciences. Springer New York, 2006.

S.M. Lynch.
Introduction to Applied Bayesian Statistics and Estimation for Social Scientists.
Statistics for Social and Behavioral Sciences. Springer New York, 2007.

Niklas Mevenkamp.
Inverse modeling in electron probe microanalysis based on deterministic transport equations, 09 2013.

A. M. Stuart.
Inverse problems: A bayesian perspective.
Acta Numerica, 19:451–559, 2010.

Claus T.
Application of the adjoint method in gradient-based optimization to the m1-model in electron beam microanalysis, 2018.

A. Tarantola.
Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation.
Elsevier Science, 1987.