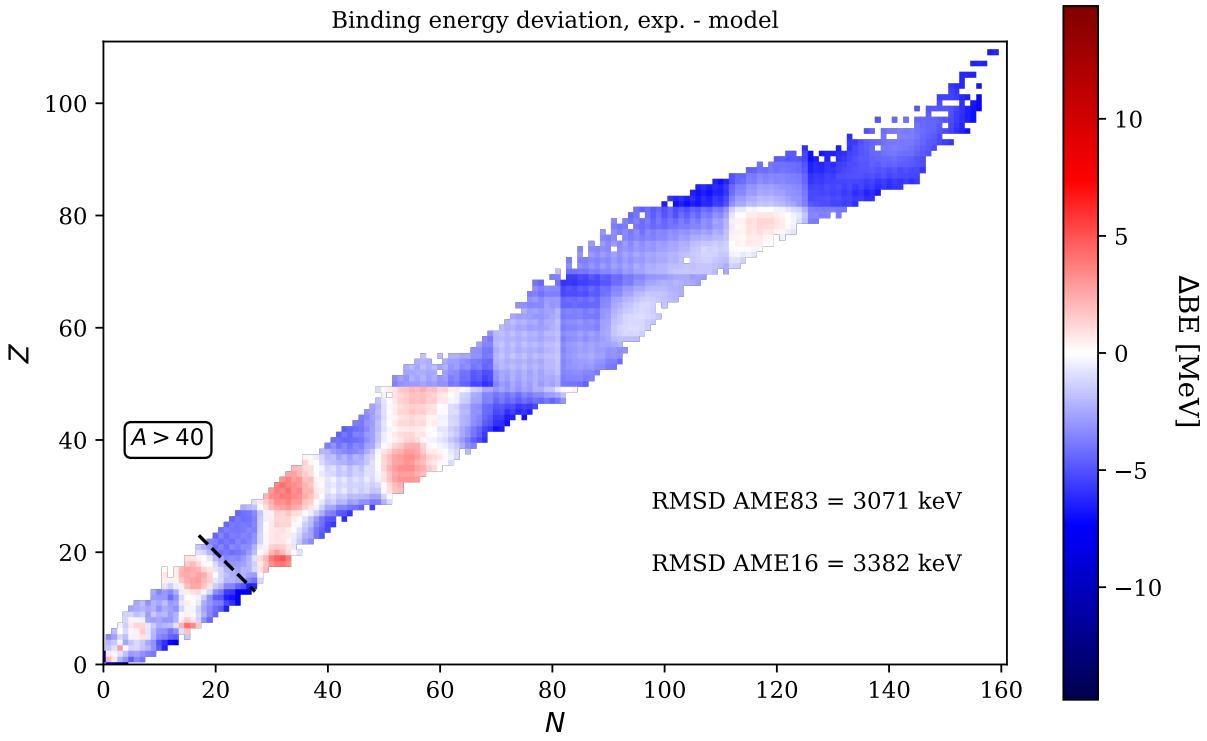
CHALMERS
UNIVERSITY OF TECHNOLOGY

α_1 = 0.68, α_2 = 17.32, α_3 = 14.83, α_4 = 36.83, α_5 = 53.96, α_6 = 0.05, α_7 = -0.49, α_8 = 0.02, α_{10} = 5.13



Bayesian Model Averaging of Nuclear Mass Models

And the Limits of Stability of Visible Matter

Master's thesis in Physics and Astronomy

Sebastian Lundquist

DEPARTMENT OF PHYSICS

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2020
www.chalmers.se

Master's thesis 2020

Bayesian Model Averaging of Nuclear Mass Models

And the Limits of Stability of Visible Matter

Sebastian Lundquist

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Physics
Division of Subatomic, High Energy and Plasma Physics
Chalmers University of Technology
Gothenburg, Sweden 2020

Bayesian Model Averaging of Nuclear Mass Models
And the Limits of Stability of Visible Matter
Sebastian Lundquist

Supervisor and examiner: Andreas Ekström, Department of Physics

Master's Thesis 2020
Department of Physics
Division of Subatomic, High Energy and Plasma Physics
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Comparison of Atomic Mass Evaluation 2016 and prediction of nuclear binding energies using a Bayesian calibration of parameters for the Duflo Zuker 10 parameter model after calculating posterior model probability and selecting the most plausible model, thus omitting the 9th term. ΔBE = experiment − model.

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2020

Bayesian Model Averaging of Nuclear Mass Models
And the Limits of Stability of Visible Matter
Sebastian Lundquist
Department of Physics
Chalmers University of Technology

Abstract

In this thesis we investigate the performance of Bayesian inference and Bayesian model averaging (BMA) applied to two nuclear mass models, the Duflo-Zuker 10 parameter model (DZ10) and the semi-empirical mass formula (SEMF). The DZ10 and SEMF models both have theoretically and experimentally motivated terms but the relative importance of them is less clear. Using Bayesian inference and BMA we have attempted to quantify model uncertainties and improve inference about nuclear masses. To explore the robustness of our BMA analysis we compare the results using different choices of parameter priors, and vary the assumed model discrepancy. The main focus is on the DZ10 model. We employ the Atomic Mass Evaluation from 1983 (AME83) for parameter estimation, then we evaluate the predictive power of the model using the Atomic Mass Evaluation from 2016 (AME16). In an attempt to determine the limits of stability of visible matter we also make a prediction for the neutron drip lines, in the tin isotopic chain ($Z = 50$) using the DZ10 model trained on AME16. The 1-neutron drip line is predicted to neutron number $N = 123^{125}_{95}$, and the 2-neutron drip line at $N = 115^{125}_{103}$. Where the error bar corresponds to a 68 % degree of belief.

Keywords: Bayesian Model Averaging, Duflo-Zuker 10, nuclear mass models, neutron drip line

Acknowledgements

I must express my gratitude to my supervisor (and examiner) Andreas Ekström for guidance and support throughout this project. I thank also the research group for their hospitality, even hosting virtual coffee breaks during the times of Corona pandemia that have been. Many thanks to my fellow students Jacob Olander, Marcus Lassila and Joel Postonen for fruitful discussions and inputs during the project.

Sebastian Lundquist, Gothenburg, June 2020

Contents

List of Symbols

Herein we will adopt a convention of using the Latin alphabet for variables connected with data and measurements, and use the Greek alphabet for variables dealing with models. We will also use bold fonts to denote vectors and matrices.

$D/\boldsymbol{d}$	Data set or vector
$\boldsymbol{E}$	Data covariance matrix
α	Model parameter(s)
$\boldsymbol{\Sigma}/\boldsymbol{\Psi}$	Parameter covariance matrix
$\boldsymbol{\Phi}/\phi$	Model design matrix/basis function
δ	Model discrepancy, inability to follow data
p(A)	Probability/Plausibility of proposition 'A'
$\pi(A)$	Prior probability/plausibility of proposition 'A'

1

Introduction

Theoretical modelling is a foundation for science and research, but often there can exist several theoretical models that equally well describe experimental data. Most often, a preferred model is selected, and typically one proceeds as if this model has generated the data. This has several drawbacks, most important is overconfident inference since this approach ignores model uncertainty. Bayesian model averaging (BMA) offers another route; use the entire set of candidate models, and average all model predictions weighted by their posterior model probability.[1, 2] This way, BMA can account for underlying model uncertainty.

In a Bayesian approach to statistics we can ask questions like 'how credible is this model?' or 'how plausible is this parameter?', which cannot be done in a frequentist sense. In this thesis we will focus on a BMA analysis of two nuclear mass models: the Duflo-Zuker 10 parameter model (DZ10)[3, 4] and the well-known semi-empirical mass formula (SEMF). By investigating such models and examining the credibility intervals of their parameters we can learn much about the underlying physics, and possibly better predict the stability of atomic nuclei.

Nuclear physics is a lively area of research and there exist different models for describing nuclear masses; e.g. a) *ab initio* based models,[5, and references therein] b) models based on nuclear density functional theory,[6] c) microscopically inspired phenomenological models such as the Duflo-Zuker (DZ) models,[3] d) purely phenomenological model e.g. the semi-empirical mass formula (SEMF). Models from category (c) and (d) are advantageous since they are easier to implement and do not require vast computational resources. Moreover, these models are typically linear in their parameters. This feature facilitates analytical expressions for the Bayesian marginalisation integrals. However, predictions from such mass models are typically made without accounting for the inherent model uncertainty. Theory is made to match experimental measurements in a pool of calibration data, but every model is, to some degree, approximate and truncated. To increase the reliability of mass predictions, one must account for theoretical deficiencies.

Several attempts have been made at predicting the neutron drip line and stability of visible matter,[6, 7] recently also employing Bayesian methods and BMA.[8–10] In this thesis we investigate the performance of BMA applied to two nuclear mass models, the DZ10 and the SEMF. They both have theoretically and experimentally motivated terms but the relative importance of them is less clear. Analysing them using Bayesian inference and BMA we have attempted to quantify model uncertainties and improve inference about

nuclear masses. To explore the robustness we compare the results of different choices of parameter priors, and vary the a priori assigned model uncertainty. The main focus will be on the DZ10, using the Atomic Mass Evaluation from 1983 (AME83)[11, 12] as training data, then evaluating the predictive power of the models using the Atomic Mass Evaluation from 2016 (AME16)[13, 14]. In an attempt at pushing the analysis of the limits of stability of visible matter we also make a prediction for the neutron drip lines in the isotopic chain, using the DZ10 model trained on AME16. We have developed the necessary Python code for achieving the goal of performing BMA as a part of the project.

2

Basics of Bayesian Inference and Nuclear Mass Models

In this chapter we will present the necessary theory for performing Bayesian model averaging of nuclear mass models. This will include a review of the statistical aspects, along with the mathematical tools for Bayesian inference. Then follows the basic theory regarding the phenomenological nuclear mass models of interest, the semi-empirical mass formula and Duflo Zuker 10 parameter model.

2.1 Bayesian Statistics and Model Averaging

Herein we will briefly introduce the Bayesian approach to statistics, then continue with how parameter estimation is performed. After that we review of advantages of analysing linear models, including a presentation of the equations that play a central role in BMA. Then follows a section which describes how BMA is performed.

2.1.1 Introduction to Bayesian Inference

The calculations in this paper will be based on *Bayesian* statistical inference. It differs from the more ubiquitous *frequentist* statistical inference, where the probability of a proposition is the relative frequency by which it occurs when repeating the experiment under identical conditions for a long time (infinitely many times). In this approach the proposition is limited to being about random variables, e.g. number of heads when flipping a coin N times. Using a Bayesian approach we expand the notion of probability of a proposition and relieve the restriction to random variables, and we can define e.g. the probability of a political party winning an election.[15–17] It is argued that Bayesian inference is subjective. But all theories are subjective, since they are limited to the available prior information, i.e. the current knowledge. As E.T. Jaynes stated, among others, objectivity only requires that two individuals with the same knowledge reach the same conclusion.[18]

Using the rules of *Boolean logic* and ordinary algebra R.T Cox found that mathematical consistency is only ensured if the real numbers attached to beliefs in propositions obey

the following rules of probability theory:[15]

$$\text{sum rule:} \quad \mathrm{p}(A) + \mathrm{p}\left(\bar{A}\right) = 1 \tag{2.1.1}$$

$$\text{product rule:} \quad \mathrm{p}(A,\, B) = \begin{cases} \mathrm{p}(A \,|\, B) \times \mathrm{p}(B) \\ \mathrm{p}(B \,|\, A) \times \mathrm{p}(A) \end{cases} \tag{2.1.2}$$

where A and B denote propositions, and $\bar{A}$ denotes A being false. $\mathrm{p}(A,\, B)$ denotes the joint probability of both propositions, A and B, being true. $\mathrm{p}(A \,|\, B)$ denotes the probability of proposition A being true conditional on B being true. From the product rule follows *Bayes' theorem*

$$\mathrm{p}(A \,|\, B) = \frac{\mathrm{p}(B \,|\, A) \times \mathrm{p}(A)}{\mathrm{p}(B)}. \tag{2.1.3}$$

Furthermore we can define *marginalisation* from these rules. If $\{B_i\}$ form a mutually exclusive and exhaustive set of possibilities, i.e. if B_i is satisfied then B_j, $j \neq i$, cannot be satisfied and $\sum_i \mathrm{p}(B_i) = 1$, marginalisation is defined as

$$\mathrm{p}(A) = \sum_k \mathrm{p}(A,\, B_k) \rightarrow \int \mathrm{d}B \; \mathrm{p}(A,\, B)\,, \tag{2.1.4}$$

where in the continuous limit the summation becomes an integration, where k runs over the set of propositions $\{B_k\}$ and the integration is performed over the corresponding space of propositions. Marginalisation can be a powerful tool in eliminating *nuisance parameters*, i.e. those which we are not interested in. For an example if we are interested in the frequency of a signal the amplitude would be a nuisance parameter.

We can rewrite Bayes' theorem in a more informative way, especially for the purposes of this report:

$$\mathrm{p}(\text{hypothesis} \,|\, \text{data},\, I) = \frac{\mathrm{p}(\text{data} \,|\, \text{hypothesis},\, I)\, \pi(\text{hypothesis} \,|\, I)}{\mathrm{p}(\text{data} \,|\, I)}, \tag{2.1.5}$$

where '$\pi(\text{hypothesis} \,|\, I)$' is called *prior probability* or *prior plausibility* and states our belief in the hypothesis before analysing the data. Here we have explicitly included the dependence on additional information, I, which is a description of the current knowledge. The quantity '$\mathrm{p}(\text{data} \,|\, \text{hypothesis}, I)$' is called *likelihood* and is a measure of how well our hypothesis predicts the observation. '$\mathrm{p}(\text{hypothesis} \,|\, \text{data}, I)$' is called *posterior probability* and is a measure of our belief in the hypothesis after we have analysed the data. The denominator '$\mathrm{p}(\text{data} \,|\, I)$' describes the marginal likelihood of observing the data, effectively it is a normalisation factor. In words Bayes' theorem states

> 'the posterior probability of the hypothesis being true, given the data, is proportional to the likelihood of the data, conditional on the hypothesis being true, multiplied by the prior probability of the hypothesis'.

Here it is manifest that our current knowledge and information I plays a role in assigning probability to propositions. It is also apparent that Bayes' theorem describes the process of learning from data, i.e. we update our belief of the hypothesis based on observed data. While it may seem that the conclusion we draw from given data is subjective as it depends on our prior belief in the hypothesis, this is partially true but the interpretation is debatable. First of all, all science is subjective in some sense as it is conducted by individuals. Different priors represent different states of knowledge. Secondly, if the choice of prior greatly affects our posterior probability it suggests our data is insufficient to draw a definite conclusion. If the data at hand is sufficient the choice of prior should not be a decisive factor regarding the conclusion we can draw from the data.

Further on in this thesis we will omit the explicit dependence on prior information I, as $\mathrm{p}(x \,|\, I) \to \mathrm{p}(x)$, but always remember that the dependence remains but implicit. Another point to have in mind is that the probability does not hold any real measure of the 'true value' of any proposition, but rather our belief in the proposition being true.

2.1.2 Bayesian Parameter Estimation

In data analysis, and science in general, estimating parameters has a central role. For example we can be interested in the mass of a planet or the fairness of a coin. In this section we will develop the foundations for our primary interest of model averaging. When estimating a parameter we are asking ourselves 'what is the true value of this parameter?'. In Bayesian theory we try to answer this by stating our belief in the value of a parameter. So, considering a theory of a continuous parameter α and a data set D, we want to determine the probability density function (pdf) $\mathrm{p}(\alpha \,|\, D)$. Using Bayes' theorem (2.1.3) we express this in terms of a prior and a likelihood:

$$\mathrm{p}(\alpha \,|\, D) \propto \pi(\alpha) \times \mathrm{p}(D \,|\, \alpha) \tag{2.1.6}$$

Where we have omitted the factor of $1/\mathrm{p}(D)$, which merely acts as a normalisation constant. Later this will become of interest as we will discuss several parameters and several models. To illustrate this more concretely, consider a classical example of tossing a coin. The throws are considered independent and the probability of the outcomes are described by a binomial distribution, where α is the probability of, say, 'heads'. Then we have the likelihood of observing H heads in N tosses given by

$$\mathrm{p}(\text{`}H \text{ heads in } N \text{ tosses'} \,|\, \alpha) \propto \alpha^H (1-\alpha)^{N-H} \tag{2.1.7}$$

The maximum likelihood estimate, α_0, of α is given by,

$$\alpha_0 : \left.\frac{\mathrm{dp}(\alpha \,|\, D)}{\mathrm{d}\alpha}\right|_{\alpha_0} = 0, \quad \left.\frac{\mathrm{d}^2\mathrm{p}(\alpha \,|\, D)}{\mathrm{d}\alpha^2}\right|_{\alpha_0} < 0 \tag{2.1.8}$$

if there is only one global maximum and no local maxima. A more thorough discussion on parameter estimation can be found in any good book on statistics, e.g. [16, 17].

To illustrate parameter estimation we can consider a data set[1] $D = \{d_i\}_{i=1}^{N}$ of independent measurements and represent the measurement errors by a Gaussian distribution.

$$d_i = f_i + e_i, \quad \text{where} \quad e_i \sim \mathcal{N}(0, \sigma_i) \equiv \frac{\exp\left(-e_i^2/2\sigma_i^2\right)}{\sqrt{2\pi}\,\sigma_i} \tag{2.1.9}$$

In general the model, f, for describing the data can be any mathematical expression but we will limit ourselves to so-called *linear models*.

2.1.3 Linear Models

A linear model, $\mathcal{M}$, is given by a mathematical expression f with linear dependence on its parameters:

$$\mathcal{M} : f(x) = \sum_{\kappa} \alpha_\kappa \phi_\kappa(x) \tag{2.1.10}$$

where α_κ are the model parameters and ϕ_κ represent the model basis functions. An example of a linear model is a one-variable polynomial of maximum degree N_p

$$\mathcal{M}(N_p) : f(x) = \sum_{\kappa=0}^{N_p} \alpha_\kappa x^\kappa \tag{2.1.11}$$

When analysing a linear model the mathematics of parameter estimation and likelihood evaluations simplify and we can sometimes arrive at analytical and closed form expressions. In the following sections we will show this explicitly for certain choices of parameter prior distribution.

We want to quantify our (un)certainty in a model. Our goal is then to obtain the posterior probability of the model, $\mathcal{M}$, given a set of data $D = \{d_i\}$, i.e. $\mathrm{p}(\mathcal{M} \,|\, D)$. This sounds simple enough, but there are several steps to compute this; we need to assign model priors, compute evidence integrals, which calls for assigning parameter priors and estimating likelihoods. We start with Bayes' theorem, by which we have

$$\mathrm{p}(\mathcal{M} \,|\, D) = \frac{\mathrm{p}(D \,|\, \mathcal{M})\,\pi(\mathcal{M})}{\mathrm{p}(D)}. \tag{2.1.12}$$

Where we need to assign a model prior $\pi(\mathcal{M})$, based on our current knowledge, this become more important later on, Sec. 2.1.4. The evidence of the model is given by the marginalisation of its parameters

$$\mathrm{p}(D \,|\, \mathcal{M}) = \int \mathrm{d}\boldsymbol{\alpha}\, \pi(\boldsymbol{\alpha} \,|\, \mathcal{M})\, \mathrm{p}(D \,|\, \mathcal{M}, \boldsymbol{\alpha})\,. \tag{2.1.13}$$

So to obtain a posterior model probability we need to evaluate this integral. This calls for assigning also a parameter prior $\pi(\boldsymbol{\alpha})$, but we start with specifying and simplifying

[1] This 'set' is not a mathematical definition but rather refers to the data as being one entity. It will be represented by a vector and we will not trouble ourselves with any rigorous mathematical difference between 'set' and 'vector'. But vectors will be written in bold font.

the data likelihood $\mathrm{p}(D \,|\, \mathcal{M}, \boldsymbol{\alpha})$. Assuming a Gaussian noise for the measurements, the likelihood for the data takes the form of a multivariate Gaussian distribution

$$\mathrm{p}(D \,|\, \mathcal{M}, \boldsymbol{\alpha}) = \frac{\exp\left(-\chi^2/2\right)}{(2\pi)^{N/2}\,|\boldsymbol{E}|^{1/2}}, \quad \chi^2 = (\boldsymbol{d}-\boldsymbol{f})^{\mathrm{T}}\,\boldsymbol{E}^{-1}\,(\boldsymbol{d}-\boldsymbol{f}) \tag{2.1.14}$$

where $\boldsymbol{d}$ denotes the data vector and $\boldsymbol{E}$ denotes the *covariance matrix* of the measurement errors. The quantity $\boldsymbol{E}^{-1}$ is often referred to as the *precision matrix*. We first search for a simple expression for χ^2 which separates the part independent from $\boldsymbol{\alpha}$ and the part that is dependent, this will allow us to evaluate the integral in Eq. (2.1.13). In the case of a linear model, with K basis functions, we can enhance our notation and define a *design matrix*

$$\boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{\phi}_1 & \cdots & \boldsymbol{\phi}_K \end{pmatrix} \tag{2.1.15}$$

where the column κ is the evaluation of basis function κ for the data vector $\boldsymbol{d}$, thus we can write the $\boldsymbol{f} = \boldsymbol{\Phi}\boldsymbol{\alpha}$. Furthermore define

$$\boldsymbol{\Psi}^{-1} = \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{E}^{-1}\boldsymbol{\Phi} \tag{2.1.16}$$

then $\boldsymbol{\Psi}$ can be considered as a covariance matrix for the model, or rather its basis functions. Then

$$\chi^2 = \boldsymbol{d}\boldsymbol{E}^{-1}\boldsymbol{d} + \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Psi}^{-1}\boldsymbol{\alpha} - 2\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{E}^{-1}\boldsymbol{d} \tag{2.1.17}$$

using the symmetry of $\boldsymbol{\Psi}$. We observe the first term is independent of the model parameters $\boldsymbol{\alpha}$, which is a first step to evaluating Eq. (2.1.13). To separate the rest we will find the maximum likelihood for the parameters, which is given by the least-squares solution. Assuming a unimodal distribution, i.e. with one unique maximum and no other peaks or plateaus, this is given by

$$\begin{aligned} &\frac{\partial \chi^2}{\partial \boldsymbol{\alpha}} = 2\boldsymbol{\Psi}^{-1}\boldsymbol{\alpha} - 2\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{E}^{-1}\boldsymbol{d} \stackrel{!}{=} 0 \\ \implies &\hat{\boldsymbol{\alpha}} = \boldsymbol{\Psi}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{E}^{-1}\boldsymbol{d}. \end{aligned} \tag{2.1.18}$$

So in order to simplify the integrand in Eq. (2.1.13), we can perform a change of variables by $\boldsymbol{\delta\alpha} = \boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}$, which yields

$$\chi^2 = (\boldsymbol{d} - \boldsymbol{\Phi}\hat{\boldsymbol{\alpha}})^{\mathrm{T}}\,\boldsymbol{E}^{-1}\,(\boldsymbol{d} - \boldsymbol{\Phi}\hat{\boldsymbol{\alpha}}) + \boldsymbol{\delta\alpha}^{\mathrm{T}}\boldsymbol{\Psi}^{-1}\boldsymbol{\delta\alpha} \tag{2.1.19}$$

The first term is completely independent of the parameters and the second term describes a Gaussian distribution. Recalling that we want to compute the model posterior, Eq. (2.1.12), hence we need to calculate the model evidence Eq. (2.1.13). Now that we obtained a nice simple expression for the likelihood, $\mathrm{p}(D \,|\, \mathcal{M}, \boldsymbol{\alpha})$, we stand before the task of specifying a prior pdf for the parameters, $\pi(\boldsymbol{\alpha})$, in order to evaluate the integral in Eq. (2.1.13). We will provide two alternatives: *uniform prior* and *normal prior*; both will allow us to perform the evidence integral analytically. A small discussion of the

implications of these choices and what they mean follows in the subsequent sections.

Uniform Parameter Prior

A uniform or *flat* prior for continuous model parameters is defined as

$$\pi(\boldsymbol{\alpha}) = \prod_{\kappa} \frac{\theta[\min(\alpha_\kappa) - \alpha_\kappa] - \theta[\max(\alpha_\kappa) - \alpha_\kappa]}{|\max(\alpha_\kappa) - \min(\alpha_\kappa)|} = \frac{\Theta(\boldsymbol{\Delta\alpha})}{\mathrm{Vol}(\boldsymbol{\Delta\alpha})} \tag{2.1.20}$$

Where $\theta(x)$ is the Heaviside step-function, and we abbreviate the sum and product by $\Theta(\boldsymbol{x})$. The prior pdf of all the parameters is given by the reciprocal of the volume of the hyper-cube which they span and zero outside it. Thus we can write the evidence as

$$\begin{aligned} \mathrm{p}(D\,|\,\mathcal{M}) = {} & \frac{\exp\left[-\frac{1}{2}(\boldsymbol{d} - \boldsymbol{\Phi}\hat{\boldsymbol{\alpha}})^{\mathrm{T}} \boldsymbol{E}^{-1} (\boldsymbol{d} - \boldsymbol{\Phi}\hat{\boldsymbol{\alpha}})\right]}{(2\pi)^{N/2} |\boldsymbol{E}|^{1/2}} \\ & \times \frac{1}{\mathrm{Vol}(\boldsymbol{\Delta\alpha})} \int_{\mathrm{Vol}(\boldsymbol{\Delta\alpha})} \mathrm{d}\boldsymbol{\alpha}\ \exp\left(-\frac{\boldsymbol{\delta\alpha}^{\mathrm{T}} \boldsymbol{\Psi}^{-1} \boldsymbol{\delta\alpha}}{2}\right) \end{aligned} \tag{2.1.21}$$

A quick remark: the first part, being the maximum likelihood estimator (MLE), is a multivariate Gaussian distribution, with covariance $\boldsymbol{E}$ and expectation value $\boldsymbol{\Phi}\hat{\boldsymbol{\alpha}}$. Furthermore we can identify the integrand to describe a posterior pdf for the parameters, as $\boldsymbol{\delta\alpha} = \boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}$,

$$\boxed{\mathrm{p}(\boldsymbol{\alpha}_k\,|\,\mathcal{M}, D) = \mathcal{N}\left(\boldsymbol{\alpha}\,|\,\hat{\boldsymbol{\alpha}}, \boldsymbol{\Psi}\right)} \tag{2.1.22}$$

Although truncated due to our prior. Now, since $\boldsymbol{\Psi}$ and of course $\boldsymbol{\Psi}^{-1}$ is a real symmetric matrix, we can decompose

$$\boldsymbol{\Psi}^{-1} = \boldsymbol{V}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}} \tag{2.1.23}$$

where $\boldsymbol{V}$ are real orthogonal matrices which columns are the eigenvectors of $\boldsymbol{\Psi}^{-1}$ and $\boldsymbol{S}$ is a diagonal matrix with the corresponding eigenvalues, λ_κ. Thus we can perform a orthogonal transformation, noting the Jacobian equals unity, into $\boldsymbol{\delta\alpha} = \boldsymbol{V}\boldsymbol{x}$ and

$$\boldsymbol{\delta\alpha}^{\mathrm{T}} \boldsymbol{\Psi}^{-1} \delta\boldsymbol{\alpha} = \sum_{\kappa} \lambda_\kappa x_\kappa^2. \tag{2.1.24}$$

With this transformation the integral can be performed analytically. If the chosen interval for the parameters, the prior distribution, is wide enough the integral can be well approximated by the result from the infinite case. Then the evidence the model is given

by

$$\boxed{\begin{aligned} \mathrm{p}(D\,|\,\mathcal{M}) &\approx \left[\frac{(2\pi)^{K/2}|\boldsymbol{\Psi}|^{1/2}}{\mathrm{Vol}(\boldsymbol{\Delta\alpha})}\right] \frac{\exp\left(-\chi^2_{\mathrm{min}}/2\right)}{(2\pi)^{N/2}|\boldsymbol{E}|^{1/2}} \\ &\equiv \Omega_{\mathcal{M}}\mathcal{L}_{\mathrm{max}} \end{aligned}} \tag{2.1.25}$$

The first factor is called *Occam* factor, as it accounts for the complexity of the model with dimension K i.e. number of parameters, in the sense that a model with more parameters that span a larger prior volume is considered to be more complex,

$$\Omega_{\mathcal{M}} = \frac{(2\pi)^{K/2}|\boldsymbol{\Psi}|^{1/2}}{\mathrm{Vol}(\boldsymbol{\Delta\alpha})}. \tag{2.1.26}$$

Here it is manifest that an unnecessary large range for the parameters will negatively contribute to the evidence for the model, as it appears in the denominator. A model with few parameters and if we are inclined to believe a certain narrow interval will receive low penalty and is thus considered more likely. A risk of narrow parameter interval is that we then exclude any values outside of that range, which means that we should use well informed ranges. The second factor in Eq. (2.1.25) is the maximum of the likelihood function, given by the inverse of the determinant of the data covariance matrix a constant factor of 2π's and an exponent

$$\chi^2_{\mathrm{min}} = (\boldsymbol{d} - \boldsymbol{\Phi}_k\hat{\boldsymbol{\alpha}}_k)^{\mathrm{T}}\boldsymbol{E}^{-1}(\boldsymbol{d} - \boldsymbol{\Phi}_k\hat{\boldsymbol{\alpha}}_k). \tag{2.1.27}$$

Now that we know how to calculate the likelihood of any model, when assigning uniform parameter priors, we can determine the posterior probability distribution Eq. (2.1.12), i.e. how much we believe in the model after considering the data at hand. But let us also discuss another choice of parameter prior.

Gaussian Parameter Prior

So picking up from Eq. (2.1.13) with the intent of calculating the model posterior Eq. (2.1.12), we assign a Gaussian distribution for the parameter prior. Gaussian distributions are conjugate to themselves, which means that assigning a Gaussian parameter prior, while having a Gaussian likelihood, the posterior will also be a Gaussian distribution. In the end we will reach an analytical expression for the model posterior as well as a parameter posterior, but we must work through quite a bit of algebra in order to do so. This discussion will be more general so we temporarily adopt a, in statistics, more standard notation.

The standard one-variable Gaussian distribution for a variable x is described by a mean, or expected, value μ and a variance σ^2:

$$\mathcal{N}\left(x\,|\,\mu,\,\sigma^2\right) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \tag{2.1.28}$$

The multivariate extension to a N-dimensional variable $\boldsymbol{x}$ is described by a expectation

value $\boldsymbol{\mu}$ in vector form and a covariance matrix $\boldsymbol{E}$

$$\mathcal{N}\left(\boldsymbol{x} \,|\, \boldsymbol{\mu}, \boldsymbol{E}\right) = \frac{1}{(2\pi)^{N/2}|\boldsymbol{E}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{E}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right] \tag{2.1.29}$$

where $|\boldsymbol{E}|$ denotes the determinant. In this form we can partition our variable, its expectation value and the covariance matrix, without loss of generality, according to

$$\boldsymbol{x} = \begin{pmatrix}\boldsymbol{x}_a \\ \boldsymbol{x}_b\end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix}\boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b\end{pmatrix}, \quad \boldsymbol{E} = \begin{pmatrix}\boldsymbol{E}_{aa} & \boldsymbol{E}_{ab} \\ \boldsymbol{E}_{ba} & \boldsymbol{E}_{bb}\end{pmatrix} \tag{2.1.30}$$

Where a and b correspond to some partitioning, for example if $\boldsymbol{x}$ is a vector of size n then $\boldsymbol{x}_a$ could denote the choice of the m first points and $\boldsymbol{x}_b$ is then the remaining $n-m$ points. Likewise for $\boldsymbol{\mu}_a$ and $\boldsymbol{\mu}_b$, and $\boldsymbol{E}_{aa}$ would be the $n \times n$ upper left part of $\boldsymbol{E}$ and $\boldsymbol{E}_{ab}$ would be the $n \times m$ upper right part. Considering only the quadratic form in the exponent in Eq. (2.1.29) we can isolate $\boldsymbol{x}_a$, or $\boldsymbol{x}_b$, in a new quadratic form. Dropping the $-\frac{1}{2}$ factor and for simplicity express in terms of the precision matrix $\boldsymbol{L} \equiv \boldsymbol{E}^{-1}$. Note: $\boldsymbol{L}_{aa}$ is not in general equal to $(\boldsymbol{E}_{aa})^{-1}$.

$$\begin{aligned}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{L}(\boldsymbol{x}-\boldsymbol{\mu}) = (\boldsymbol{x}_a-\boldsymbol{\mu}_a)^{\mathrm{T}}\boldsymbol{L}_{aa}(\boldsymbol{x}_a-\boldsymbol{\mu}_a) + (\boldsymbol{x}_a-\boldsymbol{\mu}_a)^{\mathrm{T}}\boldsymbol{L}_{ab}(\boldsymbol{x}_b-\boldsymbol{\mu}_b) \\ + (\boldsymbol{x}_b-\boldsymbol{\mu}_b)^{\mathrm{T}}\boldsymbol{L}_{ba}(\boldsymbol{x}_a-\boldsymbol{\mu}_a) + (\boldsymbol{x}_b-\boldsymbol{\mu}_b)^{\mathrm{T}}\boldsymbol{L}_{bb}(\boldsymbol{x}_b-\boldsymbol{\mu}_b)\end{aligned} \tag{2.1.31}$$

Since isolating any partition results in a new quadratic form this implies the conditional probabilities, e.g. $\mathrm{p}(\boldsymbol{x}_a \,|\, \boldsymbol{x}_b)$, are Gaussian. The quadratic form in the exponent of a Gaussian distribution appears as, omitting the constant factor,

$$(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{E}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{E}^{-1}\boldsymbol{x} - \boldsymbol{x}^{\mathrm{T}}\boldsymbol{E}^{-1}\boldsymbol{\mu} + \text{constant} \tag{2.1.32}$$

Further omitting all terms not dependent on $\boldsymbol{x}$. Identifying these terms allows us to identify the covariance and expectation value of distributions. Eq. (2.1.31) can be written in terms of order of, e.g., $\boldsymbol{x}_a$

$$\begin{aligned}&\mathcal{O}(\boldsymbol{x}_a^2): \quad \boldsymbol{x}_a^{\mathrm{T}}\boldsymbol{L}_{aa}\boldsymbol{x}_a \\ &\mathcal{O}(\boldsymbol{x}_a): \quad -\boldsymbol{x}_a^{\mathrm{T}}\left[\boldsymbol{L}_{aa}\boldsymbol{\mu}_a + \boldsymbol{L}_{ab}(\boldsymbol{x}_b-\boldsymbol{\mu}_b)\right] + \text{transpose}\end{aligned} \tag{2.1.33}$$

This implies that the conditional distribution is a Gaussian with

$$\begin{aligned}\mathrm{E}[\boldsymbol{x}_a \mid \boldsymbol{x}_b] &= \boldsymbol{\mu}_a + \boldsymbol{L}_{aa}^{-1}\boldsymbol{L}_{ab}(\boldsymbol{x}_b-\boldsymbol{\mu}_b) \\ \mathrm{cov}[\boldsymbol{x}_a \mid \boldsymbol{x}_b] &= \boldsymbol{L}_{aa}^{-1}\end{aligned} \tag{2.1.34}$$

This process can performed in reverse, which is in our interest as have a conditional normal pdf and a normal prior pdf. Defining a new variable $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{y})$, where $\mathrm{p}(\boldsymbol{x}) = \mathcal{N}\left(\boldsymbol{x} \,|\, \boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$ and $\mathrm{p}(\boldsymbol{y} \,|\, \boldsymbol{x}) = \mathcal{N}\left(\boldsymbol{y} \,|\, \boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}, \boldsymbol{E}\right)$. The quadratic form appearing in $\mathrm{p}(\boldsymbol{z}) = \mathrm{p}(\boldsymbol{x})\,\mathrm{p}(\boldsymbol{y} \,|\, \boldsymbol{x})$, omitting the constant factor of $-1/2$ is then

$$(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) + (\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})^{\mathrm{T}}\boldsymbol{E}^{-1}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}) \tag{2.1.35}$$

Which manifestly is a quadratic form in $\boldsymbol{y}$, thus $\mathrm{p}(\boldsymbol{y})$ is another normal distribution. Rewriting in terms of $\boldsymbol{z}$ we find the terms quadratic and linear in $\boldsymbol{z}$

$$\boldsymbol{z}^{\mathrm{T}} \begin{pmatrix} \boldsymbol{\Sigma}^{-1} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{E}^{-1}\boldsymbol{A} & -\boldsymbol{A}^{\mathrm{T}}\boldsymbol{E}^{-1} \\ -\boldsymbol{E}^{-1}\boldsymbol{A} & \boldsymbol{E}^{-1} \end{pmatrix} \boldsymbol{z} - \boldsymbol{z}^{\mathrm{T}} \begin{pmatrix} \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{E}^{-1}\boldsymbol{b} \\ \boldsymbol{E}^{-1}\boldsymbol{b} \end{pmatrix} \tag{2.1.36}$$

By Eq. (2.1.32) we find

$$\mathrm{p}(\boldsymbol{z}) = \mathcal{N}\left(\boldsymbol{z} \,|\, \boldsymbol{R}^{-1}\boldsymbol{M},\, \boldsymbol{R}^{-1}\right) \tag{2.1.37}$$

where we define the precision matrix and expectation value of $\boldsymbol{z}$ according to

$$\boldsymbol{R} \equiv \begin{pmatrix} \boldsymbol{\Sigma}^{-1} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{E}^{-1}\boldsymbol{A} & -\boldsymbol{A}^{\mathrm{T}}\boldsymbol{E}^{-1} \\ -\boldsymbol{E}^{-1}\boldsymbol{A} & \boldsymbol{E}^{-1} \end{pmatrix}, \quad \boldsymbol{M} \equiv \begin{pmatrix} \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{E}^{-1}\boldsymbol{b} \\ \boldsymbol{E}^{-1}\boldsymbol{b} \end{pmatrix} \tag{2.1.38}$$

The inverse of a partitioned matrix is found by the *Schur complement*, $\boldsymbol{S}/\boldsymbol{D}$, of block $\boldsymbol{D}$ of matrix $\boldsymbol{S}$

$$\begin{aligned} \boldsymbol{S} &\equiv \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix}, \qquad \boldsymbol{S}/\boldsymbol{D} \equiv \boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C} \\ \implies \boldsymbol{S}^{-1} &= \begin{pmatrix} (\boldsymbol{S}/\boldsymbol{D})^{-1} & -(\boldsymbol{S}/\boldsymbol{D})^{-1}\boldsymbol{B}\boldsymbol{D}^{-1} \\ -\boldsymbol{D}^{-1}\boldsymbol{C}(\boldsymbol{S}/\boldsymbol{D})^{-1} & \boldsymbol{D}^{-1} + \boldsymbol{D}^{-1}\boldsymbol{C}(\boldsymbol{S}/\boldsymbol{D})^{-1}\boldsymbol{B}\boldsymbol{D}^{-1} \end{pmatrix}, \end{aligned} \tag{2.1.39}$$

assuming the necessary properties of $\boldsymbol{D}$ and its Schur complement being invertible. Using this identity we find the inverse of our matrix $\boldsymbol{R}$ quite easily

$$\mathrm{cov}[\boldsymbol{z}] = \boldsymbol{R}^{-1} = \begin{pmatrix} \boldsymbol{\Sigma}_k & \boldsymbol{\Sigma}_k\boldsymbol{A}^{\mathrm{T}} \\ \boldsymbol{A}\boldsymbol{\Sigma}_k & \boldsymbol{E} + \boldsymbol{A}\boldsymbol{\Sigma}_k\boldsymbol{A}^{\mathrm{T}} \end{pmatrix} \tag{2.1.40}$$

which is the covariance matrix for the variable $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{y})$, and furthermore its expectation value

$$\mathrm{E}[z] = \boldsymbol{R}^{-1}\boldsymbol{M} = \begin{pmatrix} \boldsymbol{\mu}_k \\ \boldsymbol{A}\boldsymbol{\mu}_k + \boldsymbol{b} \end{pmatrix} \tag{2.1.41}$$

We can thus identify the distributions for the partitioned variables.[19] In our case the pdf of interest is the marginal distribution of the variable $\boldsymbol{y}$. Given a marginal Gaussian distribution for a variable $\boldsymbol{x}$ and a conditional Gaussian distribution for $\boldsymbol{y}$ given $\boldsymbol{x}$ in the form

$$\begin{cases} \mathrm{p}(\boldsymbol{x}) = \mathcal{N}\left(\boldsymbol{x} \,|\, \boldsymbol{\mu},\, \boldsymbol{\Sigma}\right) \\ \mathrm{p}(\boldsymbol{y} \,|\, \boldsymbol{x}) = \mathcal{N}\left(\boldsymbol{y} \,|\, \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b},\, \boldsymbol{E}\right) \end{cases} \tag{2.1.42}$$

the marginal distribution of the variable $\boldsymbol{y}$ and the conditional distribution for $\boldsymbol{x}$ given $\boldsymbol{y}$

are given by

$$\begin{cases} \mathrm{p}(\boldsymbol{y}) = \mathcal{N}\left(\boldsymbol{y} \,|\, \boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b},\, \boldsymbol{E} + \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^{\mathrm{T}}\right) \\ \mathrm{p}(\boldsymbol{x} \,|\, \boldsymbol{y}) = \mathcal{N}\left(\boldsymbol{x} \,|\, \boldsymbol{\Lambda}[\boldsymbol{A}^{\mathrm{T}}\boldsymbol{E}^{-1}(\boldsymbol{y} - \boldsymbol{b}) + \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}],\, \boldsymbol{\Lambda}\right) \end{cases} \tag{2.1.43}$$

where $\boldsymbol{\Lambda} = (\boldsymbol{\Sigma}^{-1} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{E}^{-1}\boldsymbol{A})^{-1}$. We have during this entire derivation neglected all normalisation factors. The reason is simple: we know the end result must be a probability distribution, thus normalised. The dependence of the variable takes the shape of a quadratic form in an exponent, i.e. a Gaussian function. Thus the final pdf must be a Gaussian distribution. Which means we have found analytical, closed form expressions relating Gaussian marginal, $\mathrm{p}(\boldsymbol{x})$, and conditional probabilities, $\mathrm{p}(\boldsymbol{y} \,|\, \boldsymbol{x})$, of propositions with their converses, $\mathrm{p}(\boldsymbol{y})$ and $\mathrm{p}(\boldsymbol{x} \,|\, \boldsymbol{y})$. This is just our problem we seek to solve.

To make it explicit for our intents, repeating ourselves: we are interested in the marginal likelihood

$$\mathrm{p}(D \,|\, \mathcal{M}) = \int \pi(\boldsymbol{\alpha} \,|\, \mathcal{M}) \mathrm{d}\boldsymbol{\alpha}\; \mathrm{p}(D \,|\, \boldsymbol{\alpha}, \mathcal{M}) \tag{2.1.44}$$

Where we assume a Gaussian error in the data, and our model is linear, thus

$$\mathrm{p}(D \,|\, \boldsymbol{\alpha}, \mathcal{M}) = \mathcal{N}\left(\boldsymbol{d} \,|\, \boldsymbol{\Phi}\boldsymbol{\alpha}, \boldsymbol{E}\right). \tag{2.1.45}$$

Furthermore we adopt a Gaussian distribution for the parameter prior according to

$$\pi(\boldsymbol{\alpha} \,|\, \mathcal{M}) = \mathcal{N}\left(\boldsymbol{\alpha} \,|\, \boldsymbol{\mu}, \boldsymbol{\Sigma}\right) \tag{2.1.46}$$

which naturally raises the question of choice of the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. This will be dealt with in the next section. Then the marginal likelihood of the data conditional on $\mathcal{M}$ and the posterior parameter probability is found.

Having known distributions

$$\begin{cases} \mathrm{p}(D \,|\, \boldsymbol{\alpha}, \mathcal{M}) = \mathcal{N}\left(\boldsymbol{d} \,|\, \boldsymbol{\Phi}\boldsymbol{\alpha}, \boldsymbol{E}\right) \\ \pi(\boldsymbol{\alpha} \,|\, \mathcal{M}) = \mathcal{N}\left(\boldsymbol{\alpha} \,|\, \boldsymbol{\mu}, \boldsymbol{\Sigma}\right) \end{cases}$$

we then obtain

$$\begin{cases} \mathrm{p}(D \,|\, \mathcal{M}) = \mathcal{N}\left(\boldsymbol{d} \,|\, \boldsymbol{\Phi}\boldsymbol{\mu}, \boldsymbol{E} + \boldsymbol{\Phi}\boldsymbol{\Sigma}\boldsymbol{\Phi}^{\mathrm{T}}\right) \\ \mathrm{p}(\boldsymbol{\alpha} \,|\, \mathcal{M}, D) = \mathcal{N}\left(\boldsymbol{\alpha} \,|\, \boldsymbol{\Lambda}\left[\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{E}^{-1}\boldsymbol{d} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right], \boldsymbol{\Lambda}\right) \end{cases}$$

where $\boldsymbol{\Lambda} = \left(\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{E}^{-1}\boldsymbol{\Phi}\right)^{-1}$ (2.1.47)

Parameter Prior Probability

The choice of prior can greatly affect the posterior pdf if the data at hand is limited. Or rather if the prior has a great influence on posterior it implies either our question or our theory is ill defined, or the data at hand is insufficient. Importantly, the choice of prior is a statement about our current knowledge, or rather lack thereof, to describe the theory. When performing parameter estimation we encountered the need of parameter prior pdf. We will focus on two choices; the uniform prior, and the Gaussian (or normal) prior. The latter is conjugate to the Gaussian likelihood. The advantage of a uniform prior is that it is constant over a range, the equally large drawback is that it neglects all parameter space outside this range. This is reflected in the posterior distribution. If our ignorance about the region in which the parameter takes its true value is total, i.e. excluding it, the posterior will always be zero, which could be a fatal flaw. If the parameter values are completely unknown, a finite uniform prior may be a poor choice.

The advantage of a Gaussian prior, as seen in Sec. 2.1.3, is that with a Gaussian error also the posterior distributions are Gaussian. This makes computations easier as the calculations can be done analytically and the results are analytical standard expressions. In Sec. 2.1.3 we adopted a Gaussian distribution for the parameter prior, $\mathrm{p}(\boldsymbol{\alpha}) = \mathcal{N}(\boldsymbol{\alpha} \,|\, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the question of assignment of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ appears. One choice is ignorance in location, i.e. setting the expectation value to zero, together with a Zellner g-prior for the covariance matrix

$$\boldsymbol{\mu} = \mathbf{0} \quad \text{and} \quad \boldsymbol{\Sigma} = g\left(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}\right)^{-1} \tag{2.1.48}$$

So the covariance matrix is a scalar multiple of the *Fisher information matrix*.[20] Typical values used are $g = 1$ and $g =$# data points, we will use $g =$#data points, in our analysis. We will focus on three types of parameter priors: uniform prior, Zellner g-prior, Gaussian prior.

$$\boxed{\begin{aligned} \pi_{\mathrm{U}}(\boldsymbol{\alpha}) &= \frac{\boldsymbol{\Theta}(\boldsymbol{\Delta\alpha})}{\mathrm{Vol}(\boldsymbol{\Delta\alpha})} \\ \pi_{\mathrm{N,\,Z}}(\boldsymbol{\alpha}) &= \mathcal{N}\left(\boldsymbol{\alpha} \,|\, \mathbf{0}, g\left[\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}\right]^{-1}\right) \\ \pi_{\mathrm{N}}(\boldsymbol{\alpha}) &= \mathcal{N}\left(\boldsymbol{\alpha} \,|\, \mathbf{0}, \sigma^2\mathbf{1}\right) \\ \pi_{\mathrm{N,\,phys.}}(\boldsymbol{\alpha}) &= \mathcal{N}(\boldsymbol{\alpha} \,|\, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \end{aligned}} \tag{2.1.49}$$

Where the size of the uniform hyper-cube in $\pi_{\mathrm{U}}(\boldsymbol{\alpha})$ can be set to dismiss negative parameter values as well as values of absurd magnitude. The g-prior, $\pi_{\mathrm{N,\,Z}}(\boldsymbol{\alpha})$, uses the values of the basis functions over the region of interest to normalise the parameter priors. This allows a simple and fast assignment for many parameters where there is lack of knowledge for a reasonable region for the parameter values. It also considers the presence of other parameters; a difficulty which arises when hand-picking the priors. The Gaussian prior, $\pi_{\mathrm{N}}(\boldsymbol{\alpha})$, will be independent and identically distributed for all parameters, centred around

0 and with a large standard deviation, σ, allowing for varying solutions of parameters. The physically inspired prior, $\pi_{\text{N, phys.}}(\boldsymbol{\alpha})$, will be unique for the model under study. Using theoretical arguments and previous information both the mean values and standard deviation will be assigned for each parameter.

2.1.4 Model Averaging

An often encountered problem in science is the choice of model, as several models can equally well describe data. Instead of only choosing one model, based on plausibility or other measures, we can average predictions using several models. In order to do this we must first start with a clarification of the situation we consider. Experiments measure quantities with errors e_i, which we will assume to be Gaussian but not necessarily independent,

$$d_i = y(x_i) + e_i. \tag{2.1.50}$$

Depending on the quality of the measurements, the data almost describes the physical process, $y(x)$, where x is some variable. This measurement error, e is typically small. We also adopt a *model discrepancy*, $\delta(x)$, which takes into consideration the inability of the model to correctly describe the physical processes. [21–24]

$$y(x) = f(x) + \delta(x) \tag{2.1.51}$$

This model discrepancy, which is independent of measurements, may well be order of magnitudes larger than the measurement error. The model discrepancy is deterministic, but unknown, and we will employ a Gaussian distribution for inference. A Gaussian maximises entropy,[16, 17] i.e. it is the most conservative, if we specify only the variance of the model discrepancy. Then we can combine the errors into $\sigma \sim \sqrt{e^2 + \delta^2} \approx \delta$. Thus, limiting ourselves to a linear model $f(x) = \boldsymbol{\phi}\boldsymbol{\alpha}$,

$$d_i = \boldsymbol{\phi}(x_i)\boldsymbol{\alpha} + \sigma \quad \Longrightarrow \quad \boxed{\mathrm{p}(\boldsymbol{d} \,|\, \alpha, \mathcal{M}) = \mathcal{N}\left(\boldsymbol{d} \,|\, \boldsymbol{\Phi}\boldsymbol{\alpha}, \sigma^2\mathbf{1}\right)} \tag{2.1.52}$$

i.e. we have an independent multivariate distribution for the data.

Thus far we have only considered one model for describing the data, but Bayesian inference is much more powerful. We can instead consider the model in the previous part as one from a set of models, i.e. $\mathcal{M} \to \mathcal{M}_k \in \{\mathcal{M}_k\}$. The above calculations, Sec. 2.1.3, hold just fine, we merely replace $\mathcal{M}$ by $\mathcal{M}_k$ and $\boldsymbol{\alpha}$ by $\boldsymbol{\alpha}_k$. In this thesis we will consider a *global model* $\boldsymbol{\mathcal{M}}$ with K parameters, and the set of models is given by all combinations of the parameters, i.e. $\mathcal{M}_k \subseteq \boldsymbol{\mathcal{M}}$ so $\mathcal{M}_k \in \{\mathcal{M}_k\}_{k=1}^{2^K-1}$. Explicitly, if $\boldsymbol{\mathcal{M}}: \ f(x) = \alpha_0 + \alpha_1 x$ then the set of models can consist of $\mathcal{M}_1: \ \alpha_0$, $\mathcal{M}_2: \ \alpha_1 x$ and $\mathcal{M}_3: \ \alpha_0 + \alpha_1 x$. Employing Bayesian methods to accomplish model averaging can have positive effects on predictive power as well as error estimation.[1, 2, 8, 21, 22]

Starting from a linear model it is possible to construct several different models, by simply choosing a subset of the terms that appear. This indicates the presence of an underlying model uncertainty, which can be quantified using BMA. The dimensionality of the set of

models grows exponentially with the number of parameters, K. To tackle this we exploit the linearity of all considered models, which allows derivation of an analytical expression for the model evidence, see Sec. 2.1.3. With the posterior model probability known we can average the posterior distribution of any quantity of interest Δ, e.g. nuclear binding energy, according to

$$\mathrm{p}(\Delta \,|\, D) = \sum_k \mathrm{p}(\Delta \,|\, \mathcal{M}_k,\, D)\, \mathrm{p}(\mathcal{M}_k \,|\, D)\,. \tag{2.1.53}$$

How to calculate the posterior model probability, $\mathrm{p}(\mathcal{M}_k \,|\, D)$, as well as the posterior parameter distribution, $\mathrm{p}(\boldsymbol{\alpha}_k \,|\, \mathcal{M}_k, D)$, was discussed earlier in Sec. 2.1.3. If we are interested in predicting measurements, i.e. $\Delta = d^\star$, then this is performed by marginalisation of the parameters

$$\mathrm{p}(\Delta = d^\star \,|\, \mathcal{M}_k,\, D) = \int \mathrm{p}(\boldsymbol{\alpha}_k \,|\, \mathcal{M}_k, D)\mathrm{d}\boldsymbol{\alpha}_k\; \mathrm{p}(d^\star \,|\, \mathcal{M}_k, \boldsymbol{\alpha}_k, D) \tag{2.1.54}$$

So the prediction, and its uncertainty, is averaged and weighted by the corresponding posterior model probabilities. Averaging over several models can improve the predictive ability in the sense of more realistic uncertainties.

Furthermore to improve the estimation of various parameters in the models we can calculate the weighted posterior distribution of the parameter α_κ given the data

$$\mathrm{p}(\alpha_\kappa \,|\, D) = \sum_k \mathrm{p}(\alpha_\kappa \,|\, \mathcal{M}_k, D)\, \mathrm{p}(\mathcal{M}_k \,|\, D)\,. \tag{2.1.55}$$

Model Prior Probability

The parameter posterior probability is of great interest, but so is the posterior model probability. So let us recall Eq. (2.1.12), and realise we now know how to compute the model evidence, Eq. (2.1.13). We even have analytical closed form solutions when adopting uniform parameter priors, Eq. (2.1.25), and Gaussian parameter priors, Eq. (2.1.47), which are the priors we will employ. The quantity that remains is the model prior probability, $\pi(\mathcal{M}_k)$. If we are naive a natural choice would be a uniform distribution, i.e. for a set of models $\{\mathcal{M}_k\}$ being all combinations of a global model with K parameters, we have

$$\pi(\mathcal{M}_k) = \frac{1}{2^K}. \tag{2.1.56}$$

While this choice seems intuitive and reasonable it may be rushed, as it does not consider the correlation between model parameters. This correlation, depending on basis functions, can be very high. An example of highly correlating basis functions are polynomials. A solution to this is to use a so-called *dilution prior* which deals with correlation among model parameters.[25] Dilution priors can be achieved by multiple methods or considerations; *tessellation*, *collinearity*, *model distance*. Adopting a dilution prior based on the collinearity of the models such that it reduces to the uniform case when the models are uncorrelated, is a possibility of reducing the number of redundant models. Let R_k be the correlation matrix of model k, i.e. $R_k \propto \boldsymbol{\Phi}_{k,\kappa}\boldsymbol{\Phi}_{k,\lambda}$, and note that if the design matrix is

orthogonal then $|R_k| = 1$ and as the basis functions correlate more $|R_k|$ reduces to zero. So we choose

$$\pi(\mathcal{M}_k) \propto h\left(|R_k|\right) \frac{1}{2^K} \tag{2.1.57}$$

for some monotonic function satisfying $h(1) = 1$ and $h(0) = 0$. E. I. George suggests $h(r) = r^p$ for two choices: $p = 1/2$, and $p = 1$.[25]

$$\boxed{\pi_D(\mathcal{M}_k) = \frac{|\boldsymbol{R}_k|^p \; 2^{-K}}{\sum_l |R_l|^p \; 2^{-K}}} \tag{2.1.58}$$

2.2 Review of Nuclear Mass Models

In Ch. 4 we will analyse the SEMF and the DZ10 nuclear mass models. So here follows a brief description.

2.2.1 SEMF

The semi-empirical mass formula, also known as Bethe-Weizsäcker mass formula, models the nucleus as a liquid drop. The formula contains physically motivated terms but the 5 parameters α_i are inferred from nuclear mass-data. The binding energy of a nucleus with a total of $A = Z + N$ nucleons, Z protons and N neutrons, is given by

$$\mathrm{BE}_{\mathrm{SEMF}} = \alpha_V A - \alpha_S A^{2/3} - \alpha_C \frac{Z(Z-1)}{A^{1/3}} - \alpha_A \frac{(N-Z)^2}{A} + \alpha_P \frac{\delta(A,Z)}{A^{1/2}}. \tag{2.2.1}$$

Where the terms, in order, correspond to effects that scale with the volume, surface, Coulomb, asymmetry, and pairwise interactions. The last term considers the pairing of nucleons of the same type and is simply given by

$$\delta(A,Z) = \begin{cases} +1, \; Z \text{ even}, \; N \text{ even} \\ 0, \; A \text{ odd} \\ -1, \; Z \text{ odd}, \; N \text{ odd} \end{cases} \tag{2.2.2}$$

Noteworthy is the existence of different versions of the SEMF, omitting the pairing term, or the Coulomb term being proportional to Z^2, or the asymmetry term in terms of $T(T+1)$ for $T = (N-Z)/A$. The first three terms (volume, surface, and Coulomb) can be seen as expansion in $A^{1/3} \sim$ radius, i.e. proportional to the size of the nucleus. The fourth term (asymmetry) is also proportional to the squared neutron excess, i.e. neutron-proton asymmetry, hence the name. This is expansion in terms of radius referred to as a *leptodermous expansion*.

The terms all have natural explanations. Using a Fermi gas model together with experimental data and physical arguments we can estimate the magnitude of the volume term and the asymmetry term. $\alpha_V \sim 10\,\mathrm{MeV}$ and $\alpha_A \sim 50\,\mathrm{MeV}$. Furthermore the surface

term is expected to be of the same order as the volume term. The Coulomb term can be theoretically estimated from electrodynamics as

$$\alpha_C = \frac{3}{5}\frac{e^2}{4\pi\epsilon_0 r_0} \approx 0.69\,\mathrm{MeV} \tag{2.2.3}$$

Where e is elementary charge, ϵ_0 the vacuum permittivity. $r_0 \approx 1.25\,\mathrm{fm}$ is the nuclear radius constant relating the atomic radius and the mass number as $R = r_0 A^{1/3}$. The SEMF can be used for guiding estimates of the binding energy

A major flaw of the SEMF is its inability to predict so-called *magic numbers*, i.e. the pattern of higher binding energy per nucleon, i.e. particularly stable nuclei, for certain numbers of protons and neutrons.

2.2.2 Duflo-Zuker Mass Models

We will only briefly describe the physical reasoning behind the DZ mass models as they are, although microscopically inspired, quite phenomenological.

The DZ mass model assumes a shell-structure of the nucleus, based on the harmonic oscillator (HO) model, and writes the Hamiltonian as a sum of a monopole part and a multipole part.[3, 4, 26–28] The binding energy can be written as

$$\mathrm{BE_{DZ}} = \langle \mathrm{H}_m \rangle + E_\mathrm{C} + E_\mathrm{sym} + E_\mathrm{pairing} \tag{2.2.4}$$

where E_C, E_sym and E_P are energy due to Coulomb interaction (which is negative, i.e. repulsive), (a)symmetry (negative) and pairing, respectively, reminiscent of the terms in the SEMF but not identical. The monopole term, H_m, describes a mean field based on an interacting shell model. Instead of solving for the mean field for every nucleus, Duflo and Zuker parameterised the energy as a sum of phenomenological terms $\mathrm{H}_m = \mathrm{H}_{Ma} + \mathrm{H}_s + \mathrm{H}_d$ to reflect energy variations due to breaking of spherical symmetry. These terms, H_{Ma}, H_s and H_d, are parameterised to reflect sequential filling of a pre-assumed harmonic oscillator shell-structure of the nucleus. In a HO model, the number of neutrons and protons are given by

$$N = \sum_{p=0} n_p \quad \text{and} \quad Z = \sum_{p=0} z_p \tag{2.2.5}$$

where p is the principal quantum number of the HO major shell, and n_p (z_p) is the number of neutrons (protons) in this shell. The degeneracy of a major shell is $D_p = (p+1)(p+2)$, for a fully filled shell we have $n_p = D_p$, and likewise for protons. The monopole part, which only contains quadratic forms of the number, $m_p = n_p + z_p$, and isospin operators, $T_p = |n_p - z_p|/2$, can be written

$$\mathrm{H}_m = \sum_{p,\,q} a_{pq} m_p \left(m_q - \delta_{pq}\right) + b_{pq}\left(T_p \cdot T_q - \frac{3m_p}{4}\delta_{pq}\right) \tag{2.2.6}$$

where the sum runs twice over all orbits p, thus including interactions. By geometric

properties of realistic forces, Duflo & Zuker reach a general rule regarding the master monopole terms: they 'must be symmetric quadratic forms of properly scaled operators'.[3] Duflo & Zuker construct in total 12 symmetric quadratics which belong to the master monopole part H_{Ma}, along with a $4T(T+1)$ term, a multipole pairing term included as its form is equal for all nuclei, and a Coulomb term which contains multipole effects but these can be ignored. But when comparing to data only some terms show significant contribution.[3] The terms appearing in the final expression for the binding energy are several and the true importance of each term is ill-understood, mostly measured by the error in the prediction or fit to data. In Ch. 4 we will employ Bayesian methods to analyse the actual evidence of the DZ models and its various terms.

Before we delve into the specific formulations of the DZ models, we will make an observation: In the SEMF there is an implicit radius $r = A^{1/3}$. In the DZ mass models this is replaced by

$$R = A^{1/3}\left[1 - \left(\frac{T}{A}\right)^2\right]^2 \tag{2.2.7}$$

for all terms except the Coulomb term, which is scaled by

$$R_C = A^{1/3}\left[1 - \left(\frac{T}{A}\right)^2\right]. \tag{2.2.8}$$

Note that both reduce to $r = A^{1/3}$ when the neutron number and proton number coincide $N = Z$.

DZ10 Model

One of the DZ models often used is the DZ10 version with 10 parameters. First we will give these in a cryptic short-hand notation which is consistent with the formalism of Duflo & Zuker, and then we will, in short, explain these terms.

$$\begin{aligned}\mathrm{BE_{DZ}} = &-\alpha_1\,(\text{Coulomb}) + \alpha_2\,(\text{FM volume} + \text{SO}) - \alpha_3\,(\text{FM surface})\\ &-\alpha_4\,(\text{isospin/symmetry volume})\\ &+\alpha_5\,(\text{isospin/symmetry surface} + \text{Wigner})\\ &+\alpha_6\,(\mathrm{S_3}\text{ volume}) - \alpha_7\,(\mathrm{S_3}\text{ surface})\\ &+\alpha_8\,(\text{QQ spherical}) + \alpha_9\,(\text{QQ deformed}) + \alpha_{10}\,(\text{pairing})\end{aligned} \tag{2.2.9}$$

We will write these out explicitly, in order, note the signs of expressions, the binding energy is given by the sum of the following expressions. The Coulomb energy is given by

$$\boxed{E_{\mathrm{C}} = -\alpha_1 \frac{Z(Z-1) - 0.76\left[Z(Z-1)\right]^{2/3}}{A^{1/3}\left[1 - \left(\frac{T}{A}\right)^2\right]}} \tag{2.2.10}$$

which resembles the Coulomb term appearing in the SEMF, Eq. (2.2.1). The so-called 'master term', FM volume+SO, describes the dominant term in the monopole Hamiltonian, corrected for spin-orbit (SO) effects so that it exhibits the experimentally observed *extruder-intruder* behaviour. The first two terms are the master monopole and the two following are SO terms to ensure the observed magicity instead of HO shell-structure,

$$\begin{aligned}\boxed{E_{\text{FM vol.+SO}} = \frac{\alpha_2}{R}\Bigg\{ & \left(\sum_p \frac{n_p}{\sqrt{D_p}}\right)^2 + \left(\sum_p \frac{z_p}{\sqrt{D_p}}\right)^2 \\ & + \sum_p \left(\frac{p n_{jp}}{2} - n_{rp}\right)\left[\left(1 + \frac{n_p}{\sqrt{D_p}}\right)\frac{p^2}{D_p^{3/2}} + \left(1 - \frac{n_p}{\sqrt{D_p}}\right)\frac{4p-5}{D_p^{3/2}}\right] \\ & + \sum_p \left(\frac{p z_{jp}}{2} - z_{rp}\right)\left[\left(1 + \frac{z_p}{\sqrt{D_p}}\right)\frac{p^2}{D_p^{3/2}} + \left(1 - \frac{z_p}{\sqrt{D_p}}\right)\frac{4p-5}{D_p^{3/2}}\right]\Bigg\}}\end{aligned} \tag{2.2.11}$$

which is a rather lengthy and sophisticated expression. The master surface term is a bit shorter

$$\boxed{E_{\text{FM surf.}} = -\frac{\alpha_3}{R^2}\left[\left(\sum_p \frac{n_p}{\sqrt{D_p}}\right)^2 + \left(\sum_p \frac{z_p}{\sqrt{D_p}}\right)^2\right]} \tag{2.2.12}$$

The isospin or symmetry terms, volume and surface corrected by Wigner, is given by

$$\boxed{E_{\text{sym}} = -\frac{\alpha_4}{R}\frac{4T(T+1)}{A^{2/3}R} + \frac{\alpha_5}{R^2}\left(\frac{4T(T+1)}{A^{2/3}R^2} - \frac{4T\left(T-\frac{1}{2}\right)}{AR^3}\right)} \tag{2.2.13}$$

The spherical part of the Hamiltonian is given by the S_3 volume term, and corresponding surface term and the 'QQ spherical' term,

$$\boxed{\begin{aligned} E_{\text{spherical}} = \frac{\alpha_6}{R}\left(\frac{n\bar{n}(n-\bar{n})}{D_\nu} + \frac{z\bar{z}(z-\bar{z})}{D_\pi}\right) - \frac{\alpha_7}{R^2}\left(\frac{n\bar{n}(n-\bar{n})}{D_\nu} + \frac{z\bar{z}(z-\bar{z})}{D_\pi}\right) \\ + \frac{\alpha_8}{R}\frac{2^{\sqrt{p_\nu}}n\bar{n}}{D_\nu}\cdot\frac{2^{\sqrt{p_\pi}}n\bar{n}}{D_\pi} \end{aligned}} \tag{2.2.14}$$

where n is the number of valence neutrons and $\bar{n} = D_\nu - n$ is the number of corresponding valence holes in the open SO major shell with principal quantum number p_ν and degeneracy D_ν. z is the number of valence protons and $\bar{z} = D_\pi - z$ is the number of corresponding valence holes holes in the open SO major shell with principal quantum number p_π and degeneracy D_π. So far the terms neglect the deformation of nuclei, this resides in the 'QQ deformed' part,

$$\boxed{E_{\text{deformed}} = \frac{\alpha_9}{R}\frac{n'\bar{n}'}{D_\nu^{3/2}}\cdot\frac{z'\bar{z}'}{D_\pi^{3/2}}} \tag{2.2.15}$$

where $n' = n - JU$ which accounts for the promotion of JU neutrons with principal quantum number p_ν to a higher level and $\bar{n}' = D_\nu - n'$, and $z' = z - JU$ which accounts for the promotion of JU protons with principal quantum number p_π to a higher level and $\bar{z}' = D_\pi - z'$. From *Nilsson diagrams*, Duflo & Zuker, find $JU_\nu = JU_\pi = JU = 4$.[3] Remaining is the pairing term which can be written

$$\boxed{E_{\text{pairing}} = \frac{\alpha_{10}}{R}\begin{cases} 2T/A & N \text{ odd, } Z \text{ odd} \\ 1 - 2T/A & N \text{ even } > Z \text{ odd} \\ 1 & N \text{ odd } > Z \text{ even} \\ 1 & N \text{ even } < Z \text{ odd} \\ 1 - 2T/A & N \text{ odd } < Z \text{ even} \\ 2 - 2T/A & N \text{ even, } Z \text{ even} \end{cases}} \tag{2.2.16}$$

The DZ10 mass model is quite a lengthy formula, and for reference printed in sequence in its entirety in App. A and a Fortran code for calculating the binding energy in App. B

To be perfectly clear, the DZ10 model is actually not entirely linear, as can be seen from the last couple lines in the code, App. B. The DZ10 code calculates two binding energies, one assuming a spherical nuclei and one assuming a deformed nuclei, then chooses the max:

$$\text{BE}_{\text{DZ10}} = \max\left(\left\langle \text{H}_{\text{monopole}}\right\rangle + \left\langle \text{H}_{\text{spherical}}\right\rangle, \left\langle \text{H}_{\text{monopole}}\right\rangle + \left\langle \text{H}_{\text{deformed}}\right\rangle\right). \tag{2.2.17}$$

For proton number $Z \leq 50$ the DZ10 code, App. B, always uses the spherical Hamiltonian, which actually is a linear model with 9 parameters. We will neglect the non-linearity and construct a design matrix under the assumption the DZ10 mass model is linear.

3

Bayesian Model Averaging

In this chapter we will continue upon the theoretical base which we established in the previous chapter. We will use a toy model to describe the underlying theory in detail. Particular emphasis is put on the choice of priors.

3.1 A Toy Model Example

Thus far we have described some of the theory and methods to analyse data using Bayesian statistics. We have also, under the assumption of linear models, derived rather simple, analytical expressions for various posterior distributions. In this section we will give an example of parameter estimation, model selection and BMA, using synthetic data and polynomial toy models.

In polynomial regression we seek the parameters of a polynomial model,

$$\mathcal{M}: \quad \sum_{\kappa=0}^{K-1} \alpha_\kappa x^\kappa = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \cdots + \alpha_{K-1} x^{K-1} \tag{3.1.1}$$

which best describes the data. This polynomial of degree $K-1$ is a linear model, see Eq. (2.1.10), with K parameters. $\{\alpha_\kappa\}_{\kappa=0}^{K-1}$. From this model we can construct a set of models by choosing a subset of the basis functions, e.g. a model only containing the constant term or a model only containing the quadratic term. We will construct a set of all possible combinations of basis functions. The dimension of this model space is $2^K - 1$. Then given a set of data, in the form of a vector, $\boldsymbol{d}$, we can follow the steps of Sec. 2.1.3 to calculate posterior model and parameter probabilities. This allows us to discuss the credibility of the basis functions of the global model.

Polynomial fitting is an ordinary least-squares problem, which there exist numerous tools for, e.g. NumPy's *polyfit*. These will trust the user to know which degree of polynomial to be used; analysis can be made by examining the residuals or rank of the coefficient matrix. The issue of overfitting appears, we can use an overly complex polynomial and perfectly describe data, but having no proper way of describing the credibility of the model.

This issue can be resolved in a Bayesian approach to parameter estimation and calculating model posterior probabilities. We can quantify more rigorously the certainty in models, and answer which one is most plausible and not just which one fits the data best. Even

though higher order models may have a better fit, the added complexity makes them less plausible. A numerical example to illustrate the rejection of more complex models, in Eq. (2.1.25): The addition of one extra parameter with range $[-500, 500]$ implies a multiplicative factor of $\sim 1 \times 10^{-3}$. Thus the goodness of fit must be at least a factor $\sim 1 \times 10^{3}$ better to be comparable to the more simple model with one parameter less. Bayesian parameter estimation and model averaging can therefore be more conservative than standard least-squares solver.

Let us start with fitting the 5 parameters of a fourth order polynomial to synthetic data generated by

$$y = x + \frac{1}{2}\sin(4\pi x). \tag{3.1.2}$$

Gaussian noise is then added to simulate measurement error. Furthermore we adopt a model discrepancy, as per Eq. (2.1.51), $\delta = 0.5$. We do this, since when we analyse mass models, Ch. 4, we will also study the robustness of BMA when varying σ in Eq. (2.1.52) Now let us consider two (rather extreme) cases: Imagine we have only a few, 5 data points from equidistant $x \in [-0.25, 0.5]$, and low precision, the added noise is $e_i \sim \mathcal{N}(0, 0.01)$, i.e. an uncertainty of $0.1 \sim 10\,\%$ of $\max(y)$ over the range, see Fig. 3.1.1.

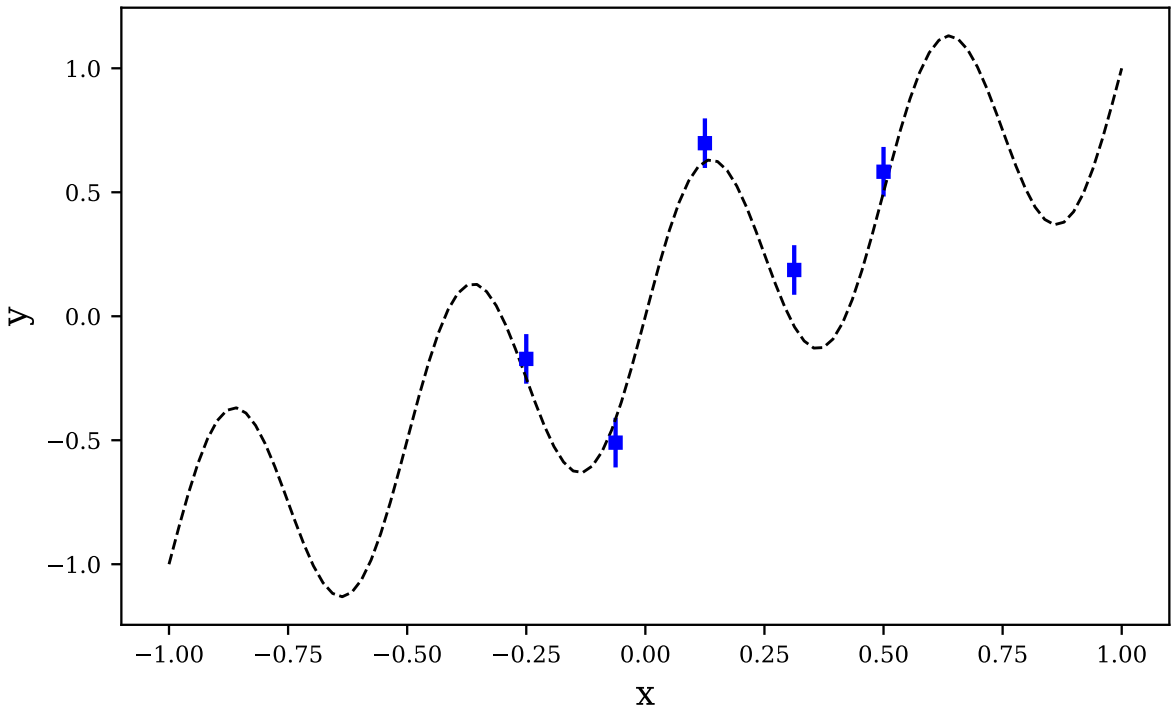


Figure 3.1.1: Toy model (dashed line) and simulated data points (blue squares) with error bars showing 1 standard deviation. 5 data points, $e \sim \mathcal{N}(0, 0.01)$

The other scenario is a large number of measurements, $n = 100$ data points using equidistant $x \in [-0.25, 0.75]$, and high precision, the added noise is $e_i \sim \mathcal{N}(0, 0.001)$, an uncertainty of $0.01 \sim 1\,\%$ of $\max(y)$ over the range, see Fig. 3.1.2. An important question

is which parameter values will make the model best fit the data. Then we can follow the procedure in Sec. 2.1 and Sec. 2.1.2, in evaluating the posterior model probabilities, Eq. (2.1.12) and (2.1.13) and (2.1.14). In order to do so we need to assign a parameter prior, this must also be done before looking at the data.

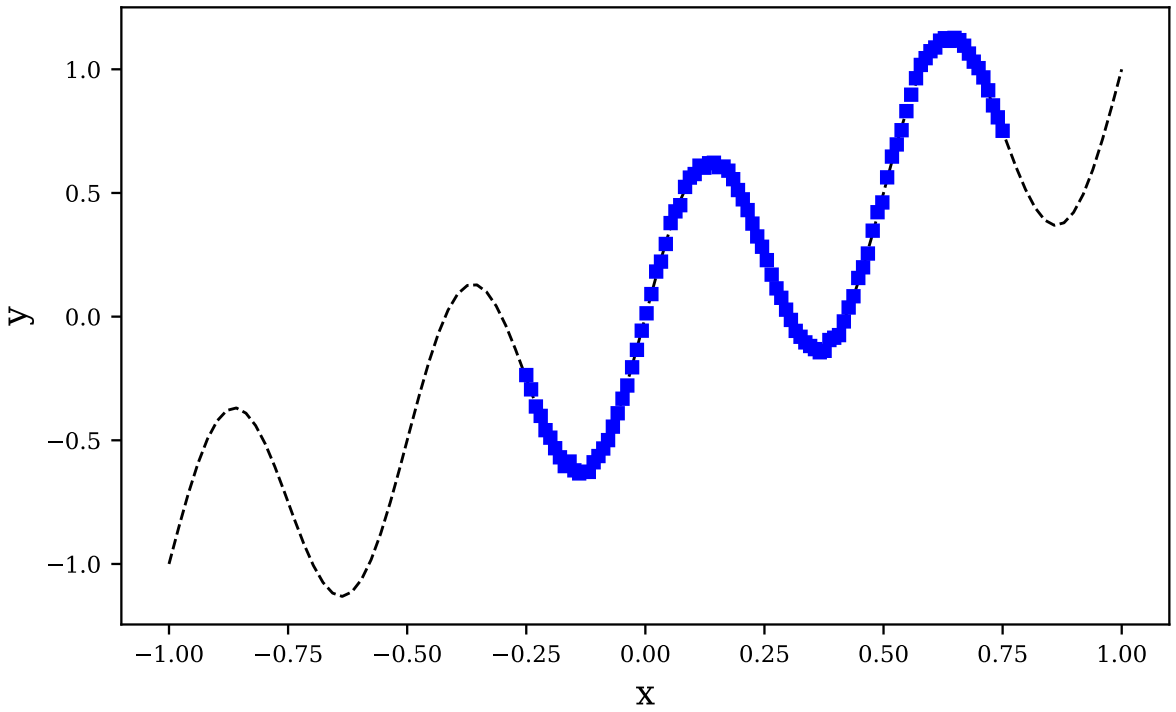


Figure 3.1.2: Toy model (dashed line) and simulated data points (blue squares) with error bars showing 1 standard deviation. 100 data points, $e \sim \mathcal{N}(0, 0.001)$

3.1.1 Choice of Parameter Prior

Depending on the data set the choice of parameter prior can have large effect. This we can illustrate by comparing the results using different priors, i.e. how well the posterior parameter distributions describe the data. Let us use an ignorant uniform prior, an ignorant Gaussian prior, and a Zellner g-prior with $g = n$:

$$\begin{cases} \pi_U(\boldsymbol{\alpha}) = \mathbf{U}[-10, 10] \\ \pi_{N,\ I}(\boldsymbol{\alpha}) = \mathcal{N}(\boldsymbol{\alpha} \,|\, \mathbf{0}, 10 \cdot \mathbf{1}) \\ \pi_{N,\ Z}(\boldsymbol{\alpha}) = \mathcal{N}\left(\boldsymbol{\alpha} \,|\, \mathbf{0}, g\left[\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}\right]^{-1}\right) \end{cases} \tag{3.1.3}$$

Focusing on a 5 parameter polynomial model, we can then use Eq. (2.1.22) and Eq. (2.1.47) to extract posterior distributions for our parameters. That is, fitting $f(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3 + \alpha_4 x^4$ to the data. Then we can compare this to the data and see how our analysis performs. In Figs. 3.1.3, 3.1.4 and 3.1.5 we show the model that fits best with the data, i.e. the one with highest posterior model probability.

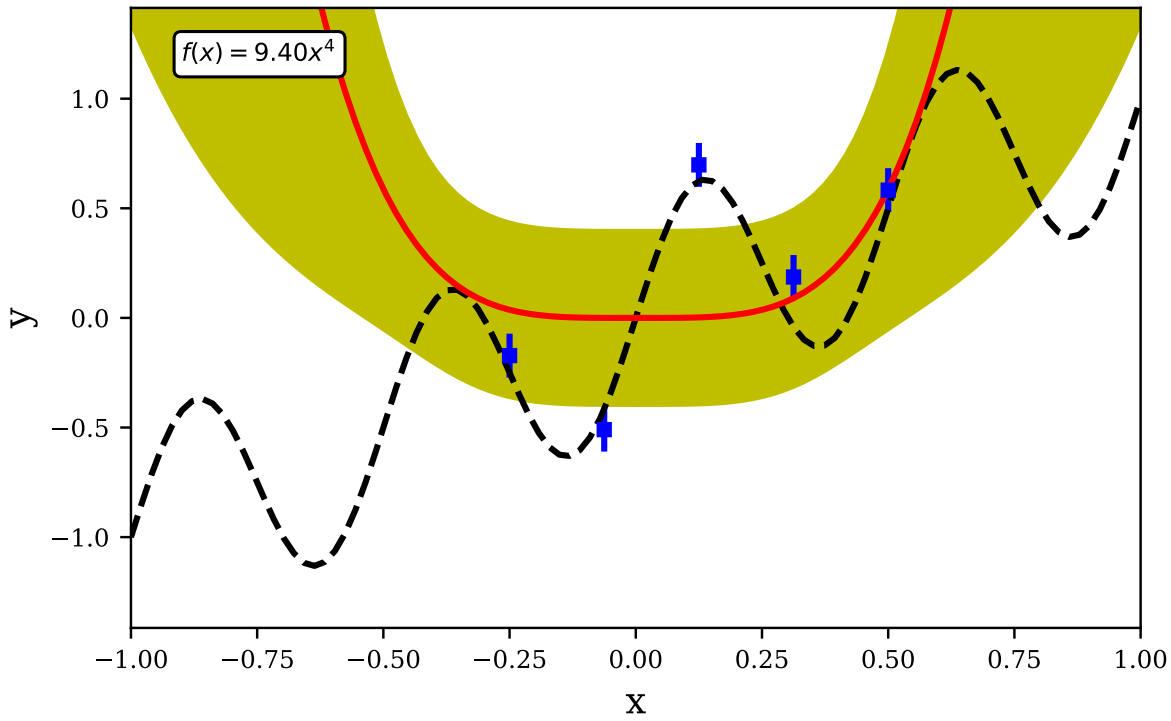


Figure 3.1.3: Fit using a uniform prior, shown with the expected value (red line) and a shaded region (yellow band) of 1 standard deviation depicting model uncertainty. The 'true model' shown dashed. As expected from a polynomial model it interpolates quite well but extrapolates very poorly.

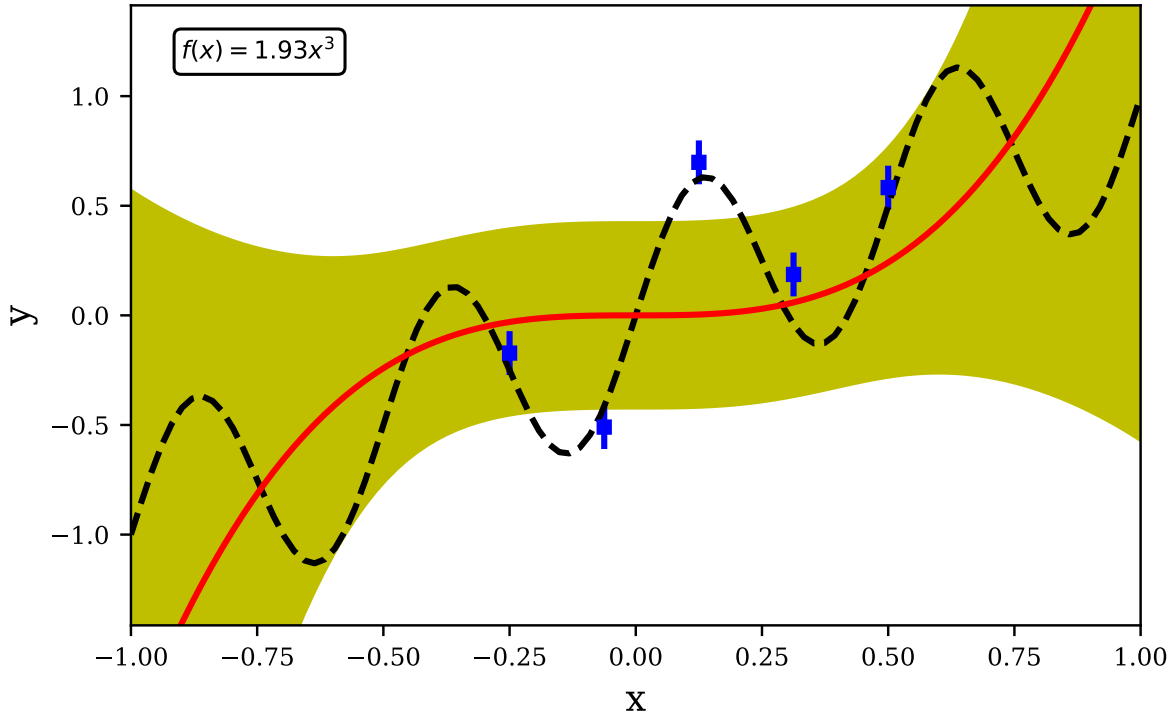


Figure 3.1.4: Fit using a Gaussian prior, shown with the expected value (red line) and a shaded region (yellow band) of 1 standard deviations depicting model uncertainty. The 'true model' shown dashed. As expected from a polynomial model it interpolates quite well but extrapolates very poorly.

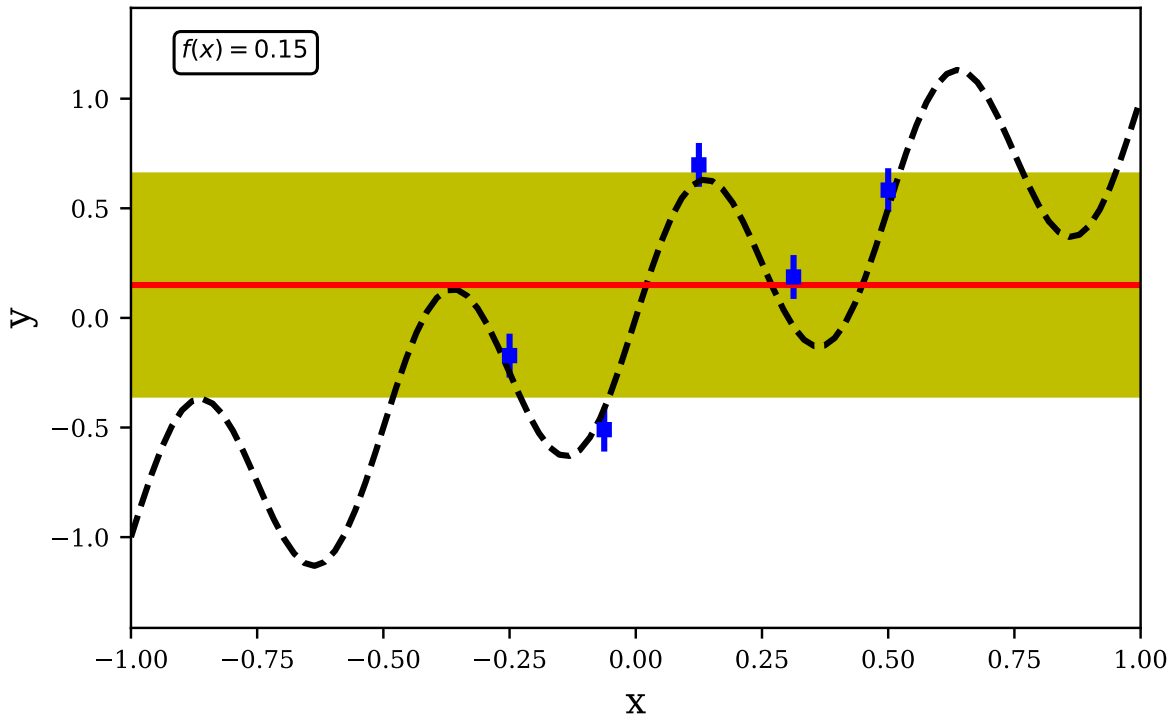


Figure 3.1.5: Fit using Zellner's g-prior, shown with the expected value (red line) and a shaded region (yellow band) of 1 standard deviations depicting model uncertainty. The 'true model' shown dashed. As expected from a polynomial model it interpolates quite well but extrapolates very poorly.

In reality we will not know the 'true model', we will not even know how far we actually are from it, and we have limited means of checking if our model actually represents reality. The only thing we can do is evaluate our model to data, and try to predict values prior to performing experiments and then compare the results. In the examples above we can see that for a small sample, different choices of parameter prior can alter the posterior drastically.

3.1.2 Different Models

We often have more than one model. For a polynomial model we can create a 'family' of models which correspond to all the combinations of the basis functions. Our numbering convention of models is based on the binary representation acting as a Boolean array. So for our toy model $f(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3 + \alpha_4 x^4$ with 5 parameters there exists $2^5 - 1 = 31$ different combinations thus allowing a 5 digit binary number to represent each model. For example:

$$11_{10} = 01011_2 \rightarrow \cancel{\alpha_0}, \alpha_1, \cancel{\alpha_2}, \alpha_3, \alpha_4 = \alpha_1, \alpha_3, \alpha_4 \tag{3.1.4}$$

would be model 11. A very important property of Bayesian statistics is that we can quantify our prior and posterior beliefs in the model we use. By Bayes theorem, Eq. (2.1.12), we can calculate a posterior distribution for the models, Fig. 3.1.6, 3.1.7 and 3.1.8, here

assuming a uniform model prior.

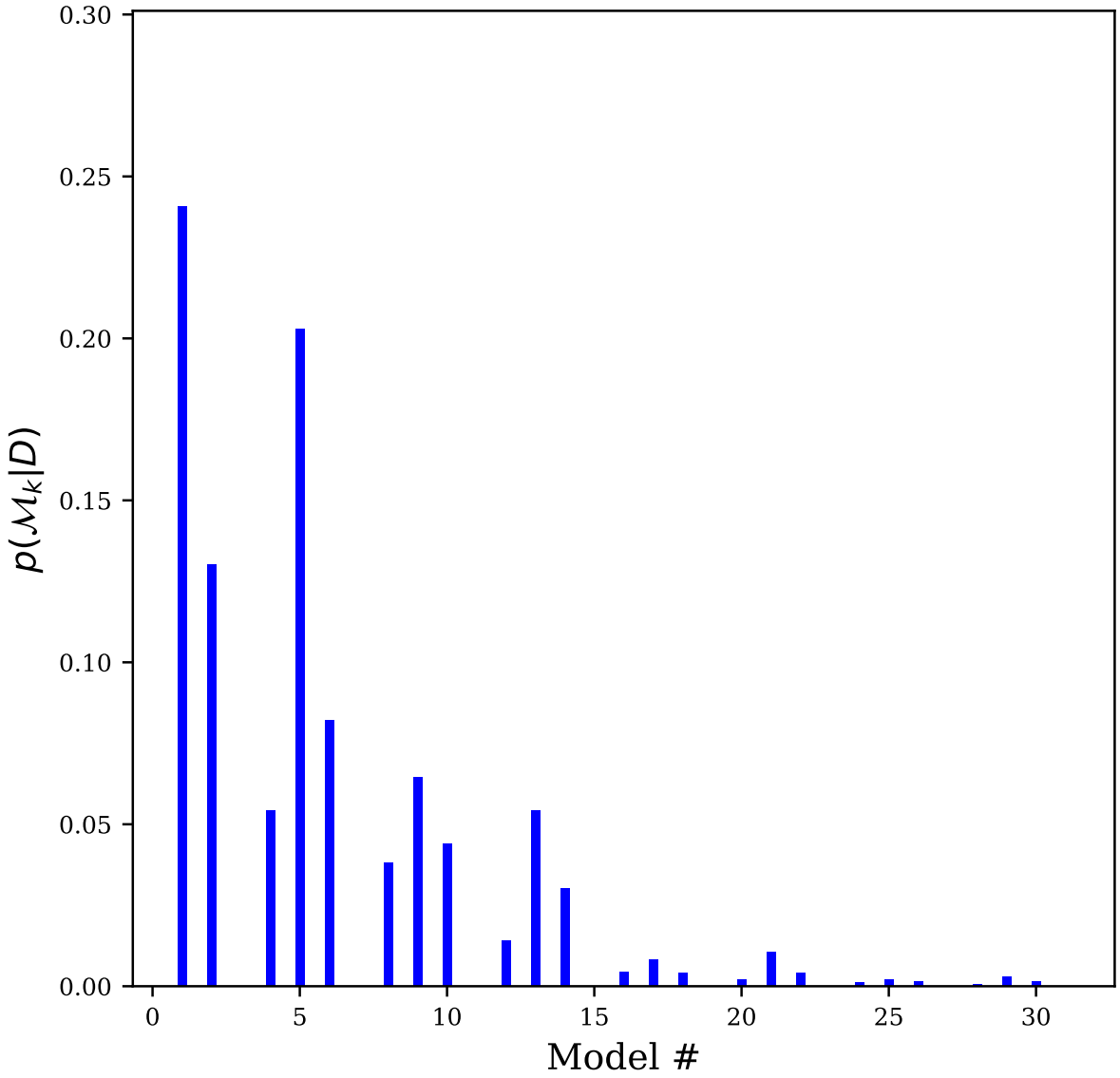


Figure 3.1.6: Distribution of posterior model probability using a uniform parameter prior, and a uniform model prior.

As we can discern in Fig. 3.1.6, 3.1.7 and 3.1.8 there can be several competing models that in a sense equally well describe the data without being too complex. But here we have neglected the correlation of basis functions, which for a polynomial is severe. Here a model dilution prior can be employed to mitigate using highly correlated basis functions, see Sec. 2.1.4.

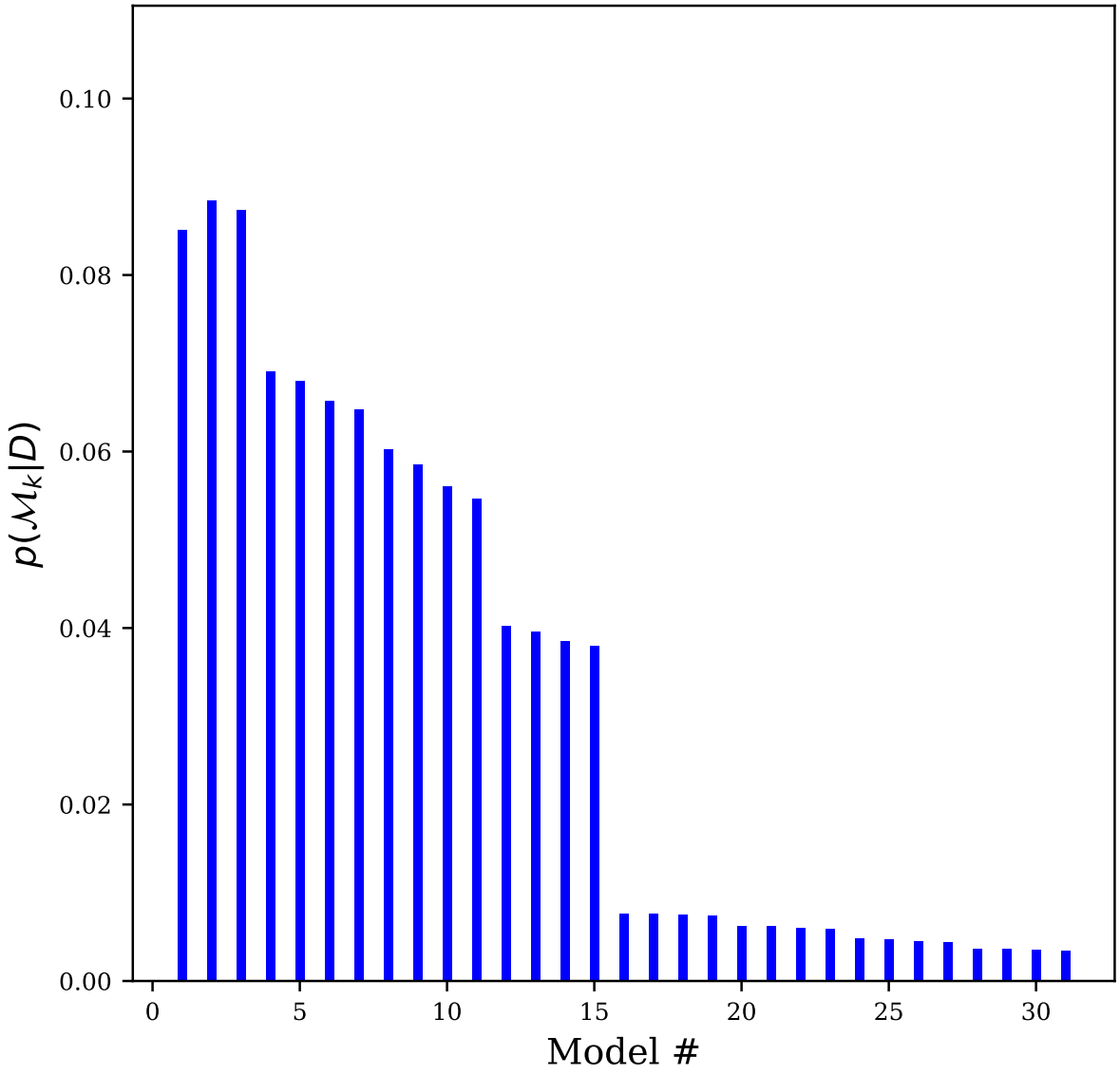


Figure 3.1.7: Distribution of posterior model probability using a Gaussian parameter prior, and a uniform model prior.

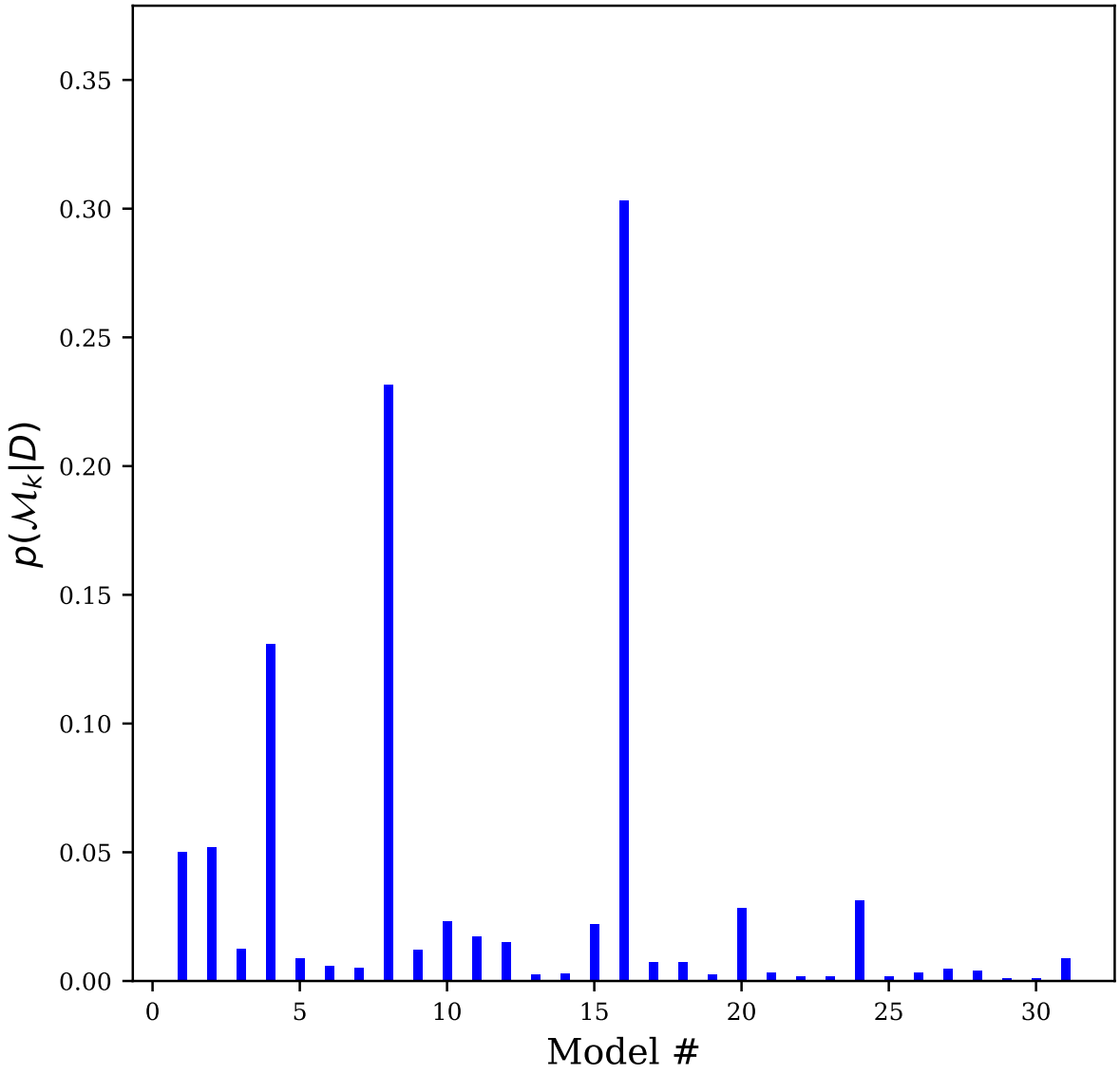


Figure 3.1.8: Distribution of posterior model probability using Zellner's g-prior for the parameters, and a uniform model prior.

3.1.3 BMA

From above we know that several models can describe the data. Which model should we select? A beautiful solution appears: model averaging, Sec. 2.1.4. We can combine the predictions from all the models at our disposal and weight them according to how credible the models are. See Fig. 3.1.9, 3.1.10, and 3.1.11.

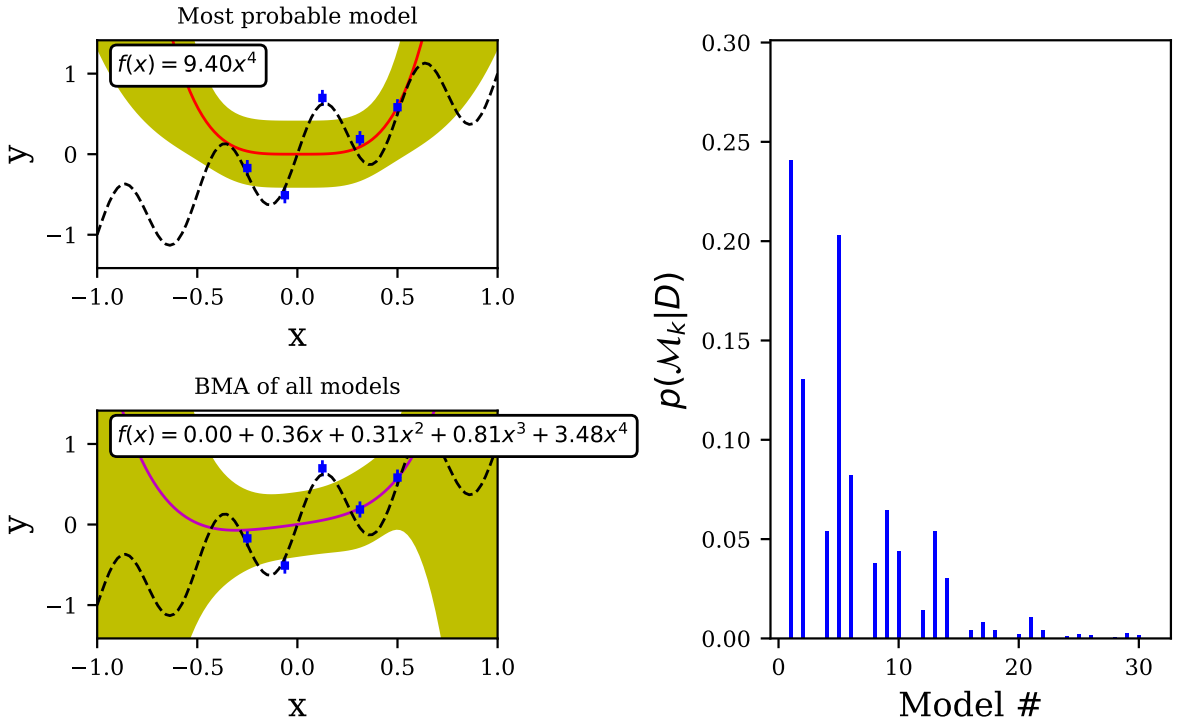


Figure 3.1.9: Fit of most probable model in top left, and BMA fit bottom left, using the posterior model probability shown on the right. All using a uniform parameter prior.

Since the Occam factor is different for different parameter priors the posterior distributions for the models are different. An important observation is the increase of error band seen in all Figs. 3.1.9, 3.1.10, and 3.1.11 when using BMA; most notably in 3.1.11. This extreme error band is due to having competing models, which describe the data at hand equally well, but extrapolate in vastly different directions.

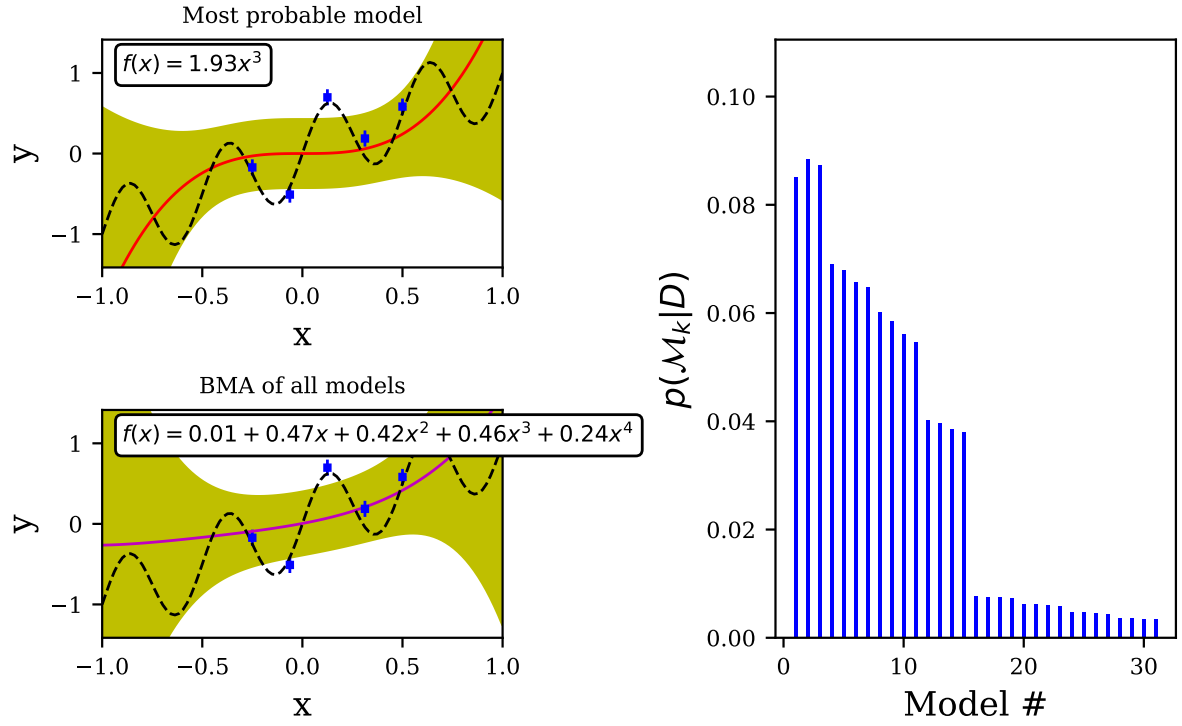


Figure 3.1.10: Fit of most probable model in top left, and BMA fit bottom left, using the posterior model probability shown on the right. All using a Gaussian parameter prior.

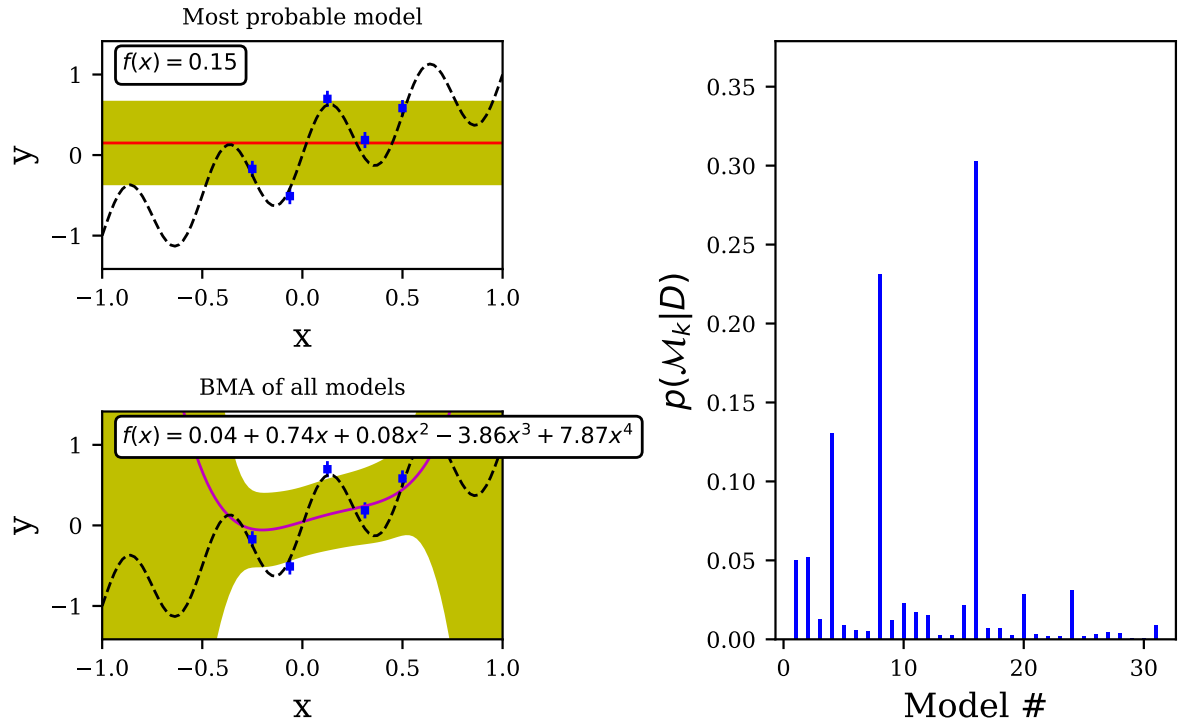


Figure 3.1.11: Fit of most probable model in top left, and BMA fit bottom left, using the posterior model probability shown on the right. All using a Zellner's g prior for the parameters.

3.1.4 Importance of Having Enough Data

How will the analysis be different if we have a larger data set? An increase in the number of data points will certainly make us more certain of any behaviour we observe. See Fig. 3.1.12, 3.1.13, and 3.1.14 for the analysis performed on a large data set with small data uncertainty.

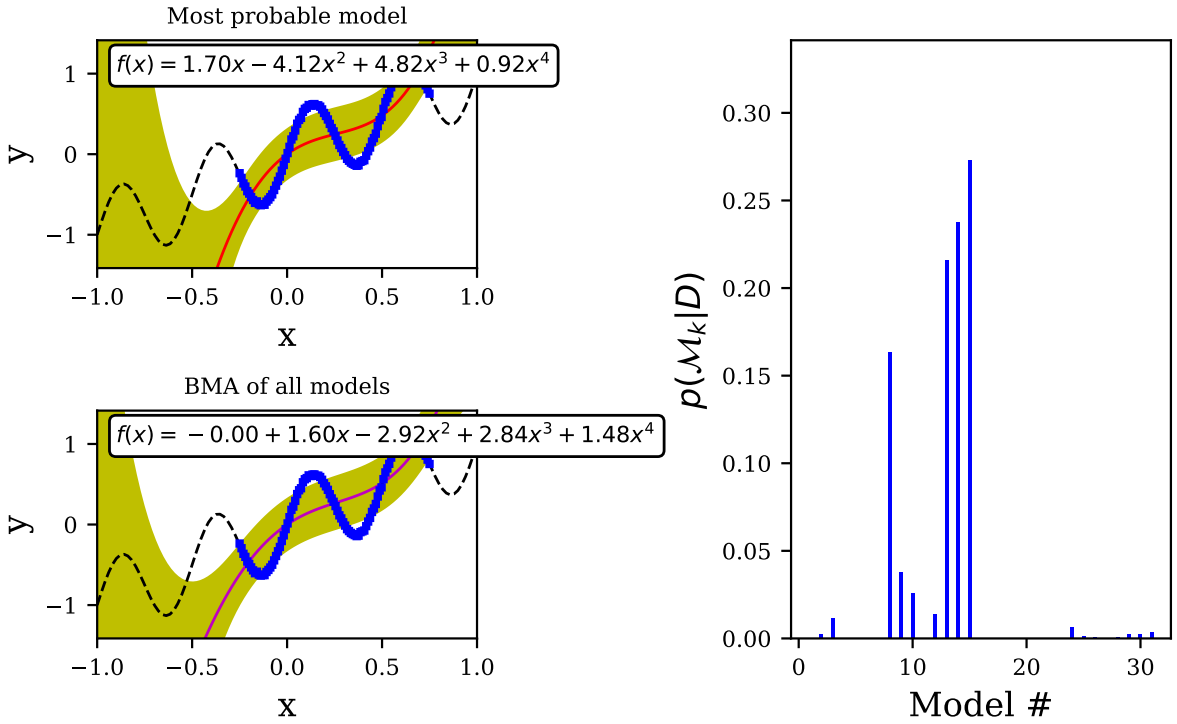


Figure 3.1.12: Fit of most probable model in top left, and BMA fit bottom left, using the posterior model probability shown on the right. All using a uniform parameter prior.

As expected from a polynomial model, of rather low degree, it has poor prediction. However the error bands of the most probable models seem to indicate we are quite certain. Using BMA we can instead place a more reliable uncertainty on our fit, especially for the extrapolation which is very poor. It still catches the behaviour of the data in general.

An important note here is that we use a model discrepancy, as Eq. (2.1.50) and Eq. (2.1.51). This means we expect an error and therefore accept an error. This explains why the fits in Fig. 3.1.12, 3.1.13 and 3.1.14 capture the linear dependency but neglects the $\sin(4\pi x)$. We know the models are insufficient, and rule out overfitted models.

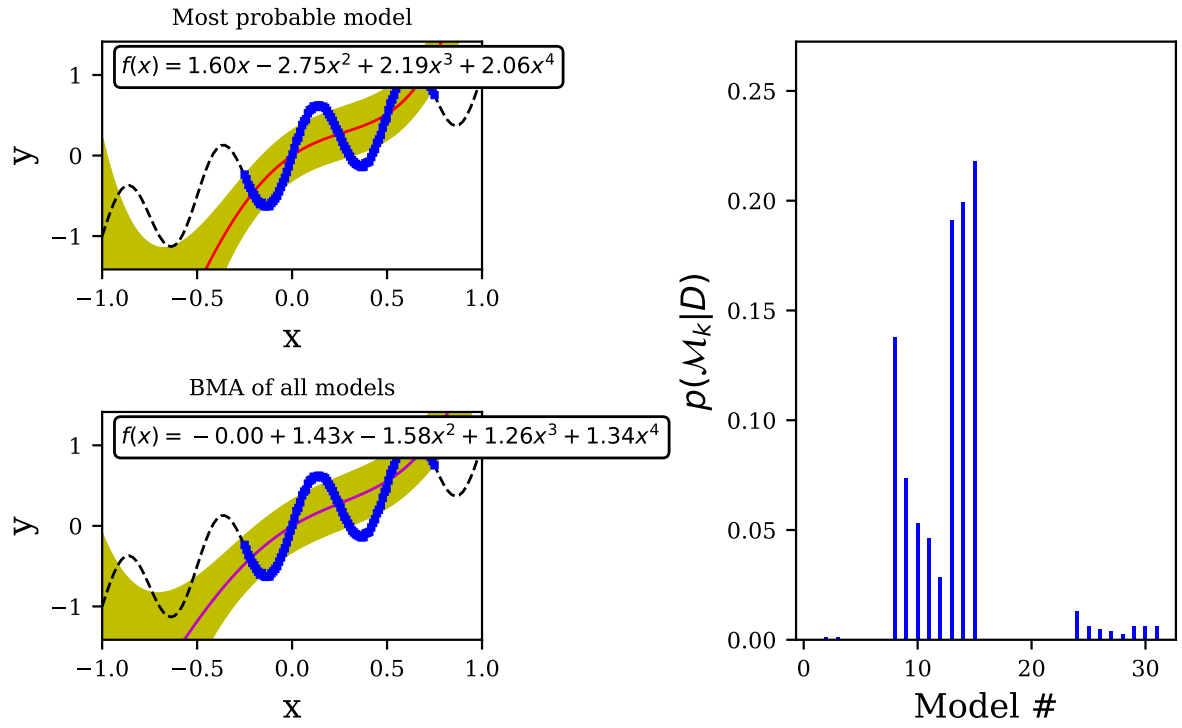


Figure 3.1.13: Fit of most probable model in top left, and BMA fit bottom left, using the posterior model probability shown on the right. All using a Gaussian parameter prior.

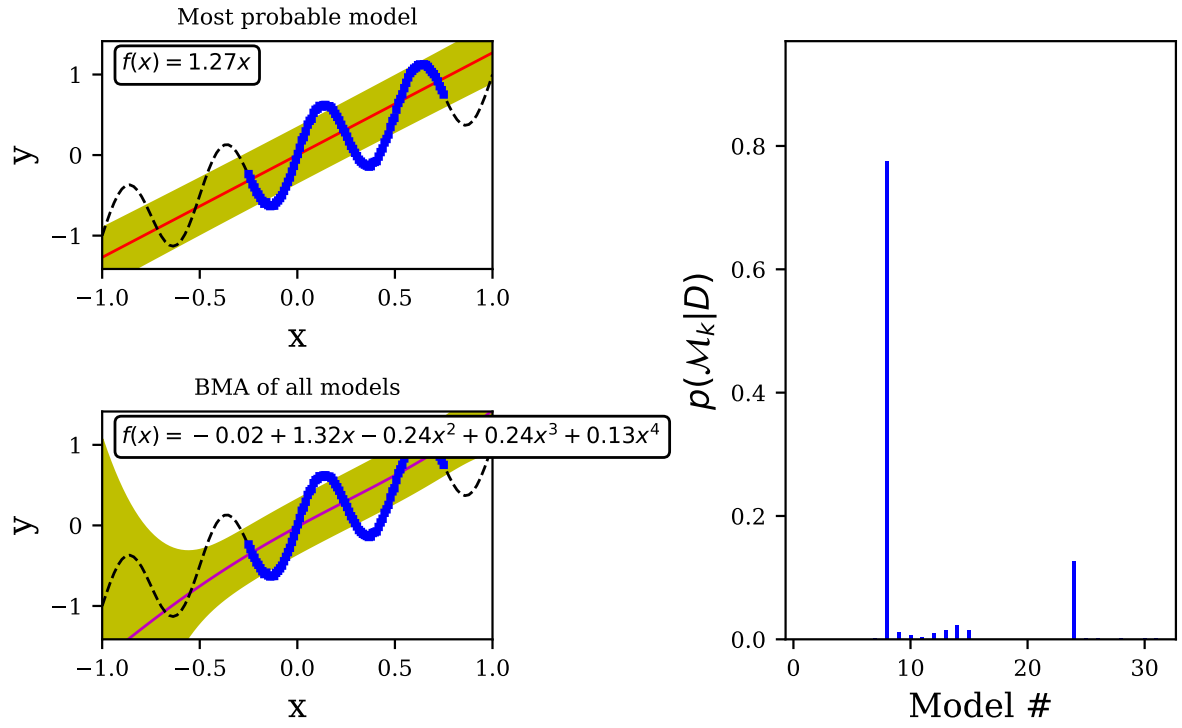


Figure 3.1.14: Fit of most probable model in top left, and BMA fit bottom left, using the posterior model probability shown on the right. All using a Zellner's g prior for the parameters.

4

Bayesian Analysis of the SEMF and DZ10 Models

We have calibrated the parameters of the mass models using AME 1983,[11, 12] and then evaluated the predictive capabilities of our models using the most recent mass data from 2016, i.e. AME 2016[13, 14]. We have also restricted ourselves to heavier nuclei, $A > 40$, when analysing the SEMF as it is a leptodermous expansion; approximating the nucleus as a liquid drop is less valid for light nuclei. The SEMF is an interesting model but not our focus so we only include it for comparison. We now turn our focus to the more complex DZ10 model, which we will assume to be a truly linear model. Choosing the parameter priors

- wide uniform distribution,
  $\pi_{\mathrm{U}}(\boldsymbol{\alpha}) : \ \alpha_i \in [0,\, 300]$ for $i = 1, \ldots 10$
- wide Gaussian distribution,
  $\pi_{\mathrm{N,\ ignorant}}(\boldsymbol{\alpha}) = \mathcal{N}\left(\mathbf{0},\, 400 \cdot \mathbf{1}\right)$
- Zellner's g-prior,
  $\pi_{\mathrm{N,\ Zellner}}(\boldsymbol{\alpha}) = \mathcal{N}\left(\mathbf{0},\, g\left[\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}\right]^{-1}\right), \quad g = \#\text{data points}$

we can investigate the robustness of parameter estimation, and posterior distributions. The number $\sigma^2 = (20\,\mathrm{MeV})^2$ is to limit the reasonable parameter values to within 60 MeV with 99.7 % degree of belief. Recall that for a large enough data set, the choice of parameter prior distribution should not play a considerable role, see Sec. 3.1. As the AME83($A > 40$) contains mass data for 1399 isotopes with typically small experimental uncertainty, of the order 10 keV, it is expected the parameter estimation will be rather robust. For comparison the AME16($A > 40$) contains mass data for 2241 isotopes with slightly higher experimental precision. There is of course a possibility of adopting a physically inspired parameter prior but we first want to compare simpler priors.

In Sec. 2.1 we discussed the process of performing parameter estimation and calculating posterior distributions for parameters, and even models. So we are fit for the task of performing analysis of especially the DZ10 mass model using different priors. We denote the binding energy measurements from AME83($A > 40$) with $\boldsymbol{d}$. We also combine the measurement uncertainty and model discrepancy δ into a diagonal covariance matrix $\boldsymbol{E}$.

Varying the factor δ allows us to study the behaviour and robustness of the evidence calculations. We have investigated values

$$\delta = 0,\, 1,\, 2,\, 3,\, 4,\, 5\,\text{MeV} \tag{4.0.1}$$

For the uniform parameter prior, $\pi_{\text{U}}(\boldsymbol{\alpha}) : \ \alpha_i \in [0,\, 300]$ for $i = 1, \ldots 10$, we have Eq. (2.1.22) and (2.1.25), yielding the parameter posterior and model posterior distributions, respectively. The wide Gaussian parameter prior, $\pi_{\text{N, wide}}(\boldsymbol{\alpha}) = \mathcal{N}(\mathbf{0},\, 400 \cdot \mathbf{1})$, and Zellner g-prior, $\pi_{\text{N, Zellner}}(\boldsymbol{\alpha}) = \mathcal{N}\left(\mathbf{0},\, g\left[\boldsymbol{\Phi}^{\text{T}}\boldsymbol{\Phi}\right]^{-1}\right)$, also have closed form solutions for all necessary posteriors, given by Eq. (2.1.47).

When we have calculated the posterior distributions for the parameters we can start making predictions. We begin our study by first restricting ourselves to the most plausible model, i.e. we have done model selection. Later we will choose all models and use BMA to make predictions. We compare these predictions of nuclear binding energies with data from AME16.

The primary goal is to study robustness of Bayesian analysis applied to nuclear mass models and the influence of BMA.

To *a posteriori* estimate the model discrepancy we use the root-mean-square deviation (RMSD) with respect to the training data

$$\text{RMSD} \equiv \sqrt{\frac{(\boldsymbol{d} - \boldsymbol{\Phi}\boldsymbol{\alpha})^2}{\#\text{data points}}}\,. \tag{4.0.2}$$

This is evaluated for each model after performing parameter estimation. We then update our belief in the model discrepancy, during predictions of nuclear binding energies.

4.1 Posterior Probability Distributions

We found that the distribution of the model posterior probabilities is rather narrow for the DZ10 mass model, and almost all choices of parameter prior and model discrepancy favour a near full DZ10 model. The agreement between a uniform parameter prior and a Gaussian parameter prior is expected as they have, by design, similar and large ranges within which we assume the parameters to have their 'true', physical, values. This yields a Bayesian framework that knows very little about the physics and the posterior values will be close to those that maximises the likelihood.

Our numbering convention of models is based on the binary representation acting as a Boolean array. For a model with K parameters there exists $2^K - 1$ different combinations thus allowing a K digit binary number to represent each model. For example:

$$359_{10} = 0101100111_2 \rightarrow \alpha_2,\, \alpha_4,\, \alpha_5,\, \alpha_8,\, \alpha_9,\, \alpha_{10} \tag{4.1.1}$$

For a wide uniform parameter prior, model number 1015 is by far the most plausible for a model discrepancy in the range $\delta = 0,\ 1,\ 2,\ 3\,\mathrm{MeV}$. That means including the terms proportional to α_1, α_2, α_3, α_4, α_5, α_6, α_8, α_9, α_{10}, only omitting the, spherical, S_3 surface term. Then for the $\delta = 3\,\mathrm{MeV}$ the 1011 model attains a non-negligible probability. For $\delta = 4,\ 5\,\mathrm{MeV}$ this model dominates. It includes the terms proportional to α_1, α_2, α_3, α_4, α_5, α_6, α_9, α_{10}, now omitting also the spherical QQ term.

The wide Gaussian parameter prior favours, without competition, model 1021 for a model discrepancy of $\delta = 0,\ 1,\ 2,\ 3,\ 4\,\mathrm{MeV}$ and is a close runner up even for 5 MeV. This corresponds to α_1, α_2, α_3, α_4, α_5, α_6, α_7, α_8, α_{10}, omitting the QQ deformed term of the Hamiltonian. For $\delta = 5\,\mathrm{MeV}$ the 1005 is slightly favoured over the 1021 model, both with about 40 % posterior probability and a third model 1007 at around 10 %. Model 1005 corresponds to α_1, α_2, α_3, α_4, α_5, α_7, α_8, α_{10}, i.e. omitting the QQ deformed term as for lower model discrepancy but now also the spherical S_3 volume term from the Hamiltonian. Model 1007 corresponds to α_1, α_2, α_3, α_4, α_5, α_7, α_8, α_9, α_{10}, i.e. allowing the QQ deformed term again and only omitting the spherical S_3 volume term.

Onto the g-prior, which for a model discrepancy $\delta = 0\,\mathrm{MeV}$ model 1021 dominates. Same as for the wide Gaussian parameter prior, excluding the QQ deformed term from the Hamiltonian. But increasing to $\delta = 1\,\mathrm{MeV}$ model 895 dominates. Which corresponds to α_1, α_2, α_4, α_5, α_6, α_7, α_8, α_9, α_{10}, i.e. including the QQ deformed term and now omitting the master FM surface term of the Hamiltonian. Then for $\delta = 2,\ 3\,\mathrm{MeV}$ model 894 is dominating. This corresponds to α_1, α_2, α_4, α_5, α_6, α_7, α_8, α_9, now besides the FM surface terms also omitting the pairing term. For $\delta = 4\,\mathrm{MeV}$ model 886 dominates. Corresponding to α_1, α_2, α_4, α_5, α_6, α_8, α_9, i.e. omitting the FM surface, the pairing terms and also the spherical S_3 surface term is excluded. For a discrepancy of $\delta = 5\,\mathrm{MeV}$ model 870 dominates. Which corresponds to α_1, α_2, α_4, α_5, α_8, α_9, now omitting the FM surface, S_3 surface and pairing terms also the S_3 volume term is rejected.

This behaviour of eliminating parameters when allowing a larger model discrepancy is anticipated, and to some extent wanted, since it displays the uncertainty in those parameters of the DZ10 mass model. An interesting result is that the full DZ10 model is never close to being favoured. This implies a redundancy, or overfitting in the DZ10 model. Overall the different parameter priors and prior model discrepancies always keep the α_1, α_2, α_4, α_5, α_8 parameters, i.e. the Coulomb term, master FM and SO term, isospin/symmetry volume and Wigner term, isospin/symmetry surface term, QQ spherical term. The Coulomb, the FM part and isospin/symmetry terms exist also in the SEMF model; the SO part and isospin/symmetry term are correctional terms and the QQ spherical term has no counterpart in the SEMF.

4.2 Predicting Binding Energies

Using the posterior model distributions to either choose a model or average over them all, when more than one model has non-negligible model posterior, we can with their respective parameter posterior distributions predict nuclear binding energies. Then ap-

proximating the model discrepancy using the RMSD, Eq. (4.0.2) with respect to training data, the posterior predictive distribution of nuclear binding energies can be evaluated, using Eq. (2.1.54) and Eq. (2.1.47). In Fig. 4.2.1 the prediction of nuclear binding energies is compared to the stable nuclei in AME16, by plotting the deviation

$$\Delta\text{BE} = \text{experimental value} - \text{model prediction}. \tag{4.2.1}$$

The RMSD values with respect to AME83 and AME16 are also shown along with the estimation of parameters used. We can note that the value of posterior RMSD with respect to training data, 3071 keV, is close to the assigned prior model discrepancy, 3000 keV. We can note the apparent overestimation of binding energy, but only speculate the reasons. It may very well come from the model being fitted to AME83 and compared to AME16. But most likely it is indicative of the inability of the model to describe nuclear masses far away from magic numbers.

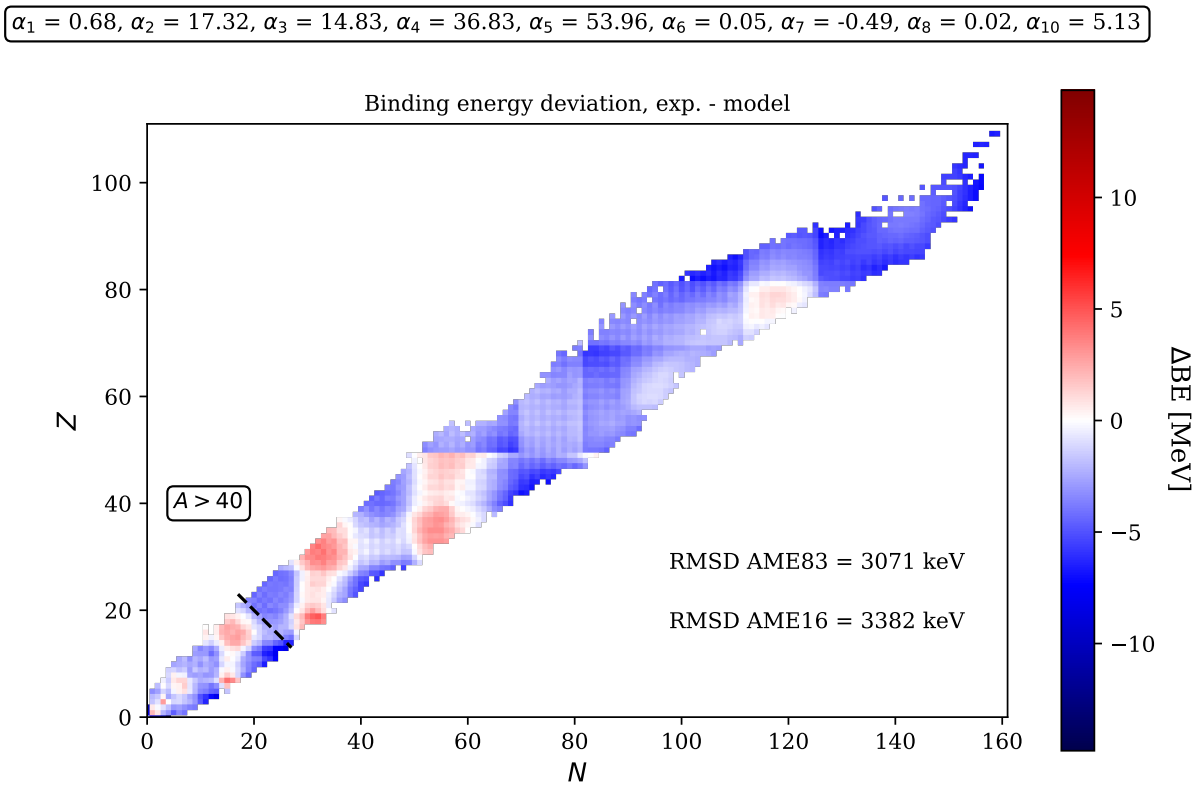


Figure 4.2.1: Comparison of AME16 and predicted binding energy using DZ10, trained on AME83 heavy nuclei ($A > 40$) using an ignorant Gaussian parameter prior and a prior model discrepancy of 3 MeV, ΔBE = AME16 − model. The posterior discrepancy is estimated by RMSD, shown in figure when compared to training data AME83 and also AME16. The parameter values for the present model are shown above the plot. Note the scale of ΔBE.

For a uniform parameter prior the posterior value of RMSD remains within 2.2 MeV–3.5 MeV with a mean of 2.9 MeV. Using a wide Gaussian parameter prior we instead get

a posterior RMSD in the region of 2.3 MeV–3.8 MeV with a mean of 3.0 MeV. The g-prior have a much higher posterior RMSD value 3.8 MeV–7.2 MeV with a mean 6.3 MeV. All of these stand in contrast to the RMSD of magnitude 700 keV claimed in [3]. The major difference is that Duflo and Zuker created a Hamiltonian based on physics, and then extracted the parameters values from their models fit to data. Herein, adopting Bayesian approach, we instead rely on which model we believe in given our priors, rather than the purely best fitting model. The most probable models, along with their parameters and the posterior RMSD with respect to AME16 for different choices of parameter prior and assigned prior model discrepancy is presented in Tab. 4.2.1

For comparison the corresponding scenario is shown for the SEMF in Fig. 4.2.2, but note the different scale of ΔBE. The pattern of failure to describe the underlying shell structure of nuclei is clearly visible in the regions of underestimated binding energy. The DZ10 model incorporates based on a harmonic oscillator some shell structure. A notable result is that all parameters of the SEMF are retained in the BMA, in contrast to the more complex DZ10 model.

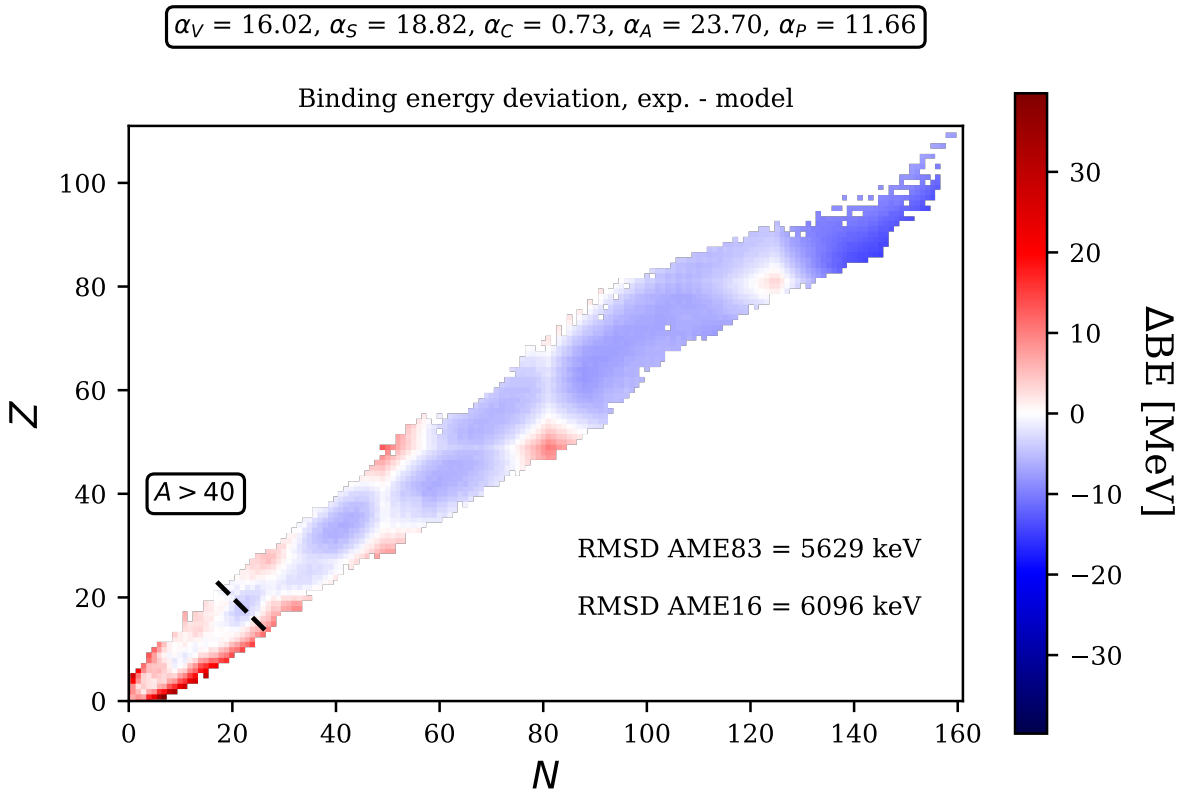


Figure 4.2.2: Comparison of AME16 and predicted binding energy using SEMF, trained on AME83 heavy nuclei ($A > 40$) using a wide Gaussian parameter prior and a model discrepancy of 3 MeV, ΔBE = AME16 − model. The posterior discrepancy is estimated by RMSD, shown in figure, compared to training data AME83 and also AME16. The parameter values for the present model are shown above the plot. Note the scale of ΔBE.

Table 4.2.1: Posterior RMSD with respect to prediction on AME16, and with respect to training data, AME83 ($A > 40$), and the presence of parameters, with different choices of parameter prior and when varying assigned prior model discrepancy.

Parameter prior	Prior δ [MeV]	RMSD 16 [keV]	RMSD 83($A > 40$) [keV]	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9	α_{10}
Uniform	0	4477	3694	●	●	●	●	●	●	○	●	●	●
	1	3635	3473	●	●	●	●	●	●	○	●	●	●
	2	3627	3458	●	●	●	●	●	●	○	●	●	●
	3	3622	3450	●	●	●	●	●	●	○	●	●	●
	4	2829	2226	●	●	●	●	●	●	○	○	●	●
	5	2829	2226	●	●	●	●	●	●	○	○	●	●
	Mean RMSD	3503	3088										
Gaussian	0	3051	2458	●	●	●	●	●	●	●	●	○	●
	1	3216	2871	●	●	●	●	●	●	●	●	○	●
	2	3391	3062	●	●	●	●	●	●	●	●	○	●
	3	3382	3071	●	●	●	●	●	●	●	●	○	●
	4	3358	3068	●	●	●	●	●	●	●	●	○	●
	5	2279	2356	●	●	●	●	●	○	●	●	○	●
	Mean RMSD	3113	2814										
Zellner g	0	3051	2458	●	●	●	●	●	●	●	●	○	●
	1	12 100	6351	●	●	○	●	●	●	●	●	●	●
	2	13 382	6806	●	●	○	●	●	●	●	●	●	○
	3	13 445	6586	●	●	○	●	●	●	●	●	●	○
	4	13 383	6637	●	●	○	●	●	●	○	●	●	○
	5	13 460	7190	●	●	○	●	●	○	○	●	●	○
	Mean RMSD	11 470	6004										

4.3 Binding Energies Across the Tin Isotopic Chain

We now focus on the tin isotopic chain, $_{50}$Sn which has a magic number of protons. In Fig. 4.3.1 the binding energy per nucleon is shown, together with the parameters for the model prediction and the associated RMSD. Data from AME16 is shown in blue with error bars corresponding to 3 standard deviations, though they are not visible at this scale. The top plot also shows the most probable model, in yellow, with 3 standard deviation error bars, and filled markers represent predictions for isotopes present in AME83 and hollow markers represent new isotopes only in AME16. The model does not seem to display any difference in predictive power for interpolation or extrapolation, though the extrapolation is not far. There is no noteworthy difference when applying BMA as can be seen in

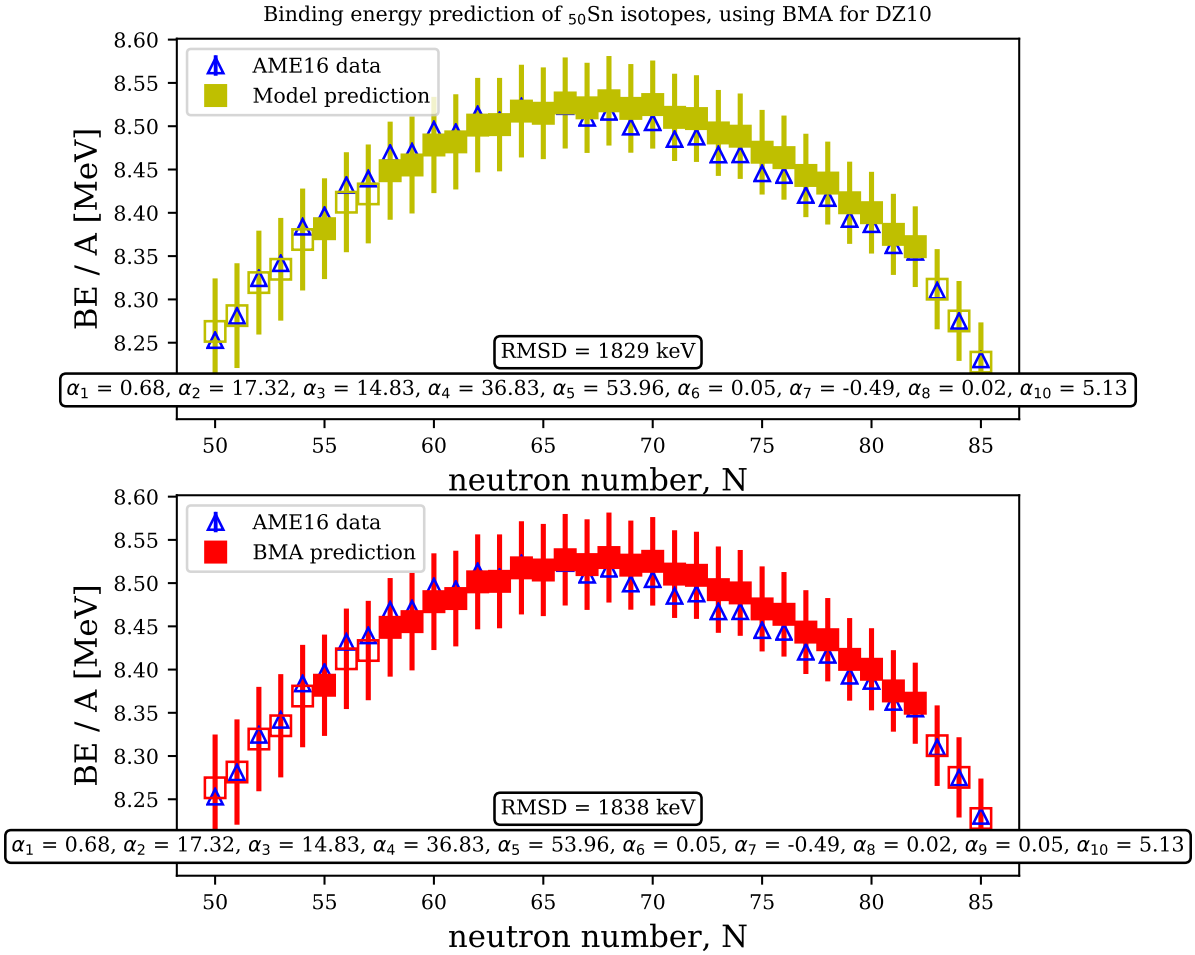


Figure 4.3.1: Comparison of AME16, blue hollow triangles, and predicted binding energy per nucleon for tin isotopes. Models trained on AME83 heavy nuclei ($A > 40$) using a wide Gaussian parameter prior and a model discrepancy of 3 MeV. Filled squares denote interpolated prediction and hollow squares extrapolation. Both data and prediction is shown with error bars of 3 standard deviations. Top panel: most probable model. Bottom panel: BMA.

Fig. 4.3.1 in bottom plot. This is due to the most probable model by far dominates over the other models.

An important observation is that even though we omit α_9, i.e. the QQ deformed term, we still describe the behaviour of the binding energy. But magic cores are not expected to be deformed, so it would have been a surprise to see anything else. However, this may very well mean reduced performance when extrapolating to nuclei far from stability.

4.4 The Neutron Drip Line

We have some trust in our parameter estimation although we have quite large uncertainties in the predictions. These uncertainties describe a credibility region for predicted nuclear binding energies. Calibrating on AME16 now, we can use the present day knowledge to search for the location of the neutron drip line. In AME16 there is listed a total of 36 tin isotopes, ranging from the doubly magic isotope with $N = 50$ neutrons to the most neutron rich isotope with $N = 85$ neutrons, which does not mark the neutron drip lines.

Negative neutron separation energies, i.e.

$$S_{1n} = \mathrm{BE}(Z,\, N) - \mathrm{BE}(Z,\, N-1) \tag{4.4.1}$$

$$S_{2n} = \mathrm{BE}(Z,\, N) - \mathrm{BE}(Z,\, N-2) \tag{4.4.2}$$

define the neutron drip line.

In Fig. 4.4.1 the prediction of the neutron drip line is shown using AME16 as training data and a Gaussian parameter prior together with a prior model discrepancy of 3 MeV. This is near the mean of posterior RMSD with respect to training data, Tab. 4.2.1, when calibrating using AME83. It is also close to the posterior RMSD with respect to prediction onto AME16, when calibrating using AME83. Using the RMSD with respect to calibration data can give a misleadingly small value, this is seen in Tab. 4.2.1 as the RMSD with respect to prediction of AME16 is almost always larger than RMSD with respect to AME83 ($A > 40$). The posterior RMSD with respect to calibration data AME16 is 2003 keV. Error bars corresponding to 3 standard deviations are shown, i.e. approximately 99.7 % credibility region for the most probable model in the top panel and for BMA in the bottom panel. The most plausible 1-neutron drip line seems to be located at $N = 123$, but within errors the drip line might enter as early as $N = 69$ and the very slight possibility of a small island of stability near $N = 221$. The 2-neutron drip line is predicted to be at $= 115$, but possibly earlier at $N = 81$ or later $N = 153$ is plausible. Limiting to a region of 1 standard deviation, i.e. a 68 % credibility interval, the region of the 1 neutron drip line is shrunk to $N = 95$–125 and the 2 neutron drip line to $N = 103$–125. We do know that the drip line is not as early as $N = 69$, this just shows the large uncertainty of these particular nuclear mass models. BMA has only a very small effect, in the analysis of the drip line, it does not shift the expected value but it does reduce the upper region of it just slightly. The 1 neutron drip line from $N = 221$ to $N = 219$ and the 2 neutron drip line from $N = 153$ to $N = 152$.

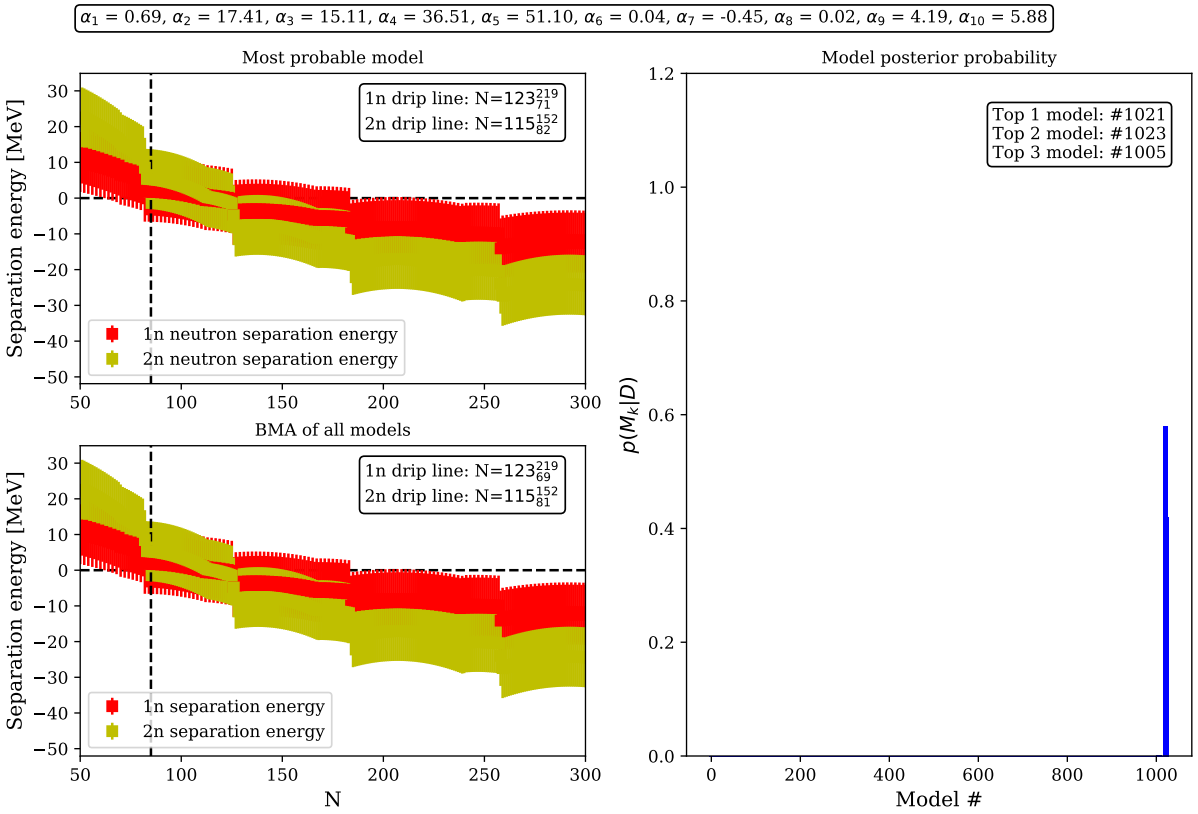


Figure 4.4.1: Left: Neutron drip line for tin ($Z = 50$), DZ10 trained on AME16 using a wide Gaussian parameter prior and assigned model discrepancy of $\delta = 3\,\text{MeV}$. The posterior RMSD with respect to AME16 is 2003 keV. Zero separation energy marked by the horizontal dashed line and the end of known stable nuclei listed in AME16, $^{135}_{50}$Sn marked by the vertical dashed line. Prediction plotted with 3 standard deviation error bars. Top left: most probable model. Bottom left: BMA. Right: Model posterior, showing two competing models, but the effects of BMA seems to be small. The parameters of the most plausible model are displayed above.

For comparison we have also predicted the neutron drip line using the SEMF when assigning a prior model discrepancy of 3 MeV, as shown in Fig. 4.4.2. The predicted separation energies clearly show how the SEMF does not describe any shell structure of the nucleus, as it describes a rather smooth decay. The fit is also quite uncertain which places a large region of plausibility on the neutron drip line, so we limit ourselves to 1 standard deviation, i.e. approximately a 68 % credibility region, to draw some conclusions. The 1 neutron drip region may be from $N = 81$ to $N = 239$ with an expected value of $N = 119$. The 2 drip region may be as low as $N = 90$ up to $N = 142$ with an expected value of $N = 110$. Since the full SEMF is the most plausible model, the effects of BMA are negligible.

We can note here that both the DZ10 model and the SEMF, using our Bayesian calibration of parameters, predict the neutron drip line to be in the vicinity of $N \approx 110$–120 and may stretch up to $N \approx 150$–240. While the lower limit may be around $N = 90$. Extremely

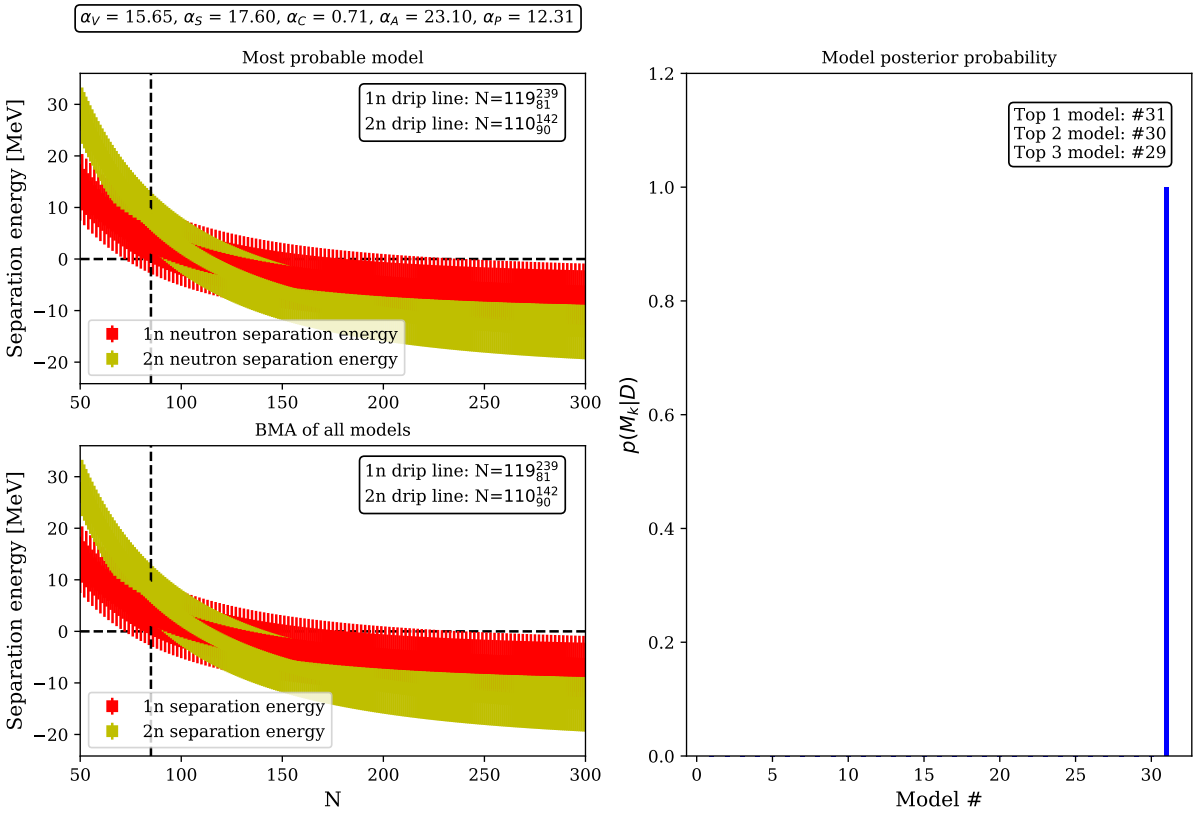


Figure 4.4.2: Left: Neutron drip line for tin ($Z = 50$), SEMF trained on AME83 using $\delta = 3\,\text{MeV}$. Zero separation energy marked by the horizontal dashed line and the end of known stable nuclei listed in AME16, $^{135}_{50}\text{Sn}$, marked by the vertical dashed line. Prediction plotted with 1 standard deviation error bars for inference. Top left: most probable model. Bottom left: BMA. Right: Model posterior, showing that one model dominates, thus rendering BMA almost redundant.

large errors, as expected, we assigned a prior model discrepancy of $\delta = 3\,\text{MeV}$ and the posterior RMSD = 6096 keV with respect to AME16.

4.5 Analysing DZ10 for Proton Number $Z \leq 50$

In the previous sections we have assumed the DZ10 model to be linear, although this is not the case. Recall Eq. (2.2.17), that it actually calculates two Hamiltonians and selects the one which yields the maximum binding energy. But, as seen in the code, App. B, Duflo & Zuker decided that isotopes with proton number $Z \leq 50$ are not associated with deformation. This allows us to correctly analyse a linear 'DZ9' mass model to assess the naive approach of assuming a truly linear DZ10 model. Comparing Fig. 4.3.1 with Fig. 4.5.1, we can see that the RMSD is slightly lower for the most probable model, top panel. But the more prominent feature is the effect of BMA, as there are more credible models available. Here BMA accomplishes a smaller RMSD while increasing the region of uncertainty. This DZ9 model seems to better interpolate for stable nuclei, while only

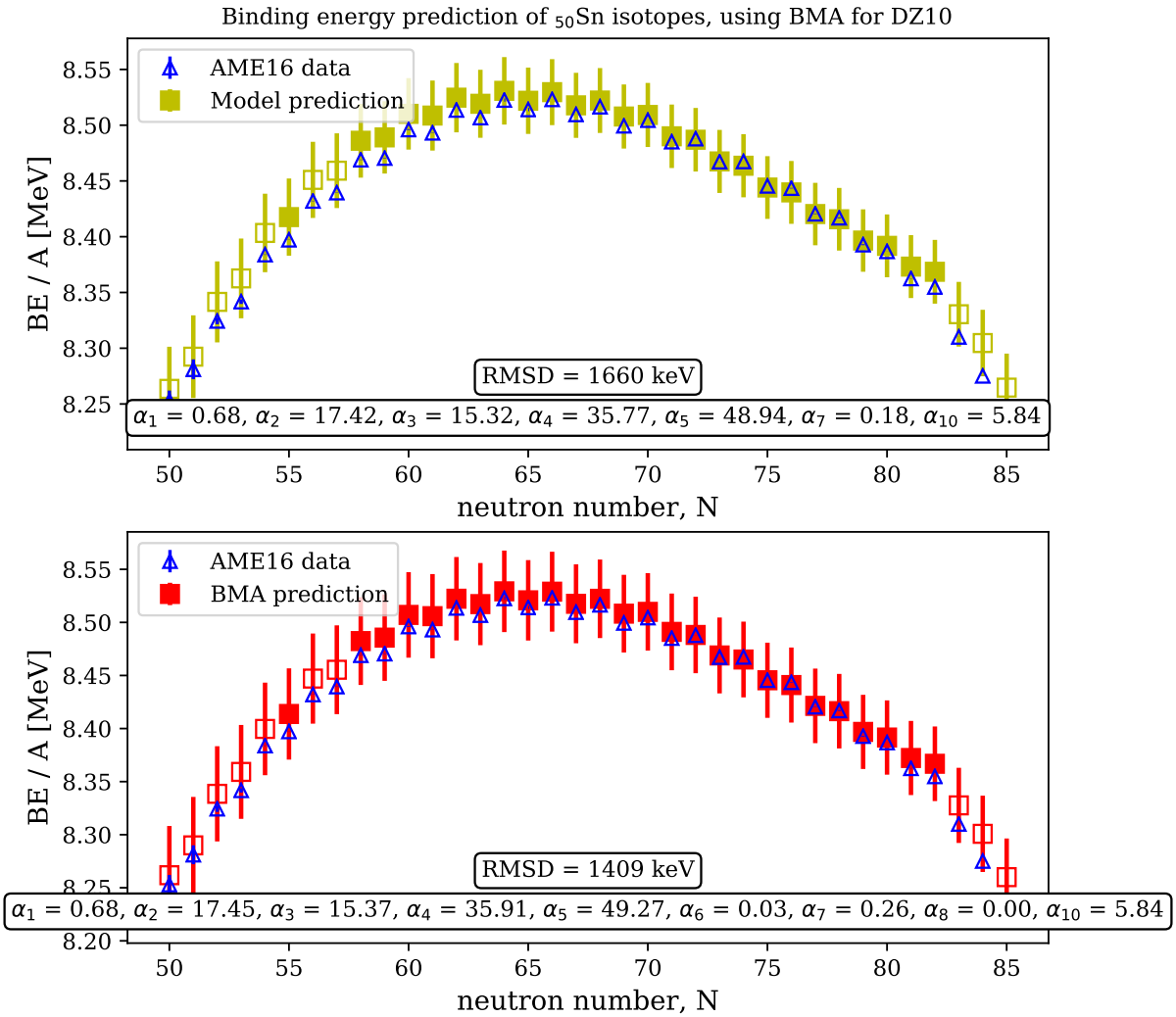


Figure 4.5.1: Comparison of AME16, blue hollow triangles, and predicted binding energy per nucleon for tin isotopes. Models trained on AME83 light nuclei ($Z \leq 50$) using a wide Gaussian parameter prior and a model discrepancy of 3 MeV. Filled squares denote interpolated prediction and hollow squares extrapolation. Both data and prediction is shown with error bars of 3 standard deviations. Top panel: most probable model. Bottom panel: BMA.

considering spherical properties.

Assuming the DZ10 model to be linear appears, of course not optimal, but adequate as an indicative tool; it captures aspects of both the spherical and the deformed Hamiltonian while this DZ9 model only captures the spherical part. Furthermore, using the Bayesian approach we have quantified the uncertainties of our models and predictions. If we were to investigate an isotopic chain with $Z > 50$ the effects are expected to be more noticeable.

5

Conclusions and Outlook

In this thesis we have studied the DZ10 mass model and the SEMF in a Bayesian framework for different choices of parameter prior, assuming both of them being linear models. We have also assigned and varied a prior model discrepancy in the models we investigate. When using wide uniform and Gaussian parameter priors, the posterior distribution for parameters and models does not vary much for different choices of prior model discrepancy. This is quite expected with the amount of data at hand and these choices are similar. The Zellner g-prior on the other hand varies much with the different choices of prior model discrepancy. It is designed to renormalise the parameter prior according to the basis functions. At the start of the project, and when analysing toy models, the g-prior seemed promising, not so different from the uniform or Gaussian priors. But the Zellner g-prior does exhibit a wanted property to an extent more than the uniform or Gaussian, in that it rejects parameters when increasing the prior model discrepancy. Dismissing parameters, as they add complexity, gives us information about the credibility of the parameters, but of course increasing posterior RMSD. The Zellner g-prior posterior RMSD is larger than the prior assigned model discrepancy. In contrast the wide uniform and Gaussian parameter priors have more or less stable posterior RMSD for all values of prior model discrepancy. A guess is that varying the model discrepancy in the range [0, 5] MeV is too much when using the g-prior, and a more narrow, informed, region is then more suited. The stability of the posterior RMSD near 3 MeV for the wide uniform and Gaussian parameter priors, when varying the *a priori* assigned model discrepancy indicates that the DZ10 has a naturally large discrepancy due to physical reasons, see Tab. 4.2.1.

In terms of posterior RMSD, with respect to calibration data as well as with respect to AME16, the wide uniform and Gaussian parameter priors outperform the Zellner g-prior, see Tab. 4.2.1. The Zellner g-prior is noticably poor at predicting compared to using Gaussian or uniform priors. But as the g-prior more fiercely omits terms we may learn more from it. In viewing the posterior model distributions it seems the belief in some terms is greater:

$$\begin{aligned} &\alpha_1,\ \alpha_2,\ \alpha_4,\ \alpha_5,\ \alpha_8 \\ &\text{Coulomb, FM and SO, symmetry volume,} \\ &\text{symmetry surface and Wigner, QQ spherical} \end{aligned} \tag{5.0.1}$$

This means a reduction from 10 parameters to 5 parameters of higher credibility, Eq. (5.0.1).

This, of course, is only a statement that their plausibility is greater than that of the terms omitted. In our analysis the following terms were deemed less credible:

$$\alpha_3,\ \alpha_6,\ \alpha_7,\ \alpha_9,\ \alpha_{10} \tag{5.0.2}$$

$$\text{FM surface, } S_3 \text{ volume, } S_3 \text{ surface, QQ deformed, pairing} \tag{5.0.3}$$

Our analysis has been restricted to using mass data of heavier nuclei, AME83 $A > 40$. Surface terms are merely corrections to the volume terms and their effects lessen with larger mass number, A. So it is reasonable surface terms are excluded in favour for volume terms. There is more mass data in AME16, and perhaps a larger number of deformed nuclei compared to AME83. This would imply the importance of the QQ deformed term may be underestimated. The exclusion of the pairing term is rather surprising but compared to the other terms its effects may be negligible, as it seems.

Several models based on the g-prior have, from a physical point of view, an unexpected posterior parameter distribution. For an example, the Coulomb term is lower than what we from standard arguments can estimate. Most likely it points to the fact that the model contains terms which equally well describe behaviour of nuclear binding energies.

Drawing a concrete conclusion regarding the neutron drip line for the tin isotopic chain is hard as the credibility region is wide. But it seems as if it is in the vicinity of the claimed magic numbers $N = 114$ and 126 with the faint possibility of having an island of stability around $N = 220$. The more reasonable $N \approx 115$–123, seems to agree with other sources.[6, 7]

As we have assumed the DZ10 model to be linear we have introduced an error, as it compares a Hamiltonian for spherical nuclei and a Hamiltonian for deformed nuclei. This could be a plausible explanation of the posterior RMSD near 3 MeV, as the two Hamiltonians differ to some extent, but not extremely. When restricting to investigating nuclei with proton number $Z \leq 50$ the DZ10 model reduces to a 9 parameter model, neglecting any deformation, as these are not associated to deform according to Duflo and Zuker, App. B. Most likely, assuming spherical nuclei, when predicting binding energies beyond stable isotopes and higher neutron excess, is sub-optimal but it serves as an indicative prediction. Furthermore the difference in predicted binding energy is not severe, comparing this DZ9 model using AME83 $Z \leq 50$ and the DZ10 model using AME83 $A > 40$. A slight decrease in RMSD, i.e. better interpolation, for the DZ9 model along the tin isotopic chain. Albeit the DZ10 model still captures some aspects of the deformation effects, thus it could possibly have better extrapolation properties. The effects of BMA on the DZ9 model is greater compared to the DZ10 model as there are more plausible models, omitting one parameter. Prediction of neutron drip line is not affected to great extent; the expected location is shifted by one or two neutrons for both the one-neutron drip line and the two-neutron drip line, well within uncertainties of using the DZ10 model. The major difference is that the importance of considering deformation is increasing with the number of neutrons as the proton number is constant, this is not taken into consideration in the DZ9 model. The prediction of separation energy using the DZ9 model does not decline as rapidly as the DZ10 model; similar to the SEMF it seems to stabilise around

some negative value. This pushes the upper bound of the location of the neutron drip lines as the error bar is wide. Yet from a physical perspective it is quite unreasonable that we would have stable isotopes for any number of neutrons.

Future Prospects

First it would be interesting to perform a Bayesian analysis of the DZ10 model for the entire chart of nuclides, not assuming linearity. However, this imposes numerical methods for the evidence calculations, and as a consequence more computational power and time.

One natural outlook is to investigate other priors. Investigating a physically inspired prior is possible, but demands an analysis of the mathematics and physics behind the formulas. A more thorough, in-depth, exploratory analysis would improve the results, in the sense that more informed priors yield more reliable posteriors. We have also assumed the amount of data to outweigh the prior dependency, in order to set a fixed value on the model discrepancy *a priori* and estimating it *a posteriori* using RMSD. This appears reasonable for our wide uniform and Gaussian priors but not the Zellner g-prior. In an extended analysis we can use a Gaussian-inverse-gamma prior which is the conjugate prior of a Gaussian likelihood with unknown mean and variance. An even more formidable prior is the Gaussian-inverse-Wishart distribution which is the conjugate to a multivariate Gaussian likelihood, allowing a unknown covariance matrix. Also, from a physics perspective we should restrict the sign of parameters, a Gaussian parameter prior is actually not fully suitable as it always allows for negative numbers. Nevertheless investigating the performance and results of using a Gaussian-inverse-Wishart distribution would indeed be interesting.

Another, very interesting, possible future project is calculating the neutron drip line for the entire chart of nuclides. Albeit an analysis of other priors should come first, investigating the drip line at the current stage could give an indication of its location.

Performing a Bayesian calibration on the DZ33 model is an interesting possibility, especially in determining the relevance of the 33 parameters. Given the large number of parameters it will certainly be able to more correctly determine nuclear binding energies. From a computational perspective one of the problems that arise is the curse of dimensionality. The computational power needed to calculate posterior distributions for models with up to 33 parameters, and to evaluate model and parameter posterior distributions, which requires matrix inversions, is vast. Especially when performing calculations for all $2^{33} - 1 = 8\,589\,934\,591$ (8 billion 589 million 934 thousand 591) models. The time it takes quickly becomes unfeasible, and numerical methods such as *Markov chain Monte Carlo model composition*, $(MC)^3$, could be used to mitigate such problems.[1]

Of course BMA can be applied to more complex, and non-linear models. However, for a non-linear model the posterior distributions may be impossible to evaluate analytically, and instead numerical evaluations are necessary, e.g. *Markov Chain Monte Carlo* (MCMC), or *nested sampling*[16, 29] must be employed.

Appendices

A

DZ10 Mass Model

$$\mathrm{BE}_{\mathrm{DZ10}} = \max\left(\left\langle \mathrm{H}_{\mathrm{monopole}}\right\rangle + \left\langle \mathrm{H}_{\mathrm{spherical}}\right\rangle, \left\langle \mathrm{H}_{\mathrm{monopole}}\right\rangle + \left\langle \mathrm{H}_{\mathrm{deformed}}\right\rangle\right) \tag{A.0.1}$$

$$\begin{aligned}
\left\langle \mathrm{H}_{\mathrm{monopole}}\right\rangle = &- \alpha_1 \frac{Z(Z-1) - 0.76\left[Z(Z-1)\right]^{2/3}}{A^{1/3}\left[1-\left(\frac{T}{A}\right)^2\right]} \\
&+ \frac{\alpha_2}{R}\left\{\left(\sum_p \frac{n_p}{\sqrt{D_p}}\right)^2 + \frac{1}{R}\left(\sum_p \frac{z_p}{\sqrt{D_p}}\right)^2\right. \\
&\quad + \sum_p \left(\frac{n_{jp}}{2} - n_{rp}\right)\left[\left(1+\frac{n_p}{\sqrt{D_p}}\right)\frac{p^2}{D_p^{3/2}} + \left(1-\frac{n_p}{\sqrt{D_p}}\right)\frac{4p-5}{D_p^{3/2}}\right] \\
&\quad \left. + \sum_p \left(\frac{z_{jp}}{2} - z_{rp}\right)\left[\left(1+\frac{z_p}{\sqrt{D_p}}\right)\frac{p^2}{D_p^{3/2}} + \left(1-\frac{z_p}{\sqrt{D_p}}\right)\frac{4p-5}{D_p^{3/2}}\right]\right\} \\
&- \frac{\alpha_3}{R^2}\left[\left(\sum_p \frac{n_p}{\sqrt{D_p}}\right)^2 + \left(\sum_p \frac{z_p}{\sqrt{D_p}}\right)^2\right] \\
&- \frac{\alpha_4}{R}\frac{4T(T+1)}{A^{2/3}R} \\
&+ \frac{\alpha_5}{R^2}\left(\frac{4T(T+1)}{A^{2/3}R^2} - \frac{4T\left(T-\frac{1}{2}\right)}{AR^3}\right) \\
&+ \frac{\alpha_{10}}{R}\begin{cases} 2T/A & N \text{ odd, } Z \text{ odd} \\ 1-2T/A & N \text{ even } > Z \text{ odd} \\ 1 & N \text{ odd } > Z \text{ even} \\ 1 & N \text{ even } < Z \text{ odd} \\ 1-2T/A & N \text{ odd } < Z \text{ even} \\ 2-2T/A & N \text{ even, } Z \text{ even} \end{cases}
\end{aligned} \tag{A.0.2}$$

$$\begin{aligned}\left\langle \mathrm{H}_{\mathrm{spherical}} \right\rangle = &+ \frac{\alpha_6}{R}\left(\frac{n\bar{n}(n-\bar{n})}{D_\nu} + \frac{z\bar{z}(z-\bar{z})}{D_\pi}\right) \\ &- \frac{\alpha_7}{R^2}\left(\frac{n\bar{n}(n-\bar{n})}{D_\nu} + \frac{z\bar{z}(z-\bar{z})}{D_\pi}\right) \\ &+ \frac{\alpha_8}{R}\frac{2^{\sqrt{p_\nu}}n\bar{n}}{D_\nu}\cdot\frac{2^{\sqrt{p_\pi}}n\bar{n}}{D_\pi}\end{aligned} \tag{A.0.3}$$

$$\left\langle \mathrm{H}_{\mathrm{deformed}} \right\rangle = +\frac{\alpha_9}{R}\frac{n'\bar{n}'}{D_\nu^{3/2}}\cdot\frac{z'\bar{z}'}{D_\pi^{3/2}} \tag{A.0.4}$$

The numbering of parameters is to conform with those of Duflo & Zuker in the original paper as well as their code. [3, 4, 30]

B

DZ10 Code

```
module dz_models

  contains
    !subroutine duflo(nn,nz,exc)
    !c-----------------------------------------------------------------------c
    !c       File "du_zu_10.feb96"                                           c
    !c                                                                       c
    !c       J. Duflo and A.P. Zuker  Feb 23, 1996  10 parameters formula    c
    !c       Reference:                                                      c
    !c             Phys. Rev. C52, 1995  (for the 28 param. formula)         c
    !c             and Private Communication to AMDC, February 1996.         c
    !c                                                                       c
    !c       Microscopic calculation of nuclear masses with 10               c
    !c       parameters. Fit to the 1810 measured masses from Helium-4       c
    !c       and up with a root-mean-square deviation (rms) of 506 keV.      c
    !c-----------------------------------------------------------------------c

    subroutine dz10(b,nx,nz,E)      ! Duflo-Zuker fevrier 1996

      implicit none

      double precision  :: a,t,r,rc,ra,e
      integer :: nn(2),nx,nz,z2,n2(2),ncum
      integer :: ndef,ju,kk,j,l,noc(18,2),i2,imax
      integer :: ip,ipm,mss,pi,k,i,id
      double precision :: px,vm,de,den,fact,moc

      !c Calculation of binding energy E (nx neutrons,nz protons)
      double precision:: b(10),dyda(10),y(2),onp(0:8,2,2) &
           ,op(2),dx(2),qx(2),os(2),oei(2),dei(2),pp(2)
      ! data b/0.7043,17.7418,16.2562,37.5562,53.9017,0.4711,2.1307, &
```

```
!       0.0210,40.5356,6.0632/
!c*********
nn(1)=nx
nn(2)=nz
a=nx+nz
t=abs(nx-nz)
r=a**(1./3.)
rc=r*(1.-.25*(t/a)**2)          !       Charge radius
ra=(rc*rc)/r
!c--------
z2=nz*(nz-1)
dyda(1)=(-z2+.76*z2**(2./3.))/rc   ! Coulomb energy
!c********                        !  beginning of main loop
do ndef=1,2                      !       ndef=1  spherical
   ju=0                             !       ndef=2  deformed
   y(ndef)=0.
   if(ndef.eq.2) ju=4               !       nucleons associated to deform.
   do kk=2,10
      dyda(kk)=0.
   enddo
   !c--------                       !  beginning of loop over N and Z
   do j=1,2
      do l=1,18
         noc(l,j)=0
      enddo
      do l=1,2
         do k=0,8
            onp(k,l,j)=0.
         enddo
      enddo
      n2(j)=2*(nn(j)/2)             !       (for pairing calculation)
      ncum=0
      i=0
      !c--------
20    i=i+1                         !    sub-shells (ssh) j and r filling
      i2=(i/2)*2
      if(i2.ne.i)then
         id=i+1                     !             for ssh j
      else
         id=i*(i-2)/4               !             for ssc r
```

```
endif
ncum=ncum+id
if(ncum.lt.nn(j))then
    noc(i,j)=id                !      nb of nucleons in each ssh
    go to 20
endif
!c--------
imax=i+1                       !      imax = last subshell nb
ip=(i-1)/2                     !      HO number (p)
ipm=i/2
pp(j)=ip
moc=nn(j)-ncum+id
noc(i,j)=moc-ju                !      nb of nucleons in last ssh
noc(i+1,j)=ju
if(i2.ne.i)then                !      ssh j
    oei(j)=moc+ip*(ip-1)       !      nb of nucleons in last EI shell
    dei(j)=ip*(ip+1)+2         !         size of the EI shell
else                           !      ssh r
    oei(j)=moc-ju              !      nb of nucleons in last EI shell
    dei(j)=(ip+1)*(ip+2)+2     !         size of the EI shell
endif
qx(j)=oei(j)*(dei(j)-oei(j)-ju)/dei(j)   ! n*(D-n)/D          S3(j)
dx(j)=qx(j)*(2*oei(j)-dei(j))            ! n*(D-n)*(2n-D)/D   Q
if(ndef.eq.2)qx(j)=qx(j)/sqrt(dei(j))    ! scaling for deformed
!c--------
do i=1,imax                              ! Amplitudes
    ip=(i-1)/2
    fact=sqrt((ip+1.)*(ip+2.))
    onp(ip,1,j)=onp(ip,1,j)+noc(i,j)/fact !     for FM term
    vm=-1.
    if((2*(i/2)).ne.i)vm=.5*ip           !      for spin-orbit term
    onp(ip,2,j)=onp(ip,2,j)+noc(i,j)*vm
enddo
!c--------
op(j)=0.
os(j)=0.
do ip=0,ipm                    !        FM and SO terms
    pi=ip
    den=((pi+1)*(pi+2))**(3./2.)
    op(j)=op(j)+onp(ip,1,j)                                  ! FM
```

```
            os(j)=os(j)+onp(ip,2,j)*(1.+onp(ip,1,j))*(pi*pi/den) &   ! SO
                +onp(ip,2,j)*(1.-onp(ip,1,j))*((4*pi-5)/den)
        enddo
        op(j)=op(j)*op(j)
    enddo
    !c--------                                  ! end of loop over N and Z
    dyda(2)=op(1)+op(2)                        !   Master term (FM): volume
    dyda(3)=-dyda(2)/ra                        !                     surface
    dyda(2)=dyda(2)+os(1)+os(2)                !   FM + SO
    dyda(4)=-t*(t+2)/(r*r)                     !   isospin term : volume
    dyda(5)=-dyda(4)/ra                        !                : surface
    if(ndef.eq.1)then                          ! sph.
        dyda(6)=dx(1)+dx(2)                     !   S3 volume
        dyda(7)=-dyda(6)/ra                     !      surface
        px=sqrt(pp(1))+sqrt(pp(2))
        dyda(8)=qx(1)*qx(2)*(2**px)             !   QQ sph.
    else                                       ! def.
        dyda(9)=qx(1)*qx(2)                     !   QQ deform.
    endif
    dyda(5)=t*(1-t)/(a*ra**3)+dyda(5)          !   "Wigner term"
    !c--------                                     !   PAIRING
    if(n2(1).ne.nn(1).and.n2(2).ne.nn(2))dyda(10)= t/a
    if(nx.gt.nz)then
        if(n2(1).eq.nn(1).and.n2(2).ne.nn(2))dyda(10)= 1-t/a
        if(n2(1).ne.nn(1).and.n2(2).eq.nn(2))dyda(10)= 1
    else
        if(n2(1).eq.nn(1).and.n2(2).ne.nn(2))dyda(10)= 1
        if(n2(1).ne.nn(1).and.n2(2).eq.nn(2))dyda(10)= 1-t/a
    endif
    if(n2(2).eq.nn(2).and.n2(1).eq.nn(1))dyda(10)= 2-t/a
    !c--------
    do mss=2,10
        dyda(mss)=dyda(mss)/ra
    enddo
    do mss=1,10
        y(ndef)=y(ndef)+dyda(mss)*b(mss)
    enddo
    !c--------                                  ! end of main loop
enddo
de=y(2)-y(1)
```

```
    E=y(2)                              ! Binding Energy for def. nuclides
    if(de.le.0..or.nz.le.50)E=y(1)      !                  spherical nuclides

    return
  end subroutine dz10
```

Bibliography

[1] Jennifer A. Hoeting et al. 'Bayesian model averaging: a tutorial (with comments by M. Clyde, David Draper and E. I. George, and a rejoinder by the authors'. In: *Statist. Sci.* 14.4 (Nov. 1999), pp. 382–417. DOI: 10.1214/ss/1009212519. URL: https://doi.org/10.1214/ss/1009212519.

[2] Larry Wasserman. 'Bayesian Model Selection and Model Averaging'. In: *Journal of Mathematical Psychology* 44.1 (2000), pp. 92–107. ISSN: 0022-2496. DOI: https://doi.org/10.1006/jmps.1999.1278. URL: http://www.sciencedirect.com/science/article/pii/S0022249699912786.

[3] J. Duflo and A.P. Zuker. 'Microscopic mass formulas'. In: *Physical Review C* 52.1 (July 1995), R23–R27. ISSN: 1089-490X. DOI: 10.1103/physrevc.52.r23. URL: http://dx.doi.org/10.1103/PhysRevC.52.R23.

[4] Chong Qi. *A short explanation of the Duflo-Zuker mass model*. 2014. arXiv: 1407.8218 [nucl-th].

[5] Petr Navrátil et al. 'Unified ab initio approaches to nuclear structure and reactions'. In: *Physica Scripta* 91.5 (Apr. 2016), p. 053002. DOI: 10.1088/0031-8949/91/5/053002. URL: https://doi.org/10.1088%5C%2F0031-8949%5C%2F91%5C%2F5%5C%2F053002.

[6] Jochen Erler et al. 'The limits of the nuclear landscape'. In: *Nature* 486 (7404 2012), pp. 509–512. DOI: 10.1038/nature11188. URL: https://doi.org/10.1038/nature11188.

[7] Rui Wang and Lie-Wen Chen. 'Positioning the neutron drip line and the r-process paths in the nuclear landscape'. In: *Phys. Rev. C* 92 (3 Sept. 2015), p. 031303. DOI: 10.1103/PhysRevC.92.031303. URL: https://link.aps.org/doi/10.1103/PhysRevC.92.031303.

[8] Léo Neufcourt et al. 'Neutron Drip Line in the Ca Region from Bayesian Model Averaging'. In: *Phys. Rev. Lett.* 122 (6 Feb. 2019), p. 062502. DOI: 10.1103/PhysRevLett.122.062502. URL: https://link.aps.org/doi/10.1103/PhysRevLett.122.062502.

[9] A. Pastore et al. 'Pairing in exotic neutron-rich nuclei near the drip line and in the crust of neutron stars'. In: *Phys. Rev. C* 88 (3 Sept. 2013), p. 034314. DOI: 10.1103/PhysRevC.88.034314. URL: https://link.aps.org/doi/10.1103/PhysRevC.88.034314.

[10] R. Utama and J. Piekarewicz. 'Refining mass formulas for astrophysical applications: A Bayesian neural network approach'. In: *Phys. Rev. C* 96 (4 Oct. 2017), p. 044308. DOI: 10.1103/PhysRevC.96.044308. URL: https://link.aps.org/doi/10.1103/PhysRevC.96.044308.

[11] A. H. Wapstra and G. Audi. 'The 1983 atomic mass evaluation. (I). Atomic mass table'. In: *Nuclear Physics, Section A* 432.1 (1985), pp. 1–54. ISSN: 03759474. DOI: 10.1016/0375-9474(85)90283-0.

[12] A. H. Wapstra and G. Audi. 'The 1983 atomic mass evaluation. (II). Nuclear-reaction and separation energies'. In: *Nuclear Physics, Section A* 432.1 (1985), pp. 55–139. ISSN: 03759474. DOI: 10.1016/0375-9474(85)90284-2.

[13] W. J. Huang et al. 'The AME2016 atomic mass evaluation (I). Evaluation of input data; And adjustment procedures'. In: *Chinese Physics C* 41.3 (2017), pp. 1–344. ISSN: 16741137. DOI: 10.1088/1674-1137/41/3/030002.

[14] Meng Wang et al. 'The AME2016 atomic mass evaluation (II). Tables, graphs and references'. In: *Chinese Physics C* 41.3 (2017). ISSN: 16741137. DOI: 10.1088/1674-1137/41/3/030003.

[15] R. T. Cox. 'Probability, Frequency and Reasonable Expectation'. In: *American Journal of Physics* 14.1 (1946), pp. 1–13. DOI: 10.1119/1.1990764. eprint: https://doi.org/10.1119/1.1990764. URL: https://doi.org/10.1119/1.1990764.

[16] Devinder S. Sivia and J. Skilling. *Data Analysis: A Bayesian Tutorial, 2e.* Oxford University Press, 2006. ISBN: 978-0-19-856832-2.

[17] Phil C. Gregory. *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with Mathematica® Support.* Cambridge University Press, 2005. ISBN: 978-0-521-15012-5.

[18] E. T. Jaynes and G. Larry Bretthorst. *Probability Theory : The Logic of Science.* Cambridge University Press, 2003. ISBN: 9781139146722. URL: http://ebookcentral.proquest.com/lib/chalmers/detail.action?docID=217750.

[19] Christopher M. Bishop. *Pattern recognition and machine learning.* Information science and statistics. Springer, 2006. ISBN: 978-1-4939-3843-8.

[20] Feng Liang et al. 'Mixtures of g Priors for Bayesian Variable Selection'. In: *Journal of the American Statistical Association* 103.481 (Mar. 2008). dummy note, pp. 410–423. DOI: 10.1198/016214507000001337.

[21] Vojtech Kejzlar et al. *Bayesian averaging of computer models with domain discrepancies: a nuclear physics perspective.* 2019. arXiv: 1904.04793 [stat.ME].

[22] Vojtech Kejzlar et al. *Statistical aspects of nuclear mass models.* 2020. arXiv: 2002.04151 [nucl-th].

[23] Marc C. Kennedy and Anthony O'Hagan. 'Bayesian calibration of computer models.' In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63.3 (2001), p. 425. ISSN: 13697412.

[24] Jenný Brynjarsdóttir and Anthony O'Hagan. 'Learning about physical parameters: the importance of model discrepancy'. In: *Inverse Problems* 30.11 (Oct. 2014), p. 114007. DOI: 10.1088/0266-5611/30/11/114007. URL: https://doi.org/10.1088%5C%2F0266-5611%5C%2F30%5C%2F11%5C%2F114007.

[25] Edward I. George. 'Dilution priors: Compensating for model space redundancy'. In: *Borrowing Strength: Theory Powering Applications – A Festschrift for Lawrence D. Brown*. Ed. by James O. Berger, T. Tony Cai and Iain M. Johnstone. Vol. Volume 6. Collections. Beachwood, Ohio, USA: Institute of Mathematical Statistics, 2010, pp. 158–165. DOI: 10.1214/10-IMSCOLL611. URL: https://doi.org/10.1214/10-IMSCOLL611.

[26] A. Abzouzi, E. Caurier and A. P. Zuker. 'Influence of saturation properties on shell-model calculations'. In: *Phys. Rev. Lett.* 66 (9 Mar. 1991), pp. 1134–1137. DOI: 10.1103/PhysRevLett.66.1134. URL: https://link.aps.org/doi/10.1103/PhysRevLett.66.1134.

[27] Andrés P. Zuker. 'On the microscopic derivation of a mass formula'. In: *Nuclear Physics A* 576.1 (1994), pp. 65–108. ISSN: 0375-9474. DOI: https://doi.org/10.1016/0375-9474(94)90738-2. URL: http://www.sciencedirect.com/science/article/pii/0375947494907382.

[28] J. Duflo and A. P. Zuker. 'The nuclear monopole Hamiltonian'. In: *Phys. Rev. C* 59 (5 May 1999), R2347–R2350. DOI: 10.1103/PhysRevC.59.R2347. URL: https://link.aps.org/doi/10.1103/PhysRevC.59.R2347.

[29] John Skilling. 'Nested Sampling'. In: *AIP Conference Proceedings* 735.November 2004 (2004), pp. 395–405. DOI: 10.1063/1.1835238. URL: https://aip.scitation.org/doi/abs/10.1063/1.1835238.

[30] *Duflo-Zuker mass model Fortran code*. Calculates the binding energies using the mass formulas of Duflo & Zuker (31 parameter and 10 parameter versions) as described in J. Duflo and A.P. Zuker , Phys. Rev. C 52 (1995) R23. Downloadable from https://www-nds.iaea.org/amdc/Duflo-Zuker-program.zip. URL: https://www-nds.iaea.org/amdc/.

DEPARTMENT OF PHYSICS
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY